JMLR: Workshop and Conference Proceedings vol 49:1–50, 2016

Best-of-K Bandits

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

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

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

Abstract

This paper studies the Best-of-K Bandit game: At each time the player chooses a subset S among all N-choose-K possible options and observes reward max(X(i) : i in S) where X is a random vector drawn from a joint distribution. The objective is to identify the subset that achieves the highest expected reward with high probability using as few queries as possible. We present distribution-dependent lower bounds based on a particular construction which force a learner to consider all N-choose-K subsets, and match naive extensions of known upper bounds in the bandit setting obtained by treating each subset as a separate arm. Nevertheless, we present evidence that exhaustive search may be avoided for certain, favorable distributions because the influence of high-order correlations may be dominated by lower order statistics. Finally, we present an algorithm and analysis for independent arms, which mitigates the surprising non-trivial information occlusion that occurs due to only observing the max in the subset. This may inform strategies for more general dependent measures, and we complement these result with independent-arm lower bounds.

1. Introduction

This paper addresses a variant of the stochastic multi-armed bandit problem, where given n arms associated with random variables $X_1, \ldots, X_n$, and some fixed $1 \leq k \leq n$, the goal is to identify the subset $S \in \binom{[n]}{k}$ that maximizes the objective $\mathbb{E}\left[\max_{i \in S} X_i\right]$. We refer to this problem as "Best-of-K" bandits to reflect the reward structure and the limited information setting where, at each round, a player queries a set S of size at most k, and only receives information about arms $X_i : i \in S$: e.g. the vector of values of all arms in S, $\{X_i : i \in S\}$ (semi-bandit), the index of a maximizer (marked bandit), or just the maximum reward over all arms $\max_{i \in S} X_i$ (bandit). The game and its valid forms of feedback are formally defined in Figure 1.

While approximating the Best-of-K problem and its generalizations have been given considerable attention from a computational angle, in the regret setting (Hofmann et al., 2011; Raman et al., 2012; Radlinski et al., 2008; Yue and Guestrin, 2011; Streeter and Golovin, 2009), this work aims at characterizing its intrinsic statistical difficulty as an identification problem. Not only do identification algorithms typically imply low (simple) regret algorithms by first exploring and then exploiting, every result in this paper can be easily extended to the PAC learning setting where we aim to find a set whose reward is within ϵ of the optimal, a pure-exploration setting of interest for science applications (Kaufmann et al., 2015; Kaufmann and Kalyanakrishnan, 2013; Hao et al., 2013).

For joint reward distributions with high-order correlations, we present distribution-dependent lower bounds which force a learner to consider all subsets $S \in \binom{[n]}{k}$ in each feedback model of interest, and match naive extensions of known upper bounds in the bandit setting obtained by treating each subset S as a separate arm. Nevertheless, we present evidence that exhaustive search may be avoided for certain, favorable distributions because the influence of high-order order correlations may be dominated by lower order statistics. Finally, we present an algorithm and analysis for independent arms, which mitigates the surprising non-trivial information occlusion that occurs in the bandit and marked bandit feedback models. This may inform strategies for more general dependent measures, and we complement these result with independent-arm lower bounds.

1.1. Motivation

In the setting where $X_i \in \{0,1\}$, one can interpret the objective $\max_{S\in\binom{[n]}{k}} \mathbb{E}[\max_{i\in S} X_i]$ as trying to find the set of items which affords the greatest coverage. For example, instead of using spread spectrum antibiotics which have come under fire for leading to drug-resistant "super bugs" (Huycke et al., 1998), consider the doctor that desires to identify the best k subset of narrow spectrum antibiotics that leads to as many favorable outcomes as possible. Here each draw from X_i represents the ith treatment working on a random patient, and for antibiotics, we may assume that there are no synergistic effects between different drugs in the treatment. Thus, the antibiotics example falls under the *bandit feedback* setting since k treatments are selected but it is only observed if at least one k-tuple of treatment led to a favorable outcome: no information is observed about any particular treatment.

Now consider content recommendation tasks where k items are suggested and the user clicks on either 1 or none. Here, each draw of $X \in \{0,1\}^n$ encodes the users joint preferences for the items in question, where $X_i = 1$ if the user likes the i-th item, and $X_i = 0$ otherwise. Because allow for dependencies between the entries of X, our model can account for complex correlations between users' preferences for different items. In this setting, we only get to observe one item which the user has clicked on, which we designate *marked-bandit* feedback.

Our final example comes from virology where multiple experiments are prepared and performed k at a time, resulting in k simultaneous, noisy responses (Hao et al., 2013); this motivates our consideration of the *semi-bandit* feedback setting.

1.2. Problem Description

We denote $[n] = \{1, 2, \ldots, n\}$. For a finite set W, we let 2^W denote its power set, $\binom{W}{p}$ denote the set of all subsets of W of size p, and write $V \sim \mathrm{Unif}[W, p]$ to denote that V is drawn uniformly from $\binom{W}{p}$. If X is a length n vector (binary, real or otherwise) and $W \subset [n]$, we let X_W denote the sub-vector indexed by entries $i \in W$.

In what follows, let $X = (X_1, \ldots, X_n)$ be a random vector drawn from the probability distribution ν over $\{0,1\}^n$. Unless otherwise stated, ν can be an arbitrary distribution over $\{0,1\}^n$ in that the entries of X may be dependent. If the entries are X are jointly independent, then we will call ν an *independent measure*.

We refer to the index $i \in [n]$ as the i-th arm, set $\mu_i = \mathbb{E}[X_i]$ to be the expected reward of the i-th arm, let ν_i denote the marginal distribution of its corresponding entry in X, e.g. $(\mathbb{E}_\nu[X])_i = \mathbb{E}_{\nu_i}[X_i] = \mu_i$. We define $\mathcal{S} := \binom{[n]}{k}$, and for a given $S \in \mathcal{S}$, we we call $\mathbb{E}[\max_{i\in S} X_i]$ the *expected reward* of S, and refer casually to the random instantiations $\max_{i\in S} X_i$ as simply the reward of S.

Best-of-k Bandits Game

for $t = 1, 2, \ldots$

Player picks $S_t \in \mathcal{S}$ and adversary simultaneously picks $x_t \in \{0,1\}^n$

$$\text{Player observes} \begin{cases} \text{Bandit feedback:} & \max_{i \in S_t} x_{t,i} \\ \text{Marked-Bandit feedback:} & \begin{cases} \emptyset & \text{if } x_{t,i} = 0\ ,\ \ \forall i \in S_t \\ \text{unif}(\arg\max_{i \in S_t} x_{t,i}) & \text{otherwise.} \end{cases} \\ \text{Semi-bandit feedback:} & x_{t,i}\ \ \forall i \in S_t \end{cases}$$

Figure 1: Best-of-k Bandits game for the different types of feedback considered. While this work is primarily interested in stochastic adversaries, our lower bound construction also has consequences for non-stochastic adversaries. Moreover, in marked feedback, we might consider non-uniform and even adversarial marking.

At each time t, nature draws a rewards vector $x_t = X$ where X is i.i.d from ν. Simultaneously, our algorithm *queries* a subset $S_t \in \mathcal{S}$ of k arms, and we refer to the entries $i \in S_t$ as the arms *pulled* by the query. As we will describe later, this problem has previously been studied in a *regret* framework, where a time horizon $T \in \mathbb{N}$ is fixed and an algorithm's objective is to minimize its regret

$$R_\nu(T) = T \max_{S \in \mathcal{S}} \mathbb{E}_\nu[\max_{i \in S} X_i] - \mathbb{E}_\nu[\sum_{t=1}^{T} \max_{i \in S_t} X_i]. \tag{1}$$

In this work, we are more concerned with the problem of identifying the best subset of k arms. More precisely, for a given measure ν, denote the optimal subset

$$S^* := \arg\max_{S \in \mathcal{S}} \mathbb{E}_\nu \left[\max_{i \in S} X_i \right] \tag{2}$$

and let T_S denote the (possibly random) number of times a particular subset $S \in \binom{[n]}{k}$ has been played before our algorithm terminates. The identification problem is then

Definition 1 (Best-of-K Subset Identification) *For any measure ν and fixed $\delta \in (0,1)$, return an estimate $\widehat{S}$ such that $\mathbb{P}_\nu(\widehat{S} \neq S^*) \leq \delta$, and which minimizes the sum $\sum_{S \in \binom{[n]}{k}} T_S$ either in expectation, or with high probability.*

Again, we remind the reader that an algorithm for Best-of-K Subset Identification can be extended to active PAC learning algorithm, and to an online learning algorithm with low regret (with high probability) (Kaufmann et al., 2015; Kaufmann and Kalyanakrishnan, 2013; Hao et al., 2013).

1.3. Related Work

Variants of Best-of-K have been studied extensively in the context of online recommendation and ad placement (Yue and Guestrin, 2011; Hofmann et al., 2011; Raman et al., 2012). For example, Radlinski et al. (2008) introduces "Ranked Bandits" where the arms X_i are stochastic random variables, which take a value 1 if the t-th user finds item i relevant, and 0 otherwise. The goal is to recommend an ordered list of items $S = (i_1, \dots, i_k)$ which maximizes the probability of a click on any item in the list, i.e. $\max_{i \in S} X_i$, and observes the first item (if any) that the user clicked on.

Streeter and Golovin (2009) generalizes the Best-of-K problem to the online maximization of a sequence of monotone, submodular function $\{F_t(S)\}_{1 \le t \le T}$ subject to knap-sack constraints $|S| \le k$, under a variety of feedback models. Since the function $S \mapsto \max_{i \in S} X_i$ is submodular, identifying S^* corresponds to special case of optimizing the monotone, submodular function $F(S) := \mathbb{E}[\max_{i \in S} X_i]$ subject to these same constraints. Streeter and Golovin (2009) also consider a variety of feedback models, including the ranked/cascading feedback described above.

Streeter and Golovin (2009), Yue and Guestrin (2011), and Radlinski et al. (2008) propose online variants of a well-known greedy offline submodular optimization algorithm (see, for example Iyer and Bilmes (2013)) , which attain $(1 - \frac{1}{e})$-approximate regret guarantees of the form

$$\sum_{t=1}^{T} F_t(S_t) - \left(1 - \frac{1}{e}\right) \max_{S^* : |S^*| \le k} F_t(S^*) \le R(T) \tag{3}$$

where $R(T)$ is some regret term that decays as $O(\text{poly}(n,k)) \cdot o(T)$. Computationally, this $1 - \frac{1}{e}$ is the best one could hope for in general: Best-of-K and Ranked Bandits are online variants of the Max-K-Coverage problem, which cannot be approximated to within a factor of $1 - \frac{1}{e} + \epsilon$ for any fixed $\epsilon > 0$ under standard hardness assumptions (Vazirani, 2013). For completeness, we provide a formal reduction from Best-of-K identification to Max-K-Coverage in Appendix A.

For independent measures, however, one can circumvent the $1 - \frac{1}{e}$ approximation factor. For example, Kveton et al. (2015) introduces non-approximate low-regret UCB-like algorithms for the ranked bandits game from Radlinski et al. (2008), which they call "cascading bandits", under the assumption that the rewards X_i are binary and independent. These guarantees are related to our upper bounds for independent distributions, but with two key differences: first, their feedback model sits strictly between bandit and semi-bandit feedback (figure 1), but is not compararable to our "marked-bandit" game, in which feedback is given uniformly at random, rather than based on the ordering of a list.

Finally, Gopalan et al. (2014) obtains regret upper bounds for the Best-of-K bandit setting as a corrolary of a general Thompson Sampling scheme. They obtain regret on the order of $O(\binom{n-1}{k} \log T)$ for bandits, and $O(n \log T)$ for the semi-bandit setting. However, these regret upper bounds seem to require the assumption that the entries X_i are independent. In light of our pure exploration upper bounds for independent measures, we conjecture that their results are quite loose, and should actually scale like $O(n \log T)$ for bandit feedback and $O(\frac{n}{k} \log T)$ for semi-bandits.

1.4. Our Contributions

Focusing on the stochastic pure-exploration setting with binary rewards, our contributions are as follows:

- We propose a family of joint distributions such that any algorithm that solves the best of k identification problem with high probability must essentially query all $\binom{n}{k}$ combinations of arms. Our lower bounds for the bandit case are nearly matched by trivial identification and regret algorithms that treat each k-subset as an independent arm. For semi-bandit feedback, our lower bounds are exponentially smaller in k than those for bandit feedback (though still requiring exhaustive search). To better understand this gap, we sketch an upper bound that achieves the lower bound for a particular instance of our construction. While in the general binary case, the difficulty of marked bandit feedback is sandwiched between bandit and semi-bandit feedback, in our particular construction we show that marked bandit feedback has no benefit over bandit feedback. In particular, for worst-case instances, our lower bounds for marked bandits are matched by upper bounds based on algorithms which *only take advantage of bandit feedback*.

- Our construction plants a k-wise dependent set S^* among $\binom{n}{k} - 1$ k-wise independent sets, creating a needle-in-a-haystack scenario. One weakness of this construction is that the gap between the rewards of the best and second best subset are exponentially small in k. This is particular to our construction, but not to our analysis: We present a partial converse which establishes that, for any two $k-1$-wise independent distributions defined over $\{0,1\}^k$ with identical marginal means μ, the difference in expected reward is exponentially small in k[1]. This begs the question: can low order correlation statistics allows us to neglect higher order dependencies? And can this property be exploited to avoid combinatorially large sample complexity in favorable scenarios with moderate gaps?

- We lay the groundwork for algorithms for identification under favorable, though still dependent, measures by designing a computationally efficient algorithm for *independent* measures for the marked, semi-bandit, and bandit feedback models. Though independent semi-bandits is straightforward (Jun et al., 2016), special care needs to be taken in order to address the information occlusion that occurs in the bandit and marked-bandit models, even in this simplified setting. We provide nearly matching lower bounds, and conclude that even for independent measures, bandit feedback may require exponentially (in k) more samples than in the semi-bandit setting.

2. Lower Bound for Dependent Arms

Intuitively, the best-of-k problem is hard for the dependent case because the high reward subsets may appear as a collection of individually low-pay off arms if not sampled together. For instance, for $k = 2$, if $X_1 = \text{Bernoulli}(1/2)$, $X_2 = 1 - X_1$, and $X_i = \text{Bernoulli}(3/4)$ for all $3 \leq i \leq n$, then clearly $\mathbb{E}[\max\{X_1, X_2\}] = 1$ is the best subset because $\mathbb{E}[\max\{X_1, X_i\}] = 1 - (1/2)(1/4) = 7/8$ and $\mathbb{E}[\max\{X_i, X_j\}] = 1 - (1/4)^2 = 15/16$ for all $3 \leq i \leq j \leq n$. However, identifying set $\{1, 2\}$ appears difficult as presumably one would have to consider all $\binom{n}{2}$ sets since if X_1 and X_2 are not queried together, they appear as $\text{Binomial}(1/2)$.

Our lower bound generalizes this construction by introducing a measure ν such that (1) the arms in the optimal set S^* are dependent but (2) the arms in every other non-optimal subset of arms $S \in \mathcal{S} - S^*$ are mutually independent. This construction amounts to hiding a "needle-in-a-haystack" S^* among all other $\binom{n}{k} - 1$ subsets, requiring any possibly identification to examine most elements of $\mathcal{S}$.

