

Measure Theory notes

Dave Neary

July 17, 2020

1 Motivation for Lebesgue integral

1.1 Overview of the Riemann integral

The Riemann integral of a continuous function $f : \mathbb{R} \rightarrow \mathbb{R}$ is defined as:

$$\int_a^b f(x)dx = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_i)\Delta x$$

where $x_i = a + (i - 1)\Delta x$ and $\Delta x = \frac{b-a}{n}$

In other words, we partition the domain of the function into small slices, and calculate the area under the curve by multiplying the width of the slices by the value of the function at the beginning of the slice.

This works well for a certain class of functions, called Riemann-integrable functions. These functions must satisfy the condition that the domain is \mathbb{R} , and that limit above exists.

More generally, we can calculate an upper Riemann sum $U(f)$ by summing the areas using $\sup f(x)$ on each partition, and a lower sum $L(f)$ by using $\inf f(x)$ for each interval. The function f is Riemann integrable when $\lim L(f) = \lim U(f)$.

1.2 Limitations of the Riemann integral

It is possible to generalize the Riemann integral to two or more dimensions, but the problem of finding an appropriate partition of the domain means that for dimensions of the real numbers which are higher than \mathbb{R}^2 , the Riemann integral is limited. In addition, we would like to consider other classes of domains than the reals for functions - for example, probability spaces or generic Hilbert spaces - where some alternative idea of the area under the curve (or more generally, the volume of a set) may make sense. Another limitation of the Riemann integral is that there are useful classes of functions for which it does not converge, but for which a reasonable value for the integral exists.

Another limitation of the Riemann integral is that there is only a very limited set of functions for which it is possible to say

$$\int \sum_n f_n(x) dx = \sum_n \int f_n(x) dx$$

Namely, $f_n(x)$ must converge uniformly to $f(x)$, which is a very strong constraint.

As a result of these limitations, the idea of the Lebesgue integral is to partition the function range instead of the domain. We then identify the subsets of the domain for specific values of $f(x)$, and calculate their volume using a generic measure function μ . By taking finer and finer intervals of the range, we can get better and better estimates of the volume under the function with respect to the domain and the measure.

The remainder of this document will describe the characteristics of a domain, the constraints required for a measure, which types of functions we can integrate, and a precise definition of the Lebesgue integral. We will also include a selection of proofs and problems which we can use the Lebesgue integral to solve.

2 Lebesgue Measure

Working backwards, to define what we mean by an integrable function, we will need to first define how to measure the volume of a subset of a domain (a measure), and to define a measure, we must first define the types of sets which will be measurable.

2.1 Measurable spaces

Starting from a set X , a collection of subsets of X , \mathcal{A} , is called a σ -algebra if it satisfies the following conditions:

1. $X \in \mathcal{A}$
2. For each $A \in \mathcal{A}$, $X \setminus A \in \mathcal{A}$
3. For a countable sequence of subsets $(A_n)_{n \in \mathbb{N}} \in \mathcal{A}$,

$$\bigcup_n A_n \in \mathcal{A}$$

We will see when we define a measure why this is called a σ -algebra.

The pair (X, \mathcal{A}) is called a measurable space.

Given any collection of subsets \mathcal{C} of subsets of X , we can generate a smallest σ -algebra which contains \mathcal{C} . That is, there is a σ -algebra \mathcal{A} such that if \mathcal{B} is a σ -algebra containing \mathcal{C} , then $\mathcal{A} \subseteq \mathcal{B}$. We say that such a σ -algebra \mathcal{A} is generated by \mathcal{C} .

2.1.1 Examples

1. **Exercise:** If $X = \{1, 2, 3, 4\}$, and the σ -algebra \mathcal{A} is generated by $\{\{1, 2\}, \{2, 3\}\}$, what are the other members of \mathcal{A} ?

