Alexander von Humboldt Stiftung/Foundation

Zukunfts-investitions-programm zip

Probabilistic Modeling in Philosophy

Stephan Hartmann

with Luc Bovens

LSE - London

PPM Group - Konstanz

Motivation

- Models are very popular in the sciences.
- Aims of this tutorial:

  1. Demonstrate that modeling techniques can also be used to tackle problems of philosophical interest.

  2. Show how to do it!
- The method: Probabilistic modeling (esp. Bayesian Networks)
- The problems: Epistemology and philosophy of science

Modeling in Science and Philosophy

- Shift from theories to models in science
- Features of theories

  - Examples: Newton‘s Mechanics, Quantum Mechanics
  - General and unversal in scope
  - Abstract
  - No idealizations involved (ideally...)
- Features of models

  - Examples: the Bohr model of the atom, gas models,...
  - Specific and limited in scope
  - Concrete
  - Involve idealized assumptions

Modeling in Science (cont’d)

- Why this shift?

  - distrust general theories

  - one often needs a model to apply a theory

  - models can be solved

  - models are more intuitive

- An important distinction

  - models of theories: e.g. pendulum (model within theoretical framework)

  - phenomenological models: e.g. the liquid drop model of the atomic nucleus (no theoretical framework)

Modeling in Philosophy

- Glymour‘s distintction

  - **Platonic tradition** in philosophy: formulate general and universal theories which aim at explaining or accounting for everything... and end up explaining nothing!

  - **Euclidian tradition** in philosophy: make idealized assumptions and explore the consequences of these assumptions... Just like scientists do it!

- Should we go for the Euclidian program in philosophy?

  - perhaps not always, but sometimes

  - pluralism of methods

Probabilistic Modeling

- There is a long tradition of logic modeling in philosophy.
- <u>But:</u> There are alternative methods that turn out to be better suited for many philosophical problems, most importantly probabilistic methods.
- We adopt Bayesianism as a framework and construct models within this framework (cf. "models of a theory").
- While traditional textbook Bayesianism (a là Howson and Urbach) has the problem of being often "too far away" from real science, our modeling approach has the advantage of being able to *bridge the gap* between a general theory (i.e. Bayesianism) and the scientific practice.

Overview

I. Lecture 1: Bayesian Networks

1. Probability Theory
2. Bayesian Networks
3. Modeling Partially Reliable Information Sources

II. Lecture 2: Applications in Epistemology

1. Is Coherence Truth-Conducive?
2. How Can one Measure the Coherence of an Information Set?
3. Open Problems

III. Lecture 3: Applications in Philosophy of Science

1. Does the Variety-of-Evidence Thesis Hold?
2. What Is a Scientific Theory?
3. Open Problems

Lecture 1

Bayesian Networks

1. Probability Theory

- Let S be a collection of sentences, and P is a probability function. It satisfies the *Kolmogorov axioms*:

  1. $P(A) \geq 0$
  2. $P(A) = 1$ if A true in all models
  3. $P(A \vee B) = P(A) + P(B)$ if A, B mutually exclusive

- Some consequences:

  1. $P(\neg A) = 1 - P(A)$
  2. $P(A) = P(B)$ if (in all models) $A \Leftrightarrow B$
  3. $P(A \vee B) = P(A) + P(B) - P(A, B)$; note: $P(A, B) := P(A \wedge B)$

Conditional Probabilities

- Definition:

  $P(A|B) = P(A, B)/P(B) \quad \text{if} \quad P(B) \neq 0$

- Bayes' Theorem:

  $P(B|A) = P(A|B)\, P(B)/P(A)$

  $= P(A|B)\, P(B)/[P(A|B)\, P(B) + P(A|\neg B)\, P(\neg B)]$

  $= P(B)/[P(B) + P(\neg B)\, x]$

with the *likelihood ratio*

  $x := P(A|\neg B)/P(A|B)$

Conditional Independence

- A and B are (unconditionally) independent iff
  $P(A, B) = P(A)\,P(B) \Leftrightarrow P(A|B) = P(A) \Leftrightarrow P(B|A) = P(B)$
- Example: throw a dice; A = observe even number, B = observe number ≤ 4.
- A is conditionally independent of B given C iff
  $P(A|B, C) = P(A|C)$
- Example: A=yellow fingers, B=lung cancer, C=smoking
- In symbols: $A \perp\!\!\!\perp B | C$
- Note: The relation $A \perp\!\!\!\perp B | C$ is symmetrical:

$$A \perp\!\!\!\perp B | C \Leftrightarrow B \perp\!\!\!\perp A | C$$

The Semi-Graphoid Axioms

- The conditional independence relation satisfies the semi-graphoid axioms:

  1. Symmetry: $X \perp\!\!\!\perp Y | Z \leftrightarrow Y \perp\!\!\!\perp X | Z$

  2. Decomposition: $X \perp\!\!\!\perp Y, W | Z \rightarrow X \perp\!\!\!\perp Y | Z$

  3. Weak Union: $X \perp\!\!\!\perp Y, W | Z \rightarrow X \perp\!\!\!\perp Y, Z | W$

  4. Contraction: $X \perp\!\!\!\perp W | Y, Z \ \& \ X \perp\!\!\!\perp Y | Z \rightarrow X \perp\!\!\!\perp Y, W | Z$

- With these axioms, new conditional independencies can be obtained from known independencies.

