Expectation and Lebesgue Integral

5

5.1   Definition of Expectation: A Problem

In the previous chapter we dealt with the concept of probability in the context of any event space Ω. We described how to proceed appropriately to define a probability as a measure of a set. Now we are focusing on the determination of expectations and variances.

Why is the calculation of expected values a problem at all? Let us continue with the example of the dice roll. There are six possible states and to each of them we can assign a random variable $X(\omega)$. The expectation of these random variables can now be determined very easily by multiplying each realization by the probability of their occurrence and then adding the six values,

$$E[X] := \sum_{\omega=1}^{6} X(\omega) \cdot \frac{1}{6}. \tag{5.1}$$

Calculating the expectation gets more complicated when dealing with a larger state space like the real numbers, $\Omega = \mathbb{R}$. A summation of the form

$$\sum_{x \in \mathbb{R}} \tag{5.2}$$

simply will not work since the real numbers cannot be enumerated exhaustively.[1] The summation rule does not make any sense.

One might be inclined to use the Riemann[2] integral as a sensible alternative. Before realizing that this does not work either, we will discuss the construction of

---

[1] For an explanation see page 106 f.

[2] Georg Friedrich Bernhard Riemann (1826–1866, German mathematician).

the Riemann integral in necessary detail. For ease of presentation we will restrict the discussion to strictly monotonously growing functions over the real numbers,

$$f : \mathbb{R} \rightarrow \mathbb{R}. \tag{5.3}$$

5.2   Riemann Integral

The definite Riemann integral over the interval $[a, b]$ is constructed by splitting this interval into many small subintervals $[t_i, t_{i+1}]$. The i index runs from $i = 1$ to $i = n$. A rectangle with a width of $t_{i+1} - t_i$ is placed over each subinterval. Several options exist for selecting the height of such a rectangle: you can use the lower function value $f(t_i)$, the upper function value $f(t_{i+1})$ or any value in between, that is $f(t_i^*)$ with $t_i < t_i^* < t_{i+1}$. When using the left function values in determining the area of each rectangle and adding all rectangles, we obtain their lower sum:

$$\text{lower sum}_n = \sum_{i=1}^{n} f(t_i)(t_{i+1} - t_i). \tag{5.4}$$

If we use the right function values, we obtain the upper sum:

$$\text{upper sum}_n = \sum_{i=1}^{n} f(t_{i+1})(t_{i+1} - t_i). \tag{5.5}$$

If the integral shall represent the area below the function, the upper sum (for a monotonically growing function) will be larger and the lower sum will be smaller than the area below the function. If one allows the decomposition to become ever finer ($n \rightarrow \infty$), it all depends on how the two sums will behave. Riemann succeeded in proving that the choice of the function value for certain functions is irrelevant. Regardless of which function value is selected, the resulting sum of all rectangles converges to the same value if the number of subintervals n goes to infinity.[3] This result is known as the Riemann integral

$$\int_a^b f(t)\,dt := \lim_{n\to\infty} \text{upper sum}_n = \lim_{n\to\infty} \text{lower sum}_n. \tag{5.6}$$

Figure 5.1 illustrates the process of constructing the Riemann integral for the case of a triple, a sixfold, and finally an infinite segmentation of the interval $[a, b]$.

---

[3]This applies, for example, to continuous functions. Meanwhile, a function is *defined* as Riemann integrable exactly when upper and lower sum converge to each other at any decomposition.
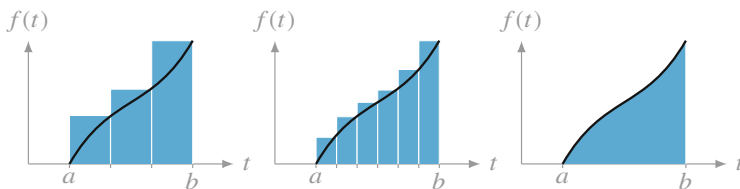
Fig. 5.1 Illustration of the upper sums of the Riemann integral with three, six, and infinitely many subintervals

In order to apply the Riemann integral, certain requirements must be met. In particular, the definition range of the function to be integrated must be a closed interval from the set of real numbers, because no other interval can be divided into an infinite number of subintervals. The function $f(\omega)$ of the dice roll example from page 65 has no such definition range.
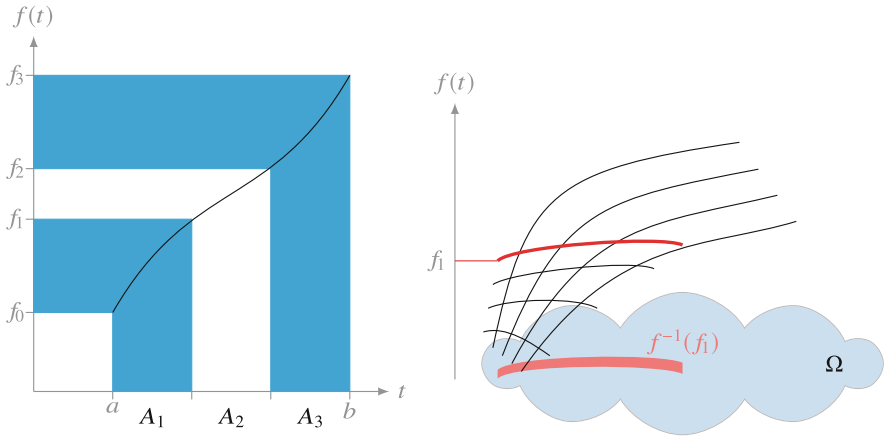
5.3 Lebesgue Integral

