Bayesian Networks

Or directed graphical models

Wilker Aziz

w.aziz@uva.nl

https://probabll.github.io

Outline and goals

Module 1 introduces *Bayesian networks* (BNs; Chapter 3).

ILOs After this module the student

- can map BNs to distributions and vice-versa;
- recognises reasoning patterns in BNs;
- recognises the flow of probabilistic influence in BNs;
- recognises independence and directed-separation.

Textbook for this course: Koller and Friedman [1].

Overview of Module 1

HC1: BNs – semantics.

LC1: BNs in code.

HC2: BNs – reasoning and influence.

WC1: exercises (semantics, reasoning and influence).

Table of contents

Semantics

Outline for this section

We will introduce BNs via an example from the textbook (the *Student* example from Section 3.1.3.1).

After presenting the complete example, we will introduce BNs in full generality.

The *Student* example (section 3.2.1)

A student scores a passing

Grade (excellent g^1, good g^2, or satisfactory g^3)

in a course of a certain

Difficulty level (hard d^1, or easy d^0)

and obtains a

Letter of recommendation (strong l^1, or weak l^0))

from the course coordinator.
This kind of information, along with the student's

SAT scores (high s^1, or low s^0),

is what some company uses to support hiring decisions, with these aspects taken as informative of the candidate's

Intelligence (high i^1, or low i^0).

The *Student* example (section 3.2.1)

- Grade $\quad \mathrm{Val}(G) = \{g^1, g^2, g^3\}$
- Difficulty $\quad \mathrm{Val}(D) = \{d^0, d^1\}$
- Intelligence $\quad \mathrm{Val}(I) = \{i^0, i^1\}$
- SAT $\quad \mathrm{Val}(S) = \{s^0, s^1\}$
- Ref. Letter $\quad \mathrm{Val}(R) = \{l^0, l^1\}$

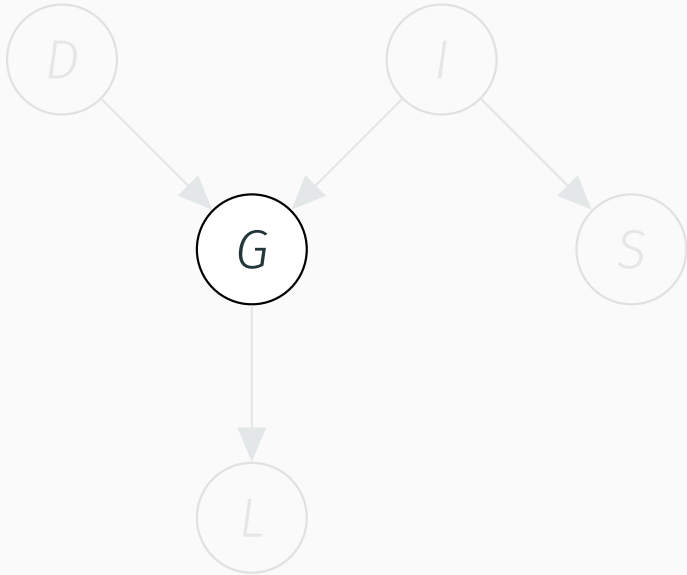


Figure 1: The BN graph for the *Student* example (Fig 3.3 of textbook).

The BN structure captures our assumptions about the world in terms of: **what variables matter and how they depend on one another**.

The possible outcomes of an a random variable X is denoted $\mathrm{Val}(X)$; see section 2.1.3.2 for more.

The *Student* example (section 3.2.1)

- Grade $\quad \mathbf{Val}(G) = \{g^1, g^2, g^3\}$
- Difficulty $\quad \mathbf{Val}(D) = \{d^0, d^1\}$
- Intelligence $\quad \mathbf{Val}(I) = \{i^0, i^1\}$
- SAT $\quad \mathbf{Val}(S) = \{s^0, s^1\}$
- Ref. Letter $\quad \mathbf{Val}(R) = \{l^0, l^1\}$

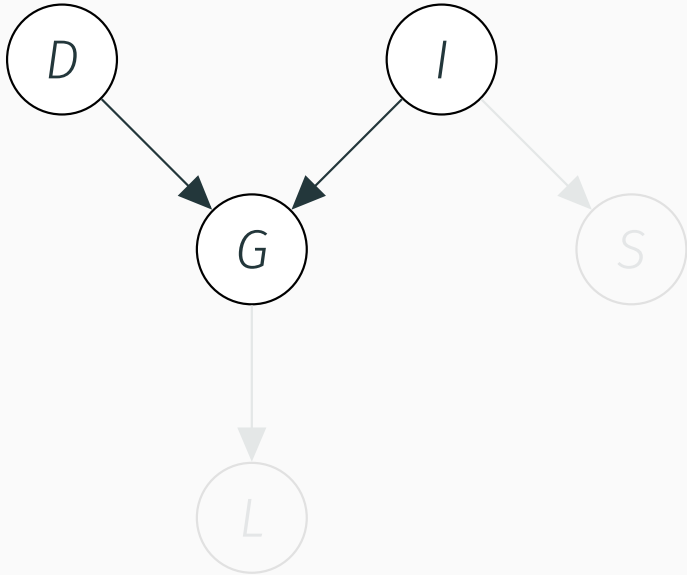


Figure 1: The BN graph for the *Student* example (Fig 3.3 of textbook).

The BN structure captures our assumptions about the world in terms of: **what variables matter and how they depend on one another**.

The possible outcomes of an a random variable X is denoted $\mathbf{Val}(X)$; see section 2.1.3.2 for more.

The *Student* example (section 3.2.1)

- **G**rade $\quad \mathbf{Val}(G) = \{g^1, g^2, g^3\}$
- **D**ifficulty $\quad \mathbf{Val}(D) = \{d^0, d^1\}$
- **I**ntelligence $\quad \mathbf{Val}(I) = \{i^0, i^1\}$
- **S**AT $\quad \mathbf{Val}(S) = \{s^0, s^1\}$
- Ref. **L**etter $\quad \mathbf{Val}(R) = \{l^0, l^1\}$

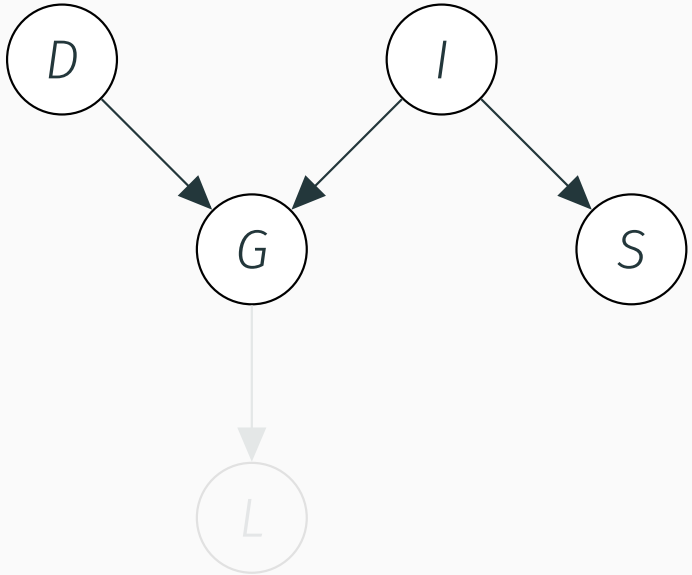


Figure 1: The BN graph for the *Student* example (Fig 3.3 of textbook).

The BN structure captures our assumptions about the world in terms of: **what variables matter and how they depend on one another**.

The possible outcomes of an a random variable X is denoted $\mathbf{Val}(X)$; see section 2.1.3.2 for more.

The *Student* example (section 3.2.1)

- **G**rade $\quad \mathbf{Val}(G) = \{g^1, g^2, g^3\}$
- **D**ifficulty $\quad \mathbf{Val}(D) = \{d^0, d^1\}$
- **I**ntelligence $\quad \mathbf{Val}(I) = \{i^0, i^1\}$
- **S**AT $\quad \mathbf{Val}(S) = \{s^0, s^1\}$
- Ref. **L**etter $\quad \mathbf{Val}(R) = \{l^0, l^1\}$

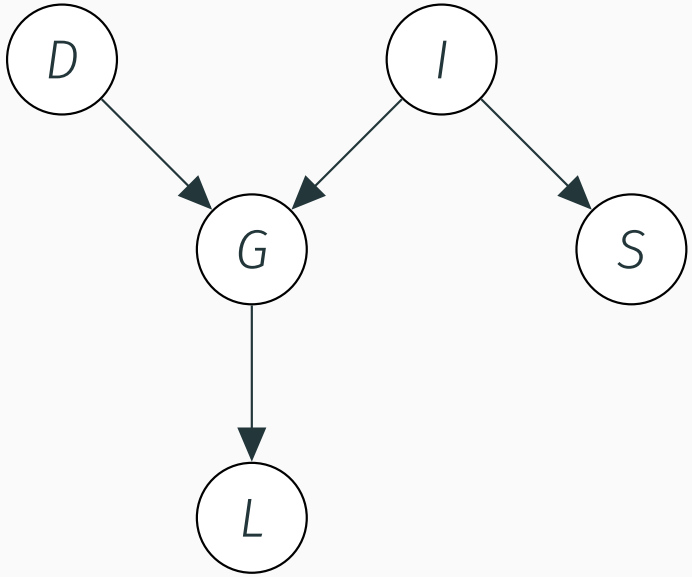


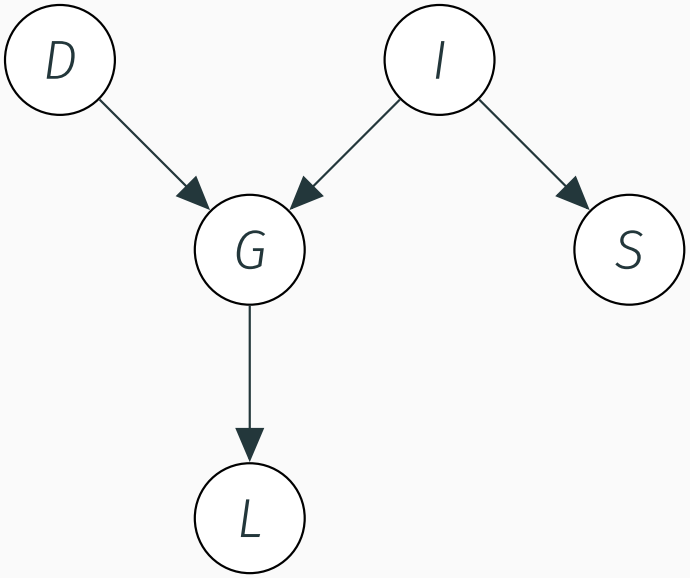
Figure 1: The BN graph for the *Student* example (Fig 3.3 of textbook).

The BN structure captures our assumptions about the world in terms of: **what variables matter and how they depend on one another**.

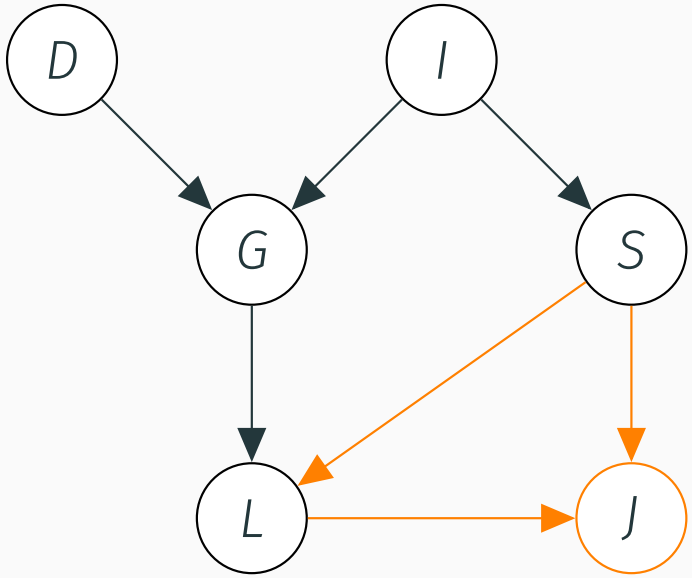
The possible outcomes of an a random variable X is denoted $\mathbf{Val}(X)$; see section 2.1.3.2 for more.

A model is an assumption about how the world works

We can always motivate different assumptions:



Student example (Fig 3.3 of textbook)

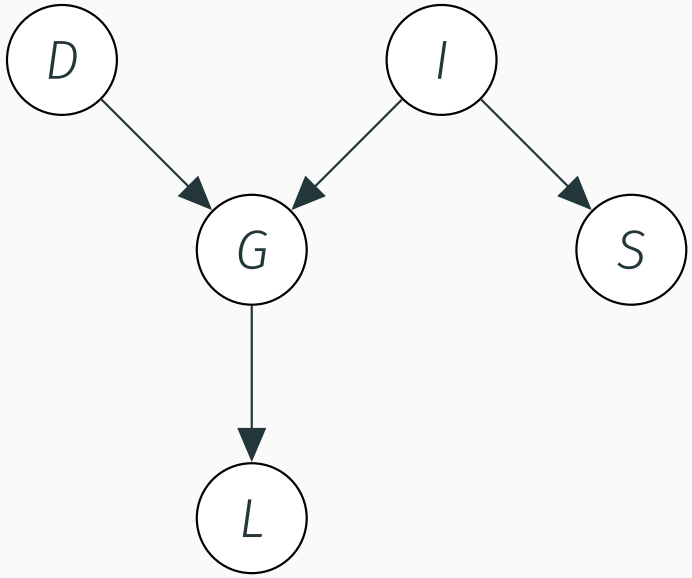


Variant: the Letter is affected by *S*, and we introduce a new variable (e.g., **J**ob).

We can now prescribe probability models that represent our uncertainty about these rvs, while reflecting interactions in the graph.

A model is an assumption about how the world works

We can always motivate different assumptions:



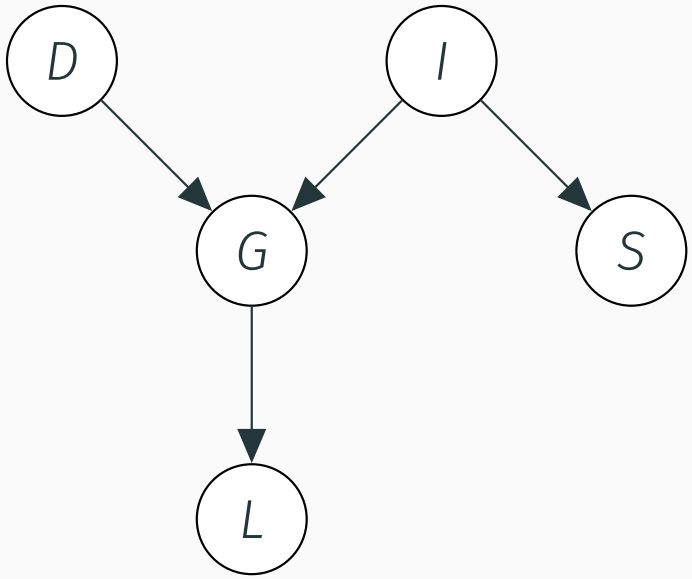
Student example (Fig 3.3 of textbook)

Variant: the Letter is affected by *S*, and we introduce a new variable (e.g., Job).

We can now prescribe probability models that represent our uncertainty about these rvs, while reflecting interactions in the graph.

A model is an assumption about how the world works

We can always motivate different assumptions:



Student example (Fig 3.3 of textbook)

Variant: the Letter is affected by *S*, and we introduce a new variable (e.g., Job).

We can now prescribe probability models that represent our uncertainty about these rvs, while reflecting interactions in the graph.

CPDs for the *Student* example (Fig 3.4 of textbook)

A conditional probability distribution (or CPD) represents our uncertainty about the outcome of a variable in a given context.

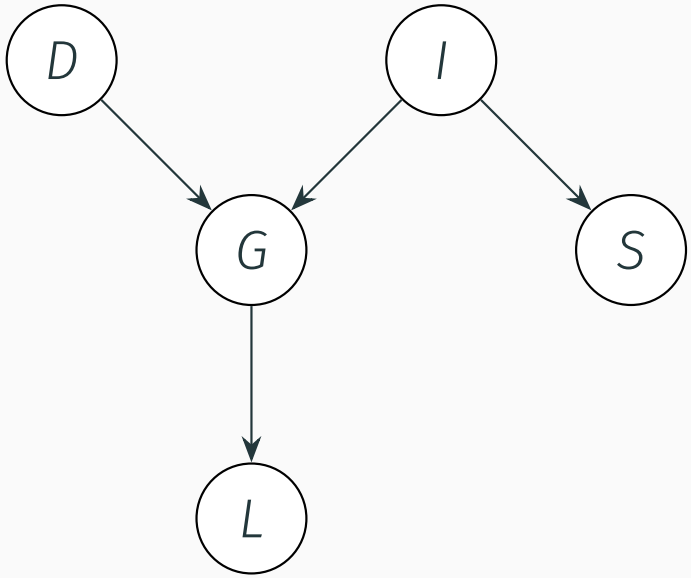


Figure 2: Tabular CPDs for the Student example (Fig 3.4 of textbook)

CPDs for the *Student* example (Fig 3.4 of textbook)

A conditional probability distribution (or CPD) represents our uncertainty about the outcome of a variable in a given context.

d^0	d^1
0.6	0.4

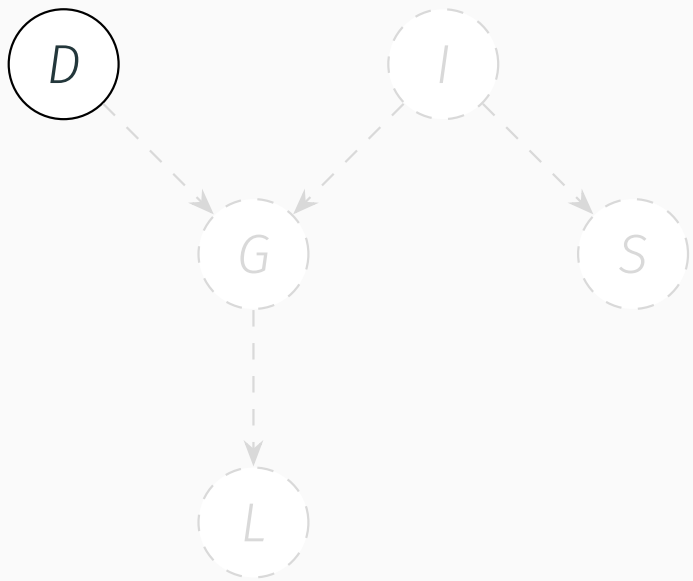


Figure 2: Tabular CPDs for the Student example (Fig 3.4 of textbook)

CPDs for the *Student* example (Fig 3.4 of textbook)

A conditional probability distribution (or CPD) represents our uncertainty about the outcome of a variable in a given context.

d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

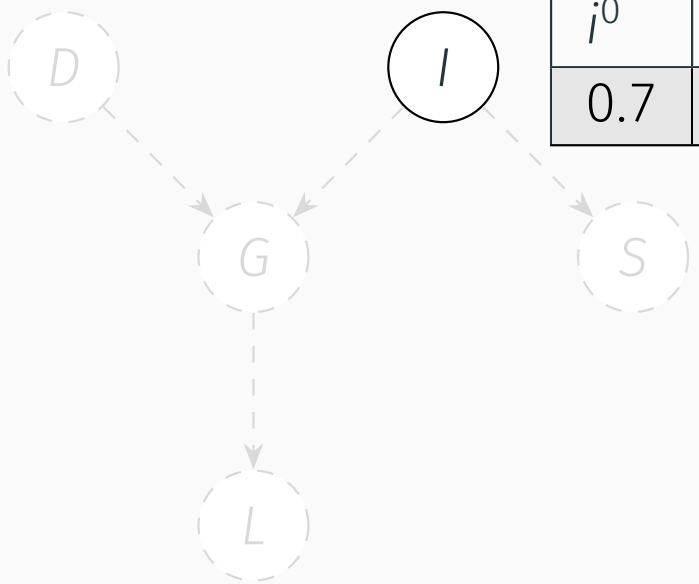
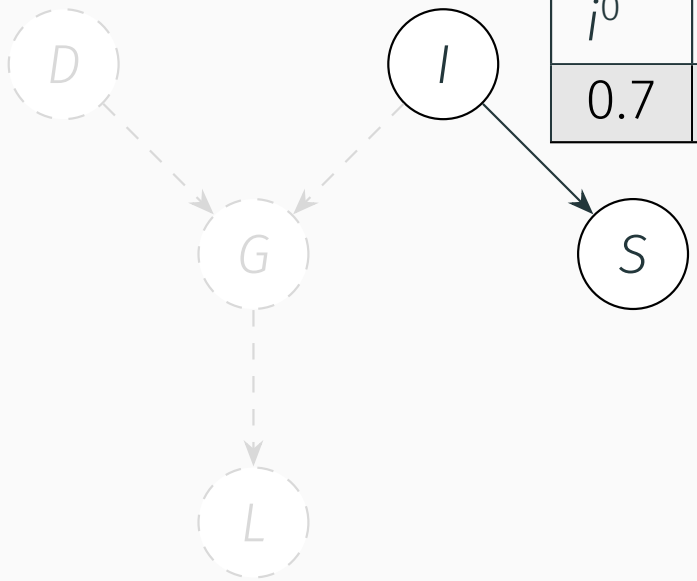


Figure 2: Tabular CPDs for the Student example (Fig 3.4 of textbook)

CPDs for the *Student* example (Fig 3.4 of textbook)

A conditional probability distribution (or CPD) represents our uncertainty about the outcome of a variable in a given context.

d^0	d^1
0.6	0.4



i^0	i^1
0.7	0.3

I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

Figure 2: Tabular CPDs for the Student example (Fig 3.4 of textbook)

CPDs for the *Student* example (Fig 3.4 of textbook)

A conditional probability distribution (or CPD) represents our uncertainty about the outcome of a variable in a given context.

d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

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

I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

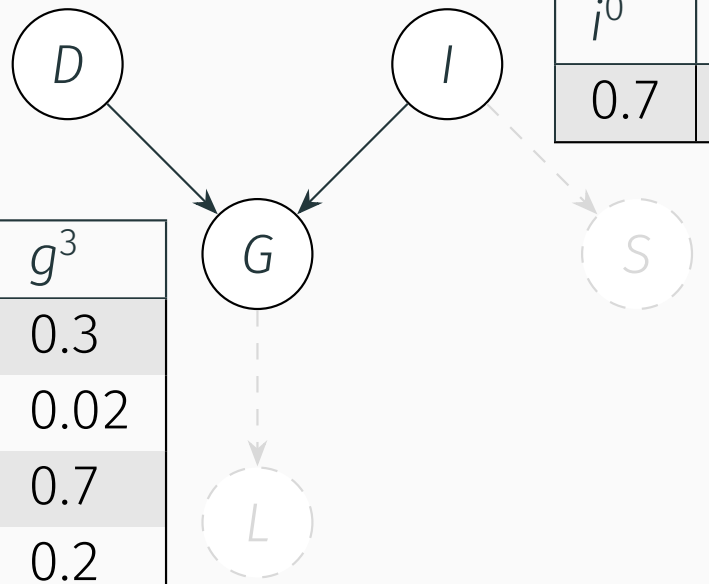
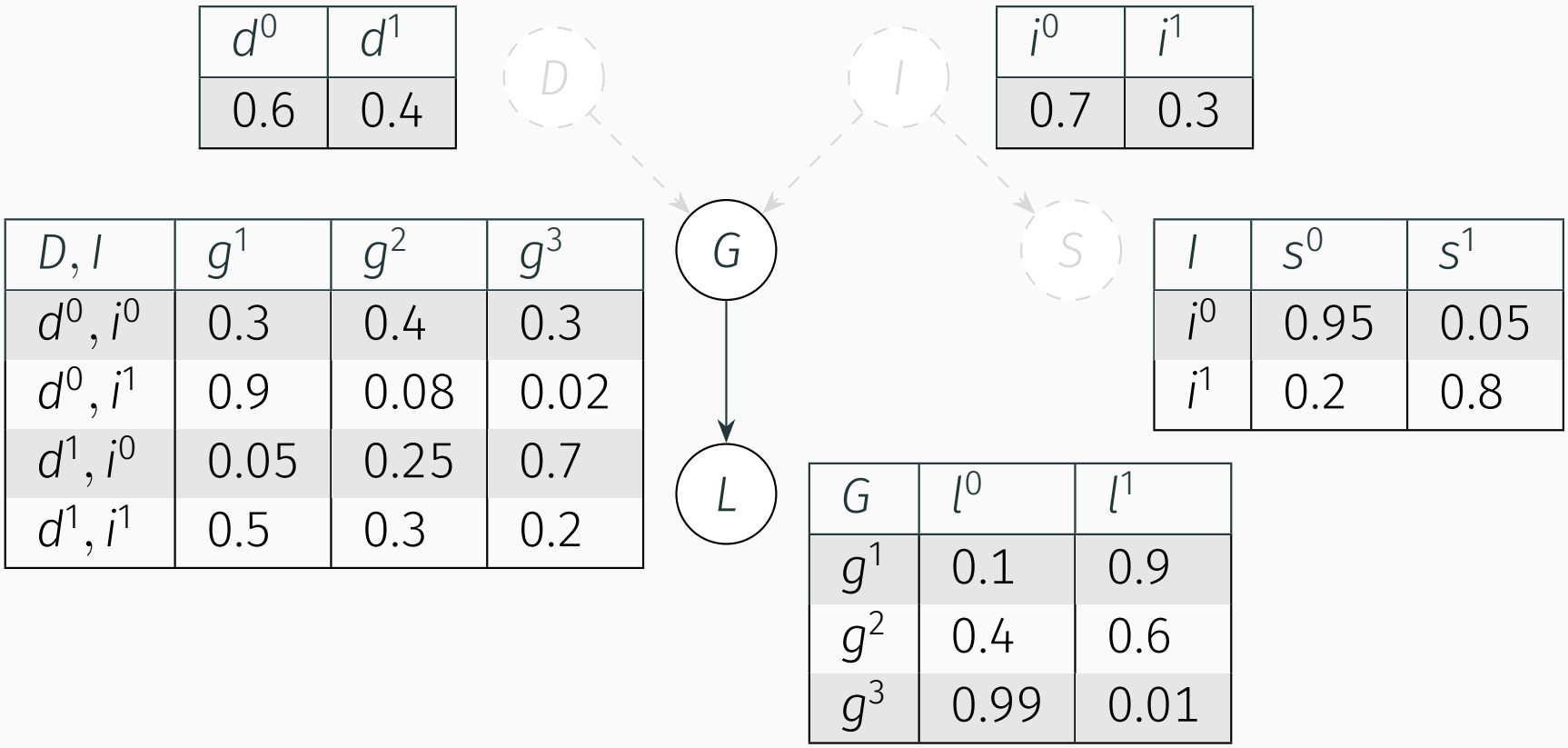


Figure 2: Tabular CPDs for the Student example (Fig 3.4 of textbook)

CPDs for the *Student* example (Fig 3.4 of textbook)

A conditional probability distribution (or CPD) represents our uncertainty about the outcome of a variable in a given context.



