Structural inference of independence relations

Peter Antal
antal@mit.bme.hu

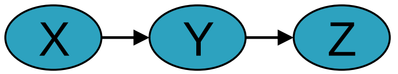
Outline

- Can we represent exactly (in)dependencies by a BN?
  - From a causal model? Suff.&nec.?

The independence model of a distribution

The independence map (model) M of a distribution P is the set of the valid independence triplets:

$M_P = \{I_{P,1}(X_1;Y_1|Z_1),\ldots, I_{P,K}(X_K;Y_K|Z_K)\}$



If P(X,Y,Z) is a Markov chain, then

$M_P=\{D(X;Y), D(Y;Z), I(X;Z|Y)\}$

Normally/almost always: D(X;Z)

Exceptionally: I(X;Z)

The semi-graphoid axioms

1. Symmetry: The observational probabilistic conditional independence is symmetric.

$$I_p(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{Z})\ iff\ I_p(\boldsymbol{Y};\boldsymbol{X}|\boldsymbol{Z})$$

2. Decomposition: Any part of an irrelevant information is irrelevant.

$$I_p(\boldsymbol{X};\boldsymbol{Y}\cup\boldsymbol{W}|\boldsymbol{Z}) \Rightarrow\ I_p(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{Z})\ and\ I_p(\boldsymbol{X};\boldsymbol{W}|\boldsymbol{Z})$$

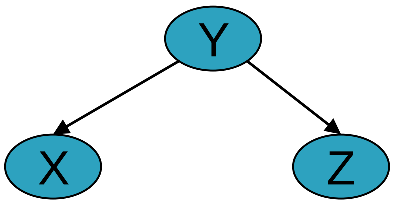
3. Weak union: Irrelevant information remains irrelevant after learning (other) irrelevant information.

$$I_p(\boldsymbol{X};\boldsymbol{Y}\cup\boldsymbol{W}|\boldsymbol{Z}) \Rightarrow\ I_p(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{Z}\cup\boldsymbol{W})$$

4. Contraction: Irrelevant information remains irrelevant after forgetting (other) irrelevant information.

$$I_p(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{Z})\ and\ I_p(\boldsymbol{X};\boldsymbol{W}|\boldsymbol{Z}\cup\boldsymbol{Y}) \Rightarrow\ I_p(\boldsymbol{X};\boldsymbol{Y}\cup\boldsymbol{W}|\boldsymbol{Z})$$

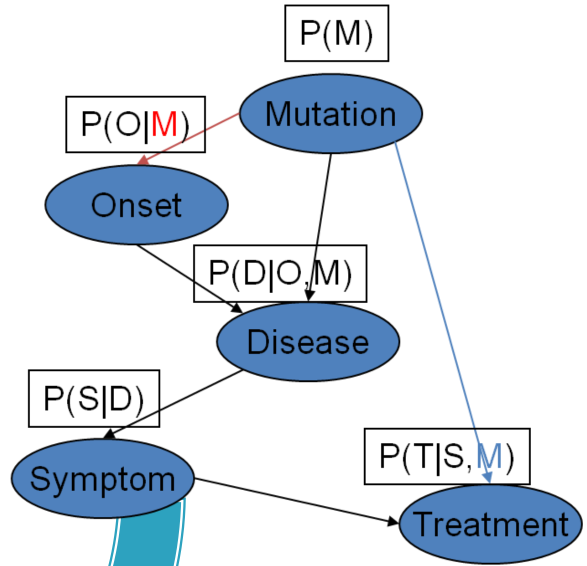
The independence map of a N-BN



If $P(Y,X,Z)$ is a naive Bayesian network, then
$M_P=\{D(X;Y), D(Y;Z), I(X;Z|Y)\}$
Normally/almost always: $D(X;Z)$
Exceptionally: $I(X;Z)$

Bayesian networks

- Directed acyclic graph (DAG)
  - nodes – random variables/domain entities
  - edges – direct probabilistic dependencies
    (edges– causal relations
- Local models – $P(X_i|Pa(X_i))$
- Three interpretations:



