Introduction to Measure and Integration

S. Kesavan

The Institute of Mathematical Sciences
CIT Campus, Taramani
Chennai - 600 113.

e mail: kesh@imsc.res.in

Abstract

In this lecture, we will look at the important properties of the Riemann integral and also study its drawbacks, thus motivating the need for a more general notion of the integral. This will lead us to the definition of the Lebesgue integral and its important results will be surveyed.

1 Introduction

Form the time of the Greeks, the problem of computing the area enclosed by a curve has been exercising the minds of scientific thinkers. This crucial question, at the base of the theory of integral calculus, was treated as early as the third century B.C. by Archimedes, who calculated the area of a circular disc, the area of a segment of a parabola and other such figures. He used the 'method of exhaustion'. The basic idea was to exhaust the given area by a sequence of polygonal domains and calculate the area as the limit of the area of the inscribed polygons. During the seventeenth century, many such areas were calculated and in each case the problem was solved by an ingenious device specially suited for the case in hand. One of the achievements of calculus was to develop a general and powerful method to replace these special restricted procedures.

From the time of Archimedes until the time of Gauss, the attitude was that the area was an intuitively obvious entity which need not be *defined*, but which had to be *computed*. Lebesgue, whose theory of measure and integration we are trying here to motivate, describes the situation as follows. Before Cauchy, there was no definition of the integral in the precise sense of the term. One was often limited to saying which areas one had to add, or subtract, to get the integral.

Cauchy, with his concern for rigour, which is characteristic of modern mathematics, defined continuous functions and their integrals in much the same way as we do now. To arrive at the integral of a continuous function f defined on an interval $[a, b]$ of the real line, he looked at sums of the form

$$S = \sum_i f(\xi_i)(x_{i+1} - x_i)$$

where $a \leq x_0 < x_1 < ... < x_i < x_{i+1} < ... < x_N = b$ is a partition of $[a, b]$ and $\xi_i \in [x_i, x_{i+1}]$. He then deduced the value of the integral

$$\int_a^b f(x)dx$$

by a suitable passage to the limit. For a long time, certain discontinuous functions were integrated; Cauchy's definition still applied to these integrals, but it was natural to investigate, as Riemann did, the exact scope of this definition.

2 The Riemann Integral

Let $[a, b] \subset \mathbb{R}$ be a finite interval and let $f : [a, b] \to \mathbb{R}$ be a bounded function. Let $\mathcal{P} = \{a = x_0 < x_1 < ... < x_N = b\}$ be a partition of the interval. Set

$$m_i \ = \ \inf_{x \in [x_{i-1}, x_i]} f(x) \ \text{ and } \ M_i \ = \ \sup_{x \in [x_{i-1}, x_i]} f(x).$$

Then, we define the lower and upper (Darboux) sums associated to the function f and the partition $\mathcal{P}$ by

$$\mathcal{L}(\mathcal{P}, f) \ = \ \textstyle\sum_{i=1}^{N} m_i (x_i - x_{i-1})$$

$$\mathcal{U}(\mathcal{P}, f) \ = \ \textstyle\sum_{i=1}^{N} M_i (x_i - x_{i-1}).$$

Then, we define the lower and upper integrals of f by

$$\textstyle\int_{\underline{a}}^{b} f(x) dx \ = \ \sup_{\mathcal{P}} \mathcal{L}(\mathcal{P}, f)$$

$$\textstyle\int_{a}^{\bar{b}} f(x) dx \ = \ \inf_{\mathcal{P}} \mathcal{U}(\mathcal{P}, f)$$

where the supremum and infimum are taken over all possible partititions of $[a, b]$. The function f is said to be **Riemann integrable** over $[a, b]$ if its lower and upper integrals are equal and the common value, called the **Riemann integral** of f over $[a, b]$ is denoted by the symbol

$$\int_a^b f(x) dx.$$

Remark 2.1 Since f is bounded, we have $m \leq f(x) \leq M$ for all $x \in [a, b]$ and it is immediate to see that

$$m(b - a) \ \leq \ \mathcal{L}(\mathcal{P}, f) \ \leq \ \mathcal{U}(\mathcal{P}, f) \ \leq \ M(b - a)$$

for all partitions $\mathcal{P}$. Thus, the lower and upper integrals of f always exist but the question of their being equal is a delicate one. ■

