CPS 590.4
Learning in games

Vincent Conitzer
conitzer@cs.duke.edu
“2/3 of the average” game

• Everyone writes down a number between 0 and 100
• Person closest to 2/3 of the average wins
• Example:
  – A says 50
  – B says 10
  – C says 90
  – Average(50, 10, 90) = 50
  – 2/3 of average = 33.33
  – A is closest (|50-33.33| = 16.67), so A wins
“2/3 of the average” game revisited

\[
\begin{align*}
100 & \quad \text{dominated} \\
(2/3) \times 100 & \quad \text{dominated after removal of} \\
(2/3) \times (2/3) \times 100 & \quad (originally) \text{dominated strategies} \\
\vdots & \\
0 & \\
\end{align*}
\]
Learning in (normal-form) games

• Approach we have taken so far when playing a game: just compute an optimal/equilibrium strategy

• Another approach: learn how to play a game by
  – playing it many times, and
  – updating your strategy based on experience

• Why?
  – Some of the game’s utilities (especially the other players’) may be unknown to you
  – The other players may not be playing an equilibrium strategy
  – Computing an optimal strategy can be hard
  – Learning is what humans typically do
  – …

• Learning strategies ~ strategies for the repeated game
• Does learning converge to equilibrium?
Iterated best response

• In the first round, play something arbitrary
• In each following round, play a best response against what the other players played in the previous round
• If all players play this, it can converge (i.e., we reach an equilibrium) or cycle

\[
\begin{array}{ccc}
0, 0 & -1, 1 & 1, -1 \\
1, -1 & 0, 0 & -1, 1 \\
-1, 1 & 1, -1 & 0, 0 \\
\end{array}
\]

\[
\begin{array}{cc}
-1, -1 & 0, 0 \\
0, 0 & -1, -1 \\
\end{array}
\]

*rock-paper-scissors*

• Alternating best response: players alternatingly change strategies: one player best-responds each odd round, the other best-responds each even round
Fictitious play [Brown 1951]

- In the first round, play something arbitrary
- In each following round, play a best response against the empirical distribution of the other players’ play
  - I.e., as if other player randomly selects from his past actions
- Again, if this converges, we have a Nash equilibrium
- Can still fail to converge…

\[
\begin{array}{ccc}
0, 0 & -1, 1 & 1, -1 \\
1, -1 & 0, 0 & -1, 1 \\
-1, 1 & 1, -1 & 0, 0 \\
\end{array}
\]

\[
\begin{array}{cc}
-1, -1 & 0, 0 \\
0, 0 & -1, -1 \\
\end{array}
\]

*rock-paper-scissors*

* a simple congestion game
Fictitious play on rock-paper-scissors

<table>
<thead>
<tr>
<th></th>
<th>Row</th>
<th>Column</th>
</tr>
</thead>
<tbody>
<tr>
<td>0, 0</td>
<td>-1, 1</td>
<td>1, -1</td>
</tr>
<tr>
<td>1, -1</td>
<td>0, 0</td>
<td>-1, 1</td>
</tr>
<tr>
<td>-1, 1</td>
<td>1, -1</td>
<td>0, 0</td>
</tr>
</tbody>
</table>

30% R, 50% P, 20% S
30% R, 20% P, 50% S
Does the empirical distribution of play converge to equilibrium?

- … for iterated best response?
- … for fictitious play?

<table>
<thead>
<tr>
<th></th>
<th>3, 0</th>
<th>1, 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2, 1</td>
<td></td>
</tr>
</tbody>
</table>
Fictitious play is guaranteed to converge in...

- Two-player zero-sum games [Robinson 1951]
- Generic 2x2 games [Miyasawa 1961]
- Games solvable by iterated strict dominance [Nachbar 1990]
- Weighted potential games [Monderer & Shapley 1996]
- **Not** in general [Shapley 1964]
- But, fictitious play always converges to the set of $\frac{1}{2}$-approximate equilibria [Conitzer 2009; more detailed analysis by Goldberg, Savani, Sørensen, Ventre 2011]
Shapley’s game on which fictitious play does not converge

- starting with \((U, M)\):

<table>
<thead>
<tr>
<th></th>
<th>0, 0</th>
<th>0, 1</th>
<th>1, 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0, 0</td>
<td>1, 0</td>
<td>0, 0</td>
<td>0, 1</td>
</tr>
<tr>
<td>0, 1</td>
<td>1, 0</td>
<td>0, 0</td>
<td>0, 0</td>
</tr>
</tbody>
</table>
Regret

