2. INTEGRATION

§2.1. A Review of Integral Calculus

An (indefinite) **integral** (or **antiderivative**) of a function f(x), where one exists, is a function g(x) whose derivative is f(x). A function is defined to be **integrable** if it has an integral. Any two integrals differ by a constant and we write them as $\int f(x)dx = g(x) + c$, where c is called an **arbitrary constant**.

If u, v are functions of x and k is a constant then

$$\int(u + v)dx = \int u\,dx + \int v\,dx \text{ and}$$

$$\int ku\,dx = k\int u\,dx\,.$$

Polynomial-like functions are integrable and their integrals can be obtained from the fact that:

$$\int x^n\,dx = \begin{cases} \dfrac{1}{n+1}x^{n+1} + c & \text{if } n \neq -1 \\ \log x + c & \text{if } n = -1 \end{cases}.$$

Example 1:

$$\int\left(4x^3 - \sqrt{x} - \frac{7}{x} + \frac{3}{x^2}\right)dx = x^4 - \frac{2}{3}x^{3/2} - 7\log x - \frac{3}{x} + c$$

The integral of e^x is e^x and the integral of sin x is − cos x.

If $a \le b$ and $g(x) = \int f(x)\,dx$ then $\int_a^b f(x)\,dx = [g(x)]_a^b$,

where this denotes $g(b) - g(a)$.

This is called a **definite integral** and if $f(x) \ge 0$ on the interval $[a, b]$ it gives the area between the graph $y = f(x)$ and the x-axis between $x = a$ and $x = b$.

§2.2. The Riemann Integral

We've defined integrals as antiderivatives. This is fine, up to a point, but it means that we can only prove that an integral exists if we can actually find it. Many integrals can't be expressed in terms of functions that we already know about, such as polynomials, trigonometric, logarithmic or exponential functions. There's no way we can define new functions as antiderivatives unless we happen to know the answer.

For example we might define a new function as $\phi(x) = \int e^{x^2}\,dx$. But how do you make a definition in terms of something until you know what that something is? To get around this difficulty we now define an integral in a way that's independent of knowing what the integral is. There are several definitions but they all give the same answer. It's just that some integrals can be defined in circumstances where others don't exist. Here we'll deal with the simplest of these integrals, the **Riemann integral**. This is named after the important German mathematician Bernhard Riemann [1826–1866].

Bernhard Riemann

Suppose f(x) is continuous on the closed interval [a, b].

We define a **partition** of [a, b] to be a finite sequence:
$\Delta = (x_0, x_1, ..., x_n)$ such that
$a = x_0 < x_1 < ... < x_n = b$.

We define the **norm** of the Δ to be the maximum of $\{x_{i+1} - x_i\}$ and we denote it by $|\Delta|$.

The **Riemann lower sum** corresponding to this partition is

$$s_\Delta(f) = \sum_{i=0}^{n-1} (x_{i+1} - x_i)m_i$$

where m_i is the minimum value of f(x) on $[x_i , x_{i+1}]$.

The **Riemann upper sum** corresponding to the partition is

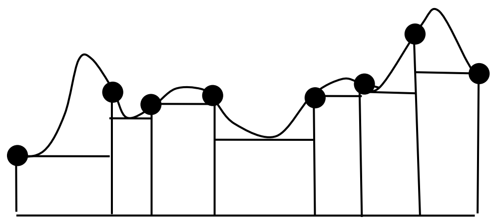
$$S_\Delta(f) = \sum_{i=0}^{n-1} (x_{i+1} - x_i)M_i .$$

where M_i is the maximum value of f(x) on [x_i , x_{i+1}].

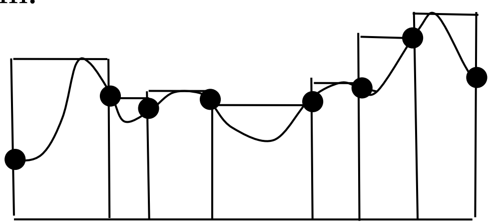
Now what exactly is going on here? We are leading up to a definition of the integral as the (signed) area under the curve. For convenience in this explanation suppose that f(x) ≥ 0 for $x \in [a, b]$, though everything works without this assumption.

Suppose we draw rectangles up from the x-axis using ordinates at the x_i and suppose that the heights are the values of m_i.

The sum of the areas in the following diagram is the lower Riemann sum.



The sum of the areas in the following diagram is the upper Riemann sum.



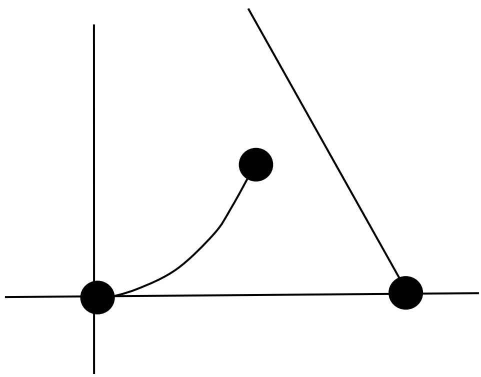
Note that we don't insist on a dissection with equal widths. The total area of the rectangles will be approximately the area under the curve. In fact the area under the curve will lie between the lower Riemann sum and the upper one.

If the limits of $s_\Delta(f)$ and $S_\Delta(f)$, as $|\Delta|$ approaches 0 exist and are equal we define this value to be the **integral** of f(x) over the interval [a, b]. We write this value as $\int_a^b f(x)\,dx$. We say that f(x) is **integrable on [*a*, *b*]** if such an integral exists and it is **integrable** if it is integrable over every relevant closed interval.

Example 2:

The function $f(x) = \begin{cases} x^2 \text{ if } x \le 1 \\ 4 - 2x \text{ if } x > 1 \end{cases}$ is not continuous on [0, 2] but it is integrable on [0, 2].

Solution:



Take the partition $\Delta_{m,n}$:

$$0 < \frac{1}{n} < \frac{2}{n} < \ldots < 1 < \frac{n-1}{n} < 1 + \frac{1}{m} < 1 + \frac{2}{m} < \ldots < 1 + \frac{m-1}{m} < 2$$

for some positive integers m, n. (Here I've chosen to divide the interval [0, 2] into $m + n$ parts. There's no good reason for different widths for each interval, except to emphasise that we don't need to have a uniform partition.)

Here $|\Delta_{m,n}| = \text{MAX}\left(\frac{1}{m}, \frac{1}{n}\right)$.

Recall that $\sum_{r=1}^{n} r = \frac{1}{2} n(n + 1)$ and

$$\sum_{r=1}^{n} r^2 = \frac{1}{6} n(n + 1)(2n + 1).$$

The lower Riemann sum for this partition is:

$$\sum_{r=0}^{n-1} \left(\frac{1}{n}\right)\left(\frac{r^2}{n^2}\right) + \sum_{r=0}^{m-1} \left(\frac{1}{m}\right)\left(2 - \frac{2(r+1)}{m}\right)$$

$$= \frac{1}{n^3} \sum_{r=0}^{n-1} r^2 + \frac{2m}{m} - \frac{2}{m^2} \sum_{r=0}^{m-1} (r + 1)$$

$$= \frac{(n-1)n(2n-1)}{6n^3} + 2 - \left(\frac{2}{m^2}\right)\left(\frac{m^2 + m}{2}\right)$$

$$= \frac{(n-1)(2n-1)}{6n^2} + 1 - \frac{1}{m} = \frac{4}{3} - \frac{3n-1}{6n^2} - \frac{1}{m}.$$

The upper Riemann sum for this partition is

$$\sum_{r=0}^{n-1}\left(\frac{1}{n}\right)\left(\frac{(r+1)^2}{n^2}\right) + \sum_{r=0}^{m-1}\left(\frac{1}{m}\right)\left(2-\frac{2r}{m}\right)$$

$$= \frac{1}{n^3}\sum_{r=0}^{n-1}(r+1)^2 + \frac{2m}{m} - \frac{2}{m^2}\sum_{r=0}^{m-1} r$$

$$= \frac{n(n+1)(2n+1)}{6n^3} + 2 - \left(\frac{2}{m^2}\right)\left(\frac{m^2-m}{2}\right)$$

$$= \frac{(n+1)(2n+1)}{6n^2} + 1 + \frac{1}{m}\ .$$

$$= \frac{1}{3} + \frac{3n+1}{6n^2} + 1 + \frac{1}{m}$$

$$= \frac{4}{3} + \frac{3n+1}{6n^2} + \frac{1}{m}\ .$$

Now $|\Delta_{m,n}| = \text{MAX}\left(\frac{1}{m}, \frac{1}{n}\right)$.

As this approaches zero, both of these sums approach $\frac{4}{3}$ so we conclude that $\int_0^2 f(x)\,dx = \frac{4}{3}$.

