Empirical-evidence Equilibria in Stochastic Games

Nicolas Dudebout

Outline

- Stochastic games
- Empirical-evidence equilibria (EEEs)
- Open questions in EEEs

Stochastic Games

- Game theory
- Markov decision processes

Game Theory

Decision making

$$u : \mathscr{A} \to \mathbb{R} \implies a^* \in \arg\max_{a \in \mathscr{A}} u(a)$$

Game theory

$$u_1 : \mathscr{A}_1 \times \mathscr{A}_2 \to \mathbb{R}$$
$$u_2 : \mathscr{A}_1 \times \mathscr{A}_2 \to \mathbb{R}$$

Nash Equilibrium

$$\begin{cases} a_1^* \in \arg\max_{a_1 \in \mathscr{A}_1} u_1\left(a_1, a_2^*\right) \\ a_2^* \in \arg\max_{a_2 \in \mathscr{A}_2} u_2\left(a_1^*, a_2\right) \end{cases}$$

Example: Battle of the Sexes

	F	O
F	2, 2	0, 1
O	0, 0	1, 3

Nash equilibria

- (F, F)
- (O, O)
- $(\frac{3}{4}F\ \frac{1}{4}O, \frac{1}{3}F\ \frac{2}{3}O)$

Markov Decision Process (MDP)

Dynamic $x^+ \sim f(x,a) \iff x^{t+1} \sim f\left(x^t, a^t\right)$

Stage cost $u(x,a)$

History $h^t = \left(x^0, x^1, \dots, x^t, a^0, a^1, \dots, a^t\right)$

Strategy $\sigma : \mathscr{H} \to \mathscr{A}$

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

Bellman's equation

$$U^*(x) = \max_{a \in \mathscr{A}} \left\{ u(x,a) + \delta \mathbb{E}_f \left[U^*\left(x^+\right) \,\middle|\, x, a \right] \right\}$$

Dynamic programming use knowledge of f

Reinforcement learning learn f from repeated interaction

Markov Decision Process (MDP)

Dynamic $x^+ \sim f(x,a) \iff x^{t+1} \sim f\left(x^t, a^t\right)$

Stage cost $u(x,a)$

History $h^t = \left(x^0, x^1, \ldots, x^t, a^0, a^1, \ldots, a^t\right)$

Strategy $\sigma : \mathscr{H} \to \mathscr{A}$

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

Bellman's equation

$$U^*(x) = \max_{a \in \mathscr{A}} \left\{ u(x,a) + \delta \mathbb{E}_f \left[U^*\left(x^+\right) \,\middle|\, x, a \right] \right\}$$

Dynamic programming use knowledge of f

Reinforcement learning learn f from repeated interaction

Markov Decision Process (MDP)

Dynamic $x^+ \sim f(x,a) \iff x^{t+1} \sim f\left(x^t, a^t\right)$

Stage cost $u(x,a)$

History $h^t = \left(x^0, x^1, \ldots, x^t, a^0, a^1, \ldots, a^t\right)$

Strategy $\sigma : \mathscr{X} \to \mathscr{A}$

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

Bellman's equation

$$U^*(x) = \max_{a\in\mathscr{A}}\left\{u(x,a) + \delta \mathbb{E}_f\left[U^*\left(x^+\right) \,\middle|\, x, a\right]\right\}$$

Dynamic programming use knowledge of f

Reinforcement learning learn f from repeated interaction

Imperfect Information (POMDP)

Dynamic $w^+ \sim n(w, a)$

Signal $s \sim \nu(w)$

History $h^t = \left(s^0, s^1, \ldots, s^t, a^0, a^1, \ldots, a^t\right)$

Strategy $\sigma : \mathscr{H} \to \mathscr{A}$

Belief $\mathbb{P}_{n,\nu,\sigma}[w \mid h]$

Imperfect Information (POMDP)

Dynamic $w^+ \sim n(w, a)$

Signal $s \sim \nu(w)$

History $h^t = \left(s^0, s^1, \ldots, s^t, a^0, a^1, \ldots, a^t\right)$

Strategy $\sigma : \mathscr{H} \to \mathscr{A}$

Belief $\mathbb{P}_{n,\nu,\sigma}[w \mid h]$

Imperfect Information (POMDP)

