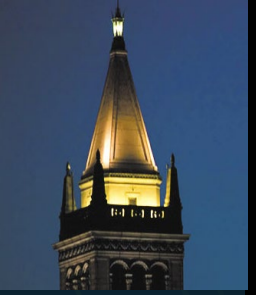
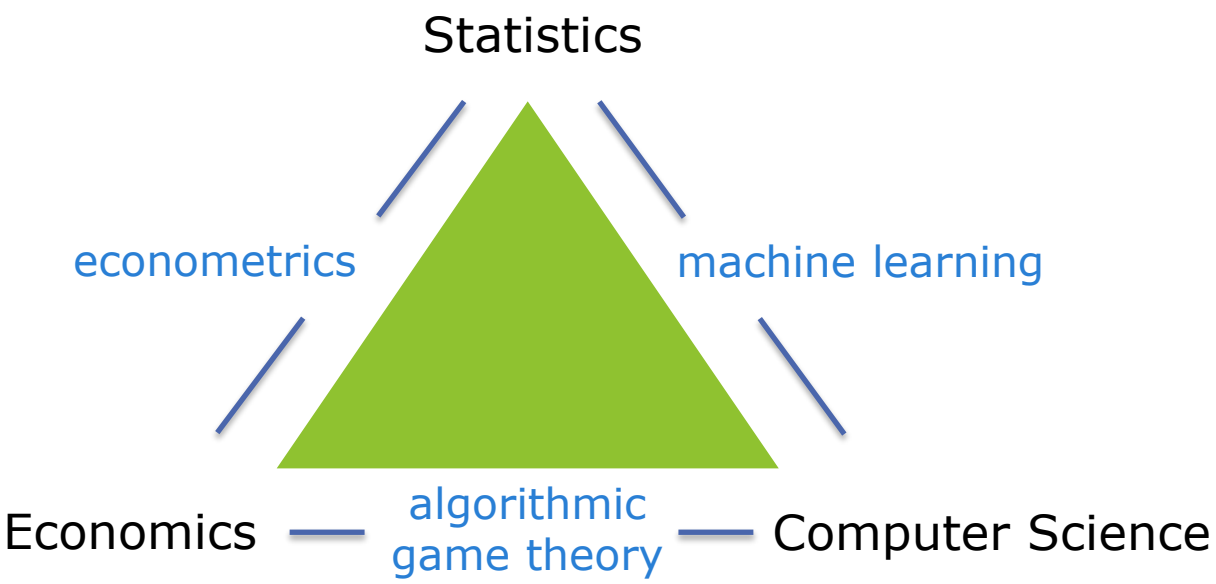


On Learning-Aware Mechanism Design

Michael Jordan
University of California, Berkeley

Three Foundational Disciplines



The Two Sides of Machine Learning

- The current era of machine learning has focused on pattern recognition
  - platforms such as TensorFlow and PyTorch have arisen to help turn pattern recognition into a commodity
- The decision-making side of machine learning will be a focus in the future
  - individual high-stake decisions
  - explanations for decisions, and dialog about decisions
  - multiple decisions
  - decisions in the context of multiple decision-makers
  - market mechanisms

Decisions

- It’s not just a matter of a threshold
- Real-world decisions with consequences
  - counterfactuals, provenance, relevance, dialog
- Sets of decisions across a network over time
  - streaming, asynchronous decisions
- Decisions when there is scarcity and competition
  - what counts as a “good decision” depends on what other decision-makers are doing, which is something that a good decision-maker will model
  - data itself can be scarce and under competitive pressure

Consider Classical Recommendation Systems

- A record is kept of each customer's purchases
- Customers are "similar" if they buy similar sets of items
- Items are "similar" are they are bought together by multiple customers
- Recommendations are made on the basis of these similarities
- These systems have become a commodity
- They are on the prediction side of ML

Multiple Decisions with Competition

- Suppose that recommending a certain movie is a good business decision (e.g., because it's very popular)
- Is it OK to recommend the same movie to everyone?
- Is it OK to recommend the same book to everyone?
- Is it OK to recommend the same restaurant to everyone?
- Is it OK to recommend the same street to every driver?
- Is it OK to recommend the same stock purchase to everyone?

The Alternative: Create a Market

- A two-way market between consumers and producers
  - based on recommendation systems on both sides
- E.g., diners are one side of the market, and restaurants on the other side
- E.g., drivers are one side of the market, and street segments on the other side

- This isn't just classical microeconomics; the use of recommendation systems, and thus statistics, is key

Music in the Data Age

- Use data to structure a two-sided market; e.g., by providing a dashboard to musicians, letting them learn where their audience is
- The musician can give shows where they have an audience
- I.e., consumers and producers become linked, and value flows: a market is created
  - the company that creates this market profits simply by taking a cut from the transactions

- The company *United Masters* is doing precisely this; www.unitedmasters.com

Some Problems at the Interface of ML and Econ

- Relationships among optima, equilibria, and dynamics
- Multi-way markets in which the individual agents need to explore to learn their preferences
- Large-scale multi-way markets in which agents view other sides of the market via recommendation systems
- Uncertainty quantification for black box and adversarial settings
- Mechanism design with learned preferences
- Contracts and statistical experiments
- Incentive-aware classification and evaluation
- Fairness, privacy, and social good

