CHAPTER

3

MEASURING UNCERTAINTY

Farrokh Alemi

This chapter describes how probability can quantify the degree of uncertainty one feels about future events. It answers the following questions:

- What is probability?
- What is the difference between objective and subjective sources of data for probabilities?
- What is Bayes's theorem?
- What are independence and conditional independence?
- How does one verify independence?

Measuring uncertainty is important because it allows one to make trade-offs among uncertain events, and to act in uncertain environments. Decision makers may not be sure about a business outcome, but if they know the chances are good, they may risk it and reap the benefits.

Probability

When it is certain that an event will occur, it has a probability of 1. When it is certain that an event will not occur, it has a probability of 0. When there is uncertainty that an event will occur, it has a probability of 0.5—or, a 50/50 chance of occurrence. All other values between 0 and 1 measure the uncertainty about the occurrence of an event.

The best way to think of probability is as the ratio of all ways an event may occur divided by all possible outcomes. In short, *probability* is the prevalence of the target event among the possible events. For example, the probability of a small business failing is the number of small businesses that fail divided by the total number of small businesses. Or, the probability of an iatrogenic infection in the last month in a hospital is the number of patients who last month had an iatrogenic infection in the hospital divided by the number of patients in the hospital during last month. The basic probability formula is

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 3.

$$P(A) = \frac{\text{Number of occurences of event } A}{\text{Total number of possible events}}.$$

Figure 3.1 shows a visual representation of probability. The rectangle represents the number of possible events, and the circle represents all ways in which event A might occur; the ratio of the circle to the rectangle is the probability of A.

Probability of Multiple Events

The rules of probability allow you to calculate the probability of multiple events. For example, the probability of either A or B occurring is calculated by first summing all the possible ways in which event A will occur and all the ways in which event B will occur, minus all the possible ways in which both event A and B will occur together (this is subtracted to avoid double counting). This sum is divided by all possible outcomes. This concept is shown in the Venn diagram in Figure 3.2. This concept is represented in mathematical terms as

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B).$$

The definition of probability gives you a simple calculus for combining the uncertainty of two events. You can now ask questions such as "What is the probability that frail elderly (age > 75 years old) or infant patients will join our HMO?" According to the previous formula, this can be calculated as

$$P(\text{Frail elderly or Infant}) = \\ P(\text{Frail elderly}) + P(\text{Infant}) - P(\text{Frail elderly and Infant}).$$

Because the chance of being both a frail elderly person and an infant is 0 (i.e., the two events are mutually exclusive), the formula can be rewritten as

$$P(\text{Frail elderly or Infant}) = P(\text{Frail elderly}) + P(\text{Infant}).$$

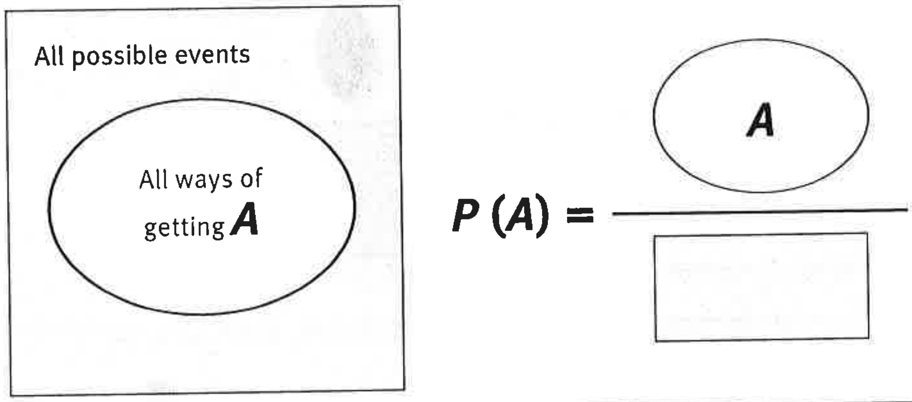


FIGURE 3.1
A Visual Representation of Probability

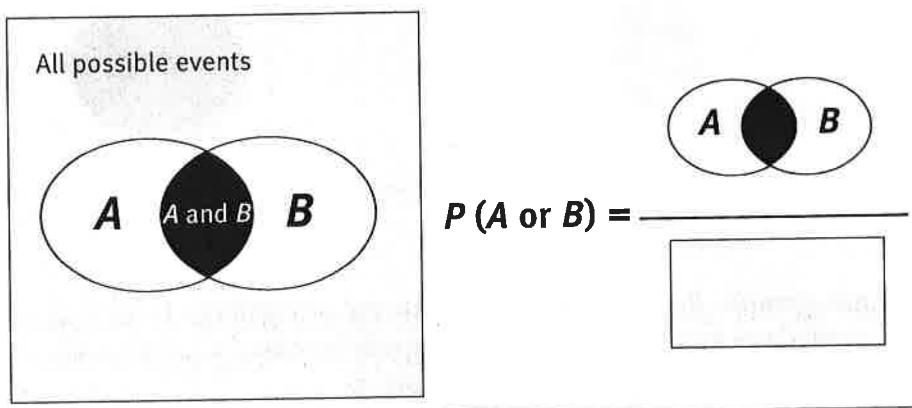


FIGURE 3.2
Visual Representation of Probability *A* or *B*

This definition of probability can also be used to measure the probability of two events co-occurring (probability of event *A* and event *B*). Note that the overlap between *A* and *B* is shaded in Figure 3.2; this area represents all the ways *A* and *B* might occur together. Figure 3.3 shows how the probability of *A* and *B* occurring together is calculated by dividing this shaded area by all possible outcomes.

