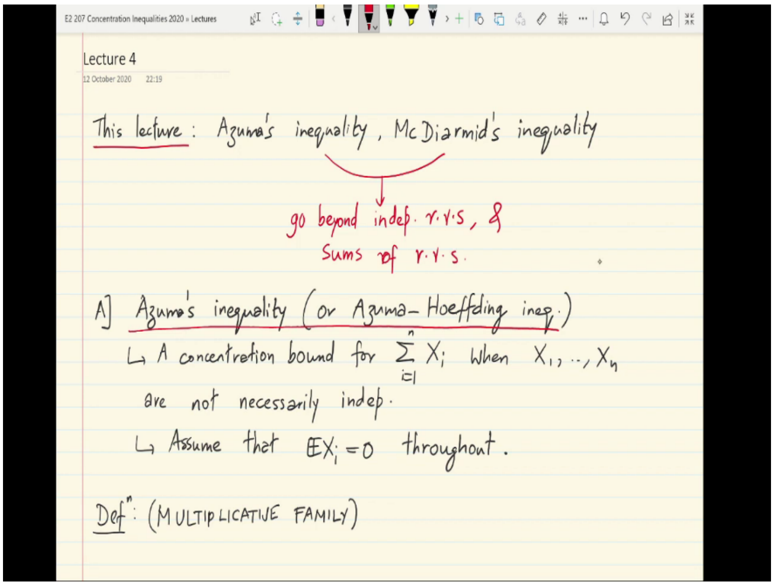
Concentration Inequalities
Prof. Aditya Gopalan
Prof. Himanshu Tyagi
Department of Electrical Communication Engineering
Indian Institute of Science, Bengaluru

Lecture - 05
Azuma and McDiarmid inequalities

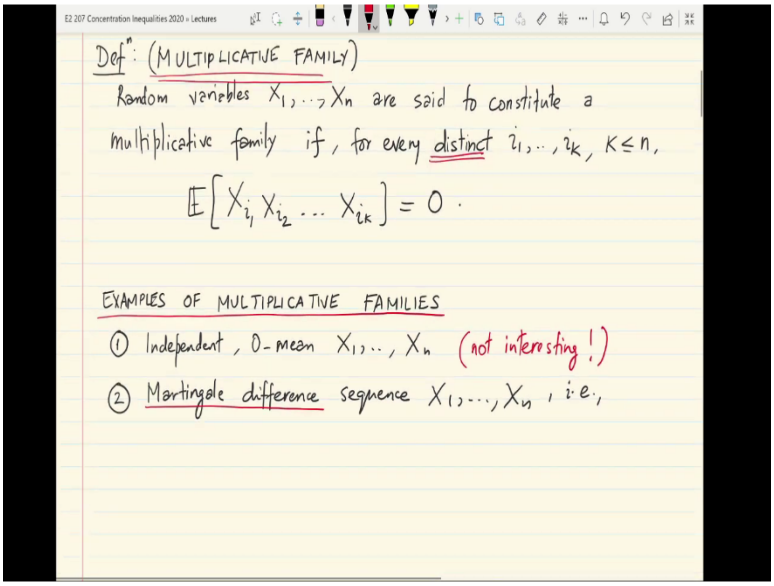
(Refer Slide Time: 00:01)



Hi all. In this lecture we will prove two very useful Concentration Inequalities that help us go beyond independent random variables. So, this will help us go beyond independent random variables and functions that are just the Σ of individual random variables ok. So, we will go beyond we will be able to go beyond just Σ ming random variables and controlling the deviations ok. These are called Azuma's inequality and MacDiarmid's inequality. Azuma's inequality is also called the Azuma Hoeffding inequality ok.

So, let us dive into the first part which is to derive Azuma's inequality or the Azuma Hoeffding inequality. So, Azuma's inequality is essentially a concentration inequality for the Σ of random variables X i which are not necessarily independent. Recall that we have already shown how we can control the deviations of a Σ of such random variables when each of them is asΣ ed to be independent of the others. We will asΣ e that all the random variables are centered throughout that is the there zero mean.

(Refer Slide Time: 01:26)



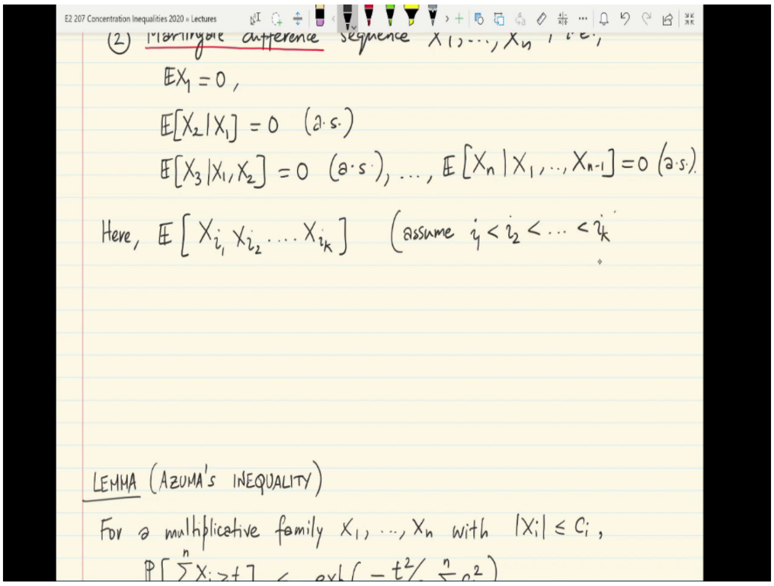
So, its useful to first define a property of these random variables which is what is called a multiplicative family. So, a collection. So, random variables X_1 through X_n are set to constitute a multiplicative family if the following condition holds ok. So, if for every distinct collection of indices i_1 through i_K where K is at most n. So, distinct is useful here is important here. So, think of a bunch of distinct indices i_1 through i_K each index lying between 1 and n.

We have that the expected value of the product of the random variables described / these indices. So, Xi_1 into Xi_2 all the way up to X_i k is 0 ok. So, you take any k any subset of random variables from this bag of n random variables and if you × to them together and take their expectation then the expectation is 0 ok.

So, this; obviously, means that each of them individually is 0 mean ok as the special case. So, this is what it means for a bunch of random variables to form a multiplicative family. Now you may ask what are why are multiplicative families reasonable to consider do they occur in common settings. So, its not hard to convince yourself that if of course, each exercise itself independent of the others and zero mean then trivially you have a multiplicative family, but this is not really the interesting setting to which Azuma's inequality will apply.

So, more non trivial example of a multiplicative family is what is called a martingale difference sequence. So, in order to define this. So, let us consider random variables X_1 through X_n with the following properties.

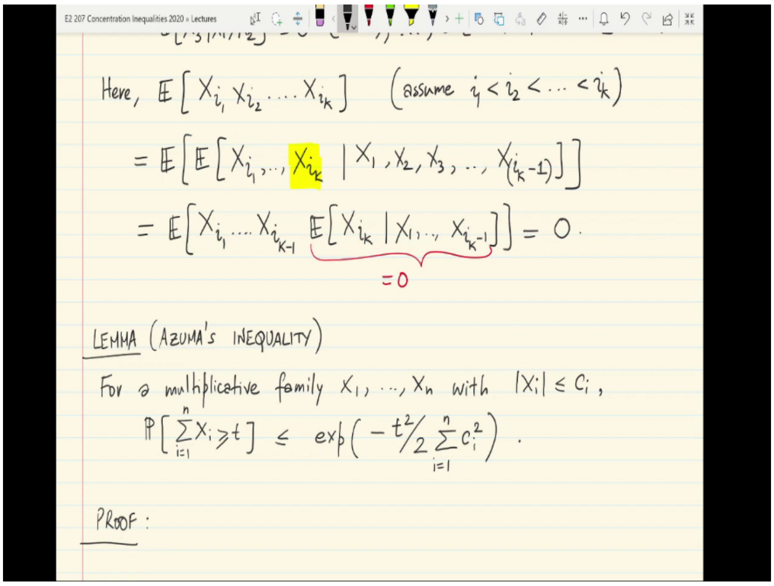
(Refer Slide Time: 03:47)



So, all of them have infinite mean firstly, and let us say expected value of X_1 is 0 the first one has 0 expectation ok the second one given the first as 0 expectation. This is as a random variable almost surely let see the third one. So, every random variable given the previous random variables has mean 0 ok conditional expectation 0 and so, on until you reach expected value of X_n given the first n - 1 random variables is 0 almost surely ok.

