PHILIPPE RIGOLLET: --of our limiting distribution, which happen to be Gaussian. But if the central limit theorem told us that the limiting distribution of some average was something that looked like a Poisson or an [? exponential, ?] then we would just have in the same way taken the quintiles of the exponential distribution.

So let's go back to what we had. So generically if you have a set of observations X1 to Xn. So remember for the kiss example they were denoted by R1 to Rn, because they were turning the head to the right, but let's just go back. We say X1 to Xn, and in this case I'm going to assume they're IID, and I'm going to make them Bernoulli with [INAUDIBLE] p, and p is unknown, right?

So what did we do from here? Well, we said p is the expectation of Xi, and actually we didn't even think about it too much. We said, well, if I need to estimate the proportion of people who turn their head to the right when they kiss, I just basically I'm going to compute the average. So our p hat was just Xn bar, which was just 1 over n sum from i over 1 2n of the Xi. The average of the observations was their estimate.

And then we wanted to build some confidence intervals around this. So what we wanted to understand is, how much that this p hat fluctuates. This is a random variable. It's an average of random variables. It's a random variable, so we want to know what the distribution is. And if we know what the distribution is, then we actually know, well, where it fluctuates. What the expectation is. Around which value it tends to fluctuate et cetera.

And so what the central limit theorem told us was if I take square root of n times Xn bar minus p, which is its average. And then I divide it by the standard deviation. Then this thing here converges as n goes to infinity, and we will say a little bit more about what it means in distribution to some standard normal random variable.

So that was the central limit theorem. So what it means is that when I think of this as a random variable, when n is large enough it's going to look like this. And so I understand perfectly its fluctuations. I know that this thing here has-- I know the probability of being in this zone. I know that this number here is 0. I know a bunch of things.

And then, in particular, what I was interested in was that the probability, that's the absolute value of a Gaussian random variable, exceeds q alpha over 2, q alpha over 2. We said that

this was equal to what? Anybody? What was that?

AUDIENCE: [INAUDIBLE]

PHILIPPE RIGOLLET: Alpha, right? So that's the probability. That's my random variable. So this is by definition q alpha over 2 is the number. So that to the right of it is alpha over 2. And this is a negative q alpha over 2 by symmetry. And so the probability that i exceeds-- well, it's not very symmetric, but the probability that i exceeds this value, q alpha over 2, is just the sum of the two gray areas. All right?

So now I said that this thing was approximately equal, due to the central limit theorem, to the probability, that square root of n. Xn bar minus p divided by square root p 1 minus p. Well, absolute value was larger than q alpha over 2. Well, then this thing by default is actually approximately equal to alpha, just because of virtue of the central limit theorem. And then we just said, well, I'll solve for p. Has anyone attempted to solve the degree two equation for p in the homework? Everybody has tried it?

So essentially, this is going to be an equation in p. Sometimes we don't want to solve it. Some of the p's we will replace by their worst possible value. For example, we said one of the tricks we had was that this value here, square root of p 1 minus p, was always less than one half. Until we could actually get the confidence interval that was larger than all possible confidence intervals for all possible values of p, but we could solve for p. Do we all agree on the principle of what we did? So that's how you build confidence intervals.

Now let's step back for a second, and see what was important in the building of this confidence interval. The really key thing is that I didn't tell you why I formed this thing, right? We started from x bar, and then I took some weird function of x bar that depended on p and n. And the reason is, because when I take this function, the central limit theorem tells me that it converges to something that I know. But this very important thing about the something that I know is that it does not depend on anything that I don't know.

For example, if I forgot to divide by square root of p 1 minus p, then this thing would have had a variance, which is the p 1 minus p. If I didn't remove this p here, the mean here would have been affected by p. And there's no table for normal p 1. Yes?

AUDIENCE: [INAUDIBLE]

PHILIPPE RIGOLLET: Oh, so the square root of n terms come from. So really you should view this. So there's a rule and sort of a quiet rule in math that you don't write a divided by b over c, right? You write c times a divided by b, because it looks nicer. But the way you want to think about this is that this is x bar minus p divided by the square root of p 1 minus p divided by n.

And the reason is, because this is actually the standard deviation of this-- oh sorry, x bar n. This is actually the standard deviation of this guy, and the square root of n comes from the [INAUDIBLE] average.

So the key thing was that this thing, this limiting distribution did not depend on anything I don't know. And this is actually called a pivotal distribution. It's pivotal. I don't need anything. I don't need to know anything, and I can read it in a table.

Sometimes there's going to be complicated things, but now we have computers. The beauty about Gaussian is that people have studied them to death, and you can open any stats textbook, and you will see a table again that will tell you for each value of alpha you're interested in, it will tell you what q alpha over 2 is. But there might be some crazy distributions, but as long as they don't depend on anything, we might actually be able to simulate from them, and in particular compute what q alpha over 2 is for any possible value [INAUDIBLE].

And so that's what we're going to be trying to do. Finding pivotal distributions. How do we take this Xn bar, which is a good estimate, and turn it into something which may be exactly or asymptotically does not depend on any unknown parameter. So here is one way we can actually-- so that's what we did for the kiss example, right? And here I mentioned, for example, in the extreme case, when n was equal to 3 we would get a different thing, but here the CLT would not be valid.

And what that means is that my pivotal distribution is actually not the normal distribution, but it might be something else. And I said we can make take exact computations. Well, let's see what it is, right? If I have three observations, so I'm going to have X1, X2, X3. So now I take the average of those guys. OK, so that's my estimate. How many values can this guy take? It's a little bit of counting.

Four values. How did you get to that number? OK, so each of these guys can take value 0, 1, right? So the number of values that it can take, I mean, it's a little annoying, because then I have to sum them, right? So basically, I have to count the number of 1's. So how many 1's can I get, right? Sorry I have to-- yeah, so this is the number of 1's that I-- OK, so let's look at that.

So we get 0, 0, 0. 0, 0, 1. And then I get basically three of them that have just the one in there, right?

So there's three of them. How many of them have exactly two 1's? 2. Sorry, 3, right? So it's just this guy where I replaced the 0's and the 1. OK, so now I get-- so here I get three that take the value 1, and one that gets the value 0. And then I get three that take the value 2, and then one that takes the value 1. The value [? 0 ?] 1's, right? OK, so everybody knows what I'm missing here is just the ones here where I replaced the 0's by 1's. So the number of values that this thing can take is 1, 2, 3, 4. So someone is counting much faster than me.