Conditional, Joint, and Marginal Probabilities

The definition of probability also helps in the calculation of the probability of an event conditioned on the occurrence of other events. In mathematical terms, *conditional probability* is shown as $P(A|B)$ and read as probability of *A* given *B*. When an event occurs, the remaining list of possible outcomes is reduced. There is no longer the need to track events that

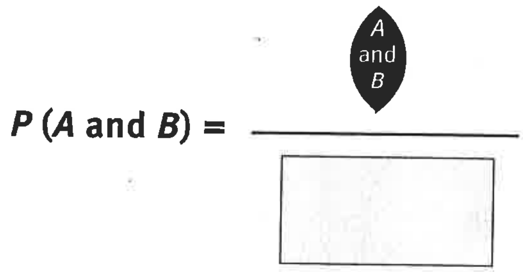


FIGURE 3.3 A Visual Representation of Joint Probability *A* or *B*

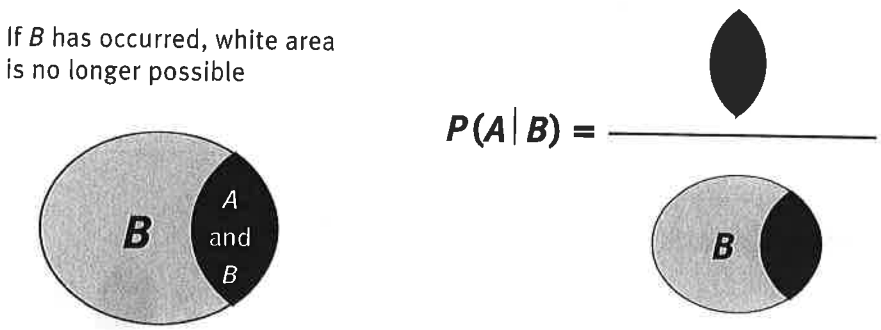


FIGURE 3.4 Probability of *A* Given *B* Is Calculated by Reducing the Possibilities

are not possible. You can calculate conditional probabilities by restricting the possibilities to only those events that you know have occurred, as shown in Figure 3.4. This is shown mathematically as

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

For example, you can now calculate the probability that a frail elderly patient who has already joined the HMO will be hospitalized. Instead of looking at the hospitalization rate among all frail elderly patients, you need to restrict the possibilities to only the frail elderly patients who have joined the HMO. Then, the probability is calculated as the ratio of the number of hospitalizations among frail elderly patients in the HMO to the number of frail elderly patients in the HMO:

$$P(\text{Hospitalized}|\text{Joined HMO}) = \frac{P(\text{Hospitalized and Joined HMO})}{P(\text{Joined HMO})}.$$

Analysts need to make sure that decision makers distinguish between joint probability, or the probability of A and B occurring together, and conditional probability, or the probability of B occurring after A has occurred. Joint probabilities, shown as $P(A \text{ and } B)$, are symmetrical and not time based. In contrast, conditional probabilities, shown as $P(A|B)$, are asymmetrical and do rely on the passage of time. For example, the probability of a frail elderly person being hospitalized is different from the probability of finding a frail elderly person among people who have been hospitalized.

For an example calculation of conditional probabilities from joint probabilities, assume that an analysis has produced the joint probabilities in Table 3.1 for the patient being either in substance abuse treatment or in probation.

Table 3.1 provides joint and marginal probabilities by dividing the observed frequency of days by the total number of days examined. *Marginal probability* refers to the probability of one event; in Table 3.1, these are provided in the row and column labeled "Total." For example, the marginal probability of a probation day, regardless of whether it is also a treatment day, is 0.56. *Joint probability* refers to the probability of two events occurring at same time; in Table 3.1, these are provided in the remaining rows and columns. For example, the joint probability of having both a probation day and a treatment day is 0.51. This probability is calculated by dividing the number of days in which both probation and treatment occur by the total number of days examined.

If an analyst wishes to calculate a conditional probability, the total universe of possible days must be reduced to the days that meet the condition. This is a very important concept to keep in mind:

> Conditional probability is a reduction in the universe of possibilities.

Suppose the analyst wants to calculate the conditional probability of being in treatment given that the patient is already in probation. In this case, the universe is reduced to all days in which the patient has been in probation. In this reduced universe, the total number of days of treatment

TABLE 3.1 Joint Probability of Treatment and Probation

	Probation Day	*Not a Probation Day*	*Total*
Treatment Day	0.51	0.39	0.90
Not a Treatment Day	0.05	0.05	0.10
Total	0.56	0.44	1.00

becomes the number of days of having both treatment and probation. Therefore, the conditional probability of treatment given probation is

$$P(\text{Treatment}|\text{Probation}) = \frac{\text{Number of days in both treatment and probation}}{\text{Number of days in probation}}.$$

Because Table 3.1 provides the joint and marginal probabilities, the previous formula can be described in terms of joint and marginal probabilities:

$$P(\text{Treatment}|\text{Probation}) = \frac{P(\text{Treatment and Probation})}{P(\text{Probation})} = \frac{0.51}{0.56} = 0.93.$$

The point of this example is that conditional probabilities can be easily calculated by reducing the universe of possibilities to only those situations that meet the condition. You can calculate conditional probabilities from marginal and joint probabilities by keeping in mind how the condition has reduced the universe of possibility.

