Chapter Objectives

- Know the several sources of uncertainty in rule-based systems.
- Know that both probability-based and heuristic methods are employed to reason about an uncertain world.
- Know how fuzzy logic is used in a rule-based system to reason about uncertainty.
- Know how Bayes theorem is used in a rule-based system to reason about uncertainty.
- Know the conditions that must be met for Bayes theorem to be used for reasoning about uncertainty.
In this chapter we introduce methods used in rule-based systems to reason about uncertain information. In Section 13.1 we discuss sources of uncertainty and overview general techniques for dealing with uncertainty in rule-based systems. In Section 13.2 we detail fuzzy logic as a strategy for reasoning about uncertainty. In Section 13.3 we describe a Bayesian approach to handling uncertain information. A complete treatment of uncertainty is beyond the scope of this book. For a comprehensive description of uncertainty in rule-based systems see Giarratano and Riley (1998).

13.1 Uncertainty: Sources and Solutions

We live in a world filled with uncertainty. Meteorologists provide us with next-day forecasts that are, on average, 85% accurate. An analyst recommends a stock for purchase after which the stock price declines. We work hard at our jobs, yet we cannot be certain about our employment status a week, a month, or a year from now. Because of our uncertain world, we become experts at dealing with the unexpected.

Sources of Uncertainty

As expert systems are built to model human reasoning, many expert system applications have the capability to reason about uncertain facts and events. From a rule-based perspective, the prospect of uncertainty leads to many issues. To illustrate this point, consider the following production rule.

**Rule 1: Large Package Rule**

IF package size is large

THEN send package UPS

At least three forms of uncertainty are built into this simple rule. First, there is uncertainty associated with the rule antecedent. Exactly what is the definition of a large package? A second source of uncertainty is the level of confidence we have in the rule itself. That is, provided we agree that the size of a package is large, to what degree can we be sure that UPS is the best medium for package delivery? A third issue requires a methodology for combining pieces of uncertain information. For example, suppose that we are 80% certain that the size of the package we wish to mail is large, and we are 90% certain that the large package rule is valid. How do we use these values to compute an overall level of confidence for shipping the package via UPS?

New questions arise when we have rules with multiple antecedent conditions and/or multiple rules with the same consequent. For example, consider rule 2.
Rule 2: Heavy and Speedy Rule

IF package weight is heavy
AND speedy delivery is required
THEN send package UPS

Rule 2 tells us to ship a package by UPS provided the package is heavy and that a speedy delivery is a must. But what if one or both of the antecedent conditions for rule 2 are less than 100% certain? We must devise a way to combine the evidence for the two antecedent conditions so as to give an overall confidence for the rule antecedent. Once computed, the confidence of the rule antecedent can be applied to obtain a confidence value for the rule consequent.

Next, consider a heavy, large package that requires speedy delivery. These conditions activate rule 1 and rule 2. We must now have a method to combine the information contained in each rule to compute an overall level of confidence for UPS delivery. New scenarios develop if we consider uncertainties associated with the terms heavy and speedy. As you can see, the problem of how to deal with uncertainty is not easily solved.

General Methods for Dealing with Uncertainty

Several methods have been implemented to handle the issues surrounding uncertainty. These methods are described as probability-based methods and heuristic methods. Probability-based methods are exact reasoning strategies that make use of standard probability theory. Probability-based methods can be further subdivided into objective, experimental, and subjective probability techniques.

**Objective probability** assigns probability values for well-defined problems. The following are three problems that can be solved with objective probability techniques:

1. What is the probability of rolling a die and seeing a six or a two?
2. What is the probability of picking four consecutive face cards from a deck of 52 cards?
3. What is the probability that someone will win the New York lottery?

Problems such as these exist in an ideal world void of the subjectivity inherent in our environment. These problems are interesting to study, but provide little guidance to help us with the decisions we make each day.

**Experimental probabilities** are based on frequency distributions obtained through sampling. Experimental probability techniques are used to develop tables for life, health, and accident insurance. Experimental probability is also the basis for many types of estimates obtained from polling a population. **Subjective probabilities** are probability
values based on personal opinion. If the opinions are those of a human expert, subjective probabilities are useful because they are backed by years of experiential learning.

