Journal of Machine Learning Research 1 (2018) 1-48 Submitted 4/00; Published 10/00

Best Arm Identification for Contaminated Bandits

Jason Altschuler JASONALT@MIT.EDU
Victor-Emmanuel Brunel VEBRUNEL@MIT.EDU
Alan Malek AMALEK@MIT.EDU
Massachusetts Institute of Technology
77 Massachusetts Ave, Cambridge, MA 02139

Editor: ??

Abstract

This paper studies active learning in the context of robust statistics. Specifically, we propose a variant of the Best Arm Identification problem for *contaminated bandits*, where each arm pull has probability ε of generating a sample from an arbitrary contamination distribution instead of the true underlying distribution. The goal is to identify the best (or approximately best) true distribution with high probability, with a secondary goal of providing guarantees on the quality of this distribution. The primary challenge of the contaminated bandit setting is that the true distributions are only partially identifiable, even with infinite samples. We first present tight, non-asymptotic sample complexity bounds for high-probability estimation of the first two robust moments (median and median absolute deviation) from contaminated samples, which may be of independent interest. Building on these results, we adapt several classical Best Arm Identification algorithms to the contaminated bandit setting and derive sample complexity upper bounds for our problem. Finally, we provide matching information-theoretic lower bounds on the sample complexity (up to a small logarithmic factor). Our results suggest an inherent robustness of classical Best Arm Identification algorithms.

Keywords: multi-armed bandits, best arm identification, robust statistics, contamination model, partial identifiability

1. Introduction

Consider Pat, an aspiring machine learning researcher deciding between working in a statistics, mathematics, or computer science department. Pat's sole criterion is salary, so Pat surveys current academics with the goal of finding the department with highest median income. However, some subset of the data will be inaccurate: some respondents obscure their salaries for privacy reasons, some convert currency incorrectly, and some do not read the question and report yearly instead of monthly salary, etc. How should Pat target his or her surveys, in an adaptive (online) fashion, to find the highest paying department with high probability in the presence of contaminated data?

In this paper, we study the Best Arm Identification (BAI) problem for multi-armed bandits where observed rewards are not completely trustworthy. The multi-armed bandit problem has received extensive study in the last three decades, see e.g. (Lai and Robbins, 1985; Bubeck and Cesa-Bianchi, 2012). We study the *fixed confidence*, or (α, δ)-PAC, BAI problem, in which the learner must identify an α-suboptimal arm with probability at least $1 - \delta$, using as few samples as possible. Most BAI algorithms for the fixed-confidence

setting assume i.i.d. rewards from distributions with relatively strict control on the tails, such as boundedness or more generally sub-Gaussianity (Jamieson et al., 2014). However, for Pat, the data neither are i.i.d. nor are from a distribution with controllable tails. How can we model this data, and how can we optimally explore these arms?

To answer the first question, we turn to robust statistics, which has studied such questions for over fifty years. In a seminal paper (Huber, 1964), Huber introduced the contamination model, which we adapt to the multi-armed bandit model by proposing the *Contaminated Best Arm Identification problem* (CBAI). This is formally defined in Section 2, but is informally described as follows. There are $k \geq 2$ arms, each endowed with a fixed base distribution F_i and arbitrary contamination distributions $G_{i,t}$ for $t \geq 1$. We place absolutely no assumptions on $G_{i,t}$. When arm i is pulled in round t, the learner receives a sample that with probability $1 - \varepsilon$ is drawn from the arm's true distribution F_i, and with the remaining probability ε is drawn from an arbitrary contamination distribution $G_{i,t}$. A key point is that suboptimality of the arms is based on the quality of the underlying true distributions F_i, not the contaminated distributions $\tilde{F}_{i,t}$ of the observed samples. Note that existing BAI algorithms, fed with samples from $\tilde{F}_{i,t}$, will not necessarily work.

This contamination model nicely fits Pat's problem: samples are usually trustworthy, but sometimes they are completely off and cannot be modeled by a distribution with controlled tails. Additionally, the nature of contamination changes with the respondent, and hence $G_{i,t}$ should be considered as time varying, which completely breaks the usual i.i.d. data assumption. Finally, Pat wants to determine the department with highest true median salary, not highest contaminated median salary. The contaminated bandit setup also naturally models many other situations that the classical bandit setup cannot, such as:

- measuring drug responses where samples can be corrupted or test results incorrectly recorded;
- testing new software features, where yet-unfixed bugs may distort responses but will be fixed before release;
- online advertisement, where a small portion of users respond differently to novel ads; and
- conducting surveys with randomized responses, where a subject's response is randomly negated in order to preserve privacy (e.g. surveys for sensitive issues (Warner, 1965)).

Importantly, note that the CBAI problem is nontrivially harder than the BAI problem since there are no consistent estimators for statistics (e.g. the mean or median) of F_i if the contaminations are allowed to be arbitrary. Under some mild technical assumptions on F_i, the contamination can cause the median of $\tilde{F}_{i,t}$ to be anywhere in an $\Theta(\varepsilon)$-neighborhood of the median of F_i, and hence we can only determine the median of F_i up to some unavoidable estimation bias, U_i, of order ε (see Section 2 for details). This leads us to later generalize our study to the more abstract *Partially Identifiable Best Arm Identification* (PIBAI) problem (defined formally in Section 4), which includes CBAI as a special case.

This PIBAI problem can be seen as an active-learning version of the classical problem of estimation under partial identifiability, which has been studied for most of the last century in the econometrics literature, see e.g. (Marschak and Andrews, 1944; Manski, 2009). A canonical example in this field is trying to learn the age distribution of a population by only

asking in which decade each subject was born; clearly the median age can only be learned up to certain unidentifiability regions.

1.1 Our Contributions and Outline

To the best of our knowledge, this is the first paper to consider the Best Arm Identification problem with arbitrary contaminations. This contaminated bandit setup models many practical problems that the classical bandit setup cannot. We begin in Section 2 by formally defining the CBAI problem and then discussing its primary challenge: partial identifiability of the arms' true distributions. Indeed, we show that the adversary's ability to inject arbitrary contaminations can render "similar" underlying distributions indistinguishable, even with access to infinite samples.

Also in Section 2, we describe the three models of the adversary's power we consider: the *oblivious* adversary chooses all contamination distributions a priori; the *prescient* adversary may choose the contamination distributions as a function of all realizations (past and future) of the true rewards and knowledge of when the learner will observe contaminated samples; and the *malicious* adversary may, in addition, correlate when the learner observes contaminated samples with the true rewards. (See Section 2.1 for formal definitions.)

Our technical contributions can be divided into three parts:

(i) we prove tight, non-asymptotic sample complexity bounds for estimation of the first two robust moments (median and median absolute deviation) from contaminated samples,

(ii) we develop efficient algorithms for the CBAI problem using these guarantees and provide sample complexity upper bounds for the fixed-confidence setting, and

(iii) we prove matching information-theoretic lower bounds showing that our algorithms have optimal sample complexity (up to small logarithmic factors).

Section 3 presents contribution (i), which may be of independent interest in the robust statistics community. We consider estimating statistics of a single arm from contaminated samples and show that although estimation of standard moments (mean, variance) is impossible, estimation of robust moments (median, median absolute deviation) is possible. Specifically, for each of the three adversarial models, we show that with probability at least $1-\delta$, the empirical median of the contaminated samples lies in a region around the true median of width $U_i + E_{n,\delta}$, where U_i is some unavoidable bias that depends on quantiles of F_i and the power of the adversary, n is the number of samples, and $E_{n,\delta}$ is a confidence-interval term that decreases at the optimal $\sqrt{\frac{\log 1/\delta}{n}}$ rate. Our results neatly capture the effect of the adversary's power by deriving different U_i for each scenario, thereby precisely quantifying the hardness of the three different adversarial settings. We then present non-asymptotic sample complexity guarantees for estimation of the second robust moment, often called the Median Absolute Deviation (MAD), under all three adversarial settings. The MAD is a robust measure of the spread of F_i and controls the width U_i of the median's unidentifiability region.

Section 4 presents contribution (ii). We show that, surprisingly, several classical BAI algorithms are readily adaptable to the PIBAI problem. This suggests a certain inher-

ent robustness of these classical bandit algorithms. We then combine these results with our estimation guarantees from (i) to obtain PAC algorithms for CBAI. We give fixed-confidence sample complexity guarantees that mirror the sample complexity guarantees for BAI in the classical stochastic multi-armed bandit setup. The main difference is that BAI sample complexities depend on the suboptimality "gaps" $\Delta_i := p_{i^*} - p_i$ between the statistics of the optimal arm i^* and each suboptimal arm i, whereas our CBAI sample complexities depend on the suboptimality "*effective gaps*" $\tilde{\Delta}_i := (p_{i^*} - U_{i^*}) - (p_i + U_i) = \Delta_i - (U_{i^*} + U_i)$, which account for the unavoidable estimation uncertainties in the most pessimistic way. We also show how to apply the MAD estimations results from (i) to obtain guarantees on the quality of the underlying distribution of the selected arm.

Section 5 presents contribution (iii). We prove matching information-theoretic lower bounds (up to a small logarithmic factor) on the sample complexity of CBAI via a reduction to classical lower bounds for the stochastic multi-armed bandit problem. We argue that for CBAI the effective gap $\tilde{\Delta}_i$ is the right analog of the traditional gap since it appears in matching ways in both our upper and lower bounds.

1.2 Previous Work

The Best Arm Identification problem in the fixed-confidence setting has a long history, see e.g. (Bechhofer et al., 1968; Lai and Robbins, 1985). Recent interest in the learning theory community was sparked by the seminal paper by (Even-Dar et al., 2002), which proposed two algorithms we study: Successive Elimination and Median Elimination. Since then, there has been significant work on the algorithmic side (see Kalyanakrishnan et al., 2012; Gabillon et al., 2012; Karnin et al., 2013; Jamieson et al., 2014). Concurrently, a parallel line of work has focused on improving lower bounds, starting with the 2-armed setting (Chernoff, 1972; Anthony and Bartlett, 2009), extending to the multi-armed setting (Mannor and Tsitsiklis, 2004), and, more recently, continuing with more finely tuned lower bounds that include properties of the arm distributions aside from the gaps (Chen and Li, 2015; Kaufmann et al., 2016; Garivier and Kaufmann, 2016).

In the cumulative regret setting, the online learning literature has considered both stochastic bandits with mild tail assumptions (for example (Bubeck et al., 2013) only assumed the existence of a $(1+\varepsilon)$ moment) and algorithms that obtain near-optimal regret guarantees if the environment is stochastic or adversarial (Bubeck and Cesa-Bianchi, 2012; Seldin and Slivkins, 2014). The partial monitoring problem is also loosely similar to the CBAI problem in the sense that both problems feature partial identification; there, the learner only knows the loss up to some subset (Bartók et al., 2014).

The existing literature closest to our work studies the BAI problem in settings more general than i.i.d. arms, for example stochastic but non-stationary distributions (Allesiardo et al., 2017; Allesiardo and Féraud, 2017) or arbitrary rewards where each arm converges to a limit (Jamieson and Talwalkar, 2016; Li et al., 2016). However, neither setting fits the contamination model or allows for arbitrary perturbations.

This paper also makes connections between several long bodies of work. The contamination model (Huber, 1964) has a long history of more than fifty years in robust statistics, see e.g. (Hampel, 1974; Maronna and Yohai, 1976; Rousseeuw and Leroy, 2005; Hampel et al., 2011). Contamination models and malicious errors have also been studied in the

computer science community; see, for example, the classical papers of (Valiant, 1985; Kearns and Li, 1993) and a recent burst of results on algorithms that handle estimation of means and variances (Lai et al., 2016), efficient estimation in high dimensions (Diakonikolas et al., 2018), PCA (Cherapanamjeri et al., 2017), and general learning (Charikar et al., 2017) in the presence of outliers or corrupted data. Finally, the partial identification literature from econometrics also has a rich history (Marschak and Andrews, 1944; Horowitz and Manski, 1995; Manski, 2009; Romano and Shaikh, 2010; Bontemps et al., 2012).

1.3 Notation

Let F be a distribution. We denote its left and right quantiles, respectively, by $Q_{L,F}(p) := \inf\{x \in \mathbb{R} : F(x) \geq p\}$ and $Q_{R,F}(p) := \inf\{x \in \mathbb{R} : F(x) > p\}$. The left and right medians of F, respectively, are then defined as $m_{1,L}(F) := Q_{L,F}(\frac{1}{2})$ and $m_{1,R}(F) := Q_{R,F}(\frac{1}{2})$. We denote the set of medians of F by $m_1(F) := [m_{1,L}(F), m_{1,R}(F)]$. When F has unique median, we overload $m_1(F)$ to be just this point rather than a singleton set containing it. When F has a unique median, we denote the median absolute deviation (MAD) of F by $m_2(F) := m_1(|X - m_1(F)|)$ where $X \sim F$. When $m_2(F)$ is unique, we further define $m_4(F) := m_1(|X - m_1(F)| - m_2(F)|)$. Note that $m_1(F)$, $m_2(F)$, and $m_4(F)$ are robust analogues of centered first (mean), second (variance), and fourth (kurtosis) moments, respectively. For clarity, we use hats throughout whenever we discuss an empirical median $\hat{m}_1(\cdot)$ or empirical MAD $\hat{m}_2(\cdot)$. When F has a unique median, we denote by H_F the "folded distribution of F", i.e., the distribution of $|Y - m_1(F)|$ where $Y \sim F$.

We denote the Dirac measure at a point $x \in \mathbb{R}$ by δ_x, the Bernoulli distribution with parameter $p \in [0, 1]$ by $\mathrm{Ber}(p)$, and the uniform distribution over an interval $[a, b]$ by $\mathrm{Unif}([a, b])$. For $\varepsilon \in [0, 1]$ and distributions F and G, we denote by $(1-\varepsilon)F+\varepsilon G$ the mixture model under which variables have law $(1 - D)Y + DZ$, where $D \sim \mathrm{Ber}(\varepsilon)$, $Y \sim F$, and $Z \sim G$ are all independent.

The interval $[a - b, a + b]$ is denoted by $[a\pm b]$, the set of non-negative real numbers by $\mathbb{R}_{\geq 0}$, the set of positive integers by $\mathbb{N}$, and the set $\{1, \ldots, n\}$ by $[n]$ for $n \in \mathbb{N}$. We denote the k-th order statistic of a (possibly random) real-valued sequence $x_1, \ldots, x_n \in \mathbb{R}$ by $x_{(k)}$, and the median of this sequence by $x_{(\mathrm{med})}$: if n is odd, this is the middle value; and if n is even, it is the average of the middle two values. We abbreviate "with high probability" by "w.h.p." and "cumulative distribution function" by "cdf".

2. Best Arm Identification for Contaminated Bandits

Here we formally define the *Contaminated Best Arm Identification problem (CBAI)*. Let $k \geq 2$ be the number of arms, $\varepsilon \in (0, \frac{1}{2})$ be the contamination level, $\{F_i\}_{i\in[k]}$ be the true but unknown distributions, and $\{G_{i,t}\}_{i\in[k],t\in\mathbb{N}}$ be arbitrary contamination distributions. This induces contaminated distributions $\tilde{F}_{i,t}$, samples from which have laws $(1-D_{i,t})Y_{i,t}+D_{i,t}Z_{i,t}$, where $D_{i,t} \sim \mathrm{Ber}(\varepsilon)$, $Y_{i,t} \sim F_i$, $Z_{i,t} \sim G_{i,t}$. Note that if all of these random variables are independent, then each $\tilde{F}_{i,t}$ is simply equal to the contaminated mixture model $(1-\varepsilon)F_i+\varepsilon G_{i,t}$. However, we generalize by also considering the setting when the $Y_{i,t}, D_{i,t}, Z_{i,t}$ are not all independent. This allows an adversary to further obfuscate samples by adapting the

distributions of the $D_{i,t}$ and $Z_{i,t}$ based on the realizations of the $Y_{i,t}$ (that is, by coupling these random variables); see below for details.

At each iteration t, a CBAI algorithm chooses an arm $I_t \in [k]$ to pull and receives a sample $X_{I_t,t}$ distributed according to the corresponding contaminated distribution $\tilde{F}_{I_t,t}$. After T iterations (a possibly random stopping time that the algorithm may choose), the algorithm outputs an arm $\hat{I} \in [k]$. For $\alpha \geq 0$ and $\delta \in (0,1)$, the algorithm is said to be (α, δ)-*PAC* if with probability at least $1-\delta$, $\hat{I}$ has median within $\alpha + U$ of the optimal:

$$\mathbb{P}\left(m_1(F_{\hat{I}}) \geq \max_{i \in [k]} m_1(F_i) - (\alpha + U) \right) \geq 1 - \delta, \tag{1}$$

where U is the unavoidable uncertainty term in estimation that is induced by partial identifiability (see Section 2.2 for a discussion of U, and see Section 3 for an explicit computation of this quantity). Thus, the goal is to find an algorithm achieving the PAC-guarantee in (1) with small sample complexity T, either in expectation or with high probability.

