Details: The Sampling Distribution of the Sample Proportion

We might not have time to discuss this in class…or maybe we rushed through it too quickly for you to follow! Here are the details about the sampling distribution of the sample proportion.

First of all, we need to get our heads straight about where proportions come from. Let's say that we're looking at blue M-M's in a bag. How do you determine the proportion of blue M-M's?

That's right—count the number of blue, and divide by the number in the bag. Now, really think about that…*count* the number of blue. There are a *fixed number* of M-M's in the bag. Each M-M in the bag is either *blue*, or *not blue*. Does any of this sound familiar?

When we count the number of blue, we're working with a **binomial random variable**. This variable is divided by the sample size to obtain the sample proportion (another random variable).

In symbols: $\hat{p} = \frac{x}{n}$.

Measuring the Center

Recall that $\mu_{kX} = k\mu_X$. Thus, $\mu_{\hat{p}} = \mu_{\frac{x}{n}} = \frac{1}{n}\mu_x$. That x is binomial—what's the mean of a binomial random variable?

Yes! $\mu_x = n \cdot p$. So $\mu_{\hat{p}} = \frac{1}{n}\mu_x = \frac{1}{n}(n \cdot p) = p$.

Measuring the Spread

Recall that $\sigma^2_{kX} = k^2\sigma^2_X$. Thus, $\sigma^2_{\hat{p}} = \sigma^2_{\frac{x}{n}} = \frac{1}{n^2}\sigma^2_x$. The variance of a binomial random variable is $\sigma^2_x = n \cdot p(1-p)$. That makes $\sigma^2_{\hat{p}} = \frac{1}{n^2}\sigma^2_x = \frac{1}{n^2}\left(n \cdot p(1-p)\right) = \frac{p(1-p)}{n}$.

This is the variance; we really want standard deviation, so take the square root!

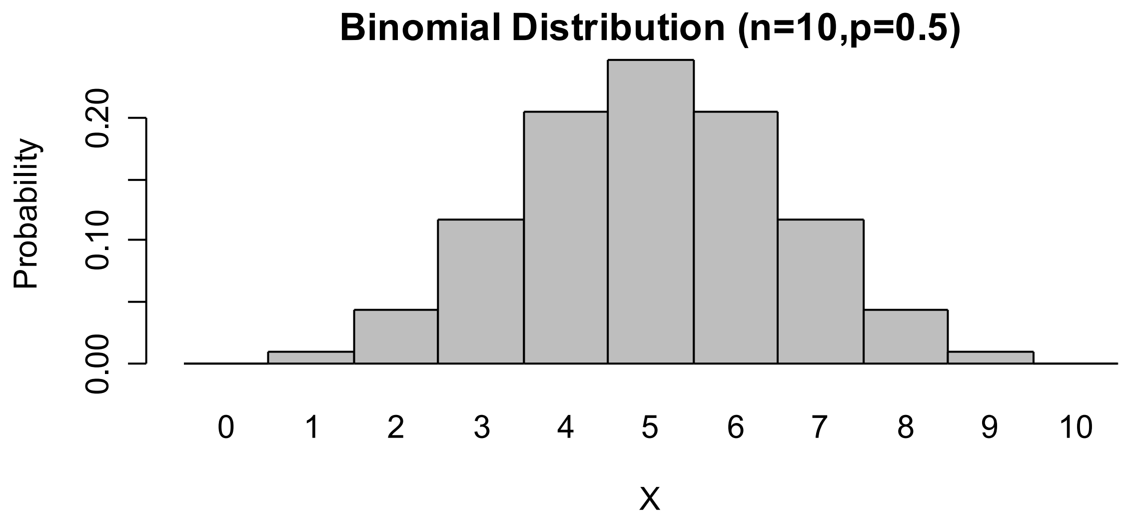
$$\sigma^2_{\hat{p}} = \frac{p(1-p)}{n} \Rightarrow \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Finding the Shape of the Sampling Distribution