**Heuristic techniques**, also know as inexact techniques, are needed when the assumptions necessary to correctly apply a probability-based approach cannot be met. Because of the inexact nature of data, heuristic techniques are more often the method of choice for rule-based systems. The first well-known heuristic technique for dealing with uncertainty in rule-based systems used **certainty factors**. Certainty factors were initially developed for the MYCIN expert system when researchers discovered that doctors were unwilling to state degrees of truth as traditional probabilities. A certainty factor (CF) is a number between –1 and 1 associated with a fact or rule. A fact or rule with a CF of 1 is believed to be true with 100% certainty. A CF of –1 indicates that the fact or rule is false. A CF of 0 offers no evidence to support or refute the truth value of a fact or rule. Many of today’s rule-based expert system shells use certainty factors to reason with uncertain information.

**Fuzzy logic** is an increasingly popular heuristic technique used to reason about uncertainty in rule-based systems. Fuzzy logic was first introduced by Zadeh (1965) as a problem-solving method for dealing with the uncertainty inherent in words with ambiguous meanings. Everyday words such as young, large, tall, more, some, high, and low, as well as thousands of others fit the category of terms appropriate for fuzzy reasoning. Fuzzy logic has been applied to many fields, including mechanical device control, medical diagnosis, and psychology to name a few (Maiers and Sherif, 1985). Here we limit our discussion to the application of fuzzy logic to rule-based reasoning.

### 13.2 Fuzzy Rule-Based Systems

Fuzzy logic has its foundation in mathematical set theory. However, traditional set theory uses two-valued boolean logic to represent set membership. That is, an element either is or is not a member of a specific set. To illustrate, we return to the familiar credit card promotion database. Consider the following rule:

**Rule 1: A Crisp Rule**

IF age < 30
AND previous accepts >= 3
THEN life insurance promotion is yes

Rule 1 states that a person under the age of 30 who has accepted three or more previous promotions belongs to the set of candidates likely to accept a new life insurance promotional offering. Boolean reasoning is *crisp* in that it does not allow room for speculation about levels of possible set membership. Therefore, with boolean logic, someone
aged 30 who has accepted five previous promotions does not belong to the set of candidate individuals for a new life insurance promotion. To contrast, fuzzy logic relies on special functions that allow us to associate varying degrees of membership with the values of numeric attributes such as age and previous accepts. Here is a fuzzified form of rule 1.

**Rule 2: A Fuzzy Rule**

**IF** age is young  
**AND** previous_accepts are several  
**THEN** life_insurance_accept is high

The terms age, previous_accepts, and life_insurance_accept are known as linguistic variables. The values associated with a linguistic variable are called linguistic values. Each linguistic value has an associated fuzzy set. A fuzzy set tells us the degree of membership a specific numeric value has within the set representing the associated linguistic value. For example, suppose the linguistic variable age has three linguistic values: young, middle_aged, and old. The fuzzy set associated with age is young maps each numeric age to a value between 0 and 1. The output of the mapping tells us the probability that each age is a member of the fuzzy set age is young. Similar statements hold true for the fuzzy sets associated with age is middle_aged and age is old. Let’s take a closer look at the concept of fuzzy set.

**Fuzzy Sets**

A fuzzy set is characterized by a membership function that may be continuous or discrete. We first consider the continuous case. Figure 13.1 displays four fuzzy sets for the linguistic variable distance_from x. Figure 13.1(a) shows a membership function for the linguistic value far_from x. Figure 13.1(b) offers a membership function for close_to x. Figures 13.1(c) and (d) give fuzzy sets for the respective linguistic values approaching and moving_from x. Notice that a numeric value half way between 0 and x has a degree of membership well above 0.5 in the fuzzy set far_from x. However, the degree of membership for the same point is below 0.5 for the fuzzy sets approaching x, and moving_from x.

Fuzzy sets can also be implemented using discrete sets. For example, a fuzzy set for age is young may be shown as:

\{(0/1.0), (5/0.95), (10/0.75), (15/0.50), (20/0.35), (30/0.10), (50/0.0)\}

where \((x/y)\) indicates a set membership value of \(y\) for an individual with age \(x\). That is, we are 95% sure that a five-year-old individual belongs to the set age is young. However, our confidence that someone aged 30 is a member of the same set is only 10%.
It is obvious that a finite fuzzy set for age is young cannot represent all possible numeric values for age. So how do we associate a probability of membership with an age not included in the fuzzy set definition? One common method is to interpolate a membership score using values contained within the set. A second approach is to train a neural network using the values defined within the finite set. Once trained, the network is used to compute the membership scores for new values not included in the original data.

