

Game-Theoretic Learning:

Regret Minimization vs. Utility Maximization

Amy Greenwald

with David Gondek, Amir Jafari, and Casey Marks

Brown University

Workshop for Women in Machine Learning

October 4, 2006

Background

No-external-regret learning converges to the set of minimax equilibria in zero-sum games. [e.g., Freund and Schapire 1996]

No-internal-regret learning converges to the set of correlated equilibria in general-sum games. [e.g., Foster and Vohra 1997]

Foreground

1. Definitions

- A continuum of no-regret properties, called no- Φ -regret.
- A continuum of game-theoretic equilibria, called Φ -equilibria.

2. Existence Theorem

- Constructive proof: No- Φ -regret learning algorithms exist, $\forall \Phi$.

3. Convergence Theorem

- No- Φ -regret learning converges to the set of Φ -equilibria, $\forall \Phi$.

4. Surprising Result

- No-internal-regret is the strongest form of no- Φ -regret learning.
- Therefore, no no- Φ -regret algorithm learns Nash equilibria.

Outline

- Game Theory
- Single Agent Learning Model
- Multiagent Learning & Game-Theoretic Equilibria

Game Theory: A Crash Course

1. General-Sum Games

- Nash Equilibrium
- Correlated Equilibrium

2. Zero-Sum Games

- Minimax Equilibrium

Single Agent Learning Model

- set of actions $A = \{1, \dots, n\}$
- for all times t ,
 - play mixed strategy q^t
 - sample pure action $a^t \sim q^t$
 - earn reward vector $r^t \in [0, 1]^n$

A **learning algorithm** \mathcal{A} is a sequence of functions $q^t : \text{History}^{t-1} \rightarrow \Delta(A)$, where a **History** is a sequence of action-reward pairs $(a^1, r^1), (a^2, r^2), \dots$

Action Transformations

Definition

A set Φ of action transformations is comprised of functions $\phi : A \rightarrow \Delta(A)$.

Examples

$$\phi_{\text{EXT}}^{(2)} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \in \Phi_{\text{EXT}} \quad \phi_{\text{INT}}^{(23)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \in \Phi_{\text{INT}}$$

Regret Vector

Given Φ , a regret vector $\rho \in \mathbb{R}^\Phi$ with $\rho_\phi(r, a) = r \cdot a\phi - r \cdot a$.

Blackwell's Approachability Theorem

Definition

Given Φ , a **no- Φ -regret** learning algorithm is one whose time-averaged regret vector $\bar{\rho}^t$ “approaches” the negative orthant \mathbb{R}_-^Φ .

Theorem

The negative orthant \mathbb{R}_-^Φ is “approachable” if there exists a learning algorithm $\mathcal{A} = q^1, q^2, \dots$ s.t. for all times t , for any sequence of rewards r^1, r^2, \dots ,

$$\rho(r^{t+1}, q^{t+1}) \cdot g(\bar{\rho}^t) \leq 0 \quad (1)$$

where $g : \mathbb{R}^\Phi \rightarrow \mathbb{R}_+^\Phi$.

Moreover, this procedure can be used to approach the negative orthant \mathbb{R}_-^Φ :

- if $\bar{\rho}^t \in \mathbb{R}_-^\Phi$, play arbitrarily;
- if $\bar{\rho}^t \in \mathbb{R}^\Phi \setminus \mathbb{R}_-^\Phi$, play according to \mathcal{A} .

Regret Matching Algorithm

Given Φ

Given $Y \in \mathbb{R}_+^{\Phi}$

If $\sum_{\phi \in \Phi} Y_{\phi} = 0$, play arbitrarily

If $\sum_{\phi \in \Phi} Y_{\phi} > 0$, define stochastic matrix

$$A \equiv A(\Phi, Y) = \frac{\sum_{\phi \in \Phi} \phi Y_{\phi}}{\sum_{\phi \in \Phi} Y_{\phi}} \quad (2)$$

play mixed strategy $q = qA$

Regret Matching Theorem

Regret matching satisfies the generalized Blackwell condition:

$$\rho(r, q) \cdot Y = 0$$

Proof

$$\rho(r, q) \cdot Y = \sum_{\phi \in \Phi} \rho_{\phi}(r, q) Y_{\phi} \quad (3)$$

