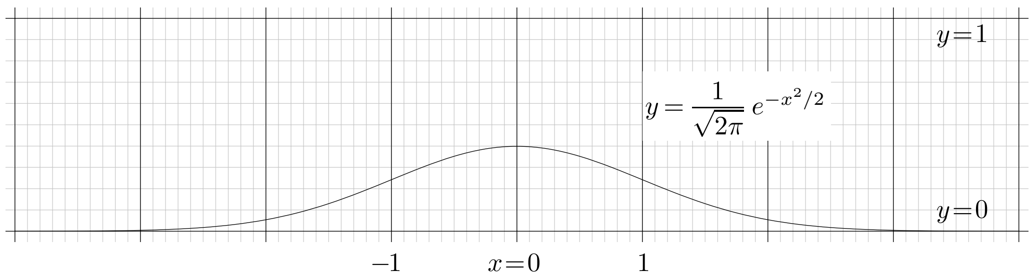
Mathematics 103 — Spring 2000

Bell curves

An important family of functions are those giving rise to the 'bell curves'. There is a whole family of them, parametrized by two numbers specifying a curve's centre and spread.

The standard curve

The **standard bell curve** is the graph



The function in the graph was first encountered when describing the way experimental errors were distributed around a true value, and for this reason it is often called the **error function**. Explicitly, we shall write for convenience

$$\mathrm{erf}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, .$$

Since $\mathrm{erf}(x)$ depends only on x^2, $\mathrm{erf}(-x) = \mathrm{erf}(x)$. So its graph is symmetric around the y-axis. For the same reason, its centre is at $x = 0$. What is not at all apparent is that the total area underneath its graph is exactly 1, and indeed that's why the factor $1/\sqrt{2\pi}$ is there. In other words, the area under the graph of $y = e^{-x^2/2}$ is equal to $\sqrt{2\pi}$, and therefore if we scale this function by $1/\sqrt{2\pi}$ we get an area of 1. That the area underneath the graph $y = e^{-x^2/2}$ is equal to $1/\sqrt{2\pi}$ is not at all an elementary fact, and we shall not attempt to explain here why it is so.

The function $\mathrm{erf}(x)$ is one of the most important functions in all of science. It plays an important role in almost any phenomenon involving a large number of objects behaving more or less randomly—including atoms emitting light, molecules diffusing in a chemical solution, and crowds of people run amok. In practice, one of the most important question you will see asked about this function is of the type: *What is the area underneath this graph between $x = a$ and $x = b$* for various numbers a and b? This area is of course just the integral

$$\frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx$$

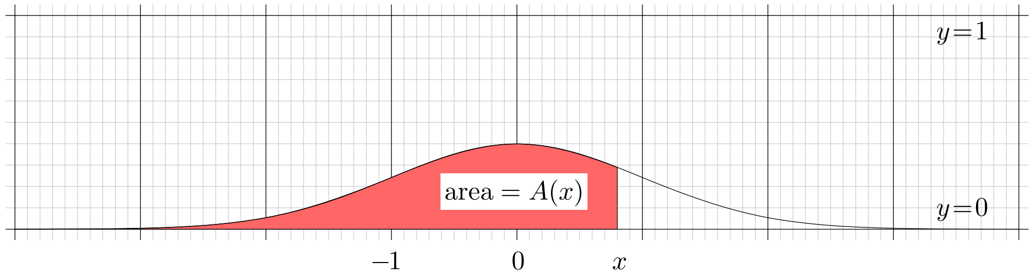
but this doesn't help too much, because

- *There is no good expression for the indefinite integral of* $\mathrm{erf}(x)$.

The only way to calculate the definite integral is to use a computer to do it directly or to use tables which have been computed already. An appendix to this Chapter contains a table of values for

$$A(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$$

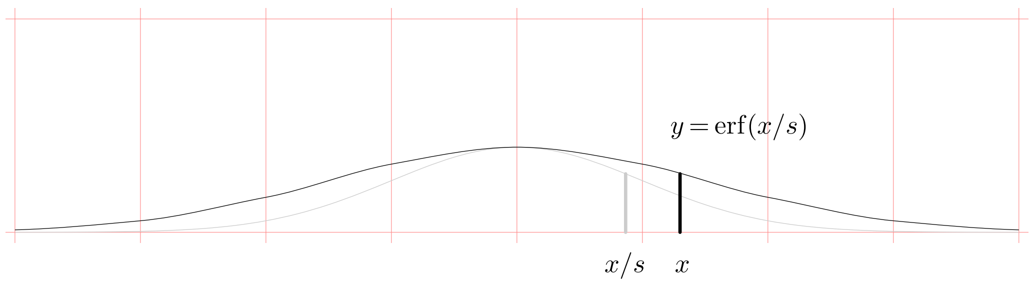
where x varies in steps of size 0.01 in the range -3 to 3.



For intermediate values of x, various kinds of interpolation can be used.

