Reasoning under Uncertainty

Course: CS40022
Instructor: Dr. Pallab Dasgupta

Department of Computer Science & Engineering
Indian Institute of Technology Kharagpur

Handling uncertain knowledge

- $\forall p$ Symptom(p, Toothache) $\Rightarrow$ Disease(p, Cavity)
  - Not correct – toothache can be caused in many other cases
  - $\forall p$ Symptom(p, Toothache) $\Rightarrow$ Disease(p, Cavity) $\vee$ Disease(p, GumDisease) $\vee$ Disease(p, ImpactedWisdom) $\vee$ …

Handling uncertain knowledge

- $\forall p$ Disease(p, Cavity)
  $\Rightarrow$ Symptom(p, Toothache)
  - This is not correct either, since all cavities do not cause toothache

Reasons for using probability

- Specification becomes too large
  - It is too much work to list the complete set of antecedents or consequents needed to ensure an exception-less rule
- Theoretical ignorance
  - The complete set of antecedents is not known
- Practical ignorance
  - The truth of the antecedents is not known, but we still wish to reason

Axioms of Probability

1. All prob are between 0 and 1: $0 \leq P(A) \leq 1$
2. P(True) = 1 and P(False) = 0
3. $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

Bayes' Rule

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

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

$$P(B \mid A) = \frac{P(A \mid B) P(B)}{P(A)}$$

Belief Networks

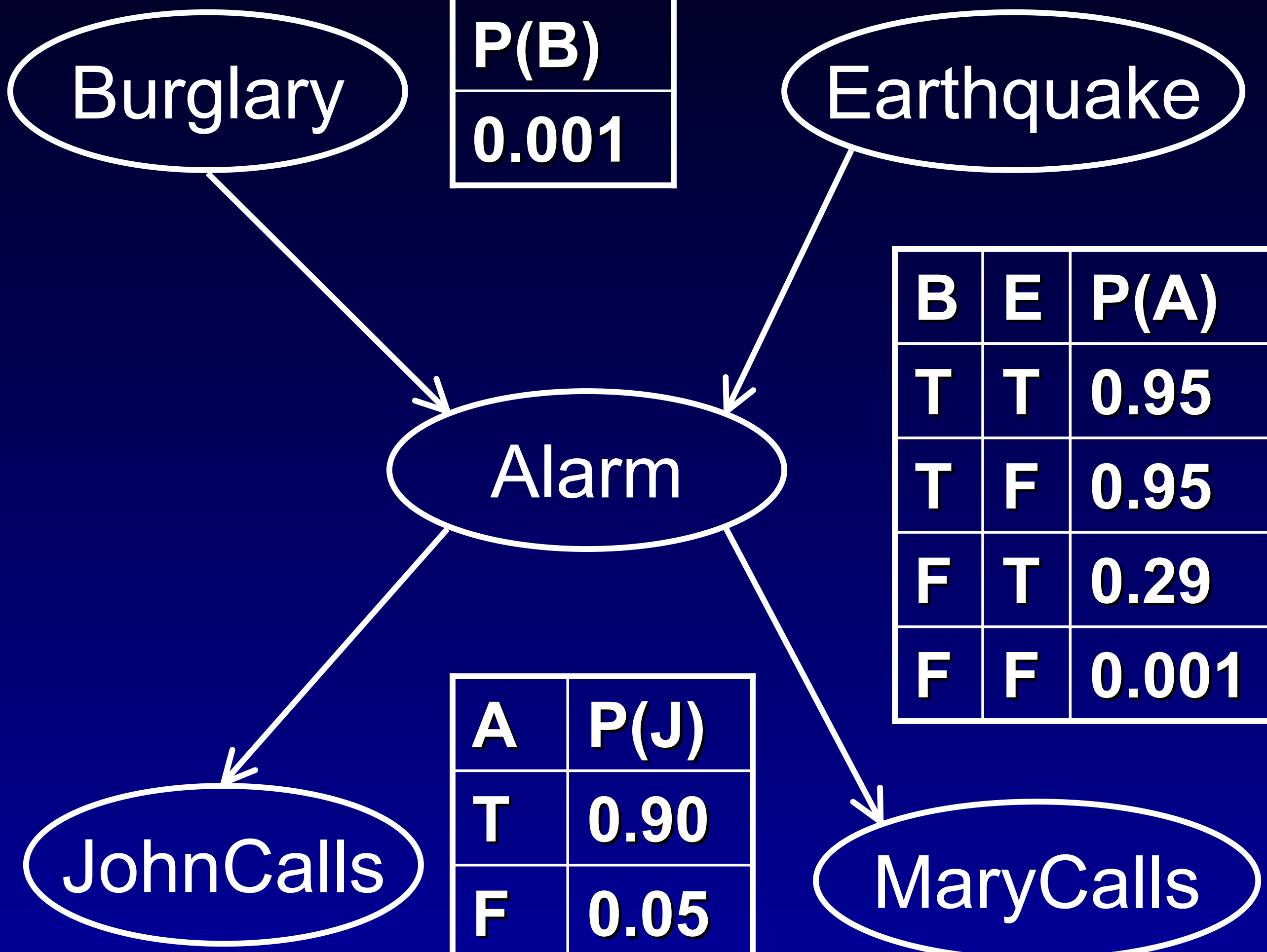
A belief network is a graph with the following:

