On Finding the Largest Mean Among Many

Kevin Jamieson[†], Matthew Malloy[†][*] Robert Nowak[†], and Sébastien Bubeck[‡]

[†]Department of Electrical and Computer Engineering,
University of Wisconsin-Madison
[‡]Princeton University,
Department of Operations Research and Financial Engineering

Abstract

Sampling from distributions to find the one with the largest mean arises in a broad range of applications, and it can be mathematically modeled as a multi-armed bandit problem in which each distribution is associated with an arm. This paper studies the sample complexity of identifying the best arm (largest mean) in a multi-armed bandit problem. Motivated by large-scale applications, we are especially interested in identifying situations where the total number of samples that are necessary and sufficient to find the best arm scale linearly with the number of arms. We present a single-parameter multi-armed bandit model that spans the range from linear to superlinear sample complexity. We also give a new algorithm for best arm identification, called PRISM, with linear sample complexity for a wide range of mean distributions. The algorithm, like most exploration procedures for multi-armed bandits, is adaptive in the sense that the next arms to sample are selected based on previous samples. We compare the sample complexity of adaptive procedures with simpler non-adaptive procedures using new lower bounds. For many problem instances, the increased sample complexity required by non-adaptive procedures is a polynomial factor of the number of arms.

1 Introduction

This paper studies the sample complexity of finding the best arm in a multi-armed bandit problem. Consider $n + 1$ arms with mean payoffs $\mu_0 > \mu_1 > \cdots > \mu_n$. The mean values and the ordering of the arms are unknown. The goal is to identify the arm with the largest mean (i.e., the "best arm") by sampling the arms. A sample of arm i is an independent realization of a random variable $X_i \in [0, 1]$ with mean $\mu_i \in [0, 1]$ (it is straightforward to extend all results presented in this paper to sub-Gaussian realizations with bounded means and variances).

The main focus of this paper is identifying necessary and sufficient conditions under which the sample complexity (total number of samples) of finding the best arm grows linearly in the number of arms. This is motivated by applications involving very large numbers of arms, such as virus replication experiments testing thousands of cell strains [3], cognitive radio problems searching over hundreds of communication channels [1, 2], and network surveillance of large social networks. These applications are time-consuming and/or costly, so minimizing the number of samples required to find the most influential genes, best channels, or malicious agents is crucial. This paper quantifies the minimum number of samples needed in such applications and gives a new algorithm called PRISM that succeeds using a total number of samples within a negligible factor of the minimum.

---

[*]The first two authors are listed in alphabetical order as both contributed equally.

Mannor and Tsitsiklis [4] showed that for any procedure that finds the best arm with probability at least $1 - \delta$ requires on the order of $\mathbf{H}\log(1/\delta)$ samples, where $\mathbf{H} := \sum_{i=1}^{n}(\mu_0 - \mu_i)^{-2}$. This lower bound shows that the sample complexity can be much greater than the number of arms. For example, if the gap between μ_0 and μ_1 is $1/n$, then $\mathbf{H}$, and the sample complexity, are at least $O(n^2)$. On the other hand, scenarios can arise in which $\mathbf{H}$ grows linearly with n. For instance, if $\mu_0 - \mu_i$ is greater than a positive constant for all i, then $\mathbf{H} = O(n)$. This case is "sparse" in the sense that μ_0 is bounded away from all others, and recent work has shown that $O(n)$ samples are sufficient in such cases [5].

Most bandit exploration algorithms are sequential and adaptive in the sense that the selection of the arms to sample next is based on previous samples. This is necessary in order to achieve linear sample complexity in the sparse case mentioned above. Non-adaptive methods, in which every arm is sampled an equal number of times, require at least $O(n \log n)$[1] samples [6]. The factor of $\log n$ is significant if n is large, which is one motivation for adaptive strategies like the one used in [3]. Contrasting the differences between adaptive and non-adaptive sampling is a second focus of this paper.

Of particular interest here is the scaling of the sample complexity as a function of the number of arms and the behavior of the gaps between their means. The sparse model discussed above is an enlightening idealization, but unlikely to arise in practice. A smoothly decaying distribution of means may be a more reasonable model for the biological and radio applications discussed above. Fig. 1 depicts the means in three different arm configurations. The left plot (a) represents a sparse model in which μ_0 is bounded away from all others means by a fixed constant (all gaps greater than a fixed constant). In this case, the sample complexities of non-adaptive and adaptive strategies differs by a factor of $\log n$. The other two plots represent cases in which the gaps between means are shrinking as n increases. In Figure 1(b) $\mu_0 - \mu_i = (\frac{i}{n})^{.49}$ and in (c) $\mu_0 - \mu_i = \frac{i}{n}$. The difference between the sample complexities is much more significant for these non-sparse cases. Also note that there are non-sparse cases in which adaptive strategies can find the best arm in $O(n)$ samples like the one shown in Fig. 1(b).
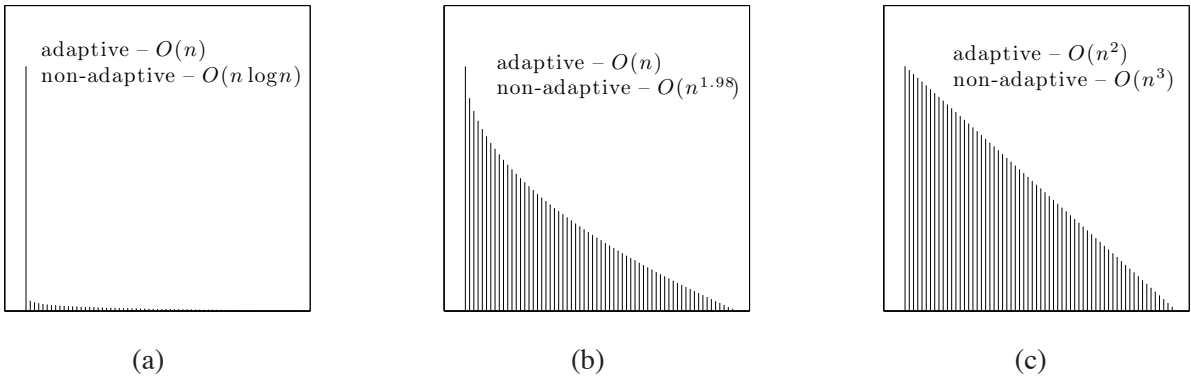

adaptive $- O(n)$ non-adaptive $- O(n \log n)$	adaptive $- O(n)$ non-adaptive $- O(n^{1.98})$	adaptive $- O(n^2)$ non-adaptive $- O(n^3)$
(a)	(b)	(c)