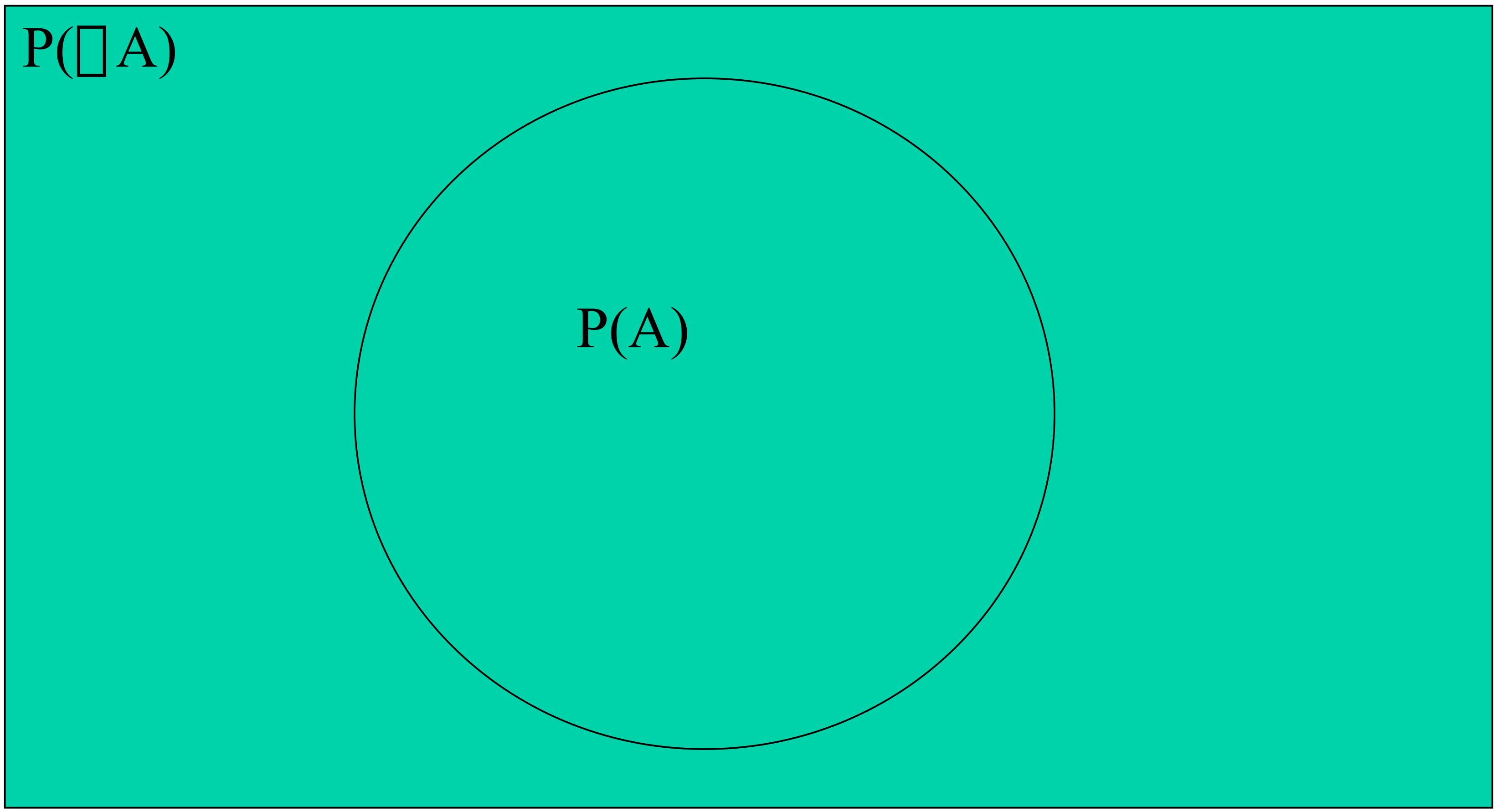
Joint and Marginal Probability

- To specify the *joint probability distribution* over two *binary propositional variables A, B*, three probabilities have to be specified (if we do not have any further knowledge):

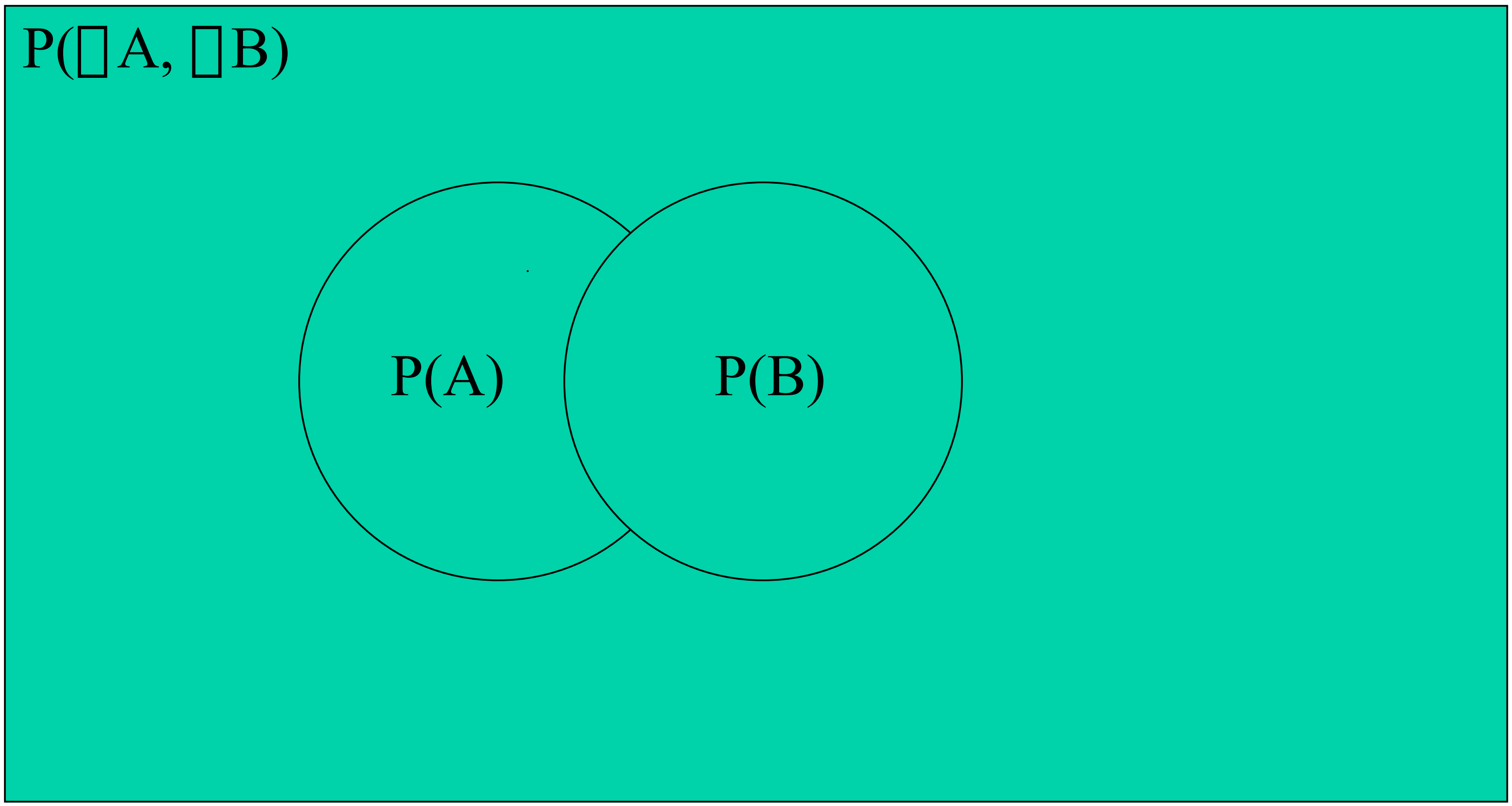
  E.g. $P(A, B) = .2$, $P(A, \neg B) = .1$, $P(\neg A, B) = .6$
- Note: $\sum_{A,\,B} P(A, B) = 1 \Rightarrow P(\neg A, \neg B) = .1$
- *Marginal probability*: $P(A) = \sum_{B} P(A, B)$
- This allows us to calculate whatever marginal or conditional probability we are interested in from the joint probability:
- $P(A_1,\ldots A_m | A_{m+1},\ldots A_n) = P(A_1,\ldots A_n)/P(A_{m+1},\ldots A_n)$

