Causal Inference: Classical and Quantum

Lecturer: Robert Spekkens
July 6, 2026

Journeys,
SAIFR, ICTP, Brazil

Causarum Investigatio
“Investigate the causes”

Simpson’s paradox in statistics and its causal resolution

P(recovery | drug) > P(recovery | no drug)

P(recovery | drug, male) < P(recovery | no drug, male)

P(recovery | drug, female) < P(recovery | no drug, female)

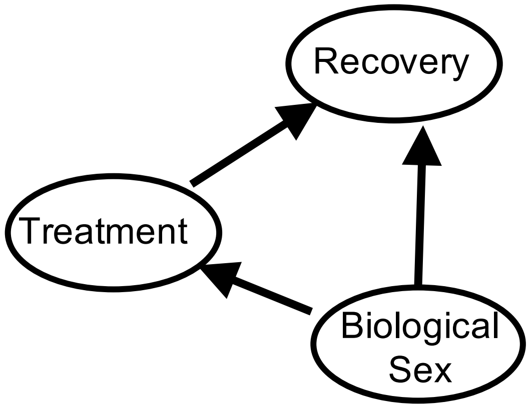
Recovery probability

	drug	**no drug**
male	180/300 = 60%	70/100 = 70%
female	20/100 = 20%	90/300 = 30%
combined	200/400 = 50%	160/400 = 40%

P(recovery | drug) > P(recovery | no drug)

P(recovery | drug, male) < P(recovery | no drug, male) ✓

P(recovery | drug, female) < P(recovery | no drug, female) ✓



Therefore: stratify the data by the common cause

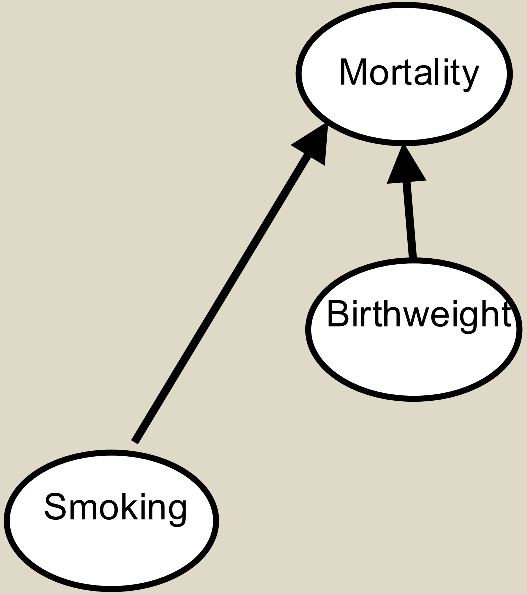
Berkson’s paradox in statistics and its causal resolution

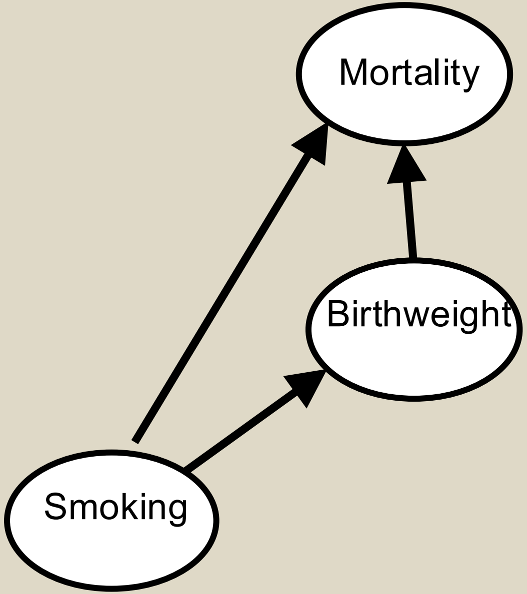
Birth weight paradox

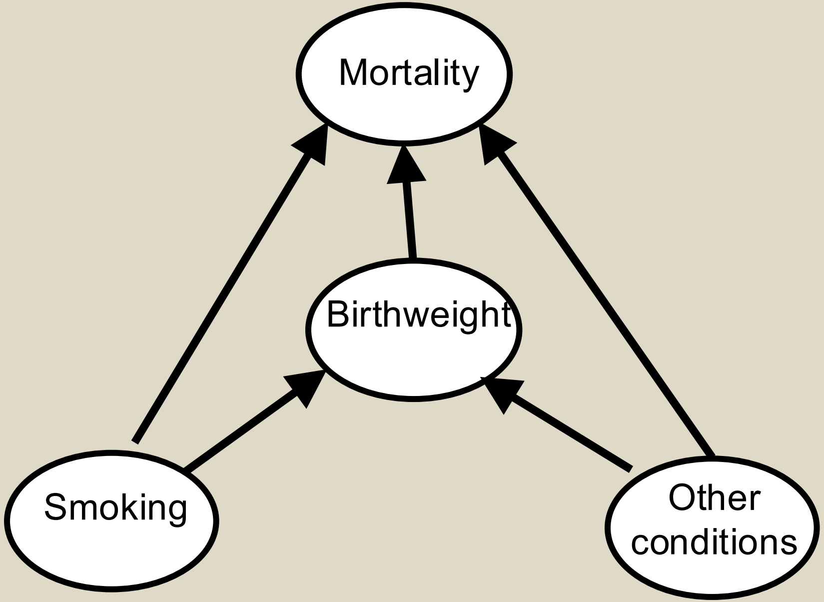
$$P(\text{mortality} \mid \text{born to smoker}) > P(\text{mortality} \mid \text{born to nonsmoker})$$

$$P(\text{mortality} \mid \text{born to smoker, LBW}) < P(\text{mortality} \mid \text{born to nonsmoker, LBW})$$

Can a mother being a smoker really reduce the risk of mortality for low birth weight children?


Mortality
Birthweight
Smoking

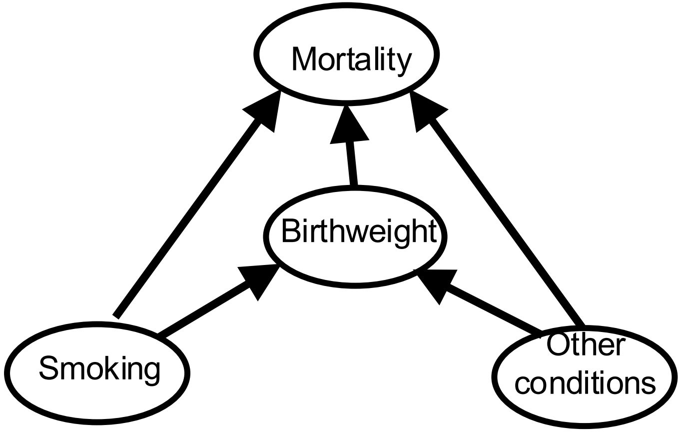

Mortality
Birthweight
Smoking


Mortality
Birthweight
Smoking
Other conditions

Birth weight paradox

P(mortality | born to smoker) > P(mortality | born to nonsmoker) ✓

P(mortality | born to smoker, LBW) < P(mortality | born to nonsmoker, LBW)



Therefore: *marginalize* over colliders on the "backdoor path"

Estimating causal effects requires knowledge of when correlations between variables can be induced or removed by conditioning and marginalizing over other variables

Broad problem of **deducing conditional independence relations from causal structure**

Preliminaries

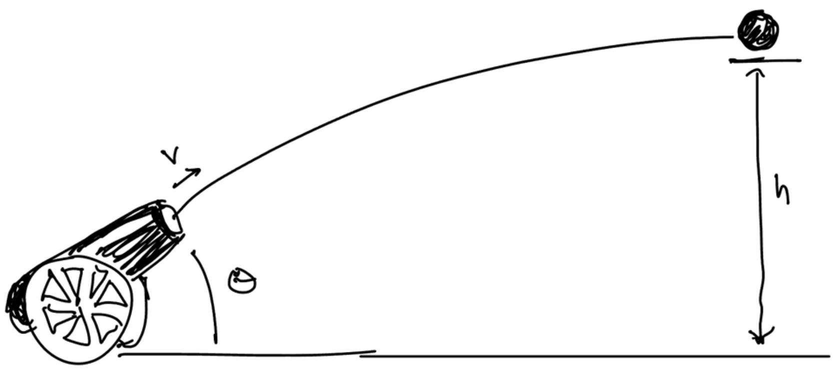
Here, we take:

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

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

Causal theory

The law of projectile motion

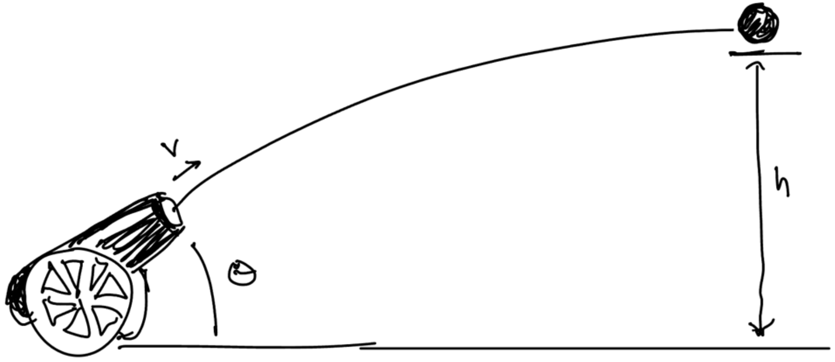


$$v = \frac{\sqrt{2gh}}{\sin\theta}$$

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

$$\theta = \arcsin\left(\frac{\sqrt{2gh}}{v}\right)$$

The law of projectile motion



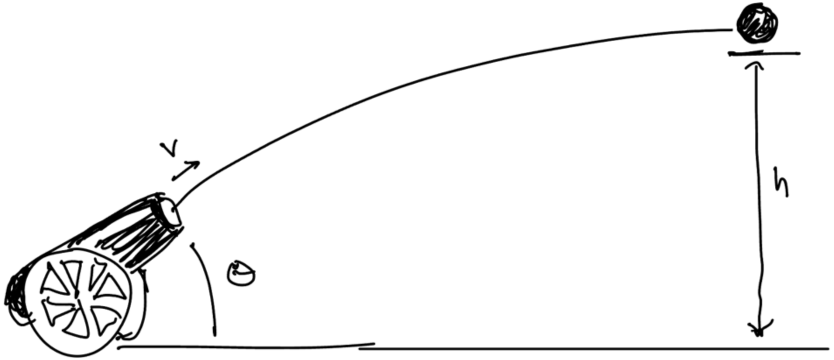
$$v = \frac{\sqrt{2gh}}{\sin\theta}$$

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

$$\theta = \arcsin\left(\frac{\sqrt{2gh}}{v}\right)$$

inferential relations

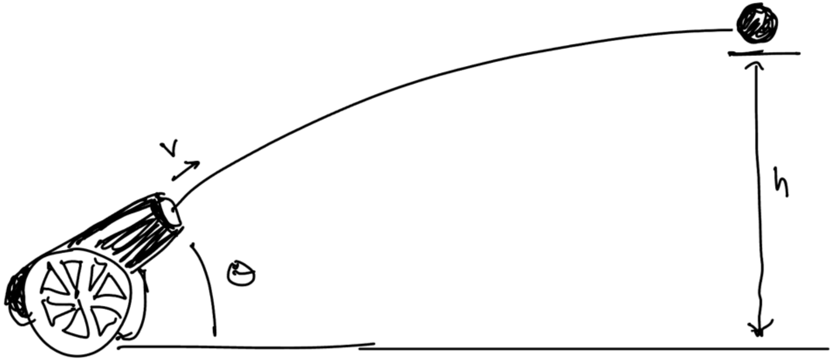
The law of projectile motion



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

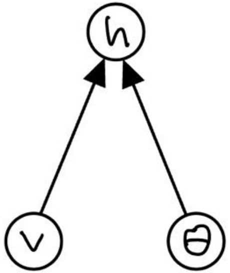
causal relation

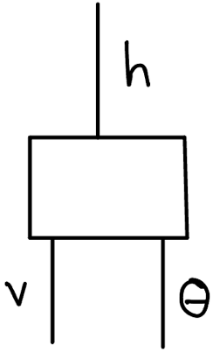
The law of projectile motion



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

causal relation





Inferential theory

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 $P_A = \delta_{A,0}$

Definition: Joint Probability Distribution on variables A and B

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

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

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

Let $\mathcal{P}_{AB}$ denote the set of distributions on AB.

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

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

Definition: Marginalizing over B

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

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

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

Note: 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:

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

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

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

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

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

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}:

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

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

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

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

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

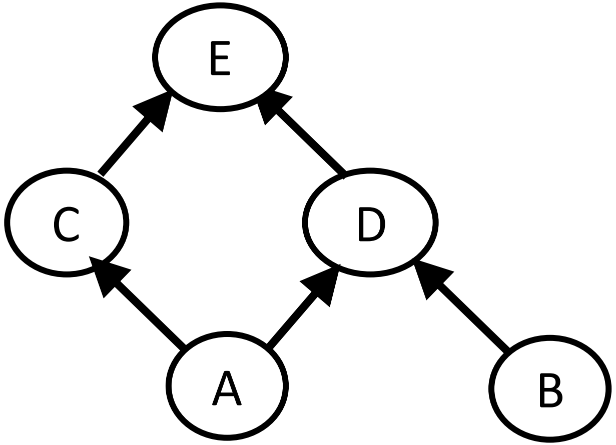
$$\forall a, b, c : P_{ABC}(abc) \left(\sum_{a'b'} P_{ABC}(a'b'c) \right) = \left(\sum_{b'} P_{ABC}(ab'c) \right) \left(\sum_{a'} P_{ABC}(a'bc) \right)$$

Semi-graphoid axioms

$$
\begin{aligned}
\text{Symmetry:} \quad & (X \perp Y \mid Z) \Leftrightarrow (Y \perp X \mid Z) \\
\text{Decomposition:} \quad & (X \perp YW \mid Z) \Rightarrow (X \perp Y \mid Z) \\
\text{Weak Union:} \quad & (X \perp YW \mid Z) \Rightarrow (X \perp Y \mid ZW) \\
\text{Contraction:} \quad & (X \perp Y \mid Z) \text{ and } (X \perp W \mid ZY) \\
& \Rightarrow (X \perp YW \mid Z)
\end{aligned}
$$

Compatibility constraints for latent-free causal structures

Causal structure



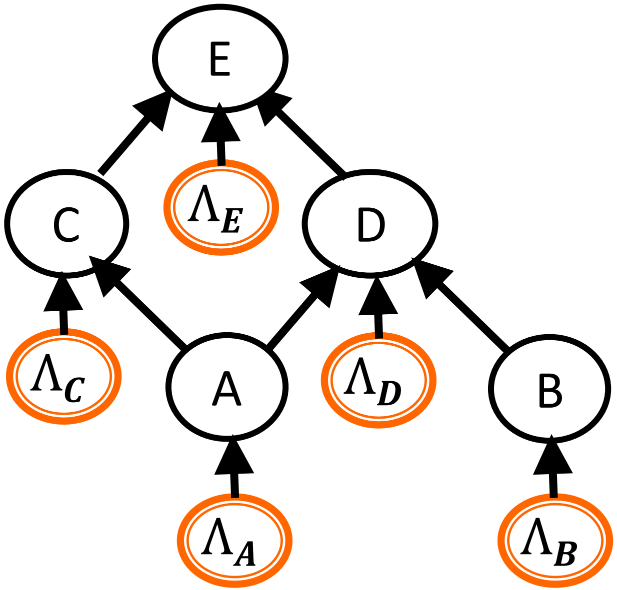
Parameters

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

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

A distribution is said to be compatible with a given causal structure if there are parameters that yield it

Causal structure

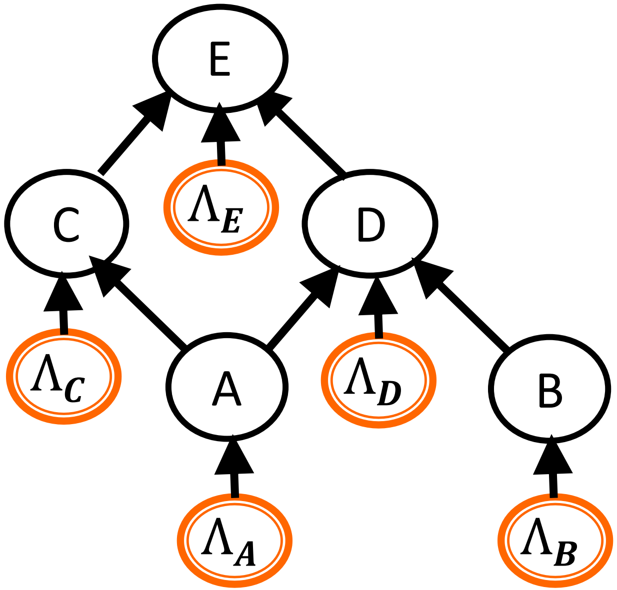


Parameters

$E = f_E(C, D, \Lambda_E)$
$D = f_D(A, B, \Lambda_D)$
$C = f_C(A, \Lambda_C)$
$B = f_B(\Lambda_B)$
$A = f_A(\Lambda_A)$
P_{Λ_E}
P_{Λ_D}
P_{Λ_C}
P_{Λ_B}
P_{Λ_A}

$$P_{ABCDE} = \sum_{\Lambda_E,\ldots,\Lambda_A} \delta_{E,f_E(C,D,\Lambda_E)}\delta_{D,f_D(A,B,\Lambda_D)}\delta_{C,f_C(A,\Lambda_C)}\delta_{B,f_B(\Lambda_B)}\delta_{A,f_A(\Lambda_A)} \times P_{\Lambda_E}P_{\Lambda_D}P_{\Lambda_C}P_{\Lambda_B}P_{\Lambda_A}$$

