CHAPTER

4

MODELING UNCERTAINTY

Farrokh Alemi and David H. Gustafson

This chapter shows the use of Bayesian probability models in forecasting market demand for healthcare services. Bayesian probability models can be used to help decision makers manage their uncertainties about future events. This chapter builds on the material covered in Chapter 3. While the previous chapter mostly focused on mathematical formulas, this chapter has a behavioral focus and describes how to interact with the experts. It shows how probability models can be constructed from expert opinions and used to forecast future events.

Statisticians can predict the future by looking at the past; they have developed tools to forecast the likelihood of future events from historical trends. For example, future sales can be predicted based on historical sales figures. Sometimes, however, analysts must forecast unique events that lack antecedents. Other times, the environment has changed so radically that previous trends are irrelevant. In these circumstances, the traditional statistical tools are of little use and an alternative method must be found. This chapter provides a methodology for analyzing and forecasting events when historical data are not available. This approach is based on Bayesian subjective probability models.

To motivate this approach, suppose an analyst has to predict the demand for a new health maintenance organization (HMO). Employees in HMOs are required to consult a primary care physician before visiting a specialist; the primary care physician has financial incentives to reduce the inappropriate use of services. Experience with HMOs shows they can cut costs by reducing unnecessary hospitalization. Suppose an investigator wants to know what will happen if a new HMO is set up in which primary care physicians have e-mail contact with their patients. At first, predicting demand for the proposed HMO seems relatively easy because there is a great deal of national experience with HMOs. But the proposed HMO uses technology to set it apart from the crowd: The member will initiate contact with the HMO via the computer, which will interview the member and send a summary to the primary care physician, who would then consult the patient's record and decide whether the patient should

This book has a companion web site that features narrated presentations, animated examples, PowerPoint slides, online tools, web links, additional readings, and examples of students' work. To access this chapter's learning tools, go to ache.org/DecisionAnalysis and select Chapter 4.

1. wait and see what happens;
2. have specific tests done before visiting;
3. take a prescription, which would be phoned to a nearby pharmacy, and wait to see if the symptoms disappear;
4. come in for a visit; or
5. bypass the primary care system and visit a specialist.

When the decision is made, the computer will inform the patient about the primary care physician's recommendations. If the doctor does not recommend a visit, the computer will automatically call the member a few days later to see if the symptoms have diminished. All care will be supervised by the patient's primary care physician.

This is not the kind of HMO with which anyone has much experience, but let's blend a few more uncertainties into the mix. Assume that the local insurance market has changed radically in recent years—competition has increased, and businesses have organized powerful coalitions to control healthcare costs. At the federal level, national health insurance is again under discussion. With such radical changes on the horizon, even data that are only two years old may be irrelevant. As if these constraints were not enough, the analyst needs to produce the forecast in a hurry. What can be done? How can an analyst predict demand for an unprecedented product? The following sections present ten steps for using Bayesian probability models to accomplish this task.

Step 1: Select the Target Event

To use the calculus of probabilities in forecasting a target event, the analyst needs to make sure that the events of interest are mutually exclusive (i.e., the events cannot occur simultaneously) and exhaustive (i.e., one event in the set must happen). Thus, in regards to the proposed HMO, the analyst might start with the following list of mutually exclusive events:

- More than 75 percent of the employees will join the HMO.
- Between 51 percent and 75 percent of the employees will join the HMO.

- Between 26 percent and 50 percent of the employees will join the HMO.
- Fewer than 25 percent of the employees will join the HMO.

Note that the events listed above are exhaustive (they cover all possibilities) and exclusive (no two events can co-occur). When gathering experts' opinions, the events being forecasted should be related to the experts' daily experiences and should be expressed in terms they are familiar with. For example, benefit managers are more comfortable thinking about individuals; if you plan to tap the intuitions of a benefit manager about the proposed HMO, you should discuss the probability in terms of one employee. To accomplish this, the list of target events might change to the following:

- The employee will join the proposed HMO.
- The employee will not join the proposed HMO.

It makes no difference in the analysis if you estimate the probability of one employee joining or the percent of the entire population joining; these are mathematically equivalent. It may make a big difference to the experts, however, so be sure to define the events of interest in terms familiar to the experts. Expertise is selective; if you ask experts about situations even slightly outside of their specific area or frame of reference, they will give either no answer or erroneous responses. Expertise has its limits; for example, some weather forecasters might be able to predict tomorrow's weather but not the weather for two day after.

It is preferable to forecast as few events as possible. However, many analysts and decision makers tend to work with a great deal of complexity in real situations. More complex target events for the proposed HMO might include the following:

- The employee will never join the proposed HMO.
- The employee will not join the HMO in the first year, but will join in the second year.
- The employee will join the HMO in the first year, but will withdraw in the second year.
- The employee will join the HMO in the first year and will stay.

Again, the events are mutually exclusive and exhaustive, but now they are more complex. The forecasts deal not only with the applicants' decisions but also with the stability of those decisions. People may join when they are sick and withdraw when they are well. Turnover rates affect administration and utilization costs, so information about the stability of the risk pool is important. In spite of the utility of such a categorization, it is difficult to combine two predictions; therefore, for reasons of simplicity and accuracy, it is preferable to design a separate model for each event: one model to predict whether the employee will join the HMO, and another to predict whether an enrollee will remain a member.

The target events must be chosen carefully because a failure to minimize their number may indicate that the analyst has not captured the essence of the uncertainty. One way of ensuring that the underlying uncertainty is being addressed is to examine the link between the forecasted events and the actions the decision maker is contemplating. Unless these actions differ radically from one another, some of the events should be combined. A model of uncertainty needs to be based on no more than two events unless there are solid reasons for the contrary. Even then, it is often best to build multiple models to forecast more than two events.

