Chapter 2

Probability Theory

Everybody speaks of probability,
but no one is able to say what
it is, in a way which is
satisfactory for others.

(Garrett Birkhoff)

2.1 Introduction

In this chapter, we turn to models of discrete data. We refer to discrete data when a measurement's output yields a *natural* number. For example, let us consider the following problems:

- We throw a die 300 times. How many times does the number 4 occur?
- We randomly select 100 men. How many of them are more than 173 cm tall?

In these cases, the measurement is taken once. If the measurement is taken a second time, then the results differ from those of the first measurement. In the second example, if the first sample randomly includes many gymnasts and the second measurement randomly involves many basketball players, then there will very probably be differences in the measured number of individuals more than 173 cm tall. But if the measurement from the second example is repeated 1000 times, then the number of individuals taller than 173 cm to be *expected* should emerge.

But can we theoretically predict how large the value of a future measurement will be? To make such a prediction, we need *models* for the measurements. We assume the die we toss is "fair". Then we expect that with 1.2 million tosses, the number of times 4 appears is about 200.000. But it is absolutely possible, albeit very unlikely, that in all these tosses, we never toss a 4. So, we have made an idealization, which in this case is sensible. We also talk about *modeling* the problem. The predictions we can make, however, are subject to uncertainties, and we can only state the *probability* that a result occurs. In this case, the probability that we roll a 4 is $\frac{1}{6}$.

Following is a physical example of these considerations, which should also serve as our standard example.

The Alpha Decay

A radioactive nucleus that emits helium-4 atomic nuclei when decaying is called an *alpha emitter*. The helium-4 nucleus emitted from the decaying nucleus is called an *alpha particle*. This alpha particle is measurable. The emission of alpha particles from a radioactive source within a given time unit is *not constant* but fluctuates *randomly*. So, the emitter can emit 10 alpha particles within a 10-second interval, then no particles at all during the next 10 seconds, followed by 100 particles during the next. A typical alpha emitter is americium-241, a decay product of plutonium-241 that is common in radioactive waste.

Number of decays observed	Number of experiments with observation
0-2	18
3	28
4	56
5	105
6	126
7	146
8	164
9	161
10	123
11	101
12	74
13	53
14	23
15	15
16	9
17+	5

Table 2.1: Number of decays within 10 seconds and number of experiments (out of 1207 total) in which the corresponding number of decays was observed.

An experiment measured the number of decays of americium-241 within a 10-second interval. The experiment was repeated 1207 times, and each time the number of decays within 10 seconds was measured. In Table 2.1, the first column lists the number of decays, and the second column shows how often this number of decays was observed out of the 1207 experiments. For example, 0, 1 or 2 alpha particles were measured in 18 of the 1207 experiments. 3 alpha particles were measured in 28 of the 1207 experiments.

No physical law can forecast how many decays occur within a particular time interval. But we can try to give the probability of each number of decays occurring within a 10-second interval. So, we strive to *design* a probability model for the alpha decay.

The question then arises of how to *estimate* the decay probabilities based on the data. For example, what is the probability of measuring 12 decays within 10 seconds? Let us represent the frequencies of the numbers of decays graphically. Let us consider for this the histogram of our data (see the left side of Figure 2.1). We thus obtain a so-called *frequency distribution*. We can for example read from this histogram that in 74 experiments, 12 decays were observed. In this case, we speak of *absolute frequency*. This number 74 referring to the observed number of experiments, however,

