Optimal PAC Multiple Arm Identification with Applications to Crowdsourcing

Yuan Zhou
Carnegie Mellon University
yuanzhou@cs.cmu.edu

Xi Chen
UC Berkeley
xichen1987@berkeley.edu

Jian Li
Tsinghua University
lijian83@mail.tsinghua.edu.cn

April 24, 2014

Abstract

We study the problem of selecting K arms with the highest expected rewards in a stochastic N-armed bandit game. Instead of using existing evaluation metrics (e.g., misidentification probability (Bubeck et al., 2013) or the metric in EXPLORE-K(Kalyanakrishnan & Stone, 2010)), we propose to use *the aggregate regret*, which is defined as the gap between the average reward of the optimal solution and that of our solution. Besides being a natural metric by itself, we argue that in many applications, such as our motivating example from the crowdsourcing, the aggregate regret bound is more suitable. We propose a new PAC algorithm, which, with probability at least $1-\delta$, identifies a set of K arms with regret at most ϵ. We provide a detailed analysis on the sample complexity of our algorithm. To complement, we establish a lower bound on the expected number of samples for Bernoulli bandits and show that the sample complexity of our algorithm matches the lower bound. Finally, we report experimental results on both synthetic and real data sets, which demonstrates the superior performance of the proposed algorithm.

1 Introduction

We study the multiple arm identification problem in a stochastic multi-armed bandit game. More formally, assume that we are facing a bandit with n alternative arms, where the ith arm is associated with an unknown reward distribution supported on $[0, 1]$ with mean θ_i. Upon each sample (or "pull") of a particular arm, the reward is an *i.i.d.* sample from the reward distribution. We sequentially decide which arm to pull next and then collect the reward by sampling that arm. The goal of our "top-K arm identification" problem is to identify a subset of K arms with the maximum total mean. The problem finds applications in a variety of areas, such as in industrial engineering (Koenig & Law, 1985), evolutionary computation (Schmidt et al., 2006) and medical domains (Thompson, 1933). Here, we highlight another application in *crowdsourcing*. In recent years, crowdsourcing services become increasingly popular for collecting labels of the data for many machine learning, data mining and analytical tasks. The readers may refer to (Raykar et al., 2010; Welinder et al., 2010; Karger et al., 2012; Zhou et al., 2012; Ho et al., 2013; Chen et al., 2013; Liu et al., 2013) and references therein for recent work on machine learning in crowdsourcing. In a typical crowdsourced labeling task, the requestor submits a batch of microtasks (e.g., unlabeled data) and the workers from the crowd are asked to complete the tasks. Upon each task completion, a worker receives a small monetary reward. Since some workers from the crowd can be highly noisy and unreliable, it is important to first exclude those unreliable workers in order to obtain high quality labels. An effective strategy for this purpose is to test each worker by a few golden samples (i.e., data with the known labels), which are usually labeled by domain experts and hence expensive to acquire. Therefore, it is desirable to select the best K workers with the minimum number of queries. This problem can be cast into our top-K arm identification problem, where each worker corresponds to an arm and the mean θ_i characterizes the ith worker's underlying reliability/quality.

More formally, assume that the arms are ordered by their means: $\theta_1 > \theta_2 > \ldots > \theta_n$ and let T be the set of selected arms with size $|T| = K$. We define the *aggregate regret* (or *regret* for short) of T as:

$$\mathcal{L}_T = \frac{1}{K}\left(\sum_{i=1}^{K}\theta_i - \sum_{i\in T}\theta_i\right). \tag{1}$$

Our goal is to design an algorithm with low sample complexity and PAC (Probably Approximately Correct) style bounds. More specifically, given any fixed positive constants ϵ, δ, the algorithm should be able to identify a set T of K arms with $\mathcal{L}_T \leq \epsilon$ (we call such a solution an ϵ-optimal solution), with probability at least $1-\delta$.

We first note that our problem strictly generalizes the previous work by (Even-Dar et al., 2006; Mannor & Tsitsiklis, 2004) for $K = 1$ to arbitrary positive integer K and hence is referred to as multiple arm identification problem. Although the problem of choosing multiple arms has been studied in some existing work, e.g., (Bubeck et al., 2013; Audibert et al., 2013; Kalyanakrishnan & Stone, 2010; Kalyanakrishnan et al., 2012), our notion of aggregate regret is inherently different from previously studied evaluation metrics such as misidentification probability (Misid-prob) (Bubeck et al., 2013) and Explore-K(Kalyanakrishnan & Stone, 2010; Kalyanakrishnan et al., 2012). In particular, Misid-prob controls the probability that the output set T is not exactly the same as the top-K arms; and Explore-Krequires to return a set T where, with high confidence, the mean of each arm in T is ϵ-close to the K-th best arm. As we will explain in the related work section, our evaluation metric is a more suitable objective for many real applications, especially for the aforementioned crowdsourcing application.

We summarize our results in this paper as follows:

1. Section 3 & 4: We develop a new PAC algorithm with sample complexity $O\left(\frac{n}{\epsilon^2}\left(1+\frac{\ln(1/\delta)}{K}\right)\right)$ for any positive constants ϵ, δ, and any $1 \leq K \leq n/2$. For $n/2 \leq K < n$, the sample complexity becomes $O\Big(\frac{n-K}{K}\cdot\frac{n}{\epsilon^2}\Big)\Big(\frac{n-K}{K}+\frac{\ln 1/\delta}{K}\Big)$. The analysis of the algorithm is presented in Section 4. It is interesting to compare this bound with the optimal $O(\frac{n}{\epsilon^2}\ln\left(\frac{1}{\delta}\right))$ bound for $K = 1$ in (Even-Dar et al., 2006; Mannor & Tsitsiklis, 2004). For $K = 1$ (i.e., selecting the best arm), our result matches theirs. Interestingly, when K is larger, our algorithm suggests that even less samples are needed. Intuitively, a larger K leads to a less stringent constraint for an ϵ-optimal solution and thus can tolerate more mistakes. Let us consider the following toy example. Assume all the arms have the same mean $1/2$, except for a random one with mean $1/2 + 2\epsilon$. If $K = 1$, to obtain an ϵ-optimal solution, we essentially need to identify the special arm and thus need a lot of samples. However, if K is large, any subset of K arms would work fine since the regret is at most $2\epsilon/K$. Our algorithm bears some similarity with previous work, such as the halving technique in (Even-Dar et al., 2006; Kalyanakrishnan & Stone, 2010; Karnin et al., 2013) and idea of accept-reject in (Bubeck et al., 2013). However, the analysis is more involved than the case for $K = 1$ and needs to be done more carefully in order to achieve the above sample complexity.

2. Section 5: To complement the upper bound, we further establish a matching lower bound for Bernoulli bandits: for $1 \leq K \leq n/2$, any (deterministic or randomized) algorithm requires at least $\Omega\left(\frac{n}{\epsilon^2}\left(1+\frac{\ln(1/\delta)}{K}\right)\right)$ samples to obtain an ϵ-optimal solution with the confidence $1-\delta$; for $n/2 \leq K < n$, the lower bound becomes $\Omega\Big(\frac{n-K}{K}\cdot\frac{n}{\epsilon^2}\Big)\Big(\frac{n-K}{K}+\frac{\ln 1/\delta}{K}\Big)$. This shows that our algorithm achieves the optimal sample complexity for Bernoulli bandits and for all values of ϵ, δ and K. To this end, we show two different lower bounds for $1 \leq K \leq n/2$: $\Omega(\frac{n}{\epsilon^2})$ and $\Omega(\frac{n}{\epsilon^2}\frac{\ln(1/\delta)}{K})$. The first bound is established via an interesting reduction from our problem to the basic problem of distinguishing two similar Bernoulli arms (with means $1/2$ and $1/2+\epsilon$ respectively). The second one can be shown via a generalization of the argument

in (Mannor & Tsitsiklis, 2004) for $K = 1$. The lower bound for $n/2 \leq K < n$ can be easily derived by a reduction to the case for $1 \leq K \leq n/2$.

3. Section 6: Finally, we conduct extensive experiments with both simulated and real data sets. The experimental results demonstrate that, using the same number of samples, our algorithm not only achieves lower regrets but also higher precisions than existing methods . Morever, using our algorithm, the maximum number of samples taken from any individual arm is much smaller than that in the SAR algorithm (Bubeck et al., 2013). This property is particularly desirable for crowdsourcing applications since it can be quite problematic, at least time-consuming, to test a single worker with too samples.

2 Related Works

Multi-armed bandit problems have been extensively studied in the machine learning community over the past decade (see for example (Auer et al., 2002a,b; Beygelzimer et al., 2011; Bubeck & Cesa-Bianchi, 2012) and the references therein). In recent years, the multiple arm identification problem has received much attention and has been investigated under different setups. For example, the work (Even-Dar et al., 2006; Mannor & Tsitsiklis, 2004; Audibert et al., 2010; Karnin et al., 2013) studied the the special case when $K = 1$. When $K > 1$, Bubeck et al.(2013) proposed a SAR (Successive Accepts and Rejects) algorithm which minimizes the *misidentification probability*, (i.e., $\Pr(T \neq \{1, \ldots, K\})$, denoted as MISID-PROB), given a fixed budget (of queries). Another line of research (Kalyanakrishnan et al., 2012; Kalyanakrishnan & Stone, 2010) proposed to select a subset T of arms, such that with high probability, for all arms $i \in T$, $\theta_i > \theta_K - \epsilon$, where θ_K is the mean of the K-th best arm. We refer this metric to as the EXPLORE-K metric.

Our notion of aggregate regret is inherently different from MISID-PROB and EXPLORE-K, and is a more suitable objective for many real applications. For example, MISID-PROB requires to identify the exact top-K arms, which is more stringent. When the gap of any consecutive pair θ_i and θ_{i+1} among the first $2K$ arms is extremely small (e.g., $o(\frac{1}{n})$), it requires a huge amount (e.g., $\omega(n^2)$) of samples to make the misidentification probability less than ϵ (Bubeck et al., 2013). While in our metric, any K arms among the first $2K$ arms consititute an ϵ-optimal solution. In crowdsourcing applications, our main goal is not to select the exact top-K workers, but a pool of good enough workers with a small number of samples. We note that the *expected simple regret*, $\frac{1}{K}(\sum_{i=1}^{K} \theta_i - \mathbf{E}[\sum_{i \in T} \theta_i])$, was also considered in a number of prior works (Audibert et al., 2010; Bubeck et al., 2013; Audibert et al., 2013). In (Audibert et al., 2010; Bubeck et al., 2013), the expected simple regret was shown to be sandwiched by Δ·MISID-PROB and MISID-PROB (for $K = 1$), where $\Delta = \theta_1 - \theta_2$. However, Δ can be arbitrarily small, hence MISID-PROB can be an arbitrarily bad bound for the simple regret. It is worthwhile noting that it is possible to obtain an expected simple regret of ϵ with at most $O(n^2/\epsilon)$ samples, using the semi-bandit regret bound in (Audibert et al., 2013)[1]. In constrast, the goal of this paper is to develop an efficient algorithm to achieve an ϵ-regret with high probability, which is a stronger requirement than obtaining an ϵ-expected simple regret.

To compare our aggregate regret with the EXPLORE-K metric, let us consider another example where $\theta_1, \ldots, \theta_{K-1}$ are much larger than θ_K and $\theta_{K+i} > \theta_K - \epsilon$ for $i = 1, \ldots, K$. It is easy to see that the set $T = \{K+1, \ldots, 2K\}$ also satisfies the requirement of EXPLORE-K [2]. However, the set T is far away from the optimal set with the aggregate regret much larger than ϵ. In crowdsourcing, the labeling performance can downgrade to a significant extent if the best set of workers (e.g., $\theta_1, \ldots, \theta_{K-1}$ in the example) is left out of the solution.

[1] The result in (Audibert et al., 2013) was stated in terms of expected accumulative regret (i.e., the expected regret over Z time slots). By setting the number of time slots Z to be $\frac{n}{K\epsilon^2}$, and choosing a random action as the final solution among Z actions (see e.g., (Bubeck et al., 2009)), one can get an expected simple regret of ϵ.

[2] For this particular instance, it is unlikely that the algorithms proposed in (Kalyanakrishnan et al., 2012; Kalyanakrishnan & Stone, 2010) would choose $\{K+1, \ldots, 2K\}$ as the solution, even though it is a valid solution under their EXPLORE-K metric. However, it is not clear, from their theoretical analysis, how good their solution is, collectively, as compared with the best K arms.

Algorithm 1 Optimal Multiple Arm Identification (OptMAI)

1: **Input:** n, K, Q.
2: **Initialization:** Active set of arms $S_0 = \{1, \ldots, n\}$; $\beta = e^{0.2} \cdot \frac{3}{4}$; set of top arms $T_0 = \emptyset$. Let $r = 0$
3: **while** $|T_r| < K$ and $|S_r| > 0$ **do**
4:     **if** $|S_r| \geq 4K$ **then**
5:         $S_{r+1} = \mathrm{QE}(S_r, K, \beta^r(1-\beta)Q)$
6:         $T_{r+1} = \emptyset$
7:     **else**
8:         $(S_{r+1}, T_{r+1}) = \mathrm{AR}(S_r, T_r, K, \beta^r(1-\beta)Q)$
9:     **end if**
10:     $r = r + 1$.
11: **end while**
12: **Output:** The set of the selected K-arms T_r.

Algorithm 2 Quartile-Elimination(QE) (S, K, Q)

Input: S, K, Q.
Sample each arm $i \in S$ for $Q_0 = \frac{Q}{|S|}$ times and let $\widehat{\theta}_i$ be the empirical mean of the i-th arm.
Find the first quartile (lower quartile) of the empirical mean $\widehat{\theta}_a$, denoted by $\widehat{q}$.
Output: The set $V = S \backslash \{i \in S : \widehat{\theta}_i < \widehat{q}\}$.

Algorithm 3 Accept-Reject(AR) (S, T, K, Q)

Input: S, T, K, Q and $s = |S|$.
Sample each arm $i \in S$ for $Q_0 = \frac{Q}{|S|}$ times and let $\widehat{\theta}_i$ be the empirical mean of the i-th arm.
Let $K' = K - |T|$. Let $\widehat{\theta}_{(K')}$ and $\widehat{\theta}_{(K'+1)}$ be the K'-th and $(K'+1)$-th largest empirical means, respectively. Define the empirical gap for each arm $i \in S$:

$$\widehat{\Delta}_i = \max(\widehat{\theta}_i - \widehat{\theta}_{(K'+1)}, \widehat{\theta}_{(K')} - \widehat{\theta}_i) \tag{2}$$

while $|T| < K$ and $|S| > 3s/4$ **do**
    Let $a \in \arg\max_{i \in S} \widehat{\Delta}_i$ and set $S = S \backslash \{a\}$.
    if $\widehat{\theta}_a \geq \widehat{\theta}_{(K'+1)}$ **then**
        Set $T = T \cup \{a\}$.
    end if
end while
Output: The set S and T.

3 Algorithm

In this section, we describe our algorithm for the multiple arm identification problem. Our algorithm OptMAI (Algorithm 1) takes three positive integers n, K, Q as the input, where n is the total number of arms, K is the number of arms we want to choose and Q is an upper bound on the total number of samples [3]. In Section 4, we show that $Q = O\left(\frac{n}{\epsilon^2}\left(1 + \frac{\ln(1/\delta)}{K}\right)\right)$ suffices to obtain an ϵ-optimal solution with probability at least $1 - \delta$. OptMAI consists of two stages, the *Quartile-Elimination (QE) stage* (line 4-6) and the *Accept-Reject (AR) stage* (line 8).