In the example being followed here, the analyst was interested in predicting how many employees will join the HMO (and also how many will not join), because this was the key uncertainty that investors needed to judge the business plan. To predict the number of employees who will join, the analyst can calculate the probability that an individual employee will join, $P(\text{Joining})$. If the total number of employees is n, then the number who will join is $n \times P(\text{Joining})$. Likewise, the number of employees who will not join is $n \times [1 - P(\text{Joining})]$.

Step 2: Divide and Conquer

The demand for the proposed HMO could be assessed by asking experts, "Out of 100 employees, how many will join?" However, this question is somewhat rhetorical, and an expert might answer, "Who knows? Some people will join the proposed HMO, others will not—it all depends on many factors." If the question is posed in these terms, it is too general to have a reasonable answer. When the predictions are complex (i.e., many contradictory clues, or components, must be evaluated), experts' predictions can be way off the mark. When talking with experts, the first task is to understand whether they can make the desired forecast with confidence and without reservation. If they can, then the analyst can rely on their forecasts and can save time. When they cannot, however, the analyst can disassemble the forecasts into judgments about the individual clues and forecast from these clues.

Errors in predictions, or judgments, may be reduced by breaking complex judgments into several components, or clues. Then the expert can specify how each clue affects the forecast, and the analyst can judge individual situations based on the clues. It is not necessary to make a direct estimate of the probability of the complex event; its probability can be derived from the clues and their influences on the complex judgments.

Let's take the example of the online HMO to see how one might follow this proposed approach. Nothing is totally new, and even the most radical health plan has components that resemble aspects of established plans. Though the proposed HMO is novel, experience offers clues to help the analyst predict the reaction to it. The success of the HMO will depend on factors that have influenced demand for services in other circumstances. Experience shows that the plan's success depends on the composition of the potential enrollees. In other words, some people have characteristics that dispose them toward or against joining the HMO. As a first approximation, the plan might be more attractive to young employees who are familiar with computers, to high-salary employees who want to save time, to employees comfortable with delayed communications on telephone answering machines, and to patients who want more control over their care. If most employees are members of these groups, you might reasonably project a high demand.

One thing is for certain: Each individual will have some characteristics (i.e., clues) that suggest a likelihood to join the health plan and some that suggest the reverse. Seldom will all clues point to one conclusion. In these circumstances, the various characteristics should be weighted relative to each other before the analyst can predict if an employee will join the health plan. Bayes's probability theorem provides one way for doing this. As discussed in Chapter 3, *Bayes's theorem* is a formally optimal model for revising existing opinions (sometimes called prior opinions) in light of new evidence or clues. The theorem states

$$\frac{P(H \mid C_1, \ldots, C_n)}{P(N \mid C_1, \ldots, C_n)} = \frac{P(C_1, \ldots, C_n \mid H)}{P(C_1, \ldots, C_n \mid N)} \times \frac{P(H)}{P(N)},$$

where

- $P(\ )$ designates probability;
- H marks a target event or hypothesis occurring;
- N designates the same event not occurring;
- $P(H)$ is the probability of H occurring before considering the clues;
- $P(N)$ is the probability of H not occurring before considering the clues;

- $C_1, \ldots, C_n$ mark the clues 1 through n;
- $P(H|C_1, \ldots, C_n)$ is the posterior probability of hypothesis H occurring given clues 1 through n;
- $P(N|C_1, \ldots, C_n)$ is the posterior probability of hypothesis H not happening given clues 1 through n;
- $P(C_1, \ldots, C_n|H)$ is the prevalence (likelihood) of the clues among the situations where hypothesis H has occurred;
- $P(C_1, \ldots, C_n|N)$ is the prevalence (likelihood) of the clues among situation where hypothesis H has not occurred;
- $P(H|C_1, \ldots, C_n) \div P(N|C_1, \ldots, C_n)$ is the posterior odds for the hypothesis occurring given the various clues;
- $P(H|C_1, \ldots, C_n) \div P(N|C_1, \ldots, C_n)$ is the likelihood ratio that measures the diagnostic value of the clues; and
- $P(H) \div P(N)$ is referred to as the prior odds and shows the forecast before considering various clues.

In other words, Bayes's theorem states that

Posterior odds after review of clues =
Likelihood ratio associated with the clues × Prior odds.

Using Bayes's theorem, if C_1 through C_n reflect the various clues, the forecast regarding the HMO can be rewritten as

$$\frac{P(\text{Joining}|C_1, \ldots, C_n)}{P(\text{Not Joining}|C_1, \ldots, C_n)} = \frac{P(C_1, \ldots, C_n|\text{Joining})}{P(C_1, \ldots, C_n|\text{Not Joining})} \times \frac{P(\text{Joining})}{P(\text{Not joining})}.$$

The difference between $P(\text{Joining}|C_1, \ldots, C_n)$ and $P(\text{Joining})$ is the knowledge of clues C_1 through C_n. Bayes's theorem shows how an opinion about an employee's reaction to the plan will be modified by the knowledge of his characteristics. Because Bayes's theorem prescribes how opinions should be revised to reflect new data, it is a tool for consistent and systematic processing of opinions.

Step 3: Identify Clues

An analyst can work with an expert to specify the appropriate clues in a forecast. The identification of clues starts with the published literature. Even if the task seems unique, it is always surprising how much has been published about related topics. There is a great deal of literature on predicting decisions to join an HMO, and even though these studies do not concern HMOs with the unique characteristics of the HMO in this example, reading them can help the analyst think more carefully about possible clues.

One will seldom find exactly what is needed in the literature. Once the literature search is completed, one should use experts to help identify clues for a forecast. Even if there is extensive literature on a subject, an analyst cannot expect to select the most important variables or to discern all important clues without the assistance of an expert. In a few telephone interviews with experts, an analyst can determine the key variables, get suggestions on measuring each one, and identify two or three superior journal articles.

