Lecture 1

Graphical Models for Dependence Relationships

An Introduction

Causal Inference Using Graphs
August 6, 2019

Adam Glynn
Department of Political Science and QTM
Emory University

Acknowledgements

Many of the slides from lectures 1 and 2 today were written by Kevin Quinn.

Goals and Objectives for This Course:

- Introduce causal graphs
- Show how causal graphs help us think about standard techniques
- Show how causal graphs help us think about not yet standard techniques

How I Use Graphs for Causal Inference

Two Possible Metaphors:

1. Draw with graphs, ink with potential outcomes
2. Sketch with graphs, paint with potential outcomes

How I Use Graphs for Causal Inference

Two Possible Metaphors:

1. Draw with graphs, ink with potential outcomes
2. Sketch with graphs, paint with potential outcomes

Many of the mistakes I see in seminar could be avoided if the author(s) sketched a bit before starting to paint.

Goals and Objectives for This Morning:

- Introduce graphical notation and terminology
- Build intuition about properties of probabilistic systems represented as directed graphs
- Provide some motivating examples

Overview

1. **Why Study Graphical Models?**

2. **Definitions and Terminology**
   - Basic Definitions
   - Graph Drawing Conventions
   - Some Additional Definitions
   - Familial Relations Specific to Directed Graphs

3. **Directed Acyclic Graphs (DAGs)**
   - Using DAGs to Represent Complicated Joint Distributions
   - Observational Equivalence
   - Conditional Independence and *d*-Separation
   - Faithfulness

What Application Areas Make Use of Graphical Models?

Graphical models are widely used in a variety of application areas:

- Social network analysis
- Causal inference
- Pattern recognition and machine learning
- Information retrieval
- Document summarization
- Multivariate analysis
- ...

Some Questions that are Easy to Answer Using Graphical Models

- Are X and Y marginally independent?

Some Questions that are Easy to Answer Using Graphical Models

- Are X and Y marginally independent?
- Is X conditionally independent of Y given Z?

Some Questions that are Easy to Answer Using Graphical Models

- Are X and Y marginally independent?
- Is X conditionally independent of Y given Z?
- How influential is a particular actor in a network?

Some Questions that are Easy to Answer Using Graphical Models

- Are X and Y marginally independent?
- Is X conditionally independent of Y given Z?
- How influential is a particular actor in a network?
- What's the full conditional distribution of $[\theta_1 | y, \theta_2]$?

Some Questions that are Easy to Answer Using Graphical Models

- Are X and Y marginally independent?
- Is X conditionally independent of Y given Z?
- How influential is a particular actor in a network?
- What's the full conditional distribution of $[\theta_1 | y, \theta_2]$?
- What will happen to Y if a variable X is set to x by outside intervention?

Some Questions that are Easy to Answer Using Graphical Models

- Are X and Y marginally independent?
- Is X conditionally independent of Y given Z?
- How influential is a particular actor in a network?
- What's the full conditional distribution of $[\theta_1 | y, \theta_2]$?
- What will happen to Y if a variable X is set to x by outside intervention?
- Which background variables need to be adjusted for in order to get a consistent estimate of the causal effect of X on Y?

What and Why of Graphical Models

In essence, a graphical model (as we will use the term in this course) is a particular visual representation of a probabilistic system.

What and Why of Graphical Models

In essence, a graphical model (as we will use the term in this course) is a particular visual representation of a probabilistic system.

Graphical models are widely used because they allow one to:

What and Why of Graphical Models

In essence, a graphical model (as we will use the term in this course) is a particular visual representation of a probabilistic system.

Graphical models are widely used because they allow one to:

- express complex mathematical systems with an equivalent visual representation

What and Why of Graphical Models

In essence, a graphical model (as we will use the term in this course) is a particular visual representation of a probabilistic system.

Graphical models are widely used because they allow one to:

- express complex mathematical systems with an equivalent visual representation
- read off non-trivial mathematical properties of the system directly from the graph

What and Why of Graphical Models

In essence, a graphical model (as we will use the term in this course) is a particular visual representation of a probabilistic system.

Graphical models are widely used because they allow one to:

- express complex mathematical systems with an equivalent visual representation
- read off non-trivial mathematical properties of the system directly from the graph
- simplify complicated computations by taking account of the structure of the graph

What and Why of Graphical Models

In essence, a graphical model (as we will use the term in this course) is a particular visual representation of a probabilistic system.

Graphical models are widely used because they allow one to:

- express complex mathematical systems with an equivalent visual representation
- read off non-trivial mathematical properties of the system directly from the graph
- simplify complicated computations by taking account of the structure of the graph
- present substantive assumptions in a transparent manner

Basic Definitions

Definition (Directed Graph)

A directed graph $\mathcal{G}$ is a pair $\langle \mathcal{V}, \mathcal{E} \rangle$ where $\mathcal{V}$ is a finite set of vertices (a.k.a. nodes) and $\mathcal{E}$ is the set of directed edges (a.k.a. directed arcs or directed links).

Each directed edge in $\mathcal{E}$ is an ordered pair of distinct vertices from $\mathcal{V} \times \mathcal{V}$.

A directed edge $(V_i, V_j) \in \mathcal{E}$ is also denoted $V_i \rightarrow V_j$.

Basic Definitions

Definition (Directed Graph)

A directed graph $\mathcal{G}$ is a pair $\langle \mathcal{V}, \mathcal{E} \rangle$ where $\mathcal{V}$ is a finite set of vertices (a.k.a. nodes) and $\mathcal{E}$ is the set of directed edges (a.k.a. directed arcs or directed links).

Each directed edge in $\mathcal{E}$ is an ordered pair of distinct vertices from $\mathcal{V} \times \mathcal{V}$.

A directed edge $(V_i, V_j) \in \mathcal{E}$ is also denoted $V_i \to V_j$.

Note that this definition does not allow edges from a node to itself or multiple copies of edges.

Basic Definitions

Definition (Directed Graph)

A directed graph $\mathcal{G}$ is a pair $\langle \mathcal{V}, \mathcal{E} \rangle$ where $\mathcal{V}$ is a finite set of vertices (a.k.a. nodes) and $\mathcal{E}$ is the set of directed edges (a.k.a. directed arcs or directed links).

Each directed edge in $\mathcal{E}$ is an ordered pair of distinct vertices from $\mathcal{V} \times \mathcal{V}$.

A directed edge $(V_i, V_j) \in \mathcal{E}$ is also denoted $V_i \to V_j$.

Note that this definition does not allow edges from a node to itself or multiple copies of edges.

In this class, we will think of each $V \in \mathcal{V}$ as being a (possibly non-scalar) random variable and each directed edge $(V_i, V_j) \in \mathcal{E}$ as some relationship between V_i and V_j.

This will be made more precise below.

Graph Drawing Conventions

There are many different conventions that tend to vary both within and across application areas.

Graph Drawing Conventions

There are many different conventions that tend to vary both within and across application areas.

The conventions that we will use in these notes are roughly consistent with those used by Judea Pearl in much (but not all) of his work.

Graph Drawing Conventions

There are many different conventions that tend to vary both within and across application areas.

The conventions that we will use in these notes are roughly consistent with those used by Judea Pearl in much (but not all) of his work.

1. Represent each unobservable vertex (i.e., latent variable) with an open circle.

Graph Drawing Conventions

There are many different conventions that tend to vary both within and across application areas.

The conventions that we will use in these notes are roughly consistent with those used by Judea Pearl in much (but not all) of his work.

1. Represent each unobservable vertex (i.e., latent variable) with an open circle.
2. Represent each observable vertex (i.e., manifest variable) with a closed circle.

Graph Drawing Conventions

There are many different conventions that tend to vary both within and across application areas.

The conventions that we will use in these notes are roughly consistent with those used by Judea Pearl in much (but not all) of his work.

1. Represent each unobservable vertex (i.e., latent variable) with an open circle.
2. Represent each observable vertex (i.e., manifest variable) with a closed circle.
3. For each directed edge $(V_i, V_j) \in \mathcal{E}$ draw a solid arrow from V_i to V_j when both nodes are observable and a dashed arrow from V_i to V_j when one or both of the nodes are unobservable.

Graph Drawing Conventions

There are many different conventions that tend to vary both within and across application areas.

The conventions that we will use in these notes are roughly consistent with those used by Judea Pearl in much (but not all) of his work.

1. Represent each unobservable vertex (i.e., latent variable) with an open circle.
2. Represent each observable vertex (i.e., manifest variable) with a closed circle.
3. For each directed edge $(V_i, V_j) \in \mathcal{E}$ draw a solid arrow from V_i to V_j when both nodes are observable and a dashed arrow from V_i to V_j when one or both of the nodes are unobservable.

Note the difference between "(un)observed" and "(un)observable".

Example

Example (Drawing a Simple Directed Graph)

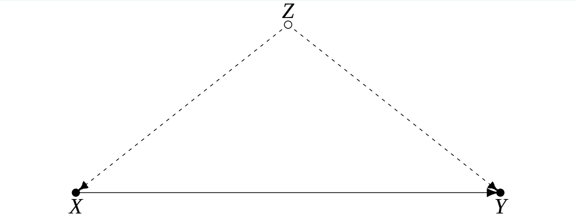
Consider the simple directed graph with vertices $\{X, Y, Z\}$ and edges $\{X \rightarrow Y,\ Z \rightarrow X,\ Z \rightarrow Y\}$ with X and Y observable and Z unobservable.

Example

Example (Drawing a Simple Directed Graph)

Consider the simple directed graph with vertices $\{X, Y, Z\}$ and edges $\{X \to Y,\ Z \to X,\ Z \to Y\}$ with X and Y observable and Z unobservable.

We would express this graph visually as:



Some Additional Definitions

Definition (Path)

A path from V_i to V_j in a graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ is a sequence of distinct nodes $V_i = X_0, \ldots, X_n = V_j$ such that $(X_{k-1}, X_k) \in \mathcal{E}$ or $(X_k, X_{k-1}) \in \mathcal{E}$ for each $k = 1, \ldots, n$. We write $V_i \sim V_j$ to denote a path from V_i to V_j.

