Uncertainty Measure

Most intelligent systems have some degree of uncertainty associated with them.

Uncertainty may occur in KBS because of the problems with the data.

- Data might be missing or unavailable.
- Data might be present but unreliable or ambiguous due to measurement errors, multiple conflicting measurements etc.
- The representation of the data may be imprecise or inconsistent.
- Data may just be expert's best guess.
- Data may be based on defaults and the defaults may have exceptions.

- Given numerous sources of errors, the most KBS requires the incorporation of some form of uncertainty management.

- For any form of uncertainty scheme, we must be concerned with three issues.
  - How to represent uncertain data?
  - How to combine two or more pieces of uncertain data?
  - How to draw inference using uncertain data?

- Probability is the oldest theory with strong mathematical basis.
- Other methods for handling uncertainty are **Bayesian belief network, Certainty factor theory ….**etc.

1.Probability Theory

- Probability is a way of turning opinion or expectation into numbers.
- It lies between 0 to 1 that reflects the likelihood of an event.

- The chance that a particular event will occur = the number of ways the event can occur divided by the total number of all possible events.

P(A)= (No. of outcomes favourable to A)/
(Total no. of possible outcomes)

Example: The probability of throwing two successive heads with a fair coin is 0.25

- Total of **four** possible outcomes are :
  HH, HT, TH & TT
- Since there is only one way of getting HH,
  probability = ¼ = 0.25

Axioms of Probability

- Let S be a sample space, A and B are events.
  - $P(A) \geq 0$
  - $P(S) = 1$
  - $P(A') = 1 - P(A)$
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - If events A and B are mutually exclusive, then

    $P(A \cup B) = P(A) + P(B)$,
- In general, for mutually exclusive events A1,…,An in S

  $P(A1 \cup A2 \cup \ldots \cup An) = P(A1) + P(A2) + \ldots + P(An)$

1.1.Joint Probability

- Joint Probability of the occurrence of two independent events is written as P (A and B) and is defined by

  $$P(A \text{ and } B) = P(A \cap B) = P(A) * P(B)$$

Example: We toss **two** fair coins separately.

Let $P(A) = 0.5$, Probability of getting Head of first coin

$P(B) = 0.5$, Probability of getting Head of second coin

- Probability (Joint probability) of getting Heads on both the coins is

  $= P(A \text{ and } B)$

  $= P(A) * P(B) = 0.5 \text{ X } 0.5 = 0.25$

- The probability of getting Heads on **one** or on **both** of the coins i.e. the union of the probabilities P(A) and P(B) is expressed as

$$P(A \cup B) = P(A) + P(B) – P(A \cap B)$$

$$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(A) * P(B)$$

$$= 0.5 \text{ X } 0.5 - 0.25$$

$$= 0.75$$

1.2.Conditional Probability

- It relates the probability of one event to the occurrence of another i.e. probability of the occurrence of an event H given that an event E is known to have occurred.
- Probability of an event H (Hypothesis), given the occurrence of an event E (evidence) is denoted by $P(H \mid E)$ and is defined as follows:

$$P(H \mid E) = \frac{\text{Number of events favorable to H which are also favorable to E}}{\text{No. of events favorable to E}}$$

$$= \frac{P(H \text{ and } E)}{P(E)}$$

Examples

- What is the probability of a person to be male if person chosen at random is 80 years old?
- The following probabilities are given
  - Any person chosen at random being male is about 0.50
  - probability of a given person be 80 years old chosen at random is equal to 0.005
  - probability that a given person chosen at random is both male and 80 years old may be =0.002

- The probability that an 80 years old person chosen at random is male is calculated as follows:

  P(X is male | Age of X is 80)

  = [P(X is male and the age of X is 80)] / [P(Age of X is 80)]

  = 0.002 / 0.005 = 0.4

1.3.Bayes' Theorem

- Bayes theorem provides a mathematical model for this type of reasoning where prior beliefs are combined with evidence to get estimates of uncertainty.
- This approach relies on the concept that one should incorporate the prior probability of an event into the interpretation of a situation.
- It relates the conditional probabilities of events.

- It allows us to express the probability P(H | E) in terms of the probabilities of P(E|H), P(H) and P(E).

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)},$$

H=hypothesis
E=evidence

- P(E|H)=0.50, P(H)=0.25 and P(E)=0.40, i.e P(H/E)=0.3125

Proof of Bayes' Theorem

- Bayes' theorem is derived from conditional probability.

