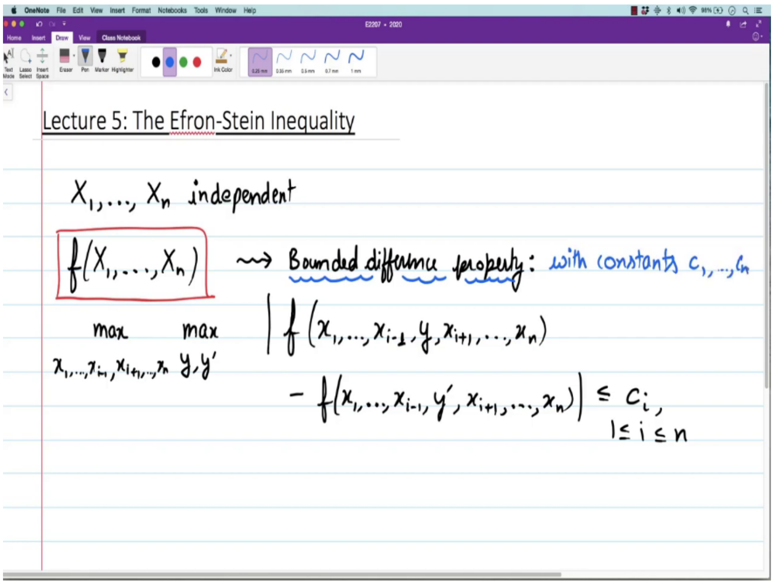
Concentration Inequalities
Prof. Aditya Gopalan
Prof. Himanshu Tyagi
Department of Electrical Communication Engineering
Indian Institute of Science, Bengaluru

Lecture - 06
Bounding variance using the Efron-Stein inequality

(Refer Slide Time: 00:21)



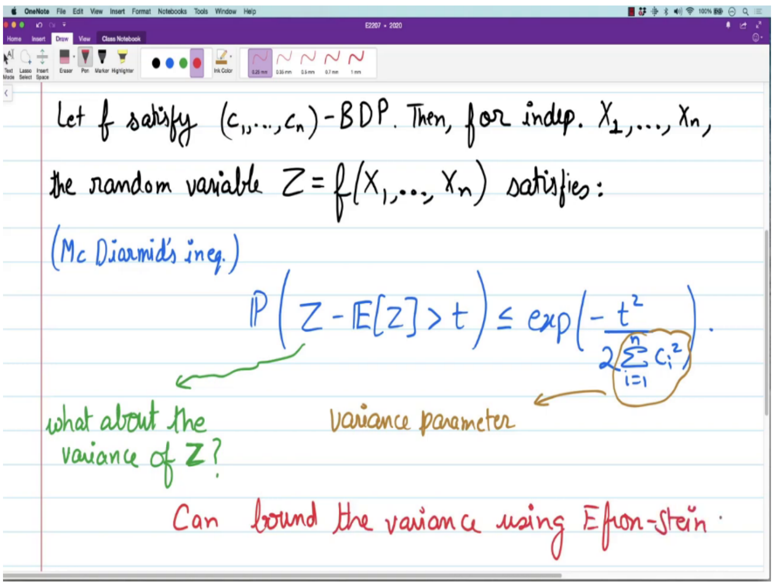
Hi. So, my name is Himanshu Tyagi, as we discussed in the beginning, a part of this course will be covered by Aditya and the other part will be covered by me. So, this from this week onwards for the next 2 weeks, I will be presenting the lectures. Just before I begin, let me quickly try to review what Aditya was doing the last time. So, Aditya was talking about this i i d random variables, just independent random variables X 1 to X n ok, they are all independent and he was talking about this function f of X 1 to X n.

And he was talking about the concentration properties of this function and he related those things to something called the bounded difference property. What was this bounded difference property? It said that if you take this function f and fix all the coordinates, let us say x 1 to x i - 1 and just allow just one coordinate to change, the i^{th} coordinate and see how much can this function change when you modify just this one coordinate, ok.

So, these are two different y and y prime, I fixed all the other coordinates and just modified this one guy. And say you take max over all y and y prime and let us say you also take max over all of the remaining coordinate. So, you take x 1 to x i - 1 and x i + 1 to x n. Suppose, so this is my function f. Suppose this function cannot change / more than C i; here this i denotes the ith coordinate and this holds true for all i's between 1 and n.

So, you have one C i for each coordinate. So, then we say that this function satisfies this bounded difference property ok with constants c 1 to c n, so with constant c 1 to c n. And what Aditya showed was something called the McDiarmid inequality, which says that if you have any function.

(Refer Slide Time: 03:20)



So, let f satisfy and just abbreviate this property above as c 1 to c n bounded difference property ok; then for independent, actually perhaps used a milder assumption there, but let us say independent X 1 to X n. If you look at the probability that, let me bring in notation. So, for independent X 1 to X n the random variable. So, the random variable of interest to us is this random variable Z, which is simply the f of X 1 to X n.

This Z satisfies the following concentration bar; it satisfies this McDiarmid's inequality, it is McDiarmid's inequality. It says that the probability that this random variable Z exceeds its expected value / more than t that probability is less than = e to the power - t ^{2} / 2 summation i

= 1 to n c i square. So, this inequality basically is a generalization of Hoeffding's inequality, where we had a similar bound holding for sum of i i d random, sum of independent random variables with each of which was bounded.

And now just from the sum of independent random variables, we can go to any function of those random variables and we get a similar concentration bound. And this concentration bound if you, I am assuming you are using this terminology; it is a sub-Gaussian concentration bound and this guy here is the variance parameter for the sub-Gaussian concentration bound, ok.

So, it is as if this random variable Z is a Gaussian with variance given by this guy ok, that is what this boundary says. So, you might as well just imagined heuristically as a Gaussian; that is what this McDiarmid inequality says. So, this was a concentration bound and today what we would like to talk about is the variance of this random variable.

