Competing Bandits: Learning under Competition

Yishay Mansour, Aleksandrs Slivkins, Zhiwei Steven Wu

manuscript on arXiv (2017)

Speaker: Joseph Chuang-Chieh Lin

Institute of Information Science
Academia Sinica
Taiwan

21 July 2017

Steven Wu

Computer and Information Science
University of Pennsylvania

Email:
steven7woo [at] gmail.com

About Me

My name is Zhiwei Steven Wu (吴志威). I am a fifth-year PhD student in the CIS Department at Penn, where I am fortunate to be co-advised by Michael Kearns and Aaron Roth.

I am broadly interested in algorithms and machine learning, especially in the areas of differential privacy, fairness in machine learning, and algorithmic economics.

My CV can be found here.

News

- June, 2017 - Starting in fall 2018, I will be joining the University of Minnesota as an Assistant Professor in the Computer Science & Engineering Department. Before that, I will be a Postdoctoral Researcher at Microsoft Research-New York City (MSR-NYC).
- June, 2017 - I defended my thesis and received the 2017 Morris and Dorothy Rubinoff Dissertation Award!

Microsoft | Technologies | Documentation | Resources | Search Microsoft Research

Research | Research areas | Products & Downloads | Programs & Events | People | Careers | Blogs | Labs & Locations

Microsoft Research Lab – New York City

"Microsoft Research New York City investigates computational social science, algorithmic economics and prediction markets, machine learning, and information retrieval. The researchers in our lab interact deeply with the vibrant academic and tech communities in the New York metropolitan area. Our primary goal is to advance the state of the art in interdisciplinary research, and our research also enhances Microsoft products and services, through direct transfer of technology and through impact on Microsoft strategy."

— Jennifer Chayes, Managing Director, Microsoft Research New England and Microsoft Research New York City

Outline

Introduction

- Modern systems strive to learn from interactions with users, and many engage in exploration.
  - product recommendations, web search, spam detection, ...
- Interplay b/w *exploration* and *competition*.
  - To balance the exploration for learning and the competition for users.
- Users' roles:
  - **customers:** generate revenue.
  - **sources of data:** for learning
  - **self-interested agents:** choosing among the competing systems.
- Actually, here "systems" $\Rightarrow$ MAB algorithms.

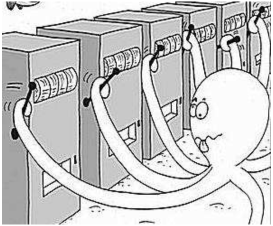
Introduction

- Modern systems strive to learn from interactions with users, and many engage in exploration.
  - product recommendations, web search, spam detection, . . .
- Interplay b/w *exploration* and *competition*.
  - To balance the exploration for learning and the competition for users.
- Users' roles:
  - **customers:** generate revenue.
  - **sources of data:** for learning
  - **self-interested agents:** choosing among the competing systems.
- Actually, here "systems" $\Rightarrow$ MAB algorithms.

Introduction

- Modern systems strive to learn from interactions with users, and many engage in exploration.
  - product recommendations, web search, spam detection, . . .
- Interplay b/w *exploration* and *competition*.
  - To balance the exploration for learning and the competition for users.
- Users' roles:
  - **customers:** generate revenue.
  - **sources of data:** for learning
  - **self-interested agents:** choosing among the competing systems.
- Actually, here "systems" $\Rightarrow$ MAB algorithms.

Multi-armed bandits (MAB)



Introduction (contribution)

- **Question:** Whether and to which extent competition incentivizes innovation.
  - Innovation: adoption of **better** algorithm.
- Competition vs. innovation relationship.
  - Well-studied in economics.
- Users' "decision rule" for choosing among the firms:
  - relates to users' rationality;
  - controls the severity of competition.

Principles & agents

- Two firms (principals) simultaneously engage in exploration and compete for T users (agents).
- In each round, a new agent arrives and chooses one of the two principals.
- The principle chooses a recommendation: an action $a_t \in A = [K]$, where A is a fixed set of actions (same for both principals and all rounds).
- The agent follows this recommendation, receives a reward $r_t \in [0, 1]$, and reports it back to the principal.

⋆ Principals simultaneously announce their learning algorithms *before* the agents start arriving, and cannot change them afterwards.

⋆ Principals' utility: the *number* of agents choosing it.

