Title: Causal Inference Lecture - 230313

Speakers: Robert Spekkens

Collection: Causal Inference: Classical and Quantum

Date: March 13, 2023 - 10:00 AM

URL: https://pirsa.org/23030071

Abstract: zoom link: https://pitp.zoom.us/j/94143784665?pwd=VFJpajVIMEtvYmRabFYzYnNRSVAvZz09

What distributions are compatible with a given causal structure?

Causal structure ⇨ Constraint on observed probability distribution

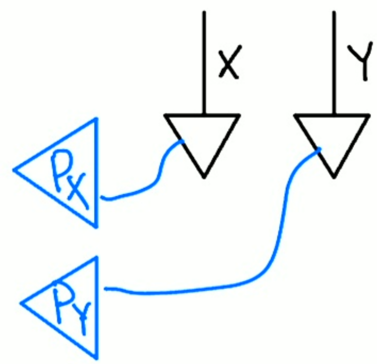
Falsification of causal structure ⇦ Violation of constraint on observed probability distribution

How the causal structure impacts the statistical parameters of the model

If X and Y have no common ancestry, then

$$X \perp Y$$

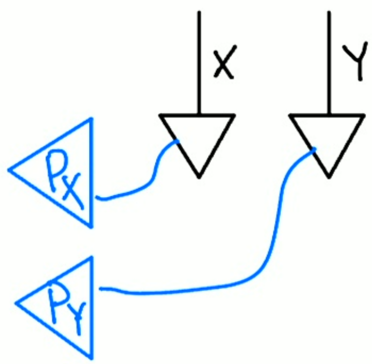
$$P_{XY} = P_X P_Y$$



If X and Y have no common ancestry, then

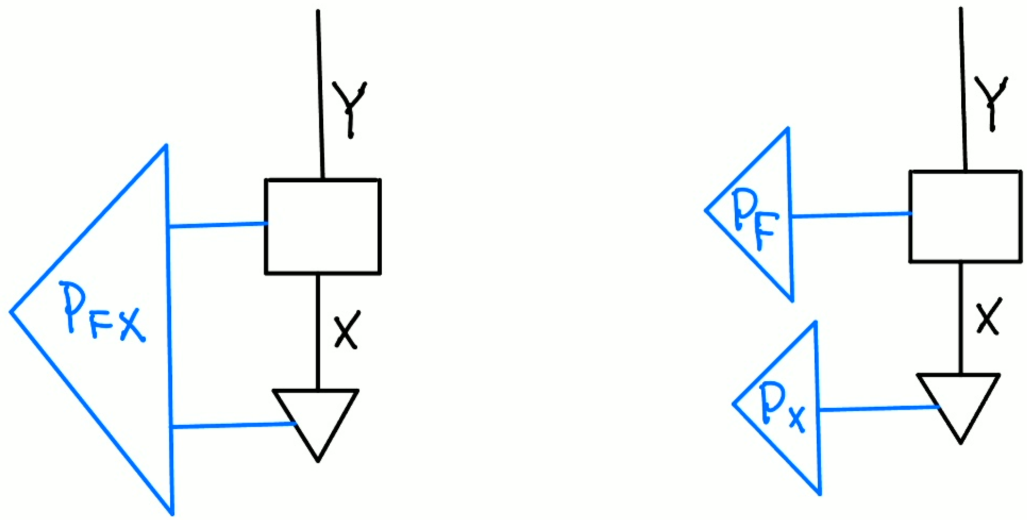
$$X \perp Y$$

$$P_{XY} = P_X P_Y$$



i.e., need a reason to posit correlated statistical sources

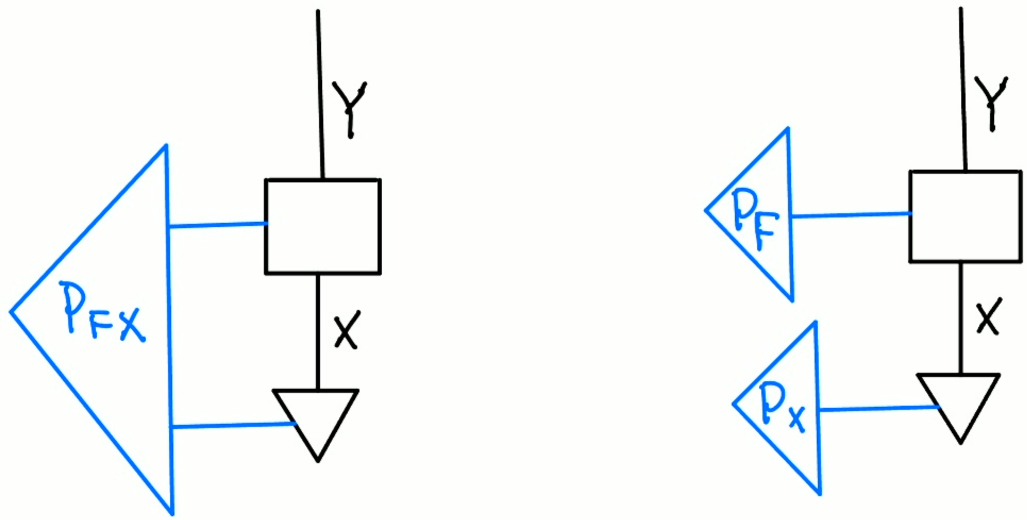
Recall: one generally assumes the mechanism fixing F is independent from that fixing X



This can be justified by the independence of causally disconnected nodes

Variables with no common ancestry are modelled by a factorizing source

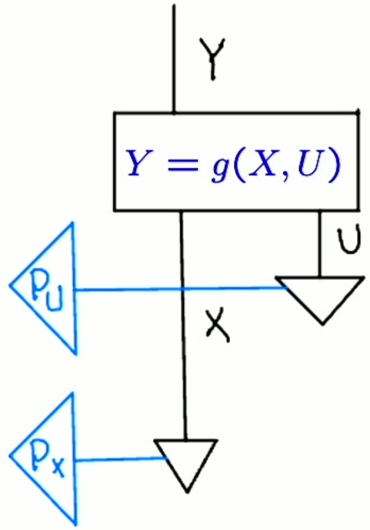
Recall: one generally assumes the mechanism fixing F is independent from that fixing X



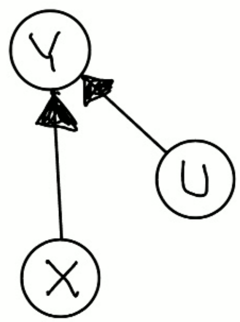
This can be justified by the independence of causally disconnected nodes

structural equation model for X causes Y

Circuit version

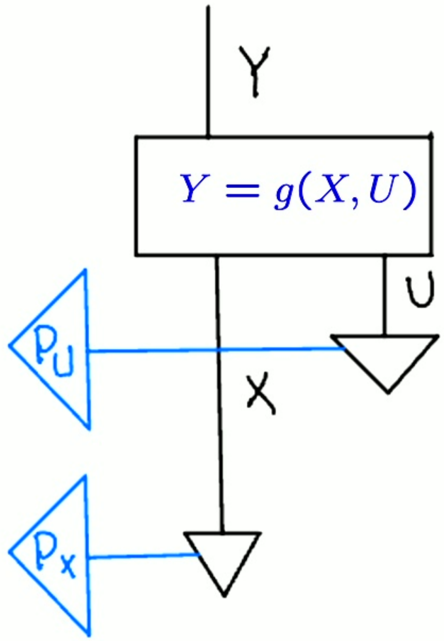


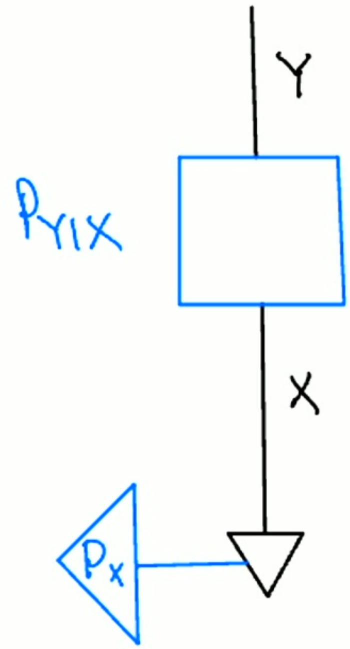
DAG version



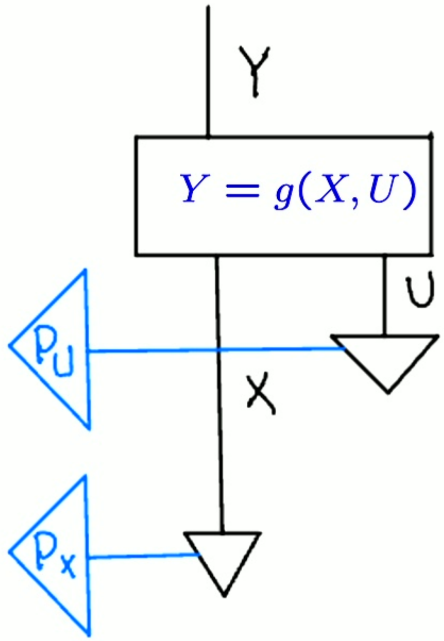
$$Y = g(X, U)$$
$$P_U$$
$$P_X$$

$$P_{Y|X} = \sum_U \delta_{Y,g(X,U)} P_U$$

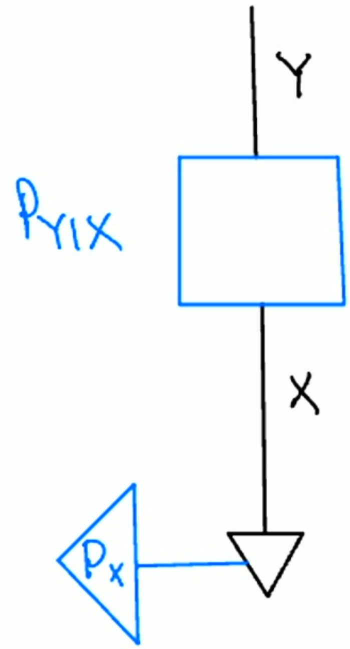




$$P_{Y|X} = \sum_U \delta_{Y,g(X,U)} P_U$$

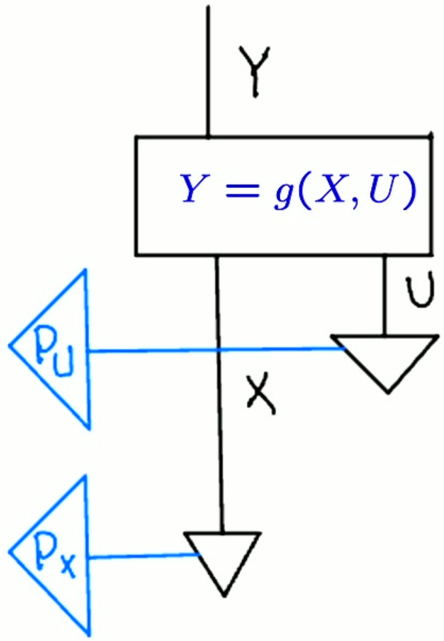


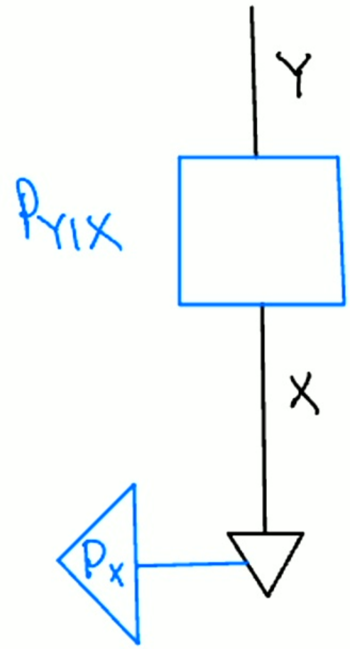
structural equation model
for X causes Y



Bayesian causal model
for X causes Y

$$P_{Y|X} = \sum_U \delta_{Y,g(X,U)} P_U$$

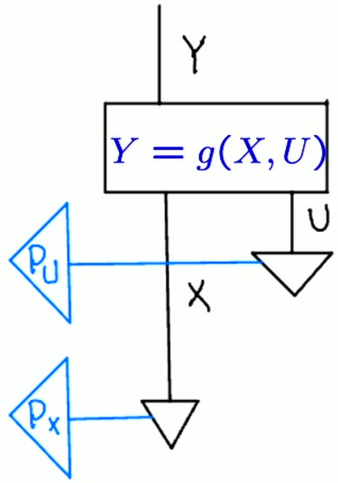
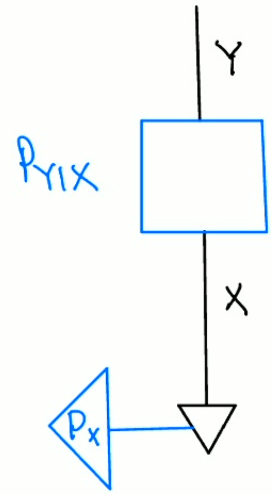
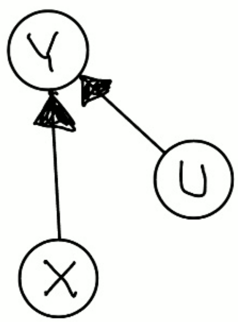




structural equation model
for X causes Y

"Quotienting"

Bayesian causal model
for X causes Y

	structural equation model	Bayesian causal model
Circuit version		
DAG version	 $Y = g(X,U)$ P_U P_X	 $P_{Y\|X}$ P_X

