

### 3.10: The Integral With Respect to a Measure

Probability density functions have very different interpretations for discrete distributions as opposed to continuous distributions. For a discrete distribution, the probability of an event is computed by summing the density function over the outcomes in the event, while for a continuous distribution, the probability is computed by integrating the density function over the outcomes. For a mixed distributions, we have partial discrete and continuous density functions and the probability of an event is computed by summing and integrating. The various types of density functions can unified under a general theory of integration, which is the subject of this section. This theory has enormous importance in probability, far beyond just density functions. Expected value, which we consider in the next chapter, can be interpreted as an integral with respect to a probability measure. Beyond probability, the general theory of integration is of fundamental importance in many areas of mathematics.

#### Basic Theory

##### Definitions

Our starting point is a measure space  $(S, \mathcal{S}, \mu)$ . That is,  $S$  is a set,  $\mathcal{S}$  is a  $\sigma$ -algebra of subsets of  $S$ , and  $\mu$  is a positive measure on  $\mathcal{S}$ . As usual, the most important special cases are

- *Euclidean space:*  $S = \mathbb{R}^n$  for some  $n \in \mathbb{N}_+$ ,  $\mathcal{S} = \mathcal{R}_n$ , the  $\sigma$ -algebra of Lebesgue measurable subsets of  $\mathbb{R}^n$ , and  $\mu = \lambda_n$ , standard  $n$ -dimensional Lebesgue measure.
- *Discrete space:*  $S$  is a countable set,  $\mathcal{S}$  is the collection of all subsets of  $S$ , and  $\mu = \#$ , counting measure.
- *Probability space:*  $S$  is the set of outcomes of a random experiment,  $\mathcal{S}$  is the  $\sigma$ -algebra of events, and  $\mu = \mathbb{P}$ , a probability measure.

The following definition reflects the fact that in measure theory, sets of measure 0 are often considered unimportant.

Consider a statement with  $x \in S$  as a free variable. Technically such a statement is a *predicate* on  $S$ . Suppose that  $A \in \mathcal{S}$ .

1. The statement holds *on*  $A$  if it is true for every  $x \in A$ .
2. The statement holds *almost everywhere on*  $A$  (with respect to  $\mu$ ) if there exists  $B \in \mathcal{S}$  with  $B \subseteq A$  such that the statement holds on  $B$  and  $\mu(A \setminus B) = 0$ .

A typical statement that we have in mind is an equation or an inequality with  $x \in S$  as a free variable. Our goal is to define the *integral* of certain measurable functions  $f : S \rightarrow \mathbb{R}$ , with respect to the measure  $\mu$ . The integral may exist as a number in  $\mathbb{R}$  (in which case we say that  $f$  is *integrable*), or may exist as  $\infty$  or  $-\infty$ , or may not exist at all. When it exists, the integral is denoted variously by

$$\int_S f d\mu, \int_S f(x) d\mu(x), \int_S f(x) \mu(dx) \quad (3.10.1)$$

We will use the first two.

Since the set of extended real numbers  $\mathbb{R}^* = \mathbb{R} \cup \{-\infty, \infty\}$  plays an important role in the theory, we need to recall the arithmetic of  $\infty$  and  $-\infty$ . Here are the conventions that are appropriate for integration:

##### Arithmetic on $\mathbb{R}^*$

1. If  $a \in (0, \infty]$  then  $a \cdot \infty = \infty$  and  $a \cdot (-\infty) = -\infty$
2. If  $a \in [-\infty, 0)$  then  $a \cdot \infty = -\infty$  and  $a \cdot (-\infty) = \infty$
3.  $0 \cdot \infty = 0$  and  $0 \cdot (-\infty) = 0$
4. If  $a \in \mathbb{R}$  then  $a + \infty = \infty$  and  $a + (-\infty) = -\infty$
5.  $\infty + \infty = \infty$
6.  $-\infty + (-\infty) = -\infty$

However,  $\infty - \infty$  is not defined (because it does not make consistent sense) and we must be careful never to produce this indeterminate form. You might recall from calculus that  $0 \cdot \infty$  is also an indeterminate form. However, for the theory of integration, the convention that  $0 \cdot \infty = 0$  is convenient and consistent. In terms of order of course,  $-\infty < a < \infty$  for  $a \in \mathbb{R}$ .

We also need to extend topology and measure to  $\mathbb{R}^*$ . In terms of the first,  $(a, \infty]$  is an open neighborhood of  $\infty$  and  $[-\infty, a)$  is an open neighborhood of  $-\infty$  for every  $a \in \mathbb{R}$ . This ensures that if  $x_n \in \mathbb{R}$  for  $n \in \mathbb{N}_+$  then  $x_n \rightarrow \infty$  or  $x_n \rightarrow -\infty$  as  $n \rightarrow \infty$  has its usual calculus meaning. Technically this topology results in the *two-point compactification* of  $\mathbb{R}$ . Now we can give  $\mathbb{R}^*$  the Borel  $\sigma$ -algebra  $\mathcal{R}^*$ , that is, the  $\sigma$ -algebra generated by the topology. Basically, this simply means that if  $A \in \mathcal{R}$  then  $A \cup \{\infty\}$ ,  $A \cup \{-\infty\}$ , and  $A \cup \{-\infty, \infty\}$  are all in  $\mathcal{R}^*$ .

#### Desired Properties

As motivation for the definition, every version of integration should satisfy some basic properties. First, the integral of the indicator function of a measurable set should simply be the size of the set, as measured by  $\mu$ . This gives our first definition:

If  $A \in \mathcal{S}$  then  $\int_S \mathbf{1}_A d\mu = \mu(A)$ .

This definition hints at the intimate relationship between measure and integration. We will construct the integral from the measure  $\mu$  in this section, but this first property shows that if we started with the integral, we could recover the measure. This property also shows why we need  $\infty$  as a possible value of the integral, and coupled with some of the properties below, why  $-\infty$  is also needed. Here is a simple corollary of our first definition.

$$\int_S 0 d\mu = 0$$

Proof

Note that  $\int_S 0 d\mu = \int_S \mathbf{1}_\emptyset d\mu = \mu(\emptyset) = 0$ .

We give three more essential properties that we want. First are the *linearity properties* in two parts—part (a) is the *additive property* and part (b) is the *scaling property*.

If  $f, g : S \rightarrow \mathbb{R}$  are measurable functions whose integrals exist, and  $c \in \mathbb{R}$ , then

1.  $\int_S (f + g) d\mu = \int_S f d\mu + \int_S g d\mu$  as long as the right side is not of the form  $\infty - \infty$
2.  $\int_S cf d\mu = c \int_S f d\mu$ .

The additive property almost implies the scaling property

The steps below do not constitute a proof because questions of the existence of the integrals are ignored and because the limit interchange in the last step is not justified. Still, the argument shows the close relationship between the additive property and the scaling property.

1. If  $n \in \mathbb{N}_+$ , then by (a) and induction,  $\int_S nf d\mu = n \int_S f d\mu$ .
2. From step (1), if  $n \in \mathbb{N}_+$  then  $\int_S f d\mu = \int_S n \frac{1}{n} f d\mu = n \int_S \frac{1}{n} f d\mu$  so  $\int_S \frac{1}{n} f d\mu = \frac{1}{n} \int_S f d\mu$ .
3. If  $m, n \in \mathbb{N}_+$  then from steps (1) and (2)  $\int_S \frac{m}{n} f d\mu = m \int_S \frac{1}{n} f d\mu = \frac{m}{n} \int_S f d\mu$ .
4.  $0 = \int_S 0 d\mu = \int_S (f - f) d\mu = \int_S f d\mu + \int_S -f d\mu$  so  $\int_S -f d\mu = - \int_S f d\mu$ .
5. By steps (3) and (4),  $\int_S cf d\mu = c \int_S f d\mu$  for every  $c \in \mathbb{Q}$  (the set of rational real numbers).
6. If  $c \in \mathbb{R}$  there exists  $c_n \in \mathbb{Q}$  for  $n \in \mathbb{N}_+$  with  $c_n \rightarrow c$  as  $n \rightarrow \infty$ . By step (5),  $\int_S c_n f d\mu = c_n \int_S f d\mu$ .
7. Taking limits in step (6) suggests  $\int_S cf d\mu = c \int_S f d\mu$ .