- The goal is to discover new principles to build healthy (e.g., fair) learning-based markets that are stabilized over long stretches of time

Outline

- Who Leads and Who Follows in Strategic Classification?
- Learning Equilibria in Matching Markets from Bandit Feedback
- Statistical Inference via Contract Theory
- Robust Learning of Optimal Auctions

Strategic Classification

Tijana Zrnic

Eric Mazumdar

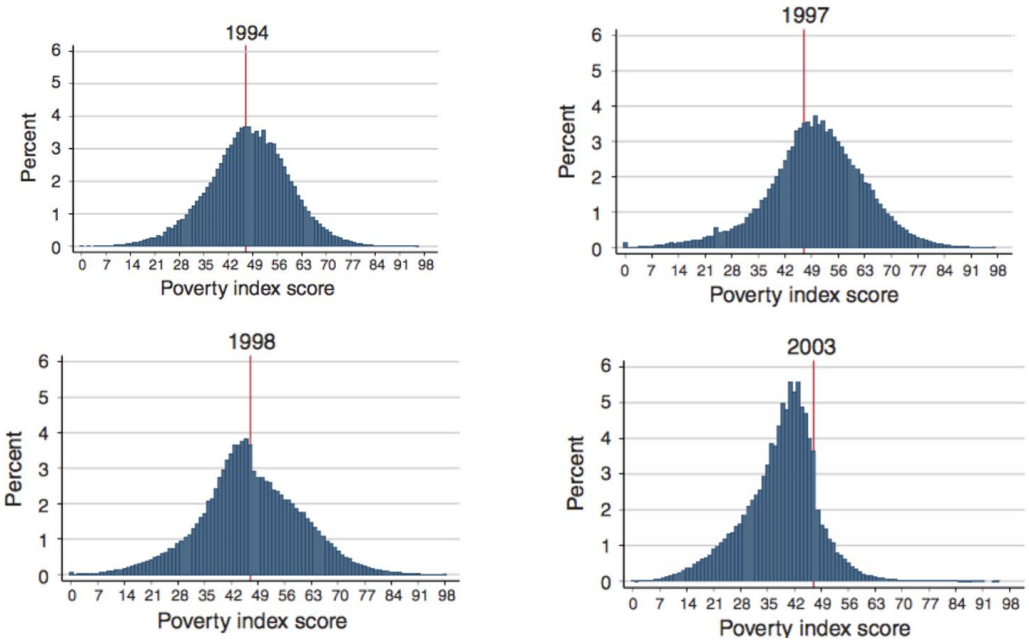
University of California, Berkeley

Decision-Making in the Face of Strategic Behavior

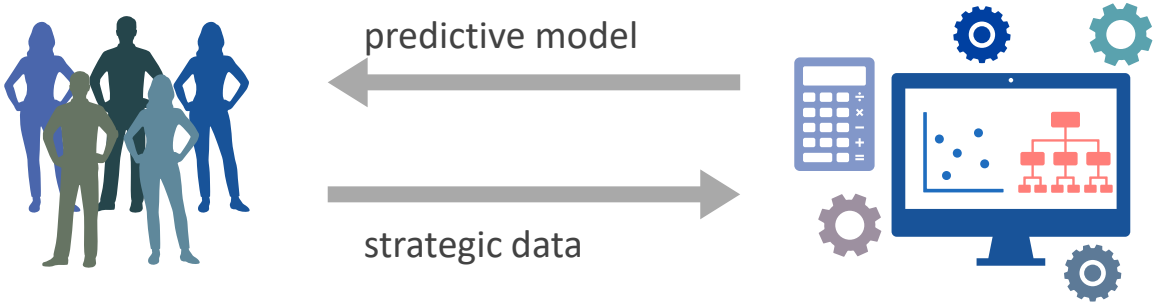
As predictive models are deployed in social settings, they must contend with strategic behaviors from people

Decision-Making in the Face of Strategic Behavior

As predictive models are deployed in social settings, they must contend with strategic behaviors from people



Feedback Loops in Learning



strategic agents

want favorable prediction

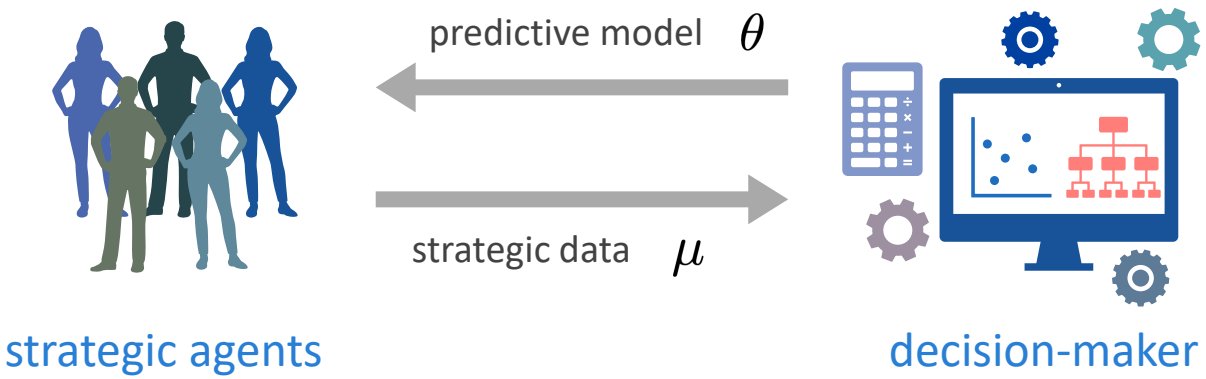
decision-maker

want to minimize prediction loss

Strategic agents and decision-maker **adapt to each other's actions**

What is the equilibrium solution and how is it achieved?

