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

Lecture - 45
Regret Analysis of SLB - I

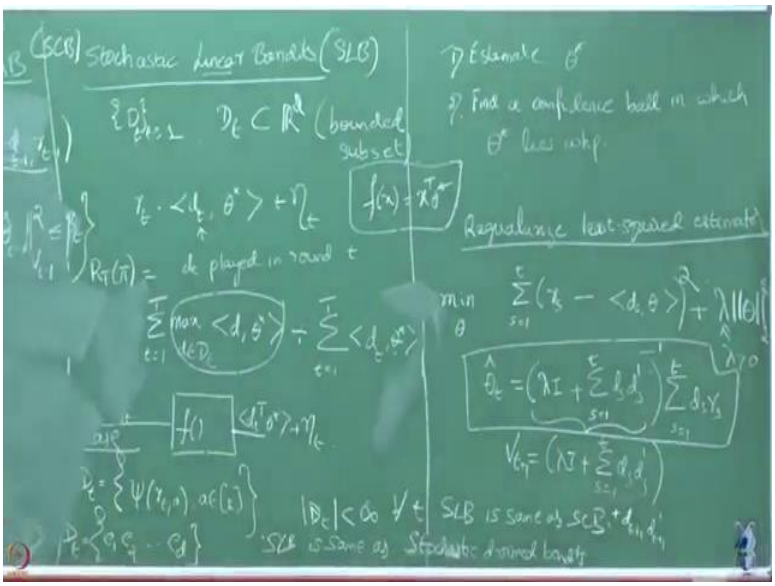
So, first thing the way we did, now drawing parallel with multi armed bandits; how we did, how you go about it? We constructed, we had two things there right; for every arm we maintained an estimate and the second thing we maintained a confidence interval about each arm.

So, we are going to do a similar thing here. First thing is, based on my past observation, I am going to estimate my theta star and then I am going to maintain a confidence interval around it. So, in the standard multi armed bandit, what did my confidence interval tell? My confidence interval told, within this region of the estimated value, your mean is going to be lie with high probability that is what we did.

So, there for each mean we did that. So, there the interval was, there we looked at interval; because I was trying to estimate the real numbers. So, on a real numbers, so I maintained an interval there. But, now I am trying to estimate a theta star which is a vector.

So, now what we will do is, we will try to maintain a ball around my whatever estimate, such that my theta star lies in that ball with high probability. So, because there earlier it was a real number, I used to just maintain a interval. But now I have a vector, so I am going to maintain a ball around it, ok.

(Refer Slide Time: 02:11)



First thing and then. So, I want to do these two things. So, first focus on estimating theta star. So, what is a good way to do estimation of theta star?

Student: Gradient descent.

Gradient descent; no just estimation, not like optimizing estimation, I am talking about estimation. So, yeah.

Student: Maximum less.

Yeah.

Student: (Refer Time: 03:10).

Maximum less, one thing we could do is maximum likelihood do it; but this is like a linear case right, that do you see any simpler thing. So

Student: (Refer Time: 03:26).

You guys know regression; linear regression? Can we do linear regression here, because it is just like a linear function, right. So, how to do a linear regression here?

Student: (Refer Time: 03:40) parameter.

Yeah. So, theta star is my parameter, I want to estimate. I am asking, how to do, how to go about it? How you pose it? How you are going to set it up?

Student: (Refer Time: 03:48).

What is the formula that gives you theta star or how you obtain it?

Student: We do not have a (Refer Time: 03:57) and.

I want you to give me the current estimate. I want you to give an estimate in every round, yeah.

Student: (Refer Time: 04:04).

How, I mean, instead of giving close now just tell me; how you got it?

Student: Differentiate with respect to (Refer Time: 04:13).

Differentiate.

Student: With respect to theta I (Refer Time: 04:16).

Yeah.

Student: (Refer Time: 04:18).

What you are going to differentiate?

Student: Mean square (Refer Time: 04:21).

Mean square.

Student: (Refer Time: 04:23).

