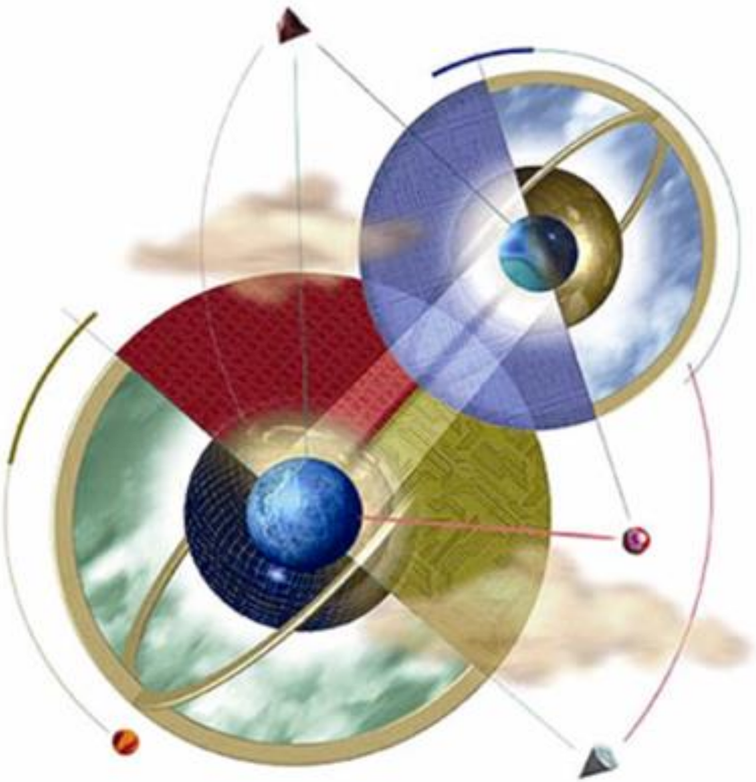


Chapter 4: Reasoning Under Uncertainty

Expert Systems: Principles and Programming, Fourth Edition

Objectives

- Learn the meaning of uncertainty and explore some theories designed to deal with it
- Find out what types of errors can be attributed to uncertainty and induction
- Learn about classical probability, experimental, and subjective probability, and conditional probability
- Explore hypothetical reasoning and backward induction

Objectives

- Examine temporal reasoning and Markov chains
- Define odds of belief, sufficiency, and necessity
- Determine the role of uncertainty in inference chains
- Explore the implications of combining evidence
- Look at the role of inference nets in expert systems and see how probabilities are propagated

How to Expert Systems Deal with Uncertainty?

- Expert systems provide an advantage when dealing with uncertainty as compared to decision trees.
- With decision trees, all the facts must be known to arrive at an outcome.
- Probability theory is devoted to dealing with theories of uncertainty.
- There are many theories of probability – each with advantages and disadvantages.

Theories to Deal with Uncertainty

- Bayesian Probability
- Hartley Theory
- Shannon Theory
- Dempster-Shafer Theory
- Markov Models
- Zadeh’s Fuzzy Theory

What is Uncertainty?

- Uncertainty is essentially lack of information to formulate a decision.
- Uncertainty may result in making poor or bad decisions.
- As living creatures, we are accustomed to dealing with uncertainty – that's how we survive.
- Dealing with uncertainty requires reasoning under uncertainty along with possessing a lot of common sense.

Dealing with Uncertainty

- Deductive reasoning – deals with exact facts and exact conclusions
- Inductive reasoning – not as strong as deductive – premises support the conclusion but do not guarantee it.
- There are a number of methods to pick the best solution in light of uncertainty.
- When dealing with uncertainty, we may have to settle for just a good solution.

Choosing Uncertainty

- Sometimes, uncertainty is chosen
  - Certain decision is expensive.
    - Prospector
  - Certain decision takes time.
    - Mycin

