The Games People Play: Learning and Equilibria

Peter Key
peter.key@microsoft.com

Microsoft Research

How to act/learn in large systems?

Large Systems of Interacting agents

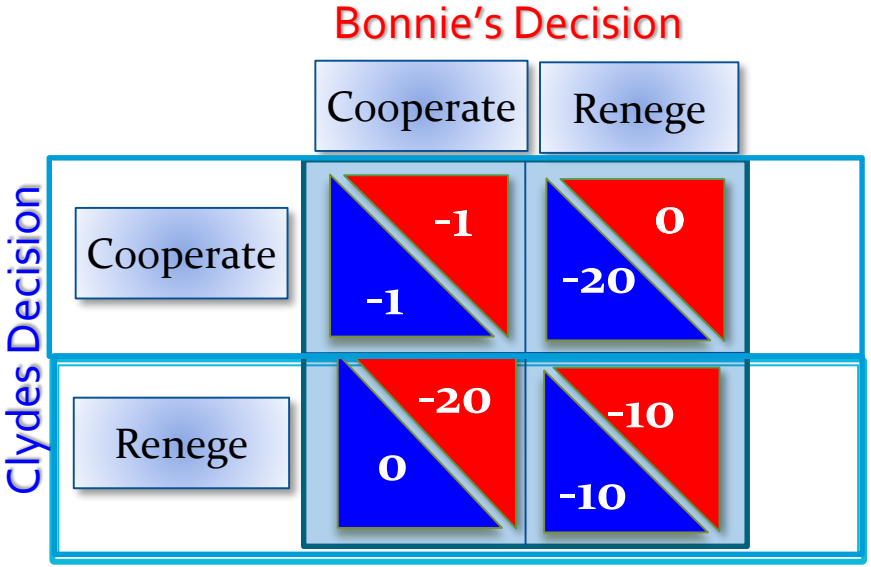
Outline

- Context
- Descriptive: Some empirical evidence
- Prescriptive: Mean Field Equilibria
- Example: bidding in auctions with budgets

Background

Nash equilibrium

Prisoner's dilemma



Nash: $U_i(s_i, s_{-i}) \geq U_i(s'_i, s_{-i})$
$\forall s'_i \in S_i, \forall i$

- Action $A = A_1 \times \cdots \times A_N$
- Strategies $S = S_1 \times \cdots \times S_N = S_i \times S_{-i}$
  $s_i \sim \Pi(A_i)$
  Π is set of all probability distr.,
- Utility $U_i(\cdot): S \mapsto \mathbb{R}$ for player *i*

Bayes Nash $EU_i(s_i, s_{-i}) \geq EU_i(s'_i, s_{-i})$

- Epistemic *types* θ_i
- *Common prior* $p(\theta)$ on $\theta = \theta_1 \times \cdots \times \theta_N$

"Renege" is a dominant strategy

Renege-renege is the only Nash Equilibrium

Nash Equilibrium $\leftrightarrow$ Fixed Point

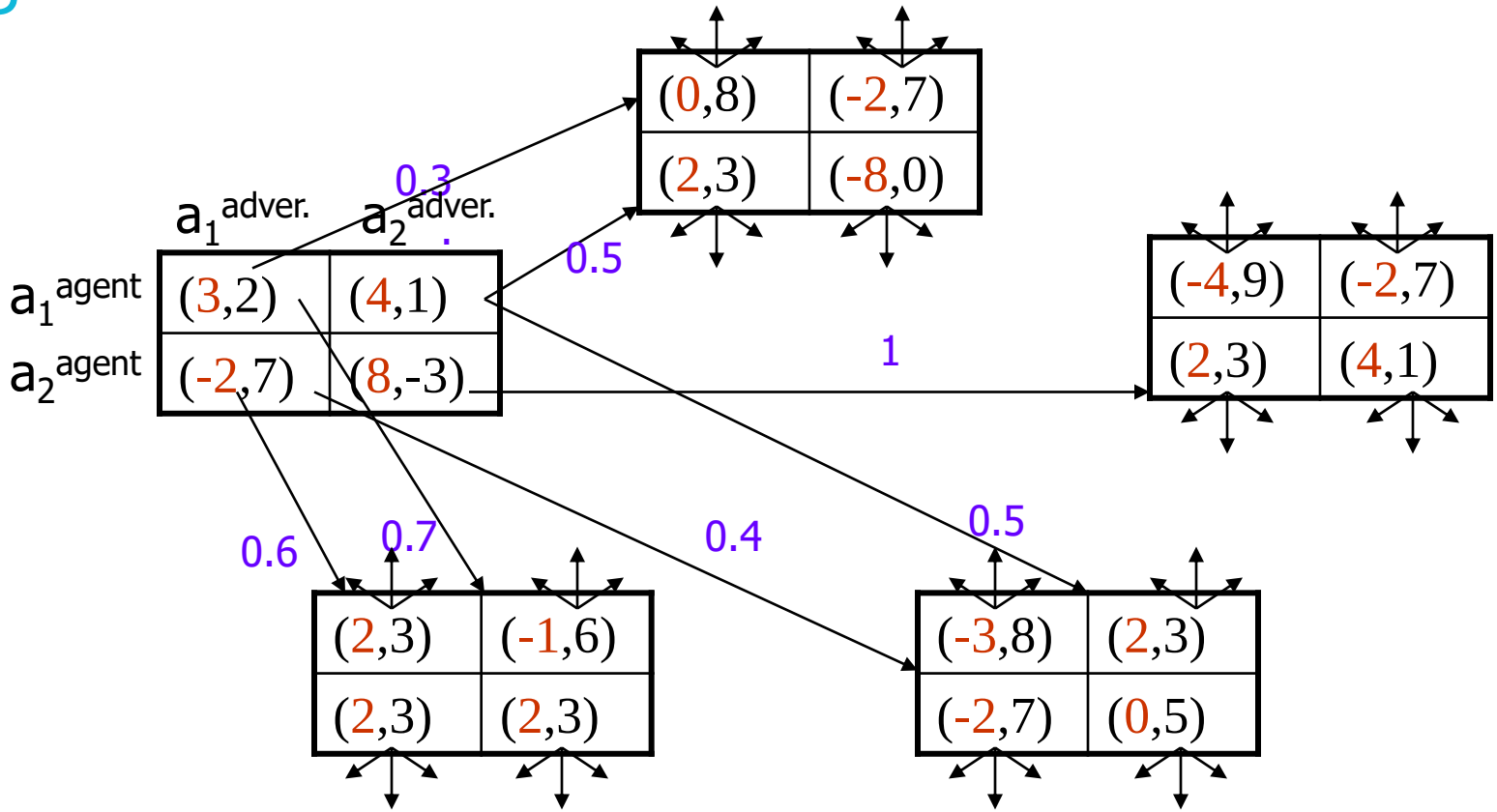
- Best response $r_i(s_{-i}) \triangleq \arg\max_{s_i} U_i(s_i, s_{-i})$