And so those numbers, you've probably seen them before, right? 1, 3, 3, 1, remember? And so essentially those guys, it takes only three values, which are either 1/3, 1. Sorry, 1/3. Oh OK, so it's 0, sorry. 1/3, 2/3, and 1. Those are the four possible values you can take. And so now-- which is probably much easier to count like that-- and so now all I have to tell you if I want to describe the distribution of this probability of this random variable, is just the probability that it takes each of these values. So X bar 3 takes the value 0 probability that X bar 3 takes the value 1/3, et cetera.

If I give you each of these possible values, then you will be able to know exactly what the distribution is, and hopefully maybe to turn it into something you can compute. Now the thing is that those values will actually depend on the unknown p. What is the unknown p here? What is the probability that X bar 3 is equal to 0 for example? I'm sorry?

AUDIENCE: [INAUDIBLE]

PHILIPPE RIGOLLET: Yeah, OK. So let's write it without making the computation So 1/8 is probably not the right answer, right? For example, if p is equal to 0, what is this probability? 1. If p is 1, what is this probability? 0. So it will depend on p.

So the probability that this thing is equal to 0, is just the probability that all three of those guys are equal to 0. The probability that X1 is equal to 0, and X2 is equal to 0, and X3 is equal to 0. Now my things are independent, so I do what I actually want to do, which say the probability of the intersection is the product of the probabilities, right? So it's just the probability that each of them is equal to 0 to the power of 3. And the probability that each of them, or say one of them is equal to 0, is just 1 minus p.

And then for this guy I just get the probability-- well, it's more complicated, because I have to

decide which one it is. But those things are just the probability of some binomial random variables, right? This is just a binomial, X bar 3. So if I look at X bar 3, and then I multiply it by 3, it's just this sum of independent Bernoulli's with parameter p. So this is actually a binomial with parameter 3 and p. And there's tables for binomials, and they tell you all this.

Now the thing is I want to invert this guy, right? Somehow. This thing depends on p. I don't like it, so I'm going to have to find ways to get this things depending on p, and I could make all these nasty computations, and spend hours doing this. But there's tricks to go around this. There's upper bounds. Just like we just said, well, maybe I don't want to solve the second degree equation in p, because it's just going to capture maybe smaller order terms, right? Things that maybe won't make a huge difference numerically.

You can check that in your problem set one. Does it make a huge difference numerically to solve the second degree equation, or to just use the [INAUDIBLE] p 1 minus p or even to plug in p hat instead of p. Those are going to be the-- problem set one is to make sure that you see what magnitude of changes you get by changing from one method to the other.

So what I wanted to go to is something where we can use something, which is just a little more brute force. So the probability that-- so here is this Hoeffding's inequality. We saw that. That's what we've finished on last time. So Hoeffding's inequality is actually one of the most useful inequalities. If any one of you is doing anything really to algorithms, you've seen that inequality before. It's extremely convenient that it tells you something about bounded random variables, and if you do algorithms typically with things bounded. And that's the case of Bernoulli's random variables, right? They're bounded between 0 and 1.

And so when I do this thing, when I do Hoeffding's inequality, what this thing is telling me is for any given epsilon here, for any given epsilon, what is the probability that Xn bar goes away from its expectation? All right, then we saw that it decreases somewhat similarly to the way a Gaussian would look like. So essentially what Hoeffding's inequality is telling me, is that I have this picture, when I have a Gaussian with mean u, I know it looks like this, right?

What Hoeffding's inequality is telling me is that if I actually take the average of some bounded random variables, then their probability distribution function or maybe math function-- this thing might not even have [INAUDIBLE] the density, but let's think of it as being a density just for simplicity-- it's going to be something that's going to look like this. It's going to be somewhat-- well, sometimes it's going to have to escape just for the sake of having integral 1.

But it's essentially telling me that those guys stay below those guys. The probability that Xn bar exceeds mu is bounded by something that decays like to tail of Gaussian.

So really that's the picture you should have in mind. When I average bounded random variables, I actually have something that might be really rugged. It might not be smooth like a Gaussian, but I know that it's always bounded by a Gaussian. And what's nice about it is that when I actually start computing probability that exceeds some number, say alpha over 2, then I know that this I can actually get a number, which is just-- sorry, the probability that it exceeds, yeah. So this number that I get here is actually going to be somewhat smaller, right?

So that's going to be the q alpha over 2 for the Gaussian, and that's going to be the-- I don't know, r alpha over 2 for this [? Bernoulli ?] random variable. Like q prime or different q. So I can actually do this without actually taking any limits, right? This is valid for any n. I don't need to actually go to infinity. Now this seems a bit magical, right? I mean, I just said we need n to be, we discussed that we wanted n to be larger than 30 last time for the central limit theorem to kick in, and this one seems to tell me I can do it for any n.

Now there will be a price to pay is that I pick up this 2 over b minus alpha squared. So that's the variance of the Gaussian that I have, right? Sort of. That's telling me what the variance should be, and this is actually not as nice. I pick factor 4 compared to the Gaussian that I would get for that. So let's try to solve it for our case. So I just told you try it. Did anybody try to do it?

So we started from this last time, right? And the reason was that we could say that the probability that this thing exceeds q alpha over 2 is alpha. So that was using CLT, so let's just keep it here, and see what we would do differently. What Hoeffding tells me is that the probability that Xn bar minus-- well, what is mu in this case? It's p, right? It's just notation here. Mu was the average, but we call it p in the case of Bernoulli's, exceeds-- let's just call it epsilon for a second.

So we said that this was bounded by what? So Hoeffding tells me that this is bounded by 2 times exponential minus 2. Now the nice thing is that I pick up a factor n here, epsilon squared. And what is b minus a squared for the Bernoulli's? 1. So I don't have a denominator here. And I'm going to do exactly what I did here. I'm going to set this guy to be equal to alpha. So that if I get alpha here, then that means that just solving for epsilon, I'm going to have some number, which will play the role of q alpha over 2, and then I'm going to be able to just

say that p is between X bar and minus epsilon, and X bar n plus epsilon. OK, so let's do it.

So we have to solve the equation. 2 exponential minus 2n epsilon squared equals alpha, which means that-- so here I'm going to get, there's a 2 right here. So that means that I get alpha over 2 here. Then I take the logs on both sides, and now let me just write it. And then I want to solve for epsilon. So that means that epsilon is equal to square root log q over alpha divided by 2n. Yes?

