Measures and Probabilities

3

Continuous-time theory makes use of a sophisticated functional analytical apparatus. If you really want to understand what a Brownian motion is and how to use it, you have no choice but to first deal with measurement theory and general integration theory.

3.1   Basic Problem of Measurement Theory

In everyday life it is often said that something is measured. Therefore, every reader probably has a certain idea of what a measure is. If you are not a mathematician, you might even ask yourself why you need a theory for such a "simple object" as a measure at all. Characteristically, a measure is a number that describes a property of an object, such as its volume, weight, or length. Probabilities are also numbers which measure something: probabilities provide information about the intensity with which someone expects a possible future development. They play a decisive role in the theory of stochastic processes. And hardly anyone will deny that probabilities are not quite as easy to comprehend as the distance between two points on a plane.

We hope that our readers can follow us better when we state that it is necessary to engage in measurement theory. This theory attempts to discuss in a general way the properties of numbers which are intended to capture characteristics of the diverse objects of interest.

Properties of Measures   An elementary introduction to measurement theory could simply be imagined in such a way that each subset of the event space is assigned a number, namely its measure. A measure μ would then be a mapping of each subset

of Ω into the real numbers or formally[1]

$$\mu : \mathcal{P}(\Omega) \rightarrow \mathbb{R}. \tag{3.1}$$

If we think of the dice again, a number has to be assigned to each of the 64 subsets. If we think of a probability measure, we would assign the relative frequency $\frac{1}{6}$ to each elementary event of an ideal dice. A subset with n elements[2] has probability $\frac{n}{6}$. Unfortunately, the conditions are much more complicated when dealing with event spaces that contain an infinite number of elements. Under these circumstances, the number of conceivable share prices within an arbitrarily large closed interval is infinite. This forces us to pursue a different approach.

It is obvious to demand that a measure has reasonable properties. You have to be careful. It can easily happen that with the formulation of desirable properties one gets entangled in logical contradictions without even realizing. In the following we will show that this is indeed the case. We will subsequently reflect on the conclusions to be drawn.

To understand how readily one can get caught in contradictions, let us look at a specific example: we concentrate on the event space $\Omega = \mathbb{R}$ which includes the real numbers, and try to construct a probability measure μ on Ω. We will present a number of properties that should be thought of being useful or at least unproblematic.

Existence: The first property that we want to propose seems perfectly natural. We require that a measure $\mu(A)$ can be assigned to each set $A \subset \Omega$. Some readers may wonder why such a trivial feature has to be mentioned at all. At the end of this section we will see that exactly this property will turn out to be problematic.

Nonnegativity: In the introductory remarks we had suggested that a measure could be understood as something like a volume, a length, or a probability. Against this background it seems obvious to postulate that a measure is non-negative,[3]

$$\forall A \subset \Omega \qquad \mu(A) \geq 0. \tag{3.2}$$

This is immediately plausible for probabilities. If one limits oneself to classical physics, masses and lengths will also be nonnegative. The area of the plane also has no negative contents.[4]

---

[1]We have described the set of all subsets of Ω as power set $\mathcal{P}(\Omega)$ with the details being discussed on the pages 20 ff.

[2]This is an event with n different results from rolling a dice only once.

[3]The symbol $\forall A$ means "for all A applies. . ."

[4]However, it is conceivable that in more advanced considerations these parameters could also become negative. In this case, the measurement theory must be expanded. One speaks then of the so-called signed measures, a topic we will not pursue further.

Additivity: Furthermore, we require that in the case of two disjoint subsets which are combined, the corresponding measures must be added,

$$\forall A, B \subset \Omega \qquad A \cap B = \emptyset \Rightarrow \mu(A) + \mu(B) = \mu(A \cup B). \qquad (3.3)$$

The measure must be additive. This requirement will come as no surprise to anyone who thinks in terms of area, space, or volume. It should also apply when you are dealing with probabilities. In this case the prerequisite of Eq. (3.3) means that the events A and B are mutually exclusive.

Before we turn to further properties of measures, we will deal with a statement about measures that can be derived directly from (3.3).

From this condition it applies, for example, that a subset cannot have a larger measure than its supersets. If $A \subset B$ applies, it follows that

$$\forall A \subset B \subset \Omega \qquad B = B \backslash A \cup A \Rightarrow \mu(B) = \mu(B \backslash A) + \mu(A) \geq \mu(A). \qquad (3.4)$$

A First Exercise (Additivity) In order to gain experience with measures we want to prove two characteristics. We will not need the following theorem for our further considerations. However, the proof of the theorem is suitable for a better understanding of the interplay of the various properties of measures.[5] We propose the following:

Proposition 3.1 *If A and B are arbitrary two subsets of Ω, the following two properties are equivalent:*

1. *The measure is additive, see Eq. (3.3).*
2. *For the measure applies*

$$\mu(A) + \mu(B) = \mu(A \cap B) + \mu(A \cup B) \qquad (3.5)$$

*(for arbitrary sets!) and $\mu(\emptyset) = 0$.*

The merit of Eq. (3.5) can be realized by considering Fig. 3.1. This figure shows three separate areas. You see the set $A \backslash (A \cap B)$ on the left, $(A \cap B)$ in the middle, and $B \backslash (A \cap B)$ on the right. Note that the intersection $(A \cap B)$ belongs to both A and B.

Let us look at Eq. (3.5). With the sum $\mu(A) + \mu(B)$ we capture the measure of A, i.e., the left as well as the middle set,

$$A = A \backslash (A \cap B) \cup (A \cap B) \qquad (3.6)$$

---

[5]If you want to skip this exercise, continue reading on
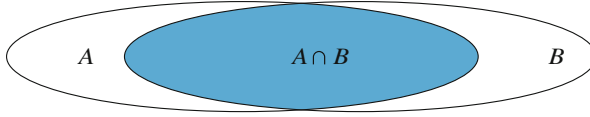
Fig. 3.1 Intuition of property (3.5) of a measure

and the measure of B, i.e., the middle and the right set,

$$B = B \backslash (A \cap B) \cup (A \cap B). \tag{3.7}$$

Obviously, the middle set $(A \cap B)$ here is "counted" twice.

Let us concentrate on the right side in Eq. (3.5). Counting is different here. In the sum $\mu(A \cap B) + \mu(A \cup B)$ we capture the measure of $A \cup B$ and thus the measure of the left, middle, and right set. Subsequently, the measure of $A \cap B$, i.e., the measure of the middle set, is added. But this is exactly the same area we calculated before. We come to the formal proof.

Proof Part $2 \Rightarrow 1$ is trivial, see Eq. (3.3). The opposite is a little more complicated. Since (3.3) must apply to any set A, B we use $A = B = \emptyset$, get $\mu(\emptyset) = 0$ and thus a part of the result. We prove the second part by referring to the exercise of the chapter on set theory.[6] Accordingly it follows from (2.10) that for any sets A and B (even if they are not disjoint)

$$A \cup B = (A \backslash B) \ \cup \ B \tag{3.8}$$

must be fulfilled. If we apply Eq. (3.3) we get

$$\mu(A \cup B) = \mu(A \backslash B) + \mu(B). \tag{3.9}$$

We also realize that for any set A and B

$$A = (A \backslash B) \ \cup \ (A \cap B) \tag{3.10}$$

and again the two sets on the right side of this equation are disjoint. Hence

$$\mu(A) = \mu(A \backslash B) + \mu(A \cap B) \tag{3.11}$$

also applies. From Eqs. (3.9) and (3.11) follows the claim, if $\mu(A \backslash B)$ is eliminated. ■

---

[6]See page 18.