1. Nodes: Set of random variables
2. Directed links: The intuitive meaning of a link from node X to node Y is that X has a direct influence on Y
3. Each node has a conditional probability table that quantifies the effects that the parent have on the node.
4. The graph has no directed cycles (DAG)

Example

- Burglar alarm at home
  - Fairly reliable at detecting a burglary
  - Responds at times to minor earthquakes
- Two neighbors, on hearing alarm, calls police
  - John always calls when he hears the alarm, but sometimes confuses the telephone ringing with the alarm and calls then, too.
  - Mary likes loud music and sometimes misses the alarm altogether

Belief Network Example



P(B)
0.001

P(E)
0.002

B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	P(J)
T	0.90
F	0.05

A	P(M)
T	0.70
F	0.01

The joint probability distribution

- A generic entry in the joint probability distribution $P(x_1, \ldots, x_n)$ is given by:

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

The joint probability distribution

- Probability of the event that the alarm has sounded but neither a burglary nor an earthquake has occurred, and both Mary and John call:

$$P(J \wedge M \wedge A \wedge \neg B \wedge \neg E)$$
$$= P(J \mid A)\ P(M \mid A)\ P(A \mid \neg B \wedge \neg E)\ P(\neg B)\ P(\neg E)$$
$$= 0.9 \text{ X } 0.7 \text{ X } 0.001 \text{ X } 0.999 \text{ X } 0.998$$
$$= 0.00062$$

Conditional independence

$$
\begin{aligned}
P(x_1,\ldots,x_n) &= P(x_n \mid x_{n-1},\ldots,x_1)P(x_{n-1},\ldots,x_1) \\
&= P(x_n \mid x_{n-1},\ldots,x_1)P(x_{n-1} \mid x_{n-2},\ldots,x_1) \\
&\qquad \ldots P(x_2 \mid x_1)P(x_1) \\
&= \prod_{i=1}^{n} P(x_i \mid x_{i-1},\ldots,x_1)
\end{aligned}
$$

- The belief network represents conditional independence:

$$P(X_i \mid X_i,\ldots,X_1) = P(X_i \mid Parents(X_i))$$

Incremental Network Construction

1. Choose the set of relevant variables X_i that describe the domain

2. Choose an ordering for the variables (*very important step*)

3. While there are variables left:

   a) Pick a variable X and add a node for it

   b) Set Parents(X) to some minimal set of existing nodes such that the conditional independence property is satisfied