AUDIENCE: [INAUDIBLE]

PHILIPPE RIGOLLET: Why is b minus a 1?

Well, let's just look, right? X lives in the interval b minus a. So I can take b to be 25, and a to be my negative 42. But I'm going to try to be as sharp as I can. All right, so what is the smallest value you can think of such that a Bernoulli random variable is larger than or equal to this value? What values does a Bernoulli random variable take? 0 and 1. So it takes values between 0 and 1. It just maxes the value. Actually, this is the worst possible case for the Hoeffding inequality.

So now I just get this one, and so now you tell me that I have this thing. So when I solve this guy over there. So combining this thing and this thing implies that the probability that p lives between Xn bar minus square root log 2 over alpha divided by 2n and X bar plus the square root log 2 over alpha divided by 2n is equal to? I mean, is at least. What is it at least equal to?

Here this controls the probability of them outside of this interval, right? It tells me the probability that Xn bar is far from p by more than epsilon. So there's a probability that they're actually outside of the interval that I just wrote. So it's 1 minus the probability of being in the interval. So this is at least 1 minus alpha. So I just use the fact that a probability of the complement is 1 minus the probability of the set. And since I have an upper bound on the probability of the set, I have a lower bound on the probability of the complement.

So now it's a bit different. Before, we actually wrote something that was-- so let me get it back. So if we go back to the example where we took the [INAUDIBLE] over p, we got this guy. q alpha over square root of-- over 2 square root n. So we had Xn bar plus minus q alpha over 2 square root n. Actually, that was q alpha over 2n, I'm sorry about that.

And so now we have something that replaces this q alpha, and it's essentially square root of 2

log 2 over alpha. Because if I replace q alpha by square root of 2 log 2 over alpha, I actually get exactly this thing here. And so the question is, what would you guess? Is this number, this margin, square root of log 2 over alpha divided by 2n, is it smaller or larger than this guy? q alpha all over 2/3n. Yes? Larger. Everybody agrees with this? Just qualitatively? Right, because we just made a very conservative statement. We do not use anything. This is true always. So it can only be better.

The reason in statistics where you use those assumptions that n is large enough, that you have this independence that you like so much, and so you can actually have the central limit theorem kick in, all these things are for you to have enough assumptions so that you can actually make sharper and sharper decisions. More and more confident statement.

And that's why there's all this junk science out there, because people make too much assumptions for their own good. They're saying, well, let's assume that everything is the way I love it, so that I can for sure any minor change, I will be able to say that's because I made an important scientific discovery rather than, well, that was just [INAUDIBLE] OK?

So now here's the fun moment. And actually let me tell you why we look at this thing. So there's actually-- who has seen different types of convergence in the probability statistic class? [INAUDIBLE] students. And so there's different types of-- in the real numbers there's very simple. There's one convergence, Xn turns to X. To start thinking about functions, well, maybe you have uniform convergence, you have pointwise convergence. So if you've done some real analysis, you know there's different types of convergence you can think of.

And in the convergence of random variables, there's also different types, but for different reasons. It's just because the question is, what do you do with the randomness? When you see that something converges to something, it probably means that you're willing to tolerate low probability things happening or where it doesn't happen, and on how you handle those, creates different types of convergence. So to be fair, in statistics the only convergence we care about is the convergence in distribution. That's this one. The one that comes from the central limit theorem.

That's actually the weakest possible you could make. Which is good, because that means it's going to happen more often. And so why do we need this thing? Because the only thing we really need to do is to say that when I start computing probabilities on this random variable, they're going to look like probabilities on that random variable. All right, so for example, think

of the following two random variables, x and minus x.

So this is the same random variable, and this one is negative. When I look at those two random variables, think of them as a sequence, a constant sequence. These two constant sequences do not go to the same number, right? One is plus-- one is x, the other one is minus x. So unless x is the random variable always equal to 0, those two things are different.

However, when I compute probabilities on this guy, and when I compute probabilities on that guy, they're the same. Because x and minus x have the same distribution just by symmetry of the gaps in random variables. And so you can see this is very weak. I'm not saying anything about the two random variables being close to each other every time I'm going to flip my coin, right?

Maybe I'm going to press my computer and say, what is x? Well, it's 1.2. Negative x is going to be negative 1.2. Those things are far apart, and it doesn't matter, because in average those things are going to have the same probabilities that's happening. And that's all we care about in statistics. You need to realize that this is what's important, and why you need to know. Because you have it really good. If your problem is you really care more about convergence almost surely, which is probably the strongest you can think of. So what we're going to do is talk about different types of convergence not to just reflect on the fact on how our life is good. It's just that the problem is that when the convergence in distribution is so weak that it cannot do anything I want with it. In particular, I cannot say that if X converges, Xn converges in distribution, and Yn converges in distribution, then Xn plus Yn converge in distribution to the sum of their limits. I cannot do that. It's just too weak.

Think of this example and it's preventing you to do quite a lot of things. So this is converge in distribution to sum n 0, 1. This is converge in distribution to sum n 0, 1. But their sum is 0, and it's certainly not-- it doesn't look like the sum of two independent Gaussian random variables, right? And so what we need is to have stronger conditions here and there, so that we can actually put things together. And we're going to have more complicated formulas. One of the formulas, for example, is if I replace p by p hat in this denominator. We mentioned doing this at some point.

So I would need that p hat goes to p, but I need stronger than n distributions so that this happens. I actually need this to happen in a stronger sense. So here are the first two strongest sense in which random variables can converge. The first one is almost surely. And who has

already seen this notation little omega when they're talking about random variables? All right, so very few. So this little omega is-- so what is a random variable? A random variable is something that you measure on something that's random.

So the example I like to think of is, if you take a ball of snow, and put it in the sun for some time. You come back. It's going to have a random shape, right? It's going to be a random blurb of something. But there's still a bunch of things you can measure on it. You can measure its volume. You can measure its inner temperature. You can measure its surface area. All these things are random variables, but the ball itself is omega. That's the thing on which you make your measurement. And so a random variable is just a function of those omegas.

Now why do we make all these things fancy? Because you cannot take any function. This function has to be what's called measurable, and there's entire courses on measure theory, and not everything is measurable. And so that's why you have to be a little careful why not everything is measurable, because you need some sort of nice property. So that the measure of something, the union of two things, is less than the sum of the measures, things like that. And so almost surely is telling you that for most of the balls, for most of the omegas, that's the right-hand side. The probability of omega is such that those things converge to each other is actually equal to 1.