Causal structure



Parameters

$$P_{E|CD} = \sum_{\Lambda_E} \delta_{E,f_E(C,D,\Lambda_E)} P_{\Lambda_E}$$

$$P_{D|AB} = \sum_{\Lambda_D} \delta_{D,f_D(A,B,\Lambda_D)} P_{\Lambda_D}$$

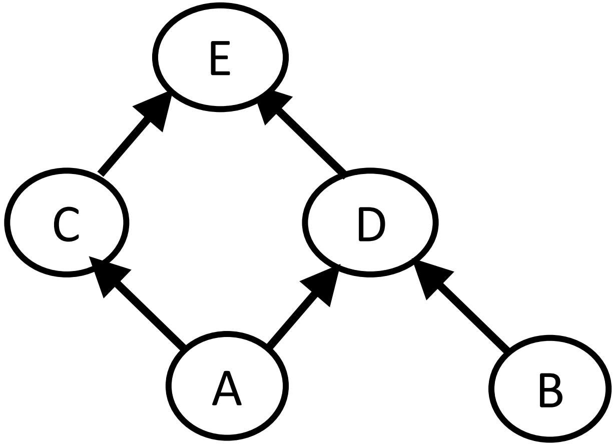
$$P_{C|A} = \sum_{\Lambda_C} \delta_{C,f_C(A,\Lambda_C)} P_{\Lambda_C}$$

$$P_B = \sum_{\Lambda_B} \delta_{B,f_B(\Lambda_B)} P_{\Lambda_B}$$

$$P_A = \sum_{\Lambda_A} \delta_{A,f_A(\Lambda_A)} P_{\Lambda_A}$$

$$\begin{aligned} P_{ABCDE} &= \sum_{\Lambda_E,\ldots,\Lambda_A} \delta_{E,f_E(C,D,\Lambda_E)} \delta_{D,f_D(A,B,\Lambda_D)} \delta_{C,f_C(A,\Lambda_C)} \delta_{B,f_B(\Lambda_B)} \delta_{A,f_A(\Lambda_A)} \\ &\qquad \times P_{\Lambda_E} P_{\Lambda_D} P_{\Lambda_C} P_{\Lambda_B} P_{\Lambda_A} \\ &= P_{E|CD} P_{D|AB} P_{C|A} P_B P_A \end{aligned}$$

Causal structure



Parameters

$P_{E|CD}$

$P_{D|AB}$

$P_{C|A}$

P_B

P_A

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

If $X_1, X_2, \ldots, X_n$ 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 causal Markov condition

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

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

Proof: $$P_{X_1,\ldots,X_n} = \sum_{\Lambda_1,\ldots,\Lambda_n} P_{X_1,\ldots,X_n,\Lambda_1,\ldots,\Lambda_n}$$
$$= \sum_{\Lambda_1,\ldots,\Lambda_n} P_{X_1,\ldots,X_n|\Lambda_1,\ldots,\Lambda_n} P_{\Lambda_1,\ldots,\Lambda_n}$$ By prob. theory

$$P_{X_1,\ldots,X_n|\Lambda_1,\ldots,\Lambda_n} = \prod_{i=1}^n \delta_{X_i,f_i(\mathrm{Pa}_G(X_i),\Lambda_i)}$$ By assumption of determinism

$$P_{\Lambda_1,\ldots,\Lambda_n} = \prod_{i=1}^n P_{\Lambda_i}$$ By independence in the DAG

$$P_{X_1,\ldots,X_n} = \sum_{\Lambda_1,\ldots,\Lambda_n} \prod_{i=1}^n \left(\delta_{X_i,f_i(\mathrm{Pa}_G(X_i),\Lambda_i)} P_{\Lambda_i}\right)$$
$$= \prod_{i=1}^n \left(\sum_{\Lambda_i} \delta_{X_i,f_i(\mathrm{Pa}_G(X_i),\Lambda_i)} P_{\Lambda_i}\right)$$
$$= \prod_{i=1}^n P_{X_i|\mathrm{Pa}_G(X_i)}$$

Another form of the causal Markov condition

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

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

Proof:

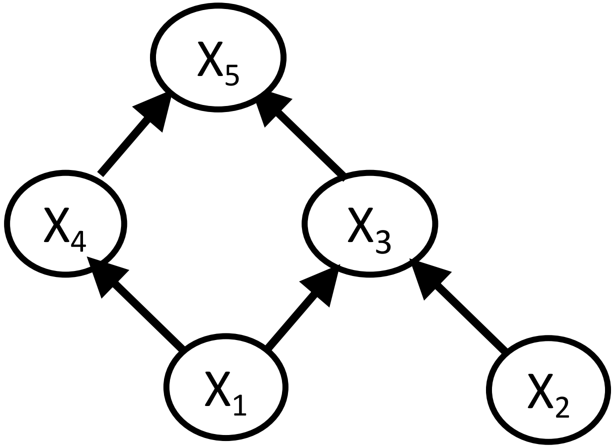
Recall the causal Markov condition:

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

$$\begin{aligned}
P(X|\text{Pa}(X),\text{Nd}(X)) &= \frac{P(X,\text{Pa}(X),\text{Nd}(X))}{P(\text{Pa}(X),\text{Nd}(X))}, \quad \text{by Prob. theory} \\
&= \frac{P(X|\text{Pa}(X))\prod_{Y\in\text{Pa}(X),\text{Nd}(X)} P(Y|\text{Pa}(Y))}{\prod_{Y\in\text{Pa}(X),\text{Nd}(X)} P(Y|\text{Pa}(Y))}, \quad \text{by Markov} \\
&= P(X|\text{Pa}(X)).
\end{aligned}$$

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

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



$(X_1 \perp X_2)$

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

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

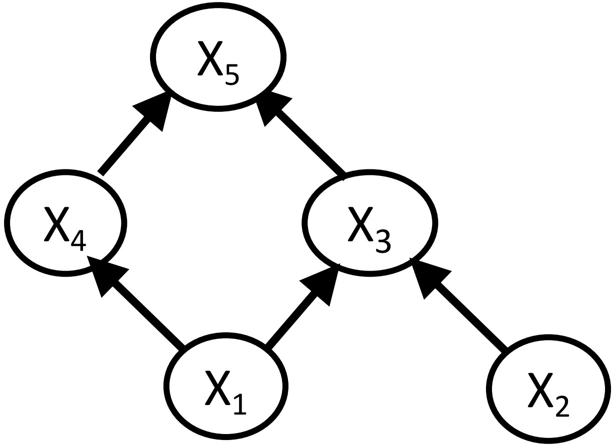
$(X_4 \perp \{X_2, X_3\} \mid X_1)$

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

Semi-graphoid axioms

$$\begin{aligned}
\text{Symmetry:} \quad & (X \perp Y \mid Z) \Leftrightarrow (Y \perp X \mid Z) \\
\text{Decomposition:} \quad & (X \perp YW \mid Z) \Rightarrow (X \perp Y | Z) \\
\text{Weak Union:} \quad & (X \perp YW \mid Z) \Rightarrow (X \perp Y \mid ZW) \\
\text{Contraction:} \quad & (X \perp Y \mid Z) \text{ and } (X \perp W \mid ZY) \\
& \Rightarrow (X \perp YW \mid Z)
\end{aligned}$$

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



$(X_1 \perp X_2)$

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

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

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

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

The semi-graphoid axioms then imply

$(X_4 \perp X_2 \mid X_1)$

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

$\dots$

d-separation

There is a graphical criterion for testing conditional independences called "d-separation"

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

implies

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

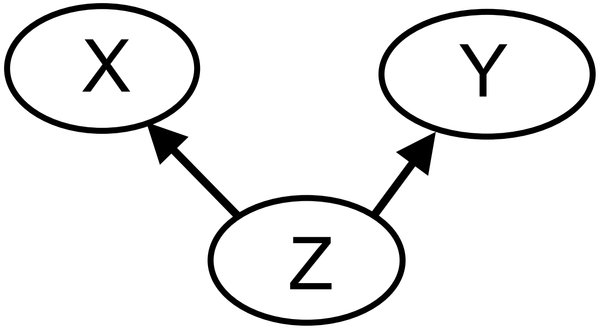
in every probability distribution
compatible with G

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 noncollider node in **Z** (i.e., the mediary in a chain or the tail of a fork)
2. The path contains a collider node **not** in **Z** and no descendant of which is in **Z**.

