EXACT AND APPROXIMATE CONDORCET WINNER IDENTIFICATION IN DUELING BANDITS

Anonymous authors
Paper under double-blind review

ABSTRACT

We study fixed-confidence identification in dueling bandits under the assumption that the preference matrix admits an approximate Condorcet winner (CW). We first propose DUELELIM, a phased elimination algorithm whose pivot is selected through a *median-of-quantiles* estimator of each arm's preference column, obtained by sub-sampling a constant number of random opponents, and prove that, if it exists, it returns the CW with probability at least $1-\delta$ using $O\big(\sum_{k\neq k^*}\Delta_k^{-2}(\ln\frac{K}{\delta}+\ln\ln\frac{1}{\Delta_k})\big)$ comparisons, where Δ_k is the Condorcet gap of arm k. Our main contribution concerns *approximate* identification, previously studied only under transitivity or parametric assumptions. When some arm beats every other arm up to a slack ε, a two-stage variant returns an ε'-approximate Condorcet winner with $\tilde{O}\big((1+N)\sum_k \max(\Delta_k,\varepsilon')^{-2}\big)$ comparisons, where N is the number of arms near-tied with the winner. This is optimal up to logarithmic factors when $N=0$, and a second lower bound shows that without transitivity the factor N cannot be removed: unlike in ε-best-arm identification, $\sum_k \max(\Delta_k,\varepsilon')^{-2}$ alone does not govern the sample complexity. Finally, relaxing the minimum in the definition of a Condorcet winner into a quantile yields an algorithm whose sample complexity $\tilde{O}\big(\frac{1}{\rho\eta\varepsilon^2}\big)$ does not depend on the number of arms and improves with the proportion ρ of good arms.

1 INTRODUCTION

In many applications it is easier to ask people which of two options they prefer than to ask them for a numerical score. A user can say which of two rankings she likes better (Yue & Joachims, 2009), an annotator which of two model answers is better (Christiano et al., 2017), and pairwise votes have become the standard way of evaluating language models (Chiang et al., 2024). The *dueling bandit* model (Yue et al., 2012) is the natural framework for learning from such feedback. There are K arms, and a preference matrix $\mathbf{P}\in[0,1]^{K\times K}$ with $P(i,j)+P(j,i)=1$ describes how they compare: at each round the learner picks a pair (i,j) and observes an independent Bernoulli$(P(i,j))$ outcome, which tells it which of the two arms won. We refer to Bengs et al. (2021) for a survey on dueling bandits.

A first question in this model is what "the best arm" should mean. When preferences are not transitive, and in aggregated human judgements they usually are not, a natural answer is the *Condorcet winner* (CW): an arm k^* that beats every other arm, in the sense that $P(k^*,j)>\frac{1}{2}$ for every $j\neq k^*$. Such an arm need not exist, but assuming that it does is a much weaker requirement than stochastic transitivity or a Bradley–Terry–Luce model, and it is the assumption under which most regret minimisation algorithms are analysed (Zoghi et al., 2014). Other notions of winner, such as the Borda, Copeland or von Neumann winners, always exist; we discuss them in App. A.

Objective. In this paper we study the *fixed-confidence identification* problem: the learner must return the Condorcet winner with probability at least $1-\delta$, using as few comparisons as possible. To measure this cost, write $\Delta_k := P(k^*,k)-\frac{1}{2}>0$ for the *Condorcet gap* of an arm $k\neq k^*$, and consider $H_{\mathrm{cw}}(\delta) := \ln\frac{1}{\delta}\sum_{k\neq k^*}\frac{1}{\Delta_k^2}$.

This quantity is the cost of merely *certifying* that k^* is the winner once it is known, and it is a lower bound on the cost of any correct algorithm (Zhou et al., 2017). It plays the same role as $\sum_k \Delta_k^{-2}\ln\frac{1}{\delta}$ in best-arm identification for classical bandits (Kaufmann et al., 2016). The difficulty specific to dueling bandits is that k^* is defined through its whole row of the matrix. Without transitivity, an arm that beats many others may still lose to an arm that loses to almost everyone; as a consequence, no arm can be discarded after a few comparisons against arbitrary opponents.

What is known. The first identification algorithms, SAVAGE (Urvoy et al., 2013) and PBR (Busa-Fekete et al., 2013), estimate the entire preference matrix. Their cost is therefore a sum over *pairs* of arms, which can be a factor K larger than $H_{\mathrm{cw}}(\delta)$. Later works improved on this. The explore-then-verify scheme of Karnin (2016) is asymptotically optimal in expectation as $\delta\to 0$ but of order K^2 for fixed δ; Haddenhorst et al. (2021) obtain the worst-case rate $K\ln K/\Delta_{\min}^2$; and, more recently, Maiti et al. (2024) and Saad et al. (2026) reach $H_{\mathrm{cw}}(\delta)$ up to logarithmic factors

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

in K. All of these works, however, address *exact* identification. Note that Maiti et al. (2024) study the strictly more general problem of identifying the pure Nash equilibrium of a zero-sum matrix game with an $n \times m$ payoff matrix; the dueling Condorcet identification problem just becomes a special case of their result for the skew-symmetric case $n = m = K$. The approximate (PAC) version of the problem, in which one is satisfied with an arm that no other arm beats by more than ε, is well understood in classical bandits, where its complexity is $\sum_k \max(\Delta_k, \varepsilon)^{-2} \ln \frac{1}{\delta}$ up to logarithms (Kalyanakrishnan et al., 2012). In dueling bandits it has only been studied under transitivity or parametric assumptions (Falahatgar et al., 2017), and under the Condorcet assumption alone it was left open (Karnin, 2016). This is unfortunate, because in real preference data an exact Condorcet winner is typically either absent or tied with a few competitors, so that the approximate problem is the one that matters in practice.

Contributions. We address these questions with three algorithms and the corresponding lower bounds. All our results hold for arbitrary preference matrices, are non-asymptotic, and hold with probability at least $1 - \delta$.

- **Exact identification** (Sec. 3). As a preliminary step, we propose DUELELIM, a phased elimination algorithm in the spirit of Median Elimination (Even-Dar et al., 2002): in each phase it selects a pivot arm and eliminates the arms that lose convincingly to it. The pivot is the arm whose preference column has the largest central quantile, and this quantile is estimated with $O(\varepsilon^{-2} \ln \frac{1}{\delta})$ comparisons, *whatever the number of alive arms*, by a median-of-quantiles estimator (Sec. 2.2) that only sub-samples a constant number of random opponents. Thm. 3.1 recovers the rate $\sum_{k \neq k^*} \Delta_k^{-2}$ of Maiti et al. (2024) and Saad et al. (2026) up to logarithmic factors, and improves their bounds by a factor $\ln K$ in the worst case. This gain comes from a different algorithmic structure, in which correctness is guaranteed by the elimination test itself rather than by a separate verification step.
- **Approximate identification** (Sec. 4). We then relax the assumption and only require that some arm beats every other arm up to a slack ε. A two-stage algorithm (a recentred elimination followed by a certification stage) returns an ε'-approximate Condorcet winner, for any target $\varepsilon' \geq 2\varepsilon$, using $O\big((1 + N(2\varepsilon')) \sum_k \max(\Delta_k, \varepsilon')^{-2} (\ln \frac{K}{\delta} + \ln \ln \frac{1}{\varepsilon'})\big)$ comparisons, where $N(x)$ counts the arms within x of the winner (Thm. 4.3). This has the form of the usual ε-best-arm identification bounds, except for the factor $1 + N$ in front. Two lower bounds show that this form is the right one. When the near-winner is isolated ($N = 0$), a change-of-measure argument shows that $\sum_k \max(\Delta_k, \varepsilon')^{-2} \ln \frac{1}{\delta}$ comparisons are necessary (Proposition 4.4), so the bound is tight up to logarithms. When near-tied arms exist, on the other hand, a multiplicative dependence on their number cannot be avoided (Proposition 4.5): we exhibit a K-armed family on which every correct algorithm needs $\Omega(K^2)$ comparisons although $\sum_k \max(\Delta_k, \varepsilon')^{-2} = O(K)$. In other words, and unlike in ε-best-arm identification, the quantity $\sum_k \max(\Delta_k, \varepsilon')^{-2}$ alone cannot be the complexity of the problem. The reason is that certifying an approximate Condorcet winner is a needle-in-a-haystack problem: a single hidden weak spot of the candidate has to be found among its $K - 1$ opponents.
- **Quantile Condorcet winners** (Sec. 5). We show both difficulties disappear if the *minimum* over the row is relaxed into a *quantile*: we call an arm a (c, η)-Condorcet winner if it loses by c or more to at most a fraction η of its opponents. Our algorithm QUANTILECW returns a $(2\varepsilon, 2\eta)$-Condorcet winner with $\tilde{O}\big(\frac{1}{\rho\eta\varepsilon^2}\big)$ comparisons (Thm. 5.2), where ρ is the unknown proportion of (ε, η)-Condorcet winners. To our knowledge, this is the first identification guarantee for Condorcet-Identification on arbitrary preference matrices whose cost does not grow with K.
- **Experiments (App. G–App. J).** Finally, we validate all three algorithms on synthetic preference matrices. On the benchmark of Maiti et al. (2024), DUELELIM-3T returns k^* with a formal δ-guarantee in every run within 3–4 H_1 at $d = 1024$, whereas the certified algorithms of Maiti et al. (2024) and Saad et al. (2026) require $>1000\, H_1$ and $>100\, H_1$ respectively. In the approximate regime, DUELELIM-ε-3T is $3.5\times$ cheaper than exact identification when $\Delta_{\min} \ll \varepsilon' \ll \beta$, confirming the prediction of Thm. 4.3. For the quantile problem, QUANTILECW is flat in K from $K = 10^3$ to $K = 10^5$ while the natural row-scan baseline grows linearly, crossing at $K \approx 10^3$.