• For each player $i$, action $a_i$ and time $t$, define the regret $r_i(a_i, t)$ as
  \[
  (\Sigma_{1 \leq t' \leq t-1} u_i(a_i, a_{-i,t'}) - u_i(a_i,t', a_{-i,t'}))/(t-1)
  \]
• An algorithm has zero regret if for each $a_i$, the regret for $a_i$ becomes nonpositive as $t$ goes to infinity (almost surely) against any opponents
• Regret matching [Hart & Mas-Colell 00]: at time $t$, play an action that has positive regret $r_i(a_i, t)$ with probability proportional to $r_i(a_i, t)$
  – If none of the actions have positive regret, play uniformly at random
• Regret matching has zero regret
• If all players use regret matching, then play converges to the set of weak correlated equilibria
  – Weak correlated equilibrium: playing according to joint distribution is at least as good as any strategy that does not depend on the signal
• Variants of this converge to the set of correlated equilibria
• Smooth fictitious play [Fudenberg & Levine 95] also gives no regret
  – Instead of just best-responding to history, assign some small value to having a more “mixed” distribution
Targeted learning

- Assume that there is a limited set of possible opponents
- Try to do well against these
- Example: is there a learning algorithm that
  - learns to best-respond against any stationary opponent (one that always plays the same mixed strategy), and
  - converges to a Nash equilibrium (in actual strategies, not historical distribution) when playing against a copy of itself (so-called self-play)?
- [Bowling and Veloso AlJ02]: yes, if it is a 2-player 2x2 game and mixed strategies are observable
- [Conitzer and Sandholm ML06]: yes (without those assumptions)
  - AWESOME algorithm (Adapt When Everybody is Stationary, Otherwise Move to Equilibrium): (very) rough sketch:

```
play according to equilibrium strategy

not all players appear to be playing equilibrium

best-respond to recent history

not all players appear to be playing stationary strategies
```
“Teaching”

• Suppose you are playing against a player that uses one of these strategies
  – Fictitious play, anything with no regret, AWESOME, …
• Also suppose you are very patient, i.e., you only care about what happens in the long run
• How will you (the row player) play in the following repeated games?
  – Hint: the other player will eventually best-respond to whatever you do

<table>
<thead>
<tr>
<th></th>
<th>4, 4</th>
<th>3, 5</th>
<th>1, 0</th>
<th>3, 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>5, 3</td>
<td>0, 0</td>
<td>2, 1</td>
<td>4, 0</td>
<td></td>
</tr>
</tbody>
</table>

• Note relationship to optimal strategies to commit to
• There is some work on learning strategies that are in equilibrium with each other [Brafman & Tennenholtz AIJ04]
Evolutionary game theory

• Given: a symmetric game

\[
\begin{array}{c|cc}
  & \text{dove} & \text{hawk} \\
\hline
\text{dove} & 1, 1 & 0, 2 \\
\text{hawk} & 2, 0 & -1, -1 \\
\end{array}
\]

Nash equilibria: (d, h), (h, d), ((.5, .5), (.5, .5))

• A large population of players plays this game, players are randomly matched to play with each other

• Each player plays a pure strategy
  – Fraction of players playing strategy \(s = p_s\)
  – \(p\) is vector of all fractions \(p_s\) (the state)

• Utility for playing \(s\) is \(u(s, p) = \sum_s p_s \cdot u(s, s')\)

• Players reproduce at a rate that is proportional to their utility, their offspring play the same strategy
  – Replicator dynamic

\[
\frac{dp_s(t)}{dt} = p_s(t)(u(s, p(t)) - \sum_{s'} p_s \cdot u(s', p(t)))
\]

• What are the steady states of this?
### Stability

<table>
<thead>
<tr>
<th></th>
<th>dove</th>
<th>hawk</th>
</tr>
</thead>
<tbody>
<tr>
<td>dove</td>
<td>1, 1</td>
<td>0, 2</td>
</tr>
<tr>
<td>hawk</td>
<td>2, 0</td>
<td>-1, -1</td>
</tr>
</tbody>
</table>

- A steady state is **stable** if slightly perturbing the state will not cause us to move far away from the state.
- E.g. everyone playing dove is not stable, because if a few hawks are added their percentage will grow.
- What about the mixed steady state?
- Proposition: every stable steady state is a Nash equilibrium of the symmetric game.
- Slightly stronger criterion: a state is **asymptotically stable** if it is stable, and after slightly perturbing this state, we will (in the limit) return to this state.
Evolutionarily stable strategies

- Now suppose players play mixed strategies
- A (single) mixed strategy $\sigma$ is evolutionarily stable if the following is true:
  - Suppose all players play $\sigma$
  - Then, whenever a very small number of invaders enters that play a different strategy $\sigma'$,
  - the players playing $\sigma$ must get strictly higher utility than those playing $\sigma'$ (i.e., $\sigma$ must be able to repel invaders)
- $\sigma$ will be evolutionarily stable if and only if for all $\sigma'$
  - $u(\sigma, \sigma) > u(\sigma', \sigma)$, or:
  - $u(\sigma, \sigma) = u(\sigma', \sigma)$ and $u(\sigma, \sigma') > u(\sigma', \sigma')$
- Proposition: every evolutionarily stable strategy is asymptotically stable under the replicator dynamic