1 Linear and quadratic approximation

**Goal:** To approximate hard to compute functions by easier functions.

**Reading:** More details are given in the reading from the supplementary notes: [http://math.mit.edu/suppnotes/suppnotes01-01a/01a.pdf](http://math.mit.edu/suppnotes/suppnotes01-01a/01a.pdf)

**Vocabulary:**
- Linear approximation = linearization
- Quadratic approximation
- Geometric series
- Binomial theorem

**Basic idea:** If $h$ is small then $h^2$ is really small and $h^3$ is really, really small.

**Example 1.1.** Approximate $f(x) = 3 + 4x + 5x^2 + 7x^3$ for $x$ near 0.

The simplest approximation is the **best linear approximation**. For $x$ small we can ignore the higher powers of $x$:

$$f(x) \approx 3 + 4x.$$  

Here the wavy equal sign '$\approx$' is read as 'approximately equals'.

If we want a more accurate approximation we can use the **best quadratic approximation**:

$$f(x) \approx 3 + 4x + 5x^2.$$  

Here we kept the first two powers of $x$ and dropped the others.

To see why these are the best approximations we turn to calculus and draw some pictures. While we’re at it we’ll work near an arbitrary base point $x = a$.

### 1.1 Basic linear formulas

A1. $f(x) \approx f(a) + f'(a)(x - a)$ for $x \approx a$.

A2. $1/(1 - x) \approx 1 + x$ for $x \approx 0$.

A3. $(1 + x)^r \approx 1 + rx$ for $x \approx 0$.

A4. $\sin x \approx x$ for $x \approx 0$. 
A1 is a theoretical statement valid for all \( f(x) \). A2 - 4 are statements about specific functions. These will be some of our building blocks for more complicated functions.

To be perfectly rigorous we should say, all \( f(x) \) that have a continuous first derivative near \( x = a \). In 18.01 that will be all functions that aren’t a disaster at \( x = a \), e.g. \( 1/(x - 1) \) at \( x = 1 \).

In class we will prove A2-4 using A1. We give two proofs of A1 right now.

**Algebraic proof of A1**

The definition of the derivative tells us that if \( y = f(x) \) then

\[
f'(a) \approx \frac{\Delta y}{\Delta x}, \quad \text{where } \Delta x = x - a, \Delta y = f(x) - f(a) \text{ and } x \approx a.
\]

Rearranging this we get

\[
\Delta y \approx f'(a) \Delta x.
\]

Using \( \Delta y = f(x) - f(a) \) and \( \Delta x = x - a \) this becomes

\[
f(x) - f(a) \approx f'(a)(x - a).
\]

One more step (moving \( f(a) \) from the left-side to the right-side gives A1:

\[
f(x) \approx f(a) + f'(a)(x - a).
\]

**Geometric proof of A1** (The tangent line approximates the graph.)

The formula for the tangent line is shown in the figure above. Geometrically we see that near the point \((a, f(a))\) the blue tangent line approximates the graph. That is

\[
y = f(x) \approx y_{\text{tangent}} = f(a) + f'(a)(x - a)
\]

This is exactly the approximation formula A1.

### 1.2 Examples of linear approximation

**Example 1.2.** Find the best linear approximation of \( f(x) = (1 + x)^{99}(1 + 3x)^{77} \) for \( x \approx 0 \).

**Answer:** This example is intended to convince you that formulas A2-4 and some algebra are often easier than trying to apply A1 directly. Since linear approximation ignores the higher order terms we can replace each of the factors of \( f(x) \) by its linear approximation given by A3. (More details will be given in class.)

\[
(1+x)^{99} \approx 1+99x; \ (1+3x)^{77} \approx 1+77\cdot3x; \text{ so } f(x) \approx (1+99x)(1+77\cdot3x) = 1+330x+99\cdot231x^2 \approx 1+330x.
\]
In the last step we again dropped the higher order term $231x^2$.

**Example 1.3.** Approximate $f(x) = 1/(1 - \sin x)^2$ for $x \approx 0$.

**answer:** $\sin x \approx x$ $\Rightarrow$ $f(x) \approx 1/(1 - x)^2 \approx (1 + x)^2 = 1 + 2x + x^2 \approx 1 + 2x$.

**Example 1.4.** Approximate $f(x) = e^x$ for $x$ near 0.

**answer:** None of the formulas A2-4 helps here, so we use A1 directly: $f'(x) = e^x$, so $f(0) = 1$, $f'(0) = 1$. Thus

$$e^x \approx 1 + x.$$ 

**Example 1.5.** Approximate $f(x) = e^x (1 - x)$ near $x = 0$.

**answer:** $f(x) \approx \frac{1 + x}{1 - x} \approx (1 + x)(1 + x) \approx 1 + 2x$.

### 1.3 Basic quadratic formulas

A5. $f(x) \approx f(a) + f'(a)(x - a) + \frac{f''(a)}{2}(x - a)^2$ for $x \approx a$.

A6. $\frac{1}{1 - x} \approx 1 + x + x^2$ for $x \approx 0$.

A7. $(1 + x)^r \approx 1 + rx + \frac{r(r - 1)}{2}x^2$ for $x \approx 0$.

A8. $\sin x \approx x$ for $x \approx 0$.

A9. $\cos x \approx 1 - \frac{x^2}{2}$ for $x \approx 0$.

Similar to linear approximations, A5 is theoretical and can be used to prove the explicit formulas A6-9.

**Example 1.6.** Find the best quad. approximation to $\sqrt{1 + 3x}$ near 0.

**answer:** $\sqrt{1 + 3x} = (1 + 3x)^{1/2} = 1 + \frac{1}{2}(3x) - \frac{1}{8}(3x)^2$.

**Example 1.7.** Find the best quad. approximation to $f(\theta) = \tan \theta = \frac{\sin \theta}{\cos \theta}$ near 0.

**answer:** $f(\theta) \approx \frac{\theta}{1 - \theta^2/2} \approx \theta(1 + \theta^2/2) \approx \theta$.

More examples:

**Example 1.8.** Find the best linear approximation of $\sqrt{a + bx}$ in two ways. First by using formula (A1) and second using the basic formulas and algebra.

**answer:** (i) Give the function a name: $f(x) = \sqrt{a + bx}$ and then find the pieces of (A1).

$f(0) = \sqrt{a};$ $f'(x) = \frac{b}{2\sqrt{a + bx}} \Rightarrow f'(0) = b/2\sqrt{a}$.

Using (A1): $f(x) \approx \sqrt{a} + \frac{b}{2\sqrt{a}} x$, for $x \approx 0$.

(ii) $f(x) = \sqrt{a} \left(1 + \frac{bx}{a}\right)^{1/2} \approx \sqrt{a} \left(1 + \frac{1}{2} \frac{bx}{a}\right)$ (same as (i)).
Example 1.9. Find the best quadratic approximation for $e^x$ for $x \approx 0$.

\textbf{answer:} $f(0) = 1; \ f'(x) = e^x \Rightarrow f'(0); \ f''(x) = e^x \Rightarrow f''(0) = 1.$

\[ e^x \approx 1 + x + \frac{x^2}{2}. \]

Example 1.10. Find the quadratic approximation for $f(x) = \frac{1}{1-x}$ for $x \approx 1/2$.

\textbf{answer:} Find the pieces for (A5) (here, $a = \frac{1}{2}$).

\[ f\left(\frac{1}{2}\right) = 2; \ f'(x) = \frac{1}{(1-x)^2} \Rightarrow f'(\frac{1}{2}) = 4; \ f''(x) = \frac{2}{(1-x)^3} \Rightarrow f''\left(\frac{1}{2}\right) = 16. \]

Using (A5): $f(x) \approx 2 + 4(x - \frac{1}{2}) + 8(x - \frac{1}{2})^2.$

Example 1.11. Same problem as above, finding the answer using algebra:

\textbf{answer:} Let $y = f(x)$.

Let $u = x - \frac{1}{2}$, (so $x \approx \frac{1}{2} \iff u \approx 0$).

\[ y = \frac{1}{1/2 - u} = \frac{2}{1 - 2u} \approx 2(1 + 2u + 4u^2) = 2 + 4(x - \frac{1}{2}) + 8(x - \frac{1}{2})^2. \]

(The first approximation comes using (A4).)

Example 1.12. (Special relativity: example 3 in notes §A.)

Special relativity tells us that the mass $m$ of an object moving with respect to an inertial frame of reference is bigger than its rest mass $m_0$. The formula relating the two is

\[ m = m_0c/\sqrt{c^2 - v^2} \]

where $v$ is the speed of the mass and $c$ is the speed of light. What $v$ is needed to produce 1% increase in mass?

\textbf{answer:} An increase of 1% means we want $m/m_0 = 1.01$. Here's a case where we need to prepare by using some algebra to find the right way to express $m/m_0$ so we can use our approximation formulas.

\[ \frac{m}{m_0} = c/\sqrt{c^2 - v^2} = (1 - (v/c)^2)^{-1/2} \approx 1 + \frac{1}{2}(v/c)^2. \]

Let $u = v/c$, so we want $1.01 = 1 + \frac{1}{2}u^2 \Rightarrow .02 = u^2 \Rightarrow u \approx \frac{1}{7} \Rightarrow v \approx c/7 \approx 27000 \text{ mi/sec}.$

Example 1.13. Suppose you have $1000 in bank at 2% continuous interest. Approximately how much money is in the bank after 1 year? After 2 years?

\textbf{answer:} For this we need to know that continuous interest leads to exponential growth in your bank balance. So if we let $f(t)$ be the balance we have

\[ f(t) = 1000e^{.02t} \approx 1000(1 + .02t + (.02t)^2/2). \]

Plugging in $t = 1$ and $t = 2$ gives:

\[ f(1) \approx 1000(1 + .02 + .0002) = 1020.20 \text{ (exact: } f(1) = 1020.2013). \]
\( f(2) \approx 1000(1 + .04 + .0008) = 1040.80 \) (exact: \( f(2) = 1040.8108 \)).

**Example 1.14.** Find the best quadratic approximation of \( f(x) = \ln(1 + x) \) near \( x = 0 \).

**Answer:** \( f(0) = 0; \ f'(x) = 1/(1 + x) \Rightarrow f'(0) = 1; \ f''(x) = -1/(1 + x)^2 \Rightarrow f''(0) = -1. \)
\[
\ln(1 + x) \approx x - \frac{x^2}{2}.
\]

### 1.4 Algebraic substitution rules

1. Can substitute a linear (quadratic) approx for any factor or divisor as long the divisor has a non-zero constant term.

2. Once you make a linear substitution you can never recover the best quadratic approximation.

**Example 1.15.** (Why we need to have a constant term) In each of the following examples the denominator has no constant term. If we don’t cancel the extra factors of \( x \) in the numerator and denominator we get spurious results.

1. \( \frac{x(1+x)}{x(2+x)} \neq \frac{x}{2x} \).

2. \( \frac{\ln(1 + x)}{xe^x} \neq \frac{x}{x} = 1. \) Instead = \( \frac{\ln(1+x)/x}{e^x} \approx \frac{1-x/2}{1+x} \approx (1-x/2)(1-x) \approx 1 - 3x/2. \)
   (Note: this would be hard to do by differentiation.)

**Example 1.16.** (Why we can’t get the best quadratic approximation after a linear substitution.) Consider the function
\[
f(x) = (1 + x + x^2 + x^3)(1 + 2x + 3x^2).
\]
Multiplying this out and keeping just terms up to order 2 we get the quadratic approximation near 0:
\[
f(x) \approx (1 + x + x^2)(1 + 2x + 3x^2) \approx 1 + 3x + 0x^2
\]
If first made linear approximations of each factor we get:
\[
f(x) \approx (1 + x)(1 + 2x) = 1 + 3x + 2x^2
\]
which is not **THE BEST** quadratic approximation of \( f(x) \). What happened is that by throwing away the quadratic terms in each factor they are not included in the product the way they should be.
2 Higher order approximations, Taylor series

2.1 Higher order approximations and Taylor series

Why stop at quadratic approximations?

Going to the cubic approximation near \( a \):

\[
f(x) \approx f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \frac{f'''(a)}{3!}(x - a)^3.
\]

Here 3! is read as '3 factorial' and means 3 \( \cdot \) 2 \( \cdot \) 1. Likewise 4! = 4 \( \cdot \) 3 \( \cdot \) 2 \( \cdot \) 2 \( \cdot \) 1 etc.

By convention 0! = 1.

A fourth order approximation near \( a \) is given by

\[
f(x) \approx f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \frac{f'''(a)}{3!}(x - a)^3 + \frac{f^{(4)}(a)}{4!}(x - a)^4.
\]

In general there is the Taylor series for \( f(x) \) near \( a \) which keeps an infinite number of terms. (To emphasize the pattern we keep the 0! and the 1!).

\[
f(x) \approx f(a) + \frac{f'(a)}{1!}(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \frac{f'''(a)}{3!}(x - a)^3 + \ldots + \frac{f^{(n)}(a)}{n!}(x - a)^n + \ldots
\]

Example 2.1. (all taking \( a = 0 \)) These are important, you should learn them.

1. Exponential function

\[
e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots
\]

We can use this to approximate \( e = e^1 \approx 1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} \approx 2.71 \).

The amount the by which the approximation differs from the true value is called the approximation error. A good rule of thumb is that the error is approximately the first unused term of the Taylor series. For example, if we use a linear approximation then the error is approximately the quadratic term and if we use a 6th order approximation then the error is approximately the 7th order term.

In this example we used up to the 4th power to approximate \( e^1 \) so we estimated the error is given by the 5th order term: error \( \approx 1/5! \approx .01 \).

Using more terms of the Taylor series we find that to many decimal places \( e = 2.7182818284590451 \).

2. Geometric series

\[
1 + x + x^2 + \ldots = \frac{1}{1 - x} \text{ valid for } -1 < x < 1.
\]

The left hand side is called a geometric series with ratio \( x \). You probably learned this formula for a geometric series in high school. If we have time in class we’ll discuss the high school proof. In 18.01A we can easily see that the formula follows because the geometric series is just the Taylor series for \( \frac{1}{1 - x} \).
3. \( \sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots \) (found using the general formula for Taylor series).

4. \( \cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \ldots \) (found using the general formula for Taylor series).

5. \( \ln(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots \) (Trick to get this in a moment.)

### 2.2 Tricks for computing Taylor series

There are a number of useful tricks for using known Taylor series to compute new ones. We illustrate these tricks with examples.

1. **Algebra:** We know 
   \[ e^x = 1 + x + \frac{x^2}{2!} + \ldots \]
   so replacing \( x \) everywhere by \( x^2 \) we get 
   \[ e^{-x^2} = 1 - x^2 + \frac{x^4}{2!} - \frac{x^6}{3!} + \ldots \]

2. **Differentiation:** We know the geometric series 
   \[ \frac{1}{1-x} = 1 + x + x^2 + x^3 + \ldots \]
   By differentiating both sides of this equation we get 
   \[ \frac{1}{(1-x)^2} = \frac{d}{dx} \left( \frac{1}{1-x} \right) = 1 + 2x + 3x^2 + 4x^3 + \ldots \]

3. **Antidifferentiation (also called integration):** We know 
   \[ \frac{d \ln(1 + x)}{dx} = \frac{1}{1+x} = 1 - x + x^2 - x^3 + \ldots \]
   So we find the Taylor series for \( \ln(1 + x) \) given in example 2.1 (5) above 
   \[ \ln(1 + x) = f(0) + x - x^2 + \frac{x^3}{3} - \ldots = x - x^2 + \frac{x^3}{3} - \ldots \]

### 2.3 Mean-Value Theorem (MVT)

The mean-value theorem plays two roles in calculus:

1. It helps us make careful estimates of the size of the error in an approximation.
2. It is the key theorem underlying much of the theory of calculus. We won’t do very much theoretical work with the MVT, but you should know that it plays a big role in proving many of the main theorems.
**Statement 1 (slope form):** If \( f(x) \) is differentiable on \([a, b]\) then there is a number \( c \) with \( a < c < b \) such that \( \frac{f(b) - f(a)}{b - a} = f'(c) \).

The picture illustrates the statement of the MVT and shows why you should believe it’s true: the theorem says that there is a \( c \) between \( a \) and \( b \) such that the tangent line at \( c \) has the same slope as the secant between \((a, f(a))\) and \((b, f(b))\).

**Statement 2 (analytic form):** If \( f \) differentiable then there is a \( c \) between \( a \) and \( x \) such that \( f(x) = f(a) + f'(c)(x - a) \).

**Proof of statement 2:** Simple algebra shows this is equivalent to statement 1.

**Example 2.2.** The following examples all illustrate uses of the MVT. Again, we don’t do formal proofs in 18.01A, but I want you to see that in mathematics even a completely obvious statement like ‘if \( f'(x) > 0 \) then \( f \) is increasing’ requires a proof and in this case we get one using the MVT.

1. Show \( e^x > 1 + x \) for \( x > 0 \).
   - **answer:** Use the analytic form with \( f(x) = e^x \) and \( a = 0 \):
     Since \( c > 0 \) we know \( e^c > 1 \) \( \Rightarrow \) \( f(x) = 1 + f'(c)x = 1 + e^c x > 1 + x \).
     With more algebra can show this holds for \( x < 0 \) also:
     \( x < 0 \) \( \Rightarrow \) \( c < 0 \) \( \Rightarrow \) \( e^c < 1 \) \( \Rightarrow \) \( e^c x > x \) (less negative) \( \Rightarrow \) \( 1 + e^c x > 1 + x \).

2. Show if \( f'(x) > 0 \) on \([a, b]\) then \( f \) is increasing.
   - **answer:** Suppose \( a < x_1 < x_2 < b \). We need to show \( f(x_1) < f(x_2) \).
     MVT (with \( x_1 \) in place of \( a \)) \( \Rightarrow \) \( f(x_2) = f(x_1) + f'(c)(x_2 - x_1) \) for some \( x_1 < c < x_2 \).
     Since \( f'(c) \) and \( x_2 - x_1 \) are both positive this shows \( f(x_2) > f(x_1) \). \( \blacksquare \)

3. Show if \( f'(x) = 0 \) on \([a, b]\) then \( f \) is constant.
   - **answer:** MVT \( \Rightarrow \) \( f(x) = f(a) + f'(c)(x - a) = f(a) + 0 \cdot (x - a) = f(a) \). \( \blacksquare \)

4. Show if \( f'(x) < 0 \) \( \Rightarrow \) \( f \) decreasing.
   - **answer:** Same as example 2.2 (2)

5. Find \( c \) (as in the MVT) for \( f(x) = x^3 \) on \([0, 1]\).
   - **answer:** \( x^3 = 0 + 3c^2x \) \( \Rightarrow \) \( c = x/\sqrt{3} \).

6. Show \( \ln x \leq x - 1 \) for \( x > 0 \).
   - **answer:** Let \( a = 1 \). MVT \( \Rightarrow \) \( \ln x = 0 + \frac{1}{c}(x - 1) \).
     We examine the cases i) \( x < 1 \) and ii) \( x > 1 \) separately.
3 Indeterminate forms, L’Hospital’s rule

Warning: Pay attention to remark on p. 407 of the textbook about ’The L’Hospital Habit’.

(Warmup) Often we can find limits by simply plugging in the limiting value of x:

Example 3.1. \( \lim_{x \to 0} \frac{5x + 1}{2x + 3} = \frac{1}{3}. \)

Example 3.2. \( \lim_{x \to 0} \frac{\sin x}{2x + 3} = \frac{0}{3}. \)

When the numerator and denominator are both 0 we have what is called an indeterminant form.

Example 3.3. Indeterminate form \( \lim_{x \to 0} \frac{\sin x}{x} = \frac{0}{0}. \)

It is called indeterminant, because in the given form we don’t if the limit exists or, if it does exist, what it equals. L’Hospital’s theorem gives us a straightforward way to compute the limit.
L’Hospital’s Theorem for the indeterminant form \( \frac{0}{0} \):

If \( f(x), g(x) \) are differentiable and \( f(a) = g(a) = 0 \) then
\[
\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f'(x)}{g'(x)}.
\]

Idea: The idea for why we would expect this to be true comes from our work with approximations. We simply replace \( f(x) \) and \( g(x) \) by their linear approximation if \( f(a) = g(a) = 0 \) then
\[
\frac{f(x)}{g(x)} \approx \frac{f(a) + f'(a)(x-a)}{g(a) + g'(a)(x-a)} = \frac{f'(a)}{g'(a)}.
\]

Example 3.4. \[\lim_{x \to 0} \frac{\sin(x)}{x} = \lim_{x \to 0} \frac{\cos x}{1} = 1.\]
(This particular example turns out to be circular, since we need this limit to show the derivative of \( \sin(x) \) is \( \cos(x) \), but we will not go into the details of this.)

Example 3.5. \[\lim_{x \to 1} \frac{x-1}{3x-3} = \lim_{x \to 1} \frac{1}{3} = \frac{1}{3}\]
(We could also can use some simple algebra to see this.)

'Proof’ of l’Hospital: The proof relies on the definition of the derivative.
\[
\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{(f(x) - f(a))/(x-a)}{(g(x) - g(a))/(x-a)} = \frac{f'(a)}{g'(a)}.
\]

Note: the first equality is only true because \( f(a) = 0 \) and \( g(a) = 0 \).

Example 3.6. \[\lim_{x \to 0} \frac{\sin^2 x}{x^2} = \lim_{x \to 0} \frac{2 \sin x \cos x}{2x} = \lim_{x \to 0} \frac{2(\cos^2 x - \sin^2 x)}{2x} = 1.\]
or use algebra: \((\lim_{x \to 0} \frac{\sin x}{x})^2 = 1^2 = 1.\)

3.1 Other indeterminant forms

There are other indefinite forms:
\[
\frac{\infty}{\infty}, \quad 0 \cdot \infty, \quad \infty - \infty, \quad 0^0, \quad \infty^0, \quad 1^\infty.
\]
L’Hospital only works for \( \frac{0}{0} \) and \( \frac{\infty}{\infty} \). For the others must manipulate them into the form to
\( \frac{0}{0} \) or \( \frac{\infty}{\infty} \). We’ll try to make all this clear using examples.

