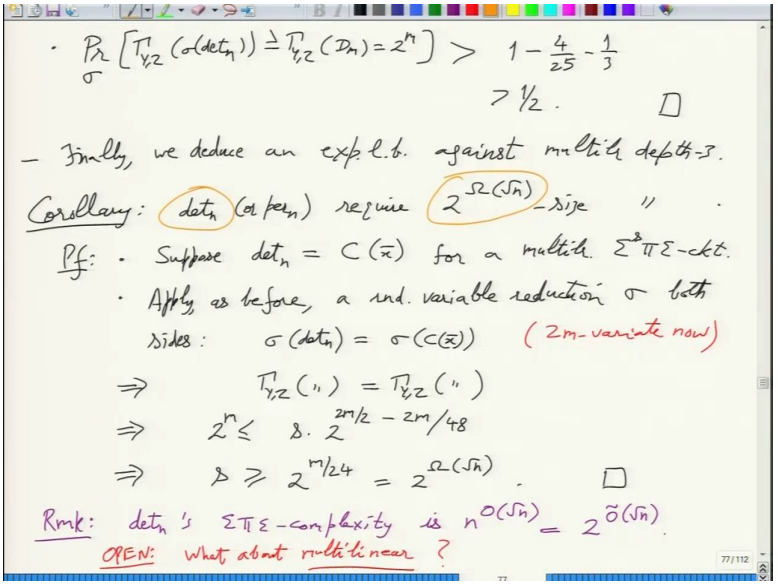
Arithmetic Circuit Complexity
Prof. Nitin Saxena
Department of Computer Science and Engineering
Indian Institute of Technology – Kanpur

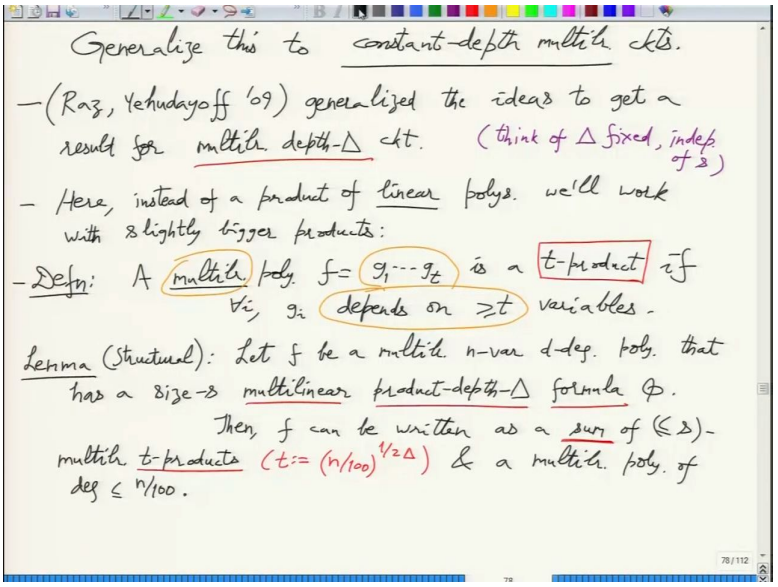
Lecture-17
Arithmetic Circuit Complexity

(Refer Slide Time: 00:14)



Last time we finished the proof of an exponential lower bound for determinant for multi linear depth-3 circuit. We showed that determinant or permanent $n \times n$ requires $2^{\Omega\sqrt{n}}$ size and this is field independent. This works for any field. And we also noted that we actually do not have such a representation. We do not really know $2^{\Omega\sqrt{n}}$ sized multi linear depth-3 representation but it is quite closed to 2^n.

(Refer Slide Time: 00:59)



We call it exponential. Now, we will try to generalize these techniques, which basically means that we want to continue working with this measure and the matrix where the matrix is we first partition the variables X into Y and Z and then look at the Y monomials vs the Z monomials. And look at the; consider the rank of that matrix. This was this generalization was done by Raz and Ran Raz Amir Yehudayoff, they generalized the prior ideas to get a result for multi linear depth-3 circuit, multi linear depth - Δ, depth-3 we have done.

