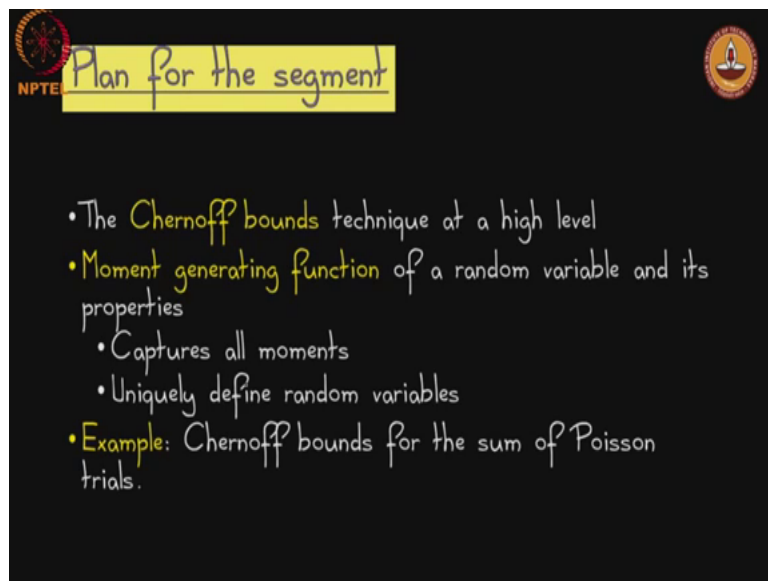


Probability & Computing
Prof. John Augustine
Department of Computer Science and Engineering
Indian Institute of Technology Madras

Module – 03
Tails Bounds I
Lecture - 19

Segment 5: Moment Generating Functions and Chernoff Bounds

(Refer Slide Time: 00:18)



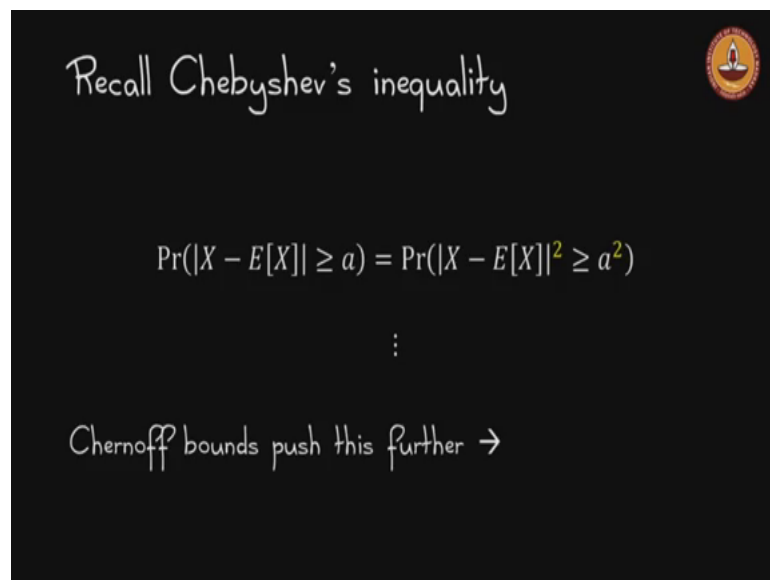
Three and now we are going to see Chernoff bounds it is a family of bounds it is not just one bound, but family of bounds they are tied together by the technique the way we would derive these bounds and an important concept along the way we will try to understand the notion called moment generating functions. These moment generating functions they are quite nice; because they capture all the moments. If you recall the moments the n th moment is expectation of X to the n this one moment generating function is like a bag that just collects all of them together.

So, if you recall the Markov's inequality required the first moment, Chebyshev's required the second moment. Why is Chernoff? So, powerful because it includes information of all the moments and as it turns out if you have all the moments you actually uniquely define the random variable. So, it is this moment generating function actually capture. So, characterizes random variables uniquely. So, that is why this is the significantly more powerful than Chebyshev's inequality and we will work out Chernoff

bounds for the sum of Poisson trials it is basically a generalization of some Bernoulli trials which is what we call binomial distribution it is a slightly more general version for which we will work out the Chernoff bound ok.

So, here is the Chernoff bound technique at a very very high level. Let us set that in the backdrop of Chebyshev's. How did we prove Chebyshev's inequality? We took a random variable X ok.

(Refer Slide Time: 02:14)



Yes, what is X minus E of X the probability of X minus E of X greater than or equal to a well we did not know how to handle that, but we squared it and when we squared if you recall the event stage this essentially the same, but now you are able to apply the second moment right and that is what we did for Chebyshev's ok.

(Refer Slide Time: 02:44)

Chernoff bounds Technique

For any $t > 0$,


$$\Pr(X \geq a) = \Pr(e^{tX} \geq e^{ta})$$

$$\leq \frac{E[e^{tX}]}{e^{ta}} \quad (\text{Markov's inequality})$$

For left tail,
choose $t < 0$

Chernoff bounds technique:

1. Bound $E[e^{tX}] \rightarrow$
2. Choose appropriate t value
3. Derive suitable "Chernoff bound."



We are just going to push this further to the extent possible. What we are going to do is not square it, but take the exponential. So, now, let us say you want the probability that X exceeds a certain quantity a ; what we do is we pick a suitable t which is some quantity greater than 0; and we ask when we say you know we morph this event over here X greater than or equal to a to E to the t X greater than or equal to E to the t a ; so, basically where it takes exponentiating on both sides.

