Probability and Statistics for ISP

Course Notes for Statistics 383

T. A. Severini

2025-12-31

Table of contents

Preface

This is a Quarto book.

To learn more about Quarto books visit https://quarto.org/docs/books.

```
1 + 1
```

```
[1] 2
```

1 Probability

1.1 Experiments

The starting point for probability theory is the concept of an *experiment*. The term experiment may actually refer to a physical experiment in the usual sense, but more generally we will refer to something as an experiment when it has the following properties:

- There is a well-defined set of possible outcomes of the experiment;
- Each time the experiment is performed exactly one of the possible outcomes occurs;
- The outcome that occurs is governed by some chance mechanism.

Let Ω denote the *sample space* of the experiment, the set of possible outcomes of the experiment; the term *outcome space* is also used. We will refer to the elements of Ω as *basic outcomes* and use the symbol ω to denote a generic basic outcome.

Consider the experiment in which we roll a die. Then

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

where, for example, 1 denotes the outcome that we roll a 1.

Consider the experiment in which we choose a number from the interval $(0, 1)$. Then $\Omega = (0, 1)$.

Suppose that we have an urn that contains three balls, two red balls and one black ball. Consider the experiment in which we successively choose two balls from the urn, that is, the balls are chosen in such a way that we know which ball was chosen first.

Then the sample space for the experiment can be written

$$\Omega = \{(R, R),\ (R, B),\ (B, R)\},$$

where, for example, (R, B) means that the first ball selected is red and the second ball selected is black.

Now suppose that the order in which the balls were selected is not recorded. Then the sample space of the experiment is given by

$$\Omega = \{\{R, R\},\ \{R, B\}\},$$

where, for example, $\{R, B\}$ means that one red ball is selected and one black ball is selected.

1.2 Events

Consider an experiment with sample space Ω. A subset A of Ω is called an *event*. Let A be an event. Then, for each $\omega \in \Omega$, either $\omega \in A$ or $\omega \notin A$. That is, when the experiment is performed, either A occurs (the observed outcome is in A) or it doesn't occur (the observed outcome is not in A).

Consider the experiment in which we roll a die. Then

$$A = \{2, 4, 6\}$$

is the event that we roll an even number.

The event that we roll a number less than or equal to 3 is given by

$$B = \{1, 2, 3\}.$$

The event that we roll a 5 is given by

$$C = \{5\}.$$

The event that we do not roll an even number, that is, that we roll an odd number is given by A^c, the complement of A. Thus,

$$A^c = \{1, 3, 5\}.$$

Consider the experiment in which two balls are drawn successively from an urn containing two red balls and one black ball; the sample space for this experiment is given in Example [urn3].

Let A denote the event that exactly one red ball is selected. Then

$$A = \{(R, B),\ (B, R)\}.$$

Because events are defined in terms of sets, sets play a central role in probability theory. Here are few basic properties. Let A, B, C be subsets of a sample space Ω; that is, let A, B, C be events. Recall that $A \cup B$, the *union* of A and B, is the set consisting of all elements that are either in A, in B, or in both A and B; $A \cap B$, the *intersection* of A and B, is the set consisting of all elements that are in both A and B. Then

- $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$
- $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$
- $(A \cup B)^c = A^c \cap B^c$
- $(A \cap B)^c = A^c \cup B^c$.

If these properties are unfamiliar, you can show why they hold using Venn diagrams. E.g., consider $(A \cup B)^c = A^c \cap B^c$. Figures 1.1 – 1.3 contain Venn diagrams of $(A \cup B)^c$, A^c, and B^c, respectively. From these diagrams, we can see that $(A \cup B)^c = A^c \cap B^c$.

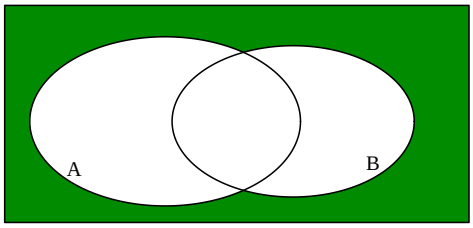

Figure 1.1: $(A \cup B)^c$

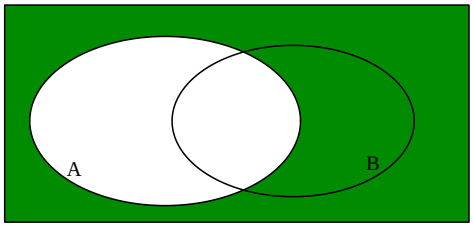

Figure 1.2: A^c

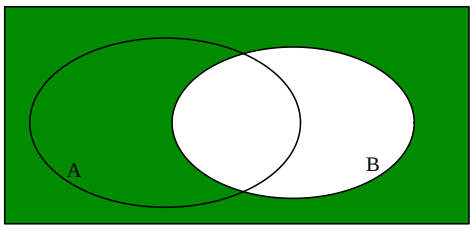

Figure 1.3: B^c

1.3 Probability Functions

Consider an experiment with sample space Ω; recall that the outcome of an experiment depends on some "chance mechanism". It follows that whether or not an event A occurs depends on that chance mechanism and we use probability theory to describe the likelihood that a given event occurs.

