Expressive Power of Graphical Models

Michael Gutmann

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

Spring Semester 2020

Recap

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

Program

1. Independency maps (I-maps)

2. Equivalence of I-maps (I-equivalence)

3. Minimal I-maps

4. (Lossy) conversion between directed and undirected I-maps

Program

1. Independency maps (I-maps)
    - Definition of I-maps and perfect maps
    - I-maps and factorisation
    - Examples and no guarantee for perfect maps
2. Equivalence of I-maps (I-equivalence)
3. Minimal I-maps
4. (Lossy) conversion between directed and undirected I-maps

I-map

- We have seen that graphs represent independencies. We say that they are independency maps (I-maps).
- *Definition*: Let $\mathcal{U}$ be a set of independencies that random variables $\mathbf{x} = (x_1, \ldots x_d)$ satisfy. A DAG or undirected graph K with nodes x_i is said to be an independency map (I-map) for $\mathcal{U}$ if the independencies $\mathcal{I}(K)$ asserted by the graph are part of $\mathcal{U}$:
$$\mathcal{I}(K) \subseteq \mathcal{U}$$
- *Definition*: K is said to be a perfect I-map (or P-map) if $\mathcal{I}(K) = \mathcal{U}$.
- A I-map is a "directed I-map" if K is a DAG, and an "undirected I-map" if K is an undirected graph.

I-map

The set of independencies $\mathcal{U}$ can be specified in different ways. For example:

- as a list of independencies, e.g.

$$\mathcal{U} = \{x_1 \perp\!\!\!\perp x_2\}$$

- as the independencies implied by a graph K_0

$$\mathcal{U} = \mathcal{I}(K_0)$$

- denoting by $\mathcal{I}(p)$ all the independencies satisfied by a specific distribution p, we can have

$$\mathcal{U} = \mathcal{I}(p)$$

I-maps and factorisation

- Assume p factorises over a DAG or undirected graph K, i.e $p(\mathbf{x})$ can be written as

$$p(\mathbf{x}) = \prod_i p(x_i|\mathrm{pa}_i) \qquad \text{or} \qquad p(\mathbf{x}) \propto \prod_c \phi_c(\mathcal{X}_c)$$

- We have previously found that all independencies asserted by the graph K hold for p.
- This means that

$$\mathcal{I}(K) \subseteq \mathcal{I}(p)$$

  and K is an I-map for $\mathcal{I}(p)$
- But K is not guaranteed to be a perfect map for $\mathcal{I}(p)$ since, as we have seen, $\mathcal{I}(K)$ may miss some independencies that hold for p.

Perfect maps and factorisation

For what set $\mathcal{U}$ of independencies is a graph K a perfect map?

- Let K be a DAG or an undirected graph. We have seen that: *if* X are Y and not (d-)separated by Z *then* $X \not\perp\!\!\!\perp Y|Z$ for some p that factorises over K (some $\equiv$ not all)
- Contrapositive: (Reminder: $A \Rightarrow B \Leftrightarrow \bar{B} \Rightarrow \bar{A}$)

  if $X \perp\!\!\!\perp Y|Z$ for all p that factorise over K *then* X and Y are (d-)separated by Z
- Denote by $\mathcal{P}_K$ the set of all p that factorise over K. We thus have:

$$\left[\bigcap_{p \in \mathcal{P}_K} \mathcal{I}(p)\right] \subseteq \mathcal{I}(K)$$

Perfect maps and factorisation

For what set $\mathcal{U}$ of independencies is a graph K a perfect map?

- Since for individual p we have $\mathcal{I}(K) \subseteq \mathcal{I}(p)$, this means that

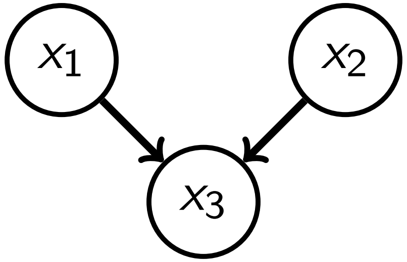
$$\mathcal{I}(K) = \bigcap_{p \in \mathcal{P}_K} \mathcal{I}(p)$$

- In plain English: K is a perfect map for the independencies that hold for all p that factorise over the graph.

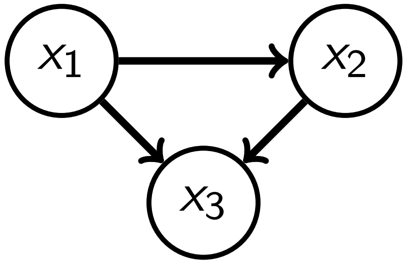
Independencies with a directed but w/o undirected P-map

For $\mathbf{x} = (x_1, x_2, x_3)$, consider $\mathcal{U} = \{x_1 \perp\!\!\!\perp x_2\}$

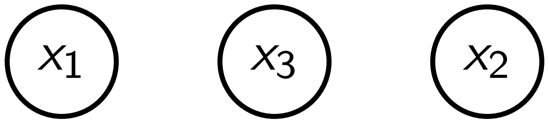
- Perfect I-map: $\mathcal{I}(G) = \mathcal{U}$