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

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

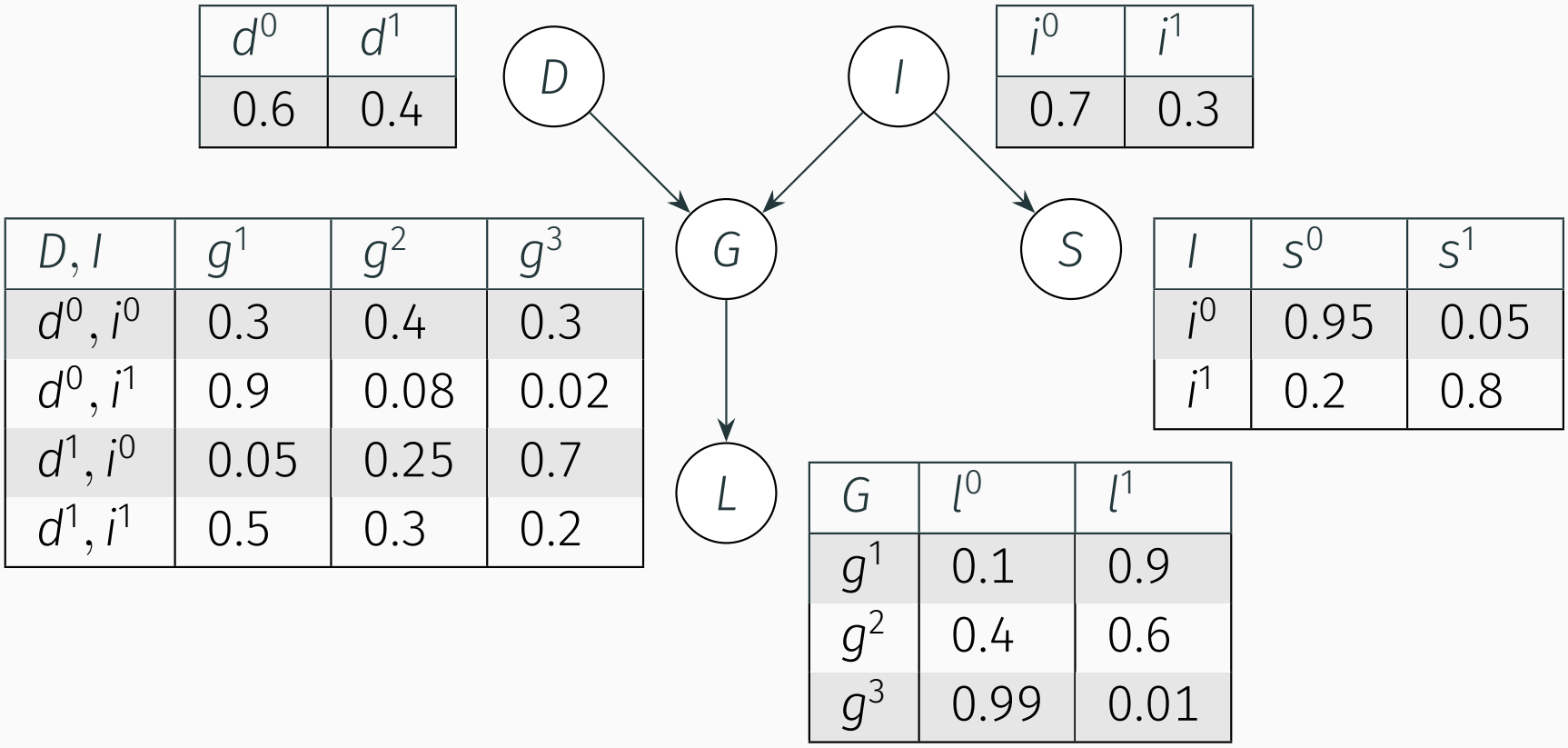
I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

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

Figure 2: Tabular CPDs for the Student example (Fig 3.4 of textbook)

CPDs for the *Student* example (Fig 3.4 of textbook)

A conditional probability distribution (or CPD) represents our uncertainty about the outcome of a variable in a given context.



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

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

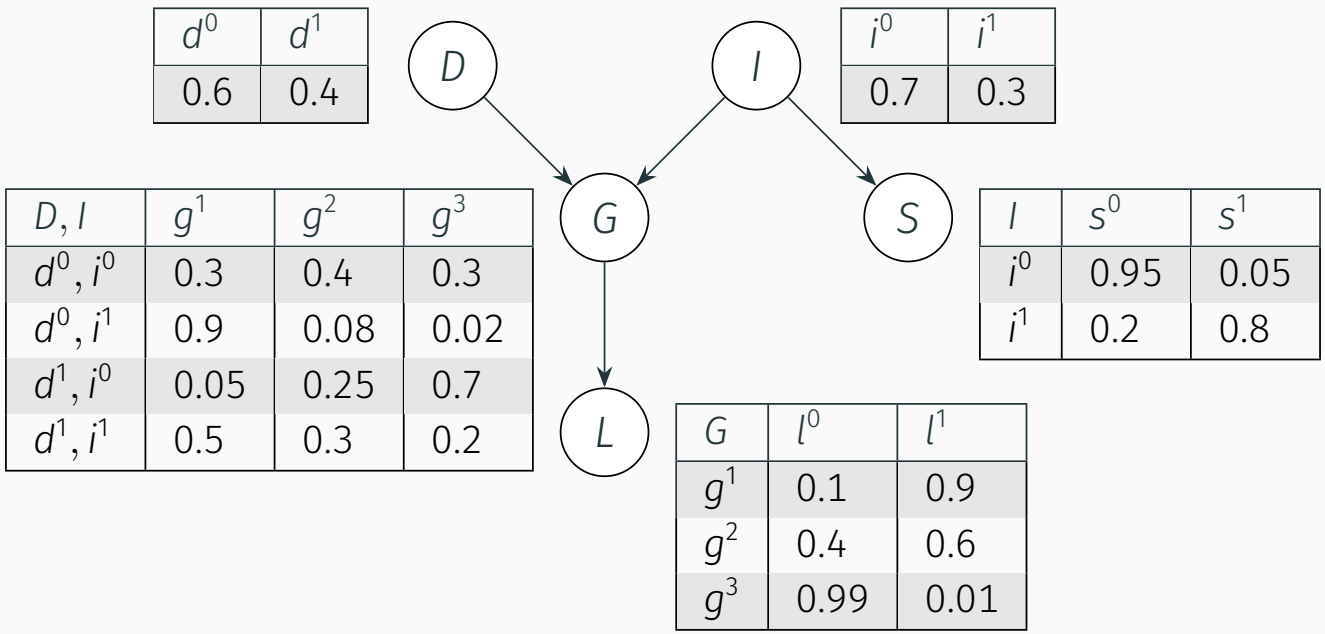
I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

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

Figure 2: Tabular CPDs for the Student example (Fig 3.4 of textbook)

These are *local prob. models*, prescribed separately for each node.

Can you evaluate these joint probabilities?



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

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

I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

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

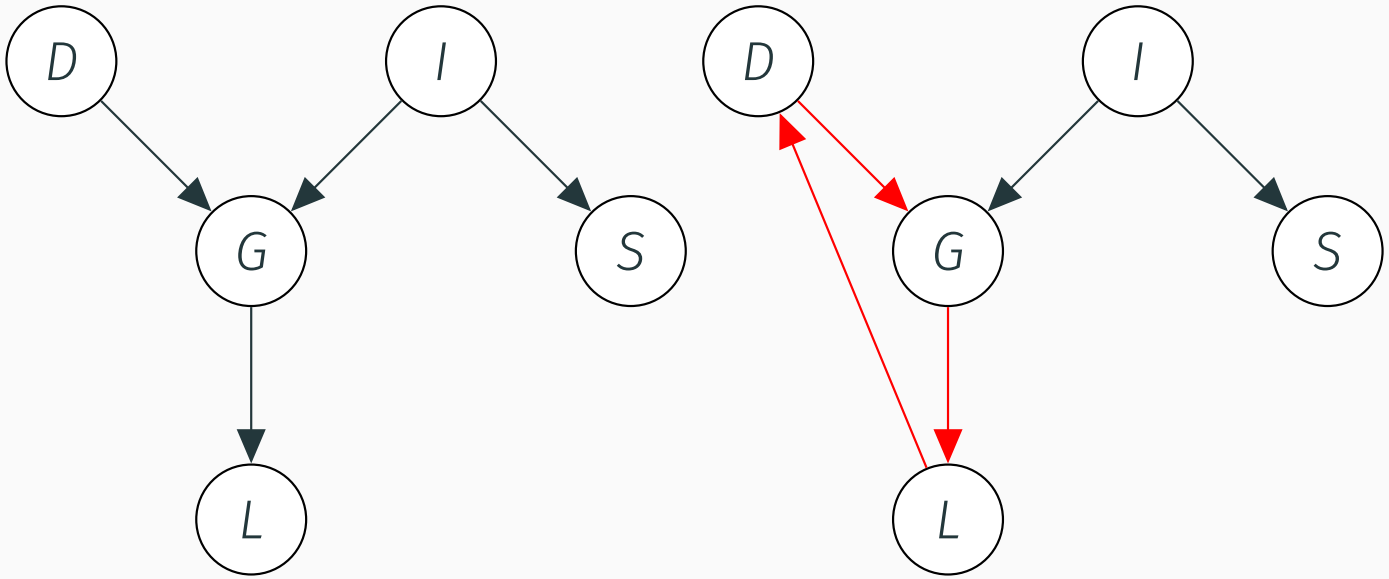
CPDs for the *Student* example (Fig 3.4 of textbook)

1. $P(D = d^1, I = i^1, G = g^2, S = s^1, L = l^1) =$
2. $P(D = d^0, I = i^1, G = g^2, S = s^1, L = l^1) =$
3. $P(D = d^0, I = i^1, G = g^3, S = s^1, L = l^1) =$

BN Structure

A BN is built on a directed acyclic graph (DAG):

- nodes represent random variables (rvs)
- edges represent direct dependence
- there are no **directed cycles**

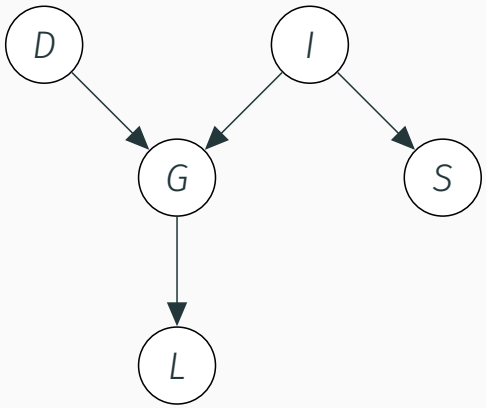


Left: DAG.

Right: **Not a DAG**.

Semantics of Bayesian networks

A DAG $\mathcal{G}$ represents a collection of *conditional independence* statements. Amongst these, some are readily recognisable from the graph structure: **each rv** is conditionally independent of its **non-descendants** given its **parents**.



Student example

rv	descendants	non-descendants	parents
D			
I			
S			
G	$\{L\}$	$\{D, I, S\}$	$\{D, I\}$
L			

Can you complete the table?

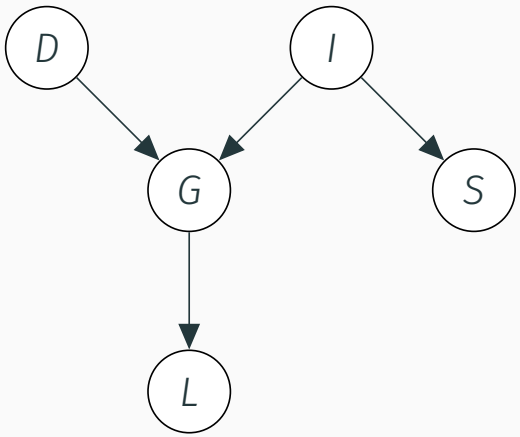
Definition 2.18 of textbook: *descendants* of X are nodes reachable by paths that begin at X; *non-descendants* of X are all nodes except X itself and its descendants; *parents* of X are non-descendants directly connected to X.

Local Independencies $\mathcal{I}_l(\mathcal{G})$ (definition 3.1)

The *local independencies* coded in the BN structure, denoted $\mathcal{I}_l(\mathcal{G})$, is the set of statements of the kind:

$$\text{For each variable } X_i: \quad X_i \perp \text{NonDesc}_{\mathcal{G}}(X_i) \mid \text{Pa}_{\mathcal{G}}(X_i)$$

In the Student example:

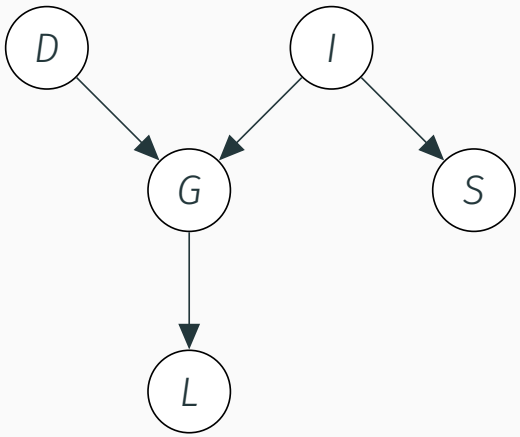


Local Independencies $\mathcal{I}_l(\mathcal{G})$ (definition 3.1)

The *local independencies* coded in the BN structure, denoted $\mathcal{I}_l(\mathcal{G})$, is the set of statements of the kind:

$$\text{For each variable } X_i: \quad X_i \perp \text{NonDesc}_{\mathcal{G}}(X_i) \mid \text{Pa}_{\mathcal{G}}(X_i)$$

In the Student example:



- $D \perp I, S$
- $I \perp D$
- $S \perp D, G, L \mid I$
- $G \perp S \mid D, I$
- $L \perp D, I, S \mid G$

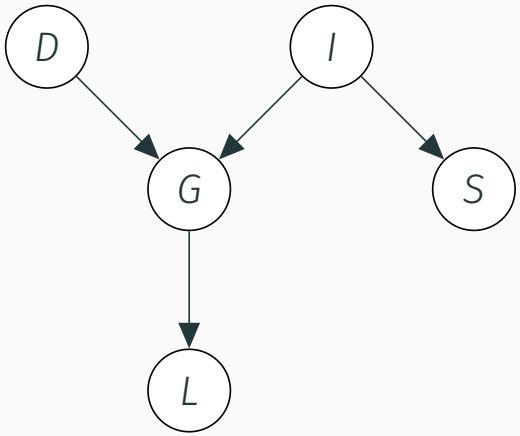
Watch out: it makes no sense to write $G \perp \overline{D, I, S}^{\text{NonDesc}} \mid \overline{D, I}^{\text{Pa}}$ as G cannot be independent of its parents D, I, the *non-descendants* that we separate ($\perp$) from G, given the parents, are the non-descendants who are not its parents.

Local Independencies $\mathcal{I}_l(\mathcal{G})$ (definition 3.1)

The *local independencies* coded in the BN structure, denoted $\mathcal{I}_l(\mathcal{G})$, is the set of statements of the kind:

$$\text{For each variable } X_i: \quad X_i \perp \text{NonDesc}_{\mathcal{G}}(X_i) \mid \text{Pa}_{\mathcal{G}}(X_i)$$

In the Student example:



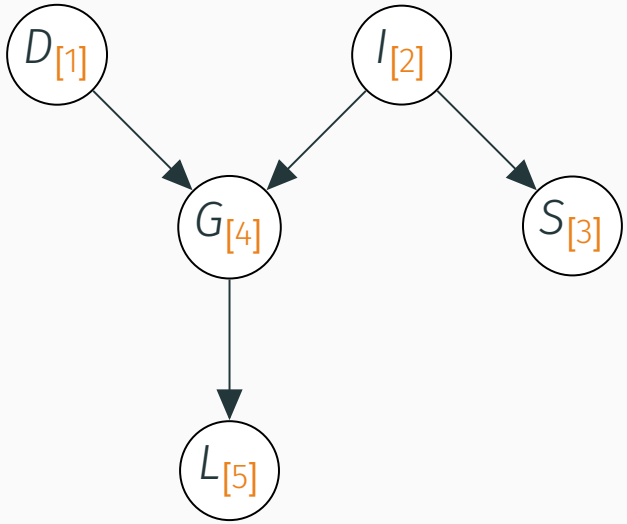
- $D \perp I, S$
- $I \perp D$
- $S \perp D, G, L \mid I$
- $G \perp S \mid D, I$
- $L \perp D, I, S \mid G$

Now let's try to use this in order to state a joint distribution.

Watch out: it makes no sense to write $G \perp \overline{D, I, S}^{\text{NonDesc}} \mid \overline{D, I}^{\text{Pa}}$ as G cannot be independent of its parents D, I, the *non-descendants* that we separate ($\perp$) from G, given the parents, are the non-descendants who are not its parents.

Topological Ordering (definition 2.19)

Index the nodes such that for any X_i, its parents are in the set $X_{<i} = \{X_j : j < i\}$. By definition then, $X_{<i} \subseteq \text{NonDesc}_{\mathcal{G}}(X_i)$.

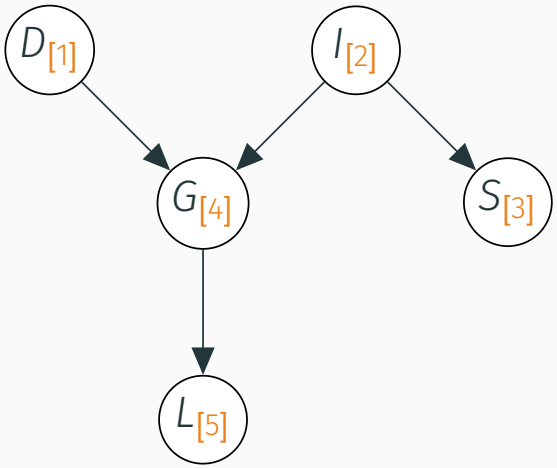


Nodes of the Student example indexed in a topological order

There can be multiple such orders, but all deliver the result $X_{<i} \subseteq \text{NonDesc}_{\mathcal{G}}(X_i)$. Topological sorting is always possible for DAGs.

Joint probability distribution for the *Student example*

No matter how many rvs and in what order we enumerate them the **chain rule of probability** always holds: $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_1 | X_{<i})$.

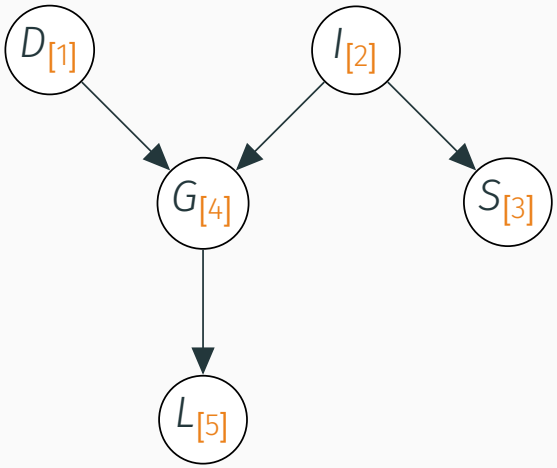


Let's apply **chain rule in a topological order**:

$$P(D, I, G, S, L) =$$
$$P(D)P(I|D)P(S|D, I)P(G|D, I, S)P(L|D, I, S, G)$$

Joint probability distribution for the *Student example*

No matter how many rvs and in what order we enumerate them the **chain rule of probability** always holds: $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_1 | X_{<i})$.



Let's apply **chain rule in a topological order**:

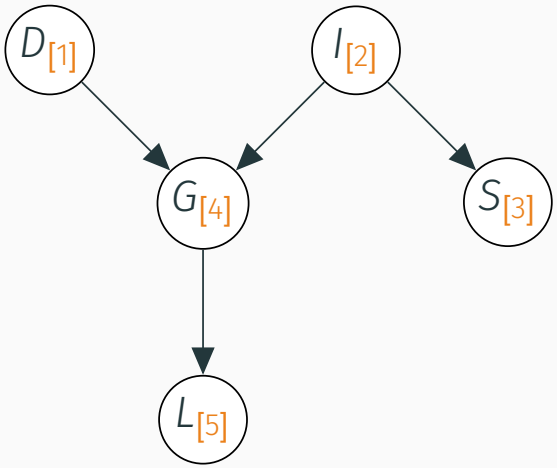
$$P(D, I, G, S, L) = P(D)P(I|D)P(S|D, I)P(G|D, I, S)P(L|D, I, S, G)$$

But the BN codes local indep. $X_i \perp \text{NonDesc}_{\mathcal{G}}(X_i) \mid \text{Pa}_{\mathcal{G}}(X_i)$

- $I \perp D$ hence $P(I|D) = P(I)$

Joint probability distribution for the *Student example*

No matter how many rvs and in what order we enumerate them the **chain rule of probability** always holds: $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_1 | X_{<i})$.



Let's apply **chain rule in a topological order**:

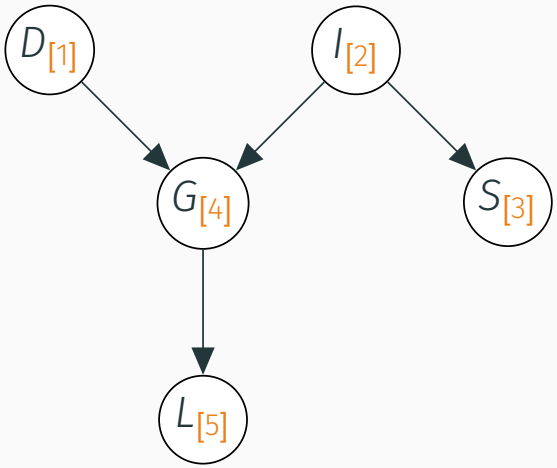
$$P(D, I, G, S, L) = P(D)P(I|D)P(S|D, I)P(G|D, I, S)P(L|D, I, S, G)$$

But the BN codes local indep. $X_i \perp \text{NonDesc}_{\mathcal{G}}(X_i) \mid \text{Pa}_{\mathcal{G}}(X_i)$

- $I \perp D$ hence $P(I|D) = P(I)$
- $S \perp D \mid I$ hence $P(S|D, I) = P(S|I)$

Joint probability distribution for the *Student example*

No matter how many rvs and in what order we enumerate them the **chain rule of probability** always holds: $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_1 | X_{<i})$.