Example 3.7. Find \( \lim_{x \to 0} \frac{\ln x}{1/x} \)

answer: Setting \( x = 0 \) we see this is indeterminant of the form \( \frac{-\infty}{\infty} \). Since L’Hospital’s rule applies to this case we take derivatives of the numerator and denominator. Written out completely we have:
\[
\lim_{x \to 0} \frac{\ln x}{1/x} = \lim_{x \to 0} -\frac{x}{1/x} = \lim_{x \to 0} -\frac{x}{x^2} = 0.
\]
8. Find $\lim_{x \to 0^+} x \ln x$.

**answer:** We see that $\lim_{x \to 0^+} x \ln x = 0^{-\infty}$, which is indeterminate, but not of a form suitable for L'Hospital's rule. We will use a simple algebraic manipulation to put it in a form where we can use L'Hospital.

$$\lim_{x \to 0^+} x \ln x = \lim_{x \to 0^+} \frac{\ln x}{1/x} = 0.$$  

(The limit equals 0 by the Example 3.7 just above.)

You have to be careful not to use L'Hospital when it is not valid to do so.

**Example 3.8.** (When not to use L'Hospital)

1. Find $\lim_{x \to 0} \frac{x^2}{2x^2 + 5}$.

**answer:** Don’t be fooled by the 0 in the numerator: because the denominator is not 0, this is not an indeterminate form.

$$\lim_{x \to 0} \frac{x^2}{2x^2 + 5} = \frac{0}{5} = 0.$$  

2. Find $\lim_{x \to 0} \frac{2x^2 + 5}{x^2}$.

**answer:** This has a 0 in the denominator but not the numerator, so it is not indeterminate:

$$\lim_{x \to 0} \frac{2x^2 + 5}{x^2} = \frac{5}{0}$$

This has no limit. We say the limit does not converge.

3. Find $\lim_{x \to 0} \frac{\sin x}{x^2}$.

**answer:** Applying L'Hospital can give you another indeterminate form. This might or might not have a limit:

$$\lim_{x \to 0} \frac{\sin x}{x^2} \not\equiv \lim_{x \to 0} \frac{\cos x}{2x} = \frac{1}{0}.$$  

This has no limit, i.e. it does not converge.

L'Hospital's rule also works for limits as $x \to \infty$.

**Proof:** Suppose

$$\lim_{x \to \infty} \frac{f(x)}{g(x)} = 0.$$  

Let $u = 1/x$ then

$$\lim_{x \to \infty} \frac{f(x)}{g(x)} = \lim_{u \to 0^+} \frac{f(1/u)}{g(1/u)} = \lim_{u \to 0^+} \frac{f'(1/u)(-1/u^2)}{g'(1/u)(-1/u^2)} = \lim_{x \to \infty} \frac{f'(1/u)}{g'(1/u)} = \lim_{x \to \infty} \frac{f'(x)}{g'(x)}.$$  

**Example 3.9.** Find $\lim_{x \to \infty} \frac{x}{e^x}$. 


**answer:** This is indefinite of the form \(\infty/\infty\), so we can apply L'Hospital's rule by taking derivatives on the top and bottom.

\[
\lim_{x \to \infty} \frac{x}{e^x} = \lim_{x \to \infty} \frac{1}{e^x} = 0.
\]

When \(f(x)\) and \(g(x)\) it is usually easier to use algebra than L'Hopital.

**Example 3.10.** Find \(\lim_{x \to \infty} \frac{x^3 + x^2 + 1}{(2x + 1)^3}\).

**answer:** If we use L'Hospital we'll need to apply it 3 times. An algebraic solution is to divide both numerator and denominator by \(x^3\):

\[
\lim_{x \to \infty} \frac{1 + 1/x + 1/x^3}{(2 + 1/x)^3} = \frac{1}{8}
\]

**Example 3.11.** Find \(\lim_{x \to \infty} x^{1/x}\).

**answer:** When \(x = \infty\) we get the indeterminant form \(\lim_{x \to \infty} x^{1/x} = \infty^0\).

We must do some algebraic manipulation before we can apply L'Hospital. To do this we introduce a new variable \(y\) defined by

\[y = \lim_{x \to \infty} x^{1/x}\]

Then

\[\ln y = \lim_{x \to \infty} \frac{1}{x} \ln x = 0 \Rightarrow y = 1.\]

This last limit is just our old friend from Example 7.

**Example 3.12.** \(\lim_{x \to 0} x^{1/x} = 0^\infty = 0\). Why did this not require L'Hospital’s rule.

### 3.2 Growth rates

In this subsection we will assume that \(a > 0\) and \(b > 0\) are positive constants.

All of the functions \(\ln x, x^a, e^{bx}\) grow to infinity as \(x\) gets large. But they do not grow at the same rate. Instead we have the important size relationship:

As \(x\) gets large we have

\[\ln x \ll x^a \ll e^{bx}\text{, for any } a > 0, b > 0\]

The double less than signs, \(\ll\), should be read as ‘grows much slower than’. So the size relationship says

\[\ln x\text{ grows much slower than } x^a\text{ grows much slower than } e^{bx}\text{.}\]

All we mean by this is that

\[\lim_{x \to \infty} \frac{\ln x}{x^a} = 0\text{ and } \lim_{x \to \infty} \frac{x^a}{e^{bx}} = 0\]
3.3 Dominant terms

Needs to be written.

4 Definite integrals; fundamental theorem

Definition: The definite integral is the area between the graph and the $x$-axis. This is denoted with an elongated ‘S’ as shown in the caption to the figure below.

\[
\int_{a}^{b} f(x) \, dx = \text{area under curve} = \text{the integral of } f(x) \text{ from } a \text{ to } b
\]

Examples: For simple graphs we can compute the area directly:

Example 4.1. Compute the definite integral $\int_{0}^{1} x \, dx$.

\[
\int_{0}^{1} x \, dx = \text{shaded area in the figure below} = \frac{1}{2}.
\]

Definite integral = area under $y = x$, over $x$-axis, between $x = 0$ and $x = 3$.

Dummy variables: We can use any symbol for the variable used in integration. That is

\[
\int_{0}^{1} x \, dx = \int_{0}^{1} u \, du = \int_{0}^{1} t \, dt \quad \text{or} \quad \int_{a}^{b} f(x) \, dx = \int_{a}^{b} f(u) \, du
\]

We often refer to the variable used in integration as a dummy variable because it’s just there for integration and doesn’t play a role in the rest of a problem or argument.

Area below the $x$-axis counts negative: If the curve is below the $x$-axis, i.e. the value of the function is negative, then we will count the area as negative in the integral.

Example 4.2.
4. DEFINITE INTEGRALS; FUNDAMENTAL THEOREM

Area is all below the x-axis: \( \int_{-1}^{0} x \, dx = -\frac{1}{2} \)

Area is both above and below the x-axis: \( \int_{-1}^{1} x \, dx = 0 \)

4.1 Summation notation

We will often use sums with many terms that show a pattern.

Example 4.3.
1. \( 1 + 2 + 3 + \ldots + 999 + 1000 \).
2. \( 1 + 1/2 + 1/3 + \ldots + 1/999 + 1/1000 \).
3. \( 1^2 + 2^2 + \ldots + N^2 \).

The ellipsis indicates we didn’t write down every term. Often this is okay since we can see the pattern. But, this notation is not always clear. One way to be fully specific and to be more compact is to use summation notation.

Here are the previous examples written in summation notation.

1. \( 1 + 2 + \ldots + 1000 = \sum_{n=1}^{1000} n \).
2. \( 1/1 + 1/2 + \ldots + 1/1000 = \sum_{n=1}^{1000} 1/n \).
3. \( 1 + 2^2 + \ldots + N^2 = \sum_{n=1}^{N} n^2 \).

The letter \( \sum \) is the uppercase Greek letter sigma –for summation.

In the examples, the letter \( n \) is the index and the terms above and below the \( \sum \) are the limits of the sum. The formula

\[
\sum_{n=1}^{1000} n^2
\]

is read as ‘the sum from \( n = 1 \) to 1000 of \( n^2 \).’
There is no real mystery to this notation. If it’s hard to understand a formula just write it out long-hand

**Example 4.4.** Compute $\sum_{j=2}^{5} j^2$.

**answer:** $\sum_{j=2}^{5} j^2 = 2^2 + 3^2 + 4^2 + 5^2 = 54$. Notice that to get this we just marched the index $j$ along from 2 to 5.

Once you see the pattern you can write a series using summation notation.

**Example 4.5.** Write $1 \cdot 2 + 2 \cdot 3 + 3 \cdot 4 + \ldots + 100 \cdot 101$ in summation notation.

**answer:** $\sum_{k=1}^{100} k \cdot (k + 1)$.

**Example 4.6.** Write the sum from $k = 7$ to 23 of $\sin(k\pi/100)$ in summation notation.

**answer:** $\sum_{k=7}^{23} \sin(k\pi/100)$.

### 4.2 Computing areas under curves by the method of Exhaustion

(We exhaust the area, hopefully you’ll be exhilarated.) This is one of the main points in 18.01A. We will use it many, many times to set up integrals. The idea is to approximate the region by a lot of thin rectangles and approximate the area by summing the area of all the rectangles.

In this section we will look at an explicit example using $y = x^2$. Later we will work abstractly with a general function $y = f(x)$.

**Example 4.7.** Approximate the integral $\int_{1}^{2} x^2 \, dx$ by approximating the area with 2 rectangles. Repeat this with 4 and then 8 rectangles.

**answer:** The figures below covers the region under the curve by 2, 4 and 8 rectangles respectively.

The area under $y = x^2$ covered by 2, 4 and 8 rectangles.

We go through the cases one at a time.

**2 rectangles:** The width of each rectangle is 0.5. The height of each rectangle is the height of the curve at the right-side of the rectangle, e.g. the right side of first rectangle is
at \( x = 1.5 \) and the rectangle has height \( y = (1.5)^2 \).

We approximate the area under the curve by summing the area of the rectangles:

\[
\int_{1}^{2} x^2 \, dx \approx (0.5)(1.5)^2 + (0.5)(2)^2 = 3.125.
\]

Since both rectangles stick up above the curve this is clearly an over estimate.

4 rectangles: The middle figure covers the region under the curve by 4 rectangles. The width of each rectangle is now 1/4 and the heights are again given by the height of the curve at the right-side of the rectangle.

Summing the area of the rectangles:

\[
\int_{1}^{2} x^2 \, dx \approx (0.25)(1.25)^2 + (0.25)(1.5)^2 + (0.25)(1.75)^2 + (0.25)(2)^2 = 2.7185.
\]

8 rectangles: The third figure covers the region under the curve by 8 rectangles. Each rectangle has width 1/8. (We use fractions instead of decimals simply to fit them in the picture.)

Summing the area of the rectangles:

\[
f_{1}^{2} x^2 \, dx \approx \frac{1}{8} \cdot \left( \frac{9}{8} \right)^2 + \frac{1}{8} \cdot \left( \frac{10}{8} \right)^2 + \frac{1}{8} \cdot \left( \frac{11}{8} \right)^2 + \frac{1}{8} \cdot \left( \frac{12}{8} \right)^2 + \frac{1}{8} \cdot \left( \frac{13}{8} \right)^2 + \frac{1}{8} \cdot \left( \frac{14}{8} \right)^2 + \frac{1}{8} \cdot \left( \frac{15}{8} \right)^2 + \frac{1}{8} \cdot 2^2 \\
\approx 2.523
\]

More, thinner rectangles gives a better estimate. Notice that in all of the figures the rectangles stick up over the curve and so our estimates are all overestimates. Also notice that as we use more, thinner rectangles they do a better and better job of approximating the region under the curve.

We call this approach the method of exhaustion because as we increase the number of rectangles the error in the estimate goes to 0, that is we ‘exhaust’ the error.

4.3 The method of exhaustion for arbitrary functions

Here are the steps for estimating the definite integral, \( \int_{a}^{b} f(x) \, dx \) of a function \( f(x) \). Don’t forget that the definite integral is just the area under the graph of \( f(x) \).

1. Divide \([a, b]\) into \( n \) equal intervals. So each interval has width \( \Delta x = \frac{b-a}{n} \)
2. Pick any point \( c_i \) in the \( i \)-th interval.
3. The height of the \( i \)-th rectangle is \( f(c_i) \), i.e. the height of the curve at \( x = c_i \).
4. Given the width and height we know the area of the \( i \)-th rectangle is

\[
\text{Area of } i\text{-th rectangle} = f(c_i) \Delta x
\]
5. Approximate the area under the curve by summing the area of the rectangles:

\[ \int_a^b f(x) \, dx = \text{area under curve} \approx \text{sum of area of rectangles} = \sum_{1}^{n} f(c_i) \Delta x \]

The following figure shows all the pieces of this algorithm.

We will call the sum \( \sum_{1}^{n} f(c_i) \Delta x \) a Riemann Sum.

To summarize:

- The Riemann sum approximates the integral
  \[ \int_a^b f(x) \, dx \approx \sum_{1}^{n} f(c_i) \Delta x \]

- In the limit as the number of rectangles \( n \) goes to \( \infty \) the Riemann sum goes to the value of the integral, i.e. the true area under the curve.
- As \( n \to \infty \) the width of each rectangle \( \Delta x \to 0 \).

### 4.4 How to choose \( c_i \)

The general algorithm is intentionally vague about how to choose the point \( c_i \) in each interval. This is because we can do it any way we want. We won’t give a formal proof, but this should seem plausible, since as the rectangles get thinner the height is basically the same for any choice of \( c_i \). For the record we state this as a theorem.

**Theorem.** No matter how you choose \( c_i \) in the limit as \( n \to \infty \) the Riemann sum goes to the value of the integral.

**Proof.** One proof uses the mean value theorem. You can find it in the textbook.

There are many ways to **choose the point \( c_i \) in each interval**. Typical choices are:
• Choose the left endpoint of each interval
• Choose the right endpoint if each interval
• Choose the midpoint of each interval
• Choose the point in each interval that maximizes the height of the rectangle.
• Choose the point in each interval that minimizes the height of the rectangle.
• Choose a random point in each interval.

The corresponding Riemann sums are the called respectively the left, right, mid, upper, lower and random Riemann sum.

Example 4.8. Approximate \( \int_0^1 x \, dx \) using a right Riemann sum. (Give the general formula for \( n \) rectangles.)

**answer:** The figure shows \([0,1]\) divided into \( n \) equal intervals.

The right endpoints are \( x_1 = 1/n, x_2 = 2/n, \ldots x_i = i/n, \ldots \). The height of the \( i \)-th rectangle is \( f(x_i) = i/n \). Therefore

\[
\int_0^1 x \, dx \approx \sum_{i=1}^n \left( \frac{i}{n} \right) \cdot \frac{1}{n} \quad \text{(Here } \Delta x = \frac{1}{n}, \text{ right endpt } = \frac{i}{n})
\]

\[
= \frac{1}{n^2} \sum_{i=1}^n i = \frac{1}{n^2} \cdot \frac{n(n+1)}{2} = \frac{1+1/n}{2}.
\]

In the limit as \( n \to \infty \) we see the Riemann sum becomes \( 1/2 \), which is the area under the curve.

**Note:** The textbook by Simmons computes \( \int x \, dx, \int x^2 \, dx, \int x^4 \, dx \). The computation relies on formulas for \( \sum i, \sum i^2, \sum i^4 \).

### 4.5 First Fundamental Theorem of Calculus

The definite integral is defined as an area. So far our only method of computing it is to use the rather tiring 'method of exhaustion'. Fortunately there is a much easier way to compute integrals.
Theorem. (The first fundamental theorem of calculus) If \( f(x) \) is continuous and \( F'(x) = f(x) \) then
\[
\int_a^b f(x) \, dx = F(b) - F(a) = \text{(definition)} \ F(x)|_a^b.
\]
This says that finding the area is equivalent to finding an anti-derivative. This is a BIG idea.

Notes: 1. In the box we introduced a shorthand notation, instead of \( F(b) - F(a) \) we can right \( F(x)|_a^b \). They mean the same thing.
2. When you see something called ‘The Fundamental Theorem’ you should assume it’s important. In this case, it warrants a lot of attention and we will sketch three proofs.

Example 4.9. (We can’t possibly do all these in class.)

1. Use the first fundamental theorem to compute \( \int_0^1 x^3 \, dx \)
\[
\text{answer: } \int_0^1 x^3 \, dx = x^4|_0^1 = \frac{1}{4} \text{ (draw your own picture).}
\]
2. Same question for \( \int_0^{\pi/a} \sin ax \, dx \).
\[
\text{answer: } \int_0^{\pi/a} \sin ax \, dx = -\frac{1}{a} \cos ax|_0^{\pi/a} = \frac{2}{a}. \quad (\text{Note the picture below shows us that the integral is positive –it’s easy to mess up signs.)}
\]
3. Show \( \int_0^{2\pi} \sin x \, dx = 0. \)
\[
\text{answer: } \int_0^{2\pi} \sin x \, dx = -\cos x|_0^{2\pi} = 0.
\]

4. Compute \( \int_1^2 \frac{1}{x} \, dx \).
\[
\text{answer: } \int_1^2 \frac{1}{x} \, dx = \ln x|_1^2 = \ln 2.
\]
5. Given a rod of length 2 m with density \( \delta(x) = 2 - (x-1)^2 \) g/m, find the total mass of the rod. (Be sure you understand why it’s okay to speak of density of a rod in mass/length.)
\[
\text{answer: } \text{This is a problem where we first need to ‘build’ the integral as a Riemann sum: To do this we divide the rod into } n \text{ small segments:}
\]
The figure shows a rod of length two divided into \( n \) equal segments, a point \( c_i \) is chosen in each segment. As usual we’ll call the length of each segment \( \Delta x \).

Now, the key step in building integrals is the following observation: If the \( i \)-th segment is small enough then the density of the segment is essentially constant and equal to \( \delta(c_i) \). This means that the mass of the \( i \)-th segment is approximately \( \delta(c_i) \Delta x \). Adding up the mass of each segment we get

\[
\text{mass of the rod} = M \approx \sum_{i=1}^{n} \delta(c_i) \Delta x.
\]

This is a Riemann sum! In the limit as \( n \to \infty \) the sum becomes an integral and we get the exact formula:

\[
M = \int_{0}^{2} \delta(x) \, dx = \int_{0}^{2} 2 - (x - 1)^2 \, dx = 2x - (x - 1)^3/3\big|_{0}^{2} = 4 - 2/3 = 10/3.
\]

4.6 An important convention

So far we’ve always had \( a < b \), the following will be quite useful.

\[
\int_{b}^{a} f(x) \, dx = -\int_{a}^{b} f(x) \, dx.
\]

4.7 Properties of definite integrals

1. \( \int_{a}^{b} f(x) + g(x) \, dx = \int_{a}^{b} f(x) \, dx + \int_{a}^{b} g(x) \, dx \).

2. \( \int_{a}^{c} f(x) \, dx + \int_{c}^{b} f(x) \, dx = \int_{a}^{b} f(x) \, dx \).

3. \( \int_{a}^{b} cf(x) \, dx = c \int_{a}^{b} f(x) \, dx \).

4. \( \left| \int_{a}^{b} f(x) \, dx \right| \leq \int_{a}^{b} |f(x)| \, dx \).

(All of these properties follow from the definition of integral as area.)

Example 4.10. (properties 1 and 3)

\[
\int_{1}^{2} 3x^3 + 4x \, dx = 3 \int_{1}^{2} x^3 \, dx + 4 \int_{1}^{2} x \, dx = 3x^4/4\big|_{1}^{2} + 4x^2/2\big|_{1}^{2} = 17\frac{1}{4}.
\]

Example 4.11. (property 2)
\[
\int_{-1}^{1} x^3 \, dx = \int_{-1}^{0} x^3 \, dx + \int_{0}^{1} x^3 \, dx = x^4/4|_0^1 + x^4/4|_{-1}^0 = -1/4 + 1/4 = 0.
\]

**Example 4.12.** (property 2)
\[
\int_{-1}^{1} |x| \, dx = \int_{-1}^{0} |x| \, dx + \int_{0}^{1} |x| \, dx = \int_{-1}^{0} -x \, dx + \int_{0}^{1} x \, dx = -x^2/2|_0^0 + x^2/2|_{-1}^1 = 1/2 + 1/2 = 1.
\]

**Example 4.13.** (property 4)
\[
\int_{-1}^{1} x^3 \, dx = x^4/4|_{-1}^1 = 0.
\]
\[
\int_{-1}^{1} |x^3| \, dx = \int_{-1}^{0} -x^3 \, dx + \int_{0}^{1} x^3 \, dx = 1/2 \Rightarrow \left| \int_{-1}^{1} x^3 \, dx \right| \leq \int_{-1}^{1} |x^3| \, dx.
\]

For each of these examples you should be able to draw a picture and understand the algebraic manipulations in terms of areas.

### 4.8 Proofs of the first fundamental theorem

Here are the three promised proofs of the first fundamental theorem.

**Proof 1:** (velocity and distance) We’ll prove the fundamental theorem by thinking of the function \( F(t) \) as the position of a mass moving along a line. Then \( F'(t) = f(t) \) is the velocity of the mass and \( F(b) - F(a) \) is the net displacement over the time interval \([a, b]\). (We use displacement because if the mass starts and ends at the same place its displacement \( F(b) - F(a) = 0 \).)

The fundamental theorem says that \( \int_{a}^{b} f(t) \, dt = F(b) - F(a) \). In words it says that the integral of velocity is displacement. We’ll show that this is the case by building the integral from a Riemann sum.

