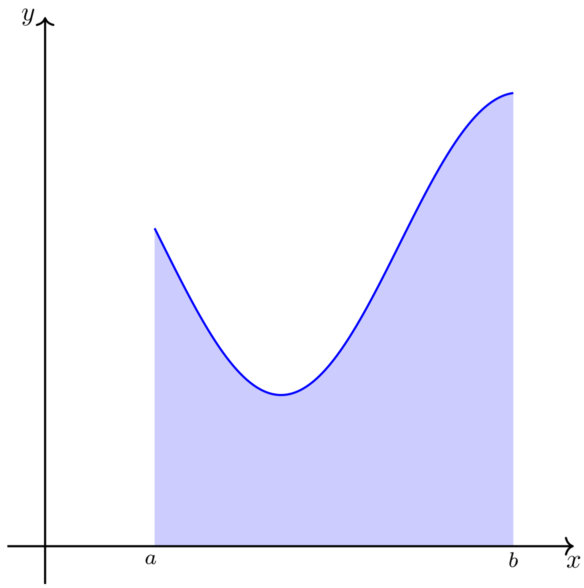
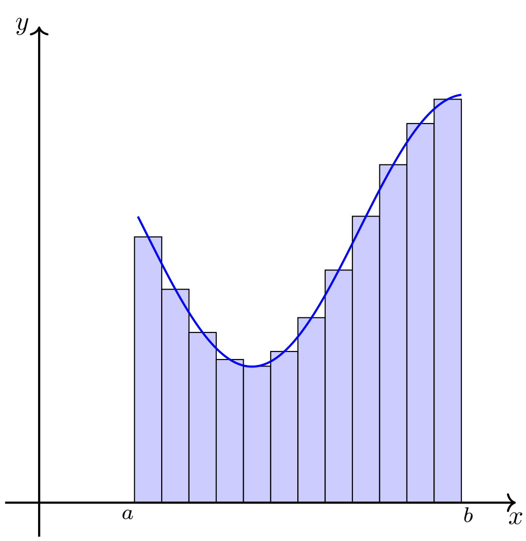
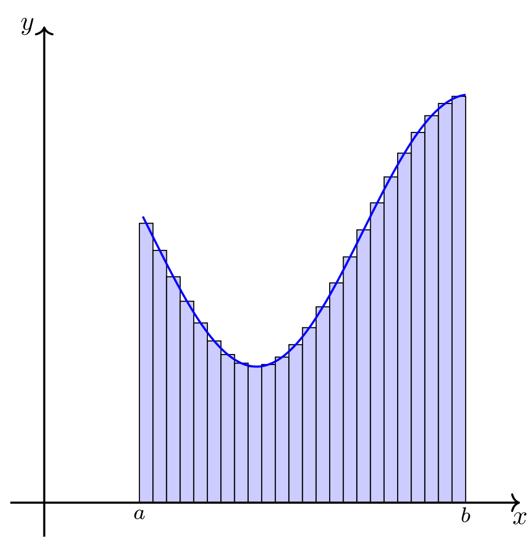
1 The Definite Integral

Motivation: *Given the graph of a function $y = f(x)$, what is the net area (the area above the x-axis and under the curve f minus the area below the x-axis and above the curve of f) of the graph between the points a and b?*



Strategy: We will divide the region $[a, b]$ into n subintervals and approximate the area by rectangles. The approximation improves by taking n larger and larger.





Definition 1. The general Riemann sum for f on the interval $[a, b]$ with n uniform subintervals is given by

