Reasoning under Uncertainty

The intelligent way to handle the unknown

COURSE: CS60045

Pallab Dasgupta
Professor,
Dept. of Computer Sc & Engg

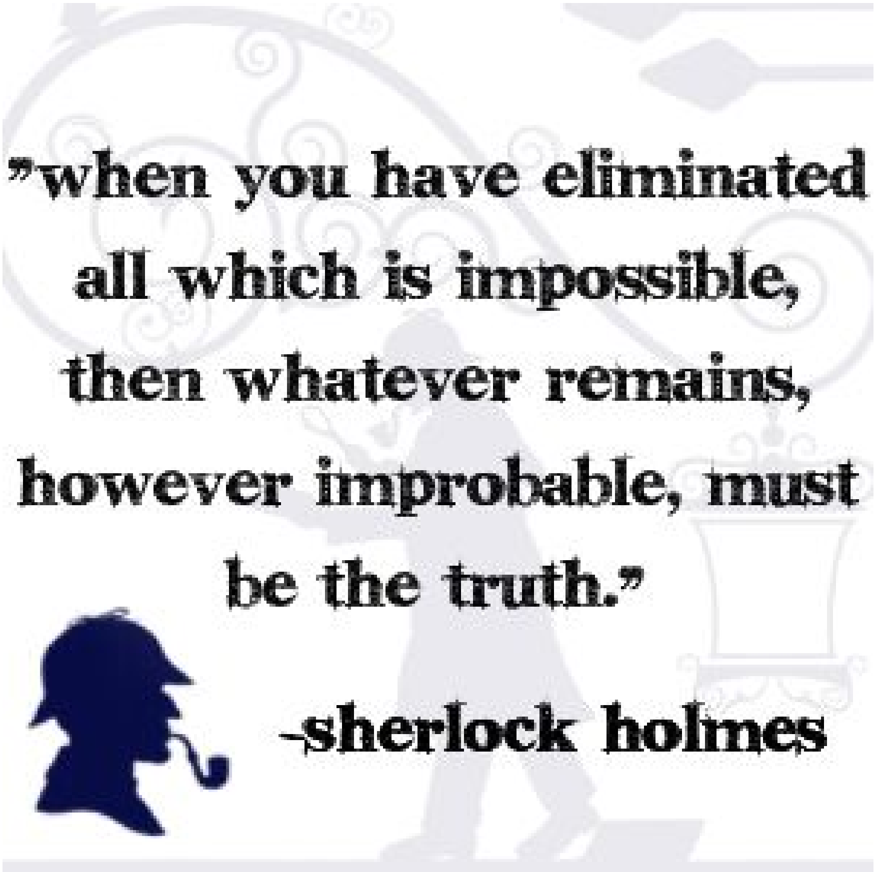
Logical Deduction versus Induction

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



Handling uncertain knowledge

- Classical first order logic has no room for uncertainty

$$\forall p \; Symptom(p, Toothache) \Rightarrow Disease(p, Cavity)$$

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

$$\forall p \; Symptom(p, Toothache) \Rightarrow Disease(p, Cavity) \lor Disease(p, GumDisease) \lor Disease(p, ImpactedWisdom) \lor \ldots$$

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

$$\forall p \; Disease(p, Cavity) \Rightarrow Symptom(p, Toothache)$$

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. $P(True) = 1$ and $P(False) = 0$
3. $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

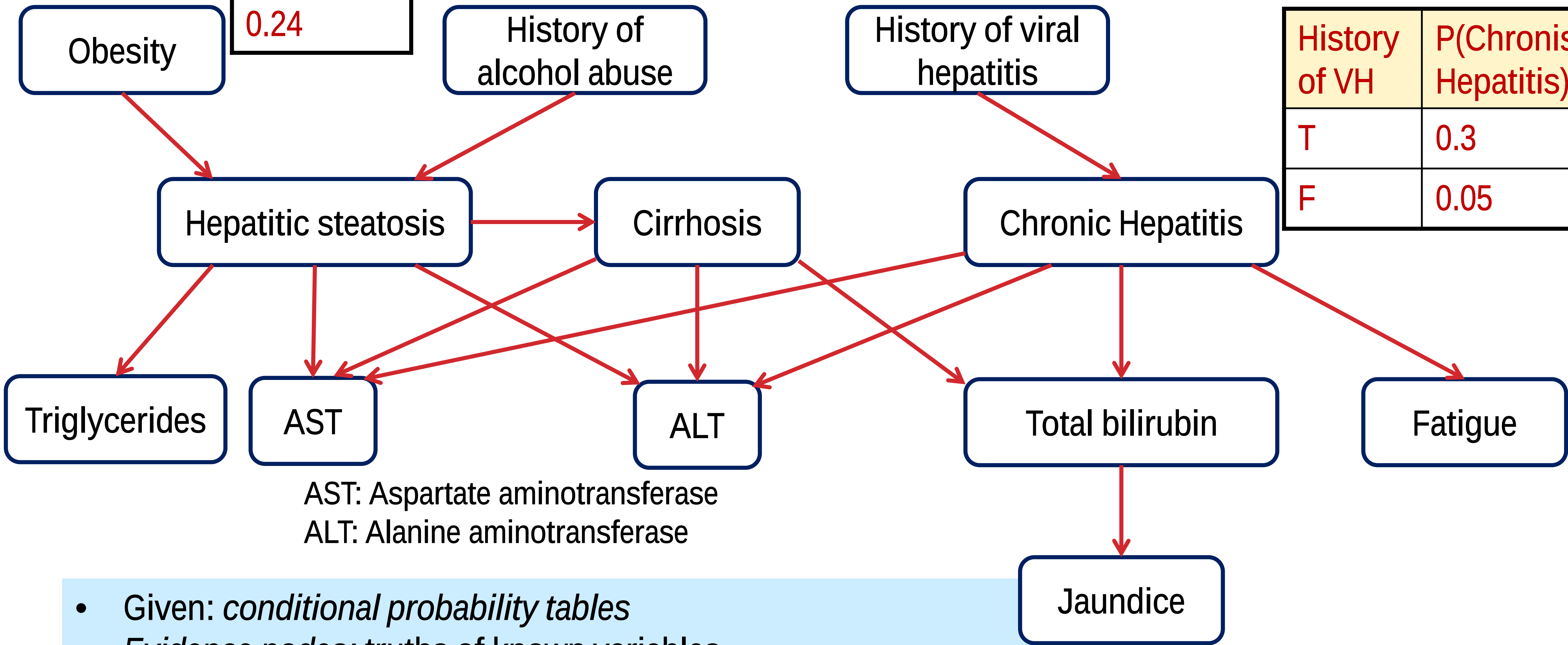
Bayes' Rule

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

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

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

Bayesian Belief Network



P(Obesity)
0.24

History of VH	P(Chronis Hepatitis)
T	0.3
F	0.05

AST: Aspartate aminotransferase
ALT: Alanine aminotransferase

- Given: *conditional probability tables*
- *Evidence nodes:* truths of known variables
- Goal: *Find probabilities of other variables and/or their combinations*

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

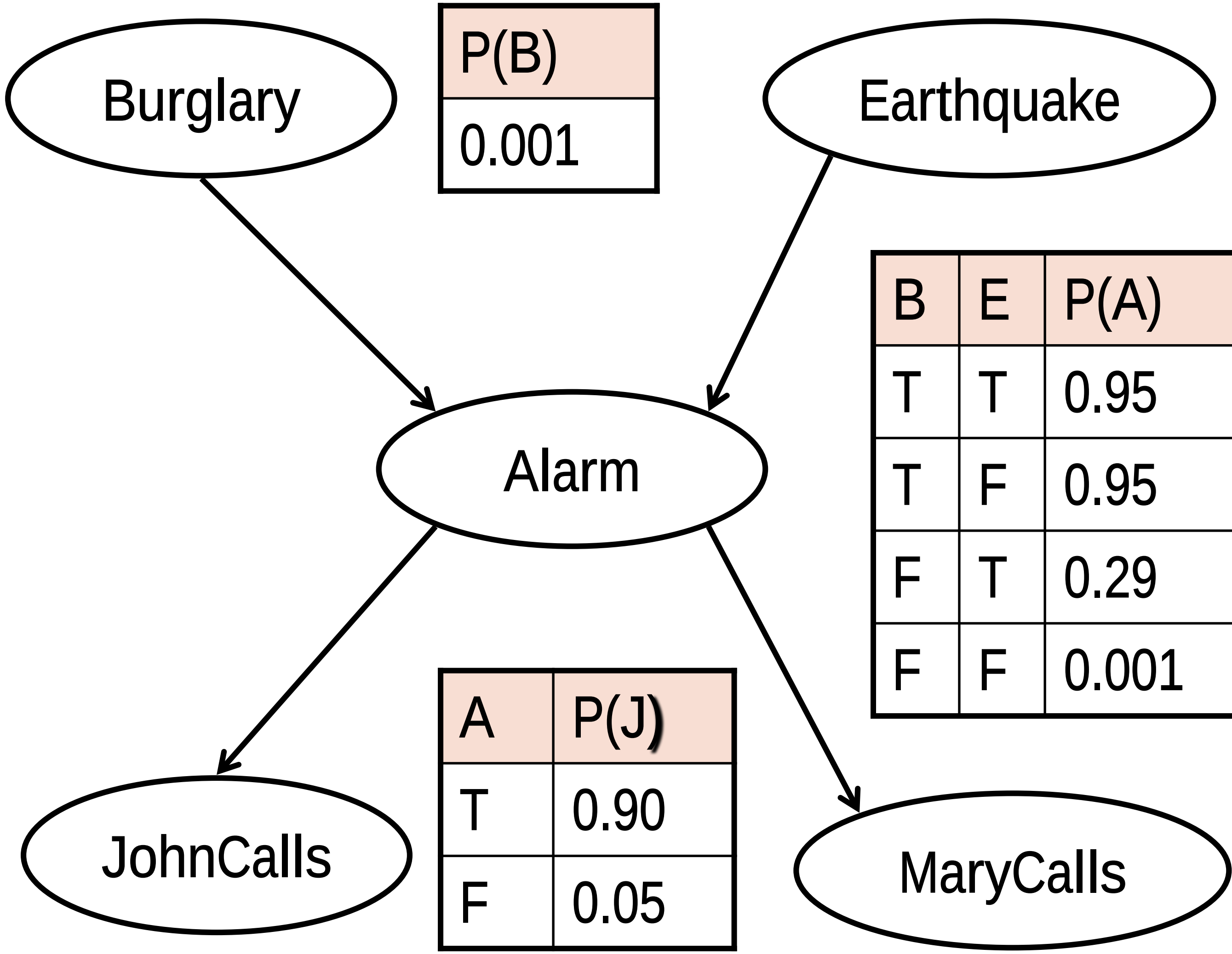
Each node has a conditional probability table that quantifies the effects that the parent have on the node.