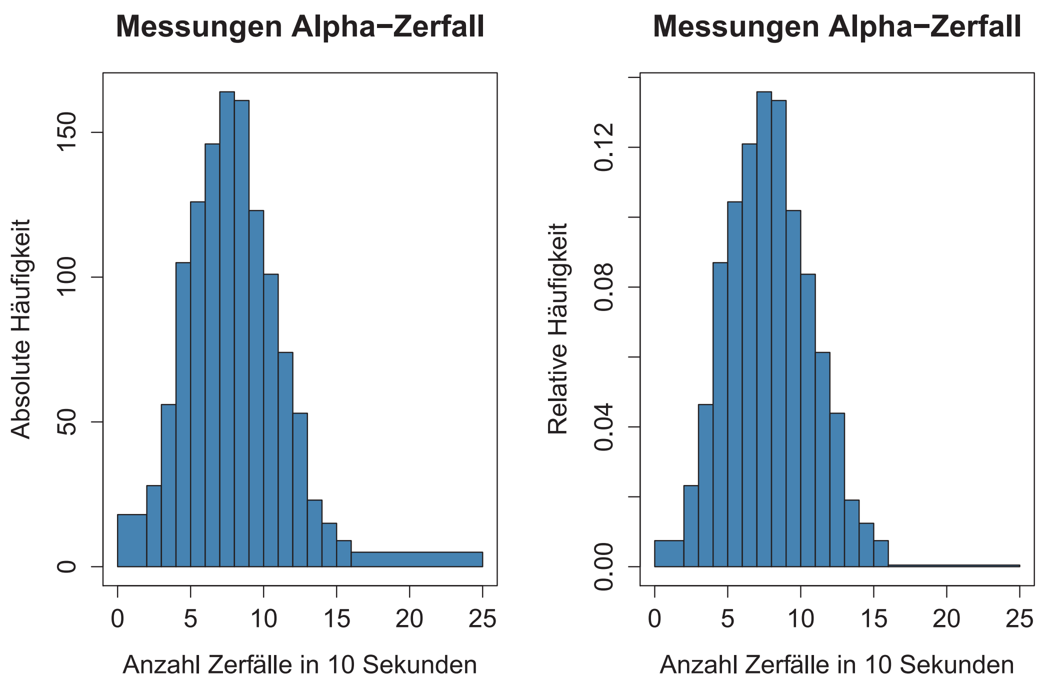

Figure 2.1: Frequency distribution of alpha decay.

is rather meaningless apart from the mention of the total number of 1207 observed experiments. This number can only be compared with difficulty to another series of experiments, for example one involving 2389 experiments.

The *relative frequency* provides more information on a particular number of decays. For this, we divide the (absolute) frequency of the experiments in which a particular number of decays was observed by the total number of experiments carried out. This then gives the *percentage* of experiments in which a particular number of decays was observed.

As we have seen in the chapter on histograms, the area of a bar corresponds to the relative frequency of measurement values within the corresponding interval. We thus read on the right side of Figure 2.1 that 7 decays happen within 10 seconds in about 12 % of the experiments. In this case, the total area of the bars yields 1.

This naturally leads us next to the concept of the probability of decay. So, we interpret the relative frequencies of the observed numbers of decays per 10-second intervals as the (estimated) *probabilities*: the probability that the respective number of decays is observed within 10 seconds. In mathematics, probabilities are generally not given as percentages but with number between 0 and 1. So, the probability that 7 decays happen within 10 seconds is 0.12. The more experiments are carried out, the closer this estimated probability is to the real probability. Furthermore, we determine that all decays put together add up to 1.

We next wonder: How are these probabilities distributed by number? Does a mathematical model fit the observed distribution of the relative frequencies and allow us to then make predictions?

2.2 Probability Models

2.2.1 Basic Concepts

Let us consider **random experiments** where the output is not exactly predictable. Examples can be:

- the number of decays of an alpha emitter
- the result of the toss of a die

A **probability model** describes what results are possible with such an experiment and what probabilities the different results have. The possible results on a die are $1, 2, 3, 4, 5, 6$; the probability of tossing one of these numbers is $1/6$, as long as the die is fair.

A probability model then lets us make certain predictions, which we can verify experimentally. For example, we can develop a good game strategy when gambling. A probability model has the following components:

- **Sample space** Ω, comprising the **elementary events** ω,
- **Events** $A, B, C, \ldots,$
- **Probabilities** P

Elementary events are possible results or outcomes of the experiments, which combine to form the sample space:

$$\Omega = \{\underbrace{\text{possible elementary events } \omega}_{\text{possible outcomes/results}}\}$$

Example 2.2.1

In the case of tossing the die, the random experiment's possible results form the sample space:

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

The element $\omega = 2$ is an elementary event. It means the number 2 appeared when the die was tossed. ◀

Example 2.2.2

In the case of the alpha decay, the sample space is defined by:

$$\Omega = \{0, 1, 2, 3, 4, \dots\}$$

since any number of decays is possible within a 10-second interval. The elementary event $\omega = 6$ means 6 decays were measured within 10 seconds. ◀

During an experiment, an elementary event is randomly "selected" from the set of all elementary events (sample space).

Example 2.2.3 Tossing a coin twice

Let us use H for "head" and T for "tail" (this applies to the rest of the chapter). All possible results when tossing a coin twice are then defined by:

$$\Omega = \{HH, HT, TH, TT\}$$

where Ω is the sample space. An elementary event can be, for example, $\omega = HT$. ◀

An **event** A is understood to be a subset of Ω:

$$A \subseteq \Omega$$

"An event A occurs" means the result ω of the experiment belongs to A.

Example 2.2.4 Tossing a coin twice

Let us toss a coin twice. Let us now consider the event A, where exactly one H is tossed. This event comprises the elementary events HT and TH. Event A is then defined by the set

$$A = \{HT, TH\}$$

If our tosses result in TT, then event A does *not* occur. ◀

Example 2.2.5 Tossing a die

Let the event A be "an odd number is tossed." Then,

$$A = \{1, 3, 5\}$$

Event A occurs when, for example, the number 5 appears on the die.