To be more explicit, we want the additivity property (a) to hold if at least one of the integrals on the right is finite, or if both are  $\infty$  or if both are  $-\infty$ . What is ruled out are the two cases where one integral is  $\infty$  and the other is  $-\infty$ , and this is what is meant by the indeterminate form  $\infty - \infty$ . Our next essential properties are the *order properties*, again in two parts—part (a) is the *positive property* and part (b) is the *increasing property*.

Suppose that  $f, g : S \rightarrow \mathbb{R}$  are measurable.

1. If  $f \geq 0$  on  $S$  then  $\int_S f d\mu \geq 0$ .
2. If the integrals of  $f$  and  $g$  exist and  $f \leq g$  on  $S$  then  $\int_S f d\mu \leq \int_S g d\mu$

The positive property and the additive property imply the increasing property

Implicit in part (a) is that the integral of a nonnegative, measurable function always exists in  $[0, \infty]$ . Suppose that the integrals of  $f$  and  $g$  exist and  $f \leq g$  on  $S$ . Then  $g - f \geq 0$  on  $S$  and  $g = f + (g - f)$ . If  $\int_S f d\mu = -\infty$ , then trivially,  $\int_S f d\mu \leq \int_S g d\mu$ . Otherwise, by the additivity property,

$$\int_S g d\mu = \int_S f d\mu + \int_S (g - f) d\mu \quad (3.10.2)$$

But  $\int_S (g - f) d\mu \geq 0$  (so in particular the right side is not  $-\infty + \infty$ ), and hence  $\int_S g d\mu \geq \int_S f d\mu$

Our last essential property is perhaps the least intuitive, but is a type of *continuity property* of integration, and is closely related to the continuity property of positive measure. The official name is the *monotone convergence theorem*.

Suppose that  $f_n : S \rightarrow [0, \infty)$  is measurable for  $n \in \mathbb{N}_+$  and that  $f_n$  is increasing in  $n$ . Then

$$\int_S \lim_{n \rightarrow \infty} f_n d\mu = \lim_{n \rightarrow \infty} \int_S f_n d\mu \quad (3.10.3)$$

Note that since  $f_n$  is increasing in  $n$ ,  $\lim_{n \rightarrow \infty} f_n(x)$  exists in  $\mathbb{R} \cup \{\infty\}$  for each  $x \in \mathbb{R}$  (and the limit defines a measurable function). This property shows that it is sometimes convenient to allow nonnegative functions to take the value  $\infty$ . Note also that by the [increasing property](#),  $\int_S f_n d\mu$  is increasing in  $n$  and hence also has a limit in  $\mathbb{R} \cup \{\infty\}$ .

To see the connection with measure, suppose that  $(A_1, A_2, \dots)$  is an increasing sequence of sets in  $\mathcal{S}$ , and let  $A = \bigcup_{i=1}^{\infty} A_i$ . Note that  $\mathbf{1}_{A_n}$  is increasing in  $n \in \mathbb{N}_+$  and  $\mathbf{1}_{A_n} \rightarrow \mathbf{1}_A$  as  $n \rightarrow \infty$ . For this reason, the union  $A$  is sometimes called the *limit* of  $A_n$  as  $n \rightarrow \infty$ . The continuity theorem of positive measure states that  $\mu(A_n) \rightarrow \mu(A)$  as  $n \rightarrow \infty$ . Equivalently,  $\int_S \mathbf{1}_{A_n} d\mu \rightarrow \int_S \mathbf{1}_A d\mu$  as  $n \rightarrow \infty$ , so the continuity theorem of positive measure is a special case of the monotone convergence theorem.

Armed with the properties that we want, the definition of the integral is fairly straightforward, and proceeds in stages. We give the definition successively for

1. Nonnegative simple functions
2. Nonnegative measurable functions
3. Measurable real-valued functions

Of course, each definition should agree with the previous one on the functions that are in both collections.

## Simple Functions

A *simple function* on  $S$  is simply a measurable, real-valued function with finite range. Simple functions are usually expressed as linear combinations of indicator functions.

### Representations of simple functions

1. Suppose that  $I$  is a finite index set,  $a_i \in \mathbb{R}$  for each  $i \in I$ , and  $\{A_i : i \in I\}$  is a collection of sets in  $\mathcal{S}$  that partition  $S$ . Then  $f = \sum_{i \in I} a_i \mathbf{1}_{A_i}$  is a simple function. Expressing a simple function in this form is a *representation* of  $f$ .
2. A simple function  $f$  has a unique representation as  $f = \sum_{j \in J} b_j \mathbf{1}_{B_j}$  where  $J$  is a finite index set,  $\{b_j : j \in J\}$  is a set of distinct real numbers, and  $\{B_j : j \in J\}$  is a collection of nonempty sets in  $\mathcal{S}$  that partition  $S$ . This representation is known as the *canonical representation*.

#### Proof

1. Note that  $f$  is measurable since  $A_i \in \mathcal{S}$  for each  $i \in I$ . Also  $f$  has finite range since  $I$  is finite. Specifically, the range of  $f$  consists of the distinct  $a_i$  for  $i \in I$  with  $A_i \neq \emptyset$ .
2. Suppose that  $f$  is simple. Let  $\{b_j : j \in J\}$  denote the (distinct) values in the range of  $f$  and let  $B_j = f^{-1}\{b_j\}$  for  $j \in J$ . Then  $J$  is finite,  $\{B_j : j \in J\}$  is a collection of nonempty sets in  $\mathcal{S}$  that partition  $S$ , and  $f = \sum_{j \in J} b_j \mathbf{1}_{B_j}$ . Conversely, suppose that  $f$  has a representation of this form. Then  $\{b_j : j \in J\}$  is the range of  $f$  and  $B_j = f^{-1}\{b_j\}$  so the representation is unique.

You might wonder why we don't just always use the canonical representation for simple functions. The problem is that even if we start with canonical representations, when we combine simple functions in various ways, the resulting representations may not be canonical. The collection of simple functions is closed under the basic arithmetic operations, and in particular, forms a vector space.

Suppose that  $f$  and  $g$  are simple functions with representations  $f = \sum_{i \in I} a_i \mathbf{1}_{A_i}$  and  $g = \sum_{j \in J} b_j \mathbf{1}_{B_j}$ , and that  $c \in \mathbb{R}$ . Then

1.  $f + g$  is simple, with representation  $f + g = \sum_{(i,j) \in I \times J} (a_i + b_j) \mathbf{1}_{A_i \cap B_j}$ .
2.  $fg$  is simple, with representation  $fg = \sum_{(i,j) \in I \times J} (a_i b_j) \mathbf{1}_{A_i \cap B_j}$ .
3.  $cf$  is simple, with representation  $cf = \sum_{i \in I} ca_i \mathbf{1}_{A_i}$ .

#### Proof

Since  $f$  and  $g$  are measurable, so are  $f + g$ ,  $fg$ , and  $cf$ . Moreover, since  $f$  and  $g$  have finite range, so do  $f + g$ ,  $fg$ , and  $cf$ . For the representations in parts (a) and (b), note that  $I \times J$  is finite,  $\{A_i \cap B_j : (i, j) \in I \times J\}$  is a collection of sets in  $\mathcal{S}$  that partition  $S$ , and on  $A_i \cap B_j$ ,  $f + g = a_i + b_j$  and  $fg = a_i b_j$ .

As we alluded to earlier, note that even if the representations of  $f$  and  $g$  are canonical, the representations for  $f + g$  and  $fg$  may not be. The next result treats composition, and will be important for the change of variables theorem in the next section.

Suppose that  $(T, \mathcal{T})$  is another measurable space, and that  $f : S \rightarrow T$  is measurable. If  $g$  is a simple function on  $T$  with representation  $g = \sum_{i \in I} b_i \mathbf{1}_{B_i}$ , then  $g \circ f$  is a simple function on  $S$  with representation  $g \circ f = \sum_{i \in I} b_i \mathbf{1}_{f^{-1}(B_i)}$ .