  $= P(A_1,\ldots A_n)/\sum_{A_{m+1},\, A_n} P(A_1,\ldots A_n)$

Representing a Joint Probability Distribution



Representing… (cont’d)



2. Bayesian Networks

- Venn diagrams and the specification of all entries in $P(A_1, \ldots, A_n)$ are not the most efficient ways to represent a joint probability distribution.
- There is also a problem of computational complexity: Specifying the joint probability distribution over n variables requires the specification of 2^n-1 numbers.
- <u>The trick:</u> Use information about conditional independencies that hold between (sets of) variables. This will reduce the number of numbers that have to be specified (and stored in a computer).

An Example from Medicine

- T: Patient has tuberculosis
- X: Positive X-ray
- Given information:

  $t := P(T) = .01$

  $p := P(X|T) = .95 = 1 - P(\neg X|T) = 1 -$ rate of *false negatives*

  $q := P(X|\neg T) = .02 =$ rate of *false positives*

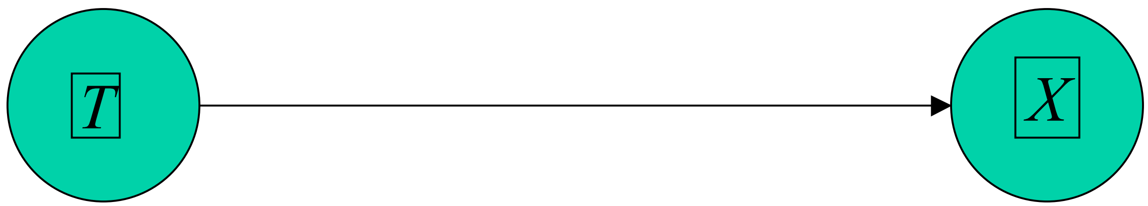
- What is $P(T|X)$? $\Rightarrow$ Apply Bayes' Theorem

  $P(T|X) = P(X|T)\, P(T) / [P(X|T)\, P(T) + P(X|\neg T)\, P(\neg T)] =$

  $= p\, t / [p\, t + q\, (1-t)] = t/[t + (1-t)\, x]$ with $x := q/p$

  $= .32$

A Bayesian Network Representation



$P(T) = .1$

$P(X|T) = .95$

$P(X|\neg T) = .2$

Parlance:

"T causes *X"*

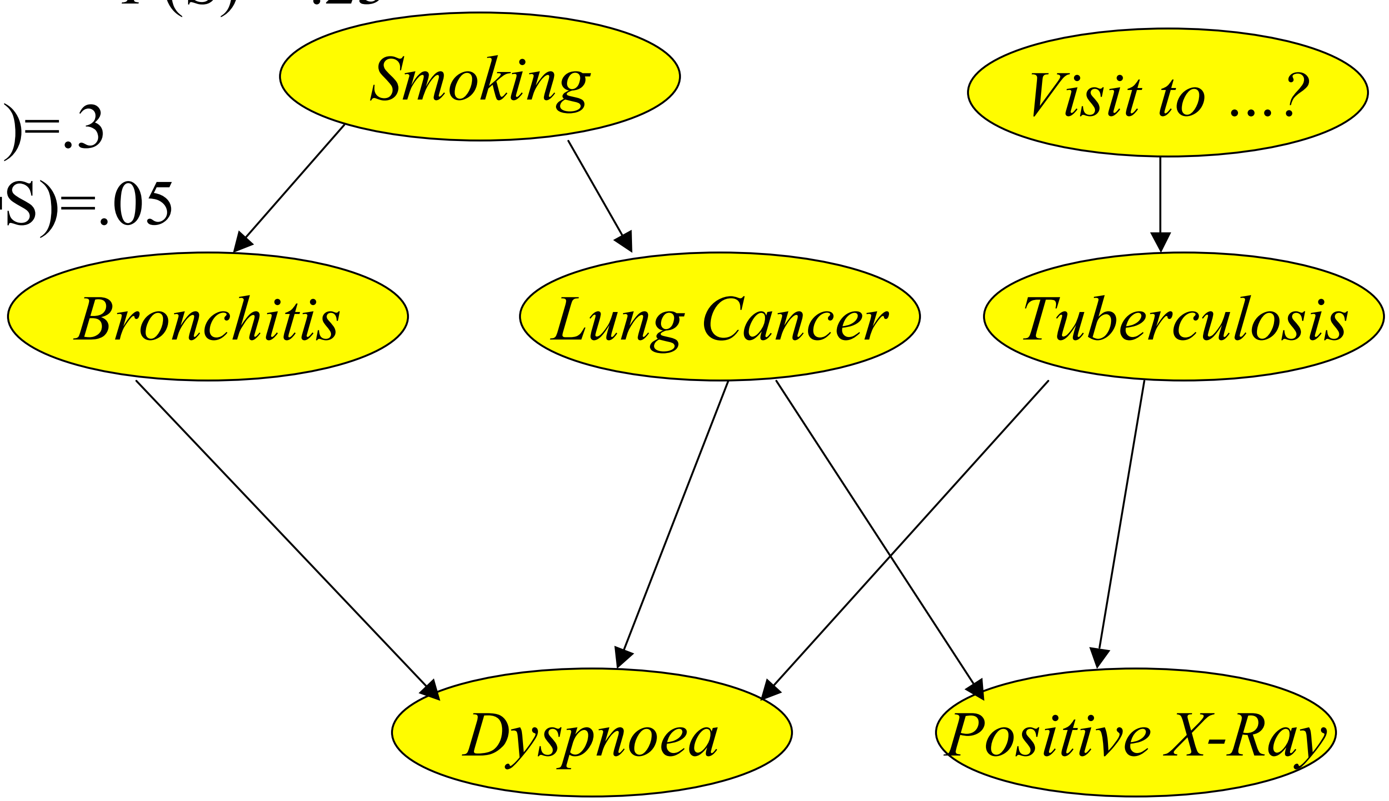
"T directly influences *X"*

A More Complicated (=Realistic) Scenario

$P(S) = .23$

$P(B|S)=.3$

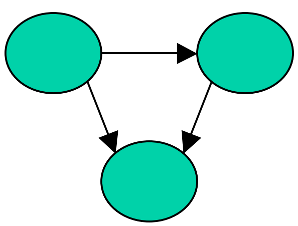
$P(B|\neg S)=.05$

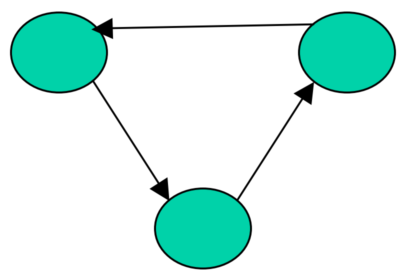


Directed Acyclic Graphs

- A directed graph G (V, E) consists of a finite set of nodes V and an irreflexive binary relation E on V.

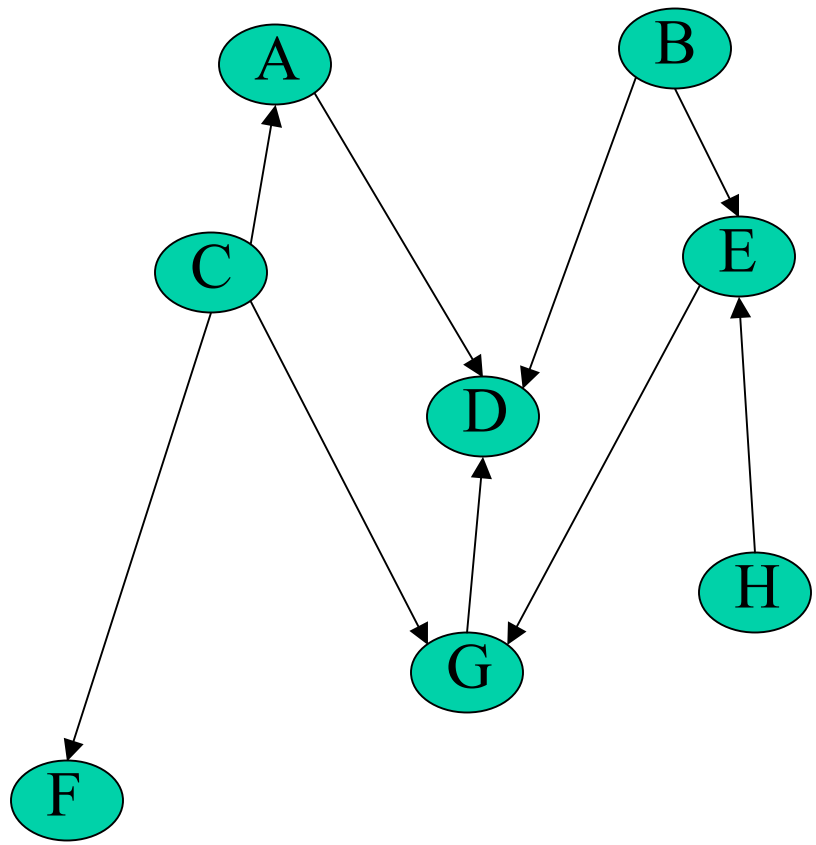
- A directed acyclic graph (DAG) is a directed graph which does not contain cycles.





Some Vocabulary

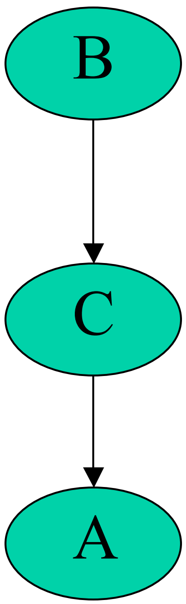
- Parents of node A: par(A)
- Ancestor
- Child node
- Descendents
- Non-Descendents
- Root node



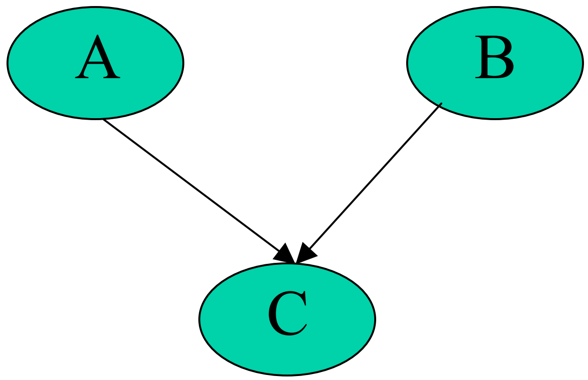
The Parental Markov Condition

- The Parental Markov Condition (PMC) requires that a variable is conditionally independent of its non-descendents given its parents.
- <u>Definition:</u> A Bayesian Network is a DAG with a probability distribution which respects the PMC.