The three algorithms rely on the same two tools, which we present in Sec. 2: a time-uniform one-sided Bernoulli KL test, derived from the mixture-martingale bound of Kaufmann & Koolen (2021), and the median-of-quantiles estimator. All proofs are given in the appendix, and App. A discusses further related work (Borda, Copeland and von Neumann winners, PAC maxing under transitivity, and quantile identification in classical bandits).

2 SETTING AND TWO ELEMENTARY TOOLS

Setting. A dueling bandit instance is a matrix $\mathbf{P} \in [0, 1]^{K \times K}$ with $P(i, j) + P(j, i) = 1$ for $i \neq j$ and $P(i, i) = \frac{1}{2}$. Comparing arms i and j returns an independent Bernoulli$(P(i, j))$ variable, which equals 1 if i wins. For any pair we write $\Delta(j, k) := P(j, k) - \frac{1}{2} \in [-\frac{1}{2}, \frac{1}{2}]$. A *Condorcet winner* is an arm k^* with $\Delta(k^*, k) > 0$ for all $k \neq k^*$. It is unique when it exists, and we write $\Delta_k := \Delta(k^*, k)$ for its gaps and $\Delta_{(1)} \leq \cdots \leq \Delta_{(K-1)}$ for the same gaps in sorted order. We will also need order statistics restricted to a set of alive arms at some phase $\ell \geq 1$ of the

Algorithm 1 $\mathtt{MME}(\{\nu_i\}_{i\in[K]}, \delta, \varepsilon, k_-, k_+)$ – median-of-quantiles estimator

$M \leftarrow \lceil 12K^2/(k_+ - k_-)^2 \rceil, \quad z \leftarrow \lceil (1 - \frac{k_+ + k_-}{2K})M \rceil, \quad N \leftarrow \lceil 6/\varepsilon^2 \rceil, \quad I \leftarrow \lceil 0.6 \ln(1/\delta) \rceil$
for $i = 1, \dots, I$ **do**
  Draw a batch B_i of M indices uniformly at random with replacement from $[K]$
  For each $j \in B_i$ draw N samples from ν_j and let $\hat{p}_{i,j}$ be their mean
  $v_i \leftarrow$ the z-th largest value of $\{\hat{p}_{i,j} : j \in B_i\}$ (with multiplicity)
end for
return $\hat{m} \leftarrow \mathrm{Median}\{v_1, \dots, v_I\}$

algorithm: for a set $\mathcal{A}_\ell \subseteq [K]$ of size K_ℓ, an arm $k \in \mathcal{A}_\ell$ and $j \le K_\ell - 1$, $\Delta_\ell(k, j)$ denotes the j-th smallest value of $\{\Delta(k, i) : i \in \mathcal{A}_\ell \setminus \{k\}\}$. Since removing arms cannot decrease an order statistic, $\Delta_\ell(k^*, j) \ge \Delta_{(j)}$.

All our algorithms keep, for every pair $\{i, j\}$, the number $n_{i,j}$ of comparisons made so far and the empirical win frequency $\hat{P}(i, j) = 1 - \hat{P}(j, i)$. These counters are never reset, so the comparisons of a pair always form a single i.i.d. sequence, whatever the stage that requested them, with a random number of comparisons that depends on past observations. Throughout, $K \ge 2$ and $\delta \in (0, e^{-8})$.

2.1 A TIME-UNIFORM ONE-SIDED TEST

Let $\mathrm{kl}(x\|q) := x \ln \frac{x}{q} + (1 - x) \ln \frac{1-x}{1-q}$ be the Bernoulli Kullback–Leibler divergence and, for $\lambda > 0$,

$$f_\lambda(n) \;:=\; 3\ln(1 + \ln n) + \tfrac{3}{2}\lambda + 10, \qquad \text{with } \lambda := \ln \tfrac{2K(K-1)}{\delta} \text{ and we denote } f := f_\lambda \text{ in the main text.} \tag{1}$$

Our first tool is a deviation inequality that holds uniformly over the sample size. It follows from the mixture-martingale bound of Kaufmann & Koolen (2021, Thm 7), specialised to a single Bernoulli arm, and is proved in App. B: if $\hat{p}_n$ is the mean of n i.i.d. Bernoulli(p) variables and $\delta' \in (0, e^{-8})$, then

$$\mathbb{P}\big(\exists n \ge 1 :\; n\,\mathrm{kl}(\hat{p}_n\|p) \ge f_{\ln(1/\delta')}(n)\big) \;\le\; \delta'. \tag{2}$$

The KL form (Garivier & Cappé, 2011) has a useful consequence: a single event per pair is enough to run tests against *any* reference point, which Sec. 4 will exploit. In the rest of the paper we only use the following consequence of (2).

Lemma 2.1 (One-sided KL tests)**.** *Let $\mathcal{E} := \{\forall n \ge 1 :\; n\,\mathrm{kl}(\hat{p}_n\|p) < f_{\ln(1/\delta')}(n)\}$, so that $\mathbb{P}(\mathcal{E}) \ge 1 - \delta'$. On $\mathcal{E}$, for every $q \in (0, 1)$ and $n \ge 1$:*

(i) if $\hat{p}_n > q$ and $n\,\mathrm{kl}(\hat{p}_n\|q) \ge f_{\ln(1/\delta')}(n)$ then $p > q$, and symmetrically if $\hat{p}_n < q$;
(ii) if $p \ge q + \gamma$ with $\gamma > 0$, then $\hat{p}_n > q$ and $n\,\mathrm{kl}(\hat{p}_n\|q) \ge f_{\ln(1/\delta')}(n)$ for every n with $n\gamma^2 \ge 2f_{\ln(1/\delta')}(n)$, and symmetrically if $p \le q - \gamma$. Moreover $n \mapsto f_{\ln(1/\delta')}(n)/n$ is non-increasing, and for $g \in (0, 1)$ and $\lambda \ge 8$,

$$n(g) := \min\{n \ge 1 :\; ng^2 \ge 2f_\lambda(n)\} \;\le\; \frac{52}{g^2}\Big(\lambda + \ln\ln\frac{6}{g}\Big). \tag{3}$$

In words, a test that fires is never wrong under the high probability event $\mathcal{E}$, and it fires after $\tilde{O}(\gamma^{-2}\lambda)$ comparisons when the true mean is at distance γ from the reference point. In our algorithms we take $\delta' = \delta/(2K(K-1))$, so that a union bound over the $K(K-1)/2$ pairs costs $\delta/4$ and $f_{\ln(1/\delta')} = f$.

2.2 MEDIAN-OF-QUANTILES ESTIMATION

Our second tool estimates a quantile of a vector of unknown means. Let $p \in [0, 1]^K$ be unknown, with sorted entries $p_{(1)} \le \cdots \le p_{(K)}$, and suppose that each p_i can only be accessed through Bernoulli(p_i) samples. Given $1 \le k_- \le k_+ \le K$, we want an estimate that lies in $[p_{(k_-)} - \varepsilon,\, p_{(k_+)} + \varepsilon]$ with probability $1 - \delta$, and we want the number of samples not to depend on K. This is exactly what pivot selection in Sec. 3 needs, with p a column of $\mathbf{P}$.