Let us call B the event that a number less than 7 is tossed. Of course, that is always true. Therefore, in this case,

$$B = \Omega$$

We call this a *certain* event.

Furthermore, let C be the event "the number 7 is tossed." That is impossible, and we write

$$C = \emptyset$$

The symbol $\emptyset$ represents the empty set, which does not contain any element. In such a case, we call the event an *impossible* event. ◀

When dealing with events, it is useful to recall the operations of set theory as well as what they mean. In theory, operations (union, intersection, complement) are used to obtain new events from predefined events (see Table 2.2).

Name	Symbol	Meaning
Union	$A \cup B$	A **or** B
Intersection	$A \cap B$	A **and** B
Complement	$\overline{A}$	**not** A
Difference	$A \backslash B = A \cap \overline{B}$	A **without** B

Table 2.2: Operations in set theory.

Remarks:

i. The "or" for the set operator *union* is not exclusive: So, an element can be in A *and* in B. Colloquially, "or" is usually interpreted as "either... or..." However, that does not apply when using "or" for the set operator union.

ii. The complement $\overline{A}$ can also be written as A^c. ♦

Sets A and B are called **disjoint** if A and B are mutually exclusive and therefore cannot occur simultaneously. In this case,

$$A \cap B = \emptyset$$

Hence this event is impossible.

De Morgan's laws also apply:

1. $\overline{A \cap B} = \overline{A} \cup \overline{B}$
2. $\overline{A \cup B} = \overline{A} \cap \overline{B}$

Venn diagrams illustrate all these terms, operations, and rules quite simply (see Figure 2.2).

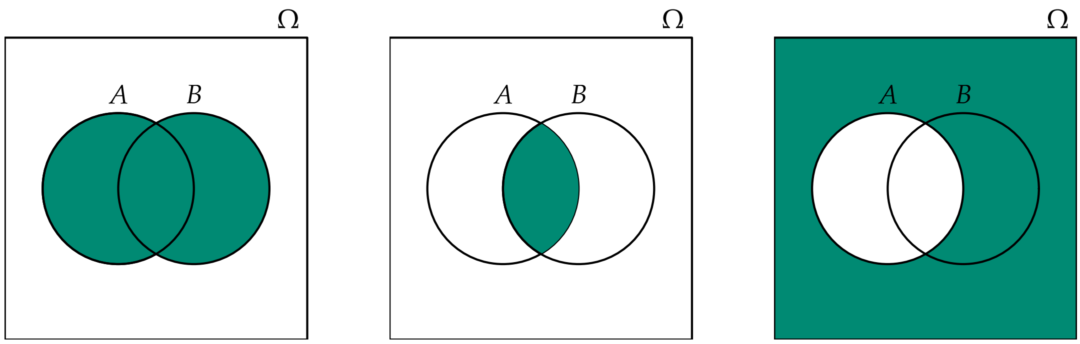

Figure 2.2: *Left:* $A \cup B$, *Center:* $A \cap B$, and *Right:* $\overline{A}$

Example 2.2.6

Let us consider the events

- A: "the Sun will shine tomorrow"
- B: "it will rain tomorrow"

The operations then have the following meanings:

1. $A \cup B$: "the Sun will shine tomorrow *or* it will rain tomorrow", and it can also mean: "the Sun will shine tomorrow *and* it will rain tomorrow".
2. $A \cap B$: "the Sun will shine tomorrow *and* it will rain tomorrow"

3. $\overline{A}$: "the Sun will not shine tomorrow"

◀

2.2.2 Probability

We have seen in the introduction that it is rather difficult to define the concept of probability precisely. It seems that in order to define it, one has to know already what it means. In order to circumvent this difficulty, among other things, mathematicians proceed *axiomatically*. This means that they do not even try to define explicitly *what* a probability is. Instead certain properties (so-called *axioms*), which seem plausible, are assumed, and one works with all statements wich follow from these axioms logically. With this procedure, one has *implicitly* fixed the meaning of the concept of probability, and any interpretation of it is fine, as long as it is consistent with the postulated axioms.