Proof: Using conditional probability

$$P(H|E) = P(H \text{ and } E) \;/\; P(E)$$

$$\Rightarrow \quad P(H|E) * P(E) = P(H \text{ and } E) \qquad (1)$$

Also $$P(E|H) = P(E \text{ and } H) \;/\; P(H)$$

$$\Rightarrow \quad P(E|H) * P(H) = P(E \text{ and } H) \qquad (2)$$

From Equation (1) and (2), we get

$$P(H|E) * P(E) = P(E|H) * P(H)$$

Hence, we obtain

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

1.4. Extension of Bayes' Theorem

- Consider one hypothesis H and two evidences E1 and E2.
- The probability of H if both E1 and E2 are true is calculated by using the following formula:

$$P(H|E1 \text{ and } E2) = \frac{P(E1|H) * P(E2|H) * P(H)}{P(E1 \text{ and } E2)}$$

- Consider one hypothesis H and Multiple evidences E1,…., En.
- The probability of H if E1,…, En are true is calculated by using the following formula:

$$P(H|E1 \text{ and } \dots \text{ and } En) = \frac{P(E1|H) * \dots * P(En|H) * P(H)}{P(E1 \text{ and } \dots \text{ and } En)}$$

Example

- Find whether Bob has a cold (hypotheses) given that he sneezes (the evidence) i.e., calculate P(H | E).

- Suppose that we know / given the following.

P(H) = P (Bob has a cold) = 0.2

P(E | H)= P(Bob was observed sneezing | Bob has a cold) = 0.75

P(E | ~H)= P(Bob was observed sneezing | Bob does not have a cold) = 0.2

Now

P(H | E) = P(Bob has a cold | Bob was observed sneezing)

= [P(E | H) * P(H)] / P(E)

Now calculate P(E) value as follows….

- We can compute P(E) as follows:

$$
\begin{aligned}
P(E) &= P(E \text{ and } H) + P(E \text{ and } \sim H) \\
&= P(E \mid H) * P(H) + P(E \mid \sim H) * P(\sim H) \\
&= (0.75)(0.2) + (0.2)(0.8) = 0.31
\end{aligned}
$$

  - Hence $P(H \mid E) = [(0.75 * 0.2)] / 0.31 = 0.48387$ or 0.5

  Note: $P(H)=0.2$ then $P(\sim H)=1-0.2=0.8$

  - We can conclude that "Bob's probability of having a cold given that he sneezes" is about 0.5

- Further it can also determine what is his probability of having a cold if he was not sneezing?

$$
\begin{aligned}
P(H \mid \sim E) &= [P(\sim E \mid H) * P(H)] / P(\sim E) \\
&= [(1 - 0.75) * 0.2] / (1 - 0.31) \\
&= 0.05 / 0.69 = 0.072
\end{aligned}
$$

  - Hence "Bob's probability of having a cold if he was not sneezing" is 0.072

2.Bayesian Belief Networks

- Joint probability distribution for n variables require 2^n entries with all possible combinations.
- The time and storage requirements for such computations become impractical as n grows.
- Joint probability distribution of two variables A and B are given in the following Table

Joint Probability	A	A’
B	0.20	0.12
B’	0.65	0.03

- Inferring with such large numbers of probabilities does not seem to model human process of reasoning.

- Human tends to single out few propositions which are known to be causally linked when reasoning with uncertain beliefs.

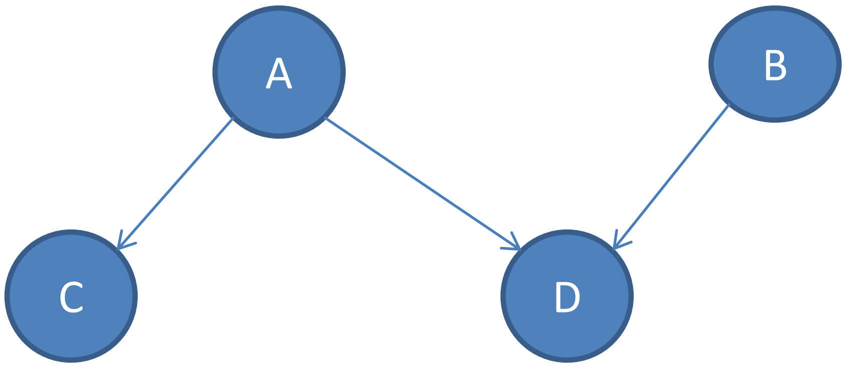
- This leads to the concept of forming belief network called a Bayesian belief network.