- I-map: $\mathcal{I}(G) = \{\}$



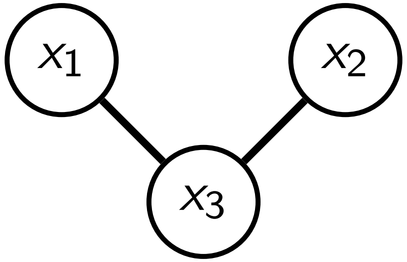
- Not an I-map: graph e.g. wrongly asserts $x_2 \perp\!\!\!\perp x_3$



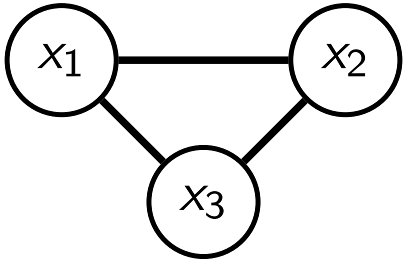
Independencies with a directed but w/o undirected P-map

For $\mathbf{x} = (x_1, x_2, x_3)$, consider $\mathcal{U} = \{x_1 \perp\!\!\!\perp x_2\}$

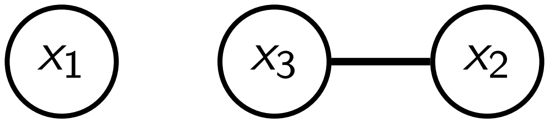
- Not an I-map: graph wrongly asserts $x_1 \perp\!\!\!\perp x_2 \mid x_3$



- I-map: $\mathcal{I}(H) = \{\}$



- Not an I-map: graph e.g. wrongly asserts $x_1 \perp\!\!\!\perp x_3$

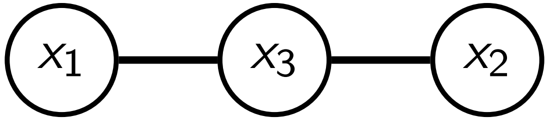


- Going through all undirected graphs shows that there is no undirected perfect I-map for $\mathcal{U}$.

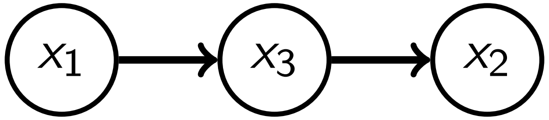
Independencies with multiple equivalent I-maps

Consider now $\mathcal{U} = \{x_1 \perp\!\!\!\perp x_2,\ x_1 \perp\!\!\!\perp x_2 | x_3,\ x_2 \perp\!\!\!\perp x_3,\ x_2 \perp\!\!\!\perp x_3 | x_1\}$

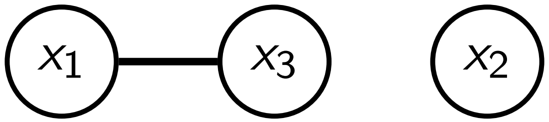
- I-map: $\mathcal{I}(H) = \{x_1 \perp\!\!\!\perp x_2 | x_3\} \subset \mathcal{U}$



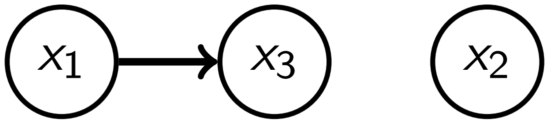
- I-map: $\mathcal{I}(G) = \{x_1 \perp\!\!\!\perp x_2 | x_3\} \subset \mathcal{U}$



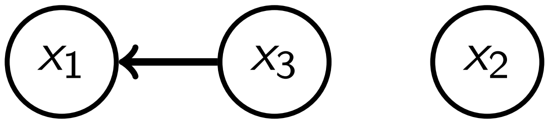
- Perfect I-map: $\mathcal{I}(H) = \mathcal{U}$



- Perfect I-map: $\mathcal{I}(G) = \mathcal{U}$



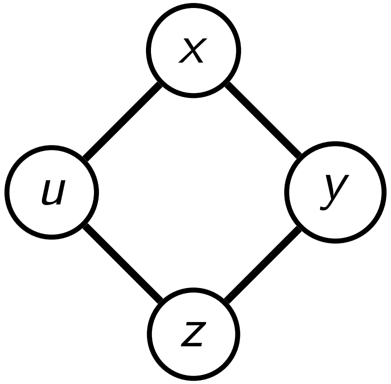
- Perfect I-map: $\mathcal{I}(G) = \mathcal{U}$



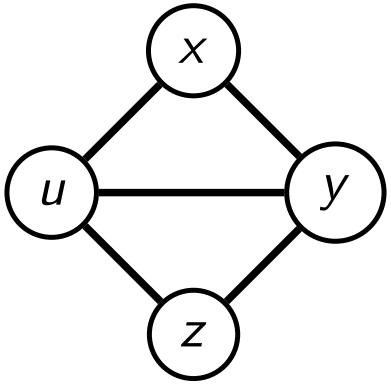
Independencies with undirected but w/o directed P-map

For random variables (x, y, z, u), $\mathcal{U} = \{x \perp\!\!\!\perp z | u, y,\; u \perp\!\!\!\perp y | x, z\}$