Given a partition $\mathcal{P}$ as above, we set

$$\mu(\mathcal{P}) \ = \ \max_{1 \leq i \leq N} (x_i - x_{i-1}).$$

Let $t_i \in [x_{i-1}, x_i]$ for $1 \leq i \leq N$. Denote

$$S(\mathcal{P}, f) \;=\; \sum_{i=1}^{N} f(t_i)(x_i - x_{i-1}).$$

Remark 2.2 The above notation is incomplete. The sum $S(\mathcal{P}, f)$ depends not only on the partition $\mathcal{P}$ and the function f, but also on the choice of the points t_i. But in order to avoid cumbersome notation, we will leave it as it is. ■

Definition 2.1 *We say that*

$$\lim_{\mu(\mathcal{P})\to 0} S(\mathcal{P}, f) \;=\; A$$

if, for every $\varepsilon > 0$, there exists a $\delta > 0$ such that, for all partitions $\mathcal{P}$ such that $\mu(\mathcal{P}) < \delta$, and for all choices of points t_i compatible with the partition, we have

$$|S(\mathcal{P}, f) - A| \;<\varepsilon.\blacksquare$$

Theorem 2.1 *The function f is Riemann integrable, if and only if, the limit defined in the above definition exists and, in this case,*

$$\int_a^b f(x)dx \;=\; \lim_{\mu(\mathcal{P})\to 0} S(\mathcal{P}, f).\blacksquare$$

Thus, we see that the requirement that a function be Riemann integrable is a very strong one. We have the following result.

Theorem 2.2 *If f is continuous, or if f has at most a countable number of discontinuities, then f is Riemann integrable.* ■

Example 2.1 Let us consider the unit interval $[0, 1]$. Let us choose some numbering of all the rational numbers in this interval and write them as $r_1, r_2, ...$ Define

$$f_n(x) \;=\; \begin{cases} 1, & \text{if } \; x = r_1, r_2, ..., r_n \\ 0, & \text{otherwise.} \end{cases}$$

The function f_n is discontinuous only at the points $r_1, ..., r_n$ which are finite in number and so, by the previous theorem, f_n is Riemann integrable. In fact, it is a simple exercise to check this fact directly using the definition of Riemann integrability and show that the integral is equal to zero.

Let us now consider the function $f(x) = \lim_{n\to\infty} f_n(x)$. It is easy to see that

$$f(x) \;=\; \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational.} \end{cases}$$

This function is discontinuous everywhere. Given any partition $\mathcal{P}$, it is easy to see that $m_i = 0$ and that $M_i = 1$ for all $1 \le i \le N$. Thus $\mathcal{L}(\mathcal{P}, f) = 0$ and $\mathcal{U}(\mathcal{P}, f) = 1$. Thus the lower integral is zero while the upper integral is unity and so f fails to be Riemann integrable. ■

This brings us to a major drawback of the Riemann integral. The limit of a sequence of Riemann integrable functions need not be Riemann integrable. Even if the limit is a Riemann integrable function, the limit of the integrals need not be the integral of the limit, as the following example shows.

Example 2.2 Let $f_n(x) = n^2x(1-x^2)^n$ for $x \in [0,1]$. Then $f_n(x) \to 0$ as $n \to \infty$ (why?). Now,

$$\int_0^1 x(1-x^2)^n dx \;=\; \frac{1}{2n+2}.$$

Thus,

$$\int_0^1 f_n(x)dx \;=\; \frac{n^2}{2n+2} \;\to\; \infty$$

while, since $f \equiv 0$, we have $\int_0^1 f(x)dx = 0$. Similarly, if we define

$$f_n(x) \;=\; nx(1-x^2)^n,$$

again $f_n \to f \equiv 0$ pointwise but $\int_0^1 f_n(x)dx \to 1/2 \neq 0$.■

So, when do the two limit processes - the pointwise limit of functions and Riemann integration (which has been defined as a limit of sums as shown in Theorem 2.1) - commute?

