Neuroinformatics Lecture (L5)

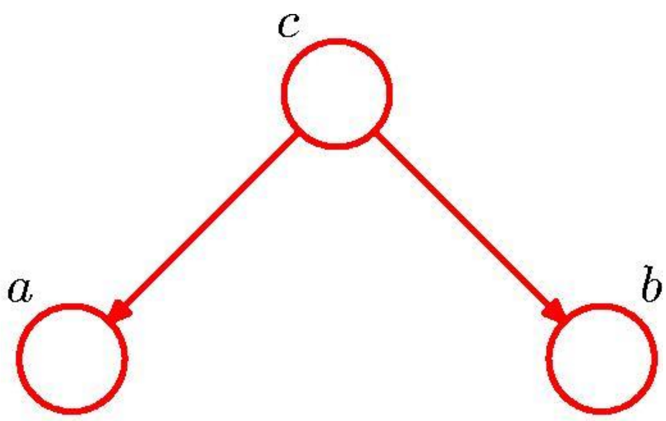
Prof. Dr. Gordon Pipa

Institute of Cognitive Science, University of Osnabrück

R

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 factories.

$$If\ p(x_1, x_2, \ldots, x_N) = p(x_1)p(x_2) \cdot \ldots \cdot p(x_N), then\ and\ only\ then\ x_1\ to\ x_N\ are\ independent.$$

Statistical independence:

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

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

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

Statistical conditional independence:

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

We use the following symbols in that case:

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

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

Independence

Stochastic variables are independent if and only if the joint probability factories.

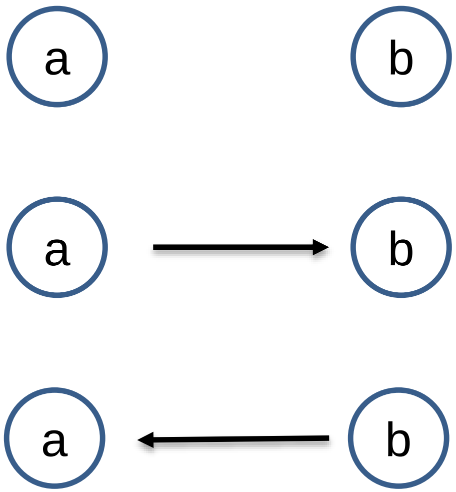
$$If\ p(x_1, x_2, \dots, x_N) = p(x_1)p(x_2) \cdot \dots \cdot p(x_N), then\ and\ only\ then\ x_1\ to\ x_N\ are\ independent$$

The joint distribution (pa$_k$ denotes the set of parents of x_k):

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

are a an b dependent ?

Is the denpendence causal ?



Independence

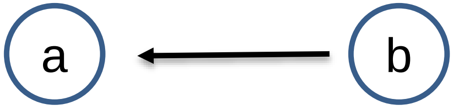
Stochastic variables are independent if and only if the joint probability factories.

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

The joint distribution (pa_k denotes the set of parents of x_k): $$p(\mathbf{x}) = \prod_{k=1}^{K} p(x_k|\mathrm{pa}_k)$$

		are a an b dependent ?	Is the denpen-dence causal ?
	$p(a,b) = p(a)p(b)$	No	No

1: The probability immediately factorizes → independence
2: Since the nodes are independent the cant influence each other



Independence

Stochastic variables are independent if and only if the joint probability factories.

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

The joint distribution (pa_k denotes the set of parents of x_k): $$p(\mathbf{x}) = \prod_{k=1}^{K} p(x_k|\text{pa}_k)$$

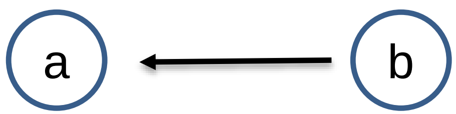
are a an b dependent ? | Is the denpen-dence causal ?

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

No | No

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

Yes



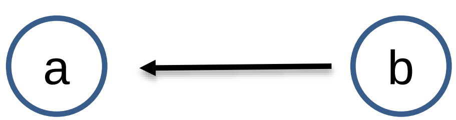
1: The probability does not factorize → dependence

Independence

Stochastic variables are independent if and only if the joint probability factories.

$$If\ p(x_1, x_2, \ldots, x_N) = p(x_1)p(x_2) \cdot \ldots \cdot p(x_N), then\ and\ only\ then\ x_1\ to\ x_N\ are\ independent$$

The joint distribution (pa_k denotes the set of parents of x_k): $$p(\mathbf{x}) = \prod_{k=1}^{K} p(x_k|\mathrm{pa}_k)$$

		are a an b dependent ?	Is the denpen-dence causal ?
	$p(a,b) = p(a)p(b)$	No	No
	$p(a,b) = p(a)p(b\|a)$	Yes	
	$p(a,b) = p(a\|b)p(b)$	Yes	

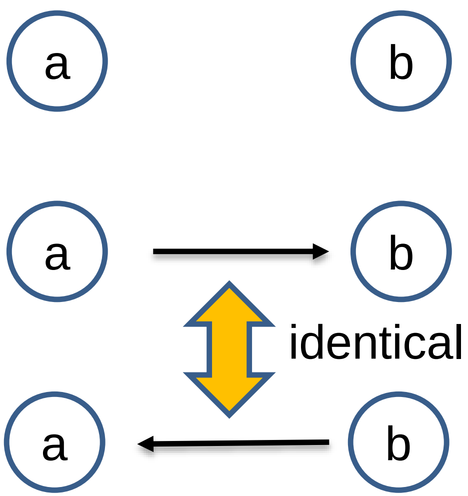
1: The probability does not factorize → dependence

Independence

Stochastic variables are independent if and only if the joint probability factories.

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

The joint distribution (pa$_k$ denotes the set of parents of x_k): $$p(\mathbf{x}) = \prod_{k=1}^{K} p(x_k|\mathrm{pa}_k)$$



	are a an b dependent ?	Is the denpen-dence causal ?
$p(a,b) = p(a)p(b)$	No	No
$p(a,b) = p(a)p(b\|a)$	Yes	No
$p(a,b) = p(a\|b)p(b)$	Yes	No

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

Can we infer the causal influence ?

The joint distribution (pa_k denotes the set of parents of x_k): $$p(\mathbf{x}) = \prod_{k=1}^{K} p(x_k|\mathrm{pa}_k)$$

$$a \longrightarrow b \longrightarrow c$$

Does that mean:
a drives b and b drives c ? NO

It means: a,b are dependent and b,c are dependent

Proof:

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

Can we infer the causal influence ?

The joint distribution (pa$_k$ denotes the set of parents of x_k):

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

$$a \longrightarrow b \longrightarrow c$$

$$a \longleftarrow b \longrightarrow c$$

Does that mean:
a drives b and b drives c ?   NO

It means: a,b are dependent and b,c are dependent

Proof:

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

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

Can we infer the causal influence ?

The joint distribution (pa$_k$ denotes the set of parents of x_k):

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

$a \rightarrow b \rightarrow c$

$a \leftarrow b \rightarrow c$

$a \leftarrow b \leftarrow c$

Does that mean:
a drives b and b drives c ? NO

It means: a,b are dependent and b,c are dependent

Proof:

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

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

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

Can we infer the causal influence ?

The joint distribution (pa$_k$ denotes the set of parents of x_k): $$p(\mathbf{x}) = \prod_{k=1}^{K} p(x_k|\mathrm{pa}_k)$$

$$a \longrightarrow b \longrightarrow c$$

$$a \longleftarrow b \longrightarrow c$$

$$a \longleftarrow b \longleftarrow c$$

Interpretation:

a,b are dependent and c,b are dependent

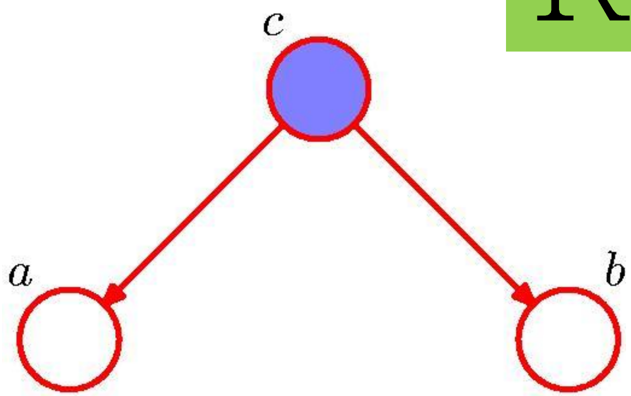
That can be the case because of:

1) a drives b
2) b drives a
3) c drives b
4) b drives c

We do not know which of the cases is true, but we know at least one of each pairwise relations ships is true.

