6 Taylor Series

You are probably already familiar with infinite series. Power series (and the Taylor series we will be studying) are a special type of infinite series where each term is proportional to a power of a variable x.

- They can be used to approximate values of a function $f(x)$ near a point $x = a$ (say) using the derivatives of the function at that point, that is, $f'(a)$, $f''(a)$, $f'''(a)$ etc.
- Power series can also be used to approximate functions themselves or, more usually, solutions to differential equations. Their advantage in solving such equations issues from the ease with which power series can be differentiated and integrated, since term by term they are simply powers x^n.
- Power series are also ubiquitous throughout numerical analysis. Every time you press a function button on a calculator (e.g. to find $\sin(\pi/5)$), you are usually summing up a rapidly convergent power series. The series is truncated after a few terms when the result has sufficient accuracy.

6.1 Approximating a function

We want to approximate a function $f(x)$ near a point $x = a$ using its derivatives at that point.

A very crude estimate for f evaluated at a point near to $x = a$ can be obtained by a linear approximation:

$$f(a+h) \approx f(a) + hf'(a) , \tag{234}$$

where h is the distance to the nearby point. This follows by first noting that $\tan\theta = f'(a)$, and then by simple trigonometry (see figure).

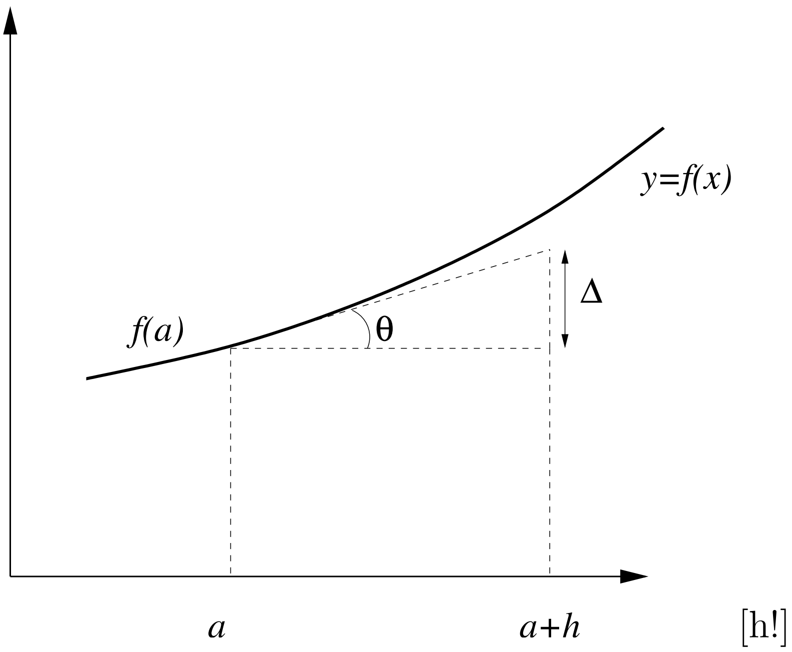

Figure 42: First approximation to $f(a+h)$ given $f(a)$ and $f'(a)$.

Of course, the smaller h is, then the better the approximation. But let's keep h fixed and try and improve the approximation.

We can rewrite (234) in a slightly different way as

$$f(x) \approx f(a) + (x-a)f'(a) \ . \tag{235}$$

Now we are approximating the function itself $f(x)$ rather than its value at a specific point.

We can do the same with $f'(x)$, as

$$f'(x) \approx f'(a) + (x-a)f''(a). \tag{236}$$

Now remember the Fundamental Theorem of Calculus,

$$\int_a^{a+h} f'(x)\,\mathrm{d}x = f(a+h) - f(a) \ . \tag{237}$$

Next substitute the approximation (236) into (237), to give

$$\begin{aligned}\int_a^{a+h} f'(x)\,\mathrm{d}x &\approx \int_a^{a+h} [f'(a) + (x-a)f''(a)]\,\mathrm{d}x \\ &= \left[xf'(a) + \frac{(x-a)^2}{2} f''(a) \right]_a^{a+h} \\ &= hf'(a) + \frac{h^2}{2} f''(a)\ , \qquad (238)\end{aligned}$$

and putting approximation (238) into the left of (237) then gives

$$f(a+h) \approx f(a) + hf'(a) + \frac{h^2}{2} f''(a)\ . \qquad (239)$$

Equation (239) is a *second-order* approximation for $f(a+h)$ (since it involves h squared), and it is an *improvement* over the first-order approximation (234). We can derive higher-order approximations: if we use (239) to give a second-order approximation for $f'(x)$ in the form

$$f'(x) \approx f'(a) + (x-a)f''(a) + \frac{(x-a)^2}{2} f'''(a)\ , \qquad (240)$$

and putting (240) into the left of (237) then gives

$$\begin{aligned}\int_a^{a+h} f'(x)\,\mathrm{d}x &\approx \int_a^{a+h} \left[f'(a) + (x-a)f''(a) + \frac{(x-a)^2}{2} f'''(a) \right] \mathrm{d}x \\ &= hf'(a) + \frac{h^2}{2} f''(a) + \frac{h^3}{6} f'''(a)\ . \qquad (241)\end{aligned}$$

Now putting approximation (241) into (237) gives

$$f(a+h) \approx f(a) + hf'(a) + \frac{h^2}{2} f''(a) + \frac{h^3}{6} f'''(a)\ . \qquad (242)$$

Equation (242) is a *third-order* approximation for $f(a+h)$ (since it involves h to the third power) and it is, in turn, an improvement over the second-order approximation (239).

We can carry on doing this for as many times as $f(x)$ is differentiable at $x = a$, and find that the nth order approximation is

$$\boxed{\begin{aligned} f(a+h) \approx\ & f(a) + hf'(a) + \tfrac{h^2}{2}f''(a) + \tfrac{h^3}{6}f'''(a) + \tfrac{h^4}{24}f^{(4)}(a) \\ & + \tfrac{h^5}{120}f^{(5)}(a) + \cdots + \tfrac{h^{n-1}}{(n-1)!}f^{(n-1)}(a) + \tfrac{h^n}{n!}f^{(n)}(a)\ . \end{aligned}} \qquad (243)$$

6.2 Taylor's theorem

Equation (243) is an approximate result but (hopefully) the error involved gets smaller as more and more term are included.

This is in fact guaranteed by **Taylor's Theorem**. We state the exact result

$$\begin{aligned} f(a+h) &= f(a) + hf'(a) + \frac{h^2}{2}f''(a) + \frac{h^3}{6}f'''(a) + \frac{h^4}{24}f^{(4)}(a) \\ &+ \frac{h^5}{120}f^{(5)}(a) + \cdots + \frac{h^n}{n!}f^{(n)}(a) + R_{n+1}\ , \end{aligned} \qquad (244)$$

where R_{n+1} is the **remainder term**, which is unknown. Taylor's Theorem states that, provided f can be differentiated $n+1$ times, there exists some point $x = \zeta$ which lies in the range $a < \zeta < a + h$ such that

$$R_{n+1} = \frac{h^{n+1}}{(n+1)!}f^{(n+1)}(\zeta)\ . \qquad (245)$$

What this means is that the error in approximating $f(a + h)$ by the nth order approximation (243) is R_{n+1}, and that the size of this error is proportional to h^{n+1}.

Equation (244) is a Taylor expansion about the point a. It is often written in the alternative, but completely equivalent, way as

$$\boxed{\begin{aligned} f(x) =\ & f(a) + (x-a)f'(a) + \tfrac{(x-a)^2}{2}f''(a) + \tfrac{(x-a)^3}{6}f'''(a) \\ & + \tfrac{(x-a)^4}{24}f^{(4)}(a) + \tfrac{(x-a)^5}{120}f^{(5)}(a) + ... + \tfrac{(x-a)^n}{n!}f^{(n)}(a) + R_{n+1} \end{aligned}} \qquad (246)$$

Here we are rewriting the function itself as a sum, rather than its value at a specific point. This sum (minus the remainder) is called a *Taylor polynomial*. As an example, let's derive an approximation for $e^{1/2}$, an actual number.

Write $f(x) = \exp(x)$, $a = 0$ and $h = 1/2$. First note

$$f'(0) = f''(0) = f'''(0) = ... = f^{(n)}(0) = e^0 = 1 . \qquad (247)$$

Plugging all this into (244) gives

$$\exp(1/2) = 1 + \frac{1}{2} + \frac{1}{2}\left(\frac{1}{2}\right)^2 + \frac{1}{6}\left(\frac{1}{2}\right)^3 + \frac{1}{24}\left(\frac{1}{2}\right)^4 + ... + \frac{1}{n!}\left(\frac{1}{2}\right)^n + R_{n+1} , \qquad (248)$$

where the remainder is now

$$R_{n+1} = \frac{1}{(n+1)!}\left(\frac{1}{2}\right)^{n+1} \exp(\zeta) \qquad (249)$$

for some $0 < \zeta < 1/2$.

Though we don't know what ζ actually, we can still estimate the worst case error. So for instance, if we wish to estimate $\exp(1/2)$ with a relative error of no more than 10^{-6}, how high does n have to be? The relative error (the ratio of the error to the exact result) associated with the nth order approximation is

$$\frac{R_{n+1}}{\exp(1/2)} = \frac{1}{(n+1)!}\left(\frac{1}{2}\right)^{n+1} \frac{\exp(\zeta)}{\exp(1/2)} . \qquad (250)$$

Since $0 < \zeta < 1/2$, the biggest possible value of $\exp(\zeta)$ is $\exp(1/2)$, so from (250) the relative error is at worst

$$\frac{1}{(n+1)!}\left(\frac{1}{2}\right)^{n+1} . \qquad (251)$$

Experimenting with a calculator shows that this maximum relative error is 1.55×10^{-6} for $n = 6$, and 9.69×10^{-8} for $n = 7$, so that to get a relative error of no more that 10^{-6} we require the seventh order approximation, $n = 7$.

Example 6.1 Find the nth order Taylor sum for $\exp x$ about $x = a$.

So we set $f(x) = e^x$, and note that $f^{(r)}(x) = e^x$ for all positive integers r. Thus $f^{(r)}(a) = e^a$. We next apply the formula:

$$f(x) = e^a + (x-a)e^a + \frac{1}{2}e^a(x-a)^2 + \cdots + \frac{1}{n!}e^a(x-a)^n + R_{n+1},$$

where the remainder term is given by

$$R_{n+1} = \frac{(x-a)^{n+1}}{(n+1)!}f^{(n+1)}(\zeta) = \frac{(x-a)^{n+1}}{(n+1)!}e^{\zeta},$$

and ζ is some number between a and x. Note that if x is much larger than a then the remainder might also be inconveniently big, unless of course n, the number of terms, is also sufficiently large.

6.3 Taylor series

If a function $f(x)$ is infinitely differentiable, and its remainder R_{n+1} goes to zero as $n \to \infty$, then we can continue adding successive terms to our Taylor polynomial forever. Ultimately we can represent $f(x)$ **exactly** as an *infinite series*, with the remainder term R_{n+1} dropped.

This is called a **Taylor series**. The Taylor series of $f(x)$ about $x = a$ is

$$\boxed{\begin{aligned} f(x) &= f(a) + (x-a)f'(a) + \tfrac{(x-a)^2}{2!}f''(a) + \tfrac{(x-a)^3}{3!}f'''(a) + \ldots \\ &\quad + \tfrac{(x-a)^n}{n!}f^{(n)}(a) + \ldots . \end{aligned}}$$

The special case of the Taylor series about $x = 0$ is called a **Maclaurin series**:

$$f(x) = f(0) + xf'(0) + \tfrac{x^2}{2!}f''(0) + \tfrac{x^3}{3!}f'''(0) + \ldots + \tfrac{x^n}{n!}f^{(n)}(0) + \ldots . \quad (252)$$

Finally, we formally define the term **power series**. It refers to an infinite series where each term is proportional to a power of a variable x:

$$\sum_{n=0}^{\infty} b_n x^n = b_0 + b_1 x + b_2 x^2 + b_3 x^3 + ..., \tag{253}$$

for some coefficients b_n. If a power series converges it is then necessarily the Taylor series of some function (cf. Borel's Theorem). Consequently, we will use the terminology power and Taylor series interchangeably.

6.4 Taylor series of the exponential function and its relatives

As a simple example, consider the power series expansion of $\exp x$ about $x = 0$. All the derivatives of $\exp(x)$ are equal to 1 at $x = 0$. Equation (252) then gives us

$$\boxed{\exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + ... + \frac{x^n}{n!} + ... = \sum_{n=0}^{\infty} \frac{x^n}{n!}\ .} \tag{254}$$

By replacing x by $-x$ in (254) we find that

$$\boxed{\exp(-x) = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + ... + \frac{(-1)^n x^n}{n!} + ... = \sum_{n=0}^{\infty} \frac{(-1)^n x^n}{n!}\ ,} \tag{255}$$

and so

$$\cosh x = \frac{1}{2}\left(1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + ... + \left[1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \frac{x^4}{4!} - .....\right]\right)\ . \tag{256}$$

The terms in even powers of x add together, but the odd terms cancel:

$$\boxed{\cosh(x) = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \frac{x^6}{6!}.... + \frac{x^{2n}}{(2n)!} + ... = \sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!}\ .} \tag{257}$$

Notice here how the general term is an even power of x.

In the same way, we can find the power series expansion for $\sinh x$:

$$\sinh x = \frac{1}{2}\left(1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + ... - \left[1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \frac{x^4}{4!} - .....\right]\right). \tag{258}$$

The terms in odd powers of x now add together, while the even terms cancel:

$$\boxed{\sinh(x) = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \frac{x^7}{7!} .... + \frac{x^{2n+1}}{(2n+1)!} + ... = \sum_{n=0}^{\infty} \frac{x^{2n+1}}{(2n+1)!}.} \tag{259}$$

Notice here how the general term is an odd power of x.

Example 6.2 Find the first three nonzero terms in the power series for $\tanh x$ about $x = 0$.

