Chapter 2 Review of Probability

2.1 Background

The material in this chapter is meant to be a brief review of the basic elements of probability. More complete treatments of the basics of probability can be found in any good introductory statistics text, such as Dixon and Massey's (1983) *Introduction to Statistical Analysis* or Winkler's (1972) *Introduction to Bayesian Inference and Decision.*

Our uncertainty about the atmosphere, or about any other system, for that matter, is of different degrees in different instances. For example, you cannot be completely certain whether rain will occur at your home tomorrow, or whether the average temperature next month will be greater or less than the average temperature last month. But it is likely that you are more sure about the latter question than about the former one.

It is not sufficient, or even particularly informative, to say that a particular event is uncertain. Rather, one is faced with the problem of expressing or characterizing degrees of uncertainty. A possible approach is to use qualitative descriptors such as "likely," "unlikely," "possible," or "chance of." Conveying uncertainty through such phrases, however, is ambiguous and open to varying interpretations (Murphy and Brown, 1983). For example, it is not clear whether "rain likely" or "rain probable" indicates less uncertainty about the prospects for rain.

It is generally preferable to express uncertainty quantitatively, and this is done using numbers called *probabilities.* In a limited sense, probability is no more than an abstract mathematical system that can be developed logically from three premises called the *axioms of probability*. This system would be uninteresting to many people, including perhaps yourself, except that the resulting abstract concepts are relevant to real-world systems involving uncertainty. Before presenting the axioms of probability and a few of their more important implications, it is necessary to define some terminology.

2.2 The Elements of Probability

2.2.1 Events

An *event* is a set, or class, or group of possible uncertain outcomes. Events can be of two kinds. A *compound event* can be decomposed into two or more (sub)events, whereas an *elementary event* cannot. As a simple example, think about rolling an ordinary six-sided die. The event "an even number of spots comes up" is a compound event, since it will occur if either two, four, or six spots appear. The event "six spots come up" is an elementary event.

In simple situations like rolling dice it is usually obvious which events are simple and which are compound. But more generally, just what is defined to be elementary or compound often depends on the problem at hand and the purposes for which an analysis is being conducted. For example, the event "precipitation occurs tomorrow" could be an elementary event to be distinguished from the elementary event "precipitation does not occur tomorrow." But if it is important to distinguish further between forms of precipitation, "precipitation occurs" would be regarded as a compound event, possibly comprising the three elementary events "liquid precipitation," "frozen precipitation," and "both liquid and frozen precipitation." If one were interested further in how much precipitation will occur, these three events would themselves be regarded as compound, each composed of at least two elementary events. In this case, for example, the compound event "frozen precipitation" would occur if either of the elementary events "frozen precipitation containing at least 0.01-in. water equivalent" or "frozen precipitation containing less than 0.01-in. water equivalent" were to occur.

2.2.2 The Sample Space

The *sample space* or *event space* is the set of all possible elementary events. Thus the sample space represents the universe of all possible outcomes or events. Equivalently, it is the largest possible compound event.

The relationships among events in a sample space can be represented geometrically, using what is called a *Venn diagram*. Often the sample space is drawn as a rectangle and the events within it are drawn as circles, as in Fig. 2.1a. Here the sample space is the rectangle labeled S, which might be the set of possible precipitation outcomes for tomorrow. Four elementary events are depicted within the boundaries of the three circles. The "No precipitation" circle is drawn not overlapping the others because neither liquid nor frozen precipitation can occur if no precipitation occurs (i.e., in the absence of precipitation). The hatched area common to both "Liquid precipitation" and "Frozen precipitation" represents the event "both liquid and frozen precipitation." That part of S in Fig. 2.1a not surrounded by circles is interpreted as representing the "null event," which cannot occur.

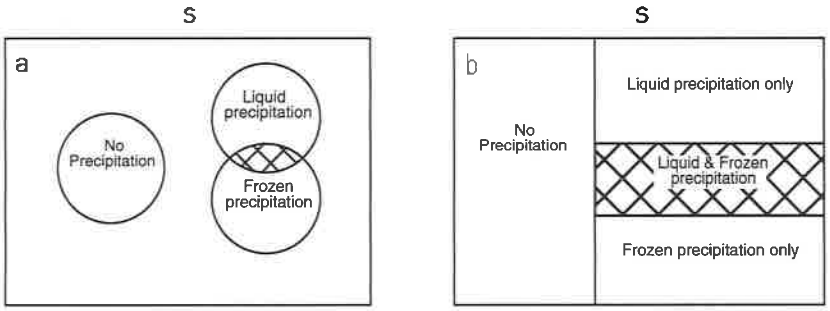

Fig. 2.1 (a) A Venn diagram representing the relationships of selected precipitation events. The hatched region represents the event "both liquid and frozen precipitation." (b) An equivalent Venn diagram drawn to emphasize that S is composed of four MECE events.

