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

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

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

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

Source Slides @AI slides-IIT KGP

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

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

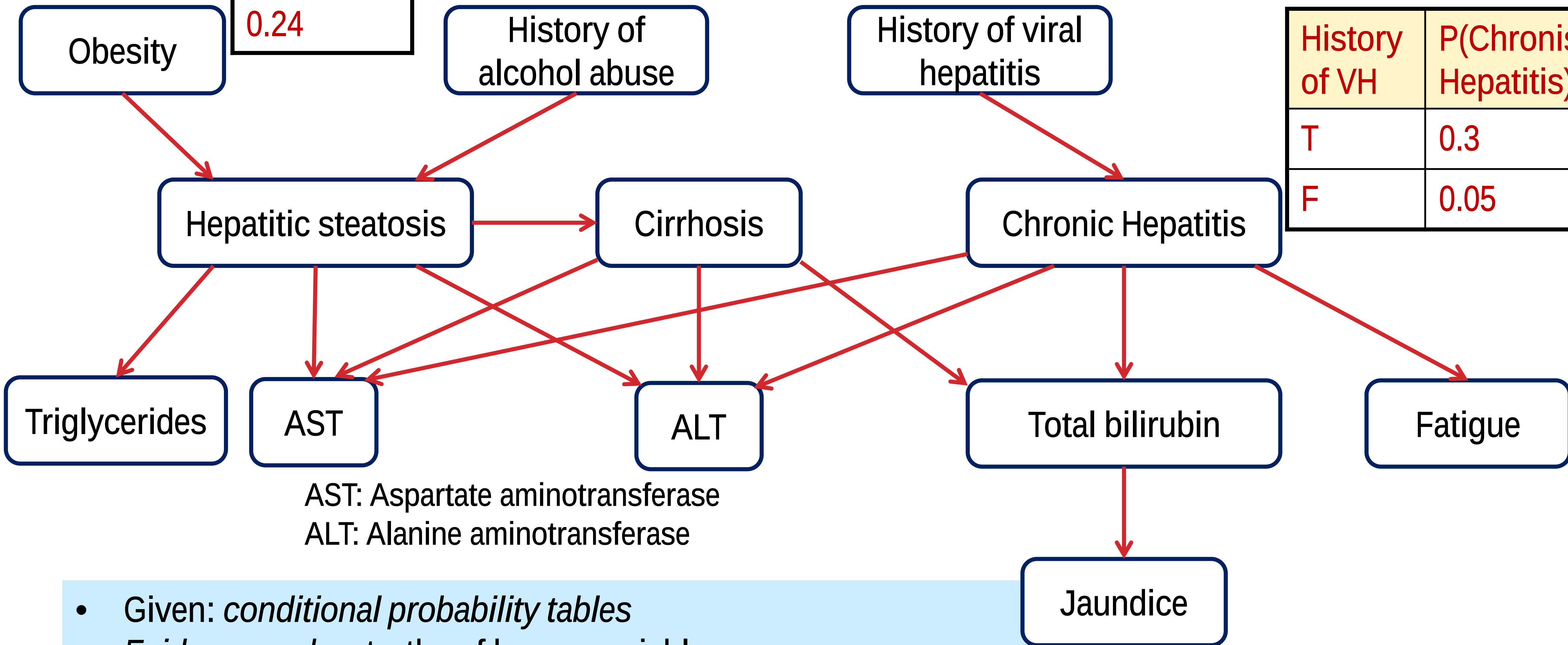
$$P(A \wedge B) = P(B \mid A)\, P(A)$$

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

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.

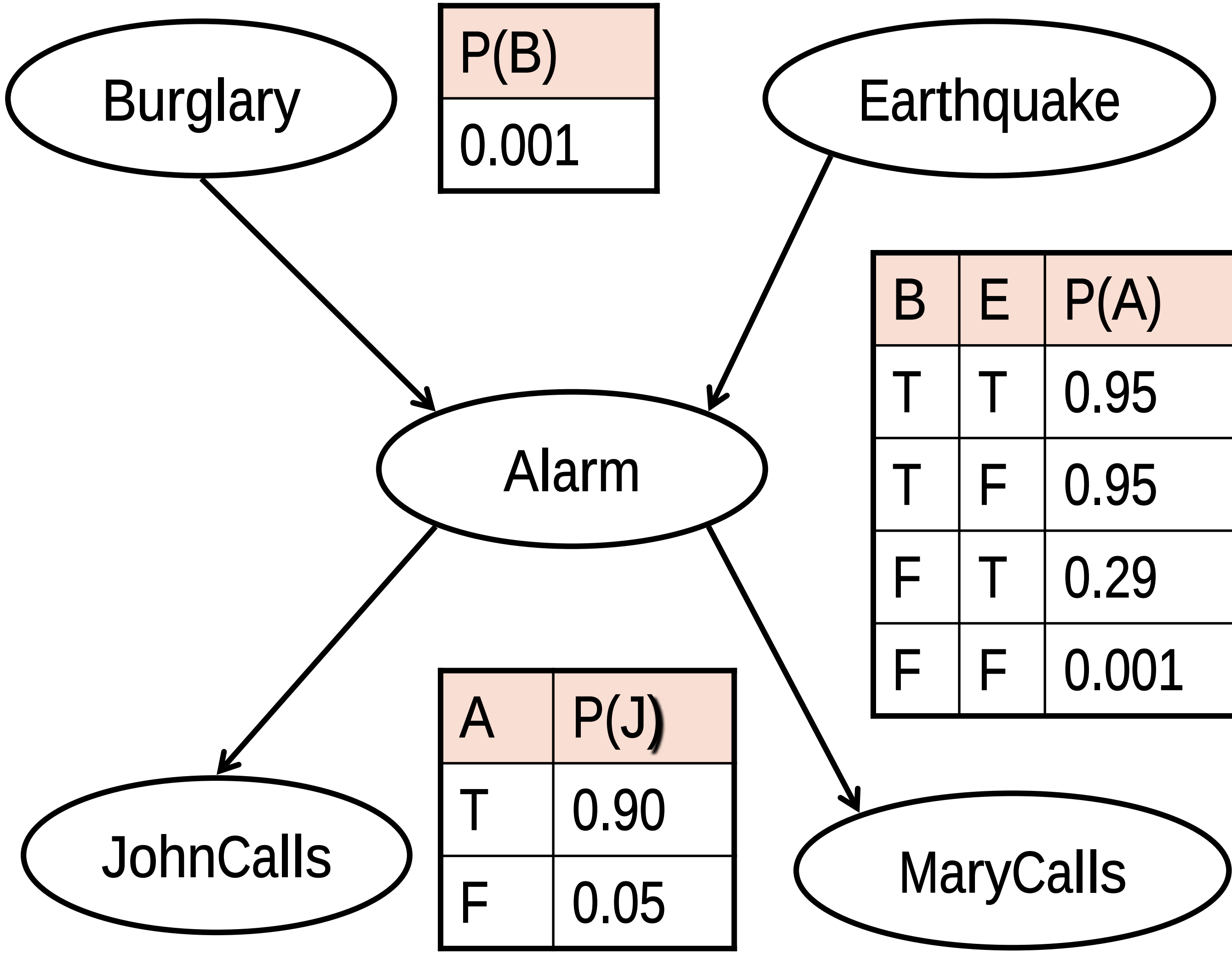
The graph has no directed cycles. It is a *directed acyclic graph* (DAG).

