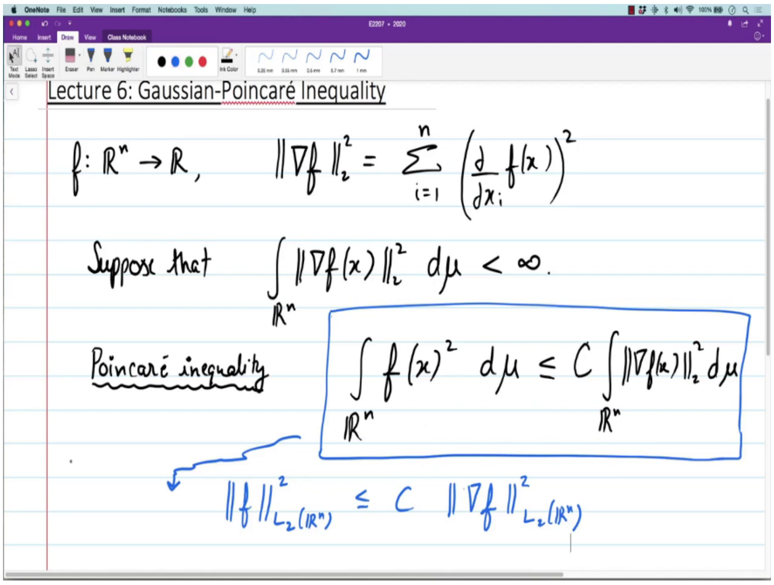
Concentration Inequalities
Prof. Aditya Gopalan
Prof. Himanshu Tyagi
Department of Electrical Communication Engineering
Indian Institute of Science, Bengaluru

Lecture - 07
The Gaussian-Poincare inequality

(Refer Slide Time: 00:21)



In this lecture, we will talk about the Gaussian Poincare inequality. Before I describe this Gaussian Poincare inequality maybe I can quickly review what is the basic classical classic Poincare inequality, which is not connected to this course, but perhaps that will help you understand this name.

So, consider a function f sorry for this consider a function f from R n to R it is a real valued function with n-dimensional domain such that, if you take its gradient; its gradient of this function is an n-dimensional vector. So, you can take its you can take its l 2 norm ok.

So, this is defined as if you remember this, this is summation i = 1 to n f x whole ^{2} ok, that is what this. So, this is the ith dimension of the gradient and that this norm is just sum of all the that the Euclidean length ^{2} of the gradient that is what this quantity is. So, suppose that if you integrate this gradient norm ok, this is the function of x.

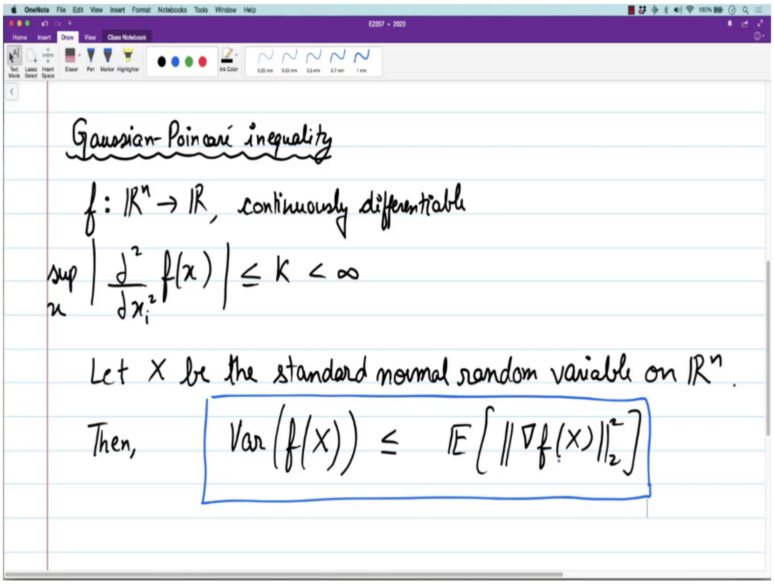
So, I will just put it just to become just to be explicit I will just put it here. And, then you take your let us take the Lebesgue measure here. So, this is all integrating over entire R n suppose that this is finite ok. So, f is a differentiable function continuously differentiable function and this norm the integration of two norm is finite.