So, the inequality stays the same. So, the event is really capturing the same thing. So, to have equal probabilities; and So you can easily see the analogy between this and Chebyshev's inequality it is essentially the same idea, but here we are pushing into the limits not just squaring it with exponential and the nice thing here is now for the right tail this is what we have here is a right tail X we were asking what is the probability that X exceeds some position a some value a and that is giving you the right end of the tail.

For the left end of the tail all you need to do is pick a t that is less than 0 and when you say pick a t that is less than 0 then you can ask what will happen what you do instead of greater than or equal to is you say what is X was the probability of X less than or equal to a . So, that will be the left tail and for that remember your t is going to be less than zero. So, what will happen over here when you apply this inequality is well you how do you convert the less than or equal to sorry to a greater than or equal to you will raise it to the power E to the tx in this context, but t being less than 0 your what happens over here is you will have a negative in the exponent ok.

So, the inequality will appropriately fall into the greater than and which means that again you can apply Markov's inequality and so on and so forth right and that is exactly what we are doing here. So, for the moment let us not worry about the left tail let us continue to focus on the right tail. So, if you what you can now do is apply the Markov's inequality. So, what we have is a random variable here and we are asking; what is the probability that it is more than some positive constant and you will you can then apply Markov's inequality? So, you see how Markov's inequality is still the basis upon which all the other tail bounds have already arrived, but nevertheless we were able to do fancy things now ok.

So, now, we have to complete the story. So, we need a good bound for this numerator over here the expectation of E to the $t X$ and that is something we will focus our efforts on ok, but if we have the right bounds for the numerator what we can do is we can play with the value of t . So, this is true for any t greater than 0 which means that for the best bounds you may need to play with the choose the right value of t ok.


And then hopefully you know the terms will cancel out nicely; and you will get a bound that is a Chernoff bound. Basically, this is a technique a general technique to deriving these sort of bounds and any bounded derived using this general technique of taking the random variable exponentiate; and asking what is the probability that it exceeds a particular value; then taking the exponentiation on both sides of that event and applying Markov's inequality and bounding the numerator. So, these this is a very standard technique this technique is called the Chernoff bound technique and any bound derived from it is the Chernoff bound ok.

So, just to recall, now we have to worry about this expectation of E to the $t X$ ok.

(Refer Slide Time: 07:15)

Moment Generating Function

The moment generating function of a random variable X is

$$M_X(t) \triangleq E[e^{tX}].$$


And as it turns out that fits very nicely with a well known notion called the moment generating function what is the moment generating function of a random variable X it is exactly what we want it is denoted M_X whether this has to be a capital X M_X of t is defined this equal to the delta on top of it is defined is defined as the expectation of E to the tX , this is exactly what we want and we have a quantity, we have a well defined notion the moment generating function. Let us try to understand this object this moment generating function.


(Refer Slide Time: 07:51)

Theorem: $M_X(t)$ captures all moments

Under the assumption that exchanging expectation and differentiation operands is legitimate, for all $n \geq 1$,

$$E[X^n] = M_X^{(n)}(0),$$

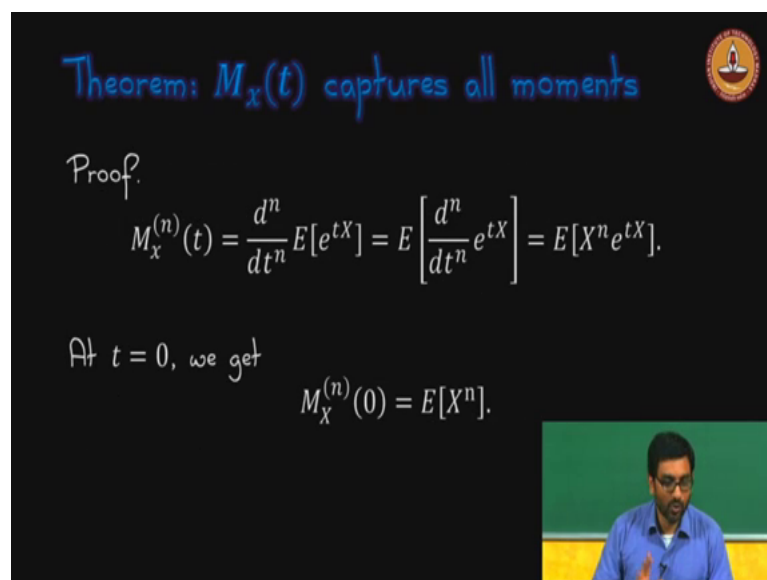
Where,

$$M_X^{(n)}(0) \text{ is } \frac{d^n}{dt^n} E[e^{tX}] \text{ evaluated at } t = 0.$$


First theorem about the moment generating function; it captures all the moments this is something that we talked about a little while ago right the Chernoff bounds why is it. So,

powerful it is capturing all the moments not just the first in a second moment. So, why is that the case and it is under some conditions. So, here the condition is that when we have expectation and differentiation we should be able to interchange there the way we apply them. If we can interchange them then we have this theorem ok. So, basically what are we claiming here; this expectation of X to the n this is the n th moment you simply get that by taking the n th derivative of the moment generating function; this n this is the n th derivative of the moment generating function. Then evaluate it at the position 0 ok. Then you basically by doing that you get the n th moment of the random variable X ; that is the nice thing. So, basically what is happening here is that the moment generating function in it is its somehow capturing all the moments when you take as you take the derivatives each derivative evaluated at 0 gives you the appropriate n th moment.

