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

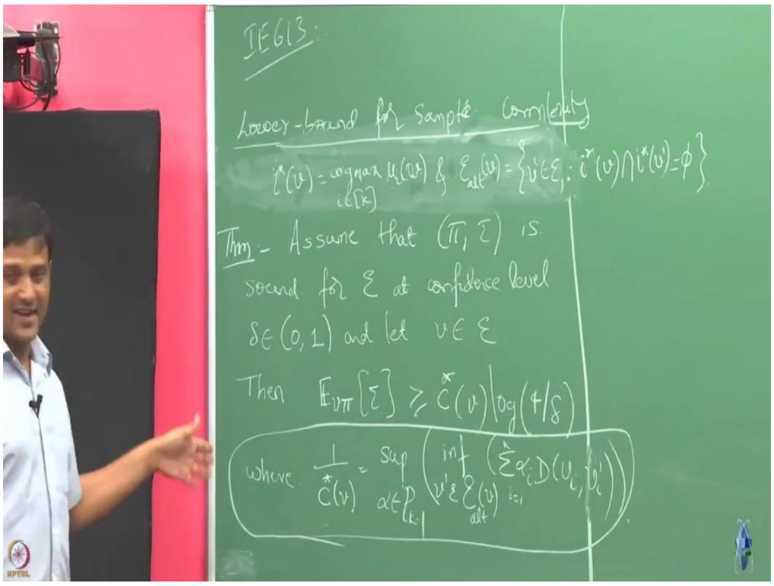
Lecture – 60
Lower Bound for Pure Exploration Problem

So, today we are going to continue this Problem of Pure Exploration. So, I will just try to wind up this today. We will just going to discuss the lower bound today and just see how this lower bound compares to the lower bound of upper bound of what we got for Kll Ucb. So, we started looking into the best arm identification problem with a fixed confidence that is if I am given δ, how should I am able to find the best arm with this much of confidence with fewer sample complexity.

The two algorithm we have seen. We so, that at least the Kll Ucb was intuitive, it is just trying to identify the best 'm' arms. By most of the time it is try to distinguish whether m th is there any ambiguity between m th and (m +1)th arm. Just try to resolve that and when it is sufficiently confident that m th and (m +1)th are separated it stopped.

Now, the question is in general any algorithm if I apply it on any instance on an average how many number of samples it needs before it gives me the correct answer.

(Refer Slide Time: 01:50)



So, we are going to look into the lower bound for sample complexity. So, the bound we are going to see this its bit involved, right now it is not clear why this bound look like that, but let us write it and then try to understand a bit more, ok. I am going to write it as a theorem, ok.

So, let us say you are given an environment ν for which the associated means are denoted as $\mu_i(\nu)$ ok. And for that environment ν let us say the optimal arm I am going to denote it as $i^*(\nu)$, fine. It is just saying that if you give me an environment ν then the $i^*(\nu)$ is just the optimal arm for that setting. Now, I am defining the set of alternative environments $\left(\grave{o}_{alt}(\nu) = \{\nu' \in \grave{o}; i^*(\nu) \cap i^*(\nu') = \varnothing\}\right)$, such that the optimal arm in this environment is going to be different from the optimal arm in the original environment.

So, what is this saying? So, ν environment is given to me, $i^*(\nu)$ is the optimal arm in this environment ν, now I am looking at all this other environment ν'; such that the optimal arm in this ν environment will be different from this. Is it fine? So, I am just basically looking at such all this environments where my optimal arm is different from that in the environment ν. Now, this lower bound is defined in terms of that ok.

Now, assume that this pair (π, τ) is sound. So, we have already defined what this sound means ok, for some environments set $\grave{o}$ at confidence level δ. So, we have say what we are defined the soundness. We have said that soundness at confidence level δ we mean that the probability that it halts the stopping time is finite and whenever it halts that it outputs a non-optimal arm.

Student: Less than.

Is less than delta? So, let us take; let us take an any policy that is sound that is any policy and the stopping time pair that is sound. Then we are saying that on any environment ν the expected sample complexity of that algorithm is going to be lower bounded by some problem dependent constant $c^*(\nu)$ multiplied by $\log(4/\delta)$.

