Mathematics Learning Centre

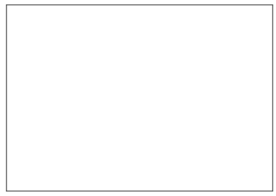
The University of Sydney

Integration: The definite integral

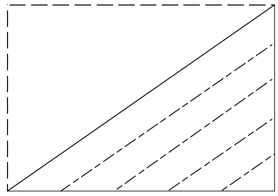
Mary Barnes

1 Finding Areas

Areas of plane (i.e. flat!) figures are fairly easy to calculate *if they are bounded by straight lines.*
The area of a rectangle is clearly the length times the breadth.
The area of a right-angled triangle can be seen to be half the area of a rectangle (see the diagram) and so is half the base times the height.

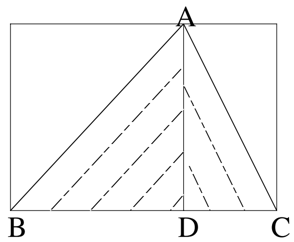


Area of rectangle
=length × breadth

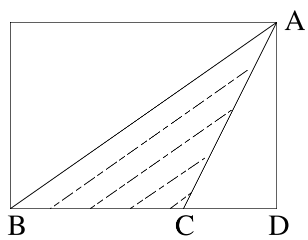


Area of triangle
$=\frac{1}{2}$ area of rectangle
$=\frac{1}{2}$ length × breadth

The areas of other triangles can be found by expressing them as the sum or the difference of the areas of right angled triangles, and from this it is clear that for any triangle this area is half the base times the height.

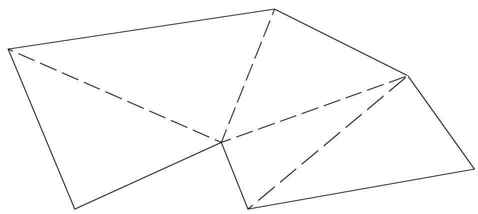


Area of △ ABC
= area of △ ABD
+ area of △ ACD



Area of △ ABC
= area of △ ABD
− area of △ ACD

Using this, we can find the area of any figure bounded by straight lines, by dividing it up into triangles (as shown).



Areas *bounded by curved lines* are a much more difficult problem, however. In fact, although we all feel we know intuitively what we mean by the area of a curvilinear figure, it is actually quite difficult to define precisely. The area of a figure is quantified by asking 'how many units of area would be needed to cover it?' We need to have some unit of area in mind (e.g. one square centimetre or one square millimetre) and imagine trying to cover the figure with little square tiles. We can also imagine cutting these tiles in halves, quarters etc. In this way a rectangle, and hence any figure bounded by straight lines, can be dealt with, but a curvilinear figure can never be covered exactly.

We are therefore forced to rely on the notion of limit in order to define areas of curvilinear figures.

To do this, we make some simple assumptions which most people will accept as intuitively obvious. These are:-

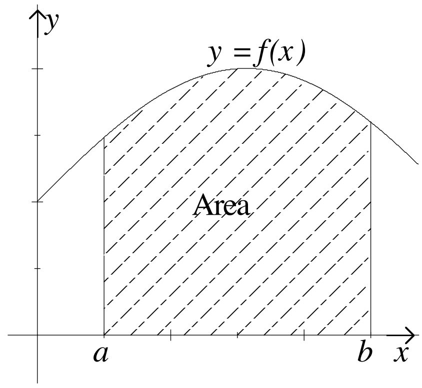
1. If one figure is a subset of a second figure, then the area of the first will be less than or equal to that of the second.

2. If a figure is divided up into non-overlapping pieces, the area of the whole will be the sum of the areas of the pieces.

