

The Sampling Complexity of Condorcet Winner Identification in Dueling Bandits

El Mehdi Saad^{1,†}

Victor Thuot^{2,†}

Nicolas Verzelen²

¹UM6P College of Computing, Rabat, Morocco

²MISTEA, INRAE, Institut Agro, Univ. Montpellier, France

[†]Equal contribution



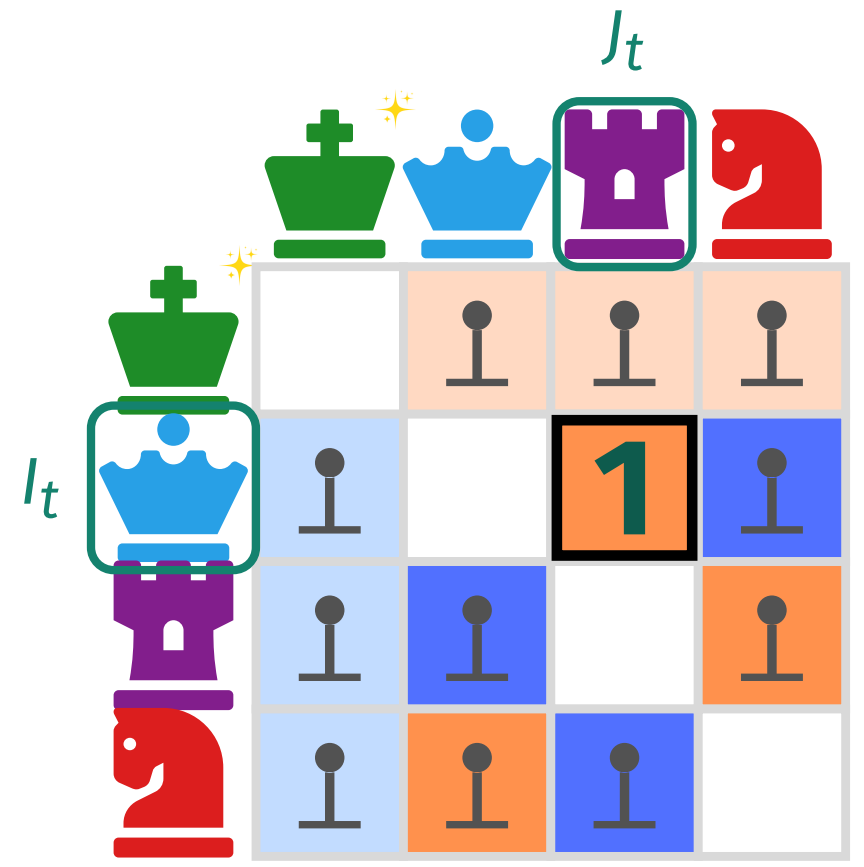
Paper (arXiv)

Motivation

Comparing is easier than scoring $\Rightarrow$ use pairwise comparisons
Observations are costly $\Rightarrow$ choose pairs **sequentially and adaptively**.

Model: dueling bandits

Unknown **gap matrix** between K items (e.g. players): $\Delta_{ij} := \mathbb{P}(i \text{ beats } j) - \frac{1}{2}$.



Assumption $\triangleright$ the **sole** assumption
a **Condorcet winner** i^* : $\Delta_{i^*j} > 0, \forall j \neq i^*$
i.e. i^* beats **every other item** in pairwise duels in expectation

skew-symmetry: $\Delta_{ij} = -\Delta_{ji}, \Delta_{ii} = 0$
number of defeats of i : $K_{i,<0} := |\{j : \Delta_{ij} < 0\}|$
ordered defeats: $\Delta_{i,(1)} \leq \dots \leq \Delta_{i,(K_{i,<0})} < 0$
dark: clear win, light: near tie
1 observation $X_t = 1$ of the duel (i_t, j_t)
1 Δ **unknown**: each entry is a **duel** we can play