The state space $\Omega = \mathbb{R}$ is a very special prerequisite. How should the idea of integration be applied to a situation where the definition range of a function does not cover a closed interval of real numbers? Earlier we have pointed out that state spaces other than those covering the real numbers do exist.[4] The state space $\Omega = C[0, \infty)$ includes all continuous functions starting at zero. An important question must be addressed. How should this set be divided into equal-sized subintervals? We can order real numbers by their value and thus form intervals; with continuous functions, however, such a procedure is not possible. Since the Riemann integral cannot be used we must explore a different way of calculating the expected value of random variables over $C[0, \infty)$.

The French mathematician Lebesgue had the ingenious idea of how to proceed. He suggested splitting *the ordinate* into subintervals rather than the abscissa. Regardless of the characteristics of the state space Ω each corresponding function maps into the real numbers. The specific segmentation of the ordinate, however, depends on the actual random variable. The procedure is described below and illustrated in Fig. 5.2 for the same function used in Fig. 5.1.

With Lebesgue integration the ordinate is split into subintervals. In Fig. 5.2 we divide the interval $[f_0, f_3)$ into three subintervals $[f_0, f_1)$, $[f_1, f_2)$, and $[f_2, f_3)$. Doing so allows us to identify three subsets on the abscissa. The subsets A_1, A_2,

---

[4]See page 27.

Fig. 5.2 Lebesgue integral: sets of inverse images of a function to be integrated $f(t)$

and A_3 result in principle from the inverse images of the function f, thus

$$A_i := f^{-1}([f_{i-1}, f_i)), \qquad i = 1, 2, 3. \tag{5.7}$$

As shown in Fig. 5.2 the interval $[f_2, f_3)$ of the ordinate is assigned the inverse image A_3 on the abscissa. Similarly, the intervals $[f_1, f_2)$ and $[f_0, f_1)$ map into the inverse images A_2 and A_1 respectively. In order to be able to integrate, the subsets represented by the inverse images must be measurable, i.e., come from the σ-algebra.

We want to understand the implication for the $f(\cdot)$ function. Will every arbitrary function be integrable using this idea? If we divide the ordinate into subintervals, the corresponding subsets are automatically created on the abscissa. If the interval $[f_{i-1}, f_i)$ is part of a segmentation, the corresponding subset on the abscissa is defined by

$$\begin{aligned} A_i := &\{\omega \ : \ f_{i-1} \le f(\omega) < f_i\} \\ = &\{\omega \ : \ f(\omega) < f_i\} \cap \Omega \setminus \{\omega \ : \ f(\omega) < f_i\}. \end{aligned} \tag{5.8}$$

It is therefore sufficient to require that the inverse image $\{\omega \ : \ f(\omega) < a\}$ of the $f(\cdot)$ function is a measurable set for all real numbers a.[5] Note that this property defines a measurable function.[6] Therefore, every measurable function is Lebesgue integrable.

---

[5]If the set $\{\omega \ : \ f(\omega) < a\}$ is measurable, then this also applies to the complement $\Omega \setminus \{\omega \ : \ f(\omega) < a\}$ as well as to the intersection of this subset with another measurable subset. Otherwise we would not have a σ-algebra.

[6]See Definition 4.1 on page 65.

Fig. 5.3 Lebesgue integral: illustration of the upper sum
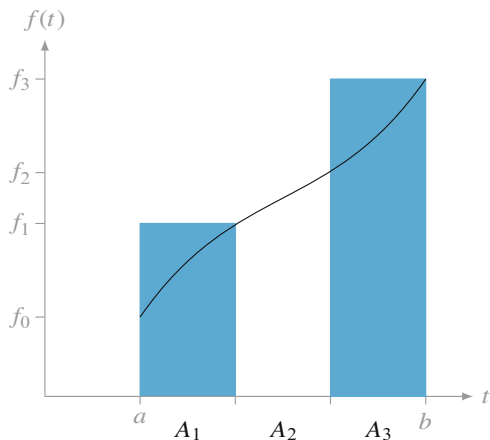
After these considerations we can present the idea of Lebesgues in its entirety. Analog to the Riemann integral (which measures the area under a function) we will approximate this area by using upper and lower sums again. If a function $f(\cdot)$ is measurable the integral can be approximated by the "upper sum"

$$\text{upper sum} := f_1 \cdot \mu(A_1) + f_2 \cdot \mu(A_2) + f_3 \cdot \mu(A_3). \tag{5.9}$$