σ-Additivity So far, we have restricted ourselves to the union of two, three, and in a few cases to four sets and formed their intersections and determined the associated measures. However, the number of sets involved has always been finite. It should have become clear how to proceed if the number of sets continues to increase, but still remains finite. Sometimes, however, it is necessary to deal with the union of an infinite series of sets and to determine their measure. It is by no means obvious how to proceed under these circumstances. A relevant property of measures in this context is called σ-additivity. That is what we are going to discuss now.

Consider an infinite sequence of sets $A_1, A_2, \ldots$ This is supposed to be a sequence of subsets, i.e.,

$$A_1 \subset A_2 \subset A_3 \subset \ldots. \tag{3.12}$$

Obviously, the sets grow with an increasing index. We form the infinite union or the set containing all elements of the A_n and call it $\bigcup_{n=1}^{\infty} A_n$. Figure 2.4 on page 19 illustrates this situation.

Each of these sets A_n has the measure $\mu(A_n)$. What can one meaningfully say about the measure of $\bigcup_{n=1}^{\infty} A_n$? To answer this question, we consider any finite number $n < m$ and break the union at m, $\bigcup_{n=1}^{m} A_n$. This set differs from $\bigcup_{n=1}^{\infty} A_n$ by those elements which are only contained in the "later" sets $A_{m+1}, A_{m+2}, \ldots$. With increasing m this "residual set" gets smaller and smaller. All we are asking is that the measure of this residual set disappears entirely when $m \to \infty$.

Thus, we require that the measures $\mu(A_n)$ converge to the measure of the set of infinite union $\mu\left(\bigcup_{n=1}^{\infty} A_n\right)$,

$$A_1 \subset A_2 \subset A_3 \subset \ldots \Rightarrow \lim_{n \to \infty} \mu(A_n) = \mu\left(\bigcup_{n=1}^{\infty} A_n\right). \tag{3.13}$$

And that is exactly what the σ-additivity is supposed to mean.

Return to our interval example from page 20. We know that sets $\left[\frac{1}{n}, 1 - \frac{1}{n}\right]$ "cling" close as possible to the open interval $(0, 1)$ when $n \to \infty$. Between these closed intervals and the limit $(0, 1)$ there is "nothing." There is no number in $(0, 1)$ that cannot be found in any one of the A_n. Now look at the measures $\mu(A_n)$. If the limit of these sets would not go to $\mu\big((0, 1)\big)$, then quite obviously a part of the measure either "disappeared" or "arose from nowhere." Property (3.13) prevents exactly that. Our measure is σ-additive.
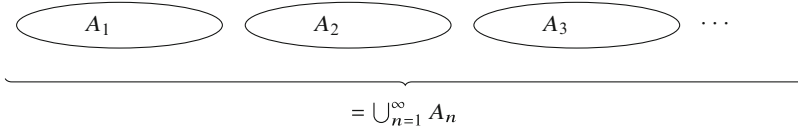
Fig. 3.2 Pairwise disjoint sets as in Proposition 3.2

You can easily come up with a "measure" which violates the condition (3.13). To this end, we define the following measure μ on the set of real numbers,[7]

$$\mu(A) = \begin{cases} 1 & A = \mathbb{R}, \\ 0 & \text{else.} \end{cases} \tag{3.14}$$

With this measure, the full probability is assigned only to the set of all real numbers with other sets being impossible. Now look at the sets $A_n = (-\infty, n]$, which contain all real numbers up to n. These sets form an ascending sequence. The following applies

$$\lim_{n \to \infty} \mu(A_n) = 0 \neq 1 = \mu\left(\bigcup_{n=1}^{\infty} A_n\right). \tag{3.15}$$

σ-additivity does not hold.

Another Exercise (σ-Additivity) Let us concentrate on σ-additivity a bit further.[8] We just looked at a series of sets, each being a subset of its predecessor. Now we turn our attention to the case of an infinite number of sets that are pairwise disjoint.[9] Then the following applies:

Proposition 3.2 *Let A_n be a sequence of pairwise disjoint sets. Furthermore, the measure is additive and σ-additive. Then the following applies:*

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n). \tag{3.16}$$

The prerequisite of Proposition 3.2 states that the sets of a sequence never overlap. To obtain a descriptive idea of what is asserted here look at Fig. 3.2. The

---

[7]This is even a probability measure.

[8]Anyone wanting to skip the exercise may continue reading from page 35 following the material after the keyword "probability measure of the event space."

[9]"Pairwise disjoint" means that every two sets (every pair, so to say) are disjoint, i.e., do not have a common element.

Proposition 3.2 states that the measure of the total set $\bigcup_{n=1}^{\infty} A_n$ is as large as the (infinite) sum of the individual measures $\mu(A_n)$.

Proof The proof's challenge is that the σ-additivity deals with ascending sets, while the sets under consideration are pairwise disjoint. We show how to cope with the pairwise disjoint sets in such a way that you end up with increasing sets. You can easily find such an ascending sequence by combining the first m sets A_m into a new set.

We start with a finite number of sets and define

$$B_n := \bigcup_{m=1}^{n} A_m. \tag{3.17}$$

Since $B_1 \subset B_2 \subset \ldots$, the sets B_n represent an ascending sequence. Thus, according to (3.13)

$$\mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \lim_{n \to \infty} \mu(B_n). \tag{3.18}$$

Remember that the union of all B_n is the same as the union of all A_n, and therefore we have[10]

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \to \infty} \mu(B_n). \tag{3.19}$$

Looking at (3.3) on page 31, the right side of the last equation can be written as

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \to \infty} \sum_{m=1}^{n} \mu(A_m). \tag{3.20}$$

That was to be shown. ∎

Probability measure of the event space: In the context of probabilities it is reasonable to assume that the decision-maker has a complete picture of all conceivable events. Therefore, probability of any event occurring is obviously one. In formal notation

$$\mu(\Omega) = 1. \tag{3.21}$$

---

[10]One reason for this is that $A \cup A = A$ always applies.

Shift invariance: One or two more properties will be added to those noted before. We request that the measurement of a set remains unchanged if it is shifted by one unit.

It is rather difficult to get a clear idea of this property when you think of a probability measure. With area measures, however, the demand for shift invariance is immediately obvious. A circle with a certain diameter finally has the same are everywhere on the plane; and a cylinder with a certain diameter and height has the same volume everywhere no matter where it is located it in space. By analogy, we require that the measure of an interval $[0, 1)$ equals the measure of the shifted interval $[x, x + 1)$ no matter how large x is. We note

$$\forall A \subset \Omega, x \in \mathbb{R} \qquad \mu(A) = \mu(A + x). \qquad (3.22)$$

The reader will probably understand that area measures should be shift-invariant. But why this should also apply to probability measures is not obvious. We will address this point later.

Contradiction Following from Our Properties After having presented the six properties of probability measures we get to the core of the matter. We intend to show the reader that a measure with the six characteristics described leads to a serious problem.

