Title: Causal Inference Lecture - 230308

Speakers: Robert Spekkens

Collection: Causal Inference: Classical and Quantum

Date: March 08, 2023 - 10:00 AM

URL: https://pirsa.org/23030070

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

How to define causation

“We have no other notion of cause and effect, but that of certain objects, which have always conjoin'd together, and which in all past instances have been found inseparable. We cannot penetrate into the reason of the conjunction. We only observe the thing itself, and always find that from the constant conjunction the objects acquire an union in the imagination.”
---David Hume

First candidate definition of A causes B

$$P_{B|A} \neq P_B$$

Equivalent to:

$$P_{A|B} \neq P_A$$

$$P_{AB} \neq P_A P_B$$

Problem: the condition is symmetric between A and B

$$P(\text{crows}|\text{sunrise}) \neq P(\text{crows})$$

$$P(\text{sunrise}|\text{crows}) \neq P(\text{sunrise})$$

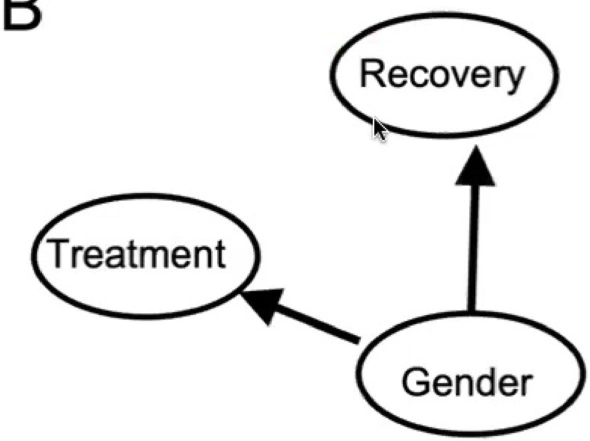
correlation is a symmetric relation

Second candidate definition of A causes B

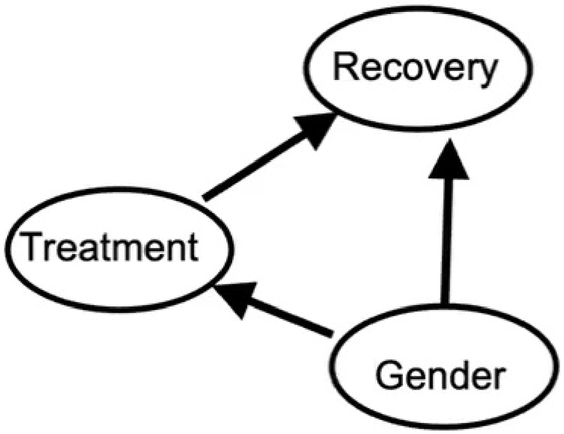
$$P_{B|A} \neq P_B$$

where is A is antecedent to B in time

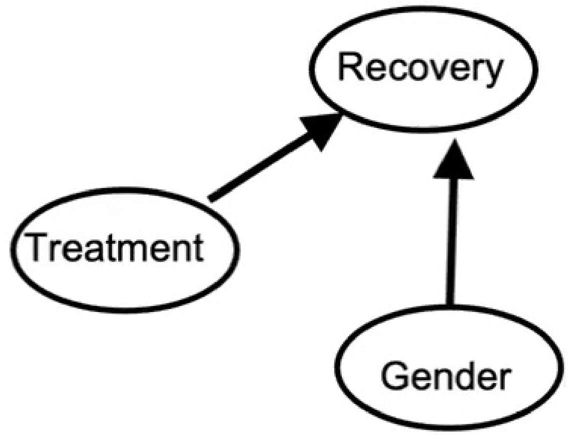
But correlation can be achieved even though A does not cause B



Actual world



Counterfactual



Standard conditional

$$P_{R|T} = \sum_G P_{R|TG} P_{G|T}$$

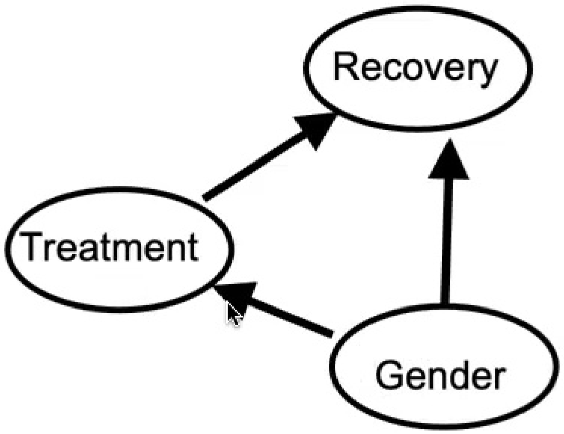
"Do conditional"

$$P_{R|\mathrm{do}T} = \sum_G P_{R|TG} P_G$$

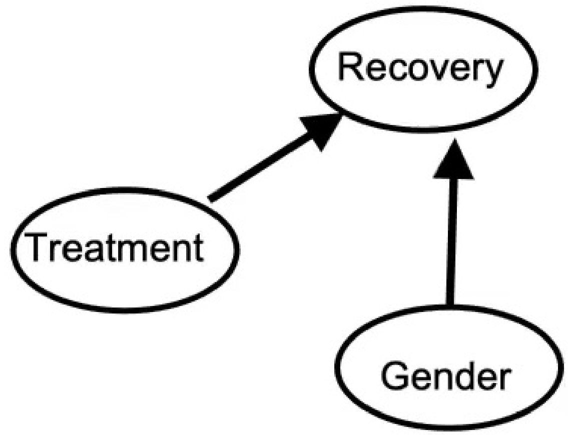
Third candidate definition of A causes B

$$P_{B|\mathrm{do}A} \neq P_B$$

Actual world



Counterfactual



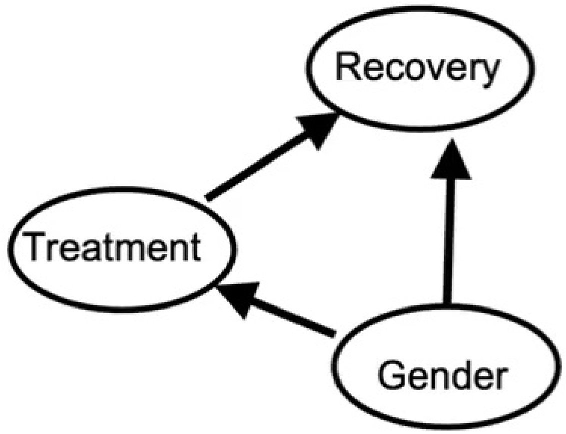
Standard conditional

$$P_{R|T} = \sum_G P_{R|TG} P_{G|T}$$

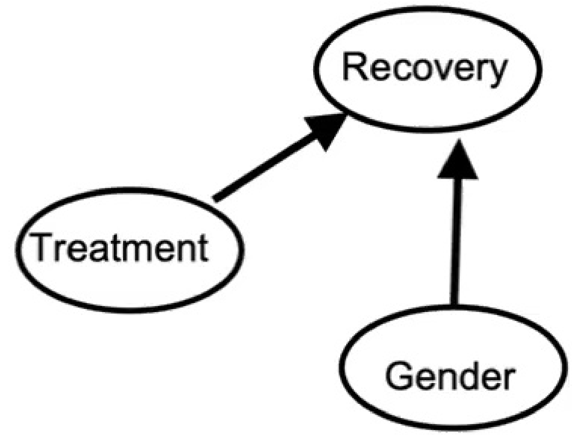
"Do conditional"

$$P_{R|\mathrm{do}T} = \sum_G P_{R|TG} P_G$$

Actual world



Counterfactual



Standard conditional

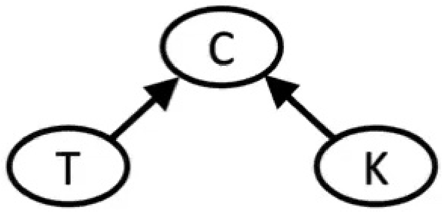
$$P_{R|T} = \sum_G P_{R|TG} P_{G|T}$$

"Do conditional"

$$P_{R|\mathrm{do}T} = \sum_G P_{R|TG} P_G$$

Vernam cypher

C = cyphertext
T = plaintext
K = key



$$C = (T + K)\mathrm{mod}2$$

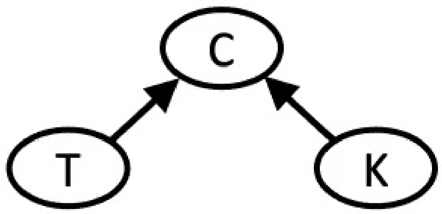
$$P_K = ½[0]_K + ½[1]_K$$

Vernam cypher

C = cyphertext

T = plaintext

K = key



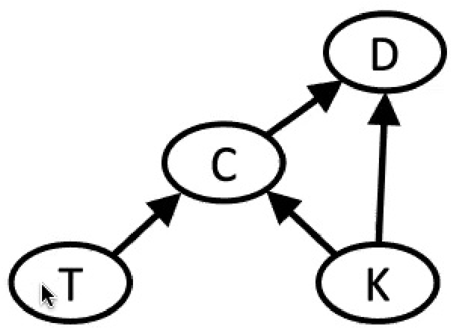
$C = (T + K)\text{mod}2$

$P_K = ½[0]_K + ½[1]_K$

$$P_{C|\text{do }T} = ½[0]_C + ½[1]_C$$
$$= P_C$$

yet T causes C!

If not, we could not decode T from C using K



D = decoded text

$D = (C + K)\text{mod}2$

Here, we take:

causal relationships between systems to be facts about the world (ontic)

probabilities to be the degrees of belief of a rational agent (epistemic)

Definition: Probability Distribution on variable A

Let A denote the variable and its set of values

$$P_A : A \to \mathbb{R} :: a \mapsto P_A(a)$$

$$\forall a \in A : 0 \leq P_A(a) \leq 1$$

$$\sum_{a \in A} P_A(a) = 1$$

Let $\mathcal{P}_A$ denote the set of distributions on A.

A distribution defines a vector on the reals

$$\vec{P}_A \in \mathbb{R}^{|A|}$$

$$\vec{P}_A = (P_A(0), P_A(1), \ldots, P_A(|A|))$$

Definition: Probability Distribution on variable A

Let A denote the variable and its set of values

$$P_A : A \to \mathbb{R} :: a \mapsto P_A(a)$$

$$\forall a \in A : 0 \leq P_A(a) \leq 1$$

$$\sum_{a \in A} P_A(a) = 1$$

Let $\mathcal{P}_A$ denote the set of distributions on A.