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



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

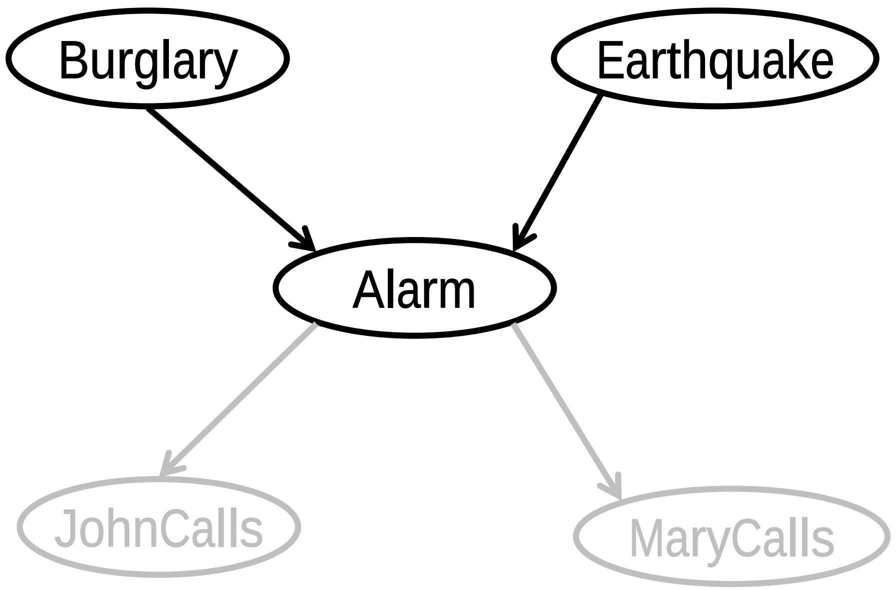
A	P(J)
T	0.90
F	0.05

A	P(M)
T	0.70
F	0.01

The joint probability distribution

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

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

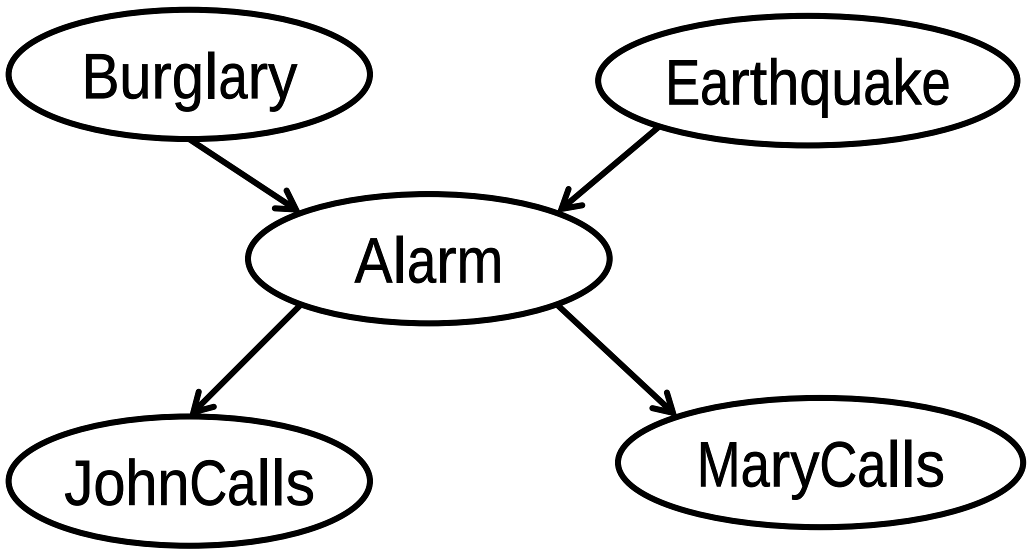


B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

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:

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



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:

$$P(B) = 0.001$$
$$P(B') = 1 - P(B) = 0.999$$

$$P(E) = 0.002$$
$$P(E') = 1 - P(E) = 0.998$$

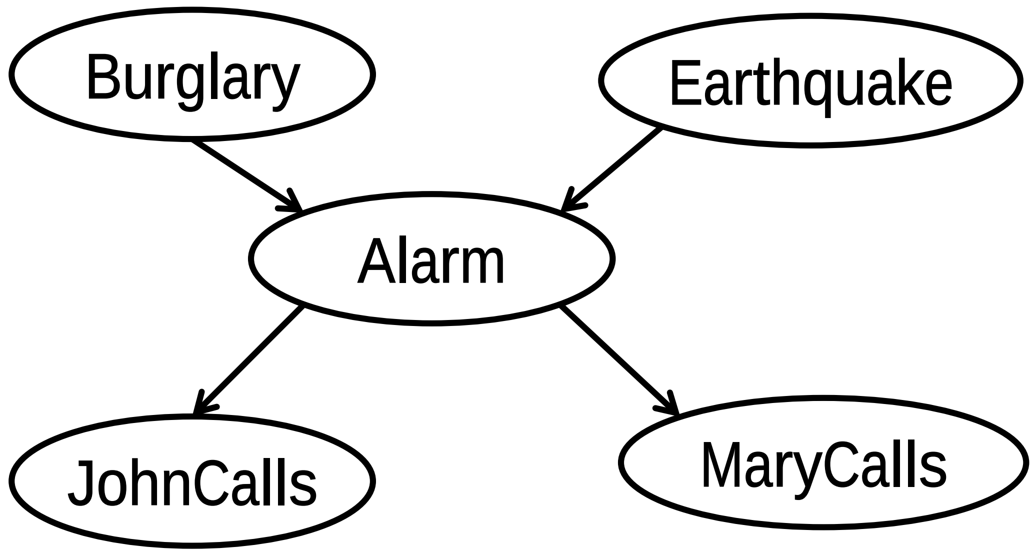
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

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:

$$P(A) = P(AB'E') + P(AB'E) + P(ABE') + P(ABE)$$

$$= P(A \mid B'E').P(B'E') + P(A \mid B'E).P(B'E) + P(A \mid BE').P(BE') + P(A \mid BE).P(BE)$$

$$= 0.001 \times 0.999 \times 0.998 + 0.29 \times 0.999 \times 0.002 + 0.95 \times 0.001 \times 0.998 + 0.95 \times 0.001 \times 0.002$$

$$= 0.001 + 0.0006 + 0.0009 = 0.0025$$

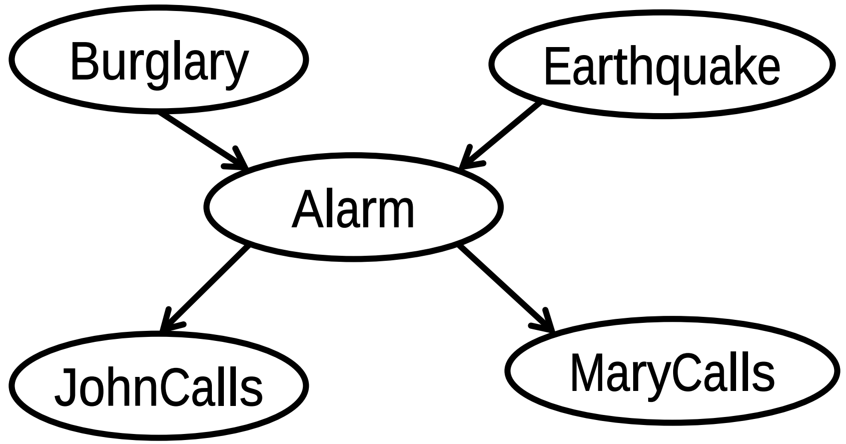
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

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(J)

$$P(J) = P(JA) + P(JA')$$

$$= P(J \mid A).P(A) + P(J \mid A').P(A')$$

$$= 0.9 \times 0.0025 + 0.05 \times (1 - 0.0025)$$

$$= 0.052125$$

$$P(AB) = P(ABE) + P(ABE') = 0.95 \times 0.001 \times 0.002 + 0.95 \times 0.001 \times 0.998$$

$$= 0.00095$$

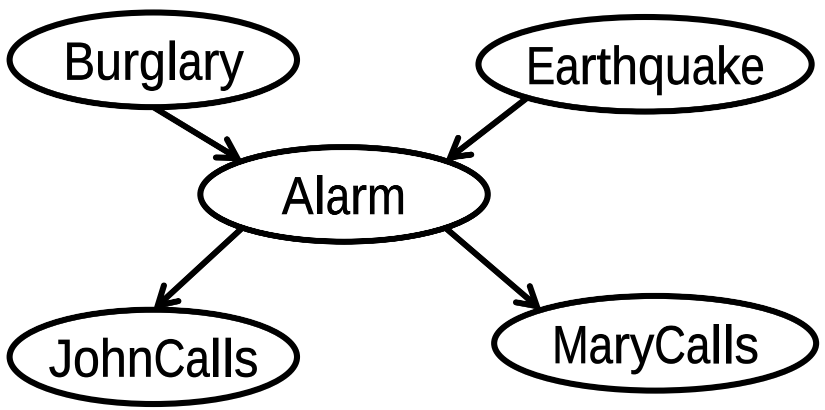
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

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(A'B) *and* P(AE)

$$P(A'B) = P(A'BE) + P(A'BE')$$
$$= P(A' \mid BE).P(BE) + P(A' \mid BE').P(BE')$$
$$= (1 - 0.95) \times 0.001 \times 0.002$$
$$+ (1 - 0.95) \times 0.001 \times 0.998$$
$$= 0.00005$$

$$P(AE) = P(AEB) + P(AEB')$$
$$= 0.95 \times 0.001 \times 0.002 + 0.29 \times 0.999 \times 0.002 = 0.00058$$

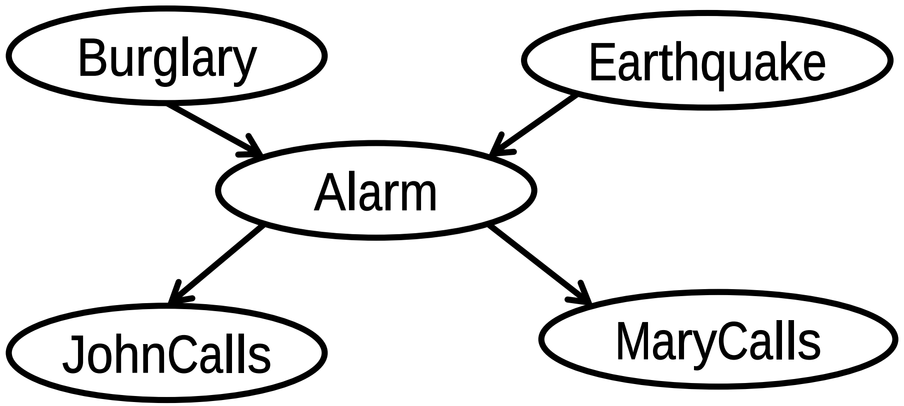
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

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

$$P(AE') = P(AE'B) + P(AE'B')$$
$$= 0.95 \times 0.001 \times 0.998 + 0.001 \times 0.999 \times 0.998$$
$$= 0.001945$$

$$P(A'E') = P(A'E'B) + P(A'E'B')$$
$$= P(A' \mid BE').P(BE') + P(A' \mid B'E').P(B'E')$$
$$= (1 - 0.95) \times 0.001 \times 0.998 + (1 - 0.001) \times 0.999 \times 0.998 = 0.996$$

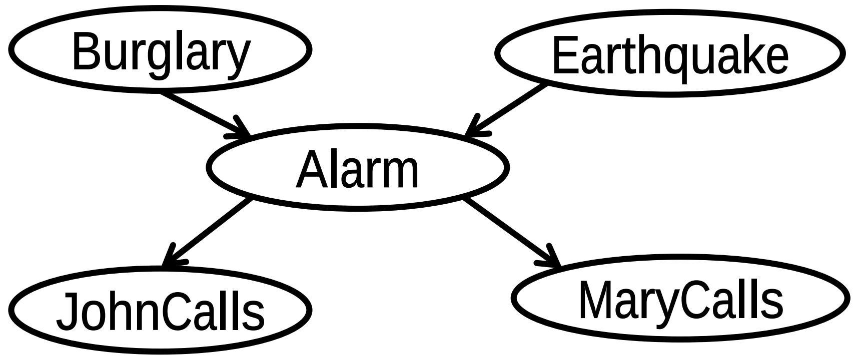
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

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)

$$P(JB) = P(JBA) + P(JBA')$$

$$= P(J \mid AB).P(AB) + P(J \mid A'B).P(A'B)$$

$$= P(J \mid A).P(AB) + P(J \mid A').P(A'B)$$

$$= 0.9 \times 0.00095 + 0.05 \times 0.00005$$

$$= 0.00086$$

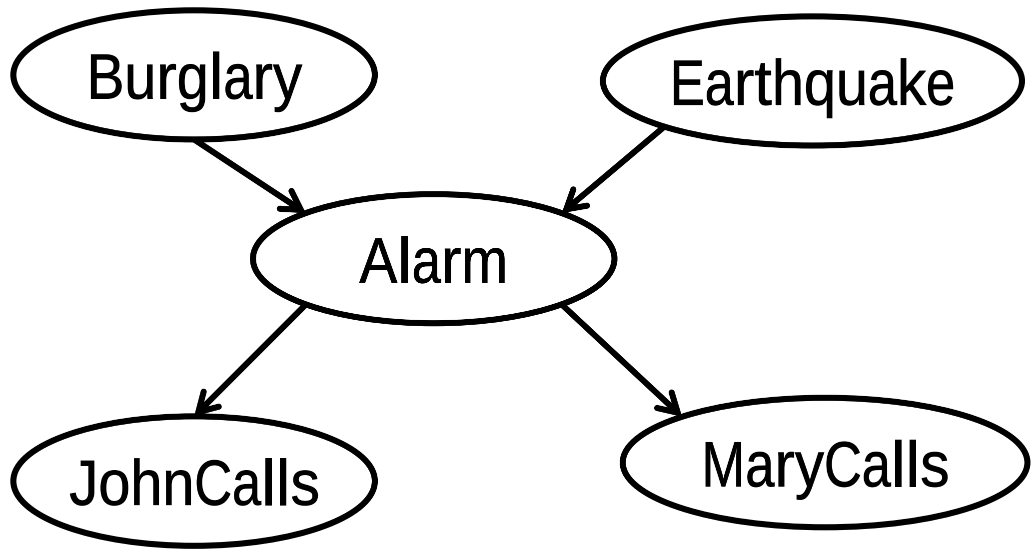
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