So, if you have such a structure such a conditional independence structure, then its we can easily compute explicitly that the expected value of. Let us take any sub collection of k of these n random variables Xi_1, Xi_2 and this is the product up to X i k ok. So, let us asΣ e also that these are ordered that is i 1 is less than i_2 less than i_k ok.

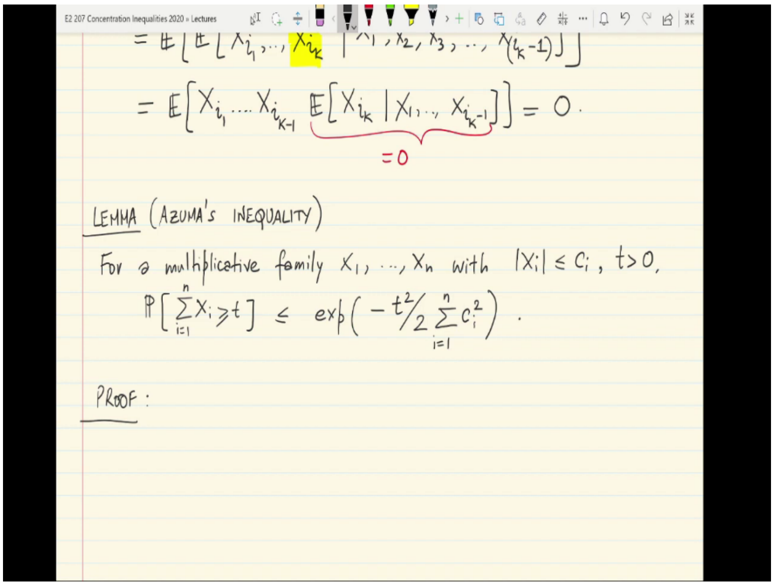
(Refer Slide Time: 05:13)



So, if you wanted to evaluate such a the expected value of such a product, here is what you could do. So, by the iterated property of the expectation let us condition inside on all, but the first all the first i_{k-1} random variable. So, X_{i_1} up to X_{i_k} condition on X_1, X_2, X_3 all the way up to $X_{i_{k-1}}$ ok and you can easily see that this inner expectation. So, the only random variable in this product X_{i_1} through X_{i_k} that is not present in the conditioning is this random variable X_{i_k} the last one.

Everyone else is actually being condition inside ok. So, this is i_{k-1} ok and so, you get the X_{i_1} product up to $X_{i_{k-1}}$ and then you have the expect what is remaining is X_{i_k} given the first i_{k-1} and you can easily show by the properties asΣ ed above that this conditional expectation is 0 ok. So, this is = 0 ok.

(Refer Slide Time: 07:31)

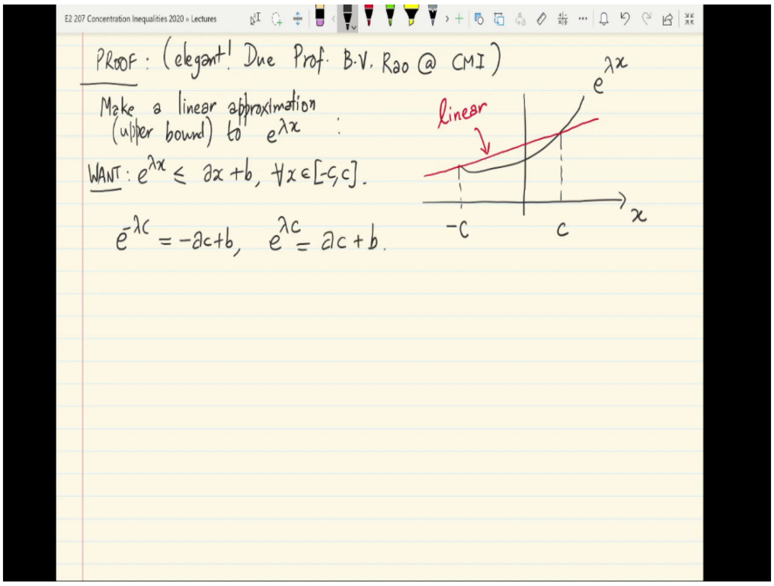


So, with this let us move on to Azuma's inequality. Azuma's inequality says the following. If you have a multiplicative family of random variables X_1 through X_n with each of them being bounded absolutely / the number C i for every I, then the probability that their Σ deviates larger than the positive number t.

So, I should add here that p greater than 0 is at most e raised to $- t^2 / 2 \, \Sigma \, i = 1$ to n c_i^2. So, you essentially get the same type of bound as what Hoeffding's inequality gives you, but for more general families of random variables X_1 through X_n namely random variables that need not only be independent.

So, they could they can be a more general multiplicative family ok. And this is a very powerful result because as we will see it applies very nicely to general processes stochastic processes called martingales. Before we do that let us go ahead and give a very nice and elegant proof of this of Azuma's inequality.

(Refer Slide Time: 08:13)

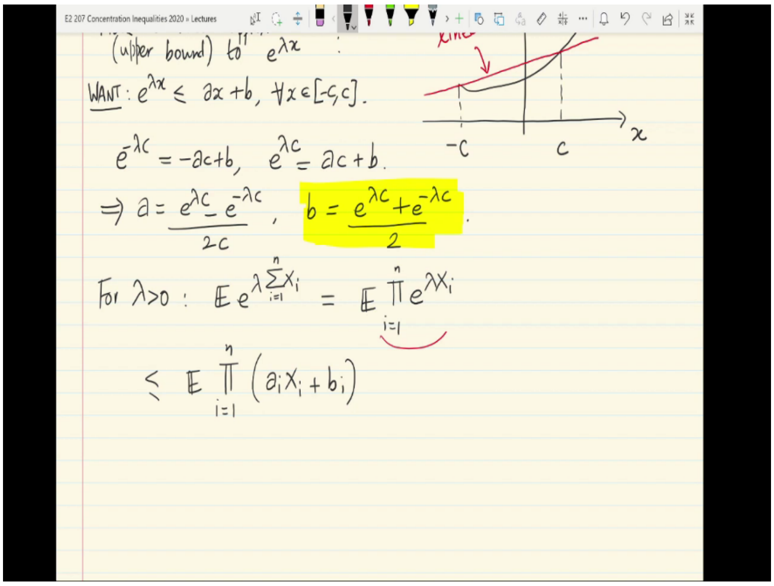


So, I must mentioned here that the first time I saw this particular proof very beautiful proof was after watching Professor B.V. Rao from the Chennai Mathematical Institute derive this inequality ok in a workshop on concentration inequalities. So, the basic idea of this proof is as follows, you want to basically make a linear approximation to the exponential function ok. In fact, an upper bound ok to the function e to the λ x ok. So, let me draw a diagram here.

So, let say that you have x here and you have the function e to the λ x for λ some positive number and you are interested in bounding this exponential function between the range - c to c and you seek a linear upper bound ok. So, you want to connect these two endpoints with the line which will form our linear upper bound.

So, how can we get this linear function? Which so, we want the following property for this linear function e to the λx is upper bounded / this function a x + b for all x in - c to c ok. So, moments glance should convince you that the way to solve for this is to equate the values at the endpoints. So, $e^{-\lambda c}$ = - a c + b and e to the λ c = a c + b.