Figure 1: Three different configurations of means, each ordered $\mu_0 > \mu_1 > \cdots > \mu_n$. In (a) $\mu_0 - \mu_i \geq 0.95$, in (b) $\mu_0 - \mu_i = (\frac{i}{n})^{.49}$, and in (c) $\mu_0 - \mu_i = \frac{i}{n}$, for $i = 1, \ldots, n$. The necessary sample complexities (sufficient to within $\log \log n$ factors) of non-adaptive and adaptive strategies are indicated in each case. Finding the best arm becomes increasingly difficult as the gaps between the means decrease, but in all cases adaptive strategies have significantly lower sample complexities.

---

[1]In this paper our focus is on how the number of samples scales with n, not necessarily the probability of failure δ. For instance, if we were to say an algorithm requires just $O(n)$ samples then it is understood that this many samples suffices to find the best arm with a fixed probability of error. However, in the theorem statements the dependence on δ is explicit.

1.1 Contributions and Organization

The paper is organized as follows. In Sec. 2, we present a single-parameter model for the distribution of means that spans the range from linear to superlinear sample complexity. In Sec. 3 we present an algorithm for best arm identification with sample complexity $O(\mathbf{H}\log(1/\delta))$ for a wide range of mean distributions. In particular, we show that the algorithm has linear sample complexity for all of the single-parameter distributions satisfying $\mathbf{H} = O(n)$. Our bounds apply to the PAC (probably approximately correct) setting (generalizing the algorithm to the fixed budget case of [7, 8] is a challenging open problem). While completing this paper[2], we became aware of independent work on the best arm problem to appear at ICML 2013 [9]. The algorithm and theoretical analysis in that paper are essentially the same as ours, but in fact the upper bound on sample complexity bound given in [9] is slightly tighter.

Sec. 4 discusses the limitations of non-adaptive sampling strategies and shows that any non-adaptive sampling strategy may require drastically more samples than our adaptive algorithm. Using new lower bounds for non-adaptive sampling procedures, we show that there exist problems in which the difference between the sample complexities of non-adaptive and adaptive procedures grows *polynomially* with the number of arms. This is somewhat surprising, as the as the advantage of adaptivity in the sparse setting (where the gaps are bound by a fixed constant) is known to be a factor of $\log n$ at best. To be more concrete, consider the following. We demonstrate problem instances where adaptive procedures, for example, succeed with just $O(n)$ samples, but all non-adaptive procedures fail without at least $O(n^{1.98})$ samples. This observation is crucial since it shows that adaptive designs can be vastly superior to simpler non-adaptive methods often used in practice (e.g., biological applications mentioned above). The take-away message is that the added implementational burden of adaptive sampling methods may be well worth the investment.

Notation, in general, follows convention. Since the order and means are unknown, we denote the index of the best arm as i^* throughout the paper. An estimate of the best arm is denoted as $\hat{i}$. Proofs of all Theorems are found in the Appendix.

2   A Single-Parameter Family of Mean Distributions

A lower bound on the sample complexity of finding the best arm follows in a straightforward way from [4, Theorem 5]; see Theorem 3 in Sec. 4 for the derivation. To find the best arm with probability at least $1 - \delta$ requires at least

$$c_1 \mathbf{H} \log(1/\delta)$$

samples, where $c_1 > 0$ is a universal constant. The quantity $\mathbf{H}$, refereed to as the hardness of the problem, is given by

$$\mathbf{H} = \sum_{i=1}^{n} \Delta_i^{-2}. \tag{1}$$

where

$$\Delta_i = \mu_0 - \mu_i \tag{2}$$

is the gap between the best arm and the ith arm.

In this paper we focus on a specific parametric family that spans the hardness of the best arm problem with a single parameter. Consider a model in which the means are given by

$$\mu_i = \mu_0 - (i/n)^\alpha \qquad (3)$$

for $i = 1, \ldots, n$ and some $\alpha \geq 0$. We refer to this model as an α-*parameterization*. Under this model, $\Delta_i = (i/n)^\alpha$. The α-parameterization spans the range from "hard" problems in which the gaps Δ_i are shrinking quickly as n grows, to "easy" or sparse cases in which the gaps are greater than a constant (when $\alpha = 0$). Theorem 3 in Sec. 4 yields the following lower bounds on the sample complexity of identifying the best arm (ignoring constant and $\log(1/\delta)$ factors):

$$\text{sample complexity} \geq \begin{cases} n & \text{if } \alpha < 1/2 \\ n \log n & \text{if } \alpha = 1/2 \\ n^{2\alpha} & \text{if } \alpha > 1/2. \end{cases} \qquad (4)$$

The focus of this paper is on finding sufficient conditions under which the best arm can be found with $O(n)$ samples. If the gaps follow the α-parameterization with $\alpha \geq 1/2$, (4) implies that the best arm *cannot* be found with $O(n)$ samples. Conversely, if the gaps satisfy $\Delta_i \geq C(\frac{i}{n})^\alpha$, for $\alpha < 1/2$, the lower bound in (4) does not preclude the possibility that order n samples are sufficient to find the best arm. In the next section, we show that when $\alpha < 1/2$, order n samples are indeed sufficient.