- It is a probabilistic graphical model that encodes probabilistic relationships among set of variables with their probabilistic dependencies.

- This belief network is an efficient structure for storing joint probability distribution.

Example

- The following graph is a Bayesian belief network.
  - Here there are four nodes with E={A, B} representing evidences and H= {C, D} representing hypotheses.
  - A and B are unconditional nodes and C and D are conditional nodes.
- Fig: Bayesian Belief Network



To describe above Bayesian network, we should specify the following probabilities

$P(A)$	=	0.3
$P(B)$	=	0.6
$P(C \mid A)$	=	0.4
$P(C \mid \sim A)$	=	0.2
$P(D \mid A, B)$	=	0.7
$P(D \mid A, \sim B)$	=	0.4
$P(D \mid \sim A, B)$	=	0.2
$P(D \mid \sim A, \sim B)$	=	0.01

- They can also be expressed as conditional probability tables as follows:

Conditional Probability Tables

P(A)	P(B)
0.3	0.6

A	P(C)
T	0.4
F	0.2

A	B	P(D)
T	T	0.7
T	F	0.4
F	T	0.2
F	F	0.01

●Using Bayesian belief network on previous slide, only 8 probability values in contrast to 16 values are required in general for 4 variables {A, B, C, D} in joint distribution probability.

●Joint probability using Bayesian Belief Network is computed as follows:

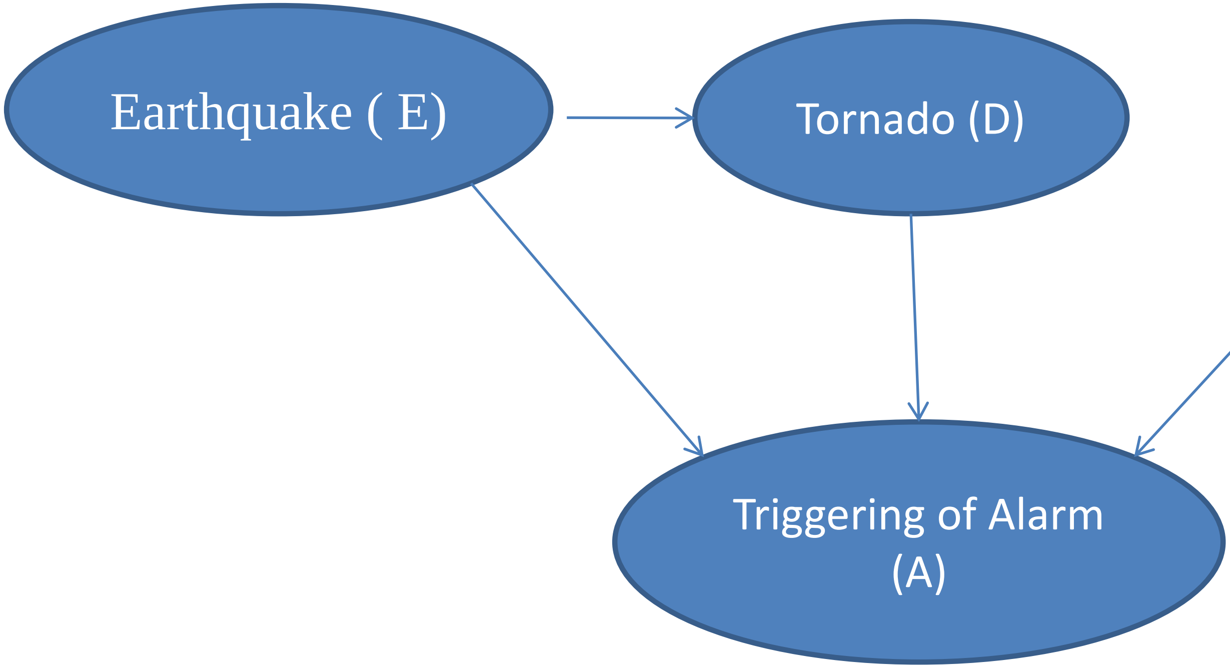
$$P(A, B, C, D) = P(D|A, B) * P(C|A) * P(B) * P(A)$$

$$= 0.7 * 0.4 * 0.6 * 0.3 = 0.0504$$

Example of Simple B-Network

- Suppose that there are three events namely earthquake, burglary or tornado which could cause ringing of alarm in a house.
- This situation can be modeled with Bayesian network as follows.
- All four variables have two possible values T (for true) and F (for false).
  - Here the names of the variables have been abbreviated to A = *Alarm*, E = *Earthquake*, and B = *Burglary* and T = *Tornado*.