   c) Define the conditional prob table for X

Conditional Independence Relations

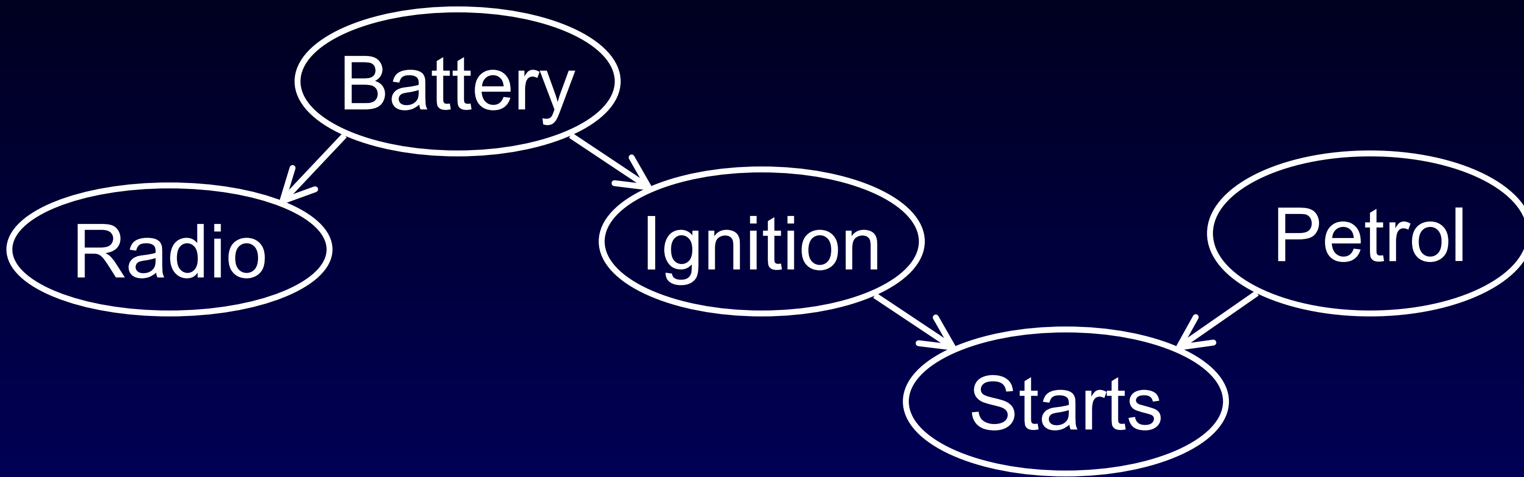
- *If every undirected path from a node in X to a node in Y is d-separated by a given set of evidence nodes E, then X and Y are conditionally independent given E.*

- A set of nodes E *d-separates* two sets of nodes X and Y if every undirected path from a node in X to a node in Y is *blocked* given E.

Conditional Independence Relations

- A path is blocked given a set of nodes E if there is a node Z on the path for which one of three conditions holds:
  1. Z is in E and Z has one arrow on the path leading in and one arrow out
  2. Z is in E and Z has both path arrows leading out
  3. Neither Z nor any descendant of Z is in E, and both path arrows lead in to Z