Let's apply **chain rule in a topological order**:

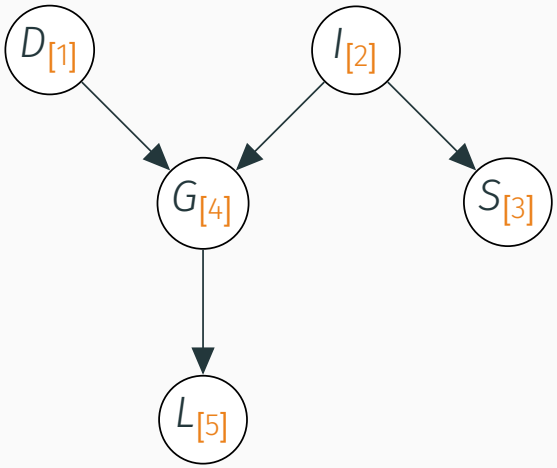
$$P(D, I, G, S, L) = P(D)P(I|D)P(S|D, I)P(G|D, I, S)P(L|D, I, S, G)$$

But the BN codes local indep. $X_i \perp \text{NonDesc}_{\mathcal{G}}(X_i) \mid \text{Pa}_{\mathcal{G}}(X_i)$

- $I \perp D$ hence $P(I|D) = P(I)$
- $S \perp D \mid I$ hence $P(S|D, I) = P(S|I)$
- $G \perp S \mid D, I$ hence $P(G|D, I, S) = P(G|D, I)$

Joint probability distribution for the *Student example*

No matter how many rvs and in what order we enumerate them the **chain rule of probability** always holds: $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_1 | X_{<i})$.



Let's apply **chain rule in a topological order**:

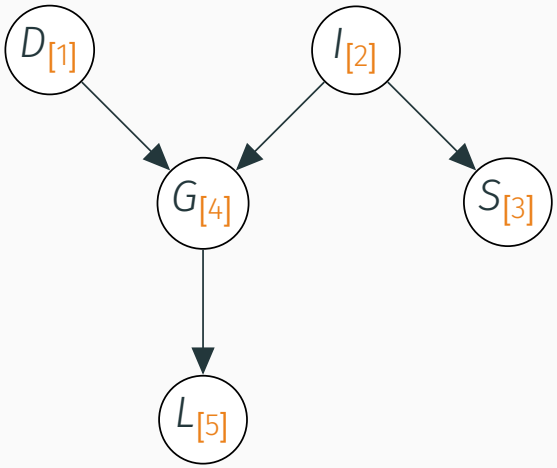
$$P(D, I, G, S, L) = P(D)P(I|D)P(S|D, I)P(G|D, I, S)P(L|D, I, S, G)$$

But the BN codes local indep. $X_i \perp \text{NonDesc}_{\mathcal{G}}(X_i) \mid \text{Pa}_{\mathcal{G}}(X_i)$

- $I \perp D$ hence $P(I|D) = P(I)$
- $S \perp D \mid I$ hence $P(S|D, I) = P(S|I)$
- $G \perp S \mid D, I$ hence $P(G|D, I, S) = P(G|D, I)$
- $L \perp D, I, S \mid G$ hence $P(L|D, I, S, G) = P(L|G)$

Joint probability distribution for the *Student example*

No matter how many rvs and in what order we enumerate them the **chain rule of probability** always holds: $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_1 | X_{<i})$.



Let's apply **chain rule in a topological order**:

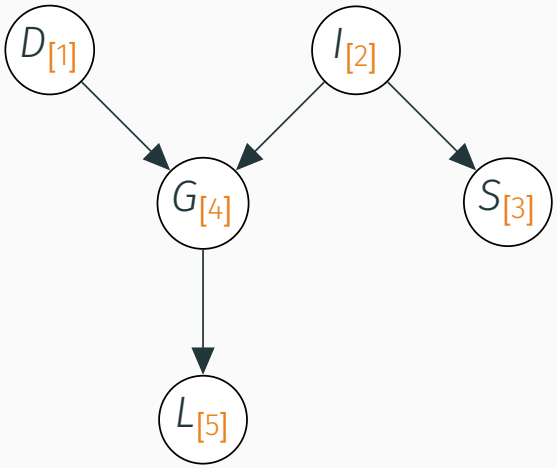
$$P(D, I, G, S, L) = P(D)P(I|D)P(S|D, I)P(G|D, I, S)P(L|D, I, S, G)$$

But the BN codes local indep. $X_i \perp \text{NonDesc}_{\mathcal{G}}(X_i) \mid \text{Pa}_{\mathcal{G}}(X_i)$

- $I \perp D$ hence $P(I|D) = P(I)$
- $S \perp D \mid I$ hence $P(S|D, I) = P(S|I)$
- $G \perp S \mid D, I$ hence $P(G|D, I, S) = P(G|D, I)$
- $L \perp D, I, S \mid G$ hence $P(L|D, I, S, G) = P(L|G)$

Joint probability distribution for the *Student example*

No matter how many rvs and in what order we enumerate them the **chain rule of probability** always holds: $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_1 | X_{<i})$.



Let's apply **chain rule in a topological order**:

$$P(D, I, G, S, L) = P(D)P(I|D)P(S|D, I)P(G|D, I, S)P(L|D, I, S, G)$$

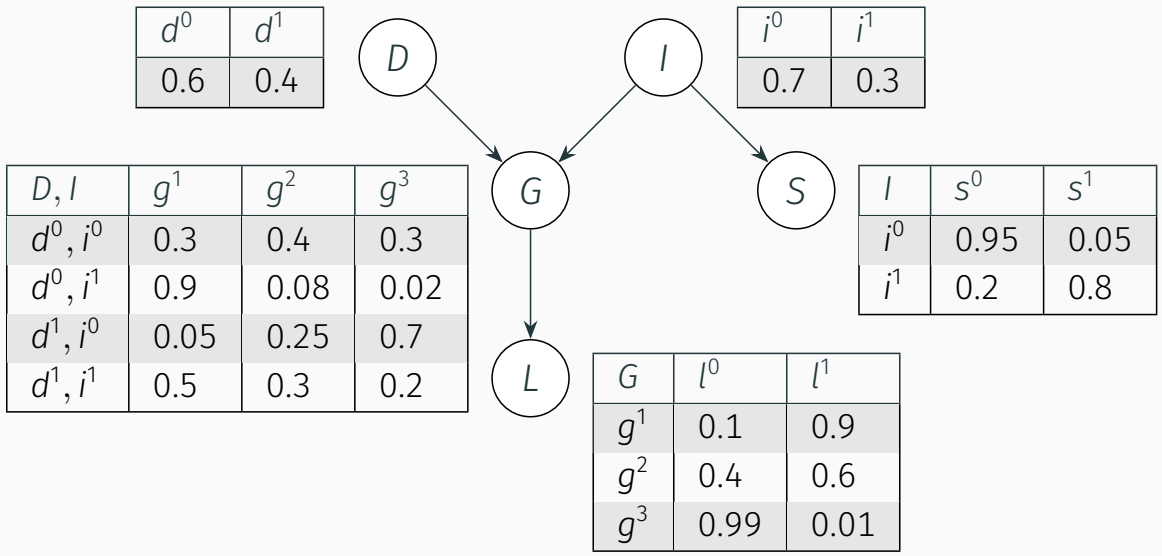
But the BN codes local indep. $X_i \perp \text{NonDesc}_{\mathcal{G}}(X_i) \mid \text{Pa}_{\mathcal{G}}(X_i)$

- $I \perp D$ hence $P(I|D) = P(I)$
- $S \perp D \mid I$ hence $P(S|D, I) = P(S|I)$
- $G \perp S \mid D, I$ hence $P(G|D, I, S) = P(G|D, I)$
- $L \perp D, I, S \mid G$ hence $P(L|D, I, S, G) = P(L|G)$

And hence: $P(D, I, G, S, L) \overset{\mathcal{G}}{=} P(D)P(I)P(S|I)P(G|D, I)P(L|G)$

A Representation of our Uncertainty over D, I, G, S, L

A *joint probability distribution* is a means to represent our uncertainty over the *possible* assignments of the rvs of interest.



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

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

I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

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

Student example (Fig 3.4 of textbook)

With its DAG and CPDs, a BN identifies one such distribution.

The probability of any one of the possible outcomes of (D, I, G, S, L) is given by $P(D, I, G, S, L) = P(D)P(I)P(G|D, I)P(S|I)P(L|G)$.

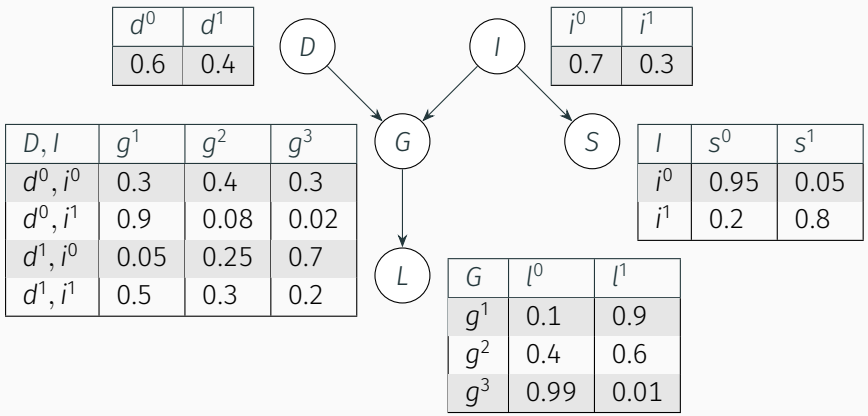
D	I	G	S	L	$P(D, I, G, S, L)$
d^0	i^0	g^1	s^0	l^0	0.01197
d^0	i^0	g^1	s^0	l^1	0.10773
d^0	i^0	g^1	s^1	l^0	0.00063
d^0	i^0	g^1	s^1	l^1	0.00567
d^0	i^0	g^2	s^0	l^0	0.06384
d^0	i^0	g^2	s^0	l^1	0.09576
d^0	i^0	g^2	s^1	l^0	0.00336
d^0	i^0	g^2	s^1	l^1	0.00504
d^0	i^0	g^3	s^0	l^0	0.11850
d^0	i^0	g^3	s^0	l^1	0.00120
d^0	i^0	g^3	s^1	l^0	0.00624
d^0	i^0	g^3	s^1	l^1	0.00006
d^0	i^1	g^1	s^0	l^0	0.00324
d^0	i^1	g^1	s^0	l^1	0.02916
d^0	i^1	g^1	s^1	l^0	0.01296
d^0	i^1	g^1	s^1	l^1	0.11664
d^0	i^1	g^2	s^0	l^0	0.00115
d^0	i^1	g^2	s^0	l^1	0.00173
d^0	i^1	g^2	s^1	l^0	0.00461
d^0	i^1	g^2	s^1	l^1	0.00691
d^0	i^1	g^3	s^0	l^0	0.00071
d^0	i^1	g^3	s^0	l^1	0.00001
d^0	i^1	g^3	s^1	l^0	0.00285
d^0	i^1	g^3	s^1	l^1	0.00003
d^1	i^0	g^1	s^0	l^0	0.00133
d^1	i^0	g^1	s^0	l^1	0.01197
d^1	i^0	g^1	s^1	l^0	0.00007
d^1	i^0	g^1	s^1	l^1	0.00063
d^1	i^0	g^2	s^0	l^0	0.02660
d^1	i^0	g^2	s^0	l^1	0.03990
d^1	i^0	g^2	s^1	l^0	0.00140
d^1	i^0	g^2	s^1	l^1	0.00210
d^1	i^0	g^3	s^0	l^0	0.18434
d^1	i^0	g^3	s^0	l^1	0.00186
d^1	i^0	g^3	s^1	l^0	0.00970
d^1	i^0	g^3	s^1	l^1	0.00010
d^1	i^1	g^1	s^0	l^0	0.00120
d^1	i^1	g^1	s^0	l^1	0.01080
d^1	i^1	g^1	s^1	l^0	0.00480
d^1	i^1	g^1	s^1	l^1	0.04320
d^1	i^1	g^2	s^0	l^0	0.00288
d^1	i^1	g^2	s^0	l^1	0.00432
d^1	i^1	g^2	s^1	l^0	0.01152
d^1	i^1	g^2	s^1	l^1	0.01728
d^1	i^1	g^3	s^0	l^0	0.00475
d^1	i^1	g^3	s^0	l^1	0.00005
d^1	i^1	g^3	s^1	l^0	0.01901
d^1	i^1	g^3	s^1	l^1	0.00019

Table 1: Joint distribution over $\mathbf{Val}(D) \times \mathbf{Val}(I) \times \mathbf{Val}(G) \times \mathbf{Val}(S) \times \mathbf{Val}(L)$.

Compactness

The BN indeed *identifies* the joint distribution, but not by independently representing one probability value per joint outcome.

The probability value of any one of the joint outcomes is expressed via a product of probabilities assigned by *local probability models* to subsets of directly interacting variables.



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

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

I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

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

We infer all 48 joint probabilities from the probabilities in the CPDs $P(D), P(I), P(S|I), P(G|D, I), P(G|L)$. Their tabular representations require $2 + 2 + 2 \times 2 + 4 \times 3 + 3 \times 2 = 26$ parameters, rather than 48.

Bayesian Network (BN) [in full generality]

A BN model is a graph $\mathcal{G}$ and a collection $\mathcal{C}$ of CPDs such that:

- $\mathcal{G}$ is a directed acyclic graph (DAG);
- the nodes in $\mathcal{G}$ represent the random variables $X_1, \ldots, X_n$;
- the edges in $\mathcal{G}$ indicate *direct* conditional dependence of a random variable X_i on its parents $\mathrm{Pa}_{\mathcal{G}}(X_i)$;
- for each node X_i, $\mathcal{C}$ contains a CPD $P(X_i|\mathrm{Pa}_{\mathcal{G}}(X_i))$ specifying the probabilities of the outcomes of X_i given outcomes of $\mathrm{Pa}_{\mathcal{G}}(X_i)$.

The BN represents a joint distribution via the *chain rule of probabilities for Bayesian networks*:

$$P(X_1, \ldots, X_n) \overset{\mathcal{G}}{=} \prod_{i=1}^{n} P(X_i|\mathrm{Pa}_{\mathcal{G}}(X_i)) \tag{1}$$

Graphs and Distributions

We say that a joint distribution P **factorises** according to $\mathcal{G}$ if

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | \mathrm{Pa}_{\mathcal{G}}(X_i))$$

The BN structure and the factorisation imply one another.

To fully identify a distribution, the BN associates the factors $P(X_i|\mathrm{Pa}_{\mathcal{G}}(X_i))$ with local probability models such as tabular CPDs.

The word *factorisation* means **decomposition into a product of factors**.

Graphs and Distributions

We say that a joint distribution P **factorises** according to $\mathcal{G}$ if

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | \mathrm{Pa}_{\mathcal{G}}(X_i))$$

The BN structure and the factorisation imply one another.

To fully identify a distribution, the BN associates the factors $P(X_i|\mathrm{Pa}_{\mathcal{G}}(X_i))$ with local probability models such as tabular CPDs.

Once one such joint distribution is in place, we can manipulate it via probability calculus to infer the result of *any* probability query about marginals and conditionals of P.

The word *factorisation* means **decomposition into a product of factors**.

Summary

- A BN uses a DAG $\mathcal{G}$ to encode statements of (conditional) independence, and these must hold in any distribution P that factorises over $\mathcal{G}$.
- We can read the factorisation readily from the BN structure: a node is independent of its non-descendants given its parents.
- To each node, a BN associates a local probability model (such as a tabular CPD).
- With a CPD attached to each node, a BN is a complete representation of a joint distribution P.

What's Next?

LC1: BN in code.

HC1b: BN – reasoning and influence.

WC1: exercises (semantics, reasoning and influence).

Reasoning

Outline for this section

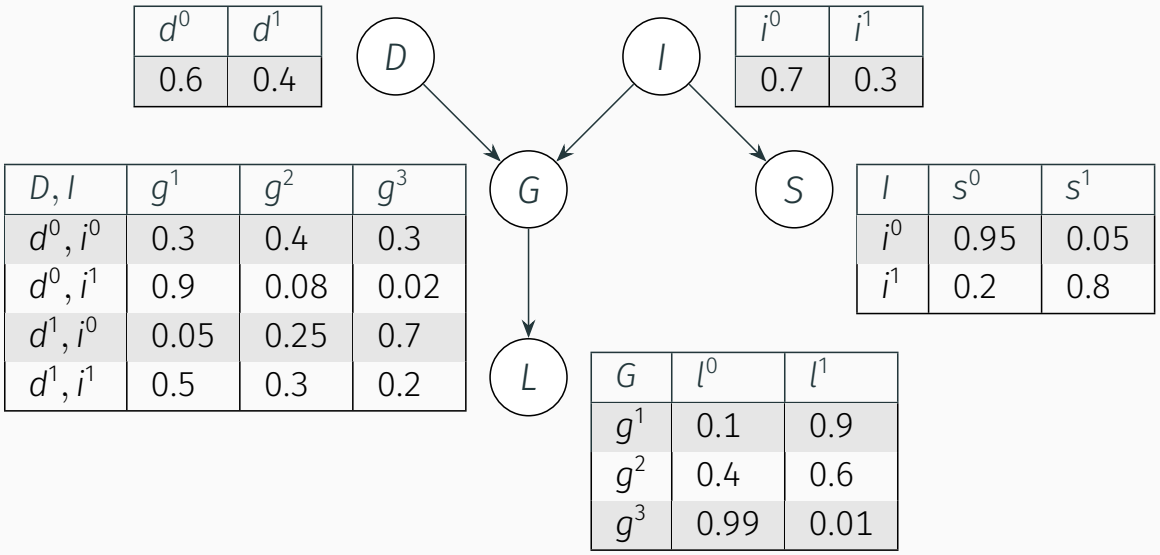
We begin by restating three of the most fundamental results of probability theory.

We then exploit those results and the BN representation to respond to queries about subsets of random variables and sets of observations.

We will start to see a connection between probabilistic inference and manipulation of the BN structure. This connection will be exploited for fully later in the course.

Beliefs

A *joint probability distribution* is a means to represent our uncertainty over the *possible* assignments of the rvs of interest.



$$P(D, I, G, S, L) = P(D)P(I)P(G|D, I)P(S|I)P(L|G)$$

With its DAG and CPDs, a BN identifies one such distribution.

Having a joint distribution, we can **reason about subsets of variables** and **given observed data**.

Probability Calculus

Chain rule

$$P(A, B) = P(A)P(B|A) = P(B)P(A|B) \tag{2}$$

Marginalisation

$$P(A) = \sum_{B} P(A, B) \tag{3}$$

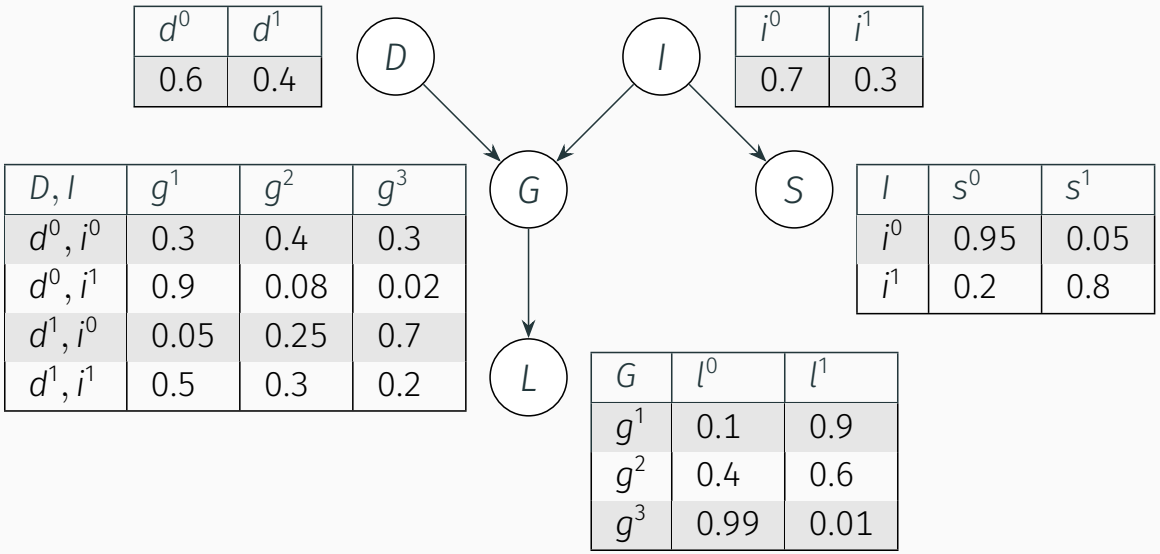
Conditioning

$$P(B|A) = \frac{P(A, B)}{P(A)} \tag{4}$$

These results extend to the cases where A and/or B are sets of rvs.

Marginal Queries

We can reason about *subsets* of rvs via *marginalisation*.



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

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

I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

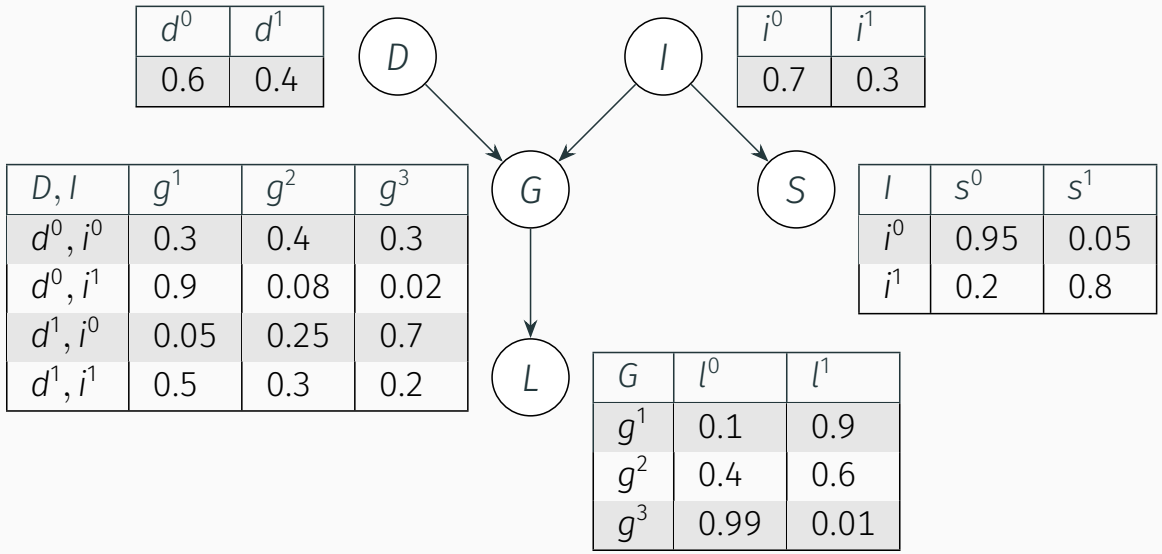
G	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

$$P(D, I, G, S, L) = P(D)P(I)P(G|D, I)P(S|I)P(L|G)$$

Evaluate $P(S = s^1)$

Marginal Queries

We can reason about *subsets* of rvs via *marginalisation*.