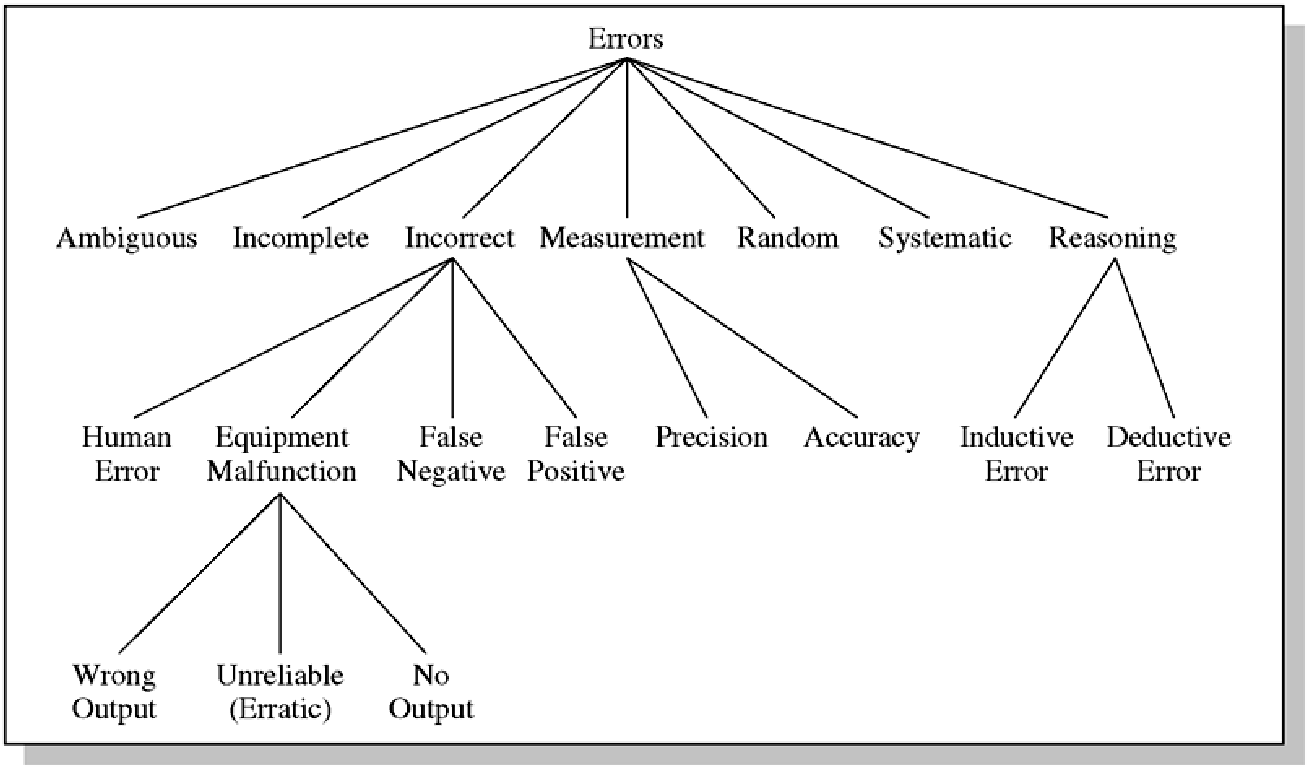
Errors Related to Hypothesis

- Many types of errors contribute to uncertainty.
  - Type I Error – accepting a hypothesis when it is not true – False Positive.
  - Type II Error – Rejecting a hypothesis when it is true – False Negative

Errors Related to Measurement

- Errors of precision – how well the truth is known
- Errors of accuracy – whether something is true or not
  - *Accuracy vs. Precision: Ruler Example.*
- Unreliability stems from faulty measurement of data – results in erratic data.
- Random fluctuations – termed random error
- Systematic errors result from bias

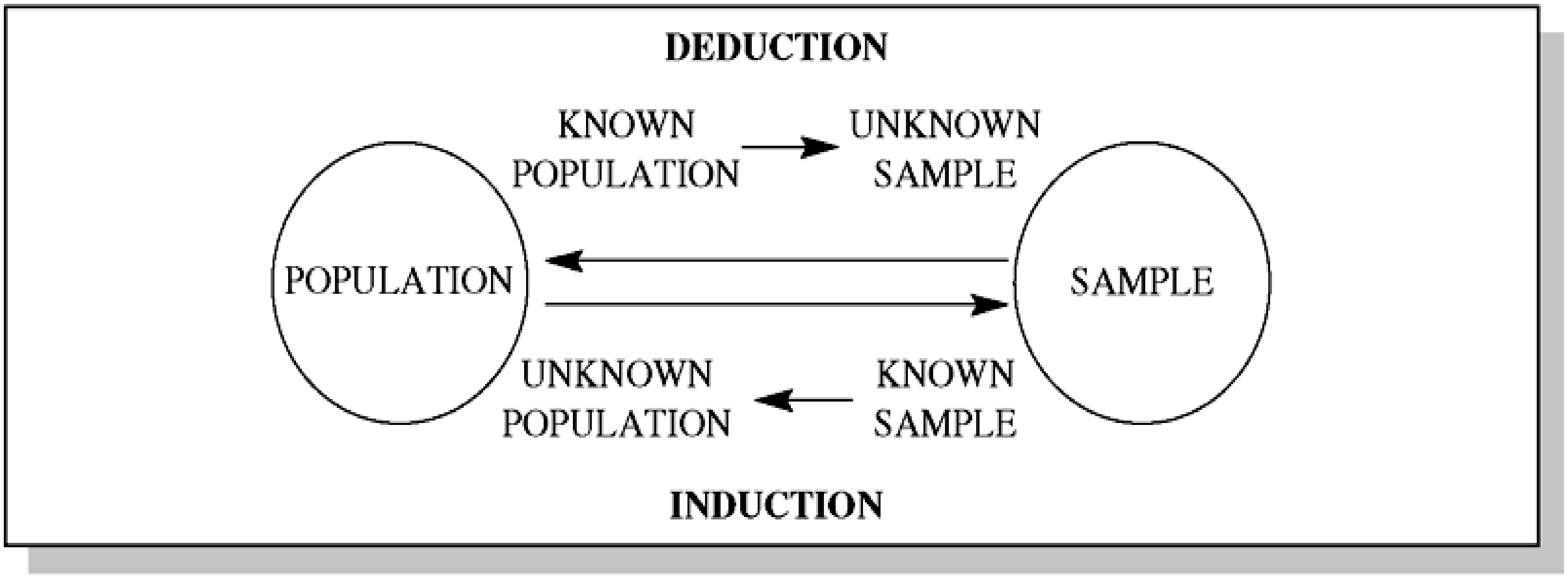
Figure 4.1 Types of Errors



Errors in Induction

- Where deduction proceeds from general to specific, induction proceeds from specific to general.
  - *Type X hard drives never fail. ➔ My type X HDD doesn't fail.*
- Inductive arguments can never be proven correct (except in mathematical induction).
  - *None of our type X HDDs have failed by now ➔ This type X HDD won't fail today.*
  - *Automatic Learning Results are usually of this type.*
- When rules are based on heuristics, there will be uncertainty.