- Perfect map: $\mathcal{I}(H) = \mathcal{U}$



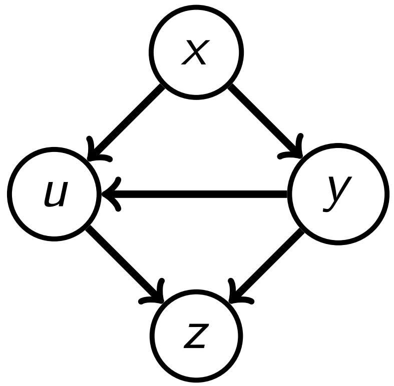
- I-map: $\mathcal{I}(H) = \{x \perp\!\!\!\perp z | u, y\} \subset \mathcal{U}$



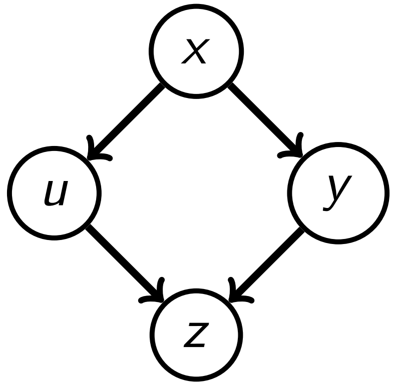
Independencies with undirected but w/o directed P-map

For random variables (x, y, z, u), $\mathcal{U} = \{x \perp\!\!\!\perp z | u, y, \; u \perp\!\!\!\perp y | x, z\}$

- I-map: $\mathcal{I}(G) = \{x \perp\!\!\!\perp z | u, y\} \subset \mathcal{U}$



- Not an I-map: graph wrongly asserts $u \perp\!\!\!\perp y | x$



- Going through all DAGs shows that there is no directed perfect I-map for $\mathcal{U}$.

Remarks

The examples illustrate a number of important points:

- Multiple graphs may make the same independency assertions.
  $\Rightarrow$ I-equivalency: When do we have $\mathcal{I}(K_1) = \mathcal{I}(K_2)$?
- The fully connected graph is always an I-map.
  $\Rightarrow$ Minimal I-maps: sparsest graph that is still an I-map?
- Perfect maps may not exist, and some independencies are better represented with directed than with undirected graphs, and vice versa.
  $\Rightarrow$ Pros/cons of directed and undirected graphs and conversion between them?

Program

1. Independency maps (I-maps)

2. Equivalence of I-maps (I-equivalence)
   - I-equivalence for DAGs: check the skeletons and the immoralities
   - I-equivalence for undirected graphs: check the skeletons

3. Minimal I-maps

4. (Lossy) conversion between directed and undirected I-maps

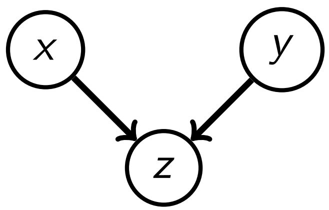
I-equivalence for DAGs

- How do we determine whether two DAGs 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

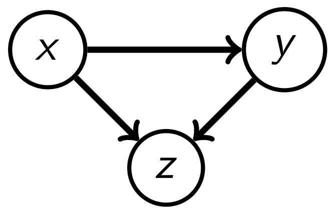
Connection	$p(x, y)$	$p(x, y \mid z)$
$x \rightarrow z \rightarrow y$	$x \not\perp\!\!\!\perp y$	$x \perp\!\!\!\perp y \mid z$
$x \leftarrow z \rightarrow y$	$x \not\perp\!\!\!\perp y$	$x \perp\!\!\!\perp y \mid z$
$x \rightarrow z \leftarrow y$	$x \perp\!\!\!\perp y$	$x \not\perp\!\!\!\perp y \mid z$

I-equivalence for DAGs

- The situation $x \perp\!\!\!\perp y$ and $x \not\perp\!\!\!\perp y \mid z$ can only happen if we have colliders without "covering edge" $x \rightarrow y$ or $x \leftarrow y$, that is when parents of the collider node are not directly connected.
- Colliders without covering edge are called "immoralities"
- Theorem: For two DAGs G_1 and G_2:
  $G1$ and G_2 are I-equivalent $\iff$ G_1 and G_2 have the same skeleton and the same set of immoralities.
  (for a proof, see e.g. Theorem 4.4, Koski and Noble, 2009; not examinable)



$x \perp\!\!\!\perp y$ and $x \not\perp\!\!\!\perp y \mid z$
Collider w/o covering edge

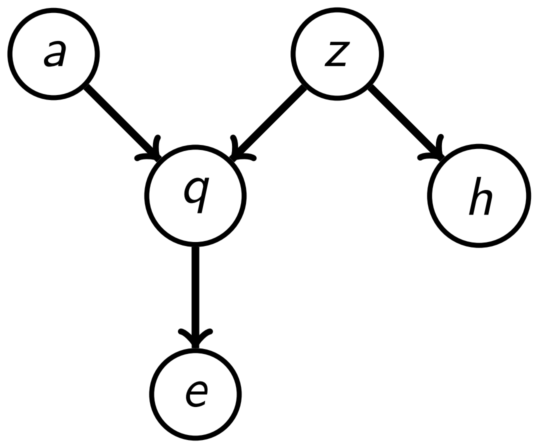


$x \not\perp\!\!\!\perp y$ and $x \not\perp\!\!\!\perp y \mid z$
Collider with covering edge