Dynamic $w^+ \sim n(w, a)$

Signal $s \sim \nu(w)$

History $h^t = \left(s^0, s^1, \ldots, s^t, a^0, a^1, \ldots, a^t\right)$

Strategy $\sigma : \Delta(\mathscr{W}) \to \mathscr{A}$

Belief $\mathbb{P}_{n,\nu,\sigma}[w \mid h]$

Stochastic Games

Dynamic $w^+ \sim n\big(w, a_1, a_2\big)$

Signals $\begin{cases} s_1 \sim \nu_1(w) \\ s_2 \sim \nu_2(w) \end{cases}$

Histories $\begin{cases} h_1^t = \big({s_1}^0, {s_1}^1, \dots, s_1^t, {a_1}^0, {a_1}^1, \dots, a_1^t\big) \\ h_2^t = \big({s_2}^0, {s_2}^1, \dots, s_2^t, {a_2}^0, {a_2}^1, \dots, a_2^t\big) \end{cases}$

Strategies $\begin{cases} \sigma_1 : \mathscr{H}_1 \to \mathscr{A}_1 \\ \sigma_2 : \mathscr{H}_2 \to \mathscr{A}_2 \end{cases}$

Beliefs $\begin{cases} \mathbb{P}_{n,\nu_1,\sigma_1,\nu_2,\sigma_2}\big[w, h_2 \,\big|\, h_1\big] \\ \mathbb{P}_{n,\nu_1,\sigma_1,\nu_2,\sigma_2}\big[w, h_1 \,\big|\, h_2\big] \end{cases}$

Stochastic Games

Dynamic $w^+ \sim n\big(w, a_1, a_2\big)$

Signals $\begin{cases} s_1 \sim \nu_1(w) \\ s_2 \sim \nu_2(w) \end{cases}$

Histories $\begin{cases} h_1^t = \big(s_1{}^0, s_1{}^1, \ldots, s_1^t, a_1{}^0, a_1{}^1, \ldots, a_1^t\big) \\ h_2^t = \big(s_2{}^0, s_2{}^1, \ldots, s_2^t, a_2{}^0, a_2{}^1, \ldots, a_2^t\big) \end{cases}$

Strategies $\begin{cases} \sigma_1 : \mathscr{H}_1 \to \mathscr{A}_1 \\ \sigma_2 : \mathscr{H}_2 \to \mathscr{A}_2 \end{cases}$

Beliefs $\begin{cases} \mathbb{P}_{n,\nu_1,\sigma_1,\nu_2,\sigma_2}\big[w, h_2 \,\big|\, h_1\big] \\ \mathbb{P}_{n,\nu_1,\sigma_1,\nu_2,\sigma_2}\big[w, h_1 \,\big|\, h_2\big] \end{cases}$

Existing Approaches

- (Weakly) belief-free equilibrium
- Mean-field equilibrium
- Incomplete theories

Empirical-evidence Equilibria

Motivation

0. Pick arbitrary strategies
1. Formulate simple but *consistent* models
2. Design strategies optimal w.r.t. models, then, back to 1.

Empirical-evidence equilibrium is a fixed point:

- Strategies optimal w.r.t. models
- Models consistent with strategies

Example: Asset Management

Trading one asset on the stock market

Model based on

- information published by the company
- observed trading activity

Model very different for each agent

Multiple to Single Agent

Multiple to Single Agent

Single Agent Setup

Single Agent Setup

$x^+ \sim f(x, a, s)$

Nature

Single Agent Setup

$$\boxed{x^+ \sim f(x, a, s)} \xleftarrow{\quad s \quad} \boxed{\text{Nature}}$$

Single Agent Setup

$$\boxed{x^+ \sim f(x, a, s)} \xleftarrow{\quad s \quad} \boxed{\begin{array}{c} w^+ \sim n(w, x, a) \\ s \sim \nu(w) \end{array}}$$

Example: Asset Management

$$\boxed{x^+ \sim f(x,a,s)} \xleftarrow{\quad s \quad} \boxed{\begin{array}{c} w^+ \sim n(w,x,a) \\ s \sim \nu(w) \end{array}}$$

State	holding $x \in \{0 \,..\, M\}$
Action	sell one, hold, or buy one $a \in \{-1, 0, 1\}$
Signal	price $p \in \{\text{Low}, \text{High}\}$
Stage cost	$p \cdot a$
Nature	w represents market sentiment, political climate, other traders