2.1 Power of the Adversary

As is typical in online learning problems, it is important to define the power of the adversary since this affects the complexity of the resulting problem. Interestingly, CBAI is still possible even when we grant the adversary significant power. (Although of course, the adversary's power is reflected in the corresponding rates.) We consider three settings, presented in increasing order of adversarial power. The key differences between these different types of adversaries are twofold: (1) whether they can choose the contaminated distributions "presciently" based on all other realizations $\{Y_{i,t}, D_{i,t}\}_{i \in [k], t \geq 1}$ both past and future; and (2) whether they can "maliciously" couple the distributions of each $D_{i,t}$ with the corresponding $Y_{i,t}$, subject only to the constraint that the marginals $D_{i,t} \sim \mathrm{Ber}(\varepsilon)$ and $Y_{i,t} \sim F_i$ stay correct.

- **Oblivious adversary.** For all $i \in [k]$, the triples $\{(Y_{i,t}, D_{i,t}, Z_{i,t})\}_{t \geq 1}$ are independent, and for all $t \geq 1$, $Y_{i,t} \sim F_i$, $D_{i,t} \sim \mathrm{Ber}(\varepsilon)$, and $Y_{i,t}$ and $D_{i,t}$ are independent.

- **Prescient adversary.** For all $i \in [k]$, the pairs $\{(Y_{i,t}, D_{i,t})\}_{t \geq 1}$ are independent, and for all $t \geq 1$, $Y_{i,t} \sim F_i$, $D_{i,t} \sim \mathrm{Ber}(\varepsilon)$, $Y_{i,t}$ and $D_{i,t}$ are independent, and $Z_{i,t}$ may depend on all $\{Y_{i',t'}, D_{i',t'}, Z_{i',t'}\}_{i' \in [k], t' \geq 1}$.

- **Malicious adversary.** For all $i \in [k]$, the pairs $\{(Y_{i,t}, D_{i,t})\}_{t \geq 1}$ are independent, and for all $t \geq 1$, $Y_{i,t} \sim F_i$, $D_{i,t} \sim \mathrm{Ber}(\varepsilon)$, and $Z_{i,t}$ may depend on all $\{Y_{i',t'}, D_{i',t'}, Z_{i',t'}\}_{i' \in [k], t' \geq 1}$.

Perhaps surprisingly, we will show that the sample complexity is the same for both oblivious and prescient adversaries. Indeed for these two settings, we will prove our upper bounds for the (more powerful) setting of prescient adversaries, and our lower bounds for the (less powerful) setting of oblivious adversaries. We also note that the rate for malicious adversaries is only worse by at most a (quantile) "factor of 2"; see Sections 3.1 and 4.3.1 for a precise statement.

2.2 Challenge of Partial Identifiability

In the introduction, we emphasized the point that a contaminating adversary can render different underlying distributions of an arm statistically indistinguishable. That is, even with infinite samples, it is impossible to estimate statistics of an arm's true distribution exactly. Consider, for example, the problem of estimating the median of a single arm with true distribution F against an oblivious adversary, which is the weakest of our three adversarial settings. Define S to be the set of all distributions F' for which there exists adversarially chosen distributions G and G' such that $(1-\varepsilon)F + \varepsilon G = (1-\varepsilon)F' + \varepsilon G'$. How large is this set S, and in particular, how far can the medians of distributions in S be from the median of F?

The following simple example shows that S is non-trivial (and thus in particular contains more than just F). Let F be the uniform distribution on the interval $[-1, 1]$, and let G be the uniform distribution on $[-1-c, -1] \cup [1, 1+c]$, where $c = \varepsilon(1-\varepsilon)^{-1}$. The contaminated distribution $\tilde{F} := (1-\varepsilon)F + \varepsilon G$ is the uniform distribution on $[-1-c, 1+c]$. However, for any $p \in [-c, c]$, $\tilde{F}$ is also equal to $(1-\varepsilon)F(\cdot - p) + \varepsilon G_p$, where G_p is the uniform distribution over $[-1-c, 1+c] \setminus [-1+p, 1+p]$. We conclude that F is statistically indistinguishable from any of the translations $\{F(\cdot - p)\}_{p\in[-c,c]}$. Hence, S is non-trivial and contains distributions with medians at least $O(c)$ away from $m_1(F)$. Even in this simple setting, infinite samples only allow us to identify the median of F at most up to the unidentifiability region $[-c, c]$.

In fact, this toy example captures the correct dependence on ε of how far the median of the contaminated distribution can be shifted, as formalized by the following simple lemma.

Lemma 1 *For any $\varepsilon \in (0, \frac{1}{2})$, any distribution F, and any median $m_1 \in m_1(F)$,*

$$\sup_{\substack{\text{distribution } G,\\ \tilde{m}_1 \in m_1((1-\varepsilon)F+\varepsilon G)}} |\tilde{m}_1 - m_1| = \max\left\{ Q_{R,F}\left(\frac{1}{2(1-\varepsilon)}\right) - m_1, m_1 - Q_{L,F}\left(\frac{1-2\varepsilon}{2(1-\varepsilon)}\right)\right\}.$$

Proof The fact that the left hand side is no smaller than the right hand side is straightforward: to shift the median to the right (resp. left), let $\{G_n\}_{n\in\mathbb{N}}$ be a sequence of distributions which are Dirac measures at n (resp. $-n$).

We now prove the reverse direction, the "$\leq$" inequality for any fixed distribution G. For shorthand, denote the contaminated distribution by $\tilde{F} := (1-\varepsilon)F + \varepsilon G$, and denote its left and right medians by $\tilde{m}_{1,L}$ and $\tilde{m}_{1,R}$, respectively. Since every median of $\tilde{F}$ lies within $[\tilde{m}_{1,L}, \tilde{m}_{1,R}]$, it suffices to show that $\tilde{m}_{1,L} \geq Q_{L,F}(\frac{1-2\varepsilon}{2(1-\varepsilon)})$ and $\tilde{m}_{1,R} \leq Q_{R,F}(\frac{1}{2(1-\varepsilon)})$. We bound $\tilde{m}_{1,L}$ presently, as bounding $\tilde{m}_{1,R}$ follows by a similar argument or by simply applying the $\tilde{m}_{1,L}$ bound to $F(-\cdot)$. By definition of $\tilde{m}_{1,L}$, for all $t > 0$, $\frac{1}{2} \leq \tilde{F}(\tilde{m}_{1,L} + t) = (1-\varepsilon)F(\tilde{m}_{1,L} + t) + \varepsilon G(\tilde{m}_{1,L} + t) \leq (1-\varepsilon)F(\tilde{m}_{1,L} + t) + \varepsilon$, where the last step is because $G(\tilde{m}_{1,L} + t) \leq 1$ since G is a distribution. Rearranging yields $F(\tilde{m}_{1,L} + t) \geq \frac{1-2\varepsilon}{2(1-\varepsilon)}$, which implies that $\tilde{m}_{1,L} \geq Q_{L,F}(\frac{1-2\varepsilon}{2(1-\varepsilon)})$. ■

3. Estimation From Contaminated Samples

The core difficulty of the contaminated bandit setup is that statistics of the true distributions are only partial identifiable (see discussion in Section 2.2 above). Therefore in order to

obtain algorithms for the CBAI problem, we must first approach the problem of estimation from contaminated samples. This is the goal of this section. Throughout the section, we will only consider a single arm, and its true distribution will be denoted by F.

Lemma 1 implies that some control of the quantiles of F is necessary to obtain guarantees for estimation of $m_1(F)$ from contaminated samples. We begin by considering F in the following family of distributions, which we stress is the minimum possible requirement needed for median estimation.

Definition 2 *For any $\bar{t} \in (0, \frac{1}{2})$ and any non-decreasing function $R : [0, \bar{t}] \to \mathbb{R}_{\geq 0}$, define $\mathcal{H}_{\bar{t},R}$ to be the family of all distributions F satisfying*

$$R(t) \geq \max\left\{Q_{R,F}\left(\frac{1}{2}+t\right) - m_1, m_1 - Q_{L,F}\left(\frac{1}{2}-t\right)\right\} \tag{2}$$

for any $t \in [0, \bar{t}]$ and any median $m_1 \in m_1(F)$.

The class $\mathcal{H}_{\bar{t},R}$ can only handle contamination levels $\varepsilon \leq \bar{\varepsilon}(\bar{t}) := \frac{2\bar{t}}{1+2\bar{t}}$; this requirement is equivalent to $\bar{t} \geq \frac{\varepsilon}{2(1-\varepsilon)}$, which, by Lemma 1, is the largest possible deviation from the $\frac{1}{2}$-quantile.

Combining Lemma 1 and Definition 2 produces a tight bound on how far the contaminated median can be moved from the true median for any $F \in \mathcal{H}_{\bar{t},R}$.

Corollary 3 *For any $\bar{t} \in (0, \frac{1}{2})$, any $\varepsilon \in (0, \bar{\varepsilon}(\bar{t}))$, and any non-decreasing function $R : [0, \bar{t}] \to \mathbb{R}_{\geq 0}$,*

$$\sup_{\substack{F \in \mathcal{H}_{\bar{t},R}, \\ \text{distribution } G, \\ \tilde{m}_1 \in m_1((1-\varepsilon)F + \varepsilon G)}} |\tilde{m}_1 - m_1(F)| = R\left(\frac{\varepsilon}{2(1-\varepsilon)}\right)$$

The rest of the section is organized as follows. Section 3.1 studies the general family of distributions in $\mathcal{H}_{\bar{t},R}$ and provides, for all three adversarial settings, tight upper and lower bounds on the sample complexity of median estimation from contaminated samples. Next, in Section 3.2 we obtain more algorithmically useful guarantees by specializing these results to a family of distributions where the cdfs increase at least linearly in a neighborhood of the median (this is essentially $\mathcal{H}_{\bar{t},R}$ for R linear in a small neighborhood around 0), a very common assumption in the robust estimation literature. For this family of distributions, we explicitly compute for all three adversarial settings the unavoidable bias terms mentioned in Section 2 for median estimation from contaminated samples. It is worth noting that these results precisely quantify the effect of the three different adversarial strengths on the complexity of the problem of median estimation in the contamination model. Finally, in Section 3.3 we provide finite-sample guarantees for estimation of the the second robust moment, the MAD, from contaminated samples.

3.1 Median Estimation Results for $\mathcal{H}_{\bar{t},R}$

3.1.1 Oblivious and Prescient Adversaries

We begin with guarantees for oblivious and prescient adversaries. It turns out that the rates are the same for both settings. In particular, we show that with probability at least $1-\delta$, the

estimation error is bounded above by $R(\frac{\varepsilon}{2(1-\varepsilon)} + O(\sqrt{\frac{\log 1/\delta}{n}}))$. Note that if R is Lipschitz, then this quantity is bounded above by the unavoidable uncertainty term $R(\frac{\varepsilon}{2(1-\varepsilon)})$ (see Corollary 3) plus an error term that decays quickly with $n^{-1/2}$ rate and sub-Gaussian tails. We remark that this latter confidence-interval term is of optimal order, since it is tight even for estimating the mean (and thus median) of a Gaussian random variable with known unit variance from *uncontaminated* samples.

Lemma 4 *Let $\bar{t} \in (0, \frac{1}{2})$, $\varepsilon \in (0, \bar{\varepsilon}(\bar{t}))$, and $F \in \mathcal{H}_{\bar{t},R}$. Let $Y_i \sim F$ and $D_i \sim Ber(\varepsilon)$, for $i \in [n]$, all be drawn independently. Let $\{Z_i\}_{i\in[n]}$ be arbitrary random variables possibly depending on $\{Y_i, D_i\}_{i\in[n]}$, and define $X_i = (1-D_i)Y_i + D_i Z_i$. Then, for any confidence level $\delta \in (0,1)$ and sample complexity $n \geq 2\left(\bar{t} - \frac{\varepsilon}{2(1-\varepsilon)}\right)^{-2} \log\frac{2}{\delta}$, we have*

$$\mathbb{P}\left(|\hat{m}_1(X_1,\ldots,X_n) - m_1(F)| \leq R\left(\frac{\varepsilon}{2(1-\varepsilon)} + \sqrt{\frac{2\log(2/\delta)}{n}}\right)\right) \geq 1-\delta.$$

Note that the minimum sample complexity n must grow as $\frac{\varepsilon}{2(1-\varepsilon)}$ approaches $\bar{t}$; this is because our class $\mathcal{H}_{\bar{t},R}$ only assumes control on the $[\frac{1}{2} \pm \bar{t}]$ quantiles.

Proof For each $i \in [n]$, define the the indicator random variable

$$L_i := \mathbb{1}\left\{(D_i = 1) \text{ or } \left(D_i = 0 \text{ and } Y_i \geq Q_{R,F}\left(\tfrac{1}{2(1-\varepsilon)} + a\right)\right)\right\},$$

where $a := \frac{\sqrt{\log(2/\delta)}}{(1-\varepsilon)\sqrt{2n}} < \sqrt{\frac{2\log(2/\delta)}{n}}$. By independence of D_i and Y_i, L_i has mean

$$\mathbb{E}[L_i] = \varepsilon + (1-\varepsilon)\left(1 - \left(\tfrac{1}{2(1-\varepsilon)} + a\right)\right) = \tfrac{1}{2} - (1-\varepsilon)a.$$

Moreover, the $\{L_i\}_{i\in[n]}$ are independent, and thus by Hoeffding's inequality,

$$\mathbb{P}\left(\hat{m}_1 \geq Q_{R,F}\left(\frac{1}{2(1-\varepsilon)} + a\right)\right) \leq \mathbb{P}\left(\sum_{i=1}^n L_i \geq \frac{n}{2}\right) \leq \exp\left(-2n(1-\varepsilon)^2 a^2\right) = \frac{\delta}{2}.$$

Therefore, with probability at least $1 - \frac{\delta}{2}$,

$$\hat{m}_1 - m_1 < Q_{R,F}\left(\frac{1}{2(1-\varepsilon)} + a\right) - Q_{R,F}\left(\frac{1}{2}\right) \leq R\left(\frac{\varepsilon}{2(1-\varepsilon)} + a\right),$$

where the final inequality is due to (2), which we may invoke since $\frac{1}{2} + \frac{\varepsilon}{2(1-\varepsilon)} + a \leq \frac{1}{2} + \bar{t}$ by our choice of n. An identical argument (or by symmetry with $-F$) yields the analogous result for the lower tail of $\hat{m}$, namely that $m_1 - \hat{m}_1 < R(\frac{\varepsilon}{2(1-\varepsilon)} + a)$ with probability at least $1 - \frac{\delta}{2}$. The lemma statement follows by a union bound. ■

3.1.2 Malicious Adversaries

We now turn to malicious adversaries and derive analogous tight, non-asymptotic sample complexity bounds for median estimation from contaminated samples. In this malicious adversarial setting, we will show that it is only possible to obtain estimation accuracy of $R(\varepsilon)$ (Lemma 5), which is weaker that the accuracy of $R(\frac{\varepsilon}{2(1-\varepsilon)})$ obtained against oblivious and prescient adversaries. However, we will also see that this dependency is tight and unavoidable (Lemma 6). Moreover, the $O(\sqrt{\frac{\log 1/\delta}{n}})$ error term in Lemma 5 is tight for the same reason as it was in Lemma 4; see the discussion there. Finally, we remark that the upper bound on ε and the lower bound on the sample complexity n in Lemma 5 are exactly the analogues of the corresponding bounds in Lemma 4; the only difference is that malicious adversaries can force the contaminated distributions to have medians at roughly $F^{-1}(\frac{1}{2}\pm\varepsilon)$, resulting in our need for control of these further quantiles.

Lemma 5 *Let $\bar{t} \in (0, \frac{1}{2})$, $\varepsilon \in (0, \bar{t})$, and $F \in \mathcal{H}_{\bar{t},R}$. Let (Y_i, D_i), for $i \in [n]$, be drawn independently with marginals $Y_i \sim F$ and $D_i \sim Ber(\varepsilon)$. Let $\{Z_i\}_{i\in[n]}$ be arbitrary random variables possibly depending on $\{Y_i, D_i\}_{i\in[n]}$, and define $X_i = (1-D_i)Y_i + D_i Z_i$. Then for any confidence level $\delta \in (0,1)$ and sample complexity $n \geq 2\left(\bar{t}-\varepsilon\right)^{-2} \log \frac{3}{\delta}$,*

$$\mathbb{P}\left(|\hat{m}_1(X_1,\ldots,X_n) - m_1(F)| \leq R\left(\varepsilon + \sqrt{\frac{2\log(3/\delta)}{n}}\right)\right) \geq 1-\delta.$$

Lemma 6 *Let $\bar{t} \in (0, \frac{1}{2})$, $\varepsilon \in (0, \bar{t})$, $\delta \in (0,1)$, $n \geq \frac{1}{2}\varepsilon^{-2}\log\frac{1}{\delta}$, and $R : [0, \bar{t}] \to \mathbb{R}_{\geq 0}$ be any strictly increasing function. Then there exists a distribution $F \in \mathcal{H}_{\bar{t},R}$ and a joint distribution on (D, Y, Z) with marginals $D \sim Ber(\varepsilon)$ and $Y \sim F$, such that*

$$\mathbb{P}\left(|\hat{m}_1(X_1,\ldots,X_n) - m_1(F)| \geq R\left(\varepsilon - \sqrt{\frac{\log(1/\delta)}{2n}}\right)\right) \geq 1-\delta,$$

where each $\{(D_i, Y_i, Z_i)\}_{i\in[n]}$ is drawn independently from the joint distribution, and $X_i := (1-D_i)Y_i + D_i Z_i$.

