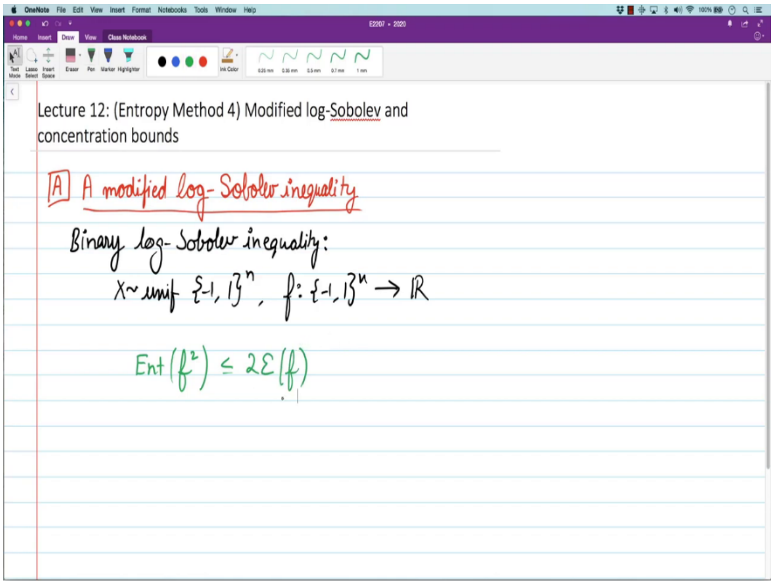
Concentration Inequalities
Prof. Aditya Gopalan
Prof. Himanshu Tyagi
Department of Electrical Communication Engineering
Indian Institute of Science, Bengaluru

Lecture - 13
A modified log-Sobolev inequality and concentration

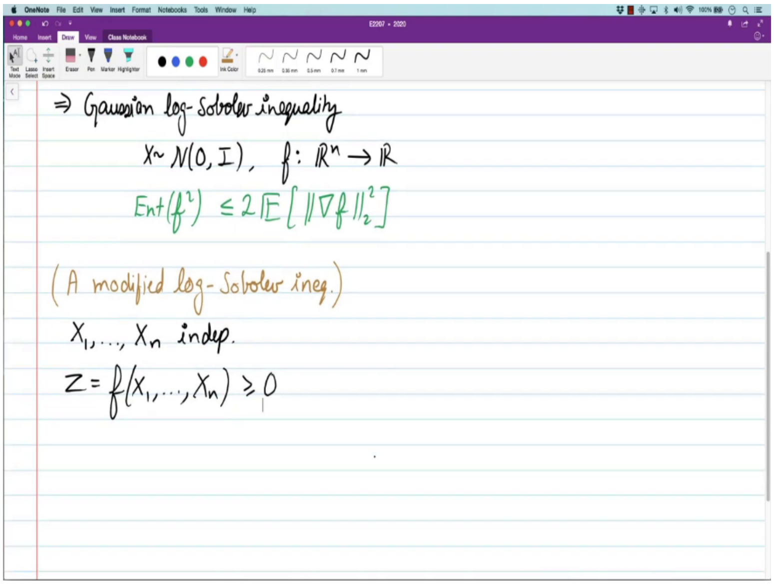
(Refer Slide Time: 00:21)



Earlier we saw the binary log Sobolev inequality log Sobolev inequality which said the following. Let X be distributed uniformly over the Boolean hypercube is the sine version of it and we looked at this function f from so 1, 1 n to R let us say for non-negative function or let us say positive function. Then actually just R and then what we showed was that the entropy of f^2 is ≤2 times is e of f which was the sort of which was what we called the Efron-Stein variance estimate for f ok.

This was the binary log Sobolev inequality we saw.