Ok. Mean square error, you want to differentiate. What is the mean square error in this case?

Student: So the actual (Refer Time: 04:31) actual reward (Refer Time: 04:34).

Yes.

Student: And the.

Observed.

Student: (Refer Time: 04:36).

What you actually observed?

Student: (Refer Time: 04:40)?

And how that observation is related to your parameter? Ok. So, you note that every time you are observing r t that actually depends on theta star. So, r t contains the information of theta star ok; but only thing is r t you are going to observe, it is perturb by this noise, ok. So, how we could do; one possibility is, go with regularized least square estimator.

So, what is this regularized least square estimator? So forget regularizer. What is the least square estimator, how it is going to look like? So, the least square going to, it is going to define the loss over period; let us say over time t, you have observed r_s by playing d_t let us say. You know that you observed r_s, but all you know is it has actually come from.

Student: (Refer Time: 06: 06).

D_s theta star, we do not know this. We are going to, for if you do not know right now theta star, right. Let us say you are assuming for some theta, things are going on. You are getting this over apriority, but this whatever value you are is getting is this is nothing but this value; but this value is a noisy version of this.

Now you want to see which is that theta, which best approximates your observations r s? You will see that, what do you want to do? You want to take this squared error, this is the error right; want to take the square error and want to minimize it over? You want to minimize it over theta. But, if you just do this kind of minimization, there could be some issues like this theta hat you are going to get as an estimate out of it; it may not be unique always, ok.

So, just to avoid those cases you are going to regularize it. And by this kind of regularization you can verify this; this function is a strictly convex function in theta. So, once it is a strictly convex function in theta, what we know?

Student: (Refer Time: 07:55).

It is going to have an unique optimal value. So, let us call that as theta t hat. And what is this solution is going to look like, theta t hat? It is going to look like. I am going to this lambda is some number positive.

So, now, you should do this, it has a closed formed solution which looks like. So, this is one of the natural estimator, right. So, in the multi armed bandit, what was the natural estimator for that? We just average the samples, we have observed for that arm, ok. But, here what we are doing? We are just fitting our observation to what we are supposed to have gotten, ok.

So, we are just trying to, in other words we are trying to minimize this least square error and try to find a theta which turns out to be having this close form solution, fine. So, we how one thing now, we have an estimate now. So, the next question for us is. So, this is just theta hat we have obtained right, based on our observations so far.

So, what have been, we have been observing? We have been observing the reward I have got in that round. So, this is only t, small t and also do not confuse by prime I mean transpose. Sometimes I also for transpose, I may end up writing t.

So, I have been observing d_s for all s 1 to t till round t; that is the one actually I played right, and I have observed the corresponding rewards. I have all this information, from that I have estimated this. But this was based on a some set of samples. See this r_s is still a noisy quantity, it has noise in it. So, theta hat itself is a random quantity now, right. So, I need to have a confidence about this noise estimate which is a noisy. So, how to construct a confidence ball around it, so that this theta hat lies in that confidence with high probability?

Student: Travels which shown (Refer Time: 11:15) before.

No, we fix it, we up priory choose it something and then fix it; right now we have not saying anything like this is a some tuning parameter. You can just take some lambda and you will see that the purpose of choosing this lambda is just to make sure that it is theta hat is unique or that this theta hat is invertible.

Student: (Refer Time: 11:46).

Yeah.

Student: (Refer Time: 11:50).

Sorry yeah this matrix is invertible; hat you have chosen lambda to be 0, I was now we we are not sure this guy is going to be invertible. But now given that we have used this regularizer term lambda here; it is just for that technicality that we want to ensure that this matrix is invertible. So, you can choose lambda to be some very small number. I mean I do not see like there is a any optimal value that you can choose, you can optimally choose lambda here.

Now, in all the stochastic linear bandits, the big thing is about constructing these confidence balls. So, if you remember in multi armed bandits; the we have seen two algorithms mainly right, which we studied a bit. What are those? One was UCB and another was what

Student: (Refer Time: 12:43)

KLUCB right, we saw KLUCB.

