MAT220 Lecture 11: The Central Limit Theorem

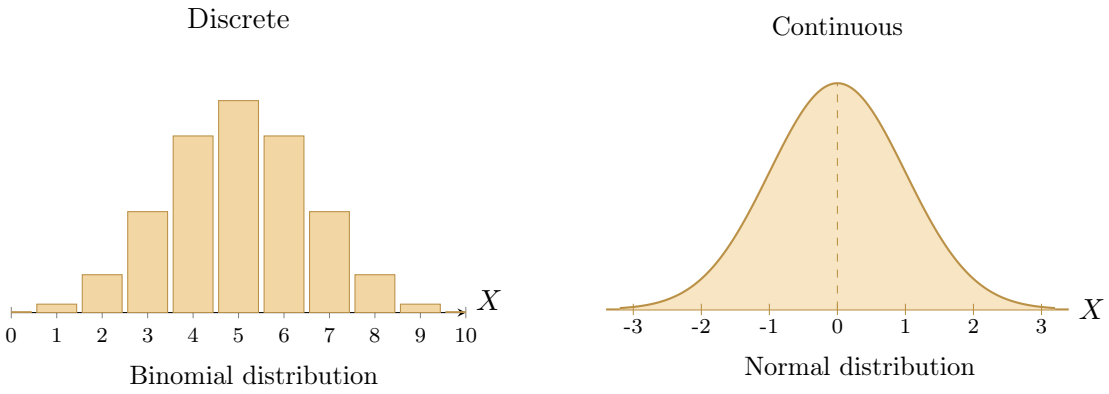
Contents

In Lecture 10, we studied the normal distribution and the use of z-scores. In this lecture, we explain why the normal distribution appears so often in statistics. The key idea is the so-called *Central Limit Theorem*, which gives a method for associating a normal distribution to sample means from many systems, even if the original system is not normally distributed to begin with. In fact, this result is so important that we will dedicate this entire lecture to it. First we will take some care to properly state the result. Then, we will use it to demonstrate how binomial distributions may be approximated with normal curves.

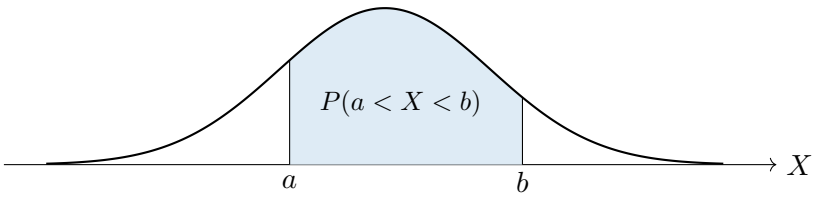
1 A Brief Review of the Normal Distribution

1.1 Discrete and continuous random variables

Last lecture we introduced *normal distributions*, which were probability distributions built from continuous random variables, instead of discrete ones. This looks very different from the binomial distributions that we have seen previously:



Here, the fundamental difference is in the possible values on the x-axis. In the case of a binomial distribution, the random variable counts the number of successes. This gives discrete values to the random variable, and then there are probabilities associated to each individual possible value that X can take. In contrast, for a normal distribution the random variable X can take *any real value*. Instead of taking heights of bars, questions about probabilities are then phrased in terms of *areas*: the probability that X takes a value in some region is given by the size of the area under the curve, as pictured below.



Here, the probability of finding our random variable X within the region between a and b is given by calculating the size of the blue shaded region.

1.2 Describing a normal distribution

Generally speaking, the mean and the standard deviation represent two different ways to describe certain features of a data set. In the case of the normal distribution, these two numbers are the only pieces of information that we need to define the entire distribution. This means that we can write down a normal distribution in mathematical notation as follows.

Normal distribution

A normal distribution is a continuous probability distribution with notation

$$X \sim \mathcal{N}(\mu, \sigma^2).$$

Here μ is the mean, σ is the standard deviation, and σ^2 is the variance.

The mean and standard deviation are enough to fully describe a normal distribution. As with data sets, the mean μ tells us where the centre of the distribution is, and the standard deviation σ tells us how spread out the distribution is. Normal distributions always satisfy the following useful properties.

Four important properties

A normal distribution has the following properties.

1. The curve is bell-shaped, with its highest point over the mean μ.

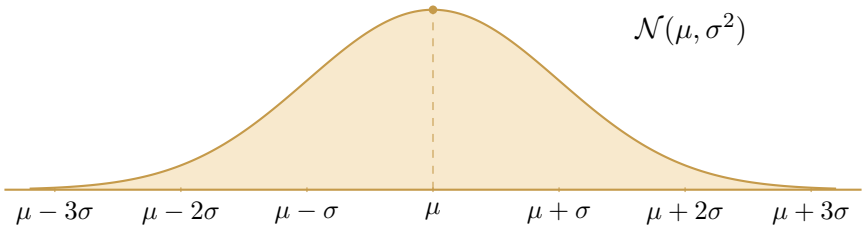
2. The curve is symmetric about the vertical line through μ.

