Bayesian Networks

Representation

Professor Ajoodha

Lecture 2

School of Computer Science and Applied Mathematics
The University of the Witwatersrand, Johannesburg

ExplainableAI Lab
— MODELLING. DECISION MAKING. CAUSALITY —

Problem Statement

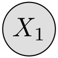
Problem

Represent a joint distribution (P) over a set of random variables $\mathcal{X} = \{X_1, \ldots, X_n\}$*. i.e.* $P(X_1, \ldots, X_n)$.

1. We can do this using the **chain rule** for probabilities.
2. The Chain Rule:

$$P(X_1, \ldots, X_n) = P(X_1)P(X_2|X_1)\ldots P(X_n|X_1, X_{n-1})$$

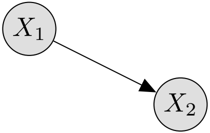
Problem Statement - $P(X_1, \dots, X_n)$



$$P(X_1, \dots, X_n) = P(X_1) \dots$$

Number of independent parameters: $2^1 - 1 = \boxed{1}$.

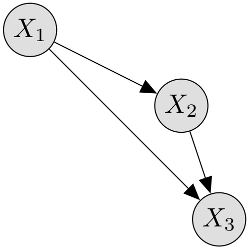
Problem Statement - $P(X_1, \ldots, X_n)$



$$P(X_1, \ldots, X_n) = P(X_1)P(X_2|X_1)\ldots$$

Number of independent parameters: $2^2 - 1 = 3$.

Problem Statement - $P(X_1, \ldots, X_n)$

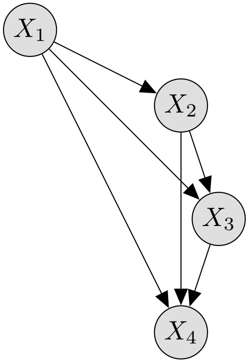


$$P(X_1, \ldots, X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\ldots$$

Number of independent parameters: $2^3 - 1 = \boxed{7}$.

Problem Statement - $P(X_1, \ldots, X_n)$

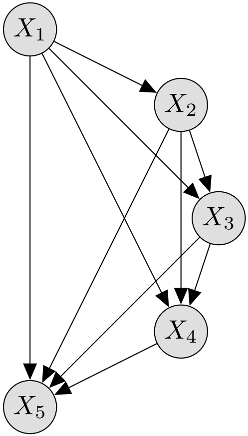
Bayesian Networks
Professor Ajoodha
Problem Statement
Bayesian Network
Independence
D-Separation
I-Equivalence
I-Maps
Min I-Maps
P-Maps



$$P(X_1, \ldots, X_n) = \ldots P(X_4 | X_1, X_2, X_3) \ldots$$

Number of independent parameters: $2^4 - 1 = 15$.

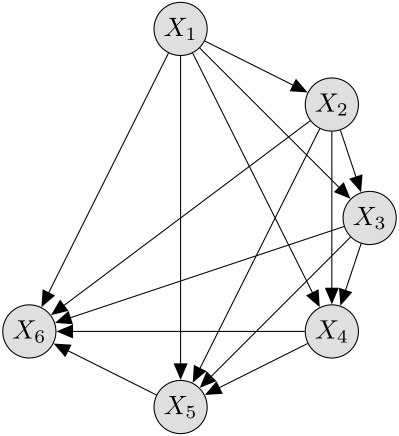
Problem Statement - $P(X_1, \ldots, X_n)$



$$P(X_1, \ldots, X_n) = \ldots P(X_5 | X_1, X_2, X_3, X_4) \ldots$$

`Number of independent parameters:` $2^5 - 1 = 31$.

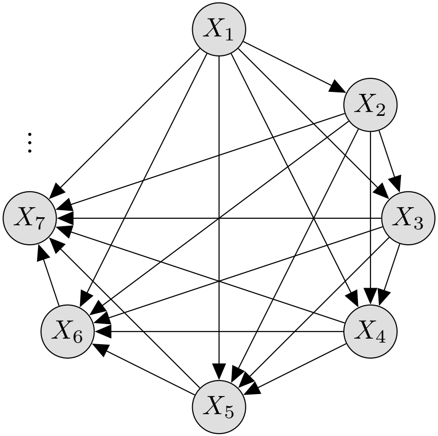
Problem Statement - $P(X_1, \ldots, X_n)$



$$P(X_1, \ldots, X_n) = \ldots P(X_6 | X_1, X_2, X_3, X_4, X_5) \ldots$$

Number of independent parameters: $2^6 - 1 = 63$.

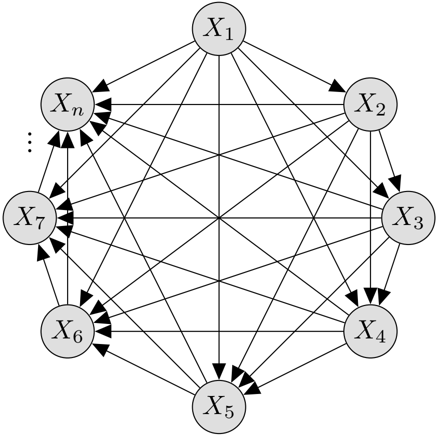
Problem Statement - $P(X_1, \dots, X_n)$



$$P(X_1, \dots, X_n) = \dots P(X_7 | X_1, X_2, X_3, X_4, X_5, X_6) \dots$$

Number of independent parameters: $2^7 - 1 = 127$.

Problem Statement - $P(X_1, \ldots, X_n)$



$$P(X_1, \ldots, X_n) = \ldots P(X_n | X_1, \ldots, X_{n-1}).$$

`Number of independent parameters:` $2^n - 1$

Problem Statement

How much is that? Why is it a problem? How many probabilities do we need to represent?

10^1 ≈ 15 **string quartets** by Beethoven

10^{10} ≈ 5.610^{10} web pages indexed by Google

10^{11} ≈ Stars in our Galaxy

10^{16} ≈ **Ants on Earth**

10^{21} ≈ $6,670,903,752,021,072,936,960$ Sudoku Grids

10^{21} ≈ Stars in the observable universe.

10^{30} ≈ The number of **bacterial cells** on Earth

10^{46} ≈ Number of legal chess positions

10^{50} ≈ Estimated number of **atoms in Earth**.

10^{63} ≈ Grains fit into Cosmos

10^{170} ≈ legal **Go positions**.

Problem Statement

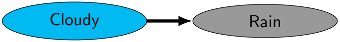
Representing the Joint distribution this way is unmanageable from every perspective.

1. Expensive to manipulate.
2. Too large to store in memory.
3. Impossible to acquire so many numbers from a human expert.
4. Values are too small for a human to reasonably contemplate.
5. Large amount of data required to make a robust estimation.

   These barriers kept probabilistic models from the spotlight for many years! (Its hard to say exactly when since progress is gradual.)

Two Simplifying Assumptions

1. The representation of independence properties of the distribution.

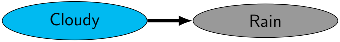


instead of $P(C,R)$ = we do this $P(C)P(R|C)$

2. The use of alternate parametrization that allows us to exploit finer-grained independencies.

Prior of Cloudy

Cloudy	
Yes	No
0.8	0.2



Conditional Probability distribution (CPD) of Rain given Cloudy

	Rain	
Cloudy	Yes	No
Yes	0.7	0.3
No	0.05	0.95

The Naïve Bayes model

Naïve Bayes, or idiot Bayes, is the simplest example of using

1. conditional **independence assumptions** and
2. a **conditional parametrization**

to model a high-dimensional probability distribution.

Assumptions:

1. Instances fall into a number of **mutually exclusive** and **exhaustive** classes.
2. The instances are **conditionally independent given the class** of the instance. $(X_i \perp \mathbf{X}_{-i} \mid C)$.

$$P(C, X_1, \ldots, X_n) = P(C)P(X_1|C)\ldots P(X_n \mid C).$$

$$P(C, X_1, \ldots, X_n) = P(C)\prod_{i=0}^{n} P(X_i \mid C).$$

What is a Bayesian Network?

A Bayesian Network models a distribution in the form of a graph ($\mathcal{G}$), where

1. $\mathcal{G}$ is a data structure that provides the skeleton for **representing a joint distribution** compactly in a factorised way.
2. $\mathcal{G}$ is a compact set of **conditional independence assumptions** about the joint distribution.

The Student Example

A The Student Example is a popular Bayesian network used to model student performance at a university.

Suppose you have the following variables:

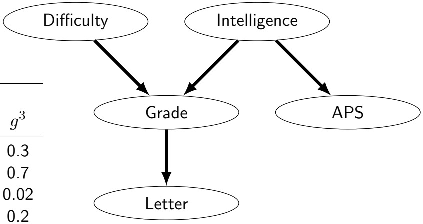
1. **Difficulty:** An indication of how hard the course is.
2. **Intelligence:** The intelligence of the student.
3. **Grade:** The grade the student gets.
4. **Letter:** The recommendation letter from the students professor.
5. **APS:** The Admission Points Score (APS) of the student.

Can you draw a simple Bayesian network to indicate the influence of these variables? Using arrows to indicate influence.

Here is **one way** to do it (with associated probabilities):

Difficulty	
d^0	d^1
0.6	0.4

Intelligence	
i^0	i^1
0.7	0.3



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

	APS	
i	a^0	a^1
i^0	0.95	0.05
i^1	0.2	0.8

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

Conditional Independence Assumptions

1) (L $\perp I, D, A \mid G$)
2) (A $\perp D, G, L \mid I$)
3) (G $\perp A \mid I, D$)
4) (D $\perp I, A$)

No. Ind Para = 15

Full joint Para = 47

Properties of the Student Example

Note that in the joint distribution there would be **48 parameters**. However, we only need **26 parameters** for this representation.

Local probability model:

1. $P(D)$: Prior over difficulty.
2. $P(I)$: Prior over intelligence.
3. $P(G|D,I)$: Probability of grade **given** difficulty and intelligence.
4. $P(L|G)$: Probability of the recommendation letter **given** the grade.
5. $P(A|I)$: Probability of the APS **given** the intelligence.