(Refer Slide Time: 09:20)



Theorem: $M_X(t)$ captures all moments

Proof.

$$M_X^{(n)}(t) = \frac{d^n}{dt^n} E[e^{tX}] = E \left[\frac{d^n}{dt^n} e^{tX} \right] = E[X^n e^{tX}].$$

At $t = 0$, we get

$$M_X^{(n)}(0) = E[X^n].$$

The proof is quite simple; you look at the n th derivative of the moment generating function and remember this one important thing to notice is that you are you are taking the n th derivative not with respect to X , but with respect to t ok. So, dn by $d dt$ to the n . So, that is the end of the derivative of well; what is the moment generating function? It is expectation of E to the $t X$ right. So, you apply that and remember we have said that differentiation and expectation can be interchanged. So, I am just interchanging it and taking the differentiation inside and taking the derivative. So, here derivative the n th derivative of E to the $t X$ is X to the n E to the $t X$ this is remember definite the derivative with respect to t ok.

And now you; so now that you have this form you can just up start applying or evaluating this expression at t equal to 0. So, when you evaluate at t equal to 0 essentially this term will vanish away you will get E to the 0 X and that is just a 1 ok. So, you will just be left with E to the expectation of X to the n and that is what we have over here. So, the nth derivative evaluated at 0 is expectation of X to the n which is the end.

(Refer Slide Time: 10:40)

Bernoulli Random Variable

$$M_X(t) = E[e^{tX}] = pe^t + (1-p) = 1 + p(e^t - 1) \leq e^{p(e^t - 1)}$$

$1 + x \leq e^x$

$$M_X^{(1)}(t) = pe^t, \text{ so at } t = 0, M_X^{(1)}(0) = p = E[X].$$

$$M_X^{(2)}(t) = pe^t, \text{ so again at } t = 0, M_X^{(2)}(0) = p = E[X^2].$$

Let us look at one more one specific example where this is applied ok. So, now, let us go back to our favourite Bernoulli random variable. We just want to confirm this; basically what are we doing here.

(Refer Slide Time: 10:53)


Theorem: $M_x(t)$ captures all moments

Under the assumption that exchanging expectation and differentiation operands is legitimate, for all $n \geq 1$,

$$E[X^n] = M_x^{(n)}(0),$$

Where,

$M_x^{(n)}(0)$ is $\frac{d^n}{dt^n} E[e^{tX}]$ evaluated at $t = 0$.



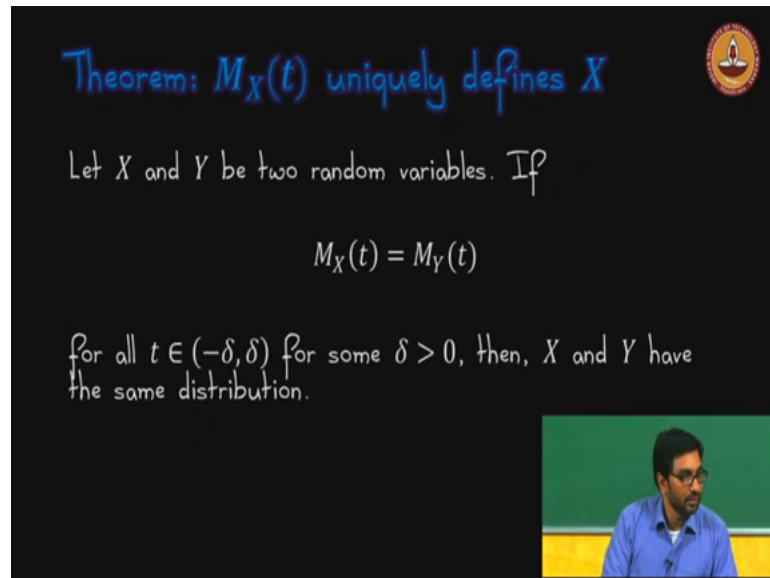
We are taking the n th derivative up evaluating at it 0 we need to be able to get the moments ok; and let us just test it out for a very simple case this is the simplest possible case that you can think of. So, what is the moment generating function that is expectation of E to the tX and that; what is this expectation of E to the tX ?

