Neuroinformatics Lecture (L4)

Prof. Dr. Gordon Pipa
Institute of Cognitive Science University of Osnabrück

Transformed Densities: Example 1 (proposed by the TAs)

R

You want to organize a party. You know the probability distribution for the number of people coming (x=h_c: head count): $p_x(x)$

Moreover you heard from a friend that he experienced that the number of drinks per person depends on the head count. From him you got the following functional relation between the head count and the number drinks (y) consumed. .

$g(x): y = x^2$

Different example from previous lecture !

Given this you want to know the probability distribution of the number of drinks per person.

$p_y(y)$

Transformed Densities: Example 2 (proposed by the TAs)

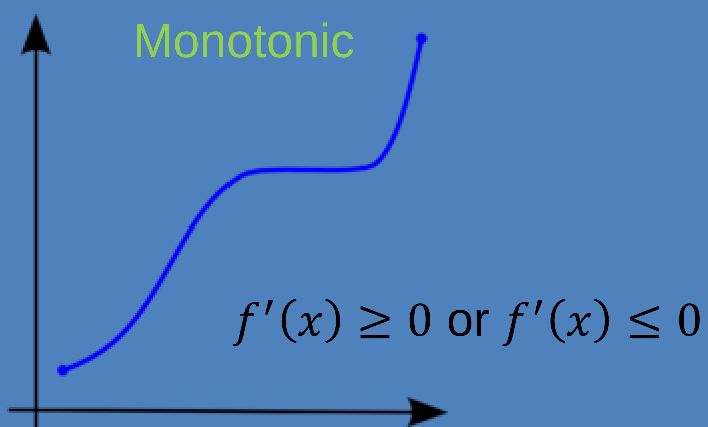
We have: $p_x(x)$ $\quad g(x): y = x^2$ $\quad$ We want to know: $p_y(y)$

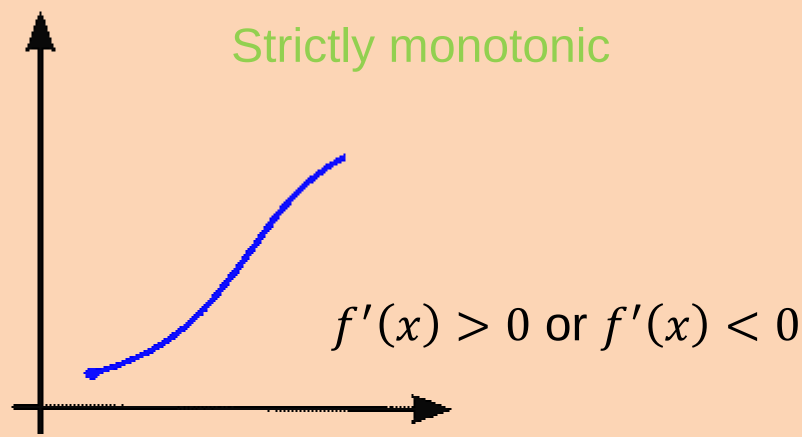
Step 1: We need to check whether g is a strictly monotonic function on the interval of interest.

We compute the derivative:

$$\frac{d}{dx} g(x) = \frac{d}{dx} x^2 = 2x$$

On the interval x>0 this slope of the function is larger 0 → strictly monotonic on x>0





Transformed Densities: Example 2 (proposed by the TAs)

We have: $p_x(x)$ $g(x): y = x^2$ We want to know: $p_y(y)$

Step 2: We know y(x). To compute the probability *p(y)* given *p(x)* we need to substitute *x* by *y* . → We need to compute the inverse function g^{-1} of *g*

$$g(x): y = x^2 \quad \Leftrightarrow g(y)^{-1}: x = \sqrt{y}$$

Step 3: We substitute x by y, and multiply with the derivative of g^{-1}

$$p(y) = p_x(g^{-1}(y))\left|\frac{dg^{-1}(y)}{dy}\right| = p_x(x)\left|\frac{dg^{-1}(y)}{dy}\right|$$

$$p(y) = p_x(\sqrt{y})\left|\frac{dg^{-1}(y)}{dy}\right|$$

$$p(y) = p_x(\sqrt{y})\left|\frac{1}{2\sqrt{y}}\right|$$

Transformed Densities: Example 2 (proposed by the TAs)

We have: $p_x(x)$ $g(x): y = x^2$ We want to know: $p_y(y)$

Lets say $p_x(x)$ is approximately Normal (Gaussian) distributed

$$p_x(x) = \frac{1}{norm} e^{\left(-\frac{(x-\mu_x)^2}{\sigma_x^{\;2}}\right)}$$

Step 3: Now we can compute p(y)

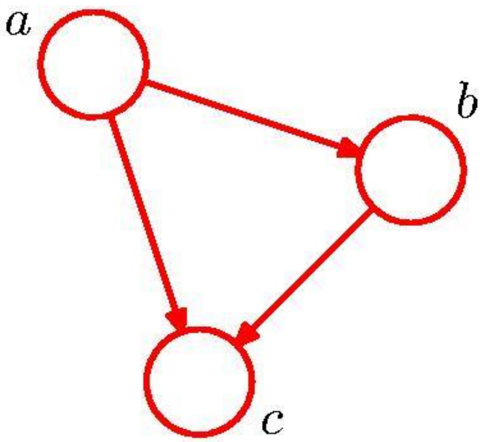
$$p(y) = p_x(\sqrt{y})\left|\frac{1}{2\sqrt{y}}\right|$$

$$p(y) = \left|\frac{1}{2\sqrt{y}}\right|\frac{1}{norm} e^{\left(-\frac{(\sqrt{y}-\mu_x)^2}{\sigma_x^{\;2}}\right)}$$

A graph comprises *nodes* (also called *vertices*) connected by *links* (also known as *edges* or *arcs*).

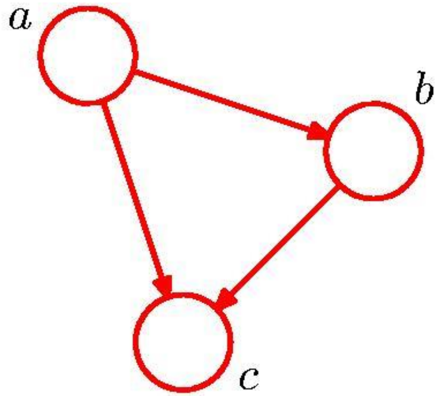
In a probabilistic graphical model, each node represents a random variable (or a group of random variables), and the links express probabilistic relationships between these variables.

The graph then captures the way in which the joint distribution over all of the random variables can be decomposed into a product of factors, each depending only on a subset of the variables.



Conditional Models for K Variables

- The directed graphs that we are considering are subject to an important restriction namely that there must be no ***directed cycles*.**
- In other words, there are no closed paths within the graph such that we can move from node to node along links following the direction of the arrows and end up back at the starting node.
- Such graphs are also called *directed acyclic graphs*, or ***DAGs***. This is equivalent to the statement that there exists an ordering of the nodes such that there are no links that go from any node to any lower-numbered node.



Conditional Models for K Variables

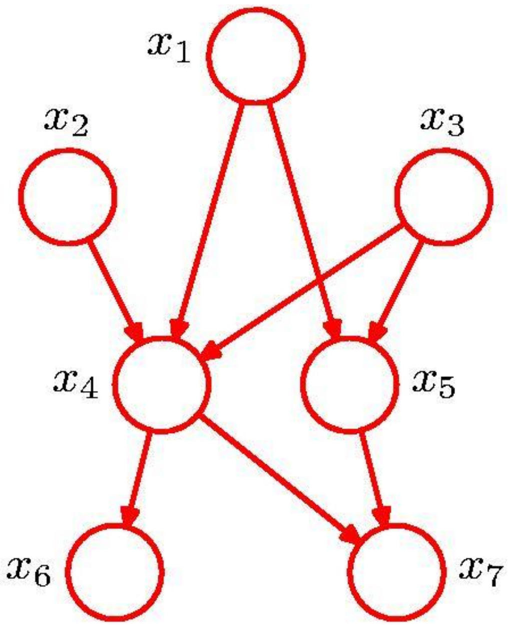
For K=3 variables:

$$p(a,b,c) = p(c|a,b)p(a,b) = p(c|a,b)p(b|a)p(a)$$

For K variables in general:

$$p(x_1, x_2, \ldots, x_K) = p(x_K | x_1, x_2, \ldots, x_{K-1}) \cdot p(x_{K-1} | x_1, x_2, \ldots, x_{K-2}) \cdot \ldots \cdot p(x_2 | x_1) p(x_1)$$

For a given choice of *K*, we can again represent this as a directed graph having *K* nodes, one for each conditional distribution on the right-hand side, with each node having incoming links from all lower-numbered nodes. We say that this graph is fully connected because there is a link between every pair of nodes.