Principles & agents

- Two firms (principals) simultaneously engage in exploration and compete for T users (agents).
- In each round, a new agent arrives and chooses one of the two principals.
- The principle chooses a recommendation: an action $a_t \in A = [K]$, where A is a fixed set of actions (same for both principals and all rounds).
- The agent follows this recommendation, receives a reward $r_t \in [0, 1]$, and reports it back to the principal.

⋆ Principals simultaneously announce their learning algorithms *before* the agents start arriving, and cannot change them afterwards.

⋆ Principals' utility: the *number* of agents choosing it.

Principles & agents (the common prior)

- For each action $a \in A$, there is a parametric family $\psi_a(\cdot)$ of reward distributions, parameterized by the mean reward μ_a.
- The mean reward vector $\mu = (\mu_a : a \in A)$ is drawn from prior distribution $\mathcal{P}_{\mathrm{mean}}$ before round 1.
- Whenever $a \in A$ is chosen, the reward is drawn independently from $\psi_a(\mu_a)$.

⋆ The Bayesian prior on rewards $\mathcal{P}$ is comprised of:
  - the prior $\mathcal{P}_{\mathrm{mean}}$ & the distributions $(\psi_a(\cdot) : a \in A)$.

Principles & agents (the information structure)

- The prior $\mathcal{P}$ is known to everyone.
- The mean rewards $\{\mu_a\}_{a \in A}$ are not revealed to anybody.
- Each principal is completely unaware of the rounds when the other is chosen.

Bayesian-expected rewards

- alg_i, the algorithm of principal i, $i \in \{1, 2\}$.
- $n_i(t)$: the number of rounds before t in which this principal is chosen.
- $\text{rew}_i(n)$: alg_i's Bayesian-expected reward for the n-th step.
  - Without competition, just as a bandit algorithm.
- $\mathbf{E}[r_t \mid \text{principal } i \text{ is chosen in round } t \text{ and } n_i(t) = n] = \text{rew}_i(n+1)$.

Agents' response

- Each agent t chooses principal i_t:
  - It chooses a distribution over the principals (p_t: prob. of choosing principal 1);
  - then draws independently from this distribution.
- $\mathcal{I}_t$: the information available to agent t before the round.
- For each principal i, its posterior mean reward:

$$\mathrm{PMR}_i(t) := \mathbf{E}[r_t \mid \mathcal{I}_t \text{ and } i_t = i] = \mathbf{E}[\mathrm{rew}_i(n_i(t)+1) \mid \mathcal{I}_t] = \mathbf{E}_{n\sim\mathcal{N}_{i,t}}[\mathrm{rew}_i(n+1)].$$

  $\mathcal{N}_{i,t}$: the posterior for $n_{i,t}$.
- Response function $p_t = f_{\mathrm{resp}}(\mathrm{PMR}_1(t) - \mathrm{PMR}_2(t))$.
  - $f_{\mathrm{resp}}(\cdot) : [-1,1] \mapsto [0,1]$.
  - **Assumption:** The same for all agents, and known to all agents.

Agents' response

- Each agent t chooses principal i_t:
  - It chooses a distribution over the principals (p_t: prob. of choosing principal 1);
  - then draws independently from this distribution.

- $\mathcal{I}_t$: the information available to agent t before the round.
- For each principal i, its posterior mean reward:

$$\mathrm{PMR}_i(t) := \mathbf{E}[r_t \mid \mathcal{I}_t \text{ and } i_t = i] = \mathbf{E}[\mathrm{rew}_i(n_i(t)+1) \mid \mathcal{I}_t] = \mathbf{E}_{n \sim \mathcal{N}_{i,t}}[\mathrm{rew}_i(n+1)].$$

$\mathcal{N}_{i,t}$: the posterior for $n_{i,t}$.

- Response function $p_t = f_{\mathrm{resp}}(\mathrm{PMR}_1(t) - \mathrm{PMR}_2(t))$.
  - $f_{\mathrm{resp}}(\cdot) : [-1, 1] \mapsto [0, 1]$.
  - **Assumption:** The same for all agents, and known to all agents.

Agents' response

- Each agent t chooses principal i_t:
  - It chooses a distribution over the principals (p_t: prob. of choosing principal 1);
  - then draws independently from this distribution.

