Web-Mining Agents

Prof. Dr. Ralf Möller
Dr. Özgür Özçep
Universität zu Lübeck
Institut für Informationssysteme

Structural Causal Models

slides prepared by Özgür Özçep

Part I: Basic Notions

(SCMs, d-separation)

Literature

- J. Pearl, M. Glymour, N. P. Jewell: Causal inference in statistics – A primer, Wiley, 2016.

(Main Reference)

- J. Pearl: Causality, CUP, 2000.

(The book on causality from the perspective of probabilistic graphical models)

- J. Pearl, D. Mackenzie: The book of why, Basic Books, 2018

(Popular science level, but definitely worth a read)

Color Conventions for part on SCMs

- Formulae will be encoded in this greenish color

- Newly introduced terminology and definitions will be given in blue

- Important results (observations, theorems) as well as emphasizing some aspects will be given in red

- Examples will be given with standard orange

- Comments and notes are given with

  post-it-yellow background

Motivation

- Usual warning:

  „Correlation is not causation“

- But sometimes (if not very often) one needs causation to understand statistical data

A remarkable correlation? A simple causality!

(Mentioned in a feauture called the „Mops des Monats“ in
the book radio show „Lesart“
on Deutschlandfunk Kultur)

Simpson's Paradox (Example)

- Record recovery rates of 700 patients given access to a drug

	Recovery rate with drug	Recovery rate without drug
Men	81/87 (93%)	234/270 (87%)
Women	192/263 (73%)	55/80 (69%)
Combined	273/350 (78%)	289/350 (83%)

- Paradox:
  - For men, taking drugs has benefit
  - For women, taking drugs has benefit, too.
  - But: for all persons taking drugs has no benefit

Resolving the Paradox (Informally)

- We have to understand the causal mechanisms that lead to the data in order to resolve the paradox

- In drug example
  - Why has taking drug less benefit for women?

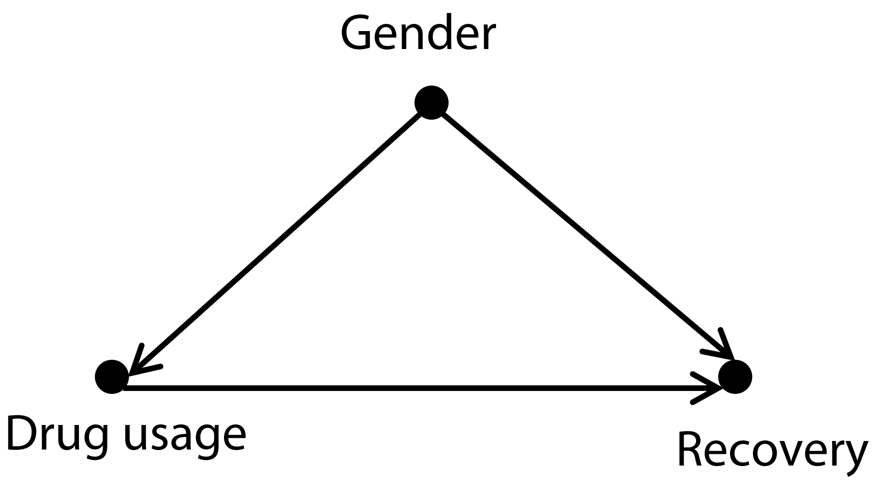
    Answer: Estrogen has negative effect on recovery
  - Data: Women more likely to take drug than men
  - Choosing randomly any person taking drugs will rather give a woman – and for these recovery is less beneficial

- In this case: Have to consider segregated data

  (not aggregated data)

Resolving the Paradox Formally (Look ahead)

- We have to understand the causal mechanisms that lead to the data in order to resolve the paradox



- Drug usage and recovery have common cause
- Gender is a confounder

Simpson's Paradox (Again)

- Record recovery rates of 700 patients given access to a drug w.r.t. blood pressure (BP) segregation

	Recovery rate **without** drug	Recovery rate **with** drug
Low BP	81/87 (93%)	234/270 (87%)
High BP	192/263 (73%)	55/80 (69%)
Combined	273/350 (78%)	289/350 (83%)

- BP recorded at end of experiment
- This time segregated data recommend not using drug whereas aggregated data does

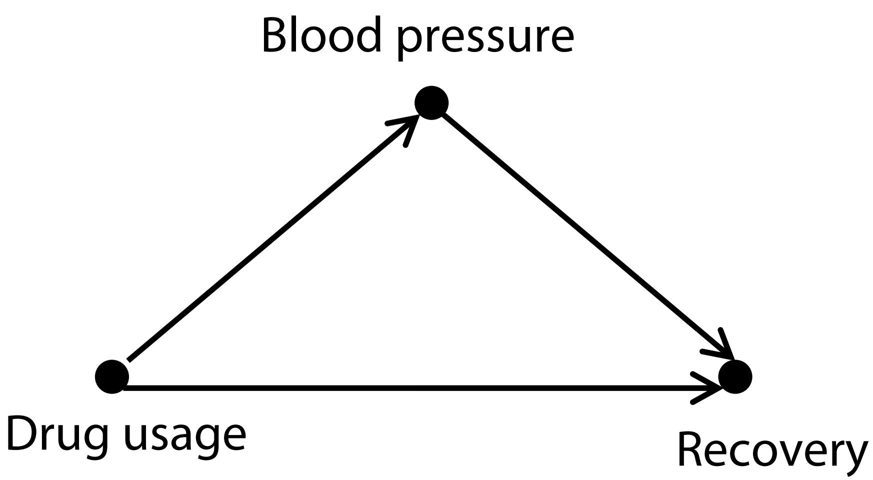
Resolving the Paradox (Informally)

- We have to understand the causal mechanisms that lead to the data in order to resolve the paradox

- In this example
  - Drug effect is: lowering blood pressure (but may have toxic effects)
  - Hence: In aggregated population drug usage recommended
  - In segregated data one sees only toxic effects

Resolving the Paradox Formally (Lookahead)

- We have to understand the causal mechanisms that lead to the data in order to resolve the paradox (look ahead)



Ingredients of a Statistical Theory of Causality

- Working definition of causation
- Method for creating causal models
- Method for linking causal models with features of data
- Method for reasoning over model and data

Working Definition

A (random) variable X is a cause of a (random) variable Y if Y - in any way - relies on X for its value

Structural Causal Model: Definition

Definition

A structural causal model (SCM) consists of

- A set U of exogenous variables
- A set V of endogenous variables
- A set of functions f assigning each variable in V a value based on values of other variables from V ∪ U

- Only endogenous variables are those that are descendants of other variables
- Exogenous variables are roots of model.
- Value instantiations of exogenous variables completely determine values of all variables in SCM

Causality in SCMs

Definition

1. X is a direct cause of Y iff $Y = f(\ldots,X,\ldots)$ for some f.
2. X is a cause of Y iff X is a direct cause of Y or there is Z s.t. X is a direct cause of Z and Z is a cause of Y.