Divide \([a, b]\) into \( n \) equal intervals and choose a value \( c_i \) in each interval. Over a small time slice the velocity is approximately constant and we can compute the distance traveled as velocity \( \times \) time. If this is negative the displacement is to the left and if it’s positive the displacement is to the right.

Thus in the \( i \)-th interval the distance traveled is approximately \( f(c_i) \Delta t \).

\[
\begin{array}{cccccccc}
& t_0 & c_1 & t_1 & c_2 & t_2 & c_3 & t_3 & \cdots & t_{n-1} & c_n & t_n \\\n\hline
a & & & & & & & & & & & b & \\
\end{array}
\]

The net displacement is the the sum of the displacements over all the small intervals. Thus

\[
\text{Net displacement} = F(b) - F(a) \approx \sum_{i=1}^{n} f(c_i) \Delta t.
\]

The sum on the right is a Riemann sum. If we let the number of intervals \( n \) go to infinity then the Riemann sum becomes an integral and the approximation becomes exact. That is

\[
\text{Net displacement} = F(b) - F(a) = \int_{a}^{b} f(t) \, dt
\]

This is the formula we wanted to prove, so we are done!
Proof 2: (Mean Value Theorem)

\[ F(b) - F(a) = \sum_{i=1}^{n} F(t_i) - F(t_{i-1}) \]
\[ = \sum_{i=1}^{n} F'(c_i)(t_i - t_{i-1}), \text{ where } c_i \text{ is from the MVT} \]
\[ = \sum_{i=1}^{n} f(c_i) \Delta t. \]

As always, this sum \( \to \int_{a}^{b} f(t) \, dt \) as \( \Delta t \to 0 \).

We won’t mention this proof again.

Proof 3: See the textbook by Simmons §6.6

### 4.9 Sums approximate integrals and integrals approximate sums

So far we have use Riemann sums to approximate integrals. We can turn this around and use integrals to approximate sums. The reason to do this is that sums can be hard to compute, but, thanks to the fundamental theorem, integrals are often easy to compute.

**Example 4.14.** A suspension bridge of length 100 m. has a parabolic support given by

\[ f(x) = x(100 - x) \]

The vertical steel cables are placed 1/2 meter apart. Approximate the total length of cable used.

**Answer:** The exact length of cable is

\[ L = f(1/2) + f(1) + f(3/2) + \ldots + f(100) = \sum_{n=1}^{200} f(n/2). \]

This is a sum, which we will approximate by an integral.

It is typical that what we want to compute is not exactly something we know but can be manipulated into a form we recognize. As long as we keep track of the manipulations we can then go backwards and compute what we want.

In this case the sum for \( L \) is close to a Riemann sum for \( \int_{0}^{100} f(x) \, dx \). So, if we divide the interval \([0, 100]\) into 200 subintervals then \( \Delta x = 1/2 \) and we have the right Riemann sum.

\[ \int_{0}^{100} f(x) \, dx \approx \frac{1}{2} \sum_{n=1}^{200} f(n/2). \]

The sum on the right hand side is just 1/2 the sum for \( L \), so we have

\[ L \approx 2 \int_{0}^{100} f(x) \, dx = 2 \left[ 50x^2 - x^3/3 \right]_{0}^{100} = \frac{10}{3} \times 10^5 \approx 3.33 \times 10^5. \]

(Using a computer we find the exact value is 333,325.)
5  The second fundamental theorem

5.1  Change of variable

In our notation for integrals \( \int_{a}^{b} f(x) \, dx \) why do we include the \( dx \)?

One answer is that we arrived at the integral via a Riemann sum \( \sum_{i=1}^{n} f(c_i) \Delta x \), so the \( dx \) is there to remind us of the \( \Delta x \).

A second answer is that the \( dx \) will help when we make a change of variable.

**Example 5.1.** (Change of variable)

Compute \( \int_{0}^{1} \frac{1}{\sqrt{1-x^2}} \, dx \) by making a change of variable.

**answer:** In a few days we will begin a systematic study of integration techniques. For now we'll make a magic choice of a new variable. You just need to see that we then convert everything to the new variable. That is, we change integrand (the function being integrated), the limits of integration and the differential \( dx \).

Let \( x = \sin(u) \). We then have:

- Differential: \( \frac{dx}{du} = \cos(u) \), so \( dx = \cos(u) \, du \)
- Limits: when \( x = 0 \) we have \( u = 0 \) and when \( x = 1 \) we have \( u = \pi/2 \).
- Integrand: \( \frac{1}{\sqrt{1-x^2}} = \frac{1}{\sqrt{1-\sin^2(u)}} = \frac{1}{\cos(u)} \).

Now we change the original integral in \( x \) to one in \( u \) by substituting in all these pieces:

\[
\int_{0}^{1} \frac{1}{\sqrt{1-x^2}} \, dx = \int_{0}^{\pi/2} \cos(u) \frac{du}{\cos(u)} = \int_{0}^{\pi/2} du = \pi/2.
\]
The First Fundamental Theorem of Calculus says that you can compute the definite integral of a function by finding its antiderivative:

$$\text{If } F' = f \text{ then } \int_a^b f(x) \, dx = F(b) - F(a)$$

This leads us to the question: Given $f(x)$ does an antiderivative $F(x)$ always exist? The answer is yes! We state this as the Second Fundamental of Calculus.

**The Second Fundamental Theorem of Calculus:**

If $f(x)$ is a continuous function then the function $F(x) = \int_a^x f(u) \, du$ has derivative $F'(x) = f(x)$.

In words, this says that $f(x)$ always has an antiderivative which is given by an integral.

### 5.3 Subtleties

There are a few subtleties that we should highlight:

1. We have defined a NEW function $F(x)$ by using the variable $x$ as a limit in the definite integral –as $x$ varies $F(x)$ will also vary. Note that we needed a dummy variable $u$ for integration because $x$ was used as a limit.

2. For any continuous function there is an antiderivative. We might not know it in closed form but we can always write it as a definite integral with a variable limit, i.e., an antiderivative of $f(x)$ is

$$F(x) = \int_a^x f(u) \, du$$

This can be quite useful since Riemann sums let us compute definite integrals as accurately as we wish.

**Example 5.2.** $F_0(x) = \int_0^x t^2 \, dt = x^3/3, \quad F_1(x) = \int_1^x t^2 \, dt = x^3/3 - 1/3$.

**Example 5.3.** The following are not elementary functions but they are functions.

- $F(x) = \int_0^x e^{-t^2} \, dt$ statistics.
- $\text{Li}(x) = \int_2^x \frac{1}{\ln t} \, dt$ number theory.
- $\text{Si}(x) = \int_0^x \sin(t^2) \, dt$ optics.

### 5.4 Natural logarithm as a definite integral

Because we know the derivative of $\ln(x) = 1/x$ we know that $\ln(x) = \int_1^x \frac{1}{t} \, dt$.

Let’s see how we can show that $\ln(x)$ obeys all the rules of logarithms just from this integral. This is an important example of how to derive properties of functions defined as integrals.

**Properties of $\ln(x)$:**
1. $\ln 1 = 0.$
   **Proof:** $\ln(1) = \int_1^1 \frac{1}{t} \, dt = 0.$

2. $\ln(ab) = \ln a + \ln b.$
   **Proof:** (uses change of variable and properties of integrals)
   \[
   \ln(ab) = \int_1^{ab} \frac{1}{t} \, dt = \int_1^a \frac{1}{t} \, dt + \int_a^{ab} \frac{1}{t} \, dt
   \]
   For the second integral in right hand expression we’ll make the change of variable $au = t.$ This gives differential and limits:
   \[a \, du = dt, \quad \text{when } t = a \text{ we have } u = 1, \text{ and when } t = ab \text{ we have } u = b.\]
   Thus
   \[
   \ln(ab) = \int_1^a \frac{1}{t} \, dt + \int_1^b \frac{1}{au} \, a \, du = \int_1^a \frac{1}{t} \, dt + \int_1^b \frac{1}{u} \, du = \ln a + \ln b
   \]

3. $\ln x$ is increasing for $x > 0.$
   **Proof:** The derivative $= \frac{1}{x} > 0.$

4. $\ln(1/a) = -\ln a.$
   **Proof:** Using properties 1 and 2: $0 = \ln 1 = \ln(a \cdot \frac{1}{a}) = \ln a + \ln(1/a)$.

5. $\ln x^n = n \ln x.$
   **Proof:** Using property 2: $\ln x^n = \ln(x \cdot x \cdot \ldots x)$

6. $\ln x \to \infty$ as $x \to \infty.$
   **Proof:** First we know that $\ln 2^n = n \ln 2$ goes to infinity as $n$ increases. Since $\ln(x)$ is increasing this means that it too must go to infinity.

### 5.5 More uses of the second fundamental theorem

**Example 5.4.** Sketch the graph of the function $F(x) = \int_0^x \frac{x^5 - 1}{1 + u^2} \, du.$

**Answer:** The point here is that graphing requires derivatives and that’s an easy thing to do for a function defined as an integral.

Critical points: $F'(x) = \frac{x^5 - 1}{1 + x^2} = 0$ gives $x = 1$ is the only critical point.

Easy to compute values: $F(0) = 0.$

\[
\begin{array}{c|c|c}
F' & F' > 0 & F' < 0 \\
1 & 1 & 1
\end{array}
\]

Sign of $F'(x)$

$y = F(x)$

$F(0) = 0.$
Example 5.5. Let \( \int_0^x f(t) \, dt = 2x(\sin(x) + 1) \) find \( f(\pi/2) \).

**answer:** \( f(x) \) is the derivative of the integral, therefore:

\[
f(x) = \frac{d}{dx} 2x(\sin(x) + 1) = 2(\sin(x) + 1) + 2x(\cos(x)).
\]

Setting \( x = \pi/2 \) we get \( f(\pi/2) = 4 \).

Example 5.6. (Chain rule) Let \( F(x) = \int_1^{x^2} \sin(u^3) \, du \) Find \( F'(x) \).

**answer:** Using the chain rule: \( F'(x) = \sin(x^6) \cdot 2x \).

Example 5.7. If \( \int_0^{x/2} f(t) \, dt = 2x(\sin(x) + 1) \) find \( f(\pi/2) \).

**answer:** Taking derivatives we have

\[
\frac{d}{dx} \int_0^{x/2} f(t) \, dt = \frac{d}{dx} 2x(\sin(x) + 1).
\]

The upper limit of \( x/2 \) in the integral means we need to use the chain rule. So we get

\[
\frac{1}{2} f(x/2) = 2(\sin(x) + 1) + 2x(\cos(x))
\]

Thus, letting \( x = \pi \) we get

\[
\frac{1}{2} f(\pi/2) = 2 + 2\pi(-1) = 2 - 2\pi \Rightarrow f(\pi/2) = 4 - 4\pi.
\]

Example 5.8. (change of variable): Compute \( \int_1^e \frac{\sqrt{\ln(x)}}{x} \, dx \).

**answer:** Substitute: \( u = \ln(x) \Rightarrow du = \frac{1}{x} \, dx \), \( x = 1 \leftrightarrow u = 0 \), \( x = e \leftrightarrow u = 1 \).

\( \Rightarrow \) integral = \( \int_0^1 \sqrt{u} \, du = 2/3 \).

Example 5.9. Compute \( \int_0^\pi \frac{\sin(x)}{(2+\cos(x))^3} \, dx \).

**answer:** Substitute: \( u = \cos(x) \Rightarrow du = -\sin(x) \, dx \), \( x = 0 \leftrightarrow u = 1 \), \( x = \pi \leftrightarrow u = -1 \).

\( \Rightarrow \) integral = \( \int_1^{-1} -\frac{1}{(2+u)^3} \, du = \int_{-1}^{1} \frac{1}{(2+u)^3} \, du = -\frac{1}{2}(2 + u)^{-2}|_{-1}^{1} = 4/9. \)
5.6 Proof of the Second Fundamental Theorem

By definition of the derivative

\[ F'(x) = \lim_{\Delta x \to 0} \frac{\Delta F}{\Delta x} = \lim_{\Delta x \to 0} \frac{F(x + \Delta) - F(x)}{\Delta x}. \]

Since \( F(x) \) is defined as a definite integral it represents area under a graph. So

\[
F(x + \Delta x) - F(x) = \int_{0}^{x+\Delta x} f(x) \, dx - \int_{0}^{x} f(x) \, dx
= \int_{x}^{x+\Delta x} f(x) \, dx
= \text{area in figure at right}
\approx f(x) \Delta x \text{ (valid when } \Delta x \text{ is small)}
\]

Using this we get

\[
\frac{F(x + \Delta) - F(x)}{\Delta x} \approx \frac{f(x) \Delta x}{\Delta x} = \Delta.
\]

In the limit as \( \Delta x \to 0 \) this becomes exact: \( \frac{dF}{dx} = f(x) \). QED

6 Geometric applications

6.1 Main idea

In this topic we are going to develop some geometric applications of integration. That is, we will find the volume, area and length of various solids, surfaces and curves. All of these applications, as well as the physical applications in the next topic, will follow the same general pattern. We’ll describe it for finding volumes of solids, but the same pattern will hold for areas of surfaces etc.

To find the volume of a solid:

1. Divide the solid into little (infinitesimal) pieces, for each of which we have an easy formula for the volume.
2. The total volume is the ‘sum’ of all the pieces. Since in the limit we will have an infinite number of pieces the ‘sum’ will really be an integral.

6.2 Area between curves

Finding the area between two curves is nearly identical to finding the area between a curve and the x-axis. (In fact, if we think of it, the x-axis is a curve.) First we slice the region
into rectangles. The width of each rectangle is $dx$ and its height is the difference between the heights of the curves.

Here are a number of examples. We won't do all of them in class.

**Example 6.1.** Find the area between the curves $y = x + 2$ and $y = x^2$.

**answer:** The steps are given below the figure. The figure illustrates all the steps.

First find points of intersection:

$$x + 2 = x^2 \iff x^2 - x - 2 = 0 \iff x = -1, 2.$$  

This gives the limits of integration.

By graphing figure out which curve is above the other:

$x + 2$ is above $x^2$.

Find the area of one slice:

area of slice $= (x + 2 - x^2) \, dx$.

Find the area by integrating ('summing') the areas of the slices:

area of region $= \int_{-1}^{2} (x + 2) - x^2 \, dx = \frac{9}{2}$.

**Example 6.2.** Find the area between curves $y = x^3$ and $y = 4x$.

**answer:** The curves intersect at $x = -2, 0, 2$.

There are two pieces:

1. For $-2 < x < 0$ the curve $y = x^3$ is on top.
2. For $0 < x < 2$ the curve $y = 4x$ is on top.

Therefore the area has to be computed in two pieces:

$$\text{area} = \int_{-2}^{0} x^3 - 4x \, dx + \int_{0}^{2} 4x - x^3 \, dx = 8.$$  

**Example 6.3.** Find the area between $y = 1 - x^2$ and $y = x + 1$.

Intersection: $1 - x^2 = x + 1 \Rightarrow x = 0, -1$.

On top: $1 - x^2$ is above $x + 1$.

Area $= \int_{-1}^{0} (1 - x^2) - (x + 1) \, dx = 1/6$. 
Example 6.4. Find the area bounded by \( y = \cos x \), \( y = \sin 2x \), between \( 0 \leq x \leq \pi/2 \).

Intersection: \( \cos x = \sin 2x = 2 \sin x \cos x \)
\[ \Rightarrow \cos x = 0 \text{ or } 2 \sin x = 1 \Rightarrow x = \pi/2, \pi/6. \]

In \([0, \pi/6]\) \( \cos x \) is on top.
In \([\pi/6, \pi/2]\) \( \sin 2x \) is on top.

\[
\text{Area} = \int_0^{\pi/6} \cos x - \sin 2x \, dx + \int_{\pi/6}^{\pi/2} \sin 2x - \cos x \, dx
\]
\[
= \left[ \sin x + \frac{1}{2} \cos 2x \right]_0^{\pi/6} - \left[ -\frac{1}{2} \cos 2x - \sin x \right]_{\pi/6}^{\pi/2}
\]
\[
= \left( \frac{1}{2} - 0 + \frac{1}{4} - \frac{1}{2} \right) + \left( \frac{1}{2} - 1 + \frac{1}{4} + \frac{1}{2} \right)
\]
\[
= \frac{1}{2}.
\]

6.3 Volumes of solids

Example 6.5. Volumes by slices

Find the volume of the sphere of radius \( R \).

We place the sphere with center at the origin and slice perpendicular to the \( x \)-axis. The figure at right illustrates that the slice is (approximately) a disk with radius \( r = \sqrt{R^2 - x^2} \) and thickness \( dx \).

Therefore

\[
\text{Volume of slice } dV = \pi r^2 \, dx = \pi (R^2 - x^2) \, dx.
\]

Total Volume = ‘sum’ of slices:
\[
V = \int_{-R}^{R} \pi (R^2 - x^2) \, dx = \pi (R^2 x - x^3/3) \bigg|_{-R}^{R} = \frac{4}{3} \pi R^3.
\]

Example 6.6. Volume of revolution: disk method

Revolve the region under the graph of \( y = f(x) \) and between \( x = a \) and \( x = b \) around the \( x \)-axis: compute the volume of the solid of revolution by vertical slices (disks).

**answer:** The figure below on the left shows a small slice of the region perpendicular to the \( x \)-axis. The figure on the right shows the ‘disk’ formed when that slice is rotated around the \( x \)-axis. The disk has radius \( y \) and thickness \( dx \). Therefore: Volume of slice:
\[
dV = \pi y^2 \, dx = \pi f(x)^2 \, dx
\]

Volume of solid = ‘sum’ of slices: \( V = \int_a^b \pi f(x)^2 \, dx \)
Example 6.7. Find the volume of revolution of the curve $y = 4 - x^2$ between 0 and 2 revolved around the $x$-axis.

**answer:** This is the same as the previous example (with a specific function).

Volume of slice: $dV = \pi y^2 \, dx = \pi (4 - x^2)^2 \, dx$.

Volume = 'sum' of slices:

$$V = \int_0^2 \pi 16 - 8x^2 + x^4 \, dx$$

$$= \pi \left( 16x - \frac{8}{3}x^3 + \frac{1}{5}x^5 \right)_0^2$$

$$= \pi (32 - 64/3 + 32/5).$$

Example 6.8. **Volume by Shells**

Find the volume of revolution of the same curve ($y = 4 - x^2$) around the $y$-axis.

**answer:** Because we often have everything defined in terms of $x$ we want to be able to compute the volumes by slicing in the $x$ direction.

A vertical rectangular slice in the $xy$-plane sweeps out a cylindrical shell when revolved around the $y$-axis. It has the following dimensions (see figure).

- thickness = $dx$
- height = $y$
- radius = $x$

So the volume of a shell is $dV = 2\pi xy \, dx$. Thus the total volume of revolution = 'sum' of shells:

$$V = \int_0^2 2\pi x(4 - x^2) \, dx$$

$$= 2\pi (2x^2 - x^4/4)_0^2 = 8\pi.$$

Example 6.9. **Method of washers**

Revolve area between the curves $y_1 = f_1(x)$ and $y_2 = f_2(x)$ around the $x$-axis.

**answer:** This is just like the method of disks except a smaller disk is removed from the middle of the larger disk. We have (see figures below):

Volume of slice = $dV = \pi (y_1^2 - y_2^2) \, dx$ and Volume = 'sum' of slices = $\int_a^b dV$. 
The same ideas work for volumes of revolution around other lines. **Example 6.10.** Find the volume of revolution of the curve \( y = 4 - x^2 \) between \( x = 0 \) and \( x = 2 \) around the line \( y = -1 \).

**answer:** This is the same as the example on the method of disks above, except instead of rotating around the line \( y = 0 \) (the \( x \)-axis) we rotate around \( y = -1 \).

The volume of a thin disk of rotation is
\[
dV = \pi r^2 dx = \pi (y + 1)^2 dx = \pi (5 - x^2)^2 dx.
\]

Volume = 'sum' of slices:
\[
V = \int_0^2 \pi (25 - 10x^2 + x^4) dx
= \pi (25x - \frac{10}{3}x^3 + \frac{1}{5}x^5)|_0^2
= \pi (50 - \frac{80}{3} + \frac{32}{5}).
\]

6.4 **Arclength**

The arclength of a curve is the distance along the curve. To find the arclength we slice the curve into little pieces each of length \( \Delta s \) (see figure at right). \( \Delta s \) is the distance along the curve, but for a small slice it is approximately straight and the Pythagorean theorem tells us
\[
\Delta s \approx \sqrt{(\Delta x)^2 + (\Delta y)^2}
\]

In the limit we get the **basic formula for arclength:**
\[
ds = \sqrt{(dx)^2 + (dy)^2}
\]

We can algebraically manipulate this formula into a useful form:
\[
ds = \sqrt{(dx)^2 + (dy)^2} \frac{dx}{dx} = \sqrt{1 + (dy/dx)^2} dx
\]
We can ‘sum’ the slices to get the arclength along a curve:

\[
\text{Arclength } = L = \int ds = \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2} \, dx.
\]

**Example 6.11.** Find the arclength of the curve \(y^2 = x^3\) between \((0, 0)\) and \((4, 8)\).

**answer:** First we solve for \(y\) as a function of \(x\): \(y = x^{3/2}\). Taking the derivative we get,
\[
\frac{dy}{dx} = \frac{3}{2} x^{1/2}, \quad \text{and } ds = \sqrt{1 + \frac{9}{4} x} \, dx.
\]
Thus
\[
\text{Arclength } = L = \int_0^4 \sqrt{1 + \frac{9}{4} x} \, dx = \frac{8}{27} (1 + \frac{9}{4} x^{3/2}) \bigg|_0^4 = \frac{8}{27} (10^{3/2} - 1).
\]

**Example 6.12.** Find the arclength of \(y = \sin x\) for \(x\) in \([0, \pi]\).

**answer:** We have \(\frac{dy}{dx} = \cos x\), so \(ds = \sqrt{1 + \cos^2 x} \, dx\). This gives
\[
\text{Arclength } = L = \int_0^\pi \sqrt{1 + \cos^2 x} \, dx.
\]

It is not possible to compute this integral in terms of elementary functions. Later we will approximate it using Riemann sums. As an aside: mathematicians know call this an **elliptic integral**.

