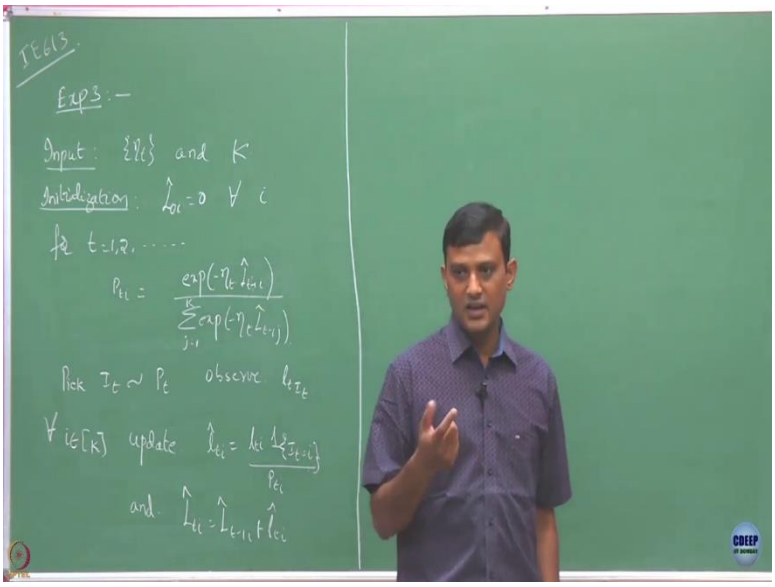
**Bandit Algorithm (Online Machine Learning)**
**Prof. Manjesh Hanawal**
**Industrial Engineering and Operations Research**
**Indian Institute of Technology, Bombay**

**Lecture – 16**
**Regret Bound of Exp3 (Contd.)**

So, let us start. So, today we will going to just complete our discussion on what I have been discussing under adversarial multi arm bandits. So, today we will just complete the proof we left last time, and then going to introduce two variants of that which will help us give a probability not in expectation, but in high probability. And so last time we discussed about this Exp3 algorithm, we said that its regret or pseudo regret can be bounded as like what is that bound some square root n k times log k right, there was some factor of square root 2 there.

So, we needed to show this finally. Before we do that I am going to slightly whatever the Exp3 algorithm we have written the pseudo code, I am going to write it the same pseudo code, but in a bit more compact way than what we have written just for our reference.

(Refer Slide Time: 01:26)




So, what was Exp3 stands for, right, Exponentially Weighted Exploration and Exploitation. So, what was the algorithm? We said input for that is some sequence of t and K, where K is the number of arms right. And then we had an initialization. So, here we have started initializing like t1 be uniform distribution. Instead of that, we will start saying that for all.

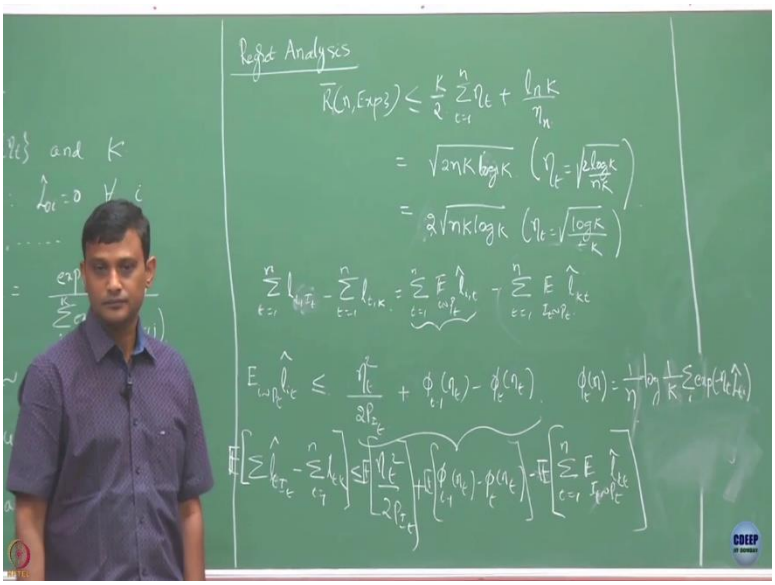
So, what was our notation for Li? Yeah this is the cumulative estimated loss right. So, we will assume that initially it is all 0. So, our notation was 0 here. So, 0 stands for the 0th round.