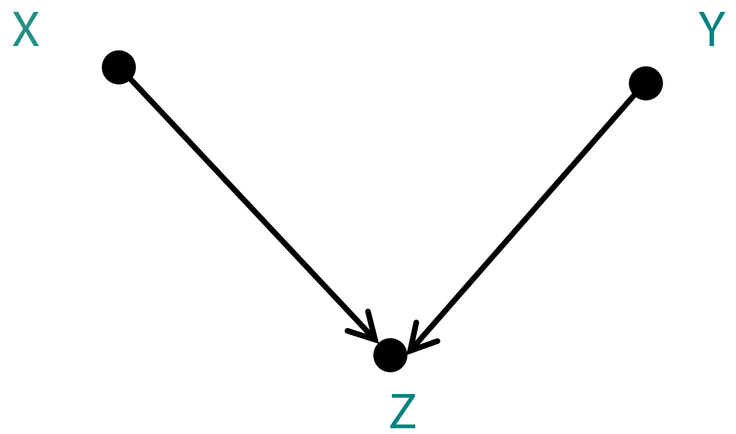
Graphical Causal Model

- Graphical causal model associated with SCM
  - Nodes = variables
  - Edges = from X to Y if $Y = f(\ldots, X, \ldots)$

- Example SCM
  - $U = \{X, Y\}$
  - $V = \{Z\}$
  - $F = \{f_Z\}$
  - $f_Z : Z = 2X + 3Y$

(Z = salary, X = years of experience,
Y = years of profession)

- Associated graph



Graphical Models

- Graphical models capture only partially SCMs
- But very intuitive and still allow for conserving much of causal information of SCM
- Convention for the next lectures: Consider only Directed Acyclic Graphs (DAGs)

SCMs and Probabilities

- Consider SCMs where all variables are random variables (RVs)

- Full specification of functions f not always possible

- Instead: Use conditional probabilities as in BNs
  - $f_X(\ldots Y \ldots)$ becomes $P(X \mid \ldots Y \ldots)$
  - Technically: Non-measurable RV U models (probabilistic) indeterminism:

    $P(X \mid \ldots . Y \ldots .) = f_X(\ldots Y \ldots, U)$

U not mentioned here

SCMs and Probabilities

- Product rule as in Bayesian networks (BNs) used for full specification of joint distribution of all RVs $X_1, \ldots, X_n$

$$P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{1 \leq i \leq n} P(\, x_i \mid \text{parentsof}(x_i)\,)$$

- Can make same considerations on (probabilistic) (in)dependence of RVs.
- Will be done in the following systematically

Bayesian Networks vs. SCMs

- BNs model statistical dependencies
  - Directed, but not necessarily cause-relation
  - Inherently statistical
  - Default application: discrete variables
- SCMs model causal relations
  - SCMs with random variables (RVs) induce BNs
  - Assumption: There is hidden causal (deterministic) structure behind statistical data
  - More expressive than BNs: Every BN can be modeled by SCMs but not vice versa
  - Default application: continuous variables

Reminder: Conditional Independence

- Event A independent of event B iff $P(A \mid B) = P(A)$
- RV X is independent of RV Y iff

  $P(X \mid Y) = P(X)$ iff

  for every x-value of X and for every y-value Y

  event $X = x$ is independent of event $Y = y$

  Notation: $(X \perp\!\!\!\perp Y)_P$ or even shorter: $(X \perp\!\!\!\perp Y)$
- X is conditionally independent of Y given Z iff

  $P(X \mid Y, Z) = P(X \mid Z)$

  Notation: $(X \perp\!\!\!\perp Y \mid Z)_P$ or even shorter: $(X \perp\!\!\!\perp Y|Z)$

Compatible distributions

Definition

Given a DAG G on RVs $\mathbf{X} = X_1, \ldots, X_n$, and a probability distribution P over $\mathbf{X}$ we say that P is compatible with G iff it allows a factorization of the form

$$\mathbf{P}(X_1, \ldots, X_n) = \boldsymbol{\pi}_{i=1}\, \mathbf{P}(X_i \mid Parents(X_i))$$

Independence in SCM graphs

- Almost all interesting independences of RVs in an SCM can be identified in its associated graph

- Relevant graph theoretical notion: d-separation

Property

X is independent of Y conditioned on Z for all distributions compatible with G
iff
X is d-separated from Y by Z in G

- D-separation in turn rests on 3 basic graph patterns
  - Chains
  - Forks
  - Colliders

Independence in SCM graphs

Property
X is independent of Y conditioned on Z for all distributions compatible with G
iff
X is d-separated from Y by Z in G

Note the scope of the all-quantifier.

Markov condition (<- direction):

If X is d-separated from Y by Z

then X is independent of Y conditioned on Z for all probability distributions compatible with G

Independence in SCM graphs

Property

X is independent of Y conditioned on Z for all distributions compatible with G
iff
X is d-separated from Y by Z in G

Faithfulness (-> direction):

If	X is **not** d-separated from Y by Z
then	X is **not** independent of Y conditioned on Z for **some** probability distribution compatible with G

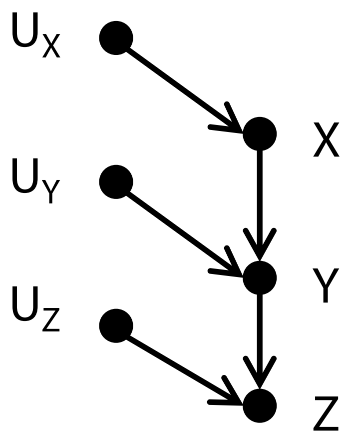
- Note: We do not have: "For all distributions"; but usually the "some" is stronger: "for many"

Chains

Example (SCM 1)

(X = school funding of high school , Y = its average SAT score, Z = average college acceptance)

- $V = \{X,Y,Z\}$ $\quad U = \{U_X,U_Y,U_Z\}$ $\quad F = \{f_X,f_Y,f_Z\}$
- f_X: $X = U_X$ $\quad f_Y$: $Y = x/3 + U_Y$ $\quad f_Z$: $Z = y/16 + U_Z$

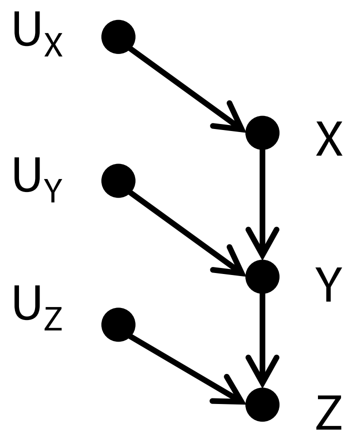


Chains

Example (SCM 2)

(X = switch, Y = circuit, Z = light bulb)