[1] Note that our construction requires all subset of $k-1$ of S^* to be independent

We now state our theorem, which characterizes the difficulty of recovering S^* arms in terms of the gap Δ between the expected reward of S^* and of the second best subset

$$\Delta := \mathbb{E}_\nu \left[\max_{i \in S^*} X_i\right] - \max_{S \in \mathcal{S} \setminus S^*} \mathbb{E}_\nu \left[\max_{i \in S} X_i\right] \tag{4}$$

Theorem 2.1 (Dependent) *Fix $k, n \in \mathbb{N}$ such that $2 \le k < n$. For any $\epsilon \in (0, 1]$ and $\mu \in (0, 1/2]$ there exists a distribution ν with $\Delta = \epsilon\mu^k$ such that any algorithm that identifies S^* with probability at least $1 - \delta$ requires, in expectation, at least*

$$(i) \quad \frac{4(1 - \epsilon(\frac{\mu}{1-\mu})^k)}{3} \cdot \left(1 - (1-\mu)^k\right)(1-\mu)^k \binom{n}{k} \Delta^{-2} \log(\tfrac{1}{2\delta}) \quad \textit{(marked-)bandit, or}$$

$$(ii) \quad \frac{2}{3}\mu^{2k}(1-\epsilon)\binom{n}{k}\Delta^{-2}\log(\tfrac{1}{2\delta}) \quad \textit{semi-bandit}$$

observations. In particular, for any $0 < \xi \le (2k)^{-k}$ there exists a distribution ν with $\Delta = \xi$ that requires at least $\frac{1}{3}\binom{n}{k}\Delta^{-2}\log(\frac{1}{2\delta})$ (marked-)bandit observations. And for any $0 < \xi \le 2^{-k-1}$ there exists a distribution ν with $\Delta = \xi$ that requires at least $\frac{1}{3}2^{-2k}\binom{n}{k}\Delta^{-2}\log(\frac{1}{2\delta})$ semi-bandit observations.

Remark 2.1 *Marked-bandit feedback provides strictly less information than semi-bandit feedback but at least as much as bandit feedback. The above lower bound for marked-bandit feedback and the nearly matching upper bound for bandit feedback remarked on below suggests that marked-bandit feedback may provide no more information than bandit feedback. However, the lower bound holds for just a particular construction and in Section 3 we show that there exist instances in which marked-bandit feedback provides substantially more information than merely bandit feedback.*

In the construction of the lower bound, $S^* = [k]$ and all other subsets behave like completely independent arms. Each individual arm has mean μ, i.e. $\mathbb{E}_\nu[X_i] = \mu$ for all i, so each $S \ne S^*$ has a bandit reward of $\mathbb{E}_\nu[\max_{i \in S} X_i] = 1 - (1-\mu)^k$. The scaling $(1 - (1-\mu)^k)(1-\mu)^k$ in the number of bandit and marked-bandit observations corresponds to the variance of this reward and captures the property that the number of times a set needs to be sampled to accurately predict its reward is proportional to its variance. Since $\mu \le 1/2$, we note that the term $1 - \epsilon(\frac{\mu}{1-\mu})^k$ is typically very close to 1, unless μ is nearly $1/2$ *and* ϵ is nearly 1.

While the lower bound construction makes it necessary to consider each subset $S \in \binom{[n]}{k}$ individually for all forms of feedback feedback, semi-bandit feedback presumably allows one to detect dependencies much faster than bandit or marked-bandit feedback, resulting in an exponentially smaller bound in k. Indeed, Remark E.2 describes an algorithm that uses the parity of the observed rewards that nearly achieves the lower bound for semi-bandits for the constructed instance when $\mu = 1/2$. However, the authors are unaware of more general matching upper bounds for the semi-bandit setting and consider this a possible future avenue of research.

2.1. Comparison with Known Upper Bounds

By treating each set $S \in \mathcal{S}$ as an independent arm, standard best-arm identification algorithms can be applied to identify S^*. The KL-based *LUCB* algorithm from Kaufmann and Kalyanakrishnan

(2013) requires $O(\Delta^2(1-(1-\mu)^k)(1-\mu)^k\binom{n}{k}\cdot k\log n)$ samples, matching our bandit lower bound up to a a multiplicative factor of $k\log n$ (which is typically dwarfed by $\binom{n}{k}$). The *lil'UCB* algorithm of Jamieson et al. (2014) avoids paying this multiplicative $k\log n$ factor, but at the cost of not adapting to the variance term $(1-(1-\mu)^k)(1-\mu)^k$. Perhaps a KL- or variance-adaptive extension of *lil-UCB* could attain the best of both worlds.

From a regret perspective, the exact construction as used in the proof of Theorem 2.1 can be used in Theorem 17 of Kaufmann et al. (2015) to state a lower bound on the regret after $T=\sum_{S\in\mathcal{S}}T_s$ bandit observations. Specifically, if an algorithm obtain a stochastic regret $R_T(\nu)=o(T^\alpha)$ for all $\alpha\in(0,1]$, then for all $S\in\binom{[n]}{k}-S^*$, we have $\liminf_{T\to\infty}\frac{\mathbb{E}_\nu[T_S]}{\log(T)}\geq\frac{(1-(1-\mu)^k)(1-\mu)^k}{\Delta^2}$ where Δ is given in Theorem 2.1. Alternatively, in an adversarial setting, the above construction with $\mu=1/2$ also implies a lower bound of $\sqrt{2^{-O(k)}\binom{n}{k}T}=\sqrt{\binom{\Omega(n)}{k}T}$ for any algorithm over a time budget T. Both of these regret bounds are matched by upper bounds found in Bubeck and Cesa-Bianchi (2012).

2.2. Do Large Sample Complexities Require Small Gaps?

While Theorem 2.1 proves the existence of a family of instances in which $\binom{n}{k}\Delta^{-2}\log(1/\delta)$ samples are necessary to identify the best k-subset, the possible gaps Δ are restricted to be no larger than $\min\{\mu^k,(1-\mu)^k\}$. The following theorem demonstrates that, for the construction in our lower bounds, these restrictions on the gaps are tight. More precisely, we show that when all subsets of size $k-1$ have identical lower-order statistics, then the gaps must be exponentially small in k:

Theorem 2.2 *Let $X=(X_1,\dots,X_k)$ be a random variable supported on $\{0,1\}^k$ with $k-1$-wise independent marginal distributions, such that $\mathbb{E}[X_i]=\mu\in[0,1]$ for all $i\in\{1,\dots,k\}$. Then there is a one-to-one correspondence between joint distributions over X and probability assignments $\mathbb{P}(X_1=\dots=X_k=0)$. When $\mu<1/2$, all such assignments lie in the range*

$$(1-\mu)^k\left(1-\left(\frac{\mu}{1-\mu}\right)^{k_{even}}\right)\leq\mathbb{P}(X_1=\dots=X_k=0)\leq(1-\mu)^k\left(1+\left(\frac{\mu}{1-\mu}\right)^{k_{odd}}\right)\quad(5)$$

Here, k_{odd} is the largest odd integer $\leq k$, and k_{even} the largest even integer $\leq k$. Moreover, when $\mu\geq1/2$, all such assignments lie in the range

$$0\leq\mathbb{P}(X_1=\dots=X_k=0)\leq(1-\mu)^{k-1}\quad(6)$$

The constrapositive of the above theorem implies that, when the gaps are not exponentially small in k, there must be differences in the lower order statistics. If there are substantial differences between lower order statistics, then we we might be able to design optimisitc algorithms which can use lower order information to circumvent the need for exhaustive search.

Example 2.1 *Consider the case where $k=2$, and fix an index $i^*\in[n]$. Suppose that $\mu^*:=\mathbb{E}[X_{i^*}]$, and for $i\in[n]-\{i^*\}$, $\mu_i:=\mathbb{E}[X_i]<\frac{\mu^*}{2}$. Then, any set $S\subset[n]$ containing i^* must have $\mathbb{E}[\max_{i\in S}X_1]\geq\mathbb{E}[X_{i^*}]=\mu^*$, whereas any subset S not containing i^* must have $\mathbb{E}[\max_{i\in S}X_1]\leq\mathbb{E}[\sum_{i\in S}X_i]<2\frac{\mu^*}{2}=\mu^*$. Thus,* regardless of the dependencies between the arms, *we know that S^* must contain i^*.*

Even if one did not know the index i^*, one could then use this fact to design an algorithm whose sample complexity for this instance is linear *in n by first trying to identify the arm i^* with the* *unsually high mean, and then searching over all sets of the form $S = \{i^*, j\}$ for $i \in [n] - \{i^*\}$ to find S^*. It is also conceivable that optimistic algorithms could adapt to this unknown structure.*

The above example points to a more general phenomenon: when X are drawn iid from a joint distribution over $\{0,1\}^n$, lower order cross moments of X enforce very strict contraints on higher order moments. These lower order statistics can be directly estimated in the semi-bandits setting, or in a natural relaxation of the bandit game when you are allowed to play $k' < k$ arms. With exceptionally gaps, one may also be able to estimate useful confidence intervals on lower-order statistics using strict bandit feedback which mandates pulling exactly k arms per query.

Ultimately, we might hope that to design an optimistic algorithm which avoids exhaustive search by estimating lower order statistics before moving on to higher order ones, ruling out large collections of subsets as it goes along. By the same token, we believe that a more general version of Theorem 2.2 - one which precisely characterizes the sizes of differences between lower order statistics for problem instances which have very large gaps - would lead to a sharper characterization of the difficulty of Best-of-K in benign problem instances.

3. Best of K with Independent Arms

While the dependent case is of considerable practical interest, the remainder of this paper investigates the best-of-k problem where ν is assumed to be a product distribution of n independent Bernoulli distributions. We show that even in this presumably much simpler setting, there remain highly nontrivial algorithm design challenges related to the information occlusion that occurs in the bandit and marked-bandit feedback settings. We present an algorithm and analysis which tries to mitigate information occlusion which we hope can inform strategies for favorable instances of dependent measures.

Under the independent Bernoulli assumption, each arm is associated with a mean $\mu_i \in [0,1)$ and the expected reward of playing any set $S \in \binom{[n]}{2}$ is equal to $1 - \prod_{i \in S}(1-\mu_i)$ and hence best subset of k arms is precisely the set of arms with the greatest k means μ_i.

3.1. Results

Without loss of generality, suppose the means are ordered $\mu_1 \geq \dots \mu_k > \mu_{k+1} \geq \dots \mu_n$. Assuming $\mu_k \neq \mu_{k+1}$ ensures that the set of top k means is unique, though our results could be easily extended to a PAC Learning setting with little effort. Define the gaps and variances via

$$\Delta_i := \begin{cases} \mu_i - \mu_{k+1} & \text{if } i \leq k \\ \mu_k - \mu_i & \text{if } i > k \end{cases} \quad \text{and} \quad V_i := \mu_i(1-\mu_i) \tag{7}$$

For $\tau > 0$, introduce the transformation

$$\mathcal{T}_{n,\delta}(\tau) := \tau \log\left(\frac{16n\log_2 e}{\delta}\log\left(\frac{8n\tau\log_2 e}{\delta}\right)\right) = \tilde{\Theta}\left(\tau\log\left(\frac{n}{\delta}\right)\right) \tag{8}$$

where $\tilde{\Theta}(\cdot)$ hides logarithmic factors of its argument. We present guarantees for the Stagewise Elimination of Algorithm 3 in our three feedback models of interest; the broad brush strokes of our analysis are addressed in Appendix B, and the details are fleshed in the Appendices C and B.2. Our first result holds for semi-bandits, which slightly improves upon the best known result for the k-batch setting (Jun et al., 2016) by adapting to unknown variances:

Theorem 3.1 (Semi Bandit) *With probability $1-\delta$, Algorithm 3 with semi-bandit feedback returns the arms with the top k means using no more than*

$$8\mathcal{T}_{n,\delta}(\tau_{\sigma(1)}) + \frac{4}{k}\sum_{i=k+1}^{n}\mathcal{T}_{n,\delta}(\tau_{\sigma(i)}) = \tilde{O}\left(\left(\tau_{\sigma(1)} + \frac{1}{k}\sum_{i=k+1}^{n}\tau_{\sigma(i)}\right)\log\left(\frac{n}{\delta}\right)\right) \tag{9}$$

queries where

$$\tau_i := \frac{56}{\Delta_i} + \frac{256}{\Delta_i^2}\begin{cases}\max\{V_i, \max_{j>k} V_j\} & i \le k\\ \max\{V_i, \max_{j\le k} V_j\} & i > k\end{cases} \tag{10}$$

and σ is a permutation so that $\tau_{\sigma(1)} \ge \tau_{\sigma(2)} \ge \dots \tau_{\sigma(n)}$.

The above result also holds in the more general setting where the rewards have arbitrary distributions bounded in $[0,1]$ almost surely (where V_i is just the variance of arm i.)

In the marked-bandit and bandit settings, our upper bounds incur a dependence on information-sharing terms H^M (marked) and H^B (bandit) which capture the extent to which the $\max$ operator occludes information about the rewards of arms in each query.

Theorem 3.2 (Marked Bandit) *Suppose we require each query to pull exactly k arms. Then Algorithm 3 with marked bandit feedback returns the arms with the top k means with probability at least $1-\delta$ using no more than*

$$16\mathcal{T}_{n,\delta}\left(\frac{\tau^M_{\sigma(1)}}{H^M}\right) + \frac{8}{k}\sum_{i=k+1}^{n}\mathcal{T}_{n,\delta}\left(\frac{\tau^M_{\sigma(i)}}{H^M}\right) = \tilde{O}\left(\frac{\log(n/\delta)}{H^M}\left(\tau^M_{\sigma(1)} + \frac{1}{k}\sum_{i=k+1}^{n}\tau^M_{\sigma(i)}\right)\right) \tag{11}$$

queries. Here, τ_i^M is given by

$$\tau_i^M := \frac{56}{\Delta_i} + \frac{256}{\Delta_i^2}\begin{cases}\mu_i & i \le k\\ \mu_k & i > k\end{cases} \tag{12}$$

σ is a permutation so that $\tau_{\sigma(1)} \ge \tau_{\sigma(2)} \ge \dots \tau_{\sigma(n)}$, and H^M is an "information sharing term" given by

$$H^M := \mathbb{E}_{X_1,\dots,X_{k-1}}\left[\frac{1}{1+\sum_{\ell\in[k-1]}\mathbb{I}(X_\ell = 1)}\right] \tag{13}$$

If we can pull fewer than k arms per round, then we can achieve

$$8\max_{i\in[k-1]} i\mathcal{T}(\tau^M_{\sigma(i)}) + \frac{8}{kH^M}\sum_{i=2}^{n}\mathcal{T}_{n,\delta}\left(\tau^M_{\sigma(i)}\right) = \tilde{O}\left(\left(\max_{i\in\{1,k-1\}} i\tau^M_{\sigma(1)} + \frac{1}{kH^M}\sum_{i=2}^{n}\tau^M_{\sigma(i)}\right)\log\left(\frac{n}{\delta}\right)\right) \tag{14}$$

We remark that as long as the means are at no more than $1-c$, $\tau_i \le \frac{1}{c}\tau_i^M$, and thus the two differ by a constant factor when the means are not too close to 1 (this difference comes from loosing $(1-\mu)$ term in a Bernoulli variance in the marked case). Furthermore, note that $H^M \ge \frac{1}{k}$. Hence, when we are allowed to pull fewer than k arms per round, Stagewise Elimination with marked-bandit feedback does no worse than a standard LUCB algorithms for stochastic best arm identification.

When the means μ_i are on the order of $1/k$, then $H^M = \Omega(1)$, and thus Stagewise Eliminations gives the same guarantees for marked bandits as for semi bandits. The reason is that, when the means are $O(1/k)$, we can expect each query S to have only a constant number of arms $\ell \in S$ for which $X_\ell = 1$, and so not much information is being lost by observing only one of them.

Finally, we note that our guarantees depend crucially on the fact that the marking is uniform. We conjecture that adversarial marking is as challenging as the bandit setting, whose guarantees are as follows:

Theorem 3.3 (Bandit) *Suppose we require each query to pull exactly k arms, $n \geq 7k/2$, and $\forall i : \mu_i < 1$. Then Algorithm 3 with bandit feedback returns the arms with the top k means with probability at least $1-\delta$ using no more than*

$$20\mathcal{T}_{n,\delta}\left(\frac{\tau^B_{\sigma(1)}}{H^B}\right) + \frac{5}{k}\sum_{i=k+1}^{n}\mathcal{T}_{n,\delta}\left(\frac{\tau^B_{\sigma(i)}}{H^B}\right) = \tilde{O}\left(\frac{\log(n/\delta)}{H^B}\left(\tau^B_{\sigma(1)} + \frac{1}{k}\sum_{i=k+1}^{n}\tau^B_{\sigma(i)}\right)\right) \tag{15}$$

queries where $H^B := \prod_{\ell\in[k-1]}(1-\mu_\ell)$ is an "information sharing term",

$$\tau_i^B \leq \frac{66}{\Delta_i} + \frac{2560}{\Delta_i^2}\begin{cases} 2(1-\mu_{k+1})\mu_i + (1-\mu_{k+1})^2(1-H^B) & i \leq k \\ 2(1-\mu_i)\mu_{k+1} + (1-\mu_i)^2(1-H^B) & i > k \end{cases}$$

and σ is a permutation so that $\tau^B_{\sigma(1)} \geq \tau^B_{\sigma(2)} \geq \dots \tau^B_{\sigma(n)}$.

