9

Reasoning in Uncertain Situations

- Uncertainty results from:
- the use of abductive inference
- attempts to reason with missing or unreliable data

Uncertainty

- Uncertainty can be considered as the lack of adequate information to make a decision.
- Uncertainty is a problem because it may prevent us from making the best decision and may even cause a bad decision.

Unsound Inference

- In expert systems, correct conclusions must often be drawn from poorly formed and uncertain evidence using unsound inference rules.
- Unsound inference rules, such as abduction, are often used to solving problems.

Abduction rule:

From P → Q and Q, it is possible to infer P.

ity, Dempster-Shafer theory, and Zadeh's fuzzy theory.

TYPES OF ERRORS

Many different types of **errors** can contributes to uncertainty. Different theories of uncertainty attempt to resolve some or all of these errors to provide the most reliable inference. Figure 4-1 illustrates a simplified classification scheme for errors.

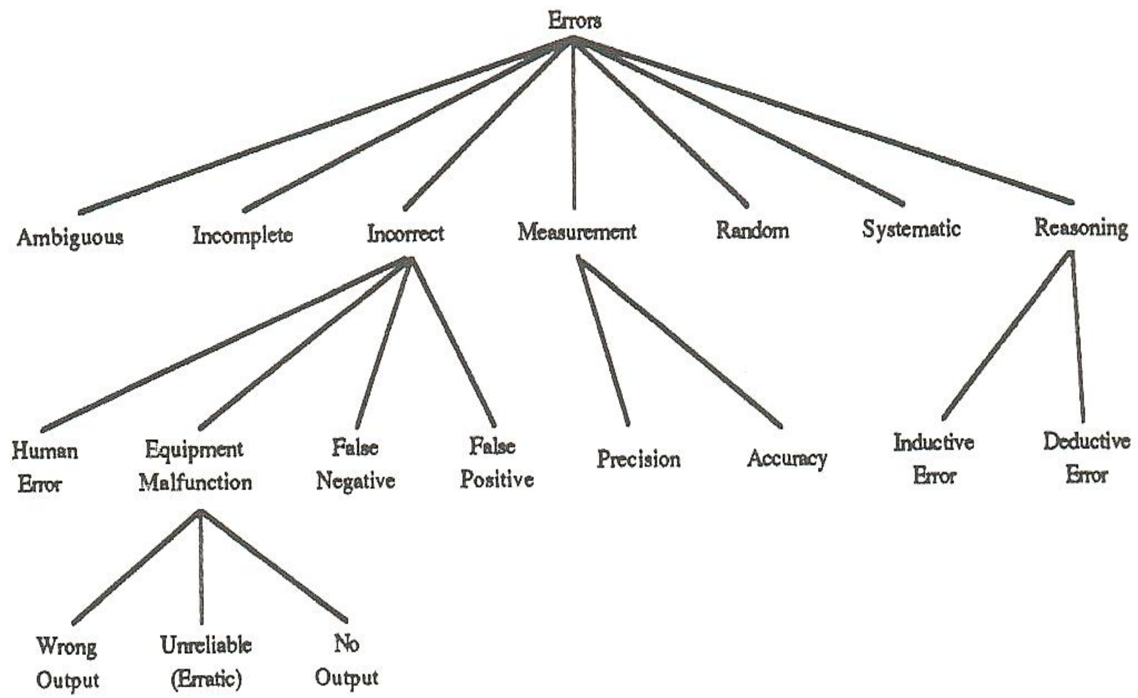


Figure 4-1
Types of Errors

Example	*Error*	*Reason*
Turn the valve off	Ambiguous	What valve?
Turn valve-1	Incomplete	Which way?
Turn valve-1 off	Incorrect	Correct is on
Valve is stuck	False positive	Valve is not stuck
Valve is not stuck	False negative	Valve is stuck
Turn valve-1 to 5	Imprecise	Correct is 5.4
Turn valve-1 to 5.4	Inaccurate	Correct is 9.2
Turn valve-1 to 5.4 or 6 or 0	Unreliable	Equipment error
Valve-1 setting is 5.4 or 5.5 or 5.1	Random Error	Statistical Fluctuation
Valve-1 setting is 7.5	Systematic Error	Miscalibration
Valve-1 is not stuck because its never been stuck before	Invalid Induction	Valve is stuck
Output is normal and so valve-1 is in good condition	Invalid deduction	Valve is stuck in open position

Table 4-1

Some ways of managing the uncertainty

1. Bayesian probability theory
2. Stanford certainty theory
3. Zadeh's fuzzy set theory
4. Dempster/Shafer theory of evidential reasoning
5. nonmonotonic reasoning

Bayesian probability theory

- Assuming random distribution of events, probability theory allows the calculation of more complex probabilities from previously known results.
- Based on the following assumptions:
- all the statistical data on the right are known
- all $P(E|Hk)$ are independent

Names	*Formula*	*Characteristics*
a priori (classical, theoretical, mathematical, symmetric equiprobable equal-likelihood)	$P(E) = \frac{W}{N}$ where W is the number of outcomes of event E for a total of N possible outcomes	Repeatable events Equally likely outcomes Exact math form known Not based on experiment All possible events and outcomes known
a posteriori (experimental, empirical, scientific, relative frequency, statistical) $P(E) \approx \frac{f(E)}{N}$	$P(E) = \lim_{N \to \infty} \frac{f(E)}{N}$ where f(E) is the frequency, f, that event, E, is observed for N total outcomes	Repeatable events based on experiments Approximated by a finite number of experiments Exact math form unknown
subjective (personal)		Nonrepeatable events Exact math form unknown Relative frequency method not possible Based on expert's experience judgement, opinion, or belief