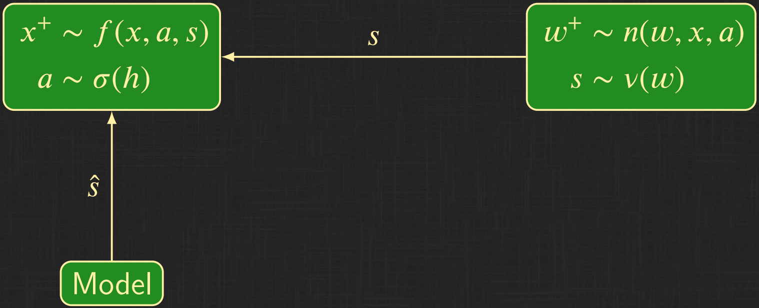
Single Agent Setup

$$\begin{tikzcd} \boxed{x^+ \sim f(x, a, s)} & \boxed{\begin{array}{c} w^+ \sim n(w, x, a) \\ s \sim \nu(w) \end{array}} \arrow[l, "s"'] \end{tikzcd}$$

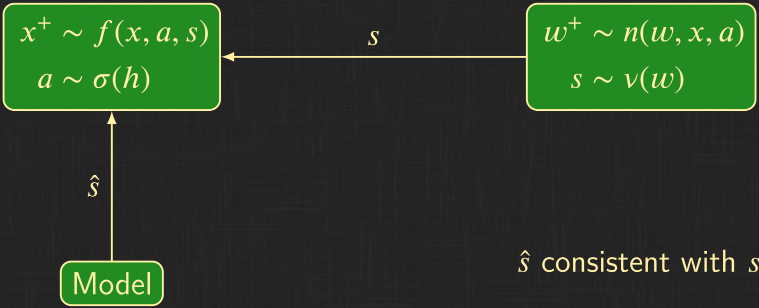
Single Agent Setup

$$\boxed{\begin{array}{c} x^+ \sim f(x, a, s) \\ a \sim \sigma(h) \end{array}} \xleftarrow{\quad s \quad} \boxed{\begin{array}{c} w^+ \sim n(w, x, a) \\ s \sim \nu(w) \end{array}}$$

Single Agent Setup



Single Agent Setup



$\hat{s}$ consistent with s

Single Agent Setup

$$\begin{tikzcd}
\boxed{\begin{array}{c} x^+ \sim f(x, a, s) \\ a \sim \sigma(h) \end{array}} & & \boxed{\begin{array}{c} w^+ \sim n(w, x, a) \\ s \sim \nu(w) \end{array}} \arrow[ll, "s"'] \\
\boxed{\text{Model}} \arrow[u, "\hat{s}"] & &
\end{tikzcd}$$

$\hat{s}$ consistent with σ

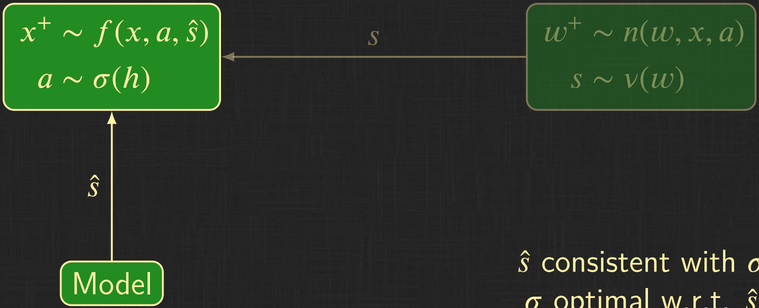
Single Agent Setup

$$x^+ \sim f(x, a, \hat{s})$$
$$a \sim \sigma(h)$$

$$\text{Model} \xrightarrow{\hat{s}} \begin{array}{c} x^+ \sim f(x, a, \hat{s}) \\ a \sim \sigma(h) \end{array} \xleftarrow{s} \begin{array}{c} w^+ \sim n(w, x, a) \\ s \sim \nu(w) \end{array}$$

$\hat{s}$ consistent with σ

Single Agent Setup



$\hat{s}$ consistent with σ
σ optimal w.r.t. $\hat{s}$

Depth-k Consistency

Consider a binary stochastic process s
0100010001001010010110111010000111010101...

Depth-k Consistency

Consider a binary stochastic process s
0100010001001010010110111010000111010101...