(Refer Slide Time: 01:24)

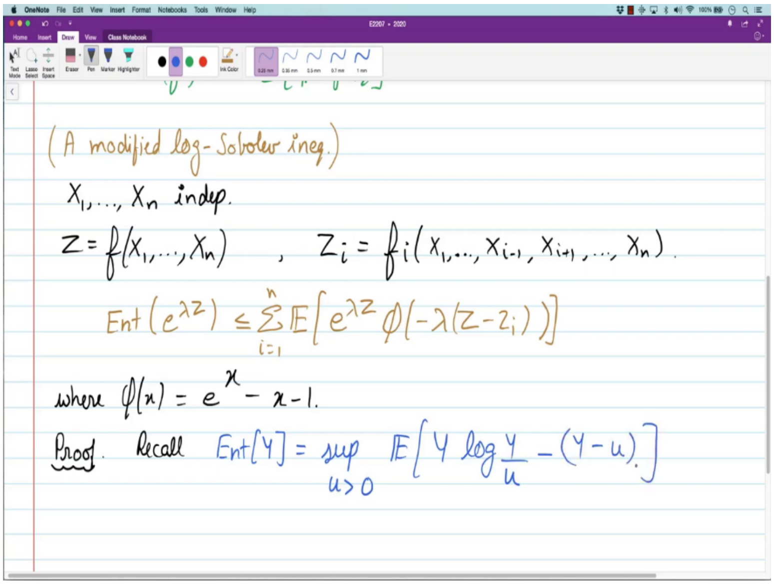


And, this implies we saw that this gives the Gaussian log-Sobolev inequality which again now it looked at X being Gaussian with mean 0 and identity variance and again f was this function from R n to R. So, this inequality said that the entropy of f^2 is ≤expected value of gradient of f^2 times that and so, these two inequalities we had derived earlier. And, what we came was that once you have these inequalities you can establish sub Gaussian concentration bounds.

So, using these two inequalities we were able to establish sub Gaussian concentration bounds for in this case let us say Lipschitz function of Gaussian random variables and here for functions for which this the Efron-Stein the Efron-Stein variance bound variance estimate was bounded. That is what we were able to do. Our goal now is to extend these bound and subsequently the concentration bounds for of similar form to general random variables that is what we want to do.

So, in the first part of this lecture we will derive a modified log Sobolev inequality which extends both of them to general random variables. So, here is our modified log Sobolev inequality. So, now, once again we have this random variables X 1 to X n that are independent and we have function f from these n random variables we will just write it as f of X 1 to X n this is a say a non negative function.

(Refer Slide Time: 04:12)



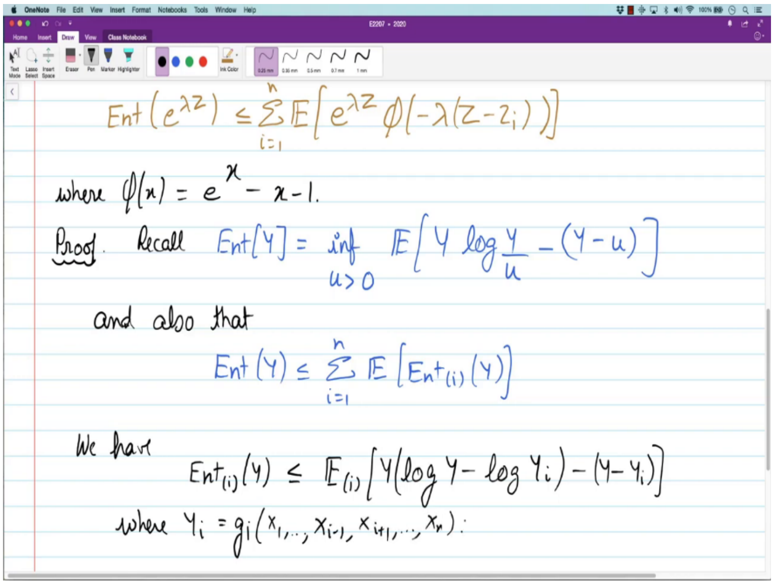
And, we define another random variable Z i which is just which is some function let us say f i of X 1 to X i - 1 X i + 1 to X n can be any function, ok and the log Sobolev inequality the modified log Sobolev inequality that we are after says that the entropy of e to the power λ f yeah, actually it is $e^{\lambda f}$ is there. So, we do not really need this to be non negative.

