6 Elements of Probability Theory

Let us first consider some simple examples to understand the need for probability theory. Often one needs to make a decision whether to carry an umbrella or not when leaving the house; a company might wonder whether to introduce a new advertisement to possibly increase sales or to continue with their current advertisement; or someone may want to choose a restaurant based on where he can get his favourite dish. In all these situations, randomness is involved. For example, the decision of whether to carry an umbrella or not is based on the possibility or chance of rain. The sales of the company may increase, decrease, or remain unchanged with a new advertisement. The investment in a new advertising campaign may therefore only be useful if the probability of its success is higher than that of the current advertisement. Similarly, one may choose the restaurant where one is most confident of getting the food of one's choice. In all such cases, an event may be happening or not and depending on its likelihood, actions are taken. The purpose of this chapter is to learn how to calculate such likelihoods of events happening and not happening.

6.1 Basic Concepts and Set Theory

A simple (not rigorous) definition of a **random experiment** requires that the experiment can be repeated any number of times under the same set of conditions, and its outcome is known only after the completion of the experiment. A simple and classical example of a random experiment is the tossing of a coin or the rolling of a die. When tossing a coin, it is unknown what the outcome will be, head or tail, until the coin is tossed. The experiment can be repeated and different outcomes may be observed in each repetition. Similarly, when rolling a die, it is unknown how many dots will appear on the upper surface until the die is rolled. Again, the die can be rolled repeatedly and different numbers of dots are obtained in each trial. A possible

C. Heumann et al., *Introduction to Statistics and Data Analysis*,
DOI 10.1007/978-3-319-46162-5_6

outcome of a random experiment is called a **simple event** (or **elementary event**) and denoted by ω_i. The set of all possible outcomes, $\{\omega_1, \omega_2, \ldots, \omega_k\}$, is called the **sample space** and is denoted as Ω, i.e. $\Omega = \{\omega_1, \omega_2, \ldots, \omega_k\}$. Subsets of Ω are called **events** and are denoted by capital letters such as A, B, C. The set of all simple events that are contained in the event A is denoted by Ω_A. The event $\bar{A}$ refers to the non-occurring of A and is called **a composite or complementary event**. Also Ω is an event. Since it contains all possible outcomes, we say that Ω will always occur and we call it a **sure event** or **certain event**. On the other hand, if we consider the null set $\emptyset = \{\}$ as an event, then this event can never occur and we call it an **impossible event**. The sure event therefore is the set of all elementary events, and the impossible event is the set with no elementary events.

The above concepts of "events" form the basis of a definition of "probability". Once we understand the concept of probability, we can develop a framework to make conclusions about the population of interest, using a sample of data.

Example 6.1.1 (Rolling a die) If a die is rolled once, then the possible outcomes are the number of dots on the upper surface: $1, 2, \ldots, 6$. Therefore, the sample space is the set of simple events $\omega_1 =$ "1", $\omega_2 =$ "2", …, $\omega_6 =$ "6" and $\Omega = \{\omega_1, \omega_2, \ldots, \omega_6\}$. Any subset of Ω can be used to define an event. For example, an event A may be "an even number of dots on the upper surface of the die". There are three possibilities that this event occurs: ω_2, ω_4, or ω_6. If an odd number shows up, then the composite event $\bar{A}$ occurs instead of A. If an event is defined to observe only one particular number, say $\omega_1 =$ "1", then it is an elementary event. An example of a sure event is "a number which is greater than or equal to 1" because any number between 1 and 6 is greater than or equal to 1. An impossible event is "the number is 7".

Example 6.1.2 (Rolling two dice) Suppose we throw two dice simultaneously and an event is defined as the "number of dots observed on the upper surface of both the dice"; then, there are 36 simple events defined as (number of dots on first die, number of dots on second die), i.e. $\omega_1 = (1, 1), \omega_2 = (1, 2), \ldots, \omega_{36} = (6, 6)$. Therefore Ω is

$$\Omega = \begin{matrix} \{(1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6) \\ (2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6) \\ (3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6) \\ (4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6) \\ (5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6) \\ (6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6)\}. \end{matrix}$$

One can define different events and their corresponding sample spaces. For example, if an event A is defined as "upper faces of both the dice contain the same number of dots", then the sample space is $\Omega_A = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$. If another event B is defined as "the sum of numbers on the upper faces is 6", then