Conditional Independence

General procedure to test on conditional independence:



We start by writing: $p(a,b \mid c)\,p(c) = p(a,b,c)$

Then 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 equation above:

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

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

D-separation

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.

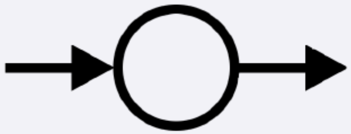
- A path from A to B is blocked if it contains a node such that either
  1. the arrows on the path meet either head-to-tail or tail-to-tail at the node, and the node is in the set C, or
  2. the arrows meet head-to-head at the node, and neither the node nor any of its descendants are in the set C.
 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



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

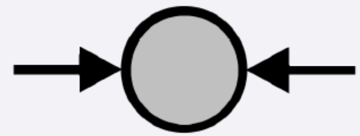
head-to-tail (fixed)

path blocked

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

d-separated

head-to-head (fixed)

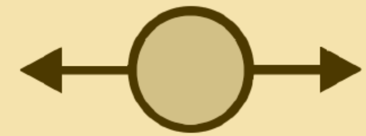


path free

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

not d-separated

tail-to-tail (fixed)



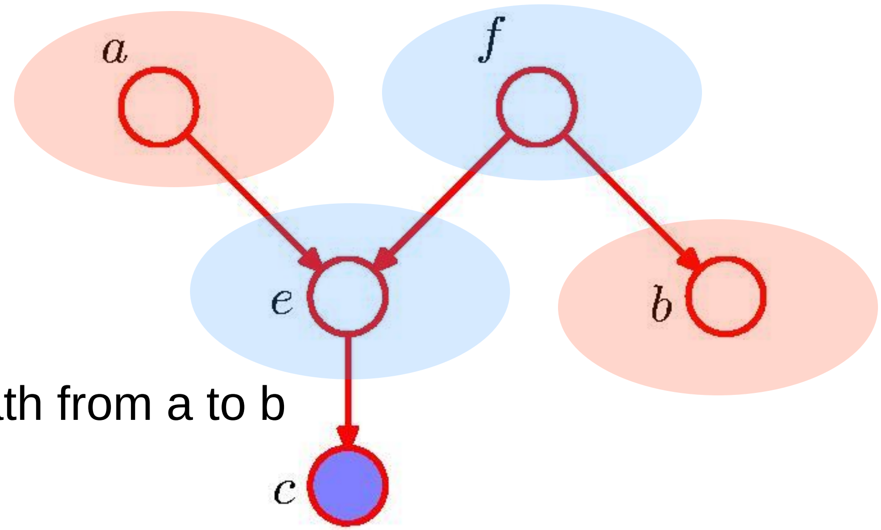
path blocked

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

d-separated

= path is blocked

D-separation – Example 1



Path from a to b

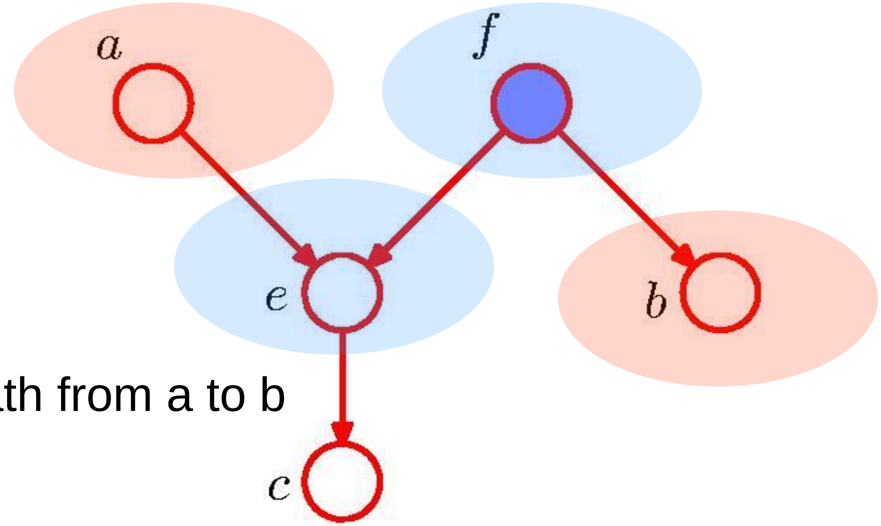
$$a \not\perp\!\!\!\perp b \mid c$$

- 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.

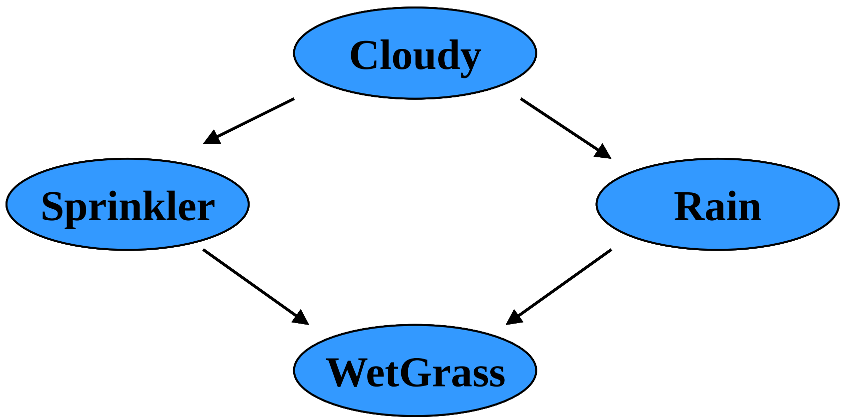
We say that node *y* is a ***descendant*** of node *x* if there is a path from *x* to *y* in which each step of the path follows the directions of the arrows.

- 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.

Representing Knowledge Example

$$p(C)$$

$$p(S, C) = p(S|C)\mathrm{p(c)}$$



$$p(R, C) = p(R|C)\mathrm{p(c)}$$

In general:

$$p(S, C, R, W) = p(W|S, C, R)p(R|C, S)p(S|C)p(C)$$

using conditional independence

$$p(S, C, R, W) = p(W|S, R)p(R|C)p(S|C)p(C)$$

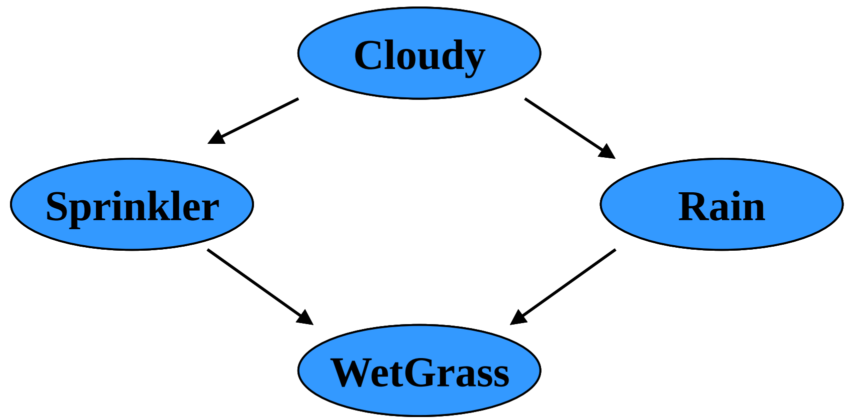
The joint distribution (pa_k denotes the set of parents of x_k):