Definition (d-separation) Given a DAG G with vertices **V**, two disjoint sets 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 by the set of vertices **Z**.

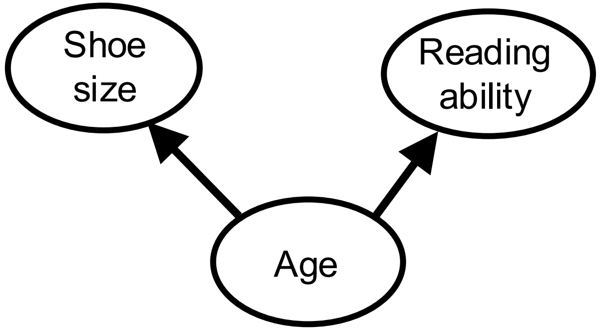
Fork



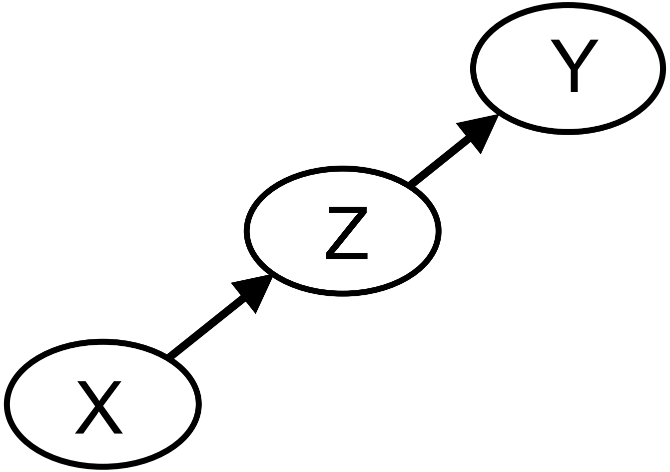
Z is a complete common cause of X and Y, therefore

$$X \perp Y | Z$$

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


Shoe size
Reading ability
Age

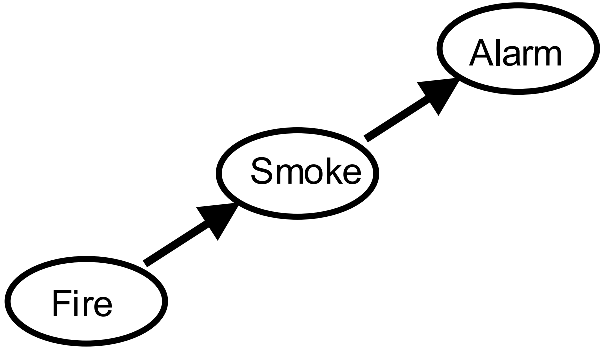
Chain



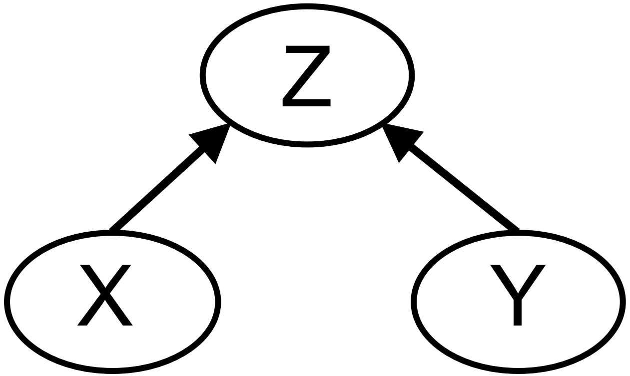
Z is the full set of parents of Y, while X is a nondescendent, therefore

$$X \perp Y | Z$$

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


Alarm
Smoke
Fire

Collider

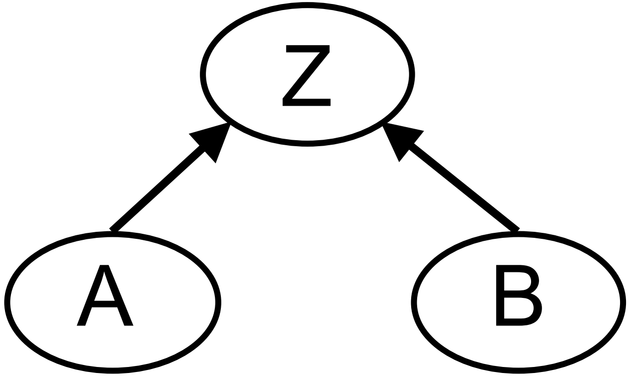


X and Y have no ancestors in common, therefore

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

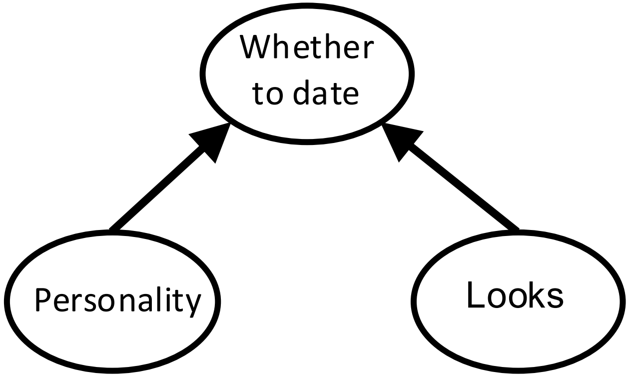


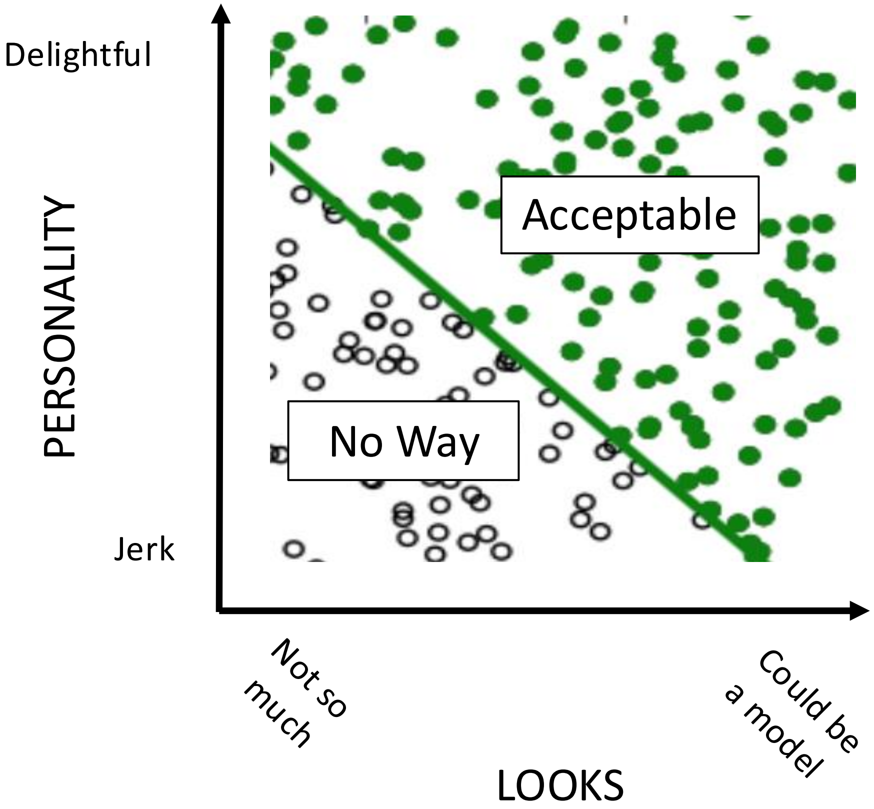
$$P_{Z|AB} = \delta_{Z,AB}$$
$$P_A = \tfrac{1}{2}[0]_A + \tfrac{1}{2}[1]_A$$
$$P_B = \tfrac{1}{2}[0]_B + \tfrac{1}{2}[1]_B$$

