Chapter 1
Theory of Probability

Abstract The theory of probability is the mathematical framework for the study of the probability of occurrence of events. The first step is to establish a method to assign the probability of an event, for example, the probability that a coin lands heads up after a toss. The *frequentist*—or empirical—approach and the *subjective*—or Bayesian— approach are two methods that can be used to calculate probabilities. The fact that there is more than one method available for this purpose should not be viewed as a limitation of the theory, but rather as the fact that for certain parts of the theory of probability, and even more so for statistics, there is an element of subjectivity that enters the analysis and the interpretation of the results. It is therefore the task of the statistician to keep track of any assumptions made in the analysis, and to account for them in the interpretation of the results. Once a method for assigning probabilities is established, the Kolmogorov axioms are introduced as the "rules" required to manipulate probabilities. Fundamental results known as Bayes' theorem and the theorem of total probability are used to define and interpret the concepts of statistical independence and of conditional probability, which play a central role in much of the material presented in this book.

1.1 Experiments, Events, and the Sample Space

Every experiment has a number of possible outcomes. For example, the experiment consisting of the roll of a die can have six possible outcomes, according to the number that shows after the die lands. The *sample space* Ω is defined as the set of all possible outcomes of the experiment, in this case $\Omega = \{1, 2, 3, 4, 5, 6\}$. An *event* A is a subset of Ω, $A \subset \Omega$, and it represents a number of possible outcomes for the experiment. For example, the event "even number" is represented by $A = \{2, 4, 6\}$, and the event "odd number" as $B = \{1, 3, 5\}$. For each experiment, two events always exist: the sample space itself, Ω, comprising all possible outcomes, and $A = \emptyset$, called the *impossible event*, or the event that contains no possible outcome.

M. Bonamente, *Statistics and Analysis of Scientific Data*, Graduate Texts in Physics, DOI 10.1007/978-1-4939-6572-4_1

Events are conveniently studied using set theory, and the following definitions are very common in theory of probability:

- The complementary $\overline{A}$ of an event A is the set of all possible outcomes except those in A. For example, the complementary of the event "odd number" is the event "even number."
- Given two events A and B, the union $C = A \cup B$ is the event comprising all outcomes of A and those of B. In the roll of a die, the union of odd and even numbers is the sample space itself, consisting of all possible outcomes.
- The intersection of two events $C = A \cap B$ is the event comprising all outcomes of A that are also outcomes of B. When $A \cap B = \emptyset$, the events are said to be *mutually exclusive*. The union and intersection can be naturally extended to more than two events.
- A number of events A_i are said to be a *partition* of the sample space if they are mutually exclusive, and if their union is the sample space itself, $\cup A_i = \Omega$.
- When all outcomes in A are comprised in B, we will say that $A \subset B$ or $B \supset A$.

1.2 Probability of Events

The probability P of an event describes the odds of occurrence of an event in a single trial of the experiment. The probability is a number between 0 and 1, where $P = 0$ corresponds to an impossible event, and $P = 1$ to a certain event. Therefore the operation of "probability" can be thought of as a function that transforms each possible event into a real number between 0 and 1.

1.2.1 The Kolmogorov Axioms

The first step to determine the probability of the events associated with a given experiment is to establish a number of basic rules that capture the meaning of probability. The probability of an event is required to satisfy the three axioms defined by Kolmogorov [26]:

1. The probability of an event A is a non-negative number, $P(A) \geq 0$;
2. The probability of all possible outcomes, or sample space, is normalized to the value of unity, $P(\Omega) = 1$;
3. If $A \subset \Omega$ and $B \subset \Omega$ are *mutually exclusive* events, then

$$P(A \cup B) = P(A) + P(B) \tag{1.1}$$

Figure 1.1 illustrates this property using set diagrams. For events that are not mutually exclusive, this property does not apply. The probability of the union is