Stackelberg Games

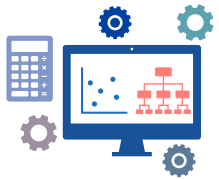


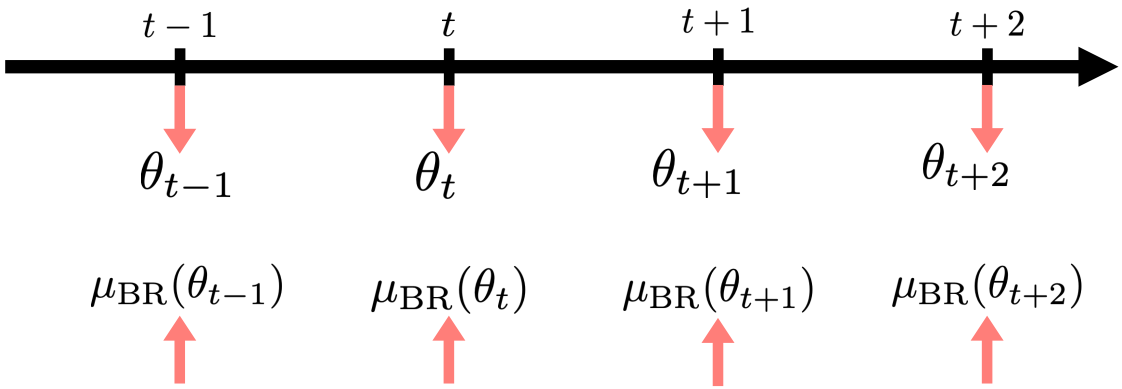
- We will model this as a Stackelberg game is a game where one player (*"leader")* moves first, and the other player (*"follower"*) moves second

Classically, the decision-maker is **assumed to be the leader**

Solution: Learning Dynamics

Decision-maker repeatedly interacts with the agents to find a Stackelberg equilibrium

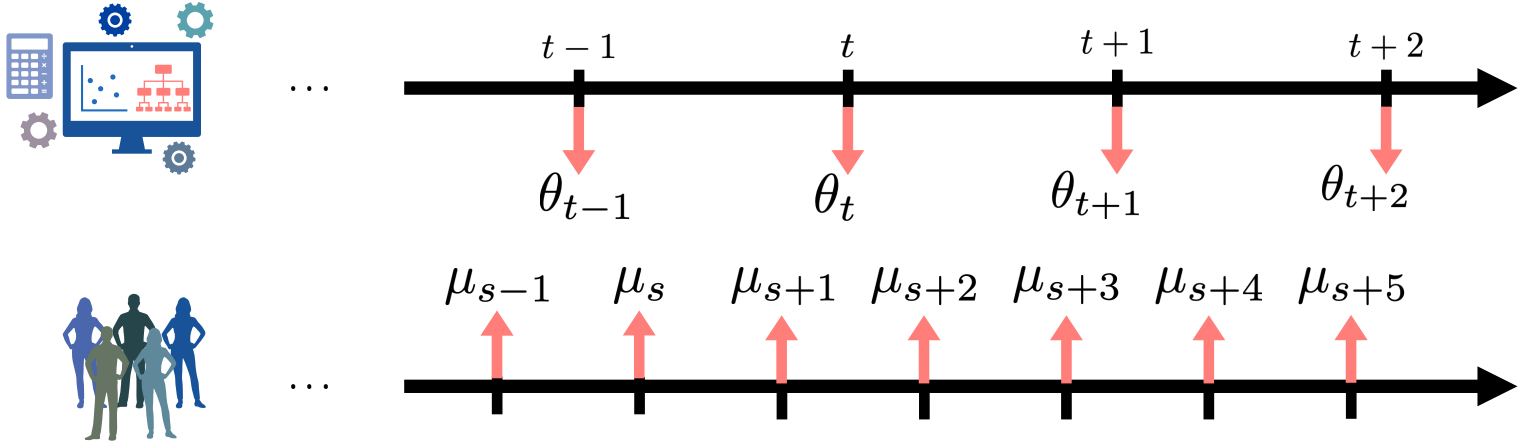




best-responds *instantaneously*

Decoupled Time Scales

We generalize the standard model to allow both players to gradually learn on their own timescale



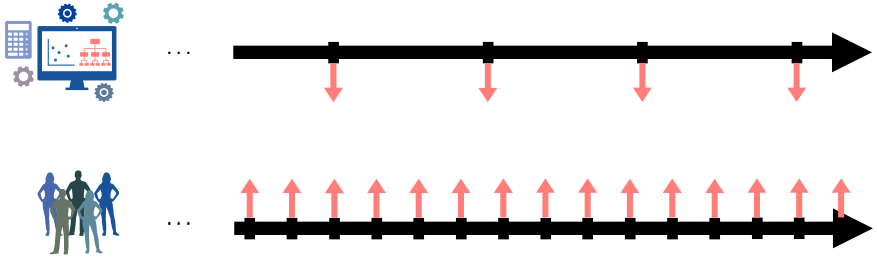
timescale $\approx$ update frequency

In such repeated interactions it is not always rational to play the best response!