Experts should be chosen on the basis of accessibility and expertise. To forecast HMO enrollment, appropriate experts might be people with firsthand knowledge of the employees, such as benefit managers, actuaries in other insurance companies, and personnel from the local planning agency. It is useful to begin discussions with experts by asking broad questions designed to help the experts talk about themselves:

> Analyst: Would you tell me a little about your experience with employees' choice in health plans?

The expert might respond with an anecdote about irrational choices by employees, implying that a rational system cannot predict everyone's behavior. Or, the expert might mention how difficult it is to forecast or how many years she has spent studying these phenomena. The analyst should understand what is occurring here: In these early responses, the expert is expressing a sense of the importance and value of her experience and input. It is vital to acknowledge this hidden message and to allow ample time for the expert to describe their contributions.

After the expert has been primed by recalling these experiences, the analyst should ask about characteristics that might suggest an employee's decision to join or not join the plan:

> Analyst: Suppose you had to decide whether an employee is likely to join, but you could not contact the employee.
> I was chosen to be your eyes and ears. What should I look for?

After a few queries of this type, the analyst should ask more focused questions:

> Analyst: What is an example of a characteristic that would increase the chance an employee will join the HMO?

The second question is referred to as a *positive prompt* because it elicits clues that would increase the chance of joining. *Negative prompts* seek clues that decrease the probability. An example of a negative prompt is the following:

Analyst: Describe an employee who is unlikely to join the proposed HMO.

Research shows that positive and negative prompts yield different sets of clues (Rothman and Salovey 1997; Rivers et al. 2005). Thus, modeling should start with clues that support the forecast, and then the clues that oppose it should be explored. Then responses can be combined so the model contains both sets of clues.

It is important to get opinions of several experts on what clues are important in the forecast. Each expert has access to a unique set of information; using more than one expert enables you to pool information and improve the accuracy of the recall of clues. At least three experts should be interviewed, each for about one hour. After a preliminary list of factors is collected during the interviews, the experts should have a chance to revise their lists, either by telephone, by mail, or in a meeting. If time and resources allow, the integrative group process is preferable for identifying the clues. This process is described in Chapter 6.

Suppose that the experts identified the following clues for predicting an employee's decision to join the new HMO:

- Age
- Income and value of time to the employee
- Gender
- Computer literacy
- Tendency to join an HMO

Step 4: Describe the Levels of Each Clue

A clue's *level* measures the extent to which it is present. At the simplest, there are two levels—presence or absence—but sometimes there are more. Gender has two levels: male and female. But age may be described in terms of six discrete levels, each corresponding to a decade: younger than 21 years, 21–30 years old, 31–40 years old, 41–50 years old, 51–60 years old, and older than 60 years.

In principle, it is possible to accommodate both discrete and continuous variables in a Bayesian model. *Discrete variables* have a determinable number of variables (e.g., gender has two levels: male and female).

Continuous variables have an infinite or near-infinite number of levels (e.g., age). In practice, discrete clues are used more frequently for two reasons:

1. Experts seem to have more difficulty estimating likelihood ratios associated with continuous clues.
2. In the health and social service areas, most clues tend to be discrete. Virtually all continuous clues can be transformed to discrete clues.

As with defining the forecast events, the primary rule for creating discrete levels is to minimize the number of categories. Rarely are more than five or six categories required, and frequently two or three suffice.

It is preferable to identify levels for various clues by asking the experts to describe a level at which the clue will increase the probability of the forecast events in question. Thus, the analyst may have the following conversation:

Analyst: What would be an example of an age that would favor joining the HMO?
Expert: Young people are more likely to join than older people.

Analyst: How do you define "young" employees?
Expert: Who knows? It all depends. But if you really push me, I would say younger than 30 years is different from older than 30 years. This is probably why the young used to say "never trust anybody over 30." This age marks real differences in life outlook.

Analyst: What age reduces the chance of joining the HMO?
Expert: Employees older than 40 years are different, too; they are pretty much settled in their ways. Of course, you understand we are just talking in general—there are many exceptions.

Analyst: Sure, I understand. But we are trying to model these general trends.

In all cases, each category or division should represent a different chance of joining the HMO. One way to check this is as follows:

Analyst: Do you think a 50-year-old employee is substantially less likely to join than a 40-year-old employee?
Expert: Yes, but the difference is not great.

After much interaction with the experts, the analyst might devise the following levels for each of the clues previously identified:

- Age: younger than 30 years old, 31–40 years old, and older than 41 years old
- Income and value of time to the employee: income over $50,000, income between $30,000 and $50,000, and income less than $30,000
- Gender: male and female
- Computer literacy: programs computers, frequently uses a computer, routinely uses output of a computer, and has no interaction with a computer
- Tendency to join an HMO: is enrolled in an HMO and is not enrolled in an HMO

In describing the levels of each clue, analysts also think through measurement issues. It is convenient to let the analyst decide on the measurement issues, but the measures should be reviewed by the expert to make sure they fit the expert's intentions. For example, income, hence hourly wage, can be used as a surrogate measure for "value of time to employee." But such surrogate measures may be misleading. If income is not a good surrogate for value of time, the analyst has wrecked the model by taking the easy way out. There is a story about a man who lost his keys in the street but was searching for them in his house. When asked why he was looking there, he responded with a certain pinched logic: "The street is dark; the light's better in the house." The lessons are that surrogate measures must be chosen carefully to preserve the value of the clue, and that the easy way is not always the best way.