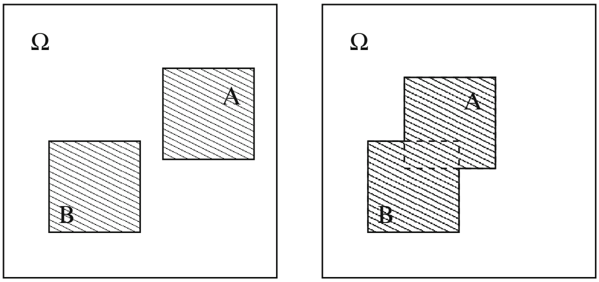

Fig. 1.1 The probability of the event $P(A \cup B)$ is the sum of the two individual probabilities, only if the two events are mutually exclusive. This property enables the interpretation of probability as the "area" of a given event within the sample space

represented by the area of $A \cup B$, and the outcomes that overlap both events are not double-counted.

These axioms should be regarded as the basic "ground rules" of probability, but they provide no unique specification on how event probabilities should be assigned. Two major avenues are available for the assignment of probabilities. One is based on the repetition of the experiments a large number of times under the same conditions, and goes under the name of the frequentist or classical method. The other is based on a more theoretical knowledge of the experiment, but without the experimental requirement, and is referred to as the Bayesian approach.

1.2.2 Frequentist or Classical Method

Consider performing an experiment for a number $N \gg 1$ of times, under the same experimental conditions, and measuring the occurrence of the event A as the number $N(A)$. The probability of event A is given by

$$P(A) = \lim_{N \to \infty} \frac{N(A)}{N}; \tag{1.2}$$

that is, the probability is the relative frequency of occurrence of a given event from many repetitions of the same experiment. The obvious limitation of this definition is the need to perform the experiment an infinite number of times, which is not only time consuming, but also requires the experiment to be repeatable in the first place, which may or may not be possible.

The limitation of this method is evident by considering a coin toss: no matter the number of tosses, the occurrence of heads up will never be exactly 50 %, which is what one would expect based on a knowledge of the experiment at hand.

1.2.3 Bayesian or Empirical Method

Another method to assign probabilities is to use the knowledge of the experiment and the event, and the probability one assigns represents the degree of belief that the event will occur in a given try of the experiment. This method implies an element of subjectivity, which will become more evident in Bayes' theorem (see Sect. 1.7). The Bayesian probability is assigned based on a quantitative understanding of the nature of the experiment, and in accord with the Kolmogorov axioms. It is sometimes referred to as *empirical* probability, in recognition of the fact that sometimes the probability of an event is assigned based upon a practical knowledge of the experiment, although without the classical requirement of repeating the experiment for a large number of times. This method is named after the Rev. Thomas Bayes, who pioneered the development of the theory of probability [3].

Example 1.1 (Coin Toss Experiment) In the coin toss experiment, the determination of the empirical probability for events "heads up" or "tails up" relies on the knowledge that the coin is unbiased, and that therefore it must be true that $P(tails) = P(heads)$. This empirical statement signifies the use of the Bayesian method to determine probabilities. With this information, we can then simply use the Kolmogorov axioms to state that $P(tails) + P(heads) = 1$, and therefore obtain the intuitive result that $P(tails) = P(heads) = 1/2$. ◇

1.3 Fundamental Properties of Probability

The following properties are useful to improve our ability to assign and manipulate event probabilities. They are somewhat intuitive, but it is instructive to derive them formally from the Kolmogorov axioms.