Before we knew about the Riemann integral we would have written

$$\int_a^b f(x)\,dx = \int_0^1 x^2\,dx + \int_1^2 (4-2x)\,dx$$

$$= \left[\frac{x^3}{3}\right]_0^1 + [4x - x^2]_1^2$$

$$= \frac{1}{3} + (8 - 4) - (4 - 1) = \frac{4}{3}$$

That is because we thought of integrals as antiderivatives. And once we prove the Fundamental Theorem of Calculus we can drop back into our old habits. So why all the fuss about the Riemann integral? The answer is that there are integrals, such as $\int e^{-x^2}\,dx$,where we can't find an existing function whose derivative is e^{-x^2}. So we might define a new function $\phi(x) = \int_0^x e^{-t^2}\,dt$. But this involves circular reasoning. We need to define the integral independently of differentiation, and this is what the Riemann integral does.

Now our example shows that a function can be integrable even if it has a discontinuity or two. Too many discontinuities, however, may render the function non integrable.

Example 3: The function $f(x) = \begin{cases} 1 \text{ if } x \text{ is rational} \\ 0 \text{ if } x \text{ is irrational} \end{cases}$ is not integrable on [0, 1]

Solution: For any partition the lower Riemann sum will be 0 and the upper Riemann sum will be 1. Hence, although the limits of these Riemann sums exist, they are not equal.

The following theorem guarantees the integrability of most of the functions we find useful. The proof can be found in books on Real Analysis.

Theorem 1: If f(x) is continuous on the interval [a, b] then it's integrable on [a, b].

Example 4: The function f(x) = e^{-x^2} is integrable because it's continuous. The fact that we can't express the integral in terms of the elementary functions doesn't affect its integrability.
We really should have used the term 'Riemann integrable' instead of 'integrable' because there are generalisations of the Riemann integral, such as the Lesbegue integral, which can provide integrals for some functions for which the Riemann integral doesn't exist. We won't discuss this here as it properly belongs to an advanced area of mathematics called Measure Theory.

At this stage we could prove all the familiar properties of the integral directly from the Riemann integral definition. However I don't think it very instructive to do so. You'll have seen proofs of these using the properties of the derivative, based on the fact that integration is anti-differentiation. This fact is enshrined in the so-called Fundamental Theorem of Calculus.

§2.3. The Fundamental Theorem of Calculus

Theorem 2A (Fundamental Theorem of Calculus A): Suppose f(x) is continuous on the interval [a, b].

Let F(x) be the Riemann integral $\int_a^x f(x)\, dx$ for $x \in [a, b]$.

Then F(x) is differentiable on [a, b] and $\frac{dF(x)}{dx} = f(x)$.

Proof: Let y = F(x).

Then $y + \Delta y = F(x + \Delta x) = \int_a^{x+\Delta x} f(x)\, dx$.

$$\text{So } F(x + \Delta x) - F(x) = \int_a^{x+\Delta x} f(x)\, dx - \int_a^{x} f(x)\, dx$$

$$= \int_x^{x+\Delta x} f(x)\, dx$$

By the Mean Value Theorem for Integrals:

$$\int_x^{x+\Delta x} f(x)\,dx = f(c).\Delta x \text{ for some } c(x,\Delta x) \text{ with}$$

$$x < c(x,\Delta x) < x + \Delta x.$$

Hence $\dfrac{F(x + \Delta x) - F(x)}{\Delta x} = f(c(x,\Delta x))$

So $\dfrac{dF(x)}{dx} = \lim\limits_{\Delta x\to 0} \dfrac{F(x + \Delta x) - F(x)}{\Delta x} = \lim\limits_{\Delta x\to 0} f(c(x,\Delta x)).$

Since $x < c(x,\Delta x) < x + \Delta x$, by the Squeeze Law,

$\lim\limits_{\Delta x\to 0} f(c(x,\Delta x)) = \lim\limits_{\Delta x\to 0} f(x) = f(x).$

Hence $\dfrac{dF(x)}{dx} = f(x)$. ✋☺

Theorem 2B (Fundamental Theorem of Calculus B): If f(x) is differentiable on [a, b] then f ′(x) is integrable and

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

Proof: The proof of part B is somewhat longer and so I will omit the proof. ✋

The following table of integrals is obtained by checking that the derivative of the integral is the function itself, using the Fundamental Theorem of Calculus. (The ‘+ c’ is omitted.

f(x)	$\int$f(x) dx
x^n for $n \neq -1$	$\dfrac{x^{n+1}}{n+1}$
$\dfrac{1}{x}$	log x
e^x	e^x
log x	x log $x - x$
sin x	$-$ cos x
cos x	sin x
tan x	log(sec x)

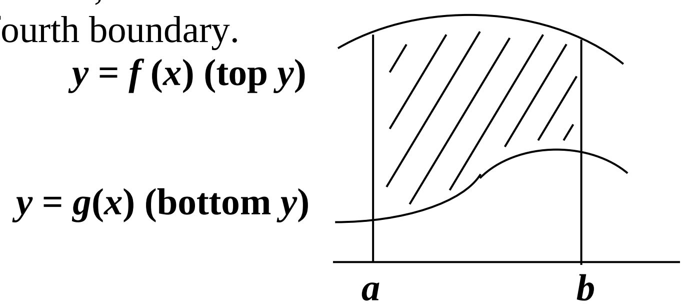
§2.4 Areas Between Curves

From the definition of the Riemann Integral we can see that if $f(x) \geq 0$ for all $x \in [a, b]$ then $\int_b^a f(x)\,dx$ is the area enclosed between the curve $y = f(x)$, the x-axis and the lines $x = a$ and $x = b$. Traditionally we've called this the area *below* the curve.

But if $f(x) \leq 0$ on $[a, b]$ the m_i and M_i in the definition of the Riemann Integral are negative, or zero. Since our concept of area is essentially something that's positive, or zero, we'd have $\int_{b}^{a} f(x)\, dx$ as representing minus the area enclosed between the curve, the *x*-axis and the two ordinates. And here the integral gives minus the area *above* the curve.

So areas above the *x*-axis are positive and areas below are negative, right? Wrong! We lied! Well, when you first learn about integration it's a convenient fiction that's true in a certain sense. But now we'll level with you and tell you the whole truth.

The boundaries for an area that's given by an integral are the graph of a function, two vertical lines and, up to now the fourth boundary has been the *x*-axis. We're now going to generalise it that so that the fourth boundary is the graph of a second function. You can still take the second function to be $y = 0$ and that gives the *x*-axis. You can still have the *x*-axis as a boundary if you like. But you can, instead, have a curved fourth boundary.



We're going to develop a formula for the area between two curves. There'll be a 'top y' and a 'bottom y'. 'Top' and 'bottom' refer to which has the greater *y* values. Of course curves may cross over and this makes the problem just a little bit more complicated, but not much.
The shaded area is the area we want.

Now $\int_a^b f(x)\,\mathrm{d}x$ and $\int_a^b g(x)\,\mathrm{d}x$ respectively give the area under the top curve, right down to the *x*-axis and the area under the bottom curve, right down to the *x*-axis. Clearly the area we want is the difference between the two, that is, $\int_a^b f(x)\,\mathrm{d}x - \int_a^b g(x)\,\mathrm{d}x = \int_a^b [f(x) - g(x)]\,\mathrm{d}x$. We have thus shown the following.

Theorem 3: If $f(x) \geq g(x)$ for $x \in [a, b]$ the area enclosed between $y = f(x)$, $y = g(x)$, $x = a$ and $x = b$ is

$$\int_a^b [f(x) - g(x)]\,\mathrm{d}x$$ ✋☺

To remember this, just remember the integral as:

$$\int_{\textbf{bottom limit}}^{\textbf{top limit}} \textbf{(top } y \textbf{ – bottom } y\textbf{) d}x$$

If we have just one curve $y = f(x)$ which lies *above* the x-axis from $x = a$ to $x = b$ and the bottom boundary of the area is the x-axis, then the top y is $y = f(x)$ and the bottom y is $y = 0$. So the area under the graph is:

$$\int_a^b [f(x) - 0]\,dx) = \int_a^b f(x)\,dx \text{ , as before.}$$

But if $y = f(x)$ lies *below* the x-axis this becomes the bottom boundary and the x-axis, $y = 0$ becomes the top boundary. The area becomes, in this case:

$$\int_a^b [0 - f(x)]\,dx = -\int_a^b f(x)\,dx \text{ .}$$

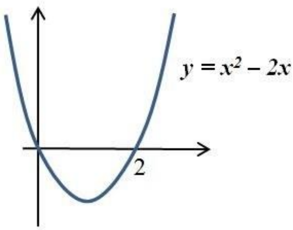
Previously we explained this minus sign by saying that "areas below the x-axis are negative". Now we're in a position to set the record straight. Areas are *never* negative, but integrals may be. Where the curve is below the x-axis the area is positive but the integral is negative. That's the right way to look at it.