So, UCB how did we derive its bound or how did we find it its confidence interval? That came from Hoeffdings inequality, right. And in KLUCB, how did the confidence intervals come from?

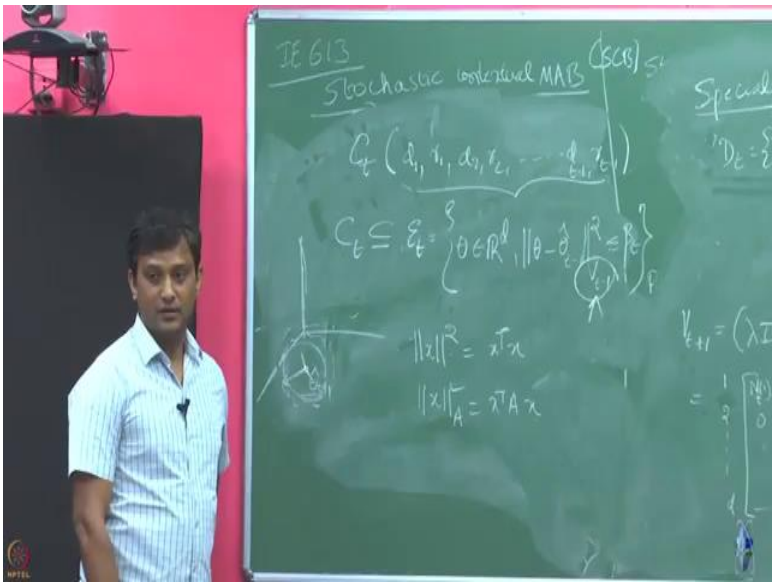
Student: Sir, some Chernoff.

So, it was bit a better version of a Chernoff. What was that Chernoff Hoeffdings inequality or just we said Chernoff?

Student: (Refer Time: 13:13).

Yeah I think it was Chernoff, for a special case of Bernoulli, right. So, for a Bernoulli case we had a tighter confidence bound and we exported in coming them. So, all those algorithms depended on how you constructed this confidence ball. So, here also now; how we are going to construct the confidence bounds that is going to change how your algorithm is going to perform? Ok. So, right now we will talk later about how we are going to choose this confidence bounds and how the parameters involved them and chosen.

(Refer Slide Time: 13:57)



But right now, let us assume a generic structure. Let us say I am going to assume a generic structures; some C t confidence interval that I am going to construct using my observations so far. That is my x, how I am denoting my d_1 and how did?

So, this is my action right; d_1. I played d_1, I observed r_1; I played d_2, I observed r_2 all the way up to d_{t-1}, r_{t-1}. So, whatever I have observed till t till round t minus 1 based on that I am going to construct a confidence ball. And we are going to assume that my theta star is going to lie in that confidence set with high probability.

So, we will later see how to define this confidence sets. A specific or a generic example is we are going to choose this C_t. So, C_t depends on all these things, but now right I am going to write just as C t; which is going to be usually the subset of some epsilon t which is going to look like this.

So, right now beta t is some parameter, I have not, we will specify this how it looks like. But what we are going to assume is, it is set of all theta which are going to lie around theta t hat within a radius of beta t. You understand what is the, how the set looks like.

So, suppose you have. So, this is like in your let us, let us visualize this in your 3D dimension, ok. Let us say this is some. So, this is where your theta t hat lies. So, what you are going to look is, you are going to look for all theta which are going to lie around this theta star within a radius of beta t. But you see that it is not a simple Euclidean

distance here; it has it is a Euclidean distance with respect to this matrix V_t. What is V_t? This V_t is exactly this quantity; C_t is a confidence set in which with high probability I like my theta start to lie in.

Student: (Refer Time: 17:44).

Yes.

Student: (Refer Time: 17:48).

It is with respect to theta t minus 1.

Student: (Refer Time: 17:56).