3 PRISM Algorithm for Best Arm Identification

To show that a linear number of pulls is sufficient for a number of problem instances, we propose and analyze the algorithm for best arm identification outlined in Fig. 2. The algorithm follow a multi-phase approach, with a specific allocation of confidence and sampling budgets across phases. The algorithm relies on the output of Median Elimination [10] to establish a threshold on each phase. We mention again independent work to appear at ICML 2013 [9], which proposes and analyzes essentially the same algorithm.
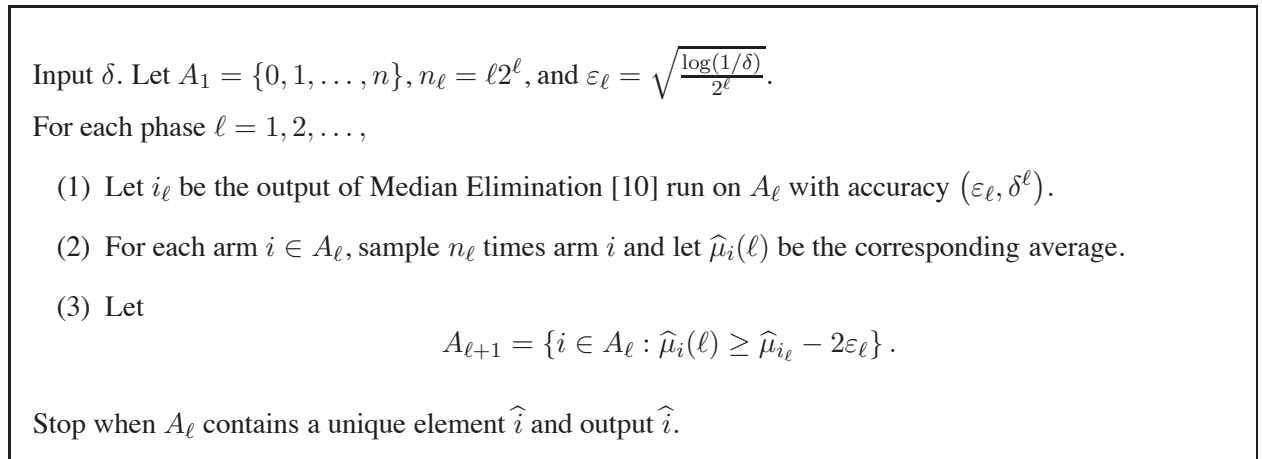

Input δ. Let $A_1 = \{0, 1, \ldots, n\}$, $n_\ell = \ell 2^\ell$, and $\varepsilon_\ell = \sqrt{\frac{\log(1/\delta)}{2^\ell}}$.

For each phase $\ell = 1, 2, \ldots,$

   (1) Let i_ℓ be the output of Median Elimination [10] run on A_ℓ with accuracy $(\varepsilon_\ell, \delta^\ell)$.

   (2) For each arm $i \in A_\ell$, sample n_ℓ times arm i and let $\widehat{\mu}_i(\ell)$ be the corresponding average.

   (3) Let
$$A_{\ell+1} = \{i \in A_\ell : \widehat{\mu}_i(\ell) \geq \widehat{\mu}_{i_\ell} - 2\varepsilon_\ell\}.$$

Stop when A_ℓ contains a unique element $\widehat{i}$ and output $\widehat{i}$.

Figure 2: PRISM algorithm for the best arm identification problem.

The following theorem is our main result. The sample complexity of identifying the best arm with probability at least $1 - \delta$ is bounded in terms of $\mathbf{H}$ and a novel measure of complexity denoted by $\mathbf{G} :=$

$\sum_{i=1}^{n} \Delta_i^{-2} \log_2(\Delta_i^{-2})$. In general $\mathbf{H} \le \mathbf{G} \le \mathbf{H} \log(\mathbf{H})$ but in many cases, $\mathbf{G} = \mathbf{H}$ and the bound implies that the best arm can be found using $O(\mathbf{H})$ samples with a fixed probability of error. This is the best known bound for the best arm problem. The proof is left to the appendix.

Theorem 1. *Let $\delta \in (0,1)$. Let $\mathbf{H} = \sum_{i=1}^{n} \Delta_i^{-2}$ where Δ_1 is the minimum gap. Then with probability at least $1 - \frac{3\delta^2}{1-\delta^2} - \frac{\delta}{1-\delta} - \frac{4\delta^2}{(1-\delta^2)^2}$, the PRISM algorithm of Fig. 2 stops after at most*

$$O\left(\log(1/\delta) \left[\mathbf{H} \log(\log(1/\delta)) + \sum_{i=1}^{n} \Delta_i^{-2} \log_2(\Delta_i^{-2}) \right] \right)$$

samples and outputs arm $\widehat{i} = i^*$.

Corollary 1. *Consider the problem instance $\mu_0 = 1$ and $\mu_i = 1 - (i/n)^\alpha$, for $i = 1, \ldots, n$ and some $0 < \alpha$. Then for a fixed probability error, using the PRISM algorithm of Fig. 2 we have that*

$$\text{number of total samples} = \begin{cases} O(n) & \text{if } \alpha < 1/2 \\ O(n \log^2 n) & \text{if } \alpha = 1/2 \\ O(n^{2\alpha} \log(n)) & \text{if } \alpha > 1/2. \end{cases} \tag{5}$$

and the algorithm outputs arm $\widehat{i} = i^*$.

Proof. For $\Delta_i = \left(\frac{i}{n}\right)^\alpha$ we have the relevant quantities given in the following table (ignoring lower order terms):

	$\mathbf{H}$	$\sum_{i=1}^{n} \Delta_i^{-2} \log_2(\Delta_i^{-2})$
$\alpha > 1/2$	$\frac{2\alpha}{2\alpha-1} n^{2\alpha}$	$\frac{(2\alpha)^2}{2\alpha-1} n^{2\alpha} \log(n)$
$\alpha = 1/2$	$n \log(n)$	$2\alpha n \log^2(n)$
$\alpha < 1/2$	$\frac{2\alpha}{1-2\alpha} n$	$\frac{2\alpha}{(1-2\alpha)^2} n$

The result follows from plugging the above quantities into Theorem 1. $\qquad\square$

Conservative PRISM. Consider the algorithm of Fig. 2, but in the first line set $n_\ell = 2^\ell$ and $\varepsilon_\ell = \sqrt{\log(\ell^2/\delta)/2^\ell}$ and for item (2), run Median Elimination with input $(\varepsilon_\ell, \delta/\ell^2)$.