Set $f(x) = \tanh x$, and note that $f(0) = 0$. We next work out its derivatives evaluated at $x = 0$:

$$\begin{aligned}
f'(x) &= \text{sech}^2 x, \\
f''(x) &= \frac{d(\cosh x)^{-2}}{dx} = -2\sinh x(\cosh x)^{-3}, \\
f'''(x) &= -2\cosh x(\cosh x)^{-3} + 6\sinh^2 x(\cosh x)^{-4}, \\
f^{(4)}(x) &= 4\sinh x(\cosh x)^{-3} + 12\sinh x(\cosh x)^{-3} - 24\sinh^3 x(\cosh x)^{-5}, \\
f^{(5)}(x) &= -48\sinh^2 x(\cosh x)^{-4} + 16(\cosh x)^{-2} \\
&\qquad - 72\sinh^2 x(\cosh x)^{-4} + 120\sinh^4 x(\cosh x)^{-4},
\end{aligned}$$

with

$$f'(0) = 1, \quad f''(0) = 0, \quad , f'''(0) = -2, \quad f^{(4)}(0) = 0, \quad f^{(5)}(0) = 16.$$

If one still has one's sanity, we can then apply the formula to get the first three terms:

$$\begin{aligned}\tanh x &= x + \frac{1}{3!}(-2)x^3 + \frac{1}{5!}(16)x^5 + \ldots, \\ &= x - \frac{x^3}{3} + \frac{2x^5}{15} - \ldots.\end{aligned}$$

6.5 Power series of trigonometric functions

Consider $f(x) = \sin x$. Note that

$$\begin{aligned}\frac{\mathrm{d}\sin x}{\mathrm{d}x} &= \cos x &= \sin\left(x + \tfrac{\pi}{2}\right) \\ \frac{\mathrm{d}^2\sin x}{\mathrm{d}x^2} &= -\sin x &= \sin(x + \pi) \\ \frac{\mathrm{d}^3\sin x}{\mathrm{d}x^3} &= -\cos x &= \sin\left(x + \tfrac{3\pi}{2}\right) \\ \frac{\mathrm{d}^4\sin x}{\mathrm{d}x^4} &= \sin x &= \sin(x + 2\pi) \\ \vdots &= \vdots &= \vdots\end{aligned}$$

which leads us to the useful general result

$$\boxed{\frac{\mathrm{d}^n\sin x}{\mathrm{d}x^n} = \sin\left(x + \frac{n\pi}{2}\right)\ .} \tag{260}$$

Therefore

$$\begin{aligned}f(0) &= \sin(0) = 0 \qquad f'(0) = \sin(\pi/2) = 1 \qquad f''(0) = \sin(\pi) = 0 \\ f'''(0) &= \sin(3\pi/2) = -1 \quad f^{(4)}(0) = \sin(2\pi) = 0....\ .\end{aligned}$$

Putting these results together:

$$\boxed{\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!}..... + \frac{(-1)^n x^{2n+1}}{(2n+1)!} + ... = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!}\ .} \tag{261}$$

In the same way we can find a power series for $f(x) = \cos x$. Note that

$$\begin{aligned}
\frac{\mathrm{d}\cos x}{\mathrm{d}x} &= -\sin x &=& \cos\left(x+\frac{\pi}{2}\right)\\
\frac{\mathrm{d}^2\cos x}{\mathrm{d}x^2} &= -\cos x &=& \cos\left(x+\pi\right)\\
\frac{\mathrm{d}^3\cos x}{\mathrm{d}x^3} &= \sin x &=& \cos\left(x+\frac{3\pi}{2}\right)\\
\frac{\mathrm{d}^4\cos x}{\mathrm{d}x^4} &= \cos x &=& \cos\left(x+2\pi\right)\\
\vdots &= \vdots &=& \vdots
\end{aligned}$$

which leads us to another useful general result

$$\boxed{\frac{\mathrm{d}^n\cos x}{\mathrm{d}x^n} = \cos\left(x+\frac{n\pi}{2}\right)}\ . \tag{262}$$

Therefore

$$\begin{aligned}
f(0) &= \cos(0) = 1 & f'(0) &= \cos(\pi/2) = 0 & f''(0) &= \cos(\pi) = -1\\
f'''(0) &= \cos(3\pi/2) = 0 & f^{(4)}(0) &= \cos(2\pi) = 1
\end{aligned} \tag{263}$$

Putting these results together

$$\boxed{\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!}\ldots.. + \frac{(-1)^n x^{2n}}{(2n)!} + \ldots. = \sum_{n=0}^{\infty}\frac{(-1)^n x^{2n}}{(2n)!}} \tag{264}$$

Example 6.3 Find the power series for $\sin x$ about $x = a$.

Once again we set $f(x) = \sin x$, and note that the formula for the power series around $x = a$ is $f(x) = \sum_n (f^{(n)}(a)/n!)(x-a)^n$. From earlier

$$f^{(n)}(a) = \left.\frac{d^n\sin x}{dx^n}\right|_{x=a} = \sin\left(a+\frac{n\pi}{2}\right).$$

Therefore:

$$\begin{aligned}
\sin x &= \sum_{n=0}^{\infty}\frac{\sin\left(a+\frac{n\pi}{2}\right)}{n!}(x-a)^n,\\
&= \sin a + \cos a\,(x-a) - \frac{\sin a}{2}(x-a)^2 - \frac{\cos a}{3!}(x-a)^3 + \frac{\sin a}{4!}(x-a)^4 + \ldots
\end{aligned}$$

An alternative way to this result is the following. Set $z = x - a$. Then

$$\sin x = \sin(z + a) = \sin z \cos a + \cos z \sin a.$$

Next we expand $\sin z$ and $\cos z$ around $z = 0$ using the canonical formulas (261) and (264), and reorganise the terms.

6.6 Euler's formula revisited

In Section 2.3 we made considerable use of the unproven result

$$\exp(i\theta) = \cos\theta + i\sin\theta\ . \tag{265}$$

Now we can prove it. First note from (254) that

$$\begin{aligned}\exp(\mathrm{i}\theta) &= 1 + \mathrm{i}\theta + \frac{(\mathrm{i}\theta)^2}{2!} + \frac{(\mathrm{i}\theta)^3}{3!} + \frac{(\mathrm{i}\theta)^4}{4!} + \frac{(\mathrm{i}\theta)^5}{5!} + \frac{(\mathrm{i}\theta)^6}{6!} + ...\\ &= 1 + \mathrm{i}\theta - \frac{\theta^2}{2!} - \frac{\mathrm{i}\theta^3}{3!} + \frac{\theta^4}{4!} + \frac{\mathrm{i}\theta^5}{5!} - \frac{\theta^6}{6!} + ...\\ &= 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + ... + \mathrm{i}\left[\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - ...\right],\end{aligned} \tag{266}$$

which we now compare with the power series for $\cos\theta$ (equation 264) and the power series for $\sin\theta$ (equation 261).

It is clear that the real part of the right hand side of (266) is $\cos\theta$ and the imaginary part is $\sin\theta$, thereby proving (265).

6.7 Power series for logarithms

We can't derive a power series for $\ln x$ about $x = 0$, because $\ln 0$ is undefined (so the first term in (252) is undefined).

However, we can find a power series for $\ln(1+x)$ about $x = 0$. Writing $f(x) = \ln(1+x)$ we see that $f(0) = \ln 1 = 0$, while

$$\begin{array}{rclrclrcl} f'(x) & = & \frac{1}{1+x} & f''(x) & = & -\frac{1}{(1+x)^2} & f'''(x) & = & \frac{2}{(1+x)^3} \\ f^{(4)}(x) & = & -\frac{6}{(1+x)^4} & f^{(5)}(x) & = & \frac{24}{(1+x)^5} & f^{(n)}(x) & = & (-1)^{n-1}\frac{(n-1)!}{(1+x)^n} . \end{array}$$

It follows that the r'th term in the power series is

$$\frac{f^{(n)}(0)}{n!}x^n = \frac{(-1)^{n-1}(n-1)!}{n!}x^n = \frac{(-1)^{n-1}}{n}x^n . \tag{267}$$

The power series for $\ln(1+x)$ about $x = 0$ is therefore

$$\boxed{\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \frac{x^5}{5} + ...(-1)^{n-1}\frac{x^n}{n} + ...} \tag{268}$$

By replacing x by $-x$ we can also find the power series for $\ln(1-x)$ about $x = 0$,

$$\ln(1-x) = -\left[x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \frac{x^5}{5} + ... + \frac{x^n}{n} + ...\right] \tag{269}$$

The power series for $\sin x$, $\cos x$, $\exp x$ etc are valid for *any* real value of x. However, the power series for $\ln(1 \pm x)$ have a limited range of validity: the power series for $\ln(1+x)$ is valid on the real axis for $-1 < x \leqslant 1$, and the power series for $\ln(1-x)$ is valid on the real axis for $-1 \leqslant x < 1$.

6.8 The binomial expansion

We now consider the function

$$f(x) = (1+x)^\alpha ,$$

where α is a real number (not necessarily an integer).

By successive differentiation we find

$$\begin{aligned} f'(x) &= \alpha(1+x)^{\alpha-1} \\ f''(x) &= \alpha(\alpha-1)(1+x)^{\alpha-2} \\ f'''(x) &= \alpha(\alpha-1)(\alpha-2)(1+x)^{\alpha-3} \\ f^{(4)}(x) &= \alpha(\alpha-1)(\alpha-2)(\alpha-3)(1+x)^{\alpha-4} \\ \vdots &= \vdots \\ f^{(n)}(x) &= \alpha(\alpha-1)(\alpha-2)(\alpha-3)...(\alpha-n+1)(1+x)^{\alpha-n} . \end{aligned} \tag{270}$$

We can use this information to find the power series of $(1+x)^\alpha$ about $x=0$:

$$\boxed{\begin{aligned} (1+x)^\alpha = & \quad 1+\alpha x+\tfrac{\alpha(\alpha-1)}{2!}x^2+\tfrac{\alpha(\alpha-1)(\alpha-2)}{3!}x^3 \\ & +\ldots \tfrac{\alpha(\alpha-1)(\alpha-2)(\alpha-3)...(\alpha-n+1)}{n!}x^n+.... \end{aligned}} \tag{271}$$

The power series (271) is valid in the range $-1 < x < 1$ for general α.

In the special case when α is a positive integer, say $\alpha = N$, then the power series stops after a finite number of terms. Specifically, the coefficient of x^{N+1} is

$$\frac{\alpha(\alpha-1)(\alpha-2)(\alpha-3)...(\alpha-N+1)(\alpha-N)}{(N+1)!}x^{N+1} , \tag{272}$$

but since $\alpha = N$ the final factor in (272) is precisely zero. Hence the term for x^{N+1} vanishes, as do the terms for any higher powers of x, since all such higher terms also contain the factor $\alpha - N$.

When $\alpha = N$ the power series (271) reduces to the polynomial:

$$\begin{aligned} (1+x)^N &= 1+Nx+\frac{N(N-1)}{2!}x^2+\frac{N(N-1)(N-2)}{3!}x^3+..... \\ &+ \frac{N(N-1)(N-2)(N-3)...(1)}{N!}x^N . \end{aligned} \tag{273}$$

The general term in this sum is

$$\frac{N(N-1)(N-2)...(N-r+1)}{r!}x^r ,$$

which we can rearrange to be

$$\frac{N!}{(N-r)!r!}x^r \,, \tag{274}$$

or in other words the usual binomial coefficient. So when α is a positive integer (271) agrees with the familiar binomial expansion you have seen before:

$$(1+x)^N = \sum_{r=0}^{N} \binom{N}{r} x^r \,.$$

Finally, you will often encounter the alternative notations

$$\frac{N!}{(N-r)!r!} \equiv {}^N C_r \equiv C_r^N \equiv {}_N C_r \,. \tag{275}$$

Example 6.4 Find the power series expansion about $x=0$ of $(2+x)^{-1/2}$.

We re-express the function so it is easier to apply the binomial expansion:

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

We see that $\alpha = -1/2$ and so we have:

$$\begin{aligned}(2+x)^{-1/2} &= 2^{-1/2}\left[1 - \frac{1}{2}\left(\frac{x}{2}\right) + \frac{1}{2!}\cdot\frac{-1}{2}\cdot\frac{-3}{2}\cdot\left(\frac{x}{2}\right)^2 \right.\\ &\qquad \left. + \frac{1}{3!}\cdot\frac{-1}{2}\cdot\frac{-3}{2}\cdot\frac{-5}{2}\cdot\left(\frac{x}{2}\right)^3 + \dots\right],\\ &= 2^{-1/2}\left[1 - \frac{x}{4} + \frac{3x^2}{32} - \frac{5x^3}{128} + \dots\right]\end{aligned}$$

6.9 The Newton-Raphson method

We finish by outlining a method to approximately solve nonlinear algebraic equations such as $f(x) = 0$, where f is a nonlinear function.

- Suppose we have a rough guess for what the solution is, x_0, so that $f(x_0) \approx 0$. But we want to improve its accuracy. In other words, generate a new better approximation to the solution (call it x_1).
- We write $x_1 = x_0 + h$. We want to find h.
- So we set $f(x_1) = f(x_0 + h) = 0$ and then Taylor expand f around x_0 truncating at first order

$$0 = f(x_0 + h) \approx f(x_0) + hf'(x_0)$$

- We solve this approximate equation to get h and hence x_1, our better approximation:

$$\boxed{x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}} \tag{276}$$

- We can then repeat the procedure to get x_2, an even better approximation. And we continue this till we converge on to the exact solution (within some specified accuracy)

Example: find the solution to $f(x) = x^2 - \ln(x + 5) = 0$.

