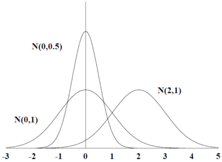
MATH7501: Mathematics for Data Science I

Final Lecture: The Normal Distribution

The Normal Distribution

Definition (Gaussian Distribution)

We say that X has a *Gaussian/normal distribution with parameters μ and σ^2* on $\mathbb{R}$ (denoted $X \sim \mathrm{N}(\mu, \sigma^2)$) if its pdf is given by $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$.



Properties of the Gaussian Distribution

If $X \sim \mathrm{N}(\mu, \sigma^2)$ then

- $\mathbb{E}[X] = \mu$
- $\mathrm{Var}[X] = \sigma^2$.

The Normal Distribution

Important Observation

The function $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ is actually a pdf, i.e., it is continuous, positive, and integrates to 1.

Proof (Part 1 of 3 – Changing Variables)

Let $u = \frac{x-\mu}{\sigma}$, so that $\frac{du}{dx} = \frac{1}{\sigma}$. Then

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(u)^2} du,$$

but $g(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(u)^2}$ is precisely the pdf of the normal distribution with $\mu = 0$ and $\sigma = 1$. It therefore suffices to prove the result in this case.

The Normal Distribution

Important Observation

The function $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ is actually a pdf, i.e., it is continuous, positive, and integrates to 1.

Proof (Part 2 of 3 – Introducing a Volume Integral)

Recall we are trying to show that $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx = 1$. This is equivalent to showing that $\left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\, dx\right)^2 = 2\pi$. But

$$\begin{aligned}
\left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\, dx\right)^2 &= \left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\, dx\right)\left(\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}\, dy\right) \\
&= \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \left(\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}\, dy\right) dx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}}\, dy dx.
\end{aligned}$$

The Normal Distribution

Important Observation

The function $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ is actually a pdf, i.e., it is continuous, positive, and integrates to 1.

Proof (Part 3 of 3 – Evaluating Volume Integral via Polar)

Using polar co-ordinates, we have

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dydx = \int_{0}^{2\pi} \int_{0}^{\infty} e^{-\frac{r^2}{2}} \mathbf{r} dr d\theta.$$

Now using integration by substitution, we have $\int_0^\infty re^{-\frac{r^2}{2}} dr = 1$. Therefore $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dydx = \int_0^{2\pi} \int_0^\infty e^{-\frac{r^2}{2}} \mathbf{r} dr d\theta = 2\pi$, as required.

The Normal Distribution

Question

Suppose that $X \sim \mathrm{N}(\mu, \sigma^2)$. Show that $\mathbb{E}[X] = \mu$.

Solution

Since the pdf is $\frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{x-\mu}{\sigma}\right)^2}$, we have

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\infty} \frac{x}{\sigma\sqrt{2\pi}} e^{-\left(\frac{x-\mu}{\sigma}\right)^2} dx \\
&= \int_{-\infty}^{\infty} \frac{y+\mu}{\sigma\sqrt{2\pi}} e^{-\left(\frac{y}{\sigma}\right)^2} dy \\
&= \int_{-\infty}^{\infty} \frac{\mu}{\sigma\sqrt{2\pi}} e^{-\left(\frac{y}{\sigma}\right)^2} dy + \int_{-\infty}^{\infty} \frac{y}{\sigma\sqrt{2\pi}} e^{-\left(\frac{y}{\sigma}\right)^2} dy \\
&= \mu + \int_{-\infty}^{0} \frac{y}{\sigma\sqrt{2\pi}} e^{-\left(\frac{y}{\sigma}\right)^2} dy + \int_{0}^{\infty} \frac{y}{\sigma\sqrt{2\pi}} e^{-\left(\frac{y}{\sigma}\right)^2} dy \\
&= \mu
\end{aligned}$$

The Normal Distribution

Question

Suppose that $X \sim \mathrm{N}(\mu, \sigma^2)$. Show that $Var[X] = \sigma^2$.

Solution (Sketch)

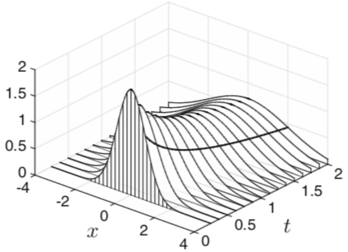
We have to show that $\int_{-\infty}^{\infty} \frac{(x-\mu)^2}{\sigma\sqrt{2\pi}} e^{-\left(\frac{x-\mu}{\sigma}\right)^2} dx = \sigma^2$. After substituting $u = \frac{x-\mu}{2\sigma}$, we get

$$\int_{-\infty}^{\infty} \frac{(x-\mu)^2}{\sigma\sqrt{2\pi}} e^{-\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{\sigma^2}{2\sqrt{\pi}} \int_{-\infty}^{\infty} u^2 e^{-u^2} du.$$

Using integration by parts, we obtain the answer.

The Normal Distribution

The Gaussian distribution comes up in physics. Indeed, if we think of time t as the same as σ^2 and set $\mu = 0$, then we get a time evolving function that describes how heat spreads out.



https://math.stackexchange.com/questions/2871506/convexity-and-concavity-in-a-textbook-figure-of-the-fundamental-solution-of-the

This function $f(t,x)$ satisfies the so-called *heat equation*:

$$\frac{\partial f}{\partial t} = \frac{\partial^2 f}{\partial x^2}.$$

The Normal Distribution

Example (Brownian Motion)

Let P be a particle travelling according to Brownian motion in $\mathbb{R}$, with "speed" equal to 1. Suppose the particle starts at position $\mu \in \mathbb{R}$.

Then the random variable X_t which describes the location of the particle P after time t is distributed as $X_t \sim \mathrm{N}(\mu, t)$.

Conclusion

Thankyou! I hope this course has prepared you for your future in data science!