Week 3

27 May 2021

11 The binomial distribution

We're now going to look at a generalization of flipping a coin (aka, doing an action multiple times where the actions are independent of one another). We saw this partially when we talked about the Bernoulli p distribution.

Suppose that I have a fair coin and I flip it multiple times. I don't necessarily care about the order that the flips come in, but I do care about the end result. For example, if I flip the coin four times and I get (H, T, T, H), I just care that I got two heads and two tails. Let's see how many possibilities we have when keep flipping the coin.

The first time I flip it I can either get a heads or a tails:

H T

The second time I flip, I again can either get a heads or tails. But since we don't care about order (H, T) and (T, H) are the same. So we have

(H, H) (T, H)
(H, T) (T, T)

HH 2 HT TT

Let's flip a third time. We end up getting:

HHH 3 HHT 3 HTT TTT

Notice that this just gives us the binomial distribution which we talked about in the counting section in the first week. We can draw these numbers using *Pascal's triangle*:

Wikipedia: Pascal's Triangle

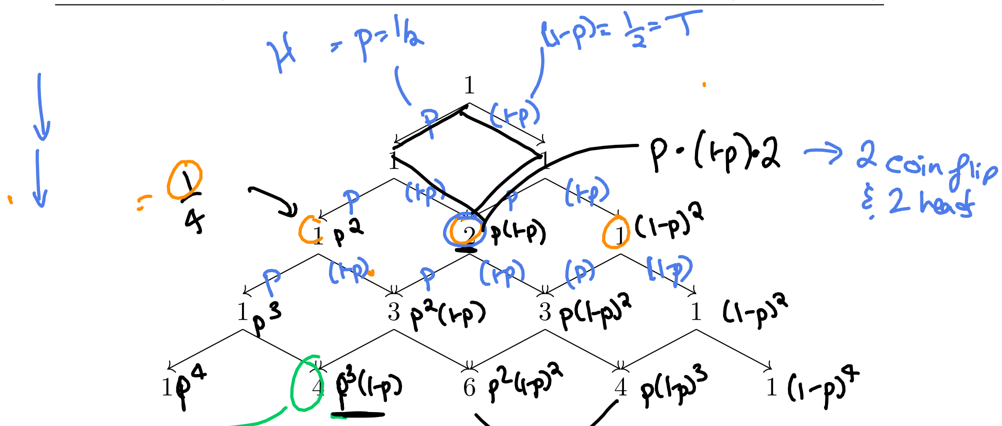


Figure 3.1: Pascal's Triangle

n 4 coinflip
k 3 heads
6+4=10

We can now use this to figure out probabilities given an arbitrary (not necessarily fair) coin. If the coin has probability p of landing on heads, then we can label each edge with p for the probability we flip a heads and a $q = 1 - p$ to represent a tails flip. So for example, if I flip 3 times, the probability that I will get 2 heads and 1 tails is $3p^2q$.

The name "binomial" comes from the expansion of the binomial $(x + y)^n$. For example:

$$(x+y)^4 = x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4$$

$p = x$
$1-p = y$

We can keep extending Pascal's triangle infinitely in order to calculate any n that we want. And so we might wonder, how does this help with flipping coins? This is where it helps to recall that the numbers in Pascal's triangle are given by the choice function. In particular if we go to the nth row and look at the kth number, that number is equal to $\binom{n}{k} = \frac{n!}{(n-k)!k!}$. In other words, if I flip my coin n times and k of them turn out to be heads then we know the probability is equal to:

$$\binom{n}{k} p^k (1-p)^{n-k}$$

This is known as the binomial distribution.

To be more precise, if we have n independent trials with probability of p that a particular trial is successful, then the *binomial probability formula* is given by:

$$p(k \text{ successes in } n \text{ trials}) = \binom{n}{k} p^k (1-p)^{n-k}$$

"k heads in n coin flips"

For a fixed n and p, the binomial probabilities for all k give a probability distribution over the numbers $\{0, 1, \ldots, n\}$ called the *binomial distribution*.

Wikipedia: Binomial Distribution
Note that the book uses the term "Binomial (n, p) distribution".

Notice how we have 0 in our distribution! This is why we let $0! = 1$; or else we couldn't work with factorial properly.

