Approximation Algorithm

Prof. Palash Dey

Department of Computer Science and Engineering

Indian Institute of Technology, Kharagpur

Week – 09

Lecture 41

Lecture 41 : Chernoff Bound

Welcome, in this lecture we will study a very powerful class of probabilistic inequalities and they are called Chernoff bounds. And then we will see how this Chernoff bounds can be used for designing randomized approximation algorithms by randomized rounding of linear programs. So, today's topic is review of Chernoff bounds. So, it is not one inequality, but a class of inequalities which are called Chernoff bounds collectively. So, in essence all these inequalities Chernoff bounds essentially show that it is very likely that the sum of n independent $\{0,1\}$ random variables are not far away. from the expected value from the expectation of the sum.

So, this is the core idea or message of all these Chernoff bounds. So, let us first see the most basic version and most useful version of Chernoff bound. most popularly used theorem. Let $X_1, X_2, \ldots, X_n$ be n independent this is very important $\{0,1\}$ random variables, this kind of random variables are also called Bernoulli random variables, they take value 0 with certain probability and value 1 with the remaining probability.

So, for Chernow bound this n random variables need to be independent that is very important although the $\{0,1\}$ this can be relaxed to certain extent actually it can be relaxed to any bounded random variable, but the proof and the idea is similar. So, let us first see the most basic version which is $\{0,1\}$ independent random variables, but they need not be identically distributed. not necessarily identically distributed then for if you denote the $\sum X_i$ if this is the sum random variable and μ is expectation of X and we have a lower bound and upper bound on μ. So, L is a lower bound on μ and U is an upper bound on μ and δ greater than 0 for this what we have is 2 inequalities probability that X takes value greater than equal to $(1+\delta)$ times the upper bound is less than equal or strictly less than $\left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{U}$.

This is the probability that X takes more than $(1+\delta)$ times upper bound value. And the lower side also we have similar inequality probability that X is less than equal to $(1-\delta)L$ this is less than $\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^L$ ok. So, this is the most common version of Chernoff bound which bounds the probabilities that X takes value more than $(1+\delta)$ times upper bound of the mean and probability that X takes value $(1-\delta)$ times lower bound of the mean. A very easy generalization of standard generalization is to generalize this $\{0,1\}$ random variable to $\{0,a_i\}$ random variable. So, we have so, let us write that theorem we have $0\le a_i\le 1$ for all $i\in[n]$.

ok and the everything else remains same only is X_i is $\{0,a_i\}$ random variable ok. That means, it takes value 0 with certain probability and value a_i with remaining probability everything else remaining the same as earlier theorem. what we have? We can prove the same bound that probability that X is greater than equal to $(1+\delta)U$. is less than $\left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^U$ the probability that X is less than equal to $\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^L$ ok. So, we will prove this more general Chernoff bound this is theorem 2 let us call it theorem 2 and the earlier one is theorem 1.

So, theorem 2 implies theorem 1. So, we will prove theorem 2, but for that we will need a more basic probabilistic inequality which is called Markov's inequality let us write Markov's inequality. let X be a non-negative random variable that means, it takes negative value any negative for any negative value it takes ah that value with probability 0 for discrete random variable. For continuous random variable it means that entire probability mask is on positive part of real line ah. Then for any a positive probability that X takes value greater than equal to a is less than equal to $\frac{E[X]}{a}$ ok.

So, proof I am leaving it as homework easy proof. you may assume that X is discrete random variable taking only finitely many values which is enough for our purposes proof is homework ok. Now, with this Markov inequality let me prove theorem 2 proof of theorem 2. So, in theorem 2 we have two inequalities one is upper bounding X another is lower bounding X again we prove only one say the first one because the second one the proof is exactly analogous. So, proof of the first inequality ok.

First you observe that if expectation of X is 0 there is nothing to prove. If expectation of X is 0 because X is a $\sum X_i$ and each X_i takes value $\{0,a_i\}$. If expectation of X is 0 that means, X takes value 0 with probability 1 that means, each X_i takes value 0 with probability 1. So, if expectation of X is 0 then X equal to 0 and the bound the inequality holds ok. So, assume without loss of generality that expectation of X is greater than 0.

if expectation of X is greater than 0, then that means, there exists an $i \in [n]$ such that expectation of X_i is greater than 0. So, we ignore all X_i whose expectation is 0. we ignore all X_j $j \in [n]$ such that expectation of $X_j$is 0. So, without loss of generality we can assume by renaming that expectation of X_i is greater than 0 for all $i=1,\ldots,n$. So, assume without loss of generality that expectation of X_i equal to p_i which is probability that X_i takes value or let us assume expectation of X_i is $a_i p_i$ which is a_i times the probability that X_i takes value a_i.

