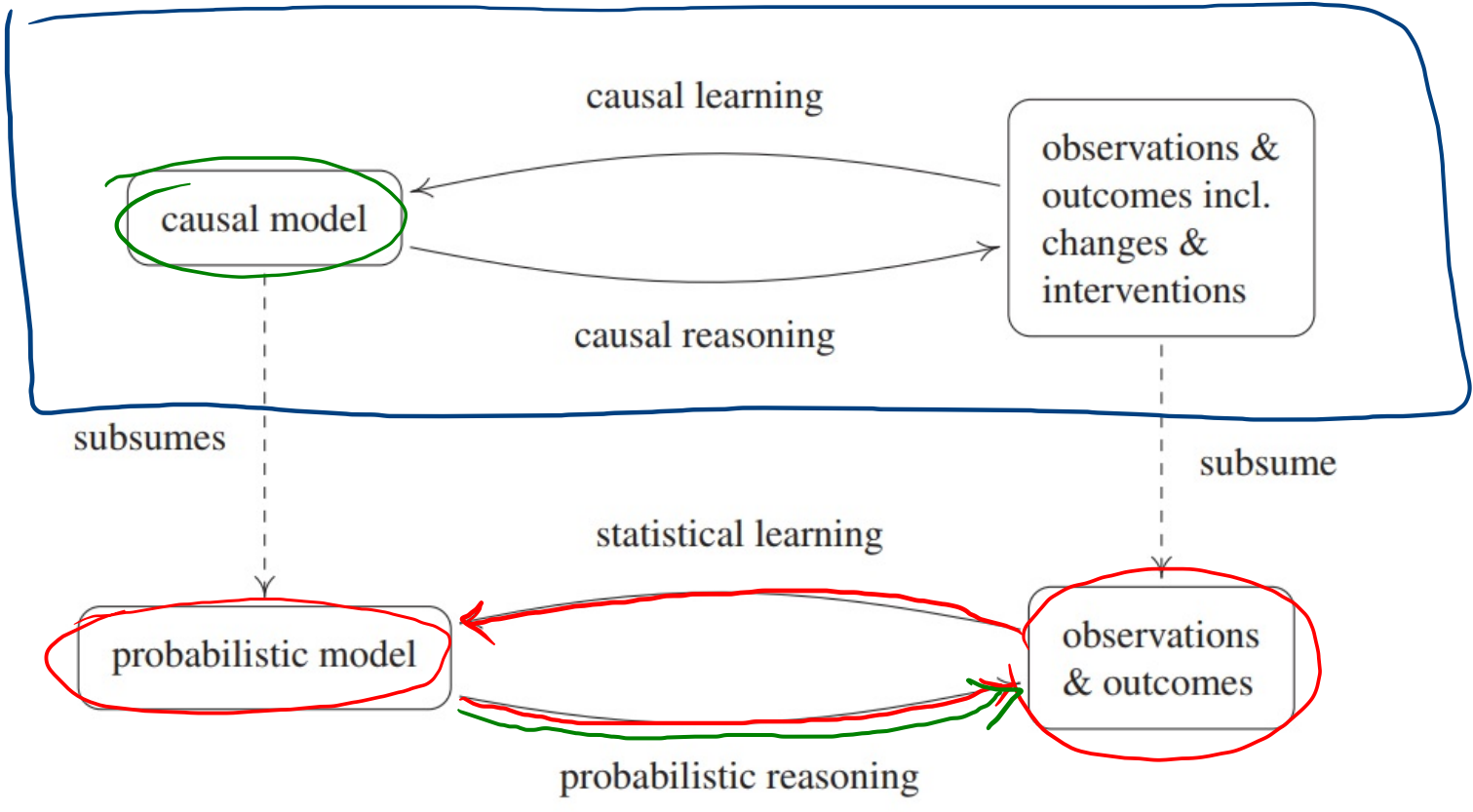
Causality

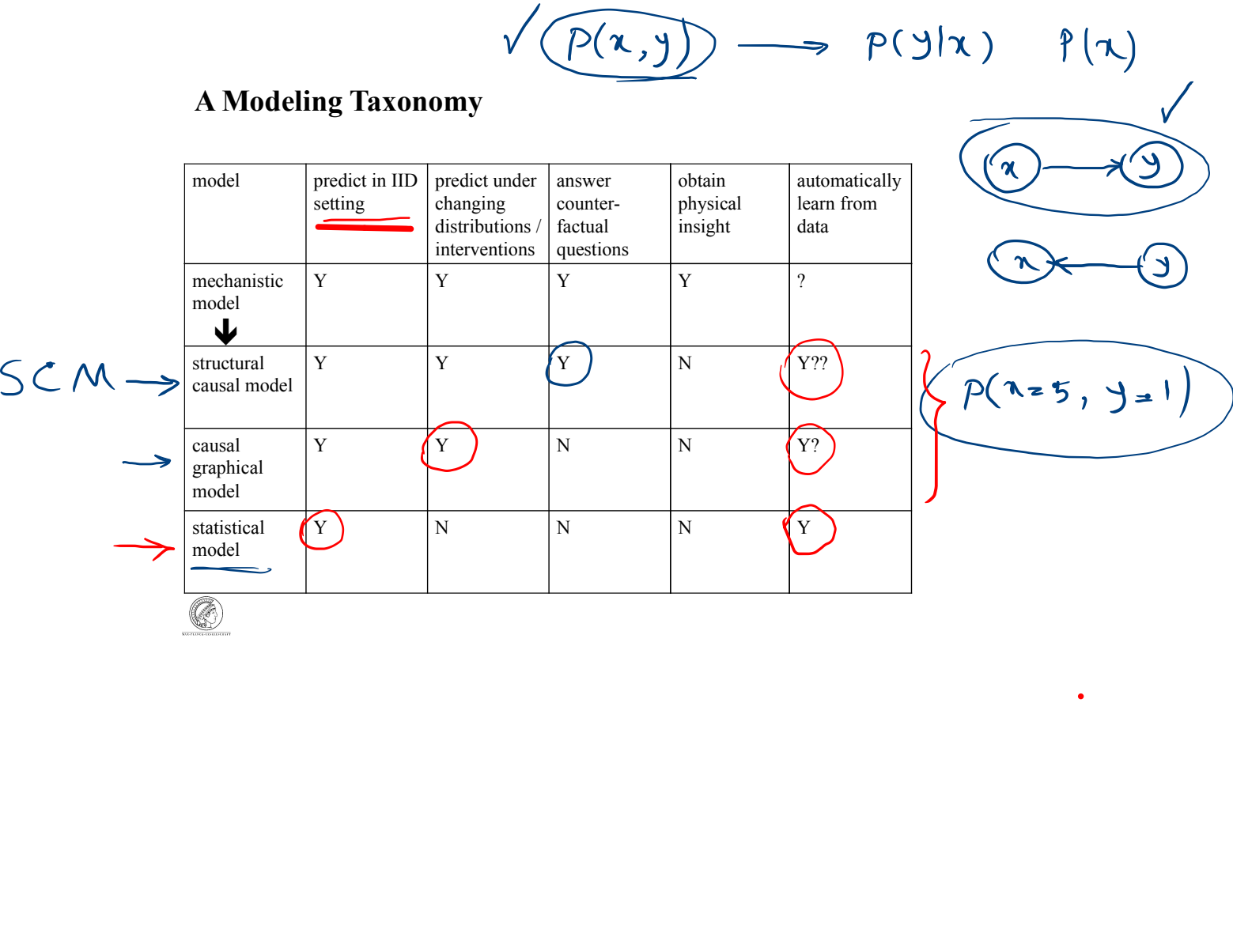
Dominik Janzing & Bernhard Schölkopf
Max Planck Institute for Intelligent Systems
Tübingen, Germany