Conditional probabilities are a very useful concept. They allow you to think through an uncertain sequence of events. If each event can be conditioned on its predecessor, a chain of events can be examined. Then, if one component of the chain changes, you can calculate the effect of the change throughout the chain. In this sense, conditional probabilities show how a series of clues might forecast a future event. For example, in predicting who will join the HMO, the patient's demographics (age, gender, income level) can be used to infer the probability of joining. In this case, the probability of joining the HMO is the target event. The clues are the patient's age, gender, and income level. The objective is to predict the probability of joining the HMO given the patient's demographics—in other words, $P(\text{Join HMO}|\text{Age, gender, income level})$.

The calculus of probability is an easy way to track the overall uncertainty of several events. The calculus is appropriate if the following simple assumptions are met:

1. The probability of an event is a positive number between 0 and 1.
2. One event certainly will happen, so the sum of the probabilities of all events is 1.
3. The probability of any two mutually exclusive events occurring equals the sum of the probability of each occurring.

Most decision makers are willing to accept these three assumptions, often referred to by mathematicians as *probability axioms.*

If a set of numbers assigned to uncertain events meet these three principles, then it is a probability function and the numbers assigned in this fashion must follow the algebra of probabilities.

Sources of Data

There are two ways to measure the probability of an event:

1. One can observe the objective frequency of the event. For example, you can see how many out of 100 people who were approached about joining an HMO expressed intent to do so.
2. The alternative is to rely on subjective opinions of an expert. In these circumstances, ask an expert to estimate the strength of her belief that the event of interest might happen. For example, you might ask a venture capitalist who is familiar with new businesses the following question: On a scale from 0 to 100, where 100 is for sure, how strongly do you feel that the average employee will join an HMO?

Both approaches measure the degree of uncertainty about the success of the HMO, but there is a major difference between them: One approach is objective while the other is based on opinion. Objective frequencies are based on observations of the history of the event, while a measurement of strength of belief is based on an individual's opinion, even about events that have no history (e.g., What is the chance that there will be a terrorist attack in our hospital?).

Subjective Probability

More than half a century ago, Savage (1954) and de Finetti (1937) argued that the rules of probabilities can work with uncertainties expressed as strength of opinion. Savage termed the strength of a decision maker's convictions "subjective probability" and used the calculus of probability to analyze these convictions. Subjective probability remains a popular method for analyzing experts' judgments and opinions (Jeffrey 2004). Reviews of the field show that under certain circumstances, experts and nonexperts can reliably assess subjective probabilities that correspond to objective reality (Wallsten and Budescu 1983). Subjective probability can be measured along two different concepts: (1) intensity of feelings and (2) hypothetical frequency. Subjective probability based on intensity of feelings can be measured by asking the experts to rate their certainty on a scale of 0 percent to 100 percent. Subjective probability based on hypothetical frequency can be measured by asking the expert to estimate how many times the target event will occur out of 100 possible situations.

Suppose an analyst wants to measure the probability that an employee will join the HMO. Using the first method, an analyst would ask an expert on the local healthcare market about the intensity of his feelings:

> Analyst: Do you think employees will join the plan? On a scale from 0 to 100, with 100 being certain, how strongly do you feel you are right?

When measuring according to hypothetical frequencies, the expert would be asked to imagine what she expects the frequency would be, even though the event has not occurred repeatedly:

> Analyst: Out of 100 employees, how many do you think will join the plan?

Subjective Probability as a Probability Function

If both the subjective and the objective methods produce a probability for the event, then the calculus of probabilities can be used to make new inferences from these data. It makes no difference whether the frequency is objectively observed through historical precedents or subjectively described by an expert; the resulting number should follow the rules of probability.

Even though subjective probabilities measured as intensity of feelings are not actually probability functions, they should be treated as such. Returning to the formal definition of a probability measure, a probability function is defined by the following characteristics:

1. The probability of an event is a positive number between 0 and 1.
2. One event certainly will happen, so the sum of the probabilities of all events is 1.
3. The probability of any two mutually exclusive events occurring equals the sum of the probability of each occurring.

These assumptions are at the root of all mathematical work in probability, so any beliefs expressed as probability must follow them. Furthermore, if these three assumptions are met, then the numbers produced in this fashion will follow all rules of probabilities. Are these three assumptions met when the data are subjective? The first assumption is always true, because you can assign numbers to beliefs so they are always positive.

But the second and third assumptions are not always true, and people do hold beliefs that violate them. However, analysts can take steps to ensure that these two assumptions are also met. For example, when the estimates of all possibilities (e.g., probability of success and failure) do not total 1, the analyst can revise the estimates to do so. When the estimated probabilities of two mutually exclusive events do not equal the sum of their separate probabilities, the analyst can ask whether they should and adjust them as necessary.

