Why Lebesgue Integration

R. Scott McIntire

Nov 1, 2024

1 Better Riemann Integration Convergence Theorems

Suppose the sequence of functions, f_n, converge point-wise to a function f. We wish to find minimal conditions to place on f_n so that the following occurs:

Theorem?

$$\lim_{n \to \infty} \int f_n(x)\, dx \quad = \quad \int \lim_{n \to \infty} f_n(x)\, dx \tag{1}$$

Like any good technology this simplifies the work of mathematicians and applied scientists. For instance, suppose the f_n converges to the function f which is the zero function. Also suppose that it is difficult or impossible to find an analytic expression for some, most, or all of the integrals of f_n. An improved convergence theorem could help us to conclude that the limit of the integrals of the functions is just 0 without ever having to attempt the integrations; or, derive upper bounds on the integrals to show that the sequence of integrals goes to 0.

However, a typical condition is that the f_n converge uniformly to f on a close bounded interval. But this rarely happens in many real world examples and you can imagine that mathematicians back in the day were getting complaints about such restrictive conditions. The people who want a better convergence theorem say that many times they don't have this strong condition and yet (1) holds. Unfortunately, other times it does not. So it seems that there are conditions less rigid than uniform convergence of the sequence of functions which allow (1) to hold.

There are better theorems and they come from a different kind of integration called the Lebesgue (*measure*) Theory of Integration.

Let's stick with Riemann integrals and see if we can find better convergence criterion.

2 Looking for better Convergence Criterion

Examining how Riemann integration works, we see that we need to bound the upper and lower Darboux sums for a given partition of the domain of a function. Let's look for an example where things fail; that is,

where a sequence of functions leads to a function for which we "lose control" constraining the upper and lower bounds of the sub-intervals in a way that the convergence theorem fails. Then we might see what we need to add/impose – in terms of conditions – on the sequence to make theorem (1) true.

Since we are using Riemann functions, let's start with discontinuous functions and see what can go wrong – this should be easier. This will also free us from trying to make complicated continuous functions that behave badly in the limit.

We restrict ourselves to functions in the domain $[0, 1]$. Consider the following sequence of functions:

$$f_n(x) = \begin{cases} 1 & \text{if } x \text{ can be expressed as the fraction, } \frac{p}{q} \text{ for some } p, q \in \mathcal{N} \text{ with } p, q \leq n \text{ ;} \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

You can see that the sequence, f_n, converges point-wise to the function

$$f(x) = \begin{cases} 1 & \text{if } x \text{ is rational;} \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

What can we say about the limit of the integrals of f_n?

Well, each f_n is non-zero at a finite number of points so its integral is 0. Therefore the limit of integrals is 0.

What about the integral of f? We will show that this integral is bounded above and below by 0; meaning that the integral is 0. The below part is obvious. For the upper bound one usually creates a sequence of functions which bound f from above and yet as you move out in the sequence their integrals go to 0. We will do something a little different. We instead construct a sequence of functions g_n which *represent* f in increasing refined ways whose integrals are more amenable to estimation from above.

First, let $\{r_i\}$ be a 1-1 mapping of the positive integers to the rationals. We know we can do this as the rationals are countable.

Consider the sets, $\{E_{i,n}\}$, defined by:

$E_{i,n} = \left[r_i - \frac{1}{n2^i}, r_i + \frac{1}{n2^i} \right]$

A given rational number lies in potentially many of the $E_{i,n}$.

Define the functions, S_i by

$$S_i(x) = \begin{cases} 1 & \text{if } x = r_i; \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Below we make use of the notation:

$$I_X(x) = \begin{cases} 1 & \text{if } x \in X; \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

2

Finally, define, g_n by

$$g_n(x) \;=\; \sum_{i=1}^{\infty} S_i(x) I_{E_{i,n}}(x) \tag{6}$$

It is easy to see that each g_n is a *representation* of the function f; that is, $f \equiv g_n$.