Informally, the proof of Lemma 5 proceeds by (i) showing that $\hat{m}_1(X_1,\ldots,X_n)$ is deterministically bounded within the order statistics $Y_{(\lfloor\frac{n}{2}\rfloor \pm \sum_{i=1}^n D_i)}$ (Lemma 7), (ii) applying Hoeffding's inequality to argue that w.h.p., at most $\sum_{i=1}^n D_i \approx \varepsilon n$ samples are contaminated, and (iii) reusing the techniques of Lemma 4 to argue that w.h.p., the order statistics of $Y_{(\lfloor\frac{n}{2}\rfloor \pm \varepsilon n)}$ are within the desired error range from $m_1(F)$. Since each of these steps is tight up to a small amount of slack, the proof of the converse Lemma 6 proceeds essentially by just showing that each of these steps occurs also in the opposite direction w.h.p.

The following lemma will be helpful for step (i). Its proof is straightforward by induction on s and is thus omitted.

Lemma 7 *Let $x_i := d_i y_i + (1-d_i)z_i$, where $y_1 \leq \cdots \leq y_n$ and $z_1, \ldots, z_n$ are arbitrary real-valued sequences, and $d_1, \ldots, d_n$ is an arbitrary binary-valued sequence satisfying $s := \sum_{i=1}^n d_i < \frac{n}{2}$. Then*

$$y_{\left(\lfloor\frac{n}{2}\rfloor - s\right)} \leq x_{(med)} \leq y_{\left(\lceil\frac{n}{2}\rceil + s\right)}.$$

Proof [Proof of Lemma 5] Define for shorthand $a := \sqrt{\frac{\log(3/\delta)}{2n}}$. By Hoeffding's inequality, the event $E := \{\sum_{i=1}^n D_i \le (\varepsilon + a)n\}$ occurs with probability at least $\mathbb{P}(E) \ge 1 - \frac{\delta}{3}$. Since $\varepsilon + a \le \bar{t} < \frac{1}{2}$ by our choice of n, Lemma 7 implies that, whenever E occurs,

$$Y_{\left(\lfloor \frac{n}{2} \rfloor - \lfloor (\varepsilon+a)n \rfloor\right)} \le \hat{m}_1(X_1, \dots, X_n) \le Y_{\left(\lceil \frac{n}{2} \rceil + \lfloor (\varepsilon+a)n \rfloor\right)}$$

also occurs. Now, define for each $i \in [n]$ the indicator random variable $L_i := \mathbb{1}(Y_i > Q_{R,F}(\frac{1}{2} + \varepsilon + 2a))$. Then $\mathbb{E}[L_i] \le \frac{1}{2} - \varepsilon - 2a$, so by Hoeffding's inequality,

$$\mathbb{P}\left(Y_{\left(\lceil \frac{n}{2} \rceil + \lfloor (\varepsilon+a)n \rfloor\right)} > Q_{R,F}(\tfrac{1}{2} + \varepsilon + 2a)\right) \le \mathbb{P}\left(\sum_{i=1}^n L_i \ge (\tfrac{1}{2} - \varepsilon - a)n\right) \le \exp(-2na^2) = \frac{\delta}{3}.$$

An identical argument (or simply by symmetry on $F(-\cdot)$) also yields that $Y_{\left(\lfloor \frac{n}{2} \rfloor - \lfloor (\varepsilon+a)n \rfloor\right)} < Q_{L,F}(\frac{1}{2} - \varepsilon - 2a)$ with probability at most $\frac{\delta}{3}$. We conclude by a union bound that with probability at least $1 - \delta$,

$$Q_{L,F}\left(\tfrac{1}{2} - (\varepsilon + 2a)\right) \le \hat{m}_1(X_1, \dots, X_n) \le Q_{R,F}\left(\tfrac{1}{2} + (\varepsilon + 2a)\right).$$

Whenever this occurs, we have by virtue of (2) that $|\hat{m}_1(X_1, \dots, X_n) - m_1(F)| \le R(\varepsilon + 2a)$, since both $\frac{1}{2} \pm (\varepsilon + 2a) \in [\frac{1}{2} \pm \bar{t}]$ by our choice of n. ■

Proof [Proof of Lemma 6] Consider any distribution F with unique median 0 satisfying $R(t) = Q_{R,F}(\frac{1}{2} + t) = -Q_{L,F}(\frac{1}{2} - t)$ for each $t \in \bar{t}$. Such an F can be constructed by starting with a Dirac measure δ_0 at zero and then pushing mass to the tails as far (2) allows. Next, consider the joint distribution on (D, Y, Z) where $Y \sim F$, the conditional distribution of D given Y, is $\mathrm{Ber}(2\varepsilon \cdot \mathbb{1}(Y \le 0))$ and $Z \sim \delta_{Q_{R,F}(\frac{1}{2} + \varepsilon)}$. The marginal of D is easily seen to be correct, since

$$\mathbb{P}(D = 1) = \mathbb{P}(D = 1 | Y \le 0) \cdot \mathbb{P}(Y \le 0) + \mathbb{P}(D = 1 | Y > 0) \cdot \mathbb{P}(Y > 0) = 2\varepsilon \cdot \tfrac{1}{2} = \varepsilon.$$

Let $a := \sqrt{\frac{\log(1/\delta)}{2n}}$, and note that $a < \varepsilon$ by our choice of n. For each $i \in [n]$, define the indicator r.v.

$$L_i := \mathbb{1}\Big((Y_i > R(\varepsilon - a)) \text{ or } (Y_i \le 0 \text{ and } D_i = 1)\Big).$$

Each of these has mean

$$\mathbb{E}L_i \le 1 - (\tfrac{1}{2} + (\varepsilon - a)) + \tfrac{1}{2}(2\varepsilon) = \tfrac{1}{2} + a.$$

Therefore, by Hoeffding's inequality, we conclude that

$$\begin{aligned}\mathbb{P}\Big(\, |\hat{m}_1(X_1, \dots, X_n) - m_1(F)| < R\left(\varepsilon - a\right) \Big) &\le \mathbb{P}\Big(\hat{m}_1(X_1, \dots, X_n) < m_1(F) + R(\varepsilon - a)\Big) \\ &\le \mathbb{P}\Big(\sum_{i=1}^n L_i \ge \frac{n}{2}\Big) \le \exp\left(-2a^2 n\right) = \delta.\end{aligned}$$

■

3.2 Median Estimation Results for $\mathcal{F}_{\bar{t},B}$

In order to obtain more algorithmically useful rates for median estimation from contaminated samples, we now introduce a more specific class of cdfs F that increase at least linearly in a neighborhood around the median. This ensures that F is not "too flat" in this neighborhood, which we stress is a very standard assumption for median estimation.

Definition 8 *For $\bar{t} \in (0, \frac{1}{2})$ and $B > 0$, let $\mathcal{F}_{\bar{t},B}$ be the family of distributions F satisfying*

$$|F(x_1) - F(x_2)| \geq \frac{1}{Bm_2(F)}|x_1 - x_2| \tag{3}$$

for all $x_1, x_2 \in I_{F,\bar{t}} := \big[Q_{L,F}(\frac{1}{2} - \bar{t}), Q_{R,F}(\frac{1}{2} + \bar{t})\big]$.

We make a few remarks about the definition. i) Requiring the right-hand side of (3) to scale inversely in the median absolute deviation (MAD) $m_2(F)$ ensures closure of $\mathcal{F}_{\bar{t},B}$ under scaling; see below. We also mention that $m_2(F)$ is a robust measure of the spread of F (it is the "median moment" analogue of variance), and controls the width of the median's unidentifiability region. ii) If $F \in \mathcal{F}_{\bar{t},B}$, then $F \in \mathcal{H}_{\bar{t},R}$ for $R(t) := Bm_2(F)t$. iii) It is not clear a priori that $m_2(F)$ is even well-defined for F, since it is not clear $m_1(F)$ is unique; however (3) ensures that $m_1(F)$ is unique for all $F \in \mathcal{F}_{\bar{t},B}$ (proven in Lemma 28). iv) Note that distributions in $\mathcal{F}_{\bar{t},B}$ are not required to have densities, nor even be continuous. v) The family $\mathcal{F}_{\bar{t},B}$ has many natural and expected properties, such as closure under scaling and translation. These properties are gathered and proved in Lemma 28, which is deferred to Appendix A for brevity of the main text.

Remark 9 *Most common distributions belong to $\mathcal{F}_{\bar{t},B}$ for some values of the parameters $\bar{t}$ and B. Moreover, by Lemma 28, if a distribution is in $\mathcal{F}_{\bar{t},B}$, then all scaled and translated versions are as well. A short list includes: i) the Normal distribution for $B \geq \frac{q_{3/4}}{\phi(q_{1/2+\bar{t}})}$, where ϕ is the standard Gaussian density and q_α is the corresponding α-quantile, ii) the uniform distribution on any interval with $\bar{t} \in (0, 1/2)$ and $B = 4$, iii) any fixed continuous distribution F with positive density $f = F'$ and $B \geq \Big(m_2(F) \min_{x \in I_{F,\bar{t}}} f(x)\Big)^{-1}$.*

We now apply the estimation results for $\mathcal{H}_{\bar{t},R}$ to estimation for $\mathcal{F}_{\bar{t},B}$. By Lemma 28, whenever $F \in \mathcal{F}_{\bar{t},B}$, then $F \in \mathcal{H}_{\bar{t},R}$ for $R(t) := tBm_2(F)$. Corollary 3 immediately yields a bound on the size of the contamination region in terms of the quantity

$$U_{\varepsilon,B,m_2} := Bm_2 \frac{\varepsilon}{2(1-\varepsilon)}.$$

Corollary 10 *For any $\bar{t} \in (0, \frac{1}{2})$, any $\varepsilon \in (0, \bar{\varepsilon}(\bar{t}))$, and any $F \in \mathcal{F}_{\bar{t},B}$,*

$$\sup_{\substack{\text{distribution } G, \\ \tilde{m}_1 \in m_1((1-\varepsilon)F + \varepsilon G)}} |\tilde{m}_1 - m_1(F)| \leq U_{\varepsilon,B,m_2(F)}.$$

Remark 11 (Tightness of Corollary 10) *For any $a > 0$, let F be the uniform distribution over $[0, a]$. Then $m_1(F) = \frac{a}{2}$ and $m_2(F) = \frac{a}{4}$, and thus $F \in \mathcal{F}_{\bar{t},B}$ for any $\bar{t} \in (0, \frac{1}{2})$ and $B = 4$. Now for any $\varepsilon \in (0, \bar{\varepsilon}(\bar{t})) \subset (0, \frac{1}{2})$, let G be the uniform distribution over $[a, \frac{a}{1-\varepsilon}]$. Then $\tilde{F} := (1-\varepsilon)F + \varepsilon G$ is the uniform distribution over $[0, \frac{a}{1-\varepsilon}]$, and has median $m_1(\tilde{F}) = \frac{a}{2} + \frac{\varepsilon}{2(1-\varepsilon)}a = m_1(F) + U_{\varepsilon,B,m_2(F)}$. An identical argument on $F(-\cdot)$ and $G(-\cdot)$ yields an example where the median decreases by $-U_{\varepsilon,B,m_2(F)}$, so the bound in Corollary 10 is tight.*

Now we turn to median estimation bounds. Since $R(t) = tBm_2(F)$ is clearly Lipschitz, the discussion preceding Lemmas 4 and 5 about the optimality of the following error bounds applies. That is, the error decomposes into the sum of the unavoidable uncertainty term $U_{\varepsilon,B,m_2(F)}$ plus a confidence-interval term decaying at the rate of $n^{-1/2}$ with sub-Gaussian tails, both of which are optimal. For brevity, we omit the proofs of the following corollaries since they follow immediately from Lemmas 4, 5, and 28.

Corollary 12 *Consider the same setup as in Lemma 4, except with the added restriction that $F \in \mathcal{F}_{\bar{t},B}$. Then, for any confidence level $\delta \in (0,1)$, error level $E > 0$, and sample complexity $n \geq 2\max\left(\frac{B^2 m_2^2(F)}{E^2}, \left(\bar{t} - \frac{\varepsilon}{2(1-\varepsilon)}\right)^{-2}\right)\log\frac{2}{\delta}$, we have*

$$\mathbb{P}\left(|\hat{m}_1(X_1, \ldots, X_n) - m_1(F)| \leq U_{\varepsilon,B,m_2(F)} + E\right) \geq 1 - \delta.$$

Corollary 13 *Consider the same setup as in Lemma 5, except with the added restriction that $F \in \mathcal{F}_{\bar{t},B}$. Then for any confidence level $\delta \in (0,1)$, error level $E > 0$, and sample complexity $n \geq 2\max\left(\frac{B^2 m_2^2(F)}{E^2}, (\bar{t} - \varepsilon)^{-2}\right)\log\frac{3}{\delta}$,*

$$\mathbb{P}\left(|\hat{m}_1(X_1, \ldots, X_n) - m_1(F)| \leq U^{(\text{MALICIOUS})}_{\varepsilon,B,m_2(F)} + E\right) \geq 1 - \delta.$$

In Corollary 13 above, we have defined $U^{(\text{MALICIOUS})}_{\varepsilon,B,m_2} := Bm_2\varepsilon$, which is a tight bound on the uncertainty in median estimation that a malicious adversary can induce (see Lemma 6).

3.3 Controlling Quantiles by Estimating the MAD

Our algorithms for CBAI in Section 4 focus on the case when all arms have distributions in $\mathcal{F}_{\bar{t},B}$ and use the confidence region guarantees from Section 3.2 as primitives. As we show there, this is enough for us to select the best arm with high probability. However, in many applications, it is desirable also to provide guarantees on the quality of the selected arm of the form: "with probability at least 80%, a new random variable $Y \sim F_{\hat{I}}$ is at least 10," say.

Such a guarantee could be accomplished directly using the machinery developed in the previous subsection by, for example, estimating the 0.2 quantile in addition to the median (the 0.5 quantile). However, this approach has two problems. First, if we would like to obtain such a guarantee for multiple probability levels (such as $55\%, 60\%, \ldots,$) we would have to perform a separate estimation for each of the corresponding quantiles. Second, and

more importantly, estimation of a quantile q from contaminated samples requires control of quantiles in a $\frac{\varepsilon}{2(1-\varepsilon)}$ neighborhood of q (this follows by an identical argument as in Lemma 1), which can severely restrict the range of quantiles that are possible to estimate.

We circument both of these problems by estimating the *median absolute deviation* (MAD) in addition to the median. The MAD describes the scale of the tails of a distribution away from the median, and is the appropriate analogue to variance (which may not exist for F and, we stress, is not estimatable from contaminated samples) for this contamination model setting.

3.3.1 $\mathcal{F}_{\bar{t},B,\bar{m}_2,\kappa}$: A Class of Distributions with Estimatable MADs

Like the median, the MAD is not fully identifiable from contaminated samples (see discussion in Section 2.2). However, we can estimate the MAD up to a reasonable region of uncertainty with only a few additional assumptions.

Definition 14 *For any $\bar{t} \in (0, \frac{1}{2})$, $B > 0$, $\bar{m}_2 > 0$, and $\kappa \geq 2$, let $\mathcal{F}_{\bar{t},B,\bar{m}_2,\kappa}$ be the family of distributions F satisfying*

(i) (3) *holds for all $x_1, x_2 \in I_{F,\bar{t}} \cup [m_1(F) \pm 2m_2(F)]$.*

(ii) $m_2(F) \leq \bar{m}_2$.

(iii) $m_2(F) \leq \kappa m_4(F)$.

Let us make a few remarks on this definition. i) We emphasize that our CBAI algorithms already have optimality guarantees when the arm distributions are in $\mathcal{F}_{\bar{t},B}$ (see Section 4.3.1); the additional assumption that the arm distributions are in $\mathcal{F}_{\bar{t},B,\bar{m}_2,\kappa}$ allows us to also obtain quantile control of the arm returned by the algorithm (see Section 4.3.2). ii) It is not clear a priori that $m_2(F)$ and $m_4(F)$ are even well-defined for F, since it is not clear that $m_1(F)$ and $m_2(F)$ are unique; however $F \in \mathcal{F}_{\bar{t},B,\bar{m}_2,\kappa} \subseteq \mathcal{F}_{\bar{t},B}$ ensures that $m_1(F)$ is unique, and the possible interval extension in (i) ensures that $m_2(F)$ is also unique for $F \in \mathcal{F}_{\bar{t},B,\bar{m}_2,\kappa}$ (proven in Lemma 28). iii) Property (iii) of the definition requires bounds on the higher robust moments, namely the fourth median moment. Informally, higher robust moments are necessary to control the error of estimation of lower robust moments, analogous to how bounds on variance (resp. kurtosis) are typically necessary for estimation of a distribution's mean (resp. variance). iv) Note that distributions in $\mathcal{F}_{\bar{t},B,\bar{m}_2,\kappa}$ are not required to have densities, nor even be continuous. v) As we will show in Lemma 17 below, the inequality $m_4(F) \leq 2m_2(F)$ holds for any distribution with well-defined $m_4(F)$ and $m_2(F)$; this is why we impose the parameter κ to be at least 2. vi) The family $\mathcal{F}_{\bar{t},B,\bar{m}_2,\kappa}$ has many natural and expected properties, such as closure under scaling and translation. These properties are gathered and proved in Lemma 28, which is deferred to Appendix A for brevity of the main text.