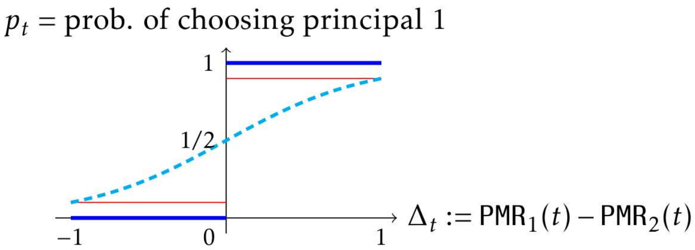
- $\mathcal{I}_t$: the information available to agent t before the round.
- For each principal i, its posterior mean reward:

$$\mathrm{PMR}_i(t) := \mathbf{E}[r_t \mid \mathcal{I}_t \text{ and } i_t = i] = \mathbf{E}[\mathrm{rew}_i(n_i(t)+1) \mid \mathcal{I}_t] = \mathbf{E}_{n \sim \mathcal{N}_{i,t}}[\mathrm{rew}_i(n+1)].$$

$\mathcal{N}_{i,t}$: the posterior for $n_{i,t}$.

- Response function $p_t = f_{\mathrm{resp}}(\mathrm{PMR}_1(t) - \mathrm{PMR}_2(t))$.
  - $f_{\mathrm{resp}}(\cdot) : [-1, 1] \mapsto [0, 1]$.
  - **Assumption:** The same for all agents, and known to all agents.

Response functions



- HardMax.
- HardMax&Random
- SoftMax.

The Bayesian Instataneous Regret

Bayesian Instataneous Regret (BIR)

$$\mathrm{BIR}_i(n) := \mathbf{E}_{\mu \sim \mathcal{P}_{\mathrm{mean}}} \left[\max_{a \in A} \mu_a \right] - \mathrm{rew}_i(n).$$

Quality of MAB algorithms in terms of BIR

- Smart MAB algorithms, such as UCB1 [Auer et al. 2002], Successive Elimination [Even-Dar et al. 2006], ...
  - $\mathrm{BIR}(n) = \tilde{O}(n^{-1/2})$.
- Naïve MAB algorithms that separate exploration and exploitation, such as Explore-then-Exploit, ϵ-Greedy, ...
  - $\mathrm{BIR}(n) = \tilde{O}(n^{-1/3})$.
- DynamicGreedy: at each step, recommends the currently best posterior action (i.e., $\arg\max_a\{\mathbf{E}[\mu_a \mid \mathcal{I}]\}$, $\mathcal{I}$: the information available so far).
  - $\mathrm{BIR}(n) = \Omega(1)$.
- StaticGreedy: always recommends the prior best action (i.e., $\arg\max_a\{\mathbf{E}_{\mu\sim\mathcal{P}_{\mathrm{mean}}}[\mu_a]\}$).
  - $\mathrm{BIR}(n) = \Omega(1)$.

Assumptions

- We focus on monotone MAB algorithms (BIR(n) is non-increasing).

⋆ DynamicGreedy is monotone (proof ignored).

- Each action has a chance to be the best:
  $\forall a \in A,\ \Pr_{\mu \sim \mathcal{P}_{\text{mean}}}[\mu_a > \mu_{a'}, \forall a' \in A \setminus \{a\}] > 0.$

- Posterior mean rewards of actions are pairwise distinct.

- Prior mean rewards of actions are also pairwise distinct.

Deviation of two algorithms

Two MAB algorithms deviate at a step n if

- $\exists a \in A$ and a realization h of step-n history, such that h is feasible for both algorithms;
- under h the two algoirthms choose a with different probability.

On full rationality

Theorem 4.1

Assume

- HardMax response function with fair tie-breaking (i.e., $f_{\mathrm{resp}}(0) = 1/2$);
- alg_1 is DynamicGreedy and alg_2 deviates from DynamicGreedy starting from some step $n_0 < T$.

Then all agents in rounds $t \geq n_0$ select principal 1.

Corollary 4.2

The competition game b/w principals has a unique Nash equilibrium:

- ▷ both principals choose DynamicGreedy.

Proof of Theorem 4.1

Lemma 4.4

With algorithms as in Theorem 4.1, we have $\mathrm{rew}_1(n_0) > \mathrm{rew}_2(n_0)$.

Lemma 4.5

Suppose alg_1 is monotone, and $\mathrm{PMR}_1(t_0) > \mathrm{PMR}_2(t_0)$ for some round t_0. Then, $\mathrm{PMR}_1(t) > \mathrm{PMR}_2(t)$ for all subsequent rounds t.