The condition that $\mu_i < 1$ ensures identifiability (see Remark B.11). The condition $n \geq 7k/2$ is an artifact of using a Balancing Set B defined in Algorithm 4; without B, our algorithm succeeds for all $n \geq k$, albeit with slightly looser guarantees (see Remark B.9).

Remark 3.1 *Suppose the means are greater than $\alpha(k)/k$ where $\alpha(k) \geq C\log k$ and C is a constant; for example, think $\alpha(k) = k/2$. Then $H^B \leq (1-\frac{\alpha(k)}{k})^k = O(\exp(-\alpha(k))) \ll 1/k$. Hence, Successive Elimination requires on the order of $\frac{1}{k}\cdot\frac{1}{H^B} = \frac{\exp(\Omega(\alpha(k))}{k}$ more queries to identify the top k-arms than the classic stochastic MAB setting where you get to pull 1-arm at a time,* despite the seeming advantage that the bandit setting lets you pull k arms per query. *When $\alpha(k) \geq C\log k$, then $\frac{\exp(\Omega(\alpha(k)))}{k}$ is at least polynomially large in k, and when $\alpha = \Omega(k)$, is exponentially large in k (e.g, $\alpha(k) = k/2$).*

On the other hand, when the means are all on the order of α/k for $\alpha = O(1)$, then $H^B = \Omega(1)$, but the term $1-H^B$ is at least $\Omega(\alpha)$. For this case, our sample complexity looks like

$$\tilde{O}(\frac{\log(n/\delta)}{k}\sum_i \frac{\alpha/k+\alpha}{\Delta_i^2} + \frac{1}{\Delta_i}) = \tilde{O}(\log(n/\delta)\sum_i \frac{\alpha}{\Delta_i^2}) \tag{16}$$

which matches, but does not out-perform, the standard 1-arm-per-query MAB guarantees, with variance adaptation (e.g., Theorem 3.1 with $k = 1$, note that α captures the variance). Hence, when the means are all roughly on the same order, it's never worse to pull 1 arm at a time and observe its reward, than to pull k and observe their max. Once the means vary wildly, however, this is certainly not true; we direct the reader to Remark B.12 for further discussion.

3.2. Algorithm

At each stage $t \in \{0, 1, 2, \dots\}$, our algorithm maintains an accept set $A_t \subset [n]$ of arms which we are confident lie in the top k, a reject set $R_t \subset [n]$ of arms which we are confident lie in the bottom $n - k$, and an undecided set U_t containing arms for which we have not yet rendered a decision. The main obstacle is to obtain estimates of the relative performance of $i \in U_t$, since the bandit and marked bandit observation models occlude isolated information about any one given arm in a pull. The key observation is that, if we sample $S \sim \mathrm{Unif}[U_t, k]$, then for $i, j \in U_t$, the following differences have the same sign as $\mu_i - \mu_j$ (stated formally in Lemma B.2):

$$\begin{aligned} &\mathbb{E}[\max_{\ell \in S} X_\ell \big| i \in S] - \mathbb{E}[\max_{\ell \in S} X_\ell \big| j \in S] \quad \text{(bandits)} \qquad \text{and} \\ &\mathbb{P}(\text{ observe } X_i = 1 \big| i \in S) - \mathbb{P}(\text{ observe } X_j = 1 \big| j \in S) \text{ (marked/semi-bandits)} \end{aligned} \tag{17}$$

This motivates a sampling strategy where we partition U_t uniformly at random into subsets $S_1, S_2, \dots, S_p$ of size k, and query each S_q, $q \in \{1, \dots, p\}$. We record all arms $\ell \in S_q$ for which $X_\ell = 1$ in the semi/marked-bandit settings (Algorithm 1, Line 3), and, in the bandit setting, mark down all arms in S_q if we observe $\max_{\ell \in S_q} X_\ell = 1$ - i.e, we observe a reward of 1 (Algorithm 1, Line 4). This recording procedure is summarized in Algorithm 1:

Algorithm 1: PlayAndRecord(S, S^+, Y)

```
1 Input $S, S^+ \subset [n]$, $Y \in \mathbb{R}^n$
2 Play $S \cup S^+$
3     Semi/Marked Bandit Setting: $Y_\ell \leftarrow 1$ for all $\ell \in S$ for which we observe $X_\ell = 1$
4     Bandit Bandit Setting: If $A$ returns a reward of 1, $Y_\ell \leftarrow 1$ for all $\ell \in S$
5 Return $Y$
```

Note that PlayAndRecord$[S, S^+, Y]$ plays a the union of S and S^+, but only records entries of Y whose indices lie in S. UniformPlay (Algorithm 2) outlines our sampling strategy. Each call to UniformPlay$[U, A, R, k^{(1)}]$ returns a vector $Y \in \mathbb{R}^n$, supported on entries $i \in U$, for which

$$\mathbb{E}[Y_i] = \begin{cases} \mathbb{P}_{S,S^+}(\text{ observe } X_i = 1 \big| i \in S \cup S^+) & \text{marked/semi-bandit} \\ \mathbb{P}_{S,S^+}(\max_{\ell \in S \cup S^+} X_\ell = 1 \big| i \in S \cup S^+) & \text{bandit} \end{cases} \tag{18}$$

where $S \sim \mathrm{Unif}[U, k^{(1)}]$ and S^+ is empty unless $|U_t| < k$ or we are allowed to pull fewer than k arms per query in which case elements of S^+ are drawn from $A \cup R$ as outlined in Algorithm 2, Line 3 otherwise.

There are a couple nuances worth mentioning. When $|U| < k$, we cannot sample k arms from the undecided set U; hence UniformPlay pulls only $k^{(1)}$ from U per query. If we are forced to pull exactly k arms per query, UniformPlay adds in a "Top-Off" set of an additional $k - k^{(1)}$ arms, from R and A (Lines 3-9). Furthermore, observe that lines 13-15 in UniformPlay carefully handle divisibility issues so as to not "double mark" entries $i \in U$, thus ensuring the correctness of Equation 18. Finally, note that each call to UniformPlay makes exactly $\lceil |U|/k^{(1)} \rceil$ queries.

We deploy the passive sampling in UniformPlay in a stagewise successive elimination procedure formalized in Algorithm 3. At each round $t = \{1, 2, \dots\}$, use a doubling sample size to $T(t) := 2^t$, and set the $k^{(1)}$ parameter for UniformPlay to be $\min\{|U_t|, k\}$ (line 3). Next, we construct the sets

Algorithm 2: UniformPlay$(U, A, R, k^{(1)})$

```
Inputs: $U$, $A$, $R$, sample size $k^{(1)}$
Uniformly at random, partition $U$ into $p := \lfloor |U|/k^{(1)} \rfloor$ sets $S^{(1)}, \dots, S^{(p)}$ of size $k^{(1)}$ and place
    remainders in $S^{(0)}$ // thus $S^{(1)}, \dots, S^{(p)} \sim \mathrm{Unif}[U, k^{(1)}]$[2]
If Require $k$ Arms per Pull and $k^{(1)} < k$ // Construct Top-Off Set $S^+$
    $k^{(2)} \leftarrow k - k^{(1)}$ // $k^{(2)} = |S^+|$
    $S^+ \leftarrow \mathrm{Unif}[R, \min\{|R|, k^{(2)}\}]$ //sample as many items from reject as possible
    If $|R| < k^{(2)}$: // sample remaining items from accept
        $S^+ \leftarrow R \cup \mathrm{Unif}[A, k^{(2)} - |R|]$
Else // Top-Off set unnecessary
    $S^+ \leftarrow \emptyset$, $k^{(2)} \leftarrow 0$
Initalize rewards vector $Y \leftarrow \mathbf{0} \in \mathbb{R}^n$
For $q = 1, \dots, p$
    $Y \leftarrow$ PlayAndRecord$[S^{(q)}, S^+, Y]$ // only mark $S^{(q)}$
If $|S^{(0)}| > 0$ // if remainder
    Draw $S^{(0,+)} \sim \mathrm{Unif}[U - S^{(0)}, k^{(1)} - |S^{(0)}|]$ // thus $S^{(0)} \cup S^{(0,+)} \sim \mathrm{Unif}[U, k^{(1)}]$
    $Y \leftarrow$ PlayAndRecord$[S^{(0)}, S^{(0,+)} \cup S^+, Y]$ // only mark $S^{(0)}$ to avoid duplicate marking
Return $Y$
```

(U'_t, R'_t) from which UniformPlay samples: in the marked and semi-bandit setting, these are just (U_t, A_t, R_t) (Line 4), while in the bandit setting, they are obtained by from Algorithm 4 which transfers a couple low mean arms from R_t into U'_t (Line 5). This procedure ameliorates the effect of information occlusion for the bandit case.

Line 7 through 9 average together $T(t) := 2^t$ independent, and identically distributed samples from UniformPlay$[U'_t, R'_t, A_t, k^{(1)}]$ to produce unbiased estimates $\hat{\mu}_{i,t}$ of the quantity $\mathbb{E}[Y_i]$ defined in Equation 18. $\hat{\mu}_{i,t}$ are Binomial, so we apply an empirical Bernstein's inequality from Maurer and Pontil (2009) to build tight $1-\delta$ confidence intervals

$$\widehat{C}_{i,t} := \sqrt{\frac{2\hat{V}\log(8nt^2/\delta)}{T(t)}} + \frac{8\log(8nt^2/\delta)}{3(T(t)-1)} \quad \text{where} \quad \hat{V}_{i,t} := \frac{T(t)\hat{\mu}_{i,t}(1-\hat{\mu}_{i,t})}{T(t)-1} \tag{19}$$

Note that $\hat{V}_{i,t}$ coincide with the canonical definition of *sample* variance. The variance-dependence of our confidence intervals is crucial; see Remarks B.7 and B.8 for more details. For any $\ell \le |U_t|$ let

$$\max_{j \in U_t}^{\ell} = \ell\text{-th largest element} \tag{20}$$

As mentioned above, Lemma B.2 ensures $\mathbb{E}[\hat{\mu}_{i,t}] > \mathbb{E}[\hat{\mu}_{j,t}]$ if and only if $\mu_i > \mu_j$. Thus, accepting an arm for $\hat{\mu}_{i,t}$ is in the top k.

The Balance Procedure is described in Algorithm 4, and ensures that U'_t contains sufficiently many arms that don't have very high (top $k+1$) means. The motivation for the procedure is somewhat subtle, and we defer its discussion to the analysis in Appendix B.3.3, following Remark B.8:

Algorithm 3: Stagewise Elimination(S, k, δ)

```
**Input** $S_1 = [n]$, Batch Size $k$, $t = 0$
**While** $|A_t| < k$ // fewer than k arms accepted
    Sample Size $T(t) \leftarrow 2^t$, Rewards Vector $Y^{(t)} \leftarrow \mathbf{0} \in \mathbb{R}^n$, $k^{(1)} \leftarrow \min\{|U_t|, k\}$
    $(U'_t, R'_t) \leftarrow (U_t, R_t)$     // Sampling Sets for UniformPlay, identical to $U_t$ and $R_t$ in
        marked/semi bandits
    **If** Bandit Setting     // Add low mean arms from $R_t$ to $U_t$
        $(U'_t, R'_t) \leftarrow$ Balance$(U_t, R_t)$
    **For** $s = 1, 2, \dots, T(t)$
        $Y^{(t)} \leftarrow Y^{(t)} +$ UniformPlay$[U'_t, R'_t, A_t, k^{(1)}]$     // get fresh samples
    $\hat{\mu}_{i,t} \leftarrow \frac{1}{T(t)} \cdot Y^{(t)}$     // normalize
    $k_t \leftarrow k - |A_t|$
    $A_{t+1} \leftarrow A_t \cup \{i \in U_t : \hat{\mu}_{i,t} - \hat{C}_{i,t} > \max^{k_t+1}_{j \in U_t} \hat{\mu}_{j,t} + \hat{C}_{j,t}\}$     // Equation 19
    $R_{t+1} \leftarrow R_t \cup \{i \in U_t : \hat{\mu}_{i,t} + \hat{C}_{i,t} < \max^{k_t}_{j \in U_t} \hat{\mu}_{j,t} - \hat{C}_{j,t}\}$
    **If** $|R_t| = n - k$ //$n - k$ arms rejected
        $A_{t+1} \leftarrow A_{t+1} \cup U_t$
    $U_{t+1} \leftarrow U_t - \{A_{t+1} \cup R_{t+1}\}$
    $t \leftarrow t + 1$
```

Algorithm 4: Balance(U, R)

```
**Input** $U, R$
$B \sim \text{Unif}[R, \max\{0, \lceil \frac{5k^{(1)}}{2} - |U| - \frac{1}{2} \rceil\}]$ //Balancing Set
$U' \leftarrow U \cup B$ , $R' \leftarrow R - B$ // Transfer $B$ from $R$ to $U$
**Return** $(U', R')$
```

4. Lower bound for Independent Arms

In the bandit and marked-bandit settings, the upper bounds of the previous section depended on "information sharing" terms that quantified the degree to which other arms occlude the performance of a particular arm in a played set. Indeed, great care was taken in the design of the algorithm to minimize impact of this information sharing. The next theorem shows that the upper bounds of the previous section for bandit and semi-bandit feedback are nearly tight up to a similarly defined information sharing term.

Theorem 4.1 (Independent) *Fix $1 \leq p \leq k \leq n$. Let $\nu = \prod_{i=1}^n \nu_i$ be a product distribution where each ν_i is an independent Bernoulli with mean μ_i. Assume $\mu_1 \geq \dots \geq \mu_k > \mu_{k+1} \geq \dots \geq \mu_n$ (the ordering is unknown to any algorithm). At each time the algorithm queries a set $S' \in \binom{[n]}{p}$ and observes $\mathbb{E}[\max_{i \in S'} X_i]$. Then any algorithm that identifies the top k arms with probability at least $1 - \delta$ requires at least*

$$\Big(\max_{j=1,\dots,n} \tau_j + \frac{1}{p} \sum_{j=1}^{n} \tau_j\Big) \log(\tfrac{1}{2\delta})$$

observations where

$$(i) \quad \tau_j = \begin{cases} \frac{(1-\mu_j-\Delta_j)}{\Delta_j^2}\frac{1-h_j+\mu_j h_j}{h_j} & \textit{if } j > k \\ \frac{(1-\mu_j)}{\Delta_j^2}\frac{1-h_j+(\mu_j-\Delta_j)h_j}{h_j} & \textit{if } j \le k \end{cases} \quad \textit{for bandit observations, and}$$

$$(ii) \quad \tau_j = \begin{cases} \frac{(1-\mu_j-\Delta_j)\mu_j}{\Delta_j^2} & \textit{if } j > k \\ \frac{(1-\mu_j)(\mu_j-\Delta_j)}{\Delta_j^2} & \textit{if } j \le k \end{cases} \quad \textit{for semi-bandit observations.}$$

where $h_j = \max_{S\in\binom{[n]-j}{p-1}} \prod_{i\in a\setminus j}(1-\mu_i)$.

Our lower bounds apply to our upper bounds when $p = k$. In the bandit setting, considering $p < k$ reveals a trade-off between the information sharing term, which decreases with larger p, with the benefit of a $\frac{1}{p}$ factor gained from querying p arms at once. One can construct different instances that are optimized by the entire range of $1 \le p \le k$. Future research may consider varying the subset size in an adaptive setting to optimize this trade off.