### 6.5 Surface area of revolution

This is trickier than volume. The main idea is that a line segment revolved around the \(x\)-axis gives the **frustrum of a cone**. You can read the textbook §7.6 for details. It derives the following formula:

A slice of the curve \(y = f(x)\) rotated around the \(x\)-axis has area
\[
dA = 2\pi y \, ds = 2\pi y \sqrt{1 + (y')^2} \, dx.
\]

In general, for a element \(ds\) rotated in a circle of radius \(r\) we get \(dA = 2\pi r \, ds\).
**Example 6.13.** (i) Find the area of the surface of revolution of \( y = 1/x \) between 1 and \( b \) revolved around the \( x \)-axis.

**Answer:** First we find \( dA \):

\[
y' = -1/x^2 \\
ds = \sqrt{1 + (y')^2} \, dx = \sqrt{1 + (-1/x^2)^2} \, dx \\
dA = 2\pi y \, ds = 2\pi \frac{1}{x} \sqrt{1 + (-1/x^2)^2} \, dx.
\]

Now, the formula for surface area gives

\[
\text{Surface area } A = \int_1^b 2\pi \frac{1}{x} \sqrt{1 + 1/x^4} \, dx.
\]

This is hard to compute but we can make some estimates. Since \( \sqrt{1 + 1/x^4} > 1 \) we have

\[
A > \int_1^b 2\pi \frac{1}{x} \, dx = 2\pi \ln b.
\]

Letting \( b \) get large we see that the area \( A \) goes to infinity as \( b \to \infty \).

(ii) Find the volume of revolution of the same curve.

Volume of slice \( dV = \pi y^2 \, dx = \pi \frac{1}{x^2} \, dx \). Thus,

\[
V = \int_1^b \frac{\pi}{x^2} \, dx = -\frac{\pi}{x} \bigg|_1^b = \pi(1 - \frac{1}{b}).
\]

Letting \( b \) go to infinity we see that \( V \to \pi \) as \( b \to \infty \).

So, the curve \( y = 1/x \) for \( x \) in \([1, \infty]\) has the odd feature that its volume of revolution is finite and its surface area of revolution is infinite. What happens if you fill the volume of volume of revolution with paint?

---

### 7 Applications: work and average value

#### 7.1 Average Value

We’ll start with a discussion of averages which will lead us to the geometric definition of the average value of a function. After some examples we’ll see the connection of average value with Riemann sums.

To explain the term average value we’ll start by examining the usual meaning of an average. Suppose there are 10 apples to be divided among 4 people then the average number of apples per person is 2.5. One way to think of this is that the average of 2.5 is the number that makes a total of 10 apples when each person is given an equal amount. This is illustrated in the following picture:
The graphs shows two ways to divide 10 apples among 4 people. Both ways total 10 apples; when everyone gets the average value the graph is a horizontal line.

This motivates our idea of the average value of a function. The average value of \( f(x) \) is the constant function that has the same total area under the curve as \( f(x) \).

Average value of \( f(x) \) for \( x \) between \( a \) and \( b \): the area of the blue rectangle is the same as the area of the orange region under the curve.

This leads directly to the definition of average value

**Definition.** The average value \( A \) of the function \( f(x) \) with respect to \( x \) over \([a, b]\) is

\[
A = \frac{1}{b-a} \int_{a}^{b} f(x) \, dx \quad \text{(average value)}
\]

Let’s verify that the area of the blue rectangle in the above figure is the same as the area under the curve. Using the above definition of \( A \) we have

\[
\text{Area of rectangle} = (b-a)A
\]

\[
= (b-a) \left( \frac{1}{b-a} \int_{a}^{b} f(x) \, dx \right)
\]

\[
= \int_{a}^{b} f(x) \, dx = \text{Area under the curve}
\]
Applying this definition is straightforward:

**Example 7.1.** Find the average values of \( y = x^2 \) with respect to \( x \) for \( x \) in \([5, 10]\)

**answer:** Average value \( = \frac{1}{5} \int_5^{10} x \, dx = 15/2. \)

**Example 7.2.** Find the average value of \( x \) on \([0, 3]\)

**answer:** In this problem we have to assume from context that the average is of the function \( y = x \) with respect to \( x \).

\[
\text{Average} = \frac{1}{3} \int_0^3 x \, dx = 3/2.
\]

**Example 7.3.** Find the average value of \( \sin x \) on \([0, \pi]\).

**answer:** Average \( = \frac{1}{\pi} \int_0^\pi \sin x \, dx = 2/\pi. \)

**Average value as a (limit of) Riemann sums**

Using the definition of a Riemann sum and a little bit of algebra we’ll see that the average value of a function (which is an integral) is approximated by the average of \( n \) values of the function (which is a sum).

To take the average of \( n \) numbers you sum them up and divide by \( n \), e.g.

\[
\text{Average of } y_1, y_2, \ldots y_n = \frac{y_1 + y_2 + \ldots + y_n}{n}.
\]

This looks like a lot like a Riemann sum. For example the right Riemann sum for \( \int_a^b f(x) \, dx \) is

\[
\int_a^b f(x) \, dx \approx \sum_{i=1}^{n} f(x_i) \Delta x = \sum_{i=1}^{n} f(x_i) \frac{b-a}{n}.
\]

To get the last expression you have to substitute \( \Delta x = \frac{b-a}{n} \). Dividing both sides of the above by \( (b-a) \) gives us

\[
\frac{1}{b-a} \int_a^b f(x) \, dx \approx \frac{1}{n} \sum_{i=1}^{n} f(x_i)
\]

The left hand side is our definition of the average value of \( f(x) \). The right hand side is the average of \( n \) values of \( f(x) \).

**What you average with respect to is important**

**Example 7.4.** Suppose a car moves during the time interval \([0, 9]\) and that its position with respect to time is given by \( x = \frac{1}{2} t^2 \).

(i) Compute the average velocity with respect to time.

(ii) Compute the average velocity with respect to position.

Note: question (i) is about averaging velocity over equally spaced time intervals and (ii) is about averaging velocity over equally spaced positions. In terms of the car scenario we can say:
Question (i) is like checking the speedometer on your car every second and averaging the resulting data. Question (ii) is like placing people every 50 meters along the road, having each of them measure your speed as you drive past and then averaging all that data.

**answer:** (i) We easily compute velocity as a function of \( t \) is \( v = x' = t \). The average of velocity with respect to \( t \) is therefore

\[
\frac{1}{9} \int_0^9 t \, dt = \frac{9}{2}.
\]

(ii) Over the time interval \([0,9]\) the car goes from \( x_0 = 0 \) to \( x_1 = \frac{81}{2} \). Using \( x = \frac{t^2}{2} \) and velocity \( v = t \) we get that \( x = \frac{v^2}{2} \). So, velocity as a function of \( x \) is

\[
v = \sqrt{2x}
\]

The average value of \( v \) with respect to w.r.t. \( x \) is

\[
\frac{1}{\frac{81}{2}} \int_0^{\frac{81}{2}} \sqrt{2x} \, dx = \frac{2\sqrt{2} \left( \frac{81}{2} \right)^{3/2}}{3 \frac{81}{2}} = 6.
\]

Can you find a simple explanation for why the average with respect to distance was greater than the average with respect to time?

**Example 7.5.** (This will be done at the end of class if there’s time.)

A point is chosen at random on the \( x \)-axis between -1 and 1; call it \( P \). What is the average length of the vertical line from \( P \) to the unit circle?

**answer:** Random on the \( x \)-axis means averaging over \( x \).

If \( P \) is at the point \( x \) on the \( x \)-axis then the vertical line to the circle has length \( \sqrt{1 - x^2} \). Averaging this with respect to \( x \) gives

\[
\text{average length} = \frac{1}{2} \int_{-1}^{1} \sqrt{1 - x^2} \, dx
\]

\[
= \frac{1}{2} \text{ area of semicircle } = \frac{\pi}{4}.
\]

![Diagram](image)

**Example 7.6.** Repeat the previous problem except now take \( Q \) randomly on the circumference of the unit semicircle. What is the average length of the vertical line from \( Q \) to the \( x \)-axis?

**answer:** Now we must average over \( \theta \) (see above figure). For fixed \( \theta \) the vertical line from \( Q \) to the \( x \)-axis has length \( \sin \theta \). Averaging with respect to \( \theta \) gives:

\[
\text{average value} = \frac{1}{\pi} \int_0^\pi \sin \theta \, d\theta = \frac{2}{\pi}.
\]
This average is different from that of the previous example!

**Example 7.7.** (Using integrals to approximate a finite sum.)

Given a harp as shown below, what is the average length of a string?

\[ y = 4 - x^2 \]

**Answer:** This is an inverse kind of problem in that we will use an integral to approximate a finite sum.

Let \( y = 4 - x^2 \) and call the number of strings \( n \). We label the points in the figure as shown below. Note we changed the scale of the \( x \)-axis just to fit in the labels.

This is really just a repeat of the section above labeled ‘average value as a (limit of) Riemann sums’. We redo the argument because the indexing here is slightly different. Let \( A \) be the average length of the strings, so \( A = \frac{1}{n} \sum_{i=0}^{n-1} y_i \). (We leave off \( y_n \) because it’s zero.) We can manipulate the sum to be a Riemann sum and approximate it by an integral as follows.

\[
n A \Delta x = \sum_{i=0}^{n-1} y_i \Delta x \approx \int_0^2 4 - x^2 \, dx = \frac{16}{3}.
\]

The final step is to remember that \( \Delta x = 2/n \) or \( n \Delta x = 2 \). So

\[
A = \frac{1}{n \Delta x} \cdot \frac{16}{3} = \frac{1}{2} \cdot \frac{16}{3} = \frac{8}{3}.
\]

### 7.2 Work

For a constant force in the direction of motion: work = force \( \times \) distance.

Typical units for work (or energy) are: ft-lb, joule = newton-meter, erg = dyne-cm.

**Hooke’s law:** \( F = -kx \). Here \( F \) is the force needed to compress or stretch a spring \( x \) units from its unstretched (equilibrium) position and \( k \) is the **spring constant** (the bigger \( k \) the stiffer the spring).
Gravitation: $F = \frac{GmM}{x^2}$ where $F$ is the gravitational attraction of two masses $m$ and $M$ a distance $x$ apart. $G$ is the universal constant of gravitation.

Example 7.8. A 100 ft long chain weighing 5 lb/ft hangs off a building. How much work is expended to haul it up to the roof?

**answer:** We’ll work the problem twice using two different ways of slicing and summing.

**Method 1:** slice into small pieces (masses) of chain. The mass of little segment of length $\Delta x$ is $\Delta m = 5 \Delta x$.

If we let $x = $ distance from the roof to the little segment. Then the work needed to haul a little segment of chain to the roof is

$$\Delta W = x \Delta m = 5 \Delta x$$

So the Total work $W \approx \sum \Delta W = \sum 5x \Delta x$.

Letting $\Delta x$ go to 0 this becomes an integral:

$$W = \int_0^{100} 5x \, dx = 25,000 \text{ lb-feet}.$$  

**Method 2:** slice into small movements of a length of chain. Let $dW$ be the work needed to move all of chain that is still hanging a distance $dx$. We can now make the calculation:

Let $x = $ length of chain still hanging.

So, the mass of remaining chain is $5x$.

The work needed to lift this chain a distance $dx$ is $dW = 5x \, dx$.

Therefore the total work to haul in the chain is

$$W = \int_0^{100} 5x \, dx = 25,000 \text{ lb-feet}.$$  

Example 7.9. Suppose a spring has natural length 10 in. and it takes 12 lb. of force stretches it 1/2 in. Find the work done in stretching it from 10 to 18 in.

**answer:** First we use the data given to find the value of $k$

$$12 \text{ lb} = k \cdot \frac{1}{2} \text{ in.} \Rightarrow k = 24 \text{ lb/in.}$$
By Hooke’s law the work done stretching it from \( x \) to \( x + dx \) is \( dW = kx \, dx \). Thus the total work done is found by ‘summing’ all the infinitesimal bits of work

\[
W = \int_{10}^{18} kx \, dx = \int_{10}^{18} 24x \, dx = 12x^2 \bigg|_{10}^{18} = 2688 \text{ lb-in} = 224 \text{ lb-ft}.
\]

(Note: You should always be careful with units.)

**Example 7.10. (Gravitation)**

How much work is done in moving a mass \( m \) from the surface of a planet of mass \( M \) and radius \( R \) to a height \( H \) above the surface of the planet?

**answer:** Units aren’t given so we’ll assume that the universal gravitation constant \( G \) is given in compatible units. The weight of the mass at a height \( h \) above the surface is \( \frac{GMm}{(R+h)^2} \).

So the work \( \Delta W \) done in going from \( h \) to \( h + \Delta h \) is approximated by

\[
\Delta W \approx \frac{GMm}{(R+h)^2} \Delta h.
\]

The total work is is given by summing all the \( \Delta W \) pieces. In the limit as \( \Delta h \to 0 \) the approximation for \( \Delta W \) becomes exact and the sum becomes an integral:

\[
W = \int_0^H \frac{GMm}{(R+h)^2} \, dh = -\frac{GMm}{R+h} \bigg|_0^H = GMm \left( \frac{1}{R} - \frac{1}{R+H} \right).
\]

A high point of Western civilization!

---

8 Integration: u-substitution, trig-substitution

**Integration techniques** Only practice will make perfect. These techniques are important, but *not* the intellectual heart of the class.

1. Inspection: \( \int x^2 \, dx = \frac{x^3}{3} + C \).
2. Guess/memorize: \( \int \sec \theta \, d\theta = \ln |\sec \theta + \tan \theta| + C \). **Memorize this!**
3. Direct substitution: \( u = g(x) \Rightarrow du = g'(x) \, dx \).

**Examples:**

(a) Compute \( \int \sin^4 x \cos x \, dx \).

**answer:** Let \( u = \sin x \Rightarrow du = \cos x \, dx \).

Substitute for all pieces in integral: \( \Rightarrow \int u^4 \, du = \frac{u^5}{5} + C \).
Back substitution: integral = $\sin^5 x + C$.

(b) Compute $\int \sin^5 x \, dx$.

answer: $\int \sin^5 x \, dx = \int \sin^4 x \sin x \, dx = \int (1 - \cos^2 x)^2 \sin x \, dx$.

Let $u = \cos x \Rightarrow \, du = -\sin x \, dx$.

Substitute: integral = $-\int (1 - u^2)^2 \, du$ (Easy to compute and back substitute)

Trig formulas you have to know:

1. $\sin^2 x + \cos^2 x = 1$.

2. Double angle: $\sin 2x = 2 \sin x \cos x$  $\cos 2x = \cos^2 x - \sin^2 x$.

3. Half angle: $\cos^2 x = \frac{1}{2}(1 + \cos 2x)$  $\sin^2 x = \frac{1}{2}(1 - \cos 2x)$.

4. $1 + \tan^2 x = \sec^2 x$  $\sec^2 x - 1 = \tan^2 x$.

5. $\frac{d}{dx} \sin x = \cos x$  $\frac{d}{dx} \cos x = -\sin x$.  $\frac{d}{dx} \tan x = \sec^2 x$  $\frac{d}{dx} \sec x = \sec x \tan x$.

Examples:

1. Compute $\int \sin^4 x \, dx$.

answer: $\sin^4 x = \left(\frac{1 - \cos 2x}{2}\right)^2 = \frac{1}{4} - \frac{1}{2} \cos 2x + \frac{1}{4} \cos^2 2x = \frac{1}{4} - \frac{1}{2} \cos 2x + \frac{1}{8}(1 + \cos 4x)$.

This last expression is easy to integrate.

2. Compute $\int \tan^2 x \sec^2 x \, dx$.

answer: Notice $\sec^2 x$ is the derivative of $\tan x$.

$\Rightarrow$ substitute $u = \tan x$, $du = \sec^2 x \, dx$.

$\Rightarrow$ integral = $\int u^2 \, du = u^3/3 + C$.

Back substitute: integral = $\tan^3 x/3 + C$.

If you’re comfortable with it, you can write this without the $u$.

$\int \tan^2 x \sec^2 x \, dx = \int \tan^2 x \, d(\tan x) = \frac{\tan^3 x}{3} + C$.

3. Compute $\int \tan^2 x \sec^4 x \, dx$.

answer: Notice $\sec^4 x = (1 + \tan^2 x) \sec^2 x = (1 + \tan^2 x) \, d\tan x$.

Substitute $u = \tan x$
\[ \Rightarrow \text{integral} = \int u^2(1 + u^2) \, du. \text{ (Easy to integrate and back substitute.)} \]

**Inverse trig substitution**

**Most important example:** \((\tan^{-1} x)\)

Compute \(\int \frac{1}{1 + x^2} \, dx.\)

**answer:** Substitute \(x = \tan \theta \Rightarrow dx = \sec^2 \theta \, d\theta.\)

\[ \Rightarrow \text{integral} = \int \frac{1}{1 + \tan^2 \theta} \sec^2 \theta \, d\theta = \int \sec^2 \theta \, d\theta = \int d\theta = \theta + C \]

Back substitution (use \(\theta = \tan^{-1} x\)) gives,

\[
\int \frac{1}{1 + x^2} \, dx = \tan^{-1} x + C.
\]

**Other examples:**

1. Compute \(\int \frac{1}{\sqrt{1-x^2}}.\)

**answer:** Substitute \(x = \sin \theta \Rightarrow dx = \cos \theta \, d\theta.\)

\[ \Rightarrow \text{integral} = \int \frac{1}{\sqrt{1 - \sin^2 \theta}} \cos \theta \, d\theta. \]

(Notice how easy the substitution is.)

Trig identities \(\Rightarrow\) integral \(= \int \frac{1}{\cos \theta} \cos \theta \, d\theta = \int d\theta = \theta + C.\)

Back substitution (Use \(\theta = \sin^{-1} x\)) \(\Rightarrow\) integral \(= \sin^{-1} x + C.\)

Check answer by differentiation.

2. Compute \(\int \frac{1}{\sqrt{a^2 - x^2}}.\)

**answer:** Substitute \(x = a \sin \theta \Rightarrow dx = a \cos \theta \, d\theta.\)

This is now similar to example 1.

3. Compute \(\int \frac{1}{\sqrt{a^2 + x^2}}.\)

**answer:** Substitute \(x = a \tan \theta \Rightarrow dx = a \sec^2 \theta \, d\theta.\)

\[ \Rightarrow \text{integral} = \int \frac{1}{\sqrt{a^2 + a^2 \tan^2 \theta}} a \sec^2 \theta \, d\theta. \]

Using algebra and trig identities we get:

\[
\text{integral} = \int \frac{1}{\sec \theta} \sec^2 \theta \, d\theta = \int \sec \theta \, d\theta = \ln |\sec \theta + \tan \theta| + C
\]

Back substitution (Use \(\sec \theta = \sqrt{1 + (x/a)^2}\)) gives

\[
\text{integral} = \ln |\sqrt{1 + (x/a)^2} + x/a| + C = \ln |\sqrt{a^2 + x^2} + x| - \ln a + C
\]

\[ = \ln |\sqrt{a^2 + x^2} + x| + C. \]

(In the last equality we replaced the constant \(\ln a + C\) by \(C\).)

You should check the answer by differentiation.

**Example:** Moment of inertia of uniform disk of radius \(a\) around a diameter.
**Answer:** Moment of inertia of point mass about an axis is \( I = md^2 \), where \( m \) is the mass and \( d \) is the distance from the axis.

Choose the diameter to be along the \( y \)-axis.

Let total mass of disk = \( M \) (so uniform density \( \delta = M/(\pi a^2) \)).

Usual slice and sum technique:
Vertical strip is (approximately) a distance \( x \) from axis.
Area of strip = \( dA = 2y \, dx \).
Mass of strip = \( dm = \delta \, dA = 2\delta y \, dx \).
Moment of inertia of strip = \( dI = x^2 \, dm = x^2 \delta y \, dx \).
Total moment of inertia = \( I = \int 2\delta x^2 y \, dx \).

Always use symmetry – it suffices to compute for half disk (and multiply by 2).

Adding limits and \( y = \sqrt{a^2 - x^2} \): \( I = 2 \int_0^a 2\delta x^2 \sqrt{a^2 - x^2} \, dx \).

Substitute \( x = a \sin \theta \), \( dx = a \cos \theta \, d\theta \).
This implies, \( x = 0 \leftrightarrow \theta = 0 \), \( x = a \leftrightarrow \theta = \pi/2 \).

\[ I = 4\delta \int_0^{\pi/2} a^2 \sin^2 \theta \sqrt{a^2 - a^2 \sin^2 \theta} \, a \cos \theta \, d\theta = 4\delta \int_0^{\pi/2} a^4 \sin^2 \theta \cos^2 \theta \, d\theta. \]

Trig identities: \( \sin^2 \theta \cos^2 \theta = 1/8(1 - \cos 4\theta) \)

\[ I = \frac{1}{2} a^4 \delta \int_0^{\pi/2} 1 - \cos 4\theta \, d\theta = \frac{1}{2} a^4 \delta (\theta - \sin(4\theta)/4) \bigg|_0^{\pi/2} = a^4 \delta \pi/4 = Ma^2/4. \]

**Completing the square:**
(This is how you derive the quadratic formula)

**Example:** Compute \( \int \frac{1}{x^2 + 2x + 5} \, dx \).

Complete the square: \( x^2 + 2x + 5 = x^2 + 2x + 1 + 4 - 1 = (x + 1)^2 + 4 \).

\[ \Rightarrow \int \frac{1}{x^2 + 2x + 5} \, dx = \int \frac{1}{(x + 1)^2 + 4} \, dx \]

Substitute \( x + 1 = 2 \tan u \), \( dx = 2 \sec^2 u \, du \)

\[ \Rightarrow \text{integral} = \int \frac{2 \sec^2 u}{4 \sec^2 u} \, du = \frac{1}{2} u + C = \frac{1}{2} \tan^{-1}(\frac{x + 1}{2}) + C \]
Example: 5D-11. Compute \( \int \frac{x}{\sqrt{-8 + 6x - x^2}} \, dx \).