Decision makers, left to their own devices, may not follow the calculus of probability. Experts' opinions also may not follow the rules of probability, but if experts agree with the aforementioned three principles, then such opinions should follow the rules of probability.

Probabilities and beliefs are not identical constructs; rather, probabilities provide a context in which beliefs can be studied. That is, if beliefs are expressed as probabilities, then the rules of probability provide a systematic and orderly method of examining the implications of these beliefs.

Bayes's Theorem

From the definition of conditional probability, one can derive the *Bayes's theorem*, an optimal model for revising existing opinion (sometimes called prior opinion) in light of new evidence or clues. The theorem states

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

where

- $P(\ )$ designates the probability of the event within the parentheses;
- H marks a target event or hypothesis occurring;
- N designates the same event not occurring;
- $C_1, \ldots, C_n$ mark the clues 1 through n;
- $P(H|C_1, \ldots, C_n)$ is the probability of hypothesis H occurring given clues 1 through n;
- $P(N|C_1, \ldots, C_n)$ is the probability of hypothesis H not occurring given clues 1 through n;
- $P(C_1, \ldots, C_n|H)$ is the prevalence of the clues among the situations where hypothesis H has occurred and is referred to as the likelihood of the various clues given H has occurred; and
- $P(C_1, \ldots, C_n|N)$ is the prevalence of the clues among situation where hypothesis H has not occurred. This term is also referred to as the likelihood of the various clues given H has not occurred.

In other words, Bayes's theorem states that

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

The difference between the left and right terms is the knowledge of clues. Thus, the theorem shows how opinions should change after examining clues 1 through n. 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.

Bayes's theorem claims that prior odds of an event are multiplied by the likelihood ratio associated with various clues to obtain the posterior odds for the event. At first glance, it might seem strange to multiply rather than add. You might question why other probabilities besides prior odds and likelihood ratios are not included. The following section makes the logical case for Bayes's theorem.

Rationale for Bayes's Theorem

Bayes's theorem sets a norm for decision makers regarding how they should revise their opinions. But who says this norm is reasonable? In this section, Bayes's theorem is shown to be logical and based on simple assumptions that most people agree with. Therefore, to remain logically consistent, everyone should accept Bayes's theorem as a norm.

Bayes's theorem was first proven mathematically by Thomas Bayes, an English mathematician, although he never submitted his paper for publication. Using Bayes's notes, Price presented a proof of Bayes's theorem (Bayes 1963). The following presentation of Bayes's argument differs from the original and is based on the work of de Finetti (1937). Suppose you want to predict the probability of joining an HMO based on whether the individual is frail elderly. You could establish four groups:

1. A group of size a joins the HMO and is frail elderly.
2. A group of size b joins the HMO and is not frail elderly.
3. A group of size c does not join the HMO and is frail elderly.
4. A group of size d does not join the HMO and is not frail elderly.

Suppose the HMO is offered to $a + b + c + d$ Medicare beneficiaries (see Table 3.2). The probability of an event is defined as the number of ways the event occurs divided by the total possibilities. Thus, since the total number of beneficiaries is $a + b + c + d$, the probability of any of them joining the HMO is the number of people who join divided by the total number of beneficiaries:

$$P(\text{Joining}) = \frac{a + b}{a + b + c + d}.$$

TABLE 3.2 Partitioning Groups Among Frail Elderly Who Will Join the HMO

	Frail Elderly	*Not Frail Elderly*	*Total*
Joins the HMO	a	b	$a + b$
Does not join the HMO	c	d	$c + d$
Total	$a + c$	$b + d$	$a + b + c + d$

Similarly, the chance of finding a frail elderly, P (Frail elderly), is the total number of frail elderly, $a + c$, divided by the total number of beneficiaries:

$$P(\text{Frail elderly}) = \frac{a + c}{a + b + c + d} .$$

Now consider a special situation in which one focuses only on those beneficiaries who are frail elderly. Given that the focus is on this subset, the total number of possibilities is now reduced from the total number of beneficiaries to the number who are frail elderly (i.e., $a + c$). If you focus only on the frail elderly, the probability of one of these beneficiaries joining is

$$P(\text{Joining} \mid \text{Frail elderly}) = \frac{a}{a + c} .$$

Similarly, the likelihood that you will find frail elderly among joiners is given by reducing the total possibilities to only those beneficiaries who join the HMO and then by counting how many were frail elderly:

$$P(\text{Frail elderly} \mid \text{Joining}) = \frac{a}{a + c} .$$

From the above four formulas, you can see that

$$P(\text{Joining} \mid \text{Frail elderly}) = P(\text{Frail elderly} \mid \text{Joining}) \times \frac{P(\text{Joining})}{P(\text{Frail elderly})} .$$

Repeating the procedure for not joining the HMO, you find that

$$P(\text{Not joining} \mid \text{Frail elderly}) = P(\text{Frail elderly} \mid \text{Not joining}) \times \frac{P(\text{Not joining})}{P(\text{Frail elderly})} .$$

Dividing the above two equations results in the odds form of the Bayes's theorem:

$$\frac{P(\text{Joining} \mid \text{Frail elderly})}{P(\text{Not joining} \mid \text{Frail elderly})} = \frac{P(\text{Joining} \mid \text{Frail elderly})}{P(\text{Frail elderly} \mid \text{Not joining})} \times \frac{P(\text{Joining})}{P(\text{Not joining})} .$$

As the above has shown, the Bayes's theorem follows from very reasonable, simple assumptions. If beneficiaries are partitioned into the four groups, the numbers in each group are counted, and the probability of an event is defined as the count of the event divided by number of possibilities, then Bayes's theorem follows. Most readers will agree that these assumptions are reasonable and therefore that the implication of these assumptions (i.e., the Bayes's theorem) should also be reasonable.