Example 11.1 Let's look at an example of something other than flipping a coin. Say we have a six sided dice and we want to figure out what the probability of getting 3 sixes out of 7 rolls. First, we know that each roll is independent and so we know we can use the binomial probability function. We know that $p = \frac{1}{6}$ and since we are doing 7 rolls we have $n = 7$. As we want 3 sixes we also know that $k = 3$. Therefore we have:

$$\binom{n}{k}(p)^k(1-p)^{n-k} = \binom{7}{3}\left(\frac{1}{6}\right)^3\left(1-\frac{1}{6}\right)^{7-3} = \left(\frac{7!}{(7-3)!(3)!}\right)\frac{1}{6^3}\frac{5^4}{6^4}$$

$$= \frac{7\cdot\cancel{6}\cdot 5}{\cancel{3\cdot 2}}\cdot\frac{5^4}{6^7} = \frac{7\cdot 5^5}{6^7} = 0.07814 \approx 7.8\%$$

Example 11.2 Now we get a little more complicated. Say we're in a coin flipping tournament and there are five contestants. Each contestant must flip six coins and are allowed into the next round of the competition if they get more than three heads. What are the chances that at least three contestants proceed to the next round?

① probability of getting >3 heads w/ 6 coin flips

$p = \frac{1}{2}$ $n = 6$ $k \in \{4, 5, 6\}$

$$P(k) = \binom{n}{k}p^k(1-p)^{n-k} = \binom{6}{k}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{6-k} = \frac{6!}{(6-k)!\,k!}\cdot\frac{1}{2^6}$$

prob of getting exactly k heads

$$P(4) = \frac{6!}{(6-4)!\,4!}\cdot\frac{1}{2^6} = \frac{\overset{3}{\cancel{6}}\cdot 5}{\cancel{2}\cdot 1\cdot 2^6} = \frac{15}{2^6}$$

$$P(6) = \frac{\cancel{6!}}{\underbrace{(6-6)!}_{0!=1}\cancel{6!}}\cdot\frac{1}{2^6} = \frac{1}{2^6}$$

$$P(5) = \frac{6!}{(6-5)!\,5!}\cdot\frac{1}{2^6} = \frac{6}{2^6}$$

$$P(4)+P(5)+P(6) = \frac{15}{2^6}+\frac{6}{2^6}+\frac{1}{2^6} = \frac{22}{2^6} = \frac{11}{2^5} = \frac{11}{32}$$

2
4
8
16
32

probability of proceeding to the next round.

② 3 out of 5 where $p = \frac{11}{32}$

$k = 3$, $n = 5$

$$\binom{n}{k}p^k(1-p)^{n-k} = \binom{5}{3}\left(\frac{11}{32}\right)^3\left(1-\frac{11}{32}\right)^{5-3} = \frac{5!}{(5-3)!\,3!}\cdot\frac{11^3}{32^3}\cdot\left(\frac{32}{32}-\frac{11}{32}\right)^2$$

$$P(3) = \frac{5\cdot\overset{2}{\cancel{4}}\cdot\cancel{3\cdot 2\cdot 1}}{\cancel{2}!\,\cancel{(3\cdot 2\cdot 1)}}\cdot\frac{11^3}{32^3}\cdot\frac{21^2}{32^2} = \frac{5\cdot 2\cdot 11^3\cdot 21^2}{32^5}$$

$$P(5) = \binom{5}{5}\left(\frac{11}{32}\right)^5\left(1-\frac{11}{32}\right)^{5-5} = \frac{5!}{0!\,5!}\cdot\frac{11^5}{32^5}\cdot 1 = \frac{11^5}{32^5}$$

$$P(4) = \binom{5}{4}\left(\frac{11}{32}\right)^4\left(1-\frac{11}{32}\right)^{5-4} = \frac{5!}{1!\,4!}\cdot\frac{11^4}{32^4}\cdot\frac{21}{32} = \frac{5\cdot 11^4\cdot 21}{32^5}$$

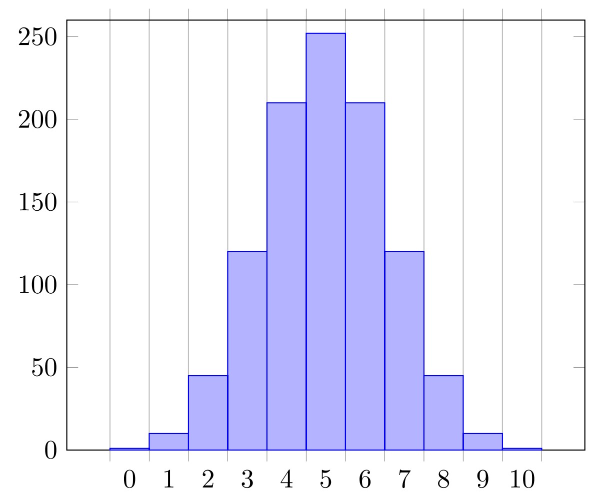
$$p(3)+p(4)+p(5) = \frac{5\cdot 2\cdot 11^3\cdot 21^2 + 5\cdot 11^4\cdot 21 + 11^5}{32^5} = 0.226 \to 22.6\%$$