So, entropy of $e^{\lambda f}$ is < = expected value of e to the power λ f entropy of λ Z is < = expected value of $e^{\lambda Z}$ φof - λ to Z - Z i Σ i = 1 to n, ok. So, this is the modified log Sobolev inequality. This looks this looks a bit like the log Sobolev inequality, but we are applying it to, but we are applying it to this function $e^{\lambda Z}$.

And, in fact, if you remember the proofs of concentration bounds from this from both these binary and Gaussian log Sobolev inequality we did apply them to the function f that was given by e^{λ} f λ Z. So, in that sense this inequality directly gets us there and hopefully will serve that purpose, ok. So, that is the main inequality we will show where I never define this function φ where φ of x = e to the power x - x - 1 ok.

So, let us prove this inequality; λ non-negative λ > 0 actually. So, the proof relies on the formula for with the formula the variational formula for entropy that we saw in the previous class. So, recall that the entropy of a non-negative random variable Y can be expressed as sup over say u that is > 0 expected value of Y log Y / u - Y - u. So, this is the formula we saw in the last class.

(Refer Slide Time: 07:20)

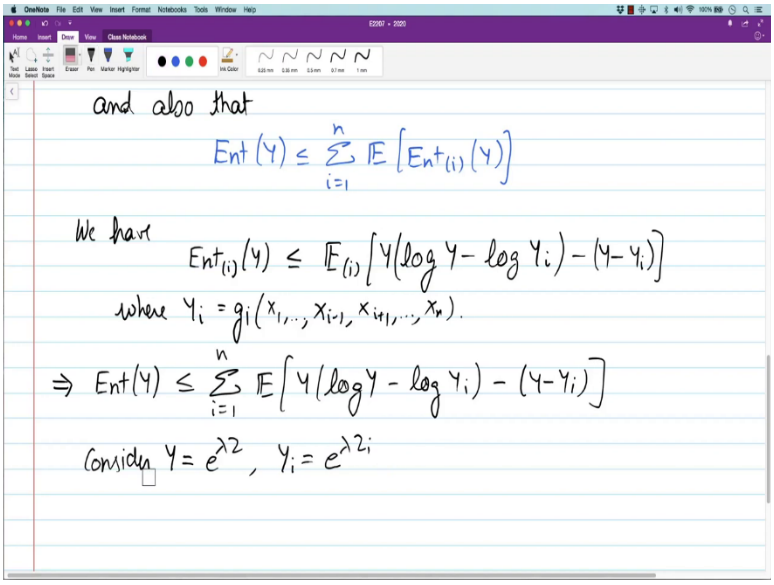


Also and also this tensorization formula for entropy, it said that entropy of a random variable y which is a function of X 1 to X n is ≤expected entropy all give, but the all, but the i-th random variable of Y, ok. So, we so, we apply this formula to this condition entropy. We have this is a general recipe for proof of all the log Sobolev inequality where we reduce the bound for N-dimensional case to the bound for one dimension case. So, we are doing the same thing here.

We will derive a bound for entropy i here. So, in this in this entropy everything, but the i-th random variable X i is fixed. So, this expectation is over just the random variable X i ok and this is = or let us say this is sorry, this is inf, also it is important, right. So, this guy is < or = I can choose any u here I will choose a particular u this expectation is with respect to everything, but the with respect to the i-th random variable. So, everything else is fixed.

And, I will fix u to some function Y i of all the other random variables ok where Y i is some appropriate function g i of all the other random variables, ok. So, we have this formula.

