Directed Graphical Models

Michael Gutmann

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

Spring semester 2019

Recap

- Statistical independence assumptions facilitate the efficient representation of probabilistic models by limiting the number of variables that are allowed to directly interact with each other.
- Statistical independencies lead to a (partial) factorisation of pdfs/pmfs
- Equivalence between factorisation and ordered Markov property
- Visualisation of pdfs/pmfs as directed graph

Program

1. Definition of directed graphical models

2. Three canonical connections in a DAG and their properties

3. Independencies in directed graphical models

Program

1. Definition of directed graphical models
   - Definition via factorisation according to the graph
   - Definition via ordered Markov property
   - Derive independencies from the ordered Markov property with different topological orderings
2. Three canonical connections in a DAG and their properties
3. Independencies in directed graphical models

Directed graphical model

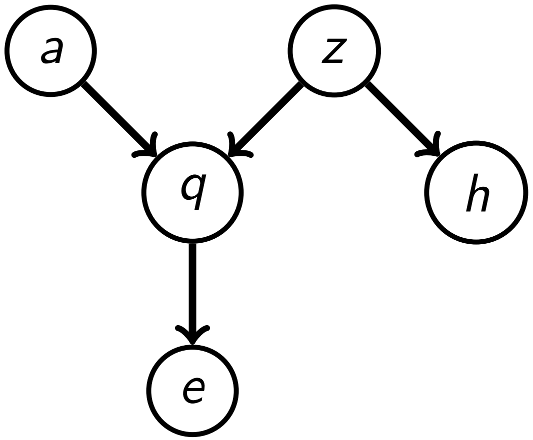
- We started with a pdf/pmf, wrote it in factorised form according to some ordering, and associated a DAG with it.
- We can also go the other way around and start with a DAG.
- *Definition (via factorisation property)* A directed graphical model based on a DAG with d nodes and associated random variables x_i is the set of pdfs/pmfs that factorise as

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

  where pa_i denotes the parents of x_i in the graph.
- Other names for directed graphical models: belief network, Bayesian network, Bayes network.

Example

DAG:



Random variables: a, z, q, e, h

Parent sets: $\mathrm{pa}_a = \mathrm{pa}_z = \varnothing, \mathrm{pa}_q = \{a, z\}, \mathrm{pa}_e = \{q\}, \mathrm{pa}_h = \{z\}$.

All models defined by the DAG factorise as:

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

Alternative definition of directed graphical models

- For any DAG with d nodes we can always find an ordering of the associated random variables that is topological to the DAG. Re-label the nodes accordingly as $x_1, \ldots, x_d$.
- In all topological orderings the parents come before the children.
- Hence: $\mathrm{pa}_i \subseteq \mathrm{pre}_i$ (recall: $\mathrm{pre}_i = \{x_1, \ldots, x_{i-1}\}$)
- Previous result on equivalence of factorisation and ordered Markov property gives

$$p(\mathbf{x}) = \prod_{i=1}^{d} p(x_i | \mathrm{pa}_i) \Longleftrightarrow x_i \perp\!\!\!\perp (\mathrm{pre}_i \setminus \mathrm{pa}_i) \mid \mathrm{pa}_i \text{ for all } i$$

- Provides an alternative definition of directed graphical models

Directed graphical model

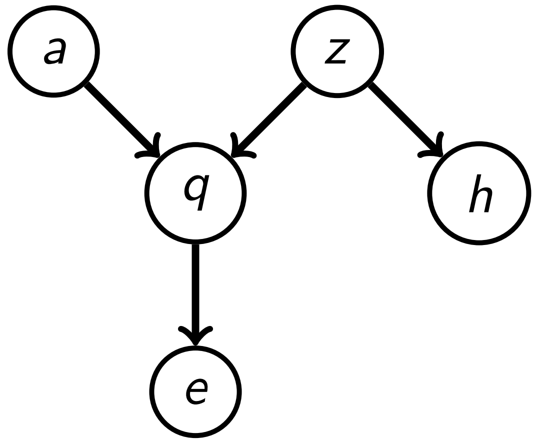
- *Definition (via ordered Markov property)* A directed graphical model based on a DAG with d nodes and associated random variables x_i is the set of pdfs/pmfs that satisfy the ordered Markov property

$$x_i \perp\!\!\!\perp (\mathrm{pre}_i \setminus \mathrm{pa}_i) \mid \mathrm{pa}_i \text{ for all } i$$

  for any ordering $x_1, \ldots, x_d$ of the x_i that is topological to the DAG.
- Remark: the notation is as before:
  pre_i are the predecessors of x_i *in the topological ordering*
  pa_i are the parents of x_i *in the graph*

Example

DAG:



Random variables: a, z, q, e, h
Ordering: (a, z, q, e, h) (meaning: $x_1 = a, x_2 = z, x_3 = q, x_4 = e, x_5 = h$)

