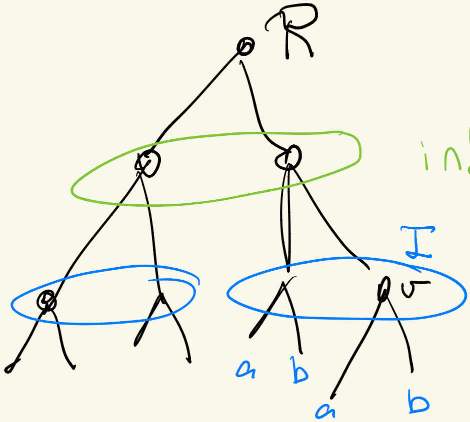
Extensive form



info sets player 1

player 2

perfect recall

player i strategy = prob distr for each information set

$x_v(a)$ = prob of choosing a at node v

& $x_v(a) = x_w(a)$ if $v, w \in I$

Run separate learning alg at each info set

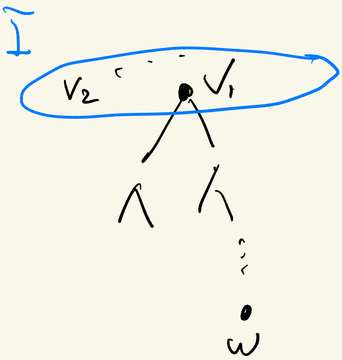
Assume today: full info on distribution used by all players

value $u_i(v, \sigma)$

v node in tree

σ distribution play used

$$u_i(v,\sigma) = \sum_{\omega} u_i(\omega)\, \pi^{\sigma}(v,\omega)$$



$\pi^{\sigma}(v,\omega)$ = prob reaching ω from v using σ

$u_i(\omega)$ = value for i at ω

$u_i(I,\sigma)$

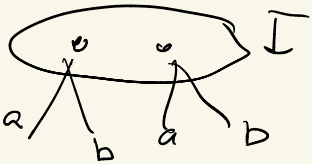
$\pi^{\sigma}_{-i}(v)$ = prob reaching v using σ for $-i$ if i makes decisions to reach I

$$u_i(I,\sigma) = \left(\sum_{v\in I} \pi^{\sigma}_{-i}(v)\, u_i(v,\sigma)\right) \Big/ \underbrace{\sum_{v\in I} \pi^{\sigma}_{-i}(v)}_{\pi^{\sigma}_{-i}(I)}$$

goal for learner at ~~node~~ info set I

$$\max \sum_{t} \pi^{\sigma^t}_{-i}(I)\, u_i(I,\sigma^t)$$

local regret at ~~we~~ info set I



local regret at I
regret for changing
$\sigma \to \sigma|_{I \to a}$

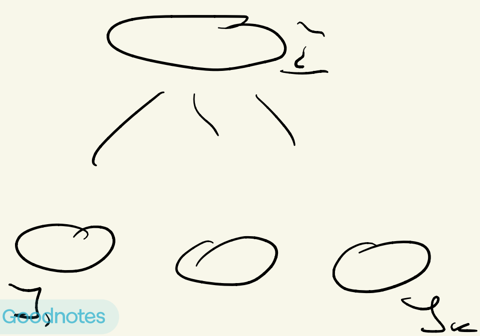
$$\mathrm{Reg}_{in}(I) = \pi_{-i}^{\sigma}(I)\left[u_i(I,\sigma) - u_i(I,\sigma|_{I\to a})\right]$$

Theorem: Reg for i a pure strategy

$$\le \sum_{I \text{ i's info set}} \max_a \mathrm{Reg}_{in}(I,a)$$

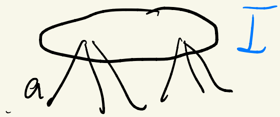
Reg(I) = regret for any pure strategy conditioned on reaching I

Claim $\mathrm{Reg}(I) \le \sum_{J \text{ info set for } i \text{ under } I} \max_a \mathrm{Reg}_{in}(J,a)$



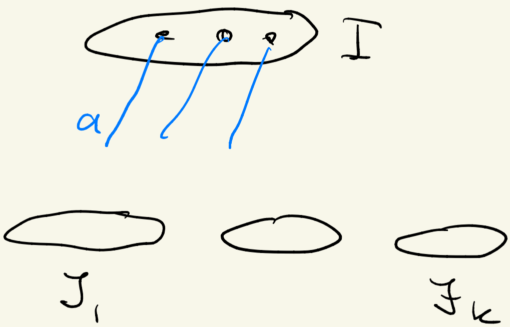
Reg(I) also uses weighted utility by $\pi_{-i}^{\sigma^t}(I)$

Proof: base case



$$Reg_{in} = Reg$$

induction step



$$u_i(I,\sigma^t) - u_i(I,\sigma^t|_{I\to a})$$

$$+ u_i(I,\sigma^t|_{I\to a}) - u_i(J, \sigma^t|_{\text{det after } I})$$

in more details (added after class)

$$u_i(I|\sigma^t|_{I\to a})\,\pi^\sigma_{-i}(I) = \sum_{w,v\in I} u_i(w)\,\pi^\sigma(v,w)$$

by definition

each $v \to w$ path goes through one of the nodes s in one of the info sets J_e

so we can write

$$= \sum_e \sum\sum_{v,w,s \in \mathcal{I}_e} \underbrace{u_i(w)} \cdot \pi^{\sigma_{I\to a}}(v,s) \cdot \underbrace{\pi^{\sigma}(s,w)}$$

sum for a given s in $u_i(s,\sigma)$

$$= \sum_e \sum_{v,s \in \mathcal{I}_e} \pi^{\sigma_{I\to a}}(v,s) \cdot u_i(s,\sigma)$$

$$= \sum_e u_i(\mathcal{I}_e,\sigma) \cdot \pi^{\sigma}_{-i}(\mathcal{I}_e,\sigma)$$

note choosing $I\to a$ is what it takes to get to the $\mathcal{I}_e$ info sets

To finish the proof, use the above for a deterministic strategy A with first decision at I as a, we get

$Reg_i(I,\sigma,A)$ regret for strategy A at I for distribution σ, utility weighted $u(I,\sigma)$

we get

$$Reg_i(I,\sigma,A) = u_i(I,\sigma)\pi^{\sigma}_{-i}(I) - u_i(I,\sigma|_{I\to a})\pi^{\sigma}_{-i}(I)$$

$$+ \pi_{-i}(I)\, u_i(I,\sigma|_{I\to a}) - \pi_{-i}(I)\, u_i(I,\sigma,A)$$

the second pair of terms by the above is

$$= \sum_{\ell} \boxed{\left(\Pi_{-i}(\mathcal{I}_e)\, u_i(\mathcal{I}_e, \sigma) - \Pi_{-i}(\mathcal{I}_e)\, u_i(\mathcal{I}_e, \sigma, A_e) \right)}$$

where A_e is the part of A starting from info set $\mathcal{I}_e$

Now summing over time t & distributions used σ^t

the green part is bounded by $Reg_{im}(I, \sigma)$

the blue parts are bounded by $Reg(\mathcal{I}_e, \sigma)$ which by the induction hypothesis

is $\leq \sum_{\substack{K \text{ info node for } i \\ \text{under } \mathcal{I}_e}} Reg_{im}(K)$

proving the claim.