1 What is integration

Let $S$ be a family of subsets of $\mathbb{R}$ (to be determined) and let $\mathcal{F}$ be a family of real-valued functions defined on $\bigcup S$ (to be determined), with $\mathcal{F}$ closed under constant scaling and addition. We are looking for a function

$$\int : \mathcal{F} \times S \to \mathbb{R} : (f, X) \mapsto \int_X f$$

with the following nice properties:

- (Linearity) For all $X \in S$, all $f, g \in \mathcal{F}$, and all $\lambda \in \mathbb{R}$, we should have $\int_X \lambda f = \lambda \int_X f$ and $\int_X (f + g) = \int_X f + \int_X g$.

- (Monotonicity) For all $X \in S$ and all $f, g \in \mathcal{F}$, if $f \leq g$ on $X$, then $\int_X f \leq \int_X g$.

- (Additivity over almost-disjoint domains): For all $X_1, X_2 \in F$, and all $f \in F$, if $X_1 \cap X_2$ is empty or contains only a single point, then $\int_{X_1 \cup X_2} f = \int_{X_1} f + \int_{X_2} f$.

These are the main properties we want from an integral (on $\mathbb{R}$). Another important property that will be true of our integrals is that for $f = c$ (a constant) and $X = [a, b]$ (an interval with $a \leq b$) we should have

$$\int_{[a, b]} c = c(b - a)$$

We will present two major theories of integration, the Riemann–Darboux integral in a slightly more modern/adaptable treatment, and the more advanced and general Lebesgue integral. The choices of $S$ and $\mathcal{F}$ in these theories differ considerably. That being said, both theories are developed along a similar blueprint: We begin by defining the integral for a certain family of easily-integrated functions, step functions in the RD theory and simple functions in the L theory. The integral for more general classes of functions is then based on these easier integrals and the desirable properties of the integral are then proved.

2 The Riemann–Darboux integral

2.1 Step functions

If $S \subseteq \mathbb{R}$, the indicator function of $S$ is the function $1_S : \mathbb{R} \to \{0, 1\}$ given by

$$1_S(x) = \begin{cases} 1 & \text{if } x \in S, \\ 0 & \text{if } x \notin S. \end{cases}$$

A step function is a finite linear combination of indicator functions of bounded intervals. That is, a step function is a function of the form

$$\phi(x) = \sum_{k=1}^n c_k \cdot 1_{I_k}(x)$$
for some \(c_1, \ldots, c_n \in \mathbb{R}\) and some bounded intervals \(I_1, \ldots, I_n \subseteq \mathbb{R}\). These intervals may be open, closed, half-open/closed, degenerate (consisting of a single point like \(\{a\} = [a, a]\)), even empty, and they may overlap. Note that a step function may have many representations like the above—there is absolutely no guarantee (or need) of uniqueness.

For most of this section, we will be working with functions on a fixed closed interval \([a, b]\) \((a < b)\). Thus, a \textit{step function} on \([a, b]\) is a function \(\phi : [a, b] \rightarrow \mathbb{R}\) that arises from restricting a step function as above to the domain \([a, b]\). Note that every such step function can be written as \(\sum c_k \cdot 1_{I_k}(x)\) for subintervals \(I_1, \ldots, I_k \subseteq [a, b]\). We will denote the set of step functions on \([a, b]\) by \(\text{Step}([a, b])\). Note that this set is closed under scaling and addition (algebraically speaking, it is a vector space over \(\mathbb{R}\)).

2.2 Integrals of step functions

It is an easy consequence of the definition of step functions that a step function can only take on finitely many values. Furthermore, for each \(y \in \mathbb{R}\), the preimage \(\phi^{-1}(y)\) is a union of finitely many disjoint subintervals of \([a, b]\). In particular, the sets \(\phi^{-1}\{y\}\) all have well-defined length (in the “naïve” sense where the length of an interval \(I\) is \(|I| = \sup(I) - \inf(I)\), and if \(I_1\) and \(I_2\) are disjoint intervals, then \(|I_1 \cup I_2| = |I_1| + |I_2|\)).

We define the integral of a step function \(\phi : [a, b] \rightarrow \mathbb{R}\) by

\[
\int_{a}^{b} \phi = \sum_{y \in \mathbb{R}} y \cdot |\phi^{-1}(y)|
\]

Note that the sum is finite, since \(\phi^{-1}(y)\) is empty except for the finitely many \(y \in \phi([a, b])\).

