PROFESSOR: We're going to finish today our discussion of limit theorems. I'm going to remind you what the central limit theorem is, which we introduced briefly last time. We're going to discuss what exactly it says and its implications. And then we're going to apply to a couple of examples, mostly on the binomial distribution.

OK, so the situation is that we are dealing with a large number of independent, identically distributed random variables. And we want to look at the sum of them and say something about the distribution of the sum. We might want to say that the sum is distributed approximately as a normal random variable, although, formally, this is not quite right. As n goes to infinity, the distribution of the sum becomes very spread out, and it doesn't converge to a limiting distribution.

In order to get an interesting limit, we need first to take the sum and standardize it. By standardizing it, what we mean is to subtract the mean and then divide by the standard deviation. Now, the mean is, of course, n times the expected value of each one of the X's. And the standard deviation is the square root of the variance. The variance is n times sigma squared, where sigma is the variance of the X's -- so that's the standard deviation.

And after we do this, we obtain a random variable that has 0 mean -- its centered -- and the variance is equal to 1. And so the variance stays the same, no matter how large n is going to be.

So the distribution of Zn keeps changing with n, but it cannot change too much. It stays in place. The mean is 0, and the width remains also roughly the same because the variance is 1. The surprising thing is that, as n grows, that distribution of Zn kind of settles in a certain asymptotic shape. And that's the shape of a standard normal random variable. So standard normal means that it has 0 mean and unit variance.

More precisely, what the central limit theorem tells us is a relation between the cumulative distribution function of Zn and its relation to the cumulative distribution function of the standard

normal. So for any given number, c, the probability that Zn is less than or equal to c, in the limit, becomes the same as the probability that the standard normal becomes less than or equal to c. And of course, this is useful because these probabilities are available from the normal tables, whereas the distribution of Zn might be a very complicated expression if you were to calculate it exactly.

So some comments about the central limit theorem. First thing is that it's quite amazing that it's universal. It doesn't matter what the distribution of the X's is. It can be any distribution whatsoever, as long as it has finite mean and finite variance. And when you go and do your approximations using the central limit theorem, the only thing that you need to know about the distribution of the X's are the mean and the variance. You need those in order to standardize Sn. I mean -- to subtract the mean and divide by the standard deviation -- you need to know the mean and the variance. But these are the only things that you need to know in order to apply it.

In addition, it's a very accurate computational shortcut. So the distribution of this Zn's, in principle, you can calculate it by convolution of the distribution of the X's with itself many, many times. But this is tedious, and if you try to do it analytically, it might be a very complicated expression. Whereas by just appealing to the standard normal table for the standard normal random variable, things are done in a very quick way. So it's a nice computational shortcut if you don't want to get an exact answer to a probability problem.

Now, at a more philosophical level, it justifies why we are really interested in normal random variables. Whenever you have a phenomenon which is noisy, and the noise that you observe is created by adding the lots of little pieces of randomness that are independent of each other, the overall effect that you're going to observe can be described by a normal random variable. So in a classic example that goes 100 years back or so, suppose that you have a fluid, and inside that fluid, there's a little particle of dust or whatever that's suspended in there. That little particle gets hit by molecules completely at random -- and so what you're going to see is that particle kind of moving randomly inside that liquid.

Now that random motion, if you ask, after one second, how much is my particle displaced, let's say, in the x-axis along the x direction. That displacement is very, very well modeled by a normal random variable. And the reason is that the position of that particle is decided by the cumulative effect of lots of random hits by molecules that hit that particle.

So that's a sort of celebrated physical model that goes under the name of Brownian motion. And it's the same model that some people use to describe the movement in the financial markets. The argument might go that the movement of prices has to do with lots of little decisions and lots of little events by many, many different actors that are involved in the market. So the distribution of stock prices might be well described by normal random variables. At least that's what people wanted to believe until somewhat recently.