So it tells me that for almost all omegas, all the omegas, if I put them together, I get something that has probability of 1. It might be that there are other ones that have probability 0, but what it's telling is that this thing happens for all possible realization of the underlying thing. That's very strong. It essentially says randomness does not matter, because it's happening always.

Now convergence in probability allows you to squeeze a little bit of probability under the rock. It tells you I want the convergence to hold, but I'm willing to let go of some little epsilon. So I'm willing to allow Tn to be less than epsilon. Tn minus T to be-- sorry, to be larger than epsilon. But the problem is they want this to go to 0 as well as n goes to infinity, but for each n this thing does not have to be 0, which is different from here, right?

So this probability here is fine. So it's a little weaker, but it's a slightly different one. I'm not going to ask you to learn and show that one is weaker than the other one. But just know that these are two different types. This one is actually much easier to check than this one.

Then there's something called convergence in Lp. So this one is the fact that it embodies the following fact. If I give you a random variable with mean 0, and I tell you that its variance is

going to 0, right? You have a sequence of random variables, their mean is 0, their expectation is 0, but their variance is going to 0. So think of Gaussian random variables with mean 0, and a variance that shrinks to 0. And this random variable converges to a spike at 0, so it converges to 0, right? And so what I mean by that is that to have this convergence, all I had to tell you was that the variance was going to 0.

And so in L2 this is really what it's telling you. It's telling you, well, if the variance is going to 0-- well, it's for any random variable T, so here what I describe was for a deterministic. So Tn goes to a random variable T. If you look at the square-- the expectation of the square distance, and it goes to 0. But you don't have to limit yourself to the square. You can take power of three. You can take power 67.6, power of 9 pi. You take whatever power you want, it can be fractional. It has to be lower than 1, and that's the convergence in Lp. But we mostly care about integer p.

And then here's our star, the convergence in distribution, and that's just the one that tells you that when I start computing probabilities on the Tn, they're going to look very close to the probabilities on the T. So that was our Tn with this guy, for example, and T was this standard Gaussian distribution. Now here, this is not any probability. This is just the probability then less than or equal to x. But if you remember your probability class, if you can compute those probabilities, you can compute any probabilities just by subtracting and just building things together.

Well, I need this for all x's, so I want this for each x, So you fix x, and then you make the limit go to infinity. You make n go to infinity, and I want this for the point x's at which the cumulative distribution function of T is continuous. There might be jumps, and that I don't actually care for those. All right, so here I mentioned it for random variables. If you're interested, there's also random vectors. A random vector is just a table of random variables. You can talk about random matrices. And you can talk about random whatever you want. Every time you have an object that's just collecting real numbers, you can just plug random variables in there.

And so there's all these definitions that [? extend. ?] So where I see you see an absolute value, we'll see a norm. Things like this. So I'm sure this might look scary a little bit, but really what we are going to use is only the last one, which as you can see is just telling you that the probabilities converge to the probabilities. But I'm going to need the other ones every once in a while. And the reason is, well, OK, so here I'm actually going to the important characterizations of the convergence in distribution, which is R convergence style.

So i converge in distribution if and only if for any function that's continuous and bounded, when I look at the expectation of f of Tn, this converges to the expectation of f of T. OK, so this is just those two things are actually equivalent. Sometimes it's easier to check one, easier to check the other, but in this class you won't have to prove that something converges in distribution other than just combining our existing convergence results.

And then the last one which is equivalent to the above two is, anybody knows what the name of this quantity is? This expectation here? What is it called? The characteristic function, right? And so this i is the complex i, and is the complex number. And so it's essentially telling me that, well, rather than actually looking at all bounded and continuous but real functions, I can actually look at one specific family of complex functions, which are the functions that maps T to E to the ixT for x and R. That's a much smaller family of functions. All possible continuous embedded functions has many more elements than just the real element.

And so now I can show that if I limit myself to do it, it's actually sufficient. So those three things are used all over the literature just to show things. In particular, if you're interested in deep digging a little more mathematically, the central limit theorem is going to be so important. Maybe you want to read about how to prove it. We're not going to prove it in this class. There's probably at least five different ways of proving it, but the most canonical one, the one that you find in textbooks, is the one that actually uses the third element.

So you just look at the characteristic function of the square root of n Xn bar minus say mu, and you just expand the thing, and this is what you get. And you will see that in the end, you will get the characteristic function of a Gaussian. Why a Gaussian? Why does it kick in? Well, because what is the characteristic function of a Gaussian? Does anybody remember the characteristic function of a standard Gaussian?

AUDIENCE: [INAUDIBLE]

PHILIPPE RIGOLLET: Yeah, well, I mean there's two pi's and stuff that goes away, right? A Gaussian is a random variable. A characteristic function is a function, and so it's not really itself. It looks like itself.

Anybody knows what the actual formula is? Yeah.

AUDIENCE: [INAUDIBLE]

PHILIPPE E to the minus?

RIGOLLET:

AUDIENCE: E to the minus x squared over 2.

PHILIPPE RIGOLLET: Exactly. E to the minus x squared over 2. But this x squared over 2 is actually just the second order expansion in the Taylor expansion. And that's why the Gaussian is so important. It's just the second order Taylor expansion. And so you can check it out. I think Terry Tao has some stuff on his blog, and there's a bunch of different proofs. But if you want to prove convergence in distribution, you very likely are going to use one this three right here.

So let's move on. This is when I said that this convergence is weaker than that convergence. This is what I meant. If you have convergence in one style, it implies convergence in the other stuff. So the first [INAUDIBLE] is that if Tn converges almost surely, this a dot s dot means almost surely, then it also converges in probability and actually the two limits, which are this random variable T, are equal almost surely. Basically what it means is that whatever you measure one is going to be the same that you measure on the other one. So that's very strong.

So that means that convergence almost surely is stronger than convergence in probability. If you're converge in Lp then you also converge in Lq for sum q less than p. So if you converge in L2, you'll also converge in L1. If you converge in L67, you converge in L2. If you're converge in L infinity, you converge in Lp for anything. And so, again, limits are equal.

And then when you converge in distribution, when you converge in probability, you also converge in distribution. OK, so almost surely implies probability. Lp implies probability. Probability implies distribution. And here note that I did not write, and the limits are equal almost surely. Why? Because the convergence in distribution is actually not telling you that your random variable is converging to another random variable. It's telling you that the distribution of your random variable is converging to a distribution.