Then Poincare inequality actually this I am just writing a very simple consequence of Poincare inequality. Poincare inequality actually extends to any p th norm of this vector lp norm of this vector. It says that if you look at the pth norm of the function itself. So, what is that? So, this function you can take this function, you can look at the value of the function at you can look at the value of the function at x take its ^{2} and then do d mu this is the two norm of ^{2} of the function.

This is ≤ some constant times the pth norm or the two norm in this case ^{2} of this guy. So, this is the Poincare inequality. The integration when we look at this inequality. So, maybe I will first I will write it compactly a little bit more compactly. So, what this says is that look at the function f look at its L 2 norm^2 and this guy is ≤ some C times look at this new function which is the gradient of f ok.

Look at the look at this norm think of this guy here and I will just write it as L 2 R n that is just slightly compact form of slightly more compact form with the same inequality ok. So, this is the Poincare inequality. The Gaussian Poincare inequality that you want to show says that a similar inequality holds, when instead of this Lebesgue measure mu you consider a Gaussian measure ok on R n. So, that is the Gaussian Poincare inequality let me write the statement.

(Refer Slide Time: 04:27)

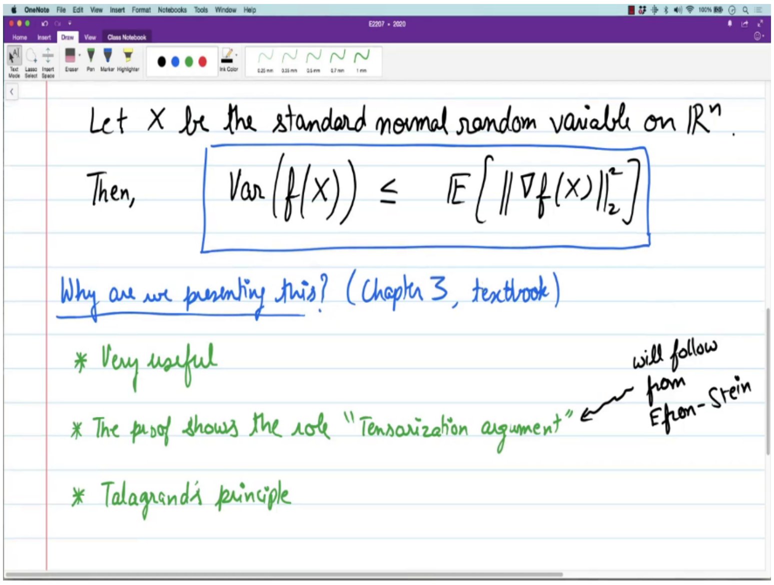


So, once again f is a function from R n to R and it is continuously differentiable. We also assume that its second derivative is bounded. So, let us do that if you look at. So, I will just completely put down what we require to be bounded fix maybe I will just put it this way. So, look at the partial derivative the second derivative along the ith dimension and take sup over x.

So, we assume this guy is ≤ K. So, second derivative is bounded that is what we assume, which is finite ok. So, this is the continuously differentiable function. And let X be the standard normal random variable on R n. So, this is just the Gaussian random variable with mean 0 and variance 1 along all directions ok. So, the inequality is then, if you look at the variance of f of X that must be < the expected value of norm of gradient of f of X whole ^{2}

So, pretty clean actually ok, this is the Gaussian Poincare inequality. So, it looks very much like the Poincare inequality here and you have you just have replaced all these integrals with integrals over the Gaussian measure that is what the inequality claim. So, why are we covering this inequality? The reason we want to cover this inequality is first.

(Refer Slide Time: 07:14)



So, why are we presenting this? by the way this is from the chapter this is from chapter 3 of the textbook ok. The Boucheron, Lugosi, Massart textbook this is chapter 3. So, I have representing this. There are three reasons for me to present this.