Proactive and Reactive Decision-Makers

We focus on two relevant modes of relative timescales:

1. decision-maker is “slow” relative to agents



We call such decision-makers proactive

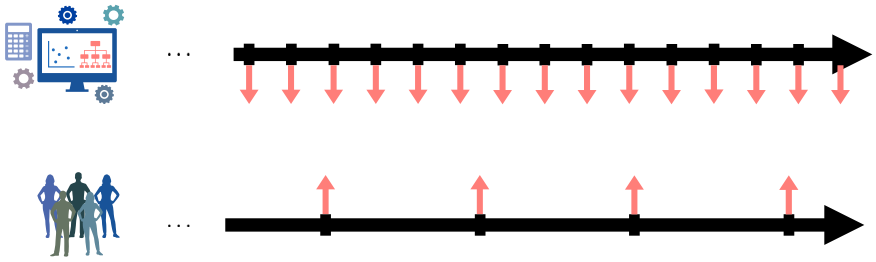
Example:

- college admissions, credit scoring

Proactive and Reactive Decision-Makers

We focus on two relevant modes of relative timescales:

2. decision-maker is “fast” relative to agents



We call such decision-makers **reactive**

Example:

- online platforms

YouTube

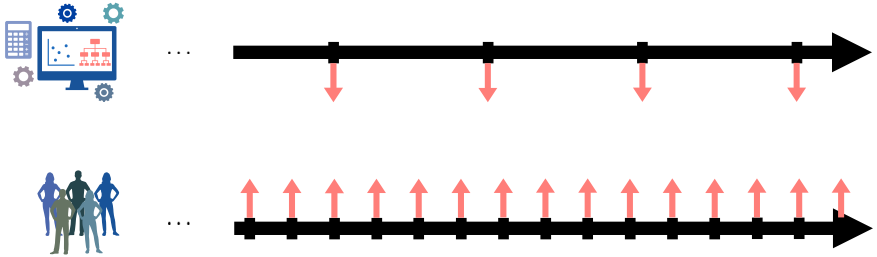
Instagram

Uber

Proactive and Reactive Decision-Makers

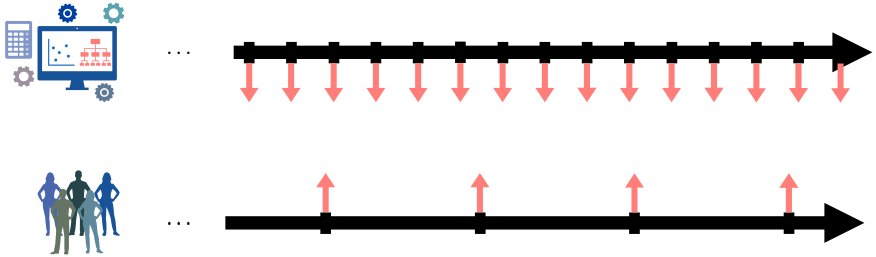
We focus on two relevant modes of relative timescales:

1. decision-maker is “slow” relative to agents



We call such decision-makers **proactive**

2. decision-maker is “fast” relative to agents



We call such decision-makers **reactive**

Decision-makers can often **choose** whether to be proactive or reactive

Results

Theorem 1 (informal)

By tuning their update frequency appropriately, the decision-maker can drive natural learning dynamics with rational strategic agents to a Stackelberg equilibrium **with either order of play**

Theorem 2 (informal)

In several standard statistical settings, **both** players prefer the equilibrium where the strategic agents **lead** and the decision-maker **follows**

