BOOK CHAPTER PRESENTATION: LEARNING, REGRET MINIMIZATION AND EQUILIBRIA.

VISHVAS VASUKI

Part 1. Outline

1. Intro about the paper

Book chapter presentation: Learning, Regret Minimization and Equilibria. [1] Authors: Avrim Blum, Yishay Mansour. A few results appeared first in COLT 2005: From external to internal regret.

We will see some important results and techniques from this work.

2. Plan

20 minutes available. 8 minutes for introducing the model. 10 minutes for proving result about making an internal regret alg from an external regret alg. 2 minutes remarking about connection with game theory and equilibria.

Part 2. Introducing the model

3. Players, strategies, utilities

Players $P = \{p_i\}$. A strategy is not a move but an algorithm to make moves. S_i: strategy set of p_i. Strategy vector (strategy profile): $s = (s_1, ..s_n)$. s_{-i}: s sans s_i.

3.1. **Mixed/ randomized strategies.** Independent mixed strategy of i: a Prob Distr over $S_i : D_i$.

Mixed strategy profile, perhaps p_i coordinated: Probability distribution over $\times_i S_i$: D.

3.2. **Utility.** Preference ordering of outcomes for i: Cost, utility of strategy: $c_i(s) = -u_i(s)$.

3.2.1. *ϵ dominated strategy.* s_i dominated by s_i' if : $u_i(s_i', s_{-i}) \geq u_i(s_i, s_{-i}) + \epsilon$.

4. Repeated games with partial info about utilities

p_1 in uncertain environment (p_{-1}); utilities of p_{-1} not known. Eg: Choosing a route to go to school.

4.1. **Model.** Same game repeated T times; At time t: p_1 uses online alg H to pick distr $D_H^{(t)}$ over S_1. p_1 picks action $k_1^{(t)}$ from $D_H^{(t)}$. Loss/ cost function for p_1: $c_1 : \times_i S_i \to [0, 1]$. $c_1^{(t)}(k_1^{(t)}) := c_1(k_1^{(t)}, D_{-1}^{(t)})$, $c_1(D) := E_{x\sim D}[c_1(x)]$.

4.1.1. *Model with full info about costs.* H gets cost vector $c_1^{(t)} \in [0,1]^{|S_1|}$, pays cost $c_1(D_H^{(t)}, D_{-1}^{(t)}) = E_{k_1^{(t)} \sim D_H^{(t)}}[c_1(k_1^{(t)}, D_{-1}^{(t)})] = E_{k_1^{(t)} \sim D_H^{(t)}}[c_1^{(t)}(k_1^{(t)})]$.

Total loss for H: $L_H^{(T)} = \sum c_1(D_H^{(t)}, D_{-1}^{(t)})$.

4.1.2. *Model with partial info about costs.* Aka Multi Armed Bandit (MAB) model. p_1 (or H) pays cost for $k_1^{(t)}$: $c_1(k_1^{(t)}, D_{-1}^{(t)}) = c_1^{(t)}(k_1^{(t)})$.

Total loss for H: $L_H^{(T)} = \sum c_1(k_1^{(t)}, D_{-1}^{(t)})$.

4.1.3. *Goal.* Minimize $\frac{L_?^{(T)}}{T}$. Maybe other p_i do the same. $D_{-1}^{(t)}$ and $c_1^{(t)}$ can vary arbitrarily over time; so, model is adversarial.

4.2. **Regret analysis.** H incurs loss $L_H^{(T)}$; p_1 sees simple policy π would have had much lower loss. Comparison class of algs G. π best alg in G: $L_\pi^{(T)} = min_{g \in G} L_g^{(T)}$. Regret $R_G = L_H^{(T)} - L_\pi^{(T)} = max_{g\in G}(L_H^{(T)} - L_g^{(T)})$.

4.2.1. *Goal.* Minimize R_G.

4.2.2. *Lower bound for regret wrt all policies.* $G_{all} = \{g : T \to S_1\}$: $\exists$ sequence of loss vectors $c_1^{(t)}$: $R_{G_{all}} \geq T(1 - |S_1|^{-1})$.

So, must restrict G.

4.3. **External regret.** Aka Combining Expert Advice. $G = \left\{i^T : i \in S_1\right\}$, policies where all $k_1^{(t)}$ are the same; π is best single action. $L_\pi^{(T)} = \sum c_1(\pi, D_{-1}^{(t)})$.

If H has low external regret bound: H matches performance of offline alg. [**Find proof**]. H comparable to optimal prediction rule from some large hyp class H. [**Find proof**].

