

# The Lebesgue Integral

Andrew Paul

6/27/2020

## 1 Introduction

Integration is a powerful mathematical tool stemming from the desire to formalize and generalize the intuitive notions of volume. The earliest form of integration was by the Riemann integral. The Riemann integral does a good job of agreeing very well of our intuitive conceptions of volume. It works by partitioning  $\mathbb{R}^n$  (the domain), taking the product of the infimum and supremum of the function being integrated in each piece of the partition, multiplying those values by the volume of the piece of the partition to form the lower sum and upper sum respectively, and hoping that the two converge to a common limit as the partition is made finer.

The notion of the volume of a set itself comes from integrating the indicator functions of the relevant sets. For many practical purposes, the partition used can be, for instance, a dyadic partition, where it is quite simple to find the volumes of the constituent dyadic cubes (just multiply the side lengths). A Riemann integral, if it exists, can be shown to be independent of the partition chosen. However, the idea of using an indicator function to isolate sets over which we wish to integrate brings up an interesting point.

Riemann integration, in the way we have specified above, is only defined for functions that are bounded with bounded support. But many integrals of practical interest are over unbounded supports: for instance, the Gaussian integral from statistics, integrals of potential in physics, and more. This leads us to one of the first advantages of the Lebesgue integral. It is robust enough to work over unbounded supports. Furthermore, the Lebesgue integral is much more well-behaved under limits. The dominated convergence theorem for Lebesgue integration is a simple statement that allows us to interchange limits with the integral. Furthermore, it leads to the interchangeability of differentiation and integration when they are composed with each other under certain conditions. This is well-known as *Feynman's trick*.

The standard approach for defining the Lebesgue integral involves partitioning the *codomain*, and summing the products of the function with the *measure* of the preimage of the partition of the codomain. We will avoid this approach, along with any discussion of measure theory (except for measure zero), and instead define the Lebesgue integral as an infinite sum of Riemann integrals. The main aim here is not to fully develop measure theory in a rigorous manner. We will not prove various key results, such as the existence of the Lebesgue integral. Instead, we will try to provide

the motivation for the Lebesgue integral and show its utility in a variety of contexts.

I will be using a few results and proofs from *Vector Calculus, Linear Algebra, and Differential Forms*, my homework from MATH 31CH and PHYS 4B at UC San Diego, and a problem from David Altizio's Problem Stash.

## 2 Measure Zero

We begin with a definition.

**Definition.** A set  $X \subset \mathbb{R}^n$  has measure zero if  $\forall \epsilon > 0$ , there exists an infinite sequence of open boxes  $B_i$  such that

$$X \subset \bigcup_{i=1}^{\infty} B_i \text{ and } \sum_{i=1}^{\infty} \text{vol}_n(B_i) < \epsilon.$$

As we will see, by allowing an *infinite* sequence of boxes to become arbitrarily small, we arrive at a notion that is fundamentally different from having zero volume. Obviously, by the definition given above, if a set has volume 0, then it also has measure 0. But, it is also possible for a set to have measure 0 while having an undefined volume. The classic example of this is the set of rational numbers, say in the interval  $[0, 1]$ .

Since the rationals are dense in  $[0, 1]$ , the upper and lower sums of the indicator function  $\mathbf{1}_{\mathbb{Q} \cap [0, 1]}(x)$  refuse to converge as we partition  $[0, 1]$  finer. The supremum over each partition is always 1 and the infimum is always 0. So the volume of the set  $\mathbb{Q} \cap [0, 1]$  is undefined.

On the other hand, we can list the elements of the set as follows:

$$1, \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, \frac{1}{5}, \dots$$

Clearly, there are numbers that repeat in this list. Next, we define  $B_i$  to be the interval of length  $\frac{\epsilon}{2^i}$  centered at the  $i^{\text{th}}$  number in the list. There are obviously some pairwise intersections between the  $B_i$  so the length of the union is less than the sum of the lengths of the  $B_i$ . Regardless, we have,

$$\mathbb{Q} \cap [0, 1] \subset \bigcup_{i=1}^{\infty} B_i \text{ and } \sum_{i=1}^{\infty} \text{vol}_n(B_i),$$

so the set of rationals in  $[0, 1]$  has measure 0.

A well-known integrability condition is that if a function that is bounded with bounded support is continuous except on a set of volume 0, then it is integrable. While this is a sufficient condition for integrability, it is not necessary. There exists functions that are continuous except on a set that does not have a volume of 0 that are still integrable. To strengthen this condition to make it both sufficient and necessary for integrability, we must use the notion of measure 0.

**Theorem** (Integrability Condition). *Let  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  be bounded with bounded support. Then  $f$  is integrable if and only if it is continuous except on a set of measure 0.*

It turns out that we can ignore a lot of bizarre behavior as long as it occurs on a set of measure 0. In fact, this theme shows up so often, we give it a name.

**Definition.** We say that  $\vec{x}$  satisfies a property almost everywhere if it satisfies the property at all  $\vec{x}$  except those on a set of measure 0.

The classic example of a function that is discontinuous on a set without volume 0, yet is still integrable is Thomae's function.<sup>1</sup> Restricted to  $[0, 1]$ , it is

$$f(x) = \begin{cases} \frac{1}{q} & \text{if } 0 \leq x = \frac{p}{q} \leq 1, \text{ with } p, q \in \mathbb{Z} \text{ coprime} \\ 0 & \text{otherwise.} \end{cases}$$

This function is continuous at the irrational numbers. Around every irrational number in  $[0, 1]$ , we can construct a small enough neighborhood so that the denominators of rational numbers in that neighborhood are large, and thus the function can be made arbitrarily close to zero around every point where it actually is zero.

The discontinuities can then be shown to occur precisely when  $x \in \mathbb{Q} \cap [0, 1]$ . But, as we have shown, the rational numbers in  $[0, 1]$  form a set with measure 0 (but has undefined volume). Hence, by our theorem for integrability, this  $f$  is integrable.

