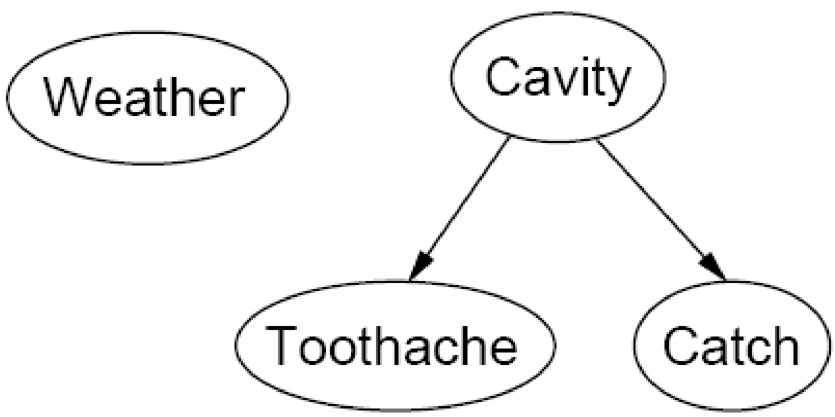
# Bayesian Networks

- Syntax
- Semantics
- Parametrized Distributions

Includes many slides by Russell & Norvig

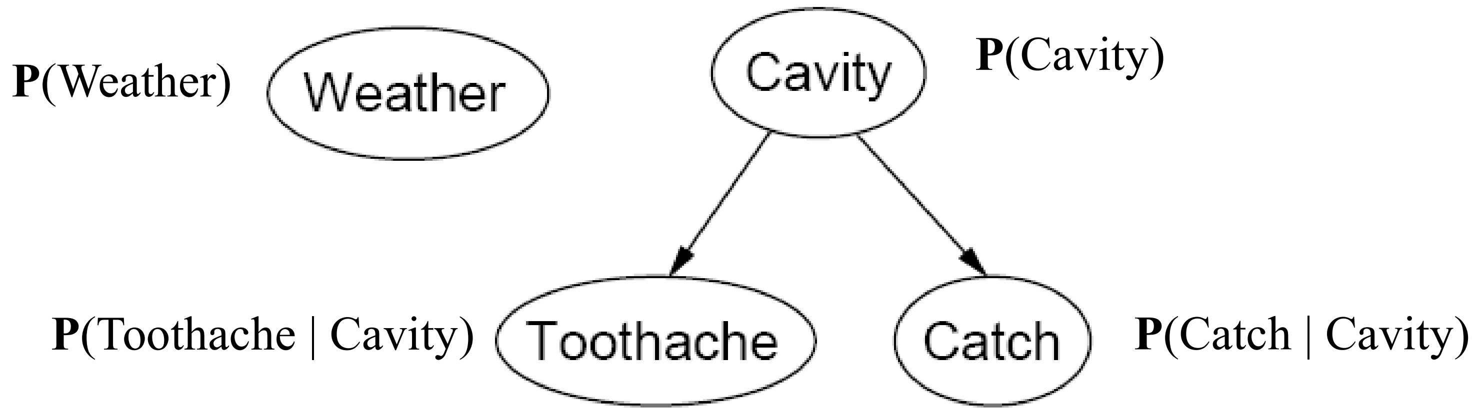
# Bayesian Networks - Structure

- Are a simple, graphical notation for conditional independence assertions
  - hence for compact specifications of full joint distributions
- A BN is a directed graph with the following components:
  - **Nodes:** one node for each variable
  - **Edges:** a directed edge from node $N_i$ to node $N_j$ indicates that variable $X_i$ has a direct influence upon variable $X_j$



# Bayesian Networks - Probabilities

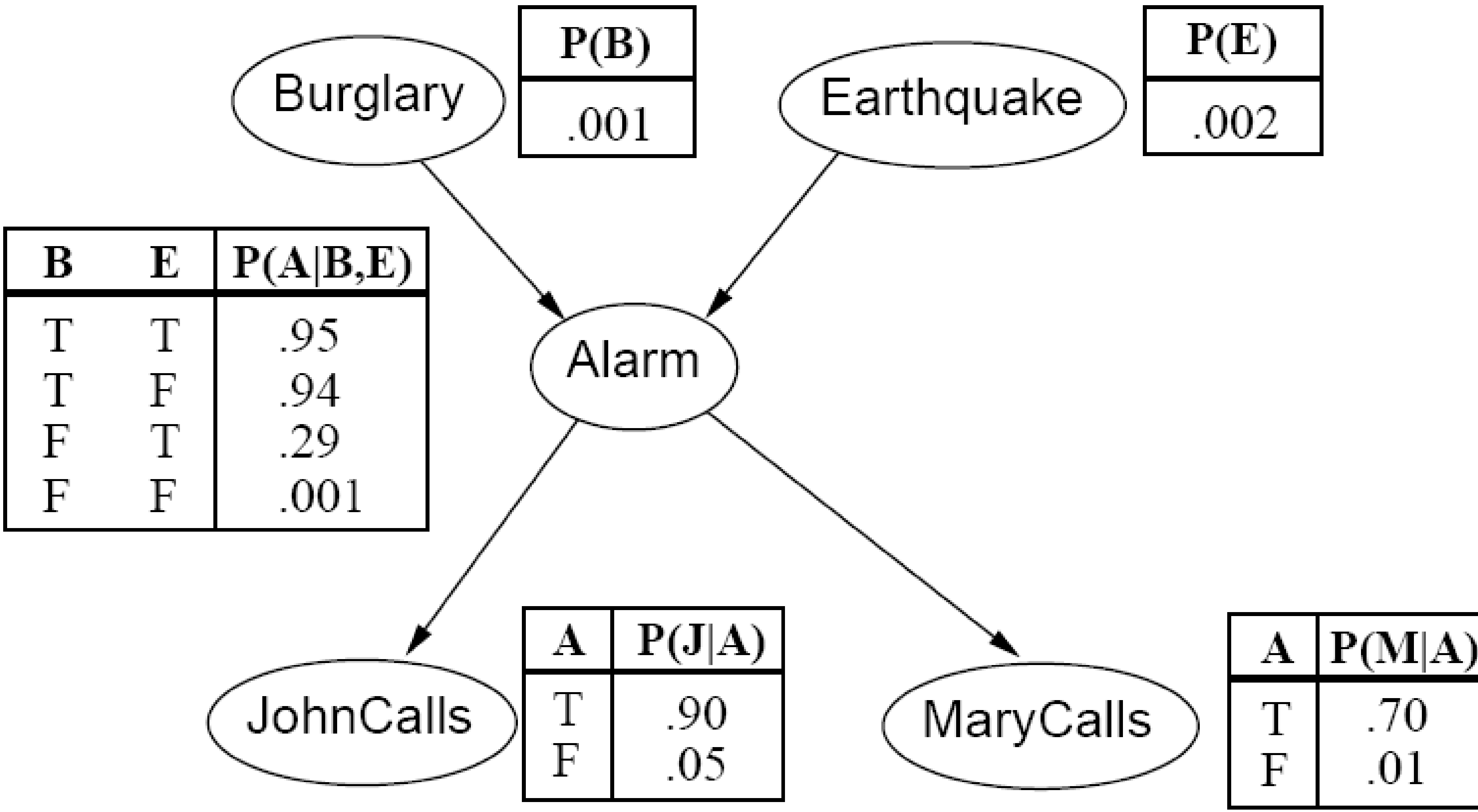
- In addition to the structure, we need a **conditional probability distribution** for the random variable of each node given the random variables of its parents.
  - i.e. we need $\mathbf{P}(X_i \mid \text{Parents}(X_i))$
- nodes/variables that are not connected are (conditionally) independent:
  - Weather is independent of Cavity
  - Toothache is independent of Catch given Cavity



# Running Example: Alarm

- Situation:
  - I'm at work
  - John calls to say that the in my house alarm went off
    - but Mary (my neighbor) did not call
  - The alarm will usually be set off by burglars
    - but sometimes it may also go off because of minor earthquakes
- Variables:
  - *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, *MaryCalls*
- Network topology reflects causal knowledge:
  - A burglar can set the alarm off
  - An earthquake can set the alarm off
  - The alarm can cause Mary to call
  - The alarm can cause John to call

# Alarm Example




| P(B) |
|---|
| .001 |

| P(E) |
|---|
| .002 |

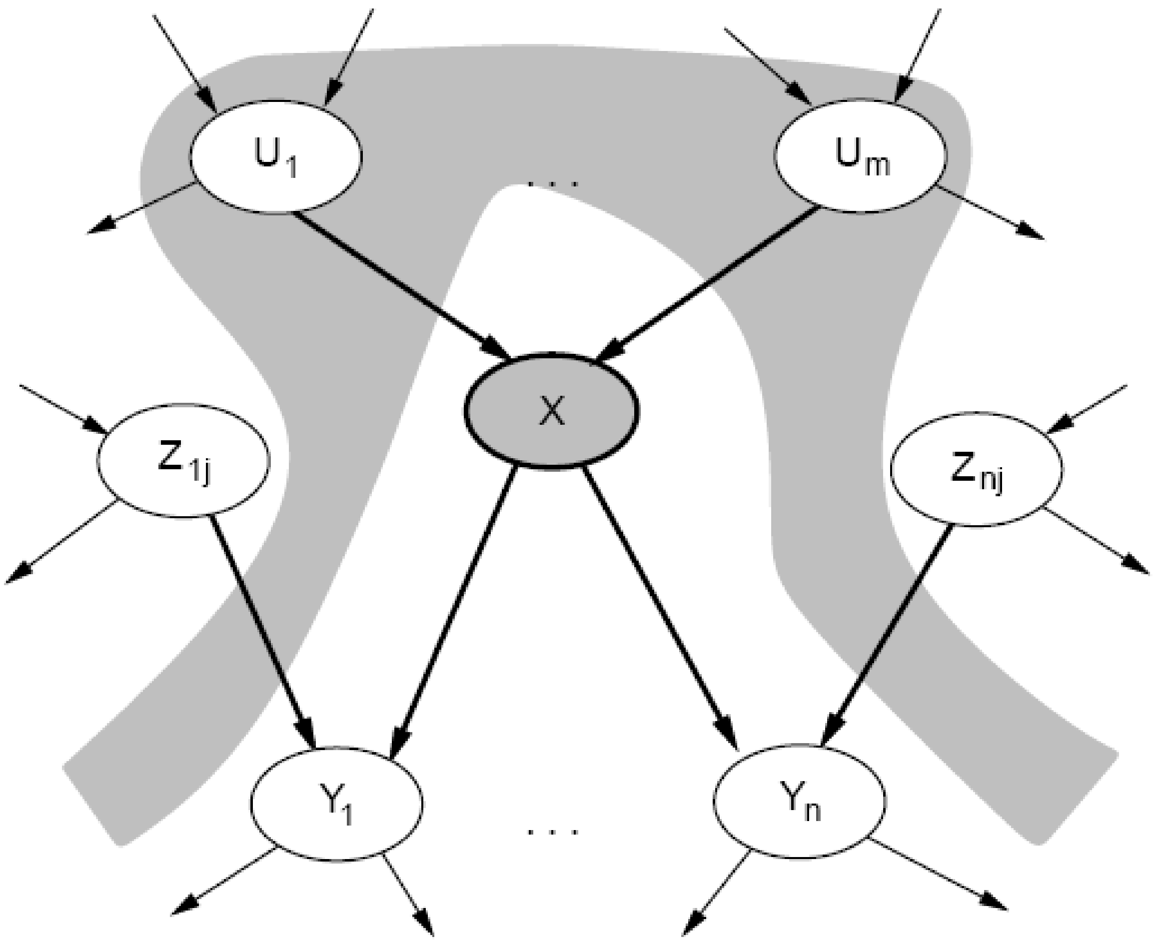
| B | E | P(A\|B,E) |
|---|---|---|
| T | T | .95 |
| T | F | .94 |
| F | T | .29 |
| F | F | .001 |