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

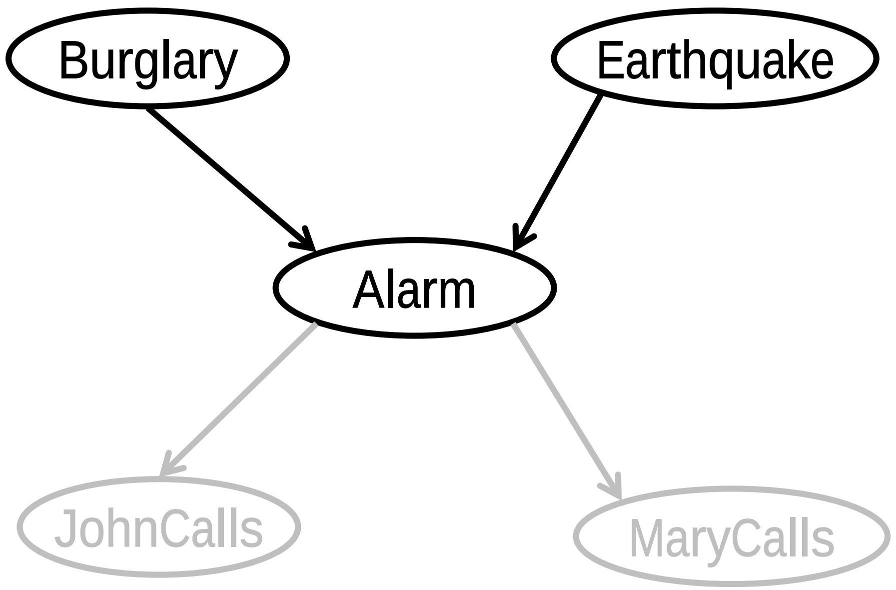
A	P(J)
T	0.90
F	0.05

A	P(M)
T	0.70
F	0.01

The joint probability distribution

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

Conditional probability

- Conditional or posterior probabilities
  e.g., P(*cavity* | *toothache*) = 0.8
  i.e., given that *toothache* is all I know there is 80% chance of cavity

- Notation for conditional distributions:
  P(*Cavity* | *Toothache*) = 2-element vector of 2-element vectors)

- If we know more, e.g., *cavity* is also given, then we have
  P(*cavity* | *toothache, cavity*) = 1

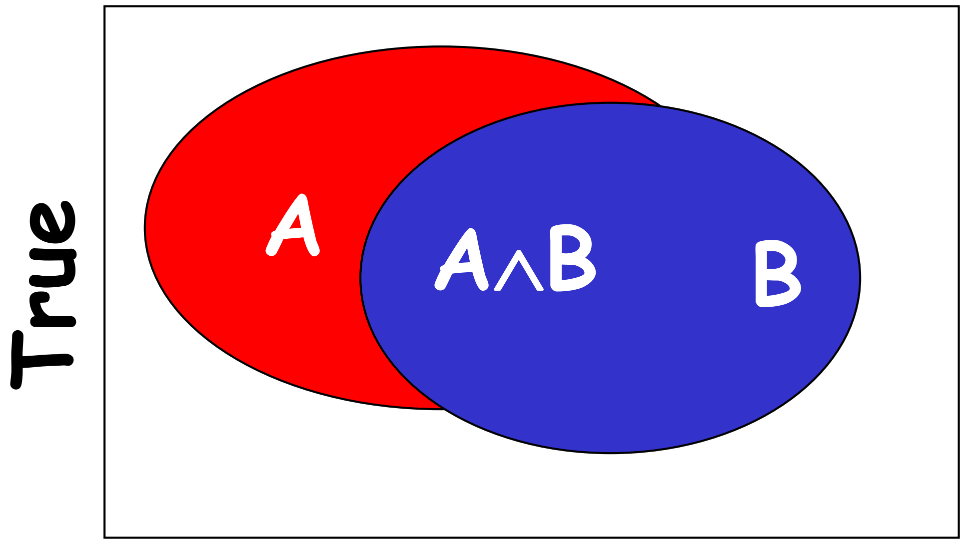
- New evidence may be irrelevant, allowing simplification:
  P(*cavity* | *toothache, sunny*) = P(*cavity* | *toothache*) = 0.8

- This kind of inference, sanctioned by domain knowledge, is crucial

Conditional Probability

- P(*A* | *B*) is the probability of *A* given *B*
- Assumes that *B* is the only info known.
- Defined by:

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



Independence

- A and B are *independent* iff:

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

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

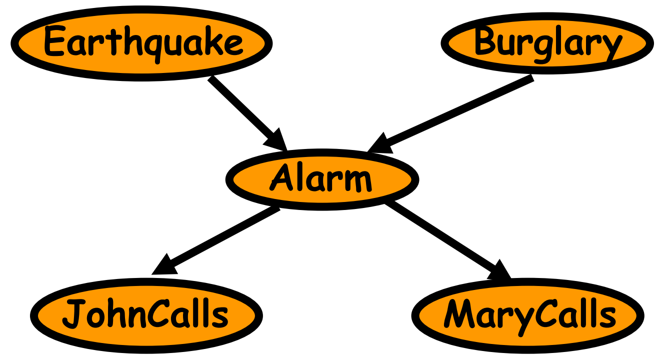
These two constraints are logically equivalent

- Therefore, if A and B are independent:

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

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

Earthquake Example (cont'd)

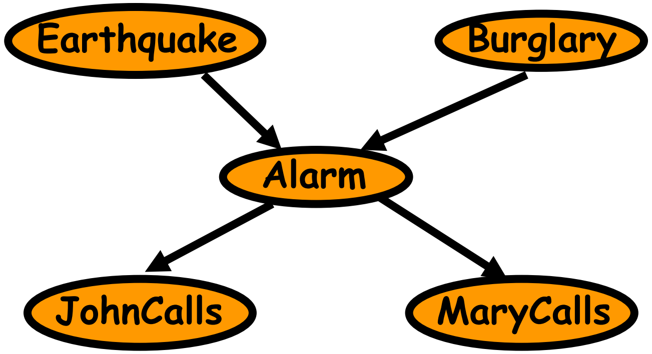


- **If we know *Alarm*, no other evidence influences our degree of belief in *JohnCalls***
  - $P(JC|MC,A,E,B) = P(JC|A)$
  - also: $P(MC|JC,A,E,B) = P(MC|A)$ and $P(E|B) = P(E)$
- **By the chain rule we have**

$$P(JC,MC,A,E,B) = P(JC|MC,A,E,B) \cdot P(MC|A,E,B) \cdot P(A|E,B) \cdot P(E|B) \cdot P(B)$$

$$= \mathbf{P(JC|A) \cdot P(MC|A) \cdot P(A|B,E) \cdot P(E) \cdot P(B)}$$

Earthquake Example (Global Semantics)



- **We just proved**

  $P(JC,MC,A,E,B) = \mathbf{P(JC|A)} \cdot \mathbf{P(MC|A)} \cdot \mathbf{P(A|B,E)} \cdot \mathbf{P(E)} \cdot \mathbf{P(B)}$

- **In general full joint distribution of a Bayes net is defined as**

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid Par(X_i))$$

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:

$$P(J \wedge M \wedge A \wedge \neg B \wedge \neg E)$$

$$= P(J \mid A)\ P(M \mid A)\ P(A \mid \neg B \wedge \neg E)\ P(\neg B)\ P(\neg E)$$

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

$$= 0.00062$$

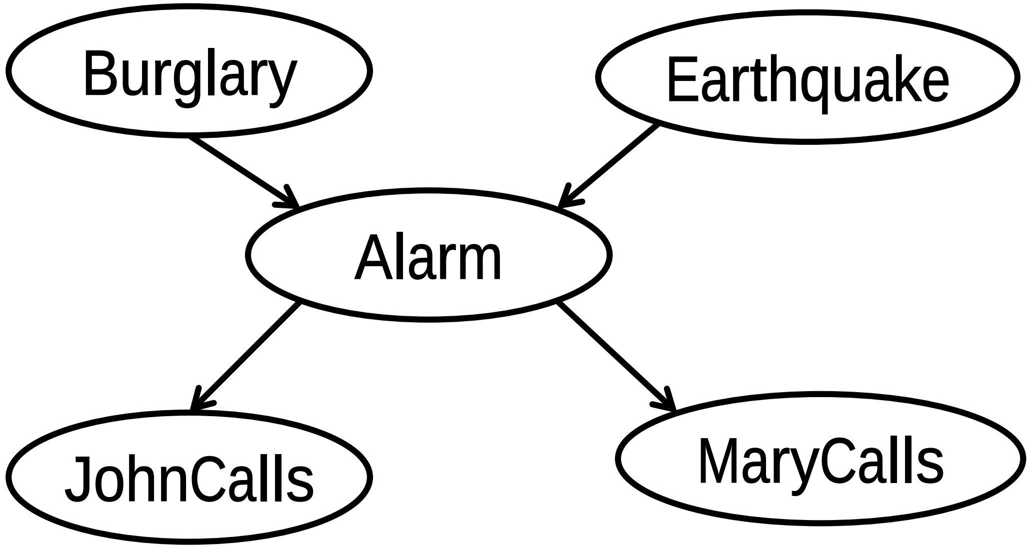
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

