Proceedings of Machine Learning Research vol 65:1–41, 2017

The Simulator: Understanding Adaptive Sampling in the Moderate-Confidence Regime

Max Simchowitz MSIMCHOW@BERKELEY.EDU
University of California, Berkeley, CA 94720 USA

Kevin Jamieson KJAMIESON@BERKELEY.EDU
University of California, Berkeley, CA 94720 USA

Benjamin Recht BRECHT@BERKELEY.EDU
University of California, Berkeley, CA 94720 USA

Abstract

We propose a novel technique for analyzing adaptive sampling called the *Simulator*. Our approach differs from the existing methods by considering not how much information could be gathered by any fixed sampling strategy, but how difficult it is to distinguish a good sampling strategy from a bad one given the limited amount of data collected up to any given time. This change of perspective allows us to match the strength of both Fano and change-of-measure techniques, without succumbing to the limitations of either method. For concreteness, we apply our techniques to a structured multi-arm bandit problem in the fixed-confidence pure exploration setting, where we show that the constraints on the means imply a substantial gap between the moderate-confidence sample complexity, and the asymptotic sample complexity as $\delta \to 0$ found in the literature. We also prove the first instance-based lower bounds for the top-k problem which incorporate the appropriate log-factors. Moreover, our lower bounds zero-in on the number of times each *individual* arm needs to be pulled, uncovering new phenomena which are drowned out in the aggregate sample complexity. Our new analysis inspires a simple and near-optimal algorithm for the best-arm and top-k identification, the first *practical* algorithm of its kind for the latter problem which removes extraneous log factors, and outperforms the state-of-the-art in experiments.

1. Introduction

The goal of adaptive sampling is to estimate some unknown property S^* about the world, using as few measurements from a set of possible measurement actions $[n] = \{1, \dots, n\}$[1]. At each time step $t = 1, 2, \dots$, a learner chooses a measurement action $a_t \in [n]$ based on past observations, and receives an observation $X_{a_t,t} \in \mathbb{R}$. We assume that the observations are drawn i.i.d from a distribution ν_a over $\mathbb{R}$, which is unknown to the learner. In particular, the vector of distributions $\nu = (\nu_1, \dots, \nu_n)$, called the *instance*, encodes the distribution of all possible measurement actions. The instance ν can be thought of as describing the state of the world, and that our property of interest $S^* = S^*(\nu)$ is a function of the instance. We focus on what is called the *fixed-confidence pure-exploration* setting, where the algorithm decides to stop at some (possibly random) time T, and returns an output $\widehat{S}$ which is allowed to differ from $S^*(\nu)$ with probability at most δ on any instance ν. Since T is exactly equal to the number of measurements taken, the goal of adaptive

[1] We only work with finitely many measurement actions, but this may be generalized as in Arias-Castro et al. Arias-Castro et al. (2013)

pure-exploration problems is to design algorithms for which T is as small as possible, either in expectation or with high probability.

Crucially, we often expect the instance ν to lie in a known constraining set $\mathcal{S}$. This allows us to encode a broad range of problems of interest as pure-exploration multi-arm bandit (MAB) problems (Bechhofer, 1958; Even-Dar et al., 2006) with structural constraints. As an example, the adaptive linear prediction problem of (Soare et al., 2014; Lattimore and Szepesvari, 2016) (known in the literature as *linear bandits*), is equivalent to MAB, subject to the constraint that the mean vector $\mu = (\mu_1, \dots, \mu_n)$ (where $\mu_a := \mathbb{E}_{X_a \sim \nu_a}[X_a]$) lies in the subspace spanned by the rows of $V = \begin{bmatrix} v_1 & | & v_2 & | \dots | & v_n \end{bmatrix}$, where $v_1, \dots, v_n \in \mathbb{R}^d$ are the vector-valued features associated with arms 1 through n. The noisy combinatorial optimization problems of Yue and Guestrin (2011); Simchowitz et al. (2016); Gopalan et al. (2014) can be also be cast in this fashion. Moreover, by considering properties $S^*(\nu)$ other than the top mean, one can use the above framework to model signal recovery and compressed sensing (Arias-Castro et al., 2013; Castro, 2014), subset-selection (Kalyanakrishnan et al., 2012), and additional variants of combinatorial optimization (Chen et al., 2014, 2016; Kveton et al., 2014).

The purpose of this paper is to present new machinery to better understand the consequences of structural constraints $\mathcal{S}$, and types of objectives $S^*(\nu)$ on the sample complexity of adaptive learning problems. This paper presents bounds for some structured adaptive sampling problems which characterize the sample complexity in the regime where the probability of error δ is a moderately small constant (e.g. $\delta = .05$, or even inverse-polynomial in the number of measurements). In contrast, prior work has addressed the sample complexity of adaptive samplings problems in the asymptotic regime that $\delta \to 0$, where such problems often admit algorithms whose asymptotic dependence on δ matches lower bounds *for each ground-truth instance*, even matching the exact instance-dependent leading constant (Garivier and Kaufmann, 2016; Russo, 2016; Luedtke et al., 2016). Analogous asymptotically-sharp and instance-specific results (even for structured problems) also hold in the regret setting where the time horizon $T \to \infty$ (Lai and Robbins, 1985; Gopalan et al., 2014; Magureanu et al., 2014; Combes et al., 2015; Talebi and Proutiere, 2016).

The upper and lower bounds in this paper demonstrate that the $\delta \to 0$ asymptotics can paint a highly misleading picture of the true sample complexity when δ is not-too-small. This occurs for two reasons:

1. Asymptotic characterizations of the sample complexity of adaptive estimation problems occur on a time horizon where the learner can learn an optimal measurement allocation tailored to the ground truth instance ν. In the short run, however, learning favorable measurement allocations is extremely costly, and the learning good allocations requires considerably more samples to learn than it itself would prescribe.

2. Asymptotic characterizations are governed by the complexity of discriminating the ground truth ν from any single, alternative hypothesis. This neglects multiple-hypothesis and suprema-of-empirical-process effects that are ubiquitous in high-dimensional statistics and learning theory (e.g. those reflected in Fano-style bounds).

To understand these effects, we introduce a new framework for analyzing adaptive sampling called the “Simulator”. Our approach differs from the existing methods by considering not how much information could be gathered by any fixed sampling strategy, but how difficult it is to distinguish a good sampling strategy from a bad one, given any limited amount of data collected up to any given time. Our framework allows us to characterize granular, instance dependent properties that any successful adaptive learning algorithm must have. In particular, these insights inspire a new, theoretically near-optimal, and practically state-of-the-art algorithm for the top-k subset selection problem. We emphasize that the Simulator framework is concerned with how an algorithm samples, rather than its final objective. Thus, we believe that the techniques in this paper can be applied more broadly to a wide class of problems in the active learning community.

After defining terms and the setting of interest in Section 2, Section 3 reviews the state-of-the-art lower bounds and their limitations, and then presents our novel lower bounds for the special case when the means are known up to a permutation. Section 3.1 explores conditions under which $\log$ factors appear in lower bounds and we leverage these observations to prove instance-specific lower bounds for top-k subset selection in Section 3.2. Inspired by the lower bounds, we introduce LUCB++, the first practical, minimax-optimal algorithm for top-k subset selection in Section 4 (proofs of sample complexity guarantees are deferred to Appendix E). The Simulator framework and its application to the lower bound when the means are known up to a permutation is presented in Sections 5 and 6, with some proofs being deferred to Appendix A. The lower bounds for top-k subset slection require more careful analysis, and are deferred to the Appendices B, C, and D. Finally, we make concluding remarks in Section 7.

2. Preliminaries

As alluded to in the introduction, the adaptive estimation problems in this paper can be formalized as multi-arm bandits problems, where the instances $\nu = (\nu_1, \dots, \nu_n)$ lie in an appropriate constraint set $\mathcal{S}$, called an instance class (e.g., the mean vectors $(\mu_1, \dots, \mu_n)$, where $\mu_a := \mathbb{E}_{X_a \sim \nu_a}[X_a]$ lie in some specified polytope). We use the term *arms* to refer both to the indices $a \in [n]$ and distributions ν_a they index. The stochastic multi-arm bandit formulation has been studied extensively in the pure-exploration setting considered in this work (Bechhofer, 1958; Even-Dar et al., 2006; Kalyanakrishnan et al., 2012; Karnin et al., 2013; Jamieson et al., 2014; Chen and Li, 2015; Garivier and Kaufmann, 2016; Russo, 2016). At each time $t = 1, 2, \dots$, a learner plays an action $a_t \in [n]$, and observes an observation $X_{a_t,t} \in \mathbb{R}$ drawn i.i.d from ν_{a_t}. At some time T, the learner decides to end the game and return some output. Formally, let $\mathcal{F}_t$ denote the sigma-algebra generated by $\{X_{a_s,s}\}_{1 \le s \le t}$, and some additional randomness ξ_{Alg} independent of all the samples (this represents randomization internal to the algorithm). A *sequential sampling algorithm* consists of

1. A sampling rule $(a_t)_{t \in \mathbb{N}}$, where $a_t \in [n]$ is $\mathcal{F}_{t-1}$ measurable.

2. A stopping time T, which is $\{\mathcal{F}_t\}_{t \in \mathbb{N}}$-measurable.

3. An output rule $\widehat{S} \subset [n]$, which is $\mathcal{F}_T$-measurable.

We let $N_a(t) = \sum_{s=1}^{t} \mathbb{I}(a_s = a)$ denote the samples collected from arm $a \in \mathcal{A}$ by time t. In particular, $N_a(T)$ is the number of times arm a is pulled by the algorithm before terminating, and $\sum_{a=1}^{n} N_a(T) = T$. A MAB algorithm corresponds to the case where the decision rule is a singleton $\widehat{S} \in \binom{[n]}{1}$, and, more generally, a TopK algorithm specifies a $\widehat{S} \in \binom{[n]}{k}$. We will use Alg as a variable which describes a particular algorithm, and use the notation $\mathbb{P}_{\nu,\mathsf{Alg}}[\cdot]$ and $\mathbb{E}_{\nu,\mathsf{Alg}}[\cdot]$ to denote probabilities and expectations which are taken with respect to the samples drawn from ν, and the (possibly randomized) sampling, stopping, and output decisions made by Alg. Finally, we adopt the following notion of correctness, which corresponds to the "fixed-confidence" setting in the active learning literature:

Definition 1 *We say that a* MAB *algorithm is* δ*-correct for a best-arm mapping* $a^* : \mathcal{S} \to [n]$ *(resp* δ*-correct for a* TopK *mapping* $S^* : \mathcal{S} \to \binom{[n]}{k}$*) over an instance class* $\mathcal{S}$ *if for all* $\nu \in \mathcal{S}$, $\mathbb{P}_{\nu,\mathsf{Alg}}[\widehat{S} = a^*(\nu)] \geq 1 - \delta$ *(resp.* $\mathbb{P}_{\nu,\mathsf{Alg}}[\widehat{S} = S^*(\nu)] \geq 1 - \delta$*).*

Typically, the best arm mapping is defined as the arm with the highest mean $a^* = \arg\max_{a \in [n]} \mu_a$, and top k mapping as the arms with the k-largest means $\arg\max_{S \in \binom{[n]}{k}} \sum_{a \in S} \mu_a$, which captures the notion of the arm/set of arms that yield the highest reward. When the best-arm mapping returns the highest-mean arm, and the observations X_b are sub-Gaussian[2], the problem complexity for MAB is typically parameterized in terms of the "gaps" between the means $\Delta_b := \mu_{a^*} - \mu_b$ (Mannor and Tsitsiklis, 2004). More generally, sample complexity is parametrized in terms of the $\mathrm{KL}(\nu_b, \nu_{a^*})$, the KL divergences between the measures ν_{a^*} and ν_b. For ease of exposition, we will present our high-level contributions in terms of gaps, but the body of the work will also present more general results in terms of KL's. Finally, our theorem statements will use $\gtrsim$ and $\lesssim$ to denote inequalities up to constant factors. In the text, we shall occasionally use $\gtrsim, \lesssim, \approx$ more informally, hiding doubly-logarithmic factors in problem parameters.

3. Statements of Lower Bound Results

Typically, lower bounds in the bandit and adaptive sampling literature are obtained by the change of measure technique (Mannor and Tsitsiklis, 2004; Castro, 2014; Garivier and Kaufmann, 2016). To contextualize our findings, we begin by stating the state-of-the-art change-measure-lower bounds, as it appears in Garivier et al. (2016). For a class of instances $\mathcal{S}$, let $\mathrm{Alt}(\nu)$ denote the set of instances $\widetilde{\nu} \in \mathcal{S}$ such that, $a^*(\widetilde{\nu}) \neq a^*(\nu)$. Then:

Proposition 2 (Theorem 1 (Garivier and Kaufmann, 2016)) *If* Alg *is* δ *correct for all* $\nu \in \mathcal{S}$*, then the expected number of samples* Alg *collects under* ν, $\mathbb{E}_{\nu,\mathsf{Alg}}[T]$*, is bounded below by the solution to the following optimization problem*

$$\min_{\tau \in \mathbb{R}^n_{\geq 0}} \sum_{a=1}^n \tau_a \quad \text{subject to} \quad \inf_{\tilde{\nu} \in \mathrm{Alt}(\nu)} \sum_{a=1}^n \tau_a \mathrm{KL}(\nu_a, \tilde{\nu}_a) \geq \mathrm{kl}(\delta, 1-\delta) \tag{1}$$

where $\mathrm{kl}(\delta, 1-\delta) := \delta \log(\frac{\delta}{1-\delta}) + (1-\delta)\log(\frac{1-\delta}{\delta})$*, which scales like* $\log(1/\delta)$ *as* $\delta \to 0$*.*

The above proposition says that the expected sample complexity $\mathbb{E}_{\nu,\mathsf{Alg}}[T]$ is lower bounded by the following, non-adaptive experiment design problem: minimize the total number of samples $\sum_a \tau_a$ subject to the constraint that these samples can distinguish between a null hypothesis $H_0 = \nu$, and any alternative hypothesis $H_1 = \tilde{\nu}$ for $\tilde{\nu} \in \nu$, with Type-I and Type-II errors at most δ. We will call the optimization problem in Equation 1 the *Oracle Lower Bound*, because it captures the best sampling complexity that could be attained by a powerful "oracle" who knows how to optimally sample under ν.

Unlike the oracle, a real learner would never have access to the true instance ν. Indeed, for MAB instances with sufficient structure, Equation 1 gives a misleading view of the instrinsic difficulty of

[2] Formally, X_b is σ^2-sub-Gaussian if $\mathbb{E}_{X_b \sim \nu_b}[e^{\lambda(X_b - \mu_b)}] \leq \exp(\lambda^2\sigma^2/2)$

the problem. For example, let $\mathcal{S}$ denote the class of instances ν where $\nu_a = \mathcal{N}(\mu_a, 1)$, and μ lies in the simplex, i.e. $\mu_a \geq 0$ and $\sum_{a \in \mathcal{A}} \mu_a = 1$. If the ground truth instance ν^* has $\mu_{a^*} = .9$ for some $a^* \in [n]$, then any oracle which uses the knowledge of the ground truth to construct a sampling allocation can simply put all of its samples on arm a^*. Indeed, the simplex constraint implies that a^* is indeed the best arm of ν, and that any instance $\widetilde{\nu}$ which has a best arm other than a^* must have $\widetilde{\nu}_{a^*} < .5$. Thus, $\forall\, \widetilde{\nu} \in \mathrm{Alt}(\nu)$, $\mathrm{KL}(\nu^*_{a^*}, \widetilde{\nu}^*) \geq \frac{(.9-.5)^2}{2} = \Omega(1)$. In other words, the sampling vector

$$\tau_a = \begin{cases} (.08)^{-1}\mathrm{kl}(\delta, 1-\delta) & a = a^* \\ 0 & a \neq a^* \end{cases} \tag{2}$$

is feasible for Equation 1 which means that the optimal number of samples predicted by Equation 1 is no more than $\sum_a \tau_a = \tau_{a^*} = O(\log(1/\delta))$. But this predicted sample complexity doesn't depend on the number of arms!

So how how hard is the simplex really? To address this question, we prove the first lower bound in the literature which, to the author's knowledge [3], accurately characterizes the complexity a strictly easier problem: when the means are known up to a permutation. Because the theorem holds when the measures are known up to a permutation, it also holds in the more general setting when the measures satisfy any permutation-invariant constraints, including when **a)** the means lie on the simplex **b)** the means lie in an l_p ball or **c)** the vector $\mu_{(1)} \geq \mu_{(2)} \geq \ldots \mu_{(n)}$ of sorted means satisfy arbitrary constraints (e.g. weighted l_p constraints on the sorted means (Bogdan et al., 2013)).

In what follows, let $\mathbf{S}_n$ denote the group of permutations on $[n]$ elements and $\pi(j)$ denote the index which j is mapped to under π. For an instance $\nu = (\nu_1, \ldots, \nu_n)$, we let $\pi(\nu) = \{\nu_{\pi(1)}, \ldots, \nu_{\pi(n)}\}$, and define the instance class $\mathbf{S}_n(\nu) := \{\pi(\nu), \pi \in \mathbf{S}_n\}$. Moreover, we use the notation $\pi \sim \mathbf{S}_n$ to denote that π is drawn uniformly at random. With this notation, $N_{\pi(b)}$ is the number of times we pull the arm indexed by $\pi(b) \in [n]$, i.e. the samples from $\nu_{\pi(b)}$. And $\mathbb{E}_{\pi \sim \mathbf{S}_n}[N_{\pi(b)}(T)]$ is the expected number of samples from ν_b since $(\pi(\nu))_{\pi(b)}$ is always equal to ν_b, and *not* the distribution $\nu_{\pi(b)}$. The following theorem essentially says that if the instance is randomly permuted before the start of the game, no δ-correct algorithm can avoid taking a substantial number of samples from ν_b for any $b \in [n]$.

Theorem 3 (Lower bounds on Permutations) *Let ν be an* MAB *instance with unique best arm a^*, and for $b \neq a^*$, define $\tau_b = \frac{1}{\mathrm{KL}(\nu_{a^*}, \nu_b) + \mathrm{KL}(\nu_b, \nu_{a^*})}$. If* Alg *is δ-correct over $\mathbf{S}_n(\nu)$ then*

$$\mathbb{E}_{\pi \sim \mathbf{S}_n} \mathbb{P}_{\pi(\nu), \mathsf{Alg}}[N_{\pi(b)}(T) > \tau_b \log(1/4\eta)] \geq \eta - \delta \tag{3}$$

for any $\nu \in (\delta, 1/4)$, and by Markov's inequality

$$\mathbb{E}_{\pi \sim \mathbf{S}_n} \mathbb{E}_{\pi(\nu), \mathsf{Alg}}[T] = \mathbb{E}_{\pi \sim \mathbf{S}_n}\left[\sum_{b \neq a^*} \mathbb{E}_{\pi(\nu), \mathsf{Alg}}[N_{\pi(b)}(T)]\right] \geq \sup_{\eta \in [\delta, 1/4]} (\eta - \delta) \log(1/4\eta) \sum_{b \neq a^*} \tau_b. \tag{4}$$

In particular, if Alg *is $\delta \leq 1/8$-correct, then* $\mathbb{E}_{\pi \sim \mathbf{S}_n} \mathbb{E}_{\pi(\nu), \mathsf{Alg}}[T] \gtrsim \sum_{b \neq a^*} \frac{1}{\mathrm{KL}(\nu_{a^*}, \nu_b) + \mathrm{KL}(\nu_b, \nu_{a^*})}$.

[3] Before publication this work, but after a preprint was available online, this result was obtained independently by Chen et al. (2017)

The proof of the above result is found in Section 6.2 and follows from the application of the Simulator introduced in Section 5, and machinery developed throughout Section 6. When the reward distributions are $\nu_b = \mathcal{N}(\mu_b, 1)$, $\mathrm{KL}(\nu_{a^*}, \nu_b) = \mathrm{KL}(\nu_b, \nu_{a^*}) = \frac{1}{2}\Delta_b^2$ (recall $\Delta_b = \mu_{a^*} - \mu_b$). Moreover, applying the oracle bound of Proposition 2 to permutations implies a lower bound of $\gtrsim \max_{b \neq a^*} \Delta_b^{-2} \log(1/\delta)$. Indeed, for each $b \in [n] \setminus \{a^*\}$, one would need to take enough samples to distinguish ν from the alternative instance where the means μ_{a^*} and μ_b are swapped, with probability of error at most $1 - \delta$. Hence, combining this oracle lower bound with Theorem 3 yields

$$\mathbb{E}_{\pi \sim \mathbf{S}_n} \mathbb{E}_{\pi(\nu), \mathsf{Alg}}[T] \gtrsim \max\{\max_{b \neq a^*} \Delta_b^{-2} \log(1/\delta), \sum_{b \neq a^*} \Delta_b^{-2}\} \,. \tag{5}$$

For comparison, the bound of Proposition 2 only implies a lower bound of $\gtrsim \max_{b \neq a^*} \Delta_b^{-2} \log(1/\delta)$, since an oracle who knows how to sample could place all their samples on a^*. Thus, for constant $\log(1/\delta)$, our lower bound differs from the bound in Proposition 2 by up to a factor of n, the number of arms. In particular, when the gaps are all on the same order, the $\delta \to 0$ asymptotics only paint an accurate picture of the sample complexity once δ is exponentially-small in n.