Take as our initial guess $x_0 = 2$. Using a calculator, this yields $f(x_0) = 2.0541$, which is not very good. Let us see if the Newton-Raphson method can get us a better approximation:

i	x_i	$f(x_i)$	h
0	2.0	2.0541	-0.53254
1	1.4675	0.28665	-0.10310
2	1.3644	0.010758	-0.0041835
3	1.3602	1.7718×10^{-5}	-6.9125×10^{-6}

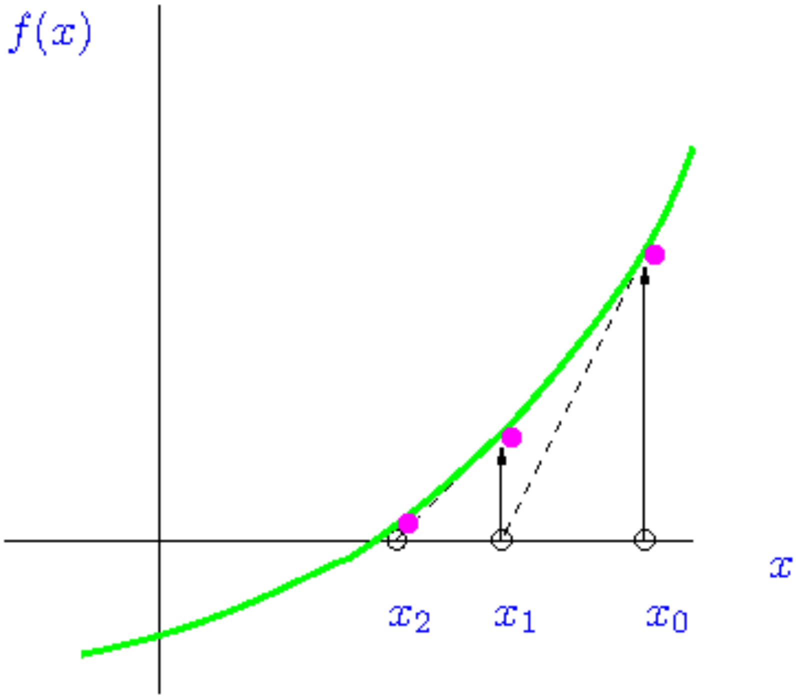

Figure 43: Successive approximations using the Newton-Raphson method.

The Newton-Raphson method has a graphical interpretation, because solving $f(x) = 0$ is the same as finding the x-intercept of the curve $y = f(x)$.

Convergence of Newton-Raphson is very rapid when x_0 is near the solution. But if f has a turning point between x_0 and the exact solution then there is the danger that the method fails.

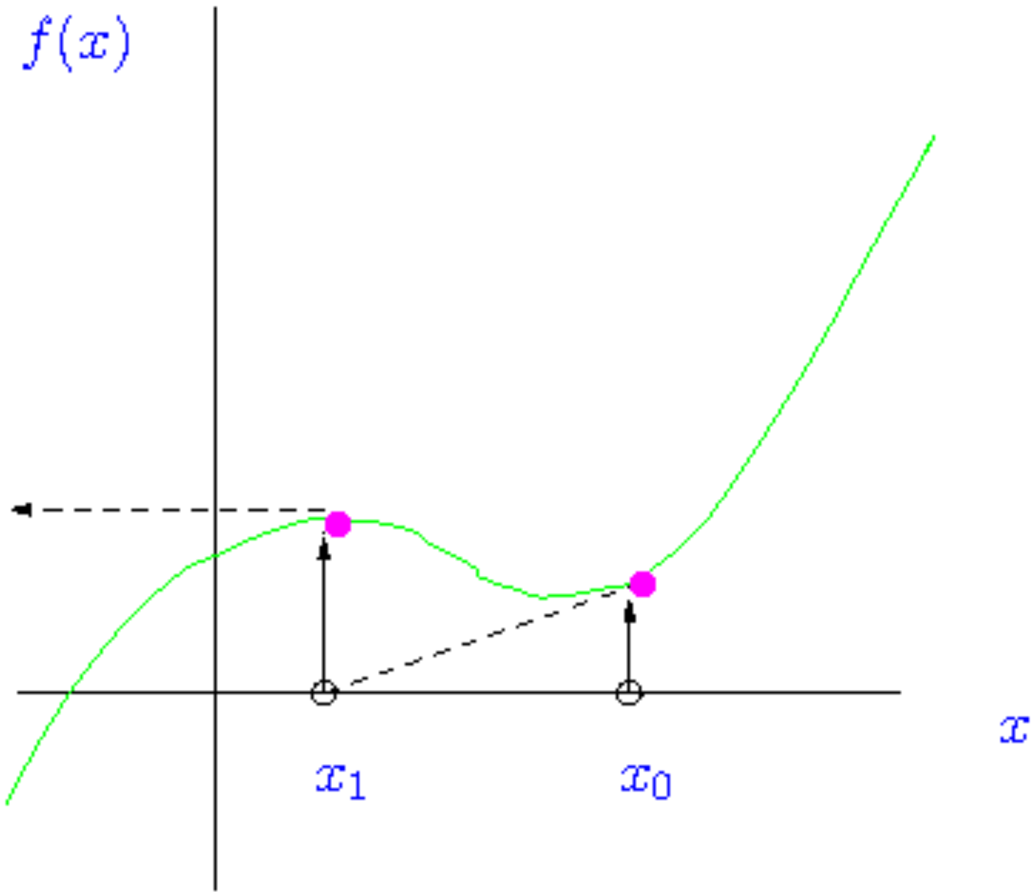

Figure 44: Example of when the Newton-Raphson may fail.

Example 6.5 [2004, paper 2, question 8D; 2006, paper 1, questions 9E]

(a) Find, by any method, the first three non-zero terms in the Taylor expansion about $x = 0$ of the following functions.

$$\frac{\log(1+x)}{1-x} \qquad \frac{1}{1+\sin x} \qquad \log(\cos x).$$

(b) Let

$$f(x) = \sum_{i=0}^{\infty} a_i x^i \quad \text{and} \quad g(x) = \sum_{i=0}^{\infty} b_i x^i$$

and let $\sum_{i=0}^{\infty} c_i x^i$ be the Taylor expansion about $x = 0$ of the function $f(x)g(x)$.

1. Find c_0, c_1 and c_2 in terms of a_0, a_1, a_2, b_0, b_1 and b_2.

2. Give a general expression for c_i as a finite sum of products of the coefficients a_j and b_j.

(c) Find the first four terms in the Taylor expansion around $x = 1$ of $\tan^{-1} x$

In (a) rather than work out all the complicated derivatives of these functions, let us write down the Taylor series of the component functions.

For the first one, recall

$$\ln(1 + x) = x - \tfrac{1}{2}x^2 + \tfrac{1}{3}x^3 - \tfrac{1}{4}x^4 + \tfrac{1}{5}x^5 - \ldots$$

and that

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \ldots,$$

the infinite geometric series with ratio x (or the binomial expansion with $\alpha = -1$). In both expansions we are assuming that x is sufficiently small so that the series converge. Combining these two:

$$\begin{aligned}\frac{\ln(1+x)}{1-x} &= (x - \tfrac{1}{2}x^2 + \tfrac{1}{3}x^3 - \tfrac{1}{4}x^4 + \ldots)(1 + x + x^2 + x^3 + \ldots), \\ &= x + x^2(1 - \tfrac{1}{2}) + x^3(\tfrac{1}{3} - \tfrac{1}{2} + 1) + \ldots, \\ &= x - \tfrac{1}{2}x^2 + \tfrac{5}{6}x^3 + \ldots .\end{aligned}$$

For the second one, introduce a new variable $y = \sin x$. If we are to expand around $x = 0$ then we expand around $y = 0$. Now

$$\frac{1}{1 + \sin x} = \frac{1}{1+y} = 1 - y + y^2 - y^3 + \ldots,$$

the last equality coming about using either the binomial expansion with $\alpha = -1$ or recognising the infinite geometric series with ratio $-y$. (This series converges for all x except those values for which $\sin x = \pm 1$.) We next insert the Taylor series for $\sin x = x - x^3/3! + x^5/5! - \dots$ and regroup the terms:

$$\begin{aligned}\frac{1}{1+\sin x} &= 1 - \sin x + (\sin x)^2 - (\sin x)^3 + \dots \\ &= 1 - (x - x^3/3! + x^5/5! - \dots) + (x - x^3/3! + x^5/5! - \dots)^2 - \dots \\ &= 1 - x + x^2 + \dots\end{aligned}$$

Finally, for the third one, we play a similar trick and set $\cos x = 1 + z$, with the new variable z. Note that expanding around $x = 0$ means expanding around $z = 0$. First we have

$$\ln(\cos x) = \ln(1+z) = z - \tfrac{1}{2}z^2 + \tfrac{1}{3}z^3 - \dots.$$

(This converges for sufficiently small z, hence sufficiently small x.) We next insert the Taylor series for $\cos x = 1 - x^2/2! + x^4/4! - x^6/6! + \dots$, recognising that $z = -x^2/2! + x^4/4! - x^6/6! + \dots$:

$$\begin{aligned}\ln(\cos x) &= (-x^2/2! + x^4/4! - x^6/6! + \dots) - \frac{1}{2}(-x^2/2! + x^4/4! - x^6/6! + \dots)^2 \\ &\qquad + \frac{1}{3}(-x^2/2! + x^4/4! - x^6/6! + \dots)^3 + \dots, \\ &= -x^2/2 + x^4/24 - x^6/720 - \frac{1}{2}(x^4/4 - x^6/24 + \dots) \\ &\qquad + \frac{1}{3}(-x^6/8 + \dots) + \dots, \\ &= -\frac{x^2}{2} - \frac{x^4}{12} - \frac{x^6}{45}.\end{aligned}$$

In part (b) we have a bonanza of terms to multiply ....

$$\begin{aligned}f(x)g(x) &= (a_0 + a_1x + a_2x^2 + \dots)(b_0 + b_1x + b_2x^2 + \dots), \\ &= a_0b_0 + (a_1b_0 + a_0b_1)x + (a_2b_0 + a_1b_1 + a_0b_2)x^2 + \dots \\ &= c_0 + c_1x + c_2x^2 + \dots,\end{aligned}$$

with $c_0 = a_0b_0$, $c_1 = a_1b_0 + a_0b_1$), and $c_2 = a_2b_0 + a_1b_1 + a_0b_2$.

We realise that the term proportional to x^l consists of a sum of all the products $a_i b_j x^{i+j}$ so that $i + j = l$. So we can write down the formula:

$$c_l = \sum_{i=0}^{l} a_i b_{l-i}.$$

Question (c) is a more straightforward one. Set $f(x) = \tan^{-1} x$. Then $f(1) = \pi/4$. And we have

$$\begin{aligned} f'(x) &= \frac{1}{1+x^2}, \\ f''(x) &= \frac{-2x}{(1+x^2)^2}, \\ f'''(x) &= \frac{-2(1+x^2)^2 - (-2x)[4x(1+x^2)]}{(1+x^2)^4} = \frac{2(3x^2-1)}{(1+x^2)^3}. \end{aligned}$$

Then $f'(1) = 1/2$, $f''(1) = -1/2$, and $f'''(1) = 1/2$, and finally:

$$\tan^{-1} x = \frac{\pi}{4} + \frac{x-1}{2} - \frac{(x-1)^2}{4} + \frac{(x-1)^3}{12} + \ldots .$$

7 Elementary Probability

Probability theory is essential in every field of science, even if it is not always used explicitly. It supplies a precise set of rules for exercising logic when we have incomplete information.

- Experiments and observations always involve random error (and usually a systematic error as well!), and so comparison with theory is inevitably probabilistic. Theory can only ever be proved up to a high level of significance.
- Some physical systems are inherently probabilistic, most notably in quantum mechanics, where we only ever compute probabilities because of the 'unknowable' nature of the subatomic world.
- Other systems, though formally deterministic, are so impossibly complicated that only a probabilistic or statistical description is feasible. Examples include: chaotic dynamical systems, fluid turbulence, the kinetic theories of gases and of reaction rates in chemistry.
- Probability theory has many other applications: the social sciences; clinical trials, epidemiology, and public health; insurance; risk assessment in share trading and commodity markets; determining reliability in consumer products, such as cars and computers; etc.
- It is also useful when playing cards and other games of chance, which is why it was originally developed: in 1654, a dispute about a popular game of dice, between Pascal and Fermat, initiated the modern study of probability, and Abraham de Moivre's book on gaming in 1711 laid down its foundations.

Probability theory and statistics differ in the sense that probability deals with the likelihood of future events (given that we understand the underlying process creating them), while statistics takes as its task the analysis of previous events, in order to determine the underlying process giving rise to these events.

7.1 Basic concepts

7.1.1 Random experiments

We will be concerned with the outcomes of **random experiments**, that is, trials or observations which can be repeated many times but which contain an element of chance.

- **Outcomes:** The possible results of the experiments are called the *outcomes*. The outcomes must be *mutually exclusive* and we can label them (say) ω_1, ω_2, ... etc.

  For example, when throwing a six-sided die the outcomes are just the numbers from one to six, $\omega_1 = 1$, $\omega_2 = 2$, ..., $\omega_6 = 6$.

  If the outcomes are described as "fair" or "unbiased" then they are equally likely.

- **Sample space:** The set of all possible outcomes of the experiment is called the *sample space*, $S = \{\omega_1, \omega_2, ...\}$.

  For the example of the die, $S = \{1, 2, 3, 4, 5, 6\}$.

- **Events:** An *event* A is a subset of the sample space S (so that $A \subset S$). An event may contain more than one outcome.

  For example, the event A might be 'throw an even number with the die', with $A = \{2, 4, 6\}$.

7.1.2 Elementary set theory

We are often concerned about whether or not two or more different events can happen together (simultaneously or consecutively) — and thus must deal with the relationship of two or more sets.

For the two events A and B we define the following sets:

1. $A \cap B$: the **intersection** of A and B, i.e. both events A and B occur;

2. $A \cup B$: the **union** of A and B, i.e. either event A, or event B, or both, occur.

3. $\overline{A}$: the **complement** of A, i.e. A does not occur. Other notations for the complement include A^c and A'.

4. $A - B$: outcomes in A which are not in B. Note that

$$A - B = A \cap \overline{B} \,. \tag{277}$$

A good way to represent these sets is via **Venn diagrams**.

