Robustness Guarantees for Mode Estimation with an Application to Bandits

Aldo Pacchiano,[1] **Heinrich Jiang,**[2] **Michael I. Jordan**[1]
[1]UC Berkeley
[2]Google Research
pacchiano@berkeley.edu, heinrichj@google.com, jordan@cs.berkeley.edu

Abstract

Mode estimation is a classical problem in statistics with a wide range of applications in machine learning. Despite this, there is little understanding in its robustness properties under possibly adversarial data contamination. In this paper, we give precise robustness guarantees as well as privacy guarantees under simple randomization. We then introduce a theory for multi-armed bandits where the values are the modes of the reward distributions instead of the mean. We prove regret guarantees for the problems of top arm identification, top m-arms identification, contextual modal bandits, and infinite continuous arms top arm recovery. We show in simulations that our algorithms are robust to perturbation of the arms by adversarial noise sequences, thus rendering modal bandits an attractive choice in situations where the rewards may have outliers or adversarial corruptions.

Introduction

Work in mode estimation has received much attention (e.g. (Parzen 1962; Chernoff 1964; Yamato 1971; Silverman 1981; Tsybakov 1990; Vieu 1996; Dasgupta and Kpotufe 2014)) with practical applications including clustering (Cheng 1995; Sheikh, Khan, and Kanade 2007; Vedaldi and Soatto 2008; Jiang and Kpotufe 2017), control (Madani and Benallegue 2007; Hofbaur and Williams 2002), power systems (Williams, Chung, and Gupta 2001; Sarmadi and Venkatasubramanian 2013), bioinformatics (Hedges and Shah 2003), and computer vision (Yin et al. 2003; Tao, Jin, and Zhang 2007; Collins 2003); however, to the best of our knowledge, little is known about the statistical robustness of mode estimation procedures despite the popularity of mode estimation and the increasing need for robustness in modern data analysis (Dwork, Roth et al. 2014). Such robustness is important if mode-estimation based learning systems need results to be less sensitive to possibly adversarial data corruption. Moreover, data sources may be more likely to release data to the learner if it can be guaranteed for each source that their additional data will not change the final outcome by much– in other words robustness is also intimately tied to another important theme of privacy (Dwork and Lei 2009).

We then provide a new application of mode estimation to the problem of the multi-armed bandits (MAB) (Robbins 1985), called *modal bandits*. MABs have been used extensively in a wide range of practical applications and have been extensively analyzed theoretically. The vast majority of works presume that the value of an arm is the expected value of a reward distribution. In this paper, we present an alternative: where the reward is a function of the modes of the distribution of an arm. This leads to a bandit technique that is more robust and better uses the information from the shape of the arm's distribution as well as other nuances that may be lost with the mean (see Figure 1).

Using the mean of the reward distribution can present serious limitations when the observations are biased, potentially due to adversarial interference. We show quantitatively that whenever this is the case, our mode-based bandit algorithms present an alternative to mean-based ones.

Another situation where modal bandits are useful is when the agent already has samples from the arms, but has only one shot to select an arm to pull. Here, the agent may be more interested in optimizing what is $likely$ to happen rather than the choice that is optimal in expectation. For example, when a risk-averse agent needs to decide between a decision that is likely to have small gains and a decision that has a small chance of high gains but large chance of no effect and prefers the former.

In this paper we assume each arm is a distribution over vectors in $\mathbb{R}^D$ with density f and a set of modes $\mathbb{M} \subset \mathbb{R}^D$. We model the reward of an arm as given by a 'score' function that takes $\mathbb{M}$ as input and outputs a value in $[0, 1]$. Although our results can be extended to other more general definitions and more complex modal behaviors, such as scoring functions depending on the value of the m-th most likely mode or the distance between the smallest and the largest mode, in this paper we focus on the case when scoring functions depend only on the most likely mode. Details of a more general setting involving scoring functions depending on multiple modes is laid out in the Appendix[1]. We proceed to define the notion of *mode* and *score function* we will use to analyze the modal bandit problem.

Definition 0.1 (Mode)**.** *Suppose that f is a density over $\mathbb{R}^D$.*

[1]See https://arxiv.org/abs/2003.02932 for full paper with Appendix.

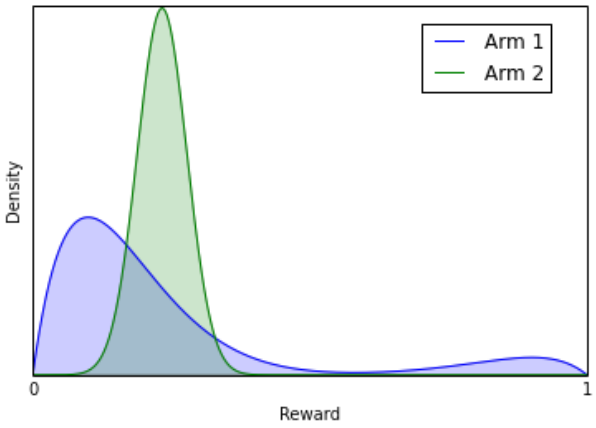

Figure 1: Two distributions with the same mean. Their underlying structure can be quite different.

x is a mode of f if $f(x') \leq f(x)$ for all $x' \in \mathbb{R}^D$.