Figure 4.4 Deductive and Inductive Reasoning about Populations and Samples



Classical Probability

- First proposed by Pascal and Fermat in 1654
- Also called *a priori* probability because it deals with ideal games or systems:
  - Assumes all possible events are known
  - Each event is equally likely to happen
- Fundamental theorem for classical probability is $P = W / N$, where W is the number of wins and N is the number of equally possible events.

Backgammon Hint!

- Dices are not fair:
  - 1: 0.155
  - 2: 0.159
  - 3: 0.164
  - 4: 0.169
  - 5: 0.174
  - 6: 0.179

Deterministic vs. Nondeterministic Systems

- When repeated trials give the exact same results, the system is deterministic.

- Otherwise, the system is nondeterministic.

- Nondeterministic does not necessarily mean random – could just be more than one way to meet one of the goals given the same input.
  - Search Engine Results.
  - Random selection of rules with similar favorability.

Some Terms

- Sample Space / Sample Point
  - {1,2,3,4,5,6} , 4.
- Event: a subset of sample space.
  - Simple/Compund

Three Axioms of Formal Theory of Probability

1. $0 \le P(E) \le 1$

2. $\sum_i P(E_i) = 1$

3. $P(E_1 \cup E_2) = P(E_1) + P(E_2)$

   for mutually exclusive events.

Experimental and Subjective Probabilities (vs. a priori probability)

- Experimental probability defines the probability of an event, as the limit of a frequency distribution:

$$P(E) = \lim_{N \to \infty} \frac{f(E)}{N}$$

- Subjective probability deals with events that are not reproducible and have no historical basis on which to extrapolate.
  - Belief/Opinion of an expert.

Table 4.5 Types of Probabilities

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)	$P(E) = \lim_{N\to\infty} \frac{f(E)}{N}$ 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)	See Section 4.12	Nonrepeatable events Exact math form unknown Relative frequency method not possible Based on expert's opinion experience, judgment, belief

$$P(E) \approx \frac{f(E)}{N}$$

Compound Probabilities

- Compound probabilities can be expressed by:

$$P(A \cap B) = \frac{n(A \cap B)}{n(S)}$$

S is the sample space and *A* and *B* are events.

- Independent events are events that do not affect each other. For pairwise independent events,

$$P(A \cap B) = P(A)\, P(B)$$

Additive Law

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

$$\begin{aligned} P(A \cup B \cup C) = {} & P(A) + P(B) + P(C) \\ & - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ & + P(A \cap B \cap C) \end{aligned}$$

Conditional Probabilities

- The probability of an event *A* occurring, given that event *B* has already occurred is called conditional probability:

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

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

Bayes' Theorem

- This is the inverse of conditional probability.
- Find the probability of an earlier event given that a later one occurred.

$$P(H_i \mid E) = \frac{P(E \cap H_i)}{\sum_j P(E \cap H_j)}$$

$$= \frac{P(E|H_i)\,P(H_i)}{\sum_j P(E|H_j)\,P(H_j)}$$

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

The Odds of Belief

- P(A) can be seen as our degree of belief in A.
  - Probability is referred to something repeatable.
  - Degree of belief refers to our certainty.
- Odds:
  - $O(H) = P(H)/P(H')$:
  - P = 0.95 , Odds = 19 to 1.
- P(A|B): The ***likelihood*** of ***hypothesis*** A, given ***event*** B.
  - The degree of belief in A, knowing B.
  - A: Hypothesis: The proposition that we want to know more about.
  - B: Evidence: What we already know.

Sufficiency and Necessity

- The likelihood of sufficiency, LS, is calculated as:

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

- The likelihood of necessity is defined as:

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

Sufficiency and Necessity

- Example:
  - *IF there are quarts-sulfide veinlets* *(E)*

    THEN there is a favorable alteration

    for the potassic zone. *(H)*

LS = 300 $P(E|H) / P(E|H')$.

LN = 0.2 $P(E'|H)/P(E'|H')$.

Table 4.10 Relationship Among Likelihood Ratio, Hypothesis, and Evidence

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

LS	Effect on Hypothesis
0	H is false when E is true or E′ is necessary for concluding H
small ($0 < LS << 1$)	E is unfavorable for concluding H
1	E has no effect on belief of H
large ($1 << LS$)	E is favorable for concluding H
∞	E is logically sufficient for H or Observing E means H must be true

Table 4.11 Relationship Among Likelihood of Necessity, Hypothesis, and Evidence

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

LN	Effect on Hypothesis
0	H is false when E is absent or E is necessary for H
small (0 < LN << 1)	Absence of E is unfavorable for concluding H
1	Absence of E has no effect on H
large (1 << LN)	Absence of E is favorable for H
∞	Absence of E is logically sufficient for H

