

Lebesgue Integral and Mathematical Expectation

3.1 Definition of the Lebesgue Integral

In this section we revisit the familiar notion of mathematical expectation, but now we define it for general (not necessarily discrete) random variables. The notion of expectation is identical to the notion of the Lebesgue integral.

Let $(\Omega, \mathcal{F}, \mu)$ be a measurable space with a finite measure. A measurable function is said to be simple if it takes a finite or countable number of values. The sum, product and quotient (when the denominator does not take the value zero) of two simple functions is again a simple function.

Theorem 3.1. *Any non-negative measurable function f is a monotone limit from below of non-negative simple functions, that is $f(\omega) = \lim_{n \rightarrow \infty} f_n(\omega)$ for every ω , where f_n are non-negative simple functions and $f_n(\omega) \leq f_{n+1}(\omega)$ for every ω . Moreover, if a function f is a limit of measurable functions for all ω , then f is measurable.*

Proof. Let f_n be defined by the relations

$$f_n(\omega) = k2^{-n} \quad \text{if} \quad k2^{-n} \leq f(\omega) < (k+1)2^{-n}, \quad k = 0, 1, \dots$$

The sequence f_n satisfies the requirements of the theorem.

We now prove the second statement. Given a function f which is the limit of measurable functions f_n , consider the subsets $A \subseteq \mathbb{R}$ for which $f^{-1}(A) \in \mathcal{F}$. It is easy to see that these subsets form a σ -algebra which we shall denote by \mathcal{R}_f . Let us prove that open intervals $A_t = (-\infty, t)$ belong to \mathcal{R}_f . Indeed it is easy to check the following relation

$$f^{-1}(A_t) = \bigcup_k \bigcup_m \bigcap_{n \geq m} \left\{ \omega : f_n(\omega) < t - \frac{1}{k} \right\}.$$

Since f_n are measurable, the sets $\{\omega : f_n(\omega) < t - \frac{1}{k}\}$ belong to \mathcal{F} , and therefore $f^{-1}(A_t) \in \mathcal{F}$. Since the smallest σ -algebra which contains all A_t is the

Borel σ -algebra on \mathbb{R} , $f^{-1}(A) \in \mathcal{F}$ for any Borel set A of the real line. This completes the proof of the theorem. \square

We now introduce the Lebesgue integral of a measurable function. When f is measurable and the measure is a probability measure, we refer to the integral as the expectation of the random variable, and denote it by Ef .

We start with the case of a simple function. Let f be a simple function taking non-negative values, which we denote by a_1, a_2, \dots . Let us define the events $C_i = \{\omega : f(\omega) = a_i\}$.

Definition 3.2. *The sum of the series $\sum_{i=1}^{\infty} a_i \mu(C_i)$, provided that the series converges, is called the Lebesgue integral of the function f . It is denoted by $\int_{\Omega} f d\mu$. If the series diverges, then it is said that the integral is equal to plus infinity.*

It is clear that the sum of the series does not depend on the order of summation. The following lemma is clear.

Lemma 3.3. *The integral of a simple non-negative function has the following properties.*

1. $\int_{\Omega} f d\mu \geq 0$.
2. $\int_{\Omega} \chi_{\Omega} d\mu = \mu(\Omega)$, where χ_{Ω} is the function identically equal to 1 on Ω .
3. $\int_{\Omega} (af_1 + bf_2) d\mu = a \int_{\Omega} f_1 d\mu + b \int_{\Omega} f_2 d\mu$ for any $a, b > 0$.
4. $\int_{\Omega} f_1 d\mu \geq \int_{\Omega} f_2 d\mu$ if $f_1 \geq f_2 \geq 0$.

Now let f be an arbitrary measurable function taking non-negative values. We consider the sequence f_n of non-negative simple functions which converge monotonically to f from below. It follows from the fourth property of the Lebesgue integral that the sequence $\int_{\Omega} f_n d\mu$ is non-decreasing and there exists a limit $\lim_{n \rightarrow \infty} \int_{\Omega} f_n d\mu$, which is possibly infinite.

Theorem 3.4. *Let f and f_n be as above. Then the value of $\lim_{n \rightarrow \infty} \int_{\Omega} f_n d\mu$ does not depend on the choice of the approximating sequence.*

We first establish the following lemma.

Lemma 3.5. *Let $g \geq 0$ be a simple function such that $g \leq f$. Assume that $f = \lim_{n \rightarrow \infty} f_n$, where f_n are non-negative simple functions such that $f_{n+1} \geq f_n$. Then $\int_{\Omega} g d\mu \leq \lim_{n \rightarrow \infty} \int_{\Omega} f_n d\mu$.*

Proof. Take an arbitrary $\varepsilon > 0$ and set $C_n = \{\omega : f_n(\omega) - g(\omega) > -\varepsilon\}$. It follows from the monotonicity of f_n that $C_n \subseteq C_{n+1}$. Since $f_n \uparrow f$ and $f \geq g$, we have $\bigcup_n C_n = \Omega$. Therefore, $\mu(C_n) \rightarrow \mu(\Omega)$ as $n \rightarrow \infty$. Let χ_{C_n} be the indicator function of the set C_n . Then $g_n = g\chi_{C_n}$ is a simple function and $g_n \leq f_n + \varepsilon$. Therefore, by the monotonicity of $\int_{\Omega} f_n d\mu$,

$$\int_{\Omega} g_n d\mu \leq \int_{\Omega} f_n d\mu + \varepsilon, \quad \int_{\Omega} g_n d\mu \leq \lim_{m \rightarrow \infty} \int_{\Omega} f_m d\mu + \varepsilon.$$

Since ε is arbitrary, we obtain $\int_{\Omega} g_n d\mu \leq \lim_{m \rightarrow \infty} \int_{\Omega} f_m d\mu$. It remains to prove that $\lim_{n \rightarrow \infty} \int_{\Omega} g_n d\mu = \int_{\Omega} g d\mu$.

We denote by b_1, b_2, \dots the values of the function g , and by B_i the set where the value b_i is taken, $i = 1, 2, \dots$. Then

$$\int_{\Omega} g d\mu = \sum_i b_i \mu(B_i), \quad \int_{\Omega} g_n d\mu = \sum_i b_i \mu(B_i \cap C_n).$$

It is clear that for all i we have $\lim_{n \rightarrow \infty} \mu(B_i \cap C_n) = \mu(B_i)$. Since the series above consists of non-negative terms and the convergence is monotonic for each i , we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \int_{\Omega} g_n d\mu &= \lim_{n \rightarrow \infty} \sum_i b_i \mu(B_i \cap C_n) \\ &= \sum_i b_i \lim_{n \rightarrow \infty} \mu(B_i \cap C_n) = \sum_i b_i \mu(B_i) = \int_{\Omega} g d\mu. \end{aligned}$$

This completes the proof of the lemma. □

It is now easy to prove the independence of $\lim_{n \rightarrow \infty} \int_{\Omega} f_n d\mu$ from the choice of the approximating sequence.