Reichenbach's principle

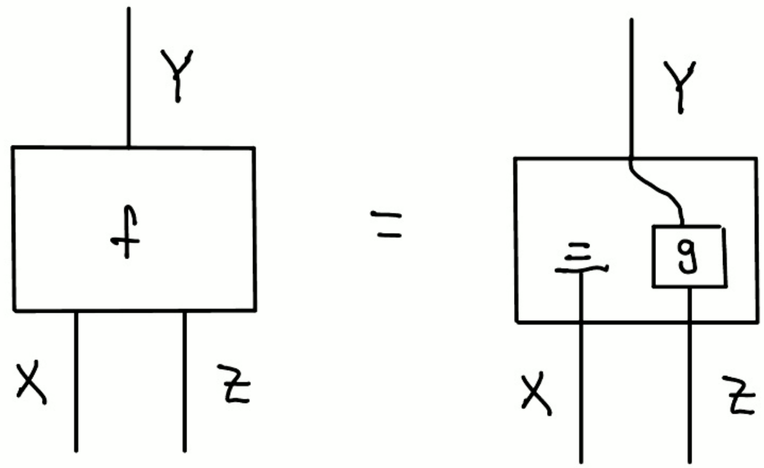
If Z is a complete common cause of X and Y, then

$$X \perp Y | Z$$

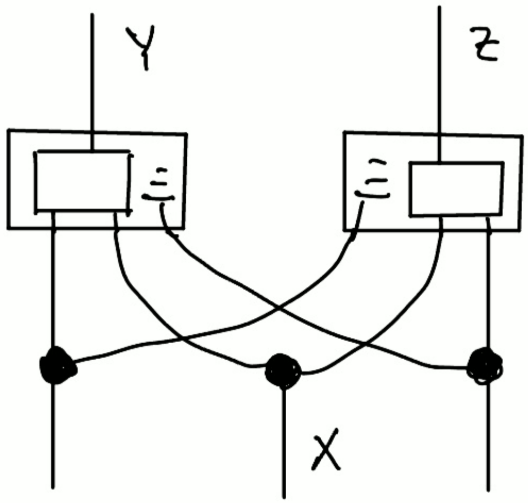
$$P_{XY|Z} = P_{X|Z} P_{Y|Z}$$

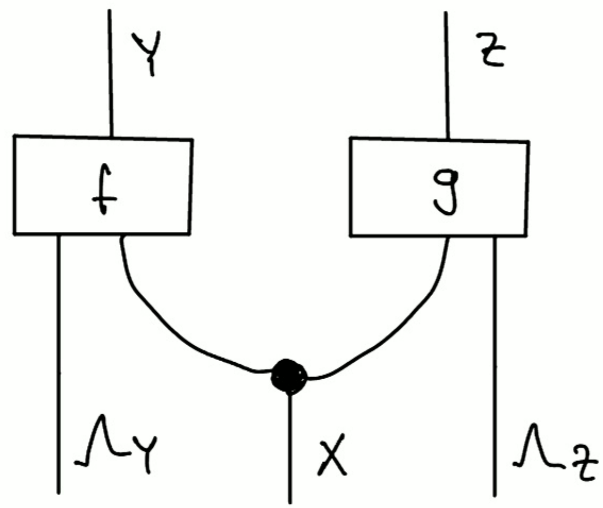
Definition: X has no influence on Y

The function f that determines Y from its causal antecedents has a trivial dependence on X

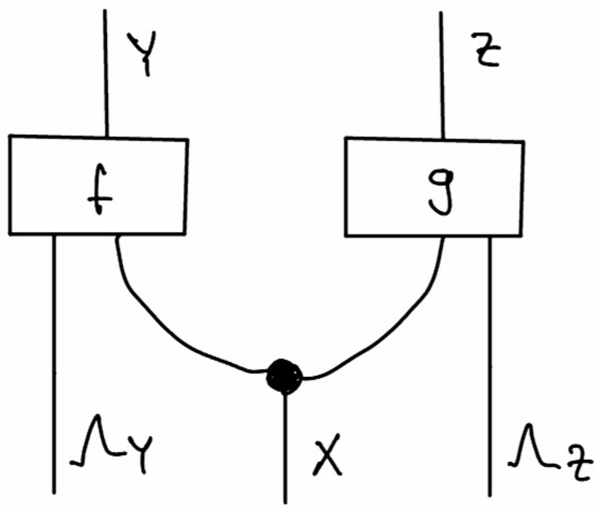


$$Y = f(X, Z) = g(Z)$$


Y
Z
X

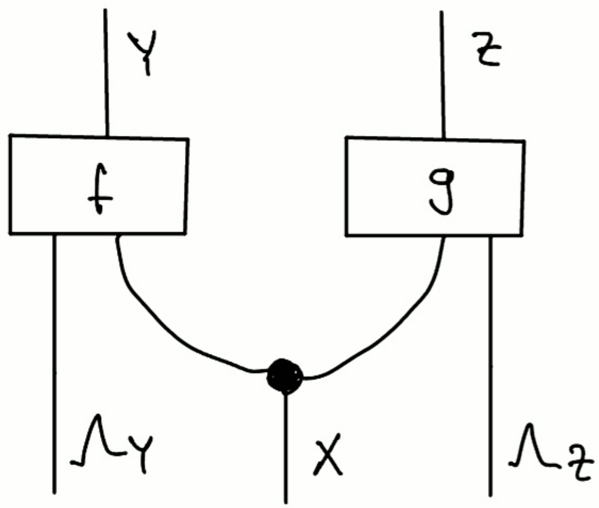


Y
Z
f
g
Λ_Y
X
Λ_Z



Why should an agent who assigns this causal structure organize their beliefs in such a way that

$$P_{XYZ} = P_{Y|X} P_{Z|X} P_X$$



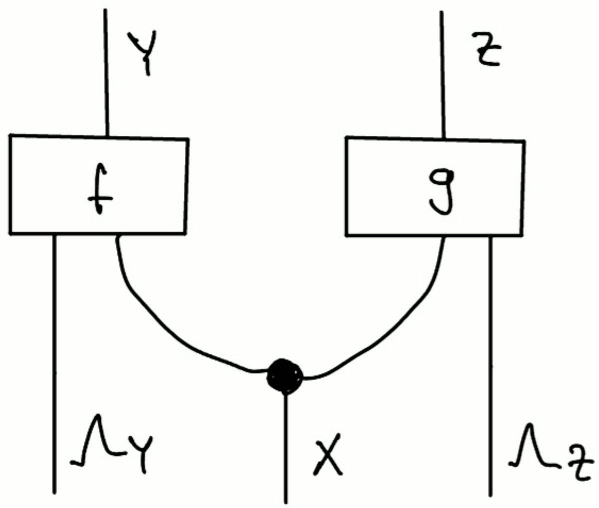
Probability theory implies

$$P_{XYZ} = \sum_{\Lambda_Y \Lambda_Z} P_{YZX\Lambda_Y\Lambda_Z}$$
$$= \sum_{\Lambda_Y \Lambda_Z} P_{YZ|X\Lambda_Y\Lambda_Z} P_{\Lambda_Y\Lambda_Z X}$$

From our causal hypothesis

$$P_{\Lambda_Y\Lambda_Z X} = P_{\Lambda_Y} P_{\Lambda_Z} P_X$$
$$P_{YZ|X\Lambda_Y\Lambda_Z} = \delta_{Y,f(X,\Lambda_X)} \delta_{Z,g(X,\Lambda_Z)}$$

$$P_{XYZ} = \sum_{\Lambda_Y \Lambda_Z} \delta_{Y,f(X,\Lambda_X)} \delta_{Z,g(X,\Lambda_Z)} P_{\Lambda_Y} P_{\Lambda_Z} P_X$$
$$= (\sum_{\Lambda_Y} \delta_{Y,f(X,\Lambda_X)} P_{\Lambda_Y})(\sum_{\Lambda_Z} \delta_{Z,g(X,\Lambda_Z)} P_{\Lambda_Z}) P_X$$



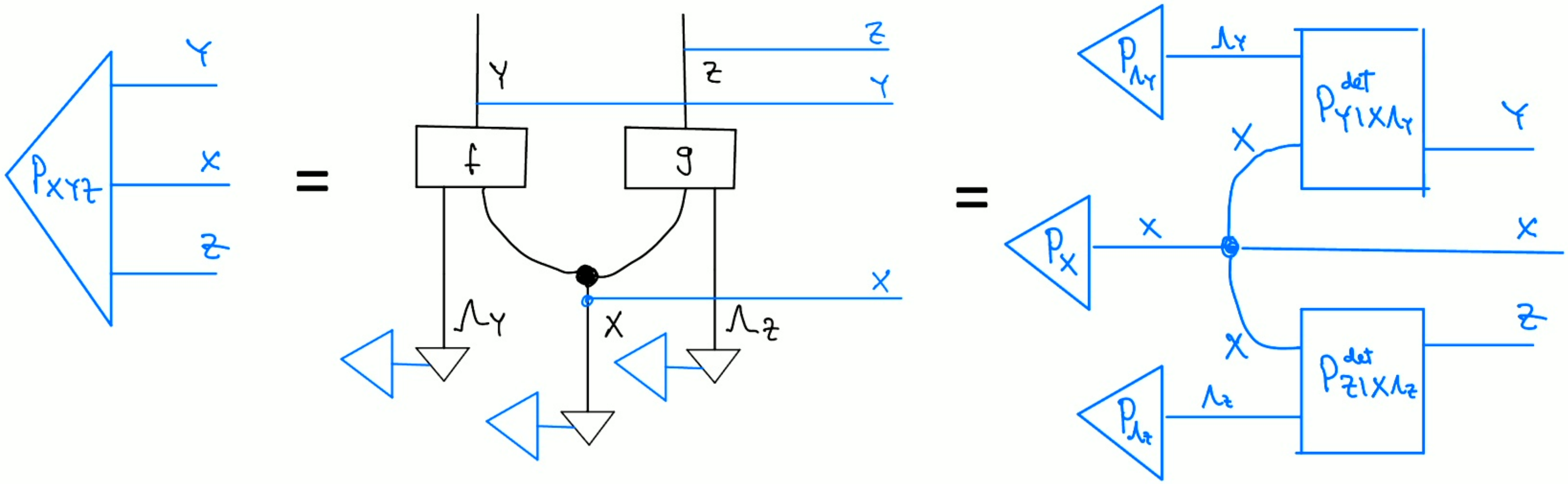
Probability theory implies

$$P_{XYZ} = \sum_{\Lambda_Y \Lambda_Z} P_{YZX\Lambda_Y\Lambda_Z}$$
$$= \sum_{\Lambda_Y \Lambda_Z} P_{YZ|X\Lambda_Y\Lambda_Z} P_{\Lambda_Y\Lambda_Z X}$$

From our causal hypothesis

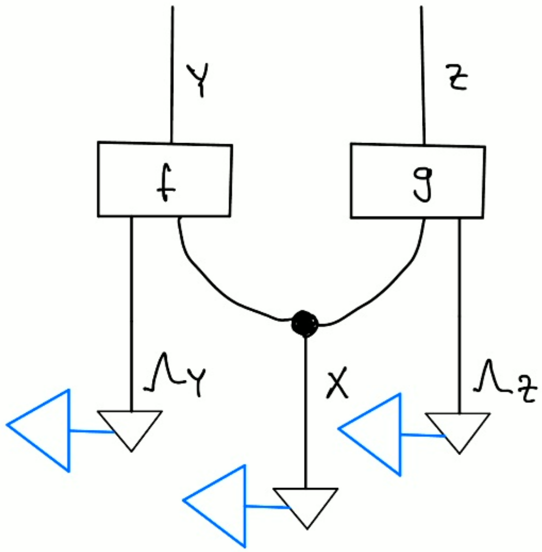
$$P_{\Lambda_Y\Lambda_Z X} = P_{\Lambda_Y} P_{\Lambda_Z} P_X$$
$$P_{YZ|X\Lambda_Y\Lambda_Z} = \delta_{Y,f(X,\Lambda_X)}\delta_{Z,g(X,\Lambda_Z)}$$

$$P_{XYZ} = \sum_{\Lambda_Y \Lambda_Z} \delta_{Y,f(X,\Lambda_X)}\delta_{Z,g(X,\Lambda_Z)} P_{\Lambda_Y} P_{\Lambda_Z} P_X$$
$$= (\sum_{\Lambda_Y} \delta_{Y,f(X,\Lambda_X)} P_{\Lambda_Y})(\sum_{\Lambda_Z} \delta_{Z,g(X,\Lambda_Z)} P_{\Lambda_Z}) P_X$$
$$= P_{Y|X} P_{Z|X} P_X$$

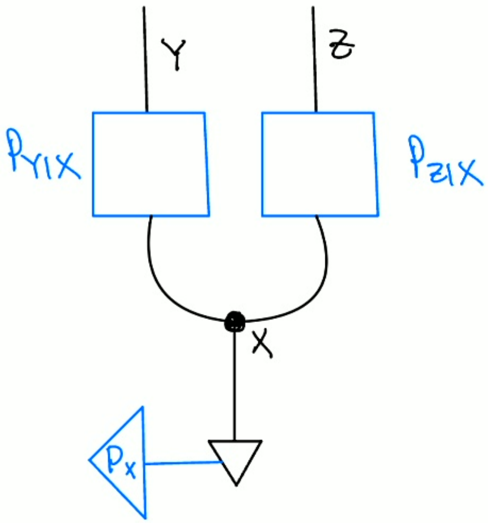


P_XYZ
Y
X
Z
=
f
g
Y
Z
Z
Y
X
Λ_Y
X
Λ_Z
=
P_ΛY
ΛY
det
P_Y|XΛY
Y
X
P_X
X
X
X
Z
P_ΛZ
ΛZ
det
P_Z|XΛZ

structural equation model
for X a complete common
cause of Y and Z



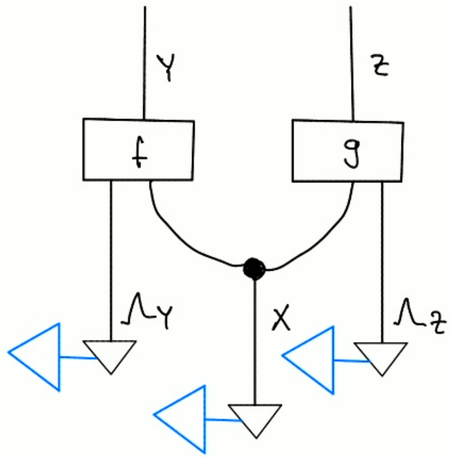
Bayesian causal model
for X a complete common
cause of Y and Z