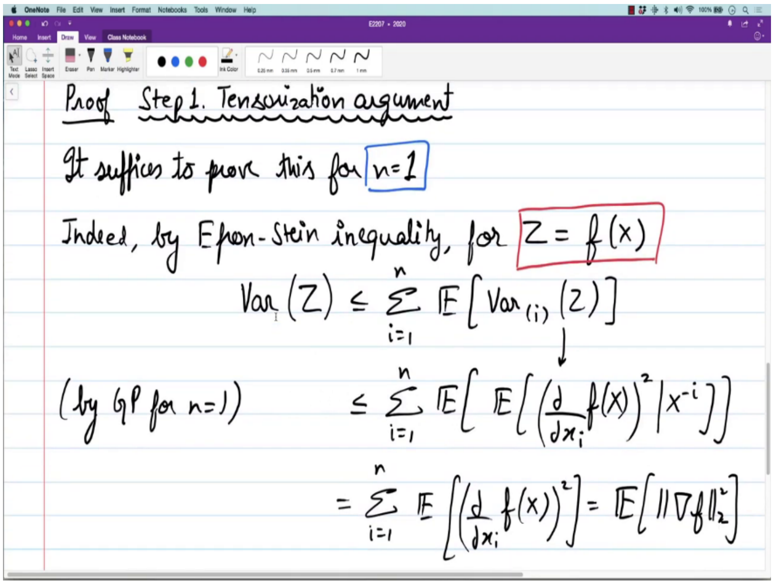
First is that I think it is very useful you may find some application, where you can use it yeah. So, that is the first reason to present this and, but perhaps more importantly, it the proof that will show it will show the role of so called tensorization argument.

So, that is something that will come again and again in this course where you will prove something for one dimension and it will extend to n-dimensions ok. And in this case the tensorization argument will follow from Efron-Stein inequality which we saw in the last lecture. Just a second and the third reason is that this gives an interesting instantiation of Talagrand principle that Aditya mentioned in the first class.

It said that functions which do not depend too much on any one dimension concentrate well around its mean. And in fact, the good measure of how well the function fluctuate in along any one direction is this gradient and if the sum of fluctuations are along all dimension. This is the sum of fluctuation across different dimensions and this gives an exact bound of variance in terms of fluctuation along different directions that is what this does.

So, this can be seen another as a this can be viewed as an instantiation of this Talagrands principle that functions which do not depend on any one dimension too much tend to concentrate around they means ok.

(Refer Slide Time: 10:14)



So, let us show the proof, we have enough build up now. So, we will prove this inequality. The proof is very cool step 1 is what I will call a tensorization argument. Now, I just say that it suffices to prove this for n = 1 ok, that is the claim, but if you show this inequality for one dimension, then automatically it holds for all dimension. So, how do we show that? Indeed, by Efron-Stein inequality the variance of this random variable Z.

So, throughout we will have Z = f of X ok that is our notation, we just I think almost throughout this course we will have this notation. But I will remind you from time to time. So, from variance of this random variable Z from Efron-Stein is ≤ i = 1 to n expected variance given all the other coordinates, but the ith -coordinate. So, you have fixed all X 1 to X i - 1 and X i + 1 to X n this is something we saw in the last lecture.

Now, if you look at this variance inside, it has all the coordinates fixed and only one coordinate is changing. So, this is exactly a variance of a function for a function, which is which is differentiable continuously differentiable this. So, this is just a one dimensional function, if you have fixed all the coordinates and therefore, if the inequality holds the

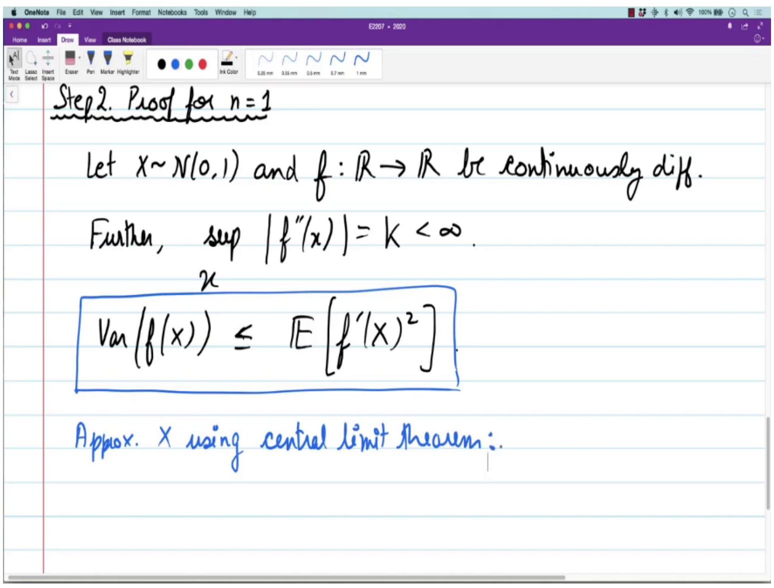
Gaussian Poincare inequality holds / Gaussian Poincare for n = 1 we can apply to this inner variance.

