Adversarial Bandits with Knapsacks

Nicole Immorlica\textsuperscript{1}, Karthik Abinav Sankararaman\textsuperscript{2}, Robert Schapire\textsuperscript{1}, Aleksandrs Slivkins\textsuperscript{1}

\textsuperscript{1}Microsoft Research, New York City \hspace{1cm} \textsuperscript{2}The University of Maryland, College Park

Overview

- BwK: general model for multi-armed bandits with resource consumption
- First algorithm for Adversarial BwK, matching lower bound.
- Subroutine: new algorithm for Stochastic BwK, with much simpler analysis.
- Modular algorithm \(\Rightarrow\) several extensions.

Motivating Examples

- Dynamic Pricing/Auctions:
  - \(d\) products, limited supply of each.
  - Seller adjusts prices (resp., auction params) over time to maximize total revenue
- Crowdsourcing markets:
  - Many similar tasks, limited budget.
  - Contractor dynamically adjusts wages to maximize #completed tasks (extension: \(d\) types of tasks, budget for each)
- Many more examples in prior work.

Prior Work on Stochastic BwK

- Special cases: Badanidiyuru + ‘12; Babaioff+ ‘12; Tran-Thanh+ ‘12; Krause & Singla ‘13; Ding+ ‘13; ...
- BwK: model & optimal algorithm:
  - Badanidiyuru, Kleinberg, Slivkins ‘13
- Extensions: Agrawal & Devanur ‘14 ‘16; Badanidiyuru, Langford, Slivkins ‘14; Agrawal, Devanur, Li ‘16; Sankararaman & Slivkins ‘18

Simultaneous work on Adversarial BwK: special cases with \(\text{ratio} = 1\) (ask us)

BwK: General Framework

- \(K\) arms, \(d\) resources, budgets \(B_1, \ldots, B_d\)
- In each round \(t \in [T]\):
  - Choose arm \(a_t \in [K]\)
  - Observe outcome vector \(\alpha_t(a_t) \in [0,1]^d\):
    - reward \(r_t\), consumption \(c_{t,j} \forall j \in [d]\)
  - Stop, if some resource runs out of budget

Goal: Maximize the total reward.

Outcome matrix: \(M_t = (\alpha_t(a) : \text{arms } a \in [K])\).

Stochastic BwK: \(M_t\) chosen IID.

Adversarial BwK: \(M_t\) chosen adversarially.

WLOG rescale s.t. all budgets are \(B = \min_j B_j\).

Benchmark

- \(\text{OPT}\) = best fixed distribution over arms (can be \(d\) times better than best fixed arm)
- \(E\left[\text{REW}\right] \geq \frac{\text{OPT}}{\text{ratio}} - \text{regret}\).
- \(\text{REW}\) = algorithm’s total reward

Lower bound for Adversarial BwK

Simple construction for \(\text{ratio} \geq \frac{1}{2}\):

- 2 arms, 1 resource, \(B = T/2\)
- Arm 1: consumption 1 in each round.
- Arm 2: 0 reward, 0 consumption.

Rew. for Arm 1

<table>
<thead>
<tr>
<th>t ∈ [1, T/2]</th>
<th>t ∈ (T/2, T]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instance 1</td>
<td>Low</td>
</tr>
<tr>
<td>Instance 2</td>
<td>Medium</td>
</tr>
</tbody>
</table>

More nuanced construction \(\Rightarrow\) \(\text{ratio} \geq \Omega(\log T)\).

Main Algorithm (MAIN)

- Two adversarial online learning algorithms:
  - (1) \(ALG_1\) for bandit feedback (e.g., EXP3.P)
  - (2) \(ALG_2\) for full-feedback (e.g., Hedge)

Two repeated zero-sum games.

In each round \(t \in [T]\):

- Simultaneously: \(ALG_1\) picks arm \(a_t \in [K]\), \(ALG_2\) picks resource \(j_t \in [d]\).
- Outcome vector \(\alpha_t(a_t)\) is observed.
- Reward for \(ALG_1\), cost for \(ALG_2\):
  \(L_t(a_t, j_t) = r_t + 1 - \frac{T}{d} c_{t,j_t}\)
- \(T_0 = T\) for Stochastic BwK, parameter othew.

Stochastic BwK

\(\mathbb{E}[L_t]\) is the Lagrangian for linear relaxation

\[
\begin{align*}
\text{maximize} & \quad T \cdot \sum_{a \in [K]} X(a) \mathbb{E}[r_t(a)] \\
\text{s.t.} & \quad \sum_{a \in [K]} X(a) = 1 \\
& \quad \forall j \in [d], \quad \sum_{a \in [K]} X(a) \mathbb{E}[c_{t,j}(a)] \leq B/T \\
& \quad \forall a \in [K], \quad 0 \leq X(a) \leq 1.
\end{align*}
\]

Proof Sketch: Use facts from prior work:

\(F_1\): \(\text{OPT} \leq \text{LP}\)-value.

\(F_2\): Minimax Lagrangian \(\Rightarrow\) Nash equilibrium.

\(F_3\): Average play \(\Rightarrow\) Nash equilibrium.

Adversarial BwK

\(\text{ALG}\), for \(X\) bandits \(\Rightarrow\) MAIN for \(X\) BwK, where \(X = \{\text{contextual, semi-}, \text{convex}\}\).

- No new research needed.
- Stochastic BwK: each extension was a paper (with slightly stronger regret bounds)
- Adversarial BwK: all results new.

Caveat: need \(\text{ALG}\), to have high-probability regret bound vs. adaptive adversary.

Extensions