$$= \sum_{\phi \in \Phi} (r \cdot q\phi - r \cdot q) Y_{\phi} \quad (4)$$

$$= \sum_{\phi \in \Phi} r \cdot (q\phi Y_{\phi} - qY_{\phi}) \quad (5)$$

$$= r \cdot \left(q \sum_{\phi \in \Phi} \phi Y_{\phi} - q \sum_{\phi \in \Phi} Y_{\phi} \right) \quad (6)$$

$$= \left(\sum_{\phi \in \Phi} Y_{\phi} \right) r \cdot \left(q \frac{\sum_{\phi \in \Phi} \phi Y_{\phi}}{\sum_{\phi \in \Phi} Y_{\phi}} - q \right) \quad (7)$$

$$= \left(\sum_{\phi \in \Phi} Y_{\phi} \right) r \cdot (qA - q) \quad (8)$$

$$= \left(\sum_{\phi \in \Phi} Y_{\phi} \right) r \cdot (q - q) \quad (9)$$

$$= 0 \quad (10)$$

Generic No-Regret Learning Algorithm (Φ, g)

for $t = 1, 2, \dots$

1. play mixed strategy q^t
2. realize pure action $a^t \sim q^t$
3. observe rewards $r^t \in [0, 1]^n$
4. for all $\phi \in \Phi$
 - compute instantaneous regret $\rho_\phi^t = r^t \cdot a^t \phi - r^t \cdot a^t$
 - update cumulative regret vector $X_\phi^t = X_\phi^{t-1} + \rho_\phi^t$
5. compute $Y = g(X^t)$
6. compute $A = \frac{\sum_{\phi \in \Phi} \phi Y_\phi}{\sum_{\phi \in \Phi} Y_\phi}$
7. solve for a fixed point $q^{t+1} = q^{t+1} A$

Special Cases of Regret Matching

Foster and Vohra 97 (Φ_{INT})

Hart and Mas-Colell 00 (Φ_{EXT})

Choose $G(X) = \frac{1}{2} \sum_k (X_k^+)^2$ so that $g_k(X) = X_k^+$

Freund and Schapire 95 (Φ_{EXT})

Cesa-Bianchi and Lugosi 03 (Φ_{INT})

Choose $G(X) = \frac{1}{\eta} \ln \left(\sum_k e^{\eta X_k} \right)$ so that $g_k(X) = e^{\eta X_k} / \sum_k e^{\eta X_k}$

Multiagent Model

- a set of players I ($i \in I$)
- for all players i ,
 - a set of pure actions A_i
 - a set of mixed actions $Q_i = \Delta(A_i)$
 - a reward function $r_i : A \rightarrow [0, 1]$, where $A = \prod_i A_i$
 - an expected reward function $r_i : Q \rightarrow [0, 1]$, where $Q = \Delta(A)$
s.t. for all $q \in Q$, $r_i(q) = \sum_{a \in A} q(a)r_i(a)$
 - a set Φ_i

Φ -Equilibrium

Definition

A mixed action profile $q^* \in Q$ is a Φ -equilibrium iff $r_i(\dot{\phi}_i(q^*)) \leq r_i(q^*)$, for all players i and for all $\phi_i \in \Phi_i$.

Examples

Correlated Equilibrium: $\Phi_i = \Phi_{\text{INT}}$, for all players i

Generalized Minimax Equilibrium: $\Phi_i = \Phi_{\text{EXT}}$, for all players i

Convergence Theorem

If all players i play no- Φ_i -regret learning algorithms, then the joint empirical distribution of play converges to the set of Φ -equilibria, almost surely.

Proof Sketch

For all players i , for all $\phi_i \in \Phi_i$,

$$\limsup_{t \rightarrow \infty} r_i(\tilde{\phi}_i(z^t)) - r_i(z^t) \tag{11}$$

$$= \limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=1}^t r_i(\phi_i(a_i^\tau), a_{-i}^\tau) - \frac{1}{t} \sum_{\tau=1}^t r_i(a_i^\tau, a_{-i}^\tau) \tag{12}$$

$$= \limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=1}^t (r_i(\phi_i(a_i^\tau), a_{-i}^\tau) - r_i(a_i^\tau, a_{-i}^\tau)) \tag{13}$$

$$\leq 0 \tag{14}$$

almost surely.

