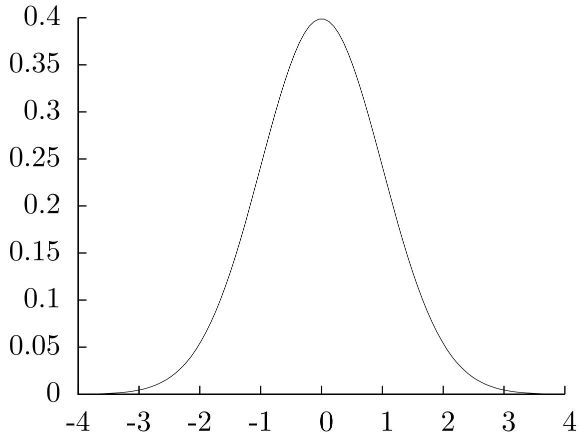

Figure 1: A normal curve with mean zero and variance one, that is $\mu = 0$ and $\sigma^2 = 1$.

9 The Gaußian distribtion

Recall how continuous probability distributions work, they are defined in terms of a density function $p(x)$ where $p(x)$ is like the probability per length, so

$$\text{Prob}(x_1 < x < x2) = \int_{x_1}^{x_2} p(x)dx \tag{1}$$

We used $f(x)$ for the probability density when we introduced it to avoid confusing it with the probabilities that are used for discrete random variables. However, like almost everything in statistics it is usually just called $p(x)$, or sometimes $p_X(x)$ if there are a few random variables around and we want to know which probability density goes with which variable.

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 is given by

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

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 $1/\sqrt{2\pi\sigma^2}$, that is there to normalize the curve:

$$\int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi\sigma^2} \tag{3}$$

[1] ß is a German letter equivalent to ss

This is confusing because we can do this particular definite integral going from minus infinity to infinity, 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.

Surprisingly it is easier to calculate the moment generating function than it is to calculate the mean and variance directly; we will do this soon; we will see that the mean is μ and the variance σ^2, just as we'd hope, that's why these particular symbols were used in the formula for the density.

The Gaußian distribution is sometimes described as $\mathcal{N}(\mu, \sigma^2)$; this notation is a little confusing, it is never really specified what 'described as' means, but roughly speaking people write $X \sim \mathcal{N}(\mu, \sigma^2)$ as a shorthand for say X is normally distributed with mean μ and variance σ^2

The moment generating function

As we noted above, obviously when the constants were named μ and σ^2 in the definition of the probability density it was because these contants correspond to the mean and variance. Here we're going to check that by working out the moment generating function for the Gaußian. This might seem a needlessly complicated way to work out the mean and the variance, but it is actually the easiest way to do it.

Recall that the moment generating function

$$m(t) = \langle e^{tX} \rangle \tag{4}$$

can be used to work out all the moments of a distribution. In the case of the Gaußian

$$m(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{xt} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \tag{5}$$

Now $e^a e^b = e^{a+b}$ so

$$m(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2} + xt} dx \tag{6}$$

and we can complete the square: first add in the new term

$$\frac{(x-\mu)^2}{2\sigma^2} - xt = \frac{1}{2\sigma^2}(x^2 - 2\mu x + \mu^2 - 2\sigma^2 xt) \tag{7}$$

Next add and take away what is needed to make a square

$$\begin{aligned} x^2 - 2\mu x + \mu^2 - 2\sigma^2 xt &= x^2 - 2(\mu + \sigma^2 t)x + (\mu + \sigma^2 t)^2 - 2\mu\sigma^2 t - \sigma^4 t^2 \\ &= (x - \mu - \sigma^2 t)^2 - 2\mu\sigma^2 t - \sigma^4 t^2 \end{aligned} \tag{8}$$

Hence

$$m(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu-\sigma^2 t)^2}{2\sigma^2} + \mu t + \frac{1}{2}\sigma^2 t^2} dx \tag{9}$$

Finally use $e^{a+b} = e^a e^b$ and move stuff with no xs outside the integral

$$m(t) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu-\sigma^2 t)^2}{2\sigma^2}} dx \right) e^{\mu t + \frac{1}{2}\sigma^2 t^2} \tag{10}$$

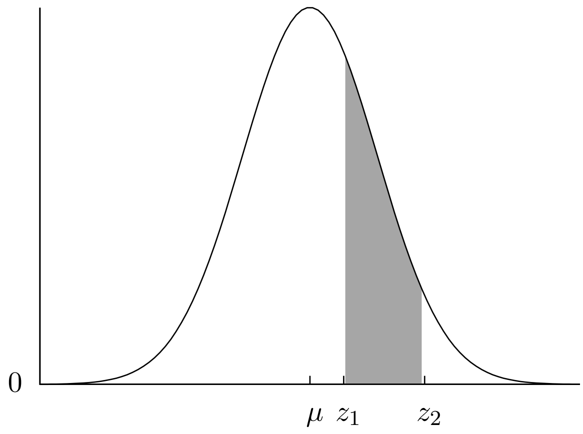

Figure 2: Working out the probability means calculating the area under the curve.