We will exploit this integrability condition to solve the following problem.

---

<sup>1</sup>This goes by many names. See here: [https://en.wikipedia.org/wiki/Thomae's\\_function](https://en.wikipedia.org/wiki/Thomae's_function).

**Problem:** Give an example of a function  $f: \mathbb{R}^2 \rightarrow \mathbb{R}$  which is Riemann integrable but such that the set:

$$S = \{x \in \mathbb{R} \mid h(y) = f(x, y): \mathbb{R} \rightarrow \mathbb{R} \text{ is not Riemann integrable}\},$$

is dense in some interval, say  $[0, 1]$ .

*Solution:* This is *almost* impossible. Let us see why.

We want a function  $f$  that is piecewise such that when  $x \in S$ ,  $f(x, y) = h(y)$ , where  $h$  is not integrable. If  $h$  is not integrable, then the supremum of  $h$  cannot be made to be arbitrarily close to the infimum of  $h$  no matter how small an interval we choose.

In  $\mathbb{R}^2$ , any dyadic cube containing a point  $x \in S$  has a cross-section along which  $f$  is identical to  $h$ . This means that over this cross-section, the infimum and the supremum of  $f$  cannot get arbitrarily close. Since the supremum over the entire cube is bounded below by the supremum over the cross-section, and the infimum over the entire cube is bounded above by the infimum over the cross-section, the supremum and infimum over the cube cannot be made to be arbitrarily close either! Which means that  $f$  cannot be integrable in  $\mathbb{R}^2$ .

So then how can an  $f$  which satisfies the given properties exist? The key is to realize that the boundary of dyadic cubes cannot contribute to the volume.<sup>2</sup>

If we define  $h(y)$  such that its supremum occurs precisely on the *boundaries* of the dyadic cubes of which it is a cross-section of, we can say that the effective suprema over those dyadic cubes *are not* bounded by the suprema on the cross-sections, since we can just take the cubes to be open, or throw away any suprema that occur on the boundaries due to the theorem we have proven which allows us to neglect the boundaries of dyadic cubes. The supremum over each cube can then possibly converge to the infimum over the cube, which would imply that  $f$  is integrable in  $\mathbb{R}^2$ .

The question then becomes, how can we construct an  $f$  such that when  $x \in S$ , we have that  $h(y)$  attains suprema on  $y$ -values that correspond to the boundaries of cubes in a dyadic paving? The first step is to identify where the boundaries of the dyadic cubes lie. They lie on lines of the form  $y = n + \frac{k}{2^N}$  for integers  $N \geq 0$ ,  $0 \leq k \leq 2^N - 1$ , and  $n$  (or the other boundaries which we will actually use, that are of the form  $x = \frac{2j-1}{2^M}$ <sup>3</sup>).

This form of  $y$  (or  $x$ ) can always be reduced to a rational number with a power of 2 in the denominator and an odd number in the numerator. So then what if defined  $h$  so that it was uniformly zero unless  $x$  was of the form

$$x = \frac{2j-1}{2^M}$$

for integers  $M, j$ ? We have to make sure that these suprema decrease over smaller and smaller

---

<sup>2</sup>Formally, this is because of a result that states that the bounded part of the graph of an integrable function has volume 0.

<sup>3</sup>Note that this is every other vertical boundary.

dyadic cubes in  $\mathbb{R}^2$ . A natural choice is to set them to  $2^{-M}$ . The larger  $M$  is, the smaller  $2^{-M}$  is, and the more numbers of the form  $\frac{2j-1}{2^M}$  we can fit in an interval. We can have that  $h(y) = 2^{-M}$  only when  $y$  is irrational, which forms a set of discontinuities with nonzero measure, which then forces the function to be not integrable in such cases. For completeness, we will show that  $[0, 1] \setminus \mathbb{Q}$  does not have measure zero.

Suppose to the contrary that  $[0, 1] \setminus \mathbb{Q}$  does have measure zero. Then, we can find some intervals  $B_i$  that contain the irrationals in  $[0, 1]$  and have total length less than  $\frac{\epsilon}{2}$ . Since we have already shown that the rationals in  $[0, 1]$  have measure zero, there are some other intervals,  $C_i$ , that contain the rationals in  $[0, 1]$  and have total length less than  $\frac{\epsilon}{2}$ . But now, the union of the  $B_i$  and the  $C_i$  must contain the union of the rationals and irrationals in  $[0, 1]$ , namely  $[0, 1]$  itself, and have a total length less than  $\epsilon$ . But a set containing  $[0, 1]$  must have a length of at least 1, and cannot be made arbitrarily small, contradiction. So the irrationals in  $[0, 1]$  have nonzero measure, as desired.

It is simple to show that the set of numbers of the form  $\frac{2j-1}{2^M}$  is dense over the interval  $[0, 1]$ .<sup>4</sup> So it is fine to define  $S$  to be the set where  $x$  is of the form  $\frac{2j-1}{2^M}$  in some interval.

Putting this all together we can define

$$f(x, y) = \begin{cases} 2^{-M} \mathbf{1}_{[0,1] \setminus \mathbb{Q}}(y) & \text{if } x \in S \text{ with } x = \frac{2j-1}{2^M} \text{ for } M, j \in \mathbb{Z} \\ 0 & \text{otherwise,} \end{cases}$$

where

$$S = \left\{ \frac{2j-1}{2^M} \mid M, j \in \mathbb{N} \text{ and } 1 \leq j \leq 2^{M-1} \right\}.$$

Holding  $x$  constant, this has discontinuities if  $x \in S$  for irrational  $y$  in  $[0, 1]$ . Since this set of irrationals form a set without measure zero, the function is not integrable holding  $x$  constant in  $S$ .