Some Additional Definitions

Definition (Path)

A path from V_i to V_j in a graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ is a sequence of distinct nodes $V_i = X_0, \ldots, X_n = V_j$ such that $(X_{k-1}, X_k) \in \mathcal{E}$ or $(X_k, X_{k-1}) \in \mathcal{E}$ for each $k = 1, \ldots, n$. We write $V_i{\sim}V_j$ to denote a path from V_i to V_j.

Note that the direction of the edges does not matter and that a path can't visit the same node more than once.

Some Additional Definitions

Definition (Path)

A path from V_i to V_j in a graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ is a sequence of distinct nodes $V_i = X_0, \ldots, X_n = V_j$ such that $(X_{k-1}, X_k) \in \mathcal{E}$ or $(X_k, X_{k-1}) \in \mathcal{E}$ for each $k = 1, \ldots, n$. We write $V_i {\sim} V_j$ to denote a path from V_i to V_j.

Note that the direction of the edges does not matter and that a path can't visit the same node more than once.

Example

A path from V_1 to V_3 exists in each of the following four graphs.

$$V_1 \rightarrow V_2 \rightarrow V_3$$
$$V_1 \leftarrow V_2 \rightarrow V_3$$
$$V_1 \rightarrow V_2 \leftarrow V_3$$
$$V_1 \leftarrow V_2 \leftarrow V_3$$

Some Additional Definitions

Definition (Directed Path)

A directed path from V_i to V_j in a graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ is a sequence of distinct nodes $V_i = X_0, \ldots, X_n = V_j$ such that $(X_{k-1}, X_k) \in \mathcal{E}$ and $(X_{k-1}, X_k) \notin \mathcal{E}$ for each $k = 1, \ldots, n$. We write $V_i \rightsquigarrow V_j$ to denote a directed path from V_i to V_j.

Definition (Directed Path)

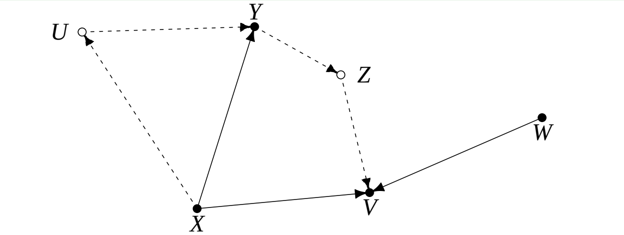
A directed path from V_i to V_j in a graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ is a sequence of distinct nodes $V_i = X_0, \ldots, X_n = V_j$ such that $(X_{k-1}, X_k) \in \mathcal{E}$ and $(X_{k-1}, X_k) \notin \mathcal{E}$ for each $k = 1, \ldots, n$. We write $V_i \leadsto V_j$ to denote a directed path from V_i to V_j.

Put more informally, a directed path from V_i to a V_j is a path from V_i to a V_j in which all the edges on the path have arrows pointing toward V_j.

Example

Example (Paths and Directed Paths)

Consider the graph depicted below.



Here there are 3 paths from X to Z:
1) $(X \to U \to Y \to Z)$,
2) $(X \to Y \to Z)$, and
3) $(X \to V \leftarrow Z)$.

Of these, only $(X \to U \to Y \to Z)$ and $(X \to Y \to Z)$ are directed paths from X to Z.

Familial Relations Specific to Directed Graphs

Definition (Parents)

In a directed graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ the set of parents of a node $V \in \mathcal{V}$ is defined to be:

$$\mathrm{pa}(V) = \{Z \in \mathcal{V} : (Z, V) \in \mathcal{E}\}$$

The set of parents of a set of nodes $\mathcal{W} \subseteq \mathcal{V}$ is defined to be:

$$\mathrm{pa}(\mathcal{W}) = \bigcup_{V \in \mathcal{W}} \mathrm{pa}(V)$$

Put more informally, the parents of V are all nodes from which there is a directed edge to V.

Familial Relations Specific to Directed Graphs

Definition (Ancestors)

In a directed graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ the set of ancestors of a node $V \in \mathcal{V}$ is defined to be:

$$\mathrm{an}(V) = \{Z \in \mathcal{V} : Z \rightsquigarrow V\}$$

The set of ancestors of a set of nodes $\mathcal{W} \subseteq \mathcal{V}$ is defined to be:

$$\mathrm{an}(\mathcal{W}) = \bigcup_{V \in \mathcal{W}} \mathrm{an}(V)$$

In words, the set of ancestors of V consists of the vertices from which there is a directed path to V.

Familial Relations Specific to Directed Graphs

Definition (Children)

In a directed graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ the set of children of a node $V \in \mathcal{V}$ is defined to be:

$$\text{ch}(V) = \{Z \in \mathcal{V} : (V, Z) \in \mathcal{E}\}$$

The set of children of a set of nodes $\mathcal{W} \subseteq \mathcal{V}$ is defined to be:

$$\text{ch}(\mathcal{W}) = \bigcup_{V \in \mathcal{W}} \text{ch}(V)$$

Put more informally, the children of V are all nodes to which there is a directed edge from V.

Familial Relations Specific to Directed Graphs

Definition (Descendents)

In a directed graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ the set of descendents of a node $V \in \mathcal{V}$ is defined to be:

$$\mathrm{de}(V) = \{Z \in \mathcal{V} : V \rightsquigarrow Z\}$$

The set of descendents of a set of nodes $\mathcal{W} \subseteq \mathcal{V}$ is defined to be:

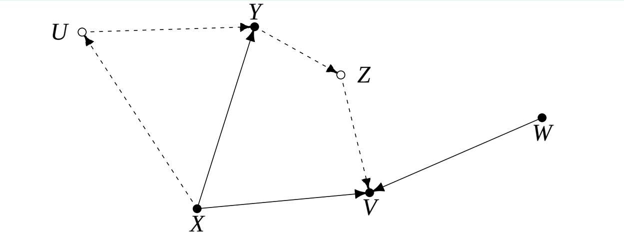
$$\mathrm{de}(\mathcal{W}) = \bigcup_{V \in \mathcal{W}} \mathrm{de}(V)$$

In words, the set of descendents of V consists of the vertices to which there exists a directed path from V.

Example

Example (Familial Relations)

Consider the graph depicted below.



Here:
$\mathrm{pa}(Z) = \{Y\}$, $\mathrm{ch}(Z) = \{V\}$, $\mathrm{an}(Z) = \{U, X, Y\}$, $\mathrm{de}(Z) = \{V\}$

and

$\mathrm{pa}(X) = \emptyset$, $\mathrm{ch}(X) = \{U, V, Y\}$, $\mathrm{an}(X) = \emptyset$,
$\mathrm{de}(X) = \{U, V, Y, Z\}$

Directed Acycylic Graphs (DAGs)

A directed graph that does not have cycles (i.e., $V \notin \text{an}(V)$ for all $V \in \mathcal{V}$) is said to be a directed acyclic graph (DAG).

Directed Acycylic Graphs (DAGs)

A directed graph that does not have cycles (i.e., $V \notin \mathrm{an}(V)$ for all $V \in \mathcal{V}$) is said to be a directed acyclic graph (DAG).

It is also possible to show that there is some ordering of the nodes of a DAG such that there is no edge from any node to any node that is earlier in the order.

Directed Acycylic Graphs (DAGs)

A directed graph that does not have cycles (i.e., $V \notin \mathrm{an}(V)$ for all $V \in \mathcal{V}$) is said to be a directed acyclic graph (DAG).

It is also possible to show that there is some ordering of the nodes of a DAG such that there is no edge from any node to any node that is earlier in the order.

DAGs are useful for a number of reasons:

Directed Acycylic Graphs (DAGs)

A directed graph that does not have cycles (i.e., $V \notin \mathrm{an}(V)$ for all $V \in \mathcal{V}$) is said to be a directed acyclic graph (DAG).

It is also possible to show that there is some ordering of the nodes of a DAG such that there is no edge from any node to any node that is earlier in the order.

DAGs are useful for a number of reasons:

- They permit a convenient factorization of the joint distribution of all of the random variables in $\mathcal{V}$

Directed Acycylic Graphs (DAGs)

A directed graph that does not have cycles (i.e., $V \notin \mathrm{an}(V)$ for all $V \in \mathcal{V}$) is said to be a directed acyclic graph (DAG).

It is also possible to show that there is some ordering of the nodes of a DAG such that there is no edge from any node to any node that is earlier in the order.

DAGs are useful for a number of reasons:

- They permit a convenient factorization of the joint distribution of all of the random variables in $\mathcal{V}$
- They allow one to build complicated joint distributions from simple parts

Directed Acycylic Graphs (DAGs)

A directed graph that does not have cycles (i.e., $V \notin \text{an}(V)$ for all $V \in \mathcal{V}$) is said to be a directed acyclic graph (DAG).

It is also possible to show that there is some ordering of the nodes of a DAG such that there is no edge from any node to any node that is earlier in the order.

DAGs are useful for a number of reasons:

- They permit a convenient factorization of the joint distribution of all of the random variables in $\mathcal{V}$
- They allow one to build complicated joint distributions from simple parts
- Relatedly, they can help one think about ways to achieve dimension reduction / data summarization

Directed Acycylic Graphs (DAGs)

A directed graph that does not have cycles (i.e., $V \notin \mathrm{an}(V)$ for all $V \in \mathcal{V}$) is said to be a directed acyclic graph (DAG).

It is also possible to show that there is some ordering of the nodes of a DAG such that there is no edge from any node to any node that is earlier in the order.