| A | P(J\|A) |
|---|---|
| T | .90 |
| F | .05 |

| A | P(M\|A) |
|---|---|
| T | .70 |
| F | .01 |

# Local Semantics of a BN

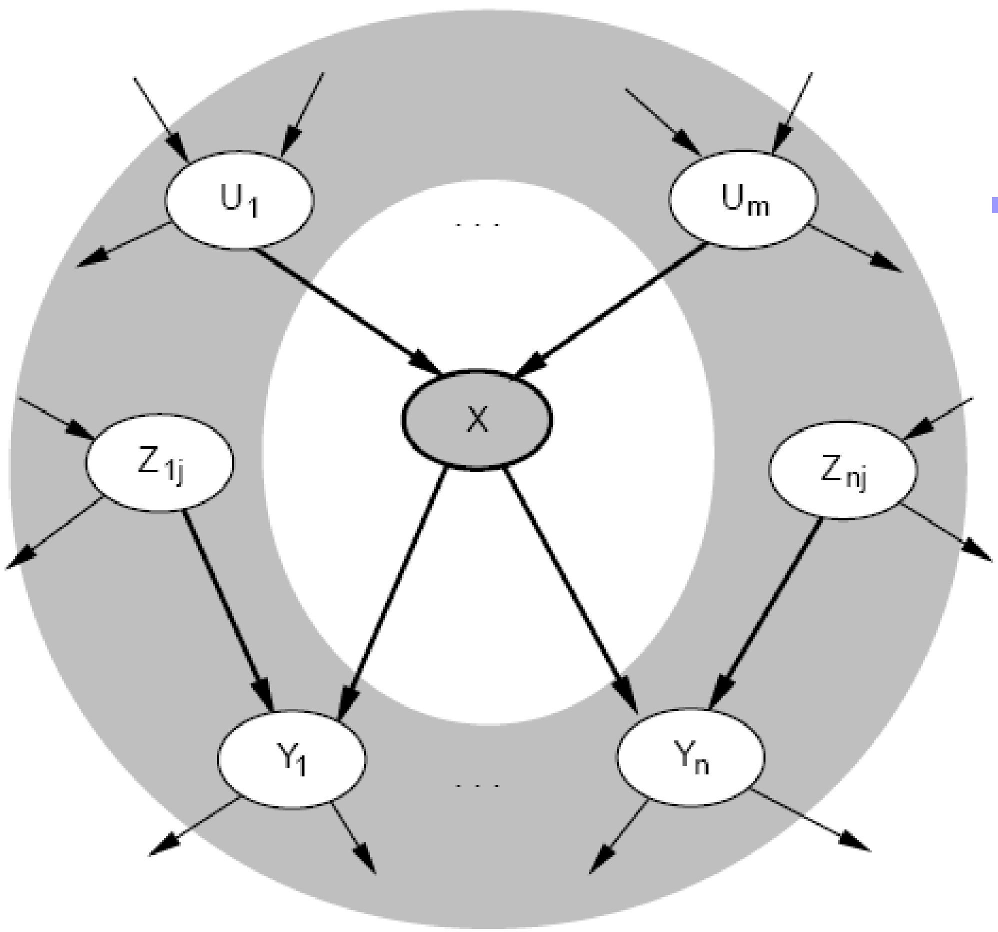
- Each node is is conditionally independent of its nondescendants given its parents




$$\mathbf{P}(X \mid U_{1}, \ldots, U_{m}, Z_{1j}, \ldots, Z_{nj}) = \mathbf{P}(X \mid U_{1}, \ldots, U_{m})$$

# Markov Blanket

- **Markov Blanket:**
  - parents + children + children's parents




- Each node is conditionally independent of all other nodes given its markov blanket

$$\mathbf{P}(X \mid U_{1,}\ldots, U_m, Y_{1,}\ldots, Y_n, Z_{1j}, \ldots, Z_{nj}) =$$
$$= \mathbf{P}(X \mid all\ variables)$$

# Global Semantics of a BN

- The conditional probability distributions define the joint probability distribution of the variables of the network

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid \text{Parents}(X_i))$$

- Example:
  - What is the probability that the alarm goes off and both John and Mary call, but there is neither a burglary nor an earthquake?

$$P(j \wedge m \wedge a \wedge \neg b \wedge \neg e) =$$

$$= P(j \mid a) \cdot P(m \mid a) \cdot P(a \mid \neg b, \neg e) \cdot P(\neg b) \cdot P(\neg e)$$

$$= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 \approx 0.00063$$

# Theorem

**Local Semantics $\Leftrightarrow$ Global Semantics**

- Proof:
  - order the variables so that parents always appear before children
  - apply chain rule
  - use conditional independence

# Constructing Bayesian Networks

Need a method such that a series of locally testable assertions of conditional independence guarantees the required global semantics

1. Choose an ordering of variables $X_1, \ldots, X_n$
2. For $i = 1$ to $n$
   add $X_i$ to the network
   select parents from $X_1, \ldots, X_{i-1}$ such that
   $\mathbf{P}(X_i | Parents(X_i)) = \mathbf{P}(X_i | X_1, \ldots, X_{i-1})$

This choice of parents guarantees the global semantics:

$$\mathbf{P}(X_1, \ldots, X_n) = \prod_{i=1}^{n} \mathbf{P}(X_i | X_1, \ldots, X_{i-1}) \quad \text{(chain rule)}$$
$$= \prod_{i=1}^{n} \mathbf{P}(X_i | Parents(X_i)) \quad \text{(by construction)}$$

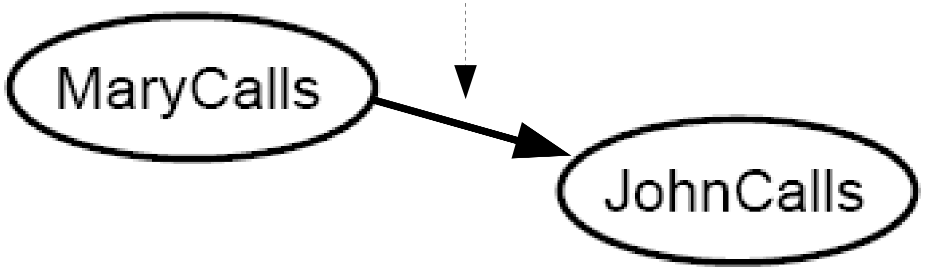
# Example

- Suppose we first select the ordering

  *MaryCalls, JohnCalls, Alarm, Burglary, Earthquake,*

$P(J \mid M) = P(J)?$ ✗

If Mary calls, it is more likely that John calls as well.



# Example

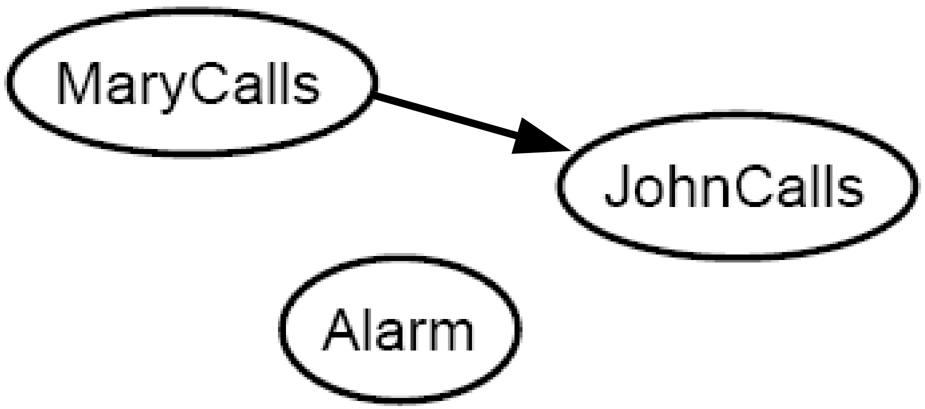
- Suppose we first select the ordering
  *MaryCalls, JohnCalls, Alarm, Burglary, Earthquake,*

$P(A \mid J, M) = P(A)?$ ✗

If Mary and John call, the probability that the alarm has gone off is larger than if they don't call.

↓

Node A needs parents J or M



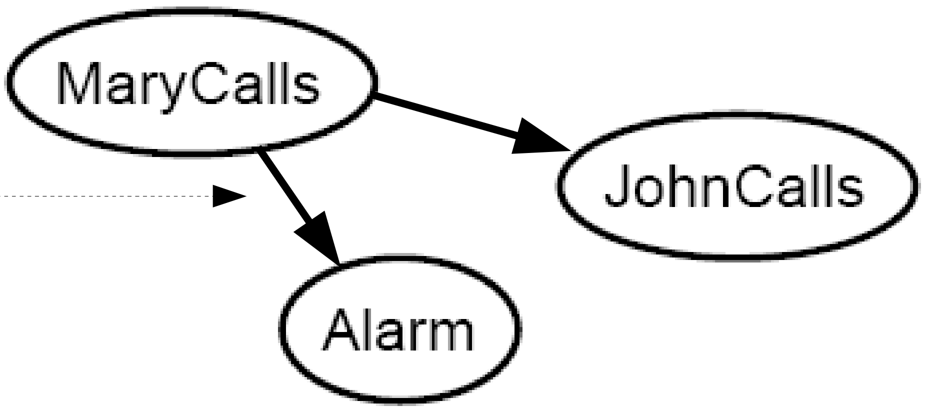
# Example

- Suppose we first select the ordering

  *MaryCalls, JohnCalls, Alarm, Burglary, Earthquake,*

$$P(A \mid J, M) = P(A \mid J)?$$

If John and Mary call, the probability that the alarm has gone off is higher than if only John calls.

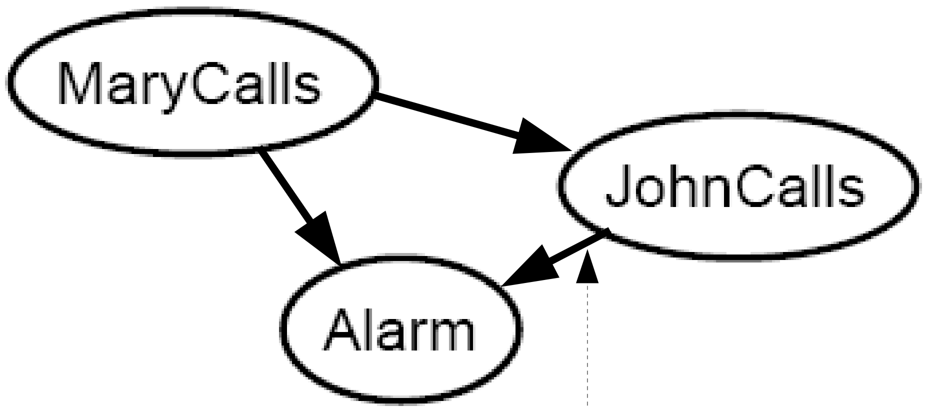


# Example

- Suppose we first select the ordering

  *MaryCalls, JohnCalls, Alarm, Burglary, Earthquake,*

$$P(A \mid J, M) = P(A \mid M)?$$ ✗




If John and Mary call, the probability that the alarm has gone off is higher than if only Mary calls.

