Probabilistic Causal Networks

The Big Objective(s)

In a wide variety of application fields two main problems need to be addressed over and over:

1. **How can (expert) knowledge of complex domains be efficiently represented?**
2. **How can inferences be carried out within these representations?**
3. **How can such representations be (automatically) extracted from collected data?**

We will deal with all three questions during the lecture.

Example 1: Planning in car manufacturing

Available information

- "Engine type e_1 can only be combined with transmission t_2 or t_5."
- "Transmission t_5 requires crankshaft c_2."
- "Convertibles have the same set of radio options as SUVs."

Possible questions/inferences:

- "Can a station wagon with engine e_4 be equipped with tire set y_6?"
- "Supplier S_8 failed to deliver on time. What production line has to be modified and how?"
- "Are there any peculiarities within the set of cars that suffered an aircondition failure?"

Example 2: Medical reasoning

Available information:

- "Malaria is much less likely than flu."
- "Flu causes cough and fever."
- "Nausea can indicate malaria as well as flu."
- "Nausea never indicated pneunomia before."

Possible questions/inferences

- "The patient has fever. How likely is he to have malaria?"
- "How much more likely does flu become if we can exclude malaria?"

Common Problems

Both scenarios share some severe problems:

- **Large Data Space**
  It is intractable to store all value combinations, i. e. all car part combinations or inter-disease dependencies.

  (Example: VW Bora has 10^{200} theoretical value combinations*)

- **Sparse Data Space**
  Even if we could handle such a space, it would be extremely sparse, i. e. it would be impossible to find good estimates for all the combinations.

  (Example: with 100 diseases and 200 symptoms, there would be about 10^{62} different scenarios for which we had to estimate the probability.*)

* The number of particles in the observable universe is estimated to be between 10^{78} and 10^{85}.

Idea to Solve the Problems

- **Given:** A large (high-dimensional) distribution δ representing the domain knowledge.

- **Desired:** A set of smaller (lower-dimensional) distributions $\{\delta_1, \ldots, \delta_s\}$ (maybe overlapping) from which the original δ *could* be reconstructed with no (or as few as possible) errors.

- With such a decomposition we can draw any conclusions from $\{\delta_1, \ldots, \delta_s\}$ that could be inferred from δ — without, however, actually reconstructing it.

Example: Car Manufacturing

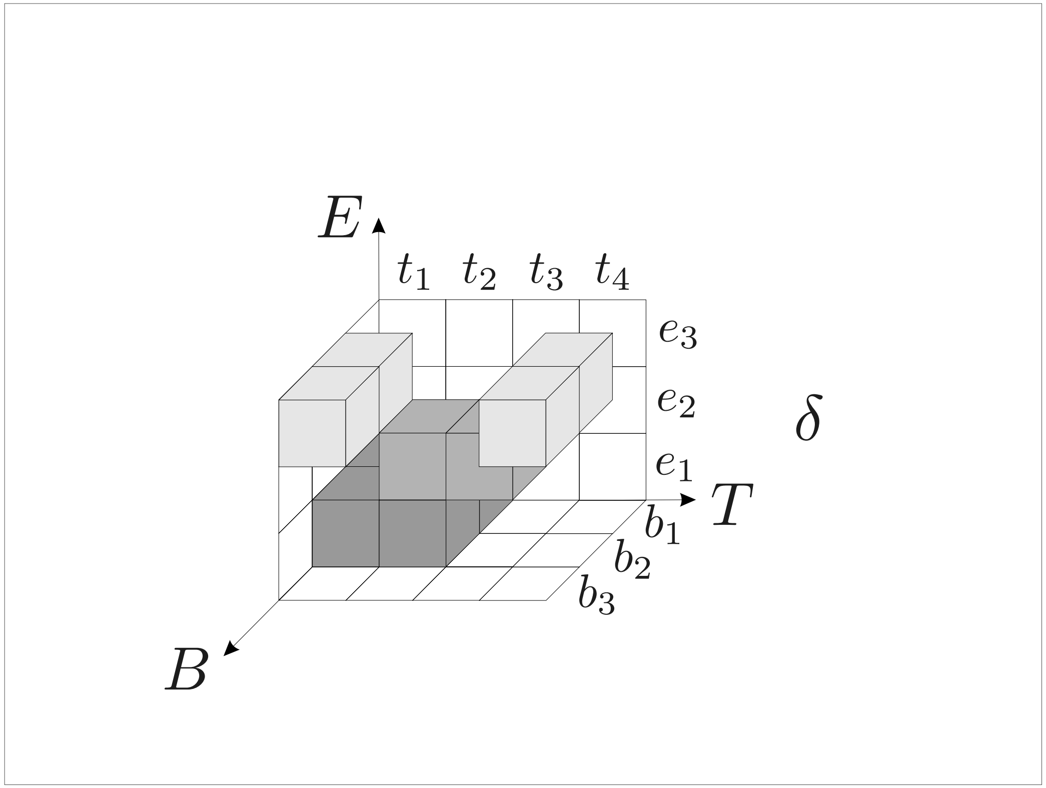
- Let us consider a car configuration is described by three attributes:
  - Engine E, $\mathrm{dom}(E) = \{e_1, e_2, e_3\}$
  - Breaks B, $\mathrm{dom}(B) = \{b_1, b_2, b_3\}$
  - Tires T, $\mathrm{dom}(T) = \{t_1, t_2, t_3, t_4\}$
- Therefore the set of all (theoretically) possible car configurations is:

$$\Omega = \mathrm{dom}(E) \times \mathrm{dom}(B) \times \mathrm{dom}(T)$$

- Since not all combinations are technically possible (or wanted by marketing) a set of rules is used to cancel out invalid combinations.

Example: Car Manufacturing

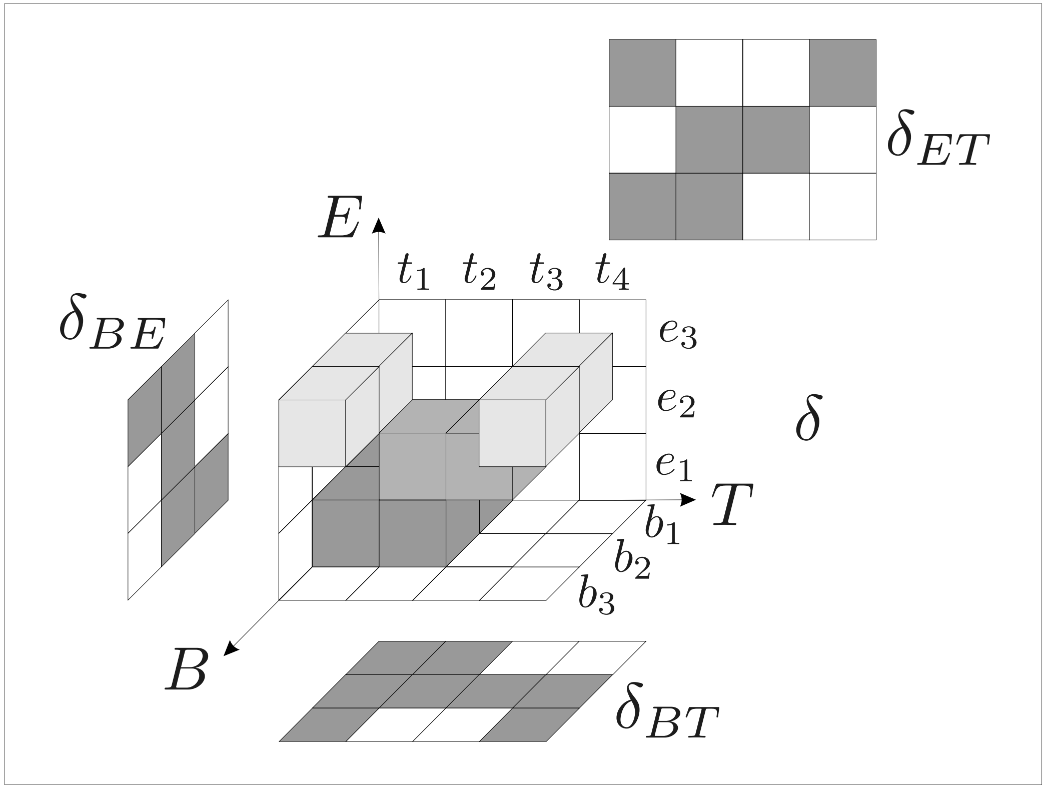
Possible car configurations



- Every cube designates a valid value combination.
- 10 car configurations in our model.
- Different colors are intended to distinguish the cubes only.