So, based on this, I am going to find a theta in round t theta and that from that I am going to decide, what is the decision I should be playing next, in the next round d t using a value from this. So, it will become more apparent later; but right now we are just saying that, whatever information we have till round t minus 1, I use this to build a confidence set in which my theta star lies with high probability, ok. So, later we will see that is the is it a possibility at all, can I come up with some beta t, such that such a set contains my theta star with high probability, ok. So, next.

Student: (Refer Time: 18:58).

Ok. So, if I just to take this the normal, what is this? This is nothing but x transpose x. If I am going to take this

Student: (Refer Time: 19:25).

x transpose, it is going to be x transpose A x.

Student: Sir.

Student: K has to be positive.

K has to be positive definite; k for positive a definite matrix, we are going to define this. So, there is a name for this right; is it called Mahalanobis matric or is it a Mahalanobis matric, fine. So, it is with respect to this matric A, we are going to define the value of x like this. So, when we this corresponds to. So, is this a special case of this, for what A.

Student: Identity.

Identity; when A is identity, we will just get this. Yeah

Student: (Refer Time: 20:20) we have positive (Refer Time: 20:22) minus 1.

It is theta hat t minus 1.

Student: How is that?

Because that is what till time t minus 1, I would have estimated that. That is the estimate I have about theta star. Now, I know that my theta, theta t hat is a random quantity, right. Now, I want to ensure that, around I know my theta star lies something around that theta t hat, theta star. I just want to ensure that, what is that region around theta t hat, theta t hat t minus 1 in which my star lies with high probability? I said this is now we are just trying to define it at every time t, right.

So, every time t before I am going to decide, what is the d_t I am going to play; based on my previous information whatever I have, I just first construct this set C_t in round t. I am going to use this set later to come up with what is the d_t I should be playing in the current round t.

So, we will see exactly how to do this. So, now, I mean this is as of now bit abstract; the way you can just think about this is, this is a ball in which the principal axes are the Eigen vectors of this V_t matrix, ok.

So, you understand, what is the principal axis? So, in this case like I have x, y, z right; but if you just look at the set in this, you can think of the directions given by the Eigen vectors of this matrix to be my basis directions for that space, ok.

And further, the length of those directions are going to be inversely proportional to the length of this eigenvalues. It is just like, so suppose you can think of this is somewhere like some ball, which is encompassing this like, which is that ball which has been described by these axes and whose length is inversely proportional to the eigenvalues of this V_t matrix. And further, I am saying this axes length is inversely proportional to the eigenvalues.

Student: (Refer Time: 23:21).

Then it is a smaller.

Student: Sir small (Refer Time: 23:25) small (Refer Time: 23:26).

Yes.

Student: Sir ball will be like (Refer Time: 23:29) one direction (Refer Time: 23:30).

Exactly. So, each direction have different length corresponding to the Eigen vector. So, now let us come back to this V_t matrix. As you see here, this V_t matrix keeps on accumulating this d_s values, right. So, if you go from t round to t plus 1th round; what is changing? You are going to add the new vector you played, right.

So, in that sense this V_t matrix is going to have more and more component. Do you think Eigen value of V_t is going to increase or decrease? So, let us say I construct, I just added one more term here d_{t+1} and d_{t+1}. So, and that will becomes, it will becomes V_{t+1}, right. So, compared to V_t, what do you think about the eigenvalues of V_{t+1}? Larger,

Student: (Refer Time: 24:43).

looks like.

Student: Sir (Refer Time: 24:47) degree technically (Refer Time: 24:49).

So, Eigen value larger, ok. So, he did a reverse engineering, any forward direction. Determinant increases, yeah determinant is nothing, but the product of eigenvalues, right. So, determinant has increased means, Eigen value should have also increased. So, does it any say anything about; if I add more term, its determinant has increased? Ok.

So, that is not obvious, right. You can think about it offline. So, if you are just going to add more terms like this. So, notice that d s, d s prime itself is a matrix; d_s is a column vector, d_s' is a row vector. Every term here is a; this is basically nothing, but some of matrices, right. So, this is one matrices, this is one matrix. What is this matrix? This is the matrix of identity, identity matrix; but the diagonals not one, but lambda. For each s, this is another matrix. Is it true that, each matrix in this sum is a PSD.