Theorem 2.2 (Median-of-quantiles)**.** *Let $1 \le k_- \le k_+ \le K$ with $k_+ - k_- \ge K/4$, $\varepsilon \in (0, 1]$ and $\delta \in (0, e^{-8})$. With probability at least $1 - \delta$, the output of Alg. 1 satisfies $p_{(k_-)} - \varepsilon \le \hat{m} \le p_{(k_+)} + \varepsilon$, and the algorithm draws at most $\frac{66K^2}{(k_+ - k_-)^2} \cdot \frac{\ln(1/\delta)}{\varepsilon^2}$ samples.*

The proof (App. C) follows the median-of-means principle (Nemirovsky & Yudin, 1983): a weak estimator that is correct with constant probability is made reliable by taking the median of $O(\ln \frac{1}{\delta})$ independent copies. Here the weak

estimator is v_i, the z-th largest of $M = O(1)$ sub-sampled empirical means. It is correct with a constant probability p_0 for two reasons: by Hoeffding's inequality, all M means are ε-accurate with constant probability; and the number of sub-sampled indices of rank above $k_\pm$ is a Binomial variable whose mean is at distance $\Theta(M(k_+ - k_-)/K)$ from z. Two features distinguish MME from classical median-of-means (Lugosi & Mendelson, 2019): the randomness comes from sub-sampling a *finite population*, and the target is an order statistic. Unlike Median Elimination (Even-Dar et al., 2002), most of the population is never touched: when $k_+ - k_- \geq K/4$ the cost is $O(\varepsilon^{-2} \ln \frac{1}{\delta})$, whereas estimating a column entry by entry, as SAVAGE and PBR do, costs K times more. The idea of estimating a quantile of a column by sub-sampling opponents is similar to the ones used by Maiti et al. (2024), whose COLMIDVAL subroutine estimates the median of a column, and by Saad et al. (2026), whose weak-opponent score estimates a quantile of the gaps of an arm against random opponents.

3 EXACT CONDORCET WINNER IDENTIFICATION

This section is a preliminary step towards our main results on approximate identification. Its goal is to recover the rate $\sum_{k \neq k^*} \Delta_k^{-2}$ of Maiti et al. (2024) and Saad et al. (2026) with an algorithm whose structure extends to the approximate setting of Sec. 4. As a by-product, this structure removes one multiplicative $\ln K$ factor that appears in the bounds of these two works.

Algorithm 2 DUELELIM

Require: confidence δ; $\mathcal{A}_1 \leftarrow [K]$, $\ell \leftarrow 1$, $m \leftarrow 1$; $n_{i,j} \leftarrow 0$, $\hat{P}(i,j) \leftarrow \frac{1}{2}$ for all i, j
while $|\mathcal{A}_\ell| > 2$ **do**
  $K_\ell \leftarrow |\mathcal{A}_\ell|$; $\varepsilon_\ell \leftarrow 2^{-m}$; $\delta_\ell \leftarrow \frac{3\delta}{\pi^2 K_\ell \ell^2}$; $n_\ell \leftarrow \lceil 12\, \varepsilon_\ell^{-2}(\lambda + \ln(1+\ell)) \rceil$; $k_{\ell,\pm}$ as in (4)
  for each $k \in \mathcal{A}_\ell$ **do** ▷ pivot identification
    $\hat{m}_\ell(k) \leftarrow \mathtt{MME}\big(\{\mathcal{B}(P(k,i)) : i \in \mathcal{A}_\ell \setminus \{k\}\}, \delta_\ell, \varepsilon_\ell, k_{\ell,-}, k_{\ell,+}\big) - \frac{1}{2}$, updating $\hat{P}(k,\cdot)$, $n_{k,\cdot}$
  end for
  $\hat{k}_\ell \leftarrow \arg\max_{k \in \mathcal{A}_\ell} \hat{m}_\ell(k)$
  for each $k \in \mathcal{A}_\ell \setminus \{\hat{k}_\ell\}$ **do** ▷ elimination against the pivot
    loop
      if $n_{\hat{k}_\ell,k} \geq 1$ **and** $\hat{P}(\hat{k}_\ell,k) > \frac{1}{2}$ **and** $n_{\hat{k}_\ell,k}\, \mathrm{kl}\big(\hat{P}(\hat{k}_\ell,k) \,\big\|\, \frac{1}{2}\big) \geq f(n_{\hat{k}_\ell,k})$ **then**
        $\mathcal{A}_\ell \leftarrow \mathcal{A}_\ell \setminus \{k\}$; **break**
      end if
      if $n_{\hat{k}_\ell,k} \geq n_\ell$ **then break**
      end if
      compare $\hat{k}_\ell$ and k; update $\hat{P}(\hat{k}_\ell,k)$ and $n_{\hat{k}_\ell,k}$
    end loop
  end for
  $\mathcal{A}_{\ell+1} \leftarrow \mathcal{A}_\ell$
  if $|\mathcal{A}_{\ell+1}| \geq 3$ **and** $\max\{2^i : 2^{i+1} \leq |\mathcal{A}_{\ell+1}| - 1\} = k_{\ell,-}$ **then** $m \leftarrow m + 1$ ▷ level unchanged: refine precision
  end if
  $\ell \leftarrow \ell + 1$
end while
if $\mathcal{A}_\ell = \{a\}$ **return** a; **else** $\mathcal{A}_\ell = \{a, b\}$: compare a and b until $n_{a,b}\, \mathrm{kl}(\hat{P}(a,b) \| \frac{1}{2}) \geq f(n_{a,b})$ and **return** the empirical winner

Algorithm. DUELELIM (Alg. 2) proceeds in phases, and each phase has two steps: choose a pivot arm, then eliminate the arms that lose to it. At phase ℓ, with $K_\ell = |\mathcal{A}_\ell| \geq 3$ alive arms, the algorithm first defines a *level* from the number $K_\ell - 1$ of opponents of an arm,

$$k_{\ell,-} := \max\{2^i :\ 2^{i+1} \leq K_\ell - 1,\ i \geq 0\}, \qquad k_{\ell,+} := 2k_{\ell,-}, \qquad i_\ell := \log_2 k_{\ell,-}, \tag{4}$$

so that $2^{i_\ell+1} \leq K_\ell - 1 < 2^{i_\ell+2}$, $k_{\ell,+} \leq K_\ell - 1$ and $K_\ell \leq 2^{i_\ell+2}$, together with a *precision* $\varepsilon_\ell = 2^{-m}$. In the first step, it runs MME on every column to obtain, for each alive arm k, a value $\hat{m}_\ell(k)$ that lies between the $k_{\ell,-}$-th and the $k_{\ell,+}$-th smallest gap of k against its current opponents, up to ε_ℓ; the *pivot* is the arm that maximises $\hat{m}_\ell$. In the second step, it compares the pivot with every other alive arm until one of two things happens: either the KL test of Lem. 2.1 certifies that the pivot beats the arm, and the arm is eliminated, or a budget $n_\ell \asymp \varepsilon_\ell^{-2}\lambda$ is reached.

Why does this make progress? Since $\Delta_\ell(k^*, k_{\ell,-}) \geq \Delta_{(k_{\ell,-})}$ and the pivot maximises $\hat{m}_\ell$, the pivot's $k_{\ell,+}$-th smallest gap is at least $\Delta_{(k_{\ell,-})} - 2\varepsilon_\ell$. When this quantity exceeds ε_ℓ, all but $k_{\ell,+} - 1$ opponents of the pivot are eliminated,

and the level decreases. The precision is halved only when a phase fails to decrease the level, which makes the cost of the phases at a given level a geometric series governed by the gap $\Delta_{(2^i)}$. When at most two arms remain, a final duel, decided by the same test, returns the winner.

Theorem 3.1 (Exact identification). *Let $\delta \in (0, e^{-8})$, $K \geq 2$, and assume that $\mathbf{P}$ admits a Condorcet winner k^*. There is an absolute constant C such that, with probability at least $1-\delta$,* DUELELIM *returns k^* after at most $C \sum_{k \neq k^*} \frac{1}{\Delta_k^2} \Big(\ln \frac{K}{\delta} + \ln \ln \frac{1}{\Delta_k} \Big)$ comparisons.*

Proof sketch. The full proof is in App. D. Let $\mathcal{G}$ be the event on which all MME calls succeed (its probability is at least $1-\delta/2$, since $\sum_\ell K_\ell \delta_\ell = \delta/2$) and the event $\mathcal{E}$ of Lem. 2.1 holds for every pair (probability at least $1-\delta/4$). On $\mathcal{G}$, three facts hold.