So, what do you get? You get the expected value of right get the square of that this is exactly the Gaussian Poincare inequality applied to one dimension. And this expectation actually is over just the ith coordinate. So, you fix everything, but the ith coordinate. So, let me make it capital X that is why you are evaluating this function and you fix all, but the ith coordinate.

So, this expectation is this one ok. So, we have just applied, you have to convince yourself that this bound I have applied a bound here and this bound is for Gaussian Poincare for n = 1 ok. If this is true then expectation of this conditional expectation is just the expectation. So, this looks like i = 1 to n expected value of this derivative of f x evaluated at this point X and the [2] of this and that is exactly if you can take the, you can take the summation inside.

So, this is exactly ok, this is exactly this ok. So, it suffices to prove it for n = 1. If you have this bound for n = 1, which you applied in the second inequality here and the bound automatically folds holds for a general n.

(Refer Slide Time: 14:02)



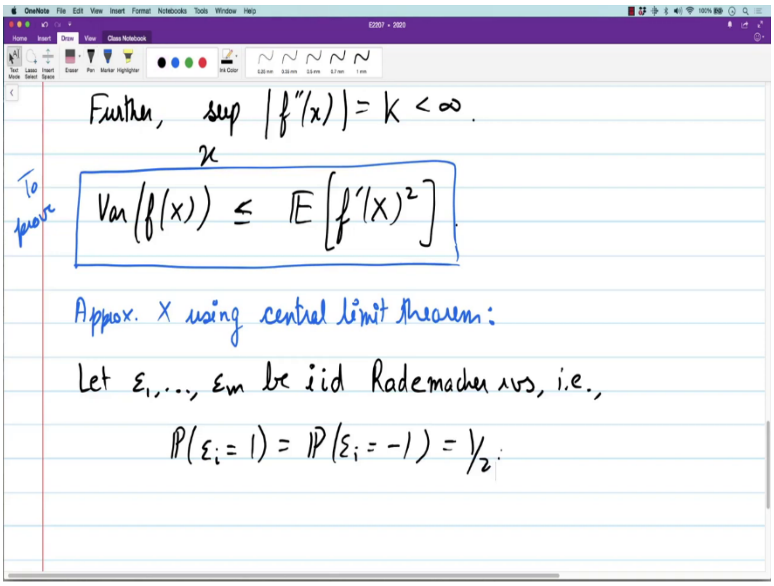
So, now, we do the proof for n = 1, proof for one dimension. This is these are these very interesting set of inequalities, where you show it for one dimension and they automatically follow for multiple dimensions and here we used Efron-Stein to establish this reduction that one dimension suffices to prove the general n dimension proof ok.

Now, how do we; how do we prove this inequality for one dimension ok? So, that proof itself is very interesting. So, what we have to prove given let X be standard normal random variable and f now we can think of it as a function from R to R be continuously differentiable, further we have a bound on the second derivative further f prime prime x this is some assumption I should be careful to write sup all the time.

Sup over x f prime prime x is yeah its some K which is finite ok that is the second assumption. And then we have to prove that expected value sorry, variance of f of X this is just the one dimensional version is $\leq$ the expected value of f prime first derivative of X^2 ok that is the one dimension version of this Gaussian Poincare inequality.