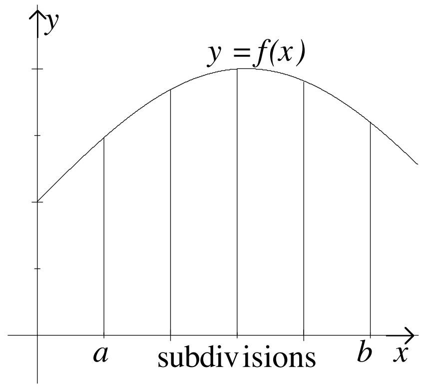
Using these assumptions, we can approximate to curved figures by means of polygons (figures with straight line boundaries), and hence *define* the area of the curved figure as the limit of the areas of the polygons as they 'approach' the curved figure (in some sense yet to be made precise).

2 Areas Under Curves

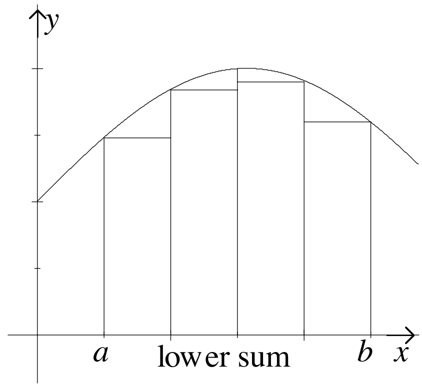
Let us suppose that we are given a positive function $f(x)$ and we want to find the area enclosed between the curve $y = f(x)$, the x-axis and the lines $x = a$ and $x = b$. (The shaded area in the diagram.)
If the graph of $y = f(x)$ is not a straight line we do not, at the moment, know how to calculate the area precisely.



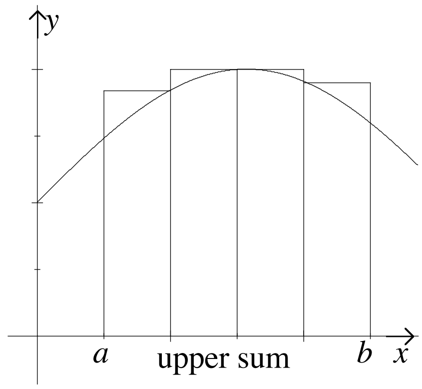
We can, however, approximate to the area as foIlows: **First** we divide the area up into strips as shown, by dividing the interval from a to b into equal subintervals, and drawing vertical lines at these points.



Next we choose the *least* value of $f(x)$ in each subinterval and construct a rectangle with that as its height (as in the diagram). The sum of the areas of these rectangles is clearly *less* than the area we are trying to find. This sum is called a *lower sum.*



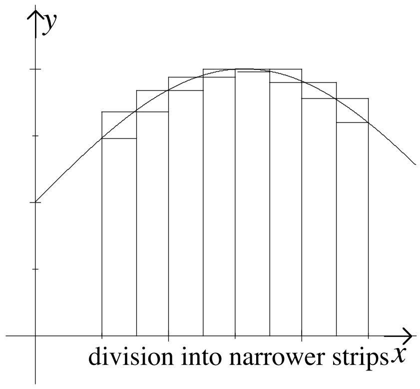
Then we choose the greatest value of $f(x)$ in each subinterval and construct a rectangle with *that* as its height (as in the diagram opposite). The sum of the areas of these rectangles is clearly *greater* than the area we are trying to find. This sum is called an *upper sum.*



Thus we have 'sandwiched' the area we want to find in between an upper sum and a lower sum. Both the upper sum and the lower sum are easily calculated because they are sums of areas of rectangles.

Although we still can't say precisely what the area under the curve is, we know between what limits it lies.

If we now increase the number of strips the area is divided into, we will get new upper and lower sums, which will be closer to one another in size and so closer to the area which we are trying to find. In fact, the larger the number of strips we take, the smaller will be the difference between the upper and lower sums, and so the better approximation either sum will be to the area under the curve.



It can be shown that if $f(x)$ is a 'nice' function (for example, a continuous function) the difference between the upper and lower sums *approaches zero* as the number of strips the area is subdivided into approaches infinity.

