9 The Gaußian Distribution

COMS10011

coms10011.github.io

November 2019

Continuous random variables

Continuous probability distributions work are defined in terms of a density function $p(x)$:

- $p(x)$is like the probability per length

So to work out a probability for an interval

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

Introducing the Gaußian distribution I

- You can tell the Gaußian distribution is important because it has different names: the *Gaußian distribution*, the *normal distribution* or *the bell curve*.
- It is a continuous distribution which is used to model a whole range of natural phenomenon.
- Much of statistics assumes almost everything has a Gaußian distribution
- There is a theorem we'll look at later, the *Central Limit Theorem*, that tells us why it is so common.

Introducing the Gaußian distribution II

The Gaussian distribution

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

where we'll show later that μ is the mean and σ^2 is the variance, as you'd expect from the choice of symbols.

Introducing the Gaußian distribution III

The Gaussian distribution

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

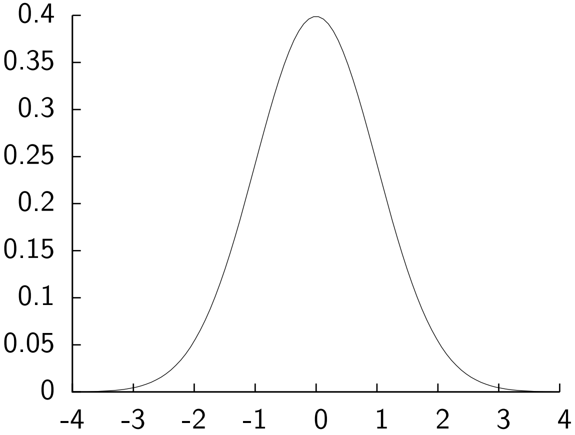
What about the $1/\sqrt{2\pi\sigma^2}$? That is there to normalize the curve:

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

- We can do this particular definite integral going from minus infinity to infinity, but the corresponding indefinite integral can't be done.
- There is a nice trick for doing the definite integral which we won't look at here for reasons of time.

Introducing the Gaußian distribution IV

The Gaussian distribution for $\mu = 0$ and $\sigma = 1$



The moment generating function I

Amazingly it is easier to work out the moment generating function than to work out the mean and variance directly. Recall

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

and this allows you to work out the moments because

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

where

$$\mu_n = \langle X^n \rangle$$

The moment generating function II

So for the Gaußian

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

Now using

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

this gives

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

The moment generating function III

Now we have

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

and write

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

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

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

The moment generating function IV

Hence

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

move the stuff with no xs outside the integral

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

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

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

The mean and variance at last I

Remember that

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

Now, rememember the chain rule

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

Using it we we get

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

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

The mean and variance at last II

Going back to

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

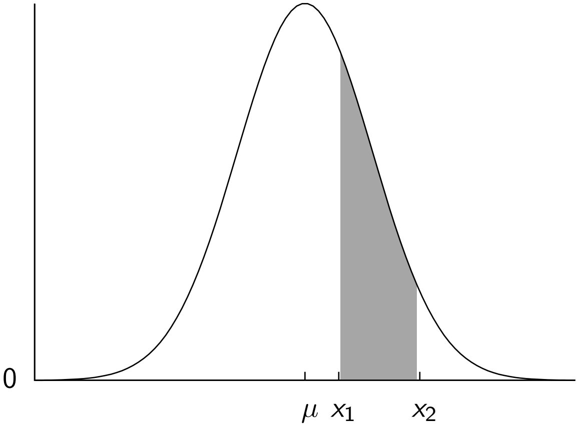
we differenciate again to get

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

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 I

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



Can't do the integral

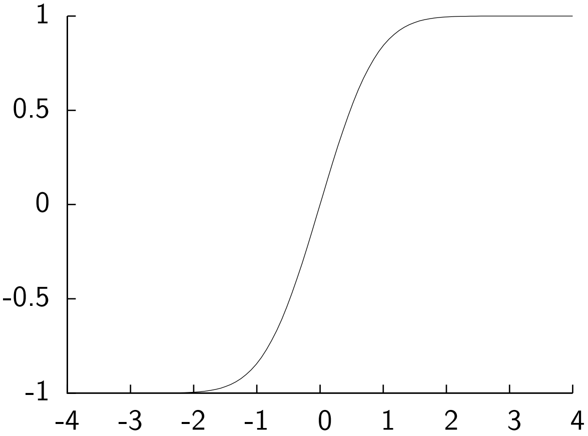
$$\text{Prob}(x_1 < x < x_2) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{x_1}^{x_2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

The problem is we can't do this integral:

- We can't write this integral in terms of functions we know.
- To get around this we define a new function specifically so we can relate the integral to that.
- This is known as a *special function*, a new function defined for some special purpose.
- The values of the special function can be calculated numerically.

The error function

$$\text{erf}\,(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^{x} e^{-y^2} dy = \frac{2}{\sqrt{\pi}} \int_{0}^{x} e^{-y^2} dy$$



Working out Gaußian probabilities II

So

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

Now do a change of variables

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

so

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

and when $x = x_1$ we have

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

and for $x = x_2$ similar.

Working out Gaußian probabilities III

Putting this into the integral we have

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

then using

$$\int_a^b f(x)dx = -\int_b^a f(x)dx$$

and

$$\int_a^b f(x)dx = \int_a^0 f(x)dx + \int_0^b f(x)dx$$

we have

$$\text{Prob}(x_1 < y < x_2) = -\frac{1}{\sqrt{\pi}} \int_0^{z_1} e^{-z^2} dz + \frac{1}{\sqrt{\pi}} \int_0^{z_2} e^{-z^2} dz$$

Working out Gaußian probabilities IV

Hence

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

An 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 the next song has loudness between 80 and 90 dB?

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

where

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

and

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

Working out numbers on a calculator or computer gives

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