Proof of Theorem 3.4. Let there be two sequences $f_n^{(1)}$ and $f_n^{(2)}$ such that $f_{n+1}^{(1)} \geq f_n^{(1)}$ and $f_{n+1}^{(2)} \geq f_n^{(2)}$ for all n , and

$$\lim_{n \rightarrow \infty} f_n^{(1)}(\omega) = \lim_{n \rightarrow \infty} f_n^{(2)}(\omega) = f(\omega) \quad \text{for every } \omega.$$

It follows from Lemma 3.5 that for any k ,

$$\int_{\Omega} f_k^{(1)} d\mu \leq \lim_{n \rightarrow \infty} \int_{\Omega} f_n^{(2)} d\mu,$$

and therefore,

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n^{(1)} d\mu \leq \lim_{n \rightarrow \infty} \int_{\Omega} f_n^{(2)} d\mu.$$

We obtain

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n^{(1)} d\mu \geq \lim_{n \rightarrow \infty} \int_{\Omega} f_n^{(2)} d\mu$$

by interchanging $f_n^{(1)}$ and $f_n^{(2)}$. Therefore,

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n^{(1)} d\mu = \lim_{n \rightarrow \infty} \int_{\Omega} f_n^{(2)} d\mu.$$

□

Definition 3.6. Let f be a non-negative measurable function and f_n a sequence of non-negative simple functions which converge monotonically to f from below. The limit $\lim_{n \rightarrow \infty} \int_{\Omega} f_n d\mu$ is called the Lebesgue integral of the function f . It is denoted by $\int_{\Omega} f d\mu$.

In the case of a simple function f , this definition agrees with the definition of the integral for a simple function, since we can take $f_n = f$ for all n .

Now let f be an arbitrary (not necessarily positive) measurable function. We introduce the indicator functions:

$$\chi_+(\omega) = \begin{cases} 1 & \text{if } f(\omega) \geq 0, \\ 0 & \text{if } f(\omega) < 0, \end{cases}$$

$$\chi_-(\omega) = \begin{cases} 1 & \text{if } f(\omega) < 0, \\ 0 & \text{if } f(\omega) \geq 0. \end{cases}$$

Then $\chi_+(\omega) + \chi_-(\omega) \equiv 1$, $f = f\chi_+ + f\chi_- = f_+ - f_-$, where $f_+ = f\chi_+$ and $f_- = -f\chi_-$. Moreover, $f_+ \geq 0$, $f_- \geq 0$, so the integrals $\int_{\Omega} f_+ d\mu$ and $\int_{\Omega} f_- d\mu$ have already been defined.

Definition 3.7. The function f is said to be integrable if $\int_{\Omega} f_+ d\mu < \infty$ and $\int_{\Omega} f_- d\mu < \infty$. In this case the integral is equal to $\int_{\Omega} f d\mu = \int_{\Omega} f_+ d\mu - \int_{\Omega} f_- d\mu$. If $\int_{\Omega} f_+ d\mu = \infty$ and $\int_{\Omega} f_- d\mu < \infty$ ($\int_{\Omega} f_+ d\mu < \infty$, $\int_{\Omega} f_- d\mu = \infty$), then $\int_{\Omega} f d\mu = \infty$ ($\int_{\Omega} f d\mu = -\infty$). If $\int_{\Omega} f_+ d\mu = \int_{\Omega} f_- d\mu = \infty$, then $\int_{\Omega} f d\mu$ is not defined.

Since $|f| = f_+ + f_-$, we have $\int_{\Omega} |f| d\mu = \int_{\Omega} f_+ d\mu + \int_{\Omega} f_- d\mu$, and so $\int_{\Omega} f d\mu$ is finite if and only if $\int_{\Omega} |f| d\mu$ is finite. The integral has Properties 2–4 listed in Lemma 3.3.

Let $A \in \mathcal{F}$ be a measurable set and f a measurable function on $(\Omega, \mathcal{F}, \mu)$. We can define the integral of f over the set A (which is a subset of Ω) in two equivalent ways. One way is to define

$$\int_A f d\mu = \int_{\Omega} f \chi_A d\mu,$$

where χ_A is the indicator function of the set A . Another way is to consider the restriction of μ from Ω to A . Namely, we consider the new σ -algebra \mathcal{F}_A , which contains all the measurable subsets of A , and the new measure μ_A on \mathcal{F}_A , which agrees with μ on all the sets from \mathcal{F}_A . Then (A, \mathcal{F}_A) is a measurable space with a measure μ_A , and we can define

$$\int_A f d\mu = \int_A f d\mu_A.$$

It can easily be seen that the above two definitions lead to the same notion of the integral over a measurable set.

Let us note another important property of the Lebesgue integral: it is a σ -additive function on \mathcal{F} . Namely, let $A = \bigcup_{i=1}^{\infty} A_i$, where A_1, A_2, \dots are

measurable sets such that $A_i \cap A_j = \emptyset$ for $i \neq j$. Let f be a measurable function such that $\int_A f d\mu$ is finite. Then

$$\int_A f d\mu = \sum_{i=1}^{\infty} \int_{A_i} f d\mu.$$

To justify this statement we can first consider f to be a non-negative simple function. Then the σ -additivity follows from the fact that in an infinite series with non-negative terms the terms can be re-arranged. For an arbitrary non-negative measurable f we use the definition of the integral as a limit of integrals of simple functions. For f which is not necessarily non-negative, we use Definition 3.7.

If f is a non-negative function, the σ -additivity of the integral implies that the function $\eta(A) = \int_A f d\mu$ is itself a measure.

The mathematical expectation (which is the same as the Lebesgue integral over a probability space) has all the properties described in Chap. 1. In particular

1. $E\xi \geq 0$ if $\xi \geq 0$.
2. $E\chi_{\Omega} = 1$ where χ_{Ω} is the random variable identically equal to 1 on Ω .
3. $E(a\xi_1 + b\xi_2) = aE\xi_1 + bE\xi_2$ if $E\xi_1$ and $E\xi_2$ are finite.

The variance of the random variable ξ is defined as $E(\xi - E\xi)^2$, and the n -th order moment is defined as $E\xi^n$. Given two random variables ξ_1 and ξ_2 , their covariance is defined as $\text{Cov}(\xi_1, \xi_2) = E(\xi_1 - E\xi_1)(\xi_2 - E\xi_2)$. The correlation coefficient of two random variables ξ_1, ξ_2 is defined as $\rho(\xi_1, \xi_2) = \text{Cov}(\xi_1, \xi_2) / \sqrt{\text{Var}\xi_1 \text{Var}\xi_2}$.

3.2 Induced Measures and Distribution Functions

Given a probability space (Ω, \mathcal{F}, P) , a measurable space $(\tilde{\Omega}, \tilde{\mathcal{F}})$ and a measurable function $f : \Omega \rightarrow \tilde{\Omega}$, we can define the induced probability measure \tilde{P} on the σ -algebra $\tilde{\mathcal{F}}$ via the formula

$$\tilde{P}(A) = P(f^{-1}(A)) \quad \text{for } A \in \tilde{\mathcal{F}}.$$

Clearly $\tilde{P}(A)$ satisfies the definition of a probability measure. The following theorem states that the change of variable is permitted in the Lebesgue integral.

Theorem 3.8. *Let $g : \tilde{\Omega} \rightarrow \mathbb{R}$ be a random variable. Then*

$$\int_{\Omega} g(f(\omega)) dP(\omega) = \int_{\tilde{\Omega}} g(\tilde{\omega}) d\tilde{P}(\tilde{\omega}).$$

The integral on the right-hand side is defined if and only if the integral on the left-hand side is defined.

Proof. Without loss of generality we can assume that g is non-negative. When g is a simple function, the theorem follows from the definition of the induced measure. For an arbitrary measurable function it suffices to note that any such function is a limit of a non-decreasing sequence of simple functions. \square

Let us examine once again the relationship between the random variables and their distribution functions. Consider the collection of all intervals:

$$\mathcal{I} = \{(a, b), [a, b), (a, b], [a, b], \text{ where } -\infty \leq a \leq b \leq \infty\}.$$

Let $m : \mathcal{I} \rightarrow \mathbb{R}$ be a σ -additive nonnegative function, that is

1. $m(I) \geq 0$ for any $I \in \mathcal{I}$.
2. If $I, I_i \in \mathcal{I}$, $i = 1, 2, \dots$, $I_i \cap I_j = \emptyset$ for $i \neq j$, and $I = \bigcup_{i=1}^{\infty} I_i$, then

$$m(I) = \sum_{i=1}^{\infty} m(I_i).$$

Although m is σ -additive, as required of a measure, it is not truly a measure since it is defined on the collection of intervals, which is not a σ -algebra.