Whenever you have to work out areas below, above or between curves always set it up as a problem involving the area *between* two curves. If you correctly identify which is the 'top y' and which is the 'bottom y' you'll never go wrong. The integral you set up will always come out as positive and will give the required area. Areas come out negative, when dealing with areas, only if you haven't set things up properly.

There's a lot of sloppy thinking with this topic and students are often taught bad habits. Alright we did start you off thinking that "areas below the axis are negative" but at the time it might have confused you to do otherwise. At least we're now setting the story right.

Some students think that since areas are positive but integrals can be negative, all you have to do is to put absolute value signs around your final answer. Even worse is the student who, on evaluating an integral to be –2, writes "= – 2 = 2". Even using absolute value signs display ignorance as to what is really going on and can lead to wrong answers if the curves cross over.

For this reason we put a modulus sign around the integral if we are finding the actual area between the curve and the x axis when the area is under the x axis.



$$\text{Area} = \left| \int_0^2 x^2 - 2x \ dx \right|$$

$$= \left| \quad \frac{-4}{3} \quad \right| \qquad = \frac{4}{3}$$

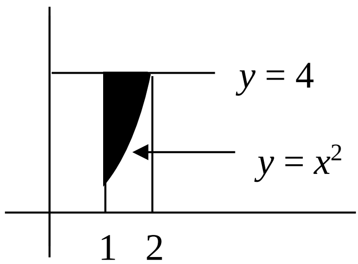
There's no actual error in this snippet that I found on the web, but it's creates the impression that you can always just integrate and then take the absolute value. I've seen many students who've ignored the fact that curves cross over and who rely on absolute values to get a positive, albeit wrong, answer.

So some guidelines to avoid such errors are:

- Never 'fudge' the sign if the integral comes out negative.

- Never use absolute value signs in connection with areas.

Example 5:
Find the area between $y = x^2$, $y = 4$ and $x = 1$.



Solution: Top y is $y = 4$ and bottom y is $y = x^2$.
The area is therefore:

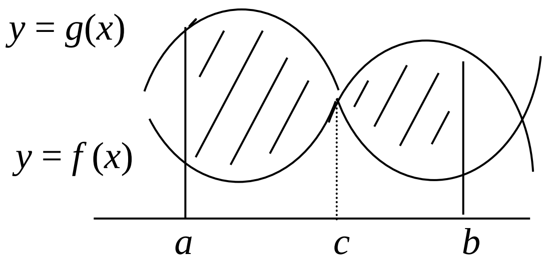
$$\int_1^2 [4 - x^2]\, dx = \left[4x - (1/3)x^3\right]_1^2 = \left(8 - \frac{8}{3}\right) - \left(4 - \frac{1}{3}\right) = \frac{5}{3}.$$

§2.5. What If Curves Cross?

It's absolutely important to determine whether curves cross within the region being considered. For if they do, what was the 'top y' will become the 'bottom y' and vice versa. If we want the area between the curve and the x-axis the interval of integration must be split up.

Suppose we want the area between the curves $y = f(x)$ and $g(x)$ between $x = a$ and $x = b$ and suppose the curves cross at some point c between a and b. Suppose that $g(\mathrm{x}) \geq f(x)$ for $x \leq c$ and $g(x) \leq f(x)$ for $x \geq c$. Then from $x = a$ to $x = c$ it's $g(x)$ which is the top y, while from $x = c$ to $x = b$ it's $f(x)$ that's on top.

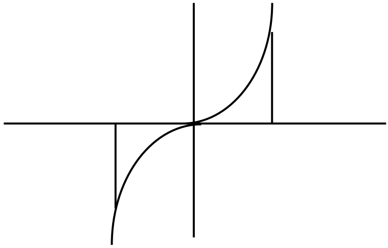
We must break the interval [a, b] into two pieces and integrate separately over these two regions.



The required area will be:

$$\int_a^c [g(x) - f(x)]\,\mathrm{d}x + \int_c^b [f(x) - g(x)]\,\mathrm{d}x\,.$$

Example 6: Find the area between $y = x^3$, the x-axis and lines $x = -1$ and $x = 1$.



Solution: Because the curve cuts the x-axis at $x = 0$ we must divide the interval [–1, 1] into two pieces.

From $x = -1$ to $x = 0$ the top y is $y = 0$ (the x-axis) while from $x = 0$ to $x = 1$ it is $y = x^3$.
The required area is:

$$\int_{-1}^{0}[0 - x^3]\,dx + \int_{0}^{1}[x^3 - 0]\,dx = \left[-\frac{x^4}{4}\right]_{-1}^{0} + \left[\frac{x^4}{4}\right]_{0}^{1}$$

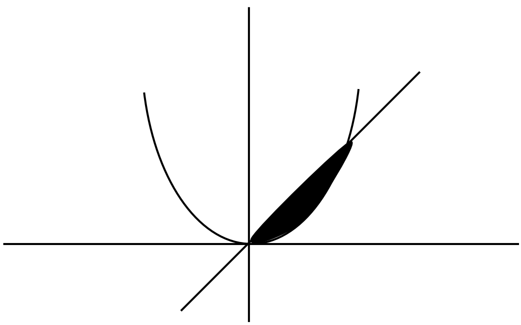
$$= \left(0 - \left(-\tfrac{1}{4}\right)\right) + \left(\tfrac{1}{4} - 0\right) = \frac{1}{2}.$$

Two things should be noted with this example.
(1) If we'd simply integrated x^3 between –1 and 1 we would have concluded that the area is:

$\int_{-1}^{1} x^3\,dx = \left[\frac{x^4}{4}\right]_{-1}^{1} = \frac{1}{4} - \frac{1}{4} = 0$, which is clearly wrong. And taking absolute values wouldn't help either. Remember to *never* use absolute values with these area questions. If you don't know which is top y and which is bottom you're likely to come up with a totally wrong answer. If you do, then you'll set up the integral correctly so that it will come out positive.

(2) In this case we could have exploited symmetry to simplify the calculation and avoid all those minus signs. The area from $x = -1$ to $x = 0$ is clearly the same as the area from $x = 0$ to $x = 1$, so we could have said that the total area $= 2\int_0^1 x^3\,dx = 2\left[\frac{x^4}{4}\right]_0^1 = \frac{1}{2}$.

Example 7: Find the area between $y = x^2$ and $y = x$. **Solution:** Here we're not given any endpoints. That's because the graphs of $y = x^2$ and $y = x$ cut in exactly two places and so they enclose a region. We need to find these two points.

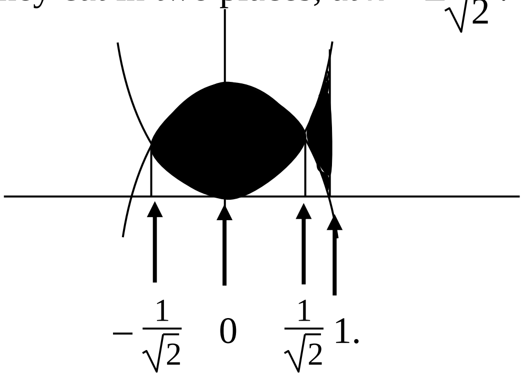


Solving $x^2 = x$ we get $x = 0$ or 1. So these are the limits of integration. The top y, between 0 and 1, is $y = x$. The area is therefore $\int_0^1 [x - x^2]\, dx$. Because we know what we're doing and have set up the integral correctly we can be sure that it will be positive.

The required area is $\int_0^1 [x - x^2]\, dx = \left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1 = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}$

Example 8: Find the area between $y = x^2$, $y = 1 - x^2$ and $x = 1$.

Solution: The curves cut when $x^2 = 1 - x^2$, that is when $2x^2 = 1$. So they cut in two places, at $x = \pm\frac{1}{\sqrt{2}}$.



We have to integrate from $-\frac{1}{\sqrt{2}}$ to $\frac{1}{\sqrt{2}}$ and then from $\frac{1}{\sqrt{2}}$

to 1. By symmetry the integral from $-\frac{1}{\sqrt{2}}$ to $\frac{1}{\sqrt{2}}$ is double that from 0 to $\frac{1}{\sqrt{2}}$.

The area is therefore:

$$2\int_0^{1/\sqrt{2}}[(1-x^2)-x^2]\,dx+\int_{1/\sqrt{2}}^{1}[x^2-(1-x^2)]\,dx$$

$$=2\int_0^{1/\sqrt{2}}[1-2x^2]\,dx+\int_{1/\sqrt{2}}^{1}[2x^2-1]\,dx$$

$$=2\left[x-\frac{2}{3}x^3\right]_0^{1/\sqrt{2}}+\left[\frac{2}{3}x^3-x\right]_{1/\sqrt{2}}^{1}$$

$$=2\left(\frac{1}{\sqrt{2}}-\frac{1}{3\sqrt{2}}\right)+\left(\frac{2}{3}-1\right)-\left(\frac{1}{3\sqrt{2}}-\frac{1}{\sqrt{2}}\right)$$

$$=\left(\frac{1}{3\sqrt{2}}\right)(6-2-1+3)-\frac{1}{3}$$

$$=\frac{6}{3\sqrt{2}}-\frac{1}{3}=\sqrt{2}-\frac{1}{3}.$$

§2.6. Volumes of Revolution

Suppose that a portion of a curve lies above the x-axis and the area below the curve is rotated about the x-axis. We shall find the volume of the solid, called the **volume of revolution**.