The information sharing terms defined in the upper and lower bounds correspond to the most pessimistic and optimistic scenarios, respectively, and result from applying coarse bounds in exchange for simpler proofs. Thus, our algorithm may fare considerably better in practice than is predicted by the upper bounds. Moreover, when $\max_i \mu_i - \min_i \mu_i$ is dominated by $\min_i \mu_i$ our upper and lower bounds differ by constant factors.

Finally, we note that our upper and lower bounds for independent measures are tailored to Bernoulli payoffs, where the best k-subset corresponds to the top k means. However, for general product distributions ν on $[0,1]^n$, this is no longer true (see Remark B.1). This leaves open the question: how difficult is Best-of-K for general, independent bounded product measures? And, in the marked feedback setting (where one receives an index of the best element in the query), is this problem even well-posed?

Acknowledgements

We thank Elad Hazan for illuminating discussions regarding the computational complexity of the Best-of-K problem, and for pointing us to resources adressing online submodularity and approximate regret. Max Simchowitz is supported by an NSF GRFP award. Ben Recht and Kevin Jamieson are generously supported by ONR awards , N00014-15-1-2620, and N00014-13-1-0129. BR is additionally generously supported by ONR award N00014-14-1-0024 and NSF awards CCF-1148243 and CCF-1217058. This research is supported in part by gifts from Amazon Web Services, Google, IBM, SAP, The Thomas and Stacey Siebel Foundation, Adatao, Adobe, Apple Inc., Blue Goji, Bosch, Cisco, Cray, Cloudera, Ericsson, Facebook, Fujitsu, Guavus, HP, Huawei, Intel, Microsoft, Pivotal, Samsung, Schlumberger, Splunk, State Farm, Virdata and VMware.

References

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

Aditya Gopalan, Shie Mannor, and Yishay Mansour. Thompson sampling for complex online problems. In *Proceedings of The 31st International Conference on Machine Learning*, pages 100–108, 2014.

Linhui Hao, Qiuling He, Zhishi Wang, Mark Craven, Michael A Newton, and Paul Ahlquist. Limited agreement of independent rnai screens for virus-required host genes owes more to false-negative than false-positive factors. *PLoS Comput Biol*, 9(9):e1003235, 2013.

Katja Hofmann, Shimon Whiteson, and Maarten de Rijke. A probabilistic method for inferring preferences from clicks. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, CIKM '11, pages 249–258, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0717-8. doi: 10.1145/2063576.2063618. URL http://doi.acm.org/10.1145/2063576.2063618.

Mark M Huycke, Daniel F Sahm, and Michael S Gilmore. Multiple-drug resistant enterococci: the nature of the problem and an agenda for the future. *Emerging infectious diseases*, 4(2):239, 1998.

Rishabh K Iyer and Jeff A Bilmes. Submodular optimization with submodular cover and submodular knapsack constraints. In *Advances in Neural Information Processing Systems*, pages 2436–2444, 2013.

Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sébastien Bubeck. lil'ucb: An optimal exploration algorithm for multi-armed bandits. In *Proceedings of The 27th Conference on Learning Theory*, pages 423–439, 2014.

Kwang-Sung Jun, Kevin Jamieson, Rob Nowak, and Xiaojin Zhu. Top arm identification in multi-armed bandits with batch arm pulls. In *The 19th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2016.

Emilie Kaufmann and Shivaram Kalyanakrishnan. Information complexity in bandit subset selection. In *Conference on Learning Theory*, pages 228–251, 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, Csaba Szepesvari, Zheng Wen, and Azin Ashkan. Cascading bandits: Learning to rank in the cascade model. In *Proceedings of the 32nd International Conference on Machine Learning (ICML-15)*, pages 767–776, 2015.

Andreas Maurer and Massimiliano Pontil. Empirical bernstein bounds and sample variance penalization. *arXiv preprint arXiv:0907.3740*, 2009.

Filip Radlinski, Robert Kleinberg, and Thorsten Joachims. Learning diverse rankings with multi-armed bandits. In *Proceedings of the 25th international conference on Machine learning*, pages 784–791. ACM, 2008.

Karthik Raman, Pannaga Shivaswamy, and Thorsten Joachims. Online learning to diversify from implicit feedback. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '12, pages 705–713, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1462-6. doi: 10.1145/2339530.2339642. URL http://doi.acm.org/10.1145/2339530.2339642.

Matthew Streeter and Daniel Golovin. An online algorithm for maximizing submodular functions. In *Advances in Neural Information Processing Systems*, pages 1577–1584, 2009.

Vijay V Vazirani. *Approximation algorithms*. Springer Science & Business Media, 2013.

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. Reduction from Max-K-Coverage to Best-of-K

As in the main text, let $X = (X_1, \ldots, X_n) \in \{0,1\}^n$ be a binary reward vector, let $\mathcal{S}^* = \{\arg\max_{S \in \binom{[n]}{k}} \mathbb{E}[\max_{i \in S} X_i]\}$ be set of all optimal k-subsets of $[n]$ (we allow for non-uniqueness), and define the gap $\Delta := \mathbb{E}_\nu \left[\max_{i \in S^*} X_i\right] - \max_{S \in \mathcal{S} \setminus \mathcal{S}^*} \mathbb{E}_\nu \left[\max_{i \in S} X_i\right]$ as the minimum gap between the rewards of an optimal and sub-optimal k-set. We say $\tilde{S}$ is $\alpha-$optimal for $\alpha \leq 1$ if $\mathbb{E}[\max_{i \in \tilde{S}} X_i] \geq \alpha \mathbb{E}[\max_{i \in S^*} X_i]$, where $S^* \in \mathcal{S}^*$. We formally introduce the classical Max-K-Coverage problem:

Definition 2 (Max-K-Coverage$(m, k, \mathcal{V})$) *A Max-K-Coverage instance is a tuple $(m, k, \mathcal{V})$, where $\mathcal{V}$ is a collection of subsets $V_1, \ldots, V_n \in 2^{[m]}$. We say $S \subset \mathcal{V}$ is a solution to Max-K-Coverage if $|S| = k$ and S maximizes $|\bigcup_{V_i \in S} V_i|$. Given $\alpha \leq 1$, we say S is an α approximation if $|\bigcup_{V_i \in S} V_i| \geq \alpha \max_{S' \in \binom{\mathcal{V}}{k}} |\bigcup_{V_i \in S'} V_i|$.*

It is well known that Max-K-Coverage in NP-Hard, and cannot be approximated to within $\alpha = 1 - \frac{1}{e} + o(1)$ under standard hardness assumptions Vazirani (2013). The following theorem gives a reduction from Best of K Indentification (under any feedback model) to Max-K-Coverage:

Theorem A.1 *Fix $\alpha \leq 1$, and let $\mathcal{A}$ be an algorithm which indentifies an α-optimal k-subset of n arms probability in time polynomial in n, k, and $1/\Delta$, with probability at least η (under any feedback model). Then there is a polynomial time α-approximation algorithm for Max-K-Coverage$[m, k, \mathcal{V}]$which succeeds with probability at least η. When $\alpha = 1$, this implies a polynomial time algorithm for exact* $\mathrm{Max-K-Coverage}[m, k, \mathcal{V}]$.

Proof Consider an instance of $\mathrm{Max-K-Coverage}[m, k, \mathcal{V}]$, and set $n = |\mathcal{V}|$. We construct a reward vector $X \in \{0,1\}^n$ as follows: At each time t, draw ω uniformly from $[m]$, and set $X_i := \mathbb{I}(\omega \in V_i)$. We run $\mathcal{A}$ on the reward vector X, and it returns a candidate set $\widehat{S} \in \binom{n}{[k]}$ which is α-optimal with probability η. We then return the sets $V_i \in \mathcal{V}$ whose indicies lie in $\widehat{S}$. We show this reduction completes in polynomial time, and if $\widehat{S}$ is α-optimal, then $\{V_i\}_{i \in \widehat{S}}$ is an α-approximation for the Max-K-Coverage instance.

Correctness: Since ω is uniform from $[m]$, the reward of a subset $S \subset [n]$ is $\mathbb{E}[\max_{i \in S} \mathbb{I}(\omega \in V_i)] = \mathbb{E}[\mathbb{I}(\omega \in \bigcup_{i \in S} V_i)] = \frac{|\bigcup_{i \in S} V_i|}{m} \propto |\bigcup_{i \in S} V_i|$. Hence, an α-optimal subset S corresponds to an α-approximation to the Max-K-Coverage instance.

Runtime: Let $R(n, k, \Delta) = O(\mathrm{poly}(n, k, 1/\Delta))$ denote an upper bound runtime of $\mathcal{A}$, and let $T(n, k, \Delta) = O(\mathrm{poly}(n, k, 1/\Delta))$ be an upper bound on the number of queries required by Algorithm $\mathcal{A}$ to return to α-optimal k-subset. Note that sampling ω takes $O(m)$ time, and setting each $X_i(\omega)$ completes in time $O(mn)$. Moreover, the expected reward of any $S \in \binom{[n]}{k}$ lies in $\{0, \frac{1}{m}, \ldots, 1\}$, so $\Delta \leq 1/m$. Thus, the runtime of our reduction is $R(n, k, \Delta) + O(mn) \cdot T(n, k, \Delta)) \leq R(n, k, 1/m) + O(mn) \cdot T(n, k, 1/m)) = O(\mathrm{poly}(n, k, m))$.

■

Remark A.1 *Note that the parameter m in the Max-K-Coverage instance shows up in the gap Δ in the runtime of the Max-K-Coverage instance. Our lower bound construction holds in the regime where $\Delta = \exp(-O(k))$, which morally corresponds to Max-K-Coverage instances in the regime where $m = \exp(\Omega(k))$.*

Appendix B. High Level Analysis for Independent Upper Bound

B.1. Preliminaries

At each stage t of Algorithm 3, there are three sources of randomness we need to account for. First, there is the randomness over all events that occurred before we start sampling from UniformPlay: this randomness determines the undecided, accept, and rejected sets U_t, A_t, and R_t, as well as their modifications U'_t, and R'_t. In what follows, we will define a so-called "Data-Tuple" $\mathcal{D}_t := (U_t, A_t, R_t, U'_t, R'_t)$ which represents the state of our algorithm, in round t, before collecting samples.

The second source of randomness comes from the uniform partitioning of U'_t into the sets $S^{(0)}, S^{(1)}, \dots, S^{(q)}$ (Algorithm 2, Line 2) and the draw of the Top-Off set S^+ (Lines 3-3), at each call to UniformPlay. Finally, there is randomness over the values that the arms $X_\ell \in S \cup S^+$ take, when pulled in PlayAndMark. To clear up any confusion, we define the probability and expectation operators

$$\mathbb{P}_{\cdot|t}\left[\cdot\right] := \mathbb{P}[\cdot \big| \mathcal{D}_t] \quad \text{and} \quad \mathbb{E}_{\cdot|t}\left[\cdot\right] := \mathbb{E}[\cdot \big| D_t] \tag{21}$$

$\mathbb{P}_{\cdot|t}\left[\cdot\right]$ and $\mathbb{E}_{\cdot|t}\left[\cdot\right]$ condition on the data in $\mathcal{D}_t$, and take expectations over the randomness in the partitioning of U'_t, draw of S^+, and the values of each arm pulled.

Treating $\mathcal{D}_t$ as fixed, we will let S denote a set with same distribution of one of the randomly partitioned subsets $S^{(1)}, \dots, S^{(q)}$ of U'_t in UniformPlay, S^+ to denote a set with the distribution of the Top-Off set chosen in UniformPlay. Recall that the purpose of S^+ is simply to ensure that we pull exact k arms per query. If either $k^{(1)} = k$, or we do not enforce exactly k-pulls per round, then $S^+ = \emptyset$. We remark that the distributions of S and S^+ are explicitly

$$S \sim \text{Unif}[U'_t, k^{(1)}] \quad \text{and} \quad S^+ \sim \begin{cases} \text{Unif}[R'_t, k^{(2)}] & |R'_t| \geq k^{(2)} \\ R'_t \cup \text{Unif}[A_t, k^{(2)} - |R'_t|] & |R'_t| < k^{(2)} \end{cases} \tag{22}$$

Note that $\mathcal{D}_t$ exactly determines $k^{(1)} := |S|$, which we recall is defined at each round as $\min\{|U_t|, k\}$ (Algorithm 2, Line 3). It also determines the size of the Top-Off set $k^{(2)}$ (Algorithm 2, Lines 4 and 9). We further note that the play $S^{(0)} \cup S^{(0,+)}$ (Algorithm 2, Lines 13-15) is also uniformly drawn as $\text{Unif}[U'_t, k^{(1)}]$, and hence has the same distribution of S. We also remark that

Claim B.1 *The sets S and S^+ are independent and disjoint under $\mathbb{P}_t$. In the marked and semi-bandit setting, there are always enough accepted/rejected arms in $|A_t \cup R_t|$ to ensure that we can fill S^+ with $k^{(2)}$ arms. In the bandit setting, there are sufficiently many accepted/rejected arms in $|A_t \cup R'_t|$ as long as $n \geq 7k/2$.*

This condition $n \geq 7k/2$ is an artifact of the balancing set in our algorithm, and is discussed in more detail in Section B.3.3.

B.2. Guarantees for General Feedback Models

The core of our analysis is common to the three feedback models. To handle bandits and marked/semi bandits settings simultaneous, we define a win function $\mathcal{W} : [n] \times 2^{[n]} \to \{0, 1\}$ which reflects the recording strategy in PlayAndRecord

$$\mathcal{W}(i,S') = \begin{cases} 1 & \text{if bandit setting and } \max_{\ell\in S'} X_\ell = 1 \\ 1 & \text{if marked/semi-bandit setting and observe } X_i = 1 \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

That is, PlayAndRecord$[S, S^+, Y]$ sets $Y_i = 1\ \forall i \in S : \mathcal{W}(i, S\cup S^+) = 1$. The following lemma characterizes the distribution of our estimations $\hat{\mu}_{i,t}$

Lemma B.2

$$\hat{\mu}_{i,t} \sim \frac{1}{T(t)}\text{Binomial}(\bar{\mu}_{i,t}, T(t)) \quad \text{and} \quad \mathbb{E}[\hat{V}_{i,t}] = V_{i,t} \tag{24}$$

where

$$\bar{\mu}_{i,t} = \mathbb{E}_t[\mathcal{W}(i, S\cup S^+)\big| i \in S] \quad \textit{and} \quad V_{i,t} := \bar{\mu}_{i,t}(1-\bar{\mu}_{i,t}) \tag{25}$$

Moreover, in semi-bandit and marked bandit settings, and if $\mu_1 \le 1$ in the bandit setting, then given $i, j \in S_t$, $\bar{\mu}_{i,t} > \bar{\mu}_{j,t}$ if and only if $\mu_i > \mu_j$.

Remark B.1 *In the partial feedback models, the property that $\bar{\mu}_{i,t} > \bar{\mu}_{j,t}$ if and only if $\mu_i > \mu_j$ is quite particular to independent Bernoulli observations. The case of dependent Bernoullis measures is adressed by Theorem 2.1. For independent, non-Bernoulli distributions, consider the setting where $n = 3$, $k = 2$, and let X_1, X_2, X_3 be independent, where $X_1 \overset{d}{=} X_2 \overset{a.s.}{=} 2/3$, and $X_3 \sim$ Bernoulli$(1/2)$. Then, $\mathbb{E}[\max(X_1, X_2)] = 2/3$, while $\mathbb{E}[\max(X_1, X_3)] = \mathbb{E}[\max(X_2, X_3)] = \frac{1}{2} + \frac{1}{3} = 5/6$. Hence, if $S \sim \text{Unif}[\{1,2,3\},2]$, $\mathbb{E}[\max_{\ell\in S} X_\ell \big| 3 \in S] > \mathbb{E}[\max_{\ell\in S} X_\ell \big| 2 \in S] = \mathbb{E}[\max_{\ell\in S} X_\ell \big| 1 \in S]$.*

The last preliminary is to define the stage-wise comparator arms $c_{i,t}$ for $i \in U_t$:

$$c_{i,t} := \begin{cases} \min\{j \in U_t : j > k\} & i \le k \\ \max\{j \in U_t : j \le k\} & i > k \end{cases} \tag{26}$$

Intuitively, the comparator arm is the arm we are mostly to falsely accept instead of i when $i \le k$, and falsely reject instead of i when $i > k$.