Sketch of the proof of Lemma 4.4

Lemma 4.4

With algorithms as in Theorem 4.1, we have $\mathrm{rew}_1(n_0) > \mathrm{rew}_2(n_0)$.

- H_{1,n_0} and H_{2,n_0} have the same distribution.
- Using coupling, WLOG assume that $H_{1,n_0} = H_{2,n_0} = H$.
- At local step n_0, DynamicGreedy chooses an action a_{1,n_0} such that for any realization $h \in \mathrm{support}(H)$ and any action $a \in A \setminus \{a_{1,n_0}\}$,

$$\mathrm{PMR}(a_{1,n_0} \mid H = h) > \mathrm{PMR}(a \mid H = h) \qquad (*).$$

- Since two algorithms deviate at step n_0, there is $h \in \mathrm{support}(H)$ and an action $a \in A$ such that

$$\Pr[a = a_{2,n_0} \neq a_{1,n_0} \mid H = h] > 0.$$

- Integrating (*) over $a \sim (a_{2,n_0} \mid H = h)$ and $h \sim H$, we obtain $\mathrm{rew}_1(n_0) > \mathrm{rew}_2(n_0)$.

Sketch of the proof of Lemma 4.4

Lemma 4.4

With algorithms as in Theorem 4.1, we have $\text{rew}_1(n_0) > \text{rew}_2(n_0)$.

- H_{1,n_0} and H_{2,n_0} have the same distribution.
- Using coupling, WLOG assume that $H_{1,n_0} = H_{2,n_0} = H$.
- At local step n_0, DynamicGreedy chooses an action a_{1,n_0} such that for any realization $h \in \text{support}(H)$ and any action $a \in A \setminus \{a_{1,n_0}\}$,

$$\text{PMR}(a_{1,n_0} \mid H = h) > \text{PMR}(a \mid H = h) \qquad (*).$$

- Since two algorithms deviate at step n_0, there is $h \in \text{support}(H)$ and an action $a \in A$ such that

$$\Pr[a = a_{2,n_0} \neq a_{1,n_0} \mid H = h] > 0.$$

- Integrating (*) over $a \sim (a_{2,n_0} \mid H = h)$ and $h \sim H$, we obtain $\text{rew}_1(n_0) > \text{rew}_2(n_0)$.

Sketch of the proof of Lemma 4.4

Lemma 4.4

With algorithms as in Theorem 4.1, we have $\text{rew}_1(n_0) > \text{rew}_2(n_0)$.

- H_{1,n_0} and H_{2,n_0} have the same distribution.
- Using coupling, WLOG assume that $H_{1,n_0} = H_{2,n_0} = H$.
- At local step n_0, DynamicGreedy chooses an action a_{1,n_0} such that for any realization $h \in \text{support}(H)$ and any action $a \in A \setminus \{a_{1,n_0}\}$,

$$\text{PMR}(a_{1,n_0} \mid H = h) > \text{PMR}(a \mid H = h) \qquad (*).$$

- Since two algorithms deviate at step n_0, there is $h \in \text{support}(H)$ and an action $a \in A$ such that

$$\Pr[a = a_{2,n_0} \neq a_{1,n_0} \mid H = h] > 0.$$

- Integrating (*) over $a \sim (a_{2,n_0} \mid H = h)$ and $h \sim H$, we obtain $\text{rew}_1(n_0) > \text{rew}_2(n_0)$.

Sketch of the proof of Lemma 4.4

Lemma 4.4

With algorithms as in Theorem 4.1, we have $\text{rew}_1(n_0) > \text{rew}_2(n_0)$.

- H_{1,n_0} and H_{2,n_0} have the same distribution.
- Using coupling, WLOG assume that $H_{1,n_0} = H_{2,n_0} = H$.
- At local step n_0, DynamicGreedy chooses an action a_{1,n_0} such that for any realization $h \in \text{support}(H)$ and any action $a \in A \setminus \{a_{1,n_0}\}$,

$$\text{PMR}(a_{1,n_0} \mid H = h) > \text{PMR}(a \mid H = h) \qquad (*).$$

- Since two algorithms deviate at step n_0, there is $h \in \text{support}(H)$ and an action $a \in A$ such that

$$\Pr[a = a_{2,n_0} \neq a_{1,n_0} \mid H = h] > 0.$$

- Integrating (*) over $a \sim (a_{2,n_0} \mid H = h)$ and $h \sim H$, we obtain $\text{rew}_1(n_0) > \text{rew}_2(n_0)$.