Theorem 4: Suppose that $f(x) \geq 0$ for $x \in [a, b]$. Suppose that the area between $y = f(x)$, the x-axis and the lines $x = a$ and $x = b$ is rotated about the x-axis. The volume of revolution so produced is $\pi \int_a^b y^2 \, dx$.

Proof: Let $V(x)$ be the volume of revolution produced when that portion of the curve between a and x is rotated about the x-axis.

Let x be incremented to $x + \Delta x$. The volume will be incremented to $V + \Delta V$.

Now the extra volume (we are thinking of Δx as being positive here) is approximately a circular disk, with radius $y = f(x)$ and thickness Δx. The volume of this disk is $\pi y^2 \Delta x$.

Hence $\Delta V \approx \pi y^2 \Delta x$ and so $\dfrac{\Delta V}{\Delta x} \approx \pi y^2$.

As $\Delta x \to 0$, $\dfrac{\Delta V}{\Delta x} \to \dfrac{dV}{dx}$ and the approximation becomes exact.

Hence $\frac{dV}{dx} = \pi y^2$ and so $V = \pi \int_a^b y^2 \, dx$.

Example 9: Suppose the portion of the line $y = ½\, x$ between $x = 0$ and $x = 6$, is rotated about the x-axis. Find the volume of revolution.

Solution: $V = \pi \int_0^6 \frac{x^2}{4} \, dx = \frac{\pi}{4} \left[\frac{x^3}{3} \right]_0^6 = \frac{\pi}{4} \cdot \frac{6^3}{3} = 18\pi.$

The solid is a cone on its side, whose base has radius 3 and whose height is 6. By a fact we learnt at school, the volume of a cone is one third times the area of the base times the perpendicular height.

In this case it is $\frac{1}{3} \times 9\pi \times 6 = 18\pi$.

Now probably you never saw a proof of the volume of a cone in high school. Well now you can construct one, by generalising the above example.

Example 10: Find the volume generated when the area below the portion of the curve $y = \sin x$ between $x = 0$ and $x = \pi$ is rotated about the x-axis.

Solution: $V = \pi \int_0^{\pi} \sin^2 x \, dx$

Now $\cos 2x = 1 - 2\sin^2 x$ so $\sin^2 x = \dfrac{1 - \cos 2x}{2}$.

$$\therefore V = \frac{\pi}{2} \int_0^{\pi} (1 - \cos 2x)\, dx$$

$$= \frac{\pi}{2}\left[x - \frac{\cos 2x}{2}\right]_0^{\pi}$$

$$= \frac{\pi}{2}\left(\pi - 1/2 - 0 + 1/2\right) = \frac{\pi^2}{2}$$

§2.7. Numerical Integration

You will know Simpson's Rule and you may even remember the Trapezium Rule that precedes it. Both are methods for estimating the area under a curve $y = f(x)$, and hence $\int_a^b f(x)\, dx$, provided $f(x) \geq 0$ for $x \in [a, b]$. Each method consists of dividing the interval $[a, b]$ into a certain number of strips and uses the ordinates at the endpoints of these strips.

In the case of the Trapezium Rule, we approximate the curve on each strip by a straight line. In the case of Simpson's Rule we use an even number of strips and on each pair of strips we approximate the curve by a parabola. Simpson's Rule is more accurate than the Trapezium Rule, for the same amount of effort, which is why the Trapezium Rule is never used.

Trapezium Rule: $\displaystyle\int_a^b \mathbf{f}(x)\,\mathbf{d}x$

$$\approx \frac{\textbf{width}}{\textbf{2}}\,\textbf{[First + Last + 2} \times \textbf{sum of others]}$$

Simpson's Rule: $\displaystyle\int_a^b \mathbf{f}(x)\,\mathbf{d}x$

$$\approx \frac{\textbf{width}}{\textbf{3}}\,\textbf{[first + last + 2(sum of other evens) + 4(sum of odds)]}$$

Here 'width' is the width of each strip = $h = \dfrac{b-a}{n}$, n is the number of strips, and the other words refer to the various ordinates (y-values), with 'first' and 'last' being self-explanatory and 'odds' and 'evens' referring to the odd-numbered and even-numbered ordinates, counting the first as y_0 (even).

Like Newton's Method, Simpson's Rule is best done in a spreadsheet, even if you are doing it by hand with the aid of a calculator. Setting the working out in table form makes for fewer errors. Like Newton's Spreadsheet, this one has 4 columns. The headings are ***x***, ***y***, ***w*** and ***wy***. The ***w***'s are the **weights**. These are 1's 2's and 4's as appropriate. The ***wy*** column contains the product of the ***w***'s and the ***y***'s. The ***y***'s are the ordinates, got by substituting the ***x***'s into the function. And the ***x***'s are evenly spaced over the interval over which we are integrating.

x	***y***	***w***	***wy***

The first thing to do is to decide how many strips you're going to use. You must use an even number of strips. The more strips you use the more work you'll have to do. But, up to a point, the more strips the more accurate will be the answer, but not always. We'll discuss the number of strips you should use later.

Having decided on the number of strips, you work out the width of each. If you're integrating from a to b and have $2n$ strips then the width is $h = \dfrac{b - a}{2n}$.

In the first row, in the ***x*** column, you put down the bottom limit of integration, a. You then add w to each x to get the next, and keep stepping out until you reach the top limit, *b*. With 2*n* strips this should give 2*n* + 1 rows. (There's always one more endpoint than strips.)

x	***y***	***w***	***wy***
a			
a + *h*			
a + 2*h*			
a + 3*h*			
…..			
b			

The next step is to substitute each of these values of *x* into the function and write down the corresponding *y*-values into the ***y***-column. The weights always follow the same pattern. The first is 1, and then they alternate 4, 2, 4, 2, … until the second last is a 4 and then the very last is a 1. The ***wy*** column is now computed by multiplying each *y*-value by the appropriate weight:

x	***y***	***w***	***wy***
A	y_0	1	y_0
a + *h*	y_1	4	$4y_1$
a + 2*h*	y_2	2	$2y_2$
a + 3*h*	y_3	4	$4y_3$
…..	…..	…..	.....
b	y_{2n}	1	y_{2n}

You then total the ***wy*** column.

x	***y***	***w***	***wy***
a	y_0	1	y_0
$a + h$	y_1	4	$4y_1$
$a + 2h$	y_2	2	$2y_2$
$a + 3h$	y_3	4	$4y_3$
…..	…..	…..	.....
b	y_{2n}	1	y_{2n}
TOTAL			TOTAL

Underneath this you write the width divided by 3. You then multiply these last two figures to get the approximation for the integral.

X	***y***	***w***	***wy***
A	y_0	1	y_0
$a + h$	y_1	4	$4y_1$
$a + 2h$	y_2	2	$2y_2$
$a + 3h$	y_3	4	$4y_3$
…..	…..	…..	.....
B	y_{2n}	1	y_{2n}
TOTAL			TOTAL
$h/3$			$h/3 \times$ TOTAL

How many strips should we choose? For a start you *must* use an even number of strips. Of course Simpson's Rule will be exact if f(x) is a quadratic, and so 2 strips will be exact. Surprisingly two strips will be exact for a cubic as well.

But usually Simpson's Rule isn't exact. The more strips you take, the more accurate the answer, at least in theory. In practice there's the phenomenon of round-off errors. Your calculator will calculate the ordinates to so many decimal places and so there are usually tiny errors for each ordinate. If you were to take an enormous number of strips, these round-off errors might very well build up and counteract the better accuracy because the parabolas are fitting better.

In most cases 6, 8 or 10 strips will be accurate enough. You may be influenced by the width of the interval. For example, integrating from $x = 3$ to $x = 5$ it would be better to use 10 strips so that $w = 0.2$. Integrating from $x = 3$ to $x = 6$ we choose just 6 strips, with $h = 0.5$.

Example 11: Use Simpson's Rule to approximate $\int_1^4 \sqrt{x}\, dx$ using 6 strips. Compare this with the exact value.