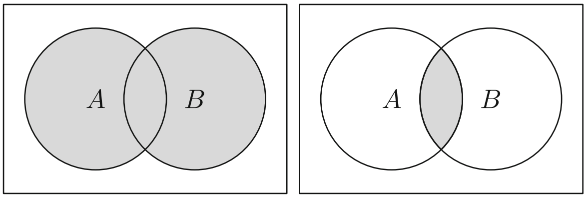

Fig. 6.1 $A \cup B$ and $A \cap B$*

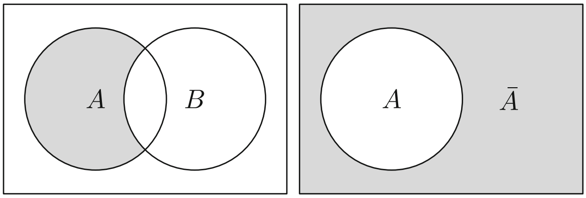

Fig. 6.2 $A \backslash B$ and $\bar{A} = \Omega \backslash A$*

the sample space is $\Omega_B = \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}$. A sure event is "get either an even number or an odd number"; an impossible event would be "the sum of the two dice is greater than 13".

It is possible to view events as sets of simple events. This helps to determine how different events relate to each other. A popular technique to visualize this approach is to use **Venn diagrams**. In Venn diagrams, two or more sets are visualized by circles. Overlapping circles imply that both events have one or more identical simple events. Separated circles mean that none of the simple events of event A are contained in the sample space of B. We use the following notations:

$A \cup B$ The union of events $A \cup B$ is the set of all simple events of A and B which occurs if at least one of the simple events of A *or* B occurs (Fig. 6.1, left side, grey shaded area). Please note that we use the word "or" from a statistical perspective: "A or B" means that either a simple event from A occurs, or a simple event from B occurs, or a simple event which is part of both A and B occurs.

$A \cap B$ The intersection of events $A \cap B$ is the set of all simple events A and B which occur when a simple event occurs that belongs to A *and* B (Fig. 6.1, right side, grey shaded area).

$A \backslash B$ The event $A \backslash B$ contains all simple events of A, which are not contained in B. The event "A but not B" or "A minus B" occurs, if A occurs but B does not occur. Also $A \backslash B = A \cap \bar{B}$ (Fig. 6.2, left side, grey shaded area).

$\bar{A}$ The event $\bar{A}$ contains all simple events of Ω, which are not contained in A. The complementary event of A (which is "Not-A" or "$\bar{A}$" occurs whenever A does not occur (Fig. 6.2, right side, grey shaded area).

$A \subseteq B$ A is a subset of B. This means that all simple events of A are also part of the sample space of B.

Example 6.1.3 Consider Example 6.1.1 where the sample space of rolling a die was determined as $\Omega = \{\omega_1, \omega_2, \ldots, \omega_6\}$ with $\omega_1 =$ "1", $\omega_2 =$ "2", ..., $\omega_6 =$ "6".

- If $A = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}$ and B is the set of all odd numbers, then $B = \{\omega_1, \omega_3, \omega_5\}$ and thus $B \subseteq A$.
- If $A = \{\omega_2, \omega_4, \omega_6\}$ is the set of even numbers and $B = \{\omega_3, \omega_6\}$ is the set of all numbers which are divisible by 3, then $A \cup B = \{\omega_2, \omega_3, \omega_4, \omega_6\}$ is the collection of simple events for which the number is either even or divisible by 3 or both.
- If $A = \{\omega_1, \omega_3, \omega_5\}$ is the set of odd numbers and $B = \{\omega_3, \omega_6\}$ is the set of the numbers which are divisible by 3, then $A \cap B = \{\omega_3\}$ is the set of simple events in which the numbers are odd and divisible by 3.
- If $A = \{\omega_1, \omega_3, \omega_5\}$ is the set of odd numbers and $B = \{\omega_3, \omega_6\}$ is the set of the numbers which are divisible by 3, then $A \backslash B = \{\omega_1, \omega_5\}$ is the set of simple events in which the numbers are odd but not divisible by 3.
- If $A = \{\omega_2, \omega_4, \omega_6\}$ is the set of even numbers, then $\bar{A} = \{\omega_1, \omega_3, \omega_5\}$ is the set of odd numbers.

Remark 6.1.1 Some textbooks also use the following notations:

$$\begin{array}{lll} A + B & \text{for} & A \cup B \\ AB & \text{for} & A \cap B \\ A - B & \text{for} & A \backslash B. \end{array}$$

We can use these definitions and notations to derive the following properties of a particular event A:

$$\begin{array}{ll} A \cup A = A & A \cap A = A \\ A \cup \Omega = \Omega & A \cap \Omega = A \\ A \cup \emptyset = A & A \cap \emptyset = \emptyset \\ A \cup \bar{A} = \Omega & A \cap \bar{A} = \emptyset. \end{array}$$

Definition 6.1.1 Two events A and B are ***disjoint*** if $A \cap B = \emptyset$ holds, i.e. if both events cannot occur simultaneously.