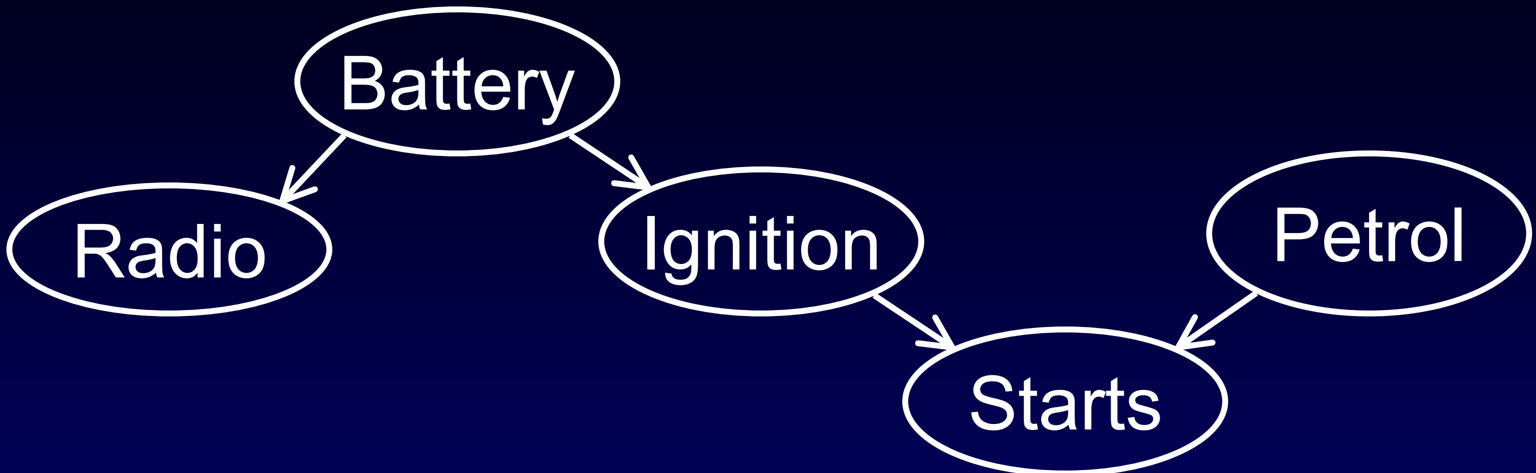
Cond Independence in belief networks



Whether there is petrol and whether the radio plays are independent given evidence about whether the ignition takes place

Petrol and Radio are independent if it is known whether the battery works

Cond Independence in belief networks



Petrol and Radio are independent given no evidence at all.

But they are dependent given evidence about whether the car starts.

If the car does not start, then the radio playing is increased evidence that we are out of petrol.

Inferences using belief networks

- Diagnostic inferences (from effects to causes)
  - Given that JohnCalls, infer that

    $P(\text{Burglary} \mid \text{JohnCalls}) = 0.016$

- Causal inferences (from causes to effects)
  - Given Burglary, infer that

    $P(\text{JohnCalls} \mid \text{Burglary}) = 0.86$ and

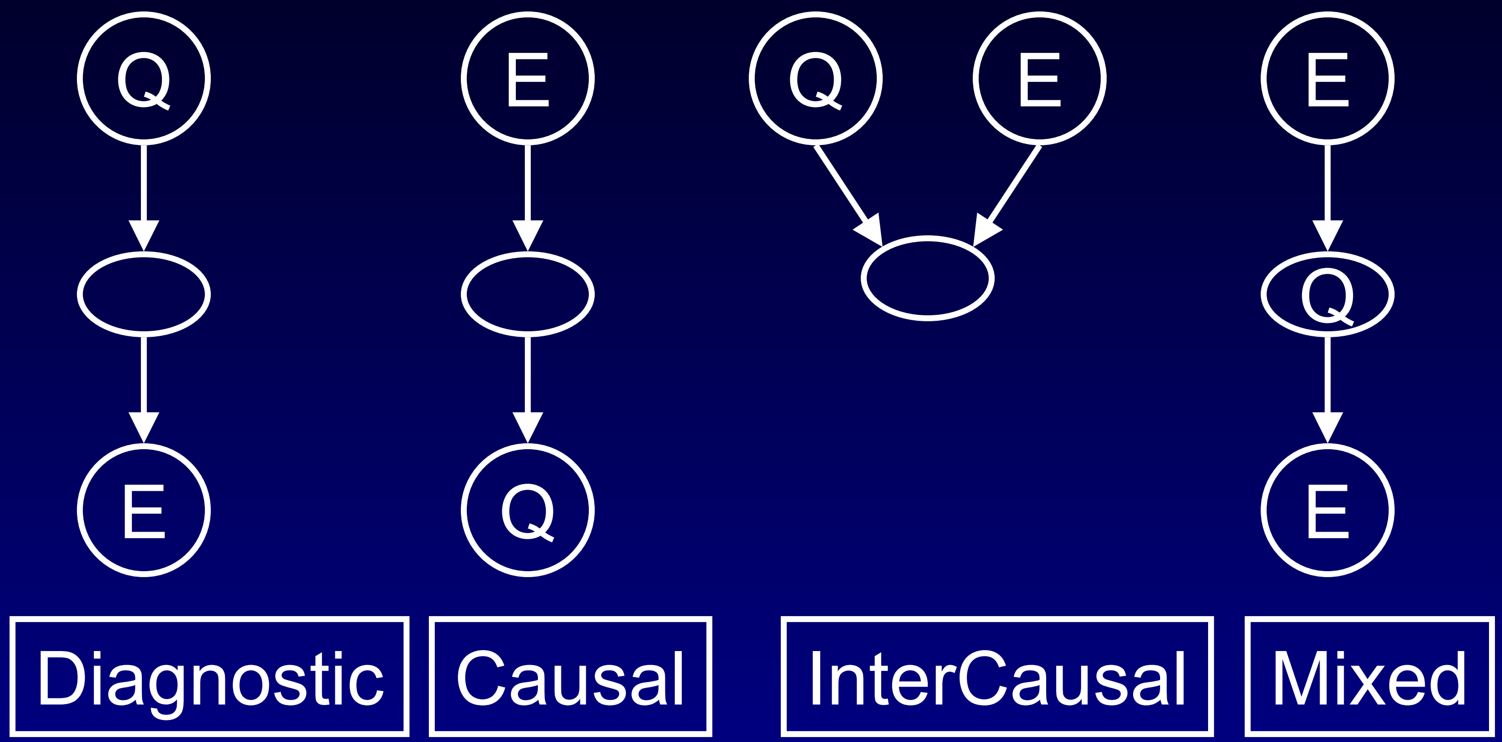
    $P(\text{MaryCalls} \mid \text{Burglary}) = 0.67$