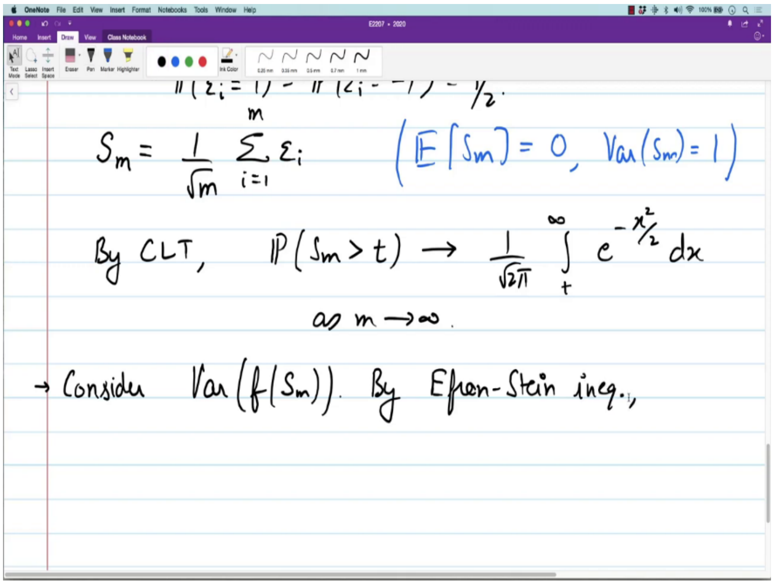
How do we show this? Well again we will use interestingly we will use the Efron-Stein more time and the idea is we will approximate this random variable X approximate X using central limit theorem. How do we do this? My god this is something we have to prove that is what we are proving now.

(Refer Slide Time: 16:26)



So, how do we do that? So, let ϵ 1 to ϵ m be independent and identically distributed Rademacher random variables Rademacher random variables. Sorry. That is each xi these ϵ i takes values - 1 and 1 and probability that ϵ i = 1 = probability that ϵ i = - 1 and that is how.

(Refer Slide Time: 17:20)



So, we take this S m as the normalized sum of these guys. So, note that the expected value of we have to note it here expected value of this S m is 0 and the variance of this S m is the sum of variances / m and what is the sum of variances here? So, this is just the variance of the individual guy. What is the variance of this Rademacher random variable? It is just 1 ok.

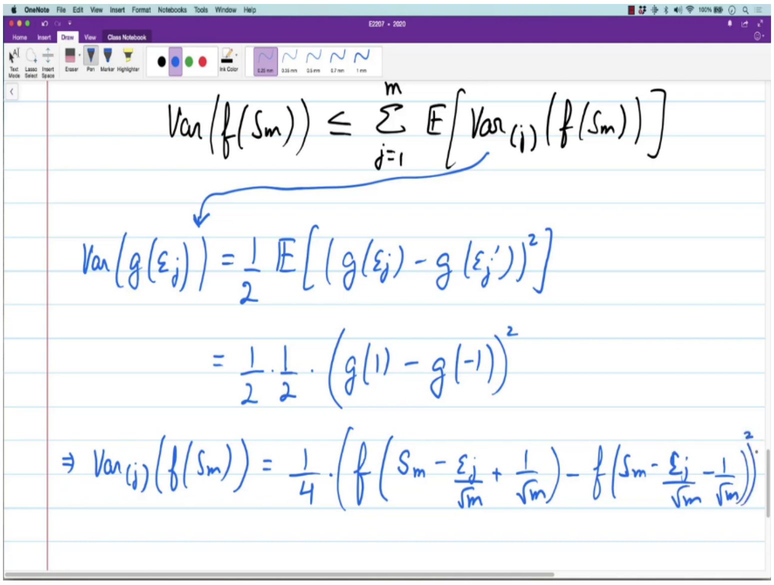
So, this is the unit variance random variable with mean 0. So, then / central limit theorem ok, probability that S m is > t for any t goes to the Gaussian tail function. So, 1 / √ 2 pi integral t to ∞ e to the power - x^2 / 2 dx as m goes to ∞. This is exactly the statement of central limit theorem, all the tails go to this Gaussian tails ok right. We will use this; we will use this central limit theorem.

But what we will do is we will look at the variance of f of S m and bounded using Efron-Stein and we will get a bound and then we will use central limit theorem to translate that bound into a bound about variance of f of X. And the claim is that if f is bounded which in this case it is because the second derivative is bounded and it is continuously differentiable.