Answer: By condition 1 above, $X = \{1, 2, 3, 4\} \in \mathcal{A}$, and by condition 2, since $X \in \mathcal{A}$, $X \setminus X = \emptyset \in \mathcal{A}$. Similarly, since $\{1, 2\}$ and $\{2, 3\} \in \mathcal{A}$, $X \setminus \{1, 2\} = \{3, 4\}$ and $X \setminus \{2, 3\} = \{1, 4\} \in \mathcal{A}$. By rule 3, $\{1, 2\} \cup \{2, 3\} = \{1, 2, 3\}$ and $\{2, 3\} \cup \{3, 4\} = \{2, 3, 4\} \in \mathcal{A}$. And by rule 2 again, $X \setminus \{1, 2, 3\} = \{4\}$ and $X \setminus \{2, 3, 4\} = \{1\} \in \mathcal{A}$. Back to rule 3, $\{4\} \cup \{1, 2\} = \{1, 2, 4\}$ and $\{1\} \cup \{3, 4\} = \{1, 3, 4\} \in \mathcal{A}$. Finally, $X \setminus \{1, 2, 4\} = \{3\}$ and $X \setminus \{1, 3, 4\} = \{2\} \in \mathcal{A}$. Since each of the individual elements of X are in a subset on their own, we can now create all possible subsets of X . $\mathcal{A} = \mathcal{P}(X)$, the power set of all subsets of X .

2. **Exercise:** Prove that for a countable sequence of subsets $(A_n)_{n \in \mathbb{N}} \in \mathcal{A}$, a σ -algebra on X , that

$$\bigcap_n A_n \in \mathcal{A}$$

Answer: Define a sequence of sets $B_n = X \setminus A_n$. Then, by condition 2, $B_n \in \mathcal{A}$ for all n . By condition 3,

$$B = \bigcup_n B_n \in \mathcal{A}$$

$$X \setminus B \in \mathcal{A}$$

By Demorgan's laws,

$$X \setminus \bigcup_n B_n = \bigcap_n (X \setminus B_n) = \bigcap_n A_n$$

So $\bigcap_n A_n \in \mathcal{A}$. QED.

3. **Exercise:** (X, \mathcal{A}) is a measurable space, with $Y \subset X$. Prove that (Y, \mathcal{A}') is a measurable space, where $\mathcal{A}' = \{A \cap Y \mid A \in \mathcal{A}\}$

Answer: Since $\emptyset \in \mathcal{A}$, $\emptyset \cap Y = \emptyset \in \mathcal{A}'$. Similarly, since $X \in \mathcal{A}$, $X \cap Y = Y \in \mathcal{A}'$.

For any $A \in \mathcal{A}$, $X \setminus A \in \mathcal{A}$, and $(X \setminus A) \cap Y = (X \cap Y) \setminus (A \cap Y) = Y \setminus (A \cap Y)$. So if $A \cap Y \in \mathcal{A}'$, then $Y \setminus (A \cap Y) \in \mathcal{A}'$

Finally, let $(A_i)_{i \in \mathbb{N}}$ be a sequence of sets in \mathcal{A} . Then

$$\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{A}$$

Define a sequence $(B_i)_{i \in \mathbb{N}}$ with $B_i = Y \cap A_i$ for all i . Then

$$\bigcup_{i \in \mathbb{N}} B_i = \bigcup_{i \in \mathbb{N}} (Y \cap A_i) = Y \cap \left(\bigcup_{i \in \mathbb{N}} A_i \right) \in \mathcal{A}' \quad (1)$$

Therefore, (Y, \mathcal{A}') is a measure space.

2.2 Measures

The extended real numbers is the set $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$.

Definition: A measure is a function $\mu : X \rightarrow \overline{\mathbb{R}}_0^+$ on a measurable space (X, \mathcal{A}) which satisfies the conditions:

1. $\mu(\emptyset) = 0$
2. if $(A_i)_{i \in \mathbb{N}} \in \mathcal{A}$ is a sequence of pairwise disjoint sets (that is, $A_i \cap A_j = \emptyset$ if $i \neq j$), then:

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

This characteristic of being able to turn a countable union of disjoint sets into a sum is why \mathcal{A} is called a σ -algebra.

In general, we can think of the measure of a set as its volume, or (for real functions) as the area under the curve.

A measurable space (X, \mathcal{A}) with a measure μ is called a measure space, and is written (X, \mathcal{A}, μ) .