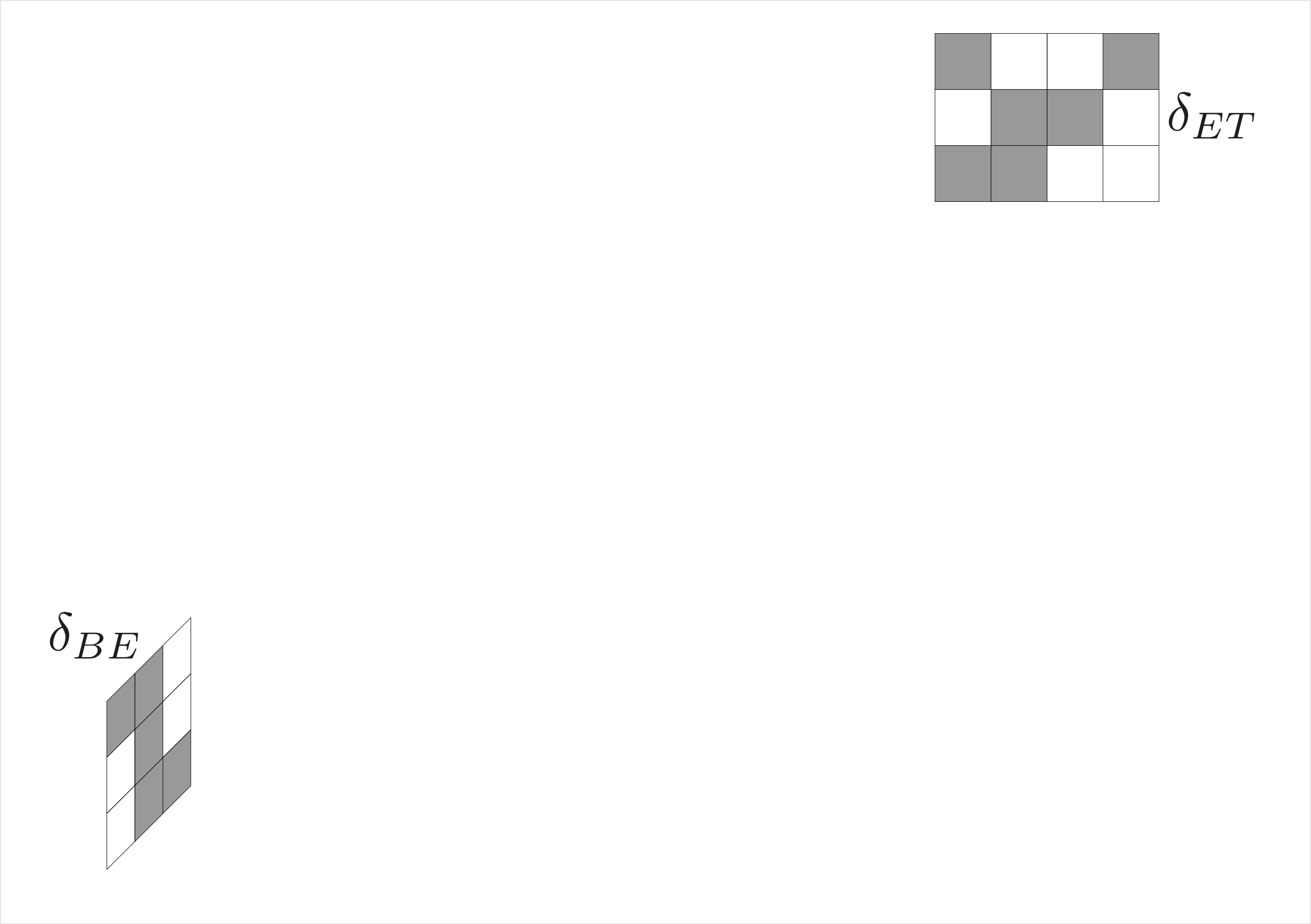
Example

2-D projections

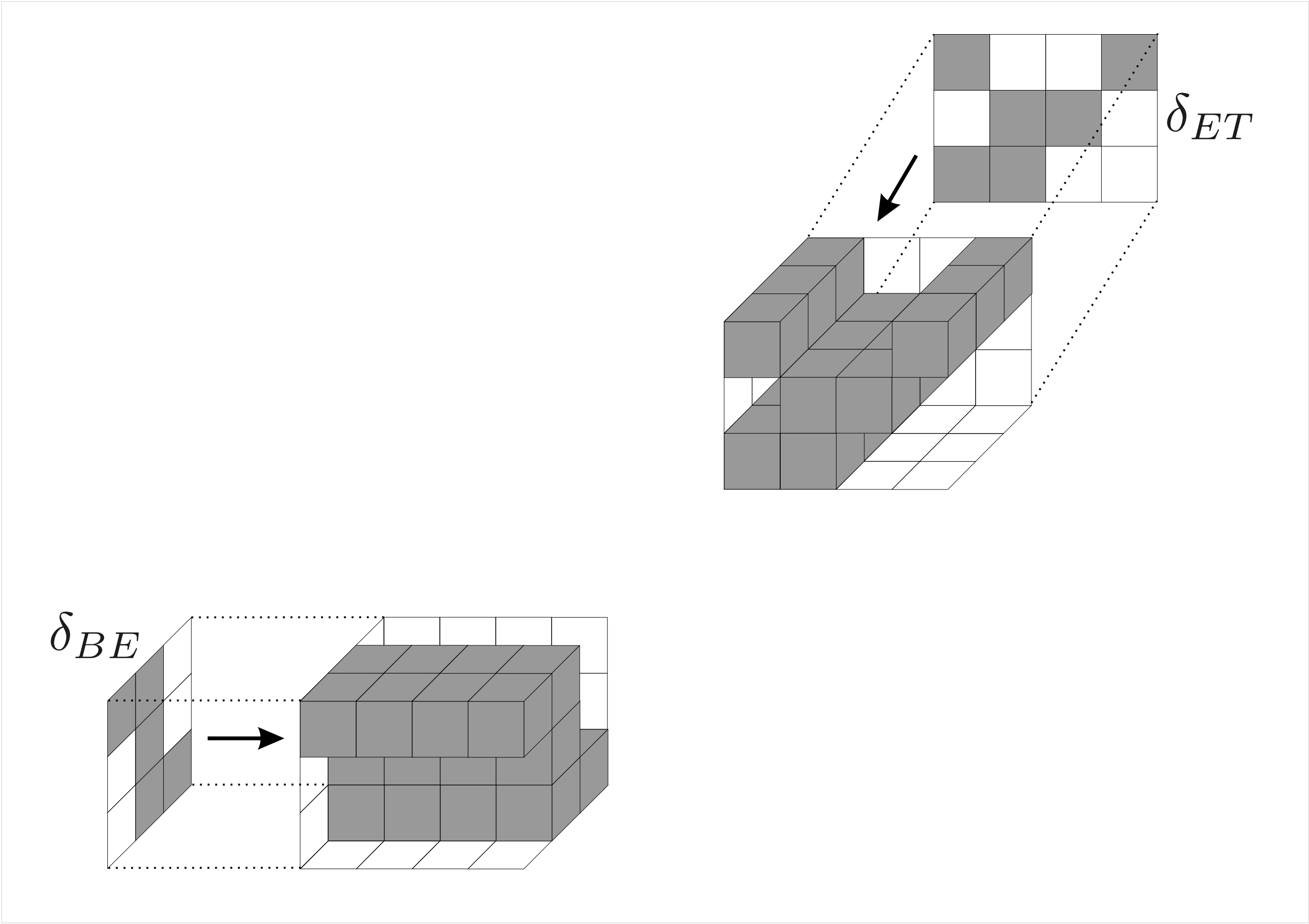


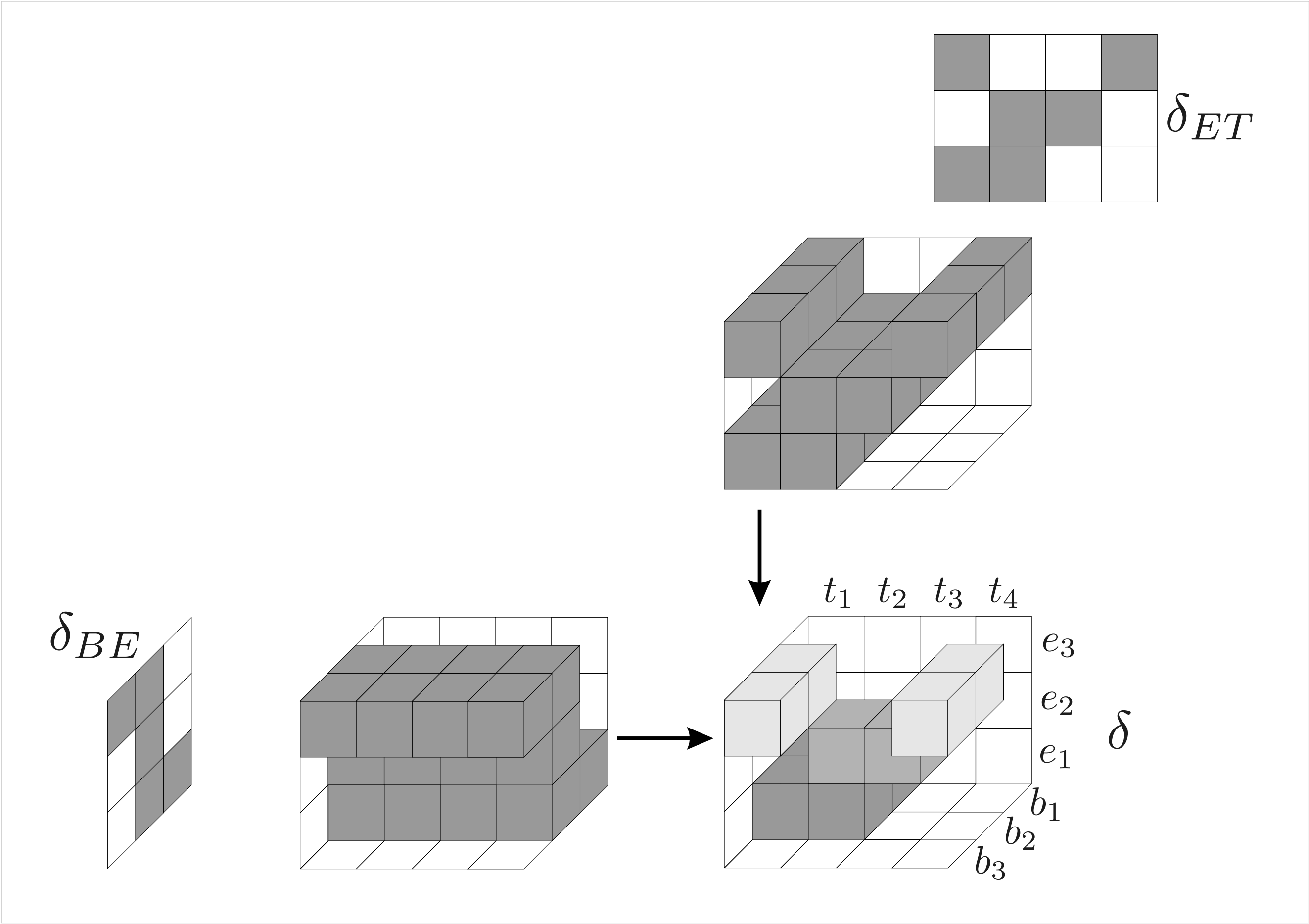
- Is it possible to reconstruct δ from the δ_i?

Example: Reconstruction of δ with δ_{BE} and δ_{ET}



Example: Reconstruction of δ with δ_{BE} and δ_{ET}





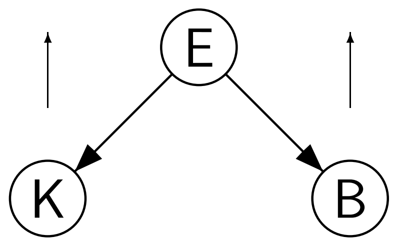
δ_{ET}
δ_{BE}
t_1 t_2 t_3 t_4
e_3
e_2
e_1
δ
b_1
b_2
b_3

Example — Qualitative Aspects

- Lecture theatre in winter: Waiting for Mr. K and Mr. B.
  Not clear whether there is ice on the roads.

- 3 variables:
  - E road condition: $\mathrm{dom}(\mathsf{E}) = \{\mathsf{ice}, \neg\mathsf{ice}\}$
  - K K had an accident: $\mathrm{dom}(\mathsf{K}) = \{\mathsf{yes}, \mathsf{no}\}$
  - B B had an accident: $\mathrm{dom}(\mathsf{B}) = \{\mathsf{yes}, \mathsf{no}\}$

- Ignorance about these states is modelled via the observer's belief.

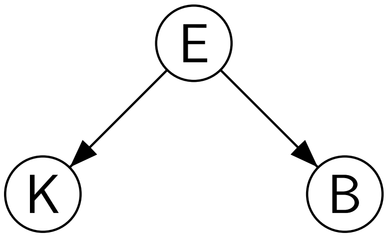


$\downarrow$ E influences K and B
(the more ice the more accidents)

$\uparrow$ Knowledge about accident increases belief in ice

Example