Predecessor sets for the ordering:
$\mathrm{pre}_a = \varnothing, \mathrm{pre}_z = \{a\}, \mathrm{pre}_q = \{a, z\}, \mathrm{pre}_e = \{a, z, q\}, \mathrm{pre}_h = \{a, z, q, e\}$

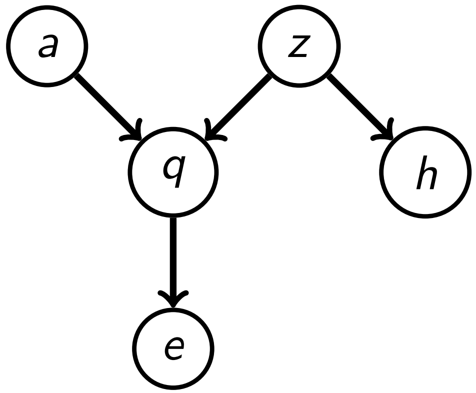
Parent sets: as before
$\mathrm{pa}_a = \mathrm{pa}_z = \varnothing, \mathrm{pa}_q = \{a, z\}, \mathrm{pa}_e = \{q\}, \mathrm{pa}_h = \{z\}$

All models in the set defined by the DAG satisfy $x_i \perp\!\!\!\perp (\mathrm{pre}_i \setminus \mathrm{pa}_i) \mid \mathrm{pa}_i$:

$$z \perp\!\!\!\perp a \qquad e \perp\!\!\!\perp \{a, z\} \mid q \qquad h \perp\!\!\!\perp \{a, q, e\} \mid z$$

Example (different topological ordering)

DAG:



Ordering: (a, z, h, q, e)

Predecessor sets for the ordering:

$\mathrm{pre}_a = \varnothing, \mathrm{pre}_z = \{a\}, \mathrm{pre}_h = \{a, z\}, \mathrm{pre}_q = \{a, z, h\}, \mathrm{pre}_e = \{a, z, h, q\}$

Parent sets: as before

$\mathrm{pa}_a = \mathrm{pa}_z = \varnothing, \mathrm{pa}_h = \{z\}, \mathrm{pa}_q = \{a, z\}, \mathrm{pa}_e = \{q\}$

All models in the set defined by the DAG satisfy $x_i \perp\!\!\!\perp (\mathrm{pre}_i \setminus \mathrm{pa}_i) \mid \mathrm{pa}_i$:

$$z \perp\!\!\!\perp a \qquad h \perp\!\!\!\perp a \mid z \qquad q \perp\!\!\!\perp h \mid a, z \qquad e \perp\!\!\!\perp \{a, z, h\} \mid q$$

Note: the models also satisfy those obtained with the previous ordering:

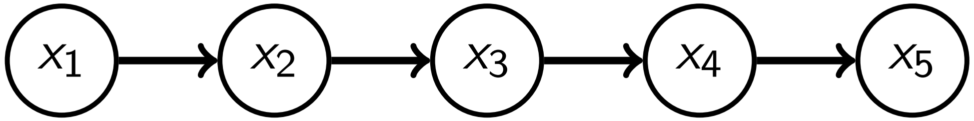
$$z \perp\!\!\!\perp a \qquad e \perp\!\!\!\perp \{a, z\} \mid q \qquad h \perp\!\!\!\perp \{a, q, e\} \mid z$$

Remarks

- Missing edges in a DAG cause the pa_i to be smaller than the pre_i, and thus lead to the independencies.
- The directed graphical model corresponds to a set of probability distributions . Two views according to the two definitions: The set includes all those distributions that you get
  - by looping over all possible conditionals $p(x_i|\mathrm{pa}_i)$,
  - by retaining, from all possible joint distributions over the x_i, those that satisfy the independencies given by the ordered Markov property
- A directed graphical model with specified conditionals is typically also called a directed graphical model.
- By using different topological orderings you can generate possibly different independence relations satisfied by the model.

Example: Markov model

DAG:



All models in the set factorise as
$p(\mathbf{x}) = p(x_1)p(x_2|x_1)p(x_3|x_2)p(x_4|x_3)p(x_5|x_4)$

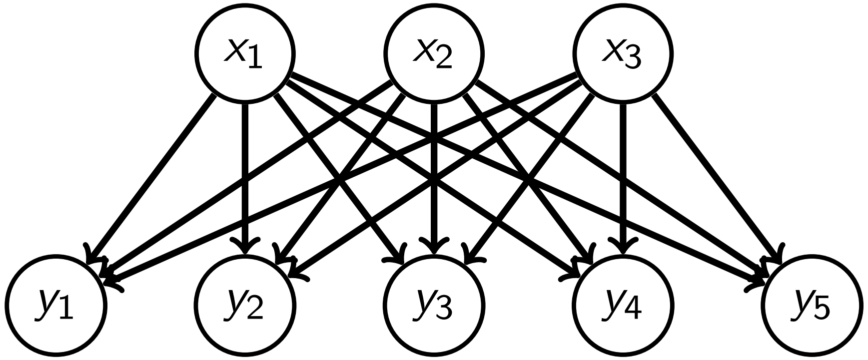
There is only one topological ordering: $(x_1, x_2, \ldots, x_5)$

By ordered Markov property: all models in the set satisfy:
$x_{i+1} \perp\!\!\!\perp x_1, \ldots, x_{i-1} \mid x_i$
(future independent of the past given the present)

Example: Probabilistic PCA, factor analysis, ICA

(PCA: principal component analysis; ICA: independent component analysis)

DAG:



Explains properties of (observed) y_i through fewer (unobserved) x_i. Different further assumptions lead to different methods (more later).