DAGs are useful for a number of reasons:

- They permit a convenient factorization of the joint distribution of all of the random variables in $\mathcal{V}$
- They allow one to build complicated joint distributions from simple parts
- Relatedly, they can help one think about ways to achieve dimension reduction / data summarization
- They are natural representation of systems evolving in time

Directed Acycylic Graphs (DAGs)

A directed graph that does not have cycles (i.e., $V \notin \text{an}(V)$ for all $V \in \mathcal{V}$) is said to be a directed acyclic graph (DAG).

It is also possible to show that there is some ordering of the nodes of a DAG such that there is no edge from any node to any node that is earlier in the order.

DAGs are useful for a number of reasons:

- They permit a convenient factorization of the joint distribution of all of the random variables in $\mathcal{V}$
- They allow one to build complicated joint distributions from simple parts
- Relatedly, they can help one think about ways to achieve dimension reduction / data summarization
- They are natural representation of systems evolving in time
- They can be given a causal interpretation

Using DAGs to Represent Complicated Joint Distributions

As an initial (simple) example, consider 3 random variables X, Y and Z. Unless otherwise stated, we will make no assumptions about these random variables or the relationships between these random variables.

Using DAGs to Represent Complicated Joint Distributions

As an initial (simple) example, consider 3 random variables X, Y and Z. Unless otherwise stated, we will make no assumptions about these random variables or the relationships between these random variables.

Using the product rule of basic probability, we can *always* write the joint density of X, Y and Z (regardless of its form) as:

$$p(x, y, z) = p(z|x, y)p(y|x)p(x)$$

Using DAGs to Represent Complicated Joint Distributions

As an initial (simple) example, consider 3 random variables X, Y and Z. Unless otherwise stated, we will make no assumptions about these random variables or the relationships between these random variables.

Using the product rule of basic probability, we can *always* write the joint density of X, Y and Z (regardless of its form) as:

$$p(x, y, z) = p(z|x, y)p(y|x)p(x)$$

Note that there are five other such factorizations that are possible.

Using DAGs to Represent Complicated Joint Distributions

Let's represent the factorization of a joint distribution with a DAG according to the following rules:

Using DAGs to Represent Complicated Joint Distributions

Let's represent the factorization of a joint distribution with a DAG according to the following rules:

1. Let the set of random variables under study (in this case $\{X, Y, Z\}$) be the set of nodes

Using DAGs to Represent Complicated Joint Distributions

Let's represent the factorization of a joint distribution with a DAG according to the following rules:

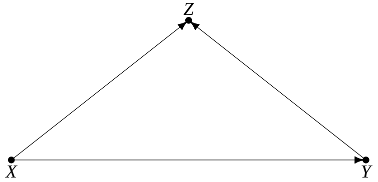
1. Let the set of random variables under study (in this case $\{X, Y, Z\}$) be the set of nodes
2. For each conditional density on the rhs of the factorization, add directed edges from each of the variables on the rhs of the conditioning bar to the variable on the lhs of the conditioning bar

Using DAGs to Represent Complicated Joint Distributions

Let's represent the factorization of a joint distribution with a DAG according to the following rules:

1. Let the set of random variables under study (in this case $\{X, Y, Z\}$) be the set of nodes
2. For each conditional density on the rhs of the factorization, add directed edges from each of the variables on the rhs of the conditioning bar to the variable on the lhs of the conditioning bar

For the 3 variable example above his gives rise to:



Note the relationship to $p(x, y, z) = p(z|x, y)p(y|x)p(x)$

Using DAGs to Represent Complicated Joint Distributions

It is easy to show, given our rules for forming graphs, that for a DAG with n vertices $V_1, \ldots, V_n$ one can write

$$p(v_1, \ldots, v_n) = \prod_{i=1}^{n} p(v_i | \mathrm{pa}(v_i))$$

Using DAGs to Represent Complicated Joint Distributions

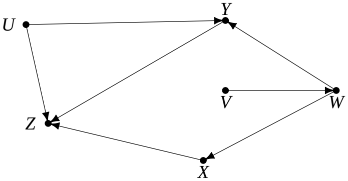
It is easy to show, given our rules for forming graphs, that for a DAG with n vertices $V_1, \ldots, V_n$ one can write

$$p(v_1, \ldots, v_n) = \prod_{i=1}^{n} p(v_i | \text{pa}(v_i))$$

This formula allows us to look at a graph and then write down the factorization that the graph implies.

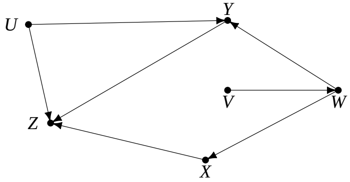
Using DAGs to Represent Complicated Joint Distributions

Let's look at a more complicated example that starts with the following graph.



Using DAGs to Represent Complicated Joint Distributions

Let's look at a more complicated example that starts with the following graph.

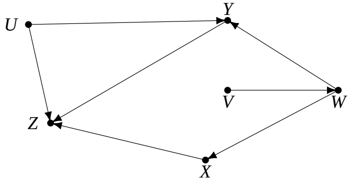


This is consistent with the factorization

$$p(u, v, w, x, y, z) = p(z|u, x, y)p(y|u, w)p(x|w)p(w|v)p(u)p(v)$$

Using DAGs to Represent Complicated Joint Distributions

Let's look at a more complicated example that starts with the following graph.



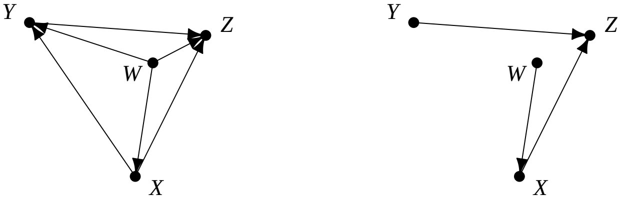
This is consistent with the factorization

$$p(u, v, w, x, y, z) = p(z|u, x, y)p(y|u, w)p(x|w)p(w|v)p(u)p(v)$$

Note that the graph embodies a lot of conditional independence assumptions.

Using DAGs to Represent Complicated Joint Distributions

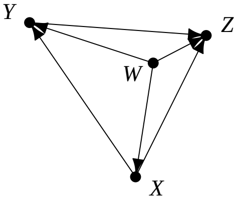
Question: Which graph below embodies more assumptions?



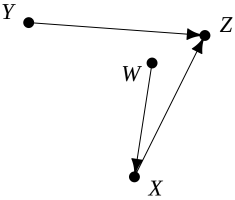
(a) (b)

Using DAGs to Represent Complicated Joint Distributions

Question: Which graph below embodies more assumptions?



(a)

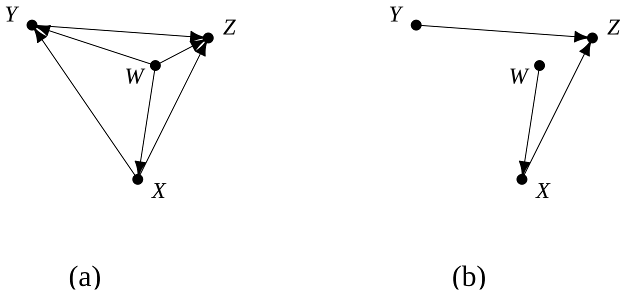


(b)

Answer: More assumptions are implied by (b).

Using DAGs to Represent Complicated Joint Distributions

Question: Which graph below embodies more assumptions?



Answer: More assumptions are implied by (b).

Missing edges correspond to conditional independence assumptions.

Using DAGs to Represent Complicated Joint Distributions

Returning to our three variable example, note that the factorization $p(x, y, z) = p(z|x, y)p(y|x)p(x)$ suggests a simple means (called the method of composition) of obtaining a sample from the joint distribution of (X, Y, Z):

1. Sample x from the marginal distribution of X

Using DAGs to Represent Complicated Joint Distributions

Returning to our three variable example, note that the factorization $p(x, y, z) = p(z|x, y)p(y|x)p(x)$ suggests a simple means (called the method of composition) of obtaining a sample from the joint distribution of (X, Y, Z):

1. Sample x from the marginal distribution of X
2. Sample y from the conditional distribution of Y given x

Using DAGs to Represent Complicated Joint Distributions

Returning to our three variable example, note that the factorization $p(x, y, z) = p(z|x, y)p(y|x)p(x)$ suggests a simple means (called the method of composition) of obtaining a sample from the joint distribution of (X, Y, Z):

1. Sample x from the marginal distribution of X
2. Sample y from the conditional distribution of Y given x
3. Sample z from the conditional distribution of Z given x and y

Using DAGs to Represent Complicated Joint Distributions

Returning to our three variable example, note that the factorization $p(x, y, z) = p(z|x, y)p(y|x)p(x)$ suggests a simple means (called the method of composition) of obtaining a sample from the joint distribution of (X, Y, Z):

1. Sample x from the marginal distribution of X
2. Sample y from the conditional distribution of Y given x
3. Sample z from the conditional distribution of Z given x and y
4. Return (x, y, z) as a draw from the joint distribution of (X, Y, Z)

Using DAGs to Represent Complicated Joint Distributions

Note that there is a graphical version of the method of composition that can be used to sample from the joint distribution of the variables governed by a DAG $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$.

Using DAGs to Represent Complicated Joint Distributions

Note that there is a graphical version of the method of composition that can be used to sample from the joint distribution of the variables governed by a DAG $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$.

This is called ancestral sampling.

Using DAGs to Represent Complicated Joint Distributions

Note that there is a graphical version of the method of composition that can be used to sample from the joint distribution of the variables governed by a DAG $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$.

This is called ancestral sampling.

1. Order the *n* variables in $\mathcal{V}$ so that no node has an edge to a lower numbered node.

Using DAGs to Represent Complicated Joint Distributions

Note that there is a graphical version of the method of composition that can be used to sample from the joint distribution of the variables governed by a DAG $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$.

This is called ancestral sampling.

1. Order the n variables in $\mathcal{V}$ so that no node has an edge to a lower numbered node.
2. Sample v_1 from the distribution with density $p(v_1)$

Using DAGs to Represent Complicated Joint Distributions

Note that there is a graphical version of the method of composition that can be used to sample from the joint distribution of the variables governed by a DAG $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$.

