PROFESSOR: So we've been talking about this chi square test. And the name chi square comes from the fact that we build a test statistic that has asymptotic distribution given by the chi square distribution. Let's just give it another shot. OK.

This test. Who has actually ever encountered the chi square test outside of a stats classroom? All right. So some people have. It's a fairly common test that you might encounter. And it was essentially to test, if given some data with a fixed probability mass function, so a discrete distribution, you wanted to test if the PMF was equal to a set value, p0, or if it was different from p0.

And the way the chi square arose here was by looking at Wald's test. And essentially if you write-- so Wald's is the one that has the chi square as the limiting distribution, and if you invert the covariance matrix, the asymptotic covariance matrix, so you compute the Fisher information, which in this particular case does not exist for the multinomial distribution, but we found the trick on how to do this. We remove the part that forbid it to be invertible, then we found this chi square distribution.

In a way we have this test statistic, which you might have learned as a black box, laundry list, but going through the math which might have been slightly unpleasant, I acknowledge, but really told you why you should do this particular normalization. So since some of you requested a little more practical examples of how those things work, let me show you a couple.

The first one is you want to answer the question, well, you know, when should I be born to be successful. Some people believe in zodiac, and so *Fortune* magazine actually collected the signs of 256 heads of the Fortune 500. Those were taken randomly. And they were collected there, and you can see the count of number of CEOs that have a particular zodiac sign.

And if this was completely uniformly distributed, you should actually get a number that's around 256 divided by 12, which in this case is 21.33. And you can see that there is numbers that are probably in the vicinity, but look at this guy. Pisces, that's 29. So who's Pisces here? All right. All right, so give me your information and we'll meet again in 10 years.

And so basically you might want to test if actually the fact that it's uniformly distributed is a valid assumption. Now this is clearly a random variable. I pick a random CEO and I measure

what his zodiac sign is. And I want to know, so it's a probability over, I don't know, 12 zodiac signs. And I want to know if it's uniform or not. Uniform sounds like it should be the status quo, if you're reasonable. And maybe there's actually something that moves away. So we could do this, in view of these data is there evidence that one is different.

Here is another example where you might want to apply the chi square test. So as I said, the benchmark distribution was the uniform distribution for the zodiac sign, and that's usually the one I give you. 1 over k, 1 over k, because well that's sort of the zero, the central point for all distributions. That's the point, the center of what we call the simplex.

But you can have another benchmark that sort of makes sense. So for example this is an actual dataset where 275 jurors were identified, racial group were collected, and you actually might want to know if you know juries in this country are actually representative of the actual population. And so here of course, the population is not uniformly distributed according to racial group.

And the way you actually do it is you actually go on Wikipedia, for example, and you look at the demographics of the United States, and you find that the proportion of white is 72%, black is 7%, Hispanic is 12, and other is about 9%. So that's a total of 1. And this is what we actually measured for some jurors. So for this guy, you can actually run the chi square test. You have the estimated proportion, which comes from this first line. You have the tested proportion, p0, that comes from the second line, and you might want to check if those things actually correspond to each other.

OK, so I'm not going to do it for you, but I sort of invite you to do it and test, and see how this compares to the quantiles of the appropriate chi square distribution and see what you can conclude from those two things.

All right. So this was the multinomial case. So this is essentially what we did. We computed the MLE under the right constraint, and that was our test statistic that converges to the chi square distribution. So if you've seen it before, that's all that was given to you. Now we know why the normalization here is p0 j and not p0 j squared or square root of p0 j, or even 1. I mean it's not clear that this should be the right normalization, but we know that's what comes from taking the right normalization, which comes from the Fisher information. All right? OK.

The thing I wanted to move onto, so we've basically covered chi square test. Are there any questions about chi square test? And for those of you who were not here on Thursday, I'm

really just-- do not pretend I just did it. That's something we did last Thursday. But are there any questions that arose when you were reading your notes, things that you didn't understand? Yes.

AUDIENCE: Is there like a formal name? Before we had talked about how what we call the Fisher information [INAUDIBLE], still has the same [INAUDIBLE] because it's the same number.

PROFESSOR: So it's not the Fisher. The Fisher information does not exist in this case. And so there's no appropriate name for this. It's the pseudoinverse of the asymptotic covariance matrix, and that's what it is. I don't know if I mentioned it last time, but there's this entire field that uses-- you know, for people who really aspire to differential geometry but are stuck in the stats department, and there's this thing called information geometry, which is essentially studying the manifolds associated to the Fisher information metric, the metric that's associated to Fisher information. And so those of course can be lower dimensional manifolds, not only distorts the geometry but forces everything to live on a lower dimension, which is what happens when your Fisher information does not exist. And so there's a bunch of things that you can study, what this manifold looks like, et cetera. But no, there's no particular terminology here about going here.

To be fair, within the scope of this class, this is the only case where you-- multinomial case is the only case where you typically see a lack of a Fisher information matrix. And that's just because we have these extra constraints that the sum of the parameters should be 1. And if you have an extra constraint that seems like it's actually remove one degree of freedom, this will happen inevitably.

And so maybe what you can do is reparameterize. So if I actually reparameterize everything function of p1 to p k minus 1, and then 1 minus the sum, this would not have happened. Because I have only a k-dimensional space. So there's tricks around this to make it exist if you want it to exist. Any other question?

All right. So let's move on to Student's t-test. We mentioned it last time. So essentially you've probably done it more even in the homework than you've done it in lectures, but just quickly this is essentially the test. That's the test when we have an actual data that comes from a normal distribution. There is no Central Limit Theorem that exists. This is really to account for the fact that for smaller sample sizes, it might be the case that it's not exactly true that when I look at xn bar minus mu divided by-- so if I look at xn bar minus mu divided by sigma times

square root of n, then this thing should have N 0, 1 distribution approximately. Right? By the Central Limit Theorem.

So that's for n large. But if n is small, then it's still true when the data is N mu, sigma squared, then it's true that square root of n-- so here it's approximately. And this is always true. But I don't know sigma in practice, right? Maybe mu, it comes from my, maybe mu comes from my mu 0, maybe something from the test statistic where mu actually is here. But for this guy I'm going to have inevitably to find an estimator.

And now in this case, for small n, this is no longer true. And what the t statistic is doing is essentially telling you what the distribution of this guy is. So what you should say is that now this guy has a t distribution with n minus 1 degrees of freedom.

That's basically the laundry list stats that you would learn. It says just look at a different table, that's what it is. But we actually defined what a t distribution was. And a t distribution is basically something that has the same distribution as some N 0, 1, divided by the square root of a chi square with d degrees of freedom divided by d. And that's a t distribution with d degrees of freedom. And those two have to be independent.