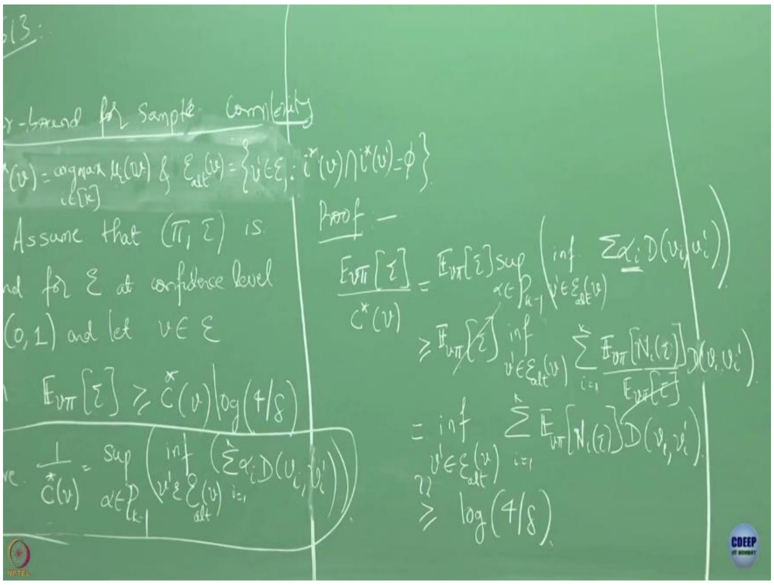
So, what is our input for this? Like we are talking about the fixed confidence right. So, how is this sample complexity depending on now this confidence term? It is $\log(4/\delta)$ here and there is a problem dependent constant. So, this problem dependent constant is with involved and it looks like this (check image above). So, the reciprocal of this problem dependent constant is the average of my Kullback Leibler divergence between the arms, right.

So, so what I am doing is I am taking environment ν' which is coming from my this set of alternative environments and now I am looking the Kullback Leibler divergence between my environment and the alternative environment, but now I am averaging them.

These α_i they are coming from a distribution. So, I am just now looking at what is this quantity. Now, this quantity depends on all possible distributions I have and also all possible alternative environments I have with respect to my given environment ν, ok.

So, given like this it is hard to interpret what this quantity means. So, let us try to anyway I try to get why this lower bound make sense with this definition.

(Refer Slide Time: 06:44)



So, to prove this we are going to start with this quantity $\left(\mathrm{E}_{\nu\pi}[\tau]/c^*(\nu)\right)$. So, I am just start looking into this ratio and now right away I am going to plug in this quantity which I have defined here $\left(1/c^*(\nu)\right)$. This is going to be; so, now, this quantity here is I am going to have taken the supremum over all possible distributions.

Now, what I will do is instead of looking at a supremum, I will look at a particular distribution. So, I am going to remove this supremum and take a particular distribution. If

I do this I am going to get a lower bound. And what is that particular distribution I will be interested in is? So, this is the expected number of times I have would have played before I stop. And what is this? This is during before I stop what is the expected number of times I would have pulled each one of these arms.

So, this forms a distribution. So, this forms a distribution and this belongs to this set of distribution that is why I have taken this one. So, notice that if the supremum is indeed achieved for this distribution then this lower bound holds with equality right.

Right now, I do not know this that is why I have written it as a lower bound, but if in if this is indeed the distribution at which this maximization occurs then we have an equality.

So, this will give us a sense of when this lower bound is going to be tight. So, the lower bound is going to be tight if at all thus distribution that this policy. So, this is all for a given policy π. This policy gets on the number of pulls before it stops if that happens to be the maximizing this divergence.

I this entire quantity then that quantity is going this lower bound is going to be tight ah, but anyway. So, we are going to get this lower bound. Then, now I am just going to do this further simplification. So, this quantity here it is does not depend on the ν'. So, let us get rid of this. So, it is then simply going to be inf over; ok. So, I will just simplify this.

Now, next what we are going to argue is what is this quantity here? So, this quantity is nothing but the weighted sum of the divergences right. So, these are like weights depending on how many times you have pulled and this is the divergence between the a pair of arms distribution.

We are next going to show that this quantity is going to be lower bounded by $\log(4/\delta)$. You need to show this. If I can show this then I am done right, when this quantity is nothing, but $c^*(\nu)\log(4/\delta)$ that is what we have here ok. Now, to prove this now you are go back to our the trick we used it in the proof of regret definition. What was the main result we used to prove the lower bound for our cumulative regret?

Student: Pinsker.

Pinsker.

Student: Yes sir.

But what portion of Pinsker?

Student: (Refer Time: 12:24). So, *p* if your *p* and *q* are the distribution.

