Why Do We Need Measure Theory?

A Math 105 Essay

Rasmus Pallisgaard

1 Introduction

Lebesgue measure theory, named after French mathematician Henri Lebesgue, defines integration via measurability of sets and functions in $\mathbb{R}^n$. This is needed, as Riemann integration which, we learn of in high school and develop further in college doesn't suffice when working with more complicated functions and subsets of $\mathbb{R}^n$. In this Essay I cover Riemann integration quickly, followed by an example of why this theory does not suffice when integrating some functions. Following this I cover the main points of Lebesgue measure theory used to develop the Lebesgue integral, and finally I use this theory to integrate one such function that Riemann integration is unable to solve.

2 Riemann Integration

While we have already, in homework 6, covered a small comparison between Riemann integration and Lebesgue integration, we find it more interesting to cover Riemann integration a bit more in depth, using [Pugh, 2015] as a reference, since this ties in nicely with his definition of Lebesgue integration.

Take a function $f : [a, b] \to \mathbb{R}$. The integral of f is the area of the undergraph of f, $U(f)$, given by

$$\int_a^b f(x)\, dx = \text{area}(U),$$

where

$$U = \{(x, y) : a \leq x \leq b, 0 \leq y < f(x)\}.$$

Using Riemann integration, this area is computed by partitioning the domain of the function, $[a, b]$, into two sets of points P, T where $P = \{x_0, ..., x_n\}$ and

$T = \{t_1, ..., t_n\}$ such that

$$a = x_0 \leq t_1 \leq x_1 \leq t_2 \leq ... \leq t_n \leq x_n = b$$

This interlacing of points t_i, x_i along with the definition of $\Delta x_i = x_i - x_{i-1}$ is used to define the Riemann sum as

$$R(f, P, T) = \sum_{i=1}^{n} f(t_i) \Delta x_i$$

This can be interpreted as every point t_i being surrounded by two x_i's such that $x_{i-1} \leq t_i \leq x_i$. Pugh defines the mesh of P as the longest existing interval $[x_{i-1}, x_i]$, and thus the Riemann integral is defined as

$$\int_a^b f(x)\, dx = I = \lim_{\text{mesh}(P) \to 0} R(f, P_h, T)$$

One note on the Riemann integral is that if $P^{(h)}$ and T are defined such that for all intervals $[x_{i-1}, x_i]$, their length are all equal to h, we can define the Riemann integral as

$$\int_a^b f(x)\, dx = I = \lim_{h \to 0} R(f, P^{(h)}, T)$$

I note this, because this is how I initially learned the definition of the Riemann integral, and the notation relates it nicely to the definition of the ordinary derivative of f.

3 Why is this not enough?

Note that in the definition of the Riemann integral we define the domain of f to be closed and bounded, i.e., $f : [a, b] \to \mathbb{R}$.

Consider example 4.1 from [Bernard R Gelbaum, 2003]. This famous construction, called the *Dirichlet function*, lets $f : [0, 1] \to \{0, 1\}$ be the characteristic function of the rational numbers, i.e.

$$f(x) = \begin{cases} 0 & x \in \mathbb{Q} \\ 1 & x \notin \mathbb{Q} \end{cases}$$

In order to analyse if the integral of this function exists we have to look at integrability, which can be defined as follows. Let

$$L(f,P) = \sum_{i=1}^{n} m_i \Delta x_i \qquad m_i = \inf\{f(t), x_{i-1} \le t \le x_i\}$$
$$U(f,P) = \sum_{i=1}^{n} M_i \Delta x_i \quad M_i = \sup\{f(t), x_{i-1} \le t \le x_i\},$$

And note that
$$L(f,P) \le R(f,P,T) \le U(f,P).$$

The upper and lower integrals of f are defined as

$$\underline{I} = \sup_P L(f,P) \qquad \bar{I} = \inf_P U(f,P)$$

over all possible partitions of $[a,b]$ into P. By Pugh theorem 3.20 if f is Riemann Integrable, then $\underline{I} = I = \bar{I}$.

Looking back at the Dirichlet function, due to its definition we find that

$$\underline{I} = 0 \text{ nad } \bar{I} = 1.$$

As $\mathbb{Q}$ is dense in $\mathbb{R}$, meaning that between every real number exists both a rational number and an irrational number. As such, for any partition of $[a,b]$ we can find both a rational and an irrational number in every interval, the lower integral will always be 0 and the upper integral will always be 1.

By theorem 3.20 this function thus isn't integrable. This is an example of an issue with Riemann integration: While it works well on functions we often meet, it fails to generalise to more complicated domains and functions we want to analyse.