In fact, our lower bound is essentially unimproveable: Appendix E.1 provides an upper bound for the setting where the top-two means are known, whose expected sample complexity on any permutation matches the on-average complexity in Equation 5 up to constant and doubly-logarithmic factors. Together, these upper and lower bounds depict two very different regimes:

1. Treating δ as a fixed constant, the lower bound of the constrained problem essentially matches known upper bounds for the *unconstrained* best-arm problem (Chen and Li, 2015; Jamieson et al., 2014). Thus, in this regime, *knowing the instance up to a permutation of the arms does not affect the sample complexity*.

2. As $\delta \to 0$, an algorithm which knows the means up to a permutation can learn to optimistically and aggressively focus its samples on the top arm, yielding an asymptotic sample complexity predicted by Proposition 2, one which is potentially far smaller than that of the unconstrained problem.[4]

These two regimes show that the Simulator and oracle lower bounds are *complementary*, and go after two different aspects of problem difficulty: In the second regime, the oracle lower bound characterizes $\lesssim \max_{b \neq a^*} \Delta_b^{-2} \log(1/\delta)$ samples sufficient to *verify* that arm a^* is the best, whereas in the first regime, the Simulator characterizes the $\gtrsim \sum_{b \neq a^*} \Delta_b^{-2}$ samples needed to learn a favorable sampling allocation[5]. We remark that Garivier et al. Garivier et al. (2016) also explores the problem of learning-to-sample by establishing the implications of Proposition 2 for finite-time regret; however, there approach does not capture any effects which aren't reflected in Proposition 2. Moreover, Bubeck and Cesa-Bianchi (Bubeck and Cesa-Bianchi, 2012) establish an *minimax*, rather than instance-specific, lower bound for regret by considering permutations of a simple MAB instance where $n - 1$ arms have the name mean, and one arm has a slightly elevated mean. Finally, we

[4] In fact, using a track-and-stop strategy similar to Garivier and Kaufmann (2016) one could design an algorithm which matches the constant factor in Proposition 2.

[5] The simulator also provides a lower bound on the *tail* of the number of pulls from a suboptimal arm since, with probability δ, arm b is pulled $\tau \log(1/8\delta)$ times. This shows that even though you can learn an oracle allocation on average, there is always a small risk of oversampling. Such affects do not appear from Proposition 2, which only control the number of samples taken in expectation

note that proving a lower bound for learning a favorable strategy in our setting must consider some sort of average or worst-case over the instances. Indeed, one could imagine an algorithm that starts off by pulling the first arm 1 until it has collected enough samples to test whether $\mu_1 = \mu_{a^*}$ (i.e. $\mu > \max_{b \neq a^*} \mu_b$), and then pulling arm 2 to test whether $\mu_2 = \mu_{a^*}$, and so on. If arm 1 is the best, this algorithm can successfully identify it without pulling any of the others, thereby matching the oracle lower bound.

3.1. Sharper Multiple-Hypothesis Lower Bounds

In contrast to change-of-measure type lower bounds like Proposition 2, the active PAC learning literature (e.g., binary classification) leverages classical tools like Fano's inequality with packing arguments (Castro and Nowak, 2008; Raginsky and Rakhlin, 2011) and other measures of class complexity such as the disagreement coefficient (Hanneke, 2009) or the eluder dimension Russo and Van Roy (2013). Because these arguments consider multiple hypotheses simultaneously, they capture effects which worst-case binary-hypothesis oracle lower bounds like Equation 1 can miss.

While the considerable gap between two-way and multiple tests is well-known in the passive setting (Tsybakov, 2009), existing techniques which capture this multiple-hypothesis complexity lead to coarse, worst- or average-case lower bounds for adaptive problems because they rely on constructions which are either artificially symmetric, or are highly pessimistic (Castro and Nowak, 2008; Raginsky and Rakhlin, 2011; Kalyanakrishnan et al., 2012). Moreover, the constructions rarely shed insights on *why* active learning algorithms seem to avoid paying the costs for multiple hypotheses that would occur in the passive setting, e.g. the folk theorem: "active learning removes log factors" (Castro, 2014).

As a first step towards understanding these effects, we prove the first instance-based lower bound which sheds light on why active learning is able to effectively reduce the number of hypotheses it needs to distinguish. To start, we prove a qualitative result for a simplified problem, using a novel reduction to Fano's inequality via the simulator. The following theorem is proved in Appendix B:

Theorem 4 *Let* Alg *be* $1/8$*-correct, consider a game with best arm* ν_1 *and* $n-1$ *arms of measure* ν_2. *Let* $S_m := \{a \in [n] : N_a(T) > \frac{1}{16}(\mathrm{KL}(\nu_1,\nu_2) + \mathrm{KL}(\nu_2,\nu_1)) \log \frac{n}{2^{16}m}\}$. *Then*

$$\mathbb{P}_{\pi \sim \mathbf{S}_n} \mathbb{P}_{\pi(\nu),\mathsf{Alg}} \left[\{\pi(1) \in S_m\} \wedge \{|S_m| \geq m\} \right] \geq \frac{3}{4} \tag{6}$$

For Gaussian rewards with unit variance, $\mathrm{KL}(\nu_1,\nu_2) + \mathrm{KL}(\nu_2,\nu_1) = \Delta^2$, where Δ is the gap between the means $\mu_1 - \mu_2$, the above proposition states that, for any $m \in [n]$, any correct MAB algorithm must sample some m arms, including the top arm, $\tau \gtrsim \Delta^{-2}\log(n/m)$ times. Thus, the number of samples allocated by the oracle of Proposition 2 are necessarily insufficient to identify the best arm for moderate δ. This is because, until sufficiently many samples has been taken, one cannot distinguish between the best arm, and other arm exhibiting large statistical deviations. Looking at exponential-gap style upper bounds (Chen and Li, 2015; Karnin et al., 2013), which halve the number of arms in consideration at each round, we see that our lower bound is qualitatively sharp for some algorithms[6]. Further, we emphasize that this set of m arms which must be pulled τ times may be random[7], depend on the random fluctuations in the samples collected, and thus cannot be

[6] We believe that UCB-style algorithms exhibit this same qualitative behavior

[7] In fact, for an algorithm with which only samples $m' = O(m)$ arms $\tau \gtrsim \Delta^{-2}\log(n/m)$, this subset of arms *must* be random. This is because for a fixed subset of m' arms, one could apply Theorem 4 to the remaining $n - m'$ arms.

determined using knowledge of the instance alone. Stated otherwise, if one sampled according to the *proporitions* as ascribed by Proposition 2, then the total number of samples one would need to collect would be suboptimal (by a factor of $\log n$). Thus, effective adaptive sampling should adapt its allocation to the statistical deviations in the collected data, not just the ground truth instance. We stress that the Simulator is indispensable for establishing this result, because it lets us characterize the stage-wise sampling allocation of adaptive algorithms.

Guided by this intuition, Appendix C employs a more involved proof strategy to establish the following guarantee for MAB with Gaussian rewards (a more general result for single-parameter exponential families is given by Theorem 19 in Appendix C.3):

Proposition 5 (Lower Bound for Gaussian MAB) *Suppose $\nu = (\nu_1, \dots, \nu_n)$ has measures $\nu_a = \mathcal{N}(\mu_a, 1)$,with $\mu_1 > \mu_2 \geq \dots \mu_n$. Then, if* Alg *is $\delta \leq 1/16$ correct over $\mathbf{S}_n(\nu)$,*

$$\mathbb{E}_{\pi \sim \mathbf{S}_n} \mathbb{E}_{\pi(\nu^{(1)}),\mathsf{Alg}}[N_{\pi(1)}(T)] \quad \gtrsim \quad \max_{2 \leq m \leq n} \Delta_m^{-2} \log(m/\delta) \quad \textit{where } \Delta_m = \mu_1 - \mu_m \tag{7}$$

In particular, when all the gaps are on the same order Δ, then the top arm must be pulled $\Omega(\Delta^{-2} \log n)$ times. When the gaps are different, $\max_{2 \leq m \leq n} \Delta_m^{-2} \log m$ trades off between larger $\log m$ factor as the inverse-gap-squared Δ_m^{-2} shrinks. As we explain in Appendix C.1, this tradeoff is best understood in the sense that the algorithm is conducting an *instance-dependent* union bound, where the union bound places more confidence on means closer to the top. The proof itself is quite involved, and constitutes the main technical contribution of this paper. We devote Section C.1 to explaining the intuition and proof roadmap. Our argument makes use of "tilted distributions", which arise in Herbst Argument in Log-Sobolev Inequalities in the concentration-of-measure literature (Raginsky and Sason, 2014). Tiltings translate the tendency of some empirical means to deviate far above their averages (i.e. to anti-concentrate) into a precise information-theoretic statement that they "look like" draws from the top arm. To the best of our knowledge, this constitutes the first use of tiltings to establish information-theoretic lower bounds, and we believe this strategy may have broader use.

3.2. Instance-Specific Lower bound for TopK

Proposition 5 readily implies the first instance-specific lower bound for the TopK. The idea is that, if I can identify an arm $j \in [k]$ as one of the top k arms, then, in particular, I can identify arm j as the best arm among $\{j\} \cup \{k+1, \dots, n\}$. Similarly, if I can reject arm ℓ as not part of the top k, then I can identify it as the "worst" arm among $\{1, \dots, k\} \cup \{\ell\}$. Section D formally proves the following lower bound using by applying the above eduction to Proposition 5:

Proposition 6 (Lower Bound for Gaussian TopK) *Suppose $\nu = (\nu_1, \dots, \nu_n)$ has measures $\nu_a = \mathcal{N}(\mu_a, \theta)$, with $\mu_1 \geq \mu_2 \geq \dots \mu_k > \mu_{k+1} \geq \dots \mu_n$. Then, if* Alg *is $\delta \leq 1/16$ correct over $\mathbf{S}_n(\nu)$,*

$$\mathbb{E}_{\pi \sim \mathbf{S}_n} \mathbb{E}_{\pi(\nu^{(1)}),\mathsf{Alg}}[N_{\pi(j)}(T)] \quad \gtrsim \quad \begin{cases} \max_{m > k} (\mu_j - \mu_m)^{-2} \log((m-k+1)/\delta) & j \leq k \\ \max_{m \leq k} (\mu_j - \mu_m)^{-2} \log((k+2-m)/\delta) & j > k \end{cases} \tag{8}$$

By taking $m = k+1$ and $m = k$ in the first and second lines of 8, our result recovers the gap-dependent bounds of Kalyanakrishnan et al. (2012) and Luedtke et al. (2016) . Moreover, when the gaps are on the same order Δ, we recover the worst-case lower bound from Kalyanakrishnan et al. (2012) of $k\Delta^{-2} \log(n-k) + (n-k)\Delta^{-2} \log k$.

3.2.1. COMPARISON WITH CHEN ET AL. CHEN ET AL. (2017)

After a manuscript of the present work was posted on one of its authors' websites, Chen et al. (2017) presented an alternative proof of Proposition 6, also by a reduction to MAB. Instead of tiltings, their argument handles different gaps by a series of careful reductions to a symmetric MAB problem, to which they apply Proposition 1. As in this paper, their proof hinges on a "simulation" argument which compares the behavior of an algorithm on an instance ν to a run of an algorithm where the reward distributions change mid-game. This seems to suggest that our simulator framework is in some sense a natural tool for these sorts of lower bounds.

While our works prove many of the same results, our papers differ considerably in emphasis. The goal for in this work is to explain *why* algorithms must incur the sample complexities that they do, rather than just sharpen logarithmic factors. In this vein, we establish Theorem 4, which has no analogue in Chen et al. (2017). Moreover, we believe that the proof of Proposition 5 based on tiltings is a step towards novel lower bounds for more sophisticated problems by translating intuitions about large-deviations into precise, information-theoretic statements. Further still, our Theorem 3 (and Proposition 16 in the appendix) imply lower bounds on the tail-deviations of the number of times suboptimal arms need to be sampled in constrained problems (see footnote 5).

4. LUCB++

The previous section showed that for TopK in the worst case, the bottom $(n-k)$ arms must be pulled in proportion to $\log(k)$ times while the top k arms must be pulled in proportion to $\log(n-k)$ times. Inspired by these new insights, the original LUCB algorithm of Kalyanakrishnan et al. (2012), and the analysis of Jamieson et al. (2014) for the MAB setting, in this section we propose a novel algorithm for TopK: LUCB++. The LUCB++ algorithm proceeds exactly like that of Kalyanakrishnan et al. (2012), the only difference being the definition of the confidence bounds used in the algorithm.

At each round $t = 1, 2, \ldots$, let $\widehat{\mu}_{a,N_a(t)}$ denote the empirical mean of all the samples from arm a collected so far. Let $U(t, \delta) \propto \sqrt{\frac{1}{t}\log(\log(t)/\delta)}$ be an anytime confidence bound based on the law of the iterated logarithm (see Kaufmann et al. (2015, Theorem 8) for explicit constants). Finally, we let TOP_t denote the set of the k arms with the largest empirical means. The algorithm is outlined in Figure 1, and satisfies the following guarantee:

Theorem 7 *Suppose that $X_a \sim \nu_a$ is $1-$subgaussian. Then, for any $\delta \in (0, 1)$, the LUCB++ algorithm is δ-correct, and the stopping time T satisfies*

$$T \le \sum_{i=1}^{k} c\Delta_i^{-2} \log\left(\tfrac{(n-k)\log(\Delta_i^{-2})}{\delta}\right) + \sum_{j=k+1}^{n} c\Delta_j^{-2} \log\left(\tfrac{k\log(\Delta_j^{-2})}{\delta}\right)$$

with probability at least $1-\delta$, where c is a universal constant.

By Propositions 6 we recognize that when the gaps are all the same the sample complexity of the LUCB++ algorithm is unimprovable up to $\log\log(\Delta_i)$ factors. This is the first *practical* algorithm that removes extraneous log factors on the sub-optimal $(n-k)$ arms Kalyanakrishnan et al. (2012); Chen et al. (2016). However, it is known that not all instances must incur a multiplicative $\log(n-k)$ on the top k arms Chen et al. (2016, 2017). Indeed, when $k = 1$ this problem is just the best-arm identification problem and the sample complexity of the above theorem, ignoring doubly

Algorithm 1: LUCB++

```
Input Set size $k$, confidence $\delta$, confidence interval $U(\cdot,\delta)$
Play Each arm $a \in [n]$ once
For rounds $t = n+1, n+2, \ldots$
    Let $\mathrm{TOP}_t = \arg\max_{S\subset[n]:|S|=k} \sum_{i\in S} \widehat{\mu}_{a,N_a(t)}$,
    If the following holds, Then return $\mathrm{TOP}_t$:
        $\min_{a\in\mathrm{TOP}_t} \widehat{\mu}_{a,N_a(t)} - U(N_a(t), \frac{\delta}{2(n-k)}) > \max_{a\in[n]-\mathrm{TOP}_t} \widehat{\mu}_{a,N_a(t)} + U(N_a(t), \frac{\delta}{2k})$    (9)
    Else pull $h_t$ and $l_t$, given by:
        $h_t := \min_{a\in\mathrm{TOP}_t} \widehat{\mu}_{a,N_a(t)} - U(N_a(t), \frac{\delta}{2(n-k)})$    $l_t := \max_{a\in[n]-\mathrm{TOP}_t} \widehat{\mu}_{a,N_a(t)} + U(N_a(t), \frac{\delta}{2k})$.
```

n	LUCB++	LUCB	Oracle	Uniform
10^1	1.0	0.99	1.60	1.67
10^2	1.0	1.17	2.00	3.4
10^3	1.0	1.50	2.51	5.32
10^4	1.0	1.89	2.90	7.12
10^5	1.0	2.09	3.32	8.49

Table 1: The number of samples taken by the algorithms before reaching their stopping condition, relative to LUCB++.

logarithimc factors, scales like $\log(n/\delta)\Delta_1^{-2} + \log(1/\delta)\sum_{i=2}^n \Delta_i^{-2}$. But there exist algorithms for this particular best-arm setting whose sample complexity is just $\log(1/\delta)\sum_{i=1}^n \Delta_i^{-2}$ exposing a case where Theorem 7 is loose Karnin et al. (2013); Jamieson et al. (2014); Chen and Li (2015); Chen et al. (2016). In general, this additional $\log(n-k)$ factor is unnecessary on the top k arms when $\sum_{i=1}^k \Delta_i^{-2} \gg \sum_{i=k+1}^n \Delta_i^{-2}$, but for large n, this is a case unlikely to be encountered in practice.

While this manuscript was in preparation, Chen et al. (2017) proposed a TopK algorithm which satisfies stronger theoretical guarantees, essentially matching the lower bound in Theorem 6. However, their algorithm (and the matroid-bandit algorithm of Chen et al. (2016)) relies on exponential-gap elimination, making it unsuitable for practical use[8]. Furthermore, our improved LUCB++ confidence intervals can be reformulated for different KL-divergences, leading to tighter bounds for non-Gaussian rewards such as Bernoullis. Moreover, we can “plug in” our LUCB++ confidence intervals into other LUCB-style algorithms, sharpening their log factors. For example, one could ammend the confidence intervals in the CLUCB algorithm of Chen et al. (2014) for combinatorial bandits, which would yield slight improvements for arbitrary decision classes, and near-optimal bounds for matroid classes considered in (Chen et al., 2016).

[8] While exponential-gap elimination algorithms might have the correct dependence on problem parameters, their constant-factors in the sample complexity are incredibly high, because they rely on the median-elimination as a subroutine (see Jamieson et al. (2014) for discussion)

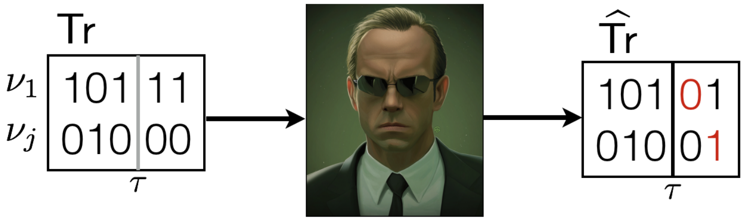


Figure 1: The Simulator acts as a man-in-the-middle between the original transcript and the transcript the algorithm receives. It leaves the transcript unchanged before some time τ, but modifies it in arbitrary ways after this time. Red denotes the samples that were changed that reduced the distance between the instances. Note that all events defined on just the first τ samples are truthful.

To demonstrate the effectiveness of our new algorithm we compare to a number of natural baselines: LUCB of Kalyanakrishnan et al. (2012), a TopK version of the oracle strategy of Garivier and Kaufmann (2016), and uniform sampling; all three use the stopping condition of Kalyanakrishnan et al. (2012) which is when the empirical top k confidence bounds[9] do not overlap with the bottom $n-k$, employing a union bound over all n arms. Consider a TopK instance for $k=5$ constructed with unit-variance Gaussian arms with $\mu_i = 0.75$ for $i \leq k$ and $\mu_i = 0.25$ otherwise. Table 1 presents the average number of samples taken by the algorithms before reaching the stopping criterion, relative to the the number of samples taken by LUCB++. For these means, the oracle strategy pulls each arm i a number of times proportional to w_i where $w_i = \frac{\sqrt{n/k-1}-1}{n-2k}$ for $i \leq k$ and $w_i = \frac{1-kw_k}{n-k}$ for $i > k$ ($w_i = 1/n$ for all i when $n = 2k$). Note that the uniform strategy is indentical to the oracle strategy, but with $w_i = 1/n$ for all i.

5. Lower Bounds via The Simulator

As alluded to in the introduction, our lower bounds treat adaptive sampling decisions made by the algorithm as hypothesis tests between different instances ν. Using a type of gadget we call a *Simulator*, we reduce lower bounds on *adaptive* sampling strategies to a family of lower bounds on different, possibly data-dependent and time-specific *non-adaptive* hypothesis testing problems.

The Simulator acts as an adversarial channel intermediating between the algorithm Alg, and i.i.d samples from the true instance ν. Given an instance ν, let $\mathsf{Tr} = \{X_{[a,s]}\}_{a\in[n],s\in\mathbb{N}} \in \mathbb{R}^{n\times\mathbb{Z}_{\geq 0}}$ denote a random transcript of an infinite sequence of samples drawn i.i.d from ν, where $\mathsf{Tr}_{a,s} = X_{[a,s]} \overset{iid}{\sim} \nu_a$. We can think of any sequential sampling algorithm Alg as operating by interacting with the transcript, where the sample $X_{a_t,t}$ is obtained by reading the sample $X_{[a_t,N_{a_t}(t)]}$ off from Tr (recall that $N_a(t)$ is the number of times arm a has been pulled at the end of round t). With this notation, we define a simulator as follows:

Definition 8 (Simulator) *A simulator* Sim *is a map which sends* Tr *to a modified transcript* $\widehat{\mathsf{Tr}} = \{\widehat{X}_{[a,s]}\}_{a\in[n],s\in\mathbb{N}}$*, which* Alg *will interact with instead of* Tr *(Figure 1). We allow this mapping to depend on the ground truth* ν *and some internal randomness* ξ_{Sim}.

[9] To avoid any effects due to the particular form of the any-time confidence bound used, we use the same finite-time law-of-the-iterated logarithm confidence bound used in (Kaufmann et al., 2015, Theorem 8) for all of the algorithms.