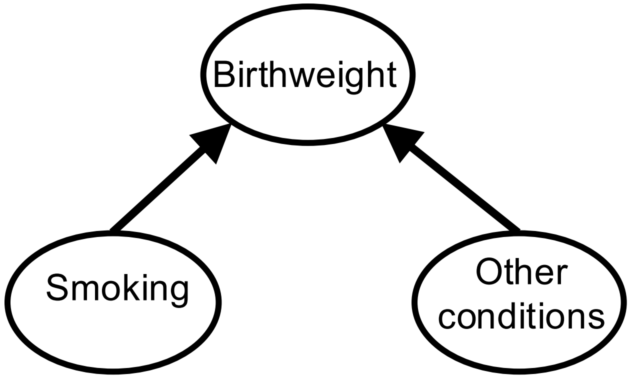
$$\begin{aligned} P_{AB} &= (\tfrac{1}{2}[0]_A + \tfrac{1}{2}[1]_A)(\tfrac{1}{2}[0]_B + \tfrac{1}{2}[1]_B) \\ &= \tfrac{1}{4}[0]_A[0]_B + \tfrac{1}{4}[0]_A[1]_B + \tfrac{1}{4}[1]_A 0]_B + \tfrac{1}{4}[1]_A[1]_B \end{aligned}$$

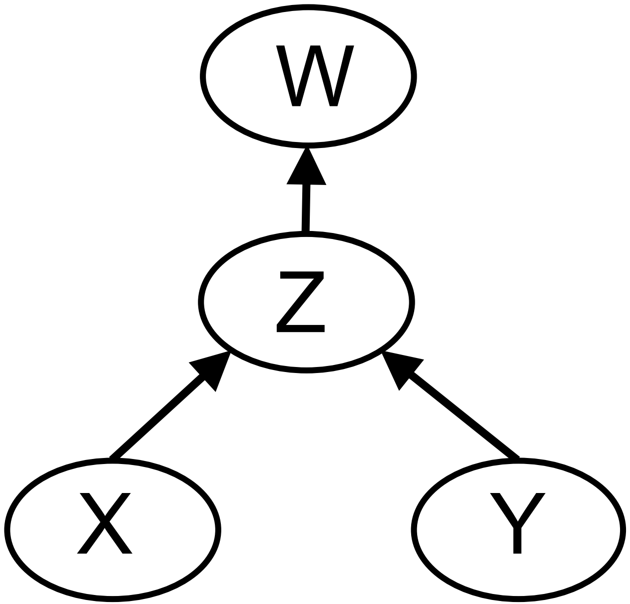
$$P_{AB|Z} = P_{Z|AB} P_{AB} P_Z^{-1}$$

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


Whether to date
Personality
Looks


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


Birthweight
Smoking
Other conditions



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

Putting it together

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 noncollider node in **Z** (i.e., the mediary in a chain or the tail of a fork)
2. The path contains a collider node **not** in **Z** and no descendant of which is in **Z**.

Definition (d-separation) Given a DAG G with vertices **V**, two disjoint sets 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 by the set of vertices **Z**.

The d-separation theorem:

Consider a causal structure G and three disjoint subsets of variables **X**, **Y** and **Z**.

Soundness

d-separation between **X** and **Y** given **Z** in G
implies $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$ in all distributions P compatible with G

Completeness

$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$ in all distributions P compatible with G
implies d-separation between **X** and **Y** given **Z** in G

The d-separation theorem:

Consider a causal structure G and three disjoint subsets of variables **X**, **Y** and **Z**.

Soundness

$$\mathbf{X} \perp_d \mathbf{Y}|\mathbf{Z} \text{ in } G \quad \Longrightarrow \quad \forall P \in \mathsf{Comp}_G : \mathbf{X} \perp \mathbf{Y}|\mathbf{Z} \text{ in } P$$

Completeness

$$\forall P \in \mathsf{Comp}_G : \mathbf{X} \perp \mathbf{Y}|\mathbf{Z} \text{ in } P \quad \Longrightarrow \quad \mathbf{X} \perp_d \mathbf{Y}|\mathbf{Z} \text{ in } G$$

Relevance to the causal compatibility problem

If all the variables in a DAG are observed,
the conditional independence relations implied by the d-separation relations are **all** the constraints on the distribution

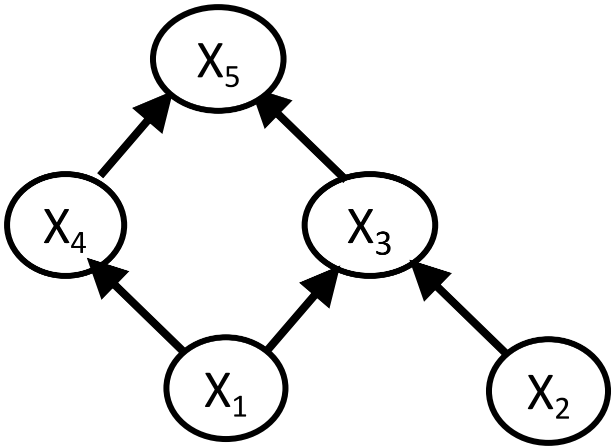
Proof:
Recall the Markov condition characterizing all compatible distributions for a latent-free DAG

We saw that this set can also be characterized by the set of conditional independence relations described by the local Markov condition together with semi-graphoid axioms

This set is equivalent to the one obtained from all d-separation relations

If **not** all the variables in a DAG are observed,
the conditional independence relations implied by the d-separation relations are merely necessary conditions on the distribution

Example:



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

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

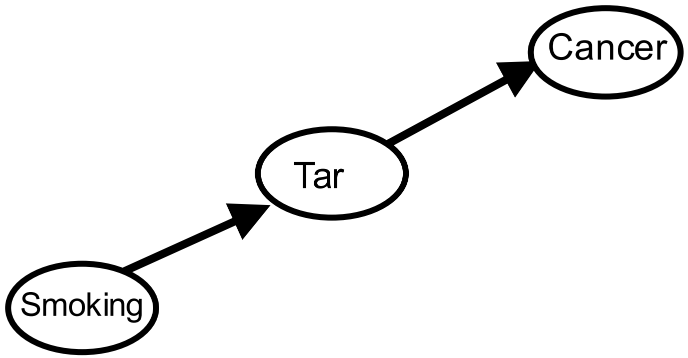
Follows from d-separation in a straightforward way

Causal structure

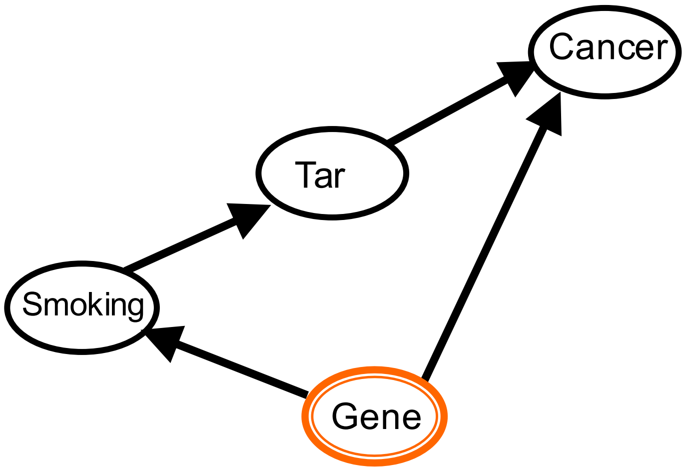
Constraint on observed probability distribution

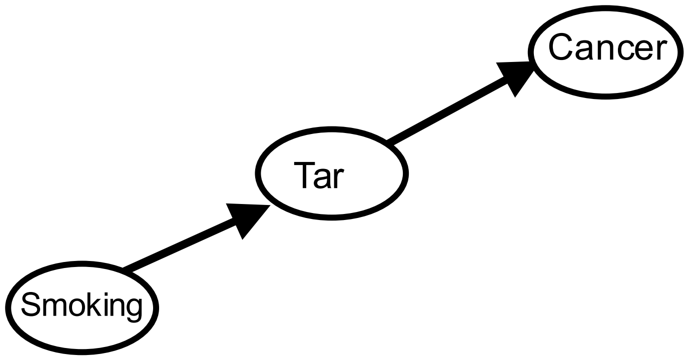
Falsification of causal structure

Violation of constraint on observed probability distribution

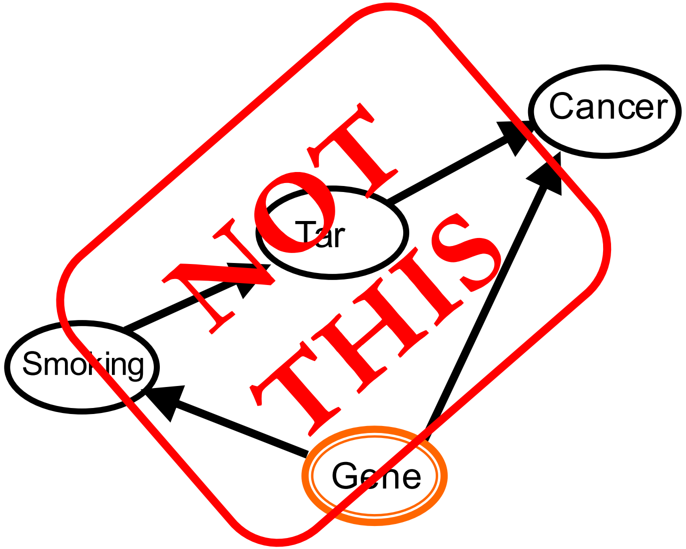


Vs.





