

# **Reasoning under Uncertainty**

*The intelligent way to handle the unknown*

**COURSE: CS60045**

Pallab Dasgupta  
Professor,  
Dept. of Computer Sc & Engg



**INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR**

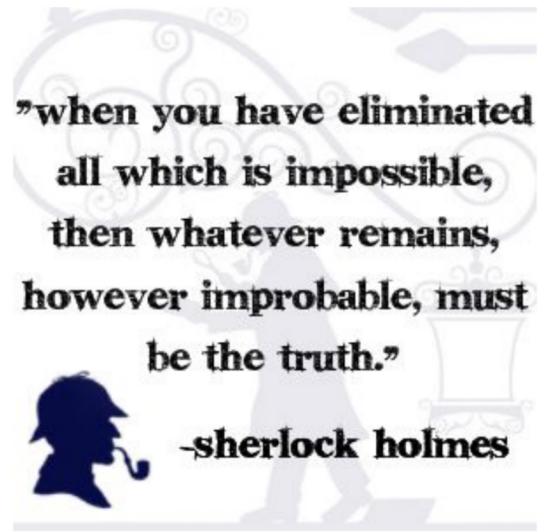
# Logical Deduction versus Induction

## DEDUCTION

- Commonly associated with *formal logic*
- Involves reasoning from known premises to a conclusion
- The conclusions reached are inevitable, certain, inescapable

## INDUCTION

- Commonly known as *informal logic* or *everyday argument*
- Involves drawing uncertain inferences based on probabilistic reasoning
- The conclusions reached are probable, reasonable, plausible, believable



**"when you have eliminated  
all which is impossible,  
then whatever remains,  
however improbable, must  
be the truth."**

**-sherlock holmes**

## Handling uncertain knowledge

- Classical first order logic has no room for uncertainty

$\forall p \text{ Symptom}(p, \text{Toothache}) \Rightarrow \text{Disease}(p, \text{Cavity})$

- Not correct – toothache can be caused in many other cases
- In first order logic we have to include all possible causes

$\forall p \text{ Symptom}(p, \text{Toothache}) \Rightarrow \text{Disease}(p, \text{Cavity}) \vee \text{Disease}(p, \text{GumDisease})$   
 $\quad \quad \quad \vee \text{Disease}(p, \text{ImpactedWisdom}) \vee \dots$

- Similarly, Cavity does not always cause Toothache, so the following is also not true

$\forall p \text{ Disease}(p, \text{Cavity}) \Rightarrow \text{Symptom}(p, \text{Toothache})$

## 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

## Predicting versus Diagnosing

- Probabilistic reasoning can be used for predicting outcomes (*from cause to effect*)
  - Given that I have a cavity, what is the chance that I will have toothache?
- Probabilistic reasoning can also be used for diagnosis (*from effect to cause*)
  - Given that I am having toothache, what is the chance that it is being caused by a cavity?

We need a methodology for reasoning that can work both ways.

## Axioms of Probability

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

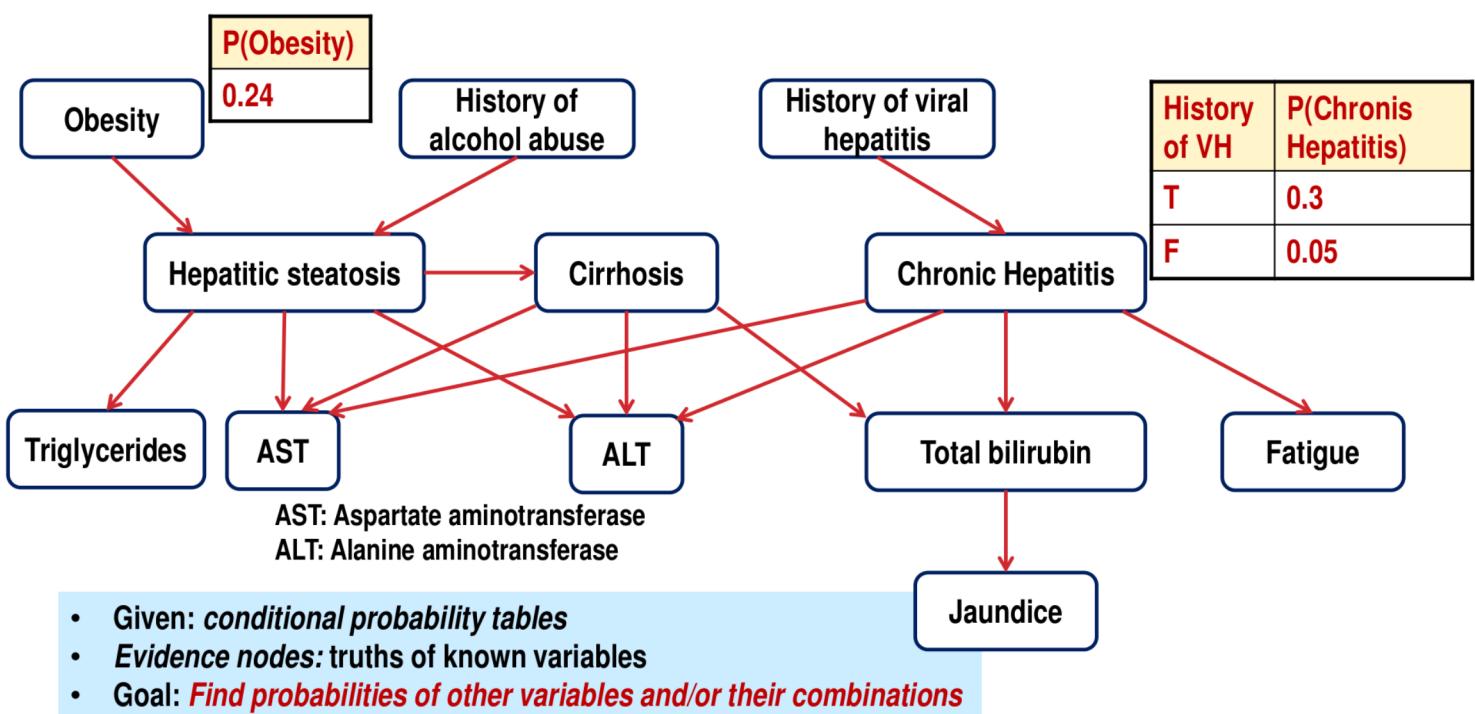
### Bayes' Rule

$$P(A \wedge B) = P(A | B) P(B)$$

$$P(A \wedge B) = P(B | A) P(A)$$

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

## Bayesian Belief Network



## Belief Networks

A belief network is a graph with the following:

- **Nodes:** Set of random variables
- **Directed links:** The intuitive meaning of a link from node X to node Y is that X has a direct influence on Y

Each node has a **conditional probability table** that quantifies the effects that the parent have on the node.