This is called ancestral sampling.

1. Order the n variables in $\mathcal{V}$ so that no node has an edge to a lower numbered node.
2. Sample v_1 from the distribution with density $p(v_1)$
3. For $i = 2, \ldots, n$

Using DAGs to Represent Complicated Joint Distributions

Note that there is a graphical version of the method of composition that can be used to sample from the joint distribution of the variables governed by a DAG $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$.

This is called ancestral sampling.

1. Order the n variables in $\mathcal{V}$ so that no node has an edge to a lower numbered node.
2. Sample v_1 from the distribution with density $p(v_1)$
3. For $i = 2, \ldots, n$
   - Sample v_i from the distribution with density $p(v_i | \text{pa}(v_i))$

Using DAGs to Represent Complicated Joint Distributions

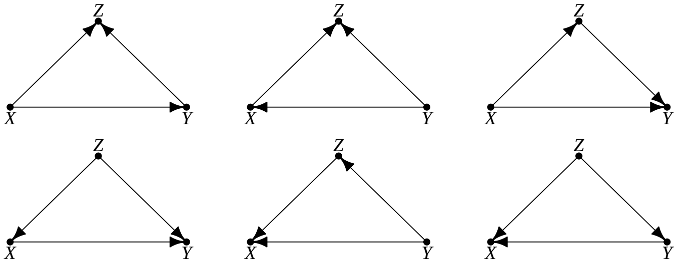
Note that there is a graphical version of the method of composition that can be used to sample from the joint distribution of the variables governed by a DAG $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$.

This is called ancestral sampling.

1. Order the n variables in $\mathcal{V}$ so that no node has an edge to a lower numbered node.
2. Sample v_1 from the distribution with density $p(v_1)$
3. For $i = 2, \ldots, n$
   - Sample v_i from the distribution with density $p(v_i | \text{pa}(v_i))$
4. Return $(v_1, \ldots, v_n)$

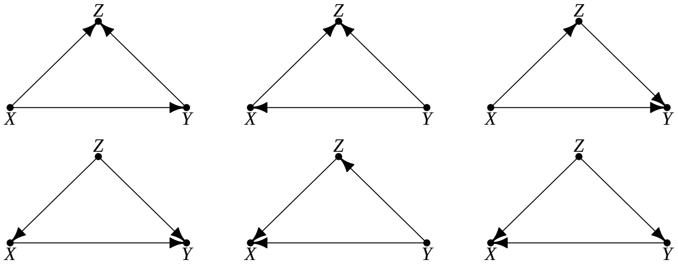
Observational Equivalence

As an aside, note that **all** of the following DAGs are consistent with any choice of $p(x, y, z)$.



Observational Equivalence

As an aside, note that all of the following DAGs are consistent with any choice of $p(x, y, z)$.



There is actually a general result here.

Observational Equivalence

Definition (Observational Equivalence)

Two DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ are said to be observationally equivalent if every probability distribution that is compatible with $\mathcal{G}_1$ is also compatible with $\mathcal{G}_2$.

Observational Equivalence

Definition (Observational Equivalence)

Two DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ are said to be observationally equivalent if every probability distribution that is compatible with $\mathcal{G}_1$ is also compatible with $\mathcal{G}_2$.

Definition (Skeleton)

The skeleton of a graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ is the object formed by removing all the arrowheads from the edges in $\mathcal{E}$.

Observational Equivalence

Definition (Observational Equivalence)

Two DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ are said to be observationally equivalent if every probability distribution that is compatible with $\mathcal{G}_1$ is also compatible with $\mathcal{G}_2$.

Definition (Skeleton)

The skeleton of a graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ is the object formed by removing all the arrowheads from the edges in $\mathcal{E}$.

Definition (*v*-Structure)

A *v*-structure in a graph $\mathcal{G}$ consists of two converging arrows whose tails are not connected by an arrow.

Observational Equivalence

Definition (Observational Equivalence)

Two DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ are said to be observationally equivalent if every probability distribution that is compatible with $\mathcal{G}_1$ is also compatible with $\mathcal{G}_2$.

Definition (Skeleton)

The skeleton of a graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ is the object formed by removing all the arrowheads from the edges in $\mathcal{E}$.

Definition (*v*-Structure)

A *v*-structure in a graph $\mathcal{G}$ consists of two converging arrows whose tails are not connected by an arrow.

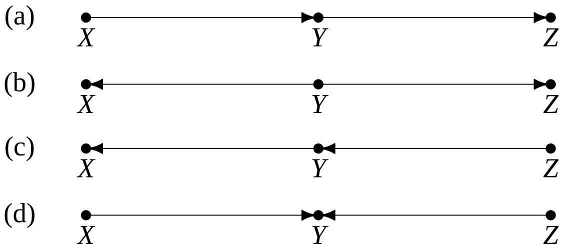
Theorem (Observational Equivalence (Verma and Pearl))

Two DAGs are observationally equivalent if and only if they have the same skeletons and the same sets of v-structures.

Example

Example (Observational Equivalence)

Consider the four graphs below.



Example

Example (Observational Equivalence)

Consider the four graphs below.

(a) $X \rightarrow Y \rightarrow Z$

(b) $X \leftarrow Y \rightarrow Z$

(c) $X \leftarrow Y \leftarrow Z$

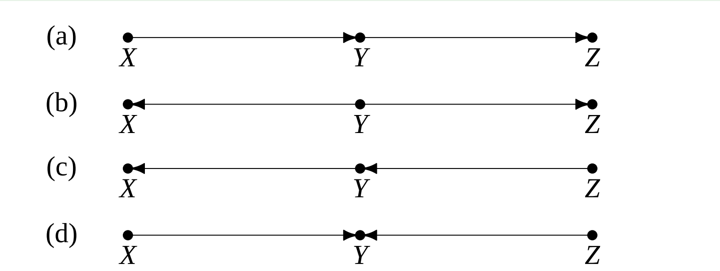
(d) $X \rightarrow Y \leftarrow Z$

Here we see that graphs (a), (b), and (c) are all observationally equivalent. *Graph (d) is not observationally equivalent to any of the other graphs.*

Example

Example (Observational Equivalence)

Consider the four graphs below.



Here we see that graphs (a), (b), and (c) are all observationally equivalent. *Graph (d) is not observationally equivalent to any of the other graphs.*

Note that graphs (a), (b), and (c) imply very different causal relationships.

Observational Equivalence

Some things to think about:

- A DAG implies a particular factorization but an unfactorized joint distribution implies a wide range of DAGs.
- Because a DAG gives a recipe for generating data from a joint distribution it is tempting to think of a DAG (by itself) in causal terms. This is a mistake because of the point immediately above.
- Interpreting a DAG causally can be reasonable—we'll talk about this extensively in later weeks—but it does require additional assumptions.

Conditional Independence

Definition (Conditional Independence)

Let $\mathcal{V} = \{V_1, \ldots, V_n\}$ be a finite set of random variables and let X, Y, and Z denote three subsets of $\mathcal{V}$. If

$$p(x|y,z) = p(x|z) \quad \text{whenever} \quad p(y,z) > 0$$

we say that X is conditionally independent of Y given Z and write

$$[X \perp\!\!\!\perp Y | Z]$$

Conditional Independence

Definition (Conditional Independence)

Let $\mathcal{V} = \{V_1, \ldots, V_n\}$ be a finite set of random variables and let X, Y, and Z denote three subsets of $\mathcal{V}$. If

$$p(x|y,z) = p(x|z) \quad \text{whenever} \quad p(y,z) > 0$$

we say that X is conditionally independent of Y given Z and write

$$[X \perp\!\!\!\perp Y | Z]$$

In words, knowing z and y provides no more information about X than knowing just z.

Conditional Independence

Definition (Conditional Independence)

Let $\mathcal{V} = \{V_1, \ldots, V_n\}$ be a finite set of random variables and let X, Y, and Z denote three subsets of $\mathcal{V}$. If

$$p(x|y,z) = p(x|z) \quad \text{whenever} \quad p(y,z) > 0$$

we say that X is conditionally independent of Y given Z and write

$$[X \perp\!\!\!\perp Y | Z]$$

In words, knowing z and y provides no more information about X than knowing just z.

Note that the equality in the definition above has to hold for all values of x, y and z for which $p(y,z) > 0$.

Conditional Independence

Definition (Conditional Independence)

Let $\mathcal{V} = \{V_1, \ldots, V_n\}$ be a finite set of random variables and let X, Y, and Z denote three subsets of $\mathcal{V}$. If

$$p(x|y,z) = p(x|z) \quad \text{whenever} \quad p(y,z) > 0$$

we say that X is conditionally independent of Y given Z and write

$$[X \perp\!\!\!\perp Y|Z]$$

In words, knowing z and y provides no more information about X than knowing just z.

Note that the equality in the definition above has to hold for all values of x, y and z for which $p(y,z) > 0$.

It is easy to see that $[X \perp\!\!\!\perp Y|Z]$ also implies that

$$p(x,y|z) = p(x|z)p(y|z)$$

Conditional Independence

The following are some properties that hold for conditionally independent random variables.

Property (Symmetry)

$$[X \perp\!\!\!\perp Y | Z] \implies [Y \perp\!\!\!\perp X | Z]$$

Property (Decomposition)

$$[X \perp\!\!\!\perp (Y, W) | Z] \implies [X \perp\!\!\!\perp Y | Z]$$

Property (Weak Union)

$$[X \perp\!\!\!\perp (Y, W) | Z] \implies [X \perp\!\!\!\perp Y | (Z, W)]$$

Property (Contraction)

$$[X \perp\!\!\!\perp Y | Z] \;\&\; [X \perp\!\!\!\perp W | (Z, Y)] \implies [X \perp\!\!\!\perp (Y, W) | Z]$$

The Importance of Conditional Independence

Conditional independence is fundamental to almost all areas of probability, statistics, and data analysis.

