The central limit theorem is absolutely remarkable. It is a very deep result, and highly nontrivial and non intuitive. There's no apparent reason why this random variable here, a standardized version of the sum of random variables, should have an approximately normal distribution.

Furthermore, it is very useful, and one key reason is that it is universal. It doesn't matter what the distribution of the X's is. No matter what the distribution is, still in the limit, this standardized version of the sum is going to behave like a normal random variable. And if we wish to apply it to particular examples or models, the only thing that we need to know about the distribution of the X's are the corresponding means and variances, as we're going to see in multiple examples.

When we apply it, it turns out to be very accurate, and it is also a very nice computational shortcut. Even if we knew, in detail, the distribution of the X's, in order to calculate the distribution of Sn, we would have to take the distribution of the X's and convolve it with itself n times, something that can be computationally tedious. Whereas the computations that are involved, when we use the central limit theorem, are very, very simple, as the examples that will be coming up will show to us.

Finally, at the philosophical level, the central limit theorem justifies why models involving normal random variables are very natural. Whenever you have a phenomenon or an object that's affected by multiple noise sources, and if these noise sources are independent, then the overall effect of those noise sources is going to be well-modeled by a normal random variable, even if the distribution of each one of these noise sources is very different from being normal. And this is a reason why, in many, many applications in many different fields, normal random variables provide very useful and accurate models.

Since the central limit theorem is so useful and important, it is worth making sure that we understand exactly what it says and to make a few comments on its mathematical content. What it says is the following. We take the sum of independent identically distributed random variables, we form this standardized version of the sum, where we subtracted the mean and divide by the standard deviation of the sum, and then what it tells us is that the CDF of this random variable, Zn, converges to the normal CDF.

So what we have is a statement about CDFs. It does not yet tell us anything specific about PDFs or PMFs. So for example, if Sn and the X's are all continuous random variable, so Zn is also continuous,

you might wonder whether the PDF of Zn converges to a normal PDF. It turns out that there are results of this kind that assert convergence of the PDF, or even PMF, of this random variable to a normal PDF, in some sense. But these results generally need a few more mathematical assumptions for the results to be valid. Nevertheless, when we show pictures of various examples, we will do this by showing pictures of PDFs and PMFs because these are easier to visualize.

Now since the result is so general and so important, it might be worth understanding to what extent it can be generalized to other contexts. Our main two assumptions are that the random variables are independent and identically distributed. Can we remove those assumptions? There are versions of the central limit theorem that apply to the case where the Xi's are not identically distributed. One just needs to make certain assumptions on the means and to the variances of the Xi's. Some conditions will be needed.

Also, the assumption of independence does not need to be literally true. There are versions of the central limit theorem that are valid when we have just weak dependence. That is, nearby X's may be dependent, but if you compare X5 with X of 1 million, then these two random variables are essentially independent. In those cases, we can still apply a suitable version of the central limit theorem.

And finally, you may be curious how this result is proved. One way of proving it, which is the way it was originally established a long time ago, for the special case of Bernoulli random variables X, in which case S is binomial. The way it was established was by carrying out algebraic manipulations on the binomial formulas. But this was a derivation that would not generalize. For the general case, the proof is obtained using so-called transform methods, which is a topic that we're not covering, but it goes as follows. We consider this function of the random variable Zn, where s is some parameter, and we show that this expectation converges to the corresponding expectation if you have the standard normal Z in the place of Zn. And this is true for all s, or at least for all s in some rich enough set.

And then, one appeals to some deep mathematical results that tell you that if this kind of expectation converges to that expectation, then the CDF of Zn must also converge to the CDF of Z. But this is a proof that involves various steps and appeals to some deep results from other fields of mathematics.

And finally, there is the practical side. What exactly does it say and how do we use it? Since the CDF of Zn can be approximated by the CDF of a standard normal, this means that in practice, we can treat the random variable Zn just as if it were a standard normal. But now, we notice that Sn is obtained as a

linear function of Zn. Namely, the definition of Zn gives us this formula.

So, Sn is a linear function of Zn. If we pretend that Zn is normal, and since linear functions of normal random variables are normal, this means that we will also pretend that Sn is normal. So in practice, the way to carry out calculations is often to just pretend that Sn is normal with the appropriate mean and variance. These are the correct means and variance of Sn, and that's the only information that we need in order to apply it. If we know mu and sigma squared, and using the normal approximation, then we have an approximate distribution for Sn, and we can go ahead.

Now in practice, can we use the central limit theorem when n is moderate? For example, if n is 30, can you apply the central limit theorem? This is a relation that's true in the limit of very large n. How large should n be? It turns out that the central limit theorem gives us very good approximations even when n has moderately small values. Now, these approximations sometimes will be better and sometimes worse. It helps if the distribution of the X's that you're starting with has some common one features with the normal distribution. If the X's are already normal, then S will be normal and there's no approximation involved. If the X's are close to normal, then for fairly small values of n, S will be very well modeled by a normal random variable.

Now, what does it mean that the distribution of the X's looks a little bit like the normal one? It helps if the distribution is symmetric around its mean, and it also helps if it is unimodal, in the sense that it has a single peak rather than multiple peaks.