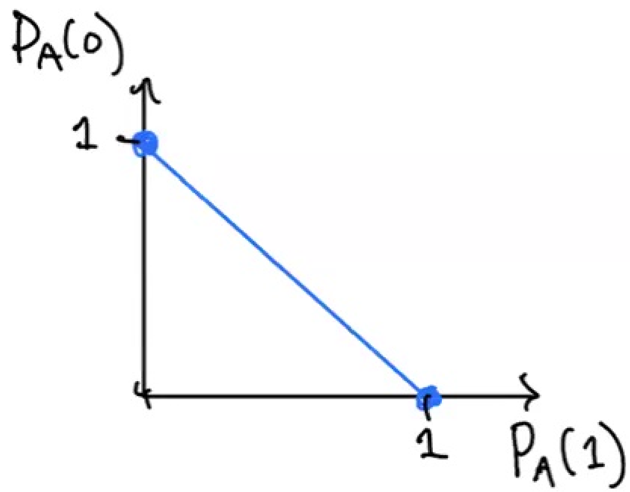
Some notation:

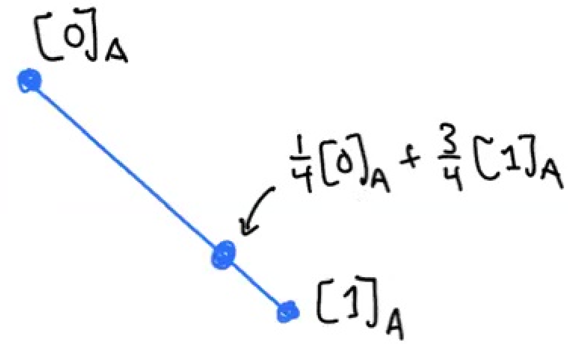
$P_A = [0]_A$ shorthand for $P_A = \delta_{A,0}$

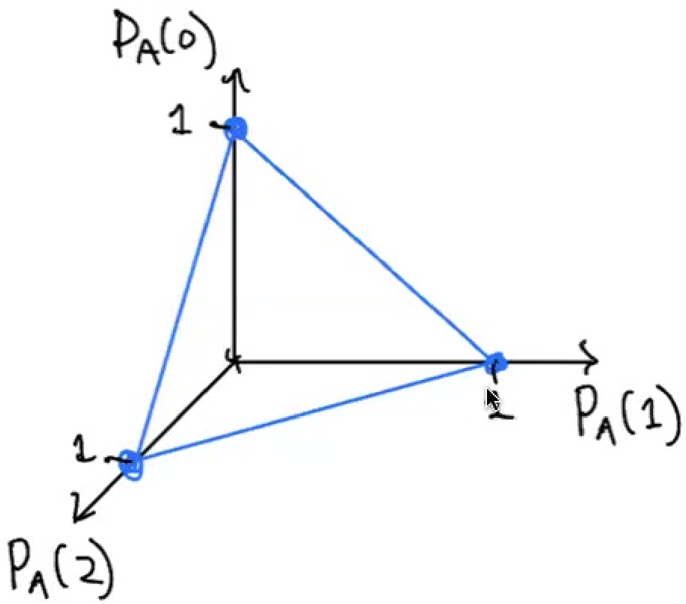
A distribution defines a vector on the reals

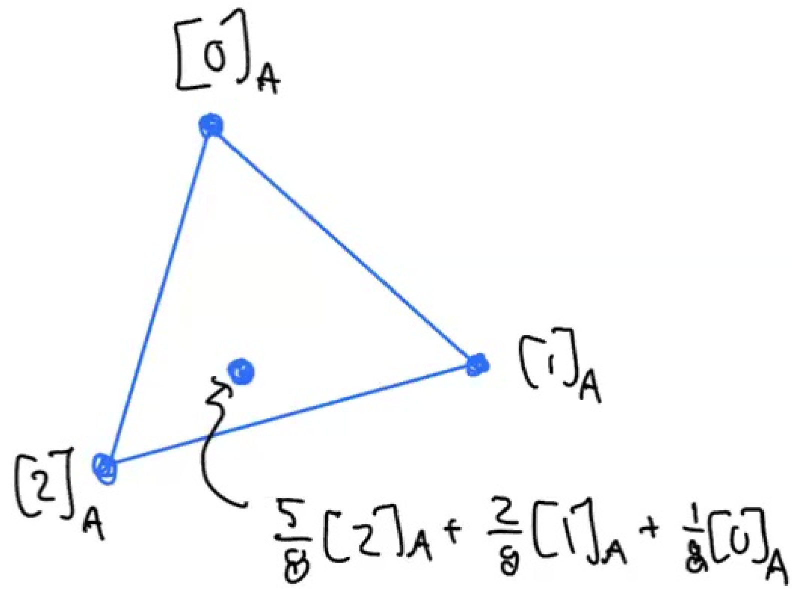
$$\vec{P}_A \in \mathbb{R}^{|A|}$$

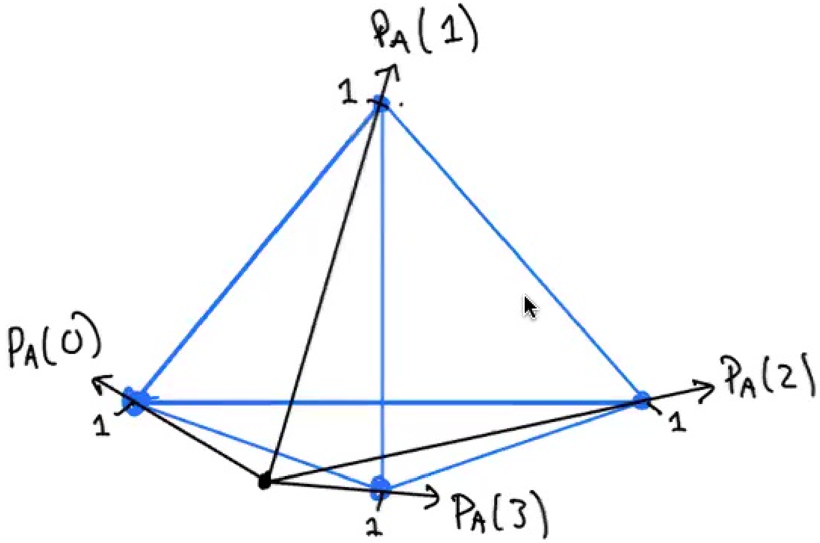
$$\vec{P}_A = (P_A(0), P_A(1), \ldots, P_A(|A|))$$



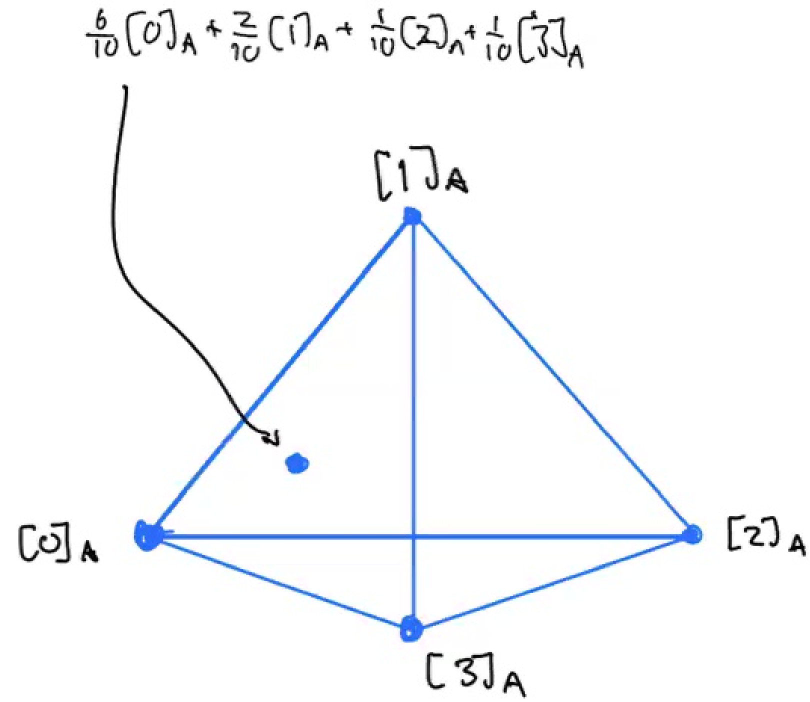



$P_A(0)$
1
1
1
$P_A(1)$
$P_A(2)$


$[0]_A$
$[1]_A$
$[2]_A$
$\frac{5}{8}[2]_A + \frac{2}{8}[1]_A + \frac{1}{8}[0]_A$



PA(1)
1
PA(0)
1
PA(2)
1
PA(3)
1



6/10 [0]A + 2/10 [1]A + 1/10 [2]A + 1/10 [3]A
[1]A
[0]A
[2]A
[3]A

Definition: Joint Probability Distribution on variables A and B

$$P_{AB} : A \times B \to \mathbb{R} :: (a,b) \mapsto P_{AB}(ab)$$

$$\forall (a,b) \in A \times B : 0 \leq P_{AB}(ab) \leq 1$$

$$\sum_{a \in A, b \in B} P_{AB}(ab) = 1$$

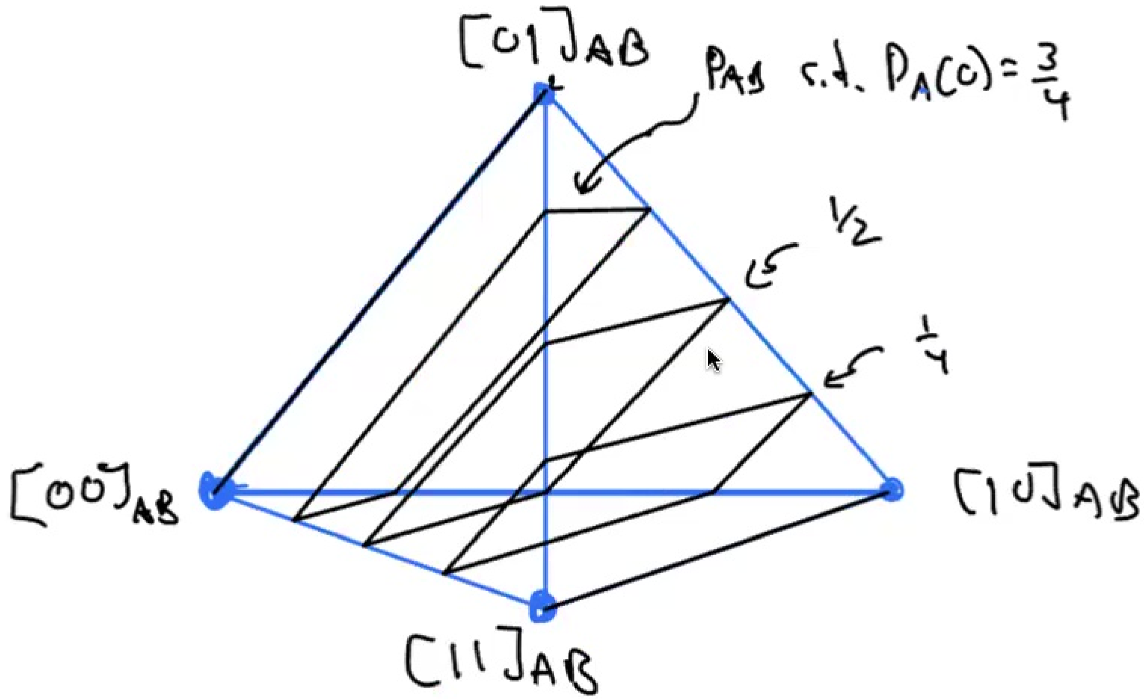
Definition: Marginal Distribution on A of a joint distribution on AB

