Journal of Machine Learning Research 20 (2019) 1-38 Submitted 6/18; Revised 2/19; Published 5/19

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: Csaba Szepesvári

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. To address this, we develop tight, non-asymptotic sample complexity bounds for high-probability estimation of the first two robust moments (median and median absolute deviation) from contaminated samples. These concentration inequalities are the main technical contributions of the paper and may be of independent interest. Using 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).

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, 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 (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 are neither i.i.d. nor 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) 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 $G_{i,t}$. The goal is to identify the arm whose true distribution has the highest (or approximately highest) median. The median is the goal rather than the mean since the distributions are not assumed to have finite first moments; and even if they do, the true distributions' means are impossible to estimate from contaminated samples since the contaminations may be arbitrary. 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 (or mean) salary.

The contaminated bandit setup also naturally models many other situations that the classical bandit setup cannot, such as any bandit problem where data may be subject to measurement or recording error with some probability. For example, consider the canonical problem of optimal experiment design where we are measuring drug responses, but in a setting where samples can be corrupted or where test results can be incorrectly measured or recorded. Another example is testing new software features, where yet-unfixed bugs may distort responses but will be fixed before release. More scenarios are discussed in Subsection 1.2 when comparing to other models in previous works, and in Subsection 2.1 after the formal definition of the CBAI problem.

Importantly, the CBAI problem is nontrivially harder than the BAI problem since there are no consistent estimators for statistics (including 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 generalize our study to the more abstract *Partially Identifiable Best Arm Identification* (PIBAI) problem (defined formally in Section 3), 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 (Marschak and Andrews, 1944; Manski, 2009). A canonical problem 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

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. However, developing algorithms for this setup requires overcoming the challenge of partial identifiability of the arms' true distributions. Indeed, it is not hard to show that the adversary's ability to inject arbitrary contaminations can render "similar" underlying distributions indistinguishable, even with access to infinite samples.

We analyze this CBAI problem under three models of the adversary's power: 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. Subsection 2.1 gives formal definitions of these adversarial settings, as well as motivating examples and applications for each.

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 use the statistical results in (i) to develop efficient algorithms for the CBAI problem 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).

We elaborate on each of these below.

Contribution (i). These concentration inequalities are the main technical contributions of the paper and may be of independent interest to 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 also 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.

Contribution (ii). We show that, surprisingly, several classical BAI algorithms are readily adaptable to the PIBAI problem. This suggests a certain inherent robustness of these classical bandit algorithms. We first present these algorithms for an abstract version of the PIBAI problem so that our algorithmic results may be easily transferrable to other application domains with different statistics of interest or different contamination models. We then combine these general results with our statistical results from (i) to obtain PAC algorithms for CBAI, the problem this paper focuses on. 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.

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, starting from work by Bechhofer et al. (1968) and Lai and Robbins (1985). Recent interest from the learning theory community was sparked by the seminal paper by Even-Dar et al. (2002), which proposed several algorithms obtaining instance-adaptive sample complexity bounds. Since then, there has been significant work on the algorithmic side (see work by Kalyanakrishnan et al., 2012; Gabillon et al., 2012; Karnin et al., 2013; Jamieson et al., 2013, 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). The partial monitoring problem, where the learner only knows the loss up to some subset (Bartók et al., 2014), is also loosely similar to the CBAI problem in the sense that both problems feature partial identification. We note, however, that minimizing cumulative

regret is not a reasonable goal in our contamination setup since the contaminations can be arbitrary and even unbounded. (Furthermore, even when boundedness is assumed, minimizing cumulative regret can still be a poor criteria; see below for concrete examples.)

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.

A related contamination model is studied by Seldin and Slivkins (2014); however, they make a boundedness assumption which makes their setup drastically different from ours.[1] Specifically, they consider cumulative regret minimization in a setting where rewards are bounded in $[0, 1]$ and contaminations can be arbitrary. Because of this $[0, 1]$ boundedness, contamination can move the means by at most ε, which is why it is reasonable that they base their algorithms on the contaminated means. However, the best contaminated mean is *not* a reasonable proxy for the best mean without the $[0, 1]$ boundedness assumption. First off, the contaminated distributions may not even have finite first moments. Moreover, when rewards are bounded but only within a large range, the contaminated mean can deviate from the true mean by ε times the size of that range. This can lead to an error bound that is extremely loose—sometimes to the point of being useless for prediction or estimation—compared to the tight bound given by quantiles (see Lemma 1) which our algorithms achieve (see Subsection 4.2).

For instance, consider the adaptive survey example mentioned earlier. There might be a wide range of (true) academic salaries, say between zero and a million dollars. As such, contaminating an ε fraction of the data could move the mean by roughly $10^6 \cdot \varepsilon$; even for a reasonable value of $\varepsilon = 0.1$, this error bound of roughly 10^5 is so large that it may drown out the actual information that the survey was trying to investigate. On the other hand, the median may be moved a significantly smaller amount since the (true) distribution of academic salaries might have reasonably narrow quantiles around the median (for example, being somewhat bell-shaped is sufficient). This performance improvement is intuitively explained by the fact that most academic salaries are not at the extremes of 0 or a million dollars, but rather are fairly regular around the median. An identical phenomenon occurs in the canonical problem of optimal experiment design where we are measuring drug responses but samples can be corrupted or where test results can be incorrectly measured or recorded. Indeed, if the measurements of interests are, say, blood pressure or weight, the values might only be bounded within a large range, but the true distribution might have reasonably narrow quantiles around the median.

We note that the need for robust bandit algorithms is further motivated by the recent work of Jun et al. (2018), which shows that in a similar model, an adversary can make certain classical bandit algorithms perform very poorly by injecting only a small amount of contamination into the observed samples.

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; for examples, see work by Hampel (1974); Maronna and Yohai (1976); Rousseeuw and Leroy

1. Another difference is that our setup and results also work for more powerful types of contaminating adversaries.

(2005); Hampel et al. (2011). Contamination models and malicious errors also have a long history in computer science, including the classical work 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

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 $\mathrm{Unif}([a,b])$. 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 $\{1,\dots,n\}$ by $[n]$ for $n \in \mathbb{N}$. We abbreviate "with high probability" by "w.h.p." and "cumulative distribution function" by "cdf".

Let F be a cdf. 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 need for this technical distinction arises when F is not strictly increasing. We denote the set of medians of F by $m_1(F) := [Q_{L,F}(\frac{1}{2}), Q_{R,F}(\frac{1}{2})]$. When F has a unique median, we overload $m_1(F)$ to be this point rather than a singleton set containing it. For shorthand, we often write $m_1(X)$ for a random variable X to denote $m_1(F)$, where F is the law of X. 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 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.

The empirical median of a (possibly random) sequence $x_1,\dots,x_n \in \mathbb{R}$ is denoted by $\hat{m}_1(x_1,\dots,x_n)$: if n is odd, this is the middle value; and if n is even, it is the average of the middle two values. The empirical MAD $\hat{m}_2(x_1,\dots,x_n)$ is then defined as $\hat{m}_1(|x_1 - \hat{m}_1(x_1,\dots,x_n)|,\dots,|x_n - \hat{m}_1(x_1,\dots,x_n)|)$.

1.4. Outline

Section 2 formally defines the CBAI problem and the power of the adversary, describes several motivating applications and examples, and discusses the primary challenge of the problem: partial identifiability. Section 3 presents our algorithmic results for the abstract setting of best arm identification under partial identifiability (the PIBAI problem). We state these algorithms in this general setting so that they may be easily transferrable later to other application domains with different statistics of interest or different contamination models. Section 4 then specializes to the CBAI problem. In order to implement the aforementioned general-purpose PIBAI algorithms for CBAI, concentration results are needed for estimation of medians given contaminated samples. These statistical results are developed in Subsection 4.1, and then used to derive upper bounds on the sample complexity of our CBAI algorithms in Subsection 4.2. Subsection 4.3 gives matching information-theoretic lower bounds on the sample complexity, showing that our algorithms are optimal

(up to small logarithmic factors). This answers the primary question the paper sets out to solve: understanding the complexity of identifying the arm with best median given contaminated samples. Section 5 then turns to our secondary goal: providing guarantees on the distribution of the selected arm. This goal requires additionally estimating the second robust moment. Subsection 5.1 develops the necessary statistical results, which are then used in Subsection 5.2 provide our algorithmic results. Section 6 concludes and discusses several open problems.