3. The tails approach the horizontal axis but never touch or cross it.

4. The total area under the entire curve is 1.

1.3 z-scores

The four properties written above mean that we can navigate a normal distribution as follows:



Here, the graph is symmetric and centered directly at μ. Since σ is defined in the same physical units as μ, we can start from the centre of the graph and "walk outwards" in increasing values of σ. This allows us to describe potential values of X in terms of their distance from μ. For instance, if X takes a value x that is far away from μ, then $x - \mu$ might be 3σ or -3σ, and if X takes a value relatively close to μ, then it could be 0.5σ away.

If we wanted to get rid of the σ in our measure of "distance from μ", then we can simply divide by σ to be left with a pure number. This number is known as the z-score:

$$z = \frac{x - \mu}{\sigma}.$$

In a nutshell, the z-score of a value x counts how far x is from μ *in terms of* σ. The larger z is, the further away x is from μ. If the overall z-score is negative, then that means the value x is to the left of μ, and if the z-score is positive, then that means x is to the right of μ.

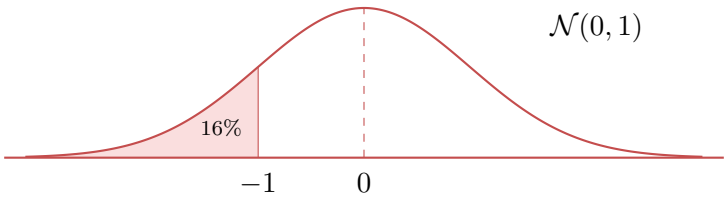
1.4 The standard normal distribution

Throughout Lecture 10 we also spent some time discussing the "standard" normal distribution. The idea here was that this is just about the simplest possible version of a normal distribution: the mean equalled 0 and the standard deviation was 1. Since this particular distribution is so well-behaved, we used the special letter Z to write its random variable. In mathematical notation, the standard normal distribution can be written $Z \sim \mathcal{N}(0, 1)$.

Conveniently, we can approximate areas (and therefore probabilities) of the standard normal distribution by using z-tables. These tables list out the areas to the left of common values of Z. For instance, a small version of this table might look like:

z	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0
$P(Z < z)$	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.98

More detailed tables can be found in the appendix of Lecture 10. Visually, entries in this table tell us the area to the *left* of a given value. For example, the table says that $P(Z < -1) \approx 0.16$, meaning that the area to the left of -1 in $\mathcal{N}(0, 1)$ is approximately 16% of the total area:



Tables such as that above only provide probabilities to the *left* of given values. Other probabilities can be calculated using the properties of the normal distribution. Some useful facts are listed below.

General rules for probabilities

For $Z \sim \mathcal{N}(0, 1)$:

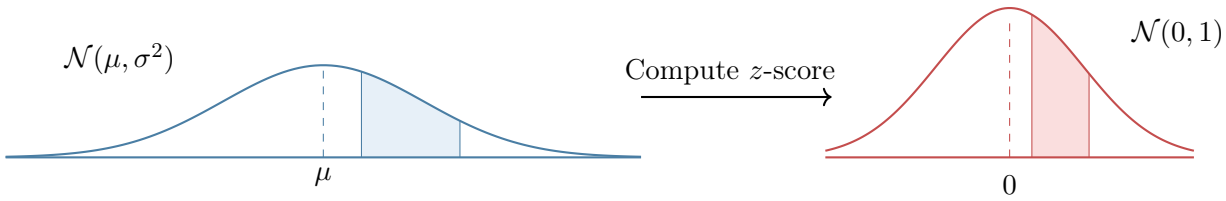
- $P(Z > z) = 1 - P(Z < z)$
- $P(z_1 < Z < z_2) = P(Z < z_2) - P(Z < z_1)$
- $P(Z > z) = P(Z < -z)$
- $P(Z = z) = 0$

1.5 Calculating other probabilities

If $X \sim \mathcal{N}(\mu, \sigma^2)$ is a general normal distribution, then we can convert a value x into the standard normal distribution by computing its z-score:

$$z = \frac{x - \mu}{\sigma}.$$

This transformation preserves area, and therefore the probability of finding X within a particular region can be calculated by first converting to the standard normal distribution, and then performing the calculation there:



Example: heights

To illustrate this point, suppose that male heights in Japan are normally distributed with mean 171 cm and standard deviation 5.6 cm. We can calculate, for example, the probability of meeting somebody taller than 180 cm by working with z-scores. Here we are looking for $P(X > 180)$, so we can compute the relevant z-score:

$$z = \frac{180 - 171}{5.6} \approx 1.61.$$

This transformation preserves the probability, so the answer to our question can be found by computing:

$$P(X > 180) \approx P(Z > 1.61),$$

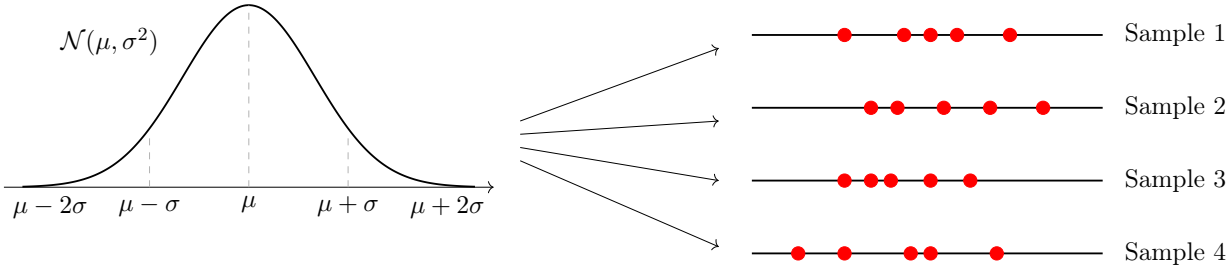
which happens to be about 5.4%.

