Dueling Bandits with Team Comparisons

Lee Cohen[1] Ulrike Schmidt-Kraepelin[2] Yishay Mansour[1,3]

NeurIPS 2021

[1] Tel Aviv University [2] TU Berlin [3] Google Research

European Research Council
Established by the European Commission

Dueling Bandits

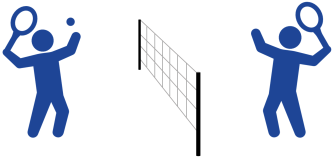
A set of **players** (arms) $[n] = \{1, \ldots, n\}$.

A learner observes noisy comparisons of pairs of **players** from $[n]$.

A **winning probability** matrix P that holds $P_{a,b} := P(a > b)$.

Common assumptions:

- **Total order** $\succ$ over players such that: For two players $a \succ b \Leftrightarrow P_{a,b} > 0.5$.
- The matrix P satisfies **strong stochastic transitivity (SST)**:
  For $a \succ b \succ c$ it holds that $P_{a,c} \geq \max\{P_{a,b}, P_{b,c}\}$



$$P = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 0.5 & 0.6 & 0.6 \\ 0.4 & 0.5 & 0.6 \\ 0.4 & 0.4 & 0.5 \end{pmatrix} \end{array}$$

[Introduced by Yue et al. 2012]

Dueling Bandits

A set of **players** (arms) $[n] = \{1, \ldots, n\}$.

A learner observes noisy comparisons of pairs of **players** from $[n]$.

A **winning probability** matrix P that holds $P_{a,b} := P(a > b)$.

Common assumptions:

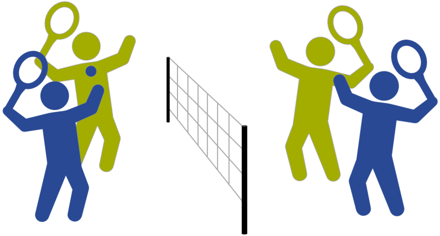
- **Total order** $\succ$ over players such that: For two players $a \succ b \Leftrightarrow P_{a,b} > 0.5$.
- The matrix P satisfies **strong stochastic transitivity (SST)**:
  For $a \succ b \succ c$ it holds that $P_{a,c} \geq \max\{P_{a,b}, P_{b,c}\}$

Possible Goals

Perform pairwise comparisons to:

- **minimize number of queries** to learn the best player a^* with high probability
- given $k \geq 1$, **minimize number of queries** to learn the top-k players with high probability, in dependence of a gap $\Delta = P_{k,k+1} - 1/2$

Dueling Bandits with Team Comparisons



$$P = \begin{array}{c} \\ ab \\ ac \\ cd \\ de \\ \dots \end{array} \begin{array}{ccccc} ab & ac & cd & de & \dots \\ \end{array}$$

$$P = \begin{pmatrix} X & X & 0.6 & 0.8 & \\ X & X & X & 0.7 & \\ 0.4 & X & X & X & \\ 0.3 & X & X & X & \\ & & & & \end{pmatrix}$$

	ab	ac	cd	de	$\dots$
ab	X	X	0.6	0.8	
ac	X	X	X	0.7	
cd	0.4	X	X	X	
de	0.3	X	X	X	
$\dots$					

A learner observes the outcome of noisy **comparisons of disjoint teams** of size k.

Dueling Bandits with Team Comparisons

Last Updated: 5th May, 2020 20:20 IST

Serena Williams Plays Against Herself In Funny TikTok Video During Coronavirus Lockdown

Serena Williams posted a hilarious TikTok video where the 23-time Grand Slam Champion can be seen playing against herself amidst the coronavirus lockdown.

Written By Sreehari Menon

$$
P = \begin{array}{c|cccccc}
 & ab & ac & cd & de & \dots \\
\hline
ab & X & X & 0.6 & 0.8 & \\
ac & X & X & X & 0.7 & \\
cd & 0.4 & X & X & X & \\
de & 0.3 & X & X & X & \\
\dots & & & & &
\end{array}
$$

A learner observes the outcome of noisy **comparisons of disjoint teams** of size k.

Our Model

- A set of **players** (arms) $[n] = \{1, \ldots, n\}$.
- A constant **team size** $k \in \mathbb{N}$, ($k \leq n/2$).
- Every k-sized subset of players is called a **team**.
- There exists a **winning probability matrix** P on the set of all teams.