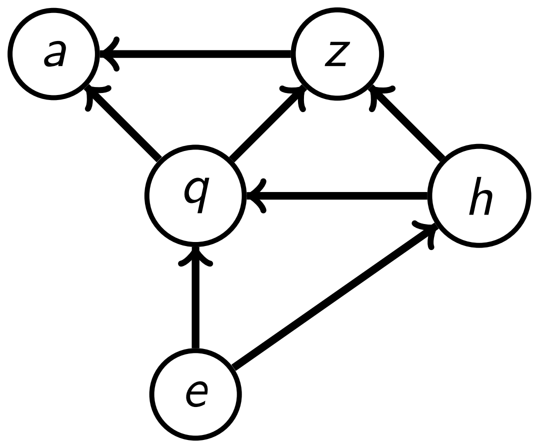
Example

Not I-equivalent because of skeleton mismatch:

G_1:



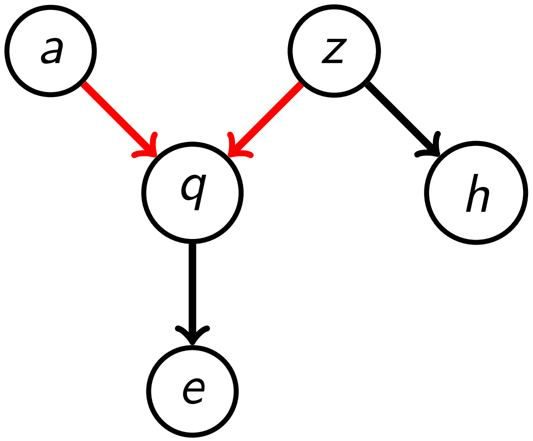
G_2:



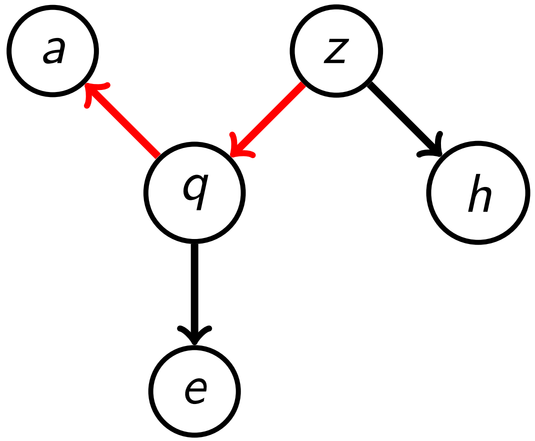
Example

Not I-equivalent because of immoralities mismatch:

G_1:



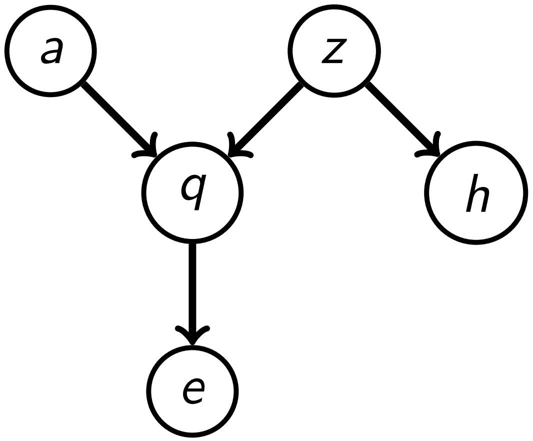
G_2:



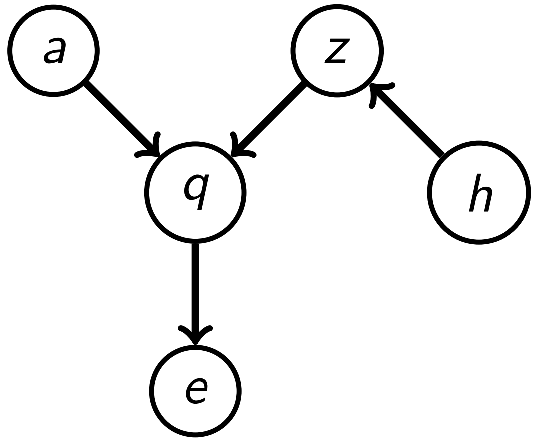
Example

I-equivalent (same skeleton, same immoralities):

G_1:

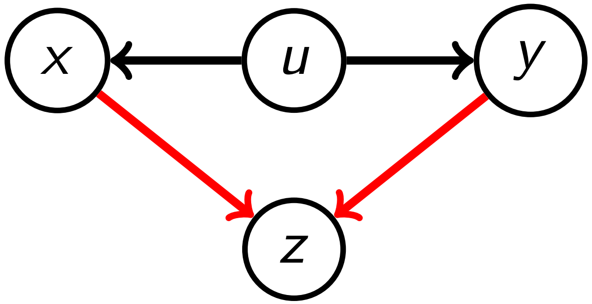


G_2:



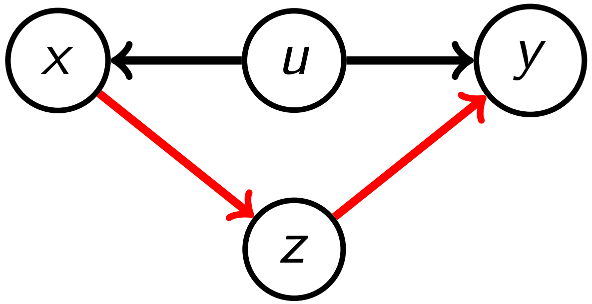
Example

Not I-equivalent (immoralities mismatch)