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

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

I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

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

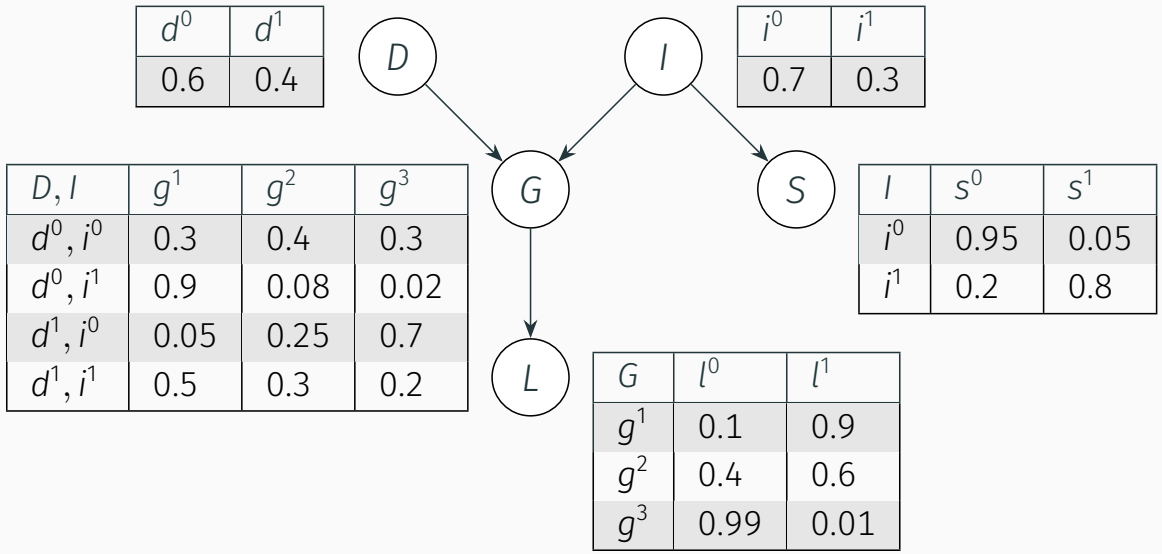
$$P(D, I, G, S, L) = P(D)P(I)P(G|D, I)P(S|I)P(L|G)$$

Evaluate $P(S = s^1) = 0.275$

Evaluate $P(G = g^1)$

Marginal Queries

We can reason about *subsets* of rvs via *marginalisation*.



$$P(D, I, G, S, L) = P(D)P(I)P(G|D, I)P(S|I)P(L|G)$$

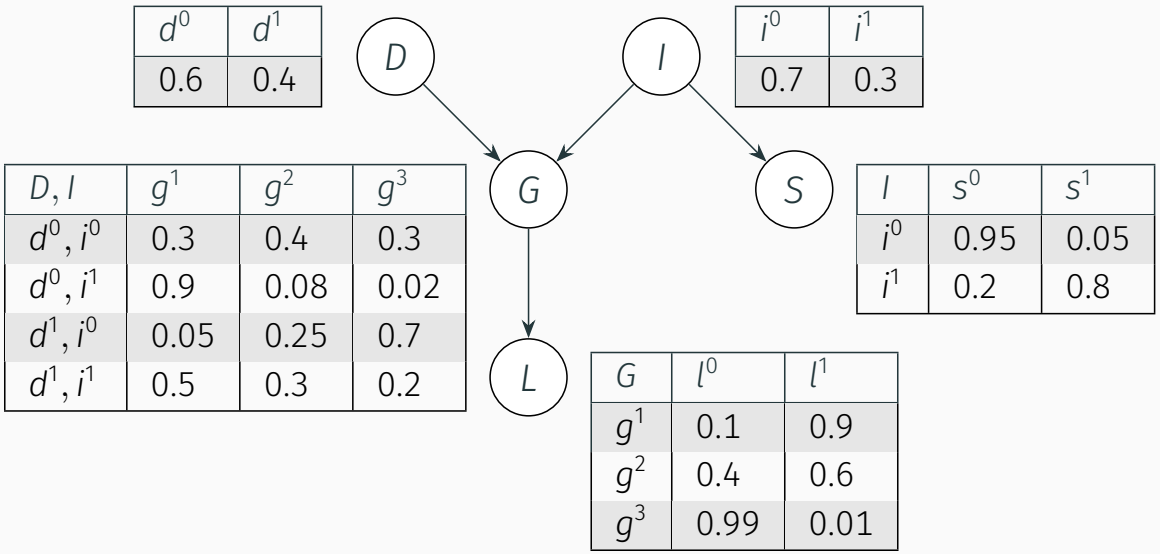
Evaluate $P(S = s^1) = 0.275$

Evaluate $P(G = g^1) = 0.3620$

Evaluate $P(L = l^1)$

Marginal Queries

We can reason about *subsets* of rvs via *marginalisation*.



$$P(D, I, G, S, L) = P(D)P(I)P(G|D, I)P(S|I)P(L|G)$$

Evaluate $P(S = s^1) = 0.275$

Evaluate $P(G = g^1) = 0.3620$

Evaluate $P(L = l^1) \approx 0.5023$

Marginal $P(S) =$

$$\sum_{D,I,G,L} P(D, I, G, S, L)$$

Marginal $P(S) =$

$$\sum_{D,I,G,L} P(D, I, G, S, L)$$

$$= \sum_{D,I,G,L} P(S|I)P(I)P(D)P(G|D, I)P(L|G)$$

Marginal $P(S) =$

$$\sum_{D,I,G,L} P(D, I, G, S, L)$$

$$= \sum_{D,I,G,L} P(S|I)P(I)P(D)P(G|D, I)P(L|G)$$

$$= \sum_{D,I,G} P(S|I)P(I)P(D)P(G|D, I) \sum_{L} P(L|G)$$

Marginal $P(S) =$

$$\sum_{D,I,G,L} P(D, I, G, S, L)$$

$$= \sum_{D,I,G,L} P(S|I)P(I)P(D)P(G|D, I)P(L|G)$$

$$= \sum_{D,I,G} P(S|I)P(I)P(D)P(G|D, I) \sum_{L} P(L|G)$$

$$= \sum_{D,I,G} P(S|I)P(I)P(D)P(G|D, I)$$

Marginal $P(S) =$

$$\sum_{D,I,G,L} P(D, I, G, S, L)$$

$$= \sum_{D,I,G,L} P(S|I)P(I)P(D)P(G|D, I)P(L|G)$$

$$= \sum_{D,I,G} P(S|I)P(I)P(D)P(G|D, I) \sum_{L} P(L|G)$$

$$= \sum_{D,I,G} P(S|I)P(I)P(D)P(G|D, I)$$

$$= \sum_{D,I} P(S|I)P(I)P(D) \sum_{G} P(G|D, I)$$

Marginal $P(S) =$

$$\sum_{D,I,G,L} P(D,I,G,S,L)$$

$$= \sum_{D,I,G,L} P(S|I)P(I)P(D)P(G|D,I)P(L|G)$$

$$= \sum_{D,I,G} P(S|I)P(I)P(D)P(G|D,I) \sum_{L} P(L|G)$$

$$= \sum_{D,I,G} P(S|I)P(I)P(D)P(G|D,I)$$

$$= \sum_{D,I} P(S|I)P(I)P(D) \sum_{G} P(G|D,I)$$

$$= \sum_{D,I} P(S|I)P(I)P(D)$$

Marginal $P(S) =$

$$\sum_{D,I,G,L} P(D, I, G, S, L)$$

$$= \sum_{D,I,G,L} P(S|I)P(I)P(D)P(G|D, I)P(L|G)$$

$$= \sum_{D,I,G} P(S|I)P(I)P(D)P(G|D, I) \sum_{L} P(L|G)$$

$$= \sum_{D,I,G} P(S|I)P(I)P(D)P(G|D, I)$$

$$= \sum_{D,I} P(S|I)P(I)P(D) \sum_{G} P(G|D, I)$$

$$= \sum_{D,I} P(S|I)P(I)P(D)$$

$$= \sum_{I} P(S|I)P(I) \sum_{D} P(D)$$

Marginal $P(S) =$

$$\sum_{D,I,G,L} P(D, I, G, S, L)$$

$$= \sum_{D,I,G,L} P(S|I)P(I)P(D)P(G|D, I)P(L|G)$$

$$= \sum_{D,I,G} P(S|I)P(I)P(D)P(G|D, I) \sum_{L} P(L|G)$$

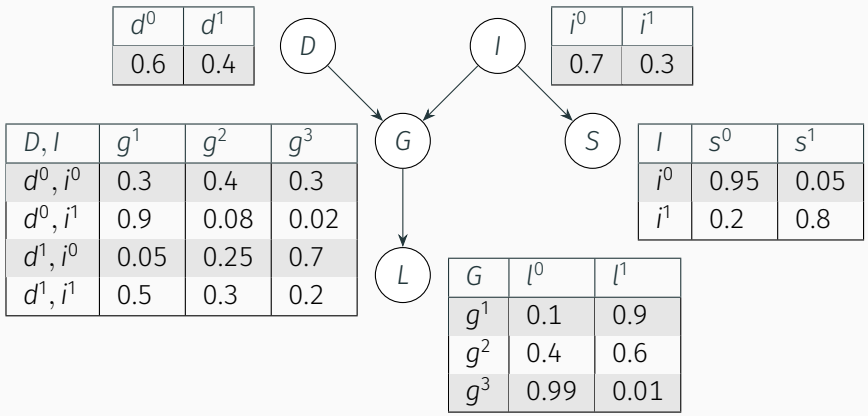
$$= \sum_{D,I,G} P(S|I)P(I)P(D)P(G|D, I)$$

$$= \sum_{D,I} P(S|I)P(I)P(D) \sum_{G} P(G|D, I)$$

$$= \sum_{D,I} P(S|I)P(I)P(D)$$

$$= \sum_{I} P(S|I)P(I) \sum_{D} P(D)$$

$$= \sum_{I} P(S|I)P(I)$$



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

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

I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

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

Marginal $P(S) =$

$$\sum_{D,I,G,L} P(D,I,G,S,L)$$

$$= \sum_{D,I,G,L} P(S|I)P(I)P(D)P(G|D,I)P(L|G)$$

$$= \sum_{D,I,G} P(S|I)P(I)P(D)P(G|D,I) \sum_{L} P(L|G)$$

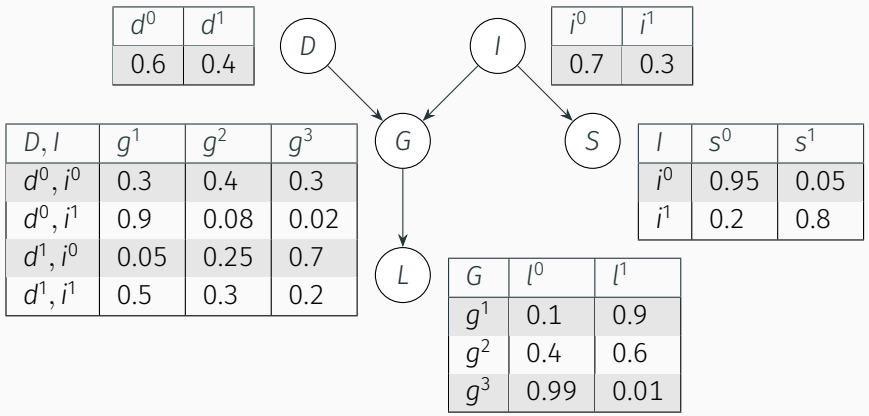
$$= \sum_{D,I,G} P(S|I)P(I)P(D)P(G|D,I)$$

$$= \sum_{D,I} P(S|I)P(I)P(D) \sum_{G} P(G|D,I)$$

$$= \sum_{D,I} P(S|I)P(I)P(D)$$

$$= \sum_{I} P(S|I)P(I) \sum_{D} P(D)$$

$$= \sum_{I} P(S|I)P(I)$$



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

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

I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

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

S	$\sum_I P(S\|I)P(I)$	$P(S)$
s^0	$0.95 \times 0.7 + 0.2 \times 0.3$	0.725
s^1	$0.05 \times 0.7 + 0.8 \times 0.3$	0.275

Marginal $P(G) =$

$$\sum_{D,I,S,L} P(D, I, G, S, L)$$

Marginal $P(G) =$

$$\sum_{D,I,S,L} P(D, I, G, S, L)$$

$$= \sum_{D,I,S,L} P(G|D, I)P(D)P(I)P(S|I)P(L|G)$$

Marginal $P(G) =$

$$\sum_{D,I,S,L} P(D, I, G, S, L)$$

$$= \sum_{D,I,S,L} P(G|D, I)P(D)P(I)P(S|I)P(L|G)$$

$$= \sum_{D,I,S} P(G|D, I)P(D)P(I)P(S|I) \sum_{L} P(L|G)$$

Marginal $P(G) =$

$$\sum_{D,I,S,L} P(D, I, G, S, L)$$

$$= \sum_{D,I,S,L} P(G|D, I)P(D)P(I)P(S|I)P(L|G)$$

$$= \sum_{D,I,S} P(G|D, I)P(D)P(I)P(S|I) \sum_{L} P(L|G)$$

$$= \sum_{D,I,S} P(G|D, I)P(D)P(I)P(S|I)$$

Marginal $P(G) =$

$$\sum_{D,I,S,L} P(D, I, G, S, L)$$

$$= \sum_{D,I,S,L} P(G|D, I)P(D)P(I)P(S|I)P(L|G)$$

$$= \sum_{D,I,S} P(G|D, I)P(D)P(I)P(S|I) \sum_{L} P(L|G)$$

$$= \sum_{D,I,S} P(G|D, I)P(D)P(I)P(S|I)$$

$$= \sum_{D,I} P(G|D, I)P(D)P(I) \sum_{S} P(S|I)$$

Marginal $P(G) =$

$$\sum_{D,I,S,L} P(D, I, G, S, L)$$

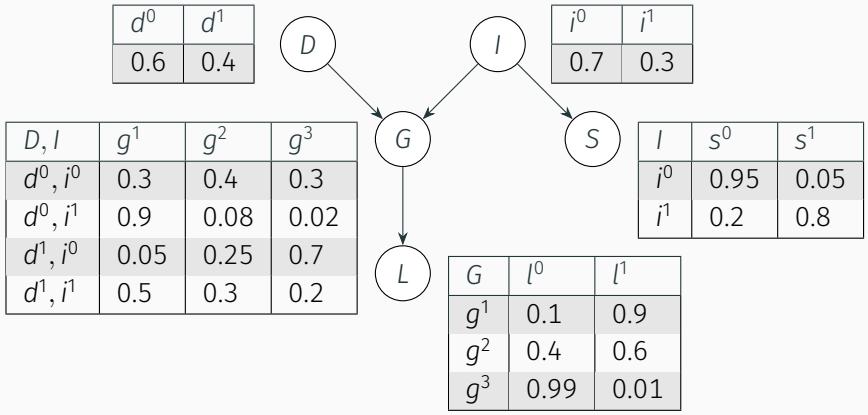
$$= \sum_{D,I,S,L} P(G|D, I)P(D)P(I)P(S|I)P(L|G)$$

$$= \sum_{D,I,S} P(G|D, I)P(D)P(I)P(S|I) \sum_{L} P(L|G)$$

$$= \sum_{D,I,S} P(G|D, I)P(D)P(I)P(S|I)$$

$$= \sum_{D,I} P(G|D, I)P(D)P(I) \sum_{S} P(S|I)$$

$$= \sum_{D,I} P(G|D, I)P(D)P(I)$$



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

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

I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

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

Marginal $P(G) =$

$$\sum_{D,I,S,L} P(D,I,G,S,L)$$

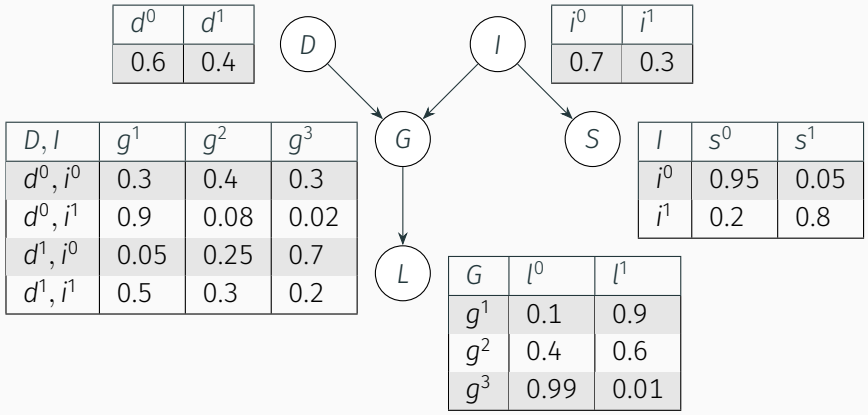
$$= \sum_{D,I,S,L} P(G|D,I)P(D)P(I)P(S|I)P(L|G)$$

$$= \sum_{D,I,S} P(G|D,I)P(D)P(I)P(S|I) \sum_{L} P(L|G)$$

$$= \sum_{D,I,S} P(G|D,I)P(D)P(I)P(S|I)$$

$$= \sum_{D,I} P(G|D,I)P(D)P(I) \sum_{S} P(S|I)$$

$$= \sum_{D,I} P(G|D,I)P(D)P(I)$$



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

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

I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

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

G	$\sum_{D,I} P(G\|D,I)P(D)P(I)$	$P(G)$
g^1	$0.3\times0.6\times0.7 + 0.90\times0.6\times0.3 + 0.05\times0.4\times0.7 + 0.5\times0.4\times0.3$	0.3620
g^2	$0.4\times0.6\times0.7 + 0.08\times0.6\times0.3 + 0.25\times0.4\times0.7 + 0.3\times0.4\times0.3$	0.2884
g^3	$0.3\times0.6\times0.7 + 0.02\times0.6\times0.3 + 0.70\times0.4\times0.7 + 0.2\times0.4\times0.3$	0.3496

Marginal $P(L) =$

$$\sum_{D,I,G,S} P(D, I, G, S, L)$$

Marginal $P(L) =$

$$\sum_{D,I,G,S} P(D, I, G, S, L)$$

$$= \sum_{D,I,G,S} P(L|G)P(D)P(I)P(G|D, I)P(S|I)$$

Marginal $P(L) =$

$$\sum_{D,I,G,S} P(D, I, G, S, L)$$

$$= \sum_{D,I,G,S} P(L|G)P(D)P(I)P(G|D, I)P(S|I)$$

$$= \sum_{D,I,G} P(L|G)P(D)P(I)P(G|D, I) \sum_{S} P(S|I)$$

Marginal $P(L) =$

$$\sum_{D,I,G,S} P(D, I, G, S, L)$$

$$= \sum_{D,I,G,S} P(L|G)P(D)P(I)P(G|D, I)P(S|I)$$

$$= \sum_{D,I,G} P(L|G)P(D)P(I)P(G|D, I) \sum_{S} P(S|I)$$

$$= \sum_{D,I,G} P(L|G)P(D)P(I)P(G|D, I)$$

Marginal $P(L) =$

$$\sum_{D,I,G,S} P(D, I, G, S, L)$$

$$= \sum_{D,I,G,S} P(L|G)P(D)P(I)P(G|D, I)P(S|I)$$

$$= \sum_{D,I,G} P(L|G)P(D)P(I)P(G|D, I) \sum_{S} P(S|I)$$

$$= \sum_{D,I,G} P(L|G)P(D)P(I)P(G|D, I)$$

$$= \sum_{G} P(L|G) \sum_{D,I} P(D)P(I)P(G|D, I)$$

Marginal $P(L) =$

$$\sum_{D,I,G,S} P(D, I, G, S, L)$$

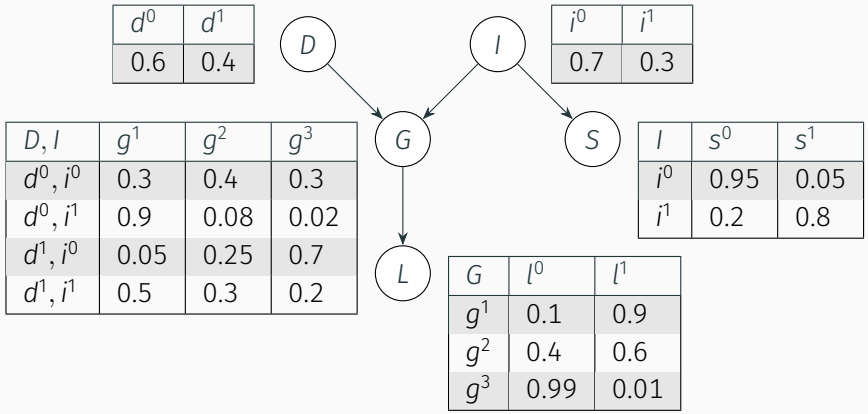
$$= \sum_{D,I,G,S} P(L|G)P(D)P(I)P(G|D, I)P(S|I)$$

$$= \sum_{D,I,G} P(L|G)P(D)P(I)P(G|D, I) \sum_{S} P(S|I)$$

$$= \sum_{D,I,G} P(L|G)P(D)P(I)P(G|D, I)$$

$$= \sum_{G} P(L|G) \sum_{D,I} P(D)P(I)P(G|D, I)$$

$$= \sum_{G} P(L|G)P(G)$$



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

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

I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

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

Marginal $P(L) =$

$$\sum_{D,I,G,S} P(D,I,G,S,L)$$

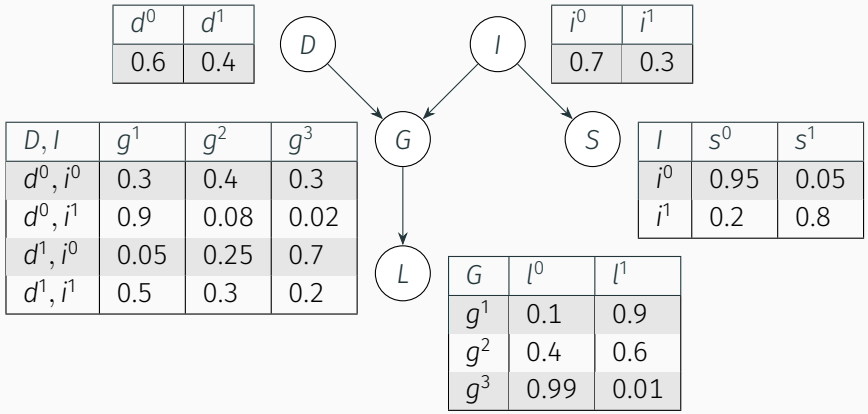
$$= \sum_{D,I,G,S} P(L|G)P(D)P(I)P(G|D,I)P(S|I)$$

$$= \sum_{D,I,G} P(L|G)P(D)P(I)P(G|D,I) \sum_{S} P(S|I)$$

$$= \sum_{D,I,G} P(L|G)P(D)P(I)P(G|D,I)$$

$$= \sum_{G} P(L|G) \sum_{D,I} P(D)P(I)P(G|D,I)$$

$$= \sum_{G} P(L|G)P(G)$$



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

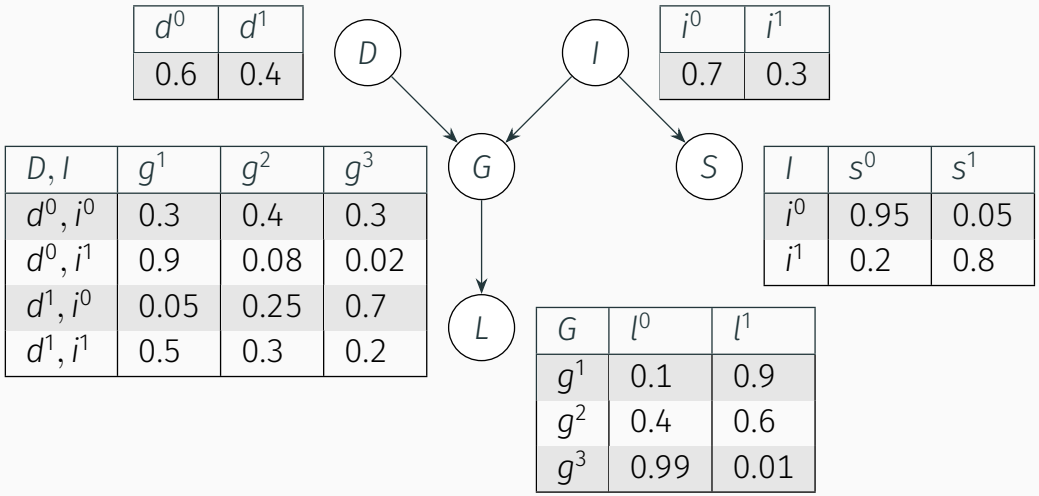
D, I	g^1	g^2	g^3
d^0, i^0	0.3	0.4	0.3
d^0, i^1	0.9	0.08	0.02
d^1, i^0	0.05	0.25	0.7
d^1, i^1	0.5	0.3	0.2

I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

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

L	$\sum_G P(L\|G)P(G)$	$P(L)$
l^0	$0.1 \times 0.3630 + 0.4 \times 0.2884 + 0.99 \times 0.3496$	0.4977
l^1	$0.9 \times 0.3630 + 0.6 \times 0.2884 + 0.01 \times 0.3496$	0.5023

Single-Variable Marginals



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

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

I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

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

$$P(D, I, G, S, L) = P(D)P(I)P(G|D, I)P(S|I)P(L|G)$$

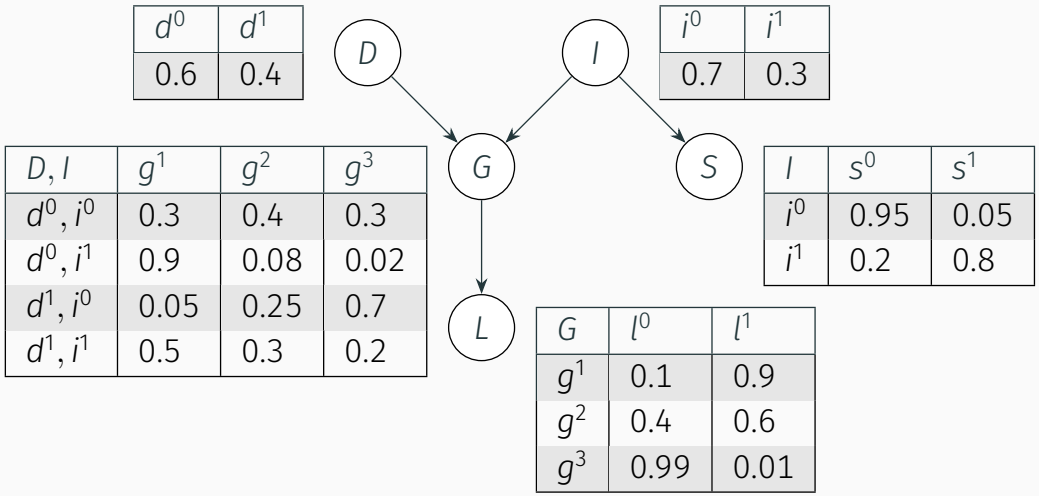
$P(D)$ and $P(I)$ are given CPDs

The other single-variable marginals are CPDs which we inferred via

- $P(S) = \sum_I P(I)P(S|I)$
- $P(G) = \sum_{D,I} P(D)P(I)P(G|D, I)$
- $P(L) = \sum_G P(L|G)P(G)$

as $P(S|I)$, $P(G|D, I)$, and $P(L|G)$ are also given CPDs.