(Refer Slide Time: 10:14)

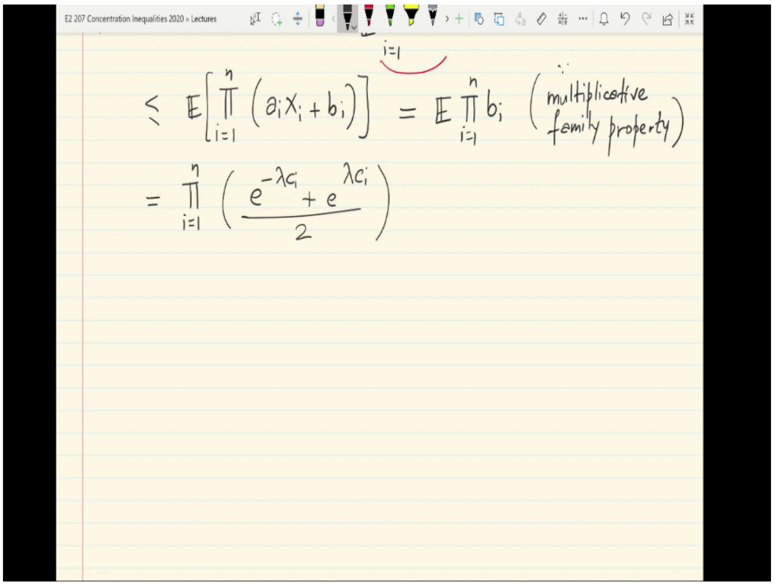


So, this gives you a = $e^{\lambda c}$ - $e^{-\lambda c}$ / 2 c and b = $e^{\lambda c}$ + $e^{-\lambda c}$ / 2. We will not really require the a for future use we will just require the b. So, just keep this b in mind this is how you get a linear approximation to a exponential ok.

So, now let us continue here let us continue with our agenda of proving Azuma's inequality. So, take any λ >0 for that let us evaluate the moment generating function of the Σ Xi's. So, expected value e to the λ Σ i = 1 to n X_i is / definition the expected value.

So, this is splits as the product of moment generating functions. Product of the expected value of the product sorry and now let us apply the linear approximation linear upper bound to each of these e to the λ Xi's.

(Refer Slide Time: 11:28)

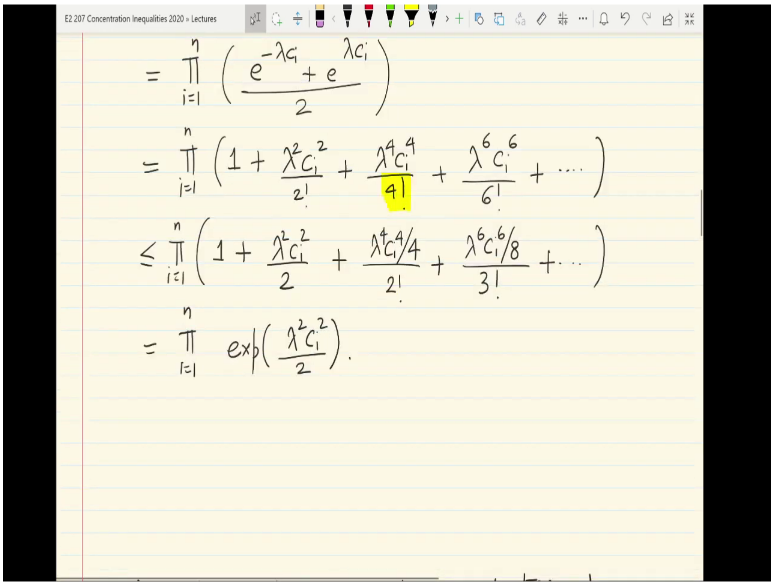


So, what that gives you is because the exponential always has a positive range one can apply this term / term. So, its the expected value of the product of a i X i + b i ok where a_i and b_i are customized according to this linear approximation depending on c i for every random variable X_i ok.

So, its the expected value of this product of a fine functions and now you can easily see that all when you take the X when you × these a fine expressions and take the expectation because of the multiplicative family property any terms involved in X_i will integrate out to 0 and you finally, are left with only expected value of the product b_i which is really not an expected value at all it just a constant ok.

So, this is by the multiplicative family property ok. So, there is really no expectation after all in the upper bound and this becomes the product substituting in the b values you have $e^{-\lambda c i} + e^{\lambda c i} / 2$.

(Refer Slide Time: 12:59)



Now, this in turn / the power series expansion of the exponential is just the product 1 + the second fourth sixth terms all cancel out leaving you with λ 2 C i 2 / 2 ! λ to the 4 C i to the 4 / 4 ! the next term just one last term that I will write out here is λ to the 6 C i to the 6 / 6 ! and so, on ok.

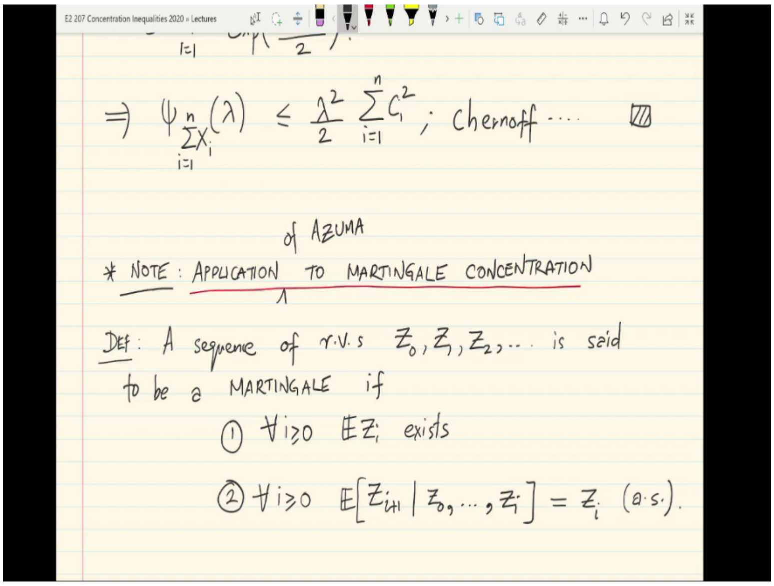
Now, we can upper bound each of these power series expansions as follows. So, λ 2 I leave this term as it is 2 ok. So, 2 ! is 2 now what I am going to do is basically. So, let us look at 4 ! here. So, 4 ! is 1 into 2 into 3 into 4, I am going to basically drop the odd numbers there.

So, 1 and 3 there. So, 1 into 2 into 3 into 4 is≥ 2 into 4. So, I will drop all the odd integers there and since I am dropping it in a denominator I only get an upper bound for the final expression and since I am left with 2 and 4, I will take the 2 common out of the odd numbers and I will basically get 2 raised to 2 into 2! ok.

So, what I get is basically λ raised to 4 C_i raised to 4 / 4 that is the 2 2 is that are come out and then I will basically have 2! the same thing if I due to the next term I will have λ 6 C_i raised to 6, I will get 3 twos as a product. So, that is 8 and then what I left with is what I am left with is 3! and so, on. So, you can see that this is basically = nothing, but the exponent of $\lambda^2 C_i^2 / 2$ ok. So, that is the bound.

So, what we have obtained is that the moment generating function of the Σ of all the X_i's is at most this quantity.

(Refer Slide Time: 15:11)

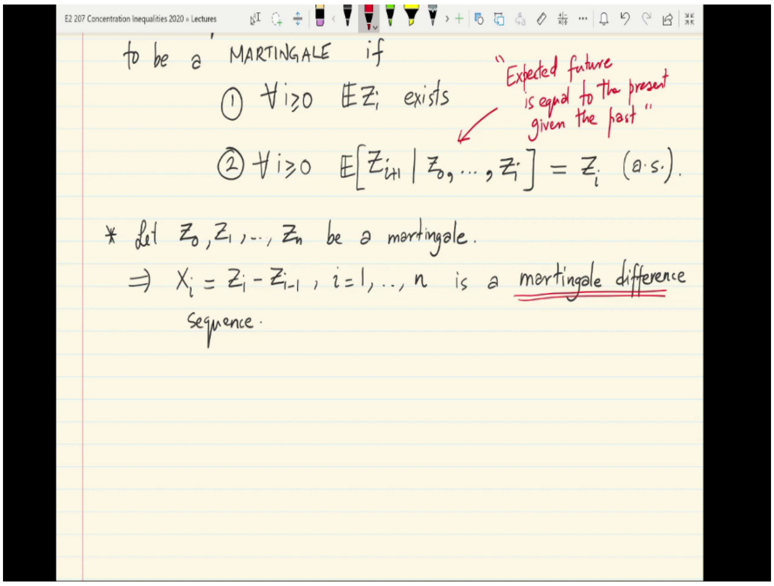


And so, it follows that C of the log moment generating function if you take logs in both sides at λ is bounded / λ 2 / 2 into the Σ of these C_i 2s ok and then just apply Chertoff to get the final bound to get Azuma's tail inequality ok. So, that is the proof of Azuma's inequality what it gives you is a tail bound for the Σ of a multiplicative family of random variables a tail deviation bound ok.

