What's an Integral, Anyway?

The main goal of this class is to develop a more in-depth understanding of the integral. To support this understanding we will first develop a sophisticated approach to the idea of *measure*, which tells us how large a set is. (You can think of this as a useful generalization of "area" or "volume".)

You probably remember the Riemann integral, from either Analysis I or Calculus II.

Definition 0.1. Let $f : [a, b] \to \mathbb{R}$. We say f is *Riemann Integrable on* $[a, b]$ if there is a number $I \in \mathbb{R}$ so that, for any $\epsilon > 0$, there is a $\delta > 0$ such that, if S is a Riemann sum corresponding to a partition of width less than δ, then $|S - I| < \epsilon$. In this case we say that I is the *Riemann Integral of* f and write $I = \int_a^b f(x)\,dx$.

This definition is perfectly serviceable, but it has a few major issues. One is just that it's incredibly awkward to state, and difficult to use to prove things.

Second, there are a lot of functions that this definition doesn't quite apply to. You may remember so-called "improper" integrals from calculus II: these are integrals either over unbounded sets, like $\int_{-\infty}^{\infty} e^{-x^2}\,dx$, or integrals of unbounded functions like $\int_0^1 \frac{1}{\sqrt{x}}\,dx$. In either case the Riemann integral does not actually converge, and we need to use an awkward limiting process to even define, let alone compute, the integral.

Third, there are many sets we can't integrate over. A Riemann integral can integrate over sets like $[1, 3]$ but not over sets like the rational numbers or the Cantor set. As something of a corollary, we can consistently integrate continuous and nearly-continuous functions, but we can't integrate messy functions like

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

Fourth, and perhaps most importantly, the Riemann integral doesn't interact well with limits of sequences of functions. If f_n is a sequence of functions, we would like to prove a theorem like

$$\lim_{n \to infty} \int f_n(x)\,dx = \int \lim_{n \to \infty} f(x)\,dx.$$

However, this is unfortunately false. An easy example is to define

$$f_n(x) = \begin{cases} 4n^2 x & 0 \leq x < \frac{1}{2n} \\ 4n - 4n^2 x & \frac{1}{2n} x < \frac{1}{n} \\ 0 & \frac{1}{n} < x \leq 1 \end{cases}.$$

This looks complicated, but the graph is just an isosceles triangle with base $frac1n$ and height $2n$, and thus total area 1. So we know that for each n, $\int_0^1 f_n(x)\,dx = 1$.

However, for any fixed $x \in [0,1]$, it's easy to see that $\lim_{n\to\infty} f_n(x) = 0$. So if f is the pointwise limit of f_n, we have $f = 0$ and $\int_0^1 f(x)\,dx = 0$. Thus

$$\lim_{n\to\infty} \int_0^1 f_n(x)\,dx = 1 \neq 0 = \int_0^1 \lim_{n\to\infty} f(x)\,dx.$$

A new definition of the integral can't fix this example; the triangles under the f_n have area 1, and the pointwise limit is zero, and no amount of redefinition will fix that. But the Lebesgue integral we will define makes it easy to see exactly why this example breaks—and makes it easy to prove that in "most" cases, our desired theorem is actually true.

In the process of building our new and improved approach to the integral, we will develop ideas that help us understand probability better. Tu solve the problems with our integral we will find a way to define the measure or volume of a set. But if we have a collection of possible events, we can treat the probability of something hpapening as the measure of the set of events in which it happens.

If we are rolling a six-sided die, each side appears with probability or measure $\frac{1}{6}$, and we don't need any sophisticated tools to establish this. But if we are choosing a real number between zero and one, how do we describe the probability of getting a rational number? This will require a bit more work. And that work is where the content of this course starts.