Our Model

- A set of **players** (arms) $[n] = \{1, \ldots, n\}$.
- A constant **team size** $k \in \mathbb{N}$, $(k \leq n/2)$.
- Every k-sized subset of players is called a **team**.
- There exists a **winning probability matrix** P on the set of all teams.

Assumptions:

- **Total order** over all teams $\succ$ such that:
  For every two teams $A, B \subset [n]$, $A \succ B \Leftrightarrow P_{A,B} > 0.5$.
- The total order is **consistent** with a **total order over single players**, $\rhd$:

$$\forall a, b \in [n]\ \forall S \subseteq [n] \setminus \{a, b\} \text{ s.t. } |S| = k-1 : a \rhd b \Leftrightarrow S \cup \{a\} \succ S \cup \{b\}$$

- The matrix P satisfies **strong stochastic transitivity (SST)**:
  For every triplet of teams such that $A \succ B \succ C$, $P_{A,C} \geq \max\{P_{A,B}, P_{B,C}\}$.

Our Model

- A set of **players** (arms) $[n] = \{1, \ldots, n\}$.
- A constant **team size** $k \in \mathbb{N}$, $(k \leq n/2)$.
- Every k-sized subset of players is called a **team**.
- There exists a **winning probability matrix** P on the set of all teams.

Assumptions:

- **Total order** over all teams $\succ$.
- The total order is **consistent** with a **total order over single players**, $\triangleright$
- The matrix P satisfies **SST**: $A \succ B \succ C \Leftarrow P_{A,C} \geq \max\{P_{A,B}, P_{B,C}\}$.

Goal 1

Return the team of the top-k (best team) players with high probability.

Results

Stochastic Setting

- **Characterization** of the deducible pairwise relations between players
- **Reduction** to any top-k identification dueling bandits setting
- After $\mathcal{O}((n + k \log k)\Delta^{-2} \max(\log \log n, \log k))$ duels in expectation, return the **top-k team** with high probability, where Δ is a gap parameter.

Deterministic Setting

We can reduce any instance from n players to $\mathcal{O}(k)$ within $\mathcal{O}(nk \log(k))$ duels.

- Identify a **Condorcet winning team** after $\mathcal{O}(nk \log(k) + k^2 \log(k) 2^{5k})$ duels.
- For additive total orders, find a **Condorcet winning team** after $\mathcal{O}(nk \log(k) + k^5)$ duels.

Example

Consider $n = 4$, $k = 2$

Total order among teams $ab \succ ac \succ ad \succ bc \succ bd \succ cd$ (hence $a \triangleright b \triangleright c \triangleright d$)

For every two teams X and Y, $P_{X,Y} = 1 \Leftrightarrow X \succ Y$

Only three feasible duels: (ab, cd), (ac, bd), and (ad, bc)

In all of them, the team that has a wins.

Even if the learner knows that $\forall\, X \cap Y = \emptyset : \; P_{X,Y} \in \{0, 1\}$*, it is impossible to distinguish between* b, c, *and* d

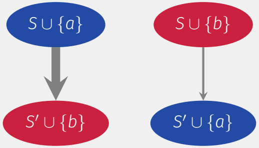
Question

When and how can a learner distinguish between two **single players**?

Subset-Subset Witness for $a \triangleright b$

Two disjoint subsets of players $S, S' \subset [n] \setminus \{a, b\}$ of size $k-1$ that hold

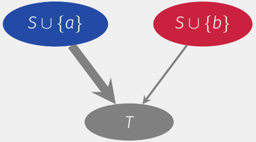
$$P_{S \cup \{a\}, S' \cup \{b\}} > P_{S \cup \{b\}, S' \cup \{a\}}$$



Subset-Team Witness for $a \triangleright b$

Two disjoint subsets of players $S, T \subset [n] \setminus \{a, b\}$, where $|S| = k-1$ and $|T| = k$, such that:

$$P_{S \cup \{a\}, T} > P_{S \cup \{b\}, T}$$



Theorem

*Player a is **provably better** than player b (written $a \triangleright^* b$) **if and only if** there exists a witness (**subset-subset** or **subset-team**) for $a \triangleright b$.*

A reduction to any top-k algorithm for Dueling Bandits

To simulate a duel between any players a, b:

- Randomly draw a triplet (S, S', T) such that $S, S', T \subset [n] \setminus \{a, b\}$ and $S \cap T = S \cap S' = \emptyset$.
- Perform duels $(S \cup \{a\}, S' \cup \{b\}), (S \cup \{b\}, S' \cup \{a\}), (S \cup \{a\} > T), (S \cup \{b\} > T)$.