Independence

In probabilities, the concept of independence has a very specific meaning. If two events are independent of each other, then the occurrence of one event does not reveal much about the occurrence of the other event. Mathematically, this condition can be presented as

$$P(A \mid B) = P(A).$$

This formula says that the probability of A occurring does not change given that B has occurred.

Independence means that the presence of one clue does not change the impact of another clue. An example might be the prevalence of diabetes and car accidents; knowing the probability of car accidents in a population will not reveal anything about the probability of diabetes.

When two events are independent, you can calculate the probability of both occurring from the marginal probabilities of each event occurring:

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

Thus, you can calculate the probability of a person with diabetes having a car accident as the product of the probability of having diabetes and the probability of having a car accident.

Conditional Independence

A related concept is conditional independence. *Conditional independence* means that, for a specific population, the presence of one clue does not change the probability of another. Mathematically, this is shown as

$$P(A \mid B, C) = P(A \mid C).$$

The above formula reads that if you know that C has occurred, telling you that B has occurred does not add any new information to the estimate of probability of A. Another way of saying this is to say that in population C, knowing B does not reveal much about the chance for A. Conditional independence also allows you to calculate joint probabilities from marginal probabilities:

$$P(A \text{ and } B|C) = P(A|C) \times P(B|C).$$

The above formula states that among the population C, the probability of both A and B occurring together is equal to the product of probability of each event occurring.

It is possible for two events to be dependent, but they may become independent of each other when conditioned on the occurrence of a third event. For example, you may think that scheduling long shifts will lead to medication errors. This can be shown as follows (≠ means "not equal to"):

$$P(\text{Medication error}) \neq P(\text{Medication error}|\text{Long shift}).$$

At the same time, you may consider that in the population of employees that are not fatigued (even though they have long shifts), the two events are independent of each other:

$$P(\text{Medication error}|\text{Long shift, Not fatigued}) = P(\text{Medication error}|\text{Not fatigued}).$$

In English, this formula says that if the nurse is not fatigued, then it does not matter if the shift is long or short; the probability of medication error does not change. This example shows that related events may become independent under certain conditions.

Use of Independence

Independence and conditional independence are often invoked to simplify the calculation of complex likelihoods involving multiple events. It has already been shown how independence facilitates the calculation of joint probabilities. The advantage of verifying independence becomes even more pronounced when examining more than two events. Recall that the use of the odds form of Bayes's theorem requires the estimation of the likelihood ratio. When multiple events are considered before revising the prior odds, the estimation of the likelihood ratio involves conditioning future events on all prior events (Eisenstein and Alemi 1994):

$$P(C_1, C_2, C_3, \ldots, C_n|H_1) = P(C_1|H_1) \times P(C_2|H_1, C_1) \times P(C_3|H_1, C_1, C_2) \times P(C_4|H_1, C_1, C_2, C_3) \times \ldots \times P(C_n|H_1, C_1, C_2, C_3, \ldots, C_{n-1}).$$

Note that each term in the above formula is conditioned on the hypothesis, or on previous events. When events are considered, the posterior odds are modified and are used to condition all subsequent events. The first term is conditioned on no additional event; the second term is conditioned on the first event; the third term is conditioned on the first and second events, and so on until the last term that is conditioned on all subsequent $n - 1$ events. Keeping in mind that conditioning is reducing the sample size to the portion of the sample that has the condition, the above formula suggests a sequence for reducing the sample size. Because there are many events, the data has to be portioned in increasingly smaller sizes. In order for data to be partitioned so many times, a large database is needed.

Conditional independence allows you to calculate likelihood ratios associated with a series of events without the need for large databases. Instead of conditioning the event on the hypothesis and all prior events, you can now ignore all prior events:

$$P(C_1, C_2, C_3, \ldots, C_n | H_1) = P(C_1|H) \times P(C_2|H) \times P(C_3|H) \times P(C_4|H) \times \ldots \times P(C_n|H).$$

Conditional independence simplifies the calculation of the likelihood ratios. Now the odds form of Bayes's theorem can be rewritten in terms of the likelihood ratio associated with each event:

$$\frac{P(H|C_1, \ldots, C_n)}{P(N|C_1, \ldots, C_n)} = \frac{P(C_1|H)}{P(C_1|N)} \times \frac{P(C_2|H)}{P(C_2|N)} \times \ldots \times \frac{P(C_n|H)}{P(C_n|N)} \times \frac{P(H)}{P(N)}.$$

In other words, the above formula states

Posterior odds = Likelihood ratio of first clue × Likelihood ratio of second clue × . . . × Likelihood ratio of *n*th clue × Prior odds.