Student: They are all rank (Refer Time: 26:15).

There are rank one matrices, right. So, if you keep on adding many PSD's; what is going to happen to its rank? And what is going to happen to its eigenvalues? So, anyway I will just leave it to you. So, just think about, what happens if you are keeping adding more and more positive semi definite matrices to a another positive semi definite matrices? Is it either eigenvalue is going to increase or not? We just see; so naturally I am now just saying that they are going to increase, verify that.

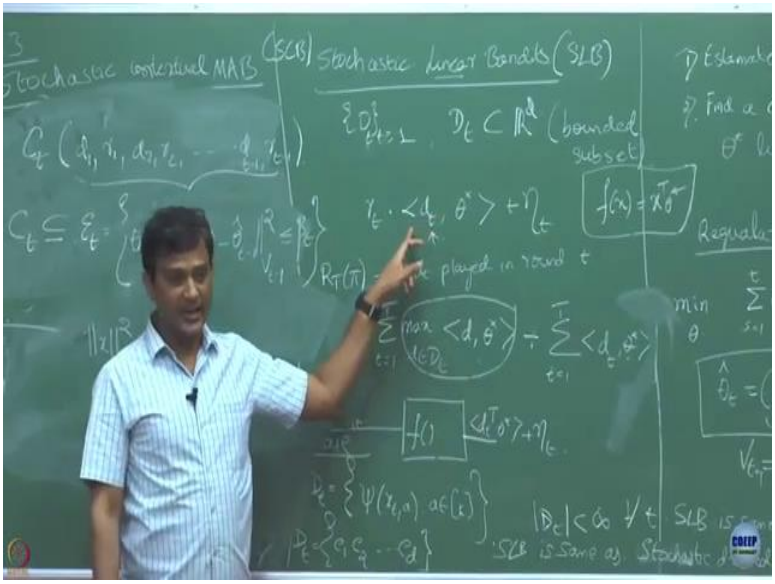
If it is going to increase what you are going to see is; as more and more data points come, I am going to get a confidence interval which is shrinking in all the directions, right. So, I should be more confident, I should be able to come up with a smaller balls in which my theta star is contained, ok.

So, let us at this point compare the difficulty in what we had in multi armed bandits and what we have here in having the confidence intervals. So, when we did a confidence interval in the multi armed bandits; how we did? We used Hoeffding's inequality, right. But in deriving Hoeffding's inequality, what we had? We had all iid samples, how did we, so just go back and now recall.

What we let us say we have pulled an arm in the standard multi armed bandits n times; we have got n samples which are iid, right. We just took their average, that will give gave me a sample mean and then try to see how that sample mean deviates from my true mean by applying Hoeffding's inequality.

So, the things were easy there, because it is easy to handle i i d samples. But now, do I have that luxury here?

(Refer Slide Time: 28:48)



Theta t are, assume they are all independent; but you see that theta star is there in every r_t right. All my rewards samples are all correlated, because this theta star is there.

And further even my choice of d_t in every round are going to be correlated; because this d_t has to be somehow dependent on, in every round depends on the past observations right, those observations themselves are correlated. So, even my choice of arm d_t here they themselves are correlated.

So, because of that, coming up with such a confidence ball against my estimation is in general hard here; because of that heavy correlation across different rounds. You see this, because even though yes good, like I have assumed that this is simple nice structure; but this structure has made all my observations correlated. So, I cannot go and simply leverage the results which I have for the iid observations.

I have to somehow take into account, the correlation across them; that is why this is all popping in, this V_t matrix, V_t has all the observations till round t minus 1, V_t is contains all the observations you have. So, far and they are going to determine how your confidence ball look, going to look like around your estimates, ok.

So, that is why when you move to this stochastic linear bandits, the estimation of theta star and building; estimation is easy we just saw, it is nothing but least square, regularized least square estimation, but confidence ball estimation is involved, ok.