12 Consecutive Odds Ratios

It turns out that for the binomial distribution, odds are actually a nicer way to look at things. Remember that odds compare two things and say what the ratio between two things occurring is. (So "Horse A has 2 : 3 odds in favour of winning against horse B" means out of ever 5 races, horse A will win 2 and horse B will win 3 on average) We're going to look at consecutive odds in the binomial distribution and see what they give.

$|A^c|$ $|A|$

Example 12.1 Let's look at the distribution when $p = \frac{1}{2}$ and $n = 10$. We end up with the following:



The distribution is given by:

$$\binom{10}{5} = \frac{10!}{5!5!}$$

1 10 45 120 210 252 210 120 45 10 1

Let's compare them consecutively:

$$\frac{10}{1} \quad \frac{45}{10} \quad \frac{120}{45} \quad \frac{210}{120} \quad \frac{252}{210} \quad \frac{210}{252} \quad \frac{45}{120} \quad \frac{10}{45} \quad \frac{1}{10}$$

$$\frac{10}{1} \quad \frac{9}{2} \quad \frac{8}{3} \quad \frac{7}{4} \quad \frac{6}{5} \quad \frac{5}{6} \quad \frac{4}{7} \quad \frac{3}{8} \quad \frac{2}{9} \quad \frac{1}{10}$$

It's easy to see that this generalizes generally:

$$\frac{\binom{n}{k}}{\binom{n}{k-1}} = \frac{\frac{n!}{k!(n-k)!}}{\frac{n!}{(k-1)!(n-k+1)!}} = \frac{\cancel{n!}}{\cancel{k!}\,(n-k)!}\cdot\frac{\cancel{(k-1)!}\,\cancel{(n-k+1)!}}{\cancel{n!}} = \frac{n-k+1}{k}$$

The above only gives us consecutive odds if the probability of success is equal to $\frac{1}{2}$, but it's not hard to see how this expands to arbitrary probability.

Theorem 12.2 (Consecutive odds for the binomial distribution) *For independent events with a probability of success p. Then the odds of k successes relative to k − 1 success are given by:*

$$\frac{p(k \text{ successes in } n \text{ trials})}{p(k-1 \text{ successes in } n \text{ trials})} : 1 = \underbrace{\frac{(n-k+1)(p)}{k(1-p)} : 1}$$

$p = 1/2 \Rightarrow (1-p) = 1/2$

Example 12.3 Let's see how this helps. Say that I have hat with five numbers in it and every week I pull out a number for seven weeks. What is the probability distribution of pulling out the number 4.

replacing #s each time

$p = \frac{1}{5}$
$n = 7$

- $P(0)$ = #4 never gets pulled

$$= \left(1-\frac{1}{5}\right)^7 = \left(\frac{4}{5}\right)^7 = \boxed{0.2097152 = p(0)}$$

- $P(1)$ = #4 appearing exactly once $= \binom{7}{1}\left(\frac{1}{5}\right)^1\left(\frac{4}{5}\right)^{7-1}$

$$\frac{P(1)}{P(0)} = \frac{(n-k+1)(p)}{k(1-p)} = \frac{(7-1+1)\frac{1}{5}}{1(\underbrace{1-\frac{1}{5}}_{\frac{4}{5}})} = \frac{7}{\cancel{5}}\cdot\frac{\cancel{5}}{4} = \frac{7}{4}$$

$$\frac{\frac{1}{5}}{\frac{4}{5}} = \frac{1}{\cancel{5}}\cdot\frac{\cancel{5}}{4} \qquad p/(1-p) = \frac{1}{4}$$

$$p(1) = \frac{7}{4}\,p(0) = \frac{7}{4}\cdot 0.2097152 = \boxed{0.3670016 = p(1)}$$

- $P(2) \Rightarrow \dfrac{P(2)}{P(1)} = \dfrac{(7-2+1)}{2}\cdot\dfrac{1}{4} = \dfrac{6}{2}\cdot\dfrac{1}{4} = \dfrac{3}{4}$ $\qquad p(2) = \frac{3}{4}p(1) = 0.2752512$