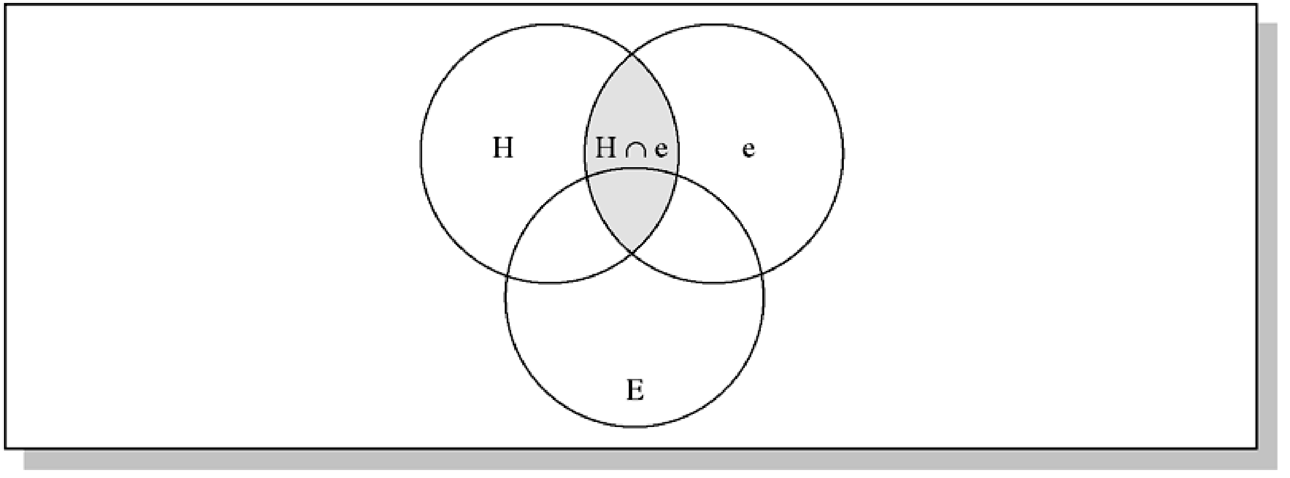
Consistency of Likelihoods

- Consistence Cases:
  - $LS > 1$ and $LN < 1$
  - $LS < 1$ and $LN > 1$
  - $LS = LN = 1$

- Expert's Opinion May Differ!

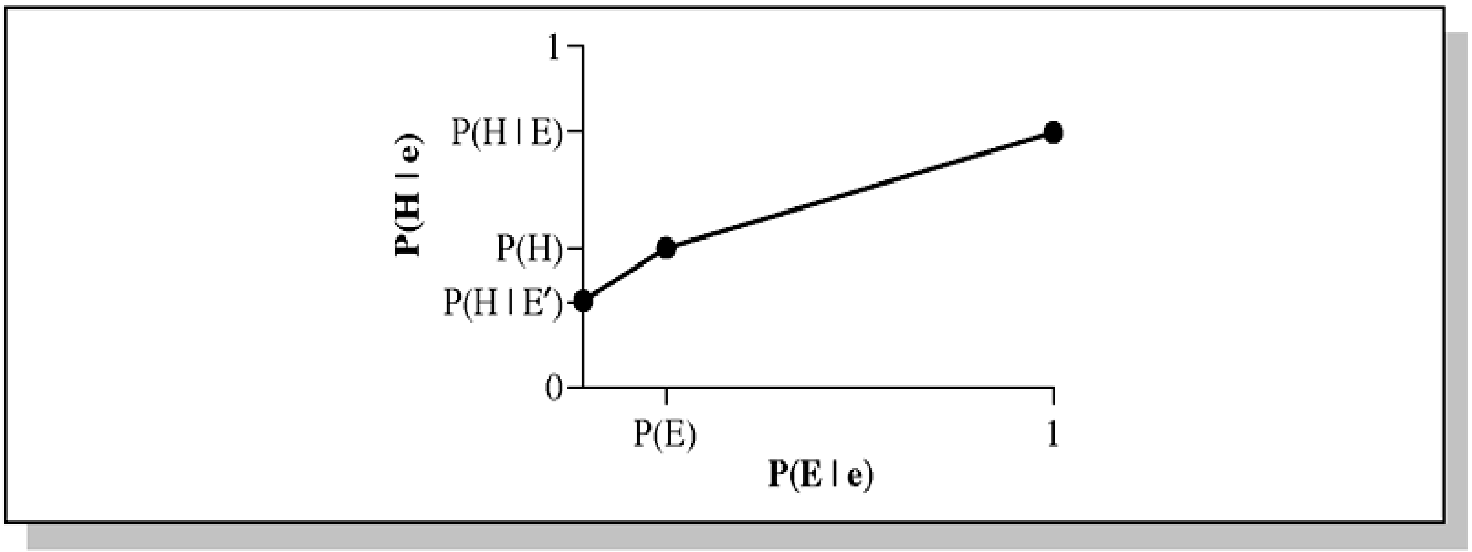
Uncertainty in Inference Chains

- Uncertainty may be present in rules, evidence used by rules, or both.
  - *How to compute $P(H|e)$ based on $P(H|E)$*
  - $P(H|e) = P(H|E \cap e).P(E|e) + P(H|E' \cap e).P(E'|e)$



Uncertainty in Inference Chains

- One way of correcting uncertainty is to assume that *P(H/e)* is a piecewise linear function.



Combination of Evidence

- The simplest type of rule is of the form:
  - IF E THEN H

  where *E* is a single piece of known evidence from which we can conclude that *H* is true.

- Not all rules may be this simple – compensation for uncertainty may be necessary.

- As the number of pieces of evidence increases, it becomes impossible to determine all the joint and prior probabilities or likelihoods.