Remark B.2 *As long as the accept set A_t only consists of arms $i \le k$, and R_t only consists of arms $i > k$, $c_{i,t}$ is guaranteed to exists. Indeed, fix $i \in U_t$, and suppose $c_{i,t}$ does not exist. If $i \le k$, then this would mean that U_t doesn't contain any rejected arms, but since A_t only contains accepted arms, all rejected arms are in R'_t, in which case Algorithm 3 will have already terminated (Line 14). A similar contradiction arises when $i > k$.*

Finally, we define the stagewise effective gaps

$$\Delta_{i,t} := |\bar{\mu}_{i,t} - \bar{\mu}_{c_{i,t},t}| \tag{27}$$

Observe that, conditioned on the data in $\mathcal{D}_t$, the means $\bar{\mu}_{i,t}$, gaps $\Delta_{i,t}$ and the variances $V_{i,t}$ are all deterministic quantities. We now have the following guarantee for Algorithm 3, which holds for the bandit, marked-bandit, and semi-bandit regimes:

Lemma B.3 (General Performance Guarantee for Successive Elimination) *In the bandit, marked-bandit, and semi-bandit settings, the following is true for all $t \in \{0, 1, \dots\}$ simultaneously with probability $1 - \delta$: Algorithm 3 never rejects i if $i \le k$ and never accepts i if $i > k$. Furthermore, if for a stage t and arm $i \in U_t$, the number of sample $T(t) := 2^t$ satisfies*

$$T(t) \ge T_{n,\delta}(\tau_{i,t}) := \tau_{i,t} \log\left(\frac{24n}{\delta}\log\left(\frac{12n\tau_{i,t}}{\delta}\right)\right) \tag{28}$$

where

$$\tau_{i,t} := \frac{56}{\Delta_{i,t}} + \frac{256\max\{V_{i,t}, V_{i,c_{i,t}}\}}{\Delta_{i,t}^2} \tag{29}$$

then by the end of stage t, i is accepted if $i \le k$ and rejected if $i > k + 1$.

Remark B.3 *The above theorem holds quite generally, and its proof abstracts out most details of best-of-k observation model. In fact, it only requires that (1) for each $i \in U_t$, $\hat{\mu}_{i,t} \sim \frac{1}{T(t)}$Binomial$(\bar{\mu}_{i,t}, T(t))$ and (2) $\bar{\mu}_{i,t} > \bar{\mu}_{j,t} \iff \mu_i > \mu_j$. In our three settings of interest, both conditions are ensured by Lemma B.2. It also holds in the semi-bandit setting when the arms have arbitrary distributions, as long as the rewards are bounded in $[0, 1]$.*

The final lemma captures the fact that each call to UniformPlay often makes fewer than $|U_t|$ queries to pull each arm in U_t:

Lemma B.4 *Suppose that, at round t, each call of uniformly play queries no more than $\alpha|U_t|/k$ times when $|U_t| \ge k$, and no more than α samples when $|U_t| \le k$. Let t_i^* be the first stage at which $i \notin U_t$. Then, Algorithm 3 makes no more than the following number of queries*

$$4\alpha T(t^*_{\sigma(1)}) + \frac{2\alpha}{k}\sum_{i=k+1}^{n} T(t^*_{\sigma(i)}) \tag{30}$$

where σ is permutation chosen so that $t^*_{\sigma(1)} \ge t^*_{\sigma(2)} \ge \dots \ge t^*_{\sigma(n)}$, and $T(t) = 2^t$, as above.

Remark B.4 *In the marked-bandit and semi-bandit settings, it is straightforward to verify that one can take $\alpha = 2$ in the above lemma. This is because Algorithm 3 always calls UniformPlay (Line 8) on $U_t' = U_t$ (Algorithm 4). Then, UniformPlay (Algorithm 8) partitions U_t into at most $\lceil |U_t|/k^{(1)} \rceil$ queries S_q^+. Recall that $k^{(1)} = \min\{|U_t|, k\}$ (Algorithm 3, Line 3) so that $\lceil |U_t|/k^{(1)} \rceil \le \lceil |U_t|/k \rceil \le 2|U_t|$ when $|U_t| \ge k$, while $|U_t|/k^{(1)} = 1 \le 2|U_t|$ once $|U_t| < k$. Controlling bound on α is slightly more involved in the bandit setting, and is addressed in Claim B.6.*

B.3. Specializing the Results

In the following sections, we again condition on the data $\mathcal{D}_t := (U_t, A_t, R_t, U_t', R_t')$. We proceed to compute the stage-wise means $\bar{\mu}_{i,t}$, variances $V_{i,t}$, and time parameters $\tau_{i,t}$ in Lemma B.3. As a warm up, let's handle the semi-bandit case:

B.3.1. SEMI-BANDITS

In Semi-Bandits, $\bar{\mu}_{i,t} = \mu_i$, and so

$$\tau_{i,t} = \tau_i = \frac{256 \max\{V_i, V_{c_{i,t}}\}}{\Delta_i^2} + \frac{56}{\Delta_i} \tag{31}$$

as in Theorem 3.1. Noting that $c_{i,t} > k$ if $i \le k$, while $c_{i,t} \le k$ if $i > k$, we can bound

$$V_{c_{i,t}} \le \begin{cases} \max_{j>k} V_j & i \le k \\ \max_{j \le k} V_j & i > k \end{cases} \tag{32}$$

Plugging the above display into Equation 31, we see that $\tau_{i,t} \le \tau_i$, as defined in Theorem 3.1. Combining this observation with Lemmas B.3 and B.4 and Remark B.4 concludes the proof of Theorem 3.1. Note that we pick up an extra factor of two, since we might end up collected at most $2\mathcal{T}_{n,\delta}(\tau_i)$ samples before either accepting, or rejected, an arm i.

B.3.2. MARKED BANDIT

In marked bandits, the limited feedback induces an “information-sharing” phenomenon between entries in the same pull. We can now define the information sharing term as:

$$H^M_{i,j,t} = \mathbb{E}_t \left[\frac{1}{1 + \sum_{\ell \in S \cup S^+ - \{i,j\}} \mathbb{I}(X_\ell = 1)} \big| i \in S \right] \tag{33}$$

where again S^+ has the distribution as S^+ in Algorithm 2, and the operator $\mathbb{E}_t$ treats the data in $\mathcal{D}_t$ as deterministic. The following remark explains the intuition behind $H^M_{i,j,t}$.

Remark B.5 *When we query a set $S \cup S^+$, marked bandit feedback uniformly selects one arm in $\{\ell \in S \cup S^+ : X_\ell = 1\}$ if its non-empty and selects no arms otherwise. Hence, the probability of receiving the feedback that $X_i = 1$ given that $i \in S$ and $X_i = 1$ is*

$$\mathbb{E}_t \left[\frac{1}{1 + \sum_{\ell \in S \cup S^+ - \{i\}} \mathbb{I}(X_\ell = 1)} \big| i \in S \right] \tag{34}$$

The above display captures how often the observation $X_i = 1$ is “suppressed” by another arm in the pull. In contrast, $H^M_{i,j,t}$ is precisely the probability of receiving feedback that $X_i = 1$, given that $X_i = 1$ and $i \in S$, but under a slightly different observation model where arm j is never marked, and instead we observe a marking uniformly from $\{\ell \in S \cup S^+ - \{j\} : X_\ell = 1\}$. Hence, we can think of $H^M_{i,j,t}$ as capturing how often arms other than j prevent us from observing $X_i = 1$. Note that the smaller $H^M_{i,j,t}$, the more the information about X_i is suppressed.

We also remark on the scaling of $H^M_{i,j,t}$:

Remark B.6 *Given $i \in S$, $|S \cup S^+ - \{i,j\}| \le k - 1$, and thus $H^M_{i,j,t} \ge 1/k$. When the means are all high, its likely that $\Omega(k)$ arms ℓ in a query will have $X_\ell = 1$, and so we should expect that $H^M_{i,j,t} = O(1/k)$. When the means are small, say $O(1/k)$, then $H^m_{i,j,t}$ can be as large as $\Omega(1)$. This is because if we observe that $X_i = 1$ from a query $S \cup S^+$, then its very likely that $X_i = 1$ in only a constant fraction of them. Stated otherwise: if the means are small, then seeing just one arm uniformly for which $X_i = 1$ as about as informative as seeing all the values of all the arms at once.*

With this definition in place, we have

Proposition B.5

$$\bar{\mu}_{i,t} - \bar{\mu}_{j,t} = (\mu_i - \mu_j) H^M_{i,j,t} \quad \textit{and} \quad V_{i,t} \le \mu_i H^M_{i,j,t} \tag{35}$$

As a consequence, we have

$$\tau_{i,t} \le \frac{1}{H^M_{i,c_{i,t},t}} \left(\frac{256 \max\{\mu_i, \mu_{c_{i,t}}\}}{\Delta_i^2} + \frac{56}{\Delta_i} \right) \le \frac{\tau_i^M}{H^M_{i,c_{i,t},t}} \tag{36}$$

where τ_i^M is as in Equation 12.

Remark B.7 *In the above proposition, the variance term $V_{i,t}$ has a factor $H^M_{i,j,t}$, which cancels out one of the $H^M_{i,j,t}$ terms from the gap $\Delta^2_{i,t}$. If we did not take advantage of a variance-adaptive confidence interval, our sample complexity would have to pay a factor of $(H^M_{i,j,t})^{-2}$ instead of just $(H^M_{i,j,t})^{-1}$.*

It is straightforward to give a worst case lower bound on $H^M_{i,j,t}$:

$$H^M_{i,j,t} \ge H^M := \mathbb{E}_{X_1,\dots,X_{k-1}} \left[\frac{1}{1 + \sum_{\ell \in [k-1]} \mathbb{I}(X_\ell)} \right] \tag{37}$$

As in the semi-bandit case, we can prove the first part Theorem 3.2 by stringing together Lemmas B.3 and B.4 and Remark B.4, using Proposition B.5 to control $\tau_{i,t}$, and Equation 37 to give a worst case bound on the information sharing term. The argument for improving the sample complexity when we can pull fewer than k arms per query (Equation 14 in Theorem 3.2) is a bit more delicate, and is deferred to section C.2.1.

B.3.3. BANDIT SETTING

Fix $i, j \in U_t'$. When UniformPlay pulls both i and j in the same query, we receive no relative information about X_i versus X_j. Moreover, when another arm X_ℓ for $\ell \in S \cup S^+ - \{i\}$ takes a value 1 (now assuming $j \notin S \cup S^+$), it masks all information about X_i. Hence the analogue of the information sharing term $H^M_{i,j,t}$ is the product $H^B_{i,j,t} \cdot \kappa_1$, where

$$\begin{aligned} H^B_{i,j,t} &:= \mathbb{P}_{\cdot|t}\left[\{X_\ell = 0 : \forall \ell \in S \cup S^+ - \{i\}\} \middle| i \in S, j \notin S \right] \quad \text{and} \\ \kappa_1 &:= \mathbb{P}_{\cdot|t}\left[j \notin S \cup S^+ \middle| i \in S \right] = \mathbb{P}_{\cdot|t}\left[j \notin S \middle| i \in S \right] \end{aligned} \tag{38}$$

We defer the interested reader to the proof of Lemma C.1 in the appendix, which transparently derives the dependence on $H^B_{i,j,t} \cdot \kappa_1$. We also show that, due the uniformity of the distribution of S, κ_1 does not depend on the particular indices i and j.

Remark B.8 *As in the Marked Bandit setting, we use a variance-adaptive confidence interval to cancel out one factor of $\kappa_1 H^B_{i,j,t}$. This turns out to incur a dependence on a parameter κ_2 - defined precisely in Section C.3 - which roughly corresponds to the inverse of the fraction of arms in U_t' whose means do not lie in the top $k+1$.*

The balancing set B is chosen precisely to control κ_1 and κ_2 It ensures that arms $i, j \in U_t$ *do not* co-occur in the same query with constant probability (thus bounding κ_1 below) and that each draw of $S \sim \text{Unif}[U_t', k^{(1)}]$ contains a good fraction of small mean arms as well (thus bounding κ_2 above). The following claim makes this precise:

Claim B.6 *Let $\kappa_1 = \mathbb{P}_{\cdot|t}\left[j \in S \middle| i \in S\right]$ and κ_2 be as in Section C.3, Equation 58. Then choice of*

$$|B| = \max\{0, \lceil \frac{5k^{(1)}}{2} - |U| - \frac{1}{2} \rceil\} \tag{39}$$

be as in Algorithm 4 ensures that $\kappa_1 \geq 1/2$, $\kappa_2 \leq 2$, and $|U'| \leq \frac{5}{2}|U|$. Moreover, as long as $n \geq \frac{7k}{2}$, Algorithm 4 can always sample B from the reject set R.

Remark B.9 (Conditions on n) *The condition $n \geq 7k/2$ ensures that the balancing set B is large enough to bound both κ_1 and κ_2. If we omit the balancing set, our algorithm can then identify the top k means for any $n \geq k$, albeit with worse sample complexity guarantees.*

Proposition B.7 (Characterization of the Gaps) *For all i, $\Delta_{i,t} \geq \Delta_i H^B_{i,c_{i,t},t}$ and*

$$\begin{aligned} \frac{\max\{V_{i,t}, V_{c_{i,t}}\}}{\Delta^2_{i,t}} &\leq (1 + 2\kappa_2)\frac{\max\{(1-\mu_i)\bar{\mu}_{i,t}, (1-\mu_{c_{i,t},t})\bar{\mu}_{c_{i,t},t}\}}{\Delta_i \; \Delta_{i,t}} \\ &\leq \frac{1+2\kappa_2}{\kappa_1 H^B_{i,c_{i,t},t}} \cdot \frac{1}{\Delta_i^2}\begin{cases} 2(1-\mu_{k+1})\mu_i + (1-\mu_{k+1})^2(1-H^B_{i,c_{i,t},t}) & i \leq k \\ 2(1-\mu_i)\mu_{k+1} + (1-\mu_i)^2(1-H^B_{i,c_{i,t},t}) & i > k \end{cases} \end{aligned} \tag{40}$$

where κ_1 and κ_2 are as in Claim B.6.

Remark B.10 *Again, the variance-adaptivity of our confidence interval reduces our dependence on information-sharing from $(H^B_{i,j,t})^{-2}$ to $(H^B_{i,j,t})^{-1}$.*

Plugging in κ_1 and κ_2 as bounded by Claim B.6,

$$\tau_{i,t} \leq \frac{56}{\Delta_i H^B_{i,c_{i,t},t}} + \frac{2560}{H^B_{i,c_{i,t},t}} \cdot \frac{1}{\Delta_i^2}\begin{cases} 2(1-\mu_{k+1})\mu_i + (1-\mu_{k+1})^2(1-H^B_{i,c_{i,t},t}) & i \leq k \\ 2(1-\mu_i)\mu_{k+1} + (1-\mu_i)^2(1-H^B_{i,c_{i,t},t}) & i > k \end{cases} \tag{41}$$

We can wrap up the proof by a straightforward lower bound on $H^B_{i,j,t}$:

$$H^B_{i,j,t} \geq H^B := \prod_{\ell \in [k-1]} (1 - \mu_\ell) \tag{42}$$

and by invoking Claim B.6 to apply Lemma B.3 with $\alpha = 5/2$ as long as $n \geq 7k/2$.