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

Representing Knowledge Example

$$p(C)$$

$$p(S,C) = p(S|C)\text{p(c)}$$



$$p(R,C) = p(R|C)\text{p(c)}$$

using conditional independence

$$p(S,C,R,W) = p(W|S,R)p(R|C)p(S|C)p(C)$$

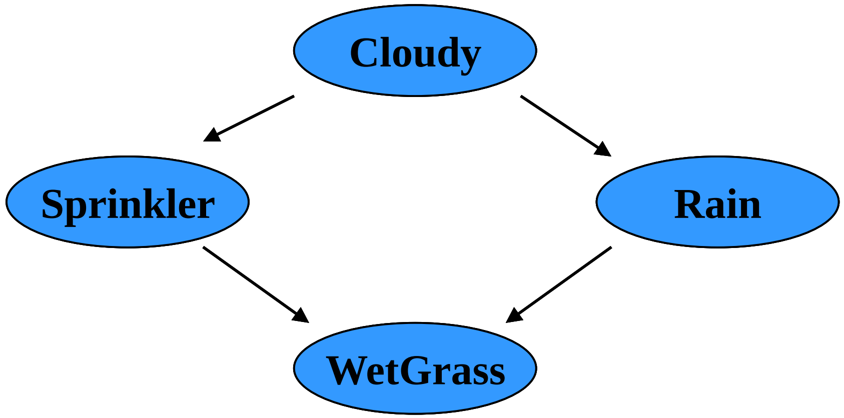
The topology of the network defines the conditional probability table (CPT) for each node.

Representing Knowledge Example

$p(C)$

P(C=F)	P(C=T)
0.5	0.5

$p(S,C) = p(S|C)\mathrm{p}(\mathrm{c})$



$p(R,C) = p(R|C)\mathrm{p}(\mathrm{c})$

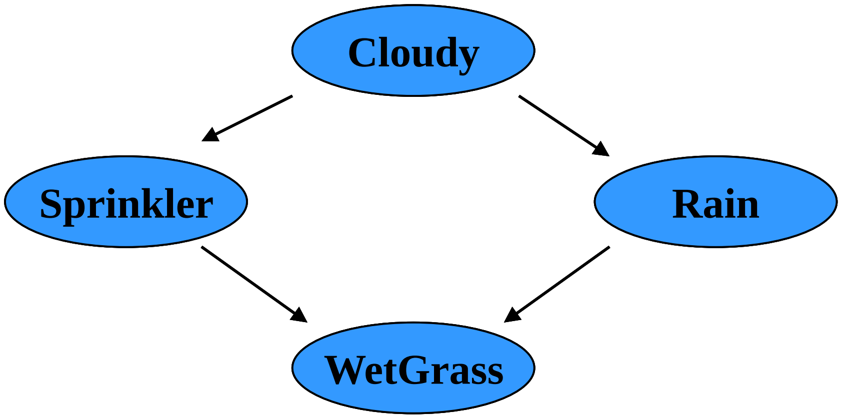
$$p(S,C,R,W) = p(W|S,R)p(R|C)p(S|C)p(C)$$

E.g. C=T means S is true ('it is cloudy') / S=F means S is false ('it is not cloudy')

Representing Knowledge Example

P(C=F)	P(C=T)
0.5	0.5

C	P(S=F)	P(S=T)
F	0.5	0.5
T	0.9	0.1



C	P(R=F)	P(R=T)
F	0.8	0.2
T	0.2	0.8

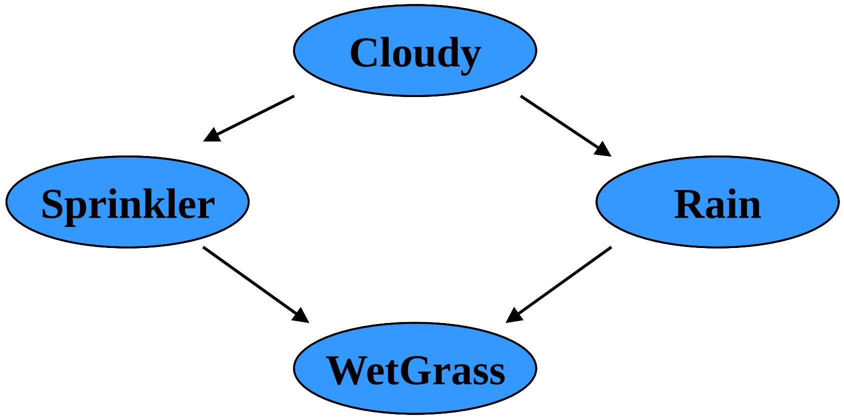
- Each row in the table contains the conditional probability of each node value for a conditioning case.
- Each row must sum to 1 because the entries represent an exhaustive set of cases for the variable.
- A conditioning case is a possible combination of values for the parent nodes.

E.g. C=T means S is true ('it is cloudy') / S=F means S is false ('it is not cloudy')

Representing Knowledge Example

P(C=F)	P(C=T)
0.5	0.5

C	P(S=F)	P(S=T)
F	0.5	0.5
T	0.9	0.1



C	P(R=F)	P(R=T)
F	0.8	0.2
T	0.2	0.8

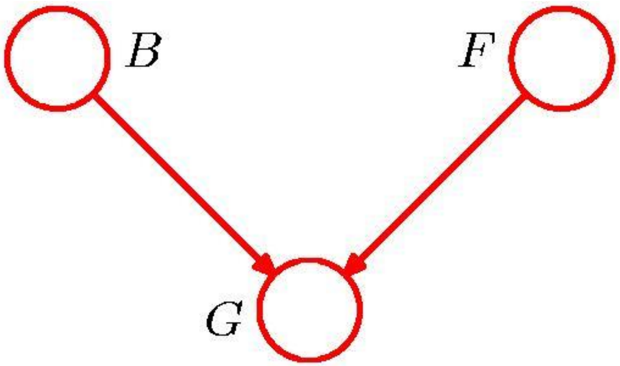
S R	P(W=F)	P(W=T)
F F	1.0	0.0
T F	0.1	0.9
F T	0.1	0.9
T T	0.01	0.99

E.g. C=T means S is true (‘it is cloudy’) / S=F means S is false (‘it is not cloudy’)

It is worth spending a moment to understand further the possible suprising behavior of the graph. Consider a particular instance of such a graph corresponding to a problem with three binary random variables relating to the fuel system on a car.

The variables are called:

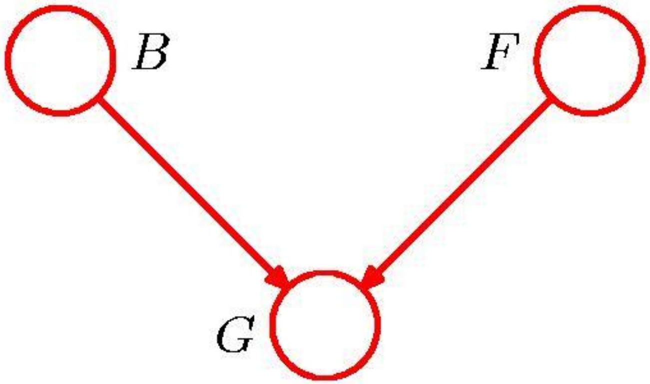
- ***B, battery state*** that is either charged (B = 1) or flat (B = 0)
- ***F, fuel tank state*** that is either full of fuel (F = 1) or empty (F = 0)
- ***G, electric fuel gauge state*** that indicates either full (G = 1) or empty (G=0)



$$
\begin{aligned}
p(B=1) &= 0.9 \\
p(F=1) &= 0.9
\end{aligned}
$$

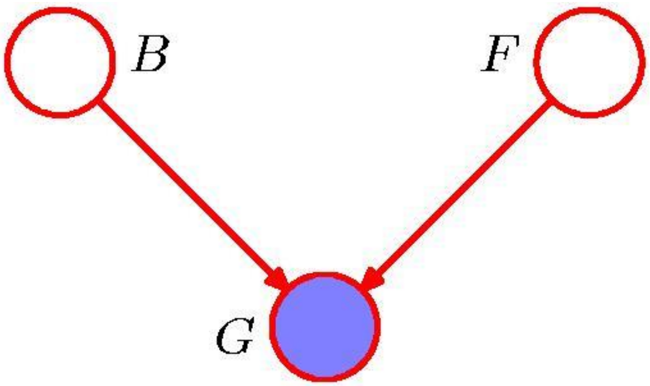
and hence

$$
p(F=0) = 0.1
$$



$$
\begin{aligned}
p(G=1|B=1,F=1) &= 0.8 \\
p(G=1|B=1,F=0) &= 0.2 \\
p(G=1|B=0,F=1) &= 0.2 \\
p(G=1|B=0,F=0) &= 0.1
\end{aligned}
$$