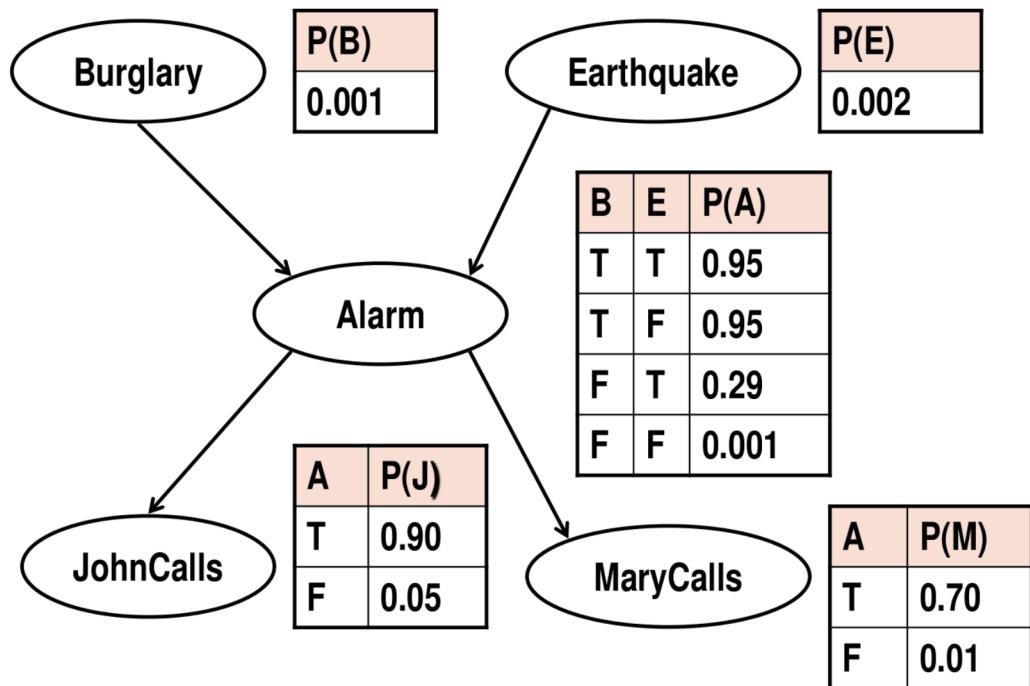
The graph has no directed cycles. It is a *directed acyclic graph* (DAG).

## Classical 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



## The joint probability distribution

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

$$P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i | Parents(X_i))$$



## 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:

$$\begin{aligned}
 & P(J \wedge M \wedge A \wedge \neg B \wedge \neg E) \\
 & = P(J | A) \ P(M | A) \ P(A | \neg B \wedge \neg E) \ P(\neg B) \ P(\neg E) \\
 & = 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 \\
 & = 0.00062
 \end{aligned}$$

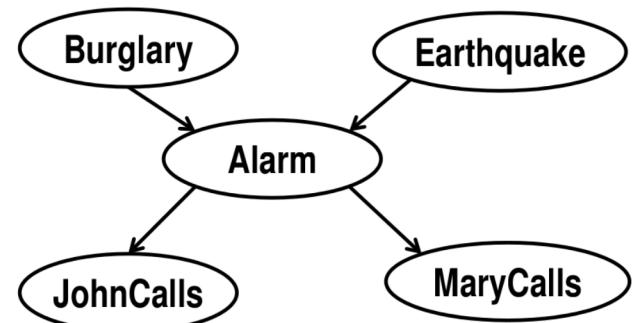
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	P(J)
T	0.90
F	0.05

A	P(M)
T	0.70
F	0.01

P(E)	P(B)
0.002	
	0.001



## The joint probability distribution

- Computation of the probabilities of several different event combinations of the Burglary-Alarm belief network example:

$$P(B) = 0.001$$

$$P(B') = 1 - P(B) = 0.999$$

$$P(E) = 0.002$$

$$P(E') = 1 - P(E) = 0.998$$

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

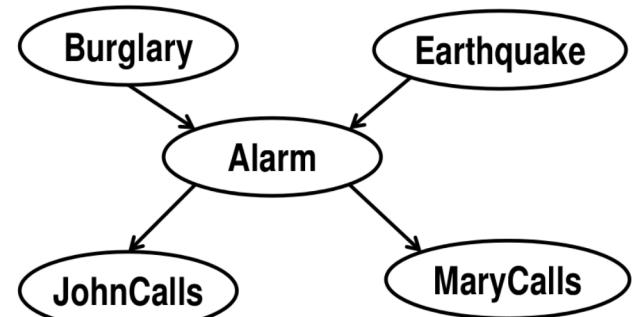
A	P(M)
T	0.70
F	0.01

P(E)
0.002

P(B)
0.001



## The joint probability distribution

- Computation of the probabilities of several different event combinations of the Burglary-Alarm belief network example:

$$\begin{aligned}
 P(A) &= P(AB'E') + P(AB'E) + P(ABE') + P(ABE) \\
 &= P(A | B'E').P(B'E') + P(A | B'E).P(B'E) + P(A | BE').P(BE') + P(A | BE).P(BE) \\
 &= 0.001 \times 0.999 \times 0.998 + 0.29 \times 0.999 \times 0.002 + 0.95 \times 0.001 \times 0.998 + 0.95 \times 0.001 \times 0.002 \\
 &= 0.001 + 0.0006 + 0.0009 = 0.0025
 \end{aligned}$$

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

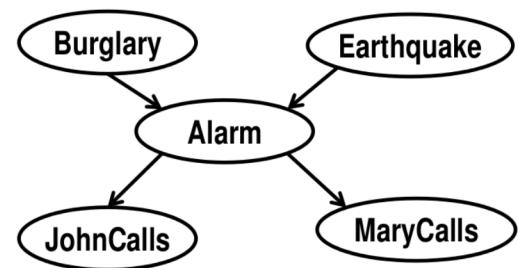
A	P(J)
T	0.90
F	0.05

A	P(M)
T	0.70
F	0.01

P(E)	P(B)
0.002	0.001



## The joint probability distribution: *Find P(J)*

$$\begin{aligned}
 P(J) &= P(JA) + P(JA') \\
 &= P(J | A).P(A) + P(J | A').P(A') \\
 &= 0.9 \times 0.0025 + 0.05 \times (1 - 0.0025) \\
 &= 0.052125
 \end{aligned}$$

$$\begin{aligned}
 P(AB) &= P(ABE) + P(ABE') = 0.95 \times 0.001 \times 0.002 + 0.95 \times 0.001 \times 0.998 \\
 &= 0.00095
 \end{aligned}$$

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

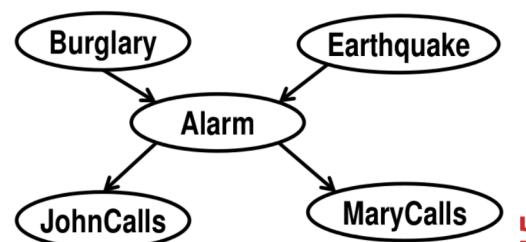
A	P(J)
T	0.90
F	0.05

A	P(M)
T	0.70
F	0.01

P(E)	P(B)
0.002	0.001



## The joint probability distribution: *Find P(A'B) and P(AE)*

$$\begin{aligned}
 P(A'B) &= P(A'BE) + P(A'BE') \\
 &= P(A' | BE).P(BE) + P(A' | BE').P(BE') \\
 &= (1 - 0.95) \times 0.001 \times 0.002 \\
 &\quad + (1 - 0.95) \times 0.001 \times 0.998 \\
 &= 0.00005
 \end{aligned}$$