http://ei.is.tuebingen.mpg.de


causal learning
causal model
observations & outcomes incl. changes & interventions
causal reasoning
subsumes
subsume
statistical learning
probabilistic model
observations & outcomes
probabilistic reasoning

A Modeling Taxonomy

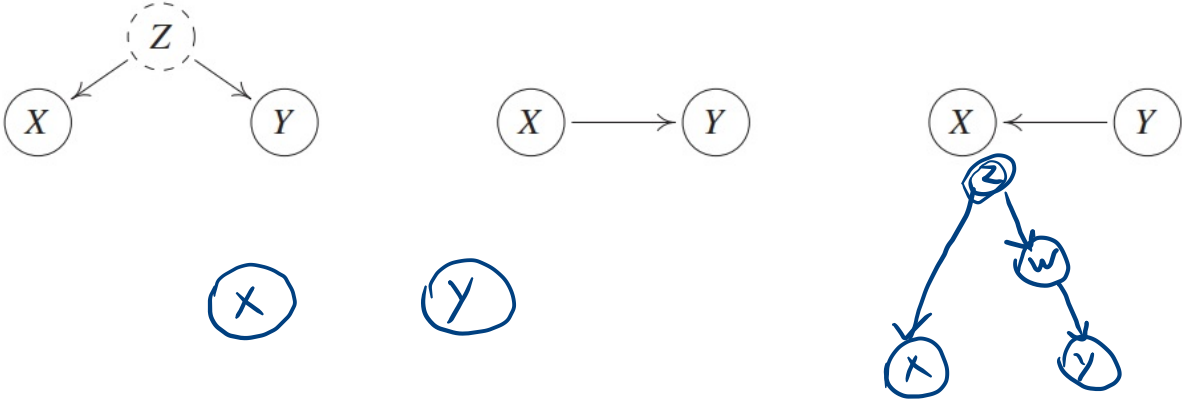
model	predict in IID setting	predict under changing distributions / interventions	answer counter-factual questions	obtain physical insight	automatically learn from data
mechanistic model	Y	Y	Y	Y	?
structural causal model	Y	Y	Y	N	Y??
causal graphical model	Y	Y	N	N	Y?
statistical model	Y	N	N	N	Y



“Correlation does not tell us anything about causality”

- Better to talk of dependence than correlation
- Most statisticians would agree that causality does tell us something about dependence
- But dependence does tell us something about causality too:

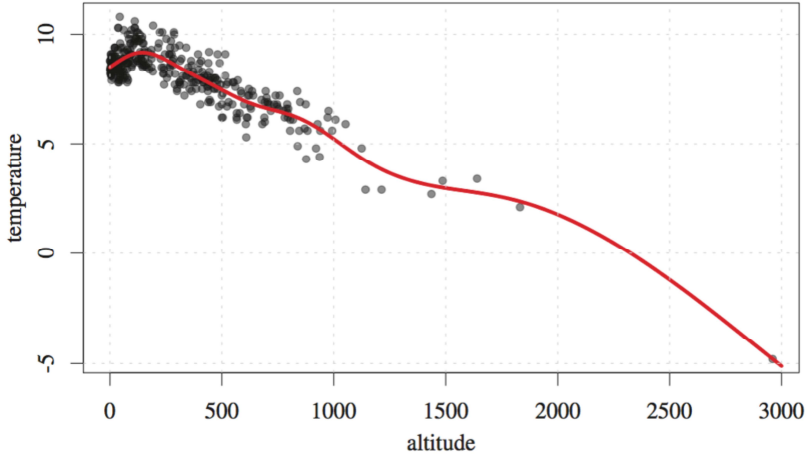
Principle 1.1 (Reichenbach's common cause principle) *If two random variables X and Y are statistically dependent ($X \not\perp\!\!\!\perp Y$), then there exists a third variable Z that* causally *influences both. (As a special case, Z may coincide with either X or Y.) Furthermore, this variable Z screens X and Y from each other in the sense that given Z, they become independent, $X \perp\!\!\!\perp Y \,|\, Z$.*



$$P(x,y) = p(x)\,p(y)$$

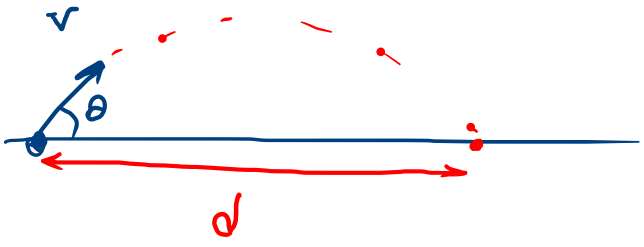
$$E[xy] = 0$$

What is cause and what is effect?



$$p(T,A) = p(A)\,p(T|A)$$

$$\begin{aligned} p(a,t) &= p(a|t)\,p(t) \qquad T \to A \\ &= p(t|a)\,p(a) \qquad A \to T \end{aligned}$$



$$p(v,d) = \underbrace{p(v)}_{M_1}\ \underbrace{p(d|v)}_{M_2}$$

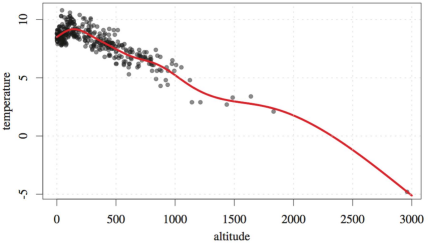
$$p(v) = \mathcal{N}(10, 2)$$

$$p(v) = \mathcal{N}(40, 5)$$

$$p(v)$$

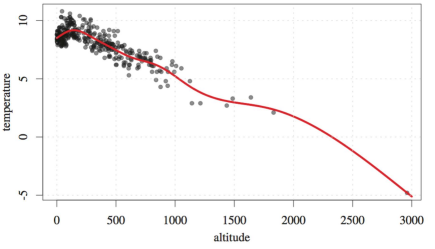
$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i \mid Pa(x_i))$$