(a) Soundness and power of the elimination step (Lem. D.2). By Lem. 2.1(i), an arm is eliminated by a pivot j only if $P(j,k) > \frac{1}{2}$, so k^* is never eliminated and the final duel returns k^*. By Lem. 2.1(ii) with $\gamma = \varepsilon_\ell$ and the choice of n_ℓ (Lem. B.2: $n_\ell \varepsilon_\ell^2 \geq 2 f(n_\ell)$), every arm k with $\Delta(\hat{k}_\ell, k) \geq \varepsilon_\ell$ is eliminated during phase ℓ.

(b) Level decrease (Lem. D.4). Since $\hat{k}_\ell$ maximises $\hat{m}_\ell$ and k^* is alive, the MME guarantees give $\Delta_\ell(\hat{k}_\ell, k_{\ell,+}) \geq \Delta_\ell(k^*, k_{\ell,-}) - 2\varepsilon_\ell \geq \Delta_{(2^i)} - 2\varepsilon_\ell$ with $i = i_\ell$. If $\Delta_{(2^i)} \geq 3 \cdot 2^{-m}$, all opponents of the pivot ranked 2^{i+1} or higher in its column have gap at least ε_ℓ and are eliminated by (a), so $|\mathcal{A}_{\ell+1}| \leq 2^{i+1}$ and the level decreases. Contrapositively, the precision is refined at level i only while $2^{-m} > \Delta_{(2^i)}/3$.

(c) Cost. A phase at precision 2^{-m} costs at most $O\big(K_\ell 4^m \ln(7K^2\ell^2/\delta)\big)$ comparisons (Lem. D.3): K_ℓ calls to MME at $O(4^m \ln(1/\delta_\ell))$ each, plus at most $K_\ell n_\ell$ elimination comparisons. Since every phase either decreases the level or refines the precision, each pair (level, precision) is visited at most once. At level i, the largest precision m_i^+ satisfies $4^{m_i^+} \leq 36/\Delta_{(2^i)}^2$ by (b), and $K_\ell \leq 2^{i+2}$; hence the phases at level i cost $O(2^i \Delta_{(2^i)}^{-2} \Lambda_i)$ in total, where $\Lambda_i \lesssim (\ln \frac{K}{\delta} + \ln \ln \Delta_{(2^i)}^{-1})$ because $\ell \leq \log_2 K + m_i^+$. It remains to turn the sum over levels into a sum over arms: the 2^{i-1} gaps of the block $\{2^{i-1}+1, \dots, 2^i\}$ are all at most $\Delta_{(2^i)}$, so $2^i \Delta_{(2^i)}^{-2} \leq 2 \sum_{k=2^{i-1}+1}^{2^i} \Delta_{(k)}^{-2}$. Summing over levels and adding the final duel (3) gives the claim. □

Discussion. Table 1 in App. A summarises the known guarantees. Compared with the pair-sum bounds of SAVAGE and PBR, Thm. 3.1 saves a factor K: thanks to MME, each phase spends $O(K_\ell \varepsilon_\ell^{-2} \lambda)$ comparisons instead of $O(K_\ell^2 \varepsilon_\ell^{-2} \lambda)$. We recover the bound of order $\tilde{O}(H_{\mathrm{cw}}(\delta))$ of Maiti et al. (2024) and Saad et al. (2026) with a $\log(K)$ factor removed. On the other hand, we do not recover the refined trade-off of Saad et al. (2026) that may be more favorable in some instances where the arms may be eliminated by stronger opponents than the Condorcet winner. The $\ln \ln \frac{1}{\Delta_k}$ term is the usual price of time-uniform tests (Jamieson et al., 2014); under strong stochastic transitivity, Ren et al. (2020) obtain the same structure with a matching lower bound.

What is new in DUELELIM. The gain of a $\ln K$ factor does not come from the estimator but from the structure of the algorithm. Both Maiti et al. (2024) and Saad et al. (2026) follow a *search–verify–restart* scheme: a search step returns a candidate, which may be wrong; a verification step then checks this candidate against all other arms; and if the check fails, the algorithm restarts with a halved precision or a doubled budget. In this scheme, correctness comes from the verification step, and the search step may drop k^*. In DUELELIM, correctness comes from the elimination test itself: an arm is eliminated only when a sequential test certifies that the pivot beats it, so k^* is never eliminated, whatever the pivot, and the last survivor is certified by construction. There is no verification pass, no restart and no global doubling; the precision is refined only when a phase fails to change the level, and counters are never reset, so each pair stops when its own test fires and arm k costs about Δ_k^{-2} comparisons. The quantile estimate is only used to choose the pivot, and only its progress guarantee is needed; the pivot may even be chosen heuristically without touching the correctness guarantee (App. H). Finally, the analysis uses only two properties: the test is sound and powerful at an arbitrary reference point, and the pivot's column is sandwiched between two of its order statistics. This is what makes the extension to approximate identification possible in the next section, where the reference point is shifted and the winner's gaps may be negative. App. A details the comparison.

4 IDENTIFICATION OF AN APPROXIMATE CONDORCET WINNER

In this section we relax the assumption that an exact Condorcet winner exists, and we ask for an approximate one instead, which always exists.

Definition 4.1 (Approximate Condorcet winner). *For $c \in (0, \frac{1}{2})$, an arm $\hat{k}$ is a* c-approximate Condorcet winner *(or c-Condorcet winner) if $P(\hat{k}, k) > \frac{1}{2} - c$ for all $k \neq \hat{k}$.*

Assumption 4.2. *For a known $\varepsilon \in (0, \frac{1}{16}]$, $\mathbf{P}$ admits an ε-Condorcet winner.*

Two remarks explain why the output must also be relaxed. First, an ε-Condorcet winner is not unique in general. Second, whether a given arm is one cannot be decided from finitely many samples, because $P(\hat{k}, k) = \frac{1}{2} - \varepsilon$ cannot be distinguished from a slightly smaller value. As in (ε, δ)-PAC bandits (Even-Dar et al., 2006), we therefore ask for an ε'-Condorcet winner for some target $\varepsilon' > \varepsilon$.

More precisely, let $\mathcal{W} := \{w : P(w, k) > \frac{1}{2} - \varepsilon \; \forall k \neq w\}$ be the (non-empty) set of ε-Condorcet winners, and fix a target $\varepsilon' > \varepsilon$. The goal is to return an ε'-approximate Condorcet winner with probability at least $1 - \delta$. To state the cost, we need some notation. For $w \in \mathcal{W}$ write $\Delta_{w,k} := P(w, k) - \frac{1}{2} \in (-\varepsilon, \frac{1}{2}]$, which may now be *negative*, and $\Delta_{w,(1)} \leq \cdots \leq \Delta_{w,(K-1)}$ for these values in sorted order. Let also

$$N_w(x) := |\{k \neq w : \; \Delta_{w,k} \leq x\}|$$

be the number of arms within x of w.

4.1 Algorithm and upper bound

Our algorithm DuelElim-ε has two stages. Both use a resolution θ and a cap L defined as

$$\theta := \max(\varepsilon, \varepsilon'/4) \; \in \; [\varepsilon, \tfrac{1}{16}], \qquad L := \lceil \log_2(1/\theta) \rceil.$$

The resolution must be at least ε for Stage 1 to be sound, i.e., to not eliminate ϵ-CW, as explained below; taking it as large as $\varepsilon'/4$ costs nothing in the final bound and makes Stage 1 cheaper.

Stage 1: recentred elimination. This stage is Alg. 2 with two changes (the full pseudocode is Alg. 4 in App. E). First, the KL test of line 9 uses the reference point $\frac{1}{2} + \theta$ instead of $\frac{1}{2}$: an arm is eliminated only when the pivot is certified to beat it by more than θ. This change is needed because, an arm may beat a member of $\mathcal{W}$ by up to $\varepsilon \leq \theta$; the recentring guarantees that $\mathcal{W}$ survives, and in exchange every eliminated arm is certified not to be a θ-Condorcet winner. Second, the precision is never refined below 2^{-L}, since gaps smaller than θ do not need to be resolved. Stage 1 stops at the first phase that fails to decrease the level at that precision, or when at most two arms remain. At that point, the number of survivors is at most four times the number of arms within $4\theta = \max(\varepsilon', 4\varepsilon) \leq 2\varepsilon'$ of any $w \in \mathcal{W}$. This is the one place where the slack limits the resolution: arms within 4ε of w cannot be told apart from w by Stage 1, whatever the target.