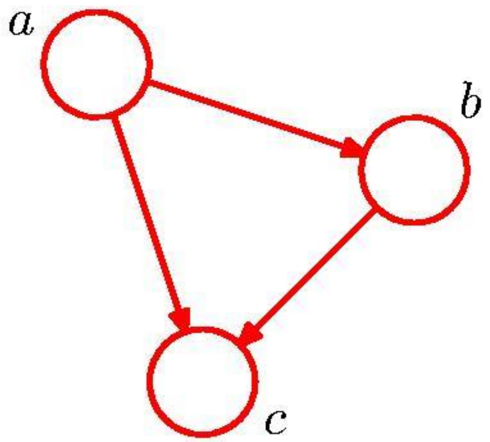
$$
\begin{aligned}
p(x_1, \ldots, x_7) \quad = \quad & p(x_1)p(x_2)p(x_3)p(x_4|x_1, x_2, x_3) \\
& p(x_5|x_1, x_3)p(x_6|x_4)p(x_7|x_4, x_5)
\end{aligned}
$$

General Factorization

The joint distribution defined by a graph is given by the product, over all of the nodes of the graph, of a conditional distribution for each node conditioned on the variables corresponding to the parents of that node in the graph. Thus, for a graph with *K* nodes, the joint distribution is given by:

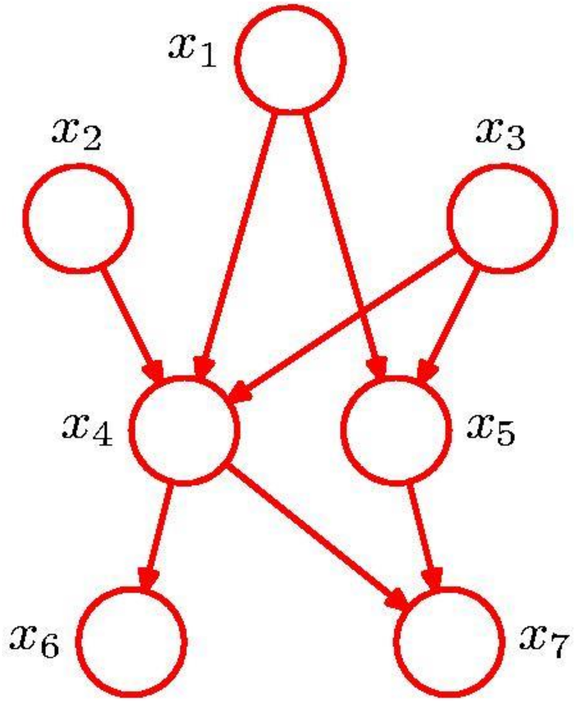
$$p(\mathbf{x}) = \prod_{k=1}^{K} p(x_k|\mathrm{pa}_k)$$

pa_k denotes the set of parents of x_k



Joint distribution ?

$$p(a,b,c) = p(c|a,b)$$

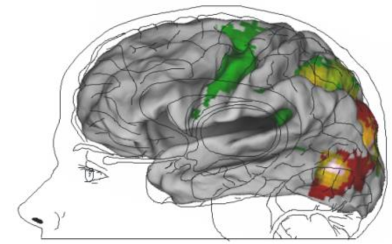


$$p(x_1, \ldots, x_7) \quad = \quad p(x_1)p(x_2)p(x_3)$$

Graphical Models – Generative Models

We can interpret such models as expressing the processes by which the observed data arose.

Complex image



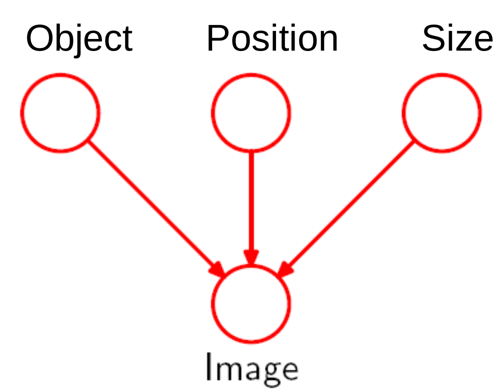
c.f. www.uniklinik-freiburg.de

For example, consider an object-recognition task in which each observed data point corresponds to an image (comprising a vector of pixel intensities) of one of the objects.

Image composed of building blocks

In this case, the latent variables might have an interpretation as the position and size of the objects.

Generative model



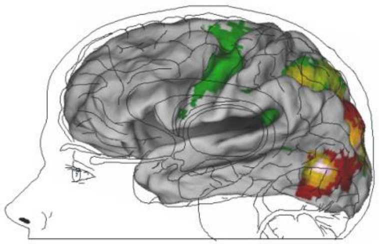
p(Image|Object,Position,Size) = p(Objects) · p(Positions) · p(Sizes)

Latent variables are variables that are not directly observed, **observed variables** are observed

The graphical model captures the *causal* process (Pearl, 1988) by which the observed data was generated. For this reason, such models are often called *generative* models.

Given a particular observed image, our goal is to find the posterior distribution over objects, in which we integrate over all possible positions and sizes.

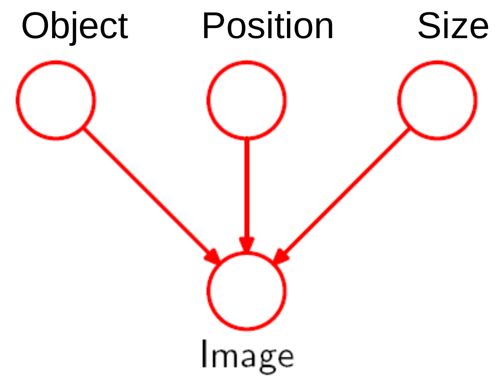
Complex image



c.f. www.uniklinik-freiburg.de

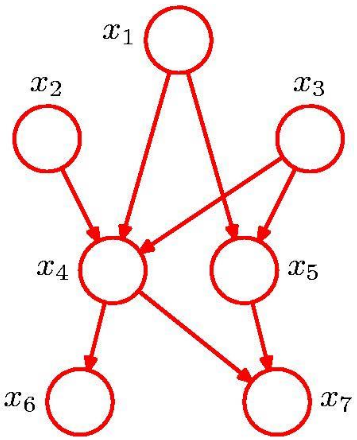
Image composed of building blocks

Generative model



p(Image|Object,Position,Size) = p(Objects) · p(Positions) · p(Sizes)

Q



$$p(\mathbf{x}) = \prod_{k=1}^{K} p(x_k|\mathrm{pa}_k)$$

3. Tick off correct statements.

(a) (4 pts) There might be more than one correct statements.

- ☐ A graph comprises nodes and links
- ☐ The entire graph describes how Joint distribution can be decomposed
- ☐ There is always only one unique graph per Joint distribution
- ☐ There can be more than one graph per joint distribution depending of the order that was used for decomposition into conditional probabilities.