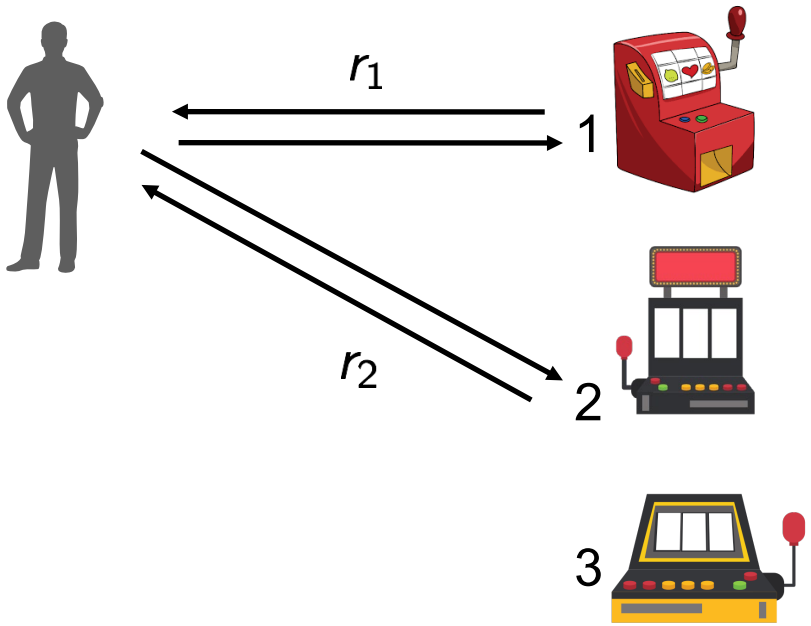
Competing Bandits in Matching Markets

Lydia Liu

Horia Mania

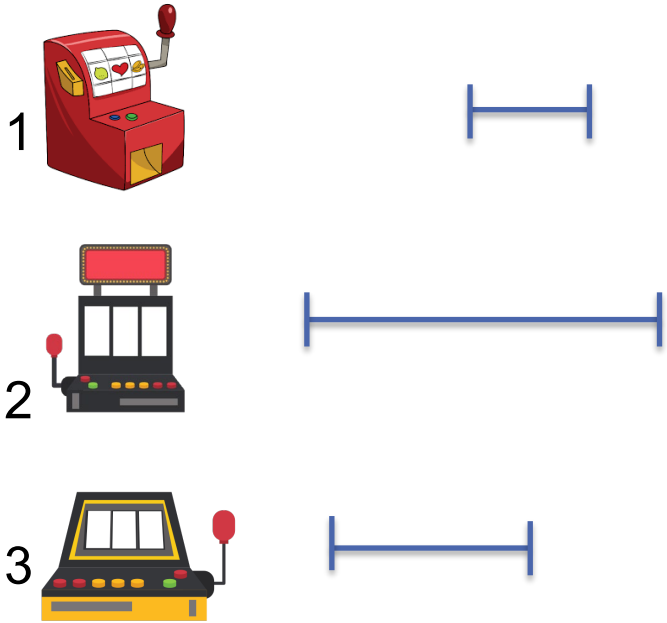
Multi-Armed Bandits

- MABs offer a natural platform to understand exploration / exploitation trade-offs



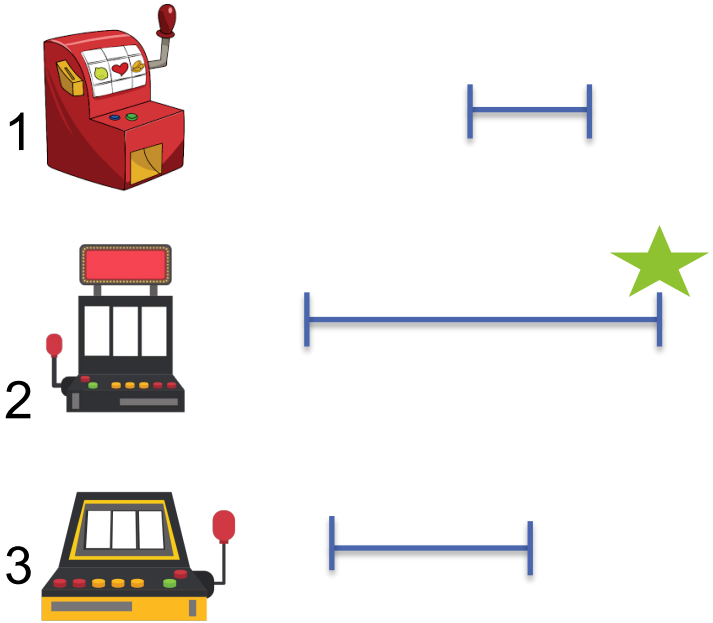
Upper Confidence Bound (UCB) Algorithm

- Maintain an upper confidence bound on reward values
- Pick the arm with the largest upper confidence bound