1. The probability of the null event is zero, $P(\emptyset) = 0$.

   Proof Start with the mutually exclusive events $\emptyset$ and Ω. Since their union is Ω, it follows from the Third Axiom that $P(\Omega) = P(\Omega) + P(\emptyset)$. From the Second Axiom we know that $P(\Omega) = 1$, from this it follows that $P(\emptyset) = 0$. □

   The following property is a generalization of the one described above:
2. The probability of the complementary event $\overline{A}$ satisfies the property

$$P(\overline{A}) = 1 - P(A). \tag{1.3}$$

   Proof By definition, it is true that $A \cup \overline{A} = \Omega$, and that A, $\overline{A}$ are mutually exclusive. Using the Second and Third axiom, $P(A \cup \overline{A}) = P(A) + P(\overline{A}) = 1$, from which it follows that $P(\overline{A}) = 1 - P(A)$. □

3. The probability of the union of two events satisfies the general property that

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{1.4}$$

This property generalizes the Third Kolmogorov axiom, and can be interpreted as the fact that outcomes in the overlap region of the two events should be counted only once, as illustrated in Fig. 1.1.

Proof First, realize that the event $A \cup B$ can be written as the union of three mutually exclusive sets, $A \cup B = (A \cap \overline{B}) \cup (B \cap \overline{A}) \cup (A \cap B)$, see Fig. 1.1. Therefore, using the Third axiom, $P(A \cup B) = P(A \cap \overline{B}) + P(B \cap \overline{A}) + P(A \cap B)$.

Then, notice that for any event A and B, it is true that $A = (A \cap \overline{B}) \cup (A \cap B)$, since $\{B, \overline{B}\}$ is a partition of Ω. This implies that $P(A) = P(A \cap B) + P(A \cap \overline{B})$ due to the fact that the two sets are again mutually exclusive, and likewise for event B. It thus follows that $P(A \cup B) = P(A) - P(A \cap B) + P(B) - P(B \cap A) + P(A \cap B) = P(A) + P(B) - P(A \cap B)$. □

Example 1.2 An experiment consists of drawing a number between 1 and 100 at random. Calculate the probability of the event: "drawing either a number greater than 50, or an odd number, at each try."

The sample space for this experiment is the set of numbers $i = 1, \ldots, 100$, and the probability of drawing number i is $P(A_i) = 1/100$, since we expect that each number will have the same probability of being drawn at each try. A_i is the event that consists of drawing number i. If we call B the event consisting of all numbers greater than 50, and C the event with all odd numbers, it is clear that $P(B) = 0.5$, and likewise $P(C) = 0.5$. The event $A \cap B$ contains all odd numbers greater than 50, and therefore $P(A \cap B) = 0.25$. Using (1.4), we find that the probability of drawing either a number greater than 50, or an odd number, is 0.75. This can be confirmed by a direct count of the possible outcomes. ◊

1.4 Statistical Independence

Statistical independence among events means that the occurrence of one event has no influence on the occurrence of other events. Consider, for example, rolling two dice, one after the other: the outcome of one die is independent of the other, and the two tosses are said to be statistically independent. On the other hand, consider the following pair of events: the first is the roll of die 1, and the second is the roll of die 1 *and* die 2, so that for the second event we are interested in the sum of the two tosses. It is clear that the outcome of the second event—e.g., the sum of both dice—depends on the first toss, and the two events are not independent.

Two events A and B are said to be statistically independent if and only if

$$P(A \cap B) = P(A) \cdot P(B). \tag{1.5}$$

At this point, it is not obvious that the concept of statistical independence is embodied by (1.5). A few examples will illustrate the meaning of this definition, which will be explored further in the following section on conditional probability.

Example 1.3 Determine the probability of obtaining two 3 when rolling two dice. This event can be decomposed in two events: $A = \{$die 1 shows 3 and die 2 shows any number$\}$ and $B = \{$die 2 shows 3 and die 1 shows any number$\}$.

It is natural to assume that $P(A) = 1/6$, $P(B) = 1/6$ and state that the two events A and B are independent by nature, since each event involves a different die, which has no knowledge of the other one. The event we are interested in is $C = A \cap B$ and the definition of probability of two statistically independent events leads to $P(C) = P(A \cap B) = P(A) \cdot P(B) = 1/36$. This result can be confirmed by the fact that there is only one combination out of 36 that gives rise to two consecutive 3. ◇

The example above highlights the importance of a proper, and sometimes extended, definition of an event. The more careful the description of the event and of the experiment that it is drawn from, the easier it is to make probabilistic calculation and the assessment of statistical independence.