So, if f is bounded then the variance of f X = the variance the variance given / this S m in the limit as m going to ∞ ok, that is what we will do. So, we consider so let me just write that part consider variance of f of S m / Efron-Stein inequality applied once again.

So, for a very different reason, the previous time we are applied it for tensorization this time we are applying it to break this X into some small parts here / Efron-Steins or let me just call it the Efron-Stein / the Efron-Stein inequality variance of f of S m is $\leq$ j = 1 to m.

(Refer Slide Time: 20:28)



Now, we are thinking of this E ϵ 1 to ϵ m as different independent random variables. So, we are applying Efron-Stein in a very different way, than what we did before. Variance fixed all ϵ i js ϵ i except the jth one and then look at this variance of f of S m, I am viewing this as a function of ϵ 1 to ϵ 1 to ϵ n m ok that is just the Efron-Stein.

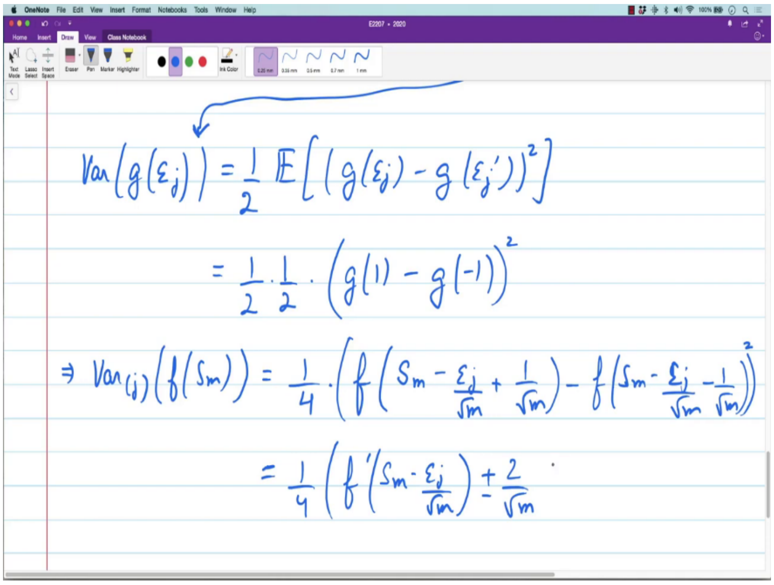
So, let us now, let us look at this variance a bit more closely. So, what is this variance remember that this is just variance of. So, we have fixed everything and we are just changing one random variable, that is we can view it as variance of g of just some ϵ j this is the only one which is changing everything else is just fixed. So, this guy = remember this is variance of this is the expected value, we have seen this before of half into g of ϵ j - g of its independent copy ϵ j prime whole ^{2}.

And, when we do it this way, this is the independent copy now these two independent random variables, if they are equal which happens with probability half. So, these two signs can equal. So, both can be 1 or both can be - 1. In this case, this will be 0 only when they are different this will be something. So, and whenever they are different you what you get is the probability that their different is half and this is just expected value of this is not there is no expectation here.

So, this is just g of 1 - g of - 1 whole ^{2} ok. So, that is for any function g and, we if we substitute it for this one. So, this implies that variance if you fix all, but the jth coordinate of is S m is 1 / 4 into f of so, S m what happens when you fix ϵ j as 1? So, first you take S m - ϵ j / √ m and add to it 1 / √ m, - S m - ϵ j / √ m - 1 / √ m that is what this function is whole ^{2} ok.