Sampling protocol

$\triangleright$ sequential & adaptive duels
for $t = 1, 2, \dots$ **do**
 choose a pair $(i_t, j_t) \in [K]^2$ $\triangleright$ based on past duels
 observe $X_t \sim \text{Bernoulli}(\frac{1}{2} + \Delta_{i_t j_t})$ $\triangleright X_t = 1$ iff i_t beats j_t
 if stop then $N_\delta \leftarrow t$, **output** $\hat{i}$ $\triangleright$ stop & answer chosen from past duels

Objective

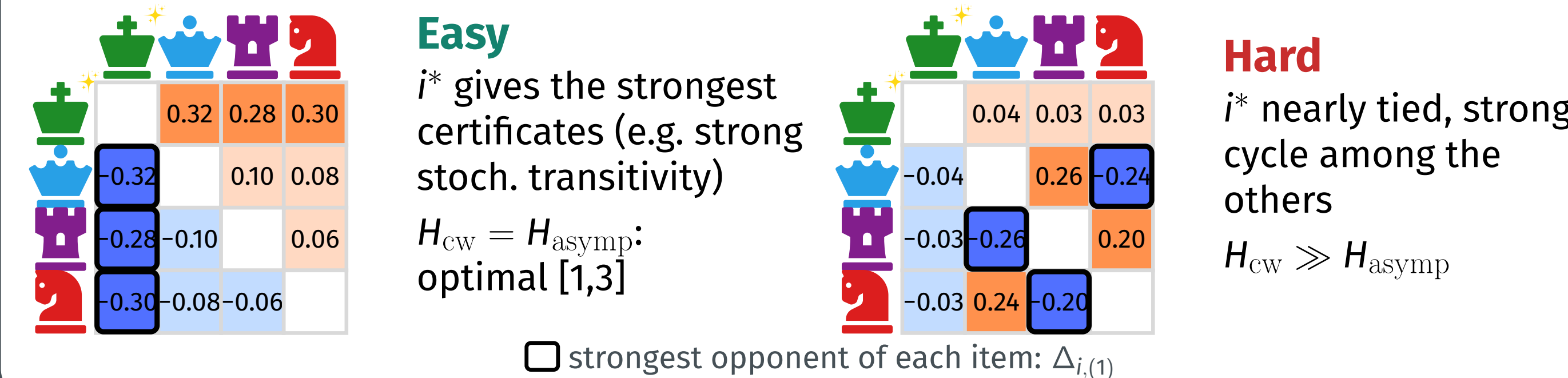
$\triangleright$ fixed confidence δ
Budget: $N_\delta =$ number of duels before stopping (random)
 δ -correct: $\mathbb{P}_\Delta(\hat{i} \neq i^*) \leq \delta$ for **every** Δ with a Condorcet winner
Goal: δ -correct, with a **minimal** $(1-\delta)$ -quantile of N_δ

State of the art: how to prove $i \neq i^*$?

Certify $\Delta_{ij} < 0$ (j : **witness** against i) at level δ : $\simeq \log(1/\delta)/\Delta_{i,j}^2$ duels.

Winner row $H_{\text{cw}}(\delta) := \sum_{i \neq i^*} \frac{\log(1/\delta)}{\Delta_{i,i^*}^2} \leftarrow \text{witness } i^*$ MidSearch [1]
 $N_\delta \lesssim H_{\text{cw}}(\delta)$ w.p. $\geq 1 - \delta$

Strongest opponent $H_{\text{asympt}}(\delta) := \sum_{i \neq i^*} \frac{\log(1/\delta)}{\Delta_{i,(1)}^2} \leftarrow \text{strongest}$ Explore-Verify [2]
 $\mathbb{E}[N_\delta] \underset{\delta \rightarrow 0}{\simeq} H_{\text{asympt}}(\delta)$ opt. [3]