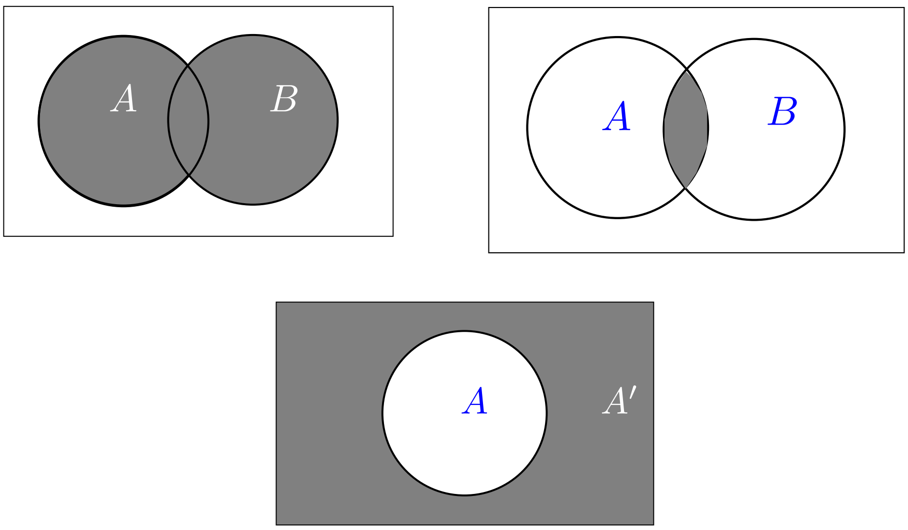

Figure 45: Panel 1 is $A \cup B$; panel 2 is $A \cap B$; lower panel is $\overline{A}$

The empty set, denoted $\varnothing$, contains no outcomes. Note that

$$A \cap \overline{A} = \varnothing, \qquad A \cup \overline{A} = S \,. \tag{278}$$

The events A and B are said to be **mutually exclusive** if they cannot both occur, i.e.

$$A \cap B = \varnothing \,. \tag{279}$$

7.1.3 Probability

The *probability* $P(A)$ expresses how likely an event A is.

Suppose we repeat our experiment a very large number N times, and find that the event A occurs N_A times. Then we define

$$P(A) = \lim_{N\to\infty} \frac{N_A}{N}. \tag{280}$$

The basic properties of probability are:

1. It is restricted between 0 and 1, i.e.

   $$0 \leqslant P(A) \leqslant 1 \ , \tag{281}$$

   with $P(S) = 1$ while $P(\varnothing) = 0$.

2. If $P(A) = 0$ then the event A is impossible.

3. For the complement of the event A we have

   $$P(\overline{A}) = 1 - P(A) \ , \tag{282}$$

   which is especially useful when $P(\overline{A})$ is easier to find than $P(A)$.

4. Additive rule for mutually exclusive outcomes. Recall that the individual outcomes $\omega_1, \omega_2, \ldots$ are mutually exclusive. So, if event $A_i = \{\omega_i\}$ for all i, and $A = \bigcup_i A_i$ is an event comprising some number of outcomes, then

   $$P(A) = \sum_i P(A_i) \tag{283}$$

5. For the union of two general events A and B (that are not necessarily mutually exclusive) we have

   $$P(A \cup B) = P(A) + P(B) - P(A \cap B) \ . \tag{284}$$

This result is easiest to see by drawing a Venn diagram - note that the term $P(A \cap B)$ is taken off the right of (284) to avoid double counting the region $A \cap B$.

If A and B are **mutually exclusive** then $P(A \cap B) = 0$ and

$$P(A \cup B) = P(A) + P(B)\,, \tag{285}$$

just as in point 4.

The result (284) can be extended to three events to give

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &- P(A \cap B) - P(B \cap C) - P(C \cap A) \\ &+ P(A \cap B \cap C)\,. \end{aligned} \tag{286}$$

This is best seen in a Venn diagram.

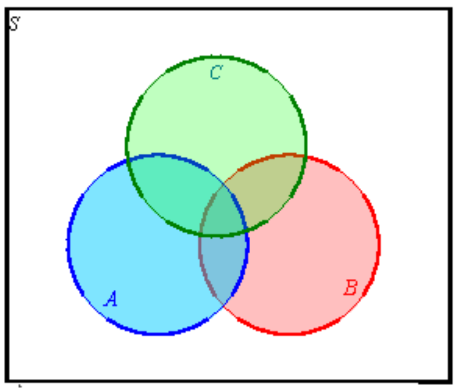

Figure 46: Venn diagram for $P(A \cup B \cup C)$.

Example 7.1 A twelve-sided die is thrown. Event A is 'the number thrown is even', event B is the number thrown is 'divisible by three', and event C is 'the number thrown is one of $6, 7, 8, 9$'. Find the probabilities of $A \cap B$, $C \cap B$, $C \cap A$, $A - B$, $A \cap B \cap C$ and $\overline{A \cup B \cup C}$.

Let us first describe, in mathematical terms, the various events. We have:

$$A = \{2,\ 4\ ,6,\ 8,\ 10,\ 12\}, \quad B = \{3,\ 6,\ 9,\ 12\}, \quad C = \{6,\ 7,\ 8,\ 9\}.$$

And since we know there are 12 equally possible outcomes in a throw, we can write down:

$$P(A) = \frac{6}{12} = \frac{1}{2}, \qquad P(B) = \frac{1}{3}, \qquad P(C) = \frac{1}{3}.$$

Consider now the event defined by the intersection of A and B. Obviously $A \cap B = \{6,\ 12\}$. And thus its probability is $P(A \cap B) = 2/12 = 1/6$. Similarly $C \cap B = \{6,\ 9\}$ and $P(C \cap B) = 1/6$.

Now $A - B$ consists of all the outcomes in A that *are not* shared by B. Thus $A - B = \{2,\ 4,\ 8,\ 10\}$. And hence $P(A - B) = 1/3$.

The intersection of A, B, and C consists of only one outcome: $A \cap B \cap C = \{6\}$. Thus $P(A \cap B \cap C) = 1/12$. On the other hand the union of the three sets is

$$A \cup B \cup C = \{2,\ 3,\ 4,\ 6,\ 7,\ 8,\ 9,\ 10,\ 12\}$$

and comprises 9 outcomes. Hence $P(A \cup B \cup C) = 9/12 = 3/4$.

Finally, recognising that $P(\overline{D}) = 1 - P(D)$ for any event D, we can write $P(\overline{A \cup B \cup C}) = 1 - P(A \cup B \cup C) = 1 - 3/4 = 1/4$. This can be checked directly by noting that $\overline{A \cup B \cup C} = \{1,\ 5,\ 11\}$, thus comprising 3 events. Its probability is then $3/12 = 1/4$, in agreement with the other approach.

Example 7.2 A card is drawn at random from a pack. Event A is 'the card is an ace', event B is 'the card is a spade (♠)', event C is 'the card is one of ace, king, queen, jack, 10'. Calculate the probability that the card has (i) at least one of these properties; (ii) all of these properties.

So we can write

$$A = \{\text{ace of } ♥,\ \text{ace of } ♦,\ \text{ace of } ♣,\ \text{ace of } ♠\}$$

and note that there are 4 outcomes, while there are 52 cards in the pack. Hence $P(A) = 4/52 = 1/13$.

Event B consists of all the spades, of which there are 13 cards (2-10, plus the 3 royals, plus the ace). Therefore $P(B) = 13/52 = 1/4$.

Event C consists of 20 cards/outcomes, because each of the fours suits contains an ace, king, queen, jack, and 10. Thus $P(C) = 20/52 = 5/13$.

Part (i) asks us to find the probability of the union of these three events: $P(A \cup B \cup C)$. It is actually a tad easier to find $P(\overline{A \cup B \cup C})$ first. The event $\overline{A \cup B \cup C}$ consists of (a) no spades and (b) only the cards $2 - 9$ in the three suits of ♥, ♦, and ♣. This gives us $8 \times 3 = 24$ cards. Thus $P(\overline{A \cup B \cup C}) = 24/52 = 6/13$. Now we can answer the question:

$$P(A \cup B \cup C) = 1 - P(\overline{A \cup B \cup C}) = 1 - \frac{6}{13} = \frac{7}{13}.$$

Part (ii) asks us to find the intersection of these sets A, B, and C. We see straightaway that $A \cap B \cap C = \{\text{ace of } ♠\}$. Thus $P(A \cap B \cap C) = 1/52$.

Example 7.3 A biased die has probability $p, 2p, 3p, 4p, 5p, 6p$ of throwing 1, 2, 3, 4, 5, 6 respectively. Find p. What is the probability of throwing an even number?

If S is the sample space, the total probability must be 1, i.e. $P(S) = 1$, or written out in full:

$$\sum_{n=1}^{6} P(\{n\}) = 1.$$

But each number rolled doesn't have an equal probability! Thus this can be rewritten

$$\sum_{n=1}^{6} np = p + 2p + 3p + 4p + 5p + 6p = 1.$$

We can solve this equation for p, and find that $p = 1/21$.

Let us denote by A the event that we roll an even number, and so $A = \{2, 4, 6\}$. We can then work out its probability directly by looking at each of the outcomes that it consists of:

$$P(A) = P(\{2\}) + P(\{4\}) + P(\{6\}) = 2p + 4p + 6p = 12p = \frac{4}{7}.$$

7.2 Conditional probability

We are often interested in determining the probability of one event given that another, possibly related, event has occurred.

The probability that event A occurs, given that event B has occurred, is denoted $P(A|B)$ and is called the **conditional probability**.

For example, event A might correspond to a student getting a first, while B corresponds to a student attending every lecture. $P(A|B)$ is the probability that a student gets a first given that they attended every lecture. $P(A|B)$ is presumably different to $P(A)$, the *unconditional* probability that a student gets a first.

Because the event B is known to have occurred, then the event B replaces S as the sample space when we try to compute the event $B|A$. This then motivates

$$P(A|B) = \frac{P(A \cap B)}{P(B)}\,, \tag{287}$$

where we have had to normalise the probability of the intersection by $P(B)$.

- If two events A and B are mutually exclusive then we have $P(A|B) = P(B|A) = 0$,

  e.g. if one card is drawn from a pack, and event A is that a red card is drawn while event B is that a club (♣) card is drawn, then B and A are mutually exclusive.

- We say two events A and B are *independent* if $P(A|B) = P(A)$,

  e.g. event B is tossing a coin and getting heads; event A is tossing the coin a second time and getting tails.

- Be careful, usually $P(A|B) \neq P(B|A)$.

As an example, we will find the probability that the single throw of a fair die results in a number less than four, given that the throw resulted in an odd number. Let B be the event of the throw being less than four, i.e. $B = \{1, 2, 3\}$, then it follows that

$$P(B) = \frac{3}{6} = \frac{1}{2} \,. \tag{288}$$

Let A be the event that the throw of the die is an odd number, i.e. $A = \{1, 3, 5\}$, then $P(A) = 1/2$ as well.

The event $B \cap A$ is a throw which is less than 4 **and** odd, i.e. $B \cap A = \{1, 3\}$, so that

$$P(B \cap A) = \frac{2}{6} = \frac{1}{3} \,. \tag{289}$$

It then follows from (287) that

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{1/3}{1/2} = \frac{2}{3} \,. \tag{290}$$

<u>Example 7.4</u> Consider drawing 2 balls out of a bag of 5 balls: 1 red, 2 green, 2 blue. What is the probability of drawing a blue ball out of the bag given that the first ball was blue if: (i) the first ball is replaced; (ii) the first ball is not replaced.

Part (i) is relatively easy. If the first blue ball is replaced then there are two blue balls in a bag of five balls, and hence the probability of drawing one of blues is

$2/5$. There is no ‘memory‘ of the first draw because the first ball was replaced. And thus there is no need to use the concepts of conditional probability here.

Part (ii), however, does require us to use conditional probability, because the first draw has impacted on subsequent draws. Let us define the following events: event A denotes getting a blue ball in the first draw. Thus $P(A) = 2/5$ (there are two blue balls out of five); event B denotes getting a blue ball in the second draw. The question asks us to find $P(B|A)$.

The event that both balls drawn are blue is just $B \cap A$, and its probability is simply $P(B \cap A) = (2/5) \times (1/4) = 1/10$, i.e. the product of getting a blue in the first draw (2/5) and of getting a blue in the second (1/4). The conditional probability (which is different to the probability of the intersection!), can then be computed

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{1/10}{2/5} = \frac{1}{4}.$$

7.2.1 Bayes' Theorem

We now come to the most important result in conditional probability, known as Bayes' Theorem.

The probability of A occurring given B is $P(A|B)$, cf. equation (287). On the other hand, the probability of event B occurring given that A has already occurred is $P(B|A)$, with

$$P(B|A) = \frac{P(B \cap A)}{P(A)} . \tag{291}$$

We then have

$$\begin{aligned} P(A \cap B) &= P(B|A)P(A) \\ P(B \cap A) &= P(A|B)P(B) . \end{aligned} \tag{292}$$

Furthermore, we know that $P(A \cap B) = P(B \cap A)$, and using (292) this leads to

$$\boxed{P(A|B) = \frac{P(A)P(B|A)}{P(B)}\ ,} \tag{293}$$

provided that $P(B) \neq 0$. This is Bayes' Theorem.

An alternative form of Bayes' Theorem can be written down by first using the *law of total probability*

$$P(B) = P(B|A)P(A) + P(B|\overline{A})P(\overline{A})\ , \tag{294}$$

which we now discuss.

This law holds because there are **two** different possible routes to getting event B.

1. The first route is that event A (probability $P(A)$) happens and then B happens (probability $P(B|A)$). The total probability for this route is

$$P(A) \times P(B|A)\ . \tag{295}$$

2. The second route is that event A does not happen (probability $P(\overline{A})$) and then B happens (probability $P(B|\overline{A})$). The total probability for this route is

$$P(\overline{A}) \times P(B|\overline{A})\ . \tag{296}$$

Adding together the probabilities in (295) and (296) then gives (294).

Putting the law of total probability into (293) gives the alternative form of Bayes' Theorem:

$$\boxed{P(A|B) = \frac{P(A)P(B|A)}{P(B|A)P(A) + P(B|\overline{A})P(\overline{A})}\ .} \tag{297}$$

This is sometimes easier to use than (293).

