Chapter 1

Events and Their Probabilities

The most important questions of life are, for the most part, really only problems of probability.

– Pierre Simon Laplace

What is probability? How to understand it? We try to explain the two problems in this chapter. First, we review briefly the history of probability. Then we give some basic concepts which are necessary to define probability. Based on these concepts, we give three different interpretations: classical probability, geometric probability and frequency. The axiomatic definition of probability is shown to unify various interpretations about probability. If some event occurs, then the probability of original event which we considered before may be changed. We discuss this kind of probability, which is called conditional probability. At last, from the definition of conditional probability, we introduce the independence of events, which is important in many probability models.

1.1 The History of Probability

Probability theory is the branch of mathematics concerned with probability, the analysis of random phenomena.

The mathematical theory of probability originates from analyzing games of chance. Gerolamo Cardano (1501-1576)'s book about games of chance, Liber de ludo aleae (in English, Book on Games of Chance), written around 1564, but not published until 1663, is believed to be the first systematic treatment of probability. Cardano used the game of throwing dice to understand the basic concepts of probability.

One of the famous probability problems is the problem of points. The problem concerns a game of chance with two players who have equal chances of winning each round. The players contribute equally to a prize pot, and agree in advance that the first player to have won a certain number of rounds will collect the entire prize. Now suppose that the game is interrupted by external circumstances before either player has achieved victory. How does one then divide the pot fairly? It is tacitly understood that the division should depend somehow on the number of rounds won by each player, and a player who is close to winning should get a larger part of the pot. But the problem is not merely one of calculation, it also includes deciding what a "fair" division should mean in the first place.

Cardano analyzed the problem but did not get the right answer. So it is generally believed that the modern probability theory was started by the French mathematicians Blaise Pascal(1623-1662) and Pierre Fermat(1601-1665) when they succeeded in deriving the problem of points. In their works, the concepts of the probability of a stochastic event and the expected or mean value of a random variable can be found. Although their investigations were concerned with problems connected with games of chance, the importance of these new concepts was clear to them, as Christiaan Huygens (1629-1695) points out in the first printed probability text, On Calculations in Games of Chance(1657), "The reader will note that we are dealing not only with games, but also that the foundations of a very interesting and profound theory are being laid here." Later, Jacob Bernoulli (1654-1705), Abraham De Moivre (1667-1754), Thomas Bayes (1702-1761), Marquis Pierre Simon Laplace (1749-1827), Johann Carl Friedrich Gauss (1777-1855), and Simeon Denis Poisson (1781-1840) contributed significantly to the development of probability theory. The notable contributors from the Russian school include Pafnuty Lvovich Chebyshev (1821-1894), and his students Andrey Andreyevich Markov (1856-1922) and Aleksandr Mikhailovich Lyapunov (1857-1918) with important works dealing with the law of large numbers.

The deductive theory based on the axiomatic definition of probability that is popular today, which is mainly attributed to Andrei Nikolaevich Kolmogorov (1903-1987), who in the 1930s along with Paul Lévy (1886-1971) found a close connection between the theory of probability and the mathematical theory of sets and functions of a real variable. Although Émile Borel(1871-1956) had arrived at these ideas earlier, putting probability theory on this modern frame work is mainly due to the early 20th century mathematicians. As is the case with all parts of mathematics, probability theory is constructed by means of the axiomatic method. So we regard probability theory as a part of mathematics.

The theory of probability has been developed steadily and applied widely in diverse fields of study since the seventeenth century. One of the most striking features of the present day is the steadily increasing use of the ideas of probability theory in a wide variety of scientific fields, involving matters as remote and different as the prediction by geneticists of the relative frequency with which various characteristics occur in groups of individuals, the calculation by telephone engineers of the density of telephone traffic, the maintenance by industrial engineers of manufactured products at a certain standard of quality, the transmission (by engineers concerned with the design of communications and automatic control systems) of signals in the presence of noise, and the study by physicists of thermal noise in electric circuits and the Brownian motion of particles immersed in a liquid or gas.

More details about the history of probability can be found in references [1, 2].

1.2 Experiment, Sample Space and Random Event

Various kinds of chance are well-known to every one of us from our everyday experience: the outcome of a coin-toss or die-roll, the length of time spent waiting in line, how meteorological phenomena will proceed. In all such situations, we are unable to predict the outcome of an "experiment" or the future course of a process. So our recognition and quantitative characterization of real-word phenomena cannot be complete (indeed being impossible in many situations) unless we account for the great role played by randomness.

A random phenomenon is a situation in which we know what outcomes could happen, but we do not know which particular outcome did or will happen. "Random" in statistics is not a synonym for "haphazard" but a description of a kind of order that emerges only in the long run. So in order to find the essence behind the "Random" phenomena, we need do enough experiments. Usually, such experiments are called random experiment. To describe the outcomes of the random experiments, we need the concepts of sample space and random events.

1.2.1 Basic Definitions

Random experiment usually has the following three characteristics.

(i) Repeatability: it can be repeated under the same conditions.

(ii) Predictability: it has more than one outcome and we know all possible outcomes before the experiment.

(iii) Uncertainty: the outcome of the experiment will not be known in advance.

In fact, our efforts are focused on the random experiments which can be repeated under the same conditions. But we also use the probability method to study some subjects which can not be repeated in application, such as lifetime etc. In this book, we shall abbreviate "random experiment" to experiment and denote it by E. Let us see some examples.

E_1 : Determination of the sex of a newborn child.

E_2 : Roll a die and observe which number appears.

E_3 : Flip two coins and observe the outcomes.

E_4 : Roll two dice and observe the outcomes.

E_5 : Observe call times for a call center.

E_6 : Measure the lifetime of cars.

Now a question is how to record the experiment data after we finish the experiment. Each possible outcome is called an ***sample point*** or ***elementary event*** and is denoted by s or ω. The set of all possible outcomes of an random experiment (if we could know all possible outcomes) is known as the ***sample space*** of the random experiment and is denoted by S or Ω. In this book, we use the notations ω and Ω.

In the following, we give examples about sample space corresponding to $\mathrm{E}_1 - \mathrm{E}_6$.

Example 1.2.1 For E_1, $\Omega_1 = \{g, b\}$, where the outcome g means that the child is a girl and b that it is a boy.

Example 1.2.2 For E_2, $\Omega_2 = \{1, 2, 3, 4, 5, 6\}$, where the outcome i means that i appeared on the die, $i = 1, 2, 3, 4, 5, 6$.

Example 1.2.3 For E_3, $\Omega_3 = \{(H, H), (H, T), (T, H), (T, T)\}$, where H means head and T means tail.

Example 1.2.4 For E_4,

$$\Omega_4 = \begin{pmatrix} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{pmatrix}$$

where the outcome (i, j) is said to occur if i appears on the first die and j appears on the second die, $i, j = 1, 2, \cdots, 6$.

Example 1.2.5 For E_5, $\Omega_5 = 0, 1, 2, \cdots$, where the outcome i is the number of call times, $i = 0, 1, \cdots$.

Example 1.2.6 For E_6, $\Omega_6 = [0, \infty)$, where the outcome t is the lifetime of a car, $0 \leqslant t < \infty$.

Any subset of the sample space S is known as an ***random event*** or ***event***. Events are usually denoted by capital letters $A, B, C, \cdots$. We say that the event A occurs when the outcome of the experiment lies in A. Those events must occur in the experiment are called the ***inevitable events***. We always denote inevitable event by S or Ω (in this book, we use Ω). Sample space Ω is an inevitable event. For Example 1.2.2, $A = \{1, 2, 3, 4, 5, 6, 7, 8\}$ is an inevitable event. Those could not happen anytime are said to be ***impossible events***. We usually denote impossible events by $\emptyset$.

In Example 1.2.1, if $A = \{g\}$, then A is the event that newborn child is a girl.

In Example 1.2.2, if $A = \{2, 4, 6\}$, then A would be the event that an even number appears on the roll.

In Example 1.2.3, if $A = \{(H, H), (H, T)\}$, then A is the event that a head appears on the first coin.

In Example 1.2.4, if $A = \{(1, 3), (2, 2), (3, 1)\}$, then A is the event that the sum of the dice equals four.

In Example 1.2.5, if $A = (2, 5)$, then A is the event that the number of the call times for a call center is 3 or 4.

In Example 1.2.6, if $A = (2, 5)$, then A is the event that the car lasts between two and five years.

1.2.2 Events as Sets

Since random events are subsets of sample space, the relationships and operations between random events could be described in term of set theory.

Let Ω be the sample space of the random experiment E and $A, B, A_i (i = 1, 2, \cdots)$ be the random events of E.

1. It is said that an event A is ***contained in*** another event B if every outcome that belongs to the subset defining the event A also belongs to the subset defining the event B. The relation between two events is expressed symbolically by $\boldsymbol{A \subset B}$. That means ***if event A occurs, then B occurs***. The relation $A \subset B$ is also expressed by saying that A is a subset of B. Equivalently, if $A \subset B$, we may say that B *contains* A and may write $B \supset A$. In the experiment E_2, suppose that $A = \{2, 4, 6\}$ and $C = \{2, 3, 4, 5, 6\}$, it follows that $A \subset C$. It should be noted that $A \subset C$ for every event A.

If two events A and B are so related that $A \subset B$ and $B \subset A$, it follows that A and B must contain exactly the same points. In other words, $A = B$.

2. ***The union*** of the event A and the event B, denoted by $\boldsymbol{A \cup B}$, consists of all outcomes that lie in either A or B. That is, the event $A \cup B$ will occur if ***either A or B occurs***. In the experiment E_1, suppose that $A = \{H\}$ and $B = \{T\}$, then $A \cup B = \{H, T\} = S$.

We also define unions of more than two events in a similar manner.

$$\bigcup_{i=1}^{n} A_i \triangleq A_1 \cup A_2 \cup \cdots \cup A_n$$

is called the union of events $A_1, A_2, \cdots, A_n$ and defined to be the event that consists of all outcomes that are in A_i for at least one value of $i = 1, 2, \cdots, n$.

$$\bigcup_{i=1}^{\infty} A_i \triangleq A_1 \cup A_2 \cup \cdots$$

is called the union of events $A_1, A_2, \cdots$.

3. ***The intersection*** of the event A and the event B, denoted by $\boldsymbol{A \cap B}$ or AB, consists of all outcomes which are both in A and B. That is, the event AB will occur only if ***both A and B occur***. In the experiment E_3, if $A = \{(H, H), (T, H), (T, T)\}$ and $B = \{(T, H), (T, T)\}$, then

$$B \subset A, AB = B = \{(T, H), (T, T)\}$$

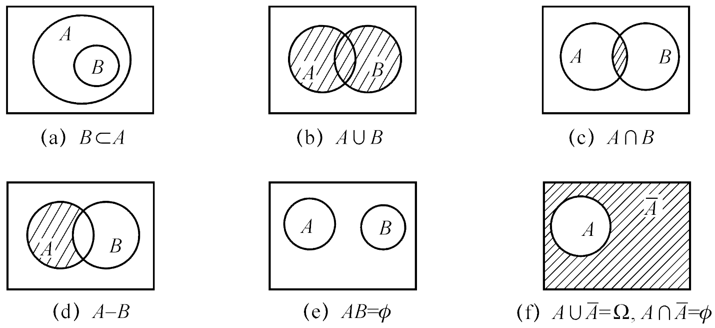

Figure 1.1: The relationships between events A and B

and thus AB would occur if the outcome is either (T, H) or (T, T).

Similarly, we can define the intersection of events $A_1, A_2, \cdots, A_n$ by the event

$$\bigcap_{i=1}^{n} A_i \triangleq A_1 \cap A_2 \cap \cdots \cap A_n$$

which consists of those outcomes that are in all of the events $A_i, i = 1, 2, \cdots, n$. And

$$\bigcap_{i=1}^{\infty} A_i \triangleq A_1 \cap A_2 \cap \cdots$$

is called the intersection of events $A_1, A_2, \cdots$.

4. ***The difference*** of events A and B, denoted by $\boldsymbol{A-B}$, consists of the outcomes that are in event A but not in event B. That means the event $A - B$ will occur if ***A occurs but B does not occur***. In the experiment E_5, if $A = (2, 12)$ and $B = (4, 13]$, then $A - B = (2, 4]$.

5. Events A and B are ***disjoint or mutually exclusive*** if $\boldsymbol{AB} = \emptyset$, which means AB would not consists of any outcome and hence could not occur. That is, ***A and B can not happen at the same time***.

More generally, the events $A_1, A_2, \cdots, A_n, \cdots$ are said to be ***pairwise disjoint*** or mutually exclusive if $A_i A_j = \emptyset$ whenever $i \neq j$.

6. Event B is said to be the ***complement*** of event A with respect to Ω if $A \cup B = \Omega$ and $AB = \emptyset$. That is, B will occur if and only if ***A does not occur***. B is usually denoted by $\overline{A}$ or A^c. In the experiment E_4, if $A = \{(1, 3), (2, 2), (3, 1)\}$, then $\overline{A}$ will occur if the sum of the dice does not equal four.

Since events could be regarded as sets, we can describe their relationships by Venn diagram, see Fig. 1.1. We know the jargons in probability theory are different from those in set theory. Here we list their connections and differences in Table1.1.

Let A, B and C be the random events of experiment E. The operations of the events will satisfy the following rules:

(i) Commutatively $A \cup B = B \cup A$, $AB = BA$.

(ii) Associatively $A \cup (B \cup C) = (A \cup B) \cup C = A \cup B \cup C$.

$A(BC) = (AB)C = ABC$.

(iii) Distributively $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$,

$A(B \cup C) = (AB) \cup (AC)$.

(iv) De Morgan's law $\overline{A \cup B} = \overline{A} \cap \overline{B}$, $\overline{A \cap B} = \overline{A} \cup \overline{B}$.

More generally, $\overline{(\bigcup_i A_i)} = \bigcap \overline{A_i}$, $\overline{(\bigcap_i A_i)} = \bigcup \overline{A_i}$.

In addition, there are some common properties, such as

(1) $\overline{\overline{A}} = A$.

(2) $A \cup B \supset A$ and $A \cup B \supset B$. In particular, if $A \subset B$, then $A \cup B = B$.

(3) $A \cap B \subset A$ and $A \cap B \subset B$. In particular, if $A \subset B$, then $A \cap B = A$.

(4) $A - B = A - AB = A\overline{B}$.

(5) $A \cup B = A \cup \overline{A}B$.

Example 1.2.7 Consider the sample space Ω consists of all positive integers less than 10, i.e., $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Let A be the event consisting of all even numbers and B be the event consisting of numbers divisible by 3. Find $\overline{A}, A \cup B, AB$ and $\overline{A}B$.

Typical notation	Set jargon	Probability jargon
Ω	Collection of objects	Sample space
ω	Member of Ω	Elementary event, outcome
A	Subset of Ω	Events that some outcome in A occurs
A or $\overline{A}$	Complement of A	Event that no outcome in A occurs
$A \cap B$	Intersection	Both A and B
$A \cup B$	Union	Either A or B or both
$A - B$	Difference	A, but no B
$A \subseteq B$	Inclusion	If A, then B
$\emptyset$	Empty set	Impossible event
Ω	Whole space	Certain set

Table 1.1: The jargons in set theory and probability theory

Solution. We have $A = \{2, 4, 6, 8\}$ and $B = \{3, 6, 9\}$. Thus
$\overline{A} = \{1, 3, 5, 7, 9\}$, $A \cup B = \{2, 3, 4, 6, 8, 9\}$, $AB = \{6\}$ and $\overline{A}B = \{3, 9\}$. ∎

Example 1.2.8 Suppose that A, B and C are three events and $D = \{$at least one of the three events will occur$\}$. Try to describe the event D by events A, B and C.