Table 4-5
Types of Probabilities

Classical probability

Theory of Probability

A formal theory of probability can made using three axioms:

axiom 1: $0 \leq P(E) \leq 1$

This axiom defines the range of probability to be the real numbers from 0 to 1. Negative probabilities are not allowed. A **certain event** is assigned probability 1 and an **impossible event** assigned probability 0.

axiom 2: $\sum_{i} P(E_i) = 1$

This axiom states that the sum of all events which do not affect each other, called **mutually exclusive events**, is 1. Mutually exclusive events have no sample point in common. For example, a computer cannot be both working correctly and not working correctly at the same time.

As a corollary of this axiom

$$P(E) + P(E') = 1$$

where E' is the complement of event E. This corollary means that the probability of an event occurring plus the probability of it not occurring is 1. That is, the occurrence and nonoccurrence of an event is a mutually exclusive and complete sample space.

$$\textit{axiom 3:} \quad P(E_1 \cup E_2) = P(E_1) + P(E_2)$$

where E_1 and E_2 are mutually exclusive events. This axiom means that if E_1 and E_2 cannot both occur simultaneously (mutually exclusive events) then the probability of one or the other occurring is the sum of their probabilities.

From these axioms, theorems can be deduced concerning the calculation of probabilities under other situations, such as nonmutually exclusive events,

Compound Probabilities

- $P(A \cap B) = P(A) * P(B)$ if A, B are pairwise independent
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Conditional probabilities:

  $P(A|B) = P(A \cap B) / P(B)$ for $P(B) \neq 0$
- Generalized Multiplicative Law:

$$P(A1 \cap A2 \cap \ldots \cap An) = P(A1|A2 \cap \ldots \cap An) * P(A2|A3 \cap \ldots \cap An) * \ldots * P(An\text{-}1|An) * P(An)$$

Bayes' Theorem

- The conditional probability, P(A|B), states the probability of event A given that B occurred.
- The inverse problem is to find the inverse probability which states the probability of an earlier event given that a later one occurred.
- This type of probability occurs often, as in the medical or equipment diagnosis where symptoms appear and the problem is to find the most likely cause.
- The solution is Bayes's Theorem:

$$P(Hi|E) = P(E|Hi)*P(Hi) / P(E)$$

Bayes' Theorem

- Is commonly used for decision tree analysis
- Is also used in the PROSPECTOR expert system to decide favorable sites for mineral exploration