And then what we will do for we will do. We will come up with this distribution which is pick $I_t$ according to distribution $P_t$; observe $l_{tIt}$, now for all i. So, it is just as same pseudo codes we wrote last time, but this is just like putting in a slightly different way. So, let us understand what is just happening. So, I initialized all cumulative loss to be 0 in the beginning. So, because of this when I started in t equals to first round what is this value? It is going to be 1/k right, and that is true for all i. So, in a way in the beginning, I am giving equal weights to all the arms that is in the first round it is like a uniform distribution.

Earlier I have we have made that explicitly by saying that in the beginning $P_1$ is uniformly distribution, but now I am doing the same thing by initializing my cumulative loss to be 0. So, after that I in the beginning of the every round, I update my probability like this, and then I am going to pick an action $I_t$ according to this distribution, and then I am going to observe the last component associated with that ok.

And then I am going to update the estimates for loss of all arms like this, and then also going to update my cumulative loss like this. It is a same thing as we did earlier. The only difference is we are making the updates right at the beginning, in the previous one we are doing it at the end of my round; but you can see that both are the same.

(Refer Slide Time: 05:40)



So, in terms of the analysis where were? We wanted to show that my pseudo regret of Exp3 is bounded by $\frac{K}{2}\sum_{\{t=1\}}^{n}\eta_t + \log K / \eta_n$ ok. So, if we showed that if it substitutes the value of $\eta_t$ here, this value we got it as $\sqrt{2nklog(k)}$ when $\eta_t = \sqrt{\frac{2\log(k)}{nk}}$. If I said $\eta$ to be always the same, if I know n this is the bound I got. And I also said that this bound is going to be what was that bound 2 $\sqrt{nklog(k)}$ let me writing it as log time if I am going to set $\eta_t = \sqrt{\frac{\log(k)}{nk}}$ in every round right ok.

So, this is all we argued we wanted to now show this ok, this is this was our goal. So, first we made this claim that this can be written as fine, this is what we have argued in terms of our basic inequalities. Then we showed that this first term here what did we show? We are able to show that finally in term using our first step. So, if we recall, we wrote this n terms of the moment generating functions, and then you are able to bound each term in the moment generating functions in step 2 and 3. I am directly going to write that here we basically are able to show that this can be expressed as $\frac{\eta_t^2}{2P_{I_t}}$ $\phi_{t-1}(\eta_t) - \phi_t(\eta_t)$. We are able to express it like this.

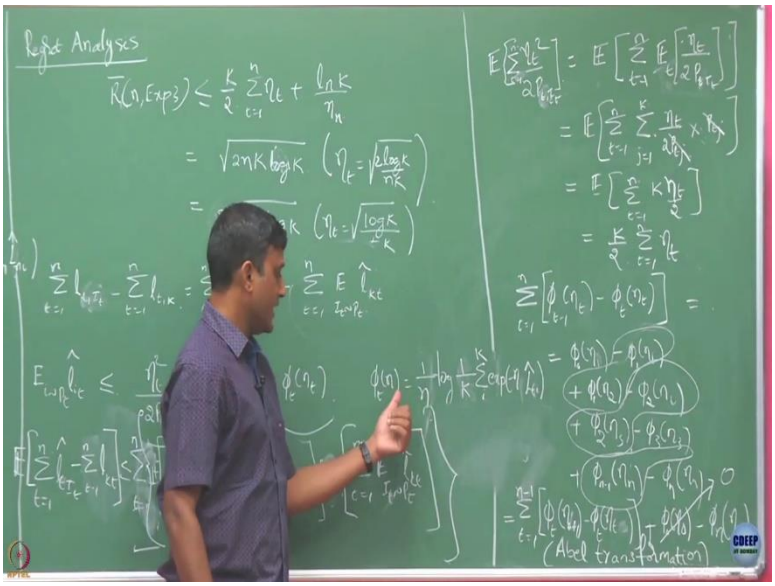
And how did we define $\phi_t(\eta_t)$ at any point t, we just said that this is nothing but $\frac{1}{\eta} log \frac{1}{k} \exp(-\eta_t \widehat{L_{ti}})$. So, this is how we have defined it fine.

Now, let us plug this quantity back here. And I am going to plug this quantity back here. Finally, what I will end up is this difference is nothing but this upper bounded by this quantity and minus this quantity whatever I have.

So, now again writing this I am just going for completeness let me write this fine. So, basically using our step 2 and step 3, we ended up this bound in the last till the last class.