Therefore, associated with each event A is a probability $\mathrm{P}(A)$. Here P is a function defined on subsets of Ω and taking values in the interval $[0, 1]$. The function P is required to have certain properties:

- $\mathrm{P}(\Omega) = 1$
- If A and B are disjoint subsets of Ω, that is, $A \cap B = \emptyset$, then $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$.
- If $A_1, A_2, \ldots,$ are disjoint subsets of Ω, then

$$\mathrm{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathrm{P}(A_n).$$

Note that when subsets of Ω are disjoint, the corresponding events are said to be *mutually exclusive*.

Suppose that $\Omega = (0, 1)$ and suppose that the probability of any interval in Ω is the length of the interval. More generally, we may take the probability of a subset A of Ω to be

$$\mathrm{P}(A) = \int_A dx.$$

For example,

$$\mathrm{P}\left((0.2, 0.7)\right) = 0.5.$$

Consider the experiment of rolling one die, as discussed in Examples [onedie1] and [onedie2] and let Ω denote the sample space of the experiment.

For $A \subset \Omega$, let $\mathrm{P}(A) = |A|/6$, the number of elements in A, divided by 6.

E.g., the probability of rolling an even number is $1/2$ and the probability of rolling a number greater than or equal to 5 is $1/3$.

Note that, when an event consists of a single basic outcome ω, we will write the probability of the event as $\mathrm{P}(\omega)$, rather than as $\mathrm{P}(\{\omega\})$, which is technically correct (because the argument of the probability function should be a set).

For instance, in the previous example, the probability of rolling a 6 will be written as $\mathrm{P}(6)$ instead of as $\mathrm{P}(\{6\})$.

When Ω is a countable set, then, by properties (P2) and (P3), the probability of any event is given by the sum of the probabilities of the basic outcomes corresponding to the event:

$$\mathrm{P}(A) = \sum_{\omega \in A} \mathrm{P}(\omega).$$

Consider an experiment with sample space

$$\Omega = \{(0,0),\ (1,0),\ (0,1),\ (1,1)\}.$$

For $\omega = (x_1, x_2) \in \Omega$, take

$$\begin{aligned}\mathrm{P}(\omega) &= \theta^{x_1}(1-\theta)^{1-x_1}\theta^{x_2}(1-\theta)^{1-x_2} \\ &= \theta^{x_1+x_2}(1-\theta)^{2-x_1-x_2}\end{aligned}$$

where $0 < \theta < 1$ is a given constant.

Thus, the four elements of Ω have probabilities $(1-\theta)^2, \theta(1-\theta), \theta(1-\theta), \theta^2$, respectively.

Let A denote the event that exactly 1 one is observed; then

$$A = \{(0,1),\ (1,0)\}.$$

It follows that the probability of A is the sum of the probabilities of the two basic outcomes in A:

$$\mathrm{P}(A) = \mathrm{P}\left((1,0) \cup (0,1)\right) = \mathrm{P}\left((1,0)\right) + \mathrm{P}\left((0,1)\right) = \theta(1-\theta) + \theta(1-\theta) = 2\theta(1-\theta).$$

Some implications of (P1) – (P3)

There are a number of straightforward consequences of properties (P1)-(P3). For instance, because $\Omega \cup \emptyset = \Omega$ and $\Omega \cap \emptyset = \emptyset$, by (P2)

$$\mathrm{P}(\Omega) = \mathrm{P}(\Omega) + \mathrm{P}(\emptyset);$$

it follows that $\mathrm{P}(\emptyset) = 0$.

Let A^c denote the complement of an set $A \subset \Omega$. Then $A \cup A^c = \Omega$ and $A \cap A^c = \emptyset$. It follows from (P2) that

$$\mathrm{P}(\Omega) = \mathrm{P}(A) + \mathrm{P}(A^c);$$

it now follows from (P1) that

$$\mathrm{P}(A^c) = 1 - \mathrm{P}(A).$$

Suppose that A_1 and A_2 are subsets of Ω that are not necessarily disjoint. Then

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

This important result is a little more difficult to prove than the others we have considered. First note that $A_1 \cup A_2$ can be written as the union of three sets, $A_1 \cap A_2$, $A_1 \cap A_2^c$, and $A_1^c \cap A_2$. Furthermore, these three sets are disjoint. An example of this fact is given in the Venn diagram in Figure 1.4. In that diagram, the blue region is $A_1 \cap A_2^c$, the yellow region is $A_1^c \cap A_2$, and the brown region is $A_1 \cap A_2$; combining these three regions forms $A_1 \cup A_2$.