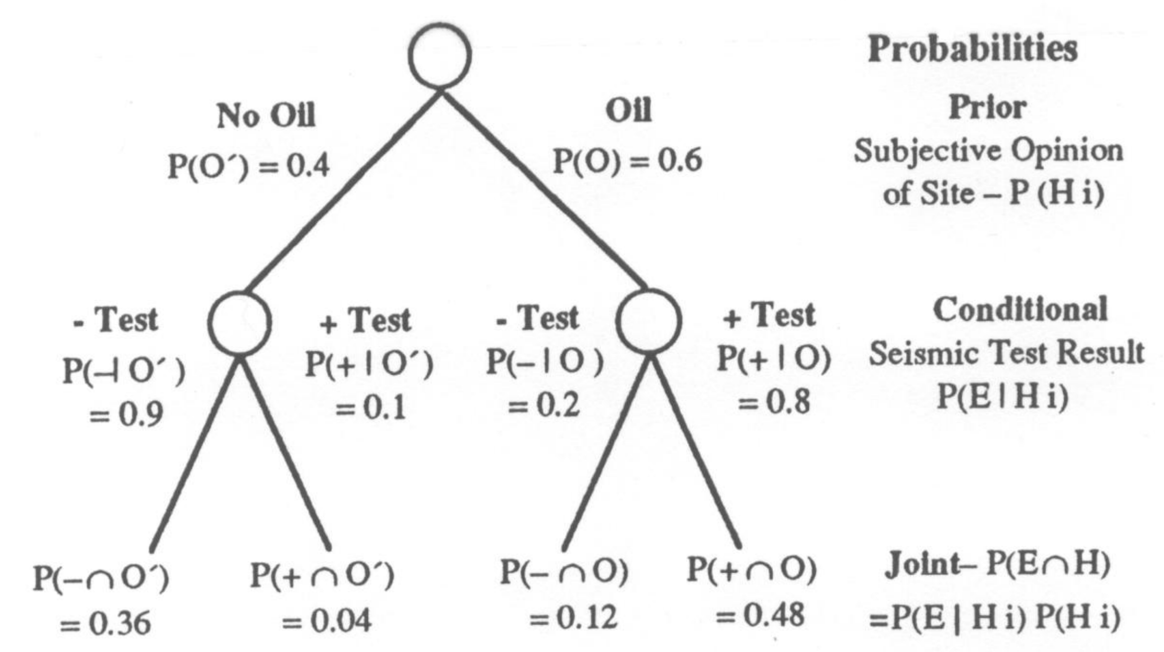


Figure 4-9
Initial Probability Tree for Oil Exploration

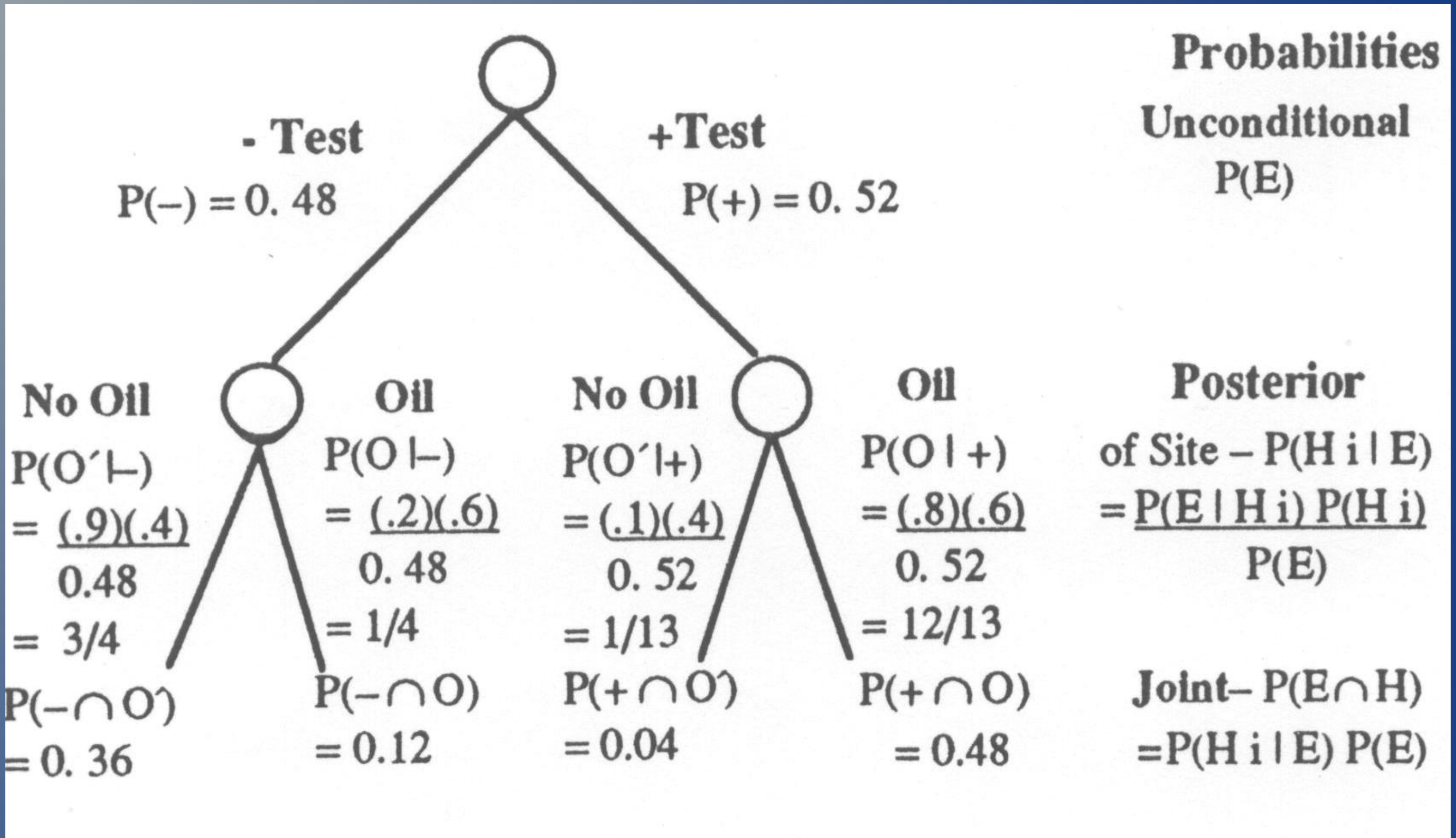


Figure 4-10
Revised Probability Tree for Oil Exploration

Limitations of the Bayesian approach:

- Correct conclusions require complete and up-to-date probabilities
- The independence of the relationships between evidence and hypotheses requires a justification
- Where these assumptions are met, Bayesian approaches offer the benefit of a well-founded and statistically correct handling of uncertainty. But, in many domains, such extensive data collection and justification are not possible. Human experts do not use the Bayesian model for problem solving.

Stanford certainty theory

- Stanford certainty theory is based on the following assumptions:
- The sum of confidence for a relationship and confidence against the same relationship must add to 1.
- The knowledge content of the rules is much more important than the algebra of confidences that holds the system together.

Call $MB(H \mid E)$ the measure of belief of a hypothesis H given evidence E.

Call $MD(H \mid E)$ the measure of disbelief of a hypothesis H given evidence E.

Now either:

$1 > MB(H \mid E) > 0$ while $MD(H \mid E) = 0$, or

$1 > MD(H \mid E) > 0$ while $MB(H \mid E) = 0$.

These two measures constrain each other in that a given piece of evidence is either for or against a particular hypothesis, an important difference between certainty theory and probability theory. Once the link between measures of belief and disbelief has been established, they may be tied together again, by:

$$CF(H \mid E) = MB(H \mid E) - MD(H \mid E).$$

The premises for each rule are formed of ands and ors of a number of facts. When a production rule is used, the certainty factors associated with each condition of the premise are combined to produce a certainty measure for the overall premise as follows. For P1 and P2, premises of the rule:

CF(P1 and P2) = MIN(CF(P1), CF(P2)), and

CF(P1 or P2) = MAX(CF(P1), CF(P2)).