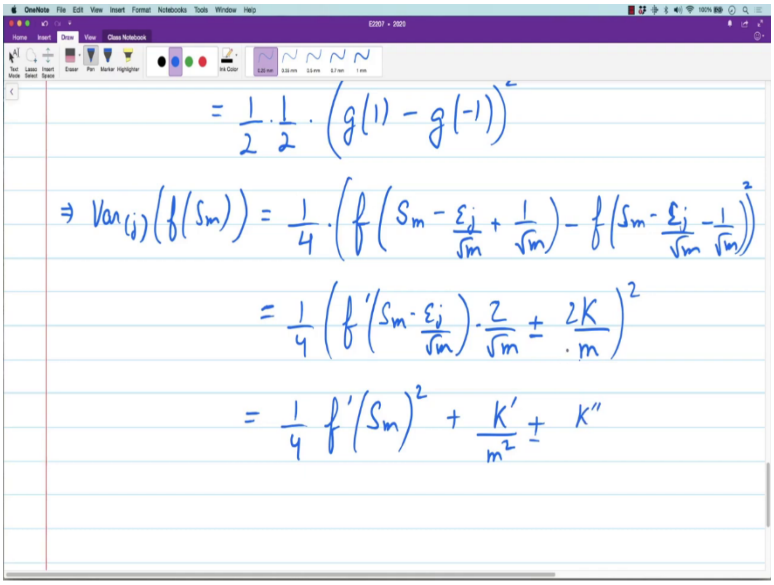
So, we get this expression here for this variance. This is the exact equality and we there is an expectation sitting outside. Now, if you look at this guy here this guy here, we are evaluating some function at this point and adding and subtracting this 1 - √ m ok.

(Refer Slide Time: 23:57)



Therefore we can bounded / this is ≤ this is = 1 / 4 f prime of S m - ϵ j / √ m ok + 2 /√m and then you have the second derivative coming in. So, its I am just putting + - this is all rough ok + - something here the second derivative comes in here, but the second derivative sorry, I am just doing Taylor series approximation.

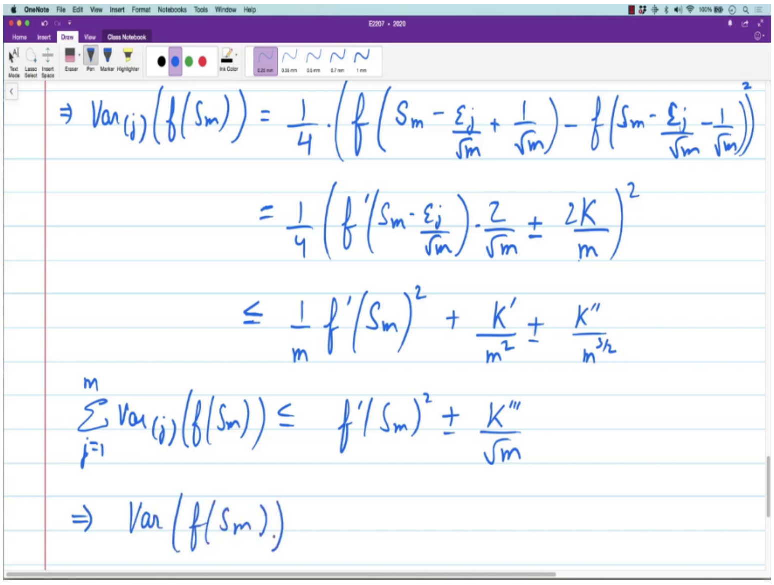
(Refer Slide Time: 24:44)



So, this is this - this is f prime this $2 / \sqrt{m}$ + now the second derivative which is bounded / K + - K and the change again is 1 / m ok K is some constant / 2 let us say 2 / m 2 / m 4 / m. So, something like that this may be wrong, this is just the Taylor series approximation / + - I mean its somewhere in this interval ok.

So, you can square it. In fact, even this guy can be taken here at some constant / m again for the second derivative part; we can take it here again. So, this is 1 / 4 f prime S m^2 + some + - some. So, + some constant K prime / m^2 + - some another constant K prime prime this is just / taking the second part here.

(Refer Slide Time: 26:05)