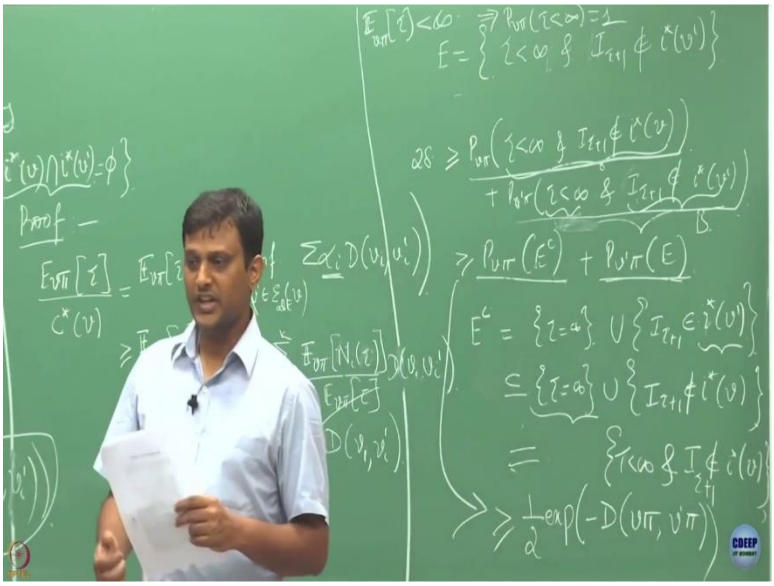
Right. So, if *p* are *q* are two distribution we wanted to say basically.

Student: *p* a plus *q* equal to (Refer Time: 12:37).

So, how thus sum of what is the mass for any event a, what is the sum of the mass of *q* on A plus *p* on A^c. So, we wanted to basically, so, that result we exploited to make sure that what is the probability that one event being confused for the other under other two another distribution.

So, we just try to use the same fact same result or same kind of idea and try to apply the same result here. To do that we need to define this set.

(Refer Slide Time: 13:26)



Let us define the set $\left(E = \{\tau < \infty \;\&\; I_{\tau+1} \notin i^*(\nu')\}\right)$. So, let us take this event E. So, if my policy π and the pair (π, τ) is sound what is the probability of this event?

Student: Less than δ.

It should be less than δ right that is the definition of my soundness. So, take any, so, here I have taken a environment ν' and I am asking when you are you stop and that you are going to output an arm which is other than my optimal arm is going to be less than δ then it is sound by definition.

So, now let us consider two events. So, I am going to look at two distributions. Let us say ν and ν' and now going to basically ask the soundness of my algorithm on this to environment. I know that by if my policy π is sound, what is this quantity? This should be upper bounded by δ and what about this quantity? No, this one P is $\nu'\pi$.

Under ν', so, I just change that the distribution. Instead of ν I am asking for ν' then. If my algorithm is sound it should be independent of what is this underlying distribution right. So.

Student: Sir, the first (Refer Time: 15:54) was the optimal arm for ν.

Yes.

Student: That is why.

And in the second case I am looking at for ν' ok. And now see like this is the probability with respect to the distribution induced by ν' and this is the probability that is induced $\nu'\pi$ ok, under which I am asking this probability. So, both should be if that is all this should be also less than δ this should be also less than δ together with the both should be less than 2δ.

See the soundness definition holds independent of what is the distributions you are looking at, as long as under that distribution that it happens that it falls it stops in finite time and whenever it happens it is going to give an arm that is other than the optimal arm should be bounded by δ under soundness.

Now, by definition we have defined this quantity to be E, the event E. Now, we are going to argue that this is also going to be this quantity here is going to be a superset of E^c that is why. So, this quantity is simply?

Student: E.

E and this we are going to $\nu\pi$ and this we are going to have E^c plus E $\nu\pi$ ok. Now, let us see let us understand this term, why is this is a superset of E^c. If this set is a superset of E^c then only I can write it as a lower bound like this right ok. Now, let us see what is E^c. If this I have defined it as E then, E^c of this is, what is that event means?

Student: (Refer Time: 18:20) belongs to the complement.

Or?

Student: (Refer Time: 18:32).

So, let us say I am going to treat this event as A and this event as B. I want A and B. A and B is complement right. $\left(A \cap B\right)^c = A^c \cup B^c$ ok. So, one more thing, if my policy π is such that my expected complexity is already infinity then this one naturally holds right like nothing to prove in that case because infinity is a bound to anything.

So, that is why we are going to I mean I am going to assume that we are going to assume that ok. If this is the case is it true that, if this expectation of the sample complexity is bounded; that means, the probability that this sample complexity is bounded is also almost surely almost sure fine.

So, let us solve some couple of things. The way all chosen ν', so, I am looking at ν and ν',and ν' I am taking from this set because I am I am interested only ν' coming from this set right, alternative environment sets. So, I am going to take that. So, if that is the case I know that this always holds that the optimal arm under ν and ν' are going to be different.

If that is the case I am going to write it as this guy is going to be whatever this is true that. So, if this is true if this guy is belongs to this it must be the case that it is going to have arm which is $i^*(\nu')$. So, we are saying that under ν', i^* is the optimal one. Now, what we are saying that? This $i^*(\nu)$ it does not belong; that means, this event already contains this.