3.3.2 MAD Estimation Results for $\mathcal{F}_{\bar{t},B,\bar{m}_2,\kappa}$.

We are now ready to present finite-sample guarantees on the speed of convergence of the empirical MAD $\hat{m}_2$ to the underlying distribution's MAD m_2 when given contaminated samples. We first present results for the oblivious and prescient adversarial settings. As

before, the estimation error decomposes into the sum of two terms: a bias term reflecting the uncertainty the adversary can inject given her contamination level, and a confidence-interval term that shrinks with an optimal $n^{-1/2}$ rate and sub-Gaussian tails.

Lemma 15 *Let $\bar{t} \in (0, \frac{1}{2})$, $\varepsilon \in (0, \min(\bar{\varepsilon}(\bar{t}), \frac{1}{B}))$, and $F \in \mathcal{F}_{\bar{t},B,\bar{m}_2,\kappa}$. Let $Y_i \sim F$ and $D_i \sim Ber(\varepsilon)$, for $i \in [n]$, all be drawn independently. Let $\{Z_i\}_{i\in[n]}$ be arbitrary random variables possibly depending on $\{Y_i, D_i\}_{i\in[n]}$, and define $X_i = (1 - D_i)Y_i + D_i Z_i$. Then, for any confidence level $\delta > 0$, error level $E > 0$, and sample complexity $n \geq 2\max\left(\frac{16\kappa^2 B^2 \bar{m}_2^2}{E^2}, \left(\min(\bar{t}, \frac{1}{B}) - \frac{\varepsilon}{2(1-\varepsilon)}\right)^{-2}\right)\log\frac{4}{\delta}$,*

$$\mathbb{P}\left(|\hat{m}_2(X_1, \ldots, X_n) - m_2(F)| \leq (1 + 2\kappa)U_{\varepsilon,B,m_2(F)} + E \right) \geq 1 - \delta,$$

where $\hat{m}_2(X_1, \ldots, X_n) := \hat{m}_1(|X_1 - \hat{m}_1(X_1, \ldots, X_n)|, \ldots, |X_n - \hat{m}_1(X_1, \ldots, X_n)|)$.

Informally, we prove Lemma 15 by decomposing the MAD estimation error into two terms, each of which resembles the error between a true median (of a distribution related to F) and an empirical median of contaminated samples, and then applying the median estimation guarantees from the previous section. Before proving Lemma 15, however, we first present two lemmas which be helpful in the proof.

Lemma 16 *For any sequence $x_1, \ldots, x_n \in \mathbb{R}$ and any $c \in \mathbb{R}$,*

$$\Big| m_1(|x_1 + c|, \ldots, |x_n + c|) - m_1(|x_1|, \ldots, |x_n|) \Big| \leq |c|.$$

Proof For any fixed vector $x \in \mathbb{R}^n$, the function $c \mapsto m_1(|x_1 + c|, \ldots, |x_n + c|)$ is 1-Lipschitz since it is the composition of 1-L_∞ Lipschitz functions: adding $c\vec{1}$, taking entrywise absolute values, and taking an order statistic. ■

Lemma 17 *For any distribution F with well-defined $m_2(F)$ and $m_4(F)$,*

$$m_4(F) \leq 2m_2(F).$$

Proof Let $Y \sim F$. Then, by an application of Lemma 16,

$$m_4(F) = m_1\left(\big|\, |Y - m_1(F)| - m_2(F)\big|\right) \leq m_1\left(|Y - m_1(F)|\right) + m_2(F) = 2m_2(F).$$

■

We are now ready to prove Lemma 15.

Proof [Proof of Lemma 15] For shorthand, denote $\hat{m}_1 := \hat{m}_1(X_1, \ldots, X_n)$, and let H denote the induced folded distribution of $|Y - m_1(F)|$ where $Y \sim F$. By the fact $m_2(F) = m_1(H)$,

Lemma 16, and the triangle inequality, the MAD estimation error is bounded above by:

$$\begin{aligned}
&\left|\hat{m}_2(X_1,\dots,X_n) - m_2(F)\right| \\
&= \left|\hat{m}_1\left(|X_1-\hat{m}_1|,\dots,|X_n-\hat{m}_1|\right) - m_1(H)\right| \\
&\le \left|\hat{m}_1\left(|X_1-m_1(F)|,\dots,|X_n-m_1(F)|\right) - m_1(H)\right| + \left|\hat{m}_1 - m_1(F)\right|. \qquad (4)
\end{aligned}$$

By Corollary 12 and our choice of n, the second error term in (4) is bounded above by $U_{\varepsilon,B,m_2(F)} + \frac{E}{2}$ with probability at least $1-\frac{\delta}{2}$. To control first error term in (4), we apply Corollary 12 with the distribution H in lieu of F and the contaminations $\tilde{Z}_i := |Z_i - m_1(F)|$ in lieu of Z_i. Thus, by combining item 8 in Lemma 28 with Corollary 12, this first error term is bounded above by $U_{\varepsilon,\kappa B,m_2(H)} + \frac{E}{2}$ with probability at least $1-\frac{\delta}{2}$ whenever we have at least $2\max\left(4B^2\kappa^2 m_2^2(H)E^{-2}, (\min(\frac{1}{2},\frac{1}{B}) - \frac{\varepsilon}{2(1-\varepsilon)})^{-2}\right)\log\frac{4}{\delta}$ samples, which is satisfied because of our choice of n and the inequality

$$m_2(H) = m_4(F) \le 2m_2(F) \qquad (5)$$

from Lemma 17. Therefore, a union bound implies that, with probability at least $1-\delta$, the MAD estimation error is at most $U_{\varepsilon,B,m_2(F)} + U_{\varepsilon,\kappa B,m_2(H)} + E$. By another application of (5), this is bounded above by $(1+2\kappa)U_{\varepsilon,B,m_2(F)} + E$, as desired. ■

Similar to median estimation above, it is possible to estimate the MAD even in the malicious adversarial setting. The proof is omitted since it is identical to the proof of Lemma 15 with uses of Corollary 12 replaced by uses of Corollary 13.

Lemma 18 *Let $\bar{t} \in (0,\frac{1}{2})$, $\varepsilon \in (0, \min(\bar{t},\frac{1}{B}))$, and $F \in \mathcal{F}_{\bar{t},B,\bar{m}_2,\kappa}$. Let (Y_i,D_i), for $i\in[n]$, be drawn independently with marginals $Y_i \sim F$ and $D_i \sim Ber(\varepsilon)$. Let $\{Z_i\}_{i\in[n]}$ be arbitrary random variables possibly depending on $\{Y_i,D_i\}_{i\in[n]}$, and define $X_i = (1-D_i)Y_i + D_iZ_i$. Then, for any confidence level $\delta > 0$, error level $E > 0$, and sample complexity $n \ge 2\max\left(\frac{16\kappa^2B^2\bar{m}_2^2}{E^2}, \left(\min(\bar{t},\frac{1}{B})-\varepsilon\right)^{-2}\right)\log\frac{6}{\delta}$,*

$$\mathbb{P}\left(|\hat{m}_2(X_1,\dots,X_n) - m_2(F)| \le (1+2\kappa)U^{(\text{MALICIOUS})}_{\varepsilon,B,m_2(F)} + E\right) \ge 1-\delta,$$

where $\hat{m}_2(X_1,\dots,X_n) := \hat{m}_1(|X_1-\hat{m}_1(X_1,\dots,X_n)|,\dots,|X_n-\hat{m}_1(X_1,\dots,X_n)|)$.

4. Algorithms

Our algorithms for CBAI actually work for the more general problem of best arm identification in *partially identified* settings. Specifically, consider any setting where the statistic (for example, the median or mean) which measures the goodness of an arm can be estimated only up to some unavoidable error term due to lack of identifiability. The main result of this section is informally that BAI algorithms for the classical stochastic multi-armed bandit setting can be adapted with little modification to partially identified settings. Perhaps surprisingly, this suggests a certain innate robustness of many existing classical BAI algorithms.

Because of this generality, we present our algorithms for this slightly more abstract problem of best arm identification under partial identifiability, which we define formally in Section 4.1. In Section 4.2, we detail how to adapt classical BAI algorithms for this setting, and in Section 4.3, we apply these guarantees to our CBAI problem.

4.1 Best Arm Identification Under Partial Identifiability

Let us formally define the setup of the *Partially Identifiable Best-Arm-Identification* problem (PIBAI). Let $k \geq 2$ be the number of arms. For each arm $i \in [k]$, consider a family of distributions $\mathcal{D}_i = \{D_i(p_i, G)\}_{G \in \mathcal{G}}$ where p_i is a parameter of interest associated with each arm and G is a nuisance parameter in some abstract space $\mathcal{G}$. We let $i^* := \arg\max_{i \in [k]} p_i$ be the best arm. We assume the existence of non-negative unavoidable biases $\{U_i\}_{i \in [k]}$ satisfying

(i) even from infinitely many independent samples $X_t, t = 1, 2, \ldots$ with $X_t \sim D_i(p_i, G_t)$ for some unknown, possibly varying $G_t \in \mathcal{G}$ $(t \geq 1)$, it is impossible to estimate p_i more precisely than the region $[p_i \pm U_i]$, and

(ii) there exists some estimator that, for any $\alpha > 0$ and $\delta \in (0, 1)$, uses $n_{\alpha,\delta} = O(\alpha^{-2} \log \frac{1}{\delta})$ i.i.d. samples[1] from D_i to output an estimate $\hat{p}_i$ satisfying

$$\mathbb{P}\Big(\hat{p}_i \in [p_i \pm (U_i + \alpha)]\Big) \geq 1 - \delta. \tag{6}$$

The PIBAI problem is then precisely the standard fixed-confidence stochastic bandit game using these distributions where in each iteration t, the algorithm chooses an arm $I_t \in [k]$, and receives a sample from $D_{I_t}(p_i, G_t)$ for some unknown $G_t \in \mathcal{G}$.

By the partial identifiability property (i), it is clear that even given infinite samples, it is impossible to distinguish between the optimal arm i^* and any suboptimal arm $i \neq i^*$ satifsying $p_i + U_i \geq p_{i^*} - U_{i^*}$. (This is made formal in the lower bound section; see Section 5.1 for details.) Therefore, we assume henceforth the statistically possible setting in which the effective gaps $\tilde{\Delta}_i = (p_{i^*} - U_{i^*}) - (p_i + U_i)$ are strictly positive for each suboptimal arm $i \neq i^*$.

For any $\alpha \geq 0$, arm i is said to be *α-suboptimal* if $\tilde{\Delta}_i \leq \alpha$. Moreover, for any $\alpha \geq 0$ and $\delta \in (0, 1)$, a PIBAI algorithm is said to be *(α, δ)-PAC* if it outputs an arm $\hat{I}$ that is α-suboptimal with probability at least $1 - \delta$. That is,

$$\mathbb{P}\big(\tilde{\Delta}_{\hat{I}} \leq \alpha\big) \geq 1 - \delta,$$

where the above probability is taken over the possible randomness of the samples, estimator from (ii), and PIBAI algorithm.

1. Using the same techniques as presented in this paper, one can also consider PIBAI for general $n_{\alpha,\delta} \neq O(\alpha^{-2} \log \frac{1}{\delta})$ and compute the resulting sample complexities. However, for simplicity of presentation, we assume that $n_{\alpha,\delta} \neq O(\alpha^{-2} \log \frac{1}{\delta})$ since anyways this is the natural (and optimal) quantity for many estimation problems such as estimating a median from contaminated samples (see Corollaries 12 and 13); or using Chernoff bounds to estimate the mean of a $[0, 1]$-supported distribution for classical stochastic MAB, etc.

$\boldsymbol{S} \leftarrow [k]$
while $|S| > 1$ **do**
  Sample each arm $i \in S$ for $n_{\alpha/2,\delta/4}$ times, and generate estimates $\hat{p}_i$
  $S \leftarrow \{i \in S \,:\, \hat{p}_i \geq \hat{m}_1(\{\hat{p}_j\}_{j\in S})\}$
  $\delta \leftarrow \frac{\delta}{2}$
end
Output the only arm left in S

Algorithm 1: Adaptation of the Median Elimination algorithm for PIBAI.

4.2 Algorithms for Best Arm Identification Under Partial Identifiability

We present three algorithms for PIBAI: a naïve Algorithm that performs uniform exploration among all arms, and adaptations of the Median Elimination and Successive Elimination algorthms of (Even-Dar et al., 2006). Pseudocode is given in accompanying figures and uses as a blackbox an estimator satisfying the above property (ii) of the PIBAI problem.

4.2.1 Naïve Algorithm for PIBAI

A simple first attempt at an (α, δ)-PAC PIBAI algorithm is the following: pull each of the k arms $n_{\alpha/2,\delta/k}$ times to create estimates $\hat{p}_i$ and output the arm $\hat{I} := \max_{i\in[k]} \hat{p}_i$ with the highest estimate. By (6) and a union bound, we have that with probability at least $1-\delta$, all estimates $\hat{p}_i \in [p_i \pm (U_i + \frac{\alpha}{2})]$. Whenever this occurs,

$$\tilde{\Delta}_{\hat{I}} = (p_{i^*} - U_{i^*}) - (p_i + U_i) \leq (\hat{p}_{i^*} + \tfrac{\alpha}{2}) - (\hat{p}_i - \tfrac{\alpha}{2}) \leq \alpha,$$

implying that $\hat{I}$ is α-suboptimal and that the (deterministic) sample complexity of the naïve algorithm is $O\left(\frac{k}{\alpha^2}\log\frac{k}{\delta}\right)$.

However, this sample complexity is suboptimal for two important reasons. First, there extra factor of k in the logarithm (introduced by the crude union bound) which is not contained in the lower bound we will prove in Section 5.1. Second, the sample complexity is not adaptive to the difficulty of the actual instance: an arm is sampled $n_{\alpha/2,\delta/k}$ times even if it is far from α-suboptimal (meaning that $\tilde{\Delta}_i \gg \alpha$) and could potentially be eliminated much more quickly. Both of these problems were remedied in the classical multi-armed bandit literature by more sophisticated algorithms; we show below how to modify these algorithms for the PIBAI setting to solve both these problems.

4.2.2 Median Elimination Algorithm for PIBAI

A simple modification of the Median Elimination algorithm (Even-Dar et al., 2006) removes the $\log k$ factor from the upper bound of the naïve algorithm. Pseudocode is given in Algorithm 1. The primary change from the original algorithm is that for PIBAI, there is no need to have geometrically decaying estimation error in the rounds. That is, the original Median Elimination Algorithm for BAI requires taking roughly $n_{(3/4)^r\alpha,2^{-r}\delta}$ samples on the round r, whereas we need only take roughly $n_{\alpha,2^{-r}\delta}$ samples.

Theorem 19 *For any $\alpha > 0$ and $\delta \in (0,1)$, Algorithm 1 is an (α,δ)-PAC* PIBAI *algorithm with sample complexity $O\left(\frac{k}{\alpha^2}\log\frac{1}{\delta}\right)$.*

$\boldsymbol{S} \leftarrow [k]$, $r \leftarrow 1$
while $|S| > 1$ **do**
  Sample each arm $i \in S$ once and produce $\hat{p}_{i,r}$ from all r past samples of it
  $S \leftarrow \{i \in S : \hat{p}_{i,r} \leq \max_{j \in S} \hat{p}_{j,r} - 2\alpha_{r,6\delta/(\pi^2 k r^2)}\}$
  $r \leftarrow r + 1$
end
Output the only arm left in S

Algorithm 2: Adaptation of Successive Elimination algorithm for PIBAI. Here, $\alpha_{r,\delta} := \sqrt{\frac{c \log \frac{1}{\delta}}{r}}$, where c is a universal constant satisfying $n_{\alpha,\delta} \leq c\alpha^{-2} \log \frac{1}{\delta}$ (see (6)).

The proof is deferred to Appendix B since it follows closely the analysis of the original algorithm (Even-Dar et al., 2006, Theorem 10), but we highlight the main differences here. In order to guarantee that Algorithm 1 returns an α-suboptimal arm for PIBAI, we must ensure that *all arms that are more than α-suboptimal are eliminated before the optimal arm is eliminated* (if it ever is). This is in contrast to the original proof for BAI in the classical multi-armed bandit setup: there, it is not problematic if the best (or even currently best) arm is eliminated at round r, so long as the best arm in consecutive rounds r and $r+1$ changes at most by a small amount.[2] At its essence, this is due to the following "additive property of suboptimality" for BAI:

> "If arm i has Δ_i suboptimality gap w.r.t. the optimal arm i^*, and if arm j has $\Delta_j^{(i)}$ suboptimality gap w.r.t. arm i, then arm j has suboptimality gap $\Delta_j = \Delta_i + \Delta_j^{(i)}$ w.r.t. the optimal arm i^*."