Remark B.11 (Conditions on μ_i) *The condition $\mu_i < 1$ ensures identifiability, since the top k arms would be indistinguishable from any subset of k arms which contains a arm i for which $\mu_i = 1$. More quantitatively, this condition ensures that the information sharing term is nonzero.*

Remark B.12 (Looseness of Equation 42) *When all the means $\mu_1, \dots, \mu_n$ are roughly on the same order, the worst case bound on $H^B_{i,j,t}$ in Equation 42 is tight up to constants. Then, as remarked 3.1, there is never an advantage to looking at k-arms at a time and receiving their* $\max$ *over testing each arm individually. On the other hand, if the means vary widely in their magnitude, then there may very well be an advantage to querying k arms at a time.*

For example, suppose there are k high means $\mu_1, \dots, \mu_k \geq 1/2$, and the remaining $n-k$ means are order $1/k$, and $n \gg k^2$. Then, in the early rounds ($|U_t| \gg k^2$), a random pull of S will contain at most a constant number of means from with top k with constant probability, and so $H^B_{i,j,t} = \Omega((1 - O(1/k))^k) = \Omega(1)$. From Lemma C.1, we see empirical means $\widehat{\mu}_{i,t}$ of the high meaned arms will be $\Omega(1)$ variance. Thus, for early stages t, $\tau_{i,t} = O(1/\Delta_i^2)$. That is, we neither pay the penalty for a small information sharing term that we pay when the means are uniformly high, nor pay a factor of k in the variance which would occur when the means are small. However, we still get to test k arms a time, and hence querying k arms at a time is roughly k times as effective as pulling 1.

Appendix C. Computing $\tau_{i,t}$ with (Marked-)Bandit Feedback

C.1. Preliminaries

We need to describe the distribution of two random subsets related to S. Again, taking the data $\mathcal{D}_t$ as given, define the sets $S_{-i\vee j}$ and $S_{-i\wedge j}$ as follows

$$S_{-i\wedge j} \sim \mathrm{Unif}[U'_t - \{i,j\}, k^{(1)} - 2] \quad \text{and} \quad S_{-i\vee j} \sim \mathrm{Unif}[U'_t - \{i,j\}, k^{(1)} - 1]] \tag{43}$$

$S_{-i\wedge j}$ (read: "S minus i *and* j") has the same distribution as $S - \{i,j\}$ given that both i *and* j are in S. Similarly, $S_{-i\vee j}$ (read: "S minus i *or* j") has the same distribution as $S - \{i,j\}$ given that either i or j are in S, but not both. Equivalently, it has the same distribution as $S - \{i\} \big| i \in S, j \notin S$, and symmetrically, as $S - \{j\} \big| j \in S, i \notin S$. We will also define the constant

$$\kappa_1 := \mathbb{P}_t(j \notin S \big| i \in S) = 1 - \frac{k^{(1)} - 1}{|U'_t| - 1} \tag{44}$$

Note that the definition of κ_1 is independent of i and j, is deterministic given the data $\mathcal{D}_t$, and is well defined since Algorithm 3 always ensures $|U'_t| > 1$[3].

C.2. Marked Bandits

In marked bandits, $U_t = U'_t$. Recall the definition

$$H^M_{i,j,t} = \mathbb{E}_{\cdot|t}\left[\frac{1}{1 + \sum_{\ell \in S \cup S^+ - \{i,j\}} \mathbb{I}(X_\ell = 1)} \big| i \in S\right] \tag{45}$$

By splitting up into the case when $j \notin S \big| i \in S$ and $j \in S \big| i \in S$, we can also express

$$\begin{aligned} H^M_{i,j,t} = \kappa_1 \mathbb{E}_{\cdot|t}\left[\frac{1}{1 + \sum_{\ell \in S_{-i\vee j} \cup S^+} \mathbb{I}(X_\ell = 1)}\right] \\ + (1 - \kappa_1)\mathbb{E}_{\cdot|t}\left[\frac{1}{1 + \sum_{\ell \in S_{-i\wedge j} \cup S^+} \mathbb{I}(X_\ell = 1)}\right] \end{aligned} \tag{46}$$

[3] the undecided set, and its modification, always contain at least two elements

Note that $S_{-i\vee j}$ is well defined except when $|U_t - \{i,j\}| = |U_t| - 2 < k^{(1)} - 1$. Since $|U_t| \geq k^{(1)}$, this issue only occurs if $|U_t| = k^{(1)} - 1$, and thus $\kappa^{(1)} = 0$. To make our notation more compact, we let $|S'|_{\mathcal{W}} = \sum_{\ell \in S'} \mathbb{I}(X_\ell = 1)$ (think "cardinality of winners"). In this notation, the above display takes the form:

$$H^M_{i,j,t} = \kappa_1 \mathbb{E}_{\cdot|t}\left[\left(1 + \left|S_{-i\vee j} \cup S^+\right|_{\mathcal{W}}\right)^{-1}\right] + (1-\kappa_1)\mathbb{E}_{\cdot|t}\left[\left(1 + \left|S_{-i\wedge j} \cup S^+\right|_{\mathcal{W}}\right)^{-1}\right] \tag{47}$$

Proof [Proof of Proposition B.5] Our goal is to bound $\bar{\mu}_{i,t} - \bar{\mu}_{j,t}$.

By the law of total probability and the definition of κ_1, we have

$$\begin{aligned}
\bar{\mu}_{i,t} &= \mu_i \mathbb{E}_{\cdot|t}\left[(1 + |S - \{i\}|_{\mathcal{W}})^{-1} \middle| i \in S\right] \\
&= \mu_i \mathbb{P}_{\cdot|t}\left[j \notin S_t \middle| i \in S\right] \mathbb{E}_{\cdot|t}\left[\left(1 + \left|S_{-i\vee j} \cup S^+\right|_{\mathcal{W}}\right)^{-1}\right] \\
&+ \mu_i \mathbb{P}_{\cdot|t}\left[j \in S \middle| i \in S\right] \mathbb{E}_{\cdot|t}\left[\left(1 + \left|\{j\} \cup S^+ \cup S_{-i\wedge j}\right|_{\mathcal{W}}\right)^{-1}\right] \\
&= \mu_i \kappa_1 \mathbb{E}_{\cdot|t}\left[\left(1 + \left|S^+ \cup S_{-i\vee j}\right|_{\mathcal{W}}\right)^{-1}\right] + \mu_i(1-\kappa_1)\mathbb{E}_{\cdot|t}\left[\left(1 + \left|\{j\} \cup S^+ \cup S_{-i\wedge j}\right|_{\mathcal{W}}\right)^{-1}\right]
\end{aligned} \tag{48}$$

By conditioning on the events when arm j takes the values of 1 or zero, respectively, we can decompose $\mathbb{E}[(1 + |\{j\} \cup S_{-i\wedge j}|_{\mathcal{W}})^{-1}]$ into

$$\mu_j \mathbb{E}_{\cdot|t}\left[(2 + \left|S^+ \cup S_{-i\wedge j}\right|_{\mathcal{W}})^{-1}\right] + (1-\mu_j)\mathbb{E}_{\cdot|t}\left[(1 + \left|S^+ \cup S_{-i\wedge j}\right|_{\mathcal{W}})^{-1}\right] \tag{49}$$

Substituting into the previous display and rearranging yields

$$\bar{\mu}_{i,t} \quad = \quad \mu_i H^M_{i,j,t} + \mu_i\mu_j(1-\kappa_1)\mathbb{E}_{\cdot|t}\left[\left(2 + \left|S^+ \cup S_{-i\wedge j}\right|_{\mathcal{W}}\right)^{-1} - \left(1 + \left|S^+ \cup S_{-i\wedge j}\right|_{\mathcal{W}}\right)\right]$$

Hence, we conclude

$$\bar{\mu}_{i,t} - \bar{\mu}_{j,t} = (\mu_i - \mu_j) H^M_{i,j,t} \tag{50}$$

To control $V_{i,t}$, we have $1 - \mu_{i,t} \leq 1$, and

$$\begin{aligned}
\bar{\mu}_{i,t} \quad &= \quad \mu_i \kappa_1 \mathbb{E}\left[\left(1 + \left|S^+ \cup S_{-i\vee j}\right|_{\mathcal{W}}\right)^{-1}\right] + \mu_i(1-\kappa_1)\mathbb{E}\left[\left(1 + \left|\{j\} \cup S^+ \cup S_{-i\wedge j}\right|_{\mathcal{W}}\right)^{-1}\right] \\
&\leq \quad \mu_i \kappa_1 \mathbb{E}\left[\left(1 + \left|S^+ \cup S_{-i\vee j}\right|_{\mathcal{W}}\right)^{-1}\right] + \mu_i(1-\kappa_1)\mathbb{E}\left[\left(1 + \left|S^+ \cup S_{-i\wedge j}\right|_{\mathcal{W}}\right)^{-1}\right] \\
&= \quad \mu_i H^M_{i,j,t}
\end{aligned}$$

■

C.2.1. IMPROVED COMPLEXITY WITH FEWER THAN k PULLS PER QUERY

In this section, we prove the second part of Theorem 3.2, which describes the setting where we permit fewer than k pulls per query.

Proof [Proof of Second Part of Theorem 3.2] We mirror the proof of Lemma B.4 in Section D.3, and adopt its notation where t_i^* be the first stage at which $i \notin U_t$, let t_0 be the first stage for which $|U_t| < k$. The same argument from Lemma B.4 show that

$$\frac{2\alpha}{k}\sum_{i=1}^{n} T(t_i^*) + \sum_{t>t_0} \mathbb{I}(|U_t| > 0) T(t) \tag{51}$$

If t_{fin} is the last stage of the algorithm for which $|U_t| > 0$, then the doubling nature of the sample size lets us bound

$$\sum_{t>t_0} \mathbb{I}(|U_t| > 0) T(t) \le 2T(t_{fin}) \tag{52}$$

and clearly $t_{fin} = \min\{t_i^* : i \in U_{t_{fin}}\}$. We now bound $\tau^M_{i,j,t_{fin}}$ for $i \in U_{t_{fin}}$ and any $j \in U_{t_{fin}}$. Indeed, recall that

$$H^M_{i,j,t} = \kappa_1 \mathbb{E}_{\cdot|t}\left[\left(1 + \left|S_{-i\vee j} \cup S^+\right|_{\mathcal{W}}\right)^{-1}\right] + (1-\kappa_1)\mathbb{E}_{\cdot|t}\left[\left(1 + \left|S_{-i\wedge j} \cup S^+\right|_{\mathcal{W}}\right)^{-1}\right] \tag{53}$$

When we are allowed to pull fewer than k arms at once, then the "Top-Off Set" S^+ is empty (Algorithm 3, Line 3), and so the above is bounded above by $\max\{|S_{-i\vee j}|, |S_{-i\wedge j}|\} \le |U_t| - 1$. Thus, we can easily bound $H^M_{i,j,t} \ge \frac{1}{|U_t|}$. In particular, this bound holds when $j = c_{i,t}$. Hence,

$$\tau_{i,c_{i,t},t_{fin}} = \frac{\tau_i}{H_{i,c_{i,t},t_{fin}}} \le |U_{t_{fin}}| \cdot \tau_i \tag{54}$$

Recalling that $\mathcal{T}_{n,\delta}(\tau)$ is monotone, and applying the easy to verify identity that

$$\mathcal{T}_{n,\delta}(\tau \cdot k') \le 2k' \mathcal{T}_{n,\delta}(\tau) \tag{55}$$

for all $k' \le n$, we have that for all $i \in U_{t_{fin}}$ that

$$T(t_i^*) \quad \le \quad 2\mathcal{T}_{n,\delta}(\tau_{i,j,t_{fin}}) \le 2\mathcal{T}_{n,\delta}(\tau_i |U_{t_{fin}}|) \le 4|U_{t_{fin}}| \mathcal{T}_{n,\delta}(\tau_i) \tag{56}$$

If σ is a permutation such that $\tau_{\sigma(1)} \ge \tau_{\sigma(2)} \ge \dots \ge \tau_{\sigma(n)}$, then for $i \in U_{t_{fin}}$, $\tau_i \le \tau_{\sigma(|U_{t_{fin}}|)}$. Hence, taking the worst case over $|U_{t_{fin}}|$, we have

$$\sum_{t>t_0} \mathbb{I}(|U_t| > 0) T(t) \le 2T(t_{fin}) \le 8|U_{t_{fin}}| \mathcal{T}(\tau_{\sigma(|U_{t_{fin}}|)}) \le 8 \max_{i \in [k-1]} i\mathcal{T}(\tau_{\sigma(i)}) \tag{57}$$

■

C.3. Bandits

In this section, we drop the dependence on t from the sets $U_t, A_t, R_t, U'_t, R'_t$, and let B be the "balancing set" from Algorithm 4; thus, $U' = U \cup B$, $A' = A - B$, and $R' = R - B$. Let $\kappa_1 = 1 - \frac{k^{(1)}-1}{|U'|-1}$ be as in Equation 44, and let

$$\kappa_2 := \frac{k^{(1)} - 1}{|U'| - 2k^{(1)}} \tag{58}$$

Finally, introduce the loss function $\mathcal{L} : 2^{[n]} \to \{0,1\}$ by $\mathcal{L}(S') = \mathbb{I}(\forall \ell \in S' : X_\ell = 0)$. Note $\mathbb{E}[\mathcal{L}(\{\ell\})] = 1 - \mu_\ell$, and if two sets $S', S'' \subset [n]$ are disjoint, then $\mathcal{L}(S' \cup S'') = \mathcal{L}(S') \cdot \mathcal{L}(S'')$. Moreover, if S' and S'' are almost-surely disjoint, random subset of $[n]$ which are independent given the data in $\mathcal{D}_t$, then $\mathbb{E}_t \mathcal{L}(S' \cup S'') = \mathbb{E}_t \mathcal{L}(S') \cdot \mathbb{E}_t \mathcal{L}(S'')$. Hence, the information sharing term can be expressed as

$$H^B_{i,j,t} := \mathbb{E}_{\cdot|t}\left[\mathcal{L}(S_{-i\vee j} \cup S^+)\right] = \mathbb{E}_{\cdot|t}\left[\mathcal{L}(S_{-i\vee j})\right] \cdot \mathbb{E}_{\cdot|t}\left[\mathcal{L}(S^+)\right] \tag{59}$$

and note that this term is nonzero as long as all the means are less than 1, since with nonzero probability, any query of a nonempty set has a nonzero probability of all its arms taking the value zero. The following lemma gives an expression of $(1 - \bar{\mu}_{i,t})$ in terms of κ_1, μ_i, $H^B_{i,j,t}$, and an error term:

Lemma C.1 (Computation of $\bar{\mu}_{i,t}$) *For any $i \neq j \in U'$, we have that*

$$1 - \bar{\mu}_{i,t} = (1 - \mu_i)\kappa_1 H^B_{i,j,t} \cdot (1 + (1 - \mu_j)\,\mathrm{Err}_{i,j,t}) \tag{60}$$

where the term

$$\mathrm{Err}_{i,j,t} := \frac{1 - \kappa_1}{\kappa_1} \cdot \frac{\mathbb{E}_{\cdot|t}\left[\mathcal{L}(S_{-i\wedge j})\right]}{\mathbb{E}_{\cdot|t}\left[\mathcal{L}[S_{-i\vee j}\right]} = \frac{k^{(1)} - 1}{|U'| - k^{(1)}} \cdot \frac{\mathbb{E}_{\cdot|t}\left[\mathcal{L}(S_{-i\wedge j})\right]}{\mathbb{E}_{\cdot|t}\left[\mathcal{L}[S_{-i\vee j}\right]} \tag{61}$$

is symmetric in i and j.