$$r \triangleq \left(r_i(s_{-i}), r_{-i}(s_i)\right): \; S \to 2^S$$

- Nash equilibrium exists $\Leftrightarrow$ r has a fixed point

where strategy profile $s = (s_i, s_{-i}) \in S$

Stochastic Game: a general environment model



Stochastic Games (Markov Game)

- *N* players, actions *A*
- *X* is a *set* of games, state x
- $P: X \times A \times X \mapsto [0,1]$ transition function
  - $p(x'|x,a)$
- $U_i: X \times A \mapsto \mathbb{R}$ (reward/utility per time)

*Definition: A **Markov Perfect Equilibrium** is a strategy profile if only has Markov strategies (stationary), and is a Nash equilibrium regardless of starting state*

Theorem: Always exists in finite player /finite state *discounted* case

But … in Large Systems???

Equilibria wish list

- Stable
- Computable
- For boundedly-rational players

Fictitious Play

Play best response to empirical distribution of opponent

Initialise Prior Beliefs
Do
    Play best response | beliefs
    Observe outcome
    Update beliefs
Until End

Empirical studies: what happens in practice?

Colonel Blotto, Facebook game

Colonel Blotto Game

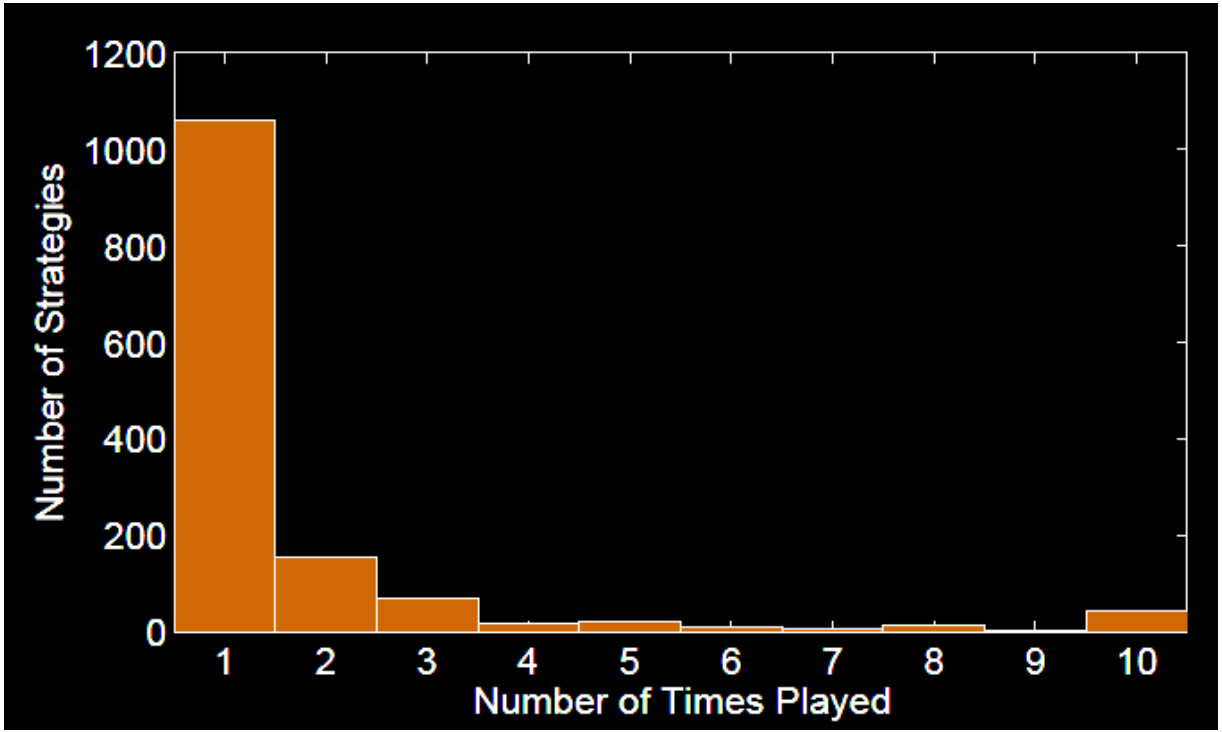
[Kohli, Bachrach, Stillwell, Kearns, Herbrich, Graepel, 2012]

- Allocate 100 troops to 5 territories
- 2 player game, player with most troops wins territory
- Nash equilibrium complex
  - Allocate [0..40] troops uniformly per territory s.t. total=100.
  - Meaning??