**Fuzzy Reasoning: An Example**

In this section we expand the familiar credit card promotion example to show the fuzzy rule-based reasoning process. Our objective is to develop a system able to com-
pute a numeric value representing a degree of confidence for an individual’s likelihood of choosing a new life insurance promotional offering. The following are three fuzzy rules for our example.

**Rule 1: Accept Is High**
IF age is young
AND previous_accepts are several
THEN life_insurance_accept is high

**Rule 2: Accept Is Moderate**
IF age is middle_aged
AND previous_accepts are some
THEN life_insurance_accept is moderate

**Rule 3: Accept Is Low**
IF age is old
THEN life_insurance_accept is low

Fuzzy sets representing values for the linguistic variables *age, previous_accepts,* and *life_insurance_accept* are shown in Figure 13.2. Notice that each linguistic variable has three fuzzy sets. For example, the linguistic variable *age* shows one fuzzy set for *age is young,* one for *age is middle_aged,* and a third for *age is old.* Also, notice that individuals in their late 20s and early 30s hold a degree of membership in the fuzzy set for *age is young* as well as the fuzzy set for *age is middle_aged.* Likewise, the fuzzy set for *previous_accepts are some* intersects with the fuzzy sets for *previous_accepts are few* and *previous_accepts are several.* Finally, the fuzzy set for *life_insurance_accept is moderate* shares values with *life_insurance_accept is low* and *life_insurance_accept is high.*

Now that we have a set of rules and fuzzy sets defining the values for the linguistic variables, we are ready to demonstrate the fuzzy inference process. The most common inference method is a four-step procedure first introduced by Mamdani and Assilian (1975). The technique is best illustrated by showing the computations for a specific set of input values. Let’s compute a *life_insurance_accept* confidence score for a 33-year-old individual who has taken advantage of five previous promotions.

**Step 1: Fuzzification**
The values input into a fuzzy rule-based system are numeric. The fuzzification process converts these crisp numeric values to their corresponding degrees of membership within each fuzzy set. For our example, Figure 13.2 shows that a customer aged 33 who has accepted five promotions has the following degrees of fuzzy set membership.
Fuzzy Set Degree of Membership

age is middle-aged 0.25
age is young 0.10
previous_accepts are some 0.20
previous_accepts are several 0.60

Figure 13.2 • Fuzzy sets for age, previous_accepts, and life_insurance_accept

(a) Fuzzy sets for age:
- Young
- Middle-Aged
- Old

(b) Fuzzy sets for previous_accepts:
- Few
- Some
- Several

(c) Fuzzy sets for life_insurance_accept:
- Low
- Moderate
- High

<table>
<thead>
<tr>
<th>Fuzzy Set</th>
<th>Degree of Membership</th>
</tr>
</thead>
<tbody>
<tr>
<td>age is middle_aged</td>
<td>0.25</td>
</tr>
<tr>
<td>age is young</td>
<td>0.10</td>
</tr>
<tr>
<td>previous_accepts are some</td>
<td>0.20</td>
</tr>
<tr>
<td>previous_accepts are several</td>
<td>0.60</td>
</tr>
</tbody>
</table>
Step 2: Rule Inference

Rule inference applies the fuzzified input values obtained in step 1 to the fuzzy rule base. Two of the rules in our rule base contain rule antecedents with more than one condition. Therefore we need a method to combine the values for the individual antecedent conditions in order to obtain an overall rule antecedent probability score. Several fuzzy set AND and OR operators have been defined to combine the antecedents of rules with multiple conditions. For our example, we use fuzzy set union for OR conditionals and fuzzy set intersection for AND conditions. Fuzzy set union outputs the maximum degree of membership for all OR conditions within a single rule. The fuzzy set AND operator outputs the minimum of all AND conditions contained within one rule. Let’s see which rules apply to our problem.

Upon examining the rules, we see that rule 1 fires because age is young and previous accepts are several each have a nonzero membership degree. Applying fuzzy set intersection to rule 1, we obtain a value of 0.10 (the minimum of 0.10 and 0.60) for the rule antecedent. The antecedent must now be applied to the conclusion component of the rule. A usual technique is to clip the rule consequent membership function at the height specified by the rule antecedent condition. Recall the consequent condition for rule 1 is life_insurance_accept is high. Figure 13.3(a) shows that the membership function for life_insurance_accept is high has been clipped at the height 0.10.