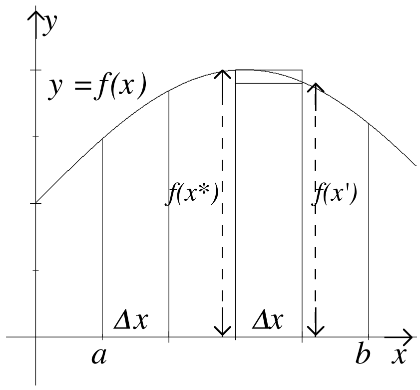
We can thus define the area under the curve to be:

> the limit of either the upper sum or the lower sum, as the number of subdivisions tends to infinity (and the width of each subdivision tends to zero).

Thus finding the area under a curve boils down to finding the *limit of a sum*.

Now let us introduce some notation so that we can talk more precisely about these concepts.

Let us suppose that the interval $[a, b]$ is divided into n equal subintervals each of width Δx. Suppose also that the greatest value of $f(x)$ in the ith subinterval is $f(x_i^*)$ and the least value is $f(x_i^{'})$.
Then the upper sum can be written as:



$$f(x_1^*)\Delta x + f(x_2^*)\Delta x + ... + f(x_n^*)\Delta x$$

or, using summation notation: $\sum_{i=1}^{n} f(x_i^*)\Delta x$.
Similarly, the lower sum can be written as:

$$f(x_1^{'})\Delta x + f(x_2^{'})\Delta x + ... + f(x_n^{'})\Delta x$$

or, using summation notation: $\sum_{i=1}^{n} f(x_i^{'})\Delta x$.

Note about summation notation

The symbol $\sum$ (pronounced 'sigma') is the capital letter S in the Greek alphabet, and stands for 'sum'.

The expression $\sum_{i=1}^{4} f(i)$ is read 'the sum of $f(i)$ from $i = 1$ to $i = 4$', or 'sigma from $i = 1$ to 4 of $f(i)$'.

In other words, we substitute 1, 2, 3 and 4 in turn for i and add the resulting expressions.

Thus,	$\sum_{i=1}^{4} x_i$	stands for	$x_1 + x_2 + x_3 + x_4$,
	$\sum_{i=1}^{5} i^2$	stands for	$1^2 + 2^2 + 3^2 + 4^2 + 5^2$,
and	$\sum_{i=1}^{2} f(x_1)\Delta x$	stands for	$f(x_1)\Delta x + f(x_2)\Delta x$.

With this notation, and letting A stand for the area under the curve $y = f(x)$ from $x = a$ to $x = b$, we can express our earlier conclusions in symbolic form.

The area lies between the lower sum and the upper sum and can be written as follows:

$$\sum_{i=1}^{n} f(x_i^{'})\Delta x \le A \le \sum_{i=1}^{n} f(x_i^{*})\Delta x.$$

The area is equal to the limit of the lower sum or the upper sum as the number of subdivisions tends to infinity and can be written as follows:

$$A = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^{'})\Delta x$$

or

$$A = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^{*})\Delta x.$$

What is the point of all this?

Well, **firstly** it enables us to define precisely what up till now has only been an imprecise intuitive concept, namely, the area of a region with curved lines forming part of its boundary.

Secondly it indicates how we may calculate approximations to such an area. By taking a fairly large value of n and finding upper or lower sums we get an approximate value for the area. The difference between the upper and lower sums tells us how accurate this approximation is. This, unfortunately, is not a very good or very practical way of approximating to the area under a curve. If you do a course in Numerical Methods you will learn much better ways, such as the Trapezoidal Rule and Simpson's Rule.

Thirdly it enables us to calculate areas *precisely*, provided we know how to find finite sums and evaluate limits. This however can be difficult and tedious, so we need to look for better ways of finding areas. This will be done in Section 4.

At this stage, many books ask students to do exercises calculating upper and lower sums and using these to estimate areas. Frequently students are also asked to find the limits of these sums as the number of subdivisions approaches infinity, and so find exact areas. We shall not ask you to do this, as it involves a great deal of computation.

