CSE 6319 Notes 4: Adaptive Decision Making

(Last updated 4/17/23 8:30 AM)

R 16/17/18; N 4; KP 18

(Best. Book. Ever. on such things . . . `https://www.amazon.com/dp/0521841089/`)

(Best. 35 pages. Ever. on such things . . . Chapter 7 of `http://www.masfoundations.org`)

(If you don't know him . . ., you should . . . `https://en.wikipedia.org/wiki/Leslie_Valiant`)

(Fun with convergence . . . `https://en.wikipedia.org/wiki/Collatz_conjecture`)

I hear the voices, and I read the front page, and I know the speculation. But I'm the decider, and I decide what is best. George W. Bush

`https://en.wikipedia.org/wiki/Fictitious_play` is an early *slow* learning approach to Nash equilibria (`berger.pdf` `robinson.pdf`). Java code is on the webpage.

4.A. Best-Response Dynamics

While the current outcome **s** is not a pure Nash equilibrium
pick an arbitrary agent i and an arbitrary beneficial deviation s'_i
update the outcome to $(s'_i, \mathbf{s}_{-i})$

Best-response dynamics converges for potential games (R p. 217), but possibly using an exponential number of iterations.

Approximate Pure Nash Equilibria in Selfish Routing Games

(R) Definitions 14.5/16.2: $\in$-Pure Nash Equilibrium

While the current outcome **s** is not a $\in$-PNE
among all agents with an $\in$-move, let i denote an
agent who can obtain the *largest* cost decrease
update the outcome to $(s'_i, \mathbf{s}_{-i})$

Theorem 16.3: Consider an atomic selfish routing game (Notes 2, p. 3) where:

1. *All players have a common source vertex and a common sink vertex.*
2. *Cost functions satisfy the "α-bounded jump condition," meaning $c_e(x+1) \in [c_e(x), \alpha \cdot c_e(x)]$ for every edge e and positive integer x.*
3. *The MaxGain variant of ϵ-best-response dynamics is used: in every iteration, among players with an ϵ-move available, the player who can obtain the biggest absolute cost decrease moves to its minimum-cost deviation.*

Then, an ϵ-PNE is reached in $(\frac{k\alpha}{\epsilon} \log \frac{\Phi(\mathbf{s}^0)}{\Phi_{\min}})$ iterations.

x - number of players using an edge

k - number of agents

ϵ - maximum deviation from PNE for any agent

Φ - potential function for atomic selfish routing game

α - from the *α-bounded jump condition*, this constant implies that when a new player is added to an edge, the cost to all players using that edge increases by a most a factor of α.

Fast Convergence for Smooth Potential Games (skip)

4.B. NO-REGRET DYNAMICS

The Model

At time $t = 1, 2, \ldots, T$:

- A decision-maker picks a mixed strategy p^t — that is, a probability distribution — over its actions A.
- An adversary picks a cost vector $c^t : A \rightarrow$ [-1, 1].
- An action a^t is chosen according to the distribution p^t, and the decision-maker incurs cost $c^t(a^t)$. The decision-maker learns the entire cost vector c^t, not just the realized cost $c^t(a^t)$.

Universally consistent = Hannan consistent = No-regret (sublinear regret)

(External) *Regret* of an action sequence $a^1, \ldots . a^T$ is (total loss - best single action)

$$\frac{1}{T}\left[\sum_{t=1}^{T} c^t(a^t) - \min(a \in A) \sum_{i=1}^{T} c^t(a)\right]$$

(Important that min is outside the summation. See R p. 232)

Definition 17.3 (No-Regret Algorithm) . . . for every $\epsilon > 0$ there exists a sufficiently large time horizon T(ϵ) . . . the expected regret is at most ϵ.

Online Decision Making

Binary prediction with expert advice and a perfect expert, KP example 18.1.1 - stock market (R Problem 17.2)

Majority of *leaders* (those without any mistakes) - guarantees no more than lg n mistakes

Follow random leader (FRL) - guarantees no more than ln n expected mistakes

Function of majority size:

Given any function $p : [1/2, 1] \to [1/2, 1]$, consider the trader algorithm A_p: When the leaders are split on their advice in proportion $(x, 1-x)$ with $x \geq 1/2$, follow the majority with probability $p(x)$.