Sketch of the proof of Lemma 4.5

Lemma 4.5

Suppose alg_1 is monotone, and $\text{PMR}_1(t_0) > \text{PMR}_2(t_0)$ for some round t_0. Then, $\text{PMR}_1(t) > \text{PMR}_2(t)$ for all subsequent rounds t.

- Induction on t, with base case $t = t_0$.
- $\mathcal{N} := \mathcal{N}_{1,t_0}$: agents' posterior distribution for n_{1,t_0}.
- By induction, all agents from t_0 to $t-1$ chose principal 1.
- $\text{PMR}_1(t) = \mathbf{E}_{n\sim\mathcal{N}}[\text{rew}_1(n+1+t-t_0)] \geq \mathbf{E}_{n\sim\mathcal{N}}[\text{rew}_1(n+1)] = \text{PMR}_1(t_0) > \text{PMR}_2(t_0) = \text{PMR}_2(t)$.

Sketch of the proof of Lemma 4.5

Lemma 4.5

Suppose alg_1 is monotone, and $\text{PMR}_1(t_0) > \text{PMR}_2(t_0)$ for some round t_0. Then, $\text{PMR}_1(t) > \text{PMR}_2(t)$ for all subsequent rounds t.

- Induction on t, with base case $t = t_0$.
- $\mathcal{N} := \mathcal{N}_{1,t_0}$: agents' posterior distribution for n_{1,t_0}.
- By induction, all agents from t_0 to $t-1$ chose principal 1.
- $\text{PMR}_1(t) = \mathbf{E}_{n\sim\mathcal{N}}[\text{rew}_1(n+1+t-t_0)] \geq \mathbf{E}_{n\sim\mathcal{N}}[\text{rew}_1(n+1)] = \text{PMR}_1(t_0) > \text{PMR}_2(t_0) = \text{PMR}_2(t)$.

Proof of Theorem 4.1

- Since the two algorithms coincide on the first $n_0 - 1$ steps, we have
  - $\text{rew}_1(n) = \text{rew}_2(n)$ for any $n < n_0$.
  - $\mathcal{N}_{1,n_0} = \mathcal{N}_{2,n_0} \triangleq \mathcal{N}$.
- By Lemma 4.4, $\text{rew}_1(n_0) > \text{rew}_2(n_0)$.
- Therefore,

$$\begin{aligned}\text{PMR}_1(n_0) = \mathbf{E}_{n\sim\mathcal{N}}[\text{rew}_1(n+1)] &= \sum_{n=0}^{n_0-1} \mathcal{N}(n) \cdot \text{rew}_1(n+1) \\ &> \mathcal{N}(n_0-1) \cdot \text{rew}_2(n_0) + \sum_{n=0}^{n_0-2} \mathcal{N}(n) \cdot \text{rew}_2(n+1) \\ &= \mathbf{E}_{n\sim\mathcal{N}}[\text{rew}_1(n+1)] = \text{PMR}_2(n_0).\end{aligned}$$

- By Lemma 4.5, all subsequent agents choose principal 1, too.

Relaxed rationality: HardMax & Random

On the relaxed rationality

- Each principal is always chosen with some positive baseline probability.
- A principal with asymptotically better BIR wins by a large margin:
    - After a "learning phase" of constant duration, all agents choose this principal with maximal possible probability $f_{\mathrm{resp}}(1)$.

Well-defined for an infinite time horizon

- Denoting $\epsilon_0 = \frac{1}{2} f_{\mathrm{resp}}(-1)$, for some constant n_0, we have

$$\forall n \geq n_0, \quad \mathrm{BIR}_1(\epsilon n)/\mathrm{BIR}_2(n) < \frac{1}{2}.$$

alg$_1$ BIR-dominates alg$_2$

- $\forall n \geq n_0$, $\mathrm{BIR}_2(n) > 2e^{-\epsilon_0 n/6}$.
  - Assumption on the "bad" algorithm.

A version of the competition game b/w the two principals

- Principals can only choose from a **finite** set $\mathcal{A}$ of monotone MAB algorithms.
- One of these algorithms is "better" than all others.
  - We call it special.
  - It BIR-dominates all other algorithms in $\mathcal{A}$.
- We call this game the restricted competition game.