Autonomous/invariant mechanisms



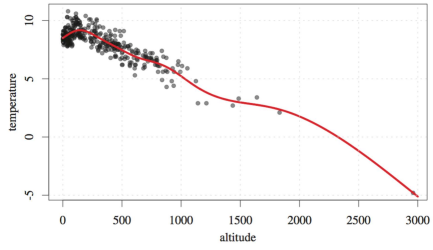
- intervention on a: raise the city, find that t changes
- hypothetical intervention on a: still expect that t changes, since we can think of a physical mechanism $p(t|a)$ that is independent of $p(a)$
- we expect that $p(t|a)$ is invariant across, say, different countries in a similar climate zone

Independence of cause & mechanism



- the conditional density $p(t|a)$ (viewed as a function of t and a) provides no information about the marginal density function $p(a)$
- this also applies if we only have a single density

Independence of noise terms



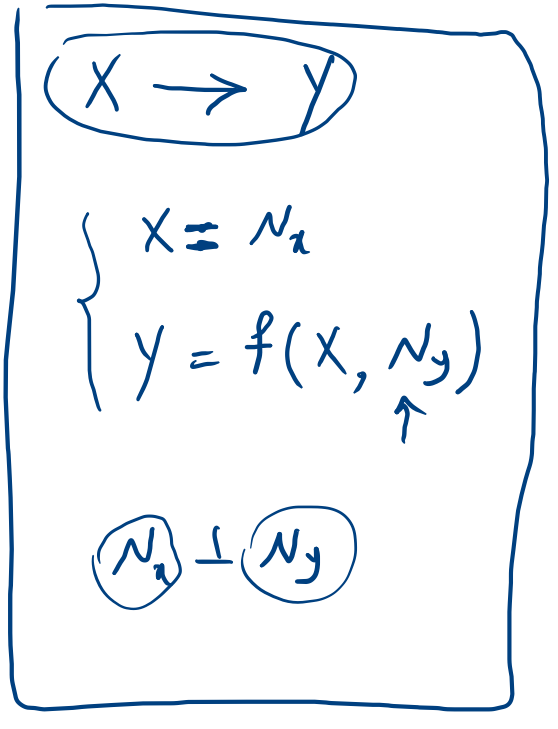
- view the distribution as entailed by a structural causal model (SCM)

$$\begin{cases} A := N_A, \\ T := f_T(A, N_T), \end{cases}$$

where $N_T \perp\!\!\!\perp N_A$

- this allows identification of the causal graph under suitable restrictions on the functional form of f_T

SCM



$X \sim P(X)$

$Y \to X$

$Y = N_Y$

$X = g(Y, N_X)$

$N_X \perp N_Y$

$p(x, y)$

$X \rightarrow Y$

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

$y = f(x, n_y)$

Pearl's do-notation

X Y

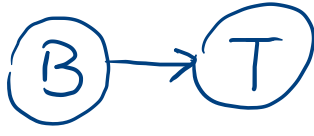
$P(y|X=x)$

$P(y|do\ X=x)$

- Motivation: goal of causality is to infer the effect of interventions

- distribution of Y given that X is set to x:

$$p(Y|do\,X=x) \quad \text{or} \quad p(Y|do\,x)$$



- don't confuse it with $P(Y|x)$
- can be computed from p and G

$P(T|B=1)$

$(B=0,\ T=20)$

$P(T|do\ B=1)$

y: یک تر ۱۰۰ سال عمر کند

x: روزی ۱۰ فنجان قهوه بخورد

Difference between seeing and doing

$$p(y|x)$$

probability that someone gets 100 years old given that we know that he/she drinks 10 cups of coffee per day

$$p(y|do\,x)$$

probability that some randomly chosen person gets 100 years old after he/she has been forced to drink 10 cups of coffee per day

The Principle of Independent Mechanisms

Principle 2.1 (Independent mechanisms) *The causal generative process of a system's variables is composed of autonomous modules that do not inform or influence each other.*

In the probabilistic case, this means that the conditional distribution of each variable given its causes (i.e., its mechanism) does not inform or influence the other conditional distributions. In case we have only two variables, this reduces to an independence between the cause distribution and the mechanism producing the effect distribution.

Example 3.2 (Cause-effect interventions) Suppose that the distribution $P_{C,E}$ is entailed by an SCM $\mathfrak{C}$

$$C := N_C$$
$$E := 4 \cdot C + N_E, \tag{3.3}$$

with $N_C, N_E \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$, and graph $C \to E$. Then,

$P(C) = \mathcal{N}(0,1)$

$P(E) = \mathcal{N}(0, 17)$

$$P(C \mid E=2) = \frac{P(C, E=2)}{P(E=2)} = \frac{\quad}{\mathcal{N}(2 \mid 0, 17)}$$