We can deduce a number of lemmas from this definition:

Lemma: If $A \subseteq B$ and $A, B \in \mathcal{A}$, then $\mu(A) \leq \mu(B)$

Lemma: For a sequence of sets $(A_i)_{i \in \mathbb{N}} \in \mathcal{A}$

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

Lemma: For a sequence of sets $(A_i)_{i \in \mathbb{N}} \in \mathcal{A}$ with $A_i \subseteq A_j$ if $i < j$ then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i \rightarrow \infty} \mu(A_i)$$

Similarly, for a sequence where $A_i \supseteq A_j$ for $i < j$, $\mu(\bigcap_{i=1}^{\infty} A_i) = \lim_{i \rightarrow \infty} \mu(A_i)$

Some examples of measures are the trivial measure $\mu(A) = 0$ for all $A \in \mathcal{A}$, the counting measure $\mu(A) = |A|$ if A is finite, or ∞ if it is infinite, and the Dirac measure $\delta_a(S) = 1$ if $a \in S$ or 0 otherwise.

Exercise: Prove that the trivial, counting, and Dirac functions are measures.

- **Trivial measure:** if $\mu(A) = 0$ for all A , then $\mu(\emptyset) = 0$, and for a pairwise disjoint collection of sets $(A_i)_{i \in \mathbb{N}} \in \mathcal{A}$,

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = 0$$

and since $\mu(A_i) = 0$ for all i ,

$$\sum_{i=1}^{\infty} \mu(A_i) = 0 = \mu\left(\bigcup_{i=1}^{\infty} A_i\right)$$

Therefore, μ is a measure.

- **Counting measure:** $\mu(\emptyset) = |\emptyset| = 0$ For a collection of pairwise disjoint sets $(A_i)_{i \in \mathbb{N}} \in \mathcal{A}$, if

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) < \infty$$

then each of $\mu(A_i)$ is finite, and there is a finite collection of subsets (A_i) of X . Each element of $\bigcup_{i=1}^{\infty} A_i$ is also an element of exactly one A_i and each element of each A_i is also an element of $\bigcup_{i=1}^{\infty} A_i$ by definition. If

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \infty$$

, then for each $a \in \bigcup_{i=1}^{\infty} A_i$, $a \in A_i$ for some i . Therefore, in either case,

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

Therefore, the counting measure $\mu(A) = |A|$ is a measure.

- **Dirac measure:** For an element $a \in X$, $\delta_a(A) = 0$ if $a \notin A$. Since $a \notin \emptyset$, $\delta_a(\emptyset) = 0$. Consider a pairwise disjoint collection of sets $(A_i)_{i \in \mathbb{N}} \in \mathcal{A}$. If $a \in \bigcup_{i=1}^{\infty} A_i$ then

$$\delta_a\left(\bigcup_{i=1}^{\infty} A_i\right) = 1$$

and $a \in A_i$ for some $i \in \mathbb{N}$, and since (A_i) are pairwise disjoint, $\delta_a(A_i) = 1$ and $\delta_a(A_j) = 0$ for $j \neq i$. Then

$$\sum_{i=1}^{\infty} \delta_a(A_i) = 1 = \delta_a\left(\bigcup_{i=1}^{\infty} A_i\right)$$

If $a \notin \bigcup_{i=1}^{\infty} A_i$ then $a \notin A_i$ for all i , and

$$\sum_{i=1}^{\infty} \delta_a(A_i) = 0 = \delta_a\left(\bigcup_{i=1}^{\infty} A_i\right)$$

Therefore, δ_a is a measure.

Exercise: Prove the lemmas above.

2.3 Measurable functions

Let $(\mathcal{X}, \mathcal{A})$ and $(\mathcal{Y}, \mathcal{B})$ be measurable spaces. A function $f : \mathcal{X} \rightarrow \mathcal{Y}$ with respect to the σ -algebras \mathcal{A}, \mathcal{B} if, for each subset $B \in \mathcal{B}$, $f^{-1}(B) \in \mathcal{A}$ (where $f^{-1}(B)$ is the pre-image of the set B under the function f , $\{x \in X : f(x) \in B\}$).

Thinking about useful collections of sets for a measure space, for functions mapping onto $\overline{\mathbb{R}}$, we can generate a σ -algebra from the set of open intervals in \mathbb{R} , plus ∞ . For the open interval $A = (a, b)$, we define the measure $\lambda(A) = b - a$. This measure is called the Lebesgue measure.

A σ -algebra \mathcal{A} generated from all of the open subsets of \mathcal{X} is called the Borel σ -algebra. It is a useful concept, because by choosing a Borel σ -algebra, \mathcal{A} is also a topology, and we inherit all of the useful theorems from topology too.