A priori knowledge	Evidence	Inferences
E unknown	B has accident	$\Rightarrow$ E $=$ ice more likely $\Rightarrow$ K has accident more likely
E $= \neg$ice	B has accident	$\Rightarrow$ no change in belief about E $\Rightarrow$ no change in belief about accident of K
E unknown E known		K and B dependent K and B independent



Causal Dependence vs. Reasoning

Rule: A entails B with certainty x: $\boxed{A \xrightarrow{x} B}$

- **Deduction** ($\rightarrow$):
  A and $A \xrightarrow{x} B$, therefore B more likely as effect (causality)

- **Abduction** ($\leftarrow$):
  B and $A \xrightarrow{x} B$, therefore A more likely as cause (no causality)

For this reason, the notion "dependency model" is to be preferred to "causal network".

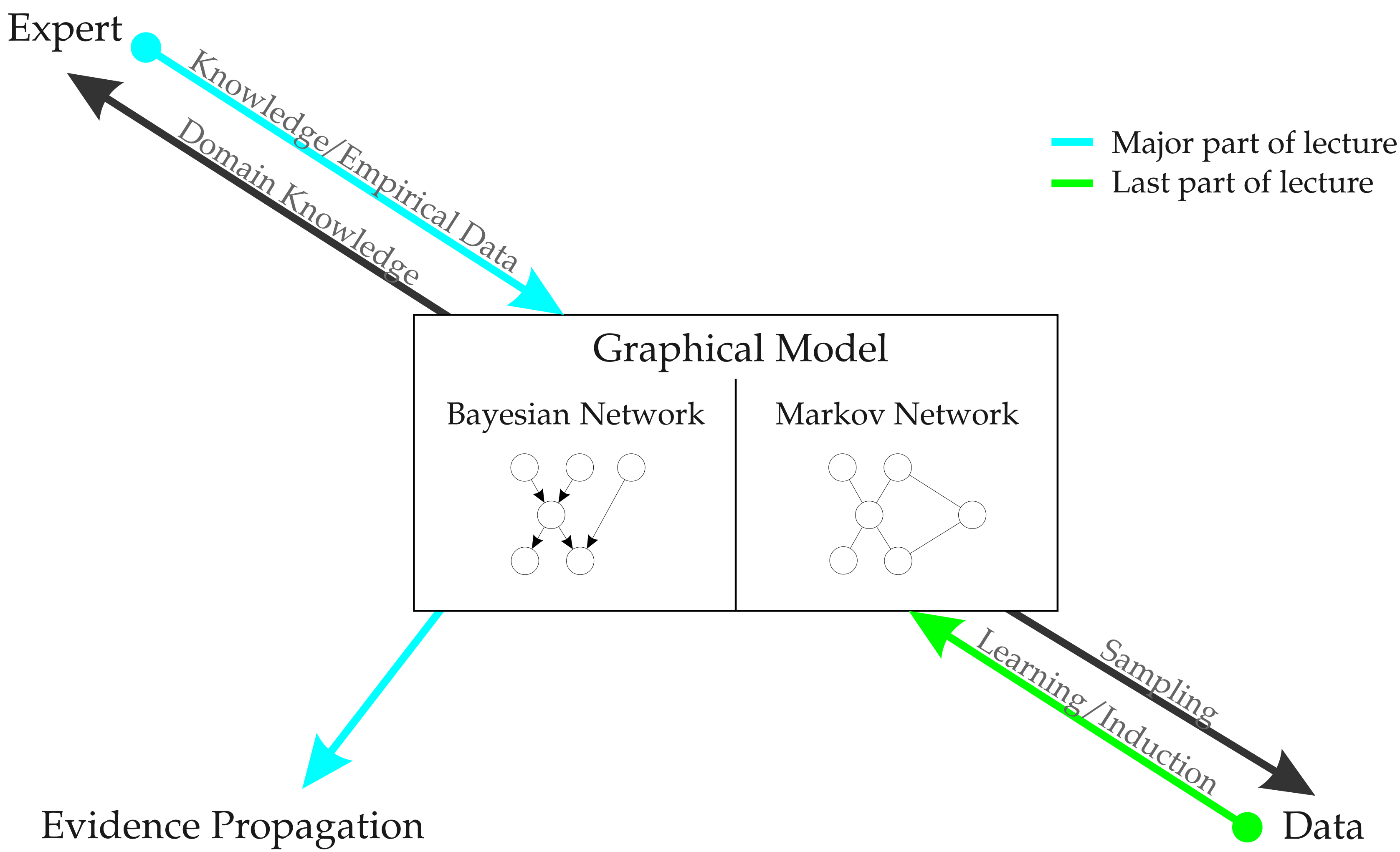
Objective

Is it possible to exploit local constraints (wherever they may come from — both structural and expert knowledge-based) in a way that allows for a decomposition of the large (intractable) distribution $P(X_1, \ldots, X_n)$ into several sub-structures $\{C_1, \ldots, C_m\}$ such that:

- The collective size of those sub-structures is much smaller than that of the original distribution P.
- The original distribution P is recomposable (with no or at least as few as possible errors) from these sub-structures in the following way:

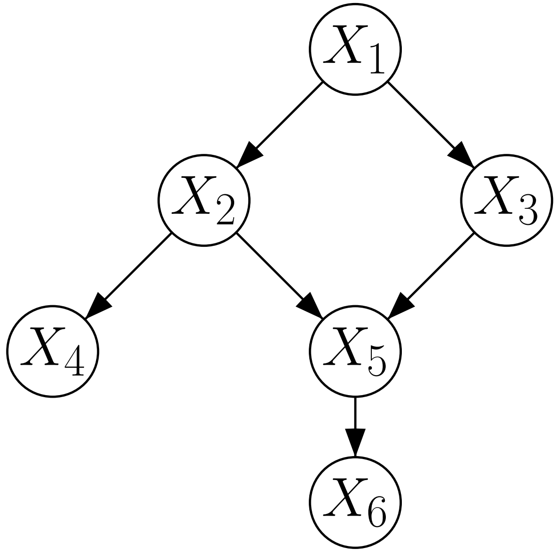
$$P(X_1, \ldots, X_n) = \prod_{i=1}^{m} \Psi_i(c_i)$$

where c_i is an instantiation of C_i and $\Psi_i(c_i) \in \mathbb{R}^+$ a *factor potential.*


Expert
Knowledge/Empirical Data
Domain Knowledge
Major part of lecture
Last part of lecture
Graphical Model
Bayesian Network
Markov Network
Sampling
Learning/Induction
Evidence Propagation
Data

Probabilistic Causal Networks

Probabilistic causal networks are directed acyclic graphs (DAGs) where the nodes represent propositions or variables and the directed edges model a direct causal dependence between the connected nodes. The strength of dependence is defined by conditional probabilities.

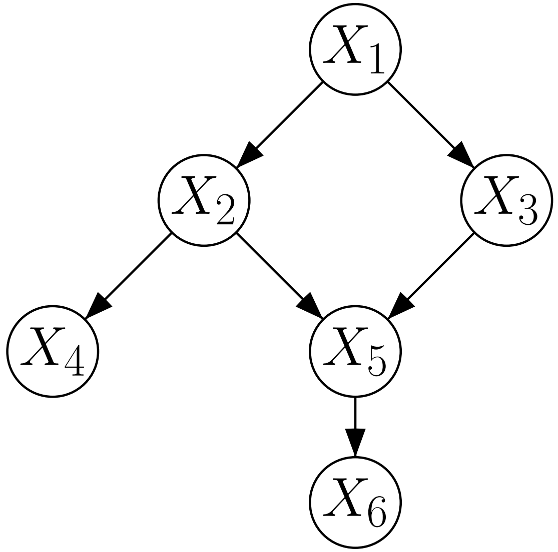


In general (according chain rule):

$$
\begin{aligned}
P(X_1, \ldots, X_6) = \; & P(X_6 \mid X_5, \ldots, X_1) \cdot \\
& P(X_5 \mid X_4, \ldots, X_1) \cdot \\
& P(X_4 \mid X_3, X_2, X_1) \cdot \\
& P(X_3 \mid X_2, X_1) \cdot \\
& P(X_2 \mid X_1) \cdot \\
& P(X_1)
\end{aligned}
$$

Probabilistic Causal Networks

Probabilistic causal networks are directed acyclic graphs (DAGs) where the nodes represent propositions or variables and the directed edges model a direct causal dependence between the connected nodes. The strength of dependence is defined by conditional probabilities.



According graph (independence structure):

$$
\begin{aligned}
P(X_1, \ldots, X_6) = \; & P(X_6 \mid X_5) \cdot \\
& P(X_5 \mid X_2, X_3) \cdot \\
& P(X_4 \mid X_2) \cdot \\
& P(X_3 \mid X_1) \cdot \\
& P(X_2 \mid X_1) \cdot \\
& P(X_1)
\end{aligned}
$$

Formal Framework

Nomenclature for the next slides:

- $X_1, \ldots, X_n$ — Variables
  (properties, attributes, random variables, propositions)

- $\Omega_1, \ldots, \Omega_n$ — respective finite domains
  (also designated with $\mathrm{dom}(X_i)$)

- $\Omega = \bigtimes_{i=1}^{n} \Omega_i$ — Universe of Discourse (tuples that characterize objects described by $X_1, \ldots, X_n$)

- $\Omega_i = \{x_i^{(1)}, \ldots, x_i^{(n_i)}\}$ — $n = 1, \ldots, n,\; n_i \in \mathbb{N}$

- Let Ω^* be the real universe of objects under consideration (e. g. population of people, collection of cars, customer transactions, etc.). Then the random vector $\vec{X} = (X_1, \ldots, X_n)$ *describes* each element $\omega^* \in \Omega^*$ in terms of the universe of discourse Ω:

$$\vec{X} : \Omega^* \to \Omega \quad \text{with} \quad \vec{X}(\omega^*) = (X_1(\omega^*), \ldots, X_n(\omega^*))$$

- If $(\Omega^*, \mathcal{E}, Q)$ is an intrinsic probability space acting in the background, then it induces — in combination with $\vec{X}$ — a probability measure P over Ω:

$$\begin{aligned}
&\forall (x_1, \ldots, x_n) \in \Omega : \\
&\qquad P(\{(x_1, \ldots, x_n)\}) = P(X_1 = x_1, \ldots, X_n = x_n) \\
&\qquad\qquad = Q(\{\omega^* \in \Omega^* \mid \bigwedge_{i=1}^{n} X_i = x_i\})
\end{aligned}$$

Formal Framework

- The product space $(\Omega, 2^\Omega, P)$ is unique iff $P(\{(x_1, \ldots, x_n)\})$ is specified for all $x_i \in \{x_i^{(1)}, \ldots, x_i^{(n_i)}\}$, $i = 1, \ldots, n$.

- When the distribution $P(X_1, \ldots, X_n)$ is given in tabular form, then $\prod_{i=1}^{n} |\Omega_i|$ entries are necessary.

- For variables with $|\Omega_i| \geq 2$ at least 2^n entries.

- The application of DAGs allows for the representation of existing (in)dependencies.

Constructing a DAG

input $P(X_1, \dots, X_n)$
output a unique DAG G

1: Set the nodes of G to $\{X_1, \dots, X_n\}$.

2: Choose a total ordering on the set of variables
(e. g. $X_1 \prec X_2 \prec \dots \prec X_n$)

3: For X_i find the smallest (uniquely determinable) set $S_i \subseteq \{X_1, \dots, X_n\}$ sucht that $P(X_i \mid S_i) = P(X_i \mid X_1 \dots, X_{i-1})$.

4: Connect all nodes in S_i with X_i and store $P(X_i \mid S_i)$ as quantization of the dependencies for that node X_i (given its parents).

5: **return** G

Belief Network

- A *Belief Network* (V, E, P) consists of a set $V = \{X_1, \ldots, X_n\}$ of random variables and a set E of directed edges between the variables.
- Each variable has a finite set of mutual exclusive and collectively exhaustive states.
- The variables in combination with the edges form a directed, acyclich graph.
- Each variable with parent nodes $B_1, \ldots, B_m$ is assigned a potential table $P(A \mid B_1, \ldots, B_m)$.
- Note, that the connections between the nodes not necessarily express a causal relationship.
- For every belief network, the following equation holds:

$$P(V) = \prod_{v \in V : P(c(v)) > 0} P(v \mid c(v))$$

with $c(v)$ being the parent nodes of v.

Example

- Let a_1, a_2, a_3 be three blood groups and b_1, b_2, b_3 three indications of a blood group test.

Variables:	A (blood group)	B (indication)
Domains:	$\Omega_A = \{a_1, a_2, a_3\}$	$\Omega_B = \{b_1, b_2, b_3\}$

- It is conjectured that there is a causal relationship between the variables.

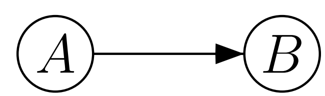
- A and B constitute random variables w. r. t. $(\Omega^*, \mathcal{E}, Q)$.

$$\Omega = \Omega_A \times \Omega_B \qquad A : \Omega^* \to \Omega_A, \quad B : \Omega^* \to \Omega_B$$

- A, B and $(\Omega^*, \mathcal{E}, Q)$ induce the probability space $(\Omega, 2^\Omega, P)$ with

$$P\Big(\{(a, b)\}\Big) = Q\Big(\{\omega^* \in \Omega^* \mid A(\omega^*) = a \wedge B(\omega^*) = b\}\Big) :$$

$P(\{(a_i, b_j)\})$	b_1	b_2	b_3	$\sum$
a_1	0.64	0.08	0.08	0.8
a_2	0.01	0.08	0.01	0.1
a_3	0.01	0.01	0.08	0.1
$\sum$	0.66	0.17	0.17	1



$P(A, B) = P(B \mid A) \cdot P(A)$

We are dealing with a belief network.