- $V = \{X,Y,Z\}$ $U = \{U_X,U_Y,U_Z\}$ $F = \{f_X,f_Y,f_Z\}$
- f_X: $X = U_X$
- $f_Y: Y = \begin{cases} \text{closed} & \text{if } (X = \text{up} \,\&\, U_Y = 0) \text{ or } (X = \text{down} \,\&\, U_Y = 1) \\ \text{open} & \text{otherwise} \end{cases}$
- $f_Z: Z = \begin{cases} \text{on} & \text{if } (Y = \text{closed} \,\&\, U_Z = 0) \text{ or } (Y = \text{open} \,\&\, U_Z = 1) \\ \text{off} & \text{otherwise} \end{cases}$

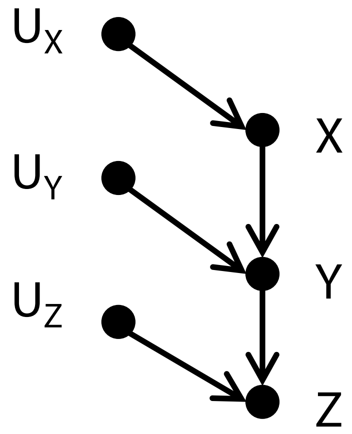


Chains

Example (SCM 3)

(X = work hours, Y = training, Z = race time)

- $V = \{X,Y,Z\}U = \{U_X,U_Y,U_Z\}$ $F = \{f_X,f_Y,f_Z\}$
- f_X: $X = U_X$
- f_Y: $Y = 84 - x + U_Y$
- f_Z: $Z = 100/y + U_Z$



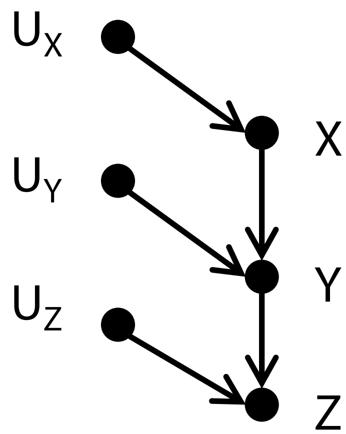
(In)Dependences in Chains

- Z and Y are likely dependent

  (For some z,y: $P(Z{=}z \mid Y = y) \neq P(Z = z)$)
- Y and X are likely dependent

  (…)
- Z and X are likely dependent
- Z and X are independent, conditioned on Y

  (For all x,z,y: $P(Z{=}z \mid X{=}x, Y = y) = P(Z = z \mid Y = y)$)



Dependence not Transitive

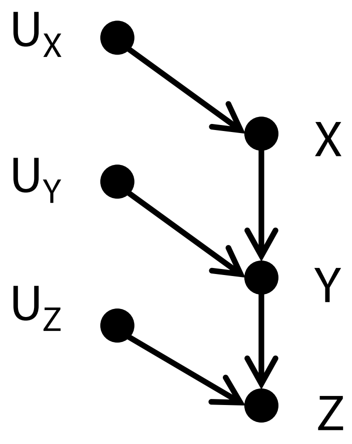
Example (SCM 4)

$V = \{X,Y,Z\}$ $U = \{U_X,U_Y,U_Z\}$ $F = \{ f_X,f_Y,f_Z \}$

– f_X: $X = U_X$

– f_Y: $Y = \begin{cases} a & \text{if } X = 1 \;\&\; U_Y = 1 \\ b & \text{if } X = 2 \;\&\; U_Y = 1 \\ c & \text{if } U_Y = 2 \end{cases}$

– f_Z: $Z = \begin{cases} i & \text{if } Y = c \quad \text{or} \quad U_Z = 1 \\ j & \text{if } Y \neq c \quad \& \quad U_Z = 2 \end{cases}$



Typo in book of Pearl et al.

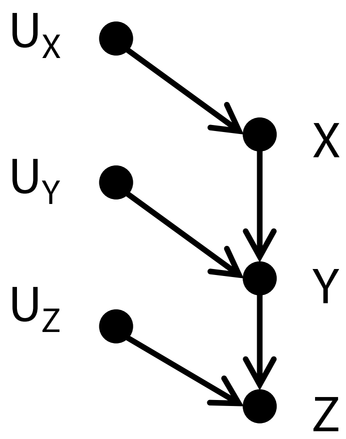
- Y depends on X, Z depends on Y **but** Z does not depend on X
- ``Variable level“ graph hides **in**dependence

Independence Rule in Chains

Rule 1 (Conditional Independence in Chains)

Variables X and Z are independent given set of variables Y iff

there is only one path between X and Z and this path is unidirectional and Y intercepts that path

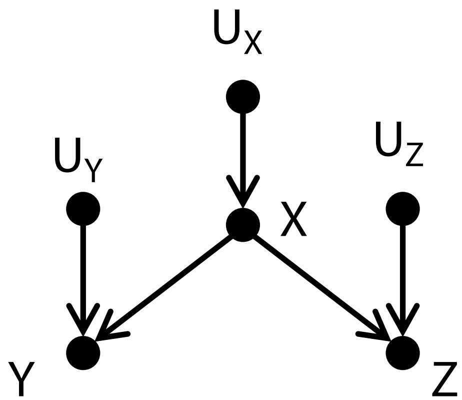


Forks

Example (SCM 5)

(X = Temperature, Y = Ice cream sale, Z = Crime)

- $V = \{X,Y,Z\}$ $U = \{U_X,U_Y,U_Z\}$ $F = \{f_X,f_Y,f_Z\}$
- f_X: $X = U_X$
- f_Y: $Y = 4x + U_y$
- f_Z: $Z = x/10 + U_Z$

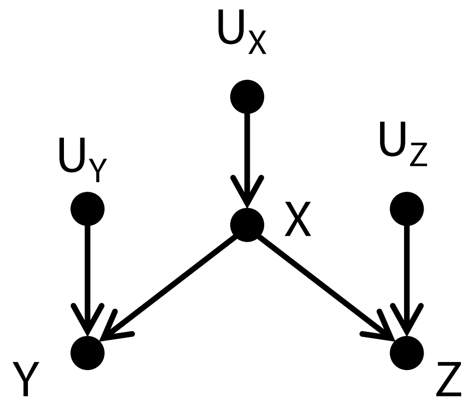


Forks

Example (SCM 5)

(X = switch, Y = light bulb 1, Z = light bulb 2)

- $V = \{X,Y,Z\}$ $U = \{U_X,U_Y,U_Z\}$ $F = \{f_X,f_Y,f_Z\}$
- f_X: $X = U_X$
- $f_Y: Y = \begin{cases} \text{on} & \text{if } (X = \text{up} \,\&\, U_Y = 0) \text{ or } (X = \text{down} \,\&\, U_Y = 1) \\ \text{off} & \text{otherwise} \end{cases}$
- $f_Z: Z = \begin{cases} \text{on} & \text{if } (X = \text{up} \,\&\, U_Z = 0) \text{ or } (X = \text{down} \,\&\, U_Z = 1) \\ \text{off} & \text{otherwise} \end{cases}$