Zero-Sum Games

Matching Pennies

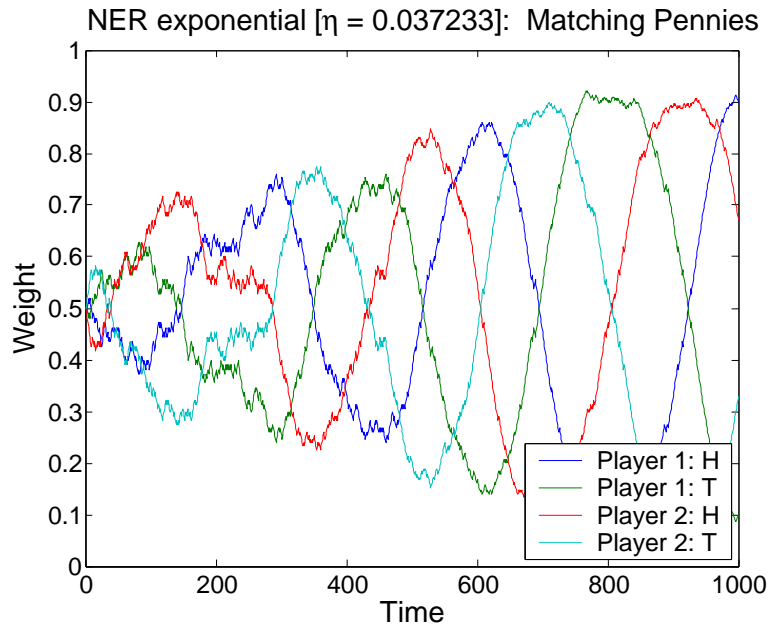
	H	T
H	$-1, 1$	$1, -1$
T	$1, -1$	$-1, 1$

Rock-Paper-Scissors

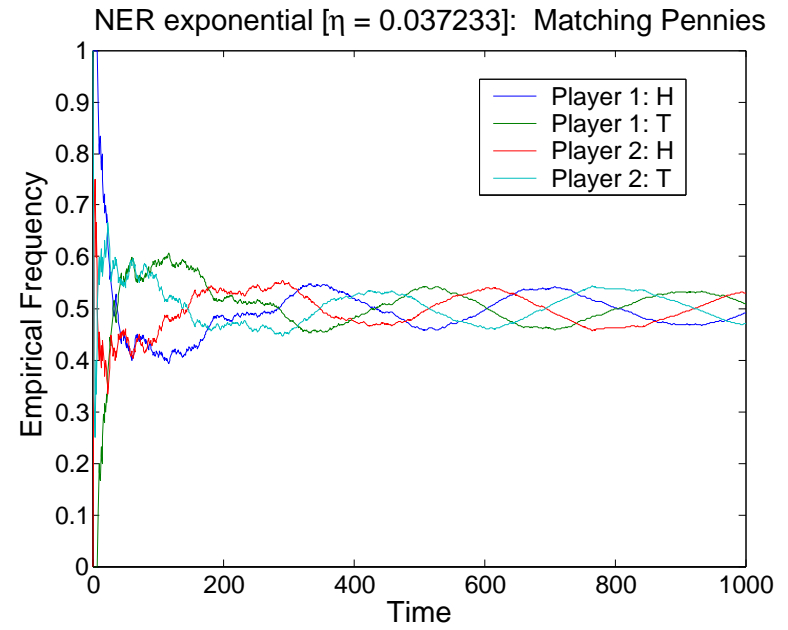
	R	P	S
R	$0, 0$	$-1, 1$	$1, -1$
P	$1, -1$	$0, 0$	$-1, 1$
S	$-1, 1$	$1, -1$	$0, 0$

