Computing Approximate Nash Equilibria and Robust Best-Responses Using Sampling

Marc Ponsen    Steven de Jong    Marc Lanctot

Paper Review by Oron Anschel

July 15, 2015
Outline

1. Introduction
   - Games
   - Best Response & Nash-Equilibrium

2. Computing Approximate Nash-Equilibrium
   - Non-Sampling methods
   - Sampling methods

3. Monte-Carlo Restricted Nash Response
   - Restricted Nash Response
   - Monte Carlo Restricted Nash Response

4. Results
   - Experiments
   - Contributions

5. Simulation
   - Game Setup
   - Results
Outline

1. Introduction
   • Games
     • Best Response & Nash-Equilibrium
2. Computing Approximate Nash-Equilibrium
   • Non-Sampling methods
   • Sampling methods
3. Monte-Carlo Restricted Nash Response
   • Restricted Nash Response
   • Monte Carlo Restricted Nash Response
4. Results
   • Experiments
   • Contributions
5. Simulation
   • Game Setup
   • Results
Games

Games Examples:
- Puzzles
- Rock-paper-scissors
- Backgammon
- Chess
- Poker
- Video games
Normal Game

- Players act simultaneously
- Represented in a **Game-Table**
- Example: Rock-paper-scissors

**Rock-paper-scissors - Table representation**

<table>
<thead>
<tr>
<th>P1\P2</th>
<th>Rock</th>
<th>Paper</th>
<th>Scissor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rock</td>
<td>[0,0]</td>
<td>[0,1]</td>
<td>[1,0]</td>
</tr>
<tr>
<td>Paper</td>
<td>[1,0]</td>
<td>[0,0]</td>
<td>[0,1]</td>
</tr>
<tr>
<td>Scissor</td>
<td>[0,1]</td>
<td>[1,0]</td>
<td>[0,0]</td>
</tr>
</tbody>
</table>
Extensive-Form Game

- Represented as a **Game-Tree**
- Examples: Chess, Backgammon, Poker, Tic Tac Toe
- Characteristics:
  - Sequential decision-making
  - Imperfect information
  - Stochastic

[Diagram of a Game-Tree]
Outline

1 Introduction
   - Games
   - Best Response & Nash-Equilibrium
2 Computing Approximate Nash-Equilibrium
   - Non-Sampling methods
   - Sampling methods
3 Monte-Carlo Restricted Nash Response
   - Restricted Nash Response
   - Monte Carlo Restricted Nash Response
4 Results
   - Experiments
   - Contributions
5 Simulation
   - Game Setup
   - Results

Paper Review by Oron Anschel
NE and RNR Using Monte Carlo Sampling
Best Response Strategy

Assume 2 players game

- $\sigma_i$: Player $i$ strategy
- $u_i$: Player $i$ game utility

Best Response Value

$$ b_1(\sigma_2) = \max_{\sigma'_1 \in \Sigma_1} u_1(\sigma'_1, \sigma_2) $$

Best Response Strategy

$$ \sigma_1 = \arg \max_{\sigma'_1 \in \Sigma_1} u_1(\sigma'_1, \sigma_2) $$
**Nash-Equilibrium Strategy**

- $\sigma_i$: Player $i$ Nash-Equilibrium strategy
- $u_i$: Player $i$ game utility

**Nash-Equilibrium**

\[
\begin{align*}
u_1(\sigma_1, \sigma_2) & \geq \max_{\sigma'_1 \in \Sigma_1} u_1(\sigma'_1, \sigma_2) \\
u_2(\sigma_2, \sigma_1) & \geq \max_{\sigma'_2 \in \Sigma_2} u_2(\sigma'_2, \sigma_1)
\end{align*}
\]

**Approximate Nash-Equilibrium**

\[
\begin{align*}
u_1(\sigma_1, \sigma_2) + \varepsilon & \geq \max_{\sigma'_1 \in \Sigma_1} u_1(\sigma'_1, \sigma_2) \\
u_2(\sigma_2, \sigma_1) + \varepsilon & \geq \max_{\sigma'_2 \in \Sigma_2} u_2(\sigma'_2, \sigma_1)
\end{align*}
\]

*How to compute a Nash-Equilibrium strategy?*
Introduction

Computing Approximate Nash-Equilibrium

MCRNR

Results

Simulation

Outline

1 Introduction
   - Games
   - Best Response & Nash-Equilibrium

2 Computing Approximate Nash-Equilibrium
   - Non-Sampling methods
   - Sampling methods

3 Monte-Carlo Restricted Nash Response
   - Restricted Nash Response
   - Monte Carlo Restricted Nash Response

4 Results
   - Experiments
   - Contributions

5 Simulation
   - Game Setup
   - Results

Paper Review by Oron Anschel

NE and RNR Using Monte Carlo Sampling
Non-Sampling methods

- Linear programming - applied to Poker (Billings et al. 2003)
- Excessive Gap Technique - applied to Poker (Hoda et al. 2010, Sandholm 2010)
Introduction