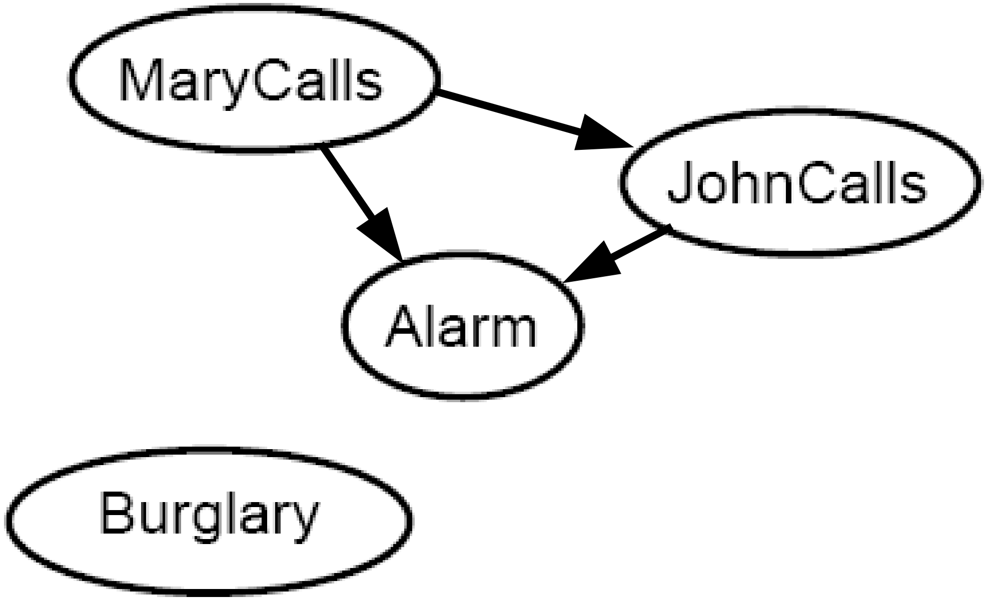
# Example

- Suppose we first select the ordering
  *MaryCalls, JohnCalls, Alarm, Burglary, Earthquake,*

$P(B \mid A, J, M) = P(B)?$ ✗

Knowing whether Mary or John called and whether the alarm went off influences my knowledge about whether there has been a burglary

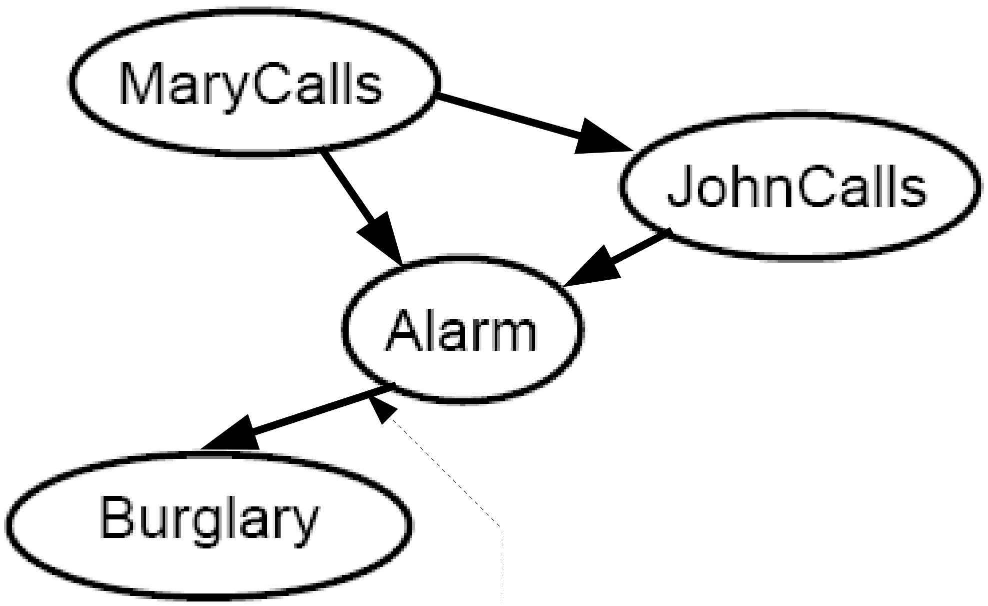
Node B needs parents A, J or M



# Example

- Suppose we first select the ordering

  *MaryCalls, JohnCalls, Alarm, Burglary, Earthquake,*




$$P(B \mid A, J, M) = P(B \mid A)?$$ ✓

If I know that the alarm has gone off, knowing that John or Mary have called does not add to my knowledge of whether there has been a burglary or not.

Thus, no edges from M and J, only from B

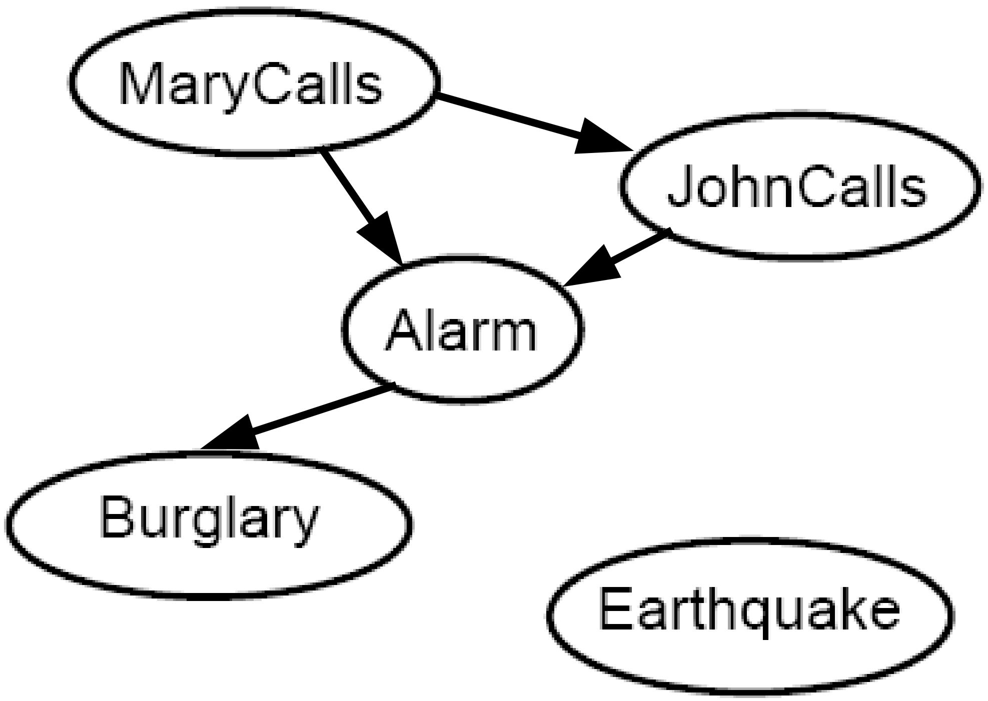
# Example

- Suppose we first select the ordering

  *MaryCalls, JohnCalls, Alarm, Burglary, Earthquake,*

$$P(E \mid B, A, J, M) = P(E \mid A)?$$ ✗

Knowing whether there has been an Earthquake does not suffice to determine the probability of an earthquake, we have to know whether there has been a burglary as well.



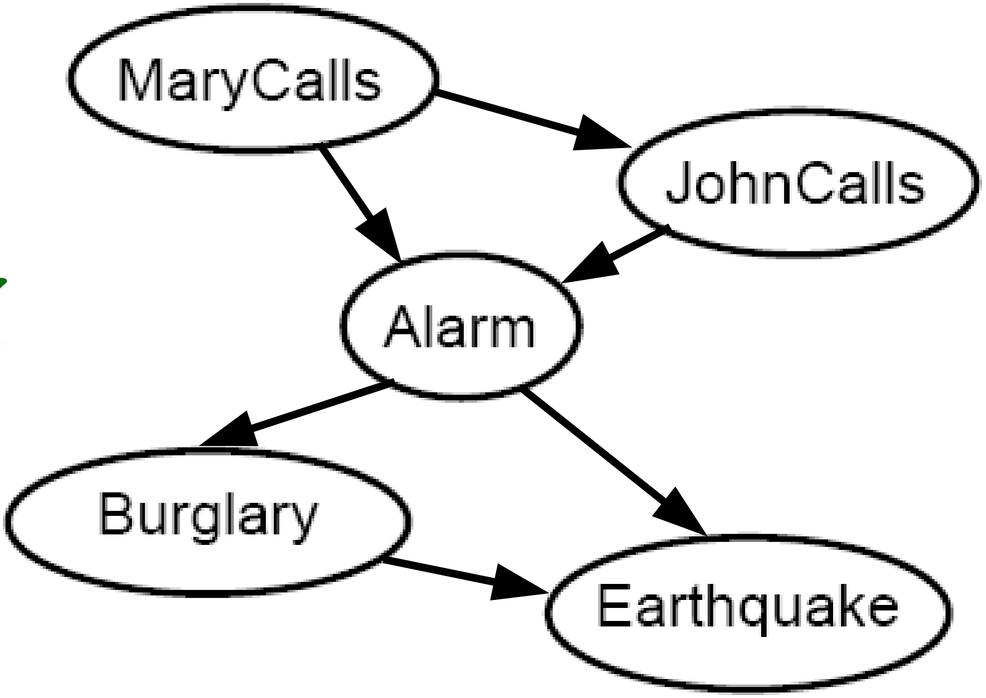
# Example

- Suppose we first select the ordering

  *MaryCalls, JohnCalls, Alarm, Burglary, Earthquake,*

$P(E \mid B, A, J, M) = P(E \mid A)?$ ✗

$P(E \mid B, A, J, M) = P(E \mid A, B)?$ ✓

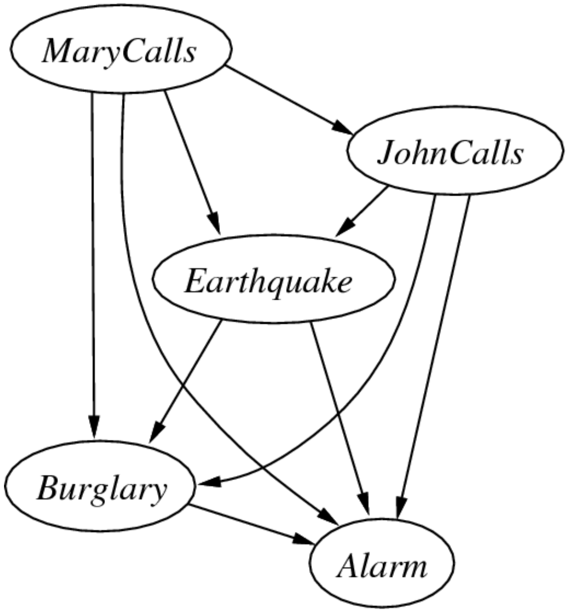



If we know whether there has been an alarm and whether there has been burglary, no other factors will determine our knowledge about whether there has been an earthquake

# Example - Discussion

- Deciding conditional independence is hard in non-causal direction
  - for assessing whether $X$ is conditionally independent of $Z$ ask the question: *If I add variable $Z$ in the condition, does it change the probabilities for $X$?*
  - causal models and conditional independence seem hardwired for humans!
- Assessing conditional probabilities is also hard in non-causal direction

- Network is less compact
  - more edges and more parameters to estimate
- Worst possible ordering
  *MaryCalls, JohnCalls*
  *Earthquake, Burglary, Alarm*
  → fully connected network