Stage 2: certification. Being an ε'-approximate Condorcet winner is a property with respect to *all* arms, including the eliminated ones, so a duel between the survivors is not enough. Instead, Certify (Alg. 5, App. E) tests each survivor $a \in \mathcal{A} \supseteq \mathcal{W}$ against every arm $j \neq a$ with two one-sided tests: the pair *passes* if $P(a, j) > \frac{1}{2} - \varepsilon'$ is certified, and a is *discarded* if $P(a, j) < \frac{1}{2} - \varepsilon$ is certified. Comparisons are allocated to the candidates in rounds of doubling budgets 2^t, and the first candidate whose $K - 1$ tests all pass is returned. This stage is the verification step of explore-then-verify schemes (Karnin, 2016), with the exact test replaced by two tests separated by $\varepsilon' - \varepsilon$; as a result, every pair resolves in $\tilde{O}((\varepsilon' - \varepsilon)^{-2})$ comparisons.

Theorem 4.3 (Approximate identification)**.** *Let $\delta \in (0, e^{-8})$, let Assumption 4.2 hold with slack ε, and let $\varepsilon' \in (\varepsilon, \frac{1}{4}]$. There is an absolute constant C such that, with probability at least $1 - \delta$,* DuelElim-ε *run with parameters $(\delta, \varepsilon, \varepsilon')$ returns an ε'-approximate Condorcet winner after at most*

$$C \left(1 + \min_{w \in \mathcal{W}} N_w(\max\{\varepsilon', 4\varepsilon\})\right) \min_{w \in \mathcal{W}} \sum_{k \neq w} \frac{1}{(\Delta_{w,k} + \varepsilon')^2} \left(\ln \frac{K}{\delta} + \ln \ln \frac{24}{(\varepsilon' - \varepsilon)}\right) \quad \textit{comparisons.}$$

The proof is postponed to App. E. The bound may seem surprising at first sight: it carries a multiplicative factor $1 + N_w(4\theta)$ that has no counterpart in Thm. 3.1, although one would expect approximate identification to be easier than exact identification. When some near-winner w is isolated, there is nothing surprising: $N_w(4\theta) = O(1)$ and Thm. 4.3 is the exact analogue of Thm. 3.1 with Δ_k replaced by $\Delta_k + \varepsilon'$, as in ε-best-arm identification (Kalyanakrishnan et al., 2012), and Proposition 4.4 below shows that it is tight up to logarithms. In particular, when an exact Condorcet winner k^* exists, Assumption 4.2 holds for every $\varepsilon > 0$, so ε and ε' can be taken small enough that $4\theta < \Delta_{(1)}$; then $N_{k^*}(4\theta) = 0$ and we recover a bound of the same form as Thm. 3.1. Assumption 4.2 is much weaker than the existence of an exact winner, however, and the factor is the price of this generality. When a cluster of near-tied arms exists, Stage 2 multiplies the certification cost by the size of the cluster: a surviving decoy may lose badly to an arm that was eliminated long ago, and only a comparison with that arm can detect this. Proposition 4.5 shows that this is not an artefact of the analysis: a multiplicative dependence on $N_w(4\theta)$ cannot be avoided in general. Note also that only Stage 1 benefits from the presence of several near-winners. Finally, θ balances the two stages – a smaller resolution makes Stage 1 more expensive but reduces the number of survivors that Stage 2 must certify – and choosing it adaptively is left for future work.

4.2 LOWER BOUNDS

Call an algorithm $(\varepsilon, \varepsilon', \delta)$-*correct*, for $\varepsilon' > \varepsilon$, if on every instance satisfying Assumption 4.2 with slack ε it stops almost surely and returns an ε'-approximate Condorcet winner with probability at least $1-\delta$. We write τ for its number of comparisons. Our first lower bound shows that the certification term of Thm. 4.3 is necessary.

Proposition 4.4 (Lower bound: certification)**.** *Let $0 < \varepsilon < \varepsilon' \le \frac{1}{4}$, $\delta \in (0, \frac{1}{2})$, and let ν be an instance with a Condorcet winner k^* whose gaps satisfy $\Delta_k > \varepsilon'$ for all $k \neq k^*$ and $P(j,k) = \frac{1}{2}$ for all $j, k \neq k^*$. Then k^* is the unique ε'-approximate Condorcet winner of ν, $N_{k^*}(\varepsilon') = 0$, and every $(\varepsilon, \varepsilon', \delta)$-correct algorithm satisfies, for an absolute constant $c > 0$,*

$$\mathbb{E}_\nu[\tau] \;\ge\; c \ln \frac{1}{2.4\delta} \sum_{k \neq k^*} \frac{1}{\max(\Delta_k, \varepsilon')^2}.$$

The proof (App. E.7) is a change-of-measure argument (Kaufmann et al., 2016). Lowering $P(k^*, k)$ to $\frac{1}{2} - \varepsilon'$ makes k the unique correct answer while keeping Assumption 4.2 true, so the pair (k^*, k) must be compared $\Omega((\Delta_k + \varepsilon')^{-2} \ln \frac{1}{\delta})$ times; the claim follows since $\Delta_k + \varepsilon' \le 2\max(\Delta_k, \varepsilon')$. The argument says nothing about arms with $\Delta_k < \varepsilon'$, and for a good reason: such an arm is an incorrect answer only if it has a *beater* j with $P(k,j) \le \frac{1}{2} - \varepsilon'$. In that case, a change of measure on the two cells (k^*, k) and (k, j) shows that every correct algorithm must either certify the winner against k, which costs $\max(\Delta_k, \varepsilon')^{-2} \ln \frac{1}{\delta}$ comparisons, or refute k through its beater, which costs $(\frac{1}{2} - P(k,j))^{-2} \ln \frac{1}{\delta}$ comparisons (Proposition E.8 in App. E.7). DUELELIM-ε always chooses the first option. The next bound shows that when many arms are within ε' of the winner and their beaters are hidden, the second option cannot be avoided either, so that $\sum_k \max(\Delta_k, \varepsilon')^{-2}$ alone cannot be the sample complexity.

Proposition 4.5 (Lower bound)**.** *Let $\varepsilon \in (0, \frac{1}{16})$, $\varepsilon' \in (\varepsilon, \frac{1}{4}]$, $\delta \in (0, \frac{1}{4}]$ and $K \ge 16$ even. There is a family $\mathcal{I}_K$ of K-armed instances, each satisfying Assumption 4.2 with slack ε and having a unique ε'-approximate Condorcet winner w_ν, with $N_{w_\nu}(4\varepsilon) = \frac{K}{2} - 1$ and $\sum_{k \neq w_\nu} \max(\Delta_k, \varepsilon')^{-2} \le \frac{K}{2\varepsilon'^2} + 8K$, such that every $(\varepsilon, \varepsilon', \delta)$-correct algorithm satisfies, for an absolute constant $c > 0$,*

$$\max_{\nu \in \mathcal{I}_K} \mathbb{E}_\nu[\tau] \;\ge\; c\,K^2 \;\ge\; c\,\varepsilon'^2 K \max_{\nu \in \mathcal{I}_K} \sum_{k \neq w_\nu} \frac{1}{\max(\Delta_k, \varepsilon')^2}.$$

To read this result, take a constant target, say $\varepsilon' = \frac{1}{4}$. On $\mathcal{I}_K$ the sum $\sum_k \max(\Delta_k, \varepsilon')^{-2}$ is of order K, while every correct algorithm needs $\Omega(K^2)$ comparisons. Therefore an upper bound of order $\sum_k \max(\Delta_k, \varepsilon')^{-2}$ up to logarithms, like the one of Thm. 3.1 for exact identification, is impossible under the weaker Assumption 4.2. Moreover, since the bound of Thm. 4.3 is $O(K \sum_k \max(\Delta_k, \varepsilon')^{-2})$ up to logarithms on $\mathcal{I}_K$, its factor $1 + N$ (here $N = \frac{K}{2} - 1$) is necessary up to logarithmic factors. This factor ε'^2 is where upper and lower bounds may disagree for small targets. The lower bound is attained by an algorithm that knows the family (it abandons a candidate row as soon as it finds its needle), but such an algorithm is not correct on other instances; a correct algorithm must certify every discard, either against w or through the decoy's needle, and DUELELIM-ε pays both. Whether the factor N can be reduced to $\varepsilon'^2 N$ is left open.

