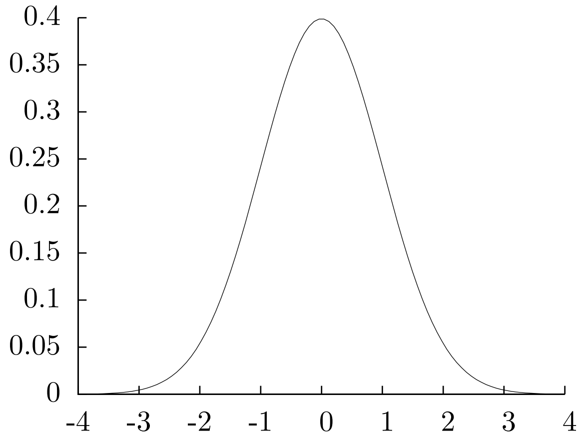

Figure 1: A normal curve with mean zero and variance one.

9 The Gaußian distribtion

The Gauß[1] or Gauss or normal or Gaußian or Gaussian or bell-curve distribution is a continuous distribution which is used to model a whole range of natural phenomenon, in fact, much of statistics and almost all statistics outside of science, assumes almost everything has a Gaußian distribution. We will see why later on, basically there is a theorem, the Central Limit Theorem, that tells us why the Gaußian distribution is as common as it is. For now though we will look at the distribution and its properties.

The Gaußian distribution with zero mean and unit variance is given by

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{1}$$

It has a classic 'bell' shape seen in Fig. 1; it looks a bit like a binomial distribution with $p = 0.5$. The slightly confusing thing is the $1/\sqrt{2\pi}$, that is there to normalize the curve:

$$\int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi} \tag{2}$$

This is confusing because we can do this definite integral, but the corresponding indefinite integral can't be done in the sense that we can't write down a formula in terms of functions we already know. There is a trick for doing the definite integral which we won't look at here for reasons of time but is very nice if you want to look it up.

First lets work out the mean and variance of this distribution, it is a maybe surprising thing that the integrals required for calculating the mean and variance are much easier than the integral mentioned above for normalizing the curve. So, mean first

$$\mu = \int_{-\infty}^{\infty} x p(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-x^2/2} dx \tag{3}$$

[1] ß is a German letter equivalent to ss

This integral is very easy because the integrand is odd: a function $f(x)$ is odd if $f(-x) = -f(x)$. Now consider the integral from minus infinity to infinity of an odd function

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{0} f(x)dx + \int_{0}^{\infty} f(x)dx \tag{4}$$

then let $x' = -x$ in the second integral. Note that $dx' = -dx$ and $x = \infty$ means $x' = -\infty$ and swithcing around the limits in an integral changes the sign:

$$\int_{b}^{a} g(x)dx = -\int_{a}^{b} g(x)dx \tag{5}$$

Finally $f(x) = f(-x') = -f(x')$ so we have

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{0} f(x)dx - \int_{0}^{\infty} f(x')dx' = 0 \tag{6}$$

Hence, noting that

$$f(x) = xe^{-x^2/2} \tag{7}$$

is odd we see $\mu = 0$

The variance is a little trickier, given the zero mean we have

$$\sigma^2 = \int_{-\infty}^{\infty} x^2 p(x)dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx \tag{8}$$

This has to be done by integrating by parts, if you don't know integration by parts, just ignore this bit, but

$$\frac{d}{dx} e^{-x^2/2} = xe^{-x^2/2} \tag{9}$$

so

$$\sigma^2 = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \frac{de^{-x^2/2}}{dx} dx \tag{10}$$

Now, applying the integration by parts formula

$$\int_{a}^{b} u \frac{dv}{dx} dx = uv|_a^b - \int_{a}^{b} v \frac{du}{dx} dx \tag{11}$$

and using the fact

$$\lim_{x\to\pm\infty} xe^{-x^2/2} = 0 \tag{12}$$

we have

$$\sigma^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = 1 \tag{13}$$

Thus the distribution

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{14}$$

is the Gaußian distribution with mean zero and variance one, it is sometimes described as $\mathcal{N}(0, 1)$; this notation is a little confusing, it is never really specified what 'described as', but roughly speaking people write $X \sim \mathcal{N}(0,1)$ as a shorthand for say X is normally distributed with mean zero and variance one, or that X has density function

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{15}$$

It is easy to work out the density function with mean μ and variance σ^2, roughly speaking this is done by translating the variable and scaling it, the only tricky bit which we will just state here is that scaling the variable changes the normalization. In short if we say a variable is normally distributed with mean μ and variance σ^2, or if we say $X \sim \mathcal{N}(\mu, \sigma^2)$ we mean it has density function

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{16}$$

Working out Gaußian probabilities

Obviously what we'd like to do is work out probabilities:

$$\text{Prob}(x \in (a,b)) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_a^b e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \tag{17}$$

There problem is that we can't do that integral, there is no way to write the integral in terms of function we already know. The solution to this problem is to define a new function, *the error function*, specifically for using to do the integral:

$$\text{erf}\,(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^{x} e^{-y^2} dy = \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2} dy \tag{18}$$

This is a so called *special function* which rougly means a function we needed to define so that we could do an integral, or solve a differential equation, we couldn't otherwise do or solve. Other examples are Bessel functions and the elliptic integrals; there are lots, especially coming from applied mathematics. Sometimes some number theory functions, like Euler's totient function, are called special functions. Either way, once a special function is defined and numerial values computed for it, then you can use it by relating other integrals to it.