Well, essentially the variable E to the tX what is the probability I mean with probability p it is going to take the value E to the t , because X remember this is E to the tX is defined based on X ; and X will be 1 with probability p and 0 with probability 1 minus p when X is 1 you are going to get E to the t 1 which is just E to the t that is. So, you get $p E$ to the t plus when X is 0 you are going to get E to the 0 which is a 1; so you when that happens with probability 1 minus. So, that is how you get this expression for the moment generating function for; and then you can play with it a little bit and apply the fact that $1 + X$ is at most E to the X and you can get it to be of the form E to the p times E to the t minus, but the most for now the form that we want is this what was what is shown in yellow ok.

So, now we have a moment generating function. Let us take the first derivative and when we take the first derivative of this expression this is with respect to t . So, you will just be left with $p e$ to the t and when you evaluate it at t equal to 0 E to the t becomes a one and. So, you are left with p and that is expectation of X if you recall this is fitting exactly the theorem that we have just proved if you take the first derivative up and apply it at value equal to 0, you get back expectation of X you can do the same thing with the second derivative ok. I will leave you to work that out on your own time, but essentially the

second derivative I mean the second moment is for the Bernoulli random variable is going to continue to be p and you are going to get the same.

(Refer Slide Time: 13:05)



Theorem: $M_X(t)$ uniquely defines X

Let X and Y be two random variables. If

$$M_X(t) = M_Y(t)$$

for all $t \in (-\delta, \delta)$ for some $\delta > 0$, then, X and Y have the same distribution.

So, the other thing is that not only does it capture all the moments when it captures all the moments it fully defines the random variable and this we are just going to state the theorem and going to skip the proof. So, what is the theorem stating if you have two random variables have X and Y ; and if they are moment generating functions in the vicinity of 0 what does that mean when t is around 0 it is between my delta and delta if both those moment generating functions are exactly equal, then we are guaranteed that both X and Y are essentially the same distribution same essential the same ok. So, this is this tells you once I mean if anything this tells you that the moment generating function really gets to the heart of what a random variable is really can characterize random variable and one more nice theorem before we get back in the Chernoff bounds ok.

(Refer Slide Time: 14:00)

Theorem: Independent Random Variables

If X and Y are independent random variables, then,

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t).$$

Proof.

$$\begin{aligned} M_{X+Y}(t) &= E[e^{t(X+Y)}] \\ &= E[e^{tX} \cdot e^{tY}] \\ &= E[e^{tX}] \cdot E[e^{tY}] \quad (\text{by independence}) \\ &= M_X(t) \cdot M_Y(t). \end{aligned}$$

What is this if you this again is a way to exploit independence of random variables ok. So, you have two random variables X and Y . X plus Y itself is a random variable. So, there is this well defined notion of the moment generating function of X plus Y and that is this left hand side right ok. The nice thing is that is just the product of the individual moment generating functions; when X and Y are independent this is not true when X and Y are not independent, but when they are independent you have this ok. This is very useful because later on when we are trying to bound say the binomial random variable or sum of Poisson random variables or whatever you will exploit independence and this is a key crucial requirement for that ok.


So, you why is it ok. So, you take the moment generating function for X plus Y you apply the formula that is nothing, but the expectation of E to the t times X plus Y and you expand it out you get E to the t X times E to the t Y , but the nice thing is now these are two random variables E to the t X and E to the t Y if X and Y are independent then E to the t X and E to the t Y are also independent and if they are independent there the expectation of their product is the product of their expectations and that is what we are over here and that is by independent which ok. What is each of these multiplicative terms? It is nothing, but the moment generating function of X and is this clear what we have talked about. So, far; any questions.

(Refer Slide Time: 15:52)

Sum of Poisson trials

- Recall: Binomial distribution is sum of Bernoulli trials. Here, all Bernoulli trials have same p values.
- Sum of Poisson trials $X = \sum_{i=1}^n X_i$ is more general. Each Poisson trial variable X_i has its own p_i value.
- Now towards Chernoff bounds for X .

→



Now, we are ready to get back into the Chernoff bounds we will work out a particular example it is basically the sum of Poisson trials this is if you recall what is the binomial distribution it is the sum of n Bernoulli trials the sum of Poisson trials it is just a generalization it is just the techniques. So, everything that we say from now on; with respect to Poisson trials is going to be applicable to Bernoulli binomial distribution as well, but; in fact it is going to be a one step more general ok.

What is the general part of it? In when you take the binomial distribution each Bernoulli trial must have the same probability of success p when we go in to sum of Poisson trials we are generalizing we are allowing each Poisson random variable X_i to be 1 with some probability p_i and 0 with some with probability $1 - p_i$ and these p_i can be different for each X_i . So, what is X equal to X_1 plus X_2 plus and so on up to X_n ; each one can have its own probability of success ok. So, that is the general nature of some of us on trial. So, of course, if all of these p_i are equal, you get back binomial distribution. Now, that we have established that to the we know what the Poisson some of us on trials is essentially a binomial distribution generalized a little bit. Let us workout Chernoff bounds for this X which is the sum of Poisson trials ok.