The calculus of probability is based on the following three axioms:

Kolmogorov's Probability Axioms

To each event A a real number $P(A) \in \mathbb{R}$ is assigned as a *probability*. This assignment must satisfy for all events A and B:

A1: $P(A) \geq 0$

A2: $P(\Omega) = 1$

A3: $P(A \cup B) = P(A) + P(B) \quad \text{if } A \cap B = \emptyset$

Remarks:

i. The notation $P(A)$ refers to the probability that event A occurs. Let A be the event that an odd number is tossed on the die. Then, if the die is fair,

$$P(A) = \frac{1}{2}$$

ii. The letter P stands for the word *probability*.

iii. Probabilities are never negative.

iv. With $P(\Omega) = 1$ one determines that the probability of a **special** event is equal to 1. Together with the other axioms this implies that the probabilities of an **arbitrary** event must be between 0 and 1. ⧫

Example 2.2.7

When tossing two coins, it is plausible that all 4 elements of

$$\Omega = \{HH, HT, TH, TT\}$$

have the same probability. Since $P(\Omega) = 1$, the probabilities must add up to 1. Hence:

$$P(HH) = P(HT) = P(TH) = P(TT) = \frac{1}{4}$$

◄

Other laws can be derived from Kolmogorov's three axioms. We list some examples:

Calculation Rules

Let A, B and $A_1, \ldots A_n$ be events, then

$$\begin{aligned} P(\overline{A}) &= 1 - P(A) && \text{for each } A \\ P(A \cup B) &= P(A) + P(B) - P(A \cap B) && \text{for all } A \text{ all } B \\ P(A_1 \cup \ldots \cup A_n) &\leq P(A_1) + \ldots + P(A_n) && \text{for all } A_1, \ldots, A_n \\ P(B) &\leq P(A) && \text{for all } A \text{ and } B \text{ where } B \subseteq A \\ P(A \backslash B) &= P(A) - P(B) && \text{for all } A \text{ and } B \text{ where } B \subseteq A \end{aligned}$$

All true statements of probability theory can be deducted logically-algebraically from the Kolmogorov-Axioms. This approach is followed by mathematicians. For us theseproofs are not so importand, but still the question whether certain statements are true. This can be intuitively made clear using Venn diagrams.

In this approach, an event is represented as a region, and its area is interpreted as probability. By definition, the total area of Ω is equal to 1, since we must have $P(\Omega) = 1$ according to A2. The first rule A1 of the box above is illustrated in Figure 2.2 on the right. Here, $P(A)$ is the area of the region A, and $P(\overline{A})$ is that of the green region. It is then evident that

$$P(A) + P(\overline{A}) = |\Omega| = 1$$

and thus

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

As an exercise, verify the other calculation rules.

This chapter covers *finite* probability models, where the sample space is finite. This is a special case of a discrete model, in which the results can be numbered using natural numbers. For example, the set

$$\Omega = \{0, 1, \ldots, 10\}$$

is *finite* and therefore discrete, whereas

$$\Omega = \mathbb{N}_0 = \{0, 1, 2, \ldots\}$$

is *infinite* but stil *discrete*.

The set $\Omega = \mathbb{R}$ is **not** discrete. We call it *continuous* instead, since the results can be varied continuously or stepless. It will later play a very important role with regard to the normal distribution.

In a finite probability space, the probability of each event determined by the probabilities of all of its elementary elements. We have the

Sum Formula

The probability of the event

$$A = \{\omega_1, \omega_2, \ldots, \omega_n\}$$

is

$$P(A) = P(\omega_1) + P(\omega_2) + \ldots + P(\omega_n) = \sum_{i=1}^{n} P(\omega_i) = \sum_{\omega \in A} P(\omega)$$

This follows by repeated application of axiom A3.

Example 2.2.8

We have a die that is not fair. The probabilities of tossing different numbers are therefore not equal. The corresponding probability for each number is given in Table 2.3.

ω	1	2	3	4	5	6
$P(\omega)$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{12}$	$\frac{1}{12}$

Table 2.3: Probabilities of an unfair die.

Hence

$$\begin{aligned} P(\Omega) &= P(1) + P(2) + P(3) + P(4) + P(5) + P(6) \\ &= \frac{1}{3} + \frac{1}{6} + \frac{1}{12} + \frac{1}{4} + \frac{1}{12} + \frac{1}{12} \\ &= 1 \end{aligned}$$

For the event $A = \{1, 2, 4\}$ we have

$$P(\Omega) = P(1) + P(2) + P(4) = \frac{1}{3} + \frac{1}{6} + \frac{1}{4} = \frac{3}{4}$$

◀

Example 2.2.9

We toss a coin twice. Then the event A: "tossing H exactly once" is

$$A = \{HT, TH\}$$

For the probability $P(A)$ that the event A occurs, we obtain

$$P(A) = P(HT) + P(TH) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

For the event B: "tossing at least one head", we obtain as the probability that this event occurs

$$P(B) = P(HT) + P(TH) + P(HH) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$$

since $B = \{HT, TH, HH\}$. In this case, this probability is easier to calculate with the so-called *complementary probability*. The complement $\overline{B}$ of B is defined as:

$$\overline{B} = \{TT\}$$

And this gives us, with the first calculation rule (see above):

$$P(B) = 1 - P(\overline{B}) = 1 - \frac{1}{4} = \frac{3}{4}$$

◀

Interpretation of Probabilities

In probability theory, often the probabilities of particular events A are defined (based on plausibilities, symmetry considerations, scientific theories, expert knowledge, and data), and then certain probabilities are calculated using the rules above.

