Monotone Multi-Armed Bandit Allocation Rules

Alex Slivkins
Microsoft Research Silicon Valley

COLT 2011
Multi-armed bandits (MAB)

- In each round, select among K “arms”, collects a reward
- Rewards are fixed in advance, but not revealed
- Goal: maximize total reward over time

Realization (of the rewards): table whose \((i,t)\)-th entry is the reward of arm \(i\) in round \(t\), if this arm is chosen.

- Realization is generated by a random process
  - in some known set of “allowed” processes

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>.</th>
<th>.</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>.</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>K</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
MAB allocation rules

- MAB allocation rule:
  - Input a vector of bids: bid $b_i$ for each arm $i$.
    Run MAB algorithm, collect rewards (raw rewards).
    Scale raw rewards from each arm $i$ by factor $b_i$.
  - Motivation: arms are ads ("Pay Per Click")
    - Each agent (advertiser) comes with one ad.
      In each round one ad is shown to a user.
      Each time ad $i$ is clicked, agent $i$ receives value $b_i$.
    - Raw rewards are clicks. Click probabilities are not known.
      Value created = total reward of the MAB allocation rule
MAB auctions

Each agent $i$ submits bid $b_i$. MAB allocation rule is run. Payments are assigned.

- The issue of incentives
  - each agent’s value-per-click is private info (not revealed)
  - agents can lie about their values if it benefits them, so they need to be incentivized to tell the truth.
- Auction is truthful if for each agent, truth-telling is no worse than lying, no matter what others do.

Devanur, Kakade EC’09
Babaioff, Sharma, Slivkins EC’09
Babaioff, Kleinberg, Slivkins EC’10
Monotone MAB allocation rules

- MAB allocation rule can be extended to truthful auction \( \iff \) it is monotone: increasing any bid \( b_i \) (fixing other bids) can only increase the total raw reward from arm \( i \).

**Problem:** For a given MAB setting, design *monotone* MAB allocation rules

- Two versions: - for each realization of the rewards - in expectation over realization (clicks)
Status of the problem

- Stochastic rewards: problem solved
  - raw reward from arm $i$ is an IID sample from distribution $D_i$
  
- UCB1 is monotone in expectation over realization (clicks)
  - UCB1 is not “monotone for each realization”, but a more sophisticated algorithm is, with same regret

- Next target: adversarial rewards
  - there is a monotone MAB allocation rule with regret $n^{2/3}$
  - how about optimal regret $n^{1/2}$?

- Ask this question about your favorite MAB setting