Because age is middle_aged and previous accepts are some have nonzero membership degrees, rule 2 also fires, giving a truth value of 0.20 for the rule antecedent. The clipped membership function for life_insurance_accept is moderate is displayed in Figure 13.3(b). Rule 3 does not fire because the membership degree for age is old is 0.

Step 3: Rule Composition

From step 2, we see that the first two rules contribute to the truth value of the linguistic variable life_insurance_accept. In step 3, we combine the truth values for all rules that have fired to form a single fuzzy set. Figure 13.4 shows the fuzzy set formed by combining the output of rule 1 and rule 2. Notice that the fuzzy set displayed in Figure 13.4 is simply a modification of the fuzzy set for life_insurance_accept shown in Figure 13.3. As rule 3 did not fire, the horizontal axis of the modified fuzzy set begins at 25 rather than 0. Also, the height of the fuzzy set reflects the clipping action seen with the execution of rules 1 and 2.

Step 4: Defuzzification

The purpose of defuzzification is to convert the fuzzy set formed in step 3 to a final crisp output score representing the degree of confidence for an individual’s likelihood
of choosing a new life insurance promotional offering. One technique is simply to associate the largest rule output value with the consequent condition. Applying this procedure to our example, we are 20% confident that the 33-year-old customer who has accepted five previous promotional offerings is a likely candidate for the life insurance promotion. You will also notice that applying this technique eliminates the need for step 3.

A second, and often more exact, technique computes the center of gravity of the fuzzy set constructed in step 3. The center of gravity is that point which vertically
divides the constructed fuzzy set into two equal masses. The mathematical formula for computing the center of gravity for continuous data is shown in Equation 13.1.

\[ \frac{\int f(x)x \, dx}{\int f(x) \, dx} \]  

(13.1)

For a sampling of data, we make use of the discrete counterpart of the equation. Specifically,

\[ \frac{\sum_i f(x_i)x_i}{\sum_i f(x_i)} \]  

(13.2)

The computation for our example is straightforward. We simply choose representative points along the horizontal axis of the fuzzy set shown in Figure 13.4 and apply the formula given in 13.2. To illustrate, let’s choose axis points 35, 45, 55, 65, 75, 85, and 95. We multiply each chosen point by its corresponding functional value to obtain the numerator value for Equation 13.2. For the denominator, we simply sum the functional values corresponding to the chosen axis points.
The result gives us a confidence value of 59% that the customer in question is a likely candidate for the life insurance promotion. As you can see, the level of confidence associated with the output attribute is noticeably increased with the center of gravity computation.

Finally, although the computations for the previous example are minimal, a larger rule set is likely to have several rules that contribute to the truth value of a rule consequent. A user-defined threshold setting can be used to avoid having several rules fire that ultimately contribute little to a final calculation. The threshold value prevents any rule whose antecedent condition does not meet a minimum setting from firing.

13.3 A Probability-Based Approach to Uncertainty

Bayes theorem offers a mathematically correct approach to reasoning about uncertainty. Therefore, when applicable, Bayes theorem is often the method of choice for reasoning about uncertainty with rule-based expert systems. Bayes theorem was first used for computing uncertainty in the expert system PROSPECTOR. PROSPECTOR helps geologists determine regions favorable for mineral exploration. We first introduced Bayes theorem in Chapter 10, where we described how it can be used for supervised data mining. Here is one form of Bayes theorem where the denominator shows the expanded formula for computing the probability of evidence \( E \).

\[
0.20 \times (35 + 45 + 55) + 0.10 \times (65 + 75 + 85 + 95) \\
0.20 \times 3.0 + 0.10 \times 4.0
\]

\[
P(H | E) = \frac{P(E | H) \times P(H)}{P(E | H) \times P(H) + P(E | \neg H) \times P(\neg H)} \quad (13.3)
\]

where

- \( P(H) \) is the a priori probability that the hypothesis to be tested is true
- \( P(\neg H) \) is the a priori probability that \( H \) is false. \( P(\neg H) = 1 - P(H) \)
- \( E \) is the evidence associated with the hypothesis. The denominator expression represents \( P(E) \)
- \( P(E | H) \) is the conditional probability of the evidence knowing the hypothesis is true
- \( P(E | \neg H) \) is the conditional probability of the evidence knowing the hypothesis is false
To see how Bayes theorem can be applied to a rule base, we return to the domain of credit card promotional offerings. Suppose a human expert has given us the following rule.