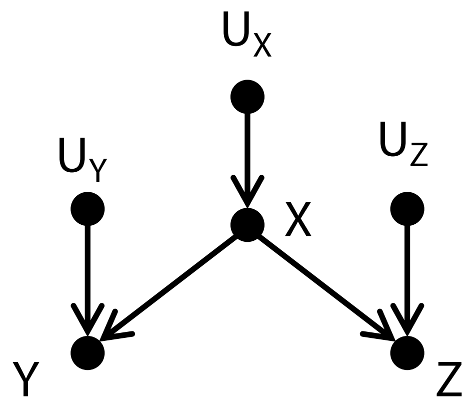
(In)Dependences in Forks

- X and Z are likely dependent

  ($\exists z,y$: $P(X=x \mid Z = z) \neq P(X = x)$)
- Y and X are likely dependent

  …
- Z and Y are likely dependent
- Y and Z are independent, conditioned on X

  ($\forall x,z,y$: $P(Y=y \mid Z=z, X = x) = P(Y = y \mid X = x)$)

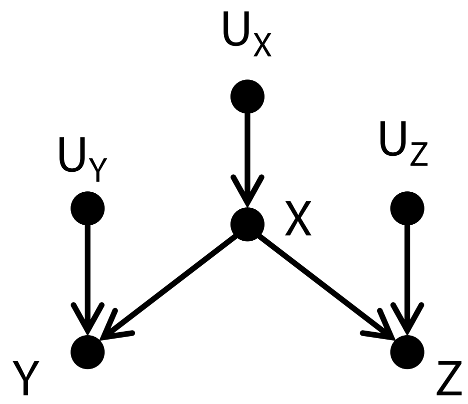


Independence Rule in Forks

Rule 2 (Conditional Independence in Forks)

If variable X is a common cause of variables Y and Z and there is only one path between Y, Z

then Y and Z are conditionally independent given X.

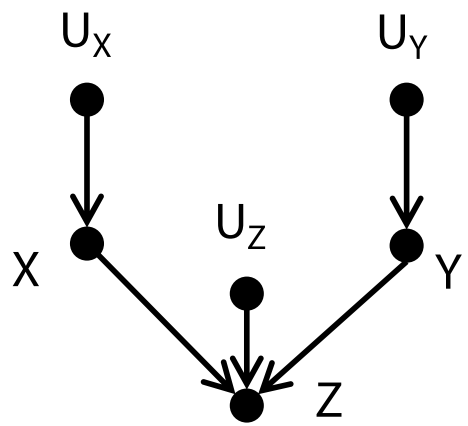


Colliders

Example (SCM 6)

(X = musical talent, Y = grade point, Z = scholarship)

- $V = \{X,Y,Z\}$ $U = \{U_X,U_Y,U_Z\}$ $F = \{f_X,f_Y,f_Z\}$
- f_X: $X = U_X$
- f_Y: $X = U_Y$
- f_Z: $Z = \begin{cases} \text{yes} & \text{if } X = \text{yes or } Y > 80\% \\ \text{no} & \text{otherwise} \end{cases}$



(In)dependence in Colliders

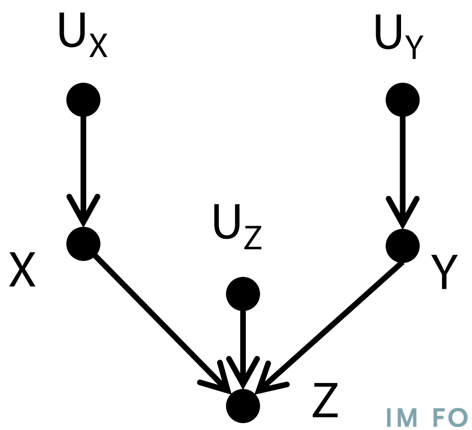
- X and Z are likely dependent

  ($\exists z,y$: $P(X=x \mid Z = z) \neq P(X = x)$)
- Y and Z are likely dependent
- X and Y are independent
- X and Y are likely conditionally dependent, given Z

$$\boxed{(\exists x,z,y: P(X= x \mid Y=y, Z = z) \neq P(X = x \mid Z = z))}$$

If scholarship received (Z)
but low grade (Y),
then must be musically talented (X)

X-Y dependence (conditioned on Z)
is statistical but not causal

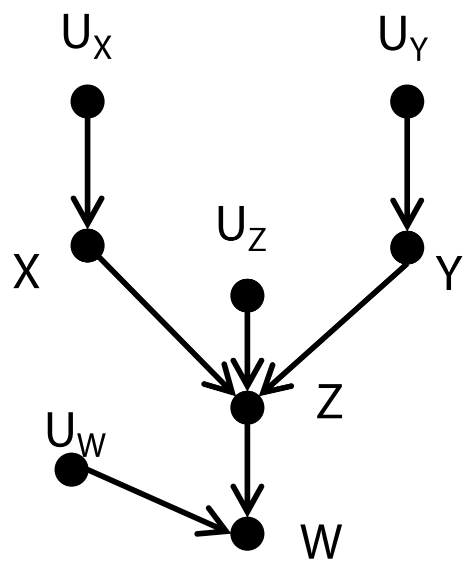


(In)dependence in Colliders (Extended)

Example (SCM 7)

(X = coin flip, Y = second coin flip,
Z = bell rings, W = bell witness)

- $V = \{X,Y,Z,W\}$ $U = \{U_X,U_Y,U_Z, U_W\}$ $F = \{f_X,f_Y,, f_W\}$
- f_X: $X = U_X$
- f_Y: $Y = U_Y$
- f_Z: $Z = \begin{cases} \text{yes} & \text{if } X = \text{head or } Y = \text{head} \\ \text{no} & \text{otherwise} \end{cases}$
- f_W: $W = \begin{cases} \text{yes} & \text{if } Z = \text{yes or } (Z = \text{no and } U_W = ½) \\ \text{no} & \text{otherwise} \end{cases}$



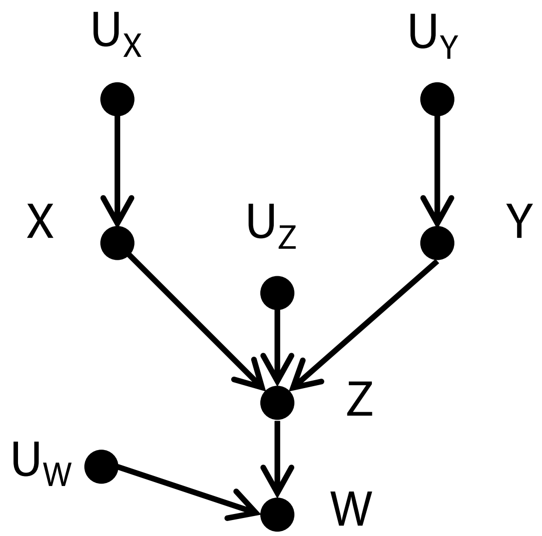
X and Y are dependent conditioned on Z
(and also on W).