Definition 2.2 *We say that* $f_n \to f$ **uniformly** *on* $[a,b]$ *if, for every* $\varepsilon > 0$, *there exists a positive integer* N *such that, for all* $x \in [a,b]$ *and for all* $n \ge N$, *we have*

$$|f_n(x) - f(x)| \;<\; \varepsilon.$$■

Theorem 2.3 *If $f_n \to f$ uniformly on $[a,b]$, and if all the f_n are Riemann integrable, then f is Riemann integrable and, further,*

$$\lim_{n\to\infty} \int_a^b f_n(x)dx \;=\; \int_a^b f(x)dx. \blacksquare$$

In the preceding example, the sequence $\{f_n\}$ failed to converge uniformly. In fact, the non-commutativity of the operations of taking the pointwise limit and the Riemann integral is a useful test to prove that a sequence of functions is not uniformly convergent.

Thus, a sequence of functions which does not converge uniformly may converge to a function which is not integrable or it can happen that the limit function is Riemann integrable but the limit of the integrals is not the integral of the limit function. But uniform convergence is a very strong condition as well.

We thus feel the need for a theory of integration, wherein a larger class of functions is integrable and such that the process of taking pointwise limits of functions commutes with the process of integration under fairly easily verifiable hypotheses. This is where the alternative approach of Lebesgue comes in useful.

The way the Riemann integral is defined, a certain amount of continuity is forced on integrable functions. As we saw in Theorem 2.1, if f is Riemann integrable, then, for all admissible choices of points t_i, the value of $S(\mathcal{P}, f)$ cannot vary too much, since the limit exists as $\mu(\mathcal{P}) \to 0$. Thus, nearby points must have nearby values 'to a large extent' and this is what Theorem 2.2 is all about. We can excuse a countable number of discontinuities. But the function which takes the value 1 on the rationals and the value 0 on irrationals is discontinuous everywhere and it fails to be Riemann integrable.

The idea of Riemann in formulating the definition of the integral is to consider the function following the abcissa. We take the values of the function as we proceed along the x-axis. Thus, we are forced to consider and compare the values of the function at nearby points and hence we are dependent on some amount of continuity.

The idea of Lebesgue is to work, not from the domain, but from the range of a function. We take a particular value and consider the set of all points where this value is assumed when we define the integral. Let us illustrate this via an example.

Example 2.3 Let $\mathcal{P}$ be a partition of the interval $[a, b]$. Let

$$f(x) \;=\; \sum_{i=1}^{N} \alpha_i \chi_{E_i}(x)$$

where $E_i = [x_{i-1}, x_i]$ and for any subset A of $\mathbb{R}$,

$$\chi_A(x) \;=\; \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A. \end{cases}$$

This function has a finite number of discontinuities and the Riemann integral is easily seen (Exercise!) to be

$$\int_a^b f(x)dx \;=\; \sum_{i=1}^{N} \alpha_i (x_i - x_{i-1}).$$

By Lebesgue's method, we will be looking at sets of the form $E_\alpha = \{x \in [a, b] \mid f(x) = \alpha\}$ for each $\alpha \in \mathbb{R}$ and multiply α by the 'length' of the set E_α and 'add' all these products. In our example, $E_\alpha = \emptyset$ if $\alpha \neq \alpha_i$ for any $1 \leq i \leq N$ and $E_{\alpha_i} = E_i$. Thus the (Lebesgue) integral is given again by the same expression as the Riemann integral, in this case. ■

