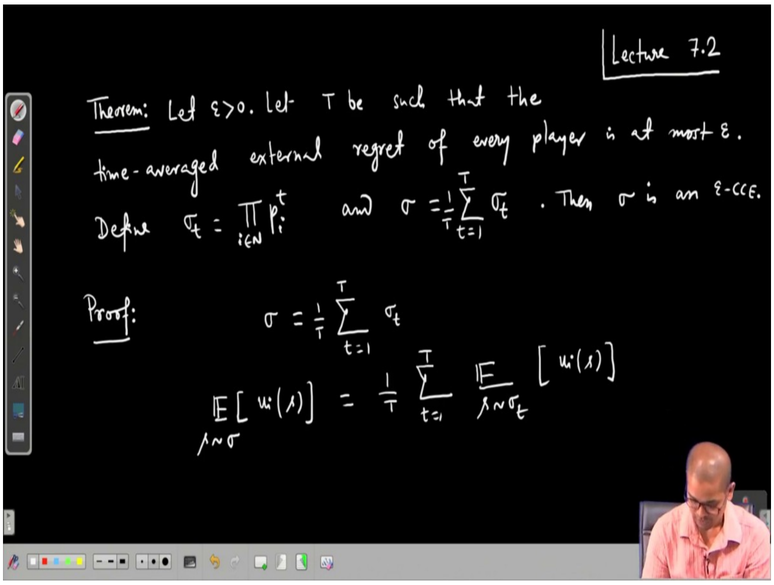
Algorithmic Game Theory
Prof. Palash Dey
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture - 32
Swap Regret and Correlated Equilibrium

Welcome. So, in the last class we were discussing the no external regret algorithm and its connection to coarse correlated equilibrium and we stated the connection so, in this class we will prove that.

(Refer Slide Time: 00:39)



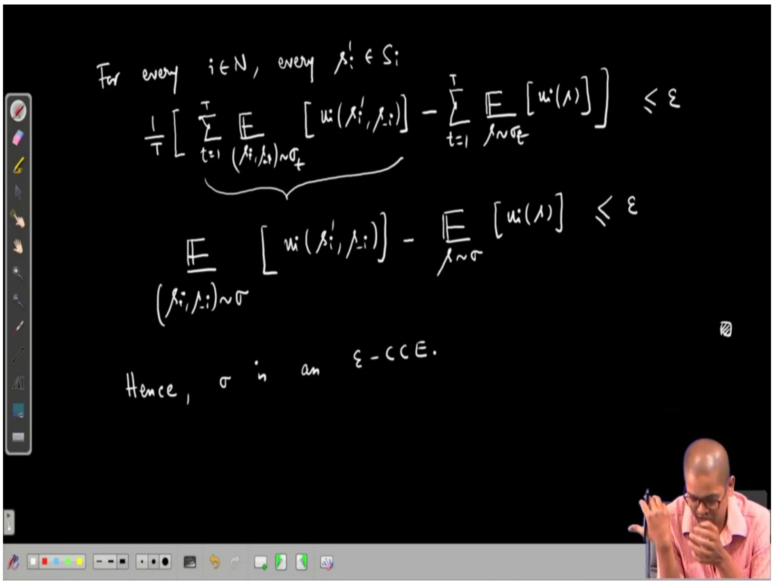
So, let us recall what was this the connection that we stated in the last class. So, let epsilon be any fraction greater than 0, let T be such that the time – averaged external regret of every player. Recall each player is running their own external no regret no external regret algorithm. So, it will go to 0, as T goes to ∞ that is the definition of no external regret algorithm the time average time average external regret of every player is at most ϵ.

Then define $\sigma_t = \prod_{i \in N} p_i^t$ p_i^t is the probability distribution that i-th player commits to in the t-th iteration $i \in N$ and σ is the average of this σ_ts, then σ is an ϵ coarse correlated equilibrium. Proof, ok. So, we start from here we have that σ by definition $\sigma = \frac{1}{T}\sum_{t=1}^{T}\sigma_t$.