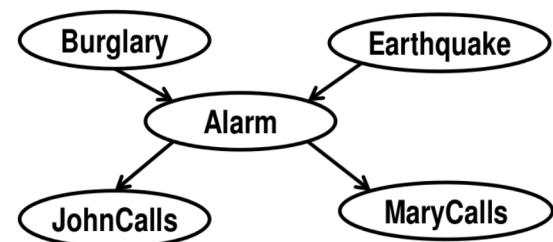
$$\begin{aligned}
 P(AE) &= P(AEB) + P(AEB') \\
 &= 0.95 \times 0.001 \times 0.002 + 0.29 \times 0.999 \times 0.002 = 0.00058
 \end{aligned}$$

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

P(E)	P(B)
0.002	
	0.001



## The joint probability distribution

$$\begin{aligned} P(AE') &= P(AE'B) + P(AE'B') \\ &= 0.95 \times 0.001 \times 0.998 + 0.001 \times 0.999 \times 0.998 \\ &= 0.001945 \end{aligned}$$

$$\begin{aligned} P(A'E') &= P(A'E'B) + P(A'E'B') \\ &= P(A' | BE').P(BE') + P(A' | B'E').P(B'E') \\ &= (1 - 0.95) \times 0.001 \times 0.998 + (1 - 0.001) \times 0.999 \times 0.998 = 0.996 \end{aligned}$$

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

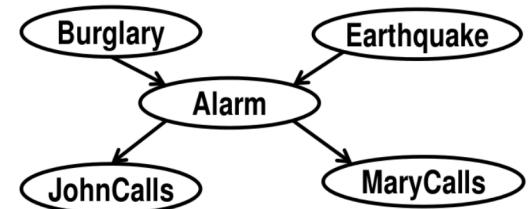
A	P(M)
T	0.70
F	0.01

P(E)
0.002

P(B)
0.001



## The joint probability distribution: *Find P(JB)*

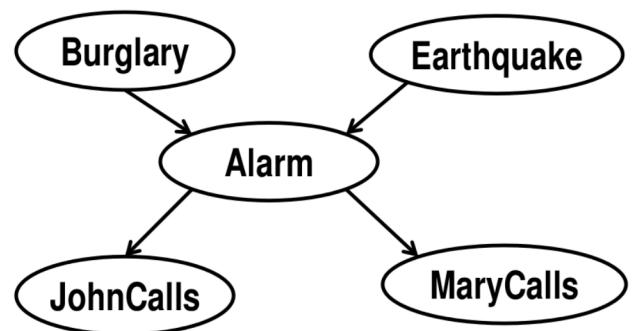
$$\begin{aligned}
 P(JB) &= P(JBA) + P(JBA') \\
 &= P(J | AB).P(AB) + P(J | A'B).P(A'B) \\
 &= P(J | A).P(AB) + P(J | A').P(A'B) \\
 &= 0.9 \times 0.00095 + 0.05 \times 0.00005 \\
 &= 0.00086
 \end{aligned}$$

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

P(E)	P(B)
0.002	
	0.001



## The joint probability distribution

- Computation of the probabilities of several different event combinations of the Burglary-Alarm belief network example:

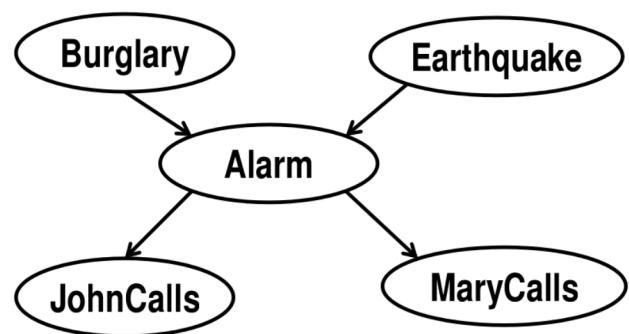
$$P(J | B) = P(JB) / P(B) = 0.00086 / 0.001 = 0.86$$

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

P(E)	P(B)
0.002	0.001



## The joint probability distribution

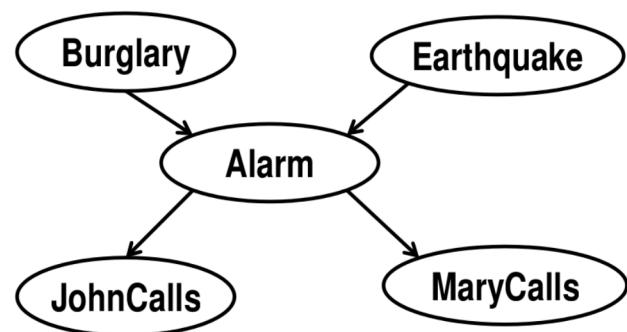
$$\begin{aligned}
 P(MB) &= P(MBA) + P(MBA') \\
 &= P(M | AB).P(AB) + P(M | A'B).P(A'B) \\
 &= P(M | A).P(AB) + P(M | A').P(A'B) \\
 &= 0.7 \times 0.00095 + 0.01 \times 0.00005 \\
 &= 0.00067
 \end{aligned}$$

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

P(E)	P(B)
T	0.002
F	0.001



## The joint probability distribution

$$P(M | B) = P(MB) / P(B) = 0.00067 / 0.001 = 0.67$$

$$P(B | J) = P(JB) / P(J) = 0.00086 / 0.052125 = 0.016$$

$$P(B | A) = P(AB) / P(A) = 0.00095 / 0.0025 = 0.38$$

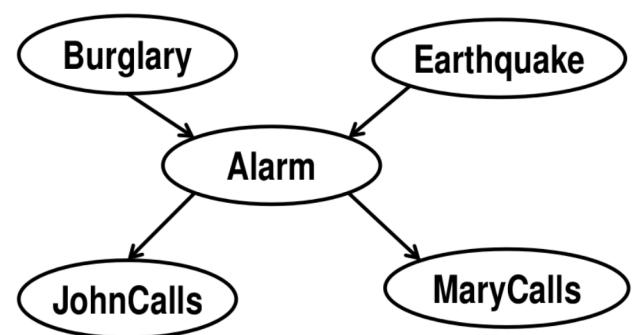
$$\begin{aligned}P(B | AE) &= P(ABE) / P(AE) = [ P(A | BE).P(BE) ] / P(AE) \\&= [ 0.95 \times 0.001 \times 0.002 ] / 0.00058 \\&= 0.003\end{aligned}$$

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

P(E)	P(B)
0.002	0.001



## The joint probability distribution

- Computation of the probabilities of several different event combinations of the Burglary-Alarm belief network example:

$$\begin{aligned} P(AJE') &= P(J | AE').P(AE') = P(J | A).P(AE') \\ &= 0.9 \times 0.001945 = 0.00175 \end{aligned}$$

$$\begin{aligned} P(A'JE') &= P(J | A'E').P(A'E') = P(J | A').P(A'E') \\ &= 0.05 \times 0.996 = 0.0498 \end{aligned}$$

$$P(JE') = P(AJE') + P(A'JE') = 0.00175 + 0.0498 = 0.05155$$

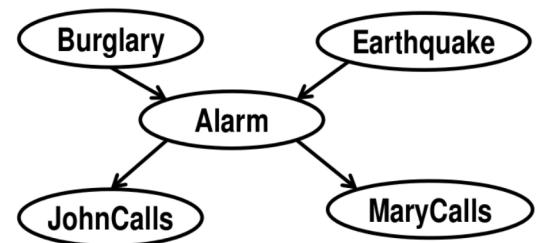
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	P(J)
T	0.90
F	0.05