Remark 2.3 Imagine a merchant in a shop wanting to add all the money he has collected from sales during a particular day. He has two methods. First, he can take the money one at a time from the till and add the amounts as he takes them out. The other is for him to sort out all the money according to each denomination, count the number of coins or notes in each denomination, multiply the number by the value of the denomination and add all these products. Both procedures will yield the same result, but the latter is more efficient, especially if it involves large quantities of money (have you seen how they count the Hundi collections in a large temple, say, Tirumala?). The approach of Riemann is like the first method where we 'take a function as it comes along the x-axis, while the approach of Lebesgue is like the second, where we sort it out according to the values in the range. Obviously, this does not say anything about the values of nearby points, and so, hopefully, will not depend on the continuity of the function.■

The Riemann integral approximates a function by another of the form

$$\sum_{i=1}^{N} f(t_i)\chi_{E_i}(x)$$

where $t_i \in E_i$ and $\mathcal{P} = \{E_i \mid 1 \leq i \leq N\}$ is a partition of $[a, b]$ and passes to the limit in sums of the form $S(\mathcal{P}, f)$.

The Lebesgue integral approximates a function by one of the form

$$\sum_{i=1}^{N} \alpha_i \chi_{A_i}(x)$$

where A_i, $1 \leq i \leq N$ are 'more general' sets than just intervals. It then defines the integral of the simpler function by

$$\sum_{i=1}^{N} \alpha_i \mu(A_i)$$

where $\mu(A)$ is the 'length' of the set A, and then passes to the limit suitably to get the integral of f.

Here is the catch. What do we mean by the 'length' of a set A which is not an interval. This brings us to the theory of measures which will generalize the notion of length (area or volume, in higher dimensions) to a fairly large class of sets.

3 Measures on Sets

In this section we will introduce the notion of a measure on an arbitrary set X.

Definition 3.1 *A collection $\mathcal{M}$ of subsets of a set X is said to be a σ-* **algebra** *on X if the following conditions are satisfied:*
(i) $X \in \mathcal{M}$.
(ii) If $A \in \mathcal{M}$, the $A^c \in \mathcal{M}$ where $A^c = X \backslash A$ is the complement of A in X.
(iii) If $A_i \in \mathcal{M}$ for $i = 1, 2, 3, ...,$ then $\cup_{i=1}^{\infty} A_i \in \mathcal{M}$. ■

The pair $(X, \mathcal{M})$ is then called a **measurable space** and the elements of $\mathcal{M}$ are called **measurable sets**.

Definition 3.2 *Let $(X, \mathcal{M})$ be a measurable space and let $F : X \to \mathbb{R}$ be a given function. It is said to be a* **measurable function** *if, for all $\alpha \in \mathbb{R}$, we have*

$$f^{-1}((\alpha, \infty)) \;=\; \{x \in X \mid f(x) > \alpha\} \;\in\; \mathcal{M}. \blacksquare$$

Proposition 3.1 *The following are equivalent:*
(i) $f^{-1}((\alpha, \infty)) \in \mathcal{M}$ for all $\alpha \in \mathbb{R}$.
(ii) $f^{-1}([\alpha, \infty)) \in \mathcal{M}$ for all $\alpha \in \mathbb{R}$.
(iii) $f^{-1}((-\infty, \alpha)) \in \mathcal{M}$ for all $\alpha \in \mathbb{R}$.
(iv) $f^{-1}((-\infty, \alpha]) \in \mathcal{M}$ for all $\alpha \in \mathbb{R}$.$\blacksquare$

Given any collection of subsets of a set X, there is the smallest σ-algebra contaiuning the collection and it is called the σ-algebra generated by the collection of subsets. If (X, τ) is a topological space, then the σ- algebra generated by the open sets is called the **Borel** σ-algebra and its members are called **Borel sets**.