The combined CF of the premises, using the above rules, is then multiplied by the CF of the rule itself to get the CF for the conclusions of the rule. For example, consider the rule in a knowledge base:

(P1 and P2) or P3 $\rightarrow$ R1 (.7) and R2 (.3)

where P1, P2, and P3 are premises and R1 and R2 are the conclusions of the rule, having CFs 0.7 and 0.3, respectively. These numbers are added to the rule when it is designed and represent the expert's confidence in the conclusion if all the premises are known with complete certainty. If the running program has produced P1, P2, and P3 with CFs of 0.6, If the running program has produced P1, P2, and P3 with CFs of 0.6, 0.4, and 0.2, respectively, then R1 and R2 may be added to the collected case-specific results with CFs 0.28 and 0.12, respectively. Here are the calculations for this example:

CF(P1(0.6) and P2(0.4)) = MIN(0.6,0.4) = 0.4.
CF((0.4) or P3(0.2)) = MAX(0.4,0.2) = 0.4.

The CF for R1 is 0.7 in the rule, so R1 is added to the set of case-specific knowledge with the associated CF of $(0.7) \times (0.4) = 0.28$.

The CF for R2 is 0.3 in the rule, so R2 is added to the set of case-specific knowledge with the associated CF of $(0.3) \times (0.4) = 0.12$.

One further measure is required: how to combine multiple CFs when two or more rules support the same result R. This rule reflects the certainty theory analog of the probability theory procedure of multiplying probability measures to combine independent evidence. By using this rule repeatedly one can combine the results of any number of rules that are used for determining a result R. Suppose CF(R1) is the present certainty factor associated with result R and a previously unused rule produces result R (again) with CF(R2); then the new CF of R is calculated by:

$CF(R1) + CF(R2) - (CF(R1) \times CF(R2))$ when CF(R1) and CF(R2) are positive,
$CF(R1) + CF(R2) + (CF(R1) \times CF(R2))$ when CF(R1) and CF(R2) are negative,

and

$$\frac{CF(R1) + CF(R2)}{1 - MIN(\,|\,CF(R1)\,|\,,\,|\,CF(R2)\,|\,)}$$

otherwise, where | X | is the absolute value of X.

CF(P1(0.6) and P2(0.4)) = MIN(0.6,0.4) = 0.4.
CF((0.4) or P3(0.2)) = MAX(0.4,0.2) = 0.4.

The CF for R1 is 0.7 in the rule, so R1 is added to the set of case-specific knowledge with the associated CF of $(0.7) \times (0.4) = 0.28$.
The CF for R2 is 0.3 in the rule, so R2 is added to the set of case-specific knowledge with the associated CF of $(0.3) \times (0.4) = 0.12$.

One further measure is required: how to combine multiple CFs when two or more rules support the same result R. This rule reflects the certainty theory analog of the probability theory procedure of multiplying probability measures to combine independent evidence. By using this rule repeatedly one can combine the results of any number of rules that are used for determining a result R. Suppose CF(R1) is the present certainty factor associated with result R and a previously unused rule produces result R (again) with CF(R2); then the new CF of R is calculated by:

$CF(R1) + CF(R2) - (CF(R1) \times CF(R2))$ when CF(R1) and CF(R2) are positive,
$CF(R1) + CF(R2) + (CF(R1) \times CF(R2))$ when CF(R1) and CF(R2) are negative,

and

$$\frac{CF(R1) + CF(R2)}{1 - MIN(\,|CF(R1)|\,,\,|CF(R2)|\,)}$$

otherwise, where | X | is the absolute value of X.

MYCIN

- MYCIN was designed to solve the problem of diagnosing and recommending treatment for meningitis and bacteremia.
- MYCIN (Middle and late 1970s, Stanford Univ.)
  - a rule-based expert system
  - goal-drive depth-first search
  - using certainty theory for heuristic inference
  - representing facts as attribute-object-value triples
  - explaining why and how questions
  - having a knowledge base editor
  - written in INTERLISP by 50 person-years
  - EMYCIN is its corresponding expert system shell.

Stanford certainty theory

- Certainty theory is excessively ad hoc. The meaning of the certainty measures is not rigorously founded.
- However, the CF is used in the heuristic search to give a priority for goals to be attempted and for cutting searching branches off in order to keep the program running, the power of the program is in the content of the rules themselves.

Numeric approaches of handling uncertain reasoning

- do not address the problem of changing data
- are unlikely used by humans

Monotonic reasoning systems

- Monotonic reasoning systems assume that axioms do not change and the conclusions drawn from them remain true, thus knowledge can only be added when we already know or through the reasoning process.

Non-monotonic logic

- Non-monotonic logic (McDermott and Doyle, 1980) is an extension to the predicate calculus, which allows statements to change their truth-values in the process of reasoning.
- When there is no information to the contrary, we may assert some fact to be true. Only later does evidence appear to the contrary, the fact must be negated.

Nonmonotonic reasoning

- Nonmonotonic reasoning systems handle uncertainty by making the most reasonable assumptions in light of uncertain information.
- It proceeds with its reasoning as if these assumptions were true.
- When inconsistencies occurs, the system change both the assumptions and all of the conclusions that depend on them.

Nonmonotonicity

- Nonmonotonicity is an important feature of human problem solving and common sense reasoning.