**answer:** Complete the square inside the square root:

\[-(x^2 - 6x + 8) = -(x^2 - 6x + 9 - 9 + 8) = -(x - 3)^2 + 1.\]

This implies,

\[
\text{integral} = \int x \sqrt{1 - (x - 3)^2} \, dx \quad \text{(Substitute } \sin u = x - 3, \cos u \, du = dx.\text{)}
\]

\[
= \int (\sin u + 3) \cos^2 u \, du.
\]

Two pieces: (i) \( \int \sin u \cos^2 u \, du = -\frac{1}{3} \cos^3 u \).

(ii) \( \int 3 \cos^2 u \, du = \frac{3}{2} \int 1 + \cos 2u \, du = \frac{3}{2} (u + \sin(2u)/2) \).

\[
\begin{array}{c}
\Delta \\
\cos^2 u \quad \sin u \\
\hline
\sqrt{1 - (x - 3)^2} \\
(x - 3)
\end{array}
\]

So,

\[
\text{integral} = -\frac{1}{3} \cos^3 u - \frac{3}{4} \sin 2u - \frac{3}{2} u + C
\]

\[
= -\frac{1}{3} (-8 + 6x - x^2)^{3/2} + \frac{3}{2} (x + 3)(-8 + 6e - x^2)^{1/2} + \frac{3}{2} \sin^{-1}(x - 3) + C.
\]

---

**9 Integration: partial fractions**

**Example 9.1.** Algebra shows that \( \frac{4}{x - 3} - \frac{1}{x - 1} = \frac{3x - 1}{x^2 - 4x + 3} \).

**Question:** How do we go backwards?

**answer:** Suppose we are given \( \frac{3x - 1}{x^2 - 4x + 3} \), then by factoring the denominator we can write

\[
\frac{3x - 1}{x^2 - 4x + 3} = \frac{3x - 1}{(x - 3)(x - 1)} = \frac{A}{x - 3} + \frac{B}{x - 1}.
\]

The goal now is to determine the values of \( A \) and \( B \).

The long method is to cross-multiply which gives

\( 3x - 1 = A(x - 1) + B(x - 3) \).

If we let \( x = 1 \) we see that \( B = -1 \).

Likewise, if \( x = 3 \) we see \( A = 4 \).
Coverup method: Doing the above without writing. This is very useful. You should read the supplementary notes §F. Here is the basic idea. The partial fractions decomposition is

\[
\frac{3x - 1}{x^2 - 4x + 3} = \frac{3x - 1}{(x - 3)(x - 1)} = \frac{A}{x - 3} + \frac{B}{x - 1}.
\]  

(1)

Multiply both sides by \(x - 3\). This produces,

\[
\frac{3x - 1}{x - 1} = A + \frac{B(x - 3)}{x - 1}.
\]  

(2)

Now set \(x = 3\) to find that \(8/2 = A\), i.e. \(A = 4\) (as before). Likewise, multiply both sides by \(x - 1\) to get:

\[
\frac{3x - 1}{x - 3} = \frac{A(x - 1)}{x - 3} + B
\]

Set \(x = 1\) to find that \(2/(-2) = B\), i.e. \(B = -1\).

Why is this called coverup? Because you can perform the calculation without writing things out. Looking at equation 2 we see that to find \(A\) all we have to do with Equation 1 is ‘coverup’ the factor of \((x - 3)\) in the denominator and substitute \(x = 3\) into what is left.

Question: Why go backwards?

answer: Because the partial fraction terms are easy to integrate!

Example 9.2. Compute \(\int \frac{1}{x^2 - 3x + 2} \, dx\).

answer: Partial fractions:

\[
\frac{1}{x^2 - 3x + 2} = \frac{1}{(x-1)(x-2)} = \frac{A}{x-1} + \frac{B}{x-2}.
\]

By coverup: \(A = -1, \ B = 1\). Thus

\[
\text{integral} = \int -\frac{1}{x-1} + \frac{1}{x-2} \, dx = -\ln(x-1) + \ln(x-2) + C.
\]

Getting fancier:

Repeated linear factors:

Example 9.3. Write \(\frac{x^2-2x+2}{(x+2)^2(x-2)}\) in terms of partial fractions.

answer: \(\frac{x^2-2x+2}{(x+2)^2(x-2)} = \frac{A}{(x+2)^2} + \frac{B}{x+2} + \frac{C}{x-2}\) (One term for each power.)

Coverup gives \(A = -5/2, \ C = 1/8\).

To find \(B\) we cross multiply and solve:

\[
x^2-2x+2 = A(x-2)+B(x+2)(x-2)+C(x+2)^2 = (B+C)x^2+(A+4C)x+(-2A-4B+4C).
\]

When two polynomials are the same, the coefficients must be equal. So,

Coefficient of \(x^2\): \(1 = B + C\). Since \(C\) is known, we get \(B = 7/8\).

Since we’ve found \(A, B\) and \(C\) there is no need to write down the other equations. We record them so you can check our answer is correct.
Coefficient of $x$: $-2 = A + 4C$.
Coefficient of 1 (constant terms): $2 = -2A - 4B + 4C$.

Quadratic factor:

**Example 9.4.** Write $\frac{x^2 + 3}{(x^2 + 1)(x - 1)}$ in terms of partial fractions.

**answer:** Since the $x^2 + 1$ term in the denominator doesn’t factor, it appears as is in the partial fractions decomposition. It’s numerator must be of the form $Bx + C$.

$$\frac{x^2 + 3}{(x^2 + 1)(x - 1)} = \frac{A}{x - 1} + \frac{Bx + C}{x^2 + 1}.$$ 

Coverup gives $A = 2$. We can’t use coverup to find $B$ or $C$, so we cross multiply and equate coefficients.

$$x^2 + 3 = A(x^2 + 1) + (Bx + C)(x - 1) = (A + B)x^2 + (C - B)x + (A - C).$$

Coefficient of $x^2$: $1 = A + B$, $A$ is known, so $B = -1$.
Coefficient of $x$: $0 = C - B$, so $C = -1$.

We have all the unknowns, so we don’t need to equate constant terms.

**Example 9.5.** (Integration)

Compute the integral $I = \int \frac{x^2 + 3}{(x^2 + 1)(x - 1)} \, dx$.

**answer:** From the previous example we know the partial fraction decomposition of the integrand. So,

$$I = \int \frac{2}{x - 1} \, dx - \int \frac{x}{x^2 + 1} \, dx - \int \frac{1}{x^2 + 1} \, dx$$

$$= 2 \ln |x - 1| - \frac{1}{2} \ln(x^2 + 1) - \tan^{-1} x + C.$$ 

(You should memorize $\int \frac{1}{x^2 + 1} = \tan^{-1} x + C$.)

**Repeated Quadratic Factor:**

**Example 9.6.** Write $\frac{2x + 1}{(x^2 + 1)^2(x - 1)}$ in terms of partial fractions.

**answer:** We need one term for each power of $(x^2 + 1)$. That is

$$\frac{2x + 1}{(x^2 + 1)^2(x - 1)} = \frac{Ax + B}{(x^2 + 1)^2} + \frac{Cx + D}{x^2 + 1} + \frac{E}{x - 1}.$$ 


We can’t get the other unknowns using coverup, so we cross multiply

$$2x + 1 = (Ax + B)(x - 1) + (Cx + D)(x^2 + 1)(x - 1) + E(x^2 + 1)^2 = (C + E)x^4 + (D - C)x^3 + (A + C - D + 2E)x^2 + (A + B - E)x.$$

Equate coefficients:
Coeficient of $x^4$: $0 = C + E$. $E$ is known, so $C = -3/4$
Coefficient of $x^3$: $0 = D - C \Rightarrow D = -3/4$
Coefficient of $x^2$: $0 = A + C - D + 2E \Rightarrow A = -3/2$
Coefficient of $x$: $2 = -A + B - C + D \Rightarrow B = 1/2$
Coefficient of $1$: $1 = -B - D + E$ (Don’t need this, but it checks out.)

**Example 9.7.** (More integration)
Compute $\int \frac{x}{(x^2 + 1)^2} \, dx$.

**Answer:** Note: this is not a partial fractions problem, because this is already a typical partial fractions term.
Substitute $u = x^2 + 1$, $du = 2x \, dx$. Thus,
\[
\text{integral} = \int \frac{1}{2u} \, du = -\frac{1}{2}u^{-1} + C = -\frac{1}{2}(x^2 + 1)^{-1} + C.
\]

**Long division**
For partial fractions the rational function must be proper, i.e. the degree of the numerator must be less than the degree of the denominator.

**Example 9.8.** Decompose $\frac{x^3 + 2x + 1}{x^2 + x - 2} = \frac{x^3 + 2x + 1}{(x + 2)(x - 1)}$ using partial fractions.

**Answer:** First use long division to write this as a polynomial plus a proper rational function.

\[
x^2 + x - 2 \quad \begin{array}{c|cc}
& x - 1 & \\
\hline
x^3 & x^3 & \quad 2x \\
x^3 & +x^2 & -2x \\
\hline
-x^2 & -x & +2 \\
-x^2 & -x & +2 \\
\hline
5x & -1
\end{array}
\]
This shows that
\[
\frac{x^3 + 2x + 1}{x^2 + x - 2} = x - 1 + \frac{5x - 1}{x^2 + x - 2},
\]

Now proceed as always.
\[
\frac{5x - 1}{x^2 + x - 2} = \frac{5x - 1}{(x + 2)(x - 1)} = \frac{A}{x + 2} + \frac{B}{x - 1}.
\]
Coverup gives $A = 11/3$ and $B = 4/3$. So the integral in question is
\[
\int \frac{x - 1 + \frac{A}{x + 2} + \frac{B}{x - 1}}{2} \, dx = x^2 - x + A \ln(|x + 2|) + B \ln(|x - 1|) + C.
\]

10 Integration by parts; numerical integration

10.1 Integration by parts
Integration by parts is just the **product rule applied to integration**. To introduce this we start with the product rule and integrate:
Product rule: \((uv)' = uv' + u'v\)

Integrate both sides: \(\int (uv)' \, dx = \int uv' \, dx + \int u'v \, dx\)

Since \(\int (uv)' = uv\):
\[
\int uv' \, dx = \int u'v \, dx
\]

Simple algebra:
\[
\int uv' \, dx = uv - \int u'v \, dx
\]

We summarize by putting the last equation in a box: the integration by parts formula is
\[
\int uv' \, dx = uv - \int u'v \, dx
\]

**Example 10.1.** Find \(\int xe^x \, dx\).

**answer:** (I suggest you learn to use the following format.)

Make the following choices for \(u\) and \(v'\):

\[
\begin{align*}
  u &= x & v' &= e^x \\
  u' &= 1 & v &= e^x
\end{align*}
\]

The integration by parts formula says
\[
\int xe^x \, dx = xe^x - \int e^x \, dx = xe^x - e^x + C.
\]

**Discussion:** Choosing \(u\) and \(v'\) is an art. Notice that in the above example \(u = x\) gets simpler when we take \(u'\) and that \(v' = e^x\) is easy to integrate.

**Example 10.2.** Find \(\int \ln(x) \, dx\).

**answer:** At first glance this integral does not look like a candidate for integration by parts because it seems to have only one part. But we can always let \(v' = 1\) for the second part:

\[
\begin{align*}
  u &= \ln x & v' &= 1 \\
  u' &= \frac{1}{x} & v &= x
\end{align*}
\]

So,
\[
\int \ln x \, dx = x \ln x - \int \frac{dx}{x} = x \ln x - x + C.
\]

**Example 10.3.** (Sometimes you have to iterate.) Find \(\int x^2e^x \, dx\).

**answer:** Choose \(u\) and \(v'\):

\[
\begin{align*}
  u &= x^2 & v' &= e^x \\
  u' &= 2x & v &= e^x
\end{align*}
\]

Thus,
\[
\int x^2e^x \, dx = x^2e^x - \int 2xe^x \, dx.
\]
Now we’ve reduced the original integral to another integral (with a smaller power of \( x \)) that we have to integrate by parts. Actually, we did this new integral in a previous example:

\[
\int 2xe^x \, dx = 2xe^x - 2e^x.
\]

So, the answer to the question is \( \int x^2e^x \, dx = x^2e^x - 2xe^x + 2e^x + C \).

**Example 10.4. (Another trick for your bag of tricks.)** Find \( \int e^x \cos(x) \, dx \).

**answer:** Both parts \( e^x \) and \( \cos(x) \) are easy to integrate and differentiate so we make a guess at how to split them:

<table>
<thead>
<tr>
<th>( u )</th>
<th>( v' )</th>
<th>( u' )</th>
<th>( v )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \cos(x) )</td>
<td>( e^x )</td>
<td>( -\sin(x) )</td>
<td>( e^x )</td>
</tr>
</tbody>
</table>

This gives

\[
\int e^x \cos(x) \, dx = e^x \cos(x) + \int e^x \sin(x) \, dx.
\]

Here’s the trick: the integral on the right is of the same form as before so we try a second integration by parts:

<table>
<thead>
<tr>
<th>( u )</th>
<th>( v' )</th>
<th>( u' )</th>
<th>( v )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sin(x) )</td>
<td>( e^x )</td>
<td>( \cos(x) )</td>
<td>( e^x )</td>
</tr>
</tbody>
</table>

Which gives

\[
\int e^x \cos(x) \, dx = e^x \cos(x) + \int e^x \sin(x) \, dx = e^x \cos(x) + e^x \sin(x) - \int e^x \cos(x) \, dx.
\]

Now the integral on the right is the same as our original integral with a minus sign. Algebra now yields:

\[
2 \int e^x \cos(x) \, dx = e^x \cos(x) + e^x \sin(x).
\]

So the answer to the question is:

\[
\int e^x \cos(x) \, dx = \frac{e^x(\cos(x) + \sin(x))}{2}
\]

**Example 10.5. (Definite integrals.)**Compute \( \int_0^\pi x \sin x \, dx \).

**answer:** Using integration by parts

<table>
<thead>
<tr>
<th>( u )</th>
<th>( v' )</th>
<th>( u' )</th>
<th>( v )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x )</td>
<td>( \sin(x) )</td>
<td>( 1 )</td>
<td>( -\cos(x) )</td>
</tr>
</tbody>
</table>

we get

\[
\int_0^\pi x \sin(x) \, dx = -x \cos(x)|_0^\pi + \int_0^\pi \cos(x) \, dx = \pi + \sin(x)|_0^\pi = \pi.
\]

Note well that both terms from the integration by parts get limits of integration. Remember that integration by parts helps you find the antiderivative and that for a definite integral you have to put the limits on the entire antiderivative.
Inverse trig functions

We know that \( \frac{d}{dx} \sin^{-1} x = \frac{1}{\sqrt{1-x^2}} \). (We’ll remind you how to find this in a minute.)

Using this formula we can use integration by parts to find \( \int \sin^{-1}(x) \, dx \).

Choose

\[
\begin{align*}
    u &= \sin^{-1} x & v &= 1 \\
    u' &= \frac{1}{\sqrt{1-x^2}} & v' &= x
\end{align*}
\]

This gives

\[
\int \sin^{-1}(x) \, dx = x \sin^{-1}(x) - \int \frac{x}{\sqrt{1-x^2}} \, dx = x \sin^{-1}(x) + \sqrt{1-x^2} + C.
\]

Let’s now recall how to **compute the derivative of** \( y = \sin^{-1}(x) \):

We have \( \sin(y) = x \). We can take the derivative of both sides with respect to \( x \) using the chain rule (remember \( y \) is a function of \( x \)) to get \( \cos(y) \frac{dy}{dx} = 1 \). Solving for \( \frac{dy}{dx} \) we get

\[
\frac{dy}{dx} = \frac{1}{\cos(y)} = \frac{1}{\sqrt{1-x^2}}.
\]

![Trig Triangle](image)

The value \( \cos(y) = \sqrt{1-x^2} \) was found using the trig-triangle above.

**A reduction formula.** **Reduction formulas** are general formulas that show how to reduce an integral with a power to a similar one with a smaller power. Here we will show how to find and use the following reduction formula.

\[
\int \sin^n(x) \, dx = -\frac{1}{n} \sin^{n-1}(x) \cos(x) + \frac{n-1}{n} \int \sin^{n-2}(x) \, dx. \tag{3}
\]

Notice that the left-hand side involves an integral with power \( n \) and the right-hand side has an integral with the smaller power \( n-2 \).

Before proving the reduction formula we show how to use it in an example.

**Example 10.6.** Find \( \int \sin^5(x) \, dx \).
answer: We use the reduction formula \(3\) repeatedly.

\[
\int \sin^5(x) \, dx = -\frac{1}{5} \sin^4(x) \cos(x) + \frac{4}{5} \int \sin^3(x) \, dx
\]

\[
= -\frac{1}{5} \sin^4(x) \cos(x) - \frac{4}{5} \cdot \frac{1}{3} \sin^2(x) \cos(x) + \frac{42}{53} \int \sin(x) \, dx
\]

\[
= -\frac{1}{5} \sin^4(x) \cos(x) - \frac{4}{5} \cdot \frac{1}{3} \sin^2(x) \cos(x) - \frac{4}{53} \cos(x) + C.
\]

Proof of \(3\). Using integration by parts choose \(u\) and \(v'\) by

\[
\begin{align*}
u &= \sin^{n-1}(x) & v' &= \sin(x) \\
u' &= (n-1) \sin^{n-2}(x) \cos(x) & v &= -\cos(x)
\end{align*}
\]

This gives

\[
\int \sin^n(x) \, dx = -\sin^{n-1}(x) \cos(x) + (n-1) \int \sin^{n-2}(x) \cos^2(x) \, dx
\]

\[
= -\sin^{n-1}(x) \cos(x) + (n-1) \int \sin^{n-2}(x)(1 - \sin^2(x)) \, dx
\]

\[
= -\sin^{n-1}(x) \cos(x) + (n-1) \int \sin^{n-2}(x) - \sin^n(x) \, dx.
\]

Moving the \(\sin^n(x)\) term on the right side to the left side we get

\[
n \int \sin^n(x) \, dx = -\sin^{n-1}(x) \cos(x) + (n-1) \int \sin^{n-2}(x) \, dx.
\]

Now dividing by \(n\) gives the reduction formula.

10.2 Numerical Integration

Numerical integration refers to a number of different ways to approximate the value of a definite integral. There are several reasons we might want to use numerical integration:

- We can use computers to evaluate the integrals. With modern computers we can easily do a calculation with tens of thousands or even millions of terms. This allows us to make extremely accurate approximations of integrals.

- For many functions there is no closed formula for the antiderivative so we are forced to turn to numerical integration to find the value. For example,

\[
\int_{-\infty}^{5} e^{-x^2} \, dx, \quad \int_{0}^{\pi} \sqrt{\sin(x)} \, dx, \quad \text{and} \quad \int_{0}^{2\pi} \sqrt{1 + \sin^2(x)} \, dx
\]

are all integrals which can’t be computed by finding an antiderivative, so we must do some type of numerical approximation.

Riemann sums. When we first defined definite integrals we introduced the notion of Riemann sums. To make a Riemann sum you divide the interval of integration into \(n\) subintervals and sum the areas of rectangles above each of these subintervals. There are many possible ways to choose the rectangles above each interval:
• **Right Riemann sum (RRS):**

\[
\int_{a}^{b} f(x) \, dx = \text{area} \approx (y_1 + y_2 + \ldots + y_n) \Delta x = \sum_{j=1}^{n} y_j \Delta x.
\]

![Right Riemann sum](image)

Figure showing the right Riemann sum approximating an integral.

For the right Riemann sum the height of a rectangle above an interval is the height of the curve above the right endpoint of the interval. Notice the right Riemann sum starts at \( j = 1 \) and ends at \( j = n \).

• **Left Riemann sum (LRS):**

\[
\int_{a}^{b} f(x) \, dx = \text{area} \approx (y_0 + y_1 + \ldots + y_{n-1}) \Delta x = \sum_{j=0}^{n-1} y_j \Delta x.
\]

![Left Riemann sum](image)

Figure showing the left Riemann sum approximating an integral.

For the left Riemann sum the height of a rectangle above an interval is the height of the curve above the left endpoint of that interval. Notice that the left Riemann sum starts at \( j = 0 \) and ends at \( j = n - 1 \).

• **Mid Riemann sum:** For the mid Riemann sum the height of a rectangle above an interval is the height of the curve above the midpoint of that interval.

• **Maximum Riemann sum:** For the maximum Riemann sum the height of a rectangle above an interval is the height of the highest point on the curve above that interval.

• **Minimum Riemann sum:** For the minimum Riemann sum the height of a rectangle is the height of the lowest point on the curve above that interval.
• **Random Riemann sum:** For the random Riemann sum the height of a rectangle above an interval is the height of a random point on the curve above that interval.

**Example 10.7.** Estimate \( \int_{0}^{1} \sqrt{1-x^3} \, dx \) using right and left Riemann sums.

**answer:** To avoid too much calculation we’ll do this with \( n = 4 \). On a computer we could use a much bigger value of \( n \). Here is one good way to organize the calculation in a table.