Definition 3.3 *Let $(X, \mathcal{M})$ be a measurable space. A* **measure** *on X is a function $\mu : \mathcal{M} \to [0, \infty]$ such that if $\{A_i\}_{i=1}^{\infty}$ are elements of $\mathcal{M}$ and are pairwise disjoint, i.e. $A_i \cap A_j = \emptyset$ if $i \neq j$, then*

$$\mu\left(\cup_{i=1}^{\infty} A_i\right) \;=\; \sum_{i=1}^{\infty} \mu(A_i). \tag{3.1}$$

The triple $(X, \mathcal{M}, \mu)$ is called a **measure space**. $\blacksquare$

Remark 3.1 A measure is the generalization of the notion of length (or area or volume). We can define it on the measurable subsets only and not on all sets. The property embodied in equation (3.1) is called **countable additivity**. If we have a disjoint set of intervals, we can say that the 'length' of their union is the sum of the individual lengths. Countable additivity is the generalization of this property. In earlier attempts to generalize the notion of length the property (3.1) was restricted to the more obvious case of finite disjoint unions. This was called finite additivity. However, the discovery that countable additivity was essential is the key to the success of the theory of Lebesgue measure and integration. $\blacksquare$

On the real line, we can construct a σ-algebra $\mathcal{M}$ which contains all the Borel sets and also define a measure μ such that, for any interval of the form $I = (a, b)$ (or, $[a, b)$, $(a, b]$, $[a, b]$), where $-\infty \leq a < b \leq +\infty$, we have

$$\mu(I) \ = \ \text{length of } I.$$

It also has the following additional properties:
(i) (**Completeness**) If $E \in \mathcal{M}$ and $\mu(E) = 0$, then for any $F \subset E$, we have $F \in \mathcal{M}$ and, *a fortiori*, $\mu(F) = 0$.
(ii) (**Translation Invariance**) If $E \in \mathcal{M}$ and if $a \in \mathbb{R}$, then

$$a + E \ = \ \{a + x \mid x \in E\} \ \in \ \mathcal{M}$$

and, further, $\mu(a + E) = \mu(E)$.
(iii) (**Regularity**) If $E \in \mathcal{M}$ and if $\varepsilon > 0$, then, there exists an open set $V \supset E$ and a compact set $K \subset E$ such that

$$\mu(V \backslash K) \ < \ \varepsilon.$$

Further,

$$\begin{aligned} \mu(E) \ &= \ \inf\{\mu(V) \mid V \supset E, \ V \text{ open}\} \\ &= \ \sup\{\mu(K) \mid K \subset E, \ K \text{ compact}\}. \end{aligned}$$

The properties (i) - (iii) above define μ uniquely upto a multiplicative constant. If we set $\mu(I)$ to be the length of an interval I, in particular, $\mu([0, 1]) = 1$, then the measure is uniquely fixed and is called the **Lebesgue measure** on $\mathbb{R}$.

Remark 3.2 In the same way, we can define the Lebesgue measure on $\mathbb{R}^N$ with the properties (i) - (iii) above and such that, if $E = \Pi_{i=1}^N [a_i, b_i)$, then

$$\mu(E) \ = \ \Pi_{i=1}^N (b_i - a_i).\blacksquare$$

4 The Lebesgue Integral

Let $(X, \mathcal{M}, \mu)$ be a measure space and let $A_i \in \mathcal{M}$ and let $\alpha_i \geq 0$, for $1 \leq i \leq m$. Consider the function

$$s(x) \ = \ \sum_{i=1}^{M} \alpha_i \chi_{A_i}(x).$$

Such a function is called a **simple function**. Define the integral of s on X with respect to the measure μ by

$$\int_X f d\mu \ = \ \sum_{i=1}^{m} \alpha_i \mu(A_i).$$

Proposition 4.1 *Let $f : X \to [0, \infty)$ be a non-negative measurable function. Then, there exists a sequence of simple functions s_n such that, for all n,*

$$0 \ \leq \ s_n \ \leq \ s_{n+1} \ \leq \ f$$

and

$$\lim_{n\to\infty} s_n \ = \ f. \blacksquare$$

In view of the above proposition, we may define, for any non-negative measurable function f, its integral over X, with respect to the measure μ, by

$$\int_X f d\mu \ = \ \sup_{\substack{0\leq s\leq f \\ s \text{ simple}}} \int_X s d\mu.$$

In case of a simple function, it is easy to verify that both the definitions coincide.