Theorem 2. *Let $\delta \in (0, 0.6]$. With probability at least $1 - 2\delta - 6\delta^2 - 6\delta^4$, Conservative PRISM stops after at most $O\left(\mathbf{H} \log\left(\frac{\log(\mathbf{H})}{\delta}\right)\right)$ pulls and outputs $\widehat{i} = i^*$.*

Theorem 1 matches the lower bound when $\alpha < 1/2$ and comes within a $\log(n)$ factor for $\alpha \ge 1/2$. On the other hand, Theorem 2 comes within a factor of $\log \log n$ of the lower bound for all $\alpha > 0$. We note that the upper bound of [9] also comes no closer than $\log \log n$ of the lower bound for $\alpha \ge 1/2$.

4 Lower Bounds on the Sample Complexity of Non-Adaptive Algorithms

Here we examine the limitations of non-adaptive sampling strategies (which sample all arms an equal number of times), since these simpler procedures are not uncommon in applications like the biological problems that partially motivate this paper. A non-adaptive procedure is any procedure that samples each arm m

times, where m is fixed a-priori, and outputs a single arm as an estimate of the best arm. We show that all non-adaptive methods may require drastically more samples than PRISM. The take-away message is that the small added difficulty (for the practitioner) in applying PRISM may be well worth the investment.

We begin by formally stating the adaptive lower bound developed in [4].

Theorem 3. Adaptive Lower Bound *[4, Theorem 5]. For every set of means, $\{\mu_i\}_{i=0}^n$, $\mu_i \in (3/8, 1/2]$, there exists a joint distribution on the arms such that the arms take mean values $\{\mu_0, \ldots, \mu_n\}$, each arm is sub-Gaussian, and any adaptive procedure with fewer than*

$$c_1 \mathbf{H} \log \frac{1}{8\delta} \tag{6}$$

samples in expectation has $\mathbb{P}(\hat{i} \neq i^*) \geq \delta$ for any $\delta \in (0, e^{-8}/8)$ and some constant c_1.

Note that the restriction on the means to $(3/8, 1/2]$ in Theorem 3 can be relaxed (see [4] for details). We proceed with the non-adaptive lower bounds which allow us to compare the best non-adaptive procedures against adaptive procedures.

Theorem 4. Non-Adaptive Lower Bound. *Consider any $\delta \in (0, e^{-3}/24)$. For every set of means $\{\mu_1, \ldots, \mu_n\}$, there exists a joint distribution on the arms such that the arms take mean values $\{\mu_1, \ldots, \mu_n\}$, each arm is sub-Guassian, and any non-adaptive procedure with fewer than*

$$\mathbf{H} \log \left(\frac{n}{25\delta} \right)$$

samples in expectation has $\mathbb{P}(\hat{i} \neq i^*) \geq \delta$. Moreover, for any value of $\mathbf{H}$ there exists a joint sub-Gaussian distribution over arms with means $\{\mu_0, \ldots, \mu_n\}$ satisfying $\sum_{i=1}^n (\mu_0 - \mu_i)^{-2} = \mathbf{H}$, such that any non-adaptive procedure with fewer than

$$\frac{\mathbf{H}n}{2} \log \left(\frac{1}{24\delta} \right)$$

samples in expectation has $\mathbb{P}(\hat{i} \neq i^*) \geq \delta$.

The lower bound of Theorem 4 consists of two statements, the first which implies that for *any* set of means, the sample complexity must be at least order $\mathbf{H} \log n$. The second statement, on the other hand, implies the existence of particular problem instances that are especially difficult for non-adaptive procedures, requiring order $\mathbf{H}n$ samples. Inspecting the proofs of the two parts of Theorem 4 one sees that the minimum gap Δ_1, not $\mathbf{H}$, is governing the query complexity for non-adaptive procedures. Using this fact we have the following Theorem which is proved in the appendix.

Theorem 5. Non-adaptive Lower Bound, α-parameterization. *Consider arms with mean values according to the parameterization of (3) for some $\alpha \geq 0$. There exists a joint sub-Gaussian distribution on the arms such that any non-adaptive procedure with fewer than*

$$\begin{cases} n \log \left(\frac{n}{25\delta} \right) & \text{if } \alpha = 0 \\ n^{2\alpha+1} \log \left(\frac{1}{24\delta} \right) & \text{if } \alpha > 0 \end{cases}$$

samples to find the best arm with a fixed probability of failure.

To see that Theorem 5 is indeed tight, it is straightforward to show that the non-adaptive procedure which chooses the arm with the largest empirical mean after sampling each arm the same number of times does indeed meet the lower bound. Letting m be the number of times each arm is sampled. Then

$$\mathbb{P}\left(\widehat{i} \neq i^*\right) \leq \sum_{i \neq i^*} \mathbb{P}\left(\widehat{\mu}_{i^*} \leq \widehat{\mu}_i\right) \leq \sum_{i \neq i^*}\left(\mathbb{P}\left(\widehat{\mu}_{i^*} \leq \mu_{i^*} - \Delta_i/2\right) + \mathbb{P}\left(\widehat{\mu}_i \geq \mu_i + \Delta_i/2\right)\right)$$

$$\leq \sum_{i \neq i^*} 2\exp(-m\Delta_i^2) = \sum_{i \neq i^*} 2\exp\left(-m\left(\frac{i}{n}\right)^{2\alpha}\right)$$

which follow from a union bound and Hoeffding's inequality. For $\alpha \neq 0$, if $m \geq n^{2\alpha}$ (which implies the total number of samples is greater than $n^{2\alpha+1}$) the above sum is convergent, and the probability that the wrong arm is returned is controlled. The case where $\alpha = 0$ is also controlled if $m \geq \log n$.

We conclude that for $\alpha \in (0, 1/2)$, when compared to adaptive procedures that require just $O(n)$ samples, any non-adaptive procedure requires a factor of $n^{2\alpha}$ more samples to identify the best arm. The implications of this observation can be somewhat surprising: for many problem instances, the improvement in the sample complexity resulting from adaptivity is polynomial in n, compared with the typical $\log(n)$ improvement observed for sparse problems ($\alpha = 0$).

References

[1] Simon Haykin. Cognitive radio: brain-empowered wireless communications. *Selected Areas in Communications, IEEE Journal on*, 23(2):201–220, 2005.

[2] David López-Pérez, Alvaro Valcarce, Guillaume De La Roche, and Jie Zhang. OFDMA femtocells: A roadmap on interference avoidance. *Communications Magazine, IEEE*, 47(9):41–48, 2009.