4 Outer Measure and Measurable Sets

The following chapters are written based on Chapter 6 of Real Mathematical Analysis by Pugh.

We can intuitively define the length of an interval (a,b) to be $b-a$, but with more complicated sets, like the Cantor set, how is this done? We define the outer measure of a set $A \subset \mathbb{R}^n$ by

$$m^*(A) = \inf\left\{\sum_k^{\infty} |B_k| : \{B_k\} \text{ is a collection of open boxes that covers A}\right\},$$

where an open box is defined as the cartesian product of open intervals: $B = \prod_i (a_i, b_i)$. Along with this definition follows the axioms of outer measures:

(a) The outer measure of the empty set is 0.

(b) if $A \subset B$, then $m^*(a) \leq m^*(b)$ (monotonicity)

(c) if $A = \cup_{n=1}^{\infty} A_n$, then $m^*(A) \leq \sum_{i=1}^{\infty} m^*(A_n)$ (countable sub-additivity)

(d) the outer measure of a closed, open, or half open half closed box $B \subset \mathbb{R}^n$ is $m^*(B) = \prod_i^n (b_i - a_i)$.

We define measurability of a set A by the Carathéodory criterion: A set $A \subseteq \mathbb{R}^n$ is measurable if for all $E \subseteq \mathbb{R}^n$, we have

$$m^*(A) = m^*(A \cap E) + m^*(A \cup E) = m^*(A \cap E) + m^*(A \cap E^c)$$

If this holds, we say $m(A) = m^*(A)$ is the measure of A. At face value this definition might seem gratuitous since most sets we think of fulfils this requirement, but in the following example we show that measurability is a nontrivial condition (example 8.11 from [Bernard R Gelbaum, 2003])

5 A counter example to measurability

In a simplified view of abstract measure theory, let S be a set. Then 2^S is the set of all subsets of S. A subset $M_S \subset 2^S$ is called a σ-algebra if

- $\emptyset \in M_S, S \in M_S$
- For all $A_1, A_2, ... \in M_S$, $\Cup_i^{\infty} A_i \in M_S$, $\Cap_i^{\infty} A_i \in M_S$, and $A_i^c \in M_S$.

A measure function μ on a measurable space (S, M_S) is a function such that

- $\mu : M_s \to [0, +\infty]$
- $\mu(\emptyset)$
- For $A_i \in M_S$, countable addition holds

A measurable space with an associated measure function (S, M_s, μ) is called a measure space. The Lebesgue measure function was described earlier as $m^*(A)$, and when $m(A)$ exists, A is Lebesgue measurable.

Let μ be a measure function define for all the sets A of real numbers, let it be finite for bounded sets, and let it be such that $\mu(x + A) = \mu(A)$ for every

$x \in \mathbb{R}$. Let the equivalence relation $\sim$ defined on $(0,1] \times (0,1]$ be such that $x \sim y$ if and only if $x - y \in \mathbb{Q}$. This binary equivalence relation can be partitioned into many disjoint equivalence classes $\mathbf{C}$. An equivalence class on an element a is the set $S_a = \{s \in S : x \sim a\}$. Applying the axiom of choice on these equivalence classes produces a set $\mathbf{B}$ that has the following properties: no two points of $\mathbf{B}$ are equivalent to each other, and every point x on $(0,1]$ is equivalent to some member of $\mathbf{B}$.

We can define for each $r \in (0,1]$ an operation on the set $\mathbf{B}$, called the *translation modulo 1* by

$$(r + \mathbf{B})(\text{mod } 1) = \{(r + \mathbf{B}) \cap (0,1]\} \cup \{((r-1) + \mathbf{B} \cap (0,1]\}$$

Claim: $\mathbf{B}$ is not measurable.

Proof: The two properties of the set $\mathbf{B}$ implies that (1) for any two sets $(r + \mathbf{B})(\text{mod } 1)$ and $(s + \mathbf{B})(\text{mod } 1)$ for distinct rational numbers $r, s \in \mathbb{Q}$ are disjoint, and (2) that every real number $x \in (0,1]$ is a member of a set $(r + \mathbf{B})(\text{mod } 1)$ for some rational number $r \in \mathbb{Q} \cap (0,1]$. Because of this, we can express the interval $(0,1]$ as a union of the pair-wise disjoint countable collection $\{(r + \mathbf{B})(\text{mod } 1)\}$, where $r \in \mathbb{Q} \cap (0,1]$. By the following:

$$\begin{aligned}
\mu\left((r + \mathbf{B})(\text{mod } 1)\right) &= \mu\left((r + \mathbf{B}) \cap (0,1]\right) + \mu\left(((r-1) + \mathbf{B} \cap (0,1]\right) \\
&= \mu\left((r + \mathbf{B}) \cap (0,1]\right) + \mu\left((r + \mathbf{B}) \cap (1,2]\right) \\
&= \mu\left((r + \mathbf{B}) \cap (0,2]\right) \\
&= \mu\left(r + \mathbf{B}\right) \\
&= \mu(\mathbf{B}),
\end{aligned}$$

showing that all sets obtained from translation modulo 1 on $\mathbf{B}$ have the same measure as $\mathbf{B}$. We note that the Lebesgue measure function has positive finite measure for all bounded intervals. Assuming that this holds for our measure function we find that

$$\mu((0,1]) = \sum_{r \in \mathbb{Q} \cap (0,1]} \mu((r + \mathbf{B})(\text{mod } 1)) = \sum_{r \in \mathbb{Q} \cap (0,1]} \mu(\mathbf{B}) = +\infty.$$

As $\in \mathbb{Q} \cap (0,1]$ is countably infinite and each $\mu(\mathbf{B})$ must be positive. This cannot hold since $(0,1]$ is bounded, and therefore $\mu(\mathbf{B}) = 0$. Thus the set $\mathbf{B}$ is not Lebesgue measurable.

$\square$

6 Properties of measurability

Apart from the properties inherited from outer measures, the following is a list of properties that hold for the Lebesgue measure of a subset $A \subset \mathbb{R}^n$:

(a) If A is measurable, then A^c is measurable.

(b) If A is measurable, then $A + x$ is measurable for an $x \in \mathbb{R}^n$.

(c) If A_1, A_2 measurable, then $A_1 \cup A_2$ and $A_1 \cap A_2$ are measurable.

(d) For a finite collection of disjoint sets $A_1, A_2, ..., A_n$, $m(\sqcup_i^n A_i) = \sum_i^n m(A_i)$ (finite additivity).

(e) For a countable collection of disjoint sets $A_1, A_2, ...$, $m(\sqcup_i^\infty A_i) = \sum_i^\infty m(A_i)$ (countable additivity).

(f) If $m^*(A) = 0$, then A is measurable.

Before describing Lebesgue Integrals we quickly cover some lemmas of Lebesgue measurability, as Lebesgue measurability is tied very strongly to Lebesgue integration. These lemmas are the following:

(a) If A, B are measurable and $A \subset B$, then $m(B \setminus A) = m(B) - m(A)$

(b) All open sets in $\mathbb{R}^n$ are countable unions of open boxes and thus measurable.

A subset $E \subset \mathbb{R}^n$ is a zero set or a nullset if $m(E) = 0$. For nullsets the following hold:

(a) $\forall A \in \mathbb{R}^n$, $m(A \cup E) = m(A)$ and $m(A \cap E^c) = m(A)$

(b) If $\bar{E} \subset E$, then $\bar{E}$ is a nullset.

(c) If $m(E) = 0$, then E is measurable.

(d) $F \subset \mathbb{R}^n$ is measurable if and only if $F \cup E$ is measurable

This leads to the following two definitions: If S is a topological space, and $u_1, u_2, ...$ is a countable collection of open sets, then $\cap_i^\infty u_i$ is called af G_δ-set. If $f_1, f_2, ...$ is a countable collection of closed sets, then $\cup_i^\infty f_i$ is called a F_σ-set. Lebesgue measurability can then be defined such that a set A is measurable if and only if there exists an F_σ-set of closed subsets F and a G_δ-set of open subsets we have G where $F \subset A \subset G$ we have $m(F \setminus G) = 0$.

The two final theorems covered are the Product theorem and the slice theorem. As can be seen, these two theorems work as opposites of each other:

The product theorem states that if $A \subset \mathbb{R}^n$ and $B \subset \mathbb{R}^k$ are measurable, then $A \times B \subset \mathbb{R}^n \times \mathbb{R}^k$ is measurable and $m(A \times B) = m(A)m(B)$. A specific case of this theorem is that $\{x\} \times \mathbb{R} \subset \mathbb{R}^2$ has measure 0 as an element in $\mathbb{R}$ has measure 0.

The slice theorem says that for a set $A \subset \mathbb{R}^n \times \mathbb{R}^k$, $m(A) = 0$ if for almost every slice $A_x = \left\{y \in \mathbb{R}^k : (x, y) \in \mathbb{R}^n \times \mathbb{R}^k\right\}$ has measure zero, i.e., $m(A_x) = 0$. Here, almost every means that this holds for all sets except for a small subset, which is a nullset.