Now, continuing from here what is this, right now, I have given a bound on this quantity, what I am interested in the expected value of this quantity right. So, I am going to take expectation of this quantity. So, when I have that I will end up taking expectation with respect to this, expectation to of this as well as expectation of this quantity ok. Now, let us try to bound expectation of each of these terms.

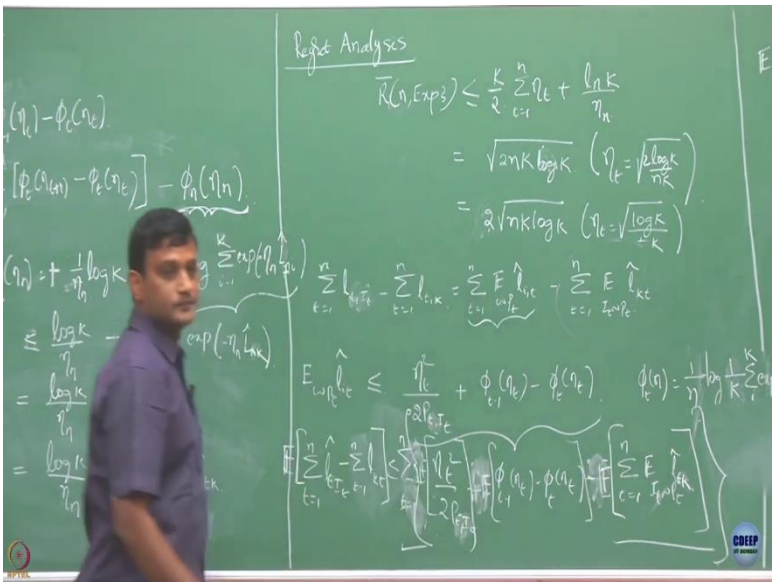
(Refer Slide Time: 13:28)




So, now, when I look into the expectation of this quantity, what is the random thing here? So, remember when I am taking this expectation, I am taking this expectation with respect to the two random quantities. One is with respect to the randomization of the player because he is going to pick an arm according to some distribution in each round, and also the adversary can also randomize his loss vector right.

So, this expectation is with respect to both this randomness. So, one thing I can do is I can split this expectation into two parts; one is this expectation, by the way I missed a summation here right, this should be a when I wrote this it there has to be a summation over here which I missed.

(Refer Slide Time: 14:33)




So, this is summation for all the quantities cost of this summation t equals to 1 to n right; maybe I should write it slightly better; maybe I will just wipe it and write it as and write like this as a summation of all these quantities ok. Is this clear? Why I have to add this summation, because I am doing the summation, alright.

Now, let me do that summation here also. So, basically I am trying to deal with the summation of a first term. So, this is I can always write it as expectation of two quantities; one is with respect to the expectation of random choice of the player and one with the random selection of a losses by the adversary. So, because of that, and sorry this should be $P_{tlt}$ ok. So, is this clear why I have split this expectation into two part?

The inner expectation is a conditional expectation, given the choice of the losses selected by the adversary conditioned on that because the learner is observing the losses that he observed by playing an actions right based on that he is going to update his probability distribution $P_t$. So, given that he will have some probability distribution. So, the inner expectation is with respect to the probability distribution of the learner. Whereas the outside is the expectation with respect to that of the adversary or the environment whatever ok.

Now, what is this distribution is this distribution here the expectation with respect to the $P_t$ , because that is the randomness with which the learner is going to play the actions right. So, if you do that, I will keep it like this. Now, let us try to work out this expectation. What

is this expectation here, I t is a random quantity that will be played according to distribution $P_t$. So, if you look into that, how can I write this, this as $E[\sum_{\{t=1\}}^{n}\sum_{\{j=1\}}^{k}\frac{\eta_t}{2P_{t_j}} \times P_{t_j}]$