4. Decompose $p(a, b, c)$ into a graphical model. Make a sketch of the model.

5. Write down the decomposed joint distribution given the graph shown on the slide.

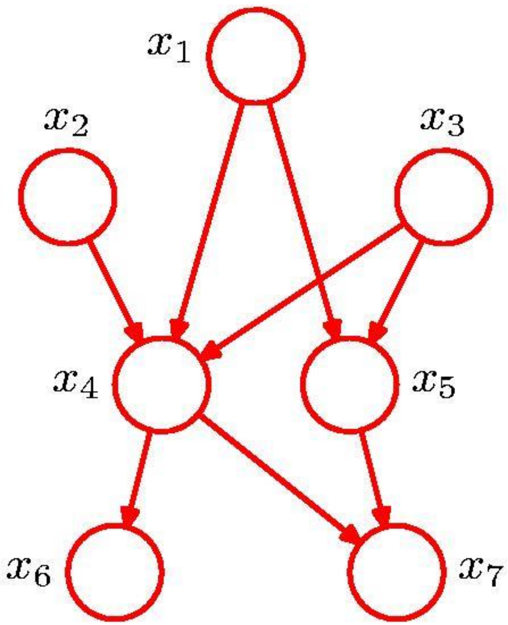
Timer (5min): Start Stop

S

3. Tick off correct statements.

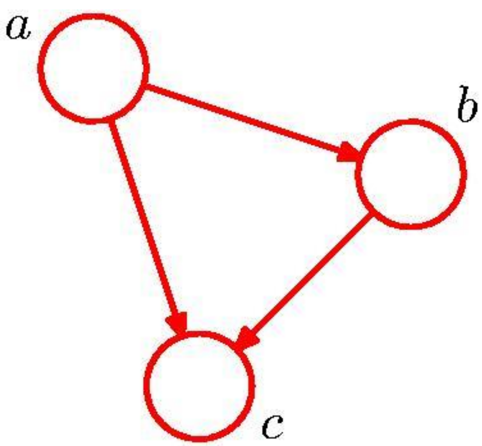
(a) (4 pts) There might be more than one correct statements.

- [X] A graph comprises nodes and links
- [X] The entire graph describes how Joint distribution can be decomposed
- [ ] There is always only one unique graph per Joint distribution
- [X] There can be more than one graph per joint distribution depending of the order that was used for decomposition into conditional probabilities.



$$
\begin{aligned}
p(x_1, \ldots, x_7) \quad = \quad & p(x_1)p(x_2)p(x_3)p(x_4|x_1, x_2, x_3) \\
& p(x_5|x_1, x_3)p(x_6|x_4)p(x_7|x_4, x_5)
\end{aligned}
$$

$$
\begin{aligned}
p(a,b,c) = & \; p(c\,|\,a,b)\,p(b\,|\,a)\,p(a) \\
& p(a|b,c)p(b|c)p(c) \\
& p(b\,|\,a,c)\,p(c\,|\,a)\,p(a)
\end{aligned}
$$



Neuroinformatics Lecture

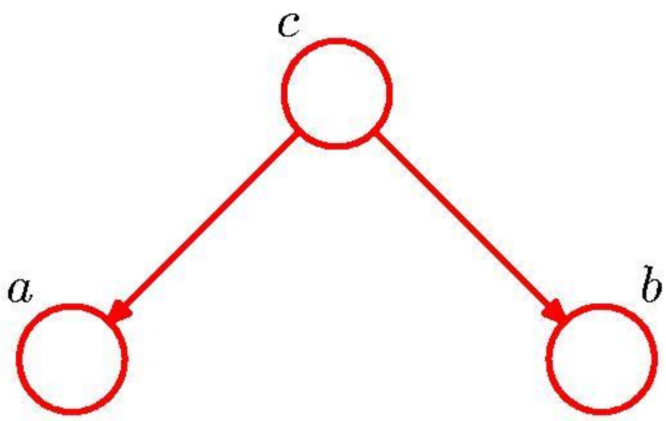
Topics:

Independence in GRAPHICAL MODELS

Statistical Independence

The most important property for graphical models is statistical independence, since it determines the structure of the graph.

Consider three variables *a*, *b*, and *c*, and suppose that the distribution of *a*, *b* and *c*, is such that it factors into a part with p(a)p(b)….



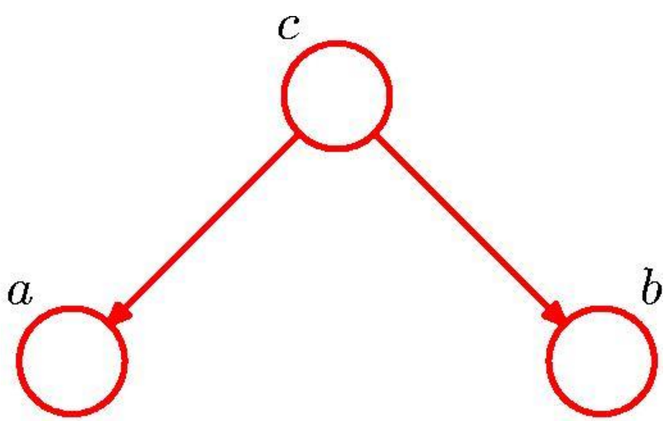
Then we say:

a is independent of b

Statistical independence:
Stochastic variables are independent if and only if the joint probability factorizes.

$$if\ p(x_1, x_2, \dots, x_N) = p(x_1)p(x_2) \cdot \dots \cdot p(x_N)\ than\ \ and\ only\ than\ x_1\ to\ x_N\ \ \ are\ independent$$

$$p(a, b, c) = p(a|c)p(b|c)p(c).$$



We can investigate whether *a* and *b* are independent by marginalizing both sides with respect to *c* to get

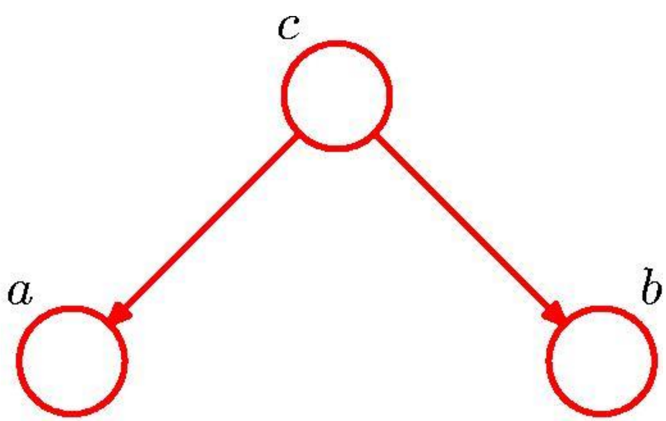
$$p(a, b) = \sum_{c} p(a|c)p(b|c)p(c).$$

We want the joint probability of just a,b

In general, this does not factorize into the product $p(a)p(b)$, and so the nodes a and b are not independent.

We say, $a \not\perp\!\!\!\perp b$ The symbol $\not\perp\!\!\!\perp$ means that the independence property does not hold.

$$p(a, b, c) = p(a|c)p(b|c)p(c).$$



We can investigate whether ***a* and *b*** are independent by marginalizing both sides with respect to *c* to give

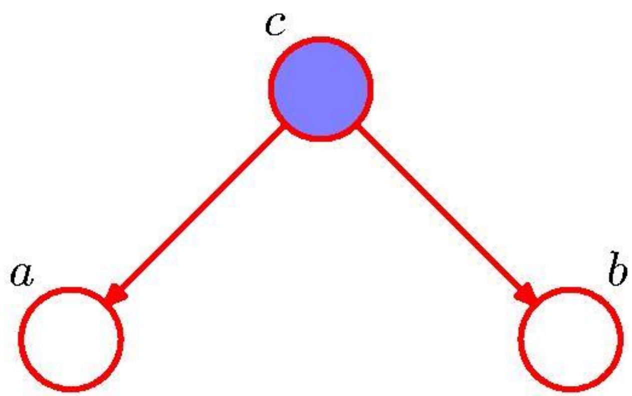
$$p(a, b) = \sum_{c} p(a|c)p(b|c)p(c).$$

In general, this does not factorize into the product $p(a)p(b)$, and so the nodes a and b are not independent.

But what if we know the state of node c ?