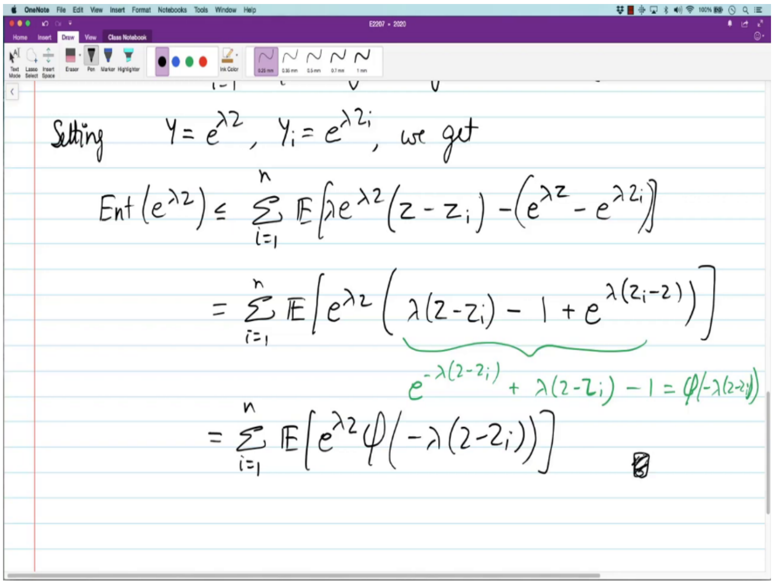
(Refer Slide Time: 09:26)



Now, so, this is true for any random variable Y. So, this implies entropy of Y is < = Σ i = 1 to n expected value of Y log Y because this is expected value of this expected value. So, that becomes expected value of Y log Y - log of Y i - Y - Y i, ok. So, this is true for all Y. So, considering Y equals to e $^{\lambda Z}$ with. So, this Y is the function of X 1 to X n.

So, this Z is also a function of X and this Z is also a function of X 1 to X n. So, this is allowed and Y i is e $^{\lambda Z i}$ we can do that, ok. So, setting these two values setting we get.

(Refer Slide Time: 10:23)



So, we just substitute these two values entropy of e to the power λ Z is < = Σ i = 1 to n expected value of Y. So, e to the power λ Z into log Y that is this λ comes out that is λ Z - Z i -, yeah ok. It is better to keep this λ inside λ this - e to the power λ Z - e to the power λ Z i ok just substitution.

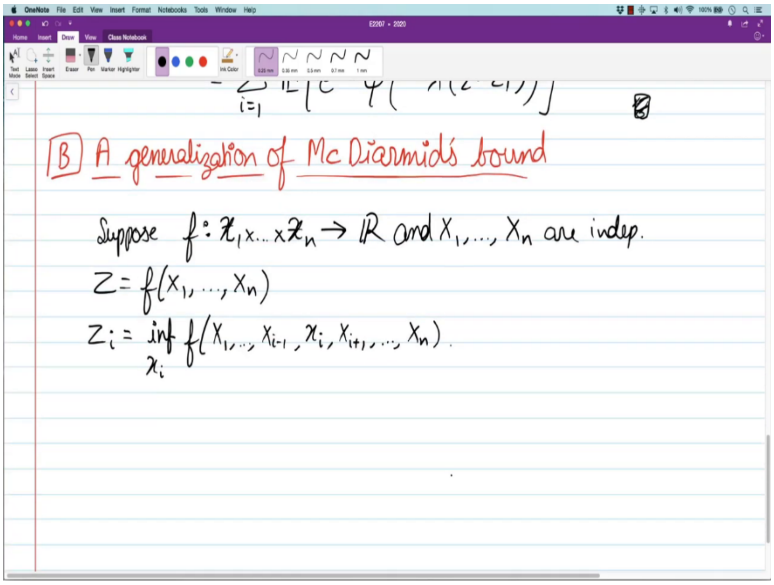
So, this is exactly = Σ i = 1 to n expected value we take e to the power λ Z outside and what we get inside is λ Z - Z i - 1 + e to the power λ Z i - Z, ok. And, so, if you look at this function here this is e $^{-\lambda Z - Z i - \lambda Z - Z i - 1}$. So, which is just φ of - λ Z - Z i ok, the function you are looking for ok.

So, we have shown that this thing is nothing, but Σ i = 1 to n expected value of e to the power λ Z φ of - λ Z - Z i. So, this general bound that is what we were after is called the this one here. So, this is called the modified log Sobolev inequality and this function φ it has some approximation properties which we will use to obtain concentration bounds, ok.

So, this well know, this is the modified log Sobolev inequality and in proving this we again use tensorization property and then we proved an inequality an elementary inequality. This so called let us say a sort of elementary inequality here that we applied. This is that variational formula for entropy that we have seen earlier, ok.