#### Proof

Recall that  $g \circ f : S \rightarrow \mathbb{R}$  and  $\text{range}(g \circ f) \subseteq \text{range}(g)$  so  $g \circ f$  has finite range.  $f$  is measurable, and inverse images preserve all set operations, so  $\{f^{-1}(B_i) : i \in I\}$  is a measurable partition of  $S$ . Finally, if  $x \in f^{-1}(B_i)$  then  $f(x) \in B_i$  so  $g[f(x)] = b_i$ .

Given the definition of the integral of an indicator function in (3) and that we want the linearity property (5) to hold, there is no question as to how we should define the integral of a nonnegative simple function.

Suppose that  $f$  is a nonnegative simple function, with the representation  $f = \sum_{i \in I} a_i \mathbf{1}_{A_i}$  where  $a_i \geq 0$  for  $i \in I$ . We define

$$\int_S f d\mu = \sum_{i \in I} a_i \mu(A_i) \quad (3.10.4)$$

The definition is consistent

*Consistency* refers to the fact that a simple function can have more than one representation as a linear combination of indicator functions, and hence we must show that all such representations lead to the same value for the integral. Let  $\{b_j : j \in J\}$  denote the set of distinct elements among the numbers  $a_i$  where  $i \in I$  and  $A_i \neq \emptyset$ . For  $j \in J$ , let  $I_j = \{i \in I : a_i = b_j\}$  and let  $B_j = \bigcup_{i \in I_j} A_i$ . Thus,  $f = \sum_{j \in J} b_j \mathbf{1}_{B_j}$ , and this is the canonical representation. Note that

$$\sum_{i \in I} a_i \mu(A_i) = \sum_{j \in J} \sum_{i \in I_j} a_i \mu(A_i) = \sum_{j \in J} b_j \sum_{i \in I_j} \mu(A_i) = \sum_{j \in J} b_j \mu(B_j) \quad (3.10.5)$$

The first sum is the integral defined in terms of the general representation  $f = \sum_{i \in I} a_i \mathbf{1}_{A_i}$  while the last sum is the integral defined in terms of the unique canonical representation  $f = \sum_{j \in J} b_j \mathbf{1}_{B_j}$ . Thus, any representation of a simple function  $f$  leads to the same value for the integral.

Note that if  $f$  is a nonnegative simple function, then  $\int_S f d\mu$  exists in  $[0, \infty]$ , so the [order properties](#) holds. We next show that the [linearity properties](#) are satisfied for nonnegative simple functions.

Suppose that  $f$  and  $g$  are nonnegative simple functions, and that  $c \in [0, \infty)$ . Then

1.  $\int_S (f+g) d\mu = \int_S f d\mu + \int_S g d\mu$
2.  $\int_S cf d\mu = c \int_S f d\mu$

**Proof**

Suppose that  $f$  and  $g$  are nonnegative simple functions with the representations  $f = \sum_{i \in I} a_i \mathbf{1}_{A_i}$  and  $g = \sum_{j \in J} b_j \mathbf{1}_{B_j}$ . Thus  $a_i \geq 0$  for  $i \in I$ ,  $b_j \geq 0$  for  $j \in J$ , and  $\int_S f d\mu = \sum_{i \in I} a_i \mu(A_i)$  and  $\int_S g d\mu = \sum_{j \in J} b_j \mu(B_j)$ .

1. As noted above,  $f+g$  has the representation

$$f+g = \sum_{(i,j) \in I \times J} (a_i + b_j) \mathbf{1}_{A_i \cap B_j} \quad (3.10.6)$$

Note that  $\{A_i \cap B_j : j \in J\}$  is a partition of  $A_i$  for each  $i \in I$ , and similarly  $\{A_i \cap B_j : i \in I\}$  is a partition of  $B_j$  for each  $j \in J$ . Hence

$$\int_S (f+g) d\mu = \sum_{(i,j) \in I \times J} (a_i + b_j) \mu(A_i \cap B_j) \quad (3.10.7)$$

$$= \sum_{i \in I} \sum_{j \in J} a_i \mu(A_i \cap B_j) + \sum_{j \in J} \sum_{i \in I} b_j \mu(A_i \cap B_j) \quad (3.10.8)$$

$$= \sum_{i \in I} a_i \mu(A_i \cap B) + \sum_{j \in J} b_j \mu(B_j \cap A) \quad (3.10.9)$$

$$= \sum_{i \in I} a_i \mu(A_i) + \sum_{j \in J} b_j \mu(B_j) = \int_S f d\mu + \int_S g d\mu \quad (3.10.10)$$

Note that all the terms are nonnegative (although some may be  $\infty$ ), so there are no problems with rearranging the order of the terms.

2. This part is easier. For  $c \in [0, \infty)$ , recall that  $cf$  has the representation  $cf = \sum_{i \in I} ca_i \mathbf{1}_{A_i}$  so

$$\int_S cf d\mu = \sum_{i \in I} ca_i \mu(A_i) = c \sum_{i \in I} a_i \mu(A_i) = c \int_S f d\mu \quad (3.10.11)$$

The [increasing property](#) holds for nonnegative simple functions.

Suppose that  $f$  and  $g$  are nonnegative simple functions and  $f \leq g$  on  $S$ . Then  $\int_S f d\mu \leq \int_S g d\mu$

**Proof**

The proof from the [additive property](#) above works. Note that  $g-f$  is a nonnegative simple function, and  $g = f + (g-f)$ . By the additivity property,  $\int_S g d\mu = \int_S f d\mu + \int_S (g-f) d\mu \geq \int_S f d\mu$ .

Next we give a version of the continuity theorem in (7) for simple functions. It's not completely general, but will be needed for the next subsection where we do prove the general version.

Suppose that  $f$  is a nonnegative simple function and that  $(A_1, A_2, \dots)$  is an increasing sequence of sets in  $\mathcal{S}$  with  $A = \bigcup_{n=1}^{\infty} A_n$ . then

$$\int_S \mathbf{1}_{A_n} f d\mu \rightarrow \int_S \mathbf{1}_A f d\mu \text{ as } n \rightarrow \infty \quad (3.10.12)$$

Proof

Suppose that  $f$  has the representation  $f = \sum_{i \in I} b_i \mathbf{1}_{B_i}$ . Then  $\mathbf{1}_{A_n} f = \sum_{i \in I} b_i \mathbf{1}_{A_n} \mathbf{1}_{B_i} = \sum_{i \in I} b_i \mathbf{1}_{A_n \cap B_i}$  and similarly,  $\mathbf{1}_A f = \sum_{i \in I} b_i \mathbf{1}_{A \cap B_i}$ . But for each  $i \in I$ ,  $B_i \cap A_n$  is increasing in  $n \in \mathbb{N}_+$  and  $\bigcup_{n=1}^{\infty} (B_i \cap A_n) = B_i \cap A$ . By the continuity theorem for positive measures,  $\mu(B_i \cap A_n) \rightarrow \mu(B_i \cap A)$  as  $n \rightarrow \infty$  for each  $i \in I$ . Since  $I$  is finite,

$$\int_{A_n} f d\mu = \sum_{i \in I} b_i \mu(A_n \cap B_i) \rightarrow \sum_{i \in I} b_i \mu(A \cap B_i) = \int_A f d\mu \text{ as } n \rightarrow \infty \quad (3.10.13)$$

Note that  $\mathbf{1}_{A_n} f$  is increasing in  $n \in \mathbb{N}_+$  and  $\mathbf{1}_{A_n} f \rightarrow \mathbf{1}_A f$  as  $n \rightarrow \infty$ , so this really is a special case of the monotone convergence theorem.

## Nonnegative Functions

Next we will consider nonnegative measurable functions on  $S$ . First we note that a function of this type is the limit of nonnegative simple functions.

Suppose that  $f : S \rightarrow [0, \infty)$  is measurable. Then there exists an increasing sequence  $(f_1, f_2, \dots)$  of nonnegative simple functions with  $f_n \rightarrow f$  on  $S$  as  $n \rightarrow \infty$ .