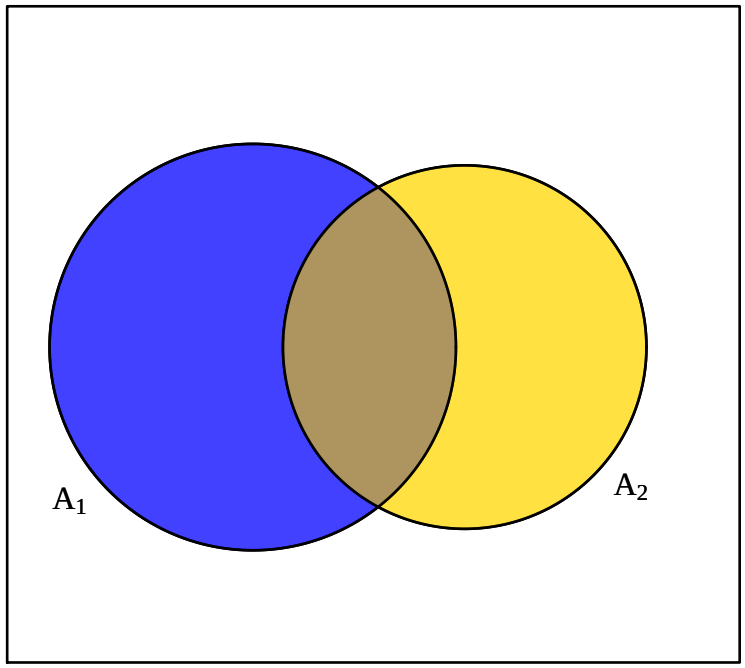

Figure 1.4: Venn Diagram Used to Illustrate $A_1 \cup A_2 = (A_1 \cap A_2) \cup (A_1 \cap A_2^c) \cup (A_1^c \cap A_2)$

From the Venn diagram we can also see that

$$A_1 = (A_1 \cap A_2^c) \cup (A_1 \cap A_2) \quad \text{and} \quad A_2 = (A_1^c \cap A_2) \cup (A_1 \cap A_2).$$

It follows from two applications of (P2) that

$$\mathrm{P}(A_1 \cup A_2) = \mathrm{P}(A_1 \cap A_2) + \mathrm{P}(A_1 \cap A_2^c) + \mathrm{P}(A_1^c \cap A_2)$$

and also that

$$\mathrm{P}(A_1) = \mathrm{P}(A_1 \cap A_2^c) + \mathrm{P}(A_1 \cap A_2) \quad \text{and} \quad \mathrm{P}(A_2) = \mathrm{P}(A_1^c \cap A_2) + \mathrm{P}(A_1 \cap A_2).$$

From the last two equations, we see that

$$\mathrm{P}(A_1 \cap A_2^c) = \mathrm{P}(A_1) - \mathrm{P}(A_1 \cap A_2) \quad \text{and} \quad \mathrm{P}(A_1^c \cap A_2) = \mathrm{P}(A_2) - \mathrm{P}(A_1 \cap A_2).$$

Substituting these expressions into the right-hand side of equation [probeq1], it follows that

$$\begin{aligned}\mathrm{P}(A_1 \cup A_2) &= \mathrm{P}(A_1 \cap A_2) + \mathrm{P}(A_1) - \mathrm{P}(A_1 \cap A_2) + \mathrm{P}(A_2) - \mathrm{P}(A_1 \cap A_2) \\ &= \mathrm{P}(A_1) + \mathrm{P}(A_2) - \mathrm{P}(A_1 \cap A_2).\end{aligned}$$

Consider an experiment with sample space Ω and events A_1, A_2. Let $A_1 \setminus A_2$ denote the elements of A_1 that are not in A_2.

Suppose that $A_2 \subset A_1$. Then

$$\mathrm{P}(A_1 \setminus A_2) = \mathrm{P}(A_1) - \mathrm{P}(A_2).$$

In general,

$$A_1 = (A_1 \cap A_2) \cup (A_1 \cap A_2^c) = (A_1 \cap A_2) \cup (A_1 \setminus A_2).$$

Note that

$$A_1 \cap A_2 \quad \text{and} \quad A_1 \cap A_2^c$$

are disjoint and that

$$A_1 \cap A_2^c = A_1 \setminus A_2.$$

Hence,

$$\mathrm{P}(A_1) = \mathrm{P}(A_1 \cap A_2) + \mathrm{P}(A_1 \setminus A_2)$$

so that

$$\mathrm{P}(A_1 \setminus A_2) = \mathrm{P}(A_1) - \mathrm{P}(A_1 \cap A_2).$$

Interpretation of probability

Although we have described the properties of a probability function, nothing has been said about what the probability function is measuring. In a mathematical sense, that is irrelevant – a probability function is defined by its properties and any function satisfying those properties can be used to calculate a "probability".

However, in order to better understand the mathematical results, and to develop some intuition regarding probability theory, it is useful to have some notion of what is meant by "probability". Several different interpretations of probability are used in applications. The most common, and the one we will use here, is the interpretation of probability as a"limiting relative frequency".

Consider an experiment with sample space Ω and let A denote an event. According to the limiting relative frequency interpretation of probability, the statement that A has probability

0.4 (for example), means that if the experiment is repeated a large number of times then in about 40% of those experiments the event A will occur.

More formally, let $N_n(A)$ denote the number of times the event A occurs in n repetitions of the experiment. Then

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

the right-hand side of this expression is often described as a "limiting relative frequency".

1.4 Sampling from a Finite Population

A particularly simple, but useful, case occurs when the sample space of the experiment, Ω, is a finite set and each $\omega \in \Omega$ has the same probability.