So, expectation of sigma by linearity of expectation we get $\frac{1}{T}\sum_{t=1}^{T}$ expected utility of player i in the strategy profile s where s is sampled from σ. This is by definition because σ is the average of $\sigma_1,\ldots,\sigma_T$ this is by definition 1 over T summation t equal to 1 to capital T expectation s sampled from σ_t $u_i(s)$.

Now, you see that each player runs a no external regret algorithm and after capital T iterations, their external regret is at most ϵ.

(Refer Slide Time: 05:00)



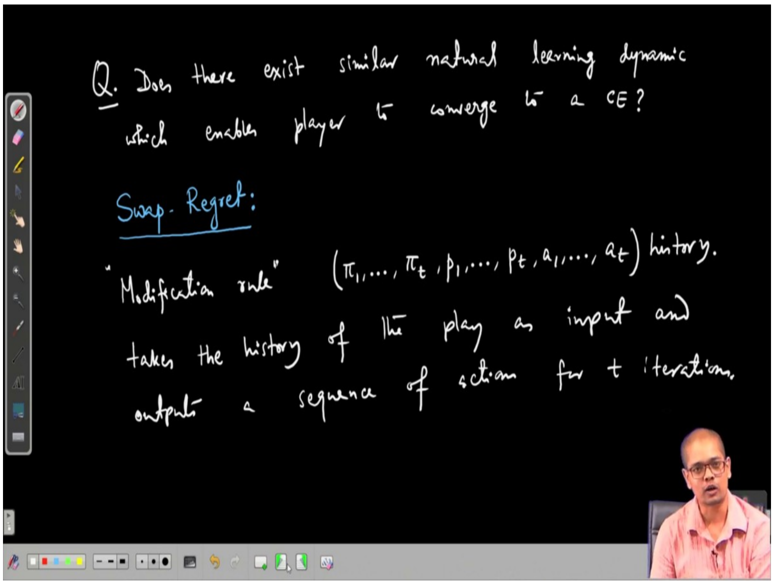
So, for every player $i \in N$ and every strategy $s_i' \in S_i$ that is by from by definition 1 over T, the time averaged external regret, what is the external regret? Summation t equal to 1 to capital T, now what is the expected utility of player i in the t-th iteration? Expectation (s_i, s_{-i}) sampled from σ_t and instead of playing s_i player i plays s_i'.

So, this is the utility that if player i plays the fixed action s_i' minus the utility that it is actually getting no s from σ_t $u_i(s)$. Now, this is at most ϵ. Recall, that in the external regret is defined as the maximum of the first term minus the second term, the second term remains fixed and if we if I take the maximum to the first term, then it is less than equal to ϵ.

In particular, if I take any s_i', then it will continue to be less than equal to ϵ and this is exactly what is that the requirement of ϵ coarse correlated equilibrium if we just replace this σ_ts with σ. This is exactly expectation of (s_i, s_{-i}) sampled from σ $u_i(s_i, s_{-i})$ minus expectation s sampled from σ u_i s by deviating to some other strategy $s_i^{'}$ the players utility can increase, but it should not increase by more than ϵ.

And, that is exactly the requirement of coarse correlated ϵ coarse correlated equilibrium. Hence, sigma is an ϵ coarse correlated equilibrium. This concludes the proof, good, very good. So, this natural no regret no external regret dynamics relate to coarse correlated equilibrium if each player plays this epsilon or a no externally regret algorithm, then they eventually jointly converge to ϵ coarse correlated equilibrium.