3 The Definition of the Definite Integral

The discussion in the previous section led to an expression of the form

$$A = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i)\Delta x \qquad (1)$$

where the interval $[a, b]$ has been divided up into n equal subintervals each of width Δx and where x_i is a point in the ith subinterval. This is a very clumsy expression, and mathematicians have developed a simpler notation for such expressions. We denote them by

$$\int_a^b f(x)dx$$

which is read as 'the integral from a to b of $f(x)dx$'.

The $\int$ sign is an elongated 's' and stands for 'sum', just as the $\sum$ did previously. The difference is that in this case it means 'the limit of a sum' rather than a finite sum. The dx comes from the Δx as we pass to the limit, just as happened in the definition of $\frac{dy}{dx}$.

Thus the definite integral is defined as *the limit of a particular type of sum* i.e. sums like that given in (1) above, as the width of each subinterval approaches zero and the number of subintervals approaches infinity.

Notes

1. Although we used the area under a curve as the motivation for making this definition, the definite integral is **not** defined to be the area under a curve but simply the limit of the sum (1).

2. Initially, when discussing areas under curves, we introduced the restriction that $f(x)$ had to be a positive function. This restriction is not necessary for the definition of a definite integral.

3. The definition can be made more general, by removing the requirement that all the subintervals have to be of equal widths, but we shall not bother with such generalisations here.

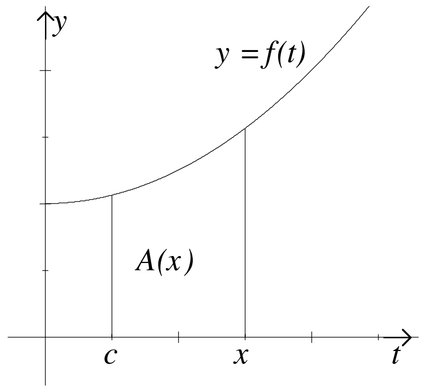
4. Sums such as (1) are called *Riemann sums* after the mathematician Georg Riemann who first gave a rigorous definition of the definite integral.

5. The definition of a definite integral requires that $f(x)$ should be *defined* everywhere in the interval $[a, b]$ and that the limit of the Riemann sums should exist. This will always be the case if f is a continuous function.

4 The Fundamental Theorem of the Calculus

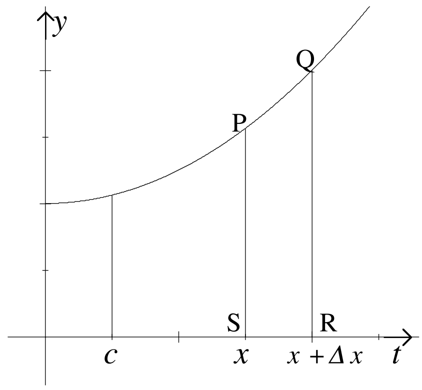
So far, we have defined definite integrals but have not given any practical way of calculating them. Nor have we shown any connection between definite integrals and differentiation.

Let us consider the special case where $f(t)$ is a continuous *positive* function, and let us consider the area under the curve $y = f(t)$ from some fixed point $t = c$ up to the variable point $t = x$. For different values of x we will get different areas. This means that the area is a function of x. Let us denote the area by $A(x)$. Clearly, $A(x)$ increases as x increases. Let us try to find the *rate* at which it increases, that is, the derivative of $A(x)$ with respect to x.



At this point, recall how we find derivatives from first principles:
Given a function $f(x)$, we let x change by an amount Δx, so that $f(x)$ changes to $f(x+\Delta x)$. The derivative of $f(x)$ is the limit of

$$\frac{f(x+\Delta x) - f(x)}{\Delta x} \text{ as } \Delta x \to 0.$$



We shall go through this process with $A(x)$ in place of $f(x)$.