The critical point is that the same argument does *not* work for PIBAI since errors propagate from adding the uncertainties U_i in the suboptimality gaps: that is, $(p_{i^*} - U_{i^*}) - (p_j + U_j) \neq [(p_{i^*} - U_{i^*}) - (p_i + U_i)] + [(p_i - U_i) - (p_j + U_j)]$. We stress that this nuance is not merely a technicality but actually fundamental to the correctness proofs for PIBAI.

4.2.3 Successive Elimination Algorithm for PIBAI

We can obtain instance-adaptive sample complexity by modifying the Successive Elimination algorithm of (Even-Dar et al., 2006). Pseudocode is given in Algorithm 2. The algorithm and analysis are almost identical to the original. As such, proof details are deferred to Appendix B.2. The main difference is that in the proof of correctness, we show the event $\{|\hat{p}_{i,r} - p_i| \leq U_i + n_{\alpha,(\delta\pi^2)/(6kr)},\ \forall r \in [R],\ \forall i \in S_r\}$ occurs with probability at least $1 - \delta$. This ensures that in each round r, each estimate $\hat{p}_{i,r}$ is accurate enough to use for the elimination step.

Note that Algorithm 2 returns the optimal arm w.h.p. (without knowing the smallest effective gap), unlike Algorithm 1 above which only returns a near-optimal arm w.h.p.

Theorem 20 *Let $\delta \in (0,1)$. With probability at least $1-\delta$, Algorithm 1 outputs the optimal arm after using at most $O\left(\sum_{i \neq i^*} \frac{1}{\tilde{\Delta}_i^2} \log\left(\frac{k}{\delta \tilde{\Delta}_i}\right)\right)$ samples.*

2. In particular, at most $2^{-r}\alpha$, since this implies that the final arm $\hat{I}$ is at most $\sum_{r=1}^{\infty} 2^{-r}\alpha = \alpha$-suboptimal.

Moreover, as noted in Remark 9 of (Even-Dar et al., 2006), Algorithm 2 is easily modified (simply terminate early) to be an (α, δ)-PAC algorithm with sample complexity

$$O\left(\frac{N_\alpha}{\alpha^2}\log\left(\frac{N_\alpha}{\delta}\right) + \sum_{i\in[k]\,:\,\tilde{\Delta}_i>\alpha}\frac{1}{\tilde{\Delta}_i^2}\log\left(\frac{k}{\delta\tilde{\Delta}_i}\right)\right),$$

where N_α is the number of α-suboptimal arms with $\tilde{\Delta}_i \le \alpha$.

Remark 21 *In the multi-armed bandit literature, the sample complexity of the Successive Elimination algorithm (tight up to a logarithmic factor) was improved upon by (Karnin et al., 2013)'s Exponential-Gap Elimination (EGE) algorithm (tight up to a doubly logarithmic factor). A natural idea is to analogously improve upon the* PIBAI *guarantee in Theorem 20 by adapting the EGE algorithm. However, this does not work. The reason for this inadaptability of the EGE – in contrast to the easy adaptability of the above-described algorithms – is that EGE relies heavily upon the "additive property of suboptimality" for MAB, which does not hold for* PIBAI *(see discussion following Theorem 19 for details).*

4.3 Algorithms for Best Arm Identification for Contaminated Bandits

In this subsection, we apply the algorithms developed in Section 4.2 for the general problem of PIBAI to the special case of CBAI. In order to implement the estimator in part (ii) of the definition of PIBAI (see Section 4.1), we make use of the guarantees proved in Section 3 for estimation from contaminated samples.

First, in Section 4.3.1 we prove PAC guarantees for CBAI; then in Section 4.3.2 we prove guarantees on the quality of the selected arm. The former task will require only median estimation from contaminated samples, and so we will just assume that the arms' distributions are in $\mathcal{F}_{\bar{t},B}$ and have robust second moments uniformly bounded[3] by some $\bar{m}_2$. The latter task will also require estimation of the MAD, and thus we will assume there that the arms' distributions are instead in $\mathcal{F}_{\bar{t},B,\bar{m}_2,\kappa}$. The parameters of these families are assumed to be known to the algorithm beforehand.

The sample complexities we prove below are in terms of the *effective gaps* $\tilde{\Delta}_i := (m_1(F_{i^*}) - U_{i^*}) - (m_1(F_i) + U_i)$ of the suboptimal arms $i \neq i^*$, where $i^* := \arg\max_{i\in[k]} m_1(F_i)$ is the arm with the highest median. Here U_i is the unavoidable uncertainty term for median estimation from contaminated samples under the given adversarial setting. Recall from Section 3 we computed these U_i explicitly: $U_i = U_{\varepsilon,B,m_2(F_i)}$ for oblivious and prescient adversaries by Corollary 12 and $U_i = U^{(\text{MALICIOUS})}_{\varepsilon,B,m_2(F_i)}$ for malicious adversaries by Corollary 13. We emphasize that this means the strength of the adversarial setting is encapsulated in the corresponding U_i.

4.3.1 PAC Guarantees

First, we discuss how Algorithm 1 yields an (α, δ)-PAC algorithm for CBAI without modification. Note that in this setting $\hat{p}_i$ is the empirical median of the payoffs of arm i. The

3. It is typically necessary to have bounds on higher order moments in order to control the error of estimating lower order moments, see e.g. (Bubeck et al., 2013).

$\boldsymbol{S} \leftarrow [k]$, $r \leftarrow 1$
Sample each arm $n_0(\delta)$ times
while $|S| > 1$ **do**
  Sample each arm $i \in S$ for $(2N+1)$ times and produce $\hat{p}_{i,r}$ from all past samples
  $S \leftarrow \{i \in S : \hat{p}_{i,r} \leq \max_{j \in S} \hat{p}_{j,r} - 2\alpha_{r,6\delta/(\pi^2 k r^2)}\}$
  $r \leftarrow r+1$
end
Output the only arm left in S

Algorithm 3: Adaptation of successive elimination algorithm (see Algorithm 2) for CBAI. Here, $\alpha_{r,\delta} := \sqrt{\frac{2B\bar{m}_2^2 \log\frac{3}{\delta}}{r}}$.

quantity $n_{\alpha,\delta}$ is equal to $2\max\left(\frac{B^2\bar{m}_2^2}{\alpha^2}, \left(\bar{t} - \frac{\varepsilon}{2(1-\varepsilon)}\right)^{-2}\right)\log\frac{2}{\delta}$ against oblivious and prescient adversaries by Corollary 12, and is equal to $2\max\left(\frac{B^2\bar{m}_2^2}{\alpha^2}, (\bar{t}-\varepsilon)^{-2}\right)\log\frac{3}{\delta}$ against malicious adversaries by Corollary 13. Theorem 19 along with the observation that the constant term in the sample complexity $n_{\alpha,\delta}$ only introduces a term that is negligible in the overall sample complexity of the Median Elimination Algorithm, immediately yields the following result.

Theorem 22 *Let $F_i \in \mathcal{F}_{\bar{t},B}$ with $m_2(F) \leq \bar{m}_2$ for each arm $i \in [n]$, and let the adversary be oblivious, prescient, or malicious. For any $\alpha > 0$ and $\delta \in (0,1)$, Algorithm 1 is an (α,δ)-PAC* CBAI *algorithm with sample complexity $O\left(\frac{k}{\alpha^2}\log\frac{1}{\delta}\right)$.*

Next, we discuss how to apply Algorithm 2. In the traditional Successive Elimination Algorithm for BAI, concentration inequalities are needed even for small sample sizes, since during the first round each arm is pulled only once. However, our guarantees from Section 2 are only valid when the sample size is greater than a certain threshold. Hence, we need to modify Algorithm 2 to obtain additional samples in an initial exploration phase. Pseudocode is given in Algorithm 3. There, $N = \left(\bar{t} - \frac{\varepsilon}{2(1-\varepsilon)}\right)^{-2}$ and $n_0(\delta) = 2N\log\frac{2}{\delta}$ against oblivious and prescient adversaries; and $N = (\bar{t}-\varepsilon)^{-2}$ and $n_0(\delta) = 2N\log\frac{3}{\delta}$ against malicious adversaries.

Theorem 23 *Let $F_i \in \mathcal{F}_{\bar{t},B}$ with $m_2(F) \leq \bar{m}_2$ for each arm $i \in [n]$, and let the adversary be either oblivious, prescient, or malicious. Let $\delta \in (0,1)$. Set $N = \left(\bar{t} - \frac{\varepsilon}{2(1-\varepsilon)}\right)^{-2}$ for the oblivious and prescient cases, and $N = (\bar{t}-\varepsilon)^{-2}$ for the malicious case. With probability at least $1-\delta$, Algorithm 3 outputs the optimal arm after using at most $O\left(kN\log\frac{1}{\delta} + \sum_{i\neq i^*}\frac{1}{\tilde{\Delta}_i^2}\log\left(\frac{k}{\delta\tilde{\Delta}_i}\right)\right)$ samples.*

4.3.2 Quality Guarantees for the Selected Arm

As discussed in Section 3.3, it is often desirable in applications not only to output the best arm, but also to output a guarantee on its quality. The previous section discussed outputing the best arm w.h.p.; we now show how the CBAI algorithms also provide a suboptimality guarantee to a certain precision "for free" without needing extra samples. We describe this

property below for our adapted Successive Elimination Algorithm 3. A nearly identical (yet technically hairier) argument yields a similar guarantee for our adapted Median Elimination Algorithm 1.

The following simple lemma will be helpful to prove the desired types of guarantees. In words, it shows that the lower tail of any distribution $F \in \mathcal{F}_{\bar{t},B,\bar{m}_2,\kappa}$ is controlled by its median and MAD.

Lemma 24 *Let $F \in \mathcal{F}_{\bar{t},B,\bar{m}_2,\kappa}$ and $Y \sim F$. Then simultaneously for all $t \in [0,\bar{t}]$,*

$$\mathbb{P}\Big(Y \geq m_1(F) - tBm_2(F)\Big) \geq \tfrac{1}{2} + t.$$

Proof By item (6) of Lemma 28, we have that for all $t \in [0,\bar{t}]$, $t = \frac{1}{2} - (\frac{1}{2} - t) \geq \frac{1}{Bm_2(F)}(F^{-1}(\frac{1}{2}) - F^{-1}(\frac{1}{2} - t))$. Rearranging yields $F^{-1}(\frac{1}{2} - t) \geq m_1(F) - Bm_2(F)t$ and completes the proof. ■

With this lemma in hand, we turn to guarantees for the arm returned by Algorithm 1. The following guarantees hold for all adversarial settings since the adversarial strength is encapsulated in the definition of the unavoidable uncertainty terms U_i.

Theorem 25 *Let $F_i \in \mathcal{F}_{\bar{t},B,\bar{m}_2,\kappa}$, and let the adversary be either oblivious, prescient, or malicious. Consider using Algorithm 1 for* CBAI *exactly as in Theorem 22. Let $\hat{I}$ denote the arm it outputs, let Y be a random variable whose conditional distribution on $\hat{I}$ is $F_{\hat{I}}$, and let $\hat{m}_1$ and $\hat{m}_2$ denote the empirical median and MAD, respectively, from the samples it has seen from $F_{\hat{I}}$. Then, simultaneously for all $t \in [0,\bar{t}]$,*

$$\mathbb{P}\left(Y \geq [\hat{m}_1 - tB\hat{m}_2] - \left[(1+Bt)\frac{4\kappa\alpha}{\sqrt{\log_2 k}} + (1+(1+2\kappa)Bt)U_{\varepsilon,B,\bar{m}_2}\right]\right) \geq \tfrac{1}{2} + t - \delta.$$

Proof By definition of Algorithm 1, $\hat{I}$ has has stayed in S for the entirety of the algorithm. Therefore, by the discussion preceding Theorem 22, $\hat{I}$ has been sampled at least $n_{\alpha,\delta}\log_2 k = 2\max\left(\frac{B^2\bar{m}_2^2}{\alpha^2}, \left(\bar{t} - \frac{\varepsilon}{2(1-\varepsilon)}\right)^{-2}\right)\log\frac{2}{\delta}\log_2 k$ times against oblivious and prescient adversaries and $n_{\alpha,\delta}\log_2 k = 2\max\left(\frac{B^2\bar{m}_2^2}{\alpha^2}, (\bar{t}-\varepsilon)^{-2}\right)\log\frac{3}{\delta}\log_2 k$ against malicious adversaries. Therefore, by Corollary 12 and Lemma 15, we have that with probability at least $1-\delta$, both of the inequalities

$$\begin{aligned} m_1(F_{\hat{I}}) &\geq \hat{m}_1 - U_{\varepsilon,B,m_2(F)} + E, \\ m_2(F_{\hat{I}}) &\leq \hat{m}_2 + (1+2\kappa)U_{\varepsilon,B,m_2(F)} + E \end{aligned}$$

hold, where $E := \frac{4\kappa\alpha}{\sqrt{\log_2 k}}$. Whenever this event occurs, we have that $m_1(F) - tBm_2(F) \geq [\hat{m}_1 - tB\hat{m}_2] - [(1+Bt)E + (1+(1+2\kappa)Bt)U_{\hat{I}}]$. We conclude by applying Lemma 24, noting that $U_{\hat{I}} \leq U_{\varepsilon,B,\bar{m}_2}$, and using a union bound. ■

It is worth noting that the bound inside the probability term in Theorem 25 can be far less conservative than the crude bound $\hat{m}_1 - tB\bar{m}_2$ if ε and α are small.

5. Lower Bounds

This section provides information-theoretic lower bounds on the sample complexity of the CBAI problem that match, up to small logarithmic factors, the algorithmic upper bounds proved in Section 4.3.1 for each of the three adversarial settings. The main insight is that we can reduce hard instances of the BAI problem to hard instances of the CBAI problem. In this way, we can leverage the sophisticated lower bounds already developed in the classical multi-armed bandit literature.

5.1 Statement of Lower Bounds

Let $i^* := \arg\max_{i\in[k]} m_1(F_i)$ denote the optimal arm. As in Section 4.3, define for each suboptimal arm $i \neq i^*$ the effective gap $\tilde{\Delta}_i := (m_1(F_{i^*}) - U_{i^*}) - (m_1(F_i) + U_i)$, where U_i is the unavoidable uncertainty term for median estimation. That is, $U_i = U_{\varepsilon,B,m_2(F_i)}$ for oblivious and prescient adversaries by Corollary 12, and $U_i = U^{(\text{PRESCIENT})}_{\varepsilon,B,m_2(F_i)}$ for malicious adversaries by Corollary 13. Since the power of the adversary is encapsulated in the U_i and therefore also the effective gaps, we can address all three adversarial settings simultaneously by proving a lower bound in terms of the corresponding effective gaps.

In Section 4.1, we argued that the PIBAI problem is impossible when there exists suboptimal arms with non-positive effective gaps; the CBAI problem clearly also has this property. Indeed, if $i \neq i^*$ satisfies $\tilde{\Delta}_i \leq 0$, then there exist distributions G_{i^*} and G_i such that the resulting contaminated distributions $\tilde{F}_i := (1-\varepsilon)F_i + \varepsilon G_i$ and $\tilde{F}_{i^*} := (1-\varepsilon)F_{i^*} + \varepsilon G_{i^*}$ have equal medians. Moreover, since the distributions F are arbitrary (CBAI makes no parameteric assumption), it is impossible to determine whether arm i or i^* has a higher true median. Thus, any CBAI algorithm, even with infinite samples, cannot succeed with probability more than $\frac{1}{2}$.

Therefore, we only consider the setting where all effective gaps $\tilde{\Delta}_i$ are strictly positive. The results in this section lower bound the sample complexity of any CBAI algorithm in terms of the effective gaps. We will focus on lower bounds for the function class $\mathcal{F}_{\bar{t},B}$ and provide lower bounds matching the upper bounds for CBAI in Section 4.3.1.

Theorem 26 *Consider* CBAI *against an oblivious, prescient, or malicious adversary. There exists positive constants δ' and B such that, for any number of arms $k \geq 2$, confidence level $\delta \in (0, \delta')$, suboptimality level $\alpha \in (0, \frac{1}{6})$, contamination level $\varepsilon \in (0, \frac{1}{15})$, regularity level $\bar{t} \in (0, \frac{1}{10})$, and (α,δ)-PAC* CBAI *algorithm, there exists a* CBAI *instance with $F_1, \ldots, F_k \in \mathcal{F}_{\bar{t},B}$, for which the algorithm uses at least the following number of samples in expectation:*

$$\mathbb{E}[T] = \Omega\left(\sum_{i\in[k]\setminus\{i^*\}} \frac{1}{\max(\tilde{\Delta}_i, \alpha)^2} \log\tfrac{1}{\delta}\right),$$

where $i^* = \arg\max_{i\in[k]} m_1(F_i)$ is the optimal arm.

Taking the limit as $\alpha \to 0$ in Theorem 26 immediately yields the following lower bound on $(0,\delta)$-PAC algorithms.

Corollary 27 *Consider the same setup as in Theorem 26. Any* $(0,\delta)$*-PAC* CBAI *algorithm uses at least the following number of samples in expectation:*

$$\mathbb{E}[T] = \Omega\left(\sum_{i\in[k]\setminus\{i^*\}} \frac{1}{\tilde{\Delta}_i^2}\log\tfrac{1}{\delta}\right).$$

5.2 Proof Sketch

We give here proofs sketches for Theorem 26 and Corollary 27. Full details are deferred to Appendix C.