First we realize why this expression is always greater than the value of the integral and therefore represents a first approximation of the area like an upper sum. For this purpose we redraw Fig. 5.2 and focus on the rectangles of the approximation $f_1 \cdot \mu(A_1), \dots, f_3 \cdot \mu(A_3)$ from (5.9) leading to Fig. 5.3. Two rectangles are highlighted. They have the widths $\mu(A_1)$ and $\mu(A_3)$ and the heights f_1 and f_3 respectively. The sum of these rectangles overstates the area because the function runs below the upper corners of the rectangles. The area we have determined by using the upper sum in expression (5.9) is greater than the integral. We can construct a lower sum in analogy

$$\text{lower sum} := f_0 \cdot \mu(A_1) + f_1 \cdot \mu(A_2) + f_2 \cdot \mu(A_3). \tag{5.10}$$

Let us suppose that the two sums converge against the same value in the limit when the subintervals on the ordinate get infinitely small. If the two limits converge to the same value for any segmentation of the ordinate, the function $f(\cdot)$ is called *Lebesgue integrable*. This value is the *Lebesgue integral* of the function and usually written in the form

$$\int_{\Omega} f(\omega) \, d\mu(\omega). \tag{5.11}$$

Fig. 5.4 Interaction of measurement and function in the Lebesgue integral
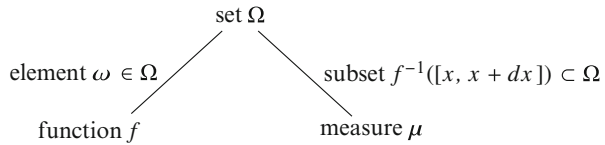
Note that in the Lebesgue integral both the f function and the measure μ refer to the basic set Ω, however, in a different way. The function $f(\omega)$ assigns a real number to each element of Ω. However, we must proceed differently when dealing with the measure $d\mu(\omega)$: a subset $f^{-1}([x, x + dx]) \subset \Omega$ and not a single element is assigned as illustrated in Fig. 5.4.

The calculation rules for the Lebesgue integral are quite similar to the Riemann integral. These rules are:

Proposition 5.1 (Calculation Rules for Lebesgue Integrals) *For Lebesgue integrable functions f and g the following applies*

$$\int_\Omega (f + g)\, d\mu = \int_\Omega f\, d\mu + \int_\Omega g\, d\mu \qquad (5.12)$$

$$\int_\Omega a \cdot f\, d\mu = a \int_\Omega f\, d\mu, \qquad \forall a \in \mathbb{R}. \qquad (5.13)$$

Applying these rules the integral over f cannot be smaller than the integral over g if $f(x) \geq g(x)$ for all $x \in \Omega$. To prove this claim we first look at the rule

$$\int_\Omega f\, d\mu - \int_\Omega g\, d\mu = \int_\Omega f - g\, d\mu. \qquad (5.14)$$

The difference $f - g$ is nonnegative according to the assumption $f(x) \geq g(x)$. However, it follows from the construction of the Lebesgue integral that we multiply and add these nonnegative differences by nonnegative values of the measure. The integral of the difference must not be negative, and this is what we have asserted.

Let us illustrate the Lebesgue integrability using three examples.

Example 5.1 (Dirichlet Function) We consider the so-called Dirichlet function[7]

$$D(x) = \begin{cases} 1, & \text{if } x \text{ rational;} \\ 0, & \text{else.} \end{cases} \qquad (5.15)$$

---

[7]Peter Gustav Lejeune Dirichlet (1805–1859, German mathematician).

We are interested in the value the Lebesgue integral has over a Stieltjes measure. We only assume $\mu(\mathbb{Q}) = 0$ and $g(1) = 1$. The first property is always true with any Stieltjes measure as shown earlier.