Example

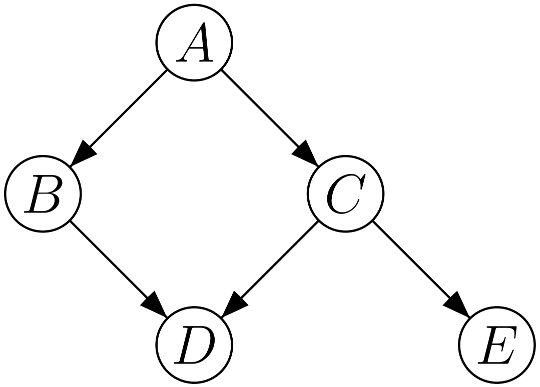
Choice of universe of discourse

	Variable	Domain
A	metastatic cancer	$\{a_1, a_2\}$
B	increased serum calcium	$\{b_1, b_2\}$
C	brain tumor	$\{c_1, c_2\}$
D	coma	$\{d_1, d_2\}$
E	headache	$\{e_1, e_2\}$

($\cdot_1$ — present, $\cdot_2$ — absent)

$$\Omega = \{a_1, a_2\} \times \cdots \times \{e_1, e_2\}$$
$$|\Omega| = 32$$

Analysis of dependencies



Example

Choice of probability parameters

$$
\begin{aligned}
P(a,b,c,d,e) &\overset{\text{abbr.}}{=} P(A=a,B=b,C=c,D=d,E=e) \\
&= P(e \mid c)P(d \mid b,c)P(c \mid a)P(b \mid a)P(a)
\end{aligned}
$$

Shorthand notation

- 11 values to store instead of 31
- Consult experts, textbooks, case studies, surveys, etc.

Calculation of conditional probabilities

Calculation of marginal probabilities

Crux of the Matter

- Knowledge acquisition (Where do the numbers come from?)
  → learning strategies

- Computational complexities
  → exploit independencies

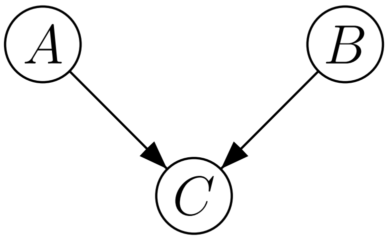
Problem:

- When does the independency of X and Y given Z hold in (V, E, P)?

- How can we determine $P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z)$ solely using the graph structure?

Dependencies

Converging Connection

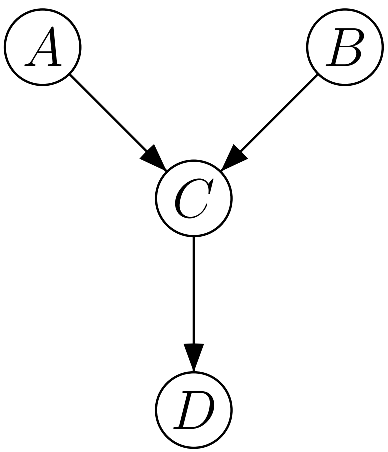


Meal quality	
A	quality of ingredients
B	cook's skill
C	meal quality

- If C is not instantiated (i. e., no value specified/observed), A and B are marginally independent.
- After instantiation (observation) of C the variables A and B become conditionally dependent given C.
- Evidence can only be transferred over a converging connection if the variable in between (or one of its successors) is initialized.

Dependencies

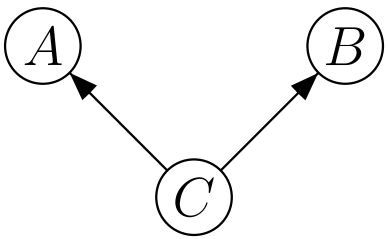
Converging Connection (cont.)



Meal quality	
A	quality of ingredients
B	cook's skill
C	meal quality
D	restaurant success

- If nothing is known about the restaurant success or meal quality or both, the cook's skills and quality of the ingredients are unrelated, thas is, *independent*.
- However, if we observe that the restaurant has no success, we can infer that the meal quality might be bad.
- If we further learn that the ingredients quality is high, we will conclude that the cook's skills must be low, thus rendering both variables *dependent*.

Diverging Connection

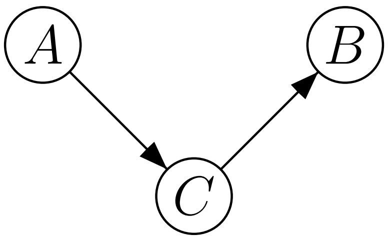


Diagnosis	
A	body temperature
B	cough
C	disease

- If C is unknown, knowledge about A ist relevant for B and vice versa, i. e. A and B are marginally dependent.
- However, if C is observed, A and B become conditionally independent given C.
- A influences B via C. If C is known it in a way blocks the information from flowing from A to B, thus rendering A and B (conditionally) independent.

Dependencies

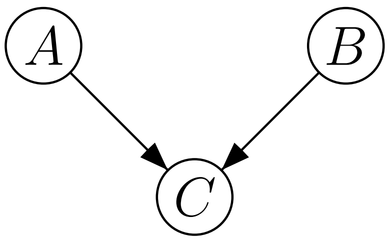
Serial Connection



Accidents	
A	rain
B	accident risk
C	road conditions

- Analog scenario to case 2
- A influences C and C influences B. Thus, A influences B.
  If C is known, it blocks the path between A and B.

Converging Connection: Marginal Independence



- Decomposition according to graph:

$$P(A, B, C) = P(C \mid A, B) \cdot P(A) \cdot P(B)$$

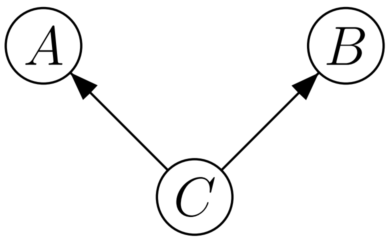
- Embedded Independence:

$$P(A, B, C) = \frac{P(A, B, C)}{P(A, B)} \cdot P(A) \cdot P(B) \qquad \text{with} \quad P(A, B) \neq 0$$

$$P(A, B) = P(A) \cdot P(B)$$

$$\Rightarrow \; A \perp\!\!\!\perp B \mid \emptyset$$

Diverging Connection: Conditional Independence



- Decomposition according to graph:

$$P(A, B, C) = P(A \mid C) \cdot P(B \mid C) \cdot P(C)$$

- Embedded Independence:

$$P(A, B \mid C) = P(A \mid C) \cdot P(B \mid C)$$

$$\Rightarrow \; A \perp\!\!\!\perp B \mid C$$

- Alternative derivation:

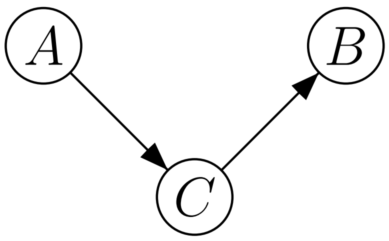
$$P(A, B, C) = P(A \mid C) \cdot P(B, C)$$

$$P(A \mid B, C) = P(A \mid C)$$

$$\Rightarrow \; A \perp\!\!\!\perp B \mid C$$

Formal Representation

Serial Connection: Conditional Independence



- Decomposition according to graph:

$$P(A, B, C) = P(B \mid C) \cdot P(C \mid A) \cdot P(A)$$

- Embedded Independence:

$$P(A, B, C) = P(B \mid C) \cdot P(C, A)$$

$$P(B \mid C, A) = P(B \mid C)$$

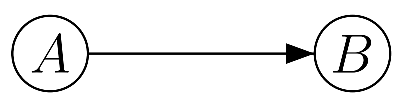
$$\Rightarrow A \perp\!\!\!\perp B \mid C$$

Trivial Cases:

- Marginal Independence:

$$P(A, B) \;=\; P(A) \cdot P(B)$$

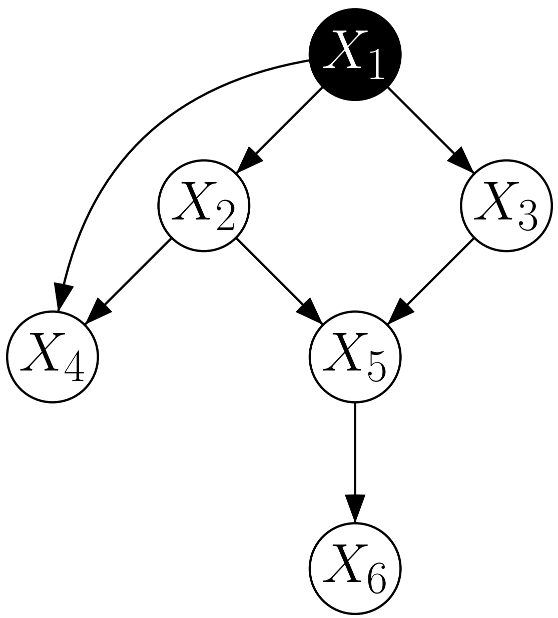
- Marginal Dependence:



$$P(A, B) \;=\; P(B \mid A) \cdot P(A)$$

Question: Are X_2 and X_3 independent given X_1?

evidence $X_1 = x_1$



d-Separation

Let $G = (V, E)$ a DAG and $X, Y, Z \in V$ three nodes.

a) A set $S \subseteq V \backslash \{X, Y\}$ *d-separates* X and Y, if S blocks all paths between X and Y. (paths may also route in opposite edge direction)

b) A path π is d-separated by S if at least one pair of consecutive edges along π is blocked. There are the following blocking conditions:

1.	$X \leftarrow Y \rightarrow Z$	tail-to-tail
2.	$X \leftarrow Y \leftarrow Z$ $X \rightarrow Y \rightarrow Z$	head-to-tail
3.	$X \rightarrow Y \leftarrow Z$	head-to-head

c) Two edges that meet tail-to-tail or head-to-tail in node Y are blocked if $Y \in S$.

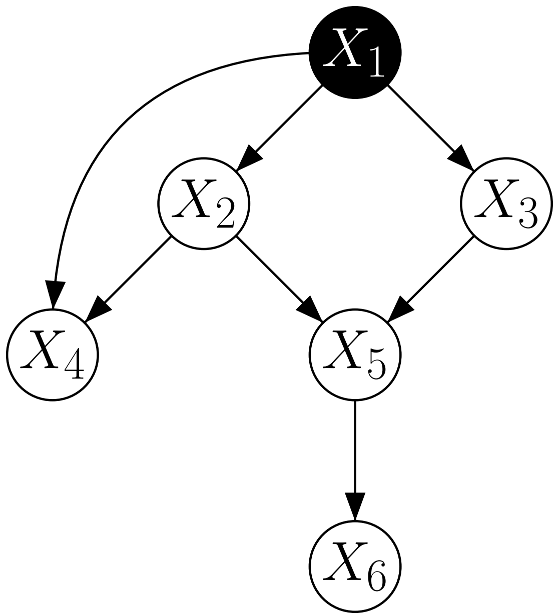
d) Two edges meeting head-to-head in Y are blocked if neither Y nor its successors are in S.

Relation to Conditional independence

If $S \subseteq V \setminus \{X, Y\}$ d-separates X and Y in a Belief network (V, E, P) then X and Y are conditionally independent given S:

$$P(X, Y \mid S) = P(X \mid S) \cdot P(Y \mid S)$$

Application to the previous example:



Paths: $\pi_1 = \langle X_2 - X_1 - X_3 \rangle$, $\pi_2 = \langle X_2 - X_5 - X_3 \rangle$
$\pi_3 = \langle X_2 - X_4 - X_1 - X_3 \rangle$, $S = \{X_1\}$

π_1 $X_2 \leftarrow X_1 \rightarrow X_3$ tail-to-tail
$X_1 \in S \;\Rightarrow\; \pi_1$ is blocked by S

π_2 $X_2 \rightarrow X_5 \leftarrow X_3$ head-to-head
$X_5, X_6 \notin S \;\Rightarrow\; \pi_2$ is blocked by S

π_3 $X_4 \leftarrow X_1 \rightarrow X_3$ tail-to-tail
$X_2 \rightarrow X_4 \leftarrow X_1$ head-to-head
both connections are blocked $\;\Rightarrow\; \pi_3$ is blocked

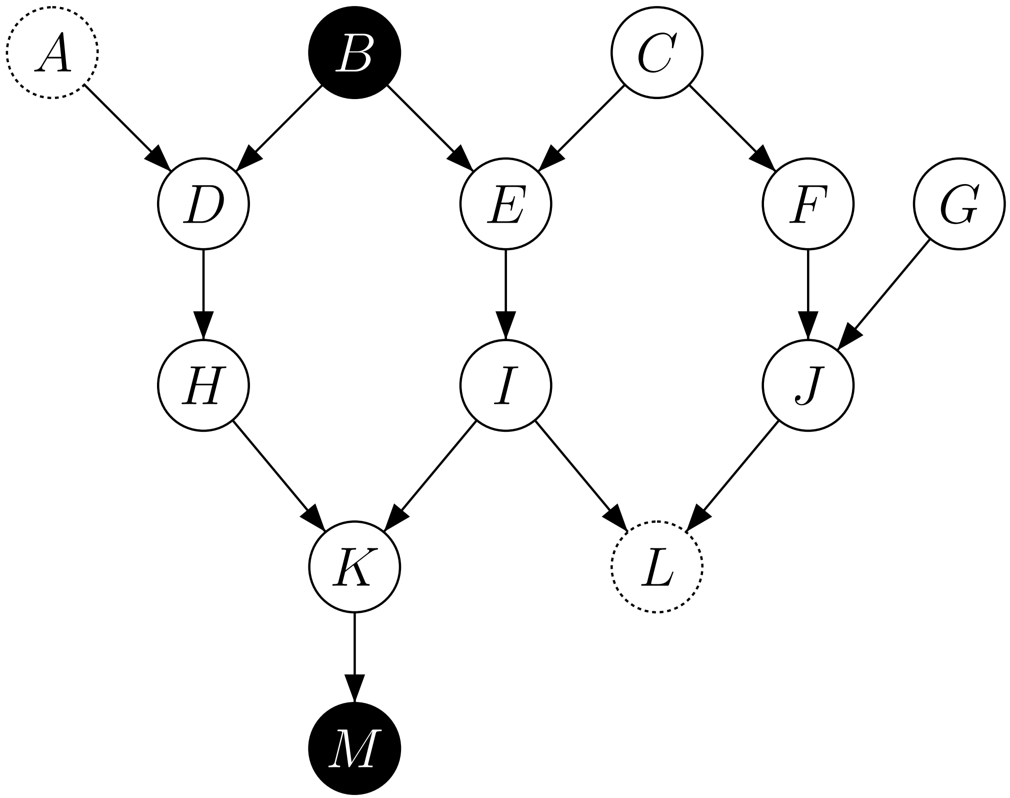
Example (cont.)

- Answer: X_2 and X_3 are d-separated via $\{X_1\}$. Therefore X_2 and X_3 become conditionally independent given X_1.

$$S = \{X_1, X_4\} \quad \Rightarrow \quad X_2 \text{ and } X_3 \text{ are d-separated by } S$$

$$S = \{X_1, X_6\} \quad \Rightarrow \quad X_2 \text{ and } X_3 \text{ are } \textit{not} \text{ d-separated by } S$$

Another Example



Are A and L conditionally independent given $\{B, M\}$?

Algebraic structure of CI statements

Question: Is it possible to use a formal scheme to infer new conditional independence (CI) statements from a set of initial CIs?

Repetition

Let $(\Omega, \mathcal{E}, P)$ be a probability space and W, X, Y, Z disjoint subsets of variables. If X and Y are conditionally independent given Z we write:

$$X \perp\!\!\!\perp_P Y \mid Z$$

Often, the following (equivalent) notation is used:

$$I_P(X \mid Z \mid Y) \quad \text{or} \quad I_P(X, Y \mid Z)$$

If the underlying space is known the index P is omitted.