4.3.1. *Rand Weighted majority alg (RWM).* Suppose $c_1^{(t)} \in \{0,1\}^{|S_1|}$. Treat S_1 as a bunch of experts: Want to put as much wt as possible on best expert. Let $|S_1| = N$. Init weights $w_i^{(1)} = 1$, total wt $W^{(1)} = N$, $Pr_{D_H^{(1)}}(i) = N^{-1}$.

If $c_1^{(t-1)}(i) = 1$, $w_i^{(t)} = w_i^{(t)}(1-\eta)$, $Pr_{D_1^{(t)}}(i) = \frac{w_i^{(t)}}{W^{(t)}}$. [**Find proof**]. Like analysis of mistake bound of panel of k experts in colt ref.

For $\eta < 2^{-1}$, $L_H^{(T)} \leq (1+\eta)\min_{i\in S_1} L_i^{(t)} + \frac{\ln N}{\eta}$. Any time H sees significant expected loss, big drop in W. $W^{(T+1)} \geq max_i w_i^{(T+1)} = (1-\eta)^{\min_i L_i^{(T)}}$. [**Incomplete**].

For $\eta = \min\left\{\sqrt{\ln N/T}, 2^{-1}\right\}$: $L_H^{(T)} \leq \min_i L_i^{(T)} + 2\sqrt{T \ln N}$. If T unknown, use 'guess and double' with const loss in regret. [**Find proof**].

4.3.2. *Polynomial weights alg.* Extension of RWM to $c_1^{(t)} \in [0,1]^{|S_1|}$. Wt update is $w_i^{(t)} = w_i^{(t)}(1 - \eta c^{(t-1)}(i))$. $L_H^{(T)} \leq \min_i L_i^{(T)} + 2\sqrt{T \ln N}$. [**Find proof**].

4.3.3. *Rand Alg Lower bounds.* If $T < \log_2 N$: For any online alg H, $\exists$ stochastic generation of losses: $E[L_H^{(T)}] = T/2$, but $\min_i L_i^{(t)} = 0$: at t=1 let N/2 actions get loss 1; at time t: half the actions which had a loss 0 at time t-1 get loss 1; so, probability mass on actions with $0 = 2^{-1}$.

If N=2, $\exists$ stochastic generation of losses: $E[L_H^{(T)} - \min_i L_i^{(T)}] = \Omega(\sqrt{T})$. [**Find proof**].

4.3.4. *Convergence to equilibrium: 2 player constant sum repeated game.* All p_i use alg H with external regret R; Value of game: (v_i). Avg loss: $\frac{L_H^{(T)}}{T} \leq v_i$. [**Find proof**]. If $R_G = O(\sqrt{T})$, convergence to v_i.

Part 3. Models to be introduced if there is time

5. Nash equilibrium

Defn: D or $\{D_i\}$ where even if all p_i know all D_i, no treachery profitable. Maybe D not unique. So each p_i can decide D_i if he knows D_{-i}.

5.1. **Randomized (mixed) strategies.** Not Pure strategy s, but distr D. Risk neutral p_i maximize $u_i(D) = E_{s\sim D}[u_i(s)]$, with $Pr_{s\sim D}(s) = \prod_i Pr_{s_i \sim D_i}(s_i)$.

5.2. **Existance of Equilibria.** Any game with $|P|, |S_i|$ finite, $\exists$ mixed strategy Nash equilib. [**Find proof**].

5.3. ϵ **Nash equilib.** A special case: $\forall i, D' : u_i(D) \geq u_i(D_i', D_{-i}) - \epsilon$

6. Correlated equilibrium D

(Aumann). Coordinator has distr D, samples s from D, tells each p_i its s_i. p_i not told s_j, but knows it is correlated to s_i; so knows all $Pr(s_{-i}|s_i)$. D known to every p_i. D is correlated equilib if it is not in any p_i's interest to deviate from s, assuming other p_i follow instructions:
$E_{s_{-i}\sim D|s_i}[u_i(s_i, s_{-i})] \geq E_{s_{-i}\sim D|s_i}[u_i(s_i', s_{-i})]$.

Mixed strategy Nash equilibrium is the special case where D_i are independently randomized (with diff coins).

6.1. **Regret defn.** $f_i : S_i \to S_i$, regret $r_i(s, f) = u_i(f_i(s_i), s_{-i}) - u_i(s)$:
$E_{s\sim D}[r_i(s, f_i)] \geq 0$.