Conditional Independence

$$p(a,b \mid c) = p(a \mid c)\, p(b \mid c)$$



In a graphical model, shaded nodes indicate *observed* or fixed *variables.*

a is independent of b given c, if we can factorize into conditional probabilities that are all conditioned only on c:

$$p(a,b \mid c) = p(a \mid c)\, p(b \mid c)$$

Than we say that the variables *a* and *b* are statistically independent, given *c*. $a \perp\!\!\!\perp b \mid c$

Trivial Conditional Independence

Consider three variables *a*, *b*, and *c*, and suppose that the conditional distribution of *a*, given *b* and *c*, is such that it **does not depend on b**.

$$p(a \mid b, c) = p(a \mid c) \quad \text{or}$$

Than we say:

a is trivially independent of b given c or $(a \perp\!\!\!\perp b \mid c)$

Summary on Conditional Independence

Consider three variables *a*, *b*, and *c*, and a conditional distribution of *a*, given *b* and *c*.

a is trivially independent of b given c ($a \perp\!\!\!\perp b \mid c$) , if e.g.:

$$p(a \mid b, c) = p(a \mid c)$$

Moreover, a is independent of b given c, if:

$$p(a, b \mid c) = p(a \mid c)\, p(b \mid c)$$

Statistical conditional independence:

Stochastic variables are conditionally independent if and only if the joint probability factorizes.

$$if\ \ p(x_1, x_2, \dots, x_N | C) = p(x_1|C)p(x_2|C) \cdot \dots \cdot p(x_N|C)$$

$$than\ \ and\ only\ than\ x_1\ to\ x_N\ \ \ are\ conditonally\ independent\ given\ C$$

Summary

Statistical independence:

$if\ p(x_1, x_2, \ldots, x_N) = p(x_1)p(x_2) \cdot \ldots \cdot p(x_N)\ than\ \ and\ only\ than\ x_1\ to\ x_N\quad are\ independent$

$a \perp\!\!\!\perp b$ *a and b are independent*

$a \not\perp\!\!\!\perp b$ *a and b are not independent*

Statistical conditional independence:

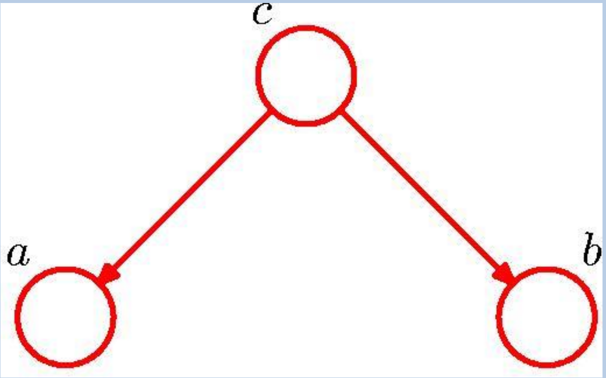
$if\ p(x_1, x_2, \ldots, x_N|C) = p(x_1|C)p(x_2|C) \cdot \ldots \cdot p(x_N|C)$
$than\ \ and\ only\ than\ x_1\ to\ x_N\quad are\ conditonally\ independent\ given\ C$

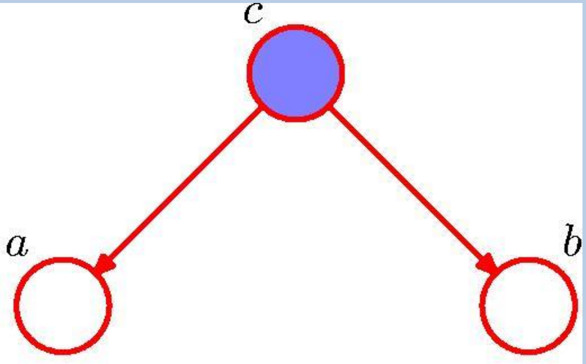
We use the following symbols in that case:

$a \perp\!\!\!\perp b \mid c$ *a and b are conditonally independent given C*

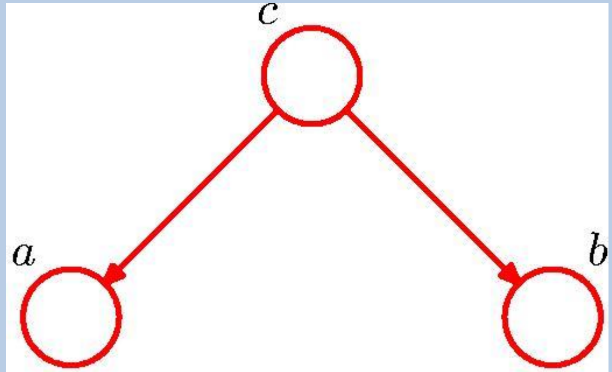
$a \not\perp\!\!\!\perp b \mid \emptyset$ *a and b are not conditonally independent given C*

Common Cause





Common Cause



a and b are dependent

$$p(a, b, c) = p(a|c)p(b|c)p(c).$$

$$p(a, b) = \sum p(a|c)p(b|c)p(c).$$

In general, this does not factorize into the product *p*(*a*)*p*(*b*), and so the nodes a and b are not independent.

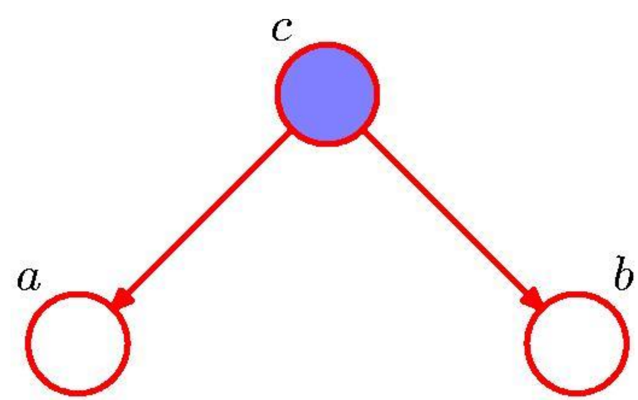
$$a \not\!\perp\!\!\!\perp b \mid \emptyset$$

Conditional Independence

General procedure to test on conditional independence:

We want to check on conditional dependence of a and b given c.

$$p(a,b \mid c) \stackrel{?}{=} p(a \mid c)\, p(b \mid c)$$



We start by writing:

$$p(a,b \mid c)\, p(c) = p(a,b,c)$$

Than we read the joint probability from the graph:

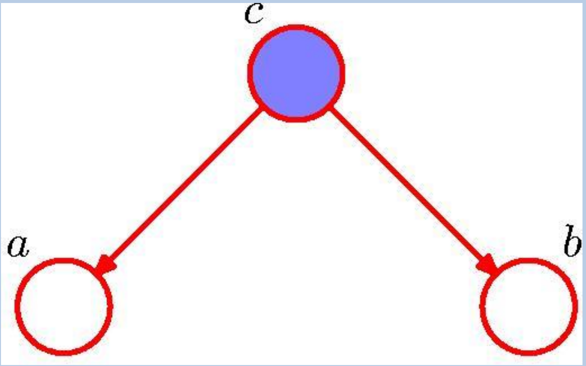
$$p(a,b,c) = p(c)\, p(a \mid c)\, p(b \mid c)$$

We use that for the Eqn. above:

$$p(a,b \mid c)\, \cancel{p(c)} = \cancel{p(c)}\, p(a \mid c)\, p(b \mid c)$$

$$\Leftrightarrow\ p(a,b \mid c) = p(a \mid c)\, p(b \mid c) \qquad\qquad a \perp\!\!\!\perp b \mid c$$

Common Cause



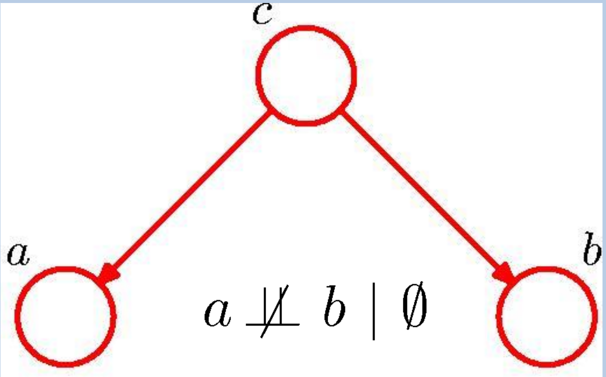
Conditioning on c blocks the paths.

