

The Normal Distribution

Based on a chapter by Chris Piech

The single most important random variable type is the **normal (a.k.a. Gaussian) random variable**, parametrized by a mean (μ) 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 quite distributed normally, but data scientists and computer scientists model them as normal distributions anyways. Why? Essentially, the normal is what we use if we know mean and variance, but nothing else. In fact, it is the most conservative¹ modeling 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:

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

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 design, a normal has $E[X] = \mu$ and $\text{Var}(X) = \sigma^2$.

Linear Transform

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

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

Projection to Standard Normal

For any normal RV X we can find a linear transform from X to the **standard normal** $N(0, 1)$. That is, if you subtract the mean (μ) of the normal and divide by the standard deviation (σ), the result is distributed according to the standard normal. We can prove this mathematically. Let $W = \frac{X-\mu}{\sigma}$:

$W = \frac{X - \mu}{\sigma}$	transform X : subtract μ and divide by σ
$= \frac{1}{\sigma}X - \frac{\mu}{\sigma}$	use algebra to rewrite the equation
$= aX + b$	where $a = \frac{1}{\sigma}$, $b = -\frac{\mu}{\sigma}$
$\sim N(a\mu + b, a^2\sigma^2)$	the linear transform of a normal is another normal
$\sim N\left(\frac{\mu}{\sigma} - \frac{\mu}{\sigma}, \frac{\sigma^2}{\sigma^2}\right)$	substituting values in for a and b
$\sim N(0, 1)$	the standard normal

¹Formally, it has the *highest entropy* $H(X) = -\int dx f(x) \log f(x)$ of any distribution given the mean and variance.

An extremely common use of this transform is to express $F_X(x)$, the CDF of X , in terms of the CDF of Z , $F_Z(x)$. Since the CDF of Z is so common it gets its own Greek symbol: $\Phi(x)$

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

Why is this useful? Well, in the days when we couldn't call `scipy.stats.norm.cdf` (or on exams, when one doesn't have a calculator), people would look up values of the CDF in a table (see the last page of these notes). Using the standard normal means you only need to build a table of one distribution, rather than an indefinite number of tables for all the different values of μ and σ !

We also have an online calculator on the CS 109 website. You should learn how to use the normal table for the exams, however!

How to Remember that Crazy PDF

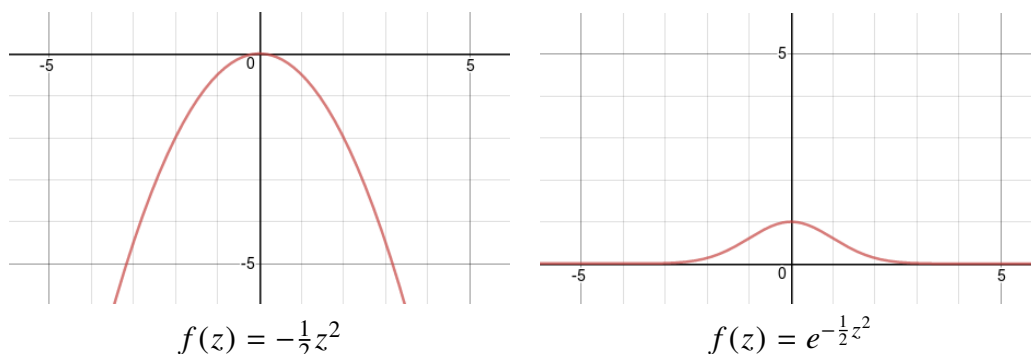
What is the PDF of the standard normal Z ? Let's plug it in:

$$f_Z(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2} = \frac{1}{1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-0}{1}\right)^2} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

This gets even better if we realize that $\frac{1}{\sqrt{2\pi}}$ is just a constant to make the whole thing integrate to 1. Call that constant C :

$$f_Z(z) = C e^{-\frac{1}{2}z^2}$$

Not so scary anymore, is it? In fact, this equation can be a rather helpful mnemonic: the normal distribution PDF is just the exponential of a parabola. What does that look like?