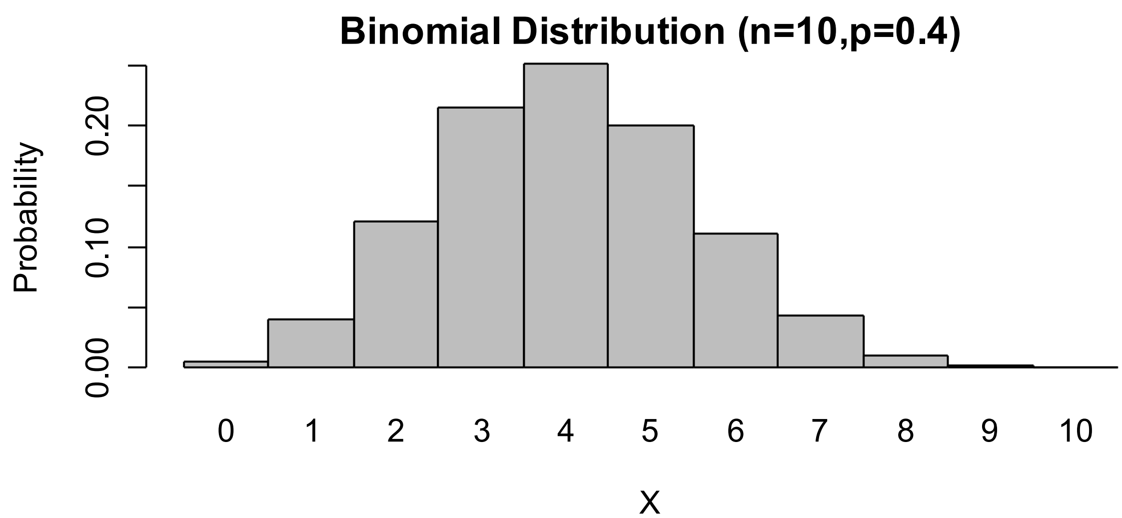
Let's look at the shape of the binominal distribution—dividing by n should have no effect on the overall shape, so we'll also get an idea of what the sampling distribution looks like.

The Effect of the Parameter

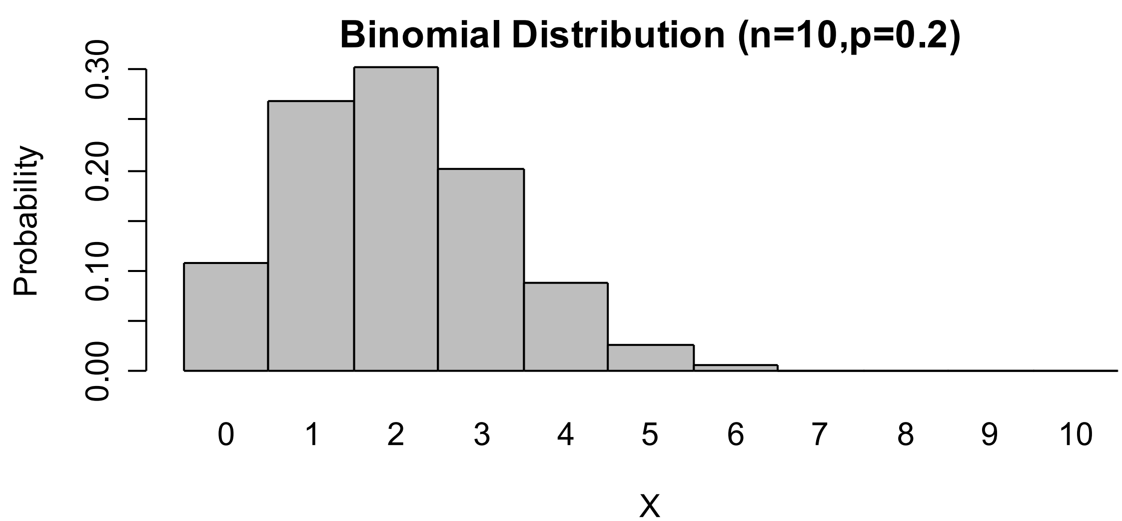
For these simulations, let's set the sample size at $n = 10$. First up: $p = 0.5$.