We consider the following example:

A screening test is 99% effective in detecting a certain disease when a person has the disease. The test yields a 'false positive' for 1% of healthy persons tested. If 0.1% of the population have the disease then what is the probability that a person whose test is positive has the disease?

Let the event A be that a person has the disease, so that $P(A) = 0.001$, and the probability that they do not have the disease is $P(\overline{A}) = 1 - P(A) = 0.999$. Let the event B be a positive test, so that $P(B|A) = 0.99$ (i.e. the probability of the test successfully detecting the disease) and $P(B|\overline{A}) = 0.01$ (i.e. the probability of a positive test on a healthy person).

From (297) we therefore have

$$
\begin{aligned}
P(A|B) &= \frac{P(A)P(B|A)}{P(B|A)P(A) + P(B|\overline{A})P(\overline{A})} \\
&= \frac{0.001 \times 0.99}{0.99 \times 0.001 + 0.01 \times 0.999} \\
&= 0.0902
\end{aligned}
$$

or roughly 9%. So the probability of someone who tests positive actually having the disease is rather low! This has arisen because the probability of a false positive, 0.01, is large compared to the probability of having the disease, 0.001.

<u>Example 7.5</u> Calculate the probability that someone who tests negative actually has the disease after all, $P(A|\overline{B})$.

So we apply Bayes' Theorem again so that

$$
P(A|\overline{B}) = \frac{P(A)P(\overline{B}|A)}{P(\overline{B}|A)P(A) + P(\overline{B}|\overline{A})P(\overline{A})}.
$$

To find $P(\overline{B}|A)$ we note that $P(\overline{B}|A) + P(B|A) = 1$, since if A happens there are only two mutually exclusive possibilities B and $\overline{B}$. We know that $P(B|A) = 0.99$, therefore $P(\overline{B}|A) = 1 - P(B|A) = 1 - 0.99 = 0.01$.

Similarly, $P(\overline{B}|\overline{A}) + P(B|\overline{A}) = 1$, and thus $P(\overline{B}|\overline{A}) = 1 - P(B|\overline{A}) = 1 - 0.01 = 0.99$.

Putting everything together now, we get

$$P(A|\overline{B}) = \frac{0.001 \times 0.01}{0.001 \times 0.01 + 0.99 \times 0.999} = \frac{0.00001}{0.98901} \approx 10^{-5},$$

which is quite a small probability.

7.3 Combinatorics

We often worry about problems which involve completing a sequence of actions or arranging/grouping elements, e.g. tossing a coin 10 times, drawing coloured balls from a bag, or arranging N indistinguishable particles into R energy levels (Bose-Einstein or Fermi-Dirac statistics).

When the order of the actions/elements matters, then we are dealing with *permutations*.

When the order of the actions/elements does not matter, then we are dealing with *combinations*.

7.3.1 Permutations

Consider n distinguishable objects, and let us select r of them without replacement and put them in order. This is a permutation.

How many different permutations are possible?

- The first time we select an object, there are n possible choices.
- The second time there are only $n - 1$ choices

- The third time there are only $n-2$ choices, and on and on until we have taken r objects.

It follows that the total number of possibilities is the product of all the choices

$$n(n-1)(n-2)\dots(n-r+1) = \frac{n!}{(n-r)!} \equiv {}^{n}P_{r}\,. \tag{298}$$

We have that ${}^{n}P_{n} = n!$, since $0! \equiv 1$.

A simple example: how many ways can all the club (♣) cards be arranged in a line?

There are $n = 13$ cards to draw from. And we are going to arrange all of them, so $r = 13$. Therefore the number of permutations is

$$^{13}P_{13} = 13! = 6,227,020,800.$$

So quite a few.

How about the number of ways to order 4 cards from the club suite?

That would be ${}^{13}P_{4} = 13!/9! = 17,160$.

Example 7.6 What is the probability that in a room of N people at least 2 have the same birthday? Take $N = 200$.

First, let us take the number of days in the year to be 365, for simplicity. Next, let A denote the event that two or more people have the same birthday. It will turn out, however, to be easier to work with its complement $\overline{A}$: the event that everyone in the room has a different birthday. Consider the number of possible permutations of *different* birthdays amongst the N people: in other words, how many lists of N different numbers can be made out of a choice of 365? This is equal to

$$^{365}P_{N} = 365 \times 364 \times 363 \times \cdots \times (365 - N + 1) = \frac{365!}{(365-N)!}.$$

On the other hand, the total number of different birthday distributions amongst the N people (allowing repetitions) is just

$$365 \times 365 \times 365 \times \cdots \times 365 = 365^N.$$

The probability $P(\overline{A})$ is the division of the first number by the second:

$$P(\overline{A}) = \frac{365!}{(365 - N)!365^N}.$$

Now we can compute $P(A)$, via

$$P(A) = 1 - P(\overline{A}) = 1 - \frac{365!}{(365 - N)!365^N}.$$

As N gets large the probability rapidly approaches 1. For example, if $N = 23$, then $P(A) \approx 1/2$. If $N = 50$, then $P(A) \approx 0.97$. Finally, when $N = 200$, we get $P(A) \approx 1 - 10^{-30} \approx 1$. In the last case it is guaranteed that at least two people share a birthday.

7.3.2 Combinations

Let us, as before, select r objects from a set of n, but let us not care about the ordering of the selection. (For example, how many hands of 5 cards are there from a full deck of 52 cards?) This is a *combination*. How many different combinations are there?

- So there are still ${}^nP_r$ ordered arrangements, as before, but we do not care about the ordering.
- Each group of r objects can be ordered $r!$ ways, but we do not want to count all of these different orderings.
- Hence we divide through ${}^nP_r$ by $r!$ to get the number of combinations:

$$\frac{{}^nP_r}{r!} = \frac{n!}{(n - r)!r!} \equiv {}^nC_r \; . \tag{299}$$

As an example, suppose we draw that hand of 5 cards from a deck of 52 cards. The number of different hands possible is

$$^{52}C_5 \equiv \frac{52!}{(52-5)!5!} = \frac{52 \times 51 \times 50 \times 49 \times 48}{120} = 2,598,960 \ . \tag{300}$$

Example 7.7 Out of 10 physics professors and 12 chemistry professors, a committee of 5 people must be chosen in which each subject has at least 2 representatives. In how many ways can this be done?

First thing to note is that there are going to be two kinds of committee: one with 3 physicists plus 2 chemists, and one with 2 physicists plus 3 chemists. We deal with each kind separately.

The first kind of committee has 3 physicists, which must be chosen from a pool of 10. But the ordering doesn't matter. Therefore the number of ways of choosing the 3 is $^{10}C_3 = 10 \times 9 \times 8/6 = 120$. On the other hand, the number of ways of choosing the 2 chemists from a pool of 12 is $^{12}C_2 = 12 \times 11/2 = 66$. Putting this information together, the total number of different committees of the first kind is $120 \times 66 = 7920$.

The second kind of committee has only 2 physicists. The number of ways of choosing these 2 is $^{10}C_2 = 45$. The number of ways of choosing 3 chemists is $^{12}C_3 = 220$. Therefore the total number of different committees of the second kind is $45 \times 220 = 9900$.

Finally, the total number of committees of the two kinds allowed = $7920+9900 = 17820$.

Binomial coefficients

The numbers ^nC_r are called the binomial coefficients, because they arise in the binomial expansion. In particular,

$$\boxed{(p+q)^n = \sum_{r=0}^{n} {}^nC_r q^r p^{n-r} \ .} \tag{301}$$

They have the interesting property that

$${}^nC_r = {}^{n-1}C_r + {}^{n-1}C_{r-1} \ . \tag{302}$$

To prove this recursion relation note that

$$\begin{aligned} {}^{n-1}C_r + {}^{n-1}C_{r-1} &= \frac{(n-1)!}{r!(n-1-r)!} + \frac{(n-1)!}{(r-1)!(n-r)!} \\ &= \frac{(n-1)!(n-r) + (n-1)!r}{r!(n-r)!} \\ &= \frac{(n-1)!(n-r+r)}{r!(n-r)!} = \frac{(n-1)!n}{r!(n-r)!} \\ &= \frac{n!}{r!(n-r)!} = {}^nC_r \ . \end{aligned} \tag{303}$$

7.4 Random variables

A **random variable** X is a variable whose value depends on the outcomes of an experiment involving some level of randomness or chance.

A random variable may take discrete values (e.g. the outcome of coin tosses) or it may take a range of continuous values (e.g. the mass of newborn babies).

- For example, let us toss a coin 3 times (our experiment) and take X to represent the number of heads (the random variable).

  The sample space of outcomes is

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

and the value of X can be 0, 1, 2 or 3. In this case, the variable X is **discrete** because it can only take discrete values 0,1,2,3. Note that more than one outcome can give the same value of the random variable.

- A different example: consider an aircraft flying through some particularly nasty clear-air turbulence that causes its velocity $\boldsymbol{v}$ to randomly vary. If we take X to be the instantaneous speed of the plane $|\boldsymbol{v}|$ we see that X can take **continuous** values between 0 and ∞, in principle. Again different outcomes can yield the same X (e.g. a sharp deviation up and a sharp and equal deviation down both generate the same speed).

7.5 Discrete probability distributions

Consider a discrete random variable X. For each value X takes, we can assign a probability.

If X takes the discrete values x_i, which have associated probabilities p_i, for $i = 1, 2, ..n$, then $P(X = x_i) = p_i$.

We can then construct a **probability function**, also called a **probability distribution**, usually just denoted $P(X)$, which is simply the probability of any event associated with X in S.

This function is normalised, as expected, according to

$$\sum_{i=1}^{n} P(X = x_i) = \sum_{i=1}^{n} p_i = 1 \ . \tag{304}$$

As an example, let us return to the coin-tossing game earlier:

- $X = 0$ corresponds only to TTT, so $P(X = 0) = 1/8$,
- $X = 1$ corresponds to the event $\{HTT, THT, TTH\}$, so $P(X = 1) = 3/8$
- $X = 2$ yields $P(X = 2) = 3/8$, and $X = 3$ yields $P(X = 3) = 1/8$.

The **cumulative probability function** (CPF), $F(x)$, is the probability that X takes a value which is less than or equal to x, i.e.

$$F(x) = P(X \leqslant x) = \sum_{x_i \leqslant x} P(X = x_i) \ . \tag{305}$$

So for a discrete random variable the CPF will be a series of steps at each value x_i, and will be constant between the steps. Note how the ultimate value of $F(x) \to 1$ as $x \to \infty$ (in the coin tossing game, $x \geqslant 3$).

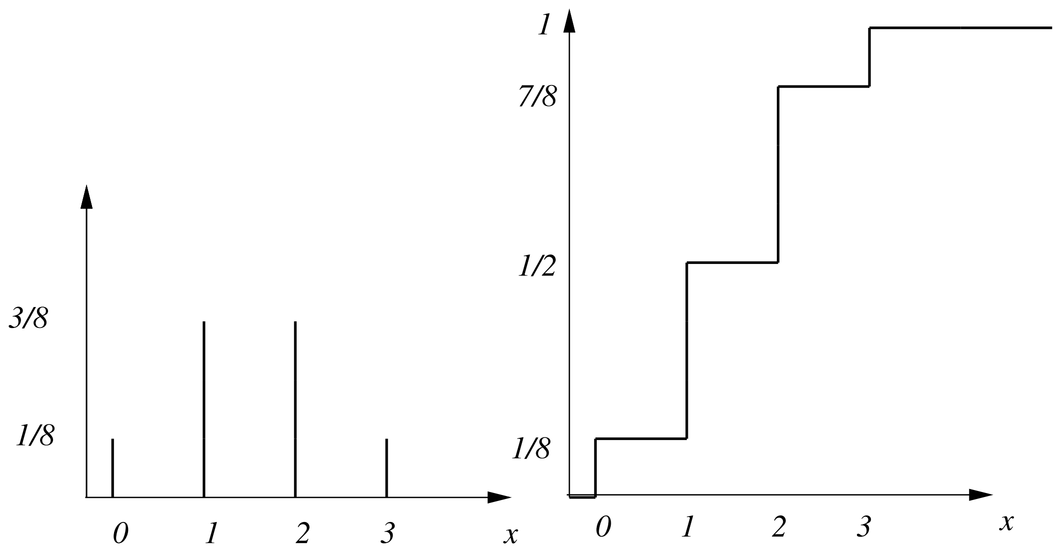

Figure 47: Graphs of $P(X)$ and $F(x)$ for the coin-tossing example.

<u>Example 7.8</u> A bag contains 6 blue balls and 4 red balls. Three balls are drawn without replacement. Find the probability function for the number of red balls drawn.

Let X be our random variable, corresponding to the number of red balls drawn. We now construct the probability distribution $P(X)$ piece by piece, by looking at $X = 0, 1, 2$ and 3 separately.

- We have $X = 3$ when we draw a red on each draw, i.e. 'RRR'. The probability of this is $4/10$ times $3/9$ times $2/8$, which is equal to 1/30. Thus $P(3) = 1/30$.
- Three different kinds of draw yield $X = 2$: RRB, RBR, and BRR (where B denotes a blue draw). The probability of RRB is $4/10 \times 3/9 \times 6/8 = 1/10$. The probability of RBR is $4/10 \times 6/9 \times 3/8 = 1/10$. The probability of BRR is $6/10 \times 4/9 \times 3/8 = 1/10$. Thus $P(2) = 1/10 + 1/10 + 1/10 = 3/10$.
- There are also three different kinds of draw to give us $X = 1$: RBB, BRB, and BBR. The probability of RBB is $4/10 \times 6/9 \times 5/8 = 1/6$. The probability of BRB is also 1/6, as is the probability of BBR. Therefore $P(1) = 1/6 + 1/6 + 1/6 = 1/2$.
- Finally, we examine $X = 0$, which corresponds only to one kind of draw: BBB. Its probability is $6/10 \times 5/9 \times 4/8 = 1/6$.

