Empirical-evidence Equilibria in Stochastic Games

Nicolas Dudebout
Multiagent problems

- stock market
- group of robots
Context

Multiagent problems

- stock market
- group of robots

Game-theoretic approach

- selfish agents
- different solution concepts

- predictive
- prescriptive
Context

Multiagent problems
- stock market
- group of robots

Game-theoretic approach
- selfish agents
- different solution concepts

- predictive
- prescriptive
Empirical-evidence Equilibrium (EEE)

Motivation
Definition
Existence
Comparison
Characterization
Predictive Use
Prescriptive Use
Graphical convention

\[ x \rightarrow f \rightarrow y \]

\[ y^+ \sim f(x) \]
Graphical convention

\[ x \rightarrow f \rightarrow y \]

\[ y^+ \sim f(x) \]

\[ x \rightarrow g \rightarrow y \]

\[ y^{t+1} \sim g(x^1, \ldots, x^t, y^1, \ldots, y^t) \]
Graphical convention

\[ x \xrightarrow{f} y \]
\[ y^{+} \sim f(x) \]

\[ x \xrightarrow{g} y \]
\[ y^{t+1} \sim g(x^{1}, \ldots, x^{t}, y^{1}, \ldots, y^{t}) \]

\[ x \xrightarrow{h} y \]
\[ y^{+} \sim h(y, x) \]
Maximal value function of a Markov Decision Process (MDP) is given by:

$$\max_{\sigma} E_{\sigma} \left[ \sum_{t=0}^{\infty} \delta^{t} \cdot u(x^{t}, a^{t}) \right]$$

where

- $\sigma$ is the policy
- $x^{+} \sim f(x, a)$ is the next state distribution
- $\sum_{t=0}^{\infty} \delta^{t}$ is the discount factor at time $t$
Markov Decision Process (MDP)

\[
\max_{\sigma} \mathbb{E}_\sigma \left[ \sum_{t=0}^{\infty} \delta^t \cdot u(x^t, a^t) \right]
\]

\[
x^+ \sim f(x, a)
\]
Markov Decision Process (MDP)

\[ \max_{\sigma} \mathbb{E}_{\sigma} \left[ \sum_{t=0}^{\infty} \delta^t \cdot u(x^t, a^t) \right] \]

\[ u(x, a), \delta \]

\[ x^+ \sim f(x, a) \]
Markov Decision Process (MDP)

\[ \max_{\sigma} \mathbb{E}_{\sigma} \left[ \sum_{t=0}^{\infty} \delta^t \cdot u(x^t, a^t) \right] \]

\[ u(x, a), \delta \]
Stochastic Game

\[ u_1(x_1, a_1), \delta_1 \]

\[ u_2(x_2, a_2), \delta_2 \]
Stochastic Game

\[ u_1(x_1, a_1, a_2), \delta_1 \]

\[ u_2(x_2, a_2, a_1), \delta_2 \]
Stochastic Game

\[ u_1(x_1, a_1, a_2), \delta_1 \]
Partially Observable Markov Decision Process (POMDP)

\[ u(x, a), \delta \]
Partially Observable Markov Decision Process (POMDP)

\[ u(x, a), \delta \]
\[ u_1(x_1, a_1, a_2), \delta_1 \]

\[ u(x, a), \delta \]
Recap
• Multiagent problems
• Game-theoretic approach
• Nash equilibrium in stochastic game $\iff$ unknown POMDPs
• Multiagent problems
• Game-theoretic approach
• Nash equilibrium in stochastic game $\iff$ unknown POMDPs

POMDP intractable
MDP solved
Stochastic Game

\[ u_1(x_1, a_1, a_2), \delta_1 \]
Stochastic Game

\[ u_1(x_1, a_1, s_1^+) , \delta_1 \]
Stochastic Game

\[ u_1(x_1, a_1, s_1^+) , \delta_1 \]
\( u(x, a, s^+), \delta \)
\[ u(x, a, s^+), \delta \]

\[
\begin{cases}
    w^+ \sim n(w, x, a) \\
    s^+ \sim \nu(w^+) 
\end{cases}
\]
\[ u(x, a, s^+), \delta \]
\( u(x, a, s^+), \delta \)
Simple Consistency

01000100010010100101101101010...
Simple Consistency

0100010001001010010110111010...