Student: There are other arms as well.

There all other arms well. At least it contains that and there could be more arms. So, that is why we are going to say this one ok. And further any of this event I know is not going to happen right because of this probability, but anyway like I mean this is bit let us say this make sense. I am going to if I am going to further ask this then the set is I had a.

Student: We have to write a it could be equal to this set.

Just a minute.

Student: (Refer Time: 22:25) tau is.

Right. It is going to be τ is going to be this event is not going to happen. So, this is going to be equal. So, now, is this clear? Now that is why because now E^c here is going to be contained in the set, which is this set here whatever I wanted because E^c is contained in the set that is why I got the lower bound because E^c is a subset of this set.

So, whatever write I wrote here this part is exactly equals to this probability here and this probability here is a lower bound on this probability ok. Now, we are ready to appeal to the high probability Pinsker inequality.

What does it says? We have such an event here where I am looking at an event E, I am looking at the probability of that happening under distribution v' P and then I am also asking the complement of that event happening under another distribution v'. So, what does the Prinsker inequality gives for such a sum?

Student: (Refer Time: 23:50).

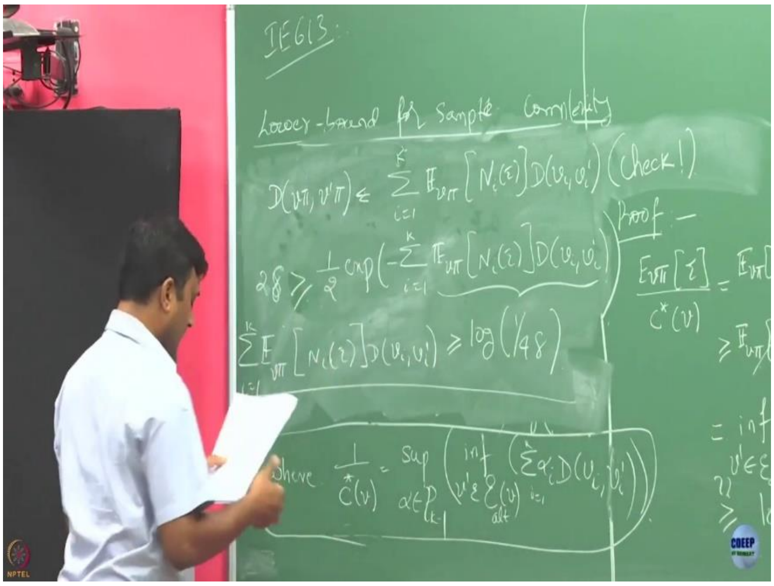
So, $P_{v\pi}\left(E^c\right)+P_{v'\pi}\left(E\right)\geq\frac{1}{2}\exp(-D(v\pi,v'\pi))$

Student: Nu prime.

So, notice that they are not just v and v', this is $v\pi$ and $v'\pi$. This is the distribution further induced by the way you have pulled your arm according to policy π. So, this kind of thing we have also already incurred when we did it in the cumulative regret case also.

Now, once we got this how did we do this? Then we say try to write down this quantity here as the weighted sum of the divergence between distributions of a pair of arms.

(Refer Slide Time: 24:59)



Now, let us try to get that. So, this step I am directly writing it, but this is exactly the way we got it in the case of the cumulative regret proof also. So, when we have looking at divergence between these two induce distribution that can be split and expressed as decomposed as the sum of the divergence between each the pairs of the distribution for each arms, but weighted in this fashion.

So, we did this. Actually we formally showed this, but it needed some more bit of notations just expressing it the pulls over all the arms before we stopped. There the stopping was like fixed that there after t rounds we have done, but here it is a random quantity τ, but whatever over the same thing, you can show this. But it so, happens that it is exactly not this it is going to happen with a think an upper bound here.

So, just check that. This is also given as one of the exercises in the book, but just check it ok. So, now, what we have from this point to this point just going to write it directly. After using this relation we have 2δ upper bounded by $\frac{1}{2}\exp\left(-\sum_{i=1}^{k} E_{v\pi}[N_i(\tau)]D(v_i, v_i')\right)$. I have just plugged in the decomposition for this divergence term which is expressed here.

Now, if you simplify this what we are going to get? I am going to get that the. So, I am just simplifying to get a bound on this. If I do this the $E_{v\pi}[N_i(\tau)]D(v_i, v_i')$ and this is going to be what? This is going to be if I am going 4δ. It is correct after if I simplify that this is what I get?

Student: Yes.

Do you have any of you have that a high probability Pinsker inequality? Can you check when I applied this point here is this constant here I got it right or it is something else?