B = Battery (0=flat, 1=fully charged)
F = Fuel Tank (0=empty, 1=full)
G = Fuel Gauge Reading
(0=empty, 1=full)



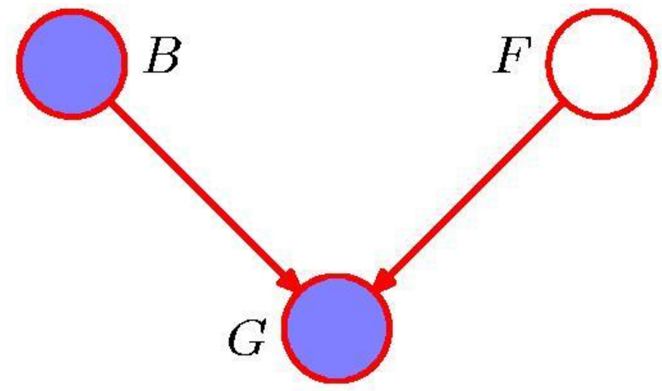
$$p(F=0) = 0.1$$

$$p(F=0|G=0) \simeq 0.257$$

Probability of an empty tank increased by observing G = 0.

$$p(F = 0|G = 0) > p(F = 0)$$

Thus observing that the gauge reads empty, makes it more likely that the tank is indeed empty.



$$
\begin{aligned}
p(F = 0) &= 0.1 \\
p(F = 0|G = 0) &\simeq 0.257 \\
p(F = 0|G = 0, B = 0) &\simeq 0.111
\end{aligned}
$$

Probability of an empty tank reduced by observing B = 0.

This is referred to as "explaining away":
Finding out that the battery is flat *explains away* the observation that the fuel gauge reads empty.

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

6. Explain the concept of 'explain away'

Timer (5min): Start — Stop

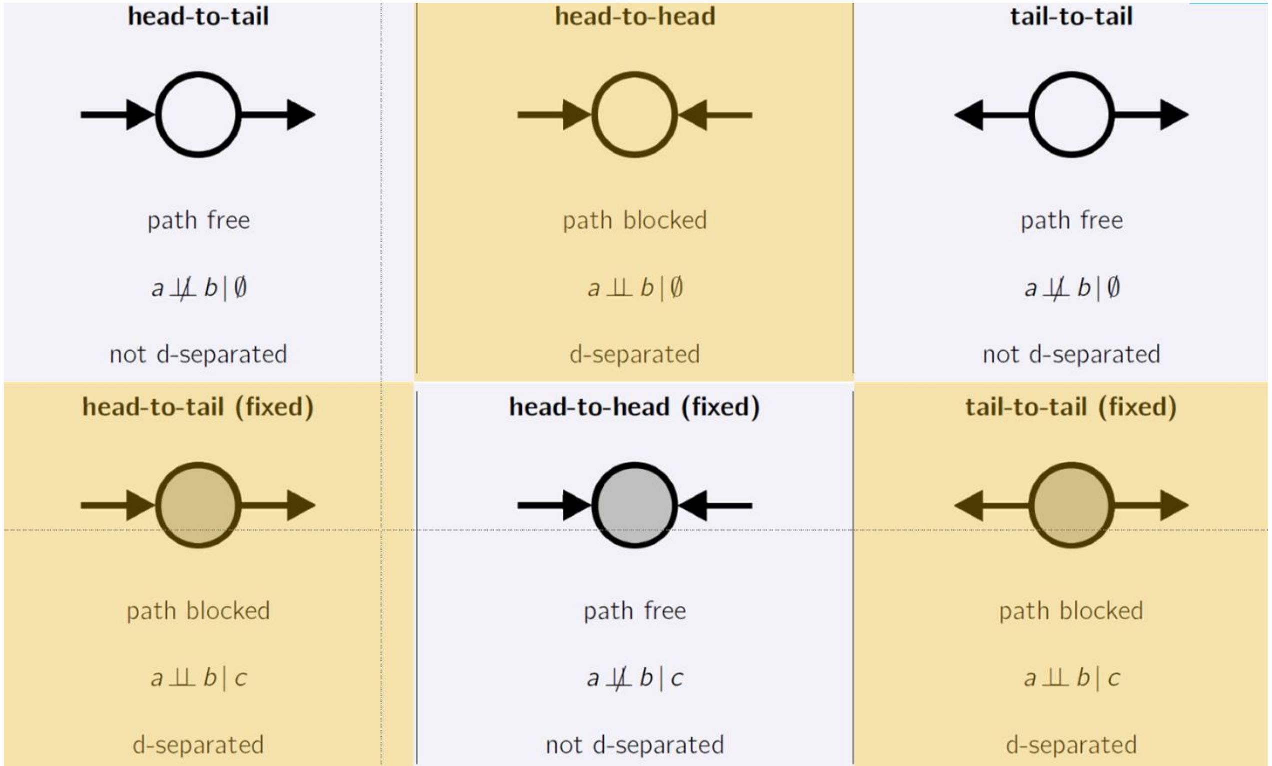
5. Tick off correct statements.

(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

6. Explain the concept of 'explain away'

- 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.

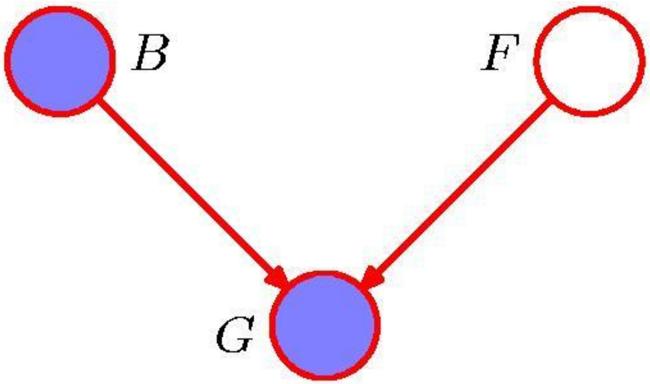


5. Tick off correct statements.

(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'



$$p(F=0) = 0.1$$

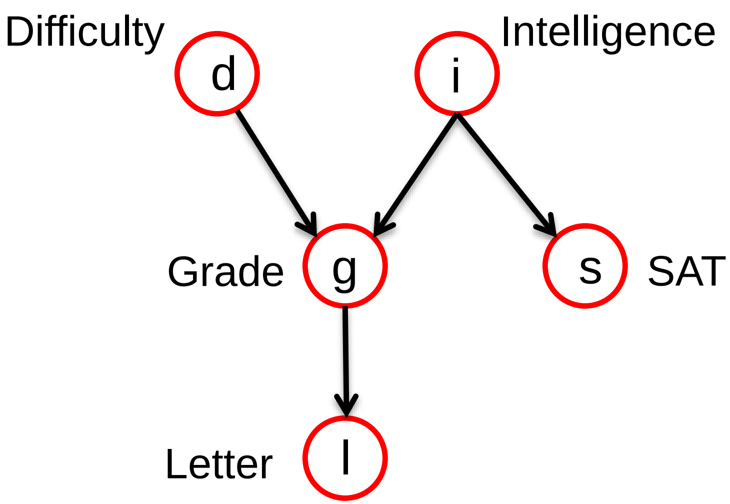
$$p(F=0|G=0) \simeq 0.257$$

$$p(F=0|G=0,B=0) \simeq 0.111$$

Probability of an empty tank reduced by observing B = 0.