In summary, our probability distribution $P(X)$ is defined through:

$$P(3) = \frac{1}{30}, \quad P(2) = \frac{3}{10}, \quad P(1) = \frac{1}{2}, \quad P(0) = \frac{1}{6}.$$

Just to check that it is normalised appropriately, we look at the sum

$$\sum_{X=0}^{3} P(X) = \frac{1}{6} + \frac{1}{2} + \frac{3}{10} + \frac{1}{30} = \frac{5 + 15 + 9 + 1}{30} = 1,$$

and all is good.

7.5.1 Mean and Variance

The **mean** of the random variable X is defined to be

$$\boxed{\mathbb{E}[X] \equiv \sum_{i=1}^{n} x_i p_i = \sum_{i=1}^{n} x_i P(X = x_i)\ .} \tag{306}$$

The mean is also often referred to as the **expectation value**, and the alternative notations $E[X]$, $\langle X \rangle$, $\overline{X}$, or μ are often used.

If an experiment is repeated a very large number of times then the average value of the associated random variable X will approach the mean (cf. the Law of Large Numbers).

For example, consider the three coin tosses in the previous subsection:

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i=1}^{n} x_i P(X = x_i) \\
&= 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} \\
&= \frac{3}{2}\ .
\end{aligned}$$

The mean has the following properties:

1. $\mathbb{E}[aX] = a\mathbb{E}[X]$, where a is a constant;

2. If X and Y are two different random variables (possibly with different probability functions), then

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]\ ; \tag{307}$$

3. If $g(X)$ is a function of the random variable X then

$$\mathbb{E}[g(X)] = \sum_{i=1}^{n} g(x_i) P(X = x_i)\ . \tag{308}$$

We are often interested in the way in which results are spread around the mean. One measure of this is the **variance** of X, which we define to be

$$\boxed{\text{var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]\ .} \qquad (309)$$

In other words, the variance is the mean value of the square of the deviation from the mean.

The **standard deviation** σ is the square root of the variance, i.e.

$$\boxed{\sigma^2 = \text{var}(X)}. \qquad (310)$$

Expanding the bracket in (309) leads to

$$\begin{aligned} \sigma^2 &= \mathbb{E}[X^2 - 2X\mu + \mu^2] \\ &= \mathbb{E}[X^2] - \mathbb{E}[2\mu X] + \mathbb{E}[\mu^2] \\ &= \mathbb{E}[X^2] - 2\mu\mathbb{E}[X] + \mu^2 \quad \text{(since } \mu \text{ is a constant)} \\ &= \mathbb{E}[X^2] - 2\mu\mu + \mu^2 \\ &= \mathbb{E}[X^2] - \mu^2\ . \end{aligned} \qquad (311)$$

From (311) we are therefore left with the very useful result that the variance σ^2 and the mean μ are related by

$$\boxed{\sigma^2 = \mathbb{E}[X^2] - \mu^2\ .} \qquad (312)$$

Note that in this expression:

$$\mathbb{E}[X^2] = \sum_{i=1}^{n} x_i^2 P(X = x_i)\ . \qquad (313)$$

For our experiment of tossing three coins with X being the number of heads, we have

$$\begin{aligned} \mathbb{E}[X^2] &= 0^2 \times \frac{1}{8} + 1^2 \times \frac{3}{8} + 2^2 \times \frac{3}{8} + 3^2 \times \frac{1}{8} \\ &= 3\ , \end{aligned}$$

so the variance is given as

$$\sigma^2 = \mathbb{E}[X^2] - \mu^2 = 3 - \left(\frac{3}{2}\right)^2 = \frac{3}{4} \,. \tag{314}$$

7.5.2 Binomial distribution

The binomial distribution arises when an experiment has only 2 possible outcomes, e.g. a single coin toss (heads or tails), probability of surviving a heart attack (yes or no), asking a random person if they can drive a tractor (yes or no).

- Let event A, labeled a 'success', denote one outcome and $B = \overline{A}$, labeled a 'failure', the other outcome.
- If the probability of event A is p then the probability of event B is $q = 1 - p$.
- Suppose that the experiment is repeated n times, and let the discrete random variable X be the number of successes, so that X takes one of the values $0, 1, 2, ..., n$. The probability distribution $P(X)$ may be written as

  $$P(X = r) \equiv B(n, p) = {}^nC_r p^r (1 - p)^{n-r} \,. \tag{315}$$

  This is the **binomial distribution**.

How did we get to the second formula?

- The sample space of n experiments contains 2^n outcomes.
- Each outcome that has r successes has a probability $p^r q^{n-r}$.
- But there are lots of outcomes with $X = r$. We need to add them all up to get our correct probability:

  $$P(X = r) = \overbrace{p^r q^{n-r} + p^r q^{n-r} + \cdots + p^r q^{n-r}}^{M} = M \cdot p^r q^{n-r}$$

  How many of them are there? That is, what is M?

- Actually, M is simply the number of ways of choosing r elements from a set of n, i.e. ${}^nC_r$.

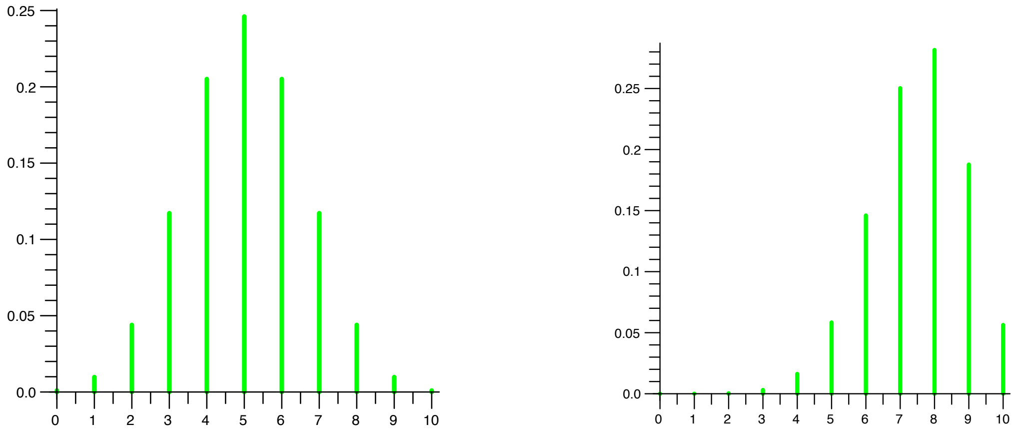

Figure 48: Binomial distributions: $B(10, 0.5)$ (left) and $B(10, 0.75)$ (right).

We can check that the binomial distribution satisfies the normalisation condition (304):

$$\sum_{r=0}^{n} P(X = r) = \sum_{r=0}^{n} {}^nC_r p^r (1-p)^{n-r} = (p + 1 - p)^n = 1 \ . \tag{316}$$

Mean

We can calculate the mean of the binomial distribution:

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{r=0}^{n} r\left[\, {}^{n}C_{r} p^{r}(1-p)^{n-r}\right] \\
&= \sum_{r=0}^{n} r\frac{n!}{r!(n-r)!}p^{r}(1-p)^{n-r} \\
&= \sum_{r=1}^{n} \frac{n!}{(r-1)!(n-r)!}p^{r}(1-p)^{n-r} \quad \text{(the } r=0 \text{ term is zero)}
\end{aligned}$$

$$\begin{aligned}
&= n\sum_{r=1}^{n} \frac{(n-1)!}{(r-1)!(n-r)!}p^{r}(1-p)^{n-r} \\
&= n\sum_{s=0}^{n-1} \frac{(n-1)!}{s!(n-1-s)!}p^{s+1}(1-p)^{n-1-s} \quad \text{(writing } r=s+1\text{)} \\
&= np\sum_{s=0}^{n-1} \frac{(n-1)!}{s!(n-1-s)!}p^{s}(1-p)^{n-1-s} \\
&= np(1-p+p)^{n-1} = np\,. \qquad (317)
\end{aligned}$$

So the mean of the binomial distribution is $\mu = np$.

Variance

To find the variance of the binomial distribution we first write

$$\begin{aligned}\mathbb{E}(X^2) &= \sum_{r=0}^{n} r^2 \left[{}^nC_r p^r (1-p)^{n-r}\right] \\ &= \sum_{r=0}^{n} r(r-1)\left[{}^nC_r p^r (1-p)^{n-r}\right] + \sum_{r=0}^{n} r\left[{}^nC_r p^r (1-p)^{n-r}\right] \qquad (318)\end{aligned}$$

The second term is exactly $\mathbb{E}(X)$ (see first line of 317), and is therefore np. The first term in (318) is

$$\begin{aligned} &= \sum_{r=0}^{n} r(r-1)\frac{n!}{r!(n-r)!}p^r(1-p)^{n-r} \\ &= \sum_{r=2}^{n} \frac{n!}{(r-2)!(n-r)!}p^r(1-p)^{n-r} \quad \text{(the } r = 0, 1 \text{ terms are zero)} \\ &= n(n-1)p^2 \sum_{r=2}^{n} \frac{(n-2)!}{(r-2)!(n-r)!}p^{r-2}(1-p)^{n-r} \\ &= n(n-1)p^2 \sum_{s=0}^{n-2} \frac{(n-2)!}{(s)!(n-2-s)!}p^{s}(1-p)^{n-2-s} \quad \text{(writing } r = s+2\text{)} \\ &= n(n-1)p^2(1-p+p)^{n-2} = n(n-1)p^2 \; . \qquad (319)\end{aligned}$$

Hence, back in (318) we have

$$\mathbb{E}(X^2) = n(n-1)p^2 + np \; , \qquad (320)$$

and using (312) and (318) we have the variance of the binomial distribution as

$$\sigma^2 = n(n-1)p^2 + np - (np)^2 = np(1-p) \; . \qquad (321)$$

So the standard deviation of the binomial distribution is $\sigma = \sqrt{np(1-p)}$.

Interestingly, as the sample size, n, gets larger the mean of X increases like n but the standard deviation, σ, increases less rapidly, like $\sqrt{n}$. Thus the relative width of the distribution $\sigma/\mu \propto 1/\sqrt{n}$, and it gets narrower and narrower the more experiments we do.

Thus many experimental measurements suppress random errors.

7.5.3 Poisson distribution

The Poisson distribution arises when the number of 'successes' in a given event is unlimited. For instance, what is the distribution of the number of photons received by a telescope per minute? How many golfers are struck by lightning every year? The number of goals scored in a football match?

If X is the number of successes (photons received, lightning strikes, etc.), then it turns out that

$$\boxed{P(X = r) = \frac{\lambda^r \exp(-\lambda)}{r!}\ ,} \tag{322}$$

which is the **Poisson distribution** with parameter λ. Here, r is the number of event occurrences, taking possible values $r = 0, 1, 2, 3, ...$ right up to infinity.

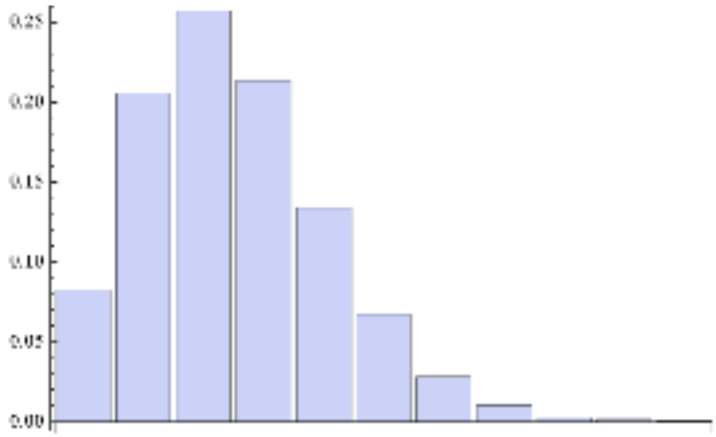

Figure 49: Poisson distribution with $\lambda = 2.5$

We can check that the Poisson distribution satisfies the normalisation condition

(304):

$$\begin{aligned}\sum_{r=0}^{\infty} P(X=r) &= \sum_{r=0}^{\infty} \frac{\lambda^r \exp(-\lambda)}{r!} \\ &= \exp(-\lambda) \sum_{r=0}^{\infty} \frac{\lambda^r}{r!} \\ &= \exp(-\lambda)\exp(\lambda) \quad . \\ &= 1\end{aligned}$$

Mean

The mean of the Poisson distribution is

$$\begin{aligned}\mathbb{E}(X) &= \sum_{r=0}^{\infty} r\frac{\lambda^r \exp(-\lambda)}{r!} \\ &= \exp(-\lambda)\sum_{r=1}^{\infty}\frac{\lambda^r}{(r-1)!} \quad \text{(the } r=0 \text{ term is zero and can be dropped)} \\ &= \exp(-\lambda)\lambda\sum_{r=1}^{\infty}\frac{\lambda^{r-1}}{(r-1)!} \\ &= \lambda\exp(-\lambda)\sum_{s=0}^{\infty}\frac{\lambda^s}{s!} \quad \text{(writing } r=s+1\text{)} \\ &= \lambda\exp(-\lambda)\exp(\lambda) \\ &= \lambda\ . \end{aligned} \tag{323}$$

So the mean of the Poisson distribution is λ.

Variance

To calculate the variance, it is easiest to find $\mathbb{E}[X^2 - X]$ first:

$$\begin{aligned}
\mathbb{E}[X^2 - X] &= \mathbb{E}[X(X-1)] \\
&= \sum_{r=0}^{\infty} r(r-1)\frac{\lambda^r \exp(-\lambda)}{r!} \\
&= \sum_{r=2}^{\infty} \frac{\lambda^r \exp(-\lambda)}{(r-2)!} \quad \text{(the } r = 0, 1 \text{ terms are zero)} \\
&= \exp(-\lambda)\lambda^2 \sum_{r=2}^{\infty} \frac{\lambda^{r-2}}{(r-2)!} \\
&= \exp(-\lambda)\lambda^2 \sum_{s=0}^{\infty} \frac{\lambda^s}{s!} \quad \text{(writing } r = s + 2) \\
&= \lambda^2 \quad . \qquad (324)
\end{aligned}$$

Hence,

$$\mathbb{E}[X^2] = \mathbb{E}[X^2 - X] + \mathbb{E}[X] = \lambda^2 + \lambda\ , \qquad (325)$$

having used (323) for the value of $\mathbb{E}[X]$. Hence, from (312)

$$\sigma^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \lambda\ . \qquad (326)$$

The Poisson and Binomial distributions are related. In fact, the Poisson distribution is the Binomial distribution in the limit $n \to \infty$, $p \to 0$ but with $np = \lambda$ held fixed, see Figure 50.