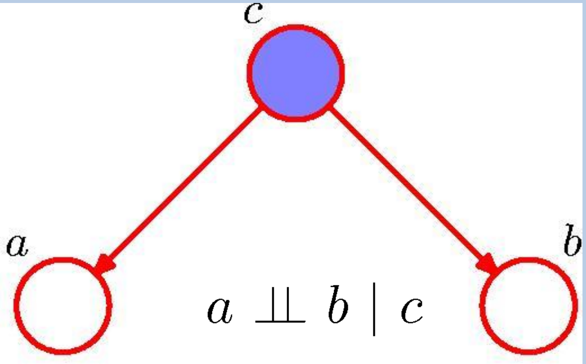
Since c has fixed state it cannot influence (‘change’) a and b anymore.

$$
\begin{aligned}
p(a, b|c) &= \frac{p(a, b, c)}{p(c)} \\
&= p(a|c)p(b|c)
\end{aligned}
$$

so we obtain the conditional independence property $a \perp\!\!\!\perp b \mid c$

Common Cause



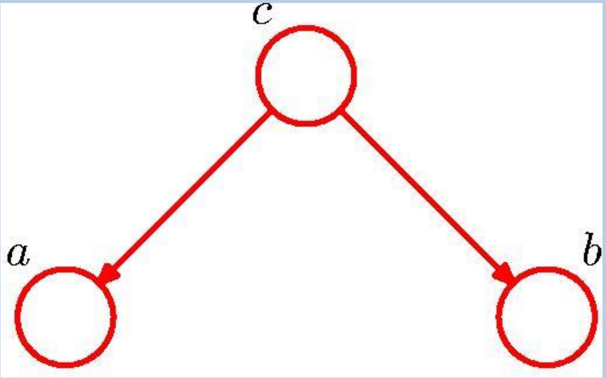


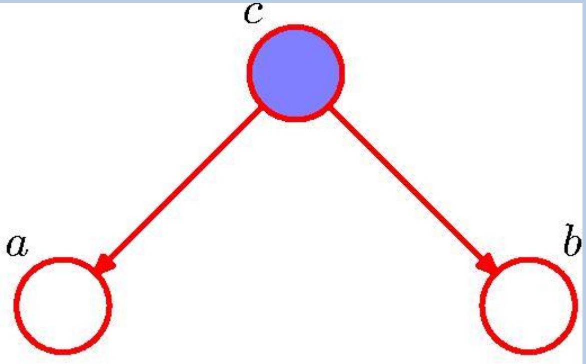
We can provide a simple graphical interpretation of this result by considering the path from **node *a* to node *b* via *c***.

The **node *c*** is said to be ***tail-to-tail*** with respect to this path because the node is connected to the tails of the two arrows.

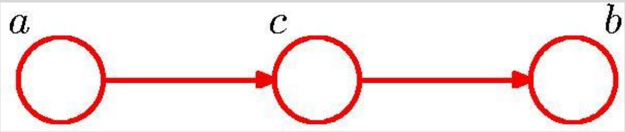
The presence of such path-connecting nodes *a* and *b* causes these nodes to be dependent. However, when we **condition on node *c***, the conditioned node 'blocks' the path from *a* to *b* and causes *a* and *b* to become **conditionally independent**.

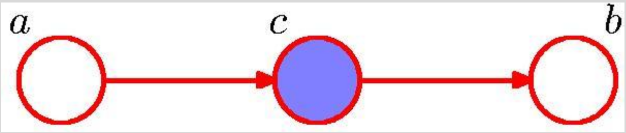
Common Cause

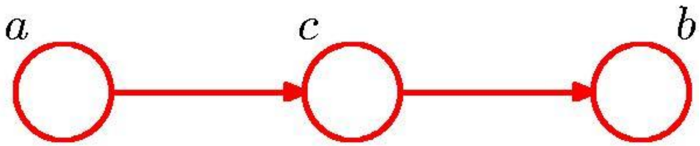




Causal Chains







Joint Distribution

$$p(a, b, c) = p(a)p(c|a)p(b|c)$$

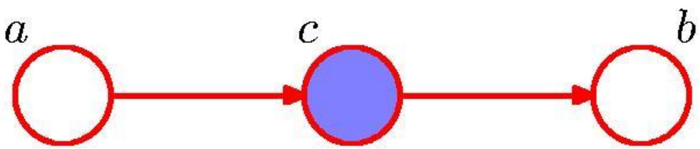
First of all, suppose that none of the variables are observed. Again, we can test to see if *a* and *b* are independent by marginalizing over *c:*

$$p(a, b) = p(a) \sum_c p(c|a)p(b|c)$$

We want the joint probability of just a,b

which in general does not factorize into *p*(*a*)*p*(*b*), and so $a \not\perp\!\!\!\perp b$

Example 2: Causal Chains



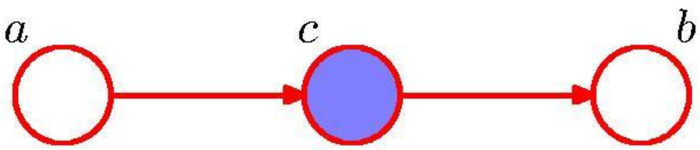
Joint Distribution

$$p(a,b|c) = \frac{p(a,b,c)}{p(c)} = \frac{p(a)p(c|a)p(b|c)}{p(c)}$$

We use Bayes rule

$$\Leftrightarrow p(a\,|\,c) = \frac{p(c\,|\,a)\,p(a)}{p(c)}$$

Example 2: Causal Chains



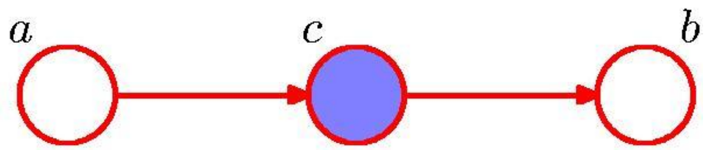
Joint Distribution

$$\begin{aligned} p(a,b|c) &= \frac{p(a,b,c)}{p(c)} \\ &= \frac{p(a)p(c|a)p(b|c)}{p(c)} \\ &= p(a|c)p(b|c) \end{aligned}$$

We use Bayes rule

$$\Leftrightarrow p(a\,|\,c) = \frac{p(c\,|\,a)\,p(a)}{p(c)}$$

Example 2: Causal Chains



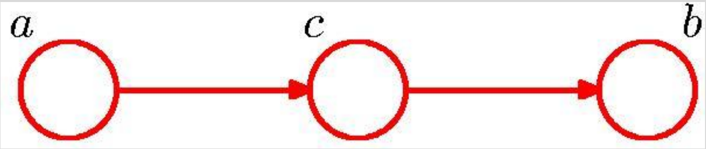
Joint Distribution

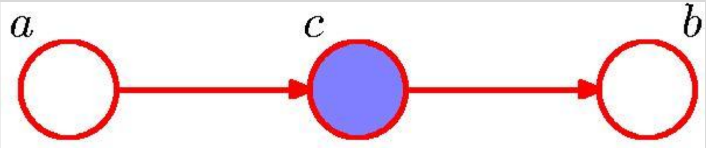
$$
\begin{aligned}
p(a,b|c) &= \frac{p(a,b,c)}{p(c)} \\
&= \frac{p(a)p(c|a)p(b|c)}{p(c)} \\
&= p(a|c)p(b|c)
\end{aligned}
$$

we obtain the conditional independence property $a \perp\!\!\!\perp b \mid c$

As before, we can interpret these results graphically. The node *c* is said to be *head-to-tail* with respect to the path from node *a* to node *b*. Such a path connects nodes *a* and *b* and renders them dependent.