We shall need the following theorem (a particular case of the theorem on the extension of a measure discussed in Sect. 3.4).

Theorem 3.9. *Let m be a σ -additive function satisfying conditions 1 and 2. Then there is a unique measure μ defined on the σ -algebra of Borel sets of the real line, which agrees with m on all the intervals, that is $\mu(I) = m(I)$ for each $I \in \mathcal{I}$.*

Consider the following three examples, which illustrate how a measure can be constructed given its values on the intervals.

Example. Let $F(x)$ be a distribution function. We define

$$m((a, b]) = F(b) - F(a), \quad m([a, b]) = F(b) - \lim_{t \uparrow a} F(t),$$

$$m((a, b)) = \lim_{t \uparrow b} F(t) - F(a), \quad m([a, b)) = \lim_{t \uparrow b} F(t) - \lim_{t \uparrow a} F(t).$$

Let us check that m is a σ -additive function. Let I, I_i , $i = 1, 2, \dots$ be intervals of the real line (open, half-open, or closed) such that $I = \bigcup_{i=1}^{\infty} I_i$ and $I_i \cap I_j = \emptyset$ if $i \neq j$. We need to check that

$$m(I) = \sum_{i=1}^{\infty} m(I_i). \tag{3.1}$$

It is clear that $m(I) \geq \sum_{i=1}^n m(I_i)$ for each n , since the intervals I_i do not intersect. Therefore, $m(I) \geq \sum_{i=1}^{\infty} m(I_i)$.

In order to prove the opposite inequality, we assume that an arbitrary $\varepsilon > 0$ is given. Consider a collection of intervals $J, J_i, i = 1, 2, \dots$ which are constructed as follows. The interval J is a closed interval, which is contained in I and satisfies $m(J) \geq m(I) - \varepsilon/2$. (In particular, if I is closed we can take $J = I$). Let J_i be an open interval, which contains I_i and satisfies $m(J_i) \leq m(I_i) + \varepsilon/2^{i+1}$. The fact that it is possible to select such intervals J and J_i follows from the definition of the function m and the continuity from the right of the function F . Note that $J \subseteq \bigcup_{i=1}^{\infty} J_i$, J is compact, and all J_i are open. Therefore, $J \subseteq \bigcup_{i=1}^n J_i$ for some n . Clearly $m(J) \leq \sum_{i=1}^n m(J_i)$. Therefore, $m(I) \leq \sum_{i=1}^n m(I_i) + \varepsilon$. Since ε is arbitrary, we obtain $m(I) \leq \sum_{i=1}^{\infty} m(I_i)$. Therefore, (3.1) holds, and m is a σ -additive function.

Thus any distribution function gives rise to a probability measure on the Borel σ -algebra of the real line. This measure will be denoted by μ_F . Sometimes, instead of writing $d\mu_F$ in the integral with respect to such a measure, we shall write dF .

Conversely, any probability measure μ on the Borel sets of the real line defines a distribution function via the formula $F(x) = \mu((-\infty, x])$. Thus there is a one-to-one correspondence between probability measures on the real line and distribution functions.

Remark 3.10. Similarly, there is a one-to-one correspondence between the distribution functions on \mathbb{R}^n and the probability measures on the Borel sets of \mathbb{R}^n . Namely, the distribution function F corresponding to a measure μ is defined by $F(x_1, \dots, x_n) = \mu((-\infty, x_1] \times \dots \times (-\infty, x_n])$.

Example. Let f be a function defined on an interval $[a, b]$ of the real line. Let $\sigma = \{t_0, t_1, \dots, t_n\}$, with $a = t_0 \leq t_1 \leq \dots \leq t_n = b$, be a partition of the interval $[a, b]$ into n subintervals. We denote the length of the largest interval by $\delta(\sigma) = \max_{1 \leq i \leq n} (t_i - t_{i-1})$. The p -th variation (with $p > 0$) of f over the partition σ is defined as

$$V_{[a,b]}^p(f, \sigma) = \sum_{i=1}^n |f(t_i) - f(t_{i-1})|^p.$$

Definition 3.11. *The following limit*

$$V_{[a,b]}^p(f) = \limsup_{\delta(\sigma) \rightarrow 0} V_{[a,b]}^p(f, \sigma),$$

is referred to as the p -th total variation of f over the interval $[a, b]$.

Now let f be a continuous function with finite first ($p = 1$) total variation defined on an interval $[a, b]$ of the real line. Then it can be represented as a difference of two continuous non-decreasing functions, namely,

$$f(x) = V_{[a,x]}^1(f) - (V_{[a,x]}^1(f) - f(x)) = F_1(x) - F_2(x).$$

Now we can repeat the construction used in the previous example to define the measures μ_{F_1} and μ_{F_2} on the Borel subsets of $[a, b]$. Namely, we can define

$$m_i((x, y]) = m_i([x, y]) = m_i((x, y)) = m_i([x, y)) = F_i(y) - F_i(x), \quad i = 1, 2,$$

and then extend m_i to the measure μ_{F_i} using Theorem 3.9. The difference $\mu_f = \mu_{F_1} - \mu_{F_2}$ is then a signed measure (see Sect. 3.6). If g is a Borel-measurable function on $[a, b]$, its integral with respect to the signed measure μ_f , denoted by $\int_a^b g(x)df(x)$ or $\int_a^b g(x)d\mu_f(x)$, is defined as the difference of the integrals with respect to the measures μ_{F_1} and μ_{F_2} ,

$$\int_a^b g(x)df(x) = \int_a^b g(x)d\mu_{F_1}(x) - \int_a^b g(x)d\mu_{F_2}(x).$$

It is called the Lebesgue-Stieltjes integral of g with respect to f .

Example. For an interval I , let $I_n = I \cap [-n, n]$. Define $m_n(I)$ as the length of I_n . As in the first example, m_n is a σ -additive function. Thus m_n gives rise to a measure on the Borel sets of the real line, which will be denoted by λ_n and referred to as the Lebesgue measure on the segment $[-n, n]$. Now for any Borel set A of the real line we can define its Lebesgue measure $\lambda(A)$ via $\lambda(A) = \lim_{n \rightarrow \infty} \lambda_n(A)$. It is easily checked that λ is a σ -additive measure which, however, may take infinite values for unbounded sets A .

Remark 3.12. The Lebesgue measure on the real line is an example of a σ -finite measure. We now give the formal definition of a σ -finite measure, although most of the measures that we deal with in this book are finite (probability) measures. An integral with respect to a σ -finite measure can be defined in the same way as an integral with respect to a finite measure.

Definition 3.13. Let (Ω, \mathcal{F}) be a measurable space. A σ -finite measure is a function μ , defined on \mathcal{F} with values in $[0, \infty]$, which satisfies the following conditions.

1. There is a sequence of measurable sets $\Omega_1 \subseteq \Omega_2 \subseteq \dots \subseteq \Omega$ such that $\mu(\Omega_i) < \infty$ for all i , and $\bigcup_{i=1}^{\infty} \Omega_i = \Omega$.
2. If $C_i \in \mathcal{F}$, $i = 1, 2, \dots$ and $C_i \cap C_j = \emptyset$ for $i \neq j$, then

$$\mu\left(\bigcup_{i=1}^{\infty} C_i\right) = \sum_{i=1}^{\infty} \mu(C_i).$$

If F_ξ is the distribution function of a random variable ξ , then the measure μ_{F_ξ} (also denoted by μ_ξ) coincides with the measure induced by the random variable ξ . Indeed, the values of the induced measure and of μ_ξ coincide on the intervals, and therefore on all the Borel sets due to the uniqueness part of Theorem 3.9.