Note the following facts:

$$\int_0^1 S_i(x) I_{E_{i,n}}(x)\, dx \leq \overbrace{1}^{\text{Max Height}} \;*\; 2\overbrace{\frac{1}{n2^i}}^{\text{Width of } E_{i,n}} \;=\; \frac{1}{n2^{i-1}}$$

$$\sum_{i=1}^{\infty} \int_0^1 S_i(x) I_{E_{i,n}}(x)\, dx \;\leq\; \sum_{i=1}^{\infty} \frac{1}{n2^{i-1}}$$

$$=\; \frac{2}{n} \sum_{i=1}^{\infty} \frac{1}{2^i}$$

$$=\; \frac{2}{n}$$

What does this say about the integral of f?

Therefore, given an $n \in \mathcal{N}^+$ for any partition of the interval, $[0,1]$, we can bound the integral of f (assuming it is Riemann integrable) by

$$\int_0^1 f(x)\, dx \;=\; \int_0^1 g_n(x)\, dx$$

$$\leq\; (\text{Darboux Upper Sum})(g_n)$$

$$\leq\; \sum_{i=1}^{\infty} 1 * \text{length}(I_{E_{i,n}}(x))$$

$$\leq\; \frac{2}{n}$$

Since this is true for all $n \in \mathcal{N}^+$, the integral of f must be ≤ 0. But, since f is non-negative, its integral must be ≥ 0. Therefore, the integral of f must be 0.

So, in trying to find an example where the theorem failed, we didn't succeed. Even using a sequence of discontinuous functions we still got convergence without requiring uniform convergence. This is suggestive that we may not, in general, need a very stringent condition for the theorem to hold.

But wait, we are assuming that the function f is a Riemann integrable function. Is it?

It turns out that the answer is NO!

If you look at any given partition of the domain, the approximating lower and upper Darboux sums, are always always 0 and 1 respectively. To see this, note that over any sub-interval of the sums the minimum

3

(infimum) value is always 0 and the maximum (supremum) is always 1 – because on any interval there is always a rational and an irrational number.

Yet the above analysis suggests that its "integral" should exist and its value should be 0.

A contrarian response to the above argument might be: "What I really want is to have a theorem for a sequence of 'nice functions'. I don't care that you have found Riemann integrable functions that don't converge to a Riemann Integrable function".

It turns out; however, that there are sequences of nice functions that converge to a non-Riemann integrable function as well!

So, in addition to the conditions we need to ensure that we can interchange limits with integrals, we also need conditions to ensure that the resulting limiting function is integrable.

This situation is similar to one that you would find yourself in if there were only rational numbers in your worldview. Let's say that at some point you discover an iterative algorithm to solve a problem for which there is no easy analytic formula. Newton's method is such an algorithm. Imagine coming up with a sequence of *numbers* based on Newton's method to find the $\sqrt{x}$. How would you formalize what it means for your sequence to converge – particularly when $x = 2$; that is, when the square root of 2 doesn't exist in your world?

We are in an analogous situation. It suggests that we try a different approach to integration before we even begin to think of better convergence theorems.

However, whatever we come up with, we should get the same value for the integrals of "nice" functions. In particular, all of the previous integration formulas should hold. For instance, it should still be true that: $\int_0^{\pi/2} \sin(x)\, dx = -\cos(x)|_0^{\pi/2} = 0 - (-1) = 1$ In other words, we don't want to lose the connection between differential and integral calculus.

3   The Beginnings of a Lebesgue Theory of Integration

Riemann integration was all about chopping up the *Domain* of a function into a partition of finite intervals and then showing that the difference between the upper and lower Darboux sums of the partition go to zero as the partition is refined. Basically, approximating the function with step functions.

Let's go the other way, based on the example from the previous section, rather than step functions, let's break up the function in terms of its *Range*. In our case there are only two "Range" values: 0 and 1. From this we get an analog of Riemann's step functions – so called characteristic functions of sets:

$$\chi_S = \begin{cases} 1 & \text{if } x \in S; \\ 0 & \text{otherwise.} \end{cases}$$

Although approximating the function, f, from the last section using step functions is unpleasant, it is very easy using characteristic functions – the complexity is buried in the underlying set, S. In our case, we can represent our function *exactly* as a characteristic function: $f(x) = \chi_R(x)$.