- 0 characteristic: $\mathbb{P}[s=0], \mathbb{P}[s=1]$
- 1 characteristic: $\begin{array}{l}\mathbb{P}\left[ss^{+}=00\right], \mathbb{P}\left[ss^{+}=10\right], \\ \mathbb{P}\left[ss^{+}=01\right], \mathbb{P}\left[ss^{+}=11\right]\end{array}$
- ...
- k characteristic: probability of strings of length $k+1$

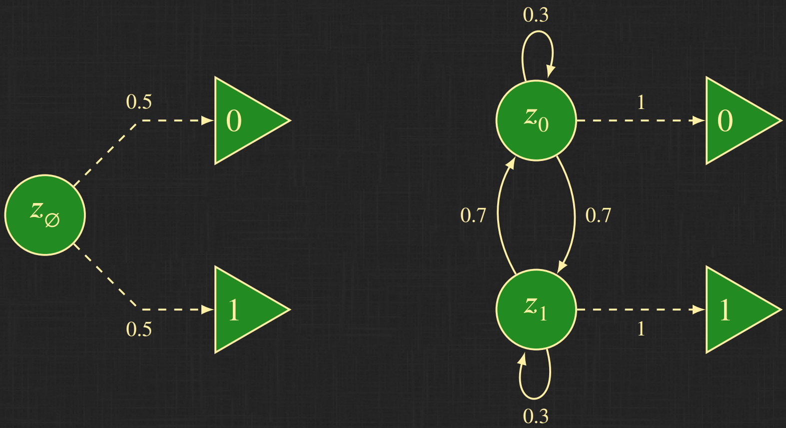
Depth-k Consistency

Consider a binary stochastic process s
0100010001001010010110111010000111010101...

- 0 characteristic: $\mathbb{P}[s = 0], \mathbb{P}[s = 1]$
- 1 characteristic: $\begin{array}{l} \mathbb{P}\left[ss^+ = 00\right], \mathbb{P}\left[ss^+ = 10\right], \\ \mathbb{P}\left[ss^+ = 01\right], \mathbb{P}\left[ss^+ = 11\right] \end{array}$
- ...
- k characteristic: probability of strings of length $k + 1$

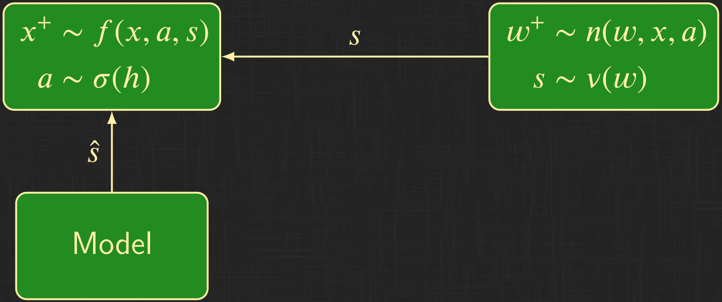
Definition Two processes s and s' are depth-k consistent if they have the same k characteristic

Depth-k Consistency: Example



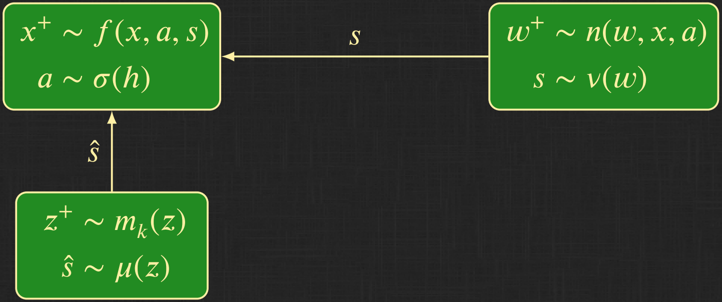
Complete picture

Fix a depth $k \in \mathbb{N}$



Complete picture

Fix a depth $k \in \mathbb{N}$



z contains the last k observed signals

$$\mu\big(z = (s_1, s_2, \ldots, s_k)\big)[s_{k+1}] = \mathbb{P}_\sigma\left[s^{t+1} = s_{k+1} \,\middle|\, s^t = s_k, \ldots, s^{t-k+1} = s_1\right]$$

Complete picture

Fix a depth $k \in \mathbb{N}$

$$
\begin{tikzcd}
\boxed{\begin{array}{c} x^+ \sim f(x, a, s) \\ a \sim \sigma(x, z) \end{array}} & & \boxed{\begin{array}{c} w^+ \sim n(w, x, a) \\ s \sim \nu(w) \end{array}} \arrow[ll, "s"'] \\
\boxed{\begin{array}{c} z^+ \sim m_k(z) \\ \hat{s} \sim \mu(z) \end{array}} \arrow[u, "\hat{s}"] & &
\end{tikzcd}
$$

z contains the last k observed signals

$$
\mu\big(z = (s_1, s_2, \ldots, s_k)\big)[s_{k+1}] = \\
\mathbb{P}_\sigma\left[s^{t+1} = s_{k+1} \,\middle|\, s^t = s_k, \ldots, s^{t-k+1} = s_1\right]
$$