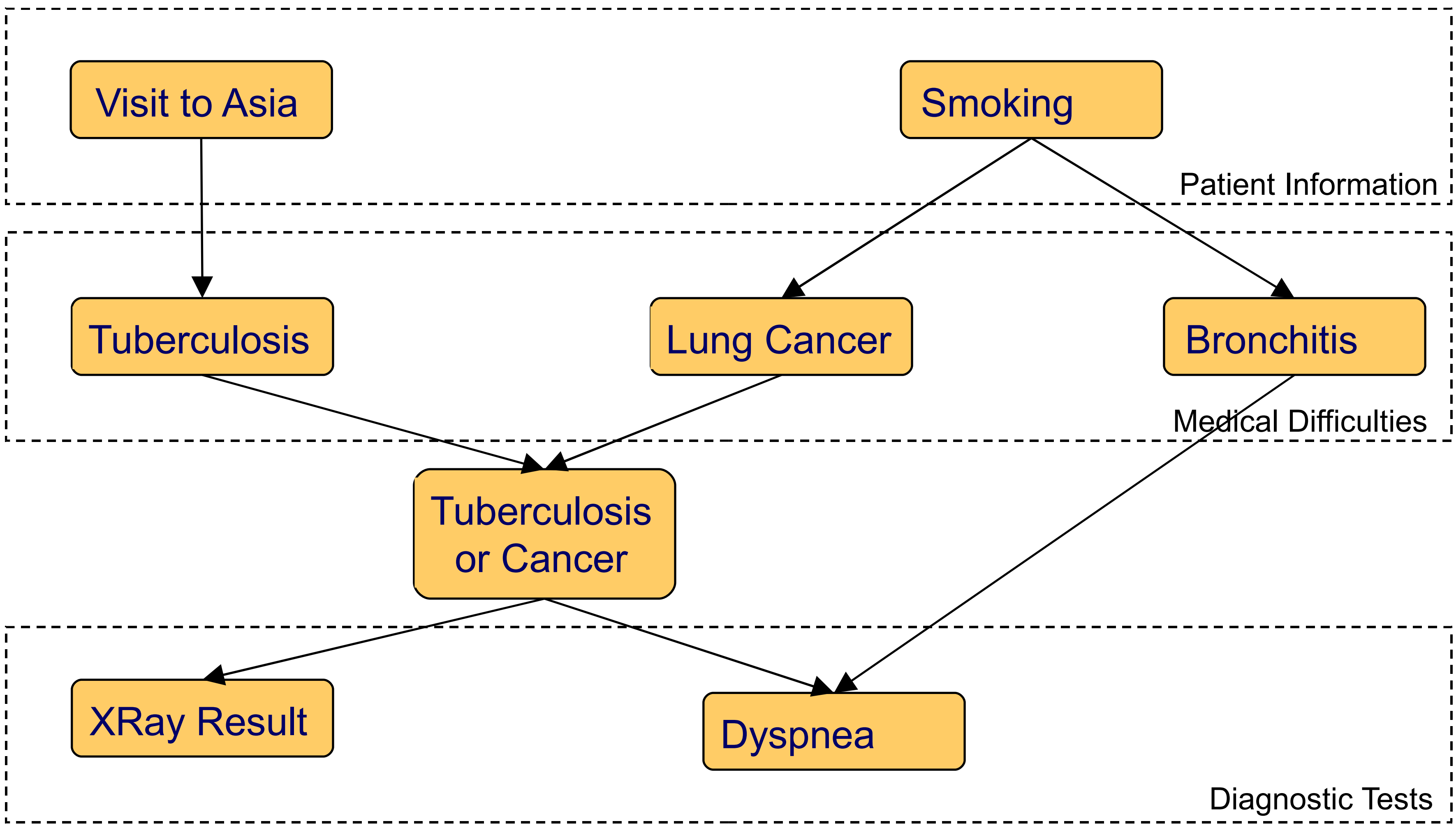


# Reasoning with Bayesian Networks

- Belief Functions (margin probabilities)
  - given the probability distribution, we can compute a margin probability at each node, which represents the belief into the truth of the proposition
    → the margin probability is also called the belief function
- New evidence can be incorporated into the network by changing the appropriate belief functions
  - this may not only happen in unconditional nodes!
- changes in the margin probabilities are then propagated through the network
  - propagation happens in forward (along the causal links) and backward direction (against them)
    - e.g., determining a symptom of a disease does not cause the disease, but changes the probability with which we believe that the patient has the disease

# Example: Medical Diagnosis

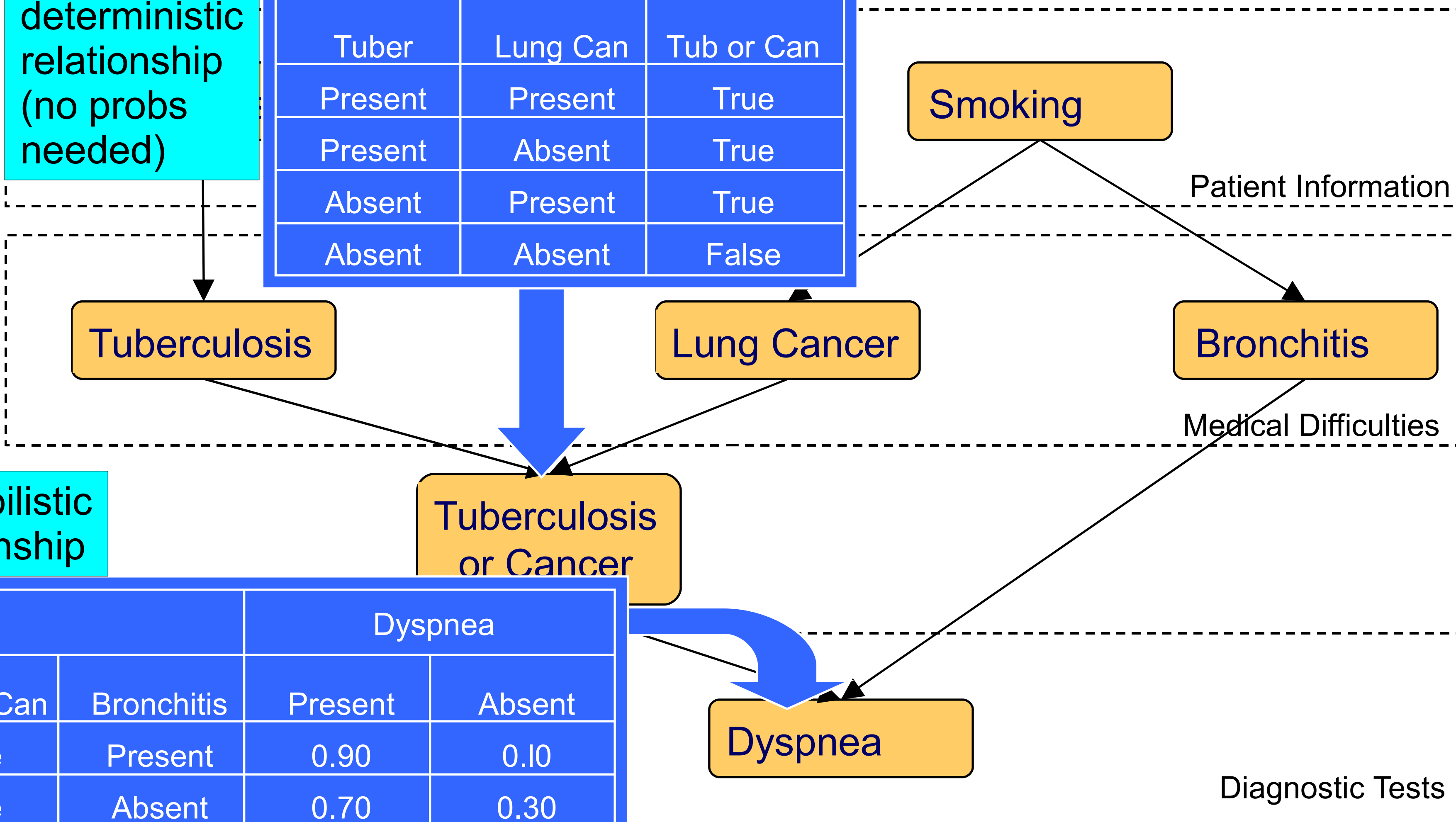
- Structure of the network




 Slide by Buede, Tatman, Bresnik, 1998 

# Example: Medical Diagnosis

- Adding Probability Distributions




deterministic relationship (no probs needed)

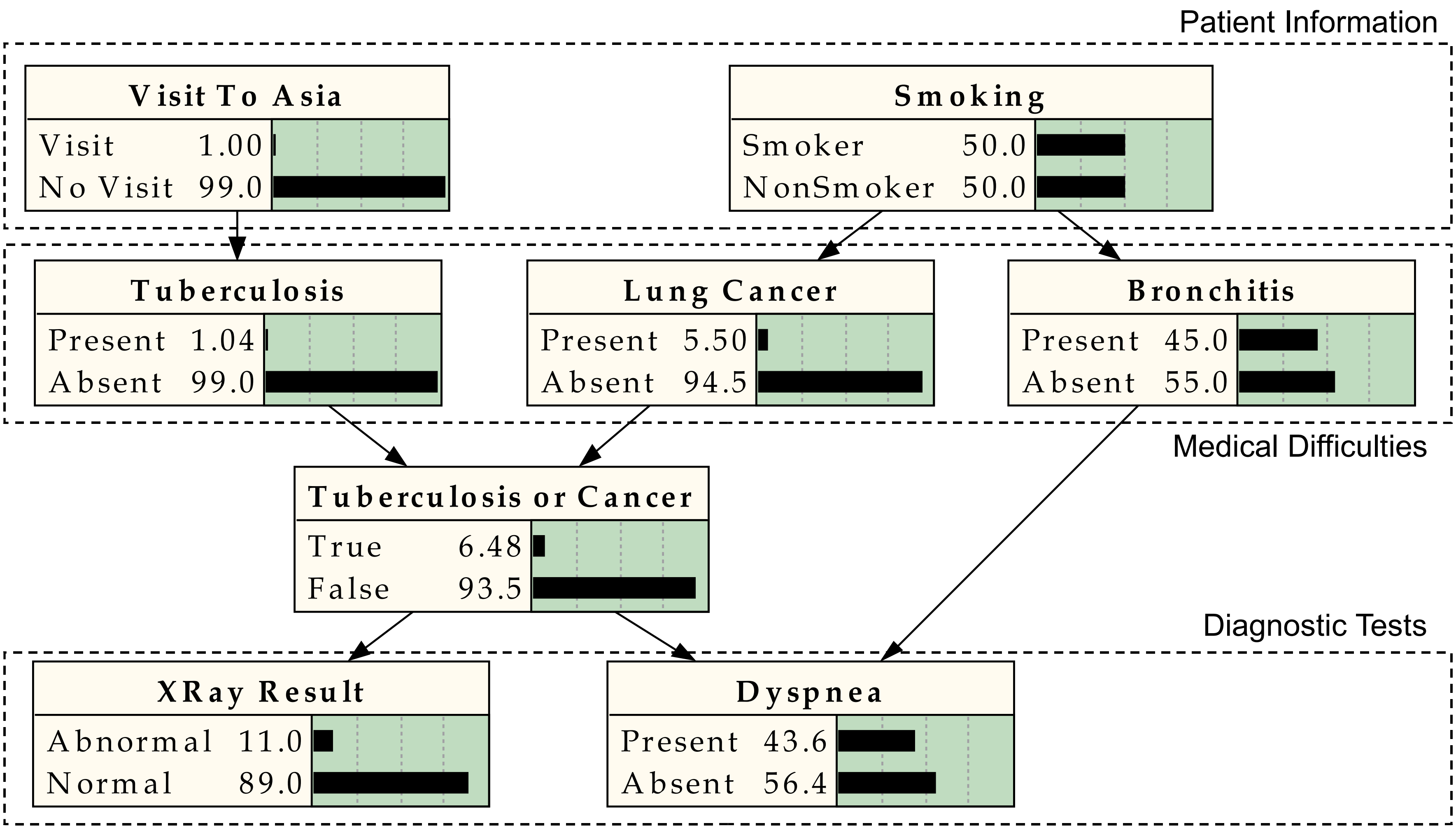
| Tuber | Lung Can | Tub or Can |
|---|---|---|
| Present | Present | True |
| Present | Absent | True |
| Absent | Present | True |
| Absent | Absent | False |

probabilistic relationship

| | | Dyspnea | |
|---|---|---|---|
| Tub or Can | Bronchitis | Present | Absent |
| True | Present | 0.90 | 0.I0 |
| True | Absent | 0.70 | 0.30 |
| False | Present | 0.80 | 0.20 |
| False | Absent | 0.10 | 0.90 |