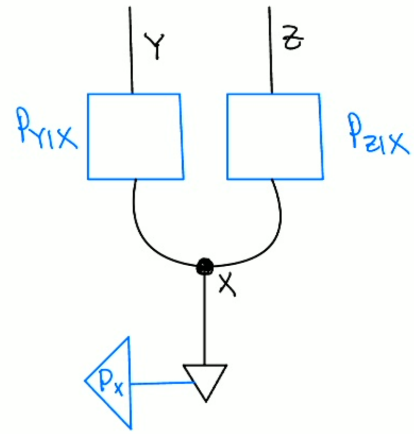


structural equation model

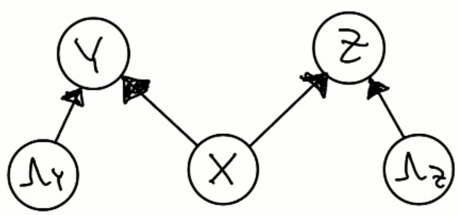
Bayesian causal model

Circuit version

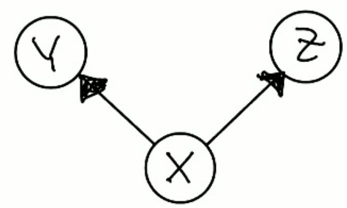




DAG version



$$Y = f(X, \Lambda_Y)$$
$$Z = g(X, \Lambda_Z)$$
$$P_{\Lambda_Y}$$
$$P_{\Lambda_Z}$$
$$P_X$$

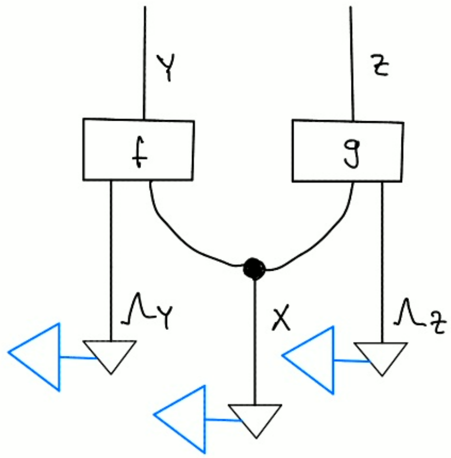


$$P_{Y|X}$$
$$P_{Z|X}$$
$$P_X$$

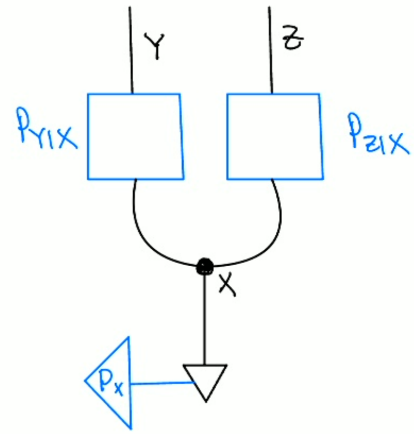
structural equation model | Bayesian causal model

Circuit version

Structural equation model:

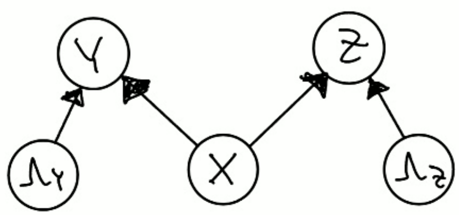


Bayesian causal model:



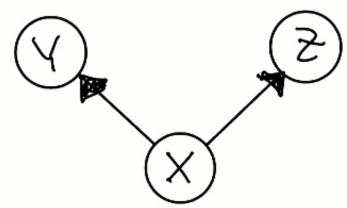
DAG version

Structural equation model:



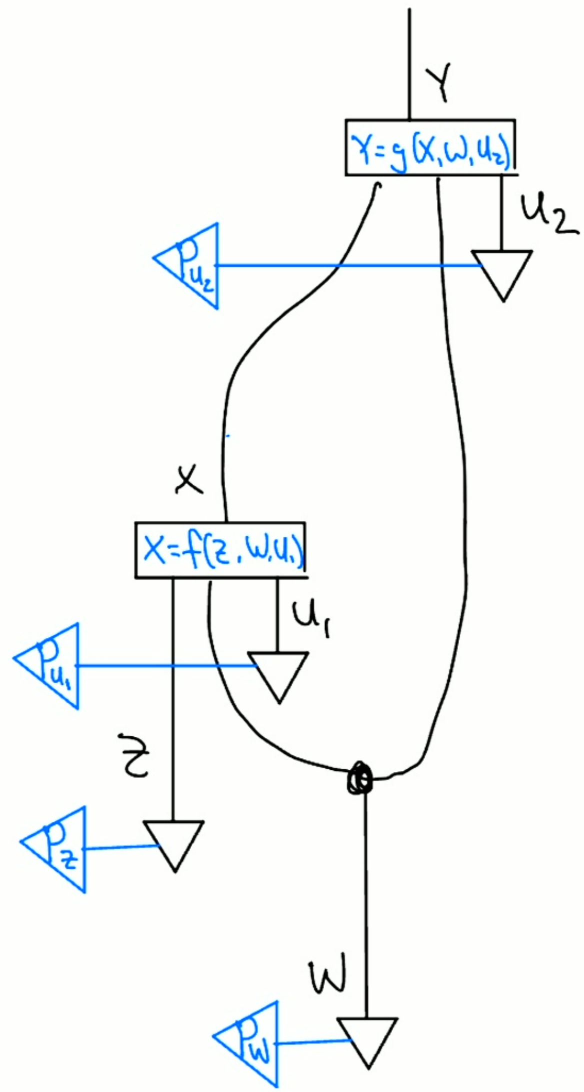
$$Y = f(X, \Lambda_Y)$$
$$Z = g(X, \Lambda_Z)$$
$$P_{\Lambda_Y}$$
$$P_{\Lambda_Z}$$
$$P_X$$

Bayesian causal model:

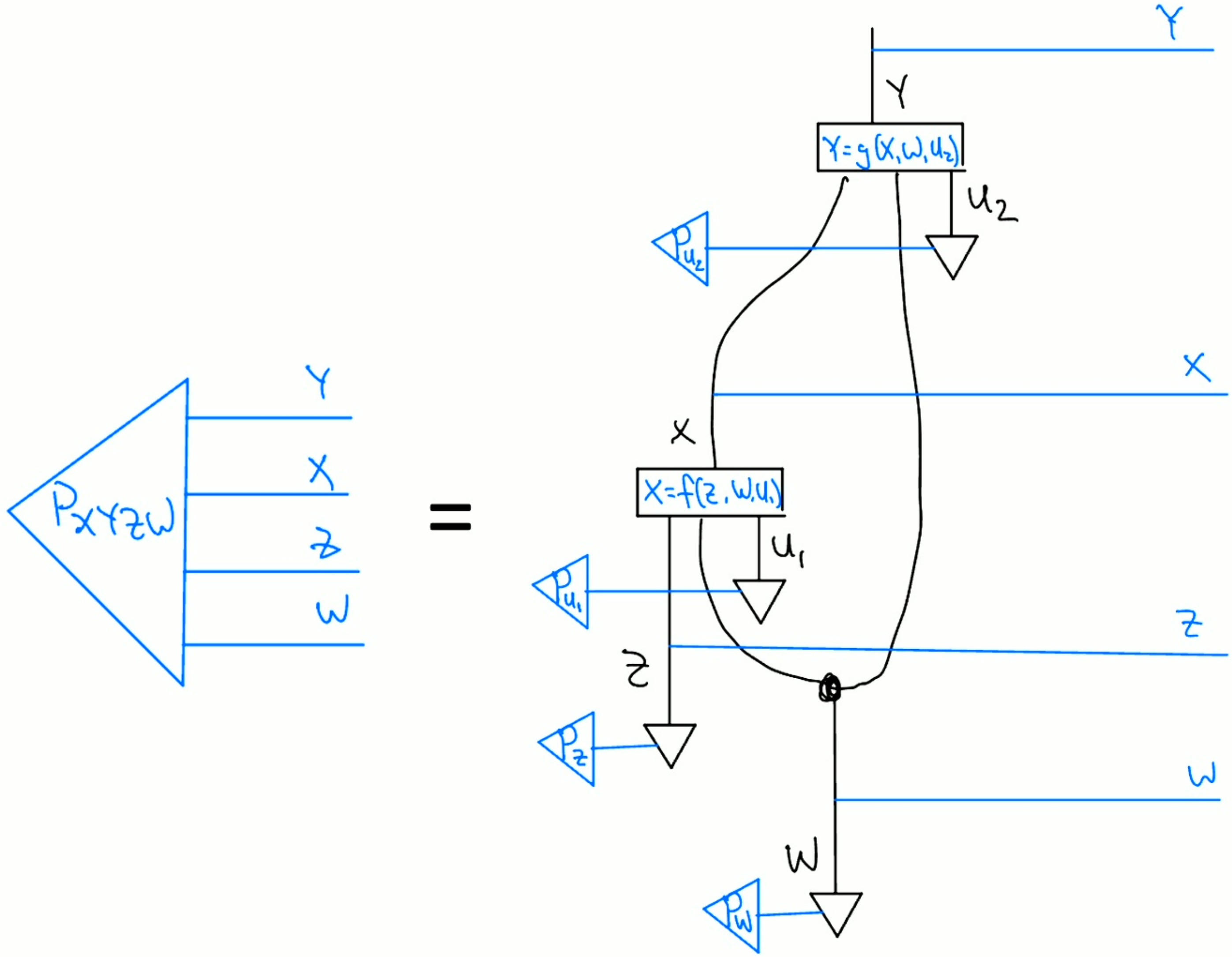


$$P_{Y|X}$$
$$P_{Z|X}$$
$$P_X$$

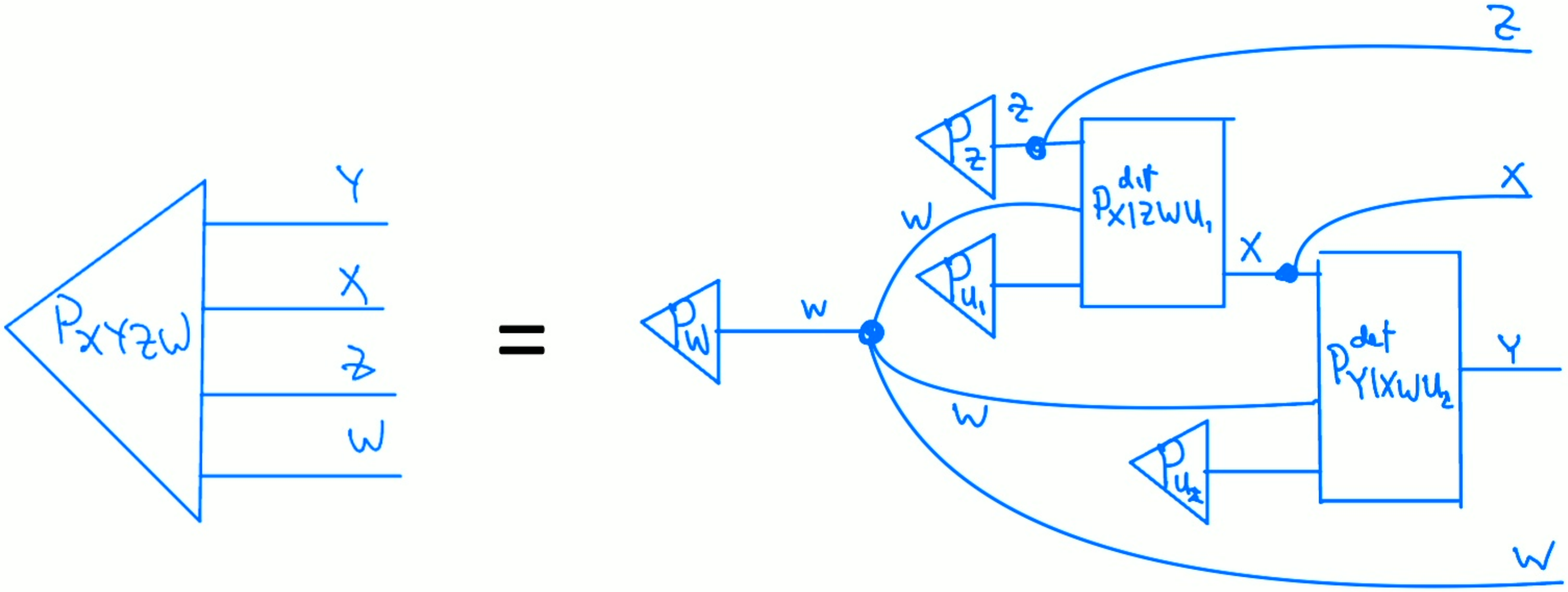
Generalizing to arbitrary causal structures