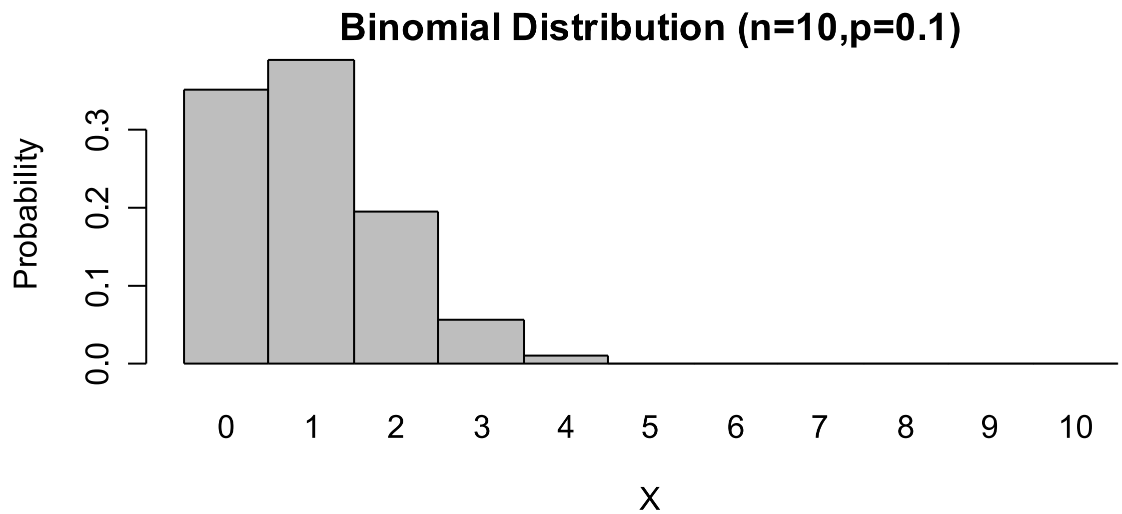
Now let $p = 0.4$.



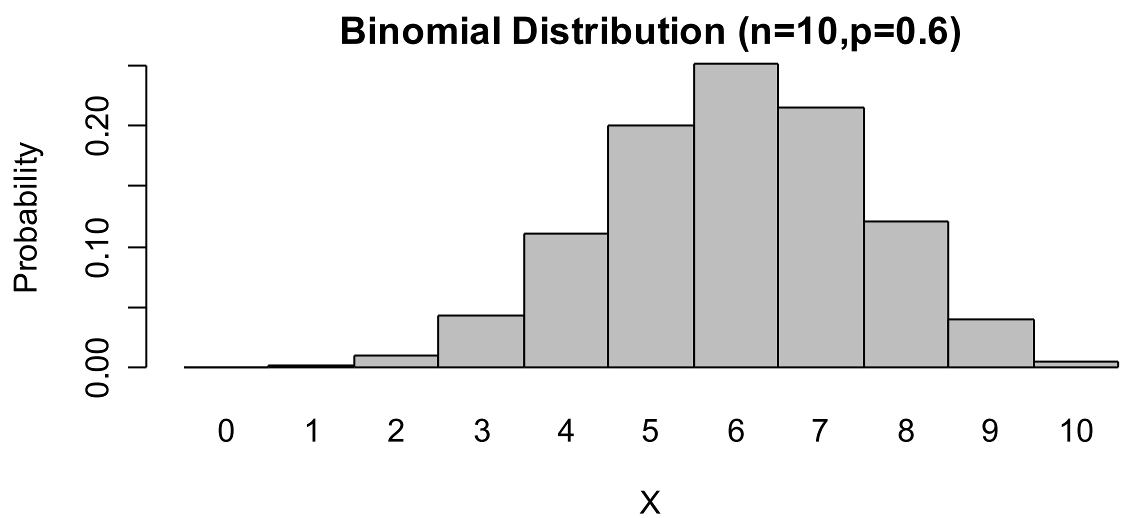
Let's try $p = 0.2$.



