Normal Distribution

The single most important random variable type is the Normal (aka Gaussian) random variable, parametrized by a mean (μ) and variance (σ^2), or sometimes equivalently written as mean and variance (σ^2). If X is a normal variable we write $X \sim N(\mu, \sigma^2)$. The normal is important for many reasons: it is generated from the summation of independent random variables and as a result it occurs often in nature. Many things in the world are not distributed normally but data scientists and computer scientists model them as Normal distributions anyways. Why? Because it is the most entropic (conservative) modelling decision that we can make for a random variable while still matching a particular expectation (average value) and variance (spread).

The Probability Density Function (PDF) for a Normal $X \sim N(\mu, \sigma^2)$ is:

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

Notice the x in the exponent of the PDF function. When x is equal to the mean (μ) then e is raised to the power of 0 and the PDF is maximized.

By definition a Normal has $\mathrm{E}[X] = \mu$ and $\mathrm{Var}(X) = \sigma^2$.

There is no closed form for the integral of the Normal PDF, and as such there is no closed form CDF. However we can use a transformation of any normal to a normal with a precomputed CDF. The result of this mathematical gymnastics is that the CDF for a Normal $X \sim N(\mu, \sigma^2)$ is:

$$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

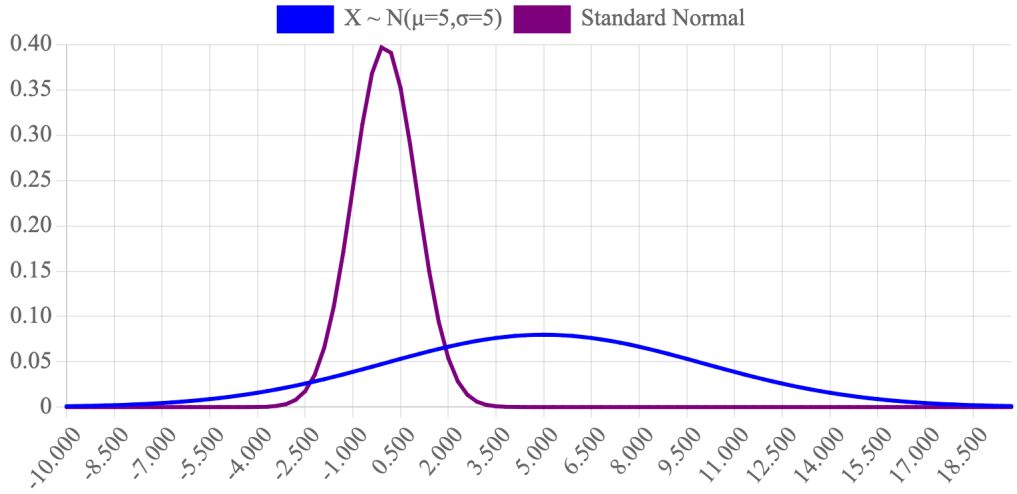
Where Φ is a precomputed function that represents that CDF of the Standard Normal.

Normal (aka Gaussian) Random Variable

Notation:	$X \sim \mathrm{N}(\mu, \sigma^2)$
Description:	A common, naturally occurring distribution.
Parameters:	$\mu \in \mathbb{R}$, the mean. $\sigma^2 \in \mathbb{R}$, the variance.
Support:	$x \in \mathbb{R}$
PDF equation:	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
CDF equation:	$F(x) = \phi(\frac{x-\mu}{\sigma})$ Where ϕ is the CDF of the standard normal
Expectation:	$\mathrm{E}[X] = \mu$
Variance:	$\mathrm{Var}(X) = \sigma^2$

PDF graph:

Parameter μ: 5 Parameter σ: 5



Linear Transform

If X is a Normal such that $X \sim N(\mu, \sigma^2)$ and Y is a linear transform of X such that $Y = aX + b$ then Y is also a Normal where:

$$Y \sim N(a\mu + b, a^2\sigma^2)$$

Projection to Standard Normal

For any Normal X we can find a linear transform from X to the standard normal $Z \sim N(0, 1)$. Note that Z is the typical notation choice for the standard normal. For any normal, if you subtract the mean (μ) of the normal and divide by the standard deviation (σ) the result is always the standard normal. We can prove this mathematically. Let $W = \frac{X-\mu}{\sigma}$:

$$
\begin{aligned}
W &= \frac{X-\mu}{\sigma} && \text{Transform X: Subtract by } \mu \text{ and diving by } \sigma \\
&= \frac{1}{\sigma}X - \frac{\mu}{\sigma} && \text{Use algebra to rewrite the equation} \\
&= aX + b && \text{Linear transform where } a = \frac{1}{\mu}, b = -\frac{\mu}{\sigma} \\
&\sim N(a\mu + b, a^2\sigma^2) && \text{The linear transform of a Normal is another Normal} \\
&\sim N(\frac{\mu}{\sigma} - \frac{\mu}{\sigma}, \frac{\sigma^2}{\sigma^2}) && \text{Substituting values in for } a \text{ and } b \\
&\sim N(0, 1) && \text{The standard normal}
\end{aligned}
$$

