THE THEORY OF INTEGRATION

Integration theory is motivated by an algebraic problem and by a geometric problem. You are already quite familiar with the geometric problem — that of finding the *area* under a curve:

The Area Problem: Given a continuous, nonnegative function $y = f(x)$, find the area of the region bounded by the x-axis, the curve $y = f(x)$, and the vertical lines $x = a$ and $x = b$.

The algebraic problem that motivates integration theory is:

The Antidifferentiation Problem: Given a continuous function $y = f(x)$, find some function $F(x)$ such that $F'(x) = f(x)$. As we have already see, this problem can be very difficult!

What is important to realize is that these two problems are really the same problem in disguise. Specifically, in attempting to solve the antidifferentiation problem geometrically, we are led to introduce the "area-so-far" function.

That is, let $y = f(x)$ be continuous and nonnegative on $[a, b]$, and for $x_0 \in [a, b]$ define $F(x_0)$ to be the area under the curve $y = f(x)$, above $y = 0$, to the right of $x = a$, and to the left of $x = x_0$. Then F can be pictured as the "area-so-far" function as x_0 varies between a and b. All you need to do now is draw a tiny rectangle with base $[x_0, x_0 + h]$ and height $f(x_0)$ to see that $F'(x_0)$ is approximately $f(x_0)$, with the approximation improving as $h \to 0$. This simple observation leads to (the first part of) the Fundamental Theorem of Calculus: namely, this "area-so-far" function $F(x)$ is differentiable and $F'(x) = f(x)$. Thus it would seem that we have solved the antidifferentiation problem.

In fact, we have simply replaced the antidifferentiation problem with a new one, namely the area problem — because it makes no sense to define $F(x_0)$ to be the "area" of a region unless we know what "area" *means*.

Our strategy for solving the area problem is based on two observations: (1) we know how to compute the area of a rectangle; (2) regions under continuous curves can be approximated to within an arbitrary degree of precision by rectangles. Our approximations using rectaingles will be called *Riemann sums*.

Let $y = f(x)$ be continuous and nonnegative on $[a, b]$, and let D be the region above the x-axis, under the curve $y = f(x)$, and between $x = a$ and $x = b$. Our goal is to find $A = Area(D)$. Our strategy will be to approximate D by finite collections of rectangles. Intuitively, these approximations should improve as we use more and more rectangles.

Each approximation will be determined by two pieces of information: (1) a *partition* of $[a, b]$, ie, a finite increasing sequence of points $a = x_0 < x_1 < \cdots < x_n = b$; and (2) a choice of *sample points* $x_k^* \in [x_{k-1}, x_k]$ in each interval for $1 \le k \le n$. The subintervals $[x_{k-1}, x_k]$ will form the bases of our approximating rectangles, and the y-values $f(x_k^*)$ will give the heights of these rectangles.

Each partition $\mathcal{P} = \{x_k\}_{0 \le k \le n}$ of $[a, b]$ and each choice of sample points $\{x_k^*\}_{1 \le k \le n}$ determines a *Riemann sum* for f on $[a, b]$. Specifically, the *Riemann sum for f on $[a, b]$ relative to the partition $\mathcal{P} = \{x_k\}_{0 \le k \le n}$ and choice of sample points $\{x_k^*\}_{1 \le k \le n}$* is defined to be the sum

$$\sum_{k=1}^{n} f(x_k^*)\Delta x_k,$$

where $\Delta x_k = x_k - x_{k-1}$ for each $1 \le k \le n$. In total, then, each Riemann sum is determined by four pieces of information:

(1) a closed interval $[a, b]$;
(2) a countinuous function f on $[a, b]$;
(3) a partition $\mathcal{P} = \{x_k\}_{0 \le k \le n}$ of $[a, b]$; and
(4) sample points $\{x_k^*\}_{1 \le k \le n}$ relative to the partition $\mathcal{P}$.