Solution. We describe the event D by three different ways:
(i) directed method: $D = A \cup B \cup C$,
(ii) decomposition method: $D = A\overline{B}\,\overline{C} \cup \overline{A}B\overline{C} \cup \overline{A}\,\overline{B}C \cup \overline{A}BC \cup A\overline{B}C \cup AB\overline{C} \cup ABC$,
(iii) inverse method: $D = \overline{\overline{D}} = \overline{\{\text{A, B and C will not occur}\}} = \overline{\overline{A}\,\overline{B}\,\overline{C}}$. ∎

1.3 Probabilities Defined on Events

The events in random experiment may occur or not. The problem of discussing the chance they occur is probability problem. In this section, we give three different interpretations about probability: classical probability, geometric probability and frequency. We also calculate many concrete probabilities of specific events in different cases.

1.3.1 Classical Probability

The classical definition or interpretation of probability is identified with the works of Jacob Bernoulli and Pierre-Simon Laplace. It is based on the concept of equally likely outcomes. For example, when a fair coin is tossed, there are two possible outcomes: a head or a tail. If we assume that the sum of these outcomes' probabilities is 1, then both the probability of a head and the probability of a tail must be 1/2 respectively. More generally, if the outcome of some experiment is one of n different outcomes, and if these n outcomes are equally likely to occur, then the probability of each outcome is $1/n$.

Generally, a random experiment E is ***classical*** if

(i) E contains only different limited basic events, that is, $\Omega = \{\omega_1, \omega_2, \cdots, \omega_n\}$. We call this kind of sample space ***simple space***, and

(ii) all outcomes are equally likely to occur.

If elementary events are assigned equal probabilities, then the probability of a disjunction of elementary events is just the number of events in the disjunction divided by the total number of elementary events. That is, for classical random experiment E, $\Omega = \{\omega_1, \omega_2, \cdots, \omega_n\}$, we define the probability of event A as

$$P(A) = \frac{\#A}{\#\Omega} \tag{1.1}$$

where $\#A$ means the number of all possible outcomes of event A, $\#\Omega$ means the number of all possible outcomes of sample space Ω. If A is comprised by k different elementary events, then $P(A) = \dfrac{k}{n}$. The probability of impossible event is defined as

$$P(\emptyset) = 0. \tag{1.2}$$

For classical random experiment E, the corresponding problems of probability belong to the category of ***classical probability***.

Theorem 1.3.1 *For classical random experiment* E, *the probability has the following properties:*

(i) *for every event* A, $P(A) \geqslant 0$,

(ii) $P(\Omega) = 1$,

(iii) *for every finite sequence of* n *disjoint events* $A_1, A_2, \cdots, A_n$,

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{i=1}^{n} P(A_i).$$

Property (iii) *is called finite additivity.*

Proof. (i) and (ii) are obvious according to the equation (1.1) and equation (1.2). Since Ω has n different elementary events, we can get $P(\Omega) = \dfrac{n}{n} = 1$.

Now let us prove (iii). Suppose that

$$A_i = \{s_1^i, s_2^i, \cdots, s_{k_i}^i\}$$

satisfying $s_l^i \neq s_r^i$, where $i = 1, 2, \cdots, m$, $1 \leqslant l, r \leqslant k_i$ and $l \neq r$.

We know that $P(A_i) = \dfrac{k_i}{n}$ from equation (1.1). It is obvious that

$$\bigcup_{i=1}^{m} A_i = \{s_1^1, s_2^1, \cdots, s_{k_1}^1; s_1^2, s_2^2, \cdots, s_{k_2}^2; \cdots; s_1^m, s_2^m, \cdots, s_{k_m}^m\}.$$

Since $A_1, A_2, \cdots$ and A_m are disjoint events, these elementary events must be different. Then we have

$$P\Big(\bigcup_{i=1}^{m} A_i\Big) = \Big(\sum_{i=1}^{m} k_i\Big)\Big/ n = \sum_{i=1}^{m} \frac{k_i}{n} = \sum_{i=1}^{m} P(A_i).$$

■

In the following we give some examples about classical probability.

Example 1.3.1 Consider the dice-roll experiment E_4 and calculate the probability of each of the possible values of the sum of the two numbers that may appear.

Solution. From Example 1.2.4, we know $\#\Omega = 36$. It can be easily found that the probability of each of the outcomes in E_4 is $1/36$. Let A_i denote the event of the sum of the two numbers is i, $i = 2, 3, \cdots, 12$. For A_2, the only outcome in Ω for which the sum is 2 is the outcome $(1, 1)$. Therefore,

$$P(A_2) = \frac{\#A_2}{\#\Omega} = 1/36.$$

By continuing in this manner, we obtain the following probability for each of the possible values of the sum:

$$P(A_2) = P(A_{12}) = \frac{1}{36}, \quad P(A_3) = P(A_{11}) = \frac{2}{36}, \quad P(A_4) = P(A_{10}) = \frac{3}{36}$$

$$P(A_5) = P(A_9) = \frac{4}{36}, \qquad P(A_6) = P(A_8) = \frac{5}{36}, \qquad P(A_7) = \frac{6}{36}.$$

■

Example 1.3.2 Suppose that a urn contains α white balls and β black balls, and that $m+n$ balls are to be chosen at random at one time ($m \leqslant \alpha, n \leqslant \beta$). We shall determine the probability that exactly m white balls will be chosen.

Solution. Since $m+n$ balls are to be chosen at random at one time, the number of all possible outcomes of sample space Ω is $\binom{\alpha+\beta}{m+n}$, and each of these $\binom{\alpha+\beta}{m+n}$ possible combinations is equally probable.

Let A be the event that contains exactly m white balls and n black balls. Now let us find the number of all possible outcomes of event A. When a combination of m white balls and n black balls is formed, the number of different combinations in which m white balls can be selected from the α available balls is $\binom{\alpha}{m}$, and the number of different combinations in which β black balls can be selected from the β available balls is $\binom{\beta}{n}$. Since each of these combinations of m white balls can be paired with each of the combinations of n black balls to form a distinct sample, the number of combinations containing exactly m white balls is $\binom{\alpha}{m}\binom{\beta}{n}$.

Therefore, the desired probability is

$$P(A) = \frac{\#A}{\#\Omega} = \frac{\binom{\alpha}{m}\binom{\beta}{n}}{\binom{\alpha+\beta}{m+n}}.$$

■

Remark: You can consider some similar questions, such as taking the ball one by one without replacement or with replacement.

Example 1.3.3 Suppose that an urn contains α white balls and β black balls, and that $k+1$ ($k+1 \leqslant \alpha+\beta$) balls are to be selected consecutively at random without replacement. Try to determine the probability that the last selected ball is exactly white ball.

Solution. In this experiment, we should distinguish outcomes with the same balls in different orders. So, there are $\binom{\alpha+\beta}{k+1}(k+1)!$ possible outcomes, i.e., $\#\Omega = \binom{\alpha+\beta}{k+1}(k+1)!$. Let A be the event that the last selected ball is exactly white ball. The last selected white ball may be any one of the α white balls. There are $\binom{\alpha+\beta-1}{k}$ ways to choose k balls from $\alpha+\beta-1$ balls. No matter which of these ways we choose, there are $k!$ ways to arrange the k balls. That is to say that $\#A = \alpha\binom{\alpha+\beta-1}{k}k!$. Then we get

$$P(A) = \frac{\alpha\binom{\alpha+\beta-1}{k}k!}{\binom{\alpha+\beta}{k+1}(k+1)!} = \frac{\alpha}{\alpha+\beta}.$$

It is worth noting that the probability has nothing to do with the value of k. ■

Example 1.3.4 Suppose that there are n people, each person will be assigned to any of the $N(n \leqslant N)$ rooms with the same probability $1/N$. We shall determine the probabilities of the following events:

A : For the given n rooms, there is exactly one person in one room.

B : There is exactly one person in one room.

C : There are exactly m people in a given room.

Solution. In this experiment, the number of all different possible outcomes are N^n.

For event A, since n rooms has been given in advance, the first people has n choices, the second people has $n-1$ choices, $\cdots$. So there are $n!$ possible outcomes in event A. We then calculate

$$P(A) = \frac{n!}{N^n}.$$

For event B, the number of the outcomes that n rooms are to be chosen at random from N rooms is $\binom{N}{n}$. The next step is the same as in event B, then

$$P(B) = \frac{\binom{N}{n} n!}{N^n}.$$

For event C, the first step is choosing m people from n people at random and let them in the given room, and the number of this process is $\binom{n}{m}$. The second step is that $n-m$ people are about to be assigned $N-1$ rooms in any way and the number of this process is $(N-1)^{n-m}$. So the probability of event C is

$$P(C) = \frac{\binom{n}{m}(N-1)^{n-m}}{N^n}.$$

■

The following example is an interesting application of the question B in Example 1.3.4.

Example 1.3.5 (**Birthday problem**) If a group consists of n people, what is the probability that at least two of them have the same birthday? Ignore leap years and assume that each day in the year is equally likely as a birthday.

Solution. Let us consider the sample space of all ordered lists of n birthdays. One such list assigns a birthday to each of the n people. All simple events in this sample space are equally likely to occur. Consider the event

$$B = \{\text{at least two people have the same birthday}\}.$$

To compute $P(B)$, we first compute $P(\overline{B})$ and then use Theorem 1.3.1(ii) and (iii), we know $P(B) = 1 - P(\overline{B})$. The complement of B is the event

$$\overline{B} = \{\text{no two people have the same birthday}\}.$$

From example 1.3.4(B), its probability is given by

$$P(\overline{B}) = \frac{365 \cdot 364 \cdots (365 - n + 1)}{365^n}.$$

Therefore,

$$P(B) = 1 - P(\overline{B}) = 1 - \frac{365 \cdot 364 \cdots (365 - n + 1)}{365^n}.$$

It can be easily found that the probability reaches 100% when the number of people reaches 367 (since there are 366 possible birthdays, including February 29). However, 99% probability is reached with just 57 people, and 50% probability with 23 people. This means that if there are 23 or more people in a room, there is more than 1 chance out of 2 that at least two people have the same birthday! ■

It is important that we do not confuse the above birthday problem with the following one. If a group of n people is in a room, what is the probability that at least one person in the group has the same birthday as you. In this case, let

$$B = \{\text{at least one person has the same birthday as you}\}.$$

We obtain in this case

$$P(B) = 1 - \Big(\frac{364}{365}\Big)^n.$$

The smallest value of n where $P(B) > 1/2$ is now $n = 253$ which is much larger than 23.

Example 1.3.6 A reception has received 12 visits in a week. Suppose that all 12 receptions are proceeded on Tuesday and Thursday. Is the reception time required?

Solution. Assume that the reception time is not specified. Receptions will be occurred at any day in a week, then, the probability of the event that all 12 receptions are proceeded on Tuesday and Thursday is

$$\frac{2^{12}}{7^{12}} = 0.0000003.$$

This is a very small probability. Practical experience shows that rare event should seldom occur (referred to as the impossible principle). Now in this example, the rare events have happened in one experiment. So there is a reason to doubt the validity of the assumption and we can reach the conclusion that the reception time is required. ■

1.3.2 Geometric Probability

The Classical definition of probability limited its application only to situations where there are a finite number of possible outcomes and all the events have equal probability. It mainly considered discrete events and its methods were mainly combinatorial. When the concept of equally likely outcomes is about to be extended to a line or 2D or 3D area, the definition of classical probability can not be used any more. For example, what is the probability that a point is selected randomly from the interval $[0,1]$? If your answer is 0, how can you get that? Is it coming from $\frac{1}{\infty}$? If so, what is the probability of choosing the interval $[0,0.5)$? Is your answer $\frac{1}{2}$? Obviously, we can not get that by classical probability.

Moreover, there are many problems similar to above examples on 2D or 3D, such as

what is the mean length of a random chord of a unit circle?

What is the chance that a needle dropped randomly onto a floor marked with equally spaced parallel lines will cross one of the lines?

What is the mean area of the polygonal regions formed when randomly oriented lines are spread over the plane?

What is the chance that three random points in the plane form an acute (rather than obtuse) triangle?

Problems of the above type and their solution techniques were first studied in the 18th century, and the general topic became known as ***geometric probability***. A random experiment E is called to be ***geometric*** if

(i) the sample space is a measurable (such as length, area, volume, etc.) region, i.e., $0 < L(\Omega) < \infty$, and

(ii) the probability of every event $A \subset \Omega$ is proportional to the measure $L(A)$ and has nothing to do with its position and shape.

In this case, we define the probability of event A as

$$P(A) = \frac{L(A)}{L(\Omega)}$$

and $P(\emptyset) = 0$.

Theorem 1.3.2 *For Geometrical random experiment* E, *the probability has the following properties:*

(i) *for every event* A, $P(A) \geqslant 0$,

(ii) $P(\Omega) = 1$, *and*

(iii) *for every countable disjoint events* $A_1, A_2, \cdots$,

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i).$$

Property (iii) *is called countable additivity.*

Proof. (i) and (ii) are obvious.

(iii) According to countable additivity in measure theory, i.e., for every countable disjoint sequence $A_1, A_2, \cdots$,

$$L(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} L(A_i),$$

then

$$P\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \frac{L\Big(\bigcup\limits_{i=1}^{\infty} A_i\Big)}{L(\Omega)} = \sum_{i=1}^{\infty} \frac{L(A_i)}{L(\Omega)} = \sum_{i=1}^{\infty} P(A_i).$$

■

In the following we give some examples about geometric probability.

Example 1.3.7 (**Lunch date problem**) You and one of your friends arrange to meet between 12:00 and 13:00. As a result, it is possible for one of you to arrive at random between 12:00 and 13:00 and waits exactly 20 minutes for another one. After 20 minutes, one of you leaves if another person has not arrived. What is the probability that you and your friend will meet?

Solution. Let x and y be the time you and your friend arriving the gate, respectively, then our sample space is a square $\Omega = \{(x, y) |\ 12 \leqslant x, y \leqslant 13\}$. Let M be the event that two of you will meet at the gate, then M will occur if and only if $|x - y| \leqslant 1/3$, i.e.,

$$M = \{(x, y) |\ 12 \leqslant x, y \leqslant 13,\ |x - y| \leqslant 1/3\}.$$

M is shown as the shaded area in Fig. 1.2. In this problem, the expression “equally likely to occur” means that the probability that the sample point is located in a special region $M \subset \Omega$ is proportional to the area of M. Again, since the certain event Ω has area 1, the probability of M is equal to the area of M, i.e., $P(M) = 1 - \left(\frac{2}{3}\right)^2 = \frac{5}{9}$. ■

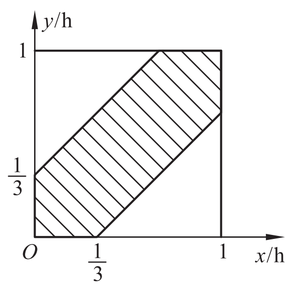

Figure 1.2: Lunch date problem

Example 1.3.8 (**Buffon’s needle problem**) Given a needle of length l dropped on a plane ruled with parallel lines d ($l < d$) units apart, see Fig. 1.3(a). What is the probability that the needle will cross a line?