To calculate the integral we divide the ordinate into the following five subintervals[8]

$$\underbrace{(\infty, 1)}_{(f_1, f_2)}, \underbrace{[1, 1]}_{[f_2, f_2]}, \underbrace{(1, 0)}_{(f_2, f_3)}, \underbrace{[0, 0]}_{[f_3, f_3]}, \underbrace{(0, -\infty)}_{(f_3, f_4)}.$$

These subintervals have inverse images of the Dirichlet function on the definition area $\mathbb{R}$, which we designate as A_1 to A_5:

$$A_1 = f^{-1}\big((\infty, 1)\big), \quad A_2 = f^{-1}\big([1, 1]\big), \quad A_3 = f^{-1}\big((1, 0)\big),$$
$$A_4 = f^{-1}\big([0, 0]\big), \quad A_5 = f^{-1}\big((0, -\infty)\big).$$

It is obvious that no function value exists in the first, third, and fifth subintervals. The corresponding images are empty, their measure is zero,

$$A_1 = A_3 = A_5 = \emptyset \quad \Rightarrow \mu(A_1) = \mu(A_3) = \mu(A_5) = 0.$$

Let us focus initially on the second and fourth intervals only. The Dirichlet function is constructed such that all rational numbers $\mathbb{Q}$ are contained in the inverse image A_2, while all irrational numbers $\mathbb{R} \setminus \mathbb{Q}$ are contained in A_4.

In order to determine the Lebesgue integral we calculate the upper sums. According to (5.10) the upper sums are

$$\int_{\mathbb{R}} D(x)\,d\mu(x) \leq \text{upper sum}$$

$$= \lim_{n \to \infty} n \cdot \mu(A_1) + 1 \cdot \mu(A_2) + \mu(A_3) + 0 \cdot \mu(A_4) + 0 \cdot \mu(A_5)$$

$$= 1\mu(\mathbb{Q}) + 0 \cdot \mu(\mathbb{R} \setminus \mathbb{Q}). \tag{5.16}$$

Each Stieltjes measure of rational numbers is zero (see page 55). From this it follows that the Stieltjes measure of the irrational numbers is one[9] resulting in

$$\int_{\mathbb{R}} D(x)\,d\mu(x) \leq 0. \tag{5.17}$$

---

[8] This procedure is not quite correct, since we should break down the ordinate into *half-open* subintervals and actually not a single interval fulfills this characteristic. It is therefore not readily clear whether the inverse images are measurable sets at all. In our case, however, this does not lead to a problem, which is why we consider our approach to be appropriate.

[9] The interval [0, 1] has the Stieltjes measure one (remember $g(1) = 1$). Since the rational numbers are countable, they have measure zero. The difference between [0, 1] and the rational numbers, i.e., the irrational numbers, must therefore have a measure of zero.

Similarly, we can determine the lower sums obtaining

$$
\int_{\mathbb{R}} D(x)\, d\mu(x) \geq \text{lower sum}
$$

$$
= 1 \cdot \mu(A_1) + 1 \cdot \mu(A_2) + 0 \cdot \mu(A_3) + 0 \cdot \mu(A_4)
$$

$$
+ \lim_{n \to -\infty} n \cdot \mu(A_5)
$$

$$
= 1 \cdot \mu(\mathbb{Q}) + 0 \cdot \mu(\mathbb{R} \setminus \mathbb{Q})
$$

$$
= 0. \tag{5.18}
$$

Hence, the Lebesgue integral of the Dirichlet function is zero.

This example is interesting for the following reason. Suppose we want to determine the classic Riemann integral $\int_a^b D(x)\, dx$. We would have to construct upper and lower sums on the interval $[a, b]$ which include the function $D(x)$. Regardless of how we break down the abscissa, the following always applies: even in any subinterval of $[a, b]$ both rational and irrational numbers exist. Thus the upper sum of the Riemann integral is always one, and the lower sum is always zero. Thus, the *Dirichlet function cannot be Riemann integrated* because the two sums do not converge against a common value. Our example illustrates the point that the Lebesgue integral can be used in situations where the Riemann integral cannot. Thus, the Lebesgue integral is far more powerful.

Example 5.2 (Power of the Lebesgue Integral) In the previous example we had used an arbitrary Stieltjes measure μ with $g(1) = 1$ and considered the particular Dirichlet function D.

Now f will be an arbitrary function with a particular measure $\mu = \delta_a$, the Dirac measure.

We calculate the integral of an arbitrary function over the Dirac measure δ_a,

$$
\int_{\Omega} f(\omega)\, d\delta_a. \tag{5.19}
$$

The Dirac measure of the set $\Omega \setminus \{a\}$ is zero. Therefore, it is meaningful to divide the ordinate into three subintervals,[10]

$$
\text{Ordinate} = \{(-\infty, f(a))\} \cup \{f(a)\} \cup \{(f(a), \infty)\}. \tag{5.20}
$$

---

[10]The middle subinterval is a closed one again. In Footnote 8 we had already pointed out that it is not strictly allowed to do this, but we do this for convenience.