(Refer Slide Time: 17:29)

Moment Generating Function $M_X(t)$


$$M_X(t) = \prod_{i=1}^n M_{X_i}(t)$$

$$\leq \prod_{i=1}^n e^{p_i(e^t - 1)}$$

$$= \exp\left\{\sum_i p_i(e^t - 1)\right\}$$

$$= e^{(e^t - 1)\mu}$$

Handwritten notes: $E[X_i] = p_i$, $\sum_i p_i \rightarrow E[X] = \mu$



So, we will need remember to apply the Chernoff technique; we will need E to the t X which means that we need the moment generating function ok. So, the nice thing now we immediately start applying the theorem based on independence. If you want the moment generating function for X , it is basically the product; because these individual X_i s are independent. The moment generating function for X is the product of the moment generating functions for the individual X_i s. So, what is the moment generating function for the individual X_i ? If you recall we worked this out a little while ago; when we worked out the Bernoulli trial with respect to probability of success p ; we worked it out to be of the form e to the p_i times E to the t minus one a few slides ago.

So, we are just applying that and so now, you have the product of several exponentials which means you can take the e to the you can take the exponential of their some of their some of their this product of the exponentials is just e to the sum of the X . And well what is that this e to the t minus one and I think this needs to be slightly clarified this it looks like this e to the t minus one is in the subscript over here it is actually p_i times e to the t minus one this; this e to the t minus one can come out of the summation because the summation is with respect to i .

So, you get e to the e to the t minus one if that if this comes out you will be left with summation i p_i what a summation i p_i . What is p_i ? P_i corresponds to random variable X_i ; expectation of X_i equal to p_i . So, what is summation of over i p_i ; expectation of X . And; and we are going to denote that by μ . So, we have got the moment generating function; which means now we can go back to the Chernoff technique and apply this.

(Refer Slide Time: 19:57)

Theorem: Chernoff bounds for X

- For any $\delta > 0$,
$$\Pr(X \geq (1 + \delta)\mu) \leq \left(\frac{e^\delta}{(1 + \delta)^{1 + \delta}} \right)^\mu$$
- For $0 < \delta \leq 1$,
$$\Pr(X \geq (1 + \delta)\mu) \leq e^{-\frac{\mu\delta^2}{3}}$$
- For $R \geq 6\mu$,
$$\Pr(X \geq R) \leq 2^{-R}$$

So, let me let us before we start applying it; let me state the bounds that we can get only one bound I am going to prove, the rest of them you you will have to proof your own I mean the textbook has it. So, you kind of have to take some time to work through them ok.

The first and if you recall; I said there are there is this one framework or technique the Chernoff bounds technique and you can derive multiple bounds. So, that is why we are going to give multiple bounds now ok. Now the first bound for any delta greater than 0 the probability; that X exceeds 1 plus delta times mu. So, just to give you give yourself a picture for what is going on here. You have a random variable X this is E of X equal to mu; you have this sum distribution and you are asking well and let us say this point is a 1 plus delta times mu. We are asking what is the probability that X will fall in this right tail; and that is at most this right hand side quantity; E to the delta divided by 1 plus delta raised to the 1 plus delta the whole raised to mu looks complicated, but there are nicer forms this is why it helps to have multiple Chernoff bounds.

So, let us look at the second form. We have here we need to restrict delta to be at most 1. What is probability X greater than 1 plus delta times mu again? Is the same story you are you are basically trying to bound the right tail it is e to the minus mu delta square by 3 Y and you hope this might look like just some mathematical jargon, but let us pause for a moment to see why this is powerful in the binomial distribution; what is your mu? Mu is

the average right. So, if you have like 10000 coin tosses what is your μ it is 5000 and where does in the tail bound probability where does μ appear μ appears in the exponential and it is e to the minus μ ok.

So, as the as the mean increases the probability decreases exponentially ok. So, that is the power of this Chernoff bounds; the probability of deviating from the expectation this tail bound it drops exponentially. So, in that example this coin flipping example that we have worked out in the past it this it will show up as e to the minus 5000 times δ^2 by 3. So, these δ^2 and 3 are also relatively small quantities, but you can immediately see as the number of coin flips goes larger and larger your probability of deviating significantly from the mean drops very very quickly ok.

So, hopefully this intuition is making sense to you because that is very very important this is this is what is telling you about the power of the Chernoff bound and one more convenient form which it is.

Student: (Refer Time: 23:32).

Well, it is positive for the numerator it is positive also for the denominator and if the denominator is dominating then it is going to be again having the same what is a third form for some R greater than 6μ . So, basically this is useful when you want to ask what is the probability of deviating beyond 6 times expectation sum in some cases this makes sense again you see this exponential drop probability if X exceeding R is at most 2 raised to the minus R .

Student: From the (Refer Time: 24:11).

Yes. So, that makes sense particularly for things like normal distribution.

Student: Yeah.

Things like that here it; so when you work with the standard deviation you are actually working with Chebyshev's inequality or some related equality and you can do that and it is usually done, because you do not have a handle on the moment generating function this is what you are saying is particularly true for a lot of fields, but in computer science; we often we define the algorithm which means that we control what random variable? What the random variables going to look like? Which means that we have more power to

in our hands? We can go in to Chernoff bounds and that is what will happen very commonly; we end up being able to exploit Chernoff bounds.