# Example: Medical Diagnosis

- Belief functions

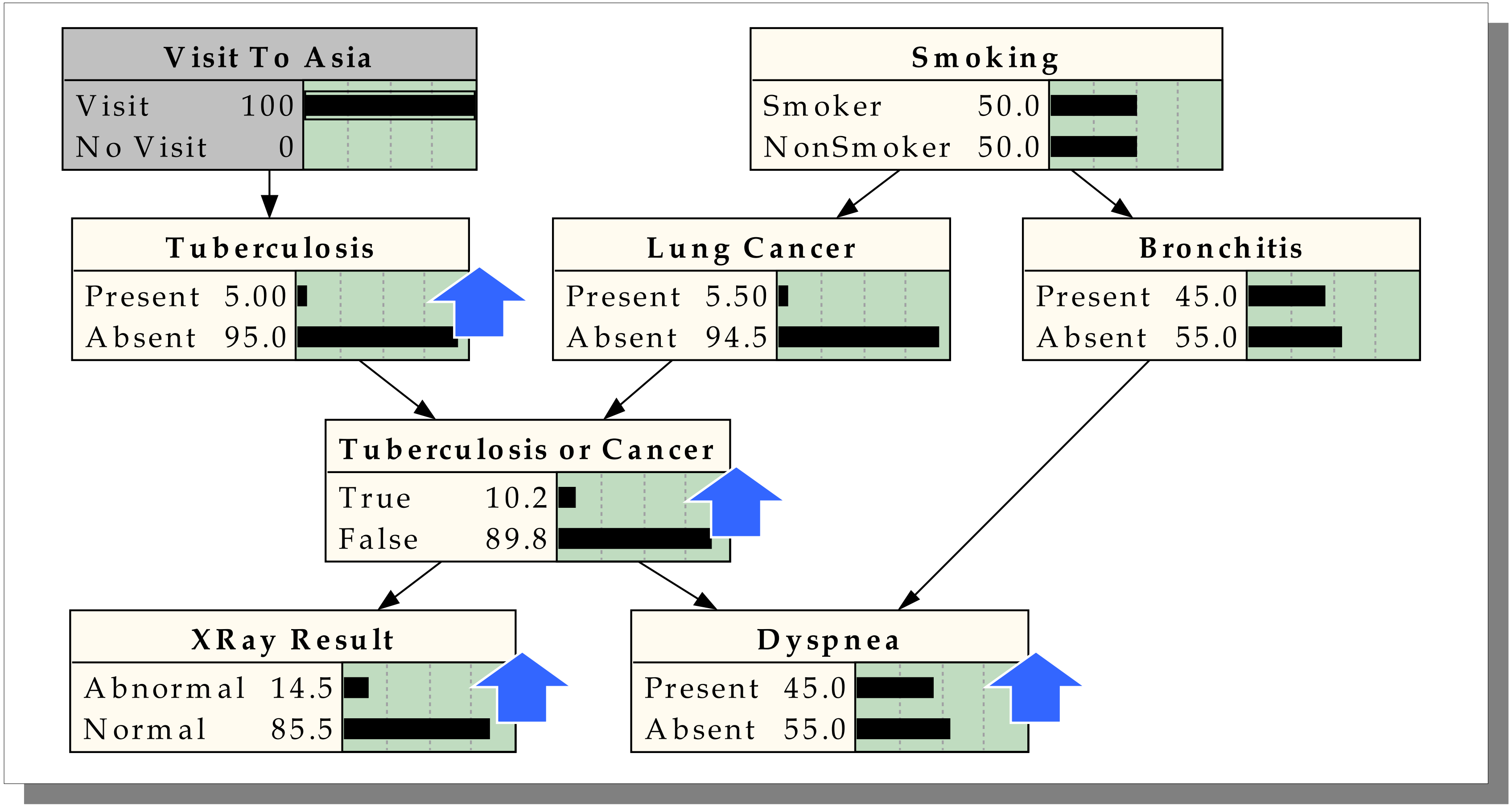




Slide by Buede, Tatman, Bresnik, 1998

# Example: Medical Diagnosis

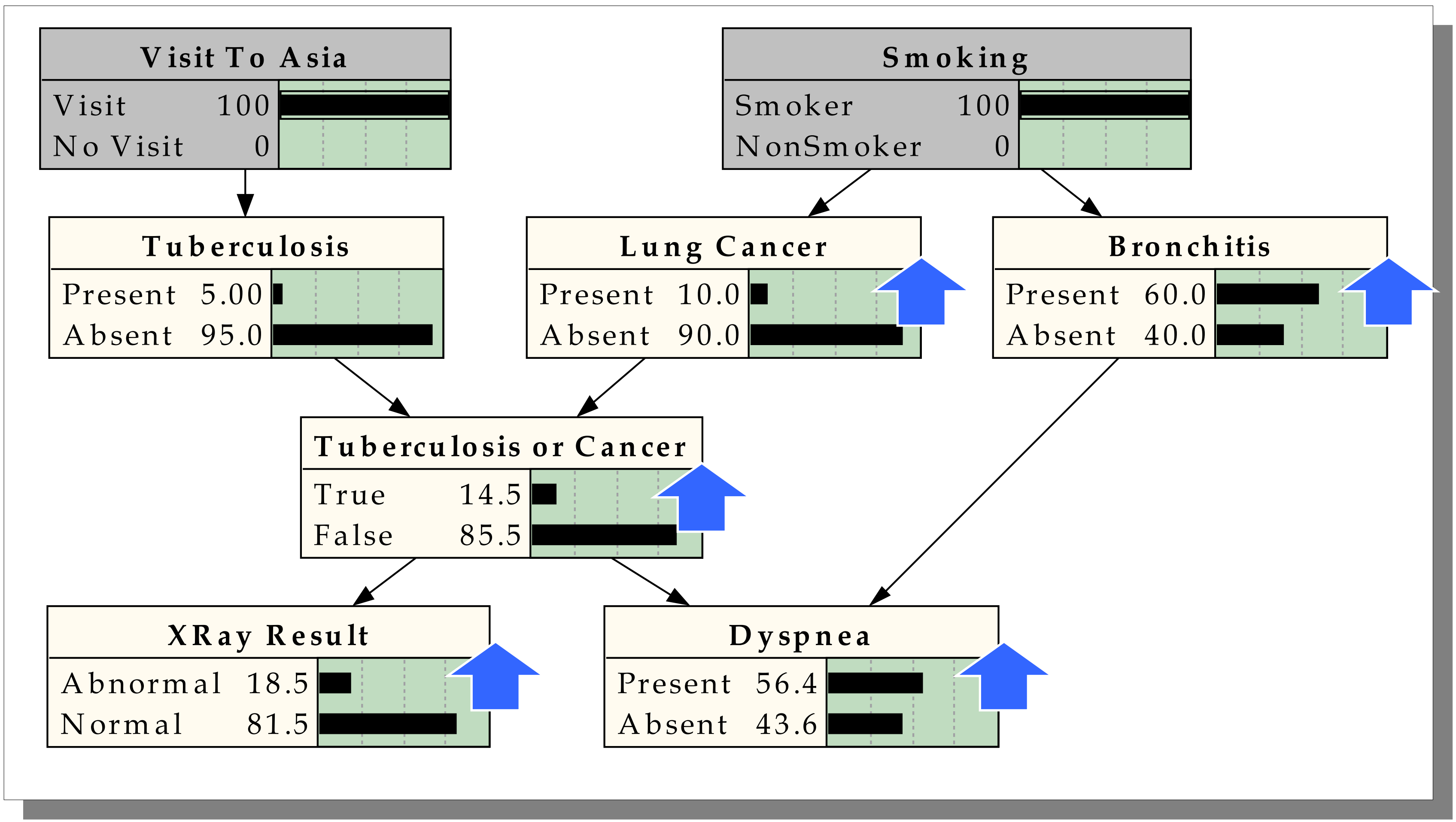
- Interviewing the patient results in change of probability for variable for "visit to Asia"




Slide by Buede, Tatman, Bresnik, 1998

# Example: Medical Diagnosis

- Patient is also a smoker...

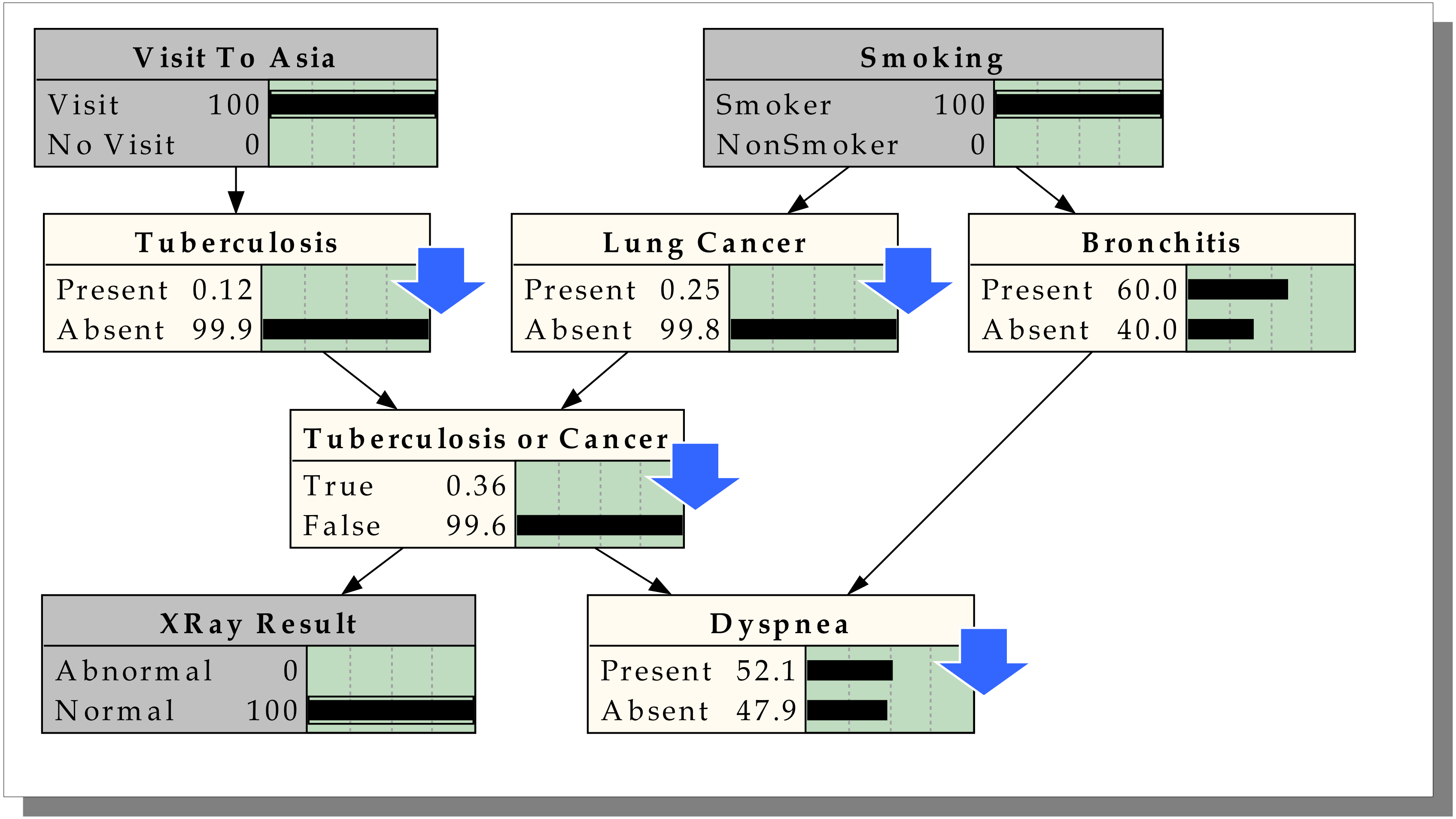




Slide by Buede, Tatman, Bresnik, 1998

# Example: Medical Diagnosis

- but fortunately the X-ray is normal...