Example 1.4 Consider the events $A = \{$die 1 shows 3 and die 2 shows any number$\}$ and $B = \{$the sum of the two dice is 9$\}$. Determine whether they are statistically independent.

In this case, we will calculate the probability of the two events, and then check whether they obey (1.5) or not. This calculation will illustrate that the two events are *not* statistically independent.

Event A has a probability $P(A) = 1/6$; in order to calculate the probability of event B, we realize that a sum of 9 is given by the following combinations of outcomes of the two rolls: (3,6), (4,5), (5,4) and (6,3). Therefore, $P(B) = 1/9$. The event $A \cap B$ is the situation in which *both* event A and B occur, which corresponds to the single combination (3,6); therefore, $P(A \cap B) = 1/36$. Since $P(A) \cdot P(B) = 1/6 \cdot 1/9 = 1/54 \neq P(A \cap B) = 1/36$, we conclude that the two events are not statistically independent. This conclusion means that one event influences the other, since a 3 in the first toss has certainly an influence on the possibility of both tosses having a total of 9. ◇

There are two important necessary (but not sufficient) conditions for statistical independence between two events. These properties can help identify whether two events are independent.

1. If $A \cap B = \emptyset$, A and B *cannot* be independent, unless one is the empty set. This property states that there must be some overlap between the two events, or else it is not possible for the events to be independent.

   Proof For A and B to be independent, it must be true that $P(A \cap B) = P(A) \cdot P(B)$, which is zero by hypothesis. This can be true only if $P(A) = 0$ or $P(B) = 0$, which in turn means $A = \emptyset$ or $B = \emptyset$ as a consequence of the Kolmogorov axioms. □

2. If $A \subset B$, then A and B *cannot* be independent, unless B is the entire sample space. This property states that the overlap between two events cannot be such that one event is included in the other, in order for statistical independence to be possible.

   Proof In order for A and B to be independent, it must be that $P(A \cap B) = P(A) \cdot P(B) = P(A)$, given that $A \subset B$. This can only be true if $B = \Omega$, since $P(\Omega) = 1$. □

Example 1.5 Consider the above Example 1.3 of the roll of two dice; each event was formulated in terms of the outcome of both rolls, to show that there was in fact overlap between two events that are independent of one another. ◇

Example 1.6 Consider the following two events: $A = \{$die 1 shows 3 and die 2 shows any number$\}$ and $B = \{$die 1 shows 3 or 2 and die 2 shows any number$\}$. It is clear that $A \subset B$, $P(A) = 1/6$ and $P(B) = 1/3$. The event $A \cap B$ is identical to A and $P(A \cap B) = 1/6$. Therefore $P(A \cap B) \neq P(A) \cdot P(B)$ and the two events are not statistically independent. This result can be easily explained by the fact that the occurrence of A implies the occurrence of B, which is a strong statement of dependence between the two events. The dependence between the two events can also be expressed with the fact that the non-occurrence of B implies the non-occurrence of A. ◇

1.5 Conditional Probability

The conditional probability describes the probability of occurrence of an event A *given* that another event B has occurred and it is indicated as $P(A/B)$. The symbol "/" indicates the statement *given that* or *knowing that*. It states that the event after the symbol is known to have occurred. When two or more events are not independent, the probability of a given event will in general depend on the occurrence of another event. For example, if one is interested in obtaining a 12 in two consecutive rolls of a die, the probability of such event does rely on the fact that the first roll was (or was not) a 6.