Proof

For  $n \in \mathbb{N}_+$  and  $k \in \{1, 2, \dots, n2^n\}$  Let  $I_{n,k} = [(k-1)/2^n, k/2^n)$  and  $I_n = [n, \infty)$ . Note that

1.  $\{I_{n,k} : k = 1, \dots, n2^n\} \cup \{I_n\}$  is a partition of  $[0, \infty)$  for each  $n \in \mathbb{N}_+$ .
2.  $I_{n,k} = I_{n+1, 2k-1} \cup I_{n+1, 2k}$  for  $k \in \{1, 2, \dots, n2^n\}$ .
3.  $I_n = \left(\bigcup_{k=n2^{n+1}+1}^{(n+1)2^{n+1}} I_{n+1,k}\right) \cup I_{n+1}$  for  $n \in \mathbb{N}_+$ .

Note that the  $n$ th partition divides the interval  $[0, n)$  into  $n2^n$  subintervals of length  $1/2^n$ . Thus, (b) follows because the  $(n+1)$ st partition divides each of the first  $2^n$  intervals of the  $n$ th partition in half, and (c) follows because the  $(n+1)$ st partition divides the interval  $[n, n+1)$  into subintervals of length  $1/2^{n+1}$ . Now let  $A_{n,k} = f^{-1}(I_{n,k})$  and  $A_n = f^{-1}(I_n)$  for  $n \in \mathbb{N}_+$  and  $k \in \{1, 2, \dots, n2^n\}$ . Since inverse images preserve all set operations, (a), (b), and (c) hold with  $A$  replacing  $I$  everywhere, and  $S$  replacing  $[0, \infty)$  in (a). Moreover, since  $f$  is measurable,  $A_n \in \mathcal{S}$  and  $A_{n,k} \in \mathcal{S}$  for each  $n$  and  $k$ . Now, define

$$f_n = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \mathbf{1}_{A_{n,k}} + n \mathbf{1}_{A_n} \quad (3.10.14)$$

Then  $f_n$  is a simple function and  $0 \leq f_n \leq f$  for each  $n \in \mathbb{N}_+$ . To show convergence, fix  $x \in S$ . If  $n > f(x)$  then  $|f(x) - f_n(x)| \leq 2^{-n}$  and hence  $f_n(x) \rightarrow f(x)$  as  $n \rightarrow \infty$ . All that remains is to show that  $f_n$  is increasing in  $n$ . Let  $x \in S$  and  $n \in \mathbb{N}_+$ . If  $x \in A_{n,k}$  for some  $k \in \{1, 2, \dots, n2^n\}$ , then  $f_n(x) = (k-1)/2^n$ . But either  $f_{n+1}(x) = (2k-2)/2^{n+1}$  or  $f_{n+1}(x) = (2k-1)/2^{n+1}$ . If  $x \in A_n$  then  $f_n(x) = n$ . But either  $f_{n+1}(x) = (k-1)/2^{n+1}$  for some  $k \in \{n2^{n+1}+1, \dots, (n+1)2^{n+1}\}$  or  $f_{n+1}(x) = n+1$ . In all cases,  $f_{n+1}(x) \geq f_n(x)$ .

The last result points the way towards the definition of the integral of a measurable function  $f : S \rightarrow [0, \infty)$  in terms of the integrals of simple functions. If  $g$  is a nonnegative simple function with  $g \leq f$ , then by the [order property](#), we need  $\int_S g d\mu \leq \int_S f d\mu$ . On the other hand, there exists a sequence of nonnegative simple function converging to  $f$ . Thus the [continuity property](#) suggests the following definition:

If  $f : S \rightarrow [0, \infty)$  is measurable, define

$$\int_S f d\mu = \sup \left\{ \int_S g d\mu : g \text{ is simple and } 0 \leq g \leq f \right\} \quad (3.10.15)$$

Note that  $\int_S f d\mu$  exists in  $[0, \infty]$  so the [positive property](#) holds. Note also that if  $f$  is simple, the new definition agrees with the old one. As always, we need to establish the essential properties. First, the [increasing property](#) holds.

If  $f, g : S \rightarrow [0, \infty)$  are measurable and  $f \leq g$  on  $S$  then  $\int_S f d\mu \leq \int_S g d\mu$ .

Proof

Note that  $\{h : h \text{ is simple and } 0 \leq h \leq f\} \subseteq \{h : h \text{ is simple and } 0 \leq h \leq g\}$ . therefore

$$\int_S f d\mu = \sup \left\{ \int_S h d\mu : h \text{ is simple and } 0 \leq h \leq f \right\} \leq \sup \left\{ \int_S h d\mu : h \text{ is simple and } 0 \leq h \leq g \right\} = \int_S g d\mu \quad (3.10.16)$$

We can now prove the [continuity property](#) known as the *monotone convergence theorem* in full generality.

Suppose that  $f_n : S \rightarrow [0, \infty)$  is measurable for  $n \in \mathbb{N}_+$  and that  $f_n$  is increasing in  $n$ . Then

$$\int_S \lim_{n \rightarrow \infty} f_n d\mu = \lim_{n \rightarrow \infty} \int_S f_n d\mu \quad (3.10.17)$$

Proof

Let  $f = \lim_{n \rightarrow \infty} f_n$ . By the order property, note that  $\int_S f_n d\mu$  is increasing in  $n \in \mathbb{N}_+$  and hence has a limit in  $\mathbb{R}^*$ , which we will denote by  $c$ . Note that  $f_n \leq f$  on  $S$  for  $n \in \mathbb{N}_+$ , so by the order property again,  $\int_S f_n d\mu \leq \int_S f d\mu$  for  $n \in \mathbb{N}_+$ . Letting  $n \rightarrow \infty$  gives  $c \leq \int_S f d\mu$ . To show that  $c \geq \int_S f d\mu$  we need to show that  $c \geq \int_S g d\mu$  for every simple function  $g$  with  $0 \leq g \leq f$ . Fix  $a \in (0, 1)$  and let  $A_n = \{x \in S : f_n(x) \geq ag(x)\}$ . Since  $f_n$  is increasing in  $n$ ,  $A_n \subseteq A_{n+1}$ . Moreover, since  $f_n \rightarrow f$  as  $n \rightarrow \infty$  on  $S$  and  $g \leq f$  on  $S$ ,  $\bigcup_{n=1}^{\infty} A_n = S$ . But by definition,  $ag \leq f_n$  on  $A_n$  so

$$\alpha \int_S \mathbf{1}_{A_n} g d\mu = \int_S \alpha \mathbf{1}_{A_n} g d\mu \leq \int_S \mathbf{1}_{A_n} f_n d\mu \leq \int_S f_n d\mu \quad (3.10.18)$$

Letting  $n \rightarrow \infty$  in the extreme parts of the displayed inequality and using the version of the monotone convergence theorem for [simple functions](#), we have  $a \int_S g d\mu \leq c$  for every  $a \in (0, 1)$ . Finally, letting  $a \uparrow 1$  gives  $\int_S g d\mu \leq c$

If  $f : S \rightarrow [0, \infty)$  is measurable, then by the [theorem above](#), there exists an increasing sequence  $(f_1, f_2, \dots)$  of simple functions with  $f_n \rightarrow f$  as  $n \rightarrow \infty$ . By the monotone convergence theorem in (18),  $\int_S f_n d\mu \rightarrow \int_S f d\mu$  as  $n \rightarrow \infty$ . These two facts can be used to establish other properties of the integral of a nonnegative function based on our knowledge that the properties hold for simple functions. This type of argument is known as *bootstrapping*. We use bootstrapping to show that the [linearity properties](#) hold:

If  $f, g : S \rightarrow [0, \infty)$  are measurable and  $c \in [0, \infty)$ , then

1.  $\int_S (f + g) d\mu = \int_S f d\mu + \int_S g d\mu$
2.  $\int_S cf d\mu = c \int_S f d\mu$