The odds form of Bayes's theorem has many applications. It is often used to estimate how various clues (events) may help revise prior probability of a target event. For example, you might use the above formula to predict the posterior odds of hospitalization for a frail elderly female patient if you accept that age and gender are conditionally independent of each other. Suppose the likelihood ratio associated with being frail elderly is $5/2$, meaning that knowing the patient is frail elderly will increase the odds of hospitalization by 2.5 times. Also suppose that knowing the patient is female reduces the odds for hospitalization by $9/10$. Now, if the prior odds for hospitalization is $1/2$, the posterior odds for hospitalization can be calculated using the following formula:

Posterior odds of hospitalization = Likelihood ratio associated with being frail elderly × Likelihood ratio associated with being female × Prior odds of hospitalization.

The posterior odds of hospitalization can now be calculated as

$$\text{Posterior odds of hospitalization} = \frac{5}{2} \times \frac{9}{10} \times \frac{1}{2} = 1.125.$$

For mutually exclusive and exhaustive events, the odds for an event can be restated as the probability of the event by using the following formula:

$$P = \frac{\text{Odds}}{1 + \text{Odds}}.$$

Using the above formula, you can calculate the probability of hospitalization:

$$P(\text{Hospitalization}) = \frac{1.125}{1 + 1.125} = 0.53.$$

Verifying Conditional Independence

There are several ways to verify conditional independence. These include (1) reducing sample size, (2) analyzing correlations, (3) asking experts, and (4) separating in causal maps.

Reducing Sample Size

If data exist, conditional independence can be verified by selecting the population that has the condition and verifying that the product of marginal probabilities is equal to the joint probability of the two events. For example, Table 3.3 presents 18 cases from a special unit prone to medication errors. The question is whether rate of medication errors is independent of length of work shift.

Using the data in Table 3.3, the probability of medication error is calculated as follows:

$$P(\text{Error}) = \frac{\text{Number of cases with errors}}{\text{Number of cases}} = \frac{6}{18} = 0.33,$$

$$P(\text{Long shift}) = \frac{\text{Number of cases seen by a provider in a long shift}}{\text{Number of cases}} = \frac{5}{18} = 0.28,$$

$$P(\text{Error and Long shift}) = \frac{\text{Number of cases with errors and long shift}}{\text{Number of cases}} = \frac{2}{18} = 0.11,$$

$$P(\text{Error and Long shift}) = 0.11 \neq .09 = 0.33 \times 0.28 = P(\text{Error}) \times P(\text{Long shift}).$$

TABLE 3.3 Medication Errors in 18 Consecutive Cases

Case	*Medication Error*	*Long Shift*	*Fatigue*
1	No	Yes	Yes
2	No	Yes	Yes
3	No	No	Yes
4	No	No	Yes
5	Yes	Yes	Yes
6	Yes	No	Yes
7	Yes	No	Yes
8	Yes	Yes	Yes
9	No	No	No
10	No	No	No
11	No	Yes	No
12	No	No	No
13	No	No	No
14	No	No	No
15	No	No	No
16	No	No	No
17	Yes	No	No
18	Yes	No	No

The previous calculations show that the probability of medication error and length of shift are not independent of each other. Knowing the length of the shift tells you something about the probability of error in that shift. However, consider the situation in which you are examining these two events among cases where the provider was fatigued. Now the population of cases you are examining is reduced to the cases 1 through 8. With this population, calculation of the probabilities yields the following:

$$P(\text{Error} \mid \text{Fatigued}) = 0.50,$$

$$P(\text{Long shift} \mid \text{Fatigued}) = 0.50,$$

$$P(\text{Error and Long shift} \mid \text{Fatigued}) = 0.25,$$

$$P(\text{Error and Long shift} \mid \text{Fatigued}) = 0.25 = 0.50 \times 0.50 = P(\text{Error} \mid \text{Fatigued}) \times P(\text{Long shift} \mid \text{Fatigued}).$$

Among fatigued providers, medication error is independent of length of work shift. The procedures used in this example, namely calculating the joint probability and examining it to see if it is approximately equal to the product of the marginal probability, is one way of verifying independence.

Independence can also be examined by calculating conditional probabilities through restricting the population size. For example, in the population of fatigued providers (i.e., in cases 1 through 8) there are several cases of working long shifts (i.e., cases 1, 2, 5, and 8). You can use this information to calculate conditional probabilities as follows:

$$P(\text{Error} \mid \text{Fatigue}) = 0.50,$$
$$P(\text{Error} \mid \text{Fatigue and Long shift}) = \frac{2}{4} = 0.50.$$

This again shows that, among fatigued workers, knowing that the work shift was long adds no information to the probability of medication error. The above procedure shows how independence can be verified by counting cases in reduced populations. If there is a considerable amount of data are available inside a database, the approach can easily be implemented by running a query that would select the condition and count the number of events of interest.

Analyzing Correlations

Another way to verify independence is to examine the correlations among the events (Streiner 2005). Two events that are correlated are dependent. For example, Table 3.4 examines the relationship between age and blood pressure by calculating the correlation between these two variables.

The correlation between age and blood pressure in the sample of data in Table 3.4 is 0.91. This correlation is relatively high and suggests that knowing something about the age of a person will tell you a great deal about the blood pressure. Therefore, age and blood pressure are dependent in this sample.