$$X_{a,b}(S, S', T) \leftarrow 1/4\Big(\mathbb{1}[S \cup \{a\} > S' \cup \{b\}] - \mathbb{1}[S \cup \{b\} > S' \cup \{a\}] + \mathbb{1}[S \cup \{a\} > T] - \mathbb{1}[S \cup \{b\} > T]\Big)$$

A reduction to any top-k algorithm for Dueling Bandits

For a, b and any triplet (S, S', T) (where (S, S') and (S, T) are possible witnesses):

$$X_{a,b}(S, S', T) = 1/4\Big(\mathbb{1}[S \cup \{a\} > S' \cup \{b\}] - \mathbb{1}[S \cup \{b\} > S' \cup \{a\}] + \mathbb{1}[S \cup \{a\} > T\,] - \mathbb{1}[S \cup \{b\} > T\,]\Big)$$

For a, b let $X_{a,b}$ be the outcome of $X_{a,b}(S, S', T)$ for some **randomly drawn triplet** (S, S', T).

Theorems

Let $P'_{a,b} = \mathbb{E}[X_{a,b}] + 1/2$.

- For every pair of players, if $a \triangleright b$ then $P'_{a,b} \geq 1/2$.
- For every pair of players it holds that $a \triangleright^* b$ if and only if $P'_{a,b} > 1/2$.
- For every triplet of players $a \triangleright b \triangleright c$ it holds that $P'_{a,c} \geq \max\{P'_{a,b}, P'_{b,c}\}$

A reduction to any top-k algorithm for Dueling Bandits

Let $\Delta = \mathbb{E}[X_{k,k+1}]$ (gap parameter).

Applying a dueling bandits algorithm by [e.g., Mohajer, Suh, and Elmahdy 2017]:

Corollary

There exists an algorithm that, with probability 0.99, returns the top-k team and requires $\mathcal{O}((n + k\log k)\frac{\max(\log\log n, \log k)}{\Delta^2})$ duels in expectation.

Deterministic Setting

Goal 2

Find a Condorcet winning team in the deterministic setting.

Deterministic Setting: $P_{A,B} \in \{0,1\}$, for all teams A, B.

A is a **Condorcet Winning Team**, if $A \succ B$ for all teams B that are **disjoint** to A.

Goal 2

Find a Condorcet winning team in the deterministic setting.

Deterministic Setting: $P_{A,B} \in \{0, 1\}$, for all teams A, B.

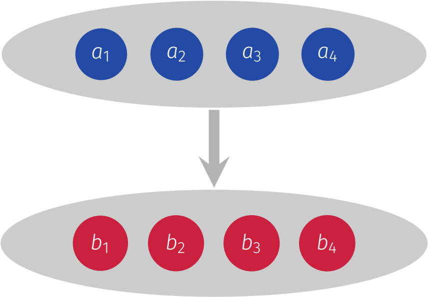
A is a **Condorcet Winning Team**, if $A \succ B$ for all teams B that are **disjoint** to A.

Example:

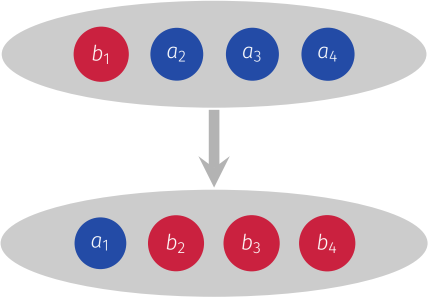
$ab \succ ac \succ ad \succ bc \succ bd \succ cd$

Condorcet winning teams are not unique!