But they generalized Razes original ideas to constant depth, Δ will be some absolute constant. As Δ grows, what will happen is the machinery will become; are the parameters will become worse? We will instead of $2^{\sqrt{n}}$ and we will get something like $2^{n^{1/\Delta}}$ but this is still exponential for constant Δ. Think of Δ constant.

So fixed; meaning independent of s. It is an absolute constant. And in this line of work, the methods also work for formulas.That will need a slight modification in the measure, which we will see in the end. Let us continue with constant depth. Now, instead of a product of linear polynomials, we will work with slightly more general products.We will call this t product. A multi linear polynomial $f = g_1 \cdots g_t$.

We will call it a t -product. If so, there are t polynomials t factors and these g_i depends on $\geq$ t variables and it is important that this is multi linear. These variables are disjoint as you

change i , so $g_1 \cdots g_t$ are basically partitioning the variable set X and each of these parts is at least t and there are t factors. So, in particular, n should be bigger than should be at least t^2 .

When we talk about a t product then we cannot say anything more than $\sqrt{n}$ product. We will need to study slightly general products because when you go from depth-3 to let us say depth-4. Already in depth-4 you can see that you have a sum of products of general polynomials. You have to study the measure on that.One common generalization seems to be t product.

And this will then continue this will allow you to go to depth-5 depth-6 and so on, up to constant depth, but then the results, the parameters will be worse. Let us see that let us see the upper bound basically well, but before we see the upper bound, we have to prove a structural result about constant depth circuits. What is the connection between depth- Δ circuits to this t product?

We have to make this connection. This is a structural lemma. Let f be multilinear n variate d-degree polynomial that has a size- s multi linear product depth, product depth Δ formula ϕ , we are seeing many things. Most important thing is that we start with a formula ϕ . Now, since we are talking about constant depth circuits, we can as well make them formula. The constant depth circuits are a special case of formulas.

So instead of circuit we will talk about formulas. And instead of depth, we will talk about product depth. We are assuming that addition the circuit is actually divided into layers and the layers addition, addition multiplication alternate. All these things we can do in the preprocessing, this does not cost much. So we are given this layered formula, the number of multiplication layers is at most Δ .

We are calling that product depth and this is there is this major restriction which is every gate computes a multi linear polynomial. This is our ϕ all of this we can achieve in preprocessing. Then f can be written as a sum of less $\leq$ (s)multi linear t products where

$t := (n/100)^{1/\Delta}$ we will use constants to make the comparisons work, but this is essentially $n^{1/\Delta}$.

This may come as a surprise. We have with we start with a depth Δ formula, but we are able to squash it into just a sum of the t products. As a sum of t products and there might be some error part which is a multi linear polynomial of low degree, degree $\leq$ n /100. So, f can be written as a sum of these things at most s many t products plus some low degree polynomials.

This is a structural result, this has nothing to do with our measure or lower bounds is the statement here. Once we have this structural result, then we will convert ϕ into this sum of t products and we will study the measure on a t product and get an upper bound and we will see that the upper bound will be smaller than the lower bound on determinant of the measures. That comparison will finish the exponential lower bound proof or constant depth multi linear formulas for circuits.

This is the definition of t product. We will apply it in the structural statement on some f and on some t no it is see is here multi liner t products. You have to invoke the definition of what is a t product.The error term very big the error term is you know will not be very big its degrees. Number of monomials is something like $n^{n/100}$. If you assume that each of you know different each of your linear terms actually has t variables or auditing variables.

Then the degree of each of the assignments will be something like t^2 and acutely other than the other term. Well, we will see. I promise you that in the end everything will be fine. That is because. The measure on this error term will give you something like $2^{n/10}$ contribution. And we will show that on this t product will the contributions will be kind of comparable. Think of $2^{n/10}$ as the final measure upper bound and the lower bound will be much higher on determinant.

