Introduction  The basic concepts of calculus are real numbers, functions from reals to reals, continuity of functions, and the derivatives and integrals of functions. Expressing these things in a functional programming language like OCAML lets us compute with these concepts and gives a concrete meaning for mathematical concepts that might seem very abstract. Thinking about the OCAML types for these things can give you a clearer understanding of them.

Integers  The integers are represented in OCAML by the type `big_int`. For brevity, I'll write `Int` for the type of integers.

Real numbers  We have seen that a real number \( x \) is a function from positive integers to rationals such that for all \( n \) and \( m \), \( |x(n) - x(m)| \leq \left( \frac{1}{n} + \frac{1}{m} \right) \). We can think of this as saying that the \( n \)th approximation \( x(n) \) is a rational number that is within distance \( \frac{1}{n} \) of the real number \( x \).

We represent a rational number using pair of integers, so it can have OCAML type `Int * Int`. So a real number can have OCAML type `Int -> Int * Int`. But we could code a pair of integers into a single integer, or, even better, we can always normalize the \( n \)th approximation \( x(n) = \frac{a_n}{b_n} \) to \( \frac{c_n}{2n} \) where \( c_n = 2n * a_n \div b_n \). Then we can just represent the real number \( x \) by the function \( \lambda n. c_n \) because the denominator of the \( n \)th approximation will always be \( 2n \).

So, a real number can have the OCAML type `Int -> Int`. Lets write `real` for whichever type we have chosen for the real numbers.

Functions of reals  What is the type of a function like \( \cosine(x) \)? Is its OCAML type `real -> real`? The answer is “yes and no”. Yes, because that is the best we can do in the OCAML type system. No, because to be a function from reals to reals it has to satisfy an additional property.
Two real numbers $x$ and $y$ are equal if their $n^{th}$ approximations are always within $\frac{2}{n}$ of each other:

$$x = y \iff |x(n) - y(n)| \leq \frac{2}{n}, \text{ for all } n$$

We call an OCAML function $f: \text{real} \to \text{real}$ an operation on real numbers. But to be a (mathematical) function on real numbers it must satisfy the property that if $x = y$ then $f(x) = f(y)$.

If that holds for all reals $x$ and $y$ in some interval $[a, b]$, then we say that the operation $f$ is a function defined on interval $[a, b]$.

$$\text{FUN}(f, a, b) \iff (x = y \Rightarrow f(x) = f(y), \text{ for all } x, y \text{ in } [a, b])$$

**Continuous functions** A function is continuous if its graph does not have any gaps or jumps. So, if $x_1$ and $x_2$ are very close together then $f(x_1)$ and $f(x_2)$ must also be close together (otherwise there would be a jump in the graph of $f$ between $x_1$ and $x_2$). To say this precisely we need an “epsilon-delta” definition, but that is not complicated. We can represent an arbitrarily small $\epsilon$ or $\delta$ as $\frac{1}{n}$ where $n$ is a positive integer. The operation $f: \text{real} \to \text{real}$ is (uniformly) continuous on the interval $[a, b]$ if for any $\epsilon$ there is a $\delta$ such that for all reals $x_1$ and $x_2$ in the interval $[a, b]$, if $x_1$ and $x_2$ are within $\delta$ of each other then $f(x_1)$ and $f(x_2)$ are within $\epsilon$ of each other.

$$\text{CONT}(f, a, b) \iff \forall n. \exists m. |x - y| \leq \frac{1}{m} \Rightarrow |f(x) - f(y)| \leq \frac{1}{n}, \text{ for all } x, y \text{ in } [a, b]$$

We can represent the fact that operation $f$ is continuous on $[a, b]$ by giving another function $\text{mc}: \text{Int} \to \text{Int}$, called the modulus of continuity of $f$, that for each $n$ gives the needed $m$. So an operation $f$ is continuous on $[a, b]$ if there is a modulus of continuity $\text{mc}$ such that, for any positive integer $n$, if $x_1$ and $x_2$ are within $\frac{1}{n\text{mc}(n)}$ of each other then $f(x_1)$ and $f(x_2)$ will be within $\frac{1}{n}$ of each other.