Fuzzy Terms

low
medium
high
very
not
more or less
little
several
few
many
more
most
about
approximately
sort of
a great deal

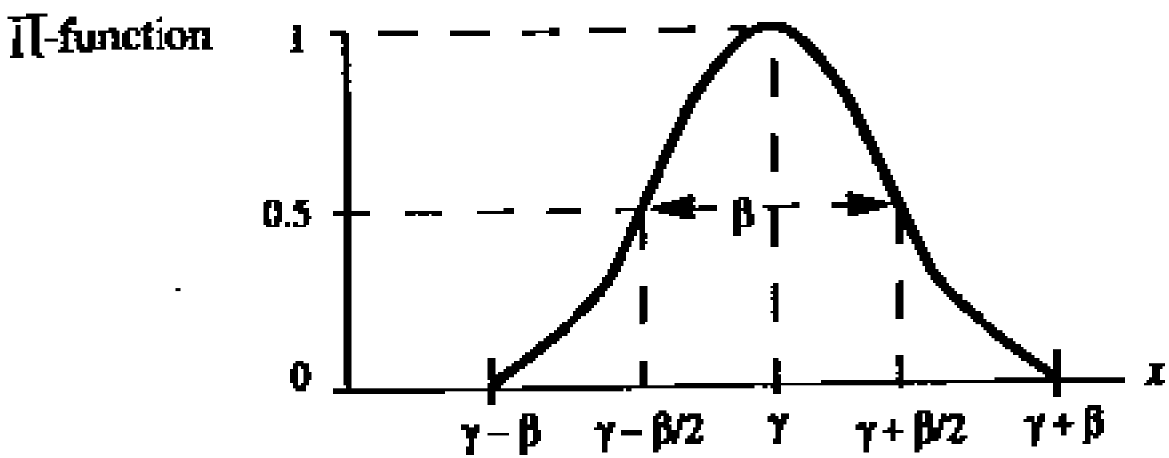
Table 5-8
Some Fuzzy Terms of Natural Language

very low
more or less low
approximately low
not low
not very low
not more or less low
medium to sort of high
higher than slightly low
low to sort of medium
most high

Table 5-9
Compound Fuzzy Natural Language Terms

Fuzzy Membership Functions

$$\Pi(x;\ \beta,\ \gamma) = \begin{cases} S(x;\ \gamma - \beta,\ \gamma - \beta/2,\ \gamma) & \text{for } x \le \gamma \\ 1 - S(x;\ \gamma,\ \gamma + \beta/2, \gamma + \beta) & \text{for } x \ge \gamma \end{cases}$$



Fuzzy Membership Functions

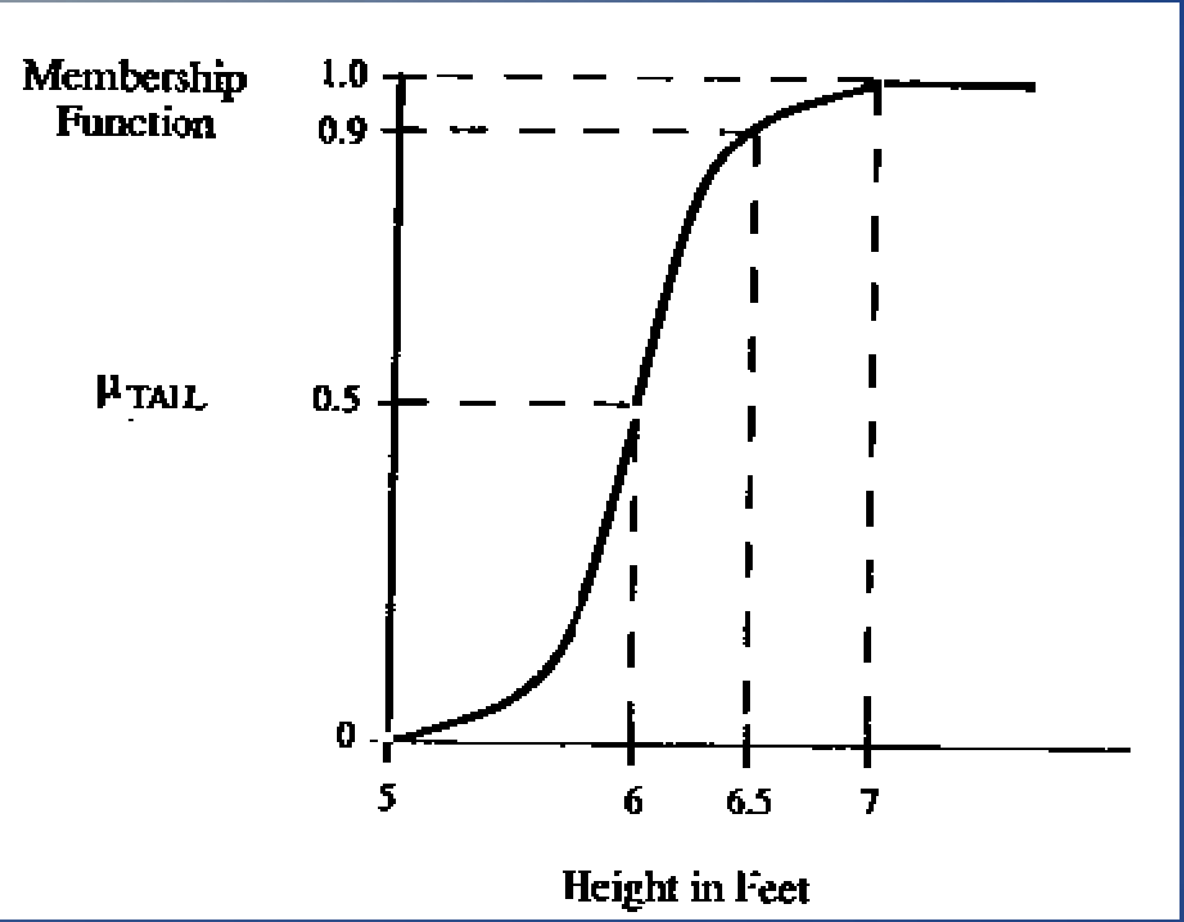


Fig 9.6 the fuzzy set representation for “small integers.”