To this end consider the half-open interval $A = [0, 1)$, which must have a measure using the first property. This measure may be denoted by $x := \mu([0, 1))$. Now we use the properties (3.2), (3.3), (3.13), and (3.22) to determine the measure of the entire real axis. We break down the real axis $\mathbb{R} = \Omega$ into infinite many half-open intervals

$$\Omega = \bigcup_{n=-\infty}^{\infty} [n, n + 1). \qquad (3.23)$$

Note that these intervals are pairwise disjoint. Then it follows that

$$
\begin{aligned}
\mu(\Omega) = \mu(\mathbb{R}) = \mu\left(\bigcup_{n \in \mathbb{Z}}[n, n+1)\right) \qquad & \text{due to (3.21) and definition} \\
= \sum_{n \in \mathbb{Z}} \mu([n, n+1)) \qquad & \text{see (3.16)} \\
= \sum_{n \in \mathbb{Z}} \mu([0, 1)) \qquad & \text{due to shift invariance (3.22)} \\
= \sum_{n \in \mathbb{Z}} x \qquad & \text{due to definition of measure} \\
= \begin{cases} 0, & \text{if} \quad x = 0, \\ \infty, & \text{else.} \end{cases} \qquad & (3.24)
\end{aligned}
$$

The following observation is decisive: regardless of the specific value x, the probability of the entire event space cannot be one: either the probability is infinite or zero. Hence, (3.24) shows the contradiction with property (3.21).

Conclusion (Measurable Sets) What conclusion must be drawn from this statement? Obviously, at least one of the properties mentioned above must be eliminated. Which of the six properties is a suitable candidate?

Let us start with shift invariance, because we have noted that there exist no obvious intuition for this property. Although removing shift invariance seems to be a good idea, it is not sufficient. It can be shown that a contradiction can be constructed even if one limits oneself to the properties of nonnegativity, additivity, and σ-additivity. The proof of the contradiction is then, however, no longer as simple as above and requires a set of advanced mathematical instruments.[11]

Thus, we have no choice other than to realize that the idea of assigning a measure to *any subset* cannot be maintained. The very first property of a measure that we developed on page 30 must be dropped. While in the finite dimensional case every elementary event will indeed have a probability, in the infinite dimensional case we must proceed with more caution. Our measurement function μ may not assign a number to any subset. Instead we must start by determining those subsets that *should be measurable* at all.

To this end the notion of a σ-algebra is introduced. There are two ways to approach this concept. One alternative is to restrict ourselves only to the properties which have to be met by measurable sets. These properties are quickly explained, so that we can understand the formal definition of a σ-algebra directly.[12] Another alternative is to provide a content-related interpretation of measurable sets which is often used when economists work with a σ-algebra.[13]

3.2 σ-Algebras and Their Formal Definition

Mathematical Basics Remember that it is not permissible to treat any subset as being measurable. Therefore, it is necessary to determine what can be measured and what cannot be measured. In most cases this choice is arbitrary.

If we want to use ideas of a measure developed on pages 30, we have to place certain minimum requirements on measurable sets. Otherwise the concept of a

---

[11]This is the proposition of Banach and Tarski from 1924. It should be noted that both scholars could even dispense with the σ-additivity of the measure for their evidence, referring only to the properties of nonnegativity and additivity. However, the proof of their theorem is only possible in at least three-dimensional space and using an axiom that is otherwise not necessary in measurement theory (axiom of choice). Under attenuated conditions, similar paradoxes can also be constructed in the plane and on the straight line.

[12]We will do that in the next section.

[13]See page 40 ff.

measurable set will lose its meaning. These minimum requirements result from mathematical considerations.

Formally, a σ-algebra contains all measurable sets. At a minimum, any σ-algebra must have the following properties:

1. It is only natural that each measure assigns the number zero to the empty set. But this presupposes that the empty set is measurable. Therefore, any σ-algebra must contain the empty set.
2. Correspondingly, any measure will naturally assign the number 1 to the entire state space. Again, this presupposes that the set Ω is measurable and must be contained in every σ-algebra.
3. We had made it clear that no measure is lost when uniting disjoint sets A, B

$$\mu(A) + \mu(B) = \mu(A \cup B), \tag{3.25}$$

see page 31. If the disjoint sets A and B are measurable, then consequently their intersection and union must also belong to the σ-algebra.
4. Consider a set $A \subset \Omega$. This set A and its complement $\Omega \backslash A$ are disjoint. The measure of the state space is

$$\mu(\Omega) = \mu(A \cup \Omega \backslash A) = \mu(A) + \mu(\Omega \backslash A). \tag{3.26}$$

Equation (3.26) implies that the complement should be included in the σ-algebra.
5. We had several examples above in which infinite unions and intersections were involved. We claim that for sets A_n also the infinite union $\bigcup_{n=1}^{\infty} A_n$ and the infinite intersection $\bigcap_{n=1}^{\infty} A_n$ are measurable.

The five properties listed are based on simple mathematical considerations. Before we interpret these properties economically we want to state the formal definition of a σ-algebra using the following two-step procedure.

A Two-Step Procedure

- The first step is to specify some sets that should be measurable.
- The second step describes the operations that can be performed with measurable sets without destroying the property of measurability. These operations include complement, union, and intersection.

Admittedly, this procedure is a bit cumbersome, because we have to check whether or not we are still dealing with a measurable set. However, it has the great advantage that one will not get entangled in logical contradictions. There is no other alternative.

Definition 3.1 (σ-Algebra) By a σ-algebra $\mathcal{F}$ we define a set of sets with the following properties[14]:

1. The empty set is a part of the algebra, $\emptyset \in \mathcal{F}$.
2. With any set B its complement $B^c = \Omega \backslash B$ is included, $\Omega \backslash B \in \mathcal{F}$.
3. Along with $B_1, B_2, \ldots$ the union $\bigcup_n B_n$ is included.

It is also said: *sets are $\mathcal{F}$-measurable if they are part of a σ-algebra.* In this context, we will also refer to the properties mentioned here as construction rules or simply rules.

The following note may be helpful. Our definition applies to *any* starting sets (subsets) of Ω. Those sets must be determined. Otherwise the properties 2 and 3 would be meaningless. The definition will usually not result in a *unique* σ-algebra. Often, different σ-algebras will exist for a given set Ω.

The reader may wonder why our definition contains statements about the union of sets, but not about their intersection. Are intersections not supposed to be included in the σ-algebra? The answer may come as a surprise. Intersections of sets are actually elements of the σ-algebra. However, we do not need to include this statement explicitly in Definition 3.1 because it follows from our definition—this result will be derived in the next paragraph. Definitions should always be as parsimonious as possible.

Measurability of Intersections To verify the statement that intersections of sets must be $\mathcal{F}$-measurable when following Definition 3.1, we focus on the third construction rule. This rule states that the union of any number of subsets $\bigcup_n B_n$ belongs to the σ-algebra. Based on the second rule the complement $\bigcup_n(\Omega \backslash B_n)$ must be $\mathcal{F}$-measurable. However, the following always applies to any set:

$$\bigcup_n (\Omega \backslash B_n) = \Omega \backslash \bigcap_n B_n \, , \tag{3.27}$$

which is illustrated by Fig. 3.3. Hence by using rule 2, $\Omega \backslash \cap_n B_n$ must also belong to the σ-algebra. It follows that not only the union but also the intersection $\cap_n B_n$ of subsets are measurable.

Measurability of the Event Space You can observe that the event space Ω is $\mathcal{F}$-measurable. The second construction rule states that $B \cup B^c = \Omega$, and according to the third rule, subset unions are measurable.

There is a vivid interpretation of what measurability means. We will discuss this in the next section.