7.6 Continuous probability distributions

We now consider a random variable X which can take any value in a continuous range, in general $-\infty < X < \infty$.

Because X is continuous we cannot assign a probability to X taking a single value $X = x$, but instead we can define the probability that X takes a value in an infinitesimally small interval, i.e. $x \leqslant X \leqslant x + \mathrm{d}x$.

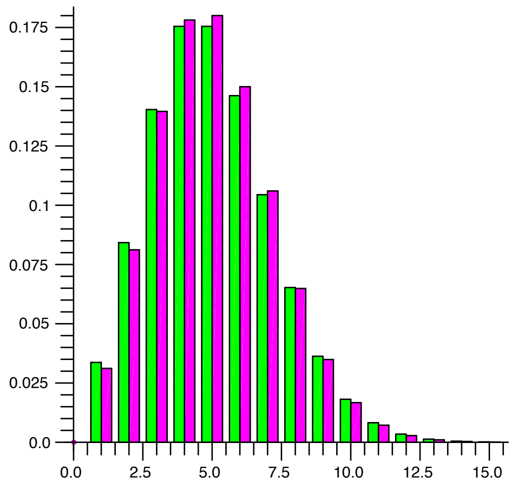

Figure 50: Binomial distribution $B(100, 0.05)$ (dark/magenta) and Poisson distribution with $\lambda = 5$ (light/green). Here $n = 100$ is large, but $np = \lambda$ is fixed.

We hence define the **probability density function** (PDF) $f(x)$, such that

$$P(x \leqslant X \leqslant x + \, dx) = f(x) \, dx \; . \tag{327}$$

- The probability that X takes a value in the **finite** range $\alpha \leqslant X \leqslant \beta$ is then given by the integral

$$P(\alpha \leqslant X \leqslant \beta) = \int_{\alpha}^{\beta} f(x) \, dx \; . \tag{328}$$

- The PDF for a continuous random variable is the equivalent of the probability function of a discrete random variable. However, $f(x)$ on its own is *not* a probability.

- We have that $f(x) \geq 0$ to ensure that all probabilities are positive. Also $f(x)$ can be larger than 1 over some subset of possible X, but the integral over any such range must be less than 1.

- The PDF must obey the normalisation condition

$$\int_{-\infty}^{\infty} f(x)\ \mathrm{d}x = 1\ , \tag{329}$$

- The cumulative probability function (CPF), $F(x)$, is defined to be the probability that $X \leqslant x$,

$$F(x) = P(X \leqslant x) = \int_{-\infty}^{x} f(x)\ \mathrm{d}x\ . \tag{330}$$

  We can see straightaway that $dF/dx = f(x)$.

7.6.1 Mean and variance

The mean and variance of a continuous random variable can be defined by a simple modification of the definitions for a discrete random variable.

Specifically, the mean is given by

$$\boxed{\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x)\ \mathrm{d}x\ .} \tag{331}$$

To calculate the variance, σ^2, we first find

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\ \mathrm{d}x\ , \tag{332}$$

and then use the result (312), i.e.

$$\sigma^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2\ . \tag{333}$$

As an example, consider the **uniform distribution** for which the random variable X is uniformly distributed between $X = 0$ and $X = 1$, i.e. takes any value in that range equally regularly. This is finite-bandwidth *white noise*, manifesting in, for example, a random number generator on a computer, quantisation error when transferring analogue to digital signals, and is also an approximation to the sound of a crashing cymbal (at least in old school drum machines).

- The PDF is hence $f(x) = 1$ for $0 \leq x \leq 1$, and 0 otherwise. Obviously $f(x)$ is correctly normalised because $\int_{-\infty}^{\infty} f(x)dx = \int_0^1 1\,dx = [x]_0^1 = 1$.
- The mean should be $1/2$ by symmetry, but let us check the formula:

$$\mu = \int_0^1 x dx = \frac{1}{2}\left[x^2\right]_0^1 = \frac{1}{2}. \tag{334}$$

- What about the variance?

$$\sigma^2 = \langle X^2 \rangle - \mu^2 = \int_0^1 x^2 dx - \left(\frac{1}{2}\right)^2 = \frac{1}{12}. \tag{335}$$

Example 7.9 Find the mean and variance of the exponential distribution

$$f(x) = \begin{cases} \lambda \exp(-\lambda x) & x \geqslant 0 \\ 0 & x < 0\,. \end{cases} \tag{336}$$

What is the probability that X takes a value in excess of two standard deviations from the mean?

We begin by computing the mean μ. Integration by parts helps us here:

$$\begin{aligned} \mu = \int_{-\infty}^{\infty} x f(x)\,dx &= \int_0^{\infty} \lambda x e^{-\lambda x}\,dx, \\ &= \left[-x e^{-\lambda x}\right]_0^{\infty} + \int_0^{\infty} e^{-\lambda x}\,dx, \\ &= 0 + \left[-\frac{1}{\lambda} e^{-\lambda x}\right]_0^{\infty} = \frac{1}{\lambda}. \end{aligned}$$

Next we compute the variance $\sigma^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2$. First we work out the first term on the right side:

$$\begin{aligned} \mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx &= \int_0^{\infty} \lambda x^2 e^{-\lambda x} dx, \\ &= \left[-x^2 e^{-\lambda x}\right]_0^{\infty} + 2\int_0^{\infty} x e^{-\lambda x}\,dx \\ &= 0 + \frac{2}{\lambda}\mu = \frac{2}{\lambda^2}. \end{aligned}$$

Therefore, $\sigma^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = 2/\lambda^2 - 1/\lambda^2 = 1/\lambda^2$.

One standard deviation is $\sigma = 1/\lambda$ and the mean is $\mu = 1/\lambda$. We are interested in finding the probability of $|X - \mu| > 2\sigma$, or in other words when $X > 1/\lambda + 2/\lambda = 3/\lambda$ or $X < 1/\lambda - 2/\lambda = -1/\lambda$. From the distribution function we see that there is zero probability of getting a negative value for X, so we discount the second case. The probability of the first case is:

$$P\left(X > \frac{3}{\lambda}\right) = \int_{3/\lambda}^{\infty} \lambda e^{-\lambda x}\, dx = \left[-e^{-\lambda x}\right]_{3/\lambda}^{\infty} = e^{-3} = 0.0498.$$

7.6.2 The normal distribution

The normal (or Gaussian or bell curve) distribution is the most important distribution in statistics. It is ubiquitous in science, as a consequence of the *central limit theorem*: the average of a huge number of random and independent experiments will be distributed increasingly like a normal distribution.

The normal distribution is defined by

$$\boxed{f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right].} \tag{337}$$

where μ is the mean and σ^2 is variance. It is often denoted $N(\mu, \sigma^2)$.

To handle the normal distribution we need the following pieces of information:

$$\int_{-\infty}^{\infty} \exp(-x^2)\, \mathrm{d}x = \sqrt{\pi}\,, \tag{338}$$

a result which can be proved using double integrals (see next term);

$$\int_{-\infty}^{\infty} x \exp(-x^2)\, \mathrm{d}x = 0\,, \tag{339}$$

a result which follows straight from the fact that the integrand is an odd function; and

$$\int_{-\infty}^{\infty} x^2 \exp(-x^2)\, \mathrm{d}x = \frac{\sqrt{\pi}}{2}\,, \tag{340}$$

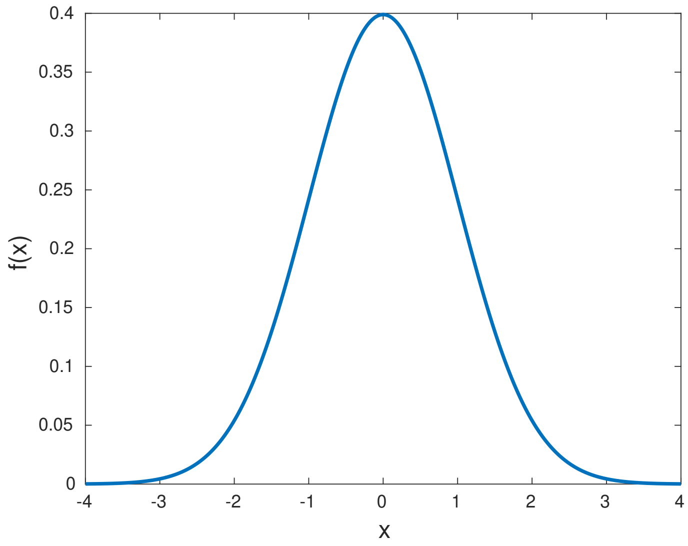

Figure 51: The normal distribution with $\mu = 0$ and $\sigma = 1$.

a result which follows by writing the integrand as $x.x\exp(-x^2)$, integrating by parts and then using (338).

We first check the normalisation condition (329):

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x &= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]\,\mathrm{d}x\\ &= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp(-y^2)\sqrt{2}\sigma\,\mathrm{d}y\\ &= 1 \quad .\end{aligned} \tag{341}$$

Mean

The mean follows from

$$
\begin{aligned}
\mathbb{E}(X) &= \int_{-\infty}^{\infty} x f(x)\ \mathrm{d}x \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]\ \mathrm{d}x \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu+\mu) \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]\ \mathrm{d}x \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \mu \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]\ \mathrm{d}x \quad \text{since the term with } x-\mu \text{ is odd} \\
&= \mu \int_{-\infty}^{\infty} f(x)\ \mathrm{d}x \\
&= \mu \quad \text{using (341)}\ .
\end{aligned}
\tag{342}
$$

Variance

To find the variance we first compute the expectation value of X^2 as

$$
\begin{aligned}
\mathbb{E}(X^2) &= \int_{-\infty}^{\infty} x^2 f(x)\ \mathrm{d}x \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]\ \mathrm{d}x \\
&= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} (\mu+\sqrt{2}\sigma y)^2 \exp(-y^2)\ \mathrm{d}y \quad \text{using substitution } x-\mu = \sqrt{2}\sigma y \\
&= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} (\mu^2 + 2\sqrt{2}\sigma\mu y + 2\sigma^2 y^2) \exp(-y^2)\ \mathrm{d}y \\
&= \frac{1}{\sqrt{\pi}} \left[\mu^2\sqrt{\pi} + \sigma^2\sqrt{\pi}\right] \quad \text{using (338, 339 and 340)} \\
&= \mu^2 + \sigma^2\ ,
\end{aligned}
\tag{343}
$$

which indeed proves that the variance is σ^2.

Cumulative probability distribution

The cumulative probability function for the normal distribution is

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right] \mathrm{d}y \ . \qquad (344)$$

The integral here cannot be written in terms of elementary functions. Instead it introduces a new special function, the *error function*, erf(x), defined to be:

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt,$$

which appears frequently in mathematics and science (especially in problems dealing with diffusion, such as the heat equation). Note that it is an *odd* function.

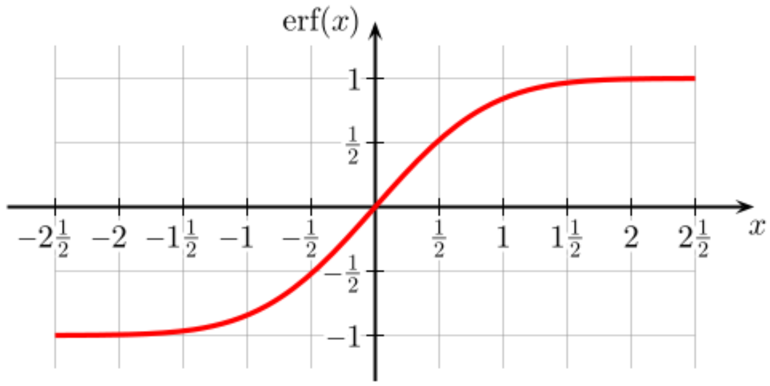

Figure 52: The error function erf(x)

Hence the cumulative probability distribution can be written as

$$F(x) = \frac{1}{2} + \frac{1}{2}\text{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right) \qquad (345)$$

As an example, suppose that a certain manufacturing process produces components whose length is normally distributed with mean 0.5cm and standard deviation 0.005cm, and suppose that a component is rejected if its length differs from the mean by more than 1%.

If we set the random variable X to be the length of the component, then the probability any given component must be rejected is

$$\begin{aligned} P(\text{reject}) = P(X > 0.5 + 0.005) &+ P(X < 0.5 - 0.005) \\ &= 1 - P(X < 0.505) + P(X < 0.495) \\ &= 1 - F(0.505) + F(0.495) \,. \end{aligned} \tag{346}$$

We now bring in the error function and get

$$\begin{aligned} P(\text{reject}) &= 1 - \frac{1}{2} - \frac{1}{2}\text{erf}\left(\frac{0.505 - 0.5}{0.005\sqrt{2}}\right) + \frac{1}{2} + \frac{1}{2}\text{erf}\left(\frac{0.495 - 0.5}{0.005\sqrt{2}}\right), \\ &= 1 - \text{erf}\left(\frac{1}{\sqrt{2}}\right), \end{aligned}$$

where we have used the oddness of the error function to simplify things.

Computing the error function (using tables or Taylor series), gives us $P(\text{reject}) = 0.3174$. So the proportion of rejected items is a massive 31.74%. Probably best to design a better manufacturing process; one that produces the components with a tighter distribution. (Suppose, in fact, that a new process is implemented and its standard deviation is $\sigma = 0.001$ cm; what is $P(\text{reject})$ now?)

Another example: The Higgs boson detection was often quoted as being at the "5-sigma level". This means the following: if we assume that the Higgs *cannot* exist ever, then the detection at CERN was the result of a random fluctuation that was 5 standard deviations away from the mean.