When we increase x by Δx, $A(x)$ increases by the area of the figure PQRS. That is, (see the diagram)

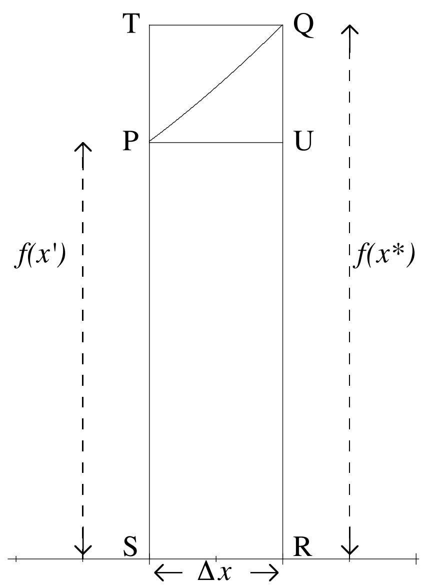
$$A(x + \Delta x) - A(x) = \text{area PQRS}.$$

Now that the area PQRS is bounded by a curved line at the top, but it can be seen to lie in between the areas of two rectangles:

$$\text{area PURS} < \text{area PQRS} < \text{area TQRS}.$$

Both of these rectangles have width Δx. Let the height of the larger rectangle be $f(x^*)$ and the height of the smaller rectangle $f(x')$. (In other words, x^* and x' are the values of x at which $f(x)$ attains its maximum and minimum values in the interval from x to $x + \Delta x$.)

Thus area PURS $= f(x')\Delta x$ and area TQRS $= f(x^*)\Delta x$.
So, $f(x')\Delta x \le A(x + \Delta x) - A(x) \le f(x^*)\Delta x$.



Now if we divide these inequalities all through by Δx, we obtain

$$f(x') \le \frac{A(x+\Delta x) - A(x)}{\Delta x} \le f(x^*).$$

Finally, if we let $\Delta x \to 0$, both $f(x')$ and $f(x^*)$ approach $f(x)$, and so the expression in the middle must also approach $f(x)$, that is, the derivative of $A(x)$, $\frac{dA}{dx} = f(x)$.

This result provides the link we need between differentiation and the definite integral.

If we recall that the area under the curve $y = f(t)$ from $t = a$ to $t = x$ is equal to $\int_a^x f(t)dt$, the result we have just proved can be stated as follows:

$$\frac{d}{dx}\int_a^x f(t)dt = f(x). \qquad (2)$$

This is the **Fundamental Theorem of the Calculus**.

In words

> If we differentiate a definite integral with respect to the upper limit of integration, the result is the function we started with.

You may not actually use this result very often, but it is important because we can derive from it the rule for calculating definite integrals:

Let us suppose that $F(x)$ is an anti-derivative of $f(x)$. That is, it is a function whose derivative is $f(x)$. If we anti-differentiate both sides of the equation (2) we obtain

$$\int_a^x f(t)dt = F(x) + c.$$

Now we can find the value of c by substituting $x = a$ in this expression.

Since $\displaystyle\int_a^a f(t)dt$ is clearly equal to zero, we obtain

$$0 = F(a) + c, \qquad \text{and so} \qquad c = -F(a).$$

Thus $\displaystyle\int_a^x f(t)dt = F(x) - F(a)$, or, letting $x = b$,

$$\int_a^b f(t)dt = F(b) - F(a).$$

This tells us how to evaluate a definite integral

- first, find an anti-derivate of the function
- then, substitute the upper and lower limits of integration into the result and subtract.

Note A convenient short-hand notation for $F(b) - F(a)$ is $[F(x)]_a^b$.

To see how this works in practice, let us look at a few examples:

i. Find $\int_0^1 x^2 dx$.

An anti-derivative of x^2 is $\frac{1}{3}x^3$, so we write

$$\int_0^1 x^2 dx = \left[\frac{1}{3}x^3\right]_0^1 = \frac{1}{3}(1)^3 - \frac{1}{3}(0)^3 = \frac{1}{3}.$$

ii. Find $\int_0^\pi \sin t dt$.