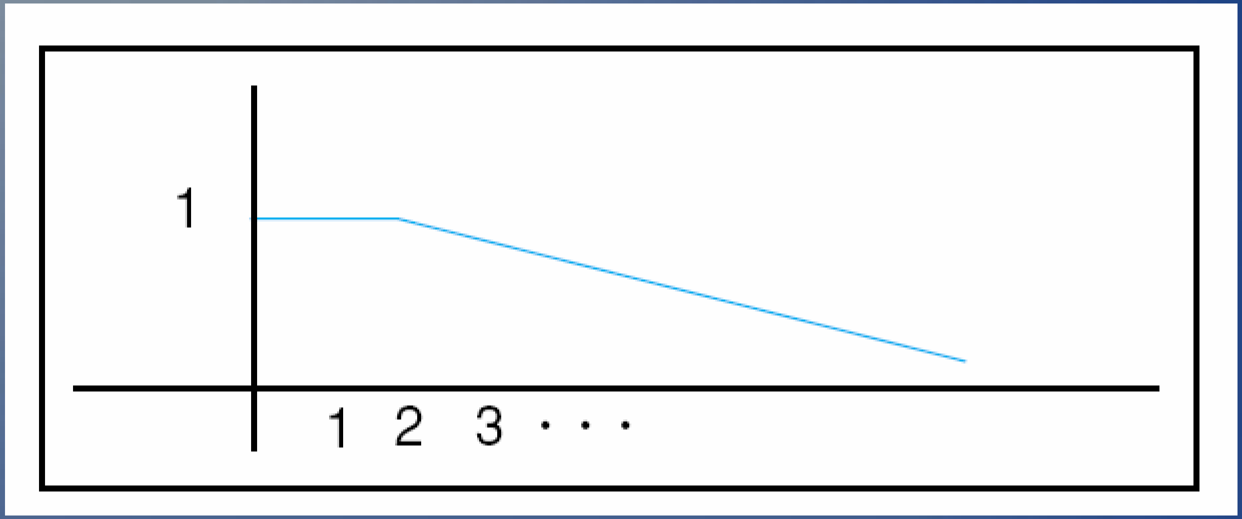


Fig 9.7 A fuzzy set representation for the sets short, medium, and tall males.

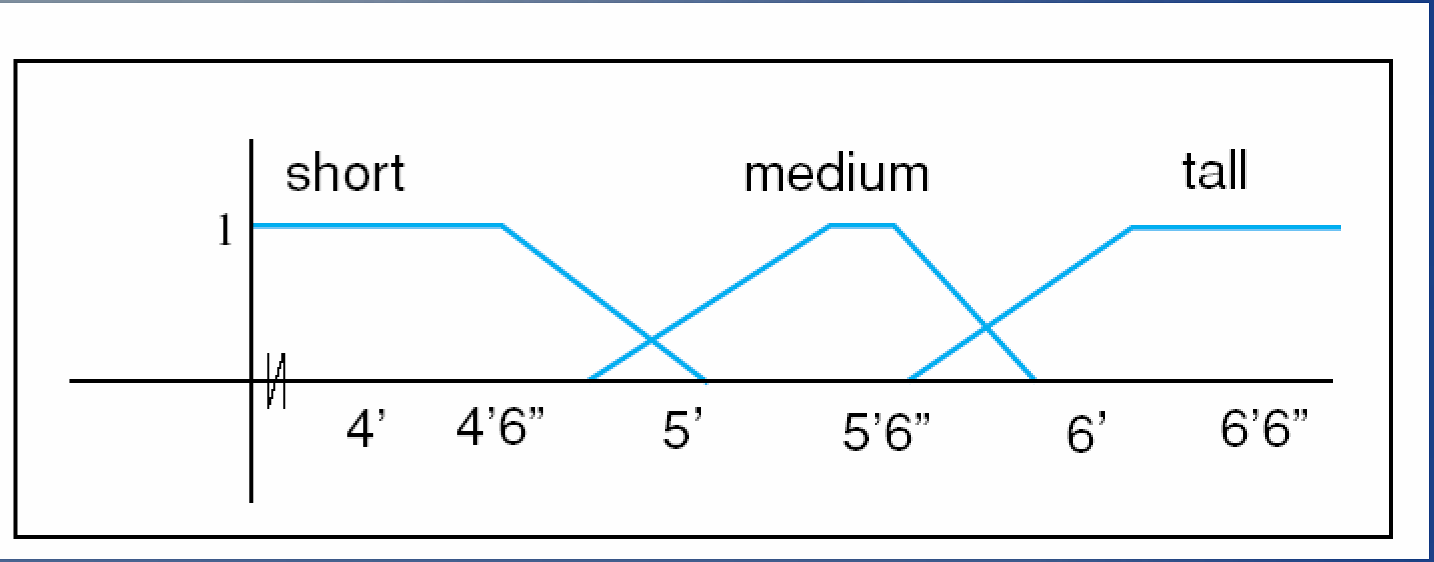


Fig 9.9 The fuzzy regions for the input values θ (a) and dθ/dt (b).