---

[14] σ-algebras are often referred to as $\mathcal{F}$. The symbol stands for the word "filtration." We will consider filtrations in more detail in Sect. 5.5.
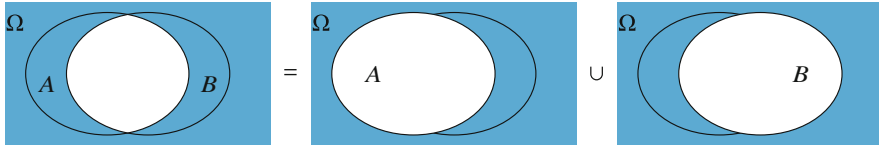
Fig. 3.3 To illustrate the identity of $\Omega\backslash(A\cap B)$ (left) and the union of $\Omega\backslash A$ and $\Omega\backslash B$ (both sets are colored blue in the images)

3.3   Examples of Measurable Sets and Their Interpretation

We will use three examples to illustrate our considerations.

Example 3.1 (Coin Toss) A σ-algebra for flipping a coin has a simple shape. First of all, we know that the σ-algebra must contain both the empty set and the total set. Thus the two sets $\emptyset$ and $\Omega = \{u, d\}$ always belong to any σ-algebra,

$$\emptyset \in \mathcal{F}, \quad \Omega \in \mathcal{F}.$$

In the case of tossing a coin the σ-algebra is either $\mathcal{F} = \{\emptyset, \Omega\}$ (and thus represents the smallest conceivable algebra) or it consists of all subsets of the event space $\mathcal{F} = \mathcal{P}(\Omega)$.[15] In the first case one speaks of a "trivial" σ-algebra. If you realize that the coin toss is the simplest uncertain situation you can imagine,[16] you might not be surprised by this result.

   The example allows a very straightforward and easy-to-understand interpretation. For this purpose we want to equate measurable events with events whose occurrence a decision-maker can "observe." The trivial σ-algebra would then be synonymous with the (almost worthless) information "a coin was tossed" without being told the result of the toss.
   In the second example, however, individual events $\{u\}$ and $\{d\}$ were also measurable. This can be understood to mean that it should be verifiable whether the coin toss resulted in heads or tails.

Example 3.2 (Dice Roll) Basically there are six possible elementary events, i.e., the sets $\{1\}$ to $\{6\}$. But let us consider the case that a person watching the dice roll is only told whether an even or an odd score was obtained. Nothing else shall be revealed. Since it is possible to check whether the dice was rolled at all, the total set $\Omega = \{1, 2, 3, 4, 5, 6\}$ and the empty set $\emptyset$ are undoubtedly among the observable events. If, moreover, it is stated whether the number of points obtained was even or odd, the sets $\{1, 3, 5\}$ and $\{2, 4, 6\}$ are also observable. This makes it possible to

---

[15]The $\mathcal{P}$ symbol denotes the power set, i.e., the set of all subsets. See page 20.

[16]After all, uncertainty can only be spoken of if there are at least two different events.

define the σ-algebra in the form

$$\mathcal{F}_1 = \Big\{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \{1, 2, 3, 4, 5, 6\}\Big\}.$$

It can easily be seen that this set indeed meets all the requirements for a σ-algebra.

Now we extend the example and assume that the exact score will be announced. Then for the σ-algebra the following applies:

$$\mathcal{F}_2 = \mathcal{P}(\{1, \ldots, 6\}),$$

where the σ-algebra is denoted by $\mathcal{F}_2$. Apparently, the σ-algebra consists of all subsets of the set $\{1, \ldots, 6\}$.

Example 3.3 (Double Dice Roll) Consider the case where a dice is rolled twice in a row and the order of the scores is important. Then an elementary event can be described by a pair such as $(1, 6)$. It should be possible to measure the event in which it is only known that the score of the second roll is exactly one point higher than the score of the first roll. Which exact scores (on the first and second roll) were achieved, however, remains hidden. Obviously, the set

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

can then be measured. The complement of this set (which contains $36 - 5 = 31$ elements) is also measurable. The same applies to the empty set and Ω. Other sets are not measurable.

Let us summarize our considerations. Measurable sets are mathematically characterized by the fact that certain operations (union, complement building) are permissible. The admissibility of these operations leads to a set of measurable sets which we call σ-algebra. Every element of this algebra is called an event. Events contain elementary events which cannot be broken down further. An event A (a measurable set or an element from the σ-algebra) can be described as follows:

> Interpretation: an event A can be measured if it is possible to observe whether or not A has occurred.

We can show that the above interpretation does not only contradict the mathematical definition but rather supports it:

1. Common sense, on which one certainly cannot always rely, tells us that for any event the negation of this event ("the opposite") should also be known. If someone can prove in court that event A has happened, he can also disprove that

event A did not happen. Exactly this shows up in the mathematics of a σ-algebra: if any set $A \in \mathcal{F}$ is selected, the complement $\Omega \backslash A$ is included in $\mathcal{F}$. The second rule of construction in the definition of the σ-algebra thus confirms common sense.

2. If events are logically linked we expect that observability is maintained. If you can prove whether or not the events A and B have taken place you will be able to tell whether or not the compound events "A and B" or "A or B" have occurred. This is ensured by the third construction rule in the definition of the σ-algebra. In our examples, the corresponding operations are transparent because the two logical links always yield only trivial results such as the sets themselves, the empty set or Ω. We note, however, that the union and intersection of two sets are always part of the algebra.

In economic contexts instead of a σ-algebra one prefers to talk about an information system. However, not all algebras can be interpreted as (meaningful or plausible) information systems; but conversely, every information system must be represented by a σ-algebra.

In summary, we can state the following: if we want to denote by σ-algebra the set of events known to and verifiable by a person, then each such algebra must meet several conditions,

There is an event: The total set Ω is part of the σ-algebra.
Negations are known: With every known event $A \in \mathcal{F}$ the complement $\Omega \backslash A$ is also located in the σ-algebra.[17]
Or/and links are known: With the events A and B being part of the σ-algebra, then the union $A \cup B$ and the intersection $A \cap B$ are also elements of the algebra.

If one imbeds also infinite unions into the set of conditions, the formal definition of a σ-algebra results.[18]

Some readers may think that there is no need to say more. That would be a mistake. In real life there exist situations where it is not sufficient that a person is informed about the existence of an event. In the case of a lawsuit, i.e., this person must also be able to convince other parties of the occurrence of the event. It must be possible to provide irrefutable evidence. The event must therefore be verifiable by a third party.

Finally, we would like to point out that information systems can also be related to one another. This can be explained by an example. With the dice roll on page 41 we had stated that at first one could only observe whether the roll resulted in an even or odd score. However, in the second σ-algebra it was also possible to verify the precise score. If the σ-algebra can be understood as an information system, it should be clear that the second system is more informative than the first one. After

---

[17]If it is known that an even number was rolled, i.e., $\{2, 4, 6\} \in \mathcal{F}$, it is also known that an odd number was not rolled, i.e., $\{1, 3, 5\} = \Omega \backslash \{2, 4, 6\} \in \mathcal{F}$.
[18]See page 39.

all, one learns something about the precise score and not only whether the score can be divided by two without any remainder. This relation of the two sets of information can be represented mathematically simply by

$$\mathcal{F}_1 \subset \mathcal{F}_2. \tag{3.28}$$

Each event observable in the information system $\mathcal{F}_1$ can also be observed in the information system $\mathcal{F}_2$. It is also said that $\mathcal{F}_2$ is "finer" than $\mathcal{F}_1$. The opposite, of course, does not apply. In this way, σ-algebras naturally reflect characteristics of information systems that otherwise can only be described with significant formal efforts.