And so what I need to check is that this guy over there is of this form. OK? So let's look at the numerator. Well, square root of n, xn bar minus mu. What is the distribution of this thing? Is it an N 0, 1?

AUDIENCE: N 0, sigma squared?

PROFESSOR: N 0, sigma squared, right. So I'm not going to put it here. So if I want this guy to be N 0, 1, I need to divide by sigma, that's what we have over there. So that's my N 0, 1 that's going to play the role of this guy here.

So if I want to go a little further, I need to just say, OK, now I need to have square root of n, and I need to find something here that looks like my square root of chi square divided by-- yeah?

AUDIENCE: Really quick question. The equals sign with the d on top, that's just defined as?

PROFESSOR: No, that's just the distribution. So, I don't know.

AUDIENCE: Then never mind.

PROFESSOR: Let's just write it like that, if you want. I mean, that's not really appropriate to have. Usually you write only one distribution on the right-hand inside of this little thing. So not just this complicated function of distributions. This is more like to explain.

OK, and so usually the thing you should say that t is equal to this X divided by square root of Z divided by d where X has normal distribution, Z has chi square distribution with d degrees of freedom.

So what do we need here? Well I need to have something which looks like my sigma hat, right? So somehow inevitably I'm going to need to have sigma hat. Now of course I need to divide this by my sigma so that my sigma goes away. And so now this thing here-- sorry, I should move on to the right, OK. And so this thing here, so sigma hat is square root of Sn.

And now I'm almost there. So this thing is actually equal to square root of n. But this thing here is actually not a-- so this thing here follows a distribution which is actually a chi square, square root of a chi square distribution divided by n. Yeah, that's the square root chi square distribution with n minus 1 degrees of freedom divided by n, because sigma hat is equal to 1 over n sum from i equal 1 to n, xi minus x bar squared. And we just said that this part here was a chi square distribution. We didn't just say it, we said it a few lectures years back, that this thing was a chi square distribution, and the fact that the presence of this x bar here was actually removing one degree of freedom from this sum. OK, so this guy here has the same distribution as a chi square n minus 1 divided by n.

So I need to actually still arrange this thing a little bit to have a t distribution. I should not see n here, but I should n minus 1. The d is the same as this d here. And so let me make the correction so that this actually happens. Well, if I actually write this to be equal to-- so if I write square root of n minus 1, as on the slide, times xn bar minus mu divided by-- well let me write it as square root of Sn, which is my sigma hat. Then what this thing is actually equal to, it follows a N 0, 1, divided by the square root of my chi square distribution with n minus 1 degrees of freedom. And here the fact that I multiply by square root of n minus 1, and I have the square root of n here, is essentially the same as dividing here by n minus 1. And that's my tn distribution. My t distribution with n minus 1 degrees of freedom. Just by definition of what this thing is. OK?

All right. Yes?

AUDIENCE: Where'd you get the square root from?

PROFESSOR: This guy? Oh sorry, that's sigma squared. Thank you. That's the estimator of the variance, not the estimator of the standard deviation. And when I want to divide it I divide by standard deviation. Thank you. Any other question or remark?

AUDIENCE: Shouldn't you divide by sigma squared? The actual. The estimator for the variance is equal to sigma squared times chi square, right?

PROFESSOR: The estimator for the variance. Oh yes, you're right. So there's a sigma squared here. Is that what you're asking?

AUDIENCE: Yeah.

PROFESSOR: Yes, absolutely. And that's where, it get cancels here. It gets canceled here. OK? So this is really a sigma squared times chi square. OK. So the fact that it's sigma squared is just because I can pull out sigma squared and just think those guys N 0, 1.

All right. So that's my t distribution. Now that I actually have a pivotal distribution, what I do is that I form the statistic. Here I called it Tn tilde. OK. And what is this thing? I know that this has a pivotal distribution. So for example, I know that the probability that Tn tilde in absolute value exceeds some number that I'm going to call q alpha over 2 for the t n minus 1, is equal to alpha.

So that's basically, remember the t distribution has the same shape as the Gaussian distribution. What I'm finding is, for this t distribution, some number q alpha over 2 of t n minus 1 and minus q alpha over 2 of t minus 1. So those are different from the Gaussian one. Such that the area under the curve here is alpha over 2 on each side so that the probability that my absolute value exceeds this number is equal to alpha.

And that's what I'm going to use to reject the test. So now my test becomes, for H0, say mu is equal to some mu 0, versus H1, mu is not equal to mu 0. The rejection region is going to be equal to the set on which square root of n minus 1 times xn bar minus mu 0 this time, divided by square root of Sn exceeds, in absolute value, exceeds q-- sorry that's already here-- exceeds q alpha over 2 of t n minus 1.

So I reject when this thing increases. The same as the Gaussian case, except that rather than reading my quantiles from the Gaussian table I read them from the Student table. It's just the same thing. So they're just going to be a little bit farther. So this guy here is just going to be a

little bigger than the one for the Gaussian one, because it's going to require me a little more evidence in my data to be able to reject because I have to account for the fluctuations of sigma hat.

So of course Student's test is used everywhere. People use only t tests, right? If you look at any data point, any output, even if you had 500 observations, if you look at the statistical software output it's going to say t test. And the reason why you see t test is because somehow it's felt like it's not asymptotic. You don't need to actually do, you know, to be particularly careful. And anyway, if n is equal to 500, since the two curves are above each other it's basically the same thing. So it doesn't really change anything. So why not use the t test?

So it's not asymptotic. It doesn't require Central Limit Theorem to kick in. And so in particular it be run if you have 15 observations. Of course, the drawback of the Student test is that it relies on the assumption that the sample is Gaussian, and that's something we really need to keep in mind. If you have a small sample size, there is no magic going on. It's not like Student t test allows you to get rid of this asymptotic normality. It sort of assumes that it's built in. It assumes that your data has a Gaussian distribution.

So if you have 15 observations, what are you going to do? You want to test if the mean is equal to 0 or not equal to 0, but you have only 15 observations. You have to somehow assume that your data is Gaussian. But if the data is given to you, this is not math, you actually have to check that it's Gaussian.

And so we're going to have to find a test that, given some data, tells us whether it's Gaussian or not. If I have 15 observations, 8 of them are equal to plus 1 and 7 of them are equal to minus 1, then it's pretty unlikely that you're going to be able to conclude that your data has a Gaussian distribution. However, if you see some sort of spread around some value, you form a histogram maybe and it sort of looks like it's a Gaussian, you might want to say it's Gaussian.