Other Marginals



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

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

I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

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

$$P(D, I, G, S, L) = P(D)P(I)P(G|D, I)P(S|I)P(L|G)$$

We can similarly express any other marginal, involving any subset of rvs, in terms of CPDs that are given:

- $P(D, I)$
- $P(D, G)$
- $P(D, L)$
- $P(I, G, S)$
- $P(S, G, L)$
- $\ldots$

Evaluate $P(S = s^1, I = i^1)$, $P(G = g^1, I = i^1)$, $P(L = l^1, I = i^1)$

Marginal $P(D, G) =$

$$\sum_{I,S,L} P(D, I, G, S, L)$$

Marginal $P(D, G) =$

$$\sum_{I,S,L} P(D, I, G, S, L)$$

$$= \sum_{I,S,L} P(D)P(I)P(G|D, I)P(S|I)P(L|G)$$

Marginal $P(D, G) =$

$$\sum_{I,S,L} P(D, I, G, S, L)$$

$$= \sum_{I,S,L} P(D)P(I)P(G|D, I)P(S|I)P(L|G)$$

$$= \sum_{I,S} P(D)P(I)P(G|D, I)P(S|I) \sum_{L} P(L|G)$$

Marginal $P(D, G) =$

$$\sum_{I,S,L} P(D, I, G, S, L)$$

$$= \sum_{I,S,L} P(D)P(I)P(G|D, I)P(S|I)P(L|G)$$

$$= \sum_{I,S} P(D)P(I)P(G|D, I)P(S|I) \sum_{L} P(L|G)$$

$$= \sum_{I} P(D)P(I)P(G|D, I) \sum_{S} P(S|I)$$

Marginal $P(D, G) =$

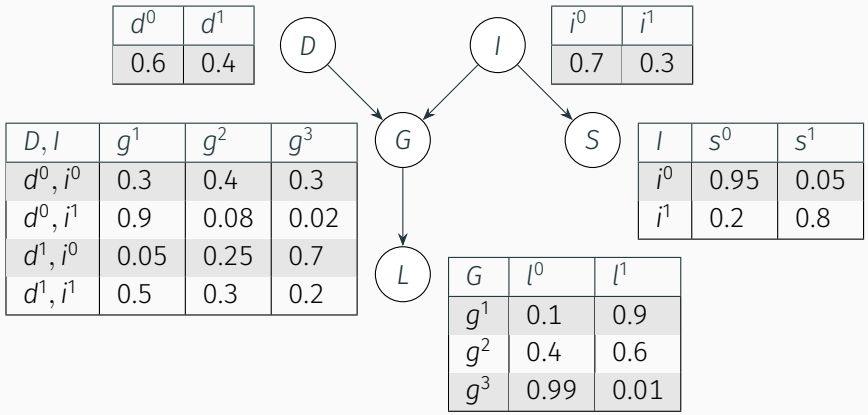
$$\sum_{I,S,L} P(D, I, G, S, L)$$

$$= \sum_{I,S,L} P(D)P(I)P(G|D, I)P(S|I)P(L|G)$$

$$= \sum_{I,S} P(D)P(I)P(G|D, I)P(S|I) \sum_{L} P(L|G)$$

$$= \sum_{I} P(D)P(I)P(G|D, I) \sum_{S} P(S|I)$$

$$= P(D) \sum_{I} P(I)P(G|D, I)$$



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

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

I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

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

Marginal $P(D, G) =$

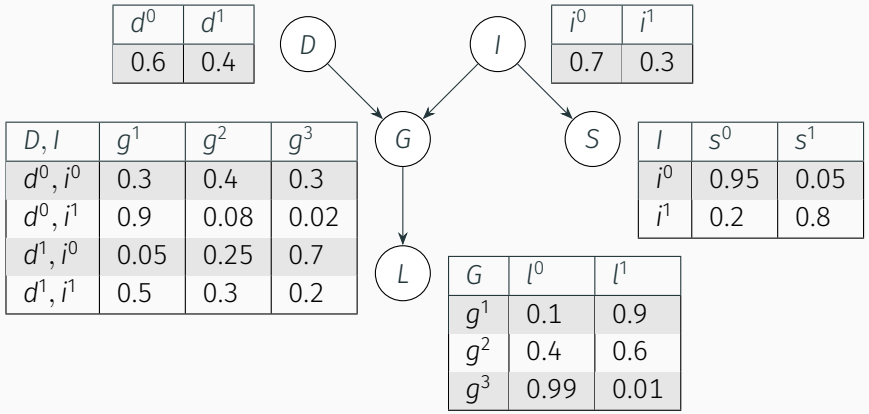
$$\sum_{I,S,L} P(D, I, G, S, L)$$

$$= \sum_{I,S,L} P(D)P(I)P(G|D, I)P(S|I)P(L|G)$$

$$= \sum_{I,S} P(D)P(I)P(G|D, I)P(S|I) \sum_{L} P(L|G)$$

$$= \sum_{I} P(D)P(I)P(G|D, I) \sum_{S} P(S|I)$$

$$= P(D) \sum_{I} P(I)P(G|D, I)$$



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

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

I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

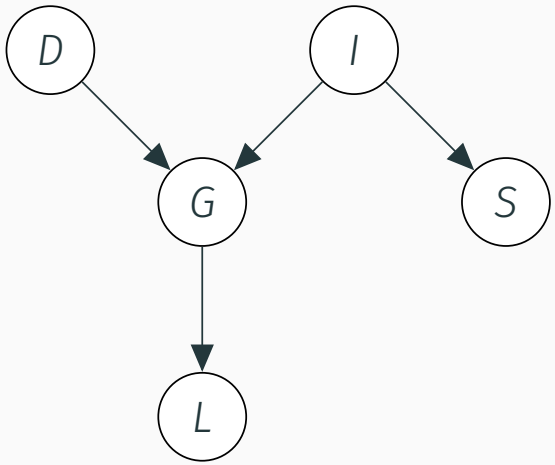
G	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

D, G	Marginalisation	$P(D, G)$
d^0, g^1	$0.6 \times (0.7 \times 0.30 + 0.3 \times 0.90)$	0.288
d^0, g^2	$0.6 \times (0.7 \times 0.40 + 0.3 \times 0.08)$	0.1824
d^0, g^3	$0.6 \times (0.7 \times 0.30 + 0.3 \times 0.02)$	0.1296
d^1, g^1	$0.4 \times (0.7 \times 0.05 + 0.3 \times 0.50)$	0.0740
d^1, g^2	$0.4 \times (0.7 \times 0.25 + 0.3 \times 0.30)$	0.1060
d^1, g^3	$0.4 \times (0.7 \times 0.70 + 0.3 \times 0.20)$	0.220

Ancestors (definition 2.18)

So far, we used probability calculus, the DAG $\mathcal{G}$ only provided us with $\mathcal{I}_l(\mathcal{G})$. Let's try to use more of the graph structure.

A is **an ancestor of** B in the DAG, if the DAG contains at least one directed path from A to B.



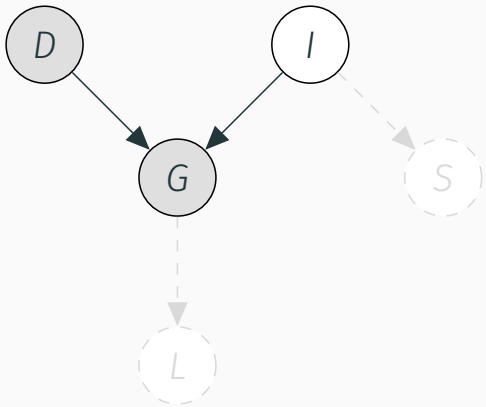
node	ancestors
D	$\emptyset$
I	$\emptyset$
S	$\{I\}$
G	
L	

Can you list the ancestors?

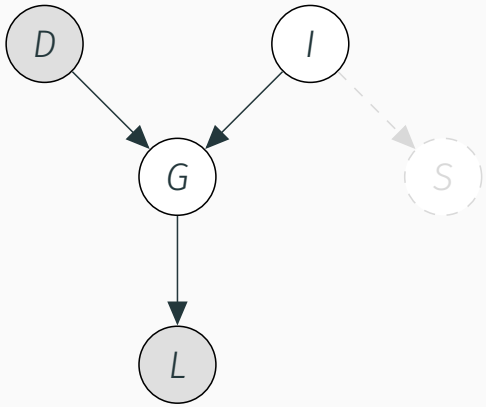
Marginalisation

To obtain marginal probabilities for a subset of nodes (shaded gray)

1. ignore their non-ancestral nodes (faded out);
2. and marginalise out the unassigned ancestors (unshaded).



To express $P(D, G)$, we can act as if L and S were not in the BN, then sum $P(D, G, I)$ over the outcome space of I: $\{i^0, i^1\}$.

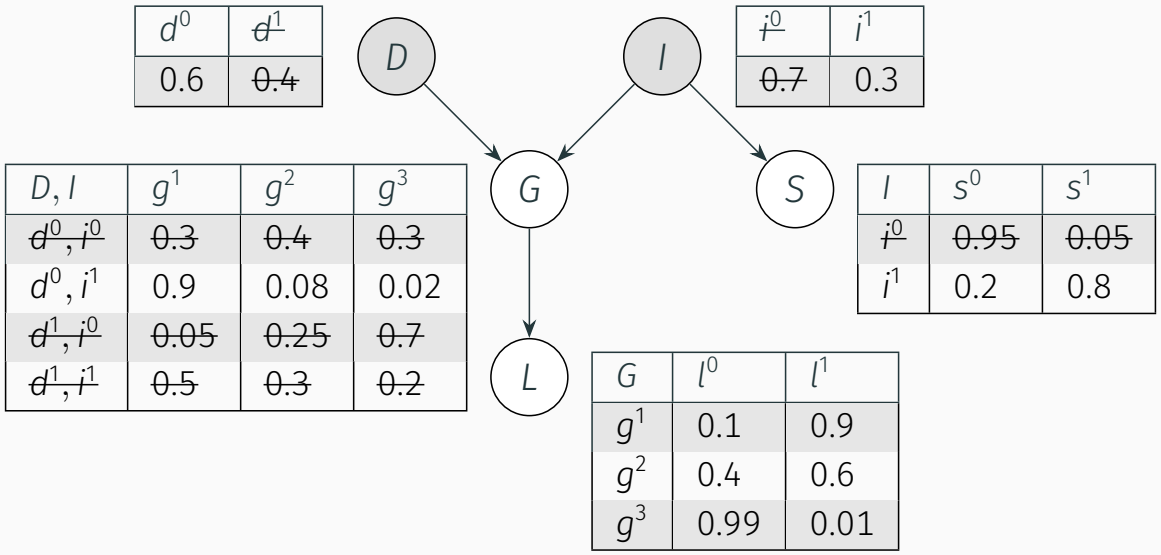


To express $P(D, L)$, we can act as if S were not in the BN, then sum $P(D, L, G, I)$ over the Cartesian product of the outcome spaces of G and I: $\{g^1, g^2, g^3\} \times \{i^0, i^1\}$

Conditional Queries

We can also reason by *conditioning* on subsets of variables.

For example, we may observe $D = d^0$ and $I = i^1$
(shading a node indicates that it has been observed)



d^0	~~d^1~~
0.6	~~0.4~~

~~i^0~~	i^1
~~0.7~~	0.3

D, I	g^1	g^2	g^3
~~d^0, i^0~~	~~0.3~~	~~0.4~~	~~0.3~~
d^0, i^1	0.9	0.08	0.02
~~d^1, i^0~~	~~0.05~~	~~0.25~~	~~0.7~~
~~d^1, i^1~~	~~0.5~~	~~0.3~~	~~0.2~~

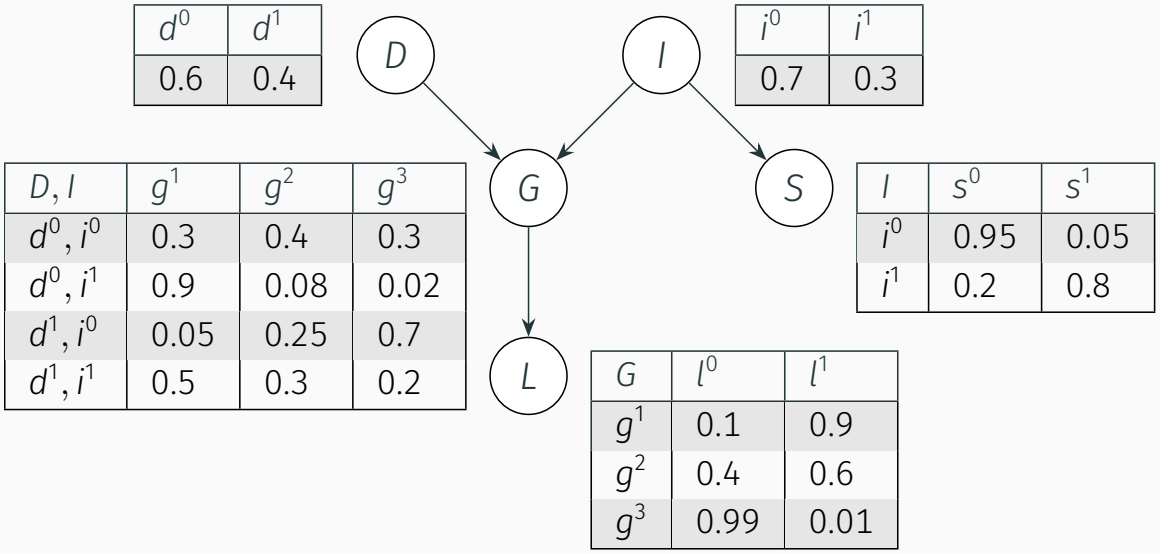
I	s^0	s^1
~~i^0~~	~~0.95~~	~~0.05~~
i^1	0.2	0.8

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

and ask ourselves what happens to $P(G|D = d^0, I = i^1)$, or $P(S|D = d^0, I = i^1)$, or $P(L|D = d^0, I = i^1)$, etc.

Conditional Queries – Causal Reasoning

When we condition on an *ancestor* of the variable of interest, we say we are reasoning 'causally'.



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

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

I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

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

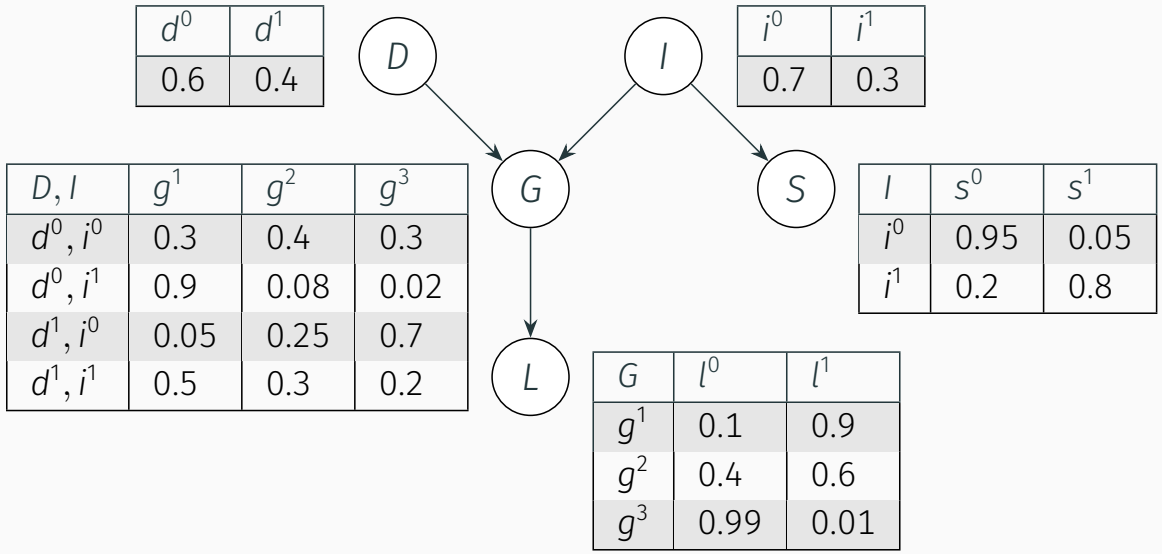
$$P(D, I, G, S, L) = P(D)P(I)P(G|D, I)P(S|I)P(L|G)$$

Evaluate $P(S = s^1|I = i^1)$

For comparison: $P(S = s^1) = 0.275$, $P(G = g^1) = 0.3620$ and $P(L = l^1) = 0.502336$

Conditional Queries – Causal Reasoning

When we condition on an *ancestor* of the variable of interest, we say we are reasoning 'causally'.



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

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

I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

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

$$P(D, I, G, S, L) = P(D)P(I)P(G|D, I)P(S|I)P(L|G)$$

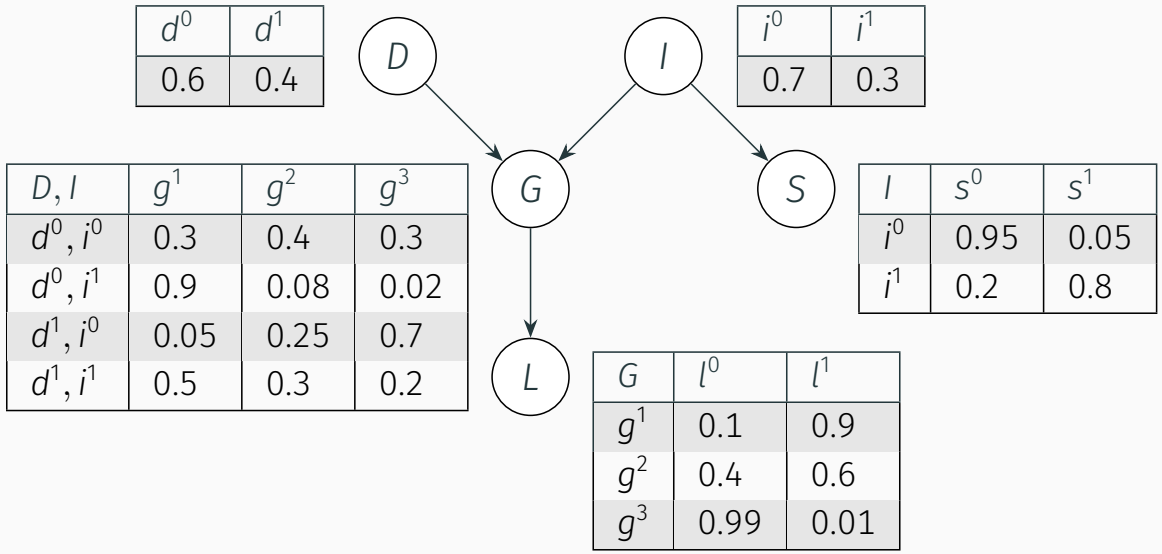
Evaluate $P(S = s^1|I = i^1) = 0.8$

Evaluate $P(G = g^1|I = i^1)$

For comparison: $P(S = s^1) = 0.275$, $P(G = g^1) = 0.3620$ and $P(L = l^1) = 0.502336$

Conditional Queries – Causal Reasoning

When we condition on an *ancestor* of the variable of interest, we say we are reasoning 'causally'.



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

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

I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

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

$$P(D, I, G, S, L) = P(D)P(I)P(G|D, I)P(S|I)P(L|G)$$

Evaluate $P(S = s^1|I = i^1) = 0.8$

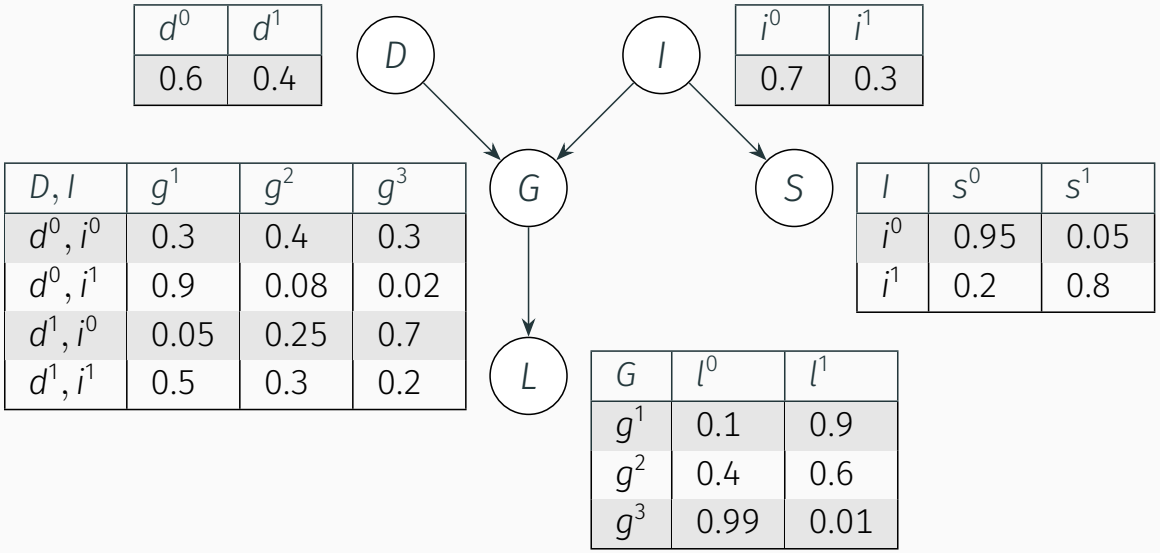
Evaluate $P(G = g^1|I = i^1) = 0.74$

Evaluate $P(L = l^1|I = i^1)$

For comparison: $P(S = s^1) = 0.275$, $P(G = g^1) = 0.3620$ and $P(L = l^1) = 0.502336$

Conditional Queries – Causal Reasoning

When we condition on an *ancestor* of the variable of interest, we say we are reasoning 'causally'.



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

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

I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

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

$$P(D, I, G, S, L) = P(D)P(I)P(G|D, I)P(S|I)P(L|G)$$

Evaluate $P(S = s^1|I = i^1) = 0.8$

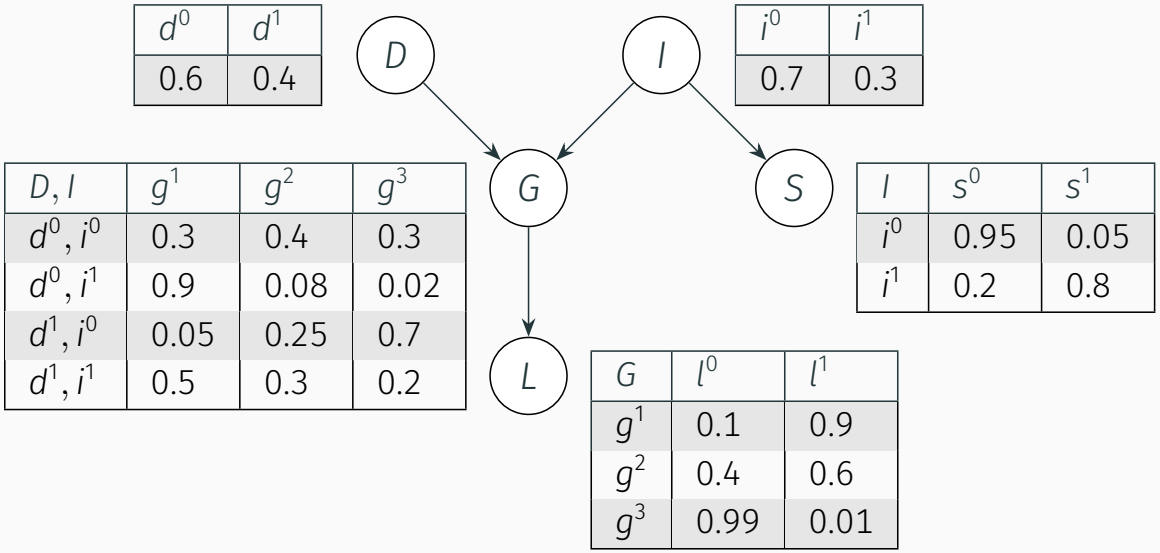
Evaluate $P(G = g^1|I = i^1) = 0.74$

Evaluate $P(L = l^1|I = i^1) \approx 0.7677$

For comparison: $P(S = s^1) = 0.275$, $P(G = g^1) = 0.3620$ and $P(L = l^1) = 0.502336$

Conditional Queries – Evidential Reasoning

When we condition on a *descendant* of the variable of interest, we say we are reasoning 'evidentially'.