3.4 Further Examples: Infinite Number of States and Times

Key Date Principle Finance theorists often analyze models in which the present ($t = 0$) and the future ($t > 0$) are considered. If situations with several future times ($t = 1, 2, \ldots, T$) are examined, there are two possible approaches. You can either work with discrete-time or continuous-time models.[19] Regardless of which approach is used a basic principle common to both must be pointed out:

> All considerations made in the context of multi-period models take place in the present ($t = 0$).

While being in $t = 0$ we think about what we *now* know about the future ($t = 1, 2, \ldots$). However, as we move in time our knowledge about the future may improve, but this aspect is of absolutely no relevance *now* (i.e., in $t = 0$).

Several Points in Time In this section we will deal with more complex σ-algebras. They comprise either several times or an infinite number of elementary events.

Example 3.4 (Binomial Model) We refer again to the example of the binomial model (see Fig. 3.4 on page 24). The model consists of exactly three points in time. The individual paths are described by sequences of u and d. There are a total of eight paths, each representing an elementary event. As can be seen at $t = 3$ only four different results are possible: the "state" uud at $t = 3$ can result from three entirely different paths: uud, udu, and duu.

---

[19]For the difference between both approaches we refer the reader to pages 23 ff.

Fig. 3.4 Binomial model
with $T = 3$
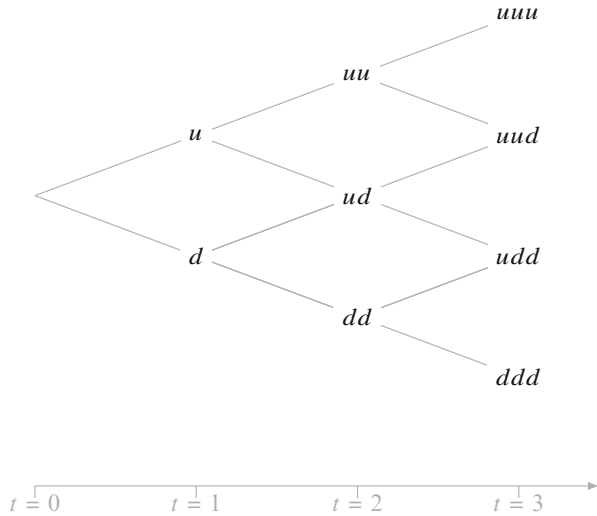
We now turn our attention to a σ-algebra, which may consist of the measurable
sets described below,

$$\mathcal{F}_2 = \Big\{ \{uuu, uud\}, \ \{udu, udd\}, \ \{duu, dud\}, \ \{ddu, ddd\}, \ \ldots \Big\}. \qquad (3.29)$$

The ... sign indicates all those sets that can be constructed by forming unions and
intersections from the four measurable sets $\{uuu, uud\}$, $\{udu, udd\}$, $\{duu, dud\}$,
and $\{ddu, ddd\}$. This means, for example, that the set $\{uuuu, uud, uuu, udd\}$ and
$\Omega \backslash \{uuuu, uud\}$ are also contained in the σ-algebra. Subsets of the above four events
are not included in the σ-algebra. Therefore the event $\{uuu\}$ is *not* measurable. The
same applies to $\{uud\}$ and $\{udu\}$.

It is also said that the σ-algebra considered here is "generated" by the four
elements $\{uuu, uud\}$, $\{udu, udd\}$, $\{duu, dud\}$, and $\{ddu, ddd\}$ mentioned above.

This σ-algebra can also be thought of as an information system. The only thing
required is to understand what makes this algebra a measurable set. Let us look, for
example, at the two measurable sets

$$\{uuu, uud\} \quad \text{and} \quad \{udu, udd\}.$$

What do these two sets have in common and what makes them different? They each consist of two elementary events, and we can assign a probability to each of the two sets. However, the following considerations are crucial:

1. Individual elementary events such as $\{uud\}$[20] are not observable. The smallest events that can be observed contain at least two elementary events.
2. If an elementary event is observable, then the same set of events also contains the elementary event which has the same two initial movements. If the measurable set contains **uu**d, it also contains **uu**u. And if **ud**u is an element of a measurable set, then this must also apply to **ud**d.

We had mentioned that σ-algebras can be interpreted as information systems. Such an information system is constructed in a way that a decision-maker can distinguish precisely which upward or downward movements will have occurred up to $t = 2$. For example, at event $\{uuu, uud\}$ the decision-maker is certain that two consecutive u-movements must have occurred, uncertainty however prevails with regard to the third movement. Similarly, at event $\{udu, udd\}$, the decision-maker is certain that up to $t = 2$ there has been one upward and one downward movement, but he does not know what the third movement will be. So we can present information about what the first two movements were, but not which movement will follow next. Thus, the σ-algebra contains the information we currently ($t = 0$) assume to have at $t = 2$, but not at time $t = 3$. The events which only differ in $t = 3$ are always combined in each measurable set. To summarize: this σ-algebra describes the information that a decision-maker today thinks he will have at $t = 2$.

We will present a further example to reinforce this idea.

Example 3.5 (Binomial Model) Let us continue with the previous example. How should a σ-algebra be constructed in order to describe the information a decision-maker will likely have in $t = 1$? Let us look at event

$$udu$$

and assume that it is part of a measurable set. At $t = 1$ the decision-maker will only know whether the first movement was up (u) or down (d). If the first movement was u, in $t = 1$ the decision-maker cannot yet distinguish whether this event or one of the three other events (udd, uuu, or uud) have occurred. Any measurable set that contains udu must also contain the three other events.

---

[20]Note that this is an event other than udu, although both paths lead to the same result at $t = 3$ as part of a recombining model. See the explanations on page 25.

Similarly, a set with event duu must also contain the three events dud, ddu, and ddd, because these four events are not yet distinguishable in $t = 1$. The generating sets of such a σ-algebra are therefore

$$\mathcal{F}_1 = \Big\{\{uuu, uud, udu, udd\}, \{duu, dud, ddu, ddd\}, \ldots \Big\}. \qquad (3.30)$$

The sign $\ldots$ is to be understood as above. However, in this simple case only two sets are added, namely the empty set $\emptyset$ and the total set Ω.

In comparing the last two examples a further reference can be made to the interpretation of a σ-algebra as an information system. While Example 3.5 describes the information available to a decision-maker at $t = 1$, Example 3.4 specifies the information that he currently believes to have at $t = 2$. Obviously, the information becomes more comprehensive as time goes by. The second σ-algebra at $t = 2$ is greater than the algebra at $t = 1$. Thus

$$\mathcal{F}_1 \subset \mathcal{F}_2. \qquad (3.31)$$

It is also said that both σ-algebras form a filtration. If one examines a binomial model with several points in time, a σ-algebra can be formulated for each t, which describes the information available at $t \geq 1$ from today's perspective. It can be stated that these algebras get "finer and finer,"

$$\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3 \ldots \qquad (3.32)$$

Economically, this corresponds to the idea that a decision-maker gains more and more knowledge over time and that no information is lost with passing time.

An Infinite Number of Share Prices   Consider the price of a stock at a future point in time and assume that the event space includes not only the options u and d, but the set of (nonnegative) real numbers, $\Omega = \mathbb{R}_+$. It is not easy to determine which events should be regarded as measurable. We will deal with this question in the following example.