**Rule 1: Age Rule**

IF age < 30

THEN life insurance promotion is yes

The rule states that if a person is less than age 30 the person will take advantage of a life insurance promotional offering. We wish to use Bayes theorem to associate a probability with the outcome of the rule. That is, we want to compute $P(\text{life insurance promotion is yes} \mid \text{age } < 30)$. To accomplish this, we need values for the following probabilities:

- $P(\text{age } < 30 \mid \text{life insurance promotion is yes})$
- $P(\text{age } < 30 \mid \text{life insurance promotion is no})$
- $P(\text{life insurance promotion is yes})$
- $P(\text{life insurance promotion is no})$

Let’s use the data in Table 13.1 to compute estimates for these values. The table data shows 14 instances. The first column displays the age of each person represented by the data instances. The second column shows the total number of promotional offerings each person has accepted. The life insurance promotion, to be distributed to a new sampling of credit card customers, was previously presented to the individuals represented by the table data. The third column indicates whether these individuals took advantage of the life insurance promotion. Here are the computed probability values:

- $P(\text{age } < 30 \mid \text{life insurance promotion is yes}) = 5/7$
- $P(\text{age } < 30 \mid \text{life insurance promotion is no}) = 3/7$
- $P(\text{life insurance promotion is yes}) = 1/2$
- $P(\text{life insurance promotion is no}) = 1/2$

The numerator term for the first listed conditional probability, $P(\text{age } < 30 \mid \text{life insurance promotion is yes})$, tells us that five of the seven individuals who said yes to the life insurance promotion are also under age 30. These five individuals are represented in the table as instances 1, 3, 6, 8, and 14. The denominator is simply the total
number of table instances showing a value of yes for life insurance promotion. The remaining three conditional probabilities are calculated in a similar fashion. We can now make the desired computation. With $H$ representing life insurance promotion = yes we have,

$$P(H \mid age < 30) = \frac{(5/7) \times (1/2)}{(5/7) \times (1/2) + (3/7) \times (1/2)} = 0.625$$

The result tells us that an individual under age 30 will accept the life insurance promotional offering with a probability of 62.5%. Also, because probability theory requires $P(H \mid E) + P(\neg H \mid E) = 1$, we conclude that the probability of life insurance promotion = no given age < 30 is 37.5%.

**Multiple Evidence with Bayes Theorem**

More often than not, experts have many pieces of evidence that must be considered when accepting or refuting a hypothesis. A main reason why experts are able to solve
complex problems is that they can combine multiple pieces of evidence, even when the evidence is uncertain and/or incomplete. We can extend Bayes theorem to provide us with the ability to reason about multiple evidence as well as multiple hypotheses (necessary for more than two hypotheses). With \( n \) possible hypotheses and \( j \) pieces of evidence, Equation 13.4 shows the extended formula for the probability of hypothesis \( H_i (1 \leq i \leq n) \).

\[
P(H_i \mid E_1 \& E_2 \& \ldots \& E_j) = \frac{P(E_1 \& E_2 \& \ldots \& E_j \mid H_i) \times P(H_i)}{\sum_{k=1}^{n} P(E_1 \& E_2 \& \ldots \& E_j \mid H_k) \times P(H_k)}
\]  

(13.4)

It is obvious that as the evidence supporting the current hypothesis increases, obtaining values for the conditional probabilities quickly becomes an issue. For example, consider asking a medical expert to supply conditional probabilities for every possible combination of symptoms a patient may or not experience given the patient has the flu!

One way to circumvent the problem is to assume that the pieces of evidence are conditionally independent. The assumption allows us to modify Equation 13.4 as follows.

\[
P(H_i \mid E_1 \& E_2 \& \ldots \& E_j) = \frac{P(E_1 \mid H_i) \times P(E_2 \mid H_i) \times \ldots \times P(E_j \mid H_i) \times P(H_i)}{\sum_{k=1}^{n} P(E_1 \mid H_k) \times P(E_2 \mid H_k) \times \ldots \times P(E_j \mid H_k) \times P(H_k)}
\]  

(13.5)

Clearly, the assumption permits us to analyze each piece of evidence solely on the basis of how it relates to the hypotheses in question. To illustrate how the formula is used, consider the following rule, which adds an additional piece of evidence to rule 1.