(Semi-)Graphoid-Axioms

Let $(\Omega, \mathcal{E}, P)$ be a probability space and W, X, Y and Z four disjoint subsets of random variables (over Ω). Then the propositions

a) Symmetry: $(X \perp\!\!\!\perp_P Y \mid Z) \;\Rightarrow\; (Y \perp\!\!\!\perp_P X \mid Z)$

b) Decomposition: $(W \cup X \perp\!\!\!\perp_P Y \mid Z) \;\Rightarrow\; (W \perp\!\!\!\perp_P Y \mid Z) \wedge (X \perp\!\!\!\perp_P Y \mid Z)$

c) Weak Union: $(W \cup X \perp\!\!\!\perp_P Y \mid Z) \;\Rightarrow\; (X \perp\!\!\!\perp_P Y \mid Z \cup W)$

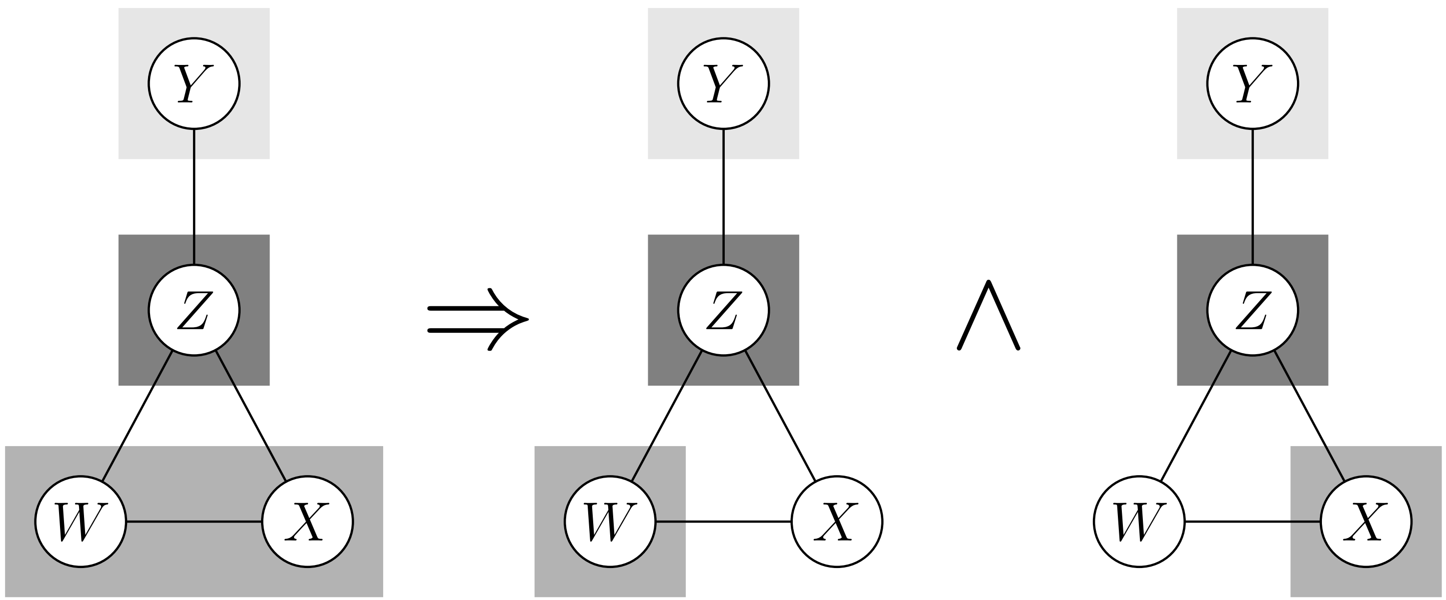
d) Contraction: $(X \perp\!\!\!\perp_P Y \mid Z \cup W) \wedge (W \perp\!\!\!\perp_P Y \mid Z) \;\Rightarrow\; (W \cup X \perp\!\!\!\perp_P Y \mid Z)$

are called the **Semi-Graphoid Axioms**. The above propositions and

e) Intersection: $(W \perp\!\!\!\perp_P Y \mid Z \cup X) \wedge (X \perp\!\!\!\perp_P Y \mid Z \cup W) \;\Rightarrow\; (W \cup X \perp\!\!\!\perp_P Y \mid Z)$

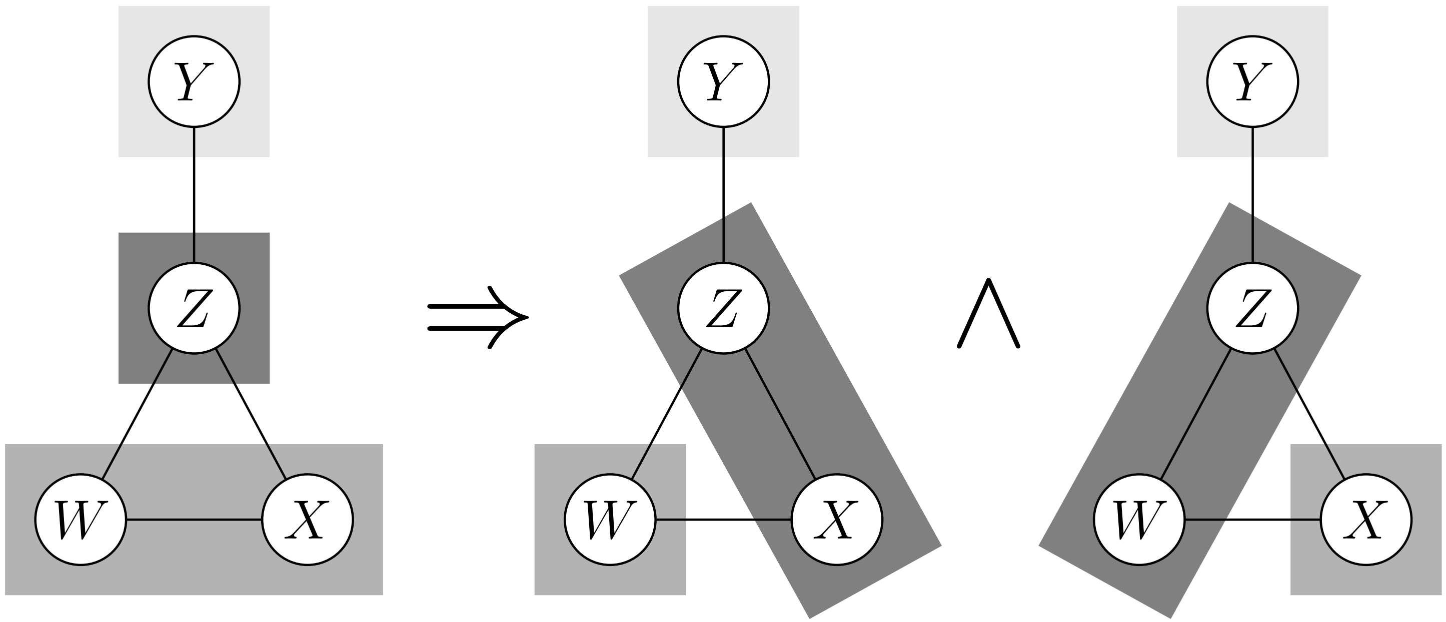
are called the **Graphoid Axioms**.

Decomposition



Drawings adapted from [Castillo *et al.* 1997].

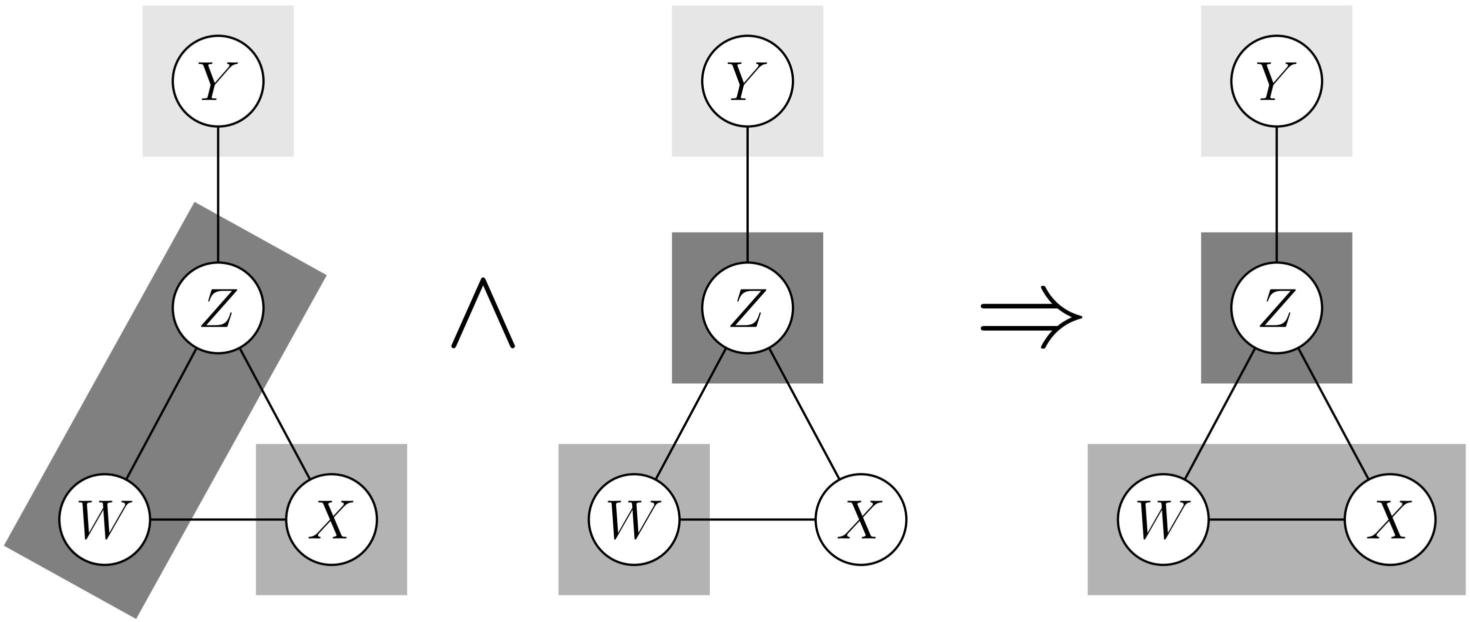
Weak Union



Learning irrelevant information W cannot render irrelevant information X relevant.