Statistics works in reverse: From data, i.e., from the information that ceertain events have occurred, one tries to draw conclusions about an unknown probability model (unknown probabilities).

There are different interpretations of the concept of probability:

- **Frequentist:** "Limit" of the relative frequency of the occurrence of the event in many independent repetitions

  By measuring alpha decay (experiment), we conclude that in 12 % of observations, 7 decays occur within 10 seconds. So, for example, in 12 experiments out of 100 total, 7 decays were observed.

- **Bayesian:** Measure for the belief that an event occurs

  "I am 90 % sure that I will pass the next exam."

Laplace-Model

In many cases, it is plausible to assume every elementary event has the same probability. An example of such a probability model is called the **Laplace model**. The standard example for this is the toss of a fair die. *Each* number then has a probability of $1/6$ of showing up on the die. In this case, the determination of probabilities reduces to counting.

The sample space Ω of the Laplace model has m elementary events. If all these elements have the same probability, then from the sum formula implies

$$P(\Omega) = \sum_{\omega \in \Omega} P(\omega) = mP(\omega) \quad \Rightarrow \quad P(\omega) = \frac{1}{m}$$

If the event E consists of g different elementary events

$$E = \{\omega_1, \omega_2, \ldots, \omega_g\}$$

then we have

$$\begin{aligned} P(E) &= P(\omega_1) + P(\omega_2) + \cdots + P(\omega_g) = \frac{1}{m} + \frac{1}{m} + \cdots + \frac{1}{m} \\ &= g \cdot \frac{1}{m} = \frac{g}{m} \end{aligned}$$

Laplace Probability

For an event E in the Laplace model we have

$$P(E) = \frac{|E|}{|\Omega|}$$

where $|A|$ denotes the number of elements of the set A.

This result is colloquially formulated as follows: we divide the number of "**favorable**" elementary events by the number of "**possible**" elementary events.

Example 2.2.10

Two fair dice are tossed. What is the probability that the sum of the numbers on the dice is 7?

An elementary event describes the numbers shown on the two dice. We can write this event in the form $(1,4)$ when 1 appears on one die and 4 appears on the other. There are 36 possible elementary events in all:

$$\Omega = \{(1,1), (1,2), \ldots, (6,5), (6,6)\}$$

Let us define as E the event that the numbers add up to 7. There are 6 such elementary events:

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

Since all elementary events are equally probable, the probability of event E is:

$$P(E) = \frac{|E|}{|\Omega|} = \frac{6}{36} = \frac{1}{6}$$

◀

2.3 Combinatorics

Laplace probability requires that elements in events can be counted. This question belongs to the field of *combinatorics*. In this section, we deal with permutations, combinations and variations. These counting methods are an important tool for solving numerous problems in probability theory and statistics and can be illustrated very clearly using the urn model.

2.3.1 Urn Model

There are n different balls in an urn, which are numbered consecutively or differ from each other in color, for example. In the next sections, we want to study the question, how many possibilities there are to draw a certain number of balls from this urn.

From the urn, k times a ball is drawn. We must distinguish between the following cases:

a) **Draw with replacement:**
Each ball drawn is returned to the urn, before the next ball is drawn, and can therefore be drawn again in subsequent draws. Each ball can therefore be drawn several times (with repetitions).

b) **Draw without replacement:**
Each ball drawn is not returned to the urn and therefore cannot be drawn again in subsequent draws. Thus, each ball can be drawn only once (without repetitions).

Furthermore, in both cases, we distinguish whether the order of the draws is taken into account or not.

1. **With consideration of the order:**
The order of the draws is taken into account, i.e., draws of the same balls are distinguished if they are drawn in a different order. We then speak of a *variation of k-th order*.

2. **Without consideration of the order:**
The order of the draws is not taken into account, i.e., draws of the same balls are considered equal, no matter in which order they were taken. We then speak of a *combination of k-th order*.

Example 2.3.1

From an urn with $n = 10$ balls, $k = 3$ are drawn. There are four variants:

1a) Draw with replacement and with consideration of the order: We note the numbers of the balls in the order in which they are drawn. This way, a list with 3 numbers grows which is mathematically called a 3-tuple. For instance, there can be a result $(4, 7, 4)$, if ball 4 is drawn first, then ball 7, and finally ball 4 again. This is possible, since the ball of the first draw was returned to the urn. The result $(7, 4, 4)$ will be considered as a different result, since the balls were drawn in a different order.
$$(4, 7, 4) \neq (7, 4, 4)$$

1b) Draw without replacement and with consideration of the order: We represent the results as in case 1a), but now no number can occur more than once, hence the result $(4, 6, 4)$ is impossible.