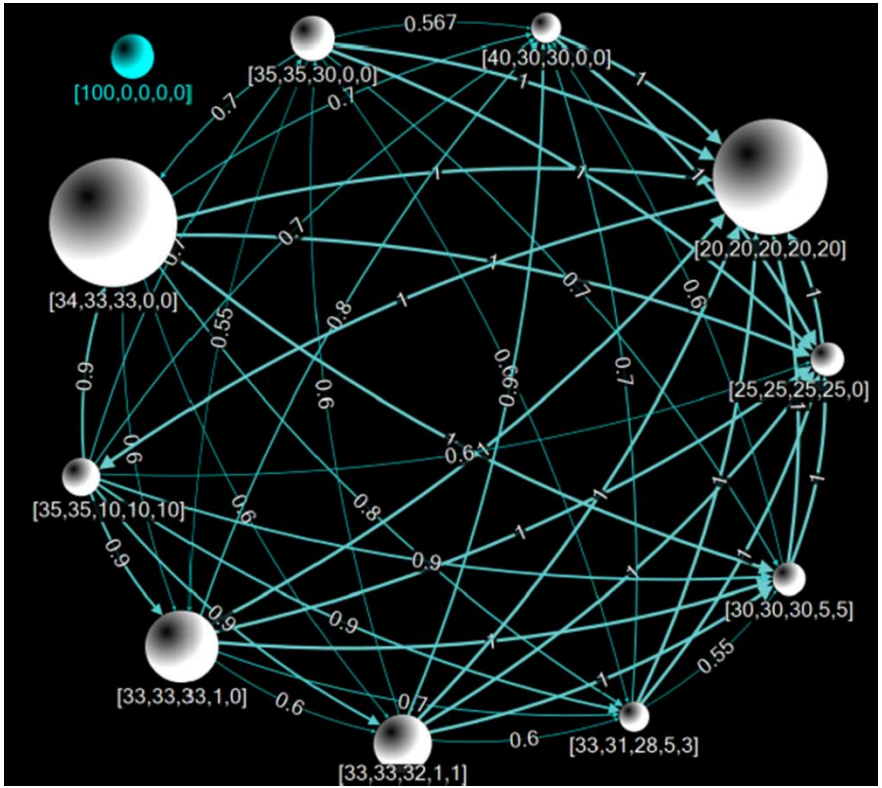
Most frequently used strategies

Strategy					Number of Times Used
0	33	0	34	33	271
20	20	20	20	20	235
33	1	33	0	33	127
1	32	33	1	33	97
35	30	35	0	0	68
0	100	0	0	0	67
10	10	35	35	10	58
25	25	25	25	0	50

Most frequently used strategies



Relationship between frequent strategies



Player strategies

- Adopt simple strategies
  - Copy winner, one-step reasoning
- But… territorial biases (left-side bias), and $\Pr(win|vs\ friend) \neq \Pr(win|vs\ stranger)$
- Other games:
  - Forza Auction house: some users approximate simple AIMD-like algorithms for setting buyout/reserve prices
  - Online auctions

All pay auction

- n players bid for single item, worth m
- Highest bid wins; all pay
- $n = 2$ players:
  - Symmetric Nash: bid uniform $[0, m]$
  - $\mathbb{E}[revenue]$=m; $\mathbb{E}[player\ utility] = 0$

The experiment: Doubloon Dash

Oren, Bachrach, Kash & K. 2013

- An online, two-player Facebook game, simulating an all-pay auction.

Each player submits a value in the range $[0, 10^4]$.
$(m = 10{,}000)$

Game and Payments

- Bid 0 to 10,000
- M-Turkers paid :

participation+ bonus

bonus based utility $\bar{u}$

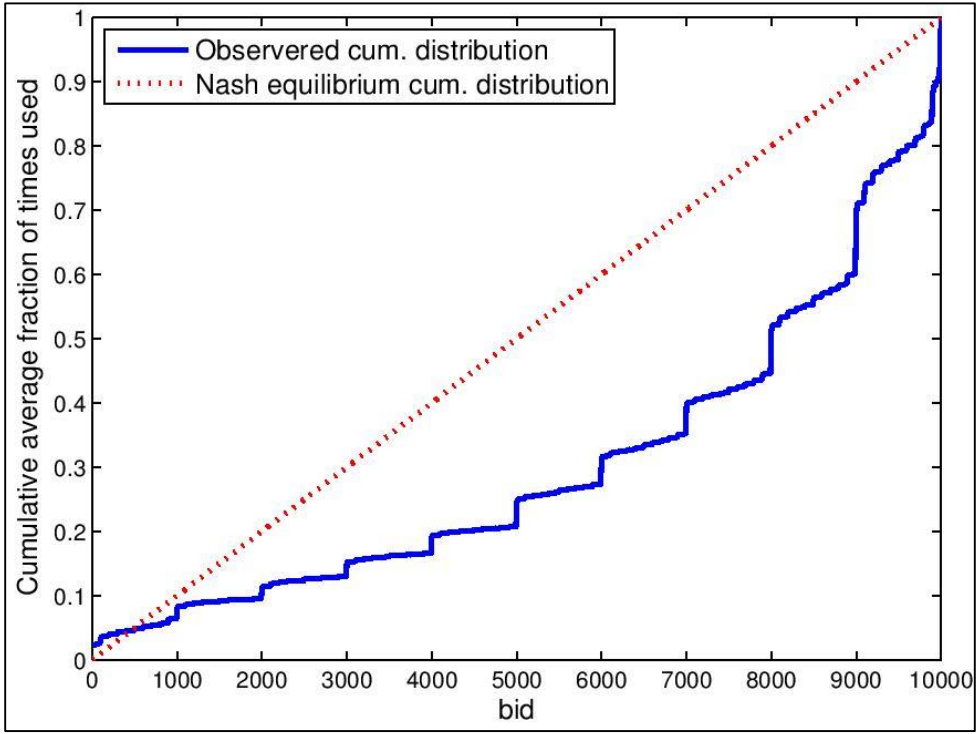
$\$\left(\frac{\bar{u}}{9{,}999}+1\right) \in \$[0,2]$