2. Formal Setup

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 are equal in distribution to $(1 - D_{i,t})Y_{i,t} + D_{i,t}Z_{i,t}$, where $D_{i,t} \sim \text{Ber}(\varepsilon)$, $Y_{i,t} \sim F_i$, and $Z_{i,t} \sim G_{i,t}$. 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, which 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; that is,

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

where U is the unavoidable uncertainty term in median estimation that is induced by partial identifiability (see Subsection 2.2 for a discussion of U, and see Subsection 4.1 for an explicit computation of this quantity). Thus, the goal is to find an algorithm achieving the PAC guarantee in (1) with small expected sample complexity T.

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. 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 $D_{i,t} \sim \text{Ber}(\varepsilon)$ 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.** Same as the oblivious adversary except that the contaminations $Z_{i,t}$ may depend on everything else (i.e. $\{Y_{j,s}, D_{j,s}, Z_{j,s}\}_{j\in[k],s\geq 1}$). Formally, 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 $\{Y_{j,s}, D_{j,s}, Z_{j,s}\}_{j\in[k],s\geq 1}$.
- **Malicious adversary.** Same as the prescient adversary except that $Y_{i,t}$ and $D_{i,t}$ do not need to be independent. Formally, 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 $\{Y_{j,s}, D_{j,s}, Z_{j,s}\}_{j\in[k],s\geq 1}$.

We note that for all three of these adversarial settings, we do not require independence between the outputs $(Y_{i,t}, D_{i,t}, Z_{i,t})$ across the arms for each fixed t.

The oblivious adversary is well-motivated, as many real-world problems where contaminations may occur fit naturally into this model. Indeed, this setting encompasses any bandit problem where data may be subject to measurement or recording error with some probability, such as measuring drug responses for clinical trials (Lai and Robbins, 1985), conducting surveys (Martin and Miron, 1992), or testing new software features with yet-unfixed bugs as mentioned in the introduction. The theory we develop extends naturally and with little modification for the prescient and malicious adversaries. Moreover, these different settings allow for the modeling of many other real-world scenarios in which the contamination models are not as simple. For instance, all data may already exist but not yet be released to the learning algorithm. If adversaries have access to all the data from the beginning, they may be able to contaminate samples depending on all the data—this is captured by the prescient setting. The malicious setting includes, for example, scenarios in which adversaries (for example, hackers) can successfully contaminate a given sample depending on the amount of effort they put into it, and the adversaries can try harder to contaminate certain samples depending on their values.

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 "factor of 2"; see Section 4 for a precise statement.

2.2. The Challenge of Partial Identifiability of the Median

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 $\Omega(c)$ 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 $m \in m_1(F)$,*

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

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 $\tilde{m}_L := Q_{L,\tilde{F}}(\frac{1}{2})$ and $\tilde{m}_R := Q_{R,\tilde{F}}(\frac{1}{2})$, respectively. Since every median of $\tilde{F}$ lies within $[\tilde{m}_L, \tilde{m}_R]$, it suffices to show that $\tilde{m}_L \geq Q_{L,F}(\frac{1-2\varepsilon}{2(1-\varepsilon)})$ and $\tilde{m}_R \leq Q_{R,F}(\frac{1}{2(1-\varepsilon)})$. We bound $\tilde{m}_L$ presently, as bounding $\tilde{m}_R$ follows by a similar argument or by simply applying the $\tilde{m}_L$ bound to $F(-\cdot)$. By definition of $\tilde{m}_L$, for all $t > 0$, $\frac{1}{2} \leq \tilde{F}(\tilde{m}_L + t) = (1-\varepsilon)F(\tilde{m}_L + t) + \varepsilon G(\tilde{m}_L + t) \leq (1-\varepsilon)F(\tilde{m}_L + t) + \varepsilon$, where the last step is because $G(\tilde{m}_L + t) \leq 1$ since G is a distribution. Rearranging yields $F(\tilde{m}_L + t) \geq \frac{1-2\varepsilon}{2(1-\varepsilon)}$, which implies that $\tilde{m}_L \geq Q_{L,F}(\frac{1-2\varepsilon}{2(1-\varepsilon)})$. ■

Finally, we note that for some simple contamination models, partial identifiability in CBAI is not a problem for identifying the arm with the best median. For example, if all arms are contaminated with a common distribution, then although the true medians are only partially identifiable, the contaminated medians are ordered in the same way as the true medians, albeit perhaps not strictly. However, for general (arbitrary) contaminations, such an ordering invariance is clearly not guaranteed.

3. Algorithms for Best Arm Identification Under Partial Identifiability

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 certain BAI algorithms for the classical stochastic multi-armed

bandit setting can be adapted with little modification to such partially identified settings. Perhaps surprisingly, this suggests a certain innate robustness of these existing classical BAI algorithms.

We present our algorithms in this section for this slightly more abstract problem of *Partially Identifiable Best-Arm-Identification* problem (PIBAI), which we define formally below. This abstraction allows our algorithmic results to later be transferred easily to other variants of the Best Arm Identification problem with different contamination models or different statistics of interest. We will show in Section 4 how to adapt this to the CBAI problem, the main focus of the paper, after first developing the necessary statistical results.

We now formally define the setup of 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 real-valued measure of the quality of arm i 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[2] from D_i to output an estimate $\hat{p}_i$ satisfying

$$\mathbb{P}\Big(\hat{p}_i \in [p_i \pm (U_i + \alpha)]\Big) \geq 1 - \delta. \tag{2}$$

The PIBAI problem is then precisely the standard fixed-confidence stochastic bandit problem using these distributions where in each iteration t, the algorithm chooses an arm $I_t \in [k]$ and receives a sample from $D_{I_t}(p_{I_t}, G_t)$ 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 $i \neq i^*$ satisfying $p_i + U_i \geq p_{i^*} - U_{i^*}$. 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, the estimator from (ii), and the PIBAI algorithm.

2. 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 $n_{\alpha,\delta} = O(\alpha^{-2} \log \frac{1}{\delta})$ since anyways this is the natural (and optimal) quantity for many estimation problems such as estimating a median from contaminated samples (see Corollaries 14 and 15), or using Chernoff bounds to estimate the mean of a $[0, 1]$-supported distribution for classical stochastic MAB, etc.

for $i \in [k]$ **do**
  Sample arm i for $n_{\alpha/2,\delta/k}$ times and produce estimate $\hat{p}_i$
end
Output $\hat{I} := \max_{i\in[k]} \hat{p}_i$.

Algorithm 1: Simple uniform exploration algorithm for PIBAI.

We now present algorithms for the PIBAI problem. First, in Subsection 3.1, we present a simple algorithm that performs uniform exploration among all arms. Next, in Subsection 3.2, we obtain more refined, instance-adaptive sample complexity bounds by adapting the Successive Elimination algorithm of (Even-Dar et al., 2006). All algorithms use as a blackbox an estimator satisfying the above property (ii) of the PIBAI problem.

3.1. Simple Algorithm

A simple (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. Pseudocode is given in Algorithm 1.

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

Proof By (2) 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,

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

implying that $\hat{I}$ is α-suboptimal. The sample complexity bound is clear. ■

3.2. Instance-adaptive Algorithms

The sample complexity of the simple Algorithm 1 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. In this subsection, we obtain such an instance-adaptive sample complexity by modifying the Successive Elimination algorithm of (Even-Dar et al., 2006), which was originally designed for the classical multi-armed bandit problem.

Pseudocode is given in Algorithm 2. At each round r, Algorithm 2 gets a single new sample from each remaining arm in order to update its estimate $\hat{p}_{i,r}$. Then, for the next round $r+1$, it only keeps the arms i whose estimates $\hat{p}_{i,r}$ are α-close to the best estimate, where the threshold α is updated at each round. The algorithm and analysis are almost identical to the original. As such, proof details are deferred to Appendix C. The main difference is that in the proof of correctness, we show the event $\{|\hat{p}_{i,r} - p_i| \le U_i + \alpha_{r,6\delta/(\pi^2 k r^2)},\ \forall r \in [R],\ \forall i \in S_r\}$ 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.

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

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 3 *Let $\delta \in (0,1)$. With probability at least $1-\delta$, Algorithm 2 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.*

Moreover, as noted in Remark 9 of (Even-Dar et al., 2006), Algorithm 2 is easily modified (by simply terminating 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 $\tilde{\Delta}_i \leq \alpha$.

Remark 4 *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 3 by adapting the EGE algorithm. However, this approach does not work. The same holds for adapting the Median Elimination algorithm of (Even-Dar et al., 2006) and the PRISM algorithm of (Jamieson et al., 2013). The reason for the inadaptability of these algorithms—in contrast to the easy adaptability of the Successive Elimination algorithm above—is that these algorithms heavily rely on the "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 PIBAI *does not have this property since errors propagate from adding the uncertainties U_i in the suboptimality gaps: that is, $(p_{i^*} - U_{i^*}) - (p_j + U_j) \neq [(p_{i^*} - U_{i^*}) - (p_i + U_i)] + [(p_i - U_i) - (p_j + U_j)]$. Because of this, in order to return 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 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.*[3] *We stress that this nuance is not merely a technicality but actually fundamental to the correctness proofs for* PIBAI.