Reminder: A topological space is a nonempty set X plus a set of subsets \mathcal{A} possessing the properties:

1. $X, \emptyset \in \mathcal{A}$
2. If $O_1 \in \mathcal{A}$ and $O_2 \in \mathcal{A}$, then $O_1 \cap O_2 \in \mathcal{A}$
3. For a sequence of sets $(O_i)_{i \in \mathbb{N}} \in \mathcal{A}$, the countable union $\bigcup_{i \in \mathbb{N}} O_i \in \mathcal{A}$

Exercise:

Exercise:

2.4 Lebesgue Integral

We can now pull all of these ideas together to define the Lebesgue integral.

We define the characteristic function $\chi_E(x)$ of the set E :

$$\chi_E(x) = \begin{cases} 1 & x \in E \\ 0 & x \notin E \end{cases}$$

A linear combination

$$\phi(x) = \sum_{i=1}^n a_i \chi_{E_i}(x)$$

is called a simple function, if ϕ is measurable with respect to the σ -algebra generated by the sets (E_i) , and assumes only a finite number of values $\{a_1, a_2, \dots, a_n\}$.

The Lebesgue integral of a simple function

$$\phi = \int_X \left(\sum a_i \chi_{A_i}(x) \right) d\mu = \sum a_i \mu(A_i)$$

is the result of defining $A_i = \phi^{-1}(a_i)$ (that is, $A_i = \{x : \phi(x) = a_i\}$) for each of the values a_i that ϕ assumes. One consequence of this definition is that (A_i) is a sequence of pairwise disjoint sets.

We can define the Lebesgue integral for a measurable non-negative function $f : X \rightarrow \overline{\mathbb{R}}^+$ with respect to a measure space (X, \mathcal{A}, μ) as:

$$\int_X f(x) d\mu = \sup \left\{ \int_X \phi(x) d\mu : 0 \leq \phi(x) \leq f(x), \phi \text{ a simple function} \right\}$$

In other words, we look over all of the simple functions that are less than f , and take the supremum across all of them.

For functions which are not non-negative, we split $f(x)$ into

$$g(x) = \max(f(x), 0)$$

and

$$h(x) = -\min(f(x), 0)$$

Then

$$\int_X f(x) d\mu = \int_X g(x) d\mu - \int_X h(x) d\mu$$

In other words, we split $f(x)$ into two non-negative functions, one representing the positive part of f , and one representing the absolute value of the negative part of f , and we can calculate the final integral by removing the negative area from the positive area.

For any continuous functions $f(x) : X \rightarrow \overline{\mathbb{R}^+}$, we can construct a sequence of simple functions $\{f_n(x)\}$ which converges pointwise to $f(x)$ as follows. For $f_n(x)$, partition the range into $2^{2n} + 1$ disjoint partitions $\{I_{n,i}\}$ with

$$I_{n,i} = \begin{cases} \left[\frac{i-1}{2^n}, \frac{i}{2^n} \right) & 1 \leq i \leq 2^{2n} \\ \left[\frac{i-1}{2^n}, \infty \right) & i = 2^{2n} + 1 \end{cases}$$

Then define $\{A_{n,i}\}_{i \leq 2^{2n}+1} = f^{-1}(I_{n,i})$, the preimage of $I_{n,i}$. The collection $\{I_{n,k}\}$ cover $[0, \infty)$ for all n . The simple function

$$f_n(x) = \sum_{i=1}^{2^{2n}+1} \frac{(i-1)}{2^n} \chi_{A_{n,i}}(x)$$

is a sequence of increasing functions which converge pointwise to $f(x)$.

Exercise: Prove that the sequence $f_n(x)$ above converges pointwise to $f(x) = x^2$ for all $x \in \overline{\mathbb{R}^+}$.

3 Lebesgue Integrals and Probability Theory

Probability distributions all share some common characteristics which allow the application of measure theory to be useful. Given a sample space Ω of possible outcomes, and an event space \mathcal{A} , which is a σ -algebra, and a probability measure P , a measure space (Ω, \mathcal{A}, P) is called a probability space if:

1. $P(\emptyset) = 0$
2. $P(\Omega) = 1$
3. if $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{A}$ is a countable collection of pairwise disjoint sets, then:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$