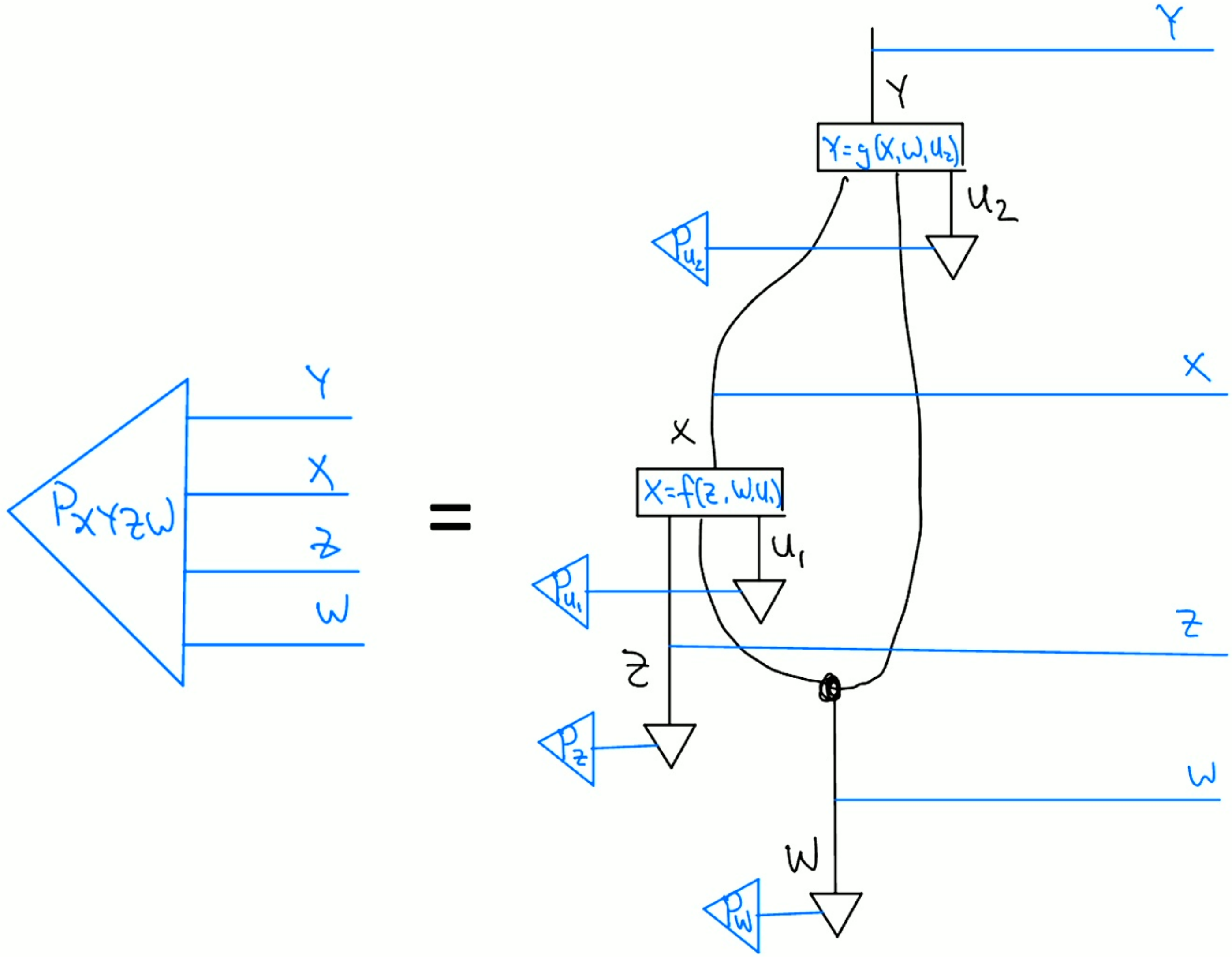


Y
Y=g(X,W,U₂)
U₂
P_{U₂}
X
X=f(Z,W,U₁)
U₁
P_{U₁}
Z
P_Z
W
P_W



P_{XYZW}
Y
X
Z
W
=
Y
Y
Y=g(X,W,U2)
U2
P_{U_2}
X
X
X=f(Z,W,U1)
U1
P_{U_1}
Z
Z
P_Z
W
W
P_W

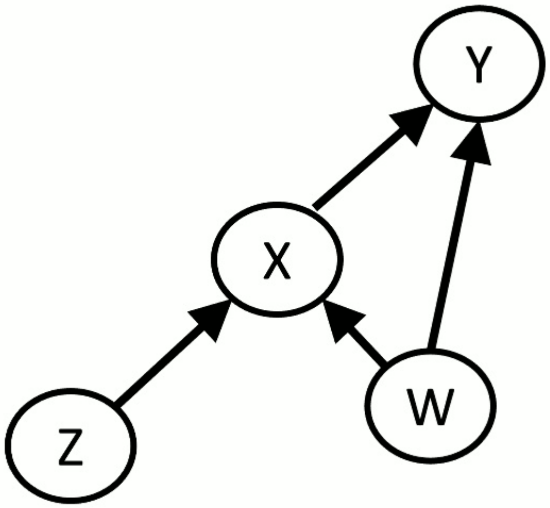

P_{XYZW}
Y
X
Z
W
=
P_W
W
W
W
P_Z
Z
P_{U_1}
$P^{det}_{X|ZWU_1}$
X
X
Z
P_{U_2}
$P^{det}_{Y|XWU_2}$
Y
W



P_{XYZW}
Y
X
Z
W
=
Y
Y
Y=g(X,W,U₂)
U₂
P_{U_2}
X
X
X=f(Z,W,U₁)
U₁
P_{U_1}
Z
Z
P_Z
W
W
P_W

Causal structure

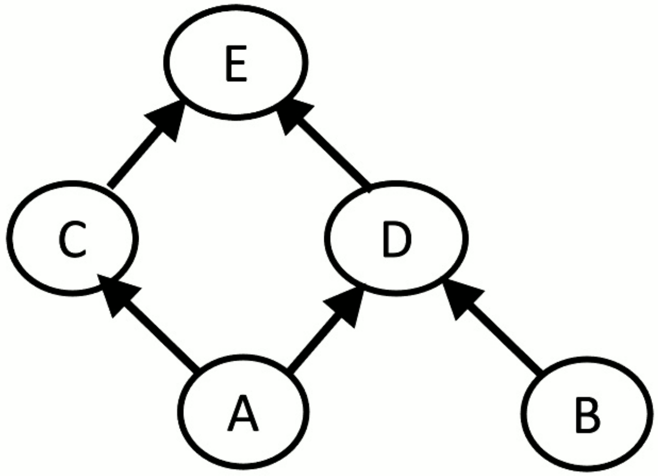
Parameters



$$P_{Y|XW}$$
$$P_{X|ZW}$$
$$P_W$$
$$P_Z$$

$$P_{YXWZ} = P_{Y|XW} P_{X|ZW} P_Z P_W$$

Causal structure



Parameters

$P_{E|CD}$
$P_{D|AB}$
$P_{C|A}$
P_B
P_A

$$P_{ABCDE} = P_{E|CD}P_{D|AB}P_{C|A}P_B P_A$$

If X_1, X_2,. . . , X_n have causal relations described by a DAG G, then the set of compatible distributions are those of the form:

$$P_{X_1,X_2,\ldots,X_n} = \prod_i P_{X_i|\mathrm{Parents}_G(X_i)}$$

This is sometimes called the Markov condition

Local Markov condition

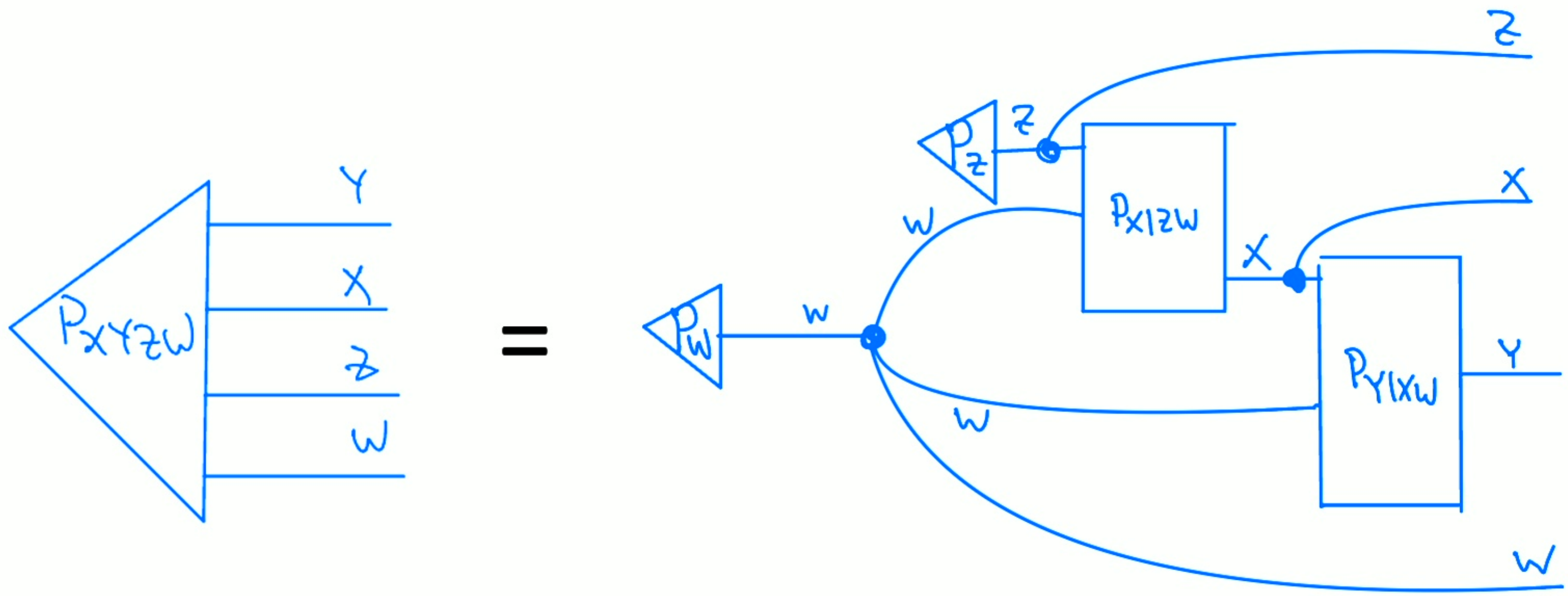
For a causal structure G, the compatible distributions are such that for every variable X

$$X \perp \mathrm{Nondescendents}_G(X) | \mathrm{Parents}_G(X)$$

If X_1, X_2,. . . , X_n have causal relations described by a DAG G, then the set of compatible distributions are those of the form:

$$P_{X_1,X_2,\ldots,X_n} = \prod_i P_{X_i|\text{Parents}_G(X_i)}$$

This is sometimes called the Markov condition



$$P_{YXWZ} = P_{Y|XW} P_{X|ZW} P_Z P_W$$

Local Markov condition

For a causal structure G, the compatible distributions are such that for every variable X

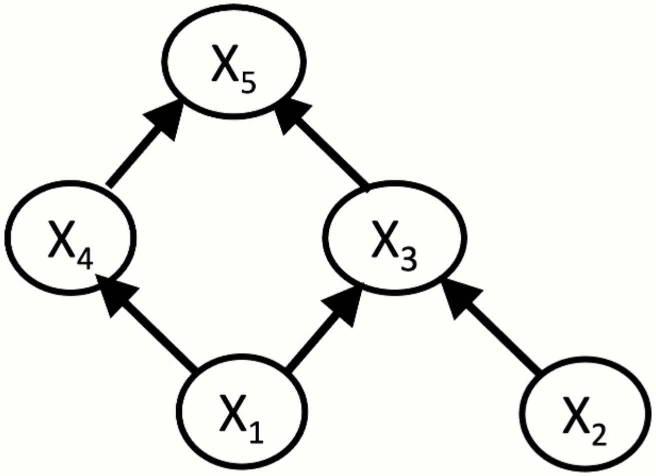
$$X \perp \mathrm{Nondescendents}_G(X) | \mathrm{Parents}_G(X)$$

$$P_{X_1,X_2,\ldots,X_n} = \prod_i P_{X_i|\mathrm{Parents}_G(X_i)}$$

$$\begin{aligned}
P(X|\mathrm{Pa}(X),\mathrm{Nd}(X)) &= \frac{P(X,\mathrm{Pa}(X),\mathrm{Nd}(X))}{P(\mathrm{Pa}(X),\mathrm{Nd}(X))}, \\
&= \frac{P(X|\mathrm{Pa}(X))\prod_{Y\in\mathrm{Pa}(X),\mathrm{Nd}(X)} P(Y|\mathrm{Pa}(Y))}{\prod_{Y\in\mathrm{Pa}(X),\mathrm{Nd}(X)} P(Y|\mathrm{Pa}(Y))}, \\
&= P(X|\mathrm{Pa}(X)).
\end{aligned}$$

$$X \perp \mathrm{Nondescendents}_G(X)|\mathrm{Parents}_G(X)$$

$$X \perp \mathrm{Nondescendents}_G(X) | \mathrm{Parents}_G(X)$$



$$(X_1 \perp X_2)$$
$$(X_2 \perp \{X_1, X_4\})$$
$$(X_3 \perp X_4 \mid \{X_1, X_2\})$$
$$(X_4 \perp \{X_2, X_3\} \mid X_1)$$
$$(X_5 \perp \{X_1, X_2\} \mid \{X_3, X_4\})$$

Semi-graphoid axioms

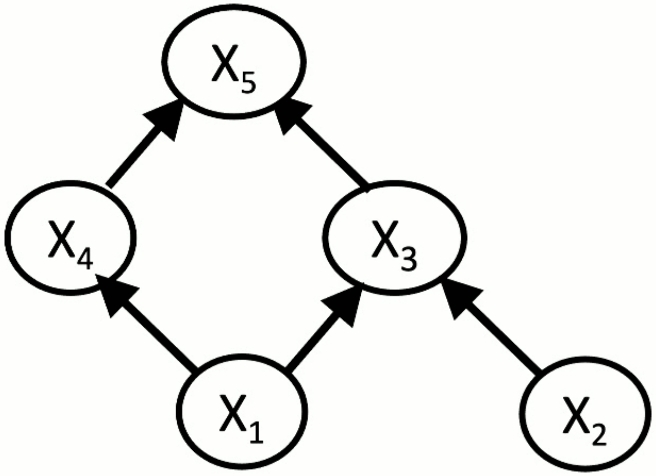
Symmetry: $(X \perp Y \mid Z) \Leftrightarrow (Y \perp X \mid Z)$

Decomposition: $(X \perp YW \mid Z) \Rightarrow (X \perp Y \mid Z)$

Weak Union: $(X \perp YW \mid Z) \Rightarrow (X \perp Y \mid ZW)$

Contraction: $(X \perp Y \mid Z)$ and $(X \perp W \mid ZY)$
$\Rightarrow (X \perp YW \mid Z)$

$$X \perp \text{Nondescendents}_G(X) | \text{Parents}_G(X)$$



$(X_1 \perp X_2)$

$(X_2 \perp \{X_1, X_4\})$

$(X_3 \perp X_4 \mid \{X_1, X_2\})$

$(X_4 \perp \{X_2, X_3\} \mid X_1)$

$(X_5 \perp \{X_1, X_2\} \mid \{X_3, X_4\})$

The semi-graphoid axioms then imply

$(X_4 \perp X_2 \mid X_1)$

$(\{X_4, X_5\} \perp X_2 \mid \{X_1, X_3\})$

$\cdots$

Semi-graphoid axioms

Symmetry: $(X \perp Y \,|\, Z) \Leftrightarrow (Y \perp X \,|\, Z)$

Decomposition: $(X \perp YW \,|\, Z) \Rightarrow (X \perp Y | \, Z)$

Weak Union: $(X \perp YW \,|\, Z) \Rightarrow (X \perp Y \,|\, ZW)$

Contraction: $(X \perp Y \,|\, Z)$ and $(X \perp W \,|\, ZY)$ $\Rightarrow (X \perp YW \,|\, Z)$

d-separation

X d-separated from **Y** by **Z**
in causal structure G

implies

$$\mathbf{X} \perp \mathbf{Y} | \mathbf{Z}$$

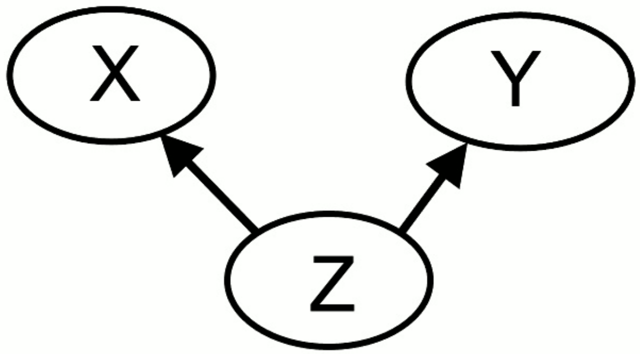
in every probability distribution
compatible with G

Definition (path blocking) A path between node X and node Y is blocked by a set of vertices **Z** if at least one of the following conditions holds:

1. The path contains a chain whose mediary node is in **Z**
2. The path contains a fork whose tail node is in **Z**

2. The path contains a collider whose head node is **not** in **Z** and no descendant of which is in **Z**.

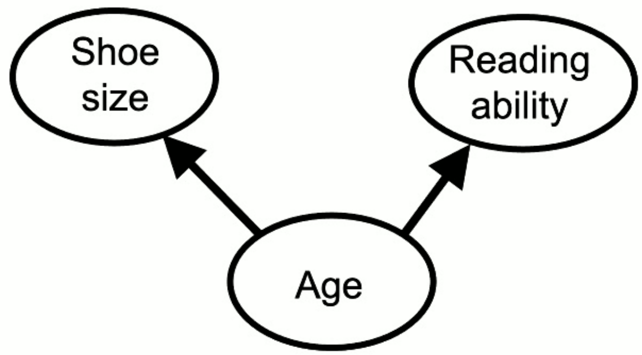
Definition (d-separation) Given a DAG G with vertices **V**, two set of vertices **X**, **Y** ∈ V are d-separated by a set of vertices **Z** ⊂ **V** if and only if for every pair of vertices, X and Y, from the sets **X** and **Y**, every path between X and Y is blocked.