4. Finding the Best Median

The previous section provided algorithms for the general PIBAI problem; in this section, these algorithms are adapted to the special case of the CBAI problem, which is the focus of the paper. This requires an estimator of the median from contaminated samples that concentrates at an exponential rate[4]. Subsection 4.1 develops these statistical results, which are of potential independent interest to the robust statistics community. Subsection 4.2 then combines these results with the results of the previous section to conclude algorithms for CBAI, and gives upper bounds on their sample complexity. Finally, Subsection 4.3 gives information-theoretic lower bounds showing that the sample complexities of these algorithms are optimal (up to a small logarithmic factor).

4.1. Estimation of Median from Contaminated Samples

In this subsection, we develop the statistical results needed to obtain algorithms for the CBAI problem. Specifically, we obtain tight, non-asymptotic sample-complexity bounds for median estimation from contaminated samples. We do this for all three adversarial models for the contamination (oblivious, prescient, and malicious). These results may be of independent interest to the robust statistics community.

The subsection is organized as follows. First, Subsection 4.1.1 studies the concentration of the empirical median under the most general distributional assumptions and provides, for all three adversarial settings, tight upper and lower bounds on the sample complexity of median estimation from contaminated samples. Next, in Subsection 4.1.2 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, 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 emphasizing 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.

Throughout this subsection, we will only consider a single arm, and its true distribution will be denoted by F. For brevity of the main text, all proofs are deferred to Appendix B.

3. In particular, at most $2^{-r}\alpha$ for most of these aforementioned algorithms, since this implies that the final arm $\hat{I}$ is at most $\sum_{r=1}^{\infty} 2^{-r}\alpha = \alpha$-suboptimal.
4. For the formal statement, see (ii) in the definition of PIBAI in Section 3.

4.1.1. General Concentration Results for Median Estimation

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 makes the most general such assumption.

Definition 5 *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*

$$R(t) \geq \max\left\{ Q_{R,F}\left(\frac{1}{2} + t\right) - m, m - Q_{L,F}\left(\frac{1}{2} - t\right) \right\} \tag{3}$$

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

In words, the function R dictates the the maximal deviation from the median that a cdf F can have for all quantiles in a small neighborhood $[\frac{1}{2} - \bar{t}, \frac{1}{2} + \bar{t}]$ around the median. Since R is an arbitrary non-decreasing function, Definition 5 gives the most general possible bound on these quantiles.

Note that if $\varepsilon > \bar{\varepsilon}(\bar{t}) := \frac{2\bar{t}}{1+2\bar{t}}$, it is impossible to control the deviation of the contaminated median from the true median for $F \in \mathcal{H}_{\bar{t},R}$. Indeed, the requirement $\varepsilon \leq \bar{\varepsilon}(\bar{t}) := \frac{2\bar{t}}{1+2\bar{t}}$ 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 with the definition of R in Definition 5 immediately yields the following 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 6 *For any $\bar{t} \in (0, \frac{1}{2})$, any $\varepsilon \in (0, \bar{\varepsilon}(\bar{t}))$, and any non-decreasing function $R : [0, \bar{t}] \to \mathbb{R}_{\geq 0}$,*

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

We now turn to estimation results. At this point it is necessary to distinguish between the three adversarial settings for the contamination. We begin with guarantees for the oblivious and prescient adversarial settings, which—somewhat surprisingly—turn out to have the same rates. 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 $R\left(\frac{\varepsilon}{2(1-\varepsilon)}\right)$ (see Corollary 6) plus an error term that decays quickly with $n^{-1/2}$ rate and has sub-Gaussian tails in δ. This 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 7 *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 number of samples* $n \geq 2\left(\bar{t} - \frac{\varepsilon}{2(1-\varepsilon)}\right)^{-2} \log\frac{2}{\delta}$ *, we have*

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

Note that because the class $\mathcal{H}_{\bar{t},R}$ only assumes control on the $[\frac{1}{2}\pm\bar{t}]$ quantiles, the minimum sample complexity n must grow as $\bar{t}$ approaches $\frac{\varepsilon}{2(1-\varepsilon)}$, since by Lemma 1 this is the largest quantile deviation that the contaminated median can be moved from the true median.

We now turn to malicious adversaries and derive analogous tight, non-asymptotic sample complexity bounds for median estimation from contaminated samples. We show that it possible to obtain estimation accuracy of $R(\varepsilon)$ in this malicious adversarial setting (Lemma 8). Note that this is weaker than the the accuracy of $R(\frac{\varepsilon}{2(1-\varepsilon)})$ obtained above against oblivious and prescient adversaries. However, we also show that this dependency is tight and unavoidable (Lemma 9). Moreover, the $O(\sqrt{\frac{\log 1/\delta}{n}})$ error term in Lemma 8 is tight for the same reason as it was in Lemma 7; see the discussion there. Finally, we remark that the upper bound on ε and the lower bound on the sample complexity n in Lemma 8 are exactly the analogues of the corresponding bounds in Lemma 7; 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 8 *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 number of samples* $n \geq 2\left(\bar{t} - \varepsilon\right)^{-2}\log\frac{3}{\delta}$*,*

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

Lemma 9 *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* F *has unique median and*

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

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

4.1.2. Improved Concentration under Quantile Control

Note that if a cdf F changes little around its median, then the neighboring quantiles are harder to distinguish from each other, and thus it is more difficult to estimate the median. 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 10 *For any $\bar{t} \in (0, \frac{1}{2})$ and $B > 0$, let $\mathcal{F}_{\bar{t},B}$ be the family of distributions F that have a unique median and satisfy*

$$|F(x_1) - F(x_2)| \geq \frac{1}{Bm_2(F)}|x_1 - x_2| \tag{4}$$

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

We make a few remarks about the definition. i) Requiring the right-hand side of Equation 4 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) Distributions in $\mathcal{F}_{\bar{t},B}$ are not required to have densities, nor even be continuous. iv) The family $\mathcal{F}_{\bar{t},B}$ has many natural and expected properties, such as closure under scaling and translation. These properties are gathered in Lemma 26, which is deferred to Appendix A for brevity of the main text.

Remark 11 (Examples) *Most common distributions belong to $\mathcal{F}_{\bar{t},B}$ for some values of the parameters $\bar{t}$ and B. Moreover, by Lemma 26, 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 Gaussian distribution, for any $\bar{t} \in (0, \frac{1}{2})$ and any $B \geq \frac{q_{3/4}}{\phi(q_{1/2+\bar{t}})}$, where ϕ is the standard Gaussian density and q_α is the corresponding α-quantile; ii) the uniform distribution on any interval, for any $\bar{t} \in (0, \frac{1}{2})$ and any $B \geq 4$; and iii) any distribution F with positive density F', for any $\bar{t} \in (0, \frac{1}{2})$ and any $B \geq \left(m_2(F) \min_{x \in [Q_{L,F}(\frac{1}{2}-\bar{t}), Q_{R,F}(\frac{1}{2}+\bar{t})]} F'(x)\right)^{-1}$.*

We now apply the estimation results for $\mathcal{H}_{\bar{t},R}$ to estimation for $\mathcal{F}_{\bar{t},B}$. By Lemma 26, whenever $F \in \mathcal{F}_{\bar{t},B}$, then $F \in \mathcal{H}_{\bar{t},R}$ for $R(t) = tBm_2(F)$. Corollary 6 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 12 *For any $\bar{t} \in (0, \frac{1}{2})$, any $\varepsilon \in (0, \bar{\varepsilon}(\bar{t}))$, and any $F \in \mathcal{F}_{\bar{t},B}$,*

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

Remark 13 (Tightness of Corollary 12) *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 12 is tight.*

Now we turn to median estimation bounds for $\mathcal{F}_{\bar{t},B}$. Since $R(t) = tBm_2(F)$ is clearly Lipschitz, the discussion preceding Lemmas 7 and 8 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 a $n^{-1/2}$ rate and with sub-Gaussian tails in δ, both of which are optimal. The $\bar{t}$ terms in the sample complexity bounds appear for the analogous reasons as in Lemma 7; see the discussion there. For brevity, we omit the proofs of the following corollaries since they follow immediately from Lemmas 7, 8, and 26.