Solution: The width is $h = 0.5$

x	***y***	***w***	***wy***
1	1	1	1
1.5	1.2247	4	4.8988
2	1.4142	2	2.8284
2.5	1.5811	4	6.3244

3	1.7320	2	3.4640
3.5	1.8708	4	7.4832
4	2	1	2.0000
TOTAL			27.9988
× *h*/3		∫=	**4.6665**

Performing the exact integration we get:

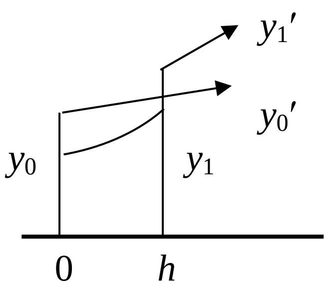
$$\int_1^4 \sqrt{x}\,\mathrm{d}x = \left[\frac{2}{3}x^{3/2}\right]_1^4 = \frac{2}{3}\,[4^{3/2} - 1^{3/2}]$$

$$= \frac{2}{3}\,[2^3 - 1] = \frac{14}{3} \approx 4.6667$$

So even with as few strips as 6 we have achieved a very good approximation by Simpson's Rule.

§2.8 The Cubic Fit Method

Why don't we fit a cubic $y = ax^3 + bx^2 + cx + d$ to each strip? This requires four pieces of information to enable us to find the four coefficients. We could take y_0 and y_1, the ordinates at the end-points, as well as y_0' and y_1', the slope at these points.



In general a cubic would be able to fit a given curve more closely than a quadratic and so give a more accurate estimate of the area. The formula for the Cubic Fit Method is:

$$\int_a^b \mathbf{f}(x)\,\mathrm{d}x \approx \underbrace{\frac{h}{2}\left[y_0 + y_n + 2(y_1 + y_2 + \ldots + y_{n-1})\right]}_{\textbf{Trapezium Rule}} - \underbrace{\frac{h^2}{12}\left[y'\right]_a^b}_{\textbf{Correction}}$$

Interestingly, the Cubic Fit Formula consists of the Trapezium Rule plus a correction factor. Yet it is not only more accurate than the Trapezium Rule, it is usually more accurate than Simpson's Rule! Even more amazing is the fact that the correction factor only depends on the slope at the two endpoints.

One advantage of the Cubic Fit Method is that you can use an odd number of strips, if that gives a simpler width. The main disadvantage of the Cubic Fit Method is that you need to be able to differentiate the function. So you wouldn't use it for functions whose derivative is particularly hard to obtain.

However what I like about it pedagogically is that it requires the student to be able to demonstrate their ability to differentiate and is not just a mindless calculation with numbers. And it is rather more accurate than Simpson's Rule with not much more effort.

Simpson's Rule is great when we don't have an algebraic expression for the function, or when the function is difficult to differentiate. But I believe that the Cubic Fit Method should be taught alongside Simpson's Rule. The fact that it was I who dreamed it up also explains why I like it so much!

Theorem 4 (Cooper): The approximation to $\int_a^b f(x)\,dx$ that is obtained by approximating the function by a series of cubics is $\frac{h}{2}[y_0 + y_n + 2(y_1 + y_2 + \ldots + y_{n-1})] - \frac{h^2}{12}[y']_a^b$ where h is the width of the strips, $y_0, y_1, \ldots, y_n$ are the successive ordinates and y′ is the derivative.

Proof: We begin by considering just one strip from $x = 0$ to $x = h$.

Suppose that the cubic $y = ax^3 + bx^2 + cx + d$ passes through $(0, y_0)$ and (h, y_1) and that its derivatives at $x = 0$ and $x = h$ are y_0' and y_1' respectively.

Then since $y' = 3ax^2 + 2bx + c$ we have:

$$y_0 = d$$
$$y_0' = c$$
$$y_1 = ah^3 + bh^2 + ch + d$$
$$y_1' = 3ah^2 + 2bh + c$$

Solving the last two equations for a, b in terms of c, d and w we get:

$3y_1 - wy_1' = bh^2 + 2ch + 3d$ which gives:

$$b = \frac{1}{h^2}[3y_1 - hy_1' - 2ch - 3d]$$

$$= \frac{1}{h^2}[3y_1 - hy_1' - 2y_0'c - 3y_0]$$

Eliminating b we get:

$2y_1 - y_1'h = - ah^3 + ch + 2d$ which gives:

$$a = -\frac{1}{h^3}(2y_1 - y_1'h - ch - 2d)$$

$$= -\frac{1}{h^3}(2y_1 - y_1'h - y_0'h - 2y_0).$$

Assembling these coefficients we have:

$$a = -\frac{1}{h^3}(2y_1 - y_1'h - y_0'h - 2y_0).$$

$$b = \frac{1}{h^2}[3y_1 - y_1'h - 2y_0'h - 3y_0]$$

$c = y_0'$

$d = y_0.$

Now the area under the cubic is $\int_0^h (ax^3 + bx^2 + cx + d)\,dx$

$$= \left[\frac{a}{4}x^4 + \frac{b}{3}x^3 + \frac{c}{2}x^2 + dx\right]_0^h$$

$$= \frac{ah^4}{4} + \frac{bh^3}{3} + \frac{ch^2}{2} + dh$$

$$= \frac{h}{12}(3ah^3 + 4bh^2 + 6ch + 12d)$$

From the above equations:

$$3ah^2 = 6y_1 - 3y_1' - 3y_0' + 6y_0$$
$$4bh^2 = 12y_1 - 4hy_1' - 8y_0' - 12y_0$$
$$6ch = 6y_0'h$$
$$12d = 12y_0$$

All this looks pretty frightening, but watch how it all simplifies.

$$\int_0^h (ax^3 + bx^2 + cx + d)\,dx = \frac{h}{12}(3ah^3 + 4bh^2 + 6ch + 12d)$$
$$= \frac{h}{12}(6y_1 + 6y_0 + y_0'h - y_1'h).$$

Suppose we repeat the above on n strips. The sum of the areas of the strips will be:

$$\frac{h}{12}(6y_0 + 6y_1 + y_0'h - y_1'h)$$
$$+ \frac{h}{12}(6y_1 + 6y_2 + y_1'h - y_2'h)$$
$$+ \frac{h}{12}(6y_2 + 6y_3 + y_2'h - y_3'h)$$
$$+ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$+ \frac{h}{12}(6y_{n-1} + 6y_n + y_{n-1}'h - y_n'h)$$

$$= \frac{h}{12} \left[6y_0 + 6y_n + 12(y_1 + y_2 + \ldots + y_{n-1}) + y_0'h - y_n'h\right]$$

$$= \frac{h}{2}\left[y_0 + y_n + 2(y_1 + y_2 + \ldots + y_{n-1})\right] - \frac{h^2}{12}\left[y_n' - y_0'\right]$$

$$= \frac{h}{2} \left[y_0 + y_n + 2(y_1 + y_2 + \ldots + y_{n-1})\right] - \frac{h^2}{12}\left[y'\right]_a^b .$$ ✋☺

Notice the way the derivatives telescope so that we only need to evaluate them at the endpoints. Also, you may recognise the first part as simply the Trapezium Rule.

Example 12: Use the Trapezium Rule, Simpson's Rule and the Cubic Fit Rule, each with with 4 strips to estimate

$$\int_1^5 \sqrt{x}\, \mathrm{d}x .$$

Solution: If $y = \sqrt{x}$ then $y' = \dfrac{1}{2\sqrt{x}}$

x	***y***	***y′***
1	1	0.5
2	1.4142	
3	1.7320	
4	2	
5	2.2361	0.2236

TRAPEZIUM RULE:

Integral = $\frac{1}{2}$ [1 + 2(1.4142 + 1.7320 + 2) + 2.2361]

= 6.7642

SIMPSON'S RULE:

Integral = $\frac{1}{3}$ [1 + 4(1.4142 + 2) + 2(1.7320) + 2.2361]

= 6.7856

Cubic Fit Correction = $\frac{1}{12}$ [0.5 – 0.2236] = 0.0230

CUBIC FIT ESTIMATE:

Integral = 6.7642 – 0.0230 = 6.7872

EXACT VALUE: Integral = 6.7869

The percentage errors, for similar amounts of computation are:

Trapezium Rule	0.3%
Simpson's Rule	0.02%
Cubic Fit Method	0.004%

Example 13: Use the Cubic Fit Method with 5 strips to estimate $\int_{2}^{7} \log x \, dx$.

Solution: If $y = \log x$ then $y' = \frac{1}{x}$.

x	y	y'
2	0.6931	0.5
3	1.0986	
4	1.3863	
5	1.6094	
6	1.7918	
7	1.9459	0.1429

TRAPEZIUM RULE:

$\frac{1}{2}$[0.6931 + 2(1.0986 + 1.3863 + 1.6094 + 1.7918)

+ 1.9459] = 7.2056

Cubic Fit Correction: $\frac{1}{12}$[0.5 – 0.1429] = 0.0298

CUBIC FIT ESTIMATE: 7.2056 – 0.0298 = 7.2354

The following table gives some advice on which rule to use:

RULE	**accur-acy**	**needs**	**# strips**	**Use when**
Trapezium	least	table of values	any number	never
Simpson	good	table of values	any even number	you only have a table of values
Cubic Fit	better	A function you can differentiate	any number	you have a formula
Integration	exact	a function you can integrate	1	you can integrate the function

EXERCISES FOR CHAPTER 2

Exercise 1: Calculate the Riemann Integral $\int_0^1 e^x \, dx$ directly from the definition, and without using the Fundamental Theorem of Calculus.