Here, R is the set of rationals on interval, $[0, 1]$.

To find the "area" under the curve "$f(x)$" we would like to multiply the height by the "width" of each of the sets. What might this mean? Well we can think of *squeezing* all of the points of each set so that there are no more holes and then taking the length of the resulting interval. But how do we make sense of this?

One way to measure the length of these complicated sets is to try different "coverings" of a set by collections of intervals and relate the "length" of a set to the lengths of the covering intervals. Specifically, as we know how to measure the length of an interval[1], an upper bound on the "length" of an arbitrary set would be, for a given covering set, the sum of the lengths of the covering intervals. We would then define the length of a set to be the infimum over all such covering bounds. We call this "length" the *Lebesgue outer measure*. It is defined more precisely by:

$$\mu^*(S) \quad = \quad \inf_{\bigcup_{i=1}^{\infty} I_i \subseteq S} \sum_{i=1}^{\infty} |I_i| \tag{7}$$

This notion of measure (length) would not be of much use if the Lebesgue outer measure of an interval was not its length. It is not hard to show that it does indeed coincide with length on intervals – and also finite collection of intervals.

Well what is the outer measure of the set R? We claim that the "length" (outer-measure) of the set of rationals on the interval $[0, 1]$ is 0.

Consider the coverings: $I_{(r_i - \frac{\epsilon}{2^i}, r_i + \frac{\epsilon}{2^i})}$. Then for any $\epsilon > 0$, $R \subset \bigcup_{i=1}^{\infty} I_{(r_i - \frac{\epsilon}{2^i}, r_i + \frac{\epsilon}{2^i})}$ Then $\forall \epsilon, \epsilon > 0$:

$$\mu^*(R) \quad \leq \quad \sum_{i=1}^{\infty} |I_{(r_i - \frac{\epsilon}{2^i}, r_i + \frac{\epsilon}{2^i})}| \tag{8}$$

$$= \quad 2\epsilon \sum_{i=1}^{\infty} \frac{1}{2^i} \tag{9}$$

$$= \quad 2\epsilon \tag{10}$$

Consequently, $\mu^*(R) = 0$.

With this we can *define* the Lebesgue Integral of the function: $f(x) = \chi_R(x)$ over the interval $[0, 1]$ as $1 * \mu^*(R) = 1 * 0 = 0$.

In general, if a function can be written as a finite linear combination of characteristic functions: $h(x) = \sum_i^N c_i \chi_{E_i}(x)$, then we would *define* the Lebesgue integral for such functions as:

$$\int h(x)\, dx \quad = \quad \sum_{i=1}^{N} c_i \, \mu^*(E_i) \tag{11}$$

Going further, we would try to define the Lebesgue Integral of an "arbitrary" function, f, as the limit of integrals of approximating "simple" functions – ones that are a finite linear combinations of characteristic functions with the property that they converge pointwise to f.

---

[1] We denote the length of an interval, I, by $|I|$.

This is the initial plan. However, there is a problem with this approach. In the next section we provide an aside that describes a potential problem with the Lebesgue Outer Measure when used as a measure of "length".

4   Outer Measure Aside: Measuring Area in the Real World

An entrepreneur wanted to measure surfaces in the wild. He had heard that the ancient Greek mathematician Archimedes measured volumes of non-regular objects by placing them in a water basin. By observing the amount the water rises, he could measure the volume of water displaced and consequently, the volume of the object.

The entrepreneur wanted to do something similar to help measure the area of non-regular surfaces. He found that a certain oil when painted on a surface would accumulate at a fixed depth. By removing the oil (a proprietary process) and measuring the volume he could compute the area by dividing by this fixed depth. He called this method the "mu" method and used the symbology $\mu(C)$ to be his measure of the surface area of an object C.

One day a friend of his pulled out his comb from his pocket and asked him to find the surface area. The technique seemed to work fine, it registered that the comb was a little more that half the area as the comb's container.