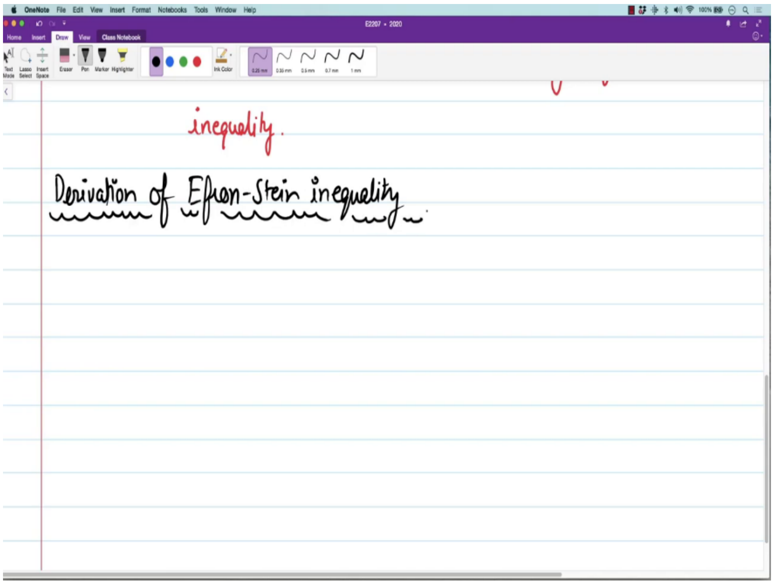
So, we just we have shown something stronger, we have shown a concentration around, it is mean with this variance parameter. But, what about the variance of this random variable? And that is what we will bound today.

So, it is a slight deviation from this concentration inequality topic that you have been seeing; it is now we are certainly talking about variance of Z. But that is as we will see that; firstly, this variance is indeed related to concentration bounds. And secondly, this is of independent interest; you may in some applications be interested in bounding the variance.

And to bound this variance, what we can use is sort of a cousin of McDiarmid's inequality and this is called the Efron Stein inequality, that is what we will see today, the Efron Stein inequality.

So, we can bound the variance using Efron Stein inequality ok, that is what we will do now; bounding the variance of these functions, which satisfy the boundary difference property using Efron Stein inequality, ok. That is what we will do in today's lecture.

(Refer Slide Time: 07:34)

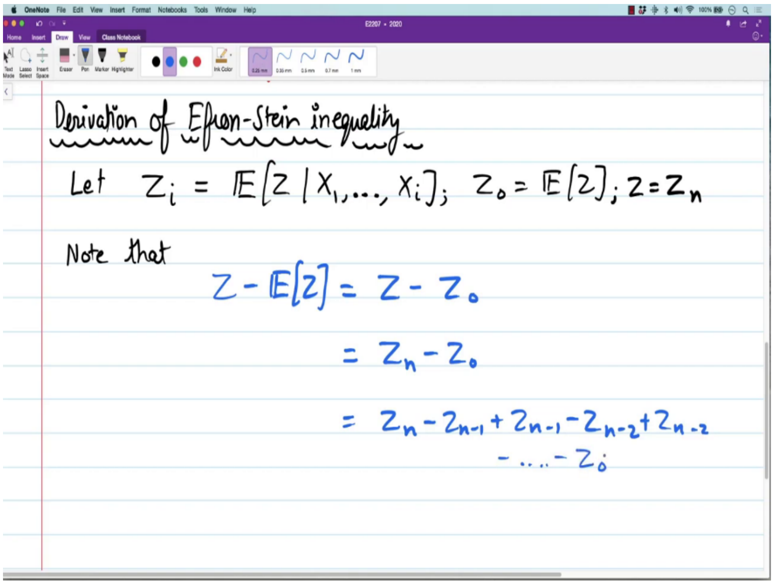


So, let me start with that, I will the first part which is this lecture, I will cover two lectures this week; in the first lecture, I will derive the Efron Stein inequality, ok. So, that is the derivation of Efron Stein inequality.

So, I will state the form at the end, let us go over the derivation of that inequality itself. How can we bound the variance of this random variable Z? As we have seen in the past that, one very basic result about sum of independent random variables is that, its variance is sum of the variance of the individual parts.

In fact, we do not even need independence for this result to hold, we only need those random variables to be uncorrelated. And so, the question now is, can we decompose this random variable Z, that is the function of X 1 to X n into various uncorrelated components? And that is what we will do and here is a trick to do that.

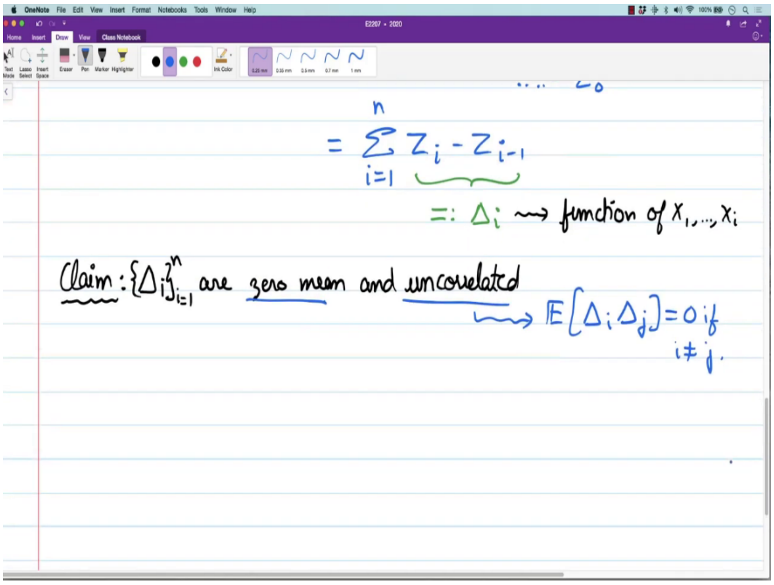
(Refer Slide Time: 09:03)



Let Z i be the expected value of your Z given X 1 to X i ok, that is what Z i is; it is the expected value given X 1 to X i. Then if you look at and let us put in some convention here Z of 0 is just the expected value of Z, it is without conditioning on anything. Then we can express this Z as Z 0; actually we can express this difference, Z - the expected value of Z, this is exactly = Z - expected value Z 0. / the way this Z also is = Z of n.

So, if you condition all the guys, you get Z, right. So, we are doing this telescopic something, so this is Z n - Z 0. So, I will add and subtract things, Z n - Z n - 1 + Z n - 1 - Z n - 2 + Z n - 2 and blah blah + - you will keep on doing and then - Z 0. So, what have I done here; it looks like a very simple trick.