$$\forall a \in A : P_A(a) := \sum_{b \in B} P_{AB}(ab)$$

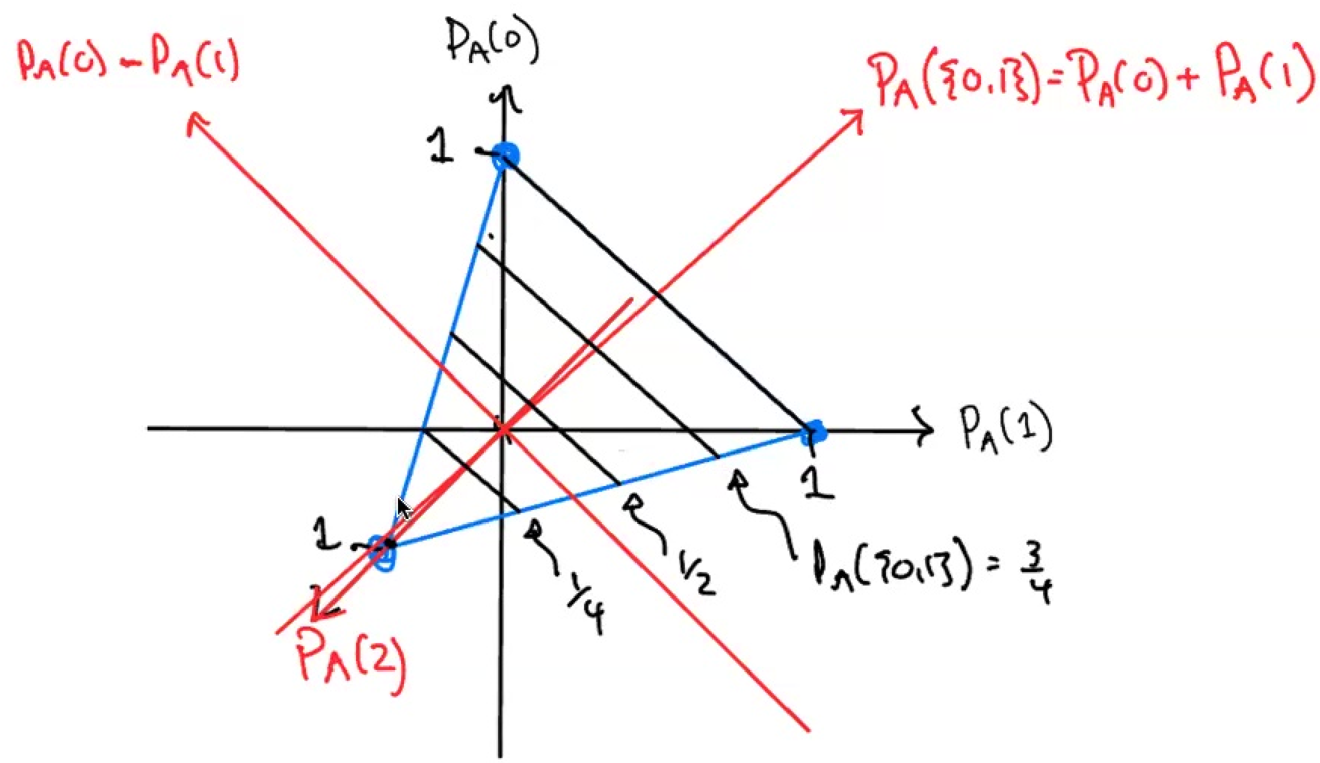
Definition: Marginalizing over B

$$\sum_B : \mathcal{P}_{AB} \to \mathcal{P}_A :: P_{AB} \mapsto P_A$$

What is marginalization geometrically?



Analogous to coarse-graining



Definition: Conditional Probability Distribution for A given B

$$P_{A|B} : A \times B \to \mathbb{R} :: (a,b) \mapsto P_{A|B}(a|b)$$

$$\forall (a,b) \in A \times B : 0 \leq P_{A|B}(a|b) \leq 1$$

$$\forall b \in B : \sum_{a \in A} P_{A|B}(a|b) = 1$$

The conditional that is defined by a joint distribution

$$P_{A|B} = P_{AB} P_B^{-1}$$

Taking Products

$$P_{ABC} = R_{ABC} S_{AB} T_C$$

means

$$\forall a, b, c : P_{ABC}(abc) = R_{ABC}(abc) S_{AB}(ab) T_C(c)$$

The conditional that is defined by a joint distribution

$$P_{A|B} = P_{AB} P_B^{-1}$$

Note: defined only for $b : P_B(b) > 0$

Check:

$$P_{A|B}(a|b) = \frac{P_{AB}(ab)}{P_B(b)} = \frac{P_{AB}(ab)}{\sum_a P_{AB}(ab)}$$

$$0 \leq P_{A|B}(a|b) \leq 1$$

$$\sum_a P_{A|B}(a|b) = 1$$

The conditional that is defined by a joint distribution

$$P_{A|B} = P_{AB} P_B^{-1}$$

The joint distribution that is defined by a conditional and a marginal

$$P_{AB} = P_{A|B} P_B$$

For any ordering of variables, one can write:

$$P_{ABCDE} = P_{A|BCDE} P_{B|CDE} P_{C|DE} P_{D|E} P_E$$

For any ordering of variables, one can write:

$$P_{ABCDE} = P_{A|BCDE} P_{B|CDE} P_{C|DE} P_{D|E} P_E$$

Proof:

$$P_{ABCDE} = P_{A|BCDE} P_{BCDE}$$
$$P_{BCDE} = P_{B|CDE} P_{CDE}$$
$$P_{CDE} = P_{C|DE} P_{DE}$$
$$P_{DE} = P_{D|E} P_E$$

Bayesian inversion

$$P_{A|B} = \frac{P_{B|A} P_A}{P_B}$$

Simple example

Disease variable

$D \in \{0, 1\}$

Test result variable

$T \in \{+, -\}$

Simple example

Disease variable

$$D \in \{0, 1\}$$

Test result variable

$$T \in \{+, -\}$$

Accuracy of test

$$P_{T|D}(+|1) = 0.95$$

False positive rate

$$P_{T|D}(+|0) = 0.02$$

Simple example

Disease variable

$$D \in \{0, 1\}$$

Test result variable

$$T \in \{+, -\}$$

Accuracy of test

$$P_{T|D}(+|1) = 0.95$$

False positive rate

$$P_{T|D}(+|0) = 0.02$$

You get a positive result. What probability should you assign to having the disease?

Simple example

Disease variable

$$D \in \{0, 1\}$$

Test result variable

$$T \in \{+, -\}$$

Accuracy of test

$$P_{T|D}(+|1) = 0.95$$

False positive rate

$$P_{T|D}(+|0) = 0.02$$

You get a positive result. What probability should you assign to having the disease?

$$P_{D|T} = \frac{P_{T|D} P_D}{P_T}$$

Simple example

Disease variable

$$D \in \{0, 1\}$$

Test result variable

$$T \in \{+, -\}$$

Accuracy of test

$$P_{T|D}(+|1) = 0.95$$

False positive rate

$$P_{T|D}(+|0) = 0.02$$

Base rate for disease

$$P_D(1) = 0.0001$$

You get a positive result. What probability should you assign to having the disease?

$$P_{D|T} = \frac{P_{T|D} P_D}{P_T}$$

$$P_{D|T}(1|+) = \frac{P_{T|D}(+|1)P_D(1)}{P_T(+)}$$

$$= \frac{P_{T|D}(+|1)P_D(1)}{P_{T|D}(+|1)P_D(1)+P_{T|D}(+|0)P_D(0)}$$

$$= \frac{0.95\times 0.0001}{0.95\times 0.0001+0.02\times 0.9999}$$

Accuracy of test

$$P_{T|D}(+|1) = 0.95$$

False positive rate

$$P_{T|D}(+|0) = 0.02$$

Base rate for disease

$$P_D(1) = 0.0001$$

$$P_{D|T}(1|+) = \frac{P_{T|D}(+|1)P_D(1)}{P_T(+)}$$

$$= \frac{P_{T|D}(+|1)P_D(1)}{P_{T|D}(+|1)P_D(1)+P_{T|D}(+|0)P_D(0)}$$

$$= \frac{0.95\times 0.0001}{0.95\times 0.0001+0.02\times 0.9999}$$

$$\simeq \frac{0.95}{0.02} \times 0.0001$$

$$\simeq 50 \times 0.0001$$

$$\simeq 0.005$$

$$\frac{1}{2}\%$$

Accuracy of test

$$P_{T|D}(+|1) = 0.95$$

False positive rate

$$P_{T|D}(+|0) = 0.02$$

Base rate for disease

$$P_D(1) = 0.0001$$

Def'n: A and B are marginally independent

$$P_{AB} = P_A P_B$$
$$P_{B|A} = P_B$$
$$P_{A|B} = P_A$$

Denote this

$$(A \perp B)$$

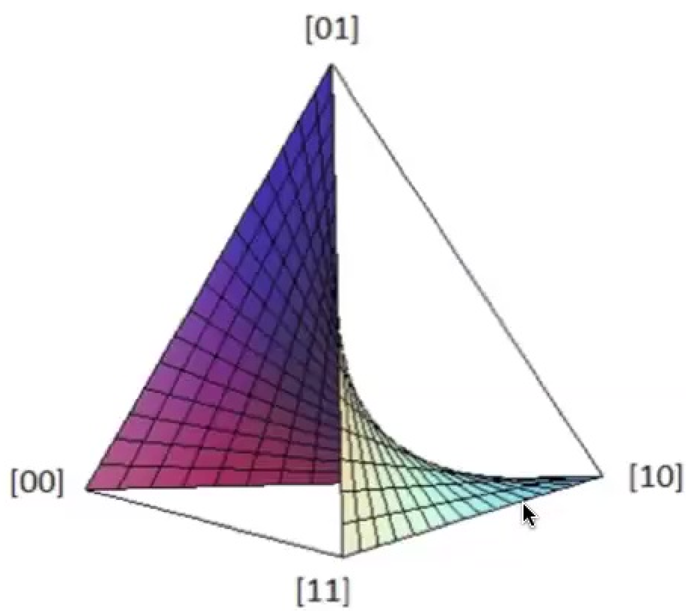
Expressed as a set of polynomial equality constraints on P_{AB}:

$$\forall a, b : P_{AB}(ab) = \left(\sum_{b'} P_{AB}(ab') \right) \left(\sum_{a'} P_{AB}(a'b) \right)$$

$$\forall a,b : P_{AB}(ab) = (\sum_b P_{AB}(ab))(\sum_a P_{AB}(ab))$$

3 independent constraints on the 4 parameters

$$(P_{AB}(00), P_{AB}(01), P_{AB}(10), P_{AB}(11))$$



Def'n: A and B are conditionally independent given C

$$P_{AB|C} = P_{A|C} P_{B|C}$$
$$P_{B|AC} = P_{B|C}$$
$$P_{A|BC} = P_{A|C}$$

Denote this
$(A \perp B|C)$

Expressed as a set of polynomial equality constraints on P_{ABC}

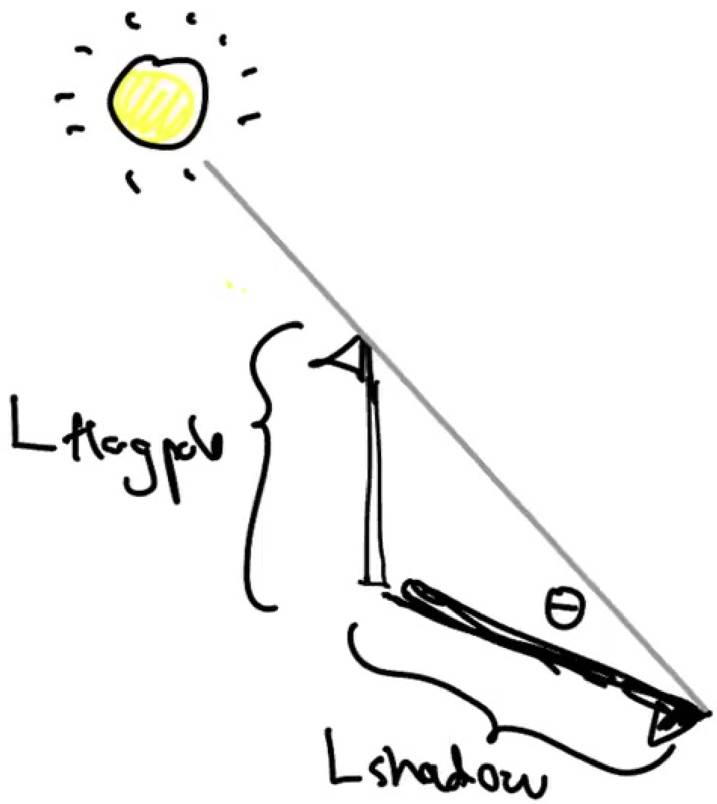
$$\forall a, b, c : P_{ABC}(abc) \left(\sum_{a'b'} P_{ABC}(a'b'c) \right) = \left(\sum_{b'} P_{ABC}(ab'c) \right) \left(\sum_{a'} P_{ABC}(a'bc) \right)$$