What is the probability that the fluctuation was actually just noise and not the Higgs, assuming the fluctuations were normally distributed?

We want to work out

$$\begin{aligned} P(|X - \mu| > 5\sigma) &= 1 - F(\mu + 5\sigma) + F(\mu - 5\sigma), \\ &= 1 - \text{erf}(5/\sqrt{2}) \approx 5.733 \times 10^{-7}. \end{aligned}$$

So the probability that the detection was just due to random noise was roughly 0.00006%. It is up to us then to decide whether this is good enough to accept that the Higgs was detected.

- The first gravitational wave detection by LIGO in 2015 of a binary black hole merger (GW150914) was regarded as a 5-sigma result. But a more recent detection of a different colliding black hole binary (GW151012) was only 2-sigma, meaning that the probability that the signal was from something else is roughly 5%. Not nearly as compelling.
- In fact, a result to 2-sigma is regarded in many fields as being sufficiently significant as proof, certainly in drug trials. But it is also means that 1 in 20 published results (using this criterion) are probably wrong . . . .
- And all this presupposes that we have a good model for the random (or other) fluctuations in the system. In other words: is the mean, against which we are basing our sigma, well constrained? A famous example: in 2014 the BICEP2 instrument in Antarctica detected a signal interpreted as evidence of gravitational waves from the primordial universe. Researchers claimed the detection was around 6 sigma. In fact, they had not accounted for the effect of dust in the Milky Way which could easily explain their signal within the probabilities, no outrageous fluctuations needed. The 'mean' was not where they thought it was!

Example 7.10 2007 Paper 1 question 4.

(a) The probability of the number n of persons passing a certain checkpoint during a day is

$$P(n;\lambda) = \frac{\lambda^n \exp(-\lambda)}{n!} ,$$

which defines a Poisson distribution with parameter λ. Show that

$$\sum_{n=0}^{\infty} P(n; \lambda) = 1 \ .$$

The probability that any given person is male is p. Show that the probability that k males and l females pass the checkpoint during a day is

$$P(k \text{ males, } l \text{ females}) = \binom{k+l}{l} \frac{p^k (1-p)^l \lambda^{(k+l)} \exp(-\lambda)}{(k+l)!} \ .$$

Hence show that the probability that k males pass (independent of the number of females passing) during the day conforms to a Poisson distribution with parameter λp.

(b) A proportion 0.1 of members of a large population have a certain viral disease, and a further proportion 0.2 are carriers of the virus. A test for the presence of the virus shows positive with probability 0.95 if the person tested has the diseases, 0.9 if the person is a carrier and 0.05 if the person in fact is free of the virus.

Calculate the probability that any given person tests positive.

Calculate the probability that a person who tests negative in fact has the virus (i.e. either has the diseases or is a carrier).

(a) The first part is just asking you to reproduce the notes:

$$\sum_{n=0}^{\infty} P(n;\, \lambda) = \sum_{n=0}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!} = e^{-\lambda} e^{\lambda} = 1,$$

where at the end there we recognised the Taylor series of the exponential function.

In the next part of the question we are interested in the case when $n = k + l$ people pass the checkpoint: k of them are male, and l are female. Putting aside gender, we know that the probability of getting n people past the checkpoint is $P(n;\,\lambda) = \lambda^n e^{-\lambda}/n!$.

Next, we need to multiply this result with the probability that out of these n people, l of them were female. This brings in the binomial distribution, because this is is equivalent to doing n experiments, with each experiment producing one of two outcomes: female or male. We are interested in the probability of getting l 'successes' (i.e. females) in n experiments, given that the probability of getting a single success is $1-p$. The binomial distribution tells us that the probability of this is

$$^{n}C_l(1-p)^l p^{n-l} = ^{k+l} C_l(1-p)^l p^k.$$

Multiplying the result with our precious probability (and setting $n = k+l$) yields

$$P(k \text{ males}, l \text{ females}) = \frac{\lambda^{k+l}e^{-\lambda}}{(k+l)!}\ ^{k+l}C_l\, p^k(1-p)^l.$$

The final part of (a) wants us to work out the probability that k males pass the checkpoint whatever the number of women who passed. One way to compute this is to work out the probability that k men pass and 0 women pass, then add to that the probability that k men pass and 1 woman passes, then add the probability that k men pass and 2 women pass, and so on and so on. We then have, first rewriting $^{k+l}C_l$ in terms of factorials:

$$\begin{aligned}
P(k \text{ males}) &= \sum_{l=0}^{\infty} \frac{\lambda^{k+l}e^{-\lambda}}{(k+l)!}{}^{k+l}C_l p^k(1-p)^l, \\
&= \sum_{l=0}^{\infty} \frac{\lambda^{k+l}e^{-\lambda}}{(k+l)!}\frac{(k+l)!}{k!l!}p^k(1-p)^l, \\
&= \frac{\lambda^k e^{-\lambda}p^k}{k!}\sum_{l=0}^{\infty}\frac{\lambda^l(1-p)^l}{l!} \\
&= \frac{(\lambda p)^k e^{-\lambda}}{k!}e^{\lambda(1-p)},
\end{aligned}$$

where in the last line we recognise that $\sum_{l=0}^{\infty} \lambda^l (1-p)^l / l!$ is the Taylor series of $e^{\lambda(1-p)}$ (treating $\lambda(1-p)$ as the x in e^x).

Finally we get

$$P(k \text{ males}) = \frac{(\lambda p)^k e^{-\lambda p}}{k!},$$

which is the Poisson distribution but with $p\lambda$ instead of λ.

(b) Okay, just to simplify notation, let i indicate 'infected' (i.e. exhibiting the diseases), c indicate 'carrier' (carrying but not exhibiting), and f indicate 'free'. In summary, we then have $P(i) = 0.1$, $P(c) = 0.2$, and $P(f) = 1 - 0.1 - 0.2 = 0.7$.

Next let us denote p as testing positive to the test, and n as testing negative to the test.

The information given us regarding the test can be summarised by:

$$P(p|i) = 0.95, \qquad P(p|c) = 0.9, \qquad P(p|f) = 0.05.$$

And thus $P(n|f) = 1 - 0.05 = 0.95$.

The sets i, c, and f are mutually exclusive, therefore we can write down the total probability of testing positive as:

$$P(p) = P(p \cap i) + P(p \cap c) + P(p \cap f).$$

Using the definition of conditional probability this may be re-expressed as

$$\begin{aligned} P(p) &= P(p|i)P(i) + P(p|c)p(c) + P(p|f)P(f), \\ &= 0.95 \cdot 0.1 + 0.9 \cdot 0.2 + 0.05 \cdot 0.7, \\ &= 0.31 \end{aligned}$$

It also follows that $P(n) = 1 - P(p) = 0.69$.

The last part of the question asks us to find the probability that someone is either a carrier or infected, given that they tested *negative*, in mathematical terms: $P(i \cup c|n)$. To ease notation we write $\overline{f} = i \cup c$. Bayes' law says:

$$P(\overline{f}|n) = \frac{P(n|\overline{f})P(\overline{f})}{P(n)}.$$

We know that $P(\overline{f}) = 0.1 + 0.2 = 0.3$ and $P(n) = 0.69$ but $P(n|\overline{f})$ requires a bit more work. We have the total probability of testing negative as

$$P(n) = P(n \cap f) + P(n \cap \overline{f}) = P(n|f)P(f) + P(n|\overline{f})P(\overline{f}),$$

which gives us

$$P(n|\overline{f})P(\overline{f}) = P(n) - P(n|f)P(f) = 0.69 - 0.95 \cdot 0.7 = 0.025.$$

Therefore

$$P(\overline{f}|n) = \frac{0.025}{0.69} \approx 0.036$$

So about 3.6 %.

Example 7.11 2007 Paper 2 question 3.

(a) The probability of an experiment that involves counting events having the result $N = n$ (where n is a non-negative integer) is

$$P(N = n) = A\rho^n ,$$

where ρ $(0 < \rho < 1)$ is given. Find the normalising constant A. Calculate the probability that $N > n$. Calculate the probability that $N > n$, conditional on $N > m$ $(n > m)$.

(b) The probability density function for a continuous random variable X is

$$f(x) = B\rho^x \equiv B\exp(-\lambda x) \qquad (\lambda = \ln(\rho^{-1}))$$

where x takes values between 0 and ∞. Find the normalising constant B. Calculate the probability that $X > x$, conditional on $X > y$ $(x > y)$. Deduce the probability density function for X, conditional on $X > y$. Calculate the variance of X, conditional on $X > y$.

(a) The total probability should be one, i.e. $\sum_{r=0}^{\infty} P(N = r) = 1$. But

$$\sum_{r=0}^{\infty} P(N = r) = A \sum_{r=0}^{\infty} \rho^r$$

is an infinite geometric series that sums to $A/(1-\rho)$. Hence $A = 1 - \rho$ if the probability distribution is normalised properly.

$P(N > n)$ is just the sum of all probabilities that $N > n$:

$$\begin{aligned}
P(N > n) = \sum_{r=n+1}^{\infty} P(N = r) &= \sum_{r=n+1}^{\infty} (1-\rho)\rho^r \\
&= \sum_{s=0}^{\infty} (1-\rho)\rho^{n+1}\rho^s, \\
&= (1-\rho)\rho^{n+1}\sum_{s=0}^{\infty} \rho^s, \\
&= (1-\rho)\rho^{n+1}\frac{1}{1-\rho} = \rho^{n+1}.
\end{aligned}$$

The last part of this section asks us to find the conditional probability $P(N > n|N > m)$. In other words, given that we have counted m events already, what is the probability that we ultimately count n ($> m$) or, put another way, another $n - m$ events. From the definition of the conditional probability this can be re-expressed as

$$P(N > n|N > m) = \frac{P((N > n) \cap (N > m))}{P(N > m)} = \frac{P(N > n)}{P(N > m)}$$

because $n > m$ and so $(N > n) \cap (N > m) = N > n$. Using the previous part of the problem, we then have

$$P(N > n|N > m) = \frac{P(N > n)}{P(N > m)} = \frac{\rho^{n+1}}{\rho^{m+1}} = \rho^{n-m}.$$

In other words: the probability of just counting another $n-m$ events more than what we have already. The number of previous events counted does not really enter directly; thus the process has no memory.

(b) First we work out the normalisation from

$$\int_0^\infty f(x)\,dx = B\int_0^\infty e^{-\lambda x}dx = B\left[\frac{-e^{\lambda x}}{\lambda}\right]_0^\infty,$$
$$= \frac{B}{\lambda} = 1.$$

Thus $B = \lambda$.

The conditional probability is similar to earlier:

$$P(X > x|X > y) = \frac{P((X > x)\cap(X > y))}{P(X > y)} = \frac{P(X > x)}{P(X > y)}$$

the last equality following from $x > y$. Given that we have counted y events already, what is the probability of counting an additional $x - y$?

Next we work out $P(X > x)$:

$$P(X > x) = \int_x^\infty \lambda e^{-\lambda z}dz = [-e^{-\lambda z}]_x^\infty = e^{-\lambda x}.$$

Thus we have

$$P(X > x|X > y) = e^{-\lambda(x-y)} = P(X > x - y).$$

This shows that the exponential distribution is 'memoryless', as before. It does not matter how many events we have already counted, the clock 'restarts' (in terms of probability) after each count.

Let us define the new 'condtitional' pdf by $g(x)$. Denote its cumulative probability function by $G(x)$. The two are related by $G(x) = \int_0^x g(x)dx$. This definition can be manipulated so that

$$G(x) = 1 - \int_x^\infty f(x)dx = 1 - P(X > x|X > y) = 1 - e^{-\lambda(x-y)}.$$

We then differentiate both sides with respect to x, noting that $dG/dx = g$, which yields $g(x) = \lambda e^{-\lambda(x-y)}$. But note that this is true only for $x > y$. For $x < y$, the pdf is not really defined, and thus on this range we set $g = 0$. In summary

$$g(x) = \begin{cases} \lambda e^{-\lambda(x-y)} & x > y \\ 0 & x < y \end{cases}$$

We first work out the mean:

$$\mathbb{E}(X|X > y) = \int_{-\infty}^{\infty} xg(x)dx = \int_{y}^{\infty} \lambda x e^{-\lambda(x-y)}dx = \int_{0}^{\infty} \lambda(y + z)e^{-\lambda z}dz,$$

where we have made the transformation $z = x - y$. This then gives us

$$\mathbb{E}(X|X > y) = \lambda y \left[\frac{e^{-\lambda z}}{-\lambda}\right]_0^{\infty} + \int_0^{\infty} \lambda z e^{-\lambda z}dz,$$

note that the second term is just the mean of the usual exponential distribution, which we computed earlier in the notes. It is just equal to $1/\lambda$. The first integral is easy to do, and we get finally:

$$\mathbb{E}(X|X > y) = y + \frac{1}{\lambda}.$$

Finally, we compute the variance from the formula

$$\text{Var}(X|X > y) = \mathbb{E}(X^2|X > y) - \mathbb{E}(X|X > y)^2.$$

We have the second term, we now need the first term. Using integration by parts

$$\begin{aligned}
\mathbb{E}(X^2|X > y) = \int_y^{\infty} x^2 \lambda e^{-\lambda(x-y)}dx &= \lambda e^{\lambda y}\left[\frac{-x^2}{\lambda}e^{-\lambda x}\right]_y^{\infty} + \lambda e^{\lambda y}\int_y^{\infty} \frac{2x}{\lambda}e^{-\lambda x}dx, \\
&= y^2 + \frac{2}{\lambda}\mathbb{E}(X|X > y), \\
&= y^2 + \frac{2}{\lambda}\left(y + \frac{1}{\lambda}\right).
\end{aligned}$$

Now putting everything together we have:

$$
\begin{aligned}
\text{Var}(X|X > y) &= \mathbb{E}(X^2|X > y) - \mathbb{E}(X|X > y)^2, \\
&= y^2 + \frac{2}{\lambda}\left(y + \frac{1}{\lambda}\right) - \left(y + \frac{1}{\lambda}\right)^2, \\
&= \frac{1}{\lambda^2}.
\end{aligned}
$$