if played >30 games

User behaviour : 518 players, ~13k games

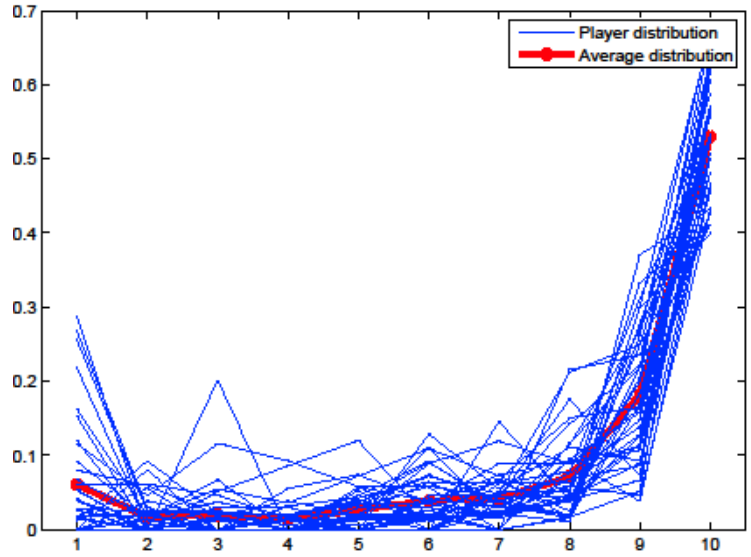
- "Spam" or "Cheat/collude"
  - 50% spammers (use few simple bids (banned))
  - Use Sybils (multiple persona)
- Have focal points
  - small bid set, concentrated on $b \equiv 0 \; mod \; 1000$, $\mathrm{b} \equiv 0 \; mod \; 10$ etc.
- Don't play Nash!
- Overbid (av. revenue $= 1.37m$, $\mathbb{E}[return] = m$)
  - Low returns $\overline{utility|no\ collude} = -0.2m$, $\mathbb{E}[utility] = 0$