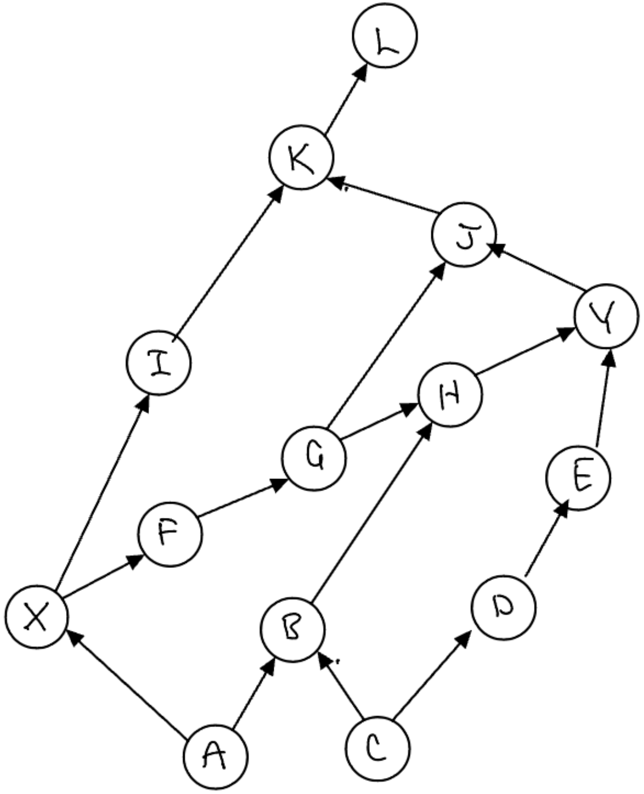
Vs.