So, now that we have our modified log Sobolev inequality, we move to the second part of the lecture where we use it to derive a concentration bound ok.

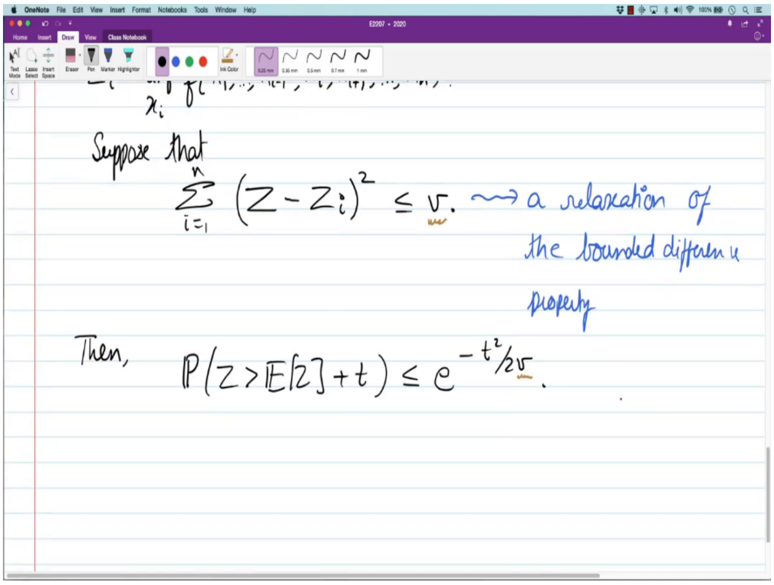
(Refer Slide Time: 14:00)



So, that is the second part of the lecture. So, generalization of McDiarmid's bound ok right. So, before we proceed we note a property of this function φ that will be used in this form ok, but maybe we state the bound. So, suppose f is a function from let us say need not be IAD. So, let us I will say X 1 to X n to r and X 1 to X n are independent ok.

And, then let us say this so, we denote this random value of the function by Z and this Z i can be any function of our choice. So, we choose this guy to be inf over X i f of X 1 to X i - 1 X i X i + 1 to X n ok, that is what this guy is ok.

(Refer Slide Time: 16:08)



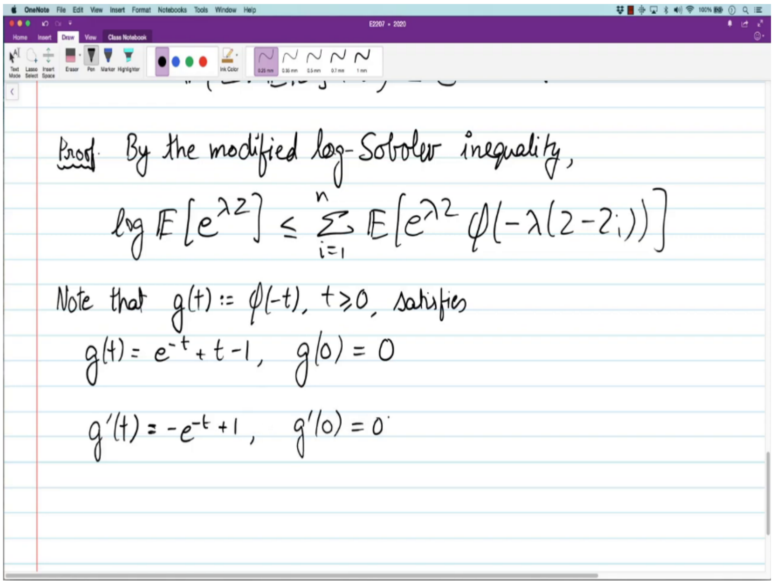
So, suppose we assume now, suppose that this - this Z i ^{2} Σ i = 1 to n. So, this is non-negative ok. Suppose this is < = v. So, as I have remarked earlier this is sort of a generalization of the bounded difference property. Earlier in the binary log Sobolev part and the Gaussian part we were taking we are taking Z i prime Z i prime which was an independent copy where we had replaced X i with its independent copy.