This is referred to as "explaining away":
Finding out that the battery is flat *explains away* the observation that the fuel gauge reads empty.

An Example



$$P(d,i,g,s,l) =$$
$$p(d)p(i)p(g\,|\,i,d)p(s\,|\,i)\boxed{p(l\,|\,g)}$$

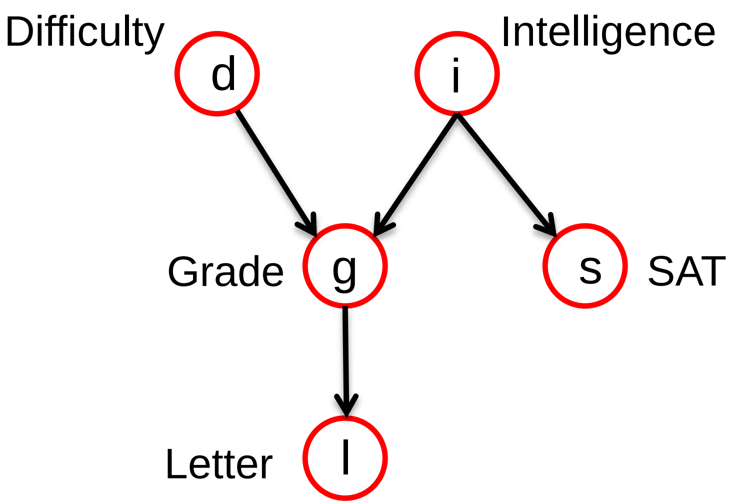
The story:

You need a recommendation letter from your professor. Before you ask for the letter, you start thinking about the different influential factors for the outcome of the letter. You realize that you need to take a probabilistic point of view since there are random or unobserved factors in the process.

- You know that your professor is going to use only your grade for the recommendation. The mapping is probabilistic, because it depends on the a relative scale, compared to other recommendation letters that were written. However, you do not know about the grades of the other students. Therefore you have incomplete knowledge about this, in principle, deterministic problem.

→ A node depends directly only on its parents, here we have just one direct parent, that is the grade.

An Example



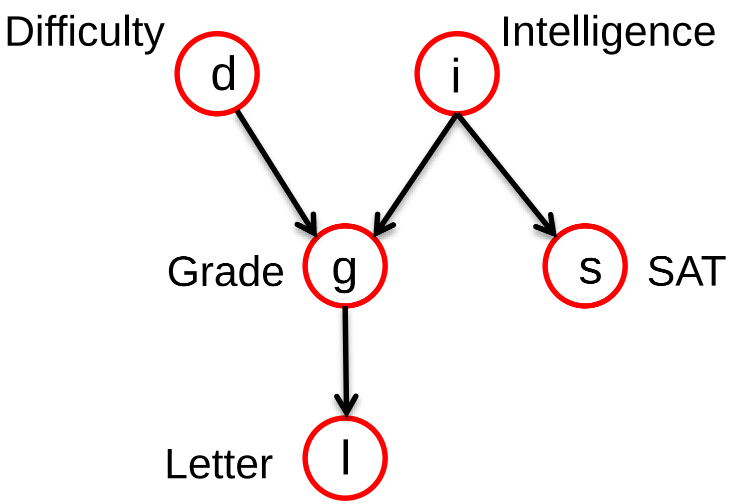
$$P(d,i,g,s,l) =$$
$$p(d)\,p(i)\,p(g\,|\,i,d)\,p(s\,|\,i)\,p(l\,|\,g)$$

The SAT is a standardized test for most college admissions in the United States.

The story:

You need a recommendation letter from your professor. Before you ask for the letter you start thinking about the different influential factors for the outcome of the letter. You realize that you need to take a probabilistic point of view, since there are random or unobserved factors in the process.

- You know that your professor is going to use only your grade for the recommendation.
- You assume that your grade has a dependence on two factors. First, your intelligence and second, the difficulty of the exam.
- Because you wanted to apply in USA, you took the SAT. You know that the SAT is standardized. Therefore you do not have to consider uncertain difficulty for the SAT.

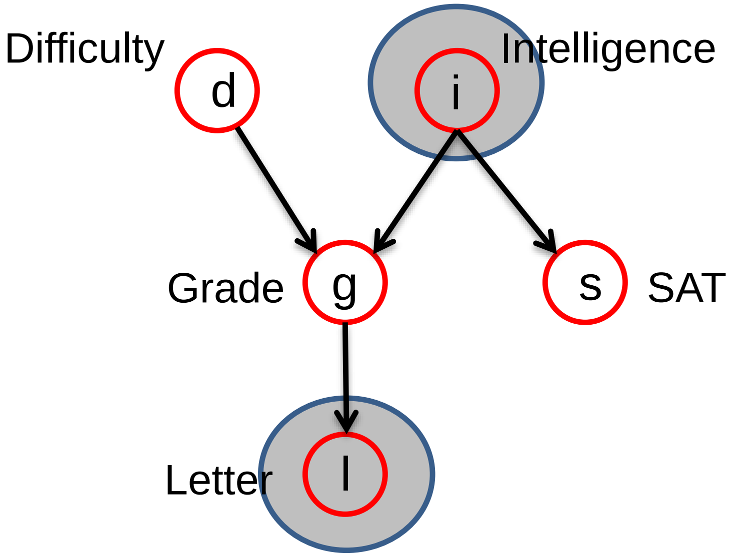


$$P(d,i,g,s,l) =$$
$$p(d)p(i)p(g \mid i,d)p(s \mid i)p(l \mid g)$$

The story:

You need a recommendation letter from your professor. Before you ask for the letter, you start thinking about the different influential factors for the outcome of the letter. You realize that you need to take a probabilistic point of view, since there are random or unobserved factors in the process.

- You know that your professor is going to use only your grade for the recommendation.
- You assume that your grade has a dependence on two factors. First, your intelligence and second, the difficulty of the exam.
- Because you wanted to apply in USA, you took the SAT. You know that the SAT is standardized. Therefore you do not have to consider uncertain difficulty for the SAT.
- You assume your intelligence and the difficulty of the exam are independent of any of the other factors.



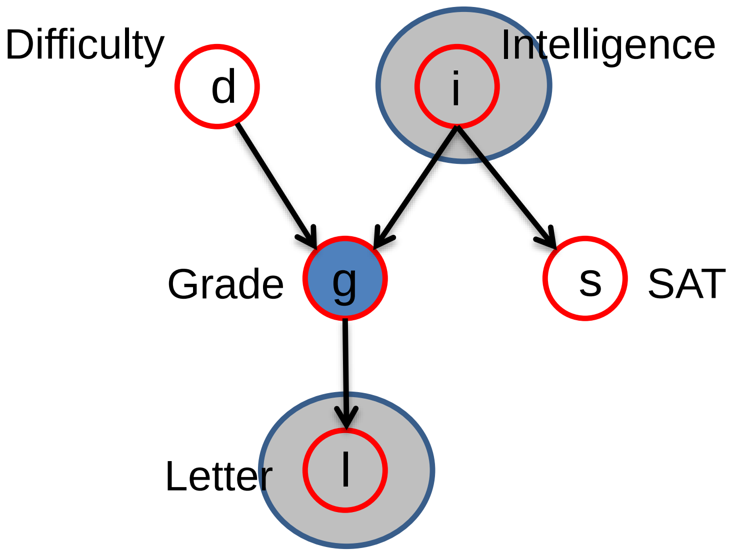
$$P(d,i,g,s,l) =$$
$$p(d)p(i)p(g \mid i,d)p(s \mid i)p(l \mid g)$$

Q: Is the letter dependent on your intelligence?

Is the path from i to l blocked?