1. Causal model

3. Concise representation of joint distributions

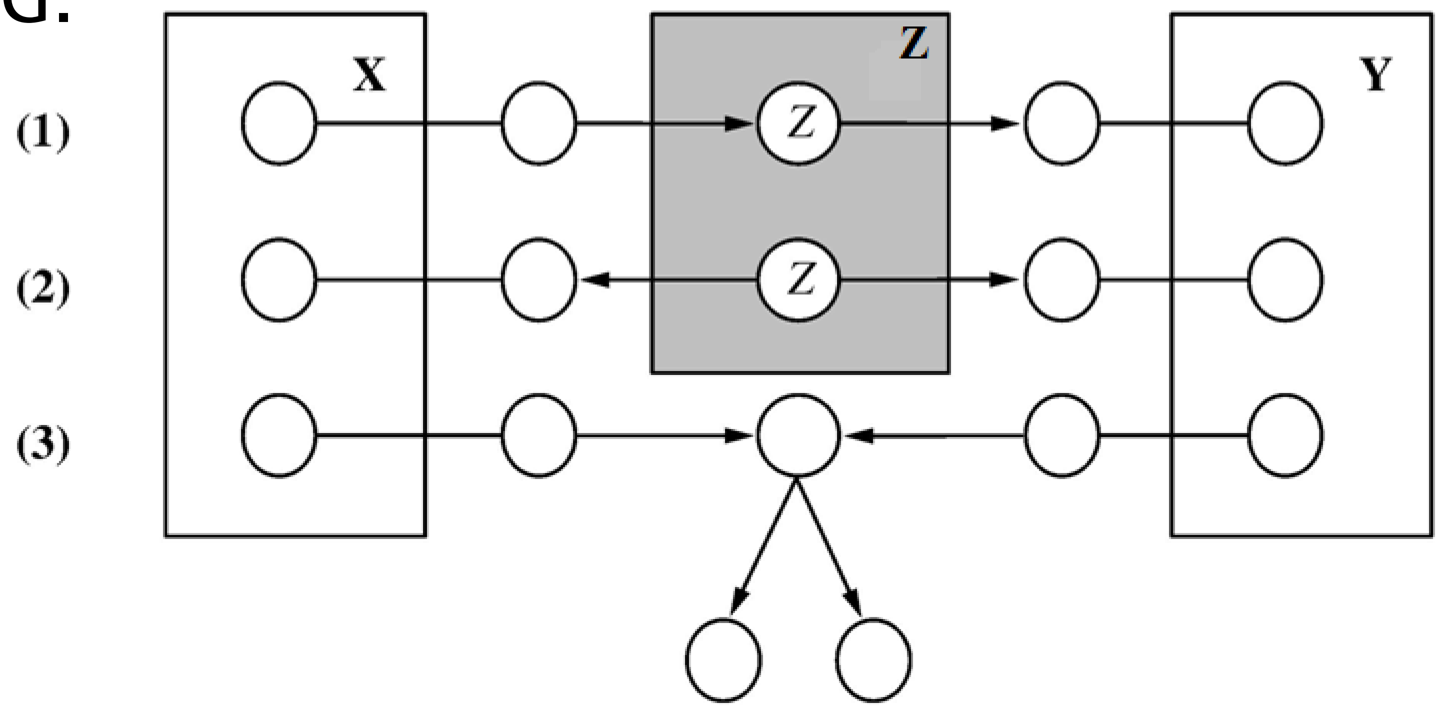
$$P(M,O,D,S,T) = P(M)P(O|M)P(D|O,M)P(S|D)P(T|S,M)$$

$$M_P=\{I_{P,1}(X_1;Y_1|Z_1),\ldots\}$$

2. Graphical representation of (in)dependencies

Inferring independencies from structure: d-separation

$I_G(X;Y|Z)$ denotes that X is d-separated (directed separated) from Y by Z in directed graph G.



d-separation and the global Markov condition

Definition 7 *A distribution $P(X_1, \ldots, X_n)$ obeys the global Markov condition w.r.t. DAG G, if*

$$\forall\, X, Y, Z \subseteq U \; (X \perp\!\!\!\perp Y | Z)_G \Rightarrow (X \perp\!\!\!\perp Y | Z)_P, \tag{9}$$

where $(X \perp\!\!\!\perp Y|Z)_G$ denotes that X and Y are d-separated by Z, that is if every path p between a node in X and a node in Y is blocked by Z as follows

1. *either path p contains a node n in Z with non-converging arrows (i.e. $\rightarrow n \rightarrow$ or $\leftarrow n \rightarrow$),*
2. *or path p contains a node n not in Z with converging arrows (i.e. $\rightarrow n \leftarrow$) and none of its descendants of n is in Z.*

Bayesian network definitions

Theorem 1 *Let $P(U)$ a probability distribution and* **G** *a DAG, then the conditions above (repeated below) are equivalent:*

- *F* P *is Markov relative* G *or* P *factorizes w.r.t* G,
- *O* P *obeys the ordered Markov condition w.r.t.* G,
- *L* P *obeys the local Markov condition w.r.t.* G,
- *G* P *obeys the global Markov condition w.r.t.* G.

Definition 8 *A directed acyclic graph* **(DAG)** G *is a* Bayesian network *of distribution* $P(U)$ *iff the variables are represented with nodes in* G *and* (G, P) *satisfies any of the conditions* F, O, L, G *such that* G *is minimal (i.e. no edge(s) can be omitted without violating a condition* F, O, L, G*).*

Representation of independencies

D-separation provides a sound and complete, computationally efficient algorithm to read off an (in)dependency model consisting the independencies that are valid in all distributions Markov relative to G, that is $\forall\ X, Y, Z \subseteq V$

$$(X \perp\!\!\!\perp Y|Z)_G \Leftrightarrow ((X \perp\!\!\!\perp Y|Z)_P \text{ in all P Markov relative to G}). \qquad (10)$$

For certain distributions exact representation is not possible by Bayesian networks, e.g.:

1. Intransitive Markov chain: X➔Y➔Z
2. Pure multivariate cause: {X,Z}➔Y
3. Diamond structure:

P(X,Y,Z,V) with M_P={D(X;Z), D(X;Y), D(V;X), D(V;Z), I(V;Y|{X,Z}), I(X;Z|{V,Y}).. }.