A	P(M)
T	0.70
F	0.01

P(E)	P(B)
0.002	
	0.001



## The joint probability distribution

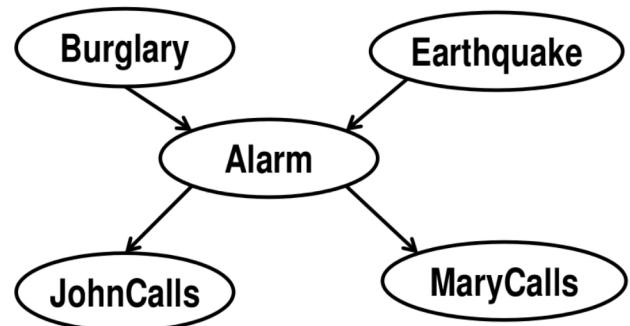
$$P(A | JE') = P(AJE') / P(JE') = 0.00175 / 0.05155 = 0.03$$

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

P(E)	P(B)
0.002	0.001



## The joint probability distribution

$$\begin{aligned}
 P(BJE') &= P(BJE'A) + P(BJE'A') \\
 &= P(J | ABE').P(ABE') + P(J | A'BE').P(A'BE') \\
 &= P(J | A).P(ABE') + P(J | A').P(A'BE') \\
 &= 0.9 \times 0.95 \times 0.001 \times 0.998 + 0.05 \times (1 - 0.95) \times 0.001 \times 0.998 \\
 &= 0.000856
 \end{aligned}$$

$$P(B | JE') = P(BJE') / P(JE') = 0.000856 / 0.05155 = 0.017$$

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

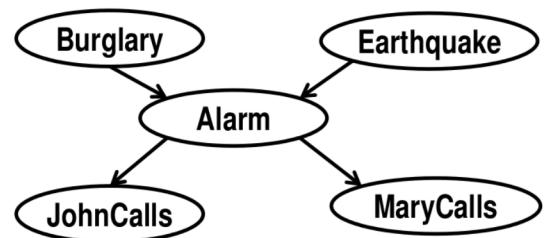
A	P(J)
T	0.90
F	0.05

A	P(M)
T	0.70
F	0.01

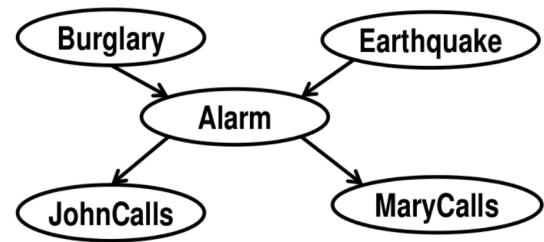
  

P(E)	P(B)
0.002	
	0.001



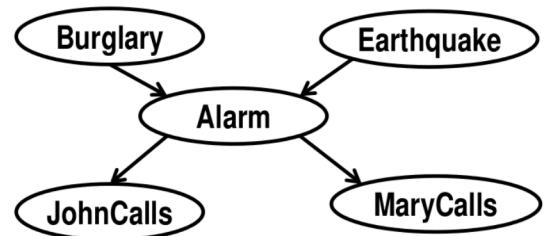
## Inferences using belief networks

- 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
$$P(\text{JohnCalls} \mid \text{Burglary}) = 0.86$$
$$P(\text{MaryCalls} \mid \text{Burglary}) = 0.67$$

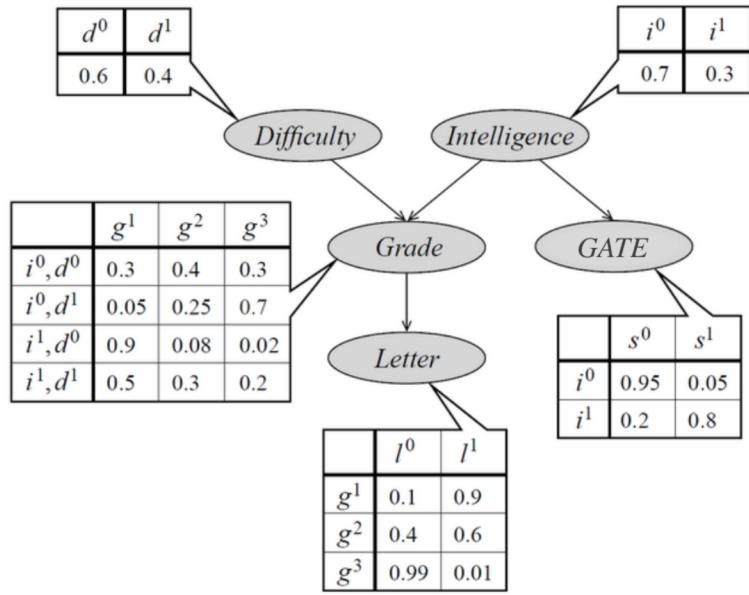


## Inferences using belief networks

- Inter-causal inferences (between causes of a common effect)
  - Given Alarm, we have  $P(\text{Burglary} \mid \text{Alarm}) = 0.376$
  - If we add evidence that Earthquake is true, then  $P(\text{Burglary} \mid \text{Alarm} \wedge \text{Earthquake}) = 0.003$
- Mixed inferences
  - Setting the effect JohnCalls to true and the cause Earthquake to false gives  
 $P(\text{Alarm} \mid \text{JohnCalls} \wedge \neg \text{Earthquake}) = 0.003$



## Exercise



## Conditional independence

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

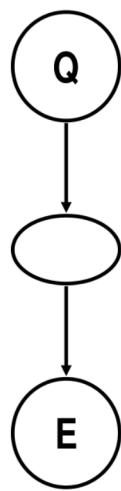
□ The belief network represents conditional independence:

$$P(X_i | X_i, \dots, X_1) = P(X_i | \text{Parents}(X_i))$$

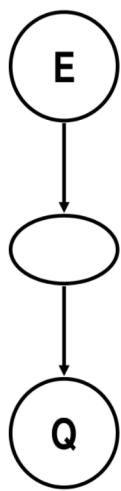
## Incremental Network Construction

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  $\text{Parents}(X)$  to some minimal set of existing nodes such that the conditional independence property is satisfied
  - c) Define the conditional probability table for  $X$