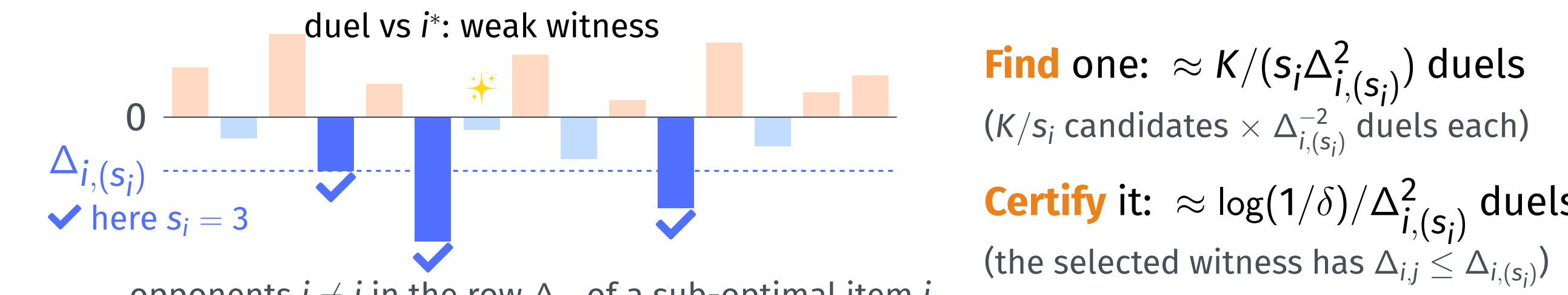


$\mathbf{i}$ H_{asympt} ignores the cost of *finding* strong witnesses (large K).

Research question. What is the **optimal** sampling complexity **beyond** the **winner's row**?

Key idea: exploration vs. certification trade-off

To eliminate $i \neq i^*$: **(i) find** a witness j , **(ii) certify** its sign $\Delta_{ij} < 0$.
Fix a level $s_i \leq K_{i,<0}$: eliminate i with **one of its s_i strongest opponents**.



Trade-off: larger s_i : easier to find, but harder to certify.
Over all items, $\mathbf{s} = (s_i)_{i \neq i^*}$:

Exploration $H_{\text{explore}}(\mathbf{s}, \delta) := \max_{i \neq i^*} \frac{K \log(1/\delta)}{s_i \Delta_{i,(s_i)}^2} + \sum_{i \neq i^*} \frac{K}{s_i \Delta_{i,(s_i)}^2}$ $\log \frac{1}{\delta}$ for the **hardest** item only

Certification $H_{\text{certify}}(\mathbf{s}, \delta) := \sum_{i \neq i^*} \frac{\log(1/\delta)}{\Delta_{i,(s_i)}^2}$ $\mathbf{s} \equiv \mathbf{1}$: $H_{\text{certify}} = H_{\text{asympt}}$, but large H_{explore}

Algorithm FC-CWI

Algorithm: Fixed-Confidence CW Identification FC-CWI(δ)

$T \leftarrow 8K \log K$ $\triangleright \Delta, \mathbf{s}^*$ unknown
while true do $\triangleright$ doubling trick: no need to know the complexity
 $(\varphi, I) \leftarrow \text{FB-CWI}(T, \delta)$ $\triangleright$ steps 1-4 below, $\leq 2T$ duels
 if φ **then return** I $\triangleright$ stops on a certificate: δ -correct
 $T \leftarrow 2T$ $\triangleright$ double the budget