A few days later the friend came by with a new comb. It was actually a comb set, consisting of two combs that fit together perfectly to form a rectangle. The teeth of the comb were much finer than the comb from a few days earlier. The retailer said is was much easier to store, you got a finer comb, and you get two combs for almost the price of one; and, better yet, they both fit in the same rectangular carrying case.

The entrepreneur applied his technique to the new combs. He found that, when fit together, they had the area of the rectangular carrying case. He determined the area of this rectangle to be A. He then tried each of the combs separately only to find that each of their areas was also A! It seems that due to the viscosity of the oil and the fine spacing between the teeth of the combs, the oil coating did not penetrate the fine crevasses of the comb, treating it as solid – consequently, the measure of the area was that of the (enclosing) rectangle.

The entrepreneur had to admit that his technique failed on the combs; there were clearly limitations to his technique. But how would he describe the limitation? Basically, the entrepreneur wanted to be able to advertise that his measurement method could be used for all "reasonable" objects. Exceptions would be objects that needed restoration that were broken into two pieces with extremely jagged edges.

What he wanted to say is that given an object, C could be μ measured *if* C was inside another object, O and the following was true:

$\mu(C) + \mu(C^c) = \mu(O)$

Here, C^c is the complementary set inside of O.

In some cases, it wasn't clear what the full object, O, consisted of. A better way to describe the limitation was to say that his measurement system was accurate when – for *any* given bounding frame X containing a set C – the measures of C and its complement in X added to the measure of X. That is,

$$\mu(X) = \mu(X \cap C) + \mu(X \cap C^c)$$

Unfortunately, this was not of much help – in other words, you know the measurements work on reconstruction only after you've done them. He stayed with his original caveat that the measurement works when some object has broken as long as the pieces are not "too" jagged.

5   Back to Outer Measure

It turns out that outer measure behaves much like the oil measurement system used in the last section. That is, there are sets, C, for which there is at least some, X ("bounding box"), so that $\mu^*(X) \neq \mu^*(X \cap C) + \mu^*(X \cap C^c)$

Sets like, C, are too jagged and its measure doesn't add properly.

We say that a set, C, is *measurable* if for all sets X the following is true:

$$\mu^*(X) \quad = \quad \mu^*(X \cap C) + \mu^*(X \cap C^c) \tag{12}$$

When a set is measurable we simply write $\mu(C)$ instead of $\mu^*(C)$.

What we're looking for is a requirement to be measurable – meaning outer measure works properly. By this we mean sets must be well behaved in the sense that the length of a set and the length of its "compliment" must add up to their union. When you think about this statement it seems clear that we very much need the "lengths" of sets and their compliments to add properly before we go any further with an integration theory – this is the least you expect/need from a new integration theory.

But why all the fuss about the intersection with "any set X". Why not just say that C and its compliment add up properly? The problem is that the compliment of C might have infinite length, because the ambient space containing C is infinite. Saying that C and its compliment add up correctly is the statement that the length of C and the length of its compliment is equal to the length of the ambient space: length(C) + ∞ = ∞ – a useless statement. This won't distinguish any sets as being special; that is, it won't help us filter out the badly behaving sets.

What we want to say, just like the combs from the previous section, is that the measure of C added with the measure of its "natural complement" – coming from a "natural" bounding box – gives back the measure of the bounding box. Without some restriction, the compliment could be an infinitely large set.

The problem now becomes: "What is the 'natural' bounding box for the set C"? Unfortunately, this is virtually impossible to say with any precision.

We can get around this problem by requiring that it be true for *all* bounding boxes. And now, we see what "X" accomplishes in the definition of measurability, (12). The "X" in this definition is a stand-in for *all* bounding boxes; but this includes, in particular, the/any bounding boxes that might be considered "natural". Consequently, (12) provides an unambiguous definition to filter out sets that behave badly, while cleanly avoiding the problem of the definition of a set's bounding box.