Games
Best Response & Nash-Equilibrium

Computing Approximate Nash-Equilibrium
Non-Sampling methods
Sampling methods

Monte-Carlo Restricted Nash Response
Restricted Nash Response
Monte Carlo Restricted Nash Response

Results
Experiments
Contributions

Simulation
Game Setup
Results
Monte Carlo Tree Search (MCTS) - Based on the UCB algorithm (B. Brügmann 1992, R. Coulom 2006, L. Kocsis and Cs. Szepesvári, S. Gelly 2008).

Monte Carlo Counterfactual Regret Minimization (MCCFR) - Based on the Regret Matching algorithm (Martin Zinkevich 2007, Marc Lanctot 2009)
Monte Carlo Tree Search (MCTS)
Monte Carlo Tree Search (MCTS)

- Convergence guarantees for **perfect information** games.

Repeat:

1. **Selection:**

   \[
   a^* \in \arg\max_{a \in A} \left( v_a + C \cdot \sqrt{\frac{\ln n_p}{n_a}} \right)
   \]

   - \(v_a\) — average simulated reward
   - \(n_a\) — visit count of action \(a\)
   - \(n_p\) — visit counts of current node

   (UCB1 algorithm)

2. **Expansion**

3. **Simulation**

4. **Backpropogation**
Monte Carlo Tree Search Cont’d

- Selection
- Expansion
- Simulation
- Backpropagation

Tree Policy

Default Policy
Monte Carlo Counterfactual Regret Minimization (MCCFR)

MCCFR
Some general results...

Average overall regret:

\[ R_i^T = \frac{1}{T} \max_{\sigma'_i \in \Sigma_i} \sum_{t=1}^{T} \left( u_i(\sigma'_i, \sigma_{-i}^t) - u_i(\sigma^t) \right) \]

Average strategy:

\[ \bar{\sigma}_i^T (a|l) = \frac{\sum_{t=1}^{T} \pi_i^{\sigma^t} (l) \sigma^t (a|l)}{\sum_{t=1}^{T} \pi_i^{\sigma^t} (l)} \]

**Theorem**

*In a zero sum game, if \( R_i^T \leq \varepsilon \) then \( \bar{\sigma}_i^T \) is a 2\( \varepsilon \) Nash-Equilibrium strategy.*
Monte Carlo Counterfactual Regret Minimization (MCCFR)

More results...

Counterfactual value:

\[ v_i(\sigma, l) = \sum_{z \in Z_I} \pi_{-i}^\sigma(z[l]) \pi^\sigma(z[l], z) u_i(z) \]

*\( Z_I \) - terminal nodes reachable from \( l \), \( z[l] \) - prefix of \( z \) in \( l \)

Intimidate Counterfactual regret:

\[
R_{i,imm}(a, l) = \frac{1}{T} \sum_{t=1}^{T} \left( v_i\left( \sigma_t^{l \rightarrow a}, l \right) - v_i(\sigma^t, l) \right)
\]

\[
R_{i,imm}(l) = \max_{a \in A(l)} R_{i,imm}(a, l)
\]

Let \( x^+ = \max(x, 0) \)

**Theorem**

\[
R_i^T \leq \sum_l R_{i,imm}^{T,^+}(l)
\]

* Using Regret Matching \( R_{i,imm}^{T,^+}(l) \) can be driven to zero!
Monte Carlo Counterfactual Regret Minimization (MCCFR)

Regret Matching:

\[
\sigma_i^t(a|l) = \frac{R_{i,imm}^{T,+}(l,a)}{\sum_a R_{i,imm}^{T,+}(l,a)}
\]

- \( R_{i,imm}^{T,+}(l,a) \) can be calculated recursively during the tree traversal.
- Can we avoid making full tree traversal?
Monte Carlo Counterfactual Regret Minimization (MCCFR)

Yes!

- MCCFR - Outcome-Sampling.
- Let $\pi^{\sigma'}(z)$ be the probability of sampling $z$.