Now, the evidence is that, actually, these distributions are a little more heavy-tailed in the sense that extreme events are a little more likely to occur that what normal random variables would seem to indicate. But as a first model, again, it could be a plausible argument to have, at least as a starting model, one that involves normal random variables. So this is the philosophical side of things.

On the more accurate, mathematical side, it's important to appreciate exactly quite kind of statement the central limit theorem is. It's a statement about the convergence of the CDF of these standardized random variables to the CDF of a normal. So it's a statement about convergence of CDFs. It's not a statement about convergence of PMFs, or convergence of PDFs.

Now, if one makes additional mathematical assumptions, there are variations of the central limit theorem that talk about PDFs and PMFs. But in general, that's not necessarily the case. And I'm going to illustrate this with-- I have a plot here which is not in your slides. But just to make the point, consider two different discrete distributions.

This discrete distribution takes values 1, 4, 7. This discrete distribution can take values 1, 2, 4, 6, and 7. So this one has sort of a periodicity of 3, this one, the range of values is a little more interesting. The numbers in these two distributions are cooked up so that they have the same mean and the same variance.

Now, what I'm going to do is to take eight independent copies of the random variable and plot the PMF of the sum of eight random variables. Now, if I plot the PMF of the sum of 8 of these, I get the plot, which corresponds to these bullets in this diagram. If I take 8 random variables, according to this distribution, and add them up and compute their PMF, the PMF I get is the one denoted here by the X's. The two PMFs look really different, at least, when you eyeball them.

On the other hand, if you were to plot the CDFs of them, then the CDFs, if you compare them

with the normal CDF, which is this continuous curve, the CDF, of course, it goes up in steps because we're looking at discrete random variables. But it's very close to the normal CDF. And if we, instead of n equal to 8, we were to take 16, then the coincidence would be even better.

So in terms of CDFs, when we add 8 or 16 of these, we get very close to the normal CDF. We would get essentially the same picture if I were to take 8 or 16 of these. So the CDFs sit, essentially, on top of each other, although the two PMFs look quite different. So this is to appreciate that, formally speaking, we only have a statement about CDFs, not about PMFs.

Now in practice, how do you use the central limit theorem? Well, it tells us that we can calculate probabilities by treating Zn as if it were a standard normal random variable. Now Zn is a linear function of Sn. Conversely, Sn is a linear function of Zn. Linear functions of normals are normal. So if I pretend that Zn is normal, it's essentially the same as if we pretend that Sn is normal. And so we can calculate probabilities that have to do with Sn as if Sn were normal. Now, the central limit theorem does not tell us that Sn is approximately normal. The formal statement is about Zn, but, practically speaking, when you use the result, you can just pretend that Sn is normal.

Finally, it's a limit theorem, so it tells us about what happens when n goes to infinity. If we are to use it in practice, of course, n is not going to be infinity. Maybe n is equal to 15. Can we use a limit theorem when n is a small number, as small as 15?

Well, it turns out that it's a very good approximation. Even for quite small values of n, it gives us very accurate answers. So n over the order of 15, or 20, or so give us very good results in practice. There are no good theorems that will give us hard guarantees because the quality of the approximation does depend on the details of the distribution of the X's. If the X's have a distribution that, from the outset, looks a little bit like the normal, then for small values of n, you are going to see, essentially, a normal distribution for the sum. If the distribution of the X's is very different from the normal, it's going to take a larger value of n for the central limit theorem to take effect.

So let's illustrates this with a few representative plots. So here, we're starting with a discrete uniform distribution that goes from 1 to 8. Let's add 2 of these random variables, 2 random variables with this PMF, and find the PMF of the sum. This is a convolution of 2 discrete uniforms, and I believe you have seen this exercise before. When you convolve this with itself, you get a triangle. So this is the PMF for the sum of two discrete uniforms.