However, we've solved one problem, but now have another. Yes, we have a clean definition, but can one use this in practice to show that a set is measurable? The definition is clean, but the burden to show a set is measurable has very much increased!

We have now attempted to motivate the restriction of the length-of-sets measuring method called outer measure to so-call measurable sets via the filtering condition, (12). This filter is called "Carathéodory's Condition", and is introduced in most text books without much motivation. And yet it is the key idea that makes Lebesgue theory work.

So, given that we restrict ourselves to these "measurable" sets it turns out that the measure behaves in the way one might want. Namely.

$$\mu(A) = \sum_{i=1}^{N} \mu(A_i) \quad \text{when } A = \bigcup_{i=1}^{N} A_i \quad \text{with } A_i \cap A_j = \emptyset \quad \text{if } i \neq j$$

$$\mu(A) = \sum_{i=1}^{\infty} \mu(A_i) \quad \text{when } A = \bigcup_{i=1}^{\infty} A_i \quad \text{with } A_i \cap A_j = \emptyset \quad \text{if } i \neq j$$

That is, the measure of a set that is composed of a collection of disjoint sets should be the sum of the measures of the disjoint sets. We need this to hold for infinite sets as well because in any definition of integration we are going to need to take limits. And, as you might imagine, the solution to determining when one can apply the outer measure effectively boils down to the above two behaviors.

It turns out that if we restrict ourselves to using only measurable sets, the resulting sets have the property that they are an *algebra* with respect to unions, intersections and complements. In fact, the set of measurable sets is a σ-*algebra*: the measurable sets are closed under all finite or countably infinite unions, intersections and the complements of sets within this algebra.

Being an algebra, we can get a good handle on what sets are measurable if we can come up with a good *base* collection of sets that are *easily* shown to be measurable. From this, we *automatically* know that a set is measurable – no matter how complicated – if it is built out of these base sets through some (possibly infinite) combination of set operations. From this, we can effectively bootstrap the set of measurable sets. Once a complicated set is determined to be measurable, it is part of the algebra and can be used along with the base measurable sets – and anything else determined to be in the algebra – to construct other measurable sets. And we can continue this process ad infinitum.

Recapping, we have realized that our outer measure – a "natural" way of measuring the length of sets – is flawed. We found a way to restrict the usage to well behaved sets. *Unfortunately*, to make the restriction precise, we made the restriction verification very difficult to use. *Fortunately*, the resulting restricted sets are a σ-*algebra* and this helps save the day. We need only do the difficult work of verifying the restriction condition, (12), on simple "base" sets and then we can claim that sets formed by complicated set operations on the base sets are automatically measurable sets. If we choose the simple base sets to be the collection of open and closed intervals, we have a collection that is easy to show satisfy, (12), and yet provide a wealth of complex sets via set operations.

With a healthy collection of measurable sets we can define "measurable functions" as any function that is a simple function or is a pointwise limit of simple functions. The connection to measurable sets is that the simple functions – remember they are based on characteristic functions of sets – use only measurable sets in their construction.

Finally, the Lebesgue integral of a measurable function, f, is defined to be the supremum of the integrals of all simple functions that are less than or equal (pointwise) to f. Again, an unambiguous definition, but one

that is hard to compute. However, it turns out that – as one might hope – this supremum is the limit of the integrals of *any* sequence of functions that are monotonically increasing and converge point-wise to f. And again, the Lebesgue integral of any base component of a simple function – a characteristic function – is 1 * (outer measure (length) of the characteristic set).

Restricting ourselves to measurable sets does have a consequence: just like Riemann integration, we will not be able to integrate all functions, only measurable functions. But, such functions turn out to be a much larger and richer class of functions than the set of Riemann integrable functions.

And now with the proper machinery – knowing that if a sequence of measurable functions converges to another measurable function – we can work to find conditions to address the theorem described at the beginning of this paper; namely, can I interchange the sequence limit with the integral?