(Refer Slide Time: 09:15)

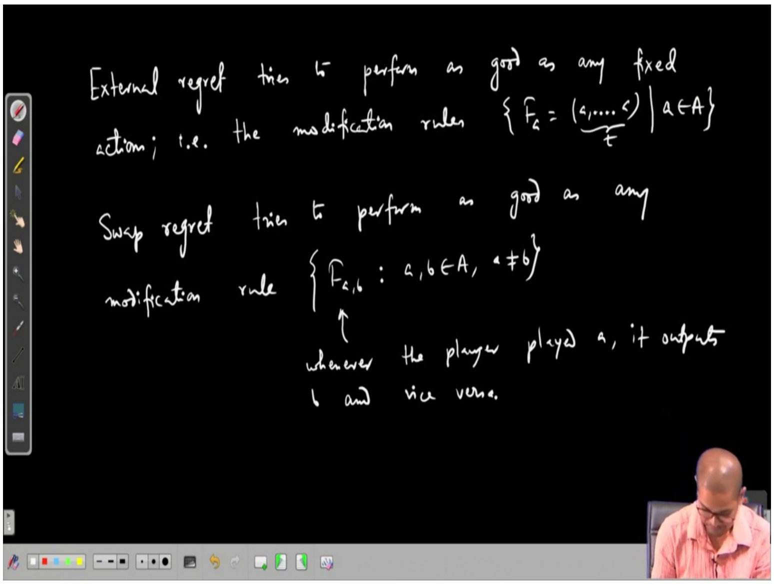


So, the next natural question is does there exist. So, question does there exist similar natural learning dynamics which enables player to converge to correlated equilibrium? So, for coarse correlated equilibrium we have no external regret dynamics. How about for correlated equilibrium, does there exist any such dynamics? And, the answer is yes and that we will study next which is called swap regret.

To motivate the idea of swap regret, let us let us view the idea of no external regret or external regret from another perspective which is called modification rule. So, modification rule it is like after playing for capital T iterations player thinks what he could have done and he would have got more utility. So, after playing T iterations all the information it has; that means, it has say $\pi_1, \ldots, \pi_t$ the payoff functions in each iteration the probability distribution the distribution it had committed $p_1, \ldots, p_t$ and the actions that it played a 1 to a t this is called history.

And, it thinks what it could have done or which other strategy it could have followed so that you know it is utility would utility would have been; would have been more. So, a modification rule modification rule takes input takes history as input; that means, and whatever has happened takes the history of the play and outputs a sequence of actions for T iterations. It is like in each iteration what action I could have played that is what is called modification rule.

(Refer Slide Time: 13:27)

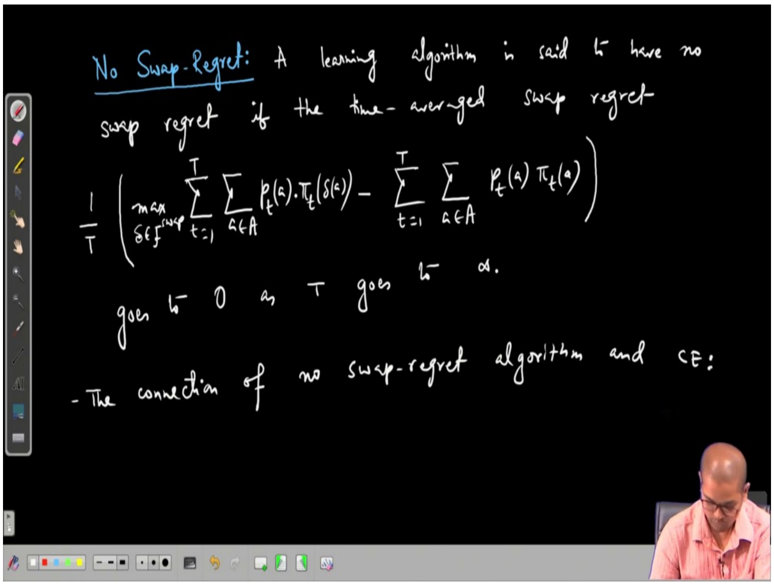


And, the notion of external regret considers only modification rules which output fixed actions. So, external regret tries to perform as good as any fixed actions that is the modification rules that it considers is of the functions of the form F a which is like it outputs only a; every iteration at play a is for all $a \in A$. The idea of swap regret makes the modification rule richer and it ask, ok.