All models in the set factorise as $p(x_1, x_2, x_3, y_1, \ldots, y_5) = p(x_1)p(x_2)p(x_3)p(y_1|x_1, x_2, x_3)p(y_2|x_1, x_2, x_3) \ldots p(y_5|x_1, x_2, x_3)$

With the ordering $(x_1, x_2, x_3, y_1, y_2, y_3, y_4, y_5)$: All satisfy:

$x_i \perp\!\!\!\perp x_j$ $\quad y_2 \perp\!\!\!\perp y_1 \mid x_1, x_2, x_3$ $\quad y_3 \perp\!\!\!\perp y_1, y_2 \mid x_1, x_2, x_3$

$y_4 \perp\!\!\!\perp y_1, y_2, y_3 \mid x_1, x_2, x_3$ $\quad y_5 \perp\!\!\!\perp y_1, y_2, y_3, y_4 | x_1, x_2, x_3$

Program

1. Definition of directed graphical models
   - Definition via factorisation according to the graph
   - Definition via ordered Markov property
   - Derive independencies from the ordered Markov property with different topological orderings
2. Three canonical connections in a DAG and their properties
3. Independencies in directed graphical models

Further independence properties?

- Parent-child links in the graph encode (conditional) independence properties.
- Ordered Markov property yields sets of independence assertions.
- Questions:
  - For any triple of random variables (x, y, z), can we determine whether $x \perp\!\!\!\perp y \mid z$ holds?
  - Does the graph induce or impose additional independencies on any probability distribution that factorises over the graph?
- Important because
  - it yields increased understanding of the properties of the model
  - we can exploit the independencies e.g. for inference and learning
- Approach: Investigate how probabilistic evidence that becomes available at a node can "flow" through the DAG and influence our belief about another node.

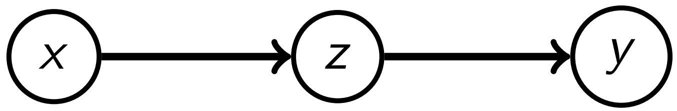
Program

1. Definition of directed graphical models
2. Three canonical connections in a DAG and their properties
   - Serial connection
   - Diverging connection
   - Converging connection
   - I-equivalence
3. Independencies in directed graphical models

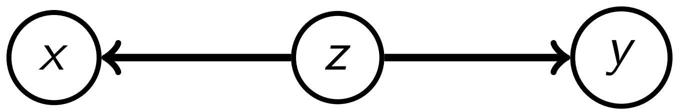
Three canonical connections in a DAG

In a DAG, two nodes x, y can be connected via a third node z in three ways:

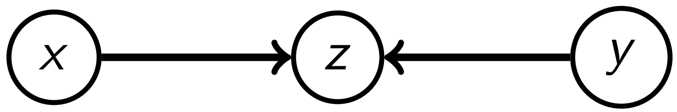
1. Serial connection (chain, head-tail or tail-head)



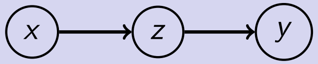
2. Diverging connection (fork, tail-tail)



3. Converging connection (collider, head-head, v-structure)

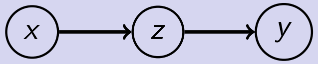


Note: in any case, the sequence x, z, y forms a trail

Serial connection 

- Markov model is made up of serial connections
- Graph: x influences z, which in turn influences y but no direct influence from x to y.
- Factorisation: $p(x, z, y) = p(x)p(z|x)p(y|z)$
- Ordered Markov property: $y \perp\!\!\!\perp x \mid z$
  If the state or value of z is known (i.e. if the random variable z is "instantiated"), evidence about x will not change our belief about y, and vice versa.

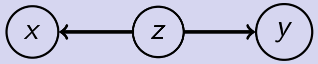
  We say that the z node is "closed" and that the trail between x and y is "blocked" by the instantiated z. In other words, knowing the value of z blocks the flow of evidence *between* x and y.

Serial connection 

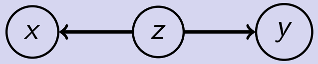
- What can we say about the marginal distribution of (x, y)?
- By sum rule, joint probability distribution of (x, y) is

$$
\begin{aligned}
p(x, y) &= \int p(x)p(z|x)p(y|z)\mathrm{d}z \\
&= p(x) \int p(z|x)p(y|z)\mathrm{d}z \\
&\neq p(x)p(y)
\end{aligned}
$$

- In a serial connection, if the state of z is unknown, then evidence or information about x will influence our belief about y, and the other way around. Evidence can flow through z between x and y.
- We say that the z node is "open" and the trail between x and y is "active".