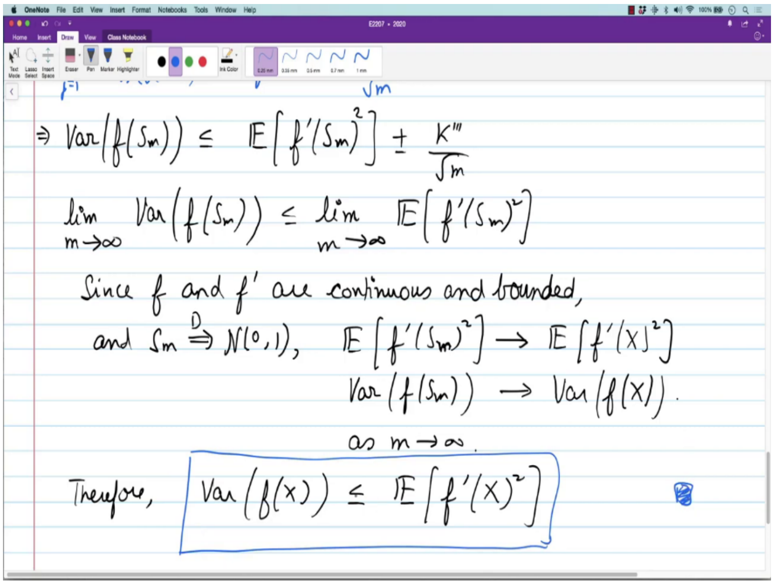
4 / m of this. So, you have another extra power hereof. So, you have m from this guy and you have √ mth from here and you have derivative here, but derivative also is bounded because the second derivative is bounded. So, this is what you get roughly. So, all these things are bounded here ok.

So, that is what we get here. Now, when we take the expectation, see this is no dependence on j here I am all abstracting out all those things / this K prime prime. If this K prime prime is obtained by using this K and again max over f prime, because the second derivative is bounded this max over f prime is also bounded that is how I am writing this ok.

The important point is that therefore, when you sum over this j from 1 to m and look at the variance of j of this f of S m this is ≤ this first term is just f prime S m^2. Since, you have multiplied with m and all the other terms I am just putting some constant here. So, this is the dominating term here so, you get this. Therefore, you can take expectation implies that the variance of f of S m this is what we had here, here, we change back to the black colour.

So, that we can relate to the previous inequality, let me show you the previous inequality yeah that is here.

(Refer Slide Time: 28:30)



So, variance of f of S m is ≤ expected value of f prime S m ^{2} + - some K prime prime / √ m that is what we get. So, taking limit m going to ∞ of this variance of f of S m is ≤ limit m going to ∞ expected value of f prime S m ^{2} + - K prime this thing goes to 0 ok. Because this is / √ m ok that is the main inequality. Now, since f and f prime are continuous functions and S m converges to this Gaussian in distribution.

So, this is convergence and distribution this is my notation for convergence and distribution. You will have that both expected value of f prime S m. So, this is also continuous function for any continuous function this is true in particular f prime^2 is continuous, this guy goes to expected value of f prime X^2 this is just by definition of convergence and distribution.

So, typically we see convergence and distribution as convergence of the law the probability law, but from the dual representation of that convergence you can also get it for as convergence of for all continuous functions bounded continuous functions ok and variance. So, you can get it to expectations, but variance is also an expectation of a continuous function and therefore, variance of f of S m because f was continuous. So, this is also variance is an expectation of some continuous function.

So, this goes to variance f of X ok. So, both these convergence happen as m goes to ∞ ok. And therefore, we have shown that variance of f of X for one-dimension case goes to sorry, is

≤ expected value of f prime X^2 which is the Gaussian Poincare inequality for n = 1 and already we have shown in the first step that it suffices to prove this thing for n = 1.

So, we are done. So, right this is the this is the very interesting proof where we use Efron-Stein twice. First, we used it for tensorization and in the next step we use it to replace a Gaussian random variable with some of iid Rademacher. Rademacher was just chosen for convenience you could have chosen any other random variable a bounded random variable would have been more convenient. And, then you got this change this Efron-Stein gives you relates the variance of f to the change in f, when you change one of those random variables.

And that we could control / the fact that f was continuously differentiable and therefore, we can use Taylor series approximation. In fact, f was twice continuously differentiable. So, we could use Taylor series approximation here. And that gives that gave you a bound for variance of fm in terms of just the f prime S m at that point + some small change which, but there was a division / 1 / √ m.

So, when you take limit m going to ∞, this extra first term goes away and you are just left with this clean inequality. And finally, we notice that since S m goes to Gaussian in distribution for any continuous function f and f prime both these things will go to variance of f X and expected value f prime X^2 ok that is the proof alright. So, this concludes the lectures for this week, I will see you in the next week.