Solution. Let A be the event that the needle will cross a line. Let x be the distance from the center of the needle to the closest line, see Fig. 1.3(b). Then x will be in the range $0 \leqslant x \leqslant d/2$. Let θ be the acute angle between the needle and the lines. Then θ will be in the range $0 \leqslant \theta \leqslant \pi$. That is,

$$\Omega = \{(x, \theta) |\ 0 \leqslant x \leqslant d/2,\ 0 \leqslant \theta \leqslant \pi\}$$

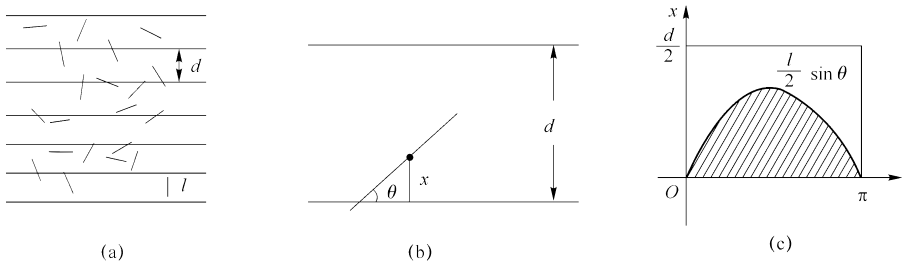

Figure 1.3: Buffon's needle problem

in the $Ox\theta$ plane. Since the needle intersects with one line if and only if $x \leqslant \frac{l\sin\theta}{2}$, i.e.,

$$A = \{(x,\theta)|\ 0 \leqslant x \leqslant d/2,\ 0 \leqslant \theta \leqslant \pi, x \leqslant \frac{l\sin\theta}{2}\}.$$

A is shown as the shaded area in Fig. 1.3(c). So the probability required is

$$P(A) = \frac{L(A)}{L(\Omega)} = \frac{\int_0^\pi \frac{l\sin\theta}{2}\mathrm{d}\theta}{\frac{d\pi}{2}} = \frac{2l}{\pi d}.$$

Buffon's needle problem is a question first posed in 1777 by Georges-Louis Leclerc, Comte de Buffon. The solution can be used to design a Monte Carlo method for approximating the number π. ■

People once thought that the definition of geometric probability is perfect. But there are some examples to show that probabilities may not be well defined. Let us see the following famous example.

Example 1.3.9 (**Bertrand paradox**) Consider an equilateral triangle inscribed in a circle. Suppose a chord of the circle is chosen at random. What is the probability that the chord is longer than a side of the triangle? Bertrand gave three arguments, all apparently valid, yet yielding different results.

(i) The "random midpoint" method: choose a point anywhere within the circle and construct a chord with the chosen point as its midpoint, see Fig. 1.4(a). The chord is longer than a side of the inscribed triangle if the chosen point falls within a concentric circle of radius 1/2 the radius of the larger circle. The area of the smaller circle is one fourth the area of the larger circle, therefore the probability a random chord is longer than a side of the inscribed triangle is 1/4.

(ii) The "random endpoints" method: choose two random points on the circumference of the circle and draw the chord joining them, see Fig. 1.4(b). To calculate the probability in question imagine the triangle rotated so its vertex coincides with one of the chord endpoints. Observe that if the other chord endpoint lies on the arc between the endpoints of the triangle side opposite the first point, the chord is longer than a side of the triangle. The length of the arc is one third of the circumference of the circle, therefore the probability that a random chord is longer than a side of the inscribed triangle is 1/3.

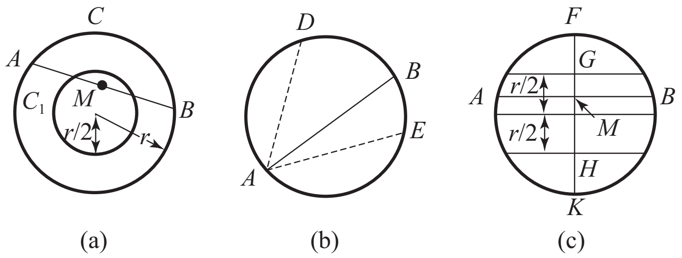

Figure 1.4: Bertrand paradox

(iii) The "random radius" method: choose a radius of the circle, choose a point on the radius and construct the chord through this point and perpendicular to the radius, see Fig. 1.4(c). To calculate the probability in question imagine the triangle rotated so a side is perpendicular to the radius. The chord is longer than a side of the triangle if the chosen point is nearer the center of the circle than the point where the side of the triangle intersects the radius. The side of the triangle bisects the radius, therefore the probability a random chord is longer than a side of the inscribed triangle is 1/2. ■

The problem's classical solution thus hinges on the method by which a chord is chosen "at random". It turns out that if, and only if, the method of random selection is specified, the problem could have a well-defined solution. There is no unique selection method, so there cannot be a unique solution. The above three solutions presented by Bertrand correspond to different selection methods, and in the absence of further information there is no reason to prefer one over another.

This and other paradoxes of the geometric interpretation of probability justified more stringent formulations, including frequency probability and subjectivist Bayesian probability.

1.3.3 The Frequency Interpretation of Probability

Suppose that some random procedure has several possible outcomes that are not necessarily equally likely. How can we define the probability $P(A)$ of any eventuality A of interest? For example, suppose that the procedure is the rolling of a die that is suspected to be weighted, or even clearly asymmetrical, in not being a perfect cube. What now is the probability of a six?

What we usually do is to roll the die a large number n of times, and let $f(n)$ be the number of sixes shown. Then (provided the rolls were made under similar conditions) the symmetry between the rolls suggests that (at least approximately)

$$P(\text{six}) = \frac{f(n)}{n} = \frac{\text{number of sixes}}{\text{number of rolls}}.$$

Furthermore, if you actually obtain an imperfect or weighted die and roll it many times, you will find that as n increases the ratio $f(n)/n$ always appears to be settling down around some asymptotic value. This provides further support for our taking $f(n)/n$ as an approximation to $P(\text{six})$.

In many problems, the probability that some specific outcome of a process will be obtained can be interpreted to mean the relative frequency with which that outcome would be obtained if the process were repeated a large number of times under similar conditions. Let E be an random experiment, A be an random event. Suppose that E was repeated n times under similar conditions. Let $f_n(A)$ be the times that A occurs. The ration

$$F_n(A) = \frac{f_n(A)}{n}$$

is said to be the ***frequency*** of event A in the n trials. If n is large enough, the probability of event A will be approximated by $F_n(A)$.

Theorem 1.3.3 *For a random experiment* E, *the frequency has the following properties:*

(i) *for every event* A, $F_n(A) \geqslant 0$,

(ii) $F_n(\Omega) = 1$,

(iii) *for every finite sequence of* n *disjoint events* $A_1, A_2, \cdots, A_n$,

$$F_n\big(\bigcup_{i=1}^{n} A_i\big) = \sum_{i=1}^{n} F_n(A_i)$$

We could demonstrate the frequency interpretation for the dice example. If we tossed the two dice 100 times, 200 times, 300 times, and so on, we would observe that the proportion of 6's would eventually settle down to the true probability of 0.139.

Thus, if we run Buffon's needle experiment a large number of times the proportion of crack crossings should be about the same as the probability of a crack crossing. More precisely, we will denote the number of crack crossings in the first n runs by f_n. In any case, if n is large, we should have

$$\pi \approx \frac{2ln}{f_n d} = \frac{2l}{F_n d}.$$

This is Buffon's famous estimate of π.

Table 1.2 are results of some experiments in this problem. Suppose that $d = 1$.

The frequency interpretation of probability is very important because the frequency can appropriately reflect the real probability and it is very simple to master. But there are some comments about it:

(i) there is no definite indication of an actual number that would be considered large enough;

(ii) this interpretation of probability rests on the important assumption that our process or experiment can be repeated many times under similar conditions. While the real-world experiment must not be completely controlled but must have some "random" features;

(iii) the probability of event A will be approximated by $F_n(A)$ do not mean $P(A)$ is the limit of $F_n(A)$;

(iv) the frequency interpretation of probability is that it applies only to a problem in which there can be, at least in principle, a large number of similar repetition of a certain process. Many important problems are not of this type. For example, the frequency interpretation of probability cannot be applied directly to the probability that a specific acquaintance will get married within the next two years or to the probability that a particular medical research project will lead to the development of a new treatment for a certain disease within a specified period of time.

Each interpretation of probability, as we can see, has its appeal and its difficulties. We need a pure mathematical definition for probability for general cases.

1.4 Probability Space

The foundations of the modern probability theory were laid by Andrey Nikolaevich Kolmogorov who combined the notion of sample space introduced by Richard von Mises and the measure theory and presented his axiom system for probability theory.

1.4.1 Axiomatic Definition of Probability

In general, we will not be concerned with the probability of individual outcomes of an experiment, but with collections of outcomes, called events. For instance, when we send a message across a channel, we will be interested in the number of errors in the received signal, rather than in the individual errors. Thus, probability will be defined as a function on sets. If sample space Ω is uncountable, then some of its subsets might be extremely ugly (really, really ugly), and we need suitable collections of sets to work with. These are called σ-algebras.

Definition 1.4.1 *A collection $\mathscr{F}$ of subsets of Ω is a σ-**algebra** if*

(i) $\Omega \in \mathscr{F}$;

(ii) $F \in \mathcal{F} \Longrightarrow \overline{F} \in \mathscr{F}$;

Experimentalist	Year	l	n	f_n	$\pi \approx \dfrac{2ln}{f_n}$
Wolf	1850	0.8	5000	2532	3.1596
Smith	1855	0.6	3204	1218.5	3.1554
De Morgan C.	1860	1.0	600	382.5	3.137
Fox	1884	0.75	1030	489	3.159
Lazzerini	1901	0.83	3408	1808	3.1415929
Reina	1925	0.5419	2520	859	3.179 5

Table 1.2: Buffon's needle experiment

(iii) *if F_n is a countable collection of sets, $n = 1, 2, \cdots$ such that $F_n \in S$ for all n, then* $\bigcup_n F_n \in \mathscr{F}$.

What is the intuition behind a σ-algebra? If Ω represents the collection of possible outcomes of an experiment, a subset of Ω is called an event. Then, a σ-algebra represents the collection of all possible, interesting events from the viewpoint of a given experiment.

Example 1.4.1 Let the experiment be a coin toss (where we blow on the coin if it stands up straight!). Then $\Omega = \{H, T\}$. Let $\mathscr{F}_1 = \{\emptyset, \{H\}, \{T\}, \{H, T\}\}$ and $\mathscr{F}_2 = \{\emptyset, \{H, T\}\}$. It is easy to prove that both $\mathscr{F}_1$ and $\mathscr{F}_2$ are σ-algebra on Ω.

Now we know that the reason that Bertrand Paradox occurs is σ-algebra on Ω is not uniquely determined.

Definition 1.4.2 *Let $P(A)(A \in \mathscr{F})$ be a non-negative set function on the σ-algebra $\mathscr{F}$. $P(A)$ is called the probability measure or* ***probability of event*** *A if it satisfies the following three axioms:*

Axiom 1. *for every $A \in \mathscr{F}, P(A) \geqslant 0$;*

Axiom 2. $P(\Omega) = 1$;

Axiom 3. (*countable additivity*) *for every infinite sequence of countable disjoint events $A_1, A_2, \cdots$,*

$$P\big(\bigcup_{i=1}^{\infty} A_i\big) = \sum_{i=1}^{\infty} P(A_i).$$

The sets in σ-algebra $\mathscr{F}$ are called events. $\mathscr{F}$ is called to be the algebra of events. The triple $(\Omega, \mathscr{F}, P)$ is a ***probability space*** *or* ***probability triple***.

You can find that the properties in Theorem 1.3.1, 1.3.2 and 1.3.3 are in accordance with the Axioms in above definition. In other words, this definition is the abstraction of three different interpretations about probability in section 1.3.

1.4.2 Properties of Probability

Now we give some properties about probability.

Theorem 1.4.1 $P(\emptyset) = 0$.

Proof. Consider the infinite sequence of events $A_1, A_2, \cdots$ such that $A_i = \emptyset$ for $i = 1, 2, \cdots$. In other words, each of the events in the sequence is just the empty set $\emptyset$. This sequence is a sequence of disjoint events, since $\emptyset \cap \emptyset = \emptyset$. Furthermore, $\cup_{i=1}^{\infty} A_i = \emptyset$. Therefore, it follows from Axiom 3 that

$$P(\emptyset) = P\big(\bigcup_{i=1}^{\infty} A_i\big) = \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^{\infty} P(\emptyset).$$

This equation states that when the number $P(\emptyset)$ is added repeated in an infinite series, the sum of that series is simply the number $P(\emptyset)$. The only real number with this property is zero. ■

Theorem 1.4.2 *For every finite sequence of countable disjoint events $A_1, A_2, \cdots, A_n$,*

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i).$$

Proof. Consider the infinite sequence of events $A_1, A_2, \cdots$, in which $A_1, A_2, \cdots, A_n$ are the n given disjoint events and $A_i = \emptyset$ for $i > n$. Then the events in this infinite sequence are disjoint and $\cup_{i=1}^{\infty} A_i = \cup_{i=1}^{n} A_i$. Therefore,

$$P(\bigcup_{i=1}^{n} A_i) = P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^{n} P(A_i) + \sum_{i=n+1}^{\infty} P(A_i) = \sum_{i=1}^{n} P(A_i) + 0 = \sum_{i=1}^{n} P(A_i)$$

holds by Axiom 3. ■

Theorem 1.4.3 *For every event A, $P(\overline{A}) = 1 - P(A)$.*

Proof. Since A and $\overline{A}$ are disjoint events and $A \cup \overline{A} = \Omega$, it follows from Theorem 1.4.2 that $P(\Omega) = P(A) + P(\overline{A})$. By Axiom 2, we know $P(\Omega) = 1$, then $P(\overline{A}) = 1 - P(A)$. ■

Theorem 1.4.4 *If $A \subset B$, then $P(B - A) = P(B) - P(A)$ and $P(A) \leqslant P(B)$.*

Proof. Since $A \subset B$, the event B can be treated as the union of the two disjoint events A and $B\overline{A}$. Therefore, $P(B) = P(A) + P(B\overline{A})$, i.e.,

$$P(B - A) = P(B\overline{A}) = P(B) - P(A).$$

Since $P(B - A) \geqslant 0$, $P(B) \geqslant P(A)$. ■

Theorem 1.4.5 *For every event A, $0 \leqslant P(A) \leqslant 1$.*

Proof. It is known from Axiom 1 that $P(A) \geqslant 0$. Since $A \subset \Omega$ for every event A, Theorem 1.3.4 implies $P(A) \leqslant P(\Omega) = 1$, by Axiom 2. ■

Theorem 1.4.6 *For every two events A and B,*

$$P(A \cup B) = P(A) + P(B) - P(AB). \tag{1.3}$$

Proof. The event $A \cup B$ may be treated as the union of the three events $A\overline{B}$, AB and $\overline{A}B$, see Fig. 1.5. Since the three events are disjoint, it follows from Theorem 1.4.2 that