**Rule 2: Age & Previous Accepts Rule**

**IF** \( \text{age} < 30 \)

AND \( \text{previous accepts} > 2 \)

**THEN** \( \text{life insurance promotion is yes} \)

Let's employ the revised equation together with the data in Table 13.1 to compute the probability of a life insurance promotion accept given that the customer is under age 30 and has accepted more than two previous promotional offerings.

\[
P(H \mid \text{age} < 30 \& \text{previous Accepts} > 2) = \frac{(5/7) \times (4/7) \times (1/2)}{(5/7) \times (4/7) \times (1/2) + (3/7) \times (2/7) \times (1/2)}
\]

\[
\approx 0.76923
\]

The value 5/7 is the conditional probability associated with \( \text{age} < 30 \) given \( \text{life insurance promotion is yes} \). Likewise, the value 4/7 represents the conditional
probability for previous accepts > 2 given life insurance promotion is yes. The value 3/7 provides the conditional probability of age < 30 given life insurance promotion is no. Finally, 2/7 tells us the conditional probability that previous accepts > 2 given life insurance promotion is no. As you can see, the new evidence has a positive effect on the rule consequent.

### Likelihood Ratios: Necessity and Sufficiency

The data in Table 13.1 make it easy for us to determine the a priori and conditional probabilities for our simple problem. However, it is often the case that we must rely on the knowledge of a human expert to supply the necessary probabilities. With this approach, the process of determining conditional and a priori probabilities becomes part of the knowledge acquisition process. Experience has shown that many experts are not comfortable providing exact values for conditional probabilities but are willing to state their belief in the truth of a piece of evidence as a likelihood ratio. For evidence $E$ and hypothesis $H$, two likelihood ratios, the **likelihood of sufficiency** (LS) and the **likelihood of necessity** (LN) are defined as shown in Equation 13.6.

$$
LS = \frac{P(E \mid H)}{P(E \mid \neg H)} \quad LN = \frac{P(\neg E \mid H)}{P(\neg E \mid \neg H)}
$$

Large values for LS favor the evidence $E$ for concluding the hypothesis. As LS approaches infinity, the evidence is said to be sufficient for concluding the hypothesis because the evidence is not seen when the hypothesis is false. Also, as LS approaches 0, the evidence becomes unfavorable for concluding $H$. A value of 1 for LS indicates that the evidence is not relevant for concluding or refuting the hypothesis. In contrast, large values of LN indicate the lack of evidence $E$ is favorable for concluding $H$. Values of LN approaching 0 indicate $E$ is necessary for concluding $H$ because a lack of the evidence supports a false hypothesis. Finally, a value of 1 for LN indicates the evidence or lack thereof is irrelevant to the truth value of $H$.

Given the definition for LS, we can state an alternative form for $P(H \mid E)$ by dividing both the numerator and denominator of Equation 13.3 by $P(E \mid \neg H)P(\neg H)$. The revised equation is shown in Equation 13.7.

$$
P(H \mid E) = \frac{LS \times O(H)}{1 + LS \times O(H)}
$$

where $O(H)$ is the prior odds on $H$ defined as $P(H)/P(\neg H)$. 
In a similar manner, we can state $P(H \mid \neg E)$ as

$$P(H \mid \neg E) = \frac{LN \times O(H)}{1 + LN \times O(H)}$$

(13.8)

To apply likelihood ratios to a rule-based system, we associate a sufficiency and necessity value with each piece of evidence. Here are modified forms of the two rules presented earlier together with additional rules having life insurance promotion is no as their consequent.

**Rule 1: Age < 30 Rule**

IF age < 30 (LS ≈ 1.67, LN = 0.5)

THEN life insurance promotion is yes

**Rule 2: Age < 30 and Previous Accepts > 2 Rule**

IF age < 30

AND previous accepts > 2 (LS = 2.0, LN = 0.6)

THEN life insurance promotion is yes

**Rule 3: Age >= 30 Rule**

IF age >= 30 (LS = 2.0, LN = 0.6)

THEN life insurance promotion is no

**Rule 4: Age < 30 and Previous Accepts <= 2 Rule**

IF age < 30

AND previous accepts <= 2 (LS = 1.67, LN = 0.5)

THEN life insurance promotion is no

The values for LS and LN can be computed directly from Table 13.1. For example, LS for age < 30 in rule 1 is obtained by dividing 5/7 by 3/7. It is worth noting that we cannot derive the value of LS knowing the value of LN and vice versa.