The following relationship defines the conditional probability:

$$P(A \cap B) = P(A/B) \cdot P(B) = P(B/A) \cdot P(A); \tag{1.6}$$

Equation (1.6) can be equivalently expressed as

$$P(A/B) = \begin{cases} \dfrac{P(A \cap B)}{P(B)} & \text{if } P(B) \neq 0 \\ 0 & \text{if } P(B) = 0. \end{cases} \tag{1.7}$$

A justification for this definition is that the occurrence of B means that the probability of occurrence of A is that of $A \cap B$. The denominator of the conditional probability is $P(B)$ because B is the set of all possible outcomes that are known to have happened. The situation is also depicted in the right-hand side panel of Fig. 1.1: knowing that B has occurred, leaves the probability of occurrence of A to the occurrence of the intersection $A \cap B$, out of all outcomes in B. It follows directly from (1.6) that if A and B are statistically independent, then the conditional probability is $P(A/B) = P(A)$, i.e., the occurrence of B has no influence on the occurrence of A. This observation further justifies the definition of statistical independence according to (1.5).

Example 1.7 Calculate the probability of obtaining 8 as the sum of two rolls of a die, given that the first roll was a 3.

Call event A the sum of 8 in two separate rolls of a die and event B the event that the first roll is a 3. Event A is given by the probability of having tosses (2,6), (3,5), (4,4), (5,3), (6,2). Since each such combination has a probability of 1/36, $P(A) = 5/36$. The probability of event B is $P(B) = 1/6$. Also, the probability of $A \cap B$ is the probability that the first roll is a 3 and the sum is 8, which can clearly occur only if a sequence of (3,5) takes place, with probability $P(A \cap B) = 1/36$.

According to the definition of conditional probability, $P(A/B) = P(A \cap B)/P(B) = 6/36 = 1/6$, and in fact only combination (5,3)—of the six available with 3 as the outcome of the second toss—gives rise to a sum of 8. The occurrence of 3 in the first roll has therefore increased the probability of A from $P(A) = 5/36$ to $P(A/B) = 1/6$, since not any outcome of the first roll would be equally conducive to a sum of 8 in two rolls. ◇

1.6 A Classic Experiment: Mendel's Law of Heredity and the Independent Assortment of Species

The experiments performed in the nineteenth century by Gregor Mendel in the monastery of Brno led to the discovery that certain properties of plants, such as seed shape and color, are determined by a pair of genes. This pair of genes, or *genotype*, is formed by the inheritance of one gene from each of the parent plants.

Mendel began by crossing two pure lines of pea plants which differed in one single characteristic. The first generation of hybrids displayed only one of the two characteristics, called the *dominant* character. For example, the first-generation plants all had round seed, although they were bred from a population of pure round seed plants and one with wrinkled seed. When the first-generation was allowed to self-fertilize itself, Mendel observed the data shown in Table 1.1 [31].

(continued)

Table 1.1 Data from G. Mendel's experiment

Character	No. of dominant	No. of recessive	Fract. of dominant
Round vs. wrinkled seed	5474	1850	0.747
Yellow vs. green seed	6022	2001	0.751
Violet-red vs. white flower	705	224	0.759
Inflated vs. constricted pod	882	299	0.747
Green vs. yellow unripe pod	428	152	0.738
Axial vs. terminal flower	651	207	0.759
Long vs. short stem	787	277	0.740

Table 1.2 Data from G. Mendel's experiment for plants with two different characters

	Yellow seed	Green seed
Round seed	315	108
Wrinkled seed	101	32

In addition, Mendel performed experiments in which two pure lines that differed by two characteristics were crossed. In particular, a line with yellow and round seed was crossed with one that had green and wrinkled seeds. As in the previous case, the first-generation plants had a 100 % occurrence of the dominant characteristics, while the second-generation was distributed according to the data in Table 1.2.