Example 6.1.4 The events A and $\bar{A}$ are disjoint events.

Definition 6.1.2 The events $A_1, A_2, \ldots, A_m$ are said to be mutually or pairwise disjoint, if $A_i \cap A_j = \emptyset$ whenever $i \neq j = 1, 2, ..., m$.

Example 6.1.5 Recall Example 6.1.1. If $A = \{\omega_1, \omega_3, \omega_5\}$ and $B = \{\omega_2, \omega_4, \omega_6\}$ are the sets of odd and even numbers, respectively, then the events A and B are disjoint.

Definition 6.1.3 The events $A_1, A_2, \ldots, A_m$ form a **complete decomposition** of Ω if and only if

$$A_1 \cup A_2 \cup \cdots \cup A_m = \Omega$$

and

$$A_i \cap A_j = \emptyset \quad \text{(for all } i \neq j).$$

Example 6.1.6 Consider Example 6.1.1. The elementary events $A_1 = \{\omega_1\}$, $A_2 = \{\omega_2\}, \ldots, A_6 = \{\omega_6\}$ form a complete decomposition. Other complete decompositions are, e.g.

- $A_1 = \{\omega_1, \omega_3, \omega_5\}, \quad A_2 = \{\omega_2, \omega_4, \omega_6\}$
- $A_1 = \{\omega_1\}, \quad A_2 = \{\omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$
- $A_1 = \{\omega_1, \omega_2, \omega_3\}, \quad A_2 = \{\omega_4, \omega_5, \omega_6\}$.

6.2 Relative Frequency and Laplace Probability

There is a close connection between the relative frequency and the probability of an event. A random experiment is described by its possible outcomes, for example getting a number between 1 and 6 when rolling a die. Suppose an experiment has m possible outcomes (events) $A_1, A_2, \ldots, A_m$ and the experiment is repeated n times. Now we can count how many times each of the possible outcome has occurred. In other words, we can calculate the absolute frequency $n_i = n(A_i)$ which is equal to the number of times an event A_i, $i = 1, 2, \ldots, m$, occurs. The relative frequency $f_i = f(A_i)$ of a random event A_i, with n repetitions of the experiment, is calculated as

$$f_i = f(A_i) = \frac{n_i}{n}. \tag{6.1}$$

Example 6.2.1 Consider roulette, a game frequently played in casinos. The roulette table consists of 37 numbers from 0 to 36. Out of these 37 numbers, 18 numbers are red, 18 are black and one (zero) is green. Players can place their bets on either a single number or a range of numbers, the colours red or black, whether the number is odd or even, among many other choices. A casino employee spins a wheel (containing pockets representing the 37 numbers) in one direction and then spins a ball over the wheel in the opposite direction. The wheel and ball gradually slow down and the ball finally settles in a pocket. The pocket number in which the ball sits down when the wheel stops is the winning number. Consider three possible outcomes A_1: "red", A_2:"black", and A_3: "green (zero)". Suppose the roulette ball is spun $n = 500$ times. All the outcomes are counted and recorded as follows: A_1 occurs 240 times, A_2 occurs 250 times and A_3 occurs 10 times. Then, the absolute frequencies are given by $n_1 = n(A_1) = 240$, $n_2 = n(A_2) = 250$, and $n_3 = n(A_3) = 10$. We therefore get the relative frequencies as

$$f_1 = f(A_1) = \frac{240}{500} = 0.48\,, \quad f_2 = f(A_2) = \frac{250}{500} = 0.5,$$
$$f_3 = f(A_3) = \frac{10}{500} = 0.02.$$

If we assume that the experiment is repeated a large number of times (mathematically, this would mean that n tends to infinity) and the experimental conditions remain the same (at least approximately) over all the repetitions, then the relative frequency $f(A)$ converges to a limiting value for A. This limiting value is interpreted as the probability of A and denoted by $P(A)$, i.e.

$$P(A) = \lim_{n \to \infty} \frac{n(A)}{n}$$

where $n(A)$ denotes the number of times an event A occurs out of n times.

Example 6.2.2 Suppose a fair coin is tossed $n = 20$ times and we observe the number of heads $n(A_1) = 8$ times and number of tails $n(A_2) = 12$ times. The meaning of a fair coin in this case is that the probabilities of head and tail are equal (i.e. 0.5). Then, the relative frequencies in the experiment are $f(A_1) = 8/20 = 0.4$ and $f(A_2) = 12/20 = 0.6$. When the coin is tossed a large number of times and n tends to infinity, then both $f(A_1)$ and $f(A_2)$ will have a limiting value 0.5 which is simply the probability of getting a head or tail in tossing a fair coin.