Inferences using belief networks

- Intercausal inferences (between causes of a common effect)
  - Given Alarm, we have

    $P(Burglary \mid Alarm) = 0.376.$
  - If we add evidence that Earthquake is true, then $P(Burglary \mid Alarm \wedge Earthquake)$ goes down to 0.003
- Mixed inferences
  - Setting the effect JohnCalls to true and the cause Earthquake to false gives $P(Alarm \mid JohnCalls \wedge \neg Earthquake) = 0.003$

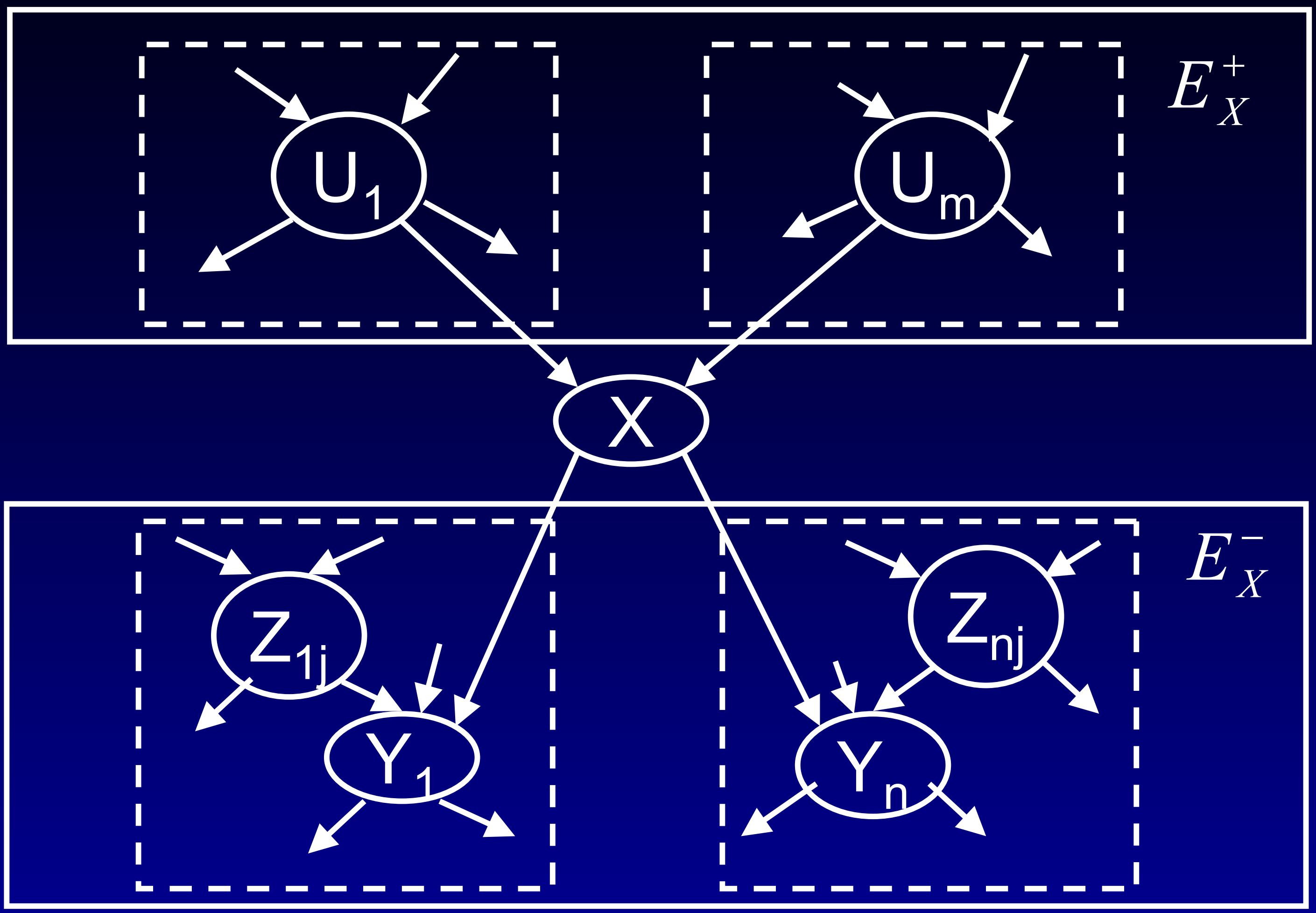
The four patterns



Answering queries

- We consider cases where the belief network is a poly-tree
  - There is at most one undirected path between any two nodes

Answering queries



Answering queries

- $U = U_1 \dots U_m$ are parents of node X
- $Y = Y_1 \dots Y_n$ are children of node X
- X is the query variable
- E is a set of evidence variables
- The aim is to compute $P(X \mid E)$

Definitions

- E_X^+ is the causal support for X
  - The evidence variables "above" X that are connected to X through its parents
- E_X^- is the evidential support for X
  - The evidence variables "below" X that are connected to X through its children
- $E_{U_i \setminus X}$ refers to all the evidence connected to node U_i except via the path from X
- $E_{Y_i \setminus X}^+$ refers to all the evidence connected to node Y_i through its parents for X

The computation of P(X|E)

$$P(X \mid E) = P(X \mid E_X^-, E_X^+)$$

$$= \frac{P(E_X^- \mid X, E_X^+)\, P(X \mid E_X^+)}{P(E_X^- \mid E_X^+)}$$

- Since X d-separates E_X^+ from E_X^-, we can use conditional independence to simplify the first term in the numerator