Student: Correct.

Its half only?

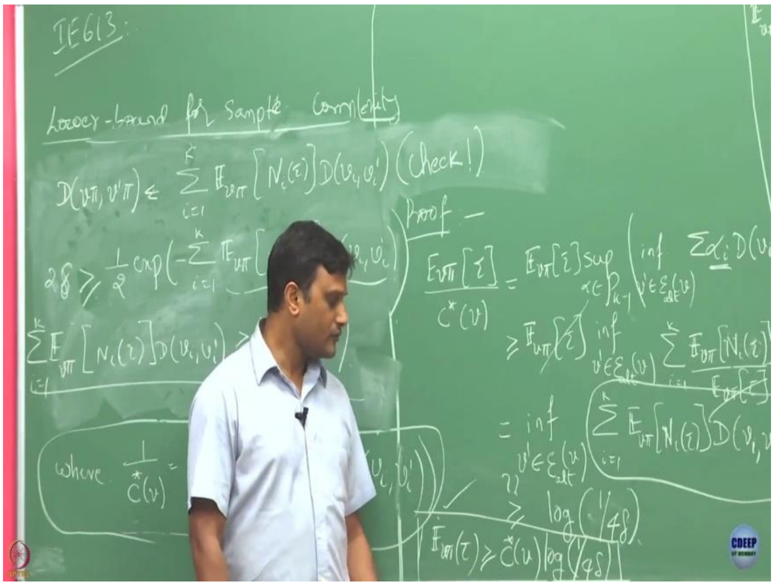
Student: Yeah.

So, in this case then I think we should be ok. So, finally, I wanted to conclude this result by saying that see now we have a lower bound on this right, which we have now actually shown it is $\log(1/4\delta)$.

This the way we have shown this is what we got. And now this is true for any v' we are going to take from this set right. We have fixed one v' and did this, but if you are going to do infimum this should hold this bound should hold even if you take the infimum or v' .

So, that it looks like there is slight confusion here the way it has been expressed here might be I do not know.

(Refer Slide Time: 29:20)



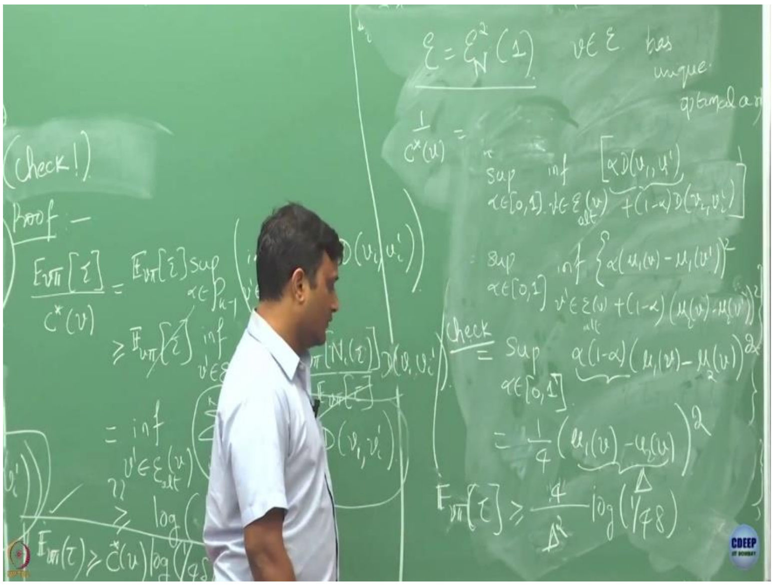
So, the so, let us stick to what we got. We have got $\log(1/4\delta)$, but the book say has written it as $\log(4/\delta)$. So, let us make it $1/4\delta$ ok. So, then we are also going to refine our what we actually showed is then? We have shown that the expected sample complexity we have shown is going to be lower bounded by $c^{*}(\nu)\log(1/4\delta)$.

So, one of you please do verify whether there is just a typo in the book or like it should be actually $1/4\delta$. So, our from our derivation I am getting it as $1/4\delta$ here, but the book it state that it is $4/\delta$. So, if see if you can verify that ok, fine. So, so let us take for time being this is our this is what we have shown. So, what we have now is, the lower sample expected sample complexity is this quantity upon 4δ.

So, if you want the confidence to be very high; that means, δ to be very small. So, this complexity is naturally going to be high. If we are happy to get it with a low confidence then this δ is going to be high and the sample complexity is going to be high. And now there is this bit a messy term here is which is a problem dependent quantity. So, in its term here what is this? But let us try to understand this quantity for some specific example ok.