Concentrating on the upper sum we obtain

$$\int_\Omega f(\omega)\, d\delta_a \;\leq\; \text{upper sum}$$

$$= f(a) \cdot \delta_a \left(\{\omega \in \Omega \,:\, f(\omega) < f(a)\}\right)$$

$$+ f(a) \cdot \delta_a \left(\{\omega \in \Omega \,:\, f(\omega) = f(a)\}\right)$$

$$+ \lim_{n\to\infty} n \cdot \delta_a \left(\{\omega \in \Omega \,:\, f(\omega) > f(a)\}\right)$$

$$= f(a) \cdot \delta_a \left(\emptyset\right) + f(a) \cdot \delta_a \left(\Omega\right) + \lim_{n\to\infty} n \cdot \delta_a \left(\emptyset\right). \qquad (5.21)$$

While the measure in the first and third terms is zero, the measure in the second term is one. This leads to $\int_\Omega f(\omega)\, d\delta_a \leq f(a)$.

Analog to the lower sum is

$$\int_\Omega f(\omega)\, d\delta_a \;\geq\; \text{lower sum}$$

$$= \lim_{n\to\infty} n \cdot \delta_a \left(\{\omega \in \Omega \,:\, f(\omega) < f(a)\}\right)$$

$$+ f(a) \cdot \delta_a \left(\{\omega \in \Omega \,:\, f(\omega) = f(a)\}\right)$$

$$+ f(a) \cdot \delta_a \left(\{\omega \in \Omega \,:\, f(\omega) > f(a)\}\right). \qquad (5.22)$$

This leads to $\int_\Omega f(\omega)\, d\delta_a \geq f(a)$.

Therefore, the Lebesgue integral equals the function value at a

$$\int_\Omega f(\omega)\, d\delta_a = f(a). \qquad (5.23)$$

The above result cannot be obtained with a Riemann integral. There is no function $g(x)$ such that

$$\int_{-\infty}^{\infty} f(x)g(x)dx = f(a) \qquad (5.24)$$

for any arbitrary function $f(\cdot)$ and any arbitrary number a: the function g would have to be infinite at a and zero otherwise.

Illustrating this we consider a function $g_n(x)$ that has value n in the neighborhood of a. Outside the neighborhood the function has value zero. The neighborhood corresponds to the interval $\left(a - \frac{1}{2n}, a + \frac{1}{2n}\right)$ which is getting smaller and smaller with increasing n. (Why we choose exactly this and no other neighborhood will become clear soon.) Figure 5.5 shows the typical course of such a function $g_n(x)$.

Let us integrate the product of $f(x)$ and $g_n(x)$. Because the product of both functions outside the neighborhood of a is zero, we can ignore this part of the
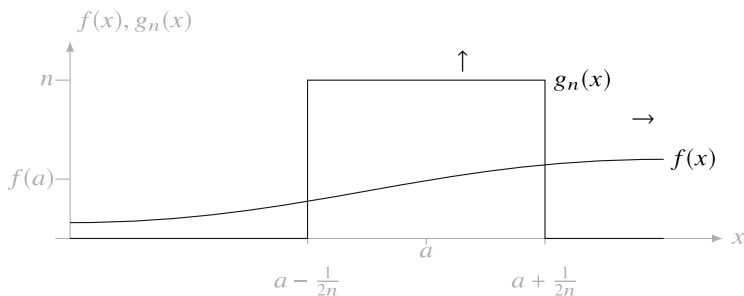
Fig. 5.5 Futile attempt to construct a Lebesgue integral with the Dirac measure using a Riemann integral

integral. From Fig. 5.5 we know the value of $g_n(x)$ in the neighborhood of a. This gives us

$$\int_{-\infty}^{\infty} f(x) \cdot g_n(x)\, dx = \int_{a-\frac{1}{2n}}^{a+\frac{1}{2n}} f(x) \cdot n\, dx. \tag{5.25}$$

Using the mean value theorem of integral calculation we can determine this integral more easily. As long as n is finite the integral corresponds approximately to the product of a value of $f(a) \cdot n$ and the length of the interval $\frac{2}{2n}$. This product equals $f(a)\, n\, \frac{2}{2n} = f(a)$. If n tends to infinity the integral converges to $f(a)$.

We conclude: anyone trying to achieve a result of the form

$$f(a) = \int_{-\infty}^{\infty} f(x)\, g(x)\, dx \tag{5.26}$$

with classic Riemann integration must use a function $g(x)$ which is zero outside a and assumes the value "infinite" at a. However, such functions do not exist in classical analysis.[11] On the contrary the result

$$f(a) = \int_{\Omega} f(\omega)\, d\mu(\omega) \tag{5.27}$$

can be obtained for any f using the Dirac measure $\mu = \delta_a$. This once again shows the power of Lebesgues' integration concept.

Example 5.3 (Lebesgue and Riemann Integral Give Identical Results) In Examples 5.1 and 5.2 we showed that a Lebesgue integral is applicable in situations where

---

[11]Functions are unique mappings into real numbers, and infinity is not a real number.

a Riemann integral is not. In this example we will show that under certain conditions a Lebesgue integral delivers a result which is identical to a Riemann integral.