2a) Draw with replacement and without consideration of the order: We put the balls back after drawing, but ignore the order in which they were drawn. It is therefore useful to note the numbers in some standard order, for instance sorted by size: If 7 is drawn once and 4 twice, then we note the result as $(4, 4, 7)$.

2b) Draw without replacement and without consideration of the order: We do not put the balls back after drawing, so repetitions are impossible. In this case, the results can be considered as sets, since there is no order and no repetitions. For instance, the results $\{4, 7, 5\}$ and $\{7, 4, 5\}$ are identical:
$$\{4, 7, 5\} = \{7, 4, 5\}$$
Again it is useful to note the numbers in sorted order: $\{4, 5, 7\}$

◀

We thus come across the terms *variation* and *combination*. In statistics, such a random sampling of k balls is called a *sample of size* k. It is called *ordered* if the order in which the sample elements (here: balls) are drawn is taken into account. However, if the order does not play a role, the sample is called *disordered*.

2.3.2 Variations

We first consider *variations*, i.e., drawing with considering the order.

Example 2.3.2

From an urn with $n = 10$ balls, $k = 3$ are drawn. The order of the draws is taken into account. In each case, we study how many different results there are.

a) Draw with replacement: In the first move we have $n = 10$ balls that we can draw, and we write down the result first. Since we replace the ball drawn, in the second move we have the same $n = 10$ possibilities. For each of the 10 results in the first move, there are 10 results in the second, so $10 \cdot 10 = 10^2$ possible draws. For each of these results, there are again 10 possibilities in the third move. Hence we have a total of

$$n \cdot n \cdot n = 10 \cdot 10 \cdot 10 = 10^3 = 1000$$

third order variations of 10 balls with repetitions.

b) Draw without replacement: In the first move we have the same $n = 10$ possibilities. Since we do not replace the drawn ball, in the second move we only have $n - 1 = 9$ possible balls. The second drawn ball reduces the number of remaining balls again, so that in the third move there are $n - 2 = 8$ possibilities. Hence, there are

$$n \cdot (n-1) \cdot (n-2) = 10 \cdot 9 \cdot 8 = 720$$

third order variations of 10 balls without repetitions.

◀

The product of the descending natural numbers in the last example can be expressed using the factorial. The *factorial* of a natural number n is defined as the product of all natural numbers from 1 to n, i.e.,

$$n! = 1 \cdot 2 \cdots (n-1) \cdot n \quad \text{for} \quad n \geq 1$$

For pragmatic reasons, one also defines

$$0! = 1$$

This means that by expansion, the result from the last example can be written as

$$10 \cdot 9 \cdot 8 = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdots 2 \cdot 1}{7 \cdot 6 \cdots 2 \cdot 1} = \frac{10!}{7!} = \frac{n!}{(n-k)!}$$

Variations of k-th Order

From an urn with n different balls, k balls are taken one after the other and arranged in the order in which they are drawn. They form a so-called *k-th order variation*.

1. **k-th order variation with repetition:** There are

$$V_w(n;k) = n^k$$

different k-th order variations with repetition, where $k > n$ is allowed.

2. **k-th order variation without repetition:** The number of k-th order variations without repetition is

$$V(n;k) = n \cdot (n-1) \cdot (n-2) \cdots (n-k+1) = \frac{n!}{(n-k)!}$$

where $k \leq n$ must hold.

Example 2.3.3

With a combination lock, you can set four digits between 0 and 9. Normally exactly one setting is correct. However, the lock has a defect, so the second digit does not have to be correct in order to open the lock. What is the probability that you can open the lock by guessing a sequence of digits to open the lock? ◀

Solution. There are 10 digits and four are set. The order is important and repetitions are possible. Therefore, a total of

$$m = V_w(10;4) = 10^4 = 10000$$

settings are possible. Since the second digit is not relevant, there are $g = 10$ favorable cases where the lock opens. Thus, the probability is

$$\frac{g}{m} = \frac{10}{10^4} = 10^{-3} = 0.001$$

that one can open the lock in the first attempt. □

Example 2.3.4

In a horse race, 12 horses start. In a three-way bet, the first three places must be predicted in the correct order. What is the probability of winning a three-way bet? ◀

Solution. From the 12 horses, 3 are drawn to occupy the first three places. The order is therefore important, but repetitions are not possible. Therefore there are

$$m = V(12;3) = \frac{12!}{(12-3)!} = 12 \cdot 11 \cdot 10 = 1320$$

possible results on the first three places. Exactly one is set, so there is $g = 1$ favorable possibility. So the probability of winning is

$$\frac{g}{m} = \frac{1}{1320} \approx 0.000758$$

□

2.3.3 Permutations

An important special case of a variation is the case $k = n$. This means that from an urn, all balls are drawn. If the order is taken into account, then one speaks of a *permuation* or *arrangement*.