And think of this, guys. x and minus x. The central limit theorem tells me that I'm converging to some standard Gaussian distribution, but am I converging to x or am I converging to minus x? It's not well identified. It's any random variable that has this distribution. So there's no way the limits are equal. Their distributions are going to be the same, but they're not the same limit. Is that clear for everyone? So in a way, convergence in distribution is really not a convergence of a random variable towards another random variable. It's just telling you the limiting distribution of your random variable [INAUDIBLE] which is enough for us.

And one thing that's actually really nice is this continuous mapping theorem, which essentially tells you that-- so this is one of the theorems that we like, because they tell us you can do what you feel like you want to do. So if I have Tn that goes to T, f of Tn goes to f of T, and this is true for any of those convergence except for Lp. But they have to have f, which is continuous, otherwise weird stuff can happen.

So this is going to be convenient, because here I don't have X to n minus p. I have a continuous function. It's between a linear function of Xn minus p, but I could think of like even crazier stuff to do, and it would still be true. If I took the square, it would converge to something that looks like its distribution. It's the same as the distribution of a square Gaussian.

So this is a mouthful, these two slides-- actually this particular slide is a mouthful. What I have in my head since I was pretty much where you're sitting, is this diagram. So what it tells me-- so it's actually voluntarily cropped, so you can start from any Lq you want large. And then as you decrease the index, you are actually implying, implying, implying until you imply convergence in probability. Convergence almost surely implies convergence in probability, and everything goes to the [? sync, ?] that is convergence in distribution. So everything implies convergence in distribution. So that's basically rather than remembering those formulas, this is really the diagram you want to remember.

All right, so why do we bother learning about those things. That's because of this limits and operations. Operations and limits. If I have a sequence of real numbers, and I know that Xn converges to X and Yn converges to Y, then I can start doing all my manipulations and things are happy. I can add stuff. I can multiply stuff. But it's not true always for convergence in distribution. But it is, what's nice, it's actually true for convergence almost surely. Convergence almost surely everything is true. It's just impossible to make it fail.

But convergence in probability is not always everything, but at least you can actually add stuff and multiply stuff. And it will still give you the sum of the n, and the product of the n. You can even take the ratio if V is not 0 of course. If the limit is not 0, then actually you need Vn to be not 0 as well. You can actually prove this last statement, right? Because it's a combination of the first statement of the second one, and the continuous mapping theorem. Because the function that maps x to 1 over x on everything but 0, is continuous. And so 1 over Vn converges to 1 over V, and then I can multiply those two things. So you actually knew that one.

But really this is not what matters, because this is something that you will do whatever happens. If I don't tell you you cannot do it, well, you will do it. But in general those things don't apply to convergence in distribution unless the pair itself is known to converge in distribution. Remember when I said that these things apply to vectors, then you need to actually say that the vector converges in distributions to the limiting factor.

Now this tells you in particular, since the cumulative distribution function is not defined for vectors, I would have to actually use one of the other distributions, one of the other criteria, which is convergence of characteristic functions or convergence of a function of bounded continuous function of the random variable. 0.2 or 0.3, but 0.1 is not going get you anywhere.

But this is something that's going to be too hard for us to deal with, so we're actually going to rely on the fact that we have something that's even better. There's something that is waiting for us at the end of his lecture, which is called Slutsky's that says that if V, in this case, converges in probability but U converge in distribution, I can actually still do that. I actually don't need both of them to converge in probability. I actually need only one of them to converge in probability to make this statement. But two sum.

So let's go to another example. So I just want to make sure that we keep on doing statistics. And every time we're going to just do a little bit too much probability, I'm going to reset the pressure, and start doing statistics again. All right, so assume you observe the times the inter-arrival time of the T at Kendall. So this is not the arrival time. It's not like 7:56, 8:15. No, it's really the inter-arrival time, right? So say the next T is arriving in six minutes. So let's say [INAUDIBLE] bound. And so you have this inter-arrival time. So those are numbers say, 3, 4, 5, 4, 3, et cetera. So I have this sequence of numbers.

So I'm going to observe this, and I'm going to try to infer what is the rate of T's going out of the station from this. So I'm going to assume that these things are mutually independent. That's probably not completely true. Again, it just means that what it would mean is that two consecutive inter-arrival times are independent. I mean, you can make it independent if you want, but again, this independent assumption is for us to be happy and safe.

Unless someone comes with overwhelming proof that it's not independent and far from being independent, then yes, you have a problem. But it might be the fact that it's actually-- if you have a T that's one hour late. If an inter-arrival time is one hour, then the other T, either they fixed it, and it's going to be just 30 seconds behind, or they haven't fixed it, then it's going to be

another hour behind. So they're not exactly independent, but they are when things work well and approximate.

And so now I need to model a random variable that's positive, maybe not upper bounded. I mean, people complain enough that this thing can be really large. And so one thing that people like for inter-arrival times is exponential distribution. So that's a positive random variable. Looks like an exponential on the right-hand slide, on the positive line. And so it decays very fast towards 0. The probability that you have very large values exponentially small, and there's a [INAUDIBLE] lambda that controls how exponential is defined. It's exponential minus lambda times something.

And so we're going to assume that they have the same distribution, the same random variable. So they're IID, because they are independent, and they're identically distributed. They all have this exponential with parameter lambda, and I'm going to try to learn something about lambda. What is the estimated value of lambda, and can I build a confidence interval for lambda.

So we observe n arrival times. So as I said, the mutual independence is plausible, but not completely justified. The fact that they're exponential is actually something that people like in all this what's called queuing theory. So exponentials arise a lot when you talk about inter-arrival times. It's not about the bus, but where it's very important is call centers, service, servers where tasks come, and people want to know how long it's going to take to serve a task. So when I call at a center, nobody knows how long I'm going to stay on the phone with this person.

But it turns out that empirically exponential distributions have been very good at modeling this. And what it means is that they're actually-- you have this memoryless property. It's kind of crazy if you think about it. What does that thing say? Let's parse it. That's the probability. So this is condition on the fact that T1 is larger than T. So T1 is just say the first arrival time. That means that conditionally on the fact that I've been waiting for the first T, well, the first [INAUDIBLE]. Well, I should probably-- the first subway for more than T conditionally-- so I've been there T minutes already.