- causal inference
- Bayesian inference
- data summarization
- Markov chains
- sufficiency
- ancillarity
- identification
- etc.

See Dawid (1979). “Conditional Independence in Statistical Theory” *JRSS B*. for additional information.

d-Separation

Definition (d-Separation)

Let $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ be a DAG and X, Y, and Z be disjoint subsets of $\mathcal{V}$.

d-Separation

Definition (d-Separation)

Let $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ be a DAG and X, Y, and Z be disjoint subsets of $\mathcal{V}$.

X is said to be *d-separated* from Y by Z in $\mathcal{G}$ (written $[X \perp\!\!\!\perp Y | Z]_{\mathcal{G}}$) if and only if Z *blocks* every path from a vertex in X to a vertex in Y.

d-Separation

Definition (d-Separation)

Let $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ be a DAG and X, Y, and Z be disjoint subsets of $\mathcal{V}$.

X is said to be *d*-separated from Y by Z in $\mathcal{G}$ (written $[X \perp\!\!\!\perp Y | Z]_{\mathcal{G}}$) if and only if Z blocks every path from a vertex in X to a vertex in Y.

A path p is said to be blocked by a set of vertices Z if and only if at least one of the following conditions hold:

1. p contains a chain structure $a \to b \to c$ or a fork structure $a \leftarrow b \to c$ where the node b is in Z
2. p contains a collider structure $a \to b \leftarrow c$ where b is *not* in Z and no descendent of b is in Z

d-Separation

Definition (d-Separation)

Let $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ be a DAG and X, Y, and Z be disjoint subsets of $\mathcal{V}$.

X is said to be d-separated from Y by Z in $\mathcal{G}$ (written $[X \perp\!\!\!\perp Y | Z]_{\mathcal{G}}$) if and only if Z blocks every path from a vertex in X to a vertex in Y.

A path p is said to be blocked by a set of vertices Z if and only if at least one of the following conditions hold:

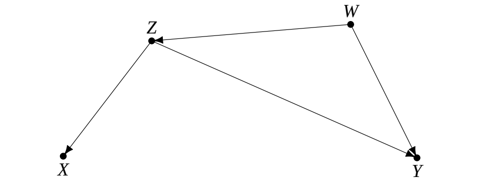
1. p contains a chain structure $a \rightarrow b \rightarrow c$ or a fork structure $a \leftarrow b \rightarrow c$ where the node b is in Z
2. p contains a collider structure $a \rightarrow b \leftarrow c$ where b is *not* in Z and no descendent of b is in Z

If X is *not* d-separated from Y by Z we say that X is d-connected to Y by Z.

Example

Example (d-Separation)

Consider the graph below.



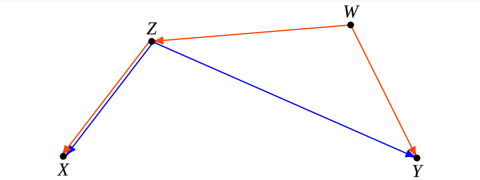
Which sets of variables (if any) d-separate X from Y?

Example

Let's draw the paths between X and Y and check to see if various conditioning sets block these paths.

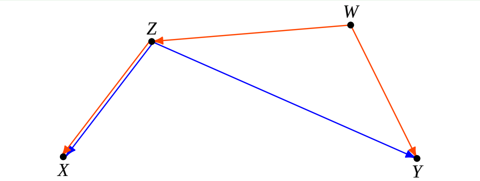
Example

Let's draw the paths between X and Y and check to see if various conditioning sets block these paths.



Example

Let's draw the paths between X and Y and check to see if various conditioning sets block these paths.

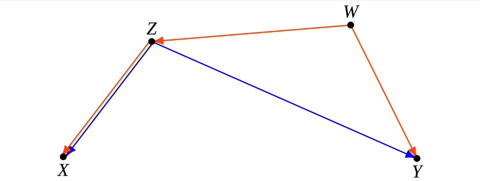


Recall that a path p is said to be blocked by a set of vertices U if and only if at least one of the following conditions hold:

1. p contains a chain $a \rightarrow b \rightarrow c$ or a fork $a \leftarrow b \rightarrow c$ where the node b is in U
2. p contains a collider $a \rightarrow b \leftarrow c$ where b is *not* in U and no descendent of b is in U

Example

Let's draw the paths between X and Y and check to see if various conditioning sets block these paths.



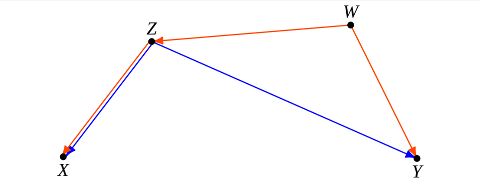
Recall that a path p is said to be blocked by a set of vertices U if and only if at least one of the following conditions hold:

1. p contains a chain $a \to b \to c$ or a fork $a \leftarrow b \to c$ where the node b is in U
2. p contains a collider $a \to b \leftarrow c$ where b is *not* in U and no descendent of b is in U

Is $\{\emptyset\}$ a sufficient conditioning set?

Example

Let's draw the paths between X and Y and check to see if various conditioning sets block these paths.



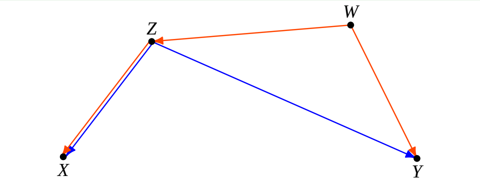
Recall that a path p is said to be blocked by a set of vertices U if and only if at least one of the following conditions hold:

1. p contains a chain $a \to b \to c$ or a fork $a \leftarrow b \to c$ where the node b is in U
2. p contains a collider $a \to b \leftarrow c$ where b is *not* in U and no descendent of b is in U

Is $\{\emptyset\}$ a sufficient conditioning set? No.

Example

Let's draw the paths between X and Y and check to see if various conditioning sets block these paths.



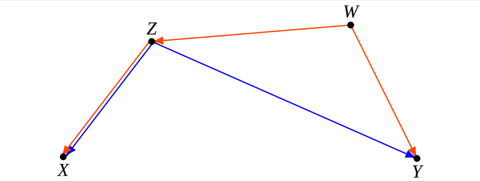
Recall that a path p is said to be blocked by a set of vertices U if and only if at least one of the following conditions hold:

1. p contains a chain $a \to b \to c$ or a fork $a \leftarrow b \to c$ where the node b is in U
2. p contains a collider $a \to b \leftarrow c$ where b is *not* in U and no descendent of b is in U

Is $\{\emptyset\}$ a sufficient conditioning set? No.
Is $\{W\}$ a sufficient conditioning set?

Example

Let's draw the paths between X and Y and check to see if various conditioning sets block these paths.



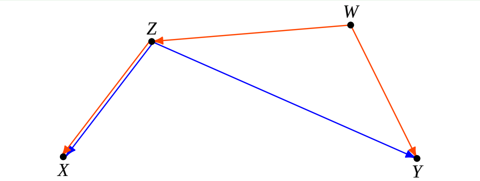
Recall that a path p is said to be blocked by a set of vertices U if and only if at least one of the following conditions hold:

1. p contains a chain $a \to b \to c$ or a fork $a \leftarrow b \to c$ where the node b is in U
2. p contains a collider $a \to b \leftarrow c$ where b is *not* in U and no descendent of b is in U

Is $\{\emptyset\}$ a sufficient conditioning set? No.
Is $\{W\}$ a sufficient conditioning set? No.

Example

Let's draw the paths between X and Y and check to see if various conditioning sets block these paths.



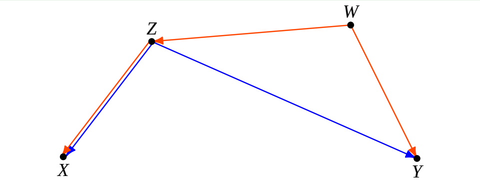
Recall that a path p is said to be blocked by a set of vertices U if and only if at least one of the following conditions hold:

1. p contains a chain $a \to b \to c$ or a fork $a \leftarrow b \to c$ where the node b is in U
2. p contains a collider $a \to b \leftarrow c$ where b is *not* in U and no descendent of b is in U

Is $\{\emptyset\}$ a sufficient conditioning set? No.
Is $\{W\}$ a sufficient conditioning set? No.
Is $\{Z\}$ a sufficient conditioning set?

Example

Let's draw the paths between X and Y and check to see if various conditioning sets block these paths.



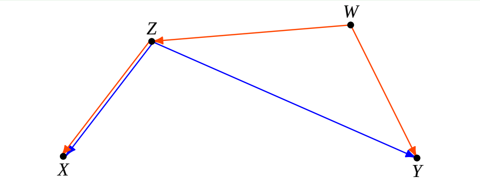
Recall that a path p is said to be blocked by a set of vertices U if and only if at least one of the following conditions hold:

1. p contains a chain $a \to b \to c$ or a fork $a \leftarrow b \to c$ where the node b is in U
2. p contains a collider $a \to b \leftarrow c$ where b is *not* in U and no descendent of b is in U

Is $\{\emptyset\}$ a sufficient conditioning set? No.
Is $\{W\}$ a sufficient conditioning set? No.
Is $\{Z\}$ a sufficient conditioning set? Yes.

Example

Let's draw the paths between X and Y and check to see if various conditioning sets block these paths.



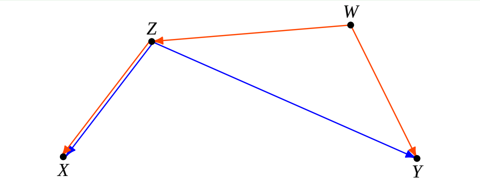
Recall that a path p is said to be blocked by a set of vertices U if and only if at least one of the following conditions hold:

1. p contains a chain $a \to b \to c$ or a fork $a \leftarrow b \to c$ where the node b is in U
2. p contains a collider $a \to b \leftarrow c$ where b is *not* in U and no descendent of b is in U

Is $\{\emptyset\}$ a sufficient conditioning set? No.
Is $\{W\}$ a sufficient conditioning set? No.
Is $\{Z\}$ a sufficient conditioning set? Yes.
Is $\{W, Z\}$ a sufficient conditioning set?