Matching Pennies

Weights

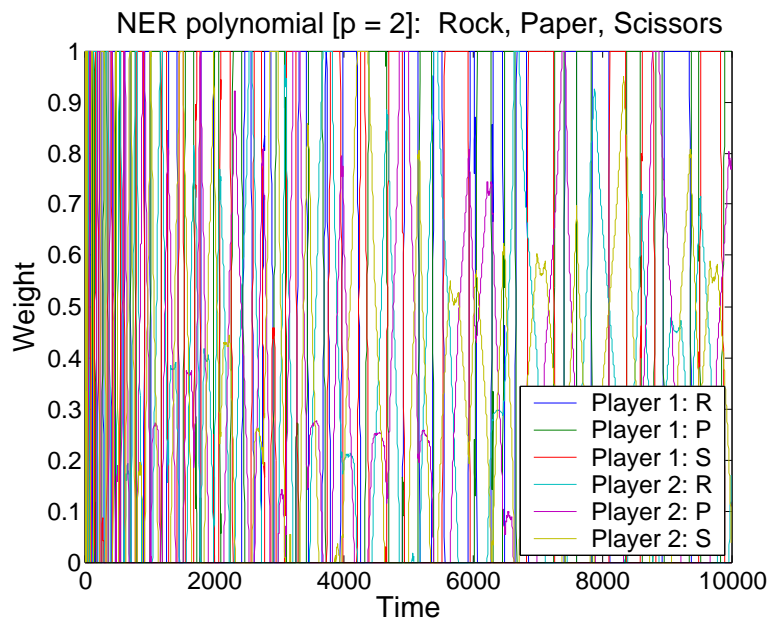


Frequencies

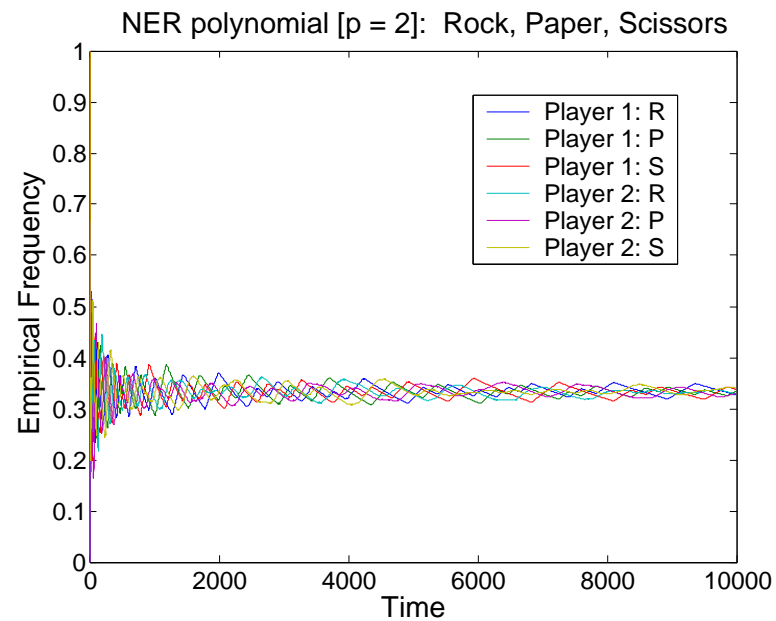


Rock-Paper-Scissors

Weights



Frequencies



General-Sum Games

Shapley Game

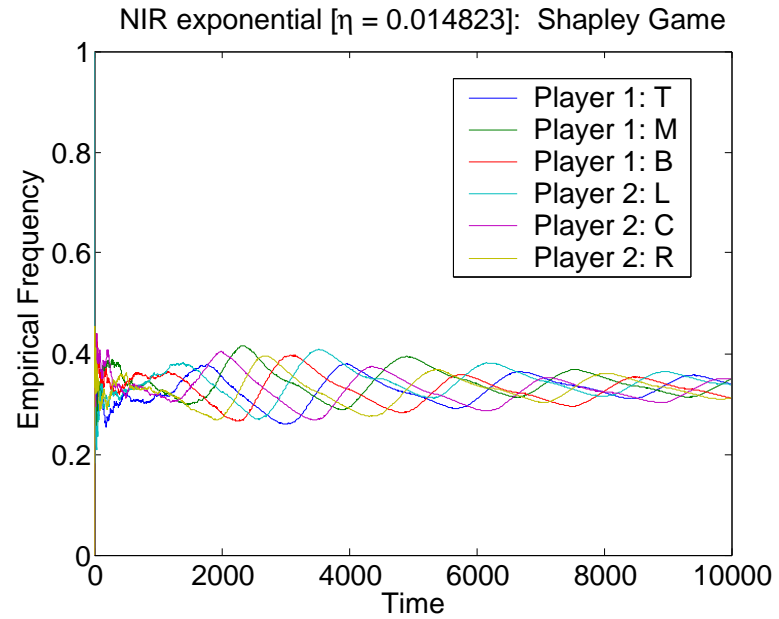
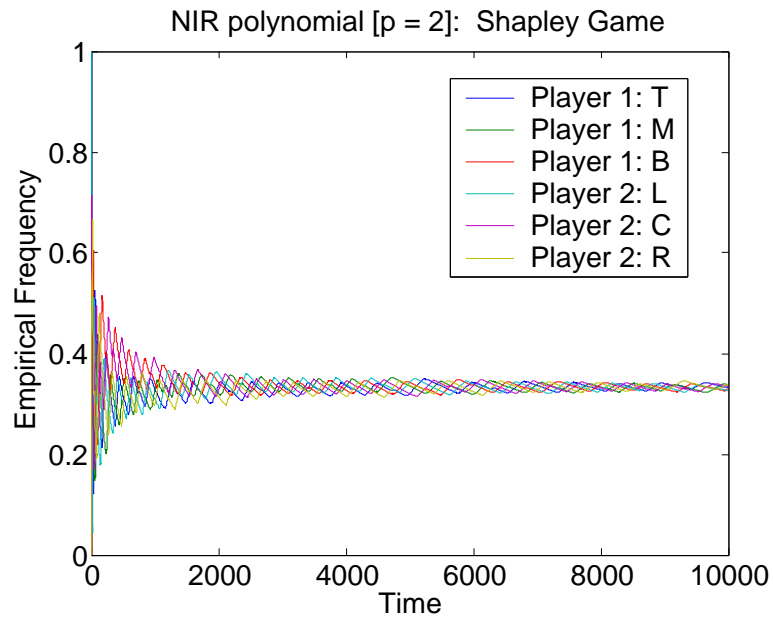
	L	C	R
T	0,0	1,0	0,1
M	0,1	0,0	1,0
B	1,0	0,1	0,0