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

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

I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

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

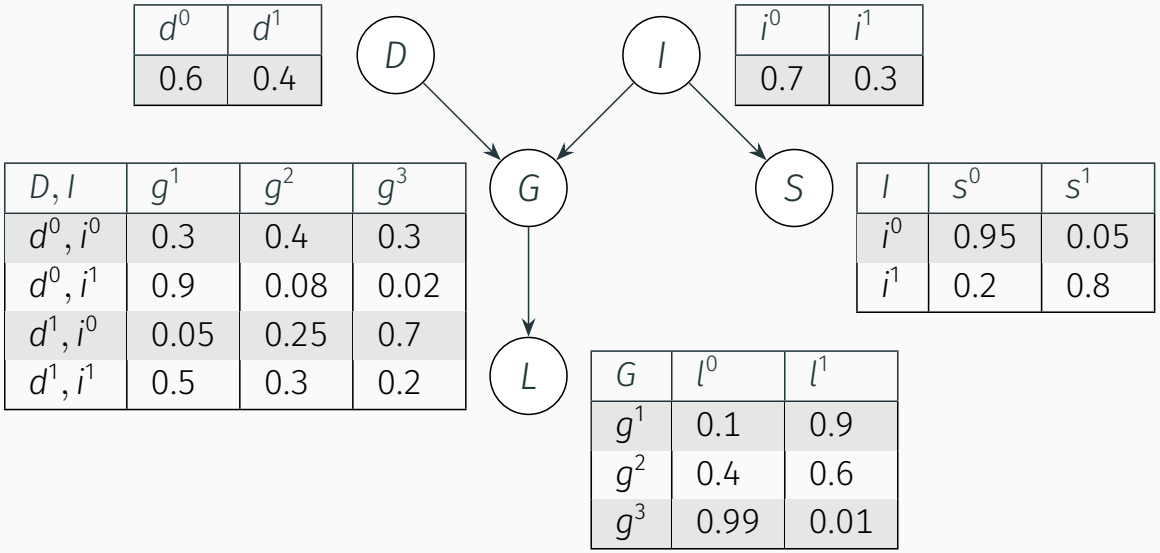
$$P(D, I, G, S, L) = P(D)P(I)P(G|D, I)P(S|I)P(L|G)$$

Evaluate $P(I = i^1|S = s^1)$

Useful marginals: $P(S = s^1) = 0.275$, $P(G = g^1) = 0.3620$ and $P(L = l^1) = 0.502336$

Conditional Queries – Evidential Reasoning

When we condition on a *descendant* of the variable of interest, we say we are reasoning 'evidentially'.



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

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

I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

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

$$P(D, I, G, S, L) = P(D)P(I)P(G|D, I)P(S|I)P(L|G)$$

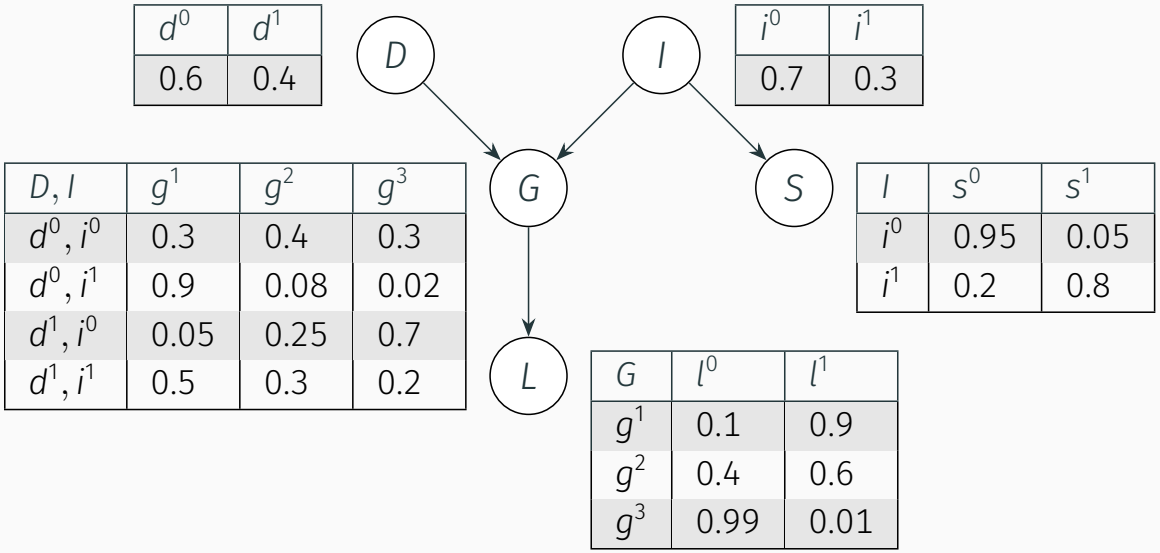
Evaluate $P(I = i^1|S = s^1) = 0.87\bar{7}$

Evaluate $P(I = i^1|G = g^1)$

Useful marginals: $P(S = s^1) = 0.275$, $P(G = g^1) = 0.3620$ and $P(L = l^1) = 0.502336$

Conditional Queries – Evidential Reasoning

When we condition on a *descendant* of the variable of interest, we say we are reasoning 'evidentially'.



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

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

I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

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

$$P(D, I, G, S, L) = P(D)P(I)P(G|D, I)P(S|I)P(L|G)$$

Evaluate $P(I = i^1|S = s^1) = 0.87\bar{7}$

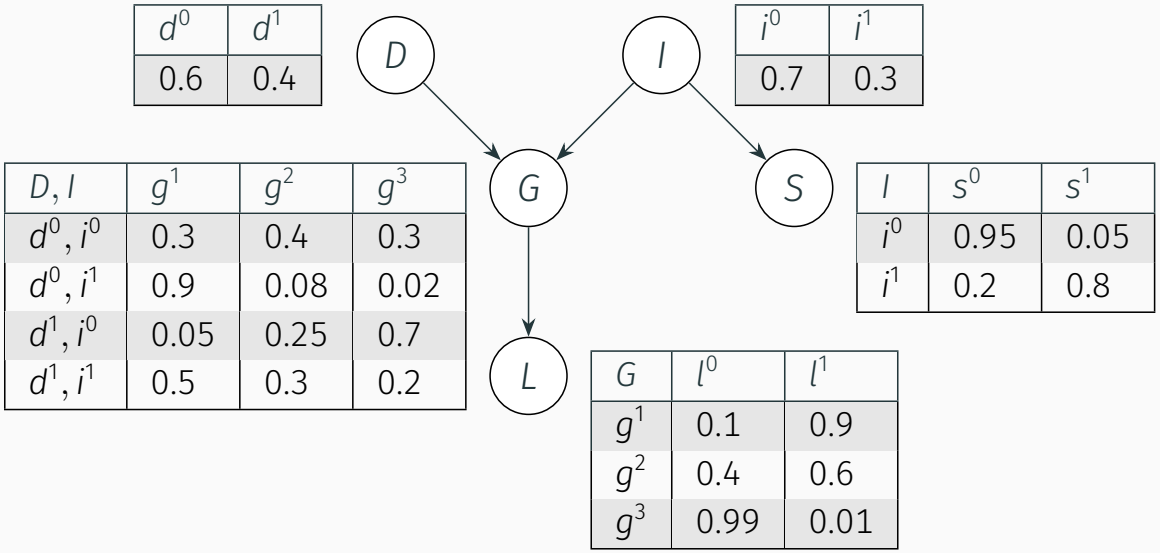
Evaluate $P(I = i^1|G = g^1) \approx 0.6133$

Evaluate $P(I = i^1|L = l^1)$

Useful marginals: $P(S = s^1) = 0.275$, $P(G = g^1) = 0.3620$ and $P(L = l^1) = 0.502336$

Conditional Queries – Evidential Reasoning

When we condition on a *descendant* of the variable of interest, we say we are reasoning 'evidentially'.



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

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

I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

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

$$P(D, I, G, S, L) = P(D)P(I)P(G|D, I)P(S|I)P(L|G)$$

Evaluate $P(I = i^1|S = s^1) = 0.87\bar{7}$

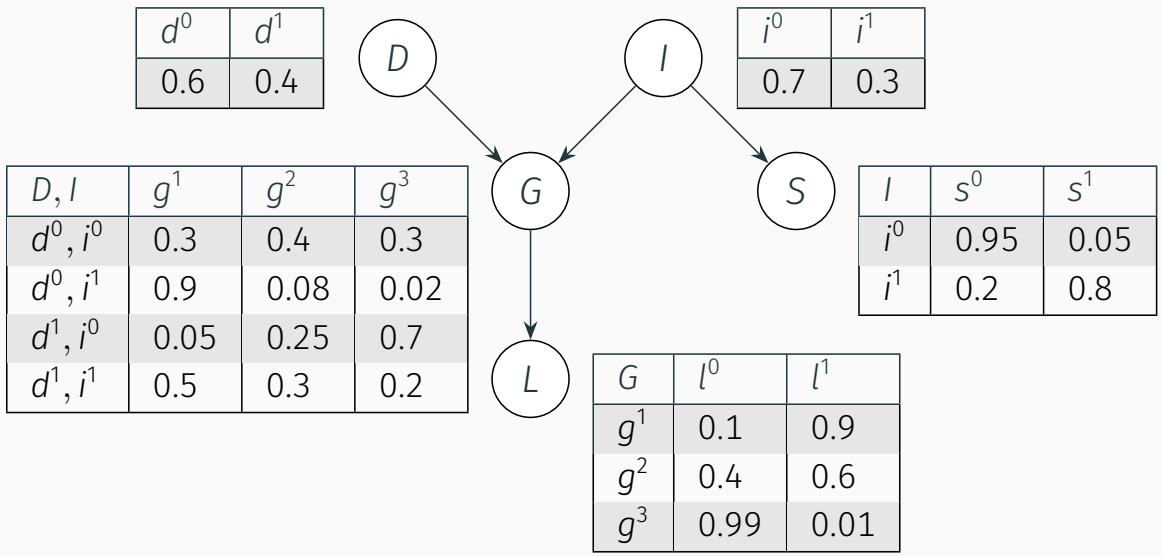
Evaluate $P(I = i^1|G = g^1) \approx 0.6133$

Evaluate $P(I = i^1|L = l^1) \approx 0.4585$

Useful marginals: $P(S = s^1) = 0.275$, $P(G = g^1) = 0.3620$ and $P(L = l^1) = 0.502336$

Conditional Queries – Intercausal Reasoning

Here D and I (the 'causes' of G) are independent of one another, but upon observing their 'effect' G, they can influence each other.



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

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

I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

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

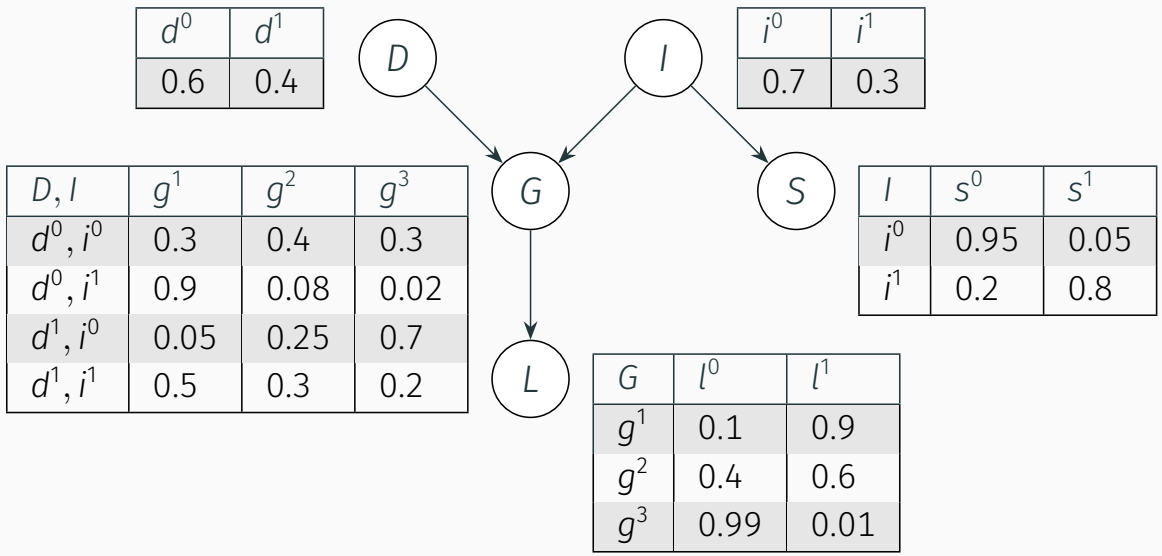
$$P(D, I, G, S, L) = P(D)P(I)P(G|D, I)P(S|I)P(L|G)$$

Evaluate $P(I = i^1|D = d^0)$

Useful marginals: $P(G = g^1) = 0.3620$ and $P(G = g^1, D = d^0) = 0.288$

Conditional Queries – Intercausal Reasoning

Here D and I (the 'causes' of G) are independent of one another, but upon observing their 'effect' G, they can influence each other.



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

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

I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

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

$$P(D, I, G, S, L) = P(D)P(I)P(G|D, I)P(S|I)P(L|G)$$

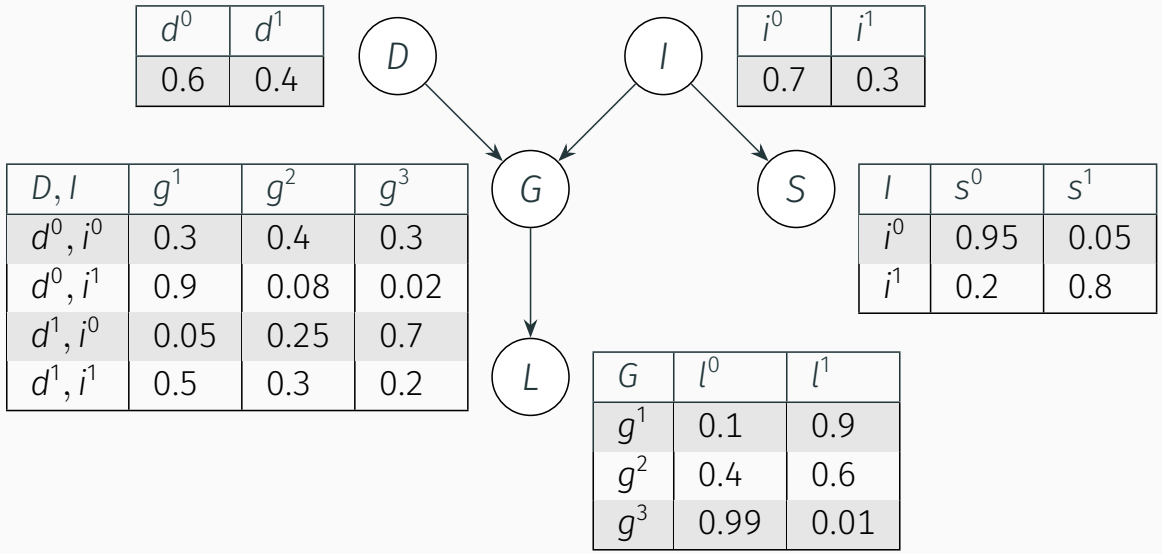
Evaluate $P(I = i^1|D = d^0) = 0.3$ — $I \perp D$ in the BN

Evaluate $P(I = i^1|G = g^1) \approx 0.6133$ — evidential reasoning

Useful marginals: $P(G = g^1) = 0.3620$ and $P(G = g^1, D = d^0) = 0.288$

Conditional Queries – Intercausal Reasoning

Here D and I (the 'causes' of G) are independent of one another, but upon observing their 'effect' G, they can influence each other.



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

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

I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

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

$$P(D, I, G, S, L) = P(D)P(I)P(G|D, I)P(S|I)P(L|G)$$

Evaluate $P(I = i^1|D = d^0) = 0.3$ — $I \perp D$ in the BN

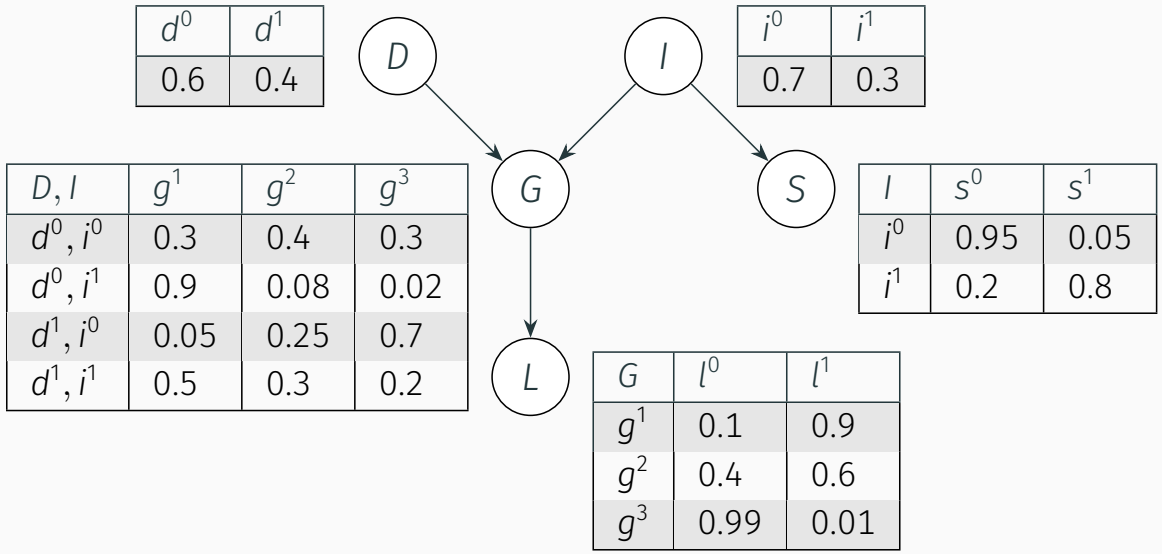
Evaluate $P(I = i^1|G = g^1) \approx 0.6133$ — evidential reasoning

Evaluate $P(I = i^1|G = g^1, D = d^0)$

Useful marginals: $P(G = g^1) = 0.3620$ and $P(G = g^1, D = d^0) = 0.288$

Conditional Queries – Intercausal Reasoning

Here D and I (the 'causes' of G) are independent of one another, but upon observing their 'effect' G, they can influence each other.



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

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

I	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

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

$$P(D, I, G, S, L) = P(D)P(I)P(G|D, I)P(S|I)P(L|G)$$

Evaluate $P(I = i^1|D = d^0) = 0.3$ — $I \perp D$ in the BN

Evaluate $P(I = i^1|G = g^1) \approx 0.6133$ — evidential reasoning

Evaluate $P(I = i^1|G = g^1, D = d^0) = 0.5625$ — intercausal

Useful marginals: $P(G = g^1) = 0.3620$ and $P(G = g^1, D = d^0) = 0.288$

Summary

- Having a BN representation of P (that is, factorisation and CPDs), we can reason about its marginals and conditionals using probability calculus.
- Observing an ancestor or a descendant triggers different reasoning patterns (causal or evidential).
- These patterns show us that the influence of an observation 'flows' through the BN affecting beliefs about other rvs (e.g., a strong letter affects the company's beliefs on high intelligence).

What's Next?

So far the DAG mostly only provided us with statements of local independence (i.e., a factorisation), and we used probability calculus for reasoning.

Next, we develop tools to characterise how influence 'flows' through BNs, enabling some forms of reasoning by analysis of the BN structure alone.

Influence

Outline for this section

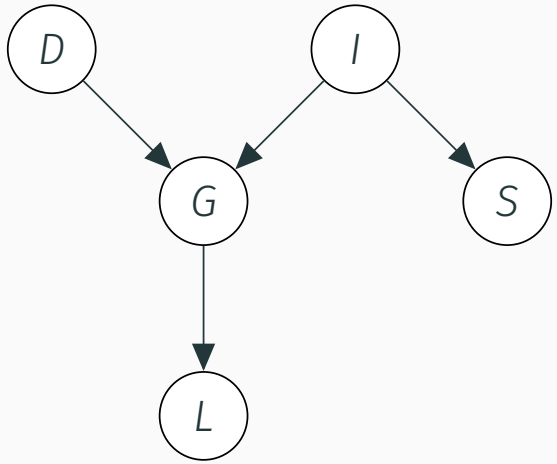
We start by discussing influence in the absence of observations, both direct and indirect influence.

We then introduce observations and review their effects on indirect influence.

We recap the notion of independence, and show how influence allows us to ascertain independence even when we do not have a complete representation of a distribution (e.g., we know its factorisation, but not its factors).

When can X influence Y?

Can knowledge about X influence our beliefs about Y?

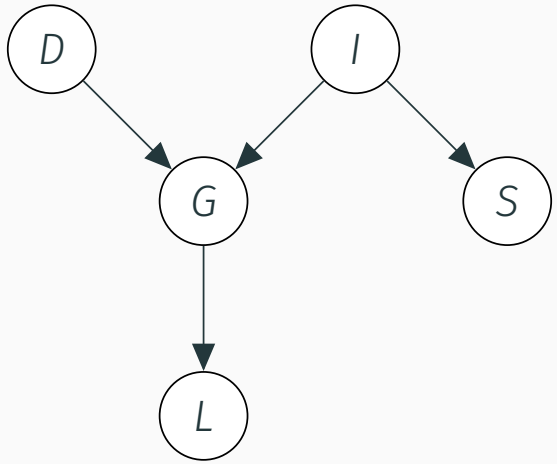


When can X influence Y?

Can knowledge about X influence our beliefs about Y?

Direct influence:

- $X \to Y$ e.g., $D \to G$

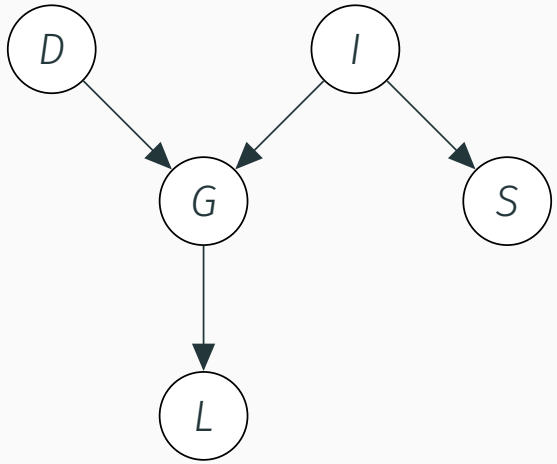


When can X influence Y?

Can knowledge about X influence our beliefs about Y?

Direct influence:

- $X \to Y$ e.g., $D \to G$ ✓ causal effect

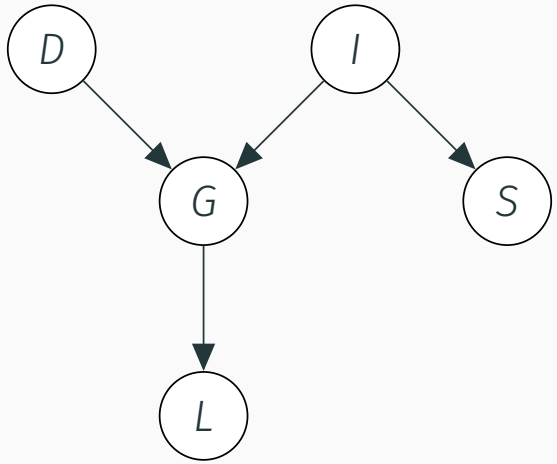


When can X influence Y?

Can knowledge about X influence our beliefs about Y?

Direct influence:

- $X \to Y$ e.g., $D \to G$ ✓ causal effect
- $X \leftarrow Y$ e.g., $G \leftarrow D$

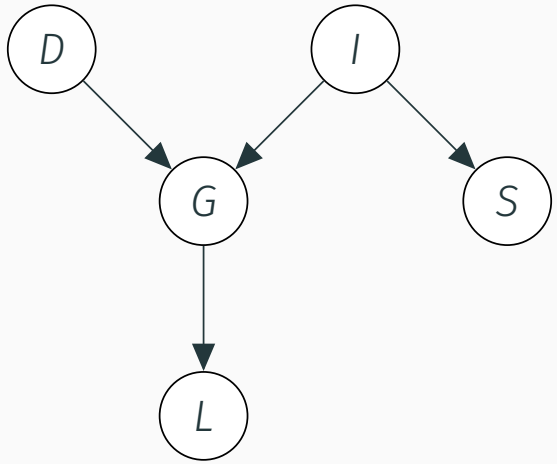


When can X influence Y?

Can knowledge about X influence our beliefs about Y?