THEOREM 18.1.4. *In the binary prediction problem with n experts including a perfect expert, consider the trader algorithm A_p that follows the majority of leaders with probability $p(x) = 1 + \log_4 x$ when that majority comprises a fraction x of leaders. Then for any horizon T, the expected number of mistakes made by A_p is at most $\log_4 n$.*

Without perfect expert (KP 18.2)

Weighted Majority Algorithm

Fix $\epsilon \in [0, 1]$. On each day t, associate a weight w_i^t with each expert i.
Initially, set $w_i^0 = 1$ for all i.
Each day t, follow the ***weighted majority*** opinion: Let U_t be the set of experts predicting up on day t, and D_t the set predicting down. Predict "up" on day t if

$$W_U(t-1) := \sum_{i \in U_t} w_i^{t-1} \geq W_D(t-1) := \sum_{i \in D_t} w_i^{t-1}$$

and "down" otherwise.

At the end of day t, for each i such that expert i predicted incorrectly on day t, set

$$w_i^t = (1-\epsilon) w_i^{t-1}.$$

Thus, $w_i^t = (1-\epsilon)^{L_i^t}$, where L_i^t is the number of mistakes made by expert i in the first t days.

THEOREM 18.2.3. *Suppose that there are n experts. Let $L(T)$ be the number of mistakes made by the Weighted Majority Algorithm in T steps with $\epsilon \leq \frac{1}{2}$, and let L_i^T be the number of mistakes made by expert i in T steps. Then for any sequence of up/down outcomes and for* every *expert i, we have*

$$L(T) \leq 2(1+\epsilon) L_i^T + \frac{2 \ln n}{\epsilon}$$

Multiplicative Weights (KP 18.3.2)

Fix $\epsilon < 1/2$ and n possible actions.
On each day t, associate a weight w_i^t with the i^{th} action.
Initially, $w_i^0 = 1$ for all i.
On day t, use the mixed strategy $\mathbf{p}^t$, where

$$p_i^t = \frac{w_i^{t-1}}{\sum_k w_k^{t-1}}.$$

For each action i, with $1 \le i \le n$, observe the loss $\ell_i^t \in [0,1]$ and update the weight w_i^t as follows:

$$w_i^t = w_i^{t-1} \exp(-\epsilon \ell_i^t).$$

THEOREM 18.3.7. *Consider the Multiplicative Weights Algorithm with n actions. Define*

$$\overline{L}_{\mathrm{MW}}^T := \sum_{t=1}^{T} \mathbf{p}^t \cdot \boldsymbol{\ell}^t,$$

where $\boldsymbol{\ell}^t \in [0,1]^n$. Then, for every loss sequence $\{\boldsymbol{\ell}^t\}_{t=1}^T$ and every action i, we have

$$\overline{L}_{\mathrm{MW}}^T \le L_i^T + \frac{T\epsilon}{8} + \frac{\log n}{\epsilon},$$

where $L_i^T = \sum_{t=1}^T \ell_i^t$. In particular, taking $\epsilon = \sqrt{\frac{8 \log n}{T}}$, we obtain that for all i,

$$\overline{L}_{\mathrm{MW}}^T \le L_i^T + \sqrt{\frac{1}{2} T \log n};$$

i.e., the regret $\mathcal{R}_T(\mathrm{MW}, \boldsymbol{\ell})$ is at most $\sqrt{\frac{1}{2} T \log n}$.

Coarse Correlated Equilibria and No-Regret Dynamics (R p. 240)

4.C. SWAP REGRET AND THE MINIMAX THEOREM

Swapping function $\delta : S_i \longrightarrow S_i$

Correlated Equilibrium: $\mathbf{E}_{\mathbf{s}\sim\sigma}[C_i(\mathbf{s})] \le \mathbf{E}_{\mathbf{s}\sim\sigma}[C_i(\delta(s_i), \mathbf{s}_{-i})]$

Definition 18.2 (Swap Regret): Defined in terms of the more general notion of swapping function mapping *all* occurences of an action to another action.

Definition 18.3 (No-Swap-Regret Algorithm): Similar to Definition 17.3

Proposition 18.4 (No-Swap-Regret Dynamics and CE)

Theorem 18-5 (Black-Box Reduction): *If there is a no-external-regret algorithm, then there is a no-swap-regret algorithm.*

The substantial element of the proof is the stationary distribution of a Markov chain.

Theorem 18.7 (Minimax Theorem): Usually proven using a fixed-point theorem, but the no-regret algorithm (e.g. multiplicative weights) reveals a mixed strategy.