Diverging connection 

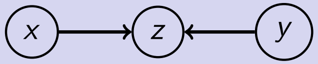
- Graph for probabilistic PCA, factor analysis, ICA has such connections (z correspond to the latents, x and y to the observed)
- Graph: z influences both x and y. No directed connection between x and y.
- Factorisation: $p(x, y, z) = p(z)p(x|z)p(y|z)$
- Ordered Markov property (with ordering z, x, y): $y \perp\!\!\!\perp x \mid z$
  If the state or value z is known, evidence about x will not change our belief about y, and vice versa.
- As in serial connection, knowing z closes the z node, which blocks the trail between x and y.

Diverging connection 

- What can we say about the marginal distribution of (x, y)?
- By sum rule, joint probability distribution of (x, y) is

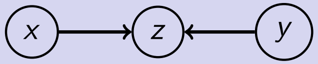
$$\begin{aligned} p(x, y) &= \int p(z)p(x|z)p(y|z)\mathrm{d}z \\ &\neq p(x)p(y) \end{aligned}$$

- In a diverging connection, as in the serial connection, if the state of z is unknown, then evidence or information about x will influence our belief about y, and the other way around. Evidence can flow through z between x and y.
- The z node is open and the trail between x and z is active.

Converging connection 

- Graph for probabilistic PCA, factor analysis, ICA has such connections (z corresponds to an observed, x and y to two latents)
- Graph: x and y influence z. No direction connection between x and y.
- Factorisation: $p(x, y, z) = p(x)p(y)p(z|x, y)$
- Ordered Markov property: $x \perp\!\!\!\perp y$
  If nothing is known about z, except what might follow from knowledge of x and y, then evidence about x will not change our belief about y, and vice versa.

  If no evidence about z is available, the z node is closed, which blocks the trail between x and y.

Converging connection 

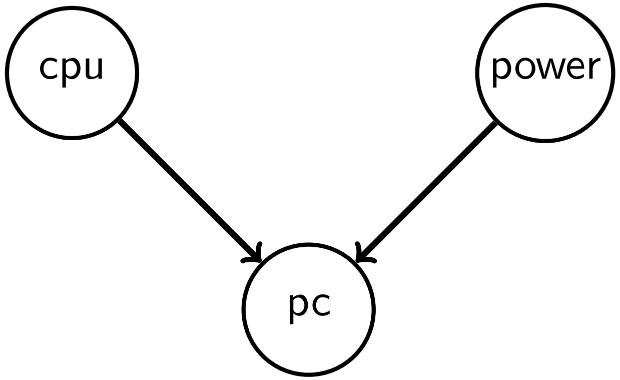
- This means that the marginal distribution of (x, y) factorises: $p(x, y) = p(x)p(y)$
- Conditional distribution of (x, y) given z?

$$p(x, y|z) = \frac{p(x, y, z)}{p(z)} = \frac{p(x)p(y)p(z|x, y)}{\int p(x)p(y)p(z|x, y)\mathrm{d}x\mathrm{d}y}$$
$$\neq p(x|z)p(y|z)$$

  This means that $x \not\perp\!\!\!\perp y \mid z$.
- If evidence or information about z is available, evidence about x will influence the belief about y, and vice versa.
- Information about z opens the z-node, and evidence can flow between x and y.
- Note: information about z means that z or one of its descendents is observed (see tutorials).
  (A node w is a descendant of z if there is a directed path from z to w.)

Explaining away

Example:



- One day your computer does not start and you bring it to a repair shop. You think the issue could be the power unit or the cpu.
- Investigating the power unit shows that it is damaged. Is the cpu fine?
- Without further information, finding out that the power unit is damaged typically reduces our belief that the cpu is damaged

$$\text{power} \not\perp\!\!\!\perp \text{cpu} \mid \text{pc}$$

- Finding out about the damage to the power unit *explains away* the observed start-issues of the computer.

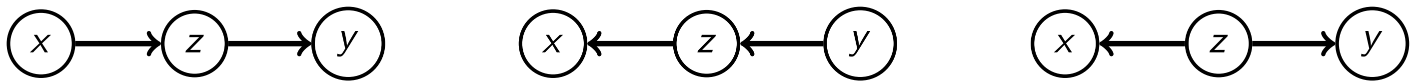
Summary

Connection	z node	$p(x, y)$	$p(x, y \mid z)$
$x \rightarrow z \rightarrow y$	default: open instantiated: closed	$x \not\!\perp\!\!\!\perp y$	$x \perp\!\!\!\perp y \mid z$
$x \leftarrow z \rightarrow y$	default: open instantiated: closed	$x \not\!\perp\!\!\!\perp y$	$x \perp\!\!\!\perp y \mid z$
$x \rightarrow z \leftarrow y$	default: closed with evidence: opens	$x \perp\!\!\!\perp y$	$x \not\!\perp\!\!\!\perp y \mid z$

Think of the z node as a valve or gate through which evidence (probability mass) can flow. Depending on the type of the connection, it's default state is either open or closed. Instantiation/evidence acts as a switch on the valve.