$P[0], P[1]$
Simple Consistency

\[010001000100101001011011010\ldots\]

\[\mathbb{P}[0], \mathbb{P}[1]\]

\[\mu(s) = \mathbb{P}[s]\]
Simple Consistency

\[ P[0], P[1] \]

\[ \mu[s] = P[s] \]

\[ u(x, a, s^+), \delta \]
Simple Consistency

\[ u(x, a, s^+), \delta \]

\[ \mu[s] = P[s] \]

01000100010010100101101101010...

\[ P[0], P[1] \]
Simple Consistency

\[ 010001000100101001011011101010... \]

\[ \mathbb{P}[0], \mathbb{P}[1] \]

\[ \mu[s] = \mathbb{P}[s] \]

\[ u(x, a, s^+), \delta \]
Simple Consistency

\[ P[0], P[1] \]

\[ \mu[s] = P[s] \]

\[ u(x, a, s^+), \delta \]
Two Systems

\[ u(x, a, s^+), \delta \]

Real System: \( R \)
Two Systems

\[ u(x, a, s^+), \delta \]

Real System: \( \mathbf{R} \)

Nature

Mockup System: \( \mathbf{M} \)
Two Systems

Real System: \( R \)

Mockup System: \( M \)

\[ u(x, a, s^+), \delta \]
Two Systems

Real System: $\mathbf{R}$

Mockup System: $\mathbf{M}$

$u(x, a, s^+), \delta$

$\sigma$ $\leftarrow$ $x$ $a$ $f$

$\sigma$ $\leftarrow$ $x$ $a$ $f$

Nature

$\sigma$ $\leftarrow$ $s$

$\delta$

$\mu$ $\leftarrow$ $s$
Two Systems

Nature

\[ u(x, a, s^+), \delta \]

\[ \sigma \]
\[ f \]
\[ x \]
\[ a \]
\[ s \]

\[ \mu \]
\[ s \]

\[ \sigma \]
\[ f \]
\[ x \]
\[ a \]

R

M
Two Systems

Nature

Real System: $R(x, a, s^+, \sigma, \mu)$

Mockup System: $M(x, a, s^+, \sigma, \mu)$
Two Systems

Nature

\[ u(x, a, s^+), \delta \]

\[ \text{Real System: } R \]

\[ \text{Mockup System: } M \]
Two Systems

\[ u(x, a, s^+), \delta \]
Two Systems

\[ u(x, a, s^+), \delta \]

\[ \sigma \]

\[ \mathbb{P}[s] \]

\[ x \]

\[ a \]

\[ f \]

\[ \sigma \]

\[ \mu \]

\[ \mathbb{P}[s] \]

\[ x \]

\[ a \]

\[ f \]

\[ \sigma \]

\[ \mathbb{P}[s] \]

\[ x \]

\[ a \]

\[ f \]

\[ \sigma \]
Two Systems

Nature

\[ u(x, a, s^+), \delta \]

\[ \mathbb{P}[s] \]

\[ f \]

\[ \sigma \]

\[ \mu \]

\[ M \]
Consistency

\[ \mu[s^+] = \mathbb{P}[s^+] \]
Consistency

\[ \mu[s^+] = \mathbb{P}[s^+] \]
\[ = \lim_{t \to \infty} \mathbb{P}[S^{t+1} = s^+] \]
Consistency

\[
\begin{align*}
\mu[s^+] &= \mathbb{P}[s^+] \\
&= \lim_{t \to \infty} \mathbb{P}[S^{t+1} = s^+]
\end{align*}
\]
Consistency

\[
\begin{align*}
\mu[s^+] &= P[s^+] \\
&= \lim_{t \to \infty} P[S^{t+1} = s^+] \\
&= \sum_{w^+, w, x, a} \nu(w^+) [s^+] \cdot \pi[w, x] \cdot \sigma(x)[a] \cdot n(w, x, a)[w^+]
\end{align*}
\]
Consistency

\[ \begin{align*}
\mu[s^+] &= \mathbb{P}[s^+] \\
&= \lim_{t \to \infty} \mathbb{P}[S^{t+1} = s^+] \\
&= \sum_{w^+,w,x,a} \nu(w^+)[s^+] \cdot \pi(w, x) \cdot \sigma(x)[a] \cdot n(w, x, a)[w^+]
\end{align*} \]
Depth-1 Consistency