Proof

1. Let  $(f_1, f_2, \dots)$  and  $(g_1, g_2, \dots)$  be increasing sequences of nonnegative simple functions with  $f_n \rightarrow f$  and  $g_n \rightarrow g$  as  $n \rightarrow \infty$ . Then  $(f_1 + g_1, f_2 + g_2, \dots)$  is also an increasing sequence of simple functions, and  $f_n + g_n \rightarrow f + g$  as  $n \rightarrow \infty$ . By the monotone convergence theorem,  $\int_S f_n d\mu \rightarrow \int_S f d\mu$ ,  $\int_S g_n d\mu \rightarrow \int_S g d\mu$ , and  $\int_S (f_n + g_n) d\mu \rightarrow \int_S (f + g) d\mu$  as  $n \rightarrow \infty$ . But  $\int_S (f_n + g_n) d\mu = \int_S f_n d\mu + \int_S g_n d\mu$  for each  $n \in \mathbb{N}_+$  so taking limits gives  $\int_S (f + g) d\mu = \int_S f d\mu + \int_S g d\mu$ .
2. Similarly,  $(cf_1, cf_2, \dots)$  is an increasing sequence of nonnegative simple functions with  $cf_n \rightarrow cf$  as  $n \rightarrow \infty$ . Again, by the MCT,  $\int_S cf_n d\mu \rightarrow \int_S cf d\mu$  and  $\int_S cf_n d\mu = c \int_S f_n d\mu$  so taking limits gives  $\int_S cf d\mu = c \int_S f d\mu$ .

## General Functions

Our final step is to define the integral of a measurable function  $f : S \rightarrow \mathbb{R}$ . First, recall the positive and negative parts of  $x \in \mathbb{R}$ :

$$x^+ = \max\{x, 0\}, \quad x^- = \max\{-x, 0\} \quad (3.10.19)$$

Note that  $x^+ \geq 0$ ,  $x^- \geq 0$ ,  $x = x^+ - x^-$ , and  $|x| = x^+ + x^-$ . Given that we want the integral to have the linearity properties in (5), there is no question as to how we should define the integral of  $f$  in terms of the integrals of  $f^+$  and  $f^-$ , which being nonnegative, are defined by the previous subsection.

If  $f : S \rightarrow \mathbb{R}$  is measurable, we define

$$\int_S f d\mu = \int_S f^+ d\mu - \int_S f^- d\mu \quad (3.10.20)$$

assuming that at least one of the integrals on the right is finite. If both are finite, then  $f$  is said to be *integrable*.

Assuming that either the integral of the positive part or the integral of the negative part is finite ensures that we do not get the dreaded indeterminate form  $\infty - \infty$ .

Suppose that  $f : S \rightarrow \mathbb{R}$  is measurable. Then  $f$  is integrable if and only if  $\int_S |f| d\mu < \infty$ .

Proof

Suppose that  $f$  is integrable. Recall that  $|f| = f^+ + f^-$ . By the additive property for nonnegative functions,  $\int_S |f| d\mu = \int_S f^+ d\mu + \int_S f^- d\mu < \infty$ . Conversely, suppose that  $\int_S |f| d\mu < \infty$ . Then  $f^+ \leq |f|$  and  $f^- \leq |f|$  so by the increasing property for nonnegative functions,  $\int_S f^+ d\mu \leq \int_S |f| d\mu < \infty$  and  $\int_S f^- d\mu \leq \int_S |f| d\mu < \infty$ .

Note that if  $f$  is nonnegative, then our new definition agrees with our old one, since  $f^+ = f$  and  $f^- = 0$ . For simple functions the integral has the same basic form as for nonnegative simple functions:

Suppose that  $f$  is a simple function with the representation  $f = \sum_{i \in I} a_i \mathbf{1}_{A_i}$ . Then

$$\int_S f d\mu = \sum_{i \in I} a_i \mu(A_i) \quad (3.10.21)$$

assuming that the sum does not have both  $\infty$  and  $-\infty$  terms.

Proof

Note that  $f^+$  and  $f^-$  are also simple, with the representations  $f^+ = \sum_{i \in I} a_i^+ \mathbf{1}_{A_i}$  and  $f^- = \sum_{i \in I} a_i^- \mathbf{1}_{A_i}$ . Hence

$$\int_S f d\mu = \sum_{i \in I} a_i^+ \mu(A_i) - \sum_{i \in I} a_i^- \mu(A_i) \quad (3.10.22)$$

as long as one of the sums is finite. Given that this is the case, we can recombine the sums to get

$$\int_S f d\mu = \sum_{i \in I} a_i \mu(A_i) \quad (3.10.23)$$

Once again, we need to establish the essential properties. Our first result is an intermediate step towards linearity.

If  $f, g: S \rightarrow [0, \infty)$  are measurable then  $\int_S (f - g) d\mu = \int_S f d\mu - \int_S g d\mu$  as long as at least one of the integrals on the right is finite.

Proof

We take cases. Suppose first that  $\int_S f d\mu < \infty$  and  $\int_S g d\mu < \infty$ . Note that  $(f - g)^+ \leq f$  and  $(f - g)^- \leq g$ . By the increasing property for nonnegative functions,  $\int_S (f - g)^+ d\mu \leq \int_S f d\mu < \infty$  and  $\int_S (f - g)^- d\mu \leq \int_S g d\mu < \infty$ . Thus  $f - g$  is integrable. Next we have  $f - g = (f - g)^+ - (f - g)^-$  and therefore  $f + (f - g)^- = g + (f - g)^+$ . All four of the functions in the last equation are nonnegative, and therefore by additivity property for nonnegative functions, we have

$$\int_S f d\mu + \int_S (f - g)^- d\mu = \int_S g d\mu + \int_S (f - g)^+ d\mu \quad (3.10.24)$$

All of these integrals are finite, and hence

$$\int_S (f - g) d\mu = \int_S (f - g)^+ d\mu - \int_S (f - g)^- d\mu = \int_S f d\mu - \int_S g d\mu \quad (3.10.25)$$

Next suppose that  $\int_S f d\mu = \infty$  and  $\int_S g d\mu < \infty$ . Then  $f - g \leq (f - g)^+$  and hence  $f \leq (f - g)^+ + g$ . Using the additivity and increasing properties for nonnegative functions, we have  $\infty = \int_S f d\mu \leq \int_S (f - g)^+ d\mu + \int_S g d\mu$ . Since  $\int_S g d\mu < \infty$  we must have  $\int_S (f - g)^+ d\mu = \infty$ . On the other hand,  $(f - g)^- \leq g$  so  $\int_S (f - g)^- d\mu \leq \int_S g d\mu < \infty$ . Hence  $\int_S (f - g) d\mu = \infty = \int_S f d\mu - \int_S g d\mu$ .

Finally, suppose that  $\int_S f d\mu < \infty$  and  $\int_S g d\mu = \infty$ . By the argument in the last paragraph, we have  $\int_S (g - f)^+ d\mu = \infty$  and  $\int_S (g - f)^- d\mu < \infty$ . Equivalently,  $\int_S (f - g)^+ d\mu < \infty$  and  $\int_S (f - g)^- d\mu = \infty$ . Hence  $\int_S (f - g) d\mu = -\infty = \int_S f d\mu - \int_S g d\mu$ .

We finally have the [linearity properties](#) in full generality.