$x \perp\!\!\!\perp y \mid u$ and $x \not\perp\!\!\!\perp y \mid u, z$

Immorality: collider w/o covering edge

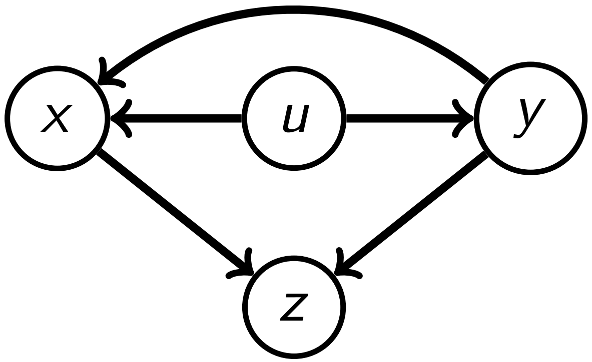


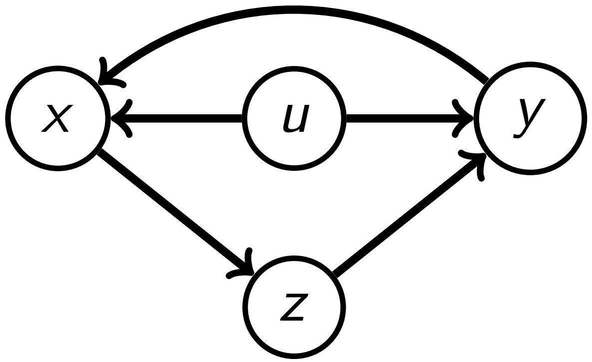
$x \not\perp\!\!\!\perp y \mid u$ and $x \perp\!\!\!\perp y \mid u, z$

Not an immorality

Example

I-equivalent (same skeleton, no immoralities)





I-equivalence for undirected graphs

- Different undirected graphs make different independence assertions.
- I-equivalent if their skeleton is the same.

Program

1. Independency maps (I-maps)
2. Equivalence of I-maps (I-equivalence)
3. Minimal I-maps
    - Definition
    - Construction of undirected minimal I-maps
    - Construction of directed minimal I-maps
4. (Lossy) conversion between directed and undirected I-maps

Minimal I-maps

- Criterion for an I-map is that the independency assertions made by the graph are true. I-maps are not concerned with the number of independency assertions made.
- I-maps of $\mathcal{U}$ may not represent ("miss") some independencies in $\mathcal{U}$.
- Full graph does not make any assertions. Empty set is trivially a subset of any $\mathcal{U}$, so that the full graph is trivially an I-map.
- *Definition*: A minimal I-map is an I-map such that if you remove an edge (more independencies), the resulting graph is not an I-map any more.

Minimal I-maps

- Intuitively, the point of minimal I-maps is to "sparsify" I-maps so that they become more useful.
  (while sparser, the independence assertions made must still be correct!)
- Sparser I-maps are more informative, easier to interpret, and they facilitate learning and inference.
- Note: A perfect map for $\mathcal{U}$ is also a minimal I-map for $\mathcal{U}$
  (being perfect is a stronger requirement than being minimal)

Constructing minimal I-maps

- If we know the factorisation of p, we can visualise p as a DAG or an undirected graph K. Since p factorises over the constructed K, K is an I-map for $\mathcal{I}(p)$ but not necessarily a minimal I-map (see before).
- We have seen that K is a perfect map for the independencies that hold for all p with a particular factorisation, but not necessarily for all the independencies that hold for the specific p.
- There are some p, for which a perfect map for $\mathcal{I}(p)$ does not exist. (see tutorial 3 for an example)
- We thus settle for obtaining minimal I-maps for $\mathcal{I}(p)$.

Constructing undirected minimal I-maps

For d random variables $\mathbf{x}$ with positive distribution $p > 0$, assume we can test whether an independency is in $\mathcal{I}(p)$, i.e. holds for p.

- Approaches based on pairwise and local Markov property
- Both yield same (unique) graph.
- For local Markov property approach: For each variable x_i:
  1. determine its Markov blanket $\mathrm{MB}(x_i)$, i.e. find minimal set of variables U such that
  $$x_i \perp\!\!\!\perp \{\text{all variables} \setminus (x_i \cup U)\} \mid U$$
  is in $\mathcal{I}(p)$
  2. we know that x_i and $\mathrm{MB}(x_i)$ must be neighbours in the graph: Connect x_i to all variables in $\mathrm{MB}(x_i)$
- We need $p > 0$ because otherwise local independencies may not imply global ones (see slides on undirected graphical models).

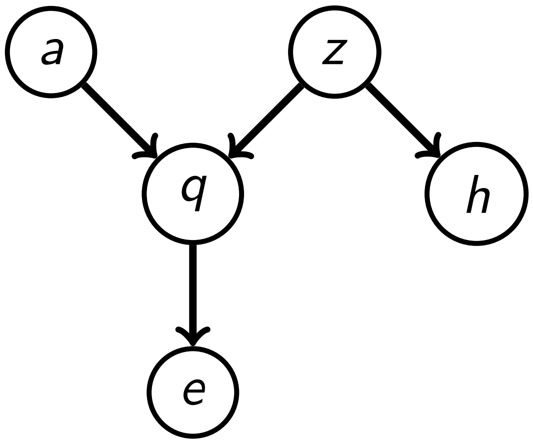
Constructing directed minimal I-maps

For d random variables $\mathbf{x}$ with distribution p, assume we can test whether an independency is in $\mathcal{I}(p)$, i.e. holds for p.