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

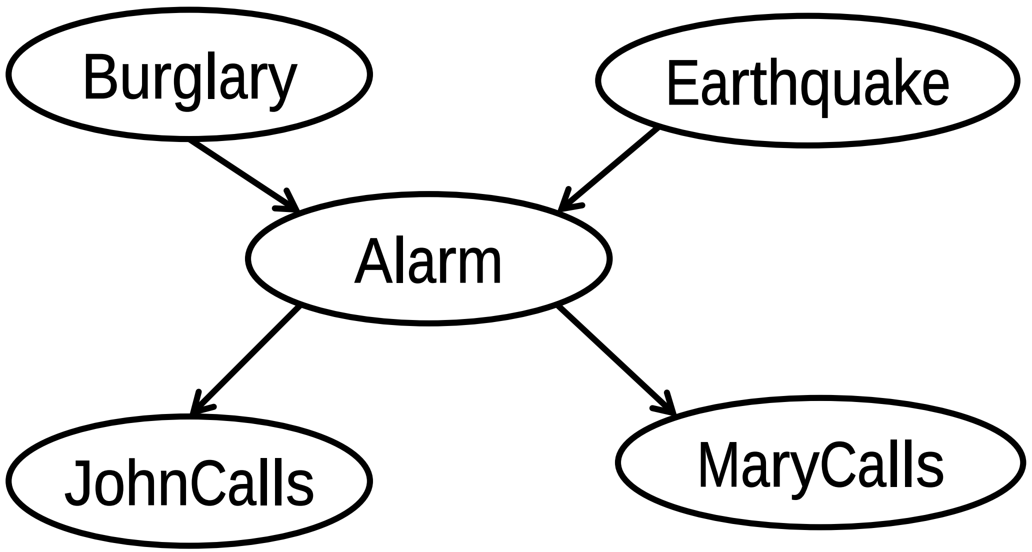
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

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:

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

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

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

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

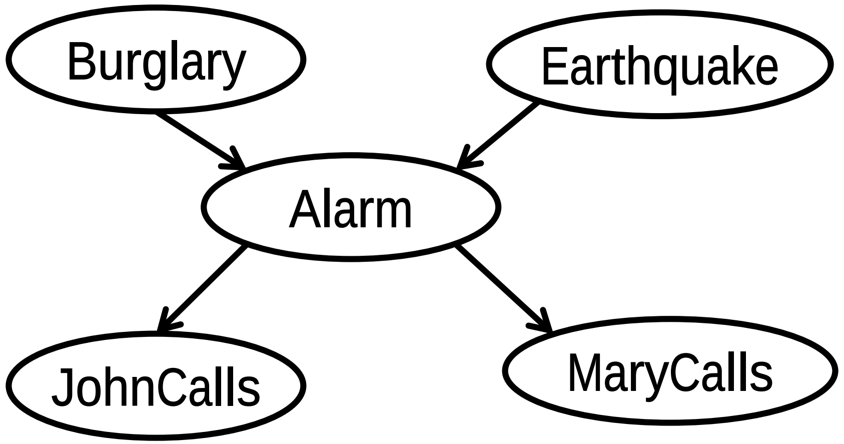
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

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)

$$P(J) = P(JA) + P(JA')$$
$$= P(J \mid A).P(A) + P(J \mid A').P(A')$$
$$= 0.9 \times 0.0025 + 0.05 \times (1 - 0.0025)$$
$$= 0.052125$$
$$P(AB) = P(ABE) + P(ABE') = 0.95 \times 0.001 \times 0.002 + 0.95 \times 0.001 \times 0.998$$
$$= 0.00095$$

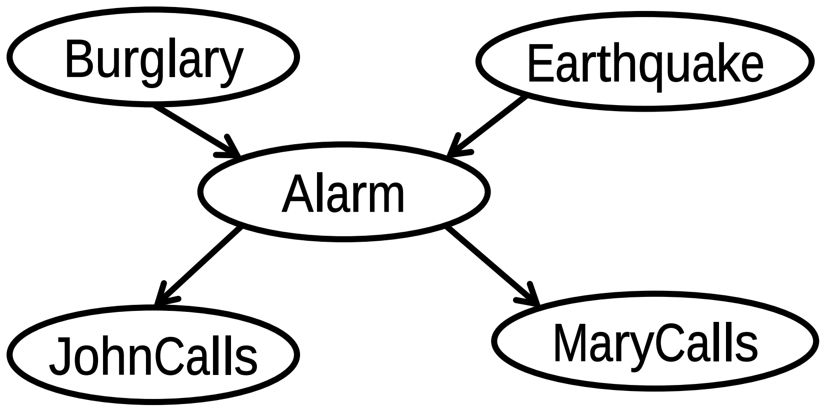
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

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’ | BE).P(BE) + P(A’ | BE’).P(BE’)

= (1 – 0.95) x 0.001 x 0.002

+ (1 – 0.95) x 0.001 x 0.998

= 0.00005

P(AE) = P(AEB) + P(AEB’)

= 0.95 x 0.001 x 0.002 + 0.29 x 0.999 x 0.002 = 0.00058

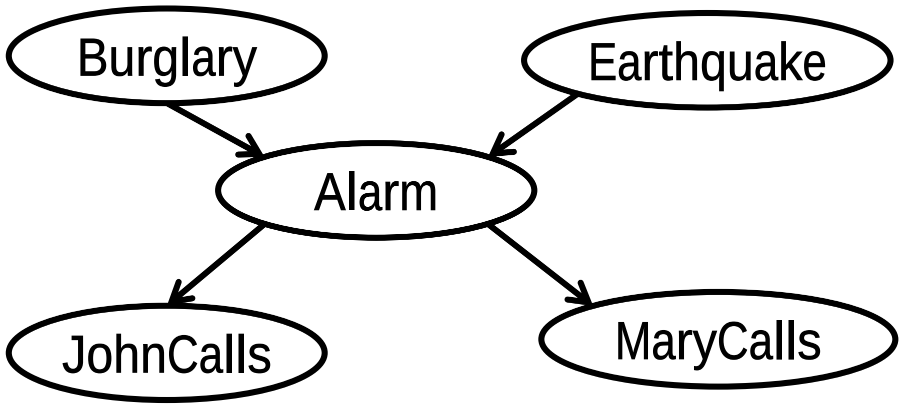
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

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

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

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

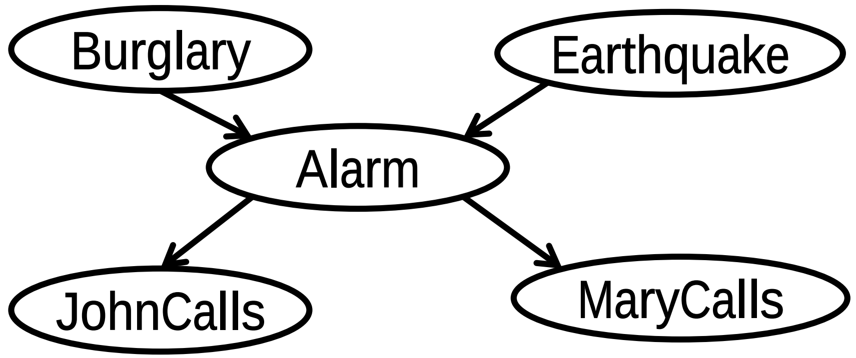
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

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

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

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

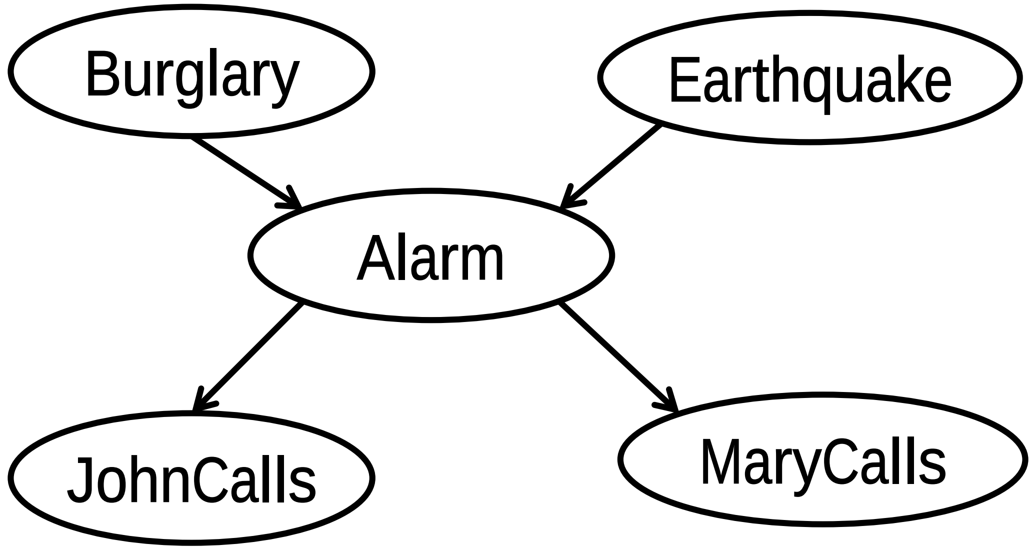
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