Using this transform we can express $F_X(x)$, the CDF of X, in terms of the known CDF of Z, $F_Z(x)$. Since the CDF of Z is so common it gets its own Greek symbol: $\Phi(x)$

$$
\begin{aligned}
F_X(x) &= P(X \leq x) \\
&= P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) \\
&= P\left(Z \leq \frac{x-\mu}{\sigma}\right) \\
&= \Phi\left(\frac{x-\mu}{\sigma}\right)
\end{aligned}
$$

The values of $\Phi(x)$ can be looked up in a table. Every modern programming language also has the ability to calculate the CDF of a normal random variable!

Example: Let $X \sim \mathcal{N}(3, 16)$, what is $P(X > 0)$?

$$
\begin{aligned}
P(X > 0) &= P\left(\frac{X-3}{4} > \frac{0-3}{4}\right) = P\left(Z > -\frac{3}{4}\right) = 1 - P\left(Z \leq -\frac{3}{4}\right) \\
&= 1 - \Phi(-\frac{3}{4}) = 1 - (1 - \Phi(\frac{3}{4})) = \Phi(\frac{3}{4}) = 0.7734
\end{aligned}
$$

What is $P(2 < X < 5)$?

$$
\begin{aligned}
P(2 < X < 5) &= P\left(\frac{2-3}{4} < \frac{X-3}{4} < \frac{5-3}{4}\right) = P\left(-\frac{1}{4} < Z < \frac{2}{4}\right) \\
&= \Phi(\frac{2}{4}) - \Phi(-\frac{1}{4}) = \Phi(\frac{1}{2}) - (1 - \Phi(\frac{1}{4})) = 0.2902
\end{aligned}
$$

Example: You send voltage of 2 or -2 on a wire to denote 1 or 0. Let X = voltage sent and let R = voltage received. $R = X + Y$, where $Y \sim \mathcal{N}(0, 1)$ is noise. When decoding, if $R \geq 0.5$ we interpret the voltage as 1, else 0. What is $P(\text{error after decoding} | \text{original bit} = 1)$?

$$
\begin{aligned}
P(X + Y < 0.5) &= P(2 + Y < 0.5) \\
&= P(Y < -1.5) \\
&= \Phi(-1.5) \\
&\approx 0.0668
\end{aligned}
$$

Example: The 67% rule of a normal within one standard deviation. What is the probability that a normal variable $X \sim \mathrm{N}(\mu, \sigma)$ has a value within one standard deviation of its mean?

$$
\begin{aligned}
\mathrm{P}(\text{Within one } \sigma \text{ of } \mu) &= \mathrm{P}(\mu - \sigma < X < \mu + \sigma) \\
&= \mathrm{P}(X < \mu + \sigma) - \mathrm{P}(X < \mu - \sigma) && \text{Prob of a range} \\
&= \Phi\Big(\frac{(\mu + \sigma) - \mu}{\sigma}\Big) - \Phi\Big(\frac{(\mu - \sigma) - \mu}{\sigma}\Big) && \text{CDF of Normal} \\
&= \Phi\Big(\frac{\sigma}{\sigma}\Big) - \Phi\Big(\frac{-\sigma}{\sigma}\Big) && \text{Cancel } \mu\text{s} \\
&= \Phi(1) - \Phi(-1) && \text{Cancel } \sigma\text{s} \\
&\approx 0.8413 - 0.1587 \approx 0.683 && \text{Plug into } \Phi
\end{aligned}
$$

We made no assumption about the value of μ or the value of σ so this will apply to every single normal random variable. Since it uses the Normal CDF this doesn't apply to other types of random variables.

CDF Calculator

To calculate the Cumulative Density Function (CDF) for a normal (aka Gaussian) random variable at a value x, also writen as $F(x)$, you can transform your distribution to the "standard normal" and look up the corresponding value in the standard normal CDF. However, most programming libraries will provide a normal cdf funciton:

Norm CDF Calculator

x	0.0
mu	0
std	1

`norm.cdf(x, mu, std)`

In python you can calculate these values using the `scipy` library

```
from scipy import stats

# get the input values
mean = 1.0
std_dev = 0.5
query = 0.1 # aka x

# calc the CDF in two lines
X = stats.norm(mean, std_dev)
p = X.cdf(query)

# calc the CDF in one line
p = stats.norm.cdf(query, mean, std_dev)
```

It is important to note that in the python library, the second parameter for the Normal distribution is standard deviation **not** variance, as it is typically defined in math notation. Recall that standard deviation is the square root of variance.