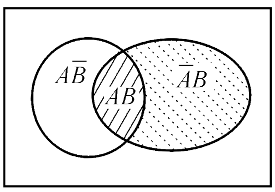

Figure 1.5: $A \cup B = A\overline{B} \cup AB \cup \overline{A}B$

$$P(A \cup B) = P(A\overline{B} \cup AB \cup \overline{A}B) = P(A\overline{B}) + P(AB) + P(\overline{A}B). \tag{1.4}$$

Furthermore, it can be seen that

$$P(A) = P(A\overline{B}) + P(AB)$$

and

$$P(B) = P(\overline{A}B) + P(AB).$$

We now use these last two equations to compute the quantity on the right-hand side of (1.3):

$$\begin{aligned} P(A) + P(B) - P(AB) &= P(A\overline{B}) + P(AB) + P(\overline{A}B) + P(AB) - P(AB) \\ &= P(A\overline{B}) + P(AB) + P(\overline{A}B), \end{aligned}$$

which is the right-hand side of (1.4). ■

Corollary 1.4.7 *For every three events A_1, A_2 and A_3,*

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) =& P(A_1) + P(A_2) + P(A_3) \\ & - [P(A_1A_2) + P(A_2A_3) + P(A_1A_3)] \\ & + P(A_1A_2A_3). \end{aligned} \tag{1.5}$$

Corollary 1.4.8 *For every n events $A_1, A_2, \cdots, A_n$,*

$$\begin{aligned} P\big(\bigcup_{i=1}^{n} A_i\big) =& \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_iA_j) + \sum_{i<j<k} P(A_iA_jA_k) \\ & - \sum_{i<j<k<l} P(A_iA_jA_kA_l) + \cdots + (-1)^{n+1} P(A_1A_2\cdots A_n). \end{aligned} \tag{1.6}$$

Proof. By Theorem 1.4.6, Corollary 1.4.8 holds for $n = 2$. Suppose that Corollary 1.4.8 holds for $n = k$. Then

$$\begin{aligned} P\big(\bigcup_{i=1}^{k+1} A_i\big) &= P\ \big(\ \big[\ \bigcup_{i=1}^{k} A_i\ \big] \cup A_{k+1}\ \big) = P\big(\bigcup_{i=1}^{k} A_i\big) + P(A_{k+1}) - P\big(\bigcup_{i=1}^{k} A_iA_{k+1}\big) \\ &= \sum_{i=1}^{k} P(A_i) - \sum_{i<j} P(A_iA_j) + \cdots + (-1)^{k+1} P(A_1A_2\cdots A_n) + P(A_{k+1}) - P\big(\bigcup_{i=1}^{k} B_i\big), \end{aligned}$$

where $B_i = A_iA_{k+1}, i = 1, 2, \cdots, k$. By applying equation (1.6) to $P(\bigcup_{i=1}^{k} B_i)$,

$$P\big(\bigcup_{i=1}^{k} B_i\big) = \sum_{i=1}^{k} P(A_iA_{k+1}) - \sum_{i<j} P(A_iA_jA_{k+1}) + \cdots + (-1)^{k+1} P(A_1A_2\cdots A_{k+1}).$$

Then the conclusion is right for $n = k + 1$. ■

Theorem 1.4.9 (i) *Let $A_1, A_2, \cdots$ be an infinite sequence of events such that $A_1 \subset A_2 \subset \cdots$. Then*

$$P\big(\bigcup_{i=1}^{\infty} A_i\big) = \lim_{n\to\infty} P(A_n). \tag{1.7}$$

(ii) *Let $A_1, A_2, \cdots$ be an infinite sequence of events such that $A_1 \supset A_2 \supset \cdots$. Then*

$$P\Big(\bigcap_{i=1}^{\infty} A_i\Big) = \lim_{n\to\infty} P(A_n). \tag{1.8}$$

Proof. (i) Let $A = \bigcup_{i=1}^{\infty} A_i$. Since $A_n \subset A_{n+1}$, we rewrite A_n by n disjoint events $A_1, A_2\overline{A_1}, \cdots, A_n\overline{A_{n-1}}$:

$$A_n = A_1 \cup A_2\overline{A_1} \cup \cdots \cup A_n\overline{A_{n-1}}.$$

Let $B_1 = A_1$, $B_m = A_m\overline{A_{m-1}}$, $m = 2, 3, \cdots$. Then

$$P(A) = P\Big(\bigcup_{m=1}^{\infty} B_m\Big) = \sum_{m=1}^{\infty} P(B_m) = \lim_{n\to\infty} \sum_{m=1}^{n} P(B_m).$$

By Theorem 1.4.2, $\sum_{m=1}^{n} P(B_m) = P(\bigcup_{m=1}^{n} B_m) = P(A_n)$. So $P\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \lim_{n\to\infty} P(A_n)$.

(ii) It can be obtained by defining $B_i = \overline{A_i}$. ■

Next, let us calculate some probabilities by applying the above properties.

Example 1.4.2 Assume A and B are two events such that $P(A) = 1/3$ and $P(B) = 1/2$. Determine the value of $P(B - A)$ satisfying each of the following conditions. (a) A and B are disjoint; (b) $A \subset B$; (c) $P(AB) = 1/8$.

Solution. Since $B = \Omega \cap B = (A \cup \overline{A})B = AB \cup \overline{A}B$, and by Theorem 1.4.2,

$$P(B) = P(AB) + P(\overline{A}B) = P(AB) + P(B - A),$$

That is,

$$P(B - A) = P(B) - P(AB). \tag{1.9}$$

(a) Since A and B are disjoint means $AB = \emptyset$, we know

$$P(B - A) = P(B) - P(AB) = P(B) = 1/2.$$

(b) If $A \subset B$, then $AB = A$. Thus $P(B - A) = P(B) - P(AB) = P(B) - P(A) = 1/6$.

(c) $P(B - A) = P(B) - P(AB) = 1/2 - 1/8 = 3/8$. ■

Example 1.4.3 Assume A and B are two events such that $P(B) = b$ and $P(A \cup B) = c$. Determine the value of $P(A\overline{B})$.

Solution. By equations (1.3) and (1.9),

$$P(A \cup B) = P(A) + P(B) - P(AB) = P(B) + [P(A) - P(AB)] = P(B) + P(A\overline{B}).$$

Thus $P(A\overline{B}) = P(A \cup B) - P(B) = c - b$. ■

Example 1.4.4 Select an integer from 1 to 1000 at random. Calculate the probability of the event that the integer is not divisible by 6 and 8.

Solution. Let A be the event that the integer is divisible by 6, and B be the event that the integer is divisible by 8. The required probability is

$$\begin{aligned} P(\overline{A}\ \overline{B}) &= P(\overline{A \cup B}) = 1 - P(A \cup B) \\ &= 1 - [P(A) + P(B) - P(AB)] \\ &= 1 - 166/1000 - 125/1000 + 41/1000 \approx 0.75. \end{aligned}$$

■

Example 1.4.5 (**The matching problem**) An absent-minded secretary prepares n letters and envelopes to send to n different people, but then randomly stuffs the letters into the envelopes. A match occurs if a letter is inserted in the proper envelope. Here we shall determine the probability of the event that at least one match occurs.

Solution. Let A_i be the event that letter i is placed in the correct envelop $(i = 1, 2, \cdots, n)$ (or a person gets his or her own hat) and we shall determine the value of $P(\bigcup_{i=1}^n A_i)$ by using equation (1.6). Since

$$\begin{aligned} P(A_i) &= \frac{(n-1)!}{n!} = \frac{1}{n}, \qquad i = 1, 2, \cdots; \\ P(A_iA_j) &= \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}, \qquad 1 \leqslant i < j \leqslant n; \\ P(A_iA_jA_k) &= \frac{(n-3)!}{n!} = \frac{1}{n(n-1)(n-2)}, \qquad 1 \leqslant i < j < k \leqslant n; \\ &\vdots \\ P(A_1 \cdots A_n) &= \frac{1}{n!}, \end{aligned}$$

we have

$$\begin{aligned} P(\bigcup_{i=1}^n A_i) &= \binom{n}{1} \cdot \frac{1}{n} - \binom{n}{2} \cdot \frac{1}{n(n-1)} + \cdots + (-1)^{n+1} \frac{1}{n!} \\ &= 1 - \frac{1}{2!} + \frac{1}{3!} + \cdots + (-1)^{n+1} \frac{1}{n!}. \end{aligned}$$

■

This probability has the following interesting features. As $n \to \infty$, the value of $P(\bigcup_{i=1}^n A_i)$ approaches the following limit:

$$1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \cdots.$$

It is shown in books on elementary calculus that the sum of the infinite series on the right side of this equation is $1 - (1/e)$, where $e = 2.71828\cdots$. Hence, $1 - (1 - 1/e) = 0.63212\cdots$. It follows that for large value of n, the probability that at least one letter will be placed in the correct envelope is approximately 0.63212.

The value of $P(\bigcup_{i=1}^{n} A_i)$ converges to the limit very rapidly. In fact, for $n = 7$ the exact value $P(\bigcup_{i=1}^{7} A_i)$ and the limiting value 0.63212 agree to four decimal places. Hence, regardless of whether seven letters are placed at random in seven or seven million letters are placed at random in seven million envelopes, the probability that at least one letter will be replaced in the correct envelope is 0.63212.

1.5 Conditional Probabilities

Suppose that an event B has occurred. How should we change the probabilities of the remaining events? We shall call the new probability for an event A the conditional probability of A given B and denote it by $P(A|B)$.

1.5.1 The Definition of Conditional Probability

Before we introduce the definition of conditional probability, let us see a probability puzzle, the Monty Hall problem. It loosely based on the American television game show "Let's Make a Deal" and named after its original host, Monty Hall. The problem was originally posed in a letter by Steve Selvin to the American Statistician in 1975. It became famous as a question from a reader's letter quoted in Marilyn Vos Savant's "Ask Marilyn" column in Parade magazine in 1990:

The Monty Hall problem: Suppose that you are on a game show, and you are given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what is behind the doors, opens another door, say No. 3, which has a goat. You are then given the option to switch your selection from Door 1 to the unopened Door 2. What is the probability that you will win the car if you switch your door selection to Door 2? Also, compute the probability that you will win the car if you do not switch.

What would you do?

Vos Savant's response was that the contestant should switch to the other door. But many readers of Vos Savant's column refused to believe that switching is beneficial.

The critical factor in this problem is that the host knows what is behind the doors in advance. The probability of the event would be changed after we could know some information. That is Conditional Probabilities.

In order to calculate the conditional probabilities, we consider a easier problem. Suppose that we flip two fair coins and that each of the 4 possible outcomes is equally likely to occur and hence has the probability 1/4. Suppose that we observe that a head appears on the flip of one coin. Then given this information, what is the probability that (H, H) occurs? To calculate this probability we reason as follows:

Given that a head appears, it follows that there can be three outcomes of our experiment, namely, $(H,H),(H,T)$ and (T,H). Since each of these outcomes originally had the same probability of occurring, they should still have equal probabilities. That is, given that a head appears on the flip of one coin, then the (conditional) probability of each of the outcomes $(H,H),(H,T),(T,H)$ is $1/3$ while the (conditional) probability of the outcome (T,T) in the sample space is 0. Hence, the desired probability will be $1/3$. In fact, this probability can be calculated as follows.

Let $A=\{(H,H)\}$ and $B=\{$ a head appears on the flip of one coin$\}$. Let $\#A$ be the number of all possible outcomes of event A.

$$P(A|B)=\frac{1}{3}=\frac{\#\{(H,H)\}}{\#\{(H,H),(H,T),(T,H)\}}=\frac{\#AB}{\#B}=\frac{\frac{\#AB}{\#S}}{\frac{\#B}{\#S}}=\frac{P(AB)}{P(B)}. \tag{1.11}$$

The above equation (1.11) helps us to define conditional probability.

Definition 1.5.1 *Given two events A and B with $P(B)>0$, the* ***conditional probability of*** *A* ***given*** *B is defined as the quotient of the joint probability of A and B, and the probability of B:*

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

Theorem 1.5.1 *If $P(B)>0$, then the conditional probability $P(A|B)$ is also a probability, that is,*

(i) *for every event $A, P(A|B)\geqslant 0$;*

(ii) $P(\Omega|B)=1$;

(iii) *for every infinite sequence of countable disjoint events $A_1,A_2,\cdots$,*

$$P\Big(\bigcup_{i=1}^{\infty}A_i\Big|B\Big)=\sum_{i=1}^{\infty}P(A_i|B).$$

Proof. (i) Since $P(B)\geqslant P(AB)$ and $P(B)>0$, we get $P(A|B)\geqslant 0$ by equation (1.12).

(ii) $P(\Omega|B)=\frac{P(\Omega B)}{P(B)}=\frac{P(B)}{P(B)}=1$.

(iii) $P\Big(\bigcup\limits_{i=1}^{\infty}A_i\Big|B\Big)=\Big(P\big(\bigcup\limits_{i=1}^{\infty}A_iB\big)\Big)\Big/P(B)=\sum\limits_{i=1}^{\infty}\frac{P(A_iB)}{P(B)}=\sum\limits_{i=1}^{\infty}P(A_i|B)$. ■

Since the conditional probability is a probability, all properties of probabilities hold for conditional probabilities. We do not prove those properties here, but just list them as follows.

Proposition 1.5.2 *If $P(B)>0$, then*

(i) $P(\emptyset|B)=0$,

(ii) *for every finite sequence of countable disjoint events $A_1,A_2,\cdots,A_n$,*

$$P\Big(\bigcup_{i=1}^{n}A_i\Big|B\Big)=\sum_{i=1}^{n}P(A_i|B),$$

(iii) $P(\overline{A}|B)=1-P(A|B)$,

(iv) *if $A\subset C$, then $P(C-A|B)=P(C|B)-P(A|B)$ and $P(A|B)\leqslant P(C|B)$,*

(v) $P(A\cup C|B)=P(A|B)+P(C|B)-P(AC|B)$.

Example 1.5.1 The probability that it is Friday and that a student is absent is 0.03. Since there are 5 school days in a week, the probability that is Friday is 0.2. What is the probability that a student is absent given that today is Friday?

Solution. $P(\text{Absent}|\text{Friday}) = \dfrac{P(\text{Friday and Absent})}{P(\text{Friday})} = \dfrac{0.03}{0.2} = 0.15 = 15\%.$ ∎

Example 1.5.2 Your neighbor has 2 children. You learn that he has a son, Joe. What is the probability that Joe's sibling is a brother?

Solution. The obvious answer that Joe's sibling is equally likely to have been born male or female suggests that the probability the other child is a boy is 1/2. This is not correct!

Consider the experiment of selecting a random family having two children and recording whether they are boys or girls. Then, the sample space is $\Omega = \{BB, BG, GB, GG\}$, where, e.g., outcome "$BG$" means that the first-born child is a boy and the second-born is a girl. Assuming boys and girls are equally likely to be born, the 4 elements of Ω are equally likely.