Three Examples

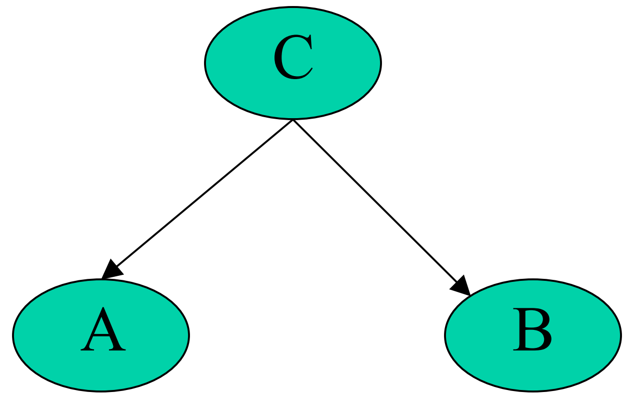


$$A \perp\!\!\!\perp B \,|\, C$$



$$A \perp\!\!\!\perp B$$

“collider”



$$A \perp\!\!\!\perp B \,|\, C$$

“common cause”

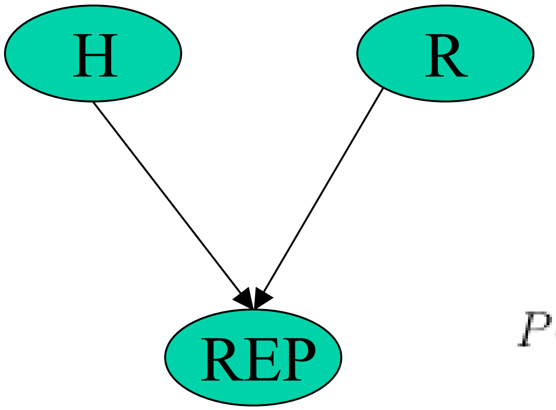
Bayesian Networks at Work

- How can one calculate probabilities with B.N.?
- Here is the answer:

$$P(A_1, \ldots, A_n) = \prod_i P(A_i | \mathrm{par}(A_i))$$

- I.e. the joint probability distribution is determined by the prior probability of the root nodes ($\mathrm{par}(A) = \varnothing$) and the conditional probabilities of all other nodes.
- This typically requires the specification of much less numbers than 2^n-1.

An Example



P(*H, R, REP*) = P(*H*) P(*R*)P(*REP*|*H, R*)

What is P(H|REP)?

$$\begin{aligned}
P(\mathrm{H}|\mathrm{REP}) &= \frac{P(\mathrm{H},\mathrm{REP})}{P(\mathrm{REP})} \\
&= \frac{\Sigma_R\, P(\mathrm{H}, R, \mathrm{REP})}{\Sigma_{H,R}\, P(H, R, \mathrm{REP})} \\
&= \frac{P(\mathrm{H})\, \Sigma_R\, P(R) P(\mathrm{REP}|\mathrm{H}, R)}{\Sigma_{H,R}\, P(H, R, \mathrm{REP})} \\
&= \frac{h(r \cdot 1 + (1-r) \cdot a)}{h(r \cdot 1 + (1-r)\cdot a) + (1-h)(r \cdot 0 + (1-r) \cdot a)} \\
&= \frac{h(r + a(1-r))}{hr + a(1-r)}
\end{aligned}$$

P(H) = *h*

P(R) = *r*

P(REP|H, R) = 1

P(REP|□H, R) = 0

P(REP|H, □R) = *a*

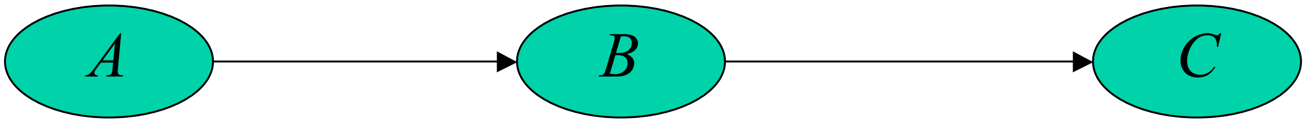
P(REP|□H, □R) = *a*

Some More Theory: *d*-Separation

- There are more independencies in a Bayesian Network than the ones accounted for by the Parental Markov Condition.
- Is there a systematic way to find all independences that hold in a given Bayesian Network?
- Yes! *d*-separation
- Let *A, B*, and *C* be sets of variables. Then:

  $A \perp\!\!\!\perp B | C$ iff *C* *d*-separates *A* from *B*.
- So what is *d*-separation?

Example 1



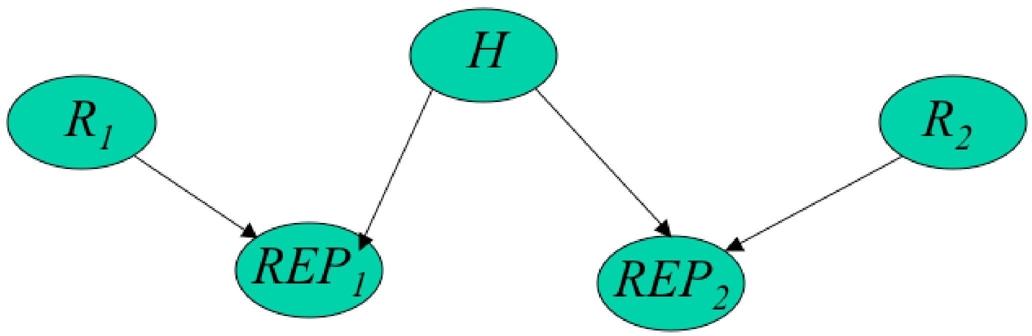
PMC $\Rightarrow$ $C \perp\!\!\!\perp A \mid B$

But it is also the case that $A \perp\!\!\!\perp C \mid B$

This does *not* follow from PMC.

It can, however, be derived from the symmetry axiom for semi-graphoids.