Now let's continue. Let's convolve this with itself. These was going to give us the PMF of a sum of 4 discrete uniforms. And we get this, which starts looking like a normal. If we go to n equal to 32, then it looks, essentially, exactly like a normal. And it's an excellent approximation. So this is the PMF of the sum of 32 discrete random variables with this uniform distribution.

If we start with a PMF which is not symmetric-- this one is symmetric around the mean. But if we start with a PMF which is non-symmetric, so this is, here, is a truncated geometric PMF, then things do not work out as nicely when I add 8 of these. That is, if I convolve this with itself 8 times, I get this PMF, which maybe resembles a little bit to the normal one.

But you can really tell that it's different from the normal if you focus at the details here and there. Here it sort of rises sharply. Here it tails off a bit slower. So there's an asymmetry here that's present, and which is a consequence of the asymmetry of the distribution we started with. If we go to 16, it looks a little better, but still you can see the asymmetry between this tail and that tail.

If you get to 32 there's still a little bit of asymmetry, but at least now it starts looking like a normal distribution. So the moral from these plots is that it might vary, a little bit, what kind of values of n you need before you get the really good approximation. But for values of n in the range 20 to 30 or so, usually you expect to get a pretty good approximation. At least that's what the visual inspection of these graphs tells us.

So now that we know that we have a good approximation in our hands, let's use it. Let's use it by revisiting an example from last time. This is the polling problem. We're interested in the fraction of population that has a certain habit been. And we try to find what f is. And the way we do it is by polling people at random and recording the answers that they give, whether they have the habit or not. So for each person, we get the Bernoulli random variable. With probability f, a person is going to respond 1, or yes, so this is with probability f. And with the remaining probability 1-f, the person responds no.

We record this number, which is how many people answered yes, divided by the total number of people. That's the fraction of the population that we asked. This is the fraction inside our sample that answered yes. And as we discussed last time, you might start with some specs for the poll. And the specs have two parameters-- the accuracy that you want and the confidence that you want to have that you did really obtain the desired accuracy. So the specs here is that we want, probability 95% that our estimate is within 1 % point from the true answer.

So the event of interest is this. That's the result of the poll minus distance from the true answer is less or bigger than 1 % point. And we're interested in calculating or approximating this particular probability.

So we want to do it using the central limit theorem. And one way of arranging the mechanics of this calculation is to take the event of interest and massage it by subtracting and dividing things from both sides of this inequality so that you bring him to the picture the standardized random variable, the Zn, and then apply the central limit theorem.

So the event of interest, let me write it in full, Mn is this quantity, so I'm putting it here, minus f, which is the same as nf divided by n. So this is the same as that event. We're going to calculate the probability of this. This is not exactly in the form in which we apply the central limit theorem. To apply the central limit theorem, we need, down here, to have sigma square root n.

So how can I put sigma square root n here? I can divide both sides of this inequality by sigma. And then I can take a factor of square root n from here and send it to the other side.

So this event is the same as that event. This will happen if and only if that will happen. So calculating the probability of this event here is the same as calculating the probability that this events happens.

And now we are in business because the random variable that we got in here is Zn, or the absolute value of Zn, and we're talking about the probability that Zn, absolute value of Zn, is bigger than a certain number. Since Zn is to be approximated by a standard normal random variable, our approximation is going to be, instead of asking for Zn being bigger than this number, we will ask for Z, absolute value of Z, being bigger than this number.

So this is the probability that we want to calculate. And now Z is a standard normal random variable. There's a small difficulty, the one that we also encountered last time. And the difficulty is that the standard deviation, sigma, of the Xi's is not known. Sigma is equal to f times-- sigma, in this example, is f times (1-f), and the only thing that we know about sigma is that it's going to be a number less than 1/2.

OK, so we're going to have to use an inequality here. We're going to use a conservative value of sigma, the value of sigma equal to 1/2 and use that instead of the exact value of sigma. And this gives us an inequality going this way.

Let's just make sure why the inequality goes this way. We got, on our axis, two numbers. One number is 0.01 square root n divided by sigma. And the other number is 0.02 square root of n. And my claim is that the numbers are related to each other in this particular way.