- We can treat the denominator as a constant

$$P(X \mid E) = \alpha\, P(E_X^- \mid X) P(X \mid E_X^+)$$

The computation of $P(X \mid E_X^+)$

We consider all possible configurations of the parents of X and how likely they are given E_X^+.

Let U be the vector of parents $U_1, \ldots, U_m$, and let u be an assignment of values to them.

$$P(X \mid E_X^+) = \sum_u P(X \mid u, E_X^+) P(u \mid E_X^+)$$

The computation of $P(X \mid E_X^+)$

$$P(X \mid E_X^+) = \sum_u P(X \mid u, E_X^+) P(u \mid E_X^+)$$

U d-separates X from E_X^+, so the first term simplifies to $P(X \mid u)$

We can simplify the second term by noting

- E_X^+ d-separates each U_i from the others,
- the probability of a conjunction of independent variables is equal to the product of their individual probabilities

$$P(X \mid E_X^+) = \sum_u P(X \mid u) \prod_i P(u_i \mid E_X^+)$$

The computation of $P(X \mid E_X^+)$

$$P(X \mid E_X^+) = \sum_u P(X \mid u) \prod_i P(u_i \mid E_X^+)$$

The last term can be simplified by partitioning E_X^+ into $E_{U1\backslash X}, \ldots, E_{Um\backslash X}$ and noting that $E_{Ui\backslash X}$ d-separates U_i from all the other evidence in E_X^+

$$P(X \mid E_X^+) = \sum_u P(X \mid u) \prod_i P(u_i \mid E_{Ui\backslash X})$$

- $P(X \mid u)$ is a lookup in the cond prob table of X
- $P(u_i \mid E_{Ui\backslash X})$ is a recursive (smaller) sub-problem

The computation of $P(E_X^- \mid X)$

Let Z_i be the parents of Y_i other than X, and let z_i be an assignment of values to the parents