Step 5: Test for Independence

Conditional independence is an important criterion that can streamline a long list of clues (Pearl 1988, 2000). Independence, as discussed in the previous chapter, means that the presence of one clue does not change the value of any other clue. Conditional independence means that for a specific population, such as employees who will join the HMO, the presence of one clue does not change the value of another. Conditional independence simplifies the forecasting task. The effect of a piece of information on the forecast is its likelihood ratio. Conditional independence allows you to write the likelihood ratio of several clues as a multiplication of the likelihood ratio of each clue. Thus, if C_1 through C_n are the clues in your forecast, the joint likelihood ratio of all the clues can be written as

$$\frac{P(C_1, \ldots, C_n \mid \text{Joining})}{P(C_1, \ldots, C_n \mid \text{Not Joining})} = \frac{P(C_1 \mid \text{Joining})}{P(C_1 \mid \text{Not Joining})} \times \frac{P(C_2 \mid \text{Joining})}{P(C_2 \mid \text{Not Joining})} \times \ldots \times \frac{P(C_n \mid \text{Joining})}{P(C_n \mid \text{Not Joining})}.$$

Assuming conditional independence, the effect of two clues is equal to the product of the effect of each clue. Conditional independence simplifies the number of estimates needed for measuring the joint influence of several pieces of information.

Let's examine whether age and gender are conditionally independent in predicting the probability of joining the HMO. Mathematically, if two clues (e.g., age and gender) are conditionally independent, then you should have

$$\frac{P(\text{Age}\,|\,\text{Gender, Joining})}{P(\text{Age}\,|\,\text{Gender, Not joining})} = \frac{P(\text{Age}\,|\,\text{Joining})}{P(\text{Age}\,|\,\text{Not joining})}.$$

This formula says that the effect of age on the forecast remains the same even when the gender of the person is known. Thus, the influence of age on the forecast does not depend on gender, and vice versa.

The chances for conditional dependence increase with the number of clues, so clues are likely to be conditionally dependent if the model contains more than six or seven clues. When clues are conditionally dependent, either one clue must be dropped from the analysis or the dependent clues must be combined into a new cluster of clues. If age and computer literacy were conditionally dependent, then either could be dropped from the analysis. As an alternative, the analyst could define a new cluster with the following combined levels:

- Younger than 30 years and programs a computer
- Younger than 30 years and frequently uses a computer
- Younger than 30 years and uses computer output
- Younger than 30 years and does not use a computer
- Older than 30 years and programs a computer
- Older than 30 years and frequently uses a computer
- Older than 30 years and uses computer output
- Older than 30 years and does not use a computer

There are statistical procedures for assessing conditional dependence. These procedures were reviewed in Chapter 3 and are further elaborated on in this chapter within the context of modeling experts' opinions. Experts will have in mind different, sometimes wrong, notions of dependence, so the words "conditional independence" should be avoided. Instead, analysts should focus on whether one clue indicates a lot about the influence of another clue. Experts are usually more likely to understand this line of questioning as opposed to understanding direct questions that ask them to verify conditional independence.

Analysts can also assess conditional independence through graphical methods (Pearl 2000). The decision maker is asked to list the possible causes and consequences (e.g., signs, symptoms, characteristics) of the condition being predicted. Only direct causes and consequences are listed, as indirect causes or consequences cannot be modeled through the odds form of Bayes's theorem. A target event is written in a node at the center of the page. All causes precede this node and are shown as arrows leading to the target node. All subsequent signs or symptoms are also shown as nodes that follow the target event node (arrows point from the target node toward the consequences). Figure 4.1 shows three causes and three consequences for the target event.

For example, in predicting who will join an HMO, you might draw the possible causes of people joining the HMO and the characteristics (consequences) that will distinguish people who have joined from those who have not. A node is created in the center and is labeled with the name of the target event. The decision maker might see time pressures and frequent travel as two reasons for joining an online HMO. The decision maker might also see that as a consequence of the HMO being joined by people who suffer time pressures and travel frequently, members of the HMO will be predominantly male, computer literate, and young. This diagram is shown in Figure 4.2.

To understand conditional dependencies implied by a graph, the following rules of dependence are applied:

1. *Connected nodes.* All nodes that are directly connected to each other are dependent on each other. The nodes in Figure 4.2 show how joining the HMO depends on the employee's time pressures, frequency of travel, age, gender, and computer usage. All five clues can be used to predict the probability of joining the HMO.
2. *Common effect.* If two or more causes that are not directly linked to each other have the same effect, then given the effect the causes are conditionally dependent on each other. Knowing that the effect was not influenced by one cause increases the probability of the remaining causes. For example, if employees who have joined the HMO have not done so because they travel often, then it is more likely that they joined because of time pressures.

The following rules apply to conditional independencies:

1. *Common cause.* If one cause has many effects, then given the cause the effects are conditionally independent from each other. This rule applies only if removing the cause would remove the link between the effects and the preceding nodes in the graph.