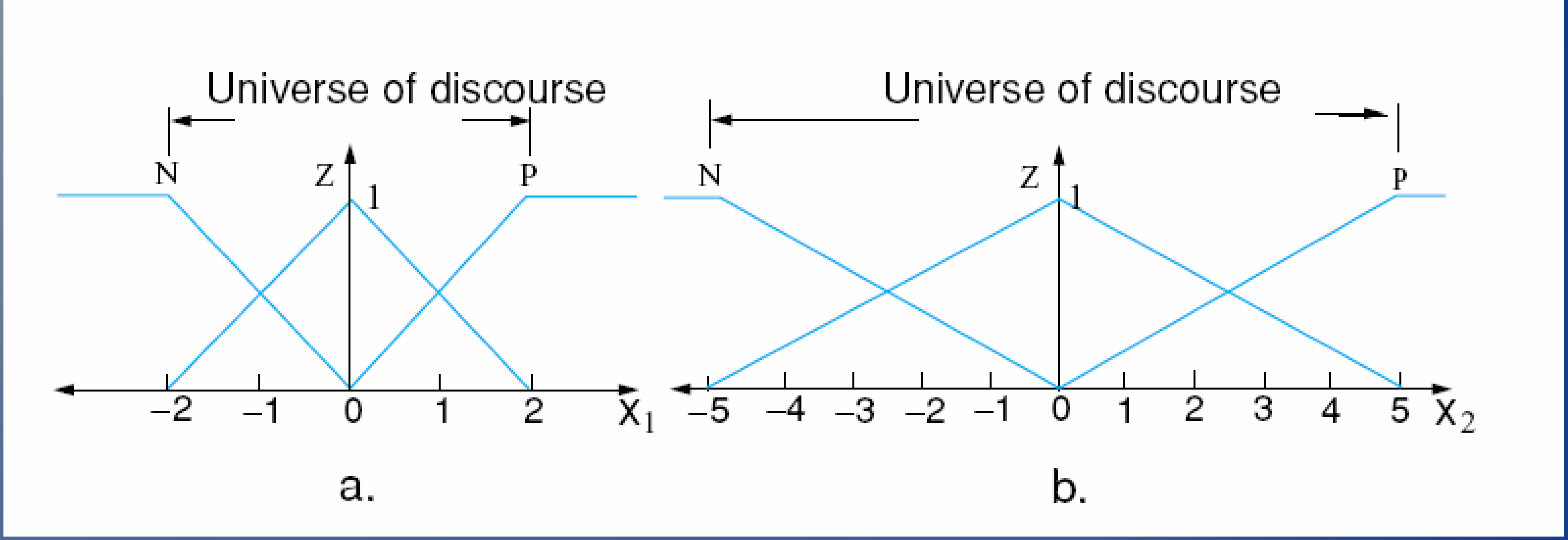
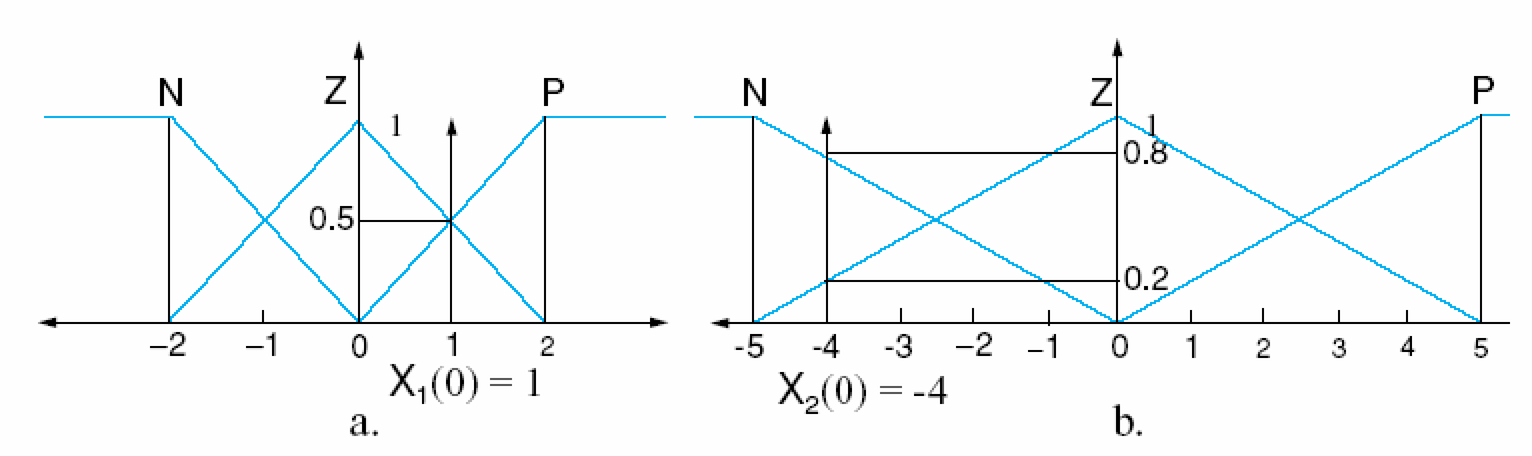


Fig 9.11 The fuzzification of the input measures $X_1 = 1$, $X_2 = -4$



Fuzzy Membership Functions

The hierarchy of the linguistic variable Appetite is illustrated in Figure 5-16. The LIGHT and HEAVY fuzzy sets are assumed to be S-functions while the MODERATE set is taken as a $\prod$-function.