<table>
<thead>
<tr>
<th>( j )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_j )</td>
<td>0</td>
<td>0.25</td>
<td>0.5</td>
<td>0.75</td>
<td>1</td>
</tr>
<tr>
<td>( y_j )</td>
<td>1</td>
<td>0.992</td>
<td>0.935</td>
<td>0.760</td>
<td>0</td>
</tr>
</tbody>
</table>

We first compute \( \Delta x = \frac{b-a}{n} = \frac{1}{4} \). Now the two Riemann sums are

- **Right Riemann sum (RRS):** \( (y_1 + y_2 + y_3 + y_4) \Delta x \approx 0.672 \)
- **Left Riemann sum (LRS):** \( (y_0 + y_1 + y_2 + y_3) \Delta x \approx 0.922 \).

Notice that for \( n = 4 \) our two approximations are fairly different. Using a computer we can compute these sums for much larger \( n \):

- with \( n = 100 \): RRS = 0.836, LRS = 0.846.
- with \( n = 500 \): RRS = 0.840, LRS = 0.842.

We can see the values of the estimates converging as \( n \) gets bigger. The “true” value of the integral = 0.8413. (This was computed numerically using Simpson’s rule with \( n = 1.6 \times 10^6 \). We will learn about Simpson’s rule shortly.)

**Trapezoidal Rule:** A more accurate rule than the ones using rectangles is the **trapezoidal rule.** The idea here is to use trapezoids above each interval in place of rectangles. The top of the trapezoid is found by connecting the points the two endpoints on the part of the curve above the interval –see the figure below. Notice how well the trapezoids approximate the area.

![Trapezoidal Rule](image.png)

Figure showing the trapezoidal rule for approximating an integral.

We know that the area of a trapezoid is

\[
\text{base} \times \text{average of the legs}
\]

Summing the area of all \( n \) trapezoids we get

\[
\int_{a}^{b} f(x) \, dx \approx \left( \frac{y_0 + y_1}{2} + \frac{y_1 + y_2}{2} + \cdots + \frac{y_{n-1} + y_n}{2} \right) \Delta x
\]

\[
= \left( \frac{1}{2} y_0 + y_1 + y_2 + \cdots + y_{n-1} + \frac{1}{2} y_n \right) \Delta x.
\]
Another way to describe the trapezoidal rule is that it’s the average of left and right Riemann sums.

**Example 10.8.** (Same integral as above.) Approximate the value of \( \int_{0}^{1} \sqrt{1-x^3} \, dx \) using the trapezoidal rule with \( n = 4 \).

**answer:** We can use the table giving \( y_0, \ldots, y_4 \) from the previous example.

\[
\text{Integral} \approx \left( \frac{1}{2} \cdot 1 + 0.992 + 0.935 + 0.760 + \frac{1}{2} \cdot 0 \right) \cdot 0.25 = 0.797.
\]

With \( n = 30 \), the trapezoidal rule approximates this integral as 0.839. The trapezoidal rule converges to the true value much faster, i.e. with smaller \( n \), than the rectangle rules.

**Simpson’s Rule:** We can get an even more accurate approximation by using parabolic caps on the area above each interval. This gives us Simpson’s rule. You can read the derivation in the textbook.

The rule requires that \( n \) be even. The formula is

\[
\int_{a}^{b} f(x) \, dx \approx \frac{1}{3} \left( y_0 + 4y_1 + 2y_2 + 4y_3 + 2y_4 + \cdots + 4y_{n-1} + y_n \right) \Delta x.
\]

(Note: there are other ways to state Simpson’s rule that don’t require \( n \) to be even. We will be consistent in using the above formulation.)

**Example 10.9.** (Same integral as above.) Approximate \( \int_{0}^{1} \sqrt{1-x^3} \, dx \) using Simpson’s rule with \( n = 4 \).

**answer:** We can once again use the table of values in the previous examples:

\[
\int_{0}^{1} \sqrt{1-x^3} \, dx \approx \frac{1}{3} \left( 1 + 4(0.992) + 2(0.935) + 4(0.760) + 0 \right) \cdot 0.25 = 0.823.
\]

Notice how even with \( n = 4 \) Simpson’s rule gives a pretty good approximation of the true value of 0.843.

With \( n = 30 \): Simpson’s rule gives \( \int_{0}^{1} \sqrt{1-x^3} \approx 0.840 \).

**Approximating sums using integrals.** As we’ve done before we can turn approximation on its head and approximate a sum using an integral.

**Example 10.10.** Use the trapezoidal rule to estimate \( 1^2 + 2^2 + 3^2 + \ldots + 100^2 \). Is the estimate too high or too low?

**answer:** Let \( \Delta x = 1 \). The trapezoidal rules says

\[
\int_{0}^{100} x^2 \, dx \approx \frac{0^2}{2} + 1^2 + 2^2 + \cdots + 99^2 + \frac{100^2}{2}.
\]

(We prefer to use the trapezoidal rule to approximate the integral because it is more accurate.) Now, the above sum is not exactly the sum we want to estimate. This is okay, often in problem solving you can get something close to what you want and then tweak it to get
exactly what you want. In our case adding \( \frac{100^2}{2} \) to the above sum gives exactly the sum we want to approximate. Thus, adding \( \frac{100^2}{2} \) to both sides of the above equation gives

\[
\int_0^{100} x^2 \, dx + \frac{100^2}{2} \approx 1^2 + 2^2 + 3^2 \ldots + 100^2.
\]

The expression on the right is exactly the sum we want to approximate. The expression on the left is easy to compute

\[
\int_0^{100} x^2 \, dx + \frac{100^2}{2} = \frac{10^6}{3} + \frac{10^4}{2} \approx 338333.33.
\]

So our approximation is \( 1^2 + 2^2 + \ldots + 100^2 \approx 338333.33 \). (The exact sum is 338350).

One final note for this particular example: since \( x^2 \) is concave up each of the trapezoids is above the curve. This means the sum is bigger than the integral used to make the estimate, so the estimate is too low.

11 Improper Integrals

Definition: \( \int_a^\infty f(x) \, dx = \lim_{b \to \infty} \int_a^b f(x) \, dx \).

If the limit exits we say the improper integral converges otherwise we say it diverges.

Example 11.1. Compute \( \int_1^\infty \frac{1}{x^2} \, dx \).

**answer:** This is an improper integral, so it equals

\[
\lim_{b \to \infty} -x^{-1}\bigg|_1^b = \lim_{b \to \infty} 1 - 1/b = 1.
\]

Thus, the integral converges to 1.

Example 11.2. Compute \( \int_1^{\infty} \frac{1}{x^{1/2}} \, dx \).

**answer:** This improper integral equals

\[
\lim_{b \to \infty} 2x^{1/2}|_1^b = \lim_{b \to \infty} 2b^{1/2} - 2.
\]

We see this diverges as \( b \to \infty \), so the integral diverges.

Example 11.3. Compute \( \int_1^{\infty} \frac{1}{x^p} \, dx \). (To avoid an annoying case, assume \( p \neq 1 \)).

**answer:** This improper integral equals

\[
\lim_{b \to \infty} \frac{x^{-p+1}}{-p+1}|_1^b = \lim_{b \to \infty} \frac{b^{-p+1}}{-p+1} + \frac{1}{p-1} = \begin{cases} \frac{1}{p-1} & \text{if } p > 1 \\ \infty & \text{if } p < 1. \end{cases}
\]
So the integral converges if \( p > 1 \) and diverges to \( \infty \) if \( p < 1 \).

If \( p = 1 \) it also diverges to \( \infty \).

This gives us the \( p \)-test

\[
\int_1^\infty \frac{1}{x^p} \, dx \to \begin{cases} p > 1 \text{ converges} \\ p \leq 1 \text{ diverges} \end{cases}
\]

You should know this. But realize that you will need it in settings where the integrals are not presented in exactly this form.

**Example 11.4.** Compute the work (energy) needed to get a mass \( m \) to a distance \( x \) from the center of the earth. (Assume \( x > R \)).

\( r = \) distance from center of earth.

\( R = \) radius of earth.

\( F = \frac{GmM}{r^2} = \frac{C}{r^2} = \) force on \( m \) at distance \( r \).

Briefly, using 'slicing and summing' we find that the work (energy) needed to get a mass \( m \) from the surface to a distance \( x \) from the center of the earth (assume \( x > R \)) is

\[
W = \int_R^x \frac{C}{r^2} \, dr = \left[ -\frac{C}{r} \right]_R^x = \frac{C}{R} - \frac{C}{x}.
\]

To escape earth’s gravity we let \( x \to \infty \). This implies \( W \to \frac{C}{R} \), i.e. we need at least \( \frac{C}{R} \) units of energy to escape.

**Example 11.5.** Consider \( f(x) = \frac{1}{x} \) for \( 1 \leq x < \infty \).

The Area under curve = \( \int_1^\infty \frac{1}{x} \, dx = \ln x|_1^\infty = \infty \).

Volume of revolution around \( x \)-axis = \( \int_1^\infty \pi(1/x)^2 \, dx = \pi \) is finite.

When we can compare two functions we have tests relating the convergence of their integrals.

**Comparison test:** Suppose \( 0 \leq f(x) \leq g(x) \)

\[
\int_a^\infty g(x) \, dx \text{ converges } \Rightarrow \int_a^\infty f(x) \, dx \text{ converges.}
\]

\[
\int_a^\infty f(x) \, dx \text{ diverges } \Rightarrow \int_a^\infty g(x) \, dx \text{ diverges.}
\]
Examples:
\[ \int_{1}^{\infty} \frac{1}{\sqrt{x^3 + 1}} \, dx < \int_{1}^{\infty} \frac{1}{x^{3/2}} \, dx \text{ converges.} \]

\[ \int_{0}^{\infty} e^{-x^2} \, dx = \int_{0}^{1} e^{-x^2} \, dx + \int_{1}^{\infty} e^{-x^2} \, dx < \int_{0}^{1} e^{-x^2} \, dx + \int_{1}^{\infty} e^{-x} \, dx \text{ converges.} \]

This is a pain. We can do better.

Limit comparison: Assume \( f, g \) are positive functions.

If \( \lim_{x \to \infty} \frac{f}{g} = 1 \) then both improper integrals converge or both diverge.

Notes:
1. We write \( f \sim g \) and say \( f \) is asymptotic with \( g \).
2. Limit comparison is also called Asymptotic comparison.

Example 11.6. If \( f = \frac{1}{\sqrt{x^3 - 1}} \), \( g = \frac{1}{x^{3/2}} \) then \( \lim_{x \to \infty} \frac{f}{g} = 1 \).

Thus, since \( \int_{2}^{\infty} \frac{1}{x^{3/2}} \, dx \) converges so does \( \int_{2}^{\infty} \frac{1}{\sqrt{x^3 - 1}} \, dx \).

Note: We need the lower limit in the integral to be greater than 1 so that \( g(x) \) is defined on the entire interval we integrate over.

"Proof" of limit comparison:

\( \lim_{x \to \infty} \frac{f}{g} = 1 \Rightarrow \) there is an \( a \) such that \( \frac{1}{2} g(x) < f(x) < 2 g(x) \) for \( x > a \).

Thus \( \frac{1}{2} \int_{a}^{\infty} g(x) \, dx < \int_{a}^{\infty} f(x) \, dx < 2 \int_{a}^{\infty} g(x) \, dx \). This shows both integrals converge or both diverge. QED

Can extend limit comparison?

Consider the integrals \( I_1 = \int_{a}^{\infty} f(x) \, dx \) and \( I_2 = \int_{a}^{\infty} g(x) \, dx \).

Suppose \( \lim_{x \to \infty} \frac{f}{g} = c \).

If \( c \neq 0 \) and \( c \neq \infty \) then both integrals converge or both diverge.

If \( c = 0 \) then \( f \) is asymptotically smaller than \( g \) so \( I_1 \) converges implies \( I_1 \) converges and \( I_1 \) diverges implies \( I_2 \) diverges.

If \( c = \infty \) then \( g \) is asymptotically smaller than \( f \) and similar conclusions hold.

Example 11.7. \( \int_{1}^{\infty} \frac{\ln x}{x^3} \, dx \) converges by asymptotic comparison with \( 1/x^2 \) since \( \lim_{x \to \infty} \frac{\ln x/x^3}{1/x^2} = 0 \).

Other improper integrals:

Example 11.8. \( \int_{0}^{1} \frac{1}{x^{1/3}} \, dx \) This is improper because \( \frac{1}{x^{1/3}} = \infty \) when \( x = 0 \).

We define \( \int_{0}^{1} \frac{1}{x^{1/3}} \, dx = \lim_{a \to 0^+} \int_{a}^{1} \frac{1}{x^{1/3}} \, dx \).

\[ \Rightarrow \text{Integral} = \lim_{a \to 0^+} \left[ \frac{3}{2} x^{2/3} \right]_{a}^{1} = \lim_{a \to 0^+} \left( \frac{3}{2} a^{2/3} - \frac{3}{2} a^{2/3} \right) = \frac{3}{2} \Rightarrow \text{the integral converges.} \]
Here’s a version of the $p$-test in this case (and the previous case).

<table>
<thead>
<tr>
<th>$p$-test: $\int_1^\infty x^p , dx$</th>
<th>$\int_1^1 x^p , dx$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p &lt; -1$ converges</td>
<td>$p &gt; 1$ converges</td>
</tr>
<tr>
<td>$p &gt; -1$ diverges</td>
<td>$p &lt; 1$ converges</td>
</tr>
<tr>
<td>$p \leq -1$ diverges</td>
<td>$p \leq 1$ diverges</td>
</tr>
</tbody>
</table>

Always think about these. It is sometimes written as:

\[
\int_1^\infty \frac{1}{x^p} \, dx \rightarrow \begin{cases} 
  p > 1 & \text{converges} \\
  p \leq 1 & \text{diverges}
\end{cases}
\]

\[
\int_0^1 \frac{1}{x^p} \, dx \rightarrow \begin{cases} 
  p < 1 & \text{converges} \\
  p \geq 1 & \text{diverges}
\end{cases}
\]

Comparison also works for the following types of improper integrals.

**Example 11.9.** \( \int_{0^+}^1 \frac{1}{\sqrt{x(x^2 + 1)}} \, dx \) converges since near \( x = 0 \), \( \frac{1}{\sqrt{2(x^2 + 1)}} \leq \frac{1}{\sqrt{x}} \).

Be careful –at first glance you might think the appropriate \( p \) for comparison is \( p = 3/2 \).

**Example 11.10.** (change of variable)

Does \( \int_{0}^{1^+} \frac{1}{\sqrt{1 - x^3}} \, dx \) converge? (The integral is improper at 1.)

Substitute \( u = 1 - x \Rightarrow 
\[
\text{Integral} = \int_{0^+}^1 \frac{1}{\sqrt{u} \sqrt{u^2 - 3u + 3}} \, du.
\]

Converges by comparison with \( 1/\sqrt{u} \).

## 12 Infinite Series

You should make sure you are comfortable with \( \sum \) notation.

**Definition of series (= sum):**

- \( a_0 + a_1 + a_2 + \ldots + a_n \) (finite series).
- \( a_0 + a_1 + a_2 + \ldots + a_n + \ldots \) (infinite series).

**Sigma notation:** \( \sum_{n=0}^{\infty} a_n = a_0 + a_1 + a_2 + \ldots \)

**Note:** The index \( n \) could be another letter, e.g. \( i, j \). It’s like the \( x \) in an integral.

**Definition.** Convergent series

If \( \lim_{N \to \infty} \sum_{n=0}^{N} a_n = S \) converges we say the infinite series converges to the sum \( S \), otherwise it diverges.

(Note the similarity to the definition of convergent improper integrals.)

**Partial sums:** Occasionally we will want a name for the finite sums.
\[ S_N = a_0 + a_1 + \ldots + a_N = \sum_{n=0}^{N} \] is called the \( N \text{th} \) partial sum.

Convergence can be written as \( \lim_{N \to \infty} S_N = S \).

**Note.** If the limit is \( \infty \) we say the series diverges to \( \infty \).

**Tests for convergence** (just like for improper integrals)

**Comparison test:**
Assume \( 0 \leq f(n) \leq g(n) \)

- If \( \sum_{n=n_0}^{\infty} g(n) \) converges then so does \( \sum_{n=n_0}^{\infty} f(n) \).
- If \( \sum_{n=n_0}^{\infty} f(n) \) diverges then so does \( \sum_{n=n_0}^{\infty} g(n) \).

**Asymptotic comparison test:**
Assume \( a_n, b_n \) are positive.

- If \( \lim_{n \to \infty} \frac{a_n}{b_n} = 1 \) then \( \sum a_n \) and \( \sum b_n \) both converge or both diverge

Just like for integrals, if \( \lim_{n \to \infty} \frac{a_n}{b_n} = c \) and \( c \neq 0 \) or \( c = \infty \) then both series converge or both diverge.

If \( c = 0 \) the implication only goes one way, e.g., if \( \sum b_n \) converges then so does \( \sum a_n \) and if \( \sum a_n \) diverges the so does \( \sum b_n \). Similar statements hold if \( c = \infty \).

**Integral Test:**
If \( f(x) \) is decreasing and \( \lim_{x \to \infty} f(x) = 0 \) then

- \( \sum_{n=n_0}^{\infty} f(n) \) and \( \int_{a}^{\infty} f(x) \, dx \) either both converge or both diverge.

N.B. the hypotheses that \( f(x) \) is decreasing and goes to 0.

**p-Test:**
\[ \sum_{n=1}^{\infty} \frac{1}{n^p} \text{ converges if } p > 1 \text{ and diverges if } p \leq 1. \]

**Proof** of the asymptotic comparison test: For large \( n \) (say for \( n \geq N \)) \( \frac{a_n}{b_n} \approx c \)

\[ \Rightarrow \text{ for large } n: \quad \frac{1}{2} c b_n < a_n < 2 c b_n. \]

So \( \sum_{n=N}^{\infty} a_n \text{ converges } \Rightarrow \sum_{n=N}^{\infty} \frac{1}{2} c b_n \text{ also converges } \Rightarrow \sum_{n=N}^{\infty} b_n \text{ converges} \).

Likewise \( \sum_{n=N}^{\infty} b_n \text{ converges } \Rightarrow \sum_{n=N}^{\infty} 2c b_n \text{ converges } \Rightarrow \sum_{n=N}^{\infty} a_n \text{ converges} \).

Note: \( \sum_{n=N}^{\infty} a_n \text{ converges } \iff \sum_{n=1}^{\infty} a_n \text{ converges} \).
**Proof** of the integral test:

Left Riemann sum = \( \sum_{n=n_0}^{\infty} f(n) > \int_{n_0}^{\infty} f(x) \, dx > \) right Riemann sum = \( \sum_{n=n_0+1}^{\infty} f(n). \)

(Probably won’t get to this in class – see proof 1 for harmonic series.)

**Proof** \( p \)-test: Use the integral test with \( \int_1^{\infty} \frac{1}{x^p} \, dx. \)

**Example 12.1. Harmonic series** (this is important)

The harmonic series is defined as \( \sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \ldots \)

**Claim:** This diverges to \( \infty. \)

**Proof 1** (integral test): We could appeal to the integral test, but we’ll set it up so you can see why the integral test is true.

\[
\sum_{n=1}^{\infty} \frac{1}{n} > \int_1^{\infty} \frac{1}{x} \, dx = \ln x \bigg|_1^{\infty} = \infty.
\]

(The inequality follows because the sum is a left Riemann sum that overestimates the area under the curve.)

![Left Riem. sum overest. integral](image1)

![Right Riem. sum underest. integral](image2)

**Proof 2:** (This is also in the book so won’t do in class.)

The sum = \((1) + (\frac{1}{2}) + (\frac{1}{3} + \frac{1}{4}) + (\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}) + \ldots \)

Each grouping of terms is greater than \( \frac{1}{2} \), for instance the one starting with \( \frac{1}{5} \) has 4 terms all bigger than \( \frac{1}{8} \). Continuing by taking twice as many terms in each successive group produces an infinite sum of groups each greater than \( \frac{1}{2} \). This implies the sum diverges to \( \infty. \)

**Example 12.2.** Show \( \sum_{n=1}^{\infty} \frac{1}{n^2} = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \ldots \) converges.

**answer:** The \( p \)-test (with \( p = 2 > 1 \)) shows this series converges.

**Example 12.3.** Does \( \sum_{n=2}^{\infty} \frac{1}{n \ln n} \) converge?

**answer:** First, \( \frac{1}{x \ln x} \) is decreasing.

\[
\int_2^{\infty} \frac{1}{x \ln x} \, dx = \int \frac{1}{\ln x} \, d(\ln x) = \ln \ln x \bigg|_2^{\infty} = \infty.
\]
So the integral test shows the series diverges.

**Example 12.4.** \( \sum \frac{1}{n^2+1} < \sum \frac{1}{n^2} \) converges by the \( p \)-test.

**Examples:** Do the following converge or diverge?

1. \( \sum \frac{2}{n^2 + n} \).

   **answer:** Converges – compare with \( \sum \frac{1}{n^2} \).

2. \( \sum \frac{n^2 + 3}{1000n^3} \).

   **answer:** Diverges – asymptotically compare with \( \sum \frac{1}{n} \).

3. \( \sum \frac{1}{(n+3)^2} \).

   Converges – compare with \( \sum \frac{1}{n^2} \).

4. \( \sum \frac{n}{\sqrt{n^2 + 2}} \).

   **answer:** Diverges – asymptotically compare with \( \sum 1 \).

5. \( \sum \frac{\tan^{-1} n}{n^3} \).

   **answer:** Converges – asymptotically compare with \( \sum \frac{1}{n^3} \) – recall \( \tan^{-1} x \) is bounded between \( -\pi/2 \) and \( \pi/2 \).

**Theorem** \( \sum a_n \) converges \( \Rightarrow \lim_{n \to \infty} a_n \to 0 \)

**Proof.**

\[
S_{n+1} = S_n + a_n \\
\downarrow \quad \quad \downarrow \quad \quad \downarrow \\
S \quad \quad S \quad \quad 0
\]

Note: The converse of this is false, i.e. if \( a_n \) goes to 0 the series might diverge. For example, the harmonic series has this property.