The event, C, that the neighbor has a son is the set $C = \{BB, BG, GB\}$. The event, D, that the neighbor has two boys (i.e., Joe has a brother) is the set $D = \{BB\}$. The desired probability is

$$P(D|C) = \frac{P(DC)}{P(C)} = \frac{P(\{BB\})}{P(\{BB, BG, GB\})} = 1/3.$$

∎

Example 1.5.3 A machine produces parts that are either good (90%), slightly defective (2%), or obviously defective (8%). Produced parts get passed through an automatic inspection machine, which is able to detect any part that is obviously defective and discard it. What is the quality of the parts that make it through the inspection machine and get shipped?

Solution. Let A (resp., B, C) be the event that a randomly chosen shipped part is good (resp., slightly defective, obviously defective). We are told that $P(A) = 0.90, P(B) = 0.02$, and $P(C) = 0.08$.

We want to compute the probability that a part is good given that it passed the inspection machine (i.e., it is not obviously defective), which is

$$P(A|\overline{C}) = \frac{P(A\overline{C})}{P(\overline{C})} = \frac{P(A)}{1 - P(C)} = \frac{0.90}{1 - 0.08} = 90/92 = 0.978.$$

∎

Example 1.5.4 Ten fair dice are rolled at one time. What is the conditional probability of the event at least two land on 1 given the event at least one of the dice lands on 1.

Solution. Let A be the event that at least one of the dice lands on 1, B be the event that at least two land on 1, and C be the event that exactly one of the dice lands on 1. We see that $B \subset A, C \subset A$ and $B = A - C$. By using the definition of classical probability,

$$P(A) = 1 - P(\overline{A}) = 1 - \frac{5^{10}}{6^{10}}, \qquad P(C) = \frac{10 \times 5^9}{6^{10}}.$$

So the required probability is

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{P(B)}{P(A)} = \frac{P(A) - P(C)}{P(A)} = \frac{1 - \dfrac{5^{10}}{6^{10}} - \dfrac{10 \times 5^9}{6^{10}}}{1 - \dfrac{5^{10}}{6^{10}}} \approx 0.615.$$

■

The following example is an application of conditional probability in actuarial science.

Example 1.5.5 In describing the survival rate and life expectancy in a certain population, let A_N denote the event of a new-born to reach the age of N years. We are given that

$$P(A_{50}) = 0.913, \qquad P(A_{55}) = 0.881, \qquad P(A_{65}) = 0.746.$$

(a) What is the probability of a 50 years old man to reach the age of 55, i.e., what is $P(A_{55}|A_{50})$?

(b) If the probability that a man who just turned 65 will die within 5 years is 0.16, what is the probability for a man to survive till his 70th birthday, i.e., what is $P(A_{70})$?

Solution. (a) Obviously, $A_{55} \cap A_{50} = A_{55}$. We have by definition,

$$P(A_{55}|A_{50}) = P(A_{55} \cap A_{50})/P(A_{50}) = P(A_{55})/P(A_{50}) \approx 0.965.$$

(b) Similarly, $P(A_{70}|A_{65}) = P(A_{70})/P(A_{65})$. So $P(A_{70}) = P(A_{65}) \cdot P(A_{70}|A_{65})$. By Proposition 1.5.2(iii) we get

$$P(A_{70}|A_{65}) = 1 - 0.16 = 0.84.$$

Therefore, $P(A_{70}) = P(A_{65}) \cdot P(A_{70}|A_{65}) = 0.746 \cdot 0.84 \approx 0.627.$ ■

Remark: The probability $P(A_N|A_M)$ is usually said to be survival probability in actuarial science. It is basic and very important to calculate annuity, premium, liability reserve and so on.

1.5.2 The Multiplication Rule

The additivity of probability in Section 1.4 helped us to calculate the probabilities of the events that at least one of the events occurs. The multiplication rule (also known as the "Law of Multiplication") we will introduce in this section also deals with many events, but it is usually used to find the probabilities of successive events (rolling one die then another, spinning a spinner twice, pulling three marbles out of a bag, etc). The formula to calculate $P(A_{70})$ in Example 1.5.5 is an application of the "Multiplication Rule". In a sense, the multiplication rule could be regarded as a perfect mathematical reflection of the Chinese phrase "proceed in an orderly way and step by step".

Theorem 1.5.3 *Assume that* $P(B) > 0$. *Then*

$$P(AB) = P(B) \cdot P(A|B). \tag{1.13}$$

Or if $P(A) > 0$, *then*

$$P(AB) = P(A) \cdot P(B|A). \tag{1.14}$$

Equations (1.13) and (1.14) are (trivially) multiplication rule for two events, which are just rewriting of the definition of conditional probability. We generalize the multiplication rule to n events as follows.

Theorem 1.5.4 *Suppose that* $A_1, A_2, \cdots, A_n$ *are events satisfying* $P(A_1A_2\cdots A_{n-1}) > 0$. *Then*

$$P(A_1A_2\cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1}). \tag{1.15}$$

Proof. The product of probabilities on the right side of equation (1.15) is equal to

$$P(A_1)\cdot\frac{P(A_1A_2)}{P(A_1)}\cdot\frac{P(A_1A_2A_3)}{P(A_1A_2)}\cdots\frac{P(A_1A_2\cdots A_n)}{P(A_1A_2\cdots A_{n-1})}.$$

Since $P(A_1A_2\cdots A_{n-1}) > 0$, each of the denominators in this product must be positive. All terms in the product cancel each other except the final numerator $P(A_1A_2\cdots A_n)$, which is the left side of equation (1.15). ■

In Theorem 1.5.4, the events A_i occur according to the index i. Theoretically speaking, the multiplication rule holds no matter which one of these events happens first. But in actual applications, we usually use the multiplication rule in the situation that the events occur successively.

Example 1.5.6 Suppose that five good fuses and two defective ones have been mixed up. To find the defective fuses, we test them one-by-one, at random and without replacement. What is the probability that we find both of the defective fuses in the first two tests?

Solution. Let D_1 (resp., D_2) be the event that we find a defective fuse in the first (resp., second) test. Then

$$P(D_1D_2) = P(D_1)P(D_2|D_1) = \frac{2}{7}\cdot\frac{1}{6} = \frac{1}{21}.$$

■

Example 1.5.7 If six cards are selected at random (without replacement) from a standard deck of 52 cards, what is the probability there will be no pairs? (two cards of the same denomination)

Solution. Let A_i be the event that the first i cards have no pair among them, $i = 1, 2, \cdots, 6$.

Of course, one can solve the problem directly using counting techniques: define the sample space to be (equally likely) ordered sequences of 6 cards; then, $\#\Omega = 52\cdot 51\cdot 50\cdots 47$, and the event A_6 has $52\cdot 48\cdot 44\cdots 32$ elements. So

$$P(A_6) = \frac{\#A}{\#\Omega} = \frac{52\cdot 48\cdot 44\cdots 32}{52\cdot 51\cdot 50\cdots 47}.$$

Here we use the multiplication rule. We want to compute $P(A_6)$, which is actually the same as $P(A_1A_2\cdots A_6)$, since $A_6 \subset A_5 \cdots \subset A_1$, implying that $A_1A_2A_3A_4A_5A_6 = A_6$. We get

$$P(A_1A_2\cdots A_6) = P(A_1)\cdot\frac{P(A_1A_2)}{P(A_1)}\cdots\frac{P(A_1A_2\cdots A_6)}{P(A_1A_2\cdots A_5)} = \frac{52}{52}\frac{48}{51}\frac{44}{50}\frac{40}{49}\frac{36}{48}\frac{32}{47}.$$

■

Example 1.5.8 (**Polya urn mode**) Suppose that an urn contains r red balls and b blue balls ($r \geqslant 2, b \geqslant 2$). Suppose that one ball is drawn randomly from the urn and its color is observed; it is then replaced in the urn, and c additional balls of the same color is added to the urn, and the selection process is repeated four times. We shall determine the probability of obtaining the sequence of outcomes red, blue, red, blue.

Solution. Let R_j denote the event that a red ball is obtained on the jth draw and B_j the event that a blue ball is obtained on the jth draw ($j = 1, 2, 3, 4$). Then

$$\begin{aligned} P(R_1B_2R_3B_4) &= P(R_1)P(B_2|R_1)P(R_3|R_1B_2)P(B_4|R_1B_2R_3) \\ &= \frac{r}{r+b}\cdot\frac{b}{r+b+c}\cdot\frac{r+c}{r+b+2c}\cdot\frac{b+c}{r+b+3c}. \end{aligned}$$

■

Since we have known that conditional probabilities behave just like probabilities, we can also have the following multiplication rule of conditional version.

Theorem 1.5.5 *Suppose that* $A_1, A_2, \cdots, A_n, B$ *are events such that* $P(A_1A_2\cdots A_{n-1}|B) > 0$. *Then*

$$P(A_1A_2\cdots A_n|B) = P(A_1|B)P(A_2|A_1B)P(A_3|A_1A_2B)\cdots P(A_n|A_1A_2\cdots A_{n-1}B).$$

1.5.3 Total Probability Formula

In daily life we often encounter an event caused by different events. Assume that A is the event caused by $B_1, B_2, \cdots, B_n$. The "Law of Total Probability" allows us to compute the probability of an event A by conditioning on cases, according to a partition $B_1, B_2, \cdots, B_n$ of the sample space.

Definition 1.5.2 *Let* Ω *denote the sample space of some experiment.* n *events* $B_1, B_2, \cdots, B_n$ *are said to form a* ***partition*** *of* Ω *if these events satisfy:*

(i) $B_1, B_2, \cdots, B_n$ *are disjoint and*

(ii) $\bigcup_{i=1}^{n} B_i = \Omega$.

For example, one way to partition Ω is to break it into sets B and $\overline{B}$, for any event B. Thus for event A, we have

$$\begin{aligned} P(A) &= P(A\Omega) = P(A\cap(B\cup\overline{B})) = P(AB\cup A\overline{B}) \\ &= P(AB) + P(A\overline{B}) = P(B)\cdot P(A|B) + P(\overline{B})\cdot P(A|\overline{B}) \end{aligned} \tag{1.16}$$

It is the simplest form of the law of total probability. More generally,

Theorem 1.5.6 *Suppose that the events $B_1, B_2, \cdots, B_n$ form a partition of the sample space Ω and $P(B_i) > 0$ for $i = 1, 2, \cdots, n$. Then, for every event A in Ω,*

$$P(A) = \sum_{i=1}^{n} P(B_i)P(A|B_i). \tag{1.17}$$

We can think of the event A as a result and the partition $B_1, B_2, \cdots, B_n$ as causes of the result A, see Fig. 1.6. The total probability formula can be thought of as a weighted average of $P(A|B_i)$, and the weights are $P(B_i)$, $i = 1, 2, \cdots$.

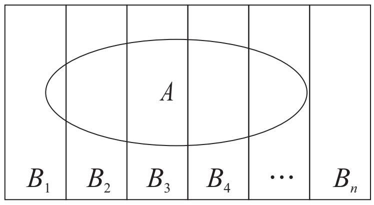

Figure 1.6: Total probability problem

Example 1.5.9 Two cards from an ordinary deck of 52 cards are missing. What is the probability that a random card drawn from this deck is a spade?

Solution. Let A be the event that the randomly drawn card is a spade. Let B_i be the event that i spades are missing from the 50-card (defective) deck, for $i = 0, 1, 2$.

By conditioning on how many spades are missing from the original (good) deck, we get

$$\begin{aligned} P(A) &= P(B_0)P(A|B_0) + P(B_1)P(A|B_1) + P(B_2)P(A|B_2) \\ &= \frac{13}{50}\frac{\binom{13}{0}\binom{39}{2}}{\binom{52}{2}} + \frac{12}{50}\frac{\binom{13}{1}\binom{39}{1}}{\binom{52}{2}} + \frac{11}{50}\frac{\binom{13}{2}\binom{39}{0}}{\binom{52}{2}} \approx \frac{1}{4}. \end{aligned}$$

■

It makes sense that the probability is simply 1/4, since each card is equally likely to be among the lost cards, why should the chance of getting a spade be changed?

Example 1.5.10 Your neighbor has 2 children. He picks one of them at random and comes by your house; he brings a boy named Joe (his son). What is the probability that Joe's sibling is a brother?

Solution. This example differs from the above Example 1.5.1: there is a mechanism by which the son was selected that gave you the information that your neighbor has a boy (the mechanism was random selection).

Think of it this way: in Example 1.5.1, we were given the event C that your neighbor has a son. Now, we are given the event C' that your neighbor randomly chose one of his 2 children, and that chosen child is a son. We note that $C' \subseteq C$ since event C' happening implies that event

C happens (if he chooses a son at random, you know for sure that he has a son!). It does not go the other way, though: C' does not imply C (just because he has a son does not mean that he chose that son at random).

We want to compute $P(D|C') = \dfrac{P(DC')}{P(C')}$, where $P(DC') = P(\{BB\}) = 1/4$ and

$$\begin{aligned} P(C') &= P(\{BB\})P(C'|\{BB\}) + P(\{BG\})P(C'|\{BG\}) \\ &\quad + P(\{GB\})P(C'|\{GB\}) + P(\{GG\})P(C'|\{GG\}) \\ &= \frac{1}{4}\cdot 1 + \frac{1}{4}\cdot\frac{1}{2} + \frac{1}{4}\cdot\frac{1}{2} + \frac{1}{4}\cdot 0 = 1/2. \end{aligned}$$

So $P(D|C') = \dfrac{P(DC')}{P(C')} = \dfrac{1/4}{1/2} = 1/2.$ ■

How to ask a sensitive question? Statisticians are sometimes confronted with how to obtain information on sensitive issues. What proportion of people use illegal drugs? How many students ever cheated on an exam? Surveying people directly and asking these types of sensitive questions are not likely to get honest responses and useful data.

Using probabilistic methods, statisticians have developed interesting ways, total probability methods, to ask sensitive questions that protect confidentiality. Here is an example.

Example 1.5.11 Respondents are given a coin and told to flip it in private, not letting anyone see the outcome. If it lands heads, they answer the sensitive question of interest (e.g. "Have you ever taken illegal drugs?"). If tails, they answer an innocuous question such as "Were you born in the first half of the year?" The respondent reports a yes or no, but does not say which question they actually answered. From a sample of such yes-no responses, how can statisticians estimate the parameter of interest, such as the proportion of people who have ever taken illegal drugs?

Solution. Let Y and N denote responses of yes and no, respectively. Let A denote the sensitive question and B the innocuous question. The unknown parameter that surveyors want to estimate is $p = P(Y|A)$, the probability that someone answers yes given that they were asked the sensitive question. We assume that the innocuous question is (i) easy to answer and (ii) has a known probability of yes and no, in this case 0.5 each.

Consider the unconditional probability $P(Y)$. By the law of total probability,

$$\begin{aligned} P(Y) &= P(A)P(Y|A) + P(B)P(Y|B) \\ &= \frac{1}{2}\cdot p + \frac{1}{2}\cdot\frac{1}{2} \\ &= \frac{p}{2} + \frac{1}{4}. \end{aligned}$$

When this survey is given to n people, the finial data will consist of n yes's and no's. The proportion of yes's is a simulated estimate of the unknown $P(Y)$. And thus

$$\frac{p}{2} + \frac{1}{4} = P(Y) \approx \frac{\text{Numbers of yes's in the sample}}{n}.$$

Solving for p gives

$$p \approx 2\left(\frac{\text{Numbers of yes's in the sample}}{n} - \frac{1}{4}\right),$$

which is the final estimate of the parameter of interest. ■

1.5.4 Bayes' Theorem

Suppose that someone told you they had a nice conversation with someone on the train. Not knowing anything else about this conversation, the probability that they were speaking to a woman is 50%. Now suppose that they also told you that this person had long hair. It is now more likely they were speaking to a woman, since women are more likely to have long hair than men. The following result, which is known as *Bayes' theorem*, can be used to calculate the probability that the person is a woman. In fact, Bayes' theorem provides a simple rule for computing the conditional probability of each partition event B_i given the event A that we have learned. It requires that we know the conditional probability of A given each event B_i and the unconditional probability of each B_i.