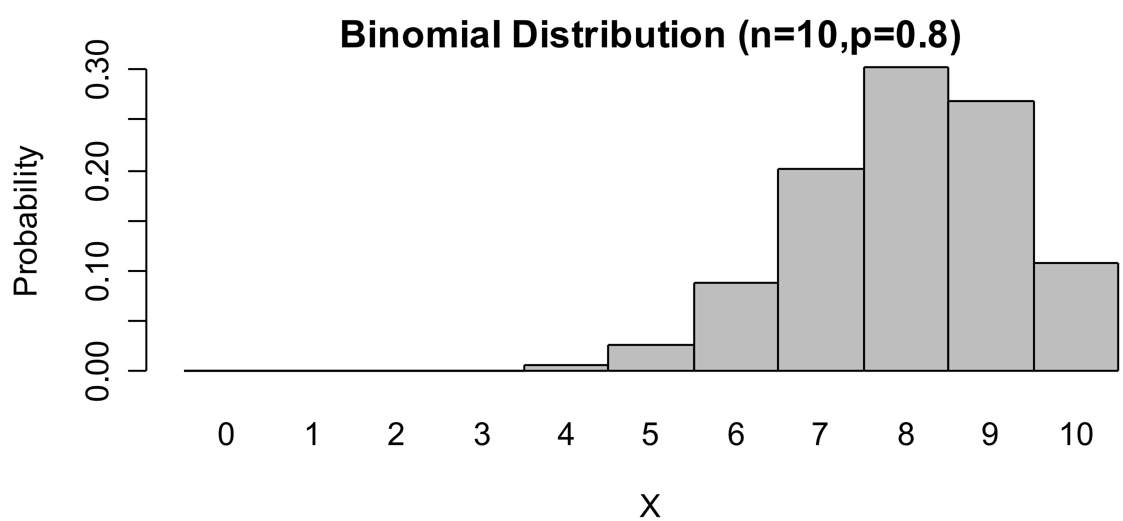
How about $p = 0.1$?



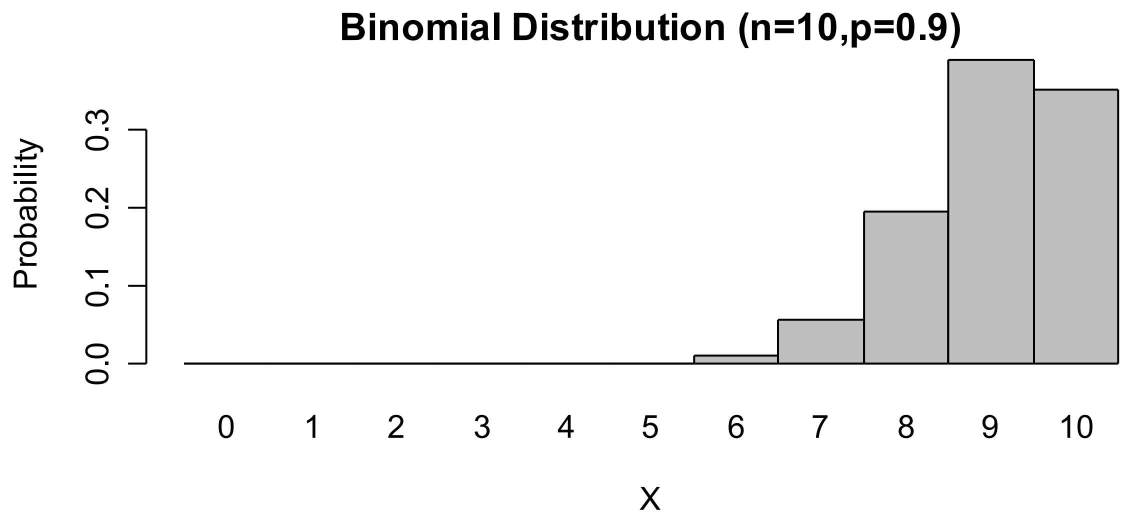
Let's go the other way: $p = 0.6$.



Now $p = 0.8$.