Theorem 3.8 together with the fact that μ_ξ coincides with the induced measure imply the following.

Theorem 3.14. *Let ξ be a random variable and g be a Borel measurable function on \mathbb{R} . Then*

$$Eg(\xi) = \int_{-\infty}^{\infty} g(x)dF_{\xi}(x).$$

Applying this theorem to the functions $g(x) = x$, $g(x) = x^p$ and $g(x) = (x - E\xi)^2$, we obtain the following.

Corollary 3.15.

$$E\xi = \int_{-\infty}^{\infty} x dF_{\xi}(x), \quad E\xi^p = \int_{-\infty}^{\infty} x^p dF_{\xi}(x), \quad \text{Var}\xi = \int_{-\infty}^{\infty} (x - E\xi)^2 dF_{\xi}(x) .$$

3.3 Types of Measures and Distribution Functions

Let μ be a finite measure on the Borel σ -algebra of the real line. We distinguish three special types of measures.

- (a) *Discrete measure.* Assume that there exists a finite or countable set $A = \{a_1, a_2, \dots\}$ such that $\mu((-\infty, \infty)) = \mu(A)$, that is A is a set of full measure. In this case μ is called a measure of discrete type.
- (b) *Singular continuous measure.* Assume that the measure of any single point is zero, $\mu(a) = 0$ for any $a \in \mathbb{R}$, and there is a Borel set B of Lebesgue measure zero which is of full measure for the measure μ , that is $\lambda(B) = 0$ and $\mu((-\infty, \infty)) = \mu(B)$. In this case μ is called a singular continuous measure.
- (c) *Absolutely continuous measure.* Assume that for every set of Lebesgue measure zero the μ measure of that set is also zero, that is $\lambda(A) = 0$ implies $\mu(A) = 0$. In this case μ is called an absolutely continuous measure.

While any given measure does not necessarily belong to one of the three classes above, the following theorem states that it can be decomposed into three components, one of which is discrete, the second singular continuous, and the third absolutely continuous.

Theorem 3.16. *Given any finite measure μ on \mathbb{R} there exist measures μ_1, μ_2 and μ_3 , the first of which is discrete, the second singular continuous and the third absolutely continuous, such that for any Borel set C of the real line we have*

$$\mu(C) = \mu_1(C) + \mu_2(C) + \mu_3(C) .$$

Such measures μ_1, μ_2 and μ_3 are determined by the measure μ uniquely.

Proof. Let A_1 be the collection of points a such that $\mu(a) \geq 1$, let A_2 be the collection of points $a \in \mathbb{R} \setminus A_1$ such that $\mu(a) \geq \frac{1}{2}$, let A_3 be the collection of points $a \in \mathbb{R} \setminus (A_1 \cup A_2)$ such that $\mu(a) \geq \frac{1}{3}$, and so on. Since the measure is finite, each set A_n contains only finitely many elements. Therefore, $A = \bigcup_n A_n$ is countable. At the same time $\mu(b) = 0$ for any $b \notin A$. Let $\mu_1(C) = \mu(C \cap A)$.

We shall now construct the measure μ_2 and a set B of zero Lebesgue measure, but of full μ_2 measure. (Note that it may turn out that $\mu_2(B) = 0$, that is μ_2 is identically zero.) First we inductively construct sets B_n , $n \geq 1$, as follows. Take B_1 to be an empty set. Assuming that B_n has been constructed, we take B_{n+1} to be any set of Lebesgue measure zero, which does not intersect $\bigcup_{i=1}^n B_i$ and satisfies

$$\mu(B_{n+1}) - \mu_1(B_{n+1}) \geq \frac{1}{m} \quad (3.2)$$

with the smallest possible m , where $m \geq 1$ is an integer. If no such m exists, then we take B_{n+1} to be the empty set. For each m there is at most a finite number of non-intersecting sets which satisfy (3.2), and therefore the set $\mathbb{R} \setminus \bigcup_{n=1}^{\infty} B_n$ contains no set C for which $\mu(C) - \mu_1(C) > 0$. We put $B = \bigcup_{n=1}^{\infty} B_n$, which is a set of Lebesgue measure zero, and define $\mu_2(C) = \mu(C \cap B) - \mu_1(C \cap B)$. Note that $\mu_2(B) = \mu_2((-\infty, \infty))$, and therefore μ_2 is singular continuous.

By the construction of μ_1 and μ_2 , we have that $\mu_3(C) = \mu(C) - \mu_1(C) - \mu_2(C)$ is a measure which is equal to zero on each set of Lebesgue measure zero. Thus we have the desired decomposition. The uniqueness part is left as an easy exercise for the reader. \square

Since there is a one-to-one correspondence between probability measures on the real line and distribution functions, we can single out the classes of distribution functions corresponding to the discrete, singular continuous and absolutely continuous measures. In the discrete case $F(x) = \mu((-\infty, x])$ is a step function. The jumps occur at the points a_i of positive μ -measure.

If the distribution function F has a Lebesgue integrable density p , that is $F(x) = \int_{-\infty}^x p(t)dt$, then F corresponds to an absolutely continuous measure. Indeed, $\mu_F(A) = \int_A p(t)dt$ for any Borel set A , since the equality is true for all intervals, and therefore it is true for all Borel sets due to the uniqueness of the extension of the measure. The value of the integral $\int_A p(t)dt$ over any set of Lebesgue measure zero is equal to zero.

The converse is also true, i.e., any absolutely continuous measure has a Lebesgue integrable density function. This follows from the Radon-Nikodym theorem, which we shall state below.

If a measure μ does not contain a discrete component, then the distribution function is continuous. Yet if the singular continuous component is present, it cannot be represented as an integral of a density. The so-called Cantor Staircase is an example of such a distribution function. Set $F(t) = 0$ for $t \leq 0$, $F(t) = 1$ for $t \geq 1$. We construct $F(t)$ for $0 < t < 1$ inductively. At the n -th step ($n \geq 0$) we have disjoint intervals of length 3^{-n} , where the function $F(t)$ is not yet defined, although it is defined at the end points of such intervals. Let us divide every such interval into three equal parts, and set $F(t)$ on the middle interval (including the end-points) to be a constant equal to the half-sum of its

values at the above-mentioned end-points. It is easy to see that the function $F(t)$ can be extended by continuity to the remaining t . The limit function is called the Cantor Staircase. It corresponds to a singular continuous probability measure. The theory of fractals is related to some classes of singular continuous measures.

3.4 Remarks on the Construction of the Lebesgue Measure

In this section we provide an abstract generalization of Theorem 3.9 on the extension of a σ -additive function. Theorem 3.9 applies to the construction of a measure on the real line which, in the case of the Lebesgue measure, can be viewed as an extension of the notion of length of an interval. In fact we can define the notion of measure starting from a σ -additive function defined on a certain collection of subsets of an abstract set.

Definition 3.17. *A collection \mathcal{G} of subsets of Ω is called a semialgebra if it has the following three properties:*

1. $\Omega \in \mathcal{G}$.
2. If $C_1, C_2 \in \mathcal{G}$, then $C_1 \cap C_2 \in \mathcal{G}$.
3. If $C_1, C_2 \in \mathcal{G}$ and $C_2 \subseteq C_1$, then there exists a finite collection of disjoint sets $A_1, \dots, A_n \in \mathcal{G}$ such that $C_2 \cap A_i = \emptyset$ for $i = 1, \dots, n$ and $C_2 \cup A_1 \cup \dots \cup A_n = C_1$.