One of the key results of these experiments goes under the name of *Law of independent assortment*, stating that a daughter plant inherits one gene from each parent plant independently of the other parent. If we denote the genotype of the dominant parent as DD (a pair of dominant genes) and that of the recessive parent as RR, then the data accumulated by Mendel support the hypothesis that the first-generation plants will have the genotype DR (the order of genes in the genome is irrelevant) and the second generation plants will have the following four genotypes: DD, DR, RD and RR, in equal proportions. Since the first three genomes will display the dominant characteristic, the ratio of appearance of the dominant characteristic is expected to be 0.75. The data appear to support in full this hypothesis.

In probabilistic terms, one expects that each second-generation plant has $P(D) = 0.5$ of drawing a dominant first gene from each parent and $P(R) = 0.5$ of drawing a recessive gene from each parent. Therefore, according to the

(continued)

hypothesis of independence in the inheritance of genes, we have

$$\begin{cases} P(DD) = P(D) \cdot P(D) = 0.25 \\ P(DR) = P(D) \cdot P(R) = 0.25 \\ P(RD) = P(R) \cdot P(D) = 0.25 \\ P(RR) = P(R) \cdot P(R) = 0.25. \end{cases} \tag{1.8}$$

When plants differing by two characteristics are crossed, as in the case of the data in Table 1.2, then each of the four events in (1.8) is independently mixed between the two characters. Therefore, there is a total of 16 possibilities, which give rise to 4 possible combinations of the two characters. For example, a display of both recessive characters will have a probability of $1/16 = 0.0625$. The data seemingly support this hypothesis with a measurement of a fraction of 0.0576.

1.7 The Total Probability Theorem and Bayes' Theorem

In this section we describe two theorems that are of great importance in a number of practical situations. They make use of a partition of the sample space Ω, consisting of n events A_i that satisfy the following two properties:

$$\begin{aligned} & A_i \cap A_j = \emptyset, \ \forall i \neq j \\ & \bigcup_{i=1}^{n} A_i = \Omega. \end{aligned} \tag{1.9}$$

For example, the outcomes 1, 2, 3, 4, 5 and 6 for the roll of a die partition the sample space into a number of events that cover all possible outcomes, without any overlap among each other.

Theorem 1.1 (Total Probability Theorem) *Given an event B and a set of events A_i with the properties (1.9),*

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B/A_i) \cdot P(A_i). \tag{1.10}$$

Proof The first equation is immediately verified given that the $B \cap A_i$ are mutually exclusive events such that $B = \cup_i (B \cap A_i)$. The second equation derives from the application of the definition of conditional probability. □

The total probability theorem is useful when the probability of an event B cannot be easily calculated and it is easier to calculate the conditional probability B/A_i given a suitable set of conditions A_i. The example at the end of Sect. 1.7 illustrates one such situation.

Theorem 1.2 (Bayes' Theorem) *Given an event B and a set of events A_i with properties (1.9),*

$$P(A_i/B) = \frac{P(B/A_i)P(A_i)}{P(B)} = \frac{P(B/A_i)P(A_i)}{\sum_{i=1}^{n} P(B \cap A_i)} \tag{1.11}$$

Proof The proof is an immediate consequence of the definition of conditional probability, (1.6), and of the Total Probability theorem, (1.10). □

Bayes' theorem is often written in a simpler form by taking into account two events only, $A_i = A$ and B:

$$P(A/B) = \frac{P(B/A)P(A)}{P(B)} \tag{1.12}$$

In this form, Bayes' theorem is just a statement of how the order of conditioning between two events can be inverted.

Equation (1.12) plays a central role in probability and statistics. What is especially important is the interpretation that each term assumes within the context of a specific experiment. Consider B as the *data* collected in a given experiment—these data can be considered as an event, containing the outcome of the experiment. The event A is a *model* that is used to describe the data. The model can be considered as an ideal outcome of the experiment, therefore both A and B are events associated with the same experiment. Following this interpretation, the quantities involved in Bayes' theorem can be interpreted as in the following:

- $P(B/A)$ is the probability, or *likelihood*, of the data given the specified model, and indicated as $\mathscr{L}$. The likelihood represents the probability of making the measurement B given that the model A is a correct description of the experiment.
- $P(A)$ is the probability of the model A, without any knowledge of the data. This term is interpreted as a *prior probability*, or the degree belief that the model is true before the measurements are made. Prior probabilities should be based upon quantitative knowledge of the experiment, but can also reflect the subjective belief of the analyst. This step in the interpretation of Bayes' theorem explicitly introduces an element of subjectivity that is characteristic of Bayesian statistics.
- $P(B)$ is the probability of collecting the dataset B. In practice, this probability acts as a normalization constant and its numerical value is typically of no practical consequence.
- Finally, $P(A/B)$ is the probability of the model after the data have been collected. This is referred to as the *posterior probability* of the model. The posterior

probability is the ultimate goal of a statistical analysis, since it describes the probability of the model based on the collection of data. According to the value of the posterior probability, a model can be accepted or discarded.

This interpretation of Bayes' theorem is the foundation of Bayesian statistics. Models of an experiment are usually described in terms of a number of parameters. One of the most common problems of statistical data analysis is to estimate what values for the parameters are permitted by the data collected from the experiment. Bayes' theorem provides a way to update the prior knowledge on the model parameters given the measurements, leading to posterior estimates of parameters. One key feature of Bayesian statistics is that the calculation of probabilities are based on a prior probability, which may rely on a subjective interpretation of what is known about the experiment before any measurements are made. Therefore, great attention must be paid to the assignment of prior probabilities and the effect of priors on the final results of the analysis.

Example 1.8 Consider a box in which there are red and blue balls, for a total of $N = 10$ balls. What is known a priori is just the total number of balls in the box. Of the first 3 balls drawn from the box, 2 are red and 1 is blue (drawing is done with re-placement of balls after drawing). We want to use Bayes' theorem to make inferences on the number of red balls (i) present in the box, i.e., we seek $P(A_i/B)$, the probability of having i red balls in the box, given that we performed the measurement $B = \{\text{Two red balls were drawn in the first three trials}\}$.

Initially, we may assume that $P(A_i) = 1/11$, meaning that there is an equal probability of having $0, 1, \ldots$ or 10 red balls in the box (for a total of 11 possibilities) before we make any measurements. Although this is a subjective statement, a uniform distribution is normally the logical assumption in the absence of other information. We can use basic combinatorial mathematics to determine that the likelihood of drawing $D = 2$ red balls out of $T = 3$ trials, given that there are i red balls (also called event A_i):

$$P(B/A_i) = \binom{T}{D} p^D q^{T-D}. \tag{1.13}$$

In this equation p is the probability of drawing one of the red balls in a given drawing assuming that there are i red balls, $p = i/N$, and q is the probability of drawing one of the blue balls, $q = 1 - p = (N - i)/N$. The distribution in (1.13) is known as the binomial distribution and it will be derived and explained in more detail in Sect. 3.1. The likelihood $P(B/A_i)$ can therefore be rewritten as

$$P(B/A_i) = \binom{3}{2} \left(\frac{i}{N}\right)^2 \left(\frac{N-i}{N}\right) \tag{1.14}$$

The probability $P(B)$ is the probability of drawing $D = 2$ red balls out of $T = 3$ trial, for all possible values of the true number of red balls, $i = 0, \ldots, 10$. This probability can be calculated from the Total Probability theorem,

$$P(B) = \sum_{i=0}^{N} P(B/A_i) \cdot P(A_i) \tag{1.15}$$

We can now put all the pieces together and determine the *posterior probability* of having i red balls, $P(A_i/B)$, using Bayes' theorem, $P(A_i/B) = P(B/A_i)P(A_i)/P(B)$.