If we now observe *c*, then this observation ‘blocks’ the path from *a* to *b* and so we obtain the conditional independence property $a \perp\!\!\!\perp b \mid c$





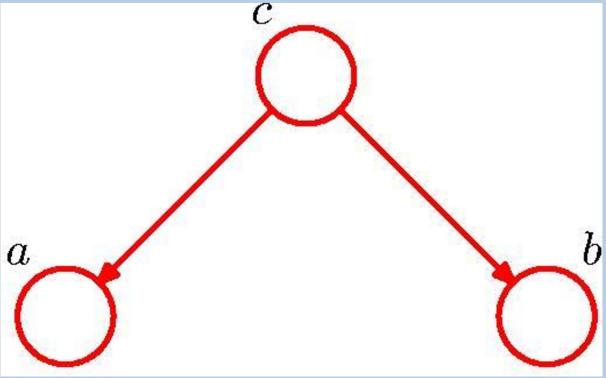
$$a \not\perp\!\!\!\perp b \mid \emptyset$$

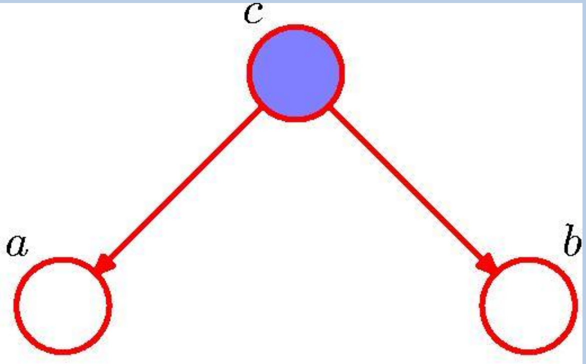
$$a \perp\!\!\!\perp b \mid c$$

The node *c* is said to be *head-to-tail* with respect to the path from node *a* to node *b*. Such a path connects nodes *a* and *b* and renders them dependent.

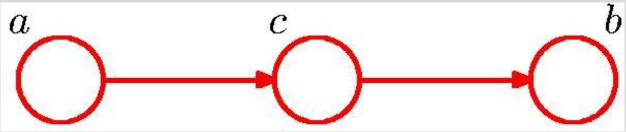
If we now observe *c*, then this observation ‘blocks’ the path from *a* to *b* and so we obtain the conditional independence property

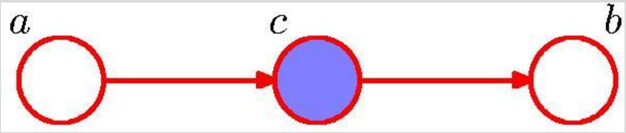
Common Cause



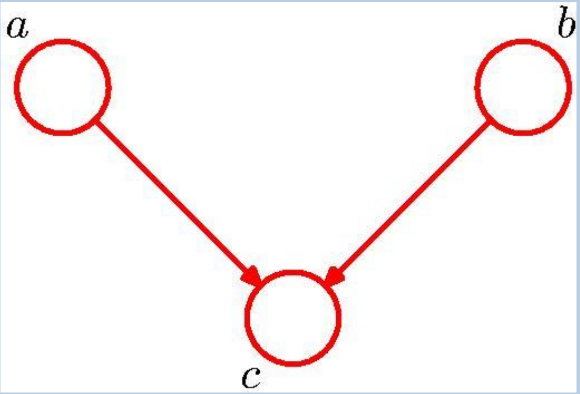


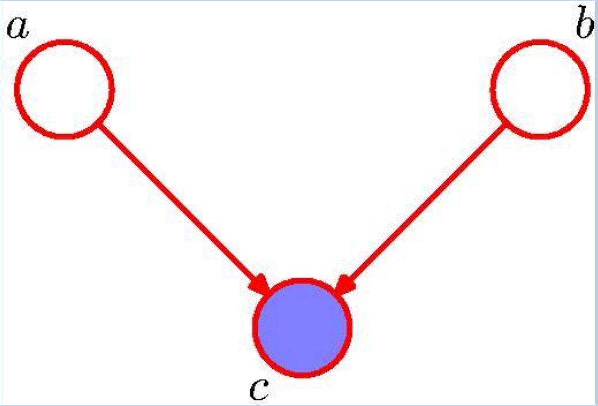
Causal Chains

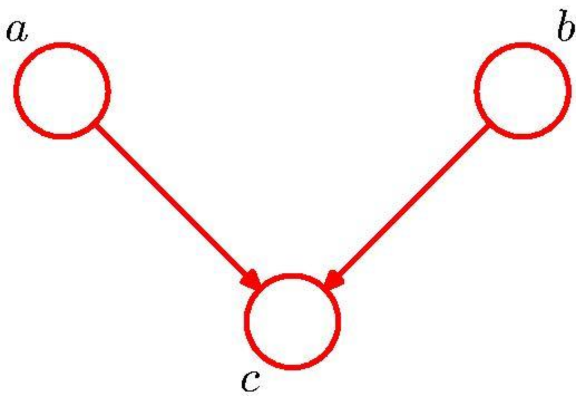




Common Effects





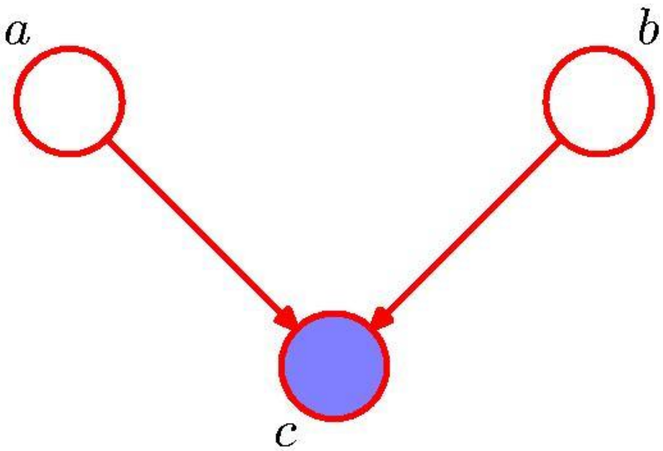


Joint Distribution

$$p(a, b, c) = p(a)p(b)p(c|a, b)$$

Marginalizing over *c*, we obtain:

$$p(a, b) = p(a)p(b) \qquad a \perp\!\!\!\perp b \mid \emptyset$$



Joint Distribution

$$
\begin{aligned}
p(a,b|c) &= \frac{p(a,b,c)}{p(c)} \\
&= \frac{p(a)p(b)p(c|a,b)}{p(c)}
\end{aligned}
$$

$$a \not\perp\!\!\!\perp b \mid c$$

Thus our third example has the opposite behaviour from the first two.

Graphically, we say that node *c* is ***head-to-head*** with respect to the **path from *a* to *b*** because it connects to the heads of the two arrows.

When node *c* is unobserved, it ‘blocks’ the path, and the variables *a* and *b* are independent. However, conditioning on *c* ‘unblocks’ the path and renders *a* and *b* dependent.

In summary:

- a tail-to-tail node or a head-to-tail node leaves a path unblocked unless it is observed, in which case it blocks the path.

- a head-to-head node blocks a path if it is unobserved, but once the node, and/or at least one of its descendants, is observed, the path becomes unblocked.

Neuroinformatics Questionnaire – Q3 – Problem 1+2

Q

1. What is the difference between the conditional independence of a and b given c, and the independence between a and b.

2. How can we determine whether a and b given c are conditionally independent.

Timer (5min): Start Stop

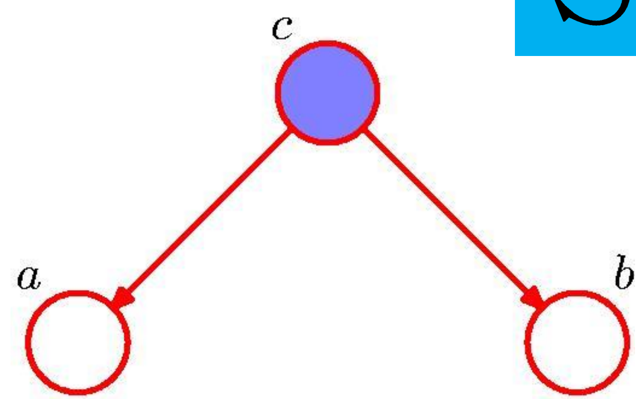
Statistical independence:
Stochastic variables are independent if and only if the joint probability factorizes.

$$if\ p(x_1, x_2, \dots, x_N) = p(x_1)p(x_2) \cdot \dots \cdot p(x_N)\ than\ \ and\ only\ than\ x_1\ to\ x_N \quad are\ independent$$

