Danger of no-regret learning

	L	R
U	1, 1	3, 0
D	0, 0	2, 1

U dominant

(U,L) Nash

row player: what should I do

Stackelberg equilibrium

choose strategy player 1 (row)

know opponent will best respond

maximize your value

mixed strategy: $x > \frac{1}{2}$ D & $1-x$ U

best response ⇒ R

more generally: leader select

random strategy α_i & $s_i \in S$

assumes other players play Nash

of the resulting game maximizing

his value

Stackelberg value = supremum

2 person game:

$\alpha_s \geq 0 \quad s \in S_1 \qquad \sum_s \alpha_s = 1$

second player $z^* = \arg\max_z \sum_s \alpha_s u_2(s,z)$

$$\sup_\alpha \sum_s \alpha_s u_1(s, z^*)$$

Claim: 2 person 0-sum

Stackelberg value = Nash value

A payoff matrix

your strategy x (payoff Ax options)

$$\arg\min Ax$$

choose x to maximize this

Recall x is Nash.

Suppose you play $\alpha_s \geq 0$ so that

$$z^* = \arg\min_z \sum_s \alpha_s u_2(s,z)$$

is unique

no-regret player:

expected value for any strategy z

after T times $\qquad T \cdot \sum_s \alpha_s u_2(s,z)$

T is large enough to have
error in expectation + error of regret
smaller than T · gap between argmin
& next

$\Rightarrow$ no-regret player must play Σ^*

Example 2.

	L	M	R
U	$0, \varepsilon$	$-2, -1$	$-2, 0$
D	$0, -1$	$-2, 1$	$2, 0$

leader

(U, L) Nash

$(\frac{1}{2}, \frac{1}{2}, R)$

Stackelberg value = 0
cannot play D with prob > $\frac{1}{2}$

If opponent is mean-based no-regret learner

Learner strategy: overall time T

U $T/2$ times ~ best response L
payoff for leader ~ 0

switch to D x times then his historic values

$L: \ \varepsilon \frac{T}{2} - x, \qquad M: -\frac{T}{2} + x$

$R: \ 0$

after $\frac{\varepsilon T}{2}$ additional steps

follow the leader chooses R

leader's value $\sim \frac{T}{2}(1-\varepsilon) \cdot 2$