01010101010101010101010101010101010101010101010101010101010101...
Depth-1 Consistency

01010101010101010101010101010101...

P[00], P[01], P[11], P[10]
Depth-1 Consistency

\[01010101010101010101010101010101...\]

\[P[00], P[01], P[11], P[10] \iff P[0 \mid 0], P[1 \mid 0], P[0 \mid 1], P[1 \mid 1]\]
Depth-1 Consistency

\[01010101010101010101010101010101...\]

\[P[00], P[01], P[11], P[10] \iff P[0|0], P[1|0], P[0|1], P[1|1]\]
Depth-1 Consistency

\[ 01010101010101010101010101010101... \]

\[ P[00], P[01], P[11], P[10] \iff P[0 \mid 0], P[1 \mid 0], P[0 \mid 1], P[1 \mid 1] \]
Two Systems

Real System: $R$

Mockup System: $M$
Two Systems

Real System: $\mathbb{R}$

Mockup System: $\mathbb{M}$
Two Systems

Real System: \( \mathbf{R} \)

Mockup System: \( \mathbf{M} \)

\[ u(x, a, s^+), \delta \]
Two Systems

Real System: \( R \)

Mockup System: \( M \)
Two Systems

Real System: $\mathbf{R}$

Mockup System: $\mathbf{M}$

$u(x, a, s^+), \delta$
Two Systems

Real System: $\mathbf{R}$

Mockup System: $\mathbf{M}$

$P[s^+ \mid z]$
Two Systems

Real System: $R$

Mockup System: $M$

$\mathbb{P}[s^+ | z]$

$u(x, a, s^+), \delta$

$\mu$

$m^1$

$z$

$s$

$x$

$a$

$\sigma$

$f$

$m^1$

$z$

$s$

$x$

$a$

$\sigma$

$f$
Two Systems

Real System: $\mathbf{R}$

Mockup System: $\mathbf{M}$
Two Systems

Real System: \( \mathbf{R} \)

\[
\mathbb{P}[s^+ | z]
\]

\[
P\left[ s^+ \mid z \right]
\]

\[
\mathcal{M} \quad m^1
\]

\[
x, a
\]

\[
\sigma
\]

\[
f
\]

\[
\mu
\]

\[
\delta
\]

Mockup System: \( \mathbf{M} \)
Two Systems

Real System: $\mathbf{R}$

Mockup System: $\mathbf{M}$

$u(x, a, s^+), \delta$

$\mathbb{P}[s^+ | z]$

$\mu$

$\sigma$

$x$

$a$

$z$

$s$
Two Systems

Real System: $\mathbf{R}$

Mockup System: $\mathbf{M}$

$u(x, a, s^+), \delta$

$\mathbb{P}[s^+ \mid z]$
Depth-k Consistency

\[
\begin{align*}
\mu(z) [s^+] &= \mathbb{P}[s^+ | z] \\
&= \lim_{t \to \infty} \mathbb{P}[S^{t+1} = s^+ | Z^t = z]
\end{align*}
\]

\[
\begin{align*}
w^+ &\sim n(w, x, a) \\
s^+ &\sim \nu(w^+)
\end{align*}
\]
Recap
Start with one agent

Arbitrarily fix a model $m^k$

Split hard problem:

- Markov chain $R$ $\Rightarrow$ consistent predictor $\mu$
- MDP $M$ $\Rightarrow$ optimal strategy $\sigma$