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

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

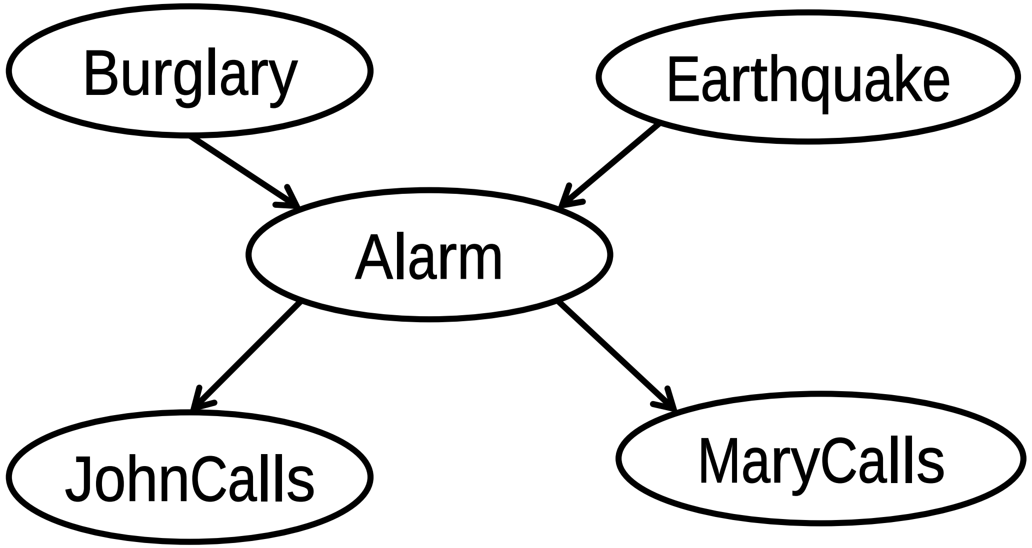
B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

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



Incremental Network Construction

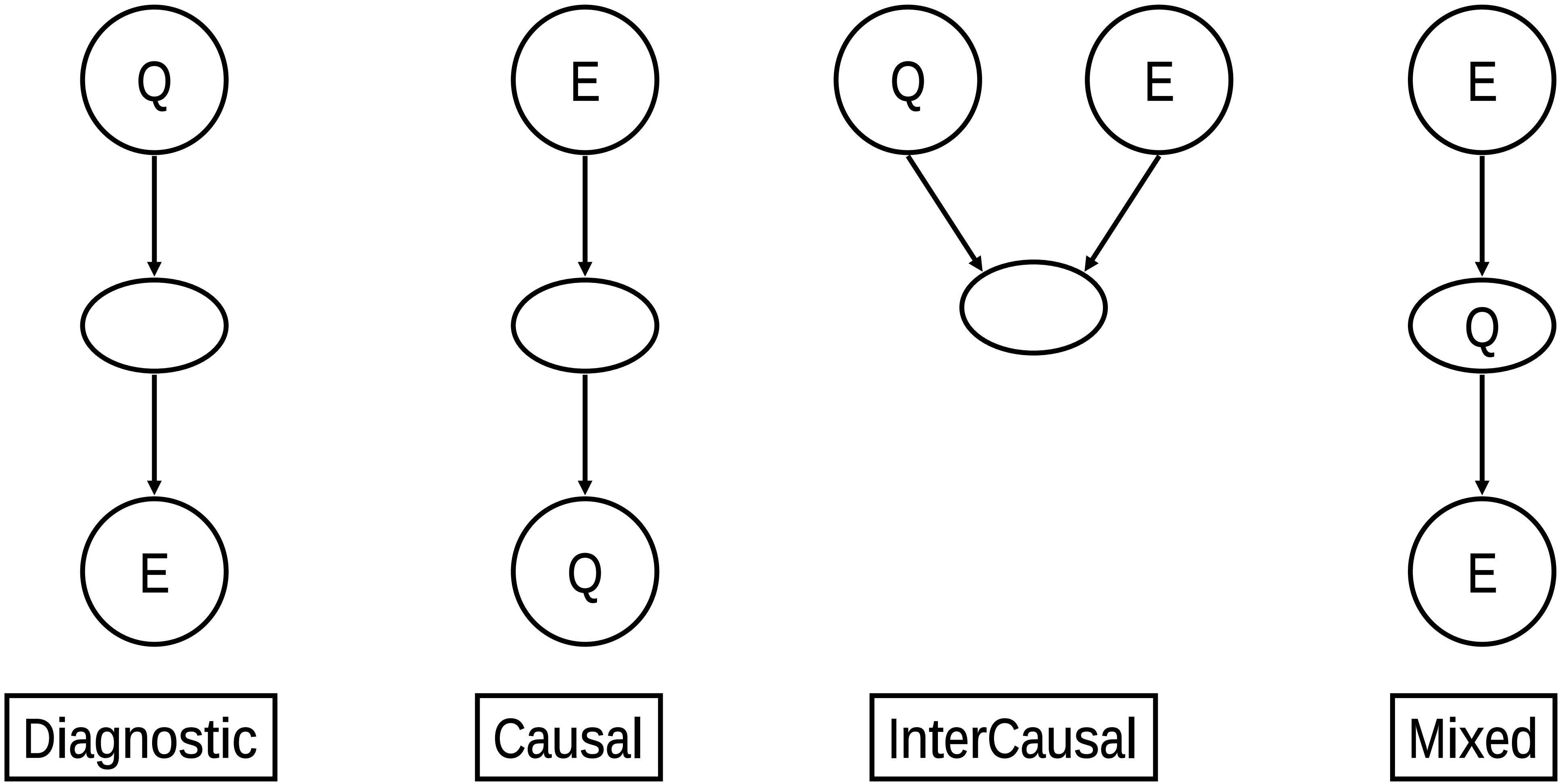
1. Choose the set of relevant variables X_i that describe the domain

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



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

- 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

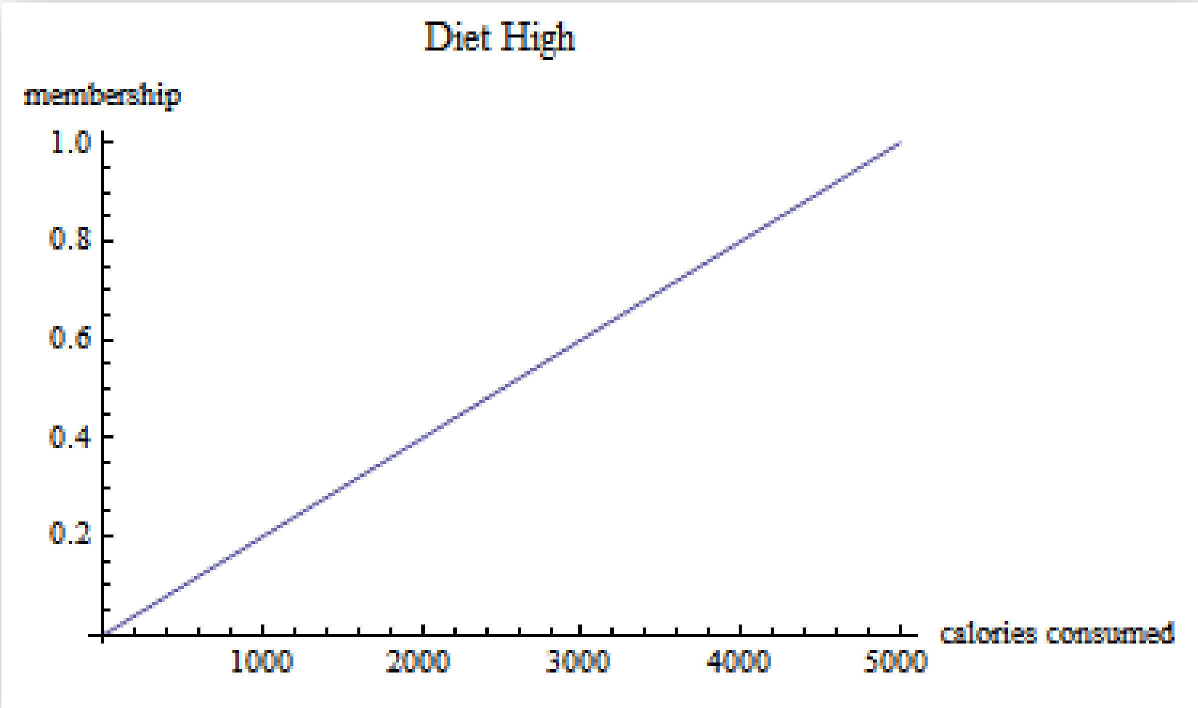
1. Diet is low AND Exercise is high $\Rightarrow$ Balanced
2. Diet is high OR Exercise is low $\Rightarrow$ Unbalanced
3. Balanced $\Rightarrow$ Risk is low
4. Unbalanced $\Rightarrow$ Risk is high