Write

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_m\};$$

note that any subset of Ω can be written as the union of sets of the form $\{\omega_j\}$, which are disjoint. Let $c = \mathrm{P}(\omega_j)$ denote the common value of the probability of each element of Ω. Then, because $\mathrm{P}(\Omega) = 1$ and

$$\mathrm{P}(\Omega) = \mathrm{P}(\omega_1) + \mathrm{P}(\omega_2) + \cdots + \mathrm{P}(\omega_m) = mc$$

we must have $c = 1/|\Omega|$ where $|\Omega|$ denotes the cardinality of Ω, that is, the number of elements in Ω.

Furthermore, for any $A \subset \Omega$,

$$\mathrm{P}(A) = \sum_{\omega\in A} \mathrm{P}(\omega) = \frac{|A|}{|\Omega|}.$$

Thus, the problem of determining $\mathrm{P}(A)$ is essentially the problem of counting the number of elements in the set A and the number of elements in Ω. The subfield of mathematics concerned with counting the number of elements in a set is known as *combinatorics*.

In some cases, such as the one in the following example, the counting needed is relatively straightforward.

Consider the experiment of rolling 2 dice, one at a time. The sample space of the experiment can be written

$$\Omega = \{(1,1), (1,2), \dots, (1,6), (2,1), \dots, (2,6), \dots, (6,1), \dots, (6,6)\}$$

so that it has 36 elements. Suppose the dice are "fair", in the sense that each element of Ω is equally likely.

Let A denote the event that result of the dice rolling is "doubles", i.e., the two numbers rolled are equal, and suppose that we are interested in the probability of A.

As noted previously, the sample space Ω has 36 elements. The event A has 6 elements, $(1,1),(2,2),\dots,(6,6)$. Thus, the probability of rolling doubles is $6/36 = 1/6$.

In other cases, the counting needed is more complicated and it is useful to apply one of the many well-known results that are used to solve such counting problems. Here we consider only a few simple ones.

Counting principle

Many results in combinatorics are based on the *counting principle*. Let A and B denote finite sets and let $A \times B$ denote the Cartesian product of A and B, that is, the set of the form

$$A \times B = \{(a,b) : \ a \in A, \ b \in B\}.$$

Then

$$|A \times B| = |A| \ |B|.$$

Thus, if a task can be completed in r stages and there are n_j ways to complete the jth stage, $j = 1, 2, \dots, r$, then there are

$$n_1 \times n_2 \times \cdots \times n_r$$

ways to complete the task.

Permutations and combinations

A *permutation* of n distinct objects is an ordering of them.

The possible permutations of a, b, c are

$$abc, \ bac, \ cab, \ acb, \ bca, \ cba.$$

The number of permutations of n distinct objects can be found using the counting principle, breaking the problem into stages. Note that there are n ways to choose the first object, $n-1$ ways to choose the second object, and so on, until there is only 1 way to choose the last object. Hence, there are

$$n \cdot (n-1) \cdots (2) \cdot (1) = n!$$

ways to order the n objects. That is, there are $n!$ possible permutations of n distinct objects.

In some cases, we might be interested in ordered *samples* from a given set.

The possible ordered samples of size 2 from the set $\{a, b, c\}$ are

$$(a,b), \ (a,c), \ (b,a), \ (b,c), \ (c,a), \ (c,b).$$

The number of possible ordered samples of size k from a set of n distinct elements can be found using the counting principle. There are n ways to choose the first element, $n-1$ ways to choose the second element, and so on. However, in contrast to permutations, here we stop selecting elements after the kth selection. Therefore, there

$$n(n-1)\cdots(n-k+1)$$

ordered samples of size k from a set of n elements; note there are k terms in this product. The expression $n(n-1)\cdots(n-k+1)$ is often denoted by $(n)_k$; it can also be written as

$$(n)_k = \frac{n!}{(n-k)!}.$$

Consider an urn containing 5 balls, 2 of which are black and 3 of which are red. Suppose that 2 balls are randomly selected from the urn, without replacement; that is, after the first ball is selected, it is not returned to the urn for the second selection. Thus, there are $(5)_2 = 5(4) = 20$ basic outcomes in Ω; "randomly selected" means that any ordered pair of 2 balls is equally likely to be selected.

Let A denote the event that a red ball is selected followed by a black ball. Find $\mathrm{P}(A)$. Because each basic outcome is assumed to have the same probability,

$$\mathrm{P}(A) = \frac{|A|}{|\Omega|}.$$

Thus, we need to count the number of basic outcomes in of A.

To find the number of basic outcomes in A, we use the facts that there are 3 ways to choose the first (red) ball and 2 ways to choose the second (black) ball. Thus, there are $(3)_2 = 3(2) = 6$ basic outcomes in A. It follows that

$$\mathrm{P}(A) = \frac{6}{20} = \frac{3}{10}.$$

Now suppose that we are interested in the possible *combinations* of k objects chosen from a set of n distinct objects. When considering combinations, the order of the objects is irrelevant; we are interested only in the set of k objects.

The possible combinations of 2 elements chosen from the set $\{a, b, c\}$ are given by

$$\{a, b\},\ \{a, c\},\ \{b, c\}.$$

The number of possible combinations of k objects chosen from a set of n distinct objects is denoted by

$$\binom{n}{k},$$

read as "n choose k".