I-equivalence

- Same independence assertions for



- The graphs have different causal interpretations

  Consider e.g. $x \equiv$ rain; $z \equiv$ street wet; $y \equiv$ car accident
- This means that based on statistical dependencies (observational data) alone, we cannot select among the graphs and thus determine what causes what.
- The three directed graphs are said to be independence-equivalent (I-equivalent).

Program

1. Definition of directed graphical models

2. Three canonical connections in a DAG and their properties
   - Serial connection
   - Diverging connection
   - Converging connection
   - I-equivalence

3. Independencies in directed graphical models

Program

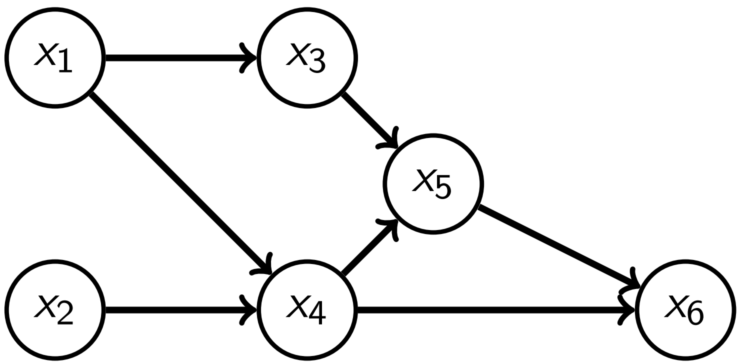
1. Definition of directed graphical models
2. Three canonical connections in a DAG and their properties
3. Independencies in directed graphical models
   - D-separation and I-map
   - Directed local Markov property
   - Equivalences of the different Markov properties and the factorisation
   - Markov blanket

Program

1. Definition of directed graphical models
2. Three canonical connections in a DAG and their properties
3. Independencies in directed graphical models
   - D-separation and I-map
   - Directed local Markov property
   - Equivalences of the different Markov properties and the factorisation
   - Markov blanket

Further independence relations

- Given the DAG below, what can we say about the independencies for the set of probability distributions that factorise over the graph?
- Is $x_1 \perp\!\!\!\perp x_2$? $\quad x_1 \perp\!\!\!\perp x_2 \mid x_6$? $\quad x_2 \perp\!\!\!\perp x_3 \mid \{x_1, x_4\}$?
- Ordered Markov properties give some independencies.
- Limitation: only allows us to condition on parent sets.
- Directed separation (d-separation) gives further independencies.



D-separation

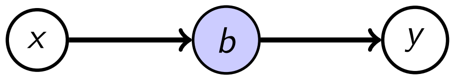
Let $X = \{x_1, \ldots, x_n\}$, $Y = \{y_1, \ldots, y_m\}$, and $Z = \{z_1, \ldots, z_r\}$ be three disjoint sets of nodes in the graph. Assume all z_i are observed (instantiated).

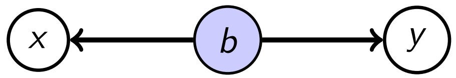
- Two nodes x_i and y_j are said to be d-separated by Z if all trails between them are blocked by Z.
- The sets X and Y are said to be d-separated by Z if every trail from any variable in X to any variable in Y is blocked by Z.

D-separation

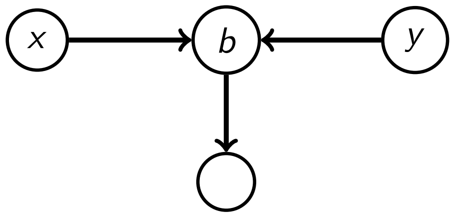
A trail between nodes x and y is blocked by Z if there is a node b on the trail such that

1. either b is part of a head-tail or tail-tail connection along the trail and b is in Z,





2. or b is part of a head-head (collider) connection along the trail and neither b nor any of its descendants are in Z.



D-separation and conditional independence

Theorem: If X and Y are d-separated by Z then $X \perp\!\!\!\perp Y \mid Z$ for all probability distributions that factorise over the DAG.

For those interested: A proof can be found in Section 2.8 of *Bayesian Networks – An Introduction* by Koski and Noble (not examinable)

Important because:

1. the theorem allows us to read out (conditional) independencies from the graph
2. no restriction on the sets X, Y, Z
3. the theorem shows that d-separation does not indicate false independence relations. It's independence assertions are sound ("soundness of d-separation").

D-separation and conditional independence

Theorem: If X and Y are not d-separated by Z
then $X \not\perp\!\!\!\perp Y \mid Z$ in some probability distributions that factorise over the DAG.

For those interested: A proof sketch can be found in Section 3.3.1 of *Probabilistic Graphical Models* by Koller and Friedman (not examinable).