In fact, I think of $2^{n/10}$ which will get for the error term, this will be much smaller than what we can proof for the t products. This contribution will not matter at all because we are working against $2^{n/2}$. The objective is to offer the upward on demand. $2^{n/2}$. is much smaller

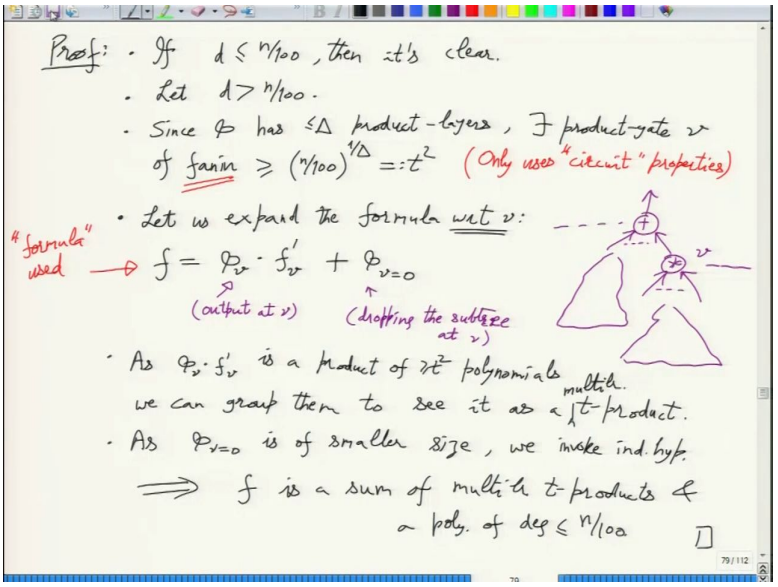
than $2^{n/2}$ .actually you like an error term is the highest measure that is what a means in general no the error term is does not matter.

This is something which we should be dropped because the measure upper bound for t products will be even bigger. We will just focus on them that will demand the most time in the proof details. This will be proved as always by induction. We will just induct on the depth of the formula will note that for product depth-1 when Δ was 1, we already did this the language now may be different, but this multi linear depth-3 circuit the product depth was 1.

And there we were studying product of linear polynomials and that is what will now try to generalize to higher depths. The higher depth case we will do inductively then you are asking sigma, we did it for $\Sigma\Pi\Sigma$ that is the base case. You have to see a product of linear polynomials as a t product. You can do it by clustering the linear polynomials actually.

So, each of your linear polynomials depends on at least 2 variables we are assuming that, you can collect t / 2 linear polynomials, multiply them call that g_1 and the next cluster g_2 and so on. By clustering you can actually see that in the base case, we did have a t product and we studied the measure on that. That was so base case is done. And now we will generalize it to higher to more product layers.

(Refer Slide Time: 17:01)



Let us do that so if the degree is n / 100 or less, well then everything is subsumed in the error part. We will assume that the degree of the resulting polynomial f is more than n / 100. Now since the degree at the root of the formula is more than n / 100 and the number of product layers is only Δ, what can you say about the degree demand at some product layer. There has to be a product layer where the degree goes up by this bound raise to $1/\Delta$.

Let us use that kind of averaging this bound raise to $1/\Delta$. Somewhere this blow up has to happen, if everywhere the blow up is smaller than this, then you can never reach n / 100. It says kind of averaging but happening in the exponent, because we are talking about multiplicative behavior. Since ϕ is a formula, since ϕ has product depth has $1 \leq \Delta$ product layers there has to exist a product gate v where the degree blows up, which basically boils down to how many inputs are there?

The product gates should have fanin at least we have to see I want to say this but is it fanin or degree? Thing this has to be degree of it has to be something else we need to just fanin there is that is true. This even fanin it cannot be degree, even fanin I am not very confident now.Let us continue and try to correct this. The point was that there will be this product gate v we can assume by induction that this product gate has that form.