The QE stage proceeds in rounds. Each QE round calls the QE subroutine in Algorithm 2, which requires three parameters S, K and Q. Here, S is the set of arms which we still want to pull and Q is the total number

[3] If Algorithm 1 stops at round $r = R$, the total number of samples is $(1 - \beta^R)Q < Q$.

of samples required in this round. We sample each arm in S for $Q/|S|$ times and then discard *a quarter* of arms with the minimum empirical mean [4]. We note that in each call of the QE subroutine, we pass different Q values (exponentially decreasing). This is critical for keeping the total number of samples linear and achieving the optimal sample complexity. See Algorithm 1 for the setting of the parameters. The QE stage repeatedly calls the QE subroutine until the number of remaining arms is at most $4K$.

Now, we enter the AR stage, which also runs in rounds. Each AR round (Algorithm 3) requires four parameters, S, T, K, Q, where S, K, Q have the same meanings as in QE and T is the set of arms that we have decided to include in our final solution and thus will not be sampled any more. In each AR subroutine (Algorithm 3), we again sample each arm for $Q/|S|$ times. We define the *empirical gap* for the i-th arm to be the absolute difference between the empirical mean of the i-th arm and the K'-th (or $K'+1$-th) largest empirical mean, where $K' = K - |T|$ (see Eq.(2)). We remove a quarter of arms with the largest empirical gaps. There are two types of those removed arms: those with the largest empirical means, which are included in our final solution set T, and those with the smallest empirical means, which are discarded from further consideration.

Remark 3.1. *We would like to mention that the naive uniform sampling algorithm, which takes the same number of samples from each arm and chooses the K arms with the largest empirical means, does not achieve the optimal sample complexity. In general, it requires at least $\Omega(n \log n)$ samples, which is* $\log n$ *factor worse than our optimal bound. See Appendix E for a detailed discussion.*

Remark 3.2. *To achieve the desired asymptotic bound on regret and sample complexity, the AR stage can be substituted by a simpler process which takes a uniform number of samples from each arm and chooses the K arms with the largest empirical means. The details can be found in Appendix D. We choose to present the AR subroutine in this section because 1) it also meets the theoretical bound in Section 4; 2) the AR stage shows a much better empirical performance.*

4 Bounding the Regret and the Sample Complexity

We analyze the regret achieved by our algorithm. Firstly, let us introduce some necessary notations. For any positive integer C, we use $[C]$ to denote the set $\{1, 2, \ldots, C\}$. For any subset S of arms, let $\mathsf{ind}_i(S)$ be the arm in S with the i-th largest mean. We use $\mathsf{val}_C(S)$ to denote the average mean of the C best arms in S, i.e., $\mathsf{val}_C(S) \triangleq \frac{1}{C}\sum_{i=1}^{C} \theta_{\mathsf{ind}_i(S)}$. Let $\mathsf{tot}_C(S) = C \cdot \mathsf{val}_C(S)$ be the total sum of the means of the C best arms in S. We first consider one QE round. Suppose S is the set of input arms and V is the output set. We first show that the average mean of the K best arms in V is at most ϵ worse that that in S, for some appropriate ϵ (depending on Q and $|S|$).

Lemma 4.1. *Assume that $K \le |S|/4$ and let V be the output of* $\mathrm{QE}(S, K, Q)$ *(Algorithm 2). For every $\delta > 0$, with probability $1-\delta$, we have that $\mathsf{val}_K(V) \ge \mathsf{val}_K(S) - \epsilon$, where $\epsilon = \sqrt{\frac{|S|}{Q}\left(10 + \frac{4\ln(2/\delta)}{K}\right)}$.*

The basic idea of the proof goes as follows. Let $p = \theta_{\mathsf{ind}_{|S|/2}(S)}$ be the median of the θ in S and $\tau = \min_{i \in V}(\widehat{\theta}_i)$ be the minimum *empirical mean* for the selected arms in V. For each arm i among the top K arms in S, we define the random variable $X_i = \mathbf{1}\{\widehat{\theta}_{\mathsf{ind}_i(S)} < p + \frac{\epsilon}{2}\}$, where $\mathbf{1}\{\cdot\}$ is the indicator function (i.e., $\mathbf{1}\{\text{true}\} = 1$ and $\mathbf{1}\{\text{false}\} = 0$). Let $X = \frac{1}{K}\sum_{i=1}^{K}(\theta_{\mathsf{ind}_i(S)} - p)X_i$. We further define two events $\mathcal{E}_1 = \{X \le \epsilon\}$ and $\mathcal{E}_2 = \{\tau < p + \frac{\epsilon}{2}\}$. Intuitively, $\mathcal{E}_2$ says that the threshold τ, as the third quartile of the empirical means, is not much larger than the expected median value (i.e., p). When $\mathcal{E}_2$ happens, $\mathcal{E}_1$ would give an upper bound on the regret $\mathsf{val}_K(S) - \mathsf{val}_K(V)$. We formalize this idea by first showing that the events $\mathcal{E}_1$ and $\mathcal{E}_2$ together imply that $\mathsf{val}_K(V) \ge \mathsf{val}_K(S) - \epsilon$. Then we prove that with the ϵ value defined in the lemma statement, Each of $\mathcal{E}_1$ and $\mathcal{E}_2$ holds with probability at least $1 - \frac{\delta}{2}$. By a simple union bound, our proof is completed. The details are presented in the appendix.

[4] The empirical mean of an arm is the average reward of the arm, over all samples in this round.

Secondly, we provide the regret bound for the AR algorithm in the following lemma with the proof presented in the appendix.

Lemma 4.2. *Let (S', T') be the output of the algorithm $\mathrm{AR}(S, T, K, Q)$. For every $\delta > 0$, with probability $1 - \delta$, we have that*

$$\mathsf{tot}_{K-|T'|}(S') + \mathsf{tot}_{|T'|}(T') \geq \mathsf{tot}_{K-|T|}(S) + \mathsf{tot}_{|T|}(T) - \epsilon K,$$

where $\epsilon = \sqrt{\frac{|S|}{Q}\left(4 + \frac{\log(2/\delta)}{K}\right)}$.

In each round of the AR-stage with $S = S_r$ and top arms $T = T_r$, the value $\frac{\mathsf{tot}_{K-|T|}(S)+\mathsf{tot}_{|T|}(T)}{K}$ is the best possible average mean of K arms. Lemma 4.2 provides an upper bound for the gap between this value on the output (T', S') by AR and the best possible one. Applying this bound over all rounds would further imply that this value of the output of Algorithm 1 is not far away from that of the real top-K arms. With Lemma 4.1 and Lemma 4.2 in place, we prove the performance of Algorithm 1 in the next theorem.

Theorem 4.3. *For every $\delta > 0$, with probability at least $1 - \delta$, the output of OptMAI algorithm T is an ϵ-optimal solution (i.e., $\mathsf{val}_K(T) \geq \mathsf{val}_K([n]) - \epsilon$) with $\epsilon = O\left(\sqrt{\frac{n}{Q}\left(1 + \frac{\ln 1/\delta}{K}\right)}\right)$.*

Theorem 4.3 also provides us the sample complexity of Algorithm 1 for any pre-fixed positive values ϵ and δ, as stated in the next corollary.

Corollary 4.4. *For any positive constants $\epsilon, \delta > 0$, it suffices to run Algorithm 1 with*

$$Q = O\left(\frac{n}{\epsilon^2}\left(1 + \frac{\ln(1/\delta)}{K}\right)\right). \tag{3}$$

in order to obtain an ϵ-optimal solution with probability $1 - \delta$. In other words, the sample complexity of the algorithm is bounded by $O\left(\frac{n}{\epsilon^2}\left(1 + \frac{\ln(1/\delta)}{K}\right)\right)$ from above.

For $n/2 \leq K < n$, we can easily obtain a better sample complexity as follows.

Theorem 4.5. *For any $\delta > 0$ and $n/2 \leq K < n$, with probability at least $1 - \delta$, there is an algorithm that can find an ϵ-optimal solution T (i.e., $\mathsf{val}_K(T) \geq \mathsf{val}_K([n]) - \epsilon$) and the number of samples used is at most at most*

$$O\Big(\frac{n-K}{K} \cdot \frac{n}{\epsilon^2}\Big)\Big(\frac{n-K}{K} + \frac{\ln 1/\delta}{K}\Big).$$

Proof. Instead of directly finding the best K arms, we attempt to find the worst $n - K$ arms. First, we can see that for any $\epsilon', \delta > 0$, we can find a set T' of $n - K$ arms such that

$$\sum_{i \in T'} \theta_i - \sum_{i=K+1}^{n} \theta_i \leq (n-K)\epsilon', \tag{4}$$

(we call such a set T' an ϵ'-*worst solution*) with probability $1-\delta$, using at most $O\left(\frac{n}{\epsilon'^2}\left(1 + \frac{\ln(1/\delta)}{n-K}\right)\right)$ samples. This can be done by constructing a new multiple arm identification instance and run OPTMAI on the new instance. In the new instance, there is an arm with mean $1-\theta_i$ if the original instance consists of an arm with mean θ_i. Sampling from this arm can be simulated by sampling from the corresponding original arm (if we get a sample of value x from the original arm, we use $1 - x$ as the sample for the new arm). It is easy to see that for any ϵ-optimal solution T' for the new instance, we have that $\sum_{i=K+1}^{n}(1-\theta_i) - \sum_{i \in T'}(1-\theta_i) \leq (n-K)\epsilon'$ (by definition), which is equivalent to an ϵ'-worst solution for the original instance.

By setting $\epsilon' = \frac{K}{n-K} \cdot \epsilon$ and $T = [n] \setminus T'$, we can see that(4) implies that $\sum_{i=1}^{K} \theta_i - \sum_{i \in T} \theta_i \leq K\epsilon$. The theorem follows. ∎

5 A Matching Lower Bound

In this section, we provide lower bounds for Bernoulli bandits where the reward of the i-th arm follows a Bernoulli distribution with mean θ_i. We prove that for any underlying $\{\theta_i\}_{i=1}^n$ and any randomized algorithm $\mathcal{A}$, the expected number of samples Q required to identify an ϵ-optimal solution with probability $1-\delta$ is at least $\max\left\{\Omega\left(\frac{n\ln(1/\delta)}{\epsilon^2 K}\right), \Omega\left(\frac{n}{\epsilon^2}\right)\right\} = \Omega\left(\frac{n}{\epsilon^2}\left(\frac{\ln(1/\delta)}{K}+1\right)\right)$. According to Corollary 4.4, for Bernoulli bandits, our algorithm achieves this lower bound of the sample complexity. In particular, we separate the proof into two parts: in the first part, we show that $Q \geq \Omega\left(\frac{n}{\epsilon^2}\right)$; and $Q \geq \Omega\left(\frac{n\ln(1/\delta)}{\epsilon^2 K}\right)$ in the second.

5.1 First Lower Bound: $Q \geq \Omega\left(\frac{n}{\epsilon^2}\right)$

Theorem 5.1. *Fix a real number ϵ, integers K, n, where $0 < \epsilon < 0.01$ and $10 \leq K \leq n/2$. Let $\mathcal{A}$ be a possibly randomized algorithm, so that for any set of n Bernoulli arms with means $\theta_1, \theta_2, \ldots, \theta_n$,*

- *$\mathcal{A}$ takes at most Q samples in expectation;*
- *with probability at least 0.8, $\mathcal{A}$ outputs a set T of size K with $\mathsf{val}_K(T) \geq \mathsf{val}_K([n]) - \epsilon$.*

Then, we have that $Q \geq \Omega(\frac{n}{\epsilon^2})$.

The high level idea of the proof of Theorem 5.1 is as follows. Suppose there is an algorithm $\mathcal{A}$ which can find an ϵ-optimal solution with probability at least 0.8 and uses at most Q samples in expectation. We show that we can use $\mathcal{A}$ as a subroutine to construct an algorithm $\mathcal{B}$, which can distinguish whether a single Bernoulli arm has mean $1/2$ or $1/2+4\epsilon$ with at most $\frac{200Q}{n}$ samples (Lemma 5.2). We utilize the well known fact that, for any algorithm (including $\mathcal{B}$), distinguishing such a Bernoulli arm requires at least $\Omega(\frac{1}{\epsilon^2})$ samples. Hence, we must have that $\frac{200Q}{n} \geq \Omega(\frac{1}{\epsilon^2})$, which gives the desired lower bound for Q.

Formally, we show in the following lemma how to construct $\mathcal{B}$, using $\mathcal{A}$ as a subroutine.

Lemma 5.2. *Let $\mathcal{A}$ be an algorithm in Theorem 5.1. There is an algorithm $\mathcal{B}$, which correctly outputs whether a Bernoulli arm X has the mean $\frac{1}{2}+4\epsilon$ or the mean $\frac{1}{2}$ with probability at least 0.51, and $\mathcal{B}$ makes at most $\frac{200Q}{n}$ samples.*

Assuming the existence of an algorithm $\mathcal{A}$ stated in Theorem 5.1, we construct the algorithm $\mathcal{B}$ as follows. Keep in mind that the goal of $\mathcal{B}$ is to distinguish whether a given Bernoulli arm (denoted as X) has mean $1/2$ or $1/2+4\epsilon$.

Algorithm 4 Algorithm $\mathcal{B}$ (which calls $\mathcal{A}$ as a subroutine)

1. Choose a random subset $S \subseteq [n]$ such that $|S| = K$ and then choose a random element $j \in S$.
2. Create n artificial arms as follows: For each $i \in [n], i \neq j$, let $\theta_i = \frac{1}{2}+4\epsilon$ if $i \in S$, let $\theta_i = \frac{1}{2}$ otherwise.
3. Simulate $\mathcal{A}$ as follows: whenever $\mathcal{A}$ samples the i-th arm:
   (1) If $i = j$, we sample the Bernoulli arm X (recall X is the arm which $\mathcal{B}$ attempts to separate);
   (2) Otherwise, we sample the arm with mean θ_i.
4. If the arm X is sampled by less than $\frac{200Q}{n}$ times and $\mathcal{A}$ returns a set T such that $j \notin T$, we decide that X has the mean of $\frac{1}{2}$; otherwise we decide that X has the mean of $\frac{1}{2}+4\epsilon$.

We note that the number of samples of $\mathcal{B}$ increases by one whenever X is sampled. Since $\mathcal{B}$ stops and outputs the mean $\frac{1}{2}+4\epsilon$ if the number of samples on X reaches $\frac{200Q}{n}$, $\mathcal{B}$ takes at most $\frac{200Q}{n}$ samples from X. The intuition why the above algorithm can separate X is as follows. If X has mean $1/2+4\epsilon$, X is no different from any other arm in S. Similarly, if X has mean $1/2$, X is the same as any other arm in $[n] \setminus S$. If $\mathcal{A}$ satisfies the requirement in Theorem 5.1, $\mathcal{A}$ can identify a significant proportion of arms with mean $1/2+4\epsilon$. So if X has mean $1/2+4\epsilon$, there is a good chance (noticeably larger than 0.5) that X will be chosen by $\mathcal{A}$. In the appendix, we formally prove the correctness of $\mathcal{B}$, i.e., it can correctly output the mean of X with probability at least 0.51; and thus conclude the proof of Lemma 5.2.