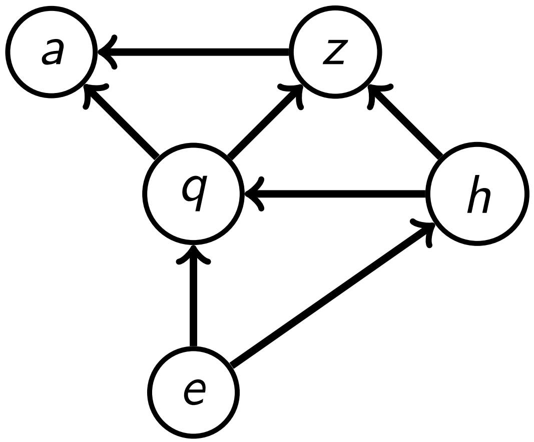
- We can use the ordered Markov property to derive a directed graph that is a minimal I-map for $\mathcal{I}(p)$.
- Procedure is exactly the same as the one used to simplify the factorisation obtained by the chain rule:
  1. Assume an ordering of the variables. Denote the ordered random variables by $x_1, \ldots, x_d$.
  2. For each i, find a minimal subset of variables $\pi_i \subseteq \mathrm{pre}_i$ such that
  $$x_i \perp\!\!\!\perp \{\mathrm{pre}_i \setminus \pi_i\} \mid \pi_i$$
  is in $\mathcal{I}(p)$.
  3. Construct a graph with parents $\mathrm{pa}_i = \pi_i$.

Directed minimal I-maps are not unique

Consider p with perfect (and hence minimal) I-map G^*



Graph G^*



Minimal I-map for ordering (e, h, q, z, a), see tutorials

- Directed (minimal) I-maps are not unique
- The minimal directed I-maps obtained with different orderings are not I-equivalent.

Program

1. Independency maps (I-maps)
2. Equivalence of I-maps (I-equivalence)
3. Minimal I-maps
4. (Lossy) conversion between directed and undirected I-maps
    - Moralisation for directed $\rightarrow$ undirected
    - Triangulation for undirected $\rightarrow$ directed
    - Strengths and weaknesses of directed and undirected graphs

Directed to undirected graphical model

Goal: given a DAG G, find an undirected minimal I-map for $\mathcal{I}(G)$.

- Probabilistic models factorises according to G as

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

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

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

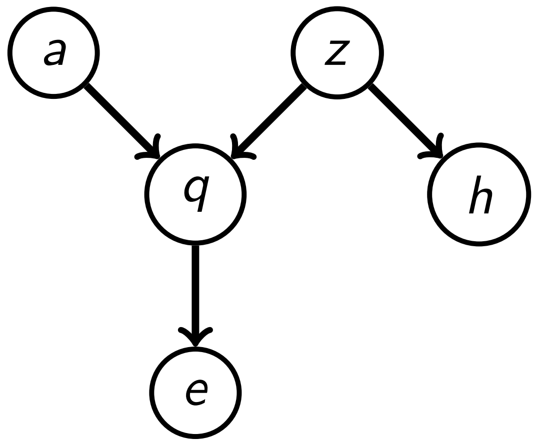
  Gibbs distribution with normalisation constant equal to one
- Visualise p as an undirected graph: form cliques for (x_i, pa_i)

Directed to undirected graphical model

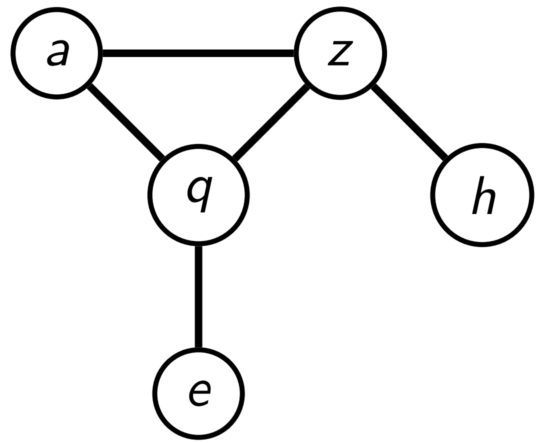
- Visualise p as an undirected graph: form cliques for (x_i, pa_i)
  $\Rightarrow$ Remove arrows, and add edges between *all* parents of x_i.
- Conversion from directed to undirected graphical model is called "moralisation" because it removes immoralities that may exist in the DAG G. Obtained undirected graph is the "moral graph" $\mathcal{M}(G)$ of G.
- Process above is equivalent to constructing the undirected minimal I-map as on slide 29 when the directed graph is used to determine the required Markov blankets.