Corollary 14 *Consider the same setup as in Lemma 7, 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 number of samples* $n \geq 2\max\left(\frac{B^2 m_2^2(F)}{E^2}, \left(\bar{t} - \frac{\varepsilon}{2(1-\varepsilon)}\right)^{-2}\right)\log\frac{2}{\delta}$, *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 15 *Consider the same setup as in Lemma 8, 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 number of samples* $n \geq 2\max\left(\frac{B^2 m_2^2(F)}{E^2}, (\bar{t}-\varepsilon)^{-2}\right)\log\frac{3}{\delta}$,

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

In Corollary 15 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 9).

4.2. Algorithms

In this subsection, we apply the algorithms developed in Section 3 for the general problem of PIBAI to the special case of CBAI. In order to implement the estimator required by part (ii) of the definition of PIBAI (see Section 3), we make use of the guarantees proved in Subsection 4.1 for median estimation from contaminated samples. Nearly matching information-theoretic lower bounds are provided in the following subsection and show that the algorithms below have optimal sample complexity (up to a small logarithmic factor).

We will only assume that the arms' distributions are in $\mathcal{F}_{\bar{t},B}$ and have robust second moments uniformly bounded[5] by some $\bar{m}_2$; this will allow us to perform median estimation

5. 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).

from contaminated samples. For simplicity, the parameters of this family are assumed to be known to the algorithm beforehand[6]. Note, however, that it is not necessary to know ε, but only an upper bound ε_0 on ε; indeed ε can be replaced with ε_0 in all our algorithms. It is very standard in the robust statistics literature to assume knowledge of such an upper bound ε_0 on the amount of contamination. The parameter $\bar{t}$ can then be arbitrarily chosen by the algorithms, so long as it is bigger than $\varepsilon_0/(2(1-\varepsilon_0))$ (or ε_0 in the malicious case) and smaller than $1/2$.

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. In Subsection 4.1, we explicitly computed $U_i = U_{\varepsilon,B,m_2(F_i)}$ for oblivious and prescient adversaries (Corollary 14) and $U_i = U^{(\text{MALICIOUS})}_{\varepsilon,B,m_2(F_i)}$ for malicious adversaries (Corollary 15). We emphasize that the strength of the adversarial setting is encapsulated in the corresponding U_i.

First, we discuss how the simple Algorithm 1 yields an (α,δ)-PAC algorithm for CBAI without modification. In this setting, $\hat{p}_i$ is the empirical median of the payoffs of arm i. The quantity $n_{\alpha,\delta}$ is equal to $2\max(\frac{B^2\bar{m}_2^2}{\alpha^2}, (\bar{t} - \frac{\varepsilon}{2(1-\varepsilon)})^{-2}) \log\frac{2}{\delta}$ against oblivious and prescient adversaries (Corollary 14) and is equal to $2\max(\frac{B^2\bar{m}_2^2}{\alpha^2}, (\bar{t}-\varepsilon)^{-2}) \log\frac{3}{\delta}$ against malicious adversaries (Corollary 15). Theorem 2 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 Algorithm 1, immediately yields the following result.

Theorem 16 *Let $F_i \in \mathcal{F}_{\bar{t},B}$ with $m_2(F) \le \bar{m}_2$ for each arm $i \in [k]$, 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 $O(kn_{\alpha/2,\delta/k}) = O\left(\frac{k}{\alpha^2}\log\frac{k}{\delta}\right)$.*

Algorithm 1 does not adapt to the difficulty of the problem instance. A natural approach is to try to apply the Successive Elimination Algorithm to BAI directly. Unfortunately, this algorithm needs concentration inequalities even for small sample sizes, whereas our guarantees are only valid when the sample size is greater than a certain threshold. Hence, we slightly modify the Successive Elimination Algorithm to obtain additional samples in an initial exploration phase; the number of samples required in this initial exploration phase is dictated by our median estimation results form Subsection 4.1. We similarly modify later rounds to take additional samples in order to produce updated median estimates at the desired certainty levels. Pseudocode is given in Algorithm 3.

6. The classes of distributions that we define have two parameters that would be hard or even impossible to estimate in practice, namely B and $\bar{m}_2$. In classical setups in which algorithms need to learn means instead of medians, it is common to assume that there is a second moment σ^2 which is known, or at least bounded by a known constant. This is because in practice σ^2 is hard to estimate from the data. Note that if the distributions F_i are known to have bounded support, $\bar{m}_2$ can be chosen as the size of the domain. It seems that knowing B is the price to pay to go from the classical to the robust setup. Indeed, even in the non-contaminated case, it is well known that the median of a distribution can be estimated at a parametric rate only when the distribution is not too flat around its median. In that case, the estimation rate depends on how "*non-flat*" the distribution is, which is exactly $Bm_2(F)$ in the present setup. In a parametric setup, i.e., when the distributions of the arms $F_i, i = 1,\dots,K$ are known up to some location and/or scale parameters, B would be known since it is translation and scale invariant (see Lemma 26).

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

Algorithm 3: Adaptation of successive elimination algorithm (see Algorithm 2) for CBAI. For oblivious and prescient adversaries, $N := 2(\bar{t} - \frac{\varepsilon}{2(1-\varepsilon)})^{-2}$; and for malicious adversaries $N := 2(\bar{t} - \varepsilon)^{-2}$. In all settings, $\alpha_{r,\delta} := \sqrt{\frac{2B\bar{m}_2^2 \log \frac{3}{\delta}}{r}}$.

Theorem 17 *Let $\delta \in (0,1)$, let $F_i \in \mathcal{F}_{\bar{t},B}$ with $m_2(F) \leq \bar{m}_2$ for each arm $i \in [k]$, and let the adversary be either oblivious, prescient, or malicious. With probability at least $1-\delta$, Algorithm 3 outputs the optimal arm after using at most $\tilde{O}\left(\sum_{i \neq i^*} \left(\frac{1}{\tilde{\Delta}_i^2} + N\right) \log\left(\frac{k}{\delta \tilde{\Delta}_i}\right)\right)$ samples.*

The proof follows from Theorem 3 and is deferred to Appendix C. Like Algorithm 2, Algorithm 3 can be stopped earlier in order to obtain an (α, δ)-PAC guarantee.

4.3. Lower Bounds

Here we provide information-theoretic lower bounds on the sample complexity of the CBAI problem that match, up to small logarithmic factors, the algorithmic upper bounds proved above in Subsection 4.2 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.

Let $i^* := \arg\max_{i \in [k]} m_1(F_i)$ denote the optimal arm. As in Section 3, $\tilde{\Delta}_i := (m_1(F_{i^*}) - U_{i^*}) - (m_1(F_i) + U_i)$ denotes the effective gap for suboptimal arms $i \neq i^*$, and U_i denotes the unavoidable uncertainty term for median estimation. We emphasize that since the power of the adversary is encapsulated in the U_i, and therefore also the effective gaps $\tilde{\Delta}_i$, we can address all three adversarial settings simultaneously by proving a lower bound in terms of the effective gaps.

In Section 3, we argued that the PIBAI problem is impossible when there exists suboptimal arms with non-positive effective gaps; the CBAI problem 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 assumptions), 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 4.2.

Theorem 18 *Consider* CBAI *against an oblivious, prescient, or malicious adversary. There exists a positive constant B such that for any number of arms $k \geq 2$, any confidence level $\delta \in (0, \frac{3}{20})$, any suboptimality level $\alpha \in (0, \frac{1}{6})$, any contamination level $\varepsilon \in (0, \frac{1}{15})$, any regularity level $\bar{t} \in (0, \frac{1}{10})$, and any effective gaps $\{\tilde{\Delta}_i\}_{i\in[k]\setminus i^*} \subset (0, \frac{1}{3})$, there exists a* CBAI *instance with $F_1, \dots, F_k \in \mathcal{F}_{\bar{t},B}$ for which any (α, δ)-PAC algorithm has expected sample complexity*

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

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

Taking the limit as $\alpha \to 0$ in Theorem 18 immediately yields the following lower bound on $(0, \delta)$-PAC algorithms.

Corollary 19 *Consider the same setup as in Theorem 18. There exists a* CBAI *instance $F_1, \dots, F_k \in \mathcal{F}_{\bar{t},B}$ for which any $(0, \delta)$-PAC algorithm has expected sample complexity*

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

Proof Sketches. We now sketch the proofs of Theorem 18 and Corollary 19. Full details are deferred to Appendix D.remove for brevity of the main text.

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 smallest possible median and the other F_i to have the largest possible median, which 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.

Note that the $\{F_i\}$ in the above 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. Ensuring the existence of such a lifting is the main obstacle, due the following technical yet important nuance: most BAI lower bounds are constructed from Bernoulli or Gaussian distributions, neither of which is 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 Equation 4 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 D for details.