[3] L. Hao, A. Sakurai, T. Watanabe, E. Sorensen, C. Nidom, M. Newton, P. Ahlquist, and Y. Kawaoka. Drosophila RNAi screen identifies host genes important for influenza virus replication. *Nature*, page 8903, 2008.

[4] Shie Mannor and John N Tsitsiklis. The sample complexity of exploration in the multi-armed bandit problem. *The Journal of Machine Learning Research*, 5:623–648, 2004.

[5] Matthew L Malloy and Robert Nowak. Sequential testing for sparse recovery. *arXiv preprint arXiv:1212.1801*, 2012.

[6] Matt Malloy and Robert Nowak. On the limits of sequential testing in high dimensions. In *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on*, pages 1245–1249. IEEE, 2011.

[7] J.-Y. Audibert, S. Bubeck, and R. Munos. Best arm identification in multi-armed bandits. In *Proceedings of the 23rd Annual Conference on Learning Theory (COLT)*, 2010.

[8] S. Bubeck, R. Munos, and G. Stoltz. Pure exploration in multi-armed bandits problems. In *Proceedings of the 20th International Conference on Algorithmic Learning Theory (ALT)*, 2009.

[9] Z. Karnin, T. Koren, and O. Somekh. Almost optimal exploration in multi-armed bandits. *Proceedings of the 30th International Conference on Machine Learning*, June 2013.

[10] Eyal Even-Dar, Shie Mannor, and Yishay Mansour. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *The Journal of Machine Learning Research*, 7:1079–1105, 2006.

[11] G. Abreu. Very simple tight bounds on the q-function. *Communications, IEEE Transactions on*, 60(9):2415–2420, 2012.

[12] Jorge Nocedal and Stephen J Wright. *Numerical optimization*. Springer Science and Business Media, 2006.

A Appendix

A.1 Proof of Theorem 1

Proof. It will be useful to consider the following 'slicing' of arms:

$$\Omega_s = \{i \in [n] : 5\sqrt{2}\,\varepsilon_{s+1} < \Delta_i \leq 5\sqrt{2}\,\varepsilon_s\}, s \geq 1.$$

Note that

$$25 \log(1/\delta) \sum_{i \in \Omega_s} \Delta_i^{-2} \leq 2^s |\Omega_s| \leq 50 \log(1/\delta) \sum_{i \in \Omega_s} \Delta_i^{-2}. \tag{7}$$

Step 1: A good event. In this step we describe the event of probability $1 - \delta$ on which we will prove the result. We want the following to hold:

$$\widehat{\mu}_{i^*}(\ell) - \mu^* \geq -\varepsilon_\ell \,, \forall \ell \geq 1 \,, \tag{8}$$

$$|\widehat{\mu}_{i_\ell}(\ell) - \mu_{i_\ell}| \leq \varepsilon_\ell \,, \forall \ell \geq 1 \,, \tag{9}$$

$$\max_{j \in A_\ell} \mu_j - \mu_{i_\ell} \leq \varepsilon_\ell \,, \forall \ell \geq 1 \,, \tag{10}$$

$$\left| \{i \in A_\ell \cap \Omega_s : \widehat{\mu}_i(\ell) \geq \widehat{\mu}_{i_\ell} - 2\varepsilon_\ell\} \right| \leq \frac{|A_\ell \cap \Omega_s|}{4} \,, \forall \ell \geq s \geq 1. \tag{11}$$

We first bound the probability that the above events do not hold. By Hoeffding's inequality we have

$$\mathbb{P}\left(\hat{\mu}_i(\ell) - \mu_i \geq \varepsilon_\ell\right) \leq \exp(-2n_\ell \varepsilon_\ell^2) = \delta^{2\ell}$$

for any i and note that an analogous inequality holds for the deviation away from its mean in the other direction. Thus, applying Hoeffding's and a union bound we have that the probability that 8) is not satisfied is less than $\frac{\delta^2}{1-\delta^2}$. The probability that (9) is not satisfied is bound in the exact same manner with an additional factor of two to satisfy both inequality directions. The only subtlety is that after the union bound one needs to condition on the value of i_ℓ before using Hoeffding's inequality, and this is possible since the random variables obtain in Step (2) of the algorithm are independent of i_ℓ. By the properties of median elimination and a union bound we have that the probability that (10) is not satisfied is less than $\frac{\delta}{1-\delta}$. Observe that by (8) and (9) one always has:

$$\widehat{\mu}_{i^*}(\ell) \geq \mu^* - \varepsilon_\ell \geq \mu_{i_\ell} - \varepsilon_\ell \geq \widehat{\mu}_{i_\ell}(\ell) - 2\varepsilon_\ell.$$

which implies that the best arm is never removed from the set.

It remains to bound the probability that (11) is not satisfied while (8), (9) and (10) are satisfied.

$$
\begin{aligned}
\mathbb{P}\left(\left|\left\{i \in A_\ell \cap \Omega_s : \widehat{\mu}_i(\ell) \geq \widehat{\mu}_{i_\ell} - 2\varepsilon_\ell\right\}\right| > \frac{|A_\ell \cap \Omega_s|}{4} \middle| (8),(9),(10), A_\ell, i_\ell\right) & \\
\leq \mathbb{P}\left(\left|\left\{i \in A_\ell \cap \Omega_s : \widehat{\mu}_i(\ell) \geq \mu^* - 4\varepsilon_\ell\right\}\right| > \frac{|A_\ell \cap \Omega_s|}{4} \middle| A_\ell\right) & \\
\leq \frac{4}{|A_\ell \cap \Omega_s|}\mathbb{E}\left(\left|\left\{i \in A_\ell \cap \Omega_s : \widehat{\mu}_i(\ell) \geq \mu^* - 4\varepsilon_\ell\right\}\right| \middle| A_\ell\right) & \\
= \frac{4}{|A_\ell \cap \Omega_s|}\sum_{i \in A_\ell \cap \Omega_s} \mathbb{P}\left(\widehat{\mu}_i(\ell) \geq \mu^* - 4\,\varepsilon_\ell \mid A_\ell\right) & \\
\leq 4\exp(-2n_\ell\varepsilon_\ell^2) = 4\delta^{2\ell}. &
\end{aligned}
$$

where the last inequality follows since all elements of Ω_s have gaps greater than $5\sqrt{2}\varepsilon_{s+1} = 5\varepsilon_s$. Summing over all $s \leq \ell$ and then over $\ell \geq 1$ gives the probability that (11) is not satisfied: $\sum_{\ell=1}^{\infty}\sum_{s\leq\ell}4\delta^{2\ell} = \frac{4\delta^2}{(1-\delta^2)^2}$. In the next steps we will assume that (8)-(9)-(10)-(11) are satisfied as they all hold with probability at least $1 - \frac{3\delta^2}{1-\delta^2} - \frac{\delta}{1-\delta} - \frac{4\delta^2}{(1-\delta^2)^2}$.