Definition 3.18. *A non-negative function with values in \mathbb{R} defined on a semialgebra \mathcal{G} is said to be σ -additive if it satisfies the following condition:*

If $C = \bigcup_{i=1}^{\infty} C_i$ with $C \in \mathcal{G}$, $C_i \in \mathcal{G}$, $i = 1, 2, \dots$, and $C_i \cap C_j = \emptyset$ for $i \neq j$, then

$$m(C) = \sum_{i=1}^{\infty} m(C_i) .$$

Theorem 3.19 (Caratheodory). *Let m be a σ -additive function defined on a semialgebra (Ω, \mathcal{G}) . Then there exists a measure μ defined on $(\Omega, \sigma(\mathcal{G}))$ such that $\mu(C) = m(C)$ for every $C \in \mathcal{G}$. The measure μ which has this property is unique.*

We shall only indicate a sequence of steps used in the proof of the theorem, without giving all the details. A more detailed exposition can be found in the textbook of Fomin and Kolmogorov “Elements of Theory of Functions and Functional Analysis”.

Step 1. Extension of the σ -additive function from the semialgebra to the algebra. Let \mathcal{A} be the collection of sets which can be obtained as finite unions of disjoint elements of \mathcal{G} , that is $A \in \mathcal{A}$ if $A = \bigcup_{i=1}^n C_i$ for some $C_i \in \mathcal{G}$,

where $C_i \cap C_j = \emptyset$ if $i \neq j$. The collection of sets \mathcal{A} is an algebra since it contains the set Ω and is closed under finite unions, intersections, differences, and symmetric differences. For $A = \bigcup_{i=1}^n C_i$ with $C_i \cap C_j = \emptyset$, $i \neq j$, we define $m(A) = \sum_{i=1}^n m(C_i)$. We can then show that m is still a σ -additive function on the algebra \mathcal{A} .

Step 2. Definition of exterior measure and of measurable sets. For any set $B \subseteq \Omega$ we can define its exterior measure as $\mu^*(B) = \inf \sum_i m(A_i)$, where the infimum is taken over all countable coverings of B by elements of the algebra \mathcal{A} . A set B is called measurable if for any $\varepsilon > 0$ there is $A \in \mathcal{A}$ such that $\mu^*(A \Delta B) \leq \varepsilon$. Recall that $A \Delta B$ is the notation for the symmetric difference of the sets A and B . If B is measurable we define its measure to be equal to the exterior measure: $\mu(B) = \mu^*(B)$. Denote the collection of all measurable sets by \mathcal{B} .

Step 3. The σ -algebra of measurable sets and σ -additivity of the measure. The main part of the proof consists of demonstrating that \mathcal{B} is a σ -algebra, and that the function μ defined on it has the properties of a measure. We can then restrict the measure to the smallest σ -algebra containing the original semialgebra. The uniqueness of the measure follows easily from the non-negativity of m and from the fact that the measure is uniquely defined on the algebra \mathcal{A} . Alternatively, see Lemma 4.14 in Chap. 4, which also implies the uniqueness of the measure.

Remark 3.20. It is often convenient to consider the measure μ on the measurable space (Ω, \mathcal{B}) , rather than to restrict the measure to the σ -algebra $\sigma(\mathcal{G})$, which is usually smaller than \mathcal{B} . The difference is that (Ω, \mathcal{B}) is always complete with respect to measure μ , while $(\Omega, \sigma(\mathcal{G}))$ does not need to be complete. We discuss the notion of completeness in the remainder of this section.

Definition 3.21. Let (Ω, \mathcal{F}) be a measurable space with a finite measure μ on it. A set $A \subseteq \Omega$ is said to be μ -negligible if there is an event $B \in \mathcal{F}$ such that $A \subseteq B$ and $\mu(B) = 0$. The space (Ω, \mathcal{F}) is said to be complete with respect to μ if all μ -negligible sets belong to \mathcal{F} .

Given an arbitrary measurable space (Ω, \mathcal{F}) with a finite measure μ on it, we can consider an extended σ -algebra $\tilde{\mathcal{F}}$. It consists of all sets $\tilde{B} \subseteq \Omega$ which can be represented as $\tilde{B} = A \cup B$, where A is a μ -negligible set and $B \in \mathcal{F}$. We define $\tilde{\mu}(\tilde{B}) = \mu(B)$. It is easy to see that $\tilde{\mu}(\tilde{B})$ does not depend on the particular representation of \tilde{B} , $(\Omega, \tilde{\mathcal{F}})$ is a measurable space, $\tilde{\mu}$ is a finite measure, and $(\Omega, \tilde{\mathcal{F}})$ is complete with respect to $\tilde{\mu}$. We shall refer to $(\Omega, \tilde{\mathcal{F}})$ as the completion of (Ω, \mathcal{F}) with respect to the measure μ .

It is not difficult to see that $\tilde{\mathcal{F}} = \sigma(\mathcal{F} \cup \mathcal{N}^\mu)$, where \mathcal{N}^μ is the collection of μ -negligible sets in Ω .

3.5 Convergence of Functions, Their Integrals, and the Fubini Theorem

Let $(\Omega, \mathcal{F}, \mu)$ be a measurable space with a finite measure. Let f and f_n , $n = 1, 2, \dots$ be measurable functions.

Definition 3.22. A sequence of functions f_n is said to converge to f uniformly if

$$\lim_{n \rightarrow \infty} \sup_{\omega \in \Omega} |f_n(\omega) - f(\omega)| = 0.$$

Definition 3.23. A sequence of functions f_n is said to converge to f in measure (or in probability, if μ is a probability measure) if for any $\delta > 0$ we have

$$\lim_{n \rightarrow \infty} \mu(\omega : |f_n(\omega) - f(\omega)| > \delta) = 0.$$

Definition 3.24. A sequence of functions f_n is said to converge to f almost everywhere (or almost surely) if there is a measurable set A with $\mu(\Omega \setminus A) = 0$ such that

$$\lim_{n \rightarrow \infty} f_n(\omega) = f(\omega) \quad \text{for } \omega \in A.$$

It is not difficult to demonstrate that convergence almost everywhere implies convergence in measure. The opposite implication is only true if we consider a certain subsequence of the original sequence f_n (see Problem 8). The following theorem relates the notions of convergence almost everywhere and uniform convergence.

Theorem 3.25 (Egorov Theorem). If a sequence of measurable functions f_n converges to a measurable function f almost everywhere, then for any $\delta > 0$ there exists a measurable set $\Omega_\delta \subseteq \Omega$ such that $\mu(\Omega_\delta) \geq \mu(\Omega) - \delta$ and f_n converges to f uniformly on Ω_δ .

Proof. Let $\delta > 0$ be fixed. Let

$$\Omega_n^m = \bigcap_{i \geq n} \left\{ \omega : |f_i(\omega) - f(\omega)| < \frac{1}{m} \right\}$$

and

$$\Omega^m = \bigcup_{n=1}^{\infty} \Omega_n^m.$$

Due to the continuity of the measure (Theorem 1.36), for every m there is $n_0(m)$ such that $\mu(\Omega^m \setminus \Omega_{n_0(m)}^m) < \delta/2^m$. We define $\Omega_\delta = \bigcap_{m=1}^{\infty} \Omega_{n_0(m)}^m$. We claim that Ω_δ satisfies the requirements of the theorem.