We focus on the case when the score function takes as input the maximum mode. For density f we denote its maximum density mode as $mode(f)$. Since the later is simply a point in $\mathbb{R}^D$:

Definition 0.2 (Score Function)**.** *A score function on a density with domain $\mathbb{R}^D$ is a map $r : \mathbb{R}^D \to [0, 1]$. The reward of an arm with associated density f equals $r(mode(f))$*

We assume that the rewards are in $[0, 1]$ but it is clear that the results can be extended to any compact interval in $\mathbb{R}$. Definition 0.1 can be relaxed to allow modes be local maximas instead of global maximas and our analysis can handle the case where the density has multiple modes and the reward is a function of these modes. We call this relaxed notion p-modes and provide analogues of our results for p-modes in the Appendix.

Contributions and Related Work

Mode Estimation

(Tsybakov 1990) gave one of the first nonparametric analyses of mode estimation using a kernel density estimator for a unimodal distribution on $\mathbb{R}^D$ and established a lower bound estimation rate of $\tilde{\Omega}(n^{1/(4+D)})$. (Dasgupta and Kpotufe 2014) gave an analysis of the k-nearest neighbor density estimator and provided a procedure based on this density estimator and nearest neighbor graphs which can recover the modes in a multimodal distribution and attained the minimax optimal rate.

In Section , we present Algorithm 1 which finds the highest density mode. In the Appendix we show this algorithm can be adapted to the case when we may care about the p-th highest modes instead. This comes from a simple modification to the mode-seeking algorithms in (Dasgupta and Kpotufe 2014; Jiang and Kpotufe 2017). We then treat the mode estimation procedure as a black box as it works without any a-priori knowledge of the density and only requires mild regularity assumptions on the density.

We build on mode estimation in the following ways. We show that our mode estimation algorithm is statistically robust to certain amounts of adversarial contamination. We then propose and analyze a differentially private mode estimation algorithm. To our knowledge, this is the first time robustness or privacy guarantees have been provided for a mode estimation procedure.

In the Appendix, we analyze the contextual modal bandit. In order to do this, we must estimate the modes of the arm's *conditional* density (conditioned on the context) given samples from the joint density. Thus, we develop a new procedure to estimate the modes of a conditional density given samples from the joint density. We show that it recovers the modes with statistical consistency guarantees and it is practical since it has similar computational complexity to that of Algorithm 1 and again treat it as a black box. Estimating the modes of a conditional density may be of independent interest because a number of nonparametric estimation problems can be formulated in this way (Chen et al. 2016).

Modal Bandits

We then apply mode estimation results to the stochastic MAB problem where the player chooses an arm index $i \in [K]$ which produces a reward from a density $f_i : [0, 1] \to \mathbb{R}^D$ with set of modes $\mathbb{M}_i$ and maximum mode θ_i. The player's objective is to choose the density -henceforth referred to as arm - with optimal modal score $r : \mathbb{R}^D \to \mathbb{R}$. We analyze different problems related to this setup. We start by introducing some results concerning mode estimation in Section . Our contributions also include analogous results to familiar ones in the classical MAB setting.

- First we study top arm identification. We present Algorthm 3, an analogue of the Upper Confidence Bound (UCB) strategy for modal bandits. Theorem 0.7 then shows that we can recover the top-arm given n pulls where n is in terms of the optimality-gap of the arms. We then present an analogue of UCB to recover the top m arms, along with guarantees on recovery of the top m arms.
- Second we introduce two new notions of regret for modal bandits. The first is an analogue of a familiar notion of pseudo-regret from the classical stochastic MAB. The second notion of regret is based on the sample mode estimates, which can be compared to familiar notion of regret computed over sample means. We then attain analogous bounds which are tight up to logarithmic factors.
- Third we introduce contextual modal bandits where the environment samples a context from $\mathbb{R}^d$ from some sampling distribution and is revealed to the learner. We show that a simple uniform sampling strategy can *directly* recover the optimal policy *uniformly* over the context space.

Other Approaches to Robust Bandits

A recent approach of Szorenyi et al. (2015) uses the quantiles of the reward distributions to value the arm. This approach indeed combats some of the limitations described above. Although using the quantiles of the reward distribution is a simple and reasonable approach in many situations

where using the mean fails, using the modes of the reward distribution has properties which are not offered by using the quantiles.

First, unlike quantiles, our method is robust against constant probability noise so as long as this noise is not too concentrated to form a new mode. Second, if the distribution has rewards concentrated around a few regions, this method *adapts* to those regions. In particular, the learner need not know the locations, shapes, or intensity of these regions– no *a priori* information about the density is needed. If one used the quantiles, then there is still the question of which quantile to choose.

In the situation where the reward depends on a *hidden* and *non-stationary* context, the mean and quantile could possibly not even converge while the modes of the reward distribution can remain stable. It is a reasonable assertion that the performance of an arm can depend on the state of the environment which the learner does not have access to. Suppose that the hidden context can take on values H_1 or H_2 sampled by the environment but not revealed to the learner. If the context is H_i, then let the reward be $\mathcal{N}(\mu_i, \sigma^2)$ where $\mathcal{N}$ denotes the normal distribution, $\mu_1 \neq \mu_2$ and $\sigma > 0$. Now suppose that the sampling distribution from which the environment chooses the hidden context is not stationary but can vary over time. In such a situation, both the mean and quantile could change drastically and the estimates of mean or quantile can possibly not converge; moreover in this situation any confidence interval typical in MAB analyses is also rendered meaningless and thus the learner would fail when using mean or quantiles. However, the modes of the reward distribution (μ_1 and μ_2) will not change.