We consider a strictly monotonous function[12] $f(x)$ over the interval [0, 1] and want to calculate the Lebesgue integral $\int_{[0,1]} f(x)\, d\mu(x)$. The measure μ is Stieltjes generated by the differentiable and strictly monotonous function $g(x)$.

Due to the strict monotonicity the function value lies in the closed interval $[f(0), f(1)]$ which we divide into n subintervals. It makes sense to use the subintervals $\left[f\left(\frac{i}{n}\right), f\left(\frac{i+1}{n}\right)\right)$ with the index i running from 0 to $n-1$. We can determine the inverse image areas of these subintervals. Due to the strict monotonicity of f the inverse function exists and the following applies:

$$f^{-1}\left(\left[f\left(\frac{i}{n}\right), f\left(\frac{i+1}{n}\right)\right)\right) = \left[\frac{i}{n}, \frac{i+1}{n}\right).\tag{5.28}$$

Looking at the lower sums of the Lebesgue integral and letting n go to infinity we get

$$\int_{[0,1]} f(x)\, d\mu(x) = \lim_{n\to\infty} \sum_{i=0}^{n-1} f\left(\frac{i}{n}\right) \mu\left(\left[\frac{i}{n}, \frac{i+1}{n}\right)\right).\tag{5.29}$$

The Stieltjes measure of this interval is determined by Eq. (3.38).[13] Therefore we get

$$\int_{[0,1]} f(x)\, d\mu(x) = \lim_{n\to\infty} \sum_{i=0}^{n-1} f\left(\frac{i}{n}\right) \left(g\left(\frac{i+1}{n}\right) - g\left(\frac{i}{n}\right)\right).\tag{5.30}$$

We rewrite this equation in a slightly more complicated form which will turn out to be suitable in a moment

$$\int_{[0,1]} f(x)\, d\mu(x) = \lim_{n\to\infty} \sum_{i=0}^{n-1} f\left(\frac{i}{n}\right) \underbrace{\frac{g\left(\frac{i+1}{n}\right) - g\left(\frac{i}{n}\right)}{\frac{i+1}{n} - \frac{i}{n}}}_{=z} \left(\frac{i+1}{n} - \frac{i}{n}\right).\tag{5.31}$$

The term marked z for $n \to \infty$ corresponds to the first derivative of g which leads to

$$\int_{[0,1]} f(x)\, d\mu(x) = \lim_{n\to\infty} \sum_{i=0}^{n-1} f\left(\frac{i}{n}\right) g'\left(\frac{i}{n}\right) \left(\frac{i+1}{n} - \frac{i}{n}\right).\tag{5.32}$$

---

[12] The following remarks also apply to non-monotonous functions f. Then, however, the proofs are more complicated.

[13] See page 52.

The right expression is the classic Riemann integral $\int_0^1 f \cdot g'\, dx$.

    Therefore the following holds:

$$\underbrace{\int_{[0,1]} f(x)\, d\mu(x)}_{\text{Lebesgue}} = \underbrace{\int_0^1 f \cdot g'\, dx}_{\text{Riemann}}. \tag{5.33}$$

To summarize: the Lebesgue integral with Stieltjes measures is a generalization of a Riemann integral.

5.4   Result: Expectation and Variance as Lebesgue Integral

On the basis of the material presented in the previous sections we are able to define the expectation and the variance of a random variable Z—even if the state space does not correspond to the real numbers. The expectation and variance are Lebesgue integrals over the probability measure of the state space Ω. Specifically, the following applies

$$\mathrm{E}[Z] := \int_{\Omega} Z(\omega)\, d\mu(\omega), \tag{5.34}$$

$$\mathrm{Var}[Z] := \int_{\Omega} \big(Z(\omega) - \mathrm{E}[Z]\big)^2 d\mu(\omega). \tag{5.35}$$

Also, the following applies:

$$\mathrm{Var}[Z] = \int_{\Omega} \big(Z^2(\omega) - 2\,\mathrm{E}[Z]Z(\omega) + \mathrm{E}^2[Z]\big)\, d\mu(\omega)$$

$$= \int_{\Omega} Z^2(\omega)\, d\mu(\omega) - 2\,\mathrm{E}[Z] \overbrace{\int_{\Omega} Z(\omega)\, d\mu(\omega)}^{=\mathrm{E}[Z]} + \mathrm{E}^2[Z] \overbrace{\int_{\Omega} d\mu(\omega)}^{=\mu(\Omega)=1}$$

$$= \int_{\Omega} Z^2(\omega)\, d\mu(\omega) - \mathrm{E}^2[Z]. \tag{5.36}$$

This is known as the decomposition theorem of variance which could also be written more concisely as $\mathrm{Var}[Z] = \mathrm{E}[Z^2] - \mathrm{E}^2[Z]$.