Complete picture

Fix a depth $k \in \mathbb{N}$

$$
\begin{tikzcd}
\boxed{\begin{array}{c} x^+ \sim f(x, a, s) \\ a \sim \sigma(x, z) \end{array}} & \boxed{\begin{array}{c} w^+ \sim n(w, x, a) \\ s \sim \nu(w) \end{array}} \arrow[l, "s"'] \\
\boxed{\begin{array}{c} z^+ \sim m_k(z) \\ \hat{s} \sim \mu(z) \end{array}} \arrow[u, "\hat{s}"]
\end{tikzcd}
$$

$\sigma \mapsto \mu$ consistent with σ

$\mu \mapsto \sigma$ optimal w.r.t. μ

z contains the last k observed signals

$$\mu\big(z = (s_1, s_2, \ldots, s_k)\big)[s_{k+1}] = \\ \mathbb{P}_\sigma\left[s^{t+1} = s_{k+1} \,\middle|\, s^t = s_k, \ldots, s^{t-k+1} = s_1\right]$$

Definition

(σ, μ) is an empirical-evidence optimum (EEO) for k iff

- σ is optimal w.r.t. μ
- μ is depth-k consistent with σ

Definition

(σ, μ) is an empirical-evidence optimum (EEO) for k iff

- σ is optimal w.r.t. μ
- μ is depth-k consistent with σ

(σ, μ) is an ϵ empirical-evidence optimum (ϵ EEO) for k iff

- σ is ϵ optimal w.r.t. μ
- μ is depth-k consistent with σ

Existence Result

Theorem
For all k and ϵ, there exists an ϵ EEO for k

Existence Result

Theorem

For all k and ϵ, there exists an ϵ EEO for k

Proof sketch

Prove continuity of $\sigma \mapsto \mu \mapsto \sigma$

$\sigma : \mathscr{X} \times \mathscr{Z} \to \Delta(\mathscr{A})$

σ parametrized over a simplex (convex and compact)

Apply Brouwer's fixed point theorem

Complete picture

Fix a depth $k \in \mathbb{N}$

$$\begin{tikzcd}
\boxed{\begin{array}{c} x^+ \sim f(x,a,s) \\ a \sim \sigma(x,z) \end{array}} & \boxed{\begin{array}{c} w^+ \sim n(w,x,a) \\ s \sim \nu(w) \end{array}} \arrow[l, "s"'] \\
\boxed{\begin{array}{c} z^+ \sim m_k(z) \\ \hat{s} \sim \mu(z) \end{array}} \arrow[u, "\hat{s}"] &
\end{tikzcd}$$

$\sigma \mapsto \mu$ consistent with σ

$\mu \mapsto \sigma$ optimal w.r.t. μ

z contains the last k observed signals

$$\mu\left(z = \left(s_1, s_2, \ldots, s_k\right)\right)[s_{k+1}] = \mathbb{P}_\sigma\left[s^{t+1} = s_{k+1} \,\middle|\, s^t = s_k, \ldots, s^{t-k+1} = s_1\right]$$

Multiagent Setting

$$x_i^+ \sim f_i(x_i, a_i, s_i)$$
$$a_i \sim \sigma_i(x_i, z_i)$$

$$\xleftarrow{\;s_i\;}$$

$$w^+ \sim n(w, x, a)$$
$$s \sim \nu(w)$$

$$\uparrow \hat{s}_i$$

$$z_i^+ \sim m_{i,k}(z_i)$$
$$\hat{s}_i \sim \mu_i(z_i)$$

$$x = (x_1, x_2, \ldots, x_N)$$
$$a = (a_1, a_2, \ldots, a_N)$$
$$s = (s_1, s_2, \ldots, s_N)$$

Empirical-evidence Equilibrium

(σ, μ) is an empirical-evidence equilibrium (EEE) for $K = \left(k_1, k_2, \ldots, k_N\right)$ iff

- for all i, σ_i is optimal w.r.t. μ_i
- for all i, μ_i is depth-k_i consistent with σ

Empirical-evidence Equilibrium

(σ, μ) is an empirical-evidence equilibrium (EEE) for $K = \left(k_1, k_2, \dots, k_N\right)$ iff