**Example 12.5.** (Telescoping series) Consider \( \sum \frac{1}{n(n+1)} \). Show it converges and find its sum.

**answer:** It converges by comparison to \( \sum \frac{1}{n^2} \).

In this case we can actually compute the sum: \( \frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1} \). We call this a telescoping series because the partial sum can be found as

\[ S_N = (1 - \frac{1}{2}) + (\frac{1}{2} - \frac{1}{3}) + \ldots + (\frac{1}{N} - \frac{1}{N+1}) = 1 - \frac{1}{N+1}. \]

It is clear \( \lim_{N \to \infty} S_N = 1 \), i.e. the series converges to 1.
13 Geometric Series, Power Series, Ratio Test

Geometric Series
A geometric series has the form

\[ 1 + r + r^2 + r^3 + \ldots = \sum_{n=0}^{\infty} r^n \]

Or more generally

\[ a + ar + ar^2 + ar^3 + \ldots = \sum_{n=0}^{\infty} ar^n \]

We call the number \( r \) the ratio of the geometric series. That is, you multiply each term by \( r \) to find the next term in the series.

**Punch line.** The geometric series

\[ 1 + r + r^2 + r^3 + \ldots \begin{cases} \text{converges to} & \frac{1}{1-r} \quad \text{if } |r| < 1 \\ \text{diverges} & \text{if } |r| \geq 1. \end{cases} \]

**Example 13.1.** Find \( \sum_{n=0}^{\infty} \left( \frac{1}{2} \right)^n \).

**answer:** Writing the sum out we see it is a geometric series with ratio \( r = 1/2 \).

\[ 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots = \frac{1}{1-1/2} = 2 \]

**Example 13.2.** Compute \( \sum_{n=0}^{\infty} \left( -\frac{1}{3} \right)^n \).

**answer:** This is a geometric series with ratio \( 1/3 \).

\[ 1 - \frac{1}{3} + \frac{1}{9} - \frac{1}{27} + \ldots = \frac{1}{1-(1/3)} = \frac{3}{4} \]

**Example 13.3.** Does \( \sum_{n=1}^{\infty} \frac{1}{(1 + (n-1)/n)^n} \) converge?

**answer:** This is not a geometric series, but we can compare it asymptotically with the geometric series \( \sum 2^{-n} \). Since this converges, so does the series in question.

(If we wanted to we could compare the series directly to \( \sum 1.5^{-n} \).)

**Variations**

\[ \sum_{n=5}^{\infty} \left( \frac{3}{4} \right)^n = \left( \frac{3}{4} \right)^5 + \left( \frac{3}{4} \right)^6 + \left( \frac{3}{4} \right)^7 + \ldots = \left( \frac{3}{4} \right)^5 \left( 1 + \frac{3}{4} + \left( \frac{3}{4} \right)^2 + \ldots \right) \]

\[ = \left( \frac{3}{4} \right)^5 \left( \frac{1}{1-3/4} \right) = \left( \frac{3}{4} \right)^5 \cdot 4. \]
\[ \sum_{n=2}^{\infty} \pi r^n = \pi r^2 (1 + r + r^2 + \ldots) = \pi r^2 \left( \frac{1}{1-r} \right) \text{ (provided } |r| < 1) . \]

\[ \sum_{n=0}^{\infty} \left( \frac{1}{2} \right)^{2n} = \sum_{n=0}^{\infty} \left( \frac{1}{4} \right)^n = \frac{4}{3} . \]

**Proof (of the sum of geometric series)**

Using the \( N \)th partial sum we have

\[ S_N = 1 + r + r^2 + \ldots + r^N \]
\[ rS_N = r + r^2 + r^3 + \ldots + r^{N+1} \]

Subtract the second equation from the first to get

\[(1 - r)S_N = 1 - r^{N+1} \]

Thus, \( S_N = \frac{1 - r^{N+1}}{1 - r} \).

Taking limits: \( \lim_{N \to \infty} S_N = \begin{cases} \frac{1}{1-r} & \text{if } |r| < 1 \\ \infty & \text{if } r \geq 1 \\ \text{doesn't exist} & \text{if } r \leq -1 \end{cases} \)

**Example 13.4.** Find a closed form for \( f(x) = x + x^3/3 + x^5/5 + \ldots \)

**answer:** Taking derivatives: \( f'(x) = 1 + x^2 + x^4 + x^6 + \ldots = 1/(1-x^2) \)

Integrating: \( f(x) = \int \frac{1}{1-x^2} \, dx \).

Using partial fractions) \( \frac{1}{1-x^2} = \frac{1/2}{1-x} + \frac{1/2}{1+x} \).

Thus \( f(x) = -\frac{1}{2} \ln(|1-x|) + \frac{1}{2} \ln(|1+x|) = \ln \sqrt{\frac{|1+x|}{|1-x|}} + C \).

We find the value of \( C \) by letting \( x = 0: f(0) = 0 \Rightarrow C = 0 \). Thus,

\[ f(x) = \ln \sqrt{\frac{|1+x|}{|1-x|}} \]

**Ratio Test.** Consider the series \( \sum a_n \). Assume: \( a_n > 0 \) for all \( n \).

Define \( L = \lim_{n \to \infty} \frac{a_{n+1}}{a_n} \).

If \( L < 1 \) then \( \sum a_n \) converges.
If \( L > 1 \) then \( \sum a_n \) diverges.
If \( L = 1 \) then the test fails.
If there is no limit \( L \) the the test fails.

**Example 13.5.** Does \( \sum_{1}^{\infty} \frac{2^n}{n^3} \) converge or diverge?
**answer:** The terms are all positive so we can try the ratio test:

\[
L = \lim_{n \to \infty} \frac{2^{n+1}/(n+1)^3}{2^n/n^3} = \lim_{n \to \infty} 2 \frac{n^3}{(n+1)^3} = 2
\]

Since \( L > 1 \), the series diverges.

**Example 13.6.** Does \( \sum_{1}^{\infty} \frac{2^n}{n!} \) converge or diverge?

**answer:** The terms are all positive so we can try the ratio test:

\[
L = \lim_{n \to \infty} \frac{2^{n+1}/(n+1)!}{2^n/n!} = \lim_{n \to \infty} 2/(n + 1) = 0.
\]

Since \( L < 1 \), the series converges.

**Example 13.7.** Does \( \sum_{1}^{\infty} \frac{1}{n^2} \) converge or diverge?

**answer:** The terms are all positive so we can try the ratio test:

\[
L = \lim_{n \to \infty} \frac{n^2}{(n + 1)^2} = 1.
\]

Since \( L = 1 \) the ratio test fails.

In this case, we know the series converges using the \( p \)-test or the integral test.

**Absolute convergence.** Series of all positive terms are better behaved than other series. As you add in more terms the series keeps growing. Either the sum reaches a limit or else it grows to infinity.

We can make any series one with positive terms by taking absolute values. This leads to the definition of absolute convergence.

Given a series \( \sum a_n \), if the ‘absolute series’ \( \sum_{n=0}^{\infty} |a_n| \) converges then we say that the series \( \sum a_n \) converges absolutely.

If a series converges but does not converge absolutely then we say it is conditionally convergent.

**Theorem.** If the series \( \sum_{n=0}^{\infty} a_n \) converges absolutely then the series \( \sum a_n \) converges in the usual sense.

**Power Series Convergence.**

Given a power (Taylor) series \( \sum_{n=0}^{\infty} a_n x^n \) we are often interested in knowing for what values of \( x \) the series converges. We can often use the following theorem to simplify matters.

**Theorem.** If \( \sum_{n=0}^{\infty} |a_n x^n| \) converges then so does \( \sum_{n=0}^{\infty} a_n x^n \).
Proof. This is just the previous theorem which says that if the series converges absolutely, then it converges in the usual sense.

Ratio Test examples for power series:

Example 13.8. Find all values of $x$ for which \( \sum_{n=1}^{\infty} \frac{x^n}{n} \) converges absolutely.

**answer:** Taking the absolute ratio of successive terms: \( \frac{|x^{n+1}/(n+1)|}{|x^n/n|} = |x| \cdot \frac{n}{n+1} \).

Taking the limit: \( L = \lim_{n \to \infty} \frac{|x| \cdot n}{n+1} = |x| \).

By the ratio test we know the series converges for \( |x| < 1 \) and diverges for \( |x| > 1 \).
(For \( |x| = 1 \) we need to look more carefully.)

Example 13.9. Find all values of $x$ for which \( \sum_{n=1}^{\infty} \frac{x^n}{n!} \) converges absolutely.

**answer:** Taking the ratio we have \( \frac{|x^{n+1}/(n+1)!|}{|x^n/n!|} = |x| \cdot \frac{1}{n+1} \).

Taking the limit: \( L = \lim_{n \to \infty} \frac{|x| \cdot 1}{n+1} = 0. \)

Since this is less than 1 for all $x$, the series converges absolutely for all $x$.

Example 13.10. For which $x$ does \( \sum_{n=1}^{\infty} \frac{x^{2n+1}}{2^n n} \) converge absolutely?

**answer:** The ratio is \( \frac{|x^{2n+3}/(2^{n+1}(n+1))|}{|x^{2n+1}/(2^n n)|} = |x|^2 \cdot \frac{n}{2(n+1)} \).

Taking the limit: \( L = \lim_{n \to \infty} \frac{|x|^2 \cdot n}{2(n+1)} = \frac{x^2}{2} \).

So, the series converges absolutely when \( \frac{x^2}{2} < 1 \) i.e. for \( |x| < \sqrt{2} \).

Alternating series test.
If \( a_n \) alternates sign and \( |a_n| \downarrow 0 \) then \( \sum a_n \) converges.

Example 13.11. (Alternating harmonic series) Show that the alternating harmonic series \( \sum_{n=1}^{\infty} \frac{(-1)^n}{n} = 1 - 1/2 + 1/3 - 1/4 + \ldots \) converges conditionally.

**answer:** Clearly the terms alternate in sign and go to 0, so the series converges. To show it converges conditionally, we have to show that it does not converge absolutely. The absolute series is \[ 1 + 1/2 + 1/3 + 1/4 + \ldots \]

This is the harmonic series, which we know diverges. QED.

**Note.** It is possible to show the alternating harmonic series converges to \( \ln(2)! \) (See the text by Simmons: bottom page 457 through example 3.)
14 Probability: Discrete Random Variables

Read the supplementary notes

14.1 Introduction

We will view probability as dealing with repeatable experiments such as flipping a coin, rolling a die or measuring a distance. Anytime there is some uncertainty as to the outcome of an experiment probability has a role to play. Gambling, polling, measuring are typical places where probability is used. In general polling and measuring involve analyzing data which is usually called statistics. In this sense statistics is just the art of applied probability.

14.2 Random outcomes and probability functions

Many ‘experiments’ have a finite number of possible outcomes each with a probability of occurring. It is often possible to list the set of all possible outcomes and give their probabilities.

The fair coin. One of the main examples used in probability is a fair coin. When we toss a fair coin we have the following:

Set of possible outcomes = \{heads, tails\}

Probability of heads = 1/2 = probability of tails.

This seems clear, though it is surprisingly difficult to make physical sense of the notion of probability. One standard interpretation is that if we flip a fair coin many times we expect that close to 1/2 of the flips will land heads.

Naturally there is a notation for probability, e.g.

\[ P(\text{heads}) = \frac{1}{2} \]

Example 14.1. Roll 1 six-sided die. What are the possible outcomes. If the die is fair what is the probability of each of these outcomes.

**answer:** Set of possible outcomes = \{1, 2, 3, 4, 5, 6\}, with \( P(j) = \frac{1}{6} \) for \( j = 1, \ldots, 6 \). That is, if we roll a die many times we expect close to 1/6 of the outcomes to be a 1 (or a 2, 3, 4, 5, 6).

Example 14.2. Roll 2 fair dice. Describe the possible outcomes and their probabilities.

**answer:** There are several choices for describing the outcomes.

1. **Ordered pairs.** Ordered pairs = \{(1,1), (1,2), \ldots, (6,6)\}. This means that we distinguish the dice, e.g. color one red and the other blue, and the outcome (2,5) means the first die lands 2 and the second lands 5. All of these 36 possible outcomes are equally likely so \( P(i, j) = \frac{1}{36} \) for any of the pairs.

2. **Total.** We can also describe the outcome of a roll as the total on the two dice. In this case the possible outcomes are:

\[ \text{Outcomes} = \{2, 3, 4, \ldots, 12\} \]
The outcomes aren’t equally likely, for example we see there is only one way to get a total of two, i.e. roll a (1,1). Therefore \( P(2) = P(1, 1) = \frac{1}{36} \).

There are two ways to get a total of 3, i.e. (1,2) and (2,1). Since each of these has probability 1/36 we have \( P(3) = P(1, 2) + P(2, 1) = \frac{2}{36} \). Proceeding in this way we get the following table:

<table>
<thead>
<tr>
<th>Total of two dice: ( x )</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
</table>

Example 14.3. Suppose that \( p \) is the fraction of voters who support candidate A. Ask a random voter if they support candidate A. Assume everyone answers yes or no. Give the possible outcomes and make a table of outcomes and probabilities:

**answer:** The possible outcomes are yes and no. The table of probabilities is:

<table>
<thead>
<tr>
<th>Outcome</th>
<th>yes</th>
<th>no</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>( p )</td>
<td>( 1-p )</td>
</tr>
</tbody>
</table>

**Note:** In all the above examples we carefully state how to run the repeatable experiment.

**General terminology.**

Suppose there are \( n \) possible outcomes, call them \( x_1, x_2, \ldots, x_n \). Then we can list the outcomes and probabilities in a table:

<table>
<thead>
<tr>
<th>Outcome</th>
<th>( x_1 )</th>
<th>( x_2 )</th>
<th>\ldots</th>
<th>( x_n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>( P(x_1) )</td>
<td>( P(x_2) )</td>
<td>\ldots</td>
<td>( P(x_n) )</td>
</tr>
</tbody>
</table>

There are a few standard names for the things we have been discussing:

- **Sample space** = set of possible outcomes = \( \{ x_1, x_2, \ldots, x_n \} \).
- **Probability function**, \( P(x_j) \) = probability of outcome \( x_j \).
- **Trial** = one run of the 'experiment'.
- **Independent**: Two trials are independent if their outcomes have no effect on each other. E.g., repeated flips of a coin are independent.

### 14.3 Probability Laws

There are a rules for probability that we used above without making them explicit. We formulate them first in words and then in symbols:

In words:
(i) Probabilities are between 0 and 1, i.e., the fraction of the time that any one outcome occurs must be between 0 and 1.
(ii) The total probability of all outcomes is 1, i.e. the probability that one of the possible outcomes occurs is 100%.
(iii) The probability that one of several outcomes occurs is just the sum of their probabilities.

More precisely, using symbols:
Suppose there are \( n \) possible outcomes and the sample space (set of all possible outcomes)
is

sample space \( S = \{x_1, x_2, \ldots, x_n\} \)

Then the probability function \( P(x_i) \) must satisfy the following three properties

(i) For each outcome \( x_i \) we have \( 0 \leq P(x_i) \leq 1 \) (probability is between 0 and 1)

(ii) \( \sum_{j=1}^{n} P(x_j) = 1 \) (total probability is 1).

(iii) \( P(x_1 \text{ or } x_2 \text{ or } x_3) = P(x_1) + P(x_2) + P(x_3) \) (probabilities add)

Example 14.4. (a) Roll two dice. Let \( A = '\text{the total is } < 4' \). What is \( P(A) \)?

answer: We know \( A = \{2, 3\} \), that is, the ‘total is less than 4’ means that the total is either 2 or 3. Using the table for the probabilities of the sum of two dice given in an earlier example we get

\[ P(A) = P(2 \text{ or } 3) = P(2) + P(3) = \frac{1}{36} + \frac{2}{36} = \frac{3}{36}. \]

(b) Let \( B = '\text{the total is odd}' \), so \( B = \{3, 5, 7, 9, 11\} \). Find \( P(B) \).

answer: \( P(B) = P(3) + P(5) + P(7) + P(9) + P(11) = \frac{2}{36} + \frac{4}{36} + \frac{6}{36} + \frac{4}{36} + \frac{2}{36} = \frac{1}{2} \).

14.4 Independence and the multiplication law

We say two random events are independent if the outcome of one does not have any effect on the outcome of the other. For example, it is reasonable to assume that the result of the first roll of two dice has no effect on the result of a second roll.

Multiplication law: Suppose we run two independent trials of an experiment and we get the outcome \( x_i \) for the first trial and \( x_j \) for the second, then the probability of the combined result is

\[ P(x_i \text{ on the first trial and } x_j \text{ on the second trial}) = P(x_i) \cdot P(x_j). \]

Example 14.5. It is reasonable to assume that different tosses of a coin are independent. Suppose we toss a fair coin twice. Describe the outcomes and give their probabilities.

answer: The following notation for the outcomes is common:

\( HH \): heads on both tosses
\( HT \): heads on the first and tails on the second toss
\( TH \): tails on the first toss and heads on the second
\( TT \): tails on both tosses.

Using the multiplication law we get

\[ P(HH) = P(H)P(H) = \frac{1}{4}; \quad P(HT) = P(H)P(T) = \frac{1}{4}; \quad P(TH) = \frac{1}{4}; \quad P(TT) = \frac{1}{4}. \]

That is, all 4 outcomes have equal probability.

Example 14.6. Toss a fair die 3 times, what is the probability of getting an odd number each time?

answer: Let \( A = \{1, 3, 5\} \). On any one toss \( P(A) = \frac{1}{2} \). Since repeated tosses are independent \( P(A \text{ then } A \text{ then } A) = P(A) \cdot P(A) \cdot P(A) = \frac{1}{8}. \)
14.5 Discrete random variables

When the outcomes of an experiment are numbers we have a random variable. More precisely: a **finite random variable** \( X \) consists of

(i) A finite list \( x_1, x_2, \ldots, x_n \) of values \( X \) can take.

(ii) A probability function \( P(x_j) \)

**Examples:**

(i) Roll a die, let \( X \) = number of spots.

(ii) Roll a die, let \( Y = (\text{number of spots})^2 \).

(iii) Toss a coin, let \( X = 1 \) if the result is heads and \( X = 0 \) if the result is tails.

There is no reason a random variable has to take only a finite number of values. If it has an infinite number of values that can be listed we call it an **infinite discrete random variable**. That is, \( X \) is an infinite discrete random variable if:

(i) \( X \) takes values \( x_1, x_2, \ldots \).

(ii) There is a probability function \( P(x_i) \) that satisfies the probability laws given above.

Later we will look at so called continuous random variables. These take values in an entire interval, e.g. \([0, 1]\) or \([0, \infty)\).

14.6 Expectation

The **expectation**, (also called **average**, **mean** or **expected value**) of the finite random variable \( X \) is defined by

\[
E(X) = x_1 P(x_1) + \ldots + x_n P(x_n) = \sum_{i=1}^{n} x_i P(x_i).
\]

**Example 14.7.** Roll a die. Define the random variable \( X = \text{number of spots} \). Compute the expected value of \( X \).

**answer:** \( E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{21}{6} = 3.5 \).

**Interpretation of expected value**

1. The expected value is the expected average over a lot of trials. To understand this, suppose I roll a die, and I pay you $1 per spot then over a lot of rolls, say 6000, I would expect to roll (about) 1000 ones, 1000 twos, etc. In that case I would pay you a total of

\[
1000 \cdot 1 + 1000 \cdot 2 + 1000 \cdot 3 + 1000 \cdot 4 + 1000 \cdot 5 + 1000 \cdot 6 \text{ dollars}
\]

So the average payment per roll would be

\[
\frac{1000 \cdot 1 + 1000 \cdot 2 + 1000 \cdot 3 + 1000 \cdot 4 + 1000 \cdot 5 + 1000 \cdot 6}{6000} = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = \frac{21}{6} = 3.5
\]

Notice the formula for the average above is exactly the formula we used to define expected value of a random variable.

Note, you would never be paid $3.5 on any one turn, it is the expected average over many turns.
Also note, expectation is a weighted average like center of mass.

**Concept question:** If I pay you $1 per spot how much would you be willing to pay to roll the die?

**Example 14.8.** Roll a die, you win $5 if it’s a 1 and lose $2 if it’s not. Model this with a random variable and its expected value.

**answer:** In detail here are the steps we need.

Define the random variable: Let $X$ the amount you win or lose on a given roll. That is, $X$ takes the values 5, -2.

Give the probability function:

$P(X = 5) = P(\text{roll a 1}) = 1/6, \quad P(X = -2) = P(\text{roll 2 to 6}) = 5/6.$

(Or in a table):

<table>
<thead>
<tr>
<th>Payment $x$</th>
<th>5</th>
<th>-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(x)$</td>
<td>1/6</td>
<td>5/6</td>
</tr>
</tbody>
</table>

Compute the expected value: $E(X) = P(5) \cdot 5 + P(-2) \cdot (-2) = \frac{5}{6} - \frac{10}{6} = -\frac{5}{6}.$

**Concept question:** Is the above bet a good one?

**Example 14.9.** (From supplementary notes) A trial consists of tossing a fair coin until it comes up heads. Let $X = \text{number of tosses.}$ Verify that the total probability of all outcomes is 1.

**answer:** First we show all the outcomes and their probabilities in a table.

<table>
<thead>
<tr>
<th>$n$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>toss pattern</td>
<td>H</td>
<td>TH</td>
<td>TTH</td>
<td>TTH</td>
<td>...</td>
</tr>
<tr>
<td>$P(X = n)$</td>
<td>1/2</td>
<td>1/4</td>
<td>1/8</td>
<td>1/16</td>
<td>...</td>
</tr>
</tbody>
</table>

So, the total probability $= \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots = 1.$ (This is a geometric series.)