The advantage of associating values of LS and LN with each piece of evidence is seen when we wish to update probabilities as evidence accumulates in an iterative fashion. This is desirable when a system is used in an interactive consultation mode. To illustrate the technique, suppose we wish to determine the probability that a new credit card customer will accept the life insurance promotion. Using the data in Table 13.1, we see that the value for $O(H)$ is 1.0 because an equal number of individuals accepted and rejected the previous life insurance promotion. Next, suppose we learn the new customer is 25 years old. This causes rule 1 to fire. We apply Equation 13.7 to compute a value for $P(\text{life insurance promotion is yes} \mid \text{age} < 30)$. 
The effect of the rule execution is threefold. First, if the consultation terminates, we conclude that the probability of this individual accepting the life insurance promotion is 62.5%. Second, if the consultation continues, the value 0.625 becomes the new a priori probability for any rule showing life insurance promotion is yes as its consequent. Third, as \( P(H | E) \) and \( P(\neg H | E) \) must sum to 1, the a priori probability of any rule consequent showing life insurance promotion is no becomes 0.375.

It is worth noting that we can also obtain the value 0.375 by applying rule 3. With \( H \) as life insurance promotion is no, and \( \neg E \) as age >= 30, we use the value of LN associated with age >= 30 shown in rule 3 to compute the new a priori probability for life insurance promotion is no.

\[
P(H | \neg E) = \frac{0.6 \times 1.0}{1 + 0.6 \times 1.0} = 0.375
\]

In any case, knowing that our credit card holder is 25 years old has allowed us to more accurately predict whether he or she will accept the life insurance promotional offering. Next, suppose we learn that our customer has accepted one previous promotional offering. That is, \( previous \ accepts = 1 \). The new information fires rule 4. With \( H = life \ insurance \ promotion \ is \ no \) and evidence \( previous \ accepts = 1 \) we have:

\[
P(H | E) = \frac{1.6 \times (0.375 / 0.625)}{1 + 1.6 \times (0.375 / 0.625)} = 0.5
\]

Notice that the probability for life insurance promotion is no increased from 0.375 to 0.5. Thus the probability for life insurance promotion is yes decreases from 0.625 to 0.5. If the consultation continues, these values become new a priori probabilities for respective rules showing life insurance promotion as their consequent. Finally, it is important to note that an identical result is obtained by simply applying Equation 13.3 to the accumulated information age <= 30 and previous accepts = 1. However, the advantage of an interactive consultation is that it gives us the ability to easily explore alternative scenarios as evidence accumulates.

**General Considerations**

Bayes theorem offers a mathematically correct approach for reasoning about uncertainty. The technique is particularly appealing when accurate statistical data, which allows con-
ditional and a priori probabilities to be easily computed, is readily available. However, certain criteria must be satisfied for valid application of the theorem. Here are a few issues that may cause problems when applying Bayes theorem to a specific problem:

1. \( P(H \mid E) + P(\sim H \mid E) \) must sum to 1. During the development of MYCIN, medical experts were unwilling to adhere to this requirement.

2. Conditional independence between multiple pieces of evidence must be assumed. Although the assumption of conditional independence improves the situation, several conditional probabilities must still be computed.

3. Although critical, prior probabilities are often unobtainable.

4. Large amounts of data must often be gathered to obtain reasonable estimates for conditional probabilities.

These issues make Bayes theorem of limited use for most rule-based applications. However, Bayes theorem has been the inspiration for many of the inexact techniques seen with today’s rule-based systems.

13.4 Chapter Summary

Many rule-based systems have the ability to reason about uncertain facts and events. At least three forms of uncertainty are built into a single rule. First, there is uncertainty associated with the rule antecedent. A second source of uncertainty is the level of confidence we have in the rule itself. A third focus requires a methodology for combining pieces of uncertain information.

Several methods have been implemented to handle uncertainty. Two general categories of solutions are probability-based methods and heuristic methods. Probability-based methods are exact reasoning strategies that make use of standard probability theory. Heuristic techniques, also known as inexact methods, are necessary when the assumptions required to correctly apply a probability-based approach cannot be met. Because of the inexact nature of data, heuristic techniques are often the method of choice for rule-based systems.

Fuzzy logic is an accepted heuristic technique for reasoning about uncertainty in rule-based systems. Fuzzy logic was first introduced by Zadeh (1965) as a problem-solving method for dealing with the uncertainty inherent in words with ambiguous meanings. Fuzzy logic relies on fuzzy set membership, which allows us to associate varying degrees of membership with the elements of a set.

Bayes theorem offers a mathematically correct approach to reasoning about uncertainty. The technique is particularly appealing when accurate statistical data, which allows conditional and a priori probabilities to be easily computed, is readily available. However,
certain criteria must be satisfied for valid application of the theorem. Bayes theorem was 
first used for computing uncertainty in the expert system PROSPECTOR.