The second step of the proof of Theorem 5.1 is a well-known lower bound on the expected sample complexity for separating a single Bernoulli arm (Chernoff, 1972; Anthony & Bartlett, 1999).

Lemma 5.3. *Fix ϵ such that $0 < \epsilon < 0.01$ and let X be a Bernoulli random variable with mean being either $\frac{1}{2} + 4\epsilon$ or $\frac{1}{2}$. If an algorithm $\mathcal{B}$ can output the correct mean of X with probability at least 0.51, then expected number of samples performed by $\mathcal{B}$ is at least $\Omega(\frac{1}{\epsilon^2})$.*

By combining Lemma 5.2 and Lemma 5.3, we have $\frac{200Q}{n} \geq \Omega(\frac{1}{\epsilon^2})$; and therefore prove the claim that $Q \geq \Omega(\frac{n}{\epsilon^2})$ in Theorem 5.1.

5.2 Second Lower Bound: $Q \geq \Omega\left(\frac{n \ln(1/\delta)}{\epsilon^2 K}\right)$

Lemma 5.4. *Fix real numbers δ, ϵ such that $0 < \delta, \epsilon \leq 0.01$, and integers K, n such that $K \leq n/2$. Let $\mathcal{A}$ be a deterministic algorithm (i.e., the only randomness comes from the arms), so that for any set of n Bernoulli arms with means $\theta_1, \theta_2, \ldots, \theta_n$,*

- *$\mathcal{A}$ makes at most Q samples in expectation;*
- *with probability at least $1 - \delta$, $\mathcal{A}$ outputs a set T of size K with $\mathsf{val}_K(T) \geq \mathsf{val}_K([n]) - \epsilon$.*

Then, we have that $Q \geq \frac{n \ln(1/\delta)}{20000 \epsilon^2 K}$.

Now, we provide a sketch of our proof, which generalizes the previous proof for the lower bound when $K = 1$ (Mannor & Tsitsiklis, 2004). Let $t = \lfloor \frac{n}{K} \rfloor \geq 2$ and we divide the first tK arms into t groups, where the j-th group consists of the arms with the indices in $[(j-1)K + 1, jK]$. We first construct t hypotheses $H_1, H_2, \ldots, H_t$ as follows. In H_1, we let $\theta_i = 1/2 + 4\epsilon$ for arms in the first group and let $\theta_i = 1/2$ for the remaining arms. In H_j, where $2 \leq j \leq t$, we let $\theta_i = 1/2 + 4\epsilon$ when i is in the first group, $\theta_i = 1/2 + 8\epsilon$ when i is in the j-th group, and $\theta_i = 1/2$ otherwise.

For each $j \in [t]$, let $\Pr_{H_j}[\cdot]$ ($\mathbf{E}_{H_j}[\cdot]$ resp.) denote the probability of the event in $[\cdot]$ (the expected value of the random variable in $[\cdot]$ resp.) under the hypothesis H_j. Let $\tilde{q}_j$ be the total number of samples taken from the arms in the j-th group. By an averaging argument, there must exist a group $j_0 \geq 2$ such that $\mathbf{E}_{H_1}[\tilde{q}_{j_0}] \leq \frac{Q}{t-1} \leq \frac{2Q}{t}$.

After fixing j_0, we focus on the hypothesis H_1 and H_{j_0}. Let $\mathsf{val}_K^{H_j}(T)$ ($\mathsf{val}_K^{H_j}([n])$ resp.) be the $\mathsf{val}_K(T)$ value ($\mathsf{val}_K([n])$ resp.) computed using θ values defined in hypothesis H_j. Note that $\mathsf{val}_K^{H_j}(T)$ (for any j) is always well defined no matter which hypothesis is the true underlying probability measure.

At a high level, our proof works as follows. We assume for contradiction that $Q < \frac{n \ln(1/\delta)}{20000\epsilon^2 K}$. Using the assumption $\Pr_{H_1}[\mathsf{val}_K^{H_1}(T) \geq \mathsf{val}_K^{H_1}([n]) - \epsilon] \geq 1 - \delta$, we can first prove that:

$$\Pr_{H_{j_0}}[\mathsf{val}_K^{H_1}(T) \geq \mathsf{val}_K^{H_1}([n]) - \epsilon] \geq \frac{\sqrt{\delta}}{4}. \tag{5}$$

We further observe that when $\mathsf{val}_K^{H_1}(T) \geq \mathsf{val}_K^{H_1}([n]) - \epsilon$, T must consist of more than $\frac{3}{4}K$ arms from the first group; while when $\mathsf{val}_K^{H_{j_0}}(T) \geq \mathsf{val}_K^{H_{j_0}}([n]) - \epsilon$, T must consist of more than $\frac{3}{4}K$ arms from the j_0-th group. Therefore the two events are mutually exclusive and we have: $\Pr_{H_{j_0}}\left[\mathsf{val}_K^{H_{j_0}}(T) \geq \mathsf{val}_K^{H_{j_0}}([n]) - \epsilon\right] \leq 1 - \Pr_{H_{j_0}}\left[\mathsf{val}_K^{H_1}(T) \geq \mathsf{val}_K^{H_1}([n]) - \epsilon\right] \leq 1 - \frac{\sqrt{\delta}}{4} \leq 1 - 2\delta$, where the last inequality is because of $\delta < 0.01$. This contradicts the performance guarantees of Algorithm $\mathcal{A}$ and thus we conclude our proof.

The most technical part is to prove (5). To this end, we construct the likelihood ratio between events under the hypothesis H_1 and H_{j_0}. The intuition is that H_1 and H_{j_0} are similar, thus for any sampling outcomes y, the probability that H_1 generates y is close to H_{j_0}. Since $\mathcal{A}$ is deterministic, the sampling outcomes determine the next action and the final decision. Using this argument, we can show that if the event $\{\mathsf{val}_K^{H_1}(T) \geq \mathsf{val}_K^{H_1}([n]) - \epsilon\}$ happens under H_1 with probability at least $1 - \delta$, it would happen under H_{j_0} with a significant probability (i.e., at least $\frac{\sqrt{\delta}}{4}$). The details are provided in the appendix.

The proof of Lemma 5.4 can be easily generalized to the case where $\mathcal{A}$ is randomized, which allows us to prove the following stronger lower bound statement. The proof of Theorem 5.5 is relegated to the appendix.

Theorem 5.5. *Fix real numbers δ, ϵ such that $0 < \delta, \epsilon \le 0.01$, and integers K, n, where $K \le n/2$. Let $\mathcal{A}$ be a (possibly randomized) algorithm so that for any set of n arms with the mean $\theta_1, \theta_2, \ldots, \theta_n$,*

- *$\mathcal{A}$ makes at most Q samples in expectation;*
- *With probability at least $1-\delta$, $\mathcal{A}$ outputs a set T of size K with $\mathsf{val}_K(T) \ge \mathsf{val}_K([n]) - \epsilon$.*

We have that $Q = \Omega\left(\frac{n \ln(1/\delta)}{\epsilon^2 K}\right)$.

6 Experiments

In this experiment, we assume that arms follow independent Bernoulli distributions with different means. To make a fair comparison, we fix the total budget Q and compare our algorithm (OptMAI) with the uniform sampling strategy and two other state-of-the-art algorithms: SAR (Bubeck et al., 2013) and LUCB (Kalyanakrishnan et al., 2012), in terms of the aggregate regret in (1).

The implementation of our algorithm is slightly different from its description in Section 3. While we choose to present a variant of our implementation only because of its simplicity of exposition, we describe the full details of the differences as follows. It is easy to check that our implementation still meets the theoretical bound proved in Section 4.

First, observe that in OptMAI, Q is an upper bound of the number of samples, while $(1-\beta^R)Q < Q$ is the actual number of samples used, where R is the total number of rounds run by the algorithm. To fully utilize the budget, we run OptMAI with a parameter slightly greater than Q to ensure that the actual number of samples roughly equals to (but no greater than) Q .

Second, in each round of QE or AR, when computing the empirical mean $\hat{\theta}_i$, our implementation uses all the samples obtained for the i-th arm (i.e. including the samples from previous rounds). This will lead to be better empirical performance especially when the budget is very limited.

Third, in each round of OptMAI, the ratio of the number of samples between two consecutive rounds is set to be $\beta = e^{0.2} \cdot 0.75 \approx 0.91$. In the real implementation, one could treat this quantity as a tuning parameter to make the algorithm more flexible (as long as $\beta \in (0.75, 1)$). In this experiment, we report the results for both $\beta = 0.8$ and $\beta = 0.9$. Based on our experimental results, one could simply set $\beta = 0.8$, which will lead to reasonably good performance under different scenarios. We propose the following strategy to tune β as a future work. In the first stage, we sample each arm for a few times and then use the empirical estimate of θ_i to generate as much simulated data as we want. Then, we choose the best β based on the simulated data. Finally, we apply the carefully tuned β to the real data using the remaining budget.

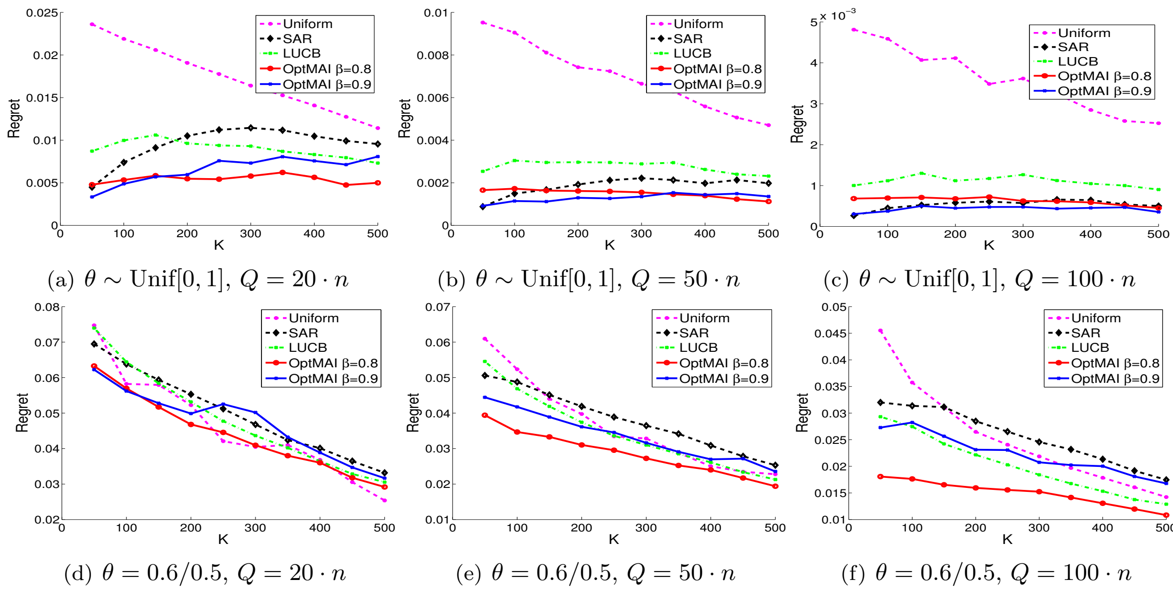
6.1 Simulated Experiments

In our simulated experiment, the number of total arms is set to $n = 1000$. We vary the total budget $Q = 20n, 50n, 100n$ and $K = 10, 20, \ldots, 500$. We use different ways to generate $\{\theta_i\}_{i=1}^n$ and report the comparison results among different algorithms:

1. $\theta_i \sim \mathrm{Unif}[0,1]$: each θ_i is uniformly distributed on $[0,1]$ (see Figure 1(a) to Figure 1(b)).
2. $\theta_i = 0.5/0.6$: $\theta_i = 0.6$ for $i = 1, \ldots, K$ and $\theta_i = 0.5$ for $i = K+1, \ldots, n$. We note that such a two level setting of θ_i is more challenging for the selection of top-K arms (see Figure 1(d) to Figure 1(f)).

It can be seen from Figure 1 that the uniform sampling performs the worst and our method outperforms SAR and LUCB in most of the scenarios. We also observe that when K is large, the setting of $\beta = 0.8$ (red line) outperforms that of $\beta = 0.9$; while for small K, $\beta = 0.9$ (blue line) is a better choice.

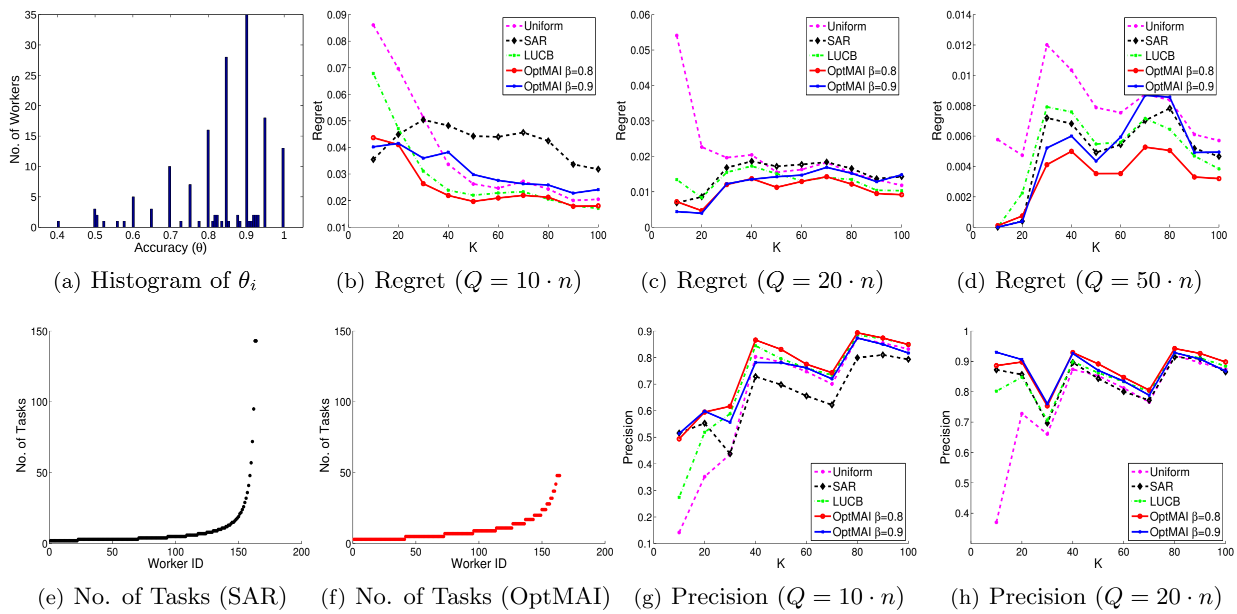
We also generate θ_i from the truncated normal distribution and the Beta distribution and have similar observations. The comparison results are presented in the appendix due to space constraints.



(a) $\theta \sim \text{Unif}[0,1]$, $Q = 20 \cdot n$ (b) $\theta \sim \text{Unif}[0,1]$, $Q = 50 \cdot n$ (c) $\theta \sim \text{Unif}[0,1]$, $Q = 100 \cdot n$

(d) $\theta = 0.6/0.5$, $Q = 20 \cdot n$ (e) $\theta = 0.6/0.5$, $Q = 50 \cdot n$ (f) $\theta = 0.6/0.5$, $Q = 100 \cdot n$

Figure 1: Performance comparison on simulated data.



(a) Histogram of θ_i (b) Regret ($Q = 10 \cdot n$) (c) Regret ($Q = 20 \cdot n$) (d) Regret ($Q = 50 \cdot n$)

(e) No. of Tasks (SAR) (f) No. of Tasks (OptMAI) (g) Precision ($Q = 10 \cdot n$) (h) Precision ($Q = 20 \cdot n$)

Figure 2: Performance comparison on the RTE data.