- $P(3) \Rightarrow \dfrac{P(3)}{P(2)} = \dfrac{7-3+1}{3}\cdot\dfrac{1}{4} = \dfrac{5}{3}\cdot\dfrac{1}{4} = \dfrac{5}{12}$ $\qquad P(3) = 0.114688$

- $P(4) \Rightarrow \dfrac{P(4)}{P(3)} = \dfrac{4}{4}\cdot\dfrac{1}{4} = \dfrac{1}{4}$ $\qquad p(4) = 0.028672$

- $P(5) \Rightarrow \dfrac{P(5)}{P(4)} = \dfrac{3}{5}\cdot\dfrac{1}{4} = \dfrac{3}{20}$ $\qquad p(5) = 0.0043008$

- $P(6) \Rightarrow \dfrac{P(6)}{P(5)} = \dfrac{2}{6}\cdot\dfrac{1}{4} = \dfrac{1}{12}$ $\qquad p(6) = 0.0003584$

- $P(7) \Rightarrow \dfrac{P(7)}{P(6)} = \dfrac{1}{7}\cdot\dfrac{1}{4} = \dfrac{1}{28}$ $\qquad p(7) = 0.0000128$

$$\underbrace{0.2097152}_{p(0)} + \underbrace{0.3670016}_{p(1)} + \underbrace{0.2752512}_{p(2)} + \underbrace{0.114688}_{p(3)} + \underbrace{0.028672}_{p(4)} + \underbrace{0.0043008}_{p(5)} + \underbrace{0.0003584}_{p(6)} + \underbrace{0.0000128}_{p(7)} = 1$$

13 Most probable outcome

Notice how in the previous example $P(1)$ had the highest value. This implies that the number 4 has the highest probability of appearing exactly 1 time. So it makes sense to ask if we can generalize this. If I have n independent trials and each trial has probability p of being successful then we might expect to have roughly np successes. (For example, if I flip a coin 4 times then I expect $4 \cdot \frac{1}{2} = 2$ heads.) If we let m be the trial which has the highest probability, *i.e.*, $P(m) > P(k)$ for all $k \neq m$, we call m the *mode* of the probability. We end up with the following:

$\lfloor 2.6\cancel{371} \rfloor = 2$

Theorem 13.1 (Mode of the binomial distribution) *If $p \in (0,1)$ where p is the probability of success of an independent trial and we suppose that we perform n trials. Then the mode of the binomial distribution is given by m where*

$$m = \lfloor (np+p) \rfloor$$

mode (highest probable outcome) $\neq$ expect value

Wikipedia: Expected value

A related idea in probability is the expected value. The *expected value* is the expected number of successes. . This is different than the mode which is the most likely number of successes. The expected value is defined to be np and it is usually denoted by μ for the binomial distribution.

The book uses the term *mean* as the expected value of a binomial distribution.

The difference between the two is best seen through examples.

Example 13.2 First let's see when they are the same. If $n = 4$ and $p = \frac{1}{2}$ (aka we flip a coin 4 times) we expect the number of successes to be:

$$\mu = np = 4 \cdot \frac{1}{2} = 2$$

This is the expected value.

The most likely number of successes (aka the mode) is given by:

$$m = \lfloor (np+p) \rfloor = \lfloor 4 \cdot \tfrac{1}{2} + \tfrac{1}{2} \rfloor = \lfloor 2 + \tfrac{1}{2} \rfloor = \lfloor 2.5 \rfloor = 2$$

Notice that they are the same!

Example 13.3 So let's look at when they aren't the same. let's flip the coin a fifth time. So $n = 5$ and $p = \frac{1}{2}$ again.

In this case, the expected number of successes is:

$$\mu = np = 5 \cdot \frac{1}{2} = 2.5$$ Expected Value

So we expect roughly two and a half heads to happen. That is the expected value.

Obviously this doesn't make sense since coin flips are static! We should have whole numbers in this case. That is where the mode kicks in and gives us an actual number. The most likely number of successes is given by:

$$m = \lfloor (np + p) \rfloor = \lfloor 5 \cdot \tfrac{1}{2} + \tfrac{1}{2} \rfloor = \lfloor 2.5 + 0.5 \rfloor = \lfloor 3 \rfloor = 3$$