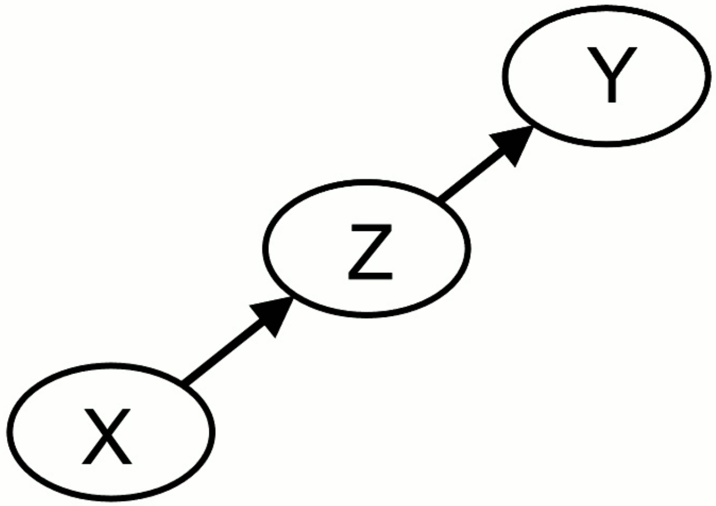


Z is a complete common cause of X and Y, therefore

$$X \perp Y | Z$$

$$P_{XY|Z} = P_{X|Z} P_{Y|Z}$$

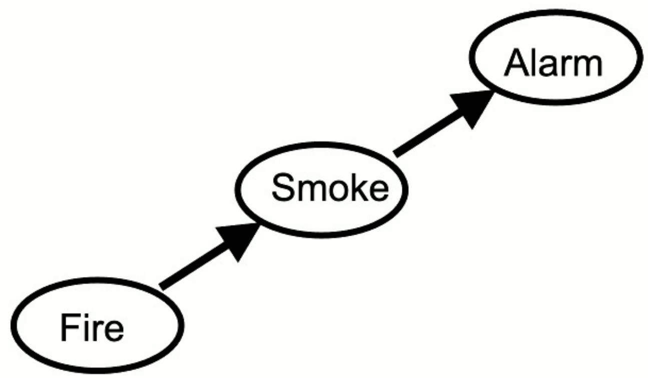

Shoe size
Reading ability
Age

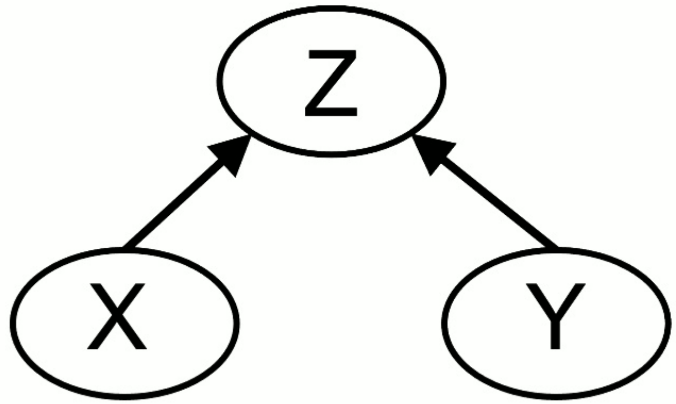


Z is the full set of parents of Y, while X is a nondescendent, therefore

$$X \perp Y | Z$$

$$P_{XY|Z} = P_{X|Z} P_{Y|Z}$$

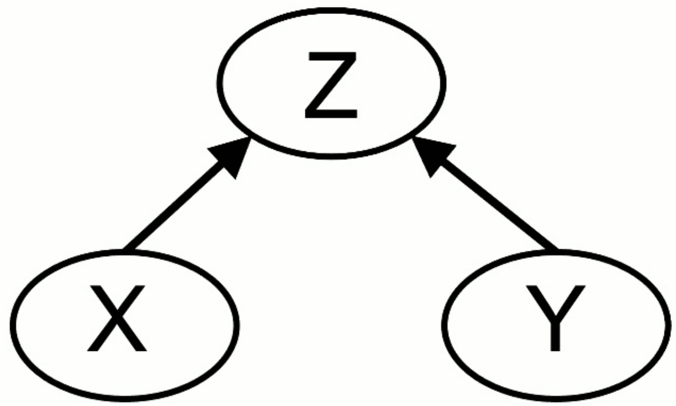

Alarm
Smoke
Fire



X and Y have no ancestors in common, therefore

X and Y are independent when one marginalizes over Z

$$X \perp Y$$

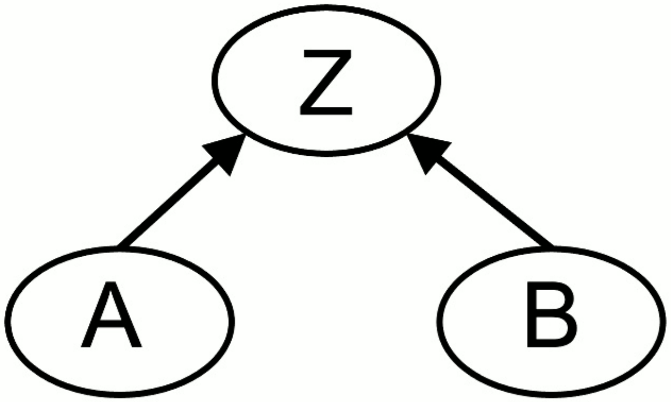


X and Y have no ancestors in common, therefore

X and Y are independent when one marginalizes over Z

$$X \perp Y$$

However, X and Y can become dependent if one conditions on Z



$$P_{Z|AB} = \delta_{Z,AB}$$

$$P_A = \tfrac{1}{2}[0]_A + \tfrac{1}{2}[1]_A$$

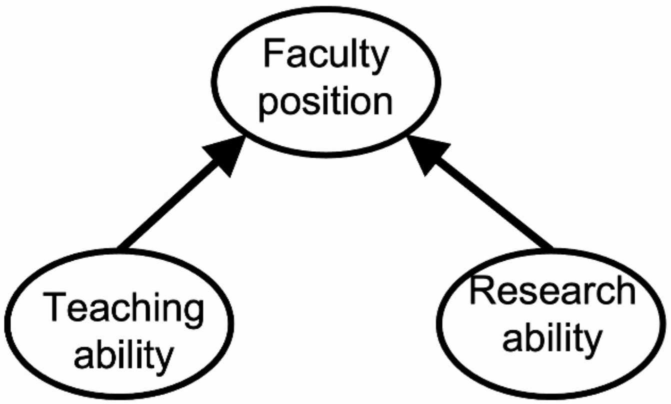
$$P_B = \tfrac{1}{2}[0]_B + \tfrac{1}{2}[1]_B$$

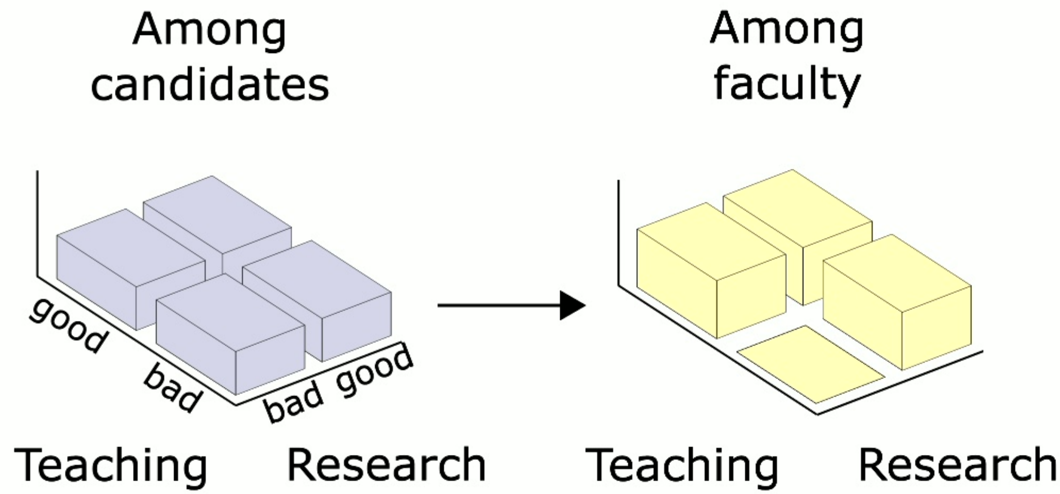
$$P_{AB} = (\tfrac{1}{2}[0]_A + \tfrac{1}{2}[1]_A)(\tfrac{1}{2}[0]_B + \tfrac{1}{2}[1]_B)$$

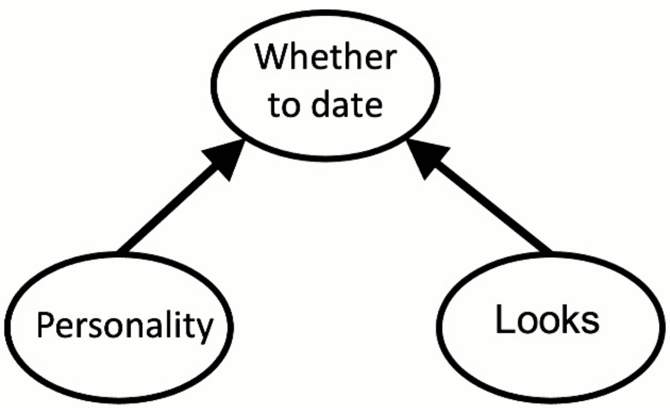
$$= \tfrac{1}{4}[0]_A[0]_B + \tfrac{1}{4}[0]_A[1]_B + \tfrac{1}{4}[1]_A 0]_B + \tfrac{1}{4}[1]_A[1]_B$$

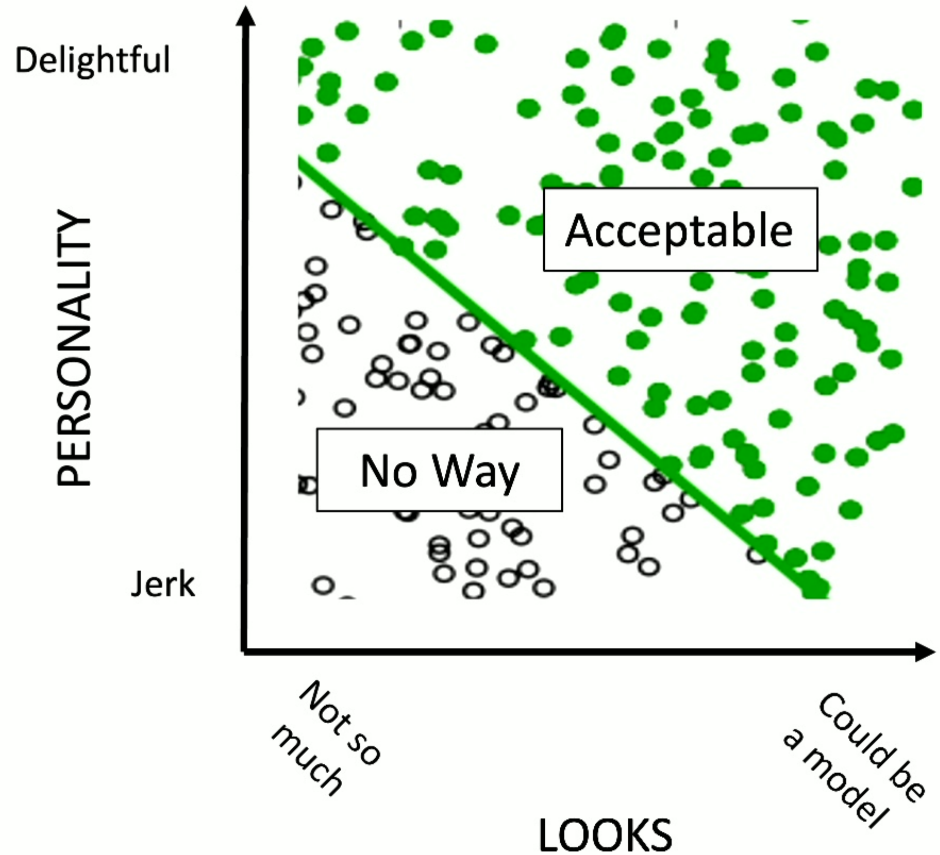
$$P_{AB|Z} = P_{Z|AB} P_{AB} P_Z^{-1}$$

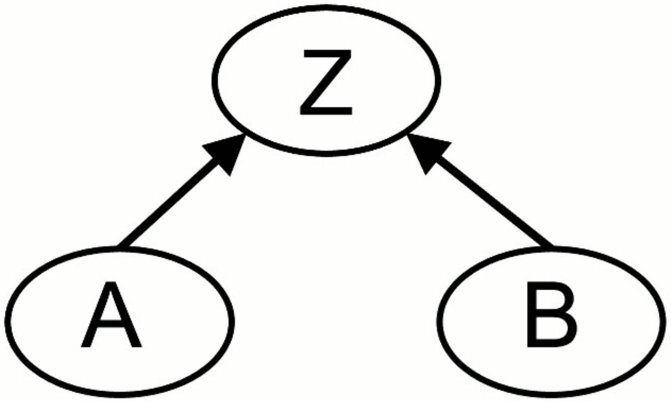
$$P_{AB|Z=0} = \tfrac{1}{3}[0]_A[0]_B + \tfrac{1}{3}[0]_A[1]_B + \tfrac{1}{3}[1]_A[0]_B$$