It is a sum of the added works on any path from the root to the leaf will be less than this at every time and then the total degree can only be less than a fraction of the first layer, in this case, the first layer that he became at the layer after that you have a case where, already k/Δ. But this does not need to multi linearity this does not use any just write. This comes from the formula no for a circuit I think repeated squaring will create a problem that automatically gives you a bound on what no fanin means 2.

Degree is 2^s but the fanin for all the multiplication layers is just 2. You will never get anything interesting for circuit. Formula is actually important and sufficient. In a formula you can talk about just look at the tree structure. And if every multiplicative fanin is smaller than this $(n/100)^{1/\Delta}$ then observe that final degree will be smaller than n / 100. That is all yes.

Because first of all what we repeatedly do is first whatever is a parameter, adequate data or $2^{s^{1/2}}$.

That is I see this is true for any circuit then. This $(n/100)^{1/\Delta}$ we have already defined this t^2 in the lemma statement and this is the node which will help us regards or induct. We will have we have this multiplication gate which is v with the t^2 inputs at least and this will feed into addition gate. There is this part of the circuit and there is this part of the circuit and there might be other in this multiplication layer you have other multiplication gates and additionally or other additional gates.

How do you want to induct? That trick you have seen before this v you delete from this graph and so that is basically setting it to 0. Look at that part and look at the v sub tree. Those are smaller strictly smaller formulas. Let us expand that way. Let us expand the formula with respect to v. That will give you f = ϕ_v is output at v something will multiply this plus, the plus part will be when you set this v part to 0 so that is kind of dropping the sub tree.

What is the multiplier here? The multiplier is essentially this derivative with respect to v. So, we will just write it as f'_v. so this is the decomposition of your formula, this respect to any node v. You have a kind of a derivative part times whatever was being computed at v plus set the v part to 0. On this weekend now induct. What do you know about these 2 parts? What do you know about ϕ_v times f'_v.

ϕ_v already has t^2 inputs, by clustering we can see that it is a t product. Because the definition of t product was very weak. It only wanted a product of t things each depending on more than t variables, at least t variables. Just by the fanin you have no so here we are, we will also use the fanin that it is a multi linear formula. These inputs have disjoint support disjoint variables.

Let us write that down. So as ϕ_v . f'_v is a product of t^2 polynomials. We can group them it as a t product. This whole part actually the part of ϕ_v and times the derivative this part is already a t product just by trivial clustering or appropriate clustering. The other part ϕ_v = 0 v

as you we use induction hypothesis that it is a sum of t products and some low degree polynomial.

This is obviously multi linear and as $\phi_{v=0}$ is of smaller size we invoke induction hypothesis. Overall both these things tell us that f is sum of multi linear t products and polynomial of degree $\leq$ n / 100 this is clear. Look at the polynomial is being fed into v no nothing is no there is no demand of linearity or anything we just want to show that ϕ_v is a product of t things, each dependent on t variables.

Sure, that is needed. It is a very weak demand. That is all this t things and each dependent on t variables. If you are feeding in t^2 polynomials and the multiplication gate is multi linear, then you can just cluster things and you get. This is in the definition then a fanin be a that is implied. No, this is a product no, g_1 as a polynomial g_1 just some of the monomial. Every monomial has to be a multi linear monomial otherwise this monomial you cannot cancel in a product.