Example 2.3.5

In how many ways can the $n = 6$ letters A, B, C, D, E, F be arranged? We number the letters according to the alphabet and we think of them as balls in an urn. We draw $k = 6$ times without putting them back. So we have

$$V(6;6) = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 6! = 720$$

permutationen of the six letters. ◀

Up to now, we have always assumed that the balls in the urn are numbered consecutively, i.e., that all balls can be distinguished from one another. Now we will drop this assumption and consider the case that certain balls look the same and are therefore indistinguishable.

Example 2.3.6

Suppose we have 7 balls in an urn, of which 3 are red, 2 are green and 2 are blue. How many arrangements of the balls are there, if we cannot tell the balls of the same color apart? If, for the matter of argument, we imagine that the balls are numbered consecutively, then with this distinction, there are

$$7!$$

permutations. But many of these are indistinguishable. E.g., if the balls 1 to 3 are red, then

$$(6, 1, 3, 7, 2, 4, 5) \quad \text{and} \quad (6, 2, 1, 7, 3, 4, 5)$$

are indistinguishable. In general, from one permutation we obtain new ones by swapping the red balls again. Thus, by permutation of the red balls, groups of 3! indistinguishable permutations are combined together. So there are

$$\frac{7!}{3!}$$

of these groups. By permuting the green balls, say balls 4 and 5, we combine 2! of these groups each in a new one. Hence there are

$$\frac{7!}{3! \cdot 2!}$$

new groups, where the blue balls are distinguished yet. By permuting these, a number of

$$\frac{7!}{3! \cdot 2! \cdot 2!} = \frac{7 \cdot 6 \cdot 5 \cdot 4}{2 \cdot 2} = 7 \cdot 6 \cdot 5 = 210$$

groups of permutations emerge, which are completely indistinguishable. ◀

Permutations of n Elements

Every possible arrangement of n elements is called a *permutation* of the elements.

1. If all n elements are pairwise distinct, then the number of permutations is
$$P(n) = 1 \cdot 2 \cdot 3 \cdots (n-1) \cdot n = n!$$
2. If among the n elements, groups of $n_1, n_2, \ldots, n_k$ each are equal, then there are
$$P(n; n_1, n_2, \ldots, n_k) = \frac{n!}{n_1! n_2! \cdots n_k!}$$
different permutations. Here, we have $n_1 + n_2 + \cdots + n_k = n$.

The factorial of n grows very quickly with n, namely exponentially. It is therefore best to calculate factorials with **Python**. The library `math` provides the method `math.factorial()` for this purpose.

```
import math

print(math.factorial(6))
print(math.factorial(7)//
      (math.factorial(3)*math.factorial(2)*math.factorial(2)))

720
210
```

Note that we have used the integer division `//` here, so that the result is of the data type `int`.

For the number of variations without repetition, there is the method `math.perm()`.

```
import math

print(math.perm(10,3))
print(math.factorial(10)//math.factorial(7))

720
720
```

2.3.4 Combinations

We now turn to drawing without taking the order into account, the so-called *combinations*.

Example 2.3.7

From an urn with $n = 10$ balls, $k = 3$ are drawn without putting them back. The order of the draws is not taken into account. If, for the sake of argument, we do take the order into account, then we have, according to Example 2.3.2 b)

$$V(10;3) = \frac{10!}{7!}$$

draws. If we permute the 3 balls in such a draw, we get one that differs only in the order. We now want to consider those as identical. Therefore, the draws are summarized in groups of 3!, so their total number is divided by 3!. Thus, we obtain

$$\frac{10!}{7! \cdot 3!} = \frac{n!}{(n-k)! \cdot k!} = 120$$

suhc groups. These correspond to combinations without repetition.

Since, in this case, the result can be interpreted as a set, this means that from a set with $n = 10$ elements

$$\frac{10!}{7! \cdot 3!} = \frac{n!}{(n-k)! \cdot k!} = 120$$

subsets with 3 elements can be constructed. ◀

The formula from the last example is so important that it deserves a name. We call

$$\binom{10}{3} = \frac{10!}{7! \cdot 3!}$$

a *binomial coefficient* and pronounces it "10 choose 3". Note that this is not a vector.