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:

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

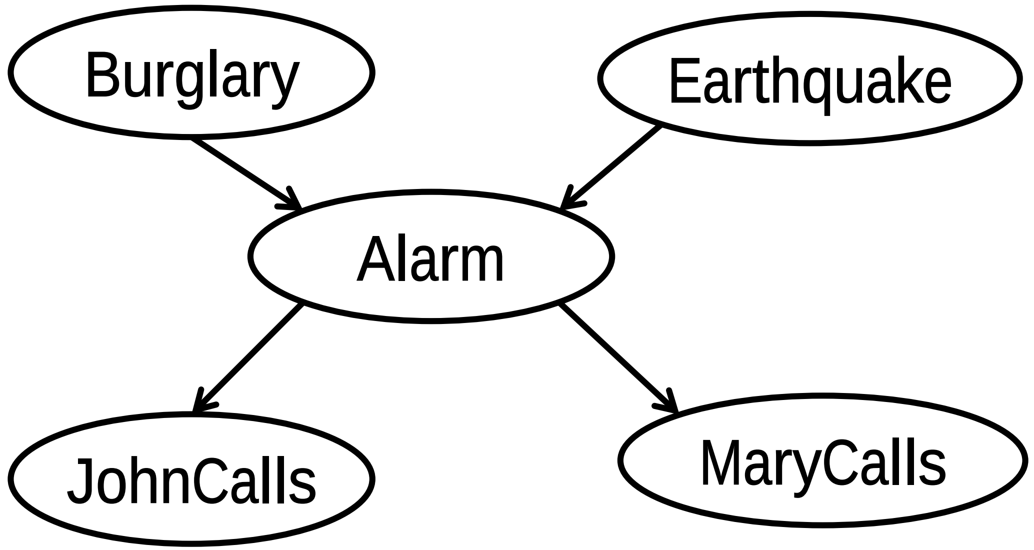
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	P(J)
T	0.90
F	0.05

A	P(M)
T	0.70
F	0.01

P(E)
0.002

P(B)
0.001



The joint probability distribution

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

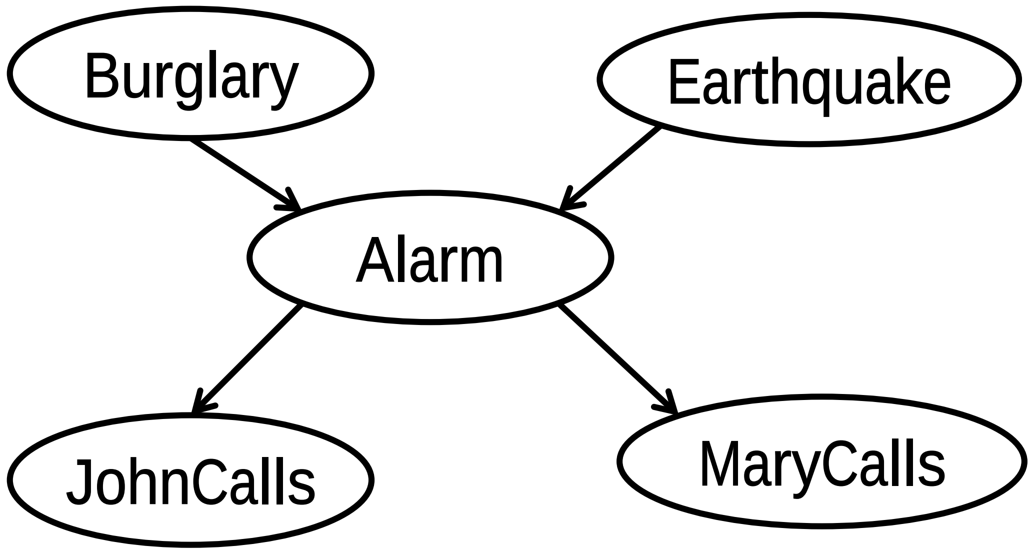
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	P(J)
T	0.90
F	0.05

A	P(M)
T	0.70
F	0.01

P(E)
0.002

P(B)
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

P(B | AE) = P(ABE) / P(AE) = [P(A | BE).P(BE)] / P(AE)

= [0.95 x 0.001 x 0.002] / 0.00058

= 0.003

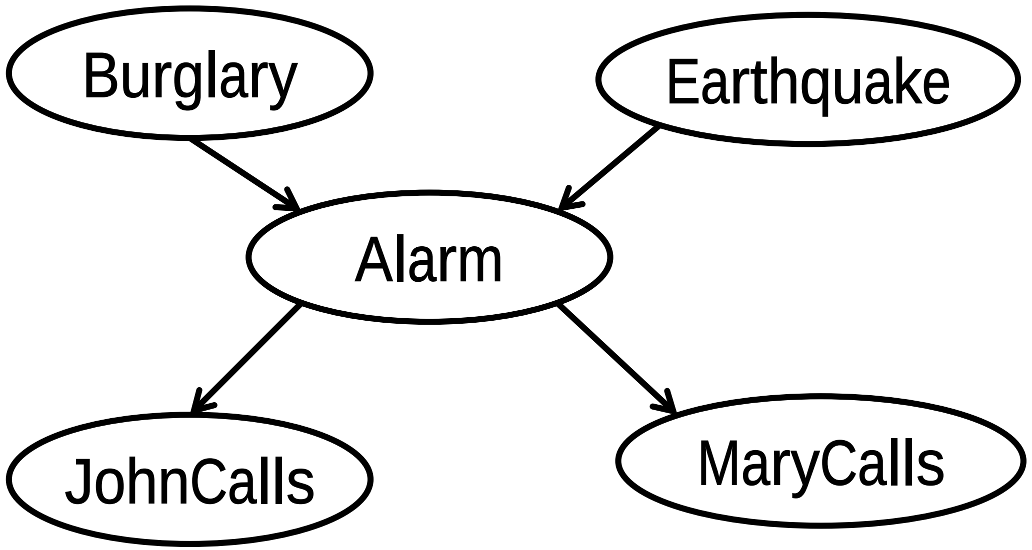
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

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:

$$P(AJE') = P(J \mid AE').P(AE') = P(J \mid A).P(AE')$$
$$= 0.9 \times 0.001945 = 0.00175$$

$$P(A'JE') = P(J \mid A'E').P(A'E') = P(J \mid A').P(A'E')$$
$$= 0.05 \times 0.996 = 0.0498$$

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

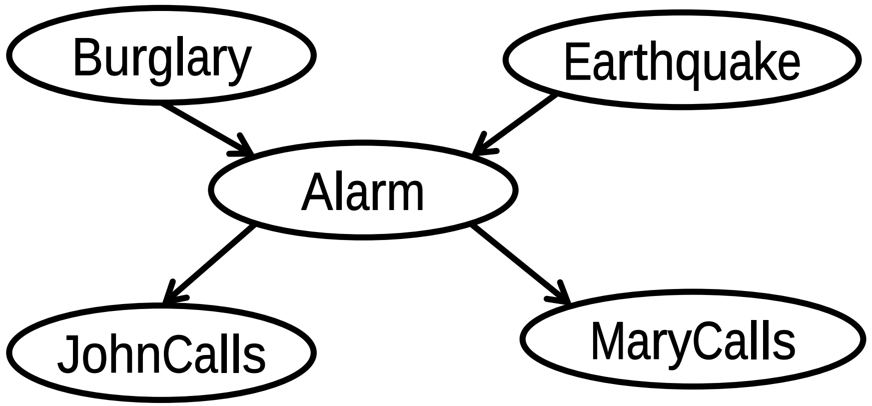
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	P(J)
T	0.90
F	0.05

A	P(M)
T	0.70
F	0.01

P(E)
0.002

P(B)
0.001



The joint probability distribution

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

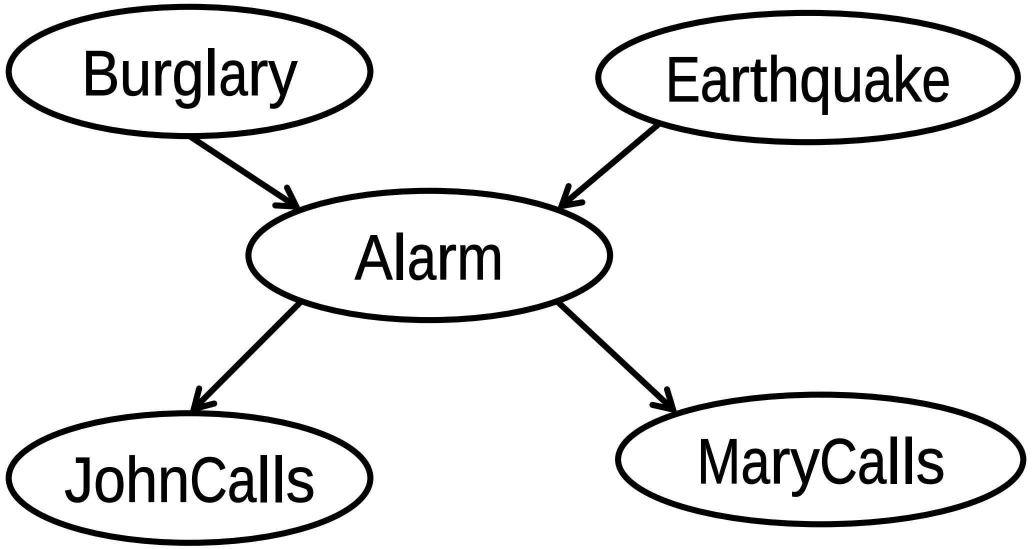
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	P(J)
T	0.90
F	0.05

A	P(M)
T	0.70
F	0.01

P(E)
0.002

P(B)
0.001



The joint probability distribution

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 x 0.95 x 0.001 x 0.998 + 0.05 x (1 – 0.95) x 0.001 x 0.998

= 0.000856

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

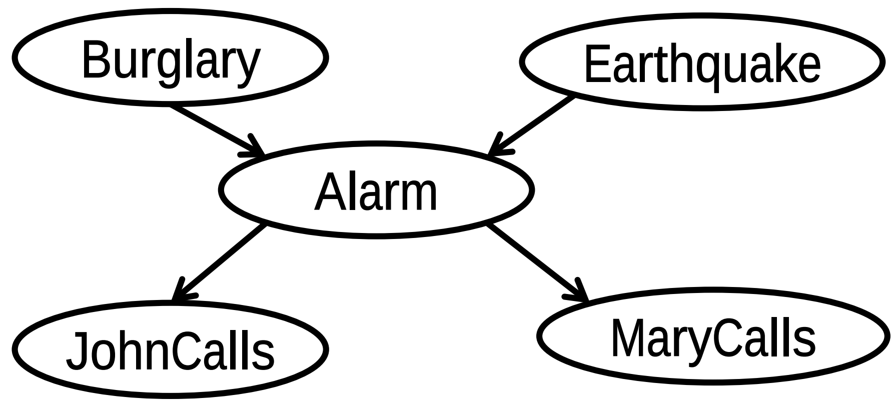
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	P(J)
T	0.90
F	0.05

A	P(M)
T	0.70
F	0.01

P(E)
0.002

P(B)
0.001

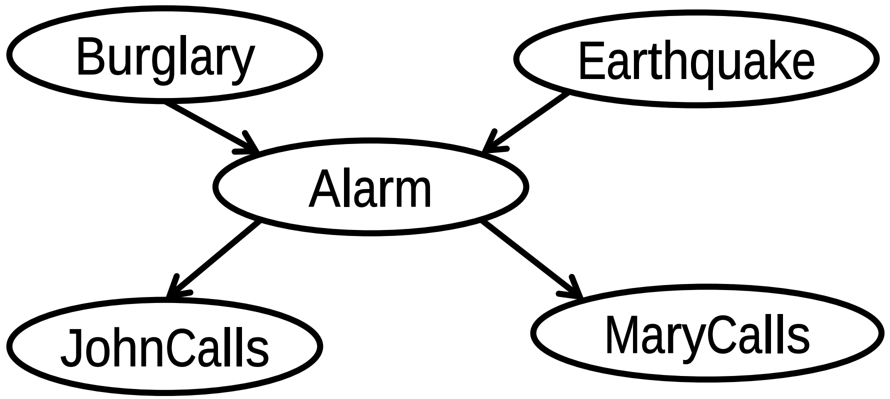


Inferences using belief networks

- Diagnostic inferences (from effects to causes)