Integration is a linear operation. If f and g are non-negative measurable functions, and if $c \in \mathbb{R}, \ c \geq 0$, then

$$\int_X (f+g) d\mu \ = \ \int_X f d\mu + \int_X g d\mu \text{ and } \int_X c f d\mu \ = \ c \int_X f d\mu.$$

Now, let f be any measurable function and set

$$f^+ \ = \ \max\{f, 0\}, \ \text{ and } \ f^- \ = \ -\min\{f, 0\}.$$

These are non-negative measurable functions as well and

$$f \ = \ f^+ - f^-, \ |f| \ = \ f^+ + f^-.$$

Definition 4.1 *We say that a measurable function f is* **integrable** *($f \in L^1(\mu)$) if*

$$\int_X |f| d\mu \ < \ +\infty. \blacksquare$$

If f is integrable, then

$$\int_X |f| d\mu = \int_X f^+ d\mu + \int_X f^- d\mu$$

is finite and so each integral on the right-hand side is finite. Thus, we can unambiguously define

$$\int_X f d\mu = \int_X f^+ d\mu - \int_X f^- d\mu.$$

Example 3.1 On the set $\mathbb{N}$ of natural numbers, we set $\mathcal{M}$ to be the σ-algebra of *all* subsets. We define the **counting measure** on $(\mathbb{N}, \mathcal{M})$ as follows: $\mu(E) = +\infty$ if E is an infinite subset and $\mu(E)$ equals the number of elements in E if E is a finite set. Then μ defines a measure. Any function $f : \mathbb{N} \to \mathbb{R}$ is then measurable and in fact is defined by a sequence $\{a_n\}$ where $f(n) = a_n$. It is easy to see that

$$\int_{\mathbb{N}} f d\mu = \sum_{n=1}^{\infty} a_n.$$

Thus, an integrable function gives rise to an absolutely convergent series and we recover the fact that an absolutely convergent series is convergent. ■

The Lebesgue integral on $\mathbb{R}$ is integration with respect to the Lebesgue measure. All Riemann integrable functions are also Lebesgue integrable and the integrals are the same. However, as we saw, the characteristic function of the set of rationals is not Riemann integrable. But the rationals form a countable set, which is measurable and so its characteristic function is a simple function and hence Lebesgue integrable (and, in fact, the integral is zero).

The Lebesgue integral behaves very nicely with respect to limit processes. We conclude by stating two very useful theorems in this regard.

Theorem 4.1 *(Monotone Convergence Theorem) Let $\{f_n\}$ be a sequence of measurable functions such that*

$$0 \leq f_1 \leq f_2 \leq \ldots \leq f_n \leq f_{n+1} \leq \ldots \leq +\infty$$

and let $f_n(x) \to f(x)$ for all $x \in X$. Then, f is measurable and

$$\lim_{n\to\infty} \int_X f_n d\mu = \int_X f d\mu. \blacksquare \tag{4.1}$$

Theorem 4.2 *(Dominated Convergence Theorem) Let* f_n *be a sequence of measurable functions and let* $f_n(x) \to f(x)$ *for all* $x \in X$*. Assume that* $|f_n| \leq g$ *for all* n*, where* g *is an integrable function. Then* f *is also integrable and*

$$\int_X |f_n - f| d\mu \to 0$$

as $n \to \infty$*. In particular (4.1) holds.*■

References

[1] Royden, H. L. Real Analysis, Macmillan.

[2] Rudin, W. Principles of Mathematical Analysis, McGraw-Hill.

[3] Rudin, W. Real and Complex Analysis,Tata McGraw - Hill.