It is not necessary to draw or think of Venn diagrams using circles to represent events. Figure 2.1b is an equivalent Venn diagram drawn using rectangles filling the entire sample space S. Drawn in this way, it is clear that S is composed of exactly four elementary events that represent the full range of outcomes that may occur. Such a collection of all possible elementary (according to whatever working definition is current) events is called *mutually exclusive and collectively exhaustive* (MECE). *Mutually exclusive* means that no more than one of the events can occur. *Collectively exhaustive* means that at least one of the events will occur. A set of MECE events completely fills a sample space.

Note that Fig. 2.1b could be modified to distinguish among precipitation amounts by adding a vertical line somewhere in the right-hand side of the rectangle. If the new rectangles on one side of this line were to represent precipitation of 0.01 in. or more, the rectangles on the other side would represent precipitation less than 0.01 in. The modified Venn diagram would then depict seven MECE events.

2.2.3 The Axioms of Probability

Having carefully defined the sample space and its constituent events, we must now associate probabilities with each event. The rules for doing so all flow logically from the three axioms of probability. While formal mathematical definitions are possible, the axioms can be stated somewhat loosely as

1. The probability of any event is nonnegative.
2. The probability of the compound event S is 1.
3. The probability of one or the other of two mutually exclusive events occurring is the sum of their two individual probabilities.

2.3 The Meaning of Probability

The axioms are the essential logical basis for the mathematics of probability. That is, the mathematical properties of probability can all be deduced from the axioms. A number of these properties are listed later in this chapter.

However, the axioms are not very informative about what probability actually means. There are two dominant views of the meaning of probability: the *frequency view* and the *Bayesian view*. Perhaps surprisingly, there has been no small controversy in the world of statistics as to which is correct. Passions have actually run so high on this issue that adherents to one view or the other have been known to launch personal (verbal) attacks on those in the opposite camp!

It is worth emphasizing that the mathematics is the same in either case. The differences are entirely in interpretation. Both of the interpretations of probability have been found to be accepted and useful in the atmospheric sciences, in much the same way that the particle–wave duality of the nature of electromagnetic radiation is accepted and useful in the field of physics.

2.3.1 Frequency Interpretation

This is the mainstream view of probability. Its development in the eighteenth century was motivated by the desire to understand games of chance. In this view, the probability of an event is exactly its long-run relative frequency. This definition is formalized in the *law of large numbers*, which states that the ratio of the number of occurrences of event $\{E\}$ to the number of opportunities for $\{E\}$ to have occurred converges to the probability of $\{E\}$, denoted $\Pr\{E\}$, as the number of opportunities increases. This idea can be written formally as

$$\Pr\left\{\left|\frac{a}{n} - \Pr\{E\}\right| \leq \epsilon\right\} \to 0 \quad \text{as} \quad n \to \infty, \tag{2.1}$$

where a is the number of occurrences, n is the number of opportunities (thus a/n is the relative frequency), and ϵ is an arbitrarily small number.

The frequency interpretation is intuitively reasonable and empirically sound. It is useful in such applications as estimating climatological probabilities by computing historical relative frequencies. For example, in the last 50 years there have been $31 \times 50 = 1550$ August days. If rain has occurred at a location of interest on 487 of those days, a natural estimate for the climatological probability of precipitation at that location during August would be $487/1550 = 0.314$.

2.3.2 Bayesian (Subjective) Interpretation

Strictly speaking, employing the frequency view of probability requires a long series of identical trials. For estimating climatological probabilities from historical weather data this approach presents essentially no problem. However, thinking about probabilities for events such as {the football team at your college or alma mater will win at least one game next season} presents some difficulty in the relative frequency framework. While one can abstractly imagine a hypothetical series of football seasons identical to the upcoming one, this suite of fictitious football seasons is of no help in actually estimating a probability for the event.