Sampled Counterfactual value:

$$\tilde{v}_i(\sigma, I) = \frac{1}{\pi^{\sigma'}(z)} \pi^{\sigma}_{-i}(z[I]) \pi^{\sigma}(z[I], z) u_i(z)$$

- We have that $E[\tilde{v}_i(\sigma, I)] = v_i(\sigma, I)$.
- Sampling based algorithm that convergence to NE.
Outline

1. Introduction
   - Games
   - Best Response & Nash-Equilibrium

2. Computing Approximate Nash-Equilibrium
   - Non-Sampling methods
   - Sampling methods

3. Monte-Carlo Restricted Nash Response
   - Restricted Nash Response
   - Monte Carlo Restricted Nash Response

4. Results
   - Experiments
   - Contributions

5. Simulation
   - Game Setup
   - Results
Restricted Nash Response

- What if the opponent doesn’t play NES?
- What is the problem in playing best response?
- Can we exploit while being robust?
- RNR (Johanson et al. 2008)
What is RNR?

- Robust best response strategy.
- Assume the opponent plays $\sigma_{fix}$ with probability $p$.
- Solve a NE for a modified game where the opponent plays $p\sigma_{fix} + (1 - p)\sigma_2$. 

Opponent plays fixed strategy

Unrestricted game

Modified game
Outline

1. Introduction
   - Games
   - Best Response & Nash-Equilibrium
2. Computing Approximate Nash-Equilibrium
   - Non-Sampling methods
   - Sampling methods
3. Monte-Carlo Restricted Nash Response
   - Restricted Nash Response
   - Monte Carlo Restricted Nash Response
4. Results
   - Experiments
   - Contributions
5. Simulation
   - Game Setup
   - Results

Paper Review by Oron Anschel
NE and RNR Using Monte Carlo Sampling
Monte Carlo Restricted Nash Response

MCRNR Algorithm:

- Evaluate $\sigma_{fix}$ for the players offline.
- Confidence parameter $p$ can be evaluated for each node/globally.
- Run MCCFR, use a modified tree as input (do not update fixed strategies nodes).
Experiments Results

- MCCFR vs MCTS in Kuhn Poker

![Graph showing comparison between MCCFR and MCTS in Kuhn Poker](image-url)
Experiments Results Cont’d

- MCCFR vs MCTS in Poker

![Graph showing performance comparison between MCCFR and MCTS in Poker](image_url)
Experiments Results Cont’d

- Playing against SparBot and POKI (benchmark machine players).
- Each 1000 online games, 5 million MCCFR/MCRNR offline iterations.
- Results obtained after 10,000 online games.

<table>
<thead>
<tr>
<th>Opponent</th>
<th>MCCFR10</th>
<th>MCRNR10</th>
<th>MCCFR100</th>
<th>MCRNR100</th>
</tr>
</thead>
<tbody>
<tr>
<td>POKI</td>
<td>0.059</td>
<td>0.369</td>
<td>0.191</td>
<td>0.482</td>
</tr>
<tr>
<td>SPARBOT</td>
<td>-0.091</td>
<td>-0.039</td>
<td>0.046</td>
<td>0.061</td>
</tr>
</tbody>
</table>
Outline

1 Introduction
   • Games
   • Best Response & Nash-Equilibrium
2 Computing Approximate Nash-Equilibrium
   • Non-Sampling methods
   • Sampling methods
3 Monte-Carlo Restricted Nash Response
   • Restricted Nash Response
   • Monte Carlo Restricted Nash Response
4 Results
   • Experiments
   • Contributions
5 Simulation
   • Game Setup
   • Results

Paper Review by Oron Anschel
NE and RNR Using Monte Carlo Sampling
Contributions

- Comparison between MCTS and MCCFR on two-player Limit Texas Hold’Em Poker.
- Introduced MCRNR algorithm for robust best response strategies.
Outline

1. Introduction
   - Games
   - Best Response & Nash-Equilibrium
2. Computing Approximate Nash-Equilibrium
   - Non-Sampling methods
   - Sampling methods
3. Monte-Carlo Restricted Nash Response
   - Restricted Nash Response
   - Monte Carlo Restricted Nash Response
4. Results
   - Experiments
   - Contributions
5. Simulation
   - Game Setup
   - Results
Penalty Kick Game:
- 2 players and a ball
Penalty Kick Game:

- Player 1: Choose start position
- Player 2: Choose shot direction
- Player 1: Move left/right/don’t move
- Result: Goal/ no goal
Outline

1 Introduction
   - Games
   - Best Response & Nash-Equilibrium

2 Computing Approximate Nash-Equilibrium
   - Non-Sampling methods
   - Sampling methods

3 Monte-Carlo Restricted Nash Response
   - Restricted Nash Response
   - Monte Carlo Restricted Nash Response

4 Results
   - Experiments
   - Contributions

5 Simulation
   - Game Setup
   - Results
Nash-Equilibrium Strategy:
- Player 1: Start at the center
- Player 2: Choose shot direction (doesn’t matter)
- Player 1: Move to shooting direction
- Result: Player 1 always stops the ball