Equivalently, $\mathsf{Sim}(\nu)$ is a measure on a *random process* $\widehat{\mathsf{Tr}} = \{\widehat{X}_{[a,s]}\}_{a\in[n],s\geq 1}$, which, unlike ν, does not require the samples $\widehat{X}_{[a,1]}, \widehat{X}_{[a,2]}, \dots$ to be i.i.d (or even independent). Hence, we use the shorthand $\mathsf{Sim}(\nu)$ to refer the measure corresponding to $\mathbb{P}_{\mathsf{Sim}(\nu),\mathsf{Alg}}$, and let $\mathbb{P}_{\mathsf{Sim}(\nu),\mathsf{Alg}}$ denote the probability taken with respect to Sim's modified transcript $\widehat{\mathsf{Tr}}$, and the internal randomness in Alg and Sim. With this notation, the quantities $\mathrm{TV}(\mathsf{Sim}(\nu), \mathsf{Sim}(\nu'))$ and $\mathrm{KL}(\mathsf{Sim}(\nu), \mathsf{Sim}(\nu'))$ are well defined as the TV and KL divergences of the random process $\widehat{\mathsf{Tr}}$ under the measures $\mathsf{Sim}(\nu)$ and $\mathsf{Sim}(\nu')$.

Note that, in general, $\mathrm{TV}(\nu, \nu') = \mathrm{KL}(\nu, \nu') = \infty$ if $\nu_a \neq \nu'_a$ for some a, since ν (resp ν') govern an infinite i.i.d sequence $\{X_{[s,a]}\} \sim \nu_a$ (resp $\sim \nu'_a$). However, in this paper we will always design our simulator so that the quantity $\mathrm{KL}(\mathsf{Sim}(\nu), \mathsf{Sim}(\nu'))$ is finite, and in fact quite small. The hope is that if the modified transcript $\widehat{\mathsf{Tr}}$ conveys too little information to distinguish between $\mathsf{Sim}(\nu^{(1)})$ and $\mathsf{Sim}(\nu^{(2)})$, then Alg will have to behave similarly on both simulated instances. Hence, we will show that if Alg behaves differently on two instances $\nu^{(1)}$ and $\nu^{(2)}$, yet Sim limits information KL between them, then Alg's behavior must differ quite a bit under $\nu^{(i)}$ versus $\mathsf{Sim}(\nu^{(i)})$, for either $i = 1$ or $i = 2$. Formally, we will show that Alg will have to "break" the simulator, in the following sense:

Definition 9 (Breaking) *Given measure ν, algorithm* Alg*, and simulator* Sim*, we say that $W \in \mathcal{F}_T$ is a truthful event under* $\mathsf{Sim}(\nu)$ *if, for all events $E \in \mathcal{F}_T$,*

$$\mathbb{P}_{\mathsf{Sim}(\nu),\mathsf{Alg}}[E \wedge W] = \mathbb{P}_{\nu,\mathsf{Alg}}[E \wedge W] \tag{10}$$

On the other hand, we will say that Alg *breaks on W^c under* $\mathsf{Sim}(\nu)$*. Recall that $\mathcal{F}_t$ is the σ-algebra generated by ξ_{Alg}, and the actions/samples collected by* Alg *up to time t.*

The key insight is that, whenever $\mathsf{Sim}(\nu)$ doesn't break (i.e. on a truthful event W), a run of Alg on ν can be perfectly simulated by running Alg on $\mathsf{Sim}(\nu)$. But if $\mathsf{Sim}(\nu)$ fudges Tr in a way that drastically limits information about ν, this means that Alg can be simulated using little information about ν, which will contradict information theoretic lower bounds. This suggests the following recipe for proving lower bounds:

1) State a claim you wish to falsify over a class of instances $\nu \in \mathcal{S}$ (e.g., the best arm is not pulled more than τ times, with some probability). **2)** Phrase your claims as candidate truthful events on each instance (e.g. $W_\nu := \{N_{a^*(\nu)}(T) \leq \tau\}$ where $a^*(\nu)$ is the best arm of ν) **3)** Construct a simulator Sim such that W_ν is truthful on $\mathsf{Sim}(\nu)$, but $\mathrm{KL}_{\mathsf{Alg}}(\mathsf{Sim}(\nu), \mathsf{Sim}(\widetilde{\nu}))$ (or TV) is small for alternative pairs $\nu, \widetilde{\nu}$. For example, if the truthful event is $\{N_{a^*(\nu)}(T) \leq \tau\}$, then simulator should only modify samples $X_{[a^*,\tau+1]}, X_{[a^*,\tau+2]}, \dots$. **4)** Apply an information-theoretic lower bound (e.g., Proposition 10 to come) to show that the simulator breaks (e.g. $\mathbb{P}_{\nu,\mathsf{Alg}}[W_\nu^c]$ is large for at least one $\nu \in \mathcal{S}$, or for a ν drawn uniformly from $\mathcal{S}$).

6. Applying the Simulator to Permutations

In what follows, we show how to use the simulator to prove Theorem 3. At a high level, our lower bound follows from considering *pairs* of instances where the best arm is swapped-out for a sub-optimal arm, and ultimately averaging over those pairs. On each such pair, we apply a version of Le Cam's method to the simulator setup (proof in Section A.1):

Proposition 10 (Simulator Le Cam) *Let $\nu^{(1)}$ and $\nu^{(2)}$ be two measures,* Sim *be a simulator, and let W_i be two truthful events under $\mathsf{Sim}(\nu^{(i)})$ for $i = 1, 2$. Then, for any algorithm* Alg

$$\sum_{i=1}^{2} \mathbb{P}_{\nu^{(i)},\mathsf{Alg}}(W_i^c) \geq \sup_{E \in \mathcal{F}_T} |\mathbb{P}_{\nu^{(1)},\mathsf{Alg}}(E) - \mathbb{P}_{\nu^{(2)},\mathsf{Alg}}(E)| - Q\left(\mathrm{KL}_{\mathsf{Alg}}\left(\mathsf{Sim}(\nu^{(1)}), \mathsf{Sim}(\nu^{(2)})\right)\right) \,, \tag{11}$$

where $Q(\beta) = \min\left\{1 - \frac{1}{2}e^{-\beta}, \sqrt{\beta/2}\right\}$. The bound also holds with $Q\left(\mathrm{KL}_{\mathsf{Alg}}\left(\mathsf{Sim}(\nu^{(1)}), \mathsf{Sim}(\nu^{(2)})\right)\right)$ replaced by $\mathrm{TV}_{\mathsf{Alg}}\left(\mathsf{Sim}(\nu^{(1)}), \mathsf{Sim}(\nu^{(2)})\right)$.

Note that Equation 11 decouples the behavior of the algorithm under ν from the information limited by the simulator. This proposition makes formal the intuition from Section 5 that the algorithm which behaves differently on two distinct instances must "break" any simulator that severely limits the information between them.

6.1. Lower Bounds on 1-Arm Swaps

The key step in proving Theorem 3 is to establish a simple lower bound that holds for pairs of instances obtained by "swapping" the best arm.

Proposition 11 *Let ν be an instance with unique best arm a^*. For $b \in [n] - \{a^*\}$, let $\nu^{(b,a^*)}$ be the instance obtained by swapping a^* and b, namely $\nu^{(b,a^*)}_{a^*} = \nu_b$, $\nu^{(b,a^*)}_b = \nu_{a^*}$, and $\nu^{(b,a^*)}_a = \nu_a$ for $a \in [n] - \{a^*, b\}$. Then, if* Alg *is δ-correct, one has that for any $\eta \in (0, 1/4)$*

$$\frac{1}{2}\left\{\mathbb{P}_{\nu,\mathsf{Alg}}\left[N_b(T) > \tau(\eta)\right] + \mathbb{P}_{\nu^{(b,a^*)},\mathsf{Alg}}\left[N_{a^*}(T) > \tau(\eta)\right]\right\} \geq \eta - \delta \,, \tag{12}$$

where $\tau(\eta) = \frac{1}{\mathrm{KL}(\nu_{a^*},\nu_b) + \mathrm{KL}(\nu_b,\nu_{a^*})} \log(1/4\eta)$

This bound implies that, if an instance $\overline{\nu}$ is drawn uniformly from $\{\nu, \nu^{(b,a^*)}\}$, then any δ-correct algorithm has to pull the suboptimal arm, namely the distribution ν_b, at least $\tau(\eta)$ times on average (over the draw of $\overline{\nu}$), with probability $\eta - \delta$. Proving this proposition requires choosing an appropriate simulator. To this end, fix a $\tau \in \mathbb{N}$, and let Sim map Tr to $\widehat{\mathsf{Tr}}$ such that,

$$\mathsf{Sim} : \widehat{X}_{[s,a]} \hookleftarrow \begin{cases} X_{[s,a]} & a \neq a^*, b \\ X_{[s,a]} & a \in \{a^*, b\}, s \leq \tau \\ \overset{iid}{\sim} \nu_{a^*} & a \in \{a^*, b\}, s > \tau \end{cases} \tag{13}$$

where for $s > \tau$ and $a \in \{a^*, b\}$, the $\widehat{X}_{[s,a]} \overset{iid}{\sim} \nu_{a^*}$ means that the samples are taken independently of everything else (in particular, independent of $X_{[s,a^*]}$ and $X_{[s,b]}$), using internal randomness ξ_{Sim}. We emphasize Sim depends crucially on ν, a^*, and b.

Note that the only entries of $\widehat{\mathsf{Tr}}$ whose distribution differs under $\mathsf{Sim}(\nu)$ and $\mathsf{Sim}(\nu^{(b,a^*)})$ are just the first τ entries from arms a^* and b, namely $\{\widehat{X}_{s,a}\}_{1 \leq s \leq \tau, a \in \{a^*,b\}}$. Hence, by a data-processing inequality

$$\mathrm{KL}_{\mathsf{Alg}}(\mathsf{Sim}(\nu), \mathsf{Sim}(\nu^{(b,a^*)})) \leq \tau\{\mathrm{KL}(\nu_{a^*}, \nu_b) + \mathrm{KL}(\nu_b, \nu_{a^*})\} \tag{14}$$

Using the notation of Proposition 10, let $\nu^{(1)} = \nu$, $\nu^{(2)} = \nu^{(b,a^*)}$, let $W_1 := \{N_b(T) \leq \tau\}$ and $W_2 := \{N_{a^*}(T) \leq \tau\}$ (i.e, under $\nu^{(i)}$ and W_i, you sample the suboptimal arm no greater than τ times). Now, Proposition 11 now follows immediately from Proposition 10, elementary manipulations, and the following claim:

Claim 1 *For $\nu^{(i)}$ and W_i defined above,* Sim *is truthful on W_i under $\nu^{(i)}$.*

Proof [Proof of Claim 1] The samples $\widehat{X}_{[s,a]}$ and $X_{[s,a]}$ have the sample distribution under $\nu^{(i)}$ and $\mathsf{Sim}(\nu^{(i)})$ for $a \notin \{a^*, b\}$ and $s \le \tau$, by construction. Moreover, the samples $\widehat{X}_{[s,a^*]}$ and $\widehat{X}_{[s,b]}$ for $s > \tau$ are also i.i.d draws from ν_{a^*}, so they have the same distribution as the samples $X_{[s,a^*]}$ and $X_{[s,b]}$ under $\nu^{(1)}$ and $\nu^{(2)}$ respectively. Thus, the only samples whose distributions are changed by the simulator are the samples $\widehat{X}_{[s,b]}$ under $\nu^{(1)}$ and $\widehat{X}_{[s,b]}$ under $\nu^{(2)}$, respectively, which Alg never accesses under under W_1 and W_2, respectively. ■

6.2. Proving Theorem 3 from Proposition 11

Theorem 3 can be proven directly using the machinery established thus far. However, we will introduce a reduction to "symmetric algorithms" which will both expedite the proof of the Theorem 3, and come in handy for additional bounds as well. For a transcript Tr, let $\pi(\mathsf{Tr})$ denote the transcript $\pi(\mathsf{Tr})_{a,s} = \mathsf{Tr}_{\pi(a),s}$, and $\mathbb{P}_{\mathsf{Alg},\mathsf{Tr}}$ denote probability taken w.r.t. the randomness of Alg acting on the fixed (deterministic) transcript Tr. For any subset $S \subset [n]$, we take $\pi(S) := \{\pi(a) : a \in S\}$.

Definition 12 (Symmetric Algorithm) *We say that an algorithm* Alg *is* symmetric *if the distribution of its sampling sequence and output commutes with permutations. That is, for any permutation π, transcript* Tr*, sequence of actions $(A_1, A_2, \dots)$, and output $\widehat{S} \subset [n]$,*

$$\mathbb{P}_{\mathsf{Alg},\mathsf{Tr}}\left[(a_1, a_2, \dots, a_T, \widehat{S}) = (A_1, A_2, \dots, A_T, S)\right]$$
$$= \mathbb{P}_{\mathsf{Alg},\pi(\mathsf{Tr})}\left[(a_1, a_2, \dots, a_T, \widehat{S}) = (\pi(A_1), \pi(A_2), \dots, \pi(A_T), \pi(S))\right] \quad (15)$$

In particular, if Alg is symmetric, then $\mathbb{P}_{\nu,\mathsf{Alg}}[N_b(\widetilde{T}) \ge \tau] = \mathbb{P}_{\pi(\nu),\mathsf{Alg}}[N_{\pi(b)}(\widetilde{T}) \ge \tau]$ for all $b \in [n]$, $\pi \in \mathbf{S}_n$, and $\{\mathcal{F}_t\}$-measurable stopping time $\widetilde{T}$. The following lemma reduces lower bounds on average complexity over permutations to lower bounds on a single instance for a symmetric algorithm (see Section A.2 for proof and discussion):

Lemma 13 (Algorithm Symmetrization) *Let* Alg *be a δ-correct algorithm over $\mathbf{S}_n(\nu)$. Then there exists a* symmetric *algorithm* $\mathsf{Alg}^{\mathbf{S}_n}$*, which is also δ correct over $\mathbf{S}_n(\nu)$, and such that, for any $\{\mathcal{F}_t\}$-measurable stopping time $\widetilde{T}$ (in particular, $\widetilde{T} = T$)*

$$\mathbb{P}_{\nu,\mathsf{Alg}^{\mathbf{S}_n}}[N_b(\widetilde{T}) \ge \tau] = \mathbb{P}_{\pi\sim\mathbf{S}_n}\mathbb{P}_{\pi(\nu),\mathsf{Alg}}[N_{\pi(b)}(\widetilde{T}) \ge \tau] \quad (16)$$

Now, we are ready to prove Theorem 3

Proof [Proof of Theorem 3] We first establish 3 for δ-correct symmetric algorithms, and use Lemma 13 to extend to all δ-correct algorithms. Again, let $\nu^{(b,a^*)}$ be the instance obtained by swapping a^* and b, and let π_b be the permutation yielding $\pi_b(\nu) = \nu^{(b,a^*)}$. Adopt the shorthand $\tau_b(\eta) = \tau_b \cdot \log(1/4\eta)$. Then assuming Alg is symmetric and noting that $\pi_b(a^*) = b$, we have

$$\begin{aligned}
\mathbb{P}_{\pi\sim\mathbf{S}_n}\mathbb{P}_{\pi(\nu),\mathsf{Alg}}[N_{\pi(b)}(T) > \tau_b(\eta)] &\overset{(i)}{=} \mathbb{P}_{\nu,\mathsf{Alg}}[N_b(T) > \tau_b(\eta)] \\
&\overset{(ii)}{=} \frac{1}{2}\left\{\mathbb{P}_{\nu,\mathsf{Alg}}[N_b(T) > \tau_b(\eta)] + \mathbb{P}_{\pi_b(\nu),\mathsf{Alg}}[N_{\pi_b(b)}(T) > \tau_b(\eta)]\right\} \\
&\overset{(iii)}{=} \frac{1}{2}\left\{\mathbb{P}_{\nu,\mathsf{Alg}}[N_b(T) > \tau_b(\eta)] + \mathbb{P}_{\nu^{(b,a^*)},\mathsf{Alg}}[N_{a^*}(T) > \tau_b(\eta)]\right\}
\end{aligned}$$

where (i) and (ii) follow from the definition of symmetric algorithms, (iii) follows from how we defined the permutation π_b. Applying Proposition 11, the above is at most $\eta - \delta$. Next, we show that Equation 3 implies Equation 4. This part of the proof need not invoke that Alg is symmetric. Applying Markov's inequality Equation 3 implies that $\mathbb{E}_{\pi \sim \mathbf{S}_n} \mathbb{E}_{\pi(\nu), \mathsf{Alg}} \geq \log(1/4\eta)(\eta - \delta)\tau_b$. Hence,

$$\begin{aligned} \mathbb{E}_{\pi \sim \mathbf{S}_n} \mathbb{E}_{\pi(\nu), \mathsf{Alg}}[T] &= \mathbb{E}_{\pi \sim \mathbf{S}_n} \mathbb{E}_{\pi(\nu), \mathsf{Alg}}[\sum_{b \in [n]} N_b(T)] = \mathbb{E}_{\pi \sim \mathbf{S}_n} \mathbb{E}_{\pi(\nu), \mathsf{Alg}}[\sum_{b \in [n]} N_{\pi(b)}(T)] \\ &\geq \mathbb{E}_{\pi \sim \mathbf{S}_n} \mathbb{E}_{\pi(\nu), \mathsf{Alg}}[\sum_{b \neq a^*} N_{\pi(b)}(T)] \geq \log(1/4\eta)(\eta - \delta) \sum_{b \neq a^*} \tau_b \end{aligned}$$

■

7. Conclusion

In the pursuit of understanding the fundamental limits of adaptive sampling in the presence of side knowledge about the problem (e.g. the means of the actions are known to lie in a known set), we unearthed fundamental limitations of the existing machinery (i.e., change of measure and Fano's inequality). In response, we developed a new framework for analyzing adaptive sampling problems – the Simulator – and applied it to the particular adaptive sampling problem of multi-armed bandits to obtain state-of-the-art lower bounds. New insights from these lower bounds led directly to formulating a new algorithm for the TOP-K problem that is state-of-the-art in both theory and practice. Armed with the tools and demonstration of their use on a simple problem, we are convinced that this recipe can be used to produce future successes for more structured adaptive sampling problems, the true goal of this work.

References

Ery Arias-Castro, Emmanuel J Candes, and Mark A Davenport. On the fundamental limits of adaptive sensing. *IEEE Transactions on Information Theory*, 59(1):472–481, 2013.

Robert E Bechhofer. A sequential multiple-decision procedure for selecting the best one of several normal populations with a common unknown variance, and its use with various experimental designs. *Biometrics*, 14(3):408–429, 1958.

Malgorzata Bogdan, Ewout van den Berg, Weijie Su, and Emmanuel Candes. Statistical estimation and testing via the sorted l1 norm. *arXiv preprint arXiv:1310.1969*, 2013.

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

Alexandra Carpentier and Andrea Locatelli. Tight (lower) bounds for the fixed budget best arm identification bandit problem. In *Proceedings of the 29th Conference on Learning Theory*, 2016.

Rui M Castro. Adaptive sensing performance lower bounds for sparse signal detection and support estimation. *Bernoulli*, 20(4):2217–2246, 2014.

Rui M Castro and Robert D Nowak. Minimax bounds for active learning. *IEEE Transactions on Information Theory*, 54(5):2339–2353, 2008.

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

Lijie Chen, Anupam Gupta, and Jian Li. Pure exploration of multi-armed bandit under matroid constraints. In *29th Annual Conference on Learning Theory*, pages 647–669, 2016.

Lijie Chen, Jian Li, and Mingda Qiao. Nearly instance optimal sample complexity bounds for top-k arm selection, 2017.

Shouyuan Chen, Tian Lin, Irwin King, Michael R Lyu, and Wei Chen. Combinatorial pure exploration of multi-armed bandits. In *Advances in Neural Information Processing Systems*, pages 379–387, 2014.

Richard Combes, Mohammad Sadegh Talebi Mazraeh Shahi, Alexandre Proutiere, et al. Combinatorial bandits revisited. In *Advances in Neural Information Processing Systems*, pages 2116–2124, 2015.

Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

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*, 7(Jun): 1079–1105, 2006.

Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *29th Annual Conference on Learning Theory*, pages 998–1027, 2016.

Aurélien Garivier, Pierre Ménard, and Gilles Stoltz. Explore first, exploit next: The true shape of regret in bandit problems. *arXiv preprint arXiv:1602.07182*, 2016.

Aditya Gopalan, Shie Mannor, and Yishay Mansour. Thompson sampling for complex online problems. 2014.

Steve Hanneke. Theoretical foundations of active learning. 2009.

Kevin G Jamieson, Matthew Malloy, Robert D Nowak, and Sébastien Bubeck. lil'ucb: An optimal exploration algorithm for multi-armed bandits. In *COLT*, volume 35, pages 423–439, 2014.

Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone. Pac subset selection in stochastic multi-armed bandits. In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, pages 655–662, 2012.

Zohar Shay Karnin, Tomer Koren, and Oren Somekh. Almost optimal exploration in multi-armed bandits. *ICML (3)*, 28:1238–1246, 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*, 2015.

Branislav Kveton, Zheng Wen, Azin Ashkan, Hoda Eydgahi, and Brian Eriksson. Matroid bandits: Fast combinatorial optimization with learning. *arXiv preprint arXiv:1403.5045*, 2014.

Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22, 1985.

T. Lattimore and C. Szepesvari. The End of Optimism? An Asymptotic Analysis of Finite-Armed Linear Bandits. *ArXiv e-prints*, October 2016.