  - Given that JohnCalls, infer that

    P(Burglary | JohnCalls) = 0.016



- Causal inferences (from causes to effects)

  - Given Burglary, infer that

    P(JohnCalls | Burglary) = 0.86

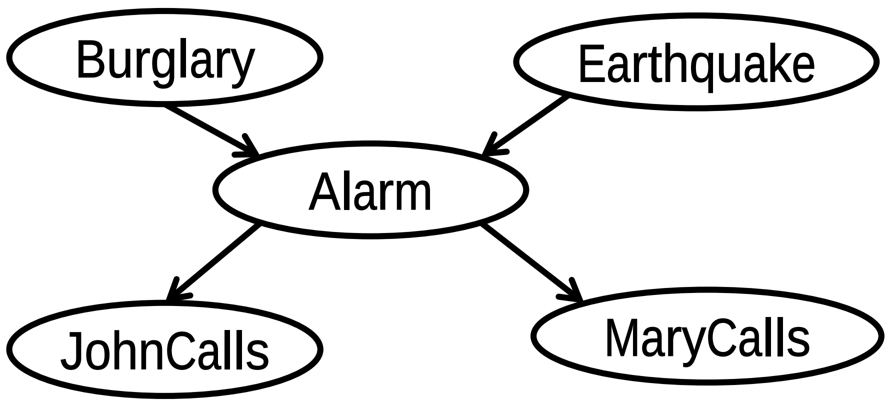
    P(MaryCalls | Burglary) = 0.67

Inferences using belief networks

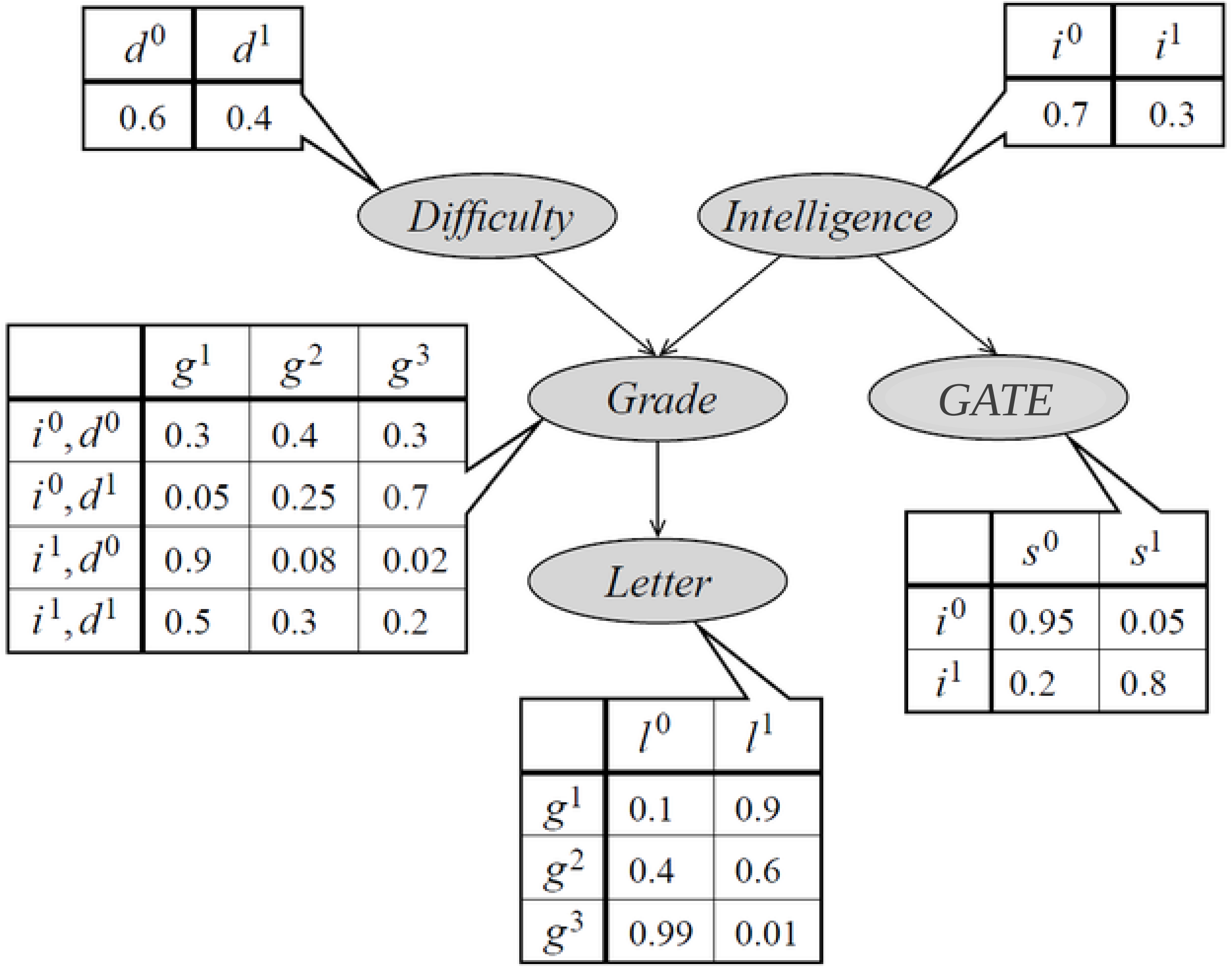
- Inter-causal inferences (between causes of a common effect)
  - Given Alarm, we have P(Burglary | Alarm) = 0.376
  - If we add evidence that Earthquake is true, then P(Burglary | Alarm ∧ Earthquake) = 0.003

- Mixed inferences
  - Setting the effect JohnCalls to true and the cause Earthquake to false gives

    P(Alarm | JohnCalls ∧ ¬ Earthquake) = 0.003



Exercise



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

	g^1	g^2	g^3
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
i^1, d^0	0.9	0.08	0.02
i^1, d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

Conditional independence

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

- 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

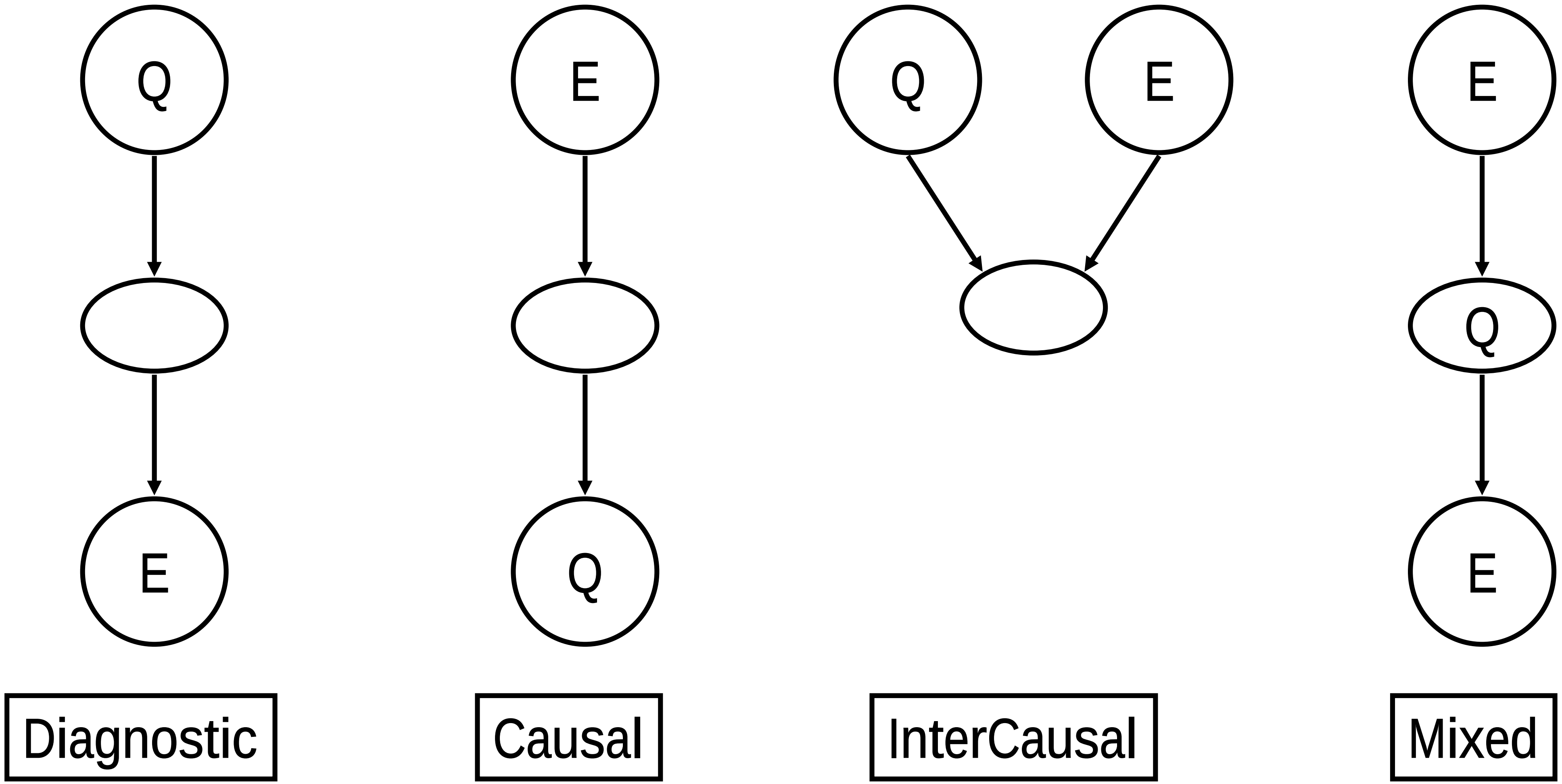
1. Choose the set of relevant variables X_i that describe the domain

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 probability table for X

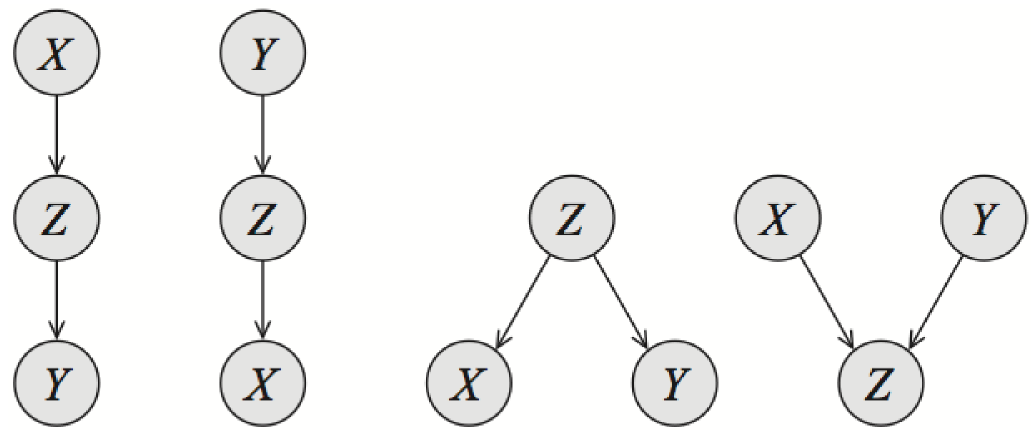
The four patterns



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:

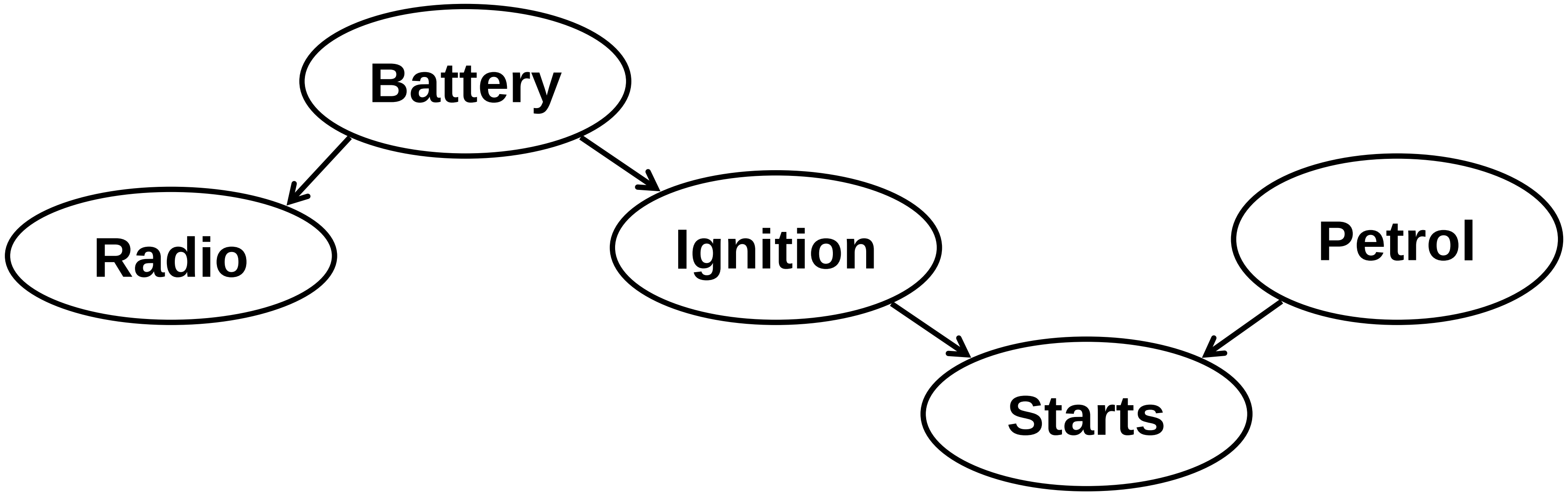
1. Z is in E and Z has one arrow on the path leading in and one arrow out (Case a and b)
2. Z is in E and Z has both path arrows leading out (Case c)
3. Neither Z nor any descendant of Z is in E, and both path arrows lead in to Z (Case d)



(a) (b) (c) (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.

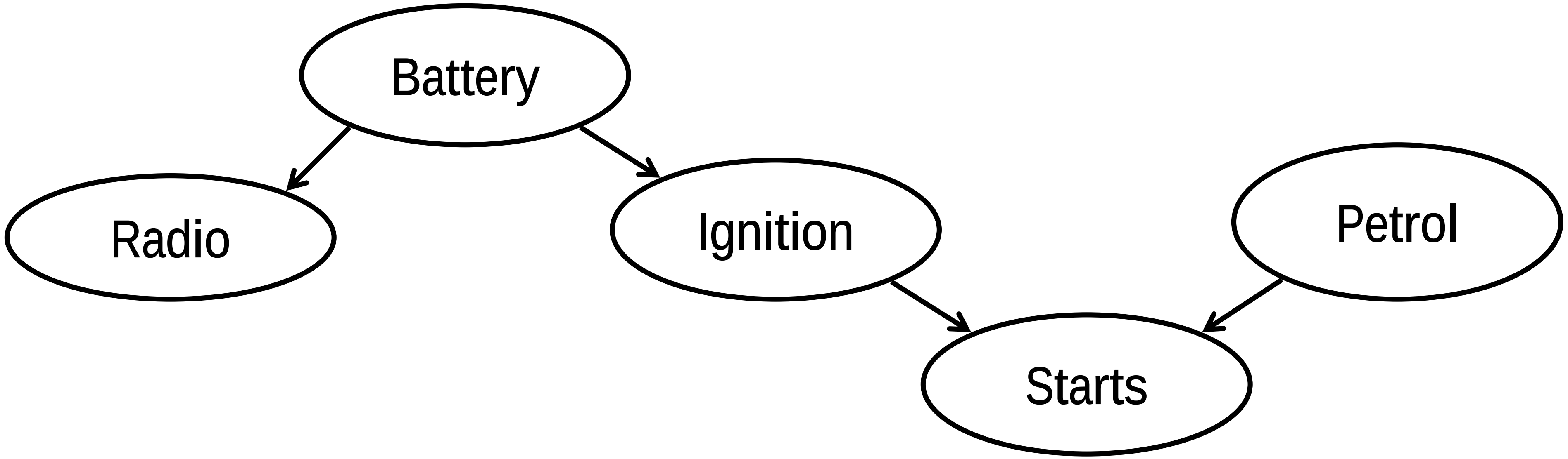
Conditional Independence in Belief Networks



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

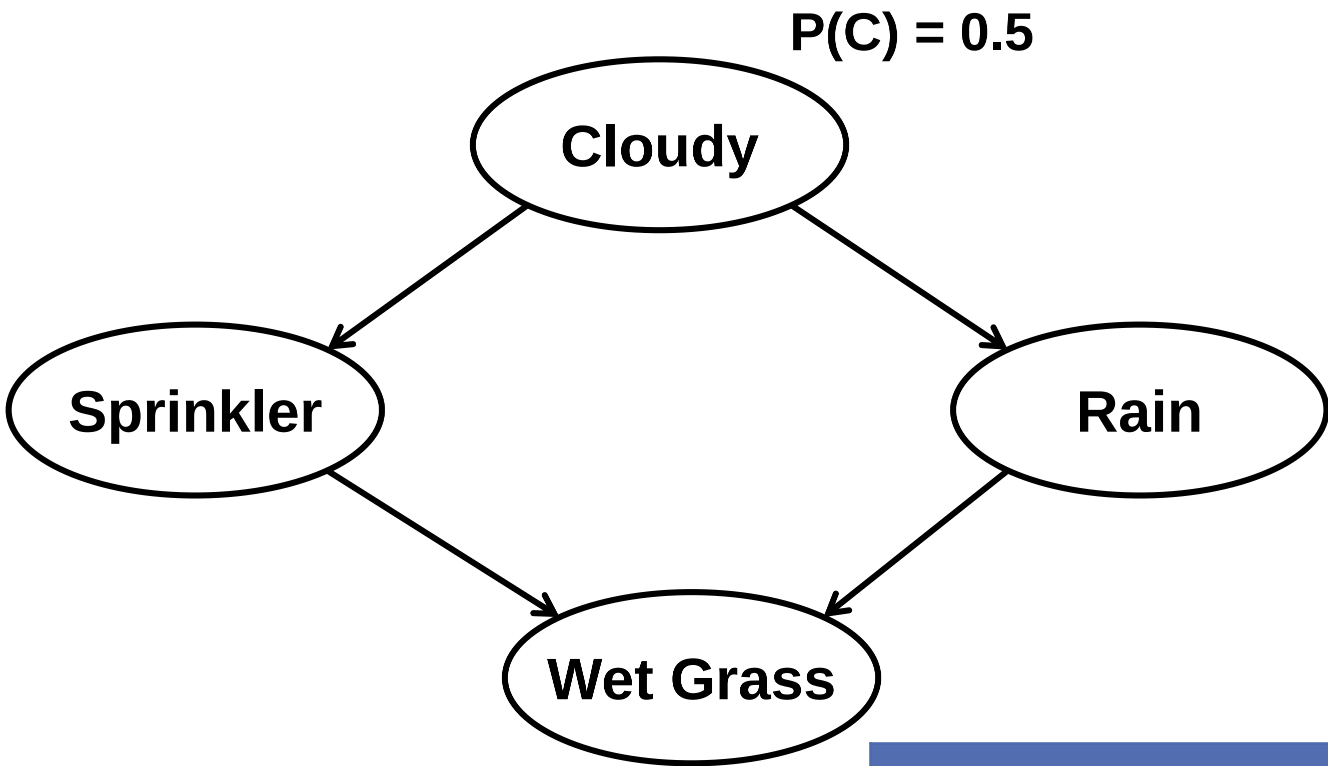
- Petrol and Radio are independent if it is known whether the battery works