6.2 Real RTE Data

We generate θ from a real recognizing textual entailment (RTE) dataset (Section 4.3 in (Snow et al., 2008)). There are 800 task and each task is a sentence pair. Each sentence pair is presented to 10 different workers to acquire binary choices of whether the second hypothesis sentence can be inferred from the first one. There are in total 164 different workers. Since there are true labels of tasks, we set each θ_i for the i-th worker to be his/her labeling accuracy. The histogram of θ_i is presented in Figure 2(a). We vary the total budget

$Q = 10n, 20n, 50n$ and K from 10 to 100 and report the comparison of the regret for different approaches in Figure 2(b) to Figure 2(d). As we can see, our method with $\beta = 0.8$ (red line) outperforms other competitors for most of K's and Q's. SAR performs the best for $K = 10$, $Q = 10n$; while our method with $\beta = 0.9$ performs the best for $K = 10$ and $Q = 20n$.

In addition, we would like to highlight an interesting property of our method from the empirical study. As shown in Figure 2(e) and Figure 2(f) with $Q = 10n$ and $K = 20$, the empirical distribution of the number of samples (i.e., tasks) assigned to a worker using SAR is much more skewed than that using our method. This property makes our method particularly suitable for crowdsourcing applications since it will be extremely time-consuming if a single worker is assigned with too many tasks (e.g., golden samples). For example, for SAR, a worker could receive up to 143 tasks (Figure 2(e)) while for our method, a worker receives at most 48 tasks (Figure 2(f)). In crowdsourcing, a single worker will take a long time and soon lose patience when performing nearly 150 testing tasks.

In Figure 2(g) and Figure 2(h), we compare different algorithms in terms of the precision, which is defined as the number of arms in T which belong to the set of the top K arms over K, i.e., $\frac{|T \cap [K]|}{K}$. As we can see, our method with $\beta = 0.8$ achieves the highest precision followed by LUCB.

7 Conclusions and Future Work

We study the problem of identifying the (approximate) top K-arms in a stochastic multi-armed bandit game. We propose to use the aggregate regret as the evaluation metric, which fits to the PAC framework. We argue that in many real applications, our metric is more suitable. Our algorithm can identify an ϵ-optimal solution with probability at least $1 - \delta$, with the sample complexity $O\left(\frac{n}{\epsilon^2}\left(1 + \frac{\ln(1/\delta)}{K}\right)\right)$ for any $1 \leq K \leq n/2$, $O\left(\frac{n-K}{K} \cdot \frac{n}{\epsilon^2}\right)\left(\frac{n-K}{K} + \frac{\ln 1/\delta}{K}\right)$ for any $n/2 < K < n$. These upper bounds match the lower bounds provided in this paper.

There are several directions that we would like to explore in the future. Firstly, our algorithm provides the guarantees for the worst case scenarios and does not depend on the actual reward distributions (i.e., the value of θ_i). In many real data sets, the means of the arms are well separated, and might be easier than the worst case instances (such as the instances constructed in our lower bound proof). Inspired by the work (Audibert et al., 2010; Bubeck et al., 2013; Karnin et al., 2013), our next step is to design new adaptive algorithms and provide more refined distribution dependent upper and lower bounds. Secondly, our lower bound instances are based on Bernoulli bandits. It would be interesting to establish the lower bound for other distributions supported on $[0, 1]$ or even more general distributions, such as sub-Gaussian. Finally, it would be of great interest to test our algorithm on real crowdsourcing platforms and apply it to many other real-world applications.

References

Anthony, M. and Bartlett, P. L. *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, 1999. 8

Audibert, J.-Y., Bubeck, S., and Lugosi, G. Regret in online combinatorial optimization. *Mathematics of Operations Research*, 2013. 2, 3

Audibert, J., Bubeck, S., and Munos, R. Best arm identification in multi-armed bandits. In *Proceedings of the Annual Conference on Learning Theory (COLT)*, 2010. 3, 11

Auer, P., Cesa-Bianchi, N., Freund, Y., and Schapire, R. The nonstochastic multiarmed bandit problem. *SIAM J. on Comput.*, 32(1):48–77, 2002a. 3

Auer, P., Cesa-Bianchi, N., and Fischer, P. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47:235–256, 2002b. 3

Beygelzimer, A., Langford, J., Li, L., Reyzin, L., and Schapire, R. E. Contextual bandit algorithms with supervised learning guarantees. In *Proceedings of International Conference on Artificial Intelligence and Statistics*, 2011. 3

Bubeck, S. and Cesa-Bianchi, N. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–122, 2012. 3

Bubeck, S., Wang, T., and Viswanathan, N. Multiple identifications in multi-armed bandits. In *Proceedings of the International Conference on Machine Learning*, 2013. 1, 2, 3, 9, 11

Bubeck, S., Munos, R., and Stoltz, G. Pure exploration in multi-armed bandits problems. In *Algorithmic Learning Theory*, pp. 23–37. Springer, 2009. 3

Chen, X., Lin, Q., and Zhou, D. Optimistic knowledge gradient for optimal budget allocation in crowdsourcing. In *Proceedings of International Conference on Machine Learning*, 2013. 1

Chernoff, H. *Sequential Analysis and Optimal Design.* Society for Industrial and Applied Mathematics (SIAM), 1972. 8

Even-Dar, E., Mannor, S., and Mansour, Y. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of machine learning research*, 7:1079–1105, 2006. 2, 3, 19

Ho, C.-J., Jabbari, S., , and Vaughan, J. W. Adaptive task assignment for crowdsourced classification. In *ICML*, 2013. 1

Kalyanakrishnan, S. and Stone, P. Efficient selection of multiple bandit arms: Theory and practice. In *Proceedings of International Conference of Machine Learning*, 2010. 1, 2, 3

Kalyanakrishnan, S., Tewari, A., Auer, P., and Stone, P. PAC subset selection in stochastic multi-armed bandits. In *Proceedings of International Conference on Machine Learning*, 2012. 2, 3, 9

Karger, D. R., Oh, S., and Shah, D. Budget-optimal task allocation for reliable crowdsourcing systems. arXiv:1110.3564v3, 11 2012. 1

Karnin, Z., Koren, T., and Somekh, O. Almost optimal exploration in multi-armed bandits. In *Proceedings of International Conference on Machine Learning*, 2013. 2, 3, 11

Koenig, L. W. and Law, A. M. A procedure for selecting a subset of size m containing the l best of k independent normal populations, with applications to simulation. *Communications in statistics. Simulation and computation*, 14:719–734, 1985. 1

Liu, Q., Steyvers, M., and Ihler, A. Scoring workers in crowdsourcing: How many control questions are enough? In *NIPS*, 2013. 1

Mannor, S. and Tsitsiklis, J. N. The sample complexity of exploration in the multi-armed bandit problem. *Journal of Machine Learning Research*, 5:623–648, 2004. 2, 3, 8

Raykar, V. C., Yu, S., Zhao, L. H., Valadez, G. H., Florin, C., Bogoni, L., and Moy, L. Learning from crowds. *Journal of Machine Learning Research*, 11:1297–1322, 2010. 1

Ronald, L. G., Donald, E. K., and Oren, P. *Concrete Mathematics; A Foundation for Computer Science.* Addison-Wesley, 1989. 20

Schmidt, C., Branke, J., and Chick, S. E. Integrating techniques from statistical ranking into evolutionary algorithms. 2006. 1

Snow, R., Connor, B. O., Jurafsky, D., and Ng., A. Y. Cheap and fast - but is it good? evaluating non-expert annotations for natural language tasks. In *EMNLP*, 2008. 10

Thompson, W. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25:285–294, 1933. 1

Welinder, P., Branson, S., Belongie, S., and Perona, P. The multidimensional wisdom of crowds. In *NIPS*, 2010. 1

Zhou, D., Basu, S., Mao, Y., and Platt, J. Learning from the wisdom of crowds by minimax conditional entropy. In *NIPS*, 2012. 1

A Proof of the Correctness of the QE Algorithm (Lemma 4.1)

Lemma 4.1 *Assume that $K \le |S|/4$ and let V be the output of* $\mathrm{QE}(S, K, Q)$ *(Algorithm 2). For every $\delta > 0$, with probability $1-\delta$, we have that $\mathsf{val}_K(V) \ge \mathsf{val}_K(S) - \epsilon$, where $\epsilon = \sqrt{\frac{|S|}{Q}\left(10 + \frac{4\ln(2/\delta)}{K}\right)}$.*

Let $p = \theta_{\mathsf{ind}_{|S|/2}(S)}$ be the median of the means of the arms in S. Let $\tau = \min_{i \in V}(\widehat{\theta}_i)$ be the minimum empirical mean for the selected arms in V. For each arm i among the top K arms in S, we define the random variable $X_i = \mathbf{1}\{\widehat{\theta}_{\mathsf{ind}_i(S)} < p + \frac{\epsilon}{2}\}$ and $X = \frac{1}{K}\sum_{i=1}^{K}(\theta_{\mathsf{ind}_i(S)} - p)X_i$, where $\mathbf{1}\{\cdot\}$ is the indicator function. We further define two events $\mathcal{E}_1 = \{X \le \epsilon\}$ and $\mathcal{E}_2 = \{\tau < p + \frac{\epsilon}{2}\}$. Our first claim is that $\mathcal{E}_1$ and $\mathcal{E}_2$ together imply our conclusion $\mathsf{val}_K(V) \ge \mathsf{val}_K(S) - \epsilon$.

Lemma A.1. *$\mathcal{E}_1$ and $\mathcal{E}_2$ imply that $\mathsf{val}_K(V) \ge \mathsf{val}_K(S) - \epsilon$.*

Proof. Suppose both $\mathcal{E}_1$ and $\mathcal{E}_2$ hold. We first claim that

$$\frac{1}{K}\sum_{i=1}^{K}\theta_{\mathsf{ind}_i(V)} \ge \frac{1}{K}\sum_{i=1}^{K}\left((1 - X_i)\theta_{\mathsf{ind}_i(S)} + X_i p\right). \tag{6}$$

To see this claim, consider arm $\mathsf{ind}_i(S)$ for some $i \in [K]$. If $X_i = 0$ (i.e., $\widehat{\theta}_{\mathsf{ind}_i(S)} \ge p + \frac{\epsilon}{2}$), together with $\mathcal{E}_2$, we have that

$$\widehat{\theta}_{\mathsf{ind}_i(S)} \ge p + \frac{\epsilon}{2} > \tau = \min_{i \in V}(\widehat{\theta}_i).$$

Hence, the arm should be included in the output set V. Moreover, since it is one of the best K arms in S, it is also one of the best K arms in V. Hence, for each term on the right hand side of (6) with $X_i = 0$, there is exactly one term with the same value on the left hand side.

Since there are $|S|/2 \ge K + |S|/4$ arms with means greater or equal to p, after removing $|S|/4$ of them, there are still at least K such arms. Therefore we know that the best K arms of V all have means greater than or equal to p. In other words, each term on the left hand side of (6) is greater than or equal to p. This proves (6). Now, we can see that

$$\begin{aligned}
\mathsf{val}_K(V) &= \frac{1}{K}\sum_{i=1}^{K}\theta_{\mathsf{ind}_i(V)} \ge \frac{1}{K}\sum_{i=1}^{K}\left((1 - X_i)\theta_{\mathsf{ind}_i(S)} + X_i p\right) \\
&= \frac{1}{K}\sum_{i=1}^{K}\theta_{\mathsf{ind}_i(S)} - \frac{1}{K}\sum_{i=1}^{K}(\theta_{\mathsf{ind}_i(S)} - p)X_i \ge \mathsf{val}_K(S) - \epsilon,
\end{aligned}$$

where the last inequality is due to $\mathcal{E}_1$. □

In light of Lemma A.1, it suffices to show that the probability that both $\mathcal{E}_1$ and $\mathcal{E}_2$ happen is at least $1 - \delta$. First, we bound $\Pr[\mathcal{E}_1]$ in the following lemma.

Lemma A.2. $\Pr[\mathcal{E}_1] \geq 1 - \dfrac{\delta}{2}$.

Before proceeding to the proof of the lemma, we state the following versions of the standard Chernoff-Hoeffding bounds, which will be useful later.

Proposition A.3. *Let $X_i (1 \leq i \leq n)$ be independent random variables with values in $[0, 1]$. Let $X = \frac{1}{n}\sum_{i=1}^{n} X_i$. The following statements hold:*

1. *For every $t > 0$, we have that*
$$\Pr\left[\, |X - \mathbf{E}[X]| > t\right] < 2\exp(-2t^2 n).$$
2. *Suppose $\mathbf{E}[X_i] < a$ for some real $0 \leq a \leq 1$. For every $t > 0$, we have that*
$$\Pr\left[X > a + t\right] < \left(\left(\frac{a}{a+t}\right)^{a+t}\left(\frac{1-a}{1-a-t}\right)^{1-a-t}\right)^{n}.$$
3. *For every $\epsilon > 0$, we have that*
$$\Pr\left[\, |X < (1-\epsilon)\,\mathbf{E}[X]\right] < \exp(-\epsilon^2 n\,\mathbf{E}[X]/2), \text{ and}$$
$$\Pr\left[\, |X > (1+\epsilon)\,\mathbf{E}[X]\right] < \exp(-\epsilon^2 n\,\mathbf{E}[X]/3).$$

Besides the above standard Chernoff-Hoeffding bounds, we also need the following Chernoff-type concentration inequality.

Proposition A.4. *Let $X_i (1 \leq i \leq K)$ be independent random variables. Each X_i takes value a_i ($a_i \geq 0$) with probability at most $\exp(-a_i^2 t)$ for some $t \geq 0$, and 0 otherwise. Let $X = \frac{1}{K}\sum_{i=1}^{K} X_i$. For every $\epsilon > 0$, when $t \geq \frac{2}{\epsilon^2}$, we have that*
$$\Pr\left[X > \epsilon\right] < \exp\left(-\epsilon^2 tK/2\right).$$

Proof of Proposition A.4. The proof is similar to that for the standard Chernoff bound. First, we observe that

$$\begin{aligned}\Pr[X > \epsilon] &= \Pr\left[\sum_{i=1}^{K} X_i > \epsilon K\right] = \Pr\left[\sum_{i=1}^{K} \epsilon t X_i > \epsilon^2 tK\right] = \Pr\left[\exp\left(\sum_{i=1}^{K} \epsilon t X_i\right) > \exp(\epsilon^2 tK)\right] \\ &\leq \mathbf{E}\left[\frac{\exp\left(\sum_{i=1}^{K} \epsilon t X_i\right)}{\exp(\epsilon^2 tK)}\right] = \frac{\prod_{i=1}^{K}\mathbf{E}[\exp(\epsilon t X_i)]}{\exp(\epsilon^2 tK)}, \end{aligned} \tag{7}$$

where the first inequality follows from Markov inequality and the last equality holds due to independence. Now, we claim that
$$\mathbf{E}[\exp(\epsilon t X_i)] \leq \exp\left(\epsilon^2 tK/2\right).$$
By the definition of X_i, combined with the fact that $a(\epsilon - a) \leq \epsilon^2/4$ for any real value a, it holds that

$$\mathbf{E}[e^{\epsilon t X_i}] \leq \exp(\epsilon a t - a^2 t) + 1 = \exp(a(\epsilon - a)t) + 1 \leq \exp\left(\epsilon^2 t/4\right) + 1.$$

When $\epsilon^2 t \geq 2$, we have $\exp\left(\epsilon^2 t/2\right) - \exp\left(\epsilon^2 t/4\right) > 1.06 > 1$ and hence $\mathbf{E}[e^{\epsilon t X_i}] \leq \exp\left(\epsilon^2 t/2\right)$. Plugging this bound into (7), we get that

$$\Pr[X > \epsilon] \leq \frac{\prod_{i=1}^{K}\exp\left(\epsilon^2 t/2\right)}{\exp(\epsilon^2 tK)} = \exp\left(-\frac{\epsilon^2 tK}{2}\right).$$

The proof is completed. $\square$

With the concentration inequality in place, now we are ready to prove Lemma A.2.