Alexander Luedtke, Emilie Kaufmann, and Antoine Chambaz. Asymptotically optimal algorithms for multiple play bandits with partial feedback. *arXiv preprint arXiv:1606.09388*, 2016.

Stefan Magureanu, Richard Combes, and Alexandre Proutiere. Lipschitz bandits: Regret lower bound and optimal algorithms. In *COLT*, pages 975–999, 2014.

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

Frank Nielsen and Vincent Garcia. Statistical exponential families: A digest with flash cards. *arXiv preprint arXiv:0911.4863*, 2009.

Maxim Raginsky and Alexander Rakhlin. Lower bounds for passive and active learning. In *Advances in Neural Information Processing Systems*, pages 1026–1034, 2011.

Maxim Raginsky and Igal Sason. *Concentration of Measure Inequalities in Information Theory, Communications, and Coding*. Now Publishers Inc., 2014.

Dan Russo and Benjamin Van Roy. Eluder dimension and the sample complexity of optimistic exploration. In *Advances in Neural Information Processing Systems*, pages 2256–2264, 2013.

Daniel Russo. Simple bayesian algorithms for best arm identification. In *29th Annual Conference on Learning Theory*, pages 1417–1418, 2016.

Max Simchowitz, Kevin Jamieson, and Benjamin Recht. Best-of-k-bandits. In *29th Annual Conference on Learning Theory*, pages 1440–1489, 2016.

Marta Soare, Alessandro Lazaric, and Rémi Munos. Best-arm identification in linear bandits. In *Advances in Neural Information Processing Systems*, pages 828–836, 2014.

Mohammad Sadegh Talebi and Alexandre Proutiere. An optimal algorithm for stochastic matroid bandit optimization. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, pages 548–556. International Foundation for Autonomous Agents and Multiagent Systems, 2016.

Alexandre B Tsybakov. Introduction to nonparametric estimation. revised and extended from the 2004 french original. translated by vladimir zaiats, 2009.

Yisong Yue and Carlos Guestrin. Linear submodular bandits and their application to diversified retrieval. In *Advances in Neural Information Processing Systems*, pages 2483–2491, 2011.

Appendix A. Proofs for Section 6

A.1. Proof of Propostion 10

First, by combining Pinkser's Inequality with the data processing inequality (Tsybakov, 2009), we arive at an elementary bound that controls TV between runs of an algorithm on simulated instances:

Lemma 14 (Pinkser's Inequality) *Let $\nu^{(1)}$ and $\nu^{(2)}$ be two measures. Then for any simulator* Sim,

$$\sup_{E \in \mathcal{F}_T} \left| \mathbb{P}_{\mathsf{Sim}(\nu^{(1)}),\mathsf{Alg}}(E) - \mathbb{P}_{\mathsf{Sim}(\nu^{(2)}),\mathsf{Alg}}(E) \right| \leq \mathrm{TV}_{\mathsf{Alg}}\left(\mathsf{Sim}(\nu^{(1)}), \mathsf{Sim}(\nu^{(2)}) \right) \tag{17}$$

$$\leq Q\left(\mathrm{KL}_{\mathsf{Alg}}\left(\mathsf{Sim}(\nu^{(1)}), \mathsf{Sim}(\nu^{(2)}) \right) \right) \tag{18}$$

Where $Q(\beta) = \min\left\{ 1 - \frac{1}{2}e^{-\beta}, \sqrt{\beta/2} \right\}$.

Note here that we only consider events $E \in \mathcal{F}_T$, which only depend on the samples $X_{a_1,1}, \dots, X_{a_T,t}$ collected from the modified $\widehat{\mathsf{Tr}}$. Now we can prove our result.
Proof [Proof of Proposition 10] By the triangle inequality

$$\begin{aligned}
&|\mathbb{P}_{\nu^{(1)},\mathsf{Alg}}(E) - \mathbb{P}_{\nu^{(2)},\mathsf{Alg}}(E)| \\
&\qquad \leq |\mathbb{P}_{\mathsf{Sim}(\nu^{(1)}),\mathsf{Alg}}(E) - \mathbb{P}_{\mathsf{Sim}(\nu^{(2)}),\mathsf{Alg}}(E)| + \sum_{i=1}^{2} |\mathbb{P}_{\mathsf{Sim}(\nu^{(i)})),\mathsf{Alg}}(E) - \mathbb{P}_{\nu^{(i)}),\mathsf{Alg}}(E)|
\end{aligned} \tag{19}$$

We can expand

$$\begin{aligned}
&\mathbb{P}_{\mathsf{Sim}(\nu^{(i)}),\mathsf{Alg}}(E) - \mathbb{P}_{\nu^{(i)},\mathsf{Alg}}(E) \\
=&\mathbb{P}_{\mathsf{Sim}(\nu^{(i)}),\mathsf{Alg}}(E \wedge W_i) + \mathbb{P}_{\mathsf{Sim}(\nu^{(i)})),\mathsf{Alg}}(E \wedge W_i^c) - (\mathbb{P}_{\nu^{(i)},\mathsf{Alg}}(E \wedge W_i) + \mathbb{P}_{\nu^{(i)},\mathsf{Alg}}(E \wedge W_i^c)) \\
=&\mathbb{P}_{\mathsf{Sim}(\nu^{(i)}),\mathsf{Alg}}(E \wedge W_i^c) - \mathbb{P}_{\nu^{(i)},\mathsf{Alg}}(E \wedge W_i^c)
\end{aligned} \tag{20}$$

where $\mathbb{P}_{\mathsf{Sim}(\nu^{(i)}),\mathsf{Alg}}(E \wedge W_i) = \mathbb{P}_{\nu^{(i)},\mathsf{Alg}}(E \wedge W_i)$ as W_i is truthful for $\nu^{(i)}$. Thus,

$$\begin{aligned}
|\mathbb{P}_{\mathsf{Sim}(\nu^{(i)}),\mathsf{Alg}}(E) - \mathbb{P}_{\nu^{(i)},\mathsf{Alg}}(E)| &= |\mathbb{P}_{\mathsf{Sim}(\nu^{(i)}),\mathsf{Alg}}(E \wedge W_i^c) - \mathbb{P}_{\nu^{(i)},\mathsf{Alg}}(E \wedge W_i^c)| \\
&\overset{(i)}{\leq} \max\{\mathbb{P}_{\mathsf{Sim}(\nu^{(i)}),\mathsf{Alg}}(E \wedge W_i^c), \mathbb{P}_{\nu^{(i)},\mathsf{Alg}}(E \wedge W_i^c)\} \\
&\overset{(ii)}{\leq} \max\{\mathbb{P}_{\mathsf{Sim}(\nu^{(i)}),\mathsf{Alg}}(W_i^c), \mathbb{P}_{\nu^{(i)},\mathsf{Alg}}(W_i^c)\} \\
&\overset{(iii)}{=} \mathbb{P}_{\nu^{(i)},\mathsf{Alg}}(W_i^c)
\end{aligned}$$

Where (i) uses the identity $|a - b| \leq \max\{a, b\}$ for $a, b \geq 0$, (ii) uses monotonicity of probability measures, and (iii) uses the fact that $\mathbb{P}_{\mathsf{Sim}(\nu^{(i)}),\mathsf{Alg}}(W_i^c) = 1 - \mathbb{P}_{\mathsf{Sim}(\nu^{(i)}),\mathsf{Alg}}(W_i) = 1 - \mathbb{P}_{\nu^{(i)},\mathsf{Alg}}(W_i) = \mathbb{P}_{\nu^{(i)},\mathsf{Alg}}(W_i^c)$, since W_i is truthful. All in all, we have

$$|\mathbb{P}_{\nu^{(1)},\mathsf{Alg}}(E) - \mathbb{P}_{\nu^{(2)},\mathsf{Alg}}(E)| \leq |\mathbb{P}_{\mathsf{Sim}(\nu^{(1)}),\mathsf{Alg}}(E) - \mathbb{P}_{\mathsf{Sim}(\nu^{(2)}),\mathsf{Alg}}(E)| + \sum_{i=1}^{2} \mathbb{P}_{\nu^{(i)}}(W_i^c) \tag{21}$$

The bound now follows from Lemma 14. ■

A.2. Proof of Lemma 13

Let Alg be a (possbily non-symmetric) algorithm. We obtain the symmetric algorithm $\mathsf{Alg}^{\mathbf{S}_n}$ by drawning a $\sigma \sim \mathbf{S}_n$, and running Alg on $\sigma(\mathsf{Tr})$ with decision rule $\sigma^{-1}(\widehat{S})$. Note then that a sample from arm a on Tr corresponds to a sample from arm $\sigma(a)$ on $\sigma(\mathsf{Tr})$. Hence, for any $\pi \in \mathbf{S}(n)$,

$$\begin{aligned}
&\mathbb{P}_{\mathsf{Alg}^{\mathbf{S}_n},\mathsf{Tr}}\left[(a_1, a_2, \dots, a_T, \widehat{S}) = (A_1, A_2, \dots, A_T, S)\right] \\
&= \frac{1}{n!}\sum_{\sigma \in \mathbf{S}_n} \mathbb{P}_{\mathsf{Alg},\sigma(\mathsf{Tr})}\left[(a_1, a_2, \dots, a_T, \sigma^{-1}(\widehat{S})) = (\sigma(A_1), \sigma(A_2), \dots, \sigma(A_T), S)\right] \\
&= \frac{1}{n!}\sum_{\sigma \in \mathbf{S}_n} \mathbb{P}_{\mathsf{Alg},\sigma(\mathsf{Tr})}\left[(a_1, a_2, \dots, a_T, \widehat{S}) = (\sigma(A_1), \sigma(A_2), \dots, \sigma(A_T), \sigma(S))\right] \\
&= \frac{1}{n!}\sum_{\sigma \in \mathbf{S}_n} \mathbb{P}_{\mathsf{Alg},\sigma\circ\pi(\mathsf{Tr})}\left[(a_1, a_2, \dots, a_T, \widehat{S}) = (\sigma\circ\pi(A_1), \sigma\circ\pi(A_2), \dots, \sigma\circ\pi(A_T), \sigma\circ\pi(S))\right] \\
&= \mathbb{P}_{\mathsf{Alg}^{\mathbf{S}_n},\pi(\mathsf{Tr})}\left[(a_1, a_2, \dots, a_T, \widehat{S}) = (\pi(A_1), \pi(A_2), \dots, \pi(A_T), \pi(S))\right]
\end{aligned} \tag{22}$$

as needed. We remark that this reduction to symmetric algorithms is also adopted in Castro (2014), but there the reduction is applied to classes of instances which themselves are highly symmetric (e.g., all the gaps are the same). Previous works on the sampling patterns lower bounds for MAB explicitly assume that algorithms satisfy weaker conditions Garivier et al. (2016); Carpentier and Locatelli (2016), whereas our reduction to symmetric algorithms still implies bounds which hold for possibly non-symmetric algorithms as well.

Appendix B. Proof of Theorem 4

In Theorem 4, we consider the simplified case $\nu_2 = \nu_3 = \dots = \nu_n$, and fix a symmetrized algorithm Alg, and the best arm has mean ν_1. We will actually prove a slightly more technical version of Theorem 4, from which the theorem follows as an immediate corollary.

Recall that the intuition behind Theorem 4 is to show that, until sufficiently many samples has been taken, one cannot differentiate between the best arm, and other arms which exhibit large statistical deviations. To this end, we construct a simulator which is truthful as long as the top arm is not sampled too often. Fix a $\tau \in \mathbb{N}$ and define the simulator Sim by

$$\mathsf{Sim} : \hat{X}_{[s,a]} \leftarrow \begin{cases} X_{[s,a]} & s \le \tau \\ \overset{i.i.d.}{\sim} \nu_2 & s > \tau \end{cases} \tag{23}$$

Since Sim only depends on the first τ samples from ν, we can use Fano's inequality to get control on events under simulated instances:

Lemma 15 *For any random, $\mathcal{F}_T$-measurable subset $\mathcal{A}$ of $[n]$ with $|\mathcal{A}| = m$,*

$$\mathbb{P}_{\pi\sim\mathbf{S}_n}\mathbb{P}_{\mathsf{Sim}(\pi(\nu)),\mathsf{Alg}}[\pi(1) \notin \mathcal{A}] \quad \ge \quad 1 - \frac{\tau\Delta^2 + \log 2}{\log(n/m)} \tag{24}$$

where $\Delta^2 := \mathrm{KL}(\nu_1, \nu_2) + \mathrm{KL}(\nu_2, \nu_1)$.

If we take $\mathcal{A} = \widehat{S}$ to be the best estimate for the top arm in the above lemma, we conclude that unless $\tau \gg \Delta^{-2}\log n$, running Alg on $\mathsf{Sim}(\nu)$ won't be able to identify the best arm. Hence, Alg will need to break the simulator by collecting more than τ samples. More subtly, we can take $\mathcal{A}$ to be the set of the first m arms pulled more than $\tau = \Delta^{-2}\log(n/m)$ times (where $|\mathcal{A}| < m$ if fewer than m arms are pulled τ times). By Lemma 15, $\mathcal{A}$ won't contain the top arm a good fraction of the time. But we know from the previous argument that the top arm is sampled at least τ times, which implies that with constant probability, there will be m arms pulled at least τ times. In summary, we arrive at the following proposition which restates Theorem 4, as well as proving that the top arm must be pulled $\Omega(\Delta^{-2}\log n)$ times:

Proposition 16 *Let* Alg *be δ-correct, consider a game with best arm ν_1 and $n-1$ arms of measure ν_2. For any $\beta \geq 0$, define* $S_{m,\beta} := \left\{a : N_a(T) > \Delta^{-2}\left(\beta\log\frac{n}{m} - \log 2\right)\right\}$. *Then,*

$$\mathbb{P}_{\pi\sim\mathbf{S}_n}\mathbb{P}_{\pi(\nu),\mathsf{Alg}}\left[N_{\pi(1)}(T) \geq \Delta^{-2}(\beta\log n - \log 2)\right] \geq 1-(\beta+\delta) \quad \text{and} \tag{25}$$

$$\mathbb{P}_{\pi\sim\mathbf{S}_n}\mathbb{P}_{\pi(\nu),\mathsf{Alg}}\left[\{\pi(1)\in S_{m,\beta}\}\wedge\{|S_{m,\beta}|\geq m\}\right] \geq 1-2\beta-\delta \tag{26}$$

Note that Theorem 4 follows from Equation by taking $\beta = 1/16$ and $\delta \leq 1/8$.

Proof [Proof of Proposition 16] Throughout the proof, will use the elementary inequality that for any events A and B, $\mathbb{P}[A] \leq \mathbb{P}[A\cap B] + \mathbb{P}[B^c]$ without comment. Let's start by proving Equation 25. Define $W_\pi = \{N_{\pi(1)}(T) \leq \tau\}$, and let W to be corresponding events when π is taken to be the identity. We see W_π is $\mathcal{F}_T$-measurable, and if $\mathsf{Alg}^{\mathbf{S}_n}$ is the symmetrized algorithm obtained from Alg, then

$$\mathbb{P}_{\pi\sim\mathbf{S}_n}\mathbb{P}_{\pi(\nu),\mathsf{Alg}}[W_\pi] = \mathbb{P}_{\nu,\mathsf{Alg}^{\mathbf{S}_n}}[W] \tag{27}$$

Hence, it suffices to assume that Alg is symmetric and work with π being the identity. Since the first τ samples from arm 1 under $\mathsf{Sim}(\nu)$ are i.i.d from ν_1, and since all samples from all other arms are i.i.d from ν_2, we see that

Claim 2 *W is a truthful event for* $\mathsf{Sim}(\nu)$.

This implies that

$$\begin{aligned}
\mathbb{P}_{\nu,\mathsf{Alg}}[W] &\leq \mathbb{P}_{\nu,\mathsf{Alg}}[W\wedge\{\hat{a}=1\}] + \mathbb{P}_{\nu,\mathsf{Alg}}[\{\hat{a}\neq 1\}] \\
&\leq \mathbb{P}_{\nu,\mathsf{Alg}}[W\wedge\hat{a}=1] + \delta \\
&\overset{(i)}{=} \mathbb{P}_{\mathsf{Sim}(\nu),\mathsf{Alg}}[W\wedge\hat{a}=1] + \delta \\
&\leq \mathbb{P}_{\mathsf{Sim}(\nu),\mathsf{Alg}}[\hat{a}=1] + \delta\ ,
\end{aligned}$$

where (i) follows from the following Claim 2. Hence, Lemma 15 implies

$$\mathbb{P}_{\mathsf{Sim}(\nu),\mathsf{Alg}}[\{\hat{a}=1\}] \leq \frac{\Delta^2\tau + \log 2}{\log n}\ . \tag{28}$$

For the next part, we may also assume without loss of generality that Alg is symmetric. Define the set $A_t = \{i : N_i((t+1)\wedge T) > \tau\}$ (these are the set of arms that have been pulled more than τ times), and let $S_m = T \wedge \sup\{t : |A_t| \leq m\}$ (S_m is the last time that A_t is no larger than m). Note that S_m is indeed a stopping time wrt to $\{\mathcal{F}_t\}$, since the $t+1$-th arm to be sampled is determined by all the samples seen up to time t, and internal randomness in Alg. Hence, we have that

$$\begin{aligned}\mathbb{P}_{\nu,\mathsf{Alg}}\left[\{|\{a : N_a(T) > \tau\}| \le m\}\right] &\le \mathbb{P}_{\nu,\mathsf{Alg}}\left[W^c \cap \{|\{a : N_a(T) > \tau\}| \le m\}\right] + \mathbb{P}_{\nu,\mathsf{Alg}}[W] \\ &\overset{(i)}{\le} \mathbb{P}_{\nu,\mathsf{Alg}}[1 \in A_{S_m}] + \mathbb{P}_{\nu,\mathsf{Alg}}[W] \;,\end{aligned}$$

where (i) follows because, under the event $\{|\{a : N_a(T) > \tau\}| \le m\}$, then $S_m = T$, and thus $A_{S_m} = \{a : N_a(T) > \tau\}$. But on W^c, $N_1(T) > \tau$, and thus $1 \in A_{S_m}$. $\mathbb{P}_{\nu,\mathsf{Alg}}[W]$ is already bounded by part 1; for part 2 we need the following claim to invoke a reduction:

Claim 3 $\mathbb{P}_{\mathsf{Sim}(\nu),\mathsf{Alg}}[1 \in A_{S_m}] = \mathbb{P}_{\nu,\mathsf{Alg}}[1 \in A_{S_m}]$. *Moreover, if* Alg *is symmetrized, then* $\mathbb{P}_{\pi\sim\mathbf{S}_n}\mathbb{P}_{\mathsf{Sim}(\pi(\nu)),\mathsf{Alg}}[\pi(1) \in A_{S_m}] = \mathbb{P}_{\mathsf{Sim}(\nu),\mathsf{Alg}}[1 \in A_{S_m}]$.

The first part of this claim holds because then event $\{1 \in A_{S_m}\}$ depends only on the first τ samples drawn from arm 1, and the first τ samples from arm 1 are i.i.d from ν_1 under both the simulator and the true measure. The second part of the claim follows directly from the definition of symmetry, since the even $1 \in A_{S_m}$ does not depend on how the arms are labeled. Thus, invoking Lemma 15,

$$\mathbb{P}_{\pi\sim\mathbf{S}_n}\mathbb{P}_{\mathsf{Sim}(\pi(\nu)),\mathsf{Alg}}[\pi(1) \in A_{S_m}] \le \frac{\Delta^2\tau + \log 2}{\log n/m} \;. \tag{29}$$

Putting pieces together, we conclude that

$$\begin{aligned}\mathbb{P}_{\nu,\mathsf{Alg}}\left[\{|\{i : N_i(T) > \tau\}| \le m\}\right] &\le \delta + (\Delta^2\tau + \log 2)\left(\frac{1}{\log n} + \frac{1}{\log(n/m)}\right) \\ &\le \delta + 2\frac{\Delta^2\tau + \log 2}{\log(n/m)} \;.\end{aligned}$$

Setting $\tau = \Delta^{-2}(\beta\log(n/m) - \log 2)$ concludes. ■

B.1. Proof of Lemma 15

For $i \in \{1,\dots,n\}$, let $\nu^{(i)}$ denote the instance where $\nu_i = \nu_1$, and $\nu_j = \nu_2$ for $j \ne i$. Let $\mathbb{P}_i$ denote the law of the transcipt $\mathsf{Sim}(\nu^{(i)})$. We beging by applying a slight generalization of Fanos Inequality:

Lemma 17 (Inexact Fano) *Let X be a random variable, and let E be a binary random variable, and suppose that Y is a random variables such that X and E are conditionally independent given Y (i.e. $X \to Y \to E$ form a Markov Chain). Then,*

$$P(E = 1) \ge 1 - \frac{I(X;Y) + \log(2)}{H(X) - H(X|E = 0, \hat{X})} \tag{30}$$

where $I(X;Y)$ denotes the mutual information between X and Y, $H(X)$ denotes the entropy of X, and $H(X|E = 0)$ denotes the conditional entropy of X given $E = 0$ (for details, seeCover and Thomas (2012))

To apply the bound, let $\pi \sim \mathbf{S}_n$, let $X = \pi(1)$, let Y denote the transcript $\widehat{\mathsf{Tr}}$ under the distribution $\mathsf{Sim}(\nu^{(\pi(1))}$, and let $E = \mathbb{I}(\{\pi(1) \in \mathcal{A}\}) = \mathbb{I}(\{X \in \mathcal{A}\})$. Then $X \to Y \to E$ forms a markov chain. Since $|\mathcal{A}| = m$, on the event $E = 0$, X can take at most m values, namely those in $\mathcal{A}$. Hence, using a standard entropy bound Cover and Thomas (2012), $H(X|E) \le \log m$. On the other hand, since X is uniform, $H(X) = \log n$, and thus $H(X) - H(X|E = 0, \hat{X}) \ge \log n/m$.