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



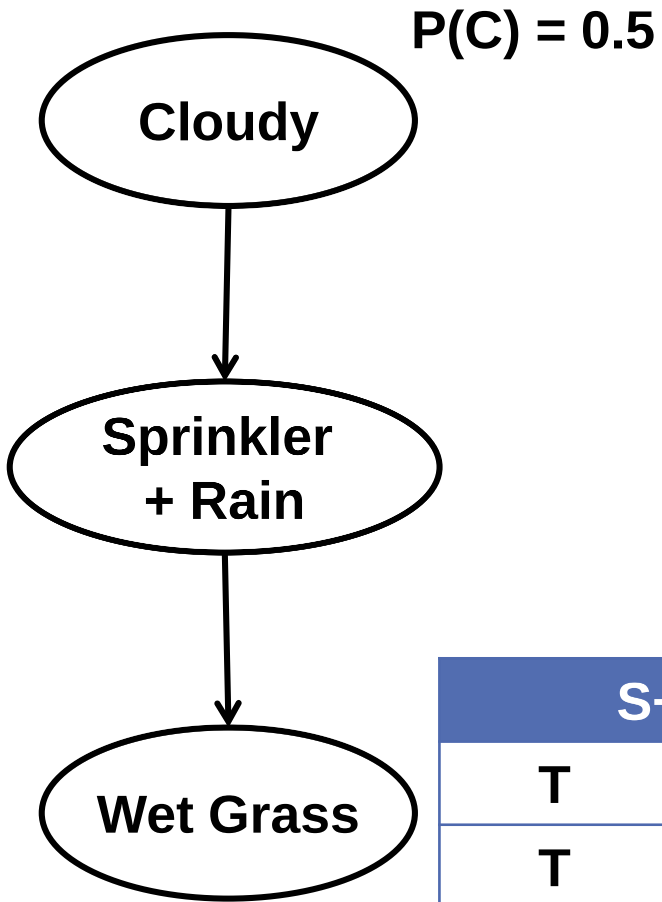
C	P(S)
T	0.10
F	0.50

C	P(R)
T	0.80
F	0.20

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

Clustering methods

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

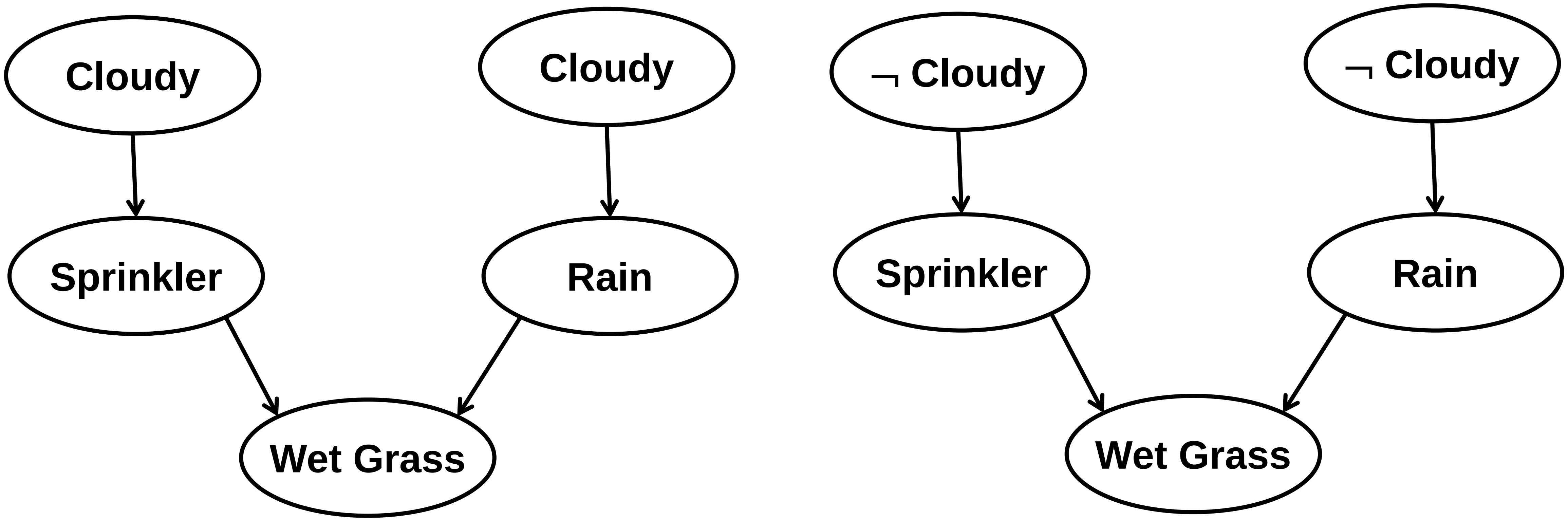


C	P(S+R = x)			
	TT	TF	FT	FF
T	0.08	0.02	0.72	0.18
F	0.40	0.10	0.40	0.10

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

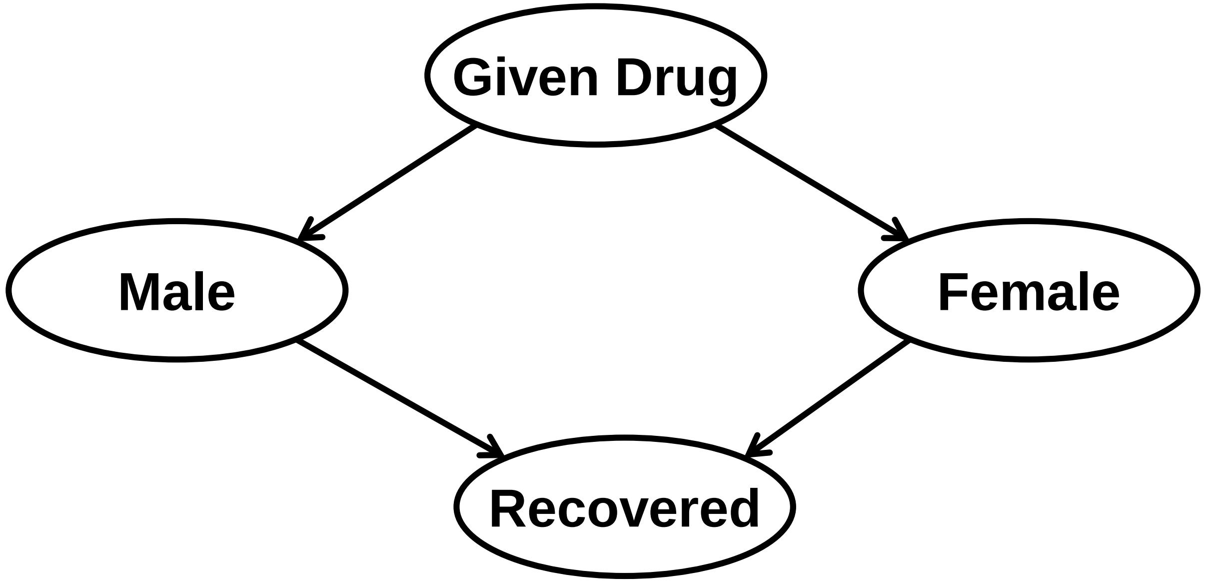
Cutset conditioning Methods

- 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(¬X) = 0

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

  Bel(Heads) = Bel(¬Heads) = 0

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

  Bel(Heads) = 0.9 X 0.5 = 0.45
  Bel(¬Heads) = 0.9 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

- 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

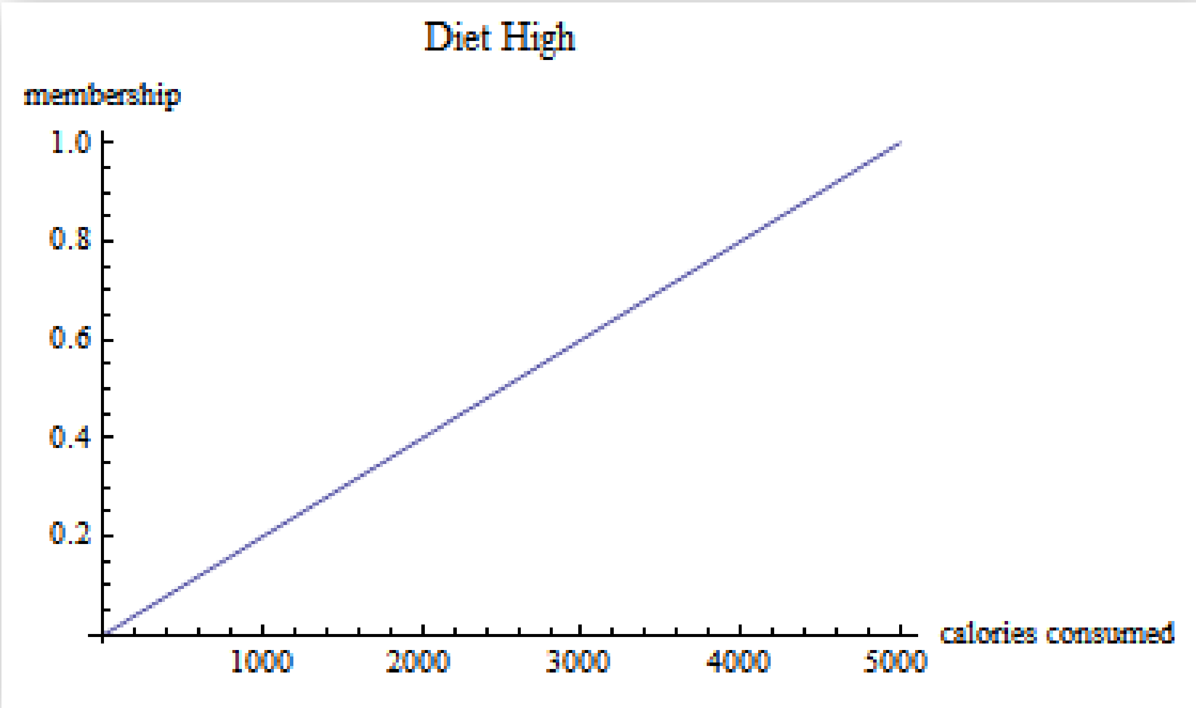
1. Diet is low AND Exercise is high $\Rightarrow$ Balanced
2. Diet is high OR Exercise is low $\Rightarrow$ Unbalanced
3. Balanced $\Rightarrow$ Risk is low
4. Unbalanced $\Rightarrow$ Risk is high

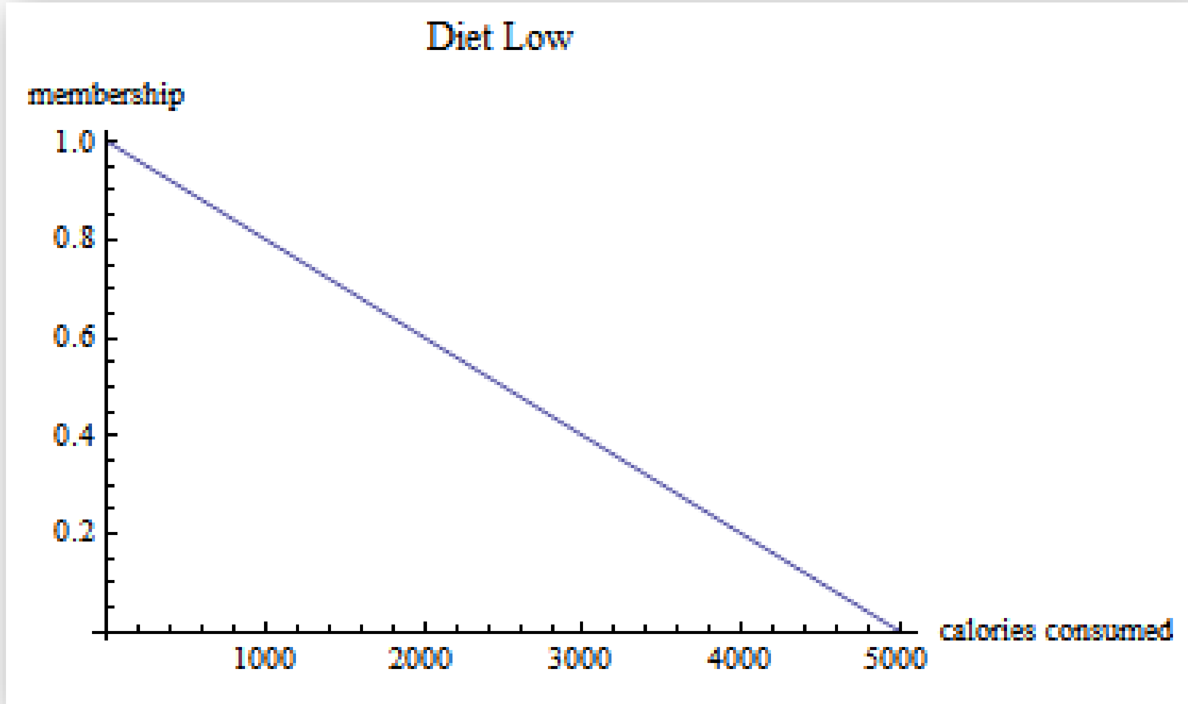
For a person it is given that:

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

What is the risk of heart disease?

Membership Functions





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

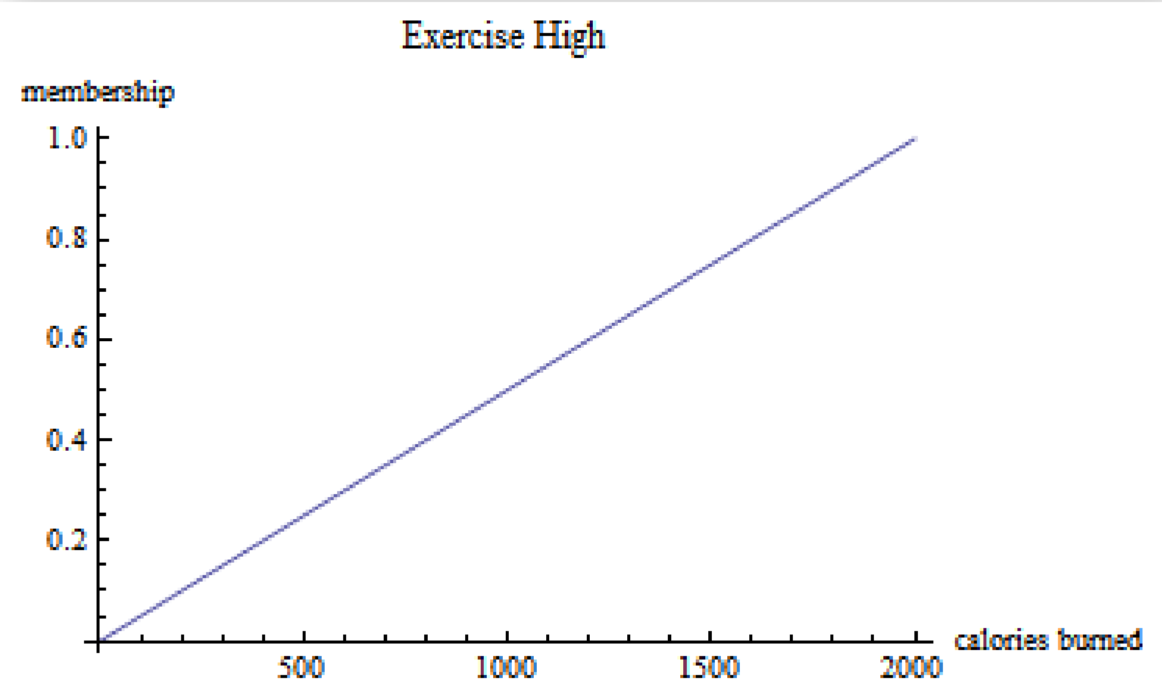
$$f_{diet\ low}(x) = 1 - \frac{1}{5000}x$$