So, I said that, the eigenvectors of this matrix V_t, they are going to define the principal axis of this ball I am looking at. And I also said that, the length of the principal axis is going to be inversely proportional to the value of the Eigen, eigenvalue corresponding to that eigenvector. So, each eigenvalue I have a corresponding eigenvectors, right.

So, let us take this particular eigenvector direction, the length of this Eigen in the, of the axes in this direction; this is going to be inversely proportional to the eigenvalue in that direction or the corresponding eigenvalue of that eigenvector. And we are just saying that, eigenvalues are increasing; it is not necessary that all the eigenvalues are increasing, some may stay constant.

Because see, this matrix has how? If this is invertible, how many eigenvalues it will have? If it is invertible that is a why full rank matrix, right. How many eigenvalues it will have, non zero eigenvalues?

Student: d.

It will have d. So, some of them is may not be increasing, but some may be increasing.

So, none of them is not decreased, again see like we have to go back and see that; if I have adding a positive semi definite matrix like this, whether its eigenvalues are going to increase, so we have to go back to this analysis, if you want to make that argument. But let us say with this they are going to increase; if not all, some are increasing. So, because of that, this principal axis lines are shrinking, right.

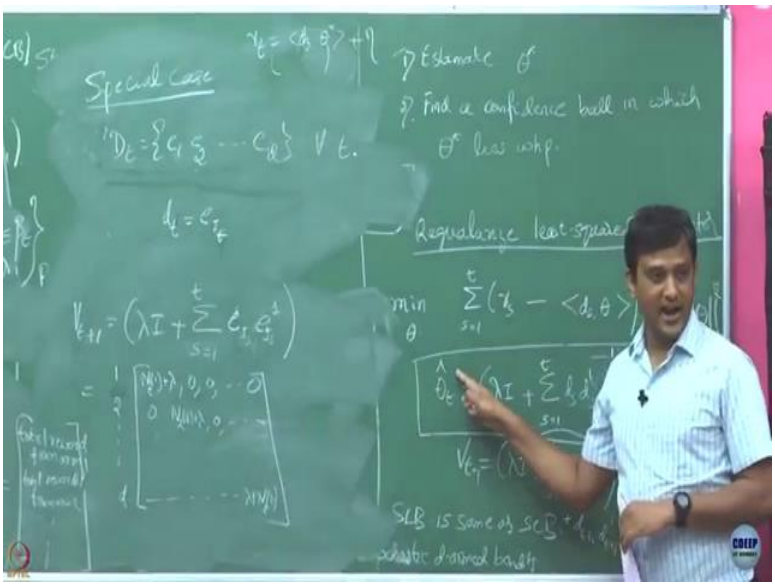
Student: (Refer Time: 33:15).

Yeah.

Student: (Refer Time: 33:17).

So, yes, if the eigenvalues of this matrix have increased; then the size of the ball has decreased in the axis, in the principal axis along which the eigenvalue has increased. And we are saying that eventually every principal axis along all the directions that eigenvalue is increasing. So, because of that it this interval keeps on shrinking. So, let us try as a special case, the case where my d_t is unit vectors, ok.

(Refer Slide Time: 34:05)



Let us as a special case. So, I have told you already if D_t is this, it is nothing but stochastic D armed bandit, right. So, we already know how my confidence intervals looks for this case, right.

So, let us see, let me denote d_s; now instead of d_s, I am going to just write it as let us say e_s; e_s is nothing but let me write it as e_{I_s} simply.

So, this is I_s is what. So, d_s is the decision I made, instead of let me say. So, d_t is the arm I pulled in round t from the set D_t. So, now, this is this let me call that whatever that d_t is simply e_{I_t}; that means, I pulled an arm whose only I_t th component is which is I_t only I_t - th component is 1.

So, now let us go back and plug it here. So, what you are going to get. So, what is this? e_i 's is a just a unit vector only in which I s component is nonzero, right. Now what is this is going to tell you? What is this going to look like? If I am going to ok; let me write it as a matrix now.

Student: (Refer Time: 36:25).