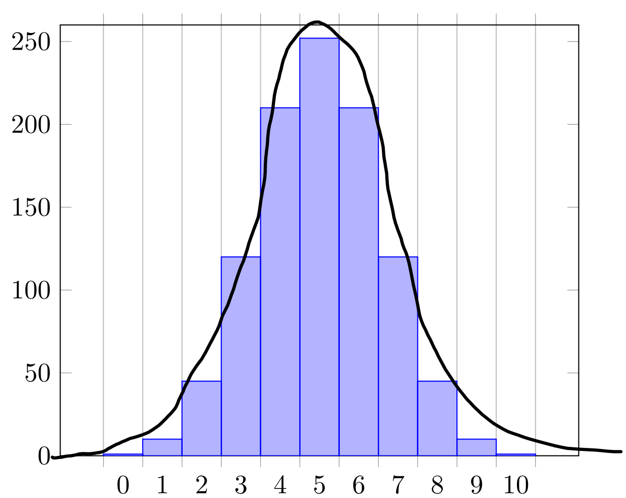
So we expect roughly 3 heads to occur.

This will be more discussed in more detail when we hit chapter three in the text.

14 Normal Distribution

Recall when we were looking at the binomial distribution we had the following example.

Example 14.1 The binomial distribution when $p = \frac{1}{2}$ and $n = 10$ is given by:



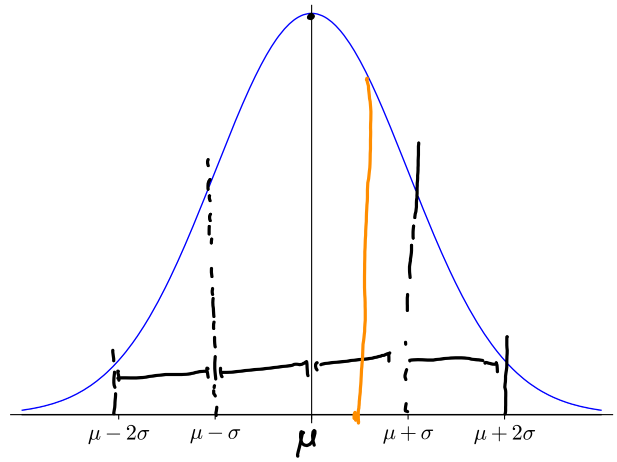
If you notice, we can kind of create a curve from these bars. This gives us what's called the normal curve.

The *normal curve* is the curve with equation:

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

The normal curve is often called the *bell curve* in the non-math world.

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



Five of these symbols you should know:

- x is a variable
- y is a variable (sometimes represented as $f(x)$).
- π is a constant roughly equal to $3.14159265358979\ldots$
- e is a constant roughly equal to $2.7182818285\ldots$
- μ is the *expected value*. $= np$

Wikipedia: Standard deviation

The only variable we haven't discussed so far is σ which is known as the *standard deviation*. Note that μ can be *any* real number while σ must be a strictly positive real number.

"Infinitesimally small width" should make you think of integration/calculus.

The best way to think about the normal curve is to think about it like a continuous histogram. In essence, it's when we start taking our boxes and making their with infinitesimally small. The μ tells us where the peak of the distribution is and the σ tells us how it's spread out.

Note that the constant $\frac{1}{\sigma\sqrt{2\pi}}$ is only there to make sure that the area under the curve is equal to 1.