Fig: Bayesian Network



●Table contains the probability values representing complete Bayesian belief network.

●Prior probability of ‘earthquake’ is 0.4 and if it is earthquake then probability of ‘tornado’ is 0.8. and if not then the probability of ‘earthquake’ is 0.5

Conditional Probability Tables

P(E)	P(B)
0.4	0.7

E	P(D)
T	0.8
F	0.5

E	B	D	P(A)
T	T	T	1.0
T	T	F	0.9
T	F	T	0.95
T	F	F	0.85
F	T	T	0.89
F	T	F	0.7
F	F	T	0.87
F	F	F	0.3

- The joint probability is computed as follows:

  $P(E, B, T, A) = P(A| E, B, T) * P(T|E) * P(E) * P(B)$

  $= 1.0 * 0.8 * 0.4 * 0.7 = 0.214$

- Using this model one can answer questions using the conditional probability formula as follows:
  - "What is the probability that it is earthquake, given the alarm is ringing?" $P(E|A)$
  - "What is the probability of burglary, given the alarm is ringing?" $P(B|A)$
  - "What is the probability of ringing alarm if both earthquake and burglary happens?" $P(A|E, B)$

Fuzzy Set

- Zadeh developed the concept of 'fuzzy sets' in mid 60's to account for numerous concepts used in human reasoning which are vague and imprecise e.g. tall, old
- Fuzzy set is very convenient method for representing some form of uncertainty.
- Later Zadeh developed 'fuzzy logic' to account for the imprecision of natural language quantities e.g. (many) and statements (e.g. not very likely).
- In Fuzzy logic, a statement can be both true or false and also can be neither true nor false. Fuzzy logic is non monotonic logic.
- Law of excluded middle does not hold true in fuzzy logic.
  - $A \vee \sim A = True$; $A \wedge \sim A = False$ -- do not hold

- Well known paradoxes can not be solved using classical logic.
- **Russell's paradox**
  - "All of the men in this town either shaved themselves or were shaved by the barber.
  - And the barber only shaved the men who did not shave themselves"
  - Answer to question: " Who shaves the barber ? " is contradictory
  - Assume that **he did shave himself**. But we see from the story that he shaved only those men who did not shave themselves. Therefore, **he did not shave himself**.
  - But we notice that every man either shaved himself or was shaved by the barber. So **he did shave himself**. We have a contradiction.

Example - Paradox

- “ All Cretans are liars”, said the Cretan
- If the Cretan is liar then his claim can not be believed and so is not a liar.
- If he is not liar then he is telling truth. But because he is Cretan, he must therefore a liar.
- The main idea behind Fuzzy systems is that truth values (in fuzzy logic) or membership values are indicated by a value in the range [0,1] with 0 for absolute false and 1 for absolute truth.
- Fuzzy sets are often incorrectly assumed to indicate some form of probability.
- Even though they can take on similar values, it is important to realize that membership grades are not probabilities.

Example

- Represent "Helen is old" using probability theory and fuzzy set. Assume that Helen's age is 75.

Probability approach:

- We may assign the statement "Helen is old" the truth value of 0.95. The interpretation is that there is 95% chance of Helen is old

Fuzzy approach:

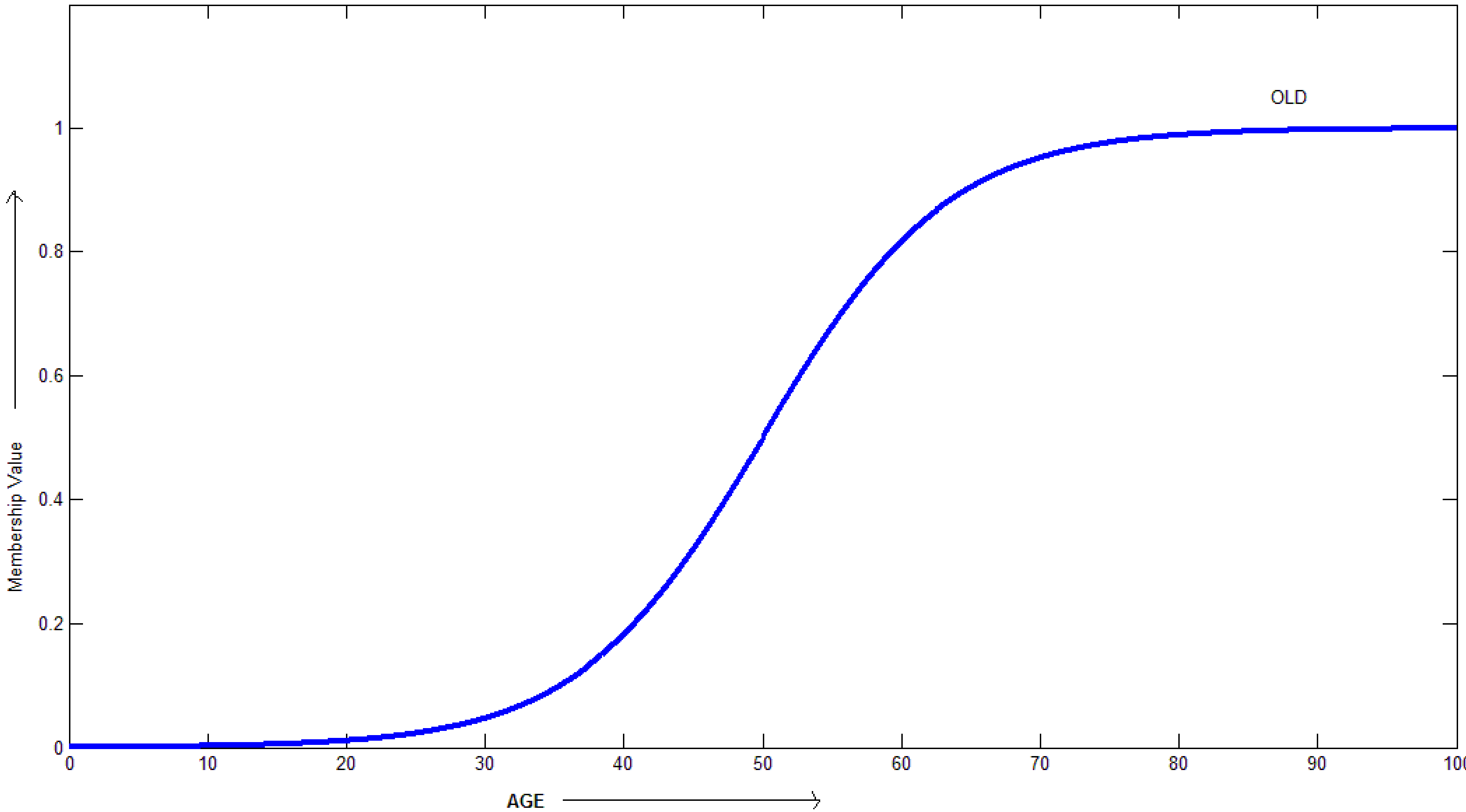
- The statement could be translated into fuzzy set terminology as follows:
- Helen is a member of the set of old people.
- It could be expressed in symbolic notation of fuzzy set as $\lambda_{OLD}(Helen) = 0.95$ i.e., Helen's degree of membership within the set of old people = 0.95

