Why Lebesgue Integration

R. Scott McIntire

Aug 16, 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. They 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. The applied people and other mathematicians want better convergence theorems.

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 “loose 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:

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

Finally, define, g_n by

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

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:

$$\begin{aligned}
\int_0^1 S_i(x) I_{E_{i,n}}(x)\, dx \leq \overbrace{1}^{\text{Max Height}} \quad * \quad \overbrace{2\frac{1}{n2^i}}^{\text{Width of } E_{i,n}} \quad &= \quad \frac{1}{n2^{i-1}} \\
\sum_{i=1}^{\infty} \int_0^1 S_i(x) I_{E_{i,n}}(x)\, dx \quad &\leq \quad \sum_{i=1}^{\infty} \frac{1}{n2^{i-1}} \\
&= \quad \frac{2}{n} \sum_{i=1}^{\infty} \frac{1}{2^i} \\
&= \quad \frac{2}{n}
\end{aligned}$$

What does this say about the integral of f?

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

$$\begin{aligned}
\int_0^1 f(x)\, dx \quad &= \quad \int_0^1 g_n(x)\, dx \\
&\leq \quad \text{(Darboux Upper Sum)}(g_n) \\
&\leq \quad \sum_{i=1}^{\infty} 1 * \text{length}(I_{E_{i,n}}(x)) \\
&\leq \quad \frac{2}{n}
\end{aligned}$$

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 (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.

Some might respond to the above argument saying: "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?

This 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 nice pieces and then showing that the upper and lower Darboux sums of a partition went to zero as the partition was 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* using a linear combination of characteristic functions:

$f(x) = \chi_R(x) + \chi_{IR}(x)$

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

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 intervals. 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. As we know how to measure the length of an interval, the "length" of an arbitrary set would be the infimum of the sum of the lengths over all such coverings. We call this "length" the *Lebesgue outer measure*. It is defined more precisely by:

$$\mu^*(S) = \inf_{\{\cup I_i\}_{i=1}^{\infty}} \sum_{i=1}^{\infty} \text{length}(I_i) \tag{7}$$

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 is that the "length" (outer-measure) of the set of rationals on the interval $[0,1]$ is 0.

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

Consequently, $\mu^*(R) = 0$. This approach will not work with the irrationals as they are not countable. In fact, any countably infinite cover of the irrationals with intervals must have a total length of at least 1. If not, there would be some open interval $[0,1]$, that is not covered. But any such open interval must contain an irrational; therefore all covers must have a total length of at least 1. Since the cover $\{(0,1), (-\epsilon, \epsilon), (1-\epsilon, 1+\epsilon)\}$ is a cover (for an infinite cover, let all other covers sets be set to $(0,1)$), with total length $1+2\epsilon$. Consequently, we see that $\mu^*(IR) = 1$.

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

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

$$\int_a^b h(x)\,dx = \sum_{i=1}^{N} c_i \mu^*(E_i) \tag{8}$$

Going further, we would try to define the Lebesgue Integral of an "arbitrary" function as the limit of integrals of approximating functions that are a finite linear combinations of characteristic functions.

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.

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 carrying case.

The entrepreneur applied his technique to the new combs. He found that, when fit together, they had the area of the resulting rectangle. 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 consisted of. A better way to describe the limitation was to say that his measurement system was accurate when *any* 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 they 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, 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:

$$\mu * (X) = \mu^*(X \cap C) + \mu^*(X \cap C^c) \tag{9}$$

What we're asking for is that any disjoint pieces constructed using C and its compliment must add up properly. 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 with a new integration theory.

But why all the fuss about the intersection with "any set X". Well we could remove X from this definition, but the definition would be meaningless if the ambient space we are dealing with is unbounded. For instance, when working with functions on R, the complement of a bounded set will have infinite length and then (9) would be the statement: $\infty =$ "Finite value" $+ \infty$. This is a useless statement. 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" gives back the measure of the bounding box – like the rectangular container of the combs. Without some restriction, the compliment is an infinite set. The problem then becomes: "What is the 'natural' bounding box for the set C"? We get around that problem by requiring that it be true for all bounding boxes. And then realizing that the "box" doesn't have to be bounded, we remove that restriction. This leaves a clean unambiguous definition for the measure of a set. But also makes it, potentially more difficult to use later in practice.

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

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

That is, the measure of a set that is composed of a bunch of other disjoint sets should be the sum of the measure of the disjoint sets.

And, as you might imagine, the solution to determining when one can apply the outer measure effectively boils done to the above two behaviors.

It turns out that if you restrict to measurable sets, then 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 these.

When a set, C, is known to be measurable we write $\mu(C)$ rather than $\mu^*(C)$.

6 Conclusion

The path to creating a definition of a Lebesgue integral proceeds as mentioned in a previous section – 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 the Riemann integral.

One way to find such an approximation 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 have less stringent convergence criterion than with Riemann integration that doesn't depend on "uniform" convergence. In fact, essentially, you get a convergence theorem that requires only that a sequence of functions converge point-wise to a function 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 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 consisting of all functions f (say over the interval $[0, 1]$) that satisfy:

$$\int_0^1 f^2(x)\, dx < \infty \tag{10}$$

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

$$\langle f, g\rangle \;\equiv\; \int_0^1 f(x)\, g(x)\, dx \tag{11}$$

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 have problems.

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: $a * d = b * c$.

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