– The evidence in each Y_i box is conditionally independent of the others given X

$$P(E_X^- \mid X) = \prod_i P(E_{Y_i \setminus X} \mid X)$$

The computation of $P(E_X^- \mid X)$

$$P(E_X^- \mid X) = \prod_i P(E_{Y_i \setminus X} \mid X)$$

Averaging over Y_i and z_i yields:

$$P(E_X^- \mid X) = \prod_i \sum_{y_i} \sum_{z_i} P(E_{Y_i \setminus X} \mid X, y_i, z_i) P(y_i, z_i \mid X)$$

The computation of $P(E_X^- \mid X)$

$$P(E_X^- \mid X) = \prod_i \sum_{y_i} \sum_{z_i} P(E_{Yi\backslash X} \mid X, y_i, z_i) P(y_i, z_i \mid X)$$

Breaking $E_{Yi\backslash X}$ into the two independent components E_{Yi}^- and $E_{Yi\backslash X}^+$

$$P(E_X^- \mid X) = \prod_i \sum_{y_i} \sum_{z_i} P(E_{Yi}^- \mid X, y_i, z_i)$$

$$P(E_{Yi\backslash X}^+ \mid X, y_i, z_i) P(y_i, z_i \mid X)$$

The computation of $P(E_X^- \mid X)$

$$P(E_X^- \mid X) = \prod_i \sum_{y_i} \sum_{z_i} P(E_{Yi}^- \mid X, y_i, z_i)$$

$$P(E_{Yi\backslash X}^+ \mid X, y_i, z_i)\, P(y_i, z_i \mid X)$$

E_{Yi}^- is independent of X and z_i given y_i, and $E_{Yi\backslash X}^+$ is independent of X and y_i

$$P(E_X^- \mid X) = \prod_i \sum_{y_i} P(E_{Yi}^- \mid y_i) \sum_{z_i} P(E_{Yi\backslash X}^+ \mid z_i) P(y_i, z_i \mid X)$$

The computation of $P(E_X^- \mid X)$

$$P(E_X^- \mid X) = \prod_i \sum_{y_i} P(E_{Yi}^- \mid y_i) \sum_{z_i} P(E_{Yi\backslash X}^+ \mid z_i) P(y_i, z_i \mid X)$$

Apply Bayes' rule to $P(E_{Yi\backslash X}^+ \mid z_i)$:

$$P(E_X^- \mid X) =$$

$$\prod_i \sum_{y_i} P(E_{Yi}^- \mid y_i) \sum_{z_i} \frac{P(z_i \mid E_{Yi\backslash X}^+) P(E_{Yi\backslash X}^+)}{P(z_i)} P(y_i, z_i \mid X)$$

The computation of $P(E_X^- \mid X)$

$$P(E_X^- \mid X) =$$

$$\prod_i \sum_{y_i} P(E_{Y_i}^- \mid y_i) \sum_{z_i} \frac{P(z_i \mid E_{Y_i \backslash X}^+) P(E_{Y_i \backslash X}^+)}{P(z_i)} P(y_i, z_i \mid X)$$

- Rewriting the conjunction of Y_i and z_i:

$$P(E_X^- \mid X) = \prod_i \sum_{y_i} P(E_{Y_i}^- \mid y_i)$$

$$\sum_{z_i} \frac{P(z_i \mid E_{Y_i \backslash X}^+) P(E_{Y_i \backslash X}^+)}{P(z_i)} P(y_i \mid X, z_i) P(z_i \mid X)$$

The computation of $P(E_X^- \mid X)$

$$P(E_X^- \mid X) = \prod_i \sum_{y_i} P(E_{Y_i}^- \mid y_i)$$

$$\sum_{z_i} \frac{P(z_i \mid E_{Y_i \backslash X}^+) P(E_{Y_i \backslash X}^+)}{P(z_i)} P(y_i \mid X, z_i) P(z_i \mid X)$$

$P(z_i \mid X) = P(z_i)$ because Z and X are d-separated. Also $P(E_{Y_i \backslash X}^+)$ is a constant

$$P(E_X^- \mid X) =$$

$$\prod_i \sum_{y_i} P(E_{Y_i}^- \mid y_i) \sum_{z_i} \beta_i P(z_i \mid E_{Y_i \backslash X}^+) P(y_i \mid X, z_i)$$