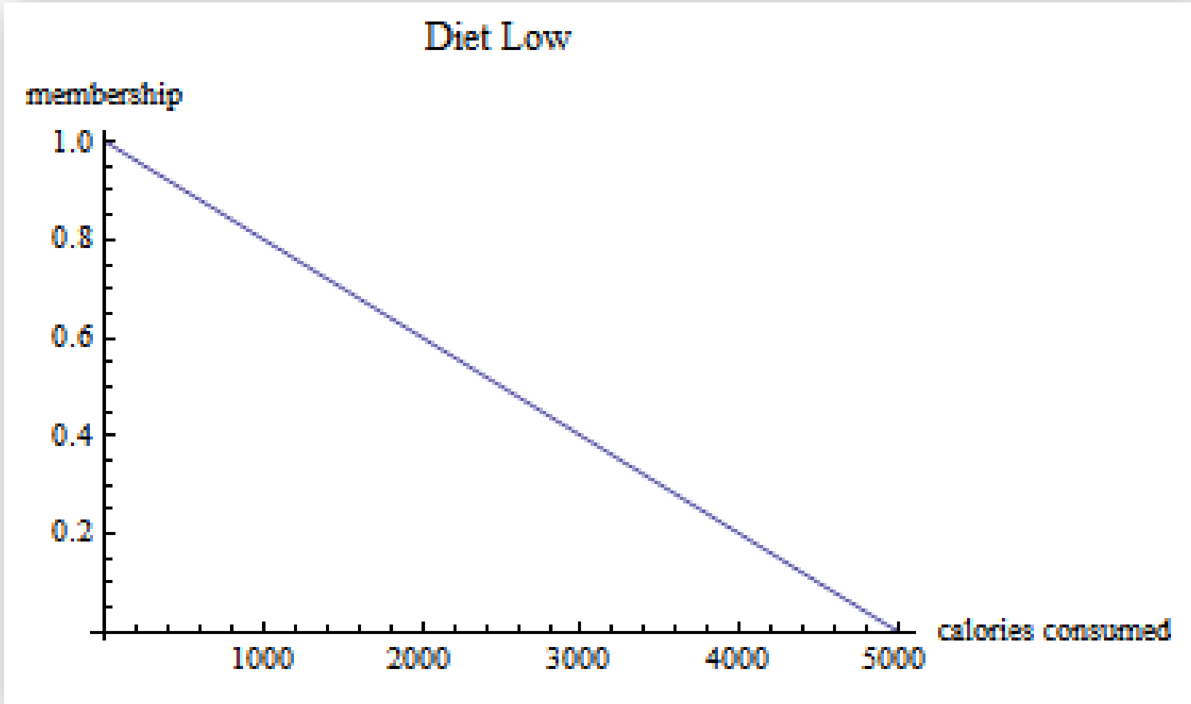
For a person it is given that:

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

What is the risk of heart disease?

Membership Functions





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

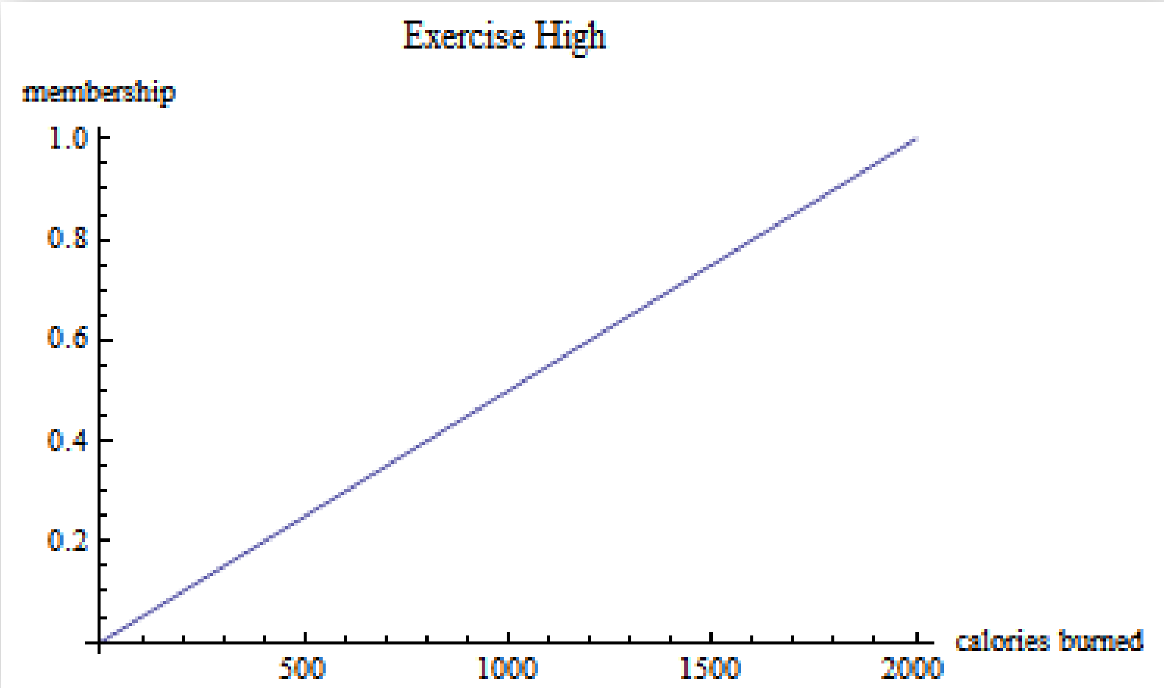
$$f_{diet\ low}(x) = 1 - \frac{1}{5000}x$$

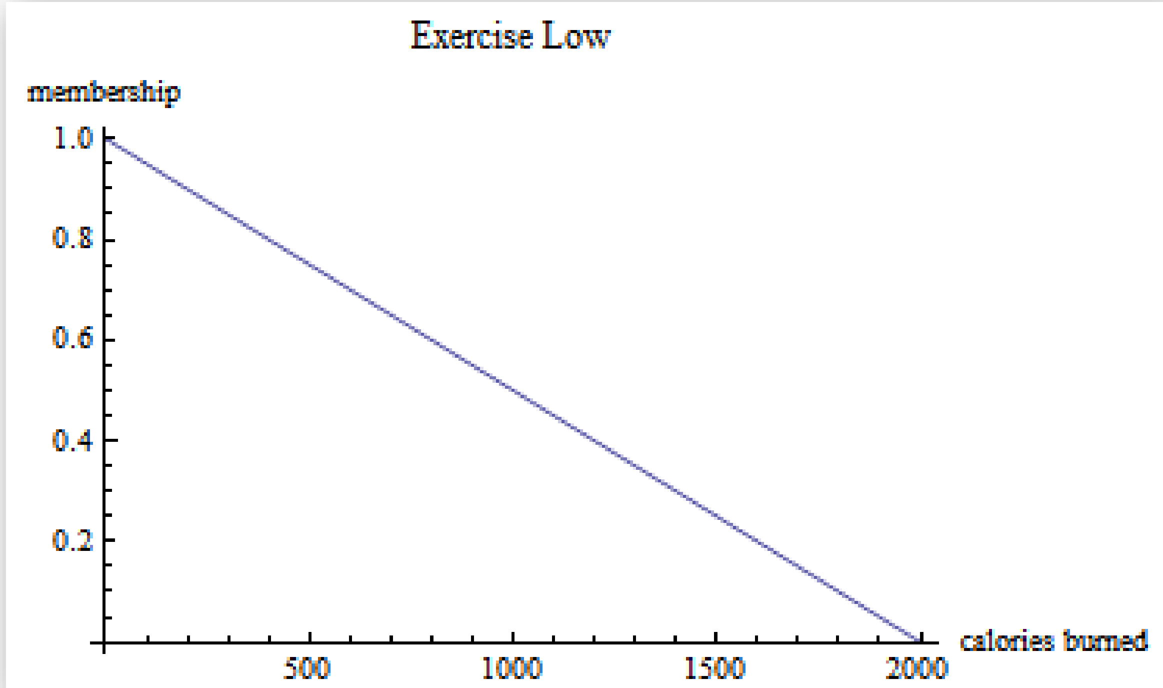
For daily calorie intake of 3000:

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 $\Rightarrow$ Balanced

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

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

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

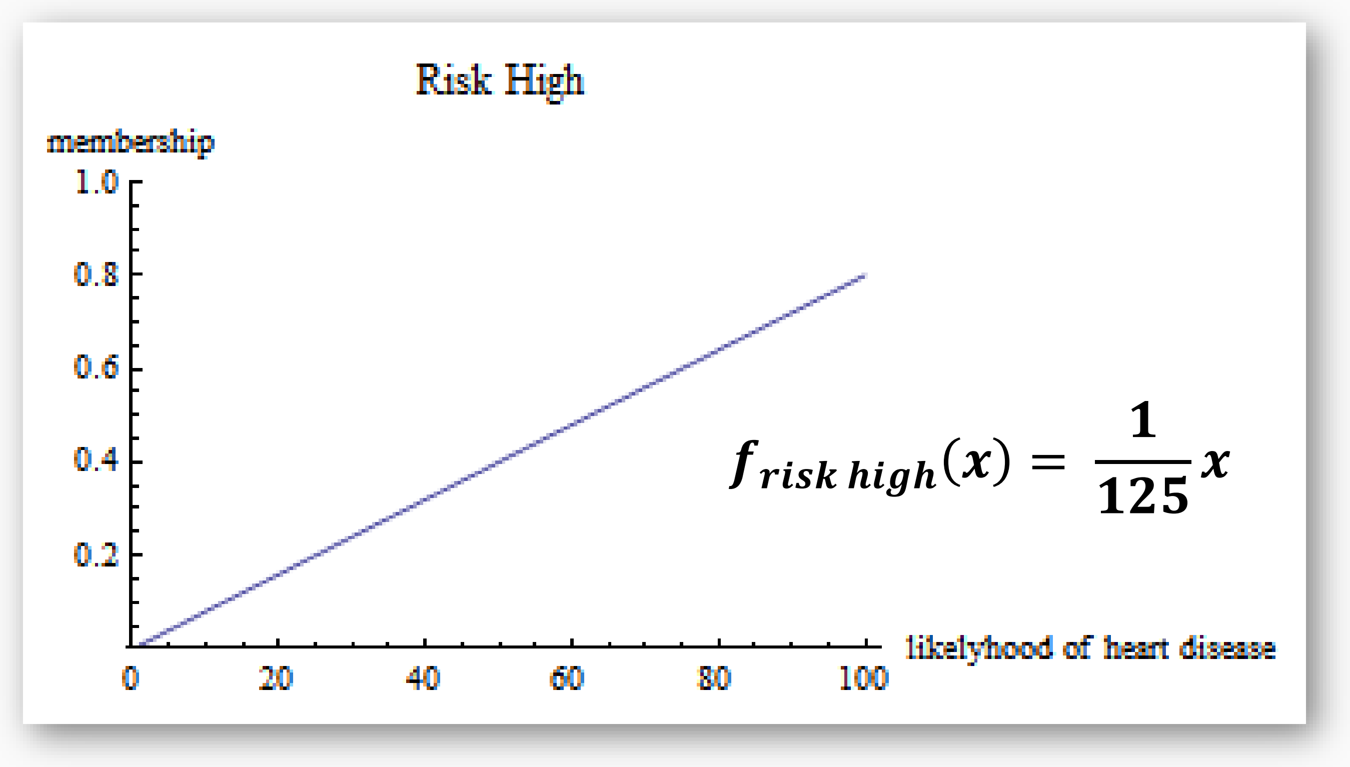
Balanced $\Rightarrow$ Risk is low

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