$$P(C,E) = P(C)\,P(E \mid C) = \mathcal{N}(0,1)\,\mathcal{N}(4C, 1)$$

$$P(C \mid do\ E=2) = P(C)$$

$P(C)$ $P(E \mid C)$

Example: Smoking and Teeth Color

S : سیگاری باشد

T : سفید بودن دندان

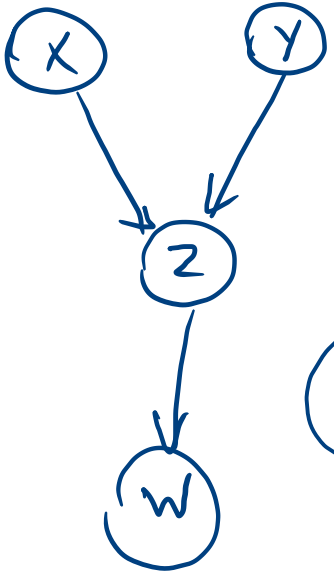
$P(S)$ $P(T|S)$

$P(S|T=1)$

$P(S|do\ T=1)$

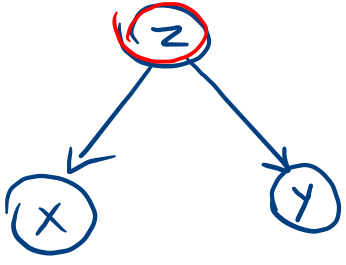
$$P(C \mid do\ E = e) = P(C)$$

$$P(E \mid do\ C = c) = \boxed{P(E \mid C)}$$



$$p(x, y \mid do\ W = w) = p(x, y)$$

$$p(z \mid do\ y = y) = p(z \mid y)$$



$$P(x,y,z) = P(z)\,P(x|z)\,P(y|z)$$

$$P(x \mid do\ y=y) = P(x)$$

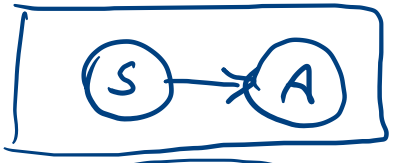
$$P(y \mid do\ x=x) = P(y)$$

$$P(x) = \sum_{z} P(x,z)$$

$$P(x|y) \neq P(x)$$

S: سیگاری

A: سرطان

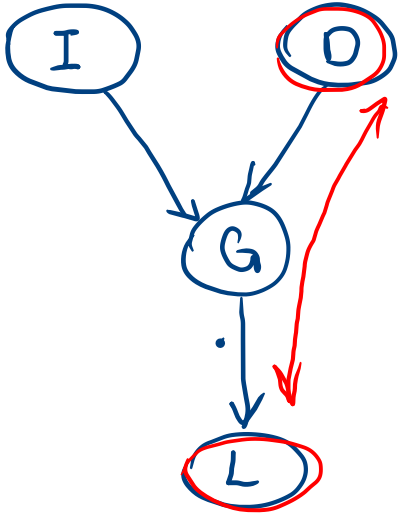


$P(A \mid S)$

$P(A \mid do\ S = 1)$

Observational Dist

Interventional ≠



$P(D \mid do\ L = \ell) = P(D)$

$(i, d, g, \cancel{\ell})$

ℓ'

$P(X_1, \dots, X_n)$

Causal Graph

SCM : Structural Causal Model

$$X_1 = U_1$$

$$X_2 = f_2(X_1, U_2)$$

$$X_3 = f_3(X_1, X_2, U_3)$$

$$\vdots$$

$U_1 \perp U_2 \perp \dots \perp U_n$

counterfactual Quest.

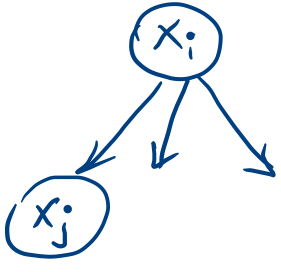
Computing $p(X_k|do\,x_i)$

$$\prod_{i=1}^{n} P(X_i \mid Pa(X_i))$$

$do\ X_i = x_i$

summation over x_i yields

$$p(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n|do\,x_i) = \prod_{j \neq i} p(X_j|PA_j(x_i))\,.$$



- distribution of X_j with $j \neq i$ is given by dropping $p(X_i|PA_i)$ and substituting x_i into PA_j to get $PA_j(x_i)$.
- obtain $p(X_k|do\,x_i)$ by marginalization

$P(x_j \mid X_i) \longrightarrow P(x_j \mid x_i)$

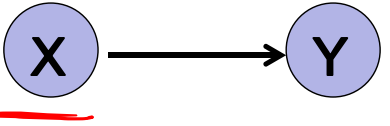
Computing $p(X_k|do\, x_i)$

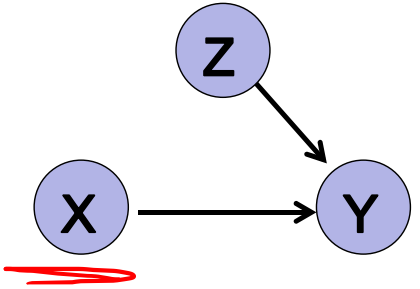
summation over x_i yields

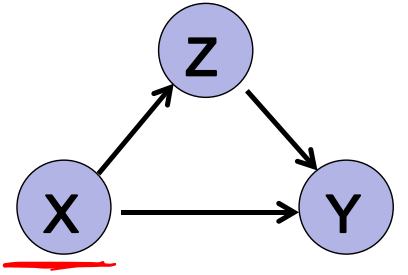
$$p(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n|do\, x_i) = \prod_{j \neq i} p(X_j|PA_j(x_i))\,.$$

- distribution of X_j with $j \neq i$ is given by dropping $p(X_i|PA_i)$ and substituting x_i into PA_j to get $PA_j(x_i)$.
- obtain $p(X_k|do\, x_i)$ by marginalization

Examples for $p(.|do\,x) = p(.|x)$

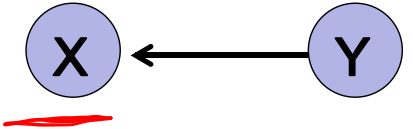




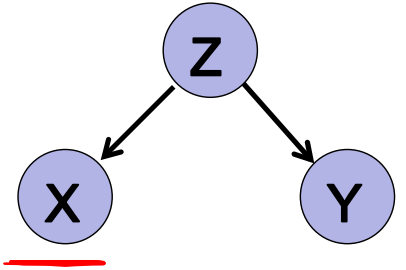


Examples for $p(.|do\,x) \neq p(.|x)$

- $p(Y|do\,x) = P(Y) \neq P(Y|x)$



- $p(Y|do\,x) = P(Y) \neq P(Y|x)$



Counterfactuals

Identifiability problem

e.g. Tian & Pearl (2002)

- given the causal DAG G and two nodes X_i, X_j

- which nodes need to be observed to compute $p(X_i|do\, x_j)$?