Independence Rule in Colliders

Rule 3 (Conditional Independence in Colliders)

If a variable Z is the collision node between variables X and Y and there is only one path between X, Y,

then X and Y are unconditionally independent, but are dependent conditional on Z and any descendant of Z



D-separation

Property
X independent of Y (conditioned on Z) for all compatible distributions iff
X d-separated from Y by Z in graph

Definition (informal)
X is d-separated from Y by Z iff
Z blocks every possible path between X and Y

- Z (possibly a set of variables) prohibits the ``flow'' of statistical effects/dependence between X and Y
  - Must block every path
  - Need only one blocking variable for each path

Pipeline metaphor

Blocking Conditions

Definition (formal)

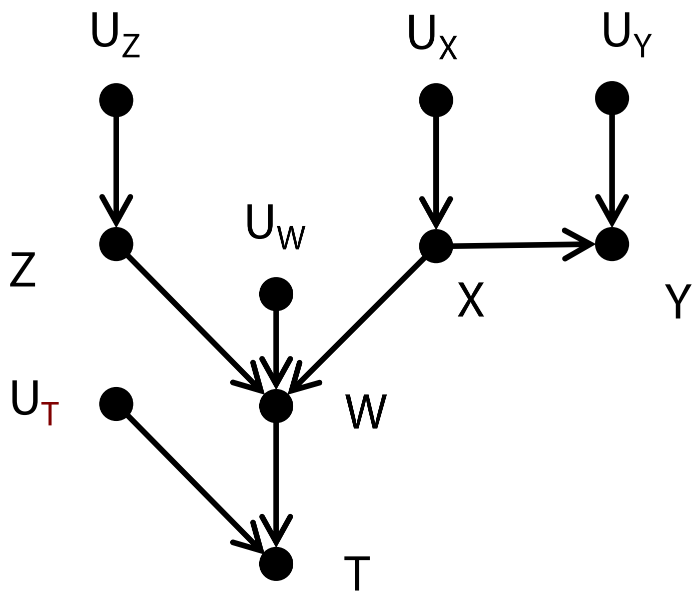
A path p in G (between X and Y) is blocked by Z iff

1. p contains chain $A \longrightarrow B \longrightarrow C$ or fork $A \longleftarrow B \longrightarrow C$ s.t. $B \in Z$ or
2. p contains collider $A \longrightarrow B \longleftarrow C$ s.t. $B \notin Z$ and all descendants of B are $\notin Z$

If Z blocks every path between X and Y, then X and Y are d-separated conditional on Z, for short: $(X \perp\!\!\!\perp Y \mid Z)_G$

In particular: X and Y are unconditionally independent iff X-Y paths contain collider.

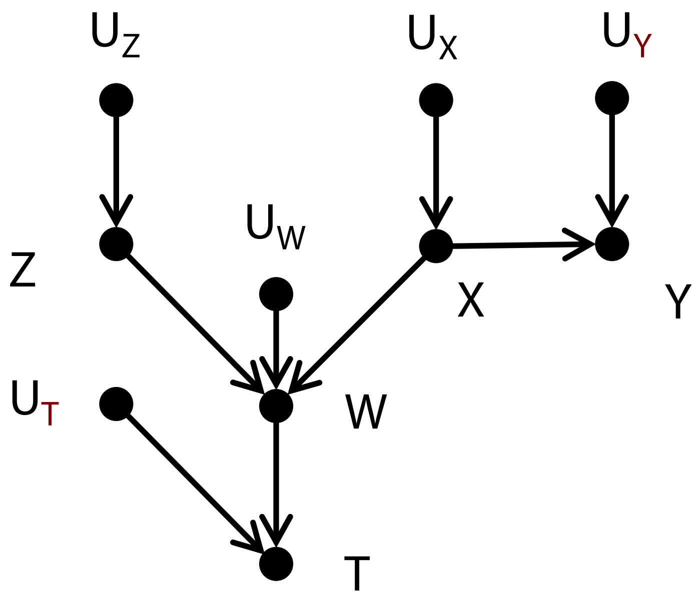
Example 1 (d-separation)



- Unconditional relation between Z and Y ?
  - D-separated because of collider on single Z-Y path.

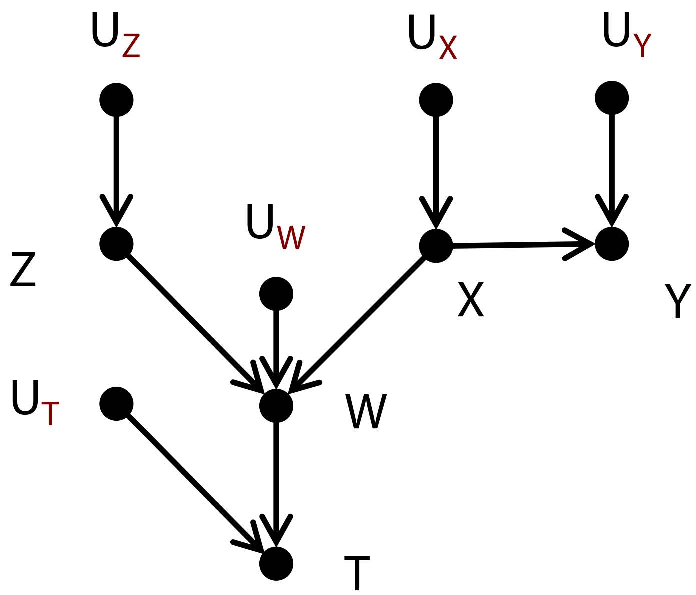
    Hence unconditionally independent

Example 1 (d-separation)



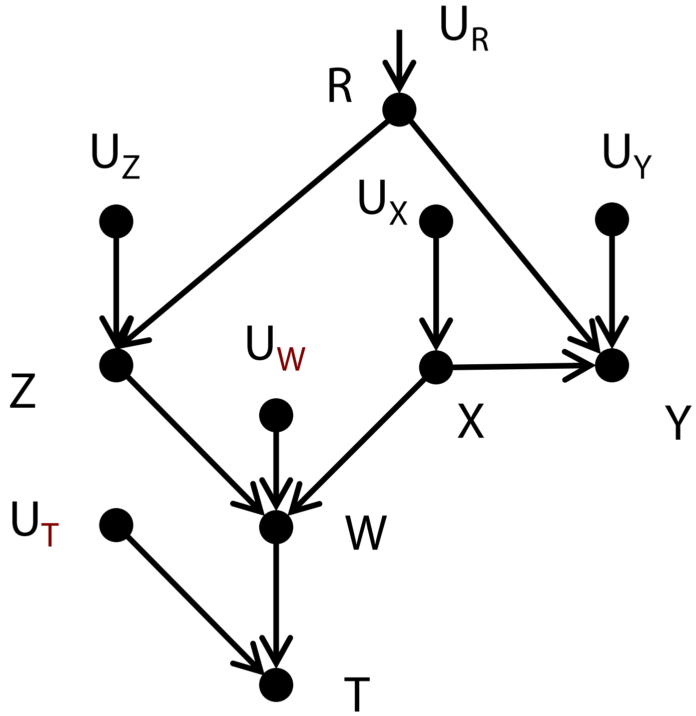
- Relation between Z and Y conditioned on {W}?
  - Not d-separated
    - because fork $X \notin \{W\}$
    - and collider $\in \{W\}$
  - Hence conditionally dependent on {W} (and {T})