Proof of Lemma A.2. Recall that $Q_0 = \frac{Q}{|S|}$ is the number of samples taken from each arm in S. By the definition of ϵ in Lemma A.2, we trivially have that $\epsilon \geq \max\left\{\sqrt{\frac{10}{Q_0}}, \sqrt{\frac{4\ln(2/\delta)}{Q_0 K}}\right\}$. For each $i \in [K]$, let $\eta_i = \max\{\theta_{\mathsf{ind}_i(S)} - p - \frac{\epsilon}{2}, 0\}$ and let $Y_i = \eta_i X_i$. By Proposition A.3(1), we have that

$$\begin{aligned}\Pr[Y_i = \eta_i] = \Pr[X_i = 1] = \Pr\left[\widehat{\theta}_{\mathsf{ind}_i(S)} < p + \frac{\epsilon}{2}\right] = \Pr\left[\widehat{\theta}_{\mathsf{ind}_i(S)} < \theta_{\mathsf{ind}_i(S)} - (\theta_{\mathsf{ind}_i(S)} - p - \frac{\epsilon}{2})\right] \\ \leq \exp\left(-2\left(\theta_{\mathsf{ind}_i(S)} - p - \frac{\epsilon}{2}\right)^2 \cdot Q_0\right) \leq \exp\left(-\eta_i^2 \cdot 2Q_0\right).\end{aligned}$$

Applying Proposition A.4 on Y_i's, we can get that

$$\Pr\left[\frac{1}{K}\sum_{i=1}^{K} Y_i > \frac{\epsilon}{2}\right] \leq \exp\left(-\frac{\epsilon^2 Q_0 K}{4}\right) \leq \frac{\delta}{2},$$

where the last inequality holds because $\epsilon \geq \sqrt{\frac{4\ln(2/\delta)}{Q_0 K}}$. Observe that $Y_i \geq (\theta_{\mathsf{ind}_S(i)} - p)X_i - \frac{\epsilon}{2}$ for all $i \in [K]$. Therefore, with probability at least $1 - \frac{\delta}{2}$, we have that

$$X = \frac{1}{K}\sum_{i=1}^{K}(\theta_{\mathsf{ind}_i(S)} - p)X_i \leq \frac{1}{K}\sum_{i=1}^{K} Y_i + \frac{\epsilon}{2} \leq \epsilon.$$

This completes the proof of Lemma A.2. □

Next, we bound the probability that $\mathcal{E}_2$ happens in the following lemma.

Lemma A.5. $\Pr[\mathcal{E}_2] \geq 1 - \dfrac{\delta}{2}$.

Proof. First, we can see that $\mathcal{E}_2$ holds if and only if there are no more than $3|S|/4$ arms with empirical mean larger than $p + \frac{\epsilon}{2}$. Define the indictor random variable $Z_i = \mathbf{1}\{\widehat{\theta}_{\mathsf{ind}_i(S)} \geq p + \frac{\epsilon}{2}\}$. Hence, it suffice to show that $\Pr\left[\sum_{i=|S|/2}^{|S|} Z_i < \frac{|S|}{4}\right] \geq 1 - \frac{\delta}{2}$.

Let us only consider the arms with indices $i \in [|S|/2, |S|]$ (i.e., $\theta_i \leq p$). By Proposition A.3(1), we have

$$\Pr\left[\widehat{\theta}_{\mathsf{ind}_i(S)} \geq p + \frac{\epsilon}{2}\right] \leq \Pr\left[\widehat{\theta}_{\mathsf{ind}_i(S)} \geq \theta_{\mathsf{ind}_i(S)} + \frac{\epsilon}{2}\right] \leq \exp\left(-\frac{\epsilon^2}{2} \cdot Q_0\right). \tag{8}$$

From (8), we can see that $\mathbf{E}[Z_i] < \exp\left(-\frac{\epsilon^2}{2} \cdot Q_0\right)$. Let $\mu = \max_{i \in [|S|/2, |S|]} \mathbf{E}[Z_i]$ and we have $\mu < \exp\left(-\frac{\epsilon^2}{2} \cdot Q_0\right)$. Then, by Proposition A.3(2), we have

$$\begin{aligned}\Pr\left[\sum_{i=|S|/2}^{|S|} Z_i > \frac{|S|}{4}\right] &\leq \left(\left(\frac{\mu}{1/2}\right)^{1/2}\left(\frac{1-\mu}{1/2}\right)^{1/2}\right)^{|S|/2} \leq \left(\sqrt{2\mu} \cdot \sqrt{2}\right)^{|S|/2} \\ &\leq \exp\left(\frac{|S|}{2}\left(\ln(2) - \frac{\epsilon^2}{2} \cdot Q_0\right)\right) \leq \exp\left(-\frac{|S|}{2} \cdot \frac{\epsilon^2}{4} \cdot Q_0\right) \leq \exp\left(-\frac{\epsilon^2 K}{2} \cdot Q_0\right) \leq \frac{\delta}{2}.\end{aligned}$$

where the third to last inequality follows because of $\epsilon > \sqrt{\frac{10}{Q_0}}$, the second to last inequality uses the assumption that $|S|/4 \geq K$ and the last inequality holds because we assume that $\epsilon \geq \sqrt{\frac{4\ln(2/\delta)}{Q_0 K}}$. □

Proof of Lemma 4.1. By Lemma A.2, Lemma A.5, and a union bound, we have $\Pr[\mathcal{E}_1 \text{ and } \mathcal{E}_2] \geq 1 - \delta$. By Lemma A.1, we have $\Pr[\mathsf{val}_K(T) \geq \mathsf{val}_K(S) - \epsilon] \geq \Pr[\mathcal{E}_1 \text{ and } \mathcal{E}_2]$ and then the lemma follows. □

B Proof of the Correctness of the AR Algorithm (Lemma 4.2 in the Main Text)

Lemma 4.2 *Let (S', T') be the output of the algorithm* $\mathrm{AR}(S, T, K, Q)$. *For every $\delta > 0$, with probability $1-\delta$, we have that*

$$\mathsf{tot}_{K-|T'|}(S') + \mathsf{tot}_{|T'|}(T') \geq \mathsf{tot}_{K-|T|}(S) + \mathsf{tot}_{|T|}(T) - \epsilon K,$$

where $\epsilon = \sqrt{\frac{|S|}{Q}\left(4 + \frac{\log(2/\delta)}{K}\right)}$.

Proof. Recall that $Q_0 = \frac{Q}{|S|}$ is the number of samples taken from each arm in S. Also recall that $K' = K - |T|$.

We need to define a few notations. Let $U_1 = T' \setminus T$ denote the set of arms we added to T' in this round. Let $U_2 = S \setminus (S' \cup U_1)$ be the set of arms we discarded in this round. Let U_1^* be the set of $|U_1|$ arms in S with largest θ_i's; let U_2^* be the set of $|U_2|$ arms in S with smallest θ_i's. Ideally, if $U_1 = U_1^*$ and $U_2 = U_2^*$, we do not lose anything in this round (i.e., $\mathsf{tot}_{K-|T'|}(S') + \mathsf{tot}_{|T'|}(T') = \mathsf{tot}_{K-|T|}(S) + \mathsf{tot}_{|T|}(T)$). When $U_1 \neq U_1^*$ and/or $U_2 \neq U_2^*$, we can bound the difference between $\mathsf{tot}_{K-|T'|}(S') + \mathsf{tot}_{|T'|}(T')$ and $\mathsf{tot}_{K-|T|}(S) + \mathsf{tot}_{|T|}(T)$ by the sum of the difference between U_1^* and U_1, and the difference between U_2^* and U_2. More concretely, we claim that

$$\left(\mathsf{tot}_{K-|T'|}(S') + \mathsf{tot}_{|T'|}(T')\right) - \left(\mathsf{tot}_{K-|T|}(S) + \mathsf{tot}_{|T|}(T)\right) \geq \left(\sum_{i \in U_2^*} \theta_i - \sum_{i \in U_2} \theta_i\right) - \left(\sum_{i \in U_1} \theta_i - \sum_{i \in U_1^*} \theta_i\right). \quad (9)$$

The proof of (9) is not difficult, but somewhat tedious, and we present it at the end of this section. From now on, we assume (9) is true.

For every $t \leq K$, for every set $U \subseteq S$ of t arms (i.e. $|U| = t$), by Proposition A.3, we have

$$\Pr\left[\left|\sum_{i \in U} \hat{\theta}_i - \sum_{i \in U} \theta_i\right| > \frac{\epsilon K}{4}\right] \leq 2\exp\left(-\frac{\epsilon^2}{8} \cdot Q_0 \frac{K^2}{t}\right) \leq 2\exp\left(-\frac{\epsilon^2}{8} \cdot Q_0 K\right).$$

By a union bound over all subset of size at most K, we have that

$$\begin{aligned}\Pr\left[\forall U \subseteq S, |U| \leq K : \left|\sum_{i \in U} \hat{\theta}_i - \sum_{i \in U} \theta_i\right| \leq \frac{\epsilon K}{4}\right] &\geq 1 - 2 \cdot 2^{|S|} \exp\left(-\frac{\epsilon^2}{8} \cdot Q_0 K\right) \\ &\geq 1 - 2\exp\left(|S| - \frac{\epsilon^2}{8} \cdot Q_0 K\right) \geq 1 - \delta,\end{aligned}$$

where we used the facts that $|S| < 4K$ and $\epsilon \geq \sqrt{\frac{1}{Q_0}\left(4 + \frac{\ln(2/\delta)}{K}\right)}$.

Thus, with probability at least $1 - \delta$, all of the following four inequalities hold:

$$\left|\sum_{i \in U_1} \hat{\theta}_i - \sum_{i \in U_1} \theta_i\right| \leq \frac{\epsilon K}{4}, \qquad \left|\sum_{i \in U_1^*} \hat{\theta}_i - \sum_{i \in U_1^*} \theta_i\right| \leq \frac{\epsilon K}{4},$$

$$\left|\sum_{i \in U_2} \hat{\theta}_i - \sum_{i \in U_2} \theta_i\right| \leq \frac{\epsilon K}{4}, \qquad \left|\sum_{i \in U_2^*} \hat{\theta}_i - \sum_{i \in U_2^*} \theta_i\right| \leq \frac{\epsilon K}{4}.$$

Therefore we have

$$\sum_{i \in U_1} \theta_i \geq \sum_{i \in U_1} \hat{\theta}_i - \frac{\epsilon K}{4} \geq \sum_{i \in U_1^*} \hat{\theta}_i - \frac{\epsilon K}{4} \geq \sum_{i \in U_1^*} \theta_i - \frac{\epsilon K}{2}, \quad \text{and} \quad (10)$$

$$\sum_{i \in U_2} \theta_i \le \sum_{i \in U_2} \hat{\theta}_i - \frac{\epsilon K}{4} \le \sum_{i \in U_2^*} \hat{\theta}_i - \frac{\epsilon K}{4} \le \sum_{i \in U_2^*} \theta_i - \frac{\epsilon K}{2}. \tag{11}$$

Combining (9), (10) and (11), we get (9) $\ge -\epsilon K$, which concludes the proof. $\square$

Proof of (9). For ease of notation, for any subset S of arms, we let $\theta(S) = \sum_{i \in S} \theta_i$. One can easily see that

$$\begin{aligned}
&\big(\mathsf{tot}_{K-|T'|}(S') + \mathsf{tot}_{|T'|}(T')\big) - \big(\mathsf{tot}_{K-|T|}(S) + \mathsf{tot}_{|T|}(T)\big) \\
&\quad = \mathsf{tot}_{K-|T'|}(S') - \mathsf{tot}_{K-|T|}(S) + \theta(U_1) \\
&\quad = \mathsf{tot}_{K-|T'|}(S') - \mathsf{tot}_{K-|T'|}(S \setminus U_1^*) + (\theta(U_1) - \theta(U_1^*)) \\
&\quad \ge \mathsf{tot}_{K-|T'|}(S') - \mathsf{tot}_{K-|T'|}(S \setminus U_1) + (\theta(U_1) - \theta(U_1^*)) \,.
\end{aligned} \tag{12}$$

Let $\widetilde{U}_2$ be the $|U_2|$ arms with the smallest means in $S \setminus U_1$. By definition we have 1) $|\widetilde{U}_2| = |U_2^*|$; 2) $\widetilde{U}_2 \cap U_1 = U_2 \cap U_1 = \emptyset$; 3) $\theta(\widetilde{U}_2) \ge \theta(U_2^*)$.