Unbalanced $\Rightarrow$ Risk is high

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

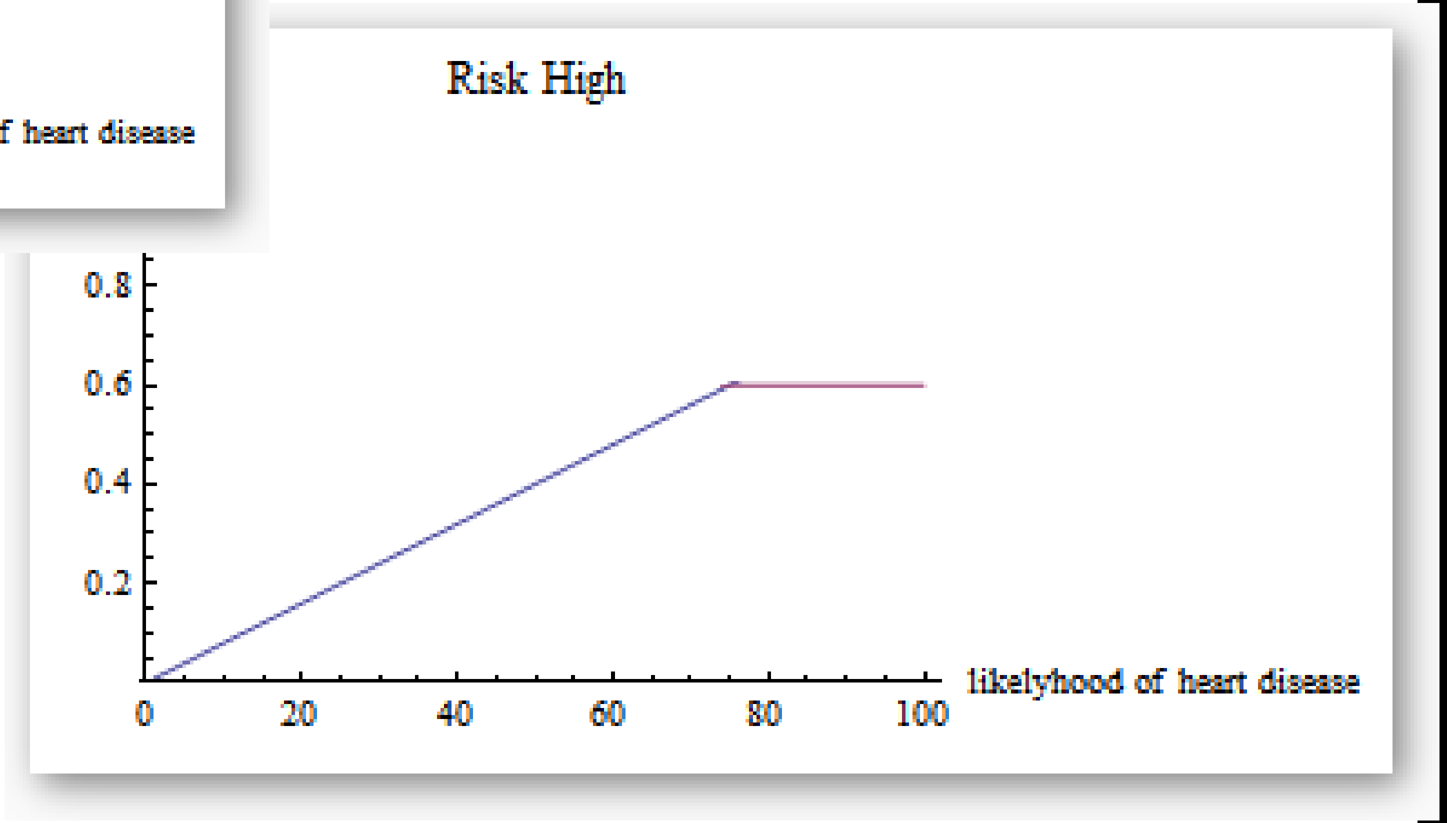
Risk-High Evaluation



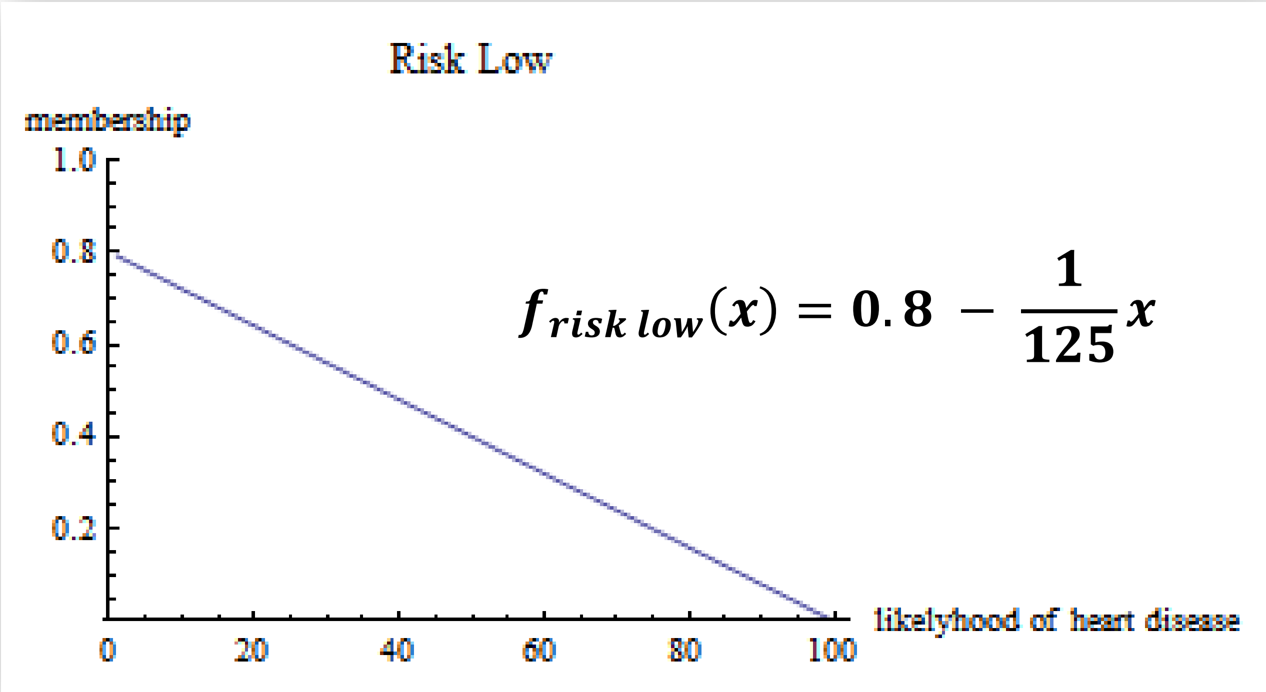
- Truth(Risk-High) = 0.6
- Therefore:

0.6 = x / 125

or, x = 75



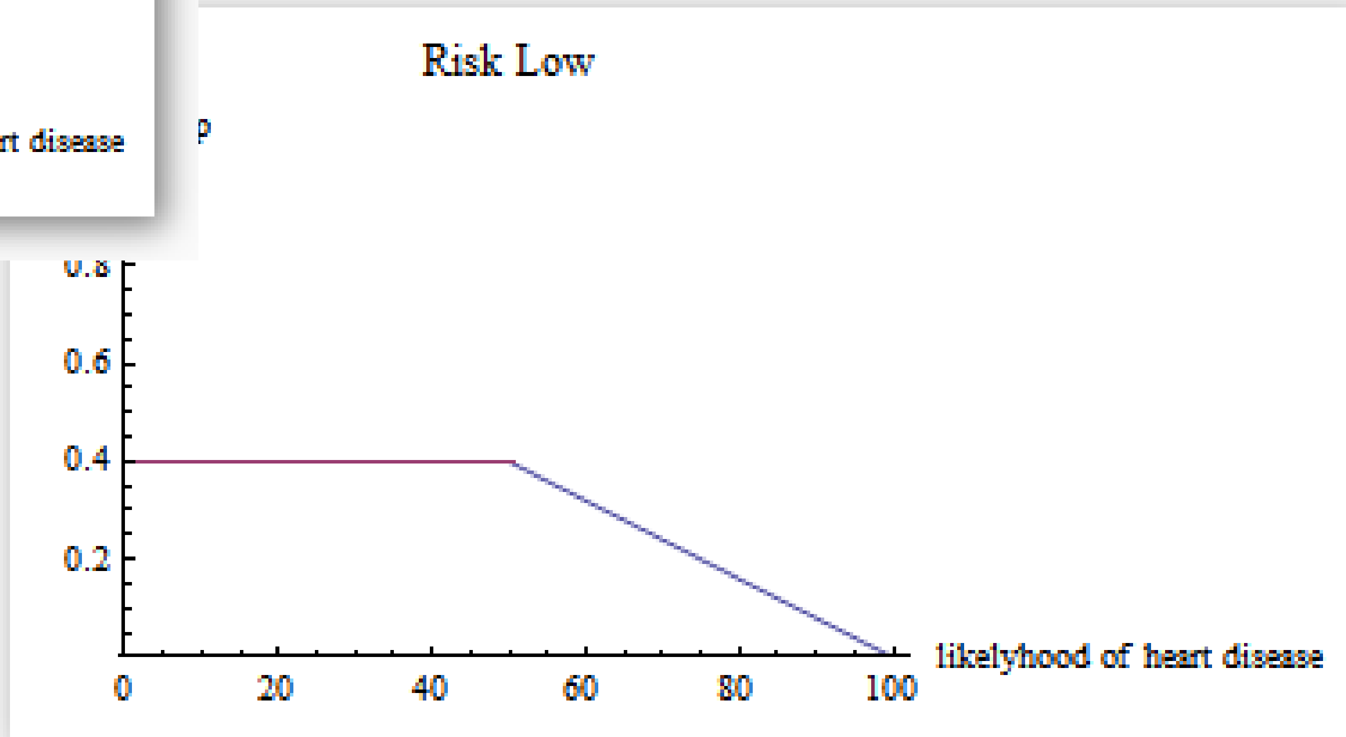
Risk-Low Evaluation



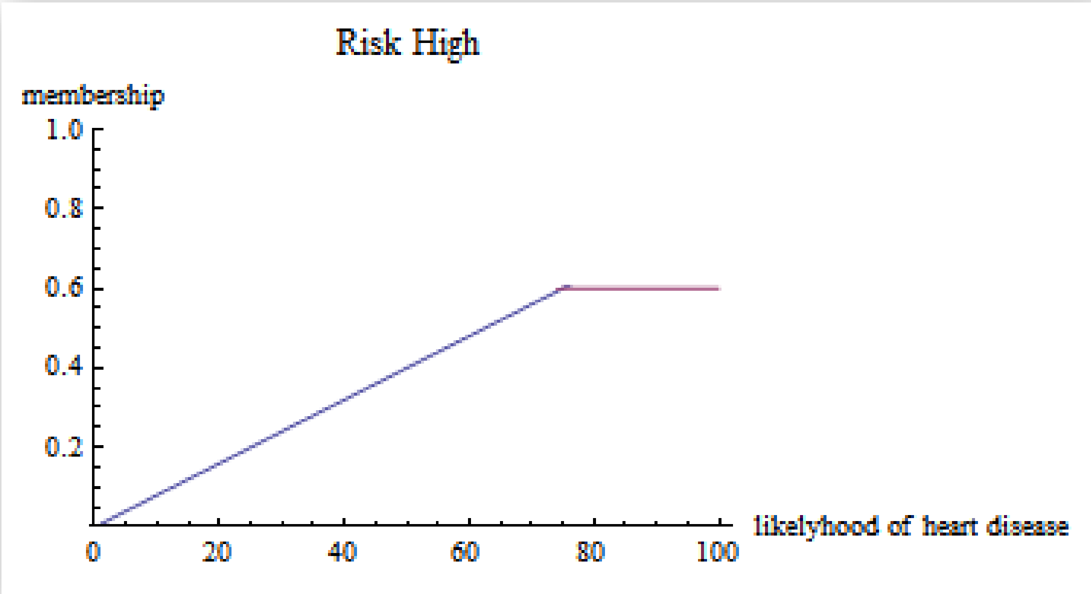
- Truth(Risk-Low) = 0.4
- Therefore:

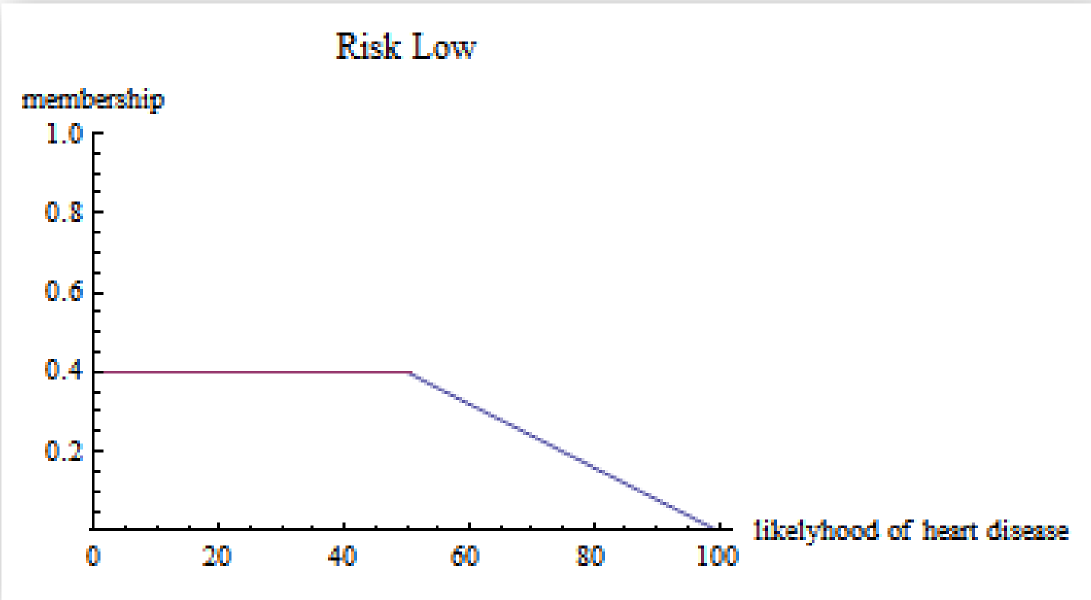
$$0.4 = 0.8 - x / 125$$

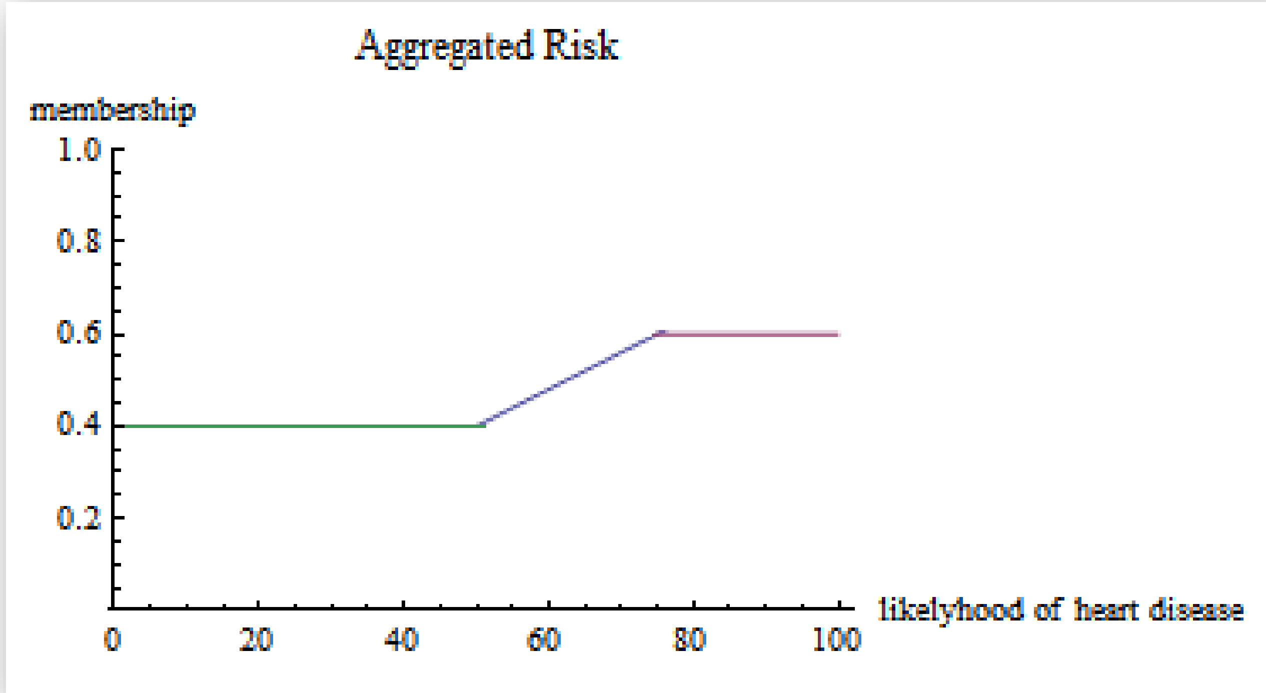
or, $x = 50$



Aggregated Risk Function







Defuzzification

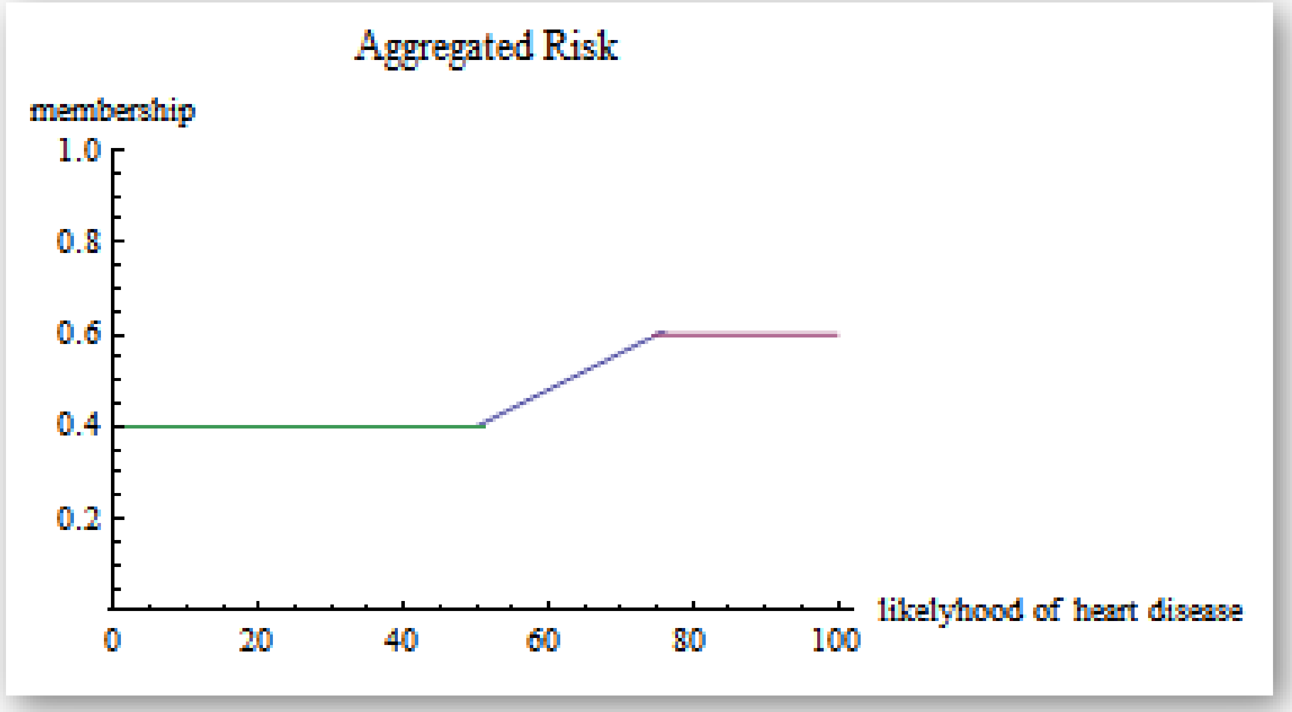
$$\int_0^{100} f_{aggregated\,risk}.\,dx$$

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

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

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

$$= 47.5$$



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