The only ingredient remaining in the proof is to prove that there are hard instances for BAI with SBer-distributed arms, which follows easily from Lemma 1 and Remark 5 in Kaufmann et al. (2016).

5. Quality Guarantees for the Selected Arm

In the previous Section 4, we developed algorithms for CBAI and proved that they select the best (or an approximately best) arm with high probability. However, in many applications, it is desirable to also provide guarantees on the quality of the selected arm such as "with probability at least 80%, a new random variable $Y \sim F_{\hat{I}}$ is at least 10."

Such a quality guarantee could be accomplished directly using the machinery developed earlier in Subsection 4.1 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 $55\%, 60\%, \ldots$, etc.) 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 estimable from contaminated samples) for this contamination model setting.

This section is organized as follows. Subsection 5.1 develops these statistical results for estimation of the MAD from contaminated samples. Subsection 5.2 then uses these results to prove that the CBAI algorithms presented earlier in Subsection 4.2—with no additional modifications—can also provide quality guarantees for the selected arm. That is, we show that in addition to outputting the best (or an approximately best) arm, these algorithms also provide quality guarantees to a certain precision "for free" without needing extra samples.

5.1. Estimation of Second Robust Moment from Contaminated Samples

In this subsection, we develop the statistical results needed to obtain quality guarantees for the selected arm in the CBAI problem. Specifically, we consider only a single arm and obtain non-asymptotic sample-complexity bounds for estimation of the second robust moment (the MAD) from contaminated samples. We do this for all three adversarial models for the contamination (oblivious, prescient, and malicious). These results may be of independent interest to the robust statistics community.

The subsection is organized as follows. First, Subsection 5.1.1 introduces a class of distributions for which the MAD is estimable from contaminated samples. The few assumptions we make are common in the statistics literature, and we give multiple examples showing that many common distributions are included in this class. Subsection 5.1.2 then proves finite-sample guarantees for MAD estimation from contaminated samples for this class of distributions and for all three adversarial contamination settings.

5.1.1. A Class of Distributions with Estimable Second Robust Moment

Like the median, the MAD is not fully identifiable from contaminated samples (see the 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 20 *For any $\bar{t} \in (0, \frac{1}{2})$, $B > 0$, $\bar{m}_2 > 0$, and $\kappa \geq 0$, let $\mathcal{F}_{\bar{t},B,\bar{m}_2,\kappa}$ be the family of distributions F with unique robust moments $m_1(F)$, $m_2(F)$, and $m_4(F)$ satisfying:*

(i) (4) *holds for all $x_1, x_2 \in \left[Q_{L,F}(\frac{1}{2} - \bar{t}), Q_{R,F}(\frac{1}{2} + \bar{t})\right] \cup [m_1(F) \pm 2m_2(F)]$.*

(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 4.2); the additional assumption that the arm distributions are in $\mathcal{F}_{\bar{t},B,\bar{m}_2,\kappa}$ allows us to also obtain quantile guarantees on the arm returned by the algorithm (see Section 5.2). ii) Property (iii) of the definition requires a bound on the fourth robust moment, the strength of which is dictated by the parameter κ. 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). iii) Note that distributions in $\mathcal{F}_{\bar{t},B,\bar{m}_2,\kappa}$ are not required to have densities, nor even be continuous. iv) The family $\mathcal{F}_{\bar{t},B,\bar{m}_2,\kappa}$ satisfies some natural and expected properties, such as closure under translation. These properties are gathered and proved in Lemma 26, which is deferred to Appendix A for brevity of the main text.

Remark 21 (Examples) *Many common distributions (e.g. the normal distribution, the uniform distribution, the Cauchy distribution, etc.) belong to $\mathcal{F}_{\bar{t},B,\bar{m}_2,\kappa}$ for some values of the parameters $\bar{t}, B, \bar{m}_2$, and κ. For instance, the uniform distribution on the interval $[a, b]$ is in $\mathcal{F}_{\bar{t},B,\bar{m}_2,\kappa}$ for any $\bar{t} \in (0, \frac{1}{2})$, any $B \geq 4$, any $\kappa \geq 2$, and any $\bar{m}_2 \geq \frac{b-a}{4}$. Another example is that the Cauchy distribution with scale parameter a belongs to $\mathcal{F}_{\bar{t},B,\bar{m}_2,\kappa}$ for any $\bar{t} \in (0, \frac{1}{2})$,*

any $B \geq \pi(1 + \max(1, \tan(\pi\bar{t})^2))$*, and* $\kappa \geq \sqrt{3} - 1$*, and any* $\bar{m}_2 \geq a$*. We recall that a Cauchy distribution with scale parameter* $a > 0$ *has density* $\frac{1}{\pi a} \frac{1}{1 + \left(\frac{x - x_0}{a}\right)^2}, x \in \mathbb{R}$*, where* $x_0 \in \mathbb{R}$ *is a location parameter. For such a distribution, it is easy to check that* $m_1(F) = x_0$*,* $m_2(F) = a$ *and* $m_4(F) = a(\sqrt{3} - 1)$*. Note that neither the mean nor the variance exists for the Cauchy distribution.*

5.1.2. Concentration Results for Estimation of Second Robust Moment

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. For brevity, proofs are deferred to Appendix B.3.

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 has sub-Gaussian tails in δ.

Lemma 22 *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 number of samples* $n \geq 2\max\left(\frac{16\kappa^2 B^2 \bar{m}_2^2}{E^2}, \left(\min(\bar{t}, \frac{1}{B}) - \frac{\varepsilon}{2(1-\varepsilon)}\right)^{-2}\right) \log\frac{4}{\delta}$*,*

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

Similar to median estimation above, it is possible to estimate the MAD even in the malicious adversarial setting.

Lemma 23 *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* $X_i = (1 - D_i)Y_i + D_i Z_i$*. Then, for any confidence level* $\delta > 0$*, error level* $E > 0$*, and number of samples* $n \geq 2\max\left(\frac{16\kappa^2 B^2 \bar{m}_2^2}{E^2}, \left(\min(\bar{t}, \frac{1}{B}) - \varepsilon\right)^{-2}\right) \log\frac{6}{\delta}$*,*

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

5.2. Algorithmic Guarantees

We now apply these statistical results for MAD estimation to show that the CBAI algorithms presented earlier in Subsection 4.2—with no additional modifications—can also provide quality guarantees for the arm they select. That is, we show that in addition to outputting the best (or an approximately best) arm, these algorithms also provide quality guarantees to a certain precision "for free" without needing extra samples.

For simplicity, we will focus on error guarantees for Algorithm 1. A nearly identical (yet technically hairier) argument yields a similar guarantee for our adaptation of the Successive Elimination algorithm for CBAI (Algorithm 3).

We will make the assumption that the arms distributions are in $\mathcal{F}_{\bar{t},B,\bar{m}_2,\kappa}$ so that the MAD estimation results proved above will apply. The parameters of this family are assumed to be known to the algorithm beforehand. Distributions in $\mathcal{F}_{\bar{t},B,\bar{m}_2,\kappa}$ have the following property that their lower tails are controlled by their 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 26, 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 $F_1, \dots, F_k \in \mathcal{F}_{\bar{t},B,\bar{m}_2,\kappa}$. Let $\varepsilon \in (0, \min(\frac{2\bar{t}}{1+2\bar{t}}, \frac{1}{B}))$ and $\bar{U} := U_{\varepsilon,B,\bar{m}_2}$ if the adversary is oblivious or prescient, or let $\varepsilon \in (0, \min(\bar{t}, \frac{1}{B}))$ and $\bar{U} := U^{(\text{MALICIOUS})}_{\varepsilon,B,\bar{m}_2}$ if the adversary is malicious. Consider using Algorithm 1 for* CBAI *exactly as in Theorem 2. 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 $t \in [0,\bar{t}]$,*

$$\mathbb{P}\Big(Y \geq (\hat{m}_1 - tB\hat{m}_2) - \big((\tfrac{1}{2} + 2\kappa tB)\alpha + (1 + (1+2\kappa)Bt)\bar{U}\big)\Big) \geq \tfrac{1}{2} + t - \tfrac{3\delta}{k}.$$

Proof By definition of Algorithm 1, $\hat{I}$ is sampled $n_{\alpha/2,\delta/k} = 2\max(\frac{4B^2\bar{m}_2^2}{\alpha^2}, (\bar{t} - \frac{\varepsilon}{2(1-\varepsilon)})^{-2}) \log\frac{2k}{\delta}$ times against oblivious and prescient adversaries, and $n_{\alpha/2,\delta/k} = 2\max\left(\frac{4B^2\bar{m}_2^2}{\alpha^2}, (\bar{t} - \varepsilon)^{-2}\right)\log\frac{3k}{\delta}$ times against malicious adversaries. Therefore—by Corollary 14 and Lemma 22 for oblivious and prescient adversaries, or similarly by Corollary 15 and Lemma 23 for malicious adversaries—we have by a union bound that with probability at least $1 - \frac{3\delta}{k}$, both of the following inequalities hold:

$$\begin{aligned} m_1(F_{\hat{I}}) &\geq \hat{m}_1 - \bar{U} - \tfrac{\alpha}{2}, \quad \text{and} \\ m_2(F_{\hat{I}}) &\leq \hat{m}_2 + (1+2\kappa)\bar{U} + 2\kappa\alpha. \end{aligned}$$

Whenever this event occurs, we have

$$m_1(F_{\hat{I}}) - tBm_2(F_{\hat{I}}) \geq (\hat{m}_1 - tB\hat{m}_2) - \big((\tfrac{1}{2} + 2\kappa tB)\alpha + (1 + (1+2\kappa)Bt)\bar{U}\big).$$

Applying Lemma 24 and taking a union bound completes the proof. ■

Note 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.