The uniform convergence follows from the fact that $|f_i(\omega) - f(\omega)| < 1/m$ for all $\omega \in \Omega_\delta$ if $i > n_0(m)$. In order to estimate the measure of Ω_δ , we note

that $f_n(\omega)$ does not converge to $f(\omega)$ if ω is outside of the set Ω^m for some m . Therefore, $\mu(\Omega \setminus \Omega^m) = 0$. This implies

$$\mu(\Omega \setminus \Omega_{n_0(m)}^m) = \mu(\Omega^m \setminus \Omega_{n_0(m)}^m) < \frac{\delta}{2^m}.$$

Therefore,

$$\mu(\Omega \setminus \Omega_\delta) = \mu\left(\bigcup_{m=1}^{\infty} (\Omega \setminus \Omega_{n_0(m)}^m)\right) \leq \sum_{m=1}^{\infty} \mu(\Omega \setminus \Omega_{n_0(m)}^m) < \sum_{m=1}^{\infty} \frac{\delta}{2^m} = \delta,$$

which completes the proof of the theorem. \square

The following theorem justifies passage to the limit under the sign of the integral.

Theorem 3.26 (Lebesgue Dominated Convergence Theorem). *If a sequence of measurable functions f_n converges to a measurable function f almost everywhere and*

$$|f_n| \leq \varphi,$$

where φ is integrable on Ω , then the function f is integrable on Ω and

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n d\mu = \int_{\Omega} f d\mu.$$

Proof. Let some $\varepsilon > 0$ be fixed. It is easily seen that $|f(\omega)| \leq \varphi(\omega)$ for almost all ω . Therefore, as follows from the elementary properties of the integral, the function f is integrable. Let $\Omega_k = \{\omega : k - 1 \leq \varphi(\omega) < k\}$. Since the integral is a σ -additive function,

$$\int_{\Omega} \varphi d\mu = \sum_{k=1}^{\infty} \int_{\Omega_k} \varphi d\mu$$

Let $m > 0$ be such that $\sum_{k=m}^{\infty} \int_{\Omega_k} \varphi d\mu < \varepsilon/5$. Let $A = \bigcup_{k=m}^{\infty} \Omega_k$. By the Egorov Theorem, we can select a set $B \subseteq \Omega \setminus A$ such that $\mu(B) \leq \varepsilon/5m$ and f_n converges to f uniformly on the set $C = (\Omega \setminus A) \setminus B$. Finally,

$$\begin{aligned} \left| \int_{\Omega} f_n d\mu - \int_{\Omega} f d\mu \right| &\leq \left| \int_A f_n d\mu - \int_A f d\mu \right| \\ &+ \left| \int_B f_n d\mu - \int_B f d\mu \right| + \left| \int_C f_n d\mu - \int_C f d\mu \right|. \end{aligned}$$

The first term on the right-hand side can be estimated from above by $2\varepsilon/5$, since $\int_A |f_n| d\mu, \int_A |f| d\mu \leq \int_A \varphi d\mu < \varepsilon/5$. The second term does not exceed

$\mu(B) \sup_{\omega \in B} (|f_n(\omega)| + |f(\omega)|) \leq 2\varepsilon/5$. The last term can be made smaller than $\varepsilon/5$ for sufficiently large n due to the uniform convergence of f_n to f on the set C . Therefore, $|\int_{\Omega} f_n d\mu - \int_{\Omega} f d\mu| \leq \varepsilon$ for sufficiently large n , which completes the proof of the theorem. \square

From the Lebesgue Dominated Convergence Theorem it is easy to derive the following two statements, which we provide here without proof.

Theorem 3.27 (Levi Monotonic Convergence Theorem). *Let a sequence of measurable functions be non-decreasing almost surely, that is*

$$f_1(\omega) \leq f_2(\omega) \leq \dots \leq f_n(\omega) \leq \dots$$

almost surely. Assume that the integrals are bounded:

$$\int_{\Omega} f_n d\mu \leq K \quad \text{for all } n.$$

Then, almost surely, there exists a finite limit

$$f(\omega) = \lim_{n \rightarrow \infty} f_n(\omega),$$

the function f is integrable, and $\int_{\Omega} f d\mu = \lim_{n \rightarrow \infty} \int_{\Omega} f_n d\mu$.

Lemma 3.28 (Fatou Lemma). *If f_n is a sequence of non-negative measurable functions, then*

$$\int_{\Omega} \liminf_{n \rightarrow \infty} f_n d\mu \leq \liminf_{n \rightarrow \infty} \int_{\Omega} f_n d\mu \leq \infty.$$

Let us discuss products of σ -algebras and measures. Let $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$ be two measurable spaces with finite measures. We shall define the product space with the product measure $(\Omega, \mathcal{F}, \mu)$ as follows. The set Ω is just a set of ordered pairs $\Omega = \Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2), \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$.

In order to define the product σ -algebra, we first consider the collection of rectangles $\mathcal{R} = \{A \times B, A \in \mathcal{F}_1, B \in \mathcal{F}_2\}$. Then \mathcal{F} is defined as the smallest σ -algebra containing all the elements of \mathcal{R} .

Note that \mathcal{R} is a semialgebra. The product measure μ on \mathcal{F} is defined to be the extension to the σ -algebra of the function m defined on \mathcal{R} via $m(A \times B) = \mu_1(A)\mu_2(B)$. In order to justify this extension, we need to prove that m is a σ -additive function on \mathcal{R} .

Lemma 3.29. *The function $m(A \times B) = \mu_1(A)\mu_2(B)$ is a σ -additive function on the semialgebra \mathcal{R} .*

Proof. Let $A_1 \times B_1, A_2 \times B_2, \dots$ be a sequence of non-intersecting rectangles such that $A \times B = \bigcup_{n=1}^{\infty} A_n \times B_n$. Consider the sequence of functions

$f_n(\omega_1) = \sum_{i=1}^n \chi_{A_i}(\omega_1)\mu_2(B_i)$, where χ_{A_i} is the indicator function of the set A_i . Similarly, let $f(\omega_1) = \chi_A(\omega_1)\mu_2(B)$. Note that $f_n \leq \mu_2(B)$ for all n and $\lim_{n \rightarrow \infty} f_n(\omega_1) = f(\omega_1)$. Therefore, the Lebesgue Dominated Convergence Theorem applies. We have

$$\begin{aligned} \lim_{n \rightarrow \infty} \sum_{i=1}^n m(A_i \times B_i) &= \lim_{n \rightarrow \infty} \sum_{i=1}^n \mu_1(A_i)\mu_2(B_i) = \lim_{n \rightarrow \infty} \int_{\Omega_1} f_n(\omega_1) d\mu_1(\omega_1) \\ &= \int_{\Omega_1} f(\omega_1) d\mu_1(\omega_1) = \mu_1(A)\mu_2(B) = m(A \times B). \end{aligned}$$

□

We are now in a position to state the Fubini Theorem. If $(\Omega, \mathcal{F}, \mu)$ is a measurable space with a finite measure, and f is defined on a set of full measure $A \in \mathcal{F}$, then $\int_{\Omega} f d\mu$ will mean $\int_A f d\mu$.

Theorem 3.30 (Fubini Theorem). *Let $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$ be two measurable spaces with finite measures, and let $(\Omega, \mathcal{F}, \mu)$ be the product space with the product measure. If a function $f(\omega_1, \omega_2)$ is integrable with respect to the measure μ , then*

$$\begin{aligned} &\int_{\Omega} f(\omega_1, \omega_2) d\mu(\omega_1, \omega_2) \\ &= \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) d\mu_2(\omega_2) \right) d\mu_1(\omega_1) \\ &= \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2) d\mu_1(\omega_1) \right) d\mu_2(\omega_2). \end{aligned} \tag{3.3}$$

In particular, the integrals inside the brackets are finite almost surely and are integrable functions of the exterior variable.