Drawings adapted from [Castillo *et al.* 1997].

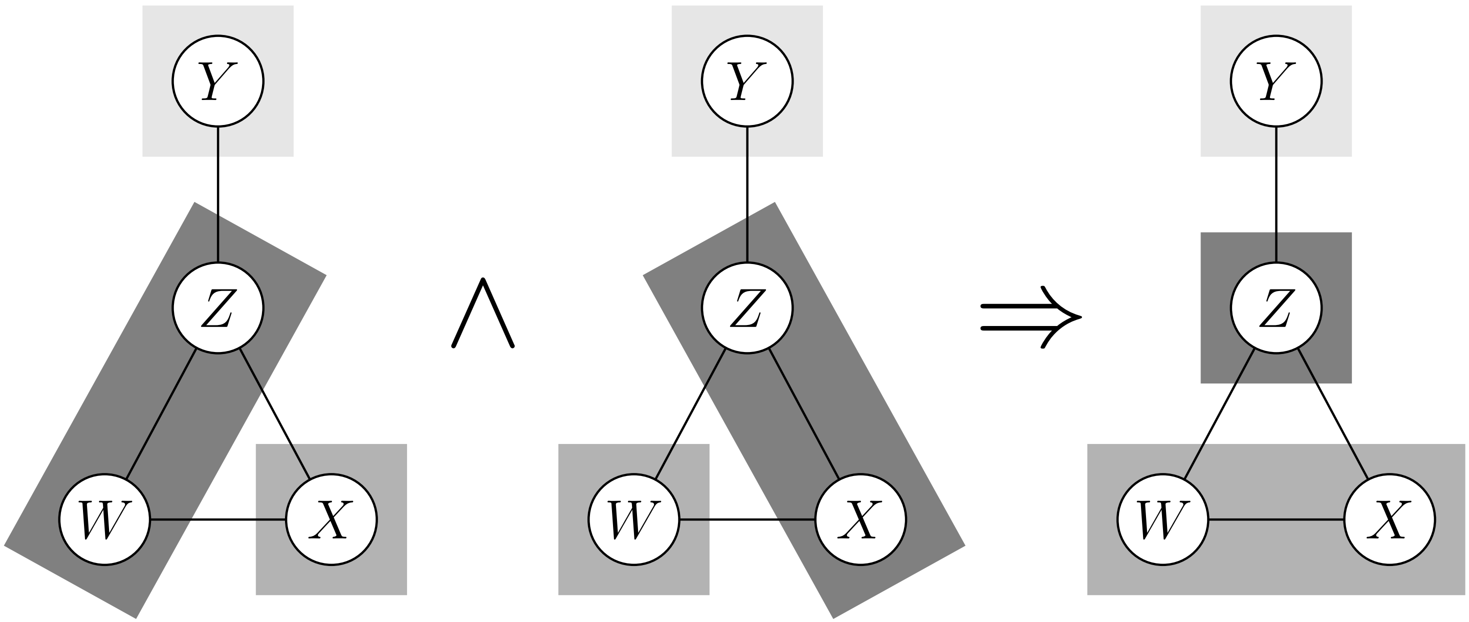
Contraction



If X is irrelevant (to Y) after having learnt some irrelevant information W, then X must have been irrelevant before.

Drawings adapted from [Castillo *et al.* 1997].

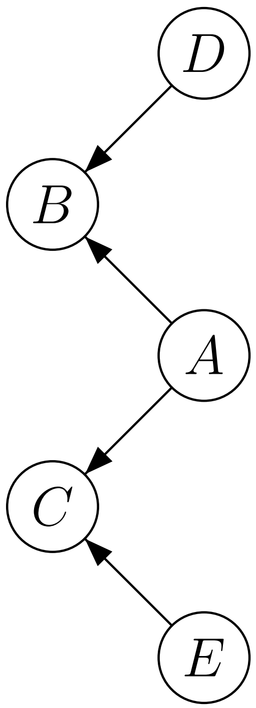
Intersection



Unless W affects Y when X is known or X affects Y when W is known, neither X nor W nor their combination can affect Y.

Drawings adapted from [Castillo *et al.* 1997].

Example



$$\text{Proposition:} \quad B \perp\!\!\!\perp C \mid A$$

$$\begin{aligned}
\text{Proof:} \quad & D \perp\!\!\!\perp A, C \mid \emptyset \quad \wedge \quad B \perp\!\!\!\perp C \mid A, D \\
\xRightarrow{\text{w. union}} \quad & D \perp\!\!\!\perp C \mid A \quad \wedge \quad B \perp\!\!\!\perp C \mid A, D \\
\xLeftrightarrow{\text{symm.}} \quad & C \perp\!\!\!\perp D \mid A \quad \wedge \quad C \perp\!\!\!\perp B \mid A, D \\
\xRightarrow{\text{contr.}} \quad & C \perp\!\!\!\perp B, D \mid A \\
\xRightarrow{\text{decomp.}} \quad & C \perp\!\!\!\perp B \mid A \\
\xLeftrightarrow{\text{symm.}} \quad & B \perp\!\!\!\perp C \mid A
\end{aligned}$$

Conditional (In)Dependence Graphs

Definition: Let $(\cdot \perp\!\!\!\perp_\delta \cdot \mid \cdot)$ be a three-place relation representing the set of conditional independence statements that hold in a given distribution δ over a set U of attributes. An undirected graph $G = (U, E)$ over U is called a **conditional dependence graph** or a **dependence map** w.r.t. δ, iff for all disjoint subsets $X, Y, Z \subseteq U$ of attributes

$$X \perp\!\!\!\perp_\delta Y \mid Z \;\Rightarrow\; \langle X \mid Z \mid Y \rangle_G,$$

i.e., if G captures by u-separation all (conditional) independences that hold in δ and thus represents only valid (conditional) dependences. Similarly, G is called a **conditional independence graph** or an **independence map** w.r.t. δ, iff for all disjoint subsets $X, Y, Z \subseteq U$ of attributes

$$\langle X \mid Z \mid Y \rangle_G \;\Rightarrow\; X \perp\!\!\!\perp_\delta Y \mid Z,$$

i.e., if G captures by u-separation only (conditional) independences that are valid in δ. G is said to be a **perfect map** of the conditional (in)dependences in δ, if it is both a dependence map and an independence map.

Definition: An undirected graph $G = (U, E)$ over a set U of attributes is said to have (w.r.t. a distribution δ) the

pairwise Markov property,

iff in δ any pair of attributes which are nonadjacent in the graph are conditionally independent given all remaining attributes, i.e., iff

$$\forall A, B \in U, A \neq B : \quad (A, B) \notin E \;\Rightarrow\; A \perp\!\!\!\perp_\delta B \mid U - \{A, B\},$$

local Markov property,

iff in δ any attribute is conditionally independent of all remaining attributes given its neighbors, i.e., iff

$$\forall A \in U : \quad A \perp\!\!\!\perp_\delta U - \text{closure}(A) \mid \text{boundary}(A),$$

global Markov property,

iff in δ any two sets of attributes which are u-separated by a third are conditionally independent given the attributes in the third set, i.e., iff

$$\forall X, Y, Z \subseteq U : \quad \langle X \mid Z \mid Y \rangle_G \;\Rightarrow\; X \perp\!\!\!\perp_\delta Y \mid Z.$$

Markov Properties of Directed Acyclic Graphs

Definition: A directed acyclic graph $\vec{G} = (U, \vec{E})$ over a set U of attributes is said to have (w.r.t. a distribution δ) the

pairwise Markov property,

iff in δ any attribute is conditionally independent of any non-descendant not among its parents given all remaining non-descendants, i.e., iff

$$\forall A, B \in U: \; B \in \text{non-descs}(A) - \text{parents}(A) \;\Rightarrow\; A \perp\!\!\!\perp_\delta B \mid \text{non-descs}(A) - \{B\},$$

local Markov property,

iff in δ any attribute is conditionally independent of all remaining non-descendants given its parents, i.e., iff

$$\forall A \in U: \quad A \perp\!\!\!\perp_\delta \text{non-descs}(A) - \text{parents}(A) \mid \text{parents}(A),$$

global Markov property,

iff in δ any two sets of attributes which are d-separated by a third are conditionally independent given the attributes in the third set, i.e., iff

$$\forall X, Y, Z \subseteq U: \quad \langle X \mid Z \mid Y \rangle_{\vec{G}} \;\Rightarrow\; X \perp\!\!\!\perp_\delta Y \mid Z.$$