Example

Given: directed graph G



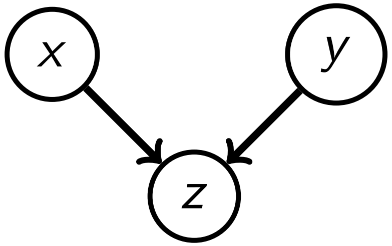
Moral graph $H = \mathcal{M}(G)$:



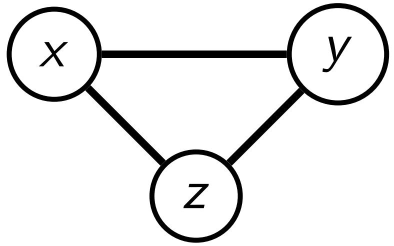
Note: We have $\mathcal{I}(H) \subset \mathcal{I}(G)$. The independency $a \perp\!\!\!\perp z \notin \mathcal{I}(H)$.
We lost that information.

Canonical example

Given: directed graph G



Moral graph $H = \mathcal{M}(G)$:



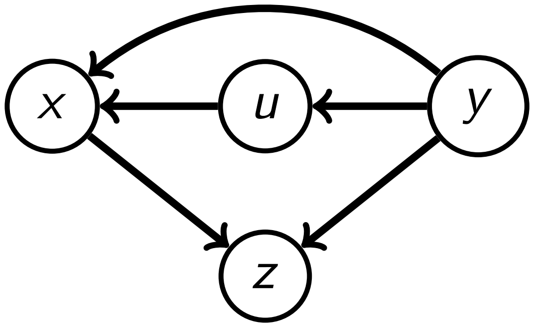
- The fully connected graph is the only minimal undirected I-map for $\mathcal{I}(G)$.
- We lost information: $\mathcal{I}(H) \subset \mathcal{I}(G)$. The independency $x \perp\!\!\!\perp y \notin \mathcal{I}(H)$. See before: there is no undirected P-map for $\mathcal{I}(G)$.
- Loss of information is due to presence of the immorality in G.

Lossless conversion for DAGs without immoralities

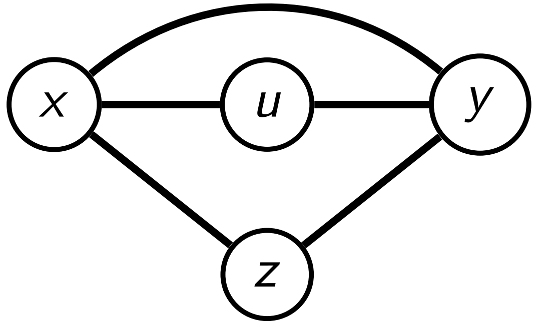
- Immoralities allow DAGs to represent independencies that cannot be represented with undirected graphs (e.g. $x \perp\!\!\!\perp y$ without enforcing $x \perp\!\!\!\perp y | z$ in the example above)
- We loose these kind of independencies when moralising a DAG.
- For a DAG G without immoralities, moralisation does not lead to a loss of information: $\mathcal{M}(G)$ is an undirected perfect map for $\mathcal{I}(G)$. (for a proof, see Section 4.5.1 in Koller and Friedman, 2009, not examinable)
- Other way to understand this result: for DAGs without immoralities, only the skeleton is relevant for I-equivalence. Since the orientation of the arrows does not matter, we can just drop them to obtain an I-equivalent undirected graph.

Example

Given: directed graph G:



Moral graph $H = \mathcal{M}(G)$



- We have $\mathcal{I}(H) = \mathcal{I}(G) = \{u \perp\!\!\!\perp z | x, y\}$.
- H is a perfect map for $\mathcal{I}(G)$.
- H and G are I-equivalent.

Undirected to directed graphical model

Goal: given an undirected graph H, find a directed minimal I-map for $\mathcal{I}(H)$.

- We can construct the directed minimal I-map with the procedure on slide 30 but use H to determine the required independencies: Instead of checking that

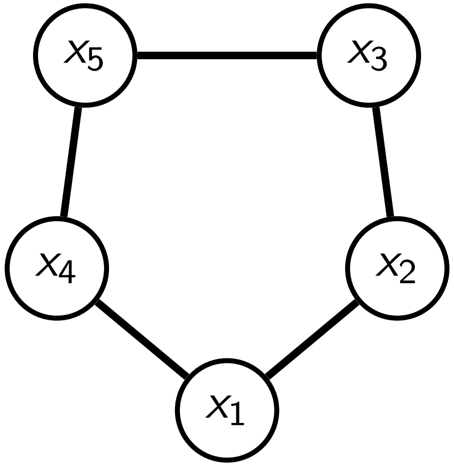
$$x_i \perp\!\!\!\perp \{\text{pre}_i \setminus \pi_i\} \mid \pi_i$$

is in $\mathcal{I}(p)$, we check whether it is in $\mathcal{I}(H)$.

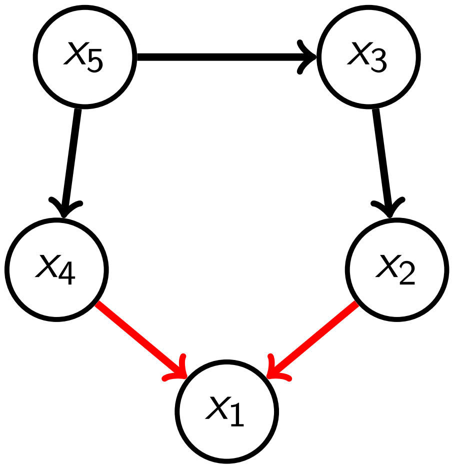
- Directed minimal I-map will not have any immoralities.
(for a proof, see e.g. Theorem 4.10 in Koller and Friedman's book; not examinable)
- Results in chordal/triangulated graphs (longest loop without shortcuts is a triangle), because otherwise we would have an immorality.