Why is this? Sigma is less than 2. So 1/sigma is bigger than 2. So since 1/sigma is bigger than 2 this means that this numbers sits to the right of that number. So here we have the probability that Z is bigger than this number. The probability of falling out there is less than the probability of falling in this interval.

So that's what that last inequality is saying-- this probability is smaller than that probability. This is the probability that we're interested in, but since we don't know sigma, we take the conservative value, and we use an upper bound in terms of the probability of this interval here.

And now we are in business. We can start using our normal tables to calculate probabilities of interest. So for example, let's say that's we take n to be 10,000. How is the calculation going to go? We want to calculate the probability that the absolute value of Z is bigger than 0.2 times 1000, which is the probability that the absolute value of Z is larger than or equal to 2.

And here let's do some mechanics, just to stay in shape. The probability that you're larger than or equal to 2 in absolute value, since the normal is symmetric around the mean, this is going to be twice the probability that Z is larger than or equal to 2.

Can we use the cumulative distribution function of Z to calculate this? Well, almost the cumulative gives us probabilities of being less than something, not bigger than something. So we need one more step and write this as 1 minus the probability that Z is less than or equal to 2.

And this probability, now, you can read off from the normal tables. And the normal tables will tell you that this probability is 0.9772. And you do get an answer. And the answer is 0.0456. OK, so we tried 10,000. And we find that our probably of error is 4.5%, so we're doing better than the spec that we had. So this tells us that maybe we have some leeway. Maybe we can use a smaller sample size and still stay without our specs.

Let's try to find how much we can push the envelope. How much smaller can we take n? To answer that question, we need to do this kind of calculation, essentially, going backwards. We're going to fix this number to be 0.05 and work backwards here to find-- did I do a mistake

here? 10,000. So I'm missing a 0 here. Ah, but I'm taking the square root, so it's 100.

Where did the 0.02 come in from? Ah, from here. OK, all right. 0.02 times 100, that gives us 2. OK, all right. Very good, OK. So we'll have to do this calculation now backwards, figure out if this is 0.05, what kind of number we're going to need here and then here, and from this we will be able to tell what value of n do we need.

OK, so we want to find n such that the probability that Z is bigger than 0.02 square root n is 0.05. OK, so Z is a standard normal random variable. And we want the probability that we are outside this range. We want the probability of those two tails together. Those two tails together should have probability of 0.05. This means that this tail, by itself, should have probability 0.025. And this means that this probability should be 0.975.

Now, if this probability is to be 0.975, what should that number be? You go to the normal tables, and you find which is the entry that corresponds to that number. I actually brought a normal table with me. And 0.975 is down here. And it tells you that to the number that corresponds to it is 1.96.

So this tells us that this number should be equal to 1.96. And now, from here, you do the calculations. And you find that n is 9604. So with a sample of 10,000, we got probability of error 4.5%. With a slightly smaller sample size of 9,600, we can get the probability of a mistake to be 0.05, which was exactly our spec.

So these are essentially the two ways that you're going to be using the central limit theorem. Either you're given n and you try to calculate probabilities. Or you're given the probabilities, and you want to work backwards to find n itself.

So in this example, the random variable that we dealt with was, of course, a binomial random variable. The X_i's were Bernoulli, so the sum of the X_i's were binomial. So the central limit theorem certainly applies to the binomial distribution. To be more precise, of course, it applies to the standardized version of the binomial random variable.

So here's what we did, essentially, in the previous example. We fixed the number p, which is the probability of success in our experiments. p corresponds to f in the previous example. Let every X_i a Bernoulli random variable and are standing assumption is that these random variables are independent. When we add them, we get a random variable that has a binomial distribution. We know the mean and the variance of the binomial, so we take S_n, we subtract

the mean, which is this, divide by the standard deviation. The central limit theorem tells us that the cumulative distribution function of this random variable is a standard normal random variable in the limit.