Exercise 2:
Find the area between the parabolas $y = x^2$, $y = 1 - x^2$ and the y-axis. Give your answer as an exact value, in terms of $\sqrt{2}$, as well as an approximation to 4 decimal places.

Exercise 3: Find the area enclosed between the curves

$$y = x^2 - 2x \text{ and } x^3 - 2x^2.$$

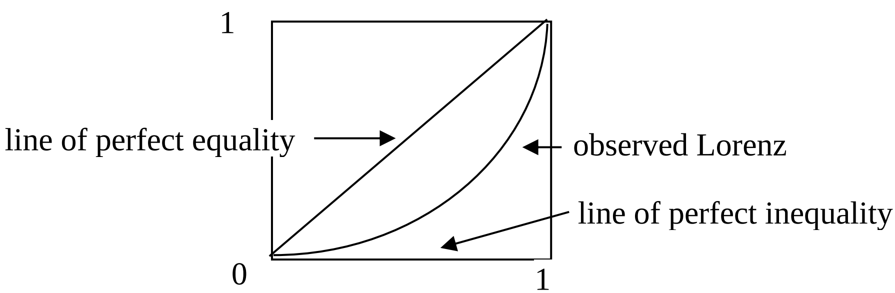
Exercise 4: Find the area enclosed between the curves $y = e^x$ and $y = 6x - x^2 - 20$ between $x = 0$ and $x = 3$. Give your answer as an exact value involving e^3 as well as an approximate numerical value to 4 decimal places.

Exercise 5: Find the area enclosed between the curve $y = \dfrac{6}{x}$ and the line $y = 5 - x$. Give the answer to 4 decimal places without using your calculator, but instead using the fact that, to 4 decimal places, log 1.5 = 0.4055.

Exercise 6: Use integration to find the area between the lines $y = 1 + 2x$ and $y = 1 - x$ between $x = -5$ and $x = 1$. Check your answer using simple geometry.

Exercise 7: Find the area enclosed between the curves $y = \sqrt{4x + 5}$, $y = \sqrt{5x + 4}$ and the x-axis.

Exercise 8: In Economics, the *Lorenz Curve* is a graph that's used in economics that shows the distribution of income. If the bottom $100x$% of households represents the bottom $100y$% of income, the point (x, y) lies on the Lorenz curve. The line y = x corresponds to perfect equality, but in practice the curve lies below it. The most extreme case, (perfect inequality) would be where one household earns all of the income and the rest earn nothing. This corresponds to the line $y = 0$ (with a spike at $x = 1$). *The Gini Coefficient* is the ratio of the area between the line of perfect equality and the observed Lorenz curve, and the area between the line of perfect equality and the line pf perfect inequality. It is a measure of inequality. The higher the Gini coefficient, the more unequal is the distribution of income. Find the Gini coefficient if the Lorenz curve is $y = x^2$.



For Exercises 9 – 16, use The Trapezium Rule, Simpson's Rule and the Cubic Fit Method, each with 8 strips, to approximate the following definite integrals. Work to 4 decimal places. Then work out the percentage error with each of these methods. The value obtained from integrating is given.

Exercise 9: $\int_{0}^{8} x^2 \, dx = 170.6667$

Exercise 10: $\int_{1}^{9} \sqrt{x} \, dx = 17.3333$

Exercise 11: $\int_{1}^{5} \frac{dx}{x} = 1.6094$

Exercise 12: $\int_{2}^{4} e^{x^2} \, dx = 1.14938 \times 10^6$

Exercise 13: $\int_{10}^{30} \log x \, dx = 59.0101$

Exercise 14: $\int_0^2 2^x \, dx = 4.3281$ [**HINT:** $2^x = e^{(\log 2)x}$]

Exercise 15: $\int_0^1 \frac{1}{1 + x^2} \, dx = 0.7854$

Exercise 16: $\int_1^{17} x^3 - 8\sqrt{x} + 2 \, dx = 20511.5051$

In exercises 17 – 20 use the 'best' method to approximate the given definite integrals. Comment on the accuracy of your answers.

Exercise 17: $\int_1^8 x^3 - x^2 + 7 \, dx$.

Exercise 18: $\int_0^{12} \sqrt{x^3 + 1} \, dx$.

Exercise 19: $\int_1^4 \sqrt{x^2 - x} \, dx$.

Exercise 20: $\int_0^6 f(x)\,dx$ where f(x) fits the following table of values.

x	0	1	2	3	4	5	6
$f(x)$	2	3	5	5	4	2	1

SOLUTIONS FOR CHAPTER 2

Exercise 1: Partition the interval [0. 1] into n intervals of equal length. Then, using the notation of the definition, $m_k = e^{(k-1)/n}$ and $M_k = e^{k/n}$.

Hence the lower Riemann sum = $\displaystyle\sum_{k=1}^{n} \frac{e^{(k-1)/n}}{n}$ and the upper Riemann sum = $\displaystyle\sum_{k=1}^{n} \frac{e^{k/n}}{n}$.

Using the GP formula, the lower Riemann sum is

$$\frac{(e^{1/n})^n - 1}{n(e^{1/n} - 1)} = \frac{e - 1}{n(e^{1/n} - 1)}.$$

$$\lim_{n\to\infty}\left[\frac{e-1}{n(e^{1/n}-1)}\right] = (e-1)\lim_{n\to\infty}\left[\frac{1}{n(e^{1/n}-1)}\right].$$

Putting $x = 1/n$ this becomes $(e-1)\lim_{x\to 0}\left[\frac{x}{e^x - 1}\right]$.

Using L'Hôpital's Rule this becomes $(e-1)\lim_{x\to 0}\left[\frac{1}{e^x}\right] =$ $e - 1$.

The upper Riemann sum $= \dfrac{e^{1/n}((e^{1/n})^n - 1)}{n(e^{1/n}-1)} = \dfrac{e^{1/n}(e-1)}{n(e^{1/n}-1)}$

$$\lim_{n\to\infty}\frac{e^{1/n}(e-1)}{n(e^{1/n}-1)} = (e-1)\lim_{n\to\infty}\left[\frac{e^{1/n}}{n(e^{1/n}-1)}\right].$$

Putting $x = 1/n$ this becomes $(e-1)\lim_{x\to 0}\left[\frac{xe^x}{e^x - 1}\right]$.

Using L'Hôpital's Rule this becomes

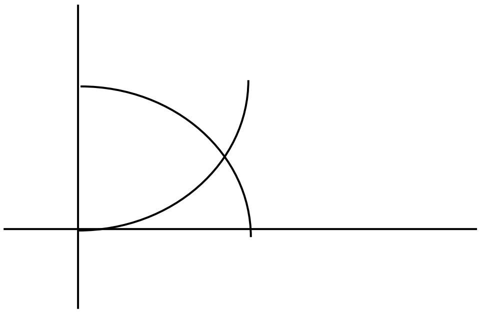
$$(e-1)\lim_{x\to 0}\left[\frac{e^x + xe^x}{e^x}\right] = e - 1.$$

Since the upper and lower Riemann sums approach the same limit that limit, namely $e - 1$, is the required integral.

Just as a check, using anti-derivatives we have

$$\int_0^1 e^x\,dx = \left[e^x\right]_0^1 = e - 1.$$

Exercise 2: The curves cut when $x^2 = 1 - x^2$,
that is, when $2x^2 = 1$ or $x = \frac{1}{\sqrt{2}}$.



Over the range from $x = 0$ to $x = \frac{1}{\sqrt{2}}$ the top y is $y = 1 - x^2$.

So the area is $\int_0^{1/\sqrt{2}} (1 - x^2) - x^2 \, dx$

$$= \int_0^{1/\sqrt{2}} 1 - 2x^2 \, dx = [x - x^2]_0^{1/\sqrt{2}} = \frac{1}{\sqrt{2}} - \frac{1}{2} \approx 0.2071.$$

Exercise 3: The curves cross if $x^2 - 2x = x^3 - 2x^2$.
Solving, we get $x^3 - 3x^2 + 2x = 0$.
Factorising we get $x(x - 1)(x - 2) = 0$, so the curves cross when $x = 0$, 1 and 2. The area enclosed is the region between the curves from $x = 0$ to $x = 2$. But since they

cross at $x = 1$ we need to split the interval [0, 2] into two pieces.