The key idea in the proof is to "lift" hard BAI instances to hard CBAI problem instances. Specifically, let $\{P_i\}_{i\in[k]}$ be distributions for arms in a BAI problem, let $\{U_i\}_{i\in[k]}$ be the corresponding unavoidable uncertainties for median estimation of P_i in CBAI , and let i^* denote the best arm. We say $\{P_i\}_{i\in[k]}$ admits an "$(\varepsilon, \mathcal{F}_{\bar{t},B})$*CBAI-lifting*" to $\{F_i\}_{i\in[k]}$ if (i) $F_i \in \mathcal{F}_{\bar{t},B}$ for each i, (ii) there exists some adversarial distributions $\{G_i\}_{i\in[k]}$ such that each $P_i = (1-\varepsilon)F_i + \varepsilon G_i$; and (iii) the effective gaps $\tilde{\Delta}_i$ for the $\{F_i\}_{i\in[k]}$ are equal to the gaps Δ_i in the original BAI problem. Intuitively, such a lifting can be thought of as choosing F_{i^*} to have the largest possible median and the other F_i to have the smallest possible median, while still being consistent with P_i and remaining in the uncertainty region corresponding to the U_i.

Armed with this informal definition, we now outline the central idea behind the reduction. Let $\{P_i\}_{i\in[k]}$ be a "hard" BAI instance with best arm i^*, and assume it admits an $(\varepsilon, \mathcal{F}_{\bar{t},B})$-CBAI-lifting to $\{F_i\}_{i\in[k]}$. Consider running a CBAI algorithm $\mathcal{A}$ on the samples obtained from $\{P_i\}$; we claim that if $\mathcal{A}$ is (α,δ)-PAC, then it must output an α-suboptimal arm for the original BAI problem. Indeed, by (ii), the samples from the BAI instance have the same law as the samples that would be obtained from the CBAI problem with $\{F_i\}$, and by (iii), the effective gaps in the CBAI problem are equal to the gaps in the original problem. Therefore, $\mathcal{A}$ must have sample complexity for this problem that is no smaller than the sample complexity of the best BAI algorithm.

Let us make a few remarks about this proof. First, note that the $\{F_i\}$ in the lifting need only *exist*, as we do not explicitly use these distributions but instead simply run $\mathcal{A}$ on samples from the original bandit instance. Thus ensuring the existence of such a lifting is the main obstacle.

This raises a technical yet important nuance: most BAI lower bounds are constructed from Bernoulli or Gaussian distributions which are not quite compatible with the reduction described above. The problem with Bernoulli arms is that they do not have liftings to $\mathcal{F}_{\bar{t},B}$: any resulting F_i would not satisfy (3) and thus cannot be in $\mathcal{F}_{\bar{t},B}$. Gaussian arms run into a different problem: it is not clear how to shift Gaussian arms up or down far enough to change the median by exactly the maximum uncertainty amount U_i. We overcome this nuance by considering arms with *smoothed Bernoulli distribution* $\mathrm{SBer}(p)$, which we define to be the uniform mixture between a Bernoulli distribution with parameter p and a uniform distribution over $[0,1]$. Indeed, this distribution is $(\varepsilon, \mathcal{F}_{\bar{t},B})$-CBAI-liftable: unlike the Bernoulli distribution, it is smooth enough to have liftings in $\mathcal{F}_{\bar{t},B}$; and unlike the Gaussian distribution, the median of the appropriate lifting of $\mathrm{SBer}(p)$ is exactly U_i away

from the median of $\mathrm{SBer}(p)$. Both of these facts are simple calculations; see Appendix C for details.

The only ingredient remaining in the proof is to prove that there are hard instances for BAI with SBer-distributed arms. Intuitively, this can be argued as follows. Let $\vec{p} \in [0,1]^k$ be such that $\{\mathrm{Ber}(p_i)\}_{i\in[k]}$ is a hard instance for BAI (such as the one from (Mannor and Tsitsiklis, 2004)); we argue that $\{\mathrm{SBer}(p_i)\}_{i\in[k]}$ is also hard. First, note that the suboptimality gaps do not change, and thus the lower bound we are trying to prove is the same. Second, any arm pull that generates a sample from the uniform distribution is not helpful; that is, $\mathrm{SBer}(p_i)$ can be thought of as a noisy version of $\mathrm{Ber}(p_i)$. This implies that any algorithm obtains less information from pulling $\mathrm{SBer}(p_i)$ than from pulling $\mathrm{Ber}(p_i)$, and so the optimal sample complexity can only increase. We make this intuition rigorous in Appendix C by adapting the elegant change-of-measure lower-bound argument from (Mannor and Tsitsiklis, 2004), which was originally designed for Bernoulli arms.

6. Conclusion

In this paper, we proposed the Best Arm Identification problem for contaminated bandits (CBAI). This setup can model many practical applications that cannot be modeled by the classical bandit setup. On the way to efficient algorithms for CBAI, we developed tight, non-asymptotic sample-complexity bounds for estimation of the first two robust moments (median and median absolute deviation) from contaminated samples. These results may be of independent interest, perhaps as ingredients for adapting other online learning techniques to similar contaminated settings.

We formalized the contaminated bandit setup as a special case of the more general partially identifiable bandit problem (PIBAI), and presented ways to adapt celebrated Best Arm Identification algorithms for the classical bandit setting to this PIBAI problem. The sample complexity is essentially changed only by replacing the suboptimality "gaps" with suboptimality "effective gaps" to adjust for the challenge of partial identifiability. We then showed how these algorithms apply to the special case of CBAI by making use of the aforementioned guarantees for estimation of robust moments from contaminated samples. Finally, we proved nearly matching information-theoretic lower bounds on the sample complexity of CBAI, showing that (up to a small logarithmic factor) our algorithms are optimal.

Our contributions suggest several potential directions for future work, including the following three.

- We have developed algorithms for online learning in the presence of partial identifiability. How far does this toolkit extend? In particular, do our results apply to more complicated or more general feedback structures such as partial monitoring (see e.g. (Bartók et al., 2014)) or graph feedback (see e.g. (Alon et al., 2017))?

- The contaminated bandit setup models many real-world problems that cannot be modeled by the classical bandit setup. Is it applicable and approachable to formulate other classical online-learning problems in similar contamination setups?

- More abstractly, we think that problems at the intersection of online learning and robust statistics are not only mathematically rich, but also are increasingly relevant

for modern applications given the recent influx of active learning tasks with data that is not completely trustworthy. It may be valuable to use techniques from one of the fields to approach problems in the other, as we did here.

Acknowledgments

We thank Marco Avella Medina and Philippe Rigollet for helpful discussions. JA is supported by NSF Graduate Research Fellowship 1122374.

Appendix A. Properties of $\mathcal{F}_{\bar{t},B}$ and $\mathcal{F}_{\bar{t},B,\bar{m}_2,\kappa}$

The following lemma lists some simple properties of $\mathcal{F}_{\bar{t},B}$ and $\mathcal{F}_{\bar{t},B,\bar{m}_2,\kappa}$, which we use often throughout the paper.

Lemma 28 *If $F \in \mathcal{F}_{\bar{t},B}$, then:*

1. *The distribution $F(a \cdot + b)$ is also in $\mathcal{F}_{B,\bar{t}}$ for any $a \neq 0$.*
2. *F is monotonically increasing in $I_{F,\bar{t}}$.*
3. *$Q_{L,F}(t) = Q_{R,F}(t)$, for all $t \in F(I_{F,\bar{t}}) = [\frac{1}{2} \pm \bar{t}]$.*
4. *$Q_{L,F}(F(x)) = x$ for all $x \in I_{F,\bar{t}}$. Thus in particular $m_1(F)$ is unique and $m_2(F)$ is well-defined, and also we write F^{-1} for $Q_{L,F} = Q_{R,F}$.*
5. *The left and right quantiles of F are equal in the interval $F(I_{F,\bar{t}}) = [\frac{1}{2} - \bar{t}, \frac{1}{2} + \bar{t}]$.*
6. *For any $u_1, u_2 \in F(I_{F,\bar{t}}) = [\frac{1}{2} - \bar{t}, \frac{1}{2} + \bar{t}]$,*

$$|u_1 - u_2| \geq \frac{1}{Bm_2(F)} |F^{-1}(u_1) - F^{-1}(u_2)|.$$

Moreover if also $F \in \mathcal{F}_{\bar{t},B,\bar{m}_2,\kappa}$, then

7. *The distribution $F(a \cdot + b)$ is in $\mathcal{F}_{B,\bar{t},a\bar{m}_2,\kappa}$ for any $a \neq 0$.*
8. *The folded distribution[4] H_F of F satisfies $H_F \in \mathcal{F}_{\bar{t}_H,B_H}$, where $B_H := \kappa B$ and $\bar{t}_H := \min(\frac{1}{2}, \frac{2}{B})$.*
9. *$m_2(F)$ is unique and thus $m_4(F)$ is well-defined.*

Proof When proved in order, all of these statements follow easily from the definition of the function class and the earlier statements. The only part requiring effort is item 8, which we prove presently. Let H be the folded distribution of H; that is the induced distribution of

4. Recall that H_F is the distribution of $|Y - m|$, where $Y \sim F$.

$|Y - m|$ where $Y \sim F$. Define also $B_H := \kappa B$ and $\bar{t}_H := \min(\frac{1}{2}, \frac{2}{B})$. Fix any $r_1, r_2 \in I_{H,\bar{t}_H}$ where without loss of generality $r_1 \geq r_2 \geq 0$. Then

$$\begin{aligned} H(r_1) - H(r_2) &= \mathbb{P}(|Y - m_1(F)| \leq r_1) - \mathbb{P}(|Y - m_1(F)| \leq r_2) \\ &= [\mathbb{P}(Y \leq m_1(F) + r_1) - \mathbb{P}(Y \leq m_1(F) + r_2)] \\ &\quad + [\mathbb{P}(Y < m_1(F) - r_2) - \mathbb{P}(Y < m_1(F) - r_1)] \\ &\geq F(m_1(F) + r_1) - F(m_1(F) + r_2) \\ &\geq \frac{|r_1 - r_2|}{B m_2(F)} \qquad (7) \\ &\geq \frac{|r_1 - r_2|}{\kappa B m_4(F)} \\ &= \frac{|r_1 - r_2|}{B_H m_2(H)}. \end{aligned}$$

The only step requiring justification is the inequality in (7): this is evident by (3) if $r_1, r_2 \in [m_1(F) \pm 2m_2(F)]$, but this condition must be checked. Once we show this condition is met, however, we are immediately done.

Therefore, it is now sufficient to prove $r_1, r_2 \in [m_1(F) \pm 2m_2(F)]$. Since $r_1, r_2 \geq 0$, it suffices to show that the largest value in $I_{H,\bar{t}_H}$, namely $Q_{R,H}(\frac{1}{2} + \bar{t}_H)$, is at most $2m_2(F)$. And to show this, it suffices to show $\frac{1}{2} + \bar{t}_H < H(2m_2(F))$. We show this last inequality presently: by an similar argument as in the first few lines of the above display,

$$\begin{aligned} H(2m_2(F)) - \tfrac{1}{2} &= H(2m_2(F)) - H(m_2(F)) \\ &= [\mathbb{P}(Y \leq m_1(F) + 2m_2(F)) - \mathbb{P}(Y \leq m_1(F) + m_2(F))] \\ &\quad + \mathbb{P}(Y \in [m_1(F) - 2m_2(F), m_1(F) - m_2(F))) \\ &> F(m_1(F) + 2m_2(F)) - F(m_1(F) + m_2(F)) \qquad (8) \\ &\geq \frac{1}{B} \qquad (9) \\ &\geq \bar{t}_H, \end{aligned}$$

where (8) is because F is monotonically increasing in $[m_1(F) - 2m_2(F), m_1(F) + 2m_2(F)]$ by property (i) in the definition of $\mathcal{F}_{\bar{t},B,\bar{m}_2,\kappa}$, and (9) is due to (3). This completes the proof. ■

Appendix B. Deferred Proofs for Algorithms

Throughout this section, we let c be a constant such that $n_{\alpha,\delta} \leq \frac{c}{\alpha^2} \log \frac{1}{\delta}$ for all $\alpha > 0$ and $\delta \in (0, 1)$. Such a constant clearly exists by definition of PIBAI; see (6).

B.1 Adaptation of the Median Elimination Algorithm

The proof of Theorem 19 follows closely the analysis of the orginal Median Elimination Algorithm (see (Even-Dar et al., 2006, Theorem 10)). See the discussion in Section 4.1 for the main difference between the two proofs.

Let us define some notation. Let R denote the number of total rounds in Algorithm 1; note that clearly R is deterministic and $R = O(\log_2 k)$. For each round $r \in [R]$, denote by S_r the set of all arms that are still in S when entering round r; and for each $i \in [S_r]$, denote by $\hat{p}_{i,r}$ the algorithm's estimate of p_i in round r. Finally, for each round $r \in [R]$, denote by $\delta_r = 2^{-r}\delta$ the value of δ in the beginning of the round.

Proof [Proof of Theorem 19] We show (α, δ)-PAC correctness first, and then prove the sample complexity bound afterwards.

Proof of (α, δ)-PAC correctness. Without loss of generality, assume that the best arm is $i^* = 1$. For each round $r \in [R]$, define the event E_r: "At round r of Algorithm 1, arm 1 is eliminated and at least one arm i with $\tilde{\Delta}_i > \alpha$ is kept for the next round". Note that if Algorithm 1 fails (i.e. does not output an α-suboptimal arm), then E_r must occur for some round $r \in [R]$. In other words, if the algorithm fails, then there exists some round $r \geq 1$ for which $1 \in S_r$, $1 \notin S_{r+1}$ and $|N_r| \geq |S_r|/2$, where $N_r := \{i \in S_r : \tilde{\Delta}_i > \alpha, \hat{p}_{i,r} > \hat{p}_{1,r}\}$. Then, the failure probability δ_F of Algorithm 1 is bounded from above by

$$\begin{aligned}
\delta_F &\leq \mathbb{P}\left(\exists r \geq 1, 1 \in S_r \text{ and } |N_r| \geq \frac{|S_r|}{2}\right) \\
&\leq \sum_{r=1}^{R} \mathbb{P}\left(1 \in S_r \text{ and } |N_r| \geq \frac{|S_r|}{2}\right) \\
&= \sum_{r=1}^{R} \mathbb{E}\left[\mathbb{P}\left(|N_r| \geq \frac{|S_r|}{2} \;\middle|\; S_r\right) \mathbb{1}_{1 \in S_r}\right]. \qquad (10)
\end{aligned}$$

Now, let $r \in [R]$ and assume $1 \in S_r$ (so that $\hat{p}_{1,r}$ is well defined). By Markov's inequality,

$$\mathbb{P}\left(|N_r| \geq \frac{|S_r|}{2} \;\middle|\; S_r\right) \leq \frac{2}{|S_r|}\mathbb{E}\left[|N_r| \;\middle|\; S_r\right] = \frac{2}{|S_r|} \sum_{i \in S_r : \tilde{\Delta}_i > \alpha} \mathbb{P}\left(\hat{p}_{i,r} > \hat{p}_{1,r}\right).$$

So consider any $i \in S_r$ such that $\tilde{\Delta}_i > \alpha$. By (6), the definition of $n_{\delta_r/4,\alpha/2}$, and a union bound, we have that both $|\hat{p}_{i,r} - p_i| \leq \frac{\alpha}{2}$ and $|\hat{p}_{1,r} - p_1| \leq \frac{\alpha}{2}$ simultaneously occur with probability at least $1 - \frac{\delta_r}{2}$. Whenever this occurs, $\hat{p}_{i,r} - \hat{p}_{1,r} \leq \left(p_i + \frac{\alpha}{2}\right) - \left(p_1 - \frac{\alpha}{2}\right) < 0$. Thus $\mathbb{P}\left(\hat{p}_{i,r} > \hat{p}_{1,r}\right) \leq \frac{\delta_r}{2}$ for any such i. Combining this with the above two displays yields

$$\delta_F \leq \sum_{r=1}^{R} \delta_r \leq \sum_{r=1}^{\infty} \delta_r = \sum_{r=1}^{\infty} 2^{-r}\delta = \delta,$$

Proof of sample complexity. The total number of arm pulls in Algorithm 1 is equal to $\sum_{r=1}^{R} |S_r| n_{\delta_r/4,\alpha/2}$. Now since $|S_{r+1}| \leq |S_r|/2$ for each round r, thus the sample complexity is bounded above by

$$k \sum_{r=1}^{R} 2^{-r} n_{\delta_r/4,\alpha/2} \leq \frac{4ck}{\alpha^2} \sum_{r=1}^{R} 2^{-r} \log\left(\frac{4}{\delta_r}\right) = \frac{4ck}{\alpha^2} \sum_{r=1}^{R} 2^{-r} \log\left(\frac{2^{r+2}}{\delta}\right) = O\left(\frac{k}{\alpha^2} \log \tfrac{1}{\delta}\right).$$

■

B.2 Adaptation of the Successive Elimination Algorithm

Let us first define some notation. As in the previous subsection, let R denote the number of total rounds in Algorithm 2 before termination; and for each round $r \in [R]$, let S_r denote the set of all arms still in S when entering round r. Denote for succintness also $\delta_r := \frac{6\delta}{\pi^2 k r^2}$.