And, now we have taken inf over all X i, this is slightly different form for what from what we saw for the binary case and the Gaussian case, but that is what we assume. It is still a condition which relaxes. So, this is a relaxation of the bounded difference property ok because if you go by bounded difference property you need a bound for individual terms.

So, you need sort of a bound on L infinity norm of this vector, but now you need a bound only on the L 2 norm. So, suppose this is true then, the probability that set X Z is expected value by t is exponentially small, ok and this v this variance vector v is the same as this variance vector v ok. So, this is a sub Gaussian concentration bound the one we have been after throughout the throughout our treatment of entropy method

So, this can be one concrete attainment of a complete inequality which you obtained using the entropy method. You were able to generalize McDiarmid just this assumption instead of requiring assumption on each and every term now we only have an assumption on a total difference and we can establish McDiarmid's inequality. So, how do we show this? Ok.

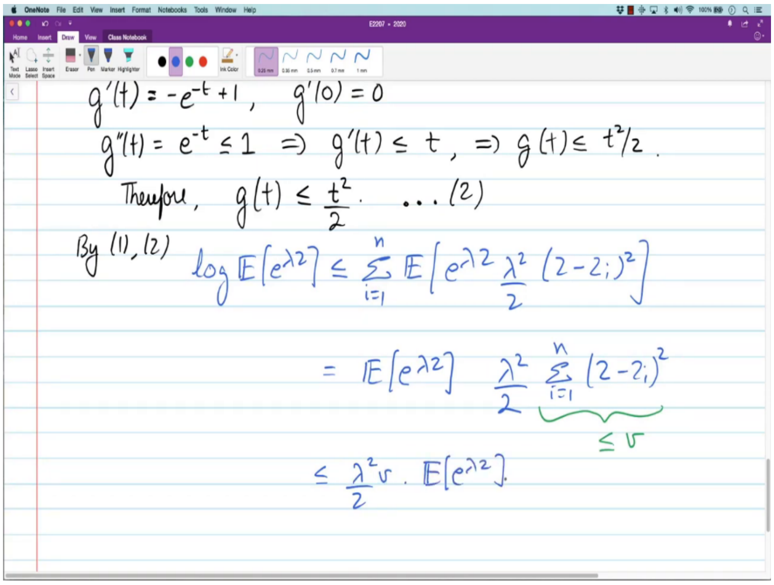
(Refer Slide Time: 18:28)



So, this is the main thing. Proof so, we start with this modified log Sobolev inequality log expected value of e to the power. I am sorry yeah that is correct λ Z, Z is < = Σ i = 1 to n expected value e to the power λ Z φ - λ Z - Z i. If this holds because Z i here is a function of all, but all, but the i-th coordinate ok. So, if you fix X 1 to X i - 1 and X i + 1 to X n you can fix you can view the Z i as a function of these guys, ok.

So, this is true. Now, note that the function g t defined as φ of - t for t > = 0 satisfies. So, let us see. So, this g t is e to the power - t + t - 1. So, g of 0 is 1 - 1 and 0. So, 0 g prime of t is - e to the power - t + 1. So, g prime of 0 will also 0.

(Refer Slide Time: 20:18)



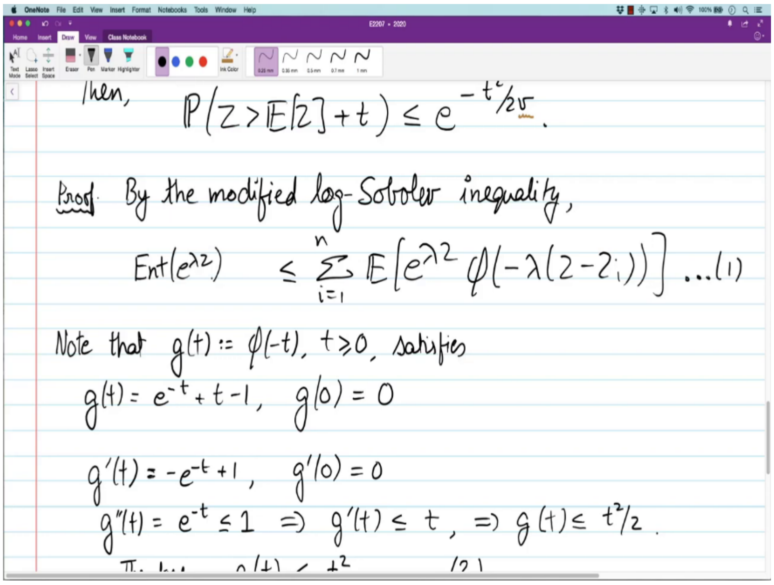
And, second derivative is e to the power - t ok. So, that is prime of t and therefore, if you want this is ≤let us just keep it this way. So, before g t is ≤So, this guy is < = 1 / is≤ 1. So, g t is ≤ $e^{2}/2$ so, right. So, its derivative is always ≤t and then this implies g t is ≤$t^2/2$, that is it ok.