Example 2.3.8

Another argumentation for the result from Example 2.3.9 uses the result about permutations of elements, some of which are indistinguishable. It can be argued that by drawing $k = 3$ from $n = 10$ balls, these are divided into two groups: the drawn and the undrawn. those drawn and those not drawn. Instead of saying, for example, that balls 3, 4 and 7 were drawn, you can also place three red balls at squares 3, 4 and 7, and one green ball each on all other 7 squares. Hence, the number of possibilities corresponds to the number of permutations of $n = 10$ balls, of which $n_1 = 3$ are red and $n_2 = 7$ are green, these are

$$P(10; 3, 7) = \frac{n!}{n_1! \cdot n_2!} = \frac{10!}{3! \cdot 7!} = \frac{n!}{k! \cdot (n-k)!} = \binom{n}{k}$$

◀

The most difficult case is the case of combinations with repetition. We could try, in a similar way as above, to first count under consideration of the order and then group together the indistinguishable draws. However, as there may be repetitions, these groups are not all of the same size. For example, in the draw $(6, 4, 8)$, by permutation of the 3 different elements, $3! = 6$ draws are combined, but in the draw $(6, 4, 4)$ there are only 3 different ones.

Example 2.3.9

From an urn with $n = 10$ balls, $k = 3$ are drawn with putting them back. The order of the draws is not taken into account. Therefore, we can sort the drawn numbers and obtain a sorted sequence of three numbers between 1 and 10, possibly with repetitions, for example $(3, 3, 9)$. If we now add 1 to the second number and 2 to the third number, we obtain three different numbers between 1 and 12, in our example

$(3, 4, 9)$. So the total number of draws is equal to the number of sorted sequences of 3 numbers between 1 and 12. This is the number of third order combinations of 12 elements **without** repetition, and of these there are

$$\binom{12}{3} = \binom{n+k-1}{k} = \frac{12!}{9!3!} = 220$$

◀

Combinations of k-th Order

From an urn with n different balls, k balls are drawn one after the other, where we do not regard the order in which they were drawn. The k balls drawn (arranged in any order) then form a *combination of k-th order*.

1. **Combinations without repetition**
   There are
   $$C(n; k) = \binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$$
   combinations k-th order without repetition, where $k \leq n$ must hold.
2. **Combinations with repetition**
   There are
   $$C_w(n; k) = \binom{n+k-1}{k}$$
   different k-th order combinations with repetition, where $k > n$ is possible.

For binomial coefficients we have the method `math.comb()`.

```
import math

print(math.comb(10,3))
print(math.factorial(10)//
      (math.factorial(7)*math.factorial(3)))
print(math.comb(12,3))

120
120
220
```

Example 2.3.10

We consider the lottery, where 6 out of 42 are drawn without putting them back. The order of the numbers is irrelevant.

a) What is the probability of a six, i.e., that all the numbers picked are correct?

b) What is the probability of three correct numbers?

◀

Solution. a) We consider combinations with repetition, so there are a total of

$$C(42;6) = \binom{42}{6}$$

results. Precisely one of these was was picked, so the probability for a six is

$$\frac{1}{\binom{42}{6}}$$

We calculate it with **Python** :

```
import math

print(1/math.comb(42,6))

1.9062920218247562e-07
```

Thus, the probability for a six is roughly $2 \cdot 10^{-7}$, hence 0.2 Millionsth.

b) As above, we have

$$m = \binom{42}{6}$$

possible draws. If (exactly) three of the picked numbers are to be correct, then you can select 3 different numbers from the 6 picked ones, which should be correct. The order of this choice does not matter, and repetitions are not possible. Therefore, there are for $\binom{6}{3}$ possibilities. The remaining 3 numbers must be wrong, otherwise you would have more than 3 correct numbers. For each choice of the 3 correct numbers therefore 3 can be drawn again from the remaining $42 - 6 = 36$ incorrect numbers. For this there are $\binom{36}{3}$ possibilities. So there are

$$g = \binom{6}{3} \cdot \binom{36}{3}$$

favorable possibilities. Thus, the probability for three correct numbers is

$$\frac{g}{m} = \frac{\binom{6}{3} \cdot \binom{36}{3}}{\binom{42}{6}}$$

Again we calculate it with **Python** :

```
import math

print(math.comb(6,3)*math.comb(36,3))
print(math.comb(6,3)*math.comb(36,3)/math.comb(42,6))

142800
0.027221850071657516
```

Hence, the probability for three correct numbers is more than 140′000 times the one for a six.

□

Educational objectives

- □ You know the three axioms of probability theory and their most important consequences.
- □ You can determine the correctness of statements about probabilities with the help of venn diagrams.
- □ You can deal with events, their set operations, and their probabilities.
- □ You know the urn model and are able to distinguish between drawing with/without replacement and with/without considering the order of the draws.
- □ You can calculate the number of variations, permutations, and combinations and with that probabilities.

Computer-based educational objectives

You are able ...

- □ to simulate the repeated execution of a random experiment using **Python** and represent relative frequenceies using bar plots.
- □ to calculate combinatorial quantities and related probabilities with **Python** .