WHAT IS AN INTEGRAL? .... BRAILEY SIMS

Department of Mathematics, University of New England.

Summary

An integral is presented as a real-valued function with domain a subset of the real-valued functions of a real variable, which is required to satisfy a number of axioms, capturing our intuitive expectations for 'areas under curves'.

The fundamental results of integration theory are shown to follow easily from these axioms and a brief discussion of the various ways by which integrals may be 'constructed' is included.

It is suggested that the axiomatic approach has the advantage of placing integration in a proper mathematical perspective, reinforcing students' earlier ideas of the function concept and allowing many associated topics, such as approximation (Simpson's Rule) to be approached more easily and naturally.

Question: *Find* $\int_1^{e^3} 1/x\,dx$.

Answer: $\int_1^{e^3} 1/x\,dx = \ldots = \ldots = 3$ SQUARE UNITS.

Recently I had the disturbing experience of seeing just how many of our H.S.C. students answered the above question in the way indicated (about 50% of those students who could answer the question at all). Of course the answer is numerically quite correct, but the inclusion of units indicates a basic misconception of the meaning of integration.

It is true that we often use the problem of finding the area *under* a curve to introduce students to the idea of an integral, and had the question been

Find the area of $\{(x, y) : 0 \leqslant y \leqslant 1/x$ *where* $1 \leqslant x \leqslant e^3\}$,

then the above answer would be right.

Integrals however arise naturally in many contexts and it is the context, not some intrinsic property of the integral, which determines what units, if any, should be attached to the answer. For example:

(i) a particle moving along a straight line, stationary at time $t = 1$ second, subject to an acceleration of t^{-1} cm/sec^2 would have velocity $\int_1^{e^3} 1/t\,dt = 3$ cm/second after e^3 seconds;

(ii) after e^3 seconds the temperature (heat accumulated) in a reservoir at 0°C after 1 second and into which heat flows at a rate varying inversely with time might be $\int_1^{e^3} t^{-1}\,dt = 3$°C.

To the mathematician however $\int_1^{e^3} 1/x\,dx$ is just a number, in this case the number 3.

So if 'areas under curves', velocities etc. are merely interpretations of integrals, we are led to ask, "What then is an integral?".

There are as many different answers to this question (all of them essentially equivalent of course) as there are approaches to the theory of integration. In the next few pages I will briefly outline what my answer might be, suppressing many of the less essential details.

If $\mathscr{F}$ denotes the set of real valued functions of a real variable, then essentially *an integral is a real-valued function whose domain is a subset of* $\mathscr{F}$; technically such a function is known as a *functional*. It is of course a very particular function satisfying a number of important axioms which will be listed below after introducing the necessary notation.

For $f,g \in \mathscr{F}$ and $\lambda \in \mathscr{R}$ (the real numbers), the sum $f+g$ and scalar multiple λf will denote those elements of $\mathscr{F}$ defined respectively by

$$\left.\begin{aligned}(f+g)(x) &= f(x)+g(x),\\ (\lambda f)(x) &= \lambda f(x)\end{aligned}\right\} \quad \text{for all } x \in \mathscr{R}.$$

With any $f \in \mathscr{F}$ and $a,b \in \mathscr{R}$ with $a \leqslant b$ associate ${}_af_b$, where

$${}_af_b(x) = \begin{cases} f(x) & \text{for } a < x < b \\ 0 & \text{for } x < a \text{ or } x > b \\ \text{anything} & \text{for } x = a \text{ or } x = b \end{cases}$$

Thus, for example, if I denotes the constant function assigning I to every $x \in \mathscr{R}$, then

$${}_0I_1(x) = 1 \text{ for } 0 < x < 1$$
$$= 0 \text{ for } x < 0 \text{ or } x > 1.$$

We now offer the following

DEFINITION: *$\mathscr{T} \subseteq \mathscr{F}$ is a set of integrable functions if*

(Ai) ${}_aI_b \in \mathscr{T}$ *for all* $a \leqslant b \in \mathscr{R}$;

(Aii) ${}_af_b \in \mathscr{T}$ *if* $f \in \mathscr{T}$;

(Aiii) $f+g \in \mathscr{T}$ *if* $f,g \in \mathscr{T}$;

(Aiv) $\lambda f \in \mathscr{T}$ *if* $f \in \mathscr{T}, \lambda \in \mathscr{R}$; } *i.e., $\mathscr{T}$ is a linear space;*

and if there exists a function $I : \mathscr{T} \to \mathscr{R}$ *satisfying*

(Av) *for* $f,g \in \mathscr{T}$ *and* $\lambda \in \mathscr{R}$, $I(f+\lambda g) = I(f) + \lambda I(g)$; *i.e., I is linear;*

(Avi) *if $f \in \mathscr{T}$ is such that $f(x) \geqslant 0$ for all $x \in \mathscr{R}$, then $I(f) \geqslant 0$; i.e., I is a positive mapping.*

(Avii) $I({}_aI_b) = b - a$ *for all* $a \leqslant b \in \mathscr{R}$.