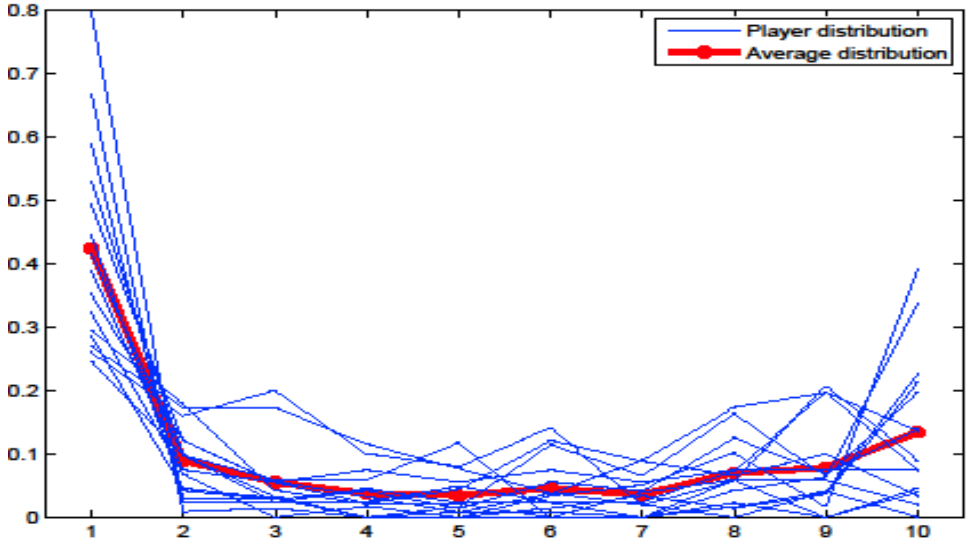
Bid Distribution



A More Refined Analysis of Bid Distribution

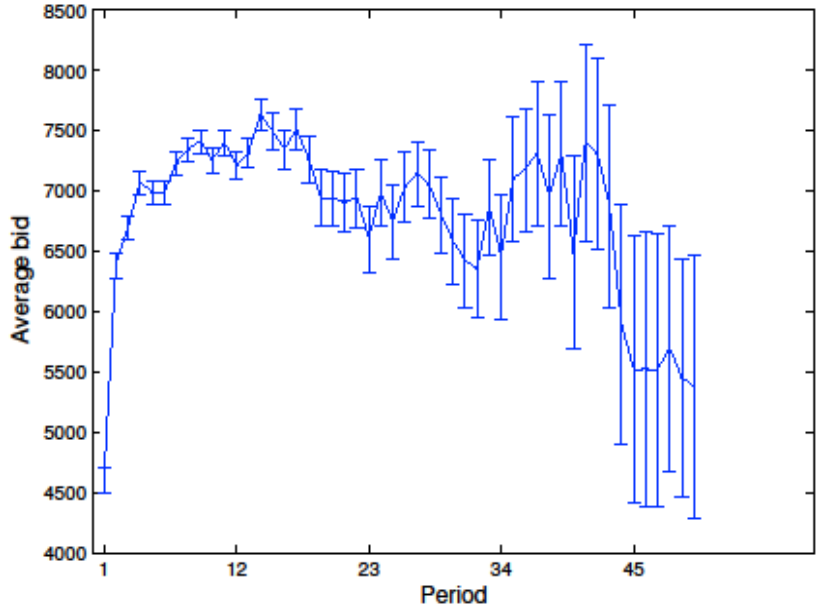
- Bin each player's bid bins of width 1,000 each..
- Cluster players based on their bin frequency vectors
- Players can be thought of belonging to different behavioural classes.





Bids per over time

- Mean bids over 4 consecutive games
- Suggests learning?



Dynamic Learning

- Players affected by winning / losing
  - Behave differently (unless bid is small, <1000)
  - Change their bids
    - Increase after losing
      - (almost all increase 2/3 of the time, conditional on bid change)
    - Decrease after winning
  - Intuitive explanation: If a player wins a game, she will try to increase her utility by lowering her bid, while still winning
  - But … little evidence of users following standard limited myopia learning, such as limited recall best response

Learning Algorithms: Dumb Heuristic

- A simple incremental heuristic.
  - On each game t:
    - If the previous game won with bid b^{t-1}: bid $b^t = b^{t-1} - c$
    - If previous game lost: bid $b^t = b^{t-1} + c$.
  - Test: run on sequences of observed opponent bids with $c = 1000$
  - Mean payoff : $-2{,}308$, standard error of 89.2.
  - The heuristic provides comparable performance to the empirically observed performance!

Limited Recall Dynamics

- Limited Recall Best Response
  - Select window of length d
  - Choose bid : optimal bid for average return of previous d opponents bids (o_t)

$$b_t = \arg\max_b \frac{\sum_{s=t-d}^{t-1} U(b, o_s)}{d}$$

Multiplicative Weights Update

- Choose discrete bid set $C = \{b_1, \dots, b_B\}$
- Initialise: $t \to 1, w_i(t) = 1 \ \forall i$
- REPEAT
  - Select bid b_i with probability $\propto w_i(t)$
  - Observe opponent bid o calculate normalised payoff $u_i = u(b_i, o) \in [-1, 1]$ for each i
  - $$w_i(t+1) \to \begin{cases} w_i(t)(1+\lambda)^{u_i} & if\ u_i \geq 0 \\ w_i(t)(1-\lambda)^{u_i} & u_i < 0 \end{cases}$$
  - $t \to t + 1$
  - UNTIL STOP

Learning Algorithms

- Users appear not to use "standard" myopic learning techniques, such as
  - Limited recall
  - Limited Best response
  - Multiplicative weights
- All these would outperform users!

Mean field equilibria

Mean Field Approach

- In Stochastic game, agent *i* wants to maximize expected discounted return $V_i(x\,|s_i, s_{-i}\,)$. **Congizant** strategies

- Mean Field Approximation: assume large number of agents, and agent

  i only reacts to average state distribution $\bar{x}$, $\widehat{V}_i(x_i\,|\bar{x}\,, \bar{s})$

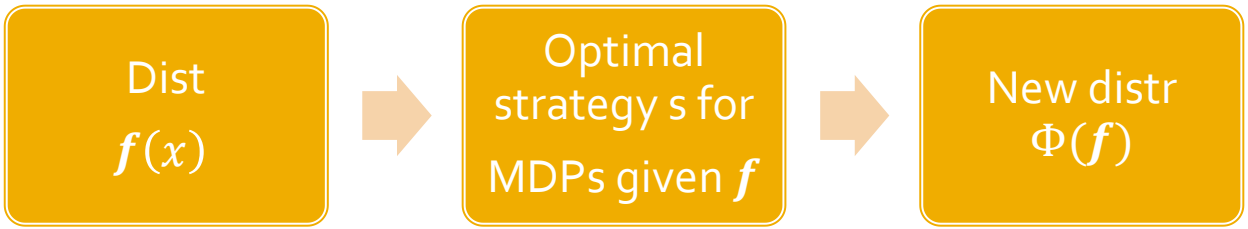
  - Where $\bar{s}$ is an ***oblivious strategy****: depends only on* x_t

Mean Field Equilibria (MFE)

- Write *distribution* of states $\bar{x}$ as $\boldsymbol{f}(x)$
- Def: A strategy s and distribution of states $\boldsymbol{f}(x)$ is a *mean field equilibrium if*