Example 1 (d-separation)



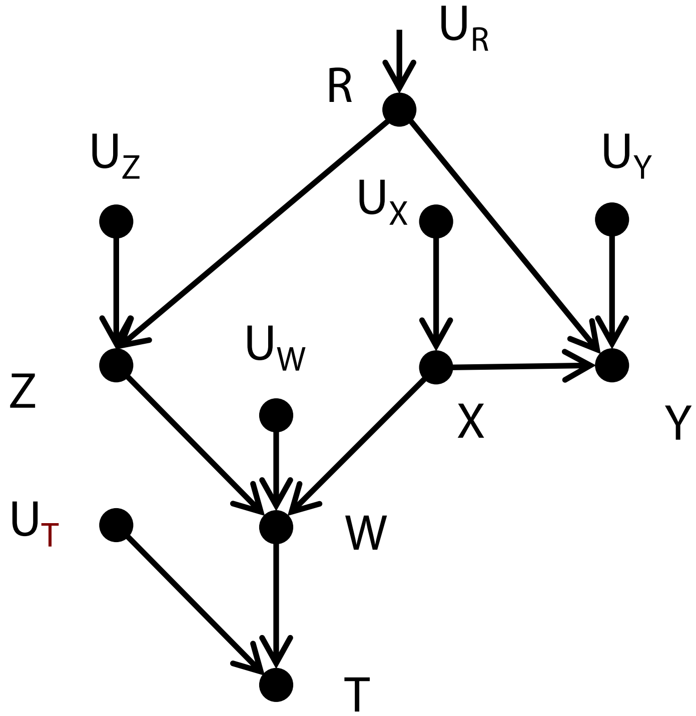
- Relation between Z and Y conditional on {W,X}?
  - d-separated
    - Because fork X blocks
  - Hence conditionally independent on {W,X}

Example 2 (d-separation)



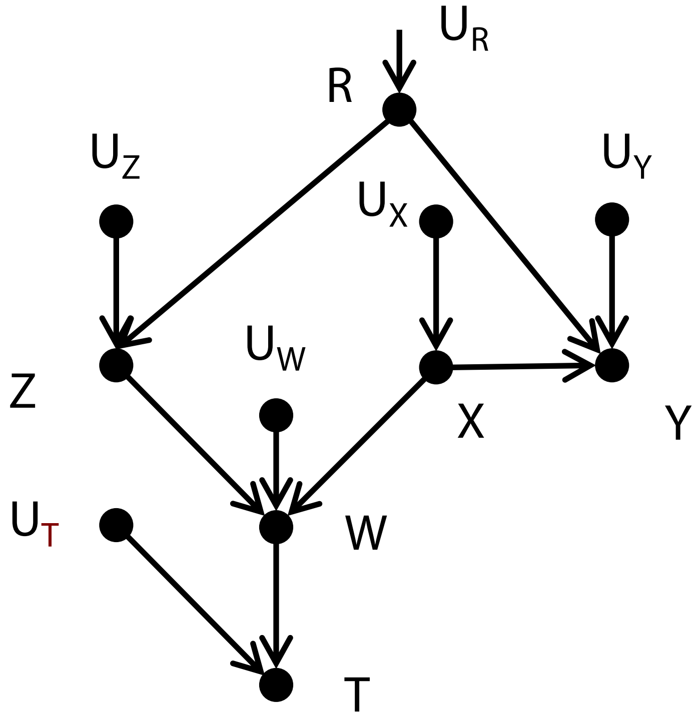
- Relation between Z and Y?
  - Not d-separated because second path not blocked (no collider)
  - Hence not unconditionally independent

Example 2 (d-separation)



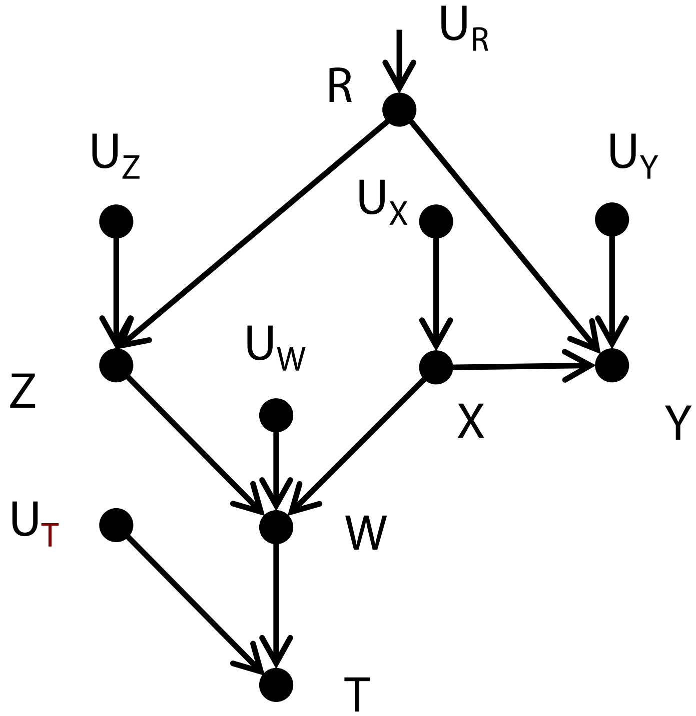
- Relation between Z and Y conditionally on {R}?
  - d-separated by {R} because
    - First path blocked by fork R
    - second path blocked by collider $W \notin \{R\}$)
  - Hence independent conditioned on {R}

Example 2 (d-separation)



- Relation between Z and Y conditionally on {R,W}?
  - Not d-separated by {R,W} because W unblocks second path
  - Hence not independent conditioned on {R,W}

Example 2 (d-separation)



- Relation between Z and Y conditionally on {R,W,X}?
  - d-separated by {R,W,X} because
    - Now second path blocked by fork X
  - Hence independent conditioned on {R,W,X}