Step 2: Bound on the total number of phases. It suffices to bound the number of phases given that (8), (9), (10) and (11) hold. Let $\mathcal{L}$ denote the first phase such that $|A_\ell| = 1$ (if there is no such phase then $\mathcal{L} = +\infty$). Observe that using (11) one can show by induction that

$$
|A_\ell| \leq 1 + \sum_{s=1}^{\ell}\frac{|\Omega_s|}{4^{\ell-s}} + \sum_{s=\ell+1}^{+\infty}|\Omega_s|. \tag{12}
$$

Define $s^* = \log_2(\Delta_1^{-2}\log(1/\delta))$ so that $\Omega_s = \emptyset$ for all $s > s^*$. By definition, when $\ell \geq s^*$ the third term in the equation immediately above is equal to zero so that

$$
|A_\ell| \leq 1 + 2^{-\ell}\sum_{s=1}^{s^*}2^{s-s^*}2^s|\Omega_s| \qquad \forall \ell \geq s^*
$$

where we have

$$
\begin{aligned}
\sum_{s=1}^{s^*}2^{s-s^*}2^s|\Omega_s| &= \frac{\Delta_1^2}{\log(1/\delta)}\sum_{s=1}^{\infty}2^{2s}|\Omega_s| \\
&= \frac{\Delta_1^2}{\log(1/\delta)}\sum_{s=1}^{\infty}2^{2s}\sum_{i=1}^{n}\mathbf{1}\left\{5\sqrt{2}\sqrt{\frac{\log(1/\delta)}{2^{s+1}}} < \Delta_i \leq 5\sqrt{2}\sqrt{\frac{\log(1/\delta)}{2^s}}\right\} \\
&\leq \frac{\Delta_1^2}{\log(1/\delta)}\sum_{i=1}^{n}\frac{50^2\log(1/\delta)^2}{\Delta_i^2} = 50^2\log(1/\delta)\mathbf{H}\Delta_1^2.
\end{aligned}
$$

We conclude that for

$$
\mathcal{L} := 1 + \max\left\{s^*, \log_2(50^2\log(1/\delta)\mathbf{H}\Delta_1^2)\right\} = \log_2(2\log(1/\delta)) + \max\left\{\log_2(\Delta_1^{-2}), \log_2(50^2\mathbf{H}\Delta_1^2)\right\}
$$

we have that $|A_\ell| < 2$ whenever $\ell \geq \mathcal{L}$. Hence, $\mathcal{L}$ is an upper bound on the stopping time.

Step 3: Bound on the total number of pulls. Recall that Median Elimination applied to a set A_ℓ with parameters $\varepsilon_\ell, \delta^\ell$ takes no more than $\frac{c_{\mathrm{ME}}}{\varepsilon_\ell^2}|A_\ell|\ell\log(1/\delta) = c_{\mathrm{ME}}\ell 2^\ell|A_\ell|$ pulls. Thus, the total number of pulls

on phase ℓ is bounded by $c\ell 2^\ell |A_\ell|$ pulls with $c = c_{\text{ME}} + 1$. Using the results from the previous step (in particular (12) and the stopping time $\mathcal{L}$) one has that the total number of pulls is bounded from above by

$$
\begin{aligned}
\sum_{\ell=1}^{\mathcal{L}} c\ell 2^\ell |A_\ell| &\leq \sum_{\ell=1}^{\mathcal{L}} c\ell 2^\ell \left(1 + \sum_{s=1}^{\ell} \frac{|\Omega_s|}{4^{\ell-s}} + \sum_{s=\ell+1}^{\infty} |\Omega_s| \right) \\
&\leq c\mathcal{L}2^{\mathcal{L}+1} + c \sum_{\ell=1}^{+\infty} \ell \sum_{s=1}^{+\infty} \left(\frac{2^s}{2^\ell} \mathbb{1}_{s\leq\ell} + \frac{2^\ell}{2^s} \mathbb{1}_{s>\ell} \right) 2^s |\Omega_s| \\
&\leq c\mathcal{L}2^{\mathcal{L}+1} + 3c \sum_{s=1}^{+\infty} s 2^s |\Omega_s| + 2c(n-1) \\
&= c\mathcal{L}2^{\mathcal{L}+1} + 150c \log(1/\delta) \left[\log_2(50 \log(1/\delta)) \mathbf{H} + \mathbf{G} \right] + 2c(n-1)
\end{aligned}
$$

where

$$
\mathbf{H} \leq \mathbf{G} := \sum_{i=1}^{n} \Delta_i^{-2} \log_2 \left(\Delta_i^{-2} \right) \leq \mathbf{H} \log(\mathbf{H})
$$

which follows directly from

$$
\begin{aligned}
\sum_{s=1}^{\infty} s 2^s |\Omega_s| &= \sum_{s=1}^{\infty} s 2^s \sum_{i=1}^{n} \mathbb{1} \left\{ 5\sqrt{2} \sqrt{\frac{\log(1/\delta)}{2^{s+1}}} < \Delta_i \leq 5\sqrt{2} \sqrt{\frac{\log(1/\delta)}{s 2^s}} \right\} \\
&\leq 50 \log(1/\delta) \sum_{i=1}^{n} \frac{1}{\Delta_i^2} \log_2 \left(\frac{50 \log(1/\delta)}{\Delta_i^2} \right).
\end{aligned}
$$

Evaluating $\mathcal{L}2^{\mathcal{L}+1}$ and collecting terms obtains the result. $\qquad\square$