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 estimating the median from contaminated samples. We complemented these results with nearly matching information-theoretic lower bounds on the sample complexity of CBAI, showing that (up to a small logarithmic factor) our algorithms are optimal. This answers the question this paper set out to solve: determining the complexity of finding the arm with highest median given contaminated samples. Finally, we used our statistical results on estimation of the second robust moment from contaminated samples to show that without additional samples, our algorithms can also output quality guarantees on the selected arm.

Our paper suggests several potential directions for future work.

- Our upper and lower bounds on the sample complexity of CBAI are off by a logarithmic factor in the number of arms. It is an interesting open question to close this gap, especially since many of the tricks for doing this in the classical BAI setup do not seem to extend to this partially identifiable setting (see Remark 4). One concrete possibility is to obtain a tighter upper bound by adapting the algorithm of (Jamieson et al., 2014) to the CBAI problem. This would require proving a non-asymptotic version of the Law of the Iterated Logarithm for empirical medians, which would be of independent interest.

- 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, given the recent influx of active learning tasks with data that are 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 are grateful to the anonymous reviewers for their insightful comments. We also 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 $\mathcal{F}_{\bar{t},B}$ (defined in Definition 10) and $\mathcal{F}_{\bar{t},B,\bar{m}_2,\kappa}$ (defined in Definition 20), which we use often throughout the paper.

For shorthand, we denote $I_{F,\bar{t}} := \big[Q_{L,F}(\frac{1}{2}-\bar{t}), Q_{R,F}(\frac{1}{2}+\bar{t})\big]$ for a distribution F and a real number $\bar{t} \in (0, \frac{1}{2})$.

Lemma 26 *If $F \in \mathcal{F}_{\bar{t},B}$, then:*

1. *For any $a \neq 0$ and $b \in \mathbb{R}$, the distribution $F(a \cdot + b)$ is also in $\mathcal{F}_{B,\bar{t}}$.*
2. *F is strictly 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 $x \in I_{F,\bar{t}}$ and 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 $u_1, u_2 \in F(I_{F,\bar{t}}) = [\frac{1}{2}-\bar{t}, \frac{1}{2}+\bar{t}]$,*

$$|u_1 - u_2| \geq \frac{1}{B m_2(F)} |F^{-1}(u_1) - F^{-1}(u_2)|.$$

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

7. *For any $a \neq 0$ and $b \in \mathbb{R}$, the distribution $F(a \cdot + b)$ is in $\mathcal{F}_{B,\bar{t},a\bar{m}_2,\kappa}$.*
8. *Let H be the distribution of $|Y - m|$, where $Y \sim F$. Then $H \in \mathcal{F}_{\bar{t}_H,B_H}$, where $B_H := \kappa B$ and $\bar{t}_H := \min(\frac{1}{2}, \frac{2}{B})$.*

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

now verify. 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

$$\begin{aligned} H(r_1) - H(r_2) &= \mathbb{P}(|Y - m_1(F)| \leq r_1) - \mathbb{P}(|Y - m_1(F)| \leq r_2) \\ &= [\mathbb{P}(Y \leq m_1(F) + r_1) - \mathbb{P}(Y \leq m_1(F) + r_2)] \\ &\quad + [\mathbb{P}(Y < m_1(F) - r_2) - \mathbb{P}(Y < m_1(F) - r_1)] \\ &\geq F(m_1(F) + r_1) - F(m_1(F) + r_2) \\ &\geq \frac{|r_1 - r_2|}{B m_2(F)} \qquad (5) \\ &\geq \frac{|r_1 - r_2|}{\kappa B m_4(F)} \\ &= \frac{|r_1 - r_2|}{B_H m_2(H)}. \end{aligned}$$

The only step requiring justification is the inequality in (5): this is evident by (4) 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, we are immediately done.

Therefore, it is now sufficient to prove that $r_1, r_2 \in [m_1(F) \pm 2m_2(F)]$. 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,

$$\begin{aligned} H(2m_2(F)) - \tfrac{1}{2} &= H(2m_2(F)) - H(m_2(F)) \\ &= [\mathbb{P}(Y \leq m_1(F) + 2m_2(F)) - \mathbb{P}(Y \leq m_1(F) + m_2(F))] \\ &\quad + \mathbb{P}(Y \in [m_1(F) - 2m_2(F), m_1(F) - m_2(F))) \\ &> F(m_1(F) + 2m_2(F)) - F(m_1(F) + m_2(F)) \qquad (6) \\ &\geq \frac{1}{B} \qquad (7) \\ &\geq \bar{t}_H, \end{aligned}$$

where (6) 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 (7) is due to (4). This completes the proof. ■

Appendix B. Proofs for Estimation from Contaminated Samples

Throughout, we denote the indicator random variable for an event $\mathcal{E}$ by $\mathbb{1}\{\mathcal{E}\}$.

B.1. Estimation of Median for Oblivious and Prescient Adversaries

Proof of Lemma 7 For shorthand, let $\hat{m} := \hat{m}_1(X_1, \ldots, X_n)$ and let $m \in m_1(F)$ be any median of F. For each $i \in [n]$, define 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,

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

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

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

where the final inequality is due to (3), 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 $m - \hat{m} < R(\frac{\varepsilon}{2(1-\varepsilon)} + a)$ with probability at least $1 - \frac{\delta}{2}$. The lemma statement follows by a union bound. ■

B.2. Estimation of Median for Malicious Adversaries

We now turn to providing the proofs of Lemma 8 and 9. Below, it will be convenient to denote by $y_{(k)}$ the k-th order statistic of a (possibly random) sequence $y_1, \ldots, y_n \in \mathbb{R}$.

Informally, the proof of Lemma 8 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 27), (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 7 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 9 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 27 *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*

$$y_{(\lfloor \frac{n}{2} \rfloor - s)} \leq \hat{m}_1(x_1, \ldots, x_n) \leq y_{(\lceil \frac{n}{2} \rceil + s)}.$$

Proof of Lemma 8 Define for shorthand $a := \sqrt{\frac{\log(3/\delta)}{2n}}$. By Hoeffding's inequality, the event $E := \{\sum_{i=1}^n D_i \leq (\varepsilon + a)n\}$ occurs with probability at least $\mathbb{P}(E) \geq 1 - \frac{\delta}{3}$. Since $\varepsilon + a \leq \bar{t} < \frac{1}{2}$ by our choice of n, Lemma 27 implies that, whenever E occurs,

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

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

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

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 $1 - \delta$,

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

Whenever this occurs, we have by virtue of (3) that $\sup_{m \in m_1(F)} |\hat{m}_1(X_1, \ldots, X_n) - m| \leq R\left(\varepsilon + 2a\right)$, since both $\frac{1}{2} \pm (\varepsilon + 2a) \in [\frac{1}{2} \pm \bar{t}]$ by our choice of n. ■

Proof of Lemma 9 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 as Equation 3 allows. Next, consider the joint distribution on (D, Y, Z) where $Y \sim F$, the conditional distribution of D given Y is $\mathrm{Ber}(2\varepsilon \cdot \mathbb{1}\{Y \leq 0\})$ and $Z \sim \delta_{Q_{R,F}(\frac{1}{2}+\varepsilon)}$. The marginal of D is easily seen to be correct, since

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

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 random variable

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

Each of these has mean

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

Therefore, by Hoeffding's inequality, we conclude that

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

■

B.3. Estimation of Second Robust Moment

Here, we prove Lemma 22. This is done 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. We begin by proving several helpful lemmas.