(Refer Slide Time: 10:40)



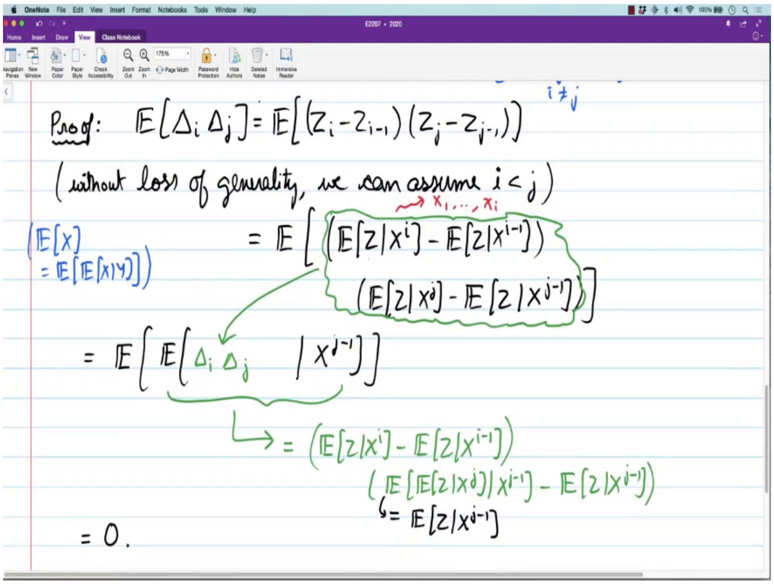
So, I do summation I = 1 to n Z of i - Z of i - 1. And this guy here I will define this guy to be δ i. So, what can we say about this δi? So, that is the, that is something we would like to see, what is so special about this δ i. An important thing about this δ i and this is the claim we have here claim.

Note that this δ i is a random variables by the way and this Z - expected value Z is also a random variable and this is increment when you condition on the first i coordinate and - conditioning on the first i - 1 coordinate, those increments are also random variables. If you do not remember this, just it should take just one second to recall that this conditional expectations are actually random variables and they are functions of what you are conditioning on.

So, they are this is a random variable, which is a function of X 1 to X i, ok. Δ i therefore, also, this to this guy the function of X 1 to X i, this guy is a function of X 1 to X i - 1; therefore, the overall δ i becomes a function of X 1 to X i, ok. So, that is just something more δ i. Claim this random variables δ i are 0 mean and uncorrelated.

So, these random variables δ i for i = 1 to n, actually i = 0 to, i = 1 to n are 0 mean. So, their expected value is 0 and they are uncorrelated. What is uncorrelated? This just means that, the expected value of δ i δ j is = 0, if i is not = j. So, they are uncorrelated, ok.

(Refer Slide Time: 13:32)



How do we prove this? So, this proof is actually an exercise just in conditional expectation, but let us do it for completeness. Let us do expected value of δ i δ j for distinct i and j; this is = the expected value of expected value of Z given X 1 to X i times expected value of Z given X 1 to X j and without loss of generality, we can assume that i is less than j.

So, this has this j is conditioning on more random variables in this guy, sorry should have yeah, should be a little bit more careful here yeah. So, this is expected value of Z i - Z i - 1, Z i - Z i - 1 - Z j - Z j -; that is what δ i is. And this guy equals to the expected value of expected value of Z given X 1 to X i, I will just abbreviate this / X 1 this superscript i, this says X 1 to X i, this is the vector X 1 to X i, ok. Just a short hand, this is my shorthand from here on.

So, this is now expected value of this given this and then you have expected value of. So, this times the expected value of Z given X j - the expected value of Z given Z j - 1, ok. So, note that all these terms here, because j is greater than i are functions of X j, all these terms here.

So, I will use this formula for expectation, where I can first. So, just I will write this formula here, expected value of any random variable X is = expected value of X given any random variable Y ok, that is a formula which you should be familiar with.

So, I will apply this formula and the random variable Y I will choose is X 1 to X j and then everything inside will be a function of X 1 to X j, X 1 to X j - 1. So, this guy here becomes expected value of expected value given X 1 to j - 1, ok. Of what function? Of this guy here;

remember this guy is just δ i δ j, ok. So, that is what I will do, so the expected, so I am just using this formula.

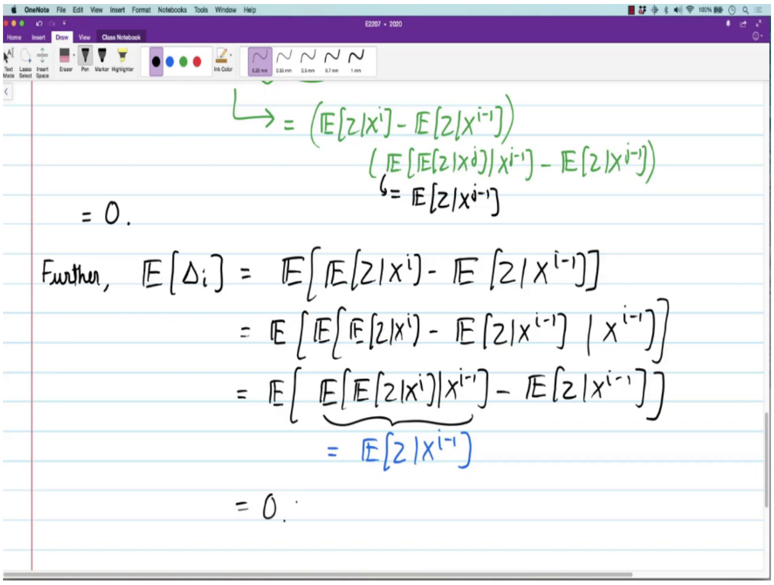
Now, when you condition on j - 1, this guy the first term here becomes a constant, the second term here becomes a constant; because you are conditioning on more than i random variables, more than or = i random variables, j - 1 is greater than or = i. So, these two terms are constant this term is also a constant; this is just j - 1, so this just depends on. So, it is a function of j - 1, so it becomes a constant.

The only random variable that you are left with here is just this guy, ok. So, this expectation here is = expected value of Z given X i; because it becomes a constant value of condition on j - 1 into expected value of Z given X i - 1 into expected value. Now, this is the term which does not remain a constant. So, it becomes expected value of Z given X j given X j - 1 - expected value of Z given X j - 1, because this is again a function.

So, the only thing which changes when you do this expectation, this is a constant, this is a constant, this is the constant; the only thing is that you have you take this further conditional expectation. So, do you know, what is this conditional expectation? Here you are conditioning one more and you are averaging over less. So, this is this is expected value of Z given a b, given expected value Z given a b; the conditional expectation given a, that is just term outside that will remain.

So, this is, this is exactly =, this part here is exactly = expected value of Z given X j - 1, which is the same as this. So, this whole term becomes 0 and therefore, this is just 0. So, these things are indeed uncorrelated, this becomes 0 and let me zoom out, so that you can see the whole proof, let us see 175, right. So, this thing is just 0. And similarly what we will see is that, the expected value of each of this δ i is also 0.

