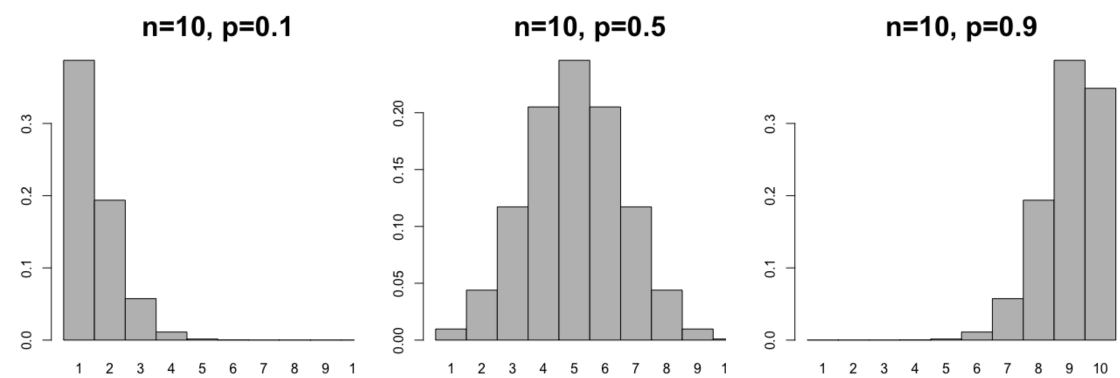
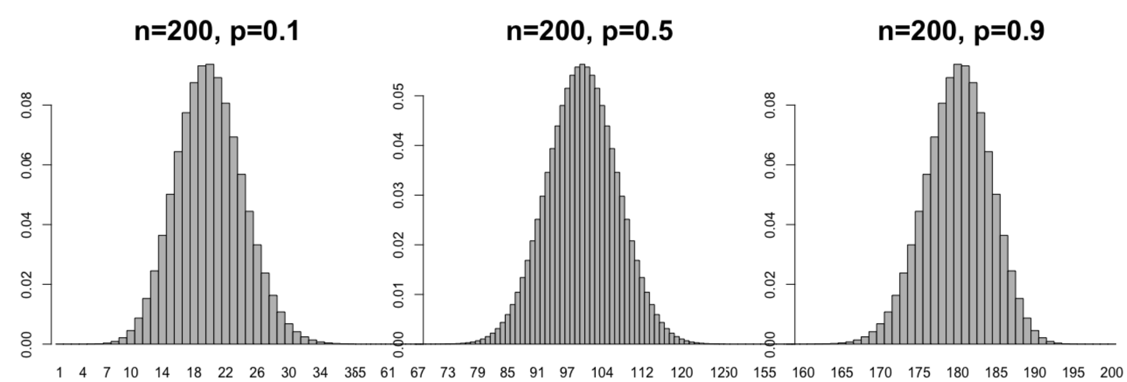
3.6.6 Normal Approximation to the Binomial Distribution

Sometimes we may want to approximate a discrete binomial random variable with a continuous normal random variable. This is nice because the binomial formula becomes computationally heavy for large values of n. This will become extremely important later when we explore statistical inference.

Let's first consider a binomial random variable with several different values of p while holding n constant at $n = 10$.



When $p = 0.5$ it seems like the distribution is symmetrical and roughly bell shaped, when $p = 0.1$ it becomes right skewed, and $p = 0.9$ produces a left skew distribution. It seems that we may only be able to use a normal distribution as an approximation if $p = 0.5$. However, let's now keep our three values of p, and now increase the size of n to $n = 200$.



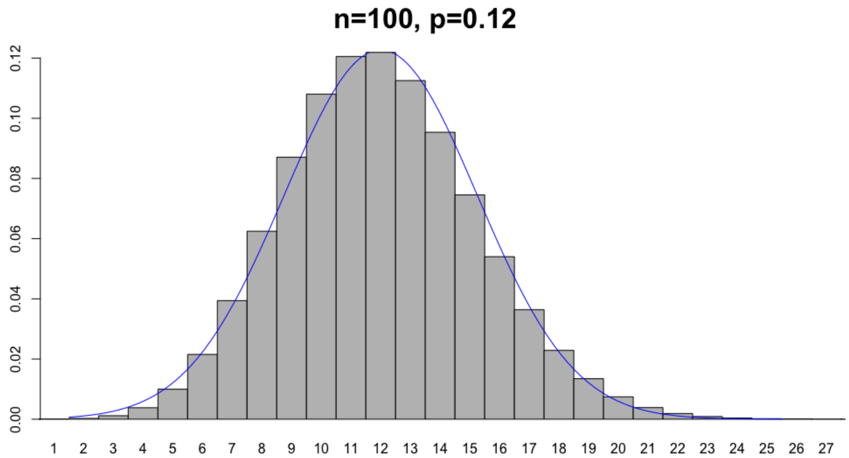
This is pretty cool, as n gets large our binomial distribution will look more and more normal. For a sufficiently large value of n any binomial distribution can be approximated with a normal distribution. We will put a rough guideline to decide how to determine what value of n is 'sufficient'.

$$\text{If} \quad np \geq 10 \quad \text{and} \quad n(1-p) \geq 10$$

We may approximate a binomial distribution with a normal distribution. This can also be seen as 10 'successes' and 10 'failures' for our binomial random variable. Also note that for our approximation we can use $\mu = np$ and $\sigma = \sqrt{np(1-p)}$ which is the mean and standard deviation for our original binomial random variable.

Example 7: Suppose that a stormtrooper has a 0.12 chance of making a shot when blasting rebel scum. Use a normal distribution to estimate the probability that a stormtrooper makes 18 or more shots in 100 successive shots.