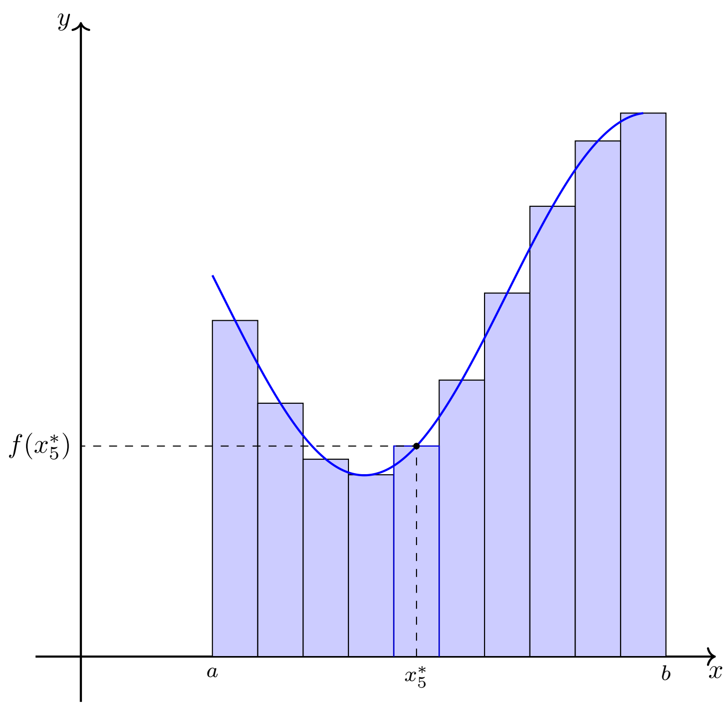
$$S_n = \sum_{i=1}^{n} f(x_i^*)\Delta x$$

where $\Delta x = \frac{b-a}{n}$ and $x_i^* \in [a + (i-1)\Delta x, a + i\Delta x]$. The approximate net area of the graph f is given by the Riemann Sum approximation.

Remark. We usually sample our function f at the right endpoint, midpoint, or left endpoint of each interval. The formula for x_i in each of these cases is given by:

1. Right Riemann Sum: $x_i^* = a + i\Delta x$
2. Midpoint Riemann Sum: $x_i^* = a + (i - \frac{1}{2})\Delta x$
3. Left Riemann Sum: $x_i^* = a + (i-1)\Delta x$.

The midpoint approximation can be visualized below



The limit of Riemann sum approximations as $n \to \infty$ gives the net area of the graph f on the interval $[a, b]$. This limit is called the definite integral of $f(x)$ on $[a, b]$.

Definition 2. We call the quantity $\int_a^b f(x)\,dx$ the *definite integral* of f on $[a, b]$, and it is defined by

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\Delta x$$

where $\Delta x = \frac{(b-a)}{n}$ and $x_i^* \in [a + (i-1)\Delta x, a + i\Delta x]$. If the number $\int_a^b f(x)\,dx$ exists[1], then we say f is *integrable* on $[a, b]$.

1.1 Useful Formulas

The following formulas for the partial sums of a number will be useful to compute the Riemann Sums of certain functions

1. Sum of first n constants:

$$\sum_{i=1}^{n} 1 = n. \tag{1}$$

[1]The limit has to exist and must all be identical for all choices of samples x_i^*.

2. Sum of first n integers:
$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}. \tag{2}$$

3. Sum of first n squares:
$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}. \tag{3}$$

4. Sum of first n cubes:
$$\sum_{i=1}^{n} i^3 = \Big(\frac{n(n+1)}{2}\Big)^2. \tag{4}$$

5. Geometric series:
$$\sum_{i=1}^{n} r^i = r\Big(\frac{1-r^n}{1-r}\Big). \tag{5}$$

1.2 Example Problems

Problem 1. ($\star$) Approximate the value of $\int_1^2 \ln(x)\,dx$ by using a left endpoint Riemann sum and 4 uniform subintervals.

Solution 1. We take $f(x) = \ln(x)$, $a = 1$, $b = 2$, and $n = 4$ in Definition 1. Since we are sampling at the left endpoints, we choose $x_i = 1 + (i-1)\Delta x$ where $\Delta x = \frac{b-a}{n} = \frac{1}{4}$. Using our formula, we have

$$\begin{aligned} S_4 = \sum_{i=1}^{4} f\big(1 + (i-1)\Delta x\big)\Delta x &= \sum_{i=1}^{4} f\Big(1 + \frac{i-1}{4}\Big)\frac{1}{4} \\ &= \frac{1}{4}\sum_{i=1}^{4} \ln\Big(1 + \frac{i-1}{4}\Big) \\ &= \frac{1}{4}\Big(\ln(1) + \ln(1.25) + \ln(1.5) + \ln(1.75)\Big) \approx 0.2970\ldots \end{aligned}$$

Problem 2. ($\star$) Approximate the area under the curve $y = 2x$ above the x-axis on the interval $[0, 10]$ using 10 uniform subintervals and samples at the right endpoint of each interval.