Direct influence:

- $X \rightarrow Y$ e.g., $D \rightarrow G$ ✓ causal effect
- $X \leftarrow Y$ e.g., $G \leftarrow D$ ✓ evidential effect



When can X influence Y?

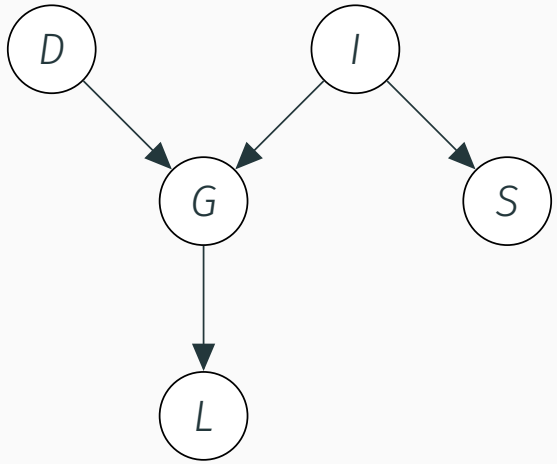
Can knowledge about X influence our beliefs about Y?

Direct influence:

- $X \to Y$ e.g., $D \to G$ ✓ causal effect
- $X \leftarrow Y$ e.g., $G \leftarrow D$ ✓ evidential effect

Indirect influence:

- $X \to W \to Y$ e.g., $D \to G \to L$



When can X influence Y?

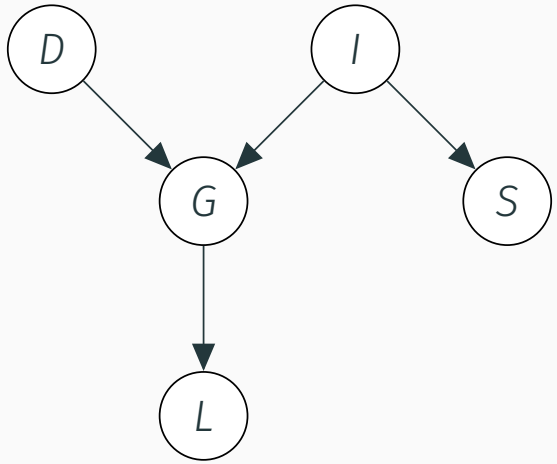
Can knowledge about X influence our beliefs about Y?

Direct influence:

- $X \to Y$ e.g., $D \to G$ ✓ causal effect
- $X \leftarrow Y$ e.g., $G \leftarrow D$ ✓ evidential effect

Indirect influence:

- $X \to W \to Y$ e.g., $D \to G \to L$ ✓
  indirect causal effect



When can X influence Y?

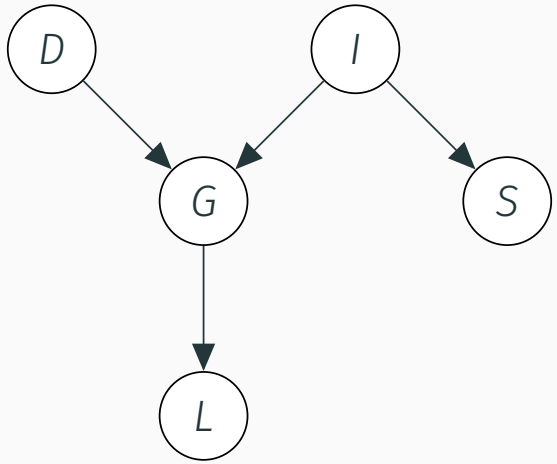
Can knowledge about X influence our beliefs about Y?

Direct influence:

- $X \to Y$ e.g., $D \to G$ ✓ causal effect
- $X \leftarrow Y$ e.g., $G \leftarrow D$ ✓ evidential effect

Indirect influence:

- $X \to W \to Y$ e.g., $D \to G \to L$ ✓ indirect causal effect
- $X \leftarrow W \leftarrow Y$ e.g., $L \leftarrow G \leftarrow D$



When can X influence Y?

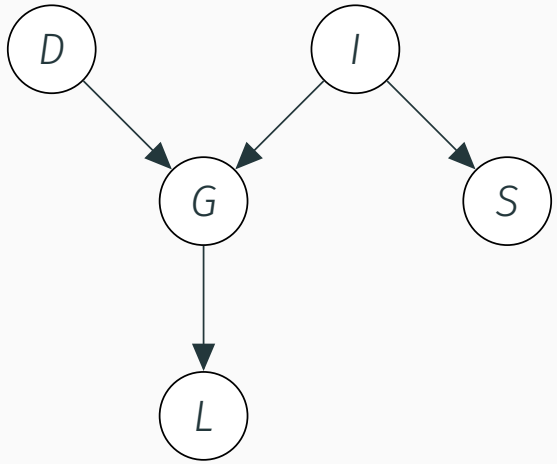
Can knowledge about X influence our beliefs about Y?

Direct influence:

- $X \rightarrow Y$ e.g., $D \rightarrow G$ ✓ causal effect
- $X \leftarrow Y$ e.g., $G \leftarrow D$ ✓ evidential effect

Indirect influence:

- $X \rightarrow W \rightarrow Y$ e.g., $D \rightarrow G \rightarrow L$ ✓
  indirect causal effect
- $X \leftarrow W \leftarrow Y$ e.g., $L \leftarrow G \leftarrow D$ ✓
  indirect evidential effect



When can X influence Y?

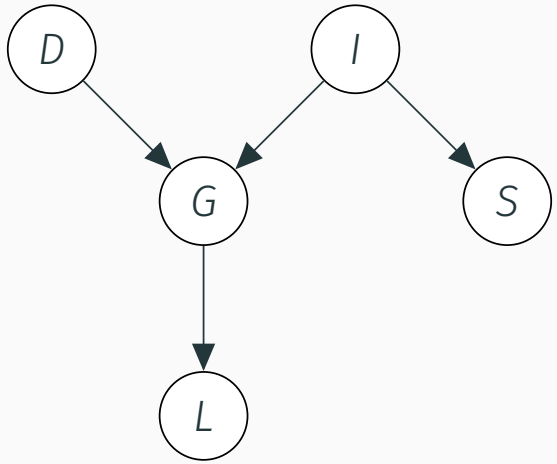
Can knowledge about X influence our beliefs about Y?

Direct influence:

- $X \rightarrow Y$ e.g., $D \rightarrow G$ ✓ causal effect
- $X \leftarrow Y$ e.g., $G \leftarrow D$ ✓ evidential effect

Indirect influence:

- $X \rightarrow W \rightarrow Y$ e.g., $D \rightarrow G \rightarrow L$ ✓
  indirect causal effect
- $X \leftarrow W \leftarrow Y$ e.g., $L \leftarrow G \leftarrow D$ ✓
  indirect evidential effect
- $X \leftarrow W \rightarrow Y$ e.g., $G \leftarrow I \rightarrow S$



When can X influence Y?

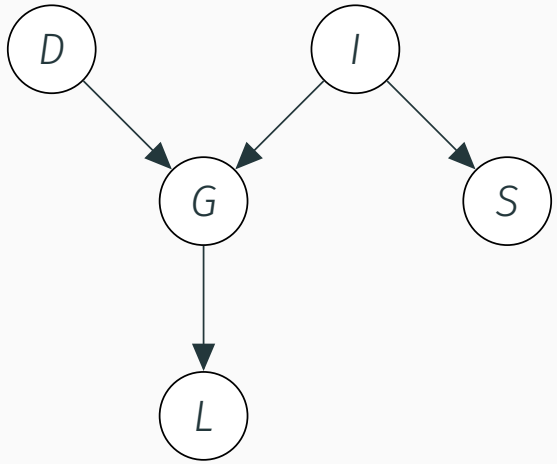
Can knowledge about X influence our beliefs about Y?

Direct influence:

- $X \rightarrow Y$ e.g., $D \rightarrow G$ ✓ causal effect
- $X \leftarrow Y$ e.g., $G \leftarrow D$ ✓ evidential effect

Indirect influence:

- $X \rightarrow W \rightarrow Y$ e.g., $D \rightarrow G \rightarrow L$ ✓
  indirect causal effect
- $X \leftarrow W \leftarrow Y$ e.g., $L \leftarrow G \leftarrow D$ ✓
  indirect evidential effect
- $X \leftarrow W \rightarrow Y$ e.g., $G \leftarrow I \rightarrow S$ ✓
  X and Y are joint effects of a common cause W



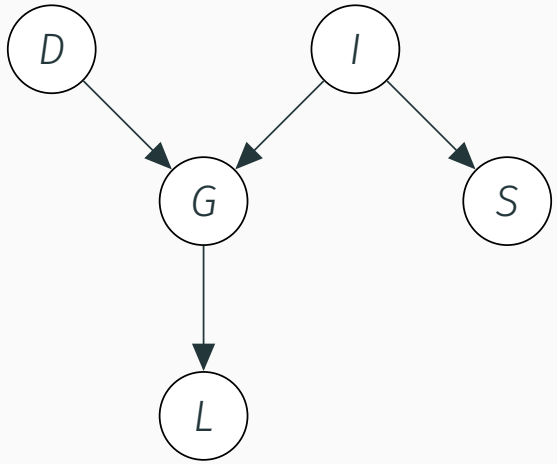
When can X influence Y?

Can knowledge about X influence our beliefs about Y?

Direct influence:

- $X \to Y$ e.g., $D \to G$ ✓ causal effect
- $X \leftarrow Y$ e.g., $G \leftarrow D$ ✓ evidential effect

Indirect influence:



- $X \to W \to Y$ e.g., $D \to G \to L$ ✓
  indirect causal effect
- $X \leftarrow W \leftarrow Y$ e.g., $L \leftarrow G \leftarrow D$ ✓
  indirect evidential effect
- $X \leftarrow W \to Y$ e.g., $G \leftarrow I \to S$ ✓
  X and Y are joint effects of a common cause W
- $X \to W \leftarrow Y$ e.g., $D \to G \leftarrow I$

When can X influence Y?

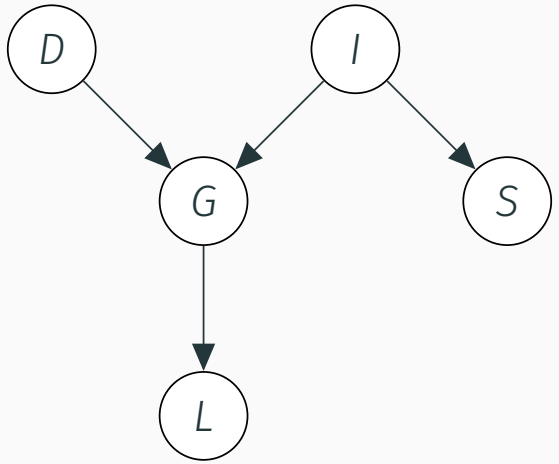
Can knowledge about X influence our beliefs about Y?

Direct influence:

- $X \to Y$ e.g., $D \to G$ ✓ causal effect
- $X \leftarrow Y$ e.g., $G \leftarrow D$ ✓ evidential effect

Indirect influence:

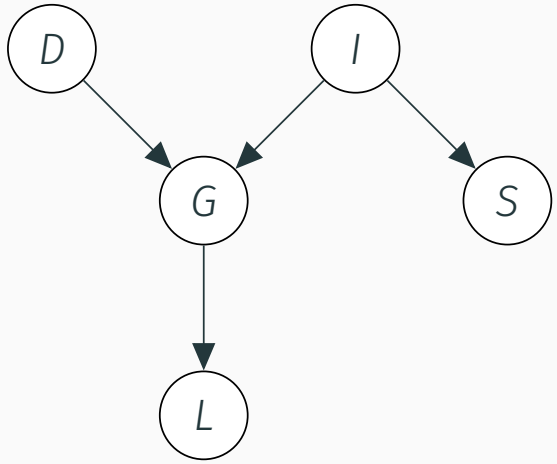
- $X \to W \to Y$ e.g., $D \to G \to L$ ✓
  indirect causal effect
- $X \leftarrow W \leftarrow Y$ e.g., $L \leftarrow G \leftarrow D$ ✓
  indirect evidential effect
- $X \leftarrow W \to Y$ e.g., $G \leftarrow I \to S$ ✓
  X and Y are joint effects of a common cause W
- $X \to W \leftarrow Y$ e.g., $D \to G \leftarrow I$ ✗
  X and Y are joint causes of a common effect W



v-structure

Can X influence Y *given* evidence about $\mathbf{Z}$?

Let us see if anything changes when we also have knowledge about an additional set of variables $\mathbf{Z}$.

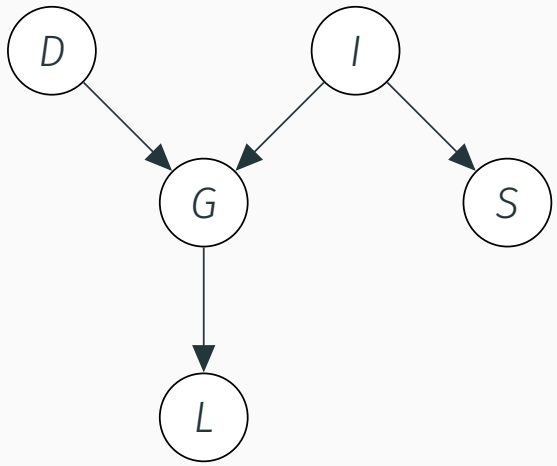


Can X influence Y *given* evidence about Z?

Let us see if anything changes when we also have knowledge about an additional set of variables **Z**.

Direct influence: unaffected by other evidence

- $X \rightarrow Y$ e.g., $D \rightarrow G$ ✓
- $X \leftarrow Y$ e.g., $G \leftarrow D$ ✓

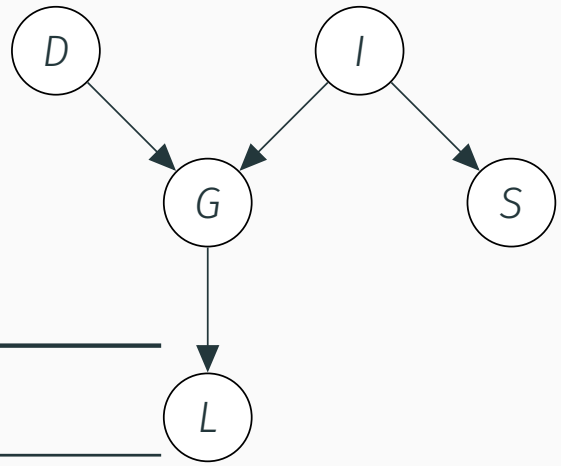


Can X influence Y *given* evidence about $\mathbf{Z}$?

Let us see if anything changes when we also have knowledge about an additional set of variables $\mathbf{Z}$.

Direct influence: unaffected by other evidence

- $X \to Y$ e.g., $D \to G$ ✓
- $X \leftarrow Y$ e.g., $G \leftarrow D$ ✓



Indirect influence: depends on W and $\mathbf{Z}$

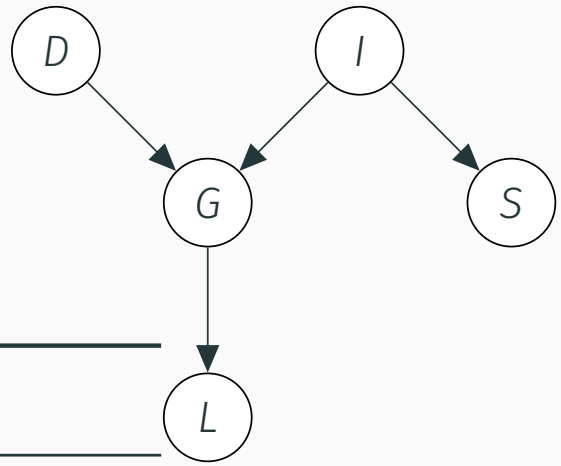
pattern	example	$W \in \mathbf{Z}$	$W \notin \mathbf{Z}$
$X \to W \to Y$	$D \to G \to L$		
$X \leftarrow W \leftarrow Y$	$L \leftarrow G \leftarrow D$		
$X \leftarrow W \to Y$	$G \leftarrow I \to S$		
$X \to W \leftarrow Y$	$D \to G \leftarrow I$		

Can X influence Y *given* evidence about $\mathbf{Z}$?

Let us see if anything changes when we also have knowledge about an additional set of variables $\mathbf{Z}$.

Direct influence: unaffected by other evidence

- $X \to Y$ e.g., $D \to G$ ✓
- $X \leftarrow Y$ e.g., $G \leftarrow D$ ✓



Indirect influence: depends on W and $\mathbf{Z}$

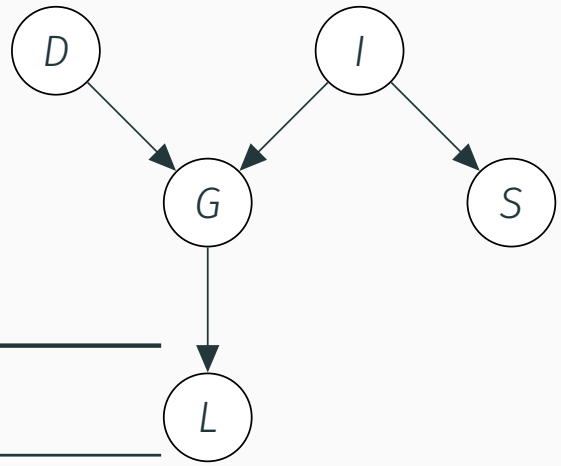
pattern	example	$W \in \mathbf{Z}$	$W \notin \mathbf{Z}$
$X \to W \to Y$	$D \to G \to L$	✗	✓
$X \leftarrow W \leftarrow Y$	$L \leftarrow G \leftarrow D$		
$X \leftarrow W \to Y$	$G \leftarrow I \to S$		
$X \to W \leftarrow Y$	$D \to G \leftarrow I$		

Can X influence Y *given* evidence about $\mathbf{Z}$?

Let us see if anything changes when we also have knowledge about an additional set of variables $\mathbf{Z}$.

Direct influence: unaffected by other evidence

- $X \to Y$ e.g., $D \to G$ ✓
- $X \leftarrow Y$ e.g., $G \leftarrow D$ ✓



Indirect influence: depends on W and $\mathbf{Z}$

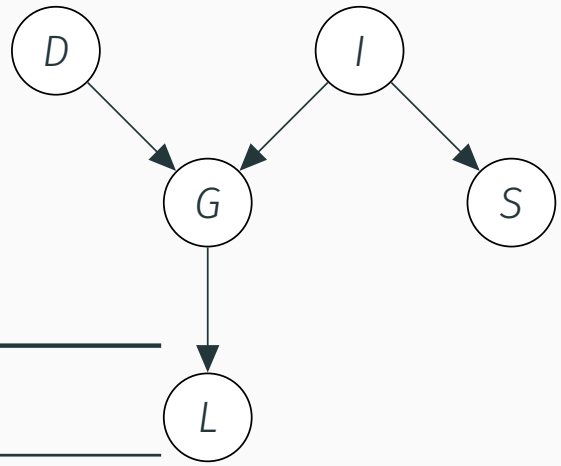
pattern	example	$W \in \mathbf{Z}$	$W \notin \mathbf{Z}$
$X \to W \to Y$	$D \to G \to L$	✗	✓
$X \leftarrow W \leftarrow Y$	$L \leftarrow G \leftarrow D$	✗	✓
$X \leftarrow W \to Y$	$G \leftarrow I \to S$		
$X \to W \leftarrow Y$	$D \to G \leftarrow I$		

Can X influence Y *given* evidence about $\mathbf{Z}$?

Let us see if anything changes when we also have knowledge about an additional set of variables $\mathbf{Z}$.

Direct influence: unaffected by other evidence

- $X \to Y$ e.g., $D \to G$ ✓
- $X \leftarrow Y$ e.g., $G \leftarrow D$ ✓



Indirect influence: depends on W and $\mathbf{Z}$

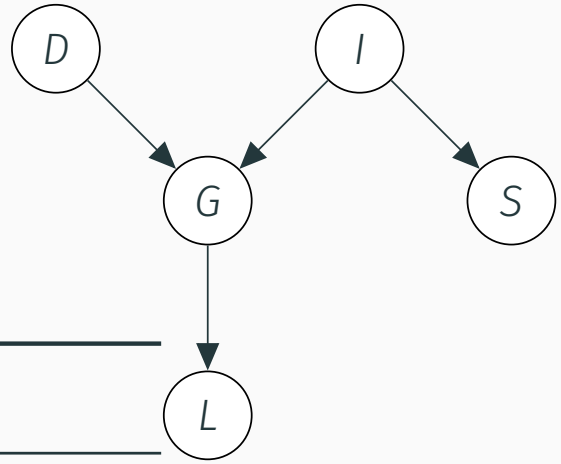
pattern	example	$W \in \mathbf{Z}$	$W \notin \mathbf{Z}$
$X \to W \to Y$	$D \to G \to L$	✗	✓
$X \leftarrow W \leftarrow Y$	$L \leftarrow G \leftarrow D$	✗	✓
$X \leftarrow W \to Y$	$G \leftarrow I \to S$	✗	✓
$X \to W \leftarrow Y$	$D \to G \leftarrow I$		

Can X influence Y *given* evidence about $\mathbf{Z}$?

Let us see if anything changes when we also have knowledge about an additional set of variables $\mathbf{Z}$.

Direct influence: unaffected by other evidence

- $X \to Y$ e.g., $D \to G$ ✓
- $X \leftarrow Y$ e.g., $G \leftarrow D$ ✓



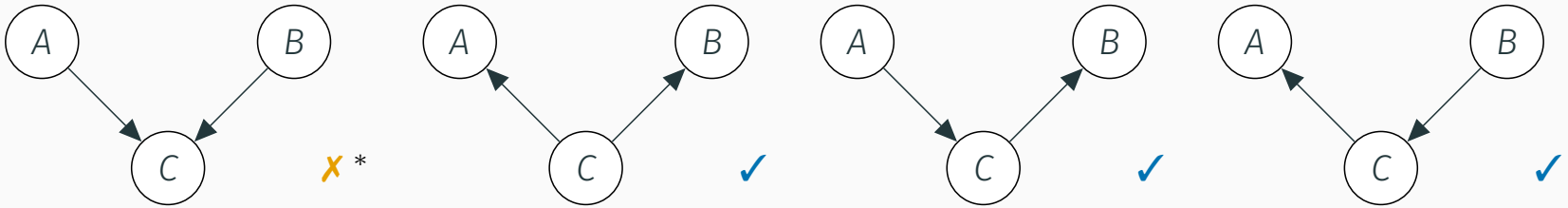
Indirect influence: depends on W and $\mathbf{Z}$

pattern	example	$W \in \mathbf{Z}$	$W \notin \mathbf{Z}$
$X \to W \to Y$	$D \to G \to L$	✗	✓
$X \leftarrow W \leftarrow Y$	$L \leftarrow G \leftarrow D$	✗	✓
$X \leftarrow W \to Y$	$G \leftarrow I \to S$	✗	✓
$X \to W \leftarrow Y$	$D \to G \leftarrow I$	✓	✗ e.g., with $Z = \{S\}$ ✓ e.g., with $Z = \{L\}$

The v-structure enables indirect influence of X on Y through W when $\mathbf{Z}$ contains W or any of its descendants.

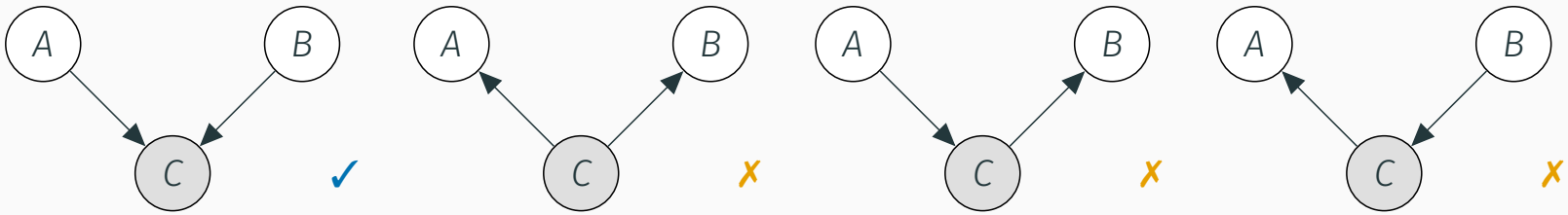
Colliders

C is a **collider** in the v-structure (left) and a **non-collider** in the rest.



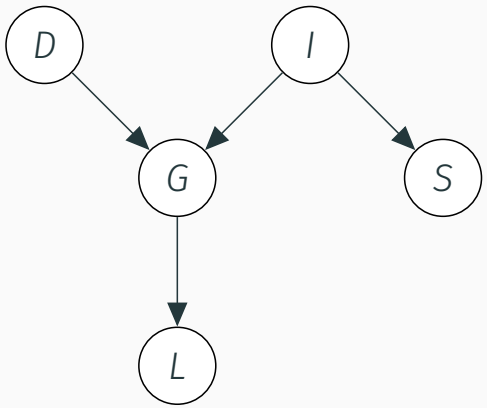
When *C* is a collider, observing it or any of its descendants **activates** the flow of influence between *A* and *B*.