  1) s is an optimal oblivious strategy given $\boldsymbol{f}(x)$

  2) $\boldsymbol{f}(x)$ is a steady state distribution of s



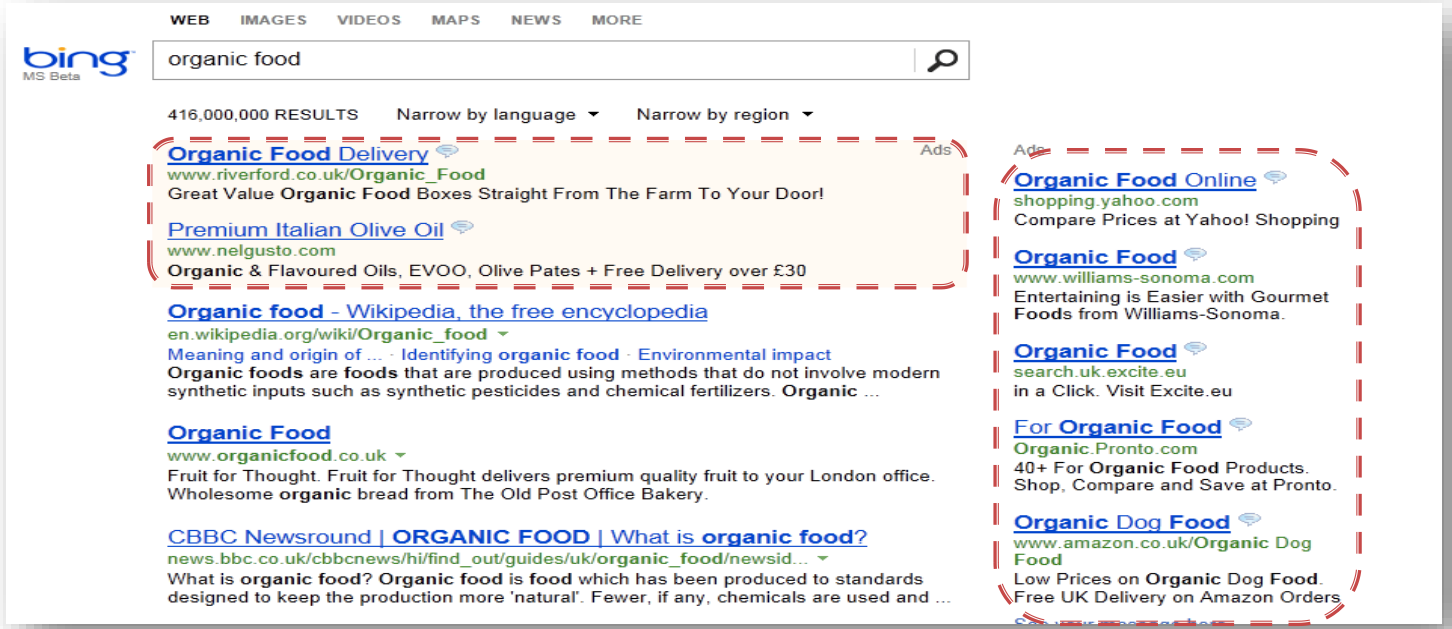
Mean Field Equilibria (MFE)

- A MFE distribution $\boldsymbol{f}$ is a fixed point of

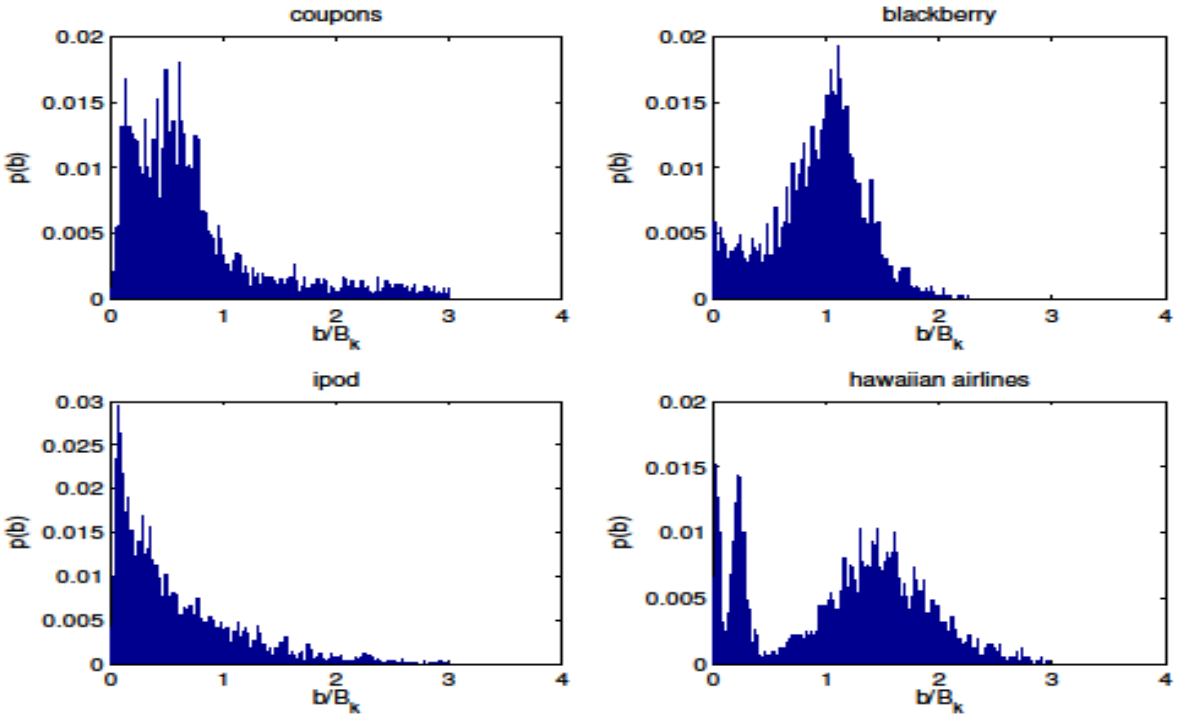
$$\boldsymbol{f} = \boldsymbol{\Phi}(\boldsymbol{f})$$

Example: bidding with budgets

Sponsored ad-auction



Bids and budgets



Traditional Approach: Game Theory

- Assume click-through weights are known
- Number of bidders is known
- Bidders solve a Nash equilibrium. Unrealistic!