And so how do we make this more quantitative? Well, the sad answer to this question is that there will be some tests that make it quantitative, but here, if you think about it for one second, what is going to be your null hypothesis? Your null hypothesis, since it's one point, it's going to be that it's Gaussian, and then the alternative is going to be that it's not Gaussian.

So what it means is that, for the first time in your statistician life, you're going to want to

conclude that H0 is the true one. You're definitely not going to want to say that it's not Gaussian, because then everything you know is sort of falling apart. And so it's kind of a weird thing where you're sort of going to be seeking tests that have no power basically. You're going to want to test that, and that's the nature. The amount of alternatives, the number of ways you can be not Gaussian, is so huge that all tests are sort of bound to have very low power. And so that's why people are pretty happy with the idea that things are Gaussian, because it's very hard to find a test that's going to reject this hypothesis.

And so we're even going to find some tests that are visual, where you're going to be able to say, well, sort of looks Gaussian to me. It allows you to deal with the borderline cases pretty efficiently. We'll see actually a particular example.

All right, so this theory of testing whether data comes from a particular distribution is called goodness of fit. Is this distribution a good fit for my data? That's the goodness of fit test. We have just seen a goodness of fit test. What was it? Yeah. The chi square test, right? The case square test, we were given a candidate PMF and we were testing if this was a good fit for our data. That was a goodness of fit test.

So of course multinomial is one example, but really what we have in the back of our mind is I want to test if my data is Gaussian. That's basically the usual thing. And just like you always see t test as the standard output from statistical software whether you ask for it or not, there will be a test for normality whether you ask it or not from any statistical software app.

All right. So a goodness of fit test looks as follows. There's a random variable X and you're given i.i.d. copies of X, X1 to Xn, they come from the same distribution. And you're going to ask the following question: does X have a standard normal distribution? So for t distribution that's definitely the kind of questions you may want to ask. Does X have a uniform distribution on 0, 1? That's different from the distribution 1 over k, 1 over k, it's the continuous notion of uniformity.

And for example, you might want to test that-- so there's actually a nice exercise, which is if you look at the p-values. So we've defined what the p-values were. And the p-value's a number between 0 and 1, right? And you could actually ask yourself, what is the distribution of the p-value under the null? So the p-value is a random number. It's the probability-- so the p-value-- let's look at the following test.

H0, mu is equal to 0, versus H1, mu is not equal to 0. And I know that the p-value is-- so I'm

going to form what? I'm going to look at Xn bar minus mu times square root of n divided by-- let's say that we know sigma for one second. Then the p-value is the probability that this is larger then square root of n little xn bar minus mu, minus 0 actually in this case, divided by sigma, where this guy is the observed.

OK. So now you could say, well, how is that a random variable? It's just a number. It's just a probability of something. But then I can view this as a function of this guy here when I plug it back to be a random variable. So what I mean by this is that if I look at this value here, if I say that phi is the CDF of N 0, 1, so the p-value is the probability that it exceeds this. So that's the probability that I'm either here or here.

AUDIENCE: [INAUDIBLE]

PROFESSOR: No, it's not, right?

AUDIENCE: [INAUDIBLE]

PROFESSOR: This is a big X and this is a small x. This is just where you plug in your data. The p-value is the probability that you have more evidence against your null than what you already have.

OK, so now I can write it in terms of cumulative distribution functions. So this is what? This is phi of this guy, which is minus this thing here. Well it's basically 2 times this guy, phi of minus square root of n, Xn bar divided by sigma. That's my p-value. If you give me data, I'm going to compute the average and plug it in there, and it can spit out the p-value. Everybody agrees?

So now I can view this, if I start now looking back I say, well, where does this data come from? Well, it could be a random variable. It came from the realization of this thing. So I can try to, I can think of this value, where now this is a random variable because I just plugged in a random variable in here. So now I view my p-value as a random variable. So I keep switching from small x to large X. Everybody agrees what I'm doing here? So I just wrote it as a deterministic function of some deterministic number, and now the function stays deterministic but the number becomes random.

And so I can think of this as some statistic of my data. And I could say, well, what is the distribution of this random variable? Now if my data is actually normally distributed, so I'm actually under the null, so under the null, that means that Xn bar times square root of n divided by sigma has what distribution? Normal? Well it was sigma, I assume I knew it. So it's N 0, 1,

right? I divided by sigma here. OK?

So now I have this random variable. And so my random variable is now 2 phi of minus absolute value of a Gaussian. And I'm actually interested in the distribution of this thing. I could ask that. Anybody has an idea of how you would want to tackle this thing? If I ask you, what is the distribution of a random variable, how do you tackle this question?

There's basically two ways. One is to try to find something that looks like the expectation of h of x for all h. And you try to write this using change of variables and something that looks like integral of h of x p of x dx. And then you say, well, that's the density. If you can read this for any h, then that's the way you would do it.

But there's a simpler way that does not involve changing variables, et cetera, you just try to compute the cumulative distribution function. So let's try to compute the probability that 2 phi minus N 0, 1, is less than t. And maybe we can find something we know.

OK. Well that's equal to what? That's the probability that a minus N 0, well let's say that an N 0, 1-- sorry, N 0, 1 absolute value is greater than minus phi inverse of t over 2. And that's what? Well, it's just the same thing that we had before. It's equal to-- so if I look again, this is the probability that I'm actually on this side or that side of this number. And this number is what? It's minus phi of t over 2. Why do I have a minus here? That's fine, OK.

So it's actually not this, it's actually the probability that my absolute value-- oh, because phi inverse. OK. Because phi inverse is-- so I'm going to look at t between 0 and-- so this number is ranging between 0 and 1. So it means that this number is ranging between 0-- well, the probability that something is less than t should be ranging between the numbers that this guy takes, so that's between 0 and 2. Because this thing takes values between 0 and 2. I want to see 0 and 1, though.

AUDIENCE: Negative absolute value is always less than [INAUDIBLE].

PROFESSOR: Yeah. You're right, thank you. So this is always some number which is less than 0, so the probability that the Gaussian is less than this number is always less than the probability it's less than 0, which is 1/2, so t only has to be between 0 and 1. Thank you.

And so now for t between 0 and 1, then this guy is actually becoming something which is positive, for the same reason as before. And so that's what? That's just basically 2 times phi of phi inverse of t over 2. That's just playing with the symmetry a little bit. You can look at the

areas under the curve. And so what it means is that those two guys cancel. This is the identity. And so this is equal to t.

So which distribution has a density-- sorry, which distribution has a cumulative distribution function which is equal to t for t between 0 and 1? That's the uniform distribution, right? So it means that this guy follows a uniform distribution on the interval 0, 1.