Statistical conditional independence:
Stochastic variables are conditionally independent if and only if the joint probability factorizes.

$$if\ p(x_1, x_2, \dots, x_N|C) = p(x_1|C)p(x_2|C) \cdot \dots \cdot p(x_N|C)$$
$$than\ \ and\ only\ than\ x_1\ to\ x_N \quad are\ conditonally\ independent\ given\ C$$

Conditional Independence



General procedure to test on conditional independence:

We want to check on conditional dependence of a and b given c.

$$p(a,b \mid c) \stackrel{?}{=} p(a \mid c)\, p(b \mid c)$$

We start by writing:

$$p(a,b \mid c)\, p(c) = p(a,b,c)$$

Than we read the joint probability from the graph:

$$p(a,b,c) = p(c)\, p(a \mid c)\, p(b \mid c)$$

We use that for the Eqn. above:

$$p(a,b \mid c)\, \cancel{p(c)} = \cancel{p(c)}\, p(a \mid c)\, p(b \mid c)$$

$$\Leftrightarrow\; p(a,b \mid c) = p(a \mid c)\, p(b \mid c) \qquad\qquad a \perp\!\!\!\perp b \mid c$$

Neuroinformatics Lecture

Topics:

Independence in GRAPHICAL MODELS II

We wish to ascertain whether a particular conditional independence statement $A \perp\!\!\!\perp B \mid C$ is implied by a given directed acyclic graph. To do so, we consider all possible paths from any node in *A* to any node in *B*.

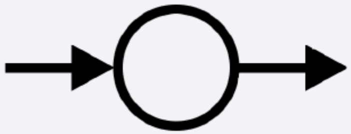
- A, B, and C are non-intersecting subsets of nodes in a directed graph.

- A path from A to B is blocked if it contains a node such that either
  1. the arrows on the path meet either head-to-tail or tail-to-tail at the node, and the node is in the set C, or
  2. the arrows meet head-to-head at the node, and neither the node nor any of its descendants are in the set C.
  3. We say that node *y* is a ***descendant*** of node *x* if there is a path from ***x* to *y*** in which each step of the path **follows the directions of the arrows**.

- If all paths from A to B are blocked, A is said to be d-separated from B by C.

- If A is d-separated from B by C, the joint distribution over all variables in the graph satisfies $A \perp\!\!\!\perp B \mid C$, i. e. that A and B are conditionally independent.

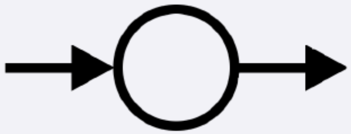
D-separation

head-to-tail	**head-to-head**	**tail-to-tail**
		
path free	path blocked	path free
$a \not\perp\!\!\!\perp b \mid \emptyset$	$a \perp\!\!\!\perp b \mid \emptyset$	$a \not\perp\!\!\!\perp b \mid \emptyset$
not d-separated	d-separated	not d-separated

path is blocked

D-separation

head-to-tail



path free

$a \not\perp\!\!\!\perp b \mid \emptyset$

not d-separated

head-to-head

path blocked

$a \perp\!\!\!\perp b \mid \emptyset$

d-separated

tail-to-tail

path free

$a \not\perp\!\!\!\perp b \mid \emptyset$

not d-separated

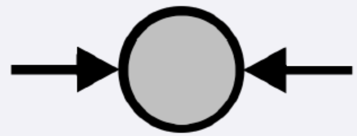
head-to-tail (fixed)

path blocked

$a \perp\!\!\!\perp b \mid c$

d-separated

head-to-head (fixed)



path free

$a \not\perp\!\!\!\perp b \mid c$

not d-separated

tail-to-tail (fixed)

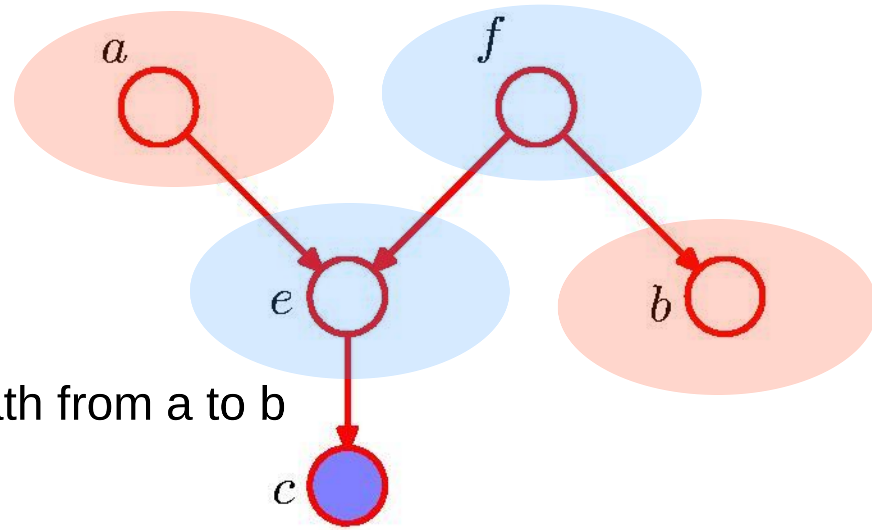
path blocked

$a \perp\!\!\!\perp b \mid c$

d-separated

path is blocked

D-separation – Example 1



Path from a to b

$$a \not\perp\!\!\!\perp b \mid c$$

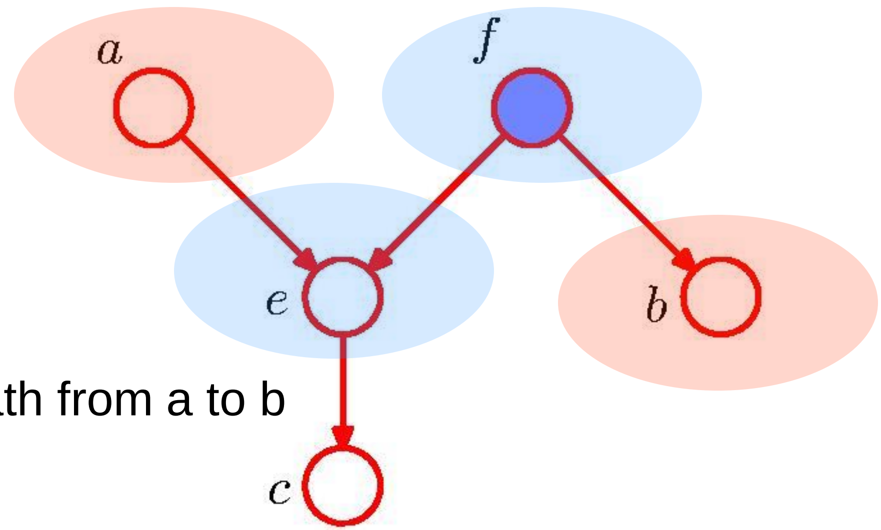
- not blocked by node *f* because it is a tail-to-tail node and not observed

- not blocked by node *e* because it is a head-to-head node, but
  it has a descendant *c* that is in the conditioning set.

  We say that node *y* is a ***descendant*** of node *x* if there is a path from *x* to *y* in which each step of the path follows the directions of the arrows.

- A, B, and C are non-intersecting subsets of nodes in a directed graph. C is the set of observed nodes.
- A path from A to B is blocked if it contains a node such that either
  1. the arrows on the path meet either head-to-tail or tail-to-tail at the node, and the node is in the set C of observed nodes, or
  2. the arrows meet head-to-head at the node, and neither the node nor any of its descendants are in the set C.
- If all paths from A to B are blocked, A is said to be d-separated from B by C.

D-separation – Example 2



Path from a to b

$$a \perp\!\!\!\perp b \mid f$$

- blocked by node *f*, because this is a tail-to-tail node that is observed
- blocked by node *e*, because this is a head-to-head node and neither the node nor any of its descendants are in the set C.

We say that node *y* is a ***descendant*** of node *x* if there is a path from *x* to *y* in which each step of the path follows the directions of the arrows.