Solution 2. We take $f(x) = 2x$, $a = 0$, $b = 10$, and $n = 10$ in Definition 1. Since we are sampling at the right endpoints, we choose $x_i^* = i\Delta x$ where $\Delta x = \frac{b-a}{n} = 1$. Using our formula, we have

$$\begin{aligned} S_{10} = \sum_{i=1}^{10} f(i\Delta x)\Delta x = \sum_{i=1}^{10} 2i &= 2\sum_{i=1}^{10} i \\ &= 2\frac{10(10+1)}{2} = 110. \qquad \text{since } \sum_{i=1}^{n} i = \frac{n(n+1)}{2}. \end{aligned}$$

Problem 3. ($\star\star$) Approximate the area under the curve $y = 2x$ above the x-axis on the interval $[0, 10]$ using n uniform subintervals and samples at the right endpoint of each interval. What does the sum converge to when we take $n \to \infty$?

Solution 3. We take $f(x) = 2x$, $a = 0$, $b = 10$, with variable n in Definition 1. Since we are sampling at the right endpoints, we choose $x_i^* = i\Delta x$ where $\Delta x = \frac{10-0}{n} = \frac{10}{n}$. Using our formula, we have

$$\begin{aligned} S_n = \sum_{i=1}^{n} f\big(i\Delta x\big)\Delta x = \sum_{i=1}^{n} 2 \cdot \frac{10i}{n} \cdot \frac{10}{n} &= \frac{200}{n^2}\sum_{i=1}^{n} i \\ &= \frac{200}{n^2} \cdot \frac{n(n+1)}{2} = 100 \cdot \frac{n+1}{n}. \qquad \text{since } \sum_{i=1}^{n} i = \frac{n(n+1)}{2} \end{aligned}$$

Taking $n \to \infty$, we have

$$\lim_{n\to\infty} S_n = \lim_{n\to\infty} 100 \cdot \frac{n+1}{n} = 100.$$

Remark. The final answer is the same as

$$\int_0^{10} 2x\,dx = x^2\Big|_{x=0}^{x=10} = 100.$$

This is also the same as the area of a triangle with base 10 and height 20.

Problem 4. ($\star\star$) Approximate the area under the curve $y = x^2$ above the x-axis on the interval $[0, 1]$ using 100 uniform subintervals and samples at the left endpoint of each interval.

Solution 4. We take $f(x) = x^2$, $a = 0$, $b = 1$, and $n = 100$ in Definition 1. Since we are sampling at the left endpoints, we choose $x_i^* = (i-1)\Delta x$ where $\Delta x = \frac{1}{100}$. Using our formula, we have

$$\begin{aligned} S_{100} = \sum_{i=1}^{100} f\big((i-1)\Delta x\big)\Delta x &= \sum_{i=1}^{100} \Big(\frac{i-1}{100}\Big)^2 \frac{1}{100} \\ &= \frac{1}{100^3}\sum_{i=1}^{100}(i-1)^2 \\ &= \frac{1}{100^3}\sum_{i=0}^{99} j^2 && \text{by reindexing } j = i-1 \\ &= \frac{1}{100^3} \cdot \frac{99(100)(199)}{6} && \text{since } \sum_{j=0}^{n} j^2 = \sum_{j=1}^{n} j^2 = \frac{n(n+1)(2n+1)}{6} \\ &= 0.32835. \end{aligned}$$

Problem 5. ($\star\star$) Approximate the area under the curve $y = x^2$ above the x-axis on the interval $[1, 5]$ using 100 uniform subintervals and samples at the right endpoint of each interval.