Partial correlations can also be used to verify conditional independence (Scheines 2002). If two events are conditionally independent from each other, then the partial correlation between the two events given the condition should be zero; this is called a *vanishing partial correlation*. Partial correlation between a and b given c can be calculated from pairwise correlations:

1. R_{ab} is the correlation between a and b.
2. R_{ac} is the correlation between events a and c.
3. R_{cb} is the correlation between event c and b.

Events a and b are conditionally independent of each other if the vanishing partial correlation condition holds. This condition states

$$R_{ab} = R_{ac} \times R_{cb}.$$

Using the data in Table 3.4, you can calculate the following correlations:

TABLE 3.4 Relationship Between Age and Blood Pressure in Seven Patients

Case	*Age*	*Blood Pressure*	*Weight*
1	35	140	200
2	30	130	185
3	19	120	180
4	20	111	175
5	17	105	170
6	16	103	165
7	20	102	155

$$R_{\text{age, blood pressure}} = 0.91,$$

$$R_{\text{age, weight}} = 0.82,$$

$$R_{\text{weight, blood pressure}} = 0.95.$$

Examination of the data shows that the vanishing partial correlation holds (≈ means approximate equality):

$$R_{\text{age, blood pressure}} = 0.91 \approx 0.82 \times 0.95 = R_{\text{age, weight}} \times R_{\text{weight, blood pressure}}.$$

Therefore, you can conclude that, given a patient's weight, the variables of age and blood pressure are independent of each other because they have a partial correlation of zero.

Asking Experts

It is not always possible to gather data. Sometimes, independence must be verified subjectively by asking a knowledgeable expert about the relationship among the variables. Independence can be verified by asking the expert to tell if knowledge of one event will tell you a lot about the likelihood of another. Conditional independence can be verified by repeating the same task, but within specific populations. Gustafson and his colleagues (1973) described a procedure for assessing independence by directly querying experts as follows (see also Ludke, Stauss, and Gustafson 1977; Jeffrey 2004):

1. Write each event on a 3" × 5" card.
2. Ask each expert to assume a specific population in which the target event has occurred.
3. Ask the expert to pair the cards if knowing the value of one clue will alter the affect of another clue in predicting the target event.
4. Repeat these steps for other populations.
5. If several experts are involved, ask them to present their clustering of cards to each other.
6. Have experts discuss any areas of disagreement, and remind them that only major dependencies should be clustered.

7. Use majority rule to choose the final clusters. (To be accepted, a cluster must be approved by the majority of experts.)

Experts will have in mind different, sometimes wrong, notions of dependence, so the words "conditional dependence" should be avoided. Instead, focus on whether one clue tells you a lot about the influence of another clue in specific populations. Experts are more likely to understand this line of questioning as opposed to directly asking them to verify conditional independence.

Strictly speaking, when an expert says that knowledge of one clue does not change the impact of another, we could interpret this to mean

$$\frac{P(C_1|H, C_2)}{P(C_1|\text{Not } H, \text{Not } C_2)} = \frac{P(C_1|H)}{P(C_1|\text{Not } H)}$$

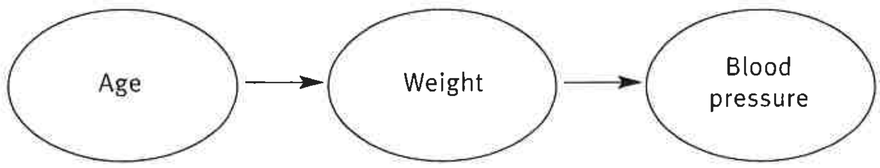
It says that the likelihood ratio of clue #1 does not depend on the occurrence of clue #2. This is a stronger condition that conditional independence because it requires conditional independence both in the population where event H has occurred and in the population where it has not. Experts can make these judgments easily, even though they may not be aware of the probabilistic implications.

Separate in Causal Maps

One can assess dependencies through analyzing maps of causal relationships (Pearl 2000; Greenland, Pearl, and Robins 1999). In a causal network, each node describes an event. The directed arcs between the nodes depict how one event causes another. Causal networks work for situations where there is no cyclical relationship among the variables; it is not possible to start from a node and follow the arcs and return to the same node. An expert is asked to draw a causal network of the events. If the expert can do so, then conditional dependence can be verified by the position of the nodes and the arcs. Several rules can be used to identify conditional dependencies in a causal network, including the following (Pearl 1988):

1. Any two nodes connected by an arrow are dependent. Cause and immediate consequence are dependent.
2. Multiple causes of same effect are dependent, as knowing the effect and one of the causes will indicate more about the probability of other causes.
3. If a cause always leads to an intermediary event that subsequently affects a consequence, then the consequence is independent of the cause given the intermediary event.
4. If one cause leads to multiple consequences, then the consequences are conditionally independent of each other given the cause.

FIGURE 3.5
Causal Map for Age, Weight, and Blood Pressure



In the above rules, it is assumed that removing the condition will actually remove the path between the independent events. For example, think of event A leading to event B and then to event C. Imagine that the relationships are shown by a directed arrow from nodes A to B and B to C. If removal of node C renders nodes A and B disconnected from each other, then A and B are proclaimed independent from each other given C. Another way to say this is to observe that event C is always between events A and B, and there is no way of following the arcs from A to B without passing through C. In this situation, A is independent of B given C:

$$P(A \mid B, C) = P(A \mid C).$$

For example, an expert may provide the map in Figure 3.5 for the relationships among age, weight, and blood pressure.