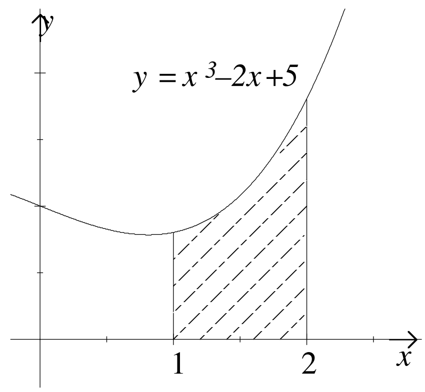
$$\begin{aligned}\int_0^\pi \sin t dt &= [-\cos t]_0^\pi \\ &= -\cos(\pi) + \cos 0 \\ &= -(-1) + 1 \\ &= 2.\end{aligned}$$

iii. Find the area enclosed between the x-axis, the curve $y = x^3 - 2x + 5$ and the ordinates $x = 1$ and $x = 2$.

In a question like this it is always a good idea to draw a rough sketch of the graph of the function and the area you are asked to find. (See below)

If the required area is A square units, then

$$\begin{aligned}A &= \int_1^2 \left(x^3 - 2x + 5\right) dx \\ &= \left[\frac{x^4}{4} - x^2 + 5x\right]_1^2 \\ &= (4 - 4 + 10) - \left(\frac{1}{4} - 1 + 5\right) \\ &= 5\frac{3}{4}.\end{aligned}$$



Exercises 4

1. **a.** $\left[2x^3\right]_2^4$

b. $\left[\frac{1}{x^2}\right]_1^3$

c. $\left[\sqrt{x}\right]_9^{16}$

d. $[\ln x]_2^4$

2. **a.** $\int_4^9 \frac{1}{\sqrt{x}} dx$

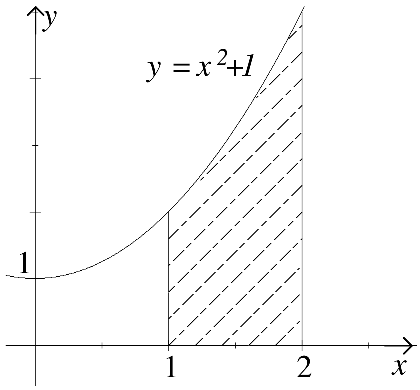
b. $\int_{\frac{-\pi}{2}}^{\frac{\pi}{2}} \cos t dt$

c. $\displaystyle\int_1^2 \frac{1}{y^2}dy$

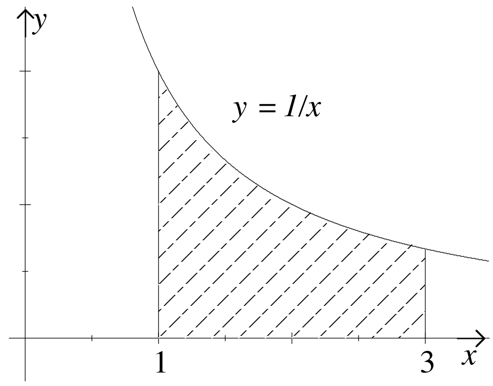
d. $\displaystyle\int_{-2}^{-1}(s^2+2s+2)ds$

3. Find the area of the shaded region in each of the diagrams below:

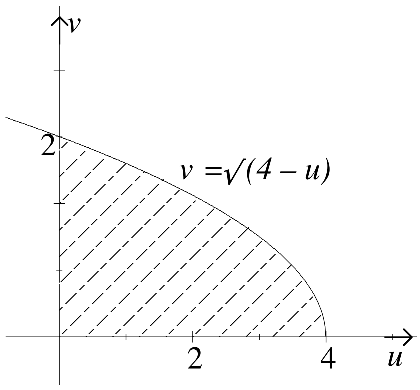
a.



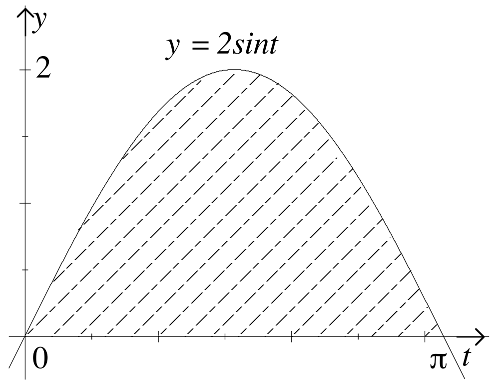
b.



c.



d.