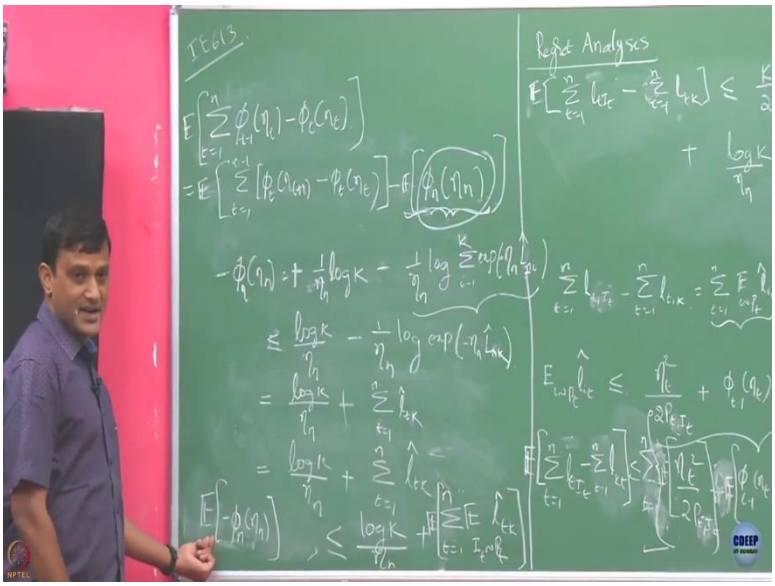
Now, if you simplify this, what are you are going to get? This quantity get cancels; inside term is just a constant. And this will what will this quantity is? it is simply this is going to be K times this quantity or like now this quantity inside term is a constant. This is going to be the same irrespective of what is the choice of losses that adversary would have made, and so I am going to pull this out and simply write it as $\frac{K}{2}\sum_{\{t=1\}}^{n}\eta_t$ ok.

So, now we are able to deal with this term. Now, let us to try to analyze this formula. What is this? We $\sum_{\{t=1\}}^{n}\phi_{t-1}(\eta_t) - \phi_t(\eta_t)$. ok. So, let us try to first expand this and see how this looks like. I am going to just expand. So, this is going to look like $(\phi_0(\eta_1) - \phi_1(\eta_1))$ + $(\phi_1(\eta_2) - \phi_2(\eta_2))$ + $\phi_2(\eta_3) - \phi_3(\eta_3)$ + … , I have just expanded this. What will be the last term? $(\phi_{n-1}(\eta_n) - \phi_n(\eta_n))$ ok.

So, now what I will do is, I am going to club these two terms, and also these two terms, and also these two terms. And I can do keep on clubbing like this right. If I can keep clubbing like this, I can rewrite this summation in a different form that is ok. Can I re-express this summation here in this format, ok? And this is this kind of this is called usually Abel transformation ok. And by the way the way we have defined this $\phi_0$, the way we have defined $\phi_0$ function what is $\phi_0$ is going to be?

If you take $\phi_0$ that is if you substitute t equals to 0, all this losses are initialized to be 0. So, this all terms are going to be 1, this summation is going to be K. So, this entire summation divided by K is going to be 1, and log of 1 is going to be 0. So, this $\phi_0$ this term is going to be 0, right. And what we will end up with the remaining part here.

(Refer Slide Time: 23:36)



I am just going to be write that $\sum_{\{t=1\}}^{n} \phi_{t-1}(\eta_t) - \phi_t(\eta_t)$. We have just shown this to be equal to or maybe other way round, it is $\sum_{\{t=1\}}^{n-1}[\phi_t(\eta_{t+1}) - \phi_t(\eta_t)] - \phi_n(\eta_n)$ ok. So, they are just like this bound we are trying to play with each of these terms ok.

Next, let us try to see what is this quantity $\phi_n$. At least we know then I am going to deal with the last bound n, my $\eta_t$ the way I have defined $\eta_t$, it is going to be simply $\sqrt{\left\{\frac{\log k}{nk}\right\}}$, for t equals to n. So, let us substitute this and see what we are going to get here. So, $\phi_n(\eta_n)$ is going to be, I am going to just separate out this the first term is going to be what? $-\frac{1}{n}\log K$, I am just taking this part and other part is going to be $-\frac{1}{\eta}\log \sum_{\{i=1\}}^{n} \exp(-\eta_n \widehat{L_{ni}})$

So, it is going to be eta n right, because we are taking it at t equals n, because it is $\phi_n(\eta_n)$ ok. So, fine, it should be like well, I am going to take it simply as $\phi_t(n)$, this is the definition of $P_t$ of n for a given n this is (Refer Time: 26:18) we were going to define. So, what is t affecting? t is affecting the cumulative loss we are going to look at. And this variable input variable $\eta$ is telling you with what you are going to multiply this loss with.