And you could actually check that. For any test you're going to come up with, this is going to be the case. Your p-value under the null will have a distribution which is uniform. So now if somebody shows up and says, here's my test, it's awesome, it just works great. I'm not going to explain to you how I built it, it's a complicated statistics that involve moments of order 27. And I'm like, OK, you know, how am I going to test that your test statistic actually makes sense? Well one thing I can do is to run a bunch of data, draw a bunch of samples, compute your test statistic, compute the p-value, and check if my p-value has a uniform distribution on the interval 0, 1. But for that I need to have a test that, given a bunch of observations, can tell me whether they're actually distributed uniformly on the interval 0, 1. And again one thing I could do is build a histogram and see if it looks like that of a uniform, but I could also try to be slightly more quantitative about this.

AUDIENCE: Why does the [INAUDIBLE] have to be for a [INAUDIBLE]?

PROFESSOR: For two tests?

AUDIENCE: For each test. Why does the p-value have to be normal? I mean, uniform.

PROFESSOR: It's uniform under the null. So because my test statistic was built under the null, and so I have to be able to plug in the right value in there, otherwise it's going to shift everything for this particular test.

AUDIENCE: At the beginning while your probabilities were of big Xn, that thing. That thing is the p-value.

PROFESSOR: That's the p-value, right? That's the definition of the p-value.

AUDIENCE: OK.

PROFESSOR: So it's the probability that my test statistic exceeds what I've actually observed.

AUDIENCE: So how you run the test is basically you have your observations and plug them into the cumulative distribution function for a normal, and then see if it falls under the given--

PROFESSOR: Yeah. So my p-value is just this number when I just plug in the values that I observe here. That's one number. For every dataset you're going to give me, it's going to be one number. Now what I can do is generate a bunch of datasets of size n, like 200 of them. And then I'm going to have a new sample of say 200, which is just the sample of 200 p-values. And I want to test if those p-values have a uniform distribution. OK? Because that's the distribution they should be having. All right?

OK. This one we've already seen. Does x have a PMF with 30%, 50%, and 20%? That's something I could try to test. That looks like your grade point distribution for this class. Well not exactly, but that looks like it.

So all these things are known as goodness of fit tests. The goodness of fit test is something that you want to know if the data that you have at hand follows the hypothesized distribution. So it's not a parametric test. It's not a test that says, is my mean equal to 25 or not. Is my proportion of heads larger than 1/2 or not? It's something that says, my distribution this particular thing. So I'm going to write them as goodness of fit, G-O-F here. You don't need to have parametric modeling to do that.

So how do I work? So if I don't have any parametric modeling, I need to have something which is somewhat non-parametric, something that goes beyond computing the mean and the standard deviation, something that computes some intrinsic non-parametric aspect of my data. And just like here we made this computation, what we did is we said well, if I actually check that the CDF of my data, that my p-value is uniform, then I know it's uniform. So it means that the cumulative distribution function has an intrinsic value about it that captures the entire distribution. Everything I need to know about my distribution is captured by the cumulative distribution function.

Now I have an empirical way of computing, I have a data-driven way of computing an estimate for the cumulative distribution function, which is using the old statistical trick which consists of replacing expectations by averages.

So as I said, the cumulative distribution function for any distribution, for any random variable, is-- so F of t is the probability that X is less than or equal to t, which is equal to the expectation of the indicator that X is less than or equal to t. That's the definition of a probability. And so here I'm just going to replace expectation by the average. That's my usual statistical trick.

And so my estimator Fn for-- the distribution is going to be 1 over n sum from i equal 1 to n of these indicators. And this is called the empirical CDF. It's just the data version of the CDF. So I just replaced this expectation here by an average.

Now when I sum indicators, I'm actually counting the number of them that satisfy something. So if you look at what this guy is, this is the number of X i's that is less than t, right? And so if I divide by n, it's the proportion of observations I have that are less than t. That's what the empirical distribution is.

That's what's written here, the number of data points that are less than t. And so this is going to be something that's sort of trying to estimate one or the other. And the law of large number actually tells me that for any given t, if n is large enough, Fn of t should be close to F of t. Because it's an average. And this entire thing, this entire statistical trick, which consists of replacing expectations by averages, is justified by the law of large number. Every time we used it, that was because the law of large number sort of guaranteed to us that the average was close to the expectation.

OK. So law of large numbers tell me that Fn of t converges, so that's the strong law, says that almost surely actually Fn of t goes to F of t. And that's just for any given t. Is there any question about this? That averages converge to expectation, that's the law of large number. And almost surely we could say in probability it's the same, that would be the weak law of large number.

Now this is fine. For any given t, the average converges to the true. It just happens that this random variable is indexed by t, and I could do it for t equals 1 or 2 or 25, and just check it again. But I might want to check it for all t's at once. And that's actually a different result. That's called a uniform result. I want this to hold for all t at the same time.

And it may be the case that it works for each t individually but not for all t's at the same time. What could happen is that for t equals 1 it converges at a certain rate, and for t equals 2 it converges at a bit of a slower rate, and for t equals 3 at a slower rate and slower rate. And so as t goes to infinity, the rate is going to vanish and nothing is going to converge. That could happen. I could make this happen at a finite point. There's many ways where it could make this happen.

Let's see how that could work. I could say, well, actually no. I still need to have this at infinity for some reason. It turns out that this is still true uniformly, and this is actually a much more

complicated result than the law of large number. It's called Glivenko-Cantelli Theorem. And the Glivenko-Cantelli Theorem tells me that, for all t's at once, Fn converges to F.

So let me just show you quickly why this is just a little bit stronger than the one that we had. If sup is confusing you, think of max. It's just the max over an infinite set. And so what we know is that Fn of t goes to F of t as n goes to infinity. And that's almost surely. And that's the law of large numbers. Which is equivalent to saying that Fn of t minus F of t as n goes to infinity converges almost surely to 0, right? This is the same thing.

Now I want this to happen for all t's at once. So what I'm going to do-- oh, and this is actually equivalent to this. And so what I'm going to do is I'm going to make it a little stronger. So here the arrow only goes one way. And this is where the sup for t in R of Fn of t. And you could actually show that this happens also almost surely.

Now maybe almost surely is a bit more difficult to get a grasp on. Does anybody want to see, like why this statement for this sup is strictly stronger than the one that holds individually for all t's? You want to see that? OK, so let's do that.

So forget about it almost surely for one second. Let's just do it in probability. The fact that Fn of t converges to F of t for all t, in probability means that this goes to 0 as n goes to infinity for any epsilon. For any epsilon in t we know we have this. That's the convergence in probability.