**Integral of a function** If $f$ is a continuous function on $[a, b]$ then $\int_a^b f(x)dx$ is the (signed) area under the graph of $f$ between $a$ and $b$. To get the $n^{th}$ approximation of this real number, we partition the interval $[a, b]$ into parts of length $s = \frac{b-a}{k}$, by letting $p_0 \leq p_1 \leq p_2 \cdots \leq p_k$ be $p_i = a + (i * s)$ so $p_0 = a$ and $p_k = b$. Then we add up the areas of the rectangles $f(p_i) * s$ for $i = 0, 1, \ldots k - 1$. We need to choose $k$ big enough and approximate the $f(p_i)$ close enough so that what we get is withing $\frac{1}{n}$ of the true area.
If $mc$ is a modulus of continuity for $f$ and $c$ is an integer $c$ such that $2n \cdot (b - a) \leq c$, then we let $m = mc(c)$ and choose $k$ so that $s = \frac{b - a}{k} \leq \frac{1}{m}$. Then for any $x$ in the small interval $[p_i, p_{i+1}]$ we will have $|f(x) - f(p_i)| \leq \frac{1}{c}$. So the difference between the area of the rectangle $f(p_i) \cdot s$ and the true area under the curve between $p_i$ and $p_{i+1}$ will be at most $\frac{s}{c}$. If we add up all of these errors we get at most $k \cdot \frac{s}{c}$ which equals $\frac{b - a}{m}$ which is $\leq \frac{1}{2n}$. The sum of the areas of the rectangles is the real number $\text{riemann\_sum} f a b k = s \cdot (f(a) + f(a + s) + \ldots f(a + (k - 1) \cdot s))$. The $(2n)^{th}$ approximation of that real number is within $\frac{1}{2n}$ of the area of the rectangles which is within $\frac{1}{2n}$ of the true area. So it is within $\frac{1}{n}$ of the true area.

**OCAML code for integral**

```ocaml
let riemann_sum f a b k =
  let x = rdiv_int (rsubtract b a) k in
  let g i =
    let aa = rmul (bigint2real (sub_big_int k i)) a in
    let bb = rmul (bigint2real i) b in
    rdiv_int (radd aa bb) k in
  let s = rsum (fun i -> f (p i)) zero_big_int (pred_big_int k) in
  rmul s x

let integral mc f a b:real =
  fun n ->
    let nn = mult_int_big_int 2 n in
    let len = canonical_bound (rsubtract b a) in
    let c = mult_big_int nn len in
    let m = mc c in
    let k = mult_big_int m len in
    riemann_sum f a b k nn:q
```

When we load this code into OCAML we get this:

```ocaml
val integral: (Big_int.big_int -> Big_int.big_int) ->
  (real -> real) -> real -> real -> real = <fun>
```

The inputs to the integral are the modulus of continuity, the operation, and the endpoints. The output is a real.

**How to get a modulus of continuity** If a function $f$ has a derivative $f'$ then the *Mean Value Theorem* says that $\frac{|f(x) - f(y)|}{|x - y|} = f'(c)$ for some $c$.
between $x$ and $y$. So, if the maximum of the absolute value of $f'$ on the interval $[a, b]$ is bounded by an integer $k$, we can use \texttt{fun n -> k*n} for a modulus of continuity for $f$ on $[a, b]$.

**Running the integral code**  We can try this out. To calculate the area under the curve $y = x^2$ between 1 and 2 we use

\[
\text{integral (fun n -> 4*n) (fun x -> rmul x x) (int2real 1) (int2real 2)}
\]

Since we know that the derivative of $x^2$ is $2x$, the maximum of the derivative on the interval $[1,2]$ is 4. That is why we can use \texttt{fun n -> 4*n} for the modulus of continuity.

To get two digits of accuracy, we need to approximate the integral within $\frac{1}{100}$ so we apply it to 100. We get 2.33, reasonably fast. But to get three digits accuracy we apply to 1000, and it takes more than a minute to get 2.333.

Let’s compute the area under the curve $y = \sin(x)$ between 0 and 3. Since the derivative of $\sin(x)$ is $\cos(x)$ and that is bounded by 1, we can use the modulus of continuity \texttt{fun n -> n}. So we use

\[
\text{integral (fun n -> n) (fun x -> sine x) (int2real 0) (int2real 3)}
\]

We get two digits of accuracy by applying it to 100 and get 1.98 but it takes about a minute.

**Fundamental theorem of Calculus** Function $F$ is an anti-derivative of function of $f$ if $F'(x) = f(x)$.

The fundamental theorem of calculus says that in that case, $\int_a^b f(x)dx = F(b) - F(a)$.

We can use this to compute $\int_1^2 x^2dx$ quickly because an anti-derivative is $\frac{x^3}{3}$ so the answer is $\frac{2^3}{3} - \frac{1^3}{3} = \frac{7}{3} = 2.3333333333333333\ldots$

For $\int_0^3 \sin(x)dx$, the anti-derivative is $-\cos(x)$, so the answer is $-\cos(3) + \cos(0)$ which is $1 - \cos(3)$. We can compute 20 digits of this very quickly and get 1.9899249660044545727. This agrees with the mealy two digits of accuracy we got after a minute using the integral code.

Moral: *The fundamental theorem of calculus is an efficiency result.* It says that a labor-intensive summation can be computed by evaluating an anti-derivative at two points.