Path is free →
Letter is dependent on your intelligence

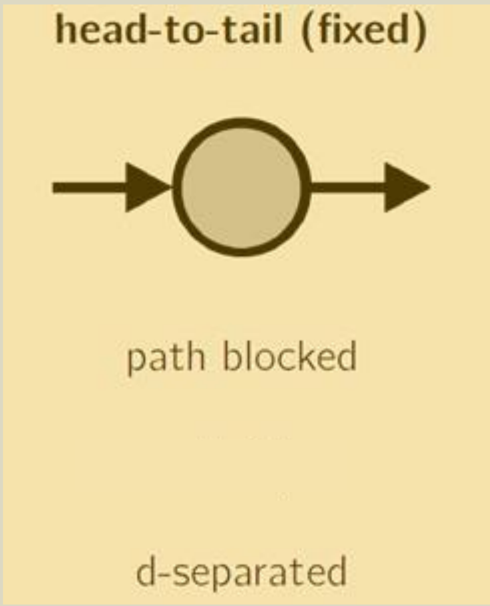
An Example



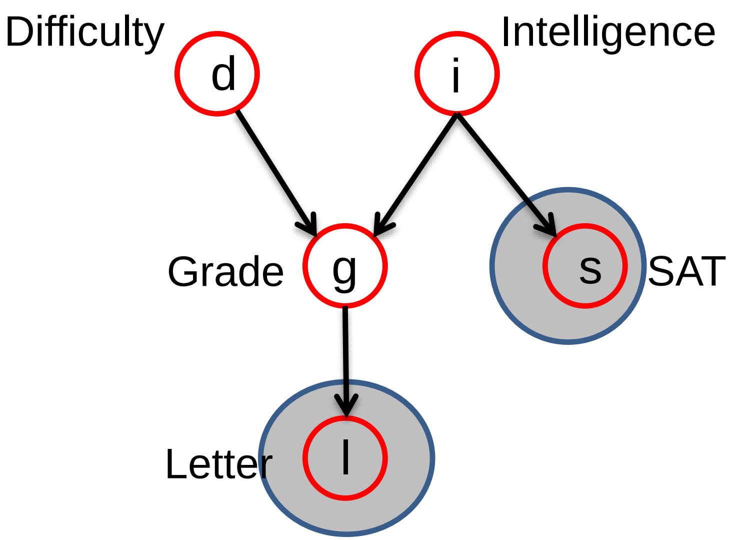
$$P(d,i,g,s,l) =$$
$$p(d)p(i)p(g\,|\,i,d)p(s\,|\,i)p(l\,|\,g)$$

Q: Is the letter dependent on your intelligence, after you wrote the exam?

Is the path from i to l blocked, given g?



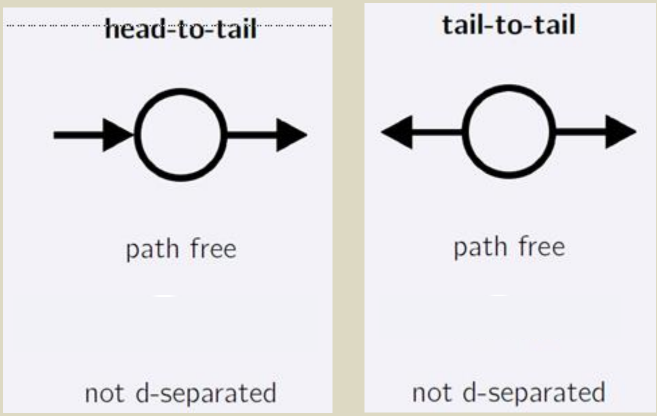
Path is blocked→
Letter is conditionally independent on your intelligence, given the outcome of the exam.



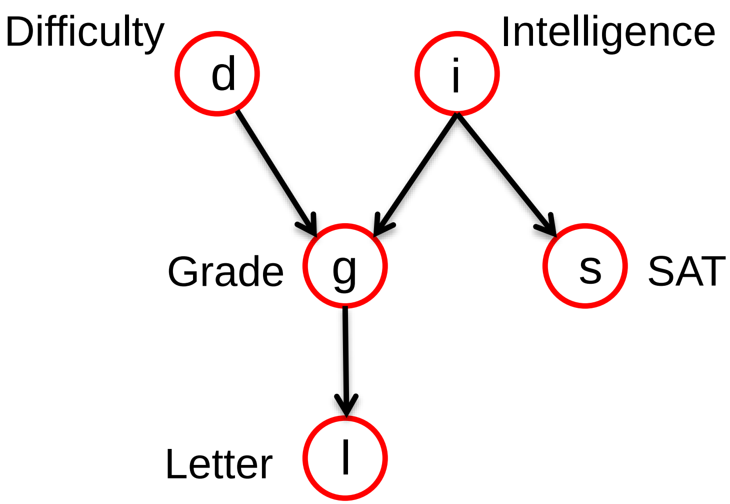
$$P(d,i,g,s,l) =$$
$$p(d)p(i)p(g\,|\,i,d)p(s\,|\,i)p(l\,|\,g)$$

Q: Is the letter dependent on your SAT score?

Is the path from s to l blocked?



Path is free →
Letter is dependent on your SAT



$$P(d,i,g,s,l) =$$
$$p(d)p(i)p(g\,|\,i,d)p(s\,|\,i)p(l\,|\,g)$$

Q: Is the letter dependent on your SAT score?

Is the path from s to l blocked?

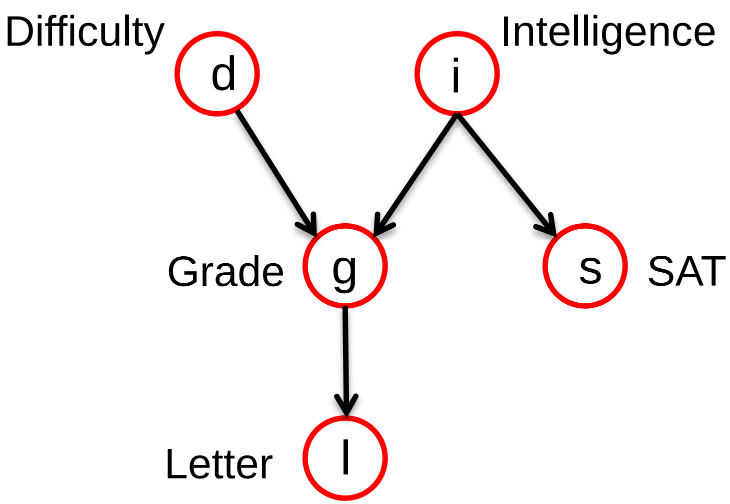
Path is free →
Letter is dependent on your SAT

How is that possible, we said that the letter is only dependent on the grade.

Yes, but the grade is dependent on the intelligence, and a high intelligence makes both a good grade and a high SAT score more likely.

Or with other words, knowing the SAT tells us something about both the intelligence and the grade, since they have a common cause.

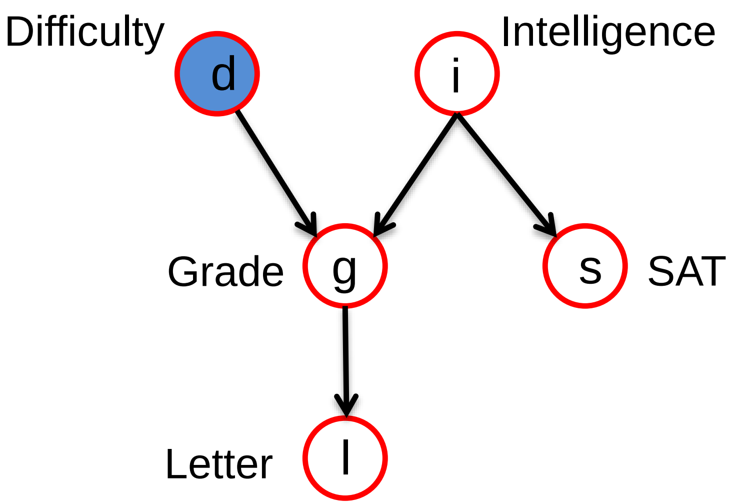
Therefore the letter is dependent on the SAT score.