Correlated Equilibrium

	L	C	R
T	0	1/6	1/6
M	1/6	0	1/6
B	1/6	1/6	0

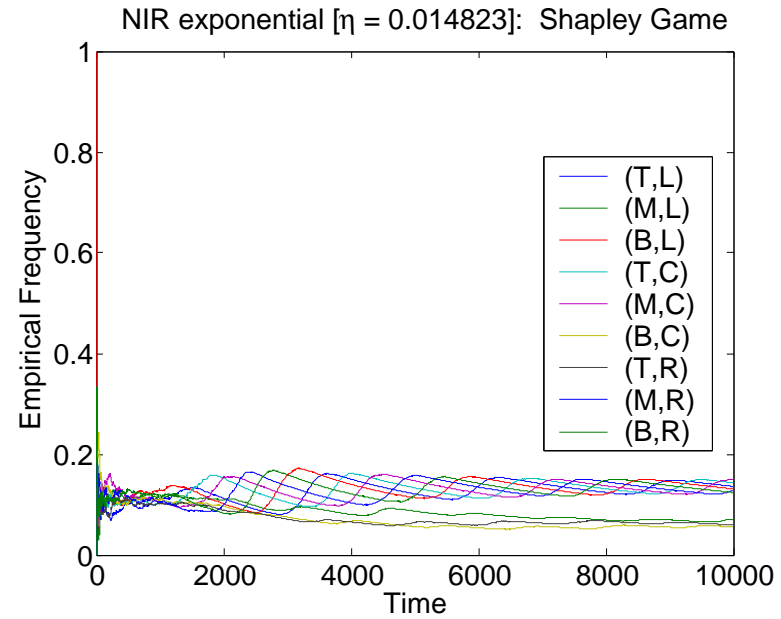
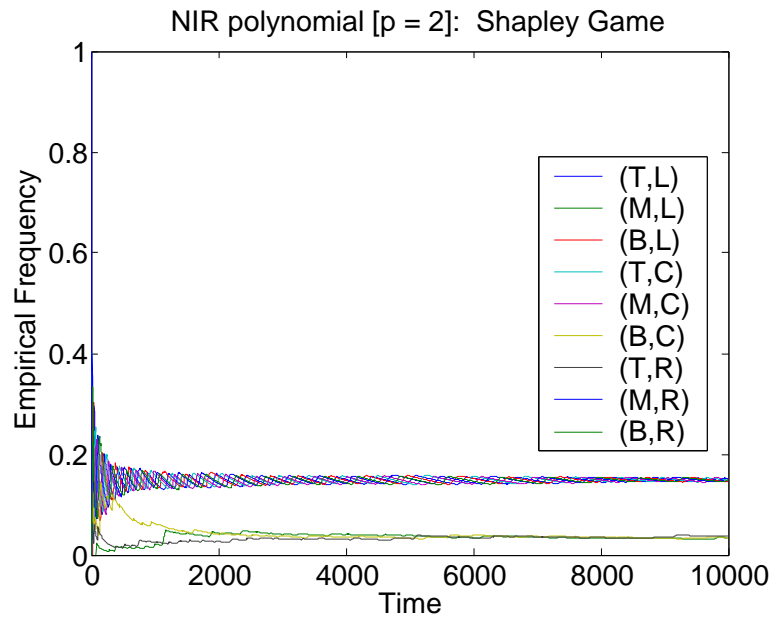
Shapley Game: No Internal Regret Learning