Mode Estimation

Algorithm and Analysis

In this section, we show how to estimate the mode of a distribution given i.i.d. samples. The results are primarily adapted from known results about nonparametric mode estimation (Dasgupta and Kpotufe 2014; Jiang and Kpotufe 2017). The density and mode assumptions are borrowed from (Dasgupta and Kpotufe 2014).

Assumption A1 ***(Modal Structure)*** A local maxima of f is a connected region M such that the density is constant on M and decays around its boundaries. Assume that each local maxima of f is a point, which we call a mode. Let $\mathcal{M}$ be the modes of f, which we assume is finite. Then further assume that f is twice differentiable around a neighborhood of each $x \in \mathcal{M}$ and f has a negative-definite Hessian at each $x \in \mathcal{M}$ and those neighborhoods are pairwise disjoint.

Theorem 0.1. *Suppose Assumption A1 holds and f is a unimodal density. There exists N_f depending on f such that for $n \geq N_f$, setting $k = n^{4/(4+D)}$, we have*

$$\mathbb{P}\left(|\hat{x} - \textit{mode}(f)| \geq \frac{\sqrt{\log(1/\delta)}\log n}{n^{1/(4+D)}}\right) \leq \delta,$$

which matches the optimal rate for mode estimation up to log factors for fixed δ. Where $|\cdot|$ denotes the l_2 norm of $\mathbb{R}^D$.

Algorithm 1 Estimating the mode

Input: k and sample points $X = \{X_1, ..., X_n\}$.
Define $r_k(x) := \inf\{r : |B(x, r) \cap X| \geq k\}$.
Return $\operatorname{argmin}_{x \in X} r_k(x)$

For the rest of the paper, we will assume these choices and thus Algorithm 1 can be treated as a black-box mode estimation procedure. Thus, we define the following notion of sample mode:

Definition 0.3. *For any set S of i.i.d. samples let $\widehat{mode}(S)$ be the estimated mode of S from applying Algorithm 1 under the settings of Theorem 0.1. In particular, the computation of $\widehat{mode}$ on a set of points is understood to be w.r.t. a confidence setting δ.*

Let $r : \mathbb{R}^D \to \mathbb{R}$ be a score function. If r is 1-Lipschitz, the following corollary holds:

Corollary 0.2. *Assuming the same setup as Theorem 0.1, then:*

$$\mathbb{P}\left(|r(\hat{x}) - r(\textit{mode}(f))| \geq \frac{\sqrt{\log(1/\delta)}\log n}{n^{1/(4+D)}}\right) \leq \delta$$

Although all of our results hold for densities over $\mathbb{R}^D$, and L-Lipschitz score functions r, in the spirit of simplicity, in the main paper we mostly discuss the case $D = 1$, score function $r(x) = x$ and density f having domain $[0, 1]$.

Robustness of Mode Estimator

We show that our mode estimation procedure is robust to arbitrary perturbations of the arm's samples. It is already clear that the mode estimates are robust to any perturbation which is sufficiently far away from the mode estimate $\widehat{x}$ and that perturbations don't create high-intensity regions (i.e. there are no samples whose k-NN radius is smaller than that of $\widehat{x}$). In such a situation, it is clear that such perturbations will not change the mode estimator.

The result below provides insight into the situation where the perturbation can be chosen adversarially and in particular when such perturbation can be chosen near the original mode estimate. Specifically, we assume there are ℓ additional points added to the dataset and the result bounds how much the mode estimate can change. We require $\ell < k$, because otherwise, an adversary can place the ℓ points close together anywhere and create a new mode estimate arbitrarily far away from the original mode estimate when using Algorithm 1.

Theorem 0.3 (Robustness)**.** *Suppose that f is a unimodal density with compact support $\mathcal{X} \subseteq \mathbb{R}^D$ and f satisfies Assumption A1. Then there exists constants C, C_1, C_2, depending on f such that the following holds for n sufficiently large depending on f. Let $0 < \delta < 1$ and $\ell > 0$ be the number of samples inserted by an adversary. Let $\hat{x}$ be the mode estimate of Algorithm 1 on n i.i.d. samples drawn from f and $\widetilde{x}$ be the mode estimate by Algorithm 1 on that sample along*

with the ℓ inserted adversarial samples. If k satisfies the following,

$$k \geq C_1 \log(1/\delta)^2 \log n + \ell$$
$$k \leq C_2 \log(1/\delta)^{2D/(4+D)} (\log n)^{D(4+D)} \cdot n^{4/(4+D)}.$$

then with probability at least $1-\delta$, we have

$$|\hat{x} - \widetilde{x}| \leq C\sqrt{\log(1/\delta)} \cdot (\log n)^{1/4} \cdot (k-\ell)^{-1/4}.$$

Proof. Let x_0 be the true mode of f. It suffices to show that for appropriately chosen $\widetilde{r}$, we have

$$\sup_{x \in B(x_0, r_0)} r_k(x) < \inf_{x \notin B(x_0, \widetilde{r})} r_{k-\ell}(x),$$

where $r_k(x)$ is the k-NN radius of any point x and r_0 is the distance of x_0 to the closest sample drawn from f. This is because when inserting ℓ points, the adversary can only decrease the k-NN distance of any point up to its $(k-\ell)$-NN distance. Thus, if we can show that the above holds, then it will imply that $|\hat{x} - \widetilde{x}| \leq \widetilde{r}$.