The subjective interpretation is that probability represents the degree of belief, or quantified judgment, of a particular individual about the occurrence of an uncertain event. For example, there is now a long history of weather forecasters routinely (and very skillfully) assessing probabilities for events such as precipitation occurrence on days in the near future. If your college or alma mater is a large enough school that professional gamblers take an interest in the outcomes of its football games, probabilities regarding those outcomes are also regularly assessed—subjectively.

Two individuals can have different subjective probabilities for an event without either necessarily being wrong. This does not mean that an individual is free to choose any numbers and call them probabilities. The quantified judgment must be a consistent judgment. This means, among other things, that subjective probabilities must be consistent with the axioms of probability, and thus with the properties of probability implied by the axioms.

2.4 Some Properties of Probability

One reason Venn diagrams can be so useful is that they allow probabilities to be visualized geometrically as areas. One's familiarity with the geometric relationships in the physical world can then be used to better grasp the more ethereal world of probability. Imagine that the area of the rectangle in Fig. 2.1b is 1, according to the second axiom. The first axiom says that no areas can be negative. The third axiom says that the total area of nonoverlapping parts is the sum of the areas of those parts.

A number of mathematical properties of probability that follow logically from the axioms are listed in this section. The geometric analog for probability provided by a Venn diagram can be used to help visualize them.

2.4.1 Domain, Subsets, Complements, and Unions

Together, the first and second axiom imply that the probability of any event will be between zero and one, inclusive. This limitation on the domain of probability can be expressed mathematically as

$$0 \leq \Pr\{E\} \leq 1. \tag{2.2}$$

If $\Pr\{E\} = 0$ the event will not occur. If $\Pr\{E\} = 1$ the event is absolutely sure to occur.

If event $\{E_1\}$ necessarily occurs whenever event $\{E_2\}$ occurs, $\{E_1\}$ is said to be a subset of $\{E_2\}$. For example, $\{E_1\}$ and $\{E_2\}$ might denote occurrence of frozen precipitation, and occurrence of precipitation of any form, respectively. The third axiom implies

$$\Pr\{E_1\} \leq \Pr\{E_2\}. \tag{2.3}$$

The complement of event $\{E\}$ is the (generally compound) event that $\{E\}$ does not occur. In Fig. 2.1b, for example, the complement of the event "liquid and frozen precipitation" is the compound event "either no precipitation, liquid precipitation only, or frozen precipitation only." The complement of the event "at least 0.01 in. of liquid equivalent" is "either zero precipitation or less than 0.01 in. of precipitation." Together the second and third axioms imply

$$\Pr\{E\}^C = 1 - \Pr\{E\}, \tag{2.4}$$

where $\{E\}^C$ denotes the complement of $\{E\}$. Many authors use an overbar as an alternative notation to represent complements. This use of the overbar is very different from its most common statistical meaning, which is to denote an arithmetic average [Eq. (1.1)].

The union of two events is the compound event that one or the other, or both, of the events occur. In set notation, unions are denoted by the symbol $\cup$. As a consequence of the third axiom, probabilities for unions can be computed using

$$\begin{aligned} \Pr\{E_1 \cup E_2\} &= \Pr\{E_1 + E_2\} = \Pr\{E_1 \text{ or } E_2 \text{ or both}\} \\ &= \Pr\{E_1\} + \Pr\{E_2\} - \Pr\{E_1 \cap E_2\}. \end{aligned} \tag{2.5}$$

The symbol $\cap$ is called the the intersection operator, and

$$\Pr\{E_1 \cap E_2\} = \Pr\{E_1, E_2\} = \Pr\{E_1 \text{ and } E_2\} \tag{2.6}$$

is the event that both $\{E_1\}$ and $\{E_2\}$ occur. The notation $\{E_1, E_2\}$ is equivalent to $\{E_1 \cap E_2\}$. Another name for $\Pr\{E_1, E_2\}$ is the joint probability of $\{E_1\}$ and $\{E_2\}$. Equation (2.5) is sometimes called the *additive law of probability*. It holds regardless of whether $\{E_1\}$ and $\{E_2\}$ are mutually exclusive. However, if the two events are mutually exclusive, the probability of their intersection is zero, since mutually exclusive events cannot both occur.