In Figure 3.5, age and weight are shown to depend on each other. Age and blood pressure are show to be conditionally independent of each other, because there is no way of going from one to the other without passing through the weight node. Note that if there were an arc between age and blood pressure (i.e., if the expert believed there was a direct relationship between these two variables), then conditional independence would be violated. Analysis of causal maps can help identify a large number of independencies among the events being considered. More details and examples for using causal models to verify independence will be presented in Chapter 4.

Summary

One way of measuring uncertainty is through the use of the concept of probability. This chapter defines what probability is and how its calculus can be used to keep track of the probability of multiple events co-occurring, the probability of one or the other event occurring, and the probability of an event that is conditioned on the occurrence of other events. Probability is often thought of as an objective, mathematical process; however, it can also be applied to the subjective opinions and convictions of experts regarding the likelihood of events. Bayes's theorem is introduced as a means of revising subjective probabilities or existing opinions based upon new evidence. The concept of conditional probability is described in terms of reducing the sample space. Conditional independence makes the calculation of Bayes's thorem easier. The chapter provides different methods for testing for conditional independence, including graphical methods, correlation methods, and sample reduction methods.

Review What You Know

1. What is the daily probability of an event that has occurred once in the last year?
2. What is the daily probability of an event that last occurred 3 months ago?
3. What assumption did you make in answering question 2?
4. Using Table 3.5, what is the probability of hospitalization given that you are male?
5. Using Table 3.5, is insurance independent of age?
6. Using Table 3.5, what is the likelihood associated with being older than 65 years among hospitalized patients?
7. Using Table 3.5, in predicting hospitalization, what is the likelihood ratio associated with being 65 years old?
8. What are the prior odds for hospitalization before any other information is available?
9. Analyze the data in the Table 3.5 and report if any two variables are conditionally independent of each other in predicting probability of hospitalization? To accomplish this you need to calculate the likelihood ratio associated with the following clues:
   a. Male
   b. > 65 years old

TABLE 3.5
Sample Cases

Case	*Hospitalized*	*Gender*	*Age*	*Insured*
1	Yes	Male	> 65	Yes
2	Yes	Male	< 65	Yes
3	Yes	Female	> 65	Yes
4	Yes	Female	< 65	No
5	No	Male	> 65	No
6	No	Male	< 65	No
7	No	Female	> 65	No

c. Insured
d. Male and > 65 years old
e. Male and insured
f. > 65 years old and insured

Then you can see if adding a piece of information changes the likelihood ratio. Keep in mind that because the number of cases are too few, many ratios cannot be calculated.

10. Draw what causes medication errors on a piece of paper, with each cause in a separate node and arrows showing the direction of causality. List all causes and their immediate effects until the effects lead to a medication error. Repeat this until all paths to medication errors are listed. It would be helpful if you number the paths.
11. Analyze the graph you have produced and list all conditional dependencies inherent in the graph.

Audio/Visual Chapter Aids

To help you understand the concepts of measuring uncertainty, visit this book's companion web site at ache.org/DecisionAnalysis, go to Chapter 3, and view the audio/visual chapter aids.

References

Bayes, T. 1963. "Essays Toward Solving a Problem in the Doctrine of Changes." *Philosophical Translation of Royal Society* 53:370–418.

Eisenstein, E. L., and F. Alemi. 1994. "An Evaluation of Factors Influencing Bayesian Learning Systems." *Journal of the American Medical Informatics Association* 1 (3): 272–84.

de Finetti, B. 1937. "Foresight: Its Logical Laws, Its Subjective Sources." Translated by H. E. Kyburg, Jr. In *Studies in Subjective Probability*, edited by H. E. Kyburg, Jr. and H. E. Smokler, pp. 93–158. New York: Wiley, 1964.

Jeffrey, R. 2004. *Subjective Probability: The Real Thing*. Cambridge, England: Cambridge University Press.

Greenland, S., J. Pearl, and J. M. Robins. 1999. "Causal Diagrams for Epidemiologic Research." *Epidemiology* 10 (1): 37–48.

Gustafson, D. H., J. J. Kestly, R. L. Ludke, and F. Larson. 1973. "Probabilistic Information Processing: Implementation and Evaluation of a Semi-PIP Diagnostic System." *Computers and Biomedical Research* 6 (4): 355–70.

Ludke, R. L., F. F. Stauss, and D. H. Gustafson. 1977. "Comparison of Five Methods for Estimating Subjective Probability Distributions." *Organizational Behavior and Human Performance* 19 (1): 162–79.

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.

Savage, L. 1954. *The Foundation of Statistics.* New York: John Wiley and Sons.

Scheines, R. 2002. "Computation and Causation." *Meta Philosophy* 33 (1 and 2): 158–80.

Streiner, D. L. 2005. "Finding Our Way: An Introduction to Path Analysis." *Canadian Journal of Psychiatry* 50 (2): 115–22.

Wallsten, T. S., and D. V. Budescu. 1983. "Encoding Subjective Probabilities: A Psychological and Psychometric Review." *Management Science* 29 (2) 151–73.