A.2 Proof of Theorem 3

Proof. Assume some procedure has $\mathbb{P}(\widehat{i} \neq i^*) \leq \delta$ and requires fewer than $c_1 \mathbf{H} \log \frac{1}{8\delta}$ samples for some $\{\mu_i\}_{i=0}^{n}$, $\mu_i \in (3/8, 1/2]$. This procedure is by definition (ε, δ) PAC (probably approximately correct) for any $\varepsilon \in (0, \Delta_1)$. [4, Theorem 5] implies any (ε, δ), procedure, $\varepsilon \in (0, \Delta_1)$, requires more than

$$
c_1 \sum_{i \in \mathcal{N}} \frac{1}{\mu_{i^*} - \mu_i} \log \frac{1}{8\delta}
$$

samples in expectation, where

$$
\mathcal{N} = \left\{ i : \mu_i \leq \mu_{i^*} - \varepsilon, \mu_i \geq \frac{\varepsilon + \mu_{i^*}}{1 + \sqrt{1/2}} \right\}
$$

Since $\mu_i \in (3/8, 1/2]$, $\mathcal{N} := [n]$. Any procedure requires more than

$$
c_i \mathbf{H} \log \frac{1}{8\delta}
$$

samples in expectation. This negates the original assumption. $\qquad\square$

A.3 Proof of Theorem 4

Proof. We restrict our attention to reward distributions of the form $\mathcal{N}(\mu_i, 1)$. Assume that $\mu_i, i = 0, \ldots, n$, are know up to a permutation, and let each arm be assigned a mean uniformly at random. We first show that the test with minimum average probability of error simply picks the largest empirical mean among all arms, i.e., $\widehat{i} = \arg\max_i \widehat{\mu}_i$, where $\widehat{\mu}_i = 1/m \sum_{j=1}^m X_{i,j}$, and $X_{i,j}$ represents the reward of arm i on the jth play of that arm, and m is the total number of samples of each arm. This can be seen by considering the maximum a-posteriori (MAP) estimator of the best arm, which by definition has the smallest probably of error. Under the assumption that the arms are assigned means uniformly at random, the MAP estimator reduces to the maximum likelihood (ML) estimator:

$$\widehat{i}_{\mathrm{MAP}} = \widehat{i}_{\mathrm{ML}} = \arg\max_i \ P(X_0^m, \ldots, X_n^m | \mathcal{H}_i),$$

where $X_i^m = X_{i,1}, \ldots, X_{i,m}$ and $\mathcal{H}_i$ is event that arm i is the best arm. Consider comparing between events $\mathcal{H}_i$ and $\mathcal{H}_{i'}$, $i \neq i'$: the ML test is $P(X_0^m, \ldots, X_n^m | \mathcal{H}_i) \lessgtr_i^{i'} P(X_0^m, \ldots, X_n^m | \mathcal{H}_{i'})$. By the independence across arms, it is straightforward to show this test reduces to:

$$P(X_i^m | \mathcal{H}_i) P(X_{i'}^m | \mathcal{H}_i) \lessgtr_i^{i'} P(X_i^m | \mathcal{H}_{i'}) P(X_{i'}^m | \mathcal{H}_{i'}) \tag{13}$$

The distribution of X_i^m, given $\mathcal{H}_i$, is simply

$$P(X_i^m | \mathcal{H}_i) = \frac{1}{\sqrt{2\pi}} \exp\left(-\left\|X_i^m - \mu_{i^*}\mathbf{1}\right\|^2 / 2\right) \tag{14}$$

where $X_i^m = [X_{i,1}, \ldots, X_{i,m}]^T \in \mathbb{R}^m$. The marginal distribution on arm i', given that arm i is the largest, follows a mixture distribution:

$$P(X_{i'}^m | \mathcal{H}_i) = \frac{1}{n\sqrt{2\pi}} \sum_{j=1}^n \exp\left(-\left\|X_{i'}^m - \mu_j\mathbf{1}\right\|^2 / 2\right). \tag{15}$$

Combining (13), (14), and (15), after a number of straightforward manipulations, excluded for brevity, it can be shown that the ML estimate prefers arm i to i' if and only if $\sum_{\ell=1}^m X_{i,\ell} > \sum_{\ell=1}^m X_{i',\ell}$, or equivalently, $\widehat{\mu}_i > \widehat{\mu}_{i'}$. The estimate with minimum probability of error is simply $\widehat{i} = \arg\max_i \widehat{\mu}_i$.

We continue by bounding the probability of error of the maximum likelihood test. For any estimator,

$$\mathbb{P}(\widehat{i} \neq i^*) \ \geq \ \mathbb{P}\left(\bigcup_{i \neq i^*} \widehat{\mu}_{i^*} - \widehat{\mu}_i \leq 0\right)$$

$$= \ \mathbb{P}(\widehat{\mu}_{i^*} \geq \mu_{i^*})\, \mathbb{P}\left(\bigcup_{i \neq i^*} \widehat{\mu}_{i^*} - \widehat{\mu}_i \leq 0 \, \middle|\, \widehat{\mu}_{i^*} \geq \mu_{i^*}\right)$$

$$+ \ \mathbb{P}(\widehat{\mu}_{i^*} < \mu_{i^*})\, \mathbb{P}\left(\bigcup_{i \neq i^*} \widehat{\mu}_{i^*} - \widehat{\mu}_i \leq 0 \, \middle|\, \widehat{\mu}_{i^*} < \mu_{i^*}\right)$$

$$\geq \ \frac{1}{2}\, \mathbb{P}\left(\bigcup_{i \neq i^*} \widehat{\mu}_i \geq \mu_{i^*}\right) = \frac{1}{2}\left(1 - \mathbb{P}\left(\bigcap_{i \neq i^*} \widehat{\mu}_i \leq \mu_{i^*}\right)\right)$$

$$= \ \frac{1}{2}\left(1 - \prod_{i \neq i^*} F_{\mathcal{N}}\left(\sqrt{m\Delta_i^2}\right)\right) \geq \frac{1}{2}\left(1 - \prod_{i \neq i^*}\left(1 - \frac{1}{12}\exp\left(-m\Delta_i^2\right)\right)\right) \tag{16}$$

$$\geq \ \min_{\Delta_1, \ldots, \Delta_n : \sum_i \frac{1}{\Delta_i^2} = \mathbf{H}} \frac{1}{2}\left(1 - \prod_{i \neq i^*}\left(1 - \frac{1}{12}\exp\left(-m\Delta_i^2\right)\right)\right) \tag{17}$$