The equation above is clearly a function of i, the true number of red balls. Consider the case of $i = 0$, i.e., what is the *posterior probability* of having no red balls in the box. Since

$$P(B/A_0) = \binom{3}{2}\left(\frac{0}{N}\right)^2\left(\frac{N-0}{N}\right) = 0,$$

it follows that $P(A_o/B) = 0$, i.e., it is impossible that there are no red balls. This is obvious, since two times a red ball was in fact drawn, meaning that there is at least one red ball in the box. Other posterior probabilities can be calculated in a similar way. ◇

Summary of Key Concepts for this Chapter

- ☐ *Event:* A set of possible outcomes of an experiment.
- ☐ *Sample space*: All possible outcomes of an experiment.
- ☐ *Probability of an Event:* A number between 0 and 1 that follows the Kolmogorov axioms.
- ☐ *Frequentist or Classical approach:* A method to determine the probability of an event based on many repetitions of the experiment.
- ☐ *Bayesian or Empirical approach:* A method to determine probabilities that uses prior knowledge of the experiment.
- ☐ *Statistical independence*: Two events are statistically independent when the occurrence of one has no influence on the occurrence of the other, $P(A \cap B) = P(A)P(B)$.
- ☐ *Conditional probability:* Probability of occurrence of an event given that another event is known to have occurred, $P(A/B) = P(A \cap B)/P(B)$.
- ☐ *Total Probability theorem:* A relationship among probabilities of events that form a partition of the sample space, $P(B) = \sum P(B/A_i)P(A_i)$.
- ☐ *Bayes' theorem*: A relationship among conditional probabilities that enables the change in the order of conditioning of the events, $P(A/B) = P(B/A)P(A)/P(B)$.

Problems

1.1 Describe the sample space of the experiment consisting of flipping four coins simultaneously. Assign the probability to the event consisting of "two heads up and two tails up." In this experiment it is irrelevant to know which specific coin shows heads up or tails up.

1.2 An experiment consists of rolling two dice simultaneously and independently of one another. Find the probability of the event consisting of having either an odd number in the first roll or a total of 9 in both rolls.

1.3 In the roll of a die, find the probability of the event consisting of having either an even number or a number greater than 4.

1.4 An experiment consists of rolling two dice simultaneously and independently of one another. Show that the two events, "the sum of the two rolls is 8" and "the first roll shows 5" are not statistically independent.

1.5 An experiment consists of rolling two dice simultaneously and independently of one another. Show that the two events, "first roll is even" and "second roll is even" are statistically independent.

1.6 A box contains 5 balls, of which 3 are red and 2 are blue. Calculate (a) the probability of drawing two consecutive red balls and (b) the probability of drawing two consecutive red balls, given that the first draw is known to be a red ball. Assume that after each draw the ball is replaced in the box.

1.7 A box contains 10 balls that can be either red or blue. Of the first three draws, done with replacement, two result in the draw of a red ball. Calculate the ratio of the probability that there are 2 or just 1 red ball in the box and the ratio of probability that there are 5 or 1 red balls.

1.8 In the game of baseball a player at bat either reaches base or is retired. Consider three baseball players: player A was at bat 200 times and reached base 0.310 of times; player B was at bat 250 times, with an on-base percentage of 0.296; player C was at bat 300 times, with an on-base percentage 0.260. Find (a) the probability that when either player A, B, or C were at bat, he reached base, (b) the probability that, given that a player reached base, it was A, B, or C.

1.9 An experiment consists of rolling two dice simultaneously and independently of one another. Calculate (a) the probability of the first roll being a 1, given that the sum of both rolls was 5, (b) the probability of the sum being 5, given that the first roll was a 1 and (c) the probability of the first roll being a 1 and the sum being 5. Finally, (d) verify your results with Bayes' theorem.

1.10 Four coins labeled 1 through 4 are tossed simultaneously and independently of one another. Calculate (a) the probability of having an ordered combination heads-tails-heads-tails in the four coins, (b) the probability of having the same ordered combination given that any two coins are known to have landed heads-up and (c) the probability of having two coins land heads up given that the sequence heads-tails-heads-tails has occurred.