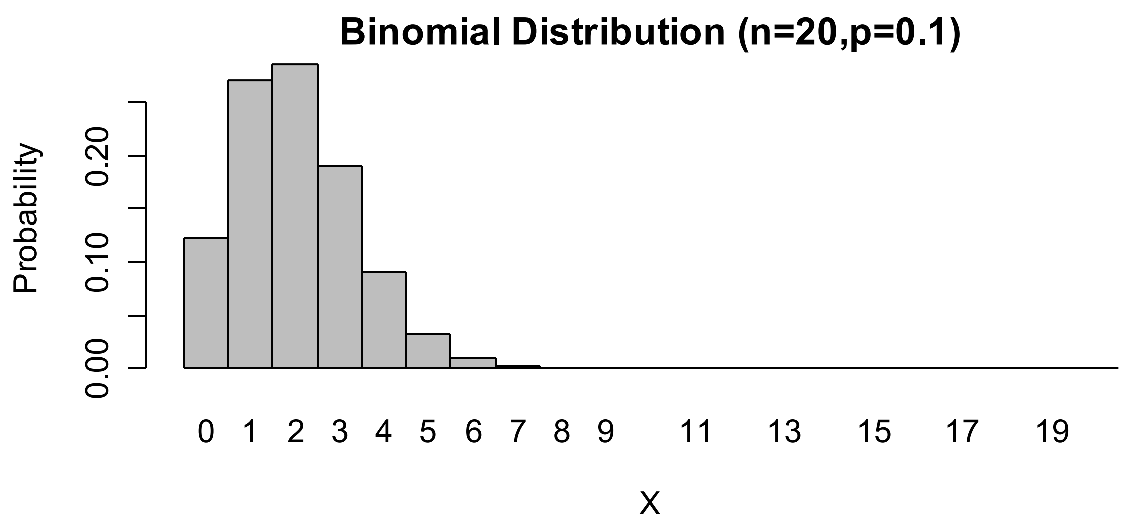
Finally, $p = 0.9$.



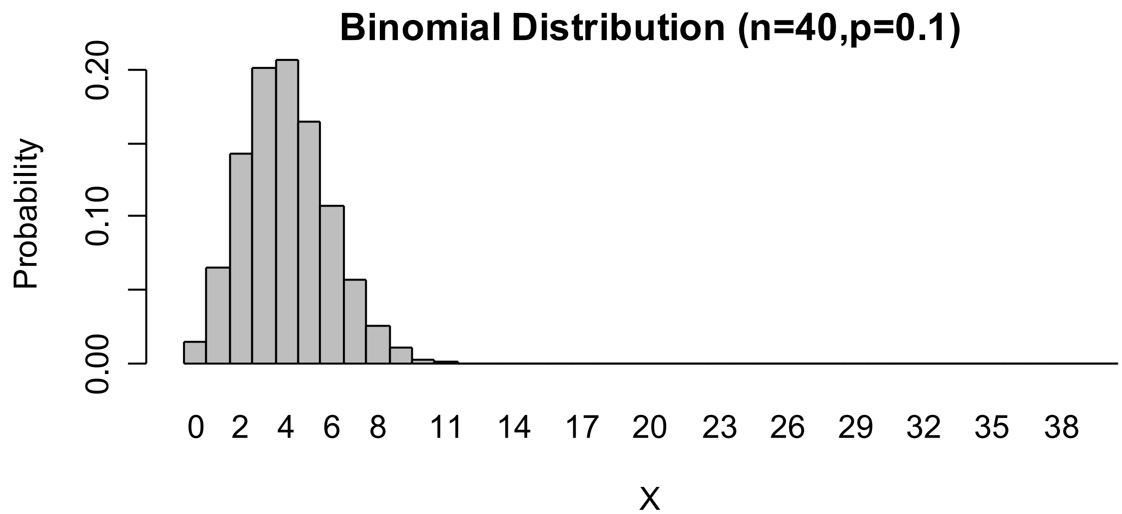
For a fixed sample size, the closer the parameter is to 0.5, the closer the binomial distribution (and thus, the sampling distribution of the sample proportion) will be to normal.