(Refer Slide Time: 20:23)

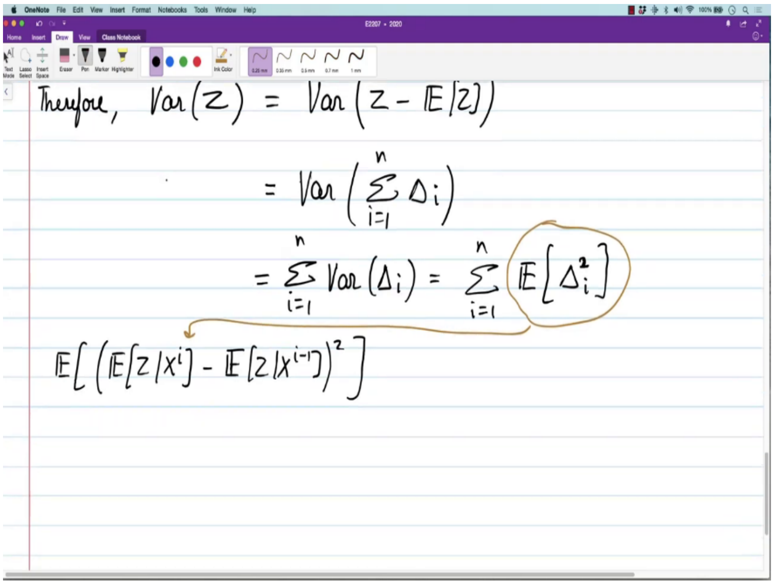


Yes, let us see that further; expected value of any of this δ i is expected value, it is a very similar proof; I will just write it of Z given X i - expected value of Z given X i - 1. And now what I will do again is, I will take conditional expectation given X i - 1 of this term.

So, that is can always do that and now what we notice is, when we take this conditional expectation; this thing since you fix X i - 1, this is a function only of X i - 1, so this is a constant. So, this is expected value of expected value of X i given X i - 1 - this thing, which is just a constant. And this term as we have seen before is = expected value of Z given X i - 1, which is = this and therefore, this whole thing becomes. So, this is 0, this expected value of 0, so this becomes 0.

So, indeed X i is a 0 mean and they are uncorrelated. So, this is very nice, then this decomposition is actually very nice. If you have this random variable Z which is a function of independent random variable, you can decompose it into independent components like this. You can decompose into 0 mean and uncorrelated components like this right; that is what we are saying here and right, ok. So, this is the first claim, we have been able to decompose this difference into uncorrelated component.

(Refer Slide Time: 22:37)



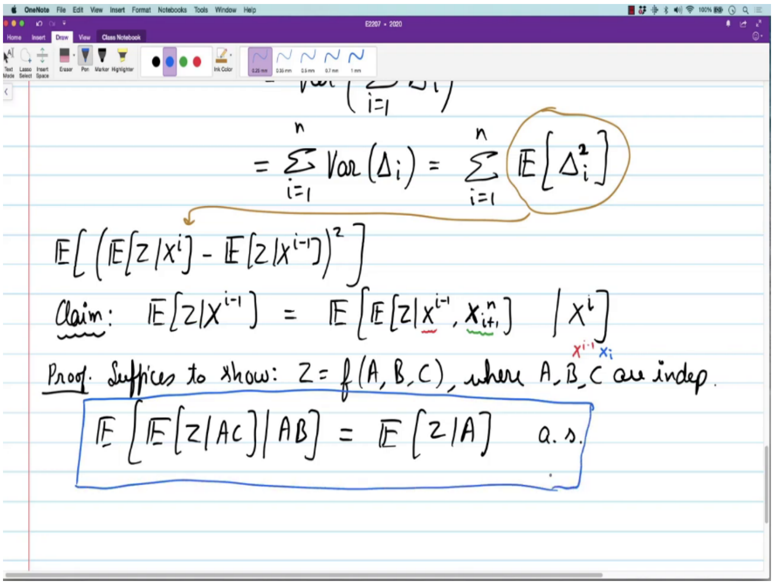
And so, we continue, therefore a goal remember was to bound the variance of Z; variance of this random variable Z is = variance of summation i = 1 to n δ i, where each of this δ i is uncorrelated and 0 mean. And this variance yeah; so maybe one step I will just write; this is just for concreteness, this is true, this is always true, ok. And this guy is summation i = 1 to n δ i; that is something we just checked where each δ i is 0 mean and uncorrelated.

And so therefore, since they are uncorrelated, the variance is additive and since this δ i's are 0 mean, this sum this variance is just expected value of δ i square, ok. So, we will examine this guy now, this expected value of δ i square. Each of these terms somehow controls fluctuation in one direction; that is very concrete here, this one direction and we want to see how much that fluctuation can be.

So, next we look at this yellow term here, the one I have circled here and this guy here this term here is just the expected value of δ i square. So, remember what was δ i; δ i was expected value of Z given X i - the expected value of Z given X i - 1 whole square, that is what this term is, this is this term here and we would like to simplify this further.

So, once again we need to use another property of conditional expectations; we have already used one, we have already used this very interesting property of conditional expectation, which I am assuming all of you are aware of that, the conditional expectation of X is expected value of conditional expectation of X given Y. This is a function of Y and this expectation of this function is = expected value of X.

(Refer Slide Time: 25:11)



Now, we need another property of conditional expectation and I will just write it in this specific context. So, the claim here is that, if you look at this second guy here, this one here; this is the conditional expectation of Z given X i - 1. Claim is that this guy equals to the conditional expectation of something given X Y, maybe I will make some more space for myself, notice that something, a little bit cleaner.

So, that is the conditional expectation of Z given all the past and all the future, ok. So, if you look at this conditional expectation, then this these two are equal that somewhat that is a result, it is a little bit, it looks a little bit complicated.

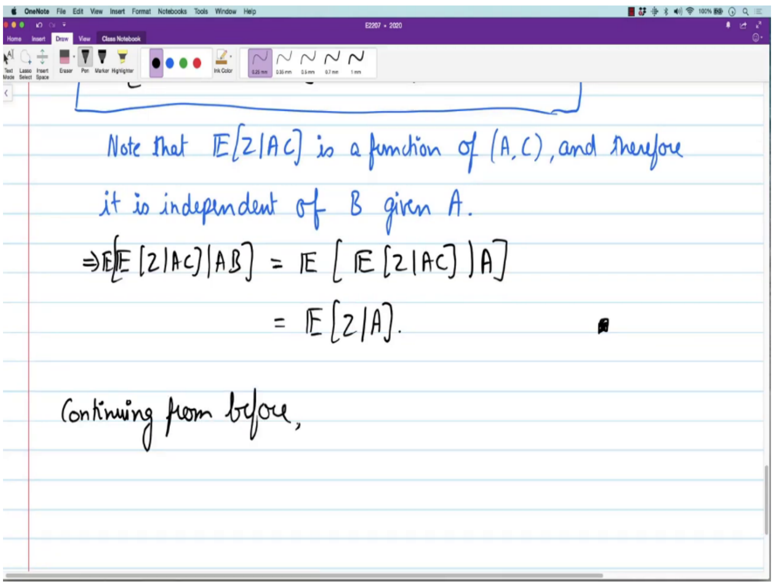
So, this part here, let us let us think of let us think about this result; we will prove it now, but let us just quickly visualize it. We are viewing this function Z as a function of three parts; this part this is the first part, then this part this is the second part. And what is the third part? And third part is X i, that is the third part.