Since $|U_1| + |U_2| + (K - |T'|) \le |S|$, the $(K - |T'|)$ arms with largest means in $S \setminus U_1$ do not intersect with the $|U_2|$ arms with smallest means in $S \setminus U_1$ (namely $\widetilde{U}_2$). Therefore, we have that

$$\mathsf{tot}_{K-|T'|}(S \setminus U_1) = \mathsf{tot}_{K-|T'|}((S \setminus U_1) \setminus \widetilde{U}_2). \tag{13}$$

On the other hand, for every set W of arms, define $\mathsf{tot}_t^{\min}(W)$ to be the sum of the t smallest means among the arms in W. Let $t = |S| - |U_1| - |U_2| - (K - |T'|)$. Since $\widetilde{U}_2$ consists of the arms with the smallest means in $S \setminus U_1$, we have

$$\mathsf{tot}_t^{\min}((S \setminus U_1) \setminus \widetilde{U}_2) \ge \mathsf{tot}_t^{\min}((S \setminus U_1) \setminus U_2).$$

Together with the facts that

$$\mathsf{tot}_t^{\min}((S \setminus U_1) \setminus \widetilde{U}_2) = \theta((S \setminus U_1) \setminus \widetilde{U}_2) - \mathsf{tot}_{K-|T'|}((S \setminus U_1) \setminus \widetilde{U}_2), \text{ and}$$

$$\mathsf{tot}_t^{\min}((S \setminus U_1) \setminus U_2) = \theta(S \setminus U_1) \setminus U_2) - \mathsf{tot}_{K-|T'|}((S \setminus U_1) \setminus U_2),$$

we can see that

$$\theta(S \setminus U_1) \setminus \widetilde{U}_2) - \mathsf{tot}_{K-|T'|}((S \setminus U_1) \setminus \widetilde{U}_2) \ge \theta(S \setminus U_1) \setminus U_2) - \mathsf{tot}_{K-|T'|}((S \setminus U_1) \setminus U_2).$$

Equivalently, we have that

$$\mathsf{tot}_{K-|T'|}((S \setminus U_1) \setminus U_2) - \mathsf{tot}_{K-|T'|}((S \setminus U_1) \setminus \widetilde{U}_2) \ge \theta(\widetilde{U}_2) - \theta(U_2). \tag{14}$$

By combining (12), (13) and (14), and the observations that $S' = (S \setminus U_1) \setminus U_2$ and $\theta(\widetilde{U}_2) \ge \theta(U_2^*)$, we have proved (9). $\square$

C Proof of the Main Theorem (Theorem 4.3 in the Main Text)

Theorem 4.3 *For every $\delta > 0$, with probability at least $1-\delta$, the output of OptMAI algorithm T is an ϵ-optimal solution (i.e., $\mathsf{val}_K(T) \ge \mathsf{val}_K([n]) - \epsilon$) with $\epsilon = O\left(\sqrt{\frac{n}{Q}\left(1 + \frac{\ln 1/\delta}{K}\right)}\right)$.*

Proof. Recall r is the counter of the number of iterations in Algorithm 1. Let r_0 be the first r such that we have $|S_r| < 4K$. Let r_1 be the final value of r. For any positive integer r, let

$$\delta_r = e^{-.1r}(1 - e^{-.1})\delta \quad \text{and} \quad \epsilon_r = O\left(\sqrt{\frac{(\frac{3}{4})^r n}{(1-\beta)\beta^r Q}\left(1 + \frac{\ln 1/\delta_r}{K}\right)}\right).$$

For $r < r_0$, by Lemma 4.1, with probability $1 - \delta_r$, we have that $\mathsf{val}_K(S_{r+1}) \geq \mathsf{val}_K(S_r) - \epsilon_r$. By union bound, with probability $1 - \sum_{r=0}^{r_0-1} \delta_r$, we have that

$$\mathsf{val}_K(S_{r_0}) \geq \mathsf{val}_K([n]) - \sum_{r=0}^{r_0-1} \epsilon_r. \tag{15}$$

For $r : r_0 \leq r < r_1$, by Lemma 4.2, with probability $1 - \delta_r$, we have that

$$\left(\mathsf{tot}_{K-|T_{r+1}|}(S_{r+1}) + \mathsf{tot}_{|T_{r+1}|}(T_{r+1})\right) - \left(\mathsf{tot}_{K-|T_r|}(S_r) + \mathsf{tot}_{|T_r|}(T_r)\right) \geq K \cdot \epsilon_r$$

Since T_{r_1} has exactly K elements and $T_{r_0} = \emptyset$, by union bound, with probability $1 - \sum_{r=r_0}^{r_1-1} \delta_r$, we can see that

$$\mathsf{val}_K(T_{r_1}) \geq \mathsf{val}_K(S_{r_0}) - \sum_{r=r_0}^{r_1-1} \epsilon_r. \tag{16}$$

Now, by a union bound over both (15) and (16), we have that, with probability $1 - \sum_{r=0}^{r_1-1} \delta_r \geq 1 - \delta$,

$$\begin{aligned}
\mathsf{val}_K([n]) - \mathsf{val}_K(T_{r_1}) &\leq \sum_{r=0}^{r_1-1} \epsilon_r = \sum_{r=0}^{r_1-1} O\left(\sqrt{\left(\frac{3/4}{\beta(1-\beta)}\right)^r \left(\frac{n}{Q}\right)\left(1 + \frac{\ln 1/\delta_r}{K}\right)}\right) \\
&\leq \sum_{r=0}^{r_1-1} O\left(\sqrt{\left(\frac{3/4}{\beta(1-\beta)}\right)^r \left(\frac{n}{Q}\right)\left(1 + \frac{\ln 1/\delta + 0.1r + \ln(1-e^{-.1})}{K}\right)}\right) \\
&= O\left(\sqrt{\frac{n}{Q}\left(1 + \frac{\ln 1/\delta}{K}\right)}\right).
\end{aligned}$$

This completes the proof of the theorem. □

D An Alternative to the AR Procedure

We can replace the AR procedure by the following uniform sampling procedure $B(S_r, K, \epsilon', \delta')$, when the number of remaining arms $|S_r|$ is at most $4K$. Using this alternative procedure, we can achieve the same asymptotic sampling complexity, and its analysis is slightly simpler. However, its performance in practice is worse than the AR procedure. Note that the condition $|S_r| \leq 4K$ is crucial for the uniform sampling procedure to achieve the desired sample complexity (otherwise, we need to pay an extra $\log n$ factor. See Section E for more information).

More specifically, the algorithm takes as input the remaining subset of arms $S_r \subseteq [n]$, an integer K, and two real numbers $\epsilon', \delta' > 0$ as input, and outputs a set $T \subseteq S_r$ such that $|T| = K$. We set $\epsilon' = \epsilon/2, \delta' = \delta/2$. Note that we only run $B(S_r, K, \epsilon', \delta')$ once and its output T is our final output of the entire algorithm. The algorithm proceeds as follows.

- Sample each arm $i \in S_r$ for

$$Q_0 = Q_B(K, \epsilon', \delta') = \frac{2(K \ln(e|S_r|/K) + \ln(2/\delta'))}{\epsilon'^2 K} = O\left(\frac{1}{\epsilon'^2}\left(1 + \frac{\log 1/\delta'}{K}\right)\right).$$

  times and let $\hat{\theta}_i$ be the empirical mean of arm i.

- Output the set $T \subseteq S_r$ which is the set of K arms with the largest empirical means.

It is easy to see the number of samples is bounded by $|S_r|Q_0 \le O\left(\frac{K}{\epsilon'^2}\left(1+\frac{\log 1/\delta'}{K}\right)\right)$. The above algorithm can achieve the following performance guarantee.

Lemma D.1. *Let T be the output of the algorithm $B(S_r, K, \epsilon', \delta')$. With probability $1-\delta'$, we have that $\mathsf{val}_K(T) \ge \mathsf{val}_K(S_r) - \epsilon'$.*

Proof. For every set $U \subseteq S_r$ of K arms (i.e. $|U| = K$), by Proposition A.3(1), we have

$$\Pr\left[\left|\frac{1}{|U|}\sum_{i\in U}\hat{\theta}_i - \frac{1}{|U|}\sum_{i\in U}\theta_i\right| > \frac{\epsilon'}{2}\right] \le 2\exp\left(-\frac{\epsilon'^2}{2}\cdot Q_0 K\right).$$

By union bound over all subsets of size K, we have that

$$\Pr\left[\forall U \subseteq S_r, |U| = K : \left|\frac{1}{|U|}\sum_{i\in U}\hat{\theta}_i - \frac{1}{|U|}\sum_{i\in U}\theta_i\right| \le \frac{\epsilon'}{2}\right] \ge 1 - 2\binom{|S_r|}{k}\exp\left(-\frac{\epsilon'^2}{2}\cdot Q_0 K\right)$$

$$\ge 1 - 2\left(\frac{e|S_r|}{K}\right)^K \exp\left(-\frac{\epsilon'^2}{2}\cdot Q_0 K\right) = 1 - 2\exp\left(K\log(e|S_r|/K) - \frac{\epsilon'^2}{2}\cdot Q_0 K\right) \ge 1-\delta'.$$

Let T^* be the set of K arms in S_r with largest θ_i's. With probability at least $1-\delta'$, we have

$$\left|\frac{1}{|T|}\sum_{i\in T}\hat{\theta}_i - \frac{1}{|T|}\sum_{i\in T}\theta_i\right| \le \frac{\epsilon'}{2}, \qquad \left|\frac{1}{|T^*|}\sum_{i\in T^*}\hat{\theta}_i - \frac{1}{|T^*|}\sum_{i\in T^*}\theta_i\right| \le \frac{\epsilon'}{2}.$$

Therefore, we can get that

$$\mathsf{val}_K(T) = \frac{1}{|T|}\sum_{i\in T}\theta_i \ge \frac{1}{|T|}\sum_{i\in T}\hat{\theta}_i - \frac{\epsilon'}{2} \ge \frac{1}{|T^*|}\sum_{i\in T^*}\hat{\theta}_i - \frac{\epsilon'}{2} \ge \frac{1}{|T^*|}\sum_{i\in T^*}\theta_i - \epsilon' = \mathsf{val}_K(S_r) - \epsilon'.$$

□

If we set $Q = O\left(\frac{n}{\epsilon^2}\left(1+\frac{\ln(1/\delta)}{K}\right)\right)$ in OPTMAI, the proof of Theorem 4.3 show that, after the QE stage, the set S_r of remaining arms satisfies that $\mathsf{val}([n]) - \mathsf{val}_K(S_r) \le \epsilon/2$ with probability at least $1-\delta/2$. Combined with the conclusion of Lemma D.1 and $\epsilon' = \epsilon/2, \delta' = \delta/2$, we get that $\mathsf{val}([n]) - \mathsf{val}_K(T) \le \epsilon$ with probability at least $1-\delta$. The number of samples used in both QE and the uniform sampling stages is at most $Q + |S_r|Q_0 = O\left(\frac{n}{\epsilon^2}\left(1+\frac{\ln(1/\delta)}{K}\right)\right)$, which is the same as the sample complexity stated in Corollary 4.4.

E Naive Uniform Sampling

As we have seen in Section D, we can use a uniform sampling procedure to replace AR. We show in this section that, simply using following naive uniform sampling as the entire algorithm is not sufficient to achieve the linear sample complexity.

Naive Uniform Sampling:

- Sample each arm $i \in S$ for Q_0 times and let $\hat{\theta}_i$ be the empirical mean of arm i.
- Output the set $T \subseteq S$ which is the set of K arms with the largest empirical means.

In fact, when $K = 1$, Even-Dar et al. (2006) showed that $Q_0 = O(\frac{1}{\epsilon^2}\log\frac{n}{\delta})$ (note this is $\log n$ factor worse than the optimal bound) is enough to identify an ϵ-optimal arm with probability at least $1-\delta$. For general K, by following the same proof of Lemma D.1, we can show that

$$Q_0 = O\left(\frac{1}{\epsilon^2}\left(\log\frac{n}{K} + \frac{\log 1/\delta}{K}\right)\right) \tag{17}$$

suffices for identifying an ϵ-optimal solution with probability at least $1-\delta$. Moreover, we can also show the bound (17) is essential tight for naive uniform sampling, as in the following theorem.

Theorem E.1. *Suppose that for any multiple arm identification problem instance and any $0 < \epsilon, \delta < 0.01$ and $1 \le K \le n/2$, the naive uniform sampling algorithm with parameter Q_0 can find an ϵ-optimal solution with probability at least $1-\delta$. Then, it must hold that $Q_0 = \max\left\{\Omega\left(\frac{1}{\epsilon^2}\log\frac{n}{K}\right), \Omega\left(\frac{1}{\epsilon^2}\frac{\log 1/\delta}{K}\right)\right\}$.*

Proof. In fact, the second lower bound $\Omega\left(\frac{1}{\epsilon^2}\frac{\log 1/\delta}{K}\right)$ holds for any algorithm (including the uniform sampling algorithm), which is proved in Lemma 5.4. So, we only focus on the first lower bound in this proof. Let C be a sufficiently large constant ($C > 2^{10000}$ suffices). First we consider the case where $1 \le K \le n/C$. Consider the instance which consists of K Bernoulli arms with mean $1/2 + 4\epsilon$ (denoted as set A) and $n-K$ Bernoulli arms with mean $1/2$ (denoted as set B). It is easy to see that any ϵ-optimal solution must contain at least $\frac{3}{4}K$ arms from A. Let $Q_0 = \frac{1}{400\epsilon^2}\log\frac{n}{K}$. Now, we show that with probability at least 0.05, there are at least $K/4$ arms from B whose empirical mean is at least $1/2 + 8\epsilon$ and at least $K/4$ arms from A whose empirical mean is smaller than $1/2 + 8\epsilon$. We denote the former event by $\mathcal{E}_1$ and the later by $\mathcal{E}_2$. Note that if the event that both $\mathcal{E}_1$ and $\mathcal{E}_2$ happen implies that we fail to find an ϵ-optimal solution.

First, let us consider $\mathcal{E}_1$. Let $Y_i = \mathbf{1}\{\hat{\theta}_i \ge 1/2 + 8\epsilon\}$. For any arm i in B, we have that

$$\Pr[Y_i = 1] = \Pr[\hat{\theta}_i \ge 1/2 + 8\epsilon] = \left(\frac{1}{2}\right)^{Q_0} \sum_{i=(1/2+8\epsilon)Q_0}^{Q_0} \binom{Q_0}{i} \ge \left(\frac{1}{2}\right)^{200Q_0\epsilon^2} \ge \left(\frac{n}{K}\right)^{-1/2},$$

where the second to last inequality follows from the fact that $\sum_{k \le \alpha m}\binom{m}{k} \ge 2^{mH(\alpha) - \log m}$ ($H(\alpha)$ is the binary entropy function) Ronald et al. (1989) and the Tylor expansion of $H(\alpha)$ around 1/2: $H(1/2-\epsilon) \simeq 1 - 2\epsilon^2/\ln 2 + o(\epsilon^2)$. Therefore, in expectation, there are at least $(n-K)\left(\frac{n}{K}\right)^{-1/2}$ arms in B whose empirical mean is at least $1/2 + 8\epsilon$, i.e., $\mathbf{E}[\sum_{i \in B} Y_i] \ge (n-K)\left(\frac{n}{K}\right)^{-1/2}$. Using Proposition A.3(3), we can see that (for $n \ge CK \ge C$)

$$\Pr\left[\sum_{i \in B} Y_i < \frac{K}{4}\right] \le \exp\left(-\left(\frac{1}{2}\right)^2 (n-K)\left(\frac{n}{K}\right)^{-1/2}/2\right) < 0.05,$$

where we use the fact that $(n-K)\left(\frac{n}{K}\right)^{-1/2} \ge K/2$ in the first inequality, and that $(n-K)\left(\frac{n}{K}\right)^{-1/2} \ge \frac{n-1}{\sqrt{n}}$ for $1 \le K \le n/2$ in the second. Hence, with probability at least 0.95, there are at least $K/4$ arms in B whose empirical mean is at least $1/2 + 8\epsilon$.

For any arm in A, using Proposition A.3(2), we can see that

$$\Pr[\hat{\theta}_i \ge 1/2 + 8\epsilon] \le \exp\left(-\frac{1}{4}\epsilon^2\left(\frac{1}{2} + 4\epsilon\right)Q_0/3\right) < 0.5,$$

in which the last inequality holds because $Q_0 \ge \frac{1}{400\epsilon^2}\log C$. Let $Z_i = \mathbf{1}\{\hat{\theta}_i \ge 1/2 + 8\epsilon\}$. Let $\mu = \exp\left(-\frac{1}{4}\epsilon^2\left(\frac{1}{2} + 4\epsilon\right)Q_0/3\right) < 0.5$. Then, by Proposition A.3(2), we have

$$\Pr[\neg\mathcal{E}_2] = \Pr\left[\sum_{i \in A} Z_i > \frac{3K}{4}\right] \quad \le \quad \left(\left(\frac{\mu}{3/4}\right)^{3/4}\left(\frac{1-\mu}{1/4}\right)^{1/4}\right)^K < 0.877.$$

The last inequality holds since $(\frac{\mu}{3/4})^{3/4}(\frac{1-\mu}{1/4})^{1/4}$ is an increasing function on $[0, 0.5]$, thus is maximized at $u = 0.5$. Hence, $\Pr[\mathcal{E}_2] \ge 0.1$. So, we have $\Pr[\mathcal{E}_1 \text{ and } \mathcal{E}_2] \ge 0.05$ and the proof is complete for the case $K \le n/C$. When, $n/C \le K \le n/2$, the desired bound becomes $\Omega(1/\epsilon^2)$, which follows from Theorem 5.1. $\square$

F Lower bounds

F.1 Proof of the First Lower Bound (Lemma 5.2 in the Main Text)

Lemma 5.2 *Let $\mathcal{A}$ be an algorithm in Theorem 5.1, then there is an algorithm $\mathcal{B}$ which correctly outputs whether a Bernoulli arm X has the mean $\frac{1}{2} + 4\epsilon$ or the mean $\frac{1}{2}$ with probability at least 0.51, and $\mathcal{B}$ makes* *at most $\frac{200Q}{n}$ samples in expectation.*

Proof. Given an algorithm $\mathcal{A}$ stated in Theorem 5.1, the construction of $\mathcal{B}$ is as follows. Recall that the goal of $\mathcal{B}$ is to distinguish whether the given Bernoulli arm X has mean $1/2$ or $1/2 + 4\epsilon$.