7 Lebesgue Integrability

Let $f : \mathbb{R} \to [0, \infty)$. We define it's undergraph, like with the Riemann integral, as

$$U(f) = \{(x, y) : 0 \leq y < f(x)\}$$

Following Pugh's definition of measurability and integrability we say that f is a measurable if the undergraph $U(f)$ is measurable. If so, then we say

$$\int f = m(U(f))$$

if $\int f < \infty$ then we say that f is integrable. This type of integral is called the Lebesgue integral. Initially this is defined only for functions with a positive range, and the definition for functions with negative range and full range follows later on. Accompanying this definition is a list of theorems regarding integrability of functions. The first is the **Monotone Convergence Theorem**, which says that given a sequence of measurable functions $\{f_n\}$, $f_n : \mathbb{R} \to [0, \infty)$, and $f_n \uparrow f$ a.e. as $n \to \infty$, then

$$\int f_n \uparrow \int f$$

If $f_n \downarrow f$ a.e. as $n \to \infty$, then

$$\int f_n \downarrow \int f$$

The second theorem is the **Dominated Convergence Theorem**, which states that given a sequence of integrable functions $\{f_n\}$ and $f_n \to f$ a.e, and g is an integrable function such that $g(x) \geq f_n(x)$. Then

$$\int f_n \to \int f$$

The third theorem is **Fatou's lemma**, which states that if f_n is a sequence of measurable functions, then

$$\int \liminf f_n \leq \liminf \int f_n$$

The following is furthermore a set of properties for integrable functions which follow from the properties of measurability:

(a) If $f \leq g$, then $\int f \leq \int g$

(b) If $f(x) = g(x)$ a.e., then $\int f = \int g$

(c) If $c \geq 0$ then $\int cf = c \int f$

(d) The integral of f is zero if and only if $f(x) = 0$ a.e.

(e) $\int (f + g) = \int f + \int g$

Next we cover Fubini's theorem and the extension of Lebesgue integrability to the real numbers.

proceeding as done in the lectures, the indicator function for a measurable set $E \subset \mathbb{R}$ is defined as

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

Given a function $f : R \to R$, we define

$$E_+ = \{x \in \mathbb{R} : f(x) > 0\} \quad E_- = \{x \in \mathbb{R} : f(x) < 0\}$$

Notice that neither of these are defined when $f(x) = 0$, as the undergraph of f does not include these points. Using E_+ and E_- we can define

$$f_+ = f \cdot I_{E_+} \quad f_- = f \cdot I_{E_-}$$

where f_+ and f_- both are non-negative functions. With this we get the following:

$$f = f_+ - f_- \qquad |f| = f_+ + f_- \qquad \int f = \int f_+ - \int f_-$$

Using the above one can prove that the properties for the Lebesgue integrals of non-negative functions extend to the real-valued ones.

Fubini's theorem is thus the final named theorem we mention in relation to Lebesgue measure theory. It states that if $f : \mathbb{R}^2 \to [0, \infty)$ is measurable then there exists $F(x), G(y)$ such that

$$F(x) = \int f(x, y)\ dy \text{ a.e.}$$
$$G(y) = \int f(x, y)\ dx \text{ a.e.}$$

and

$$\int F(x)\ dx = \int f(x, y)\ dxdy = \int G(y)\ dy$$

Finally we cover two important statements. The following theory was derived from Pughs definition of measurable function, and in corollary 41 he proves that his undergraph measurability is equivalent to the traditional definition of a function $f : \mathbb{R} \to \mathbb{R}$ being measurable if and only if $f^-1(V)$ is measurable for every set $V \in \mathbb{R}$. This is great because this means the definitions and theorems named above agree with the traditional theory. Finally, with the traditional definition of a measurable function, Tao proves that Riemann integrability implies Lebesgue integrability, and that these results are equal when a function is Riemann integrable.

8 The Lebesgue integral of the Dirichlet function

While we have covered a lot of theory we need only a sample of it to produce the Lebesgue integral of the Dirichlet function.