The instances and proof sketch. The arms are split into $n = \frac{K}{2}$ *candidates* and $m = \frac{K}{2}$ *fillers*. Candidates tie with each other, and fillers tie with each other. Every candidate beats every filler with probability $\frac{3}{4}$, except in one *special* cell per candidate row: the winner w beats its special filler with probability 1, whereas every other candidate a loses to its special filler, $P(a, \sigma(a)) = \frac{1}{2} - \varepsilon'$ (a "needle"). Thus w is a Condorcet winner, all other candidates have $\Delta_a = 0$, and each of them is disqualified by a single hidden cell among m. The proof (App. E.6) is a Bayesian argument: w and the needles are drawn uniformly at random, and a genie reveals the type of each candidate–filler cell at its first comparison, which can only help the algorithm. Conditionally on the past, the candidate rows whose special cell has not been found yet are exchangeable. Correctness with probability at least $\frac{3}{4}$ therefore forces the algorithm to discover the special cell of at least $n/12$ rows in expectation (Eq. (18)). By a coupling with the run in which the row has no special cell, each such discovery costs $m/2$ comparisons in expectation (Eq. (19)). Together, $\mathbb{E}[\tau] \ge nm/24 = K^2/96$. □

Comparison with existing hardness results. In classical bandits, near-optimal arms enter the complexity *additively*, and many good arms even make the problem easier when the output need not be certified (Katz-Samuels & Jamieson, 2020). Here, in contrast, they enter *multiplicatively*. The closest existing result is Falahatgar et al. (2018, Thm 1): under weak stochastic transitivity, a $\frac{1}{4}$-maximum requires $\Omega(K^2)$ comparisons, through a construction where every arm has a single hidden beater.

5 QUANTILE CONDORCET WINNERS: A K-FREE GUARANTEE

The previous section showed that certifying an approximate Condorcet winner is a needle-in-a-haystack problem, because the property is a *minimum* over the candidate's row. In this section we show that both the dependence on K and the factor N disappear if the minimum is relaxed into a *quantile*. The reason is simple: a quantile of a row, unlike its minimum, can be estimated by sampling a few random opponents, exactly as MME does. For $a \in [K]$ and $c > 0$, let $B_c(a) := \{j \neq a : P(a,j) \leq \frac{1}{2} - c\}$ be the set of opponents that beat a by at least c, and let $\beta_c(a) := |B_c(a)|/(K-1)$ be their fraction.

Definition 5.1 (Quantile Condorcet winner). *For $c \in (0, \frac{1}{2})$ and $\eta \in [0,1]$, an arm a is a (c,η)-*Condorcet winner *if $\beta_c(a) \leq \eta$. We write $\mathcal{W}_{c,\eta}$ for the set of such arms and $\rho_{c,\eta} := |\mathcal{W}_{c,\eta}|/K$ for their proportion.*

The parameter η interpolates between two extremes. A $(c,0)$-Condorcet winner is exactly a c-approximate Condorcet winner, and $\mathcal{W}_{c,\eta} = \mathcal{W}_{c,0}$ as soon as $\eta < 1/(K-1)$. At the other end, a $(c, \frac{1}{2})$-Condorcet winner always exists: each pair contributes at most one element to $\bigcup_a B_c(a)$, so $\sum_a \beta_c(a) \leq K/2$ (Lem. F.1). This is why Thm. 5.2 lets η range over $(0, \frac{1}{2}]$. The relaxation is in the spirit of the count-type target of SAVAGE (Urvoy et al., 2013) and of the Copeland winner (Zoghi et al., 2015): as $c \to 0$, a (c,η)-Condorcet winner is an arm with Copeland score at least $(1-\eta)(K-1)$. Unlike the Copeland target, however, ours keeps a margin c and is a *set* rather than an argmax, which is what makes it certifiable by sampling.

Given ε and η, the goal is to return a $(2\varepsilon, 2\eta)$-Condorcet winner under the assumption that $\mathcal{W}_{\varepsilon,\eta} \neq \emptyset$; the proportion $\rho := \rho_{\varepsilon,\eta}$ is unknown. Both relaxations, in c and in η, are needed for a K-free guarantee. Relaxing c and η are unavoidable for the usual reason: two values that differ slightly cannot be told apart with finitely many samples.

Algorithm 3 QUANTILECW

Require: confidence δ, slack $\varepsilon \in (0, \frac{1}{4})$, quantile $\eta \in (0, \frac{1}{2}]$; $\gamma \leftarrow \eta/32$, $N \leftarrow \lceil \frac{2}{\varepsilon^2} \ln \frac{32}{\eta} \rceil$, $t \leftarrow \frac{11}{8}\eta$

1: **for** $i = 1, 2, \ldots$ **do**
2: $\quad \delta_i \leftarrow \frac{3\delta}{\pi^2 i^2}$, $M_i \leftarrow \lceil \frac{12}{\eta} \max\{3, \ln \frac{1}{\delta_i}\} \rceil$
3: $\quad$ draw a candidate a_i uniformly in $[K]$ and opponents $j_{i,1}, \ldots, j_{i,M_i}$ i.i.d. uniformly in $[K] \setminus \{a_i\}$
4: $\quad$ compare a_i with each $j_{i,l}$ exactly N times; let $\hat{p}_{i,l}$ be the empirical win frequency of a_i
5: $\quad \hat{B}_i \leftarrow |\{l \leq M_i : \hat{p}_{i,l} < \frac{1}{2} - \frac{3}{2}\varepsilon\}|$; **if** $\hat{B}_i \leq t M_i$ **return** a_i
6: **end for**

QUANTILECW (Alg. 3) repeatedly draws a random candidate and tests a quantile of its row on a random sample of opponents. Two design choices keep the cost low. First, each label $\mathbf{1}\{\hat{p}_{i,l} < \frac{1}{2} - \frac{3}{2}\varepsilon\}$ is allowed to be wrong with probability $\gamma = \eta/32$; these errors are absorbed into the count instead of being removed by a union bound, so that N depends on η only through $\ln(32/\eta)$. Second, only soundness, that is, accepting a candidate which is not a $(2\varepsilon, 2\eta)$-winner, consumes the confidence budget δ_i; a good candidate is accepted with constant probability, which only affects the number of rounds. (A round may also stop as soon as $\hat{B}_i > t M_i$.)

Theorem 5.2 (K-free identification of a quantile Condorcet winner). *Let $\varepsilon \in (0, \frac{1}{4})$, $\eta \in (0, \frac{1}{2}]$, $\delta \in (0, \frac{1}{2}]$, and let $\mathbf{P}$ be any preference matrix with $\mathcal{W}_{\varepsilon,\eta} \neq \emptyset$; set $\rho := \rho_{\varepsilon,\eta}$ and $I_\rho := \lceil \frac{4}{3\rho} \ln \frac{2}{\delta} \rceil$. With probability at least $1-\delta$,* QUANTILECW *stops after at most I_ρ rounds, returns a $(2\varepsilon, 2\eta)$-Condorcet winner, and uses at most*

$$I_\rho \, M_{I_\rho} \, N \;\leq\; \frac{C}{\rho\,\eta\,\varepsilon^2} \ln \frac{2}{\delta} \ln \frac{4 I_\rho^2}{\delta} \ln \frac{32}{\eta} \;=\; O\!\left(\frac{1}{\rho\,\eta\,\varepsilon^2} \ln \frac{1}{\delta} \ln \frac{1}{\rho\delta} \ln \frac{1}{\eta} \right) \quad \textit{comparisons, for an absolute constant } C.$$