Example

Let's draw the paths between X and Y and check to see if various conditioning sets block these paths.



Recall that a path p is said to be blocked by a set of vertices U if and only if at least one of the following conditions hold:

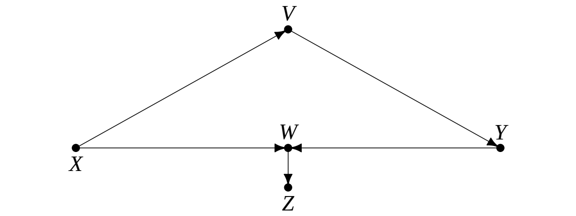
1. p contains a chain $a \rightarrow b \rightarrow c$ or a fork $a \leftarrow b \rightarrow c$ where the node b is in U
2. p contains a collider $a \rightarrow b \leftarrow c$ where b is *not* in U and no descendent of b is in U

Is $\{\emptyset\}$ a sufficient conditioning set? No.
Is $\{W\}$ a sufficient conditioning set? No.
Is $\{Z\}$ a sufficient conditioning set? Yes.
Is $\{W, Z\}$ a sufficient conditioning set? Yes.

Example

Example (d-Separation)

Consider the graph below.



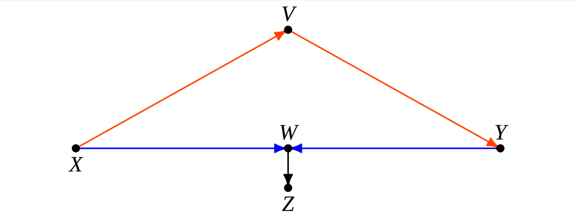
Which sets of variables (if any) d-separate X from Y?

Example

Let's draw the paths between X and Y and check to see if various conditioning sets block these paths.

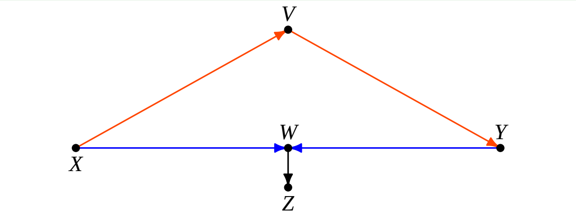
Example

Let's draw the paths between X and Y and check to see if various conditioning sets block these paths.



Example

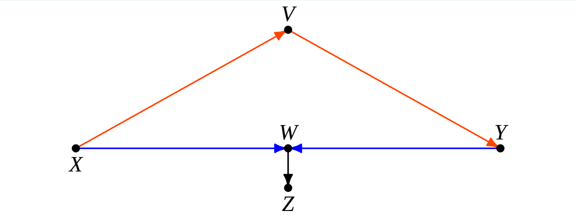
Let's draw the paths between X and Y and check to see if various conditioning sets block these paths.



Is $\{\emptyset\}$ a sufficient conditioning set?

Example

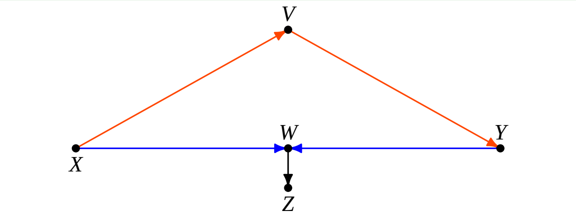
Let's draw the paths between X and Y and check to see if various conditioning sets block these paths.



Is $\{\emptyset\}$ a sufficient conditioning set? No.

Example

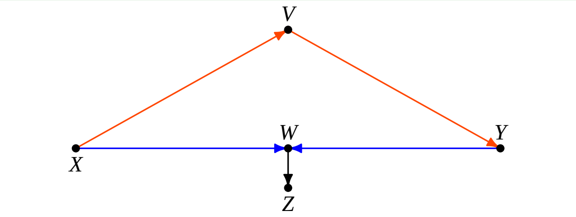
Let's draw the paths between X and Y and check to see if various conditioning sets block these paths.



Is $\{\emptyset\}$ a sufficient conditioning set? No.
Is $\{V\}$ a sufficient conditioning set?

Example

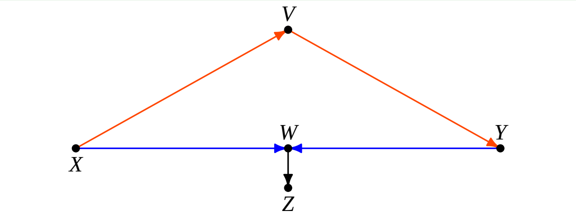
Let's draw the paths between X and Y and check to see if various conditioning sets block these paths.



Is $\{\emptyset\}$ a sufficient conditioning set? No.
Is $\{V\}$ a sufficient conditioning set? Yes.

Example

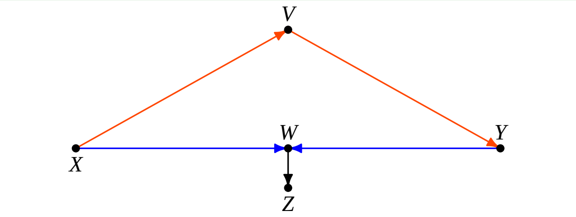
Let's draw the paths between X and Y and check to see if various conditioning sets block these paths.



Is $\{\emptyset\}$ a sufficient conditioning set? No.
Is $\{V\}$ a sufficient conditioning set? Yes.
Is $\{W\}$ a sufficient conditioning set?

Example

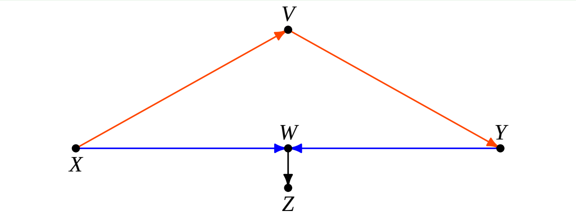
Let's draw the paths between X and Y and check to see if various conditioning sets block these paths.



Is $\{\emptyset\}$ a sufficient conditioning set? No.
Is $\{V\}$ a sufficient conditioning set? Yes.
Is $\{W\}$ a sufficient conditioning set? No.

Example

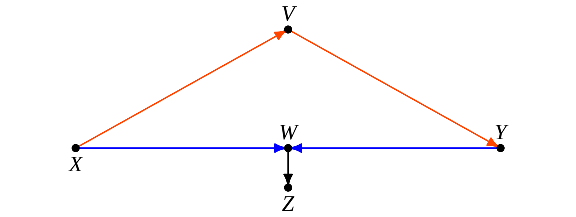
Let's draw the paths between X and Y and check to see if various conditioning sets block these paths.



Is $\{\emptyset\}$ a sufficient conditioning set? No.
Is $\{V\}$ a sufficient conditioning set? Yes.
Is $\{W\}$ a sufficient conditioning set? No.
Is $\{Z\}$ a sufficient conditioning set?

Example

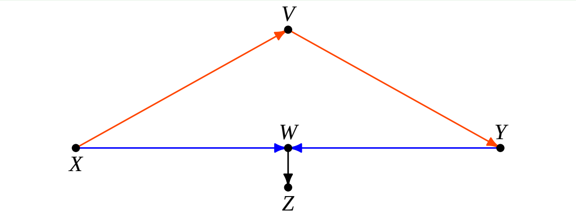
Let's draw the paths between X and Y and check to see if various conditioning sets block these paths.



Is $\{\emptyset\}$ a sufficient conditioning set? No.
Is $\{V\}$ a sufficient conditioning set? Yes.
Is $\{W\}$ a sufficient conditioning set? No.
Is $\{Z\}$ a sufficient conditioning set? No.

Example

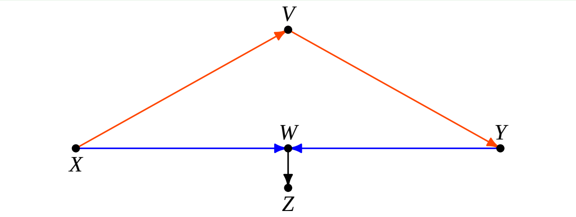
Let's draw the paths between X and Y and check to see if various conditioning sets block these paths.



Is $\{\emptyset\}$ a sufficient conditioning set? No.
Is $\{V\}$ a sufficient conditioning set? Yes.
Is $\{W\}$ a sufficient conditioning set? No.
Is $\{Z\}$ a sufficient conditioning set? No.
Is $\{V, W\}$ a sufficient conditioning set?

Example

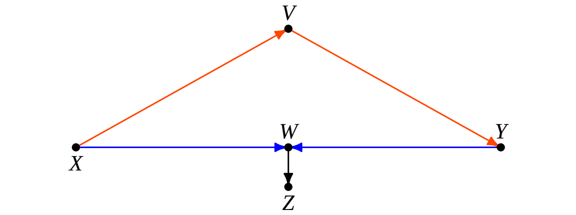
Let's draw the paths between X and Y and check to see if various conditioning sets block these paths.



Is $\{\emptyset\}$ a sufficient conditioning set? No.
Is $\{V\}$ a sufficient conditioning set? Yes.
Is $\{W\}$ a sufficient conditioning set? No.
Is $\{Z\}$ a sufficient conditioning set? No.
Is $\{V, W\}$ a sufficient conditioning set? No.

Example

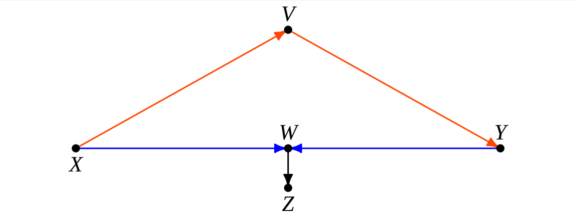
Let's draw the paths between X and Y and check to see if various conditioning sets block these paths.