On the other hand,  $f$  is 0 everywhere in  $\mathbb{R}^2$  except when  $x \in S$  and  $y \in [0, 1] \setminus \mathbb{Q}$ . But due to the way  $S$  is defined, these points that are not in the support of  $f$  are always on the boundaries of dyadic cubes for some large enough  $N$ . They do not contribute to the supremum over open cubes, and the supremum of the form  $2^{-M}$  approaches 0, the infimum over any cube, as  $N \rightarrow \infty$ . So this  $f$  is integrable in  $\mathbb{R}^2$ .  $\square$

As we have seen, a function that is continuous almost everywhere is “nice” enough to be Riemann integrated. This is just one of the places we will see that satisfying some conditions almost everywhere allows us to make powerful statements.

---

<sup>4</sup>This is essentially because we can make the distance between consecutive numbers of this form arbitrarily small by choosing  $M$  large enough.

### 3 The Weaknesses of the Riemann Integral

We already know that the Riemann integral is only defined for bounded functions with bounded support. But this egregious limitation is not the only downfall of Riemann integration. Riemann integrals are also not very well-behaved under limits. We will show what this means by introducing a theorem that shows the extent to which Riemann integrals “play nice” under limits.

**Theorem** (Convergence for Riemann Integrals). *Let  $k \mapsto f_k$  be a sequence of integrable functions  $\mathbb{R}^n \rightarrow \mathbb{R}$ , all with support in a fixed ball  $B \subset \mathbb{R}^n$ , and converging uniformly to a function  $f$ . Then,  $f$  is integrable and*

$$\int_{\mathbb{R}^n} f(\vec{x}) |d^n \vec{x}| = \lim_{k \rightarrow \infty} \int_{\mathbb{R}^n} f_k(\vec{x}) |d^n \vec{x}|.$$

In other words, we can interchange a limit and an integral, given that certain conditions are satisfied. The proof of this theorem is fairly straightforward.

*Proof.* For  $\epsilon > 0$ , choose  $K$  such that  $\forall k > K$ , we have

$$\sup_{x \in \mathbb{R}^n} |f(\vec{x}) - f_k(\vec{x})| < \frac{\epsilon}{3 \operatorname{vol}_n(B)}.$$

From this, we can make statements about the  $N^{\text{th}}$  lower sums and upper sums,  $L_N$  and  $U_N$ , respectively. For  $k > K$ , we have

$$L_N(f_k) - L_N(f) \leq |L_N(f) - L_N(f_k)| < \operatorname{vol}_n(B) \sup_{x \in \mathbb{R}^n} |f(\vec{x}) - f_k(\vec{x})| = \frac{\epsilon}{3},$$

$$U_N(f) - U_N(f_k) \leq |U_N(f) - U_N(f_k)| < \operatorname{vol}_n(B) \sup_{x \in \mathbb{R}^n} |f(\vec{x}) - f_k(\vec{x})| = \frac{\epsilon}{3}.$$

This inequalities rearrange to

$$L_N(f) > L_N(f_k) - \frac{\epsilon}{3},$$

$$U_N(f) < U_N(f_k) + \frac{\epsilon}{3}.$$

Now, since the  $f_k$  are integrable, we can choose  $N$  large enough so that  $U_N(f_k) - L_N(f_k) < \frac{\epsilon}{3}$ . So when we subtract our two inequalities, we obtain

$$U_N(f) - L_N(f) < U_N(f_k) - L_N(f_k) + \frac{2\epsilon}{3} < \epsilon.$$

□

This is quite a powerful tool, as we will see. Before demonstrating the weakness of this theorem, we will demonstrate its strength by solving a problem with it.

**Problem:** For each positive integer  $n$ , let  $U_n$  denote the area enclosed by the graph of  $|x|^n + |y|^n = 1$ . Compute

$$\lim_{n \rightarrow \infty} n^2(4 - U_n).$$

*Solution:* The graph is a special type of curve, known as a *superellipse*, or the *Lamé curve*.<sup>5</sup> Some examples are shown below.

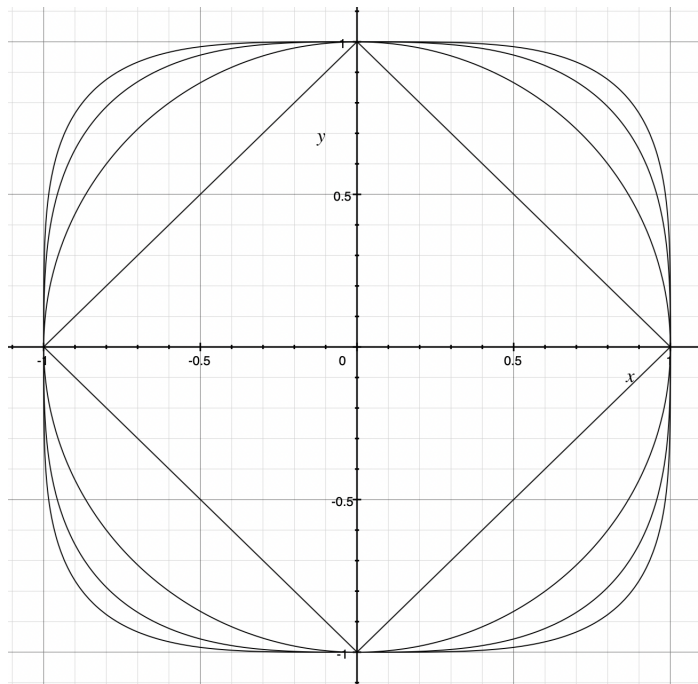


Figure 1: Superellipses for  $n = 1, 2, 3, 4$ . Observe that the  $n = 1$  case is just a square and the  $n = 2$  case is just a circle.