Proof sketch. The full proof is in App. F. Fix a round with candidate a. By Hoeffding's inequality, an opponent $j \notin B_\varepsilon(a)$ is labelled "bad" with probability at most γ, and an opponent $j \in B_{2\varepsilon}(a)$ is labelled "not bad" with probability at most γ. Since the opponents are i.i.d. and receive fresh comparisons, the labels are i.i.d. Bernoulli variables with a mean $q \in [\beta_{2\varepsilon}(a) - \gamma, \beta_\varepsilon(a) + \gamma]$, so that $\hat{B}_i \sim \mathrm{Bin}(M_i, q)$. For a good candidate, $q \leq \frac{33}{32}\eta < t$; for a candidate that is not a $(2\varepsilon, 2\eta)$-winner, $q \geq \frac{63}{32}\eta > t$. The Chernoff–KL bound $e^{-M \mathrm{kl}(t\|q)}$, combined with $\mathrm{kl}(t\|q) \geq (t-q)^2/(2\max(t,q))$, then shows that a good candidate is accepted with probability at least $\frac{3}{4}$ (because $M_i \geq 36/\eta$) and a bad one with probability at most δ_i (because $M_i \geq \frac{12}{\eta} \ln \frac{1}{\delta_i}$). Rounds are independent, and each one accepts a good candidate with probability at least $\frac{3}{4}\rho$, so the algorithm stops within I_ρ rounds with probability $1 - \delta/2$. Since $\sum_i \delta_i = \delta/2$, wrong acceptances are also controlled, and round i uses $M_i N$ comparisons. □

Discussion. The number of arms enters the bound only through the proportion ρ of good arms: the more (ε, η)-Condorcet winners there are, the cheaper the identification. This is the behaviour one expects from a problem with many optima (de Heide et al., 2021), and the opposite of Thm. 4.3. The two results are consistent: when $\eta < 1/(2(K-1))$, the output is a 2ε-approximate Condorcet winner and the bound becomes of order $K/(\rho\varepsilon^2)$, which recovers the dependence on K that Proposition 4.5 shows to be unavoidable. The algorithm itself is the dueling analogue of the "sample a random arm and certify it" strategies used for quantile identification in classical bandits (Roy Chaudhuri & Kalyanakrishnan, 2017), where the certificate of a single arm is replaced by a quantile test on its row, and where the two-confidence idea of Ren et al. (2019) is built in. The $\Omega(\frac{1}{\rho\varepsilon^2}\ln\frac{1}{\delta})$ lower bounds of that literature (Roy Chaudhuri & Kalyanakrishnan, 2017, Cor. 3.4) suggest that the dependence on ρ and ε is necessary, and a proportion test between η and 2η suggests that the factor $1/\eta$ is necessary too. Formally, however, we only have Proposition 4.4 (with $K = 2$), and a matching lower bound is left open. Finally, the fixed budget N can be replaced by the sequential tests of Lem. 2.1 at per-pair confidence γ without changing the analysis; opponents far from the threshold then cost only $O(\ln\frac{1}{\eta})$ comparisons (App. J).

6 Experiments

We report extensive experiments on synthetic preference matrices validating all three algorithms against certified and heuristic baselines. We compare against MidSearch and MidSearch+Verify (Maiti et al., 2024) and the certified algorithm FC-CWI (Saad et al., 2026), each implemented from the published pseudocode and, for MidSearch, from the authors' public repository. Due to space constraints, the complete setup, all plots and all tables are deferred to App. G–App. J; we summarise the main findings below.

- **Pivot bottleneck and the DuelElim-3T variant** (Apps. H.1 and H.2). In the analysed DuelElim (Algorithm 2), pivot selection via MME accounts for 88–99% of all comparisons; the elimination loop itself is cheap. DuelElim-3T (Algorithm 6) fixes this by trying a free "previous pivot" tier and a cheap τ-shrunk MME tier before falling back to the certified MME, inheriting the guarantee of Theorem 3.1 up to a factor 3 (Proposition H.1). Empirically it is **75–170× faster** than the analysed algorithm on the Figure-2 benchmark.
- **Exact identification vs. Maiti et al. (2024) and Saad et al. (2026)** (App. H.5, Figures 1–2, Table 2). DuelElim-3T returns k^* with a δ-guarantee in every run within 3–4 H_1 at $d = 1024$, against >1000 H_1 for MidSearch+Verify (whose ColMidVal subroutine alone requires $\approx 5.8 \times 10^9$ comparisons at $d = 1024$, far exceeding the total budget) and >100 H_1 for FC-CWI (whose FB-CWI step spends 5–25× its nominal budget at each doubling round). On the heuristic side, DuelElim-τ reaches 100% accuracy at 0.75–1 H_1 on Figure-2, outpacing MidSearch, but the two heuristics trade off on Figure-1: neither dominates, and neither carries a guarantee.
- **Approximate identification: cost vs. target** ε' (App. I.3, Figure 3, Table 3). DuelElim-ε-3T is 3.5× cheaper than exact identification in the regime $\Delta_{\min} \ll \varepsilon' \ll \beta$, confirming Theorem 4.3: Stage 1 cost is flat in ε' below β (the bad arms always cost the same), while Certify's cost grows as $\varepsilon' \to 0$. For large $\varepsilon' \geq \beta$, Certify$([K])$ alone is cheaper; the *lockstep* combination of DuelElim-ε-3T and Certify$([K])$ (Appendix I.2) stays within a factor 2 of the better algorithm over the *entire* range of ε', at the price of running both in parallel on separate statistics.
- **Approximate identification: scaling in** K (App. I.4, Figure 4, Table 4). DuelElim-ε-3T scales as $O(K)$ (Stage 1 eliminates all bad arms against a single pivot) while Certify$([K])$ scales as $O(K^{1.9})$, with a crossover already below $K = 32$.
- **K-free guarantee for QuantileCW** (Apps. J.3 and J.4, Figures 5–6, Tables 5–6). QuantileCW is flat in K from $K = 10^3$ to $K = 10^5$ on both Family A (easy pairs) and Family B (pairs close to the label threshold), while the row-scan baseline grows linearly and crosses it at $K \approx 10^3$. The QuantileCW-SEQ variant, which replaces the fixed per-pair budget with anytime KL tests (App. J.2), is 3–4× cheaper at large K by resolving easy pairs early; at $K = 3 \cdot 10^4$ it is 140× cheaper than the sequential row scan on Family A.

7 Conclusion

We gave an explicit high-probability guarantee for exact Condorcet winner identification, and the first guarantees for approximate identification under the Condorcet assumption alone. These consist of an upper bound in terms of the effective gaps $\max(\Delta_k, \varepsilon')$, which is tight up to logarithms when the near-winner is isolated; a lower bound showing that near-tied arms otherwise enter the complexity multiplicatively; and a K-free guarantee for the quantile relaxation. Several questions remain open: whether the factor N of Thm. 4.3 can be reduced to $\varepsilon'^2 N$, whether an instance-dependent guarantee is possible when the slack is unknown, whether the exact bound can be sharpened to $H_{\mathrm{cw}}(\delta)$ while remaining explicit, and whether a matching lower bound holds for quantile Condorcet winners.

REFERENCES

Maryam Aziz, Jesse Anderton, Emilie Kaufmann, and Javed Aslam. Pure exploration in infinitely-armed bandit models with fixed-confidence. In *Proceedings of the 29th International Conference on Algorithmic Learning Theory (ALT)*, volume 83 of *PMLR*, pp. 3–24, 2018. URL https://proceedings.mlr.press/v83/aziz18a.html.

Viktor Bengs, Róbert Busa-Fekete, Adil El Mesaoudi-Paul, and Eyke Hüllermeier. Preference-based online learning with dueling bandits: A survey. *Journal of Machine Learning Research*, 22(7):1–108, 2021. URL https://jmlr.org/papers/v22/18-546.html.

Róbert Busa-Fekete, Balázs Szörényi, Paul Weng, Weiwei Cheng, and Eyke Hüllermeier. Top-k selection based on adaptive sampling of noisy preferences. In *Proceedings of the 30th International Conference on Machine Learning (ICML)*, volume 28 of *PMLR*, 2013. URL https://proceedings.mlr.press/v28/busa-fekete13.html.

Wei-Lin Chiang, Lianmin Zheng, Ying Sheng, Anastasios N. Angelopoulos, Tianle Li, Dacheng Li, Banghua Zhu, Hao Zhang, Michael I. Jordan, Joseph E. Gonzalez, and Ion Stoica. Chatbot arena: An open platform for evaluating LLMs by human preference. In *Proceedings of the 41st International Conference on Machine Learning (ICML)*, 2024.

Paul F. Christiano, Jan Leike, Tom B. Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. In *Advances in Neural Information Processing Systems 30 (NIPS)*, 2017.

Rianne de Heide, James Cheshire, Pierre Ménard, and Alexandra Carpentier. Bandits with many optimal arms. In *Advances in Neural Information Processing Systems 34 (NeurIPS)*, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/bd33f02c4e28615b5af2d24703e066d5-Abstract.html.