Marginal problem

Is there a distribution on A, B, C that has the following marginals?

$$P_{AB} = \frac{1}{2}[00] + \frac{1}{2}[11]$$
$$P_{AC} = \frac{1}{2}[00] + \frac{1}{2}[11]$$
$$P_{BC} = \frac{1}{2}[00] + \frac{1}{2}[11]$$

W. Salmon's flagpole example

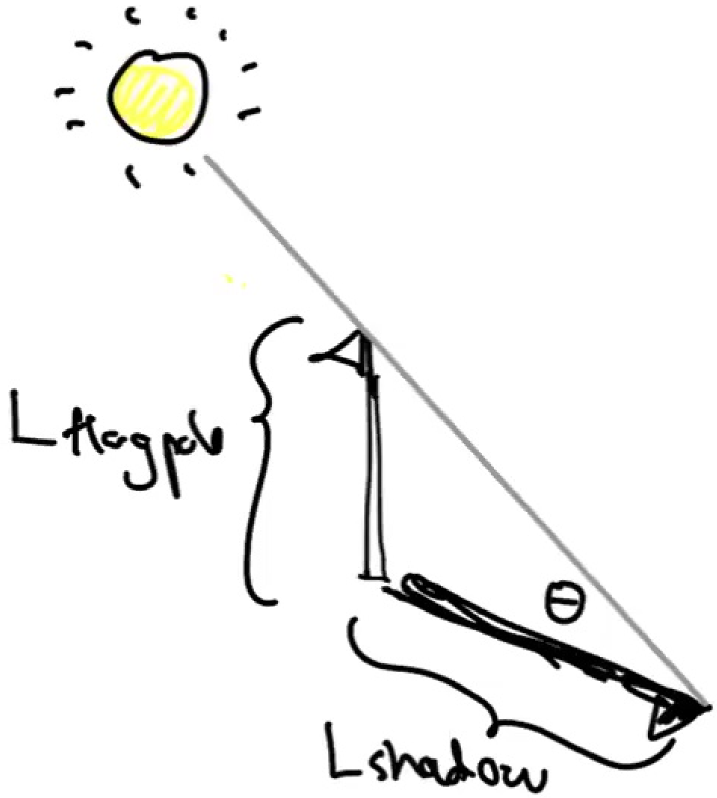


$$\frac{L_{\text{flagpole}}}{L_{\text{shadow}}} = \tan(\theta)$$

$$L_{\text{shadow}} = \frac{L_{\text{flagpole}}}{\tan(\theta)}$$

Inferential relations

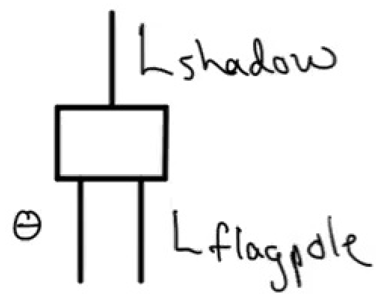
W. Salmon's flagpole example



$$\frac{L_{\text{flagpole}}}{L_{\text{shadow}}} = \tan(\theta)$$

$$L_{\text{shadow}} = \frac{L_{\text{flagpole}}}{\tan(\theta)}$$

Inferential relations



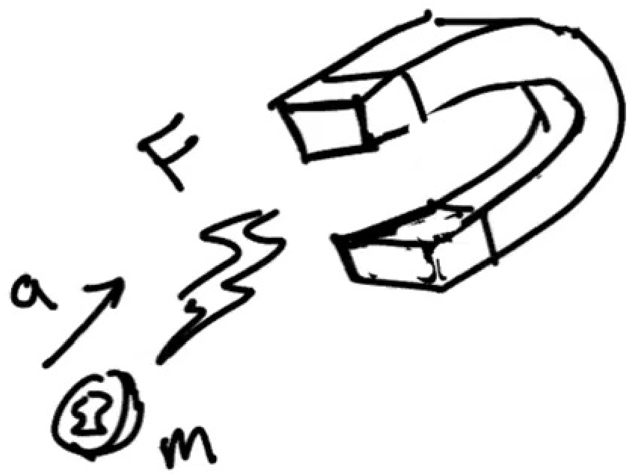
$$L_{\text{shadow}} = \frac{L_{\text{flagpole}}}{\tan(\theta)}$$

causal relation

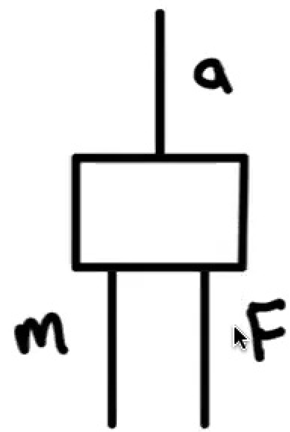
Newton's second law

$$F = ma$$

$$a = F/m$$



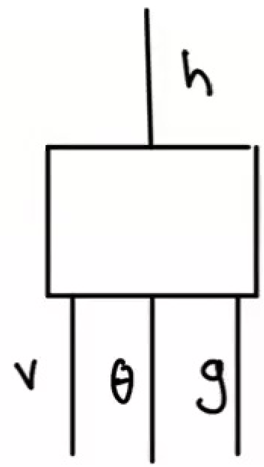
Inferential relations



$$a = F/m$$

Causal relation

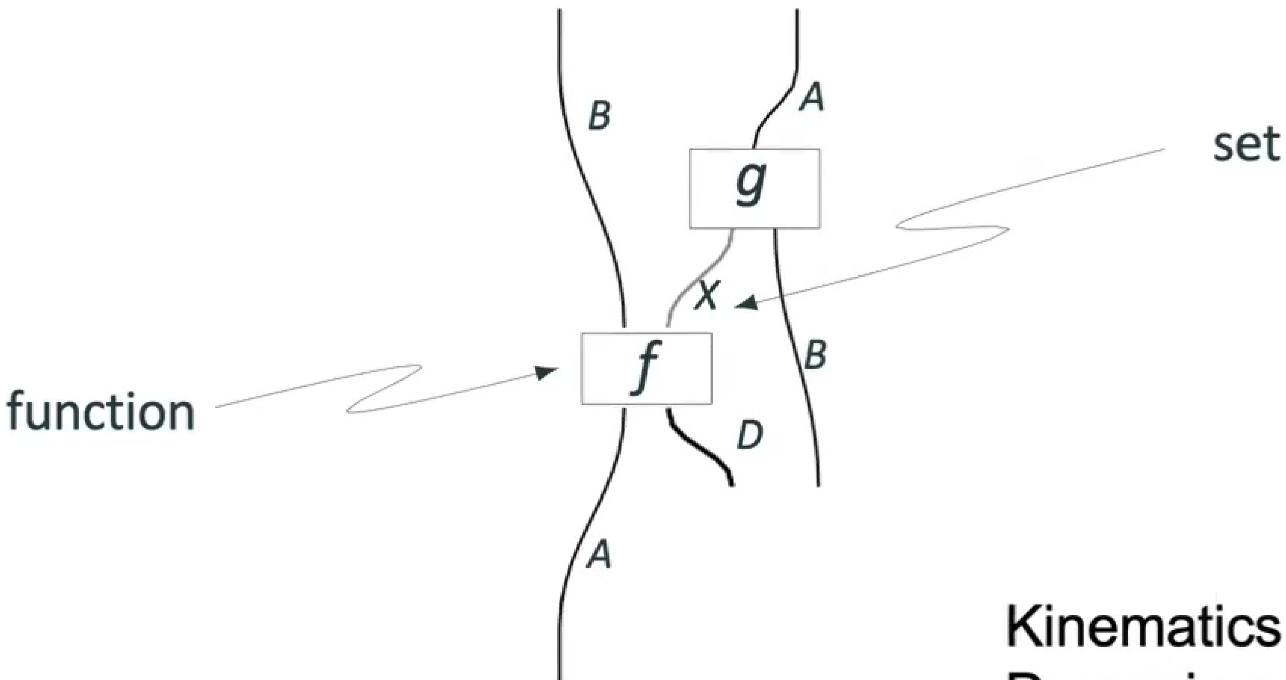
Law of projectile motion



$$h = \frac{v^2 \sin^2 \theta}{2g}$$

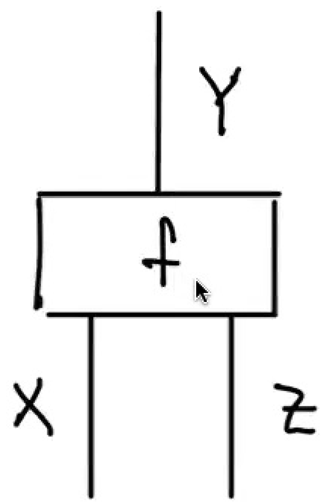
Causal laws warrant inferences about counterfactuals

Causal theory



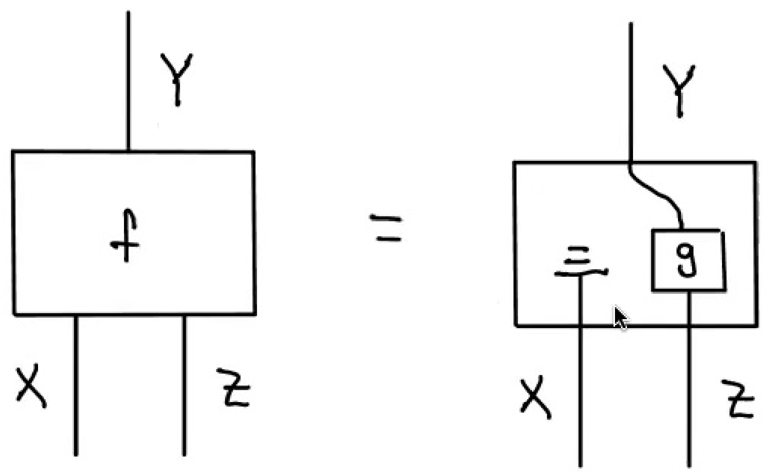
Kinematics: sets
Dynamics: functions

There can be other causal parents besides X



Definition: X has no influence on Y

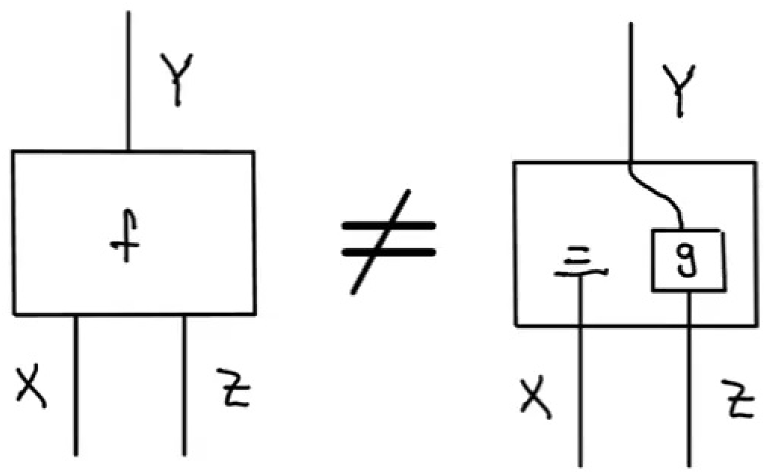
The function f that determines Y from its causal antecedents has a trivial dependence on X