First we must check that this approximation is appropriate given our values of n and p. We have $np = 100(0.12) = 12$ and $n(1-p) = 100(88) = 88$ so we may use the normal approximation. Shown below is both the actual binomial distribution for the problem, with the normal distribution used for approximation superimposed.

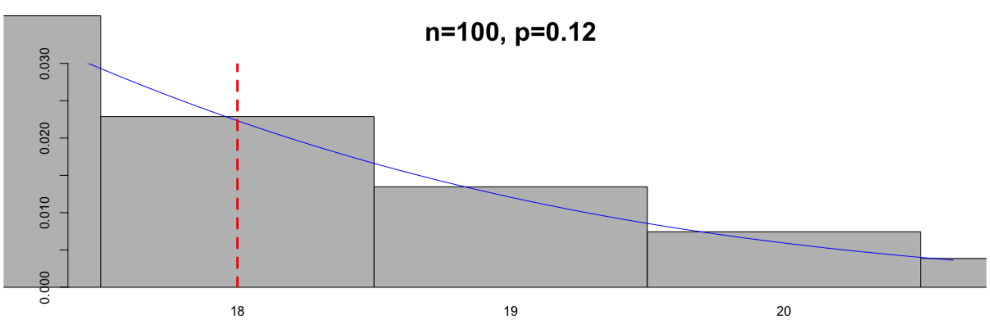


The normal approximation curve has mean $\mu = np = 100(12) = 12$, and standard deviation $\sqrt{np(1-p)} = \sqrt{100(0.12)(0.88)} = 3.249615$. We can now find a z-score for probability.

$$\begin{aligned} z &= \frac{x-\mu}{\sigma} \\ &= \frac{18-12}{3.249615} \\ &= 1.846373 \end{aligned}$$

We would like $P(Z > 1.846373) = 1 - P(Z \leq 1.846373) = 0.03241903$. There is roughly 3.2% chance of the stormtrooper making more than 18 shots.

We can also improve our approximation with a **continuity correction**. This is an adjustment to take into account that we are using a continuous distribution for our approximation instead of a discrete distribution. Consider the last example.



When we zoom in on the distribution you may notice that using the normal approximation completely disregards the left half of the $X = 18$ bar for the discrete distribution. We will get a better approximation if we use 17.5 for our approximation.

When using a normal distribution to approximate a binomial distribution we may summarize the continuity correction factor as follows.

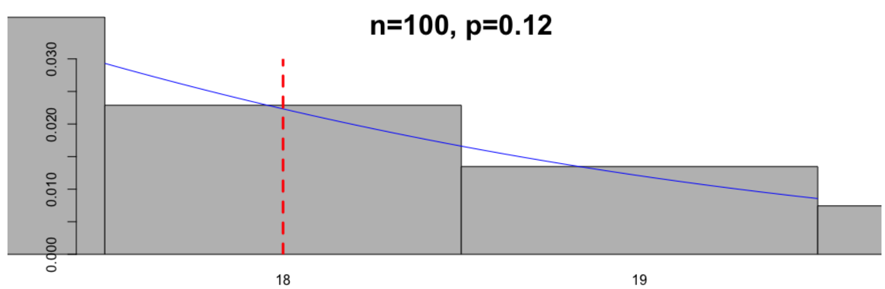
Binomial	**Normal Approximation**
$P(X = x)$	$P(x - 0.5 \leq X \leq x + 0.5)$
$P(X \leq x)$	$P(X \leq x + 0.5)$
$P(X < x)$	$P(X \leq x - 0.5)$
$P(X > x)$	$P(X \geq x + 0.5)$
$P(X \geq x)$	$P(X \geq x - 0.5)$

I do not recommend memorizing this table as the results are easy to obtain with a simple sketch of the two distributions.

Example 8: Suppose that a stormtrooper has a 0.12 chance of making a shot when blasting rebels.

(a) Use a normal distribution to estimate the probability that a stormtrooper makes exactly 18 shots in 100 successive shots. Use the continuity correction factor.

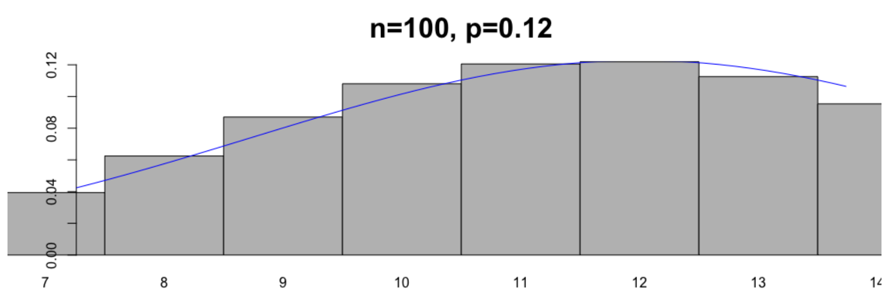
From the last problem we know that an approximation using a normal distribution is appropriate with $n = 100$ and $p = 0.12$. Lets zoom in on 18 to visualize how to use the continuity correction factor in our approximation.



Notice that to approximate the probability mass for $X = 18$, we would like the area under the pdf from 17.5 to 18.5. Let Y be our normal distribution. We want $P(17.5 \leq Y \leq 18.5) = 0.0225$ (calculated with computer). So a stormtrooper would make 18 out of 100 successive shots approximately 2.3% of the time.

(b) Use a normal distribution to approximate the probability that the stormtrooper makes between 8 and 12 shots. Use the continuity correction factor.

Let's again sketch out what's going on.



Now we want $P(7.5 \leq Y \leq 12.5) = 0.4781$. So a stormtrooper would make between 8 and 12 shots out of 100 successive shots about 4.8% of the time.