We w.t.s that the Lebesgue integral of the Dirichlet function $f : [0, 1] \to \mathbb{R}$:

$$f(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \notin \mathbb{Q} \end{cases}$$

is 0, i.e. $\int f = 0$. First we partition the undergraph of f into two sets: E_0 and E_1 defined by

$$E_0 = \{x : x \notin \mathbb{Q}\} = [0, 1] \cap (\mathbb{R} \setminus \mathbb{Q}) \qquad E_1 = \{x : x \in \mathbb{Q}\} = [0, 1] \cap \mathbb{Q}$$

Furthermore, let f_{E_1} be f defined on E_1 and f_{E_0} likewise. Since $E_1 = [0, 1] \cap \mathbb{Q}$ is a countable set, its measure is 0, since we can split into a sum of countable elements each with a measure of 0. Thus $m(U(f_{E_1}))$. The set E_0 is uncountable, but $f(x) = 0 \forall x \in E_0$, and thus $m(U(f_{E_0})) = 0$. Therefore

$$\int f = \int f_{E_0} + \int f_{E_1} = 0 + 0 = 0$$

While the result above by itself is not groundbreaking, but it is a fun example of the power of Lebesgue integration. This example, at least, helped me understand why Lebesgue integration was needed what could be done with it.

9 Lebesgue Integration and Probability Theory

One area where Lebesgue Integration reigns supreme is probability theory. From the following notes `https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.716.4061&rep=rep1&type=pdf` from the University of Western Ontario, London, Canada, The properties of Lebesgue integration in probability theory is described, and it is fascinating. We will not delve into great detail about probability theory, but we will cover enough definitions and properties to show that core theorems described in the previous sections can be used.

In section 5 we learned of the measure space, the measure function and the measurable space. Now let Ω be a set we call the sample space and let M_Ω be a sigma algebra on this space. Then $E \in M_\Omega$ is called an event of the sample space. Now let (Ω, M_Ω, P) be a measure space such that the measure function P has the property that $P(\Omega) = 1$, i.e. that the measure of the entire space is 1. This space is called a Probability space. We then define a random variable as a measurable function $X : \Omega \to \mathbb{R}$. Because the measure of space is translation variant and because of finite additivity, as well as the product theorem, if X, Y, and $\alpha \in \mathbb{R}$ are random variables, then so are

- $Z = X + Y$
- $Z = \alpha X$
- $Z = XY$

Recalling our previous definition of the indicator function I_A:

$$I_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$$

If a random variable X has finite range, then it can be expressed as

$$X(\omega) = \sum_{k=1}^{n} a_k I_{A_k}(\omega)$$

since we for each of the finite a_i can let $A_i = X^{-1}(\{a_i\})$. Following the definition of non-negative functions for Lebesgue integrals, if (Ω, M_Ω, P) is a probability space and X is a non-negative random variable that can be represented as

$$X = \sum_{k=1}^{n} a_k I_{A_k}$$

Then its expectation is defined as

$$\mathbb{E}_P(X) = \sum_{k=1}^{N} A_k P(A_k)$$

Finally, while we haven't covered it in this essay, Tao (Analysis II) covers the construction of the Lebesgue integrals using sums of simple functions. Using this definition of the Lebesgue integral, By letting $\mathbb{E}(X) = \mathbb{E}(X^+) + \mathbb{E}(X^-)$ we find that the expectation of a random variable is simply the Lebesgue integral over the sample space:

$$\mathbb{E}(X) = \int_\Omega X(\omega) dP(\omega)$$

And $\mathbb{E}(X)$ is said to be integrable. Because of this we can utilise many of the theorems covered above to describe expectation. Among these are

- The Monotone Convergence Theorem: Let X be a random variable and let $\{X_n\}$ be a sequence of non-negative random variables such that $X_n \uparrow X$. Then
$$\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$$

- The Dominated Convergence Theorem: Let Y be an integrable random variable and let $\{X_n\}$ be a sequence of non-negative random variables such that $X_n \to X$ and $|X_n| \leq Y$ for all n. Then
$$\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$$

10 Conclusion

The following Essay has covered many details around measurability and Lebesque integrability, and used it first to argue for the use of it in cases where Riemann integrability cannot be used, and afterward as a tool to describe fundamental probability theory. We have covered the main definitions, properties and theorems of measure theory and lebesgue integrability theory from the perspective of Pugh, with the addition of showing why measure theory is needed by finding a set that that fails the Carathéodory criterion. We have also shown that the theory of the Riemann integral fails to be abl to compute the integral of the Dirichlet function, and subsequently shown that using the theory of the Lebesgue integral we can accurately compute this integral. Finally we have covered initial probability theory using the theory covered before and made connections between measure theory and the Lebesgue integral and the expected value from probability theory.

References

[Bernard R Gelbaum, 2003] Bernard R Gelbaum, J. M. H. O. (2003). Dover Publications, Inc.

[Pugh, 2015] Pugh, C. C. (2015). *Real Mathematical Analysis.* Springer.