Theorem 1.5.7 (***Bayes' Theorem***) *Let the events $B_1, B_2, \cdots, B_n$ form a partition of the sample space Ω such that $P(B_i) > 0$ for $i = 1, 2, \cdots, n$, and let A be an event such that $P(A) > 0$. Then, for $i = 1, 2, \cdots, n$,*

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{n} P(B_j)P(A|B_j)} \tag{1.18}$$

Proof. By the definition of conditional probability, $P(B_i|A) = \dfrac{P(B_iA)}{P(A)}$. By the multiplication rule for conditional probabilities, equation (1.13), $P(B_iA) = P(B_i)P(A|B_i)$. According to Theorem 1.5.6, $P(A) = \sum_{j=1}^{n} P(B_j)P(A|B_j)$. The equation 1.18 holds. ■

We give some applications in the following examples.

Solution to The Monty Hall Problem. Suppose that $\Omega = \{1, 2, 3\}$, where outcome i means that the car is behind door i. Let $A = \{$ Host open Door 3 $\} = \{1, 2\}$ and $B = \{$ You choose Door 1$\}$. The probability that you win by switching to Door 2 given that he tells you Door 3 has a goat is:

$$\begin{aligned} P(\{2\}|AB) &= \frac{P(AB \cap \{2\})}{P(AB)} = \frac{P(\{2\})P(AB|\{2\})}{P(AB)} \\ &= \frac{P(\{2\})P(AB|\{2\})}{P(\{1\})P(AB|\{1\}) + P(\{2\})P(AB|\{2\}) + P(\{3\})P(AB|\{3\})} \\ &= \frac{\frac{1}{3}\cdot 1}{\frac{1}{3}\cdot\frac{1}{2} + \frac{1}{3}\cdot 1 + \frac{1}{3}\cdot 0} = \frac{2}{3}. \end{aligned}$$

While $P(\{1\}|AB) = 1 - P(\{2\}|AB) = 1/3$. □

Example 1.5.12 A new test has been devised for detecting a particular type of cancer. If the test is applied to a person who has this type of cancer, the probability that the person will have a positive reaction is 0.95 and the probability that the person will have a negative reaction is 0.05. If the test is applied to a person who does not have this type of cancer, the probability that the person will have a positive reaction is 0.05 and the probability that the person will have a negative reaction is 0.95. Suppose that in the general population, one person out of every 100,000 people has this type of cancer. If a person selected at random has a positive reaction to the test, what is the probability that he has this type of cancer?

Solution. Let A denote the event that a person has a positive reaction to the test, B_1 denote the event that he has this type of cancer and B_2 denote the event that he does not have this type of cancer. Now we have known that $P(A|B_1) = 0.95, P(\overline{A}|B_1) = 0.05, P(A|B_2) = 0.05, P(\overline{A}|B_2) = 0.95$ and $P(B_1) = 1/100000$. It now follows from Bayes' theorem that

$$\begin{aligned} P(B_1|A) &= \frac{P(B_1)P(A|B_1)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2)} \\ &= \frac{(1/100000) \cdot (0.95)}{(1/100000) \cdot (0.95) + (99999/100000) \cdot (0.05)} \approx 0.00019. \end{aligned}$$

■

It seems impossible! The result is contrary to our common sense. So in practical applications, doctors generally increased the probability $P(B_1)$ through some other medical examinations in advance to increase the accuracy of the test. For example, if $P(B_1) = 0.2$, then $P(B_1|A) \approx 0.83$. This example illustrates not only the use of Bayes' theorem, but also the importance of taking into account all of the information available in a problem.

Example 1.5.13 Three different machines M_1, M_2 and M_3 were used for producing a large batch of similar manufactured items. Suppose that 20 percent of the items were produced by machine M_1, 30 percent of the items were produced by machine M_2, and 50 percent of the items were produced by machine M_3. Suppose further that 1 percent of the items produced by machine M_1 are defective, that 2 percent of the items produced by machine M_2 are defective, and that 3 percent of the items produced by machine M_3 are defective. Finally, suppose that one item is selected at random from the entire batch, and it is found to be defective. We shall determine the probability that this item was produced by machine $M_i (i = 1, 2, 3)$.

Solution. Let B_i be the event that the selected item was produced by machine $M_i (i = 1, 2, 3)$, and let A be the event that the selected item is defective. We must evaluate the conditional probability $P(B_2|A)$. The probability $P(B_i)(i = 1, 2, 3)$ is as follows:

$$P(B_1) = 0.2, \qquad P(B_2) = 0.3, \qquad P(B_3) = 0.5.$$

Furthermore, the probability $P(A|B_i)$ that an item produced by machine M_i will be defective is: $P(A|B_1) = 0.01, P(A|B_2) = 0.02, P(A|B_3) = 0.03$. It now follows from Bayes' theorem that

$$\begin{aligned} P(B_1|A) &= \frac{P(B_1)P(A|B_1)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3)} \\ &= \frac{(0.2)(0.01)}{(0.2)(0.01) + (0.3)(0.02) + (0.5)(0.03)} = 0.087. \end{aligned}$$

By the similar way, we obtain $P(B_2|A) = 0.261, P(B_3|A) = 0.652$. ■

Remark 1.5.1 (i) *In Example 1.5.14, a probability like $P(B_i)$ is called the* ***prior probability*** *that the selected item will have been produced by machine M_i, because $P(B_i)$ is the probability of this event before the item is selected and before it is known whether the selected item is defective or nondefective. A probability like $P(B_i|A)$ is called the* ***posterior probability*** *that the selected item was produced by machine M_i, because it is the probability of this event after it is known that the selected item is defective.*

(ii) *From Example 1.5.14, we can observe that*

$$P(B_1) = 0.2 > P(B_1|A) = 0.087; \quad P(B_3) = 0.5 < P(B_3|A) = 0.652.$$

That means there is no stationary numerical relationship between prior probability and posterior probability.

In the end, let us see an interesting application of Bayes formula in detection.

Example 1.5.14 (**Witness reliability**) After a robbery the thief jumped into a taxi and disappeared. An eyewitness on the crime scene is telling the police that the cab is yellow. Now we have known the following information:

(i) In that particular city 80% of taxis are black and 20% of taxis are yellow.

(ii) Eyewitness are not always reliable and from past experience it is expected that an eyewitness is 80% accurate. He will identify the color of a taxi accurately (yellow or black) 8 out 10 times.

Equipped with this information, can we make sure eyewitness' testimony is worth something?

Solution. Let

$\{true = Y\}$ be the event that the color of the taxi was actually yellow and

$\{true = B\}$ the event that it was black. Let

$\{report = Y\}$ be the event that the eyewitness identified the color of the taxi as actually yellow and

$\{report = B\}$ the event that it was reported as black.

The goal is to compute $P(true = Y|report = Y)$. By using Bayes' formula,

$$P(true = Y|report = Y) = \frac{P(true = Y)P(report = Y|true = Y)}{P(report = Y)},$$

where $P(true = Y) = 0.2$ (this is the prior probability) and $P(report = Y|true = Y) = 0.8$ (this is the accuracy of witness testimony). Next to compute $P(report = Y)$, we need the actual color of the taxi. By using the law of total probability,

$$\begin{aligned} P(report = Y) &= P(true = Y)P(report = Y|true = Y) \\ &\quad + P(true = B)P(report = Y|true = B) \\ &= 0.2 \times 0.8 + 0.8 \times 0.2 = 0.32. \end{aligned}$$

Putting everything together we find

$$P(true = Y|report = Y) = \frac{P(true = Y)P(report = Y|true = Y)}{P(report = Y)} = \frac{0.16}{0.32} = \frac{1}{2}$$

and so the eyewitness' testimony does not provide much useful certainty. ■

Bayesians VS Frequentists

During the developing history of statistics, two major schools of thought emerged along the way and have been locked in an on-going struggle in trying to determine which one has the correct view on probability. These two schools are known as the Bayesian and Frequentist schools of thought. Both the Bayesians and the Frequentists hold a different philosophical view on what defines probability. Below are some fundamental differences between the Bayesian and Frequentist schools of thought:

Bayesians

(i) Probability is subjective and can be applied to single events based on degree of confidence or beliefs. For example, Bayesian can refer to tomorrow's weather as having 50% of rain, whereas this would not make sense to a Frequentist because tomorrow is just one unique event, and cannot be referred to as a relative frequency in a large number of trials.

(ii) Parameters are random variables that has a given distribution, and other probability statements can be made about them.

(iii) Probability has a distribution over the parameters, and point estimates are usually done by either taking the mode or the mean of the distribution.

Frequentists

(i) Probability is objective and refers to the limit of an event's relative frequency in a large number of trials. For example, a coin with a 50% probability of heads will turn up heads 50% of the time.

(ii) Parameters are all fixed and unknown constants.

(iii) Any statistical process only has interpretations based on limited frequencies. For example, a 95% C.I. of a given parameter will contain the true value of the parameter 95% of the time.

1.6 Independent of Events

Life is full of random events! You need to get a "feel" for them to be a smart and successful person. The toss of a coin, throwing dice and lottery draws are all examples of random events. Sometimes an event can affect the next event. We call those Dependent Events, because what happens depends on what happened before. Independent Events are not affected by previous events. This is an important idea! Independence is frequently invoked as a modeling assumption, and moreover, (classical) probability itself is based on the idea of independent replications of the experiment. In this section, we will discuss independence, a fundamental concept in probability theory.

1.6.1 Independence of Two Events

Consider the following examples:

(i) Landing on heads after tossing a coin AND rolling a 5 on a single 6-sided die.

(ii) Choosing a marble from a jar AND landing on heads after tossing a coin.

(iii) Choosing a 3 from a deck of cards, replacing it, AND then choosing an ace as the second card.

(iv) Rolling a 4 on a single 6-sided die, AND then rolling a 1 on a second roll of the die.

The list of examples could be extended indefinitely. In each case, we should expect to model the events as independent in some way. How should we incorporate the concept in our developing model of probability?

We take our clue from the examples above. Pairs of events are considered. The "operational independence" described indicates that knowledge that one of the events has occurred does not affect the likelihood that the other will occur. For a pair of events A and B, this is the condition

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

Occurrence of the event A is not conditioned by occurrence of the event B. If the occurrence of B does not affect the likelihood of the occurrence of A, we should be inclined to think of the events A and B as being independent in a probability sense. We take our clue from the condition $P(A|B) = P(A)$ to define the independence of two events.

Definition 1.6.1 *Two events A and B are **independent** if*

$$P(AB) = P(A)P(B). \tag{1.19}$$

It is obvious that if $P(A) > 0$ and $P(B) > 0$, then independence is equivalent to the statement that the conditional probability of one event given the other is the same as the unconditional probability of the event:

$$P(A|B) = P(A) \Longleftrightarrow P(B|A) = P(B) \Longleftrightarrow P(AB) = P(A)P(B).$$

The independence of the events in the examples at the beginning of this section are obvious. But things do not always like those.

Example 1.6.1 Let three fair coins be tossed. Let $A = \{$ all heads or all tails $\}$, $B = \{$ at least two heads $\}$, and $C = \{$ at most two tails $\}$. Of the pairs of events, (A, B), (A, C), and (B, C), which are independent and which are dependent?

Solution. The events can be written explicitly:

$$\begin{aligned} A =& \{HHH, TTT\}, \\ B =& \{HHH, HHT, HTH, THH\}, \\ C =& \{HHH, HHT, HTH, THH, HTT, THT, TTH\}. \end{aligned}$$

Then

$$\begin{aligned} P(AB) &= 1/8 = (2/8)(4/8) = P(A) \cdot P(B), \\ P(AC) &= 1/8 \neq (2/8)(7/8) = P(A) \cdot P(C), \\ P(BC) &= 4/8 \neq (4/8)(7/8) = P(B) \cdot P(C). \end{aligned}$$

Thus A and B are independent. A and C are dependent. B and C are dependent. ■

Example 1.6.2 A plant gets two independent genes for flower color, one from each parent plant. If the genes are identical, then the flowers are uniformly of that color; if they are different, then the flowers are striped in those two colors. The genes for the colors pink, crimson, and red occur in the population in the proportions $p : q : r$, where $p + q + r = 1$. A given plant's parents are selected at random; let A be the event that its flowers are at least partly pink, and let B be the event that its flowers are striped.

(a) Find $P(A)$ and $P(B)$.

(b) Show that A and B are independent if $p = 2/3$ and $r = q = 1/6$.

(c) Are these the only values of p, q, and r such that A and B are independent?

Solution. (a) With an obvious notation (P for pink, C for crimson, and R for red), we have

$$P(PP) = P(P)P(P) = p^2, \qquad \text{(by parents independence)},$$

because P occurs with probability p. Likewise,

$$P(PR) = P(R)P(P) = rp = P(RP).$$

Hence,

$$\begin{aligned} P(A) &= P(PP \cup PR \cup PR \cup PC \cup CP) \\ &= p^2 + 2pr + 2pq \\ &= 1 - (1-p)^2, \end{aligned}$$

because $p + q + r = 1$. (Can you see how to get this last expression directly?)

Similarly,

$$P(B) = P(PC \cup CP \cup PR \cup RP \cup RC \cup CR) = 2(pq + qr + rp).$$

(b) The events A and B are independent, if and only if,

$$P(A)P(B) = P(AB) = P(PC \cup CP \cup PR \cup RP) = 2(pq + pr).$$

From part (a), this is equivalent to

$$(1 - (1 - p)^2)(pq + qr + pr) = p(q + r), \tag{1.20}$$

and this is satisfied by the given values of p, q, and r.

(c) No. Rearranging equation (1.20), we see that A and B are independent for any values of q and r lying on the curve $rq = 2rq(q + r) + r^3 + q^3$, in the $r - q$ plane. You may care to amuse yourself by showing that this is a loop from the origin. Outside the loop, A and B are attractive; inside the loop, A and B are repellent. ■

The terms independent and disjoint sound vaguely similar but they are actually very different. In fact, we have following

Theorem 1.6.1 *Suppose that A and B are disjoint events for an experiment, each with positive probability. Then A and B are dependent.*

Proof. Since A and B are disjoint events, $P(AB) = P(\emptyset) = 0$. While $P(A)P(B) > 0$ because $P(A) > 0, P(B) > 0$. So $P(AB) \neq P(A)P(B)$. ■

Remark 1.6.1 *Note that disjointness is purely a set-theory concept while independence is a probability (measure-theoretic) concept. Indeed, two events can be independent relative to one probability measure and dependent relative to another. But most importantly, two disjoint events can never be independent, except in the trivial case that one of the events is null.*

If A and B are independent events in an experiment, it seems clear that any event that can be constructed from A should be independent of any event that can be constructed from B. This is the case, as the next theorem shows. Moreover, this basic idea is essential for the generalization of independence that we will consider shortly.