Algorithm 5 Algorithm $\mathcal{B}$ (which calls $\mathcal{A}$ as a subroutine)

1: Choose a random subset $S \subseteq [n]$ such that $|S| = K$ and then choose a random element $j \in S$.
2: Create n artificial arms as follows: For each $i \in [n], i \neq j$, let $\theta_i = \frac{1}{2} + 4\epsilon$ if $i \in S$, let $\theta_i = \frac{1}{2}$ otherwise.
3: Simulate $\mathcal{A}$ as follows: whenever $\mathcal{A}$ samples the i-th arm:
   (1) If $i = j$, we sample the Bernoulli arm X (recall X is the arm which $\mathcal{B}$ attempts to separate);
   (2) Otherwise, we sample the arm with mean θ_i.
4: If the arm X is sampled by less than $\frac{200Q}{n}$ times and $\mathcal{A}$ returns a set T such that $j \notin T$, we decide that X has the mean of $\frac{1}{2}$; otherwise we decide that X has the mean of $\frac{1}{2} + 4\epsilon$.

We note that in step 3(1), only when $\mathcal{A}$ attempts to sample the j-th (artificial) arm, we *actually* take a sample from X. Since $\mathcal{B}$ stops and output the mean $\frac{1}{2} + 4\epsilon$ if the number of trials on X reaches $\frac{200Q}{n}$, $\mathcal{B}$ takes at most $\frac{200Q}{n}$ samples form X. Now, we $\mathcal{B}$ can correctly output the mean of X with the probability at least 0.51.

We first show that when the Bernoulli arm X has mean $\frac{1}{2}$, $\mathcal{B}$ decides correctly with probability at least 0.51. Assuming that X has the mean $\frac{1}{2}$, among the n arms in the algorithm $\mathcal{A}$, the arms in $S \setminus \{j\}$ have mean $\frac{1}{2} + 4\epsilon$, while others have mean $\frac{1}{2}$. For each $i \in [n]$, let the random variable q_i be the number of samples taken from the i-th arm by $\mathcal{A}$. We have

$$\sum_{i \in [n]} \mathbf{E}[q_i] \leq Q.$$

Let the random variable q_X be the number of samples taken from arm X and $S' = S \setminus \{j\}$. Observe that when conditioned on S', for $\mathcal{A}$, j is the same as any other arms in $[n] \setminus S'$, hence, j is uniformly distributed among $[n] \setminus S'$. We have

$$\begin{aligned}
\mathbf{E}[q_X] &= \mathbf{E}_{S'}\left[\mathbf{E}[q_X \mid S']\right] = \mathbf{E}_{S'}\left[\frac{1}{n-K+1}\sum_{i \in [n]\setminus S'} \mathbf{E}[q_i \mid S']\right] \\
&\leq \frac{1}{n-K+1}\mathbf{E}_{S'}\left[\sum_{i \in [n]} \mathbf{E}[q_i \mid S']\right] \leq \frac{2}{n}\sum_{i \in [n]} \mathbf{E}[q_i] \leq \frac{2Q}{n},
\end{aligned}$$

where in the second equality we use the fact that j is uniformly distributed among $[n] \setminus S'$ conditioned on S', and in the second inequality we used the assumption that $K \leq n/2$. Therefore, by Markov's inequality,

$$\Pr\left[q_X \geq \frac{200Q}{n}\right] < 0.01.$$

Let T be the output of the algorithm $\mathcal{A}$. It is easy to see that $\mathsf{val}_K([n]) = \frac{1}{2} + 4\epsilon \cdot (1 - \frac{1}{K})$, and $\mathsf{val}_K(T) = \frac{1}{2} + 4\epsilon \cdot \frac{|S' \cap T|}{K}$. When $\mathcal{A}$ finds an ϵ-optimal solution (i.e., $\mathsf{val}_K(T) \geq \mathsf{val}_K([n]) - \epsilon$), we have

$$\frac{1}{2} + 4\epsilon \cdot \frac{|S' \cap T|}{K} \geq \frac{1}{2} + 4\epsilon \cdot \left(1 - \frac{1}{K}\right) - \epsilon,$$

from which we can get that $|S' \cap T| \geq \frac{3}{4}K - 1$. Since $\mathcal{A}$ can find an ϵ-optimal solution with probability at least 0.8, for any fixed subset S', we have that

$$\Pr\left[|S' \cap T| \geq \frac{3}{4}K - 1\right] \geq \Pr[\mathsf{val}_K(T) \geq \mathsf{val}_K([n]) - \epsilon] \geq 0.8.$$

Conditioned on S', j is uniformly distributed among $[n] \setminus S'$ and is independent from T. Therefore, we have

$$\begin{aligned}\Pr[j \in T] = \mathbf{E}_{S'}\left[\Pr[j \in T \mid S']\right] = \mathbf{E}_{S',T}\left[\frac{|([n] \setminus S') \cap T|}{|[n] \setminus S'|}\right] \\ \leq 0.2 \times 1 + 0.8 \cdot \frac{\frac{1}{4}K + 1}{n - K + 1} \leq 0.2 + 0.8 \times (0.25 + 0.1) = 0.48,\end{aligned}$$

where in the last inequality, we used the assumption that $10 \leq K \leq n/2$. Therefore, when X has the mean $\frac{1}{2}$, we have that

$$\Pr\left[\mathcal{B} \text{ decides that } X \text{ has mean } \frac{1}{2}\right] = \Pr\left[j \notin T \text{ and } q_X \leq \frac{200Q}{n}\right] \geq 0.52 - 0.01 = 0.51.$$

Now we assume that X has the mean $\frac{1}{2} + 4\epsilon$. The proof is similar as before. Among the n arms, the arms in S have the mean $\frac{1}{2} + 4\epsilon$, while others have the mean $\frac{1}{2}$. Again, let T be the output of the algorithm $\mathcal{A}$. Since with probability at least 0.8, we have that $\mathsf{val}_K(T) \geq \mathsf{val}_K([n]) - \epsilon$, we have

$$\Pr\left[|S \cap T| \geq \frac{3}{4}K\right] \geq 0.8.$$

Since j is a uniformly distributed in S, and conditioned on S, j is independent from T, we have that

$$\Pr[j \in T] = \mathbf{E}_S\left[\Pr[j \in T \mid S]\right] = \mathbf{E}_{S,T}\left[\frac{|S \cap T|}{|S|}\right] \geq 0.8 \cdot \frac{3}{4} \geq 0.6.$$

In sum, when X has the mean $\frac{1}{2} + 4\epsilon$, we have that

$$\Pr\left[\mathcal{B} \text{ decides that } X \text{ has mean } \frac{1}{2} + 4\epsilon\right] \geq \Pr\left[j \in T\right] \geq 0.6 > 0.51.$$

In either case, $\mathcal{B}$ makes the right decision with probability at least 0.51. $\square$

F.2 Proof of the Second Lower Bound (Lemma 5.4 and Theorem 5.5 in the Main Text)

Lemma 5.4 *Fix real numbers δ, ϵ such that $0 < \delta, \epsilon \leq 0.01$, and integers K, n such that $K \leq n/2$. Let $\mathcal{A}$ be a deterministic algorithm (i.e., the only randomness comes from the sampling the arms), so that for any set of n Bernoulli arms with means $\theta_1, \theta_2, \ldots, \theta_n$,*

- *$\mathcal{A}$ makes at most Q samples in expectation;*
- *with probability at least $1 - \delta$, $\mathcal{A}$ outputs a set T of size K with $\mathsf{val}_K(T) \geq \mathsf{val}_K([n]) - \epsilon$.*

Then, we have that $Q \geq \frac{n \ln(1/\delta)}{20000\epsilon^2 K}$.

Proof of Lemma 5.4. Let $t = \lfloor \frac{n}{K} \rfloor \geq 2$ and we divide the first tK arms into t groups. The j-th group consists of the arms with the index in $[(j-1)K + 1, jK]$ for $j \in [t]$. We first construct t hypotheses $H_1, H_2, \ldots, H_t$ as follows. In H_1, we let $\theta_i = 1/2 + 4\epsilon$ for arms in the first group and let $\theta_i = 1/2$ for the remaining arms. In H_j, where $2 \leq j \leq t$, we let $\theta_i = 1/2 + 4\epsilon$ when i is in the first group, $\theta_i = 1/2 + 8\epsilon$ when i is in the j-th group, and $\theta_i = 1/2$ otherwise. For each $j \in [t]$, let $\Pr_{H_j}[\cdot]$ denote the probability of the event in $[\cdot]$ under the hypothesis H_j and $\mathbf{E}_{H_j}[\cdot]$ the expected value of the random variable in $[\cdot]$ under the hypothesis H_j.

For each arm $i \in [n]$, let the random variable q_i be the number of times that $\mathcal{A}$ samples the i-th arm before termination. For each $j \in [t]$, let the random variable $\widetilde{q_j} = \sum_{i=(j-1)K+1}^{jK} q_i$ be the total number of trials of the arms in the j-th group. Since $\mathcal{A}$ makes at most Q samples in expectation, we know that $\mathbf{E}_{H_1}\left[\sum_{j=2}^{t}\widetilde{q}_j\right] \le \mathbf{E}_{H_1}\left[\sum_{i\in[n]} q_i\right] \le Q$. By an averaging argument, there exists j_0 with $2 \le j_0 \le t$ such that

$$\mathbf{E}_{H_1}[\widetilde{q}_{j_0}] \le \frac{Q}{t-1} \le \frac{2Q}{t}.$$

Using Markov's inequality and letting $Q_0 = \frac{8Q}{t}$, we have $\Pr_{H_1}\left[\widetilde{q}_{j_0} \ge Q_0\right] \le \frac{\mathbf{E}_{H_1}[\widetilde{q}_{j_0}]}{Q_0} \le \frac{1}{4}$, and hence,

$$\Pr_{H_1}\left[\widetilde{q}_{j_0} \le Q_0\right] \ge \frac{3}{4}. \tag{18}$$

Now we only focus on the hypotheses H_1 and H_{j_0}. Let $\mathsf{val}_K^{H_j}(T)$ ($\mathsf{val}_K^{H_j}([n])$ resp.) be $\mathsf{val}_K(T)$ ($\mathsf{val}_K([n])$ resp.) under the hypothesis H_j. In other words, $\mathsf{val}_K^{H_j}(T)$ is the average mean value of the best K arms in T, if the the means of the arms are dictated by hypothesis H_j.

Now, we assume for contradiction that $Q < \frac{n\ln(1/\delta)}{20000\epsilon^2 K}$ (i.e., $Q_0 < \frac{\ln(1/\delta)}{1250\epsilon^2}$). Let T denote the output of $\mathcal{A}$. First, using the assumption that

$$\Pr_{H_1}[\mathsf{val}_K^{H_1}(T) \ge \mathsf{val}_K^{H_1}([n]) - \epsilon] \ge 1-\delta \tag{19}$$

(i.e., if the underlying hypothesis is H_1, T is an ϵ-optimal solution), we can prove that:

$$\Pr_{H_{j_0}}[\mathsf{val}_K^{H_1}(T) \ge \mathsf{val}_K^{H_1}([n]) - \epsilon] \ge \frac{\sqrt{\delta}}{4}. \tag{20}$$

We further observe that when $\mathsf{val}_K^{H_1}(T) \ge \mathsf{val}_K^{H_1}([n]) - \epsilon$, T must consist of more than $\frac{3}{4}K$ arms from the first group; while when $\mathsf{val}_K^{H_{j_0}}(T) \ge \mathsf{val}_K^{H_{j_0}}([n]) - \epsilon$, T must consist of more than $\frac{3}{4}K$ arms from the j_0-th group. Therefore, the two events are mutually exclusive and we have:

$$\begin{aligned}\Pr_{H_{j_0}}\left[\mathsf{val}_K^{H_{j_0}}(T) \ge \mathsf{val}_K^{H_{j_0}}([n]) - \epsilon\right] &\le 1 - \Pr_{H_{j_0}}\left[\mathsf{val}_K^{H_1}(T) \ge \mathsf{val}_K^{H_1}([n]) - \epsilon\right]\\ &\le 1-\frac{\sqrt{\delta}}{4} \le 1-2\delta,\end{aligned}$$

where the last inequality holds because $\delta < 0.01$. This essentially says that if the underlying hypothesis is H_{j_0}, the probability that $\mathcal{A}$ finds an ϵ-optimal solution is not large enough, which contradicts the performance guarantees of the Algorithm $\mathcal{A}$, and thus we conclude our proof.

Therefore, the remaining task is to prove (20). We first define a sequence of random variables $Z_0, Z_1, Z_2, \ldots, Z_{Q_0}$ where $Z_0 = 0$. For each $i \in [Q_0]$, if the i-th trial of the j_0-th group by $\mathcal{A}$ results in 1, let $Z_i = Z_{i-1}+1$; if the result is 0, let $Z_i = Z_{i-1}-1$; if $\mathcal{A}$ terminates before the i-th trial of the j_0-th group, let $Z_i = Z_{i-1}$. Under hypothesis H_0, the sequence $\{Z_0, Z_1, Z_2, \ldots, Z_{Q_0}\}$ forms a martingale since arms in the j_0-th group are independent zero-mean random variables. Therefore, by Azuma-Hoeffding's inequality, we have

$$\Pr_{H_1}\left[|Z_{Q_0}| \le \sqrt{5Q_0}\right] > 1 - 2\exp\left(-\frac{(\sqrt{5Q_0})^2}{2Q_0}\right) > \frac{3}{4}. \tag{21}$$

By a union bound over (18), (19) and (21), we have

$$\Pr_{H_1}\left[\mathsf{val}_K^{H_1}(T) \ge \mathsf{val}_K^{H_1}([n]) - \epsilon \text{ and } \widetilde{q}_{j_0} \le Q_0 \text{ and } |Z_{Q_0}| \le \sqrt{5Q_0}\right] \ge 1-\delta-\frac{1}{4}-\frac{1}{4} \ge \frac{1}{4}. \tag{22}$$

For ease of notation, we use $\mathcal{E}$ to denote the event that all of the following three events happen: (1) $\mathsf{val}_K^{H_1}(T) \ge \mathsf{val}_K^{H_1}([n]) - \epsilon$, (2) $\widetilde{q}_{j_0} \le Q_0$ and (3) $|Z_{Q_0}| \le \sqrt{5Q_0}$.