			Bid 1	Bid 2	Bid 3	Bid 4
	Bid 1	Bid 1				
Bid 1	Bid 2	Bid 2				
Bid 2	Bid 3	Bid 3				
Bid 3	Bid 4	Bid 4				
Bid 4						

Bidding with budgets

- Budgets constrained bidders
- Bid to maximise return in repeated auction
- Equivalent to online stochastic knapsack

Bidding with budgets

- Single good, repeated auction
- User has budget B_i(t), value $v_i(t)$, bids $a_i(t)$
- Winner pays 2nd highest price $p(t)$
  - Private valuation $\sim v$ (observable)
  - Competing bid $\sim p$ (not observable)
- Decision variable is bid at time t: $a(t)$
  - Depends on both $v(t)$ and $B(t)$
- Gain per unit time: $g(t) = \big(v_i(t) - p(t)\big)\mathbf{1}_{a_i(t)>p(t)}$

Bidding with budgets

- Objective: maximise discounted return

$$\sum_{t=0}^{\infty} e^{-\beta t}\mathbb{E}[g(t)]$$

- Assume opponents bids $p \sim iid\ F$

- Large system scaling: Scale both "time" and "space" as $\beta \to 0$

$$V_\beta(B) \triangleq \beta v_\beta(b) = \beta v_\beta\left(B/\beta\right)$$

Main insights [Gummadi, K, Proutiere, 2012]

- *Shade* bid:

  shade(remaining balance B) $\triangleq \frac{1}{1+V'(B)}$

  so bid $\frac{v}{1+V'(B)}$

- $V(B)$ can be computed explicitly in the limit

- Extends other auctions SP, GSP, FP, etc. (any quasi linear utility)

Mean Field Games

- Large number *N* of competing bidders
- Divide into *K* classes, budget B_k, after $N \times T_k$ timeslots, budget re-initialised.
- Time *t*, select random subset $S_t \subset N$, size *S*

Definition: Restrict attentions to strategies that are the same modulo restarts. A *renewal* equilibrium, is an equilibrium in this setting.

Mean field equilibria

- As $N \to \infty$, distribution of opponents bid stationary
- Let z be dist. of highest $S-1$ bids
- $x_i(z) \triangleq$ optimal strategy of i,
- $\hat{x}_i(z) \triangleq$ empirical bid distribution of i,

A MFE exists if $z \sim Max[\hat{x}_i(z), \mathrm{i} = \pi(1), \ldots \pi(\mathrm{S}-1)]$

Theorem:

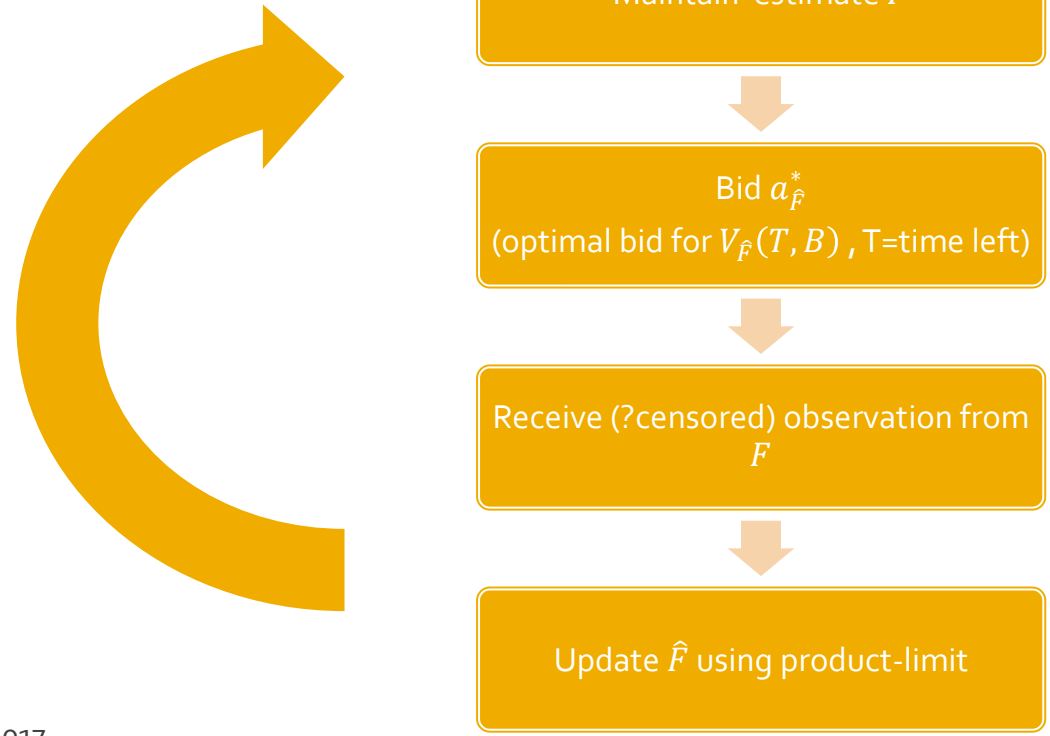
A MFE exists when there a lot of auctions per period.