Explain away

Q: You send the letter to a potential employer. The letter is not good. The employer does not know anything else than the letter:

→ Since l is dependent on i, the employer will infer a low probability p(i) that you are intelligent.



Explain away

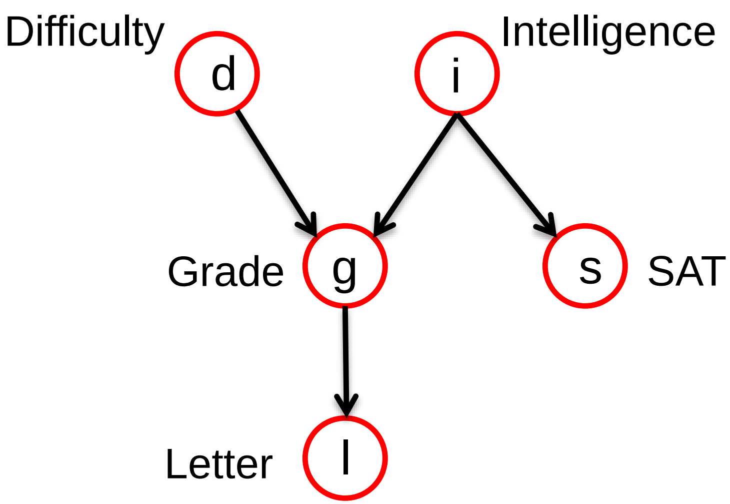
Q: You send the letter to a potential employer. The letter is not good. The employer does not know anything else than the letter:

→ Since l is dependent on i, the employer will infer a low probability p(i) that you are intelligent.

Now the employer gets to know from a friend that the exam had been very difficult in your year. This offers an alternative explanation for a bad letter.

→ Since l is dependent on i and d, and the employer knows that the difficulty was high, he will infer an increased probability p(i|l,d) that you are intelligent → the alternative explanation explained away the naïve/simple hypothesis.

An Example: Now with real p's



$$p(d,i,g,s,l) =$$
$$p(d)p(i)p(g|i,d)p(s|i)p(l|g)$$

d=0	d=1
0.6	0.4

i=0	i=1
0.7	0.3

	g=1	g=2	g=3
d=0,i=0	0.3	0.4	0.3
d=1,i=0	0.05	0.25	0.7
d=0,i=1	0.9	0.08	0.02
d=1,i=1	0.5	0.3	0.2

	s=0	s=1
i=0	0.95	0.05
i=1	0.2	0.8

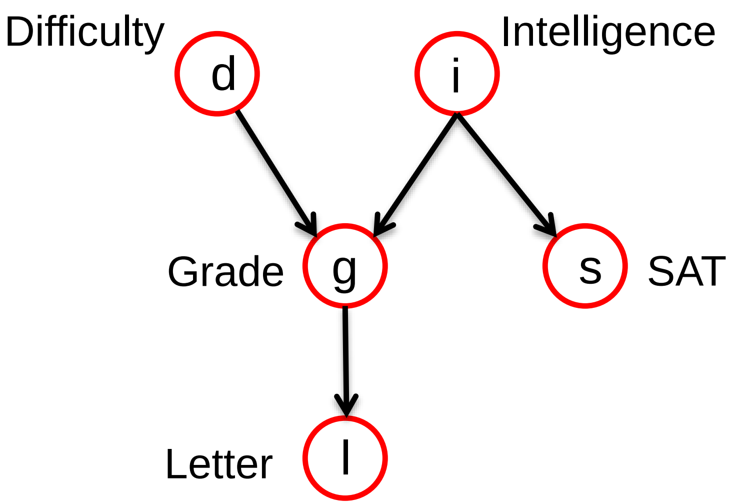
	l=0	l=1
g=1	0.1	0.9
g=2	0.4	0.6
g=3	0.99	0.01

What is the probability of a good letter p(l=1)?

What the probability of a good letter given that you are not intelligent p(l=1|i=0) ?

What the probability of a good letter given that you are not intelligent and the exam is easy p(l=1|i=0,d=0) ?

An Example: Now with real p's



$$p(d,i,g,s,l) = p(d)\,p(i)\,p(g\,|\,i,d)\,p(s\,|\,i)\,p(l\,|\,g)$$

What is the probability of a good letter p(l=1)?

We marginalize the joint probability.

$$p(l=1) = \sum_{d,i,g,s} p(d,i,g,s,l)$$

d=0	d=1
0.6	0.4

i=0	i=1
0.7	0.3

	g=1	g=2	g=3
d=0,i=0	0.3	0.4	0.3
d=1,i=0	0.05	0.25	0.7
d=0,i=1	0.9	0.08	0.02
d=1,i=1	0.5	0.3	0.2

	s=0	s=1
i=0	0.95	0.05
i=1	0.2	0.8

	l=0	l=1
g=1	0.1	0.9
g=2	0.4	0.6
g=3	0.99	0.01

An Example: Now with real p's

$$p(l=1)=\sum_{d,i,g,s} p(d)p(i)p(g\,|\,i,d)p(s\,|\,i)p(l\,|\,g)$$

$$p(l=1)=$$

$$\boxed{0.6\cdot 0.7\cdot 0.3\cdot 0.95\cdot 0.9+}$$
$$0.4\cdot 0.7\cdot 0.05\cdot 0.95\cdot 0.9+$$
$$0.6\cdot 0.3\cdot 0.9\cdot 0.2\cdot 0.9+$$
$$0.4\cdot 0.3\cdot 0.5\cdot 0.2\cdot 0.9+....$$

d=0	d=1
0.6	0.4

i=0	i=1
0.7	0.3

	g=1	g=2	g=3
d=0,i=0	0.3	0.4	0.3
d=1,i=0	0.05	0.25	0.7
d=0,i=1	0.9	0.08	0.02
d=1,i=1	0.5	0.3	0.2

	s=0	s=1
i=0	0.95	0.05
i=1	0.2	0.8

	l=0	l=1
g=1	0.1	0.9
g=2	0.4	0.6
g=3	0.99	0.01

An Example: Now with real p's

$$p(l=1)=\sum_{d,i,g,s} p(d)p(i)p(g\,|\,i,d)p(s\,|\,i)p(l\,|\,g)$$

$$p(l=1)=$$

$$0.6\cdot 0.7\cdot 0.3\cdot 0.95\cdot 0.9+$$
$$\boxed{0.4\cdot 0.7\cdot 0.05\cdot 0.95\cdot 0.9+}$$
$$0.6\cdot 0.3\cdot 0.9\cdot 0.2\cdot 0.9+$$
$$0.4\cdot 0.3\cdot 0.5\cdot 0.2\cdot 0.9+....$$

d=0	d=1
0.6	0.4

i=0	i=1
0.7	0.3

	g=1	g=2	g=3
d=0,i=0	0.3	0.4	0.3
d=1,i=0	0.05	0.25	0.7
d=0,i=1	0.9	0.08	0.02
d=1,i=1	0.5	0.3	0.2

	s=0	s=1
i=0	0.95	0.05
i=1	0.2	0.8

	l=0	l=1
g=1	0.1	0.9
g=2	0.4	0.6
g=3	0.99	0.01

An Example: Now with real p's

$$p(l=1)=\sum_{d,i,g,s} p(d)p(i)p(g\mid i,d)p(s\mid i)p(l\mid g)$$

$$p(l=1)=$$

$$0.6\cdot 0.7\cdot 0.3\cdot 0.95\cdot 0.9+$$

$$0.4\cdot 0.7\cdot 0.05\cdot 0.95\cdot 0.9+$$

$$\boxed{0.6\cdot 0.3\cdot 0.9\cdot 0.2\cdot 0.9+}$$

$$0.4\cdot 0.3\cdot 0.5\cdot 0.2\cdot 0.9+....$$

d=0	d=1
0.6	0.4

i=0	i=1
0.7	0.3

	g=1	g=2	g=3
d=0,i=0	0.3	0.4	0.3
d=1,i=0	0.05	0.25	0.7
d=0,i=1	0.9	0.08	0.02
d=1,i=1	0.5	0.3	0.2

	s=0	s=1
i=0	0.95	0.05
i=1	0.2	0.8

	l=0	l=1
g=1	0.1	0.9
g=2	0.4	0.6
g=3	0.99	0.01

An Example: Now with real p's

$$p(l=1)=\sum_{d,i,g,s} p(d)p(i)p(g\mid i,d)p(s\mid i)p(l\mid g)$$

$$p(l=1)=$$

$$0.6\cdot 0.7\cdot 0.3\cdot 0.95\cdot 0.9+$$
$$0.4\cdot 0.7\cdot 0.05\cdot 0.95\cdot 0.9+$$
$$0.6\cdot 0.3\cdot 0.9\cdot 0.2\cdot 0.9+$$
$$\boxed{0.4\cdot 0.3\cdot 0.5\cdot 0.2\cdot 0.9+....}$$