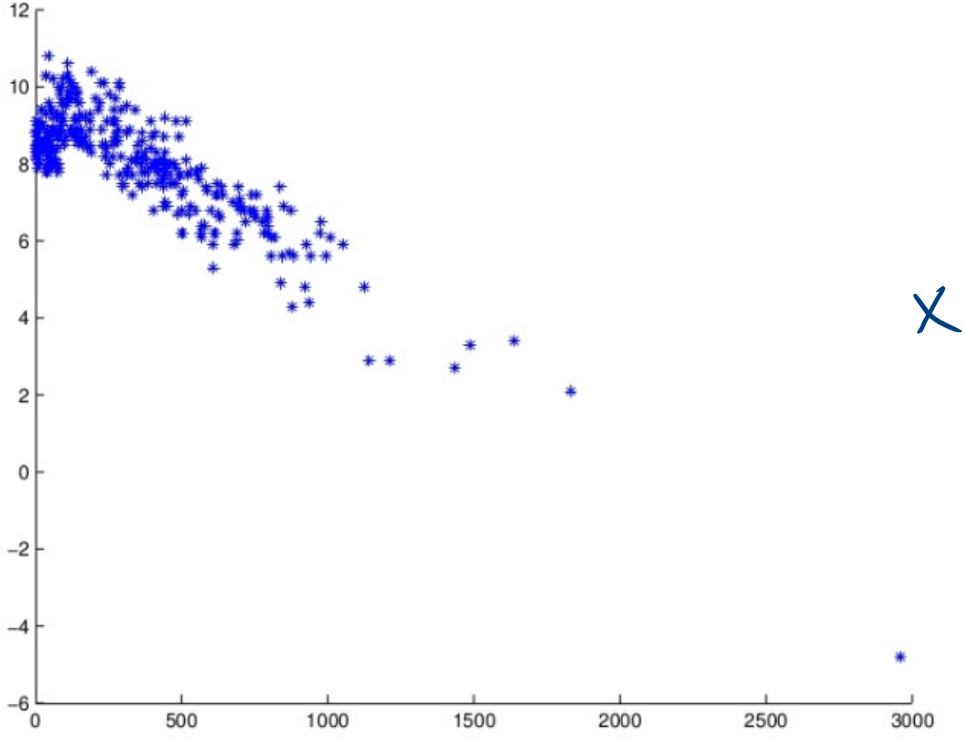
Causal Discovery

Inferring the DAG

- Key postulate: Causal Markov condition

- Essential mathematical concept: d-separation
  (describes the conditional independences required by a causal DAG)

What's the cause and what's the effect?

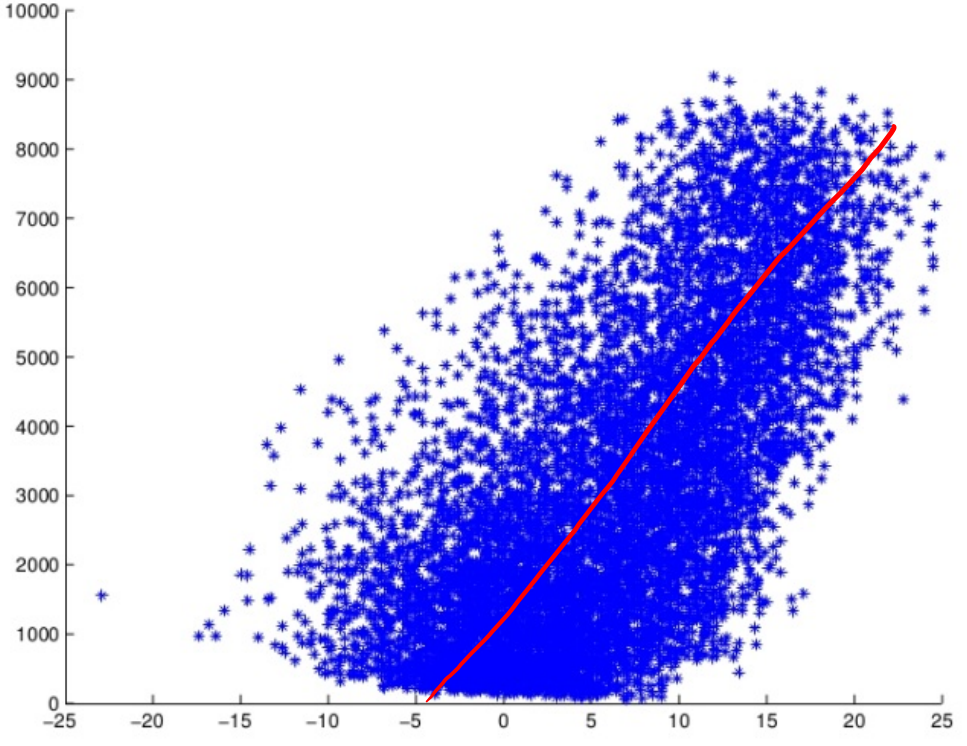


دما

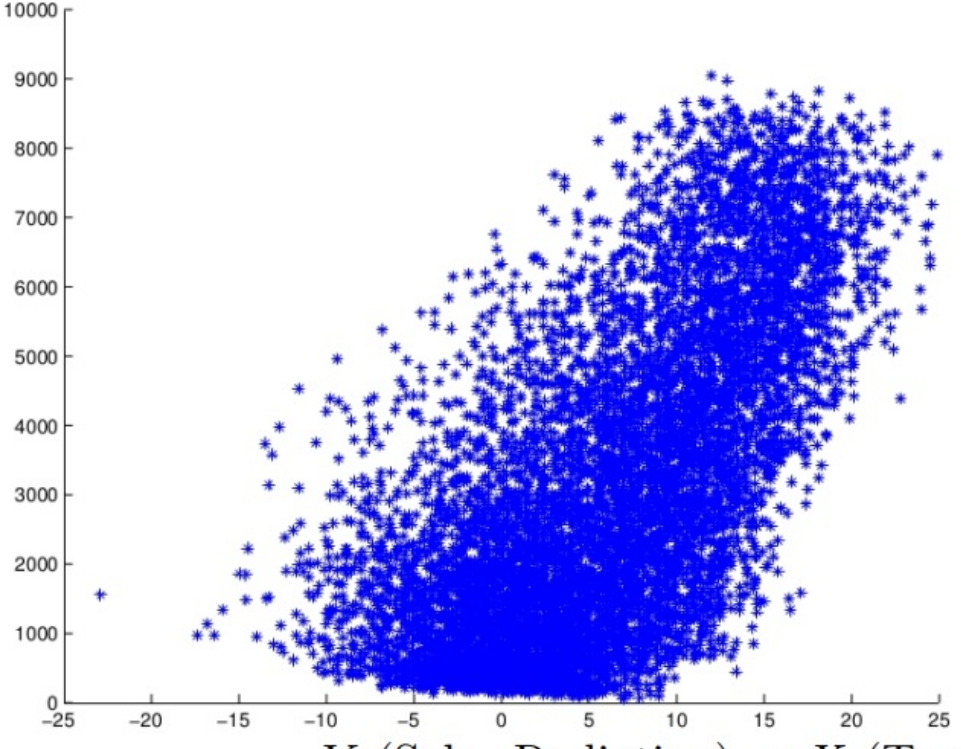
X → Y

ارتفاع

What's the cause and what's the effect?