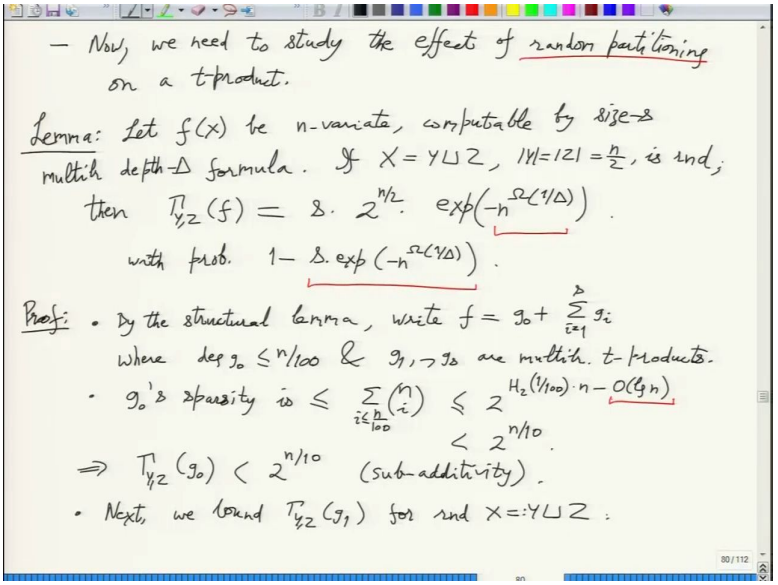
How can you cancel a non multi linear monomial it will survive that that also is implied?If this product is multi linear so, everything is implied by the multi linearity of it. So, just saying this is enough, those are actually small exercises. I will not prove it but it can be you can take it as an exercise because we are talking about product so, product is easy to study formula that is a good question. This guy needs a formula.

Formula is needed here to do this decomposition. This we had seen also when we did the brains depth reduction for formulas vs circuit depth reduction that this is brains expansion used in brains proof as well. Any other question? This terms out to be quite simple. And we use we did not write here the base case, but I guess if to be convinced that when depth is just 1 that is the depth 3.

There you are basically multiplying these linear polynomials and they are also you will cluster them so that it looks like a t product.This completes the structural lemma. Now what

we will do is just study this product using the measures the behavior of the measure on a product t product.

(Refer Slide Time: 33:24)



Now we need to study the effect of random partitioning. This is something even more basic what happens when we randomly partition the variables on a t product Let f on variable set x be n variate polynomial and computable by size s, multi linear depth Δ formula then you partition equally size of x is n and randomly. When you do this, then the $\Gamma_{Y,Z}(f)$.

And so I am not just studying product I am studying the whole model, but because of the previous lemma the whole model is just I mean, it essentially is a sum of t products as many. So ultimately, the proof will boil down to only studying t product. And what we will say what we will get in the end is that $\Gamma_{Y,Z}(f) = s.\, 2^{n/2}.\, exp(-n^{\Omega(1/\Delta)})$ so s s for the size,, ignore s because what so, look at the thing we are seeing about a t product.

We are saying that it is a measure is significantly smaller than $2^{n/2}$ where smaller means that in the exponent we are subtracting by $n^{1/\Delta}$. If you compare this with the depth-3, what we had shown in depth-3 there we had gotten n / 48 or something. That was a special case of this. In fact, it was it was the parameters are very good here the parameters are going to the exponent of n even.

The n will actually become when you take Δ to be 2 it will become smaller than $\sqrt{n}$ so, that difference has become smaller now, I mean as Δ increases difference becomes smaller, but still it is quite significant it is n to the these some constant. You are actually talking about 2^n raise to some constant and that will become your lower bound on X. It is much bigger than polynomial.

It is a nearly exponential lower bound for determinant in the end and what is the probability so, this will happen with probability because we have chosen Y randomly with probability significantly high. The error probability is if you assume that s is small then the error probability is small. This is clearly helpful in the eventual lower bound is the statement clear. We are seeing that if your formula has size smaller we are smaller means $2^{n^{1/\Delta}}$.

If it is smaller than this then with positive probability, in fact, constant probability, the measure will be significantly smaller than $2^{n/2}$, $2^{n/2}$ is the max measure to possible. That is the statement as you pick the partition in a random way, equal partition. So lets do the proof. By the structural lemma write $f = g_0 + \sum_{i=0}^{s} g_i$ where g_0 has low degree and $g_1, \cdots g_s$ are multi linear t products where this figured from the structural lemma. What is g_0 's sparsity?

How many monomials are there?We can take a shortcut because we know that g_0 is multi linear, If there are i variables, then the number of monomials can only be $\binom{n}{i}$.,Using that shortcut we will get a better bound. $\binom{n}{i}$ i going from 1 to or 0 to n / 100 and variables are something we can just stop this will come out to be quite large compared to s. This is this g_0 has to be necessarily separated.