Example 2



PMC $\Rightarrow$ $REP_1 \perp\!\!\!\perp REP_2 | H, R_1$ (*)

But not $REP_1 \perp\!\!\!\perp REP_2 | H$

However: PMC $\Rightarrow$ $R_1 \perp\!\!\!\perp H, REP_2$

Weak Union $\Rightarrow$ $R_1 \perp\!\!\!\perp REP_2 | H$ (**)

(*), (**) and *Contraction* $\Rightarrow R_1, REP_1 \perp\!\!\!\perp REP_2 | H$

Decomposition $\Rightarrow$ $REP_1 \perp\!\!\!\perp REP_2 | H$

d-Separation

- A path *p* is *d*-separated (or blocked) by (a set) Z iff there is a node *w* satisfying either:

  1. *w* has converging arrows ($u \rightarrow w \leftarrow v$) and none of *w* or its descendents are in Z.

  2. *w* does not have converging arrows and $w \in Z$.

- If Z blocks every path from X to Y, then Z *d*-separates X from Y and $X \perp\!\!\!\perp Y | Z$.

How to Construct a Bayesian Network

1. Specify all relevant variables.
2. Specify all conditional independences which hold between them.
3. Construct a Bayesian Network which exhibits these independencies.
4. Check other (perhaps unwanted) independencies with the *d*-separation criterion. Modify the networks.
5. Specify the prior probabilities of all root nodes and the conditional probabilities of all other (child) nodes given their parents.
6. Calculate any probability you are interested in.

3. Modeling Partially Reliable Information Sources

When we receive information from *independent* and *partially reliable* sources, what is our degree of confidence that this information is true?

– Independence?
– Partial reliability?

A. Independence

REP_i

is independent of

F_1, REP_1, F_{i-1}, $REP_{i\text{-}1}$, F_{i+1}, REP_{i+1}, F_n, REP_n

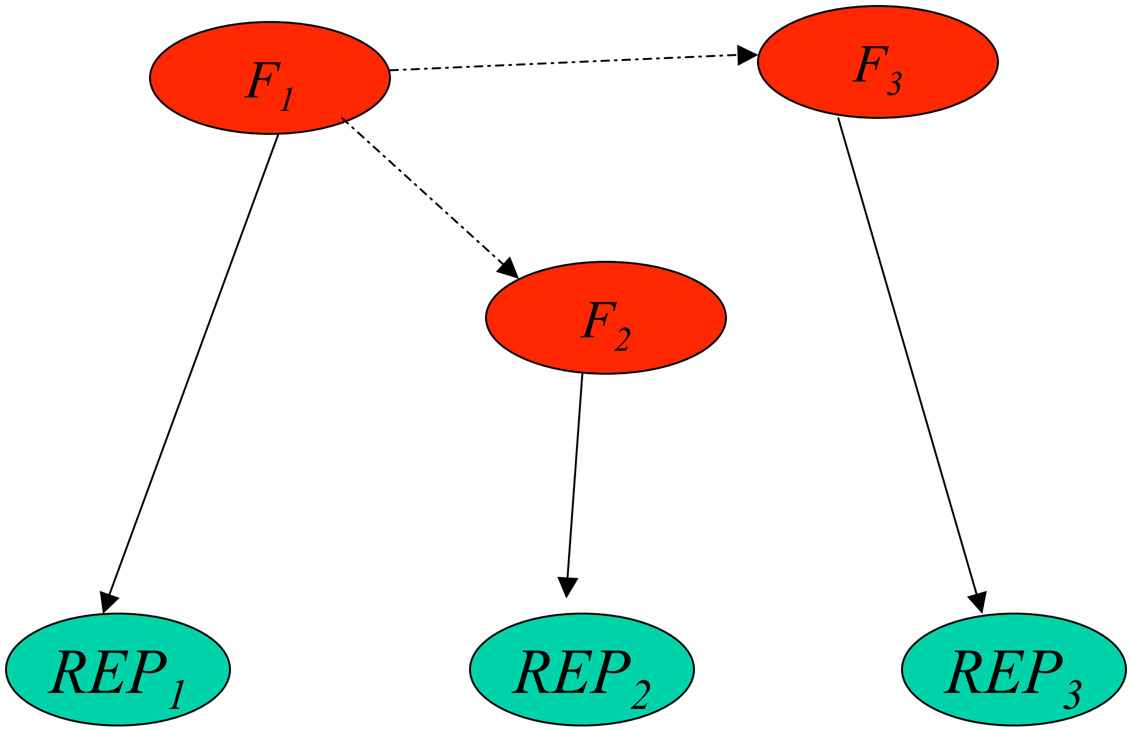
given F_i

B. Partial Reliability

To model partially reliable information sources, additional model assumptions have to be made.

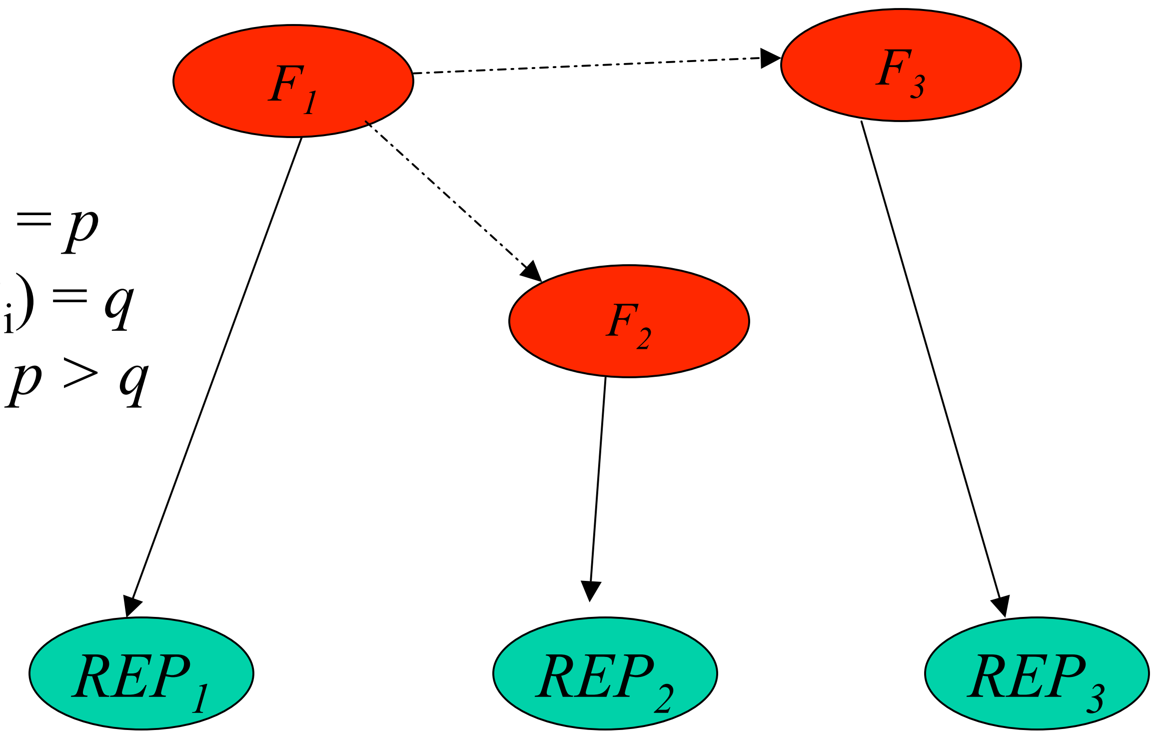
Examine *two* models!