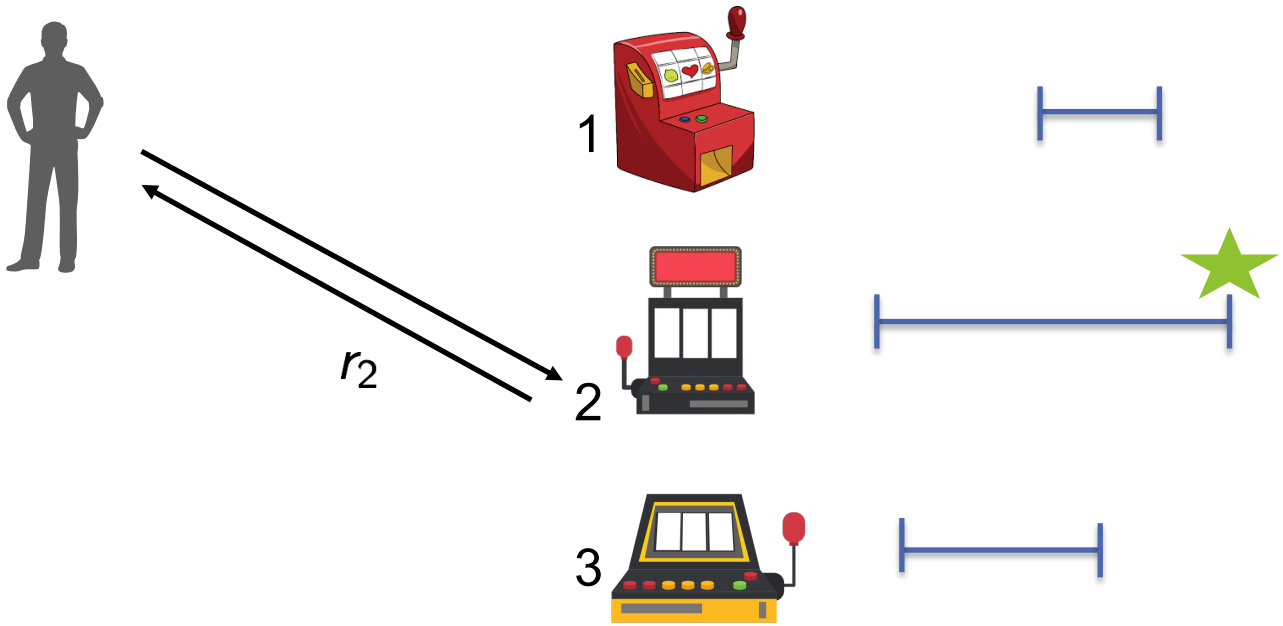
Upper Confidence Bound (UCB) Algorithm

- Maintain an upper confidence bound on reward values
- Pick the arm with the largest upper confidence bound



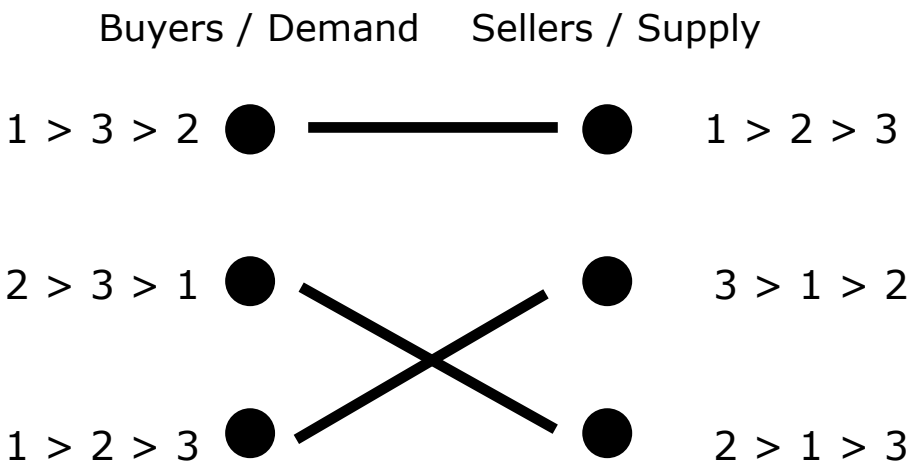
Upper Confidence Bound (UCB) Algorithm

- Maintain an upper confidence bound on reward values
- Pick the arm with the largest upper confidence bound



Matching Markets

Suppose we have a market in which the participants have preferences:



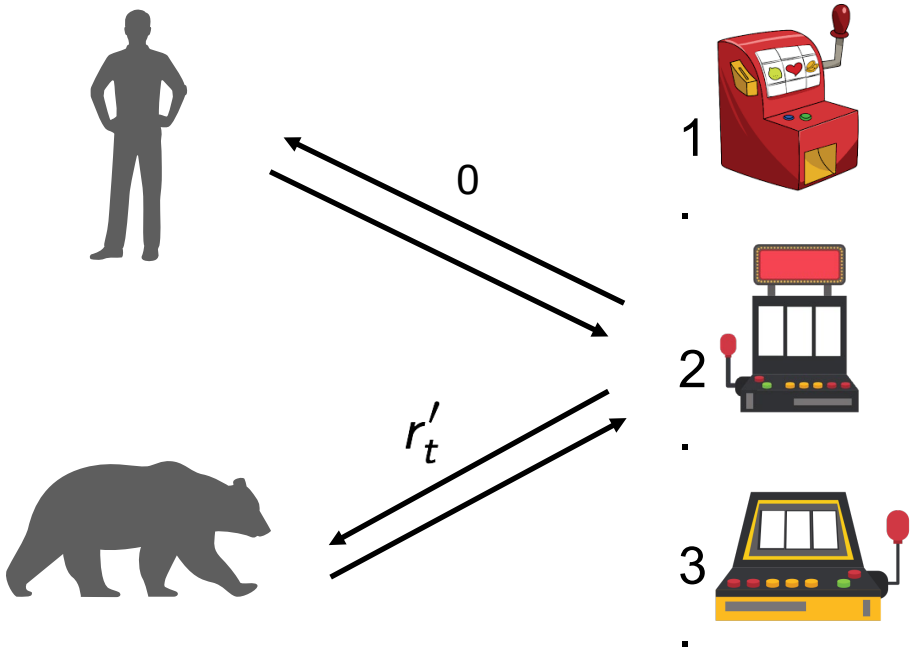
Gale and Shapley introduced this problem in 1962 and proposed a celebrated algorithm that always finds a stable match

Matching Markets Meet Bandit Learning

What if the participants in the market do not know their preferences a priori, but observe noisy utilities through repeated interactions?