Sum over all 2*2*3*2 =24 combinations

$$p(l=1)=0.502$$

The probability that you get a good letter without knowing anything about d,i,g,s is 0.502.

d=0	d=1
0.6	0.4

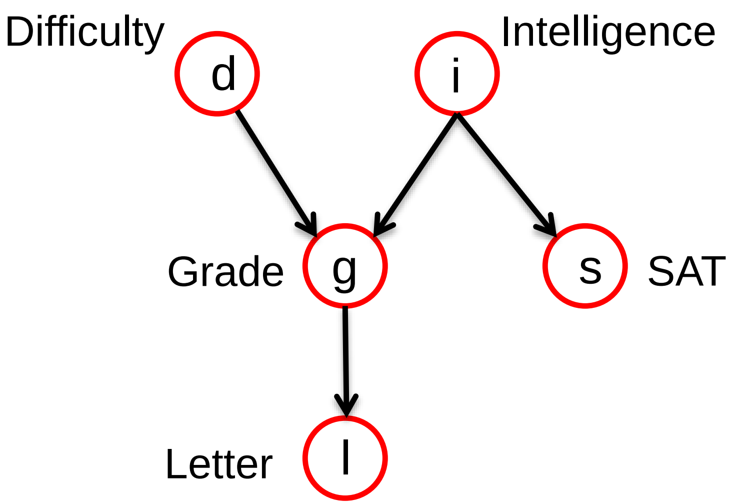
i=0	i=1
0.7	0.3

	g=1	g=2	g=3
d=0,i=0	0.3	0.4	0.3
d=1,i=0	0.05	0.25	0.7
d=0,i=1	0.9	0.08	0.02
d=1,i=1	0.5	0.3	0.2

	s=0	s=1
i=0	0.95	0.05
i=1	0.2	0.8

	l=0	l=1
g=1	0.1	0.9
g=2	0.4	0.6
g=3	0.99	0.01

An Example: Now with real p's



$$p(d,i,g,s,l) =$$
$$p(d)\,p(i)\,p(g\,|\,i,d)\,p(s\,|\,i)\,p(l\,|\,g)$$

What is the probability of a good letter given that the student is not intelligent

p(l=1|i=0)

$$p(d,i,g,s,l) = p(d,g,s,l\,|\,i)\,p(i) =$$
$$p(d)\,p(i)\,p(g\,|\,i,d)\,p(s\,|\,i)\,p(l\,|\,g)$$

d=0	d=1
0.6	0.4

i=0	i=1
0.7	0.3

	g=1	g=2	g=3
d=0,i=0	0.3	0.4	0.3
d=1,i=0	0.05	0.25	0.7
d=0,i=1	0.9	0.08	0.02
d=1,i=1	0.5	0.3	0.2

	s=0	s=1
i=0	0.95	0.05
i=1	0.2	0.8

	l=0	l=1
g=1	0.1	0.9
g=2	0.4	0.6
g=3	0.99	0.01

An Example: Now with real p's

What is the probability of a good letter given that the student is not intelligent

p(l=1|i=0)

$$p(d,g,s,l\,|\,i) = p(d)\,p(g\,|\,i,d)\,p(s\,|\,i)\,p(l\,|\,g)$$

Marginalizing the conditional probability

$$p(l=1\,|\,i) = \sum_{d,g,s} p(d)\,p(g\,|\,i,d)\,p(s\,|\,i)\,p(l\,|\,g)$$

Sum over all 2*3*2 =12 combinations

$$p(l=1\,|\,i=0) = 0.389$$

d=0	d=1
0.6	0.4

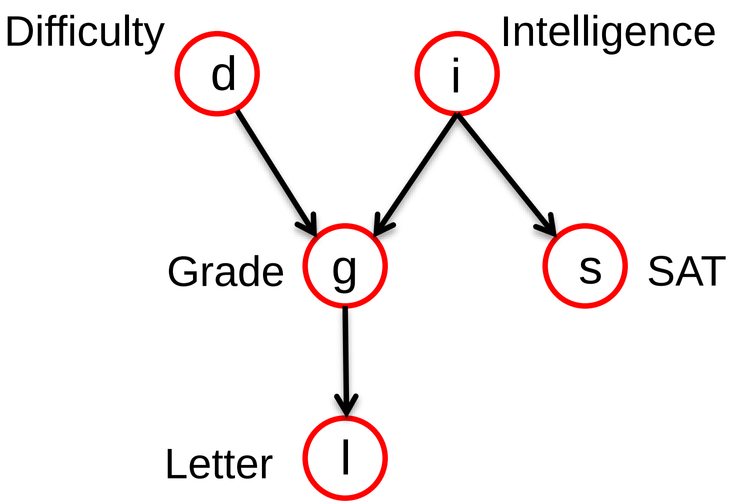
i=0	i=1
0.7	0.3

	g=1	g=2	g=3
d=0,i=0	0.3	0.4	0.3
d=1,i=0	0.05	0.25	0.7
d=0,i=1	0.9	0.08	0.02
d=1,i=1	0.5	0.3	0.2

	s=0	s=1
i=0	0.95	0.05
i=1	0.2	0.8

	l=0	l=1
g=1	0.1	0.9
g=2	0.4	0.6
g=3	0.99	0.01

An Example: Now with real p's



What is the probability of a good letter given that the student is not intelligent and the exam is easy: p(l=1|i=0,d=0)

$$p(d,i,g,s,l) = p(g,s,l \mid d,i)\,p(d,i)$$
$$= p(d)\,p(i)\,p(g \mid i,d)\,p(s \mid i)\,p(l \mid g)$$

Path from d to i is blocked. → d,i are independent and p(d,i)=p(d)p(i)

An Example: Now with real p's

What is the probability of a good letter given that the student is not intelligent and the exam is easy: p(l=1|i=0,d=0)

$$p(g,s,l \mid d,i) = p(g \mid i,d)\, p(s \mid i)\, p(l \mid g)$$

Marginalizing the conditional probability

$$p(l=1 \mid i,d) = \sum_{g,s} p(g \mid i,d)\, p(s \mid i)\, p(l \mid g)$$

Sum over all 2*3 = 6 combinations

$$p(l=1 \mid i=0, d=0) = 0.513$$

d=0	d=1
0.6	0.4

i=0	i=1
0.7	0.3

	g=1	g=2	g=3
d=0,i=0	0.3	0.4	0.3
d=1,i=0	0.05	0.25	0.7
d=0,i=1	0.9	0.08	0.02
d=1,i=1	0.5	0.3	0.2

	s=0	s=1
i=0	0.95	0.05
i=1	0.2	0.8

	l=0	l=1
g=1	0.1	0.9
g=2	0.4	0.6
g=3	0.99	0.01

What is the probability of a good letter p(l=1)?

$$p(l=1)=0.502$$

What is the probability of a good letter given that the student is not intelligent p(l=1|i=0) ?

$$p(l=1\,|\,i=0)=0.389$$

What is the probability of a good letter given that the student is not intelligent and the exam is easy: p(l=1|i=0,d=0).

$$p(l=1\,|\,i=0,d=0)=0.513$$