Distinction in two views

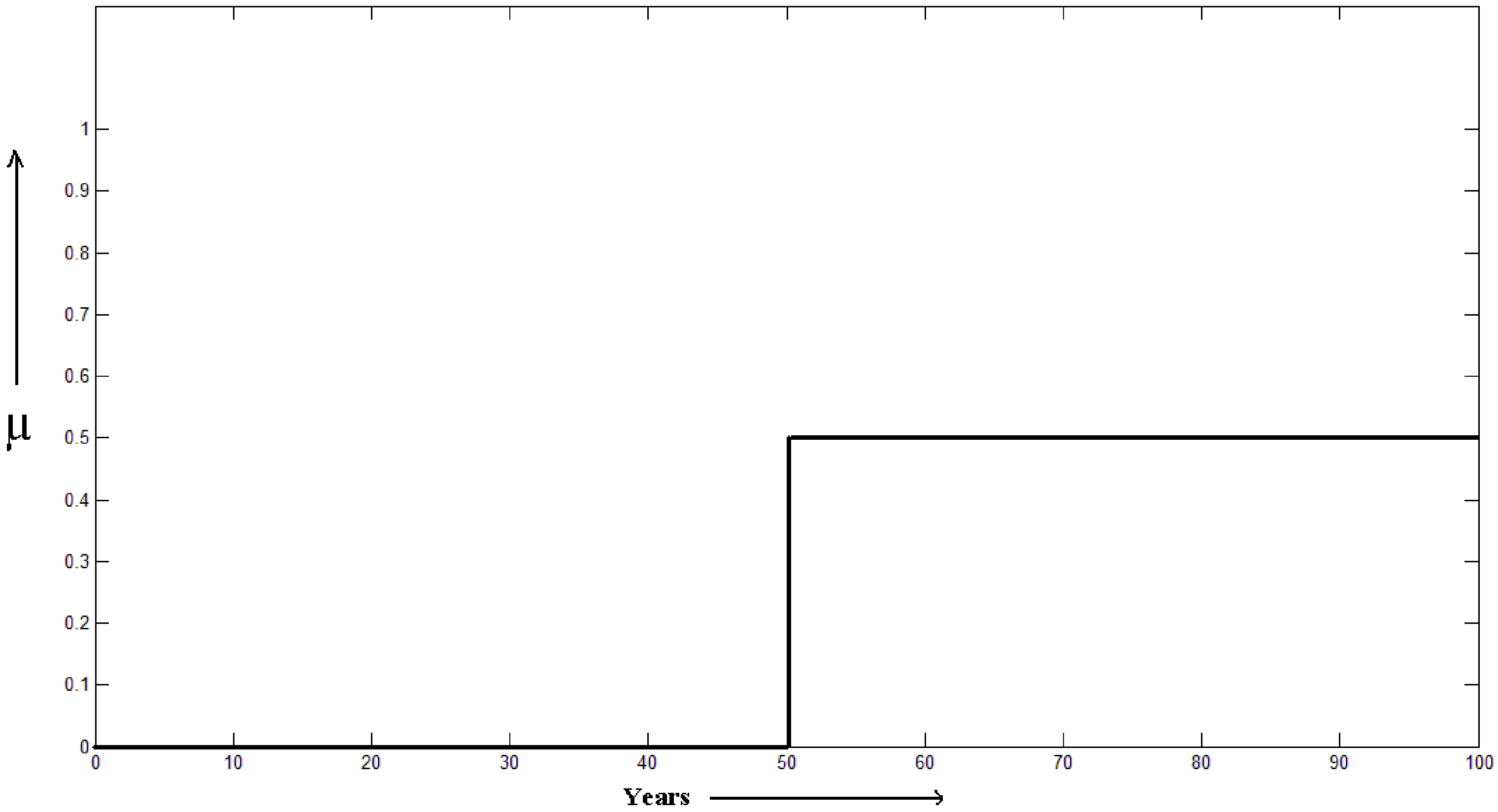
Important distinction between fuzzy systems and probability.

- Although these two statements seem similar but they actually carry different meanings.
- **First view:** There are 5% chances that Helen may not old.
- **Second view**: There is no chance of Helen being young and she is more or less old.
  - Here μ_{OLD} is a membership function operation on the fuzzy set of old people (denoted OLD) which return a value between 0 and 1.

Membership function μ_{OLD} for the fuzzy set OLD is represented as



Membership function for crisp (conventional) set older than 50 years is represented as:



Additional operations

1. Equality: A = B, if $\mu_{A(x)} = \mu_{B(x)}$, $\forall x \in X$
2. Not equal: $A \neq B$, if $\mu_{A(x)} \neq \mu_{B(x)}$ for at least one $x \in X$
3. Containment: $A \subseteq B$ if and only if $\mu_{A(x)} \leq \mu_{B(x)}$, $\forall x \in X$
4. Proper subset: If $A \subseteq B$ and $A \neq B$
5. Product: A.B is defined as $\mu_{A.B(x)} = \mu_{A(x)} \cdot \mu_{B(x)}$
6. Power : A^N is defined as: $\mu_A^{N(x)} = (\mu_{A(x)})^N$
7. Bold union : $A \oplus B$ is defined as:

   $$\mu_{A \oplus B(x)} = \text{Min}\ [1, \mu_{A(X)} + \mu_{B(x)}]$$

8. Bold intersection: A O B is defined as:

   $$\mu_{A\, \mathrm{O}\, B(x)} = \text{Max}\ [0, \mu_{A(x)} + \mu_{B(x)} - 1]$$

Various Types of Membership Functions

- S-shaped function
- Z-shaped function
- Triangular Membership Function
- Trapezoidal Membership Function
- Gaussian Distribution Function
- Pi function
- Vicinity function

Various Types of Membership Functions

- S-shaped function
- Z-shaped function
- Triangular Membership Function
- Trapezoidal Membership Function
- Gaussian Distribution Function
- Pi function
- Vicinity function

THANK YOU ALL

★ ★ ★ ★ ★

ALL THE BEST