So, each element here is a matrix; in that matrix how many components are nonzero?

Student: Only one.

Only one component; and what is that component corresponds to?

Student: (Refer Time: 36:43).

That, it is index, right ok. So, let me call this as 1, 2, so the corresponding to d. What is the first line going to look like?

Student: 1, t plus 1.

1.

Student: (Refer Time: 37:02) one (Refer Time: 37:03) one bracket small t (Refer Time: 37:07).

So, is it like.

Student: Lambda plus lambda (Refer Time: 37:15).

Lambda times number of times.

Student: Lambda (Refer Time: 37:16).

Student: Lambda.

Yes.

Student: t and substituents.

Yeah till time t, how many times arm one has been played or how many times unit vector one has been played plus lambda and everything is going to be 0 like this, right. What is the second row is going to look like?

Student: (Refer Time: 37:46) 0 (Refer Time: 37:47) minus.

And similarly, what is this going to look like? Now, if you are going to look at V_{t+1} inverse; what it is playing the role? It is just like 1 by this quantities right; it is just like a diagonal matrix right, it is going to play the role of this. Now if you are going to have played and this quantity here; then I am going to look at this V_t inverse right, this is nothing but 1 divided by this quantity, 1 divided by this for every component, right. And what is this look like? Going to look like; so this is what? This is going to be a row vector right; r is a scalar, d_s is what?

Student: Column vector.

It is a column vector, right. Now, if you are going to look at over t rounds; what this going to look like?

Student: (Refer Time: 39:11).

So, I am just going to treat summation d_s r_s, where s is 1 to t; now d s I am replacing it by d_{It} e_{It}. What it is going to look like? This is going to be a column vector; what is the first number is going to look like?

Student: Sum of the reward.

Sum of the rewards from.

Student: (Refer Time: 39:36).

First arm, right.

Student: Whenever arm (Refer Time: 39:38).

Ah.

Student: Whenever unit vector (Refer Time: 39:40).

Whenever unit vector one is played; I mean the first unit vector is played. So, that is like total reward; how to write this? So, I am just going to write it in words ok; total reward from arm 1, similarly second one is a total reward from arm 2 like that, right.

Now, in this case, in this special case, what is the total reward from arm m? It is nothing but the rewards collected from a distribution whose mean value is theta one star right, because whenever. So, I removed this right; like r_t is nothing but what I said, it is s theta star plus noise, right. When this d_s is a unit vector only I_t -th term; so then that is means only that component remains plus noise.

So that means, this is the total reward collected from arm; that is nothing but the total rewards collected from a distribution whose mean value is theta 1 star. And this is the total reward collected from a distribution whose mean reward is theta 2 star and like this. Now, what is this theta hat in that case is giving you? Now, just combine these two

information's. So, this is V_t inverse and now this is the total reward across all this. So, now you see that, what is my theta i hat is going to be?

Student: (Refer Time: 42:01).

Yeah. So, what is that is going to turn out to be?

Student: Total reward (Refer Time: 42:05).

Total reward in the i-th row divided by.

Student: Lambda n.

Lambda n, right. So, this is nothing but what we had earlier; this is the exactly estimation that we had got an earlier right, except the fact that there is a lambda has come into picture here, but that is fine. So, this is nothing, but.

Student: (Refer Time: 42:34).

Which mean theta i star divided by whatever I have. So, in a way this will gave me theta i star and this decouples; for every component I can find it like this, right. So, for i-th component, this is theta. So, overall I can now find out theta i is nothing but this quantity right; theta 1 hat, theta 2 hat nothing but theta d hat. Am I correct? So, I am just applying this theta hat formula whatever I have here, with V_t hat computed like this and this term computed it like this. So, this now just gives a sample mean right for each arm.

Now, similarly now if you just go back and apply to this; you will see that this is nothing but a similar confidence bound what we had in the case of multi armed bandit setting, with this upper bound. So, you remember in the earlier case, we have this multi armed bandit; what was the confidence term?