Example 6.2.3 In Example 6.2.1, the relative frequency of $f(\text{red}) = f(A_1)$ tends to $18/37$ as n tends to infinity because 18 out of 37 numbers are red.

The reader will gain a more theoretical understanding of how repeated experiments relate to expected quantities in the following chapters after learning the Theorem of Large Numbers described in Appendix A.3.

A different definition of probability was given by Pierre-Simon Laplace (1749–1827). We call an experiment a **Laplace experiment** if the number of possible simple events is finite and all the outcomes are equally probable. The probability of an arbitrary event A is then defined as follows:

Definition 6.2.1 The proportion

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{Number of “favourable simple events” for } A}{\text{Total number of possible simple events}} \tag{6.2}$$

is called the **Laplace probability**, where $|A|$ is the cardinal number of A, i.e. the number of simple events contained in the set A, and $|\Omega|$ is the cardinal number of Ω, i.e. the number of simple events contained in the set Ω.

The cardinal numbers $|A|$ and $|\Omega|$ are often calculated using the combinatoric rules introduced in Chap. 5.

Example 6.2.4 (Example 6.1.2 continued) The sample space contains 36 simple events. All of these simple events have equal probability 1/36. To calculate the probability of the event A that the sum of the dots on the two dice is at least 4 and at most 6, we count the favourable simple events which fulfil this condition. The simple events are $(1, 3)$, $(2, 2)$, $(3, 1)$ (sum is 4), $(1, 4)$, $(2, 3)$, $(4, 1)$, $(3, 2)$ (sum is 5) and $(1, 5)$, $(2, 4)$, $(3, 3)$, $(4, 2)$, $(5, 1)$ (sum is 6). In total, there are $(3 + 4 + 5) = 12$ favourable simple events, i.e.

$$A = \{(1,3), (2,2), (3,1), (1,4), (2,3), (4,1), (3,2), (1,5), (2,4), (3,3), (4,2), (5,1)\} .$$

The probability of the event A is therefore $12/36 = 1/3$.

6.3 The Axiomatic Definition of Probability

An important foundation for modern probability theory was established by A.N. Kolmogorov in 1933 when he proposed the following **axioms of probability**.

Axiom 1 Every random event A has a probability in the (closed) interval [0, 1], i.e.

$$0 \leq P(A) \leq 1.$$

Axiom 2 The sure event has probability 1, i.e.

$$P(\Omega) = 1.$$

Axiom 3 If A_1 and A_2 are disjoint events, then

$$P(A_1 \cup A_2) = P(A_1) + P(A_2).$$

holds.

Remark Axiom 3 also holds for three or more disjoint events and is called the **theorem of additivity of disjoint events**. For example, if A_1, A_2, and A_3 are disjoint events, then $P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3)$.

Example 6.3.1 Suppose the two events in tossing a coin are A_1: "appearance of head" and A_2: "appearance of tail" which are disjoint. The event $A_1 \cup A_2$: "appearance of head or tail" has the probability

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) = 1/2 + 1/2 = 1.$$

Example 6.3.2 Suppose an event is defined as the number of points observed on the upper surface of a die when rolling it. There are six events, i.e. the natural numbers 1, 2, 3, 4, 5, 6. These events are disjoint and they have equal probability of occurring: $P(1) = P(2) = \cdots = P(6) = 1/6$. The probability of getting an even number is then

$$P(\text{“even number”}) = P(2) + P(4) + P(6) = 1/6 + 1/6 + 1/6 = 1/2.$$

6.3.1 Corollaries Following from Kolomogorov's Axioms

We already know that $A \cup \bar{A} = \Omega$ (sure event). Since A and $\bar{A}$ are disjoint, using Axiom 3 we have

$$P(A \cup \bar{A}) = P(A) + P(\bar{A}) = 1\,.$$

Based on this, we have the following corollaries.

Corollary 1 *The probability of the complementary event of A, (i.e.* $\bar{A}$*) is*

$$P(\bar{A}) = 1 - P(A). \tag{6.3}$$

Example 6.3.3 Suppose a box of 30 chocolates contains chocolates of 6 different flavours with 5 chocolates of each flavour. Suppose an event A is defined as $A = \{\text{“marzipan flavour”}\}$. The probability of finding a marzipan chocolate (without looking into the box) is $P(\text{“marzipan”}) = 5/30$. Then, the probability of the complementary event $\bar{A}$, i.e. the probability of not finding a marzipan chocolate is therefore

$$P(\text{“no marzipan flavour”}) = 1 - P(\text{“marzipan flavour”}) = 25/30.$$

Corollary 2 *The probability of occurrence of an impossible event* $\emptyset$ *is zero:*

$$P(\emptyset) = P(\bar{\Omega}) = 1 - P(\Omega) = 0.$$

Corollary 3 *Let* A_1 *and* A_2 *be not necessarily disjoint events. The probability of occurrence of* A_1 *or* A_2 *is*

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2). \tag{6.4}$$

The rule in (6.4) is known as **the additive theorem of probability**. Again we use the word “or” in the statistical sense: either A_1 is occurring, A_2 is occurring, or both of them. This means we have to add the probabilities $P(A_1)$ and $P(A_2)$ but need to make sure that the simple events which are contained in both sets are not counted twice, thus we subtract $P(A_1 \cap A_2)$.