Here’s how to interpret the above graphically: The region measured by \(\int_{a}^{b} \phi\) is a bar graph. For each \(y\), the summand \(y \cdot |\phi^{-1}(y)|\) is the combined area of the bars that have height \(y\) (with area below the \(x\)-axis counted negatively).

This definition is somewhat cumbersome in practice. Instead, we will show that the integral of the step function \(\sum k c_k \cdot 1_{I_k}\) is simply \(\sum c_k \cdot |I_k|\). The advantage to defining the integral as above is that it does not depend on the choice of representation (and the same step function has many representations).

**Theorem 1.** Let \(\phi : [a, b] \rightarrow \mathbb{R}\) be a step function and suppose that

\[
\phi = \sum_{k=1}^{n} c_k \cdot 1_{I_k}
\]

for \(c_1, \ldots, c_n \in \mathbb{R}\) and intervals \(I_1, \ldots, I_n \subseteq [a, b]\). Then

\[
\int_{a}^{b} \phi = \sum_{k=1}^{n} c_k \cdot |I_k|.
\]

**Proof.** We will induct on \(n\). If \(n = 0\), then \(\phi = 0\) and \(\sum c_k |I_k| = 0\) (since it is the empty sum). On the other hand \(\phi^{-1}(y)\) is empty except for at \(y = 0\), where \(\phi^{-1}(y) = [a, b]\); thus, \(\int_{a}^{b} \phi = 0 \cdot (b - a) = 0\).
Suppose then that \( \phi = \sum_{k=1}^{n+1} c_k \cdot 1_{I_k} \) and let \( \phi_0 = \phi - c_{n+1} \cdot 1_{I_{n+1}} = \sum_{k=1}^{n} c_k \cdot 1_{I_k} \). Then \( \phi_0 \) is a step function and our inductive hypothesis is that
\[
\sum_{k=1}^{n} c_k \cdot |I_k| = \int_a^b \phi_0 = \sum_{y \in \mathbb{R}} y \cdot |\phi_0^{-1}(y)|
\]
Now, let \( x \in [a, b] \) and let \( y = \phi^*(x) \). If \( x \notin I_{n+1} \), then \( \phi(x) = y \). If \( x \in I_{n+1} \), then \( \phi(x) = y + c_{n+1} \). We therefore have
\[
\int_a^b \phi = \sum_{y \in \mathbb{R}} y \cdot |\phi_0^{-1}(y) \cap I_{n+1}| + \sum_{y \in \mathbb{R}} (y + c_{n+1}) \cdot |\phi_0^{-1}(y) \cap I_{n+1}|
\]
Distributing the \( (y + c_{n+1}) \) term yields
\[
\int_a^b \phi = \sum_{y \in \mathbb{R}} y \cdot |\phi_0^{-1}(y) \cap I_{n+1}| + \sum_{y \in \mathbb{R}} y \cdot |\phi_0^{-1}(y) \cap I_{n+1}| + c_{n+1} \sum_{y \in \mathbb{R}} |\phi_0^{-1}(y) \cap I_{n+1}|
\]
\[
= \sum_{y \in \mathbb{R}} y \cdot (|\phi_0^{-1}(y) \cap I_{n+1}| + |\phi_0^{-1}(y) \cap I_{n+1}|) + c_{n+1} \sum_{y \in \mathbb{R}} |\phi_0^{-1}(y) \cap I_{n+1}|
\]
\[
= \sum_{y \in \mathbb{R}} y \cdot |\phi_0^{-1}(y)| + c_{n+1} \sum_{y \in \mathbb{R}} |\phi_0^{-1}(y) \cap I_{n+1}|
\]
The first term is \( \int_a^b \phi_0 \) (by definition). The second term is \( c_{n+1} \cdot |I_{n+1}| \) since \( I_{n+1} = \bigcup_{y \in \mathbb{R}} \phi_0^{-1}(y) \cap I_{n+1} \) is a disjoint union of sets with well-defined length (almost all of whose terms are empty). Finally, we conclude that
\[
\int_a^b \phi = \int_a^b \phi_0 + c_{n+1} \cdot |I_{n+1}| = \sum_{k=1}^{n} c_k \cdot |I_k| + c_{n+1} \cdot |I_{n+1}| = \sum_{k=1}^{n+1} c_k \cdot |I_k|
\]
as desired. \( \square \)

Another way to interpret the theorem above is as
\[
\int_a^b \sum_{k=1}^{n} c_k \cdot 1_{I_k} = \sum_{k=1}^{n} c_k \cdot \int_a^b 1_{I_k}
\]
because \( \int_a^b 1_{I_k} = |I_k| \) for any subinterval \( I_k \subseteq [a, b] \). The theorem above amounts to linearity of the integral on step functions.