"not d-separated" is also called "d-connected"
$\not\perp\!\!\!\perp$ means statistically dependent

D-separation and conditional independence

- It can also be that d-connected variables are independent for some distributions.
- Example (Koller, Example 3.3): $p(x, y)$ with $x, y \in \{0, 1\}$ and

$$p(y = 0|x = 0) = a \quad p(y = 0|x = 1) = a$$

  for $a > 0$ and some non-zero $p(x = 0)$.
- Graph has arrow from x to y. Variables are not d-separated.

- $p(y = 0) = ap(x = 0) + ap(x = 1) = a$,
  which is $p(y = 0|x)$ for all x.
- $p(y = 1) = (1 - a)p(x = 0) + (1 - a)p(x = 1) = 1 - a$,
  which is $p(y = 1|x)$ for all x.
- Hence: $p(y|x) = p(y)$ so that $x \perp\!\!\!\perp y$.

D-separation and conditional independence

- This means that d-separation does generally not reveal all independencies in all probability distributions that factorise over the graph.
- In other words, individual probability distributions that factorise over the graph may have further independencies not included in the set obtained by d-separation.
- We say that d-separation is not “complete”.

I-map

- A graph is said to be an independency map (I-map) for a set of independencies $\mathcal{I}$ if the independencies asserted by the graph are part of $\mathcal{I}$.
- For a directed graph G, let $\mathcal{I}(G)$ be all the independencies that we can derive via d-separation.
- Denote the independencies that a specific distribution p satisfies by $\mathcal{I}(p)$.
- The previous results on d-separation can thus be written as

$$\mathcal{I}(G) \subseteq \mathcal{I}(p) \qquad \text{for all } p \text{ that factorise over } G$$

- As we have seen, we generally do not have $\mathcal{I}(G) = \mathcal{I}(p)$. If we have equality, the graph is said to be a perfect map (P-map) for $\mathcal{I}(p)$.

Recipe to determine whether two nodes are d-separated

1. Determine all trails between x and y (note: direction of the arrows does here not matter).
2. For each trail:
   i Determine the default state of all nodes on the trail.
      - open if part of a head-tail or a tail-tail connection
      - closed if part of a head-head connection

   ii Check whether the set of observed nodes Z switches the state of the nodes on the trail.

   iii The trail is blocked if it contains a closed node.
3. The nodes x and y are d-separated if all trails between them are closed.

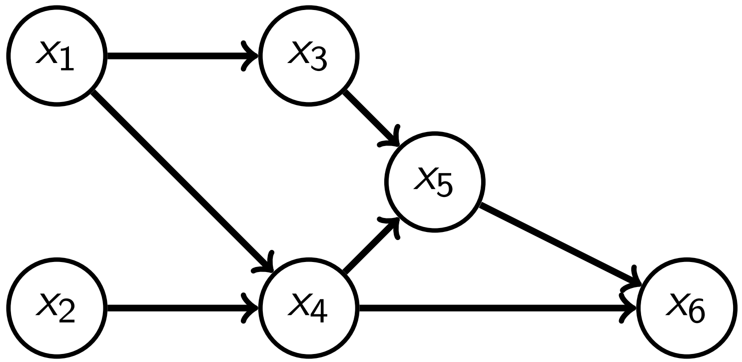
Example: Are x_1 and x_2 d-separated?

Follows from ordered Markov property, but let us answer it with d-separation.

1. Determine all trails between x_1 and x_2
2. For trail x_1, x_4, x_2
   - i default state
   - ii conditioning set is empty
   - iii $\Rightarrow$ Trail is blocked

   For trail x_1, x_3, x_5, x_4, x_2
   - i default state
   - ii conditioning set is empty
   - iii $\Rightarrow$ Trail is blocked

   Trail $x_1, x_3, x_5, x_6, x_4, x_2$ is blocked too (same arguments).
3. $\Rightarrow$ x_1 and x_2 are d-separated.

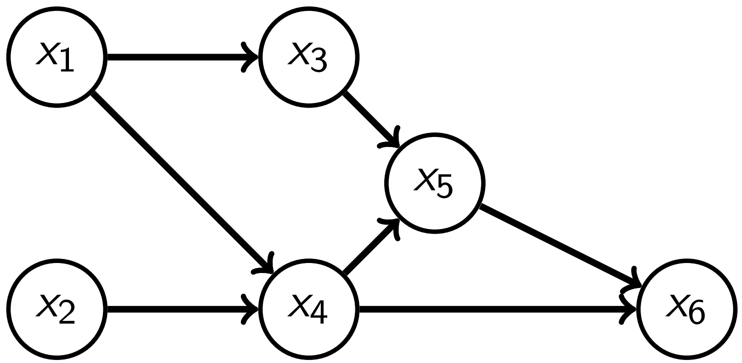


$x_1 \perp\!\!\!\perp x_2$ for all probability distributions that factorise over the graph.

Example: Are x_1 and x_2 d-separated by x_6?