Example 6.3.4 There are 10 actors acting in a play. Two actors, one of whom is male, are portraying evil characters. In total, there are 6 female actors. Let an event A describe whether the actor is male and another event B describe whether the character is evil. Suppose we want to know the probability of a randomly chosen actor being male or evil. We can then calculate

$$\begin{aligned}
&P(\text{actor is male or evil}) = \\
&\quad = P(\text{actor is male}) + P(\text{actor is evil}) - P(\text{actor is male and evil}) \\
&\quad = \frac{4}{10} + \frac{2}{10} - \frac{1}{10} = \frac{1}{2}\,.
\end{aligned}$$

Corollary 4 *If $A \subseteq B$ then $P(A) \leq P(B)$.*

Proof We use the representation $B = A \cup (\bar{A} \cap B)$ where A and $\bar{A} \cap B$ are the disjoint events. Then using Axiom 3 and Axiom 1, we get

$$P(B) = P(A) + P(\bar{A} \cap B) \geq P(A).$$

6.3.2 Calculation Rules for Probabilities

The introduced axioms and corollaries can be summarized as follows:

(1) $0 \leq P(A) \leq 1$
(2) $P(\Omega) = 1$
(3) $P(A_1 \cup A_2) = P(A_1) + P(A_2)$, if A_1 and A_2 are disjoint
(4) $P(\emptyset) = 0$
(5) $P(\bar{A}) = 1 - P(A)$
(6) $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$
(7) $P(A) \leq P(B)$, if $A \subseteq B$

6.4 Conditional Probability

Consider the following example to understand the concept of conditional probability: Suppose a new medical test is developed to diagnose a particular infection of the blood. The test is conducted on blood samples from 100 randomly selected patients and the outcomes of the tests are presented in Table 6.1.

There are the following four possible outcomes:

- The blood sample has an infection and the test diagnoses it, i.e. the test is correctly diagnosing the infection.
- The blood sample does not have an infection and the test does not diagnose it, i.e. the test is correctly diagnosing that there is no infection.
- The blood sample has an infection and the test does not diagnose it, i.e. the test is incorrect in stating that there is no infection.
- The blood sample does not have an infection but the test diagnoses it, i.e. the test is incorrect in stating that there is an infection.

Table 6.2 contains the relative frequencies of Table 6.1. In the following, we interpret the relative frequencies as probabilities, i.e. we assume that the values in Table 6.2 would be observed if the number n of patients was much larger than 100.

It can be seen that the probability that a test is positive is $P(T+) = 0.30 + 0.10 = 0.40$ and the probability that an infection is present is $P(IP) = 0.30 + 0.15 = 0.45$.

Table 6.1 Absolute frequencies of test results and infection status

		Infection		Total (row)
		Present	Absent	
Test	Positive (+)	30	10	40
	Negative (−)	15	45	60
	Total (column)	45	55	Total = 100

Table 6.2 Relative frequencies of patients and test

		Infection		Total (row)
		Present (IP)	Absent (IA)	
Test	Positive (+)	0.30	0.10	0.40
	Negative (−)	0.15	0.45	0.60
	Total (column)	0.45	0.55	Total = 1

If one already knows that the test is positive and wants to determine the probability that the infection is indeed present, then this can be achieved by the respective **conditional probability** $P(IP|T+)$ which is

$$P(IP|T+) = \frac{P(IP \cap T+)}{P(T+)} = \frac{0.3}{0.4} = 0.75.$$

Note that $IP \cap T+$ denotes the "relative frequency of blood samples in which the disease is present *and* the test is positive" which is 0.3.

More generally, recall Definition 4.1.1 from Chap. 4 where we defined conditional, joint, and marginal frequency distributions in contingency tables. The present example simply applies these rules to the contingency tables of relative frequencies and interprets the relative frequencies as an approximation to the probabilities of interest, as already explained.