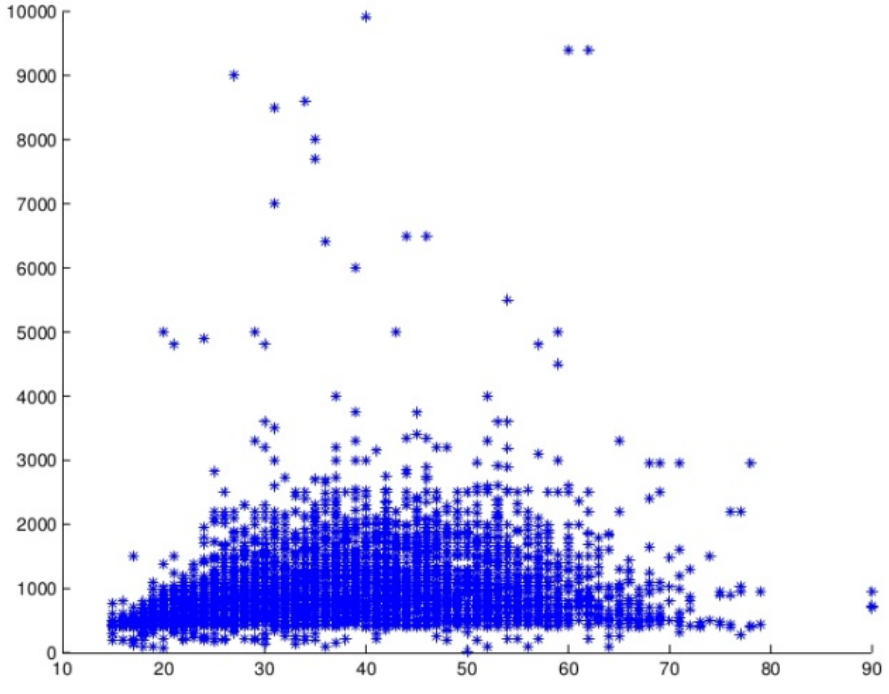


What's the cause and what's the effect?

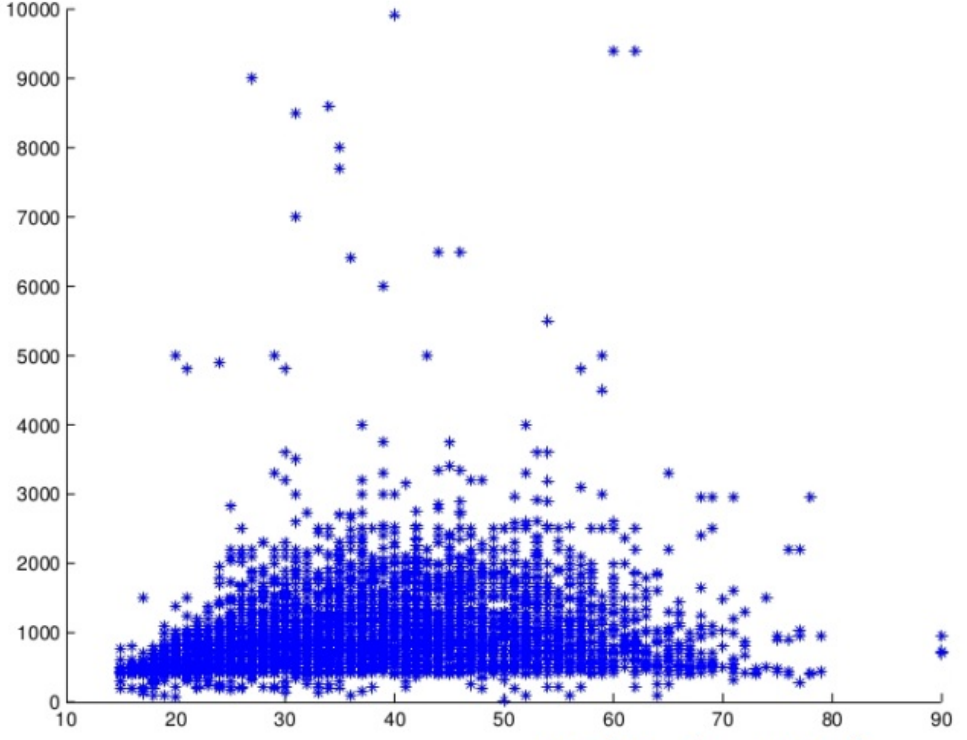


Y (Solar Radiation) $\rightarrow$ X (Temperature)

What's the cause and what's the effect?

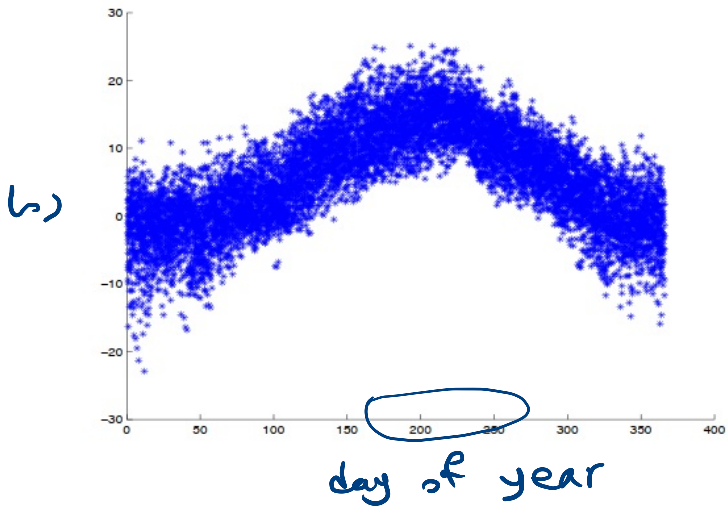


What's the cause and what's the effect?



X (Age) $\rightarrow$ Y (Income)

What's the cause and what's the effect?



x → y

Causal faithfulness

Spirtes, Glymour, Scheines

p is called faithful relative to G if only those independences hold true that are implied by the Markov condition, i.e.,

$$(X \perp\!\!\!\perp Y \,|Z)_G \quad \Leftarrow \quad (X \perp\!\!\!\perp Y \,|Z)_p$$

Recall: Markov condition reads

$$(X \perp\!\!\!\perp Y \,|Z)_G \quad \Rightarrow \quad (X \perp\!\!\!\perp Y \,|Z)_p$$

d-separation

Examples of unfaithful distributions (1)

Cancellation of direct and indirect influence in linear models

SCM

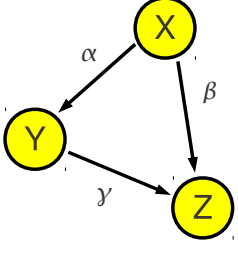
$$\begin{aligned} X &= U_X \\ Y &= \alpha X + U_Y \\ Z &= \beta X + \gamma Y + U_Z \end{aligned}$$