Frequencies



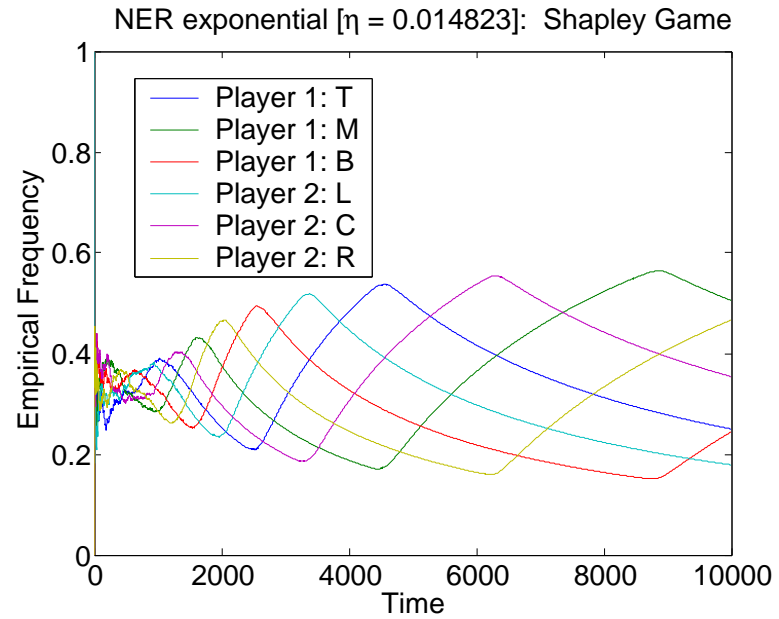
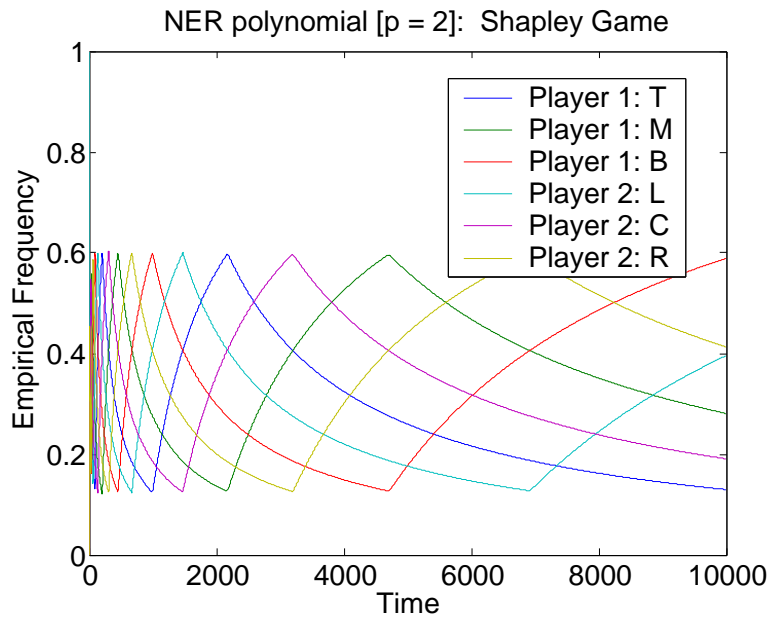
Shapley Game: No Internal Regret Learning

Joint Frequencies



Shapley Game: No External Regret Learning

Frequencies



Summary

- No-external- and no-internal-regret can be defined along one continuum, no- Φ -regret.
- No- Φ -regret learning algorithms exist, $\forall \Phi$.
- No- Φ -regret learning converges to the set of Φ -equilibria, $\forall \Phi$.
- No-internal-regret learning is the strongest form of no- Φ -regret learning. Therefore, Nash equilibrium cannot be learned via no- Φ -regret learning.

“A little rationality goes a long way” [Hart 03]

Regret Minimization vs. Utility Maximization

- RM is easy to implement.
- RM justifies randomness in actions.
- Can RM be used to explain human behavior?