Is $\{\emptyset\}$ a sufficient conditioning set? No.
Is $\{V\}$ a sufficient conditioning set? Yes.
Is $\{W\}$ a sufficient conditioning set? No.
Is $\{Z\}$ a sufficient conditioning set? No.
Is $\{V, W\}$ a sufficient conditioning set? No.
Is $\{V, Z\}$ a sufficient conditioning set?

Example

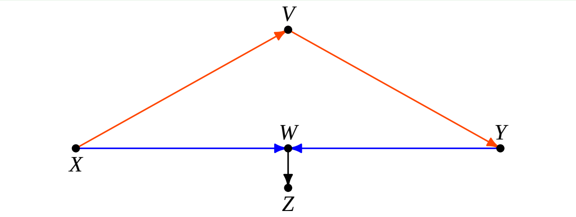
Let's draw the paths between X and Y and check to see if various conditioning sets block these paths.



Is $\{\emptyset\}$ a sufficient conditioning set? No.
Is $\{V\}$ a sufficient conditioning set? Yes.
Is $\{W\}$ a sufficient conditioning set? No.
Is $\{Z\}$ a sufficient conditioning set? No.
Is $\{V, W\}$ a sufficient conditioning set? No.
Is $\{V, Z\}$ a sufficient conditioning set? No.

Example

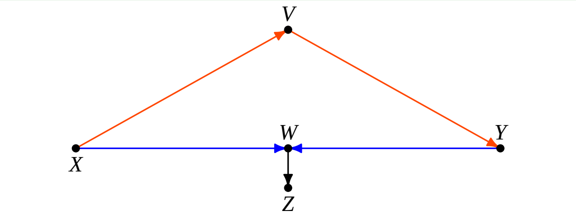
Let's draw the paths between X and Y and check to see if various conditioning sets block these paths.



Is $\{\emptyset\}$ a sufficient conditioning set? No.
Is $\{V\}$ a sufficient conditioning set? Yes.
Is $\{W\}$ a sufficient conditioning set? No.
Is $\{Z\}$ a sufficient conditioning set? No.
Is $\{V, W\}$ a sufficient conditioning set? No.
Is $\{V, Z\}$ a sufficient conditioning set? No.
Is $\{V, W, Z\}$ a sufficient conditioning set?

Example

Let's draw the paths between X and Y and check to see if various conditioning sets block these paths.



Is $\{\emptyset\}$ a sufficient conditioning set? No.
Is $\{V\}$ a sufficient conditioning set? Yes.
Is $\{W\}$ a sufficient conditioning set? No.
Is $\{Z\}$ a sufficient conditioning set? No.
Is $\{V, W\}$ a sufficient conditioning set? No.
Is $\{V, Z\}$ a sufficient conditioning set? No.
Is $\{V, W, Z\}$ a sufficient conditioning set? No.

Probabilistic Implications of *d*-Separation

Theorem (Probabilistic Implications of *d*-Separation (Verma and Pearl))

Let $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ be a DAG and X, Y, and Z be disjoint subsets of $\mathcal{V}$.

If Z d-separates X from Y in $\mathcal{G}$, then X is conditionally independent of Y given Z in every distribution compatible with $\mathcal{G}$.

Conversely, if X and Y are not d-separated by Z in $\mathcal{G}$, then X and Y are conditionally dependent given Z in at least one distribution compatible with $\mathcal{G}$.

Example

Example (d-Separation Implies Conditional Independence)

Consider a joint density that factorizes as:

$$p(v, w, x, y, z) = p(x|v, z)p(v|w)p(w|y, z)p(y)p(z)$$

Example

Example (*d*-Separation Implies Conditional Independence)

Consider a joint density that factorizes as:

$$p(v, w, x, y, z) = p(x|v, z)p(v|w)p(w|y, z)p(y)p(z)$$

Questions:
Is $[X \perp\!\!\!\perp Y | V]$ in this distribution?
Is $[X \perp\!\!\!\perp Y | W]$ in this distribution?
Is $[X \perp\!\!\!\perp Y | Z]$ in this distribution?
Is $[X \perp\!\!\!\perp Y | (V, Z)]$ in this distribution?

Example

Example (d-Separation Implies Conditional Independence)

Consider a joint density that factorizes as:

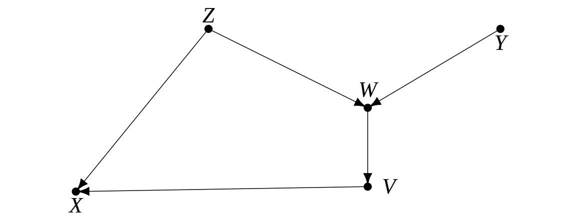
$$p(v, w, x, y, z) = p(x|v, z)p(v|w)p(w|y, z)p(y)p(z)$$

Questions:
Is $[X \perp\!\!\!\perp Y | V]$ in this distribution?
Is $[X \perp\!\!\!\perp Y | W]$ in this distribution?
Is $[X \perp\!\!\!\perp Y | Z]$ in this distribution?
Is $[X \perp\!\!\!\perp Y | (V, Z)]$ in this distribution?

Difficult to say without a lot of tedious calculations.

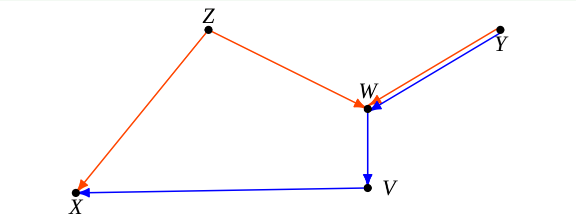
Example

Unless we draw the associated graph:



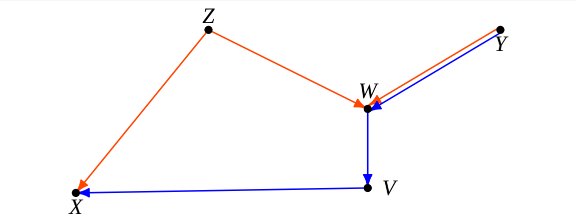
Example

And highlight the paths from X to Y



Example

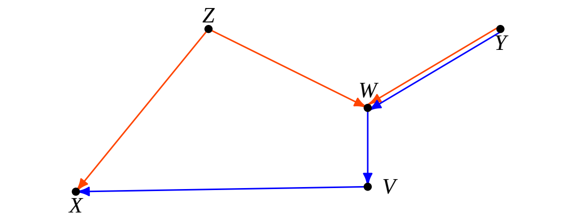
And highlight the paths from X to Y



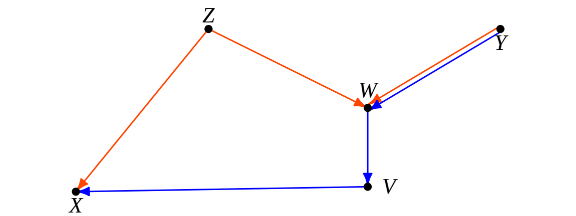
Recall that a path p is said to be blocked by a set of vertices U if and only if at least one of the following conditions hold:

1. p contains a chain $a \to b \to c$ or a fork $a \leftarrow b \to c$ where the node b is in U
2. p contains a collider $a \to b \leftarrow c$ where b is *not* in U and no descendent of b is in U

Example

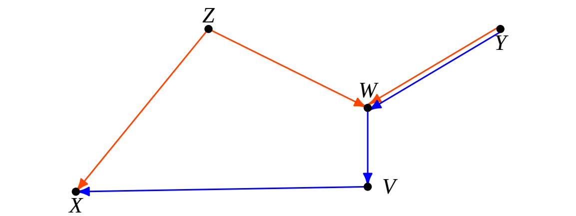


Example



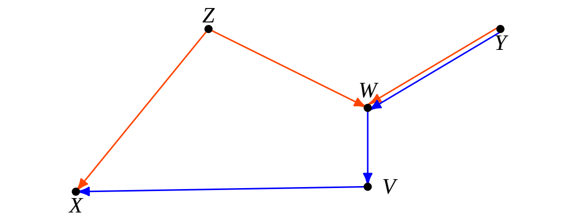
This implies that:
$[X \not\perp\!\!\!\perp Y \mid V]_{\mathcal{G}}$

Example



This implies that:
$[X \not\!\perp\!\!\!\perp Y | V]_{\mathcal{G}}$
$[X \not\!\perp\!\!\!\perp Y | W]_{\mathcal{G}}$

Example



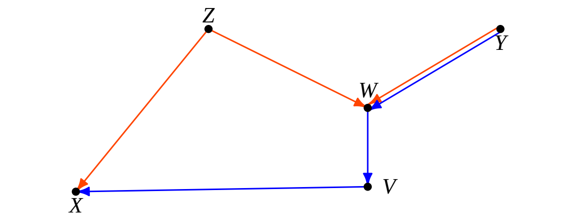
This implies that:

$[X \not\perp\!\!\!\perp Y | V]_{\mathcal{G}}$

$[X \not\perp\!\!\!\perp Y | W]_{\mathcal{G}}$

$[X \not\perp\!\!\!\perp Y | Z]_{\mathcal{G}}$

Example



This implies that:
$[X \not\perp\!\!\!\perp Y | V]_{\mathcal{G}}$
$[X \not\perp\!\!\!\perp Y | W]_{\mathcal{G}}$
$[X \not\perp\!\!\!\perp Y | Z]_{\mathcal{G}}$
and
$[X \perp\!\!\!\perp Y | (V, Z)]_{\mathcal{G}}$

Example

Example (d-Connection Does Not Imply Conditional Dependence in *All* Distributions Compatible with $\mathcal{G}$)