We use the intersection operator $\cap$ to describe events which occur for $A = a$ *and* $B = b$. This relates to the joint relative frequencies. The marginal relative frequencies (i.e. probabilities $P(A = a)$) can be observed from the column and row sums, respectively; and the conditional probabilities can be observed as the joint frequencies in relation to the marginal frequencies.

For simplicity, assume that all simple events in $\Omega = \{\omega_1, \omega_2, \ldots, \omega_k\}$ are equally probable, i.e. $P(\omega_j) = \frac{1}{k}$, $j = 1, 2, \ldots, k$. Let A and B be two events containing n_A and n_B numbers of simple events. Let further $A \cap B$ contain n_{AB} numbers of simple events. The Laplace probability using (6.2) is

$$P(A) = \frac{n_A}{k}, \quad P(B) = \frac{n_B}{k}, \quad P(A \cap B) = \frac{n_{AB}}{k}.$$

Assume that we have prior information that A has already occurred. Now we want to find out how the probability of B is to be calculated. Since A has already occurred, we know that the sample space is reduced by the number of simple events which are contained in A. There are n_A such simple events. Thus, the total sample space Ω is reduced by the sample space of A. Therefore, the simple events in $A \cap B$ are those simple events which are realized when B is realized. The Laplace probability for B under the prior information on A, or under the condition that A is known, is therefore

$$P(B|A) = \frac{n_{AB}/k}{n_A/k} = \frac{P(A \cap B)}{P(A)}. \tag{6.5}$$

This can be generalized to the case when the probabilities for simple events are unequal.

Definition 6.4.1 Let $P(A) > 0$. Then the **conditional probability** of event B occurring, given that event A has already occurred, is

$$P(B|A) = \frac{P(A \cap B)}{P(A)}. \tag{6.6}$$

The roles of A and B can be interchanged to define $P(A|B)$ as follows. Let $P(B) > 0$. The conditional probability of A given B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{6.7}$$

We now introduce a few important theorems which are relevant to calculating conditional and other probabilities.

Theorem 6.4.1 (Multiplication Theorem of Probability) *For two arbitrary events A and B, the following holds:*

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A). \tag{6.8}$$

This theorem follows directly from the two definitions (6.6) and (6.7) (but does not require that $P(A) > 0$ and $P(B) > 0$).

Theorem 6.4.2 (Law of Total Probability) *Assume that $A_1, A_2, \ldots, A_m$ are events such that $\cup_{i=1}^m A_i = \Omega$ and $A_i \cap A_j = \emptyset$ for all $i \neq j$, $P(A_i) > 0$ for all i, i.e. A_1, $A_2, \ldots, A_m$ form a complete decomposition of $\Omega = \cup_{i=1}^m A_i$ in pairwise disjoint events, then the probability of an event B can be calculated as*

$$P(B) = \sum_{i=1}^{m} P(B|A_i)P(A_i). \tag{6.9}$$

6.4.1 Bayes' Theorem

Bayes' Theorem gives a connection between $P(A|B)$ and $P(B|A)$. For events A and B with $P(A) > 0$ and $P(B) > 0$, using (6.6) and (6.7) or (6.8), we get

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A \cap B)}{P(A)} \frac{P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} . \tag{6.10}$$

Let $A_1, A_2, \ldots, A_m$ be events such that $\cup_{i=1}^m A_i = \Omega$ and $A_i \cap A_j = \emptyset$ for all $i \neq j$, $P(A_i) > 0$ for all i, and B is another event than A, then using (6.9) and (6.10), we get

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_i P(B|A_i)P(A_i)} . \tag{6.11}$$

The probabilities $P(A_i)$ are called **prior probabilities**, $P(B|A_i)$ are sometimes called **model probabilities** and $P(A_j|B)$ are called **posterior probabilities**.

Example 6.4.1 Suppose someone rents movies from two different DVD stores. Sometimes it happens that the DVD does not work because of scratches. We consider the following events: A_i $(i = 1, 2)$: "the DVD is rented from store i". Further let B denote the event that the DVD is working without any problems. Assume we know that $P(A_1) = 0.6$ and $P(A_2) = 0.4$ (note that $A_2 = \bar{A}_1$) and $P(B|A_1) = 0.95$, $P(B|A_2) = 0.75$ and we are interested in the probability that a rented DVD works fine. We can then apply the Law of Total Probability and get

$$\begin{aligned} P(B) &\overset{(6.9)}{=} P(B|A_1)P(A_1) + P(B|A_2)P(A_2) \\ &= 0.6 \cdot 0.95 + 0.4 \cdot 0.75 = 0.87. \end{aligned}$$