Now what I want is to put a sup here. The probability that the sup is lower than epsilon, might be actually always larger than, never go to 0 in some cases. It could be the case that for each given t, I can make n large enough so that this probability becomes small. But then maybe it's an n of t. So this here means that for any-- maybe I shouldn't put, let me put a delta here. So for any epsilon, for any t and for any epsilon, there exists n, which could depend on both epsilon and t, such that the probability that Fn t minus F of t exceeding delta is less than epsilon t. There exists an n and a delta. No, that's for all delta, sorry. So this is true. That's what this limit statement actually means.

But it could be the case that now when I take the sup over t, maybe that n of t is something that looks like t. Or maybe, well, integer part of t. It could be, right? I don't say anything. It's just an n that depends on t. So if this n is just t, maybe t over epsilon, because I want epsilon. Something like this. Well that means that if I want this to hold for all t's at once, I'm going to have to go for the n that works for all t's at once. But there's no such n that works for all t's at

once. The only n that works is infinity. And so I cannot make this happen for all of them.

What Glivenko-Cantelli tells you, it's actually this is not something that holds like this. That the n that depends on t, there's actually one largest n that works for all the t's at once, and that's it. OK. So just so you know why this is actually a stronger statement, and that's basically how it works. Any other question? Yeah.

AUDIENCE: So what's the position for this to have, because the random variable have a finite mean, finite variance?

PROFESSOR: No. Well the random variable does have finite mean and finite variance, because the random variable is an indicator. So it has everything you want. This is one of the nicest random variables, this is a Bernoulli random variable. So here when I say law of large number, that this holds. Where did I write this? I think I erased it. Yeah, the one over there. This is actually the law of large numbers for Bernoulli random variables. They have everything you want. They're bounded. Yes.

AUDIENCE: So I'm having trouble understanding the first statement. So it says, for all epsilon and all t, the probability of that--

PROFESSOR: So you mean this one?

AUDIENCE: Yeah.

PROFESSOR: For all epsilon and all t. So you fix them now. Then the probability that, sorry, that was delta. I changed this epsilon to delta at some point.

AUDIENCE: And then what's the second line?

PROFESSOR: Oh, so then the second line says that, so I'm just rewriting in terms of epsilon delta what this n goes to infinity means. So it means that for any a t and delta, so that's the same as this guy here, then here I'm just going back to rewriting this. It says that for any epsilon there exists an n large enough such that, well, n larger than this thing basically, such that this thing is less than epsilon.

So Glivenko-Cantelli tells us that not only is this thing a good idea pointwise, but it's also a good idea uniformly. And all it's saying is if you actually were happy with just this result, you should be even happier with that result. And both of those results only tell you one thing.

They're just telling you that the empirical CDF is a good estimator of the CDF.

Now since those indicators are Bernoulli distributions, I can actually do even more. So let me get this guy here. OK so, those guys, Fn of t, this guy is a Bernoulli distribution. What is the parameter of this Bernoulli distribution? What is the probability that it takes value 1?

AUDIENCE: F of t.

PROFESSOR: F of t, right? It's just the probability that this thing happens, which is F of t. So in particular the variance of this guy is the variance of this Bernoulli. So it's F of t 1 minus F of t. And I can use that in my Central Limit Theorem. And Central Limit Theorem is just going to tell me that if I look at the average of random variables, I remove their mean, so I look at square root of n Fn of t, which I could really write as xn bar, right? That's really just an xn bar. Minus the expectation, which is F of t, that comes from this guy. Now if I divide by square root of the variance, that's my square root p1 minus p. Then this guy, by the Central Limit Theorem, goes to some N 0, 1. Which is the same thing as you see there, except that the variance was put on the other side. OK.

Do I have the same thing uniformly in t? Can I write something that holds uniformly in t? Well, if you think about it for one second it's unlikely it's going to go too well. In the sense that it's unlikely that the supremum of those random variables over t is going to also be a Gaussian.

And the reason is that, well actually the reason is that this thing is actually a stochastic process indexed by t. A stochastic process is just a sequence in random variables that's indexed by, let's say time. The one that's the most famous is Brownian motion, and it's basically a bunch of Gaussian increments. So when you go from t to just t a little after that, you have add some Gaussian into the thing.

And here it's basically the same thing that's happening. And you would sort of expect, since each of this guy is Gaussian, you would expect to see something that looks like a Brownian motion at the end. But it's not exactly a Brownian motion, it's something that's called the Brownian bridge.

So if you've seen the Brownian motion, if I make it start at 0 for example, so this is the value of my Brownian motion. Let's write it. So this is one path, one realization of Brownian motion. Let's call it w of t as t increases. So let's say it starts at 0 and looks like something like this. So that's what Brownian motion looks like. It's just something that's pretty nasty. I mean it looks

pretty nasty, it's not continuous et cetera, but it's actually very benign in some average way. So Brownian motion is just something, you should view this as if I sum some random variable that are Gaussian, and then I look at this from farther and farther, it's going to look like this.

And so here I cannot have a Brownian motion in the n, because what is the variance of Fn of t minus F of t at t is equal to 1? Sorry, at t is equal to infinity.

AUDIENCE: 0.

PROFESSOR: It's 0, right? The variance goes from 0 at t is negative infinity, because at negative infinity F of t is going to 0. And as t goes to plus infinity, F of t is going to 1, which means that the variance of this guy as t goes from negative infinity to plus infinity is pinned to be 0 on each side. And so my Brownian motion cannot, when I describe a Brownian motion I'm just adding more and more entropy to the thing and it's going all over the place, but here what I want is that as I go back it should go back to essentially 0. It should be pinned down to a specific value at the n.

And that's actually called the Brownian bridge. It's a Brownian motion that's conditioned to come back to where it started essentially. Now you don't need to understand Brownian bridges to understand what I'm going to be telling you. The only thing I want to communicate to you is that this guy here, when I say a Brownian bridge, I can go to any probabilist and they can tell you all the probability properties of this stochastic process. It can tell me the probability that it takes any value at any point. In particular, it can tell me-- the supremum between 0 and 1 of this guy, it could tell me what the cumulative distribution function of this thing is, can tell me what the density of this thing is, can tell me everything.

So it means that if I want to compute probabilities on this object here, which is the maximum value that this guy can take over a certain period of time, which is basically this random variable. So if I look at the value here, it's a random variable that fluctuates. It can tell me where it is with hyperability, can tell me the quantiles of this thing, which is useful because I can build a table and use it to compute my quantiles and form tests from it.

So that's what actually is quite nice. It says that if I look at the square root of n Fn hat minus sup over t, I get something that looks like the sup of these Gaussians, but it's not really sup of Gaussian, it's sup of a Brownian motion.