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

Can you calculate this $P(i^1, d^0, g^2, a^1, l^0)$?

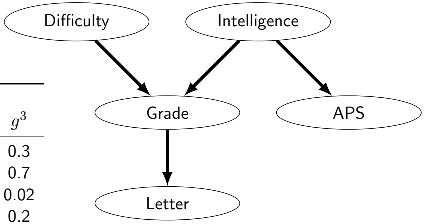
$= P(i^1)P(d^0)P(g^2|d^0,i^1)P(a^1|i^1)P(l^0|g^2)$

The Student Example

$$P(i^1, d^0, g^2, a^1, l^0) = \underset{\mathbf{0.3}}{P(i^1)} \times \underset{\mathbf{0.6}}{P(d^0)} \times \underset{\mathbf{0.08}}{P(g^2|d^0, i^1)} \times \underset{\mathbf{0.8}}{P(a^1|i^1)} \times \underset{\mathbf{0.4}}{P(l^0|g^2)} = \mathbf{0.004608}$$

Difficulty	
d^0	d^1
0.6	0.4

Intelligence	
i^0	i^1
0.7	0.3



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

	APS	
i	a^0	a^1
i^0	0.95	0.05
i^1	0.2	0.8

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

Reasoning Patterns

Suppose we would like to reason about a WITS student named Xolani.

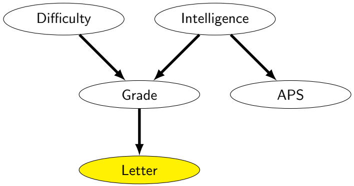
We may want to also compute queries like $P(\mathbf{Y} = \mathbf{y} | \mathbf{E} = \mathbf{e})$ about Xolani.

There are three ways we can reason about Xolani:

1. **Causal Inference:** exploring the relationship between a cause and its effect.
2. **Evidential Inference:** Manipulating and reasoning from evidential information.
3. **Intercausal Inference:** probabilistic dependence of causes of an observed common effect causes of the same effect can interact.

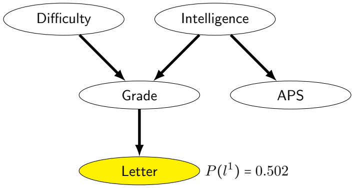
Causal Inference

1. What is the probability that Xolani gets a **good letter** from his professor?
2. Now suppose that we find out that Xolani is **not intelligent**.
3. Now suppose that we find out that the **PGM course is easy**.



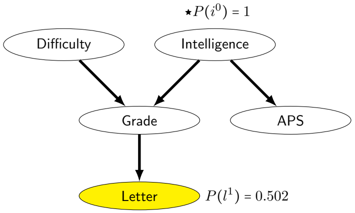
Causal Inference

1. What is the probability that Xolani gets a **good letter** from his professor?
2. Now suppose that we find out that Xolani is **not intelligent**.
3. Now suppose that we find out that the **PGM course is easy**.



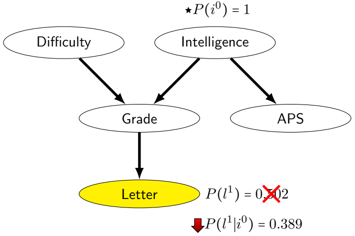
Causal Inference

1. What is the probability that Xolani gets a **good letter** from his professor?
2. Now suppose that we find out that Xolani is **not intelligent**.
3. Now suppose that we find out that the **PGM course is easy**.



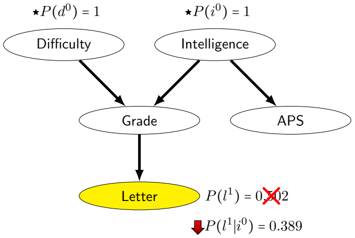
Causal Inference

1. What is the probability that Xolani gets a **good letter** from his professor?
2. Now suppose that we find out that Xolani is **not intelligent**.
3. Now suppose that we find out that the **PGM course is easy**.



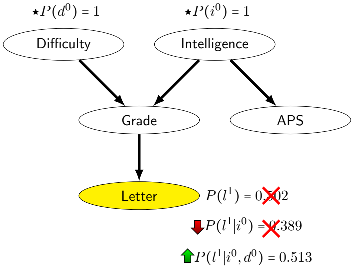
Causal Inference

1. What is the probability that Xolani gets a **good letter** from his professor?
2. Now suppose that we find out that Xolani is **not intelligent**.
3. Now suppose that we find out that the **PGM course is easy**.



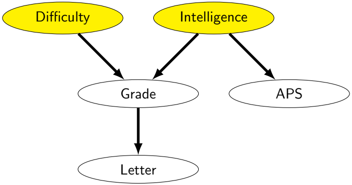
Causal Inference

1. What is the probability that Xolani gets a **good letter** from his professor?
2. Now suppose that we find out that Xolani is **not intelligent**.
3. Now suppose that we find out that the **PGM course is easy**.



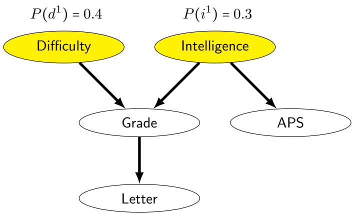
Evidential Inference (Explaining)

1. Now Suppose that we would like to consider hiring Xolani.
2. We calculate the probability that Xolani **is intelligent**.
3. We find Xolani's **grade** for the PGM course.
4. We can now **re-calculate Xolani's intelligence** (w.r.t. the grade).
   - Xolani's intelligence decreases
   - The PGM course difficulty increases
5. If we **didn't have the grade**, but had a weak recommendation letter.



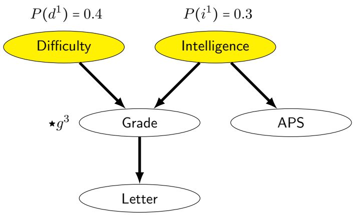
Evidential Inference (Explaining)

1. Now Suppose that we would like to consider hiring Xolani.
2. We calculate the probability that Xolani **is intelligent**.
3. We find Xolani's **grade** for the PGM course.
4. We can now **re-calculate Xolani's intelligence** (w.r.t. the grade).
   - Xolani's intelligence decreases
   - The PGM course difficulty increases
5. If we **didn't have the grade**, but had a weak recommendation letter.



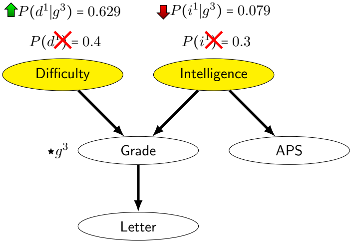
Evidential Inference (Explaining)

1. Now Suppose that we would like to consider hiring Xolani.
2. We calculate the probability that Xolani **is intelligent**.
3. We find Xolani's **grade** for the PGM course.
4. We can now **re-calculate Xolani's intelligence** (w.r.t. the grade).
    - Xolani's intelligence decreases
    - The PGM course difficulty increases
5. If we **didn't have the grade**, but had a weak recommendation letter.



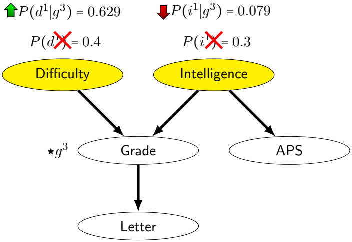
Evidential Inference (Explaining)

1. Now Suppose that we would like to consider hiring Xolani.
2. We calculate the probability that Xolani **is intelligent**.
3. We find Xolani's **grade** for the PGM course.
4. We can now **re-calculate Xolani's intelligence** (w.r.t. the grade).
   - Xolani's intelligence decreases
   - The PGM course difficulty increases
5. If we **didn't have the grade**, but had a weak recommendation letter.



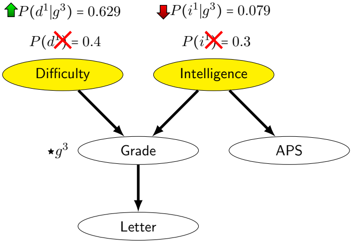
Evidential Inference (Explaining)

1. Now Suppose that we would like to consider hiring Xolani.
2. We calculate the probability that Xolani **is intelligent**.
3. We find Xolani's **grade** for the PGM course.
4. We can now **re-calculate Xolani's intelligence** (w.r.t. the grade).
    - Xolani's intelligence decreases
    - The PGM course difficulty increases
5. If we **didn't have the grade**, but had a weak recommendation letter.



Evidential Inference (Explaining)

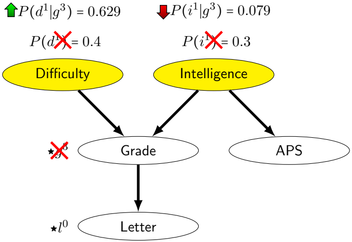
1. Now Suppose that we would like to consider hiring Xolani.
2. We calculate the probability that Xolani **is intelligent**.
3. We find Xolani's **grade** for the PGM course.
4. We can now **re-calculate Xolani's intelligence** (w.r.t. the grade).
   - Xolani's intelligence decreases
   - The PGM course difficulty increases
5. If we **didn't have the grade**, but had a weak recommendation letter.



Evidential Inference (Explaining)

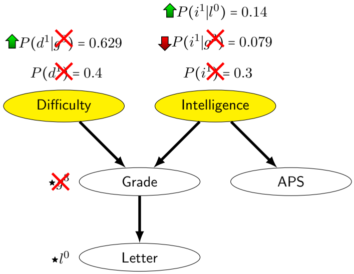
Bayesian Networks
Professor Ajoodha
Problem Statement
Bayesian Network
Independence
D-Separation
I-Equivalence
I-Maps
Min I-Maps
P-Maps

1. Now Suppose that we would like to consider hiring Xolani.
2. We calculate the probability that Xolani **is intelligent**.
3. We find Xolani's **grade** for the PGM course.
4. We can now **re-calculate Xolani's intelligence** (w.r.t. the grade).
    - Xolani's intelligence decreases
    - The PGM course difficulty increases
5. If we **didn't have the grade**, but had a weak recommendation letter.