Thus, to conclude, it suffices to show that $I(X;Y) \le \tau\Delta^2$. Let $\bar{\mathbb{P}}$ denote the marginal of Y, that is, $\mathbb{P}_X$, where $X \overset{unif}{\sim} \{1, \dots, n\}$. Then, a standard application of Jensen's inequality (see Cover and Thomas (2012) for details) gives

$$I(X;Y) \quad := \quad \sum_{i=1}^{n} \mathbb{P}(X = i)\mathrm{KL}(\mathbb{P}_i, \bar{\mathbb{P}}) \le \sum_{j,i=1}^{M} \mathbb{P}(X = j)\mathbb{P}(X = i)\mathrm{KL}(\mathbb{P}_i, \mathbb{P}_j) \tag{31}$$

For $i = j$, $\mathrm{KL}(\mathbb{P}_i, \mathbb{P}_i) = 0$. For $i \neq j$, we use the independence of the entries of the transcript to compute

$$\begin{aligned} \mathrm{KL}(\mathbb{P}_i, \mathbb{P}_j) =& \sum_{a=1}^{n} \sum_{s=1}^{\infty} \mathrm{KL}(\widehat{X}_{[a,s]} \big| \mathsf{Sim}(\nu^{(i)}), \widehat{X}_{[a,s]} \big| \mathsf{Sim}(\nu^{(j)}) \\ \overset{(i)}{=}& \sum_{s=1}^{\tau} \mathrm{KL}(\widehat{X}_{[i,s]} \big| \mathsf{Sim}(\nu^{(i)}), \widehat{X}_{[i,s]} \big| \mathsf{Sim}(\nu^{(j)}) + \mathrm{KL}(\widehat{X}_{[j,s]} \big| \mathsf{Sim}(\nu^{(i)}), \widehat{X}_{[j,s]} \big| \mathsf{Sim}(\nu^{(j)}) \\ \overset{(ii)}{=}& \sum_{s=1}^{\tau} \mathrm{KL}(\nu_1, \nu_2) + \mathrm{KL}(\nu_2, \nu_1) = \tau\Delta^2 \ , \end{aligned}$$

where (i) follows since the law of $\widehat{X}_{[a,s]}$ differs between $\mathsf{Sim}(\nu^{(i)})$ and $\mathsf{Sim}(\nu^{(j)})$ for $a \in \{i, j\}$ and $s \in \{1, \dots, \tau\}$, and (ii) follows from the construction of our simulator. Hence,

$$I(X;Y) \quad \le \quad \sum_{j,i=1}^{M} \mathbb{P}(X = j)\mathbb{P}(X = i)\mathrm{KL}(\mathbb{P}_i, \mathbb{P}_j) = \sum_{j,i=1}^{M} \frac{\tau\Delta^2}{n^2}\mathbb{I}(i \neq j) \le \tau\Delta^2 \tag{32}$$

Appendix C. Lower Bounds for Distinct Measures

C.1. High Level-Intuition For Proposition 5

As in the other results in this paper, the key step boils down to designing an effective simulator Sim. Unlike the prior bounds, we need to take a lot of care to quantify how Sim limits information between instances.

To make things concrete, suppose that the base instance is ν with best arm index 1, and where the measures ν_i are Gaussians with means μ_i and variance 1. For clarity, suppose that the gaps are on the same order, say $\Delta \le \mu_1 - \mu_b \le 2\Delta$ for all $b \ge 2$. Since our goal is to show that the best arm must be pulled $\gtrsim \Delta^{-2} \log n$ times on average, a natural choice of a truthful event is $W = \{N_1(T) \le \tau\}$ for some $\tau \gtrsim \Delta^{-2} \log n$. This suggests that our simulator should always return the true samples $X_{[a,s]}$ from Tr for all arms $a \neq 1$, and the first τ samples from arm 1.

Once τ samples are taken from arm 1, our Sim will look at the first τ samples from each arm $j \neq 1$, and pick an index $\widehat{j}$ such that the first τ samples $X_{[\widehat{j},1]}, \dots, X_{[\widehat{j},\tau]}$ "look like" they were drawn from the distribution ν_1. We do this by defining events E_j which depend on the first τ-samples from arm j, as well as some internal random bits ξ_j, and choosing $\widehat{j}$ uniformly from the arms j for which E_j holds. In other word, our simulator is given by

$$\mathsf{Sim}(\nu) : \widehat{X}_{[a,s]} \hookleftarrow \begin{cases} X_{[a,s]} & a \neq 1 \\ X_{[1,s]} & a = 1, s \le \tau \\ \overset{i.i.d}{\sim} \nu_{\widehat{j}} & a = 1, s > \tau \end{cases} \quad \text{where} \tag{33}$$

$$\widehat{j} = \begin{cases} \overset{unif}{\sim} \{j \neq 1 : E_j \text{ holds}\} & \text{if at least one } E_j \text{ holds} \\ 1 & \text{otherwise} \end{cases} \tag{34}$$

Our construction will ensure that at least one E_j will hold with constant probability. Hence, the only information which can distinguish between the arms 1 and $\widehat{j} \neq 1$ are the first τ samples from each arm. But if the first τ samples from arm $\widehat{j}$ "look" as if they were drawn from ν_1, then this information will be insufficient to tell the arms apart. In other words, we can think of Sim as forcing the learner to conduct an adversarially-chosen, *data-dependent* two-hypothesis test: is the best arm 1 or arm $\widehat{j}$?

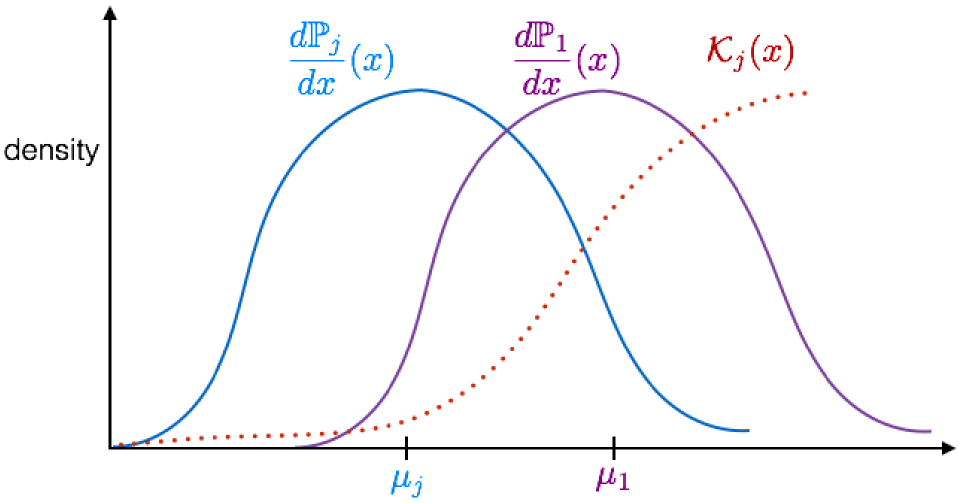
What's left is to understand why we should even expect to find an arm $\widehat{j}$ whose first $\tau = O(\Delta^{-2} \log n)$ samples resemble those from arm 1. The intuition for this is perhaps best understood in terms of Gaussian large-deviations. Indeed, consider the empirical means of each arm $\widehat{\mu}_{j,\tau} = \frac{1}{\tau} \sum_{s=1}^{\tau} X_{[j,s]}$. Then for any *fixed* $j \in [n]$, we have that $|\mu_j - \widehat{\mu}_j| \lesssim \sqrt{1/\tau}$. However, Gaussian large deviations imply that for *some* arm $\widehat{j} \in \{2, \dots, n\}$, the empirical mean will overestimate its true mean by a factor of $\approx \sqrt{\log(n)/\tau}$ (that is $\widehat{\mu}_{\widehat{j},\tau} \ge \mu_{\widehat{j}} + \Omega(\sqrt{\log(n/\delta)/\tau})$). By the assumption that $\Delta \le \mu_1 - \mu_{\widehat{j}} \le 2\Delta$, the large deviation combined with a confidence interval around arm 1 implies that unless $\tau \gtrsim \Delta^{-2} \log n$, there will be an arm $\widehat{j}$ whose empirical mean is larger the empirical mean of arm 1; thereby "looking" like the best arm.

Unfortunately, this intuition is not quite enough for a proof. Indeed, if $\tau \ll \Delta^{-2} \log n$, then with good probability the the arm with the greatest empirical mean will not be best arm. This leads to a paradox: suppose $\tau \ll \Delta^{-2} \log n$, and the learner is given a choice between two arms - one of which has the highest empirical mean, and one of which is assured to be the best arm. Then the learner should guess that the best arm is the one with the *lesser* of the two empirical means!

C.2. Tiltings

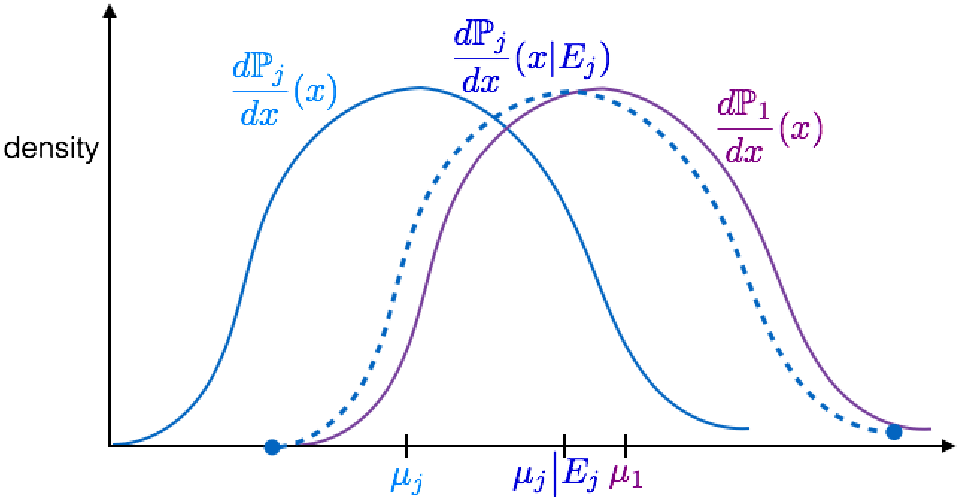
To get around this issue, we pick $\widehat{j}$ using a technique called "tilting", which is the key technical innovation behind this result. Given τ samples from arm j, and access to some random bits ξ_j, the goal is to construct an event E_j (depending on the τ samples from arm j, as well as ξ_j) such that conditioning on E_j "tilts" the distribution of the first τ samples from an arm j to "look like" samples from arm 1. Since the sample mean is a sufficient statistic for Gaussians, it is sufficient to ensure that the distribution of the sample means $\widehat{\mu}_{j,\tau}$ are close in distribution. The basic idea is captured in the following proposition:

Proposition 18 (Informal) *Let $\xi_j \sim \mathrm{Uniform}[0,1]$ and independent of everything else, and let $p \in (0,1)$. If $\tau \ll \Delta^{-2} \log(1/p)$, then there exists a deterministic function $\mathcal{K}_j : \mathbb{R} \to [0,1]$ such that the following holds: Define the event $E_j = \{\mathcal{K}_j(\widehat{\mu}_{j,\tau}) \le \xi_j\}$. Then, the conditional distribution*



(*a*)

Figure 2: Before Tilting



(*a*)

Figure 3: After Tilting

Figure 4: The event E_j depends on the samples from arm j. Thus, conditioning on E_j "tilts" the distribution of the those samples.

of $\widehat{\mu}_{j,\tau}$ on E_j "looks like" the distribution of $\widehat{\mu}_{1,\tau}$, in the sense that the $\mathrm{TV}(\widehat{\mu}_{1,\tau};\widehat{\mu}_{j,\tau}|E_j) = o(1)$. *Moreover, E_j holds with probability at least p.*

Since ξ_j is uniform, $\mathcal{K}_j(\widehat{\mu}_{j,\tau}) = \mathbb{P}_{\xi_j}(E_j|\widehat{\mu}_{j,\tau})$. Thus, up to normalization, conditioning on the event E_j reweights the density of $\widehat{\mu}_{j,\tau}$ by the value of $\mathcal{K}_j(\widehat{\mu}_{j,\tau})$, thereby tilting its shape to resemble the distribution of $\widehat{\mu}_{1,\tau}$. This is depicted in Figure 4. The random numbers ξ_j are essential to this construction, since they let us reweight the distribution of $\widehat{\mu}_{j,\tau}$ by fractional values. Since $\mathcal{K}_j$ is bounded above by one, reweighting doesn't come for free, and our major technical challenge is to choose $\mathcal{K}_j$ so as to ensure that $\mathbb{P}(E_j) = \mathbb{E}[\mathcal{K}_j(\widehat{\mu}_{j,\tau})]$ is at least p. This sort of construction is known in the probability literature as "tilting", and is used in the Herbst argument in the concentration-of-measure literature (Chapter 3 of Raginsky and Sason (2014))[10]. To the best our knowledge, this constitutes the first use of tiltings for proving information theoretic lower bounds.

To conclude our simulator argument, we apply Proposition 18 with $p = (10/n)$ and $\tau \approx \Delta^{-2}\log\frac{n-1}{10} \approx \Delta^{-2}\log n$. Then for any fixed arm j, E_j will hold with probability at least say $10/(n-1)$ (say $n \gg 10$), on which the first τ samples from arm j will "look-like" samples from ν_1, in TV distance. Hence, with probability $1-(1-10/(n-1))^{n-1} \ge 1-e^{-10} \ge .999$, there will exists an arm $\widehat{j}$ such that $E_{\widehat{j}}$ holds, and thus the first τ samples from $\widehat{j}$ "look-like" samples from ν_1, in TV. In particular, if our simulator chooses $\widehat{j}$ uniformly from the arms j such that E_j holds (and takes $\widehat{j} = 1$ otherwise), then with probability $.999$, our simulator can confuse the learner by showing her two arms the distribution of whose samples look like ν_1, as needed.

C.2.1. DATA-DEPENDENT TWO-HYPOTHESIS TESTING

Recall above that Sim forces the learner to perform a data-dependent two hypothesis test - "is the best arm 1 or $\widehat{j}$ " - chosen adversarially from the set of two-hypothesis tests "is the best arm 1 or j" for $j \in \{2,\dots,n\}$. We emphasize that the argument from Proposition 18 is very different than the familiar reductions to n-way or composite hypothesis testing problems. Observe that

1. By giving the learner the choice between only arms 1 and $\widehat{j}$, the adversarial two-hypothesis test reduces the learner's number of possible hypotheses for the best arm from n down to 2. Thus, this problem is potentially easier than the n-way hypothesis test corresponding to best-arm identification. In particular, Proposition 18 is *not implied by Fano's Equality or other n-way testing lower bounds*

2. By the same token, the adversarial two-hypothesis test is also potentially easier than the *composite* hypothesis test: is 1 the best arm, or is another arm $j \in \{2,\dots,n\}$ the best arm? Hence, Proposition 18 is *not implied by lower bounds on composite hypothesis tests.*

3. On the other hand, since $\widehat{j}$ depends on the observed data in this adversarial way, the adversarial two-way hypothesis is strictly *harder* than the standard oblivious two-hypothesis test which *fixes j in advance* and asks: is the best arm 1 and some j? Indeed, fixing the two-hypothesis test in advance does not force the learner to incur a log-factor in the sample complexity.

C.3. Statement of the Main Technical Theorem

Our main theorem is stated for single parameter exponential families(Nielsen and Garcia, 2009), which we define for the sake of completeness in Section C.4.

Theorem 19 *Let ν be a measure with best arm such that each ν_j comes from an exponential family $\{p_\theta\}_{\theta\in\Theta}$ with corresponding parameter by $\theta_j \in \Theta$, and that $[\theta_j, 2\theta_1-\theta_j] \subset \Theta$. Suppose that* Alg *is δ-correct, in the sense that for any $\pi \in \mathbf{S}_n$,* Alg *can identify the unique arm of $\pi(\nu)$ with density $\nu_1 = p_{\theta_1}$ among with probability of error at most δ. Then, for all $\alpha > 0$*

$$\begin{aligned}&\mathbb{E}_{\pi\sim\mathbf{S}_n}\mathbb{P}_{\pi(\nu),\mathsf{Alg}}[\{N_{\pi(1)}(T) > \Delta_{\text{eff}}^{-2}\log(n/\alpha)\}] \ge \sup_{\kappa\in[0,1]}\frac{1}{2}\left(1-e^{-\alpha\kappa(1-\kappa)}\right)(1-2\kappa)-\delta \\ &\text{where}\quad \Delta_{\text{eff}}^2 = \max_{j>1}\mathrm{kl}(\theta_1,\theta_j)+\mathrm{kl}(2\theta_1-\theta_j,\theta_j)\end{aligned} \tag{35}$$

[10] Unlike our construction, the Herbst argument tilting reweights by an unbounded function $\mathcal{K}_j$ (rather than a function bounded in $[0,1]$), and thus those tiltings *cannot* be interpreted as a conditioning on an event

Furthermore, observe that for Gaussian rewards with unit variance, Δ_{eff}^2 corresponds exactly with the largest squared gap $(\theta_1 - \theta_j)^2$. By considering best-arm subproblems with the top $m \le n$ arms, we arrive at the following corollary, which immediate specializes to Proposition 5 with Gaussian rewards:

Corollary 20 *In setting of Theorem 19, we have the following lower bound for every $m \le n$ and $\alpha > 0$,*

$$\mathbb{E}_{\pi \sim \mathbf{S}_n} \mathbb{P}_{\pi(\nu),\mathsf{Alg}}[\{N_{\pi(1)}(T) > \Delta_{\mathrm{eff}}(m)^{-2} \log(m/\alpha)\}] \ge \sup_{\kappa \in [0,1]} \frac{1}{2}\left(1 - e^{-\alpha\kappa(1-\kappa)}\right)(1 - 2\kappa) - \delta$$
$$\text{where} \quad \Delta_{\mathrm{eff}}(m)^2 \textit{ is the } (m-1)\textit{-th smallest value of } \{\mathrm{kl}(\theta_1, \theta_j) + \mathrm{kl}(2\theta_1 - \theta_j, \theta_j)\}_j \tag{36}$$

C.4. Censored Tilting

In this section, we are going to formally construct the events E_j. We will first illustrate the idea for a a generic collection of random variables, and then show how to specialize for bandits. For each $j \in [n]$, we consider a Markov Chain $Z_j \to E_j = 1$, where Z_j is a real valued random variable, and E_j is an event depending only on Z_j. Under suitable technical conditions, the distribution $(Z_j, \mathbb{I}(E_j))$ is then defined by a Markov Kernel $\mathcal{K}_j : \mathbb{R} \to [0,1]$, where $\mathbb{P}(E_j | Z_j = z) = \mathcal{K}_j(z)$. Conversely, any such Markov Kernel induces a joint distribution on $(Z_j, \mathbb{I}(E_j))$. To replicate the malicious adversary from Proposition 18, we can represent E_j explictly by letting $\xi_j \sim \mathrm{Uniform}[0,1]$ and independent of everything else, and setting $E_j = \{\xi_j \le \mathcal{K}_j(z)\}$.

We will say that $\mathcal{K}_j$ is *nondegenerate* if $\mathbb{P}(E_j = 1) \equiv \mathbb{E}[\mathcal{K}_j(Z_j)] > 0$. When Z_j has a density to a measure $\eta(x)$, and $\mathcal{K}_j$ is nongenerate, then Baye's rule implies

$$\frac{d\mathbb{P}_{Z_j}}{d\eta}(x | E_j) = \frac{\mathcal{K}_j(x)}{\mathbb{E}[\mathcal{K}_j(Z_j)]} \cdot \frac{d\mathbb{P}_{Z_j}}{d\eta}(x) \tag{37}$$

In other words, conditioning on the event E_j "tilts" the density of Z_j by a function $\mathcal{K}_j(x)/\mathbb{E}[\mathcal{K}_j(Z_j)]$. We will call tiltings that arise in this fashion a *censored tilting*. Indeed, imagine an observer who tries to measure Z_j. On E_j, she gets a proper measurement of Z_j, but on E_j^c she is censored. Then the censored tilting $\mathbb{P}(Z_j | E_j)$ describes the distribution of the observers non-censored measurements. Keeping with this metaphor, we will call E_j the *measuring event* induced by $\mathcal{K}_j$.