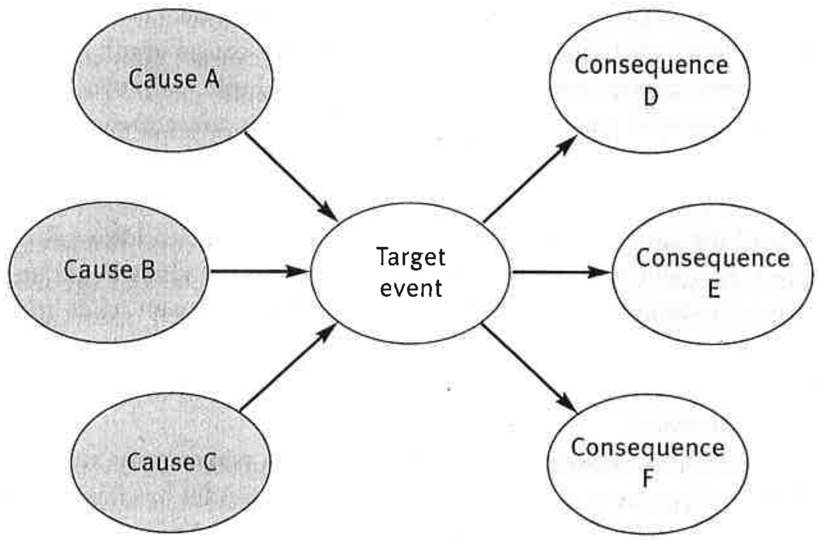

FIGURE 4.1 Causes and Consequences of Target Event Need to Be Drawn

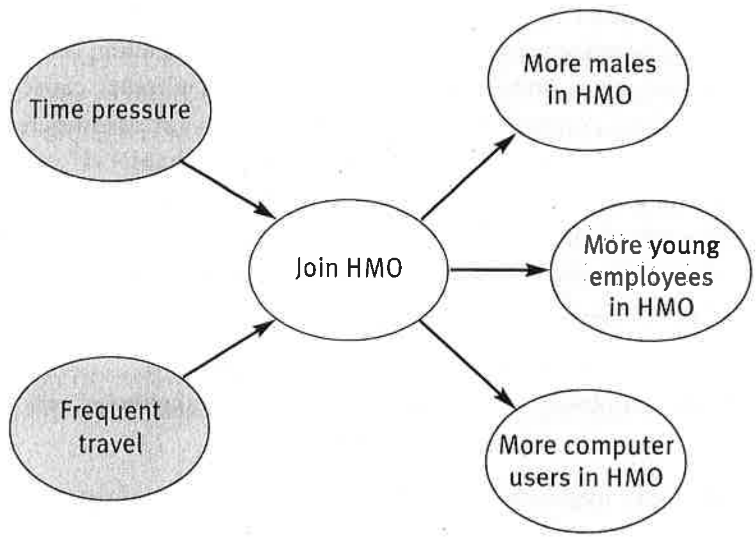

FIGURE 4.2 Causes and Three Signs (Consequences) of Joining the Online HMO

2. *Nodes arranged in a series.* For example, a cause leading to an intermediary event leading to a consequence can signal conditional independence of cause and consequence. For example, given the condition of joining the HMO, time pressure and age of employee are independent of each other because they are in a serial arrangement and removing the middle node (joining the HMO) will remove the connection between the two nodes.

If conditional dependence is found, the analyst has three choices. First, the analyst can ignore the dependencies among the clues. This would work well when the consequences depicted in the causal graph correspond directly with various causes in the graph. For example, employees who put high value on their time may select the HMO; as a consequence, the average salary of employees who join might be higher than those who do not join. In this situation, the cause can be ignored by including the consequence of the cause in the model. As consequences are conditionally independent, this will reduce the dependence among the clues in the model. Ignoring the dependencies in the clues will also work well when multiple clues point to the same conclusion and when the dependencies are small. But when this is not the case, ignoring the dependencies can lead to erroneous predictions.

Second, the analyst could help the decision maker revise the graph or use different causes or consequences so that the model has fewer dependencies. This is often done by better defining the clues. For example, an analyst could reduce the number of links among the nodes by revising the consequence so that it only occurs through the target event. If at all possible, several related causes should be combined into a single cause to reduce conditional dependencies among the clues used in predicting the target event.

Third, all conditionally dependent clues (whether cause or consequence) can be combined to predict the target event. Examples of how to combine two clues were provided earlier in this chapter. If one cause has n levels and another m levels, the new combined cause will have $n \times m$ levels. For each of these levels, separate likelihood ratios must be assessed.

For example, the Bayes's theorem for predicting the odds of joining the HMO can be presented as

$$\text{Posterior odds of joining} = \text{Likelihood ratio}_{\text{time pressure and travel frequency}} \times \text{Likelihood ratio}_{\text{age}} \times \text{Likelihood ratio}_{\text{gender}} \times \text{Likelihood ratio}_{\text{computer use}} \times \text{Prior odds of joining.}$$

Note that the likelihood ratio for the combination of time pressure and travel frequency refers to all possible combinations of levels of these two clues.

Step 6: Estimate Likelihood Ratios

This section explains how to estimate likelihood ratios from experts' subjective opinions (see also Van der Fels-Klerx et al. 2002; Otway and von Winterfeldt 1992). To estimate likelihood ratios, experts should think of the prevalence of the clue in a specific population. The importance of this point is not always appreciated. A likelihood estimate is conditioned on the forecast event, not vice versa. Thus, the effect of being young (younger than 30 years) on the probability of joining the HMO is determined by finding the number of young employees among joiners. There is a crucial distinction between this probability and the probability of joining if one is young. The first statement is conditioned on joining the HMO, the second on being young. The likelihood of individuals younger than 30 years joining is P(Young|Joining), while the probability of joining the HMO for a person younger than 30 years is P(Joining|Young). The two concepts are very different.

A likelihood ratio is estimated by asking questions about the prevalence of the clue in populations with and without the target event. For example, for forecasting the likelihood of joining the HMO, the analyst might ask the following:

> Analyst: Of 100 people who do join, how many are younger than 30 years? Of 100 people who do not join the HMO, how many are younger than 30 years?

The ratio of the answers to these two questions determines the likelihood ratio associated with being younger than 30 years. This ratio could be estimated directly by asking the expert to estimate the odds of finding the clue in populations with and without the target event:

> Analyst: Imagine two employees, one who will join the HMO and one who will not join. Who is more likely to be younger than 30 years? How many times more likely?

The likelihood ratios can be estimated by relying on experts' opinions, but the question naturally arises about whether experts can accurately estimate probabilities. Accurate probability estimation does not mean being correct in every forecast. For example, if you forecast that an employee has a 60 percent chance of joining the proposed HMO but the employee does not join, was the forecast inaccurate? Not necessarily. The accuracy of probability forecasts cannot be assessed by the occurrence of a single event. A better way to check the accuracy of a probability is to compare it against observed frequency counts. A 60 percent chance of joining is accurate if 60 of 100 employees join the proposed HMO. A single case reveals nothing about the accuracy of probability estimates.