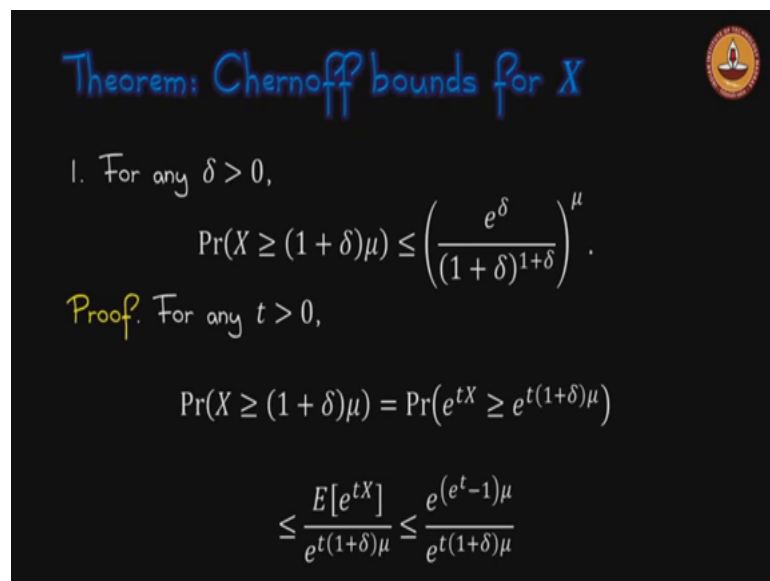
Because we have access to all the moments and therefore, implicitly we do not explicitly think about it all the time implicitly we have access to all the moments and so we will be applying the more powerful techniques. And this is significantly more powerful than just asking how many times away from; how many standard deviations away from the mean you are? So, this is I think your question is also motivated by this right.

Student: Yeah.

This is just one convenient form; it might be useful in some cases where the mean is small. If the mean is small then you can ask; what is the probability that you are exceeding 6 times the small mean? Then it might be a this might be a convenient form to use ok. And again goes to say that it is basically the same technique all these three bounds are derived using the same technique; it is just playing with the value of t, if you recall there is a t a parameter t showing up in the derivation of Chernoff bounds displaying with a value of t and other mathematical minor mathematical juggleries you get these convenient forms.

And so when you think about applying them; you just choose the most appropriate one and typically the second form is the most commonly used form of this.

(Refer Slide Time: 26:17)



Theorem: Chernoff bounds for X

1. For any $\delta > 0$,

$$\Pr(X \geq (1 + \delta)\mu) \leq \left(\frac{e^\delta}{(1 + \delta)^{1 + \delta}} \right)^\mu$$

Proof. For any $t > 0$,

$$\Pr(X \geq (1 + \delta)\mu) = \Pr(e^{tX} \geq e^{t(1 + \delta)\mu})$$
$$\leq \frac{E[e^{tX}]}{e^{t(1 + \delta)\mu}} \leq \frac{e^{(e^t - 1)\mu}}{e^{t(1 + \delta)\mu}}$$

And as I said this the same technique can also be applied in the in the left tail as well we will see that, but for now let us actually look at the proof. So, just to be clear all these three have the proofs in the textbook. I am only going to go through the first inequality. So, that we have one inequality nailed down; we know how to do it, but the other things I still expect you to go through it on your own so.

We are now focusing on the first inequality. X greater than $1 + \delta$ times μ and here δ is any value any positive value; and we will not need to prove that it is at most this right hand side ok. And again we are going to apply the exact Chernoff bound technique that we have already seen before. We are just going to say what is X greater than probability that X greater than $1 + \delta$ times μ well; you take the exponentiation on both sides probability of e to the $t X$ greater than e to the t times $1 + \delta$ times μ is standard technique. Apply the Markov's inequality. So, that is you get expectation of e to the $t X$ divided by the right hand side of this event ok.

So, but we already know this; we know remember we just a few slides ago we derived the moment generating function for the random variable X that is nothing, but e to the e to the t minus 1 times μ if you recall we did that ok.

So, I am just applying that over here. So, you get this you get this form over here. Now comes the trick because now you have to choose a particular value of t that will be convenient for you and you see a lot of e to the t s and if you if you want to cancel things out what do you do when you have E to the some lon of something you the e and the lon cancel each other out, right.

(Refer Slide Time: 28:17)

Theorem: Chernoff bounds for X


Proof contd.

If we set $t = \ln(1 + \delta) > 0$ (since $\delta > 0$),

$$\Pr(X \geq (1 + \delta)\mu) \leq \frac{e^{(e^t - 1)\mu}}{e^{t(1 + \delta)\mu}}$$

$$= \left(\frac{e^\delta}{(1 + \delta)^{1 + \delta}} \right)^\mu$$

Inequalities 2 & 3 are left as HW.



So, that is the; that is what we are going to exploit. So, we are going to set t equal to $\ln(1 + \delta)$ and because δ is greater than 0 this $\ln(1 + \delta)$ is also going to be greater than 0.