Consider the following system:

$$X = U_1$$

$$\begin{aligned} Z &= \alpha X + U_2 \\ &= \alpha U_1 + U_2 \end{aligned}$$

$$\begin{aligned} Y &= \gamma X + \beta Z + U_3 \\ &= \gamma U_1 + \beta(\alpha U_1 + U_2) + U_3 \\ &= (\gamma + \alpha\beta) U_1 + \beta U_2 + U_3 \end{aligned}$$

where it is assumed that $(U_1, U_2, U_3)' \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

Example

We can write this in matrix notation as:

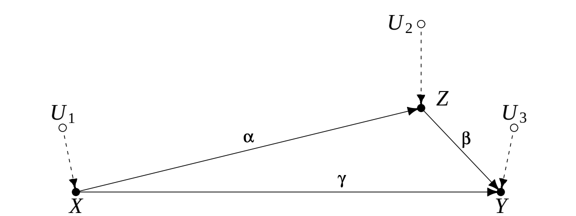
$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ (\gamma + \alpha\beta) & \beta & 1 \\ \alpha & 1 & 0 \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix}$$

Example

We can write this in matrix notation as:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ (\gamma + \alpha\beta) & \beta & 1 \\ \alpha & 1 & 0 \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix}$$

and depict it with the following DAG:

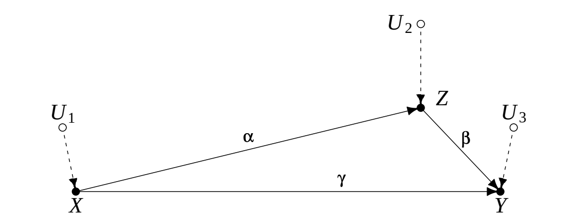


Example

We can write this in matrix notation as:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ (\gamma + \alpha\beta) & \beta & 1 \\ \alpha & 1 & 0 \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix}$$

and depict it with the following DAG:



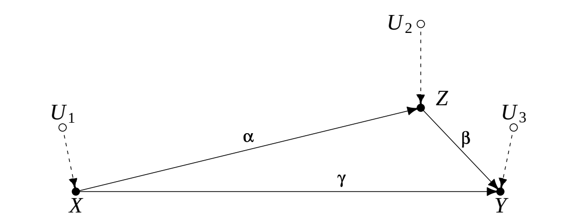
Here we see that Z does not d-separate X from Y – in other words, X and Y are d-connected given Z.

Example

We can write this in matrix notation as:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ (\gamma + \alpha\beta) & \beta & 1 \\ \alpha & 1 & 0 \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix}$$

and depict it with the following DAG:



Here we see that Z does not d-separate X from Y– in other words, X and Y are d-connected given Z.

Are they always conditionally dependent given Z?

Example

Because X, Y, and Z are linear functions of Gaussian random variables we know that their joint distribution is also Gaussian with variance-covariance matrix $\mathbf{\Sigma}$:

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & 0 & 0 \\ (\gamma + \alpha\beta) & \beta & 1 \\ \alpha & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & (\gamma + \alpha\beta) & \alpha \\ 0 & \beta & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & (\gamma + \alpha\beta) & \alpha \\ (\gamma + \alpha\beta) & \left[(\gamma + \alpha\beta)^2 + \beta^2 + 1\right] & [\alpha(\gamma + \alpha\beta) + \beta] \\ \alpha & [\alpha(\gamma + \alpha\beta) + \beta] & (\alpha^2 + 1) \end{pmatrix}$$

Example

Now that we know the joint distribution of X, Y, and Z we can use standard results from multivariate normal theory to calculate the conditional distribution of $(X, Y)|Z$. We know that this is also Gaussian with variance-covariance matrix $\mathbf{\Sigma}_{XY|Z}$:

$$\mathbf{\Sigma}_{XY|Z} = \begin{pmatrix} 1 & (\gamma + \alpha\beta) \\ (\gamma + \alpha\beta) & \left[(\gamma + \alpha\beta)^2 + \beta^2 + 1\right] \end{pmatrix} - \frac{1}{\alpha^2 + 1} \begin{pmatrix} \alpha \\ [\alpha(\gamma + \alpha\beta) + \beta] \end{pmatrix} \begin{pmatrix} \alpha & [\alpha(\gamma + \alpha\beta) + \beta] \end{pmatrix}$$

Example

Now that we know the joint distribution of X, Y, and Z we can use standard results from multivariate normal theory to calculate the conditional distribution of $(X, Y)|Z$. We know that this is also Gaussian with variance-covariance matrix $\mathbf{\Sigma}_{XY|Z}$:

$$\mathbf{\Sigma}_{XY|Z} = \begin{pmatrix} 1 & (\gamma + \alpha\beta) \\ (\gamma + \alpha\beta) & [(\gamma + \alpha\beta)^2 + \beta^2 + 1] \end{pmatrix} - \frac{1}{\alpha^2 + 1} \begin{pmatrix} \alpha \\ [\alpha(\gamma + \alpha\beta) + \beta] \end{pmatrix} \begin{pmatrix} \alpha & [\alpha(\gamma + \alpha\beta) + \beta] \end{pmatrix}$$

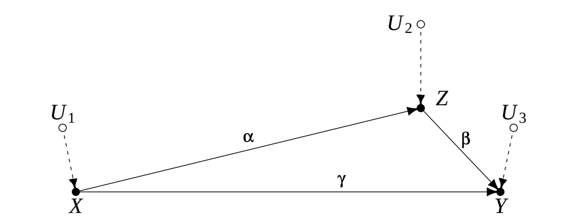
After multiplying this out and simplifying we see that

$$\mathrm{Cov}(X, Y|Z) = \frac{\gamma}{\alpha^2 + 1}$$

Thus, when $\gamma = 0$ $\mathrm{Cov}(X, Y|Z) = 0$ and because $p(x, y|z)$ is Gaussian we know that $[X \perp\!\!\!\perp Y|Z]$ when $\gamma = 0$ — even though $[X \not\perp\!\!\!\perp Y|Z]_{\mathcal{G}}$.

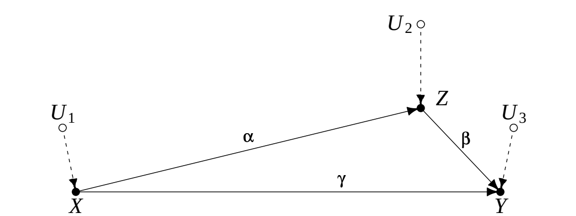
Example

It turns out that because of the linear Gaussian nature of this system we can reach the same result by looking at the edge coefficients on the graph.



Example

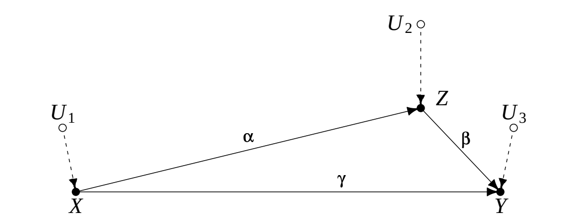
It turns out that because of the linear Gaussian nature of this system we can reach the same result by looking at the edge coefficients on the graph.



The linearity of the system implies that the coefficients on X and Z in a regression of Y on X and Z will be γ and β respectively.

Example

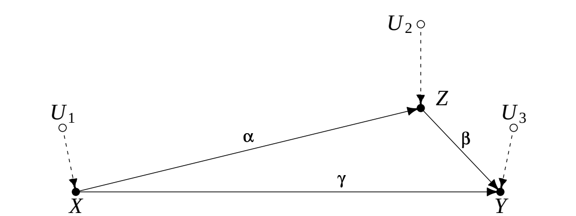
It turns out that because of the linear Gaussian nature of this system we can reach the same result by looking at the edge coefficients on the graph.



The linearity of the system implies that the coefficients on X and Z in a regression of Y on X and Z will be γ and β respectively.
Standard results from the linear model tell us that $\mathrm{Cov}(X, Y|Z)$ will be 0 when $\gamma = 0$.

Example

It turns out that because of the linear Gaussian nature of this system we can reach the same result by looking at the edge coefficients on the graph.



The linearity of the system implies that the coefficients on X and Z in a regression of Y on X and Z will be γ and β respectively.
Standard results from the linear model tell us that $\text{Cov}(X, Y|Z)$ will be 0 when $\gamma = 0$.
Because the joint distribution of (X, Y, Z) is Gaussian we know that $\text{Cov}(X, Y|Z) = 0 \implies [X \perp\!\!\!\perp Y|Z]$.

Faithfulness

Conditional dependence does not follow from *d*-connection in the example immediately above because the graph is not faithful to the distribution of (X, Y, Z).

Faithfulness

Conditional dependence does not follow from d-connection in the example immediately above because the graph is not faithful to the distribution of (X, Y, Z).

Definition (Faithfulness)

A DAG $\mathcal{G}$ and a distribution P are faithful to each other if and only if all conditional independence relations true in P (and only those conditional independence relations) are represented by the edge structure of $\mathcal{G}$.

Faithfulness

Conditional dependence does not follow from *d*-connection in the example immediately above because the graph is not faithful to the distribution of (X, Y, Z).

Definition (Faithfulness)

A DAG $\mathcal{G}$ and a distribution P are faithful to each other if and only if all conditional independence relations true in P (and only those conditional independence relations) are represented by the edge structure of $\mathcal{G}$.

Faithfulness can be violated as in the previous example where an edge exists but the effect is 0 or when non-zero effects cancel each other out. Deterministic relationships among variables can also cause problems.

Faithfulness

Conditional dependence does not follow from d-connection in the example immediately above because the graph is not faithful to the distribution of (X, Y, Z).

Definition (Faithfulness)

A DAG $\mathcal{G}$ and a distribution P are faithful to each other if and only if all conditional independence relations true in P (and only those conditional independence relations) are represented by the edge structure of $\mathcal{G}$.

Faithfulness can be violated as in the previous example where an edge exists but the effect is 0 or when non-zero effects cancel each other out. Deterministic relationships among variables can also cause problems.

Violations of faithulness tend to be knife-edge situations.