To find an expression for n choose k, consider choosing an **ordered** sample of size k from n distinct elements.

This can be done in two steps:

1. choose k elements from n
2. order the k elements

We know that

- there are $(n)_k$ ordered samples of size k from n distinct elements
- there are n choose k ways to choose k elements from n
- there are $k!$ ways to order k elements

Therefore, we must have

$$(n)_k = \binom{n}{k} k!$$

so that

$$\binom{n}{k} = \frac{(n)_k}{k!} = \frac{n!}{k!\,(n-k)!}.$$

Terms of the form $\binom{n}{k}$ are often called the *binomial coefficients*, because of the binomial formula:

$$(x+y)^n = \sum_{j=0}^{n} \binom{n}{j} x^j y^{n-j}$$

for all real numbers x, y and all positive integers n.

Consider the framework of Example [urn0]: there is an urn with 2 black balls and 3 red balls and 2 balls are randomly selected from the urn. Let B denote the event that 2 red balls are selected. Find $\mathrm{P}(B)$.

The basic outcomes here are the sets of two balls selected from the urn. Because each basic outcome is assumed to have the same probability,

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

The number of elements in Ω is the number of ways to select 2 balls from the set of 5, given by

$$\binom{5}{2} = 10.$$

The number of elements in B is the number of ways to choose 2 red balls from the set of 3 red balls, given by

$$\binom{3}{2} = 3.$$

Thus, $\mathrm{P}(B) = 3/10$.

Suppose that 5 cards are dealt from a well-shuffled deck of playing cards. Recall that, in such a deck, there are 52 cards and each card falls into one of four suits (13 cards in each suit).

What is the probability that all 5 cards are of the same suit (i.e., a "flush", in poker)?

There are

$$\binom{52}{5}$$

ways to choose 5 cards from a deck of 52. To find the number of ways in which 5 cards can be chosen from one suit, we can use the counting principle: there are 4 ways to choose the suit and, given the suit, there are

$$\binom{13}{5}$$

ways to choose the 5 cards from the suit. Therefore, there are

$$4\binom{13}{5}$$

ways to choose 5 cards from one suit.

It follows that the probability of being dealt 5 cards from one suit is

$$\begin{aligned}
\frac{4\binom{13}{5}}{\binom{52}{5}} &= \frac{4\frac{13!}{8!\,5!}}{\frac{52!}{47!\,5!}} \\
&= 4\frac{(13)(12)(11)(10)(9)}{(52)(51)(50)(49)(48)} \\
&= 0.00198.
\end{aligned}$$

In some cases, it is easier to find the probability of an event A by finding the probability of A^c and using the fact that $\mathrm{P}(A) = 1 - \mathrm{P}(A^c)$. In fact, this simple result often converts a complicated problem into a relatively easy one.

Suppose that n cards are dealt from a well-shuffled deck of playing cards. What is the probability that at least 1 face card is drawn?

We can calculate the probability of being dealt at least one face card by calculating the probability of being dealt no face cards and then subtracting that result from 1.

In a standard deck of cards, there are 12 face cards and 40 non-face cards.

To be dealt n non-face cards, n cards must be chosen from the 40 non-face cards. There are

$$\binom{40}{n}$$

ways to do this. Since there are

$$\binom{52}{n}$$

ways to choose n cards from the entire deck, the probability of being dealt no face cards is

$$\frac{\binom{40}{n}}{\binom{52}{n}}$$

and, hence, the probability of being dealt at least one face card is

$$1 - \frac{\binom{40}{n}}{\binom{52}{n}} = 1 - \frac{40 \cdot 39 \cdots (40 - n + 1)}{52 \cdot 51 \cdots (52 - n + 1)}.$$

This result holds for $n \leq 40$; otherwise the probability is 0.

1.5 Conditional Probability

Consider the dice-rolling experiment discussed in Example [dice1]: two dice are rolled, one at a time. The sample space of the experiment, Ω, has 36 elements,

$$\Omega = \{(1,1), (1,2), \ldots, (1,6), (2,1), \ldots, (2,6), \ldots, (6,1), \ldots, (6,6)\}$$

and each element of Ω is equally likely.

Let A denote the event result of the experiment includes at least 1 six; then

$$A = \{(1,6), (2,6), (3,6), (4,6), (5,6), (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)\},$$

which has 11 elements. Hence, $\mathrm{P}(A) = 11/36$.

Now suppose that we know that the sum of the dice is at least 10. Based on this information, what is the probability that the result includes at least 1 six?

Note that this probability cannot be described in terms of a single event, because it includes the condition that the sum of the dice is at least 10. It is an example of a *conditional probability.*

Let B denote the event that the sum of the dice is at least 10; then

$$B = \{(4,6), (5,5), (5,6), (6,4), (6,5), (6,6)\}.$$

We want to find $\mathrm{P}(A \mid B)$, read as the "conditional probability of A given B". It is the probability that we roll at least 1 six **given that** the sum of the dice is at least 10.

There are 6 elements in B; for 5 of these, there is at least one six. Therefore, it is reasonable to expect that $\mathrm{P}(A \mid B) = 5/6$.