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

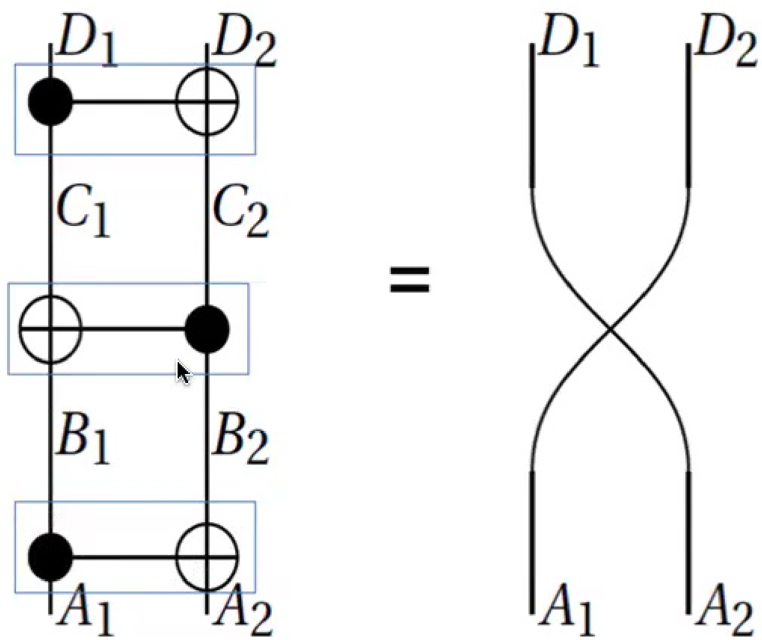
Definition: X has a nontrivial influence on Y

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

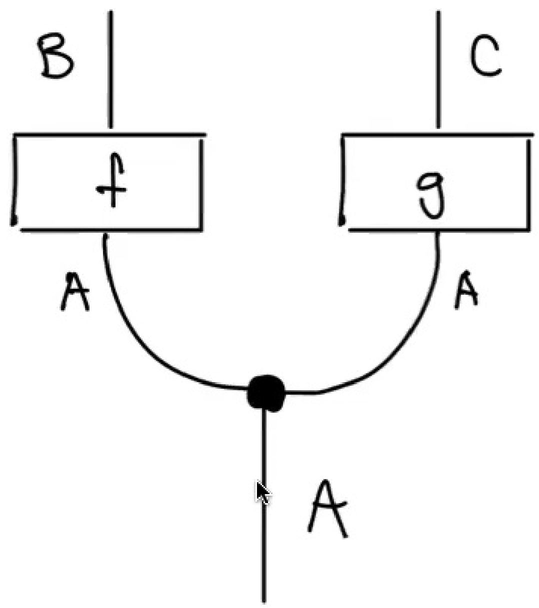


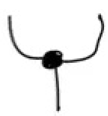
$$Y = f(X, Z) \neq g(Z)$$

Effective causal influences

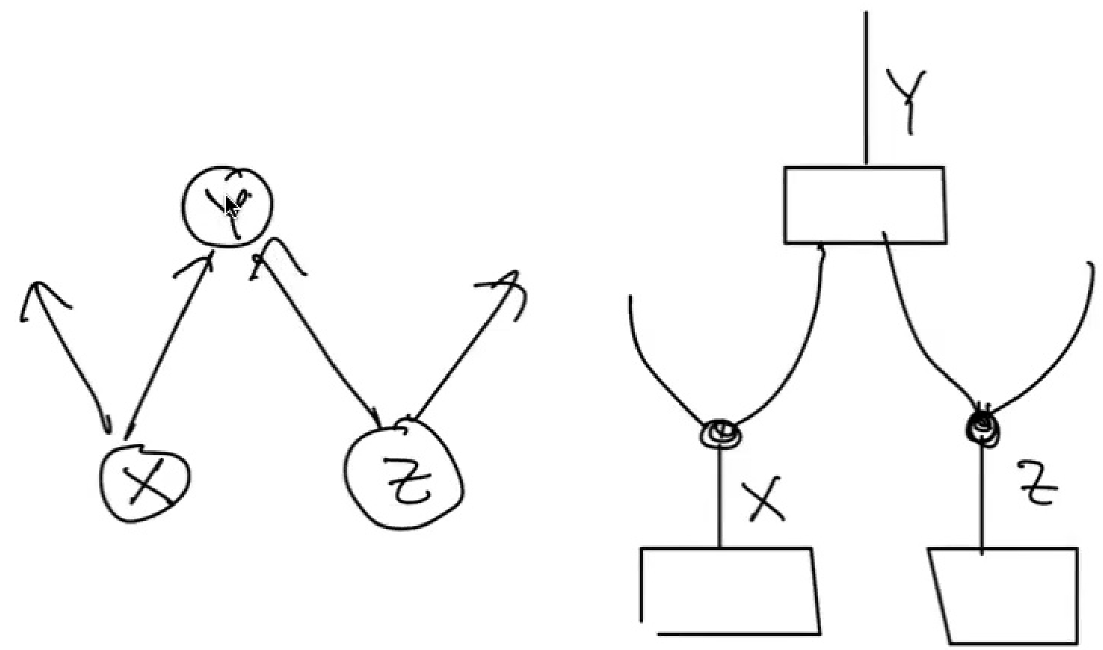


How to represent a variable that influences more than one other?



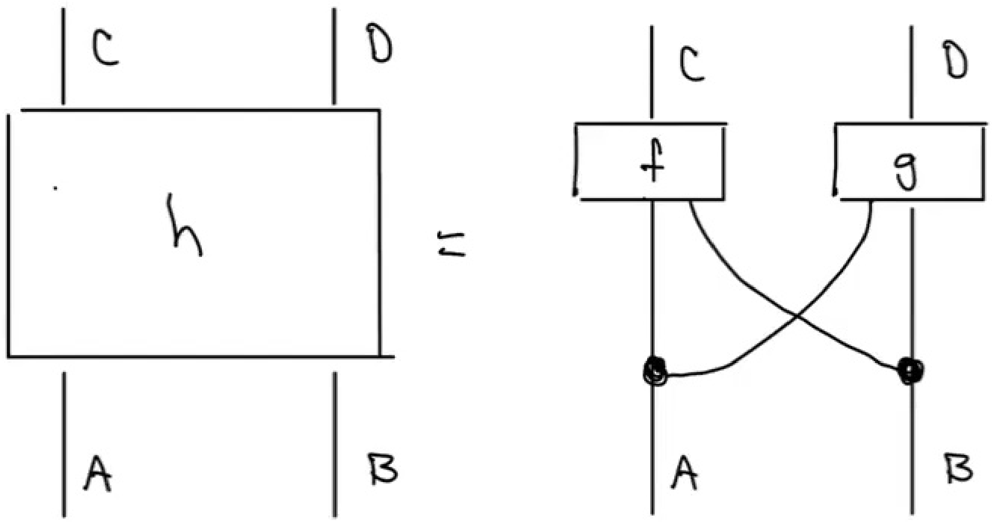
 $: A \to A \times A :: a \mapsto (a, a)$

Directed Acyclic graphs versus circuits



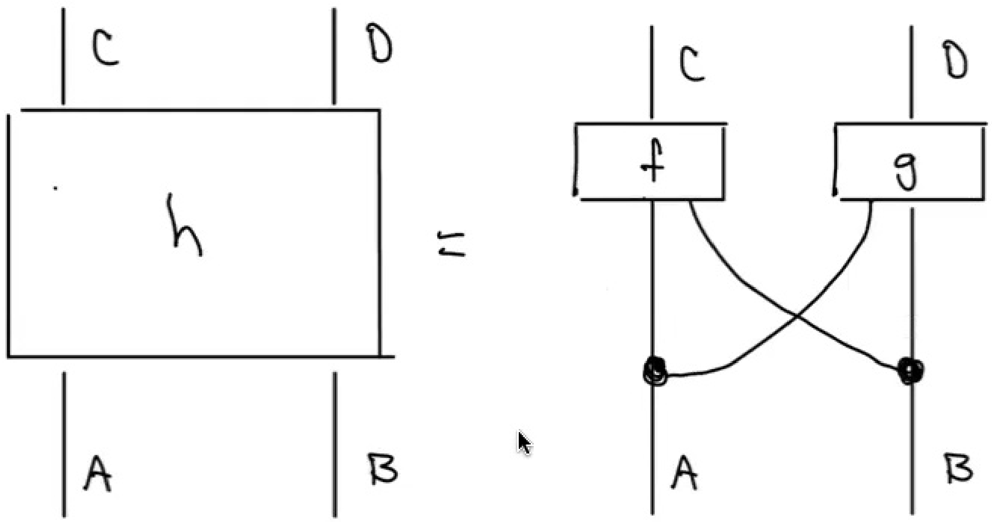
Note: A DAG usually does not imply deterministic dependences, whereas my circuits do.

Causal autonomy



where
f and g can be
specified
independently

Causal autonomy

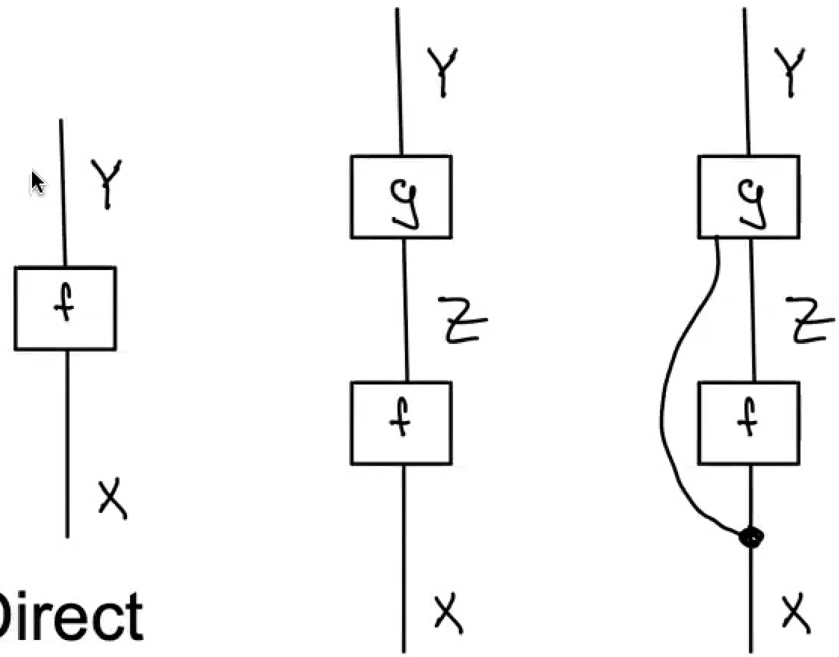


where
f and g can be
specified
independently

Causal autonomy fails in many circumstances:

- Symplectic dynamics
- Conservation laws

Direct versus indirect causal influence

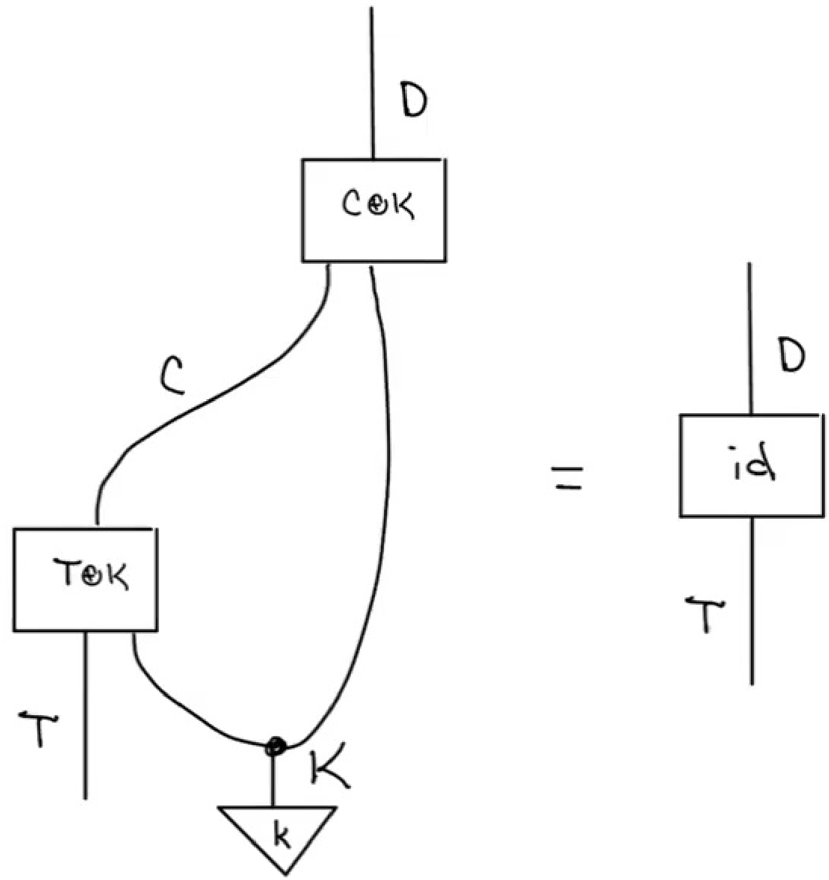


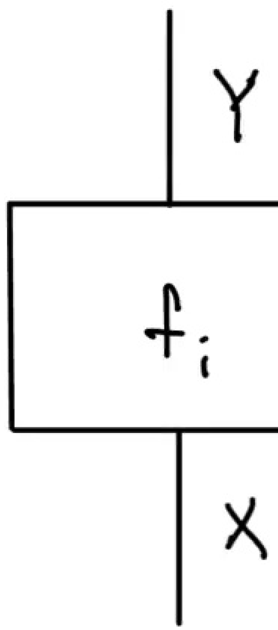
Direct

Indirect

Both direct and indirect

Causal structure of Vernam cypher





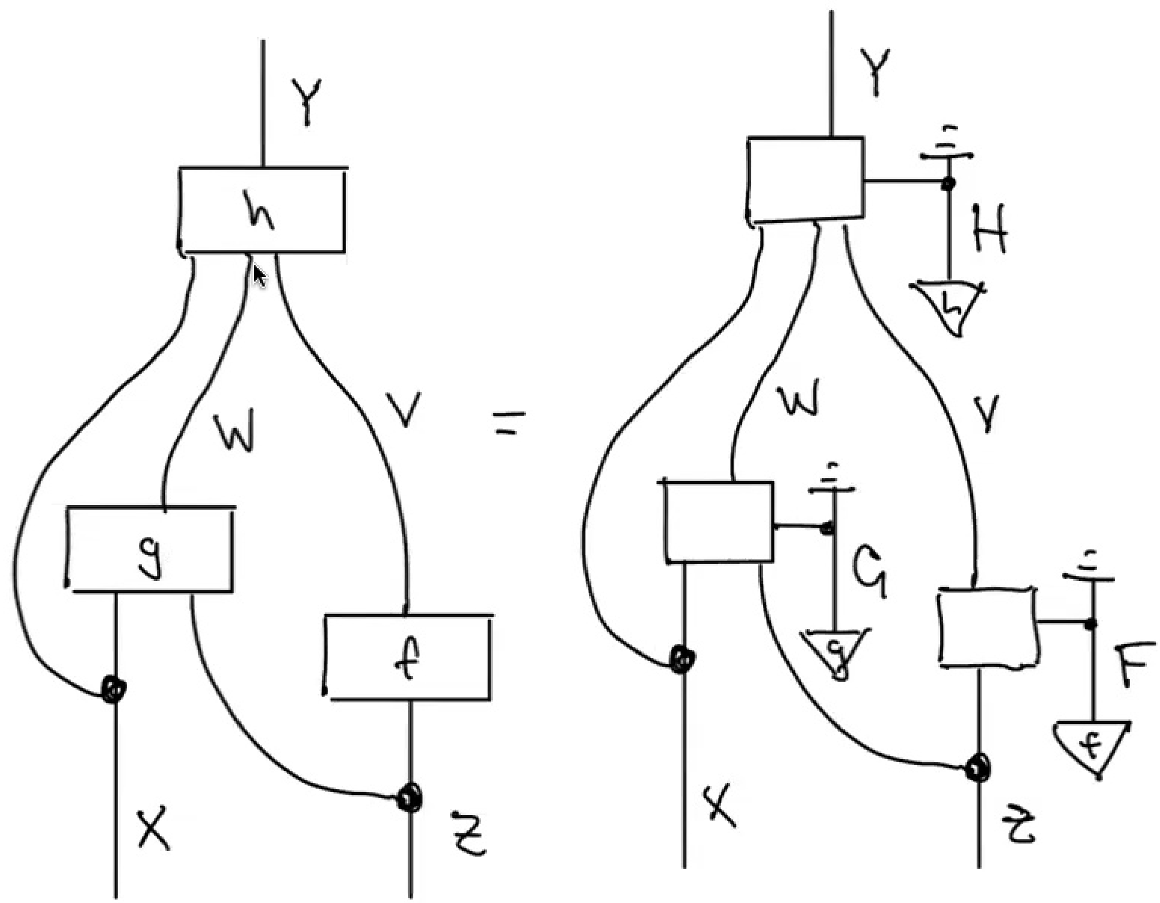
$$f_i \in \{ f_{id}, f_{flip}, f_{reset\text{-}0}, f_{reset\text{-}1} \}$$