So, we have this bound and we so, this is the first observation, this is the modified log Sobolev inequality. This is some elementary bound for this function earlier it was not clear why do we have this function but, now we have this nice property for the φ function.

So, combining 1 and 2, we get that the log expected value of $e^{\lambda Z}$ is ≤ Σ i = 1 to n expected value $e^{\lambda Z}$. Now, we substitute this bound $\lambda^2 / 2$ into Z - Z i^2 very nice.

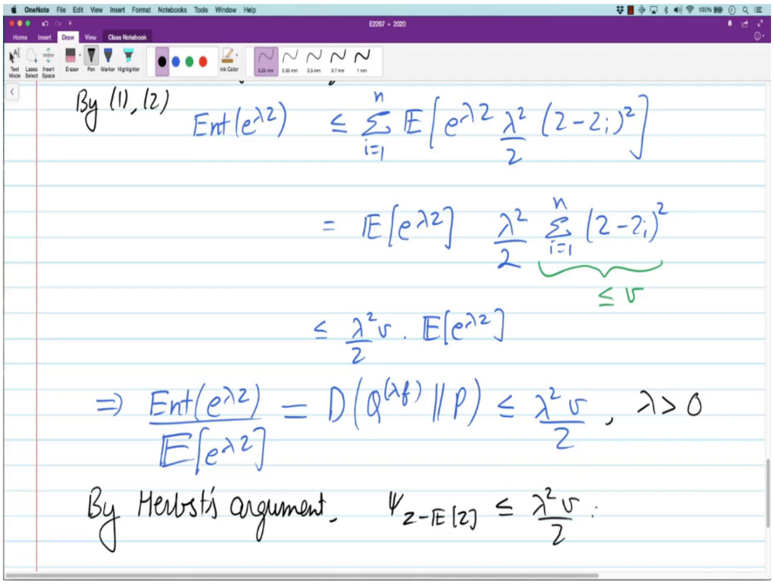
So, this expected value $e^{\lambda Z}$ comes out and we have $\lambda^2/2$ Σ i = 1 to n Z - Z i^2 the quantity that we have assumed is bounded by v ok. So, therefore, this guy here is ≤ λ^2 v/2 expected value $e^{\lambda Z}$, alright. So, now, we have our familiar ok this sorry.

(Refer Slide Time: 23:05)



This is the bound for the entropy of e $^{\lambda Z}$ that is what that is what the that is the bound that we derived and so and so, this is also the entropy of e $^{\lambda Z}$.

(Refer Slide Time: 23:19)