1. Determine all trails between x_1 and x_2
2. For trail x_1, x_4, x_2
   - i default state
   - ii influence of x_6
   - iii $\Rightarrow$ Trail not blocked

   No need to check the other trails: x_1 and x_2 are not d-separated by x_6



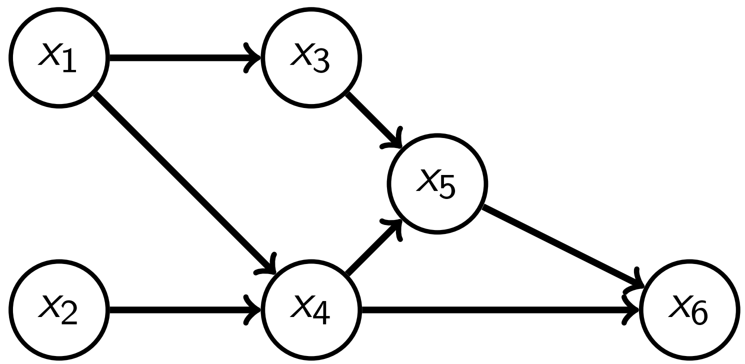
$x_1 \perp\!\!\!\perp x_2 \mid x_6$ does generally not hold for probability distributions that factorise over the graph.

Example: Are x_2 and x_3 d-separated by x_1 and x_4?

1. Determine all trails between x_2 and x_3
2. For trail x_3, x_1, x_4, x_2
   i default state
   ii influence of $\{x_1, x_4\}$
   iii $\Rightarrow$ Trail blocked

   For trail x_3, x_5, x_4, x_2
   i default state
   ii influence of $\{x_1, x_4\}$
   iii $\Rightarrow$ Trail blocked

   Trail x_3, x_5, x_6, x_4, x_2 is blocked too (same arguments).
3. $\Rightarrow$ x_2 and x_3 are d-separated by x_1 and x_4.



$x_2 \perp\!\!\!\perp x_3 \mid \{x_1, x_4\}$ for all probability distributions that factorise over the graph.

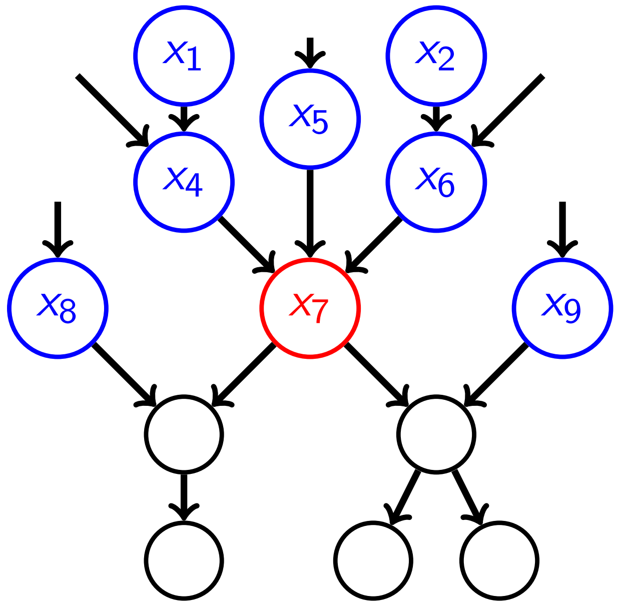
Directed local Markov property

- The independencies from the ordered Markov property depend on the topological ordering chosen.
- We now use d-separation to derive a similarly local Markov property that does not depend on the ordering, and show the equivalence for any topological ordering:

$$x_i \perp\!\!\!\perp (\mathrm{pre}_i \setminus \mathrm{pa}_i) \,|\mathrm{pa}_i \iff x_i \perp\!\!\!\perp (\mathrm{nondesc}(x_i) \setminus \mathrm{pa}_i) \,|\mathrm{pa}_i$$

where $\mathrm{nondesc}(x_i)$ denotes the non-descendants of x_i.

$$x_i \equiv x_7$$
$$\mathrm{pa}_7 = \{x_4, x_5, x_6\}$$
$$\mathrm{pre}_7 = \{x_1, x_2, \ldots, x_6\}$$
$\mathrm{nondesc}(x_7)$ in blue



Directed local Markov property

$x_i \perp\!\!\!\perp \mathrm{pre}_i \setminus \mathrm{pa}_i | \mathrm{pa}_i \Leftarrow x_i \perp\!\!\!\perp \mathrm{nondesc}(x_i) \setminus \mathrm{pa}_i | \mathrm{pa}_i$ follows because $\{x_1, \ldots, x_{i-1}\} \subseteq \mathrm{nondesc}(x_i)$ for all topological orderings