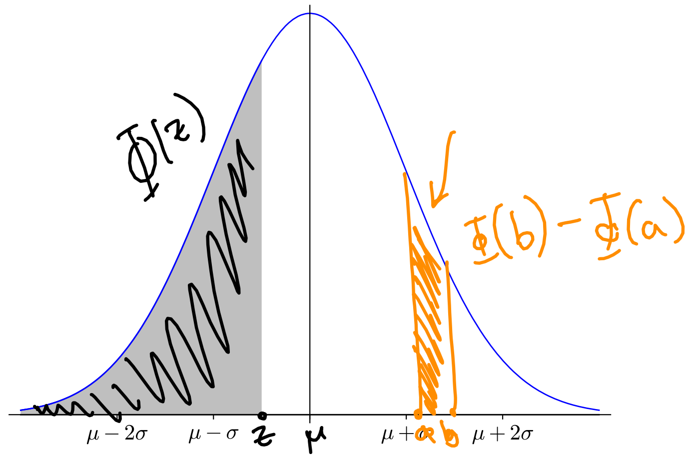
$\int_{-\infty}^{\infty} f(x)\,dx = 1$

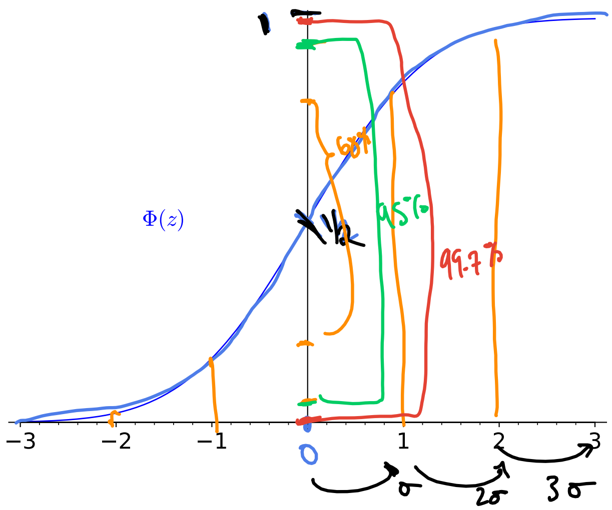
Although we have a curve, that doesn't actually define a distribution. Recall that for a distribution we need to define a probability map. The *normal distribution with expected value μ and standard deviation σ* is the distribution over the x-axis defined by the areas under the normal curve with μ and σ. If $\mu = 0$ & $\sigma = 1$ then we call this distribution the *standard normal distribution*.

Wikipedia: Normal distribution
The normal distribution is sometimes called the "Gaussian distribution".

Another way to look at the normal distribution is through a cumulative distribution function. For this we let $z = \frac{(x-\mu)}{\sigma}$ and we say that z is *x in standard units*. This terminology comes from the fact that in the standard normal distribution $z = x$. We let $\Phi(x)$ denote the probability of everything to the left of z. The function Φ is called the *standard normal cumulative distribution function*.

Wikipedia: Cumulative distribution function
We'll normally just right CDF to mean "cumulative distribution function".





Since we're looking at area under the curve, we know that the standard normal CDF is given by the formula:

$\mu = 0$

$\sigma = 1$

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx$$

The function $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ is known as the *standard normal distribution*. For a normal distribution the probability between a and b is given by:

$$\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

If you recall from the end of your calculus class, $e^{-\frac{1}{2}x^2}$ doesn't have a simple indefinite integral, so $\Phi(z)$ has no simple exact formula unfortunately. But there are methods to calculate it.

Checkout Appendix 5 if you want some values just for fun.

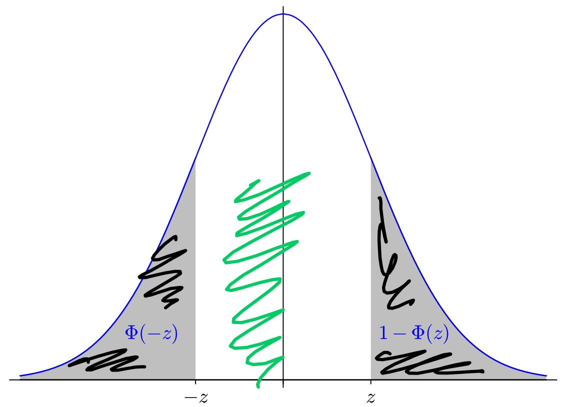
According to the book, there's a nice approximation function for $\Phi(z)$ given by:

$$\Phi(z) \approx 1 - \frac{1}{2}\left(1 + 0.196854z + 0.115194z^2 + 0.000344z^3 + 0.019527z^4\right)^{-4}$$

This approximation will give an approximation up to three significant figures for every $z \geq 0$.

Even without exact formulas there are some things we can deduce directly from the normal curve. For example, we know that

$$\Phi(-z) = 1 - \Phi(z) \text{ and } \Phi(0) = \frac{1}{2}$$



If we let $\Phi(a, b)$ denote the probability on the interval (a, b) then by the difference rule of probabilities we have:

$$\Phi(a,b) = \Phi(b) - \Phi(a)$$

We can also combine some formulas to get:

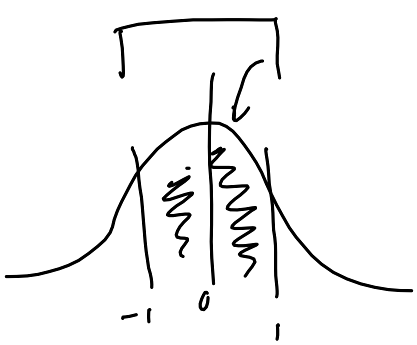
$$\Phi(-z,z) = \Phi(z) - \Phi(-z) = \Phi(z) - (1-\Phi(z)) = 2\Phi(z) - 1$$

It's best not to memorize all of these formulas. It's better to try and know how we are deriving them. You can always look them up if you need to double check.