Lemma 28 *For any (possibly random) sequence $x_1, \dots, x_n \in \mathbb{R}$ and any $c \in \mathbb{R}$,*

$$\left|\hat{m}_1(|x_1 + c|, \dots, |x_n + c|) - \hat{m}_1(|x_1|, \dots, |x_n|)\right| \le |c|$$

holds almost surely, over the possible randomness of the sequence.

Proof For any fixed vector $x \in \mathbb{R}^n$, the function $c \mapsto \hat{m}_1(|x_1 + c|, \dots, |x_n + c|)$ is 1-Lipschitz since it is the composition of the following 1-Lipschitz functions with respect to the L_∞ norm: adding $c\vec{1}$, taking entrywise absolute values, and taking an order statistic. ■

We also give the following distributional version of Lemma 28.

Lemma 29 *Consider any real-valued random variable X and any $c \in \mathbb{R}$. Assume that $m_1(|X|)$ and $m_1(|X + c|)$ are unique. Then*

$$\left|m_1(|X + c|) - m_1(|X|)\right| \le |c|.$$

Proof Assume without loss of generality that $c \ge 0$; the case $c < 0$ follows by an identical argument or by symmetry. For shorthand, let $m := m_1(|X|)$ and let $\tilde{m} := m_1(|X + c|)$. First, we show that $\tilde{m} \le m + c$, i.e., we show that $\mathbb{P}(|X + c| \le m + c) \ge \frac{1}{2}$. This is straightforward: $\mathbb{P}(|X + c| \le m + c) = \mathbb{P}(-m - 2c \le X \le m) \ge \mathbb{P}(-m \le X \le m) = \mathbb{P}(|X| \le m) \ge \frac{1}{2}$, where we the first step is by non-negativity of m and the last step is by definition of m. A nearly identical argument shows $m \le \tilde{m} + c$, namely $\mathbb{P}(|X| \le \tilde{m} + c) = \mathbb{P}(-\tilde{m} - c \le X \le \tilde{m} + c) = \mathbb{P}(-\tilde{m} \le X \le \tilde{m} + 2c) \ge \mathbb{P}(-\tilde{m} \le X \le \tilde{m}) = \mathbb{P}(|X| \le \tilde{m}) \ge \frac{1}{2}$. We therefore conclude that $|\tilde{m} - m| \le c$. ■

Lemma 30 *For any distribution F with unique $m_2(F)$ and $m_4(F)$,*

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

Proof Let $Y \sim F$. By Lemma 29,

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

■

We are now ready to prove Lemma 22.

Proof of Lemma 22 For shorthand, denote $\hat{m} := \hat{m}_1(X_1, \ldots, X_n)$, and let H denote the distribution of $|Y - m_1(F)|$ where $Y \sim F$. Combining the fact that $m_2(F) = m_1(H)$ with Lemma 28 and the triangle inequality, the MAD estimation error is bounded above by

$$\begin{aligned}
&\left|\hat{m}_2(X_1, \ldots, X_n) - m_2(F)\right| \\
&= \left|\hat{m}_1\left(|X_1 - \hat{m}|, \ldots, |X_n - \hat{m}|\right) - m_1(H)\right| \\
&\leq \left|\hat{m}_1\left(|X_1 - m_1(F)|, \ldots, |X_n - m_1(F)|\right) - m_1(H)\right| + \left|\hat{m} - m_1(F)\right|. \qquad (8)
\end{aligned}$$

By Corollary 14 and our choice of n, the second error term in (8) is bounded above by $U_{\varepsilon,B,m_2(F)} + \frac{E}{2}$ with probability at least $1 - \frac{\delta}{2}$. To control the first error term in (8), we apply Corollary 14 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 26 with Corollary 14, 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 $m_2(H) = m_4(F) \leq 2m_2(F)$ from Lemma 30. 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 the inequality $m_2(H) \leq 2m_2(F)$, this is bounded above by $(1 + 2\kappa)U_{\varepsilon,B,m_2(F)} + E$, as desired. ■

The proof of Lemma 23—the analogous result to Lemma 22 but for the *malicious* adversarial setting—is omitted since it is identical to the proof of Lemma 22 with the uses of Corollary 14 replaced by uses of Corollary 15.

Appendix C. Proofs for Adaptation of the Successive Elimination Algorithm

We first define some notation. 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 Equation 2). 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. For succintness, we also denote $\delta_r := \frac{6\delta}{\pi^2 k r^2}$.

Proof of Theorem 3 Without loss of generality, assume that the best arm is $i^* = 1$. Define the event $E := \{|\hat{p}_{i,r} - p_i| \leq U_i + \alpha_{r,\delta_r}, \forall r \geq 1, \forall i \in S_r\}$. Recall that $\hat{p}_{i,r}$ is the estimate of the measure of quality for arm i after r samples, and that arm i will stop being pulled once $i \notin S_r$. For analysis purposes, consider consider virtual estimates $\hat{p}_{i,r}$ that would be obtained if we continued to pull the eliminated arms indefinitely, so that $\hat{p}_{i,r}$ is defined for all $i \in [k]$ and $r \in \mathbb{N}$. By a union bound,

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

where above we have used (2) and the famous Basel identity.

We conclude from the above that E occurs with probability at least $1 - \delta$. Henceforth let us assume that E occurs. A simple induction argument shows that $1 \in S_r$ for each round

$r \in [R]$, which implies that Algorithm 2 returns 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. Clearly, $T = \sum_{i=1}^k T_i$. Moreover, since arm 1 is never eliminated so long as E occurs, $T \leq 2\sum_{i=2}^k T_i$. 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 $p_i + U_i + \alpha_{r,\delta_r} < p_1 - U_1 - 3\alpha_{r,\delta_r}$, or equivalently $\tilde{\Delta}_i > 4\alpha_{r,\delta_r}$. We conclude that arm i is eliminated in the first round r where

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

which occurs for $r \leq C\frac{1}{\tilde{\Delta}_i^2}\log\left(\frac{k}{\delta\tilde{\Delta}_i}\right)$ 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 of Theorem 17 The proof is nearly identical to that of Theorem 3. The main difference in the proof of correctness is that at each round $r \geq 0$, if any arm is still in S_r, it has been pulled at least

$$\begin{aligned} N\log\left(\frac{\pi^2 k}{2\delta}\right) + 2N\sum_{t=1}^r \log\left(\frac{t+1}{t}\right) + r &= N\log\left(\frac{\pi^2 k(r+1)^2}{2\delta}\right) + r \\ &\geq \left(\frac{2B^2\bar{m}_2^2}{\alpha_{r,\delta_r}^2} + N\right)\log\frac{3}{\delta_r} \\ &\geq \max\left(\frac{2B^2\bar{m}_2^2}{\alpha_{r,\delta_r}^2}, N\right)\log\frac{3}{\delta_r} \end{aligned}$$

times. Thus, we may apply Corollary 14 (for oblivious or prescient adversaries) or Corollary 15 (for malicious adversaries) to obtain the identical guarantees as needed in the proof of Theorem 3 for estimating the median up to accuracy α_{r,δ_r} with probability at least $1 - \delta_r$. Hence, the $(0,\delta)$-PAC guarantee follows easily by identical reasoning. The only difference in the sample complexity proof is that it we must keep track of the additional draws

$$\sum_{i=1}^k \left[\left\lceil N\log\left(\frac{\pi^2 k}{2\delta}\right)\right\rceil + \sum_{t=1}^{T_i}\left\lceil 2N\log\left(\frac{t+1}{t}\right)\right\rceil\right] \leq \sum_{i=1}^k \left[(T_i+1) + N\log\left(\frac{\pi^2 k(T_i+1)^2}{2\delta}\right)\right],$$

where T_i is the number of times that arms i is pulled. By what was shown in Theorem 3, the first term $\sum_{i=1}^{k}(T_i+1) = O(\sum_{i\neq i^*} \frac{1}{\tilde{\Delta}_i^2}\log(\frac{k}{\delta\tilde{\Delta}_i}))$, and the second term $\sum_{i=1}^{k} N\log(\frac{\pi^2 k(T_i+1)^2}{2\delta}) = O(N\sum_{i\neq i^*}\log(\frac{k}{\delta\tilde{\Delta}_i}\log(\frac{k}{\delta\tilde{\Delta}_i}))) = O(N\sum_{i\neq i^*}\log(\frac{k}{\delta\tilde{\Delta}_i}))$. ■

Appendix D. Proofs for Lower Bounds

In this section, we make the proof sketch in Subsection 4.3 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 (Mannor and Tsitsiklis, 2004) that all algorithms we consider have a stopping time T which is almost surely finite.