Faculty
position
Teaching
ability
Research
ability


Among
candidates
good
bad
bad
good
Teaching
Research
Among
faculty
Teaching
Research


Whether to date
Personality
Looks


Delightful
Jerk
PERSONALITY
Acceptable
No Way
Not so much
Could be a model
LOOKS

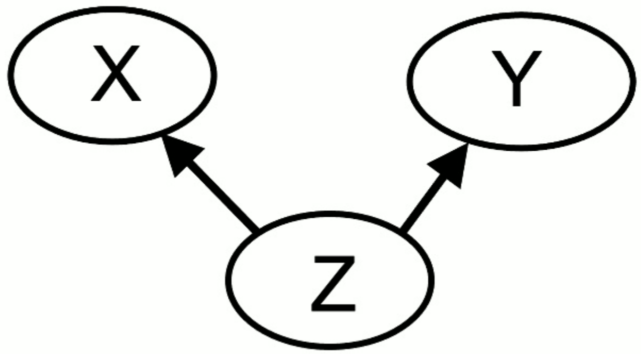


$$P_{Z|AB} = \delta_{Z,AB}$$
$$P_A = \tfrac{1}{2}[0]_A + \tfrac{1}{2}[1]_A$$
$$P_B = \tfrac{1}{2}[0]_B + \tfrac{1}{2}[1]_B$$

$$P_{AB} = (\tfrac{1}{2}[0]_A + \tfrac{1}{2}[1]_A)(\tfrac{1}{2}[0]_B + \tfrac{1}{2}[1]_B)$$
$$= \tfrac{1}{4}[0]_A[0]_B + \tfrac{1}{4}[0]_A[1]_B + \tfrac{1}{4}[1]_A 0]_B + \tfrac{1}{4}[1]_A[1]_B$$

$$P_{AB|Z} = P_{Z|AB} P_{AB} P_Z^{-1}$$

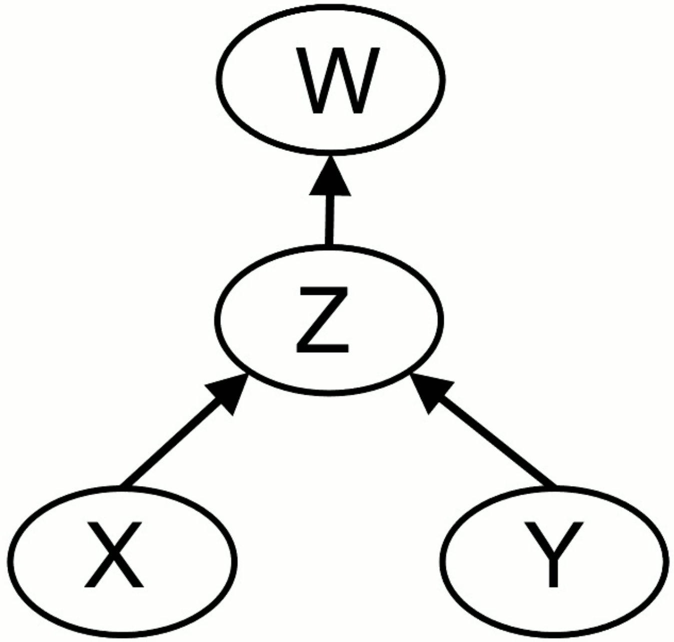
$$P_{AB|Z=0} = \tfrac{1}{3}[0]_A[0]_B + \tfrac{1}{3}[0]_A[1]_B + \tfrac{1}{3}[1]_A[0]_B$$



Z is a complete common cause of X and Y, therefore

$$X \perp Y | Z$$

$$P_{XY|Z} = P_{X|Z} P_{Y|Z}$$



If W is a **descendent** of a common effect of X and Y, then

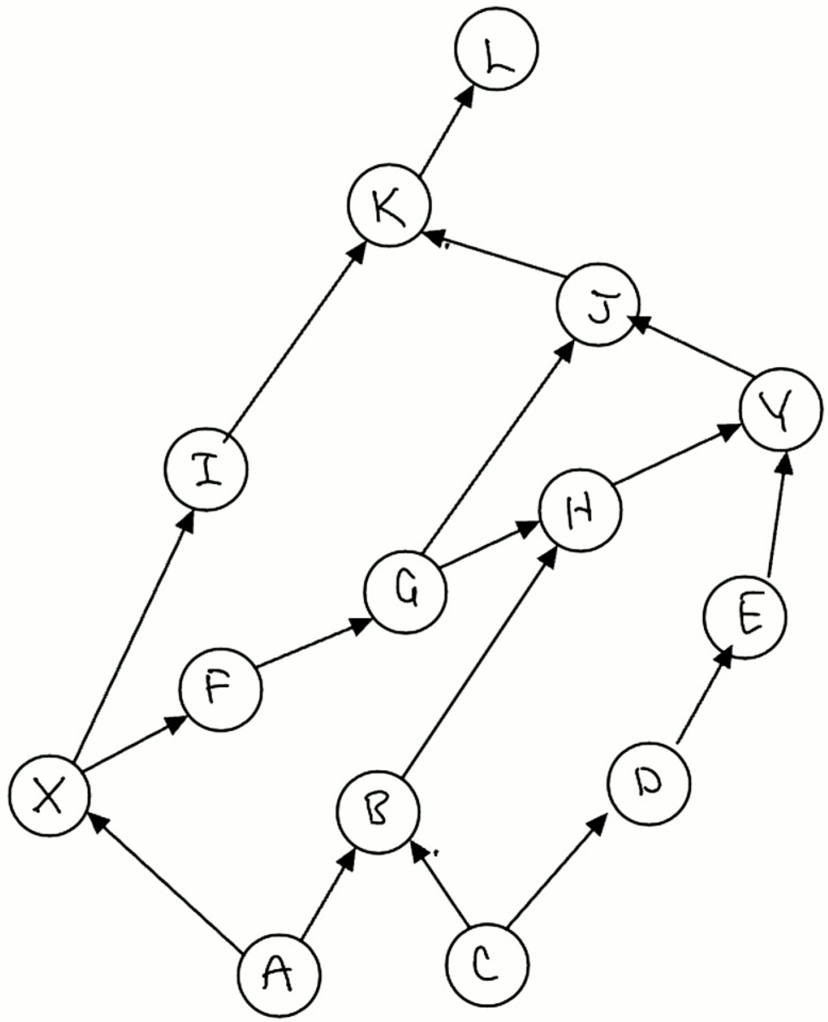
X and Y can become dependent when one conditions over W

Definition (path blocking) A path between node X and node Y is blocked by a set of vertices **Z** if at least one of the following conditions holds:

1. The path contains a chain whose mediary node is in **Z**
2. The path contains a fork whose tail node is in **Z**

2. The path contains a collider whose head node is **not** in **Z** and no descendant of which is in **Z**.

Definition (d-separation) Given a DAG G with vertices **V**, two set of vertices **X**, **Y** ∈ V are d-separated by a set of vertices **Z** ⊂ **V** if and only if for every pair of vertices, X and Y, from the sets **X** and **Y**, every path between X and Y is blocked.


L
K
J
I
Y
H
G
E
F
X
B
D
A
C

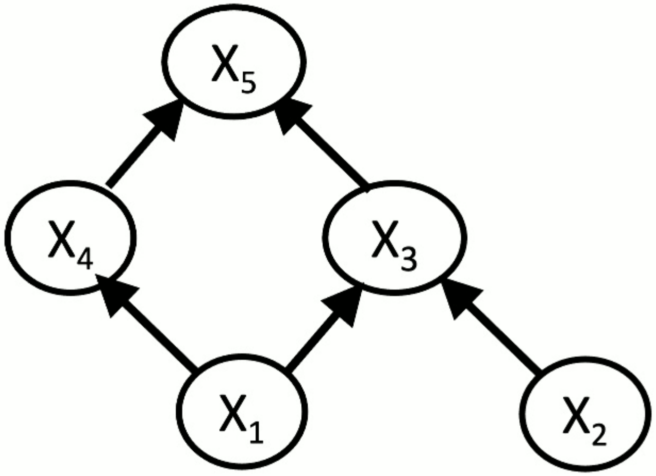
X d-separated from **Y** by **Z**
in causal structure G

implies

$$\mathbf{X} \perp \mathbf{Y} | \mathbf{Z}$$

in every probability distribution
compatible with G

If P has all the conditional independences that are implied by d-separation relations in a DAG G, then P is said to be Markov relative to G



$$\{X_4, X_5\} \perp X_2 | \{X_1, X_3\}$$

Arduous to derive this from the local Markov condition and applications of the semi-graphoid axioms

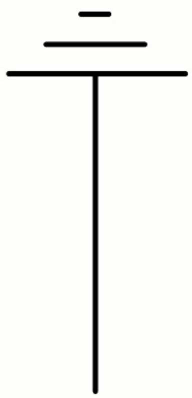
Follows from d-separation in a straightforward way

Definition (path blocking) A path between node X and node Y is blocked by a set of vertices **Z** if at least one of the following conditions holds:

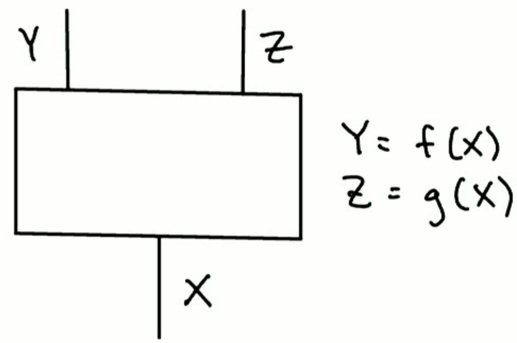
1. The path contains a chain whose mediary node is in **Z**
2. The path contains a fork whose tail node is in **Z**

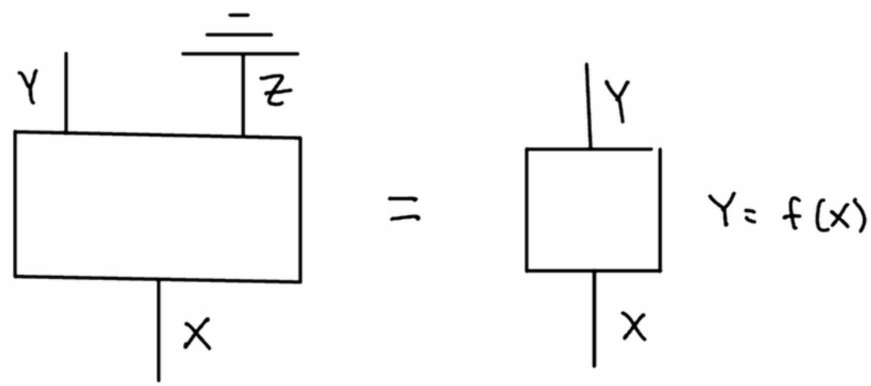
2. The path contains a collider whose head node is **not** in **Z** and no descendant of which is in **Z**.

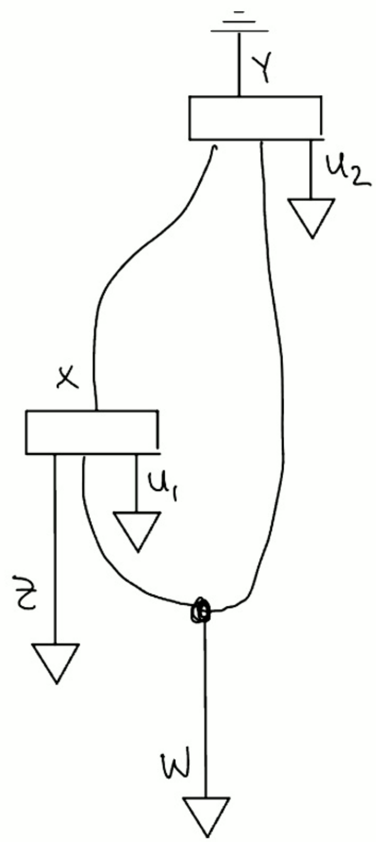
Definition (d-separation) Given a DAG G with vertices **V**, two set of vertices **X**, **Y** ∈ V are d-separated by a set of vertices **Z** ⊂ **V** if and only if for every pair of vertices, X and Y, from the sets **X** and **Y**, every path between X and Y is blocked.