Combination of Evidence

IF E_1 and E_2 then H

After E_1 and E_2 are observed, the probability of H changes from the prior P(H) to:

$$P(H \mid E_1 \cap E_2) = \frac{P(H \cap E_1 \cap E_2)}{P(E_1 \cap E_2)}$$

$$= \frac{P(H \cap E_1 \cap E_2)}{P(E_1 \cap E_2 \cap H) + P(E_1 \cap E_2 \cap H')}$$

(1) $P(H \mid E_1 \cap E_2)$

$$= \frac{P(E_1 \cap E_2 \mid H)\ P(H)}{P(E_1 \cap E_2 \mid H)\ P(H) + P(E_1 \cap E_2 \mid H')\ P(H')}$$

Combination of Evidence

has E as the conjunction of evidence, as in:

IF E_1 AND E_2 AND ... E_N THEN H

All the E_i must be true with some probability for the antecedent to be true. In the general case, each piece of evidence is based on partial evidence e. The probability of the evidence is:

$$P(E \mid e) = P(E_1 \cap E_2 \cap \ldots E_N \mid e)$$

$$= \frac{P(E_1 \cap E_2 \cap \ldots E_N \cap e)}{P(e)}$$

Combination of Evidence

$$P(E_1 \cap E_2 \mid e) = \frac{P(E_1 \cap E_2 \cap e)}{P(e)}$$

$$= \frac{P(E_2 \mid E_1 \cap e)\ P(E_1 \mid e)\ P(e)}{P(e)}$$

Using the assumption of independence,

$$P(E_2 \mid e) = P(E_2 \mid E_1 \cap e)$$

because the evidence E_1 does not contribute any knowledge toward E_2. So:

$$P(E_2 \cap E_1 \mid e) = P(E_2 \mid e)\ P(E_1 \mid e)$$

and in general:

$$P(E_1 \cap E_2 \cap \ldots E_N \mid e) = \prod_{i=1}^{N} P(E_i \mid e)$$

Combination of Evidence

$$P(E_1 \cap E_2 \mid e) = \frac{P(E_1 \cap E_2 \cap e)}{P(e)}$$

$$= \frac{P(E_2 \mid E_1 \cap e)\ P(E_1 \mid e)\ P(e)}{P(e)}$$

Using the assumption of independence,

$$P(E_2 \mid e) = P(E_2 \mid E_1 \cap e)$$

because the evidence E_1 does not contribute any knowledge toward E_2. So:

$$P(E_2 \cap E_1 \mid e) = P(E_2 \mid e)\ P(E_1 \mid e)$$

and in general:

$$P(E_1 \cap E_2 \cap \ldots E_N \mid e) = \prod_{i=1}^{N} P(E_i \mid e)$$

Combination of Evidence Continued

- If the antecedent is a logical combination of evidence, then fuzzy logic and negation rules can be used to combine evidence.

$$P(E_1 \cap E_2 \cap \ldots E_N \mid e) = \prod_{i=1}^{N} P(E_i \mid e)$$

$$P(E \mid e) = \min [P(E_i \mid e)]$$

Combination of Evidence Continued

IF E_1 OR E_2 OR ... E_N THEN H

$$P(E \mid e) = 1 - \prod_{i=1}^{N} [1 - P(E_i \mid e)]$$

$$P(E \mid e) = \max [P(E_i \mid e)]$$

Combination of Evidence Continued

IF E_1 AND (E_2 OR E_3') THEN H

then:

$$E = E_1 \text{ AND } (E_2 \text{ OR } E_3')$$
$$E = \min \{P(E_1 \mid e), \max [P(E_2 \mid e), 1 - P(E_3 \mid e)]\}$$

Inference Net

- In a real ES, the number of nodes are too many, so an organization is necessary.

Table 4.12 Some PROSPECTOR Models

Model Name	Model Description	Number of Nodes	Number Askable	Number of Rules
PCD	Porphyry Copper	200	97	135
RFU	Roll-front Uranium	185	147	169
SPB	Sandstone-hosted Lead	35	21	32

- To store the knowledge in the form of a taxonomy.

Figure 4.18 The Top-level Porphyry Copper Hypothesis of PROSPECTOR Expressed Using Certainty Factors