So let's do one more example of a calculation. Let's take n to be-- let's choose some specific numbers to work with. So in this example, first thing to do is to find the expected value of S_n, which is n times p. It's 18.

Then we need to write down the standard deviation. The variance of S_n is the sum of the variances. It's np times (1-p). And in this particular example, p times (1-p) is 1/4, n is 36, so this is 9. And that tells us that the standard deviation of this n is equal to 3.

So what we're going to do is to take the event of interest, which is S_n less than 21, and rewrite it in a way that involves the standardized random variable. So to do that, we need to subtract the mean. So we write this as S_n-3 should be less than or equal to 21-3. This is the same event. And then divide by the standard deviation, which is 3, and we end up with this. So the event itself of--

AUDIENCE: [INAUDIBLE].

Should subtract, 18, yes, which gives me a much nicer number out here, which is 1. So the event of interest, that S_n is less than 21, is the same as the event that a standard normal random variable is less than or equal to 1. And once more, you can look this up at the normal tables. And you find that the answer that you get is 0.43.

Now it's interesting to compare this answer that we got through the central limit theorem with the exact answer. The exact answer involves the exact binomial distribution. What we have here is the binomial probability that, S_n is equal to k. S_n being equal to k is given by this formula. And we add, over all values for k going from 0 up to 21, we write a two lines code to calculate this sum, and we get the exact answer, which is 0.8785. So there's a pretty good agreements between the two, although you wouldn't call that's necessarily excellent agreement.

Can we do a little better than that? OK. It turns out that we can. And here's the idea. So our random variable S_n has a mean of 18. It has a binomial distribution. It's described by a PMF that has a shape roughly like this and which keeps going on.

Using the central limit theorem is basically pretending that S_n is normal with the right mean

and variance. So pretending that Zn has 0 mean unit variance, we approximate it with Z, that has 0 mean unit variance. If you were to pretend that Sn is normal, you would approximate it with a normal that has the correct mean and correct variance. So it would still be centered at 18. And it would have the same variance as the binomial PMF.

So using the central limit theorem essentially means that we keep the mean and the variance what they are but we pretend that our distribution is normal. We want to calculate the probability that Sn is less than or equal to 21. I pretend that my random variable is normal, so I draw a line here and I calculate the area under the normal curve going up to 21. That's essentially what we did.

Now, a smart person comes around and says, Sn is a discrete random variable. So the event that Sn is less than or equal to 21 is the same as Sn being strictly less than 22 because nothing in between can happen. So I'm going to use the central limit theorem approximation by pretending again that Sn is normal and finding the probability of this event while pretending that Sn is normal. So what this person would do would be to draw a line here, at 22, and calculate the area under the normal curve all the way to 22.

Who is right? Which one is better? Well neither, but we can do better than both if we sort of split the difference. So another way of writing the same event for Sn is to write it as Sn being less than 21.5. In terms of the discrete random variable Sn, all three of these are exactly the same event. But when you do the continuous approximation, they give you different probabilities. It's a matter of whether you integrate the area under the normal curve up to here, up to the midway point, or up to 22. It turns out that integrating up to the midpoint is what gives us the better numerical results. So we take here 21 and 1/2, and we integrate the area under the normal curve up to here.

So let's do this calculation and see what we get. What would we change here? Instead of 21, we would now write 21 and 1/2. This 18 becomes, no, that 18 stays what it is. But this 21 becomes 21 and 1/2. And so this one becomes 1 + 0.5 by 3. This is 117.

So we now look up into the normal tables and ask for the probability that Z is less than 1.17. So this here gets approximated by the probability that the standard normal is less than 1.17. And the normal tables will tell us this is 0.879.

Going back to the previous slide, what we got this time with this improved approximation is

0.879. This is a really good approximation of the correct number. This is what we got using the 21. This is what we get using the 21 and 1/2. And it's an approximation that's sort of right on-- a very good one.