Miroslav Dudík, Katja Hofmann, Robert E. Schapire, Aleksandrs Slivkins, and Masrour Zoghi. Contextual dueling bandits. In *Proceedings of the 28th Conference on Learning Theory (COLT)*, volume 40 of *PMLR*, pp. 563–587, 2015. URL https://proceedings.mlr.press/v40/Dudik15.html.

Eyal Even-Dar, Shie Mannor, and Yishay Mansour. PAC bounds for multi-armed bandit and Markov decision processes. In *Proceedings of the 15th Annual Conference on Computational Learning Theory (COLT)*, pp. 255–270, 2002.

Eyal Even-Dar, Shie Mannor, and Yishay Mansour. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of Machine Learning Research*, 7:1079–1105, 2006. URL https://jmlr.org/papers/v7/evendar06a.html.

Moein Falahatgar, Alon Orlitsky, Venkatadheeraj Pichapati, and Ananda Theertha Suresh. Maximum selection and ranking under noisy comparisons. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, volume 70 of *PMLR*, 2017. URL https://arxiv.org/abs/1705.05366.

Moein Falahatgar, Ayush Jain, Alon Orlitsky, Venkatadheeraj Pichapati, and Vaishakh Ravindrakumar. The limits of maxing, ranking, and preference learning. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, volume 80 of *PMLR*, pp. 1427–1436, 2018. URL https://proceedings.mlr.press/v80/falahatgar18a.html.

Aurélien Garivier and Olivier Cappé. The KL-UCB algorithm for bounded stochastic bandits and beyond. In *Proceedings of the 24th Conference on Learning Theory (COLT)*, volume 19 of *PMLR*, pp. 359–376, 2011.

Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Proceedings of the 29th Conference on Learning Theory (COLT)*, volume 49 of *PMLR*, pp. 998–1027, 2016. URL https://proceedings.mlr.press/v49/garivier16a.html.

Björn Haddenhorst, Viktor Bengs, and Eyke Hüllermeier. Identification of the generalized Condorcet winner in multi-dueling bandits. In *Advances in Neural Information Processing Systems 34 (NeurIPS)*, 2021. URL https://proceedings.neurips.cc/paper/2021/file/d9de6a144a3cc26cb4b3c47b206a121a-Paper.pdf.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sébastien Bubeck. lil' UCB: An optimal exploration algorithm for multi-armed bandits. In *Proceedings of the 27th Conference on Learning Theory (COLT)*, volume 35 of *PMLR*, pp. 423–439, 2014. URL https://proceedings.mlr.press/v35/jamieson14.html.

Kevin Jamieson, Sumeet Katariya, Atul Deshpande, and Robert Nowak. Sparse dueling bandits. In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2015. URL https://arxiv.org/abs/1502.00133.

Marc Jourdan, Rémy Degenne, and Emilie Kaufmann. An ε-best-arm identification algorithm for fixed-confidence and beyond. In *Advances in Neural Information Processing Systems 36 (NeurIPS)*, 2023. URL https://openreview.net/forum?id=OLk3F64eSg.

Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone. PAC subset selection in stochastic multi-armed bandits. In *Proceedings of the 29th International Conference on Machine Learning (ICML)*, 2012. URL https://icml.cc/2012/papers/359.pdf.

Zohar Karnin, Tomer Koren, and Oren Somekh. Almost optimal exploration in multi-armed bandits. In *Proceedings of the 30th International Conference on Machine Learning (ICML)*, volume 28 of *PMLR*, pp. 1238–1246, 2013. URL https://proceedings.mlr.press/v28/karnin13.html.

Zohar S. Karnin. Verification based solution for structured MAB problems. In *Advances in Neural Information Processing Systems 29 (NIPS)*, 2016. URL https://proceedings.neurips.cc/paper_files/paper/2016/hash/65b9eea6e1cc6bb9f0cd2a47751a186f-Abstract.html.

Julian Katz-Samuels and Kevin Jamieson. The true sample complexity of identifying good arms. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 108 of *PMLR*, pp. 1781–1791, 2020. URL https://proceedings.mlr.press/v108/katz-samuels20a.html.

Emilie Kaufmann and Wouter M Koolen. Mixture martingales revisited with applications to sequential tests and confidence intervals. *Journal of Machine Learning Research*, 22(246):1–44, 2021.

Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best-arm identification in multi-armed bandit models. *Journal of Machine Learning Research*, 17(1):1–42, 2016. URL https://jmlr.org/papers/v17/kaufman16a.html.

Gábor Lugosi and Shahar Mendelson. Mean estimation and regression under heavy-tailed distributions: A survey. *Foundations of Computational Mathematics*, 19:1145–1190, 2019.

Arnab Maiti, Ross Boczar, Kevin Jamieson, and Lillian Ratliff. Near-optimal pure exploration in matrix games: A generalization of stochastic bandits & dueling bandits. In *International Conference on Artificial Intelligence and Statistics*, pp. 2602–2610. PMLR, 2024.

Shie Mannor and John N. Tsitsiklis. The sample complexity of exploration in the multi-armed bandit problem. *Journal of Machine Learning Research*, 5:623–648, 2004. URL https://jmlr.org/papers/v5/mannor04b.html.

Arkadi S. Nemirovsky and David B. Yudin. *Problem Complexity and Method Efficiency in Optimization*. Wiley, 1983.

Wenbo Ren, Jia Liu, and Ness B. Shroff. Exploring k out of top ρ fraction of arms in stochastic bandits. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 89 of *PMLR*, 2019. URL http://proceedings.mlr.press/v89/ren19a.html.

Wenbo Ren, Jia Liu, and Ness B. Shroff. The sample complexity of best-k items selection from pairwise comparisons. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, volume 119 of *PMLR*, 2020. URL https://arxiv.org/abs/2007.03133.

Arghya Roy Chaudhuri and Shivaram Kalyanakrishnan. PAC identification of a bandit arm relative to a reward quantile. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*, 2017.

El Mehdi Saad, Victor Thuot, and Nicolas Verzelen. The sampling complexity of condorcet winner identification in dueling bandits. *arXiv preprint arXiv:2603.15189*, 2026.

Aadirupa Saha and Aditya Gopalan. PAC battling bandits in the Plackett-Luce model. In *Proceedings of the 30th International Conference on Algorithmic Learning Theory (ALT)*, 2019. URL https://arxiv.org/abs/1808.04008.

Balázs Szörényi, Róbert Busa-Fekete, Adil Paul, and Eyke Hüllermeier. Online rank elicitation for Plackett-Luce: A dueling bandits approach. In *Advances in Neural Information Processing Systems 28 (NIPS)*, 2015. URL https://proceedings.neurips.cc/paper/2015/file/7eacb532570ff6858afd2723755ff790-Paper.pdf.

Tanguy Urvoy, Fabrice Clerot, Raphaël Féraud, and Sami Naamane. Generic exploration and K-armed voting bandits. In *Proceedings of the 30th International Conference on Machine Learning (ICML)*, volume 28 of *PMLR*, 2013. URL https://proceedings.mlr.press/v28/urvoy13.html.

Yisong Yue and Thorsten Joachims. Interactively optimizing information retrieval systems as a dueling bandits problem. In *Proceedings of the 26th International Conference on Machine Learning (ICML)*, 2009. URL https://www.cs.cornell.edu/people/tj/publications/yue_joachims_09a.pdf.

Yisong Yue, Josef Broder, Robert Kleinberg, and Thorsten Joachims. The K-armed dueling bandits problem. *Journal of Computer and System Sciences*, 78(5):1538–1556, 2012.

Yichi Zhou, Jialian Li, and Jun Zhu. Identify the nash equilibrium in static games with random payoffs. In *International Conference on Machine Learning*, pp. 4160–4169. PMLR, 2017.

Masrour Zoghi, Shimon Whiteson, Rémi Munos, and Maarten de Rijke. Relative upper confidence bound for the K-armed dueling bandit problem. In *Proceedings of the 31st International Conference on Machine Learning (ICML)*, 2014. URL https://arxiv.org/abs/1312.3393.

Masrour Zoghi, Zohar S. Karnin, Shimon Whiteson, and Maarten de Rijke. Copeland dueling bandits. In *Advances in Neural Information Processing Systems 28 (NIPS)*, 2015. URL https://proceedings.neurips.cc/paper/2015/file/9872ed9fc22fc182d371c3e9ed316094-Paper.pdf.