Theorem 1.6.2 *If A and B are independent events in an experiment, then each of the following pairs of events is independent:* (i) $\overline{A}$ *and* B*;* (ii) A *and* $\overline{B}$*;* (iii) $\overline{A}$ *and* $\overline{B}$*.*

Proof. We prove case (i). Suppose that A and B are independent. Then by the difference rule, i.e., equation (1.9),

$$P(\overline{A}B) = P(B) - P(AB) = P(B) - P(A)P(B) = [1 - P(A)]P(B) = P(\overline{A})P(B).$$

Hence $\overline{A}$ and B are independent.

(ii) and (iii) can be obtained by the similar way. ■

Example 1.6.3 Let A and B be independent events with $P(A) = 1/4$ and $P(A \cup B) = 2P(B) - P(A)$. Find the values of $P(B)$, $P(A|B)$ and $P(\overline{B}|A)$.

Solution. Since A and B are independent,

$$P(A \cup B) = P(A) + P(B) - P(AB) = P(A) + P(B) - P(A)P(B).$$

Thus,

$$1/4 + P(B) - (1/4)P(B) = 2P(B) - 1/4.$$

This implies that $P(B) = 2/5$.

Since A and B are independent,

$$P(A|B) = P(A) = 1/4.$$

By using Theorem 1.6.2(ii),

$$P(\overline{B}|A) = P(\overline{B}) = \frac{3}{5}.$$

■

1.6.2 Independence of Several Events

By multiplication rule, equation (1.19), we define independence of two events above. Now the problem is how to define the independence of three events A, B, C? How to express the occurrence of any event of the three events has nothing to do with the occurrence of the other two events?

Definition 1.6.2 *Three events A, B and C in the sample space S of a random experiment are said to be **mutually independent** (or independent) if*

$$\begin{aligned}
P(AB) &= P(A)P(B),\\
P(AC) &= P(A)P(C),\\
P(BC) &= P(B)P(C),\\
P(ABC) &= P(A)P(B)P(C).
\end{aligned}$$

*Events A, B and C are said to be **pairwise independent** if the first three equations hold.*

Example 1.6.4 Suppose that a balanced coin is independently tossed two times. Define the following events:

A: Head appears on the first toss;

B: Head appears on the second toss;

C: Both tosses yield the same outcome.

Are events A, B and C mutually independent?

Solution. The sample space Ω consists of the following outcomes: $\Omega = \{HH, HT, TH, TT\}$ and the events listed above are $A = \{HH, HT\}$, $B = \{TH, HH\}$, $C = \{HH, TT\}$.

Since the coin is balanced, that is, all outcomes are assigned the same probability, namely 1/4, we get

$$\begin{aligned} &P(A) = P(B) = P(C) = 2/4 = 1/2, \\ &P(AB) = P(AC) = P(BC) = \frac{1}{4}, \\ &P(ABC) = P(\{HH\}) = 1/4. \end{aligned}$$

Therefore,

$$P(AB) = P(A)P(B), \quad P(AC) = P(A)P(C), \quad P(BC) = P(B)P(C).$$

However,

$$P(ABC) \neq P(A)P(B)P(C).$$

So A, B and C are pairwise independent. ■

Example 1.6.5 Toss two different standard dice, white and black. Define the following events:

$A_1 = \{$first die = 1, 2 or 3$\}$;

$A_2 = \{$first die = 3, 4 or 5$\}$;

$A_3 = \{$sum of faces is 9$\}$.

Are events A_1, A_2 and A_3 mutually independent?

Solution. The sample space Ω of the outcomes consists of all ordered pairs $(i, j), i, j = 1, \cdots, 6$, i.e., $\Omega = \{(1, 1), (1, 2), ..., (6, 6)\}$. Since the two dice are standard, each point in S has probability 1/36. So the probability of the events are

$$P(A_1) = 1/2, \quad P(A_2) = 1/2, \quad P(A_3) = 1/9.$$

Obviously,

$$\begin{aligned} &A_1A_2 = \{(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)\}, \\ &A_1A_3 = \{(3, 6)\}, \ A_2A_3 = \{(3, 6), (4, 5), (5, 4)\}, \ A_1A_2A_3 = \{(3, 6)\}. \end{aligned}$$

It follows that

$$P(A_1A_2A_3) = 1/36 = (1/2)(1/2)(1/9) = P(A_1)P(A_2)P(A_3).$$

However,

$$\begin{aligned} &P(A_1A_2) = 1/6 \neq 1/4 = P(A_1)P(A_2), \\ &P(A_1A_3) = 1/36 \neq 1/18 = P(A_1)P(A_3), \\ &P(A_2A_3) = 1/12 \neq 1/18 = P(A_2)P(A_3). \end{aligned}$$

So events A_1, A_2 and A_3 are not mutually independent. ■

From Example 1.6.4, we can obtain the conclusion that pairwise independence of a given set of random events does not imply that these events are mutually independent. From Example 1.6.5, we can see that $P(ABC) = P(A)P(B)P(C)$ does not mean A, B, C are pairwise independent. So the four equations in the definition of independence of three events are all essential.

Now we extend the definition of independence to the cases of more events.

Definition 1.6.3 *Events $A_1, A_2, \cdots, A_n$ are **pairwise independent** if every possible pair of these events are independent, i.e., $P(A_jA_k) = P(A_j) \cdot P(A_k)$ for all $j, k (j \neq k)$.*

Definition 1.6.4 *The n events $A_1, A_2, \cdots, A_n$ are **independent** if, for every subset $A_{i_1}, A_{i_2}, \cdots, A_{i_k}$ of these events $(k = 2, 3, \cdots, n)$,*

$$P(A_{i_1}A_{i_2}\cdots A_{i_k}) = P(A_{i_1})P(A_{i_2})\cdots P(A_{i_k}). \tag{1.21}$$

Theorem 1.6.3 *The n events $A_1, A_2, \cdots, A_n$ are **independent** if and only if for any subset $A_{i_1}, A_{i_2}, \cdots, A_{i_k}$,*

$$P(A^*_{i_1}A^*_{i_2}\cdots A^*_{i_k}) = P(A^*_{i_1})P(A^*_{i_2})\cdots P(A^*_{i_k}), \quad (k = 2, 3, \cdots, n) \tag{1.22}$$

where A^*_i denotes either A_i or $\overline{A}_i$ (the same on both sides of the equation).

Example 1.6.6 Suppose that A, B and C are three independent events such that $P(A) = 1/4, P(B) = 1/3$ and $P(C) = 1/2$. (a) Determine the probability that none of these three events will occur; (b) Determine the probability that exact one of these three events will occur.

Solution. (a) By Theorem 1.6.3,

$$P(\overline{A}\ \overline{B}\ \overline{C}) = P(\overline{A})P(\overline{B})P(\overline{C}) = (1 - 1/4)(1 - 1/3)(1 - 1/2) = 1/4.$$

(b) By additivity of probability and Theorem 1.6.3,

$$\begin{aligned}
&P(A\overline{B}\ \overline{C} \cup \overline{A}B\overline{C} \cup \overline{A}\ \overline{B}C) \\
&= P(A\overline{B}\ \overline{C}) + P(\overline{A}B\overline{C}) + P(\overline{A}\ \overline{B}C) \\
&= P(A)P(\overline{B})P(\overline{C}) + P(\overline{A})P(B)P(\overline{C}) + P(\overline{A})P(\overline{B})P(C) \\
&= (1/4)(1 - 1/3)(1 - 1/2) + (1 - 1/4)(1/3)(1 - 1/2) + (1 - 1/4)(1 - 1/3)(1/2) \\
&= 1/12 + 1/8 + 1/4 = 11/24.
\end{aligned}$$

■

Example 1.6.7 (**Reliability analysis**) We have two systems illustrated in Fig. 1.7. Given a particular lifetime value t_0, let A_i denote the event that the lifetime of cell i exceeds $t_0 (i = 1, 2, \cdots, 6)$. Assume that the A_is are independent events (whether any particular cell lasts more than t_0 hours has no bearing on whether or not any other cell does) and that $P(A_i) = 0.9$ for every i since the cells are identical. Determine the probabilities that the systems' lifetime exceeds t_0. Which system is more reliable?

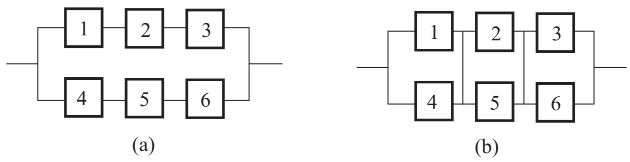

Figure 1.7: System configuration (a) series parallel, (b) total-cross-tied

Analysis. For the system in Fig. 1.7(a), there are two subsystems connected in parallel, each one containing three cells. In order for the system to function, at least one of the two parallel subsystems must work. Within each subsystem, the three cells are connected in series, so a subsystem will work only if all cells in the subsystem work.

For the total-cross-tied system shown in Fig. 1.7(b), it is obtained from the series-parallel array by connecting ties across each column of junctions. Now the system fails as soon as an entire column fails, and system lifetime exceeds t_0 only if the life of every column does so.

Solution. For the first system,

$$\begin{aligned}
&P(\text{system lifetime exceeds } t_0)\\
&= P\big((A_1 \cap A_2 \cap A_3) \cup (A_4 \cap A_5 \cap A_6)\big)\\
&= P(A_1 \cap A_2 \cap A_3) + P(A_4 \cap A_5 \cap A_6) - P\big((A_1 \cap A_2 \cap A_3) \cap (A_4 \cap A_5 \cap A_6)\big)\\
&= P(A_1)P(A_2)P(A_3) + P(A_4)P(A_5)P(A_6) - P(A_1)P(A_2)P(A_3)P(A_4)P(A_5)P(A_6)\\
&= 0.9^3 + 0.9^3 - 0.9^6 \approx 0.927.
\end{aligned}$$

Alternatively,

$$\begin{aligned}
P(\text{system lifetime exceeds } t_0) &= 1 - P(\text{both subsystem lives are } \leqslant t_0)\\
&= 1 - [P(\text{subsystem lives are } \leqslant t_0)]^2\\
&= 1 - [1 - P(\text{subsystem lives are } > t_0)]^2\\
&= 1 - [1 - (0.9)^3]^2 = 0.927.
\end{aligned}$$

For the second configuration,

$$\begin{aligned}
P(\text{system lifetime exceeds } t_0) &= P\big((A_1 \cup A_4) \cap (A_2 \cup A_5) \cap (A_3 \cup A_6)\big)\\
&= P(A_1 \cup A_4) \cdot P(A_2 \cup A_5) \cdot P(A_3 \cup A_6)\\
&= \big[P(A_1) + P(A_4) - P(A_1 A_4)\big]^3\\
&= (0.9 + 0.9 - 0.9 \cdot 0.9)^3 \approx 0.97
\end{aligned}$$

In reliability theory, we call the probability $P(\text{system lifetime exceeds } t_0)$ the reliability of the system. Since the reliability of the second system is larger than the first, the second system is more reliable. ■

1.6.3 Bernoulli Trials

The Bernoulli trials, named after Swiss mathematician Jacob Bernoulli (1654-1705), is one of the simplest yet most important trials in probability. Consider the following experiments:

(i) Flipping a coin and observing which side will appears.

(ii) Rolling a die and observing whether the number six appears.

(iii) In conducting a political opinion poll, choosing a voter at random to ascertain whether that voter will vote "yes" in an upcoming referendum.

All above examples have the same character: the experiment E that has *only two outcomes.* A ***Bernoulli experiment*** E is such kind of random experiment, the outcome of which can be classified in but one of two mutually exclusive and exhaustive ways, mainly, *success* (denoted by S) or *failure* (denoted by F) (e.g., female or male, life or death, nondefective or defective).

A sequence of Bernoulli trials E_n occurs when a Bernoulli experiment is *performed several independent times* so that *the probability of success*, say, p, *remains the same* from trial to trial. That is, in such a sequence we let p denote the probability of success on each trial. In addition, frequently $q = 1 - p$ denote the probability of failure; that is, we shall use q and $1 - p$ interchangeably.

It is easy to get the following

Theorem 1.6.4 *The outcomes of* E_n *are the* 2^n *sequences of length* n*. The number of outcomes of* E_n *that contain a exactly* k *times is given by the binomial coefficient* $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

Theorem 1.6.5 *The probability that the outcome of an experiment that consists of* n *Bernoulli trials has* k *successes and* $n - k$ *failures is given by:*

$$P_n(k) := P_n(\#S = k) = \binom{n}{k} p^k q^{n-k}.$$

Here $\#S$ *means the number of successes.*

Proof. Now, suppose that, when we performed the n trials, we first obtained k consecutive successes S and then $n - k$ consecutive failures F. By independence, we may write that the probability of this elementary outcome is

$$P(\underbrace{SS\cdots S}_{k\ times}\ \underbrace{FF\cdots F}_{(n-k)\ times}) = P(S)^k P(F)^{n-k} = p^k (1-p)^{n-k}.$$

Hence, given that we can place the k successes among the n trials in $\binom{n}{k}$ different ways, we deduce that

$$P_n(\#S = k) = \binom{n}{k} p^k q^{n-k}.$$

■

Example 1.6.8 We are monitoring calls at a switchboard in a large manufacturing firm and have determined that one third of the calls are long distance and two thirds of the calls are local. We have decided to pick ten calls at random and would like to know how many calls in the group of four are long distance. In other words, let N be a random variable indicating the number of long distance calls in the group of four. Thus, N is binomial with $n = 10$ and $p = 1/3$. It also happens that in this company, half of the individuals placing calls are women and half are men. We would also like to know how many of the group of four were calls placed by men. Let M denote the number of men placing calls; thus, M is binomial with $n = 10$ and $p = 1/2$. ■

1.7 Review

In this section, we will summarize the main content of this chapter.

1. A function $P(\cdot)$ defined on events is a probability function or probability distribution if

   – for every $A \in \mathcal{F}, P(A) \geqslant 0$,

   – $P(\Omega) = 1$,

   – $P\big(\bigcup\limits_{i=1}^{\infty} A_i\big) = \sum\limits_{i=1}^{\infty} P(A_i)$ for any disjoint events events $A_1, A_2, \cdots$.

2. The function $P(\cdot)$ obeys the following rules:

   – **Range:** $0 \leqslant P(A) \leqslant 1$.

   – **Impossible event:** $P(\emptyset) = 0$.

   – **Union:** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

   – **Complement:** $P(A) = 1 - P(\overline{A})$.

   – **Difference:** $P(A - B) = P(A) - P(AB)$.

3. **Notation:** $P(A|B)$ The conditional probability of A given B.

   – **Conditioning Rule:** $P(A|B) = P(AB)/P(B)$.

   – **Addition Rule:** $P(A \cup B|C) = P(A|C) + P(B|C)$, when $A \cap C$ and $B \cap C$ are disjoint.

   – **Multiplication Rule:** $P(ABC) = P(A|BC)P(B|C)P(C)$.

   – **Law of Total Probability:** $P(A) = \sum\limits_{i=1}^{n} P(B_i)P(A|B_i)$, where $B_1, B_2, \cdots, B_n$ form a partition of the sample space Ω.