2 Sampling Distributions

2.1 Sampling a random variable

Consider now a general normal distribution $X \sim \mathcal{N}(\mu, \sigma^2)$. As we previously mentioned, the probability of finding X within any given region can be determined by the area under the curve at that region. Therefore, we would expect that X is more likely to land in regions that correspond to larger areas.

Suppose now that we take a *sample* of random values from this distribution. For the sake of simplicity, let's pick 5 values randomly. Using the terminology of Lecture 1, this is known as a *simple random sample of size* 5. Generally, these 5 values could land anywhere on the x-axis. So, if we were to keep taking simple random samples of size 5 over and over again, they may look quite different:



For each individual sample, we can always compute a sample mean:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5}.$$

Now, here is the crucial observation: since our sample consists of random values, the exact value of the sample mean $\bar{x}$ will also be random. Therefore, we can interpret the sample mean as a random variable in its own right, and we write this random variable as $\bar{X}$. We will now explore what the distribution of the random variable $\bar{X}$ looks like.

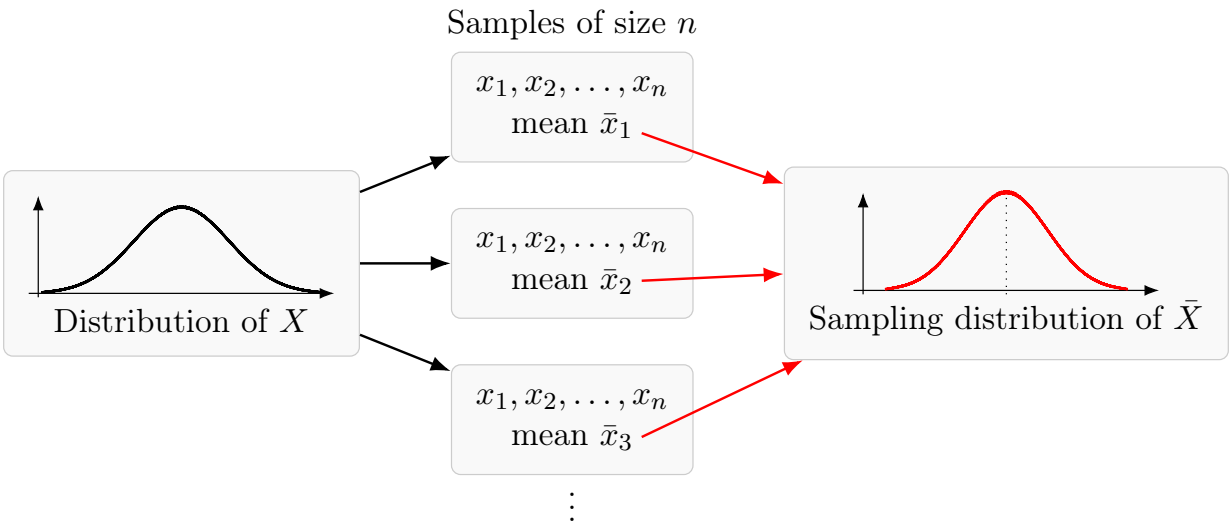
2.2 The sampling distribution of the sample mean

Imagine taking all possible simple random samples of size n from a population. We can compute the sample mean associated to each sample and collect all of these results together into a new probability distribution. This new object is known as a *sampling distribution*.

Sampling distribution of the sample mean

Given a probability distribution, the **sampling distribution of $\bar{X}$ for samples of size n** is the probability distribution formed from all sample means taken from all simple random samples of size n.

We can imagine the construction of a sampling distribution as something like this:



In other words, we start with a probability distribution, take all the simple random samples of size n, compute all of the associated sample means $\bar{x}$, and then collect them all together. Since the samples were randomly drawn to begin with, the sample means $\bar{x}$ can be treated as values of a random variable $\bar{X}$, and consequently, the collection of all possible $\bar{x}$'s has the structure of a probability distribution.

A sample of size n should be thought of as random variables $X_1, \ldots, X_n$, each with the same distribution as X, and

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n}.$$

3 The Central Limit Theorem

As we might expect, different probability distributions will generally have different sampling distributions. The Central Limit Theorem is a result that tells us how these sampling distributions will look, depending on the shape of the original distribution. We will now present two related results: one for normal distributions, and another for more general probability distributions.

Throughout this course we will not prove these theorems. Instead, our main focus will be on the *application* of the Central Limit Theorem in techniques like Estimation and Hypothesis Testing. So, our goal is simply to state and understand the result.