Example 3.6 (Share Price)   For convenience we consider an event space containing all real numbers (and not only the nonnegative ones), i.e., $\Omega = \mathbb{R}$.

Proceeding in the same way as with natural numbers and assigning a positive probability to every conceivable value leads to a serious problem. Let's assume that the German DAX is measured in real numbers and all values between 8000 and 15,000 are possible. Let us further assume that we would like to model the DAX as a rectangular distribution. If every real number between 8000 and 15,000 had

the same positive probability, the sum of these probabilities would inevitably go to infinity and not to one. Even probabilities of zero do not avoid the problem, because these probabilities sum to zero and not to one. These conclusions remain valid even when other distributions are being used.

For this reason we better not start with the requirement that the sets $\mathbb{N}$ and $\mathbb{Q}$ are measurable. But how should we proceed? If *single numerical values* must be unlikely, a sensible way to proceed is with *intervals of numbers*. As a first step we specify all closed intervals $[a, b]$ for any real numbers $a \leq b$ as measurable. Subsequently we examine which other sets are measurable if we apply the design rules from page 39 and proceed as follows:

1. By letting $a = b$ one knows immediately that the point sets $\{a\}$, which contain only the real number a, can be measured.
2. If the set $\{a\}$ is measurable, according to rule 2 its complement can also be measured. Thus, the sets $\mathbb{R}\backslash\{a\}$ and $\mathbb{R}\backslash\{b\}$ must be measurable.
3. Consequently the open intervals (a, b) can also be measured. This follows from the fact that the intersection of $\mathbb{R}\backslash\{a\}$ and $\mathbb{R}\backslash\{b\}$ with $[a, b]$ is the same as (a, b); and we had proven on page 39 that the intersection of any subsets is measurable.
4. Since point sets are measurable, the set of all rational numbers, in short $\mathbb{Q}$, must also be measurable, because it is a union of all point sets.

All sets that can be generated with the construction mechanism used here are called *Borel-measurable sets*.[21] One particular characteristic of these sets is the fact that the open intervals can be measured.[22] Based on rule 3 all sets are Borel-measurable which can be composed of a finite number of open intervals. A union of open intervals is also called an *open set*. An open set is characterized by the fact that not only point sets x but also all—however small—open intervals around x are part of the set. Open sets can be thought of sets without "borders" (such as the closed interval $[0, 1]$).

To make matters more complex we will consider not only an infinite number of values of a share price but also its continuous development. Handling both elements is what the Brownian motion is all about. Let us now describe the underlying σ-algebra.

---

[21] Félix Édouard Justin Émile Borel (1871–1956, French mathematician).

[22] It may be hard to imagine that there exist sets that are not Borel-measurable, nevertheless they can be constructed. However, the design specifications for such sets are highly complicated.
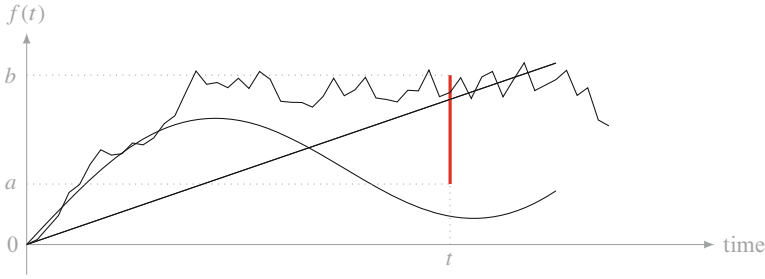
Fig. 3.5 Three elementary events in the event space $C[0, \infty)$

Example 3.7 (Share Price Evolution) With this example we will now approach the Brownian motion. Wiener[23] was the first to describe what a measurable set in $C[0, \infty)$ could look like.[24]

Since the set $C[0, \infty)$ has an infinite number of functions, characterizing the measurable sets is anything but a trivial task. One cannot expect the σ-algebra to consist only of a finite number of functions.

Constructive action has to be used again. In a first step, one describes specific measurable sets and in a second step one allows these initial sets to form their union or intersection. In the following we will concentrate on the first step, a task that is far from being elementary. The initial sets which are defined as measurable consist of the following continuous functions:

First step (one point in time)   We concentrate on one single point in time $t > 0$ and two real numbers $a < b$. The measurable set defined in the first step includes all those functions with a value being exactly in the interval $[a, b]$ at time t. This is illustrated in Fig. 3.5.[25]

At time t one can see a red vertical line running from a to b on the ordinate. You can recognize that two of the paths intersect this vertical line. The sinusoidal path, however, runs in such a way that it neither intersects nor touches the red line. Now one has to consider the set of *all* continuous functions that go through the red line, i.e.,

$$Z = \{f : \text{function } f \text{ is continuous on } [0, T] \quad \text{and} \quad f(t) \in [a, b]\}. \quad (3.33)$$

---

[23]Norbert Wiener (1894–1964, American mathematician). It is often said that Wiener was the first to define what is now known as the Wiener measure. This is not entirely precise, because Wiener published his work in 1923, but the measurement theory was put on an axiomatic basis only in 1933 by Kolmogoroff. However, Wiener described in his paper how to calculate measures of different sets of paths and is therefore with good reason called the founder of the stochastic theory of the Brownian motion.

[24]If you do not remember which event space we have designated with $C[0, \infty)$ see page 26.

[25]The same three functions were shown in Fig. 2.6 on page 27.
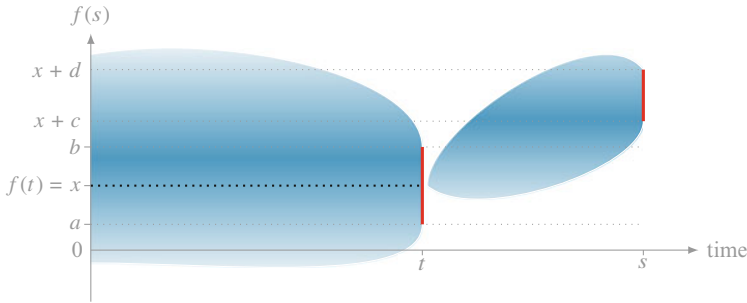
Fig. 3.6 Cylinder sets with two fixed points in time

The set Z being characterized by (3.33) is measurable; this property applies regardless of how the time t and the limits a and b are chosen.[26]

Second step (two points in time) Now we are not looking at one but at two points in time with $0 < t < s$. In addition to the numbers a and b two more numbers $c < d$ are given. The measurable set that is defined in the second step includes the paths running through the interval $[a, b]$ at t. However, there is another requirement which plays a central role when looking at the second point in time s. The development of the paths from our (to be defined) measurable set Z should not be arbitrary between t and s; rather, the difference of the function values $f(s) - f(t)$ must belong to the interval $[c, d]$.

It is not easy to express this statement precisely: each measurable event should pass the interval $[a, b]$ at time t, that is $f(t) \in [a, b]$. In addition, the relation $f(s) - f(t) \in [c, d]$ should apply for measurable events. This means the following: if, for example, the event $f(t) = x$ would happen at t, then measurable events at time s should pass through the interval $f(s) \in [x+c, x+d]$. The interval from which the function values originate is shifted with each value $f(t) = x$.