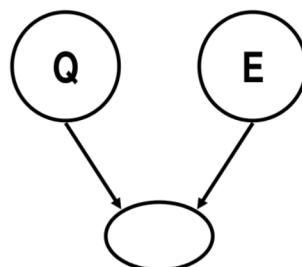
## The four patterns



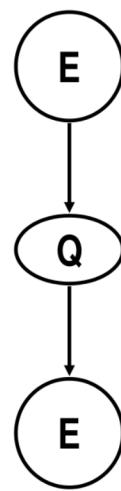
Diagnostic



Causal



InterCausal

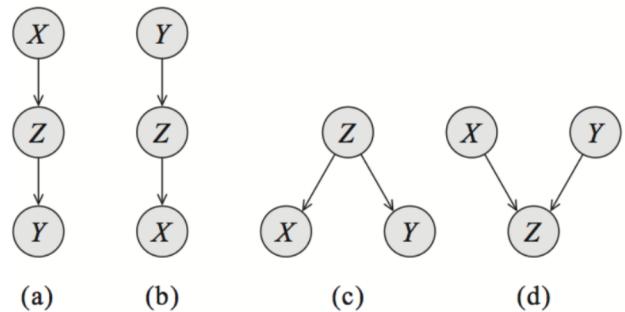


Mixed

## Conditional Independence Relations

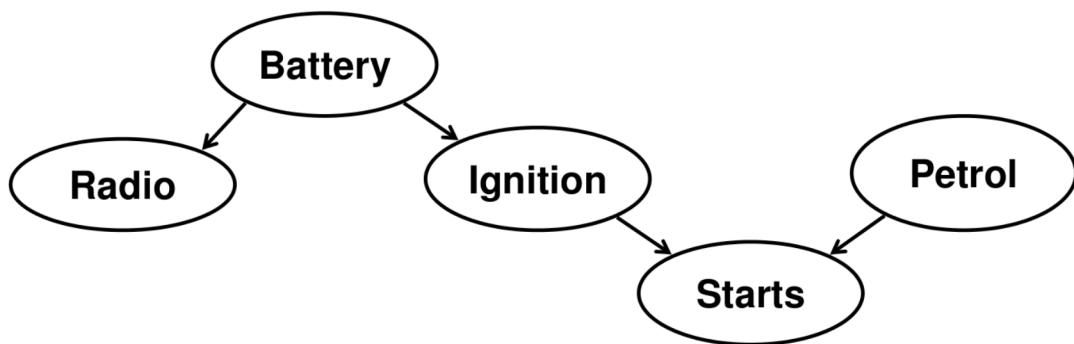
A path is blocked given a set of nodes E if there is a node Z on the path for which one of three conditions holds:

1. Z is in E and Z has one arrow on the path leading in and one arrow out (**Case a and b**)
2. Z is in E and Z has both path arrows leading out (**Case c**)
3. Neither Z nor any descendant of Z is in E, and both path arrows lead in to Z (**Case d**)



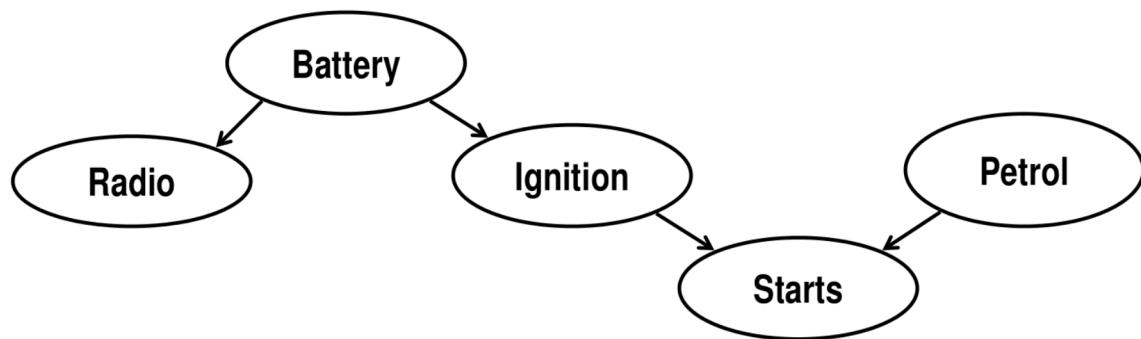
- If every undirected path from a node in X to a node in Y is d-separated by a given set of evidence nodes E, then X and Y are conditionally independent given E.
- A set of nodes E d-separates two sets of nodes X and Y if every undirected path from a node in X to a node in Y is blocked given E.

## Conditional Independence in Belief Networks



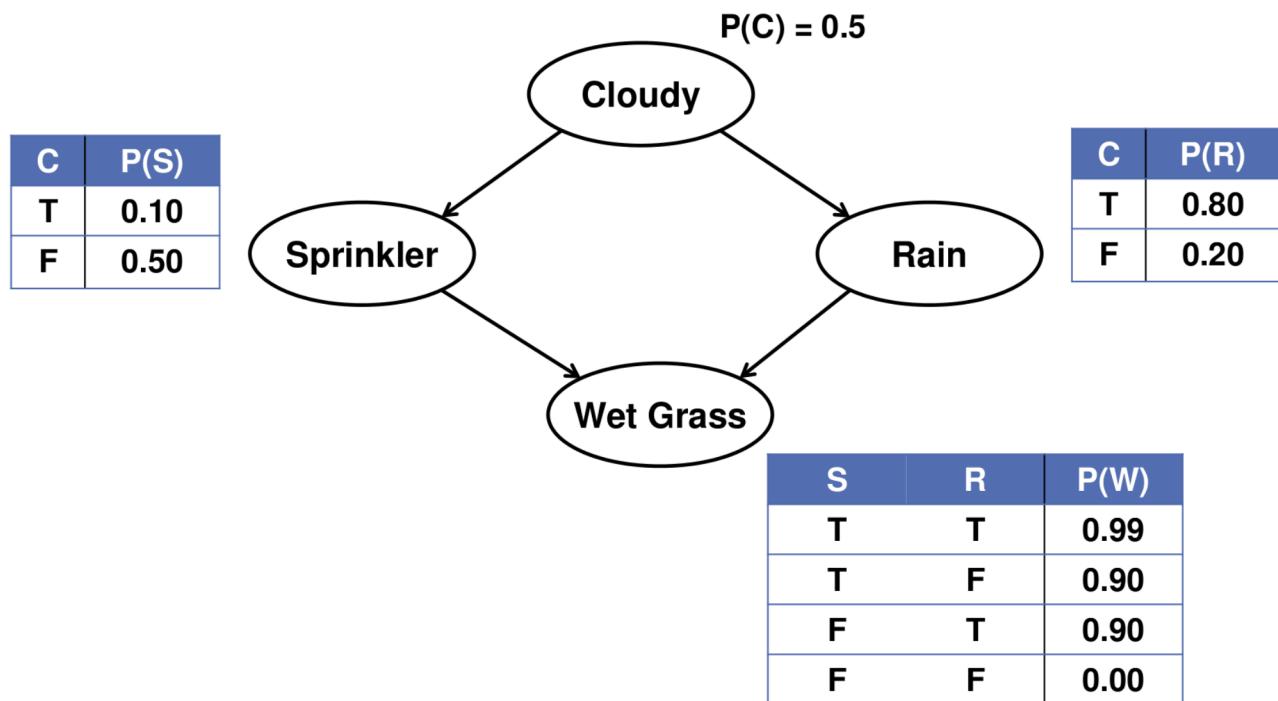
- Whether there is petrol and whether the radio plays are independent given evidence about whether the ignition takes place
- Petrol and Radio are independent if it is known whether the battery works