We have that the above is equivalent to showing the following:

$$\inf_{x \in B(x_0, r_0)} f_k(x) > \sup_{x \notin B(x_0, \widetilde{r})} f_{k-\ell}(x) \cdot \frac{k}{k-\ell},$$

where f_k is the k-NN density estimator. Using k-NN density estimation bounds, we have the following for some constants C_3, C_4:

$$\inf_{x \in B(x_0, r_0)} f_k(x) \geq f(x_0) - \frac{C_3 \cdot \log(1/\delta) \cdot \sqrt{\log n}}{\sqrt{k}},$$
$$\sup_{x \notin B(x_0, \widetilde{r})} f_{k-\ell}(x) \leq f(x_0) - C_4 \left(\widetilde{r}^2 - \frac{\log(1/\delta) \cdot \sqrt{\log n}}{\sqrt{k-\ell}} \right)$$

The result then follows by choosing

$$\widetilde{r}^2 \geq C \frac{\log(1/\delta) \cdot \sqrt{\log n}}{\sqrt{k-\ell}},$$

for appropriate C, as desired. □

Differentially-Private Mode Estimation

In some applications such as healthcare, anonymization of the procedure is necessary and there has been much interest in ensuring such privacy (Dwork et al. 2006). As it stands, Algorithm 1 does not satisfy anonymization since the output is one of the input datapoints. We use the (ϵ, δ)-differential privacy notion of (Dwork et al. 2006) (defined below) and show that a simple modification of our procedure can ensure this notion of privacy.

Definition 0.4 (Differential Privacy)**.** *A randomized mechanism $\mathcal{M} : \mathcal{D} \to \mathcal{R}$ satisfies (ϵ, δ)-differential privacy if any two adjacent inputs $d, d' \in \mathcal{D}$ (i.e. d and d' are sets which differ by at most one datapoint) if the following holds for all $S \subset \mathcal{R}$:*

$$\mathbb{P}(\mathcal{M}(d) \in S) \leq e^{\epsilon} \mathbb{P}(\mathcal{M}(d') \in S) + \delta.$$

To ensure differential privacy, we utilize the Gaussian noise mechanism (see (Dwork et al. 2006)) to the final mode estimate. We now show that this method (Algorithm 2) has differential privacy guarantees.

Algorithm 2 Differentially Private Mode Estimation

Input: k, σ, and sample points $X = \{X_1, ..., X_n\}$.
$\widehat{x} := \text{argmin}_{x \in X} r_k(x)$
Return $\widehat{x} + \mathcal{N}(0, \sigma^2 I)$

Theorem 0.4. *Suppose that f is a unimodal density with compact support $\mathcal{X} \subseteq \mathbb{R}^D$ and f satisfies Assumption A1. Then there exists constants C, C_1, C_2, depending on f such that the following holds for n sufficiently large depending on f. Let $0 < \delta < 1$ and $\epsilon > 0$. Suppose that*

$$\sigma \geq C \log(2/\delta) \cdot (\log n)^{1/4} \cdot k^{1/4} \cdot \epsilon^{-1}.$$

. If k satisfies the following,

$$k \geq C_1 \log(1/\delta)^2 \log n + \ell$$
$$k \leq C_2 \log(1/\delta)^{2D/(4+D)} (\log n)^{D(4+D)} \cdot n^{4/(4+D)}.$$

then with probability at least $1-\delta$, Algorithm 2 is (ϵ, δ)-differentially private.

Remark 0.5. *In particular, we see that taking $k = n^{4/(4+D)}$, we get that $\sigma \approx \log(1/\delta) n^{-1/(4+D)} \cdot \epsilon^{-\epsilon} \to 0$ as $n \to 0$.*

Proof. The result follows by Theorem 1 of (Okada, Fukuchi, and Sakuma 2015) and the global sensitivity of estimating the mode from Theorem 0.3. □

Remark 0.6. *For the remainder of the paper, unless noted otherwise, we assume that we use the mode estimator of Algorithm 1 as a black-box using the settings of Theorem 0.1. It is straightforward to substitute the mode estimation procedure by modify the hyperparameter settings or use a different procedure Algorithm 2 appropriately adjusting the guarantees.*

Top Arm Identification

As common to works in best-arm identification e.g. (Audibert, Bubeck, and Munos 2010; Jamieson and Nowak 2014), we characterize the difficulty of the problem based on the gaps between the value of the arms to that of the optimal arm and the sample complexity can be written in terms of these.