When *C* is a non-collider, observing it **blocks** the flow of influence between *A* and *B*.



Trails

A **trail** (definition 2.16) is a sequence of nodes $X_1, \ldots, X_k$ connected by edges, ignoring the direction of the edges.



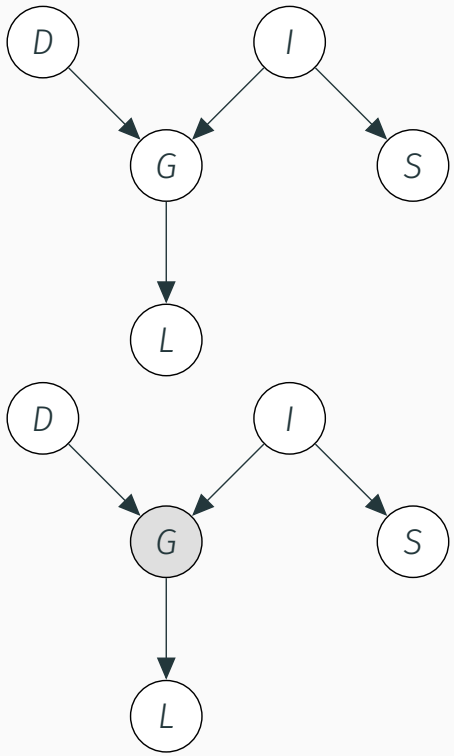
Examples

- (I, G, L)
- (D, G, L)
- (D, G, I)
- (D, G, I, S)

Active Trails (definition 3.6)

A trail is **active** if

- for any v-structure, the collider or any of its descendants is observed
- no other node along the trail is observed
  (which implies that non-colliders are not observed)



Examples

- (I, G, L) ✓
- (D, G, L) ✓
- (D, G, I) ✗
- (D, G, I, S) ✗

Examples

- (I, G, L) ✗
- (D, G, L) ✗
- (D, G, I) ✓
- (D, G, I, S) ✓

Independence

A joint distribution P **satisfies independence between** the random variables X and Y, which we denote by $P \models A \perp B$, if:

- $P(X, Y) = P(X)P(Y)$
- $P(X|Y) = P(X)$
- $P(Y|X) = P(Y)$

Independence

A joint distribution P **satisfies independence between** the random variables X and Y, which we denote by $P \models A \perp B$, if:

- $P(X, Y) = P(X)P(Y)$
- $P(X|Y) = P(X)$
- $P(Y|X) = P(Y)$

A joint distribution P **satisfies independence** of the rvs X and Y *given* Z, which we denote by $P \models (X \perp Y \mid Z)$, if:

- $P(X, Y|Z) = P(X|Z)P(Y|Z)$
- $P(X|Y, Z) = P(X|Z)$
- $P(Y|X, Z) = P(Y|Z)$

These definitions are equivalent, if one is met they are all met.

Testing independence when we know all relevant factors

I	D	$P(D\|I)$
i^0	d^0	0.6
i^0	d^1	0.4
i^1	d^0	0.6
i^1	d^1	0.4

Does $P \models D \perp I$ hold?

Testing independence when we know all relevant factors

I	D	$P(D\|I)$
i^0	d^0	0.6
i^0	d^1	0.4
i^1	d^0	0.6
i^1	d^1	0.4

Does $P \models D \perp I$ hold?

G	I	D	$P(D\|I,G)$
g^1	i^0	d^0	0.9
g^1	i^0	d^1	0.1
g^1	i^1	d^0	0.7297
g^1	i^1	d^1	0.2703
g^2	i^0	d^0	0.7059
g^2	i^0	d^1	0.2941
g^2	i^1	d^0	0.2857
g^2	i^1	d^1	0.7143
g^3	i^0	d^0	0.3913
g^3	i^0	d^1	0.6087
g^3	i^1	d^0	0.1304
g^3	i^1	d^1	0.8696

Does $P \models (D \perp I \mid G)$ hold?

Testing independence when we know all relevant factors

I	D	$P(D\|I)$
i^0	d^0	0.6
i^0	d^1	0.4
i^1	d^0	0.6
i^1	d^1	0.4

Does $P \models D \perp I$ hold?

What if we do not know these factors in tabular form, but we have a tabular view of the joint distribution $P(D, I, G, S, L)$, as shown in Table 1?

G	I	D	$P(D\|I, G)$
g^1	i^0	d^0	0.9
g^1	i^0	d^1	0.1
g^1	i^1	d^0	0.7297
g^1	i^1	d^1	0.2703
g^2	i^0	d^0	0.7059
g^2	i^0	d^1	0.2941
g^2	i^1	d^0	0.2857
g^2	i^1	d^1	0.7143
g^3	i^0	d^0	0.3913
g^3	i^0	d^1	0.6087
g^3	i^1	d^0	0.1304
g^3	i^1	d^1	0.8696

Does $P \models (D \perp I \mid G)$ hold?

Watch Out!

Whichever definition we pick, it's necessary to **test all relevant assignments**. Example:

A	B	$P(B\|A)$
a^1	b^0	0.7
a^1	b^1	0.3
a^2	b^0	0.7
a^2	b^1	0.3
a^3	b^0	0.4
a^3	b^1	0.6

Does $P \models B \perp A$ hold?

Watch Out!

Whichever definition we pick, it's necessary to **test all relevant assignments**. Example:

A	B	$P(B\|A)$
a^1	b^0	0.7
a^1	b^1	0.3
a^2	b^0	0.7
a^2	b^1	0.3
a^3	b^0	0.4
a^3	b^1	0.6

Does $P \models B \perp A$ hold?

Finding that $P_{B|A=a^1}$ and $P_{B|A=g^2}$ are the same is not enough to ascertain independence.

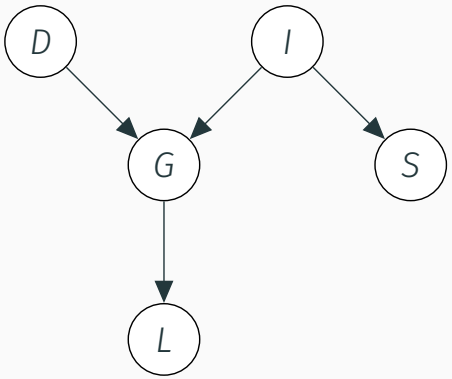
The same applies when testing *conditional* independence.

BNs and Independence

A BN structure $\mathcal{G}$ encodes a set of (conditional) independences.

The more elementary ones are readily recognisable from the graph structure, they are called the *local independencies* of $\mathcal{G}$, denoted $\mathcal{I}_l(\mathcal{G})$.

$\mathcal{I}_l(\mathcal{G})$ is the set of statements of the kind $X_i \perp \text{NonDesc}_{\mathcal{G}}(X_i) \mid \text{Pa}_{\mathcal{G}}(X_i)$



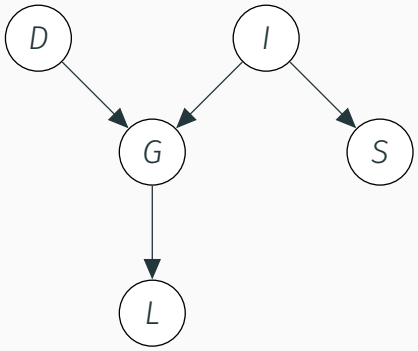
For the Student BN, $\mathcal{I}_l(\mathcal{G}) =$

- $D \perp I, S$
- $I \perp D$
- $S \perp D, G, L \mid I$
- $G \perp S \mid D, I$
- $L \perp D, I, S \mid G$

But how can we ascertain all conditional independences implied by $\mathcal{G}$? For example, if P factorises over $\mathcal{G}$, does that imply that $P \models (D \perp S \mid L)$?

BNs and Independence

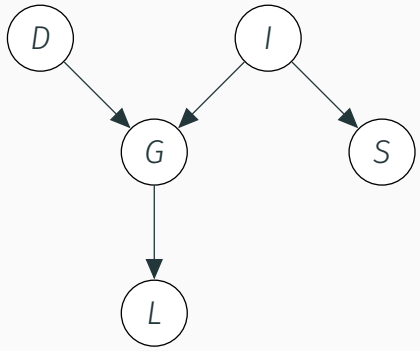
We know that P factorises over the BN $\mathcal{G}$ and we want to determine whether or not $P \models (\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z})$, for any subset of rvs $\mathbf{X}, \mathbf{Y}$ and $\mathbf{Z}$ of the joint distribution P.



If P factorises over the Student BN, does that imply that $P \models (D \perp S \mid L)$?

BNs and Independence

We know that P factorises over the BN $\mathcal{G}$ and we want to determine whether or not $P \models (\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z})$, for any subset of rvs $\mathbf{X}, \mathbf{Y}$ and $\mathbf{Z}$ of the joint distribution P.

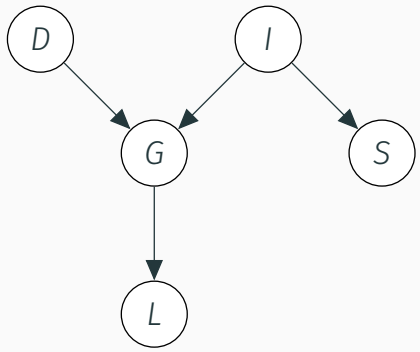


If P factorises over the Student BN, does that imply that $P \models (D \perp S \mid L)$?

First idea. Use prob. calculus and test by checking the relevant distributions against any of the definitions of independence.

BNs and Independence

We know that P factorises over the BN $\mathcal{G}$ and we want to determine whether or not $P \models (\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z})$, for any subset of rvs $\mathbf{X}, \mathbf{Y}$ and $\mathbf{Z}$ of the joint distribution P.



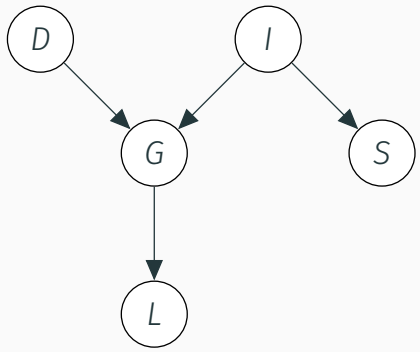
If P factorises over the Student BN, does that imply that $P \models (D \perp S \mid L)$?

First idea. Use prob. calculus and test by checking the relevant distributions against any of the definitions of independence.

Second idea. Use *flow of influence* (namely, active trails) to test for independence by analysis of the BN structure alone.

BNs and Independence

We know that P factorises over the BN $\mathcal{G}$ and we want to determine whether or not $P \models (\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z})$, for any subset of rvs $\mathbf{X}, \mathbf{Y}$ and $\mathbf{Z}$ of the joint distribution P.



If P factorises over the Student BN, does that imply that $P \models (D \perp S \mid L)$?

First idea. Use prob. calculus and test by checking the relevant distributions against any of the definitions of independence.

Second idea. Use *flow of influence* (namely, active trails) to test for independence by analysis of the BN structure alone. In the example, observing L *activates* the v-structure so $P \not\models (D \perp S \mid L)$.

Modifying the Student Example with Additional Variables

We further assume that the student gets a **J**ob (yes j^1, or no j^0) depending on S and G, and we regard the student's own **H**appiness (happy h^1, or unhappy h^0) as dependant on G and J.

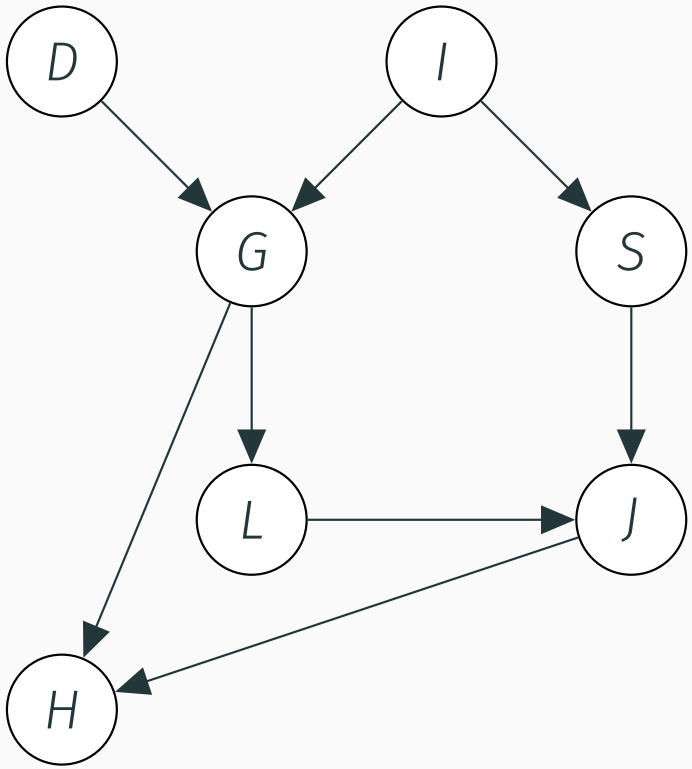


Figure 3: The *Extended* Student Example

Modifying the Student Example with Additional Variables

We further assume that the student gets a **J**ob (yes j^1, or no j^0) depending on S and G, and we regard the student's own **H**appiness (happy h^1, or unhappy h^0) as dependant on G and J.

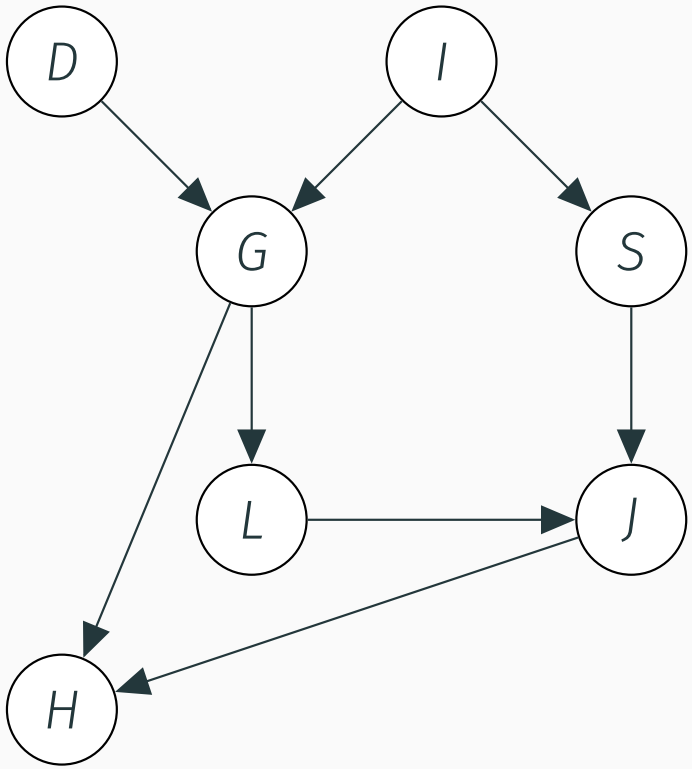


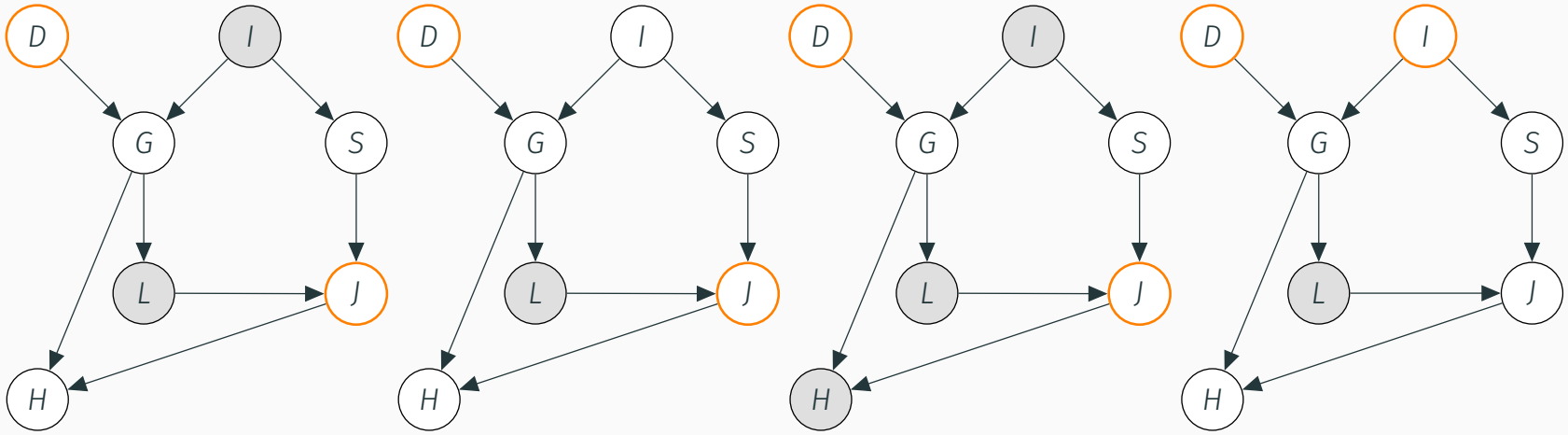
Figure 3: The *Extended* Student Example

At this point, we do not have CPDs for this DAG. Yet, we can reason about independencies.

Directed Separation (section 3.3.1, definition 3.7)

We say that $\mathbf{X}$ and $\mathbf{Y}$ are d-separated given $\mathbf{Z}$, denoted $\textbf{d-sep}_{\mathcal{G}}(\mathbf{X}; \mathbf{Y} \mid \mathbf{Z})$, **if there is no active trail** between any node $X \in \mathbf{X}$ and $Y \in \mathbf{Y}$ given $\mathbf{Z}$.

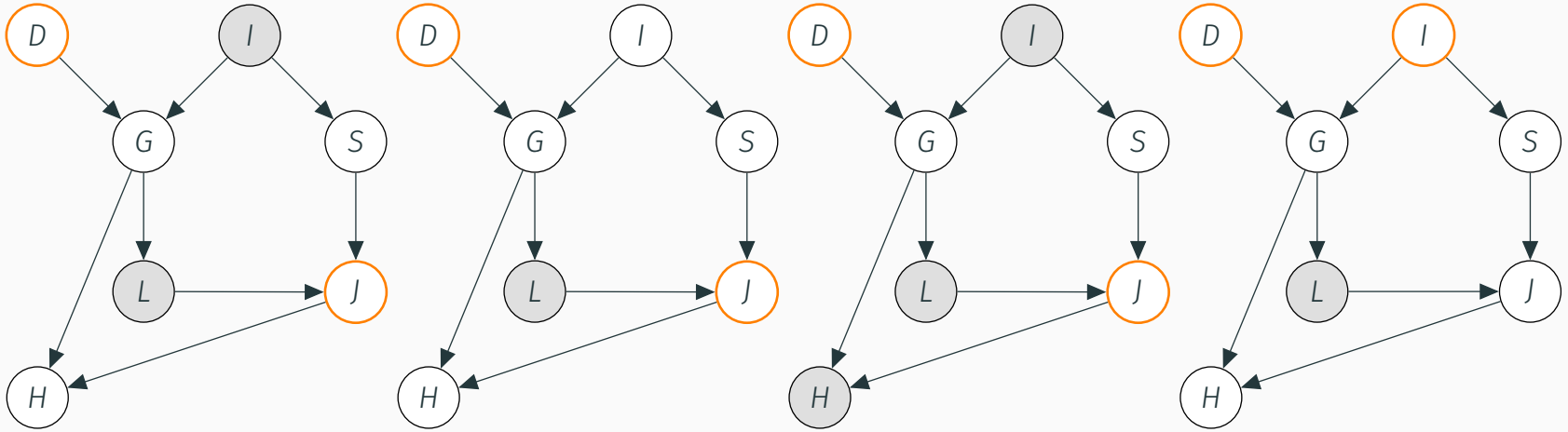
Examples:



Directed Separation (section 3.3.1, definition 3.7)

We say that $\mathbf{X}$ and $\mathbf{Y}$ are d-separated given $\mathbf{Z}$, denoted $\text{d-sep}_{\mathcal{G}}(\mathbf{X}; \mathbf{Y} \mid \mathbf{Z})$, **if there is no active trail** between any node $X \in \mathbf{X}$ and $Y \in \mathbf{Y}$ given $\mathbf{Z}$.

Examples:

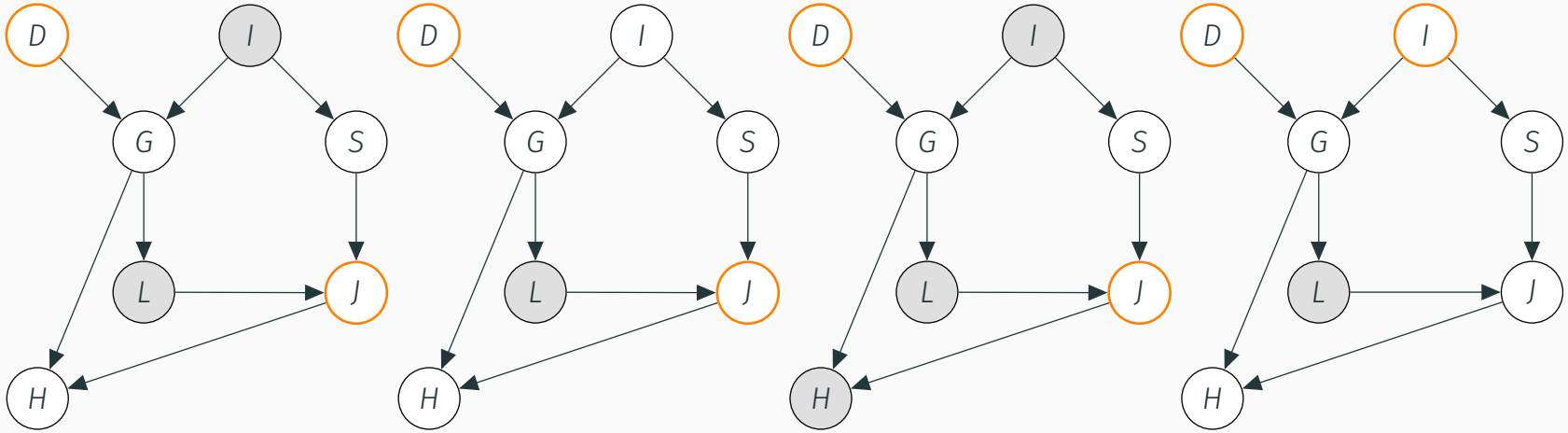


d-sep$(D; J|L, I)$ ✓ d-sep$(D; J|L)$ ✗ d-sep$(D; J|I, L, H)$ ✗ d-sep$(D; I|L)$ ✗

Directed Separation (section 3.3.1, definition 3.7)

We say that $\mathbf{X}$ and $\mathbf{Y}$ are d-separated given $\mathbf{Z}$, denoted $\text{d-sep}_{\mathcal{G}}(\mathbf{X}; \mathbf{Y} \mid \mathbf{Z})$, **if there is no active trail** between any node $X \in \mathbf{X}$ and $Y \in \mathbf{Y}$ given $\mathbf{Z}$.

Examples:



$\text{d-sep}(D; J|L, I)$ ✓ $\text{d-sep}(D; J|L)$ ✗ $\text{d-sep}(D; J|I, L, H)$ ✗ $\text{d-sep}(D; I|L)$ ✗

D-separation in $\mathcal{G}$ implies conditional independence in *any* P that factorises over $\mathcal{G}$.

I know how to test independence, why learn about d-sep?

I know how to test independence, why learn about d-sep?

D-separation in $\mathcal{G}$ gives us a complete view of all the independencies that hold for *any* distribution that factorises over $\mathcal{G}$.

This has impacts on

- inference, e.g. telling us what computations we need to perform or may safely skip when we need marginals and conditionals
- learning, e.g. telling us what parameters are necessary or superfluous

Probability Calculus or D-Separation?

Both are provably correct, so you can decide based on convenience.

- Probability calculus can be tedious and we may need access to the actual probability factors.
- D-separation uses the graph structure alone, the actual probability factors are not needed.
- Enumerating trails and testing for activation can be tedious, but graph algorithms are systematic and can be rather efficient.

Summary

- Influence may flow through trails regardless of directionality, but it is blocked by the v-structure, unless the collider or any of its descendants is observed, and it is blocked by observation of a non-collider.
- With an explicit representation of P, we can ascertain any independence via probability calculus (i.e., we infer the relevant factors from the joint distribution and test independence).
- Without requiring an explicit representation of P, and simply knowing that it factorises over the BN structure, we can ascertain any independence via d-separation.

What's Next?

WC1: exercises (semantics, reasoning and influence).

Friday: P1 deadline.

Module 2: Markov Networks.

References

[1] Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.