Evidential Inference (Explaining)

1. Now Suppose that we would like to consider hiring Xolani.
2. We calculate the probability that Xolani **is intelligent**.
3. We find Xolani's **grade** for the PGM course.
4. We can now **re-calculate Xolani's intelligence** (w.r.t. the grade).
   - Xolani's intelligence decreases
   - The PGM course difficulty increases
5. If we **didn't have the grade**, but had a weak recommendation letter.

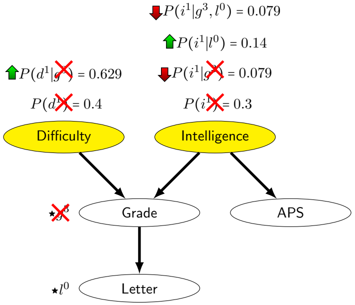


Evidential Inference (Explaining)

Bayesian Networks
Professor Ajoodha

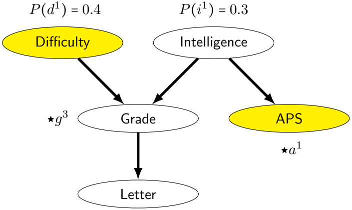
Problem Statement
Bayesian Network
Independence
D-Separation
I-Equivalence
I-Maps
Min I-Maps
P-Maps

1. Now Suppose that we would like to consider hiring Xolani.
2. We calculate the probability that Xolani **is intelligent**.
3. We find Xolani's **grade** for the PGM course.
4. We can now **re-calculate Xolani's intelligence** (w.r.t. the grade).
   - Xolani's intelligence decreases
   - The PGM course difficulty increases
5. If we **didn't have the grade**, but had a weak recommendation letter.



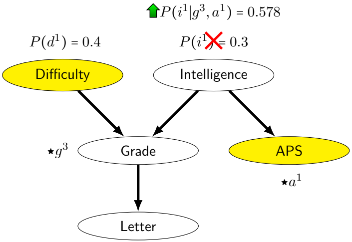
Intercausal Inference (Explaining Away)

1. Suppose Xolani obtained a good APS score in high school and a C in the PGM course.
2. Then his intelligence goes up.
   - Influence of the APS > the PGM grade.
3. The C in the PGM course increases the difficulty.
4. … and further increases when the APS evidence is introduced.
5. How does the cause **Intelligence influence Difficulty**? (Intercausal)



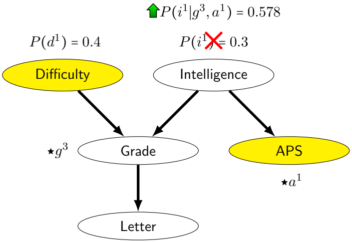
Intercausal Inference (Explaining Away)

1. Suppose Xolani obtained a good APS score in high school and a C in the PGM course.
2. Then his intelligence goes up.
   - Influence of the APS > the PGM grade.
3. The C in the PGM course increases the difficulty.
4. … and further increases when the APS evidence is introduced.
5. How does the cause **Intelligence influence Difficulty**? (Intercausal)



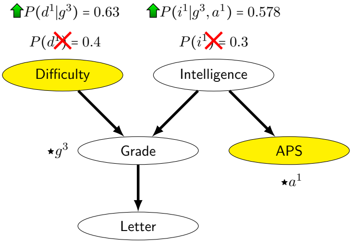
Intercausal Inference (Explaining Away)

1. Suppose Xolani obtained a good APS score in high school and a C in the PGM course.
2. Then his intelligence goes up.
   - Influence of the APS > the PGM grade.
3. The C in the PGM course increases the difficulty.
4. ... and further increases when the APS evidence is introduced.
5. How does the cause **Intelligence influence Difficulty**? (Intercausal)



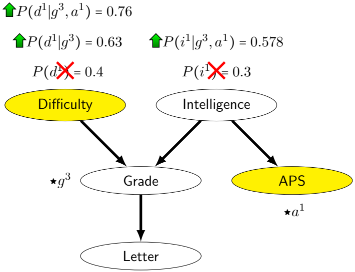
Intercausal Inference (Explaining Away)

1. Suppose Xolani obtained a good APS score in high school and a C in the PGM course.
2. Then his intelligence goes up.
   - Influence of the APS > the PGM grade.
3. The C in the PGM course increases the difficulty.
4. … and further increases when the APS evidence is introduced.
5. How does the cause **Intelligence influence Difficulty**? (Intercausal)



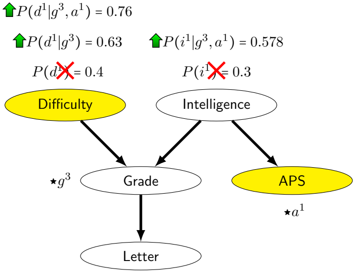
Intercausal Inference (Explaining Away)

1. Suppose Xolani obtained a good APS score in high school and a C in the PGM course.
2. Then his intelligence goes up.
   - Influence of the APS > the PGM grade.
3. The C in the PGM course increases the difficulty.
4. . . . and further increases when the APS evidence is introduced.
5. How does the cause **Intelligence influence Difficulty**? (Intercausal)



Intercausal Inference (Explaining Away)

1. Suppose Xolani obtained a good APS score in high school and a C in the PGM course.
2. Then his intelligence goes up.
    - Influence of the APS > the PGM grade.
3. The C in the PGM course increases the difficulty.
4. … and further increases when the APS evidence is introduced.
5. How does the cause **Intelligence influence Difficulty**? (Intercausal)



Bayesian Network Definition

A Bayesian Network is a pair $\mathcal{B} = (\mathcal{G}, P)$ where P factorises over $\mathcal{G}$.

1. $\mathcal{G}$ is graph over variables $X_1, \ldots, X_n$.
2. P factorises according to $\mathcal{G}$ nodes:

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

Can we guarantee independence?

When can we **guarantee independence between variables** in a Bayesian network structure?

Recall that $\mathcal{G}$ is made up of conditional independence assumptions, called $\mathcal{I}_\ell(\mathcal{G})$.

Does $(\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z})$ hold?

In order to find out if the independence holds. We could consider when it does **not hold**?

Lets consider this for:

- Direct Connections
- Indirect Connections

Direct Connection

Bayesian Networks

Professor Ajoodha

Problem Statement

Bayesian Network

Independence

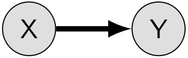
D-Separation

I-Equivalence

I-Maps

Min I-Maps

P-Maps

Supose we have the following: 

Then we will need to consider whether "X" influences "Y".

That is, whether $P(X|Y) \neq P(X)$ or $P(Y|X) \neq P(Y)$.

- Evidence for X can change Y and
- Evidence for Y can change X

Influence can flow through direct connections, despite any evidence from other variables (i.e. Z).

Indirect Connection

For indirect connections between X and Y via Z we have four possible cases:

1. Causal Trail: $X \rightarrow Z \rightarrow Y$
2. Evidential Trail: $X \leftarrow Z \leftarrow Y$
3. Common Cause: $X \leftarrow Z \rightarrow Y$
4. Common Effect: $X \rightarrow Z \leftarrow Y$

Causal Trail

G is not observed.

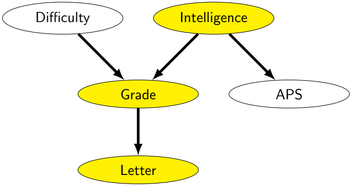
1. If student **is intelligent**
2. then chances of an **A is higher**
3. likelihood of good recommendation increases

∴ intelligence influence letter

G is observed.

1. If we know the grade
2. then intelligence increases
3. but letter can no longer influence intelligence

$\therefore (L \perp I \mid G)$



Causal Trail

G is not observed.

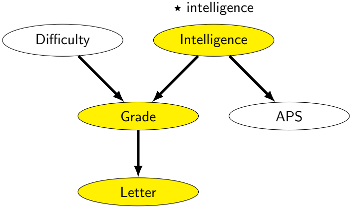
1. If student **is intelligent**
2. then chances of an **A is higher**
3. likelihood of good recommendation increases

∴ intelligence influence letter

G is observed.

1. If we know the grade
2. then intelligence increases
3. but letter can no longer influence intelligence

$\therefore (L \perp I \mid G)$



Causal Trail

G is not observed.

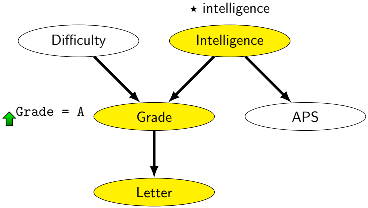
1. If student **is intelligent**
2. then chances of an **A is higher**
3. likelihood of good recommendation increases

$\therefore$ intelligence influence letter

G is observed.

1. If we know the grade
2. then intelligence increases
3. but letter can no longer influence intelligence

$\therefore (L \perp I \mid G)$



Causal Trail

G is not observed.

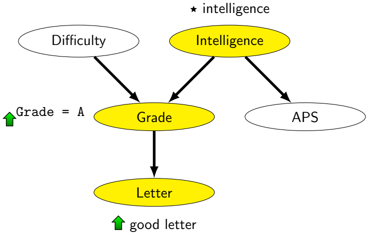
1. If student **is intelligent**
2. then chances of an **A is higher**
3. likelihood of good recommendation increases

∴ intelligence influence letter

G is observed.

1. If we know the grade
2. then intelligence increases
3. but letter can no longer influence intelligence

$\therefore (L \perp I \mid G)$



Causal Trail

G is not observed.

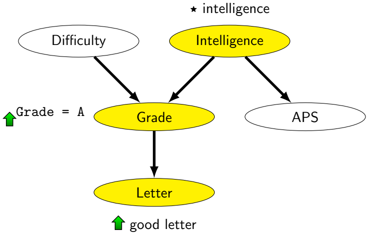
1. If student **is intelligent**
2. then chances of an **A is higher**
3. likelihood of good recommendation increases

$\therefore$ intelligence influence letter

G is observed.