Definition 0.5. *Let f_i denote the density of the i-th arm's reward distribution. Let θ_i be the top mode of f_i where $\theta_1 \geq \theta_2 \geq \cdots \geq \theta_K$. Then we can define the gap between an arm's mode and that of the optimal arm.*

$$\Delta_i := \theta_1 - \theta_i.$$

Although we've indexed the arms this way, it is clear that the algorithms in this paper are invariant to permutations of arms.

We give the Upper Confidence Bound (UCB) strategy (Algorithm 3). For each arm, we maintain a running estimate of the mode as well as a confidence band. Then at each round, we pull the arm with the highest upper confidence bound. When compared to the classical UCB strategy, we replace the running estimates of the mean and confidence

Algorithm 3 UCB Strategy

Input: Total time n and confidence parameter δ.
Define $S_i(t)$ be the rewards observed from arm i up to and include time t.
Let $T_i(t)$ be the number of times arm i was pulled up to and including time t. i.e. $|S_i(t)| = T_i(t)$.

For $t = 1, ..., n$, pull arm I_t, where I_t is the following.

$$\operatorname*{argmax}_{i=1,...,K} \left\{ \widehat{\text{mode}}(S_i(t-1)) + \frac{\log(1/\delta) \cdot \log(T_i(t-1))}{(T_i(t-1))^{1/(4+D)}} \right\}.$$

Algorithm 4 Uniform Sampling Strategy

Input: Total time n and confidence parameter δ.

for $t = 1$ to n **do**
Pull arm (where ties are broken arbitrarily)

$$I_t := \operatorname*{argmin}_{i=1,...,K} \{T_i(t-1)\}.$$

end for
$\widehat{\theta}_i := \widehat{\text{p-mode}}(S_i(n))$ for $i = 1, ..., K$
Return top k arms according to $\widehat{\theta}_i$ value.

band of the mean with the mode and the confidence band of the mode. Our sample complexities now depend on the confidence bands for mode estimation, which converge at a different rate than that of the mean.

We can then give the following result about Algorithm 3's ability to determine the best arm.

Theorem 0.7. *[Top arm identification] Suppose $\theta_1 > \theta_2$. Then there exists universal constants $C_0, C_1 > 0$ such that Algorithm 3 with n timesteps and confidence parameter δ/n satisfies the following. If*

$$n \geq \textit{PolyLog}\left(\frac{1}{\delta}, \sum_{i=2}^{K} \Delta_i^{-(4+D)}\right) \cdot \sum_{i=2}^{K} \Delta_i^{-(4+D)},$$

where PolyLog denotes some polynomial of the logarithms of its arguments,
then

$$\mathbb{P}\left(\operatorname*{argmax}_{i=1,...,K} \left|\{t : I_t = i, 1 \leq t \leq n\}\right| = 1\right) \geq 1 - \delta,$$

where N_{f_i}'s are constants depending on f_i established in Theorem 0.1.

Remark 0.8. *We can compare this to the analogous result for classical MAB (Audibert, Bubeck, and Munos 2010) whose sample complexity (ignoring logarithmic factors) is of order $\sum_{i=2}^{K} \Delta_i^{-2}$ (where the gaps here are w.r.t. the distributional means). Our sample complexity is quintic rather than quadratic in the inverse gaps due to the difficulty of recovering modes compared to recovering means. In fact, for $K = 2$, there exists two distributions such that we require sample complexity at least $\Omega(\Delta_2^{-(4+D)})$ to differentiate between the two distributions. This follows immediately from lower bounds in mode estimation as analyzed in (Tsybakov 1990). Thus, our results are tight up to log factors.*

We next introduce a simple uniform sampling strategy and give a PAC bound to obtain an ϵ-optimal arm (which means its mode is within ϵ of mode of the optimal arm) .

This result can be compared to (Even-Dar, Mannor, and Mansour 2002) for the classical MAB.

Theorem 0.9. *[ϵ-optimal arm identification] Let $\epsilon > 0$. If we run Algorithm 4 with n at least*

$$\max\left\{K(\log(K) + \log(1/\delta))^5 \epsilon^{-5} \log(\epsilon^{-5}), K \max_{i \in [K]} N_{f_i}\right\},$$

then the arm with the highest sample mode is ϵ-optimal with probability at least $1 - \delta$.

Proof. It suffices to choose n large enough such that

$$|\widehat{\text{mode}}(S_i(n)) - \theta_i| \leq \epsilon/2.$$

Indeed, if this were the case, then if arm $i \neq 1$ was selected as the top arm but not ϵ-optimal, then

$$\begin{aligned}\theta_i < \theta_1 - \epsilon &\Rightarrow \theta_i + \epsilon/2 < \theta_1 - \epsilon/2 \\ &\Rightarrow \widehat{\text{mode}}(S_i(n)) < \widehat{\text{mode}}(S_1(n)),\end{aligned}$$

a contradiction. Now from Theorem 0.1 with confidence parameter δ/K, it follows that it suffices to take

$$n \geq K(\log(K) + \log(1/\delta))^5 \epsilon^{-5} \log(\epsilon^{-5}),$$

as desired. □

Regret Analysis

We introduce the following notions of regret based on the modes.

$$\begin{aligned}\mathcal{R}(n) &= n \cdot \max_{i=1,...,K} \theta_i - \sum_{j=1}^{n} \theta_{I_j}, \\ \overline{\mathcal{R}}(n) &= \max_{i=1,...,K} n \cdot \widehat{\text{mode}}\left(\{X_{i,t} : 1 \leq t \leq n\}\right) \\ &\quad - \sum_{i=1}^{K} T_i(n) \cdot \widehat{\text{mode}}(\{X_{i,t} : I_t = i, 1 \leq t \leq n\}).\end{aligned}$$

The regret thus rewards the strategy with the mode ($\mathcal{R}_n$) or the sample mode ($\overline{\mathcal{R}}(n)$) of all trials for a particular arm rather than the mean as in classical formulations.