Remark 21 *Tiltings appear as a step in the Herbst Argument for proving concentration of measure bounds from Log-Sobolev inequalitys. In that setting, one tilts by potentially unbounded functions $g_j \ge 0$ that need only satisfy the integrability condition $\mathbb{E}[g_j(Z_j)] < \infty$. In our setting, this tilting to arise from a function $\mathcal{K}_j \in [0,1]$, since $\mathcal{K}_j$ corresponds to a conditional probability operator.*

To apply this idea to MAB, fix a measure ν with decreasing means $\mu_1 > \mu_2 \ge \dots \mu_n$. Given a transcript Tr and $\tau \in \mathbb{N}$, let $\overline{X}_{j,\tau} = \frac{1}{\tau}\sum_{s=1}^{\tau} X_{[j,s]}$. We will simply write $\overline{X}_j$ when τ is clear from context. To simplify things, we shall assume that all the measures ν_j come from a cannonical exponential family of densities $p_\theta(x) = \exp(\theta x - A(\theta))d\eta(x)$ with respect to a measure $\eta(x)$ where θ lie in a convex subset Θ of $\mathbb{R}$. It is well known that this implies that

Lemma 22 (Nielsen and Garcia (2009)) *Suppose that ν_j has density $p_{\theta_j}(x) = \exp(\theta_j x - A(\theta_j))$ with respect to a measure η. Then,*

1. *$\overline{X}_{j,\tau}$ is a sufficient statistic for $X_{[j,1]}, \dots, X_{[j,\tau]}$*

2. *There exists a measure $\eta_\tau(x)$ on $\mathbb{R}$, such that $\overline{X}_{j,\tau}$ has density $q_{\tau\theta_j}(x) := \exp(\tau\theta_j x - \tau A(\theta_j))d\eta_\tau(x)$ with respect $\eta_\tau(x)$.*

In particular, the densities $q_{\tau\theta}(x)$ for $\theta \in \Theta$ form an exponential family.

Now, for each j, consider tiltings of the form $\mathcal{K}_j(x) = \frac{e^{\tau(\theta_1-\theta_j)x}}{c_j}\mathbb{I}(e^{\tau(\theta_1-\theta_j)x} \le c_j)$. Then,

$$\frac{d\mathbb{P}_{Z_j}}{d\eta}(x|E_j) \propto e^{\tau\theta_j x} \cdot e^{\tau(\theta_1-\theta_j)x)} \cdot \mathbb{I}(e^{\tau(\theta_1-\theta_j)x} \le c_j) = e^{\tau\theta_1 x}\mathbb{I}(e^{\tau(\theta_1-\theta_j)x} \le c_j) \tag{38}$$

Since $\frac{d\mathbb{P}_{Z_j}}{d\eta}(x|E_j)$ is a density, the uniquess of normalization implies the following facts:

Lemma 23 *Let $\mathcal{K}_j(x) = \frac{e^{\tau(\theta_1-\theta_j)x}}{c_j}\mathbb{I}(e^{\tau\theta_j x} \le c_j)$, and let E_j be the corresponding measuring event. Then,*

1. *The censored tilting of $\overline{X}_j|E_j$ has the distribution of $\overline{X}_1\big|\{e^{\tau(\theta_1-\theta_j)\overline{X}_j} \le c\}$*

2. $\mathrm{TV}(\mathbb{P}_{\overline{X}_1}, \mathbb{P}_{\overline{X}_j}[\cdot\big|E_j]) = \mathbb{P}(e^{\tau(\theta_1-\theta_j)\overline{X}_1} > c)$

3. $\mathbb{P}(E_j) = \frac{1}{c}(1 - Q_j(E_j)) \cdot e^{\tau\{A(\theta_1)-A(\theta_j)\}}$

Proof The first point follows from Equation 38. The second point follows directly from Lemma 24, and the last point follows from the following computation:

$$\begin{aligned}
\mathbb{P}(E_j) &= \frac{1}{c} \cdot \mathbb{E}\left[\exp(\tau(\theta_1-\theta_j)\overline{X}_j) \cdot \mathbb{I}(e^{\tau(\theta_1-\theta_j)\overline{X}_j} \le c)\right] \\
&= \frac{1}{c} \cdot \int \exp(\tau(\theta_1-\theta_j)x)\exp(\tau\theta_j x - \tau A(\theta_j))\mathbb{I}(e^{\tau(\theta_1-\theta_j)x} \le c)d\eta_\tau(x) \\
&= \frac{1}{c} \cdot \int \exp(\tau\theta_1 x - \tau A(\theta_j)))\mathbb{I}(e^{\tau(\theta_1-\theta_j)x} \le c)d\eta_\tau(x) \\
&= \frac{1}{c} \cdot \int \exp(\tau\theta_1 x - \tau A(\theta_1) + \tau\{A(\theta_1) - A(\theta_j)\}))\mathbb{I}(e^{\tau(\theta_1-\theta_j)x} \le c)d\eta_\tau(x) \\
&= \frac{e^{\tau(A(\theta_1)-A(\theta_j)}}{c} \cdot \int \exp(\tau\theta_1 x - \tau A(\theta_1))\mathbb{I}(e^{\tau(\theta_1-\theta_j)x} \le c)d\eta_\tau(x) \\
&= \frac{e^{\tau\{A(\theta_1)-A(\theta_j)\}}}{c}\mathbb{P}[e^{\tau(\theta_1-\theta_j)\overline{X}_1} \le c]
\end{aligned}$$

■

The last point follows from the following lemma, proved in Section C.8.3:

Lemma 24 (TV **under conditioning**) *Let $\mathbb{P}$ be a probability measure on a space $(\Omega, \mathcal{F})$, and let $B \in \mathcal{F}$ have $\mathbb{P}(B) > 0$. Then,*

$$\mathrm{TV}(\mathbb{P}[\cdot], \mathbb{P}[\cdot|B]) = \mathbb{P}[B^c] \tag{39}$$

.

C.5. Building the Simulator

C.5.1. Defining the Simulator on ν

Again, let ν be an instance with best arm ν_1, and fix a $\tau \in \mathbb{N}$. Our simulator will always return the true samples $X_{[a,s]}$ from Tr for all arms $a \neq 1$, and for the first τ samples from arm 1. After τ samples are taken from arm 1, the samples will be drawn independently from the measure $\nu_{\widehat{j}}$, where $\widehat{j} \in [n]$ is a maliciously chosen index which we will define shortly, using the events E_j in the previous section. To summarize,

$$\mathsf{Sim}(\nu) : \widehat{X}_{[a,s]} \mapsfrom \begin{cases} X_{[a,s]} & a \neq 1 \\ X_{[1,s]} & a = 1, s \leq \tau \\ \overset{i.i.d}{\sim} \nu_{\widehat{j}} & a = 1, s > \tau \end{cases} \tag{40}$$

Fact 1 *If $W = \{N_1(T) \leq \tau\}$, then* Alg *is truthful on W under* $\mathsf{Sim}(\nu)$.

Proof The only samples which are altered by $\mathsf{Sim}(\nu)$ are those taken from arm 1 after arm 1 has been sampled $> \tau$ times. ■

Next, let's define $\widehat{j}$. Let $\overline{X}_j = \frac{1}{\tau}\sum_{s=1}^{\tau} X_{[j,s]}$, fix constants $c_2, \dots, c_n \in \mathbb{R}_{>0}$ to be chosen later, and let $\mathcal{K}_j$ be the corresponding Markov Kernel from Lemma 23. For each $j \in \{2, \dots, n\}$, Sim draws a i.i.d random number $\xi_j \overset{unif}{\sim} [0,1]$. The following fact is just a restatement of this definition in the language of Section C.4:

Fact 2 *Let $E_j = \{\xi_j \leq \mathcal{K}_j(\overline{X}_j)\}$. Then E_j is the measuring event corresponding to the Markov Kernel $\mathcal{K}_j$, and are mutually independent.*

We now define the index $\widehat{j}$ and corresponding "malicious events" M_j by

$$M_j := \{\widehat{j} = j\} \quad \text{where} \quad \widehat{j} = \begin{cases} \overset{unif}{\sim} \{j : E_j \text{ occurs}\} & \text{on } \bigcup_j E_j \\ 1 & \text{otherwise} \end{cases} \tag{41}$$

C.5.2. Defining Sim on Alternate Measures

The next step is to construct our alternative hypotheses. Let $\pi_{(\ell)}$ denote the permutation which swaps arms 1 and ℓ, and define the measures $\nu^{(2)}, \dots, \nu^{(n)}$, where $\nu^{(\ell)} = \pi_{(\ell)}(\nu)$ (note that $\pi_{(\ell)} = \pi_{(\ell)}^{-1}$). To define the simulator on these instances, we still let $\xi_j \overset{unif}{\sim} [0,1]$, and now define, for $j \in \{2, \dots, n\}$

$$E_j^{(\ell)} := \{\xi_j \leq \mathcal{K}_j(\overline{X}_{\pi_{(\ell)}(j)})\} \tag{42}$$

$$\widehat{j}_\ell \overset{unif}{\sim} \{j : E_j^{(\ell)} \text{ holds}\} \tag{43}$$

$$M_j^{(\ell)} := \{\widehat{j}_\ell = j\} \tag{44}$$

and set

$$\mathsf{Sim}(\nu^{(\ell)}) : X_{[a,s]} \mapsto \begin{cases} X_{[a,s]} & a \neq \ell \\ X_{[a,s]} & a = \ell, s \leq \tau \\ \overset{i.i.d}{\sim} \nu_{\widehat{j}_\ell} & a = \ell, s > \tau \end{cases} \tag{45}$$

Note that this esnsure that if $\widehat{\mathsf{Tr}}$ is a transcript from $\mathsf{Sim}(\nu^{(\ell)})$, and $\widehat{\mathsf{Tr}}^{(\ell)}$, then $\pi_{(\ell)}^{-1}(\widehat{\mathsf{Tr}}^{(\ell)})$ (that is, the transcript obtained by swapping indices 1 and ℓ in $\widehat{\mathsf{Tr}}^{(\ell)}$) has the same distribution as $\widehat{\mathsf{Tr}}$.

Our construction is symmetric in the following sense:

Fact 3 Alg *is truthful on* $W_j := \{N_{j)}(T) \le \tau\}$ *under* $\mathsf{Sim}(\nu^{(j)})$. *Moreover, for each* $j \in \{2, \dots, n\}$, $\mathbb{P}_{\mathsf{Sim}(\nu)}[W|M_j] = \mathbb{P}_{\mathsf{Sim}(\nu^{(j)})}[W_j|M_j^{(j)}]$ *and* $\mathbb{P}_{\mathsf{Sim}(\nu)}[\{\hat{y} \ne 1\}|M_j] = \mathbb{P}_{\mathsf{Sim}(\nu^{(j)})}[\{\hat{y} \ne j\}|M_j^{(j)}]$

Proof The first point just follows since $\mathsf{Sim}(\nu^{(j)})$ only changes samples once arm j has been pulled more than τ times. The second point follows since, the event $M_j^{(j)}$ (resp $\{\hat{y} \ne j\}$) and W_j correspond to the events M_j (resp $\{\hat{y} \ne 1\}$) and W if the labels of arms 1 and j are swapped. But if we swap the labels of 1 and j, distribution of $\widehat{\mathsf{Tr}}^{(j)}$ under $\mathsf{Sim}(\nu^{(j)})$ is identical to the distribution of $\widehat{\mathsf{Tr}}$ under $\mathsf{Sim}(\nu)$. ■

Using this symmetry, the total variation between the transcripts returned by $\mathsf{Sim}(\nu)$ given E_j and $\mathsf{Sim}(\nu^{(j)})$ given E_j can be bounded as follows

Fact 4 *Let* $\overline{X}_\ell$ *denote a sample with the distribution of* $\sum_{s=1}^{\tau} X_{\ell,s}$, *where each* $X_{\ell,s} \sim \nu_\ell$. *For* $j \in \{2, \dots, n\}$, $\mathrm{TV}\left[\mathsf{Sim}(\nu)\middle|M_j; \mathsf{Sim}(\nu^{(j)})\middle|M_j^{(j)}\right] \le 2\mathrm{TV}(\overline{X}_j|E_j, \overline{X}_1)$.

This fact takes a bit of care to verify, and so we defer its proof to Section C.8.1.

C.6. Coupling together ν and $\{\nu^{(j)}\}$

Facts 1 and 3, we can couple together the measures using a conditional analogue of the the Simulator Le Cam (Proposition 10), proved in Section C.8.2.

Lemma 25 (Conditional Le Cam's) *Suppose that any events* W, $\{W_j\}$ *and* M_j *satisfy the conclusions of Facts 1 and 3. Then, if* Alg *is symmetric, then for all* $j \in \{2, \dots, n\}$

$$2\mathbb{P}_{\nu,\mathsf{Alg}}\left[W^c|M_j\right] \ge 1 - 2\mathbb{P}_{\nu,\mathsf{Alg}}\left[\{\hat{y} \ne 1\}|M_j\right] - \mathrm{TV}\left[\mathsf{Sim}(\nu|M_j) - \mathsf{Sim}(\nu^{(j)}|M_j)\right] \tag{46}$$

Effectively, the above lemma paritions the space into malicious events M_j, and applies Proposition 10 on each part of the partition.

Since the events M_j are disjoint, multipling the left and right hand side of Equation 46 by $\mathbb{P}_{\nu,\mathsf{Alg}}[M_j]$, setting $\overline{M} := \bigcup_{j=2}^{M}$ and summing yields

$$\begin{aligned} 2\mathbb{P}_{\nu,\mathsf{Alg}}\left[W^c \wedge \overline{M}\right] \quad &\ge \quad \mathbb{P}_{\nu,\mathsf{Alg}}\left[\overline{M}\right] - 2\mathbb{P}_{\nu,\mathsf{Alg}}\left[\{\hat{y} \ne 1\} \wedge \overline{M}\right] \\ &- \quad \sum_j \mathbb{P}_{\nu,\mathsf{Alg}}\left[W^c \wedge M_j\right] \mathrm{TV}\left[\mathsf{Sim}(\nu|M_j) - \mathsf{Sim}(\nu^{(j)}|M_j)\right] \end{aligned}$$

We can bound $\mathbb{P}_{\nu,\mathsf{Alg}}\left[W^c \wedge \overline{M}\right] \le \mathbb{P}_{\nu,\mathsf{Alg}}\left[W^c\right]$ and, if Alg is δ-correct, then $\mathbb{P}_{\nu,\mathsf{Alg}}\left[\{\hat{y} \ne 1\} \wedge \overline{M}\right] \le \mathbb{P}_{\nu,\mathsf{Alg}}\left[\{\hat{y} \ne 1\}\right] \le \delta$. Finally, Holder's Inequality, the disjointness of M_j and Fact 4 imply

$$\begin{aligned} & \sum_j \mathbb{P}_{\nu,\mathsf{Alg}}\left[W^c \wedge M_h\right] \mathrm{TV}\left[\mathsf{Sim}(\nu \big| M_j) - \mathsf{Sim}(\nu^{(j)} \big| M_j)\right] \\ \le\ & \left(\sum_j \mathbb{P}_{\nu,\mathsf{Alg}}[M_j]\right) \cdot \max_j \mathrm{TV}\left[\mathsf{Sim}(\nu \big| M_j) - \mathsf{Sim}(\nu^{(j)} \big| M_j)\right] \\ =\ & \mathbb{P}_{\nu,\mathsf{Alg}}[\overline{M}] \cdot \max_j \mathrm{TV}\left[\mathsf{Sim}(\nu \big| M_j) - \mathsf{Sim}(\nu^{(j)} \big| M_j)\right] \\ \le\ & \mathbb{P}_{\nu,\mathsf{Alg}}[\overline{M}] \cdot 2 \max_j \mathrm{TV}(\overline{X}_j \big| E_j, \overline{X}_1) \end{aligned}$$

where the last step is a consequence of Fact 4. Combining these bounds, and noting that $\overline{M} = \bigcup_{j=2}^n M_j \equiv \bigcup_{j=2}^n E_j$ and $W = \{N_1(T) > \tau\}$ implies the following proposition:

Proposition 26 *Suppose that* Alg *is δ-correct and symmetric. Then*

$$2\mathbb{P}_{\nu,\mathsf{Alg}}[\{N_1(T) > \tau\}] \ge \mathbb{P}_{\nu}[\bigcup_{j=2}^n E_j](1 - 2\max_j \mathrm{TV}(\overline{X}_j \big| E_j, \overline{X}_1)) - 2\delta \tag{47}$$

where we note that the probability of $\bigcup_{j=2}^n E_j$ does not depend on Alg.

Our goal is now clear: choose the Kernel's $\mathcal{K}_j$ so as to balance the terms $Pr_\nu[\bigcup_{j=2}^n E_j]$ and $\max_j Q_j(E_j)$ in Equation 26.

C.7. Proving Theorem 19

To conclude Theorem 19, we first introduce the following technical lemma.

Lemma 27 *Suppose that ν_j comes from an exponential family $\{p_\theta\}_{\theta\in\Theta}$ with corresponding parameter by $\theta_j \in \Theta$. If $[\theta_j, 2\theta_1 - \theta_j] \subset \Theta$, then for any $\kappa > 0$, there exists a choice of c_j for which the corresponding kernel $\mathcal{K}_j$ has*

$$\mathrm{TV}(\overline{X}_j \big| E_j, \overline{X}_1) \le \kappa \quad \textit{and} \quad \mathbb{P}(E_j) \ge \kappa(1-\kappa)e^{-\tau\{\mathrm{kl}(\theta_1,\theta_j) + \mathrm{kl}(2\theta_1 - \theta_j, \theta_j)\}} \tag{48}$$

where $\mathrm{kl}(\theta, \widetilde{\theta})$ denotes the KL *divergence between the laws $\mathbb{P}_\theta$ and $\mathbb{P}_{\widetilde{\theta}}$.*

With this Lemma in hand, we see that taking $\kappa > 0$, and $\tau = \log(n/\alpha)(\max_j \mathrm{kl}(\theta_1, \theta_j) + \mathrm{kl}(2\theta_1 - \theta_j, \theta_j))^{-1}$ implies that

$$\begin{aligned} 2\mathbb{P}_{\nu,\mathsf{Alg}}[\{N_1(T) > \tau\}] &\ge (1 - (1 - \kappa(1-\kappa)\alpha/n)^n (1 - 2\kappa) - 2\delta \\ &\ge (1 - \left(1 - \frac{\alpha\kappa(1-\kappa)}{n}\right)^n)(1 - 2\kappa) - 2\delta \\ &\ge (1 - e^{-\alpha\kappa(1-\kappa)})(1-2\kappa) - 2\delta \end{aligned}$$

Moving from symmetrized algorithms to expecations over $\pi \sim \mathbf{S}_n$ (Lemma 13) concludes the proof of Theorem 19.

Proof [Proof of Lemma 27] By Markov's inequality and an elementary identity for the MGF of a natural exponential family,

$$Q_j(E_j) = \mathbb{P}[e^{\tau(\theta_1-\theta_j)\overline{X}_1} > c] \quad \leq \quad \frac{1}{c}\mathbb{E}[e^{\tau(\theta_1-\theta_j)\overline{X}_1}] = \frac{1}{c}\exp(\tau(A(2\theta_1-\theta_j)-A(\theta_1)) \tag{49}$$

In particular, if we set $c_j = \frac{1}{\kappa}\exp(\tau(A(2\theta_1-\theta_j)-A(\theta_1))$ then the above expression is no more than κ. With this choice of c,

$$\begin{aligned}
\mathbb{P}(E_j) &= \frac{1}{c}e^{\tau\{A(\theta_1)-A(\theta_j)\}}(1-Q_j(E_j)) \\
&= \kappa e^{\tau\{2A(\theta_1)-A(\theta_j)-A(2\theta_1-\theta_j)\}}(1-Q_j(E_j)) \\
&\geq \kappa(1-\kappa)e^{\tau\{2A(\theta_1)-A(\theta_j)-A(2\theta_1-\theta_j)\}}
\end{aligned}$$

We now invoke a well known property of exponential families

Fact 5 (Nielsen and Garcia (2009)) *Let $\{p_\theta\}_{\theta\in\Theta}$ be an exponential family. Then for $\theta, \widetilde{\theta} \in \Theta$, then $\mathrm{kl}(\theta,\widetilde{\theta}) = (\theta-\widetilde{\theta})A'(\theta) - A(\theta) + A(\widetilde{\theta})$, where $A'(\theta) = \int x p_\theta(x)d\nu(x)$ provided the integral exists.*

For ease of notation, set $d_j = \theta_1 - \theta_j$. Then,

$$\begin{aligned}
& 2A(\theta_1) - A(\theta_j) - A(2\theta_1-\theta_j) \\
= \; & A(\theta_1) - A(\theta_1-d_j) + A(\theta_1) - A(\theta_1+d_j) \\
= \; & A(\theta_1) - A(\theta_1-d_j) - A'(\theta_1)d_j + A(\theta_1) - A(\theta_1+d_j) + A'(\theta_1)d_j \\
= \; & -\mathrm{kl}(\theta_1,\theta_1-d_j) - \mathrm{kl}(\theta_1,\theta_1+d_j)
\end{aligned}$$

■

C.8. Deferred Proofs for Theorem 19

C.8.1. Proof of Fact 4

Let $\widehat{\mathsf{Tr}}$ with samples $\widehat{X}_{[a,s]}$ and denote the transcript from $\mathsf{Sim}(\nu)$ and let $\widehat{\mathsf{Tr}}^{(j)}$ with samples $\widehat{X}^{(j)}_{[a,s]}$ denote the transcript from $\mathsf{Sim}(\nu^{(j)})$.