Then the probability that I wait for s more minutes. So that's the probability that T1 is learned, and the time that we've already waited plus x. Given that I've been waiting for T minutes, really I wait for s more minutes, is actually the probability that I wait for s minutes total. It's

completely memoryless. It doesn't remember how long have you been waiting. The probability does not change. You can have waited for two hours, the probability that it takes another 10 minutes is going to be the same as if you had been waiting for zero minutes. And that's something that's actually part of your problem set. Very easy to compute.

This is just an analytical property. And you just manipulate functions, and you see that this thing just happen to be true, and that's something that people like. Because that's also something that benefit. And also what we like is that this thing is positive almost surely, which is good when you model arrival times. To be fair, we're not going to be that careful. Because sometimes we are just going to assume that something follows a normal distribution. And in particular, I mean, I don't know if we're going to go into that details, but a good thing that you can model with a Gaussian distribution are heights of students.

But technically with positive probability, you can have a negative Gaussian random variable, right? And the probability being it's probably 10 to the minus 25, but it's positive. But it's good enough for us for our modeling. So this thing is nice, but this is not going to be required. When you're modeling positive random variables, you don't always have to use positive distributions that are supported on positive numbers. You can use distributions like Gaussian.

So now this exponential distribution of T1, Tn they have the same parameter, and that means that in average they have the same inter-arrival time. So this lambda is actually the expectation. And what I'm just saying is that they're identically distributed means that I mean some sort of a stationary regime, and it's not always true. I have to look at a shorter period of time, because at rush hour and 11:00 PM clearly those average inter-arrival times are going to be different So it means that I am really focusing maybe on rush hour.

Sorry, I said it's lambda. It's actually 1 over lambda. I always mix the two. All right, so you have the density of T1. So f of T is this. So it's on the positive real line. The fact that I have strictly positive or larger [INAUDIBLE] to 0 doesn't make any difference. So this is the density. So it's lambda E to the minus lambda T. The lambda in front just ensures that when I integrate this function between 0 and infinity, I get 1.

And you can see, it decays like exponential minus lambda T. So if I were to draw it, it would just look like this. So at 0, what value does it take? Lambda. And then I decay like exponential minus lambda T. So this is 0, and this is f of T. So very small probability of being very large. Of course, it depends on lambda. Now the expectation, you can compute the expectation of this

thing, right? So you integrate T times f of T. This is part of the little sheet that I gave you last time. This is one of the things you should be able to do blindfolded. And then you get the expectation of T1 is 1 over lambda. That's what comes out.

So as I actually tell many of my students, 99% of statistics is replacing expectations by averages. And so what you're tempted to do is say, well, if in average I'm supposed to see 1 over lambda, I have 15 observations. I'm just going to average those observations, and I'm going to see something that should be close to 1 over lambda. So statistics is about replacing averages, expectations with averages, and that's we do.

So Tn bar here, which is the average of the Ti's, is a pretty good estimator for 1 over lambda. So if I want an estimate for lambda, then I need to take 1 over Tn bar. So here is one estimator. I did it without much principle except that I just want to replace expectations by averages, and then I fixed the problem that I was actually estimating 1 over lambda by lambda. But you could come up with other estimators, right?

But let's say this is my way of getting to that estimator. Just like I didn't give you any principled way of getting p hat, which is Xn bar in the kiss example. But that's the natural way to do it. Everybody is completely shocked by this approach? All right, so let's do this.

So what can I say about the properties of this estimator lambda hat? Well, I know that Tn bar is going to 1 over lambda by the law of large number. It's an average. It converges to the expectation both almost surely, and in probability. So the first one is the strong law of large number, the second one is the weak law of large number. I can apply the strong one. I have enough conditions. And hence, what do I apply so that 1 over Tn bar actually goes to lambda? So I said hence. What is hence? What is it based on?

AUDIENCE: [INAUDIBLE]

PHILIPPE RIGOLLET Yeah, continuous mapping theorem, right? So I have this function 1 over x. I just apply this function. So if it was 1 over lambda squared, I would have the same thing that would happen just because the function 1 over x is continuous away from 0.

And now the central limit theorem is also telling me something about lambda. About Tn bar, right? It's telling me that if I look at my average, I remove the expectation here. So if I do Tn bar minus my expectation, rescale by this guy here, then this thing is going to converge to some Gaussian random variable, but here I have this lambda to the negative 1-- to the

negative 2 here, and that's because they did not tell you that if you compute the variance-- so from this, you can probably extract. So if I have X that follows some exponential distribution with parameter lambda. Well, let's call it T.

So we know that T in expectation, the expectation of T is 1 over lambda. What is the variance of T? You should be able to read it from the thing here. 1 over lambda squared. That's what you actually read in the variance, because the central limit theorem is really telling you the distribution goes through this n. But this numbers and this number you can read, right? If you look at the expectation of this guy it's-- of this guy comes out. This is 1 over lambda minus 1 over lambda. That's why you read the 0.

And if you look at the variance of the dot, you get n times the variance of this average. Variance of the average is picking up a factor 1 over n. So the n cancels. And then I'm left with only one of the variances, which is 1 over lambda squared. OK, so we're not going to do that in details, because, again, this is just a pure calculus exercise. But this is if you compute integral of lambda e to the minus t lambda times t squared. Actually t minus 1 over lambda squared dt between 0 and infinity. You will see that this thing is 1 over lambda squared. How would I do this? Configuration by [INAUDIBLE] or you know it. All right.

So this is what the central limit theorem tells me. So this gives me if I solve this, and I plug in so I can multiply by lambda and solve, it would give me somewhat a confidence interval for 1 over lambda. If we just think of 1 over lambda as being the p that I had before, this would give me a central limit theorem for-- sorry, a confidence interval for 1 over lambda. So I'm hiding a little bit under the rug the fact that I have to still define it. Let's just actually go through this. I see some of you are uncomfortable with this, so let's just do it.

So what we've just proved by the central limit theorem is that the probability, that's square root of n Tn minus 1 over lambda exceeds q alpha over 2 is approximately equal to alpha, right? That's just the statement of the central limit theorem, and by approximately equal I mean as n goes to infinity. Sorry I did not write it correctly. I still have to divide by square root of 1 over lambda squared, which is the standard deviation, right? And we said that this is a bit ugly. So let's just do it the way it should be. So multiply all these things by lambda.