Suppose that $\mathcal{A}$ uses exactly Q' trials. We call a string $y = ((i_1, b_1), (i_2, b_2), \ldots, (i_{Q'}, b_{Q'}))$ a transcript for a particular execution of $\mathcal{A}$ if the r-th trial ($1 \le r \le Q'$) performed by $\mathcal{A}$ is the i_r-th arm and the result is $b_r \in \{0,1\}$. Let $\mathcal{Y}$ be the set of transcripts for $\mathcal{A}$. For each $y \in \mathcal{Y}$, we define the following quantities:

- Let $u_0^i(y)$ be the number of $(i,0)$ pairs in y and $u_1^i(y)$ be the number of $(i,1)$ pairs in y;
- Let $q_i(y) = u_0^i(y) + u_1^i(y)$ be the number of times $\mathcal{A}$ takes sample from the i-th arm in y;
- Let $\widetilde{u}_0^j(y) = \sum_{i=(j-1)K+1}^{jK} u_0^i(y)$ be the number of times that sampling from the j-th group results 0; $\widetilde{u}_1^j(y) = \sum_{i=(j-1)K+1}^{jK} u_1^i(y)$ be the number of times that sampling from the j-th group results 1;
- For all $j \in [t]$, let $\widetilde{q}_j(y) = \sum_{i=(j-1)K+1}^{jK} q_i(y) = \widetilde{u}_0^j(y) + \widetilde{u}_1^j(y)$ be the total number of samples taken from the j-th group in y.
- let $T(y)$ be the output of $\mathcal{A}$ when the transcript generated by $\mathcal{A}$ is y (note that the output of $\mathcal{A}$ is completed determined by y since $\mathcal{A}$ is deterministic).

Let the random variable Y be the transcript generated by $\mathcal{A}$. We use $\mathcal{E}_y$ to denote the event that $\mathsf{val}_K^{H_1}(T(y)) \geq \mathsf{val}^{H_1}(K) - \epsilon$ and $\widetilde{q}_{j_0}(y) \leq Q_0$ and $|\widetilde{u}_0^{j_0}(y) - \widetilde{u}_1^{j_0}(y)| \leq \sqrt{5Q_0}$. It is not hard to see that an equivalent way to write (22) is as follows:

$$\sum_{y\in\mathcal{Y}} \mathbf{1}\{\mathcal{E}_y\} \cdot \Pr_{H_1}[Y=y] \geq \frac{1}{4}. \tag{23}$$

Now, we claim that for any $y \in \mathcal{Y}$, when $\mathcal{E}_y$ is true, we have that

$$\frac{\Pr_{H_{j_0}}[Y=y]}{\Pr_{H_1}[Y=y]} \geq \sqrt{\delta}. \tag{24}$$

Therefore, we have

$$\begin{aligned}\Pr_{H_{j_0}}\left[\mathsf{val}_K^{H_1}(T) \geq \mathsf{val}_K^{H_1}([n]) - \epsilon\right] &\geq \Pr_{H_{j_0}}[\mathcal{E}] = \sum_{y\in\mathcal{Y}} \mathbf{1}\{\mathcal{E}_y\} \cdot \Pr_{H_{j_0}}[Y=y] \\ &\geq \sum_{y\in\mathcal{Y}} \mathbf{1}\{\mathcal{E}_y\} \cdot \Pr_{H_1}[Y=y] \cdot \sqrt{\delta} \geq \frac{\sqrt{\delta}}{4},\end{aligned}$$

where the second inequality is because of (24), and the last inequality is because of (23). Therefore, we finish the proof of the Eq. (20), which concludes the proof the lemma.

What remains is to prove the claim (24). Fix a $y \in \mathcal{Y}$. We first express $\Pr_{H_1}[Y=y]$ and $\Pr_{H_{j_0}}[Y=y]$ in terms of $\widetilde{u}$ and $\widetilde{q}$:

1. $\Pr_{H_1}[Y=y] = \left(\frac{1}{2}+4\epsilon\right)^{\widetilde{u}_1^1(y)}\left(\frac{1}{2}-4\epsilon\right)^{\widetilde{u}_0^1(y)}\prod_{j=2}^{t}\left(\frac{1}{2}\right)^{\widetilde{q}_j(y)}$;
2. $\Pr_{H_{j_0}}[Y=y] = \left(\frac{1}{2}+4\epsilon\right)^{\widetilde{u}_1^1(y)}\left(\frac{1}{2}-4\epsilon\right)^{\widetilde{u}_0^1(y)}(1-16\epsilon)^{\widetilde{u}_0^{j_0}(y)}(1+16\epsilon)^{\widetilde{u}_1^{j_0}(y)}\prod_{j=2}^{t}\left(\frac{1}{2}\right)^{\widetilde{q}_j(y)}$.

Taking the ratio, we obtain that

$$\begin{aligned}\frac{\Pr_{H_{j_0}}[Y=y]}{\Pr_{H_1}[Y=y]} &= (1-16\epsilon)^{\widetilde{u}_0^{j_0}(y)}(1+16\epsilon)^{\widetilde{u}_1^{j_0}(y)} = (1-16\epsilon)^{\frac{\widetilde{q}_{j_0}(y)}{2}+\frac{\widetilde{u}_0^{j_0}(y)-\widetilde{u}_1^{j_0}(y)}{2}}(1+16\epsilon)^{\frac{\widetilde{q}_{j_0}(y)}{2}-\frac{\widetilde{u}_0^{j_0}(y)-\widetilde{u}_1^{j_0}(y)}{2}} \\ &= (1-256\epsilon^2)^{\frac{\widetilde{q}_{j_0}(y)}{2}}\left(\frac{1-16\epsilon}{1+16\epsilon}\right)^{\frac{\widetilde{u}_0^{j_0}(y)-\widetilde{u}_1^{j_0}(y)}{2}} \geq (1-256\epsilon^2)^{\frac{\widetilde{q}_{j_0}(y)}{2}}(1-32\epsilon)^{\left|\frac{\widetilde{u}_0^{j_0}(y)-\widetilde{u}_1^{j_0}(y)}{2}\right|}.\end{aligned} \tag{25}$$

When both $\widetilde{q}_{j_0}(y) \leq Q_0$ and $|u_0^{j_0}(y) - u_1^{j_0}(y)| \leq \sqrt{5Q_0}$ hold, we have (recall that $Q_0 \leq \frac{\ln(1/\delta)}{1250\epsilon^2}$)

$$\begin{aligned}(1-256\epsilon^2)^{\frac{\widetilde{q}_{j_0}(y)}{2}}(1-32\epsilon)^{\left|\frac{\widetilde{u}_0^{j_0}(y)-\widetilde{u}_1^{j_0}(y)}{2}\right|} &\geq (1-256\epsilon^2)^{Q_0/2}(1-32\epsilon)^{\sqrt{5Q_0}/2} \\ &\geq (1-256\epsilon^2)^{\frac{\ln(1/\delta)}{2500\epsilon^2}}(1-32\epsilon)^{\frac{\sqrt{\ln(1/\delta)}}{30\epsilon}} \geq \delta^{1/4}\cdot\delta^{1/4} = \sqrt{\delta},\end{aligned}$$

where in the penultimate inequality we used the assumption that $0 < \delta \leq 0.01$. This proves (24). $\square$

Theorem 5.5 *Fix real numbers δ, ϵ such that $0 < \delta, \epsilon \le 0.01$, and integers K, n, where $K \le n/2$. Let $\mathcal{A}$ be a (possibly randomized) algorithm so that for any set of n arms with the mean $\theta_1, \theta_2, \ldots, \theta_n$,*

- *$\mathcal{A}$ makes at most Q samples in expectation;*
- *With probability at least $1-\delta$, $\mathcal{A}$ outputs a set T of size K with $\mathsf{val}_K(T) \ge \mathsf{val}_K([n]) - \epsilon$.*

We have that $Q = \Omega\big(\frac{n \ln(1/\delta)}{\epsilon^2 K}\big)$.

Proof of Theorem 5.5. We show that essentially the same lower bound also holds for any randomized algorithm. The following argument is standard and we include it for completeness. Fix $0 < \epsilon, \delta < 1/2$. We assume, for contradiction, that there is a randomized algorithm $\mathcal{A}$ which can achieve the same performance guarantee stated as in the theorem, but the expected number Q of samples is no more than $\frac{n \log(1/\delta)}{100000\epsilon^2 K}$. We can view the randomized algorithm $\mathcal{A}$ as a deterministic algorithm with a sequence S of random bits. We use R to denote the randomness from the arms. Note that if we fix S and R, the execution and the output of the algorithm are fixed. We use $\mathcal{A}(S,R) = 1$ to denote the event that the output of $\mathcal{A}$ is an ϵ-optimal solution. Let us use $Q(S,R)$ to denote the number of samples taken by $\mathcal{A}$. The performance guarantee of $\mathcal{A}$ is that

$$\Pr_{S,R}[\mathcal{A}(S,R)=1] = \mathbf{E}_{S,R}[\mathcal{A}(S,R)] = \mathbf{E}_S\, \mathbf{E}_R[\mathcal{A}(S,R) \mid S] \ge 1-\delta.$$

This is equivalent to say that $\mathbf{E}_S\, \mathbf{E}_R[1-\mathcal{A}(S,R) \mid S] \le \delta$. By Markov inequality, we have that $\Pr_S\Big[\mathbf{E}_R[1-\mathcal{A}(S,R) \mid S] \ge 2\delta\Big] \le 1/2$. Equivalently, we have that

$$\Pr_S\Big[\mathbf{E}_R[\mathcal{A}(S,R) \mid S] \ge 1-2\delta\Big] \ge 1/2. \tag{26}$$

By our assumption, we have $\mathbf{E}_{S,R}\, Q(S,R) \le \frac{n\log(1/\delta)}{100000\epsilon^2 K}$. So, by Markov inequality,

$$\Pr_S\Big[\mathbf{E}_R[Q(S,R) \mid S] \le \frac{n\log(1/\delta)}{40000\epsilon^2 K}\Big] \ge \frac{3}{5}. \tag{27}$$

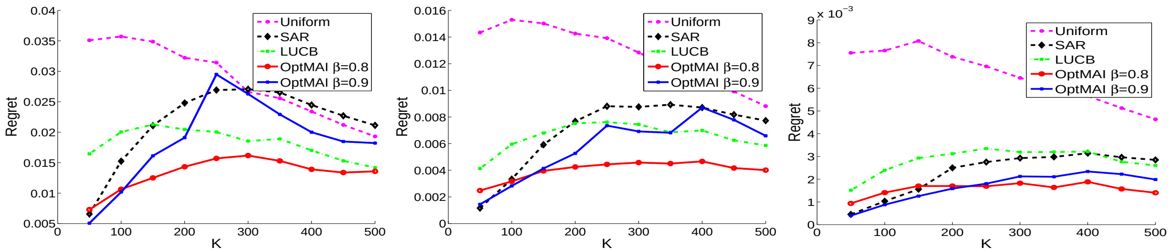
Combining (26) and (27), we know there is a particular random sequence S such that both $\mathbf{E}_R[\mathcal{A}(S,R) \mid S] \ge 1-2\delta$ and $\mathbf{E}_R[Q(S,R) \mid S] \le \frac{n\log(1/\delta)}{40000\epsilon^2 K}$ hold. Since the algorithm $\mathcal{A}$ with a particular sequence S is simply a deterministic algorithm, this contradicts the lower bound we proved for any deterministic algorithm in Lemma 5.4. ∎

Theorem F.1. *Fix real numbers δ, ϵ such that $0 < \delta, \epsilon \le 0.01$, and integers K, n, where $n/2 \le K < n$. Let $\mathcal{A}$ be a (possibly randomized) algorithm such that for any multiple arm identification instance, $\mathcal{A}$ can outputs an ϵ-optimal set T of size K, with probability at least $1-\delta$, using at most Q samples in expectation. We have that*

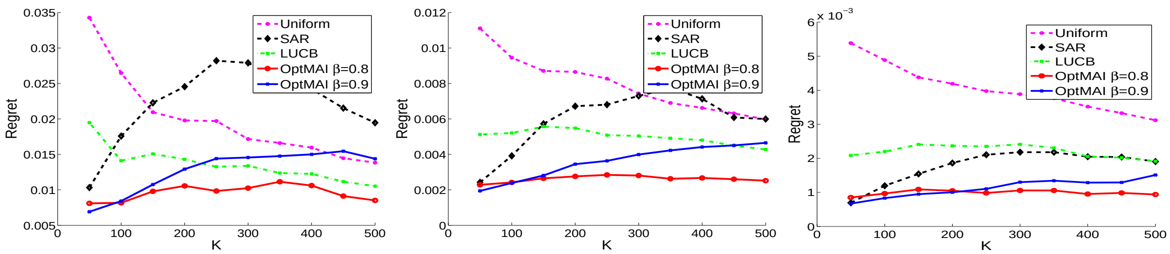
$$Q = \Omega\Big(\frac{n-K}{K} \cdot \frac{n}{\epsilon^2}\Big)\Big(\frac{n-K}{K} + \frac{\ln 1/\delta}{K}\Big).$$

.

Proof. In fact, in the proof of Theorem 4.5, we have established the equivalence between the problem of find an ϵ-optimal solution of size K and an ϵ'-optimal solution of size $n-K$, where $\epsilon' = \frac{K}{n-K} \cdot \epsilon$. Since $n-K \le n/2$, we can use the lower bounds developed in Theorem 5.1 and Theorem 5.5, which show that Q should be at least $Q = \Omega\left(\frac{n}{\epsilon'^2}\left(1 + \frac{\ln(1/\delta)}{n-K}\right)\right)$. Plugging in $\epsilon' = \frac{K}{n-K} \cdot \epsilon$, we obtain the desired lower bound. ∎



(a) $\theta \sim \mathrm{TN}(0.5, 0.2)$, $Q = 20 \cdot n$ (b) $\theta \sim \mathrm{TN}(0.5, 0.2)$, $Q = 50 \cdot n$ (c) $\theta \sim \mathrm{TN}(0.5, 0.2)$, $Q = 100 \cdot n$



(d) $\theta \sim \mathrm{Beta}(4, 1)$, $Q = 20 \cdot n$ (e) $\theta \sim \mathrm{Beta}(4, 1)$, $Q = 50 \cdot n$ (f) $\theta \sim \mathrm{Beta}(4, 1)$, $Q = 100 \cdot n$

Figure 3: Performance comparison on simulated data.

G Additional Experiments

In this section, we provide additional simulated experimental results when using the following two different ways to generate $\{\theta_i\}_{i=1}^n$:

1. $\theta \sim \mathrm{TN}(0.5, 0.2)$: each θ_i is generated from a truncated normal distribution with mean 0.5, the standard deviation 0.2 and the support $[0, 1]$ (Figure 3(a) to Figure 3(c)).

2. $\theta \sim \mathrm{Beta}(4, 1)$: each θ_i is generated from a Beta distribution with the parameters $(4, 1)$. The $\{\theta_i\}$ from $\mathrm{Beta}(4, 1)$ are close to the workers' accuracy in real crowdsourcing applications, where most workers perform reasonably well and the averaged accuracy is around 80% (Figure 3(d) to Figure 3(f)).

We note that the number of total arms is set to $n = 1000$. We vary the total budget $Q = 20n, 50n, 100n$ and $K = 10, 20, \ldots, 500$. We use different ways to generate $\{\theta_i\}_{i=1}^n$ and report the comparison among different algorithms. It can be seen from Figure 1 that our method outperforms the SAR and LUCB in most of the scenarios. In addition, we also observe that when K is large, the setting of $\beta = 0.8$ outperforms that of $\beta = 0.9$; while for small K, $\beta = 0.9$ is a better choice.