Let  $L$  be the desired limit. Observe that by symmetry (sign changes vanish in the absolute value),  $U_n$  is four times the area enclosed by the region in the first quadrant. That is,

$$U_n = 4 \int_0^1 (1 - x^n)^{\frac{1}{n}} dx.$$

We can expand the integrand as a binomial series. This gives us

$$U_n = 4 \int_0^1 \sum_{k=0}^{\infty} \binom{\frac{1}{n}}{k} (-x^n)^k dx.$$

We can compute the radius of convergence of the binomial series. We first find that the absolute

<sup>5</sup>The  $n = 4$  case is known as the *squircle* for reasons that are clear in Figure 1.



value of the ratio of consecutive terms is

$$\left| \frac{\binom{\frac{1}{n}}{k+1} (-x^n)^{k+1}}{\binom{\frac{1}{n}}{k} (-x^n)^k} \right| = \left| \frac{1 - kn}{n(k+1)} x^n \right|.$$

Now, we take the limit  $k \rightarrow \infty$  to find

$$\lim_{k \rightarrow \infty} \left| \frac{1 - kn}{n(k+1)} x^n \right| = |x^n| < 1.$$

So our binomial series converges absolutely for  $|x| < 1$ . Furthermore, the binomial series converges at the endpoint of  $x = 1$ , hence the series converges for  $x \in [0, 1]$  which is exactly the relevant  $x$  that we need for the integration. We would like to interchange the sum and the integral. To do this, we define a sequence of functions  $N \mapsto f_N$  such that

$$f_N(x) = \mathbf{1}_{[0,1]}(x) \sum_{k=0}^N \binom{\frac{1}{n}}{k} (-x^n)^k.$$

Clearly, we have  $\lim_{N \rightarrow \infty} f_N(x) = \mathbf{1}_{[0,1]}(x)(1-x^n)^{\frac{1}{n}}$ . Furthermore, by absolute convergence,  $\forall \epsilon > 0$ , we have

$$N > M(x, \epsilon) \Rightarrow \left| (1-x^n)^{\frac{1}{n}} - f_N(x) \right| < \epsilon, \quad \forall x \in [0, 1],$$

where  $M(x, \epsilon)$  is some function of  $x$  and  $\epsilon$ , and is the smallest integer for which the above statement holds true. Now we show that  $M(x)$  is bounded for any fixed  $\epsilon$ .<sup>6</sup> Fix an  $\epsilon > 0$ .

Since  $(1-x^n)^{\frac{1}{n}}$  and  $f_N(x)$  are continuous, we have that their difference is continuous, so for any  $x \in [0, 1]$ , we can find some neighborhood of  $x$  (an open set  $U$  containing  $x$ ) such that  $\forall y \in U$ , we have  $\left| (1-y^n)^{\frac{1}{n}} - f_N(y) \right| < \epsilon$ .

Doing this for every  $x \in [0, 1]$  gives us a covering of  $[0, 1]$ . But since  $[0, 1]$  is compact, by the Heine-Borel theorem, from that covering we can extract a finite subcovering of  $[0, 1]$ . The open sets that remain in our subcovering correspond to a finite set  $S = \{x_i\}$  with  $x_i \in [0, 1]$ . Hence  $M(x) \leq \max_{x_i \in S} M(x_i)$  as desired.

Now, we simply define

$$\overline{M}(\epsilon) = \sup_{x \in [0,1]} M(x, \epsilon),$$

and it is clear that as long as  $N > \overline{M}(\epsilon)$ , we have  $\left| (1-x^n)^{\frac{1}{n}} - f_N(x) \right| < \epsilon$  for all  $x \in [0, 1]$ . So  $f_N$  converges uniformly to  $\mathbf{1}_{[0,1]}(x)(1-x^n)^{\frac{1}{n}}$ . Furthermore, the supports of the  $f_N$  are obviously in a fixed ball due to the indicator function.

<sup>6</sup>See: <https://math.stackexchange.com/a/3747711/419177>.

So we can apply our theorem to write

$$\begin{aligned}
\frac{U_n}{4} &= \int_{\mathbb{R}} \mathbf{1}_{[0,1]}(x)(1-x^n)^{\frac{1}{n}} dx \\
&= \lim_{N \rightarrow \infty} \int_{\mathbb{R}} f_N(x) dx \\
&= \lim_{N \rightarrow \infty} \int_{\mathbb{R}} \mathbf{1}_{[0,1]}(x) \sum_{k=0}^N \binom{\frac{1}{n}}{k} (-x^n)^k dx \\
&= \lim_{N \rightarrow \infty} \sum_{k=0}^N \int_0^1 \binom{\frac{1}{n}}{k} (-x^n)^k dx \\
&= \sum_{k=0}^{\infty} \int_0^1 \binom{\frac{1}{n}}{k} (-x^n)^k dx.
\end{aligned}$$

Now the rest is a computation. Using the definition of the generalized binomial coefficient, we can write

$$\begin{aligned}
\frac{U_n}{4} &= 1 + \sum_{k=1}^{\infty} \int_0^1 \frac{(-x^n)^k}{k!} \prod_{j=0}^{k-1} \left(\frac{1}{n} - j\right) dx \\
&= 1 + \sum_{k=1}^{\infty} \frac{(-1)^k}{(nk+1)k!} \prod_{j=0}^{k-1} \left(\frac{1}{n} - j\right).
\end{aligned}$$

Hence,

$$\frac{1}{4}n^2(4 - U_n) = \frac{n}{n+1} + \sum_{k=2}^{\infty} \frac{(-1)^{k+1}n}{(nk+1)k!} \prod_{j=1}^{k-1} \left(\frac{1}{n} - j\right).$$

Now we can take the limit.

$$\begin{aligned}
\frac{L}{4} &= \lim_{n \rightarrow \infty} \frac{1}{4}n^2(4 - U_n) \\
&= 1 + \sum_{k=2}^{\infty} \frac{(-1)^{k+1}}{k \cdot k!} \prod_{j=1}^{k-1} (-j) \\
&= 1 + \sum_{k=2}^{\infty} \frac{(k-1)!}{k \cdot k!} \\
&= 1 + \sum_{k=2}^{\infty} \frac{1}{k^2} \\
&= 1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \dots
\end{aligned}$$

But this is just the Basel problem. By Euler, we conclude  $L = 4 \cdot \frac{\pi^2}{6} = \boxed{\frac{2\pi^2}{3}}$ .