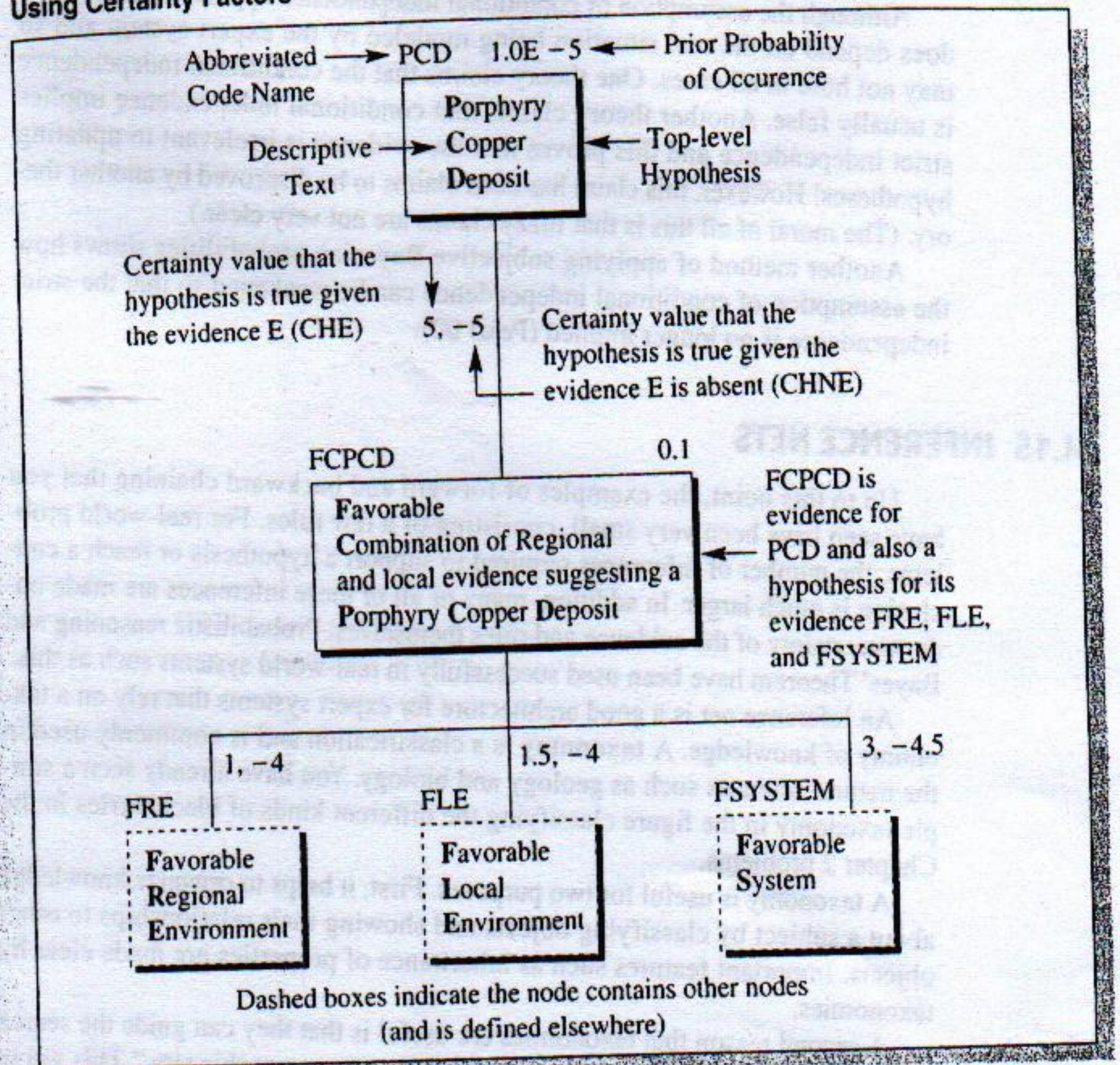
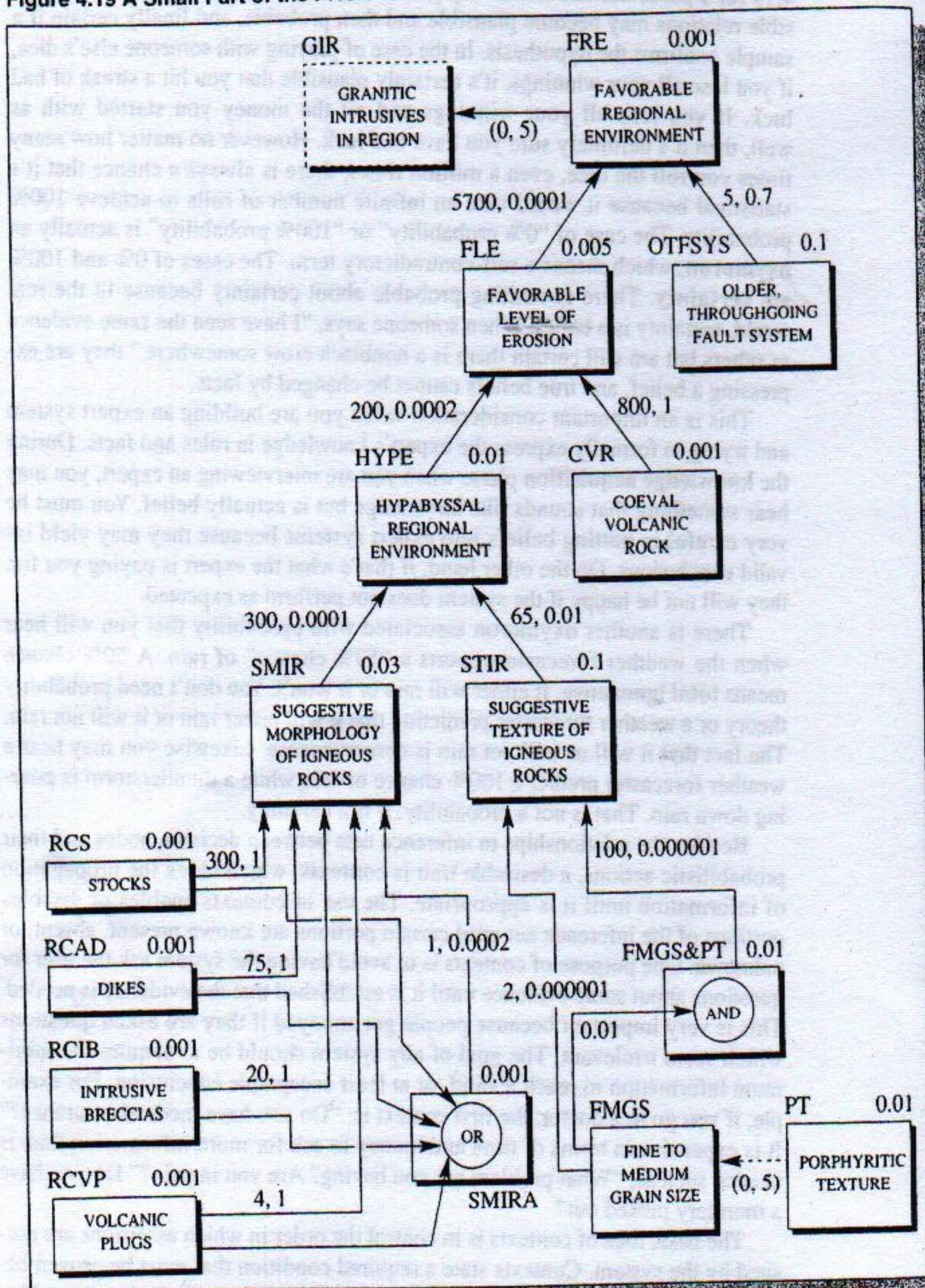


