A Brief Discussion of Lebesgue Measure Theory

Introduction

In class, we used the following function in our development of our definition of area under a curve:

\[ f(x) = \begin{cases} 
1 & \text{if } x \text{ is rational} \\
0 & \text{if } x \text{ is irrational} 
\end{cases} \]

In particular, if we compute \( \lim_{n \to \infty} R_n \) for this function over the interval \([0, 1]\), we get the wrong answer! Namely, we get 1, while the right answer is (I claimed) 0.

This was one of the points that led us to abandon \( \lim R_n \) as our definition for the area under a curve.

As a reminder, here is that computation again:

\[
\lim_{n \to \infty} R_n = \lim_{n \to \infty} \sum_{i=1}^{n} f(x_i) \Delta x \\
= \lim_{n \to \infty} \sum_{i=1}^{n} f\left(\frac{i}{n}\right) \left(\frac{1}{n}\right) \\
= \lim_{n \to \infty} \sum_{i=1}^{n} 1 \cdot \left(\frac{1}{n}\right) \\
= \lim_{n \to \infty} n \cdot 1 \cdot \left(\frac{1}{n}\right) \\
= \lim_{n \to \infty} 1 \\
= 1
\]

In the second line above, the function evaluates to 1 because it is being evaluated at \( \frac{i}{n} \), which is rational since \( i \) and \( n \) are both integers.

As I stated above, I claimed in class that this is actually the wrong answer to the question of the area under this function. The right answer can be computed with a Lebesgue integral, which I didn’t have time to discuss in class.

But in this particular instance, this Lebesgue integral actually can be computed without having to go to particularly gory depths.

The Lebesgue Integral

Roughly speaking, one of the main ideas behind the Lebesgue integral is to slice up the range instead of the domain. Strangely, this turns out to make a big difference.
Here are the specifics: Suppose we have a function $f$ defined on $[a, b]$. For the sake of simplicity, suppose we also have a bounded range given by the interval $[c, d]$.

Similarly to what we have done before, we will slice... but this time we slice up the interval $[c, d]$ into $n$ equal slices – not the interval $[a, b]$. We will also eventually consider the limit as $n$ approaches infinity. Also similarly to before, we define $y_i$ to be the upper endpoint of the $i$th such subinterval.

For a given $n$ and a given subinterval $[y_{i-1}, y_i]$, we consider the set of all points $x$ in the domain $[a, b]$ for which $f(x) \in [y_{i-1}, y_i]$. Let’s call this set $D_i$ (sometimes called the “pre-image” of the interval $[y_{i-1}, y_i]$).

Effectively, this subdivides the domain $[a, b]$ into $n$ subsets $D_1, \ldots, D_n$. Accordingly, this subdivides the area under the curve into $n$ regions as well. One such region is shown in the figure above.

The area of one of these regions is computed as the height $(y_i)$ times the width, which is the width of the set $D_i$. Note that in the case of the set $D_i$ in the figure above, the “width” of $D_i$ is the sum of the widths of the two intervals indicated.

At this point, we can see an advantage and a disadvantage of the Lebesgue approach. The clear advantage is that, because we have sliced horizontally in some sense, the height is easy to determine – unlike in the case of the Riemann integral, where we have serious problems in choosing the height. The disadvantage is that, unfortunately, we now have a nontrivial problem...
in determining the width – which of course was very easy to compute in the Riemann integral as just $\Delta x = (b - a)/n$.

Computing these widths is in general a hard problem, but can be dealt with effectively by advanced techniques I won’t describe here.

But, once that has been taken care of, then the Lebesgue integral can be written simply as

$$\lim_{n \to \infty} \sum_{i=1}^{n} y_i \cdot \text{width}(D_i)$$

which is just the sum of the products of heights times widths for each of the regions the area was divided into.

Now, for this particular function that we are interested in, there are only two values of $y$ for which there is any pre-image at all – namely, $y = 0$ (for which the pre-image is the set of irrationals between 0 and 1) and $y = 1$ (for which the pre-image is the set of rationals between 0 and 1). So, the Lebesgue integral just turns into

$$1 \cdot \text{width(rationals)} + 0 \cdot \text{width(irrationals)} = \text{width(rationals)}$$

So, the problem in this case reduces simply to finding the “width” of the set of rational numbers between 0 and 1.

This is a computation that we will be able to do, but first we have to establish a particular property of this set of rational numbers between 0 and 1.

**Countability**

A set $A$ of numbers is said to be “countable” if one can form a list

$$a_1, a_2, a_3, a_4, \ldots$$

such that every number in $A$ appears somewhere in that list.