- for all i, σ_i is optimal w.r.t. μ_i
- for all i, μ_i is depth-k_i consistent with σ

Theorem

For all K and ϵ, there exists an ϵ EEE for K

Open Questions

- endogenous model depending on action
- large number of agents
- large k
- relating EEE to other concepts (MFE, optimum)
- offline computation
- online learning using empirical evidence

Open Questions

- endogenous model depending on action
- large number of agents
- large k
- relating EEE to other concepts (MFE, optimum)
- offline computation
- online learning using empirical evidence

Open Questions

- endogenous model depending on action
- large number of agents
- large k
- relating EEE to other concepts (MFE, optimum)
- offline computation
- online learning using empirical evidence

Open Questions

- endogenous model depending on action
- large number of agents
- large k
- relating EEE to other concepts (MFE, optimum)
- offline computation
- online learning using empirical evidence

Example: Asset Management

State	holdings $x_i \in \{0 \,..\, M\}$
Action	sell one, hold, or buy one $a_i \in \{-1, 0, 1\}$
Signal	price $p \in \{\text{Low}, \text{High}\}$
Dynamic	$x_i^+ = x_i + a_i$
Stage cost	$p \cdot a_i$
Nature	market trend $b \in \{\text{Bull}, \text{Bear}\}$ $w = (b, p)$

Nature is a *sticky bear*

Example: Asset Management

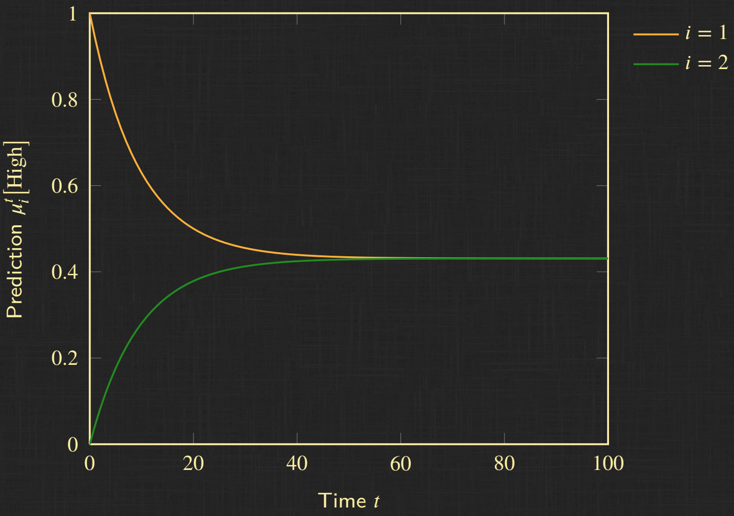
0. Pick arbitrary models μ
1. Design strategies σ optimal w.r.t. models μ
2. Formulate consistent models μ_{upd}, then, back to 1.

Depth-0 consistency:

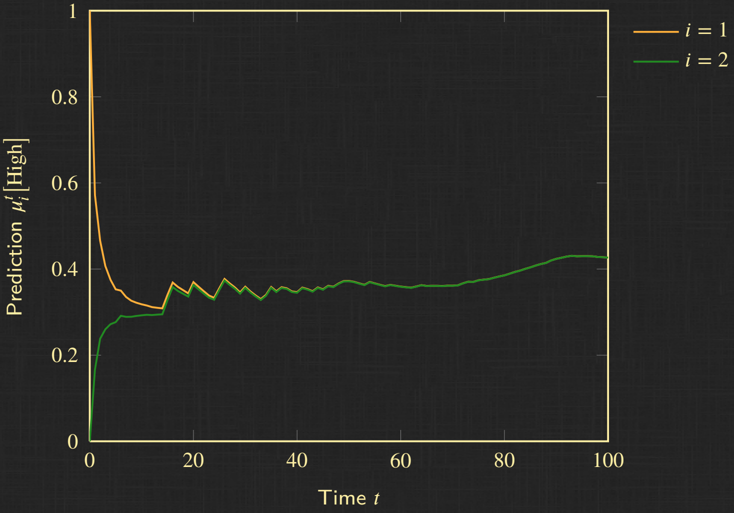
- $\mu_1 = 1$
- $\mu_2 = 0$

$$\mu_i^{t+1} = (1-\alpha)\mu_i^t + \alpha\left(\mu_{i,\text{upd}}^t - \mu_i^t\right)$$

Learning Results: Offline



Learning Results: Online



Empirical-evidence Equilibria

- Introduce
- Contrast
- Compute