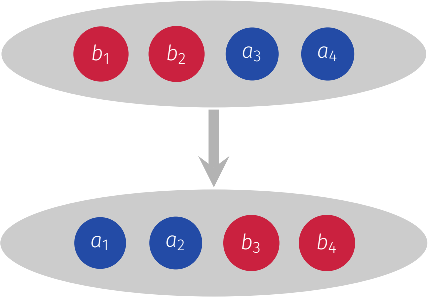
The Uncover Subroutine



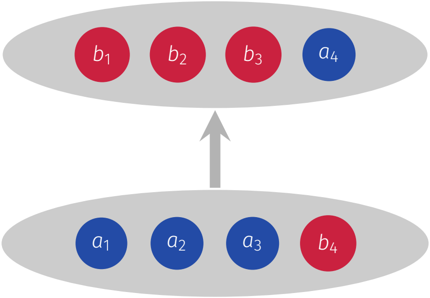
The Uncover Subroutine



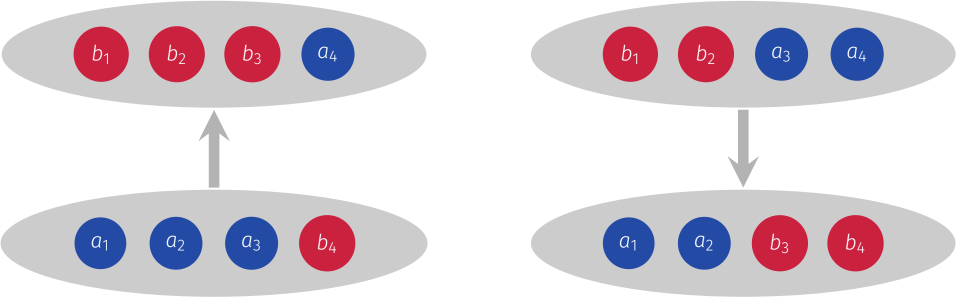
The Uncover Subroutine



The Uncover Subroutine

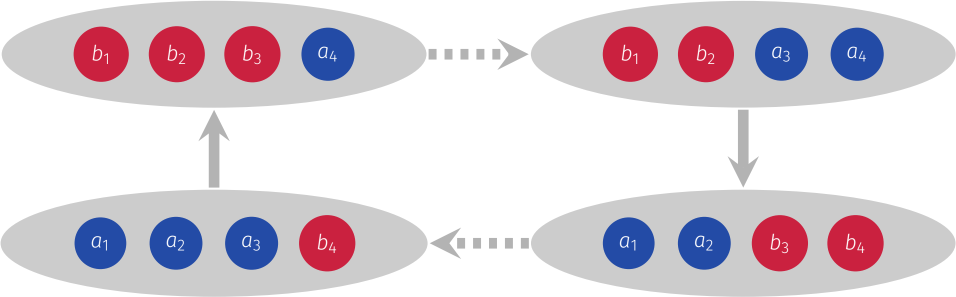


The Uncover Subroutine



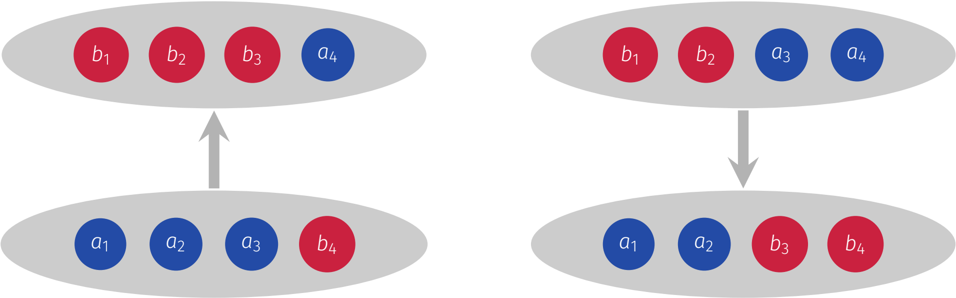
$S = \{b_1, b_2, a_4\}$ and $S' = \{a_1, a_2, b_4\}$ forms a subset-subset witness for $a_3 \triangleright b_3$.

The Uncover Subroutine



$S = \{b_1, b_2, a_4\}$ and $S' = \{a_1, a_2, b_4\}$ forms a subset-subset witness for $a_3 \triangleright b_3$.

The Uncover Subroutine

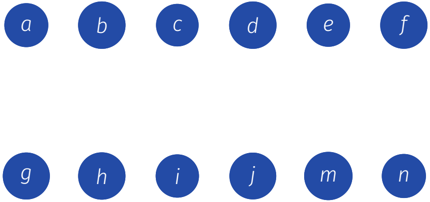


$S = \{b_1, b_2, a_4\}$ and $S' = \{a_1, a_2, b_4\}$ forms a subset-subset witness for $a_3 \triangleright b_3$.

