This tutorial is going to teach you about the normal distribution's approximation to the binomial distribution.

So let's do a little bit of review about the binomial distribution itself. If you want k number of successes in n trials, we're going to multiply n, choose k. n choose k gives the number of ways that we can achieve k successes in n trials on the tree diagram times k successes, represented by the probability of success to the k power, times n minus k, the rest, failures, represented by the probability of failure to the power of the number of failures that we want.

Using that formula, you can create a probability distribution for all the values of k, zero successes, one success, two successes, all the way up to n successes. And that, in turn, can be made into a histogram, where the x-axis are the values of k, the number of successes; and the y-axis is the relative frequency of those successes. So these buckets on zero go up to the height corresponding to the probability.

So, let's take a look. Just like all distributions, that histogram is going to have a mean and a standard deviation. The mean is pretty obvious to calculate, and we'll do it with an example real quick. Suppose you rolled a fair die six times. How many threes would you expect? What if you rolled it 60 times or 600 times? How many threes would you expect?

Hopefully, what you're thinking off the top of your head, is if you rolled it six times, you'd expect one of them to be a three. If you rolled it 60 times you'd expect about 10 threes. If you rolled it 600 times, you'd expect about 100 threes.

So, what exactly is it that you're doing? How did you come up with those so fast? You might be saying you divided by 6 because 1/6 of all the rolls will be three. Dividing by six, yes, that's great. What I'm going to submit that you were doing was you were actually multiplying by 1/6 because 1/6 was the probability. So 1/6 of six was one.

So, what we're seeing is the average, or the expected value, is going to be the number of trials times the probability of success. That's where we got 60 times 1/6 is the probability of a three gives you 10 as the expected value.

The standard deviation, we're not going to go through and derive the value, but it is fairly compact. All you do is take n times p, which actually that was the expected value, times q, which was the probability of failure. This is 1 minus p. And then take the square root.

And every distribution that we've dealt with, and every distribution that we're going to deal with, has three key features. By now, you should know that they're shape, center, and spread. Center and spread we just dealt with.

We found the mean and we found the standard deviation. But what about the shape? Well, the shape of this distribution is affected by two things. It's affected by both n and p.

Now let's take a look at this distribution here, where I did 10 trials and this was the probability of success. Notice when the probability of success is very high, the distribution is skewed very heavily to the left. When the probability of success is very low, the distribution becomes very much more skewed to the right. And when it's near 0.5, the probability of success, the distribution becomes very nearly symmetric. And so that's what we should see when we look at the binomial distribution.

All right, so how does n affect the shape of the distribution? Well, let's take a look. When we had 10 trials, and a probability of success of 0.4, it was fairly symmetric. And now look here. This is when there were a hundred trials, and it's still fairly symmetric. But let's look at what happens. Let's look at the comparison.

We decided that when the probability of success was very high, the shape would be skewed. But if you take a look, it's skewed here very heavily at 10 trials but at 100 trials it's nearly symmetric. It's a little skewed to the left, but not heavily skewed to the left.

How about when p is very low? Here this was heavily skewed to the right, whereas this distribution is only slightly skewed to the right. And if you look very closely, for the vast majority of values of p, the bottom distribution looks approximately normal. So that should be an interesting fact for you to find out is that when n is low, the skew, if any, is more prominent. And when n is high, the distribution is approximately normal. The only exceptions are when the value of p is very low or very high.

So this is a big deal. This means when that you have a large number of trials, the distribution of binomial probabilities is nearly normal, with the mean of what we found the mean to be, and standard deviation of what we found the standard deviation to be. Ultimately what we're finding, is the binomial distribution with parameters n and p, this is what makes the binomial look like what it looks like, looks a lot like the normal distribution with that mean and that standard deviation.

But it has to be large enough to satisfy two conditions. np, the mean, has to be at least 10. And nq, the expected number of failures, has to be at least 10. This means that we had to be far off of the left-hand side, far enough off the left-hand side, and far enough off the right-hand side. Remember when we had that distribution, it looked normal when we were safely in the middle of the distribution, and not near the very, very ends. So these two situations have to be satisfied, the two conditions do.

So this makes looking at a lot of these problems a whole lot easier. So suppose a baseball player gets a hit 28% of the time when he comes to bat. Probability that he gets over 30 hits in his next 95 at bats? Well, the old way, we

would have to find the probability that he gets exactly 31 hits, plus the probability that he gets exactly 32, all the way up to the probability that he gets exactly 95 hits. That's 65 individual calculations to do.

The new way, and this will give us an answer, but the new way uses the normal approximation. We'll use the mean of 26.6 and a standard deviation of 3.763 to use the normal distribution to find the answer. Both conditions are satisfied, np and nq are both bigger than 10, and so the normal distribution, or the binomial distribution, is going to look very much like this.

And so what we're going to do is we're going to use the normal distribution, calculate out a z-score, and find the probability that way. And this is almost the same as what we got using the binomial calculations.

So to recap, the normal distribution is a good approximation for the binomial under certain conditions. n has to be large, and p has to be not too extreme, not too high, not too low. We can use the mean and standard deviation of the binomial as the mean and standard deviation for the normal, and use z-scores to find the probabilities. This simplifies the problem by a whole, whole lot. So we talked about the normal approximation for the binomial distribution.

Good luck and we'll see you next time.