D.1. BAI Lower Bound

Here, we prove a gap-dependent lower bound for the BAI problem that we will make use of in the following subsection. Mannor and Tsitsiklis (2004) gave the first such lower bound by exhibiting hard instances for BAI using Bernoulli-distributed arms. However, our CBAI reduction will not work with Bernoulli distributions (see the proof sketch in Subsection 4.3), 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.

Recall that we define the smoothed Bernoulli distribution with parameter $p \in (0,1)$, denoted by $\mathrm{SBer}(p)$, to be the uniform mixture of the Bernoulli distribution with parameter p and the uniform distribution on $[0,1]$ with weights $\frac{1}{2}$ and $\frac{1}{2}$. It is easy to see that for $p, q \in (0,1)$, the Kullback-Leibler divergence between $\mathrm{SBer}(p)$ and $\mathrm{SBer}(q)$ is given by

$$\begin{aligned}\mathsf{KL}(\mathrm{SBer}(p), \mathrm{SBer}(q)) &= \frac{1}{2}\mathsf{KL}(\mathrm{Ber}(p), \mathrm{Ber}(q)) \\ &= \frac{1}{2}\left(p\log\left(\frac{p}{q}\right) + (1-p)\log\left(\frac{1-p}{1-q}\right)\right).\end{aligned}$$

In particular, a second-order Taylor expansion yields that, for all $\eta \in (0, \frac{1}{2})$, there exists some positive constant C_η such that the inequality

$$\mathsf{KL}(\mathrm{SBer}(p), \mathrm{SBer}(q)) \leq C_\eta (p-q)^2 \tag{9}$$

holds for all $p, q \in [\eta, 1-\eta]$.

With Equation 9 in hand, the following lemma becomes a direct consequence of Lemma 1 and Remark 5 from Kaufmann et al. (2016).

Lemma 31 *Let $k \geq 2$, $\eta \in (0, \frac{1}{2})$, and $p_1, \ldots, p_k \in [\eta, 1-\eta]$. Consider the instance of* BAI *where arm $i \in [k]$ has distribution SBer(p_i). Then, for any $\alpha > 0$ and any $\delta \in (0, \frac{3}{20})$, any (α,δ)-PAC* BAI *algorithm must use at least the following number of samples*

in expectation:

$$\mathbb{E}[T] \geq \frac{C_\eta}{4}\left(\sum_{i\in[k]\setminus\{i*\}} \frac{1}{\max(\Delta_i,\alpha)^2}\log\left(\frac{1}{2.4\delta}\right)\right),$$

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

D.2. CBAI Lower Bound

In this subsection, we show how to prove the lower bounds for CBAI using the technique sketched in Subsection 4.3. In particular, we will show how to prove Theorem 18 by "CBAI-lifting" the MAB instances we proved were hard in Lemma 31. The proof of Corollary 19 then follows immediately by letting $\alpha \to 0$; since an algorithm that returns the *best* arm with probability at least $1-\delta$, is by definition a $(0,\delta)$-PAC algorithm.

Proof of Theorem 18 Fix any $k \geq 2$, $\delta \in (0, \frac{3}{20})$, $\alpha \in (0, \frac{1}{6})$, $\varepsilon \in (0, \frac{1}{15})$, and $\bar{t} \in (0, \frac{1}{10})$. With $\eta = \frac{1}{3}$, Lemma 31 proves that for any $p_1, \ldots, p_k \in [\frac{1}{3}, \frac{2}{3}]$, the BAI instance with $\tilde{F}_i := \mathrm{SBer}(p_i)$ distributed-arms has the property that any (α, δ)-PAC BAI algorithm must use at least

$$\Omega\left(\sum_{i\in[k]\setminus\{i^*\}} \frac{1}{\max(\Delta_i,\alpha)^2}\log\tfrac{1}{\delta}\right) \tag{10}$$

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 $\{\mathrm{SBer}(p_i)\}_{i\in[k]}$. Without loss of generality, let us assume $1 = \arg\max_{i\in[k]} p_i$ (arm 1 is the best). At this point we now treat the different adversarial settings separately, since the lower bounds and thus also the liftings differ.

D.2.1. Lifting for Oblivious and Prescient Adversaries

Define the following distributions:

$$\begin{aligned} F_1 &:= \frac{1-2\varepsilon}{2(1-\varepsilon)}\mathrm{Ber}(r) + \frac{1}{2(1-\varepsilon)}\mathrm{Unif}([0,1]), \quad \text{and} \\ F_i &:= \frac{1-2\varepsilon}{2(1-\varepsilon)}\mathrm{Ber}(q_i) + \frac{1}{2(1-\varepsilon)}\mathrm{Unif}([0,1]) \quad \forall i \in \{2,\ldots,k\}, \end{aligned}$$

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

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

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 $m_1(\tilde{F}_1) = p_1$, $m_1(F_1) = p_1 + \varepsilon$, and $m_2(F_1) = \frac{1-\varepsilon}{2}$. Similarly, for any $i \in \{2,\ldots,k\}$, we have $m_1(\tilde{F}_i) = p_i$, $m_1(F_i) = p_i - \varepsilon$, and $m_2(F_i) = \frac{1-\varepsilon}{2}$. Moreover, for all $i \in [k]$, we have that $F_i \in \mathcal{F}_{B,\bar{t}}$ for $B = 4$ and any $\bar{t} < \frac{1}{2(1-\varepsilon)}\min(p_i + \varepsilon, 1 -$

$(p_i+\varepsilon)) \le \frac{1}{2}\cdot\frac{15}{14}\cdot\min(\frac{1}{3}, 1-(\frac{2}{3}+\frac{1}{15})) = \frac{1}{7}$. 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

$$|m_1(\tilde{F}_i) - m_1(F_i)| = \varepsilon = Bm_2(F_i)\frac{\varepsilon}{2(1-\varepsilon)} = U_{\varepsilon,B,m_2(F_i)}.$$

Therefore, we conclude that running any (α,δ)-approximate CBAI algorithm $\mathcal{A}$ on the samples obtained from the above BAI instance will result in $\mathcal{A}$ returning 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 event 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. We conclude from the lower bound in (10) that $\mathcal{A}$ must use at least $\Omega\left(\sum_{i\in[k]\setminus\{i^*\}}\frac{1}{\max(\tilde{\Delta}_i,\alpha)^2}\log\frac{1}{\delta}\right)$ samples in expectation, where $\{\tilde{\Delta}_i\}_{i\in[k]\setminus\{i^*\}}$ are the effective gaps w.r.t. the arm distributions $\{F_i\}_{i\in[k]}$.

D.2.2. Lifting for Malicious Adversaries

The idea is similar to the oblivious and prescient case done above. The difference is that malicious adversaries can shift quantiles further (see Corollary 15) and so we must exhibit a lifting that exactly matches this larger shift. Consider the distributions over the CBAI arms

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

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 $\mathrm{Ber}(\varepsilon(\frac{p_1}{2}+\varepsilon)^{-1}\mathbb{1}\{Y_1=1\})$. Similarly, for each suboptimal arm $i\in\{2,\dots,k\}$, define the joint distribution J_i over (Y_i,Z_i,D_i) to be $Y_i\sim F_i$ and $Z_i\sim\delta_1$, and the conditional distribution of D_i given Y_i is $\mathrm{Ber}(\varepsilon(\frac{1-p_1}{2}+\varepsilon)^{-1}\mathbb{1}\{Y_i=0\})$. 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 also equal to $\mathrm{Ber}(\varepsilon)$ 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 $(1-D_i)Y_i+D_iZ_i$ where $(Y_i,Z_i,D_i)\sim J_i$. It is not hard to see that

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

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.*

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 that $F_1(1) = 1$. 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 $F_1 \in \mathcal{F}_{\bar{t},B_1}$ with $B_1 m_2(F_1) = 2$. (Note 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 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 $F_i \in \mathcal{F}_{\bar{t},B_i}$ with $B_i m_2(F_i) = 2$. Therefore, we conclude that for each $i \inf[k]$, 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.An identical reduction argument as in the oblivious and prescient adversary finishes the proof.

■

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 Sebastien Bubeck. On finding the largest mean among many. *arXiv preprint arXiv:1306.3917*, 2013.

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.

Kwang-Sung Jun, Lihong Li, Yuzhe Ma, and Jerry Zhu. Adversarial attacks on stochastic bandits. In *Advances in Neural Information Processing Systems (NIPS)*, pages 3644–3653, 2018.

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)*, 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.

Margaret Martin and Straf Miron. *Principles and practices for a federal statistical agency*. National Academies Press, 1992.

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.