Proof [Proof of Theorem 20] Without loss of generality, assume that the best arm is $i^* = 1$. Consider the event $E := \{|\hat{p}_{i,r} - p_i| \leq U_i + \alpha_{r,\delta_r},\ \forall r \geq 1,\ \forall i \in S_r\}$. The following inequality is a consequence of a union bound, where we consider virtual estimates $\hat{p}_{i,r}$ for $r \leq R, i \notin S_r$ or $r > R, i \in [k]$ that could be obtained if we continued to pull the eliminated arms indefinitely:

$$\mathbb{P}\left(E^C\right) \leq \sum_{i=1}^{k}\sum_{r=1}^{\infty} \mathbb{P}\left(|\hat{p}_{i,r} - p_i| > U_i + \alpha_{r,\delta_r}\right) \leq \sum_{i=1}^{k}\sum_{r=1}^{\infty} \delta_r \leq \sum_{i=1}^{k}\sum_{r=1}^{\infty} \frac{6\delta}{\pi^2 k r^2} = \delta,$$

where above we have used (6) and the famous Basel identity.

We conclude from the above that E occurs with probability at least $1 - \delta$. Henceforth let us be in the event that E occurs. A simple induction argument shows that $1 \in S_r$ for each round $r \in [R]$, yielding that Algorithm 2 outputs the optimal arm. Indeed, at each round r for which $1 \in S_r$, E guarantees that for all $j \in S_r$,

$$\hat{p}_{1,r} \geq p_1 - U_1 - \alpha_{r,\delta_r} = \tilde{\Delta}_j + p_j + U_j - \alpha_{r,\delta_r} > p_j + U_j - \alpha_{r,\delta_r} \geq \hat{p}_{j,r} - 2\alpha_{r,\delta_r},$$

and so by definition of Algorithm 2, the optimal arm 1 is not eliminated at round r.

Now, still assuming that E occurs, let us bound the sample complexity T. For each $i \in [k]$, denote by T_i the number of times that arm i is pulled. Then clearly $T = \sum_{i=1}^{k} T_i$. Moreover, since arm 1 is never eliminated (proved above), $T \leq 2\sum_{i=2}^{k} T_i$. For each $i \geq 2$, arm i is eliminated no later than the first round r in which $\hat{p}_{i,r} < \hat{p}_{1,r} - 2\alpha_{r,\delta_r}$ which, by E, is satisfied as soon as $p_i + U_i + \alpha_{r,\delta_r} < p_1 - U_1 - 3\alpha_{r,\delta_r}$, i.e., $\tilde{\Delta}_i > 4\alpha_{r,\delta_r}$. We conclude that arm i is eliminated in the first round r where

$$\tilde{\Delta}_i > 4\alpha_{r,\delta_r} = 4\sqrt{\frac{c\log(\frac{\pi^2 k r^2}{6\delta})}{r}}.$$

Now this is granted when $r \leq C\frac{1}{\tilde{\Delta}_i^2}\log\left(\frac{k}{\delta\tilde{\Delta}_i}\right)$ for some universal constant $C > 0$. Hence,

$$T \leq \sum_{j=2}^{k} T_j = O\left(\sum_{j=2}^{k} \frac{1}{\tilde{\Delta}_i^2}\log\left(\frac{k}{\delta\tilde{\Delta}_i}\right)\right).$$

■

Proof [Proof of Theorem 23] The proof is very similar to that of Theorem 20. We only note that at each round $r \geq 0$, if any arm is still in S_r, it has been pulled at least n_r times, where

$$\begin{aligned} n_r &\geq n_0(\delta_1) + (2N+1)r \\ &\geq n_0(\delta_r) + r \\ &\geq n_0(\delta_r) + \frac{2B^2\bar{m}_2^2}{\alpha_{r,\delta_r}^2}\log\frac{s}{\delta_r}, \end{aligned}$$

where we set $s = 2$ in the oblivious and prescient cases, $s = 3$ in the malicious case. Hence, the $(0, \delta)$-PAC guarantee follows easily from the same reasoning as that of Theorem 20, and the sample complexity is only affected by the preliminary draws before the first round of the algorithm.

■

Appendix C. Deferred Proofs for Lower Bounds

In this section, we make the proof sketch in Section 5.1 formal. The proof is broken into two parts. First, we exhibit hard instances for BAI in which the arms all have smoothed Bernoulli distributions. Second, we reduce this instance into a lower bound instance for CBAI.

Throughout, we adopt the standard assumption in the multi-armed bandit literature (see e.g. (Mannor and Tsitsiklis, 2004)) and only consider algorithms where the stopping time T is almost surely finite.

C.1 BAI Lower Bound Using Smoothed Bernoulli Arms

We will follow closely the change-of-measure argument from (Mannor and Tsitsiklis, 2004) which provided the first gap-dependent lower bound for the BAI problem. Their paper exhibited hard instances for BAI using Bernoulli-distributed arms. However, our CBAI reduction will not work with Bernoulli distributions (see proof sketch in Section 5.2), and so here we exhibit hard instances for BAI using instead a distribution that will work for our CBAI reduction, namely the smoothed Bernoulli distribution.

Lemma 29 *There exists a positive constant δ' such that for every $\alpha \in (0, \frac{1}{2})$, every $\delta \in (0, \delta')$, every $p \in [0, \frac{1}{2}]^k$, and every (α, δ)-PAC* BAI *algorithm, there exists an instance of* BAI *with SBer distributions forcing the algorithm to use at the least following number of samples in expectation:*

$$\mathbb{E}[T] \geq \Omega\left(\sum_{i \in [k] \setminus \{i*\}} \frac{1}{\max(\Delta_i, \alpha)^2} \log \tfrac{1}{\delta}\right),$$

where $i^* := \arg\max_{i \in [k]} p_i$, and $\Delta_i := p_{i^*} - p_i$ denotes for each $i \in [k] \setminus \{i^*\}$.

Proof We follow closely the proof of Theorem 5 in (Mannor and Tsitsiklis, 2004) and only describe the main differences. Fix any $\underline{p} \in (0, \frac{1}{2})$; as their paper says, this will only affect the constant in our final lower bound, which is anyways absorbed by our $\Omega(\cdot)$.

We are now ready to begin. The technique is to reduce to a hypothesis-testing lower bound; our hypothesis are just slightly different since we will consider arms with SBer$(\cdot)$ distributions instead of Ber$(\cdot)$ distributions, albeit with the same parameters. Formally, let $p_1 = \max_{i \in [k]} p_i$ without loss of generality, and define the hypotheses as in (Mannor and Tsitsiklis, 2004) by

$$H_0 \ : \ q_i = p_i, \ \ \forall i \in [k],$$

and for $\ell \in [k]$:

$$H_\ell \ : \ q_\ell = p_1 + \alpha, \ q_i = p_i, \ \forall i \neq \ell.$$

If we are in hypothesis H_ℓ, then we will consider the BAI instance with arm distributions $\{\mathrm{SBer}(q_i)\}_{i\in[k]}$. (As opposed to the arm distributions $\{\mathrm{Ber}(q_i)\}_{i\in[k]}$ as in the original paper.) We will henceforth use $\mathbb{E}_\ell[\cdot]$ and $\mathbb{P}_\ell(\cdot)$ to denote the expectation and probability, respectively, with respect to the randomness of algorithm and arms under the measure induced by hypothesis H_ℓ.

Assume for sake of contradiction that there exists some arm $\ell \in [k]$ satisfying $\mathbb{E}_0[T_\ell] \le t^*_\ell$, where t^*_ℓ is defined as in their paper, and the random variable T_ℓ denote the number of times that arm ℓ is pulled. Now a slight deviation from their proof: we define $Z^\ell_1, \dots Z^\ell_{T_\ell}$ to be the sequence of samples obtained from the pulls of arm ℓ; we partition this into the sequences $X^\ell_1, \dots, X^\ell_{T_{\ell,x}}$ and $U^\ell_1, \dots, U^\ell_{T_{\ell,u}}$ of samples that were generated, respectively, from the Bernoulli and uniform sub-distributions in $\mathrm{SBer}(\cdot)$. Define also $K^\ell_t := \sum_{s=1}^t X^\ell_s$ for any $t \le T_{\ell,x}$. Here $T_{\ell,x}$ and $T_{\ell,u}$, respectively, are the number of samples that were generated from each of these sub-distributions; thus in particular $T_{\ell,x} + T_{\ell,u} = T_\ell$, and so obviously

$$T_{\ell,x} \le T_\ell. \tag{11}$$

This seemingly trivial inequality will shortly help us formalize the notion that *injecting uniform samples can only lessen the amount of information the algorithm receives about the different arms' biases.*

The next thing that changes is the most substantial change: the proof of their Lemma 6. The statement of their lemma remains the same for us; the only difference is the likelihood ratio calculation for the change-of-measure argument. Formally, let the random variable W be the history and define the likelihood functions $L_\ell(w) = \mathbb{P}_\ell(W = w)$. Denoting by $\hat{I}$ the algorithm's final decision of best arm, we wish to show that conditional upon the event

$$S_\ell := \{T_{\ell,x} \le 4t^*_\ell\} \cap \{\hat{I} \ne \ell\} \cap \{\max_{1\le t\le 4t^*_\ell} |K^\ell_t - p_\ell t| < \sqrt{t^*_\ell \log(1/\theta)}\},$$

then $\mathbb{P}_\ell(I \ne \ell) > \delta$. By the elegant change-of-measure argument in the end of the proof of their Theorem 5, it suffices to show the following two anti-concentration type inequalities: (i) the following uniform lower bound of the likelihood ratios:

$$\frac{L_\ell(W)}{L_0(W)} 1_{S_\ell} \ge 8\delta \cdot 1_{S_\ell}, \tag{12}$$

and (ii) that

$$\mathbb{P}_0(S_\ell) > 1/8. \tag{13}$$

We first prove (13) since this follows nearly identically as in their paper. We use a union bound over the three events in the definition of S_ℓ. The second and third events occur under H_0 with probability at least $\frac{1}{2}$ and $\frac{7}{8}$, respectively, by identical arguments as in their paper. To lower bound the probability of the first event, we first use the "information inequality" in (11) and then their simple Markov's inequality argument to obtain that $\mathbb{P}_0(T_{\ell,x} \le 4t^*_\ell) \ge \mathbb{P}_0(T_\ell \le 4t^*_\ell) \ge \frac{3}{4}$. Therefore by a union bound, (13) is proven. So it suffices to now just prove (12); we do this presently.

Let the random variable I_t denote the arm that the algorithm pulls at iteration t. Consider any realization of history $w = ((i_1, w_1), \dots, (i_1, w_T))$ where i_t denotes the arm

pulled at iteration t and w_t denotes the corresponding sample. We will denote the history up to time t by $w_{1:t} := ((i_1, w_1), \ldots, (i_t, w_t))$. Then since an algorithm is defined as a random mapping from history to arms, and since also the sample w_t is conditionally independent of all history, we obtain by the chain rule of probability,

$$\begin{aligned} L_\ell(W) &= \left[\prod_{t=1}^{T} \mathbb{P}_{\mathrm{ALG}}(I_t = i_t \mid W_{1:t-1} = w_{1:t-1})\right]\left[\prod_{t=1}^{T} \mathbb{P}(W_t = w_t \mid I_t = i_t,\ W_{1:t-1} = w_{1:t-1})\right] \\ &= \left[\prod_{t=1}^{T} \mathbb{P}_{\mathrm{ALG}}(I_t = i_t \mid W_{1:t-1} = w_{1:t-1})\right]\left[\prod_{t=1}^{T} \mathbb{P}(W_t = w_t \mid I_t = i_t)\right]. \end{aligned}$$

Now every arm besides ℓ has the same distribution in H_0 and H_ℓ. Thus we conclude

$$\frac{L_\ell(W = w)}{L_0(W = w)} = \frac{\prod_{t=1}^{T} \mathbb{P}(W_t = w_t \mid I_t = i_t)}{\prod_{t=1}^{T} \mathbb{P}(W_t = w_t \mid I_t = i_t)}.$$

Now note that under both hypotheses H_0 and H_ℓ, the measure $\mathrm{SBer}(q_\ell)$ is absolutely continuous w.r.t. the measure $(\delta_0 + \delta_1 + \lambda)$, where λ denotes the Lesbegue measure and δ_0 and δ_1 denote Dirac measures at 0 and 1, respectively. Therefore we can compute the Radon-Nikodym derivatives of $\mathrm{SBer}(q)$ w.r.t. $(\delta_0 + \delta_1 + \lambda)$, to obtain the densities $\mu_q(\cdot)$. Continuing from the above display, we conclude that

$$\begin{aligned} \frac{L_\ell(W = w)}{L_0(W = w)} &= \frac{\prod_{t=1}^{T_\ell} \mu_{p_1+\alpha}(Z_t^\ell)}{\prod_{t=1}^{T_\ell} \mu_{p_\ell}(Z_t^\ell)} \\ &= \frac{\prod_{t=1}^{T_\ell} \mu_{p_1+\alpha}(Z_t^\ell)}{\prod_{t=1}^{T_\ell} \mu_{p_\ell}(Z_t^\ell)} \\ &= \frac{\prod_{t=1}^{T_\ell} \left(\left[\frac{1}{2}(p_1+\alpha)\mathbb{1}_{Z_t^\ell=1} + \frac{1}{2}(1-(p_1+\alpha))\mathbb{1}_{Z_t^\ell=0}\right] + \left[\frac{1}{2}\mathbb{1}_{Z_t^\ell\in(0,1)}\right]\right)}{\prod_{t=1}^{T_\ell} \left(\left[\frac{1}{2}p_\ell\mathbb{1}_{Z_t^\ell=1} + \frac{1}{2}(1-p_\ell)\mathbb{1}_{Z_t^\ell=0}\right] + \left[\frac{1}{2}\mathbb{1}_{Z_t^\ell\in(0,1)}\right]\right)} \\ &= \frac{\prod_{t=1}^{T_\ell} \left(\left[\frac{1}{2}(p_1+\alpha)\mathbb{1}_{X_t^\ell=1} + \frac{1}{2}(1-(p_1+\alpha))\mathbb{1}_{X_t^\ell=0}\right] + \left[\frac{1}{2}\mathbb{1}_{Y_t^\ell\in(0,1)}\right]\right)}{\prod_{t=1}^{T_\ell} \left(\left[\frac{1}{2}p_\ell\mathbb{1}_{X_t^\ell=1} + \frac{1}{2}(1-p_\ell)\mathbb{1}_{X_t^\ell=0}\right] + \left[\frac{1}{2}\mathbb{1}_{Y_t^\ell\in(0,1)}\right]\right)} \\ &= \frac{\sum_{S\subseteq[T_\ell]} \left(\prod_{t\in S} a_{p_1+\alpha}(X_t^\ell) \prod_{t\notin S} b(Y_t^\ell)\right)}{\sum_{S\subseteq[T_\ell]} \left(\prod_{t\in S} a_{p_\ell}(X_t^\ell) \prod_{t\notin S} b(Y_t^\ell)\right)}, \end{aligned} \tag{14}$$

where in the last step we have denoted $a_q(X_t^\ell) := \frac{1}{2}(p_1+\alpha)\mathbb{1}_{X_t^\ell=1} + \frac{1}{2}(1-(p_1+\alpha))\mathbb{1}_{X_t^\ell=0}$ for both $q \in \{p_1+\alpha, p_\ell\}$, and also denoted $b(Y_t^\ell) := \frac{1}{2}\mathbb{1}_{Y_t^\ell\in(0,1)}$.

Now define the events H_0^{Ber} and H_ℓ^{Ber} to be exactly those in (Mannor and Tsitsiklis, 2004). That is, they are the same as our hypotheses except that samples are generated from $\mathrm{Ber}(p)$ instead of $\mathrm{SBer}(p)$. Let also $L_0^{\mathrm{Ber}}(\cdot)$ and L_ℓ^{Ber} denote the analogous likelihood functions defined. In pages 632-634 of (Mannor and Tsitsiklis, 2004), it is proved that, conditional on the event S_ℓ, the following inequality holds for any w of length at most T_ℓ

$$\frac{L_\ell^{\mathrm{Ber}}(w)}{L_0^{\mathrm{Ber}}(w)} \geq 8\delta.$$

Now expand the left hand side of the above display exactly identically as we did for the likelihood ratio $\frac{L_\ell(w)}{L_0(w)}$ above. This yields that for each $S \subseteq [T_\ell]$,

$$\frac{\prod_{t\in S} a_{p_1+\alpha}(X_t^\ell)}{\prod_{t\in S} a_{p_\ell}(X_t^\ell)} \geq 8\delta. \tag{15}$$

Now recall that the following simple fact: if $\frac{d_i}{e_i} \geq c$ for all i, then $\frac{\sum_i d_i}{\sum_i e_i} \geq c$. Therefore we conclude by combining (14) and (15), that the inequality (12) is true. This completes the proof. ■

C.2 Proving the Lower Bound for CBAI

In this subsection, we show how to prove the lower bounds for CBAI using the technique sketched in Subsection 5.1. In particular, we will show how to prove Theorem 26 by "CBAI-lifting" the MAB instances we proved were hard in Lemma 29. The proof of Corollary 27 then follows immediately by letting $\alpha \to 0$; or said differently, by realizing that an algorithm that returns the *best* arm with probability at least $1-\delta$, is by definition a $(0,\delta)$-PAC algorithm.