4. Evaluate

a. $\displaystyle\int_0^1 xe^{x^2}dx$

b. $\displaystyle\int_{-2}^{-1}\frac{1}{3-x}dx$

c. $\displaystyle\int_0^{\frac{\pi}{2}}\sin 2y dy$

d. $\displaystyle\int_1^5 \frac{t}{4+t^2}dt$

5 Solutions to exercises

Exercises 4

1. **a.** $2(4^3) - 2(2^3) = 112$

b. $\frac{1}{9} - \frac{1}{1} = -\frac{8}{9}$

c. $\sqrt{16} - \sqrt{9} = 1$

d. $\ln 4 - \ln 2 = \ln \frac{4}{2} = \ln 2$

2. **a.** $\int_4^9 x^{-\frac{1}{2}} dx = \left[2x^{\frac{1}{2}}\right]_4^9 = 2\sqrt{9} - 2\sqrt{4} = 2$

b. $\int_{\frac{-\pi}{2}}^{\frac{\pi}{2}} \cos t dt = [\sin t]_{\frac{-\pi}{2}}^{\frac{\pi}{2}} = \sin \frac{\pi}{2} - \sin\left(-\frac{\pi}{2}\right) = 1 - (-1) = 2$

c. $\int_1^2 y^{-2} dy = \left[-y^{-1}\right]_1^2 = -\frac{1}{2} - \left(-\frac{1}{1}\right) = \frac{1}{2}$

d. $\int_{-2}^{-1} (s^2 + 2s + 2) ds = \left[\frac{1}{3}s^3 + s^2 + 2s\right]_{-2}^{-1} = \left(-\frac{1}{3} + 1 - 2\right) - \left(-\frac{8}{3} + 4 - 4\right) = 1\frac{1}{3}$

3. **a.** Area $= \int_1^2 (x^2 + 1) dx = \left[\frac{1}{3}x^3 + x\right]_1^2 = \left(\frac{8}{3} + 2\right) - \left(\frac{1}{3} + 1\right) = 3\frac{1}{3}$

b. Area $= \int_1^3 \frac{1}{x} dx = [\ln x]_1^3 = \ln 3 - \ln 1 = \ln 3$

c. Area $= \int_0^4 \sqrt{(4-u)} du = -\int_0^4 (4-u)^{\frac{1}{2}}(-1) du = -\left[\frac{2}{3}(4-u)^{\frac{3}{2}}\right]_0^4 = 5\frac{1}{3}$

d. Area $= \int_0^\pi 2 \sin t dt = [-2\cos t]_0^\pi = -2\cos\pi + 2\cos 0 = 4$

4. **a.** $\int_0^1 xe^{x^2} dx = \frac{1}{2}\int_0^1 e^{x^2} \cdot 2x dx = \frac{1}{2}\left[e^{x^2}\right]_0^1 = \frac{1}{2}(e-1)$

b. $\int_{-2}^{-1} \frac{1}{3-x} dx = -\int_{-2}^{-1} \frac{1}{3-x} \cdot (-1) dx = -\left[\ln(3-x)\right]_{-2}^{-1} = -(\ln 4 - \ln 5) = \ln 5 - \ln 4 = \ln\frac{5}{4}$

c. $\int_0^{\frac{\pi}{2}} \sin 2y dy = \frac{1}{2}\int_0^{\frac{\pi}{2}} \sin 2y \cdot 2 dy = \frac{1}{2}\left[-\cos 2y\right]_0^{\frac{\pi}{2}} = \frac{1}{2}(-\cos\pi + \cos 0) = 1$

d. $\int_1^5 \frac{t}{4+t^2} dt = \frac{1}{2}\int_1^5 \frac{2t}{4+t^2} dt = \frac{1}{2}\left[\ln(4+t^2)\right]_1^5 = \frac{1}{2}\ln\frac{29}{5}$