We may also be interested in the probability that the movie was rented from store 1 *and* is working which is

$$P(B \cap A_1) \overset{(6.8)}{=} P(B|A_1)P(A_1) = 0.95 \cdot 0.6 = 0.57.$$

Now suppose we have a properly working DVD. What is the probability that it is rented from store 1? This is obtained as follows:

$$P(A_1|B) \overset{(6.7)}{=} \frac{P(A_1 \cap B)}{P(B)} = \frac{0.57}{0.87} = 0.6552.$$

Now assume we have a DVD which does not work, i.e. $\bar{B}$ occurs. The probability that a DVD is not working given that it is from store 1 is $P(\bar{B}|A_1) = 0.05$. Similarly,

$P(\bar{B}|A_2) = 0.25$ for store 2. We can now calculate the conditional probability that a DVD is from store 1 given that it is not working:

$$P(A_1|\bar{B}) \overset{(6.11)}{=} \frac{P(\bar{B}|A_1)P(A_1)}{P(\bar{B}|A_1)P(A_1) + P(\bar{B}|A_2)P(A_2)}$$
$$= \frac{0.05 \cdot 0.6}{0.05 \cdot 0.6 + 0.25 \cdot 0.4} = 0.2308.$$

The result about $P(\bar{B})$ used in the denominator can also be directly obtained by using $P(\bar{B}) = 1 - 0.87 = 0.13$.

6.5 Independence

Intuitively, two events are independent if the occurrence or non-occurrence of one event does not affect the occurrence or non-occurrence of the other event. In other words, two events A and B are independent if the probability of occurrence of B has no effect on the probability of occurrence of A. In such a situation, one expects that

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

Using this and (6.7), we can write

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$
$$= \frac{P(A \cap \bar{B})}{P(\bar{B})} = P(A|\bar{B}). \tag{6.12}$$

This yields:

$$P(A \cap B)P(\bar{B}) = P(A \cap \bar{B})P(B)$$
$$P(A \cap B)(1 - P(B)) = P(A \cap \bar{B})P(B)$$
$$P(A \cap B) = (P(A \cap \bar{B}) + P(A \cap B))P(B)$$
$$P(A \cap B) = P(A)P(B)\ . \tag{6.13}$$

This leads to the following definition of stochastic independence.

Definition 6.5.1 Two random events A and B are called **(stochastically) independent** if

$$P(A \cap B) = P(A)P(B)\ , \tag{6.14}$$

i.e. if the probability of simultaneous occurrence of both events A and B is the product of the individual probabilities of occurrence of A and B.

This definition of independence can be extended to the case of more than two events as follows:

Definition 6.5.2 The n events $A_1, A_2, \ldots, A_n$ are stochastically mutually independent, if for any subset of m events $A_{i_1}, A_{i_2}, \ldots, A_{i_m}$ $(m \leq n)$

$$P(A_{i_1} \cap A_{i_2} \cdots \cap A_{i_m}) = P(A_{i_1})P(A_{i_2}) \cdot \ldots \cdot P(A_{i_m}) \tag{6.15}$$

holds.

A weaker form of independence is pairwise independence. If condition (6.15) is fulfilled only for two arbitrary events, i.e. $m = 2$, then the events are called **pairwise independent**. The difference between pairwise independence and general stochastic independence is explained in the following example.

Example 6.5.1 Consider an urn with four balls. The following combinations of zeroes and ones are printed on the balls: 110, 101, 011, 000. One ball is drawn from the urn. Define the following events:

A_1 : The first digit on the ball is 1.

A_2 : The second digit on the ball is 1.

A_3 : The third digit on the ball is 1.

Since there are two favourable simple events for each of the events A_1, A_2 and A_3, we get

$$P(A_1) = P(A_2) = P(A_3) = \frac{2}{4} = \frac{1}{2}.$$

The probability that all the three events simultaneously occur is zero because there is no ball with 111 printed on it. Therefore, A_1, A_2, and A_3 are not stochastically independent because

$$P(A_1)P(A_2)P(A_3) = \frac{1}{8} \neq 0 = P(A_1 \cap A_2 \cap A_3).$$

However,

$$P(A_1 \cap A_2) = \frac{1}{4} = P(A_1)P(A_2)\,,$$
$$P(A_1 \cap A_3) = \frac{1}{4} = P(A_1)P(A_3)\,,$$
$$P(A_2 \cap A_3) = \frac{1}{4} = P(A_2)P(A_3)\,,$$

which means that the three events are pairwise independent.

6.6 Key Points and Further Issues

Note:

✓ We summarize some important theorems and laws:

- The Laplace probability is the ratio

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{Number of “favourable simple events” for } A}{\text{Total number of possible simple events}}.$$

- The Law of Total Probability is

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

- Bayes' Theorem is

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_i P(B|A_i)P(A_i)}.$$

- n events $A_1, A_2, \ldots, A_n$ are (stochastically) independent, if

$$P(A_1 \cap A_2 \cdots \cap A_n) = P(A_1)P(A_2) \cdot \ldots \cdot P(A_n).$$

✓ In Sect. 10.8, we present the χ^2-independence test, which can test whether discrete random variables (see Chap. 7) are independent or not.