EEE$s$ are fixed points of:
Stochastic Game

\[ u_1(x_1, a_1, s_1^+), \delta_1 \]

\[ u_2(x_2, a_2, s_2^+), \delta_2 \]
Stochastic Game

\[ u_1(x_1, a_1, s_1^+), \delta_1 \]
Stochastic Game

\[ u_1(x_1, a_1, s_1^+), \delta_1 \]
Stochastic Game

\[ u_1(x_1, a_1, s_1^+), \delta_1 \]
Stochastic Game

\[ u_1(x_1, a_1, s_1^+), \delta_1 \]
Stochastic Game

\[ u_1(x_1, a_1, s_1^+), \delta_1 \]

\[ u(x, a, s^+), \delta \]

\[ \sigma \]

\[ m^k \]

\[ x_1, a_1, x, a \]

\[ z_1, m^k, z, s \]

\[ u_1, f_1, f \]

\[ \sigma_1, \sigma \]

\[ x_1, \sigma, f \]

\[ s_1^+, s^+ \]

\[ \text{Nature} \]
Stochastic Game

\[ u_2(x_2, a_2, s_2^+), \delta_2 \]
Empirical-evidence Equilibrium

$m^{k_1}$ and $m^{k_2}$ fixed

$(\mu_1, \sigma_1, \mu_2, \sigma_2)$ is an empirical-evidence equilibrium (EEE) if:

- $\mu_1$ is consistent with $R$
- $\mu_2$ is consistent with $R$
- $\sigma_1$ is optimal for $M_1$
- $\sigma_2$ is optimal for $M_2$
Empirical-evidence Equilibrium

$m^{k_1}$ and $m^{k_2}$ fixed

$(\mu_1, \sigma_1, \mu_2, \sigma_2)$ is an $\varepsilon$ empirical-evidence equilibrium ($\varepsilon$ EEE) if:

- $\mu_1$ is consistent with $R$
- $\mu_2$ is consistent with $R$
- $\sigma_1$ is $\varepsilon$ optimal for $M_1$
- $\sigma_2$ is $\varepsilon$ optimal for $M_2$
EEE vs Nash

- optimization complexity fixed by agent not opponents
- always implementable
- each agent knows when at equilibrium
- less intrinsic to the problem
Existence of $\varepsilon$ EEEs

Theorem

For any $m^{k_1}$ and $m^{k_2}$, there exists an $\varepsilon$ EEE.
Existence of $\varepsilon$ EEEs

Theorem
For any $m^{k_1}$ and $m^{k_2}$, there exists an $\varepsilon$ EEE.

Proof.

- $\varepsilon$ and Gibbs distribution $\Rightarrow \mu_i \mapsto \sigma_i$ is a function
- $\mu \mapsto \mu$ is a continuous function
- set of predictors is compact and convex
- Brouwer’s fixed point theorem
Theorem

For any $m^{k_1}$ and $m^{k_2}$, there exists an EEE.
Theorem

For any $m^{k_1}$ and $m^{k_2}$, there exists a EEE.

Proof.

- $\mu \mapsto \mu$ is a closed-graph correspondence
- set of predictors is compact and convex
- Kakutani’s fixed point theorem
Theorem

*Exogenous EEEs in perfect-monitoring repeated games yield correlated equilibria of the underlying one-shot game.*

Repeated game:

Stochastic game without a state

Correlated equilibrium:

Nash equilibrium with common source of randomness
Recap
• multiagent EEE identical to single agent
• each agent arbitrarily picks a model $m^k$
• EEEs always exist
• EEEs induce correlated equilibria in repeated games
Asset Management Example

State holdings $x_i \in [0, M]$  
Action sell one, hold, or buy one $a_i \in \{-1, 0, 1\}$  
Signal price $p \in \{\text{Low, High}\}$  
Dynamic $x_i^+ = x_i + a_i$  
Stage cost $p \cdot a_i$  
Nature market trend $w \in \{\text{Bull, Bear}\}$  
Model depth $0$
Iterative Process

Update Rule

$$\mu_{i}^{r+1} = (1 - \alpha^r)\mu_i^r + \alpha^r (\tilde{\mu}_i - \mu_i^r)$$
Theoretical Predictor