The probability for the joint event, $\Pr\{E_1, E_2\}$ is subtracted in Eq. (2.5) to compensate for its having been counted twice when the probabilities for events $\{E_1\}$ and $\{E_2\}$ are added. This can be seen most easily by thinking about how to find the total geometric area surrounded by the two overlapping circles in Fig. 2.1a. The hatched region in Fig. 2.1a represents the intersection event {liquid precipitation and frozen precipitation}, and it is contained within each of the two circles labeled "Liquid precipitation" and "Frozen precipitation."

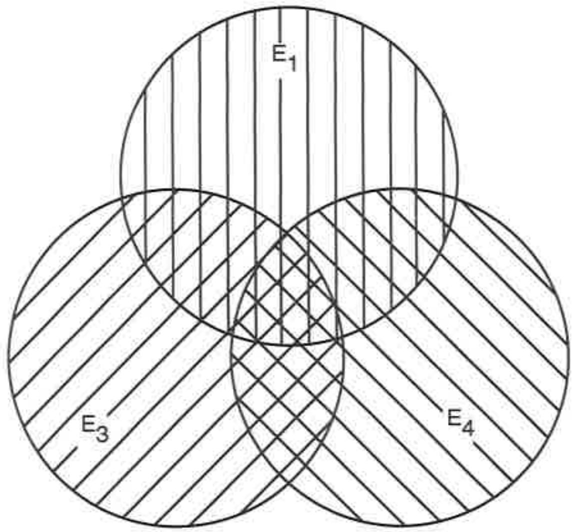

Fig. 2.2 Venn diagram illustrating computation of probability of the union of three intersecting events. The regions with two overlapping hatch patterns have been double-counted, and their areas must be subtracted to compensate. The central region with three overlapping hatch patterns has been triple-counted, but then subtracted three times when the double-counting was corrected. Its area must be added back again.

The additive law [Eq. (2.5)] can be extended to the union of three or more events by thinking of $\{E_1\}$ or $\{E_2\}$ as a compound event (i.e., a union of other events), and recursively applying Eq. (2.5). For example, if $\{E_2\} = \{E_3 \cup E_4\}$, substituting into Eq. (2.5) yields, after some rearrangement

$$\begin{aligned}\Pr\{E_1 \cup E_3 \cup E_4\} = {} & \Pr\{E_1\} + \Pr\{E_3\} + \Pr\{E_4\} \\ & - \Pr\{E_1 \cap E_3\} - \Pr\{E_1 \cap E_4\} - \Pr\{E_3 \cap E_4\} \\ & + \Pr\{E_1 \cap E_3 \cap E_4\}. \end{aligned} \tag{2.7}$$

This result may be difficult to grasp algebraically, but is fairly easy to visualize geometrically. Figure 2.2 illustrates the situation. Adding together the areas of the three circles individually [first line in Eq. (2.7)] results in double-counting of the areas with two overlapping hatch patterns, and triple-counting of the central area contained in all three circles. The second line of Eq. (2.7) corrects the double-counting, but subtracts the area of the central region three times. This area is added back a final time in the third line of Eq. (2.7).

2.4.2 Conditional Probability

It is often the case that one is interested in the probability of an event, given that some other event has occurred or will occur. For example, the probability of

freezing rain, given that precipitation occurs, may be of interest; or perhaps one needs to know the probability of coastal windspeeds above some threshold, given that a hurricane makes landfall nearby. These are examples of conditional probabilities. The event that must be "given" is called the *conditioning event*. The conventional notation for conditional probability is a vertical line, so denoting $\{E_1\}$ as the event of interest and $\{E_2\}$ as the conditioning event, one writes

$$\Pr\{E_1|E_2\} = \Pr\{E_1 \text{ given that } E_2 \text{ has occurred or will occur}\}. \tag{2.8}$$

If the event $\{E_2\}$ has occurred or will occur, the probability of $\{E_1\}$ is the conditional probability $\Pr\{E_1|E_2\}$. If the conditioning event has not occurred or will not occur, this in itself gives no information on the probability of $\{E_1\}$.