Subroutine FB-CWI(T, δ): $O(\log K)$ score-based eliminations

1 Strong opponent: search the whole row $\alpha^{(s)} \leftarrow \text{BSH}(\alpha, \frac{B_k}{4})$
(find: *exploration*) Bracketed Sequential Halving [4] over **all** $K - 1$ duels (α, β) , with $B_k \simeq T/(|A_k| \log K)$ duels per active item, finds a strong opponent $\alpha^{(s)}$; $Z_k^{(s)}(\alpha) \approx \Delta_{\alpha, \alpha^{(s)}}$.
2 Weak opponent: a lower quantile $Z_k^{(w)}(\alpha) \leftarrow \text{Range-Quantile}(\alpha, \frac{B_k}{4})$
New Range-Quantile finds a value between the $\frac{1}{8}$ - and $\frac{1}{4}$ -quantiles of $(\Delta_{\alpha, \beta})_{\beta \in A_k}$.
 $\geq |A_k|/4$ rows have $\geq |A_k|/4$ entries ≤ 0 , but $Z_k^{(w)}(i^*) > 0$.
3 Score & eliminate the bottom 1/8 $A_{k+1} \leftarrow \text{top } \frac{7}{8} \text{ of } A_k \text{ by } S_k$
 $S_k(\alpha) = \min\{Z_k^{(s)}(\alpha), 0\} + Z_k^{(w)}(\alpha)$
(trade-off) strong: sharp, but must be found; weak: always there, protects i^* .
4 Certify $\varphi \leftarrow \varphi_1 \vee \varphi_2$
 φ_1 (**frontier**): eliminated items all score $< -\tau_k$, while $S_k(i^*) \geq -\tau_k$ w.p. $\geq 1 - \delta \Rightarrow i^*$ never eliminated.
 φ_2 (**direct**): Test-CW certifies $\Delta_{ij} > 0, \forall j \neq i$, at level δ ($\leq T$ duels): the $H_{\text{cw}}(\delta)$ regime.
✓ φ true $\rightarrow$ FC-CWI returns I
✗ else: $T \leftarrow 2T$ (double the budget), back to step 1

Upper bound: FC-CWI adapts to the full gap matrix

Theorem 1. FC-CWI is δ -correct and, w.p. $\geq 1 - \delta$, up to logs:

$$N_\delta \lesssim H_{\text{cw}}(\delta) \wedge \min_{\mathbf{s}: s_i \leq K_{i,<0}} \{H_{\text{certify}}(\mathbf{s}, \delta) + H_{\text{explore}}(\mathbf{s}, \delta)\}$$

Never worse than H_{cw} [1], **much better** when i^* has small gaps. The minimizer $\mathbf{s}^*$ (effective sparsity) is reached **automatically**.

Lower bound: the trade-off is unavoidable

Theorem 2 (local minimax). For **any** δ -correct algorithm A and any gap matrix Δ , there exists $\tilde{\Delta}$ obtained by permuting entries of Δ (same winner, sign structure and $\mathbf{s}^*$) with:

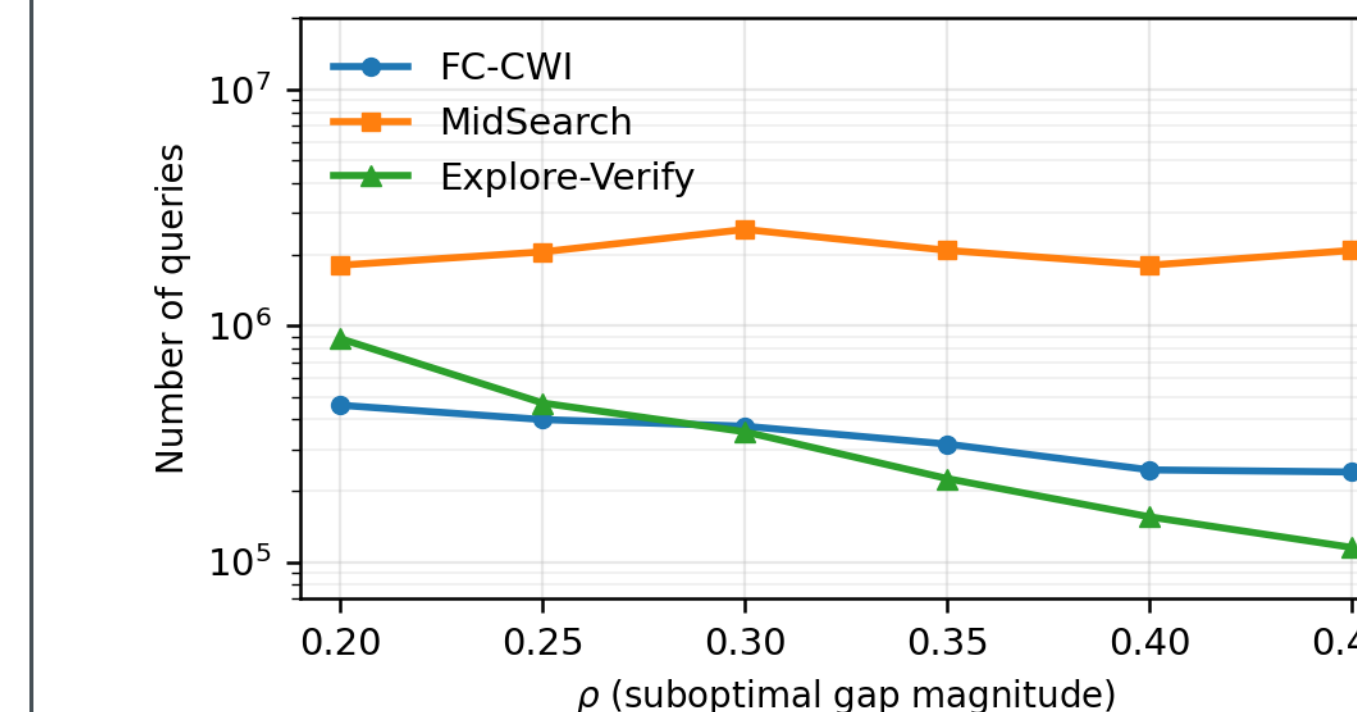
$$\mathbb{P}_{\tilde{\Delta}, A}(N_\delta \gtrsim H_{\text{certify}}(\mathbf{s}^*, \delta) + H_{\text{explore}}(\mathbf{s}^*, \delta)) \geq \delta$$

Key tool: reduction to **active multiple signal detection**.

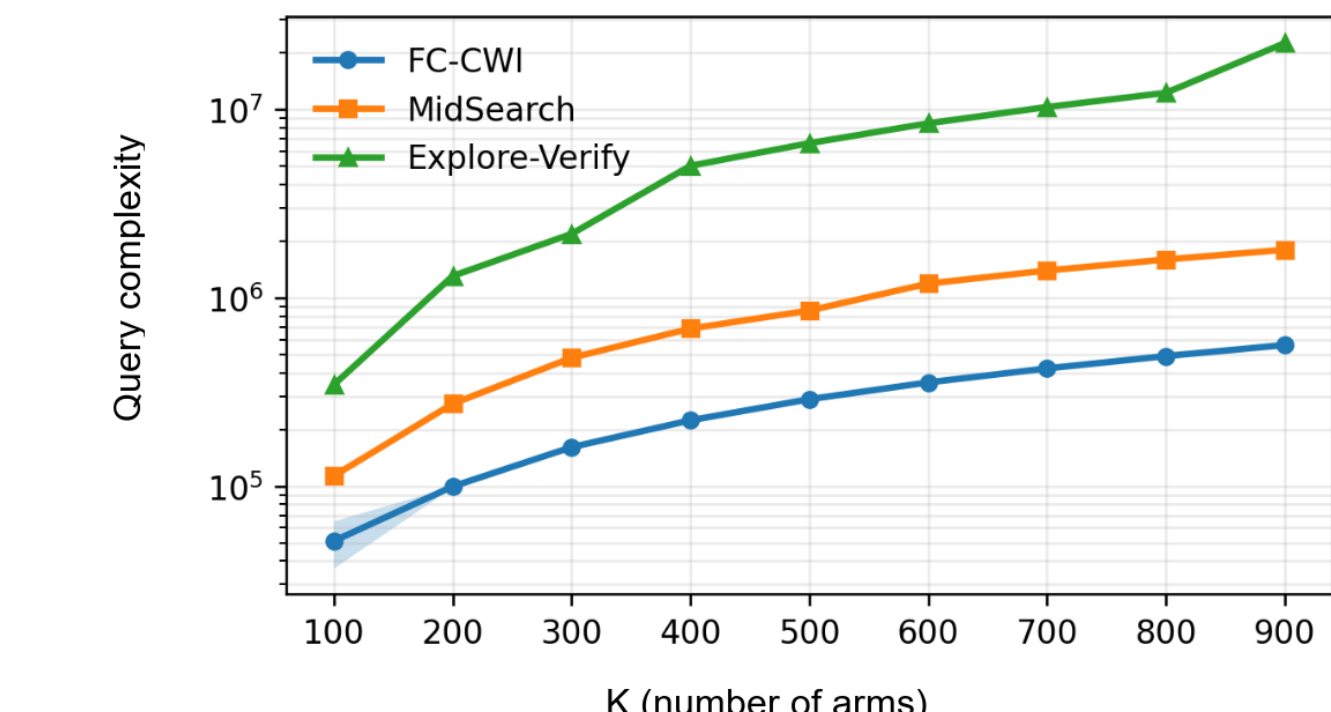
Bounds match (up to logs): FC-CWI is **locally optimal** for all K and δ .