For example, the set of positive integers is countable, as we see with the list

$$1, 2, 3, 4, \ldots$$

The set of all integers is also countable, as shown by the list

$$0, 1, -1, 2, -2, 3, -3, \ldots$$

The numbers in the list above are not in order, but that is okay – the point is that every number in the given set (in other words every integer) appears somewhere in the list.
It is tempting to think that all infinite sets are countable, but in fact this is not true. For example, the set of all real numbers is NOT countable. This can be proved without too much trouble, but I won’t include that proof here since it isn’t directly relevant.

So, since some infinite sets are countable, and some are not, we cannot simply assume that the set we are interested in (the set of rational numbers between 0 and 1, which we will refer to as $S$ throughout the remainder of this discussion) is countable. We need to verify this somehow.

The first step is to observe that we can first make an infinite array that includes all such rational numbers. This is shown below, where every numerator is the same as the column number, and every denominator is the same as the row number.

\[
\begin{array}{cccccc}
\frac{1}{1} & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdots \\
\frac{1}{2} & \frac{2}{2} & \frac{2}{3} & \frac{2}{4} & \cdots \\
\frac{1}{3} & \frac{2}{3} & \frac{3}{3} & \frac{3}{4} & \cdots \\
\frac{1}{4} & \frac{2}{4} & \frac{3}{4} & \frac{4}{4} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots \\
\end{array}
\]

Now, this includes all positive rational numbers, not just the ones in $S$ (the ones between 0 and 1); so, we scratch off the ones that are greater than one... we quickly notice that this is simply all of the entries above the diagonal.

We then notice that numbers appear multiple times in the array – for example, $1/1 = 2/2 = 3/3 = \ldots$, and $1/2 = 2/4 = 3/6 = \ldots$, and so on. Let’s scratch off the duplicates as well, leaving only the fractions that are in reduced form.

What we are left with at this point is just

\[
\begin{array}{cccccc}
\frac{1}{1} & \cdots \\
\frac{1}{2} & \cdots \\
\frac{1}{3} & \frac{2}{3} & \cdots \\
\frac{1}{4} & \frac{3}{4} & \cdots \\
\vdots & \vdots & \vdots & \ddots \\
\end{array}
\]

Since we have only removed duplicates and numbers that are not in $S$, we know that every rational number is on this array somewhere. And we furthermore notice that each row is now finite. So, we can make a list by just taking one row at a time. We get

\[
\begin{array}{cccccccc}
\frac{1}{1} & \frac{1}{2} & \frac{1}{3} & \frac{2}{3} & \frac{1}{4} & \frac{3}{4} & \frac{1}{5} & \frac{2}{5} & \frac{3}{5} & \cdots \\
\end{array}
\]
This list contains all of the elements of $S$, and so we can conclude that $S$ is countable.

The “Width” of the Rationals

Now that we have established that the set we are interested in, $S$, is countable, we can show that the “width” of $S$ is 0, with the following method.

First we will construct a new set, that we will denote by $B_k$. This set will be the union of intervals, each of which has a rational number as its center. We will start with 1 as a center, and then continue through the list that we formed in the last section.

The radii of these intervals will not be constant though... we will make them smaller as we proceed through the list. The first one will have radius $k$, the next one will have radius $k/2$, the next one $k/4$, ... etc., cutting the radius in half each time.

Putting all of this together, here is our new set $B_k$:

$$B_k = \left(\frac{1}{1} - k, \frac{1}{1} + k\right) \cup \left(\frac{1}{2} - k/2, \frac{1}{2} + k/2\right) \cup \left(\frac{1}{3} - k/4, \frac{1}{3} + k/4\right) \cup \cdots$$

We can make some observations about $B_k$. First of all, we note that $B_k$ contains $S$ as a subset, no matter what $k$ is – since in fact every point in $S$ is the center of one of the intervals that makes up $B_k$.

And since $B_k$ contains $S$ as a subset, it stands to reason that

$$\text{width}(S) \leq \text{width}(B_k)$$

Second, we can draw some conclusions about the width of $B_k$, because it is made up of a bunch of intervals, each of which has a known width. Of course some of these intervals overlap, and so we can’t conclude that the width of $B_k$ is equal to the sum of the widths of the intervals... but we can conclude that it will be less than or equal to that sum.

This gives us

$$\text{width}(B_k) \leq 2k + k + k/2 + k/4 + k/8 + \cdots$$

$$\leq 2k \left(1 + 1/2 + 1/4 + 1/8 + 1/16 + \cdots\right)$$

$$\leq 4k$$
Combining these two observations, we get

$$\text{width}(S) \leq \text{width}(B_k) \leq 4k$$

Since we know that this must be true for ALL positive values of $k$, no matter how small, the only possibility is that

$$\text{width}(S) = 0$$

Going back to the previous discussion of using the Lebesgue integral to compute area, we finally conclude that

$$\text{area under the curve} = \text{Lebesgue integral} = \text{width}(S) = 0$$