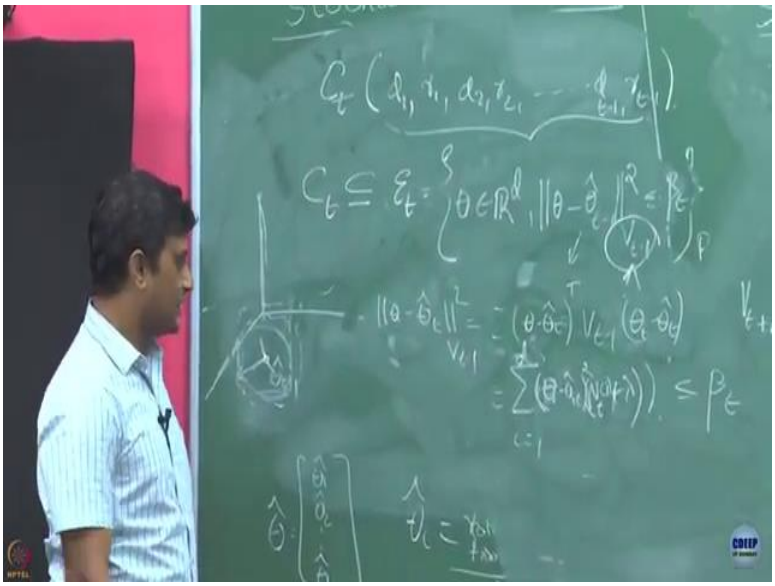
Student: (Refer Time: 44:11).

It was like some 2 log t divided by.

Student: (Refer Time: 44:14).

number of pulls, right.

(Refer Slide Time: 44:23)



You see that now because, let us just write down this; what is this? Theta minus theta t hat right. So, this is nothing by summation of theta i minus, so this is with V_{t-1} right. So, now, what, ok. So, this is by definition is nothing, but theta t, theta t hat.

Student: Transpose.

Transpose.

Student: V_{t-1}.

V_{t-1} and theta t minus theta t hat. Now, the structure of V_t is like that it is diagonal, ok. So, because of this diagonal nature; what is going to happen to this?

Student: (Refer Time: 45:08).

So, every time here, what it is going to happen is? This term remains like this, but it is going to be multiplied by the square of the corresponding differences, right. So, this is what, like so; now, if I am going to look at like this, it is going to look like what? Theta 1 minus theta 1 t hat.

Student: Squared.

Squared divided by, this is just V_t right; this is going to $N_t(1)$ plus lambda.

Student: Summation one.

Why summation?

Student: (Refer Time: 45:56) scalar (Refer Time: 45:58).

Ok. So, now what we are saying is that, total sum across all this directions is this; but does this not looking like a confidence terms we had in the bandit case, right.

Student: (Refer Time: 46:22).

Ok. If I am going to say this is now I want this to be Beta t right; this I want to be less than Beta t. So, if this is constant; now on this sum, it must be the case that for those which have large of these quantity, this difference is going to be the smaller, ok.

Student: Do they have to have that poses constrain to hold.

Yeah this constrain to hold, like this sum is going to be larger than. So, anyway what it is saying?

Student: Play more time (Refer Time: 47:07).

Student: You play more time when we when the interval (Refer Time: 47:12)

So, what I am doing? I am looking for all these thetas for such that this holds, right. So, what all this thetas will be? If something is already large; I mean the number of pulls is large, then this theta component is going to have a smaller.

Student: Range.

Range, right.

Student: (Refer Time: 47:33).

Beta is fixed, beta t is fixed; we are that is like this is for a given beta t.

Student: (Refer Time: 47:39).

Yeah, whatever.

Student: (Refer Time: 47:41).

Yeah. So, I just wanted to ensure that we have V_t here, not V_t's inverse; but at least with V_t we will see that, if some component has been explored much, then I have already high confidence and I will be, its range is going to be smaller now. Let us discuss, how the optimisticity comes in the next class ok. So, with this confidence set, we will see how to try to play an arm optimistically in the next class, ok. Let us stop here.