Sketch of the Proof. The fact that the theorem holds if f is an indicator function of a set $A \times B$, where $A \in \mathcal{F}_1, B \in \mathcal{F}_2$, follows from the construction of the Lebesgue measure on the product space. The fact that the theorem also holds if f is an indicator function of a measurable set then easily follows from Lemma 4.13 proved in the next chapter. □

Concerning f which is not necessarily an indicator function, without loss of generality we may assume that f is non-negative. If f is a simple integrable function with a finite number of values, we can represent it as a finite linear combination of indicator functions, and therefore the theorem holds for such functions. If f is any integrable function, we can approximate it by a monotonically non-decreasing sequence of simple integrable functions with finite number of values. Then from the Levi Convergence Theorem it follows that the repeated integrals are finite and are equal to the integral on the left-hand side of (3.3).

3.6 Signed Measures and the Radon-Nikodym Theorem

In this section we state, without proof, the Radon-Nikodym Theorem and the Hahn Decomposition Theorem. Both proofs can be found in the textbook of S. Fomin and A. Kolmogorov, “Elements of Theory of Functions and Functional Analysis”.

Definition 3.31. Let (Ω, \mathcal{F}) be a measurable space. A function $\eta : \mathcal{F} \rightarrow \mathbb{R}$ is called a signed measure if

$$\eta\left(\bigcup_{i=1}^{\infty} C_i\right) = \sum_{i=1}^{\infty} \eta(C_i)$$

whenever $C_i \in \mathcal{F}$, $i \geq 1$, are such that $C_i \cap C_j = \emptyset$ for $i \neq j$.

If μ is a non-negative measure on (Ω, \mathcal{F}) , then an example of a signed measure is provided by the integral of a function with respect to μ ,

$$\eta(A) = \int_A f d\mu,$$

where $f \in L^1(\Omega, \mathcal{F}, \mu)$. Later, when we talk about conditional expectations, it will be important to consider the converse problem—given a measure μ and a signed measure η , we would like to represent η as an integral of some function with respect to measure μ . In fact this is always possible, provided $\mu(A) = 0$ for a set $A \in \mathcal{F}$ implies that $\eta(A) = 0$ (which is, of course, true if $\eta(A)$ is an integral of some function over the set A).

To make our discussion more precise we introduce the following definition.

Definition 3.32. Let (Ω, \mathcal{F}) be a measurable space with a finite non-negative measure μ . A signed measure $\eta : \mathcal{F} \rightarrow \mathbb{R}$ is called absolutely continuous with respect to μ if $\mu(A) = 0$ implies that $\eta(A) = 0$ for $A \in \mathcal{F}$.

Remark 3.33. An equivalent definition of absolute continuity is as follows. A signed measure $\eta : \mathcal{F} \rightarrow \mathbb{R}$ is called absolutely continuous with respect to μ if for any $\varepsilon > 0$ there is a $\delta > 0$ such that $\mu(A) < \delta$ implies that $|\eta(A)| < \varepsilon$. (In Problem 10 the reader is asked to prove the equivalence of the definitions when η is a non-negative measure.)

Theorem 3.34 (Radon-Nikodym Theorem). Let (Ω, \mathcal{F}) be a measurable space with a finite non-negative measure μ , and η a signed measure absolutely continuous with respect to μ . Then there is an integrable function f such that

$$\eta(A) = \int_A f d\mu$$

for all $A \in \mathcal{F}$. Any two functions which have this property can be different on at most a set of μ -measure zero.

The function f is called the density or the Radon-Nikodym derivative of η with respect to the measure μ .

The following theorem implies that signed measures are simply differences of two non-negative measures.

Theorem 3.35 (Hahn Decomposition Theorem). *Let (Ω, \mathcal{F}) be a measurable space with a signed measure $\eta : \mathcal{F} \rightarrow \mathbb{R}$. Then there exist two sets $\Omega^+ \in \mathcal{F}$ and $\Omega^- \in \mathcal{F}$ such that*

1. $\Omega^+ \cup \Omega^- = \Omega$ and $\Omega^+ \cap \Omega^- = \emptyset$.
2. $\eta(A \cap \Omega^+) \geq 0$ for any $A \in \mathcal{F}$.
3. $\eta(A \cap \Omega^-) \leq 0$ for any $A \in \mathcal{F}$.

If $\tilde{\Omega}^+, \tilde{\Omega}^-$ is another pair of sets with the same properties, then $\eta(A) = 0$ for any $A \in \mathcal{F}$ such that $A \in \Omega^+ \Delta \tilde{\Omega}^+$ or $A \in \Omega^- \Delta \tilde{\Omega}^-$.

Consider two non-negative measures η^+ and η^- defined by

$$\eta^+(A) = \eta(A \cap \Omega^+) \quad \text{and} \quad \eta^-(A) = -\eta(A \cap \Omega^-).$$

These are called the positive part and the negative part of η , respectively. The measure $|\eta| = \eta^+ + \eta^-$ is called the total variation of η . It easily follows from the Hahn Decomposition Theorem that η^+, η^- , and $|\eta|$ do not depend on the particular choice of Ω^+ and Ω^- . Given a measurable function f which is integrable with respect to $|\eta|$, we can define

$$\int_{\Omega} f d\eta = \int_{\Omega} f d\eta^+ - \int_{\Omega} f d\eta^-.$$

3.7 L^p Spaces

Let $(\Omega, \mathcal{F}, \mu)$ be a space with a finite measure. We shall call two complex-valued measurable functions f and g equivalent ($f \sim g$) if $\mu(f \neq g) = 0$. Note that \sim is indeed an equivalence relationship, i.e.,

1. $f \sim f$.
2. $f \sim g$ implies that $g \sim f$.
3. $f \sim g$ and $g \sim h$ imply that $f \sim h$.

It follows from general set theory that the set of measurable functions can be viewed as a union of non-intersecting subsets, the elements of the same subset being all equivalent, and the elements which belong to different subsets not being equivalent.

We next introduce the $L^p(\Omega, \mathcal{F}, \mu)$ spaces, whose elements are some of the equivalence classes of measurable functions. We shall not distinguish between a measurable function and the equivalence class it represents.

For $1 \leq p < \infty$ we define

$$\|f\|_p = \left(\int_{\Omega} |f|^p d\mu \right)^{\frac{1}{p}}.$$

The set of functions (or rather the set of equivalence classes) for which $\|f\|_p$ is finite is denoted by $L^p(\Omega, \mathcal{F}, \mu)$ or simply L^p . It readily follows that L^p is a normed linear space, with the norm $\|\cdot\|_p$, that is

1. $\|f\|_p \geq 0$, $\|f\|_p = 0$ if and only if $f = 0$.
2. $\|\alpha f\|_p = |\alpha| \|f\|_p$ for any complex number α .
3. $\|f + g\|_p \leq \|f\|_p + \|g\|_p$.

It is also not difficult to see, and we leave it for the reader as an exercise, that all the L^p spaces are complete. We also formulate the Hölder Inequality, which states that if $f \in L^p$ and $g \in L^q$ with $p, q > 1$ such that $1/p + 1/q = 1$, then $fg \in L^1$ and

$$\|fg\|_1 \leq \|f\|_p \|g\|_q.$$

When $p = q = 2$ this is also referred to as the Cauchy-Bunyakovskii Inequality. Its proof is available in many textbooks, and thus we omit it, leaving it as an exercise for the reader.