So, what do you do what do you expect like this quantity like ok. This quantity is somehow dependent on v. Do you expect it to be like inversely dependent on the sub optimality gap? So, if it is inversely dependent on sub optimality gap then it is natural that if the gaps are small then the sample complexity is also high, but it is not obvious that this expression contains that. So, let us try to work out it for some simple case.

(Refer Slide Time: 31:32)



So, let us see this. Suppose, let us take an example. Let us take environment to be $\mathcal{E} = \mathcal{E}_N^2(1)$. What is such the notation indicates? So, it has 2 arms. The superscript is tells the number of arms. N is showing it is a set of sub Gaussian noise and with variance is equals to 1.

It is a one sub Gaussian in this case and mean we are going to take 0 because this is a sub Gaussian random variable. Now and also assume take that $\nu \in \mathcal{E}$ that has unique optima. Now, if you are not going to compute this quantity. So, since this there are only 2 arms, I had to only worry about one probability in one probability vector with just one free variable. So, I am just going to take it as a supremum over $\alpha \in [0,1]$ and then infimum over $\mathcal{E}_{alt}(\nu)$.

So, I am going to take there are only 2 arms right. If one term is α other term is going to be other term is going to be $1-\alpha$ because this α constitute a probability vector. Now, this is going to be. So, if it is a Gaussian random variable what is the divergence between two Gaussian random variables?

Student: $\mu_1 - \mu_2$ (Refer Time: 34:18).

$\mu_1 - \mu_2$ provided their variance is the same.

Student: (Refer Time: 34:24).

Ok.

Student: You are asking for a kl divergence.

kl divergence.

Student: That is not (Refer Time: 34:30).

No, it is $\mu_1 - \mu_2$, whatever the means of the distributions their difference squared.

Student: (Refer Time: 34:39).

Provided their variance is the same. So, I am not saying ok. So, I have to correct here like I am just saying this is the set of all Gaussian distributions with variance 1. It is not necessary that the mean is going to be 0 here. In this case it is going to be this term here is going to be simply $\alpha\left(\mu_1(\nu)-\mu_1(\nu')\right)^2+(1-\alpha)\left(\mu_2(\nu)-\mu_2(\nu')\right)^2$

Student: (Refer Time: 35:03).

So, this is the divergence between two Gaussian distributions with the same variance ok. Now, let us try to see if we can optimize the inner term. So, now, we have to find this quantity taking infimum or all ν'.

So, ν is given to me this is for a given ν that is fixed. So, now, suppose we want to compute this ν'coming from this set alternative set, let us check this. I am just going to write this. So, recall that what is this ν'? This coming from a set in which the optimal arm is has to be different from that of the optimal arm under the environment ν.

So, under this if you try to find the infimum of this I am just directly going to write it as $\alpha(1-\alpha)\left(\mu_1(\nu)-\mu_2(\nu)\right)^2$. So, this is going to be if I am going to find all ν', I will eventually end up with this again. Just check this, I have not checked this. So, just we have trying to understand how this quantity looks like.

Now, if you have so, now, this quantity is only sup over α. Now, α appears as $\alpha(1-\alpha)$. When is this quantity maximize $\alpha(1-\alpha)$ if you are going to maximize over between [0,1]?

Student: Half.

That is going to maximize at half. So, then this going to be $(1/4)(\mu_1(\nu)-\mu_2(\nu))^2$. And what is $\mu_1(\nu)-\mu_2(\nu)$ for this particular distribution? This is the sub optimality gap right. We have only taken 2 arms and for this is the mean of the first arm and this is the mean of the second arm and it is a squared one. Whichever is the optimal I do not know, but this is giving me the sub optimality gap.

Student: (Refer Time: 38:10).

No it is $\mu_2(\nu)$ only . So, I am already optimizing it for all ν'.

Student: (Refer Time: 38:20).

This is the after taking infimum over this I am directly writing it like this.

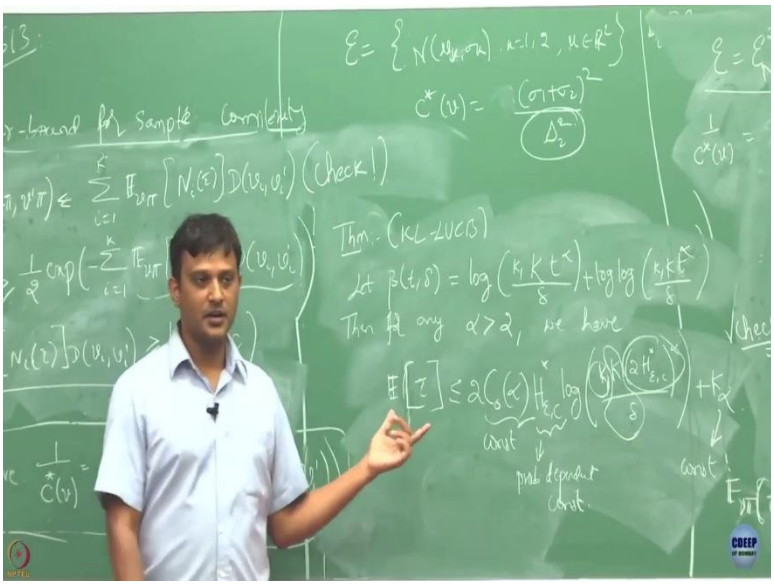
Student: (Refer Time: 38:27).