Censored learning

[Amin, Kearns, Key & Schwaighofer 2012]

- But opponents bid distribution unknown?
- Censored observations
  1. Imperfect observation of market price
     - If win with bid *b*, know price *p*
     - If lose, know price *p*>*b*
  2. Only charged if add clicked
- Use Kaplan-Meier (product limit) as part of a "Greedy-Product Limit" Algorithm
  - Achieves $O\left(\sqrt{T}\right)$ regret (need assumptions on K-M convergence)

Greedy-Product limit Algorithm



Other algorithms (with Long-Than et al 2013)

- $\epsilon - First$
  - **Exploration:** For time ϵT bid randomly within budget and observe allocation
  - **Exploitation:**
    - Estimate distribution $\hat{F}$ using KM
    - Find optimal bids for remaining times and budgets
    - For time $(\epsilon T, \mathrm{T}]$ uses these to bid (no updates of $\hat{F}$)
- Achieves $O\left(T^{2/3}\right)$ regret

Experiments

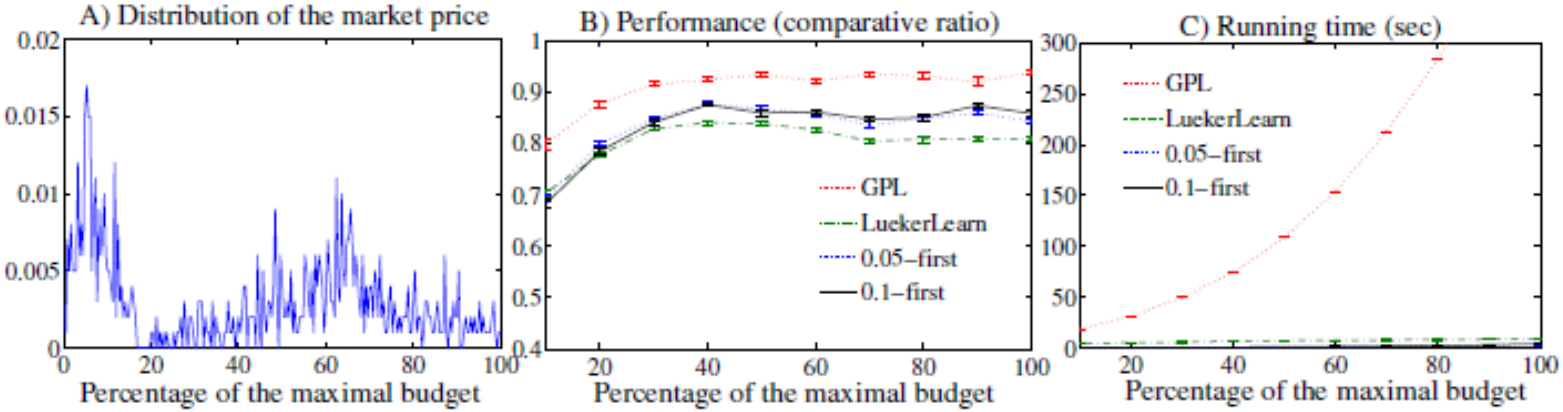
1. Distributional.
   - Construct distribution from empirical AdCenter distribution of bids.
   - Simulate model.
2. Sequential.
   - Violate Modelling Assumption.
   - Feed historical bids to algorithm sequentially.

Sequential: Compared to offline-optimal

Name	Competitive Ratio	St dev
Greed PL	**0.91**	**.117**
LuekerLearn	0.90	.115
Fixed-Price search	0.63	.140
Q-Learn	**0.59**	**.156**
Budget Smooth	0.31	.325

Performance vs. Complexity

- GPL performs well
- $\epsilon - First$ larger regret but reduced run time
  (eg, x 5 faster with 10% less accuracy)



Concluding remarks

- Mean field approaches/ Functional Laws of Large Numbers can simplify large systems
- Promising direction for learning in such systems
- Challenge:
  - Computable & credible strategies
  - Link to empirical studies of agents

References

- Facebook games
  - Pushmeet Kohli, Yoram Bachrach, David Stillwell, Michael Kearns, Ralf Herbrich, and Thore Graepel, Colonel Blotto On Facebook: The Effect of Social Relations On Strategic Interaction, in *ACM Web Sciences 2012*, ACM Conference on Web Sciences, 2012
- Mean Field Games
  - **Mean Field Games and Applications**, Olivier Guéant, J.-M. Lasry and P.-L. Lions, *in* Paris-Princeton Lectures on Mathematical Finance 2010, Ed. Springer, January 2011
  - S. Adlakha, R. Johari, G.Y. Weintraub (2011). Equilibria of dynamic games with many players: existence, approximation, and market structure. Submitted.
  - K. Iyer, R. Johari, M. Sundararajan (2012). Mean field equilibria of dynamic auctions with learning. Submitted. A preliminary version was presented at ACM Electronic Commerce (2011).
- Bidding with Budgets
  - Ramakrishna Gummadi, Peter Key, and Alexandre Proutiere, Optimal Bidding Strategies and Equilibria in Dynamic Auctions with Budget Constraints, 24 May 2012; http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2066175
  - Kareem Amin, Michael Kearns, Peter Key, and Anton Schwaighofer, Budget Optimization for Sponsored Search:Censored Learning in MDPs, in *Uncertainty in Artificial Intelligence (UAI)*, Uncertainty in Artificial Intelligence (UAI), August 2012