$= \beta X + \gamma\alpha X + \gamma U_Y + U_Z$

$= (\beta + \alpha\gamma) X + \gamma U_Y + U_Z$

with independent noise terms U_X, U_Y, U_Z

$\beta + \alpha\gamma = 0 \Rightarrow X \perp\!\!\!\perp Z$

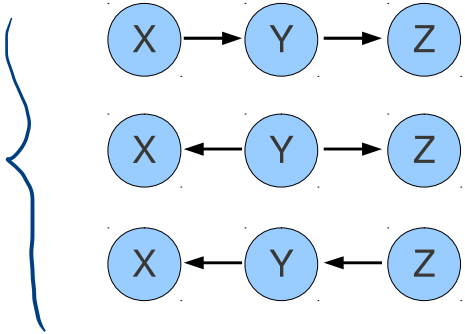


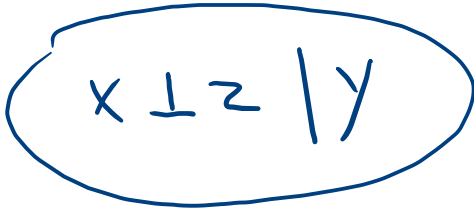
Markov equivalence class

Theorem (Verma and Pearl, 1990): two DAGs are Markov equivalent iff they have the same skeleton and the same v-structures.

skeleton: corresponding undirected graph
v-structure: substructure $X \to Y \leftarrow Z$ with no edge between X and Z

Markov equivalent DAGs

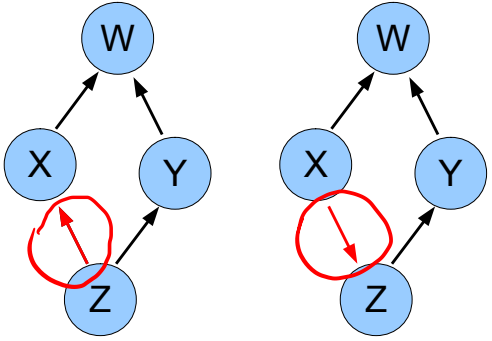


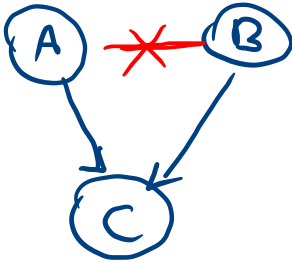


same skeleton, no v-structure

$$X \perp\!\!\!\perp Z \,|Y$$

Markov equivalent DAGs





same skeleton, same v-structure at W

Efficient construction of skeleton

PC algorithm by Spirtes & Glymour (1991)

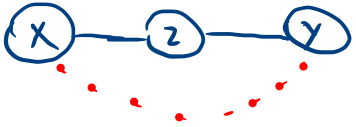
iteration over size of Sepset

1. remove all edges $X - Y$ with $X \perp\!\!\!\perp Y$
2. remove all edges $X - Y$ for which there is a neighbor $Z \neq Y$ of X with $X \perp\!\!\!\perp Y \,|Z$
3. remove all edges $X - Y$ for which there are two neighbors $Z_1, Z_2 \neq Y$ of X with $X \perp\!\!\!\perp Y \,|Z_1, Z_2$
4. ...

Advantages

- many edges can be removed already for small sets
- testing all sets S_{XY} containing the adjacencies of X is sufficient
- depending on sparseness, algorithm only requires independence tests with small conditioning tests
- polynomial for graphs of bounded degree

Find v-structures



$X \perp Y \mid Z$ ✓

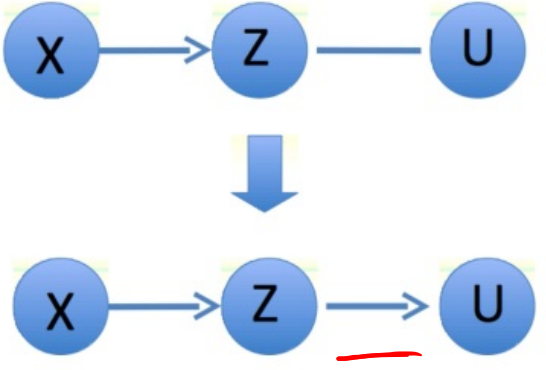
- given $X - Z - Y$ with X and Y non-adjacent
- given S_{XY} with $X \perp\!\!\!\perp Y \mid S_{XY}$

a priori, there are 4 possible orientations:

$X \to Z \to Y$ $X \leftarrow Z \to Y$ $X \leftarrow Z \leftarrow Y$	$Z \in S_{XY}$
$X \to Z \leftarrow Y$	$Z \notin S_{XY}$

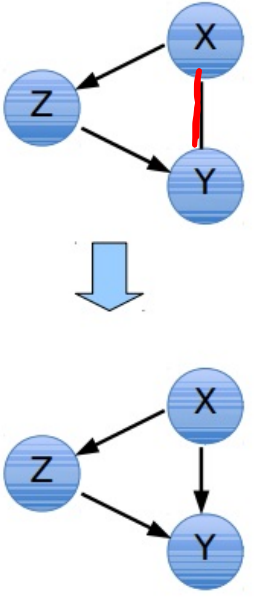
Orientation rule: create v-structure if $Z \notin S_{XY}$

Direct further edges (Rule 1)



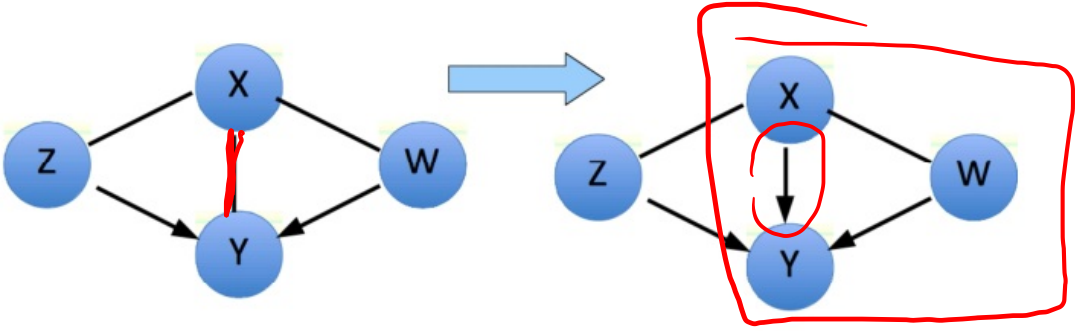
(otherwise we get a new v-structure)

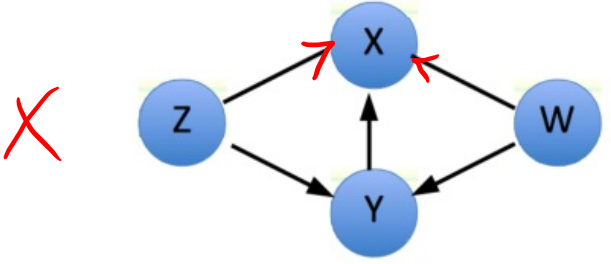
Direct further edges (Rule 2)