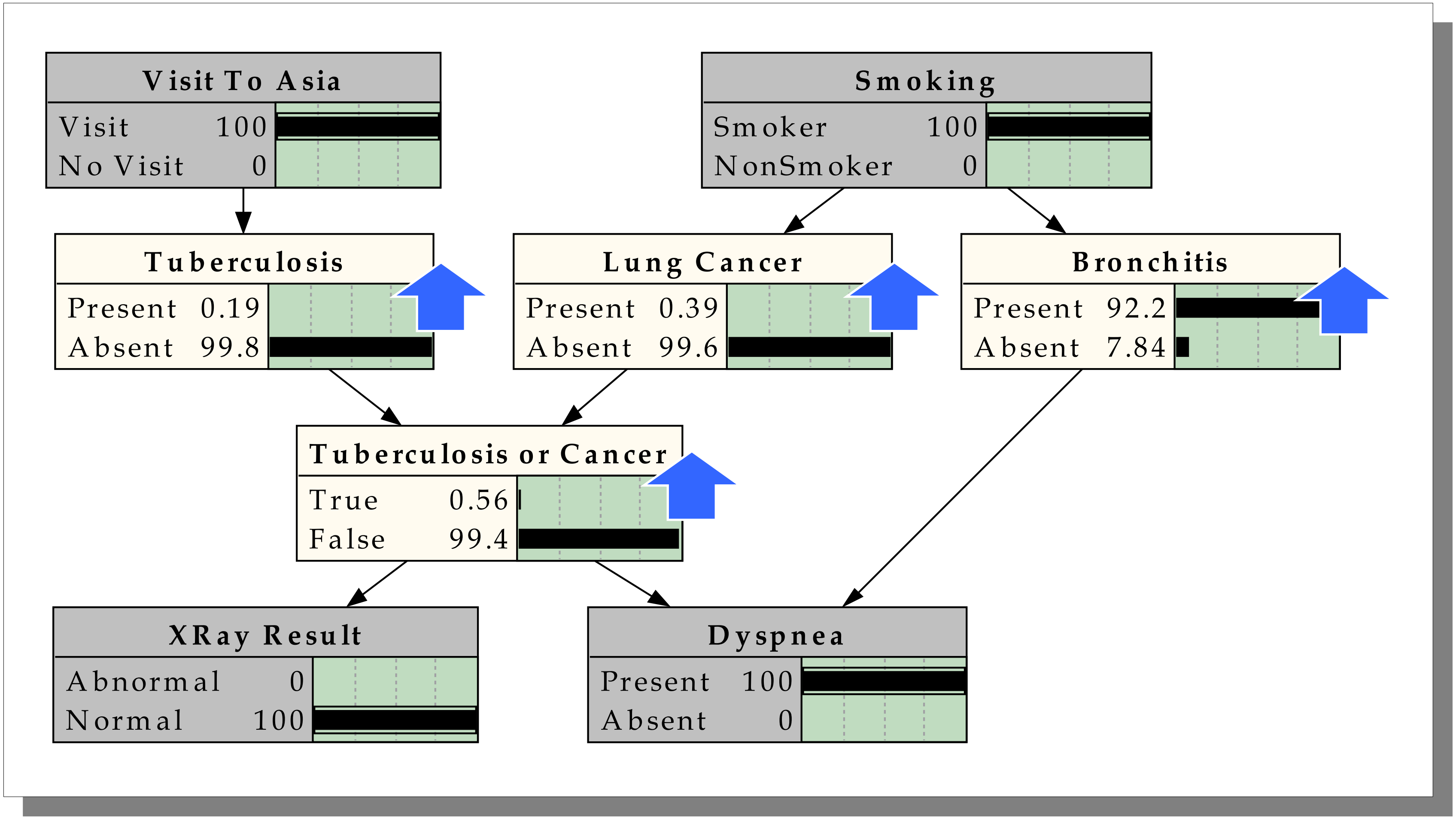




Slide by Buede, Tatman, Bresnik, 1998

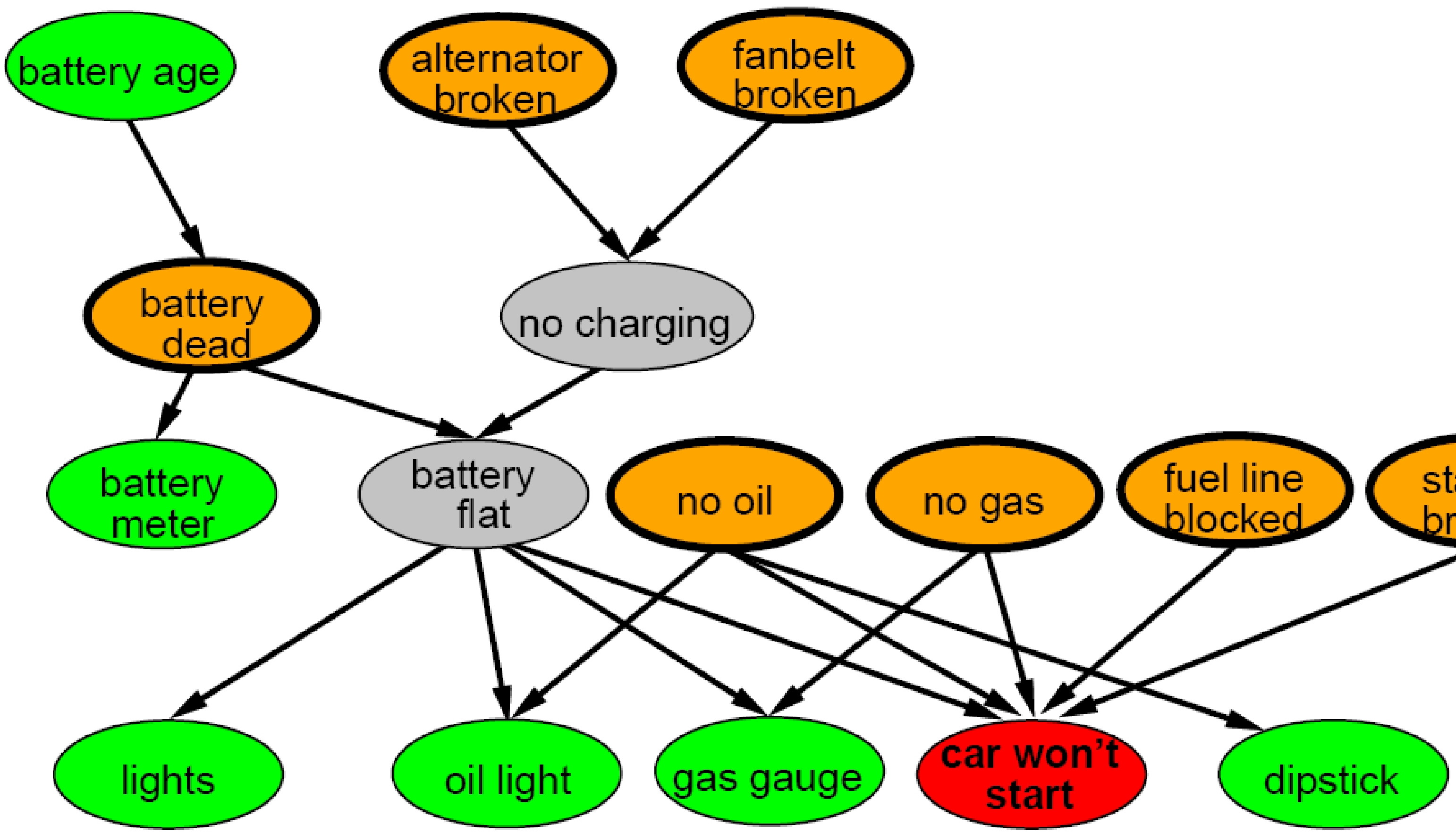
# Example: Medical Diagnosis

- but then again patient has difficulty in breathing.

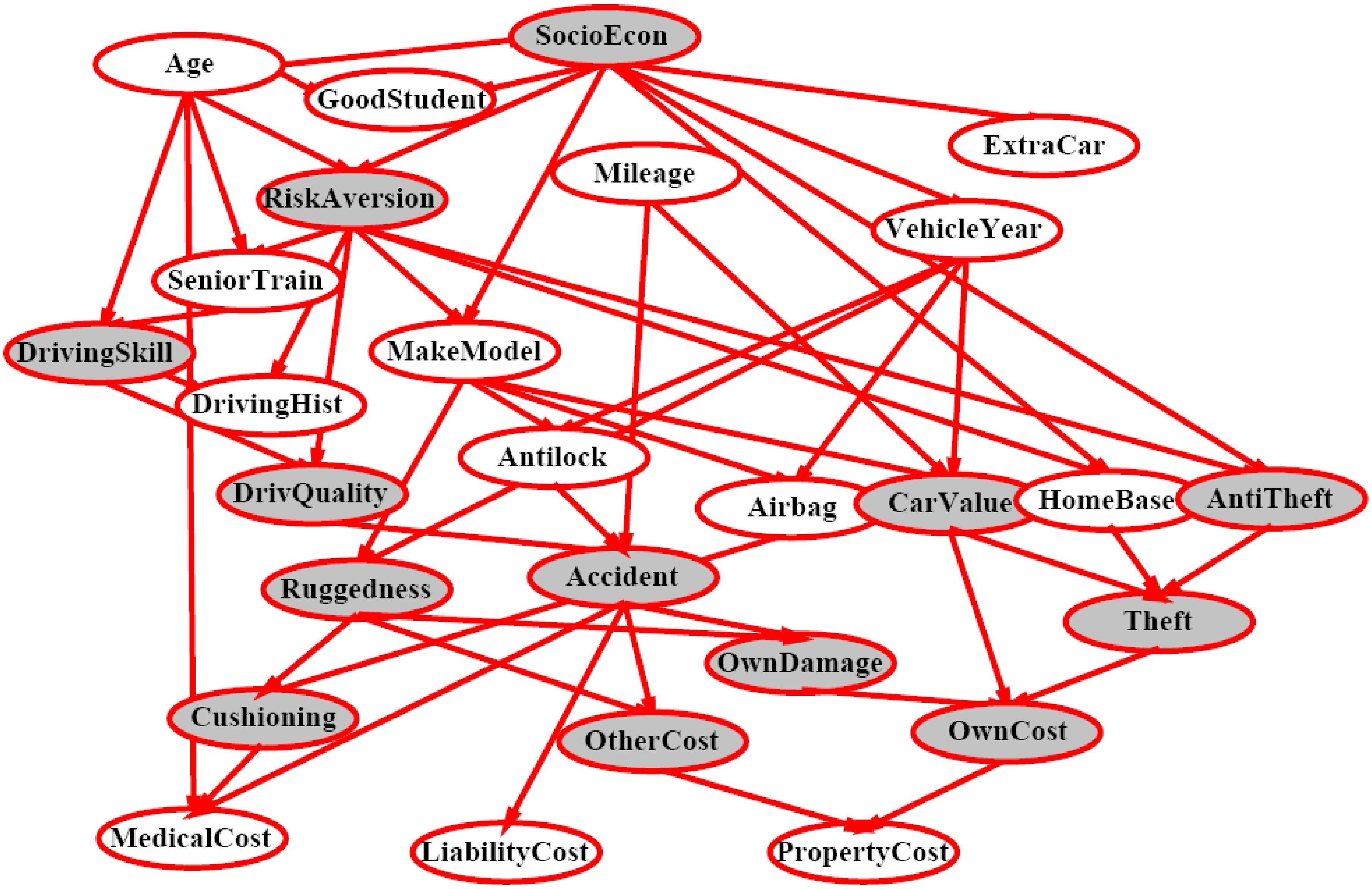


# More Complex Example: Car Diagnosis

- Initial evidence: Car does not start
- Test variables
- Variables for possible failures
- Hidden variables: ensure spare structure, reduce parameters