Solution 5. We take $f(x) = x^2$, $a = 1$, $b = 5$, and $n = 100$ in Definition 1. Since we are sampling at the right endpoints, we choose $x_i^* = 1 + i\Delta x$ where $\Delta x = \frac{5-1}{100} = \frac{1}{25}$. Using our formula, we have

$$\begin{aligned}
S_{100} &= \sum_{i=1}^{100} f\Big(1 + i\Delta x\Big)\Delta x \\
&= \sum_{i=1}^{100} \left(1 + \frac{i}{25}\right)^2 \cdot \frac{1}{25} \\
&= \frac{1}{25^3}\sum_{i=1}^{100}(25+i)^2 \\
&= \frac{1}{15625}\sum_{i=1}^{100}(625 + 50i + i^2) \\
&= \frac{1}{15625}\left(625\sum_{i=1}^{100} 1 + 50\sum_{i=1}^{100} i + \sum_{i=1}^{100} i^2\right) \\
&= \frac{1}{15625}\left(625 \cdot 100 + 50 \cdot \frac{100 \cdot 101}{2} + \frac{100(101)(201)}{6}\right) \qquad \text{formulas (1), (2), (3)} \\
&= 41.8144.
\end{aligned}$$

Problem 6. $(\star\star\star)$ Approximate the area under the curve $y = x^2$ above the x-axis on the interval $[0, 1]$ using n uniform subintervals and samples at the midpoint of each interval. What does the sum converge to when we take $n \to \infty$?

Solution 6. We take $f(x) = x^2$, $a = 0$, $b = 1$, with variable n in Definition 1. Since we are sampling at the midpoints of the intervals, we choose $x_i^* = (i - \frac{1}{2})\Delta x$ where $\Delta x = \frac{1}{n}$. Using our formula, we have

$$\begin{aligned}
S_n &= \sum_{i=1}^{n} f\left(\left(i - \frac{1}{2}\right)\Delta x\right)\Delta x \\
&= \sum_{i=1}^{n}\left(\frac{2i-1}{2n}\right)^2 \frac{1}{n} \\
&= \frac{1}{4n^3}\sum_{i=1}^{n}(2i-1)^2 \\
&= \frac{1}{4n^3}\sum_{i=1}^{n}(4i^2 - 4i + 1) \\
&= \frac{1}{4n^3}\left(4\sum_{i=1}^{n} i^2 - 4\sum_{i=1}^{n} i + \sum_{i=1}^{n} 1\right) \\
&= \frac{1}{4n^3}\left(4 \cdot \frac{n(n+1)(2n+1)}{6} - 4 \cdot \frac{n(n+1)}{2} + n\right) \qquad \text{using formulas (1), (2), (3)} \\
&= \frac{n(n+1)(2n+1)}{6n^3} - \frac{(n+1)}{2n^2} + \frac{1}{4n^2}.
\end{aligned}$$

As $n \to \infty$, we have

$$\lim_{n\to\infty} S_n = \lim_{n\to\infty}\left(\frac{n(n+1)(2n+1)}{6n^3} - \frac{(n+1)}{2n^2} + \frac{1}{4n^2}\right) = \frac{1}{3}.$$

Remark. The final answer is the same as

$$\int_0^1 x^2\,dx = \frac{x^3}{3}\Big|_{x=0}^{x=1} = \frac{1}{3}.$$

2 Properties of Definite Integrals

2.1 Basic Properties

Let f and g be integrable functions, the definite integrals obey the following properties:

1. Changing the Index:
$$\int_a^b f(x)\,dx = \int_a^b f(t)\,dt.$$

2. Linearity: If $c, d \in \mathbb{R}$
$$\int_a^b cf(x) + dg(x)\,dx = c\int_a^b f(x)\,dx + d\int_a^b g(x)\,dx.$$

3. Monotonicity: If $f(x) \le g(x)$ for all $x \in [a, b]$ then
$$\int_a^b f(x)\,dx \le \int_a^b g(x)\,dx.$$

4. Splitting the Region of Integration: For any a, b,
$$\int_a^a f(x)\,dx = 0 \qquad \text{and} \qquad \int_a^b f(x)\,dx = -\int_b^a f(x)$$
and for any c,
$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