Now there's something you should be very careful here. I cheated a little bit. I mean, I didn't cheat, I can do whatever I want. But my notation might be a little confusing. Everybody sees

that this t here is not the same as this t here? Can somebody see that? Just because, first of all, this guy's between 0 and 1. And this guy is in all of R.

What is this t here? As a function of this t here? This guy is F of this guy. So really, if I want it to be completely transparent and not save the keys of my keyboard, I would read this as sup over t of Fn t minus F of t goes to N distribution as n goes to infinity. The supremum over t, again in R, so this guy is for t in the entire real line, this guy is for t in the entire real line. But now I should write b of what? F of t, exactly. So really the t here is F of the original one.

And so that's a Brownian bridge, where when t goes to infinity the Brownian bridge goes from 0 to 1 and it looks like this. A Brownian bridge at 0 is 0, at 1 it's 0. And it does this. But it doesn't stray too far because I condition it to come back to this point. That's what a Brownian bridge is.

OK. So in particular, I can find a distribution for this guy. And I can use this to build a test which is called the Kolmogorov-Smirnov test. The idea is the following. It says, if I want to test some distribution F0, some distribution that has a particular CDF F0, and I plug it in under the null, then this guy should have pretty much the same distribution as the supremum of Brownian bridge. And so if I see this to be much larger than it should be when it's the supremum of a Brownian bridge, I'm actually going to reject my hypothesis.

So here's the test. I want to test whether H0, F is equal to F0, and you will see that most of the goodness of fit tests are formulated mathematically in terms of the cumulative distribution function. I could formulate them in terms of personality density function, or just write x follows N 0, 1, but that's the way we write it. We formulate them in terms of cumulative distribution function because that's what we have a handle on through the empirical cumulative distribution function. And then it's versus H1, F is not equal to F0.

So now I have my empirical CDF. And I hope that for all t's, Fn of t should be close to F0 of t. Let me write it like this. I put it on the exponent because otherwise that would be the empirical distribution function based on zero observations. Now I form the following test statistic.

So my test statistic is tn, which is the supremum over t in the real line of square root of n Fn of t minus F of t, sorry, F0 of t. So I can compute everything. I know this from the data, and this is the one that comes from my null hypothesis. As I can compute this thing. And I know that if this is true, this should actually be the supremum of a Brownian bridge. Pretty much.

And so the Kolmogorov-Smirnov test is simply, reject if this guy, tn, in absolute value, no actually not in absolute value. This is just already absolute valued. Then this guy should be what? It should be larger than the q alpha over 2 distribution that I have. But now rather than putting N 0, 1, or Tn, this is here whatever notation I have for supremum of Brownian bridge.

Just like I did for any pivotal distribution. That was the same recipe every single time. I formed the test statistic such that the asymptotic distribution did not depend on anything I know, and then I would just reject when this pivotal distribution was larger than something. Yes?

AUDIENCE: I'm not really sure why Brownian bridge appears.

PROFESSOR: Do you know what a Brownian bridge is, or?

AUDIENCE: Only vaguely.

PROFESSOR: OK. So this thing here, think of it as being a Gaussian. So for all t you have a Gaussian distribution. Now a Brownian motion, so if I had a Brownian motion I need to tell you what the-- so it's basically a Brownian motion is something that looks like this. It's some random variable that's indexed by t.

I want, say, the expectation of Xt could be equal to 0 for all t. And what I want is that the increments have a certain distribution. So what I want is that the expectation of Xt minus Xs follows some distribution which is N 0, t minus s. So the increments are bigger as I go farther, in terms of variability. And I also want some covariance structure between the two. So what I want is that the covariance between Xs and Xt is actually equal to the minimum of s and t. Yeah, maybe. Yeah, that should be there.

So this is, you open a probability book, that's what it's going to look like. So in particular, you can see, if I put 0 here and X0 is equal to 0, it has 0 variance. So in particular, it means that Xt, if I look only at the t-th one, it has some normal distribution with variance t. So this is something that just blows up.

So this guy here looks like it's going to be a Brownian motion because when I look at the left-hand side it has a normal distribution. Now there's a bunch of other things you need to check. It's the fact that you have this covariance, for example, which I did not tell you. But it sure look somewhat like that.

And in particular, when I look at the normal with mean 0 and variance here, then it's clear that

this guy does not have a variance that's going to go to infinity just like the variance of this guy. We know that the variance is forced to be back to 0. And so in particular we have something that has mean 0 always, whose variance has to be 0 at 0, and variance-- sorry, at t equals negative infinity, and variance 1 at t equals plus infinity. So a variance 0 at t equals plus infinity, and so I have to basically force it to be equal to 0 at each n. So the Brownian motion here tends to just go to infinity somewhere, whereas this guy forces it to come back.

Now everything I described to you is on the scale negative infinity to plus infinity, but since everything depends on F of t, I can actually just put that back into a scale, which is 0 and 1 by a simple change of variable. It's called change of time for the Brownian motion. OK? Yeah.

AUDIENCE: So does a Brownian bridge have a variance at each point that's proportional? Like it starts at 0 variance and then goes to 1/4 variance in the middle and then goes back to 0 variance? Like in the same parabolic shape?

PROFESSOR: Yeah. I mean, definitely. I mean by symmetry you can probably infer all the things.

AUDIENCE: Well I can imagine Brownian bridge with a variance that starts at 0 and stays, like, the shape of the variance as you move along.

PROFESSOR: Yeah, so I don't know if-- there is an explicit formula for this, and it's simple. That's what I can tell you, but I don't know what the explicit, off the top of my head what the explicit formula is.

AUDIENCE: But would it have to match this F of t 1 minus F of t structure? Or not?

PROFESSOR: Yeah.

AUDIENCE: Or does the fact that we're taking the supremum--

PROFESSOR: No. Well the Brownian bridge, this is the supremum-- you're right. So this will be this form for the variance for sure, because this is only marginal distributions that don't take-- right, the process is not just what is the distribution at each instant t. It's also how do those distributions interact with each other in terms of covariance. For the marginal distributions at each instance t, you're right, the variance is F of t 1 minus F of t. We're not going to escape that. But then the covariance structure between those guys is a little more complicated. But yes, you're right. For marginal that's enough. Yeah?

AUDIENCE: So the supremum of the Brownian bridge is a number between 0 and 10, let's just say.

PROFESSOR: Yeah, it could be infinity.

AUDIENCE: So it's not symmetrical with respect to 0, so why are we doing all over 2?

PROFESSOR: OK. Did say raise it? Yeah. Because here I didn't say the supremum of the absolute value of a Brownian bridge, I just said the supremum of a Brownian bridge. But you're right, let's just do this like that. And then it's probably cleaner.