This is, in fact, correct. The general formula for a conditional probability is

$$\mathrm{P}(A \mid B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)},$$

provided that $\mathrm{P}(B) > 0$. Note that $A \cap B$ represents the part of A that satisfies the condition B.

In the example,

$$A \cap B = \{(4,6), (5,6), (6,4), (6,5), (6,6)\}$$

so that $\mathrm{P}(A \cap B) = 5/36$. Using the fact that $\mathrm{P}(B) = 6/36$ yields the answer given above.

Conditional probabilities are useful because they allow us to incorporate additional information into the probability calculation.

Note that, for a given event B with $\mathrm{P}(B) > 0$, the function Q defined on subsets of Ω and given by

$$\mathrm{Q}(A) = \mathrm{P}(A \mid B)$$

is a probability function on Ω, in the sense that it satisfies all the properties of a probability function, such as

$$\mathrm{P}(A^c \mid B) = 1 - \mathrm{P}(A \mid B)$$

and

$$\mathrm{P}(A_1 \cup A_2 \mid B) = \mathrm{P}(A_1 \mid B) + \mathrm{P}(A_2 \mid B) - \mathrm{P}(A_1 \cap A_2 \mid B).$$

Consider an urn with 2 red balls and w white balls for some $w \geq 2$. Suppose that 2 balls are randomly selected from the urn. Given that the balls are the same color, what is the probability that they are red?

Define two events, A, the event that both balls are red and B, the event that the balls are the same color. Hence, we want to determine $\mathrm{P}(A \mid B)$.

There are

$$\binom{2+w}{2}$$

ways to choose 2 balls from the urn, which contains $2 + w$ balls. There is 1 way to choose 2 red balls and

$$\binom{w}{2}$$

ways to choose 2 white balls. Hence,

$$\mathrm{P}(B) = \frac{1 + \binom{w}{2}}{\binom{2+w}{2}} = \frac{w(w-1) + 2}{(w+2)(w+1)}$$

and

$$\mathrm{P}(B) = \frac{1}{\binom{2+w}{2}} = \frac{2}{(w+2)(w+1)}.$$

Note that, because $A \subset B$, $A \cup B = A$. It follows that

$$\begin{aligned}\mathrm{P}(A \,|\, B) &= \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = \frac{\mathrm{P}(A)}{\mathrm{P}(B)} \\ &= \frac{\frac{2}{(w+2)(w+1)}}{\frac{w(w-1)+2}{(w+2)(w+1)}} \\ &= \frac{2}{w(w-1)+2}.\end{aligned}$$

Multiplication law

Rewriting the expression for conditional probability yields the *multiplication law* for probabilities: for events A, B,

$$\mathrm{P}(A \cap B) = \mathrm{P}(B \,|\, A)\mathrm{P}(A) = \mathrm{P}(A \,|\, B)\mathrm{P}(B).$$

Consider an urn with r red balls and b black balls, where r and b are positive integers. Suppose that 2 balls are randomly selected from the urn, one at a time. What is the probability that the first ball is red and the second ball is black?

Define two events, A, the event that the first ball is red and B, the event that the second ball is black. We want $\mathrm{P}(A \cap B)$.

This is a case in which the expression $\mathrm{P}(B \,|\, A)\mathrm{P}(A)$ may be a convenient way to calculate $\mathrm{P}(A \cap B)$: the probability that the first ball is red is easy to determine,

$$\mathrm{P}(A) = \frac{r}{r+b},$$

and, given the result on the first ball, the probability that the second ball is black is also easy to determine.

Specifically, if the first ball is red, that leaves $r-1$ red balls and b black balls in the urn. Hence,

$$\mathrm{P}(B \,|\, A) = \frac{b}{r-1+b}.$$

It follows that

$$\mathrm{P}(A \cap B) = \frac{r}{r+b}\,\frac{b}{r-1+b}.$$

Independent events

Roughly speaking, events A and B are said to be *independent* if the occurrence of one event does not affect the probability of the other, in the sense that

$$\mathrm{P}(A \mid B) = \mathrm{P}(A) \quad \text{and} \quad \mathrm{P}(B \mid A) = \mathrm{P}(B).$$

Using the multiplication law, these can be written

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

which is taken as the definition of independence.

Consider the example of rolling two fair dice, one at a time. Let A denote the event that result includes at least 1 six and B denote the event that the sum is at least 10. Then, as we have seen in the example at the beginning of this section, $\mathrm{P}(A) = 11/36$, $\mathrm{P}(B) = 6/36$, and $\mathrm{P}(A \cap B) = 5/36$. Thus, because

$$\frac{5}{36} \neq \frac{11}{36}\frac{6}{36},$$

A and B are not independent events. That is, knowing that the sum is of the dice is at least 10 affects the probability that the result includes at least 1 six. Or, alternatively, knowing that the result includes at least 1 six affects the probablity that the sum is at least 10.

Now consider a third event, C, which denotes the event that the first die is a 4:

$$C = \{(4,1),(4,2),(4,3),(4,4),(4,5),(4,6))\,.$$

Hence $\mathrm{P}(C) = 1/6$.