We next give a regret bounds for Algorithm 3. For $\mathcal{R}(n)$, we attain a poly-logarithmic regret in the number of time steps, while for $\overline{\mathcal{R}}(n)$ we attain a regret of order $\widetilde{O}(n^{4/(4+D)})$. The extra error from the latter is incurred from the errors in the mode estimates.

Theorem 0.10. *Suppose $\theta_1 > \theta_2$. Then with probability at least $1 - \delta$, the regret of Algorithm 3 with n time steps and confidence parameter δ/n satisfies*

$$\begin{aligned}\mathcal{R}(n) &\leq \textit{PolyLog}\left(\frac{1}{\delta}, n\right) \cdot \sum_{i=2}^{K} \Delta_i^{-(3+D)} \\ \overline{\mathcal{R}}(n) &\leq \mathcal{R}(n) + O\left(\left(\textit{PolyLog}\left(\frac{1}{\delta}, n\right) + K\right) \cdot n^{\frac{3+D}{4+D}}\right).\end{aligned}$$

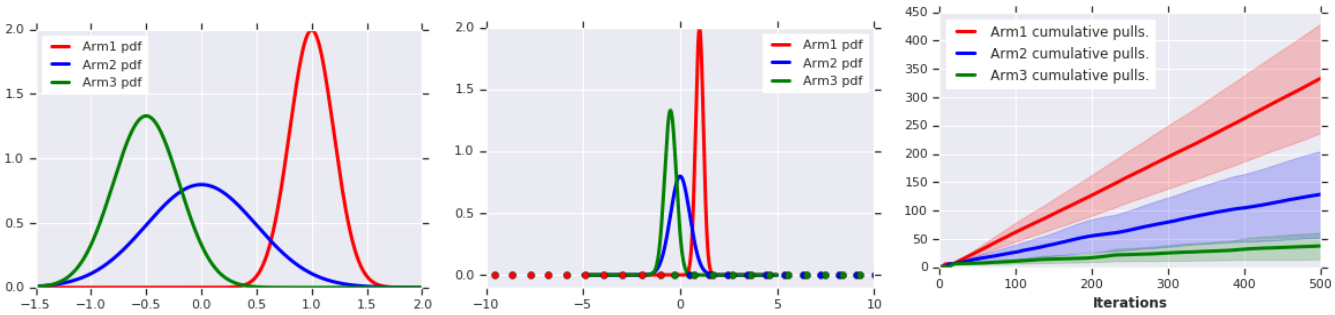

Figure 2: Left, three arms. Center, perturbations. Right, Algorithm 3 cumulative arm pulls.

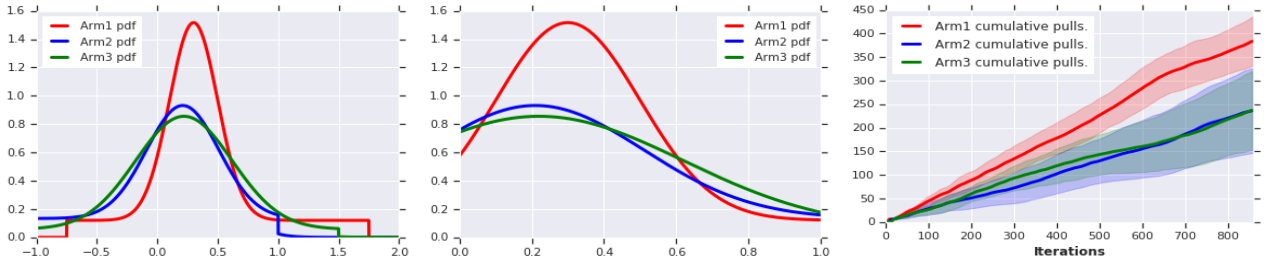

Figure 3: Left, three arms. Center, zoomed in view. Right, algorithm 3's cumulative arm pulls.