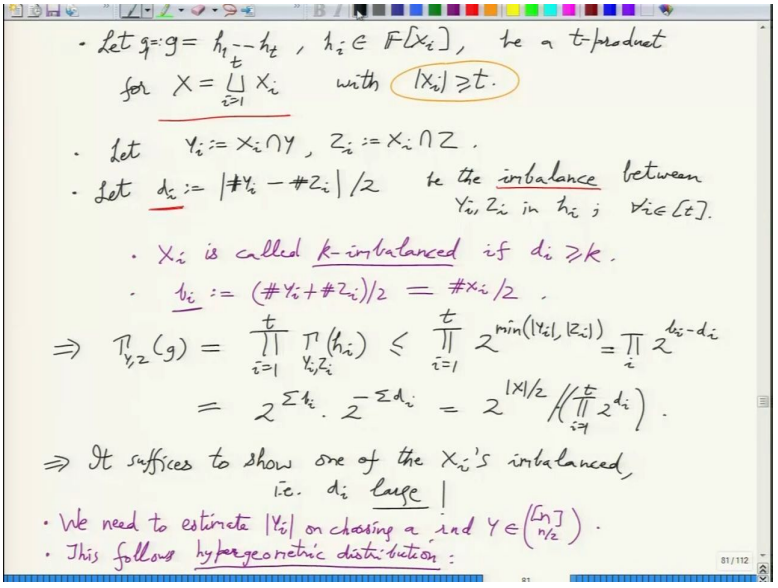
And then the number of monomials it has is Σ $\binom{n}{i}$ if you are i mean this is essentially dominated by $\binom{n}{n/100}$ which is something like $2^{n/100}$. It is in the right direction, but we can even do this exactly. This will come out to be by Stirling's approximation $2^{H_2(1/100).n - O(logn)}$.

That is not more than login up to a constant multiple so, this cannot change the main term by much. So, assuming n to be sufficiently large, this we can assume is smaller than $2^{n/10}$. Is

this fair? By Stirling's approximation we get that g_0 has at most $2^{n/10}$ monomials what is the measure on a monomial? 1 so, the measure of g_0 is at most this much by sub-additivity. I mean in our scale of things this $2^{n/10}$ is completely insignificant because we are dealing with $n/2 - n^{1/\Delta}$.

In that scale of things this $2^{n/10}$ is actually not important. We just, once we have this expression we forget about it. And let us go to the main thing which is g_1. What is the measure on g_1? Next we bound measure of g_1 for a random partition.

(Refer Slide Time: 45:00)



So what? Let us replace g_1 / g in our subsequent calculations, so let $g = h_1 \cdots h_t$. It is a t product, so, you have t factors and each factor depends on at least t variables. Let us say X_i is the variable set so the variable set of g is X and if you look at the factors they are variable sets or X_i is, each of them is at least t and you have t factors. What is the action of gamma one this? When you partition X you it will induce a partition of X_i.

Let us call that Y_i, Z_i? And recall the intuition that if all the variables say in each one go to y what is the measure contribution of H_1 just 1 right?If all the variables of H_1 are one side, then it contributes nothing and quantitatively that intuition leads you to the fact that H_1 will contribute the most or it will contribute more, if the discrepancy is more between Y_i , *and* Z_i We define discrepancy.

Let $d_i := |\neq Y_i - \neq Z_i|/2$ .Let this be the imbalance between Y_i, Z_i *in* h_i for all i and we will call X_i in k imbalanced X_i IS called k -imbalanced. If the what this part is just combinatorial I mean we are just using the graph, somewhat graph properties and then mostly probability. We will call this variable set of h_i in k imbalanced if the discrepancy is at least k and it will be useful to define $b_i := (\neq Y_i + \neq Z_i) = \neq X_i/2$, d_i is the discrepancy b_i is, is half of the size.