The Effect of Sample Size

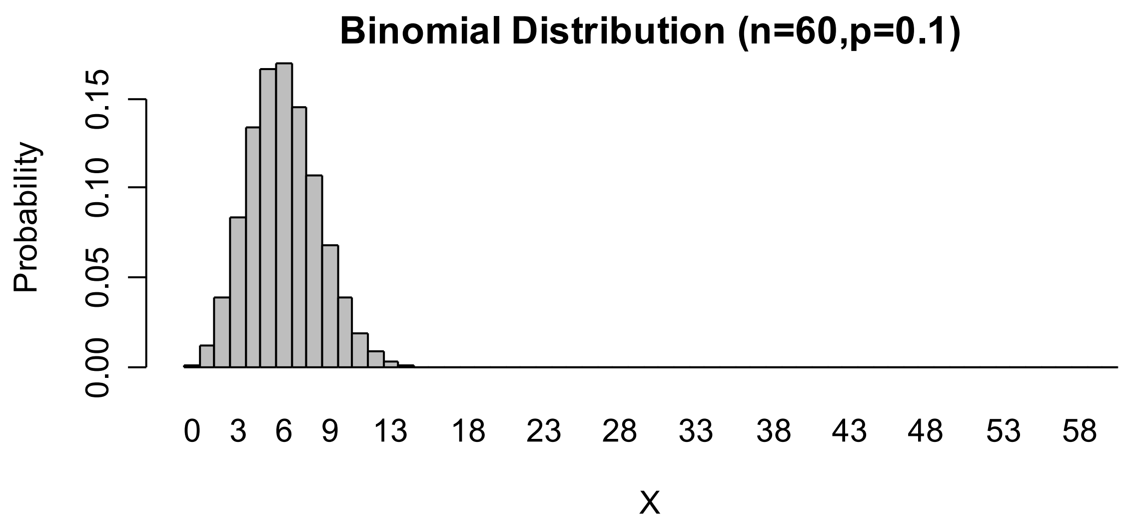
Let's set the parameter at a fairly extreme value—say, $p = 0.1$—and see what happens as the sample size increases. We'll start with 20.



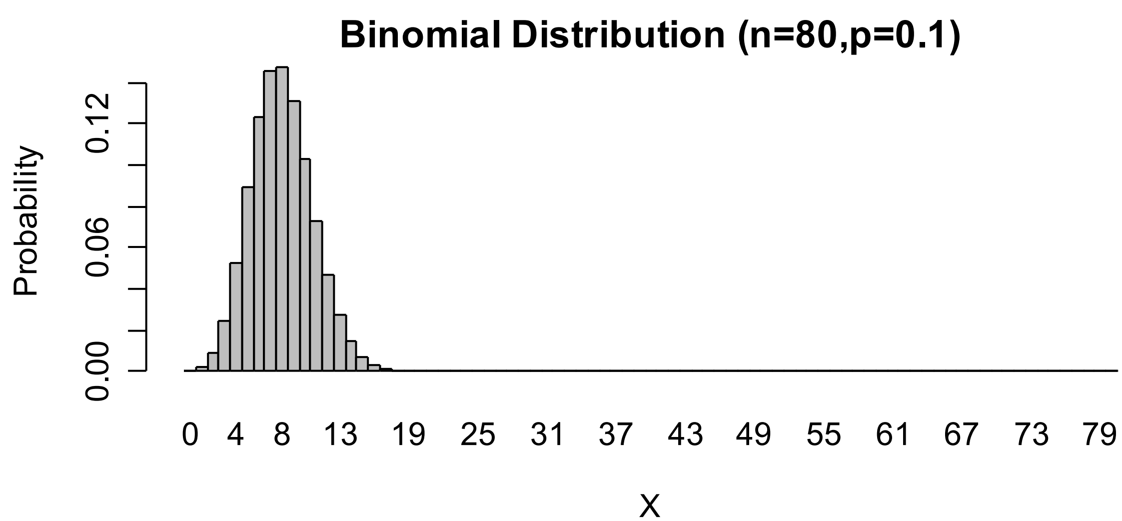
How about 40?



Let's look at $n = 60$.



One more—80.



The effect seems fairly clear—the larger the sample size, the closer the binomial (and the sampling) distribution gets to normal.