Proof Using the independence of the arms, we have

$$1 - \bar{\mu}_{i,t} = \mathbb{E}_{\cdot|t}\left[\mathcal{L}(S \cup S^+)\big| i \in S\right] = (1 - \mu_i)\mathbb{E}_{\cdot|t}\left[\mathcal{L}(S - \{i\})\big| i \in S\right] \mathbb{E}_{\cdot|t}\left[\mathcal{L}(S^+)\right]$$

For $i \neq j \in U'$, we have

$$\begin{aligned}
\mathbb{E}_{\cdot|t}\left[\mathcal{L}(S - \{i\})\big| i \in S\right] &= \kappa_1 \mathbb{E}_{\cdot|t}\left[\mathcal{L}(S - \{i\})\big| i \in S, j \notin S\right] + (1 - \kappa_1)\mathbb{E}_{\cdot|t}\left[\mathcal{L}(S - \{i\})\big| i \in S, j \in S\right] \\
&= \kappa_1 \mathbb{E}_{\cdot|t}\left[\mathcal{L}(S_{-i\vee j})\right] + (1 - \kappa_1)\mathbb{E}[\mathcal{L}(\{j\} \cup S_{-i\wedge j}] \\
&= \kappa_1 \mathbb{E}_{\cdot|t}\left[\mathcal{L}(S_{-i\vee j})\right] + (1 - \kappa_1)(1 - \mu_j)\mathbb{E}[\mathcal{L}(S_{-i\wedge j})] \\
&= \kappa_1 \mathbb{E}_{\cdot|t}\left[\mathcal{L}(S_{-i\vee j})\right] \left(1 + (1 - \mu_j)\frac{1 - \kappa_1}{\kappa_1} \cdot \frac{\mathbb{E}_{\cdot|t}\left[\mathcal{L}(S_{-i\wedge j})\right]}{\mathbb{E}_{\cdot|t}\left[\mathcal{L}(S_{-i\vee j})\right]}\right)
\end{aligned}$$

The result now follows from plugging in the above display into the first one, and using the definition of κ_1. ■

Since both $H^B_{i,j,t}$ and $\mathrm{Err}_{i,j,t}$ are symmetric in i and j, we get an exact expression for the gaps.

Corollary C.2 (Bandit Gaps)

$$\bar{\mu}_{i,t} - \bar{\mu}_{j,t} = \kappa_1 H_{i,j,t} \cdot (\mu_i - \mu_j) \tag{62}$$

In particular, $\bar{\mu}_{i,t} > \bar{\mu}_{j,t}$ if and only if $\mu_i > \mu_j$, and

$$\Delta_{i,t} = \kappa_1 H_{i,c_{i,t},t} \cdot \left|\mu_i - \mu_{c_{i,t}}\right| \tag{63}$$

To get an expression for $\tau_{i,t}$, as defined in Lemma B.3, we need to get an expression for the ration of the variance to the gap-squared, $\frac{\max\{V_{i,t},V_{c_{i,t}}\}}{\Delta_{i,t}^2}$. We decompose $V_{i,t} = (1-\bar{\mu}_{i,t})\bar{\mu}_{i,t}$, and similarly for $c_{i,t}$, and begin by bounding $(1-\bar{\mu}_{i,t})/\Delta_{i,t}$ and $(1-\bar{\mu}_{c_{i,t},t})/\Delta_{i,t}$:

Lemma C.3

$$\frac{1-\bar{\mu}_{i,t}}{\Delta_{i,t}} \le \frac{(1+2\kappa_2)(1-\mu_i)}{\Delta_i} \quad \textit{and} \quad \frac{1-\bar{\mu}_{c_{i,t},t}}{\Delta_{i,t}} \le \frac{(1+2\kappa_2)(1-\mu_{c_{i,t}})}{\Delta_i} \tag{64}$$

This result uses $1-\bar{\mu}_{i,t}$ to kill off one factor of $\kappa_1 H^B_{i,j,t}$ from the stagewise gaps $\Delta_{i,t}$, so that our final expression $\tau_{i,t}$ depends on the inverse information sharing term, and not its square. The proof of the above lemma is somewhat delicate, and we defer it to the end of this section. Next, we need an upper bound on $\bar{\mu}_{i,t}$. Clearly, we can upper bound this quantity by 1, but this can be loose when the means are small, and so we introduce the following lemma

Lemma C.4

$$\frac{\max\{(1-\mu_i)\bar{\mu}_{i,t}, (1-\mu_{c_{i,t}})\bar{\mu}_{c_{i,t},t}\}}{\Delta_{i,t}} \tag{65}$$

$$\le \frac{1}{\kappa_1 \Delta_i H^B_{i,c_{i,t},t}} \begin{cases} 2(1-\mu_{k+1})\mu_i + (1-\mu_{k+1})^2(1-H^B_{i,c_{i,t},t}) & i \le k \\ 2(1-\mu_i)\mu_{k+1} + (1-\mu_i)^2(1-H^B_{i,c_{i,t},t}) & i > k \end{cases} \tag{66}$$

Combining Corollary C.2, Lemma C.3 and C.4, establishes Proposition B.7

C.3.1. PROOF OF LEMMA C.4

We start out with a simple upper bound on $\bar{\mu}_i$ and $\bar{\mu}_{c_{i,t}}$:

Lemma C.5

$$\bar{\mu}_{i,t} \le \mu_i + \mu_{c_{i,t}} + (1-\mu_i)(1-H^B_{i,c_{i,t},t}) \tag{67}$$

and similarly when we swap i and $c_{i,t}$

Proof [Proof of Lemma C.5] Let $c = c_{i,t}$. For $S' \in 2^{[n]}$, define the "win" function $\mathcal{W}(S')$: $1-\mathcal{L}(S')$ which takes a value of 1 if $\exists \ell \in S' : X_\ell = 1$. By a union bound, $\mathbb{E}[\mathcal{W}(S' \cup S'')] \le \mathbb{E}[\mathcal{W}(S')] + \mathbb{E}[\mathcal{W}(S'')]$, even when S' and S'' are dependent. Hence,

$$\bar{\mu}_{i,t} = \mathbb{E}_{\cdot|t}\left[\mathcal{W}(S\cup\tilde{S})\middle|i \in S\right] \tag{68}$$

$$= \mathbb{E}_{\cdot|t}\left[\mathbb{I}(X_i = 1)\mathcal{W}(S\cup\tilde{S})\middle|i \in S\right] + \mathbb{E}_{\cdot|t}\left[\mathbb{I}(X_i \ne 1)\mathcal{W}(S\cup\tilde{S})\middle|i \in S\right] \tag{69}$$

$$\le \mu_i + (1-\mu_i)\mathbb{E}_{\cdot|t}\left[\mathcal{W}(S-\{i\}\cup\tilde{S})\middle|i \in S\right] \tag{70}$$

Now, using the union bound property of $\mathcal{W}$, we have

$$\mathbb{E}_{\cdot|t}\left[\mathcal{W}(S-\{i\}\cup\tilde{S})\middle|i \in S\right] \le \mu_c + \mathbb{E}_{\cdot|t}\left[\mathcal{W}(S-\{i\}-\{c\}\cup\tilde{S})\middle|i \in S\right] \tag{71}$$

Finally, by decomposing into the cases when $c \in S$ and $c \notin S$, we

$$\mathbb{E}_{\cdot|t}\left[\mathcal{W}(S - \{i\} - \{c\} \cup \tilde{S})\big| i \in S\right] = \kappa_1 \mathbb{E}_t \mathcal{W}(S_{-i\vee c}) + (1-\kappa_1)\mathbb{E}_{\cdot|t}\left[S_{-i\wedge c}\right] \tag{72}$$

Observe that $S_{-i\wedge c} \sim \mathrm{Unif}[U', k^{(1)}-2]$, whereas $S_{-i\vee c} \sim \mathrm{Unif}[U', k^{(1)}-1]$; consequently, playing $S_{-i\vee c}$ has a greater chance of yielding a win than $S_{-i\wedge c}$. Thus, we can bound

$$\mathbb{E}_{\cdot|t}\left[\mathcal{W}(S - \{i\} - \{c\} \cup \tilde{S})\big| i \in S\right] \le \mathbb{E}_{\cdot|t}\left[\mathcal{W}(S_{-i\vee j})\right] = 1 - H^B_{i,c,t} \tag{73}$$

■

Now, Lemma C.4 follows from the following claim, together with the expression for the gap $\Delta_{i,t}$ from Corollary C.2:

Claim C.6

$$\frac{\max\{(1-\mu_i)(\mu_i+\mu_{c_{i,t}}), (1-\mu_{c_{i,t}})(\mu_i+\mu_{c_{i,t}})\}}{|\mu_i - \mu_{c_{i,t}}|} \le \frac{2}{\Delta_i} \cdot \begin{cases} (1-\mu_{k+1})\mu_i & i \le k \\ (1-\mu_i)\mu_k & i > k \end{cases} \tag{74}$$

and

$$\frac{\max\{(1-\mu_i)^2, (1-\mu_{c_{i,t}})^2\}}{|\mu_i - \mu_{c_{i,t}}|} \le \frac{1}{\Delta_i} \begin{cases} (1-\mu_{k+1})^2 & i \le k \\ (1-\mu_i)^2 & i > k \end{cases} \tag{75}$$

Proof Suppose first that $i > k$, so that $(1-\mu_{c_{i,t}})(\mu_i+\mu_{c_{i,t}}) \le (1-\mu_i)(\mu_i+\mu_{c_{i,t}}) \le 2(1-\mu_i)\mu_{c_{i,t}}$. Then,

$$\begin{aligned} \frac{2(1-\mu_i)\mu_{c_{i,t}}}{|\mu_{c_{i,t}} - \mu_i|} &= \frac{2(1-\mu_i)\mu_{c_{i,t}}}{\mu_{c_{i,t}} - \mu_i} && (76) \\ &= \frac{2(1-\mu_i)}{1 - \mu_i/\mu_{c_{i,t}}} && (77) \\ &\le \frac{2(1-\mu_i)}{1-\mu_i/\mu_k} && (78) \\ &\le \frac{2(1-\mu_i)\mu_k}{\mu_k - \mu_i} && (79) \\ &\le \frac{2(1-\mu_i)\mu_k}{\Delta_i} && (80) \end{aligned}$$

The rest follows from similar arguments. ■

C.3.2. PROOF OF LEMMA C.3

Lemma C.3 follows from the expression for the gaps in Corollary C.2, and the following technical lemma:

Lemma C.7 *Fix $i \in U'$, and let $c \in U' \cap [k]$ if $i > k$ and $c \in U' - [k]$. Then,*

$$\frac{1-\bar{\mu}_{i,t}}{|\mu_i - \mu_c|} \le \frac{1-\mu_i}{\Delta_i} \cdot \kappa_1(1+2\kappa_2)H_{i,j,c} \tag{81}$$

and

$$\frac{1-\bar{\mu}_{c,t}}{|\mu_i - \mu_c|} \le \frac{1-\mu_c}{\Delta_i} \cdot \kappa_1(1+2\kappa_2)H_{i,j,c} \tag{82}$$

Proof By Lemma C.1,

$$1-\bar{\mu}_{i,t} \quad = \quad (1-\mu_i)\kappa_1 H_{i,c,t}\left(1+(1-\mu_c)\mathrm{Err}_{i,c,t}\right). \tag{83}$$

The following lemma, proved later, controls the term on $\mathrm{Err}_{i,c,t}$.

Lemma C.8 *Suppose that $j \in [k+1]$, and that the balancing set B satisfies $B \cap [k] = \emptyset$. Then, for any $i \neq c \in U$ (where possibly $j \neq c$), we have*

$$(1-\mu_j)\mathrm{Err}_{i,c,t} \quad \le \quad \kappa_2. \tag{84}$$

When $i > k$, $c \in [k]$ and $1-\bar{\mu}_{c,t} \le 1-\bar{\mu}_{i,t}$ so that

$$\begin{aligned}
\frac{1-\bar{\mu}_{c,t}}{|\bar{\mu}_{i,t}-\bar{\mu}_{c,t}|} \quad &\le \quad \frac{1-\bar{\mu}_{i,t}}{|\bar{\mu}_{i,t}-\bar{\mu}_{c,t}|} && (85)\\
&\le \quad \frac{(1-\mu_i)\kappa_1 H_{i,c,t}(1+\kappa_2)}{|\bar{\mu}_{i,t}-\bar{\mu}_{c,t}|} && (86)\\
&\le \quad \frac{(1-\mu_i)(1+\kappa_2)}{|\mu_i-\mu_c|} && (87)\\
&\le \quad \frac{(1-\mu_i)(1+\kappa_2)}{\Delta_i} && (88)
\end{aligned}$$

where (86) follows from combining (83) and Lemma C.8, (87) follows from Corollary C.2, and (88) holds by $|\mu_i - \mu_c| \ge \max\{\Delta_i, \Delta_c\}$. Moreover, swapping the roles of c and i, we have that when $i \le k$,

$$\begin{aligned}
\frac{1-\bar{\mu}_{c,t}}{|\bar{\mu}_{i,t}-\bar{\mu}_{c,t}|} \quad &\le \quad \frac{(1-\mu_c)\kappa_1 H_{i,c,t}(1+\kappa_2)}{|\bar{\mu}_{i,t}-\bar{\mu}_{c,t}|} && (89)\\
&\le \quad \frac{(1-\mu_c)(1+\kappa_2)}{\Delta_i}. && (90)
\end{aligned}$$

The final case we need to deal with is the computation of $\frac{1-\bar{\mu}_{i,t}}{|\bar{\mu}_{i,t}-\bar{\mu}_{c,t}|}$ when $i \le k$. The problem is that it might be the case that $c > k+1$, impeding the application of Lemma C.8. We get around this issue by breaking up into cases:

(1) If $1-\mu_c$ and $1-\mu_i$ are on the same order, we are not in so much trouble. Indeed, if $1-\mu_c \le 2(1-\mu_i)$, then, we have

$$\begin{aligned}
1-\bar{\mu}_{i,t} \quad &= \quad (1-\mu_i)H_{i,c,t}\left(1+(1-\mu_c)\mathrm{Err}_{i,c,t}\right)\\
&\le \quad (1-\mu_i)H_{i,c,t}\left(1+2(1-\mu_{k+1})\mathrm{Err}_{i,c,t}\right)\\
&\le \quad (1-\mu_i)H_{i,c,t}\left(1+2\kappa_2\right)
\end{aligned}$$

where the last step follows from applying Lemma C.8 with $j = k+1$.