And if you know d_i and b_i then you know the sizes of Y_i as Z_i there is a simple formula. There is some will give you in particular size of the bigger part and the difference will give you size of the big smaller part. Let us write that yes. If you look at the measure now on this product so, that will be the product of measures which is now $Y_i Z_i$ and what is the upper bound on this. That is $2^{min(|Y_i|,|Z_i|)}$.

What is $2^{\Sigma b_i}$? $2^{\Sigma b_i}$ is just X / 2 and then you are just dividing it by the product of 2^{d_i} that clear. The point of all this is that and this $2^{X/2}$ is think of this as $2^{n/2}$ and we are interested in the amount of this n / 2 will fall. So, that amount we want to get simply from sum d_i In the remaining proof what we want to show is that there is some i says that d_i is as big as the thing we need in the lemma statement.

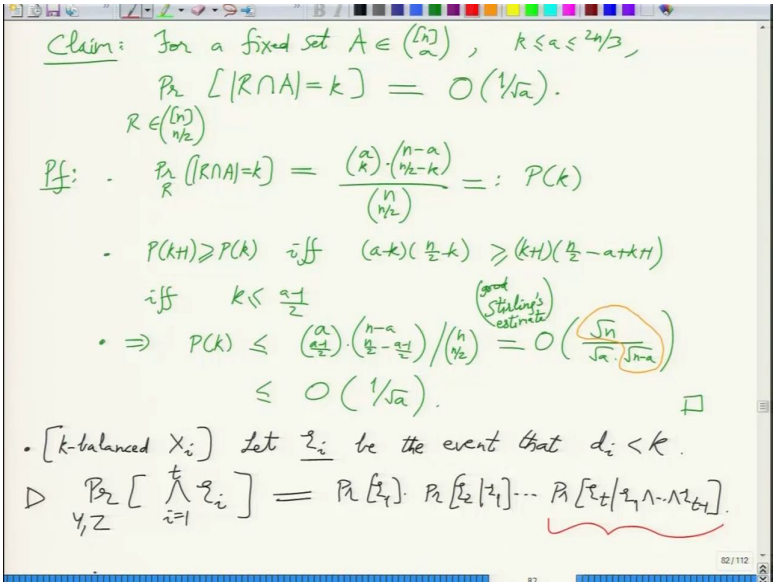
It suffices to show one of the X_i 's imbalanced that is d_i large lead to this is what we want to show and how do you show this? You do it sequentially you start with the X_1 and with your probabilistic choice of Y and Z compute the probability that X_1 is balanced. And then conditioned on this X_2 is balanced, conditioned on that X_3 is balanced and so on. Compute this probability and show that it is small. Which will mean that some x i will be imbalanced.

Whichever X_i imbalance d_i will be quantitatively large. And it will get it will give you the expression in the lemma statement. Now it is a scary probabilistic calculation. So brush up your probability for this. We need to estimate and remember that we are treating X_i as fixed rate. When we do the probability calculations X_1 to X_t are fixed and they are whenever needed we can use this lower bound they are large sets and experiment on which the probability is in that experiment we are picking why to be n / 2 size subset of n.

That is the only random choice everything else is fixed in the experiment. We need to estimate the size of Y_i $\varepsilon\binom{[n]}{n/2}$ .Yes, I do not know why but this kind of distribution is called hyper geometric distribution. It is meant to scare you further. This follows hyper geometric distribution.Again, you picking n / 2 sides subset randomly.

And you are interested in the random variable size of Y with a fixed size of Y intersected with a fixed it. How many points have you picked? Essentially from a fixed set that is your random variable. And so this is for some reason called hyper geometric distribution. And so we want to compute the probability of that number being something like k what is the probability?

(Refer Slide Time: 56:41)



Let us take a detour and prove the compute the probability. For a fixed set A of size a you forget about these bounds will need these bounds in the end. This is essentially saying that set is of a decent size it is not too small it is not too large. So for a decent sized, fixed set A, the probability that a random set intersection with a comes out to be k, k sized where R is a random n / 2 size subset of 1 to n.