The moral from this numerical example is that doing this 1 and 1/2 correction does give us better approximations. In fact, we can use this 1/2 idea to even calculate individual probabilities. So suppose you want to approximate the probability that Sn equal to 19. If you were to pretend that Sn is normal and calculate this probability, the probability that the normal random variable is equal to 19 is 0. So you don't get an interesting answer.

You get a more interesting answer by writing this event, 19 as being the same as the event of falling between 18 and 1/2 and 19 and 1/2 and using the normal approximation to calculate this probability. In terms of our previous picture, this corresponds to the following.

We are interested in the probability that Sn is equal to 19. So we're interested in the height of this bar. We're going to consider the area under the normal curve going from here to here, and use this area as an approximation for the height of that particular bar.

So what we're basically doing is, we take the probability under the normal curve that's assigned over a continuum of values and attributed it to different discrete values. Whatever is above the midpoint gets attributed to 19. Whatever is below that midpoint gets attributed to 18. So this is green area is our approximation of the value of the PMF at 19.

So similarly, if you wanted to approximate the value of the PMF at this point, you would take this interval and integrate the area under the normal curve over that interval. It turns out that this gives a very good approximation of the PMF of the binomial. And actually, this was the context in which the central limit theorem was proved in the first place, when this business started.

So this business goes back a few hundred years. And the central limit theorem was first approved by considering the PMF of a binomial random variable when p is equal to 1/2. People did the algebra, and they found out that the exact expression for the PMF is quite well approximated by that expression hat you would get from a normal distribution. Then the proof was extended to binomials for more general values of p.

So here we talk about this as a refinement of the general central limit theorem, but, historically, that refinement was where the whole business got started in the first place. All right, so let's go

through the mechanics of approximating the probability that Sn is equal to 19-- exactly 19. As we said, we're going to write this event as an event that covers an interval of unit length from 18 and 1/2 to 19 and 1/2. This is the event of interest.

First step is to massage the event of interest so that it involves our Zn random variable. So subtract 18 from all sides. Divide by the standard deviation of 3 from all sides. That's the equivalent representation of the event. This is our standardized random variable Zn. These are just these numbers.

And to do an approximation, we want to find the probability of this event, but Zn is approximately normal, so we plug in here the Z, which is the standard normal. So we want to find the probability that the standard normal falls inside this interval. You find these using CDFs because this is the probability that you're less than this but not less than that. So it's a difference between two cumulative probabilities.

Then, you look up your normal tables. You find two numbers for these quantities, and, finally, you get a numerical answer for an individual entry of the PMF of the binomial. This is a pretty good approximation, it turns out. If you were to do the calculations using the exact formula, you would get something which is pretty close-- an error in the third digit-- this is pretty good.

So I guess what we did here with our discussion of the binomial slightly contradicts what I said before-- that the central limit theorem is a statement about cumulative distribution functions. In general, it doesn't tell you what to do to approximate PMFs themselves. And that's indeed the case in general. One the other hand, for the special case of a binomial distribution, the central limit theorem approximation, with this 1/2 correction, is a very good approximation even for the individual PMF.

All right, so we spent quite a bit of time on mechanics. So let's spend the last few minutes today thinking a bit and look at a small puzzle. So the puzzle is the following. Consider Poisson process that runs over a unit interval. And where the arrival rate is equal to 1. So this is the unit interval. And let X be the number of arrivals. And this is Poisson, with mean 1.

Now, let me take this interval and divide it into n little pieces. So each piece has length 1/n. And let Xi be the number of arrivals during the Ith little interval.

OK, what do we know about the random variables Xi? Is they are themselves Poisson. It's a number of arrivals during a small interval. We also know that when n is big, so the length of

the interval is small, these Xi's are approximately Bernoulli, with mean 1/n.

Guess it doesn't matter whether we model them as Bernoulli or not. What matters is that the Xi's are independent. Why are they independent? Because, in a Poisson process, these joint intervals are independent of each other.