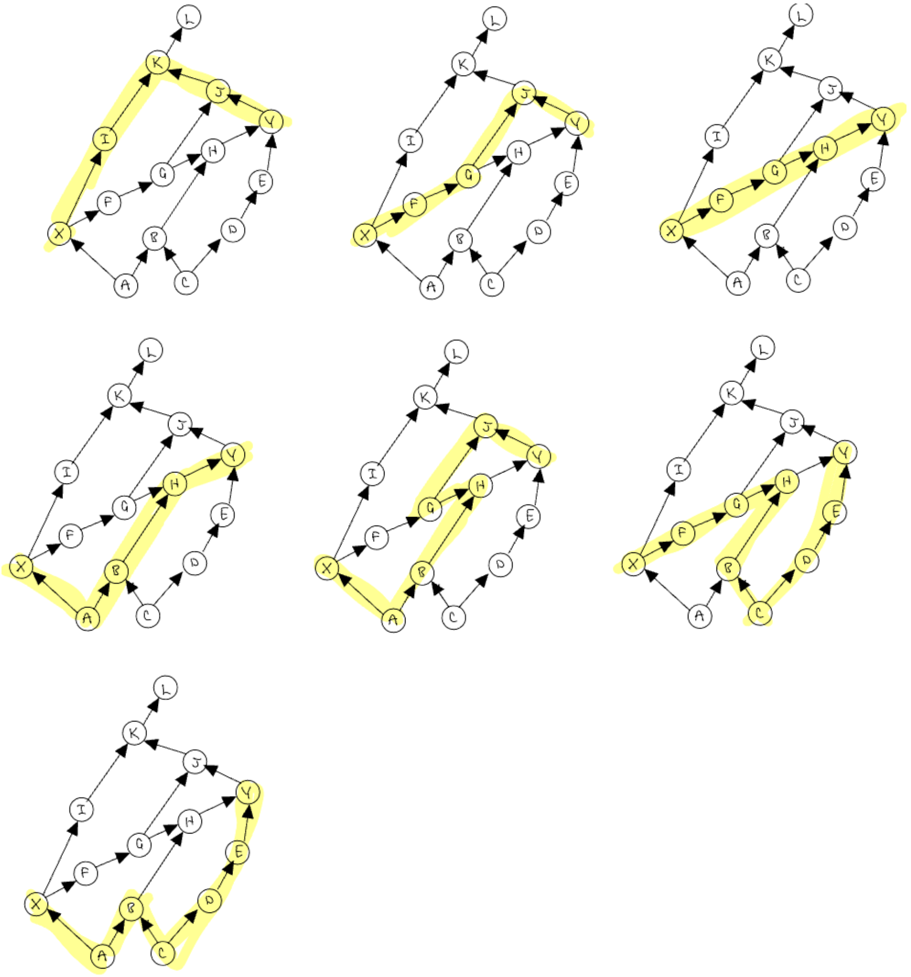
Observe P_{STC} such that

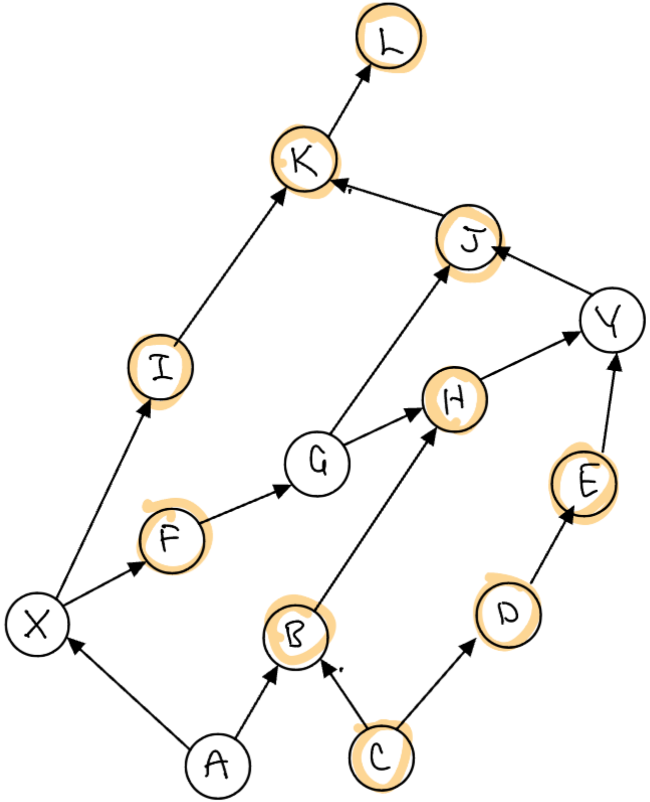
$$S \perp C | T$$

Then under assumption of no fine-tuning

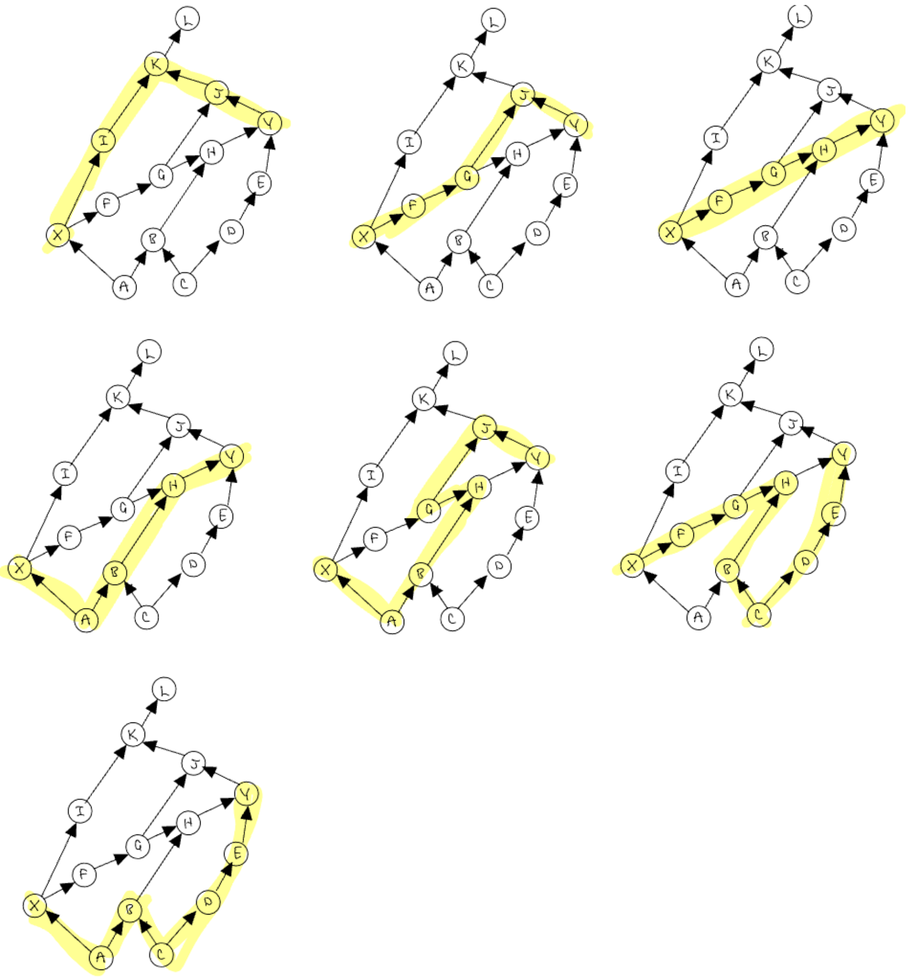


$$X \perp Y | GA$$

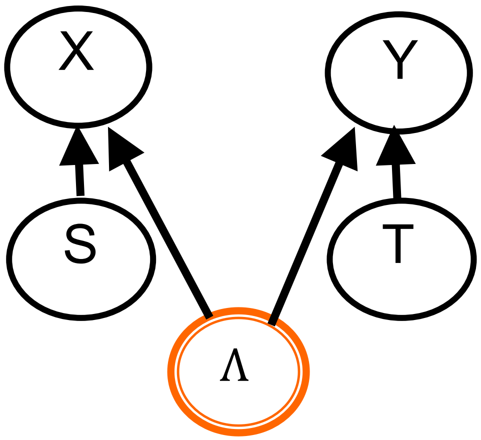




$$X \perp Y | GA$$



Applications



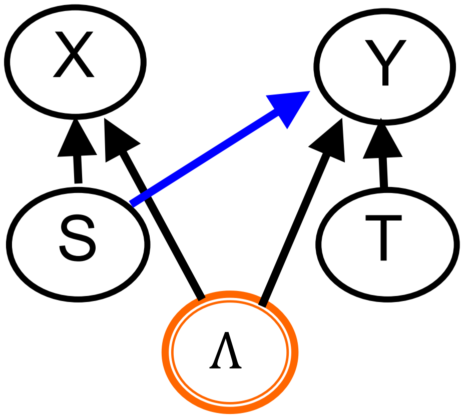
Conditional independence relations

$$X \perp T \mid S$$
$$Y \perp S \mid T$$

Selective
no-signalling

$$X \perp T$$
$$Y \perp S$$

Nonselective
no-signalling



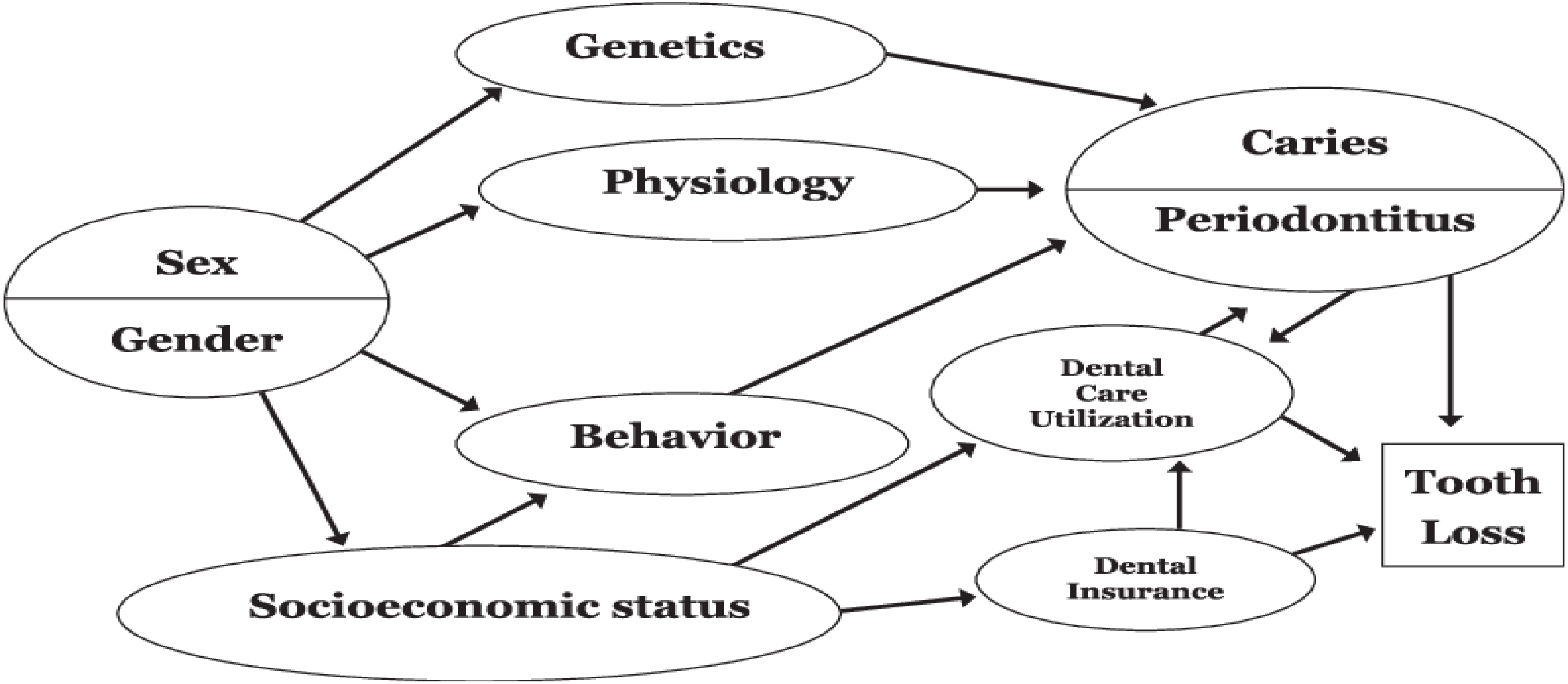
Conditional independence relations

$X \perp T \mid S$

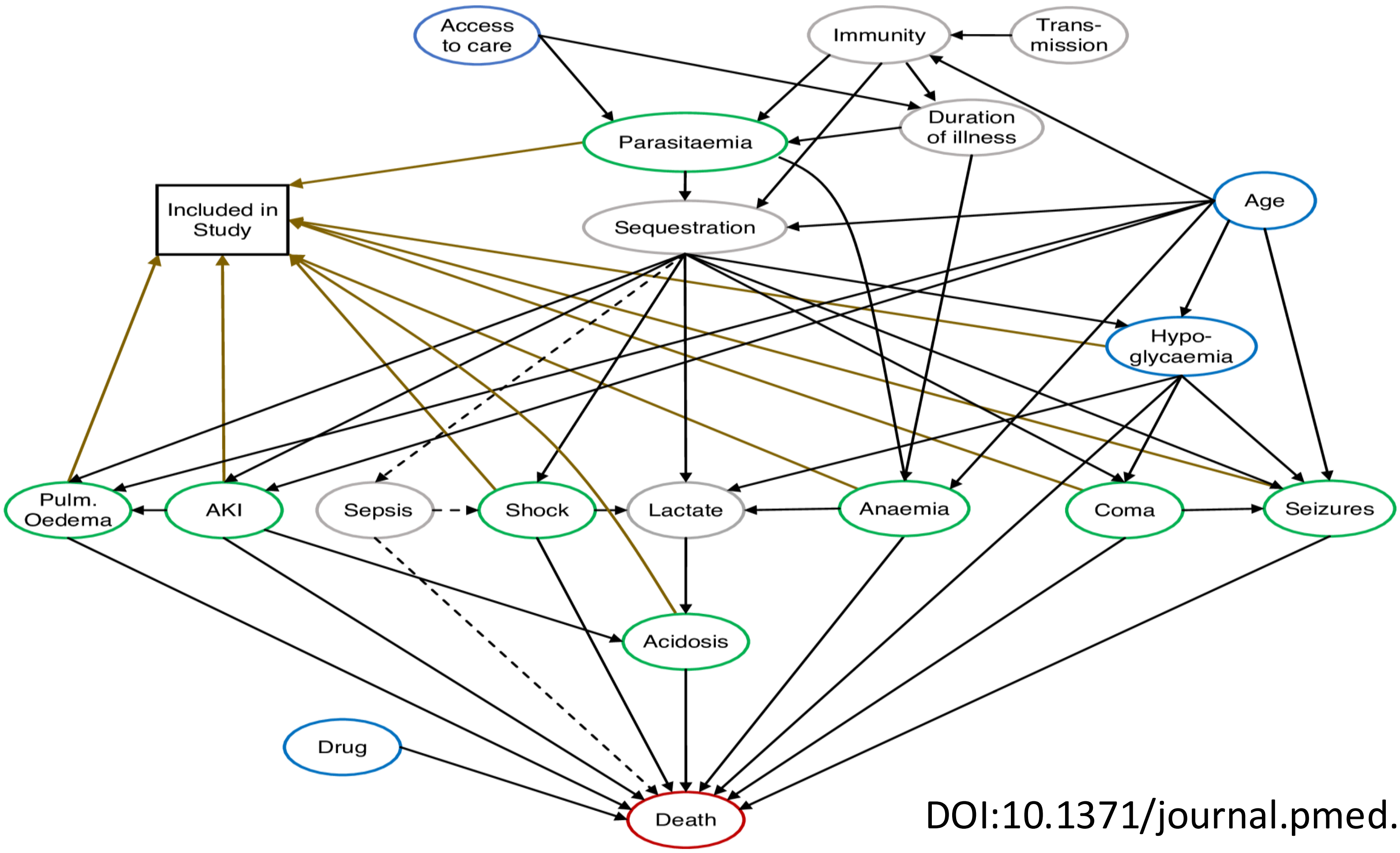
Selective no-signalling

$X \perp T$

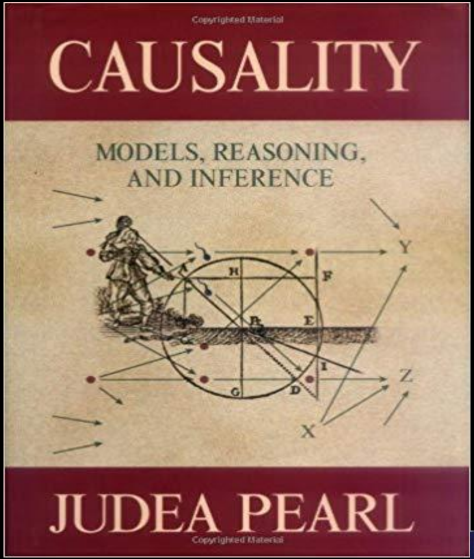
Nonselective no-signalling

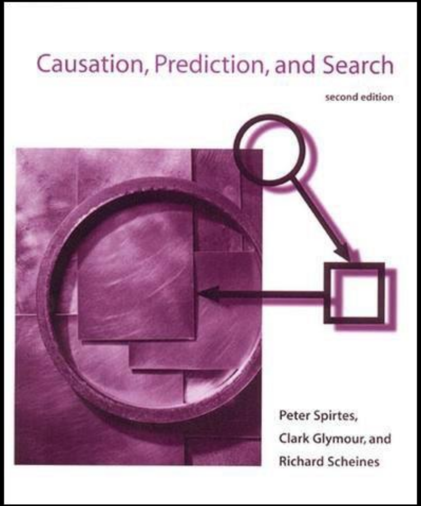


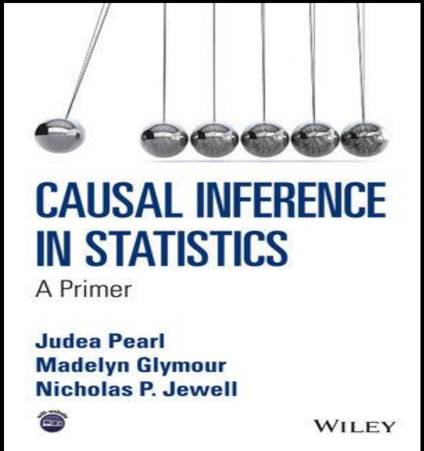
DOI:10.1016/j.cden.2013.02.006

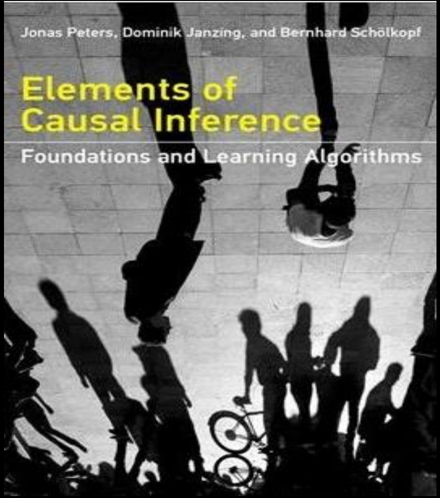


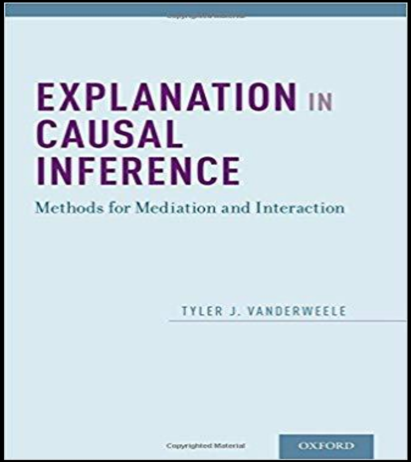
DOI:10.1371/journal.pmed.1002858

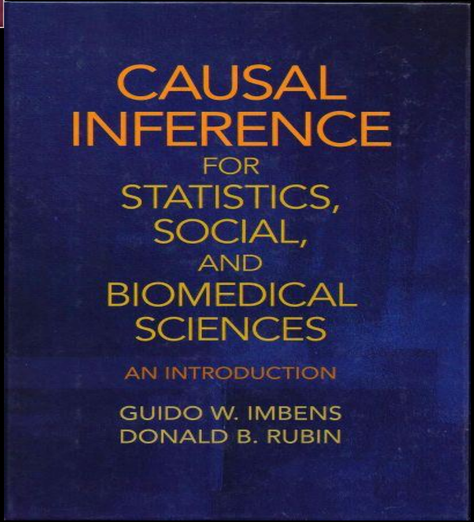

Copyrighted Material
CAUSALITY
MODELS, REASONING, AND INFERENCE