**Corollary 2.** The integral \( \int : \text{Step}([a, b]) \to \mathbb{R} \) is linear.

An interval partition of \([a, b]\) is a family \( \mathcal{P} = \{I_k\}_{k=1}^{n} \) of subintervals of \([a, b]\) that are pairwise disjoint (\( I_i \cap I_j = \emptyset \) for \( i \neq j \)) and whose union is \([a, b]\). We say that \( \phi \) is compatible with \( \mathcal{P} \) if it can be written as a linear combination over \( \{1_J\}_{J \in \mathcal{P}} \). Every step function of \([a, b]\) is compatible with some interval partition of \([a, b]\).
Corollary 3. The integral $\int : \text{Step}([a,b]) \to \mathbb{R}$ is monotonic.

Proof. Let $\phi, \psi \in \text{Step}([a,b])$ satisfy $\phi \leq \psi$ on $[a,b]$ and let $\{I_i\}_{i=1}^n$ and $\{J_j\}_{j=1}^n$ be compatible interval partitions of $[a,b]$ for $\phi$ and $\psi$, respectively. Then $\{I_i \cap J_j\}_{ij}$ is an interval partition of $[a,b]$ compatible for $\psi - \phi$. We have

$$
\psi - \phi = \sum_{ij} (d_j - c_i) \cdot 1_{I_i \cap J_j}
$$

where $c_i$ is the value of $\phi$ on $I_i$ and $d_j$ is the value of $\psi$ on $J_j$. If $I_i \cap J_j$ is nonempty, then since $\phi \leq \psi$ we must have $c_i \leq d_j$. (If $I_i \cap J_j$ is empty, then it has zero length.) Thus,

$$
\int_a^b (\psi - \phi) = \sum_{ij} (d_j - c_i) \cdot |I_i \cap J_j|
$$

has only nonnegative terms. By linearity, it follows that $\int_a^b (\psi - \phi) = \int_a^b \psi - \int_a^b \phi \geq 0$, so $\int_a^b \phi \leq \int_a^b \psi$, as desired. \qed

Combining these corollaries plus an easy, unproved third result:

**Theorem 4.** The integral $\int : \text{Step}([a,b]) \to \mathbb{R}$ is linear, monotonic, and additive on almost-disjoint intervals. This last property means

$$
\int_a^b \phi = \int_a^c \phi + \int_c^b \phi
$$

for any step function $\phi : [a,b] \to \mathbb{R}$ and any $c \in [a,b]$. That is, integrals of step functions on intervals satisfy our basic requirements for integrals!

### 2.3 Lower and upper integrals of bounded functions

Let $f : [a,b] \to \mathbb{R}$ be a bounded function. Consider the set

$$
\mathcal{L}(f, [a,b]) = \left\{ \int_a^b \phi : \phi \text{ is a step function with } \phi \leq f \text{ on } [a,b] \right\}
$$

Intuitively, this is the set of all lower estimates for the integral of $f$.

The set $\mathcal{L}(f, [a,b])$ is nonempty and bounded above by $M(b-a)$ where $f \leq M$ on $[a,b]$, so it has a supremum (the "best" lower estimate is the greatest one!). We define

$$
\int_a^b f = \sup \mathcal{L}(f, [a,b])
$$

This is the lower (Riemann–Darboux) integral of $f$ on $[a,b]$. Similarly, we define $\mathcal{U}(f, [a,b])$ to be the integrals over all $\psi \geq f$ on $[a,b]$, and the upper integral by

$$
\int_a^b f = \inf \mathcal{U}(f, [a,b])
$$
Note that the upper and lower integrals are always defined. Here’s why: If $f$ is bounded, then $-M \leq f \leq M$ for some constant $M \geq 0$. $\phi = -M$ and $\psi = M$ are step functions with integrals equal to $-M(b-a)$ and $M(b-a)$, respectively. This proves that $\mathcal{L}(f, [a, b])$ is nonempty (it contains $-M(b-a)$) and bounded above (by $M(b-a)$) so it has a supremum: the lower integral of $f$. The upper integral exists by a symmetric argument.