So, that is what not check directly written ok. So, you have to do some optimization making sure that this ν' is such that the optimal arm in it is going to be different from that of the optimal arm for ν.

So, and then we are going to get this. So, if we are going to plug it here you will see that the expected is like of the order 1 upon we are going to get it as like something ok, 1. So, this is like what we call $\mu_1(\nu)-\mu_2(\nu)$ as Δ right what we have been calling as sub optimality gap for this environment. So, this is going to be like $\frac{4}{\Delta}\log\left(\frac{1}{4\delta}\right)$.

So, anyway this complex looking quantity, it is time to capture what is the how complex the problem instance is. Here the problems how complex the problem instance is capture directly through this sub optimality gap, right once you simply this for this specific instead of the set of Gaussian distributions with variance one we exactly got this ok.

(Refer Slide Time: 40:14)



On the book they compute for couple of more examples, for I am just going to write this. Suppose, if you are going to take epsilon my environment to be. So, here in this example we fixed variance for all the arms to be 1, but even if you relax that and we are going to, $\mathcal{E}=\{N(\mu_k,\sigma_k): k=1,2,\ \mu\in\mathbb{R}^2\}$, and μ vector could be anything.

So, you take any mean vectors and any variance quantity then you can compute that. This quantity turns out to $c^*(\nu)=\frac{(\sigma_1+\sigma_2)^2}{\Delta^2}$. So, if you are allowed any variance σ_1,σ_2, so, here I in this example I have σ_1,σ_2 to be both 1. That is why I have got it has $4/\Delta^2$, but if you just make them any σ_1,σ_2 this is what we are going to get just by following the same approach.

So, again you see that this quantity is going to be inversely proportional to the square of the sub optimality gap fine. So, this is how the sample complexity look like. Now, the question is the algorithms we had earlier are they optimal, in the sense that they are regret bounds is of the same order as this lower bound.

I am just going to write the upper bound we are going to get for kll ucb. It is a big the upper bound if I have to wanted to write it in a full generality it is a very very complicated one, so, I will just try to give you the order wise flavour of that theorem ok.

I am just going to say them for kll ucb. So, the term kl lucb we have this term $\beta(t,\delta)$ right which came in the computation of our upper and lower confidence bounds. So, suppose let us say if I said that values $\beta(t,\delta)$ to be of this fashion this $H_{ò}$ here is some problem dependent constant and this k^*, k_α is another constant.

Now, if you see this, so, κ_1 is a again another constant here which I am not going to define. If you look at this the way I got this is log k by δ is what I get, but in the actual lower bound what is how was the dependency on δ.

Student: (Refer Time: 43:58).

It is $\log(1/4\delta)$ like, but the so, that term inside log did not depend on how many arms are there, but here it looks like this count guy depends on this side. In that sense this is not optimal right.

So, even though it has other constant, but if you just look at the parameters of interest, number of arms, k, the δ parameters are confidence parameter and the problem specific parameters that is all δ. So, justify just want to focus on this δ k, it looks like the extra $\delta\kappa$ has popped up here. In that sense this is not optimal even though there are other constants here.

So, we will just leave this upper bound here. What you can look into the paper for all this description. I just wanted to draw an analogy like how does the upper bounds for the algorithm we have discussed they compare with our lower bounds. So, then the question is by just looking at that you feel that this algorithm is it may be empirically performing good, but at least the bound wise you will see that it is not matching the lower bounds right.

So, in that way you want to look for a better algorithm or like hope that I may have a come up with an algorithm which has a tighter bound then this and matches the lower bound and maybe it has a better empirical performance also.