With Fig. 3.6 we try to illustrate this aspect. It should be understood that the position of the vertical line at time s depends on where the event passes the vertical line relevant for time t. In other words, the larger (smaller) x is, the higher (lower) the interval relevant at time s is located. Hence, we have not visualized all conceivable developments. Rather, we have limited ourselves to those developments which belong to a fixed value $f(t) = x$. In principle Fig. 3.6 should be extended for each $x \in [a, b]$.

We must emphasize that the blue-shaded areas in Fig. 3.6 could lead to misinterpretations because one could think that measurable events are restricted to the blue areas at all. Of course, one can imagine continuous functions that go through both vertical lines and still fall outside the blue areas. Functions with

---

[26]Because the red line can be understood as a (very thin) cylinder such a set is often called cylinder set.

these properties are also measurable. However, at times t and s they must have function values that are specified in

$$f(t) \in [a, b] \quad \text{and} \quad f(s) - f(t) \in [c, d]. \tag{3.34}$$

Indeed, the blue areas between the times t and s have neither an upper nor a lower bound. Now we can state that the set constructed in this way

$$Z = \big\{ f : \text{function } f \text{ is continuous on } [0, \ T] \quad \text{and}$$

$$f(t) \in [a, b], \ f(s) - f(t) \in [c, d] \big\} \tag{3.35}$$

must also be measurable. This property should also apply regardless of how the times t, s and the four numbers a, b, c, d are selected.

Next steps   These constructions have to be repeated for three, four, and any number of other times. However, the number of these points in time always remains finite. The resulting sets of continuous functions are measurable.

As stated in Sect. 3.2, using these measurable sets one can form unions, intersections, and complements.

Let us summarize. The measurable sets are obtained by a two-stage process. First, specific subsets (initial sets) are determined which should be measurable by definition. Additional measurable sets are formed with the help of the rules discussed above.[27] The resulting measurable sets may be different from each other depending on the initial sets which are chosen in the first step.

The symbol $\mathcal{F}$ is used to denote the sets that form a σ-algebra. The totality of basic sets, σ-algebra, and measure μ is called *measure space* $(\Omega, \mathcal{F}, \mu)$.

Usually σ-algebras are constructed as shown in our examples: we start with specific sets and add further sets by unions, intersections, and complements. It is stated that the σ-algebra is *generated* by these specific sets. For example, if the generating sets can be described by the symbol X, one can write $\sigma(X)$. In the case of Borel-measurable sets we could use the notation $\sigma\big(\underbrace{[a, \ b]_{a, \ b \in \mathbb{R}}}_{:=X}\big)$ for the σ-algebra.

3.5   Definition of a Measure

After introducing measurable sets we will define what constitutes a measure and proceed in the same way as before.

- As a first step the measure is specified for those sets which are directly measurable.

---

[27]See (3.2), (3.3), and (3.13) on page 30 f.

- All sets which are not directly measurable can be obtained by union, intersection, and complement of directly measurable sets. In the case of finite unions we use the property of additivity given in Eq. (3.3) to determine the measure of such sets and in the case of infinite unions the property of σ-additivity given in Eq. (3.13).[28]

Hence, we can assign a unique number to each measurable set which we define as its measure.

Definition 3.2 (σ-Algebra, Measure) A measure is a mapping of a σ-algebra $\mathcal{F}$ into the real numbers

$$\mu : \mathcal{F} \rightarrow \mathbb{R}. \tag{3.36}$$

The properties of nonnegativity according to (3.2), additivity according to (3.3), and σ-additivity according to (3.13) are valid.

Mind that we waive the property of shift invariance. Since we will not only look at probability measures, we allow for $\mu(\Omega) \neq 1$. On the following pages we discuss the construction of measures in the light of two examples (dice roll and, later, real numbers).

Example 3.8 (Dice Roll) Let us start with the dice roll. In the following we will use a more appropriate notation. For the set of all scores which are possible with a dice roll, we can write

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

The set of even numbers is written as Ω^e and the set of odd numbers Ω^o:

$$\Omega^e = \{2, 4, 6\} \qquad \text{and} \qquad \Omega^o = \{1, 3, 5\}.$$

The matter is very simple at $t = 1$: the only information available is whether the score is even or odd. We stipulate

$$\mu(\Omega^e) = \mu(\Omega^o) = \frac{1}{2}.$$

Events such as {4}, {5}, {6} will not have their own measure because they cannot be measured.

---

[28]See page 33.

At $t = 2$ the elementary events are also measurable. In this case it makes sense to define the measure as follows:

$$\mu(\{1\}) = \ldots = \mu(\{6\}) = \frac{1}{6}.$$

None of the above is particularly remarkable.

3.6   Stieltjes Measure

The matter gets more interesting when we look at the Borel-measurable sets of the real line.[29] We had started the construction of the measurable sets with the closed intervals $[a, b]$. Let us consider a monotonously growing and differentiable[30] function

$$g : \mathbb{R} \to \mathbb{R}. \qquad (3.37)$$

Examples for such functions are $g(a) = e^a$ or $g(a) = \Phi(a)$, where $\Phi(a)$ is the distribution function of the standard normal distribution. We stipulate that

$$\mu([a, b]) := g(b) - g(a) \qquad (3.38)$$

applies. μ is also referred to as the Stieltjes measure.[31] It is obviously defined as a measure of closed intervals.

In the following we show that we have therefore also defined the measure of the open intervals: from our considerations on page 46 we know that point sets and open intervals are also measurable. For point sets, it follows directly

$$\mu(\{a\}) = g(a) - g(a) = 0, \qquad (3.39)$$

implying they have measure zero. We can furthermore write a closed interval $[a, b]$ as union $\{a\} \cup (a, b) \cup \{b\}$, where the three subsets are pairwise disjoint. Hence, because of property of a measure (3.3)
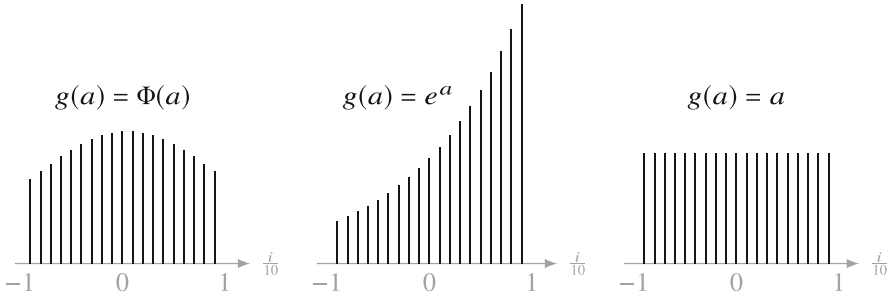
$$g(b) - g(a) = \mu([a, b]) = \mu(\{a\}) + \mu((a, b)) + \mu(\{b\})$$
$$= \mu((a, b)). \qquad (3.40)$$

---

[29]See page 47.

[30]It should be noted that every differentiable function is continuous.

[31]Thomas Jean Stieltjes (1856–1894, Dutch mathematician).

$g(a) = \Phi(a)$ $g(a) = e^a$ $g(a) = a$

Fig. 3.7 Stieltjes measures $\mu\left(\left[\frac{i}{10}, \frac{i+1}{10}\right]\right)$ for different generating functions $g(\cdot)$ depending on $\frac{i}{10}$

We recognize that the open interval has the same measure as the closed interval. It is easy to conclude that the half-open intervals $[a, b)$ and $(a, b]$ have the measure $g(b) - g(a)$ too.