1. If we know the grade
2. then intelligence increases
3. but letter can no longer influence intelligence

$\therefore (L \perp I \mid G)$



Causal Trail

G is not observed.

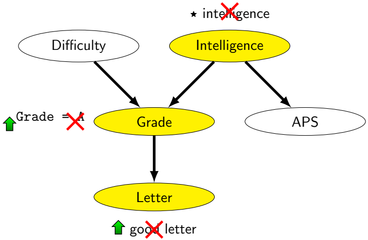
1. If student **is intelligent**
2. then chances of an **A is higher**
3. likelihood of good recommendation increases

∴ intelligence influence letter

G is observed.

1. If we know the grade
2. then intelligence increases
3. but letter can no longer influence intelligence

$\therefore (L \perp I \mid G)$



Causal Trail

G is not observed.

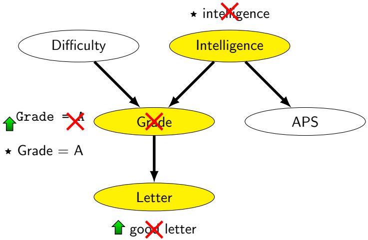
1. If student **is intelligent**
2. then chances of an **A is higher**
3. likelihood of good recommendation increases

∴ intelligence influence letter

G is observed.

1. If we know the grade
2. then intelligence increases
3. but letter can no longer influence intelligence

$\therefore (L \perp I \mid G)$



Causal Trail

G is not observed.

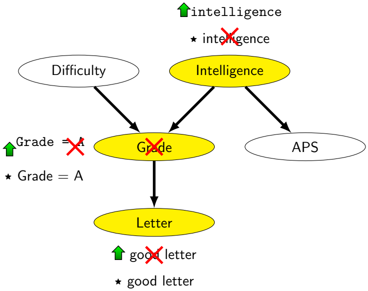
1. If student **is intelligent**
2. then chances of an **A is higher**
3. likelihood of good recommendation increases

∴ intelligence influence letter

G is observed.

1. If we know the grade
2. then intelligence increases
3. but letter can no longer influence intelligence

$\therefore (L \perp I \mid G)$



Causal Trail

G is not observed.

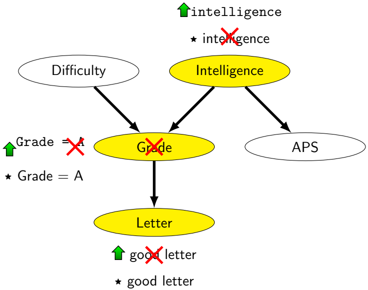
1. If student **is intelligent**
2. then chances of an **A is higher**
3. likelihood of good recommendation increases

∴ intelligence influence letter

G is observed.

1. If we know the grade
2. then intelligence increases
3. but letter can no longer influence intelligence

$\therefore (L \perp I \mid G)$



G is not observed.

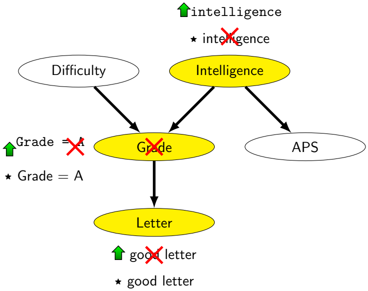
1. If student **is intelligent**
2. then chances of an **A is higher**
3. likelihood of good recommendation increases

$\therefore$ intelligence influence letter

G is observed.

1. If we know the grade
2. then intelligence increases
3. but letter can no longer influence intelligence

$$\therefore (L \perp I \mid G)$$



Evidential Trail

G is not observed.

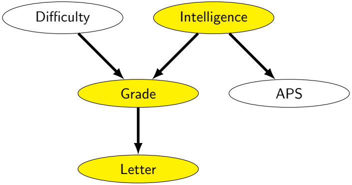
1. If student gets **good letter**
2. then chances of an **A is higher**
3. likelihood of intelligence increases

$\therefore$ letter influence intelligence

G is observed.

1. Then the grade influence intelligence
2. and the letter gives no more information about intelligence

$\therefore (I \perp L \mid G)$



Evidential Trail

G is not observed.

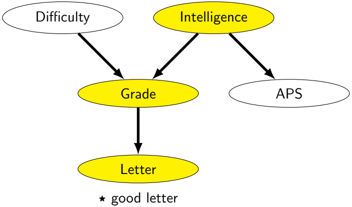
1. If student gets **good letter**
2. then chances of an **A is higher**
3. likelihood of intelligence increases

∴ letter influence intelligence

G is observed.

1. Then the grade influence intelligence
2. and the letter gives no more information about intelligence

$\therefore (I \perp L \mid G)$



Evidential Trail

G is not observed.

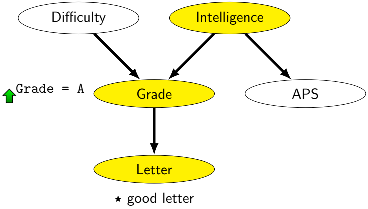
1. If student gets **good letter**
2. then chances of an **A is higher**
3. likelihood of intelligence increases

∴ letter influence intelligence

G is observed.

1. Then the grade influence intelligence
2. and the letter gives no more information about intelligence

$\therefore (I \perp L \mid G)$



Evidential Trail

G is not observed.

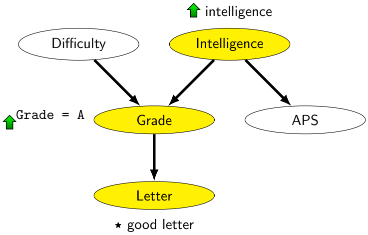
1. If student gets **good letter**
2. then chances of an **A is higher**
3. likelihood of intelligence increases

∴ letter influence intelligence

G is observed.

1. Then the grade influence intelligence
2. and the letter gives no more information about intelligence

$\therefore (I \perp L \mid G)$



Evidential Trail

G is not observed.

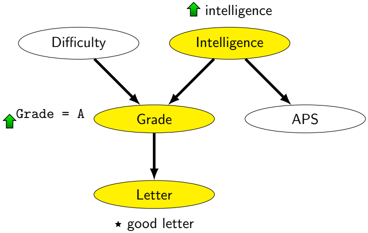
1. If student gets **good letter**
2. then chances of an **A is higher**
3. likelihood of intelligence increases

∴ letter influence intelligence

G is observed.

1. Then the grade influence intelligence
2. and the letter gives no more information about intelligence

$\therefore (I \perp L \mid G)$



Evidential Trail

Bayesian Networks

Professor Ajoodha

Problem Statement

Bayesian Network

Independence

D-Separation

I-Equivalence

I-Maps

Min I-Maps

P-Maps

G is not observed.

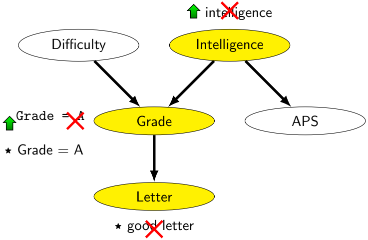
1. If student gets **good letter**
2. then chances of an **A is higher**
3. likelihood of intelligence increases

$\therefore$ letter influence intelligence

G is observed.

1. Then the grade influence intelligence
2. and the letter gives no more information about intelligence

$\therefore (I \perp L \mid G)$



Evidential Trail

G is not observed.

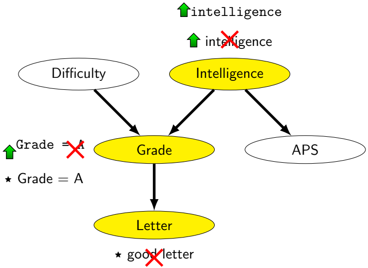
1. If student gets **good letter**
2. then chances of an **A is higher**
3. likelihood of intelligence increases

$\therefore$ letter influence intelligence

G is observed.

1. Then the grade influence intelligence
2. and the letter gives no more information about intelligence

$\therefore (I \perp L \mid G)$



Evidential Trail

G is not observed.

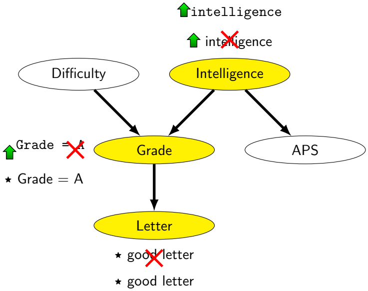
1. If student gets **good letter**
2. then chances of an **A is higher**
3. likelihood of intelligence increases

∴ letter influence intelligence

G is observed.

1. Then the grade influence intelligence
2. and the letter gives no more information about intelligence

$\therefore (I \perp L \mid G)$



Evidential Trail

G is not observed.

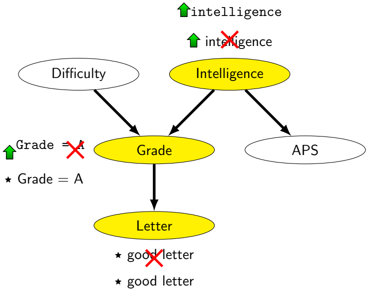
1. If student gets **good letter**
2. then chances of an **A is higher**
3. likelihood of intelligence increases

$\therefore$ letter influence intelligence

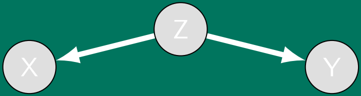
G is observed.

1. Then the grade influence intelligence
2. and the letter gives no more information about intelligence

$\therefore (I \perp L \mid G)$



Common Cause



I is not observed.

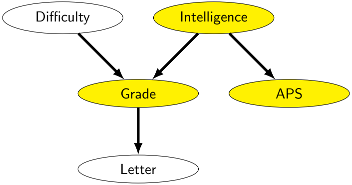
1. If student gets **good APS**
2. then chances of **intelligence increases**
3. likelihood of good **grade increases**

$\therefore$ APS influence intelligence

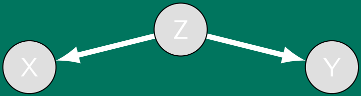
I is observed.

1. Then intelligence influence Grade/APS
2. but APS gives no more information about Grade

$\therefore (A \perp G \mid I)$



Common Cause



I is not observed.