**Proposition 5.** If $f : [a, b] \to \mathbb{R}$ is bounded, then $\int_{a}^{b} f \leq \int_{a}^{b} g$.

**Proof.** This is simply because if $\phi \leq f \leq \psi$ for step functions $\phi$ and $\psi$, we have $\int_{a}^{b} \phi \leq \int_{a}^{b} \psi$, so every element of $\mathcal{L}(f, [a, b])$ is $\leq$ every element of $\mathcal{U}(f, [a, b])$. 

We will now prove properties of the lower and upper integral that are analogous (but somewhat weaker) to the desired properties of the integral:

**Theorem 6.** Let $f, g : [a, b] \to \mathbb{R}$ be bounded functions and let $\lambda \in \mathbb{R}$.

1. (Scaling.) If $\lambda \geq 0$ then
   \[ \int_{a}^{b} \lambda f = \lambda \int_{a}^{b} f \quad \text{and} \quad \int_{a}^{b} \lambda f = \lambda \int_{a}^{b} f. \]

   If $\lambda \leq 0$ then
   \[ \int_{a}^{b} \lambda f = \lambda \int_{a}^{b} f \quad \text{and} \quad \int_{a}^{b} \lambda f = \lambda \int_{a}^{b} f. \]

2. (Addition.) $\int_{a}^{b} (f + g) \geq \int_{a}^{b} f + \int_{a}^{b} g$ and $\int_{a}^{b} (f + g) \leq \int_{a}^{b} f + \int_{a}^{b} g$.

3. (Monotonicity.) If $f \leq g$ on $[a, b]$ then $\int_{a}^{b} f \leq \int_{a}^{b} g$ and $\int_{a}^{b} f \leq \int_{a}^{b} g$.

4. (Almost-disjoint domains.) If $c \in [a, b]$, then $\int_{a}^{b} f = \int_{a}^{c} f + \int_{c}^{b} f$ and same for $\int$.

**Proof.** The proofs are all pretty similar and rely on just a couple elementary results about suprema and infima. Let $S, T \subseteq \mathbb{R}$ (and assume below that the supremum and infimum exist, where necessary):

- Define $\lambda S = \{ \lambda s : s \in S \}$. Then, if $\lambda \geq 0$, $\sup(\lambda S) = \lambda \sup(S)$ and $\inf(\lambda S) = \lambda \inf(S)$. If $\lambda \leq 0$, then $\sup(\lambda S) = \lambda \inf(S)$ and $\inf(\lambda S) = \lambda \sup(S)$.

- If $S \subseteq T$ then $\sup(S) \leq \sup(T)$ and $\inf(S) \geq \inf(T)$.

- Define $S + T = \{ s + t : s \in S, t \in T \}$. $\sup(S + T) = \sup(S) + \sup(T)$ and $\inf(S + T) = \inf(S) + \inf(T)$.

We will prove the first (in)equalities in (i), (ii), and (iii).
i. Let $\lambda \geq 0$. We have an equality of sets

$$\{ \phi : \phi \leq \lambda f \} = \{ \lambda \phi_0 : \phi_0 \leq f \}$$

Integrating the members of the left and the right, and using $\int_a^b \lambda \phi_0 = \lambda \int_a^b \phi_0$ because $\phi_0$ is a step function, we see that

$$\mathcal{L}(\lambda f, [a, b]) = \lambda \mathcal{L}(f, [a, b])$$

now we take suprema on both sides and use $\sup(\lambda S) = \lambda \sup(S)$. This establishes the first equality in (i) of the theorem.

ii. We have

$$\{ \phi_f + \phi_g : \phi_f \leq f, \phi_g \leq g \} \subseteq \{ \phi : \phi \leq f + g \}$$

Now, integrate the members of both sets above. Using $\int_a^b (\phi_f + \phi_g) = \int_a^b \phi_f + \int_a^b \phi_g$ on the left, we obtain

$$\mathcal{L}(f, [a, b]) + \mathcal{L}(g, [a, b]) \subseteq \mathcal{L}(f + g, [a, b])$$

and now we take suprema on both sides and use the properties listed at the beginning of the proof to establish the first inequality in (ii) of the theorem.

iii. If $\phi \leq f$, then $\phi \leq g$, so this immediately implies that $\mathcal{L}(f, [a, b]) \subseteq \mathcal{L}(g, [a, b])$, and this proves the first inequality in (iii) of the theorem.