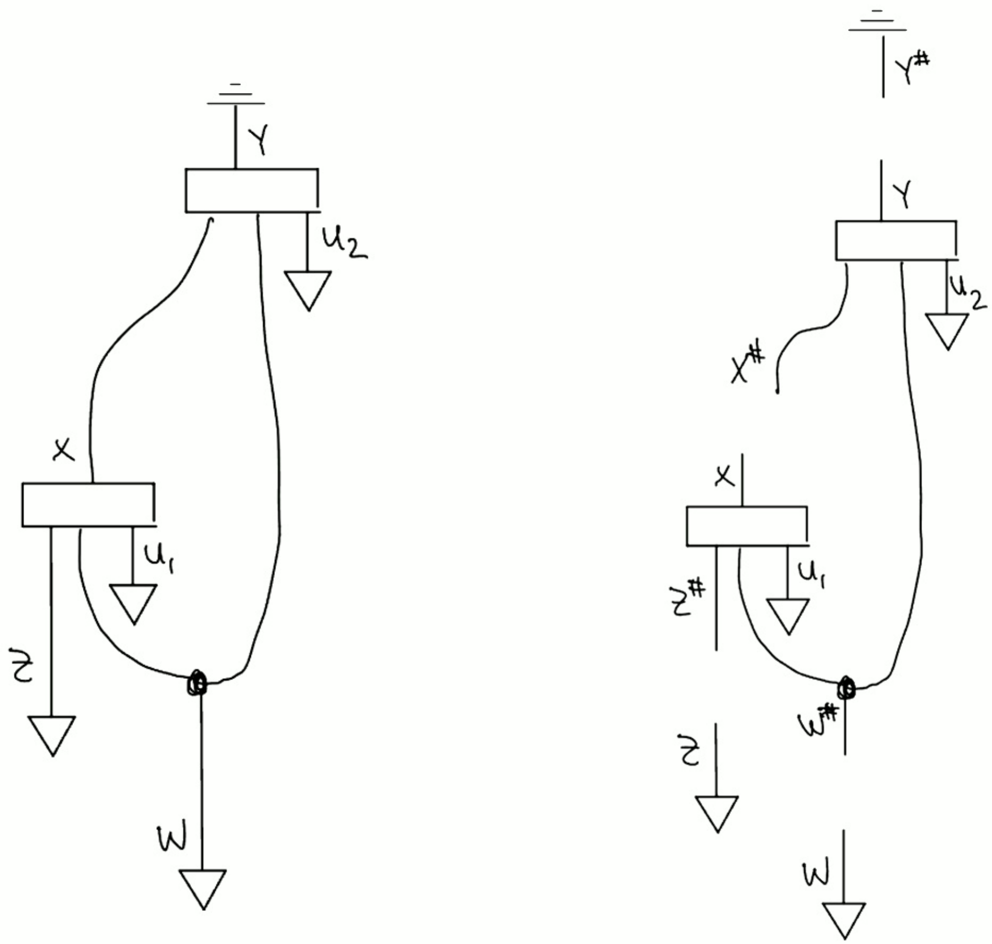


Exclude from model
(Not relevant for any future causal processes)


Y
Z
Y= f(X)
Z = g(X)
X


Y
Z
=
Y
Y= f(X)
X
X

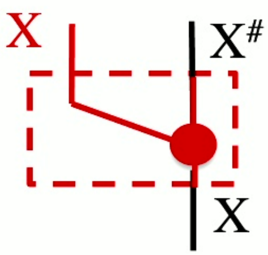

Y
u_2
X
u_1
Z
W



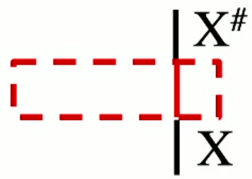
Y
u_2
X
u_1
Z
W
$Y^\#$
Y
u_2
$X^\#$
X
u_1
$Z^\#$
$W^\#$
Z
W

Causal account of different probing schemes

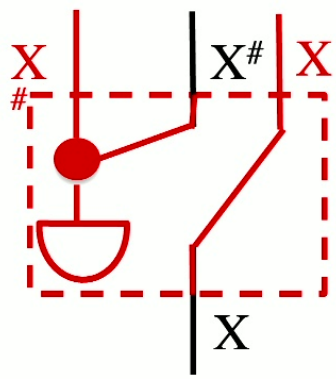
Observation



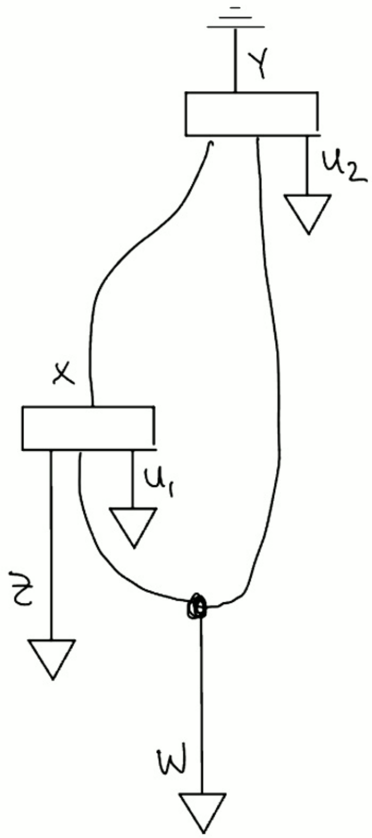
Leave it unobserved

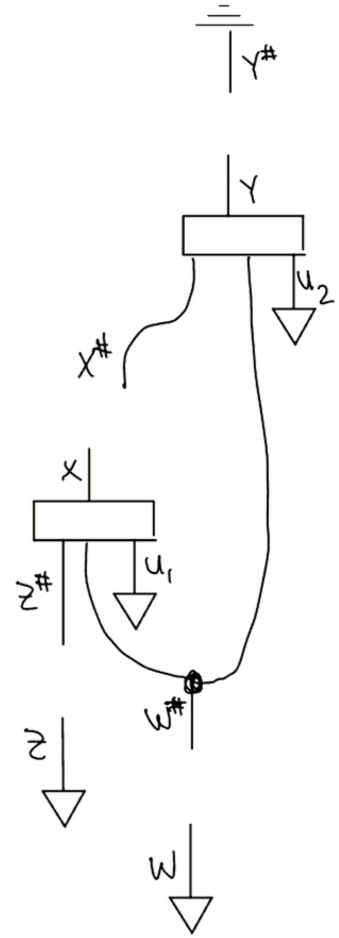


Intervention



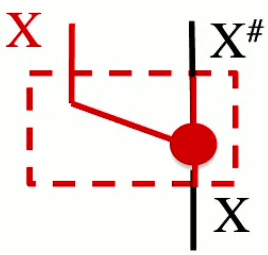
Observe and reprepare new random version with flag


Y
u2
X
u1
Z
W

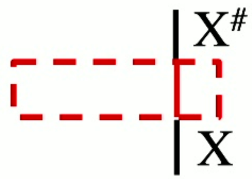

Y#
Y
u2
X#
X
Z#
u1
W#
Z
W

Causal account of different probing schemes

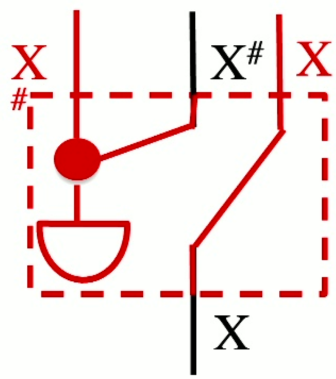
Observation



Leave it unobserved

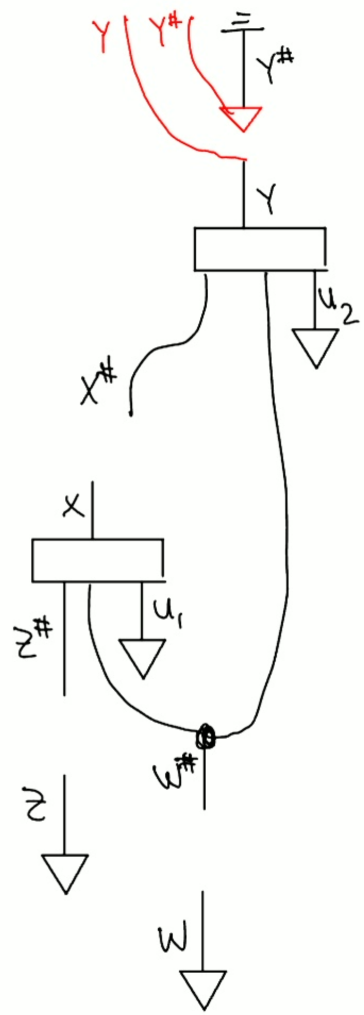


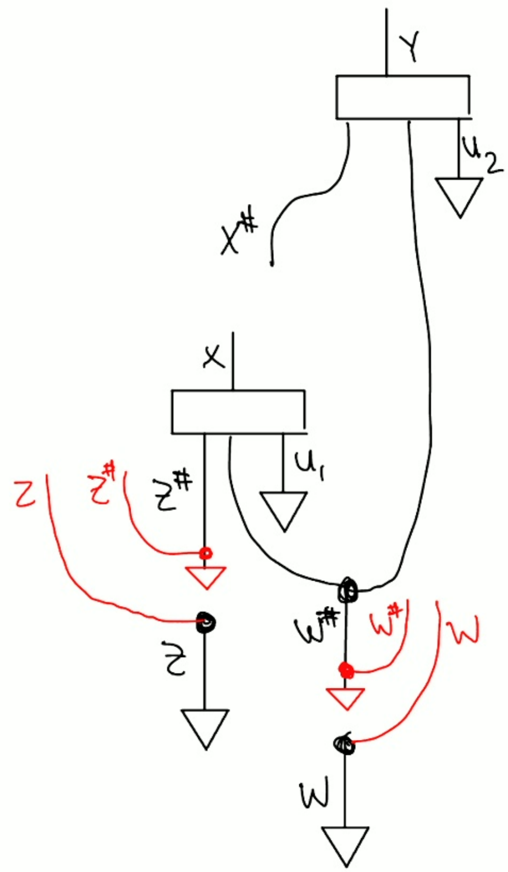
Intervention



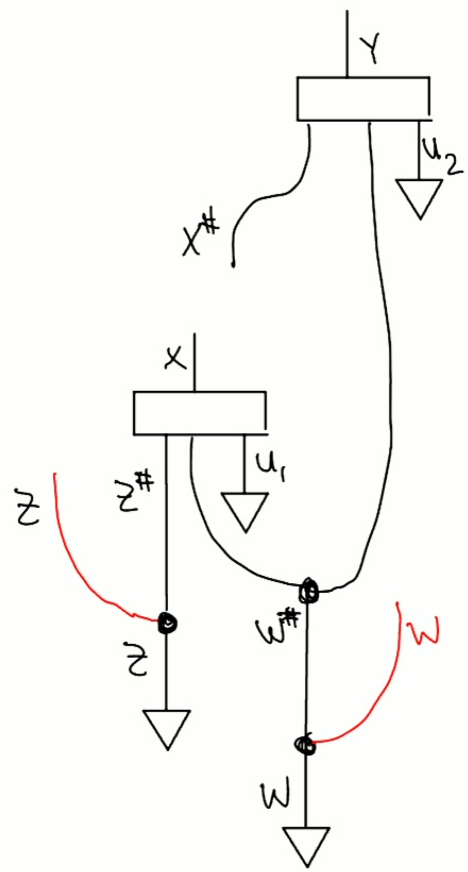
Observe and reprepare new random version with flag