What is this probability? What do you think? Is this probability large or small? Well, or what do we need in our proof for a lemma statement, what do you want? do you want an upper bound on this or a lower bound on this that you can see according to the strategy, we need

upper bound, you want to show that this is small, because we will then invoke this many times and we will show that the probability of X_1 X_2 all these things being balanced is small.

The direction we want is often upper bound will show that this is at most 1 over root a. Which again seems to be birthday paradox, but I may be naive. Again there is the $\sqrt{a}$ appearing yes, once this claim is stated the proof is very straight forward. This is just you can write down the probability exactly. $pr[|R \cap A| = k]$ is what? So you have to first decide which k elements. That will be each use k because this set is fixed.

That is your favorable choices. Once you have decided which k these you have to then forget. So n - k and no, it is not saying no this a and k respectively you remove because set a is fixed. And which k is also fixed and this divided by $\binom{n}{n/2}$ so, it is similar $\binom{a}{k}$ now introduced no there then it is the same without $\binom{a}{k}$ is that but there we did not have to k. But that was it fun? It was coming out to be 1 I think we are taking a = k then.

That was a = k then that was a = k, k and this is just general version. Call this P (k). This probability because we have to now make sense of this expression. To make sense of this expression, we will need to know how does k a and n compare. Which is why we have those bounds the boundary condition of a but it is tricky to analyze. We will start with a small observation that if you increase k what happens to P (k)?

You can see that as you increase k the probability will increase and then at some point it will stop increasing. P (k+1) >P (k), if and only if. Well, so write this expression twice. Once for k once for k + 1 cancel out things and what will remain? a - k times do you agree? That is what I wanted to hear. This a - k and k + 1 comes from $\binom{a}{k}$ and $\binom{a}{k+1}$ comparison.

And the remaining this n / 2 things come from $\binom{n-a}{n/2-k}$ comparison for k and k + 1. You get these 2 fractions and you just have to compare these fractions. You can simplify this and, let me skip the middle step ultimately you can deduce that if and only if k is less than a - 1 / 2. This will give you exactly this without any loss. Probability expression increases as long as you do not cross k does not cross half of a.

It depends on the others agree with this or not? Well guessing now is not allowed. You does have guessed before it is too late now. Now this so, this is a proof. That is your maximum. We can now deduce that P (k) can never exceed, we should make it kind of equal greater than equal to not equal to less than or equal to and then I will just pick k to be (a – 1) / 2, that their P of k reaches its maximum.

That gives you $\binom{a}{a-1/2}$ divided by $\binom{n}{n/2}$ now, who will estimate this Stirling's approximation but it will really use the most precise Stirling's approximation, because if you look at the main terms they will cancel out but Stirling approximation has this square root terms. Those square root terms will come from this a n - a and n. We will get that this is O($\frac{\sqrt{n}}{\sqrt{a}-\sqrt{n-a}}$)good Stirling's estimate.

It really boils down to these lower order terms kind of Stirling's approximation. And you are now you use this bound another boundary condition that a is not too large. Because if a becomes n then this is infinity, but they will never become n ,we have stopped at the 2n / 3. So this will come out to be $\leq$ O ($\frac{1}{\sqrt{a}}$) because these 2 things this $n/n-a$ is bounded by a constant.

Is that clear? That is the calculation.That was a detour yes, I will finish it next time. For now, remember that we want to compute the probability that X_1X_2 all these variable subsets remain balanced k balanced. In particular k balanced X_i is what we want to study. So this will denote by event E_i. Let E_i be the event. That discrepancy is less than k. And we want to compute the probability that all these events happen.

Even to E_t they happen and as you are choosing Y and Z randomly.This probability is a product of probabilities. This is product of E_1. Then E_2 conditioned on E_1 dot dot probability of E_t conditioned on the previous ones. Next time what we will do is we will basically compute this rate will we give an expression for this and hence get an expression for this product and show that it is as small as claimed in the statement. Then somewhere there is imbalance and just a single imbalance will give you already the fall in the measure.