First, note that under M_j and $M^{(j)}_j$, all samples $\widehat{X}_{[a,s]}$ and $\widehat{X}^{(j)}_{[a,s]}$ for $a \in \{1,j\}$ and $s > \tau$ are i.i.d from ν_j. Moreover, by symmetry of the construction under swapping the labels of 1 and j, its easy to see that the samples $\widehat{X}_{[a,s]}$ and $\widehat{X}^{(j)}_{[a,s]}$ for $a \notin \{1,j\}$ have the same distribution under M_j and $M^{(j)}_j$ respectively as well (even though these samples are not necessarily going to be i.i.d from ν_a because of the conditioning). Hence,

$$\begin{aligned}
&\mathrm{TV}\left[\mathsf{Sim}(\nu)\big|M_j; \mathsf{Sim}(\nu^{(j)})\big|M^{(j)}_j\right] \\
&\qquad\qquad = \mathrm{TV}\left(\{\widehat{X}_{[1,s]},\widehat{X}_{[j,s]}\}_{1\leq s\leq\tau}\big|M_j; \{\widehat{X}^{(j)}_{[1,s]},\widehat{X}^{(j)}_{[j,s]}\}_{1\leq s\leq\tau})\big|M^{(j)}_j\right)
\end{aligned} \tag{50}$$

Since Sim doesn't actually change the first τ samples, we can actually drop this $X_{[a,s]}$ notation and just use $X_{[a,s]}$. Next note that, M_j is independent of $\{X_{[1,s]}\}_{1\le s\le\tau}$ and $M_j^{(j)}$ is independent of $\{X_{[j,s]}\}_{1\le s\le\tau}$. Hence, the first τ samples from arm 1 (resp arm j) are i.i.d from ν_1, and independent from the samples $\{X_{[j,s]}\}_{1\le s\le\tau}$ (resp. $\{X_{[1,s]}\}_{1\le s\le\tau}$). Using the TV bound $\mathrm{TV}(P_1\otimes Q_1; P_2\otimes Q_2)\le \mathrm{TV}(P_1;P_2)+\mathrm{TV}(Q_1;Q_2)$, for product measures $P_i\otimes Q_i$, we find that

$$\begin{aligned}\mathrm{TV}&\left[\mathsf{Sim}(\nu)\big|M_j;\mathsf{Sim}(\nu^{(j)})\big|M_j^{(j)}\right]\\ &\qquad\le \mathrm{TV}\left(\{X_{[1,s]}\}_{1\le s\le\tau};\{X_{[1,s]}^{(j)}\}_{1\le s\le\tau}|M_j^{(j)}\right)\\ &\qquad\qquad+\mathrm{TV}\left(\{X_{[j,s]}\}_{1\le s\le\tau}|M_j;\{X_{[j,s]}^{(j)}\}_{1\le s\le\tau}\right)\end{aligned}\tag{51}$$

By symmetry of construction, and symmetry of TV distance, its easy to check that

$$\begin{aligned}\mathrm{TV}\left(\{X_{[1,s]}\}_{1\le s\le\tau};\{X_{[1,s]}^{(j)}\}_{1\le s\le\tau}|M_j^{(j)}\right) &= \mathrm{TV}\left(\{X_{[1,s]}^{(j)}\}_{1\le s\le\tau}|M_j^{(j)};\{X_{[1,s]}\}_{1\le s\le\tau}\right)\\ &= \mathrm{TV}\left(\{X_{[j,s]}\}_{1\le s\le\tau}|M_j;\{X_{[j,s]}^{(j)}\}_{1\le s\le\tau}\right)\\ &= \mathrm{TV}\left(\{X_{[1,s]}\}_{1\le s\le\tau};\{X_{[j,s]}\}_{1\le s\le\tau}|M_j\right)\end{aligned}$$

Hence, it suffices to check

$$TV\left(\{X_{[1,s]}\}_{1\le s\le\tau};\{X_{[1,s]}^{(j)}\}_{1\le s\le\tau}|M_j\right)=\mathrm{TV}(\overline{X}_1;\overline{X}_j|E_j)\tag{52}$$

We first use a sufficient statistic argument to reduce the total variation from samples to a TV between empirical means:

Claim 4

$$TV\left(\{X_{[1,s]}\}_{1\le s\le\tau};\{X_{[1,s]}^{(j)}\}_{1\le s\le\tau}|M_j\right)=\mathrm{TV}(\overline{X}_1;\overline{X}_j|M_j)\tag{53}$$

The proof is somewhat pedantic, and so we prove in just a moment. To conclude, we finally is to note that $\overline{X}_j|M_j$ has the same distribution as $\overline{X}_j|E_j$, since

$$\begin{aligned}\mathbb{P}(\overline{X}_j\in A|M_j) &= \mathbb{P}(\overline{X}_j\in A\cap M_j)/\mathbb{P}(M_j)\\ &\overset{i}{=} \mathbb{P}(\overline{X}_j\in A\cap E_j,M_j)/\mathbb{P}(M_j)\\ &= \mathbb{P}(E_j)\mathbb{P}(\overline{X}_j\in A,M_j|E_j)(\mathbb{P}(E_j)/\mathbb{P}(M_j)\\ &\overset{ii}{=} \mathbb{P}(\overline{X}_j\in A|E_j)\mathbb{P}(M_j|E_j)\mathbb{P}(E_j)(\mathbb{P}(E_j\cap M_))/\mathbb{P}(M_j)\\ &= \mathbb{P}(\overline{X}_j\in A|E_j)/\mathbb{P}(M_j)\\ &= \mathbb{P}(\overline{X}_j\in A|E_j)\end{aligned}$$

Where i follows since $M_j\implies E_j$, and ii follows since M_j and $\overline{X}_j$ are conditionally independent given E_j.

Proof [Proof of Claim 4] Define the laws P_1,P_j over the $(X_1,\dots,X_\tau)\in\mathbb{R}^\tau$ where under P_1, $(X_1,\dots,X_\tau)$ have the law of $X_{[1,1]},\dots,X_{[1,\tau]}$, and under P_j, they have the law of $X_{[j,1]},\dots,X_{[j,\tau]}$.

We use $P_j(|M_j)$ to denote the law of $X_{[j,1]},\dots,X_{[j,\tau]}$ under M_j. Since $X_{[j,1]},\dots,X_{[j,\tau]}$ are independent of M_j given $\overline{X}_j$ (recall that M_j depends only on some internal randomness and E_j, which depends only on $\overline{X}_j$). Hence, letting $\overline{X} = \sum_{s=1}^{\tau} X_s$

$$\begin{aligned} &P_j((X_1,\dots,X_\tau) = (x_1,\dots,x_\tau)|M_j) \\ &\qquad = P_j((X_1,\dots,X_\tau) = (x_1,\dots,x_\tau)|\overline{X} = \bar{x})P_j(\overline{X} = \bar{x}|M_j) \end{aligned} \tag{54}$$

Moreover, since that since ν_1, ν_j come from a one-parameter exponential family,

$$P_1(\cdot|\overline{X} = \bar{x}) = P_j(\cdot|\overline{X} = \bar{x}) \tag{55}$$

Thus, we conclude that

$$\begin{aligned} \mathrm{TV}(P_1; P_j|M_j) &= \int_{\mathbf{x}\in\mathbb{R}^\tau} |dP_1(\mathbf{x}) - dP_j(\mathbf{x}|M_j)| \\ &= \int_{\bar{x}}\int_{\mathbf{x}:\sum_s \mathbf{x}_s = \tau\bar{x}} |dP_1(\bar{x})dP_1(\mathbf{x}|\bar{x}) - dP_j(\bar{x},M_j)dP_j(\mathbf{x}|\bar{x},M_j)| \\ &\overset{i}{=} \int_{\bar{x}}\int_{\mathbf{x}:\sum_s \mathbf{x}_s = \tau\bar{x}} |dP_1(\bar{x})dP_1(\mathbf{x}|\bar{x}) - dP_j(\bar{x},M_j)dP_j(\mathbf{x}|\bar{x})| \\ &\overset{ii}{=} \int_{\bar{x}}\int_{\mathbf{x}:\sum_s \mathbf{x}_s = \tau\bar{x}} |dP_1(\bar{x})dP_1(\mathbf{x}|\bar{x}) - dP_j(\bar{x},M_j)dP_1(\mathbf{x}|\bar{x})| \\ &= \int_{\bar{x}}\int_{\mathbf{x}:\sum_s \mathbf{x}_s = \tau\bar{x}} dP_1(\mathbf{x}|\bar{x})|dP_1(\bar{x}) - dP_j(\bar{x},M_j)| \\ &\overset{iii}{=} \int_{\bar{x}} |dP_1(\bar{x}) - dP_j(\bar{x},M_j)|(\int_{\mathbf{x}:\sum_s \mathbf{x}_s = \tau\bar{x}} dP_1(\mathbf{x}|\bar{x})) \\ &= \int_{\bar{x}} |dP_1(\bar{x}) - dP_j(\bar{x},M_j)| \\ &= \mathrm{TV}(\overline{X}_1; \overline{X}_j|M_j) \end{aligned}$$

where i follows from Equation 54, ii follows from Equation 55, iii is Fubini's theorem. ■

C.8.2. PROOF OF CONDITIONAL SIMULATED LE CAM(LEMMA 25)

Proof [Proof of Lemma 25]

$$\begin{aligned} \mathbb{P}_{\nu,\mathsf{Alg}}\mathbb{P}[W^c|M_j] &\overset{(i)}{=} \frac{1}{2}(\mathbb{P}_{\nu,\mathsf{Alg}}\mathbb{P}[W^c|M_j] + \mathbb{P}_{\nu^{(j)},\mathsf{Alg}}\mathbb{P}[W_j^c|M_j^{(j)}]) \\ &\overset{(ii)}{\geq} \sup_{A\in\mathcal{F}_T} \left|\mathbb{P}_{\nu,\mathsf{Alg}}\mathbb{P}[A|M_j] - \mathbb{P}_{\nu^{(j)},\mathsf{Alg}}\mathbb{P}[A|M_j^{(j)}]\right| - \mathrm{TV}[\mathsf{Sim}(\nu|M_j),\mathsf{Sim}(\nu^{(j)}|M_j^{(j)})] \\ &\qquad\qquad \overset{(iii)}{\geq} 1 - 2\mathbb{P}_{\nu,\mathsf{Alg}}\mathbb{P}[\{\hat{y}\neq 1\}|E_j] - \mathrm{TV}[\mathsf{Sim}(\nu|M_j),\mathsf{Sim}(\nu^{(j)}|M_j^{(j)})] \end{aligned} \tag{56}$$

where (i) follows from symmetry, (ii) follows from applying Proposition 10 using the measures $\nu|M_j$ and $\nu^{(j)}\big|M_j^{(j)}$, (this time, with TV instead of KL), and (iii) follows since

$$\begin{aligned} &\sup_{A \in \mathcal{F}_T} \left|\mathbb{P}_{\nu,\mathsf{Alg}}\mathbb{P}[A|M_j] - \mathbb{P}_{\tilde{\nu}_j,\mathsf{Alg}}\mathbb{P}[A|M_j^{(j)}]\right| \\ \geq\;& \mathbb{P}_{\nu,\mathsf{Alg}}\mathbb{P}[\{\hat{y}=1\}|M_j] - \mathbb{P}_{\nu^{(j)},\mathsf{Alg}}\mathbb{P}[\{\hat{y}\neq 1\}|M_j^{(j)}] \\ \geq\;& \mathbb{P}_{\nu,\mathsf{Alg}}\mathbb{P}[\{\hat{y}=1\}|M_j] - \mathbb{P}_{\nu^{(j)},\mathsf{Alg}}\mathbb{P}[\{\hat{y}\neq j\}|M_j^{(j)}] \\ =\;& 1 - \mathbb{P}_{\nu,\mathsf{Alg}}\mathbb{P}[\{\hat{y}\neq 1\}|M_j] - \mathbb{P}_{\nu^{(j)},\mathsf{Alg}}\mathbb{P}[\{\hat{y}\neq j\}|M_j^{(j)}] \\ =\;& 1 - 2\mathbb{P}_{\nu,\mathsf{Alg}}\mathbb{P}[\{\hat{y}\neq 1\}|M_j] \end{aligned}$$

where the last line is a consequence of Fact 3. ■

C.8.3. PROOF OF LEMMA 24

Proof [Proof of Lemma 24] Let call $\mathcal{F}$ denote the σ algebra generated by X. For any measures $\mathbb{P}$ and $\mathbb{Q}$ over $\mathcal{F}$, note that $\mathbb{P}[A] - \mathbb{Q}[A] = \mathbb{Q}[A^c] - \mathbb{P}[A^c]$. Hence,

$$\begin{aligned} \mathrm{TV}(\mathbb{P},\mathbb{Q}) &= \sup_{A\in\mathcal{F}} |\mathbb{P}[A] - \mathbb{Q}[A]| \\ &= \sup_{A\in\mathcal{F}} \max\{\mathbb{P}[A] - \mathbb{Q}[A], \mathbb{P}[A^c] - \mathbb{Q}[A^c]\} \\ &= \sup_{A\in\mathcal{F}} \mathbb{P}[A] - \mathbb{Q}[A] \end{aligned}$$

Now $\mathbb{Q} = \mathbb{P}[\cdot|B]$. Since any $A \in \mathcal{F}$ can be written as $A = A_B \sqcup A_{B^c}$ here $A_B \subset B$ and $A_{B^c} \subset B$

$$\begin{aligned} \mathrm{TV}(\mathbb{P},\mathbb{Q}) &= \sup_{A\in\mathcal{F}} \mathbb{P}[A] - \mathbb{Q}[A] \\ &= \sup_{A_B \cup A_{B^c} \in \mathcal{F}} \mathbb{P}[A_B \cup A_{B^c}] - \mathbb{Q}[A \cup A_{B^c}] \\ &= \sup_{A_B \cup A_{B^c} \in \mathcal{F}} \{\mathbb{P}[A_B] - \mathbb{Q}[A_B] + \mathbb{P}[A_{B^c}] - \mathbb{Q}[A_{B^c}]\} \\ &= \sup_{A_B \subset B \in \mathcal{F}} \{\mathbb{P}[A_B] - \mathbb{Q}[A_B]\} + \sup_{A_{B^c} \subset B^c \in \mathcal{F}} \{\mathbb{P}[A_{B^c}] - \mathbb{Q}[A_{B^c}]\} \end{aligned}$$

For any $A_B \subset B$, we see $\mathbb{Q}[A_B] = \mathbb{P}[A_B \cap B]/\mathbb{P}[B] = \mathbb{P}[A_B]/\mathbb{P}[B]$, so $\mathbb{P}[A_B] - \mathbb{Q}[A_B] = (1 - \mathbb{P}[B]^{-1})\mathbb{P}(A_B) \leq 0$, and thus $\sup_{A_B \subset B \in \mathcal{F}}\{\mathbb{P}[A_B] - \mathbb{Q}[A_B]\} = 0$, by taking $A_B = \emptyset$. On the other hand, for $A_{B^c} \subset B^c$, $\mathbb{Q}[A_{B^c}] = \mathbb{P}[A_{B^c} \cap B]/\mathbb{P}[B] = 0$, and thus,

$$\sup_{A_{B^c} \subset B^c \in \mathcal{F}} \{\mathbb{P}[A_{B^c}] - \mathbb{Q}[A_{B^c}]\} = \sup_{A_{B^c} \subset B^c \in \mathcal{F}} \mathbb{P}[A_{B^c}] = \mathbb{P}[B^c] \tag{57}$$

■

Appendix D. Proof of Proposition 6

We prove Proposition 6 by arguing via "algorithmic restrictions". The basic idea is that, if a lower bound holds for one MAB or TopK problem, then it should also hold for the "simpler" MAB or TopK problem which arises by "removing" some of the arms.

Formally, let $\nu = (\nu_a)_{a \in A}$ is an instance with arms indexed by $a \in A$ (where A is finite). For $B \subset A$, we define the restriction of ν to B, denoted $\nu_{|B}$, as the instance $(\nu_b)_{b \in B}$, indexed by arms $b \in B$. We let $\mathbf{S}_A$ and $\mathbf{S}_B$ denote the groups of permutations on the elements of A and B, respectively. Finally, given a subset $S \subset A$ of "good arms", recall that we say an algorithm Alg with decision rule $\widehat{S} \subset A$ is δ correct in identifying S over $\mathbf{S}_A(\nu)$ if $\mathbb{P}_{\pi(\nu),\mathsf{Alg}}[\widehat{S} = \pi(S)] \geq 1 - \delta$ for all $\pi \in \mathbf{S}_A$.

Lemma 28 (Lower Bounds from Restrictions) *Let $\nu = (\nu_a)_{a \in A}$ be an instance, $B \subset A$, and fix $\delta > 0$ and $b \in B$. Suppose that any algorithm $\mathsf{Alg}_{|B}$ which is δ-correct in identifying $S \cap B$ over $\mathbf{S}_B(\nu_{|B})$ satisfies the lower bound*

$$\mathbb{E}_{\sigma \sim \mathbf{S}_B} \mathbb{P}_{\sigma(\nu_{|B}),\mathsf{Alg}_{|B}}[N_{\sigma(b)}(T)] \geq \tau] \geq 1 - \eta \tag{58}$$

for some $\tau, \eta > 0$ (which may depend on ν, S, B, δ and b). Then any algorithm Alg which is δ-correct in identifying S over $\mathbf{S}_A(\nu)$ satisfies the analogous lower bound

$$\mathbb{E}_{\pi \sim \mathbf{S}_A} \mathbb{P}_{\pi(\nu),\mathsf{Alg}}[N_{\pi(b)}(T)] \geq \tau] \geq 1 - \eta \tag{59}$$

for the same τ and η.

To see how this lemma implies the bound for TopK, let $A = [n]$, and for $j \in [k]$ and $\ell \in [n] \setminus [k]$, define the sets $B_j = \{j\} \cup ([n] \setminus [k])$ and $B_\ell = \{\ell\} \cup ([n] \setminus [k])$. Finally, let $S = [k]$ denotes the top k arms, and $\widetilde{S} = [n] \setminus [k]$ denote the bottom $[n-k]$. Then, any δ-correct algorithm over $\mathbf{S}_n(\nu)$ equivalently identifies S and $\widetilde{S}$ with probability of error at most δ. Moreover, $B_j \cap S = \{j\}$, and $B_\ell \cap \widetilde{S} = \{\ell\}$. Now apply Lemma 28 using the MAB lower bounds from Proposition 5 for a) the problem of identifying ν_j from permutations of $\nu_{|B_j}$ and b) the problem ν_ℓ from permutations of $\nu_{|B_\ell}$.

D.1. Proof of Lemma 28

Proof Let Alg be be δ-correct in identifying S over $\mathbf{S}_A(\nu)$. Without loss of generality, we may assume that Alg is symmetric over $\mathbf{S}_A$ (Lemma 13). We will now construct an algorithm $\mathsf{Alg}_{|B}$ which "inherits" the correctness and complexity of Alg.

Claim 5 *There exists a symmetric (over $\mathbf{S}_B$) algorithm $\mathsf{Alg}_{|B}$ with decision rule $\widehat{S}_{|B} \subset B$ which satisfies, for all $b \in B$*

$$\mathbb{P}_{\nu_{|B},\mathsf{Alg}_{|B}}[N_b(T)] \geq \tau] \quad = \quad \mathbb{P}_{\nu,\mathsf{Alg}}[N_b(T)] \geq \tau] \quad \textit{and} \tag{60}$$

$$\mathbb{P}_{\nu_{|B},\mathsf{Alg}_{|B}}[\widehat{S}_{|B} = S \cap B] \quad \geq \quad \mathbb{P}_{\nu,\mathsf{Alg}}[\widehat{S} = S] \tag{61}$$

Assume the above claim. Since $\mathsf{Alg}_{|B}$ is symmetric and Alg is δ-correct, Equation 61 implies that $\mathsf{Alg}_{|B}$ is δ-correct over $\mathbf{S}_B(\nu_{|B})$, since all $\sigma \in \mathbf{S}_B$,

$$\mathbb{P}_{\sigma(\nu_{|B}),\mathsf{Alg}_{|B}}[\widehat{S}_{|B} = \sigma(S \cap B)] \geq \tau] = \mathbb{P}_{\nu_{|B},\mathsf{Alg}_{|B}}[\hat{y} \cap B = S \cap B]] \geq \mathbb{P}_{\nu,\mathsf{Alg}}[\widehat{S} = S] \geq 1 - \delta$$

Thus, by symmety of $\mathsf{Alg}_{|B}$ and the assumption of the lemma, we find for the choice of $b \in B$ and $\delta > 0$,

$$\mathbb{P}_{\nu_{|B},\mathsf{Alg}_{|B}}[N_b(T)] \geq \tau] = \mathbb{E}_{\sigma \sim \mathbf{S}_B} \mathbb{P}_{\sigma(\nu_{|B}),\mathsf{Alg}_{|B}}[N_{\sigma(b)}(T)] \geq \tau] \geq 1 - \eta \tag{62}$$

And hence, by Equation 60 and symmetry of Alg,

$$\mathbb{E}_{\pi \sim \mathbf{S}_A} \mathbb{P}_{\pi(\nu),\mathsf{Alg}}[N_{\pi(b)}(T)] \geq \tau] = \mathbb{P}_{\nu,\mathsf{Alg}}[N_b(T)] \geq \tau] = \mathbb{P}_{\nu_{|B},\mathsf{Alg}_{|B}}[N_b(T)] \geq \tau] \geq 1 - \eta \tag{63}$$

which concludes the proof. ■

To conclude, we just need to verify that we can construct $\mathsf{Alg}_{|B}$ as in Claim 5. To do this, let $\mathsf{Tr}_{|B}$ be a transcript samples $(X_{[b,s]})_{b \in B, s \in \mathbb{N}}$. For $a \in A \setminus B$, simulate a transcript $\mathsf{Tr}_{A \setminus B}$ of samples $(\widetilde{X}_{[a,s]})$ where $\widetilde{X}_{[a,s]} \overset{iid}{\sim} \nu_b$. Finally, let $\overline{\mathsf{Tr}}$ be the transcript obtained by concatening $\mathsf{Tr}_{|B}$ with the simulated transcript $\mathsf{Tr}_{A \setminus B}$, i.e. $\overline{X}_{[b,s]} = X_{[b,s]}$ for $b \in B$, and $\overline{X}_{[a,s]} = \widetilde{X}_{[a,s]}$. Finally, let $\mathsf{Alg}_{|B}$ be algorithm obtained by running Alg on the transcript $\mathsf{Tr}_{|B}$, with decision rule $\widehat{S}_{|B} = \widehat{S} \cap B$ (where $\widehat{S}$ is the decision rule of A).