Now the participants have an exploration/exploitation problem, in the context of other participants

Competing Agents



Bandit Markets

- We conceive of a bandit market: agents on one side, arms on the other side.

Agents get noisy rewards when they pull arms.

Arms have preferences over agents (these preferences can also express agents' skill levels)

When multiple agents pull the same arm only the most preferred agent gets a reward.

Regret in Bandit Markets

Then it is natural to define the regret of agent i up to time n as:

$$R_i(n) = \underbrace{n\mu_i(m(i))}_{\text{Mean reward of stable match}} - \sum_{t=1}^{n} \underbrace{\mathbb{E}X_{i,m_t}(t)}_{\text{Reward at time t}}$$

Minimizing this regret is natural. It says that agents should expect rewards as good as their stable match in hindsight.

Regret-Minimizing Algorithm

Gale-Shapley upper confidence bounds (GS-UCB):

- Agents rank arms according to upper confidence bounds for the mean rewards.
- Agents submit rankings to a matching platform.
- The platform uses these rankings to run the Gale-Shapley algorithm to match agents and arms.
- Agents receive rewards and update upper confidence bounds.
- Repeat.

Theorem

Theorem (informal): If there are N agents and K arms and GS-UCB is run, the regret of agent i satisfies

$$R_i(n) = \mathcal{O}\left(\frac{NK\log(n)}{\boxed{\Delta^2}}\right)$$

Reward gap of possibly other agents.

- In other words, if the bear decides to explore more, the human might have higher regret.
- See paper for refinements of this bound and further discussion of exploration-exploitation trade-offs in this setting.
- Finally, we note that GS-UCB is incentive compatible. No single agent has an incentive to deviate from the method.

Learning Equilibria in Matching Markets from Bandit Feedback

Meena Jagadeesan Alex Wei Yixin Wang Jacob Steinhardt

University of California, Berkeley

Equilibria in Matching Markets

- Large-scale two-sided matching platforms must find market outcomes that align with user preferences while simultaneously learning these preferences from data
- We achieve this with transferable utilities (the Shapley-Shubik model), where the platform both selects a matching and sets monetary transfers between agents
- We introduce a novel measure of instability, which we define to be the minimum amount the platform could subsidize agents to achieve stability

A Quantitative Measure of Instability

- Our measure of instability (the minimum amount the platform could subsidize agents to achieve stability):

$$\max_{S \subseteq \mathcal{A}} \left[\left(\max_{X' \in \mathcal{X}_S} \sum_{a \in S} u_a(\mu_{X'}(a)) \right) - \left(\sum_{a \in S} u_a(\mu_X(a)) + \tau_a \right) \right]$$

- This measure can be optimized directly or can be the source of further relaxations or duality transformations

Learning Equilibria in Matching Markets from Bandit Feedback

- Using our new stability measure as a loss function, we design low-regret algorithms for learning stable matchings
  - from noisy user feedback in a multi-armed bandit model
- Recall that “optimism in the face of uncertainty” is the core principle underlying multi-armed bandit algorithms
- Our algorithmic insight is that the optimism principle applies to a primal-dual formulation of matching with transfers and leads to near-optimal regret bounds

Statistical Contract Theory

Stephen Bates

Michael Sklar

Jake Soloff

Contract Theory

principal

- Has only partial knowledge
- Must incentivize the agents

agent

- Has private information
- Strategic and self-interested

This talk: **Contract Theory meets Neyman-Pearson**

Clinical Trials

Average Cost of Clinical Trial

Department of Health and Human Services, 2014

Therapeutic Area	Phase 1	Phase 2	Phase 3
Anti-Infective	$4.2 (5)	$14.2 (6)	$22.8 (5)
Cardiovascular	$2.2 (9)	$7.0 (13)	$25.2 (3)
Central Nervous System	$3.9 (6)	$13.9 (7)	$19.2 (7)
Dermatology	$1.8 (10)	$8.9 (12)	$11.5 (13)
Endocrine	$1.4 (12)	$12.1 (10)	$17.0 (9)
Gastrointestinal	$2.4 (8)	$15.8 (4)	$14.5 (11)
Genitourinary System	$3.1 (7)	$14.6 (5)	$17.5 (8)
Hematology	$1.7 (11)	$19.6 (1)	$15.0 (10)
Immunomodulation	$6.6 (1)	$16.0 (3)	$11.9 (12)
Oncology	$4.5 (4)	$11.2 (11)	$22.1 (6)
Ophthalmology	$5.3 (2)	$13.8 (8)	$30.7 (2)
Pain and Anesthesia	$1.4 (13)	$17.0 (2)	$52.9 (1)
Respiratory System	$5.2 (3)	$12.2 (9)	$23.1 (4)

(in millions of dollars)

Immense social investment in clinical trials

How Should the FDA Test?

	type	P(approve)	P(non-approve)	
bad drugs	$\theta = 0$	0.05	0.95	(5% type-1 error)
good drugs	$\theta = 1$	0.80	0.20	(80% power)

Is this a good statistical protocol?

Case 1: small profit. $20 million cost to run trial. $200 million if approved.

$\mathbb{E}[\text{profit}|\theta = 0] =$ -$10 million

All approvals are good drugs!