11

where $F_{\mathcal{N}}(x)$ is the standard Gaussian CDF. The inequality in (16) follows since $F_{\mathcal{N}}(x) \leq 1 - \exp(-x^2)/12$ for $x \geq 0$ [11, Eqn. 13]. The next step in the proof will be showing that error probabilities smaller than a fixed constant, (17) is minimized when the gaps are equal, i.e., when $\Delta_1 = \Delta_2 = ... = \sqrt{n/\mathbf{H}}$. First define $\mathbf{z} \in \mathbb{R}_+^n$ with elements $z_i := 1/(m\Delta_i^2)$. We can recover the minimum of (17) by solving

$$\underset{\mathbf{z} \in \mathbb{R}_+^n : \mathbf{1}^T\mathbf{z} = \mathbf{H}/m}{\operatorname{argmax}} \sum_{i=1}^n \log\left(1 - \frac{1}{12}\exp\left(-1/z_i\right)\right). \tag{18}$$

Define the Lagrangian of (18) as

$$\mathcal{L}(\mathbf{z}, \lambda) = -\sum_{i=1}^n \log\left(1 - \frac{1}{12}\exp\left(z_i^{-1}\right)\right) - \lambda(\mathbf{1}^T\mathbf{z} - \mathbf{H}/m).$$

From [12, p. 321], any $\mathbf{z}$ that maximizes (18) necessarily satisfies

$$\frac{\partial \mathcal{L}}{\partial z_i} = \frac{z_i^{-2}}{12\exp(z_i^{-1}) - 1} - \lambda z_i = 0 \quad \forall\, i \tag{19}$$

$$\mathbf{1}^T\mathbf{z} = \mathbf{H}.$$

The above system of equations is satisfied by pairs $(\mathbf{z}, \lambda)$ that satisfy

$$\left\{ z_i : \lambda = \frac{z_i^{-3}}{12\exp(z_i^{-1}) - 1} \right\} \quad \forall\, i \tag{20}$$

and $\mathbf{1}^T\mathbf{z} = \mathbf{H}$ simultaneously. Differentiation of (20) shows the function $\lambda(z_i)$ is monotonically increasing in z_i for $z_i \leq 1/3$. First, consider a solution to (19) which has one or more $z_i \geq 1/3$. This would imply $m\Delta_i^2 \leq 3$ for some i, and from (17), $\mathbb{P}(\hat{i} \neq i^*) \geq \frac{1}{24}\exp(-3)$. For any λ, (20) is satisfied by at most one $z_i \in (0, 1/3]$ by the monotonicity of the function on this range; this implies implies either 1) the $\mathbf{z}$ that maximizes (18) has the form $z_1 = z_2 = \cdots = z_n$, or 2) $\mathbb{P}(\hat{i} \neq i^*) \geq \frac{1}{24}\exp(-3)$. We focus our attention on the case when $z_1 = \cdots = z_n$ (and thus $\Delta_1 = \cdots = \Delta_n$). Since $\sum_i 1/\Delta_i^2 = \mathbf{H}$, $\Delta_i = \sqrt{n/\mathbf{H}}$ for all i. (17) gives

$$\mathbb{P}(\hat{i} \neq i^*) \geq \frac{1}{2}\left(1 - \left(1 - \frac{1}{12}\exp\left(-\frac{mn}{\mathbf{H}}\right)\right)^n\right).$$

Recall the total number of samples is given by nm. If $mn \leq \mathbf{H}(\log n + \log((25\delta)^{-1}))$, then for $\delta \in (0, e^{-3}/24)$

$$\mathbb{P}(\hat{i} \neq i^*) \geq \frac{1}{2}\left(1 - \left(1 - \frac{25\delta}{12n}\right)^n\right) \tag{21}$$

$$\geq \frac{1}{2}\left(1 - \exp\left(-\frac{25\delta}{12}\right)\right) \quad \text{for all } n \geq 1 \tag{22}$$

$$\geq \delta \tag{23}$$

which completes the proof of the first statement of the theorem.

To prove the second statement of the theorem, consider the following set of gaps –

$$\Delta_i = \begin{cases} \sqrt{\frac{2}{\mathbf{H}}} & i = 1 \\ \sqrt{\frac{2(n-1)}{\mathbf{H}}} & i > 1. \end{cases}$$

Note that $\{\Delta_1, ..., \Delta_n\}$ satisfy $\sum_{i=1}^{n} 1/\Delta_i^2 = \mathbf{H}$. From (16), and by considering only the arm with the smallest gap,

$$\mathbb{P}(\widehat{i} \neq i) \geq \frac{1}{2} \left(1 - \prod_{i \neq i^*} \left(1 - \frac{1}{12} \exp\left(-m\Delta_i^2\right) \right) \right) \geq \frac{1}{24} \exp\left(-\frac{2m}{\mathbf{H}}\right). \tag{24}$$

If $m \leq \frac{\mathbf{H}}{2} \log\left(\frac{1}{24\delta}\right)$, we have $\mathbb{P}\left(\widehat{i} \neq i^*\right) \geq \delta$. This implies that if the total number of measurements is less than $\frac{\mathbf{H}n}{2} \log\left(\frac{1}{24\delta}\right)$, then $\mathbb{P}(\widehat{i} \neq i^*) \geq \delta$, completing the proof of the second statement of Thm. 4.

$\square$

A.4 Proof of Corollary 2

Proof. When $\alpha = 0$, Theorem 4 implies the result. When $\alpha > 0$, we can bound (16) by dropping all terms in the product except the term corresponding to the smallest gap. This gives

$$\mathbb{P}(\widehat{i} \neq i) \geq \frac{1}{24} \exp\left(-m\Delta_1^2\right) = \frac{1}{24} \exp\left(-mn^{-2\alpha}\right) \tag{25}$$

Setting $m \leq n^{2\alpha} \log\left(\frac{1}{24\delta}\right)$ implies the result.

$\square$