Systematic bias may exist in subjective estimates of probabilities. Research shows that subjective probabilities for rare events are inordinately low, while they are inordinately high for common events. These results have led some psychologists to conclude that cognitive limitations of the assessor inevitably distorts subjective probability estimates.

Alemi, Gustafson, and Johnson (1986) argue that the accuracy of subjective estimates can be increased through three steps. First, experts should be allowed to use familiar terminology and decision aids. Distortion of probability estimates can be seen in a group of students, but not among real experts. For example, meteorologists seem to be good probability estimators. Weather forecasters are special because they assess familiar phenomena and have access to a host of relevant and overlapping objective information and judgment aids, such as computers and satellite photos. The point is that experts can reliably estimate likelihood ratios if they are dealing with a familiar concept and have access to their usual tools.

A second way of improving experts' estimates is to train them in selected probability concepts (Dietrich 1991). In particular, experts should learn the meaning of a likelihood ratio. Ratios larger than 1 support the occurrence of the forecast event; ratios less than 1 oppose the probability of the forecast event. A ratio of 1 to 2 reduces the odds of the forecast by half; a ratio of 2 doubles the odds.

The experts should also be taught the relationship between odds and probability. Odds of 2 to 1 mean a probability of 0.67; odds of 5 to 1 mean a probability of 0.83; odds of 10 to1 mean a probability of an almost certain event. The forecaster should walk the expert through and discuss in depth the likelihood ratio for the first clue before proceeding. The first few estimates of the likelihood ratio associated with a clue can each take 20 minutes because many things are discussed and modified. Later estimates often take less than a minute.

A third step for improving experts' estimates of probabilities is to rely on more than one expert (Walker et al. 2003) and on a process of estimation, discussion, and reestimation. Relying on a group of experts increases the chance of identifying major errors. In addition, the process of individual estimation, group discussion, and individual reestimation reduces pressures for artificial consensus while promoting information exchange among the experts.

Step 7: Estimate Prior Odds

According to Bayes's theorem, forecasts require two types of estimates: (1) likelihood ratios associated with specific clues, and (2) prior odds associated with the target event. Prior odds can be assessed by finding the historical prevalence of the event. In a situation without a precedent, prior odds can be estimated by asking experts to imagine the future prevalence of the event:

> Analyst: Out of 100 employees, how many will join the HMO?

The response to this question provides the probability of joining, P(Joining), and this probability can be used to calculate the odds for joining:

$$\text{Odds for joining} = \frac{P(\text{Joining})}{1 - P(\text{Joining})}.$$

When no reasonable prior estimate is available and when a large number of clues is going to be examined, an analyst can arbitrarily assume that the prior odds for joining are 1 to 1 and then allow clues to alter posterior odds as the analyst proceeds with gathering information.

Step 8: Develop Scenarios

Decision makers use scenarios to think about alternative futures. The purpose of forecasting with scenarios is to make the decision maker sensitive to possible futures. The decision maker can work to change the possible futures. Many future predictions are self-fulfilling prophecies—a predicted event happens because steps are taken that increase the chance for it to happen. In this circumstance, predictions are less important than choosing the ideal future and working to make it come about. Scenarios help the decision maker choose a future and make it occur.

Scenarios are written as coherent and internally consistent narrative scripts. The more believable they are, the better. Scenarios are constructed by selecting various combinations of clue levels, writing a script, and adding details to make the group of clues more credible. An optimistic scenario may be constructed by choosing only clue levels that support the occurrence of the forecast event; a pessimistic scenario combines clues that oppose the event's occurrence. Realistic scenarios, on the other hand, are constructed from a mix of clue levels. In the HMO example, scenarios could describe hypothetical employees who would join the organization. A scenario describing an employee who is most likely to join is constructed by assembling all of the characteristics that support joining:

> A 29-year-old male employee earning more than $60,000. He is busy and values his time; he is familiar with computers and uses them both at work and at home. He is currently an HMO member, although he is not completely satisfied with it.

A pessimistic scenario describes the employees least likely to join:

A 55-year-old female employee earning less than $85,000. She has never used computers and has refused to join the firm's existing HMO.

More realistic scenarios combine other clue levels:

A 55-year-old female employee earning more than $60,000 has used computers but did not join the firm's existing HMO.

A large set of scenarios can be made by randomly choosing clue levels and then asking experts to throw out impossible combinations. To do this, first write each clue level on a card and make one pile for each clue. Each pile will contain all the levels of one clue. Randomly select a level from each pile, write it on a piece of paper, and return the card to the pile. Once a level for all clues are represented on the piece of paper, have an expert check the scenario and discard scenarios that are wildly improbable.

If experts are evaluating many scenarios (perhaps 100 or more), arrange the scenario text so they can understand them easily and omit frivolous detail. If experts are reviewing a few scenarios (perhaps 20 or so), add detail and write narratives to enhance the scenarios' credibility.

Because scenarios examine multiple futures, they introduce an element of uncertainty and prepare decision makers for surprises. In the HMO example, the examination of scenarios of possible members would help the decision makers understand that large segments of the population may not consider the HMO desirable. This leads to two possible changes. First, a committee could be assigned to examine the unmet needs of people unlikely to join and to make the proposal more attractive to segments not currently attracted to it. Second, another committee could examine how the proposed HMO could serve a small group of customers and still succeed.

Sometimes forecasting is complete after the decision maker has examined the scenarios. In these circumstances, making the decision maker intuitively aware of what might happen suffices. In other circumstances, decision makers may want a numerical forecast. To get to a numerical prediction, the analyst must take two more steps.

Step 9: Validate the Model

In the final analysis, any subjective probability model is just a set of opinions, and as such it should not be trusted until it passes vigorous evaluation. The evaluation of a subjective model requires answers to two related questions: Does the model reflect the experts' views? Are the experts' views accurate?