Case 2: large profit. $20 million cost to run trial. $2 billion if approved.

$\mathbb{E}[\text{profit}|\theta = 0] =$ $80 million

Many bad approved drugs!

Statistical Contract Theory

Denote the agent's private information as $\theta \in \Theta$

Present the agent with the following opt-in protocol:

our task:
design this menu

1. Agent pays R
2. Agent chooses payout function f from menu $\mathcal{F}$
3. Statistical trial yields random variable $Z \sim P_\theta$
4. Agent receives payoff $f(Z)$
   Principal receives utility $u(\theta, f(Z))$

$f : \mathcal{Z} \to \mathbb{R}_+$

$f \in \mathcal{F}$

$Z \in \mathcal{Z}$

Agent acts to maximize their payoff: $f^{\text{br}} = \underset{f \in \mathcal{F}}{\text{argmax}} \mathbb{E}_{Z \sim P_\theta}[f(Z)]$

Incentive Alignment

bad (null) agents: $\Theta_0 \subset \Theta$ $\quad u(\theta_0, f(Z)) \leq 0$, decreasing in $f(Z)$ for $\theta_0 \in \Theta_0$

good (nonnull) agents: $\Theta \backslash \Theta_0$ $\quad u(\theta_1, f(Z)) \geq 0$, increasing in $f(Z)$ for $\theta_1 \notin \Theta_0$

The principal wants to transact as much as possible with good agents

Definition (Incentive-aligned contract)

A menu $\mathcal{F}$ is *incentive-aligned* if for all $f \in \mathcal{F}$ and $\theta_0 \in \Theta_0$

$$\mathbb{E}_{Z \sim P_{\theta_0}}[f(Z) - R] \leq 0$$

agent's expected profit

note: $p \leq .05$ protocol <u>not</u> incentive aligned

On average, null drugs are not profitable, so null agents are incentized to drop out

E-values: Statistical Evidence on the Right Scale

Definition (E-value)

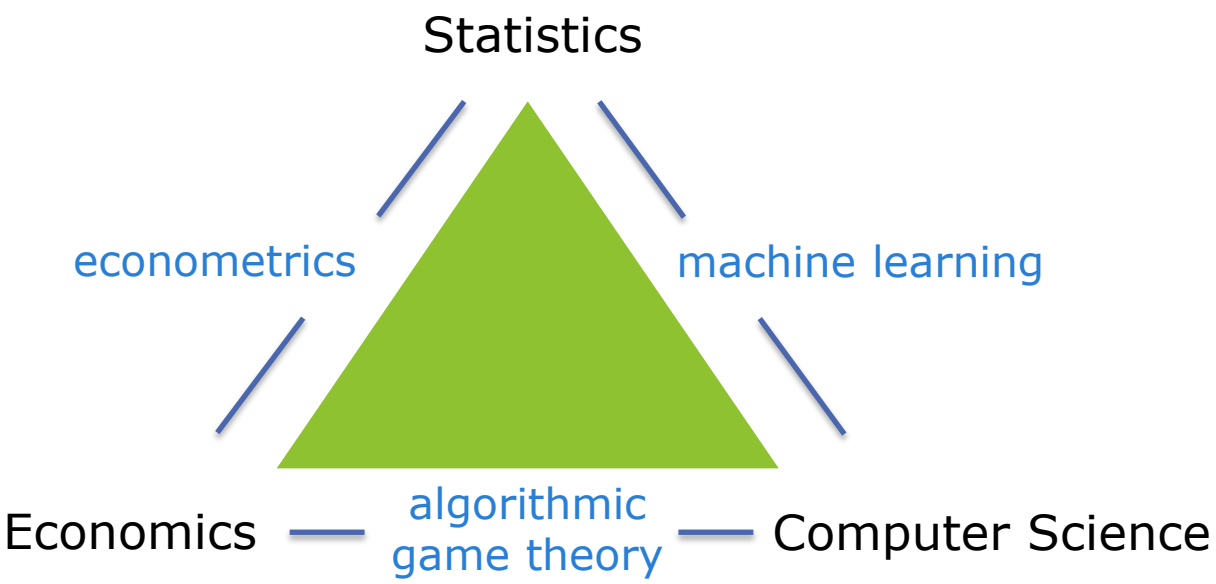
A random variable $X \geq 0$ is an *E-value* for null hypothesis Θ_0 if for all $\theta_0 \in \Theta_0$

$$\mathbb{E}_{Z \sim P_{\theta_0}}[X] \leq 1$$

Theorem

A contract is incentive-aligned if and only if all payoff functions are E-values.

Three Foundational Disciplines



A Personal View on "Machine Learning" or "AI": It is the Emergence of a New Engineering Field

- Cf. chemical engineering in the 40s and 50s
  - built on chemistry, fluid mechanics, etc
  - driven by the possibility of building chemical factories
  - new concepts and mathematical principles were needed
- Cf. electrical engineering at the turn of the last century
  - built on electromagnetism, optics, etc
  - new concepts and mathematical principles were needed
- The new field builds on inferential ideas, algorithmic ideas, and economic ideas from the past three centuries
  - what's fundamentally new is the idea of building large-scale systems based on these ideas, using data flows at planetary scale