that means, p_i equal to probability that X_i takes value a i this is greater than 0 for all $i \in [n]$ ok. So, then what is expectation of X? Expectation of X is $\sum_{i=1}^{n} a_i p_i$. Now, for any t greater than 0 probability that X is greater than $(1+\delta)\mu$ because t is positive this is same as the probability that tX is greater than $t(1+\delta)\mu$ I do not want to write μ I want to write you see here I am interested in probability of this event. Now I take exponent on both side because e^x is an increasing function this probability remains same increasing strictly increasing function and continuous e^{tX} is greater than $e^{t(1+\delta)\mu}$.

Now, we apply Markov's inequality and using Markov's inequality I write that this is Markov's inequality on e^{tX} treating it as a random variable this is expectation of e^{tX} by $e^{t(1+\delta)\mu}$. So, in the remaining part of the proof we will find out e to the what is expectation of e^{tX} a good upper bound of it and plug the value here and choose a value of t which gives the best inequality. So, these are the two steps which will give us the required bound. So, what is expectation of e^{tX}? This is expectation of e^{tX} is $\sum_{i=1}^{n} X_i$. This is expectation of $\prod_{i=1}^{n} e^{tX_i}$.

Now, because x i s are independent that is a very crucial assumption in Chernoff bound, this I can push this expectation inside the product, this is where I am using independence of $X_1,\ldots,X_n$ expectation of e^{tX_i}. Now, what is expectation of e^{tX_i} let us compute it. X_i takes value 0 X_i takes value a_i with probability p_i. So, this is with probability p_i the value is value of this function is function is $e^{t a_i p_i}$ plus X_i takes value 0 with probability $(1-p_i)$ and X_i equal to 0 value of e^{tX_i} is 1.

So, this is $(1-p_i)$ this is $1+p_i e^{t a_i - 1}$. Now, we apply standard inequality that $1+x \le e^x$ which we can which can be shown by any standard calculus based technique. So, this is less than equal to $\prod_{i=1}^{n} e^{p_i} e^{t a_i - 1}$ ok, because this is, but before that let us I want to put this a_i here.

So, let us do that first. So, this is less than equal to $1+p_i a_i e^{t-1}$. This you show that for a i greater than 0 less than equal to 1 $e^{t a_i -1}$ is less than equal to $a_i e^{t-1}$. Again this can be shown by standard calculus of 12th standard.

Now, we apply the inequality that $1+x \leq e^x$ this product $i=1,\ldots,n$. this is e^x $p_i a_i e^{t-1}$. Now, this product becomes sum in the exponent $e^{t-1}\sum_{i=1}^{n} p_i a_i$. This is nothing, but $E[X]e^{t-1}$. Now, μ is a upper bound on expectation of X.

So, this is less than equal to $e^{e^t-1}U$. Now, I want to so, what we have shown here we have shown that expectation of e^{tX} is less than equal to $e^U e^{t-1}$ for all t greater than 0. So, this upper bound we put it here. So, what we get is probability that $X \geq (1+\delta)U$ is less than equal to $e^U e^{t-1}$ by in the denominator we have $e^t(1+\delta)U$. this is $\frac{e^t-1}{e^t(1+\delta)^U}$ ok.

Now, I want to choose $t>0$ which minimizes this choose t greater to minimize these expression ok. Minimizing this function is same as minimizing log of this function because log is an increasing function. So, this is greater than 0 recall we need to pick t greater than 0.

So, by putting $t=\ln n(1+\delta)$ what we get is this is $\left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^U$. How do you choose t? You just minimize this function subject to t and pick that t. So, this is what is what gives you a nice upper bound on this. So, similarly you can show the lower bound also. probability that X takes value less than equal to $(1-\delta)L$ is strictly less than $\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^L$ ok and then, but these expressions although they are tight you know they these are very hard to visualize.

So, for that we use this lemma whose proof is again based on elementary calculus and I leave it to you that for δ in between 0 and 1 both inclusive. We can show that $\left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^U$ is less than equal to $e^{-U\delta^2/3}$ and for 0 less than equal to δ less than 1, we can show that $\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^L$ is less than equal to $e^{-L\frac{\delta^2}{2}}$. which are much more easy to work with inequality. So, this inequality is then we can augment and write this is less than equal $e^{-U\delta^2/3}$ and this is less than equal to $e^{-L\frac{\delta^2}{2}}$ ok. So, let us stop here. Thank you.