To answer the first question, design about 30 to 100 scenarios, ask the experts to rate each, and compare these ratings to model predictions. If the ratings and predictions match closely (a correlation above 0.7), then the model simulates the experts' judgments. For example, the analyst can generate 30 hypothetical employees and ask the experts to rate the probability that each will join the proposed HMO. To help the experts accomplish this, you should ask them to arrange the cases from more to less likely, to review pairs of adjacent employees to determine if the rank order is reasonable, and to change the rank orders of the employees if needed. Once the cases have been arranged in order, the experts would be asked to rate the chance of joining on a scale of 0 to 100. For each scenario, the analyst should use the Bayes's theorem to forecast whether the employee will join the HMO. Table 4.1 also shows the resulting ratings and predictions.

Next, compare the Bayes's theorem result to the average of the experts' ranking. If the rank-order correlation is higher than 0.70, the analyst may conclude that the model simulates many aspects of the experts' intuitions. Figure 4.3 shows the relationship between the model's predictions and the average experts' ratings.

The straight line of the graph shows the expected relationship. Some differences between the model's predictions and the experts' ratings should be expected, as the experts will show many idiosyncrasies and inconsistencies not found in the model. But the model's predictions and the experts' intuitions should not sharply diverge. One way to examine this is through correlations. If the correlation were lower than 0.7, then the experts' intuitions might not have been effectively modeled, in which case the model must be modified because the likelihood ratios might be too high or some important clues might have been omitted. In this example, the model's predictions and the average of the experts' ratings had a high correlation, which leads to the conclusion that the model simulate the experts' judgments.

The above procedure leaves unanswered the larger and perhaps more difficult question of the accuracy of the experts' intuitions. Experts' opinions can be validated if they can be compared to observed frequencies, but this is seldom possible (Howard 1980). In fact, if the analyst has access to observed frequencies, there is no need to consult experts to create subjective probability models. In the absence of objective data, what steps can an analyst take to reassure herself regarding her experts?

One way to increase confidence in experts' opinions is to use several experts. If the experts reach a consensus, then one should feel comfortable with a model that predicts that consensus. *Consensus* means that, after discussing the problem, experts independently rate the hypothetical

TABLE 4.1 Two Experts' Ratings and Bayes's Theorem on 30 Hypothetical Scenarios

Scenario Number	Experts' Ratings		Average Rating	Model Predictions
	Expert 1	Expert 2		
1	41	45	43.0	48
2	31	32	31.5	25
3	86	59	72.5	87
4	22	35	28.5	21
5	61	93	77.0	80
6	38	60	49.0	58
7	38	100	69.0	46
8	14	85	49.5	29
9	30	27	28.5	30
10	33	71	52.0	32
11	45	97	71.0	49
12	22	11	16.5	29
13	39	65	52.0	48
14	28	38	33.0	21
15	28	71	49.5	23
16	33	74	53.5	53
17	46	64	55.0	31
18	75	67	71.0	77
19	61	43	52.0	59
20	73	83	78.0	97
21	0	44	22.0	14
22	16	77	46.5	20
23	37	92	64.5	44
24	15	66	40.5	19
25	43	62	52.5	28
26	16	67	41.5	25
27	48	51	49.5	14
28	100	73	86.5	100
29	15	52	33.5	0
30	6	0	3.0	16

scenarios close to one another. One way of checking the degree of agreement among experts' ratings of the scenarios is to correlate the ratings of each pair of experts. Correlation values above 0.70 suggest excellent agreement; values between 0.50 and 0.70 suggest more moderate agreement. If the correlations are below 0.50, then experts differed, and it is best to examine their differences and redefine the forecast. In the previous example, the two experts had a high correlation, which suggests that they did not agree on the ratings of the scenarios.

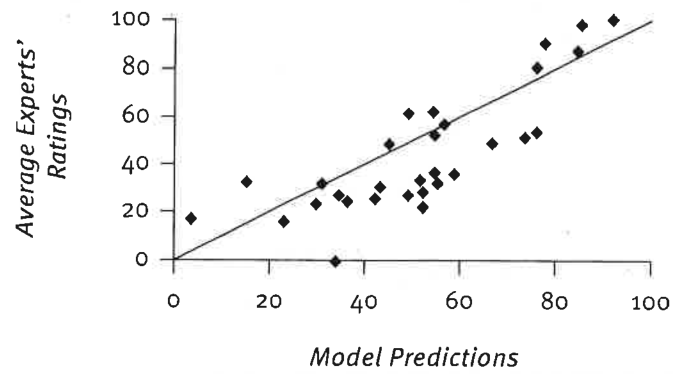

FIGURE 4.3 Validating a Model by Testing if it Simulates Experts' Judgments

Some investigators believe a model, even if it predicts the consensus of the best experts, is still not valid because only objective data can validate a model. According to this rationale, a model provides no reason to act unless it is backed by objective data. Although it is true that no model can be fully validated until its results can be compared to objective data, expert opinions are sufficient grounds for action in many circumstances. In some circumstances (e.g., surgery), people trust their lives to experts' opinions. If one is willing to trust one's life to expertise, one should be willing to accept expert opinion as a basis for business and policy action.