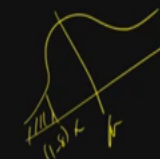
So, which means we can apply it. So, going back to this form; if you apply if you apply t equal to $\ln(1 + \delta)$; so essentially here you will be you will be getting $\ln(1 + \delta)$. So, what will you get over here you will get e to the e to the t will become $\ln(1 + \delta)$ you know it will it will e to the $1 + \delta$ will become $1 + \delta$ minus 1. So, it will be $1 + \delta$ minus 1 divided; now I M going to take the whole to the μ separately and here e to the t will become $\ln(1 + \delta)$. So, and this $1 + \delta$ will be posted over here; which is exactly the form that we; so this is the first technique. The other techniques in similar fashion will be derived and at; in fact, in this case you look at the other you basically play with this form. This is one of the best forms that you can get the other bounds are obtained by displaying with;

So, those are left as homework for you. So, now, let us any questions before I move on and how this derivation works out.

(Refer Slide Time: 29:58)

Theorem: Left Tail for X


For any $0 < \delta < 1$,



$$\Pr(X \leq (1 - \delta)\mu) \leq \left(\frac{e^{-\delta}}{(1 - \delta)^{1-\delta}} \right)^\mu$$

and

$$\Pr(X \leq (1 - \delta)\mu) \leq e^{-\frac{\mu\delta^2}{2}}.$$



This is just two forms for the left tail, that we have; again they look quite similar because they are derivations are essentially of the same pattern. Now because it is the left tail you need to consider delta values to be in the range 0 to 1 and you have these two forms again here it is X less than or equal to $1 - \delta$ times μ . So, so you will essentially what you are doing is this is your μ ; this is going to be your $1 - \delta$ times μ .

So, you are asking what is the probability of this left tail. That is again at most some quantity which is if you notice it looks very similar to the type of bounds that we have and the second form is also their probability X at most $1 - \delta$ is at most e to the minus $\mu \delta^2$ by two and usually this form is more useful and. In fact, you can take the two left in the right tail the second form of them this.


(Refer Slide Time: 31:03)

Corollary: Both tails

For any $0 < \delta < 1$,

$$\Pr(|X - \mu| \geq \delta\mu) \leq 2e^{-\frac{\mu\delta^2}{3}}.$$

Proof is simply by the Union Bound.



In this case; it is this one we have something similar in the right tail as well we can put them both together and apply the union bound and you can get a combination of both right and left tails. So, this is here I because it is combination of right hand left tails you have the absolute value of X deviating from the expectation to be greater than some delta times mu is at most 2 E to the minus mu delta square by 3.

Student: (Refer Time: 31:28).

What is the time?

Student: divided by 2 or 3.

It is divided by 2; it is this happens to not be a typo, but when you apply the union bound you are just a little bit relaxed about how you apply it. The first one is I believe the tightest, but it really is not. So, much about tightness in these cases because the tightness usually just gives you some I mean typically the benefits just in constants the you usually use the more appropriate question would be which one is the most convenient one of the use of course, in some cases one might be convenient, but might not give you the good enough bound, right.

So, the first one is usually the tightest because it is pretty close to what you want, but what is also and the reason why I even went through the proof of the Chernoff technique is that there are some cases where you actually have to redo the proof of the Chernoff technique for a specific t value that fits the application. Remember this here we applied a

particular t value that; that nicely cancels things out and it is great for a general form, but there are applications for which you really have to go through the derivation and choose the appropriate t value and all that.

(Refer Slide Time: 32:54)

Example: Coin Flips

Let X be the number of heads out of n coin flips.

$$\Pr\left(|X - \frac{n}{2}| \geq \left(\frac{1}{2}\right) \sqrt{6n \ln n}\right) \leq 2 \exp\left\{-\left(\frac{1}{3}\right) \left(\frac{n}{2}\right) \left(\frac{6 \ln n}{n}\right)\right\}$$

$$= \frac{2}{n}$$

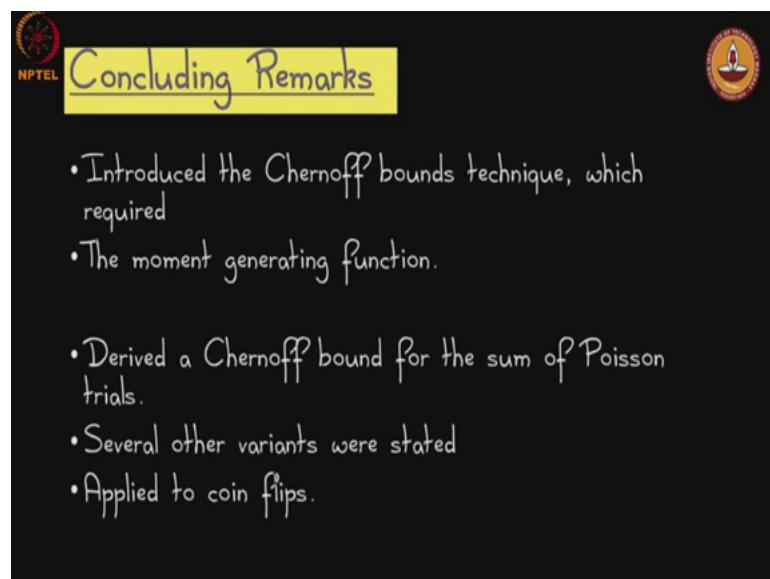
So, it is good to know the general derivation technique. So, we are now almost done; only thing I want to point out now is, how this Chernoff bounds that we have seen so far fares with respect to coin flips? This is the example that we have seen. And here; now we are going to ask what is the; so X is the number of heads out of n coin flips ok. So, this is essentially the binomial distribution and since we know even how to handle Poisson; some of Poisson trials you we can actually handle the binomial distributions just a special case.