So, with that definition, we have this quantity here. But what I am interested in not this $\phi_n$, I am interested in minus of this quantity. So, I am going to take minus here. So, because of this, this guy becomes positive, and this guy is negative here ok. Now, let us focus on

this term $\frac{1}{\eta}\log\sum_{\{i=1\}}^{n}\exp(-\eta_n\,\widehat{L_{ni}})$. If I remove some times some terms in the summation, this quantity is only going to be smaller, because each term is a positive quantity here.

Now, because of that and with this minus sign, if I remove some quantities there, I am only going to get. So, here I equals to n to k right, I am going to only written the component kth component in this. So, it is going to be $L_{ki}$. Is this correct? What I have done is I have in the summation I have written only one component. It is up to me which component I want to written and in this case I have written the kth component. And because of that this quantity becomes smaller, but this with a negative sign, the n is the last round, yeah, n is going to remain the same, yes this should be n k that is right yeah.

Now, if I just now further simplify this, this is going to be what? Log of exponential get canceled, this eta and logs of this $\eta_n$, and what I will end of is $\widehat{L_{nk}}$ . But by definition what is $\widehat{L_{nk}}$? This is a cumulative loss of the kth action till round n right, so that I can write it as $\sum_{\{t=1\}}^{n}\widehat{l_{tk}}$. So, now I also know that this quantity is nothing but from my earlier definition, this I can write it as t equals to 1 to n this is nothing but $\sum_{\{t=1\}}^{n}E_{\{I_t\sim P_t\}}\,\widehat{l_{tk}}$.

This is n k here when we are writing cumulative. So, this is the cumulative of the first n terms, so that is why we are letting p equals to 1 to n summation; l t k hat just a minute what we wanted to say l. So, let me write this is this was $l_{tk}$ here, l t k hat right. So, how did you write these terms here? What is the meaning of this, why did you write this term? This term $\widehat{l_{tk}}$ was equals to this term here right, and this is equals to this term ok. So, maybe I do not need this term, it is just this term here right.

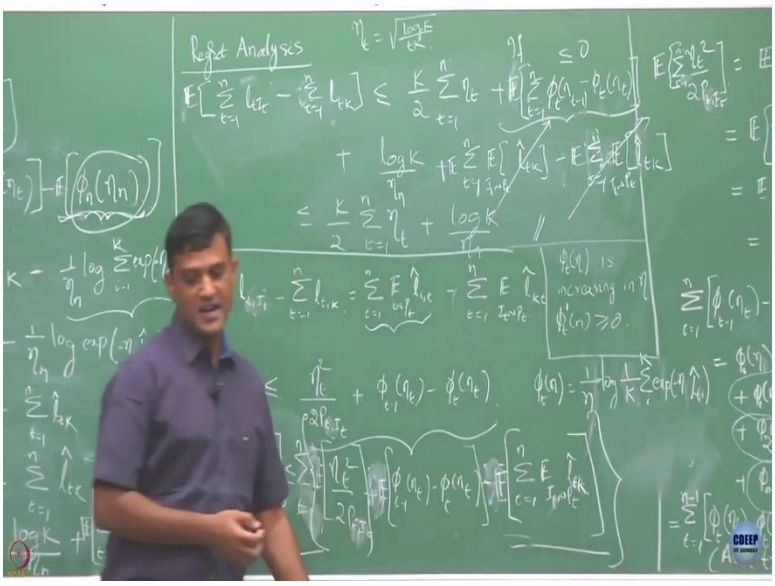
Now, what I am interested in not this quantity, but the expected value of this quantity right, because there is an expectation here. When I want to take an expectation, there is an expectation here for all these quantities right. I removed the expectation here, but there was one expectation here, and also there was an expectation ok.

Now, I would then take expectation of this quantity here. Now, if I do that, what I will finally end up is $\frac{\log K}{\eta_n}$. So, now, this was like $\widehat{l_{tk}}$. And if you are going to take the expectation of this, like the way we did I can split this expectation into two part; expectation with respect to random adversary and expectation with respect to the of the

players strategy. If I do that, now we will see that this expectation turns out to be simply. Is that right? Ok, fine.

So, now we are almost there. So, finally, now simplifying putting and clubbing all these things together, so there is a mistake we made here right like you notice that when this log knocks out this exponent, we ended up $-\eta_n \widehat{L_{nk}}$, but this was a minus and there was also minus here, because of this we will end up with the positive term here, not the negative sign ok. Now, putting all these tools together, finally, we have now a bound on this, we have a bound on this, and we have just this term.