Now, why is it useful? We will give a glimpse of why it can be used very very effectively this is section called application of Azuma's inequality to martingale concentration. So, we are going to apply Azuma to come up with a what is called martingale concentration and example of a martingale concentration result. So, to define to set that a let us first define what the martingale is.

So, sequence. So, in our context here sequence of random variables Z 0, Z 1, Z 2 and so, on is said to be a martingale, if two conditions are satisfied. Firstly, that all of these random variables are finite expectation. The second more critical property is that the expected value of any of them given all the previous ones is = the last one the most recent one Z i ok. So, expected value of Z i + 1 given all previous random variables up to Z i is = Z i ok.

(Refer Slide Time: 16:45)

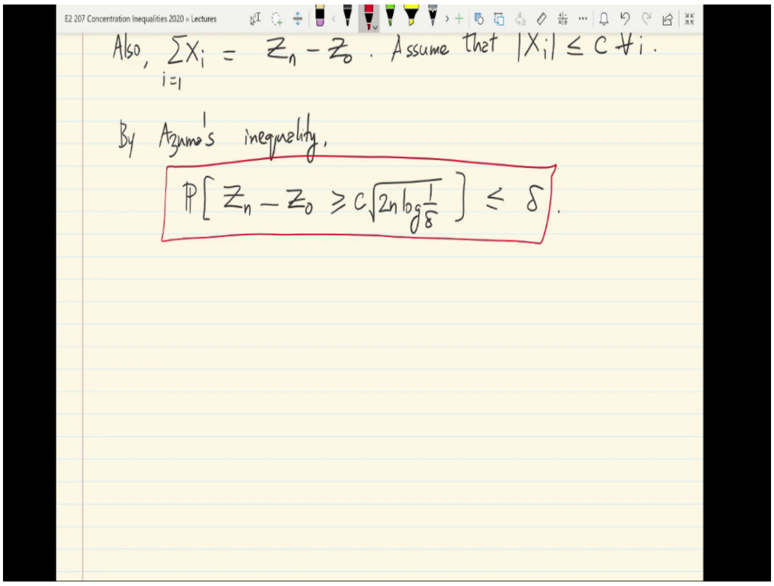


So, basically in words it means that the expected. So, very roughly speaking the expected future if you are sitting at time i the expected future value which is Z i + 1 is = the present given the past ok. So, or another way to say it is that the expected change of the future compared to what you have seen so, far is 0 ok. Its a very powerful property that arises in many stochastic processes and so, its of great interest to be able to control deviations or proof concentration inequalities about martingales very often ok.

So, with this definition of a martingale let us as Σ e. So, let us see how Azuma can be applied. So, let Z_0, Z_1 all the way up to Z n be a martingale meaning that basically the conditional expectation of each of them given the previous members of the sequences is the same as the most recent one ok. So, this means that. So, if you take differences successive differences and call them X_i. So, X_i is = Z_i - Z_{i-1} for i going from 1 to n, its easy to show that this is the martingale difference sequence as described for ok.

So, what is a martingale difference sequence? If you take the conditional expectation of each X_i with respect to the previous xs all the way up to X_{i-1} then that conditional expectation is 0 ok. So, that is what it means for the X's to be a martingale difference sequence.

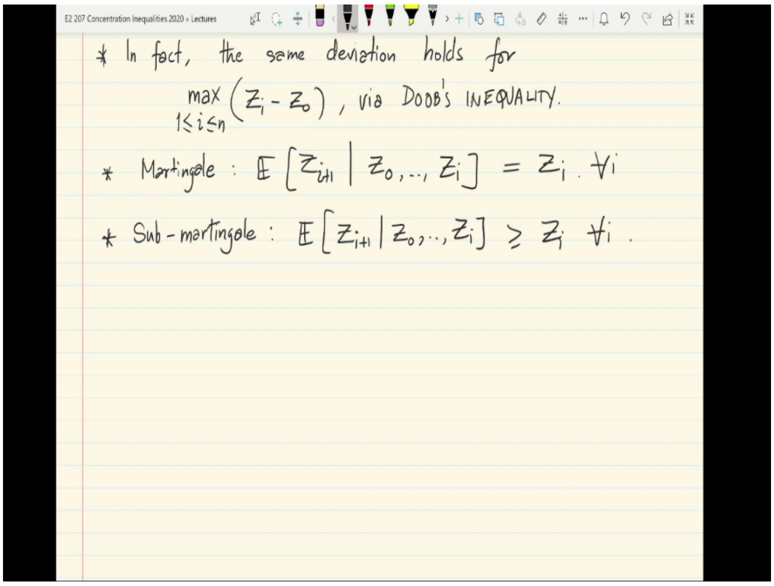
(Refer Slide Time: 18:54)



Also so, you have the important identity that the Σ of all these X i' s is precisely the final Z - the initial Z ok just my telescoping Σ s. So, let us as Σ e that all these martingale differences X_i are bounded / some numbers ok for all i. So, / Azuma's inequality what is this give us? If you apply Azuma's inequality to the Σ of all these X i's which are a multiplicative family because they are martingale they are a martingale difference sequence.

We have that the probability that the last term of the stochastic process Z_n deviates from the initial value by an amount which is C times root n roughly C √2 n log 1 / delta. So, this is a rare event ok happens with probability at most delta. So, this is a straightforward application of Azuma's inequality. In fact, something much stronger can be said about the same martingale Z n about the fluctuations of the same martingale Z_0, Z_1 all the way up to Z_n as the stochastic process by using something called Doob's inequality.

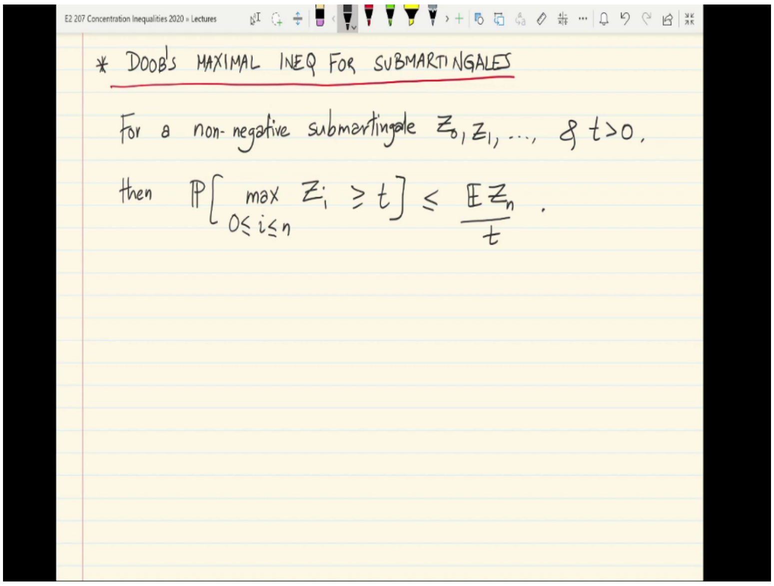
(Refer Slide Time: 20:32)



So, let me put that down here so in fact, the same deviation holds the same deviation bound holds for the quantity which is the largest possible deviation of any of the martingale terms. So, max over all i going from 1 to n of $Z_i - Z_0$ ok. So, recall that this is of course, lower bounded by $Z_n - Z_0$. In fact, this is sort of like a uniform deviation of all the terms of the martingale Z Z_i with relative to the initial value Z_0 ok. So, via what is called Doob's inequality.