Step 10: Make a Forecast

To make a forecast, an analyst should begin by describing the characteristics of the situation at hand. The likelihood ratios corresponding to the situation at hand are used to make the forecast. In the HMO example, the likelihood ratios associated with characteristics of the employee are used to calculate the probability that an employee will join the HMO. Suppose you evaluated a 29-year-old man earning $60,000 who is computer literate but is not an HMO member. Suppose the likelihood ratios associated with these characteristics are 1.2 for being young, 1.1 for being male, 1.2 for having a high hourly wage, 3.0 for being computer literate, and 0.5 for not being a member of an HMO. Assuming prior odds were equal and the characteristics are conditionally independent, this employee's posterior odds of joining are calculated as follows:

$$\text{Odds of joining} = 1.1 \times 1.2 \times 3 \times 0.5 \times 1 = 1.98.$$

The probability of a mutually exclusive and exhaustive event A can be calculated from its odds using the following formula:

$$P(A) = \frac{\text{Odds}(A)}{1 + \text{Odds}(A)}$$

The above prediction then becomes

$$P(\text{Joining}) = \frac{1.98}{1 + 1.98} = 0.66.$$

The probability of joining can be used to estimate the number of employees likely to join the new HMO (in other words, the demand for the proposed product). If the analyst expects to have 50 of this type of employee, the analyst can forecast that 33 employees (50×0.66) will join. If one does similar calculations for other types of employees, one can calculate the total demand for the proposed HMO.

Analysis of demand for the proposed HMO shows that most employees would not join, but that 12 percent of the employed population might join. Careful analysis can allow the analyst to identify a small group of employees who could be expected to support the proposed HMO, showing that a niche is available for the innovative plan.

Summary

Forecasts of unique events are useful, but they are difficult because of the lack of data. Even when events are not unique, frequency counts are often unavailable, given time and budget constraints. However, the judgments of people with substantial expertise can serve as the basis of forecasts.

In predictions where many clues are needed for forecasting, experts may not function at their best; and as the number of clues increases, the task of forecasting becomes increasingly arduous. Bayes's theorem is a mathematical formula that can be used to aggregate the effect of various clues. This approach combines the strength of human expertise (i.e., estimating the relationship between the clue and the forecast) with the consistency of a mathematical model. Validating these models poses a problem because no objective standards are available, but once the model has passed scrutiny from several experts from different backgrounds, one can feel sufficiently confident about the model to recommend action based on its forecasts.

Review What You Know

Suppose you want to build a model that can predict a patient's breast cancer risk. You can interview an expert and thereafter use the model of the experts' judgment in evaluating patients' overall risk for breast cancer. Using the advice provided here, describe the process of getting the expert to identify clues (i.e., risk factors).

What will you ask an expert if you wish to estimate the likelihood ratio associated with the clue, "age < 9 years at menarche" in predicting a patient's breast cancer risk?

What will you ask from a primary care provider to estimate the prior odds for breast cancer among her patients?

In this chapter, it is suggested that instead of predicting probability of future events, it is sometimes sufficient to generate a number of future scenarios and have the decision maker evaluate these scenarios. Speculate under what circumstances you would stop at scenario generation and not proceed to a numerical estimation. Give an example of a situation in which only reviewing scenarios will be sufficient.

Rapid-Analysis Exercises

Construct a probability model to forecast an important event at work. Select an expert who will help you construct the model. Make an appointment with the expert and construct the model. Prepare a report that answers the following questions:

1. What are the assumptions about the problem and its causes?
2. What is the uncertain outcome?
3. Why would a model be useful and to whom would it be useful?
4. Conduct research to report if similar models or studies have been done by others.
5. What clues and clue levels can be used to make the prediction?
6. What are the likelihood ratios associated with each clue level?
7. Show the application of the model in a case or scenario.
8. What is the evidence that the model is valid?
   a. Is the model based on available data?
   b. Did the expert consider the model simple to use?
   c. Did the expert consider the model to be face valid?
   d. Does the model simulate the experts' judgment on at least 15 cases or scenarios?

   e. Graph the relationship between the model scores and the experts' ratings.
   f. Report the correlation between the model score and the experts' ratings.
9. If data are available, report whether the model corresponds with other measures of the same concept (i.e., construct validity)?
10. If data are available, report whether the model predicts any objective gold standard.

Audio/Visual Chapter Aids

To help you understand the concepts of modeling uncertainty, visit this book's companion web site at ache.org/DecisionAnalysis, go to Chapter 4, and view the audio/visual chapter aids.

References

Alemi, F., D. H. Gustafson, and M. Johnson. 1986. "How to Construct a Subjective Index." *Evaluation and the Health Professions* 9 (1): 45–52.

Dietrich, C. F. 1991. *Uncertainty, Calibration and Probability: The Statistics of Scientific and Industrial Measurement.* 2nd ed. Philadelphia, PA: Thomas and Francis.

Howard, R. A. 1980. "An Assessment of Decision Analysis." *Operations Research* 28 (1): 4–27.

Otway, H., and D. von Winterfeldt. 1992. "Expert Judgment in Risk Analysis and Management: Process, Context, and Pitfalls." *Risk Analysis* 12 (1): 83–93.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* San Francisco: Morgan Kaufmann.

———. 2000. *Causality: Models, Reasoning, and Inference.* Cambridge, England: Cambridge University Press.

Rivers, S. E., P. Salovey, D. A. Pizarro, J. Pizarro, and T. R. Schneider. 2005. "Message Framing and Pap Test Utilization Among Women Attending a Community Health Clinic." *Journal of Health Psychology* 10 (1): 65–77.

Rothman, A. J., and P. Salovey. 1997. "Shaping Perceptions to Motivate Healthy Behavior: The Role of Message Framing." *Psychological Bulletin* 121 (1): 3–19.

Van der Fels-Klerx, I. H., L. H. Goossens, H. W. Saatkamp, and S. H. Horst. 2002. "Elicitation of Quantitative Data from a Heterogeneous Expert Panel: Formal Process and Application in Animal Health." *Risk Analysis* 22 (1): 67–81.

Walker, K. D., P. Catalano, J. K. Hammitt, and J. S. Evans. 2003. "Use of Expert Judgment in Exposure Assessment: Part 2. Calibration of Expert Judgments about Personal Exposures to Benzene." *Journal of Exposure Analysis and Environmental Epidemiology* 13 (1): 1–16.