Variation

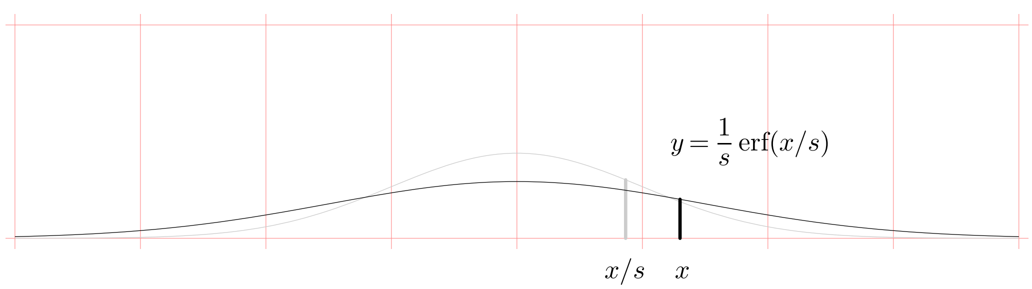
From the standard bell curve we get others by scaling it and shifting it. We start by scaling it. If we scale it horizontally by a factor s, this means stretching by horizontally by that factor. Compressing it by a factor s means scaling it horizontally by $1/s$. Now if we stretch any graph $y = f(x)$ by s, the new graph we get is $y = f(x/s)$. That is to say, the height of the new graph at x is equal to the height of the old graph at x/s. In this figure, for example, the factor is $s = 3/2$:



But if we stretch the curve by a factor of s we multiply the area by s as well. We want an area of 1, so we must now divide the height by s. The equation of our new graph is

$$y = \frac{1}{s}\,\mathrm{erf}(x/s) = \frac{1}{s\sqrt{2\pi}}\,e^{-x^2/2s^2}\ .$$

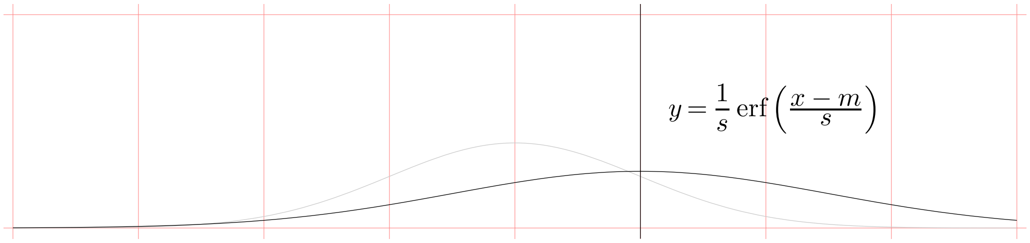
and its graph is



Now we shift it right by, say m. This changes the equation to

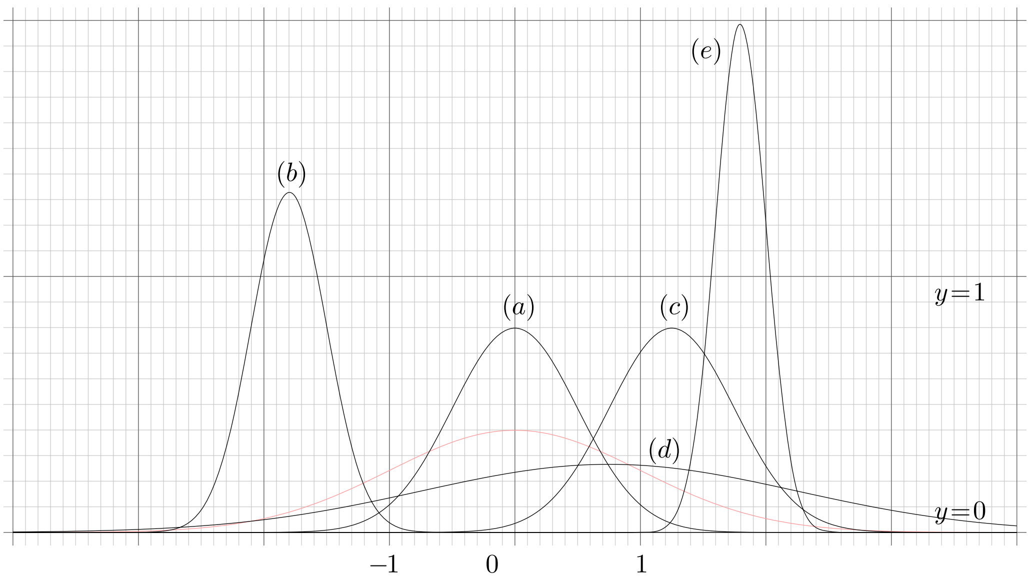
$$y = \frac{1}{s}\,\mathrm{erf}\left(\frac{x-m}{s}\right) = \frac{1}{s\sqrt{2\pi}}e^{-(x-m)^2/2s^2}$$

and its graph looks like this:



The parameter m is the **mean** value of x for this distribution, and s measures the **spread** of the graph. If s is large, then the curve is wide and flat, but if it is small the curve is thin and tall.

Exercise 1. *In the following figure is a collection of 'mystery' bell curves, alongside the standard one. Tell roughly for each one what s and m are.*



Exercise 2. *What is the area between 0 and 1 under the graph $y = \mathrm{erf}(x)$?*

Exercise 3. *What is the area between -2 and 1 under the graph $y = \mathrm{erf}(x/2)/2$?*