If we look close at what went on in this problem, we'll see that our theorem enabled us to exchange the limit and integral. But we were lucky. We only needed to integrate over  $(0, 1)$ , which happened to be precisely where our series expansion converges absolutely. This guaranteed that our sequence of functions converged uniformly to the integrand.

Furthermore, since all the action was occurring in  $(0, 1)$ , it was very easy to define that interval to be the fixed ball in which all of the functions had their supports. But this may not be the case in general.

We cannot do anything about having to have bounded supports. This is simply a fundamental limitation of Riemann integration. We can, however, relax the uniform convergence condition. For a price. The trade-off is that we must know whether our limiting function is Riemann integrable beforehand. Observe that in our previous theorem, the conditions imply that the limit of the sequence is itself integrable. This is no longer the case with the new theorem that we will introduce.

**Theorem** (Dominated Convergence for Riemann Integrals). *Let  $f_k: \mathbb{R}^n \rightarrow \mathbb{R}$  be a sequence of Riemann integrable functions. Suppose there exists  $R$  such that all  $f_k$  have their support in  $B_R(\vec{0})$  and satisfy  $|f_k| \leq R$ . Let  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  be a Riemann integrable function such that  $f(\vec{x}) = \lim_{k \rightarrow \infty} f_k(\vec{x})$  almost everywhere. Then,*

$$\int_{\mathbb{R}^n} f(\vec{x}) |d^n \vec{x}| = \lim_{k \rightarrow \infty} \int_{\mathbb{R}^n} f_k(\vec{x}) |d^n \vec{x}|.$$

The biggest issue of this form of DCT is that we must know that  $f$  is integrable beforehand. And as it turns out,  $f$  is hardly ever Riemann integrable if  $f_k$  does not converge to it uniformly, in which case we can just use the first theorem. It is somewhat of the same conundrum we have when we try to prove that a sequence converges. Using the definition of convergence to prove convergence requires knowing the limit of the sequence beforehand, which we often times do not. However, in a complete metric space, we can establish the equivalence of convergent sequences and Cauchy sequences, with the Cauchy condition having the benefit of not requiring knowledge of the limit beforehand.<sup>7</sup>

We will see that DCT takes on a much nicer form for Lebesgue integrals, which we define in the next section.

---

<sup>7</sup>See *The Contraction Mapping Theorem*.

## 4 Defining the Lebesgue Integral

Before we explicitly define the Lebesgue integral, we will establish its uniqueness. We will not be proving uniqueness here. The proof is technical and is best accessible in textbooks. So, we will briefly outline the results that imply uniqueness. From now on we will abbreviate “Riemann integrable” as *R-integrable* and “Lebesgue integrable” as *L-integrable*.

**Theorem** (Convergence Almost Everywhere). *Let  $k \mapsto f_k$  be a sequence of  $R$ -integrable functions on  $\mathbb{R}^n$  such that*

$$\sum_{k=1}^{\infty} \int_{\mathbb{R}^n} |f_k(\vec{x})| |d^n \vec{x}| < \infty.$$

*Then the series  $\sum_{k=1}^{\infty} f_k(\vec{x})$  converges almost everywhere.*

**Theorem** (Uniqueness of the Lebesgue Integral). *Let  $k \mapsto f_k$  and  $k \mapsto g_k$  be two sequences of  $R$ -integrable functions such that*

$$\sum_{k=1}^{\infty} \int_{\mathbb{R}^n} |f_k(\vec{x})| |d^n \vec{x}|, \sum_{k=1}^{\infty} \int_{\mathbb{R}^n} |g_k(\vec{x})| |d^n \vec{x}| < \infty,$$

*and  $\sum_{k=1}^{\infty} f_k(\vec{x}) = \sum_{k=1}^{\infty} g_k(\vec{x})$  almost everywhere. Then,*

$$\sum_{k=1}^{\infty} \int_{\mathbb{R}^n} f_k(\vec{x}) |d^n \vec{x}| = \sum_{k=1}^{\infty} \int_{\mathbb{R}^n} g_k(\vec{x}) |d^n \vec{x}|.$$

That is, the infinite series of the integrals of the absolute values of the functions converge, which means that the infinite series of the integrals of the functions must converge. Moreover, if two sequences of functions satisfy that property and the infinite series of the sequences are equal almost everywhere, then the infinite series of their integrals are equal. In other words, if we define

$$f = \sum_{k=1}^{\infty} f_k = \sum_{k=1}^{\infty} g_k$$

almost everywhere, then we can define an integral of  $f$  to be entirely independent of the choice of the series that converges to  $f$ . Assuming existence, the result of this operation which we will define on  $f$  is unique.

**Definition.** *Let  $k \mapsto f_k$  be a sequence of  $R$ -integrable functions such that*

$$\sum_{k=1}^{\infty} \int_{\mathbb{R}^n} |f_k(\vec{x})| |d^n \vec{x}| < \infty.$$

*Then the Lebesgue integral of the function  $f$ , where  $f$  is almost everywhere equal to  $\sum_{k=1}^{\infty} f_k$ , is*

$$\int_{\mathbb{R}^n} f(\vec{x}) |d^n \vec{x}| \equiv \sum_{k=1}^{\infty} \int_{\mathbb{R}^n} f_k(\vec{x}) |d^n \vec{x}|.$$

From this definition, we can immediately see that if a function  $f$  is R-integrable, then it is L-integrable, and the Lebesgue integral is just equal to the Riemann integral. To see this, just take  $f_1 = f$  and  $f_k = 0$  for  $k > 1$ . So the Lebesgue integral fundamentally *generalizes* the Riemann integral.

The first advantage of the Lebesgue integral that we will investigate is its ability to work on functions with unbounded supports. This is an extraordinarily important idea. Many integrals of practical interest (in statistics and physics) are over infinitely large domains. To demonstrate this in action, we solve a problem—one that is impossible with the Riemann integral.