6.7 Exercises

Exercise 6.1

(a) Suppose $\Omega = \{0, 1, \ldots, 15\}$, $A = \{0, 8\}$, $B = \{1, 2, 3, 5, 8, 10, 12\}$, $C = \{0, 4, 9, 15\}$. Determine $A \cap B$, $B \cap C$, $A \cup C$, $C \setminus A$, $\Omega \setminus (B \cup A \cup C)$.

(b) Now consider the three pairwise disjoint events E, F, G with $\Omega = E \cup F \cup G$ and $P(E) = 0.2$ and $P(F) = 0.5$. Calculate $P(\bar{F})$, $P(G)$, $P(E \cap G)$, $P(E \setminus E)$, and $P(E \cup F)$.

Exercise 6.2 A driving licence examination consists of two parts which are based on a theoretical and a practical examination. Suppose 25 % of people fail the practical examination, 15 % of people fail the theoretical examination, and 10 % of people fail both the examinations. If a person is randomly chosen, then what is the probability that this person

(a) fails at least one of the examinations?
(b) only fails the practical examination, but not the theoretical examination?
(c) successfully passes both the tests?
(d) fails any of the two examinations?

Exercise 6.3 A new board game uses a twelve-sided die. Suppose the die is rolled once, what is the probability of getting

(a) an even number?
(b) a number greater than 9?
(c) an even number greater than 9?
(d) an even number or a number greater than 9?

Exercise 6.4 The Smiths are a family of six. They are celebrating Christmas and there are 12 gifts, two for each family member. The name tags for each family member have been attached to the gifts. Unfortunately the name tags on the gifts are damaged by water. Suppose each family member draws two gifts at random. What is the probability that someone

(a) gets his/her two gifts, rather than getting the gifts for another family member?
(b) gets none of his/her gifts, but rather gets the gifts for other family members?

Exercise 6.5 A chef from a popular TV cookery show sometimes puts too much salt in his pumpkin soup and the probability of this happening is 0.2. If he is in love (which he is with probability 0.3), then the probability of using too much salt is 0.6.

(a) Create a contingency table for the probabilities of the two variables "in love" and "too much salt".
(b) Determine whether the two variables are stochastically independent or not.

Exercise 6.6 Dr. Obermeier asks his neighbour to take care of his basil plant while he is away on leave. He assumes that his neighbour does not take care of the basil with a probability of $\frac{1}{3}$. The basil dies with probability $\frac{1}{2}$ when someone takes care of it and with probability $\frac{3}{4}$ if no one takes care of it.

(a) Calculate the probability of the basil plant surviving after its owner's leave.
(b) It turns out that the basil eventually dies. What is the probability that Dr. Obermeier's neighbour did not take care of the plant?

Exercise 6.7 A bank considers changing its credit card policy. Currently 5 % of credit card owners are not able to pay their bills in any month, i.e. they never pay their bills. Among those who are generally able to pay their bills, there is still a 20 % probability that the bill is paid too late in a particular month.

(a) What is the probability that someone is not paying his bill in a particular month?
(b) A credit card owner did not pay his bill in a particular month. What is the probability that he never pays back the money?
(c) Should the bank consider blocking the credit card if a customer does not pay his bill on time?

Exercise 6.8 There are epidemics which affect animals such as cows, pigs, and others. Suppose 200 cows are tested to see whether they are infected with a virus or not. Let event A describe whether a cow has been transported by a truck recently or not and let B denote the event that a cow has been tested positive with a virus. The data are summarized in the following table:

	B	$\bar{B}$	Total
A	40	60	100
$\bar{A}$	20	80	100
Total	60	140	200

(a) What is the probability that a cow is infected and has been transported by a truck recently?
(b) What is the probability of having an infected cow given that it has been transported by the truck?
(c) Determine and interpret $P(B)$.

Exercise 6.9 A football practice target is a portable wall with two holes (which are the target) in it for training shots. Suppose there are two players A and B. The probabilities of hitting the target by A and B are 0.4 and 0.5, respectively.

(a) What is the probability that at least one of the players succeeds with his shot?
(b) What is the probability that exactly one of the players hits the target?
(c) What is the probability that only B scores?

→ Solutions to all exercises in this chapter can be found on p. 361

**Source* Toutenburg, H., Heumann, C., *Induktive Statistik*, 4th edition, 2007, Springer, Heidelberg