(iv) is left as an optional exercise.

\[ \}\]

**Definition 7.** A function $h : E \to \mathbb{R}$ (with $E \subseteq \mathbb{R}$) is called Lipschitz continuous with Lipschitz factor $K$ if for all $x, y \in E$ we have $|h(x) - h(y)| \leq K|x - y|$.

Lipschitz continuity implies uniform continuity (and hence “just continuity”). A differentiable function is Lipschitz continuous on $[a, b]$ iff it has a bounded derivative on $(a, b)$ (this can be seen by dividing both sides above by $|x - y|$ and taking limits). It follows that $\sqrt{x}$ is uniformly continuous on $[0, 1]$ but not Lipschitz continuous, since its derivative is unbounded on $(0, 1)$.

**Theorem 8.** Let $f : [a, b] \to \mathbb{R}$ be a bounded function. The functions

$$F : [a, b] \to \mathbb{R} : x \mapsto \int_a^x f \quad \text{and} \quad \overline{F} : [a, b] \to \mathbb{R} : x \mapsto \int_a^x f$$

are Lipschitz continuous on $[a, b]$.

**Proof.** We will prove the claim for the first function, which we will call $F$ so I don’t have to type as much. Fix $M$ so that $|f| \leq M$ on $[a, b]$. Let $x < y$ for $x, y \in [a, b]$. We have (by (iv) in the previous theorem):

$$F(y) - F(x) = \int_x^y f$$
Now, observe that $-M \leq f \leq M$ on $[x, y]$, so using monotonicity of the lower integral (and the easy fact that $\int_x^y c = c(y - x)$ for a constant $c$), we have

$$-M(y - x) \leq F(y) - F(x) = \int_x^y f \leq M(y - x)$$

This inequality is equivalent to $|F(x) - F(y)| \leq M|x - y|$.

The above theorem is a very primitive version of the fundamental theorem of calculus. In some sense, integration “upgrades” bounded functions to continuous functions (later on, we’ll see that it upgrades continuous functions to differentiable functions, and so on).

2.4 Integrable functions

Let $f : [a, b] \to \mathbb{R}$ be bounded. We say that $f$ is integrable if its lower and upper integrals are equal, and then we define the (Riemann–Darboux) integral to be their common value:

$$\int_a^b f = \int_a^b f = \int_a^b f$$

Otherwise, $\int_a^b f < \int_a^b f$ and $f$ is non-integrable.

**Theorem 9.** (Cauchy criterion for integrability.) A bounded function $f : [a, b] \to \mathbb{R}$ is integrable if and only if for all $\epsilon > 0$ there exist step functions $\phi, \psi : [a, b] \to \mathbb{R}$ such that $\phi \leq f \leq \psi$ and

$$\int_a^b (\psi - \phi) = \int_a^b \psi - \int_a^b \phi < \epsilon$$

**Proof.** Easy, by definitions of the lower integral, the upper integral, and integrability.

**Theorem 10.** Let $f, g : [a, b] \to \mathbb{R}$ be integrable functions and let $\lambda \in \mathbb{R}$.

i. (Scaling.) $\lambda f$ is integrable and $\int_a^b \lambda f = \lambda \int_a^b f$.

ii. (Addition.) $f + g$ is integrable and $\int_a^b (f + g) = \int_a^b f + \int_a^b g$.

iii. (Monotonicity.) If $f \leq g$ on $[a, b]$, then $\int_a^b f \leq \int_a^b g$.

iv. (Almost-disjoint domains.) If $c \in [a, b]$, then $\int_a^b f = \int_a^c f + \int_c^b f$.

**Proof.** These follow almost directly from the properties proven for the lower and upper integrals. For (i), if $\lambda \geq 0$ we have

$$\int_a^b \lambda f = \lambda \int_a^b f = \lambda \int_a^b f = \int_a^b \lambda f$$
so $\lambda f$ is integrable, and $\int_a^b \lambda f = \lambda \int_a^b f$. The proof for $\lambda \leq 0$ is similar (with lower and upper integrals switched in the middle two expressions above).

For (ii), we have

$$\int_a^b (f + g) \leq \int_a^b f + \int_a^b g = \int_a^b f + \int_a^b g \leq \int_a^b (f + g)$$

but the reverse inequality is always true: $\int_a^b (f + g) \geq \int_a^b (f + g)$. Thus, equality holds, and so $f + g$ is integrable and $\int_a^b (f + g) \leq \int_a^b f + \int_a^b g$.

(iii) and (iv) follow directly from the versions for lower and upper integrals. \qed