**Problem:** Prove that for all polynomials  $p$ , the Lebesgue integral

$$\int_{\mathbb{R}^n} p(\vec{x}) e^{-|\vec{x}|^2} |\mathrm{d}^n \vec{x}|$$

exists.

*Solution:* It is a quick exercise to show that the Lebesgue integral is linear. This means that it suffices to prove the statement for  $p(\vec{x})$  that consist of just a single term. So suppose that  $p(\vec{x})$  is a single term polynomial with degree  $d$ . Let the integrand be  $f(\vec{x})$ . Let  $B_0$  be the unit ball centered at the origin in  $\mathbb{R}^n$ . Let

$$B_1 = \{\vec{x} \in \mathbb{R}^n \mid 1 < |\vec{x}| \leq 2\}.$$

For natural numbers  $k > 1$ , define

$$B_k = \left\{ \vec{x} \in \mathbb{R}^n \mid \sqrt{k \log 3} < |\vec{x}| \leq 2\sqrt{k \log 3} \right\}.$$

Now consider the functions

$$g_k(\vec{x}) = \mathbf{1}_{B_k}(\vec{x}) f(\vec{x}).$$

Observe that since some of the  $B_k$  overlap, we must have that  $\sum_{k=0}^{\infty} g_k(\vec{x}) \geq f(\vec{x})$ . In particular, the sum is equal to  $\alpha(\vec{x}) f(\vec{x})$ , where  $\alpha(\vec{x})$  counts how many of the sets  $B_k$  that  $\vec{x}$  is an element of.

If we similarly defined some  $f_k$  to equal  $f$  times the indicator function of some partitions  $C_k$  of  $\mathbb{R}^n$  where there is no overlap between the  $C_k$ , and the  $C_k$  are determined by intervals of distances from the origin like the  $B_k$ , we would have that  $\sum_{k=1}^{\infty} f_k(\vec{x}) = f(\vec{x})$ . Since the  $B_k$  overlap with each other but the  $C_k$  do not, for any positive integer  $N$ , there exists an  $M$  such that

$$\sum_{k=1}^N \int_{\mathbb{R}^n} |f_k(\vec{x})| |\mathrm{d}^n \vec{x}| \leq \sum_{k=1}^M \int_{\mathbb{R}^n} |g_k(\vec{x})| |\mathrm{d}^n \vec{x}|.$$

In particular, we can take any  $M$  large enough such that  $\sup_{\vec{x} \in B_M} |\vec{x}| \geq \sup_{\vec{x} \in C_N} |\vec{x}|$ . So, it suffices to prove the stronger result

$$\sum_{k=1}^{\infty} \int_{\mathbb{R}^n} |g_k(\vec{x})| |\mathrm{d}^n \vec{x}| < \infty.$$

To do this, we write

$$\int_{\mathbb{R}^n} |g_k(\vec{x})| |\mathrm{d}^n \vec{x}|,$$

for  $k > 1$ . For  $k = 0$  and  $k = 1$ , this integral is well-defined, since the integrand is continuous and bounded with bounded support. For  $k > 1$ , we introduce the change of variables parameterization  $\Phi_k$  that maps  $\vec{x} \mapsto \vec{x} \sqrt{k \log 3}$ . Observe that this sends  $B_1$  to  $B_k$ . In matrix form, we can write the

derivative of  $\Phi_k$  as

$$[D\Phi_k] = I_n \sqrt{k \log 3} = \begin{bmatrix} \sqrt{k \log 3} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sqrt{k \log 3} \end{bmatrix},$$

where  $I_n$  is the  $n \times n$  identity matrix. From this, we can easily compute that  $\det [D\Phi_k] = (k \log 3)^{\frac{n}{2}}$ . So by change of variables, for  $k > 1$ , we have

$$\begin{aligned} \int_{\mathbb{R}^n} |g_k(\vec{x})| |d^n \vec{x}| &= \int_{B_k} |f(\vec{x})| |d^n \vec{x}| \\ &= |\det [D\Phi_k]| \int_{B_1} |(f \circ \Phi_k)(\vec{x})| |d^n \vec{x}| \\ &= (k \log 3)^{\frac{n}{2}} \int_{B_1} \left| p\left(\vec{x} \sqrt{k \log 3}\right) e^{-|\vec{x} \sqrt{k \log 3}|^2} \right| |d^n \vec{x}| \\ &= (k \log 3)^{\frac{n}{2}} \int_{B_1} \left| (k \log 3)^{\frac{d}{2}} p(\vec{x}) 3^{-k|\vec{x}|^2} \right| |d^n \vec{x}| \\ &= (k \log 3)^{\frac{n+d}{2}} \int_{B_1} 3^{-k|\vec{x}|^2} |p(\vec{x})| |d^n \vec{x}| \\ &\leq \frac{k^{\frac{n+d}{2}} (\log 3)^{\frac{n+d}{2}}}{3^k} \int_{B_1} |p(\vec{x})| |d^n \vec{x}|. \end{aligned}$$

<sup>8</sup> Since  $|p(\vec{x})|$  clearly integrable over  $B_1$ ,  $(\log 3)^{\frac{n+d}{2}} \int_{B_1} |p(\vec{x})| |d^n \vec{x}|$  is just some constant, say  $A$ . Hence,

$$\sum_{k=1}^{\infty} \int_{\mathbb{R}^n} |g_k(\vec{x})| |d^n \vec{x}| \leq \int_{\mathbb{R}^n} |g_0(\vec{x})| |d^n \vec{x}| + \int_{\mathbb{R}^n} |g_1(\vec{x})| |d^n \vec{x}| + A \sum_{k=2}^{\infty} \frac{k^{\frac{n+d}{2}}}{3^k}.$$

So all that is left is to check the convergence of the series

$$S = \sum_{k=2}^{\infty} \frac{k^{\frac{n+d}{2}}}{3^k}.$$

We take the limit of the absolute value of the ratio of consecutive terms and apply the ratio test:

$$\lim_{k \rightarrow \infty} \frac{1}{3} \left( \frac{k+1}{k} \right)^{\frac{n+d}{2}} = \frac{1}{3} < 1.$$

By the ratio test,  $S$  converges and we are done.  $\square$

So the Lebesgue integral is robust enough to handle functions with unbounded supports. This gives birth to the generalizations of improper integrals from elementary calculus. Next, we will see that the Lebesgue integral gives us a very useful form of DCT, and in doing so, paves the way for us to prove Feynman's trick.