## Conditional 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.

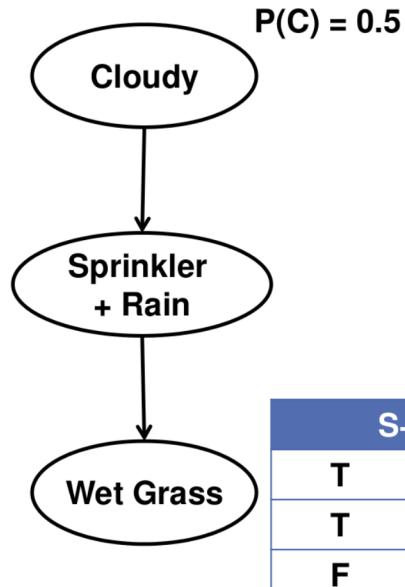
## Inference in multiply connected Belief Networks



## Clustering methods

Transform the net into a probabilistically equivalent (but topologically different) poly-tree by merging offending nodes

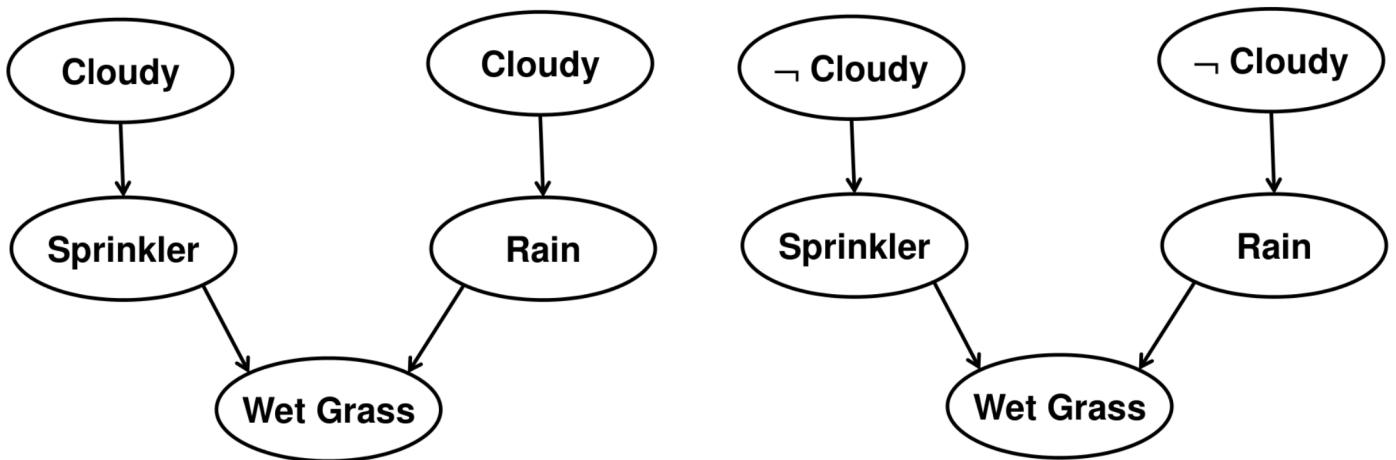
C	P(S+R = x)			
	TT	TF	FT	FF
T	0.08	0.02	0.72	0.18
F	0.40	0.10	0.40	0.10



S+R	P(W)
T T	0.99
T F	0.90
F T	0.90
F F	0.00

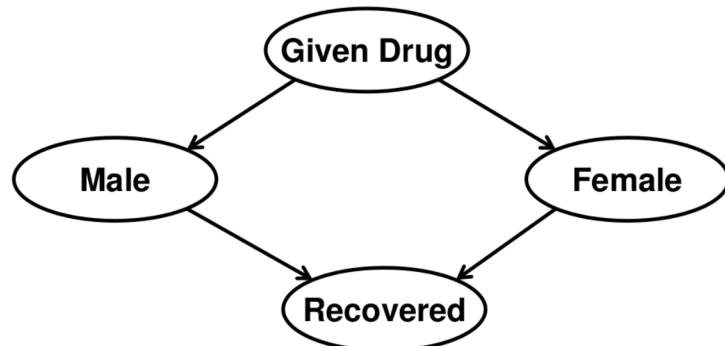
## Cutset conditioning Methods

- A set of variables that can be instantiated to yield a poly-tree is called a *cutset*
- Instantiate the cutset variables to definite values
  - 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.
  - Statistical bias can lead to misleading results – Simpson's paradox



## Simpson's Paradox

Males	Recovered	Not recovered	Rec. Rate
Given drug	18	12	60%
Not given drug	7	3	70%

Females	Recovered	Not recovered	Rec. Rate
Given drug	2	8	20%
Not given drug	9	21	30%

Combined	Recovered	Not recovered	Rec. Rate
Given drug	20	20	50%
Not given drug	16	24	40%

- Should the drug be administered, or not?

## Simpson's Paradox

Males	Recovered	Not recovered	Rec. Rate
Given drug	18	12	60%
Not given drug	7	3	70%
Females	Recovered	Not recovered	Rec. Rate
Given drug	2	8	20%
Not given drug	9	21	30%
Combined	Recovered	Not recovered	Rec. Rate
Given drug	20	20	50%
Not given drug	16	24	40%

$$P(\text{recovery} \mid \text{male} \wedge \text{given\_drug}) = 0.6$$

$$P(\text{recovery} \mid \text{female} \wedge \text{given\_drug}) = 0.2$$

$$P(\text{recovery} \mid \text{given\_drug}) = P(\text{recovery} \mid \text{male} \wedge \text{given\_drug})P(\text{given\_drug} \mid \text{male})$$

$$+ P(\text{recovery} \mid \text{female} \wedge \text{given\_drug})P(\text{given\_drug} \mid \text{female})$$

$$= (0.6 \times 30/40) + (0.2 \times 10/40) = 0.5$$

## Default reasoning

- Some conclusions are made by default unless a counter-evidence is obtained
  - Non-monotonic reasoning
- Points to ponder
  - What is 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.
- 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  $Bel(\neg X) = 0$
- For example, if we do not know whether a coin is fair, then:  
$$Bel(\text{ Heads }) = Bel(\neg\text{Heads}) = 0$$
- If we are given that the coin is fair with 90% certainty, then:  
$$Bel(\text{ Heads }) = 0.9 \times 0.5 = 0.45$$
$$Bel(\neg\text{Heads}) = 0.9 \times 0.5 = 0.45$$
  - *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
- The rules for evaluating the fuzzy truth, T, of a complex sentence are:

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

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

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

## Example: Cardiac Health Management

### Fuzzy Rules

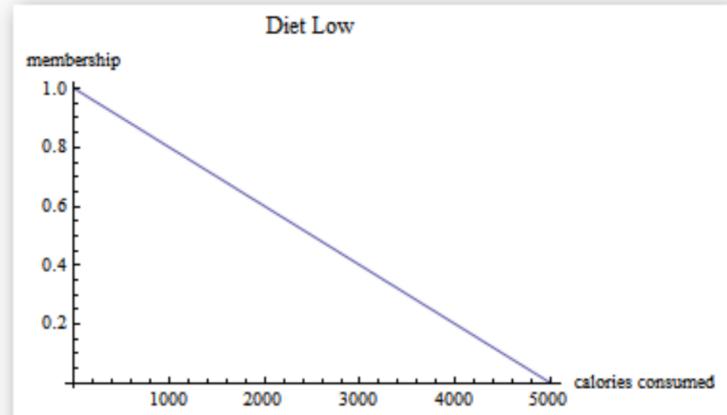
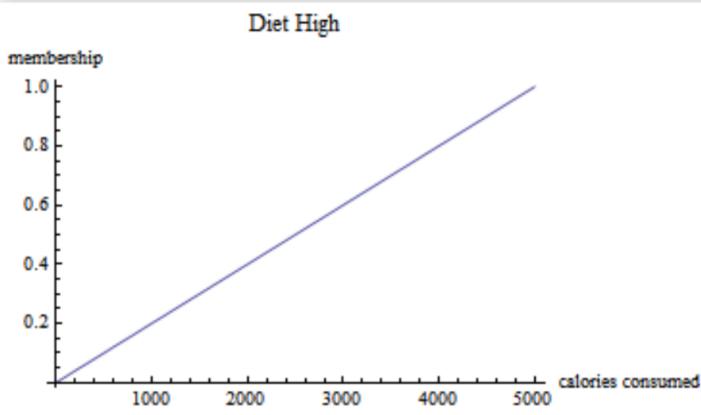
1. Diet is low AND Exercise is high  $\Rightarrow$  Balanced
2. Diet is high OR Exercise is low  $\Rightarrow$  Unbalanced
3. Balanced  $\Rightarrow$  Risk is low
4. Unbalanced  $\Rightarrow$  Risk is high

For a person it is given that:

- Diet = 3000 calories per day
- Exercise = burning 1000 calories per day

What is the risk of heart disease?

## Membership Functions



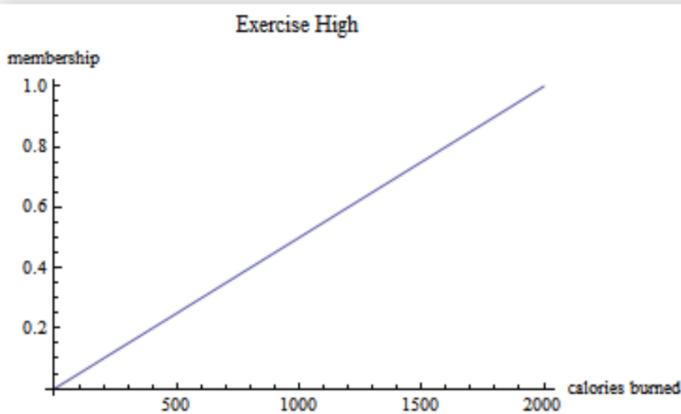
$$f_{\text{diet high}}(x) = \frac{1}{5000}x$$

$$f_{\text{diet low}}(x) = 1 - \frac{1}{5000}x$$

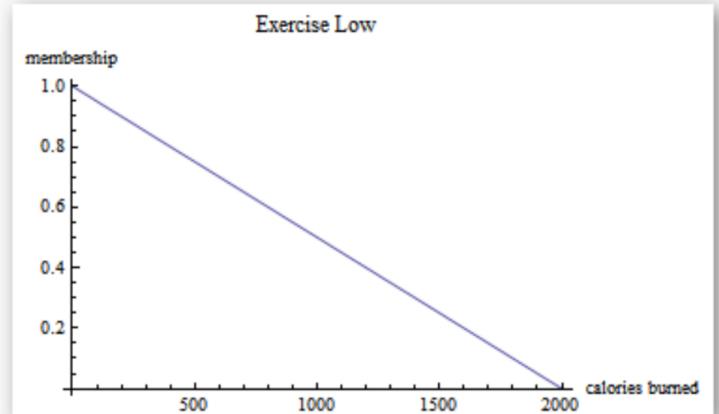
For daily calorie intake of 3000:

Membership for Diet-High =  $3000 / 5000 = 0.6$   
Membership for Diet-Low = 0.4

## Membership Functions



$$f_{\text{exercise high}}(x) = \frac{1}{2000}x$$



$$f_{\text{exercise low}}(x) = 1 - \frac{1}{2000}x$$

For daily calorie burned of 1000:

Membership for Exercise-High =  $1000 / 2000 = 0.5$   
Membership for Exercise-Low = 0.5

## Rule Evaluation

Truth( Diet-High ) = 0.6

Truth( Diet-Low ) = 0.4

Truth( Exercise-High ) = 0.5

Truth( Exercise-Low ) = 0.5

Diet is low AND Exercise is high  $\Rightarrow$  Balanced

- Truth( Balanced ) =  $\min \{ \text{Truth( Diet-Low )}, \text{Truth( Exercise-High )} \} = \min \{ 0.4, 0.5 \} = 0.4$

Diet is high OR Exercise is low  $\Rightarrow$  Unbalanced

- Truth( Unbalanced ) =  $\max \{ \text{Truth( Diet-High )}, \text{Truth( Exercise-Low )} \} = \max \{ 0.6, 0.5 \} = 0.6$

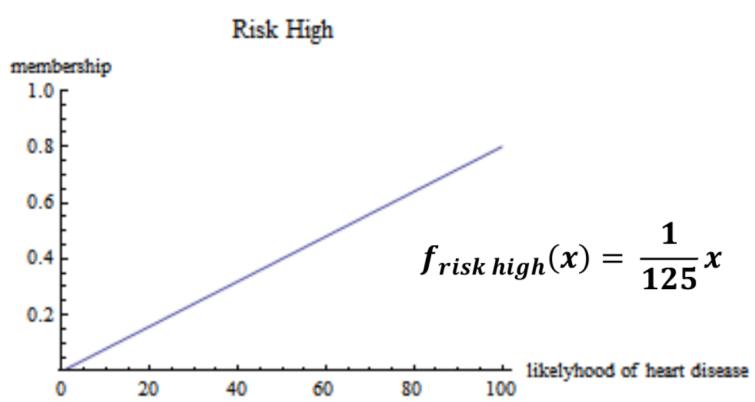
Balanced  $\Rightarrow$  Risk is low

- Truth( Risk-Low ) = Truth( Balanced ) = 0.4

Unbalanced  $\Rightarrow$  Risk is high

- Truth( Risk-High ) = Truth( Unbalanced ) = 0.6

## Risk-High Evaluation



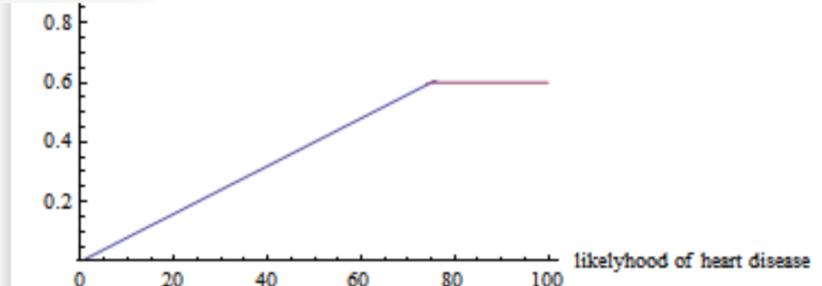
- Truth( Risk-High ) = 0.6

- Therefore:

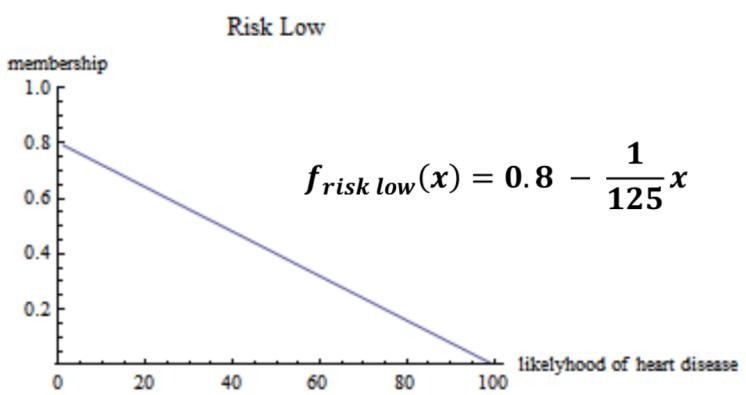
$$0.6 = x / 125$$

$$\text{or, } x = 75$$

Risk High



## Risk-Low Evaluation

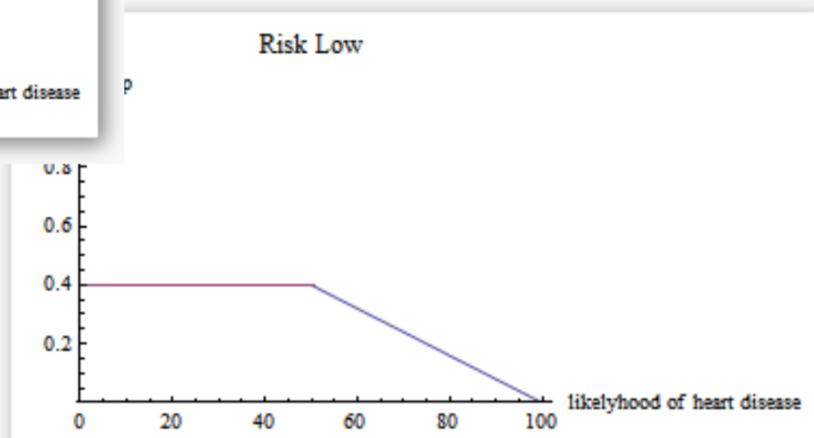


- Truth( Risk-Low ) = 0.4

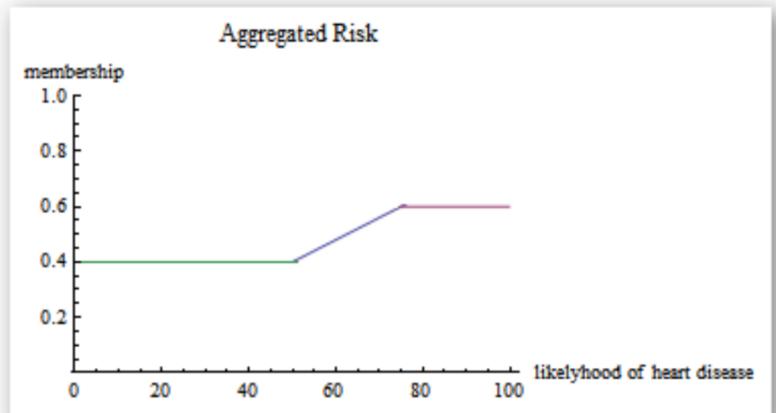
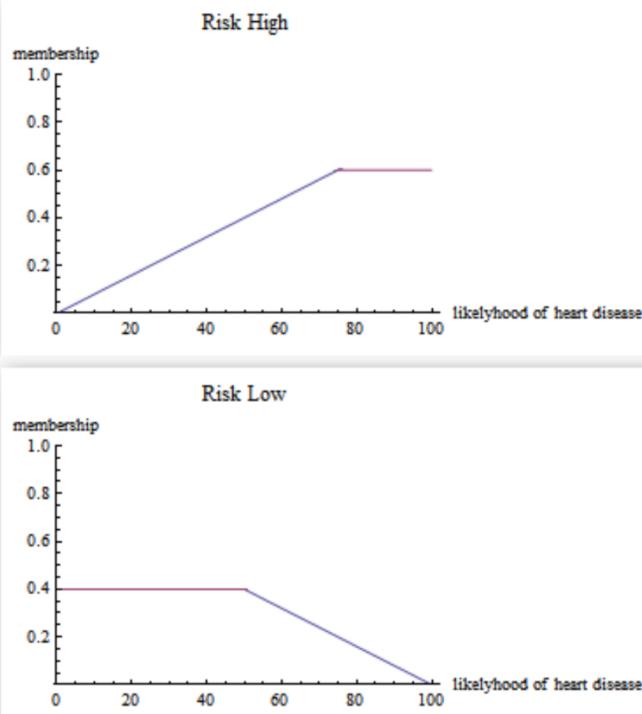
- Therefore:

$$0.4 = 0.8 - x / 125$$

$$\text{or, } x = 50$$



## Aggregated Risk Function



## Defuzzification

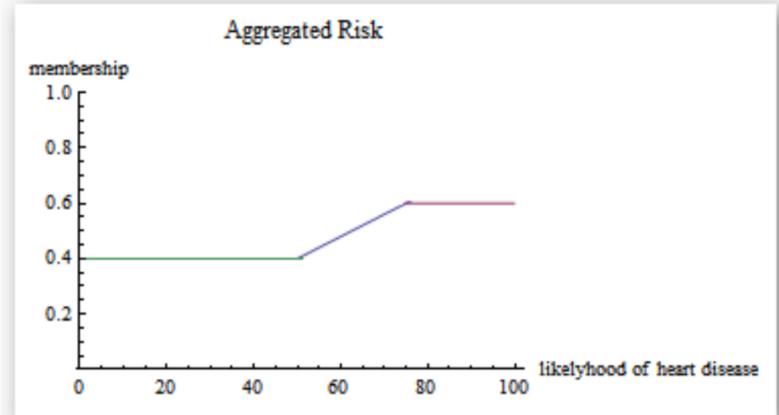
$$\int_0^{100} f_{\text{aggregated risk}} \cdot dx$$

$$= \int_0^{50} 0.4 \, dx + \int_{50}^{75} \frac{1}{125} x \, dx + \int_{75}^{100} 0.6 \, dx$$

$$= 50 \times 0.4 + \frac{1}{125} \left[ \frac{x^2}{2} \right]_{50}^{75} + 25 \times 0.6$$

$$= 20 + (75^2 - 50^2)/250 + 15$$

$$= 47.5$$



Therefore the likelihood of a heart disease for the person is 47.5%

## What next ?

- Probabilistic reasoning is an integral part of many domains of AI. We intend to study the following in future –
  - Probabilities reasoning in state machines (Markov Chains)
    - Good for modeling dynamical systems, recurrent behavior
    - Reinforcement Learning methods work with Markov Decision processes
- You may also look up some of these for further reading –
  - Bayesian optimization is an advanced method for automated problem solving under limited knowledge of the state space
  - Bayesian learning methods are gaining in popularity for making classifiers more important
    - Uncertainty needs to be factored into classifiers, so that the classifier can separate out lack of knowledge as one of the outcomes
    - For example, if a ML classifier is trained to separate wolves from huskies, it should be able to say “I don’t know” if presented with the picture of a cat
  - Structures like Stochastic AND/OR Graphs are being conceived for explainable AI (XAI)