More formally, conditional probability is defined in terms of the intersection of the event of interest and the conditioning event, according to

$$\Pr\{E_1|E_2\} = \frac{\Pr\{E_1 \cap E_2\}}{\Pr\{E_2\}}, \tag{2.9}$$

provided the probability of the conditioning event is not zero. Intuitively, it makes sense that conditional probabilities are related to the joint probability of the two events in question, $\Pr\{E_1 \cap E_2\}$. Again, this is easiest to understand through the analogy to areas in a Venn diagram, as shown in Fig. 2.3. We understand the unconditional probability of $\{E_1\}$ to be represented by that proportion of the sample space S occupied by the rectangle labeled E_1. Conditioning on $\{E_2\}$ means that we are interested only in those outcomes containing $\{E_2\}$. We are, in effect, throwing away any part of S not contained in $\{E_2\}$. This amounts to considering a new sample space, S′, that is coincident with $\{E_2\}$. The conditional

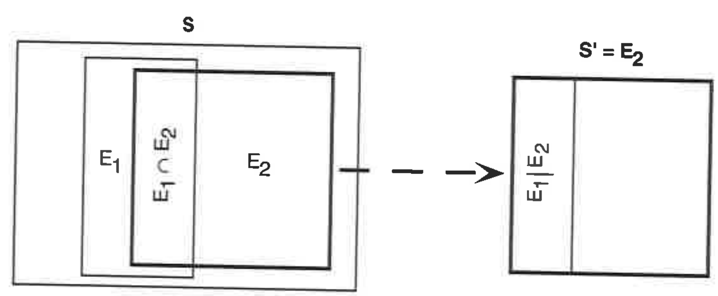

Fig. 2.3 Illustration of the definition of conditional probability. The unconditional probability of $\{E_1\}$ is that fraction of the area of S occupied by $\{E_1\}$ on the left side of the figure. Conditioning on $\{E_2\}$ amounts to considering a new sample space, S′ composed only of $\{E_2\}$, since this means that we are concerned only with occasions when $\{E_2\}$ occurs. Therefore the conditional probability $\Pr\{E_1|E_2\}$ is given by that proportion of the area of the new sample space S′ occupied by both $\{E_1\}$ and $\{E_2\}$. This proportion is computed in Eq. (2.9).

probability $\Pr\{E_1|E_2\}$ is therefore represented geometrically as that proportion of the new sample space area occupied by both $\{E_1\}$ and $\{E_2\}$. If the conditioning event and the event of interest are mutually exclusive, the conditional probability clearly must be zero, since their joint probability will be zero.

2.4.3 Independence

Rearranging the definition of conditional probability Eq. (2.9) yields the form of this expression called the *multiplicative law of probability*:

$$\Pr\{E_1 \cap E_2\} = \Pr\{E_1|E_2\}\Pr\{E_2\} = \Pr\{E_2|E_1\}\Pr\{E_1\}. \tag{2.10}$$

Two events are said to be independent if the occurrence or nonoccurrence of one does not affect the probability of the other. For example, if one rolls a red die and a white die, the probability of an outcome of interest on the red die does not depend on the outcome of the white die, and vice versa. The outcomes of rolling dice are independent. Independence between $\{E_1\}$ and $\{E_2\}$ implies $\Pr\{E_1|E_2\} = \Pr\{E_1\}$ and $\Pr\{E_2|E_1\} = \Pr\{E_2\}$. Independence of events makes the calculation of joint probabilities particularly easy, since the multiplicative law then reduces to

$$\Pr\{E_1 \cap E_2\} = \Pr\{E_1\}\Pr\{E_2\}, \qquad \text{for } \{E_1\} \text{ and } \{E_2\} \text{ independent.} \tag{2.11}$$

Equation (2.11) is extended easily to the computation of joint probabilities for more than two independent events, by simply multiplying all the probabilities of the unconditional events.

Example 2.1. Conditional Relative Frequency

Consider estimating climatological (i.e., long-run, or relative frequency) estimates of probabilities using the data set given in Table A.1 of Appendix A. Climatological probabilities conditional on other events can be computed. Such probabilities are sometimes referred to as *conditional climatological probabilities*, or *conditional climatologies.*

Suppose it is of interest to estimate the probability of at least 0.01 in. of liquid equivalent precipitation at Ithaca in January, given that the minimum temperature is at least 0° F. Physically, these two events would be expected to be related since very cold temperatures typically occur on clear nights, and precipitation occurrence requires clouds. This physical relationship would lead one to expect that these two events would be statistically related (i.e., not independent), and that the conditional probabilities of precipitation given different temperature conditions will be different from each other, and from the unconditional probability. In particular, on the basis of our understanding of the underlying physical processes, we expect the probability of precipitation given minimum temperature of 0° F or higher will be larger than the conditional probability given the complementary event of minimum temperature colder than 0° F.