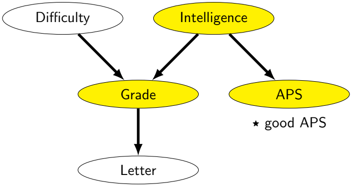
1. If student gets **good APS**
2. then chances of **intelligence increases**
3. likelihood of good **grade increases**

∴ APS influence intelligence

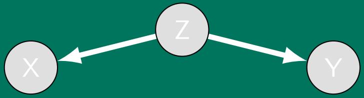
I is observed.

1. Then intelligence influence Grade/APS
2. but APS gives no more information about Grade

$\therefore (A \perp G \mid I)$



Common Cause



I is not observed.

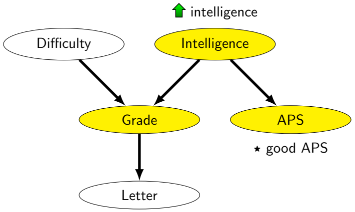
1. If student gets **good APS**
2. then chances of **intelligence increases**
3. likelihood of good **grade increases**

∴ APS influence intelligence

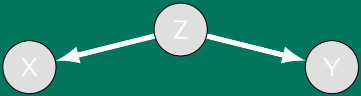
I is observed.

1. Then intelligence influence Grade/APS
2. but APS gives no more information about Grade

$\therefore (A \perp G \mid I)$



Common Cause



I is not observed.

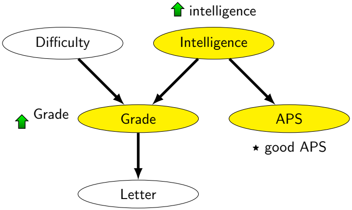
1. If student gets **good APS**
2. then chances of **intelligence increases**
3. likelihood of good **grade increases**

∴ APS influence intelligence

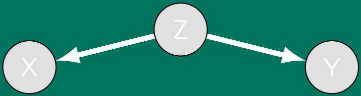
I is observed.

1. Then intelligence influence Grade/APS
2. but APS gives no more information about Grade

$\therefore (A \perp G \mid I)$



Common Cause



I is not observed.

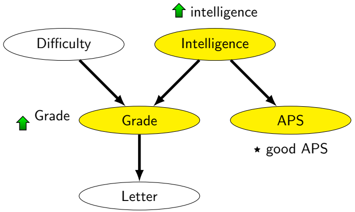
1. If student gets **good APS**
2. then chances of **intelligence increases**
3. likelihood of good **grade increases**

$\therefore$ APS influence intelligence

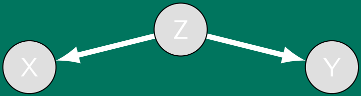
I is observed.

1. Then intelligence influence Grade/APS
2. but APS gives no more information about Grade

$\therefore (A \perp G \mid I)$



Common Cause



I is not observed.

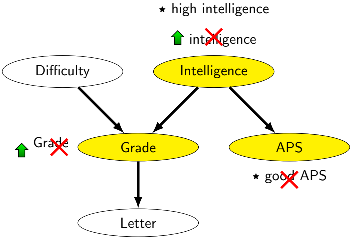
1. If student gets **good APS**
2. then chances of **intelligence increases**
3. likelihood of good **grade increases**

$\therefore$ APS influence intelligence

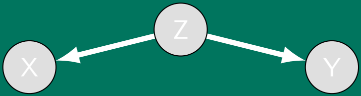
I is observed.

1. Then intelligence influence Grade/APS
2. but APS gives no more information about Grade

$\therefore (A \perp G \mid I)$



Common Cause



I is not observed.

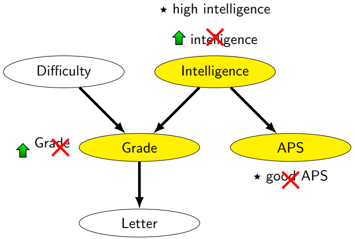
1. If student gets **good APS**
2. then chances of **intelligence increases**
3. likelihood of good **grade increases**

∴ APS influence intelligence

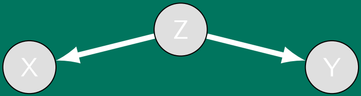
I is observed.

1. Then intelligence influence Grade/APS
2. but APS gives no more information about Grade

$\therefore (A \perp G \mid I)$



Common Cause



I is not observed.

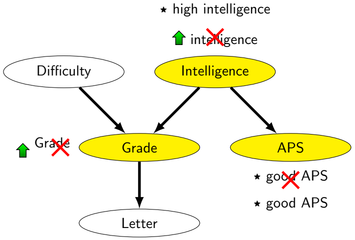
1. If student gets **good APS**
2. then chances of **intelligence increases**
3. likelihood of good **grade increases**

∴ APS influence intelligence

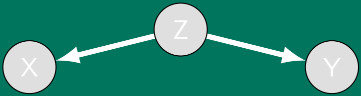
I is observed.

1. Then intelligence influence Grade/APS
2. but APS gives no more information about Grade

$\therefore (A \perp G \mid I)$



Common Cause



I is not observed.

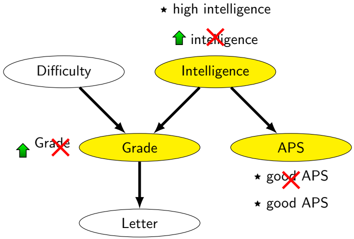
1. If student gets **good APS**
2. then chances of **intelligence increases**
3. likelihood of good **grade increases**

$\therefore$ APS influence intelligence

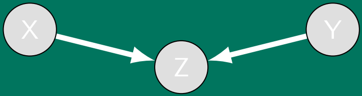
I is observed.

1. Then intelligence influence Grade/APS
2. but APS gives no more information about Grade

$\therefore (A \perp G \mid I)$



Common Effect



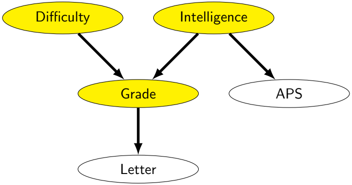
G is not observed.

1. It turns out that when G is **not observed**.
2. Intelligence and difficulty **are not** correlated

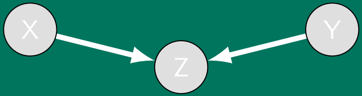
G is observed (or one of its descendants).

1. Then intelligence can influence difficulty

We call this a **v-structure**



Intuition behind V-structure

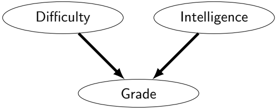


Counter-intuitive because ...

> In other scenarios, observing a variable blocks communication.

However ...

1. Without knowing the student's grade, the difficulty of the course doesn't imply anything about the student's intelligence.
2. Knowing the student received an 'A' grade, if the course is known to be difficult, it can influence our perception of the student's intelligence.



Therefore, difficulty can influence intelligence, if grade is known.

Independence in Graphical Models

This is EVERYTHING you need to know about independence in Bayesian Networks:

Tail	**Observed(Z)**	$\mathbf{X \perp Y}$
$X \rightarrow Z \rightarrow Y$	Yes	Yes
$X \rightarrow Z \rightarrow Y$	No	No
$X \leftarrow Z \leftarrow Y$	Yes	Yes
$X \leftarrow Z \leftarrow Y$	No	No
$X \leftarrow Z \rightarrow Y$	Yes	Yes
$X \leftarrow Z \rightarrow Y$	No	No
$X \rightarrow Z \leftarrow Y$	Yes***	No
$X \rightarrow Z \leftarrow Y$	No	Yes

***or one of its descendants.

General case of independence

In summary here is a definition of independence for Bayesian Networks:

1. Let $\mathcal{G}$ be a BN structure and $X_1 \rightleftharpoons \ldots \rightleftharpoons X_n$ a trail in $\mathcal{G}$.
2. The trail $X_1 \rightleftharpoons \ldots \rightleftharpoons X_n$ **is active** (influence can flow) given **Z** if:
   - Whenever there is a v-structure, $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$ then either X_i is observed or one of the descendants of X_i is observed.
   - No other nodes in the trail is in **Z**.

D-separation

Let $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ be three sets of nodes in $\mathcal{G}$.

We say that $\mathbf{X}$ and $\mathbf{Y}$ are d-separated given $\mathbf{Z}$ if:

There are no active trails between any node $X \in \mathbf{X}$ and $Y \in \mathbf{Y}$ given $\mathbf{Z}$.

Notation:

We use $\mathcal{I}(\mathcal{G})$ to denote the set of independences that correspond to d-separation.

$$\mathcal{I}(\mathcal{G}) = \{(\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z}) \ : \ \texttt{d-sep}_{\mathcal{G}}(\mathbf{X} : \mathbf{Y} \mid \mathbf{Z})\}$$

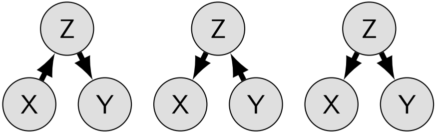
I-Equivalence

$\mathcal{I}(\mathcal{G})$ specifies independences associated with $\mathcal{G}$.

Multiple graph structures can be associated with the same set $\mathcal{I}(\mathcal{G})$.

These graph structures are called I-equivalent since the all encode $\mathcal{I}(\mathcal{G})$.

For example the following graphs are I-equivalent since they all encode $\mathcal{I}(\mathcal{G}) = \{X \perp Y \mid Z\}$:



There is no intrinsic property of the ground truth distribution, P, to favour one graph structure over another.

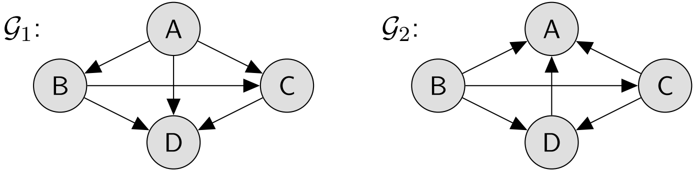
Is $\mathcal{G}_1$ and $\mathcal{G}_2$ I-equivalent?