Proof [Proof of Theorem 26] Let δ' be the constant from Lemma 29. Assume for sake of contradiction that there exists some $k \geq 2$, $\delta \in (0,\delta')$, $\alpha \in (0,\frac{1}{6})$, $\varepsilon \in (0,\frac{1}{15})$, $\bar{t} \in (0,\frac{1}{10})$, and (α,δ)-PAC CBAI algorithm $\mathcal{A}$ such that for any CBAI instance with $F_1,\dots,F_k \in \mathcal{F}_{\bar{t},B}$, the algorithm uses

$$o\left(\sum_{i\in[k]\setminus\{i^*\}} \frac{1}{\max(\tilde{\Delta}_i,\alpha)^2}\log\tfrac{1}{\delta}\right) \tag{16}$$

samples in expectation, where $i^* := \arg\max_{i\in[k]} m_1(F_i)$ and $\{\tilde{\Delta}_i\}_{i\in[k]\setminus\{i^*\}}$ are the effective gaps w.r.t. the arm distributions $\{F_i\}_{i\in[k]}$.

Now Lemma 29 asserts the existence of a BAI instance where all arms have $\tilde{F}_i := \mathrm{SBer}(p_i)$ distributions where $p_i \in [\frac{1}{3},\frac{1}{2}+\alpha] \subseteq [\frac{1}{3},\frac{2}{3}]$ and for which any (α,δ)-PAC MAB algorithm must use at least

$$\Omega\left(\sum_{i\in[k]\setminus\{i^*\}} \frac{1}{\max(\Delta_i,\alpha)^2}\log\tfrac{1}{\delta}\right) \tag{17}$$

samples in expectation, where $i^* := \arg\max_{i\in[k]} p_i$ and $\{\Delta_i\}_{i\in[k]\setminus\{i^*\}}$ are the gaps w.r.t. the arm distributions $\{\mathrm{Ber}(p_i)\}_{i\in[k]}$. Without loss of generality, let us assume $1 = \arg\max_{i\in[k]} p_i$, i.e. that arm 1 is the best. At this point we separate into the different adversarial settings, since the lower bounds and thus also the liftings differ.

Lifting for oblivious and prescient adversaries. Define the following distributions:

$$F_1 := \frac{1-2\varepsilon}{2(1-\varepsilon)}\mathrm{Ber}(r) + \frac{1}{2(1-\varepsilon)}\mathrm{Unif}([0,1])$$
$$F_i := \frac{1-2\varepsilon}{2(1-\varepsilon)}\mathrm{Ber}(q) + \frac{1}{2(1-\varepsilon)}\mathrm{Unif}([0,1]), \quad \forall i \in \{2,\dots,k\}$$

where $r := \frac{p_1}{1-2\varepsilon}$ and $q := \frac{p_i - 2\varepsilon}{1-2\varepsilon}$. It is not hard to see that:

$$\begin{aligned}\tilde{F}_1 &:= \mathrm{SBer}(p_1) = (1-\varepsilon)F_1 + \varepsilon\delta_0 \\ \tilde{F}_i &:= \mathrm{SBer}(p_i) = (1-\varepsilon)F_i + \varepsilon\delta_1, \quad \forall i \in \{2,\dots,k\}\end{aligned}$$

In other words, *samples generated from SBer(p_i) are equal in distribution to samples generated from the above contaminated mixture model* of $(1-\varepsilon)F_i$ and ε times a Dirac measure.

Next, a simple calculation shows that $m_1(\tilde{F}_1) = p_1$, $m_1(F_1) = p_1 + \varepsilon$, $m_2(F_1) = \frac{1-\varepsilon}{2}$. And similarly for any $i \in \{2,\dots,k\}$, we have $m_1(\tilde{F}_i) = p_i$, $m_1(F_i) = p_i - \varepsilon$, $m_2(F_i) = \frac{1-\varepsilon}{2}$. Moreover, for all $i \in [k]$, we have that $F_i \in \mathcal{F}_{B,\bar{t}}$ for $B = 4$ and any $\bar{t} < \frac{1}{2(1-\varepsilon)} \min(p_i + \varepsilon, 1 - (p_i + \varepsilon)) \le \frac{1}{2} \cdot \frac{15}{14} \cdot \min(\frac{1}{3}, 1 - (\frac{2}{3} + \frac{1}{15})) = \frac{1}{7}$. Therefore we conclude that for each $i \in [k]$, the change in median between the distribution F_i and the contaminated distribution $\tilde{F}_i$ is equal to

$$|m_1(\tilde{F}_i) - m_1(F_i)| = \varepsilon = B m_2(F_i) \frac{\varepsilon}{2(1-\varepsilon)} = U_{\varepsilon,B,m_2(F_i)}.$$

Therefore we conclude that running the aforementioned (α,δ)-approximate CBAI algorithm $\mathcal{A}$ on the samples obtained from the above BAI instance will result in $\mathcal{A}$ outputting arm $\hat{I}$ satisfying $m_1(F_i) \ge m_1(F_1) - (2U_{\varepsilon,B,\frac{1-\varepsilon}{2}} + \alpha)$, with probability at least $1-\delta$. Whenever this occurs, we have by the above calculations that $p_{\hat{i}} \ge p_1 - \alpha$. Therefore $\mathcal{A}$ returned an α approximate arm with probability at least $1-\delta$, for this hard BAI instance. Comparing the sample complexity of $\mathcal{A}$ in (16) with the lower bound in (17), we conclude the desired contradiction.

Lifting for malicious adversaries. The idea is similar to what we did above for the oblivious and prescient cases. The difference is that malicious adversaries can shift quantiles further (see Corollary 13) and as such we must exhibit a lifting that exactly matches this larger shift. Formally, define the following underlying distributions over the CBAI arms:

$$\begin{aligned}F_1 &:= \tfrac{1}{2}\mathrm{Ber}(p_1 + 2\varepsilon) + \tfrac{1}{2}\mathrm{Unif}([0,1]) \\ F_i &:= \tfrac{1}{2}\mathrm{Ber}(p_i - 2\varepsilon) + \tfrac{1}{2}\mathrm{Unif}([0,1]), \quad \forall i \in \{2,\dots,k\}\end{aligned}$$

We now present the malicious CBAI adversarial strategy. For the optimal arm 1, define the joint distribution J_1 over (Y_1, Z_1, D_1) where $Y_1 \sim F_1$, $Z_1 \sim \delta_0$, and the conditional distribution of D_1 given Y_1 is $\mathrm{Ber}(\varepsilon(\frac{p_1}{2} + \varepsilon)^{-1} \cdot \mathbb{1}(Y_1 = 1))$. Similarly, for each suboptimal arm $i \in \{2,\dots,k\}$, define the joint distribution J_i over (Y_i, Z_i, D_i) where $Y_i \sim F_i$, $Z_i \sim \delta_1$, and the conditional distribution of D_i given Y_i is $\mathrm{Ber}(\varepsilon(\frac{1-p_1}{2} + \varepsilon)^{-1} \cdot \mathbb{1}(Y_i = 0))$. It is simple to see that for each arm $i \in [k]$, the marginals are correct under each J_i. Indeed, a simple conditioning calculation yields:

$$\begin{aligned}\mathbb{P}(D_1 = 1) &= \mathbb{P}(D_1 = 1 | Y_1 = 1)\mathbb{P}(Y_1 = 1) + \mathbb{P}(D_1 = 1 | Y_1 \neq 1)\mathbb{P}(Y_1 \neq 1) \\ &= \varepsilon(\tfrac{p_1}{2} + \varepsilon)^{-1} \cdot \tfrac{1}{2}(p_1 + 2\varepsilon) + 0 \\ &= \varepsilon.\end{aligned}$$

An identical argument shows that the marginal distribution of D_i is equal to $\mathrm{Ber}(\varepsilon)$ also for each suboptimal arm $i \in \{2,\dots,k\}$. Now for each arm $i \in [k]$, denote by C_i the corresponding

contaminated distributions induced by $(1-D_i)Y_i + D_i Z_i$ where $(Y_i, Z_i, D_i) \sim J_i$. It is not hard to see that:

$$\tilde{F}_1 := \mathrm{SBer}(p_1) = C_1$$
$$\tilde{F}_i := \mathrm{SBer}(p_i) = C_i, \quad \forall i \in \{2, \dots, k\}$$

In other words, *samples generated from SBer(p_i) are equal in distribution to the samples generated by the malicious* CBAI *adversary's distribution C_i.*

Next, a simple calculation shows that for the optimal arm, $(F_1)^{-1}(\frac{1}{2}) = p_1 + 2\varepsilon$ and $(\tilde{F}_1)^{-1}(\frac{1}{2}) = (F_1)^{-1}(\frac{1}{2} - \varepsilon) = p_1$. Note further that F_1 has cdf satisfying $F_1(s) = \frac{1}{2}(1 - p_1 - 2\varepsilon) + \frac{1}{2}s$ for all $s \in [0, 1)$, and $F_1(1) = 1$. Thus for any $x_1, x_2 \in [Q_{L,F_1}(\frac{1}{2} - \bar{t}), Q_{R,F_1}(\frac{1}{2} + \bar{t})]$, we have that $|F(x_1) - F(x_2)| = \frac{1}{2}|x_1 - x_2|$ since $[\frac{1}{2} \pm \bar{t}] \subseteq [0.4, 0.6]$ is contained within the interval $[\frac{1}{2}(p_1 + 2\varepsilon), 1 - \frac{1}{2}(p_1 + 2\varepsilon)] \supseteq [\frac{1}{2}(\frac{2}{3} + 2 \cdot \frac{1}{15}), 1 - \frac{1}{2}(\frac{2}{3} + 2 \cdot \frac{1}{15})] = [0.4, 0.6]$. By definition, this implies that $F_1 \in \mathcal{F}_{\bar{t}, B_1}$ where $B_1 m_2(F_1) = 2$. (Note also that B_1 is a finite constant since $p_1 \in (\frac{1}{3}, \frac{2}{3}) \subset (0, 1)$ implies $m_2(F_i) > 0$.) A completely identical calculation similarly shows that each suboptimal arm $i \in \{2, \dots, k\}$ satisfies $(F_i)^{-1}(\frac{1}{2}) = p_i - 2\varepsilon$, $(\tilde{F}_i)^{-1}(\frac{1}{2}) = (F_i)^{-1}(\frac{1}{2} + \varepsilon) = p_i$, and $F_i \in \mathcal{F}_{\bar{t}, B_i}$ where $B_i m_2(F_i) = 2$. Therefore we conclude that for each $i \in [k]$, the difference in medians between the distributions F_i and $\tilde{F}_i$ is equal to

$$|m_1(F_i) - m_1(\tilde{F}_i)| = 2\varepsilon = B_i m_2(F_i)\varepsilon = U^{(\text{MALICIOUS})}_{\varepsilon, B_i, m_2(F_i)},$$

which is exactly equal to the largest possible uncertainty. Therefore we conclude by an identical contradiction argument as in the oblivious/prescient adversary proof above.

■

References

Robin Allesiardo and Raphaël Féraud. Selection of learning experts. In *International Joint Conference on Neural Networks (IJCNN)*, pages 1005–1010. IEEE, 2017.

Robin Allesiardo, Raphaël Féraud, and Odalric-Ambrym Maillard. The non-stationary stochastic multi-armed bandit problem. *International Journal of Data Science and Analytics*, (4):267–283, 2017.

Noga Alon, Nicolo Cesa-Bianchi, Claudio Gentile, Shie Mannor, Yishay Mansour, and Ohad Shamir. Nonstochastic multi-armed bandits with graph-structured feedback. *SIAM Journal on Computing*, 46(6):1785–1826, 2017.

Martin Anthony and Peter L Bartlett. *Neural network learning: Theoretical foundations*. Cambridge University Press, 2009.

Gábor Bartók, Dean P Foster, Dávid Pál, Alexander Rakhlin, and Csaba Szepesvári. Partial monitoring – classification, regret bounds, and algorithms. *Mathematics of Operations Research*, (4):967–997, 2014.

Robert Eric Bechhofer, Jack Kiefer, and Milton Sobel. *Sequential identification and ranking procedures: with special reference to Koopman-Darmois populations*. University of Chicago Press, 1968.

Christian Bontemps, Thierry Magnac, and Eric Maurin. Set identified linear models. *Econometrica*, (3):1129–1155, 2012.

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

Sébastien Bubeck, Nicolo Cesa-Bianchi, and Gábor Lugosi. Bandits with heavy tail. *IEEE Transactions on Information Theory*, (11):7711–7717, 2013.

Moses Charikar, Jacob Steinhardt, and Gregory Valiant. Learning from untrusted data. In *Symposium on Theory of Computing (STOC)*, pages 47–60. ACM, 2017.

Lijie Chen and Jian Li. On the optimal sample complexity for best arm identification. *arXiv preprint arXiv:1511.03774*, 2015.

Yeshwanth Cherapanamjeri, Prateek Jain, and Praneeth Netrapalli. Thresholding based Efficient Outlier Robust PCA. *arXiv preprint arXiv:1702.05571*, 2017.

Herman Chernoff. *Sequential analysis and optimal design*. SIAM, 1972.

Ilias Diakonikolas, Gautam Kamath, Daniel M Kane, Jerry Li, Ankur Moitra, and Alistair Stewart. Robustly learning a gaussian: Getting optimal error, efficiently. In *Symposium on Discrete Algorithms (SODA)*, pages 2683–2702. SIAM, 2018.

Eyal Even-Dar, Shie Mannor, and Yishay Mansour. PAC bounds for multi-armed bandit and markov decision processes. In *Conference on Computational Learning Theory (COLT)*, pages 255–270. Springer, 2002.

Eyal Even-Dar, Shie Mannor, and Yishay Mansour. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of Machine Learning Research (JMLR)*, (Jun):1079–1105, 2006.

Victor Gabillon, Mohammad Ghavamzadeh, and Alessandro Lazaric. Best arm identification: A unified approach to fixed budget and fixed confidence. In *Advances in Neural Information Processing Systems (NIPS)*, pages 3212–3220, 2012.

Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Conference on Learning Theory (COLT)*, pages 998–1027, 2016.

Frank R Hampel. The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, (346):383–393, 1974.

Frank R Hampel, Elvezio M Ronchetti, Peter J Rousseeuw, and Werner A Stahel. *Robust statistics: the approach based on influence functions*. John Wiley & Sons, 2011.

Joel L Horowitz and Charles F Manski. Identification and robustness with contaminated and corrupted data. *Econometrica*, pages 281–302, 1995.

Peter J Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, pages 73–101, 1964.

Kevin Jamieson and Ameet Talwalkar. Non-stochastic best arm identification and hyperparameter optimization. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 240–248, 2016.

Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sébastien Bubeck. lilucb: An optimal exploration algorithm for multi-armed bandits. In *Conference on Learning Theory (COLT)*, pages 423–439, 2014.

Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone. PAC subset selection in stochastic multi-armed bandits. In *International Conference on Machine Learning (ICML)*, pages 655–662, 2012.

Zohar Karnin, Tomer Koren, and Oren Somekh. Almost optimal exploration in multi-armed bandits. In *International Conference on Machine Learning (ICML)*, pages 1238–1246, 2013.

Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best-arm identification in multi-armed bandit models. *The Journal of Machine Learning Research (JMLR)*, (1):1–42, 2016.

Michael Kearns and Ming Li. Learning in the presence of malicious errors. *SIAM Journal on Computing*, (4):807–837, 1993.

Kevin A Lai, Anup B Rao, and Santosh Vempala. Agnostic estimation of mean and covariance. In *Foundations of Computer Science (FOCS), 2016 IEEE 57th Annual Symposium on*, pages 665–674. IEEE, 2016.

Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, (1):4–22, 1985.

Lisha Li, Kevin Jamieson, Giulia DeSalvo, Afshin Rostamizadeh, and Ameet Talwalkar. Hyperband: A novel bandit-based approach to hyperparameter optimization. *arXiv preprint arXiv:1603.06560*, 2016.

Shie Mannor and John N Tsitsiklis. The sample complexity of exploration in the multi-armed bandit problem. *Journal of Machine Learning Research (JMLR)*, (Jun):623–648, 2004.

Charles F Manski. *Identification for prediction and decision*. Harvard University Press, 2009.

Ricardo A Maronna and Víctor J Yohai. Robust estimation of multivariate location and scatter. *Journal of the American Statistical Association*, 1976.

Jacob Marschak and William H Andrews. Random simultaneous equations and the theory of production. *Econometrica*, pages 143–205, 1944.

Joseph P Romano and Azeem M Shaikh. Inference for the identified set in partially identified econometric models. *Econometrica*, (1):169–211, 2010.

Peter J Rousseeuw and Annick M Leroy. *Robust regression and outlier detection*. John Wiley & Sons, 2005.

Yevgeny Seldin and Aleksandrs Slivkins. One practical algorithm for both stochastic and adversarial bandits. In *International Conference on Machine Learning (ICML)*, pages 1287–1295, 2014.

Leslie G Valiant. Learning disjunction of conjunctions. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 560–566, 1985.

Stanley L Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, (309):63–69, 1965.