So yeah, actually well it should be q alpha. So this is basically, you're right. So think of it as being one-sided. And there's actually no symmetry for the supremum. I mean the supremum is not symmetric around 0, so you're right. I should not use alpha over 2, thank you. Any other question? This should be alpha. Yeah. I mean those slides were written with 1 minus alpha and I have not replaced all instances of 1 minus alpha by alpha. I mean, except this guy, tilde. Well, depends on how you want to call it. But this is still, the probability that Z exceeds this guy should be alpha. OK?

And this can be found in tables. And we can compute the p-value just like we did before. But we have to simulate it because it's not going to depend on the cumulative distribution function of a Gaussian, like it did for the usual Gaussian test. That's something that's more complicated, and typically you don't even try. You get the statistical software to do it for you.

So just let me skip a few lines. This is what the table looks like for the Kolmogorov-Smirnov test. So it just tells you, what is your number of observations, n. Then you want alpha to be equal to 5%, say. Let's say you have nine observations. So if square root of n absolute value of Fn of t minus F of t exceeds this thing, you reject.

Well it's pretty clear from this test is that it looks very nice, and I tell you this is how you build it. But if you think about it for one second, it's actually really an annoying thing to build because you have to take the supremum over t. This depends on computing a supremum, which in practice might be super cumbersome. I don't want to have to compute this for all values t and then to take the maximum of those guys.

It turns out that that's actually quite nice that we don't have to actually do this. What does the empirical distribution function look like? Well, this thing, remember Fn of t by definition was-- so let me go to the slide that's relevant. So Fn of t looks like this.

So what it means is that when t is between two observations, then this guy is actually keeping the same value. So if I put my observations on the real line here. So let's say I have one

observation here, one observation here, one observation here, one observation here, and one observation here, for simplicity. Then this guy is basically, up to this normalization, counting how many observations they have that are less than t. So since I normalize by n, I know that the smallest number here is going to be 0, and the largest number here is going to be 1. So let's say this looks like this. This is the value 1.

At the value, since I take it less than or equal to, when I'm at Xi, I'm actually counting it. So the jump happens at Xi. So that's the first observation, and then I jump. By how much do I jump? Yeah? One over n, right? And then this value belongs to the right. And then I do it again. I know it's not going to work out for me, but we'll see. Oh no actually, I did pretty well.

This is what my cumulative distribution looks like. Now if you look on this slide, there is this weird notation where I start putting now my indices in parentheses. X parenthesis 1, X parenthesis 2, et cetera. Those are called the ordered statistic. It's just because it might be, when my data is given to me I just call the first observation, the one that's on top of the table, but it doesn't have to be the smallest value. So it might be that this is X1 and that this is X2, and then this is X3, X4, and X5. These might be my observations. So what I do is that I call them in such a way that this is actually, I recall this guy X1, which is just really X3. This is X2, X3, X4, and X5. These are my reordered observations in such a way that the smallest one is indexed by one and the largest one is indexed by n.

So now this is actually quite nice, because what I'm trying to do is to find the largest deviation from this guy to the true cumulative distribution function. The true cumulative distribution function, let's say it's Gaussian, looks like this. It's something continuous, for a symmetric distribution it crosses this axis at 1/2, and that's what it looks like. And the Kolmogorov-Smirnov test is just telling me how far do those two curves get in the worst possible case?

So in particular here, where are they the farthest? Clearly that's this point. And so up to rescaling, this is the value I'm going to be interested in. That's how they get as far as possible from each other. Here, something just happened, right? The farthest distance that I got was exactly at one of those dots.

It turns out this is enough to look at those dots. And the reason is, well because after this dot and until the next jump, this guy does not change, but this guy increases. And so the only point where they can be the farthest apart is either to the left of a jump or to the right of a jump. That's the only place where they can be far from each other. And that means that only one

observation. Everybody sees that? The farthest points, the points at which those two curves are the farthest from each other, has to be at one of the observations. And so rather than looking at a sup over all possible t's, really all I need to do is to look at a maximum only at my observations. I just need to check at each of those points whether they're far.

Now here, notice that you did not, this is not written Fn of Xi. The reason is because I actually know what Fn of Xi is. Fn of the i-th order observation is just the number of jumps I've had until this observation. So here, I know that the value of Fn is 1 over n, here it's 2 over n, 3 over n, 4 over n, 5 over n. So I knew that the values of Fn at my observations, and those are actually the only values that Fn can take, are an integer divided by n. And that's why you see i minus 1 over n, or i over n. This is the difference just before the jump, and this is the difference at the jump.

So here the key message is that this is no longer a supremum over all t's, but it's just the maximum from 1 to n. So I really have only two n values to compute. This value and this value for each observation, that's 2n total. I look at the maximum and that's actually the value. And it's actually equal to tn. It's not an approximation. Those things are equal. That's just the only places where those guys can be maximum. Yes?

AUDIENCE: It seems like since the null hypothesis [INAUDIBLE] the entire distribution of theta, this is like strictly more powerful than just doing it [INAUDIBLE].

PROFESSOR: It's strictly less powerful.

AUDIENCE: Strictly less powerful. But is there, is that like a big trade-off that we're making when we do that? Obviously we're not certain in the first place that we want to assume normality. Does it make sense to [INAUDIBLE], the Gaussian [INAUDIBLE].

PROFESSOR: So can you, I'm not sure what question you're asking.

AUDIENCE: So when we're doing a normal test, we're just asking questions about the mus, the means of our distribution. [INAUDIBLE] This one, it seems like it would be both at the same time. [INAUDIBLE] Is this decreasing power [INAUDIBLE]?

PROFESSOR: So remember, here in this test we want to conclude to H0, in the other test we typically want to conclude to H1. So here we actually don't want power, in a way.

And you have to also assume that doing a test on the mean is probably not the only thing

you're going to end up doing on your data after you actually establish that it's normally distributed. Then you have the dataset, you've sort of established it's normally distributed, and then you can just run the arsenal of statistical studies. And we're going to see regression and all sorts of predictive things, which are not just tests if the mean is equal to something. Maybe you want to build a confidence interval for the mean. Then this is not, confidence interval is not a test. So you're going to have to first test if it's normal, and then see if you can actually use the quantiles of a Gaussian distribution or a t distribution to build this confidence interval.

So in a way you should do this as like, the flat fee to enter the Gaussian world, and then you can do whatever you want to do in the Gaussian world. We'll see actually that your question goes back to something that's a little important, is here I said F0 is fully specified. It's like an N 1, 5. But I didn't say, is it normally distributed, which is the question that everybody asks. You're not asking, is it this particular normal distribution with this particular mean and this particular variance.