   – **Bayes′s Rule (or Theorem):** $P(B_i|A) = \dfrac{P(B_i)P(A|B_i)}{\sum\limits_{j=1}^{n} P(B_j)P(A|B_j)}$.

   – **Extended Addition Rule:** $P\big(\bigcup\limits_i A_i|C\big) = \sum\limits_i P(A_i|C)$, where $A_1, A_2, \cdots$ form a partition of the sample space.

3. **Independence Rule:** A and B are independent if and only if $P(AB) = P(A)P(B)$.

   More generally, the n events $A_1, A_2, \cdots, A_n$ are **independent** if and only if for any subset $A_{i_1}, A_{i_2}, \cdots, A_{i_k}$,

   $$P(A^*_{i_1} A^*_{i_2} \cdots A^*_{i_k}) = P(A^*_{i_1}) P(A^*_{i_2}) \cdots P(A^*_{i_k}), \quad (k = 2, 3, \cdots, n)$$

   where A^*_i denotes either A_i or $\overline{A}_i$ (the same on both sides of the equation).

1.8 Exercises

- **Experiment, Sample Space and Random Event**

1.1 We consider the following random experiment: a fair die is rolled; if (and only if) a 6 is obtained, the die is rolled a second time. How many elementary outcomes are there in the sample space Ω?

1.2 An academic department has just completed voting by secret ballot for a department head. The ballot box contains four slips with votes for candidate A and three slips with votes for candidate B. Suppose that these slips are removed from the box one by one.

(a) List all possible outcomes.

(b) Suppose that a running tally is kept as slips are removed. For what outcomes does A remain ahead of B throughout the tally?

1.3 From 10 married couples, we want to select a group of 6 people that is not allowed to contain a married couple.

(a) How many choices are there?

(b) How many choices are there if the group must also consist of 3 men and 3 women?

1.4 Consider n-digit numbers where each digit is one of the 10 integers 0, 1, $\cdots$, 9. How many such numbers are there for which

(a) no two consecutive digits are equal?

(b) 0 appears as a digit a total of i times, $i = 0, \cdots, n$?

1.5 Let E, F, G be three events. Find expressions for the events that of E, F, G

(a) only F occurs,

(b) both E and F but not G occur,

(c) at least one event occurs,

(d) at least two events occur,

(e) all three events occur,

(f) none occurs,

(g) at most one occurs,

(h) at most two occur.

1.6 Express the following events in the simplest form.

(a) $(A \cup B) \cap (B \cup C)$,

(b) $(A \cup B) \cap (A \cup \overline{B})$,

(c) $(A \cup B) \cap (A \cup \overline{B}) \cap (\overline{A} \cup B)$.

1.7 Suppose that a coin is tossed ten times. Let A denote the event that a head is obtained on the first toss, and let B denote the event that a head is obtained on the sixth toss. Are A and B disjoint?

- **Probabilities Defined On Events**

1.8 A box contains three marbles: one red, one green, and one blue. Consider an experiment that consists of taking one marble from the box then replacing it in the box and drawing a second marble from the box. What is the sample space? If, at all times, each marble in the box is equally likely to be selected, what is the probability of each point in the sample space?

1.9 You roll two dice. What is the probability of the events:

(a) They show the same?

(b) Their sum is seven or eleven?

(c) They have no common factor greater than unity?

(d) The sum of the numbers is 2, 3, or 12?

(e) The sum is odd?

(f) The difference is odd?

(g) The product is odd?

(h) One number divides the other?

(i) The first die shows a smaller number than the second?

(j) Different numbers are shown and the smaller of the two numbers is $r, 1 \leqslant r \leqslant 6$?

1.10 A tea set has four cups and saucers with two cups and saucers in each of two different colors. If the cups are placed at random on the saucers, what is the probability that

(a) No cup is on a saucer of the same color?

(b) One cup is on a saucer of the same color?

(c) Two cups are on saucers of the same color?

(d) No cup is on a saucer of the same color if the set comprises four cups and saucers in four distinct colors?

1.11 A woman planning her family considers the following schemes on the assumption that boys and girls are equally likely at each delivery:

(a) Have three children.

(b) Bear children until the first girl is born or until three are born, whichever is sooner, and then stop.

(c) Bear children until there is one of each sex or until there are three, whichever is sooner, and then stop.

Let B_i denote the event that i boys are born, and let C denote the event that more girls are born than boys. Find $P(B_1)$ and $P(C)$ in each of the cases a and b.

1.12 A point (x, y) is to be selected from the square A containing all points (x, y) such that $0 \leqslant x \leqslant 1$ and $0 \leqslant y \leqslant 1$. Suppose that the probability that the selected point belong to each specified subset of A is equal to the area of that subset. Find the probability of each of the following subsets:

(a) the subset of points such that $(x - \frac{1}{2})^2 + (y - \frac{1}{2})^2 \geqslant \frac{1}{4}$;

(b) the subset of points such that $(x + \frac{1}{2})^2 + (y - \frac{1}{2})^2 \geqslant \frac{1}{4}$;

(c) the subset of points such that $y \leqslant 1 - x^2$;

(d) the subset of points such that $y = x$.

1.13 A train and a bus arrive at the station at random between 9 A.M. and 10 A.M. The train stops for 10 minutes and the bus for x minutes. Find x so that the probability that the bus and the train will meet equals 0.5.

1.14 Let a point be picked at random in unit square. Compute the probability that it is in the triangle bounded by $x = 0$, $y = 0$, and $x + y = 1$.

1.15 Let a point be picked at random in the disk of radius 1. Find the probability that it lies in the angular sector from 0 to $\pi/4$ radians.

1.16 Suppose that A and B are events in sample space Ω. Suppose that $P(A) = 1$ and $P(B) = 0$. Is A an inevitable event? Is B an impossible event?

- **Probability Space**

1.17 Suppose that $\mathscr{F}$ is a σ-algebra of Ω. Prove the following:

(a) $\emptyset \in \mathscr{F}$;

(b) If $A, B \in \mathscr{F}$, then $AB \in \mathscr{F}$ and $A - B \in \mathscr{F}$.

(c) If $A_i \in \mathscr{F}, i = 1, 2, \cdots, n$, then $\bigcup\limits_{i=1}^{n} A_i \in \mathscr{F}$ and $\bigcap\limits_{i=1}^{n} \in \mathscr{F}$.

(d) If $A_i \in \mathscr{F}, i = 1, 2, \cdots$, then $\bigcap\limits_{i=1}^{\infty} \in \mathscr{F}$.

1.18 Suppose that $P(AB) = 0$. Then what conclusion in the following is correct?

(A) AB is an impossible event; (B) AB may not be an impossible event;

(C) A and B are disjoint; (D) $P(A) = 0$ or $P(B) = 0$.

1.19 Show that if $AB = \emptyset$, then $P(A) \leqslant P(\overline{B})$.

1.20 If A, B and C are three events such that $P(A \cup B \cup C) = 0.7$, what is the probability of $P(\overline{A}\,\overline{B}\,\overline{C})$?

1.21 Show that if $P(A) = P(B) = P(AB)$, then $P(A\overline{B} \cup B\overline{A}) = 0$.

1.22 Let A and B be events with probabilities $P(A) = 3/4$ and $P(B) = 1/3$. Show that $1/12 \leqslant P(AB) \leqslant 1/3$, and give examples to show that both extremes are possible. Find corresponding bounds for $P(A \cup B)$.

1.23 Show that the probability that exactly one of the events A and B occurs is

$$P(A) + P(B) - 2P(A \cap B).$$

1.24 If $P(A) = 0.9$ and $P(B) = 0.8$, show that $P(AB) \geqslant 0.7$. In general, show that

$$P(AB) \geqslant P(A) + P(B) - 1$$

This is known as *Bonferroni's inequality.*

1.25 If $P(A) = 0.5, P(B) = 0.4$, and $P(A - B) = 0.32$. Calculate the values of $P(AB)$ and $P(A \cup B)$.

1.26 If $P(A) = p, P(B) = q$ and $P(A \cup B) = r$. Calculate the values of $P(AB)$, $P(A\,\overline{B})$ and $P(\overline{A}\,\overline{B})$.

1.27 Show if $P(A) + P(B) = 1$, then $P(AB) = P(\overline{A}\,\overline{B})$.

- **Conditional Probabilities**

1.28 If A and B are disjoint, $P(A) = P(B) = a, 0 < a < 1$, and $P(A|\overline{B}) = P(\overline{A}|\overline{B})$, then find the values of a and $P(A \cup B)$.

1.29 If $0 < P(B) < 1$, and $AB = \overline{A}\,\overline{B}$, then find the value of $P(A|\overline{B}) = P(\overline{A}|B)$.

1.30 Four individuals have responded to a request by a blood bank for blood donations. None of them has donated before, so their blood types are unknown. Suppose that only type O+ is desired and only one of the four actually has this type. If the potential donors are selected in random order for typing, what is the probability that at least three individuals must be typed to obtain the desired type?

1.31 An urn contains b blue balls and c cyan balls. A ball is drawn at random, its color is noted, and it is returned to the urn together with d further balls of the same color. This procedure is repeated indefinitely. What is the probability that:

(a) The second ball drawn is cyan?

(b) The first ball drawn is cyan given that the second ball drawn is cyan?

1.32 We have two coins; the first is fair and the second two-headed. We pick one of the coins at random, we toss it twice and heads shows both times. Find the probability that the coin picked is fair.

1.33 Box 1 contains 1000 bulbs of which 10% are defective. Box 2 contains 2000 bulbs of which 5% are defective. Two bulbs are picked from a randomly selected box.

(a) Find the probability that both bulbs are defective.

(b) Assuming that both are defective, find the probability that they came from box 1.

1.34 Incidence of a rare disease. Only 1 in 1000 adults is afflicted with a rare disease for which a diagnostic test has been developed. The test is such that when an individual actually has the disease, a positive result will occur 99% of the time, whereas an individual without the disease will show a positive test result only 2% of the time. If a randomly selected individual is tested and the result is positive, what is the probability that the individual has the disease?

1.35 Suppose that an individual is randomly selected from the population of all adult males living in the China. Let A be the event that the selected individual is over 6 ft in height, and let B be the event that the selected individual is a professional basketball player. Which do you think is larger, $P(A|B)$ or $P(B|A)$? Why?

1.36 A commuter has two vehicles, one being a compact car and the other one a minivan. Three times out of four, he uses the compact car to go to work and the remainder of the time he uses the minivan. When he uses the compact car (resp., the minivan), he gets home before 5:30 p.m. 75% (resp., 60%) of the time. However, the minivan has air conditioning. Calculate the probability that

(a) he gets home before 5:30 p.m. on a given day;

(b) he used the compact car if he did not get home before 5:30 p.m.;

(c) he uses the minivan and he gets home after 5:30 p.m.;

(d) he gets home before 5:30 p.m. on two (independent) consecutive days and he does not use the same vehicle on these two days.

1.37 A company that manufactures video cameras produces a basic model and a deluxe model. Over the past year, 40% of the cameras sold have been of the basic model. Of those buying the basic model, 30% purchase an extended warranty, whereas 50% of all deluxe purchasers do so. If you learn that a randomly selected purchaser has an extended warranty, how likely is it that he or she has a basic model?

1.38 One box contains six red balls and four green balls, and a second box contains seven red balls and three green balls. A ball is randomly chosen from the first box and placed in the second box. Then a ball is randomly selected from the second box and placed in the first box.

(a) What is the probability that a red ball is selected from the first box and a red ball is selected from the second box?

(b) At the conclusion of the selection process, what is the probability that the numbers of red and green balls in the first box are identical to the numbers at the beginning?

1.39 One half percent of the population has a particular disease. A test is developed for the disease. The test gives a false positive 3% of the time and a false negative 2% of the time.

(a) What is the probability that Joe (a random person) tests positive?

(b) Joe just got the bad news that the test came back positive; what is the probability that Joe has the disease?

1.40 A lab is attempting to stain many cells. Young cells stain properly 90% of the time and old cells stain properly 70% of the time.

(a) If 30% of the cells are young, what is the probability that a cell stains properly?

(b) If 70% of the cells are young, what is the probability that a cell stains properly?

1.41 Further study of the cell-staining problem reveals that new cells stain properly with probability 0.95, 1-day-old cells stain properly with probability 0.9, 2-day-old cells stain properly with probability 0.8, and 3-day-old cells stain properly with probability 0.5. Suppose that

$$\begin{aligned} P(\text{cell is 0 day old}) &= 0.4 \\ P(\text{cell is 1 day old}) &= 0.3 \\ P(\text{cell is 2 days old}) &= 0.2 \\ P(\text{cell is 3 days old}) &= 0.1. \end{aligned}$$

(a) Find the probability that a cell stain properly.

(b) The lab finds a way to estimate the oldest cells(more than 3 days old) from its stock. What is the probability of proper staining? Write this as a conditional probability.

- **Independent Events**

1.42 Suppose that three cities A, B and C are connected by roads. There are two roads from A to B and two roads from B to C. In winters, each of the four roads is blocked by snow with probability p $(0 < p < 1)$, independently of the others.

(a) Assume that there are no direct roads connecting A and C without passing through B. Find the probability that there is an open road from A to B given that there is no open route from A to C.

(b) If, in addition, there is another direct road from A to C and this road is blocked with the same probability p, independently of the others, then find, again, the probability that there is an open road from A to B given that there is no open route from A to C.

1.43 (a) If A is independent of itself, show that $P(A)$ is 0 or 1.

(b) If $P(A)$ is 0 or 1, show that A is independent of all events B.

1.44 Suppose that $0 < P(A) < 1$. Prove $P(B|A) = P(B|\overline{A})$ if and only if A and B are independent.

1.45 Suppose that A and B are independent, and $P(\overline{A}\ \overline{B}) = 1/9, P(A) = P(\overline{B})$. Find the values of $P(A)$.

1.46 A document is equally likely to be in any of three box files. A search of the ith box will discover the document (if it is indeed there) with probability p_i. What is the probability that the document is in the first box:

(a) Given that I have searched the first box once and not found it?

(b) Given that I have searched the first box twice and not found it?

(c) Given that I have searched all three boxes once and not found it? Assume searches are independent.

1.47 Let p represent the probability of an event A. What is the probability that

(a) A occurs at least twice in n independent trials;

(b) A occurs at least third in n independent trials?

1.48 Suppose that there are r successes in n independent Bernoulli trials. Find the conditional probability of a success on the ith trial.

1.49 A player wins \$1 if he throws two heads in succession, otherwise he loses two quarters. If the game is repeated 50 times, what is the probability that the net gain or less exceeds (a) \$l? (b)\$5 ?

1.50 Jane has three children, each of which is equally likely to be a boy or a girl independently of the others. Define the events:

A={all the children are of the same sex},

B={there is at most one boy},

C={the family includes a boy and a girl}.

(a) Show that A is independent of B, and that B is independent of C.

(b) Is A independent of C?

(c) Do these results hold if boys and girls are not equally likely?

(d) Do these results hold if Jane has four children?

1.51 A species of bird comes in three colors: red, blue, and green. Twenty percent are red, 30% are blue, and 50% are green. Females prefer red to blue to green and mate with the nest male they find.

(a) Females pick the better of the first two males they meet. What is the probability a female mates with a green bird? What did you have to assume about independence?

(b) Females pick the better of the first two males they meet. What is the probability a female mates with a blue bird and the probability a female mates with a red bird?