As it turns out, the exponential of a (downward) parabola is a familiar shape: the “bell curve”. Now bring back the fact that $Z = \frac{X-\mu}{\sigma}$, and you can see that μ determines where the “peak” of the bell curve will be, while σ tells you how wide it is. (Don't forget that C changes too!)

Example 1

Let $X \sim N(3, 16)$, what is $P(X > 0)$?

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

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

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

Example 2

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 $Y \sim N(0, 1)$ is noise. When decoding, if $R \geq 0.5$ we interpret the voltage as 1, else 0. What is $P(\text{error after decoding} | \text{original bit} = 1)$?

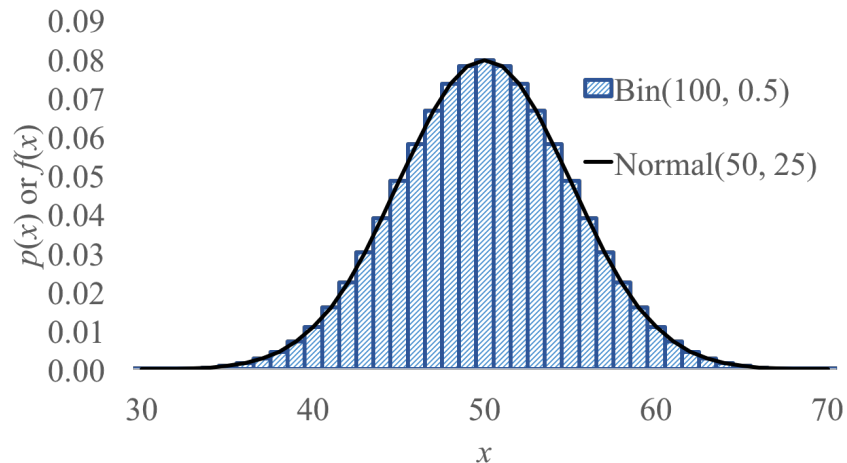
$$P(X + Y < 0.5) = P(2 + Y < 0.5) = P(Y < -1.5) = \Phi(-1.5) = 1 - \Phi(1.5) \approx 0.0668$$

Binomial Approximation

Imagine this terrible scenario. You need to solve a probability question on a binomial random variable (see the chapter on discrete distributions) with a large value for n (the number of experiments). You quickly realize that it is way too hard to compute by hand. Recall that the binomial probability mass function has an $n!$ term. You decide to turn to your computer, but after a few iterations you realize that this is too hard even for your GPU boosted mega computer (or your laptop).

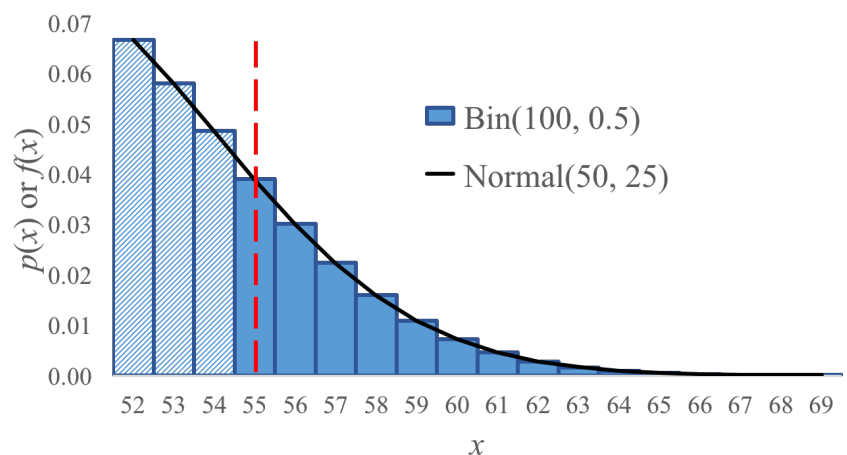
As a concrete example, imagine that in an election each person in a country with 10 million people votes in an election. Each person in the country votes for candidate A, with independent probability 0.53. You want to know the probability that candidate A gets more than 5 million votes. Yikes!