So, this Z is a function of three parts and we are conditioning on two of them. And then in the outside part we are taking another two parts; the part the red part X i - 1 and the blue part X i and this blue part is missing from this first expectation. And the claim is that, when you do that, only the first part is left, that is the claim. So, that is roughly what you would like to prove. So, proof it suffices to show Z for, Z = function of three parts A, B, C, where A, B, C are independent.

The expected value, this outer part here of let me get the inner part; first expected value of Z given AC first part and the third part. Conditional expectation of this given AB, you would like to claim is = the expected value of; just the common part is left here Z given A, that is what you would like to show. And this is almost surely with probability one. So, how do we show this claim? This is what we have to show, it is equivalent to what we have written in.

This is just some simply, this is just some simplification that we are trying to get for expression of δ i; it looks a little bit of a degradation right now, but we will see soon connected to this expression for variance, alright. So, coming back to this game; how do we show this? Well this thing here is a function of A and C and therefore, yeah and therefore, when we condition on both A and B, this is independent of B, ok. So, function of AC condition on both A and B, it is independent of B. So, we agree with that part.

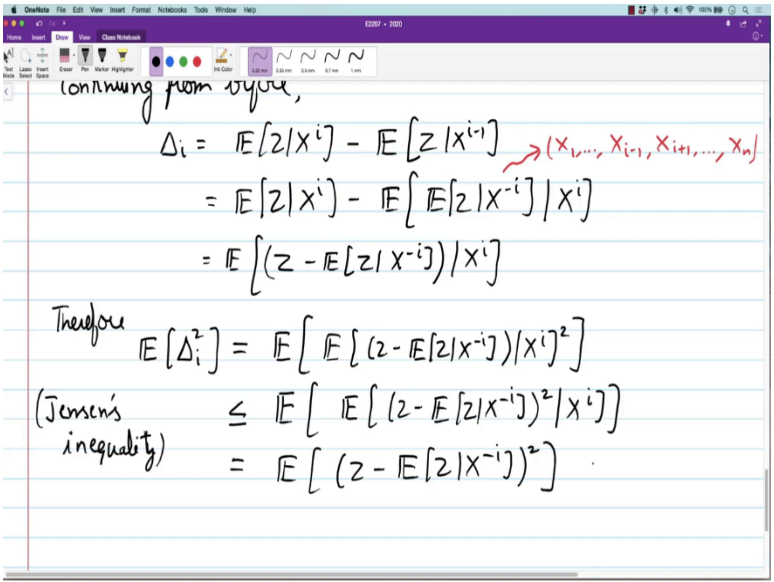
(Refer Slide Time: 29:22)



So, note that expected value of this is a function of A comma C and therefore, it is independent of B even when given condition on A. So, even when you give A, this is independent of B; because A C jointly are independent of B, ok. So, we use this. So, this implies that the expected value of Z given A C, conditional expectation given A B is exactly =; this is from the independence condition Z given A C given A, ok.

And now, this is the familiar formula that you have seen earlier, this is conditioning one more and then conditioning on a part of it. So, that is just the common part remains ok, that is the claim here. So, not much of difficult proof, but it is an interesting observation. So, what this tells us is the following. So, continuing from before, let us go back to our expression for δ i.

(Refer Slide Time: 31:14)



Δ i which was conditional expectation of Z given X i - conditional expectation of Z given X i - 1; can be written as conditional expectation of Z given X i - the conditional expectation of Z given everything, but the ith part. So, I will abbreviate this / - i ok, I will I will write it this way, given X i. That is what the previous claim was showing, where this guy is everything, but the ith part; X 1 to X i - 1 and X i + 1 to X n ok, that is what this vector is.

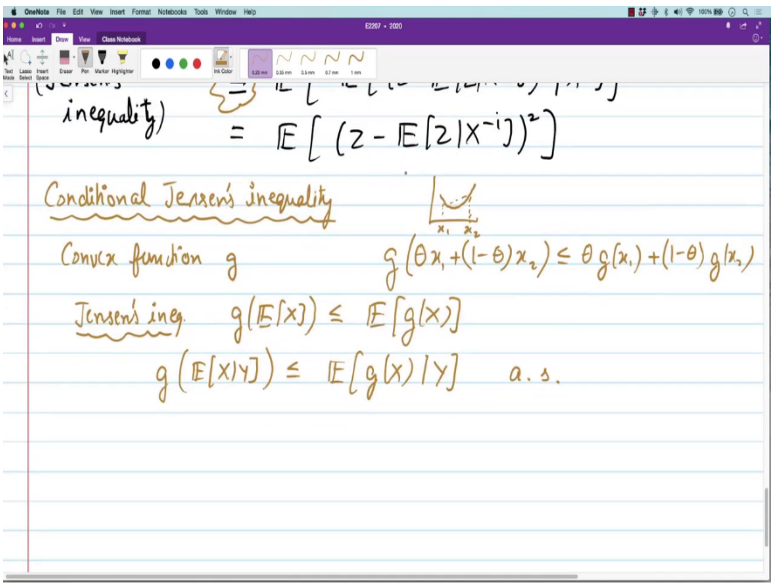
And this is just this claim, this is just this claim above that these two are equal. And so, the interesting thing about this is that, the outer expectation is same for both of them, the first term and the second term. So, this can be written as the expected value of Z - the conditional expectation of Z given X - i given X i, that is what δ i is.

Therefore, the expected value of δ i square is = the expected value of this conditional expectation square; conditional expectation of Z given Z - given X - i and then outside X i whole square.

Now, we use another inequality which is a which is also very useful, I will review it and this is what is called Jensen's inequality, ok. Actually in this case we do not know we do not need the most general one; but what this inequality says is that, the expected value of a convex function like square is less than = the. So, this is the expected value square, this is less than = the expected value of the square, that is what Jensen's inequality says.

So, maybe I will note it down later; but this is / Jensen's inequality, it is in fact the conditional version of Jensen's inequality that, this is less than = expected value of. So, I will take the square inside ok and now this is the expected value of expected value. So, that is the formula we have seen earlier. So, this is just a single expectation of Z - conditional expectation of Z - X - i whole square, ok.