3.1 Version 1: the sampling distribution when the original population is normal

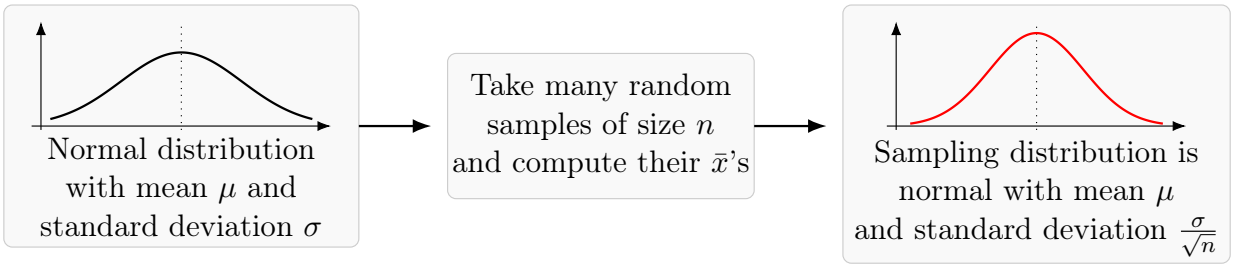
Let's first suppose that our original probability distribution is normal, i.e. $X \sim \mathcal{N}(\mu, \sigma^2)$. Our first result relates the sampling distribution of $\bar{X}$ to the parameters μ and σ.

The Central Limit Theorem: Version 1

Let X be a random variable such that $X \sim \mathcal{N}(\mu, \sigma^2)$. Then:

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

This result says that when we start with a normal distribution with mean μ and variance σ^2, the collection of all sample means will arrange themselves as another normal distribution, with the same center μ but now with variance $\frac{\sigma^2}{n}$:



Roughly speaking, if we take a sample of values from $\mathcal{N}(\mu, \sigma^2)$, then the sample mean $\bar{x}$ will average out the random noise in these individual values and return something more central. Therefore, on average, we would expect these sample means to be closer to the true center of the distribution. This idea partially explains why the variance $\frac{\sigma^2}{n}$ of the sampling distribution is smaller than the original distribution. As we can see from the formula, the larger the value of n, the more tightly packed the sampling distribution will be – this is because samples of larger size are less sensitive to individual outliers which may skew the average.

3.2 Version 2: when the original population is not necessarily normal

The first result says that if we build a sampling distribution of $\bar{X}$ from a normal distribution, then the result is another normal distribution. Amazingly, there is a much stronger version of the Central

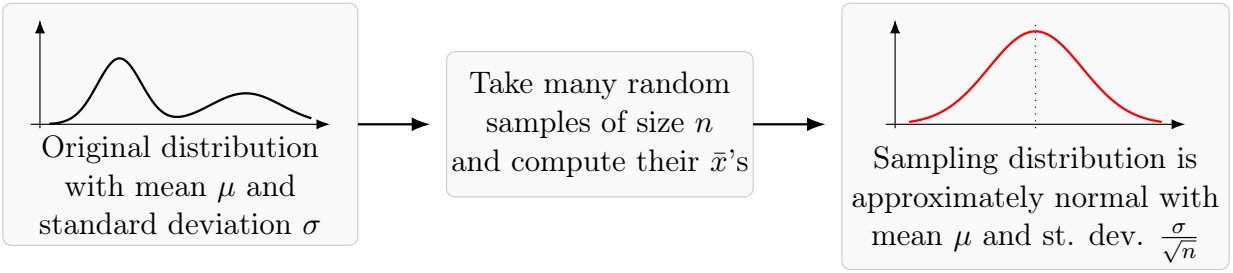
Limit Theorem:

Central Limit Theorem

Suppose X has a probability distribution with finite mean μ and finite variance σ^2. For a large enough sample size n (often $n \geq 30$), the sample mean $\bar{X}$ is approximately normally distributed, with parameters:

$$\bar{X} \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

In other words, the Central Limit Theorem says that the sampling distribution of $\bar{X}$ will look approximately normal, with mean μ and variance $\frac{\sigma^2}{n}$.



In both of our results, the standard deviation of the sampling distributions had the same general form. As a matter of fact, this form has a special name: it is called the *standard error*.

Standard error

The standard deviation of the sampling distribution of $\bar{X}$ is

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}.$$

This is called the **standard error**.

The standard error measures how much sample means vary from sample to sample. As the sample size n increases, the standard error decreases, meaning that the resulting sampling distribution will be more tightly packed around the center μ. As mentioned previously, this matches our intuition that larger samples are less sensitive to outliers, i.e. they give more stable averages.

3.3 Why the Central Limit Theorem matters

In the real world we don't usually have access to population parameters. For instance, if we were performing a study on the average height of people in Tokyo, then we cannot go out and measure the heights of the millions of people in the city. Instead, in reality, we will only ever be able to take *samples* of the population and compute some sample statistics.

The Central Limit Theorem provides a mathematical way to connect sample statistics to the true population mean. Although we may never understand the true structure of the population's distribution, we can at least expect that if we take our samples to be large enough (usually $n \geq 30$), then the distribution of all their sample means is approximately normal. In particular, the center of

the sampling distribution is still the *true* population mean. Thus, the Central Limit Theorem offers us some extra expressivity – we can now talk about the theoretical relationship between a sample mean $\bar{x}$ and the population mean μ. This idea of inferring population parameters from sample statistics lies at the beating heart of Inferential Statistics. In the next few lectures we will explore this idea in great detail: we will see some techniques for actually *estimating* μ based on some sample data.