So that means now that the absolute value, so with probability 1 minus alpha asymptotically, I have that square root of n times lambda Tn minus 1 is less than or equal to q alpha over 2. So what it means is that, oh, I have negative q alpha over 2 less than square root of n. Let me

divide by square root of n here. lambda Tn minus 1 q alpha over 2. And so now what I have is that I get that lambda is between-- that's Tn bar-- is between 1 plus q alpha over 2 divided by root n. And the whole thing is divided by Tn bar, and same thing on the other side except I have 1 minus q alpha over 2 divided by root n divided by Tn bar.

So it's kind of a weird shape, but it's still of the form 1 over Tn bar plus or minus something. But this something depends on Tn bar itself. And that's actually normal, because Tn bar is not only giving me information about the mean, but it's also giving me information about the variance. So it should definitely come in the size of my error bars. And that's the way it comes in this fairly natural way. Everybody agrees?

So now I have actually built a confidence interval. But what I want to show you with this example is, can I translate this in a central limit theorem for something that converges to lambda, right? I know that Tn bar converges to 1 over lambda, but I also know that 1 over Tn bar converges to lambda. So do I have a central limit theorem for 1 over Tn bar? Technically no, right? Central limit theorems are about averages, and 1 over an average is not an average.

But there's something that statisticians like a lot, and it's called the Delta method. The Delta method is really something that's telling you that you can actually take a function of an average, and let it go to the function of the limit, and you still have a central limit theorem. And the factor or the price to pay for this is something which depends on the derivative of the function.

And so let's just go through this, and it's, again, just like the proof of the central limit theorem. And actually in many of those asymptotic statistics results, this is actually just a Taylor expansion, and here it's not even the second order, it's actually the first order, all right? So I'm just going to do linear approximation of this function. So let's do it. So I have that g of Tn bar-- actually let's use the notation of this slide, which is Zn and theta. So what I know is that Zn minus theta square root of n goes to some Gaussian, this standard Gaussian. No, not standard.

OK, so that's the assumptions. And what I want to show is some convergence of g of Zn to g of theta. So I'm not going to multiply by root n just yet. So I'm going to do a first order Taylor expansion. So what it is telling me is that this is equal to Zn minus theta times g prime of, let's call it theta bar where theta bar is somewhere between say Zn and theta, for sum. OK, so if

theta is less than Zn you just permute those two. So that's what the Taylor first order Taylor expansion tells me. There exists a theta bar that's between the two values at which I'm expanding so that those two things are equal. Is everybody shocked? No? So that's standard Taylor expansion.

Now I'm going to multiply by root n. And so that's going to be what? That's going to be root n Zn minus theta. Ah-ha, that's something I like. Times g prime of theta bar. Now the central limit theorem tells me that this goes to what? Well, this goes to sum n 0 sigma squared, right? That was the first line over there. This guy here, well, it's not clear, right? Actually it is. Let's start with this guy.

What does theta bar go to? Well, I know that Zn is going to theta. Just because, well, that's my law of large numbers. Zn is going to theta, which means that theta bar is sandwiched between two values that converge to theta. So that means that theta bar converges to theta itself as n goes to infinity. That's just the law of large numbers. Everybody agrees? Just because it's sandwiched, right? So I have Zn. I have theta, and theta bar is somewhere here. The picture might be reversed. It might be that Zn end is larger than theta.

But the law of large number tells me that this guy is not moving, but this guy is moving that way. So you know when n is [INAUDIBLE], there's very little wiggle room for theta bar, and it can only get to theta. And I call it the sandwich theorem, or just find your favorite food in there. So this guy goes to theta, and now I need to make an extra assumption, which is that g prime is continuous. And if g prime is continuous, then g prime of theta bar goes to g prime of theta.

So this thing goes to g prime of theta. But I have an issue here. Is that now I have something that converges in distribution and something that converges in say-- I mean, this converges almost surely or saying probability just to be safe. And this one converges in distribution. And I want to combine them. But I don't have a slide that tells me I'm allowed to take the product of something that converges in distribution, and something that converges in probability. This does not exist. Actually, if anything it told me, do not do anything with things that converge in distribution.

And so that gets us to our-- OK, so I'll come back to this in a second. And that gets us to something called Slutsky's theorem. And Slutsky's theorem tells us that in very specific cases, you can do just that. So you have two sequences of random variables, Xn bar, that's Xn that converges to X. And Yn that converges to Y, but Y is not anything. Y is not any random

variable. So X converges in this distribution. Sorry, I forgot to mention, this is very important. Xn converges in distribution, Y converges in probability.

And we know that in generality we cannot combine those two things, but Slutsky tells us that if the limit of Y is a constant, meaning it's not a random variable, but it's a deterministic number 2, just a fixed number that's not a random variable, then you can combine them. Then you can sum them, and then you can multiply them. I mean, actually you can do whatever combination you want, because it actually implies that X, the vector Xn, Yn converges to the vector Xc.

OK, so here I just took two combinations. They are very convenient for us, the sum and the product so I could do other stuff like the ratio if c is not 0, things like that. So that's what Slutsky does for us. So what you're going to have to write a lot in your homework, in your mid-terms, by Slutsky. I know some people are very generous with their by Slutsky. They just do numerical applications, mu is equal to 6, and therefore by Slutsky mu square is equal to 36. All right, so don't do that. Just use, write Slutsky when you're actually using Slutsky.

But this is something that's very important for us, and it turns out that you're going to feel like you can write by Slutsky all the time, because that's going to work for us all the time. Everything we're going to see is actually going to be where we're going to have to combine stuff. Since we only rely on convergence from distribution arising from the central limit theorem, we're actually going to have to rely on something that allows us to combine them, and the only thing we know is Slutsky. So we better hope that this thing works.

So why Slutsky works for us. Can somebody tell me why Slutsky works to combine those two guys? So this one is converging in distribution. This one is converging in probability, but to a deterministic number. g prime of theta is a deterministic number. I don't know what theta is, but it's certainly deterministic. All right, so I can combine them, multiply them. So that's just the second line of that in particular. All right, everybody is with me?

So now I'm allowed to do this. You can actually-- you will see something like counterexample questions in your problem set just so that you can convince yourself. It's always a good thing. I don't like to give them, because I think it's much better for you to actually come to the counterexample yourself. Like what can go wrong if Y is not a random-- sorry, if Y is not a-- sorry, if c is not the constant, but it's a random variable. You can figure that out.