(Refer Slide Time: 34:41)



So, that is an important observation and here in this inequality here, we have used the conditional Jensen's inequality; let me quickly review this inequality. So, this applies for any let us say convex function. So, convex function let us say g. And what is a convex function?

So, here we have g of theta 1 x 1, let us say theta x 1, theta is something between 0 and 1 + 1 - theta x 2. So, you have two points x 1 and x 2.

And what is this? This thing is a straight line joining them. And a convex function is the one for which the value of the function at any point on this line is below the average of the values, ok. This kind of function, the value of the average that is here is below the average of the value, which will be in this line ok, that is a convex function. And if you since the theta is between 0 and 1, you can think of this is g of.

So, this is sometimes called Jensen's inequality, it is almost like the definition of convex functions; it says that g of expected value of X for a convex function is less than = expected value of g of X and its conditional version, instead of expected value has conditional expected value.

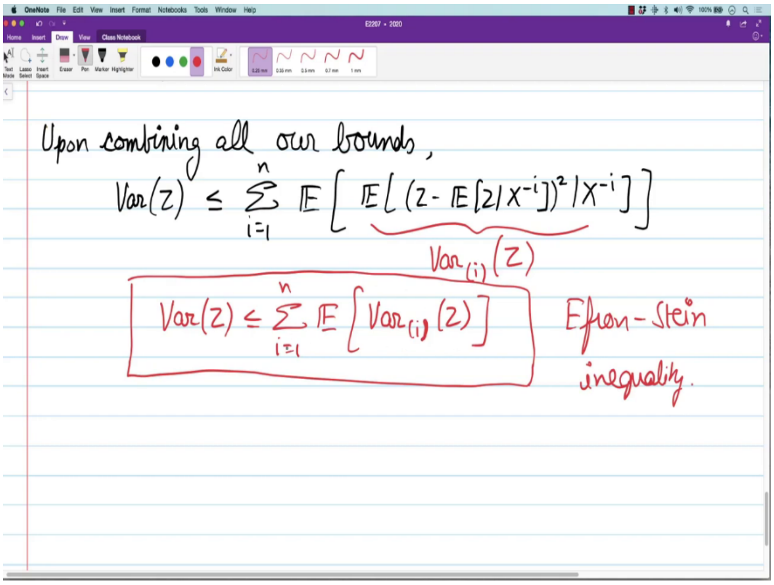
So, g of expected value of X given Y is less than = expected value of g X given Y, ok. And now, this is a random variable; this inequality must hold with some probability and we claim that this holds with probability one. So, it holds almost sure, that is what conditional Jensen's inequality is.

So, the function that we are looking at here, this is the conditional expectation. And the function that we are looking here, looking at here is the square function. And what this says is that, if you take the conditional expectation outside or if you take the square inside, the because square is a convex function, thing can only increase. So, we take the square inside in the conditional expectation outside and it only increases and this is = this, ok.

Another way is that this is the conditional mean square and this is the second moment of this random variable under condition on X i and second moment is greater than the conditional mean square; this is just like non negativity of conditional variance of this random variable, ok. So, now let us quickly see, quickly review what all we have seen. So, we had the Z - E Z and we decompose into this uncorrelated component, 0 mean uncorrelated component δ i.

And therefore, the variance of Z was expect sum of expected value of this second moments, because there was 0 mean. Then after that we noticed that, in fact each of this δ i can be expressed as expected value of some centered random variable, where the with the centering mean is when condition on X - i. So, we use Jensen's and we got this form. So, combining everything what we get is sorry.

(Refer Slide Time: 38:39)



So, upon combining all our bounds, the way conditional Jensen's is the only inequality we have used till now, everything else was exact equality. The variance of not the variance sorry yeah sorry, variance of this random variable Z is less than = summation i = 1 to n expected value of this guy, ok. And in fact, I will do conditioning on X i X - i here and express this expected value in a slightly different way; expected value of Z - conditional expectation of Z given everything else.

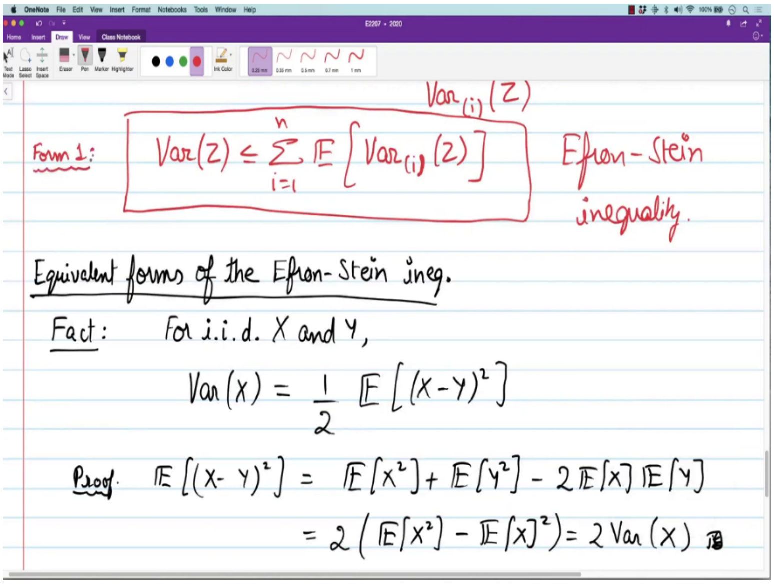
So, we have fixing everything else, but the ith coordinate and taking the square and you take conditional expectation with respect to that ok and then you take another, outside conditional expectation. So, we will abbreviate this guy, this thing here. So, here we are looking at the probability space, we have condition, where we have condition on everything, but the ith guy; we will abbreviate this / this bracket i.

So, this is the variance of this random variable Z; when you have condition on everything, but the ith coordinate and that is what we are denoting / this i. So, it is a random variable here. And what we have shown, this is the first form of Efron Stein inequality that, this variance of random variable Z is less than = summation i = 1 to n expected value of these conditional variances, variance of i given Z, right.

That is roughly that is what our main claim is, this is the Efron Stein inequality that we have shown, ok. Now, note that this is just to reiterate; this is the conditional variance of the random variable Z, when you condition on all, but the ith coordinate. So, it is like the fluctuation in the ith coordinate; you fixed everything else and only allowing ith coordinate to vary.

And when you do that, then you get this conditional variance and this variance is sort of less than = the variance or the fluctuation contributed / the individual coordinate, that is the claim of Efron Stein inequality. Now, what we will do is, we will give provide other equivalent forms of this inequality, ok.