To estimate this conditional relative frequency, one is interested only in those data records for which the Ithaca minimum temperature was at least 0° F. There are 24 such days in Table A.1. Of these 24 days, 14 show measurable precipitation (ppt), yielding the estimate $\Pr\{\text{ppt} \geq 0.01 \text{ in.}|T_{\min} \geq 0°\text{F}\} = 14/24 \approx 0.58$. The precipitation data for the 7 days on which the minimum temperature was colder than 0° F have been ignored. Since measurable precipitation was recorded on only one of these seven days, one could estimate the probability of precipitation given the complementary conditioning event of minimum temperature colder than 0° F as $\Pr\{\text{ppt} \geq 0.01 \text{ in.}|T_{\min} < 0°\text{F}\} = 1/7 \approx 0.14$. The corresponding estimate of the unconditional probability of precipitation would be $\Pr\{\text{ppt} \geq 0.01 \text{ in.}\} = 15/31 \approx 0.48$. ¤

The differences in the conditional probability estimates calculated in Example 2.1 reflect statistical dependence. Since the underlying physical processes are well understood, one would not be tempted to speculate that relatively warmer minimum temperatures somehow cause precipitation. Rather, the temperature and precipitation events show a statistical relationship because of their (different) physical relationships to clouds. When dealing with statistically dependent variables whose physical relationships may not be known, it is well to remember that statistical dependence does not necessarily imply a physical cause-and-effect relationship.

Example 2.2. Persistence as Conditional Probability

Atmospheric variables often exhibit statistical dependence with their own past or future values. In the terminology of the atmospheric sciences, this dependence through time is usually known as *persistence*. Persistence can be defined as the existence of (positive) statistical dependence among successive values of the same variable, or among successive occurrences of a given event. *Positive dependence* means that large values of the variable tend to be followed by relatively large values, and small values of the variable tend to be followed by relatively small values. It is usually the case that this statistical dependence of meteorological variables in time is positive. For example, the probability of an above-average temperature tomorrow is higher if today's temperature was above average. Thus, another name for persistence is *positive serial dependence*.

Consider characterizing the persistence of the event {precipitation occurrence} at Ithaca, again with the small data set in Table A.1 of Appendix A. Physically, serial dependence would be expected in this daily data because the typical time scale for the midlatitude synoptic waves associated with most winter precipitation at this location is several days. The statistical consequence should be that days for which measurable precipitation is reported should tend to occur in runs, as should days without measurable precipitation.

In order to evaluate serial dependence for precipitation events it is necessary to estimate conditional probabilities of the type Pr{ppt today | ppt yesterday}. Since the data set in Table A.1 contains no records for either December 31, 1986 or February 1, 1987, there are 30 "yesterday/today" data pairs to work with. In order to estimate Pr{ppt today | ppt yesterday} one needs only to count the number of days reporting precipitation (as the conditioning, or "yesterday" event) that are followed by the subsequent day reporting precipitation (as the event of interest, or "today"). When estimating this conditional probability one is not interested in what happens following days on which no precipitation is reported. Excluding January 31, there are 14 days on which precipitation is reported. Of these, 10 are followed by another day with precipitation reported, and four are followed by dry days. The conditional relative frequency estimate would therefore be Pr{ppt today | ppt yesterday} = 10/14 ≈ 0.71. Similarly, conditioning on the complementary event (no precipitation "yesterday") yields Pr{ppt today | no ppt yesterday} = 5/16 ≈ 0.31. The difference in these conditional probability estimates confirms the serial dependence in these data, and quantifies the tendency of the wet and dry days to occur in runs. These two conditional probabilities also constitute a "conditional climatology." ¤

2.4.4 Law of total probability

Sometimes probabilities must be computed indirectly because of limited information. One relationship that can be useful in such situations is the *law of total probability*. Consider a set of MECE events, $\{E_i\}$, $i = 1, \ldots, I$ on the sample space of interest. Figure 2.4 illustrates this situation for $I = 5$ events. If there is an event $\{A\}$, also defined on this sample space, its probability can be computed by summing the joint probabilities