Update Rule

\[ \mu_{i}^{r+1} = (1 - \alpha)\mu_{i}^{r} + \alpha(\bar{\mu}_{i} - \mu_{i}^{r}) \]
Empirical Predictor

\[ u_1(a_1, p) \]

\[ a_1 \]

\[ \sigma_1 \]

\[ f_1 \]

\[ x_1 \]

\[ \mu_{i}^{r+1} = (1 - \alpha^{r}) \mu_{i}^{r} + \alpha^{r} (\tilde{\mu}_{i}^{T} - \mu_{i}^{r}) \]

\[ \alpha^{r} \text{ non-summable, square-summable} \]
$P[\text{High}]$

$\mu_1^r$

$\mu_2^r$

Round $r$

$0$ $100$
Hawk-dove Game

Repeated game

<table>
<thead>
<tr>
<th></th>
<th>h</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>-1, -1</td>
<td>6, 0</td>
</tr>
<tr>
<td>D</td>
<td>0, 6</td>
<td>3, 3</td>
</tr>
</tbody>
</table>

Nash equilibria (H, d) and (D, h)

Want correlated equilibrium alternating between the two
Hawk-dove Game

Depth-2 models

Strategies:
\[ \sigma_1(d, h) = 0.999H + 0.001D \]
\[ \sigma_1(h, d) = 0.999D + 0.001H \]
\[ \sigma_1(h, h) = 0.5H + 0.5D \]
\[ \sigma_1(d, d) = 0.5H + 0.5D \]

Associated predictors:
\[ \mu_1(d, h) = 0.996d + 0.004h \]
\[ \mu_1(h, d) = 0.996h + 0.004d \]
\[ \mu_1(h, h) = 0.5h + 0.5d \]
\[ \mu_1(d, d) = 0.5h + 0.5d \]

Strategy approximately optimal as \( \delta \) close enough to one

Generalizes to any convex combination of pure Nash equilibria
Recap
Predictive given models and adaptation rule a EEE emerges
Prescriptive implement desired outcome as a EEE
Extensions

• $n$ agents
• endogenous models $z^+ \sim m(z, x, a, s)$
• notions of consistency: approximate, weak, and eventual
• convergence of empirical iterative process when theoretical one converges
Empirical-evidence Equilibrium (EEE)

Motivation: intractable problem
Definition: split into Markov chain and consistent MDPs
Existence: fixed-point theorems
Comparison: lower computational requirements
Characterization: correlated equilibrium in repeated game
Predictive Use: model to understand stock price
Prescriptive Use: desired outcome encoded as EEE
Publications


Endogenous Model

Real System: $R$

Mockup System: $M$

$$u(x, a, s^+), \delta$$
Brouwer’s Fixed-point Theorem
Kakutani’s Fixed-point Theorem
Consistency Formula

\[ \mu(z)[s^+] = \sum_{w^+} \nu(w^+)[s^+] \frac{\sum_{w,x,a} \pi_\sigma[w, x, z] \cdot \sigma(z)[a] \cdot n(w, x, a)[w^+]}{\sum_{w,x} \pi_\sigma[w, x, z]} \]
Consistency

Strong Consistency

\[ \mu(z)[s^+] = \lim_{t \to \infty} \mathbb{P}[S^{t+1} = s^+ \mid Z^t = z] \]

Weak Consistency

\[ \mu(z)[s^+] = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{P}[S^{t+1} = s^+ \mid Z^t = z] \]

Eventual Consistency

\[ \lim_{t \to \infty} \mathbb{P}[Z^t = z] > 0 \implies \mu(z)[s^+] = \lim_{t \to \infty} \mathbb{P}[S^{t+1} = s^+ \mid Z^t = z] \]
Learning Result

Theorem
Suppose the theoretical learning dynamic has a Lyapunov function. For a large enough observation window, the empirical learning dynamic converges.

Proof.
- ODE method for stochastic approximation
- Lyapunov stability of perturbed systems