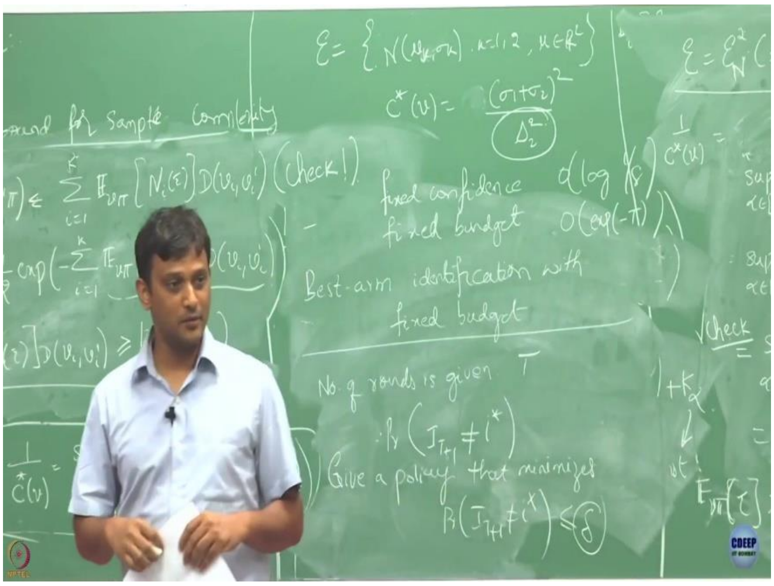
So, again as I told in the class off late many algorithms have come up on this and we do not have time to discuss all those things. But, if any of you interested, you could take study this the survey of what all the new algorithms that have come up.

You can compare their numerical performance and say which one is better and which one has the best theoretical guarantee and which one has the best empirical guarantees ok. The last thing which I do not want to delve into is there are two flavours of best arm identification problems. We said one is fixed confidence, other is what?

Student: Fixed budget.

Fixed budget. You may ask the question ok. If I have I can only explain 100 rounds within this you after this 100 rounds you tell me the best arm. So, in that case you will be goal is to output an arm which is optimal with high probability. So, in that case you want to minimize.

(Refer Slide Time: 46:48)



So, let me just write that. So, in this number of rounds is given, let us say T is given and you may have to output an arm at the end of T+1, such that the probability that this is not equal to i^* should be minimum. You want to. You get this? So, this is your criteria. So, give a policy that minimizes I just this quantity.

Now, in this line of attack, we are trying to we will show that whatever the policy would allow with what probability it is going to guarantee this. So, what do you expect this probability to be if you have a T number of rounds right? You are going to use T rounds and after that you are going to output an arm and you are going to tell me with what probability that is going to be not an optimal arm.

You have to basically bound it. You may say that I am going to whatever I give you it so, happens that it is not going to be the optimal arm and that probability is going to be bounded by δ. You want to make this delta as small as possible then your algorithm. What you feel that how this you expect this δ to be dependent on T?

Student: (Refer Time: 49:16).

Ah?

Student: 1/T.

Like 1/T.

Student: (Refer Time: 49:24).

But what could?

Student: Sir e to the power.

May be exponential, but how?

Student: (Refer Time: 49:30).

It should be expected to be $\exp(-T)$?

Student: There is a log 1 by δ for (Refer Time: 49:38) powered one.

Yeah.

Student: Then (Refer Time: 49:43).

Yeah.

Student: In this case if I just take the inverse of capacity.

So, you have basically putting δ is 1 by T?

Student: It is kind of dual problem.

Is that?

Student: I mean there is some.

So, if so, for the fixed; for the fixed confidence, you got like let us say order $o\left(\log(1/\delta)\right)$ right for so, for a given δ. Now, for fixed budget that showing to be like $o\left(\exp\left(-T\right)\right)$?

Student: Minus.

Minus T.

Student: Depend on some other problem dependent

Of course, there will be some other problem dependent once, but. So, what you are saying is I just equate this two T that is a sample.

Student: (Refer Time: 50:50).

It looks like I mean when people this fixed budget is bit more nuanced than what we have for fixed confidence in this. Definitely, this the way you are thinking this relation does not translate like this. One has to separately prove it, but there is a recent work that says that it is not necessary that we have to build altogether different algorithm for the fixed budget case.

If you have a an algorithm so, basically that work has a; basically we can come up with a common algorithm. This small difference like there will be, but it is going to use a common module, such that it can simultaneously perform better offer the fixed budget case as well as fixed confidence case.

So, they have one algorithm with that calls certain modules. What is the difference in that algorithm is the way it is going to stop, at what point it is going to stop. So, if you have given T the fix budget case, it is going to explore T rounds and then going to output an arm. If you just pass it on δ, it is making sure that it explore sufficiently many time that whatever it is going to output it happens it outputs arm which is going to optimal with probability at least 1-δ.

In a way one can think of committed with a common frame which performs on both the settings ok. But, you will not go into that like you should have interested that it is also one of the good papers to explore and see how that performs empirically as well as what all the theoretical guarantees of that algorithm. I do not have it. It is called u gap algorithm that is think NIPS 2012 paper, if you are interested to look into that ok. So, let us stop here.