So how would you do it in practice? Well you would say, I'm just going to replace the mean by the empirical mean and the variance by the empirical variance. But by doing that you're making a huge mistake because you are sort of depriving your test of the possibility to reject the Gaussian hypothesis just based on the fact that the mean is wrong or the variance is wrong. You've already stuck to your data pretty well. And so you're sort of like already tilting the game in favor of H0 big time.

So there's actually a way to arrange for this. OK, so this is about pivotal statistic. We've used this word many times. And So that's how. I'm not going to go into this test. It's really, this is a recipe on how you would actually build the table that I showed you, this table. This is basically the recipe on how to build it. There's another recipe to build it, which is just open a book at this page. That's a little faster. Or use software.

I just wanted to show you. So let's just keep in mind, anybody has a good memory? Let's just keep in mind this number. This is the threshold for the Kolmogorov-Smirnov statistic. If I have 10 observations and I want to do it at 5%, it's about 41%. So that's the number that it should be larger from.

So it turns out that if you want to test if it's normal, and not just the specific normal, this number is going to be different. Do you think the number I'm going to read in a table that's appropriate for this is going to be larger or smaller? Who says larger?

AUDIENCE: Sorry, what was the question?

PROFESSOR: So the question is, this is the number I should see if my test was, is X, say, N 0, 5. Right? That's a specific distribution with a specific F0. So that's the number, I would build the Kolmogorov-Smirnov statistic from this. I would perform a test and check if my Kolmogorov-Smirnov statistic tn is larger than this number or not. If it's larger I'm going to reject. Now I say, actually, I don't want to test if H0 is N 0, 5, but it's just a mu sigma squared for some mu and sigma squared. And in particular I'm just going to plugin mu hat and sigma hat into my F0, run the same statistic, but compare it to a different number.

So the larger the number, the more or less likely am I to reject? The less likely I am to reject, right? So if I just use that number, let's say this is a large number, I would be more tempted to say it's Gaussian. And if you look at the table you would get that if you make the appropriate correction at the same number of observations, 10, and the same level, you get 25% as opposed to 41%.

That means that you're actually much more likely if you use the appropriate test to reject the fact that it's normal, which is bad news, because that means you don't have access to the Gaussian arsenal, and nobody wants to do this. So actually this is a mistake that people do a lot. They use the Kolmogorov-Smirnov test to test for normality without adjusting for the fact that they've plugged in the estimated mean and the estimated variance.

This leads to rejecting less often, right? I mean this is almost half of the number that we had. And then they can be happy and walk home and say, well, I did the test and it was normal. So this is actually a mistake that I believe that genuinely at least a quarter of the people do make in purpose. They just say, well I want it to be Gaussian so I'm just going to make my life easier. So this is the so-called Kolmogorov Lilliefors test. We'll talk about it, well not today for sure.

There's other statistics that you can test, that you can use. And the idea is to say, well, we want to know if the empirical distribution function, the empirical CDF, is close to the true CDF. The way we did it is by forming the difference in looking at the worst possible distance they can be. That's called a sup norm, or L infinity norm, in functional analysis.

So here, this is what it looked like. The distance between Fn and F that we measured was just the supremum distance over all t's. That's one way to measure distance between two functions. But there's an infinite many ways to measure distance between functions. One is something we're much more familiar with, which is the squared L2-norm. This is nice because

this has like an inner product, it has some nice properties. And you could actually just, rather than taking the sup, you could just integrate the squared distance. And this is what leads to Cramier-Von Mises test.

And then there's another one that says, well, maybe I don't want to integrate without weights. Maybe I want to put weights that account for the variance. And this guy is called Anderson-Darling. For each of these tests you can check that the asymptotic distribution is going to be pivotal, which means that there will be a table at the back of some book that tells you what the statistic, the quantiles of square root of n times this guy are asymptotically, basically. Yeah?

AUDIENCE: For the Kolmogorov-Smirnov test, for the table that shows the value it has, it has the value for different n. But I thought we [INAUDIBLE]--

PROFESSOR: Yeah. So that's just to show you that asymptotically it's pivotal, and I can point you to one specific thing. But it turns out that this thing is actually pivotal for each n. And that's why you have this recipe to construct the entire thing, because it's actually not true for all possible n's. Also there's the n that shows up here. So no actually, this is something you should have in mind.

So basically, let me strike what I just said. This thing you can actually, this distribution will not depend on F0 for any particular n. It's just not going to be a Brownian bridge but a finite sample approximation of a Brownian bridge, and you can simulate that just drawing samples from it, building a histogram, and constructing the quantiles for this guy.

AUDIENCE: No one has actually developed a table for Brownian--

PROFESSOR: Oh, there is one. That's the table, maybe. Let's see if we see it at the bottom of the other table. Yeah. See? Over 40, over 30. So this is not the Kolmogorov-Smirnov, but that's the Kolmogorov Lilliefors. Those numbers that you see here, they are the numbers for the asymptotic thing which is some sort of Brownian bridge. Yeah?

AUDIENCE: Two questions. If I want to build the Kolmogorov-Smirnov test, it says that F0 is required to be continuous.

PROFESSOR: Yeah.

AUDIENCE: [INAUDIBLE] If we have, like, probability mass of a particular value. Like some sort of data.

PROFESSOR: So then you won't have this nice picture, right? This can happen at any point because you're going to have discontinuities in F and those things can happen everywhere. And then--

AUDIENCE: Would the supremum still work?

PROFESSOR: You mean the Brownian bridge?

AUDIENCE: Yeah. The Kolmogorov test doesn't say that you have to be able to easily calculate the supremum.

PROFESSOR: No, no, no, but you still need it. You still need it for-- so there's some finite sample versions of it that you can use that are slightly more conservative, which is in a way good news because you're going to conclude more to H0.

And there's are some, I forget the name, it's Kiefer-Wolfowitz, the Kiefer-Dvoretzky-Wolfowitz, an equality which is basically like Hoeffding's inequality. So it's basically up to bad constants telling you the same result as the Brownian bridge result, and those are true all the time. But for the exact asymptotic distribution, you need continuity. Yes.

AUDIENCE: So just a clarification. So when we are testing the Kolmogorov, we shouldn't test a particular mu and sigma squared?

PROFESSOR: Well if you know what they are you can use Kolmogorov-Smirnov, but if you don't know what they are you're going to plug in-- as soon as you're going to estimate the mean and the variance from the data, you should use the one we'll see next time, which is called Kolmogorov Lilliefors. You don't have to think about it too much. We'll talk about it on Thursday. Any other question?

So we're out of time. So I think we should stop here, and we'll resume on Thursday.