If  $f, g: S \rightarrow \mathbb{R}$  are measurable functions whose integrals exist, and  $c \in \mathbb{R}$ , then

1.  $\int_S (f + g) d\mu = \int_S f d\mu + \int_S g d\mu$  as long as the right side is not of the form  $\infty - \infty$ .
2.  $\int_S cf d\mu = c \int_S f d\mu$

Proof

1. Note that  $f + g = (f^+ - f^-) + (g^+ - g^-) = (f^+ + g^+) - (f^- + g^-)$  and the two functions in parentheses in the last expression are nonnegative. By the previous lemma and the additivity property for nonnegative functions, we have

$$\int_S (f + g) d\mu = \int_S (f^+ + g^+) d\mu - \int_S (f^- + g^-) d\mu \quad (3.10.26)$$

$$= \left( \int_S f^+ d\mu + \int_S g^+ d\mu \right) - \left( \int_S f^- d\mu + \int_S g^- d\mu \right) \quad (3.10.27)$$

assuming that either both integrals in the first parentheses are finite or both integrals in the second parentheses are finite. In either case, we can group the terms (without worrying about the dreaded  $\infty - \infty$ ) to get



$$\int_S (f+g) d\mu = \left( \int_S f^+ d\mu - \int_S f^- d\mu \right) + \left( \int_S g^+ d\mu - \int_S g^- d\mu \right) = \int_S f d\mu + \int_S g d\mu \quad (3.10.28)$$

2. Note that if  $c \geq 0$  then  $(cf)^+ = cf^+$  and  $(cf)^- = cf^-$ . Hence using the scaling property for nonnegative functions,

$$\int_S cf d\mu = \int_S (cf)^+ d\mu - \int_S (cf)^- d\mu = \int_S cf^+ d\mu - \int_S cf^- d\mu = c \int_S f^+ d\mu - c \int_S f^- d\mu = c \int_S f d\mu \quad (3.10.29)$$

On the other hand, if  $c < 0$ ,  $(cf)^+ = -cf^-$  and  $(cf)^- = -cf^+$ . Again using the scaling property for nonnegative functions,

$$\int_S cf d\mu = \int_S (cf)^+ d\mu - \int_S (cf)^- d\mu = \int_S -cf^- d\mu - \int_S -cf^+ d\mu = -c \int_S f^- d\mu + c \int_S f^+ d\mu = c \int_S f d\mu \quad (3.10.30)$$

In particular, note that if  $f$  and  $g$  are integrable, then so are  $f+g$  and  $cf$  for  $c \in \mathbb{R}$ . Thus, the set of integrable functions on  $(S, \mathcal{S}, \mu)$  forms a vector space, which is denoted  $\mathcal{L}(S, \mathcal{S}, \mu)$ . The  $\mathcal{L}$  is in honor of Henri Lebesgue, who first developed the theory. This vector space, and other related ones, will be studied in more detail in the section on function spaces.

We also have the [increasing property](#) in full generality.

If  $f, g : S \rightarrow \mathbb{R}$  are measurable functions whose integrals exist, and if  $f \leq g$  on  $S$  then  $\int_S f d\mu \leq \int_S g d\mu$

Proof

We can use the proof based on the additive property from (6). First  $g = f + (g - f)$  and  $g - f \geq 0$  on  $S$ . If  $\int_S f d\mu = -\infty$  then trivially,  $\int_S f d\mu \leq \int_S g d\mu$ . Otherwise  $\int_S (g - f) d\mu \geq 0$  and therefore  $\int_S g d\mu = \int_S f d\mu + \int_S (g - f) d\mu \geq \int_S f d\mu$ .

### The Integral Over a Set

Now that we have defined the integral of a measurable function  $f$  over all of  $S$ , there is a natural extension to the integral of  $f$  over a measurable subset

If  $f : S \rightarrow \mathbb{R}$  is measurable and  $A \in \mathcal{S}$ , we define

$$\int_A f d\mu = \int_S \mathbf{1}_A f d\mu \quad (3.10.31)$$

assuming that the integral on the right exists.

If  $f : S \rightarrow \mathbb{R}$  is a measurable function whose integral exists and  $A \in \mathcal{S}$ , then the integral of  $f$  over  $A$  exists.

Proof

Note that  $(\mathbf{1}_A f)^+ = \mathbf{1}_A f^+$  and  $(\mathbf{1}_A f)^- = \mathbf{1}_A f^-$ . Also  $\mathbf{1}_A f^+ \leq f^+$  and  $\mathbf{1}_A f^- \leq f^-$ . If  $\int_S f d\mu$  exists, then either  $\int_S f^+ d\mu < \infty$  or  $\int_S f^- d\mu < \infty$ . By the increasing property, it follows that either  $\int_S \mathbf{1}_A f^+ d\mu < \infty$  or  $\int_S \mathbf{1}_A f^- d\mu < \infty$ , so  $\int_A f d\mu$  exists.

On the other hand, it's clearly possible for  $\int_A f d\mu$  to exist for some  $A \in \mathcal{S}$ , but not  $\int_S f d\mu$ .

We could also simply think of  $\int_A f d\mu$  as the integral of a measurable function  $f : A \rightarrow \mathbb{R}$  over the measure space  $(A, \mathcal{S}_A, \mu_A)$ , where  $\mathcal{S}_A = \{B \in \mathcal{S} : B \subseteq A\} = \{C \cap A : C \in \mathcal{S}\}$  is the  $\sigma$ -algebra of measurable subsets of  $A$ , and where  $\mu_A$  is the restriction of  $\mu$  to  $\mathcal{S}_A$ . It follows that all of the essential properties hold for integrals over  $A$ : the [linearity properties](#), the [order properties](#), and the [monotone convergence theorem](#). The following property is a simple consequence of the general additive property, and is known as *additive property for disjoint domains*.

Suppose that  $f : S \rightarrow \mathbb{R}$  is a measurable function whose integral exists, and that  $A, B \in \mathcal{S}$  are disjoint. then

$$\int_{A \cup B} f d\mu = \int_A f d\mu + \int_B f d\mu \quad (3.10.32)$$

Proof

Recall that  $\mathbf{1}_{A \cup B} = \mathbf{1}_A + \mathbf{1}_B$ . Hence by the additive property and the previous result,

$$\int_{A \cup B} f d\mu = \int_S \mathbf{1}_{A \cup B} f d\mu = \int_S (\mathbf{1}_A f + \mathbf{1}_B f) d\mu = \int_S \mathbf{1}_A f d\mu + \int_S \mathbf{1}_B f d\mu = \int_A f d\mu + \int_B f d\mu \quad (3.10.33)$$

By induction, the additive property holds for a finite collection of disjoint domains. The extension to a countably infinite collection of disjoint domains will be considered in the next section on properties of the integral.



## Special Cases

### Discrete Spaces

Recall again that the measure space  $(S, \mathcal{S}, \#)$  is *discrete* if  $S$  is countable,  $\mathcal{S}$  is the collection of all subsets of  $S$ , and  $\#$  is counting measure on  $\mathcal{S}$ . Thus all functions  $f : S \rightarrow \mathbb{R}$  are measurable, and as we will see, integrals with respect to  $\#$  are simply sums.

If  $f : S \rightarrow \mathbb{R}$  then

$$\int_S f d\# = \sum_{x \in S} f(x) \quad (3.10.34)$$

as long as either the sum of the positive terms or the sum of the negative terms is finite.

Proof

The proof is a bootstrapping argument.

1. Suppose first that  $S$  is finite. In this case, every function  $f : S \rightarrow \mathbb{R}$  is simple and has the representation  $f = \sum_{x \in S} f(x) \mathbf{1}_x$  where  $\mathbf{1}_x$  is an abbreviation of  $\mathbf{1}_{\{x\}}$ . Thus the result follows from the definition of the integral.
2. Next suppose that  $S$  is countable infinite and  $f : S \rightarrow [0, \infty)$ . Let  $(A_1, A_2, \dots)$  be an increasing sequence of finite subsets of  $S$  with  $\bigcup_{i=1}^{\infty} A_i = S$ . Define  $f_n = \sum_{x \in A_n} f(x) \mathbf{1}_x$ . Then  $(f_1, f_2, \dots)$  is an increasing sequence of simple functions with  $f_n \rightarrow f$  as  $n \rightarrow \infty$ . Thus

$$\int_S f d\# = \lim_{n \rightarrow \infty} \int_S f_n d\# = \lim_{n \rightarrow \infty} \sum_{x \in A_n} f(x) \quad (3.10.35)$$

But by definition, the last limit on the right is just  $\sum_{x \in S} f(x)$ .

3. Finally consider the general case where  $S$  is countable and  $f : S \rightarrow \mathbb{R}$ . In this case the result follows from the definition of the integral as  $\int_S f d\# = \int_S f^+ d\# - \int_S f^- d\#$  as long as one of the integrals on the right is finite. By (b),  $\int_S f^+ d\#$  is the sum of the positive terms and  $-\int_S f^- d\#$  is the sum of the negative terms.