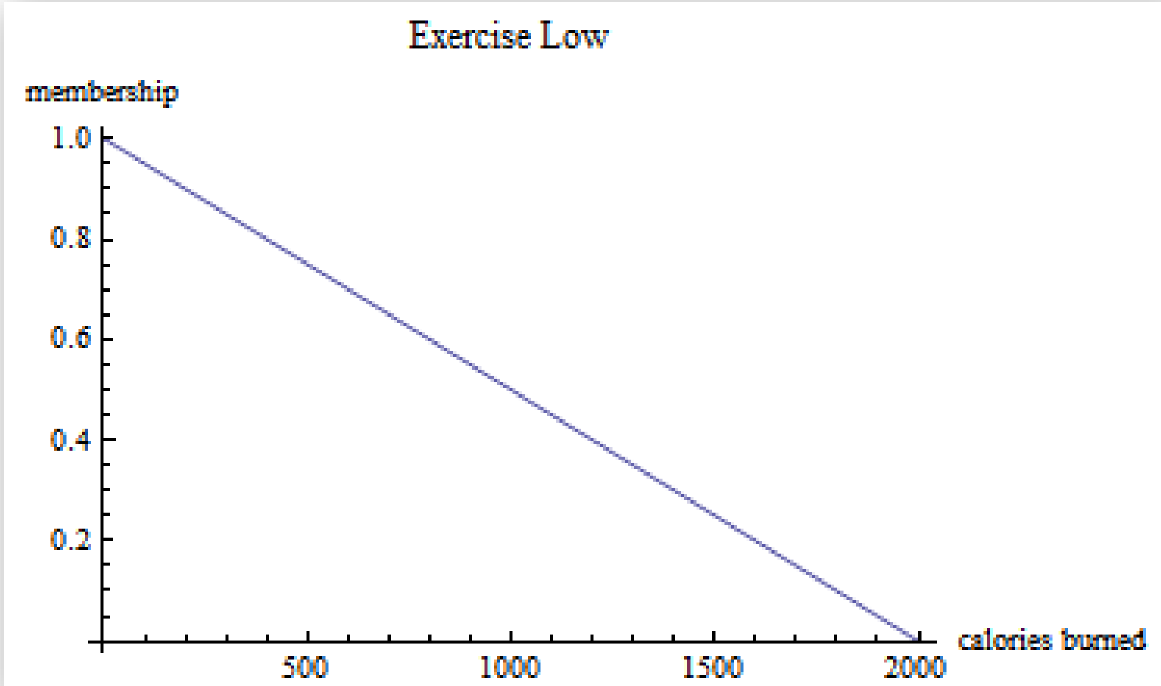
For daily calorie intake of 3000:

Membership for Diet-High = 3000 / 5000 = 0.6

Membership for Diet-Low = 0.4

Membership Functions





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

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

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 ⇒ Balanced

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

Diet is high OR Exercise is low ⇒ Unbalanced

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

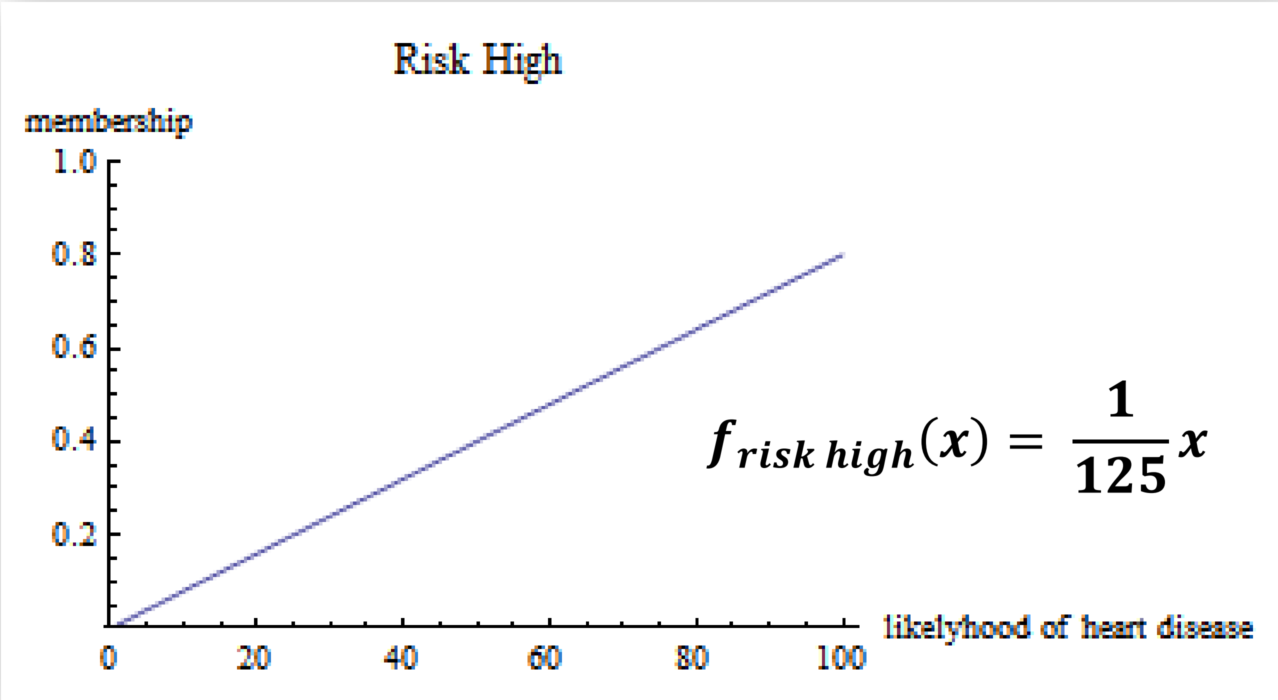
Balanced ⇒ Risk is low

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

Unbalanced ⇒ Risk is high

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

Risk-High Evaluation

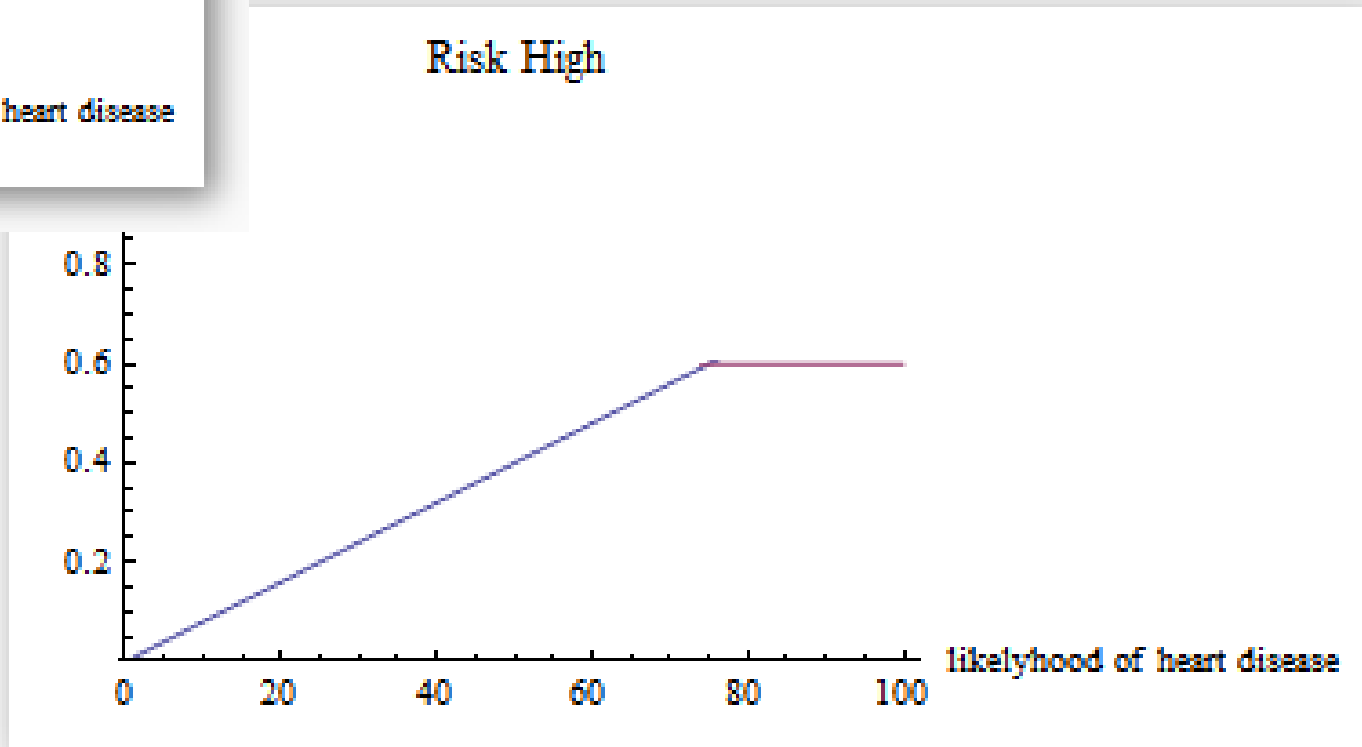


$$f_{risk\ high}(x) = \frac{1}{125}x$$

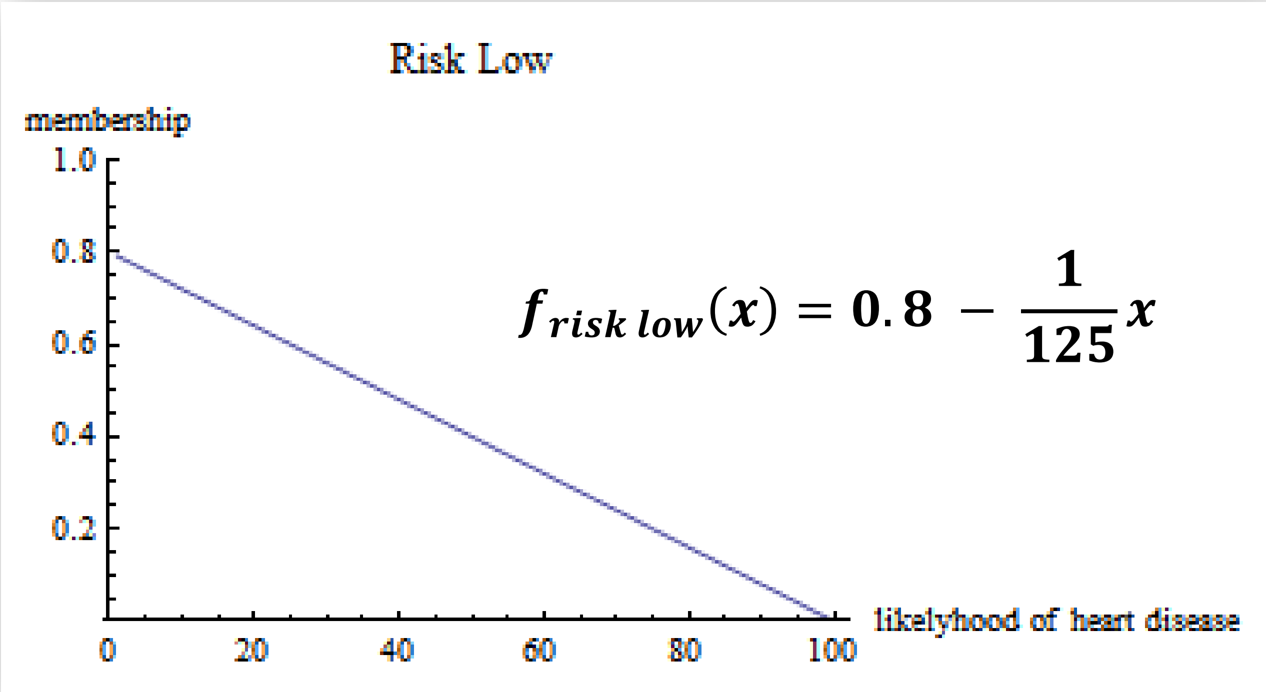
- Truth(Risk-High) = 0.6
- Therefore:

  0.6 = x / 125

  or, x = 75



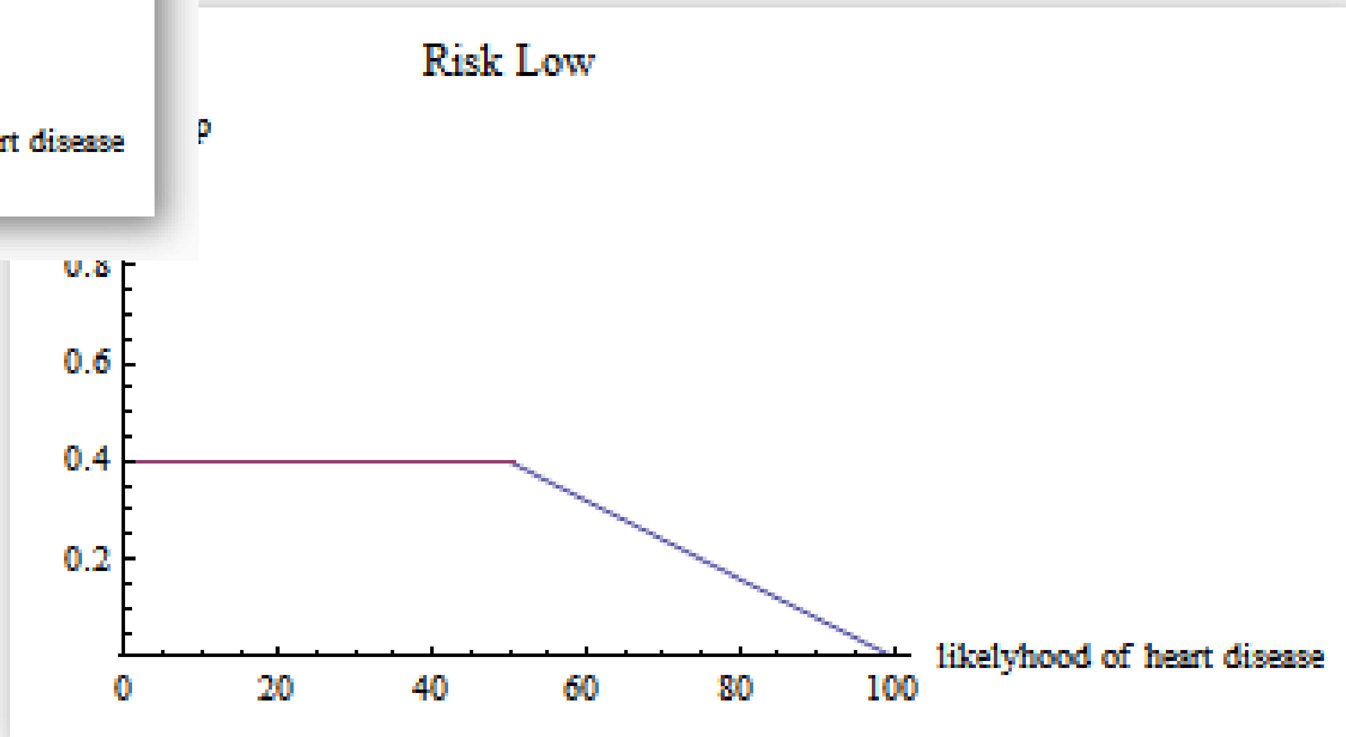
Risk-Low Evaluation



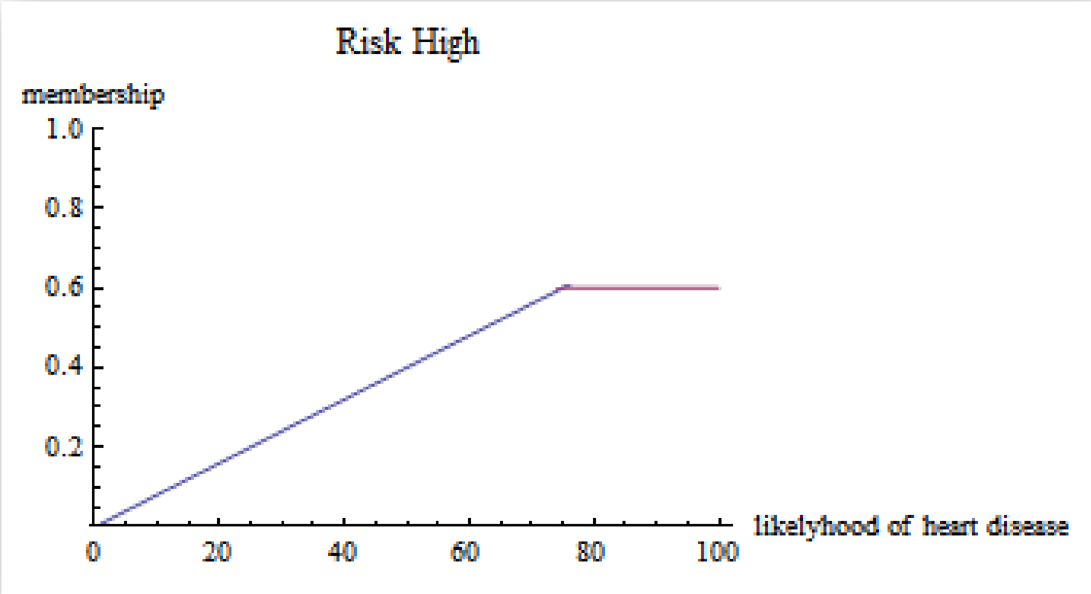
- Truth(Risk-Low) = 0.4
- Therefore:

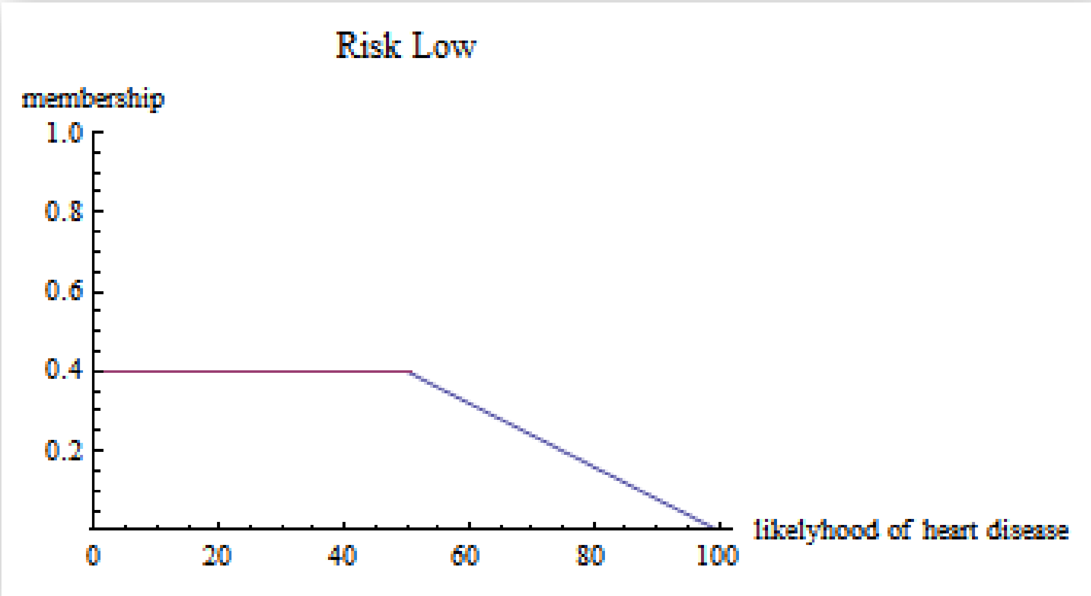
$$0.4 = 0.8 – x / 125$$

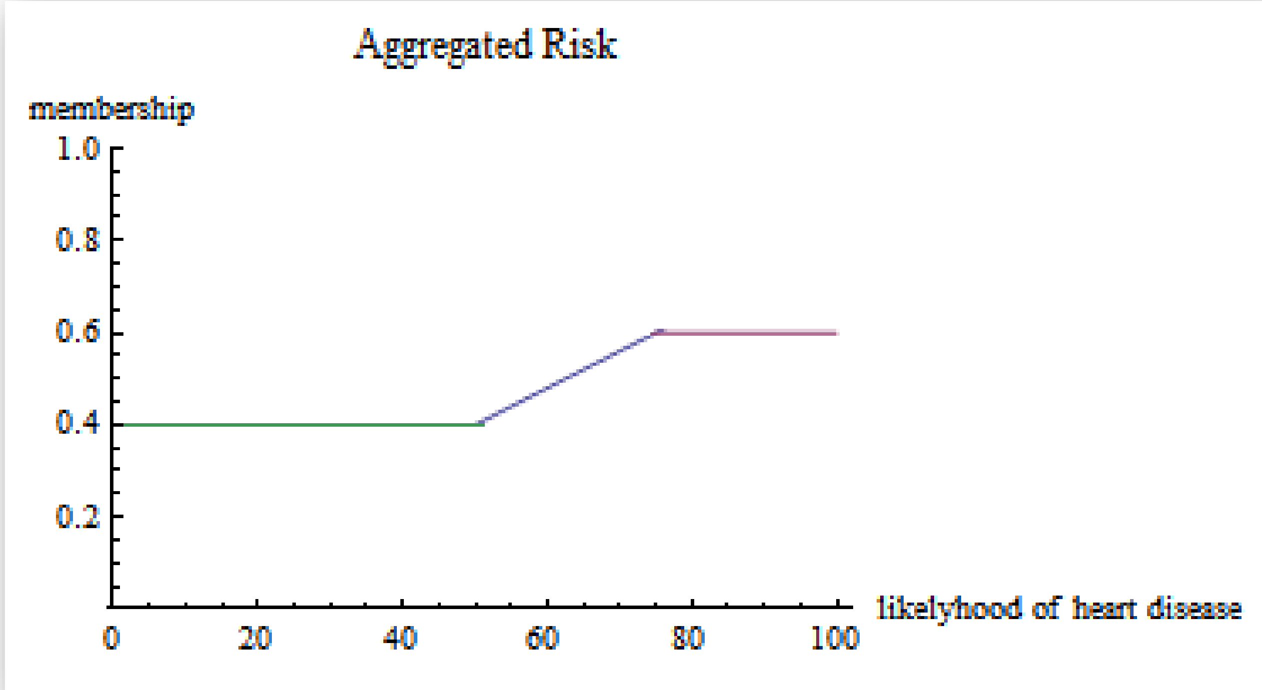
or, $x = 50$



Aggregated Risk Function







Defuzzification

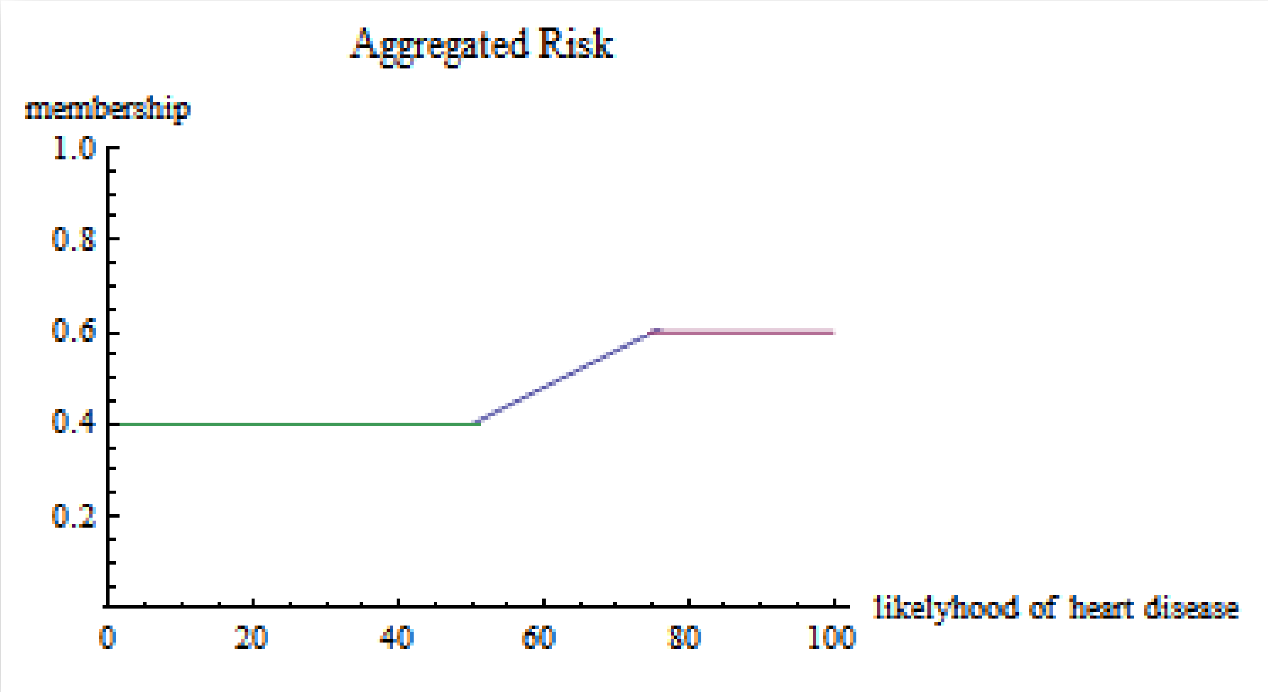
$$\int_0^{100} f_{aggregated\ risk}.\,dx$$

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

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

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

$$= 47.5$$



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)