3.4 A Worked Example: Guinea Pig Weights

Suppose that a large population of guinea pigs has mean 950g and standard deviation 150g. Assume that the population is not normally distributed. Now suppose we take a random sample of 30 guinea pigs and compute the sample mean $\bar{x}$. We will answer the question:

> What is the probability that the average weight of 30 guinea pigs is between 900g and 1000g?

Since $n = 30$, we can use the Central Limit Theorem to approximate the sampling distribution of $\bar{X}$:

$$\bar{X} \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) = \mathcal{N}\left(950, \frac{150^2}{30}\right).$$

Equivalently, the standard error is

$$\sigma_{\bar{X}} = \frac{150}{\sqrt{30}} \approx 27.39.$$

So we may also write

$$\bar{X} \approx \mathcal{N}(950, 27.39^2).$$

To answer the question above, we are looking to find the probability $P(900 \leq \bar{X} \leq 1000)$. To do so, we need to first convert both endpoints to z-scores:

$$z_1 = \frac{900 - 950}{150/\sqrt{30}} \approx -1.83,$$

and

$$z_2 = \frac{1000 - 950}{150/\sqrt{30}} \approx 1.83.$$

Therefore,

$$P(900 \leq \bar{X} \leq 1000) \approx P(-1.83 \leq Z \leq 1.83).$$

Using a standard normal table, we have:

$$P(Z < 1.83) = 0.96638 \qquad \text{and} \qquad P(Z < -1.83) = 0.03362,$$

from which we can conclude that:

$$P(900 \leq \bar{X} \leq 1000) \approx P(-1.83 \leq Z \leq 1.83) = 0.96638 - 0.03362 = 0.93276.$$

So, the probability of taking a random sample of 30 guinea pigs with an average weight between 900g and 1000g is about 93%.

3.4.1 Why this is useful

If we were forced to calculate this probability directly, then we would be left with a very awkward computation. Firstly, if we tried to list all possible samples of 30 guinea pigs from 1000 guinea pigs, the number of possible samples would be

$$\binom{1000}{30} \approx 2.43 \times 10^{57}.$$

To calculate the exact probability directly using our Golden Formula for classical probability, we would need to:

1. list all possible samples of size 30;
2. compute the sample mean for every sample;
3. count how many of those sample means lie between 900 and 1000, and then
4. divide by $C_{1000,30}$, the total number of possible samples.

Of course, this would be much more difficult than simply approximating the solution via the Central Limit Theorem.

4 Normal Approximation to the Binomial Distribution

4.1 The Problem with Large Binomial Experiments

Recall from Lecture 8 that a binomial experiment is a probabilistic system with the following properties:

1. the same trial is repeated several times,
2. there are only ever two possible outcomes to each trial (which we called *success* and *failure*),
3. each trial has the same probabilities, and
4. the probabilities are independent.

Binomial distributions were specified by two main numbers: n, which denoted the number of trials, and p, which denoted the probability of success. The central question in a binomial experiment is to figure out the exact probability of getting a particular number, or range of numbers, of successes in n attempts. To properly write this mathematically, we defined a discrete random variable X whose job was to count the number of successes in n trials. We then saw that the probability of getting r successes in n trials was given by the formula:

$$P(X = r) = C_{n,r} p^r q^{n-r},$$

where here $q = 1 - p$.

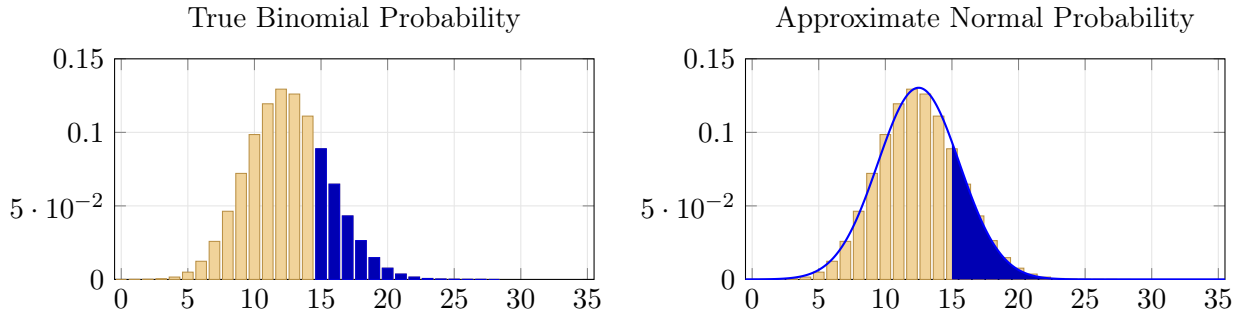
Imagine now that we have a rather large binomial experiment, say $n = 50$ with $p = 0.25$. If we were asked to compute a complicated probability such as $P(X \geq 15)$, then this would require us to compute many individual probabilities and add them up:

$$\begin{aligned} P(X \geq 15) &= P(X = 15) + P(X = 16) + \cdots + P(X = 49) + P(X = 50) \\ &= C_{50,15} p^{15} q^{35} + C_{50,16} p^{16} q^{34} + \cdots + C_{50,49} p^{49} q^1 + C_{50,50} p^{50} q^0, \end{aligned}$$

or perhaps via the complement:

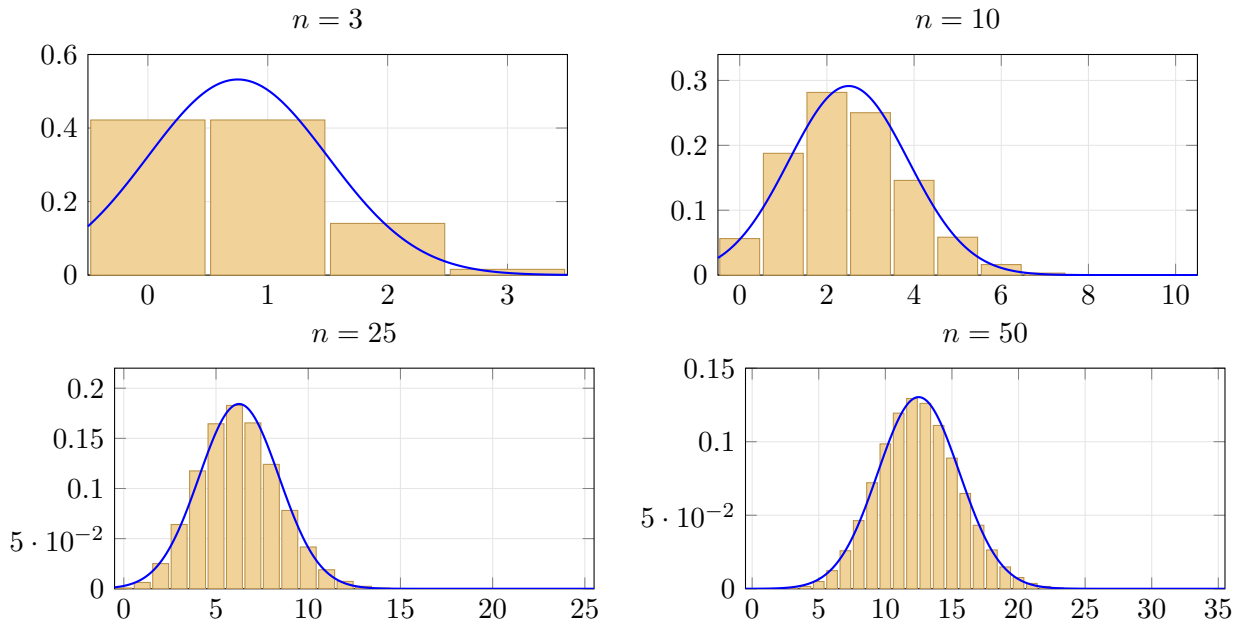
$$\begin{aligned} P(X \geq 15) &= 1 - P(X < 15) \\ &= 1 - (P(X = 0) + P(X = 1) + \cdots + P(X = 13) + P(X = 14)) \\ &= 1 - P(X = 0) - P(X = 1) - \ldots - P(X = 13) - P(X = 14) \\ &= 1 - \left(C_{50,0}p^0q^{50}\right) - \left(C_{50,1}p^1q^{49}\right) - \cdots - \left(C_{50,13}p^{13}q^{37}\right) - \left(C_{50,14}p^{14}q^{36}\right) \end{aligned}$$

Either way we try to do this, we are met with a large and difficult computation. Instead, in a situation like this, it may be easier to simply approximate the distribution with a normal curve and use z-scores to get a rough answer. Graphically, the true probability corresponds to the sum of the heights of all of the bars from 15 to 50, versus computing the area to the right of 15 underneath a normal curve that approximately matches the distribution:



Of course, this point becomes clearer if we consider a much larger binomial experiment, say $n = 1000$. Suddenly, calculating a probability like $P(X \geq 400)$ would involve 601 different computations of the formula $C_{n,r}p^rq^{n-r}$, followed by a rather large sum.

In Lecture 10 we motivated normal distributions by seeing how binomial distributions begin to resemble normal curves after appropriate scaling. Visually, taking n to be larger and larger caused the number of bars to increase until the image began to resemble a continuous curve. This means that as n increases, the binomial distribution will start to look more and more like a normal curve, and therefore the approximation of probabilities using z-scores will become more and more appropriate. In the image below we fit normal curves to smaller binomial distributions. Notice the normal curves start to fit the binomial distributions better and better as n gets larger:



We will now see how these normal curves can be created via an application of the Central Limit Theorem.