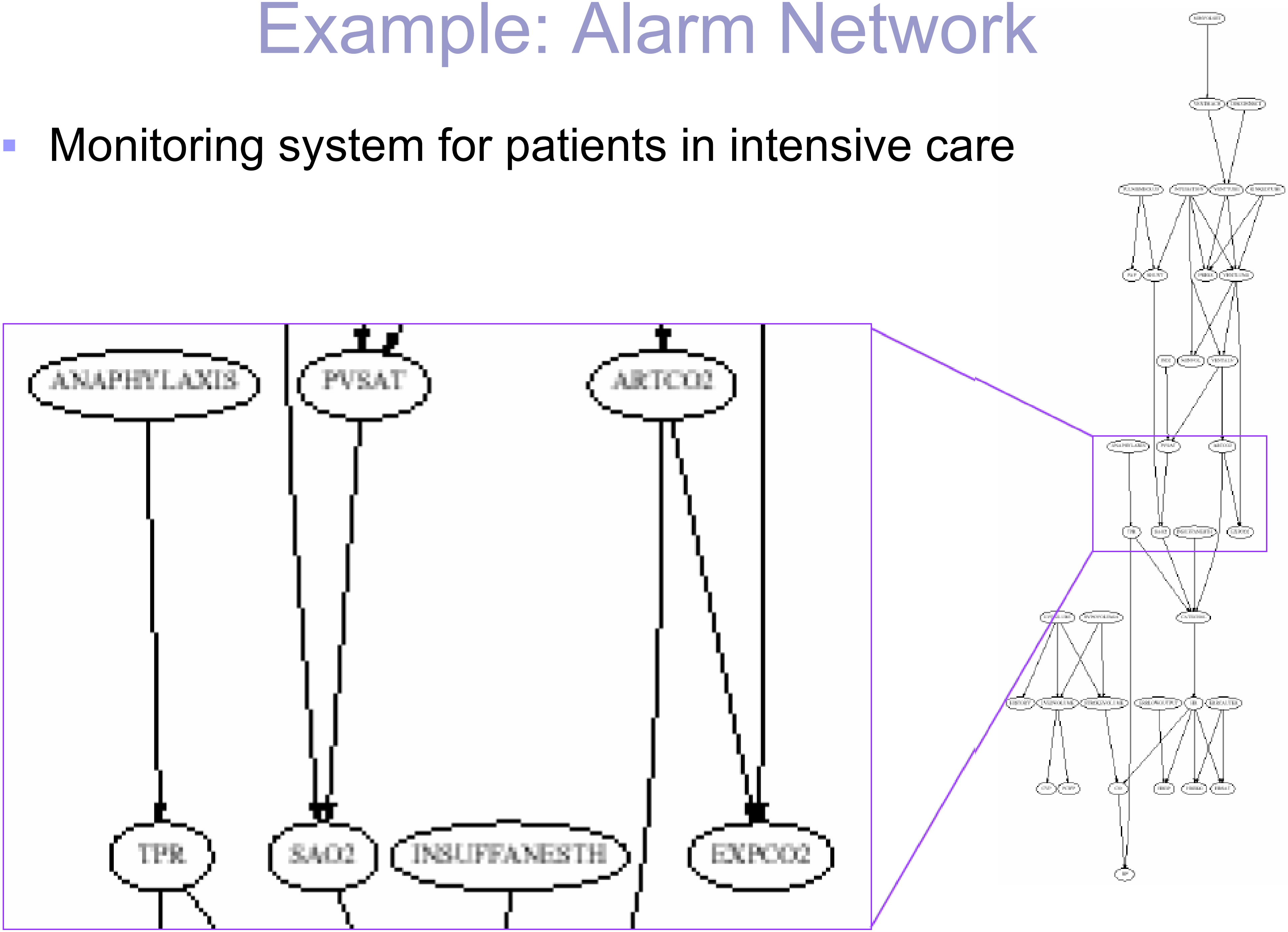


# More Complex: Car Insurance



# Example: Alarm Network

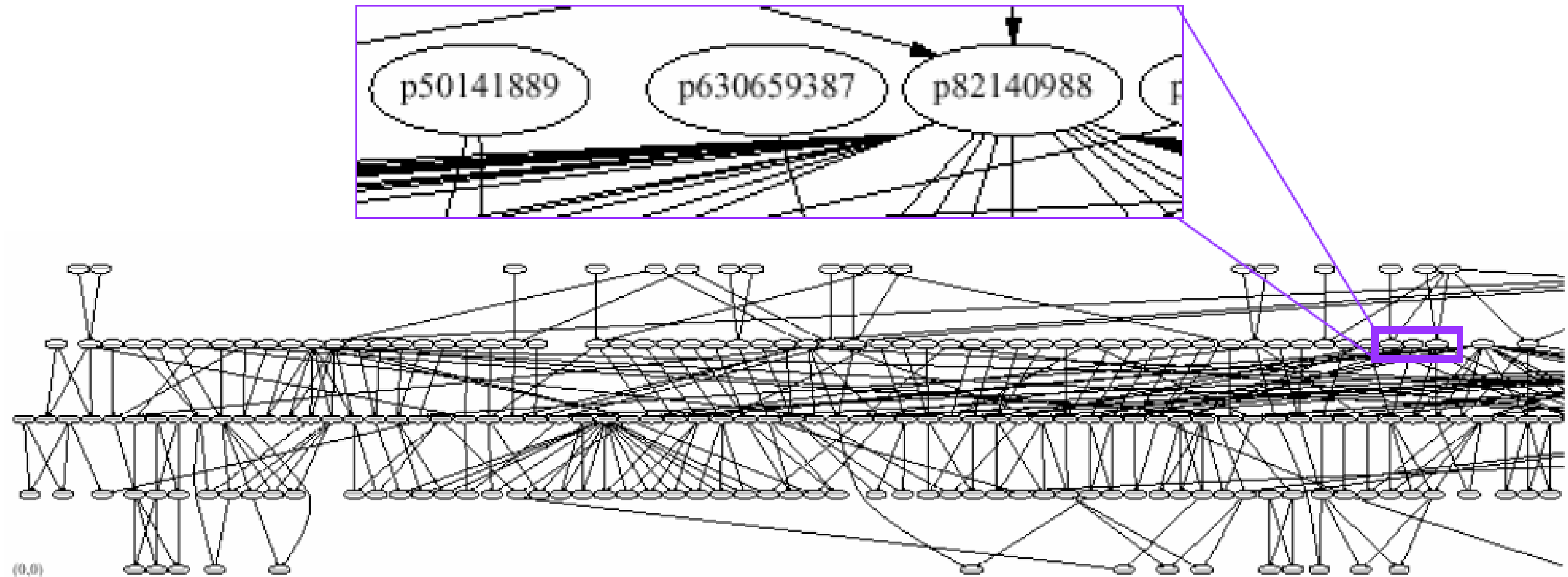
- Monitoring system for patients in intensive care




 Slide by C. Boutilier and P. Poupart, 2003 

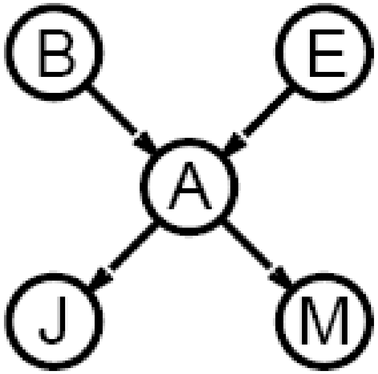
# Example: Pigs Network

- Determines pedigree of breeding pigs
  - used to diagnose PSE disease
  - half of the network structure shown here



# Compactness of a BN

A CPT for Boolean $X_i$ with $k$ Boolean parents has $2^k$ rows for the combinations of parent values




Each row requires one number $p$ for $X_i = true$ (the number for $X_i = false$ is just $1 - p$)

If each variable has no more than $k$ parents, the complete network requires $O(n \cdot 2^k)$ numbers

I.e., grows linearly with $n$, vs. $O(2^n)$ for the full joint distribution

For burglary net, $1 + 1 + 4 + 2 + 2 = 10$ numbers (vs. $2^5 - 1 = 31$)

# Compact Conditional Distributions

CPT grows exponentially with number of parents
CPT becomes infinite with continuous-valued parent or child

Solution: canonical distributions that are defined compactly

Deterministic nodes are the simplest case:

$X = f(Parents(X))$ for some function $f$

E.g., Boolean functions

$$NorthAmerican \Leftrightarrow Canadian \lor US \lor Mexican$$

E.g., numerical relationships among continuous variables

$$\frac{\partial Level}{\partial t} = \text{inflow + precipitation - outflow - evaporation}$$

# Compact Conditional Distributions Independent Causes

Noisy-OR distributions model multiple noninteracting causes

1) Parents $U_1 \ldots U_k$ include all causes (can add leak node)
2) Independent failure probability $q_i$ for each cause alone

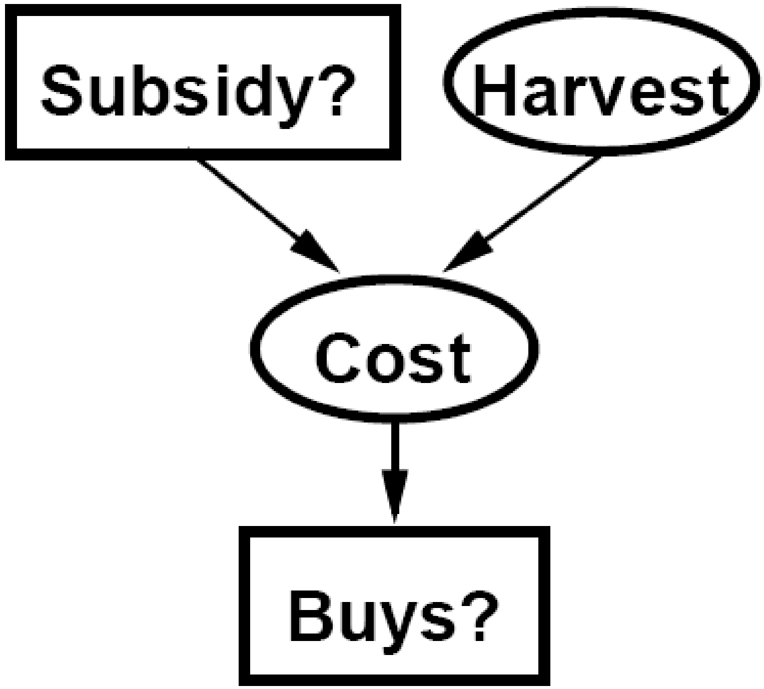
$$\Rightarrow P(X|U_1 \ldots U_j, \neg U_{j+1} \ldots \neg U_k) = 1 - \Pi_{i=1}^{j} q_i$$

| $Cold$ | $Flu$ | $Malaria$ | $P(Fever)$ | $P(\neg Fever)$ |
|---|---|---|---|---|
| F | F | F | **0.0** | 1.0 |
| F | F | T | 0.9 | **0.1** |
| F | T | F | 0.8 | **0.2** |
| F | T | T | 0.98 | $0.02 = 0.2 \times 0.1$ |
| T | F | F | 0.4 | **0.6** |
| T | F | T | 0.94 | $0.06 = 0.6 \times 0.1$ |
| T | T | F | 0.88 | $0.12 = 0.6 \times 0.2$ |
| T | T | T | 0.988 | $0.012 = 0.6 \times 0.2 \times 0.1$ |

Number of parameters **linear** in number of parents

# Hybrid Networks

Discrete (*Subsidy?* and *Buys?*); continuous (*Harvest* and *Cost*)




Option 1: discretization—possibly large errors, large CPTs
Option 2: finitely parameterized canonical families

1) Continuous variable, discrete+continuous parents (e.g., *Cost*)
2) Discrete variable, continuous parents (e.g., *Buys?*)

# Continuous Conditional Distributions

Need one conditional density function for child variable given continuous parents, for each possible assignment to discrete parents

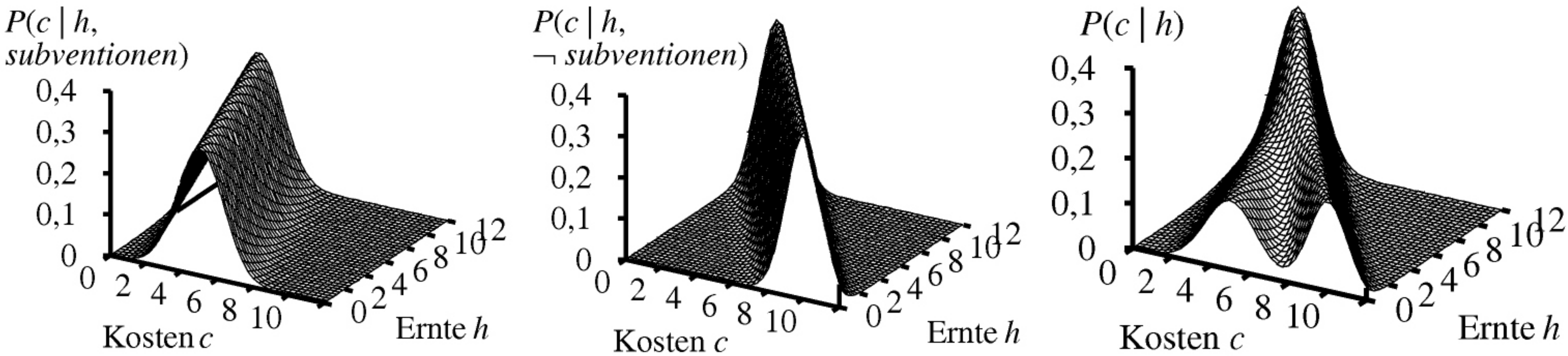
Most common is the linear Gaussian model, e.g.,:

$$
\begin{aligned}
& P(Cost = c | Harvest = h, Subsidy? = true) \\
= \; & N(a_t h + b_t, \sigma_t)(c) \\
= \; & \frac{1}{\sigma_t \sqrt{2\pi}} exp\left( -\frac{1}{2} \left( \frac{c - (a_t h + b_t)}{\sigma_t} \right)^2 \right)
\end{aligned}
$$