Example 3.9 (Real Numbers) For three specific functions g we will characterize the resulting probabilities more precisely. For this purpose we first choose the function $g(a) = \Phi(a)$, then the function $g(a) = e^a$, and finally $g(a) = a$. To understand the measure applied over any range of the real line, we focus on the closed interval $[-1, 1]$ and break it down into twenty subintervals. For each of the twenty subintervals we define a measure $\mu([\frac{i}{10}, \frac{i+1}{10}])$ with $i = -10, -9, \ldots, +9$ and plot function values. Figure 3.7 shows the effects that emerge with various functions g.[32] These are our observations:

- With $g(a) = \Phi(a)$ the measure of the interval increases as i goes to zero. The entire real line has the measure 1.
- With $g(a) = e^a$ the measure of the interval increases as i grows. The entire real line has infinite measure.
- With $g(a) = a$ the measure of the interval does not change with i changing. The whole real line has infinite measure again.

We inevitably determine the measure of a subinterval as the difference between the function values $g\left(\frac{i+1}{10}\right) - g\left(\frac{i}{10}\right)$. Therefore, it should come as no surprise that the figures look like the first derivatives of the respective measurement functions.

In the case $g(a) = a$ the result is also called *Lebesgue measure* and is denoted as λ. It corresponds to our "common" perceptions of length units. In the other two cases the lengths are "weighted," whereby the weight depends on where the interval to be measured is located on the real line.

---

[32]In case $g(\infty) = 1$ and $g(-\infty) = 0$ we have a probability measure (the entire real line has the measure 1). The graphs would then reflect the densities. This corresponds to $g(a) = \Phi(a)$.

3.7   Dirac Measure

We will later revert to an admittedly degenerated probability measure where only the number a is highly probable. In fact it is certain! All other numbers are absolutely unlikely. This measure can be called degenerate because the numbers other than a are impossible. The reader will subsequently understand why such a measure can be important in the discussion about the Lebesgue integral.

To formally define the Dirac measure we again look at the real line $\Omega = \mathbb{R}$. For the fixed number a we use

$$\mu([a, a]) = 1, \qquad \mu((-\infty, a)) = \mu((a, \infty)) = 0. \tag{3.41}$$

This measure is known as *Dirac measure* and is usually denoted by the letter δ_a.[33]

3.8   Null Sets and the Almost-Everywhere Property

The sets N having no weight for a given measure μ (i.e., $\mu(N) = 0$) are of special importance. Such sets are also called null sets. The complement $N^c = \Omega \setminus N$ has full measure (which can be infinite if Ω has no finite measure).

Null sets play an important role. To understand this consider the function

$$f(x) = n \quad \text{with } n \leq x < n + 1, \quad n \in \mathbb{N}. \tag{3.42}$$

This function resembles a staircase that jumps up one unit at any natural number and remains constant between these numbers as shown in Fig. 3.8.[34] The function $f(x)$ has discontinuities in the places of the natural numbers and is otherwise piecewise constant. This phenomenon is anything but noteworthy. However, the discontinuities must not be ignored because they are responsible for the fact that certain mathematical operations (derivations, limits, etc.) cannot readily be applied. The staircase function is, however, recalcitrant and annoying.

Null sets offer a mathematically precise way to deal with these annoying discontinuities.[35] For this purpose we look at those points on the real line where f is discontinuous which is precisely the set of natural numbers $\mathbb{N}$. Although there are infinitely many natural numbers, the entire set is rather small in comparison to the remaining real numbers. In order to cope with the problem we look at a

---

[33]Paul Adrien Maurice Dirac (1902–1984, British physicist).

[34]The graph directly indicates the name of this function, as it actually resembles a staircase. However, the representation is mathematically imprecise, because the function at $x = 1, 2, \dots$ obviously does not produce a single value but an interval of values. Of course, this is not allowed for functions. Technically speaking, the vertical lines in Fig. 3.8 should be removed.

[35]It should be emphasized that the discontinuities here are only one illustrative example against the background of which it is easy to discuss. We can also control other "unwanted" properties of a function with null sets.

Fig. 3.8 Staircase function
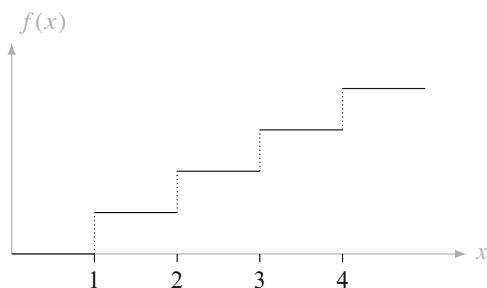
measure on the set of real numbers and try to take advantage of the described property. This is achieved by selecting the measure so that $\mu(\mathbb{N}) = 0$. Obviously, the staircase function $f(x)$ is constant outside the set $\mathbb{N}$; discontinuities are only present in $\mathbb{N}$, and this set has now measure zero. In our example, this permits the statement "The staircase function f is μ-almost everywhere continuous," because the property (here: continuity of the function) applies everywhere except for the null set. The trick is not to deny unwanted properties of a function, but to ignore them by assigning them a measure that does not matter at all.

If μ were a probability measure, we would obviously ignore events that have measure zero. These are simply unlikely events. Our above statement would then read "The staircase function f is continuous except for unlikely events." If μ measures the weight of objects, we could state "The function f is continuous except for elements without mass." Null sets do not attempt to deny the existence of disturbing properties of functions; rather null sets are used to disregard these characteristics. The staircase function remains discontinuous, but the discontinuities are unlikely, insignificant, without mass, etc., in short: a null set. We can state:

Definition 3.3 A property applies μ-almost everywhere[36] exactly when it applies to all elements of the set $N^c = \Omega \backslash N$.

Note that the choice of the measure plays a crucial role and it is very important which μ is used. If two different measures μ_1 and μ_2 are defined, it is quite possible that one and the same function μ_1-almost everywhere is continuous, while this property is lost if μ_2 is selected. It is therefore important to choose the measure μ skillfully.

Please also note that null sets of a measure can be very large, indeed infinitely large. For example, it can be shown that the set of rational numbers is a null set when a Stieltjes measure is employed. To intuitively understand the implication one should imagine all rational numbers on a real line. If one adds a point to each of these fractions, "almost" the entire real line will be drawn: for each real number

---

[36]Often, the term "f is μ-almost everywhere continuous" is abbreviated by "f is μ-a.e. continuous."

selected one can find infinitely many rational numbers which are arbitrarily close. Nevertheless, those numbers form a null set if a Stieltjes measure is used. Null sets can therefore be infinitely large and still have measure zero.

Finally, we give four statements which apply almost everywhere under a specific measure.

- The x^2 function is Lebesgue-almost everywhere positive.[37]
- Each and every number is Dirac δ_a-almost everywhere equal to a.[38]
- The absolute value function $|x|$ can Lebesgue-almost everywhere be differentiated.[39]
- The staircase function in Fig. 3.8 can Lebesgue-almost everywhere be differentiated.[40]

---

[37]$x = 0$ is the only point where the function is not positive. This set has Lebesgue measure zero.

[38]The numbers that do not equal a are given by the intervals $(-\infty, a)$ and (a, ∞). This is a very large set but its Dirac measure is nonetheless zero.

[39]The number where the function cannot be differentiated is $x = 0$. This set has the Lebesgue measure zero.

[40]The points at which the function cannot be differentiated are the set of natural numbers. This set has the Lebesgue measure zero.