Model I: Fixed Reliability
Paradigm: Medical Testing



Model I: Fixed Reliability
Paradigm: Medical Testing

$P(\mathrm{REP_i}|\mathrm{F_i}) = p$
$P(\mathrm{REP_i}|\neg\mathrm{F_i}) = q$
for $p > q$



Measuring Reliability

- Let's assume that we get positive reports.

  $p := P(\mathrm{REP}_i | \mathrm{F}_i)$ (for all $i = 1,\dots, n$)

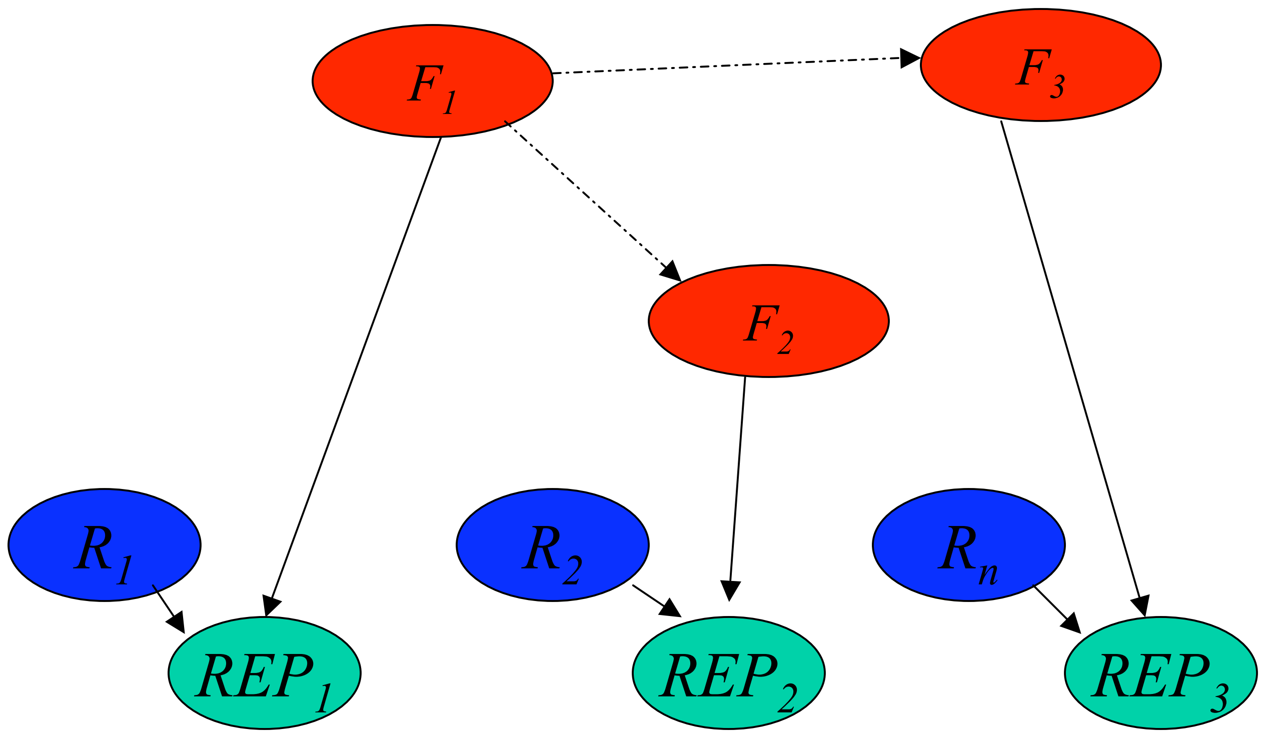
  $q := P(\mathrm{REP}_i | \neg \mathrm{F}_i)$ with $p > q$

  $r := 1 - q/p$

Randomization Full Reliability

0---1

Model II: Variable Reliability, Fixed Randomization Parameter Paradigm: Scientific Instruments



Model II: Variable Reliability, Fixed Randomization Parameter Paradigm: Scientific Instruments