Alternatively: If we require that $\mathscr{T}$ be closed under 'translation' and I translationally invariant (i.e., if $f \in I$ then $f_h \in \mathscr{T}$ and $I(f_h) = I(f)$ where $f_h(x) = f(x+h)$ all $x,h \in \mathscr{R}$), then (Ai) may be replaced by ${}_0I_1 \in \mathscr{T}$ and (Avii) by $I({}_0I_1) = 1$.

When first introducing integration to students one might avoid explicit mention of (Ai) to (Aiv), tacitly assuming the set of integrable functions being considered is closed under the operations needed to talk about integrals.

The function I is called an *integral* on $\mathscr{T}$ and we speak of its value at $f \in \mathscr{T}$, $I(f)$, as the integral of f. I have purposely chosen to write I instead of the more conventional $\int$ to help eliminate any preconceived ideas we might have about integrals.

At first sight these formal axioms may appear strange and difficult. However, if we interpret $I(f)$ as the *area* between the graph of f and the x-axis (an interpretation which we hope is sensible), then each axiom is merely the formalisation of an intuitively obvious result for such areas.

(P missing)

While (Av), with $\lambda = 1$, states that the area under a curve whose ordinate at x is $f(x) + g(x)$ is the sum of the areas under the curves with ordinates $f(x)$ and $g(x)$ at x. It is left to the reader to supply similar interpretations for the remaining axioms.

Granted that a set $\mathscr{T}$ of integrable functions can be found (more will be said on this later) many of the basic results of integration then follow easily from the axioms. We illustrate this with just two.

NOTATION: For $f \in \mathscr{T}$ and $a \leqslant b \in \mathscr{R}$ we write $I_a^b(f)$ for $I({}_af_b)$ and call this *the definite integral of f from a to b.*

INTEGRAL MEAN VALUE THEOREM: *If $f \in \mathscr{T}$ and $m,M \in \mathscr{R}$ are such that $m \leqslant f(x) \leqslant M$ for all x with $a \leqslant x \leqslant b$, then*

$$m(b-a) \leqslant I_a^b(f) \leqslant M(b-a).$$

(The more usual form follows from this and Bolzano's Theorem if f is assumed continuous.)

Proof. The function $g = {}_af_b - m_aI_b$ is such that $g(x) \geqslant 0$ for all $x \in \mathscr{R}$ and so by (Avi) $I(g) \geqslant 0$, but using (Av) and (Avii) $I(g) = I_a^b(f) - m(b-a)$, whence $m(b-a) \leqslant I_a^b(f)$.

The result that $I_a^b(f) \leqslant M(b-a)$ follows similarly and is left to the reader.

Now, since we can take ${}_af_c = {}_af_b + {}_bf_c$ where $a < b < c \in \mathscr{R}$, it follows from (Av) that $I_a^c(f) = I_a^b(f) + I_b^c(f)$ for all $f \in \mathscr{T}$.

It is consistent with this to define $I_b^a(f) = -I_a^b(f)$ when $a \leqslant b$, for then $I_a^a(f) = I_a^b(f) + I_b^a(f) = 0$ as expected.

DEFINITION: For any $f \in \mathscr{T}$ and $a \in \mathscr{R}$ we can define a new function F by $F(x) = I_a^x(f)$ for all $x \in \mathscr{R}$. Such a function will be called a *primitive* of f.

FUNDAMENTAL THEOREM OF CALCULUS: *If $f \in \mathscr{T}$ is continuous at $x_0 \in \mathscr{R}$ and F is a primitive of f, then F is differentiable at x_0 and* $F'(x_0) = f(x_0)$.

Proof. $[F(x_0+h)-F(x_0)]/h = [I_a^{x_0+h}(f) - I_a^{x_0}(f)]/h = [I_{x_0}^{x_0+h}(f)]/h$.

Thus, by the Integral Mean Value Theorem, $m \leqslant [F(x_0+h)-F(x_0)]/h \leqslant M$ where m is the minimum of f and M is the maximum of f for x between x_0 and x_0+h, which exist, for sufficiently small h, by the continuity of f at x_0 and further as $h \to 0$, $m, M \to f(x_0)$.

So $F'(x_0) = \lim_{h\to 0} [f(x_0+h)-F(x_0)]/h$ exists and equals $f(x_0)$.

One could continue in this vein developing other standard results: The Substitution Theorem, Integration by Parts Formula, and the like. However further developments follow from the theorems already established in quite orthodox ways (see any introductory calculus text book) and would serve no purpose in this note.

We have seen that given a set of integrable functions, the standard theorems of integral calculus are readily shown to hold for it. The fundamental problem of integration is then to construct (or establish the existence of, and then characterise) suitably 'large' sets of integrable functions.