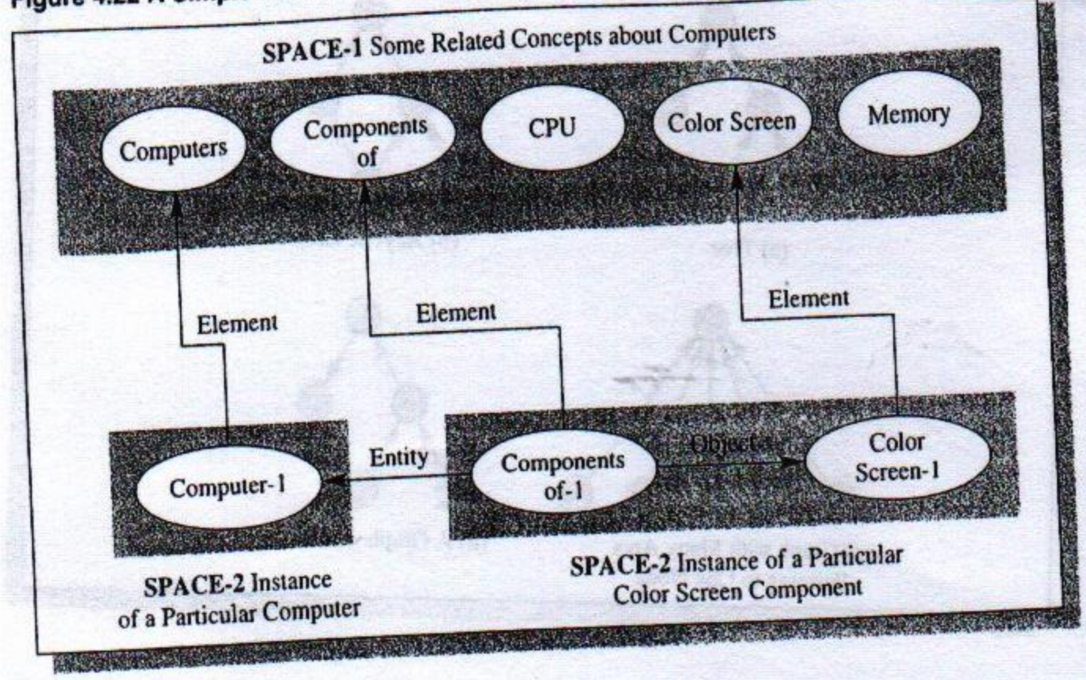
Figure 4.19 A Small Part of the PROSPECTOR Inference Net for Porphyry Copper