If the sum of the positive terms and the sum of the negative terms are both finite, then  $f$  is integrable with respect to  $\#$ , but the usual term from calculus is that the series  $\sum_{x \in S} f(x)$  is *absolutely convergent*. The result will look more familiar in the special case  $S = \mathbb{N}_+$ . Functions on  $S$  are simply sequences, so we can use the more familiar notation  $a_i$  rather than  $a(i)$  for a function  $a : S \rightarrow \mathbb{R}$ . Part (b) of the proof (with  $A_n = \{1, 2, \dots, n\}$ ) is just the definition of an infinite series of nonnegative terms as the limit of the partial sums:

$$\sum_{i=1}^{\infty} a_i = \lim_{n \rightarrow \infty} \sum_{i=1}^n a_i \quad (3.10.36)$$

Part (c) of the proof is just the definition of a general infinite series

$$\sum_{i=1}^{\infty} a_i = \sum_{i=1}^{\infty} a_i^+ - \sum_{i=1}^{\infty} a_i^- \quad (3.10.37)$$

as long as one of the series on the right is finite. Again, when both are finite, the series is absolutely convergent. In calculus we also consider *conditionally convergent* series. This means that  $\sum_{i=1}^{\infty} a_i^+ = \infty$ ,  $\sum_{i=1}^{\infty} a_i^- = \infty$ , but  $\lim_{n \rightarrow \infty} \sum_{i=1}^n a_i$  exists in  $\mathbb{R}$ . Such series have no place in general integration theory. Also, you may recall that such series are pathological in the sense that, given any number in  $\mathbb{R}^*$ , there exists a rearrangement of the terms so that the rearranged series converges to the given number.

### The Lebesgue and Riemann Integrals on $\mathbb{R}$

Consider the one-dimensional Euclidean space  $(\mathbb{R}, \mathcal{R}, \lambda)$  where  $\mathcal{R}$  is the usual  $\sigma$ -algebra of Lebesgue measurable sets and  $\lambda$  is Lebesgue measure. The theory developed above applies, of course, for the integral  $\int_A f d\mu$  of a measurable function  $f : \mathbb{R} \rightarrow \mathbb{R}$  over a set  $A \in \mathcal{R}$ . It's not surprising that in this special case, the theory of integration is referred to as *Lebesgue integration* in honor of our good friend Henri Lebesgue, who first developed the theory.

On the other hand, we already have a theory of integration on  $\mathbb{R}$ , namely the Riemann integral of calculus, named for our other good friend Georg Riemann. For a suitable function  $f$  and domain  $A$  this integral is denoted  $\int_A f(x) dx$ , as we all remember from calculus. How are the two integrals related? As we will see, the Lebesgue integral generalizes the Riemann integral.

To understand the connection we need to review the definition of the Riemann integral. Consider first the standard case where the domain of integration is a closed, bounded interval. Here are the preliminary definitions that we will need.

Suppose that  $f : [a, b] \rightarrow \mathbb{R}$ , where  $a, b \in \mathbb{R}$  and  $a < b$ .

1. A *partition*  $\mathcal{A} = \{A_i : i \in I\}$  of  $[a, b]$  is a finite collection of disjoint *subintervals* whose union is  $[a, b]$ .

2. The *norm* of a partition  $\mathcal{A}$  is  $\|\mathcal{A}\| = \max\{\lambda(A_i) : i \in I\}$ , the length of the largest subinterval of  $\mathcal{A}$ .
3. A set of points  $B = \{x_i : i \in I\}$  where  $x_i \in A_i$  for each  $i \in I$  is said to be *associated* with the partition  $\mathcal{A}$ .
4. The *Riemann sum* of  $f$  corresponding to a partition  $\mathcal{A}$  and a set  $B$  associated with  $\mathcal{A}$  is

$$R(f, \mathcal{A}, B) = \sum_{i \in I} f(x_i) \lambda(A_i) \quad (3.10.38)$$

Note that the Riemann sum is simply the integral of the simple function  $g = \sum_{i \in I} f(x_i) \mathbf{1}_{A_i}$ . Moreover, since  $A_i$  is an interval for each  $i \in I$ ,  $g$  is a *step function*, since it is constant on a finite collection of disjoint intervals. Moreover, again since  $A_i$  is an interval for each  $i \in I$ ,  $\lambda(A_i)$  is simply the length of the subinterval  $A_i$ , so of course measure theory per se is not needed for Riemann integration. Now for the definition from calculus:

$f$  is *Riemann integrable* on  $[a, b]$  if there exists  $r \in \mathbb{R}$  with the property that for every  $\epsilon > 0$  there exists  $\delta > 0$  such that if  $\mathcal{A}$  is a partition of  $[a, b]$  with  $\|\mathcal{A}\| < \delta$  then  $|r - R(f, \mathcal{A}, B)| < \epsilon$  for every set of points  $B$  associated with  $\mathcal{A}$ . Then of course we define the integral by

$$\int_a^b f(x) dx = r \quad (3.10.39)$$

Here is our main theorem of this subsection.

If  $f : [a, b] \rightarrow \mathbb{R}$  is Riemann integrable on  $[a, b]$  then  $f$  is Lebesgue integrable on  $[a, b]$  and

$$\int_{[a,b]} f d\lambda = \int_a^b f(x) dx \quad (3.10.40)$$

On the other hand, there are lots of functions that are Lebesgue integrable but not Riemann integrable. In fact there are indicator functions of this type, the simplest of functions from the point of view of Lebesgue integration.

Consider the function  $\mathbf{1}_{\mathbb{Q}}$  where as usual,  $\mathbb{Q}$  is the set of rational number in  $\mathbb{R}$ . Then

1.  $\int_{\mathbb{R}} \mathbf{1}_{\mathbb{Q}} d\lambda = 0$ .
2.  $\mathbf{1}_{\mathbb{Q}}$  is not Riemann integrable on any interval  $[a, b]$  with  $a < b$ .

*Proof*

Part (a) follows from the definition of the Lebesgue integral:

$$\int_{\mathbb{R}} \mathbf{1}_{\mathbb{Q}} d\lambda = \lambda(\mathbb{Q}) = 0 \quad (3.10.41)$$

For part (b), note that there are rational and irrational numbers in every interval of  $\mathbb{R}$  of positive length (the rational numbers and the irrational numbers are *dense* in  $\mathbb{R}$ ). Thus, given any partition  $\mathcal{A} = \{A_i : i \in I\}$  of  $[a, b]$ , no matter how small the norm, there are Riemann sums that are 0 (take  $x_i \in A_i$  irrational for each  $i \in I$ ), and Riemann sums that are  $b - a$  (take  $x_i \in A_i$  rational for each  $i \in I$ ).

The following fundamental theorem completes the picture.

$f : [a, b] \rightarrow \mathbb{R}$  is Riemann integrable on  $[a, b]$  if and only if  $f$  is bounded on  $[a, b]$  and  $f$  is continuous almost everywhere on  $[a, b]$ .

Now that the Riemann integral is defined for a closed bounded interval, it can be extended to other domains.

**Extensions of the Riemann integral.**

1. If  $f$  is defined on  $[a, b)$  and Riemann integrable on  $[a, t]$  for  $a < t < b$ , we define  $\int_a^b f(x) dx = \lim_{t \uparrow b} \int_a^t f(x) dx$  if the limit exists in  $\mathbb{R}^*$ .
2. If  $f$  is defined on  $(a, b]$  and Riemann integrable on  $[t, b]$  for  $a < t < b$ , we define  $\int_a^b f(x) dx = \lim_{t \downarrow a} \int_t^b f(x) dx$  if the limit exists in  $\mathbb{R}^*$ .
3. If  $f$  is defined on  $(a, b)$ , we select  $c \in (a, b)$  and define  $\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx$  if the integrals on the right exist in  $\mathbb{R}^*$  by (a) and (b), and are not of the form  $\infty - \infty$ .
4. If  $f$  is defined on  $[a, \infty)$  and Riemann integrable on  $[a, t]$  for  $a < t < \infty$  we define  $\int_a^\infty f(x) dx = \lim_{t \rightarrow \infty} \int_a^t f(x) dx$ .
5. If  $f$  is defined on  $(-\infty, b]$  and Riemann integrable on  $[t, b]$  for  $-\infty < t < b$  we define  $\int_{-\infty}^b f(x) dx = \lim_{t \rightarrow -\infty} \int_t^b f(x) dx$  if the limit exists in  $\mathbb{R}^*$ .
6. If  $f$  is defined on  $\mathbb{R}$  we select  $c \in \mathbb{R}$  and define  $\int_{-\infty}^\infty f(x) dx = \int_{-\infty}^c f(x) dx + \int_c^\infty f(x) dx$  if both integrals on the right exist by (d) and (e), and are not of the form  $\infty - \infty$ .

