Cooperative and Stochastic Multi-Player Multi-Armed Bandit: Optimal Regret With Neither Communication Nor Collisions

Sébastien Bubeck (Microsoft), Thomas Budzinski (ENS Lyon), Mark Sellke (Stanford): COLT 2021

① The Setup

- Problem: K -armed stochastic bandits with **multiple players**.
- K arms with reward probabilities $p_1, \dots, p_K$.
- At time $t = 1,2, \dots, T$ each of $m \le K$ players choose an arm.
- Players **cannot communicate.**
- Colliding (choosing the same arm) gives **no reward**.
- Goal: low expected regret R_T compared to top m arms.
- Motivation: cognitive radio. Arms ≈ channels.

  [Lai-Jiang-Poor 08, Liu-Zhao 10 Anandkumar-Michael-Tang-Swami 11].

② Some Existing Work

- $R_T \le \tilde{O}\left(\sqrt{T}\right)$ when $p_i \le 1 - \varepsilon$ [Lugosi-Mehrabian 18].
- $R_T \le O\left(\sum_{i=1}^{K} \frac{\log T}{\Delta_i}\right)$ [Huang-Combes-Trinh 21].
- Adversarial losses [BLPS 20]
  - Adaptive losses, no shared randomness: $R_T \ge \Omega(T)$.
  - $R_T \le \tilde{O}\left(\sqrt{T}\right)$ if collisions are explicitly announced.
  - $R_T \le \tilde{O}\left(T^{1-\frac{1}{2m}}\right)$, only losses observed.
- Theme: collisions allow **implicit communication.**
  - Relies on **observing** the lack of reward from collision.

③ Main Result: Optimal Regret with No Collisions At All

Theorem: using public shared randomness, can achieve

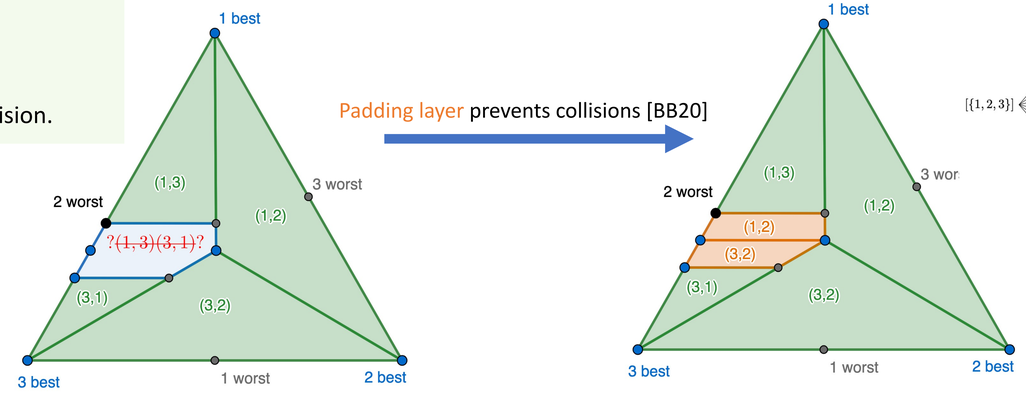
$$R_T \le O\left(mK^{5.5}\sqrt{T \log T}\right);$$

$$\mathbb{P}[\text{there is a collision}] \le O(1/T^2).$$

[BB 20]: for $(K, m) = (3,2)$, $R_T = \Theta\left(\sqrt{T \log T}\right)$ is optimal.

④ Topological Obstruction and Padding Layer for $(K, m) = (3,2)$

- Assume full-feedback. (Bandit uses similar construction, but more technical.)
- Naïve Goal: always play top 2 actions of empirical estimates.
- Naïve Strategy: everywhere in "state space", pre-label player-action matching.
- Problem: going "around a circle" leads to inevitable collision.
- Fix [BB20]: add "padding layer" as buffer. Random location ⇒ little harm.
- Player estimates close ⇒ use adjacent regions ⇒ no collision (with good labels).



⑤ New Ingredient: A Partition for General (K, m)

- Main difficulty: need a complicated generalization of the partition.
- Idea: define **tree of regions** by **ordered** set of **inequalities** between arm means.
- Separate arms that *might* be top m until top m vs bottom $K - m$ determined.
- Example for $(K, m) = (10,5)$:

  Start from the root with no inequalities

  $\{1,2,3,4,5,6,7,8,9,10\}$

  $\to \{1,2,3,4,5,6,8\} >_1 \{7,9,10\}$

  $\to \{1,3,5\} >_2 \{2,4,6,8\} >_1 \{7,9,10\}$

  $\to \{1,3,5\} >_2 \{4,8\} >_3 \{2,6\} >_1 \{7,9,10\}$

- Main regions have top m determined. These are leaves in the tree.
- Generalized padding layers used for separation if:
  1. There is a near-tie for which inequality to add next.
  2. We are close to a previous layer of padding.