So, let state. So, to state Doob's inequality, it helps to define what is called a sub martingale. So, recall that a martingale is a bunch of random variables Z_i such that essentially expected value of Z_{i+1} given Z_0 to Z_i turns out to be the same as Z i ok a generalization of this is what is called a sub martingale which says that the expected value of Z_{i+1} the same term this conditional expectation is at least Z_i ok. So, for all i ok. So, this is the generalization of a martingale, the mean the conditional expectation is at least the current value.

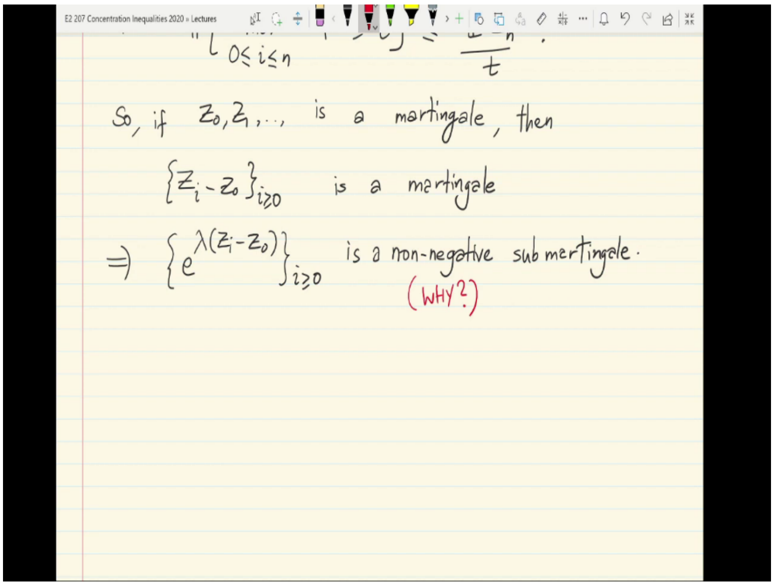
(Refer Slide Time: 22:27)



So, with this we can state what is called the famous Doob's maximal inequality for sub martingales, there are several versions of this inequality, but the one we will state here which will be useful for us is the following. So, if you have a non-negative sub martingale let say Z_0, Z_1, Z_2 and so, on and if you have t a positive number, then the probability that the largest Z exceeds t is no more than the expected value of the last element of the Z_n / t ok.

So, this is a very similar form to the basic Markov's inequality except that it is much more powerful in the sense that it applies to the uniform fluctuations or uniform deviations of an n length stochastic process ok. So, as long as you can control the right hand side which is the expected value of the nth term of the nth random variable in the stochastic process, then you can hope to control in a uniform way the deviations of this entire sample path of this martingale ok.

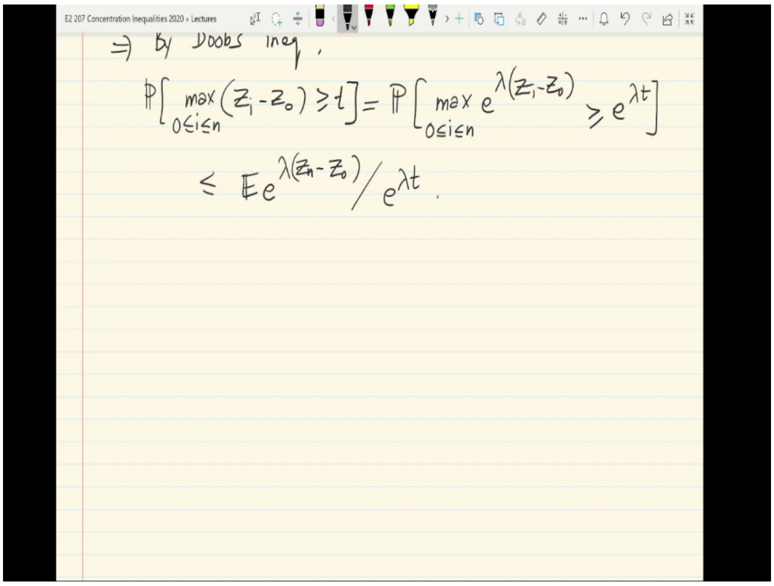
(Refer Slide Time: 24:10)



So, just to see how this is useful if Z 0, Z 1 and so, on is a martingale. So, let us take a the same set up to which we applied Azuma's inequality where the Z i's were all a martingale and you got this inequality here right. So, here all the Z_i's were as Σ e to be a martingale sequence then what we can do is we can subtract firstly, Z_0 from each of these. So, $Z_i - Z_0$ as i ranges from 0, 1, 2 and so, on is also a martingale ok one can easily check this and now what we can do is we can apply the function e raised to λ times to each of these martingale element.

So, let us consider e raised to λ into $Z_i - Z_0$ over all i, this turns out to be a so firstly, this is nonnegative its a nonnegative stochastic process; obviously, and it actually turns out to be a sub martingale ok. You can easily check this I leave this to you to check the reason for this property, but roughly I mean the abstract reason for this is if you apply any convex function to a martingale then you show that the resulting sequence is a sub martingale ok.

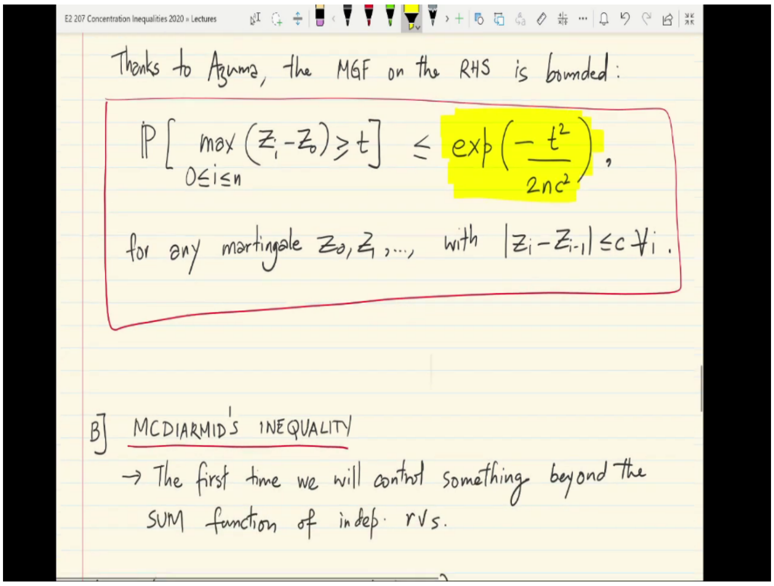
(Refer Slide Time: 25:44)



So, hence / Doob's inequality, we now have a non-negative sub martingale here and we can say that the probability that the largest deviation Z i - Z 0 exceeds t for any λ greater than 0 this is = the. So, let us exponentiate both sides largest of i of e raised to λ Z_i - $Z_0 \geq e^{\lambda t}$ and / Doob's inequality you get the bound you can just substitute the expected value of the last element and divide / e raised to λ t ok and now is where.

So, now we just reduce the supremum this max over i = 1 to n. So, just the single Z n - Z 0 whose moment generating function we can already control thanks to Azuma's inequality.