---

<sup>8</sup>In the last line we note that  $3^{-k|\vec{x}|^2}$  is maximized on  $B_1$  when  $|\vec{x}| = \inf_{\vec{x} \in B_1} |\vec{x}| = 1$ .

## 5 A Better DCT and Feynman's Trick

We introduce a much better form of DCT with the Lebesgue integral.

**Theorem** (Dominated Convergence for Lebesgue Integrals). *Let  $k \mapsto f_k$  be a sequence of L-integrable functions that converges pointwise almost everywhere to some function  $f$ . Suppose that there exists an L-integrable function  $F: \mathbb{R}^n \rightarrow \mathbb{R}$  such that  $|f_k(\vec{x})| \leq F(\vec{x})$  for almost all  $\vec{x}$ . Then  $f$  is L-integrable and its integral is*

$$\int_{\mathbb{R}^n} f(\vec{x}) |d^n \vec{x}| = \lim_{k \rightarrow \infty} \int_{\mathbb{R}^n} f_k(\vec{x}) |d^n \vec{x}|.$$

This resolves the issues we had with DCT for Riemann integrals. It does not require that we know if the limit of the sequence of  $f_k$  is integrable, *and* it allows the sequence to converge just pointwise instead of uniformly! All for the price of having each  $f_k$  bounded by some L-integrable function.

This form of DCT allows us to prove a result about differentiating an integral. We state it as follows.

**Theorem** (Feynman's Trick). *Let  $f(t, \vec{x}): \mathbb{R}^{n+1} \rightarrow \mathbb{R}$  be a function such that for each fixed  $t$ , the integral*

$$F(t) = \int_{\mathbb{R}^n} f(t, \vec{x}) |d^n \vec{x}|$$

*exists. Suppose that  $\frac{\partial f}{\partial t}$  exists for almost all  $\vec{x}$ . If there exists  $\epsilon > 0$  and an L-integrable function  $g$  such that for all  $s \neq t$ ,*

$$|s - t| < \epsilon \Rightarrow \left| \frac{f(s, \vec{x}) - f(t, \vec{x})}{s - t} \right| \leq g(\vec{x}),$$

*then  $F$  is differentiable, and its derivative is*

$$\frac{dF}{dt} = \int_{\mathbb{R}^n} \frac{\partial}{\partial t} f(t, \vec{x}) |d^n \vec{x}|.$$

In other words, we can exchange the derivative and the integral under certain circumstances. As we will see in the proof, these conditions are necessary for DCT to be applied.

*Proof.* We compute

$$\frac{dF}{dt} = \lim_{h \rightarrow 0} \frac{F(t+h) - F(t)}{h} = \lim_{h \rightarrow 0} \int_{\mathbb{R}^n} \frac{f(t+h, \vec{x}) - f(t, \vec{x})}{h} |d^n \vec{x}|.$$

Now if we define  $s_1, s_2, \dots$  to be a sequence such that  $|s_1 - t|, |s_2 - t|, \dots$  is a strictly decreasing sequence, and let

$$f_k(\vec{x}) = \frac{f(s_k, \vec{x}) - f(t, \vec{x})}{h},$$

it is clear that  $\lim_{k \rightarrow \infty} f_k(\vec{x}) = \frac{\partial f}{\partial t}$ . Furthermore, since we are given that there exists  $|f_k(\vec{x})| \leq g(\vec{x})$ , for some L-integrable  $g$ , we can apply DCT for Lebesgue integrals to continue our calculation of



$F'(t)$  by moving the limit in the integral:

$$\frac{dF}{dt} = \int_{\mathbb{R}^n} \lim_{h \rightarrow 0} \frac{f(t+h, \vec{x}) - f(t, \vec{x})}{h} |d^n \vec{x}| = \int_{\mathbb{R}^n} \frac{\partial}{\partial t} f(t, \vec{x}) |d^n \vec{x}|.$$

□

This powerful technique enables us to compute certain integrals by introducing a parameterization and differentiating with respect to the parameter. The technique, commonly known as *Feynman's trick*, can be used to find many practical integrals, such as this one which has relevance to statistical mechanics.

The Maxwell-Boltzmann distribution is a probability distribution. It is defined by

$$f(v) = \frac{4}{\sqrt{\pi}} \left( \frac{m}{2k_B T} \right)^{3/2} v^2 e^{-\frac{mv^2}{2k_B T}},$$

where  $\int_a^b f(v) dv$  gives the probability that a randomly chosen molecule in an ideal gas with molecular mass  $m$  and temperature  $T$  has a speed between  $a$  and  $b$ .  $k_B$  is the Boltzmann constant. A commonly used measure of the “average” speed of a molecule in a gas is the quadratic mean of the speeds, also known as the root-mean square (RMS) speed.<sup>9</sup> To compute this from the Maxwell-Boltzmann distribution, we must evaluate

$$v_{\text{rms}} = \sqrt{\int_0^\infty v^2 f(v) dv}.$$

To accomplish, we will find a more general result.

---

<sup>9</sup>It can be shown that  $\text{QM} \geq \text{AM}$ . This is part of the *mean inequality chain* which states  $\text{QM} \geq \text{AM} \geq \text{GM} \geq \text{HM}$ . See <https://mathwithandy.weebly.com/blog/qm-am-inequality>.