Since $\overline{\mathsf{Tr}}$ has the same distribution as a transcript from ν when $\mathsf{Tr}_{|B}$ is drawn from $\nu_{|B}$, we immediate see that

$$\begin{aligned} \mathbb{P}_{\nu_{|B},\mathsf{Alg}_{|B}}[N_b(T)] \geq \tau] &= \mathbb{P}_{\nu,\mathsf{Alg}}[N_b(T)] \geq \tau] \quad \text{and} && (64) \\ \mathbb{P}_{\nu_{|B},\mathsf{Alg}_{|B}}[\hat{y} \cap B = S \cap B] &= \mathbb{P}_{\nu,\mathsf{Alg}}[\widehat{S} \cap B = S \cap B] \geq \mathbb{P}_{\nu,\mathsf{Alg}}[\widehat{S} = S] && (65) \end{aligned}$$

which verifies Equations 60 and 61. It's also easy to check that $\mathsf{Alg}_{|B}$ is symmetric, since permuting $\mathsf{Tr}_{|B}$ under a permutation $\sigma \in \mathbf{S}_B$ amounts to permuting $\overline{\mathsf{Tr}}$ by a permutation $\pi \in \mathbf{S}_B$ which fixes elements of $B \setminus A$. Hence, symmetryof $\mathsf{Alg}_{|B}$ follows from symmetry of Alg.

Appendix E. Upper Bound Proof

Proof [Proof of Theorem 7] Observe that if $\mathrm{TOP}_t \neq [k]$ then there is at least one arm from $[k]$ in TOP_t^c. Since we play the arm $l_t \in TOP_t^c$ with the largest upper-confidence-bound, the arms in $[k] \cap \mathrm{TOP}_t^c$ will eventually rise to the top. A mirror image of this process is also happening in the top empirical arms: if $\mathrm{TOP}_t \neq [k]$ then there is at least one arm from $[k]^c$ in TOP_t and the arms in $[k]^c \cap \mathrm{TOP}_t$ will eventually fall to the bottom since the arm $h_t \in \mathrm{TOP}_t$ with the lowest-confidence-bound is played. Thus, for some sufficently large t, the arms in $[k] \cap \mathrm{TOP}_t^c$ and $[k]^c \cap \mathrm{TOP}_t$ start to concentrate around the gap between the kth and $(k+1)$th arm. Because we play an arm from each side of the gap at each time, h_t and l_t, the empirical means eventually converge to their true means revealing the true ordering.

Preliminaries

We will use the following quantities throughout the proof. Let $U(t,\delta) \propto \sqrt{\frac{1}{t}\log(\log(t)/\delta)}$ such that $\max\{\mathbf{P}(\bigcup_{t=1}^{\infty}\{\widehat{\mu}_{i,t} - \mu_i \geq U(t,\delta)\}), \mathbf{P}(\bigcup_{t=1}^{\infty}\{\widehat{\mu}_{i,t} - \mu_i \leq -U(t,\delta)\})\} \leq \delta$ (see Jamieson et al. (2014, Lemma 1) or Kaufmann et al. (2015, Theorem 8) for an explicit expression). For any $i \in [n]$ define

$$\mathcal{E}_i = \begin{cases} \{\widehat{\mu}_{i,t} - \mu_i \geq -U(t, \frac{\delta}{2k})\} & \text{if } i \in \{1, \ldots, k\} \\ \{\widehat{\mu}_{i,t} - \mu_i \leq U(t, \frac{\delta}{2(n-k)})\} & \text{if } i \in \{k+1, \ldots, n\}. \end{cases}$$

In what follows assume that $\mathcal{E}_i$ hold for all $i \in [n]$ since, by the definition of $U(t,\delta)$,

$$\mathbf{P}\left(\bigcup_{i=1}^{n} \mathcal{E}_i^c\right) \leq \sum_{i=1}^{k} \frac{\delta}{2k} + \sum_{i=k+1}^{n} \frac{\delta}{2(n-k)} \leq \delta. \tag{66}$$

For any $j \in [k]^c$ define the random variable

$$\rho_j = \sup\{\rho > 0 : \widehat{\mu}_{j,t} - \mu_j < U(t, \tfrac{\rho\delta}{2k})\ \forall t\} \tag{67}$$

and the quantity $\tau_j = \min\{t : U(t, \frac{\rho_j\delta}{2k}) < \Delta_j/2\}$. Note that on the event $\mathcal{E}_j$ we have $\rho_j \geq \frac{k}{n-k}$ which guarantees that τ_j is finite, but we will show that ρ_j is typically actually $\Omega(1)$. For any $i \in [k]$ define $\tau_i = \min\{t : U(t, \frac{\delta}{2(n-k)}) < \Delta_i/2\}$, and note that on $\mathcal{E}_i$, we have $\widehat{\mu}_{i,t} - \mu_i \geq -U(t, \frac{\delta}{2k}) \geq -U(t, \frac{\delta}{2(n-k)})$. From these definitions, we conclude that

$$\widehat{\mu}_{j,t} - \mu_j \leq \Delta_j/2 \quad \forall t \geq \tau_j, j \in [k]^c \quad \text{ and } \quad \widehat{\mu}_{i,t} - \mu_i \overset{\mathcal{E}_i}{\geq} -\Delta_i/2 \quad \forall t \geq \tau_i, i \in [k]. \tag{68}$$

We leave the τ_i random variables unspecified for now but will later upper bound their sum.

Step 0: Correctness

Suppose $\mathrm{TOP}_\tau \neq [k]$. Then there exists an $i \in \mathrm{TOP}_\tau \cap [k]^c$ and $j \in \mathrm{TOP}_\tau^c \cap [k]$ such that

$$\mu_i \overset{(i)}{\geq} \widehat{\mu}_{i,N_i(t)} - U(N_i(t), \tfrac{\delta}{2(n-k)}) \overset{(ii)}{>} \widehat{\mu}_{j,N_j(t)} + U(N_j(t), \tfrac{\delta}{2k}) \overset{(iii)}{\geq} \mu_j$$

where (i) and (iii) employ Equation 66, and (ii) holds by assumption because of the stopping time τ. This display implies $\mu_j < \mu_i$, a contradiction, since the means in $[k]$ are strictly greater than those in $[k]^c$.

Step 1: $[k]$ rise to the top

Note that

$$[k] \neq \mathrm{TOP}_t, l_t \notin [k] \implies \exists i \in [k], j \in [k]^c : \widehat{\mu}_{j,N_j(t)} + U(N_j(t), \tfrac{\delta}{2k}) \geq \widehat{\mu}_{i,N_i(t)} + U(N_i(t), \tfrac{\delta}{2k}).$$

By the definition of ρ_j in Equation 67 and the above implication,

$$\mu_j + 2U(N_j(t), \tfrac{\delta\rho_j}{2k}) \geq \widehat{\mu}_{j,N_j(t)} + U(N_j(t), \tfrac{\delta}{2k}) \geq \widehat{\mu}_{i,N_i(t)} + U(N_i(t), \tfrac{\delta}{2k}) \overset{\mathcal{E}_i}{\geq} \mu_i$$

where the last inequality holds on event $\mathcal{E}_i$. By Equation 68, if $N_j(t) \geq \tau_j$ then $U(N_j(t), \frac{\delta\rho_j}{2k}) < \Delta_j/2$, but this would imply $\mu_i \leq \mu_j + 2U(N_j(t), \frac{\delta\rho_j}{2k}) < \mu_j + \Delta_j = \mu_k \leq \mu_i$ in the above display, a contradiction. That is, for any $j \in [k]^c$, the number of times that $l_t = j$ is bounded by τ_j. Since j could be any arm in $[k]^c$, we account for this by considering the sum of all possible upper bounds to conclude that

$$\{[k] \neq \mathrm{TOP}_t\} \cap \{t \geq \sum_{j=k+1}^{n} \tau_j\} \implies \{l_t \in [k]\}.$$

Step 2: Concentration at the Top

The previous step showed that for all $t \geq \sum_{j=k+1}^{n} \tau_j$ we either have $\mathrm{TOP}_t = [k]$, or $\mathrm{TOP}_t \neq [k]$

and $l_t \in [k]$. By Equation 68, for all $i \in [k]$, on the event $\mathcal{E}_i$, we have that $U(N_i(t), \frac{2k}{\delta}) < \Delta_i/2$ when $N_i(t) \geq \tau_i$. Thus, once $t \geq \sum_{i=1}^n \tau_i$ we conclude that if $[k] \neq \mathrm{TOP}_t$ then *at least* one arm $i \in [k]$ satisfies $U(N_i(t), \frac{2k}{\delta}) < \Delta_i/2$. That is

$$\{[k] \neq \mathrm{TOP}_t\} \cap \{t \geq \sum_{i=1}^n \tau_i\} \implies \{l_t \in [k]\} \cap \{\exists i \in [k] \cap \mathrm{TOP}_t^c : \widehat{\mu}_{i,N_i(t)} > \mu_i - \Delta_i/2\}. \tag{69}$$

Step 3: $[k]^c$ **fall to the bottom**

We first claim that if $t \geq \sum_{i=1}^k \tau_i + 2\sum_{i=k+1} \tau_i$ and $\mathrm{TOP}_t \neq [k]$ then $h_t \in [k]$. To see this, we will show that the number of times any $j \in [k]^c$ is equal to h_t is bounded by τ_j. Suppose not, so that $t \geq \sum_{i=1}^n \tau_i$, $j \in [k]^c$, $N_j(t) \geq \tau_j$, $j = h_t$, and $i \in [k]$ is the i in Equation 69, then

$$\widehat{\mu}_{j,N_j(t)} \overset{(68)}{<} \mu_j + \Delta_j/2 = \frac{\mu_k + \mu_j}{2} \leq \frac{\mu_i + \mu_{k+1}}{2} \leq \mu_i - \Delta_i/2 \overset{(69)}{<} \widehat{\mu}_{i,N_i(t)}$$

but $\widehat{\mu}_{j,N_j(t)} > \widehat{\mu}_{i,N_i(t)}$ contradicts the fact that $j \in \mathrm{TOP}_t$ and $i \notin \mathrm{TOP}_t$. As above, to account for all possible values of $j \in [k]^c$ we assume that they all saturate their bounds. Thus,

$$\{[k] \neq \mathrm{TOP}_t\} \cap \{t \geq \sum_{i=1}^k \tau_i + 2\sum_{i=k+1} \tau_i\} \implies \{h_t \in [k]\}. \tag{70}$$

Now we will show that the number of times any $i \in [k]$ is equal to h_t is bounded by τ_i. Note that

$$[k] \neq \mathrm{TOP}_t, h_t \in [k] \implies \exists i \in [k], j \in [k]^c : \widehat{\mu}_{j,N_j(t)} - U(N_j(t), \tfrac{\delta}{2(n-k)}) \geq \widehat{\mu}_{i,N_i(t)} - U(N_i(t), \tfrac{\delta}{2(n-k)}).$$

On events $\mathcal{E}_j$ and $\mathcal{E}_i$ we have that

$$\begin{aligned}\mu_j \overset{\mathcal{E}_j}{\geq} \widehat{\mu}_{j,N_j(t)} - U(N_j(t), \tfrac{\delta}{2(n-k)}) \geq \widehat{\mu}_{i,N_i(t)} - U(N_i(t), \tfrac{\delta}{2(n-k)}) &\overset{\mathcal{E}_i}{\geq} \mu_i - U(N_i(t), \tfrac{\delta}{2k}) - U(N_i(t), \tfrac{\delta}{2(n-k)}) \\ &\geq \mu_i - 2U(N_i(t), \tfrac{\delta}{2(n-k)}).\end{aligned}$$

By Equation 68, if $N_i(t) \geq \tau_i$ then after simplifying the above display we have

$$\mu_j \geq \mu_i - 2U(N_i(t), \tfrac{\delta}{2(n-k)}) > \mu_i - \Delta_i/2 = \frac{\mu_i + \mu_{k+1}}{2}$$

which is a contradiction. Accounting for all values of $i \in [k]$, we conclude that

$$\{[k] \neq TOP_t\} \cap \{t \geq 2\sum_{i=1}^n \tau_i\} \implies \{h_t \notin [k]\}. \tag{71}$$

Combining Equations 70 and 71 we conclude that $TOP_t = [k]$ whenever $t \geq 2\sum_{i=1}^n \tau_i$.

Step 4: The stopping condition is met

While $TOP_t = [k]$ whenever $t \geq 2\sum_{i=1}^n \tau_i$, we still must wait until the stopping condition is met. For any $i \in [k]$, if $N_i(t) \geq \tau_i$ then

$$\widehat{\mu}_{i,N_i(t)} - U(N_i(t), \tfrac{2(n-k)}{\delta}) > \mu_i - \Delta_i/2 \geq \frac{\mu_k + \mu_{k+1}}{2} \quad \forall i \in [k].$$

And for any $j \in [k]^c$, if $N_j(t) \geq \tau_j$ then

$$\widehat{\mu}_{j,N_j(t)} + U(N_j(t), \tfrac{2k}{\delta}) < \mu_j + \Delta_j/2 \leq \frac{\mu_k + \mu_{k+1}}{2} \quad \forall j \notin [k].$$

All arms satisfy these conditions after at most an additional $\sum_{i=1}^n \tau_i$ pulls. Thus, the stopping condition of (9) is met after at most $3\sum_{i=1}^n \tau_i$ total pulls.

Step 5: Counting the number of measurements

Recall that $3\sum_{i=1}^n \tau_i$ is a random variable because ρ_j for $j \in [k]^c$ are random variables. Recalling the definitions of τ_j preceding Equation 68, we note that

$$\min\{t : U(t,s) < \Delta/2\} \leq c\Delta^{-2}\log(\log(\Delta^{-2})/s)$$

for some universal constant c. For $i \in [k]$ this means $\tau_i = \min\{t : U(t, \frac{\delta}{2(n-k)}) < \Delta_i/2\} \leq c\Delta_i^{-2}\log(2(n-k)\log(\Delta_i^{-2})/\delta)$. For $j \in [k]^c$ we have

$$\tau_j = \min\{t : U(t, \tfrac{\rho_j\delta}{2k}) < \Delta_j/2\} \leq c\Delta_j^{-2}\log(2k\log(\Delta_j^{-2})/\delta) + c\Delta_j^{-2}\log(1/\rho_j).$$

By the definition of $U(\cdot,\cdot)$ and ρ_j we have that $\mathbf{P}(\rho_j \leq \rho) \leq \frac{\rho\delta}{2k} < \rho$, so reparameterizing with $\rho = \exp(-s\Delta_j^2)$

$$\mathbf{P}(\Delta_j^{-2}\log(1/\rho_j) \geq s) \leq \exp(-s\Delta_j^2/2)$$

which implies $\Delta_j^{-2}\log(1/\rho_j)$ is an independent sub-exponential random variable. Using standard techniques for sums of independent random variables (see (Jamieson et al., 2014, Lemma 4) for an identical calculation) we observe that with probability at least $1-\delta$

$$\sum_{j=k+1}^{n} \Delta_j^{-2}\log(\tfrac{1}{\rho_j}) \leq \sum_{j=k+1}^{n} c'\Delta_j^{-2}\log(1/\delta)$$

for some universal constant c'. Combining the contributions of the deterministic components of τ_i and τ_j obtains the result. ■

E.1. Upper Bounds for Permutations

In this section, we present a nearly-matching upper bound for permutations (Theorem 29). For simplicity, we consider the setting where each measure ν_a is 1-subGaussian, and has mean μ_a. We let $\mu_{(1)} > \mu_{(2)} \geq \dots \geq \mu_{(n)}$, denote the sorted means, and set $\Delta_i = \mu_{(1)} - \mu_{(i)}$.

Theorem 29 *In the setting given above, there exists a δ- algorithm* Alg *which, given knowledge of the means $\mu_{(1)}$ and $\mu_{(2)}$, returns the top arm with expected sample complexity*

$$\mathbb{E}_{\nu,\mathsf{Alg}}[T] \lesssim \frac{\log(1/\delta)}{\Delta_2^2} + \sum_{i=1}^{n} \frac{\log\log(\min\{n, \Delta_i^{-1}\})}{\Delta_i^2} \tag{72}$$

We remark that this upper bound matches our lower bound up to the doubly-logarithmic factor $\log\log(\min\{n, \Delta_i^{-1}\})$. We believe that one could remove this factor when the means are known up to a permutation, though closing this small gap is beyond the scope of this work. To prove the above theorem, we combine the following Lemma with the best-arm algorithm from Chen and Li (2015):

Proposition 30 *Suppose that for each δ, there exists an (unconstrained)* MAB *algorithm* Alg_δ *which is δ-correct for 1-subGaussian distributions with unconstrained means, and satisfies $\mathbb{E}_{\nu,\mathsf{Alg}_\delta}[T] \le H_1(\nu) + H_2(\nu)\log(1/\delta)$. Then, there exists an an* MAB *algorithm which, give knowledge of the the best mean μ_1 and the second best mean μ_2, satisfies*

$$\mathbb{E}_{\nu,\mathsf{Alg}}[T] \lesssim \frac{\log(1/\delta)}{\Delta_2^2} + H_1(\nu) + H_2(\nu) \tag{73}$$

Proof Fix constants c_1 and c_2 to be chosen later The algorithm proceeds in stages: at round k, set $\delta_k = 10^{-k}$, and run Alg_{δ_k} to get an estimate $\hat{a}_k$ of the best arm. Then, sample $\hat{a}_k$ $\frac{c_1}{\Delta_2^2}\log(c_2k^2/\delta)$ times to get an estimate $\widehat{\mu}^k$, and return $\widehat{a} = \hat{a}_k$ if $\widehat{\mu}^k > \mu_1 - \Delta_2/2$. By a standard Chernoff bound, we can choose c_1 so that $\widehat{\mu}_k$ satisfies the following

$$\mathbb{P}(\widehat{\mu}^k > \mu_1 - \Delta_2/2 \big| \hat{a}_k = a^*) \ge 1 - 2\delta/c_2k^2 \quad \text{and} \quad \mathbb{P}(\widehat{\mu}^k > \mu_1 - \Delta_2/2 \big| \hat{a}_k \ne a^*) \le 2\delta/c_2k^2$$

Hence,

$$\begin{aligned}
\mathbb{P}(\widehat{a} \ne a^*) &\le \sum_{k=1}^{\infty} \mathbb{P}(\{\widehat{\mu}^k \ne a^*\} \wedge \{\widehat{\mu}^k > \mu_1 - \Delta_2/2\}) && (74)\\
&\le \sum_{k=1}^{\infty} \mathbb{P}(\{\widehat{\mu}^k > \mu_1 - \Delta_2/2\} \big| \{\widehat{\mu}^k \ne a^*\}) \le \frac{2\delta}{c_2}\sum_{k=1}^{\infty} k^{-2} = \frac{\pi^2}{3c_2} && (75)
\end{aligned}$$

Hence, choosing $c_2 = 3/\pi^2$ ensures that Alg is δ-correct. Moreover, we can bound

$$\mathbb{E}_{\nu,\mathsf{Alg}}[T] \le \sum_{k=1}^{\infty} \mathbb{P}(E_{k-1}) * \{\frac{c_1}{\Delta_2^2}\log(c_2k^2/\delta) + \mathbb{E}_{\nu,\mathsf{Alg}_{\delta_k}}[T]\} \tag{76}$$

where E_{k-1} is the event that the algorithm has not terminated by stage $k-1$. Note that if the algorithm has not terminated at a stage j, then it is not the case that $\hat{a}_j = a^*$ and $\{\widehat{\mu}^j > \mu_1 - \Delta_2/2\})$. By a union bound, the probability that these two events don't occur is at most $1 - \delta_k - \frac{2\delta}{c_2k^2} \le 1 - (\delta_k + 2\delta/c_2) \le 1/2$. Hence, bounding $\mathbb{E}_{\nu,\mathsf{Alg}_{10^{-k}}}[T] \le H_1(\nu) + kH_2(\nu)\log 10$, and using independence of the rounds have

$$\begin{aligned}
\mathbb{E}_{\nu,\mathsf{Alg}}[T] &\le \sum_{k=1}^{\infty} 2^{1-k} * \{\frac{c_1}{\Delta_2^2}\log(c_2k^2/\delta) + H_1(\nu) + kH_2(\nu)\log 10)\} && (77)\\
&\lesssim \frac{\log(1/\delta)}{\Delta_2^2} + H_1(\nu) + H_2(\nu) && (78)
\end{aligned}$$

■