(Refer Slide Time: 26:55)



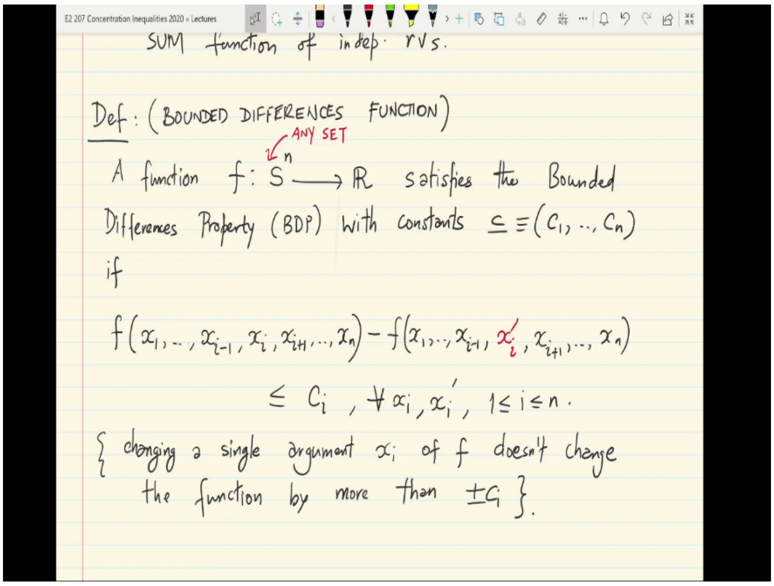
So, thanks to Azuma the MGF the moment generating function on the RHS the right hand side is bounded ok and we finally, get the bound probability that the largest deviation of any Z n Z i from its initial value is at most e raised to - t 2 / 2 n c 2 ok for any martingale process Z 0, Z 1 so, on with bounded increments ok.

So, Z i - Z i - 1 is at most c for all i ok. So, its important to note that this bound is obtained the right hand side bound is obtained / basically applying Doob's inequality first to get to the last term the Z n term the moment generating function of the Z n - Z naught.

And then applying Azuma's inequality to be able to control the moment generating function of Z n - Z naught which is essentially its log moment generating function is here to essentially give a sub Gaussian type tail here and then applying the Chertoff bound later ok. So, that is what this entire technique is ok.

So, the sub Gaussian appropriate sub Gaussian nature of the tail is what Azuma's inequality brings out and in conjunction with Doob's inequality this helps you control uniformly the deviations of a martingale along its entire sample path ok.

(Refer Slide Time: 29:13)



So, we come to the second part of this lecture which in which we will show a very powerful inequality called MacDiarmid's inequality that will for the first time allow us to control functions other than the Σ of independent random variables ok. So, this will allow us to see for the first time, how we can control something beyond the Σ of independent random variables. So, may be more complicated functions of a bunch of independent quantities. So, before we present MacDiarmid's inequality and derive it, let us define something called a bounded differences function.

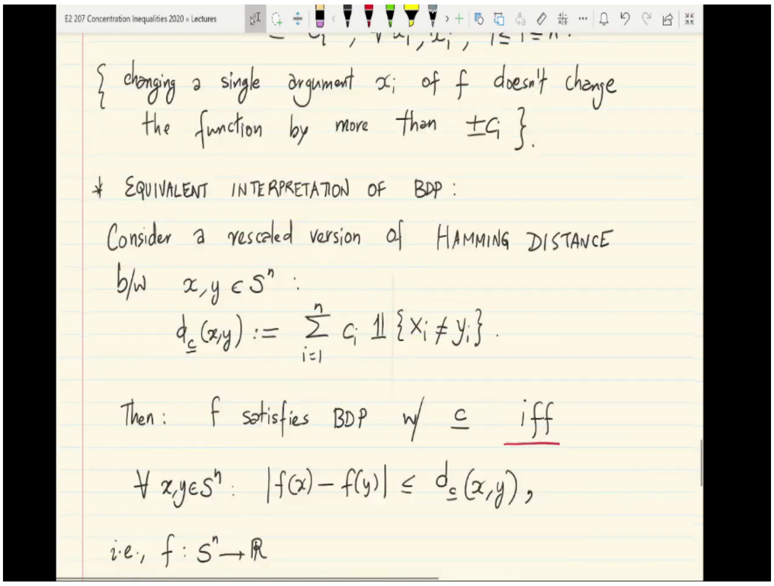
So, a function. So, let f be a function from any space S n. So, the Cartesian product of any set S. So, this set S can be very abstract ok. So, any set n copies of that set to R. So, a function f of this form is set to satisfy the bounded differences property. So, let us called it BDP with constants C. So, with the constant vector let us call it C under bar denoted as C 1 up to C n. So, these are C real numbers these are n real numbers if the following property holds. So, take the function evaluate the function at any n tuple x 1 through x n.

So, I will split it as x 1 through x i - 1 followed / x i followed / x i + 1 all the way up to x n and take its difference from the same argument except where one argument is replace the ith argument is replace. So, x 1 through x i - 1 and then instead of x i let say we put a different argument x i prime and then the remaining arguments are the same.

So, whenever you have a change of one argument then the function value deviation is no more than C i. So, this holds for all choices of x i and the different choice x i prime and for i being in 1 to n ok. So, roughly. So, inverts changing a single argument x i of f does not change the function / more than + C i or + - C i in either direction ok.

So, this is what is called a bounded differences function its relatively stable to small perturbations of its arguments and another way to express an equivalent way to express the same property is as follows.

(Refer Slide Time: 32:31)

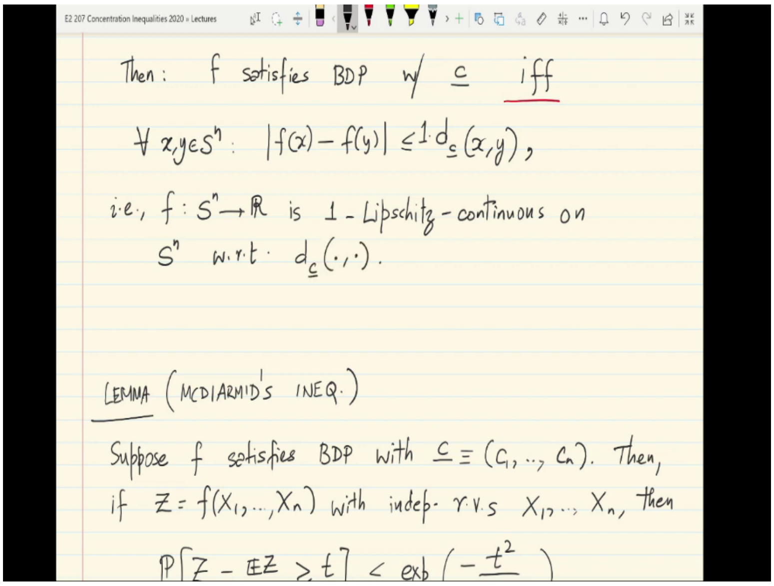


So, this is an equivalent interpretation of the BDP of the bounded differences property the more geometric interpretation. So, let us consider a rescaled version of what is called the hamming distance between two points two vectors x and y in S n ok. So, the way. So, think of a distance or a metric between discrepancy measured between x and y defined as d c x, y c is the vector of coefficients c 1 through c n of the bounded differences property. So, let this be defined as the Σ over all i of C i times the indicator that X i is not = Y i ok.

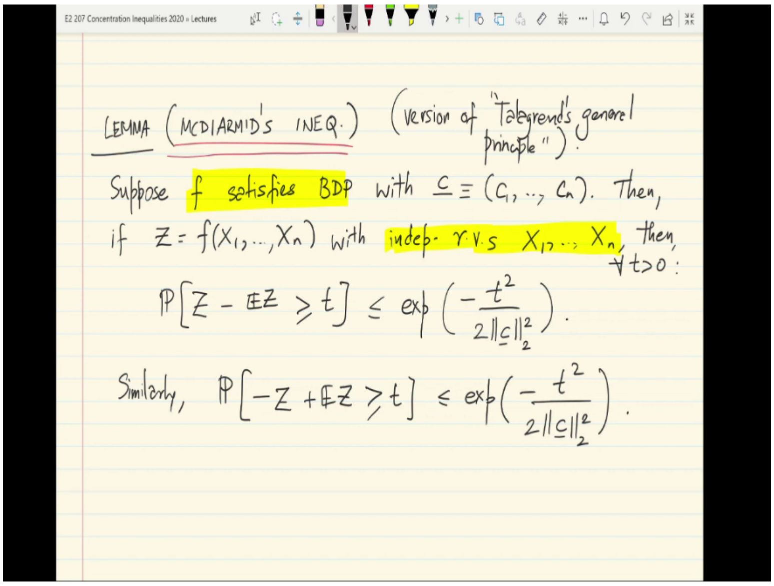
So, if the C i's are all 1 you get the usual what is called the usual hamming distance between 2 n tuples, but this is for general C i. So, then the following statements are equivalent f satisfies the bounded differences property with constants c 1 through c n denoted / c bar if and only if for all x and y in S n the modulus of f x the absolute value of f x - f y is at most d c bar x bar ok.