Experiments vs. MidSearch [1] and Explore-Verify [2]

Sparse Δ : i^* wins by $\geq \gamma$; each sub-optimal row has a fraction p of $\pm p$ entries, the rest is 0 (p : **informativeness** of sub-optimal duels).



Exp. 1: FC-CWI $\ll$ MidSearch
($K = 64, p = 0.3, \delta = 0.1$)



Exp. 2: Explore-Verify worst in K
($\gamma = 0.3, p = 0.4, \delta = 0.3$)

Main message

The budget of FC-CWI is optimal (in $(1-\delta)$ -quantile, up to logs, local minimax):

$$N_\delta \simeq H_{\text{cw}}(\delta) \wedge \min_{\mathbf{s}} \{H_{\text{certify}}(\mathbf{s}, \delta) + H_{\text{explore}}(\mathbf{s}, \delta)\}$$

(H_{cw} is optimal when i^* is the strongest opponent of every item [3])

The full gap matrix matters, not only the winner's row.

Exploration vs. certification: intrinsic, balanced **automatically**.

Large K : up to $K \times$ fewer duels than [2]; **arbitrarily** fewer than [1].

Expectation $\neq$ quantile: as $\delta \rightarrow 0$ (when H_{cw} is not the min), the optimal $(1-\delta)$ -quantile of N_δ is up to $K \times$ larger than $\mathbb{E}[N_\delta] \simeq H_{\text{asympt}}(\delta)$:

$$\log \frac{1}{\delta} \inf_{\mathbf{s}} \left[\sum_{i \neq i^*} \frac{1}{\Delta_{i,(s_i)}^2} + \max_{i \neq i^*} \frac{K}{s_i \Delta_{i,(s_i)}^2} \right]$$

[1] Maiti, Boczar, Jamieson & Ratliff. Near-optimal pure exploration in matrix games: a generalization of stochastic bandits & dueling bandits. *AISTATS*, 2024.

[2] Karnin. Verification based solution for structured MAB problems. *NeurIPS*, 2016.

[3] Haddenhorst, Bengs & Hüllermeier. Identification of the generalized Condorcet winner in multi-dueling bandits. *NeurIPS*, 2021.

[4] Zhao, Stephens, Szepesvári & Jun. Revisiting simple regret: fast rates for returning a good arm. *ICML*, 2023.