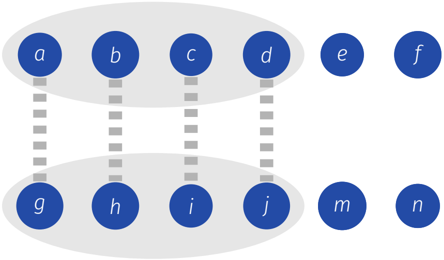
Theorem

Let $A = \{a_1, \ldots, a_k\}$ and $B = \{b_1, \ldots, b_k\}$ be two disjoint teams with $A \succ B$. After $\mathcal{O}(\log(k))$ duels, **Uncover** *returns (a_i, b_i) and a witness for $a_i \triangleright b_i$.*

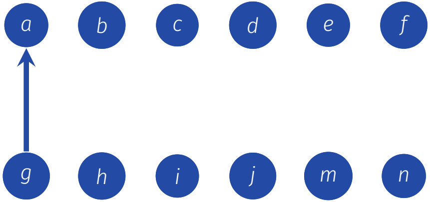
Reduction to $\mathcal{O}(k)$ players



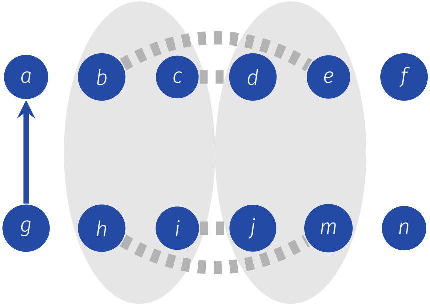
Reduction to $\mathcal{O}(k)$ players



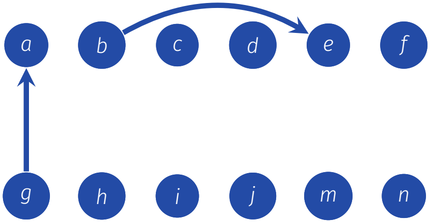
Reduction to $\mathcal{O}(k)$ players

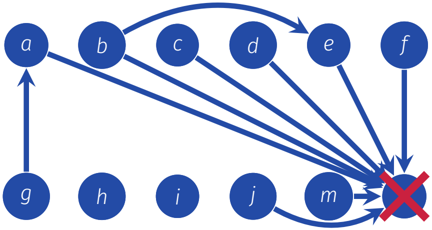


Reduction to $\mathcal{O}(k)$ players



Reduction to $\mathcal{O}(k)$ players

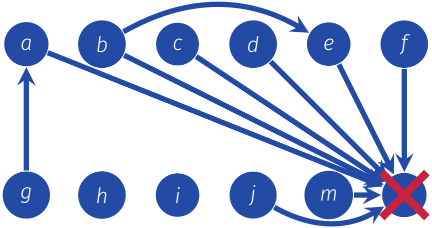




Lemma

Let $R \subseteq [n]$ including the top-$2k$ players. Let A^* be a team for which $A^* \succ B$ for all disjoint $B \subseteq R$. Then, A^* is Condorcet winning in the original instance.

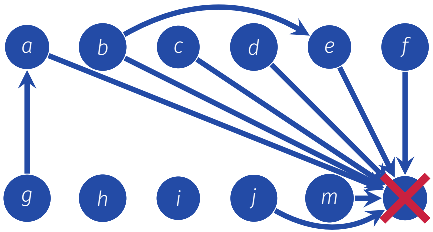
Reduction to $\mathcal{O}(k)$ players



Theorem

After $\mathcal{O}(nk\log(k))$ duels **ReducePlayers** *returns $R \subseteq [n]$ containing the top-$2k$ players and guaranteeing that $|R| < 6k$.*

Reduction to $\mathcal{O}(k)$ players



Theorem

After $\mathcal{O}(nk\log(k))$ duels **ReducePlayers** *returns $R \subseteq [n]$ containing the top-$2k$ players and guaranteeing that $|R| < 6k$.*

Theorem

There exists an algorithm that identifies a **Condorcet winning team** *within $\mathcal{O}(nk\log(k) + k^2\log(k)2^{5k})$ duels.*

The Case of Additive Linear Orders

A total order is **additive linear**, if there are values $v(a) \in \mathbb{R}$ for all $a \in [n]$, such that

$$A \succ B \Leftrightarrow \sum_{a \in A} v(a) > \sum_{b \in B} v(b) \quad \text{for all teams } A \neq B.$$

Theorem

For additive total orders, there exists an algorithm that finds a Condorcet winning team within $\mathcal{O}(nk \log(k) + k^5)$ duels.

Discussion and Open Questions

- **Regret bound:** By optimizing δ, we can easily derive a regret bound of $\mathcal{O}(n(\Delta^{-2}(\log(T) + \log\log\Delta^{-1}))$ based on the results in the stochastic setting, where T is the number of rounds.
- **Lower Bounds:** Any algorithm needs $n - 2k$ duels to identify a Condorcet winning team. Can we find better lower bounds?
- **Relaxing the assumptions:** Can we relax the total order or consistency assumptions and still get a tractable model?