Note that

$$A \cap C = \{(4,6)\} \quad \text{and} \quad B \cap C = \{(4,6)\}\,.$$

Hence, both $\mathrm{P}(A \cap C)$ and $\mathrm{P}(B \cap C)$ are $1/36$. Because

$$\frac{1}{36} \neq \frac{11}{36}\frac{1}{6},$$

it follows that A and C are not independent. However,

$$\mathrm{P}(B \cap C) = \frac{1}{36} = \frac{6}{36}\frac{1}{6} = \mathrm{P}(B)\mathrm{P}(C)$$

so that B and C are independent events.

Thus, knowing that the first die is a 4 does not affect the probability that the sum is at least 10. Note that 1/6 of the results in which the first die is a 4 have a sum of at least 10 and 1/6 of all results have a sum of at least 10.

On the other hand, knowing that the first die is a 4 changes the probability that the result includes at least one 6 from the unconditional probability of 11/36 to the conditional probability of

$$\mathrm{P}(A \,|\, C) = \frac{\mathrm{P}(A \cap C)}{\mathrm{P}(C)} = \frac{1/36}{1/6} = \frac{1}{6}.$$

The concept of independence can be extended to an arbitrary number of events. First consider three events, A_1, A_2, A_3. These events are said to be independent if

$$\mathrm{P}(A_1 \cap A_2 \cap A_3) = \mathrm{P}(A_1)\mathrm{P}(A_2)\mathrm{P}(A_3), \quad P(A_1 \cap A_2) = \mathrm{P}(A_1)\mathrm{P}(A_2),$$

$$P(A_1 \cap A_3) = \mathrm{P}(A_1)\mathrm{P}(A_3) \quad \text{and} \quad P(A_2 \cap A_3) = \mathrm{P}(A_2)\mathrm{P}(A_3).$$

More generally, a set of n events $A_1, A_2, \dots A_n$ is independent if the probability of the intersection of any subset of the events is the product of the probabilities of the events in the subset. That is, for any integer k, $1 \leq k \leq n$ and indices $1 \leq i_1 < i_2 < \cdots < i_k \leq n$,

$$\mathrm{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = \mathrm{P}(A_{i_1}) \cdots \mathrm{P}(A_{i_k}).$$

(Bernoulli trials) Consider an experiment with a sample space consisting of two basic outcomes, ω_1 and ω_2. Then there are four possible events, $\emptyset$, $\Omega = \{\omega_1, \omega_2\}$, $\{\omega_1\}$, and $\{\omega_2\}$. Thus, the probability function can be described by a single number, $\mathrm{P}(\{\omega_1\})$; it follows that $\mathrm{P}(\{\omega_2\}) = 1 - \mathrm{P}(\{\omega_1\})$. An experiment of this form is known as a *Bernoulli trial*.

Now consider an experiment consisting of n independent replications of the experiment. The sample space for this second experiment is given by

$$\Omega_n = \Omega \times \Omega \times \cdots \times \Omega.$$

Thus, an element of Ω_n can be written $(\omega_{i_1}, \omega_{i_2}, \dots \omega_{i_n})$, where $i_1, i_2, \dots, i_n$ each take values in the set $\{1, 2\}$.

Let P_n denote the probability function of the experiment. The term "independent replications" refers to the fact that

$$\mathrm{P}_n\left(\{(\omega_{i_1}, \dots, \omega_{i_n})\}\right) = \mathrm{P}(\{\omega_{i_1}\})\mathrm{P}(\{\omega_{i_2}\}) \cdots \mathrm{P}(\{\omega_{i_n}\}).$$

The experiment with sample space Ω_n and probability function P_n is known as a *sequence of Bernoulli trials*.

This is a generalization of the scenario considered in Example [binom_ex], in which $n = 2$ and ω_1 and ω_2 were denoted by 0 and 1, respectively.

The partition theorem

Consider an experiment with sample space Ω. A *partition* of Ω is a collection of disjoint subsets of Ω, $\{A_1, A_2, ...\}$ such that

$$\bigcup_i A_i \equiv A_1 \cup A_2 \cup \cdots = \Omega.$$

Such a partition can be either finite or infinite.

Note that any set $B \subset \Omega$ can be written

$$B = B \cap \Omega = B \cap \left(\bigcup_i A_i \right) = \bigcup_i B \cap A_i;$$

furthermore,

$$B \cap A_1,\ B \cap A_2,\ ...$$

are disjoint. It follows that

$$\mathrm{P}(B) = \sum_i \mathrm{P}(B \cap A_i).$$

Furthermore, if $\mathrm{P}(A_i) > 0$ for all i, then

$$\mathrm{P}(B) = \sum_i \mathrm{P}(B \,|\, A_i)\mathrm{P}(A_i).$$

The results given in [part1] and [part2] are known as the *partition theorem*; the term *law of total probability* is also used.

Consider an urn with r red balls and b black balls. Suppose that 2 balls are randomly selected from the urn, one at a time. What is the probability that the second ball is black?

We analyzed this experiment in Example [urn1]. There we saw that the probability that the second ball is black is easy to calculate if we know the result of the first ball. This suggests that it may be convenient to use the partition theorem.