A partition $\mathcal{P} = \{x_k\}_{0 \le k \le n}$ is *regular* if $\Delta x_k = \Delta x_j$ for all $1 \le j, k \le n$, ie, if each subinterval $[x_{k-1}, x_k]$ has the same length. Notice that the Riemann sum for $y = f(x)$ on $[a, b]$ using a *regular* partition containing n subintervals and *left endpoints* as sample points takes on the relatively simple form

$$\sum_{k=0}^{n-1} f\left(a + k\frac{b-a}{n}\right)\left(\frac{b-a}{n}\right),$$

and the use of *right endpoints* as sample points shifts everything one subinterval to the right to yield the sum:

$$\sum_{k=1}^{n} f\left(a + k\frac{b-a}{n}\right)\left(\frac{b-a}{n}\right).$$

Intuitively, each of the above approximations should improve as n increases, so we might simply want to define the area A to be, say,

$$\lim_{n\to\infty} \sum_{k=0}^{n-1} f\left(a + k\frac{b-a}{n}\right)\left(\frac{b-a}{n}\right).$$

However, we should really only say that the area A is well-defined if *all* types of approximations tend to the same value, no mater what kind of partition or choice of sample points we use. So in order to allow for all possible choices of partitions and sample points, we make the following rather complicated definition. (Here the *norm* of a partition $\mathcal{P} = \{x_k\}_{0\le k\le n}$, written $\|\mathcal{P}\|$, is defined to be $\max\{\Delta x_k\}$, ie, the length of the largest subinterval appearing in $\mathcal{P}$).

Definition. Let $y = f(x)$ be continuous and nonnegative on the closed interval $[a, b]$. We say that the *area* under $y = f(x)$ and above $y = 0$ between $x = a$ and $x = b$ is A if:

> for every $\epsilon > 0$, there exists $\delta > 0$ such that for every partition $\mathcal{P} = \{x_k\}_{0\le k\le n}$ of $[a, b]$ such that $\|\mathcal{P}\| < \delta$, and for every choice of sample points x_k^* relative to $\mathcal{P}$, $|\sum_{k=1}^{n} f(x_k^*)\Delta x_k - A| < \epsilon$.

The above condition is usually abbreviated: "$A = \lim_{\|\mathcal{P}\|\to 0}\left(\sum_{k=1}^{n} f(x_k^*)\Delta x_k\right)$."

Furthermore, if this area A exists, we say that f is *integrable* on $[a, b]$, and we write

$$A = \int_a^b f(x)\,dx = \lim_{\|\mathcal{P}\|\to 0}\left(\sum_{k=1}^{n} f(x_k^*)\Delta x_k\right)$$

and call $\int_a^b f(x)\,dx$ the *definite integral* of f on $[a, b]$.

We can then prove the following important results concerning the definite integral of f on $[a, b]$:

First of all, as expected, the area under a *continuous* curve is perfectly well-defined:

Theorem: *If $y = f(x)$ is continuous on $[a, b]$, then f is integrable on $[a, b]$, ie,*

$$\int_a^b f(x)\,dx = \lim_{\|\mathcal{P}\|\to 0}\left(\sum_{k=1}^{n} f(x_k^*)\Delta x_k\right) \text{ exists.}$$

Secondly, since the whole point of introducing the "area-so-far" function was to find an antiderivative of $y = f(x)$, we should not be surprised that the following is true:

Theorem [FTOC(1)]: *Suppose $y = f(x)$ is continuous on $[a, b]$. Then the function*

$$F(x) = \int_a^x f(t)\,dt$$

is continuous on $[a, b]$, differentiable on (a, b), and for all $x \in (a, b)$, $F'(x) = f(x)$.

This result is often referred to as the first part of the Fundamental Theorem of Calculus. It says that *integration and differentiation are inverse operations.*

Finally, the second part of the Fundamental Theorem of Calculus provides a computational technique for evaluating definite integrals — and hence for computing the areas of regions under curves:

Theorem [FTOC(2)]: *If $y = f(x)$ is continuous on $[a, b]$ and if $F(x)$ is any antiderivative of f on $[a, b]$, then*

$$\int_a^b f(x)\,dx = F(b) - F(a).$$