4.2 Applying the Central Limit Theorem to a Binomial Distribution

Suppose that we have a binomial experiment with fixed n, p and q. Let X be the random variable that counts the number of successes. In a way, we can understand each individual trial as a separate probabilistic system, with its own random variable attached. If we define:

$$Y_i = \begin{cases} 1 & \text{if trial } i \text{ is successful} \\ 0 & \text{if trial } i \text{ fails,} \end{cases}$$

then X can be rewritten as the sum:

$$X = Y_1 + Y_2 + \cdots + Y_{n-1} + Y_n.$$

Each term in this sum contributes either a 1 for success, or a 0 for failure. Therefore the sum X will count the total number of successes after n trials. We now notice: if we divide X by n, then we get something that looks conveniently like a sample mean:

$$\frac{X}{n} = \frac{Y_1 + Y_2 + \cdots + Y_{n-1} + Y_n}{n}.$$

The Central Limit Theorem says that the sample mean of the random variables $Y_1, \ldots, Y_n$ will arrange itself approximately into a normal distribution $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$, where μ and σ^2 are the mean and variance of a single trial variable Y_i. In this case, our original variable is one of the Y_i – the variable associated to a single trial of the experiment. When working with Binomial experiments, we know that the analogue of the mean is the expected value. For the single variable Y_i, we have:

$$Y_i = \begin{cases} 1 & \text{if the trial is successful} \\ 0 & \text{if the trial fails.} \end{cases}$$

The expected value[1] is:

$$\mathrm{E}(Y_i) = P(Y_i = 0) \cdot 0 + P(Y_i = 1) \cdot 1 = q \cdot 0 + p \cdot 1 = p.$$

Moreover, the variance is:

$$\begin{aligned}
\mathrm{Var}(Y_i) &= \mathrm{E}\left((Y_i - \mathrm{E}(Y_i))^2\right) \\
&= \mathrm{E}\left((Y_i - p)^2\right) \\
&= P(Y_i = 0)(0 - p)^2 + P(Y_i = 1)(1 - p)^2 \\
&= q(-p)^2 + p(q)^2 \\
&= qp^2 + pq^2 \\
&= pq(p + q) \\
&= pq.
\end{aligned}$$

To summarise: the variable $\frac{X}{n}$ acts as a sort-of sample mean of the outcomes of each individual trial in the binomial experiment. Each of these trials has the same expected value equal to p, and the same variance equal to pq. Applying the Central Limit Theorem tells us that $\frac{X}{n}$ is approximately normal, specifically:

$$\frac{X}{n} \approx \mathcal{N}\left(p, \frac{pq}{n}\right).$$

We can use this information to create a new approximation for our original variable X. Some complicated algebra involving normal distributions[2] gives us the following fact:

$$\frac{X}{n} \approx \mathcal{N}\left(p, \frac{pq}{n}\right) \qquad \Longrightarrow \qquad X \approx \mathcal{N}\left(np, npq\right).$$

Thus we have arrived at a normal approximation of X.

Normal approximation to a binomial distribution

If X is binomially distributed according to n trials with success probability p, and using the rough rule $np > 5$ and $nq > 5$, then X can be approximated by the normal distribution:

$$X \approx \mathcal{N}(np, npq).$$

Written differently, we say that $X \approx Y$, where $Y \sim \mathcal{N}(np, npq)$.

[1] Recall that, for a discrete random variable, the expected value of a random variable is roughly

$$\mathrm{E}(X) = \sum \text{probability} \times \text{outcome}$$

.

[2] If I must: we can substitute $\frac{x}{n}$ into the probability density function that defines the Normal distribution to see that

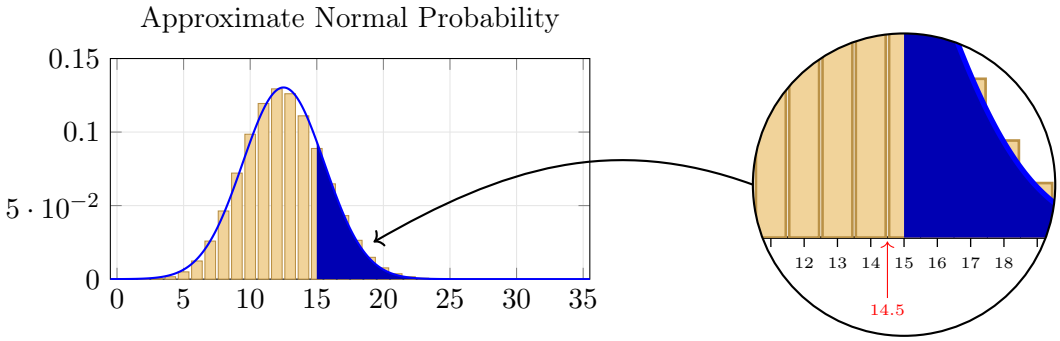
$$f\left(\frac{x}{n}\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\frac{x}{n} - \mu)^2}{2\sigma^2}\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - n\mu)^2}{2n^2\sigma^2}\right).$$

This *almost* looks like the definition of $\mathcal{N}(n\mu, n^2\sigma^2)$, except for the first term. If we add an extra $\frac{1}{n}$ to the front of the expression, then we notice that $\frac{1}{n} f\left(\frac{x}{n}\right)$ now matches perfectly with the definition of $\mathcal{N}(n\mu, n^2\sigma^2)$.