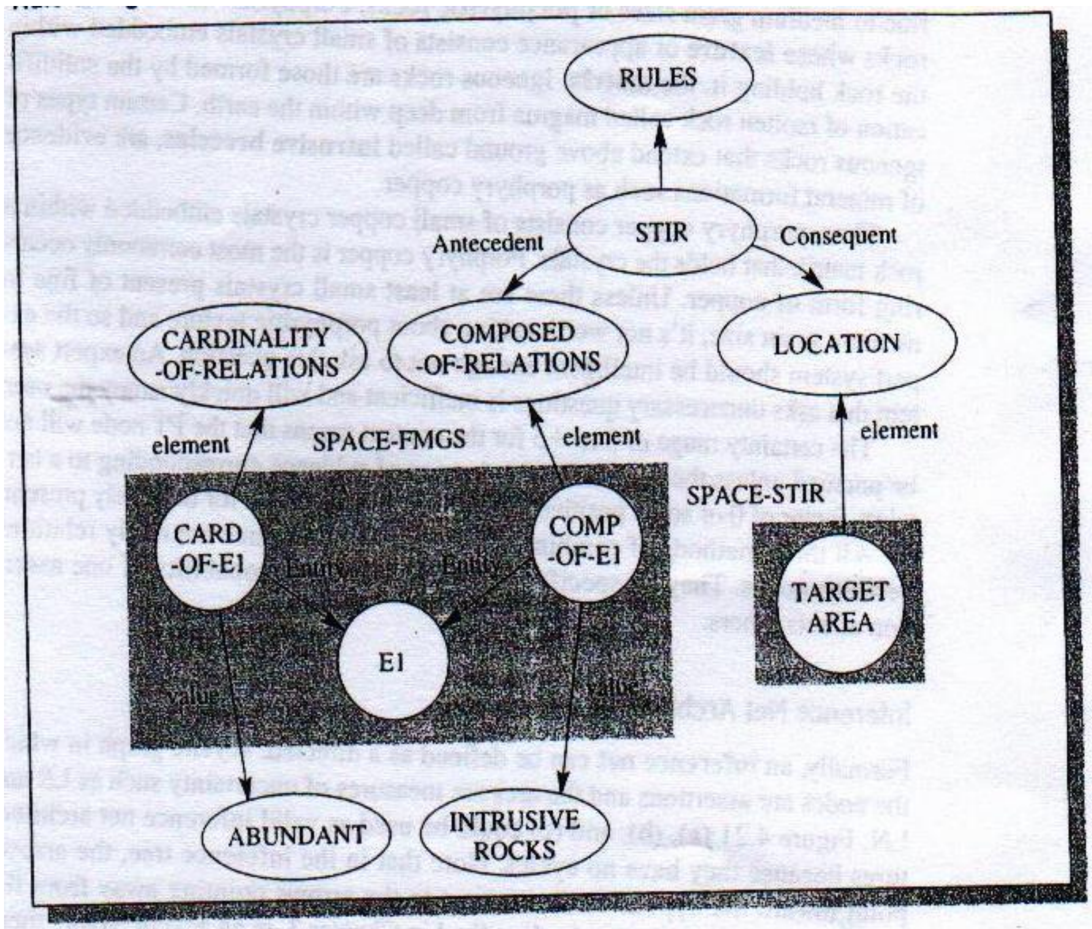
Inference Net Architecture

- Directed Acyclic Graph!
- Partitioned:

Figure 4.22 A Simple Partitioned



Inference Net Architecture



Evidence Combination Methods

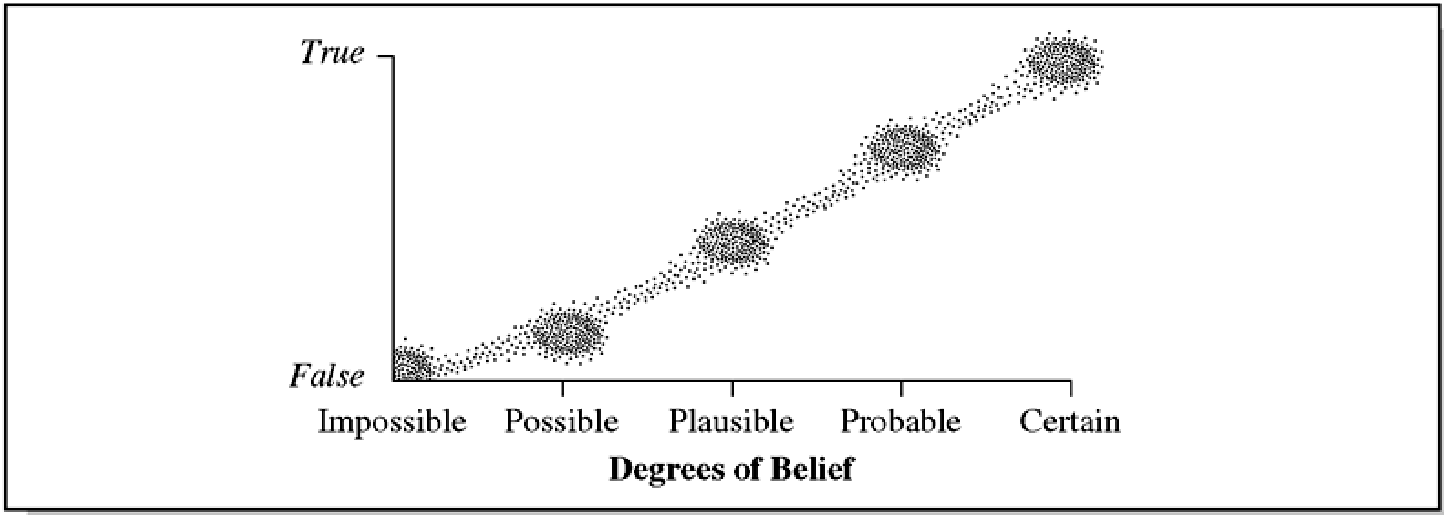
- Logical Combinations:
  - And / Or Nodes.
- Weighted Combination:
  - $O(H|E_1 \cap E_2 \cap .. \; E_n) = [\prod LS_i] \times O(H)$
  - $\log(O(H|E_1 \cap E_2 \cap .. \; E_n))$

    $= [\sum \log(LS_i)] + \log(O(H))$

Types of Belief

- Possible – no matter how remote, the hypothesis cannot be ruled out.
- Probable – there is some evidence favoring the hypothesis but not enough to prove it.
- Certain – evidence is logically true or false.
- Impossible – it is false.
- Plausible – more than a possibility exists.

Figure 4.20 Relative Meaning of Some Terms Used to Describe Evidence



Propagation of Probabilities

- The chapter examines the classic expert system PROSPECTOR to illustrate how concepts of probability are used in a real system.
- Inference nets like PROSPECTOR have a static knowledge structure.
- Common rule-based system is a dynamic knowledge structure.

Summary

- In this chapter, we began by discussing reasoning under uncertainty and the types of errors caused by uncertainty.
- Classical, experimental, and subjective probabilities were discussed.
- Methods of combining probabilities and Bayes' theorem were examined.
- PROSPECTOR was examined in detail to see how probability concepts were used in a real system.

Summary

- An expert system must be designed to fit the real world, not visa versa.
- Theories of uncertainty are based on axioms; often we don’t know the correct axioms – hence we must introduce extra factors, fuzzy logic, etc.
- We looked at different degrees of belief which are important when interviewing an expert.