On relaxed rationality: HardMax & Random

Theorem 5.1

Assume

- HardMax&Random response function;
- both algorithms are well-defined for an infinite time horizon.

Then, each agent $t \geq n_0$ chooses principal 1 with maximal possible probability $f_{\mathrm{resp}}(1)$.

Corollary 5.3

Assume HardMax&Random response function. Consider the restricted competition game with special algorithm `alg`. Then, for any sufficiently large time horizon T, this game has a unique Nash equilibrium:

- ▷ both principals choose `alg`.

Proof of Theorem 5.1

Theorem 5.1

Assume

- HardMax&Random response function;
- both algorithms are well-defined for an infinite time horizon.

Then, each agent $t \geq n_0$ chooses principal 1 with maximal possible probability $f_{\mathrm{resp}}(1)$.

- Consider round $t \geq n_0$.
- Each agent choose principal 1 with prob. $\geq f_{\mathrm{resp}}(-1) > 0$.
  - $\epsilon_0 := f_{\mathrm{resp}}(-1)/2$.
- $\mathbf{E}[n_1(t+1)] \geq 2\epsilon_0 t$.
- By Chernoff bounds, we have $n_1(t+1) \geq \epsilon_0 t$ with prob. $\geq 1 - e^{-\epsilon_0 t/6}$.

⋆ We need to prove that $\mathrm{PMR}_1(t) - \mathrm{PMR}_2(t) > 0$.

Proof of Theorem 5.1

Theorem 5.1

Assume

- HardMax&Random response function;
- both algorithms are well-defined for an infinite time horizon.

Then, each agent $t \geq n_0$ chooses principal 1 with maximal possible probability $f_{\mathrm{resp}}(1)$.

- Consider round $t \geq n_0$.
- Each agent choose principal 1 with prob. $\geq f_{\mathrm{resp}}(-1) > 0$.
  - $\epsilon_0 := f_{\mathrm{resp}}(-1)/2$.
- $\mathbf{E}[n_1(t+1)] \geq 2\epsilon_0 t$.
- By Chernoff bounds, we have $n_1(t+1) \geq \epsilon_0 t$ with prob. $\geq 1 - e^{-\epsilon_0 t/6}$.

⋆ We need to prove that $\mathrm{PMR}_1(t) - \mathrm{PMR}_2(t) > 0$.

Proof of Theorem 5.1 (contd.)

- For any m_1, m_2, consider the quantity:

$$\Delta(m_1, m_2) := \mathrm{BIR}_2(m_2 + 1) - \mathrm{BIR}_1(m_1 + 1).$$

- Whenever $m_1 \geq \epsilon_0 t - 1$ and $m_2 < t$,

$$\Delta(m_1, m_2) \geq \Delta(\epsilon_0 t,\ t) \geq \mathrm{BIR}_2(t)/2.$$

- Therefore,

$$\begin{aligned}\mathrm{PMR}_1(t) - \mathrm{PMR}_2(t) &= \mathbf{E}_{\substack{m_1 \sim \mathcal{N}_{1,t},\\ m_2 \sim \mathcal{N}_{2,t}}} [\Delta(m_1, m_2)] \\ &\geq -e^{-\epsilon_0 t/6} + \mathbf{E}_{\substack{m_1 \sim \mathcal{N}_{1,t},\\ m_2 \sim \mathcal{N}_{2,t}}} [\Delta(m_1, m_2) \mid m_1 \geq \epsilon_0 t - 1] \\ &\geq \mathrm{BIR}_2(t)/2 - e^{-\epsilon_0 t/6} \\ &> 0\end{aligned}$$

Proof of Theorem 5.1 (contd.)