(Refer Slide Time: 41:24)



So, this is the main form, we can still write it for some time and now we will provide some other equivalent form of this inequality; equivalent forms of the Efron Stein inequality. And these equivalent forms are simply obtained / writing equivalent expression for variances.

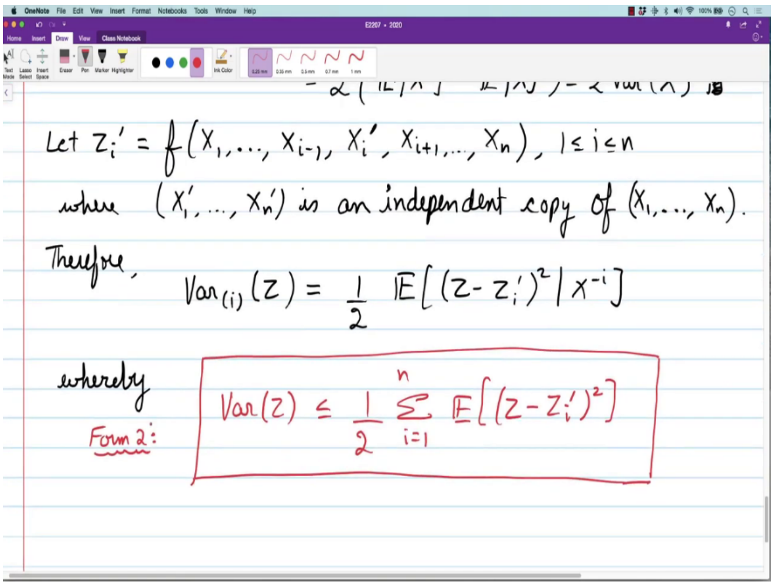
So, first observation, this is just a fact that you can verify; I will give a homework exercise to verify this. If you have a random variable X and a copy Y of it; so, for independent and identically distributed random variables X and Y, variance of X is = half of expected value of X - Y square that is a claim, ok.

So, that is something you can try to show, actually we can just show it; let us just show it ok, let me not leave it as a homework exercise. So, this is proof of this fact, expected value of X - Y square is = expected value of X square + expected value of Y square - 2 times expected value of X - expected value of Y.

So, this is an independent copy, but X and Y are identically distributed. So, this becomes 2 times expected value of X square - expected value of X whole square and this is = 2 times the variance of X, ok.

So, that is a very simple statement. So, if we use this proof here and we will substitute this in this inner variance; what is this inner variance? We have fixed everything all the entries except the ith entry, ok. And now we would like to replace ith entry with its independent copy; how do we do that?

(Refer Slide Time: 44:02)



So, we will have a notation for that, let Z i prime be f of X 1 to X i - 1. So, that those entries are fixed to X 1 to X i - 1, except that the ith entry is replaced / its independent copy. So, all the other entries of this random variable these of this form, all the other arguments of this functions are fixed as before X 1 to X i - 1, X i 1 + X n, except the ith entry is flipped is replaced with an independent copy. So, where X 1 prime to X n prime is an independent copy of X 1 to X n, ok.

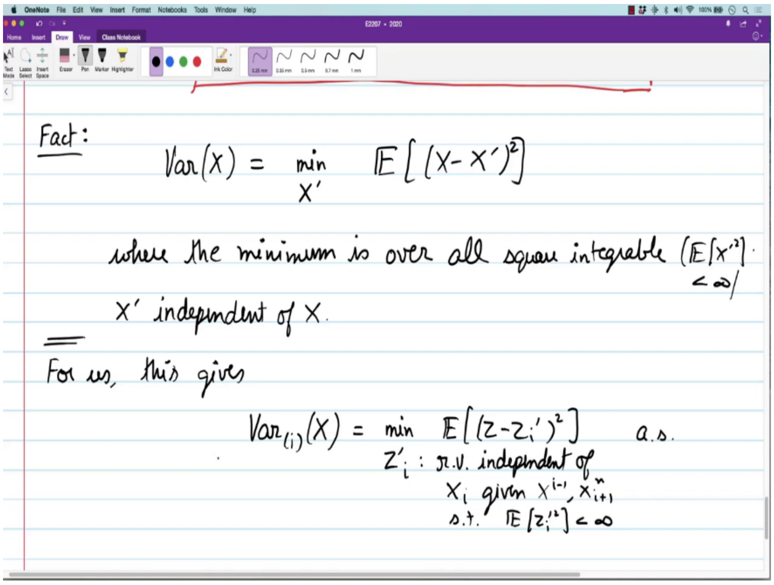
So, then, so the Z i prime is an independent copy of; so Z i prime is an independent copy of Z when you condition on everything else. And / the previous fact, if you look at the variance given i of Z, this is exactly = half into the expected value of Z - Z i prime square; of course given all, but the ith entries X - i, ok. These two are equal, this is just / this fact here.

And when we plug this into a first form of Efron Stein, we get the second form of Efron Stein inequality, which says that variance of Z is less than = half summation i = 1 to n expected value of Z - Z i prime whole square, looks pretty neat.

So, this different forms may have different applications, where remember Z i prime is obtained / replacing, Z i prime is just like Z, except that the arguments are the ith coordinate of the argument X 1 to X n is replaced with its independent copy X i prime, that is what the Z i prime is, ok.

That is the first, that is the first equivalent, that is the first equivalent form; well let me number them. So, this is form 1 of Efron Stein inequality and this is form 2. And now finally, I will give you another form; this that the final form will again be obtained / using an alternative expression for variance; I will also put this as a fact, many of you may already be aware of this fact.

(Refer Slide Time: 47:18)



So, suppose you have this random variable X, variance of X is something you are looking at; this is actually the minimum mean square error for all square integrable function. So, this variance of X is minimum over all X prime of expected value of X - X prime ok, where the minimum is over all square integrable, basically the ones with finite second moment, X prime independent of X.