It turns out that there is a condition called "Lebesgue Dominated Convergence". This theorem requires only that the sequence of functions converge "almost everywhere" (all but a set of measure 0) to a function, f, and the sequence of functions as well as the function, f, are all bound above by an integrable function, g, having a finite integral. This is the kind of result we were hoping for and did see evidence for with the functions converging to our strange function which was 1 on rationals and 0 otherwise.

6   Conclusion

The path to creating a definition of a Lebesgue integral proceeds as mentioned in the previous sections – use approximating functions that are finite linear combinations of characteristic functions. Such combinations are called "simple" functions. They take the role that step functions played for Riemann integration.

One way to create approximating simple functions is by chopping up the range of the function into evenly spaced (if the range is finite) intervals and then looking at the domain values of f that map into this range interval. The set of values in the domain could, potentially, be a complicated set. But if they are all measurable, then one can create an approximating simple lower bound for the integral of f from simple functions based on these sets. This turns out to work provided – as you should imagine – that the "domain" sets are measurable. This leads to restricting the integration to functions, f, that are "measurable functions". It turns out that all the usual functions you encountered in first year calculus as well as any piecewise continuous functions are measurable functions. But the set of measurable functions includes functions that are not Riemann integrable – including the function we started out with, f, that was 1 on the rationals and 0 on the irrationals.

One can also show that the Lebesgue integral coincides with the Riemann integral for "nice" functions.

Finally, from this theory one does indeed get less stringent convergence criterion than the Riemann counterpart. One can have a criterion that avoids "uniform" convergence. In fact, essentially, you get a convergence theorem that requires only that a sequence of functions converge pointwise to a function[2] and that the sequence of functions are "dominated" – bounded above – by a Lebesgue integrable function.

It should be noted that there is an analogy between Riemann and Lebesgue integration and the Rational and the Real number systems. In the same way that the Real number system "completes" the Rational number

---

[2]In fact, one only requires that a sequence of functions converge point-wise to a function "almost everywhere" – meaning everywhere except on a set of outer measure 0.

system. The Lebesgue integration theory "completes" the Riemann theory. In fact, with the Lebesgue theory one can create function spaces with a metric that makes them "complete" in the same sense as Real numbers; namely, Cauchy sequences always converge in that space. The function space L_2 consists of all functions f (say over the interval $[0,1]$) that satisfy:

$$\int_0^1 f^2(x)\,dx < \infty \tag{13}$$

Here, the integral is the Lebesgue integral. This space also has an inner product defined, for two functions $f, g \in L_2$ by

$$\langle f, g \rangle \equiv \int_0^1 f(x)\,g(x)\,dx \tag{14}$$

And since it is a complete space, it is what is known as a *Hilbert Space*. A Hilbert space is a vector space with an inner product that is complete with respect to the metric induced from its inner product. That is, Cauchy sequences converge to an element in this space, just like they do for the finite dimensional Hilbert spaces: $R, R^2, R^3, \ldots$. And, much of the behavior of the function space, L_2 is replicated from the finite dimensional R^n spaces. This wouldn't be the case if we used the Riemann integral for the inner product/metric. It would be like working with the vector spaces $Q, Q^1, Q^2, \ldots$, where Q is the set of rational numbers. The vector properties would be there, but not the important analytic properties – and the notion of perpendicularity would be missing.

We should mention that, technically, the function space L_2 does not consist of functions, it consists of equivalence classes of functions having the property that they are identical except possibly on a set of measure 0.

If this seems strange to you, it shouldn't. The rational numbers are, strictly speaking, equivalence classes of integer pairs. For instance, the fraction that we write as '1/2' is really the equivalence class (set) of pairs of integers: $\{(1,2), (2,4), (3, 6), (4,8), \ldots\}$. To be equivalent, any two pairs, (a, b) and (c, d) must satisfy: $a * d = b * c$. Also, note that each equivalence class of integer pairs – representing a given rational number – consists of an *infinite* number of integer pairs.

In the same way, the Real number 1 is not the Rational number 1, it is actually an equivalence class; namely, the set of all Cauchy sequences that converge to the Rational number 1 – the simplest of which is the sequence: $1, 1, 1, 1 \ldots$.

10