Now, to reader similar with the analysis and topology this will remind you of what is called Lipchitz continuity. So, another way to say the same thing is that f from S n to R is 1 Lipchitz continuous because there is essentially a 1 here × in the right hand side on S n with respect to the metric quote unquote metric or distance in d / c bar ok.

(Refer Slide Time: 35:15)



(Refer Slide Time: 35:27)



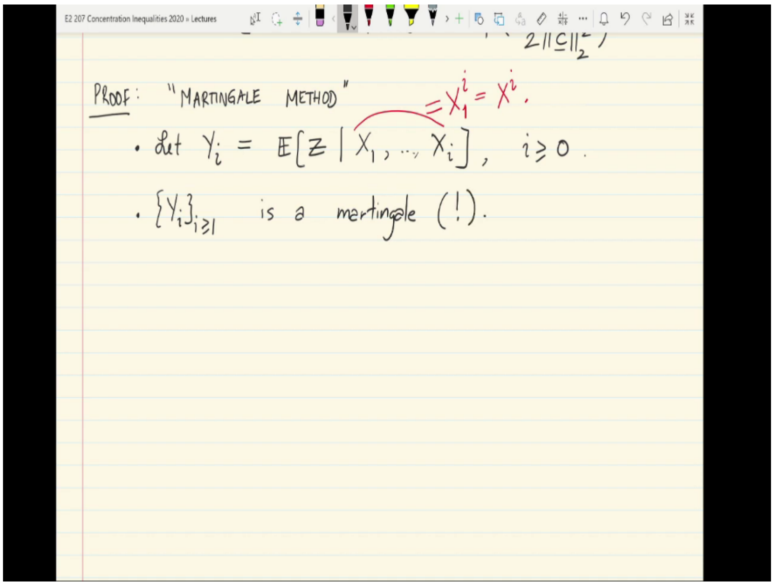
So, with this we are ready to state what is called MacDiarmid's inequality which is a very powerful inequality that applies to the fluctuations of bounded difference functions of independent random variables. So, the setup is that if you have a function f that satisfies a bounded difference property and you apply f to a bunch of independent random variables X 1 through X n then the probability that Z deviates from its expected value / a number more than t.

So, t is positive here. So, the probability that Z deviates from its own expected value / a number larger than t is at most e raised to - t 2 / 2 times the l 2 norm of the vector c whole 2. This is again an exponential concentration a sub Gaussian concentration type result but for random variables Z that can depend on a pretty in a pretty complicated way on independent random variables X 1 through X n.

And in some sense this is a version of this is one instantiation of Talagrand's principle that we pointed out in the first lecture ok which says that a stable function which depends only in a stable sense on its constituent random variables should essentially exhibit strong concentration to its mean ok.

So, similarly you can apply the same to - f and you get a property of the other tail the left hand side tail of z with the same sub Gaussian type tail d k ok of probability. So, in the last part of this lecture we will basically prove MacDiarmid's inequality and that the heart of it we will exploit the boundary differences property of it or the stability property of it.

(Refer Slide Time: 37:25)

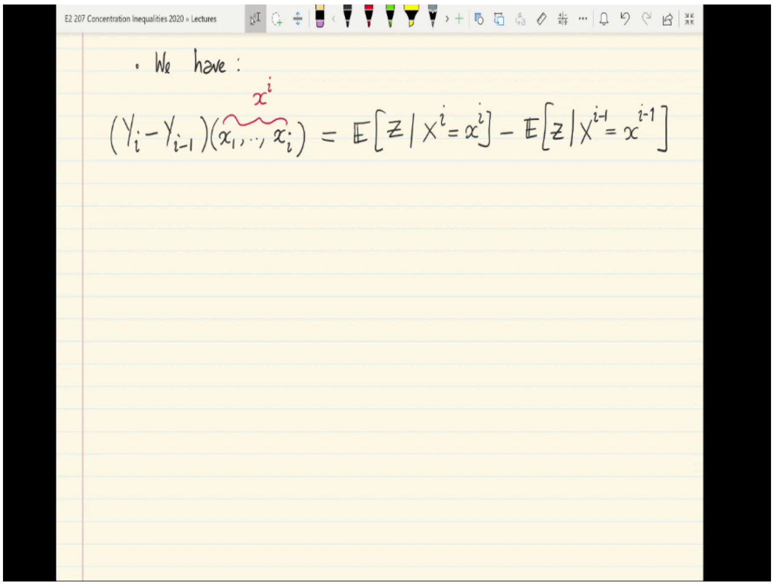


So, to prove MacDiarmid's inequality we will appeal to what is commonly called the martingale method in probability and at the heart of this method is basically constructing an appropriate martingale given a bunch of random variables to which you can apply powerful results about martingales and their deviations. So, in this case let us do the following let Y i. So, take the expected value of Z which is f of X 1 through X n conditioned on the first i of these X's ok and call it.

So, I $\geq$ 0 and call it Y i ok for notational purposes we will often refer to the collection X 1 through X i as X subscript 1 superscript i or also in some cases just as X_i ok where the subscript if not present is as Σ ed to be 1 ok. So, Y_i is essentially the conditional expectation of Z based on increasing collections of random variables X_1 up to X_i as i ranges from 0 up to n and the nice thing about this set of random variables Y is that it forms a martingale. So, this is easy to show offline.

So, the set of variables Y_i form martingale ok. This is a rather nice property that emerges just / this definition. So, let us now compute. So, if the moment you have Y i as a martingale one may be tempted to try to understand its differences and if they are bounded appropriately one can probably then go ahead and apply a nice inequality like Azuma's inequality for bounded difference for martingales with bounded differences bounded increments ok.

(Refer Slide Time: 39:39)

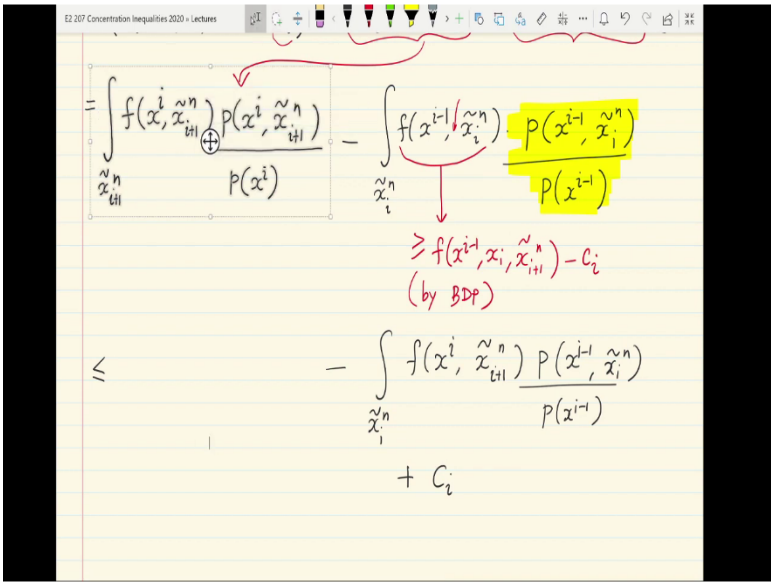


So, to that end let us investigate ah. So, we have. So, let us consider the difference $Y_i - Y_{i-1}$ ok. Now, let us think of this as a random variable. So, of course, all the random variables y depend on the basic random variables in this community space which are the X 1 through X n. So, let us think of y this difference $Y_i - Y_{i-1}$ as a function of the base basic values of all the random variables which is x 1 through x n ok. In particular we know that Y_i is going to be a function of X 1 through X_i Y_{i-1} is going to be a function only of the first $i - 1^{th}$ random variables.