From $x = 0$ to $x = 1$ the top curve is $y = x^3 - 2x^2$. Then from $x = 1$ to $x = 2$ the top curve is $y = x^2 - 2x$. So the area is:

$$A = \int_0^1 [(x^3 - 2x^2) - (x^2 - 2x)]\,dx$$

$$+ \int_1^2 [(x^2 - 2x) - (x^3 - 2x^2)]\,dx$$

$$= \left[\frac{x^4}{4} - x^3 + x^2\right]_0^1 + \left[x^3 - \frac{x^4}{4} - x^2\right]_1^2$$

$$= \left(\frac{1}{4} - 1 + 1\right) - 0 + \left(8 - \frac{16}{4} - 4\right) - \left(1 - \frac{1}{4} - 1\right) = \frac{1}{2}.$$

Exercise 4: The curves do not cross and the top y is $y = e^x$.

So the area is $\int_0^3 [e^x - (6x - x^2 - 20)]\,dx$

$$= \int_0^3 [e^x - 6x + x^2 + 20]\,dx)$$

$$= \left[e^x - 3x^2 + \frac{1}{3}x^3 + 20x\right]_0^3$$

$= (e^3 - 27 + 9 + 60) - 1$

$= e^3 + 41 \approx 61.0855.$

Exercise 5: The curves cut when $\frac{6}{x} = 5 - x$.

Solving we get $6 = 5x - x^2$ and so $x^2 - 5x + 6 = 0$.
Hence $(x - 2)(x - 3) = 0$ and so $x = 2, 3$.
Let's see what happens between $x = 2$ and $x = 3$.
When $x = 2.5$ the value of y for the curve $y = \frac{6}{x}$ is 2.4 while for the line $y = 5 - x$ it is 2.5. So between $x = 2$ and $x = 3$ the top y is $y = 5 - x$.
The area between is therefore:

$$\int_2^3 [(5 - x) - \frac{1}{x}]\, dx = [5x - x^2 - \log x]_2^3$$

$= (15 - 9 - \log 3) - (9 - 4 - \log 2)$

$= 1 - \log 3 + \log 2$

$= 1 - \log(3/2)$

$\approx 0.5945.$

Exercise 6: The lines cut at $x = 0$. When $x = -1$, the value of y for $y = 1 + 2x$ is -1 while for $y = 1 - x$ it is 2, so between $x = -5$ and $x = 0$ the top y is $y = 1 - x$.

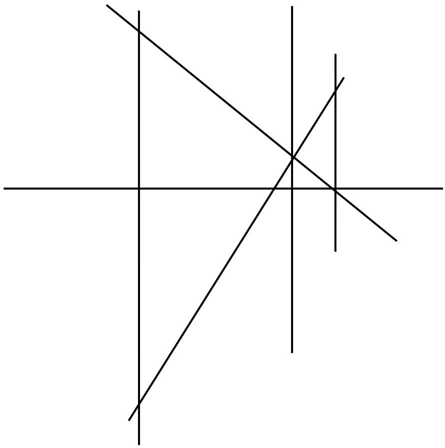
Between $x = 0$ and $x = 1$ the top y is $y = 2x + 1$.
So the enclosed area is:

$$\int_{-5}^{0}[(1 - x) - (2x + 1)]\,dx + \int_{0}^{1}[(2x + 1) - (1 - x)]\,dx$$

$$= \int_{-5}^{0}[-3x]\,dx + \int_{0}^{1}[3x]\,dx$$

$$= \left[-\frac{3}{2}x^2\right]_{-5}^{0} + \left[\frac{3}{2}x^2\right]_{0}^{1}$$

$$= (0 + \frac{3}{2}.25) + (\frac{3}{2} - 0) = \frac{75}{2} + \frac{3}{2} = \frac{78}{2} = 39.$$



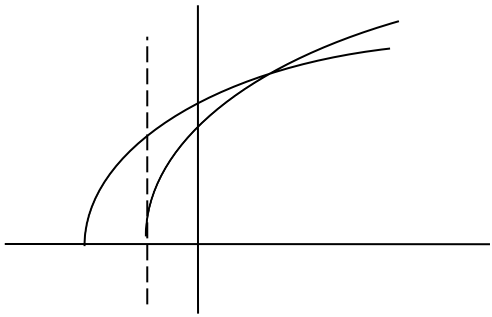
We can check this by simple geometry since the region consists of two triangles. Remember that the area of a triangle is ½ × base × perpendicular height. Here we take the base to be the vertical side.

When $x = -5$, $1 - x = 6$ while $2x + 1 = -9$. So the larger triangle has a base of 15 and a perpendicular height of 5. Its area is $\frac{15\times5}{2} = \frac{75}{2}$.

When $x = 1$, $2x + 1 = 3$ and $1 - x = 0$ so the smaller triangle has base 3 and perpendicular height of 1. Its area is $\frac{3}{2}$.

The total area is therefore $\frac{75}{2} + \frac{3}{2} = 39$.

Exercise 7: The curves cross when $4x + 5 = 5x + 4$, that is when $x = 1$.



The curve $y = \sqrt{5x + 4}$ cuts the x-axis at $x = -4/5$ while $y = \sqrt{4x + 5}$ cuts it at $x = -5/4$.

We must divide the area into two portions at $x = -4/5$. The top y between $x = -5/4$ and $x = 1$ is $\sqrt{4x + 5}$. Between $x = -5/4$ and $-4/5$ the bottom y is $y = 0$, the x-axis.

Then between $x = -4/5$ and $x = 1$ the bottom y is $y = \sqrt{5x + 4}$.

Hence the area is:

$$\int_{-5/4}^{-4/5}\sqrt{4x+5}\,dx + \int_{-4/5}^{1}[\sqrt{4x+5}-\sqrt{5x+4}]\,dx$$

$$= \left[\frac{2}{3}\cdot\frac{1}{4}(4x+5)^{3/2}\right]_{-5/4}^{-4/5}$$

$$+ \left[\frac{2}{3}\cdot\frac{1}{4}(4x+5)^{3/2} - \frac{3}{2}\cdot\frac{1}{5}(5x+4)^{3/2}\right]_{-4/5}^{1}$$

$$= \frac{1}{6}\left[\left(-\frac{16}{5}+\frac{25}{5}\right)^{3/2} - 0\right] + \frac{1}{6}\left[9^{3/2} - \left(-\frac{16}{5}+\frac{25}{5}\right)^{3/2}\right]$$

$$- \frac{3}{10}\left[9^{3/2} - 0\right]$$

$$= \frac{1}{6}\left[\left(\frac{9}{5}\right)^{3/2}\right] + \frac{1}{6}\left[9^{3/2} - \left(\frac{9}{5}\right)^{3/2}\right] - \frac{3}{10}\,9^{3/2}$$

$$= \frac{1}{6}\left[\frac{27}{5\sqrt{5}}\right] + \frac{1}{6}\left[27 - \frac{27}{5\sqrt{5}}\right] - \frac{3}{10}\,.\,27$$

$$= \frac{9}{2} - \frac{81}{10} = \frac{90-81}{20} = \frac{9}{20} = 0.45.$$

Exercise 8: The area between the line of perfect equality ($y = x$) and the observed Lorenz curve, is

$$\int_{0}^{1}(x - x^2\,dx) = \left[\frac{1}{2}x^2 - \frac{1}{3}x^3\right]_{0}^{1} = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}\,.$$

The area between the line of perfect equality ($y = x$) and the line of perfect inequality ($y = 0$) is the area of the triangle below the line $y = x$, which is clearly ½ . Hence the Gini Coefficient is $\frac{1/6}{1/2} = \frac{1}{3}$.

Exercise 9:

TRAPEZIUM

x	y
0	0
1	1
2	4
3	9
4	16
5	25
6	36
7	49
8	64
$\int$	172

SIMPSON

w	wy
1	0
4	4
2	8
4	36
2	32
4	100
2	72
4	196
1	64
$\int$	170.6667

$y' = 2x$ so Correction = $\frac{1}{12}[16 - 0] = 1.3333$

TRAPEZIUM	172.0000	0.8%
SIMPSON	170.6667	0%
CUBIC FIT	170.6667	0%
EXACT VALUE	170.6667	

Exercise 10:

TRAPEZIUM			**SIMPSON**	
x	***y***		***w***	***wy***
1	1		1	1
2	1.4142		4	5.6568
3	1.7321		2	3.4642
4	2		4	8
5	2.2361		2	4.4722
6	2.4495		4	9.798
7	2.6458		2	5.2916
8	2.8284		4	11.3136
9	3		1	3
∫	17.3061		∫	17.3321

$y' = \frac{1}{2\sqrt{x}}$ so Correction $= \frac{1}{12}[0.1667 - 0.5] = -0.0278$

TRAPEZIUM	17.3061	0.16%
SIMPSON	17.3321	0.007%
CUBIC FIT	17.3339	0.003%
EXACT VALUE	17.3333	

Exercise 11:

TRAPEZIUM

x	***y***
1	1
1.5	0.6667
2	0.5
2.5	0.4
3	0.3333
3.5	0.2857
4	0.25
4.5	0.2222
5	0.2
$\int$	1.6289