So the Xi's are independent. And they also have the same distribution. And we have that X, the total number of arrivals, is the sum over the Xn's. So the central limit theorem tells us that, approximately, the sum of independent, identically distributed random variables, when we have lots of these random variables, behaves like a normal random variable. So by using this decomposition of X into a sum of i.i.d random variables, and by using values of n that are bigger and bigger, by taking the limit, it should follow that X has a normal distribution.

On the other hand, we know that X has a Poisson distribution. So something must be wrong in this argument here. Can we really use the central limit theorem in this situation?

So what do we need for the central limit theorem? We need to have independent, identically distributed random variables. We have it here. We want them to have a finite mean and finite variance. We also have it here, means variances are finite.

What is another assumption that was never made explicit, but essentially was there? Or in other words, what is the flaw in this argument that uses the central limit theorem here? Any thoughts?

So in the central limit theorem, we said, consider-- fix a probability distribution, and let the Xi's be distributed according to that probability distribution, and add a larger and larger number or Xi's. But the underlying, unstated assumption is that we fix the distribution of the Xi's. As we let n increase, the statistics of each Xi do not change.

Whereas here, I'm playing a trick on you. As I'm taking more and more random variables, I'm actually changing what those random variables are. When I take a larger n, the Xi's are random variables with a different mean and different variance. So I'm adding more of these, but at the same time, in this example, I'm changing their distributions.

That's something that doesn't fit the setting of the central limit theorem. In the central limit theorem, you first fix the distribution of the X's. You keep it fixed, and then you consider adding more and more according to that particular fixed distribution. So that's the catch. That's why the central limit theorem does not apply to this situation. And we're lucky that it doesn't apply

because, otherwise, we would have a huge contradiction destroying probability theory.

OK, but now that's still leaves us with a little bit of a dilemma. Suppose that, here, essentially we're adding independent Bernoulli random variables. So the issue is that the central limit theorem has to do with asymptotics as n goes to infinity. And if we consider a binomial, and somebody gives us specific numbers about the parameters of that binomial, it might not necessarily be obvious what kind of approximation do we use.

In particular, we do have two different approximations for the binomial. If we fix p, then the binomial is the sum of Bernoulli's that come from a fixed distribution, we consider more and more of these. When we add them, the central limit theorem tells us that we get the normal distribution.

There's another sort of limit, which has the flavor of this example, in which we still deal with a binomial, sum of n Bernoulli's. We let that sum, the number of the Bernoulli's go to infinity. But each Bernoulli has a probability of success that goes to 0, and we do this in a way so that np, the expected number of successes, stays finite.

This is the situation that we dealt with when we first defined our Poisson process. We have a very, very large number so lots, of time slots, but during each time slot, there's a tiny probability of obtaining an arrival. Under that setting, in discrete time, we have a binomial distribution, or Bernoulli process, but when we take the limit, we obtain the Poisson process and the Poisson approximation.

So these are two equally valid approximations of the binomial. But they're valid in different asymptotic regimes. In one regime, we fixed p, let n go to infinity. In the other regime, we let both n and p change simultaneously.

Now, in real life, you're never dealing with the limiting situations. You're dealing with actual numbers. So if somebody tells you that the numbers are like this, then you should probably say that this is the situation that fits the Poisson description-- large number of slots with each slot having a tiny probability of success.

On the other hand, if p is something like this, and n is 500, then you expect to get the distribution for the number of successes. It's going to have a mean of 50 and to have a fair amount of spread around there. It turns out that the normal approximation would be better in this context.

As a rule of thumb, if n times p is bigger than 10 or 20, you can start using the normal approximation. If n times p is a small number, then you prefer to use the Poisson approximation. But there's no hard theorems or rules about how to go about this. OK, so from next time we're going to switch base again. And we're going to put together everything we learned in this class to start solving inference problems.