(Refer Slide Time: 32:54)




So, if you have finally, put all the things together ok, sorry I was this one minus expectation. So, let me club all these things. First thing is.

We have bounded this part; actually we have bonded this minus of whole of this part ok.

Ok, sorry this should be only the expectation of this quantity here ok. So, the earlier terms still remains like when you are looking at the whole term this term still remains. If I am going to take the expectation of this term, it is going to be expression of say this term expectation and also expectation of this term. What we have now shown is expectation of this term is upper bounded like this ok.

So, now, if you are going to put all these terms, this term the expected quantity of this we have shown it to be upper bounded by this quantity which is given by $\frac{K}{2}\sum_{\{t=1\}}^{n} \eta_t$ ok. The next term on this what we have is $E[\sum_{\{t=1\}}^{n} \phi_{t-1}(\eta_t) - \phi_t(\eta_t)]$., and then this term we have bound of $\frac{\log K}{\eta_n}$, and this term here $E \sum_{\{t=1\}}^{n} E\left[\widehat{l_{tk}}\right]$, where expectation is with respect to I t going P t. And we have also, This is I have just split this expectation into two parts here, and then we have this part here which is and now the last part I am writing $E \sum_{\{t=1\}}^{n} E_{I_t \sim P_t}\left[\widehat{l_{tk}}\right]$.

So, I am taking this expectation right. When I take the expectation, when I take expectation, I have to always take expectation with respect to the randomness of the adversary as well as the randomness of the learner, so that is why that expectation I have split into two parts. The inside part is with respect to the randomness of adversary, and outside part is with respect to that of adversary. Yes, same thing like here I have just write, this part is the same right, because here also I am taking the expectation with respect to a randomness both the adversary and the learner. Is that clear?

Ok, now we are almost there this part knocks off with this. This term and this term is what we have in the actual bound also right. So, finally, we need to show that this quantity is 0 or this is upper bounded by 0. If you can show that, we are done fine. Then that if this quantity here is less than or equals to 0, then we have the bound of $\frac{K}{2}\sum_{\{t=1\}}^{n} \eta_t + \frac{\log K}{\eta_n}$, and this is what we wanted to show.

Now, the question is why is that this guy is summation is less than or equals to 0? So, for that instead of now going again further details, I will leave you to verify that. If you are going to take $\phi_t$, any t, see you notice that this is a difference of $\phi_t$ function computed at $n_{t-1}$ and $n_t$; in round t, I am looking at the same $\phi_t$ function; one computed at $\eta_t$, and one computed at $\eta_{t-1}$.

Now, how are these $\eta_t$'s are chosen? The $\eta_t$'s are chosen to be either all the same quantity if you know apriori, what is the number of rounds we are going to play that is n or we have chosen such that they are decreasing in t right. So, what is $\eta_t$? Like this. So, this $\eta_t$ is decreasing in t. If you increase t, this guy is going to decrease. Because of that this $\eta_t$ is going to be smaller than $\eta_{t-1}$.

Now, you can show that this function $\phi_t$ is, so if we can show that this $\phi_t$ is increasing eta, then this difference is always going to be negative right. So, how you are going to show that? What you will do is you take the derivative of this guy $\phi_t$ with respect to $\eta$, and then you are going to argue that this guy is going to be greater than or equals to 0.

So, I will leave it as an exercise for you guys, because I have already done many steps only that this is the one step which we are not verifying. Verify that, to verify that you need to know a quantity called KL divergence which I have not yet introduced, but look into the book. But even if you do not know this, that is fine. You should be able to just differentiate it in the standard fashion and manipulate the derivative to see that this guy is always positive for whatever $\eta$ you are going to choose. So, because of that this guy is going to be negative. If this guy is going to negative, then we have this upper bound ok, so, fine. So, this kind this now completes the proof of Exp3.

So, what we are able to show that if you look into the regret bound on the pseudo regret we have this quantity which said that this will translate to regret bound of the order $\sqrt{nklogk}$. Any you have any doubts regarding this proof so far? Yeah, we have went through many steps, but most of the steps are standards, the manipulation steps were standard, even though it is not clear like why we follow this steps in this particular sequence. So, this is a kind of standard steps we will incur when you are going to deal with any proofs on the regret bounds of adversariality ok, so fine.