How can we tell if two graph structures are I-equivalent?

If $\mathcal{G}_1$ and $\mathcal{G}_2$ have the same skeleton and the same set of v-structures $\implies$ they are I-equivalent.

Def: A v-structure, $X \rightarrow Z \leftarrow Y$, is an **immorality** if there is no direct edge (covering edge) between X and Y.

For example, the following complete graphs have the same immoralities (none) but different v-structures.



Theorem: Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be two graphs over $\mathcal{X}$. Then $\mathcal{G}_1$ and $\mathcal{G}_2$ have the same skeleton and the same set of immoralities $\iff$ they are I-equivalent.

From Distributions to Graphs

If P factorises over $\mathcal{G}$, then $\mathcal{G}$ must encode independencies that hold for P.

We can use $\mathcal{G}$ to reveal the structure P.

Given a distribution P, to what extent can we construct a graph, $\mathcal{G}$, whose independencies are a reasonable surrogate for the independencies in P?

A full joint distribution is too large to represent explicitly

But we can try to get as close to P as possible!

I-Map

Def: Let $\mathcal{I}(\mathcal{K}) \subseteq \mathcal{I}$ be a set of independencies for a graph $\mathcal{K}$. Then $\mathcal{K}$ (the graph) is an **I-map** for $\mathcal{I}$.

Def: A graph $\mathcal{K}$ is a **minimal I-map** for a set of independencies, $\mathcal{I}$, if it is an I-map for $\mathcal{I}$ and if the removal of a single edge from $\mathcal{K}$ renders it not an I-map.

To represent distribution P, we could:

- **On the one hand:** just use a complete graph (chain rule for prob) but this will not exploit independencies.
- **On the other hand:** just use minimal I-map to try to represent P.

Lets try to build a minimal I-map!

Algorithm to find a minimal I-map

Algorithm Build-Minimal-I-Map

Require: ordering : $X_1, \ldots, X_n$ and set $\mathcal{I}$

Set $\mathcal{G}$ to be an empty graph over $\mathcal{X}$

for $i = 1, \ldots n$ **do**

  U $\leftarrow \{X_1, \ldots, X_{i-1}\}$

  for U' $\subseteq \{X_1, \ldots, X_{i-1}\}$ **do**

    if U' $\subseteq$ **U** & $(X_i \perp \{X_1, \ldots, X_{i-1}\} - \textbf{U'} \mid \textbf{U'}) \in \mathcal{I}$ **then**

      U $\subseteq$ **U'**

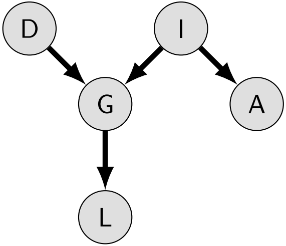
  for $X_j \in$ **U do**

    Add $X_j \rightarrow X_i$ to $\mathcal{G}$

Example 1

(1) Ordering : $D, I, A, G, L.$

Ground Truth:



Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\text{Student}}}) = \mathcal{I}(\mathcal{G}_{\text{Student}})$

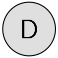
1) $(L \perp I, D, A \mid G)$
2) $(A \perp D, G, L \mid I)$
3) $(G \perp A \mid I, D)$
4) $(D \perp I, A)$

Example 1

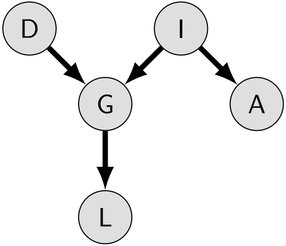
(1) Ordering: D, I, A, G, L.

(2) Add D

Learned Model:



Ground Truth:



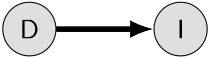
Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\text{Student}}}) = \mathcal{I}(\mathcal{G}_{\text{Student}})$

1) $(L \perp I, D, A \mid G)$
2) $(A \perp D, G, L \mid I)$
3) $(G \perp A \mid I, D)$
4) $(D \perp I, A)$

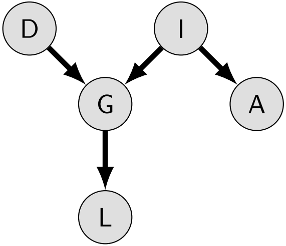
Example 1

(1) Ordering: D, I, A, G, L.
(2) **Add** D
(3) **Add** I

Learned Model:



Ground Truth:



Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\text{Student}}}) = \mathcal{I}(\mathcal{G}_{\text{Student}})$

1) $(L \perp I, D, A \mid G)$
2) $(A \perp D, G, L \mid I)$
3) $(G \perp A \mid I, D)$
4) $(D \perp I, A)$

Example 1

(1) Ordering: D, I, A, G, L.

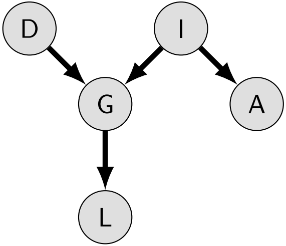
(2) **Add** D

(3) **Add** I

(4) **Remove** $D \rightarrow I$ since $(D \perp I, A) \in \mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}})$

Learned Model:

Ground Truth:



Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\text{Student}}}) = \mathcal{I}(\mathcal{G}_{\texttt{Student}})$

1) (L $\perp I, D, A \mid G$)
2) (A $\perp D, G, L \mid I$)
3) (G $\perp A \mid I, D$)
4) (D $\perp I, A$)

Example 1

(1) Ordering: D, I, A, G, L.

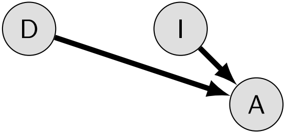
(2) **Add** D

(3) **Add** I

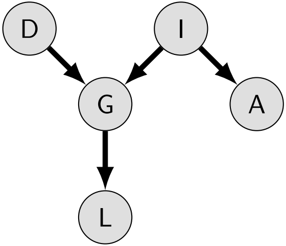
(4) **Remove** $D \to I$ since $(D \perp I, A) \in \mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}})$

(5) **Add** A

Learned Model:



Ground Truth:



Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}}) = \mathcal{I}(\mathcal{G}_{\texttt{Student}})$

1) (L $\perp I, D, A \mid G$)
2) (A $\perp D, G, L \mid I$)
3) (G $\perp A \mid I, D$)
4) (D $\perp I, A$)

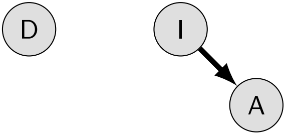
Example 1

(1) Ordering: D, I, A, G, L.

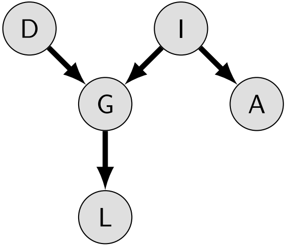
⋮

(6) **Remove** $D \to A$ since $(D \perp A) \in \mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}})$ (v-structure)

Learned Model:



Ground Truth:



Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}}) = \mathcal{I}(\mathcal{G}_{\texttt{Student}})$

1) (L $\perp I, D, A \mid G$)
2) (A $\perp D, G, L \mid I$)
3) (G $\perp A \mid I, D$)
4) (D $\perp I, A$)

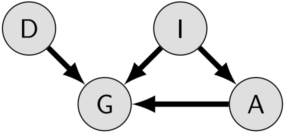
Example 1

(1) Ordering: D, I, A, G, L.

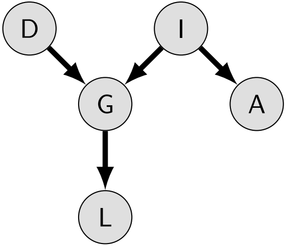
⋮

(7) **Add** G

Learned Model:



Ground Truth:



Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}}) = \mathcal{I}(\mathcal{G}_{\texttt{Student}})$

1) (L $\perp I, D, A \mid G$)
2) (A $\perp D, G, L \mid I$)
3) (G $\perp A \mid I, D$)
4) (D $\perp I, A$)

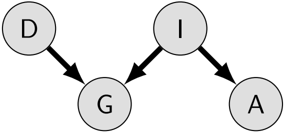
Example 1

(1) Ordering: D, I, A, G, L.

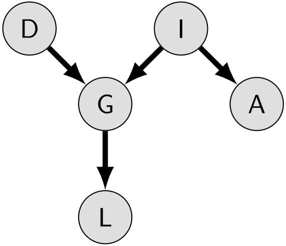
⋮

(8) **Remove** $A \to G$ since
$(G \perp A \mid I, D) \in \mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}})$

Learned Model:



Ground Truth:



Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}}) = \mathcal{I}(\mathcal{G}_{\texttt{Student}})$

1) $(\mathsf{L} \perp I, D, A \mid G)$
2) $(\mathsf{A} \perp D, G, L \mid I)$
3) $(\mathsf{G} \perp A \mid I, D)$
4) $(\mathsf{D} \perp I, A)$

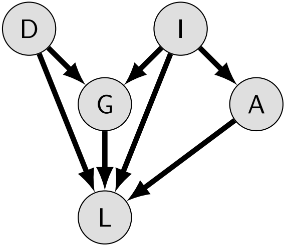
Example 1

(1) Ordering: D, I, A, G, L.

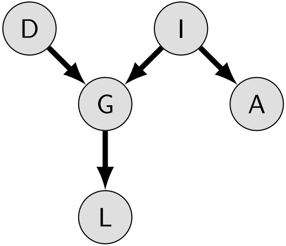
⋮

(9) **Add** L

Learned Model:



Ground Truth:



Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\text{Student}}}) = \mathcal{I}(\mathcal{G}_{\text{Student}})$

1) $(L \perp I, D, A \mid G)$
2) $(A \perp D, G, L \mid I)$
3) $(G \perp A \mid I, D)$
4) $(D \perp I, A)$

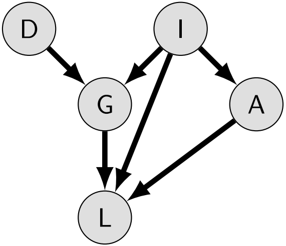
Example 1

(1) Ordering: D, I, A, G, L.