5.5   Conditional Expectation

In the previous section we have shown the process of determining the expectation of a random variable using the Lebesgue integral. In doing so we have, however, ignored an aspect which plays a major role in financial problems. Analyzing an

investment decision requires the evaluation of future cash flows that will occur over a period of several years $t = 1, 2, \ldots$.

In particular we assume that the investment decision must be made today (in $t = 0$) and cannot or should not be revised. Given these starting conditions the future cash flows the decision-maker currently expects to occur in $t = 1, 2, \ldots$ must enter the evaluation process. The expected values of these cash flows are called "classical" or "unconditional expectations."

Let us now change the perspective of the decision-maker: the decision-maker is interested from the very beginning in a flexible investment plan, i.e., he is considering also possible modifications of the original investment decision. For example, this could include that the decision-maker can build in $t = 0$ either a larger or a smaller production facility. At $t = 1$ there should also be the possibility to expand a smaller factory or to abandon the investment.

Once $t = 1$ has occurred the decision-maker will have newer and different information about the probabilities of future cash flows than he had in $t = 0$. Today he can only decide on the basis of the information available at $t = 0$. Thus, the decision-maker can at best consider those future cash flows which he believes in $t = 0$ will be realized in later periods given that certain conditions will take place. From today's perspective the future cash flow developments in $t \geq 1$ could either be influenced by a boom or a bust. Such state-dependent expectations are called "conditional expectations." It is therefore very important to distinguish between unconditional and conditional expectations and being aware of their implications.

Conditional Expectations Regarding an Event Let us clarify what distinguishes a conditional expectation from an unconditional one. In general, the expectation of a random variable is the weighted average of all possible states with the weights representing their probabilities. The expectation describes something like the average result of a random variable. Of course, you need certain information about future events to be able to calculate expectations. Therefore, we must look more closely at this information.

The information a decision-maker has available can be described in more detail using the σ-algebra $\mathcal{F}$ as shown in Sect. 3.2. Given the measure space $(\Omega, \mathcal{F}, \mu)$ we know the event space Ω, the measurable set $\mathcal{F}$, and the probability measure μ. Considering subset $A \subset \Omega$ we assume that this subset is measurable ($A \in \mathcal{F}$). In other words, it can be determined whether a specific event does or does not belong to A. One should be interested in how large the expectation of all events is, if one *limits oneself to elements of A*. This implies that only events from A are included in the calculation of the expectation and that the relevant probabilities have to be normalized such that they sum to one.

This concept can be illustrated particularly well with the help of a binomial model. For this purpose we use again Fig. 2.5 from page 25 but add specific numerical values.

Fig. 5.6 Binomial model
with cash flows *CF* to $t = 3$
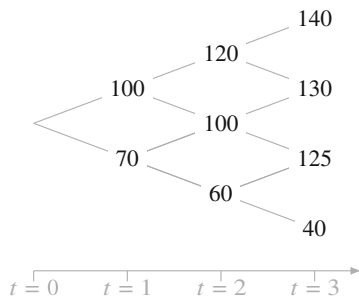
Example 5.4 (Binomial Model) Figure 5.6 shows a binomial model with three
points in time which describes future cash flows. Further, the upward and downward
movements are equally probable.

The didactic advantage of the binomial model is that questions of how individual
events can be measured do not complicate the presentation of the relevant problem.
In this example any set of events can be measured. Let us focus on the node of 125
at $t = 3$.

There exist three possible states ending in this node. These states are the
elementary events *udd*, *ddu*, and *dud*. If we want to determine the expectation
of the cash flows at time $t = 2$ assuming that the payment 125 is reached in $t = 3$,
this must be done as follows:

Starting from the condition $CF_3 = 125$, only the three events *udd*, *ddu*, and
dud are possible. These events are equally likely; their—normalized—conditional
probabilities are therefore $\frac{1}{3}$. This results in the conditional expectation of

$$\mathrm{E}[CF_2|CF_3 = 125] = \underbrace{\frac{1}{3} \cdot 100}_{udd} + \underbrace{\frac{1}{3} \cdot 100}_{dud} + \underbrace{\frac{1}{3} \cdot 60}_{ddu} \approx 86.67.$$

The following example can possibly make the approach clearer.

Example 5.5 (Dice Roll) Consider the example of a dice roll, with event *A* being
an even score, i.e.,

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

With an ideal dice every score can happen with the probability $\frac{1}{6}$. Restricting our
consideration to the even scores three cases are possible. The conditional probability
of each even score is $\frac{1}{3}$. The expectation conditional on *A* is the sum of the products
of the (even) scores with their conditional probabilities. Formally:

$$\mathrm{E}[X|A] = \frac{1}{3} \cdot 2 + \frac{1}{3} \cdot 4 + \frac{1}{3} \cdot 6 = 4.$$

The conditional expectation of an odd score being rolled, that is for the event

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

is 3.