Considering the large number of axioms such a set must satisfy, it is perhaps surprising that any such sets exist. However they do, and can be arrived at in a great variety of ways, which it would take us far beyond the scope of this note to do more than mention a few.

A 'small' but important set of integrable functions can be arrived at directly from the axioms. This is the set of *step functions* S, where $f \in S$ if there exists a finite set of points $x_0 < x_1 < \dots < x_n$ and numbers $f_1, f_2, \dots, f_n$ such that

$$f(x) = \begin{cases} 0 \text{ for } x < x_0 \text{ or } x > x_n \\ f_i \text{ for } x_{i-1} < x < x_i, \quad (i = 1, 2, \dots, n) \end{cases}$$

Since such as f can be written as $\sum_{i=1}^{n} {}_{x_{i-1}}f_{x_i}$ it follows from (Ai), (Aiii) and (Aiv) that f is integrable while (Av) and (Avii) necessitate that $I(f) = \sum_{i=1}^{n} f_i \cdot (x_i - x_{i-1})$, and so the integral is uniquely determined on S.

One way of proceeding might be to try and 'extend' S to a larger set of functions. Thus the set of $f \in \mathcal{F}$ for which there exists a sequence $f_1, f_2, \dots, f_n, \dots$ of step functions converging "uniformly" to f and for which $\mu = \lim_{n\to\infty} I(f_n)$ is finite, can, after the appropriate checking, be shown to form a set of integrable functions, for which the integral is given by $I(f) = \mu$. In this way we arrive at the integrability of the set of *regulated functions* considered by J. Dieudonné (*Foundations of Modern Analysis*, Academic, N.Y., 1960); a set, containing all the integrable continuous functions, which is sufficient to adequately serve most needs of the applied mathematician.

More ambitiously, one might show that if $f_1, f_2, \dots, f_n, \dots$ is an 'increasing' sequence of step functions for which the sequence $I(f_1), I(f_2), \dots, I(f_n), \dots$ converges, then except for points x lying in a suitably "small" set (a set of *measure zero* to be precise) $f_n(x) \to f(x)$ as $n \to \infty$, for some $f \in \mathcal{F}$. The set of all functions f arising in this way can then be shown to satisfy (Ai) to (Avii) with $I(f) = \lim_{n\to\infty} I(f_n)$ and one arrives at the set of *Lebesgue integrable functions* (see Weir, *Lebesgue Integration & Measure*, Cambridge 1973).

Alternatively we might start with a given set of functions in mind and after ensuring that it satisfied (Ai) to (Aiv) attempt to construct an integral, I, on it which satisfies the remaining axioms. This is essentially the approach taken in Riemann integration. One starts with an adequate subset of functions of 'bounded variation' and using the machinery of upper and lower sums (essentially step function devices) arrives at a suitably defined integral for it.

Lastly, as is often done in school calculus courses we may construct a limited, though valuable, set of integrable functions by taking inspiration from the fundamental theorem of the calculus established earlier.

Thus, let $f \in \mathcal{F}$ be continuous except at a finite number of points, and assume we can find a continuous function F which is an antiderivative of f except possibly at the points of discontinuity of f. That is: $F'(x) = f(x)$ for all points x at which f is continuous.

Then after some, though not very difficult, checking we can show that f is integrable with $I(f) = \lim_{z\to\infty} F(z) - F(-z)$ provided this limit exists. The important point to note is that for $f \in \mathcal{F}$ with antiderivative F we have that

$$G(x) = \begin{cases} F(a) \text{ for } x < a \\ F(x) \text{ for } a \leqslant x \leqslant b \\ F(b) \text{ for } x > b \end{cases}$$

is a suitable antiderivative of ${}_af_b$ whence ${}_af_b \in \mathcal{F}$ and from which it follows that $I_a^b(f) = F(b) - F(a)$.

In conclusion then, we see that an integral is a very distinguished function, the existence of which and in particular the connection between it and the operation of differentiation (Fundamental Theorem of Calculus) allows a great variety of problems to be solved. For example, the physical problems presented at the start of the article. Because of this importance many theories have been developed to establish the existence of "integrals" on sets other than subsets of $\mathcal{F}$. In most of these cases the starting point is an appropriately modified version of the axioms (Ai) to (Avii). For instance the Haar integral for real-valued functions from a compact topological group.

Once we see integrals in this light it becomes as silly to write $\int_1^{e^3} 1/x \, dx = 3$ SQUARE UNITS as it would be to write of the function $f \in \mathcal{F}$, defined by $f(x) = x^3$, $f(2) = 8$ CUBIC UNITS.

Apart from the likelihood of a student, who does not appreciate this, committing an error, his ability to see the central and important role played by integration theory will necessarily be severely handicapped. It is therefore essential that students studying integration are made aware of the distinction between an integral and the answer it may produce to any given problem.