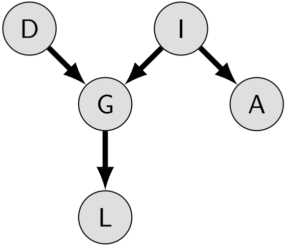
⋮

(10) **Remove** $D \rightarrow L$ since $(L \perp D \mid G) \in \mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}})$

Learned Model:



Ground Truth:



Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}}) = \mathcal{I}(\mathcal{G}_{\texttt{Student}})$

1) (L $\perp I, D, A \mid G$)
2) (A $\perp D, G, L \mid I$)
3) (G $\perp A \mid I, D$)
4) (D $\perp I, A$)

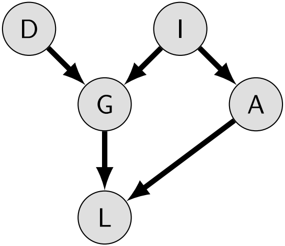
Example 1

(1) Ordering: D, I, A, G, L.

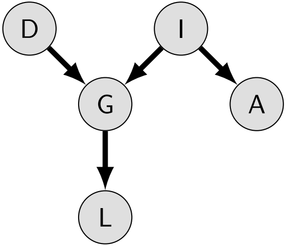
⋮

(11) **Remove** $I \to L$ since $(L \perp I \mid G) \in \mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}})$

Learned Model:



Ground Truth:



Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}}) = \mathcal{I}(\mathcal{G}_{\texttt{Student}})$

1) (L $\perp I, D, A \mid G$)
2) (A $\perp D, G, L \mid I$)
3) (G $\perp A \mid I, D$)
4) (D $\perp I, A$)

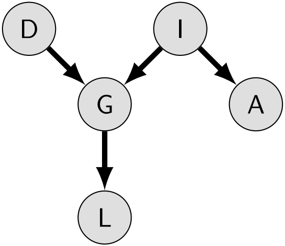
Example 1

(1) Ordering: D, I, A, G, L.

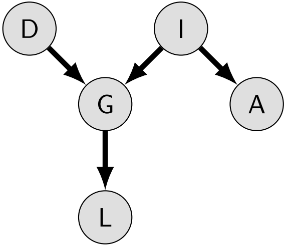
⋮

(12) **Remove** $A \to L$ since $(L \perp A \mid G) \in \mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}})$

Learned Model:



Ground Truth:



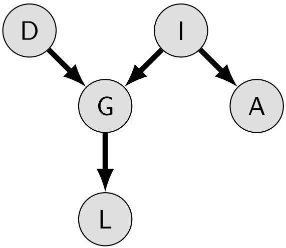
Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}}) = \mathcal{I}(\mathcal{G}_{\texttt{Student}})$

1) $(\text{L} \perp I, D, A \mid G)$
2) $(\text{A} \perp D, G, L \mid I)$
3) $(\text{G} \perp A \mid I, D)$
4) $(\text{D} \perp I, A)$

Example 2

(1) Ordering: L, A, G, I, D.

Ground Truth:



Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\text{Student}}}) = \mathcal{I}(\mathcal{G}_{\texttt{Student}})$

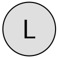
1) $(L \perp I, D, A \mid G)$
2) $(A \perp D, G, L \mid I)$
3) $(G \perp A \mid I, D)$
4) $(D \perp I, A)$

Example 2

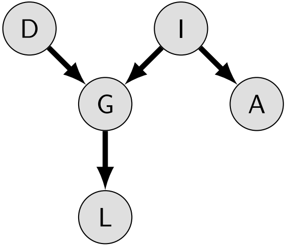
(1) Ordering: L, A, G, I, D.

(2) **Add** L

Learned Model:



Ground Truth:



Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\text{Student}}}) = \mathcal{I}(\mathcal{G}_{\text{Student}})$

1) (L $\perp I, D, A \mid G$)
2) (A $\perp D, G, L \mid I$)
3) (G $\perp A \mid I, D$)
4) (D $\perp I, A$)

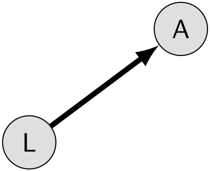
Example 2

(1) Ordering: L, A, G, I, D.

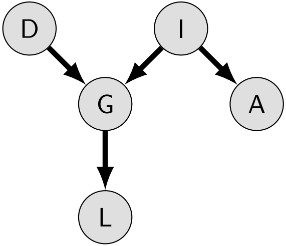
⋮

(3) **Add** A

Learned Model:



Ground Truth:



Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\mathtt{Student}}}) = \mathcal{I}(\mathcal{G}_{\mathtt{Student}})$

1) $(L \perp I, D, A \mid G)$
2) $(A \perp D, G, L \mid I)$
3) $(G \perp A \mid I, D)$
4) $(D \perp I, A)$

Example 2

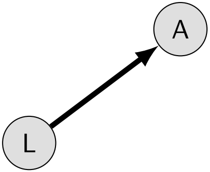
(1) Ordering: L, A, G, I, D.

⋮

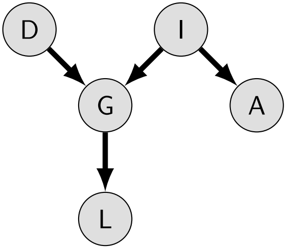
Remove?

* *A has no parent which can render it independent to L.*

Learned Model:



Ground Truth:



Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\text{Student}}}) = \mathcal{I}(\mathcal{G}_{\text{Student}})$

1) $(L \perp I, D, A \mid G)$
2) $(A \perp D, G, L \mid I)$
3) $(G \perp A \mid I, D)$
4) $(D \perp I, A)$

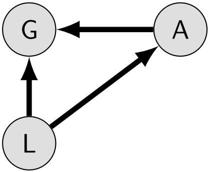
Example 2

(1) Ordering: L, A, G, I, D.

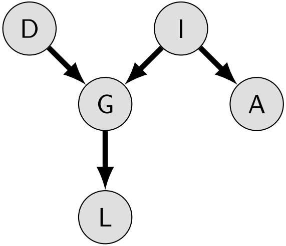
⋮

(4) **Add** G

Learned Model:



Ground Truth:



Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\text{Student}}}) = \mathcal{I}(\mathcal{G}_{\text{Student}})$

1) $(L \perp I, D, A \mid G)$
2) $(A \perp D, G, L \mid I)$
3) $(G \perp A \mid I, D)$
4) $(D \perp I, A)$

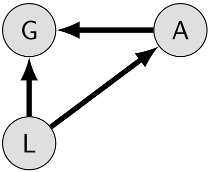
Example 2

Remove?

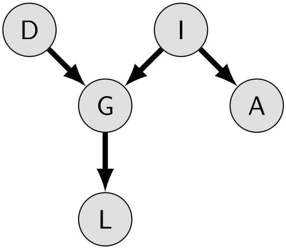
* *G is dependent on L. Keep* $L \rightarrow G$

** *Although* $(G \perp A \mid I)$, *but* $(G \not\perp A \mid L)$. *Keep* $A \rightarrow G$

Learned Model:



Ground Truth:



Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\text{Student}}}) = \mathcal{I}(\mathcal{G}_{\text{Student}})$

1) $(\text{L} \perp I, D, A \mid G)$
2) $(\text{A} \perp D, G, L \mid I)$
3) $(\text{G} \perp A \mid I, D)$
4) $(\text{D} \perp I, A)$

Example 2

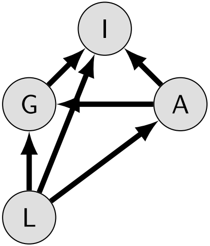
Bayesian Networks
Professor Ajoodha
Problem Statement
Bayesian Network
Independence
D-Separation
I-Equivalence
I-Maps
Min I-Maps
P-Maps

(1) Ordering: L, A, G, I, D.

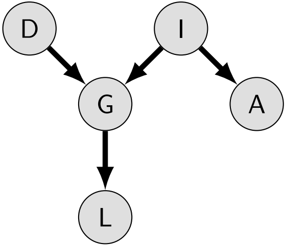
⋮

(5) **Add** I

Learned Model:



Ground Truth:



Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\text{Student}}}) = \mathcal{I}(\mathcal{G}_{\text{Student}})$

1) (L $\perp I, D, A \mid G$)
2) (A $\perp D, G, L \mid I$)
3) (G $\perp A \mid I, D$)
4) (D $\perp I, A$)

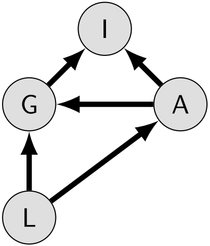
Example 2

(1) Ordering: L, A, G, I, D.

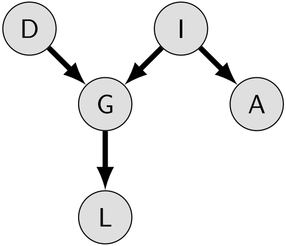
⋮

(6) **Remove** $L \to I$ since $(L \perp I \mid G) \in \mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}})$

Learned Model:



Ground Truth:



Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}}) = \mathcal{I}(\mathcal{G}_{\texttt{Student}})$

1) (L $\perp I, D, A \mid G$)
2) (A $\perp D, G, L \mid I$)
3) (G $\perp A \mid I, D$)
4) (D $\perp I, A$)

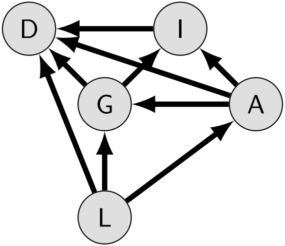
Example 2

(1) Ordering: L, A, G, I, D.

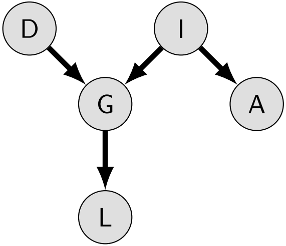
⋮

(7) **Add** D

Learned Model:



Ground Truth:



Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\text{Student}}}) = \mathcal{I}(\mathcal{G}_{\text{Student}})$

1) (L $\perp I, D, A \mid G$)
2) (A $\perp D, G, L \mid I$)
3) (G $\perp A \mid I, D$)
4) (D $\perp I, A$)

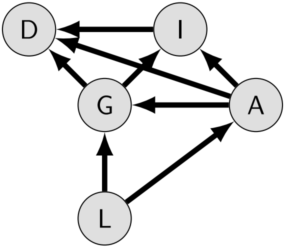
Example 2

(1) Ordering: L, A, G, I, D.