JUDEA PEARL
Copyrighted Material

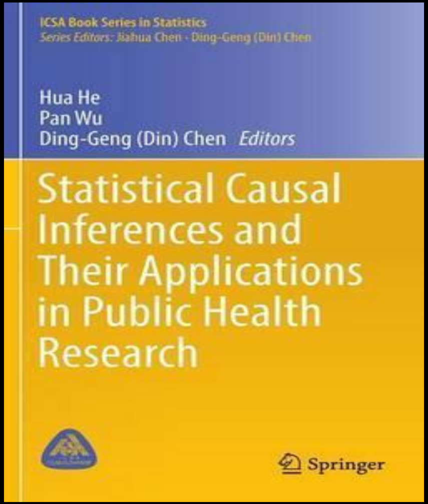

Causation, Prediction, and Search
second edition
Peter Spirtes,
Clark Glymour, and
Richard Scheines

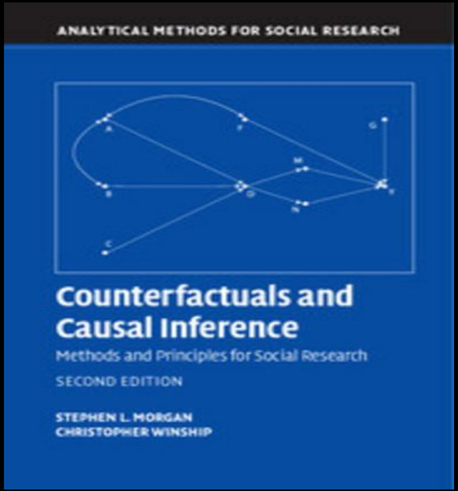

CAUSAL INFERENCE IN STATISTICS
A Primer
Judea Pearl
Madelyn Glymour
Nicholas P. Jewell
WILEY


Jonas Peters, Dominik Janzing, and Bernhard Schölkopf
Elements of Causal Inference
Foundations and Learning Algorithms


EXPLANATION IN CAUSAL INFERENCE
Methods for Mediation and Interaction
TYLER J. VANDERWEELE
Copyrighted Material
OXFORD


CAUSAL INFERENCE
FOR STATISTICS, SOCIAL, AND BIOMEDICAL SCIENCES
AN INTRODUCTION
GUIDO W. IMBENS
DONALD B. RUBIN


ICSA Book Series in Statistics
Series Editors: Jiahua Chen · Ding-Geng (Din) Chen
Hua He
Pan Wu
Ding-Geng (Din) Chen Editors
Statistical Causal Inferences and Their Applications in Public Health Research
Springer


ANALYTICAL METHODS FOR SOCIAL RESEARCH
Counterfactuals and Causal Inference
Methods and Principles for Social Research
SECOND EDITION
STEPHEN L. MORGAN
CHRISTOPHER WINSHIP

All scientific studies