So, in general this difference is going to be a function of the first i of these random variables ok and in our notation this is just x superscript i ok just for our reference. So, this by definition is the expected value of Z which is f of all the Xs. So, let just say Z given big X i is = the configuration small x_i - expected value of Z given the first i random variables take the values specific values small x_{i-1} ok. So, we are trying to express the difference $Y_i - Y_{i-1}$ in terms of the values the outcome x 1 through x_i.

So, let us write this out in some detail what I am going to have here is I am going to write each one of these terms using the basic conditional expectation formula.

(Refer Slide Time: 41:39)



So, let us do this. So, the first term is what happens when we try to compute the conditional expectation of Z given / fixing the first the values of the first i random variables X i ok. So, let us as Σ e that all these random variables have densities the calculations is the same in the more general case, but this just illustrates the structure of the problem.

So, one would have integrated f Z is basically f of the first i being fixed to x superscript i and there are the remaining random variables which we will call x tilde i + 1 all the way to n ok and we have to integrate this over all the x tilde i + 1 to n ok with respect to the conditional distribution of the x given this configuration small x_i.

So, what I mean by that is one has to × this by the conditional distribution where the first i random variable they are fixed to x_i the remaining are allowed to range in the x tildes and to normalize one has to normalize to express this conditioning event this is just p of x i ok. p of x i is shorthand for the probability that the first i random variables in the xs as Σ e the configuration small x i ok.

So, this is exactly this integral in a similars in a similar spirit we can write down the second integral. So, this the integration is performed over all random variables other than the first i - 1; that means, x tilde i going to n f x_{i-1} x tilde i n and then you have finally, a p x_{i-1} x tilde i to n and 1 has to divide / p of x_{i-1} ok.

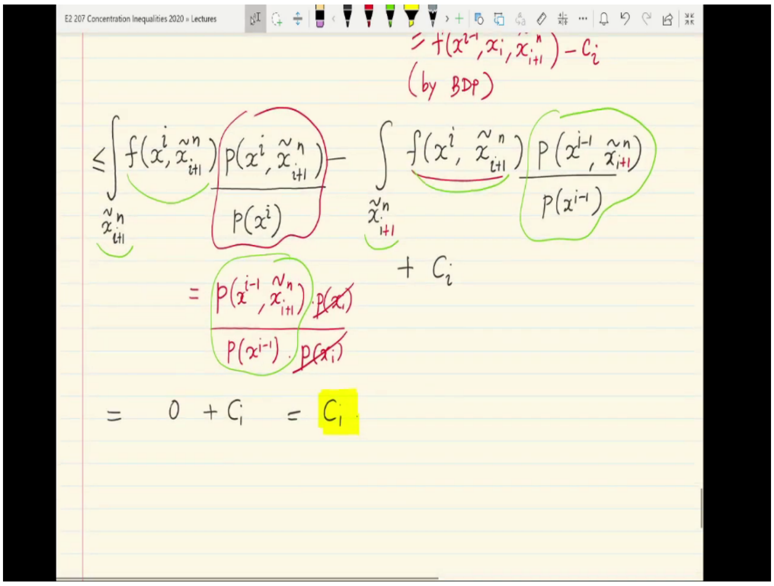
So, these are both of these conditional expectations written in detail as integrals ok. Now let us absorb that this term here ok. So, this term here imagine that x_i. So, this term has x_i tilde ok. So, the immediate next element after x_{i-1} is x_i tilde and if you were to replace that x_i tilde with x_i ok which is this particular x i here then we know / the boundary differences property that the function cannot change / more than C_i.

So, this value automatically turns out to have the lower bound. So, if I replaced x i tilde with x i with any other x_i for instance and let all the other arguments remain the same x tilde i + 1 to n then you would not deviate / a number more than C_i ok. So, this is / the bounded differences property on coordinate i. So, applying this bound here gives you that this is at most the same integral x tilde i + 1 n. So, let me just write ditto ditto the same integral here - integral over the same domain.

So, I can replace this term here with f x i x. So, the i x i gets absorbed in the x_{i-1} and becomes x superscript i x tilde i + 1 up to n followed / the usual x_{i-1} x_i tilde n divided / px_{i-1} and the - C i just gives you a + C_i at the end just as a constant because everything else.

So, if you integrate just this term ok over the x tilde i n i x tilde subscript i superscript n you will just get 1 / definition of the conditional expectations of the conditional distribution ok. So, you have some integral - some other integral + C i as a upper bound ok. Now, let us operate on the first term. So, to do that I will put back the first term here let me copy and paste this integral here ok.

(Refer Slide Time: 47:10)



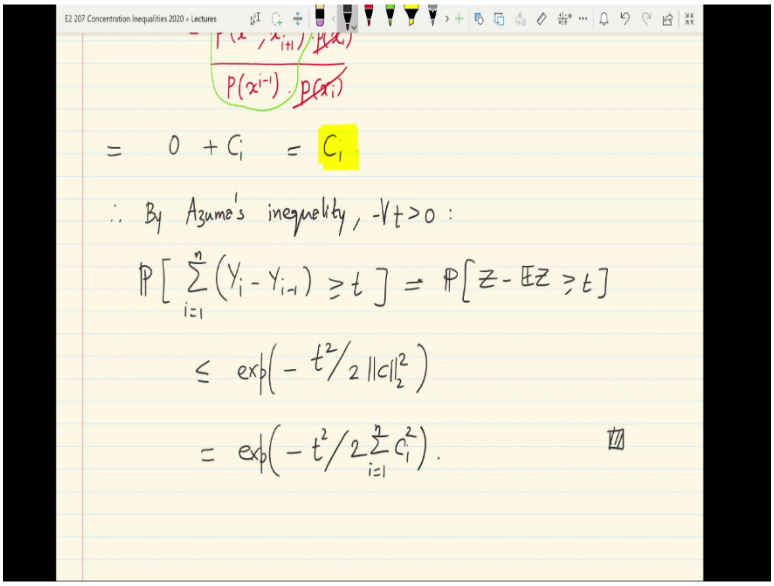
So, this was the exact first integral verbatim now what we have here is that. So, let us make two observations here. Firstly, in this right hand side integral let us complete a right hand side integral. So, if you look at the second integral here, this term does not depend on x_i tilde at all ok the only dependence on x_i tilde is here ok.

So, it follows that if you integrate you can drop the integral with respect to i i x i tilde you can just integrate from x_i tilde x_{i+1} tilde onwards and you can essentially drop that variable here because you integrate it out ok. So, what I mean by that is that you can just put i + 1 here ok that is what marginalizing out the x_i tilde has got us ok and now its time to compare both of these two expressions. So, we have the same domain of integration finally, we have the same first argument ok and what about the second argument we can write this as.

So, let us look at this term here this fraction. So, this / the independence of the xs recall that we as Σ e that x_1 through x n are independent one can always write this as p of x_i - 1 one through x_{i-1} and the remaining x tilde i + 1 to n × p x_i ok. So, I am singling out p x i because of the product structure the using the independence of the x i's / I will again single out p x_i from the denominator ok and so, this p of x_i cancels and we now have the same term as this here ok.

So, it basically means that the first two integrals are equal and so, you get that this expression is exactly $= 0 + C_i$ which is C_i. So, what we have shown is that almost surely for any realization y i - y i - 1 has the absolute bound C i ok. So, y i is a martingale whose difference is enjoy the ith difference enjoy is the bound C i ok and that is all we need to complete the argument.

(Refer Slide Time: 50:14)



So, invoking Azuma's inequality for the multiplicative family which is given / the differences of the Y is we have for t greater than 0 probability that i = 1 Y i - Y i - 1 exceeding t which is just the same in the Z languages probability Z - expected Z exceeding t is bounded / e raised to $- t^2 / 2$ norm $C^2{}_2$ which is essentially e raised to $- t^2 / 2\, \Sigma\, i\, C_i^2$ and that completes the proof of MacDiarmid's inequality.

Thank you.