Don't panic (unless you are candidate B, then sorry, this election is not for you). Did you notice how similar a normal distribution's PDF and a binomial distribution's PMF look? Let's take a side by side view:



Lets say our binomial is a random variable $X \sim \text{Bin}(100, 0.5)$ and we want to calculate $P(X \geq 55)$. We could cheat by using the closest fit normal (in this case $Y \sim N(50, 25)$). How did we chose that particular Normal? I simply selected one with a mean and variance that matches the Binomial expectation and variance. The binomial expectation is $np = 100 \cdot 0.5 = 50$. The Binomial variance is $np(1 - p) = 100 \cdot 0.5 \cdot 0.5 = 25$.

Since $Y \approx X$ then $P(X \geq 55)$ seems like it should be $\approx P(Y \geq 55)$. That is almost true. It turns out that there is a formal mathematical reason why the normal is a good approximation of the binomial as long as the Binomial parameter p is reasonable (eg in the range $[0.3 \text{ to } 0.7]$) and n is large enough. However! There was an oversight in our logic. Let’s look a bit closer at the binomial we are approximating.



Since we want to approximate $P(X \geq 55)$, our goal is to calculate the sum of all of the columns in the Binomial PMF from 55 and up (all those dark columns). If we calculate the probability that the approximating Normal random variable takes on a value greater than 55 $P(Y \geq 55)$ we will get the integral starting at the vertical dashed line. Hey! That’s not where the columns start. Really

we want the area under the curve starting half way between 54 and 55. The correct approximation would be to calculate $P(X \geq 54.5)$.

Yep, that adds an annoying layer of complexity. The simple idea is that when you approximate a discrete distribution with a continuous one, if you are not careful your approximating integral will only include half of one of your boundary values. In this case we were only adding half of the column for $P(X = 55)$). The correction is called the continuity correction.

You can use a Normal distribution to approximate a Binomial $X \sim \text{Bin}(n, p)$. To do so define a normal $Y \sim (E[X], \text{Var}(X))$. Using the Binomial formulas for expectation and variance, $Y \sim (np, np(1-p))$. This approximation holds for large n and moderate p . Since a Normal is continuous and Binomial is discrete we have to use a continuity correction to discretize the Normal.

$$P(X = k) \sim P\left(k - \frac{1}{2} < Y < k + \frac{1}{2}\right) = \Phi\left(\frac{k - np + 0.5}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - np - 0.5}{\sqrt{np(1-p)}}\right)$$

You should get comfortable deciding what continuity correction to use. Here are a few examples of discrete probability questions and the continuity correction:

Discrete (Binomial) probability question	Equivalent continuous probability question
$P(X = 6)$	$P(5.5 < X < 6.5)$
$P(X \geq 6)$	$P(X > 5.5)$
$P(X > 6)$	$P(X > 6.5)$
$P(X < 6)$	$P(X < 5.5)$
$P(X \leq 6)$	$P(X < 6.5)$

Example 3

100 visitors to your website are given a new design. Let $X = \#$ of people who were given the new design and spend more time on your website. Your CEO will endorse the new design if $X \geq 65$. What is $P(\text{CEO endorses change} | \text{it has no effect})$?

$E[X] = np = 50$. $\text{Var}(X) = np(1-p) = 25$. $\sigma = \sqrt{\text{Var}(X)} = 5$. We can thus use a Normal approximation: $Y \sim \mathcal{N}(50, 25)$.

$$P(X \geq 65) \approx P(Y > 64.5) = P\left(\frac{Y - 50}{5} > \frac{64.5 - 50}{5}\right) = 1 - \Phi(2.9) = 0.0019$$

Example 4

Stanford accepts 2480 students and each student has a 68% chance of attending. Let $X = \#$ students who will attend. $X \sim \text{Bin}(2480, 0.68)$. What is $P(X > 1745)$?

$E[X] = np = 1686.4$. $\text{Var}(X) = np(1-p) = 539.7$. $\sigma = \sqrt{\text{Var}(X)} = 23.23$. We can thus use a Normal approximation: $Y \sim \mathcal{N}(1686.4, 539.7)$.

$$P(X > 1745) \approx P(Y > 1745.5) = P\left(\frac{Y - 1686.4}{23.23} > \frac{1745.5 - 1686.4}{23.23}\right) = 1 - \Phi(2.54) = 0.0055$$