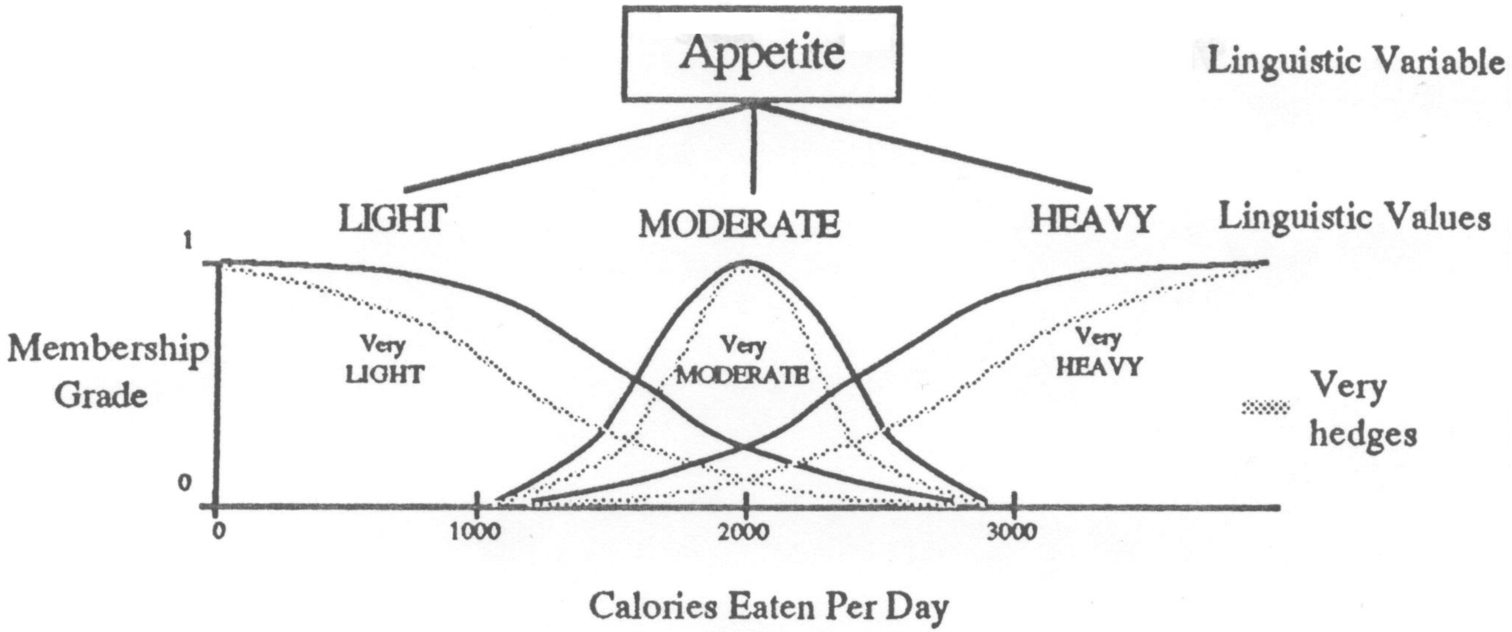


Figure 5-16
The Linguistic Variable Appetite and Its Values

$x(A')$	$= x(\text{NOT } A)$	$= 1 - \mu_A(x)$
$x(A) \wedge x(B)$	$= x(A \text{ AND } B)$	$= \min(\mu_A(x), \mu_B(x))$
$x(A) \vee x(B)$	$= x(A \text{ OR } B)$	$= \max(\mu_A(x), \mu_B(x))$
$x(A) \rightarrow x(B)$	$= x(A \rightarrow B)$	$= x((\sim A) \vee B) = \max[(1 - \mu_A(x)), \mu_B(x)]$

Table 5-15
Some Fuzzy Logic Operators

Another definition for the CHOCOLATE fuzzy set may involve **hedges** to modify the meaning of a set. For example, a CHOCOLATE fuzzy set of one type of chocolate could be defined as follows.

```
CHOCOLATE = Very CHOCOLATE + Very Very CHOCOLATE +
            More Or Less CHOCOLATE +
            Slightly CHOCOLATE +
            Plus CHOCOLATE + Not Very CHOCOLATE + ...
```

Standard hedges can be defined in terms of some fuzzy set operators and a fuzzy set, F, as shown in Table 5-13.

Hedge	*Operator Definition*
Very F	$CON(F) = F^2$
More Or Less F	$DIL(F) = F^{0.5}$
Plus F	$F^{1.25}$
Not F	$1 - F$
Not Very F	$1 - CON(F)$
Slightly F	INT [NORM (PLUS F And NOT (VERY F))]

Table 5-13
Some Linguistic Hedges and Operators

Fig 9.13 The fuzzy consequents (a) and their union (b). The centroid of the union (-2) is the crisp output.