Intervention

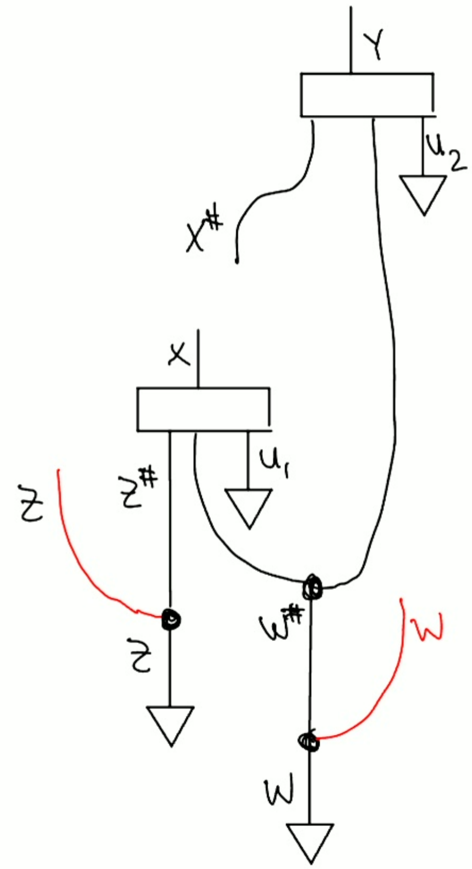




Intervention

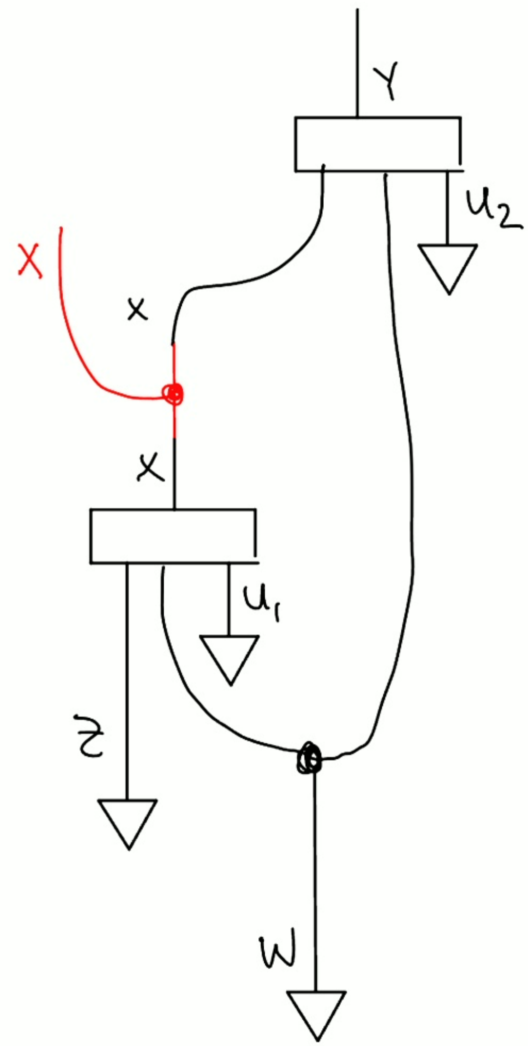


Observation

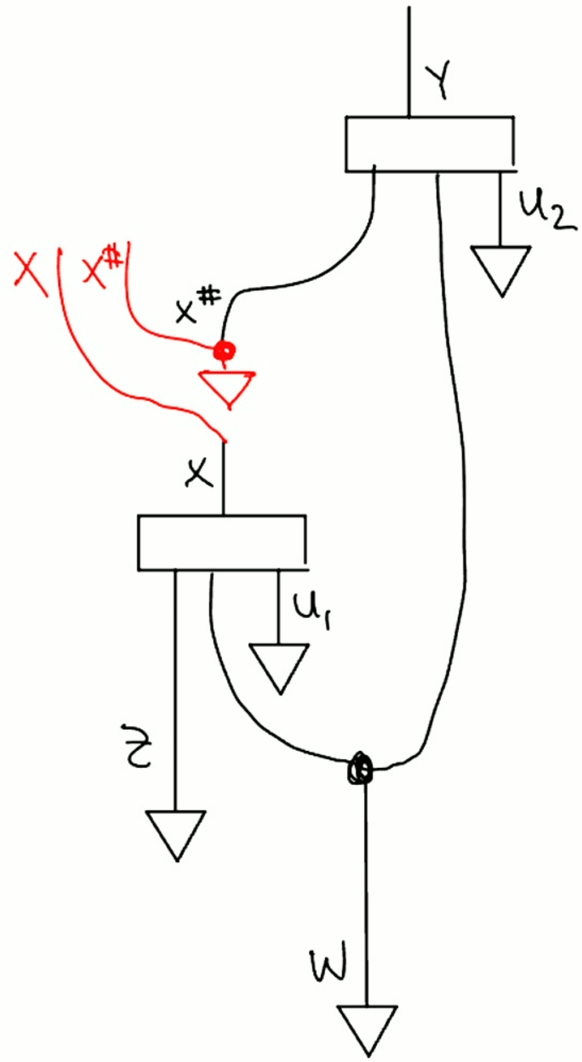


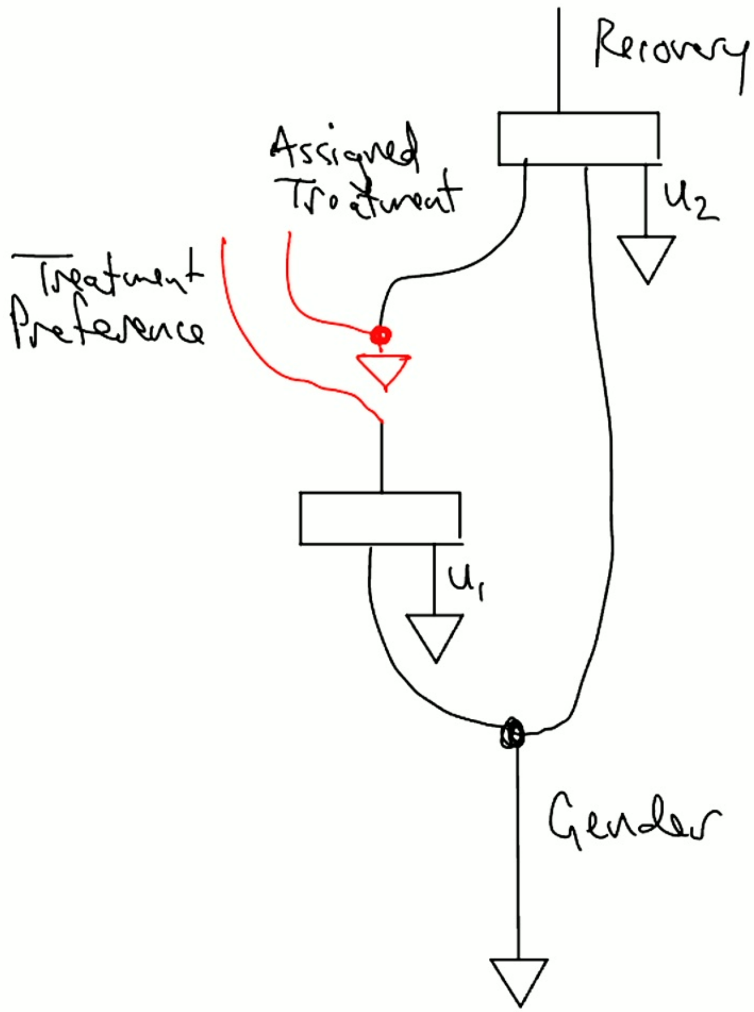
For a root node, observation is essentially just as good as intervention

Ambiguity as to whether correlations between X and Y are due to cause-effect relation or due to common-cause relation



Intervention

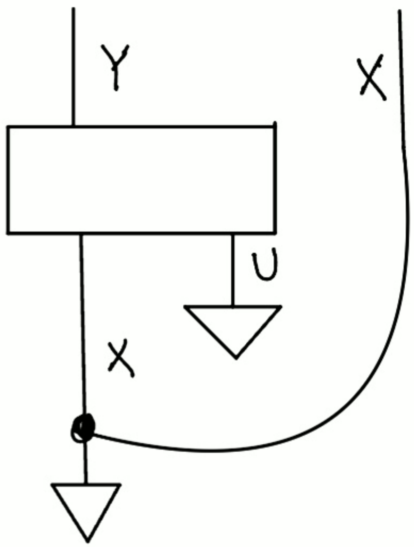


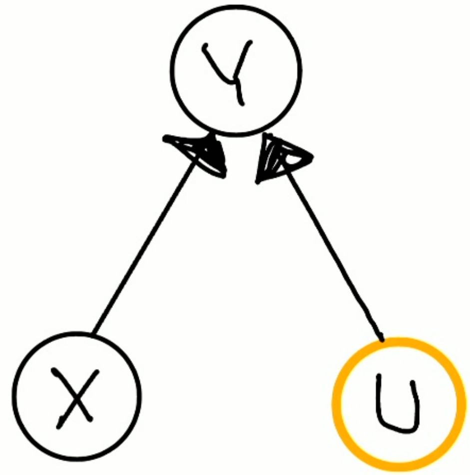

Recovery
Assigned Treatment
Treatment Preference
U_2
U_1
Gender

If there is a deterministic causal mechanism between X and Y, intervention on X can learn the function

For most pairs of variables, X and Y, the causal mechanism is not deterministic, so we learn only the do-conditional $P_{X|doY}$

The case of passive observation

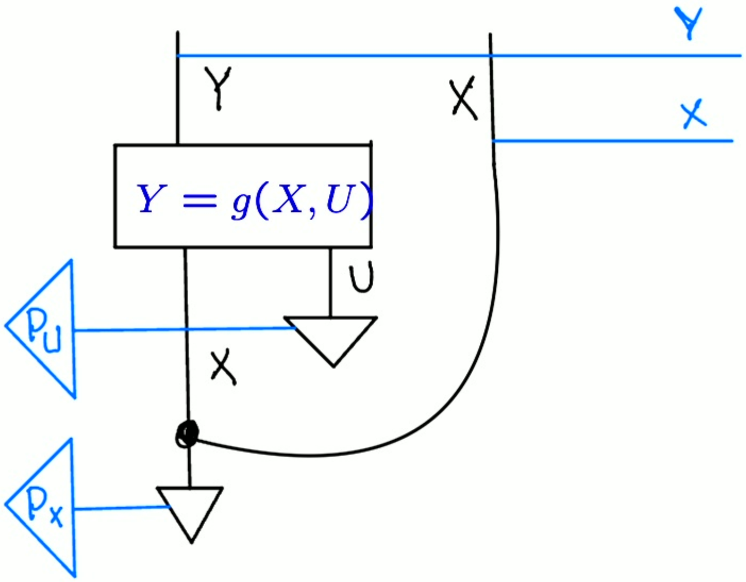


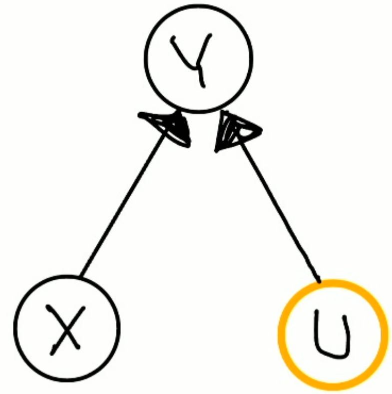


unobserved – closed wire
observed –open wire

unobserved – orange node
Observed – black node

Structural equation model

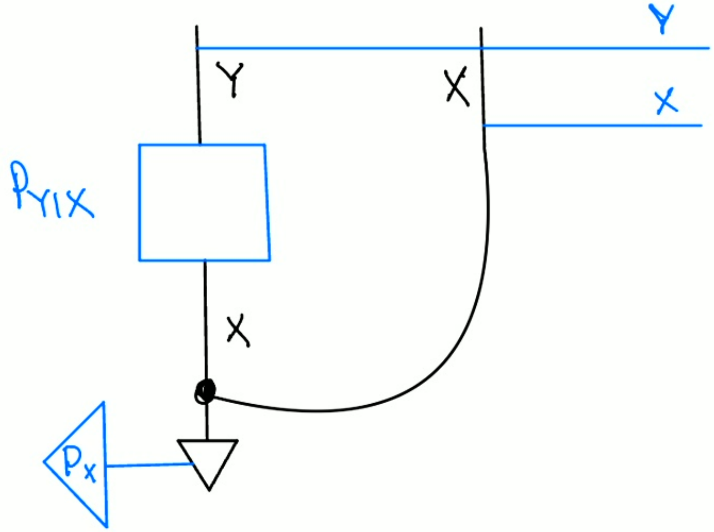


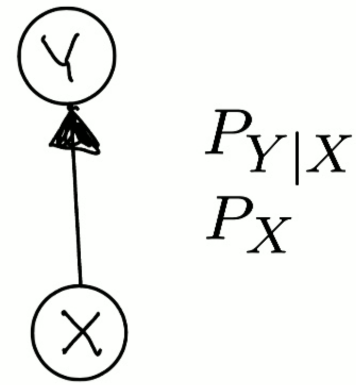


$$Y = g(X, U)$$
$$P_U$$
$$P_X$$

$$P_{XY} = \sum_U \delta_{Y,g(X,U)} P_U P_X$$

Bayesian causal model

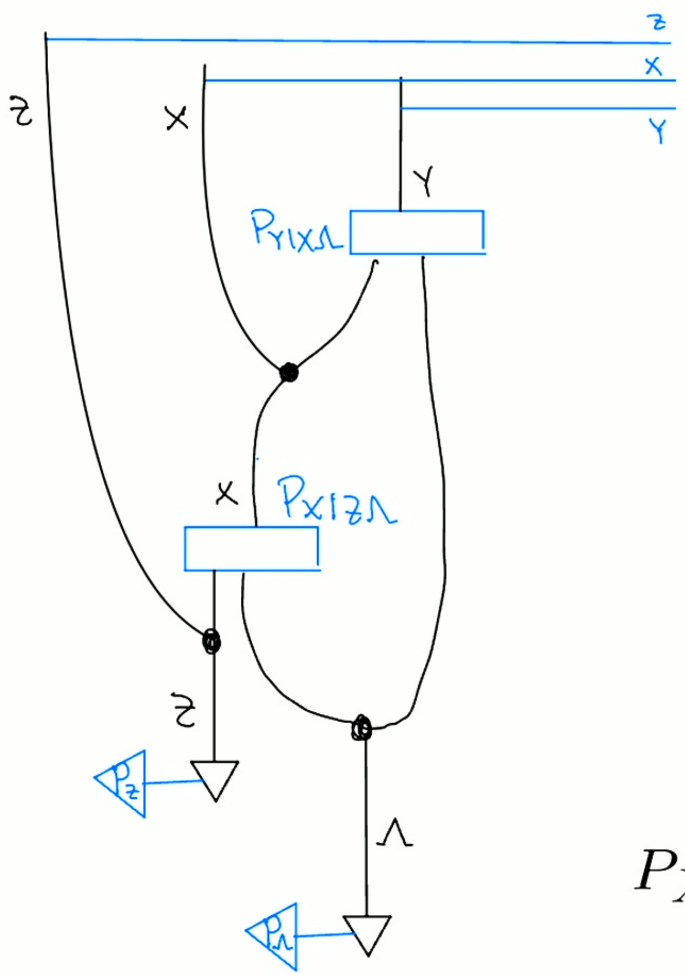


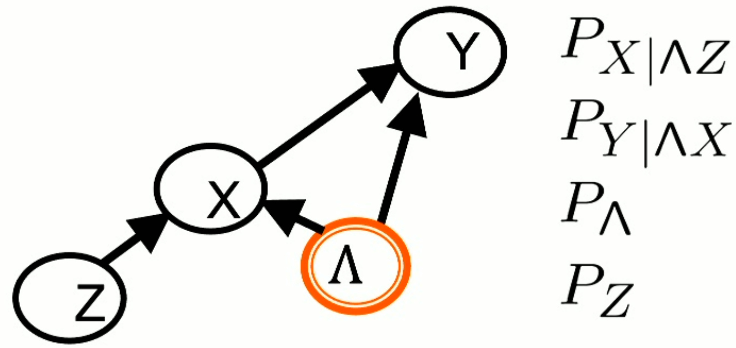


Shorthand involving scrambling of causal and inferential elements
$P_{Y|X}$ becomes a parameter

Unobserved variables that influence just one node are dropped
$P_{Y|X}$ becomes a parameter

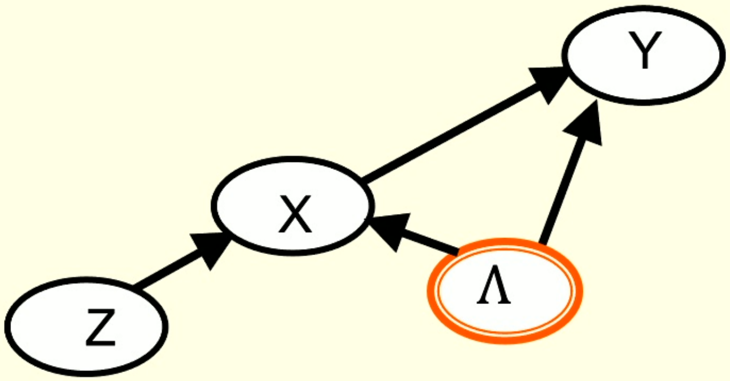
Bayesian causal model where a common cause is unobserved





$$P_{XYZ} = \sum_\Lambda P_{Y|X\Lambda} P_{X|Z\Lambda} P_\Lambda P_Z$$

Causal structure | Parameters



$$P_{X|\Lambda Z}$$
$$P_{Y|\Lambda X}$$
$$P_{\Lambda}$$
$$P_{Z}$$

$$P_{XYZ} = \sum_{\Lambda} P_{Y|X\Lambda} P_{X|Z\Lambda} P_{\Lambda} P_{Z}$$