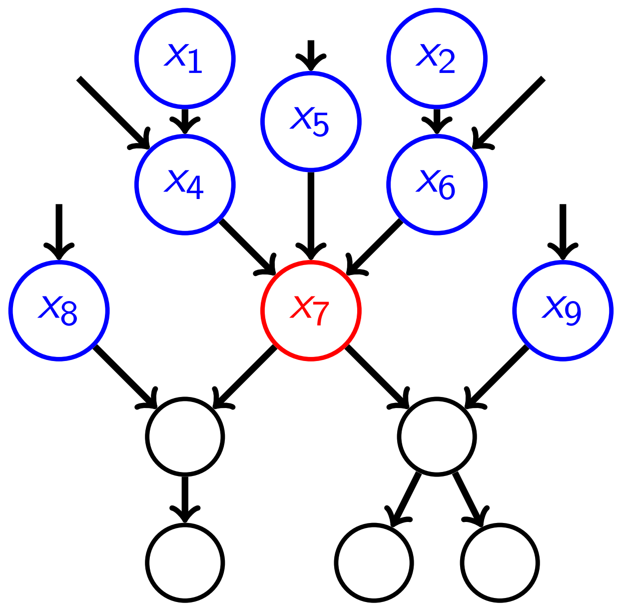
For $\Rightarrow$ consider all trails from x_i to $\{\mathrm{nondesc}(x_i) \setminus \mathrm{pa}_i\}$.

Two cases: move against or with the arrows:

(1) upward trails are blocked by the parents

(2) downward trails must contain a head-head (collider) connection because the $x_j \in \{\mathrm{nondesc}(x_i) \setminus \mathrm{pa}_i\}$ is a non-descendant. These paths are blocked because the collider node or its descendants are never part of pa_i.

The result now follows because all paths from x_i to all elements in $\{\mathrm{nondesc}(x_i) \setminus \mathrm{pa}_i\}$ are blocked.



Remarks

- The local Markov independencies do not depend on a topological ordering. They can be directly read from the graph.
- The direction "local Markov property $\Rightarrow$ ordered Markov property" implies that models that satisfy one ordered Markov property also have to satisfy all other ordered Markov properties obtained with different topological orderings.
- This means that a directed graphical model can be specified via the directed Markov properties for one topological ordering only.

Summary of the equivalences

Factorisation		$p(\mathbf{x}) = \prod_{i=1}^{d} p(x_i \vert \mathrm{pa}_i)$
	$\Updownarrow$	
ordered Markov property		$x_i \perp\!\!\!\perp \mathrm{pre}_i \setminus \mathrm{pa}_i \mid \mathrm{pa}_i$
	$\Updownarrow$	
local directed Markov property		$x_i \perp\!\!\!\perp \mathrm{nondesc}(x_i) \setminus \mathrm{pa}_i \mid \mathrm{pa}_i$
	$\Updownarrow$	
global directed Markov property		all independencies by d-separation

Broadly speaking, the graph serves two related purposes:

1. it tells us how distributions factorise
2. it represents the independence assumptions made

Markov blanket

What is the minimal set of variables such that knowing their values makes x independent from the rest?

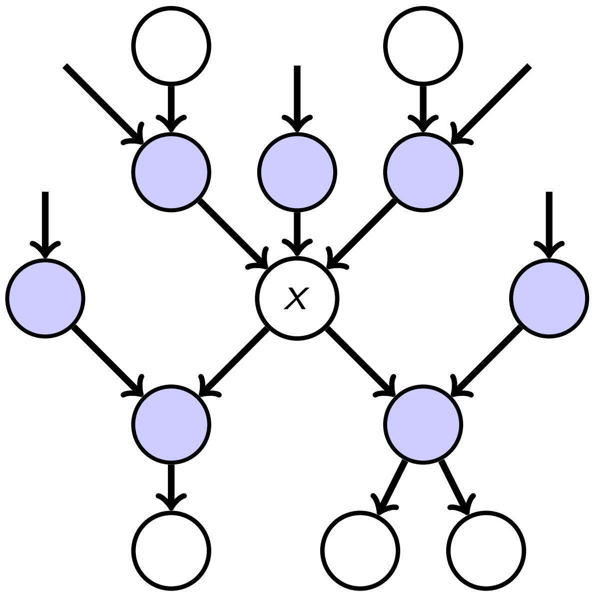
From d-separation:

- Isolate x from its ancestors

  $\Rightarrow$ condition on parents
- Isolate x from its descendants

  $\Rightarrow$ condition on children
- Deal with collider connection

  $\Rightarrow$ condition on co-parents (other parents of the children of x)



In a directed graphical model, the parents, children, and co-parents of x are called its Markov blanket, denoted by $\mathrm{MB}(x)$. We have $x \perp\!\!\!\perp \{\text{all variables} \setminus x \setminus \mathrm{MB}(x)\} \mid \mathrm{MB}(x)$.

Program recap

1. Definition of directed graphical models
   - Definition via factorisation according to the graph
   - Definition via ordered Markov property
   - Derive independencies from the ordered Markov property with different topological orderings
2. Three canonical connections in a DAG and their properties
   - Serial connection
   - Diverging connection
   - Converging connection
   - I-equivalence
3. Independencies in directed graphical models
   - D-separation and I-map
   - Directed local Markov property
   - Equivalences of the different Markov properties and the factorisation
   - Markov blanket