$$f_{id}(x) = x$$

$$f_{flip}(x) = x \oplus 1$$

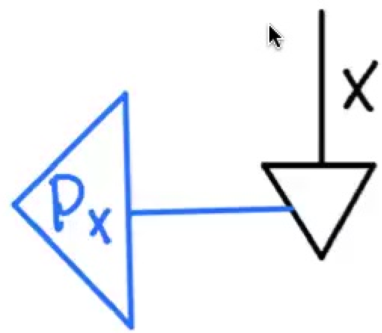
$$f_{reset\text{-}0}(x) = 0$$

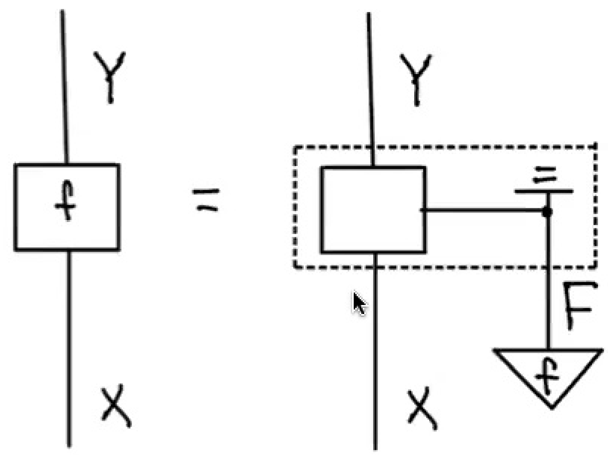
$$f_{reset\text{-}1}(x) = 1$$

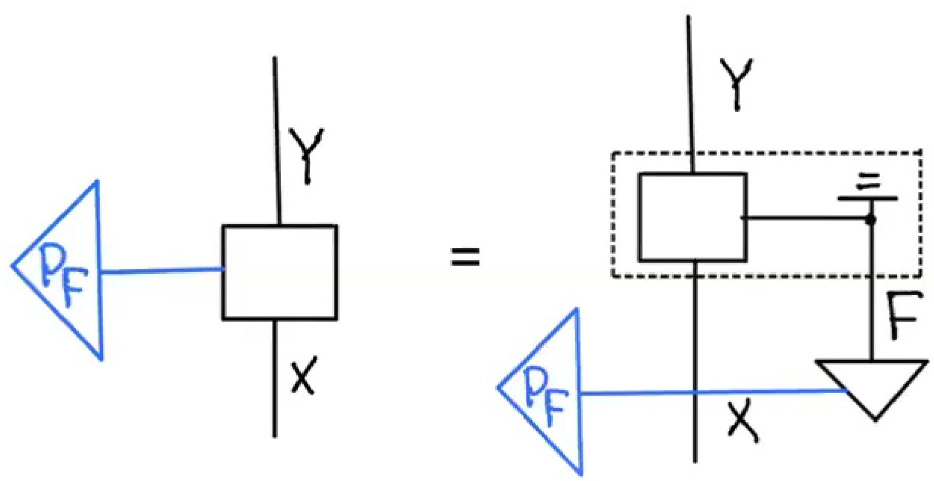


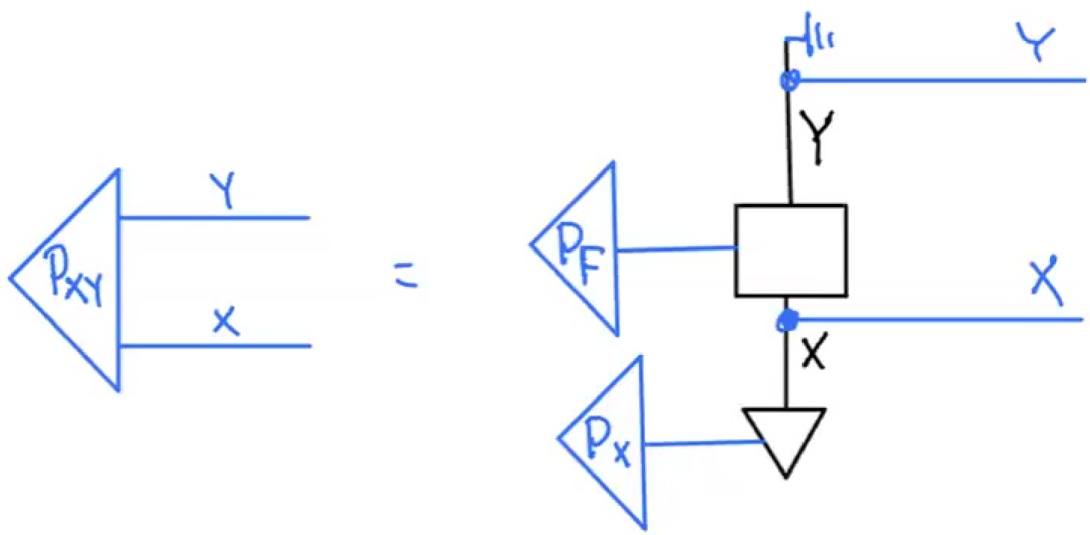
Y
h
W
V
g
f
X
Z
=
Y
H
h
W
V
G
g
F
f
X
Z

Uncertainty about causal mechanisms

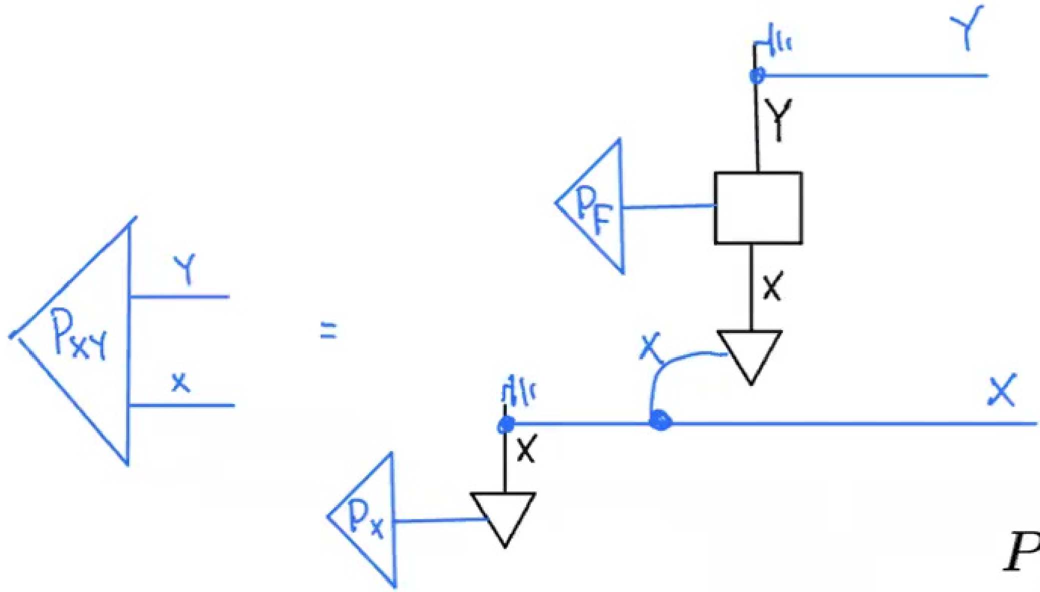

X
Px


Y
f
=
Y
=
F
f
X
X

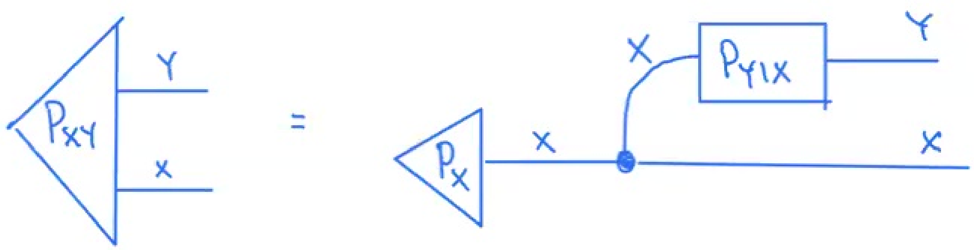

Y
P_F
=
Y
=
F
P_F
X
X



Pxy
Y
X
=
PF
Y
Y
X
X
Px



$$P_{Y|X} = \sum_f \delta_{Y,f(X)} P_F(f)$$



One's uncertainty about the causal mechanism is captured by P_F

One's uncertainty about the causal mechanism is captured by P_F

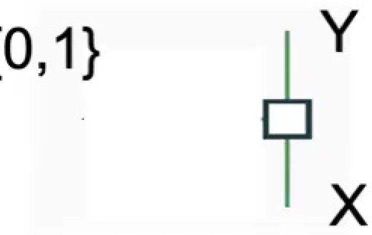
What can be estimated from statistical data is $P_{Y|doX}$

$$P_{Y|\mathrm{do}X} = \sum_f \delta_{Y,f(X)} P_F(f)$$

But there are many P_F consistent with a given $P_{Y|doX}$

Consider the four functions on the set {0,1}

$$f_{\text{id}}, f_{\text{flip}}, f_{\text{reset}-0}, f_{\text{reset}-1}$$



Now, consider two states of knowledge:

$$P_F = \tfrac{1}{2}[f_{\text{id}}] + \tfrac{1}{2}[f_{\text{flip}}] \qquad P'_F = \tfrac{1}{2}[f_{\text{reset}-0}] + \tfrac{1}{2}[f_{\text{reset}-1}]$$

$$P_{Y|\text{do}X} = \sum_f \delta_{Y,f(X)} P_F(f)$$

$$P_{Y|\text{do}X} = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad P_{Y|\text{do}X} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \qquad = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

One's uncertainty about the causal mechanism is captured by P_F

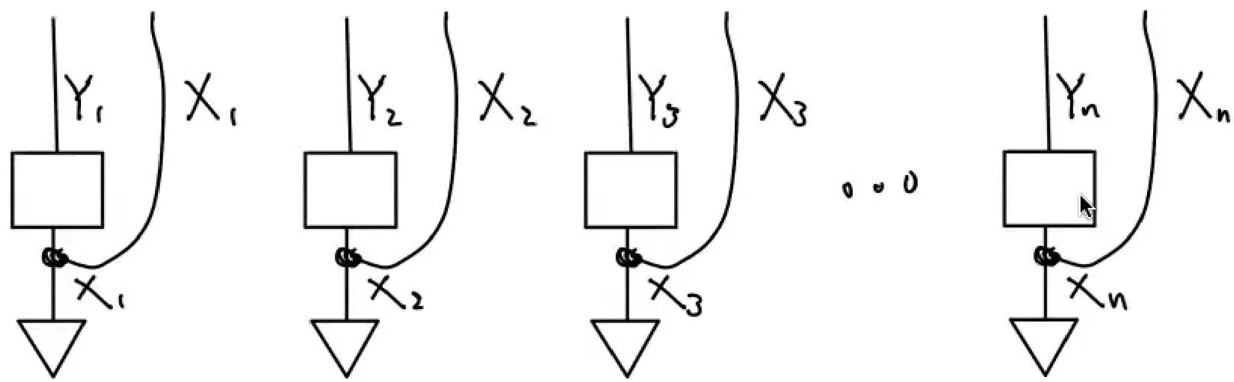
What can be estimated from statistical data is $P_{Y|\text{do}X}$

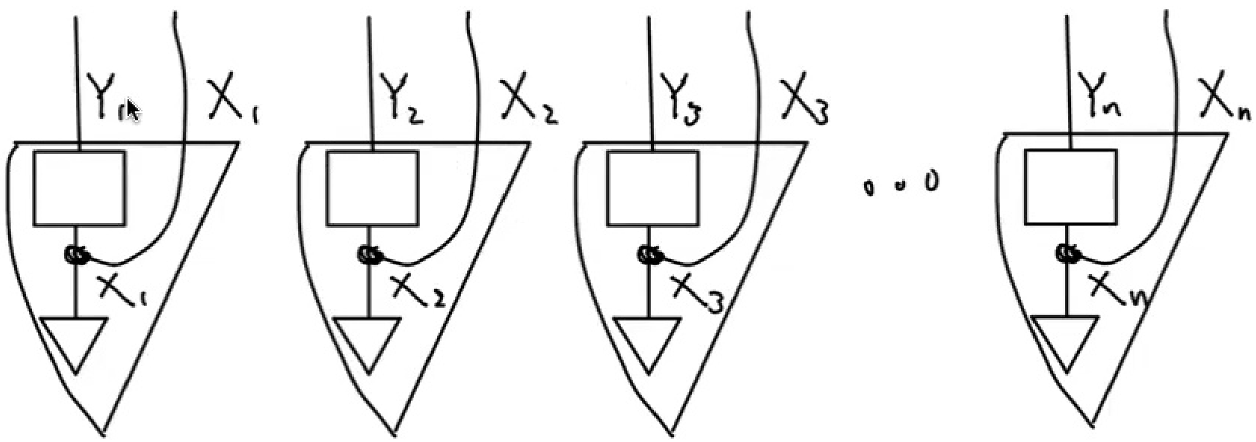
$$P_{Y|\text{do}X} = \sum_f \delta_{Y,f(X)} P_F(f)$$

But there are many P_F consistent with a given $P_{Y|\text{do}X}$

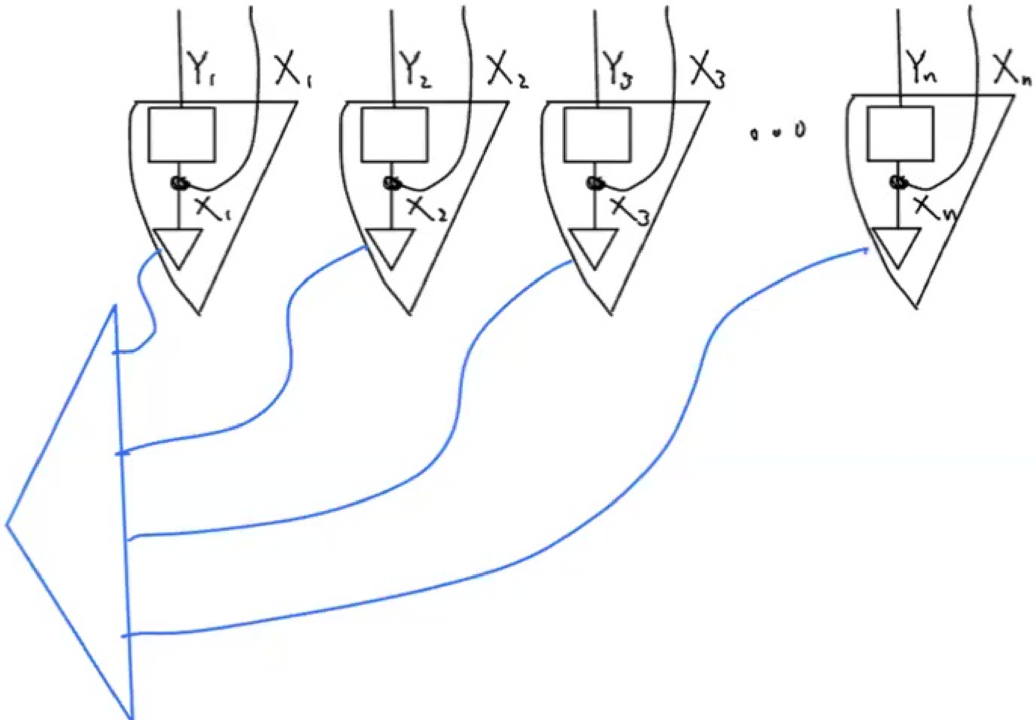
It is typical in the field of causal inference to settle for $P_{Y|\text{do}X}$