Using D-separation

- Verifying/falsifying causal models on observational data
  1. G = SCM to test for
  2. Calculate independencies I_G entailed by G using d-separation
  3. Calculate independencies I_D from data (by counting and estimating probabilities) and compare with I_G
  4. If $I_G = I_D$, SCM is a good solution. Otherwise identify problematic $I \in I_G$ and change G locally to fit corresponding $I' \in I_D$

Using D-separation

- This approach is local
  - If I_G not equal I_D, then can manipulate G w.r.t. RVs only involved in incompatibility
  - Usually seen as benefit w.r.t. global approaches via likelihood with scores, say
  - Note: In score-based approach one always considers score of whole graph

    (But: one also aims at decomposability/locality of scoring functions)
- This approach is qualitative and constraint based

- Known algorithms: PC (Spirtes) , IC (Verma & Pearl)

Equivalent Graphs

- One learns graphs that are (observationally) equivalent w.r.t. entailed independence assumptions
- Formalization
  - v(G) = v-structure of G = set of colliders in G of form $A \longrightarrow B \longleftarrow C$ where A and C not adjacent
  - sk(G) = skeleton of G = undirected graph resulting from G

Definition
G_1 is equivalent to G_2 iff $v(G_1) = v(G_2)$ and $sk(G_1) = sk(G_2)$

Equivalent Graphs

Theorem
Equivalent graphs entail same set of d-separations

Intuitively clear:

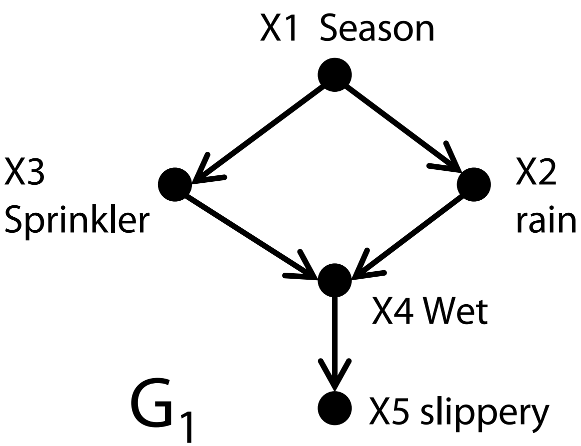
- Forks and chains have similar role w.r.t. independence (Hence forgetting about the direction in skeleton does not lead to loss of information)
- Collider has different role (hence need v-structure)

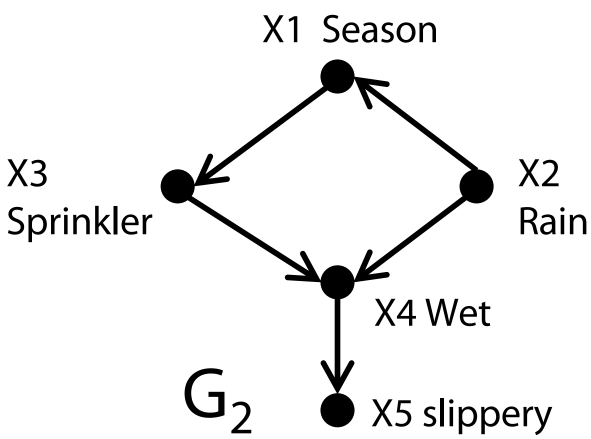
Equivalent Graphs

- v(G) = v-structure of G = set of colliders in G of form $A \rightarrow B \leftarrow C$ where A and C not adjacent
- sk(G) = skeleton of G = undirected graph resulting from G

Definition

G_1 is equivalent to G_2 iff $v(G_1) = v(G_2)$ and $sk(G_1) = sk(G_2)$





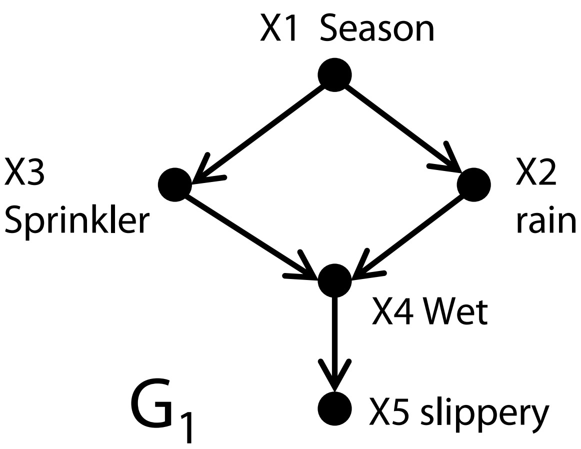
- $v(G_1) = v(G_2)$
- $sk(G_1) = sk(G_2)$
- Hence equivalent

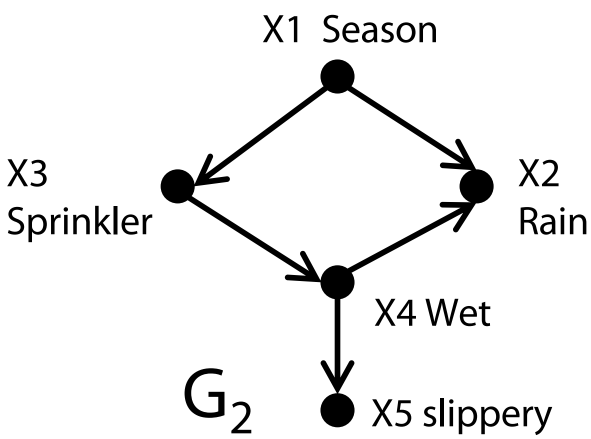
Equivalent Graphs

- v(G) = v-structure of G = set of colliders in G of form $A \rightarrow B \leftarrow C$ where A and C not adjacent
- sk(G) = skeleton of G = undirected graph resulting from G

Definition

G_1 is equivalent to G_2 iff $v(G_1) = v(G_2)$ and $sk(G_1) = sk(G_2)$





- $v(G_1) \neq v(G_2)$
- $sk(G_1) = sk(G_2)$
- Hence not equivalent

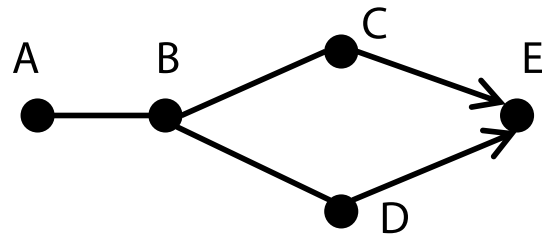
IC-Algorithm (Verma & Pearl, 1990)

Input
P resp.
P-independencies

$(C \perp\!\!\!\perp A \mid B)$
$(C \perp\!\!\!\perp D \mid B)$
$(D \perp\!\!\!\perp A \mid B)$
$(E \perp\!\!\!\perp A \mid B)$
$(E \perp\!\!\!\perp B \mid C,D)$

Algorithm
→
Steps 1-3

Output
Pattern
(represents compatible class of equivalent DAGs)



Definition

Pattern = partially directed DAG
= DAG with directed and non-directed edges

Directed edge A-> B in pattern: In any of the DAGs the edge is A->B
Undirected edge A-B in pattern: There exists (equivalent) DAGs with A->B in one and B ->A in the other

Verma, T. & Pearl, J: Equivalence and synthesis of causal models.
Proceedings of the 6. conference on Uncertainty in AI, 220-227, 1990.