All right, so let's go back. So we have now this Delta method that tells us that now I have a central limit theorem for functions of averages, and not just for averages. So the only price to

pay is this derivative there. So, for example, if g is just a linear function, then I'm going to have a constant multiplication. If g is a quadratic function, then I'm going to have theta squared that shows up there. Things like that. So just think of what kind of applications you could have for this.

Here are the functions that we're interested in, is x maps to 1 over x. What is the derivative of this guy? What is the derivative of 1 over x? Negative 1 over x squared, right? That's the thing we're going to have to put in there. And so this is what we get. So now when I'm actually going to write this, so if I want to show square root of n lambda hat minus lambda. That's my application, right? This is actually 1 over Tn, and this is 1 over 1 over lambda. So the function g of x is 1 over x in this case.

So now I have this thing. So I know that by the Delta method-- oh, and I knew that Tn, remember, square root of Tn minus 1 over lambda was going to sum normal with mean 0 and variance 1 over lambda squared, right? So the sigma square over there is 1 over lambda squared. So now this thing goes to what? Sum normal. What is going to be the mean? 0.

And what is the variance? So the variance is going-- I'm going to pick up this guy, 1 over lambda squared, and then I'm going to have to take g prime of what? Of 1 over lambda, right? That's my theta. So I have g of theta, which is 1 over theta. So I'm going to have g prime of 1 over lambda. And what is g prime of 1 over lambda? So we said that g prime is 1 over negative 1 over x squared. So it's negative 1 over 1 over lambda squared-- sorry, squared. Which is nice, because g can be decreasing. So that would be annoying to have a negative variance.

And so g prime is negative 1 over, and so what I get eventually is lambda squared up here, but then I square it again. So this whole thing here becomes what? Can somebody tell me what the final result is? Lambda squared right? So it's lambda 4 divided by lambda 2. So that's what's written there. And now I can just do my good old computation for a-- I can do a good computation for a confidence interval.

All right, so let's just go from the second line. So we know that lambda hat minus lambda is less than, we've done that several times already. So it's q alpha over 2-- sorry, I should put alpha over 2 over this thing, right? So that's really the quintile of what our alpha over 2 times lambda divided by square root of n. All right, and so that means that my confidence interval should be this, lambda hat. Lambda belongs to lambda plus or minus q alpha over 2 lambda

divided by root n, right? So that's my confidence interval.

But again, it's not very suitable, because-- sorry, that's lambda hat. Because they don't know how to compute it. So now I'm going to request from the audience some remedies for this. What do you suggest we do? What is the laziest thing I can do? Anybody? Yeah.

AUDIENCE: [INAUDIBLE]

PHILIPPE RIGOLLET Replace lambda by lambda hat. What justifies for me to do this?

AUDIENCE: [INAUDIBLE]

PHILIPPE RIGOLLET Yeah, and Slutsky tells me I can actually do it, because Slutsky tells me, where does this lambda come from, right? This lambda comes from here. That's the one that's here. So actually I could rewrite this entire thing as square root of n lambda hat minus lambda divided by lambda converges to sum n 0, 1. Now if I replace this by lambda hat, what I have is that this is actually really the original one times lambda divided by lambda hat. And this converges to n 0, 1, right?

And now what you're telling me is, well, this guy I know it converges to n 0, 1, and this guy is converging to 1 by the law of large number. But this one is converging to 1, which happens to be a constant. It converges in probability, so by Slutsky I can actually take the product and still maintain my conversion to distribution to a standard Gaussian. So you can always do this. Every time you replace some p by p hat, as long as their ratio goes to 1, which is going to be guaranteed by the law of large number, you're actually going to be fine.

And that's where we're going to use Slutsky a lot. When we do plug in, Slutsky is going to be our friend. OK, so we can do this. And that's one way. And then other ways to just solve for lambda like we did before. So the first one we got is actually-- I don't know if I still have it somewhere. Yeah, that was the one, right? So we had 1 over Tn q, and that's exactly the same that we have here. So your solution is actually giving us exactly this guy when we actually solve for lambda.

So this is what we get. Lambda hat. We replace lambda by lambda hat, and we have our asymptotic convergence theorem. And that's exactly what we did in Slutsky's theorem. Now we're getting to it at this point is just telling us that we can actually do this. Are there any questions about what we did here? So this derivation right here is exactly what I did on the board I showed you. So let me just show you with a little more space just so that we all

understand, right? So we know that square root of n lambda hat minus lambda divided by lambda, the true lambda defined converges to sum n 0, 1. So that was CLT plus Delta method.

Applying those two, we got to here. And we know that lambda hat converges to lambda in probability and almost surely, and that's what? That was law of large number plus continued mapping theorem, right? Because we only knew that one of our lambda hat converges to 1 over lambda. So we had to flip those things around. And now what I said is that I apply Slutsky, so I write square root of n lambda hat minus lambda divided by lambda hat, which is the suggestion that was made to me.

They said, I want this, but I would want to show that it converges to sum n 0, 1 so I can legitimately use q alpha over 2 in this one though. And the way we said is like, well, this thing is actually really q divided by lambda times lambda divided by lambda hat. So this thing that was proposed to me, I can decompose it in the product of those two random variables.

The first one here converges through the Gaussian from the central limit theorem. And the second one converges to 1 from this guy, but in probability this time. That was the ratio of two things in probability, we can actually get it. And so now I apply Slutsky. And Slutsky tells me that I can actually do that. But when I take the product of this thing that converges to some standard Gaussian, and this thing that converges in probability to 1, then their product actually converges to still this standard Gaussian [INAUDIBLE]

Well, that's exactly what's done here, and I think I'm getting there. So in our case, OK, so just a remark for Slutsky's theorem. So that's the last line. So in the first example we used the problem dependent trick, which was to say, well, turns out that we knew that p is between 0 and 1. So we have this p 1 minus p that was annoying to us. We just said, let's just bound it by 1/4, because that's going to be true for any value of p.

But here, lambda takes any value between 0 and infinity, so we didn't have such a trick. It's something like we could see that lambda was less than something. Maybe we know it, in which case we could use that. But then in this case, we could actually also have used Slutsky's theorem by doing plug in, right? So here this is my p 1 minus p that's replaced by p hat 1 minus p hat. And Slutsky justify, so we did that without really thinking last time. But Slutsky actually justifies the fact that this is valid, and still allows me to use this q alpha over 2 here.

All right, so that's the end of this lecture. Tonight I will post the next set of slides, chapter two. And, well, hopefully the video. I'm not sure when it's going to come out.