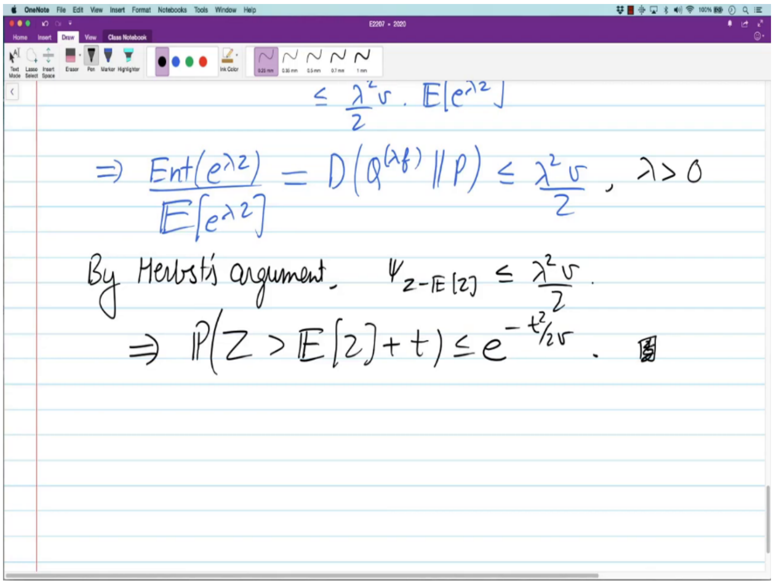


That is what we need for entropy method we need a bound for the entropy of e $^{\lambda Z}$ ok. So, what we have is that the entropy of e $^{\lambda Z}$ /expected value e $^{\lambda Z}$.

Remember that this is the; this is what this is the divergence that we were looking for this is that divergence Q λ f from p and we are showing that this is $\leq \lambda^2$ v/ 2. So, this by Herbst's argument in fact, you just have to remember that formula for log moment generating function is the integral of this guy /λ^2.

This is true for all $\lambda > 0$, the log moment generating function of this random variable Z, yeah. So, remember that you can always subtract expected value of Z from both the exponents because it is like multiplying with a constant. So, this guy here is $\leq \lambda^2$ v / 2.

(Refer Slide Time: 24:31)



And, this implies that this is a sub Gaussian random variable. So, we have a sub Gaussian tail form probability that a random variable that exceeds its expected value + t is $\leq$e t^2 /2 v, ok, alright.

So, this concludes the past few weeks of lectures for the past few lectures and so, what we have achieved at the end of this is this very nice generalization of the McDiarmid inequality and note that I have provided it as one sided tail bound the upper tail bond; if you want the lower tail bound you have to apply this to - Z.

If you apply this condition to - Z this becomes Z - the sup over X i Z i Z i sup over X i that is the Z i and then you will have the if suppose that guy is bounded then you will have the lower tail bound, ok.

So, yeah and in fact, if sup over X i - inf over X i if that thing itself is bounded squared if that thing itself is bounded then you have both side tail bounds ok. That is roughly the part. Now, to the main important thing here is this whole recipe ok. So, we observe this very nice formula for we observe this very nice formula for the log moment generating function that is where whole this entire thing started with this particular formula.

And, then it follows that if you bound this quantity by a constant then that gives a sub Gaussian bound for the log moment generating function and more importantly perhaps bounding this quantity is it can be handled in a nice way because this quantity tensorizes, ok. This quantity in particular gets related to entropy and that entropy tensorizes.

And, therefore, since entropy tensorizes whenever we want to prove a bound for entropy we can prove bound on this individual components separately and we did prove one such bound using a variational formula for entropy which we saw in the middle. And, then that is you obtain this nice inequality which is called the modified log Sobolev inequality.

And, this function φ that we obtained in this application we were only looking at one sided tail bounds under this assumption and therefore, we could for that direction we use an approximation for this φ; where is that approximation? Yeah, this guy here and therefore, this inequality gets Z - Z $i^2 \lambda^2$ and then this tensorizes.

So, this gets bounded by λ^2 v and then you have bounded this ratio effectively, and then using the formula for log moment generating function we get the sub Gaussainity bound for the log moment generating function which in turn implies a sub Gaussian concentration bound sub Gaussian tail bounds ok. So, this is the entropy of entropy method and we have illustrated this using this binary log Sobolev inequality, Gaussian log Sobolev inequality and now finally, for general random variables ok.

So, this concludes our discussion of the entropy method. In the next lecture Aditya will start with the with the transportation method which is the second major approach for proving concentration bound that we will see in this course. And, just like this entire entropy method was based on this particular formula as I have mentioned several times earlier the transportation method will be based on this formula, the Gibb's variation formula.

And, there we will use we will use the fact that that this difference here can be controlled using a transportation inequality in terms of divergence and that will yield another upper bound for the log moment generating function leading to sub Gaussian concentration bounds.

So, see you in the next lecture.