So, I could have played instead of playing a whenever I played a, I could have played b and whenever I played b I could have played a and so on, ok. So, this is the idea of swap. So, the idea of swap regret swap regret tries to perform as good as the as good as any modification as good as any modification rule $F_{a,b}$ where $a, b \in A, a \neq b$.

So, what is $F_{a,b}$? $F_{a,b}$ is whenever the player plays a player played a, it outputs b, recall that the output is a sequence and vice versa. So, of all these modification rules and their combinations that is the benchmark for swap regret. So, this so, with this motivation let us formally define no swap regret algorithm no swap regret.

(Refer Slide Time: 17:12)



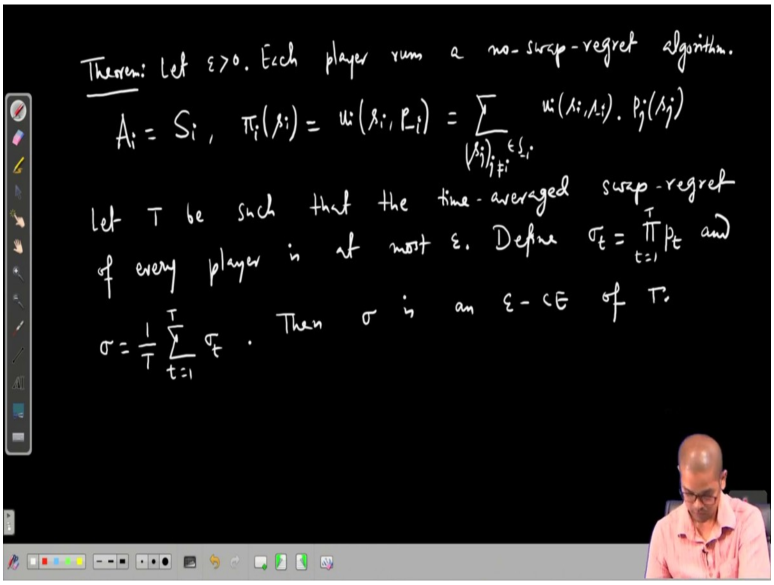
What is no swap regret? A learning dynamic or a learning algorithm is said to have no swap regret if the time average swap regret if the time averaged swap regret. So, what is time average swap regret? So, what is the utility of player i utility of the player? So, T iterations capital T iterations and in every iteration it picks an action a with probability $p_t(a)$ and gets a payoff of $\pi_t(a)$.

And, it considers instead of playing a, if I have would have played b and instead of playing b I could have played c or something like that some other permutation of the actions I could have played. So, let us call that delta. So, same thing yeah need not be permutation any function. So, $a \in A$ and whenever a was picked it ponders whether instead of playing a I could have played $\delta(a)$, ok.

So, this is then this would have been my utility and I want to maximize my utility over all actions over all functions delta in F^{swap}. F^{swap} is the modification rules allowed in the swap regret. So, this is the absolute regret and I want to I am interested in time average regret so, 1 over T. So, if the time average swap regret goes to 0, as T goes to ∞, good.

So, we will see why or how it makes sense, but before that or does there exist a no swap period algorithm that we will see later but, first let us say its connection with correlated equilibrium. So, the connection of no swap regret algorithm and correlated equilibrium that we will consider next.