**Problem:** For natural numbers  $n$ , compute

$$\int_0^\infty x^n e^{-\alpha x^2} dx.$$

*Solution:* First, we define

$$I(\alpha) = \int_0^\infty e^{-\alpha x^2} dx$$

$$J(\alpha) = \int_0^\infty x e^{-\alpha x^2} dx.$$

Observe that by Fubini's theorem,<sup>10</sup> we can rewrite  $I(\alpha)$  as

$$\begin{aligned} I(\alpha) &= \sqrt{\left(\int_0^\infty e^{-\alpha x^2} dx\right) \left(\int_0^\infty e^{-\alpha y^2} dy\right)} \\ &= \sqrt{\int_0^\infty \int_0^\infty e^{-\alpha(x^2+y^2)} dx dy} \\ &= \sqrt{\int_0^{\frac{\pi}{2}} \int_0^\infty r e^{-\alpha r^2} dr d\theta}, \end{aligned}$$

where we invoke the polar substitution in the last line. The first integral can be evaluated with the  $u$ -substitution,  $u = r^2$ . Then, we have

$$\int_0^\infty r e^{-\alpha r^2} dr = \frac{1}{2} \int_0^\infty e^{-\alpha u} du = -\frac{1}{2\alpha} e^{-\alpha u} \Big|_0^\infty = \frac{1}{2\alpha} = J(\alpha).$$

So we can continue for  $I(\alpha)$ :

$$I(\alpha) = \sqrt{\int_0^{\frac{\pi}{2}} \frac{1}{2\alpha} d\theta} = \frac{1}{2} \sqrt{\frac{\pi}{\alpha}}.$$

So both  $I(\alpha)$  and  $J(\alpha)$  have simple closed forms. We can compute the derivatives of  $I$  and  $J$  by explicitly differentiating their closed forms. Inductively, we can see that doing this gives us

$$\frac{d^m I}{d\alpha^m} = \left(-\frac{1}{2}\right)^{m+1} \sqrt{\frac{\pi}{\alpha^{2m+1}}} \prod_{i=0}^m (2i-1),$$

$$\frac{d^m J}{d\alpha^m} = (-1)^m \frac{m!}{2\alpha^{m+1}}.$$

Now, we check to see if the integral definitions of  $I(\alpha)$  and  $J(\alpha)$  allow us to use Feynman's trick.

---

<sup>10</sup>We are being handwavy here. There is a precise form of Fubini's theorem for Lebesgue integrals that is nicer than the Riemann form (similar to DCT).

For  $I$ , we wish to find an L-integrable function  $g_I(\vec{x})$  such that

$$|s - \alpha| < \epsilon \Rightarrow \left| \frac{e^{-sx^2} - e^{-\alpha x^2}}{s - \alpha} \right| \leq g_I(\vec{x}).$$

We can rearrange this to

$$\left| e^{-sx^2} - e^{-\alpha x^2} \right| \leq |s - \alpha| g_I(\vec{x}) < \epsilon g_I(\vec{x}).$$

This is true if we choose  $g_I(\vec{x}) = \frac{1}{\epsilon} \left( e^{-(\alpha-\epsilon)x^2} + e^{-\alpha x^2} \right)$ . Since  $\alpha$  must be positive for our integrals to converge, we can choose  $\epsilon$  small enough so that  $\alpha - \epsilon > 0$ . Then, a small modification of our proof in the previous section that  $p(\vec{x})e^{-|\vec{x}|^2}$  is L-integrable shows that this choice of  $g_I$  is also L-integrable. We can also find a similar function  $g_J$  for  $J$ .

Hence, we can also compute the derivatives of  $I$  and  $J$  by differentiating their integral definitions and use Feynman's trick to pull the derivatives inside of the integrals. Inductively, this gives us

$$\begin{aligned} \frac{d^m I}{d\alpha^m} &= \int_0^\infty \frac{\partial^m}{\partial \alpha^m} \left( e^{-\alpha x^2} \right) dx = (-1)^m \int_0^\infty x^{2m} e^{-\alpha x^2} dx, \\ \frac{d^m J}{d\alpha^m} &= \int_0^\infty \frac{\partial^m}{\partial \alpha^m} \left( x e^{-\alpha x^2} \right) dx = (-1)^m \int_0^\infty x^{2m+1} e^{-\alpha x^2} dx. \end{aligned}$$

Now we can equate our two forms for our derivatives of  $I$  and  $J$  to conclude that for  $n \in \mathbb{N}$ , we have

$$\boxed{\int_0^\infty x^n e^{-\alpha x^2} dx = \begin{cases} -\left(\frac{1}{2}\right)^{\frac{n}{2}+1} \sqrt{\frac{\pi}{\alpha^{n+1}}} \prod_{i=0}^{\frac{n}{2}} (2i-1) & n \text{ is even} \\ \left(\frac{1}{2\alpha}\right)^{\frac{n+1}{2}} \left(\frac{n-1}{2}\right)! & n \text{ is odd} \end{cases}}.$$

In the context of statistical mechanics, choosing  $n = 4$  gives us that  $v_{\text{rms}} = \sqrt{\frac{3k_B T}{m}}$ .

## 6 Conclusion

The Lebesgue integral is a powerful extension of the Riemann integral. Our treatment here, defining it as an infinite sum of Riemann integrals, is nonstandard. The standard approach involves a more thorough discussion of measure, specifically *nonzero measure*. Measure theory is a vast topic, and there is much more to the Lebesgue integral that we have not covered. Apart from DCT, another useful theorem that we did not discuss is the monotone convergence theorem, which presents another good example of how the Lebesgue integral is well-behaved under limits. In the solution to the last problem, we were not thorough in our application of Fubini's theorem, and in fact the theorem takes on a nicer form for Lebesgue integrals, similar to DCT.

Nonetheless, the utility of Lebesgue integration can be seen, as the integral and its related topics allowed for the solutions of four difficult problems.