5. Symmetry: If $f(x)$ is an odd function [i.e. $f(-x) = -f(x)$], then
$$\int_{-a}^a f(x)\,dx = 0,$$
and if $f(x)$ is an even function [i.e. $f(-x) = f(x)$], then
$$\int_{-a}^a f(x)\,dx = 2\int_0^a f(x)\,dx.$$

2.2 Accuracy of Riemann Sum Approximations

Without doing any computations, we can determine if the Riemann sums are over or under approximations by looking at the shape of the curve we want to estimate the area of:

$f(x)$	Left	Right	Midpoint
Increasing	Under	Over	?
Decreasing	Over	Under	?
Convex	?	?	Under
Concave	?	?	Over

For example, the table says that if $f(x)$ is increasing on $[a, b]$, then the left Riemann sum is an under approximation of the definite integral, and the right Riemann sum is an over approximation of the definite integral. The fact f is increasing does not tell us enough to determine if the midpoint is an over or under approximation in general.

2.3 Example Problems

Problem 1. ($\star$) If $\int_1^6 f(x)\,dx = 4$, $\int_1^2 f(x)\,dx = 3$, and $\int_5^6 f(x)\,dx = 7$, find $\int_2^5 f(x)\,dx$.

Solution 1. Using the fact that

$$\int_1^6 f(x)\,dx = \int_1^2 f(x)\,dx + \int_2^5 f(x)\,dx + \int_5^6 f(x)\,dx,$$

we have

$$\int_2^5 f(x)\,dx = \int_1^6 f(x)\,dx - \int_1^2 f(x)\,dx - \int_5^6 f(x)\,dx = 4 - 3 - 7 = -6.$$

Problem 2. ($\star\star$) The cumulative density function of the exponential random variable is a function defined on $[0,\infty)$ given by

$$F_\lambda(x) = \lambda \int_0^x e^{-\lambda t}\,dt,$$

where $\lambda > 0$ is a fixed constant. Express the definite integral $\int_5^9 e^{-\lambda t}\,dt$ in terms of $F_\lambda(x)$.

Solution 2. We start by splitting the region of integration

$$\int_5^9 e^{-\lambda t}\,dt = \int_5^0 e^{-\lambda t}\,dt + \int_0^9 e^{-\lambda t}\,dt = -\int_0^5 e^{-\lambda t}\,dt + \int_0^9 e^{-\lambda t}\,dt.$$

We now multiply and divide by λ to conclude

$$\int_5^9 e^{-\lambda t}\,dt = \frac{1}{\lambda}\left(\lambda \int_0^9 e^{-\lambda t}\,dt - \lambda \int_0^5 e^{-\lambda t}\,dt\right) = \frac{1}{\lambda}\cdot (F_\lambda(9) - F_\lambda(5)).$$

Problem 3. ($\star\star\star$) Prove that

$$\sum_{i=1}^N i^4 \geq \int_0^N x^4\,dx = \frac{N^5}{5}.$$

Solution 3. Notice that the sum corresponds to the right Riemann sum approximation of $f(x) = x^4$ with N uniform subintervals. We take $a = 0$, $b = N$, and $n = N$ in Definition 1. Since we are sampling at the right endpoints, we choose $x_i^* = i\Delta x$ where $\Delta x = \frac{N-0}{N} = 1$. Using our formula, we have

$$S_N = \sum_{i=1}^N f(i\Delta x)\Delta x = \sum_{i=1}^N f(i\Delta x)\Delta x = \sum_{i=1}^N f(i) = \sum_{i=1}^N i^4.$$

Since $f'(x) = 4x^3$ is increasing on $[0, N]$, the right Riemann sum is an over approximation of the definite integral, so

$$\sum_{i=1}^N f(i) \geq \int_0^N f(x)\,dx.$$

We will show in Week 2 that the definite integral is given explicitly by

$$\int_0^N x^4\,dx = \frac{x^5}{5}\Big|_{x=0}^{x=N} = \frac{N^5}{5}.$$