And so we want to ask; what is the probability that; so in the particular example that we had n was 10000. So, the mean would have been something like 5000 right. And we are asking what is the probability there it deviates away from the mean by this quantity ok? And so let us work that out. So, this is this is just this is recall this is if you ignore the constants essentially what you have is a square root of $n \ln n$ ok. So, it is basically if you for a rough understanding of what we are talking about here? What is the square root of 10000? well it is roughly a 100 right. So, we are asking what is the probability that X deviates from the mean by more than roughly a 100 ok. If you just you can ignore the constants in the \ln term; \ln is small constants are also small and let us see.

How do you work this out? So, here on the right hand side you need it to be of the form δ times μ right. So, you should work out the appropriate. So, we on the left or what we have over here is; what is exactly the tail bound that we want? We want to understand; what are these probabilities these tail probabilities? Ok, but now we have to fit it to the Chernoff bounds form ok. So, now, what is your μ ? μ is $n/2$. So, your δ if μ is $n/2$; what you have to do? To get the μ ; to get the δ is just multiply by $n/2$ and also multiply by $2/n$. If you do that you get this will be your μ this $1/n$. So, basically let me write that a little bit more carefully over here what you do is you multiply by $n/2$ and you also multiply by $2/n$ ok.

This $n/2$ is nothing, but your μ . Whatever is left is your δ ok. So, that so now we have all other δ and μ . So, we can apply the formula and if you recall the formula it is $2e^{-\mu\delta^2/3}$. So, that is what you are applying over here ok. And when you apply you get $2/n$. And for a similar range what did we get out of Chebyshev's? I think we got something like $1/9$ or something like that the probability of exceeding 5000 plus or minus 160 we got was something like $1/9$, but now what are we getting we are getting $2/n$. So, this is like $2/10000$ ok. So, you see; how the tail bound tail probability is much tighter when you use the Chernoff bounds. So, that is generally the idea here wherever possible you want to use the Chernoff bounds, but when it is not possible you live with the Chebyshev's or the Markov's inequality.

(Refer Slide Time: 37:33)

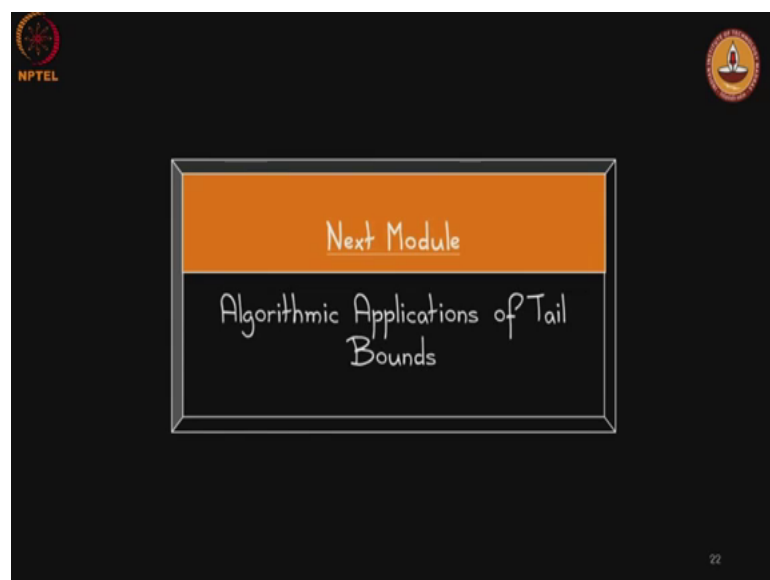


Concluding Remarks

- Introduced the Chernoff bounds technique, which required
 - The moment generating function.
- Derived a Chernoff bound for the sum of Poisson trials.
- Several other variants were stated
- Applied to coin flips.

So, with that we can conclude our segment for today. So, we introduce the Chernoff bounds technique and you know along the way, we picked up an understanding of moment generating functions we looked at some properties and we worked out the derivation of a Chernoff bound for some of Poisson trials at least one form of it there were other forms that we have stated we have not actually looked at the derivations and what we did was we applied it to the coin flips and we saw it is immediate how powerful Chernoff is compared to Chebyshev's I hope you can appreciate this difference. So, that is what we have seen so far and in this module we have seen two slightly larger segments. So, we have seen the analysis of the median algorithm in the last segment which again like slightly longer module and this Chernoff bounds technique again is slightly longer module five segments in this module 3 ok.

(Refer Slide Time: 38:42)



So, in the next module what we are going to do is we are going to focus on applications we are going to now that we know all these bounds these markers in equality Chebyshev's and Chernoff's and things like that what we are going to do is start looking at some algorithmic contexts and we are going to start applying them. So, hopefully these bounds you know come alive to you. So, you will actually see how they help in a computer science ok. So, that will be the topic for next module ok.

Thanks.