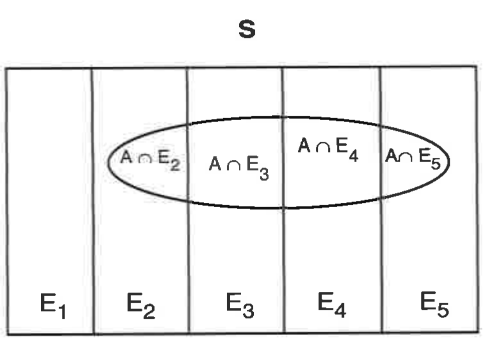

Fig. 2.4 Illustration of the law of total probability. The sample space S contains the event $\{A\}$, represented by the oval, and five MECE events, $\{E_i\}$.

$$\Pr\{A\} = \sum_{i=1}^{I} \Pr\{A \cap E_i\}. \tag{2.12}$$

Substituting the multiplicative law of probability yields

$$\Pr\{A\} = \sum_{i=1}^{I} \Pr\{A|E_i\}\Pr\{E_i\}. \tag{2.13}$$

If the unconditional probabilities $\Pr\{E_i\}$ and the conditional probabilities of $\{A\}$ given the MECE $\{E_i\}$ are known, the unconditional probability of $\{A\}$ can be computed. It is important to note that Eq. (2.13) is correct only if the events $\{E_i\}$ constitute a MECE partition of the sample space.

Example 2.3. Combining Conditional Probabilities Using the Law of Total Probability

Example 2.2 can also be viewed in terms of the law of total probability. Consider that there are only $I = 2$ MECE events partitioning the sample space: $\{E_1\}$ denotes precipitation yesterday and $\{E_2\} = \{E_1\}^C$ denotes no precipitation yesterday. Let event $\{A\}$ be the occurrence of precipitation today. If the data were not available, one could compute $\Pr\{A\}$ using the conditional probabilities through the law of total probability. That is, $\Pr\{A\} = \Pr\{A|E_1\}\Pr\{E_1\} + \Pr\{A|E_2\}\Pr\{E_2\} = (10/14)\,(14/30) + (5/16)\,(16/30) = 0.50$. Since the data are available in Appendix A, the correctness of this result can be confirmed simply by counting. ¤

2.4.5 Bayes' Theorem

Bayes' theorem is an interesting combination of the multiplicative law and the law of total probability. In a relative frequency setting, Bayes' theorem is used to "invert" conditional probabilities. That is, if $\Pr\{E_1|E_2\}$ is known, Bayes' theorem may be used to compute $\Pr\{E_2|E_1\}$. In the Bayesian framework it is used to revise or update subjective probabilities consistent with new information.

Consider again a situation such as that shown in Fig. 2.4, in which there are a defined set of MECE events $\{E_i\}$, and another event $\{A\}$. The multiplicative law [Eq. (2.10)] can be used to find two expressions for the joint probability of $\{A\}$ and any of the events $\{E_i\}$. Since

$$\begin{aligned}\Pr\{A, E_i\} &= \Pr\{A|E_i\}\Pr\{E_i\} \\ &= \Pr\{E_i|A\}\Pr\{A\},\end{aligned} \tag{2.14}$$

combining the two right-hand sides and rearranging yields

$$\Pr\{E_i|A\} = \frac{\Pr\{A|E_i\}\Pr\{E_i\}}{\Pr\{A\}} = \frac{\Pr\{A|E_i\}\Pr\{E_i\}}{\sum_{j=1}^{J} \Pr\{A|E_j\}\Pr\{E_j\}}. \tag{2.15}$$

The law of total probability has been used to rewrite the denominator. Equation (2.15) is the expression for Bayes' theorem. It is applicable separately for each of the MECE events $\{E_i\}$. Note, however, that the denominator is the same for each E_i, since $\Pr\{A\}$ is obtained each time by summing over all the events, indexed in the denominator by the subscript *j*.

Example 2.4. Bayes' Theorem from a Relative Frequency Standpoint

In Example 2.1, conditional probabilities for precipitation occurrence given minimum temperatures above or below 0° F were estimated. Bayes' theorem can be used to compute the converse conditional probabilities, concerning temperature events given that precipitation did or did not occur. Let $\{E_1\}$ represent minimum temperature of 0° F or above, and let $\{E_2\} = \{E_1\}^C$ be the complementary event that minimum temperature is colder than 0° F. Clearly the two events constitute a MECE partition of the sample space. Recall that minimum temperatures of at least 0° F were reported on 24 of the 31 days, so that the unconditional climatological estimates of probabilities for the temperature events would be $\Pr\{E_1\} = 24/31$ and $\Pr\{E_2\} = 7/31$. Recall from Example 2.1 that $\Pr\{A|E_1\} = 14/24$ and $\Pr\{A|E_2\} = 1/7$.