The norm in the L^2 space comes from the inner product, $\|f\|_2 = (f, f)^{1/2}$, where

$$(f, g) = \int_{\Omega} f \bar{g} d\mu.$$

The set L^2 equipped with this inner product is a Hilbert space.

3.8 Monte Carlo Method

Consider a bounded measurable set $U \subset \mathbb{R}^d$ and a bounded measurable function $f : U \rightarrow \mathbb{R}$. In this section we shall discuss a numerical method for evaluating the integral $I(f) = \int_U f(x) dx_1 \dots dx_d$.

One way to evaluate such an integral is based on approximating it by Riemann sums. Namely, the set U is split into measurable subsets U_1, \dots, U_n with small diameters, and a point x_i is selected in each of the subsets U_i . Then the sum $\sum_{i=1}^n f(x_i) \lambda(U_i)$, where $\lambda(U_i)$ is the measure of U_i , serves as an approximation to the integral. This method is effective provided that f does not change much for a small change of the argument (for example, if its gradient is bounded), and if we can split the set U into a reasonably small number of subsets with small diameters (so that n is not too large for a computer to handle the summation).

On the other hand, consider the case when U is a unit cube in \mathbb{R}^d , and d is large (say, $d = 20$). If we try to divide U into cubes U_i , each with the side of length $1/10$ (these may still be rather large, depending on the desired accuracy of the approximation), there will be $n = 10^{20}$ of such sub-cubes,

which shows that approximating the integral by the Riemann sums cannot be effective in high dimensions.

Now we describe the Monte Carlo method of numerical integration. Consider a homogeneous sequence of independent trials $\omega = (\omega_1, \omega_2, \dots)$, where each $\omega_i \in U$ has uniform distribution in U , that is $P(\omega_i \in V) = \lambda(V)/\lambda(U)$ for any measurable set $V \subseteq U$. If U is a unit cube, such a sequence can be implemented in practice with the help of a random number generator. Let

$$I^n(\omega) = \sum_{i=1}^n f(\omega_i).$$

We claim that I^n/n converges (in probability) to $I(f)/\lambda(U)$.

Theorem 3.36. *For every bounded measurable function f and every $\varepsilon > 0$*

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{I^n}{n} - \frac{I(f)}{\lambda(U)}\right| < \varepsilon\right) = 1.$$

Proof. Let $\varepsilon > 0$ be fixed, and assume that $|f(x)| \leq M$ for all $x \in U$ and some constant M . We split the interval $[-M, M]$ into k disjoint sub-intervals $\Delta_1, \dots, \Delta_k$, each of length not greater than $\varepsilon/3$. The number of such intervals should not need to exceed $1 + 6M/\varepsilon$. We define the sets U_j as the pre-images of Δ_j , that is $U_j = f^{-1}(\Delta_j)$. Let us fix a point a_j in each Δ_j . Let $\nu_j^n(\omega)$ be the number of those ω_i with $1 \leq i \leq n$, for which $\omega_i \in U_j$. Let $J^n(\omega) = \sum_{j=1}^k a_j \nu_j^n(\omega)$.

Since $f(x)$ does not vary by more than $\varepsilon/3$ on each of the sets U_j ,

$$\left|\frac{I^n(\omega)}{n} - \frac{J^n(\omega)}{n}\right| \leq \frac{\varepsilon}{3} \quad \text{and} \quad \frac{|I(f) - \sum_{j=1}^k a_j \lambda(U_j)|}{\lambda(U)} \leq \frac{\varepsilon}{3}.$$

Therefore, it is sufficient to demonstrate that

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{J^n}{n} - \frac{\sum_{j=1}^k a_j \lambda(U_j)}{\lambda(U)}\right| < \frac{\varepsilon}{3}\right) = 1,$$

or, equivalently,

$$\lim_{n \rightarrow \infty} P\left(\left|\sum_{j=1}^k a_j \left(\frac{\nu_j^n}{n} - \frac{\lambda(U_j)}{\lambda(U)}\right)\right| < \frac{\varepsilon}{3}\right) = 1.$$

This follows from the law of large numbers, which states that ν_j^n/n converges in probability to $\lambda(U_j)/\lambda(U)$ for each j . \square

Remark 3.37. Later we shall prove the so-called strong law of large numbers, which will imply the almost sure convergence of the approximations in the Monte Carlo method (see Chap. 7). It is important that the convergence rate (however it is defined) can be estimated in terms of $\lambda(U)$ and $\sup_{x \in U} |f(x)|$, independently of the dimension of the space and the smoothness of the function f .

3.9 Problems

1. Let f_n , $n \geq 1$, and f be measurable functions on a measurable space (Ω, \mathcal{F}) . Prove that the set $\{\omega : \lim_{n \rightarrow \infty} f_n(\omega) = f(\omega)\}$ is \mathcal{F} -measurable. Prove that the set $\{\omega : \lim_{n \rightarrow \infty} f_n(\omega) \text{ exists}\}$ is \mathcal{F} -measurable.
2. Prove that if a random variable ξ taking non-negative values is such that

$$P(\xi \geq n) \geq 1/n \quad \text{for all } n \in \mathbb{N},$$

then $E\xi = \infty$.

3. Construct a sequence of random variables ξ_n such that $\xi_n(\omega) \rightarrow 0$ for every ω , but $E\xi_n \rightarrow \infty$ as $n \rightarrow \infty$.
4. A random variable ξ takes values in the interval $[A, B]$ and $\text{Var}(\xi) = ((B - A)/2)^2$. Find the distribution of ξ .
5. Let $\{x_1, x_2, \dots\}$ be a collection of rational points from the interval $[0, 1]$. A random variable ξ takes the value x_n with probability $1/2^n$. Prove that the distribution function $F_\xi(x)$ of ξ is continuous at every irrational point x .
6. Let ξ be a random variable with a continuous density p_ξ such that $p_\xi(0) > 0$. Find the density of η , where

$$\eta(\omega) = \begin{cases} 1/\xi(\omega) & \text{if } \xi(\omega) \neq 0, \\ 0 & \text{if } \xi(\omega) = 0. \end{cases}$$

Prove that η does not have a finite expectation.

7. Let ξ_1, ξ_2, \dots be a sequence of random variables on a probability space (Ω, \mathcal{F}, P) such that $E|\xi_n| \leq 2^{-n}$. Prove that $\xi_n \rightarrow 0$ almost surely as $n \rightarrow \infty$.
8. Prove that if a sequence of measurable functions f_n converges to f almost surely as $n \rightarrow \infty$, then it also converges to f in measure. If f_n converges to f in measure, then there is a subsequence f_{n_k} which converges to f almost surely as $k \rightarrow \infty$.
9. Let $F(x)$ be a distribution function. Compute $\int_{-\infty}^{\infty} (F(x+10) - F(x)) dx$.
10. Prove that a measure η is absolutely continuous with respect to a measure μ if and only if for any $\varepsilon > 0$ there is a $\delta > 0$ such that $\mu(A) < \delta$ implies that $\eta(A) < \varepsilon$.
11. Prove that the $L^p([0, 1], \mathcal{B}, \lambda)$ spaces are complete for $1 \leq p < \infty$. Here \mathcal{B} is the σ -algebra of Borel sets, and λ is the Lebesgue measure.
12. Prove the Hölder Inequality.
13. Let ξ_1, ξ_2, \dots be a sequence of random variables on a probability space (Ω, \mathcal{F}, P) such that $E\xi_n^2 \leq c$ for some constant c . Assume that $\xi_n \rightarrow \xi$ almost surely as $n \rightarrow \infty$. Prove that $E\xi$ is finite and $E\xi_n \rightarrow E\xi$.