Mean $Cost$ varies linearly with $Harvest$, variance is fixed

Linear variation is unreasonable over the full range
but works OK if the **likely** range of $Harvest$ is narrow

# Continuous Conditional Distributions




$\mathbf{P}(Cost \mid Harvest, subsidy)$ $\mathbf{P}(Cost \mid Harvest, \neg subsidy)$ $\mathbf{P}(Cost \mid Harvest)$

$$= \mathbf{P}(Cost \mid Harvest, subsidy) + \mathbf{P}(Cost \mid Harvest, \neg subsidy)$$

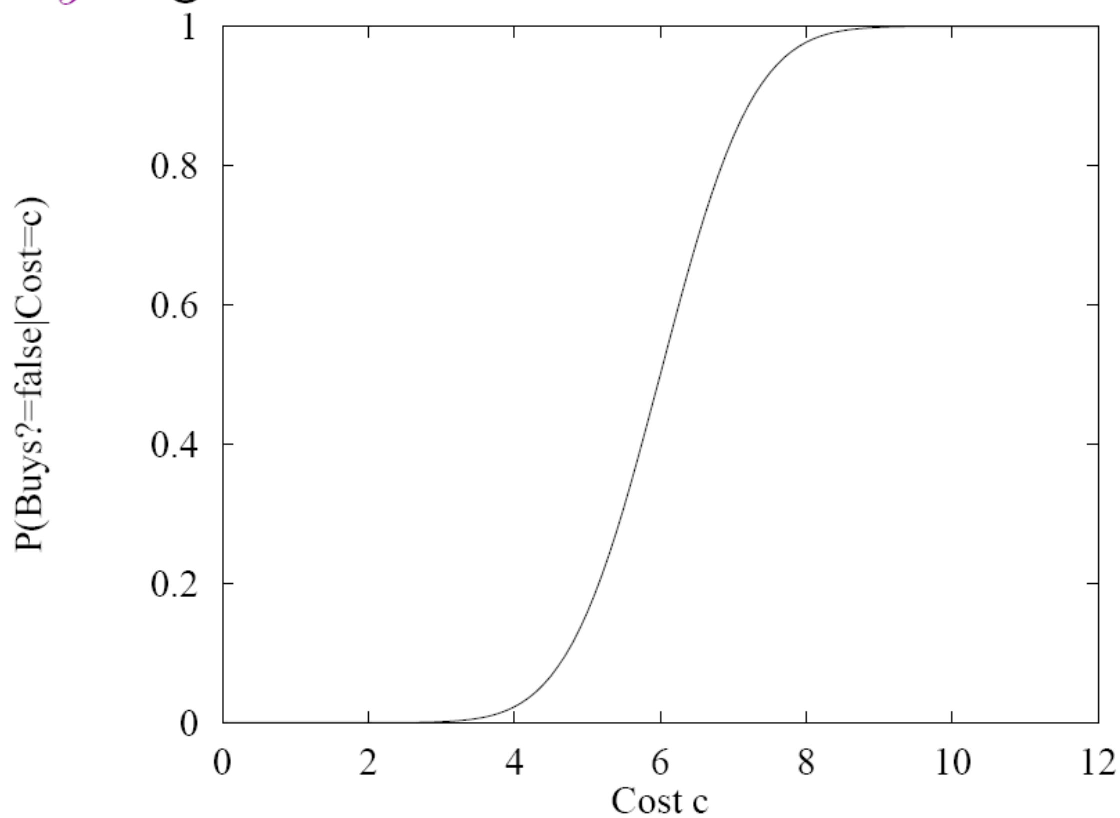
All-continuous network with LG distributions

$\Rightarrow$ full joint distribution is a multivariate Gaussian

Discrete+continuous LG network is a conditional Gaussian network i.e., a multivariate Gaussian over all continuous variables for each combination of discrete variable values

# Discrete Variables with Continuous Parents

Probability of $Buys?$ given $Cost$ should be a "soft" threshold:




Probit distribution uses integral of Gaussian:

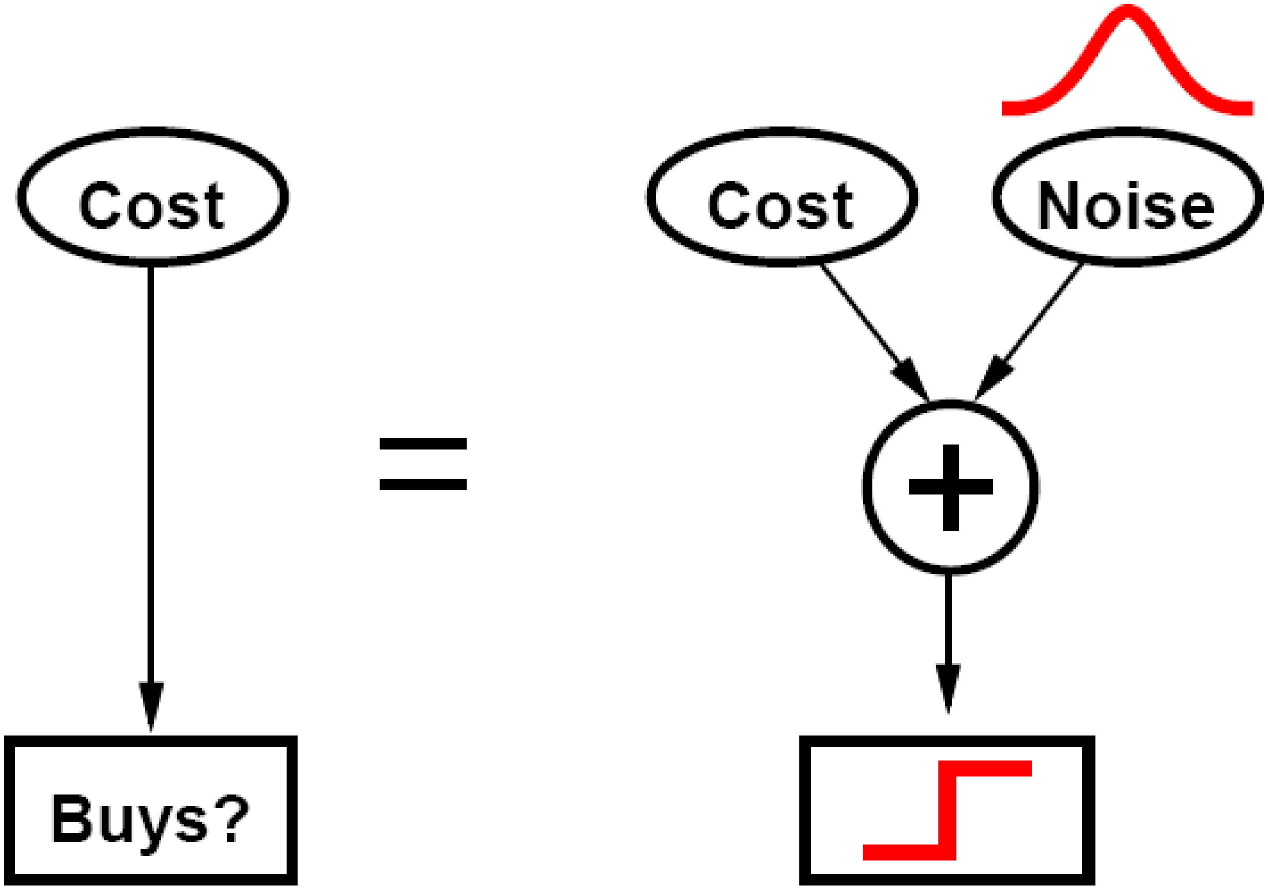
$\Phi(x) = \int_{-\infty}^{x} N(0,1)(x)dx$

$P(Buys? = true \mid Cost = c) = \Phi((-c + \mu)/\sigma)$

# Why Probit?

1. It's sort of the right shape

2. Can view as hard threshold whose location is subject to noise

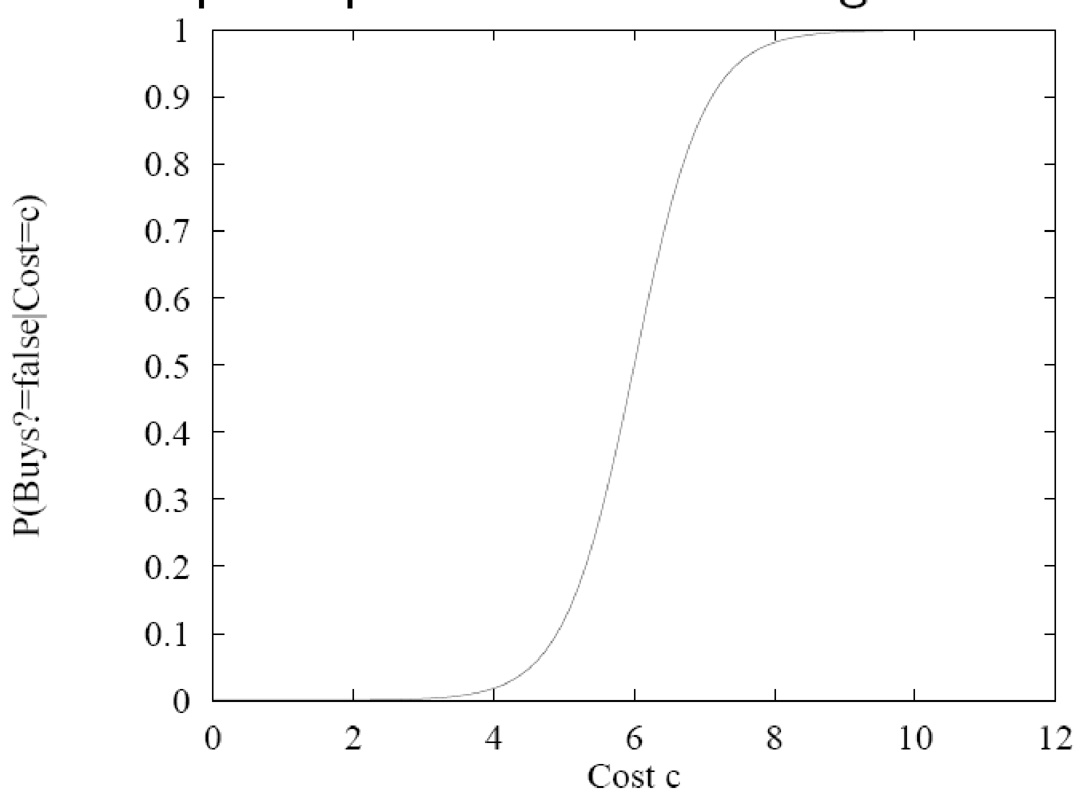


# Discrete Variables with Continuous Parents

Sigmoid (or logit) distribution also used in neural networks:

$$P(Buys? = true \mid Cost = c) = \frac{1}{1 + exp(-2\frac{-c+\mu}{\sigma})}$$

Sigmoid has similar shape to probit but much longer tails:



# Real-World Applications of BN

- Industrial
  - Processor Fault Diagnosis - by Intel
  - Auxiliary Turbine Diagnosis - GEMS by GE
  - Diagnosis of space shuttle propulsion systems - VISTA by NASA/Rockwell
  - Situation assessment for nuclear power plant – NRC
- Military
  - Automatic Target Recognition - MITRE
  - Autonomous control of unmanned underwater vehicle - Lockheed Martin
  - Assessment of Intent


Slide by Buede, Tatman, Bresnik, 1998

# Real-World Applications of BN

- Medical Diagnosis
  - Internal Medicine
  - Pathology diagnosis - Intellipath by Chapman & Hall
  - Breast Cancer Manager with Intellipath
- Commercial
  - Financial Market Analysis
  - Information Retrieval
  - Software troubleshooting and advice - Windows 95 & Office 97
  - Pregnancy and Child Care – Microsoft
  - Software debugging - American Airlines' SABRE online reservation system


Slide by Buede, Tatman, Bresnik, 1998