Estimating $P_{Y|doX}$ from statistical data


Y_1
X_1
X_1
Y_2
X_2
X_2
Y_3
X_3
X_3
Y_n
X_n
X_n

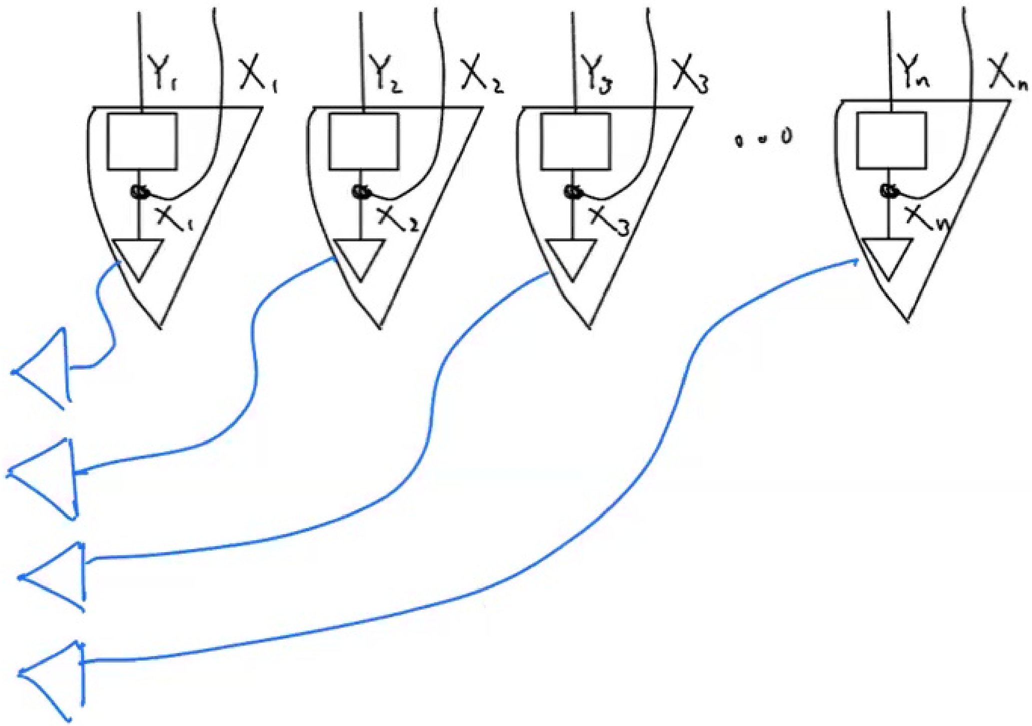


Y_1
X_1
X_1
Y_2
X_2
X_2
Y_3
X_3
X_3
...
Y_n
X_n
X_n



Under assumption of exchangeability, the distribution is de Finetti form

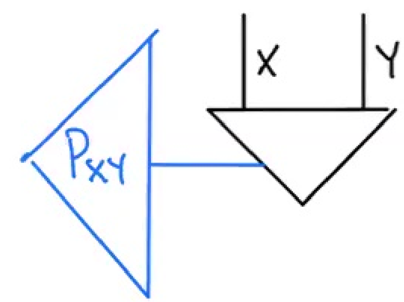
$$P_{X_1Y_1X_2Y_2\ldots X_nY_n} = \int \mathrm{d}Q_{XY}\,\mu(Q_{XY})\,Q_{X_1Y_1} \otimes Q_{X_2Y_2} \cdots \otimes Q_{X_nY_n}$$



For large data, the posterior is of the form

$$P_{X_1Y_1X_2Y_2\ldots X_nY_n} = P_{X_1Y_1} \otimes P_{X_2Y_2} \otimes \cdots \otimes P_{X_nY_n}$$

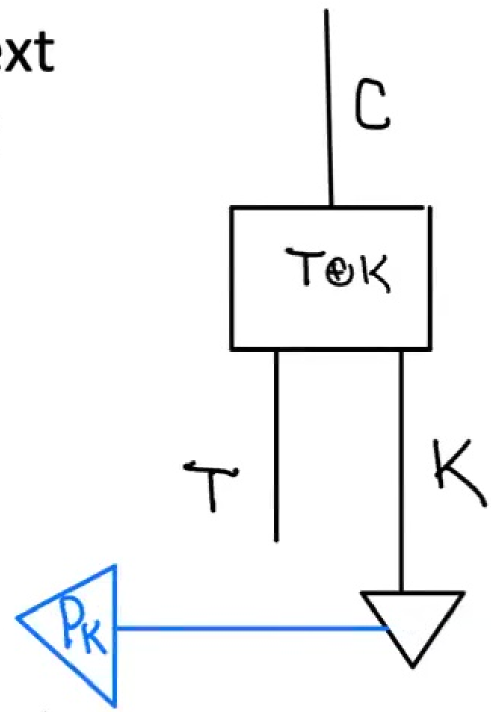
We write:



Vernam cypher

What is the causal influence of T on C?

C = cyphertext
T = plaintext
K = key



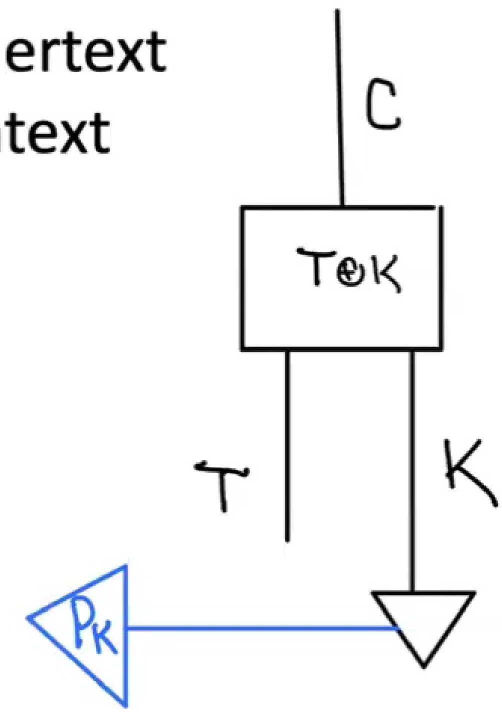
Recall: $P_{C|\mathrm{do}T} = P_C$

$P_K = \frac{1}{2}[0]_K + \frac{1}{2}[1]_K$

Vernam cypher

What is the causal influence of T on C?

C = cyphertext
T = plaintext
K = key



$$P_K = \frac{1}{2}[0]_K + \frac{1}{2}[1]_K$$

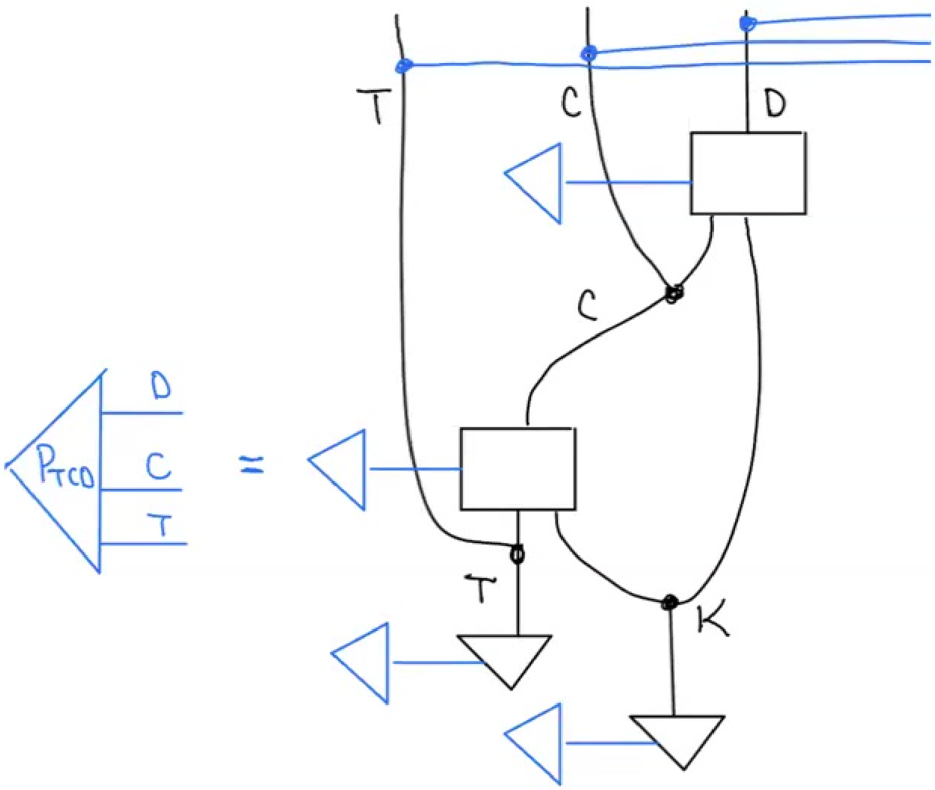
Let F denote the effective function from T to C
It is distributed as follows:

$$P_F = \frac{1}{2}[f_{\text{id}}] + \frac{1}{2}[f_{\text{flip}}]$$

If K is unobserved, this is the most we can say about influence of T on C

Vernam cypher

$$P_{TCD} = (\tfrac{1}{2}[00]_{TD} + \tfrac{1}{2}[11]_{TD})(\tfrac{1}{2}[0]_C + \tfrac{1}{2}[1]_C)$$



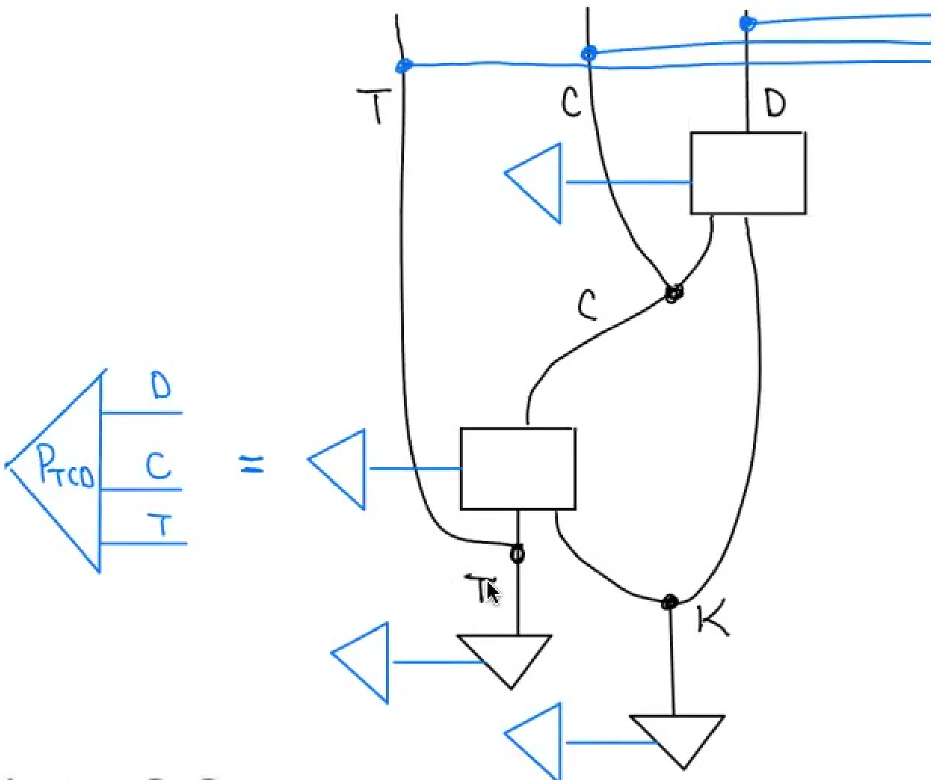
Let F be the effective function from T to C and G the effective function from C to D

The only way to account for the data is:

$$P_{FG} = \tfrac{1}{2}[f_{\text{id}}, f_{\text{id}}] + \tfrac{1}{2}[f_{\text{flip}}, f_{\text{flip}}]$$

Vernam cypher

$$P_{TCD} = (\tfrac{1}{2}[00]_{TD} + \tfrac{1}{2}[11]_{TD})(\tfrac{1}{2}[0]_C + \tfrac{1}{2}[1]_C)$$



Let F be the effective function from T to C and G the effective function from C to D

The only way to account for the data is:

$$P_{FG} = \tfrac{1}{2}[f_{\text{id}}, f_{\text{id}}] + \tfrac{1}{2}[f_{\text{flip}}, f_{\text{flip}}]$$

This implies

$$P_F = \tfrac{1}{2}[f_{\text{id}}] + \tfrac{1}{2}[f_{\text{flip}}]$$