Plugging in some numbers we see:

$\Phi(-1, 1) \approx$ 68 % $\quad \Phi(-2, 2) \approx$ 95 % $\quad \Phi(-3, 3) \approx$ 99.7 %

From these probabilities you can deduce almost any other probability.



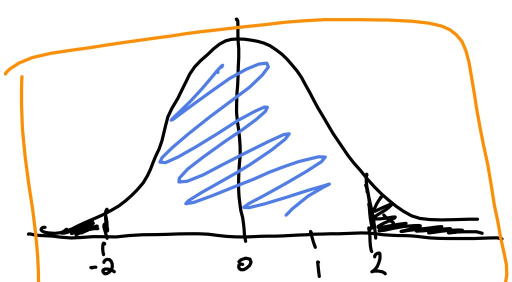
Example 14.2 What is $\Phi(0, 1)$?

$$\Phi(0,1) = \frac{1}{2}\Phi(-1,1) = \frac{1}{2}0.68 = 0.34 \quad 34\%$$

What is $\Phi(2, \infty)$?

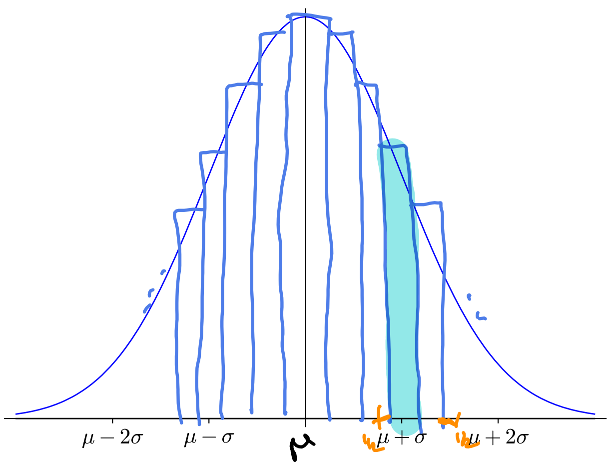
$$\Phi(2,\infty) = \frac{1}{2}\left(1 - \Phi(-2,2)\right) = \frac{1}{2}(1-0.95)$$
$$= \frac{1}{2}(0.05) = 0.025$$
$$\boxed{2.5\%}$$

$$\Phi(2,\infty) = \Phi(-2)$$
$$= 1 - \Phi(2)$$
$$= 1 - 0.9750 = 0.0250$$



15 The normal approximation to the binomial distribution

When we defined the normal curve, we used μ and σ, but we also mentioned that the idea came from looking at the curve of the binomial distribution.



Since the binomial distribution is dependent on n and p, we might ask how are the two pairs of numbers related to one another? One of them, μ, we already know. The expected value μ is just equal to np. But what about σ? We won't go into details now, but it turns out that for sufficiently large n then $\sigma = \sqrt{np(1-p)}$. Remember that we called this σ the *standard deviation*.

Recall that in the normal curve, it doesn't make sense to ask about the probability at a certain point. So instead, let $P(a \text{ to } b)$ be the probability of getting i successes where $i \in [a, b]$ in n independent trials with probability p of success. Then we have that

$P(a \text{ to } b)$ is the proportion of the area under the histogram of the binomial distⁿ b/n $a - \frac{1}{2}, b + \frac{1}{2}$

$\approx$ proportion of area under normal curve b/n $x_1 = a - \frac{1}{2}$ & $x_2 = b + \frac{1}{2}$

$=$ propⁿ of area under standard normal curve b/n $z_1 = \frac{a - \frac{1}{2} - \mu}{\sigma}$, $z_2 = \frac{b + \frac{1}{2} - \mu}{\sigma}$

In other words, the *normal approximation to the binomial distribution* for n independent trials with probability p of success is given by:

$$P(a \text{ to } b) \approx \Phi\left(\frac{b + \frac{1}{2} - \mu}{\sigma}\right) - \Phi\left(\frac{a - \frac{1}{2} - \mu}{\sigma}\right)$$

Let's see how this helps.

Example 15.1 Suppose that I randomly toss a fair coin $n = 100$ times. We want to know the probability of getting 50 heads.