SIMPSON

w	***wy***
1	1
4	2.6668
2	1
4	1.6
2	0.6666
4	1.1428
2	0.5
4	0.8888
1	0.2
$\int$	1.6108

$$y' = -\frac{1}{x^2} \text{ so Correction} = \frac{0.25}{12}[-0.04 + 1] = 0.02$$

TRAPEZIUM	1.6289	1.2%
SIMPSON	1.6108	0.09%
CUBIC FIT	1.6089	0.03%
EXACT VALUE	1.6094	

Exercise 12:

TRAPEZIUM

x	***y***
2	54.5982
2.25	157.9850
2.5	518.0128
2.75	1924.6511
3	8103.0839
3.25	38657.6514
3.5	208981.2889
3.75	1280165.5968
4	8886110.5205
$\int$	1495397.7073

SIMPSON

w	***wy***
1	54.5982
4	631.9400
2	1036.0256
4	7698.6044
2	16206.1678
4	154630.6056
2	417962.5778
4	5120662.3872
1	8886110.5205
$\int$	1217082.7856

$y' = 2xe^{x^2}$ so Correction

$$= \frac{0.25^2}{12}[71088884.1641 - 218.3926] = 370253.4676$$

TRAPEZIUM	1.49540×10^6	30%
SIMPSON	1.21708×10^6	6%
CUBIC FIT	1.12514×10^6	2%
EXACT VALUE	1.14938×10^6	

Exercise 13:

TRAPEZIUM	
x	***y***
10	2.3026
12.5	2.5257
15	2.7080
17.5	2.8622
20	2.9957
22.5	3.1135
25	3.2189
27.5	3.3142
30	3.4012
$\int$	58.9753

SIMPSON	
w	***wy***
1	2.3026
4	10.1028
2	5.4160
4	11.4488
2	5.9914
4	12.4540
2	6.4378
4	13.2568
1	3.4012
$\int$	59.0095

$$y' = \frac{1}{x} \text{ so Correction} = \frac{2.5^2}{12}[0.0333 - 0.1] = -0.0347$$

TRAPEZIUM	58.9753	0.06%
SIMPSON	59.0095	0.001%
CUBIC FIT	59.0100	0.0002%
EXACT VALUE	59.0101	

Exercise 14:

TRAPEZIUM

x	y
0	1
0.25	1.1892
0.5	1.4142
0.75	1.6818
1	2
1.25	2.3784
1.5	2.8284
1.75	3.3636
2	4
$\int$	4.3389

SIMPSON

w	wy
1	1
4	4.7568
2	2.8284
4	6.7272
2	4
4	9.5136
2	5.6568
4	13.4544
1	4
$\int$	4.3281

$y' = \log 2.2^x$ so

Correction = $\frac{0.25^2}{12}$ [2.7726 – 0.6931] = 0.0108

TRAPEZIUM	4.3389	0.12%
SIMPSON	4.3281	0%
CUBIC FIT	4.3281	0%
EXACT VALUE	4.3281	

Exercise 15:

TRAPEZIUM		SIMPSON	
x	***y***	***w***	***wy***
0	1	1	1
0.125	0.9846	4	3.9384
0.25	0.9412	2	1.8824
0.375	0.8767	4	3.5068
0.5	0.8	2	1.6
0.625	0.7191	4	2.8764
0.75	0.64	2	1.28
0.875	0.5664	4	2.2656
1	0.5	1	0.5
$\int$	0.7847	$\int$	0.7854

$y' = -\dfrac{2x}{(1+x^2)^2}$ so

Correction = $\dfrac{0.125^2}{12}[-0.5+0] = -0.0007$

TRAPEZIUM	0.7847	0.09%
SIMPSON	0.7854	0%
CUBIC FIT	0.7854	0%
EXACT VALUE	0.7854	

Exercise 16:

TRAPEZIUM			SIMPSON	
x	***y***		***w***	***wy***
1	–5		1	–5
3	15.1436		4	60.5744
5	109.1115		2	218.2230
7	323.8340		4	1295.336
9	707		2	1414
11	1306.4670		4	5225.8680
13	2170.1556		2	4340.3112
15	3346.0161		4	13384.0644
17	4882.0151		1	4882.0151
∫	20832.4707		∫	20543.5947

$y' = 3x^2 - \dfrac{4}{\sqrt{x}}$ so

Correction = $\dfrac{2^2}{12}$ [866.0299 + 1] = 289.01

TRAPEZIUM	20832.4707	1.6%
SIMPSON	20543.5947	0.16%
CUBIC FIT	20543.4607	0.16%
EXACT VALUE	20511.5051	

Exercise 17: Here it’s not difficult to find the indefinite integral: $\dfrac{x^4}{4} - \dfrac{x^3}{3} + 7x$, and evaluating it at the endpoints we

find that the integral is 909.33333 – 6.91667 = 902.4167, correct to 4 decimal places.

The Cubic Fit Method would be almost as easy to use. Clearly one strip is sufficient, seeing that we are fitting a cubic to a cubic. This will give the exact value. The Trapezium Method will give $\frac{7(455-7)}{2}$ = 1617 and the correction factor, with $y' = 3x^2 - 2x$, is $\frac{49(176-2)}{12}$ = 714.58333. The Cubic Fit Method thus gives 1617 – 714.58333 = 902.4167, correct to 4 decimal places.

Simpson's Rule would also be exact, but we would have to use 2 strips, not 1.

Exercise 18: This is not easy to integrate so we shall use a numerical method. We will use the Cubic Fit Method:

TRAPEZIUM

x	***y***
0	1
4	8.06226
8	22.64950
12	41.58125
∫	208.00954

$$y' = \frac{3x^2}{2\sqrt{x^3 + 1}} \text{ so}$$

Cubic Correction is $\frac{16}{12}$ [5.19465 – 0] = 6.92620.

Hence the Cubic Fit gives

208.00954 – 6.92620 = 201.1033 as the estimate.

But is 201 correct to the nearest integer?

Let's try it with 6 strips.

TRAPEZIUM

x	***y***
0	1
2	3
4	8.06226
6	14.73092
8	22.64950
10	31.6386
12	41.58125
∫	202.74381

The Correction this time is $\frac{6.9260}{4}$ = 1.73155 since *h* is now 2, not 4.

The Cubic Fit Method gives:

202.74381 – 1.73155 = 201.0123.

Since the improvement is only in the first decimal place it's reasonable to expect that both are correct to the nearest integer, namely 201.

Exercise 19: It is clearly too hard to find the indefinite integral, so a numerical method is called for. But before we embark on the Cubic Fit Method, consider the Correction Factor.

$y' = \dfrac{2x}{2\sqrt{x^2 - x}} = \dfrac{x}{\sqrt{x^2 - 1}}$. We will need to evaluate this at the endpoints and there is a problem at $x = 1$, the bottom end-point. This is one of the limitations of the Cubic Fit Method. If the derivative is zero at either end-point we can't use this method. We must fall back on Simpson's Rule. We'll use 6 strips to start with.

SIMPSON

x	***y***	***w***	***wy***
1	0	1	0
1.5	0.86603	4	3.46412
2	1.42421	2	2.8242
2.5	1.93649	4	7.74596
3	2.44949	2	4.89898
3.5	2.95804	4	11.83216
4	3.46410	1	3.46410
		$\int$	5.70562

Before we can be reasonably certain as to how accurate this is, we need to repeat the process with more steps to see how much the estimate changes. Let's use 12 strips.

SIMPSON

x	***y***	***w***	***wy***
1	0	1	0
1.125	0.55902	4	2.23608
1.5	0.86603	2	1.73206
1.75	1.14564	4	4.46256
2	1.41421	2	2.82842
2.25	1.67705	4	6.70820
2.5	1.93649	2	3.87298
2.75	2.19374	4	8.77496
3	2.44949	2	4.89898
3.25	2.70416	4	10.81664
3.5	2.95804	2	5.91608
3.75	3.21131	4	12.84524
4	3.46410	1	3.46410
		∫	5.73025

It would appear that 5.7 is correct to one decimal place.

Exercise 20: Here we do not have a function, only some values, so we can't find the indefinite integral. But nor can we use the Cubic Fit Method, because we have nothing to differentiate. We can only use the Trapezium Rule, or Simpson's Rule and, as we know, Simpson's

Rule is superior to the Trapezium Rule. We'll use 6 strips. (We can't use any more since we don't have the values at the intermediate points.)

SIMPSON

x	***y***	***w***	***wy***
0	2	1	2
1	3	4	12
2	5	2	10
3	5	4	20
4	4	2	8
5	2	4	8
6	1	1	1
		$\int$	20.33

It's hard to comment on the accuracy because we can't use any more strips and we don't have any more information to go by. From experience we could guess that it is accurate to the nearest integer, and possibly to one decimal place. So we might quote 20.3 as our estimate.