Define two events, A, the event that the first ball is red and B, the event that the second ball is black. We have seen that

$$\mathrm{P}(A) = \frac{r}{r+b} \quad \text{and} \quad \mathrm{P}(B \,|\, A) = \frac{b}{r-1+b}.$$

The same basic argument can be used to find $\mathrm{P}(B \,|\, A^c)$: if A^c occurs, that is, if the first ball is black, then, when the second ball is chosen there are r red balls and $b-1$ black balls in the urn so that

$$\mathrm{P}(B \,|\, A^c) = \frac{b-1}{r+b-1}.$$

Hence, we can use the partition theorem with $A_1 = A$ and $A_2 = A^c$:

$$\begin{aligned}
\mathrm{P}(B) &= \mathrm{P}(B \mid A)\mathrm{P}(A) + \mathrm{P}(B \mid A^c)\mathrm{P}(A^c) \\
&= \frac{b}{r-1+b}\,\frac{r}{r+b} + \frac{b-1}{r-1+b}\,\frac{b}{r+b} \\
&= \frac{br + (b-1)b}{(r+b)(r+b-1)} \\
&= \frac{b}{r+b}.
\end{aligned}$$

Bayes' Theorem

Consider an experiment and let A, B be events. In some cases, information is available regarding $\mathrm{P}(A \mid B)$ but we are interested in $\mathrm{P}(B \mid A)$. Fortunately, the two conditional probabilities are related through $\mathrm{P}(A \cap B)$:

$$\mathrm{P}(A \cap B) = \mathrm{P}(A \mid B)\mathrm{P}(B) = \mathrm{P}(B \mid A)\mathrm{P}(A).$$

It follows that

$$\mathrm{P}(B \mid A) = \frac{\mathrm{P}(A \mid B)\mathrm{P}(B)}{\mathrm{P}(A)},$$

provided that $\mathrm{P}(A) > 0$. This result is known as *Bayes' Theorem*.

Bayes' Theorem is often used in conjuction with the partition theorem so that, for example,

$$\mathrm{P}(B \mid A) = \frac{\mathrm{P}(A \mid B)\mathrm{P}(B)}{\mathrm{P}(A \mid B)\mathrm{P}(B) + \mathrm{P}(A \mid B^c)\mathrm{P}(B^c)}.$$

More generally, if $B_1, B_2, \ldots,$ is a partition of Ω, then

$$\mathrm{P}(B_j \mid A) = \frac{\mathrm{P}(A \mid B_j)\mathrm{P}(B_j)}{\sum_i \mathrm{P}(A \mid B_i)\mathrm{P}(B_i)}.$$

A classic example of the use of Bayes' Theorem is in the analysis of a diagnostic test.

Consider a medical diagnostic test for some specified disease. Suppose it is known that

1. 5% of all patients who take the test have the disease
2. the *specificity* of the test is 0.99. That is, a patient known to not have the disease has a 99% chance of a negative test
3. the *sensitivity* of the test is 0.98. That is, a patient known to have the disease has a 98% chance of a positive test

If a particular patient has a positive result on the test, what is the probability that the patient has the disease?

Define two events, A, the event that the patient has the disease and B, the event that the patient's test is positive. The facts (1) - (3) given above can be rewritten in probability notation:

1. $\mathrm{P}(A) = 0.05$
2. $\mathrm{P}(B^c \mid A^c) = 0.99$
3. $\mathrm{P}(B \mid A) = 0.98$

We want to find $\mathrm{P}(A \mid B)$.

Using Bayes' Theorem, together with the partition theorem,

$$\begin{aligned}\mathrm{P}(A \mid B) &= \frac{\mathrm{P}(B \mid A)\mathrm{P}(A)}{\mathrm{P}(B \mid A)\mathrm{P}(A) + \mathrm{P}(B \mid A^c)\mathrm{P}(A^c)} \\ &= \frac{(0.98)(0.05)}{(0.98)(0.05) + (0.01)(0.95)} \\ &= 0.838.\end{aligned}$$

Consider an urn with 3 red balls and 2 black balls and consider the experiment in which two balls are chosen from the urn without replacement. Let A denote the event that the first ball selected is a red ball and let B denote the event that the second ball selected is a red ball. Find $\mathrm{P}(A \mid B)$.

In this example, probabilities that are conditonal on A, that is, conditional on the outcome of the first ball, are relatively easy to calculate. For instance, if the first ball is red, then the probability that the second ball is red is $2/4 = 1/2$; if the first ball is black, then the probability that the second ball is red is $3/4$. Thus,

$$\mathrm{P}(B \mid A) = 1/2 \quad \text{and} \quad \mathrm{P}(B \mid A^c) = 3/4$$

and, using the fact that $\mathrm{P}(A) = 3/5$, it follows from the law of total probability that

$$\mathrm{P}(B) = \mathrm{P}(B \mid A)\mathrm{P}(A) + \mathrm{P}(B \mid A^c)\mathrm{P}(A^c) = \frac{1}{2}\frac{3}{5} + \frac{3}{4}\frac{2}{5} = \frac{3}{5}.$$

To find $\mathrm{P}(A \mid B)$ we can use Bayes' Theorem:

$$\mathrm{P}(A \mid B) = \frac{\mathrm{P}(B \mid A)\mathrm{P}(A)}{\mathrm{P}(B)} = \frac{(1/2)(3/5)}{3/5} = \frac{1}{2}.$$