$\binom{n}{k}p^k(1-p)^{n-k}$

By the binomial distribution we have:

$$P(50) = \binom{100}{50}\left(\frac{1}{2}\right)^{50}\left(\frac{1}{2}\right)^{50} = \frac{100!}{50!\,50!\,2^{100}} \approx \underline{0.0795892}$$

By the numerical approximation we have that $a = b = 50$ and $\mu = np = 100 \cdot \frac{1}{2} = 50.$ $\sigma = \sqrt{p(1-p)n} = \sqrt{50 \cdot \left(\frac{1}{2}\right)} = \sqrt{25} = 5$

$$P(50) \simeq \Phi\left(\frac{50+\frac{1}{2}-\mu}{\sigma}\right) - \Phi\left(\frac{50-\frac{1}{2}-\mu}{\sigma}\right)$$

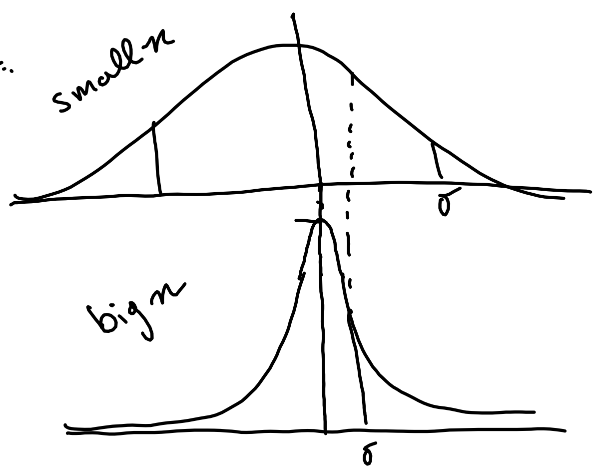
$$= \Phi\left(\frac{50.5-50}{5}\right) - \Phi\left(\frac{49.5-50}{5}\right)$$

$$= \Phi\left(\frac{1}{10}\right) - \Phi\left(-\frac{1}{10}\right)$$

$$= 2\left(\Phi\left(\frac{1}{10}\right)\right) - 1 = 2 \cdot 0.5398 - 1 = \underline{0.0796}$$

$50.5 - 50 = 0.5$

$\frac{\frac{1}{2}}{5} = \frac{1}{2 \cdot 5}$



Fluctuations If we make n grow more and more, we notice that our box sizes get smaller and smaller. This means that our approximations get better and better as n gets bigger. The size of the random fluctuations is of the order of $\sigma = \sqrt{np(1-p)}$ as it tells us how far from μ we are. So as $n \to \infty$ our σ gets larger and more of the distribution is covered.

Likewise, the proportion of successes fluctuates over time too. The typical size of random fluctuations in the relative frequency of successes is of order $\frac{\sqrt{p(1-p)n}}{n} = \sqrt{p(1-p)/n}$. Since $p(1-p)$ is maximized when $p = (1-p) = \frac{1}{2}$, we know $\sqrt{p(1-p)/n} \leq \frac{1}{\sqrt{4n}} = \frac{1}{2\sqrt{n}}$. And as $n \to \infty$ we get that $\frac{1}{\sqrt{n}} \to 0$ making it so that this order goes to 0 as n grows.

$p(1-p)$

$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

$\frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9} < \frac{1}{4}$

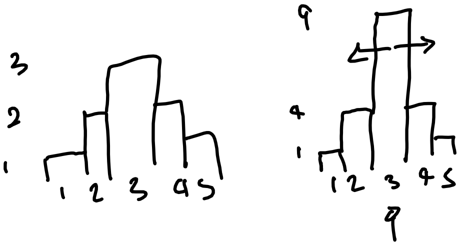
$\frac{1}{n}\left(1-\frac{1}{n}\right) = \frac{n-1}{n^2}$

Theorem 15.2 (Square root law) *In n independent trials with n sufficiently large and probability p of success, then*

- *The number of success will (with high probability) lie in a relatively small interval of numbers, centred on $\mu = np$, with width a moderate multiple of $\sqrt{n}$.*
- *The proportion of successes will (with high probability) lie in a small interval centred on p, with width a moderate multiple of $\frac{1}{\sqrt{n}}$.*

This implies one of my favourite theorems in probability:

Wikipedia: Law of large numbers



$\mu = 3$

Theorem 15.3 (Law of large numbers) *Suppose that n is the number of independent trials and that n is very large. Let p be the probability of success on each trial. Then for every $\varepsilon > 0$*

$$P(\text{ proportion of successes in } n \text{ trials differs from } p \text{ by less than } \varepsilon) \to 1$$

$p \pm \frac{1}{\sqrt{n}}$ $\quad \varepsilon \sim \frac{1}{\sqrt{n}}$

as $n \to \infty$.

What this is saying is that the more trials we conduct, the closer to p our relative frequency will become.

Note that this is sometimes known as the "strong law of large numbers".