- For any m_1, m_2, consider the quantity:

$$\Delta(m_1, m_2) := \mathrm{BIR}_2(m_2 + 1) - \mathrm{BIR}_1(m_1 + 1).$$

- Whenever $m_1 \geq \epsilon_0 t - 1$ and $m_2 < t$,

$$\Delta(m_1, m_2) \geq \Delta(\epsilon_0 t,\ t) \geq \mathrm{BIR}_2(t)/2.$$

- Therefore,

$$\begin{aligned}
\mathrm{PMR}_1(t) - \mathrm{PMR}_2(t) &= \mathbf{E}_{\substack{m_1 \sim \mathcal{N}_{1,t},\\ m_2 \sim \mathcal{N}_{2,t}}} [\Delta(m_1, m_2)] \\
&\geq -e^{-\epsilon_0 t/6} + \mathbf{E}_{\substack{m_1 \sim \mathcal{N}_{1,t},\\ m_2 \sim \mathcal{N}_{2,t}}} [\Delta(m_1, m_2) \mid m_1 \geq \epsilon_0 t - 1] \\
&\geq \mathrm{BIR}_2(t)/2 - e^{-\epsilon_0 t/6} \\
&> 0
\end{aligned}$$

SoftMax Response Function

A even more relaxed rationality

SoftMax response function

f_{resp} is SoftMax if the following conditions hold:

- $f_{\text{resp}}(\cdot) \in [\epsilon, 1-\epsilon]$ for some $\epsilon \in (0, 1/2)$ (bounded away from 0 and 1).
- $\exists \delta_0, c_0, c_0' > 0$, such that $\forall x \in [-\delta_0, \delta_0]$, $c_0 \leq f_{\text{resp}}(x) \leq c_0'$ (smooth around 0).
- $f_{\text{resp}}(0) = \frac{1}{2}$ (fair tie-breaking).

Results on SoftMax response functions

Theorem 6.2

Assume

- SoftMax response function;
- alg_1 BIR-dominates alg_2.

Then, each agent $t \geq n_0$ chooses principal 1 with probability $\geq \frac{1}{2} + \frac{c_0}{4}\mathrm{BIR}_2(t)$.

Corollary 6.3

- Assume SoftMax&Random response function.
- Consider the restricted competition game with special algorithm `alg`.
- Assume that all other algorithms satisfy $\mathrm{BReg}(n) \to \infty$.

Then, for any sufficiently large T, this game has a unique Nash equilibrium:

▷ both principals choose `alg`.

$\mathrm{BReg}(n) := \sum_{n'=1}^{n} BIR(n')$.

Results on SoftMax response functions

Theorem 6.2

Assume

- SoftMax response function;
- alg_1 BIR-dominates alg_2.

Then, each agent $t \geq n_0$ chooses principal 1 with probability $\geq \frac{1}{2} + \frac{c_0}{4}\text{BIR}_2(t)$.

Corollary 6.3

- Assume SoftMax&Random response function.
- Consider the restricted competition game with special algorithm `alg`.
- Assume that all other algorithms satisfy $\text{BReg}(n) \to \infty$.

Then, for any sufficiently large T, this game has a unique Nash equilibrium:

▷ both principals choose `alg`.

$\text{BReg}(n) := \sum_{n'=1}^{n} BIR(n')$.

Weakly BIR-domination

alg_1 weakly-BIR-dominates alg_2

For some $n_0(T) \in \text{poly}\log(T)$ and constants $\beta_0, \alpha_0 \in (0, 1/2)$,

$$\forall n \geq n_0(T), \quad \frac{\text{BIR}_1((1-\beta_0)n)}{\text{BIR}_2(n)} < 1 - \alpha_0.$$

Results on SoftMax response functions (contd.)

Theorem 6.4

Assume

- SoftMax response function;
- alg_1 weakly-BIR-dominates alg_2;
- $\exists n(\epsilon)$ such that $\text{BIR}_2(n) > e^{-\epsilon n}$ for each $n \geq n(\epsilon)$.

Then, each agent $t \geq n_0$ chooses principal 1 with probability $\geq \frac{1}{2} + \frac{c_0 \alpha_0}{4} \text{BIR}_2(t)$.

Corollary 6.5

- Assume SoftMax&Random response function.
- Consider the restricted competition game with special algorithm `alg` (weakly).
- All other algorithms satisfy $\text{BReg}(n) \to \infty$.

Then, for any sufficiently large T, this game has a unique Nash equilibrium:

▷ both principals choose `alg`.

Concluding remarks

- f_{resp} controls directly "the extent" to which agents make rational decisions.
- We measure *innovation* in terms of whether and when `alg` is chosen in an equilibrium.
  - HardMax: **no innovation**; DynamicGreedy is chosen over `alg`.
  - HardMax&Random: **some innovation**; `alg` is chosen as long as it BIR-dominates.
  - SoftMax: **more innovation**; `alg` is chosen as long as it weakly-BIR-dominates.

Thank you.