**Example 14.10.** If I paid you $n for a trial of length $n$, what would you pay to take a turn?

**answer:** We compute the expectation.

$E(X) = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{8} + \ldots = 2.$

(The supplementary notes give a nice method for finding this sum.) Assuming you can play the game many times, for any price less than $2 you would expect to win a positive amount on average.

### 14.7 Poisson random variable with parameter $m$

There are lots of random variables that model various situations. One of them is the Poisson random variable. To describe it we need to give the possible values, their probabilities, and we should say what it models.

Let $X$ be a **Poisson random variable** with parameter $m$. 
(i) Values: \( X \) takes values 0, 1, 2, 3, \ldots

(ii) Probability: \( P(X = k) = e^{-m} \frac{m^k}{k!} \). (The factor \( e^{-m} \) is chosen to give total probability 1.)

(iii) Model: A Poisson random variable models the number of times a low probability event occurs in a given time interval.

Examples: A Poisson random variable can be used to model the following counts.

(i) Number of defects in a manufacturing process in a day.

(ii) Number of errors in data transmission.

(iii) Number of cars in an hour at a rural tollbooth.

(iv) Number of chocolate chips in a cookie.

**Theorem.** \( E(X) = m \).

**Proof.**

\[
E(X) = \sum_{k=0}^{\infty} k \cdot P(X = k) = \sum_{k=0}^{\infty} ke^{-m} \frac{m^k}{k!}
\]

\[
= e^{-m} \left( m + \frac{m^2}{1!} + \frac{m^3}{2!} + \ldots \right)
\]

\[
= me^{-m} \left( 1 + \frac{m}{1!} + \frac{m^2}{2!} + \ldots \right)
\]

\[
= me^{-m} e^m
\]

\[
= m \quad \text{QED.}
\]

(To get from the second to last line to the last line, you need to remember the Taylor series for \( e^m \).)

**Example 14.11.** A manufacturer of widgets knows that 1/200 will be defective.

(a) What is the probability that a box of 144 contains no defective widgets?

**Answer:** Since a single widget has a 1/200 chance of being defective, the expected number of defective widgets in a box is 144/200. Since being defective is rare we can model the number of defective widgets in a box as a Poisson random variable \( X \) with mean \( m = 144/200 \).

The problem asks us to find \( P(X = 0) \). We know this is \( P(X = 0) = e^{-m} \frac{m^0}{0!} = e^{-144/200} = 0.487 \).

(b) What is the probability that more than 2 are defective?

**Answer:** This problem asks us to find \( P(X > 2) \). We do this using the formula \( P(X > 2) = 1 - P(X = 0, 1, 2) \).

\[
1 - P(X = 0, 1, 2) = 1 - e^{-m}(1 + m + \frac{m^2}{2!}) = 0.037.
\]

### 14.8 Histograms

If we ran many trials and made a histogram of the percentage of each outcome, the result should look like the graph of \( P(x_j) \) vs. \( x_j \).
Or we make a histogram of the count for each outcome. The histograms at right give examples.

Histogram of percentages (Poisson distr. with $\mu = 4.5$)

### 15 Probability: Continuous Random Variables

#### Read the supplementary notes

Continuous random variables take a continuum of values, e.g. all real numbers, or all positive real numbers.

**Example 15.1.** Let $X$ be the waiting time between requests at a telephone switch. Unlike our previous examples $X$ can take any positive value—it is continuous.

This is nice, we can apply the calculus we know. For instance, instead of sums we get to take integrals. This is part of our motivation for studying probability in 18.01A.

**Definition:** A **continuous random variable** $X$ is a variable that takes values $x$ together with a **probability density function** $f(x)$ such that

(i) $X$ takes values in the range $(c, d)$ (here $c$ can be $-\infty$ and $d$ can be $\infty$). (The endpoints may or may not be included.)

(ii) $f(x) \geq 0$.

(iii) $\int_c^d f(x) \, dx = 1$.

(iv) $P(a \leq X \leq b) = \int_a^b f(x) \, dx$.

The interpretation of these rules are similar to those for discrete random variables.

(i) $X$ takes values depending on the outcome of some random experiment.

(ii) This is a little tricky. By probability density we mean probability per unit of $x$ (whatever $x$ happens to be). Since probabilities are nonnegative the same must be true of probability densities. –See (iii).

(iii), (iv) Recall that for discrete random variables we summed the probabilities over all possible outcomes to get total probability is 1. For continuous functions we replace the sum by an integrals.

Let’s make an analogy. Suppose we have a length of board that runs between $x = 0$ and $x = b$. Also suppose it has a variable density $\delta(x)$ in units of kg/meter. The mass of the board between $x$ and $x + dx$ is $\delta(x) \, dx$. As usual, to get the total mass of the we need to
integrate
\[ \text{mass} = \int_0^b \delta(x) \, dx. \]

Probability works the same way. Since \( f(x) \) is a density in units of probability/unit of \( x \), the probability between \( X = x \) and \( X = x + dx \) is \( f(x) \, dx \). Naturally, the total probability between \( X = a \) and \( X = b \) is
\[ P(a \leq X \leq b) = \int_a^b f(x) \, dx. \]

Likewise the total probability over all possible values of \( X \) is
\[ \text{Total probability} = P(c \leq X \leq d) = \int_c^d f(x) \, dx = 1. \]

### 15.1 Exponential distribution with mean \( m \)

As you might guess, there are a lot of continuous random variables with names that model a variety of situations.

An exponential random variable \( X \) with mean \( m \) has the following probabilities.

(i) Values: The range of values of \( X \) is \([0, \infty)\).

(ii) Probability density: \( f(x) = \frac{e^{-x/m}}{m} \), where \( m > 0 \).

(iii) Exponential distributions are used to model waiting times.

Soon we will define and learn to compute the following for a continuous random variable. In the meantime we give their values for \( X \).

(iv) Expected value: \( E(X) = m \)

(v) Variance: \( \sigma^2(X) = m^2 \).

(vi) Cumulative distribution function: \( F(x) = 1 - e^{-x/m} \).

**Example 15.2.** Let \( X \) be an exponential random variable with mean \( m \). Show the density defined above has total probability 1.

**answer:** The range is \([0, \infty)\) and the density is \( e^{-x/m} / m \). So
\[
\text{Total probability} = \int_0^\infty f(x) \, dx = \int_0^\infty \frac{e^{-x/m}}{m} \, dx = -e^{x/m} \bigg|_0^\infty = 1
\]

**Example 15.3.** Let \( X \) be an exponential random variable with mean \( m = 1 \). Find \( P(0 \leq X \leq 1) \).
answer: \( P(0 \leq X \leq 1) = \int_0^1 e^{-x} \, dx = 1 - e^{-1} = 0.632. \)

Example 15.4. (a) Same question if \( m = 10. \)

answer: \( P(0 \leq X \leq 1) = \int_0^1 \frac{e^{-x/10}}{10} \, dx = 1 - e^{-1/10} = 0.095. \)

(b) What is \( P(X \geq 60)? \)

answer: \( P(60 \leq X) = \int_{60}^{\infty} \frac{e^{-x/10}}{10} \, dx = e^{-60/10} = 0.002. \)

15.2 Uniform random variables with range \([c, d]\)

The simplest continuous random variable is uniform, i.e. the density is constant. An uniform random variable \( X \) with range \([c, d]\) has the following probabilities.

(i) Values: The range of values is \([c, d]\).

(ii) Density: \( f(x) = \frac{1}{d-c}. \)

(iii) Expected value: \( E(X) = \frac{c + d}{2} \)

(iv) Variance: \( \sigma^2(X) = \frac{(d-c)^2}{2}. \)

Example 15.5. Check that the total probability is 1.

answer: Total probability = \( \int_c^d f(x) \, dx = \int_c^d \frac{1}{d-c} \, dx = 1 \) (easy).

Uniform density on \([0, 2]\). Shaded area = \( P(a \leq X \leq b) = \frac{b-a}{2} \)

15.3 Expected value

For a continuous random variable the expected value has the same interpretation as for a discrete random variable. That is, it is the ‘expected’ average of a large number of independent trials.

Recall, for a discrete variable the expected value is the sum over all values \( x \) of \( x \cdot P(X = x). \)

For a continuous random variable we replace \( P(X = x) \) by \( f(x) \, dx \), i.e. the probability \( X \) is in the range \( x \) to \( x + dx \). This gives the definition.

Definition. For a random variable \( X \) with probability density \( f(x) \) and range \([c, d]\), the
expected value of $X$ is

$$E(X) = \int_c^d x f(x) \, dx$$

Example 15.6. Suppose $X$ a uniform random variable on $[c, d]$. Show that the expected value is $\frac{d+c}{2}$.

answer: The range is $[c, d]$ and the density is $f(x) = \frac{1}{d-c}$. So,

$$E(X) = \int_c^d x f(x) = \int_c^d x \left( \frac{1}{d-c} \right) \, dx = \left[ \frac{x^2}{2(d-c)} \right]_c^d = \frac{d^2 - c^2}{2(d-c)} = \frac{d+c}{2}.$$ 

The last equality came by factoring $d^2 - c^2 = (d+c)(d-c)$.

Example 15.7. $X$ an exponential random variable with mean $m$. Show that the expected value is $m$. (That’s why we say, ‘with mean $m$’.)

answer: The range is $[0, \infty)$ and the density is $\frac{e^{-x/m}}{m}$. So,

$$E(X) = \int_0^\infty x \frac{e^{-x/m}}{m} \, dx = -xe^{-x/m} - me^{-x/m}\bigg|_0^\infty = m.$$ 

(The integral is done by parts, and the limit at $\infty$ is 0 because the exponent is negative.)

15.4 Cumulative Distribution Function

The cumulative distribution function $F(x)$ is just the accumulated probability from the left side of the range to the point $x$, i.e. the integral of the probability density function.

Definition. If $X$ is a continuous random variable with range $[c, d]$ and probability density $f(x)$ then its probability distribution function is

$$F(x) = \int_c^x f(u) \, du = P(X \leq x).$$

Exponential density (left) and cumulative distribution function (right). ($m = 10$)

Notes:

(i) $F(x)$ is an antiderivative of $f(x)$.

(ii) The integral starts at the left end of the range of $X$. So, $F(x)$ is just the probability $P(X \leq x)$ – see figure above.
(iii) $F(x)$ is increasing –this is because $f(x) \geq 0$, so the area increases as $x$ moves to the 
right.

(iii) $\lim_{x \to \infty} F(x) = 1$, i.e. the total probability is 1.

(iv) $P(a \leq X \leq b) = F(b) - F(a)$.

**Proof.** Since $F(x)$ is an antiderivative of $f(x)$ we have 

$$P(a \leq X \leq b) = \int_a^b f(x) \, dx = F(x)|_a^b = F(b) - F(a).$$

**Example 15.8.** Suppose $X$ an exponential random variable with mean $m$. Compute its 
cumulative distribution function.

**answer:** 

$$F(x) = \int_0^x \frac{e^{-u/m}}{m} \, du = e^{-u/m}\bigg|_0^x = 1 - e^{-x/m}. \quad \text{(See above figures.)}$$

**Example 15.9.** (2.2 in supplementary notes). A radioactive substance emits a beta-particle on 
average every 10 seconds. What’s the probability of waiting more than a minute for the next emission?

**answer:** Radioactive waiting times are modeled by an exponential distribution $X$ with 
mean $m$.

Find $m$: $m$ is the average time. So, $m = 10$ seconds.

One minute = 60 seconds, so we are asked to find $P(X > 60)$. We know $F(x) = 1 - e^{-x/10}$ 

$$P(X > 60) = \int_{60}^{\infty} f(x) \, dx = F(x)|_{60}^{\infty} = (1 - e^{-x/10})|_{60}^{\infty} = e^{-6} \approx 0.002.$$ 

(Very small probability.)

Note. Another standard way to do the computation is as 

$$P(X > 60) = 1 - P(X < 60) = 1 - F(60) = 1 - (1 - e^{-6}) = e^{-6} \approx 0.002.$$

### 15.5 Variance and standard Deviation

In general, we think of the expected value or mean of a random variable as a measure of 
location. That is, in a sense it gives the center of the probability distribution. In a precise 
sense, it is exactly like the center of mass of a mass distribution.

In this section we will discuss variance and standard deviation. These are measures of the 
spread of a probability distribution.

Recall, if $X$ is a discrete random variable taking values $x_1, x_2, \ldots$ then its mean or expected 
value is 

$$m = E(X) = \sum_i x_i P(x_i)$$

**Definition.** The variance of $X$ about its mean is 

$$\sigma^2(X) = \sum_i (x_i - m)^2 P(x_i).$$
Usually we just say ‘variance of $X’$. We say variance is a measure of the spread about the mean. This is because $x_i - m$ is the distance of the value $x_i$ from the mean. The variance is a weighted some of these distances, with the weighting depending on the probability that $x_i$ occurs.

The units of variance are units of $X$ squared. In order to get a measure in units of $X$ we take the square root.

**Definition.** The **standard deviation of $X$ about its mean** is $\sigma(X) = \sqrt{\sigma^2(X)}$.

We have the equivalent definitions for $X$ a continuous random variable with range $[c, d]$.

**Definition.** The mean or expected value of $X$ is

$$m = E(X) = \int_c^d x f(x) \, dx.$$

**Definition.** The **variance of $X$ about its mean** is

$$\sigma^2(X) = \int_c^d (x - m)^2 f(x) \, dx.$$

**Definition.** The **standard deviation of $X$ about its mean** is

$$\sigma(X) = \sqrt{\sigma^2(X)}.$$

**Theorem.** We have the formulas

$$\sigma^2(X) = \sum_i x_i^2 P(x_i) - m^2 \quad \sigma^2(X) = \int x^2 f(x) \, dx - m^2.$$

For discrete and continuous random variables respectively.

**Proof:** See the supplementary random notes.

**Notes:**
(i) Variance is always non-negative.
(ii) Variance measures the spread of the distribution around the mean.
(iii) If $X$ is a Poisson random variable with mean $m$ then $\sigma(X) = \sqrt{m}$.
(iv) If $X$ is an exponential random variable with mean $m$ then $\sigma(X) = m$.

The proofs of (iii) and (iv) are similar to those used to find $E(X)$.

Standard deviation will be very important when we study normal distributions.

### 15.6 Histograms

**Histograms are not on the Fall 2017-18 final.**

Let $X$ be an exponential distribution with mean 10. So, $f(x) = \frac{1}{m} e^{-x/m}$ is its probability density function.

The shaded area $= P(a \leq X \leq b) = \int_a^b f(x) \, dx$. 
Concept question: What is the area under the curve from $x = 0$ to $\infty$?

The next picture shows histograms made from data taken (using a statistical package) from an exponential distribution. The histograms are scaled to have total area 1.

Notice how well they approximate the density.

16 Probability: Normal Distributions

Read the supplementary notes

16.1 Standard Normal Random Variables

This is the famous bell curve, it has many uses throughout math and science. The most basic involves many measurements of the same quantity.

The following describe the standard normal distribution.

(i) Standard notation: $Z$

(ii) Values: $(-\infty, \infty)$.

(iii) Probability density function: $\phi(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}$. 
(iv) Cumulative distribution function:

\[ \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} \, dt. \]

(v) \( P(a \leq Z \leq b) = \frac{1}{\sqrt{2\pi}} \int_{a}^{b} e^{-t^2/2} \, dt = \Phi(b) - \Phi(a). \)

Notes

1. \( Z, \phi(z) \) and \( \Phi(z) \) are all standard notation.

2. We need the factor of \( \frac{1}{\sqrt{2\pi}} \) so that the total probability is 1. (We will show this in 18.02.)

3. \( \Phi(z) \) has no closed form, we must compute it numerically. See the table in the supplementary notes (it’s also on page 3 of these notes).

There are several useful formulas based on the symmetry of the density function.

4. \[
\Phi(-z) = \int_{-\infty}^{-z} \phi(t) \, dt
= \int_{-\infty}^{z} \phi(t) \, dt \quad \text{(by symmetry)}
= 1 - \Phi(z).
\]

5. \( P(-a < Z < a) = \Phi(a) - \Phi(-a) = \Phi(a) - (1 - \Phi(a)) = 2\Phi(a) - 1. \)

Theorem. \( E(Z) = 0, \quad \sigma(Z) = 1. \)
Proof: $z\phi(z)$ is an odd function, so

$$E(Z) = \int_{-\infty}^{\infty} z\phi(z) \, dz = 0.$$ 

The proof for $\sigma$ is in the pset.

**Important numbers**

The following values give useful rules of thumb. You should memorize them!

(i) $P(-1 < Z < 1) \approx 0.68$

(ii) $P(-2 < Z < 2) \approx 0.95$

That is, the probability $Z$ is within one standard deviation of the mean is 0.68 and the probability $Z$ is within two standard deviations of the mean is 0.95.

Normal Distributions. We can scale and shift $Z$. We define

$$X_{m,\sigma} = \sigma Z + m.$$ 

This is called normal with mean $m$ and standard deviation $\sigma$. So, $Z = X_{0,1}$.

Notice how the densities with bigger standard deviations are more spread out.

**Theorem.**

1. $X_{m,\sigma}$ has probability density function $\phi_{m,\sigma}(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$.
2. $E(X_{m,\sigma}) = m$, $\sigma(X_{m,\sigma}) = \sigma$.
3. $P(-\sigma < X_{m,\sigma} - m < \sigma) \approx 0.68$. 
4. \( P(-2\sigma < X_{m,\sigma} - m < 2\sigma) \approx 0.95. \)

**Proof.** This is just a change of variables argument.

We transform \( X_{m,\sigma} \) to standard normal –this is the same type of argument we will use repeatedly.

Proof of 1.

\[
P(a < X_{m,\sigma} < b) = P(a < \sigma Z + m < b) = P\left( \frac{a - m}{\sigma} < Z < \frac{b - m}{\sigma} \right) = \frac{1}{\sqrt{2\pi}} \int_{(a - m)/\sigma}^{(b - m)/\sigma} e^{-z^2/2} dx.
\]

Now make the change of variable \( x = \sigma z + m \), which gives

\[
P(a < X_{m,\sigma} < b) = \frac{1}{\sqrt{2\pi}} \int_{a}^{b} e^{-(x-m)^2/2\sigma^2} dx.
\]

Proof of 2.

\[
E(X_{m,\sigma}) = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} xe^{-(x-m)^2/2\sigma^2} dx.
\]

The same change of variable as in 1 gives

\[
E(X_{m,\sigma}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma ze^{-z^2/2} dz + \frac{m}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} dz = m.
\]

Proof of 3. This is similar to 2.

**Example 16.1.** Suppose the average height of a male in Boston is a normal r.v. with mean 68 inches and with a standard deviation of 4.

(a) What is the percentage of men over 6 feet tall (6 feet = 72 inches)?

**Answer:** Let \( X = \text{height of a randomly chosen person}. \)

We want \( P(X > 72) \).

Note that 72 = 68 + 4 = \( \mu + \sigma \).

For a normal random variable \( P(X > \mu + \sigma) = 0.16. \)

Reason 1: Look in the table \( P(Z < 1) = 0.84 \Rightarrow P(Z > 1) = 0.16. \)

Reason 2: For a normal random variable 68% of the probability is within one standard deviation of the mean. Thus, 32% is outside (above or below) 1 standard deviation from the mean.
Thus, (by symmetry) 16% are more than 1 standard deviation above the mean.
Final answer: 16%.

(b) What percentage of Boston males will be able to walk through a 76 inch door frame without ducking?

**Answer:** We want to compute \( P(X < 76) \).

76 inches is 2\( \sigma \) above the mean. We know 95% are within 2\( \sigma \) of the mean. So only 2.5% are more than 2\( \sigma \) above the mean \( \Rightarrow P(X < 76) \approx 97.5\% \).

**Table of values for \( \Phi(z) \)**

\[
\begin{array}{cccccccccc}
 z & 0 & .1 & .2 & .3 & .4 & .5 & .6 & .7 & .8 & .9 \\
 \Phi(z) & 0.5000 & 0.5398 & 0.5793 & 0.6179 & 0.6554 & 0.6915 & 0.7257 & 0.7580 & 0.7881 & 0.8159 \\
 z & 1.0 & 1.1 & 1.2 & 1.3 & 1.4 & 1.5 & 1.6 & 1.7 & 1.8 & 1.9 \\
 \Phi(z) & 0.8413 & 0.8643 & 0.8849 & 0.9032 & 0.9192 & 0.9332 & 0.9452 & 0.9554 & 0.9641 & 0.9713 \\
 z & 2.0 & 2.1 & 2.2 & 2.3 & 2.4 & 2.5 & 2.6 & 2.7 & 2.8 & 2.9 & 3.0 \\
 \Phi(z) & 0.9772 & 0.9821 & 0.9861 & 0.9893 & 0.9918 & 0.9938 & 0.9953 & 0.9965 & 0.9974 & 0.9981 & 0.9987 \\
\end{array}
\]

**Example 16.2.** (Another table lookup example). Suppose \( X \) is a normal random variable, with mean 5 and standard deviation 2. What is the probability \( X < 0 \)?

**Answer:**

\[
P(X < 0) = P \left( \frac{X - 5}{2} < -\frac{5}{2} \right) = P(Z < -2/5)
\]

\[
= \Phi(-2.5) = 1 - \Phi(2.5) = 1 - 0.9938 = 0.0062.
\]

**Example 16.3.** Suppose the lifetime in hours of a brand of flashlight batteries is a normal random variable \( X \) with mean 100 and standard deviation 12. Give an interval in which \( X \) lies 95% of the time.

**Answer:** We know \( P(m - 2\sigma < X < m + 2\sigma) \approx 0.95 \), so \( P(76 < X < 124) \approx 0.95 \).

Thus, our interval is \( 76 < X < 124 \).