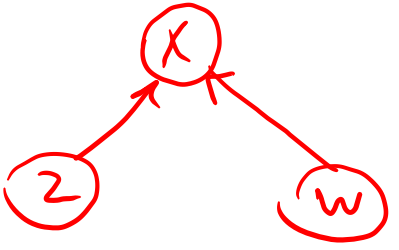
(otherwise one gets a cycle)

Direct further edges (Rule 3)



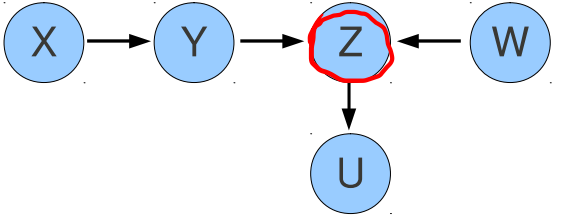


could not be completed without creating a cycle or a new v-structure

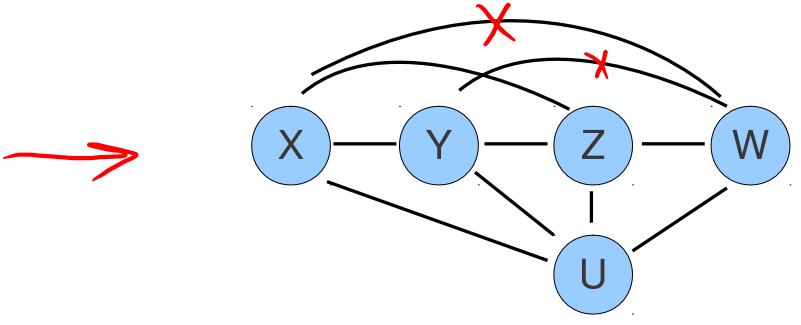


Examples

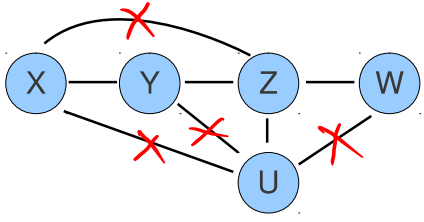
(taken from Spirtes et al, 2010)
true DAG

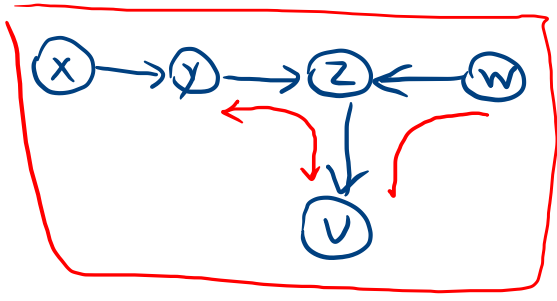


start with fully connected undirected graph



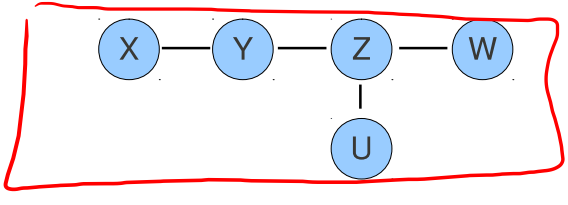
remove all edges $X - Y$ with $X \perp\!\!\!\perp Y \,|\emptyset$





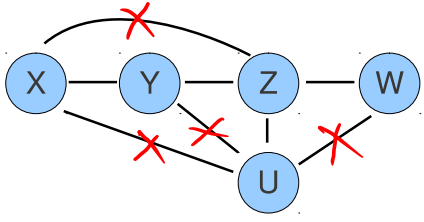
$$X \perp\!\!\!\perp W \quad Y \perp\!\!\!\perp W$$

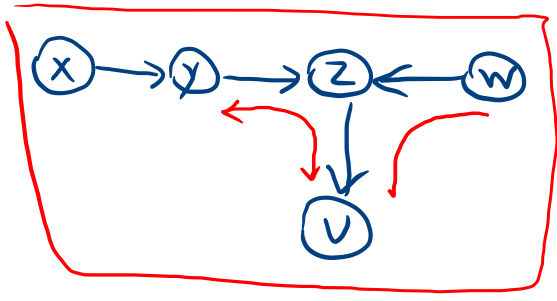
remove all edges having Sepset of size 1



$$X \perp\!\!\!\perp Z \,|Y \quad X \perp\!\!\!\perp U \,|Y \quad Y \perp\!\!\!\perp U \,|Z \quad W \perp\!\!\!\perp U \,|Z$$

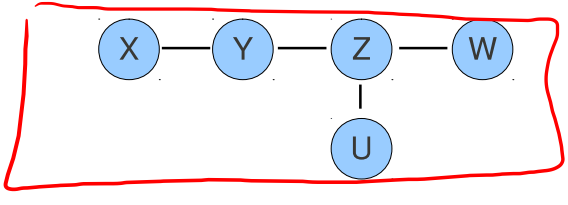
remove all edges $X - Y$ with $X \perp\!\!\!\perp Y \,|\emptyset$





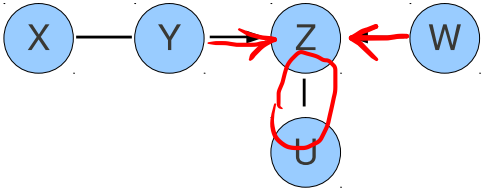
$$X \perp\!\!\!\perp W \quad Y \perp\!\!\!\perp W$$

remove all edges having Sepset of size 1



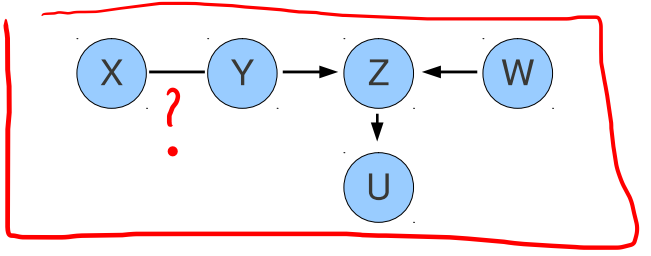
$$X \perp\!\!\!\perp Z \,|Y \quad X \perp\!\!\!\perp U \,|Y \quad Y \perp\!\!\!\perp U \,|Z \quad W \perp\!\!\!\perp U \,|Z$$

find v-structure



$$Z \notin S_{YW}$$

orient further edges (no further v-structure)



edge $X - Y$ remains undirected