Using a so-called indicator function,[14] the above results can be generalized. The expectation conditional on the measurable subset A can be expressed as

$$E[X|A] = \frac{\int_{\Omega} X \cdot 1_A d\mu(x)}{\mu(A)}. \tag{5.37}$$

Equation (5.37) makes sense only for events with a positive probability.

Conditional Expectation Regarding a σ-Algebra   Let us expand our analysis and investigate an expectation not only regarding a single event A but also regarding a whole σ-algebra. Initially, we will explain what this means in mathematical terms.

We have stated earlier that a σ-algebra can be thought of as an information system.[15] The elements of the σ-algebra describe those events which a decision-maker can observe or verify. We know how to determine the expectation of a quantity X in relation to a single observable event $A \in \mathcal{F}$: it is the conditional expectation $E[X|A]$. If at $t = 0$ a decision-maker reflects on the future he cannot restrict himself to only one elementary event of set A. Indeed he will include the fact that also the complement of A can take place. Given the possible events of the σ-algebra the decision-maker should obviously try to obtain an overview of all possible expectations of X. Thus, he calculates not only a *single* conditional expectation but also *all* conditional expectations for the conceivable sets $A \in \mathcal{F}$ of the σ-algebra. Let us illustrate this aspect in the context of a binomial model.

Example 5.6 (Binomial Model)   Referring to Fig. 5.6 on page 82 we focus on time $t = 2$. We have shown previously (page 43) how to describe the information which a decision-maker today believes he will have available at time $t = 2$. It's about the σ-algebra $\mathcal{F}_2$ which can be generated from the elements of set

$$A = \{\{uuu, uud\}, \{udu, udd\}, \{duu, dud\}, \{ddu, ddd\}\}.$$

---

[14]This function is one on A and zero otherwise, so

$$1_A(x) := \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$

[15]See page 42.

The set A contains only elementary events which can no longer be discriminated at time $t = 2$. Let us concentrate on all cash flows CF_3 given in Fig. 5.6 for time $t = 3$ and try to determine their expectations on the basis of the information the decision-maker believes at $t = 0$ he will have available at $t = 2$. For this purpose, we decompose the set A into the pairwise disjoint subsets[16]

$$A_1 = \{uuu, uud\}$$
$$A_2 = \{\{udu, udd\}, \{duu, dud\}\}$$
$$A_3 = \{ddu, ddd\}$$

with

$$A_1 \cup A_2 \cup A_3 = A.$$

Since A is *only* the set that generates the σ-algebra $\mathcal{F}_2$, $A \subset \mathcal{F}_2$ applies. It is easy to see that the expected cash flows CF_3 depend on which of the three subsets is considered. Considering the subset A_1 only the cash flows 140 and 130 associated with the elementary events uuu and uud can materialize at time $t = 3$. Their expectation is $\frac{1}{2} \cdot 140 + \frac{1}{2} \cdot 130 = 135$. Correspondingly for subset A_2 only the cash flows 130 and 125 can occur and their expectation equals $\frac{1}{2} \cdot 130 + \frac{1}{2} \cdot 125 = 127.5$. Similarly, for subset A_3 the cashflows 125 and 40 matter and lead to $\frac{1}{2} \cdot 125 + \frac{1}{2} \cdot 40 = 82.5$. Summarizing we have

$$\mathrm{E}\,[CF_3|A] = \begin{cases} 135.0, & \text{if } A = A_1, \\ 127.5, & \text{if } A = A_2, \\ 82.5, & \text{if } A = A_3. \end{cases}$$

Emphasizing the information that the decision-maker will have available at $t = 2$, the expectation can also be written in a somewhat more casual form

$$\mathrm{E}\,[CF_3|\mathcal{F}_2] = \begin{cases} 135.0, & \text{if at } t = 2 \quad uu, \\ 127.5, & \text{if at } t = 2 \quad ud \text{ or } du, \\ 82.5, & \text{if at } t = 2 \quad dd. \end{cases} \qquad (5.38)$$

Note that on the right side of Eq. (5.38) only the three possible nodes at time $t = 2$ are mentioned, while on the left side the σ-algebra $\mathcal{F}_2$ is used which includes more information than only set A. The notation of this equation does not precisely match and is therefore a bit more casual.

---

[16]Given the subsets are not empty such a segmentation is called partition.

The above example deserves two comments:

1. The conditional expectation is not just a number.[17] Rather, there exist several values because for each event A a state-dependent expectation must be calculated. While the classical expectation is written as $E[X]$, the notation for the conditional expectation

$$E[X|\mathcal{F}]$$

   highlights this difference.
2. While our example deals with only few events in generating the σ-algebra, the idea can also be implemented with large algebras.[18]

---

[17]In any case this usually applies.

[18]Since some of the relevant sets may have disappearing probabilities $\mu(\cdot) = 0$, the application of Eq. (5.37) is not permissible.