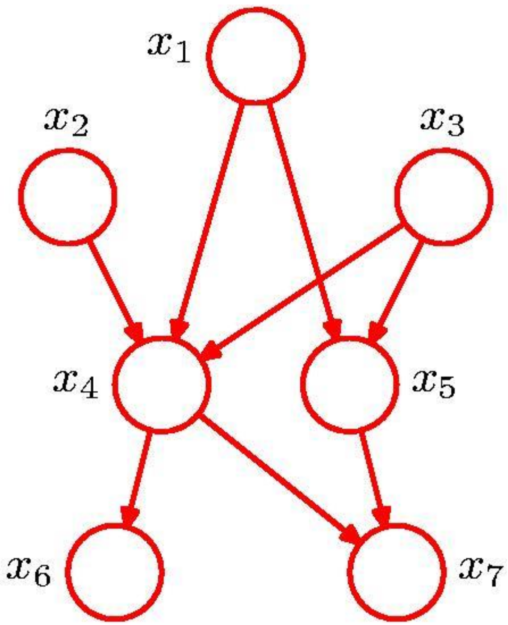
- A, B, and C are non-intersecting subsets of nodes in a directed graph. C is the set of observed nodes.
- A path from A to B is blocked if it contains a node such that either
  1. the arrows on the path meet either head-to-tail or tail-to-tail at the node, and the node is in the set C of observed nodes, or
  2. the arrows meet head-to-head at the node, and neither the node nor any of its descendants are in the set C.
- If all paths from A to B are blocked, A is said to be d-separated from B by C.

Q

3. Describe the concept of graphical model

4. Explain the d-separation property.

Timer (5min): Start Stop



$$p(x_1, \ldots, x_7) = p(x_1)p(x_2)p(x_3)p(x_4|x_1, x_2, x_3) \\ p(x_5|x_1, x_3)p(x_6|x_4)p(x_7|x_4, x_5)$$

General Factorization

The joint distribution defined by a graph is given by the product, over all of the nodes of the graph, of a conditional distribution for each node conditioned on the variables corresponding to the parents of that node in the graph. Thus, for a graph with *K* nodes, the joint distribution is given by:

$$p(\mathbf{x}) = \prod_{k=1}^{K} p(x_k|\mathrm{pa}_k)$$

pa_k denotes the set of parents of x_k

Q

5. Tick off correct statements.

(a) (6 pts) There might be more than one correct statements.

- ☐ A path from A to B is blocked if it contains a node such that either the arrows on the path meet either head-to-tail or tail-to-tail at the node, and the node is in the set C, or the arrows meet head-to-head at the node, and neither the node nor any of its descendants are in the set C.

- ☐ A path from A to B is blocked if it contains a node such that either the arrows on the path meet either head-to-head at the node, and the node is in the set C, or the arrows meet tail-to-head or tail-to-tail at the node, and neither the node nor any of its descendants are in the set C.

- ☐ A path is blocked if it connected with at least 2 other nodes

- ☐ If all paths from A to B are blocked, A is said to be d-separated from B by C

- ☐ If any paths from A to B is blocked, A is said to be d-separated from B by C

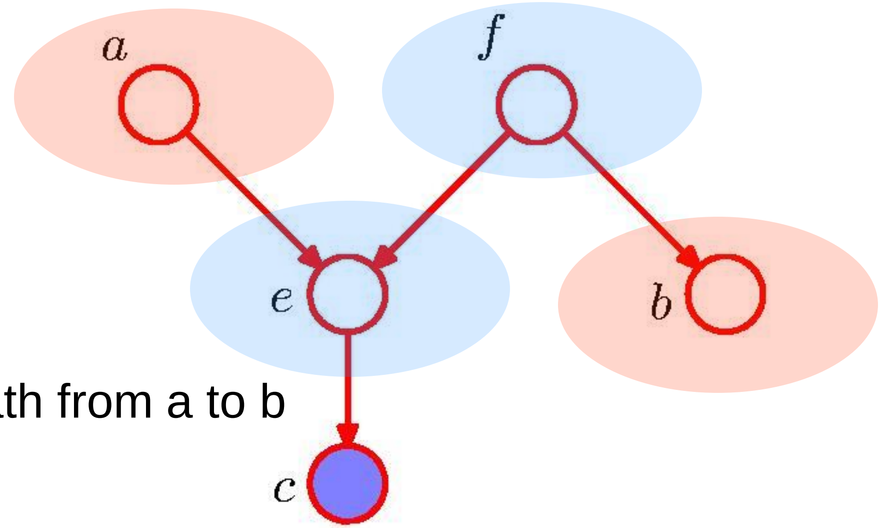
Timer (5min): Start Stop

5. Tick off correct statements.

(a) (6 pts) There might be more than one correct statements.

- [x] A path from A to B is blocked if it contains a node such that either the arrows on the path meet either head-to-tail or tail-to-tail at the node, and the node is in the set C, or the arrows meet head-to-head at the node, and neither the node nor any of its descendants are in the set C.
- [ ] A path from A to B is blocked if it contains a node such that either the arrows on the path meet either head-to-head at the node, and the node is in the set C, or the arrows meet tail-to-head or tail-to-tail at the node, and neither the node nor any of its descendants are in the set C.
- [ ] A path is blocked if it connected with at least 2 other nodes
- [ ] If all paths from A to B are blocked, A is said to be d-separated from B by C
- [ ] If any paths from A to B is blocked, A is said to be d-separated from B by C

D-separation – Example 1



Path from a to b

$$a \not\perp\!\!\!\perp b \mid c$$

- not blocked by node *f* because it is a tail-to-tail node and not observed

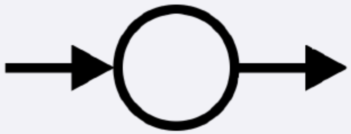
- not blocked by node *e* because it is a head-to-head node, but
  it has a descendant *c* that is in the conditioning set.

  We say that node *y* is a ***descendant*** of node *x* if there is a path from *x* to *y* in which each step of the path follows the directions of the arrows.

- A, B, and C are non-intersecting subsets of nodes in a directed graph. C is the set of observed nodes.
- A path from A to B is blocked if it contains a node such that either
  1. the arrows on the path meet either head-to-tail or tail-to-tail at the node, and the node is in the set C of observed nodes, or
  2. the arrows meet head-to-head at the node, and neither the node nor any of its descendants are in the set C.
- If all paths from A to B are blocked, A is said to be d-separated from B by C.

D-separation

head-to-tail



path free

$a \not\perp\!\!\!\perp b \mid \emptyset$

not d-separated

head-to-head

path blocked

$a \perp\!\!\!\perp b \mid \emptyset$

d-separated

tail-to-tail

path free

$a \not\perp\!\!\!\perp b \mid \emptyset$

not d-separated

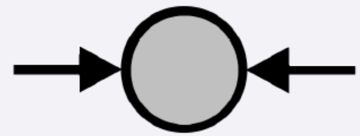
head-to-tail (fixed)

path blocked

$a \perp\!\!\!\perp b \mid c$

d-separated

head-to-head (fixed)



path free

$a \not\perp\!\!\!\perp b \mid c$

not d-separated

tail-to-tail (fixed)

path blocked

$a \perp\!\!\!\perp b \mid c$

d-separated

path is blocked

5. Tick off correct statements.

(a) (6 pts) There might be more than one correct statements.

- [x] A path from A to B is blocked if it contains a node such that either the arrows on the path meet either head-to-tail or tail-to-tail at the node, and the node is in the set C, or the arrows meet head-to-head at the node, and neither the node nor any of its descendants are in the set C.
- [ ] A path from A to B is blocked if it contains a node such that either the arrows on the path meet either head-to-head at the node, and the node is in the set C, or the arrows meet tail-to-head or tail-to-tail at the node, and neither the node nor any of its descendants are in the set C.
- [ ] A path is blocked if it connected with at least 2 other nodes
- [x] If all paths from A to B are blocked, A is said to be d-separated from B by C
- [ ] If any paths from A to B is blocked, A is said to be d-separated from B by C

6. Explain the concept of 'explain away'