Equation 2.15 can be applied separately for each of the two events $\{E_i\}$. In each case the denominator is $\Pr\{A\} = (14/24)\,(24/31) + (1/7)\,(7/31) = 15/31$. (This differs slightly from the estimate for the probability of precipitation derived in Example 2.2, since there the data for December 31 could not be included.) Using Bayes' theorem, the conditional probability for minimum temperature at least 0° F given precipitation occurrence is $(14/24)\,(24/31)/(15/31) = 14/15$. Similarly, the conditional probability for minimum temperature below 0° F given nonzero precipitation is $(1/7)\,(7/31)/(15/31) = 1/15$. Since all the data are available in Appendix A, these calculations can also be verified directly. ¤

Example 2.5. Bayes' Theorem from a Subjective Probability Standpoint

A subjective (Bayesian) probability interpretation of Example 2.4 can also be made. Suppose a weather forecast specifying the probability of the minimum temperature being at least 0° F is desired. If no more sophisticated information were available, it would be natural to use the unconditional climatological probability for the event, $\Pr\{E_1\} = 24/31$, as representing the forecaster's uncertainty or degree of belief in the outcome. In the Bayesian framework this baseline state of information is known as the *prior probability*. Suppose, however, that the forecaster could know whether precipitation will occur on that day. This information will affect the forecaster's degree of certainty in the temperature outcome. Just how much more certain the forecaster can become depends on the strength of the relationship between temperature and precipitation, expressed in the conditional probabilities of precipitation occurrence given the two minimum temperature outcomes. These conditional probabilities, $\Pr\{A|E_i\}$ in the notation of this example, are known as the *likelihoods*. If precipitation occurs, the forecaster is more certain that the minimum temperature will be at least 0° F, with the revised probability given by Eq. (2.15) as $(14/24)\,(24/31)/(15/31) = 14/15$. This modified, or updated, judgment regarding the probability of a very cold minimum temperature not occurring is called the *posterior probability*. Here the posterior probability is substantially larger than the prior probability of 24/31. Similarly, if precipitation does not occur, the forecaster is more confident that the minimum temperature will not be 0° F or warmer. Note that the differences between this example and Example 2.4 are entirely in the interpretation, and that the computations and numerical results are identical. ¤

Exercises

2.1. In the climatic record for 60 winters at a given location, single-storm snowfalls greater than 35 cm occurred in nine of those winters (define such snowfalls as event *A*), the lowest temperature was below −25° C in 36 of the winters (define this as event *B*). Both events *A* and *B* occurred in three of the winters.
   (a) Sketch a Venn diagram for a sample space appropriate to these data.
   (b) Write an expression using set notation, for the occurrence of 35-cm snowfalls, −25° C temperatures, or both. Estimate the climatological probability for this compound event.
   (c) Write an expression using set notation, for the occurrence of winters with 35-cm snowfalls in which the temperature does not fall below −25° C. Estimate the climatological probability for this compound event.
   (d) Write an expression using set notation, for the occurrence of winters having neither −25° C temperatures nor 35-cm snowfalls. Again, estimate the climatological probability.

2.2. Using the January 1987 data set in Table A.1, define event *A* as Ithaca $T_{max} >$ 32° F and event *B* as Canandaigua $T_{max} >$ 32° F.
   (a) Explain the meanings of $\Pr(A)$, $\Pr(B)$, $\Pr(A,B)$, $\Pr(A \cup B)$, $\Pr(A|B)$, and $\Pr(B|A)$.
   (b) Estimate, using relative frequencies in the data, $\Pr(A)$, $\Pr(B)$, and $\Pr(A,B)$.
   (c) Using the results from part (b), calculate $\Pr(A|B)$.
   (d) Are events *A* and *B* independent? How do you know?

2.3. Three radar sets, operating independently, are searching for "hook" echos (a radar signature associated with tornados). Suppose that each radar has a probability of 0.05 of failing to detect this signature when a tornado is present.
   (a) Sketch a Venn diagram for a sample space appropriate to this problem.