IC-Algorithm (Informally)

1. Find all pairs of variables that are dependent of each other (applying standard statistical method on the database) and eliminate indirect dependencies

2. + 3. Determine directions of dependencies

Note: „Possible“ in step 3 means: if you can find two patterns such that in the first the edge A-B becomes A->B but in the other A<-B, then do not orient.

IC-Algorithm (schema)

1. Add (undirected) edge A-B iff there is no set of RVs **Z** such that $(A \perp\!\!\!\perp B \mid \mathbf{Z})_{P}$. Otherwise let Z_{AB} denote some set **Z** with $(A \perp\!\!\!\perp B \mid \mathbf{Z})_{P}$.
2. If A–B–C and not A-C, then $A \rightarrow B \leftarrow C$ iff $B \notin Z_{AC}$
3. Orient as many of the undirected edges as possible, under the following constraints:
   - orientation should not create a new v-structure and
   - orientation should not create a directed cycle.

Steps 1 and step 3 leave out details of search
- Hierarchical refinement of step 1 gives PC algorithm (next slide)
- A refinement of step 3 possible with 4 rules (thereafter)

PC algorithm (Spirtes & Glymour, 1991)

- Remember Step 1 of IC
  1. Add (undirected) edge A-B iff there is no set of RVs **Z** such that $(A \perp\!\!\!\perp B \mid \mathbf{Z})_P$. Otherwise let Z_{AB} denote some set **Z** with $(A \perp\!\!\!\perp B \mid \mathbf{Z})_P$.
- Have to search **all** sets **Z** of RVs for given nodes A,B
  - Start with fully connected graph (with undirected edges)
  - Done systematically by sets of cardinality 0,1,2,3…
  - Remove edges from graph as soon as independence found
  - Polynomial time for graphs of finite degree (because can restrict search for **Z** to nodes adjacent to A,B)

P.Spirtes, C. Glymour: An algorithm for fast recovery of sparse causal graphs. Social Science Computer Review 9: 62-72, 1991.

IC-Algorithm (with rule-specified last step)

1. as before
2. as before
3. Orient undirected edges as follows
   - B — C into B→C if there is an arrow A→B s.t. A and C are not adjacent;
   - A — B into A→B if there is a chain A→C→B;
   - A — B into A→B if there are two chains A—C→B and A—D→B such that C and D are nonadjacent;
   - A — B into A→B if there are two chains A—C→D and C→D→B s.t. C and B are nonadjacent;

IC algorithm

Theorem

The 4 rules specified in step 3 of the IC algorithm are necessary (Verma & Pearl, 1992) and sufficient (Meek, 95) for getting a maximally oriented pattern of DAGs compatible with the input-independencies.

T. Verma and J. Pearl. An algorithm for deciding if a set of observed independencies has a causal explanation. In D. Dubois and M. P. Wellman, editors, UAI '92: Proceedings of the Eighth Annual Conference on Uncertainty in Artificial Intelligence, 1992, pages 323–330. Morgan Kaufmann, 1992.

Christopher Meek: Causal inference and causal explanation with background knowledge. UAI 1995: 403-410, 1995.

Stable Distribution

- The IC algorithm accepts stable distributions P (over set of variables) as input, i.e. distribution P s.t. there is DAG G giving exactly the P-independencies

- Extension IC* works also for sampled distributions generated by so-called latent structures
  - A latent structure (LS) specifies additionally a (subset) of observation variables for a causal structure
  - A LS not determined by independencies
  - IC* not discussed here, see, e.g.,

  J. Pearl: Causality, CUP, 2001, reprint, p. 52-54.

Criticism and further developments

Definition
The problem of ignorance denotes the fact that there are RVs A, B and sets of RVs Z such that it is not known whether $(A \perp\!\!\!\perp B \mid Z)_P$ or not $(A \perp\!\!\!\perp B \mid Z)_P$

- Problem of ignorance ubiquitous in science practice
- IC faces the problem of ignorance (Leuridan 2009)
- (Leuridan 2009) approaches this with adaptive logic

B. Leuridan. Causal discovery and the problem of ignorance: an adaptive logic approach. Journal of Applied Logic, 7(2):188–205, 2009.

Uhhh, a lecture with a hopefully useful

APPENDIX

Probability theory basics reminder

Random variable (RV)

- possible worlds defined by assignment of values to random variables.
- Boolean random variables
  e.g., Cavity (do I have a cavity?).
  Domain is < true , false >
- Discrete random variables
  e.g., possible value of Weather is one of
  < sunny, rainy, cloudy, snow >
- Domain values must be exhaustive and mutually exclusive
- Elementary propositions are constructed by assignment of a value to a random variable: e.g.,
  - Cavity = false (abbreviated as ¬cavity)
  - Cavity = true (abbreviated as cavity)
- (Complex) propositions formed from elementary propositions and standard logical connectives, e.g., Weather = sunny ∨ Cavity = false

Probabilities

- Axioms (for propositions $a, b, \top = (a \vee \neg a)$, and $\bot = \neg \top$):
  - $0 \leq P(a) \leq 1;\ P(\top) = 1;\ P(\bot) = 0$
  - $(P(a \vee b) = P(a) + P(b) - P(a \wedge b)$
- Joint probability distribution of $\mathbf{X} = \{X_1, \dots, X_n\}$
  - $\boldsymbol{P}(X_1, \dots, X_n)$
  - gives the probability of every atomic event on $\boldsymbol{X}$
- Conditional probability
  $P(a \mid b) = P(a \wedge b) / P(b)\ if\ P(b) > 0$
- Chain rule
  $$\boldsymbol{P}(X_1, \dots, X_n) = \prod_{i=1}^{n} \boldsymbol{P}(X_i | X_1, \dots, X_{i-1})$$
- Marginalization: $\boldsymbol{P}(Y) = \sum_{z \in Z} \boldsymbol{P}(Y, z)$
- Conditioning on Z:
  - $\boldsymbol{P}(Y) = \sum_{z \in Z} \boldsymbol{P}(Y|z)\boldsymbol{P}(z)$ (discrete)
  - $\boldsymbol{P}(Y) = \int \boldsymbol{P}(Y|z)\boldsymbol{P}(z)dz$ (continuous)
    $= \mathbb{E}_{\boldsymbol{z} \sim P(z)} \mathrm{P(Y|z)}$ (expected value notation)
- Bayes' Rule
  $$P(H|D) = \frac{P(D|H) \cdot P(H)}{P(D)} = \frac{P(D|H) \cdot P(H)}{\sum_h P(D|h)P(h)}$$