6.2. ϵ **correlated equilibrium.** $E_{s\sim D}[r_i(s, f_i)] \leq \epsilon$.

6.3. **Traffic light/ Chicken.** $C = \begin{pmatrix} (-100,-100) & (1,0) \\ (0,1) & (0,0) \end{pmatrix}$. s = (1, 2) and (2, 1) stable; so coordinator picks one randomly. This correlation increases payoff as the low expected utility mixed strategy $D_i = (101^{-1}, 1 - 101^{-1})$ is avoided.

Part 4. Important results

6.4. **Low external regret alg in partial cost info model.** Exploration vs exploitation tradeoff in algs.

Alg MAB: Divide time T into K blocks; in each time block $\tau + 1$: explore and get cost vector: execute action i at random time to get vector of RV's: $\hat{c}^{(\tau)}$, also exploit: use distr $D^{(\tau)}$ as strategy; pass $\hat{c}^{(\tau)}$ to full info external regret alg F with ext regret $R^{(K)}$ over K time steps; get distr $D^{(\tau+1)}$ from F.

Max Loss during exploration steps: NK. RV for total loss of F over K time blocks: $\hat{L}_F^{(T)} = \frac{T}{K}\sum_\tau p^\tau c^\tau \leq \frac{T}{K}(min_i \hat{L}_i^{(K)} + R^{(K)}$. Taking expectation, $L_{MAB}^{(T)} = E[\hat{L}_{MAB}^{(T)}] = E[\hat{L}_F^{(T)} + NK] \leq \frac{T}{K}(E[min_i \hat{L}_i^{(K)}] + R^{(K)}) + NK \leq \frac{T}{K}(min_i E[\hat{L}_i^{(K)}] + R^{(K)}) + NK \leq min_i L_i^{(T)} + \frac{T}{K}R^{(K)} + NK$.

Using the $O(\sqrt{K \log N})$ alg, with $K = (\frac{T}{K}R_K)$, we get $L_{MAB}^{(T)} \leq min_i L_i^{(T)} + O(T^{2/3}N^{1/3}\log N)$.

6.5. **Swap regret.** Comparison alg (H,g) is H with some swap fn $g : S_1 \to S_1$.

6.5.1. *Internal regret.* A special case: Swap every occurance of action b_1 with action b_2. Modification fn: $switch_i(k_i, b_1, b_2) = k_i$ except $switch_i(b_1, b_1, b_2) = b_1$.

6.5.2. *Low Internal regret alg using external regret minimization algs.* Let $N = |S_i|$; $(A_1, .., A_N)$ copies of alg with external regret bound R. Master alg H gets from A_i distr $q_i^{(t)}$ over S_i; makes matrix $Q^{(t)}$ with $q_i^{(t)}$ as rows; finds stationary distr vector $p^{(t)} = p^{(t)}Q^{(t)}$: Picking $k_i \in S_i$ same as picking A_j first, then picking $k_i \in S_i$; gets loss vector $c^{(t)}$; gives A_i loss vector $p_i^{(t)}c^{(t)}$.

$\forall j : L_{A_i} = \sum_t p_i^{(t)} \langle c^{(t)}, q_i^{(t)} \rangle \leq \sum_t p_i^{(t)} c_j^{(t)} + R$. Also, Sum of percieved losses = actual loss. So, for any swap fn g, $L_H^T \leq \sum_i \sum_t p_i^{(t)} c_{g(i)}^{(t)} + NR = L_{F,g}^{(T)} + NR$.

Thence, using polynomial weights alg, swap regret bound $O(\sqrt{|S_1| T \log |S_1|})$.

6.5.3. *Convergence to Correlated equilibrium.* Every p_i uses strategy with swap regret $\leq R$: then empirical distr Q over $\times_i S_i$ is an $\frac{R}{T}$ correlated equilibrium. $R = L_H^{(T)} - L_{H,g}^{(T)} = \sum_t E_{s^{(t)} \sim D^{(t)}}[r_i(s, g)] = T E_{s \sim Q}[r_i(s, g)]$.

Convergence if all players have sublinear swap regret.

6.5.4. *Frequency of dominated strategies.* p_1 uses alg with swap regret R over time T; w: avg over T of prob weight on ϵ dominated strategies; so $\epsilon w T \leq R$; so $w \leq \frac{R}{T\epsilon}$.

If alg minimizes external regret using polynomial weights alg, freq of doing dominated actions tends to 0.

References

[1] Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V. Vazirani. *Algorithmic Game Theory.* Cambridge University Press, New York, NY, USA, 2007.