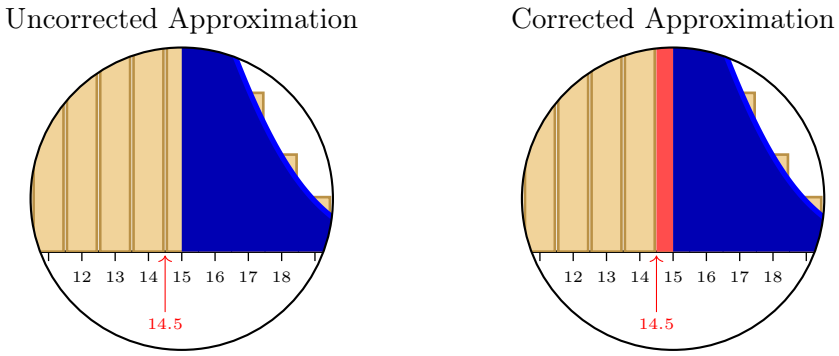
Note that the conditions $np > 5$ and $nq > 5$ are imposed in order to make sure that the original binomial distribution is large enough and not too skewed to have a meaningful approximation. Notice, for instance, the diagrams in Section 4.1: the approximation of the $n = 3, p = 0.25$ case doesn't really fit the original distribution.

4.3 The Continuity Correction

According to the previous section, when a binomial distribution is large enough and not too skewed, we can approximate it with a normal distribution $\mathcal{N}(np, npq)$. The latter then allows us to compute z-scores and derive probabilities of areas. Although potentially useful, there is a genuine geometrical issue when approximating bars with a smoother area. Let's consider again the example of $n = 50$ and $p = 0.25$ from Section 4.1. If we zoom into the diagram, we notice that the approximate continuous area happens to cut the bar corresponding to $X = 15$ in half:



Therefore, the area corresponding to the blue region is going to be smaller than the true probability. We can correct this by making our area approximation start counting from the *start* of the bar, instead of in the middle:



Generally, this will give a closer approximation than the uncorrected version. This adjustment is called the *continuity correction*.

Rules for continuity corrections

When approximating a binomial random variable X by a normal random variable Y, move each cutoff by 0.5 so that the normal area fully covers the bars of the binomial distribution.

Binomial probability		Normal approximation
$P(X \geq r)$	becomes	$P(Y \geq r - 0.5)$
$P(X > r)$	becomes	$P(Y > r + 0.5)$
$P(X \leq r)$	becomes	$P(Y \leq r + 0.5)$
$P(X < r)$	becomes	$P(Y < r - 0.5)$
$P(a \leq X \leq b)$	becomes	$P(a - 0.5 \leq Y \leq b + 0.5)$

For example, if we wanted to approximate $P(15 \leq X)$ then we would compute $P(14.5 \leq Y)$, and if we wanted to approximate $P(X \leq 12)$ then we would compute $P(Y \leq 12.5)$.

5 Worked Example: Vintage Clothes

We will now walk through an example of a binomial approximation. Suppose that I head down to Shimokitazawa to try and buy some cool vintage clothes. From past experience, I know that there is about a 25% chance that I will find something interesting to buy in any given store. We assume the stores are independent and that this probability stays the same from store to store. I walk into 25 different stores. We will try to find an approximate answer to the question:

What is the probability that 8 or more of the 25 stores will have good clothes?

Assuming that we can model this as a binomial experiment, let's define the random variable

$$X = \text{the number of stores with good clothes.}$$

By assumption, the variable X follows a binomial distribution with $n = 25$, $p = 0.25$ and $q = 1 - 0.25 = 0.75$.

5.1 Method 1: an exact binomial calculation

For future reference, we will first compute this result exactly. We are being asked to find $P(X \geq 8)$. This can be calculated directly as:

$$\begin{aligned} P(X \geq 8) &= P(X = 8) + P(X = 9) + \cdots + P(X = 25) \\ &= C_{25,8}(0.25)^8(0.75)^{17} + C_{25,9}(0.25)^9(0.75)^{16} + \cdots + C_{25,25}(0.25)^{25}(0.75)^0. \end{aligned}$$

Using a calculator, we find that:

$$P(X \geq 8) \approx 0.27349.$$

5.2 Method 2: normal approximations

In order to use a normal approximation, first we have to check the approximation conditions:

$$np = 25(0.25) = 6.25 \qquad \text{and} \qquad nq = 25(0.75) = 18.75.$$

These are both greater than 5, so a normal approximation is reasonable. Since we have $n = 25$, $p = 0.25$ and $q = 0.75$, the approximating normal distribution is:

$$X \approx \mathcal{N}(np, npq) = \mathcal{N}(6.25, 4.6875).$$

Without applying a continuity correction, we calculate an approximate probability by computing the z-score:

$$z = \frac{8 - 6.25}{\sqrt{4.6875}} \approx 0.81.$$

Therefore:

$$P(X \geq 8) \approx P(Z > 0.81) = 1 - P(Z < 0.81) \approx 1 - 0.79103 = 0.20897.$$

As we can see, this is noticeably smaller than the exact answer of 0.27349. However, if we make the continuity correction before computing the probability, we would change $P(X \geq 8)$ into $P(Y \geq 7.5)$. Plugging this new value into the z-score formula, we get:

$$z = \frac{7.5 - 6.25}{2.17} \approx 0.58 \qquad \Longrightarrow \qquad P(Y \geq 7.5) \approx P(Z \geq 0.58) \approx 0.28096.$$

Notice that this is significantly closer to the true value of 0.27349.