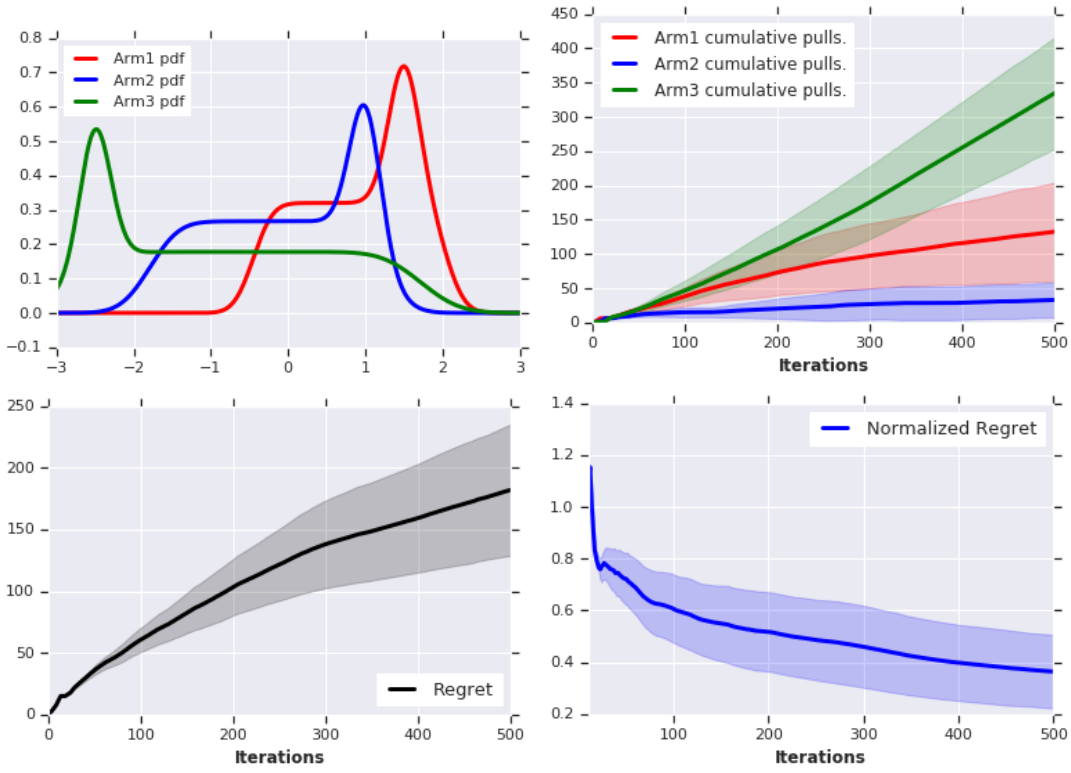

Figure 4: Upper Left, three arm densities. Upper Right, Arm pulls. Lower left, Regret. Lower right, Normalized regret.

Dataset	Mean	Median	Mode
Australian	1.22	1.32	0
Banknote	0.04	0.05	0
Blood	10.6	0.37	0
Electricity	.0017	.0106	.0011
Steel	14360	10596	6831

Table 1: Robustness results under random removals on benchmark OpenML datasets. We compare our k-NN based mode other approaches such as mean and element-wise median. For each of the datasets, we use $k = 100$ for the mode estimation and show the L_2 distance between the mean/median/mode of the features vs if randomly removed 5% of the datapoints. Results are averaged across 20 runs.

Dataset	Mean	Median	Mode
Australian	.603	.483	0
Banknote	.018	.023	0
Blood	5.82	0.28	0
Electricity	.0009	.0085	.0006
Steel	5564	5626	2291

Table 2: Robustness results under random corruption on benchmark OpenML datasets. We show results if we instead randomly choose 1% of the datapoints and add Gaussian noise, where the noise is centered at 0 and the variance for the coordinate is 10 times the variance for that coordinate in the training set.

Remark 0.11. *We can compare this result for $\mathcal{R}(n)$ to that of the classical notion of pseudo-regret, defined below, which also achieves logarithmic regret.*

$$n \max_{i=1,\dots,K} \mu_i - \sum_{j=1}^{n} E[\mu_{I_j}],$$

where μ_i is the mean of the i-th arm's reward distribution.

Experiments

Robustness. In Figure 2, we test the robustness of Algorithm 3 to perturbations of the arms. We consider the case when the score function equals the identity. The red (Arm 1) density's mode has the largest reward value. With probability 0.2 we receive a sample from a noise sequence denoted by the marked points on the x-axis. The colors of these points correspond to which arm we perturb.

Based on the reward distribution given in Figure 2, Algorithm 3 pulls Arm 1 (the arm with the highest modal score) more often despite the perturbations experienced by this arm being negative and the perturbations of the remaining arms being positive values. We average over 25 random seeds and mark the standard deviation.

We also include some results regarding the robustness of our algorithms under random removals on the OpenML benchmark datasets. The methodology and results are described in Tables 1 and 2.

Fine-grained Sensitivity. In Figure 3, we show Algorithm 3 distinguishes between arms with very close modes. We again consider the identity score function. The red (Arm 1) density's mode has the largest modal reward. We average over 25 random seeds and mark the standard deviation.

Finding arms with furthest mode. In Figure 4, we show Algorithm 3 works with score functions other than the identity. In this case we show it can find the arm whose highest density mode is furthest away from the origin– that is the score function equals the distance of the arm's most likely mode to the origin. In this setup Arm 3 is optimal. We also plot the Regret and Normalized Regret (we divide the regret by the iteration index) using the distance from the origin to the arm's mode as reward.

The plot shows Algorithm 3 learns to choose the arm with outlier behavior and does so in a way minimizing the regret captured by differences in outlier score. We average over 25 seeds and mark the standard deviation.