Immoralities and chordal/triangulated DAGs

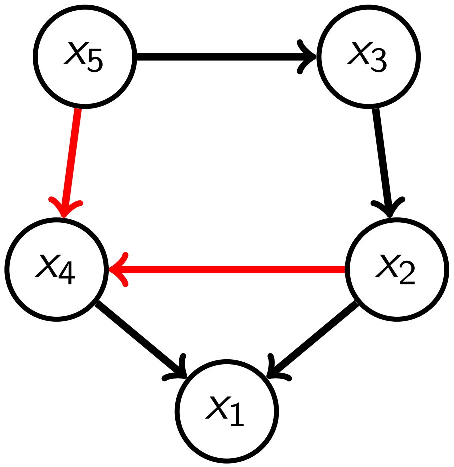
Undirected graph: (immoralities in red)



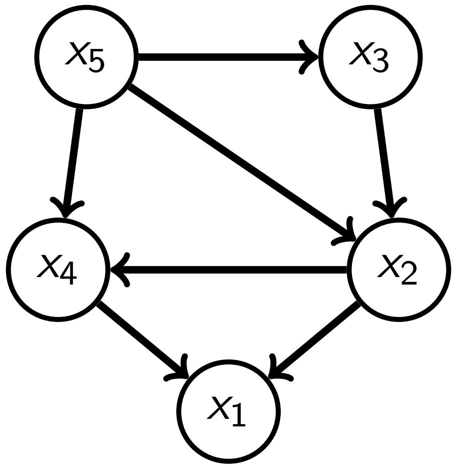
DAGs:



not chordal

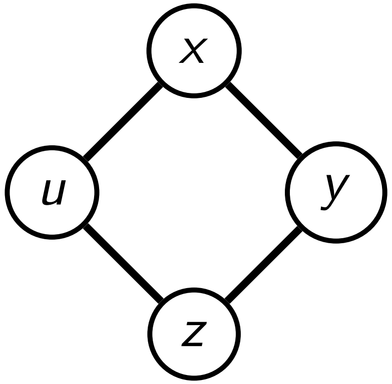


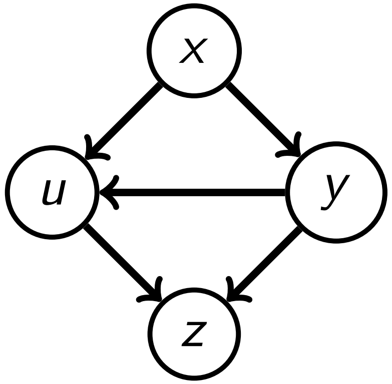
not chordal



chordal

Canonical example





Given: undirected graph H

$$x \perp\!\!\!\perp z \mid u, y$$
$$u \perp\!\!\!\perp y \mid x, z$$

G: min I-map for $\mathcal{I}(H)$
(with ordering: x, y, u, z)

$$x \perp\!\!\!\perp z \mid u, y$$
$$u \not\perp\!\!\!\perp y \mid x, z$$

- We lost information: $\mathcal{I}(G) \subset \mathcal{I}(H)$.
- Different orderings would give different directed minimal I-maps G. But there is no directed perfect map for $\mathcal{I}(H)$.
- Loss of information is due to the loop of length > 3 without a shortcut in H (H is not chordal).

Lossless conversion for chordal undirected graphs

(for proofs, see e.g. Section 4.5.3. in Koller and Friedman's book; proofs not examinable)

- Such loops allow undirected graphs represent independencies that cannot be represented with DAGs (see example above).
- We need to introduce edges (triangulate the graph) when constructing the DAG because otherwise it would not be an I-map. However, triangulation leads to a loss of information.
- If (and only if) H is a chordal/triangulated undirected graph, we can obtain a DAG G that is a perfect map for $\mathcal{I}(H)$, i.e. H and G are I-equivalent.

Strengths and weaknesses

- Some independencies are more easily represented with DAGs, others with undirected graphs.
- Both directed and undirected graphical models have strengths and weaknesses.
- Undirected graphs are suitable when interactions are symmetrical and when there is no natural ordering of the variables, but they cannot represent "explaining away" scenario (colliders).
- DAGs are suitable when we have an idea of the data generating process (e.g. what is "causing" what), but they may force directionality where there is none.
- It is possible to combine individual strengths with mixed/partially directed graphs (see e.g. Barber, Section 4.3, not examinable).

Program recap

1. Independency maps (I-maps)
   - Definition of I-maps and perfect maps
   - I-maps and factorisation
   - Examples and no guarantee for perfect maps
2. Equivalence of I-maps (I-equivalence)
   - I-equivalence for DAGs: check the skeletons and the immoralities
   - I-equivalence for undirected graphs: check the skeletons
3. Minimal I-maps
   - Definition
   - Construction of undirected minimal I-maps
   - Construction of directed minimal I-maps
4. (Lossy) conversion between directed and undirected I-maps
   - Moralisation for directed $\rightarrow$ undirected
   - Triangulation for undirected $\rightarrow$ directed
   - Strengths and weaknesses of directed and undirected graphs