13.5 Key Terms

Center of gravity. The point that vertically divides a fuzzy set into two equal masses.

Certainty factor. A value between −1 and 1 associated with a fact or rule. A certainty 
factor of 1 expresses complete confidence in the fact or rule. A value of −1 
indicates a lack of confidence. A certainty factor of 0 offers no evidence to sup-
port or refute the fact or rule.

Clipping. Cutting the height of a rule consequent membership function at a point 
specified by the truth of the rule’s antecedent condition.

Experimental probability. A probability computed from data obtained through 
sampling.

Fuzzy logic. A problem-solving method for dealing with the uncertainty inherent 
in words with ambiguous meanings.

Fuzzy set. A set associated with a linguistic value. The fuzzy set tells us the degree of 
membership that a specific numeric value has within the set representing the lin-
guistic value.

Fuzzy set intersection. A fuzzy operator to compute the overall probability of a 
rule antecedent condition with rules having one or several AND conditionals. 
The fuzzy set AND operator outputs the minimum of all AND conditions con-
tained within an individual rule.

Fuzzy set union. A fuzzy operator to compute the overall probability of a rule an-
tecedent condition with rules having one or several OR conditionals. The opera-
tor outputs the maximum degree of membership for all OR conditions in the 
antecedent of a rule.

Heuristic technique. Any method used to reason about uncertainty that does not 
make assumptions about the nature of the data.

Likelihood of necessity. The conditional probability of evidence \( E \) being false 
given hypothesis \( H \) is true divided by the conditional probability of \( E \) being false 
given \( H \) is false.

Likelihood of sufficiency. The conditional probability of evidence \( E \) being true 
given hypothesis \( H \) is true divided by the conditional probability of \( E \) being true 
given \( H \) is false.

Linguistic value. A value associated with a linguistic variable.
**Linguistic variable.** A fuzzy logic variable. Each value of a linguistic variable has an associated fuzzy set.

**Objective probability.** A crisp probability value.

**Probability-based methods.** Exact reasoning strategies that make use of standard probability theory.

**Subjective probability.** A probability value based on personal experience and personal opinion.

### 13.6 Exercises

#### Review Questions

1. Differentiate between the following:
   a. Subjective and objective probability
   b. Subjective and experimental probability
   c. Likelihood of sufficiency and likelihood of necessity
   d. Fuzzy set union and fuzzy set intersection
   e. Certainty factor and probability

2. Briefly explain each step of the fuzzy logic inference method introduced by Mamdani and Assilian (1975).

3. Computational Question #8 in Chapter 12 asks you to build goal trees to help solve several problems. For each problem, decide if a rule-based system designed to solve the problem should contain a method for reasoning about uncertainty. For those problems where you believe uncertainty is a factor, determine whether a fuzzy logic or a Bayesian approach is a better choice. Justify each decision.

#### Data Mining Questions

**LAB** 1. In the second part of Section 13.2 we defined a fuzzy set for \( age = young \) as follows:

\[
\{(0/1.0), (5/0.90), (10/0.75), (15/0.50), (20/0.35), (30/0.10), (50/0.0)\}
\]

**LAB** Denotes exercise appropriate for a laboratory setting.
Use your IDA backpropagation software to train a neural network to associate an age with a probability of membership. Test and record the network output for the following ages:

\[ Age = 2, 7, 25, 40, 60. \]

**Computational Questions**

1. Here’s an interesting problem that can be solved with simple probability theory. You are given three boxes, two are empty and one contains one million dollars. You are asked to select any one of the three boxes. Next, after you are shown the empty contents of one of the two boxes you did not choose, you must decide if you would like to change your initial selection. Should you stay with your original choice or swap your box for the remaining unopened box?

2. Use the fuzzy sets and rules defined in Section 13.2 to compute a `life_insurance_accept` confidence score for a 30-year-old individual who has taken advantage of five previous promotions.

3. Suppose that the probability of someone liking sugar given that the person eats cereal X is 75%. We also know that 30% of the population eats cereal X. Finally, the probability of a person liking sugar and not eating cereal X is 40%. Use Bayes theorem to compute the probability that a person eats cereal X given that he or she likes sugar.

4. Use the rules defined in Section 13.3 together with Bayes theorem to compute the probability that a customer aged 25 who has accepted three previous promotions will accept the life insurance promotion.