7. The integral is defined for a domain that is the union of a finite collection of disjoint intervals by the requirement that the integral be additive over disjoint domains

As another indication of its superiority, note that *none* of these convolutions is necessary for the Lebesgue integral. Once and for all, we have defined  $\int_A f(x) dx$  for a general measurable function  $f : \mathbb{R} \rightarrow \mathbb{R}$  and a general domain  $A \in \mathcal{R}$

### The Lebesgue-Stieltjes Integral

Consider again the measurable space  $(\mathbb{R}, \mathcal{R})$  where  $\mathcal{R}$  is the usual  $\sigma$ -algebra of Lebesgue measurable subsets of  $\mathbb{R}$ . Suppose that  $F : \mathbb{R} \rightarrow \mathbb{R}$  is a general distribution function, so that by definition,  $F$  is increasing and continuous from the right. Recall that the Lebesgue-Stieltjes measure  $\mu$  associated with  $F$  is the unique measure on  $\mathcal{R}$  that satisfies

$$\mu(a, b] = F(b) - F(a); \quad a, b \in \mathbb{R}, \quad a < b \quad (3.10.42)$$

Recall that  $F$  satisfies some, but not necessarily all of the properties of a probability distribution function. The properties not necessarily satisfied are the normalizing properties

- $F(x) \rightarrow 0$  as  $x \rightarrow -\infty$
- $F(x) \rightarrow 1$  as  $x \rightarrow \infty$

If  $F$  does satisfy these two additional properties, then  $\mu$  is a probability measure and  $F$  its probability distribution function.

The integral with respect to the measure  $\mu$  is, appropriately enough, referred to as the *Lebesgue-Stieltjes* integral with respect to  $F$ , and like the measure, is named for the ubiquitous Henri Lebesgue and for Thomas Stieltjes. In addition to our usual notation  $\int_S f d\mu$ , the Lebesgue-Stieltjes integral is also denoted  $\int_S f dF$  and  $\int_S f(x) dF(x)$ .

### Probability Spaces

Suppose that  $(S, \mathcal{S}, \mathbb{P})$  is a probability space, so that  $S$  is the set of outcomes of a random experiment,  $\mathcal{S}$  is the  $\sigma$ -algebra of events, and  $\mathbb{P}$  the probability measure on the sample space  $(S, \mathcal{S})$ . A measurable, real-valued function  $X$  on  $S$  is, of course, a real-valued random variable. The integral with respect to  $\mathbb{P}$ , if it exists, is the *expected value* of  $X$  and is denoted

$$\mathbb{E}(X) = \int_S X d\mathbb{P} \quad (3.10.43)$$

This concept is of fundamental importance in probability theory and is studied in detail in a separate chapter on Expected Value, mostly from an elementary point of view that does not involve abstract integration. However an advanced section treats expected value as an integral over the underlying probability measure, as above.

Suppose next that  $(T, \mathcal{T}, \#)$  is a discrete space and that  $X$  is a random variable for the experiment, taking values in  $T$ . In this case  $X$  has a discrete distribution and the probability density function  $f$  of  $X$  is given by  $f(x) = \mathbb{P}(X = x)$  for  $x \in T$ . More generally,

$$\mathbb{P}(X \in A) = \sum_{x \in A} f(x) = \int_A f d\#, \quad A \subseteq T \quad (3.10.44)$$

On the other hand, suppose that  $X$  is a random variable with values in  $\mathbb{R}^n$ , where as usual,  $(\mathbb{R}^n, \mathcal{R}_n, \lambda_n)$  is  $n$ -dimensional Euclidean space. If  $X$  has a continuous distribution, then  $f : T \rightarrow [0, \infty)$  is a probability density function of  $X$  if

$$\mathbb{P}(X \in A) = \int_A f d\lambda_n, \quad A \in \mathcal{R}^n \quad (3.10.45)$$

Technically,  $f$  is the density function of  $X$  with respect to counting measure  $\#$  in the discrete case, and  $f$  is the density function of  $X$  with respect to Lebesgue measure  $\lambda_n$  in the continuous case. In both cases, the probability of an event  $A$  is computed by integrating the density function, with respect to the appropriate measure, over  $A$ . There are still differences, however. In the discrete case, the existence of the density function with respect to counting measure is guaranteed, and indeed we have an explicit formula for it. In the continuous case, the existence of a density function with respect to Lebesgue measure is not guaranteed, and indeed there might not be one. More generally, suppose that we have a measure space  $(T, \mathcal{T}, \mu)$  and a random variable  $X$  with values in  $T$ . A measurable function  $f : T \rightarrow [0, \infty)$  is a *probability density function* of  $X$  (or more precisely, the *distribution* of  $X$ ) with respect to  $\mu$  if

$$\mathbb{P}(X \in A) = \int_A f d\mu, \quad A \in \mathcal{T} \quad (3.10.46)$$

This fundamental question of the existence of a density function will be clarified in the section on absolute continuity and density functions.

Suppose again that  $X$  is a real-valued random variable with distribution function  $F$ . Then, by definition, the distribution of  $X$  is the Lebesgue-Stieltjes measure associated with  $F$ :

$$\mathbb{P}(a < X \leq b) = F(b) - F(a), \quad a, b \in \mathbb{R}, \quad a < b \quad (3.10.47)$$

regardless of whether the distribution is discrete, continuous, or mixed. Trivially,  $\mathbb{P}(X \in A) = \int_S \mathbf{1}_A dF$  for  $A \in \mathcal{R}$  and the expected value of  $X$  defined above can also be written as  $\mathbb{E}(X) = \int_{\mathbb{R}} x dF(x)$ . Again, all of this will be explained in much more detail in the next chapter on Expected Value.

### Computational Exercises

Let  $g(x) = \frac{1}{1+x^2}$  for  $x \in \mathbb{R}$ .

1. Find  $\int_{-\infty}^{\infty} g(x) dx$ .
2. Show that  $\int_{-\infty}^{\infty} xg(x) dx$  does not exist.

Answer

1.  $\int_{-\infty}^{\infty} g(x) dx = \pi$
2.  $\int_0^{\infty} xg(x) dx = \infty$ ,  $\int_{-\infty}^0 xg(x) dx = -\infty$

You may recall that the function  $g$  in the last exercise is important in the study of the *Cauchy distribution*, named for Augustin Cauchy. You may also remember that the graph of  $g$  is known as the *witch of Agnesi*, named for Maria Agnesi.

Let  $g(x) = \frac{1}{x^b}$  for  $x \in [1, \infty)$  where  $b > 0$  is a parameter. Find  $\int_1^{\infty} g(x) dx$

Answer

$$\int_1^{\infty} g(x) dx = \begin{cases} \infty, & 0 < b \leq 1 \\ \frac{1}{b-1}, & b > 1 \end{cases}$$

You may recall that the function  $g$  in the last exercise is important in the study of the *Pareto distribution*, named for Vilfredo Pareto.

Suppose that  $f(x) = 0$  if  $x \in \mathbb{Q}$  and  $f(x) = \sin(x)$  if  $x \in \mathbb{R} - \mathbb{Q}$ .

1. Find  $\int_{[0, \pi]} f(x) d\lambda(x)$
2. Does  $\int_0^{\pi} f(x) dx$  exist?

Answer

1. 2
2. No

This page titled [3.10: The Integral With Respect to a Measure](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist](#) ([Random Services](#)) via [source content](#) that was edited to the style and standards of the LibreTexts platform.