Competing Bandits in Matching Markets



Lydia T. Liu Horia Mania Michael I. Jordan

University of California, Berkeley

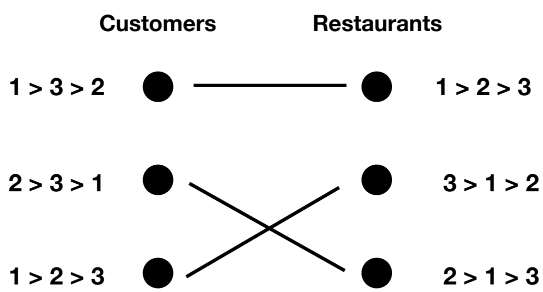
Matching Markets are Ubiquitous

Two sides of the market must be matched.
Each side has constraints: capacity, preferences, etc.

Residents		Hospitals
Students		High schools
Customers	**Modern matching markets: repeated interaction via online platforms**	Restaurants
Jobs		Job candidates

Matching Markets are Ubiquitous

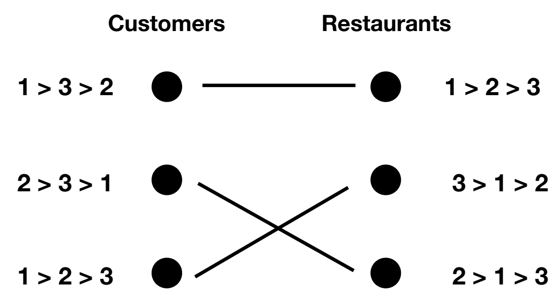
A market in which the participants have preferences over the other side:



We want to find a <u>**stable**</u> matching: no two participants prefer to be matched with each other over their respective matches.

Matching Markets are Ubiquitous

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

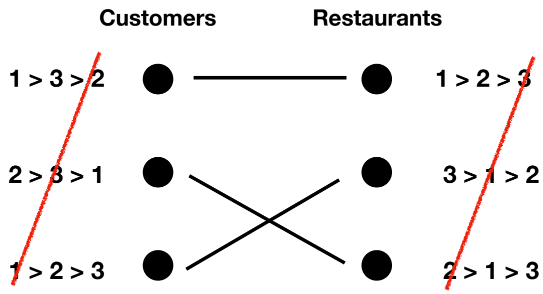


Gale and Shapley (1962) proposed the deferred acceptance algorithm that always finds a stable matching.

In this algorithm one side of the market iteratively makes proposals to the other side

Matching Markets are Ubiquitous

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

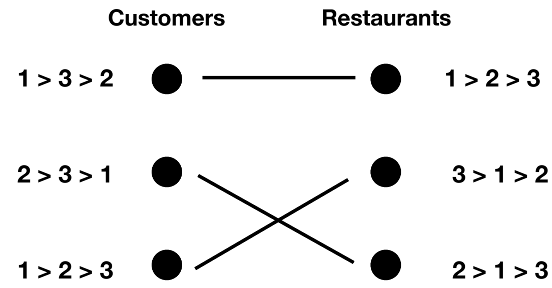


What if the participants in the market do not know their preferences a priori, but observe noisy utilities through repeated interactions?

Examples: restaurants and customers, online labor markets, load balancing in data centers.

Matching Markets are Ubiquitous

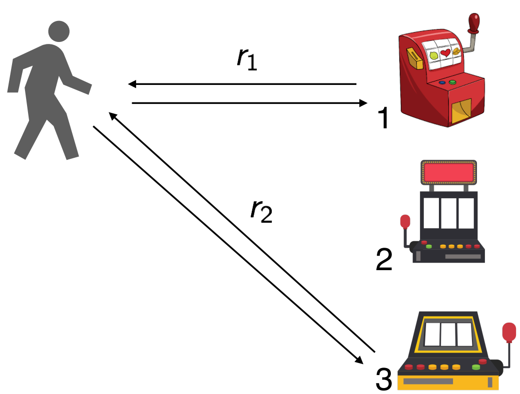
Suppose we have a market in which the participants have preferences:



Then exploration + exploitation is needed in the market!

Multi-Armed Bandits

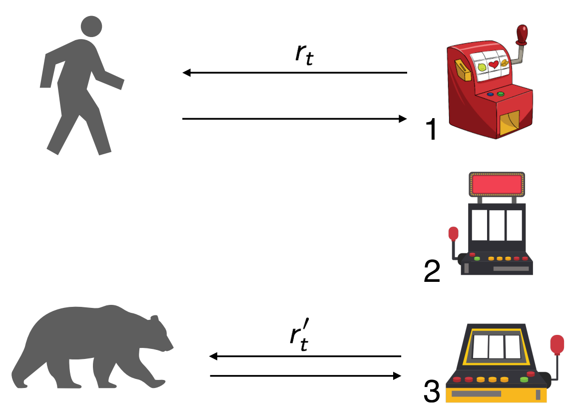
MAB provides a natural framework to understand exploration / exploitation trade-offs.



Arms = actions with stochastic rewards, mean μ_i

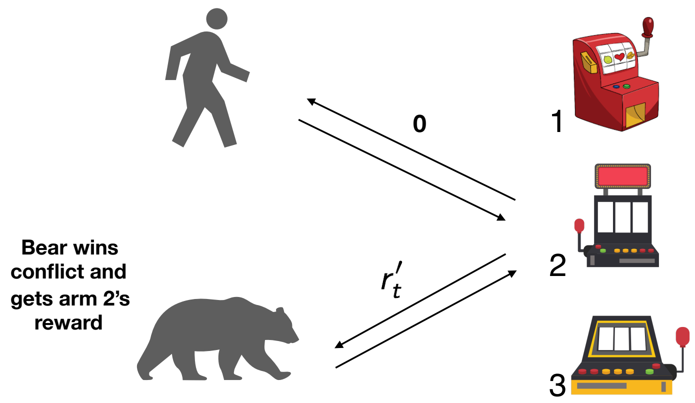
Multi-Armed Bandits

Let's add a competing player!



Multi-Armed Bandits

Let's add a competing player!



Competing Bandits in Matching Markets

In summary: we consider a bandits market with agents on one side, arms on the other.

Agents get noisy rewards when they pull arms. Same arm has different mean reward for different agents.

Arms have known 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 (competition)

Competing Bandits in Matching Markets

Define the ***stable regret*** of agent ***i*** up to time ***n*** as:

$$R_i(n) = n\mu_i(m(i)) - \sum_{t=1}^{n} \mathbb{E}X_{i,m_t}(t)$$

Mean reward of stable match

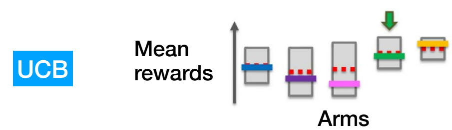
Reward at time t

This is a natural regret notion because in hindsight, agents should expect rewards as good as their stable match.

If there are multiple stable matches, a bit more care is needed. See more at our poster.

Competing Bandits in Matching Markets

Algorithm: Gale-Shapley upper confidence bound (GS-UCB)
Avoids having players conflict on the same arms, and minimizes regret of all agents



1. Agents rank arms by the UCBs of the mean rewards.
2. Agents submit rankings to a matching platform.
3. The platform runs the Gale-Shapley algorithm to match agents and arms.
4. Agents receive rewards and update UCBs.
5. Repeat.

Competing Bandits in Matching Markets

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

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

Minimum gap of arms' rewards for all agents.

In other words, if the bear has 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 submitting UCBs to platform.