and the stuff in the big brackets is just one, it is the integral of the Gaußian with mean $\mu+\sigma^2 t$, so

$$m(t) = e^{\mu t+\frac{1}{2}\sigma^2 t^2} \tag{11}$$

Now remember that

$$\frac{d^n m}{dt^n}(0) = \mu_n \tag{12}$$

where

$$\mu_n = \langle X^n \rangle \tag{13}$$

Now, using the chain rule, which in this case tells us that

$$\frac{d}{dt}e^{f(t)} = \frac{df}{dt}e^{f(t)} \tag{14}$$

we get

$$\frac{d}{dt}e^{\mu t+\frac{1}{2}\sigma^2 t^2} = (\mu+\sigma^2 t)e^{\mu t+\frac{1}{2}\sigma^2 t^2} \tag{15}$$

If we set $t=0$ this tells us that $\langle X \rangle = \mu$. Next

$$\frac{d^2}{dt^2}e^{\mu t+\frac{1}{2}\sigma^2 t^2} = \frac{d}{dt}(\mu+\sigma^2 t)e^{\mu t+\frac{1}{2}\sigma^2 t^2} = [(\mu+\sigma^2 t)^2+\sigma^2]e^{\mu t+\frac{1}{2}\sigma^2 t^2} \tag{16}$$

and if we set $t=0$ we get $\langle X^2 \rangle = \sigma^2+\mu^2$ and hence $\langle X^2 \rangle - \langle X \rangle^2 = \sigma^2$.

Working out Gaußian probabilities

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

$$\text{Prob}(x_1 < x < x_2) = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{x_1}^{x_2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx \tag{17}$$

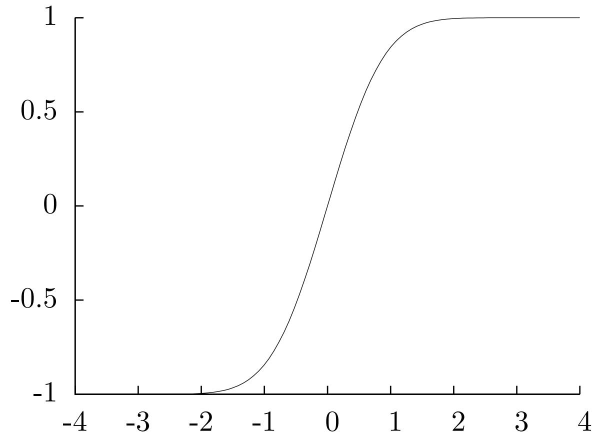

Figure 3: The error function erf(x).

as illustrated in Fig. 2. The 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. A lot of effort in the C19 was put into defining special functions and finding efficient ways to numerically calculate values; in those days, of course, these then went into big tables of values; now all of that is done for us by the C `math` library and it successors.

A graph of erf$\,(x)$ is shown in Fig. 3. We can use it to work out Gaußian probablities. Consider

$$\text{Prob}(x_1 < x < x_2) = \int_{x_!}^{x_2} p(x)dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{x_1}^{x_2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \tag{19}$$

Now let

$$z = \frac{x - \mu}{\sqrt{2}\sigma} \tag{20}$$

so

$$dz = \frac{dx}{\sqrt{2}\sigma} \tag{21}$$

and when $x = x_1$ we have

$$z = z_1 = \frac{x_1 - \mu}{\sqrt{2}\sigma} \tag{22}$$

and when $x = x_2$ we have

$$z = z_2 = \frac{x_2 - \mu}{\sqrt{2}\sigma} \tag{23}$$

Substituting this all back into the integral

$$\text{Prob}(x_1 < y < x_2) = \frac{1}{\sqrt{\pi}} \int_{z_1}^{z_2} e^{-\frac{z^2}{2}} dz = \frac{1}{\sqrt{\pi}} \int_{z_1}^{0} e^{-z^2} dz + \frac{1}{\sqrt{\pi}} \int_{0}^{z_2} e^{-z^2} dz \tag{24}$$

Hence using the usual

$$\int_a^b f(x)dx = -\int_b^a f(x)dx \tag{25}$$

we have

$$\text{Prob}(x_1 < x < x_2) = \frac{1}{2}[\text{erf}\,(z_2) - \text{erf}\,(z_1)] \tag{26}$$

Example

The loudness of songs at a concert are normally distributed with mean 75 dB and standard deviation $\sigma = 10dB$. What is the probability that a long has loudness between 80 and 90 dB? Well

$$\text{Prob}(80 < x < 90) = \frac{1}{2}[\text{erf}\,(z_2) - \text{erf}\,(z_1)] \tag{27}$$

where

$$\sqrt{2}z_1 = \frac{80 - 75}{10} = 0.5 \tag{28}$$

and

$$\sqrt{2}z_2 = \frac{90 - 75}{10} = 1.5 \tag{29}$$

Working out $\text{erf}\,(0.5/\sqrt{2})$ using your calculator or computer gives 0.38, whereas $\text{erf}\,(1.5/\sqrt{2}) = 0.87$ so

$$\text{Prob}(80 < x < 90) = 0.2417 \tag{30}$$