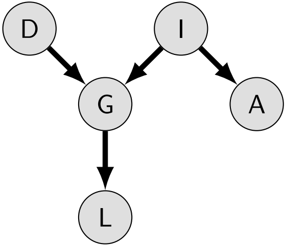
⋮

(8) **Remove** $L \rightarrow D$ since $(L \perp I, D, A \mid G) \in \mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}})$

Learned Model:



Ground Truth:



Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}}) = \mathcal{I}(\mathcal{G}_{\texttt{Student}})$

1) (L $\perp I, D, A \mid G$)
2) (A $\perp D, G, L \mid I$)
3) (G $\perp A \mid I, D$)
4) (D $\perp I, A$)

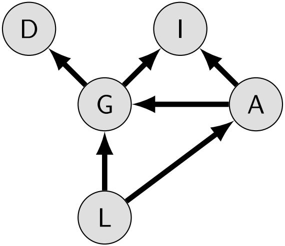
Example 2

(1) Ordering: L, A, G, I, D.

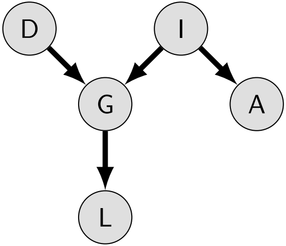
⋮

(9) **Remove** $\{A, I\} \to D$ since $(D \perp I, A) \in \mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}})$

Learned Model:



Ground Truth:



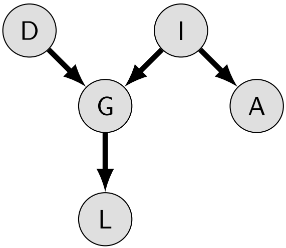
Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}}) = \mathcal{I}(\mathcal{G}_{\texttt{Student}})$

1) $(L \perp I, D, A \mid G)$
2) $(A \perp D, G, L \mid I)$
3) $(G \perp A \mid I, D)$
4) $(D \perp I, A)$

Example 3

(1) Ordering: L, D, A, I, G.

Ground Truth:



Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}}) = \mathcal{I}(\mathcal{G}_{\texttt{Student}})$

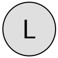
1) $(L \perp I, D, A \mid G)$
2) $(A \perp D, G, L \mid I)$
3) $(G \perp A \mid I, D)$
4) $(D \perp I, A)$

Example 3

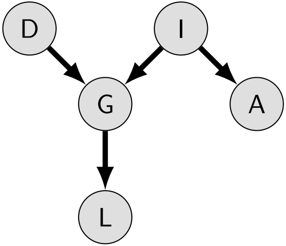
(1) Ordering: L, D, A, I, G.

(2) **Add** L

Learned Model:



Ground Truth:



Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}}) = \mathcal{I}(\mathcal{G}_{\texttt{Student}})$

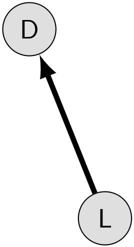
1) (L $\perp I, D, A \mid G$)
2) (A $\perp D, G, L \mid I$)
3) (G $\perp A \mid I, D$)
4) (D $\perp I, A$)

Example 3

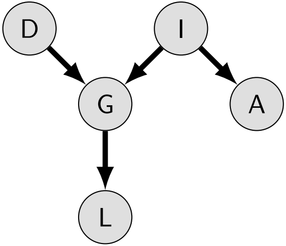
(1) Ordering: L, D, A, I, G.

(3) **Add** D

Learned Model:



Ground Truth:



Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}}) = \mathcal{I}(\mathcal{G}_{\texttt{Student}})$

1) $(L \perp I, D, A \mid G)$
2) $(A \perp D, G, L \mid I)$
3) $(G \perp A \mid I, D)$
4) $(D \perp I, A)$

Example 3

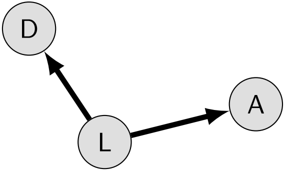
(1) Ordering: L, D, A, I, G.

$\vdots$

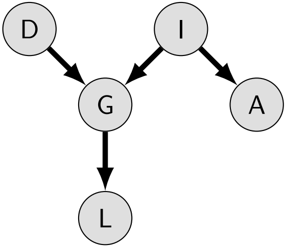
(4) **Add** A

* **Remove** $D \rightarrow A$ since $(D \perp I, A) \in \mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}})$

Learned Model:



Ground Truth:



Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}}) = \mathcal{I}(\mathcal{G}_{\texttt{Student}})$

1) (L $\perp I, D, A \mid G$)
2) (A $\perp D, G, L \mid I$)
3) (G $\perp A \mid I, D$)
4) (D $\perp I, A$)

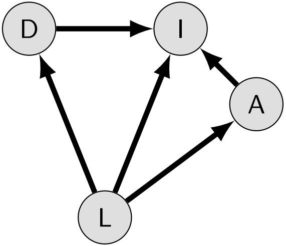
Example 3

(1) Ordering: L, D, A, I, G.

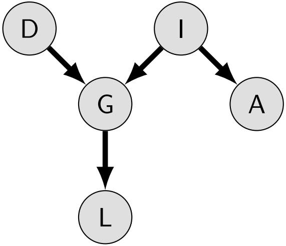
⋮

(5) **Add** I

Learned Model:



Ground Truth:



Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\text{Student}}}) = \mathcal{I}(\mathcal{G}_{\text{Student}})$

1) (L $\perp I, D, A \mid G$)
2) (A $\perp D, G, L \mid I$)
3) (G $\perp A \mid I, D$)
4) (D $\perp I, A$)

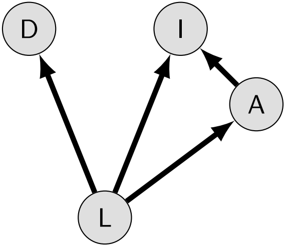
Example 3

(1) Ordering: L, D, A, I, G.

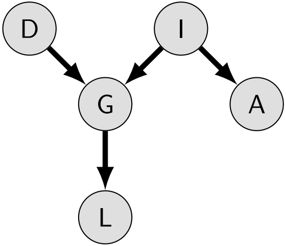
⋮

* **Remove** $D \rightarrow I$ since $(D \perp I, A) \in \mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}})$

Learned Model:



Ground Truth:



Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}}) = \mathcal{I}(\mathcal{G}_{\texttt{Student}})$

1) $(L \perp I, D, A \mid G)$
2) $(A \perp D, G, L \mid I)$
3) $(G \perp A \mid I, D)$
4) $(D \perp I, A)$

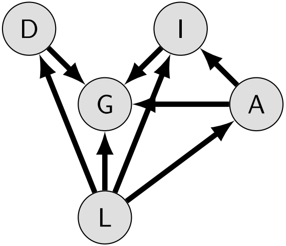
Example 3

(1) Ordering: L, D, A, I, G.

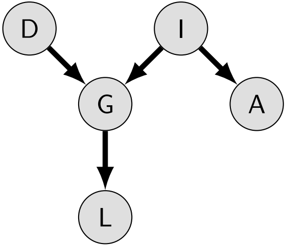
⋮

(6) **Add** G

Learned Model:



Ground Truth:



Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\mathtt{Student}}}) = \mathcal{I}(\mathcal{G}_{\mathtt{Student}})$

1) (L $\perp I, D, A \mid G$)
2) (A $\perp D, G, L \mid I$)
3) (G $\perp A \mid I, D$)
4) (D $\perp I, A$)

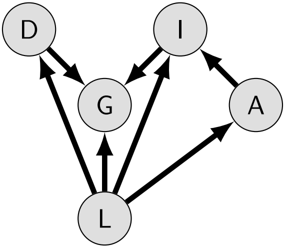
Example 3

(1) Ordering: L, D, A, I, G.

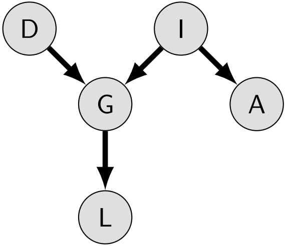
⋮

(7) **Remove** $A \to G$ since $(G \perp A \mid I) \in \mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}})$

Learned Model:



Ground Truth:



Conditional Independence Assumptions $\mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}}) = \mathcal{I}(\mathcal{G}_{\texttt{Student}})$

1) (L $\perp I, D, A \mid G$)
2) (A $\perp D, G, L \mid I$)
3) (G $\perp A \mid I, D$)
4) (D $\perp I, A$)

Limitations of Minimal I-Maps

1. All three examples where minimal I-maps.
2. However Example 2 and 3 fail to capture all of the independencies that hold in $\mathcal{I}(P_{\mathcal{B}^{\texttt{Student}}})$.
3. Therefore, even if $\mathcal{G}$ is a minimal I-Map for the true distribution P, there is **no guarantee** that $\mathcal{G}$ will capture P.

P-Maps

We say that a graph $\mathcal{K}$ is a perfect map (P-map) for a set of independencies $\mathcal{I}$ if we have that $\mathcal{I}(\mathcal{K}) = \mathcal{I}$.

We say that $\mathcal{K}$ is a P-map for P if $\mathcal{I}(\mathcal{K}) = \mathcal{I}(P)$.

E.g. $\mathcal{G}_{\texttt{Student}}$ is a P-map for $P_{\mathcal{B}^{\texttt{Student}}}$).

Does every distribution have a perfect map?

No.

1. Regularity in the parameterization which can not be captured in $\mathcal{G}$.
2. Independencies imposed by the structure are not appropriate.

For example, the following assumptions for distribution P cannot be coded in the structure of a BN.

- $A \perp C \mid B, D$
- $B \perp D \mid A, C$