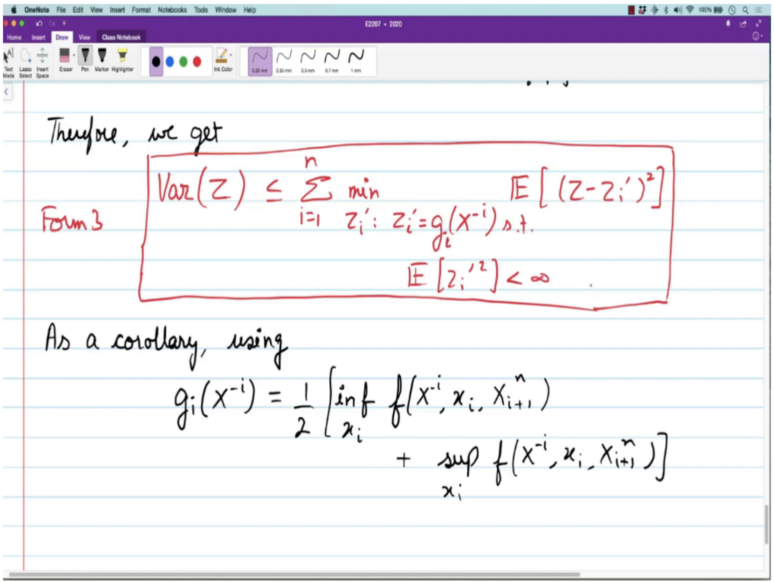
So, you can take a random variable independent of X and take minimum of this mean square error; because we are looking for an independent random variable that best approximates X.

In fact, this is obtained / the mean of X, a constant random variable is the best one here; that is what that is a result you can show, in fact, you can show it / just differentiating if you like. But there are many ways of showing this actually, a more formal proof you will do some completion of square, but this is something you may be aware of.

For us we are looking for the random variable X we are interested in is this guy, which is this random variable Z condition on all the other X i - 1. So, there what you have this X prime, it is not independent it is a random function of X 1 to X i - 1 and X i + 1 to X n and you minimize this guy over all such random functions. So, for us this gives, this is some fact that we recall; for us this gives variance over i of X, this is given all the other guys, ok. So, this is a randomized value when everything else is fixed, right.

And therefore, this is minimum over all Z prime i random variable independent of X i given everything else, given X i - 1 and X i + 1 to n that is what is fixed here ok, such that expected value of Z i prime square is finite, right. So, this is true almost surely; because yeah this is what I mean / almost surely is overall realization of X i - 1 and X i + 1 to n. And we have an expectation outside. So, we can just view this Z i prime as a randomized function of these guys.

(Refer Slide Time: 51:07)



And we can write the third form variance of a random variable Z is less than = summation i = 1 to n minimum over all functions Z i prime, where Z i prime is = g of X - i.

Some function, in fact a randomized function is also allowed, such that expected value of Z i prime square is finite ok, minimum over all such functions, expected value and this function this can depend on i's I am putting an i here of Z - Z i prime whole square. I am using a similar notation as this case here; here there is the half, there is no half here ok, that is the third form.

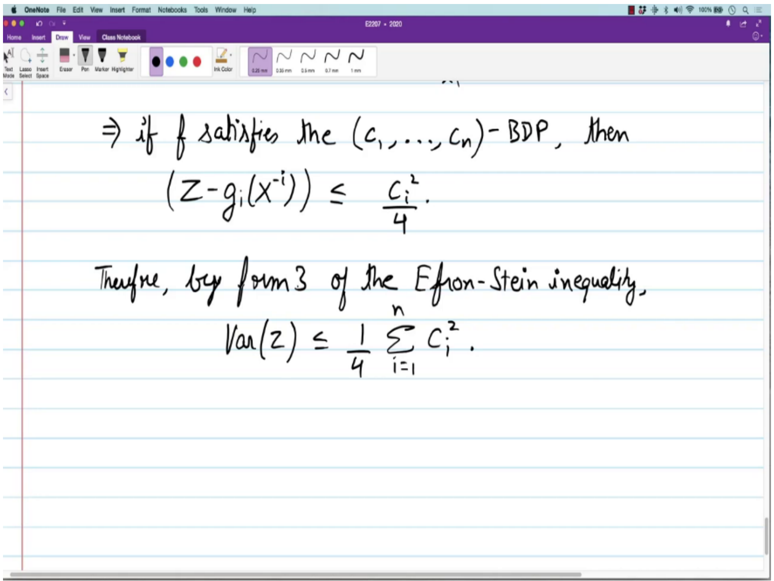
This third form is also very handy, because this is less than the minimum one; you can substitute your favorite function and get another form. These three forms are equivalent, none is stronger than the other; but let us just weaken this third form and note an important corollary as a corollary using. So, I can define any function here g i of X - i is =.

We can we would like to define a function here, so that for each coordinate, for each coordinate we get that value C i that we had seen earlier, that is what we would like to do. So, what is that function Z i? How would we like to define that function Z i?

So, that function can be defined as half of in over, it is the average of max and min over this coordinate; let us say X I, now you have f of you fix everything else X - i take min in forward this guy X i + 1 to n + sup over x i f of X - i x i X i + 1 to n, ok.

So, you can define this function this way and since this is min or inf over all such function, this function is a specific choice and I should I should have written inf here, I am being a little bit casual / saying min, that is alright. So, if you look at this function, what is Z - this function Z i prime square?

(Refer Slide Time: 54:54)



So, it is subtracting the average value from this guy and therefore, it cannot be less than it cannot be more than this implies; if f satisfies the c 1, c 2, c n bounded difference property, then Z - g i X - 1 X - i is less than = c i square / 4, ok. You can check that, it is the min of; it is the average value that you are subtracting of the two extremes and therefore, this is less than = this.

Therefore, / form 3 of the Efron Stein inequality, so this is the form 3, I have put in a specific function this one; it has the power of this method, you can put any specific function. / the way this is the means minimum mean square error of estimating Z / looking at all the other coordinate; that is what this guy is, / looking at all the other coordinates of X i.

So, looking at X - i, what is the minimum mean square error in estimating, minimum mean squared error / in estimating Z. So, now yeah, so we have this bound, if the function satisfies the bounded difference property and therefore, / form 3 variance of Z is less than = 1 / 4 summation i = 1 to n c i square, which is very much like what we have seen in the McDiarmid inequality, ok.

So, this Efron Stein to conclude this Efron Stein inequality gives a similar bound for the variance of Z itself, instead of the variance factor ok. This is what I wanted to say in this lecture; in the next lecture, I will see an application of this Efron Stein inequality to get concentration bounds for self, for this function satisfying boundary difference property. Just like McDiarmid, we can get McDiarmid concentration bound.

But before I close, just a quick review, we saw three different forms of Efron Stein inequality; the first one is this one, where we have decomposed this variance into variance along individual coordinate when you change, when you fix all the other coordinates, but only allow individual ith coordinate to move, that is this variance i.

Then equivalently we saw that, we can write an independent copy Z i prime obtained / replacing the ith coordinate of X ith coordinate of f, of domain of f of input of f with its independent copy and that is what you get here.

And then we use the MMSE expression minimum mean square error expression for variance to get another form, where the minimum is over all functions of the coordinates X - i. And you are trying to estimate the Z and the mean square error; this is also equivalent to the previous form, all these three forms are equivalent, ok. See you in the next lecture.