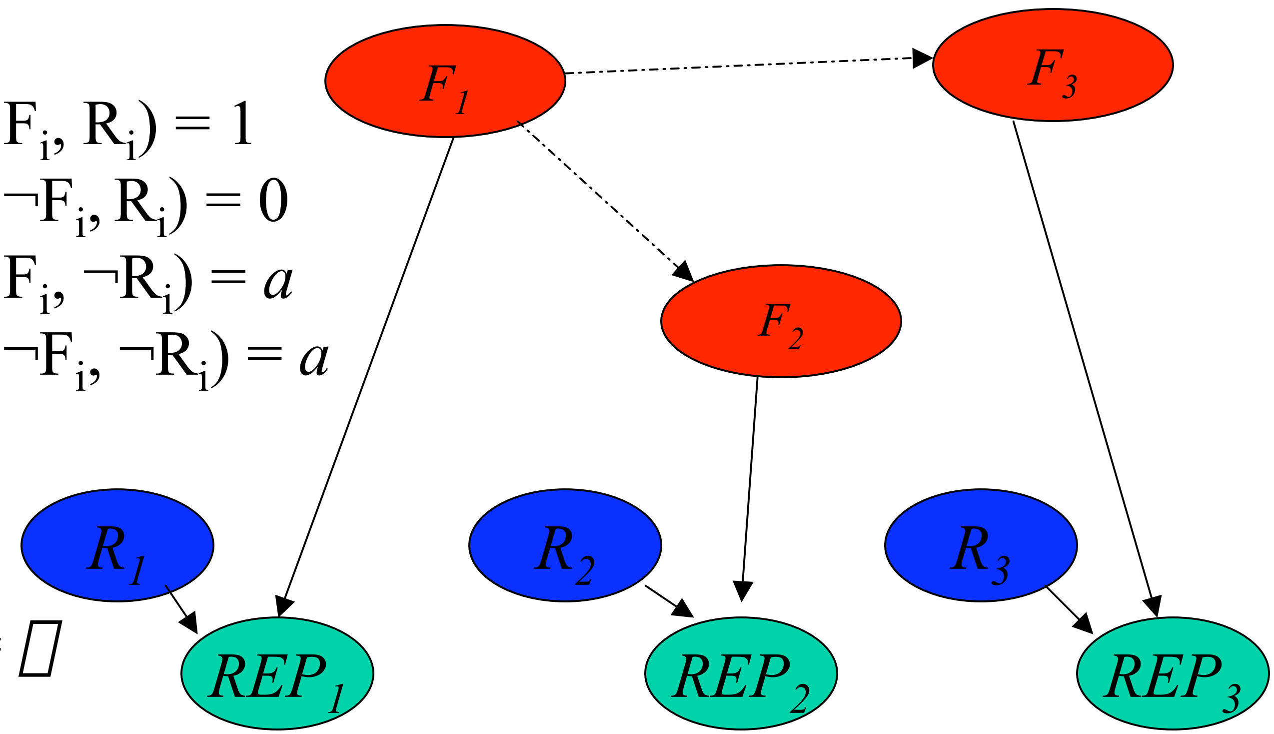
$P(\mathrm{REP_i}|\mathrm{F_i}, \mathrm{R_i}) = 1$

$P(\mathrm{REP_i}|\neg\mathrm{F_i}, \mathrm{R_i}) = 0$

$P(\mathrm{REP_i}|\mathrm{F_i}, \neg\mathrm{R_i}) = a$

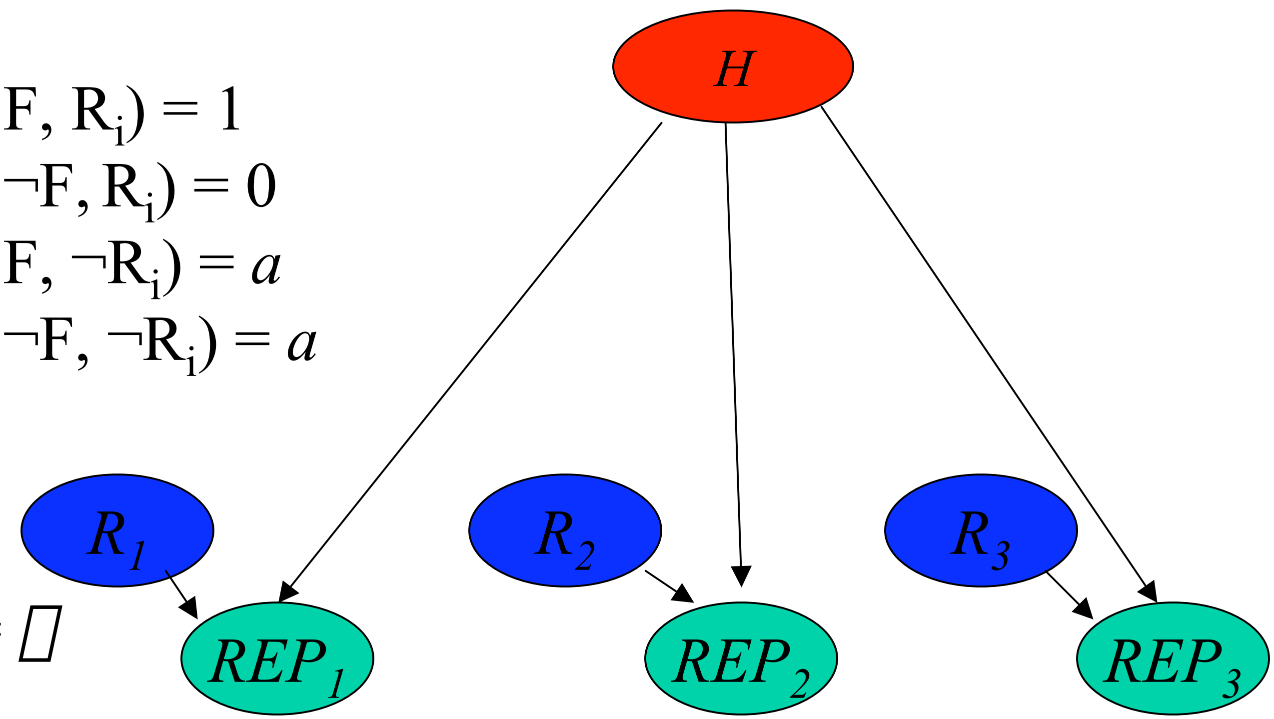
$P(\mathrm{REP_i}|\neg\mathrm{F_i}, \neg\mathrm{R_i}) = a$

$P(\mathrm{R_i}) = \square$



Model IIa: Testing One Hypothesis

$P(\mathrm{REP}_i|\mathrm{F}, \mathrm{R}_i) = 1$
$P(\mathrm{REP}_i|\neg\mathrm{F}, \mathrm{R}_i) = 0$
$P(\mathrm{REP}_i|\mathrm{F}, \neg\mathrm{R}_i) = a$
$P(\mathrm{REP}_i|\neg\mathrm{F}, \neg\mathrm{R}_i) = a$



$P(\mathrm{R}_i) = \square$

Taking Stock

- The Parental Markov Condition is part of the definition of a Bayesian Network.
- *The d*-separation criterion helps us to identify all conditional independences in a Bayesian Network.
- We constructed two basic models of partially reliable information sources:

  - reliability endogenous (medical testing)

  - reliability exogenous (scientific instruments)