The computation of $P(E_X^- \mid X)$

$$P(E_X^- \mid X) =$$

$$\prod_i \sum_{y_i} P(E_{Yi}^- \mid y_i) \sum_{z_i} \beta_i P(z_i \mid E_{Yi\backslash X}^+) P(y_i \mid X, z_i)$$

- The parents of Y_i (the Z_{ij}) are independent of each other.
- We also combine the β_i into one single β

The computation of $P(E_X^- \mid X)$

$$P(E_X^- \mid X) =$$

$$\beta \prod_i \sum_{y_i} P(E_{Y_i}^- \mid y_i) \sum_{z_i} P(y_i \mid X, z_i) \prod_j P(z_{ij} \mid E_{Z_{ij} \backslash Y_i})$$

- $P(E_{Yi}^- \mid y_i)$ is a recursive instance of $P(E_X^- \mid X)$
- $P(y_i \mid X, z_i)$ is a cond prob table entry for Y_i
- $P(z_{ij} \mid E_{Zij \backslash Yi})$ is a recursive sub-instance of the $P(X \mid E)$ calculation

Inference in multiply connected belief networks

- Clustering methods
  - Transform the net into a probabilistically equivalent (but topologically different) poly-tree by merging offending nodes
- Conditioning methods
  - Instantiate variables to definite values, and then evaluate a poly-tree for each possible instantiation

Inference in multiply connected belief networks

- Stochastic simulation methods
  - Use the network to generate a large number of concrete models of the domain that are consistent with the network distribution.
  - They give an approximation of the exact evaluation.

Default reasoning

- Some conclusions are made by default unless a counter-evidence is obtained
  - Non-monotonic reasoning
- Points to ponder
  - Whats the semantic status of default rules?
  - What happens when the evidence matches the premises of two default rules with conflicting conclusions?
  - If a belief is retracted later, how can a system keep track of which conclusions need to be retracted as a consequence?

Issues in Rule-based methods for Uncertain Reasoning

- Locality
  - In logical reasoning systems, if we have $A \Rightarrow B$, then we can conclude B given evidence A, *without worrying about any other rules*. In probabilistic systems, we need to consider *all* available evidence.

Issues in Rule-based methods for Uncertain Reasoning

- Detachment
  - Once a logical proof is found for proposition B, we can use it regardless of how it was derived (*it can be detached from its justification*). In probabilistic reasoning, the source of the evidence is important for subsequent reasoning.

Issues in Rule-based methods for Uncertain Reasoning

- Truth functionality
  - In logic, the truth of complex sentences can be computed from the truth of the components. Probability combination does not work this way, except under strong independence assumptions.

A famous example of a truth functional system for uncertain reasoning is the *certainty factors model*, developed for the Mycin medical diagnostic program

Dempster-Shafer Theory

- Designed to deal with the distinction between *uncertainty* and *ignorance*.
- We use a belief function *Bel(X)* – probability that the evidence supports the proposition
- When we do not have any evidence about X, we assign Bel(X) = 0 as well as Bel($\neg$X) = 0

Dempster-Shafer Theory

For example, if we do not know whether a coin is fair, then:

$\text{Bel}(\text{ Heads }) = \text{Bel}(\neg\text{Heads }) = 0$

If we are given that the coin is fair with 90% certainty, then:

$\text{Bel}(\text{ Heads }) = 0.9 \text{ X } 0.5 = 0.45$

$\text{Bel}(\neg\text{Heads }) = 0.9 \text{ X } 0.5 = 0.45$

Note that we still have a gap of 0.1 *that is not accounted for by the evidence*

Fuzzy Logic

- Fuzzy set theory is a means of specifying how well an object satisfies a vague description
  - Truth is a value between 0 and 1
  - Uncertainty stems from lack of evidence, but given the dimensions of a man concluding whether he is fat has no uncertainty involved

Fuzzy Logic

- The rules for evaluating the fuzzy truth, T, of a complex sentence are

$$T(A \wedge B) = \min(\ T(A), T(B)\ )$$

$$T(A \vee B) = \max(\ T(A), T(B)\ )$$

$$T(\neg A) = 1 - T(A)$$