Conclusion

In this paper, we've provided two contributions which are of independent interest: (i) robustness and privacy guarantees for mode estimation and (ii) a new application of mode estimation the bandit problem, which we call *modal bandits*. To our knowledge, we give the first robustness and privacy guarantees for mode estimation, a popular practical method with a long history of theoretical analysis. We then give an extensive analysis of the modal bandits problems, including best-arm identification and regret bound. Additionally, results for contextual modal bandits, and infinite armed bandits can be found in the appendix. We include simulations showing that modal bandits indeed can provide robustness to adversarial corruption, thus suggesting that modal bandits can be an attractive choice in settings where robustness is important.

References

Audibert, J.-Y.; Bubeck, S.; and Munos, R. 2010. Best arm identification in multi-armed bandits. In *COLT*, 41–53.

Chen, Y.-C.; Genovese, C. R.; Tibshirani, R. J.; Wasserman, L.; et al. 2016. Nonparametric modal regression. *The Annals of Statistics* 44(2): 489–514.

Cheng, Y. 1995. Mean shift, mode seeking, and clustering. *IEEE transactions on pattern analysis and machine intelligence* 17(8): 790–799.

Chernoff, H. 1964. Estimation of the mode. *Annals of the Institute of Statistical Mathematics* 16(1): 31–41.

Collins, R. T. 2003. Mean-shift blob tracking through scale space. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, volume 2, II–234. IEEE.

Dasgupta, S.; and Kpotufe, S. 2014. Optimal rates for k-NN density and mode estimation. In *Advances in Neural Information Processing Systems*, 2555–2563.

Dwork, C.; Kenthapadi, K.; McSherry, F.; Mironov, I.; and Naor, M. 2006. Our data, ourselves: Privacy via distributed noise generation. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, 486–503. Springer.

Dwork, C.; and Lei, J. 2009. Differential privacy and robust statistics. In *STOC*, volume 9, 371–380.

Dwork, C.; Roth, A.; et al. 2014. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science* 9(3–4): 211–407.

Even-Dar, E.; Mannor, S.; and Mansour, Y. 2002. PAC bounds for multi-armed bandit and Markov decision processes. In *International Conference on Computational Learning Theory*, 255–270. Springer.

Hedges, S. B.; and Shah, P. 2003. Comparison of mode estimation methods and application in molecular clock analysis. *BMC bioinformatics* 4(1): 1–11.

Hofbaur, M. W.; and Williams, B. C. 2002. Mode estimation of probabilistic hybrid systems. In *International Workshop on Hybrid Systems: Computation and Control*, 253–266. Springer.

Jamieson, K.; and Nowak, R. 2014. Best-arm identification algorithms for multi-armed bandits in the fixed confidence setting. In *Information Sciences and Systems (CISS), 2014 48th Annual Conference on*, 1–6. IEEE.

Jiang, H.; and Kpotufe, S. 2017. Modal-set estimation with an application to clustering. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, 1197–1206. PMLR.

Madani, T.; and Benallegue, A. 2007. Backstepping control with exact 2-sliding mode estimation for a quadrotor unmanned aerial vehicle. In *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 141–146. IEEE.

Okada, R.; Fukuchi, K.; and Sakuma, J. 2015. Differentially private analysis of outliers. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 458–473. Springer.

Parzen, E. 1962. On estimation of a probability density function and mode. *The annals of mathematical statistics* 33(3): 1065–1076.

Robbins, H. 1985. Some aspects of the sequential design of experiments. In *Herbert Robbins Selected Papers*, 169–177. Springer.

Sarmadi, S. N.; and Venkatasubramanian, V. 2013. Electromechanical mode estimation using recursive adaptive stochastic subspace identification. *IEEE Transactions on Power Systems* 29(1): 349–358.

Sheikh, Y. A.; Khan, E. A.; and Kanade, T. 2007. Mode-seeking by medoidshifts. In *2007 IEEE 11th International Conference on Computer Vision*, 1–8. IEEE.

Silverman, B. W. 1981. Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society: Series B (Methodological)* 43(1): 97–99.

Szorenyi, B.; Busa-Fekete, R.; Weng, P.; and Hüllermeier, E. 2015. Qualitative multi-armed bandits: A quantile-based approach. In *32nd International Conference on Machine Learning*, 1660–1668.

Tao, W.; Jin, H.; and Zhang, Y. 2007. Color image segmentation based on mean shift and normalized cuts. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 37(5): 1382–1389.

Tsybakov, A. B. 1990. Recursive estimation of the mode of a multivariate distribution. *Problemy Peredachi Informatsii* 26(1): 38–45.

Vedaldi, A.; and Soatto, S. 2008. Quick shift and kernel methods for mode seeking. In *European conference on computer vision*, 705–718. Springer.

Vieu, P. 1996. A note on density mode estimation. *Statistics & probability letters* 26(4): 297–307.

Williams, B. C.; Chung, S.; and Gupta, V. 2001. Mode estimation of model-based programs: monitoring systems with complex behavior. In *IJCAI*, 579–590.

Yamato, H. 1971. Sequential estimation of a continuous probability density function and mode. *Bull. Math. Statist* 14: 1–12.

Yin, P.; Tourapis, H.-Y.; Tourapis, A. M.; and Boyce, J. 2003. Fast mode decision and motion estimation for JVT/H. 264. In *Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429)*, volume 3, III–853. IEEE.