(Refer Slide Time: 21:29)

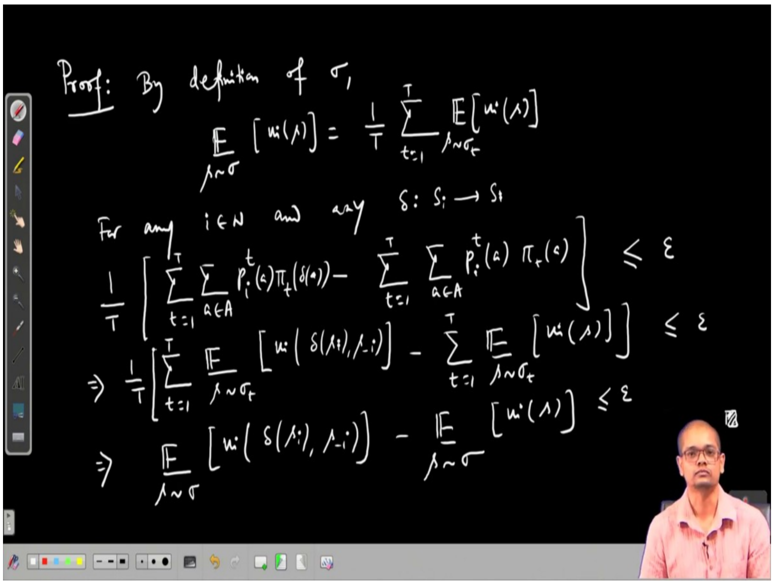


So, here is a theorem. As usual let epsilon be any fraction arbitrarily small greater than 0 and ok each player as usual as before each player runs no swap regret algorithm. Now, when I say no swap regret algorithm, what is their what is the actions? As usual before; that means, action of player i is the strategies available to player i, they need to commit to a probability distribution. So, exact the same setting remains.

And, the p of function $\pi_i(s_i)$ is exactly like utility of player i when it plays s_i and other players are playing according to their committed probability distribution recall which is $s_j, j \neq i$ utility of (s_i, s_{-i}) and the probability that this particular strategy profile is picked is $p_j(s_j)$, ok. So, the setting is exactly as before that each player runs a no swap regret algorithm.

Before they were running a no external regret algorithm, now they are just running a no swap regret algorithm, everything else remains same. And, let capital T be such that the time averaged swap regret of every player is at most epsilon. Then define as usual exactly as usual define σ_t is the $\times_{t=1}^{T} p_t$ and σ is the average of this σ_ts, then σ is an ϵ correlated equilibrium of the strategic form game Γ ok.

(Refer Slide Time: 25:15)



And, the similar proof works. So, let us prove it. Proof: so, by definition of sigma we have what is the expected utility of player i if it plays according to σ? $u_i(s)$is 1 over T is sampled from σ_t station $u_i(s)$ sorry, this is t equal to 1 to capital T and s sampled from σ_t ok. Now, since each players time average swap regret is at most ϵ after T iterations, so, for any player $i \in N$ and any switching function $\delta : S_i \rightarrow S_i$ what is the utility of player i?

The utility of player i is summation t equal to 1 to capital T and p_t is the committed probability distribution. So, $\sum_{a \in A} p_i^t(a)$ p_i^t is the i-th players commuted probability distribution in the t-th iteration. And, with a is the probability with which it is picked and p_i^t and now, it considers instead of playing a, I if would have played $\delta(a)$ then what would have happened. Capital T $\sum_{a \in A} p_i^t(a) \pi_t(\delta(a))$. So, this is the absolute difference of utility which is regret and if I divide by capital T it is time average swap regret, this is at most ϵ.

Now, by definition this is 1 over T summation t equal to 1 to capital T expectation s sampled from σ_t u_i instead of playing s_i it is playing $\delta(s_i), s_{-i}$ minus summation t equal to 1 to capital T expectation a sampled from σ_t $u_i(s)$ this is less than ϵ. I have just written what do we mean by expectation of $u_i(s)$ where a sampled from σ_t that is just this expression.

And, now the first term is nothing but expected utility of player i if it modifies it action with this using this switching function; that means, instead of playing s_i it plays $\delta(s_i)$ minus this is the expected utility of player i if it simply follows σ. And, this utility can be less, but it should not be more less than amount of ϵ which is exactly prove that sigma is a epsilon correlated equilibrium of the game. So, that concludes the proof, ok.

Thank you.