(2) What happens when $1 - \mu_c > 2(1-\mu_i)$? Then we have

$$\begin{aligned}
(\mu_i - \mu_c)^{-1}(1-\mu_c) &= \frac{1-\mu_c}{\Delta_i} \cdot \frac{\Delta_i}{\mu_i - \mu_c} \\
&= \frac{1-\mu_{k+1}}{\Delta_i} \cdot \frac{\Delta_i}{\mu_i - \mu_c} \cdot \frac{1-\mu_c}{1-\mu_{k+1}}
\end{aligned}$$

More suggestively, we can write the above as

$$\frac{1-\mu_{k+1}}{\Delta_i} \cdot \frac{(1-\mu_{k+1}) - (1-\mu_i)}{(1-\mu_c) - (1-\mu_i)} \cdot \frac{1-\mu_c}{1-\mu_{k+1}} \tag{91}$$

As soon as $(1-\mu_c) > 2(1-\mu_i)$, Equation 91 is bounded by

$$\begin{aligned}
(\mu_i - \mu_c)^{-1}(1-\mu_c) = \frac{1-\mu_{k+1}}{\Delta_i} \cdot \frac{(1-\mu_{k+1}) - (1-\mu_i)}{\frac{1}{2}(1-\mu_c)} \cdot \frac{1-\mu_c}{1-\mu_{k+1}} &= \frac{2((1-\mu_{k+1}) - (1-\mu_i))}{\Delta_i} \\
&\leq 2\frac{1-\mu_{k+1}}{\Delta_i}
\end{aligned}$$

Hence,

$$\begin{aligned}
\frac{1}{\mu_i - \mu_c} \cdot H_{i,c,t}\left(1 + (1-\mu_c)\mathrm{Err}_{i,c,t}\right) &= \frac{H_{i,c,t}}{\mu_i - \mu_c} + \frac{1-\mu_c}{\mu_i - \mu_c} \cdot \mathrm{Err}_{i,c,t} H_{i,c,t} \\
&\leq \frac{H_{i,c,t}}{\Delta_i} + \frac{2H_{i,c,t}}{\Delta_i}\left((1-\mu_{k+1})\mathrm{Err}_{i,c,t}\right) \\
&= \frac{H_{i,c,t}}{\Delta_i}\left(1 + 2(1-\mu_{k+1})\mathrm{Err}_{i,c,t}\right) \\
&\leq \frac{H_{i,c,t}}{\Delta_i}\left(1 + 2\kappa_2\right)
\end{aligned}$$

where the last line follows from Lemma C.8 with $j = k+1$. ■

Proof [Proof of Lemma C.8]

$S_{-i\vee c}$ has the same distribution $S_{-i\wedge c} \cup y$, where $y \sim \mathrm{Unif}[U' - S_{-i\wedge c} - \{i,c\}, 1]$. If $Y \sim$ Bernoulli(μ_y) then

$$\mathbb{E}_{\cdot|t}\left[\mathcal{L}(S_{-i\vee c})\right] = \mathbb{E}(1-Y)\mathcal{L}(S_{-i\wedge c}) = \mathbb{E}_{\cdot|t}\left[\mathbb{E}_{\cdot|t}\left[1 - Y \middle| S_{-i\wedge c}\right] \cdot \mathcal{L}(S_{-i\wedge c})\right] \tag{92}$$

Since $j \in [k+1]$, $\mu_j \geq \mu_\ell$ for all $\ell \notin [k]$, and thus $(1-\mu_\ell) \geq (1-\mu_j)$ for all $\ell \notin [k]$. It thus follows that

$$\begin{aligned}
\mathbb{E}_{\cdot|t}\left[1 - Y \middle| S_{-i\wedge c}\right] &= \frac{1}{|U' - \{i,c\} - S_{-i\wedge c}|} \sum_{\ell \in U' - \{i,c\} - S_{-i\wedge c}} (1-\mu_\ell) \\
&\geq \frac{1}{|U' - \{i,c\} - S_{-i\wedge c}|} \sum_{\ell \in U' - \{i,c\} - S_{-i\wedge c} - [k]} (1-\mu_\ell) \\
&\geq \frac{1}{|U' - \{i,c\} - S_{-i\wedge c}|} \sum_{\ell \in U' - \{i,c\} - S_{-i\wedge c} - [k]} (1-\mu_j) \\
&= \frac{|U' - \{i,c\} - S_{-i\wedge c} - [k]|}{|U' - \{i,c\} - S_{-i\wedge c}|}(1-\mu_j)
\end{aligned}$$

If $r \in [k] \cap U' = U \cup B$, then we must have $r \in U$, since $B \cap [k] = \emptyset$ by assumption. This implies that $|U' - \{i,c\} - S_{-i\wedge c} - [k]| \ge |U' - \{i,c\} - S_{-i\wedge c}| - \min\{k, |U|\}$. Using the fact that $|U' - \{i,c\} - S_{-i\wedge c}| = |U'| - k^{(1)}$, and that $k^{(1)} = \min\{k, |U|\}$, we conclude that

$$\mathbb{E}_{\cdot|t}\left[1 - Y \middle| S_{-i\wedge c}\right] \ge (1-\mu_j)\frac{|U'| - k^{(1)} - \min\{k,|U|\}}{|U'| - k^{(1)}} \tag{93}$$

$$= (1-\mu_j)\frac{|U'| - 2k^{(1)}}{|U'| - k^{(1)}} \tag{94}$$

Thus, this entails that $\mathbb{E}[\mathcal{L}(S_{-i\vee c})] \ge (1-\mu_j)\frac{|U'|-2k^{(1)}}{|U'|-k^{(1)}}\mathbb{E}[\mathcal{L}(S_{-i\wedge c})]$, and hence

$$(1-\mu_j)\mathrm{Err}_{i,c,t} = \frac{k^{(1)}-1}{|U'|-k^{(1)}} \cdot \frac{(1-\mu_j)\mathbb{E}\mathcal{L}(S_{-i\wedge c})}{\mathbb{E}\mathcal{L}(S_{-i\vee c})} \tag{95}$$

$$\le \frac{k^{(1)}-1}{|U'|-k^{(1)}} \cdot \frac{|U'|-k^{(1)}}{|U'|-2k^{(1)}} \tag{96}$$

$$= \frac{k^{(1)}-1}{|U'|-2k^{(1)}} := \kappa_2 \tag{97}$$

as needed. ■

C.3.3. CONTROLLING κ_1 AND κ_2

Proof [Proof of Claim B.6] For ease of notation, drop the dependence on the round t and the definitions $\kappa_1 = 1 - \frac{k^{(1)}-1}{|U'|-1}$ and $\kappa_2 = \frac{k^{(1)}-1}{|U'|-2k^{(1)}}$. Noting that $|U'| = |B| + |U|$, we see that if $\kappa_1 \ge 1/2$ is desired, we require that

$$\kappa_1 \ge 1/2 \iff |U'| - 1 \ge 2(k^{(1)} - 1) \iff |B| \ge 2k^{(1)} - |U| - 1 \tag{98}$$

Whereas

$$\kappa_2 \le 2 \iff 2(|U'| - 2k^{(1)}) \ge k^{(1)} - 1 \iff |B| \ge \frac{5}{2}k^{(1)} - |U| - \frac{1}{2} \tag{99}$$

Hence $\kappa \le 2 \implies \kappa_1 \le 1/2$, and the above display makes it clear that the choice of B in Algorithm 4 ensures that this holds. To verify the second condition, note that when $|B| = 0$, then $|U'| = |U|$. When $|B| > 0$, we have

$$|B| = \lceil \frac{5k^{(1)}}{2} - |U| - \frac{1}{2} \rceil \le \frac{5k^{(1)}}{2} - |U| \tag{100}$$

so that $|U'| = |U| + |B| \le \frac{5}{2}\min\{|U|, k\}$. Finally, in order to always sample a balance set $B \subseteq R$, we need to ensure that at each round, $|R| \ge |B|$. Again, we may assume that $|B| > 0$, so that $|U| + |B| \le \frac{5k}{2}$. Using the facts that $|R| + |A| + |U| = n$ (every item is rejected, accepted, and undecided) and $|A| \le k-1$ (k accepts ends the algorithm), we have $|R| \ge n - |U| - (k-1) \ge n - |U| - (k-1)$. But $n - |U| - (k-1) \ge |B| \iff n \ge (k-1) + |U'| \ge \frac{7k}{2}$, as needed. ■

Appendix D. Concentration Proofs for Section B.2

D.1. An Empirical Bernstein

The key technical ingredient is an empirical version of Bernstein's inequality, which lets us build variance-adaptive confidence intervals:

Theorem D.1 (Modification of Theorem 11 in Maurer and Pontil (2009)) *Let $Z := (Z_1, \dots, Z_n)$ be a sequence of independent random variables bounded by $[0,1]$. Let $\bar{Z}_n = \frac{1}{n}\sum_i Z_i$, $\bar{Z} := \mathbb{E}[\bar{Z}_n]$, let $\mathrm{Var}_n[Z]$ denote the empirical variance of Z, $\frac{1}{n-1}\sum_{i=1}^n (Z_i^2 - \bar{Z}_n^2)$, and set $\mathrm{Var}[Z] := \mathbb{E}[\mathrm{Var}_n[Z]]$. Then, with probability $1-\delta$,*

$$\begin{aligned} \left|\bar{Z} - \bar{Z}_n\right| &\leq \sqrt{\frac{2\mathrm{Var}_n[Z]\log(4/\delta)}{n}} + \frac{8\log(4/\delta)}{3(n-1)} && (101) \\ &\leq \sqrt{\frac{2\mathrm{Var}[Z]\log(4/\delta)}{n}} + \frac{14\log(4/\delta)}{3(n-1)} && (102) \end{aligned}$$

The result follows from Bernstein's Inequality, and the following concentration result regarding the square root of the empirical variance.

Lemma D.2 (Theorem 10 in Maurer and Pontil (2009)) *In the set up of Theorem D.1,*

$$\left|\sqrt{\mathbb{E}[\mathrm{Var}_n[Z]]} - \sqrt{\mathrm{Var}_n[Z]}\right| \leq \sqrt{\frac{2\log(2/\delta)}{n-1}} \tag{103}$$

hold with probability $1-\delta$.

Proof [Proof of Theorem D.1] The argument follows the proof of Theorem 11 in Maurer and Pontil (2009). Let $W := \frac{1}{n}\sum_{i=1}^n \mathrm{Var}[Z_i]$. It is straightforward to verify that $W \leq \mathbb{E}[\mathrm{Var}_n[X]]$, and hence Bernstein's inequality yields that, with probability $1-\delta$,

$$\begin{aligned} \left|\frac{1}{n}\sum_{i=1}^n Z_i - \mathbb{E}[Z_i]\right| &\leq \sqrt{\frac{2W\log(4/\delta)}{n}} + \frac{2\log(4/\delta)}{3n} \\ &\leq \sqrt{\frac{2\mathbb{E}[\mathrm{Var}_n[Z]]\log(4/\delta)}{n}} + \frac{2\log(4/\delta)}{3n} \\ &\leq \sqrt{\frac{2\log(4/\delta)}{n}} \cdot \sqrt{\mathrm{Var}_n[Z]} + \frac{2\log(4/\delta)}{\sqrt{n(n-1)}} + \frac{2\log(4/\delta)}{3n} && (104) \\ &< \sqrt{\frac{2\mathrm{Var}_n[Z]\log(4/\delta)}{n}} + \frac{8\log(4/\delta)}{3(n-1)} \\ &< \sqrt{\frac{2\mathbb{E}[\mathrm{Var}_n[Z]]\log(4/\delta)}{n}} + \frac{14\log(4/\delta)}{3(n-1)} \end{aligned}$$

which completes the proof.

■

In our algorithm, the confidence intervals $\hat{C}_{i,t}$ depend on sample variances, and are thus random. To insure they are bounded above, we define a confidence parameter $C_{i,t}$ which depends on the true (but unkown) stagewise variance parameter

$$C_{i,t} := \sqrt{\frac{2V_{i,t}\log(8nt^2/\delta)}{T(t)}} + \frac{14\log(8nt^2/\delta)}{3(T(t)-1)} \tag{105}$$

We extend our Empirical Bernstein bound to a union bound over all rounds $t \in \{1, 2, \dots\}$, showing that, uniformly over all rounds, $\hat{C}_{i,t}$ is a reasonable confidence interval and never exceeds $C_{i,t}$:

Lemma D.3 (Stagewise Iterated Logarithm Bound for Empirical Bernstein) *Let*

$$\mathcal{E} := \{\cap_{t=1}^{\infty} \cap_{i=1}^{n} \{|\hat{\mu}_{i,t} - \bar{\mu}_{i,t}| \le \hat{C}_{i,t} \le C_{i,t}\}\} \tag{106}$$

Then $\mathbb{P}(\mathcal{E}) \ge 1 - \delta$.

Proof Let $\mathcal{E}_{i,t}$ denote the event that $\{|\hat{\mu}_{i,t} - \bar{\mu}_{i,t}| \le \hat{C}_{i,t} \le C_{i,t}\}$. Conditioned on any realization of the data $\mathcal{D}_t$ at stage t, an application of Theorem D.1 shows that $\mathbb{P}(\mathcal{E}_{i,t}|\mathcal{D}_t) \le \frac{\delta}{2nt^2}$. Integrating over all such realizations, $\mathbb{P}(\mathcal{E}_{i,t}) \le \frac{\delta}{2nt^2}$. Finally, taking a union bound over all stages t and arms $i \in [n]$ shows that

$$\mathbb{P}(\mathcal{E}) \le \sum_{t=1}^{\infty}\sum_{i=1}^{n}\mathbb{P}(\mathcal{E}_{i,t}) \le \sum_{t=1}^{\infty}\sum_{i=1}^{n}\frac{\delta}{2nt} = \frac{\delta}{2}\sum_{t=1}^{\infty} t^{-2} \le \delta \tag{107}$$

■

We now invert the Iterated Logarithm via

Lemma D.4 (Inversion Lemma) *For any* $\Delta > 0$ *and* $t \ge 2$, $C_{i,t} \le \Delta$ *as long as*

$$T \ge \left(\frac{16V_{i,t}}{\Delta^2} + \frac{14}{\Delta}\right)\log\left(\frac{24n}{\delta}\log\left(\frac{12n}{\delta}\left(\frac{16V_{i,t}}{\Delta^2} + \frac{14}{\Delta}\right)\right)\right) \tag{108}$$

Proof It suffices to show that $\sqrt{\frac{2V_{i,t}\log(8nt^2/\delta)}{T(t)}} \le \Delta/2$ and $\frac{14\log(8nt^2/\delta)}{3(T(t)-1)} \le \Delta/2$. Since $t^2 = (\log_2(T))^2 \le (\log_2 e \log(T))^2$, it suffices that

$$\frac{8V_{i,t}\log(8n\log_2^2 e \log^2(T(t))/\delta))}{\Delta^2 T(t)} \le 1 \quad \text{and} \quad \frac{28\log(8n\log_2 e\log^2(T(t))/\delta)}{3\Delta(T(t)-1)} \le 1$$

As long as $t \ge 2$, so that $T(t) \ge e$, it suffices that

$$\frac{16V_{i,t}\log(8n\log_2 e\log(T(t))/\delta))}{\Delta^2 T(t)} \le 1 \quad \text{and} \quad \frac{14\log(8n\log_2 e\log(T(t))/\delta)}{\Delta T(t)} \le 1$$

Let $\alpha_1 = 16V_{i,t}/\Delta^2$, $\alpha_2 = 14/\Delta$ and $\beta = 8n\log_2 e/\delta < 12n/\delta$. Then both inequalities take the form

$$\alpha_p \log(\beta\log(T))/T \le 1 \tag{109}$$

where we simplify $T(t) = T$. Using the inversion

$$T \geq \alpha \log(2\beta \log(\alpha\beta)) \implies \alpha \log(\beta \log(T))/T \leq 1 \tag{110}$$

we obtain that it is sufficient for $T \geq (\alpha_1+\alpha_2)\log(2\beta\log(\alpha_1+\alpha_2)) \geq \max_p \alpha_p \log(2\beta\log(\alpha_p\beta))$, or simply

$$T \geq \left(\frac{16V_{i,t}}{\Delta^2} + \frac{14}{\Delta}\right) \log\left(\frac{24n}{\delta} \log\left(\frac{12n}{\delta}\left(\frac{16V_{i,t}}{\Delta^2} + \frac{14}{\Delta}\right)\right)\right) \tag{111}$$

■

D.2. Proof of Theorem B.3

We show that Theorem B.3 holds as long as the event $\mathcal{E}$ from Lemma D.3 holds. The definition of $\mathcal{E}$ and Algorithm 3 immediately imply that no arms in $[k]$ are rejected, and no arms in $[n] - [k]$ are accepted. To prove the more interesting part of the theorem, fix an index $i \in U_t$, and define

$$\mathcal{C}(i) := \begin{cases} \{j \in U_t, j > k\} & i \leq k \\ \{j \in U_t, j \leq k\} & i > k \end{cases} \tag{112}$$

Also, let $c_i = \arg\min_{j\in\mathcal{C}(i)} |\mu_i - \mu_j|$. We can think of $\mathcal{C}(i)$ as the set of all arms competing with i for either an accept or reject, and c_i as the competitor closest i in mean. For $i > k$ to be rejected, it is sufficient that, for all $j \in \mathcal{C}(i)$, $\min_{j\in\mathcal{C}(i)} \hat{\mu}_{j,t} - \hat{C}_{j,t} \geq \hat{\mu}_{i,t} + \hat{C}_{j,t}$. Under $\mathcal{E}$, $\hat{\mu}_{j,t} - \hat{C}_{j,t} \geq \bar{\mu}_{j,t} - 2C_{j,t}$, and $\bar{\mu}_{i,t} \leq \bar{\mu}_{i,t} + 2C_{i,t}$, so that it is sufficient for

$$\forall j \in \mathcal{C}(i) : \bar{\mu}_{j,t} - \bar{\mu}_{i,t} \geq 2(C_{i,t} + C_{j,t}) \tag{113}$$

Analogously, for $i \leq k$, i is accepted under $\mathcal{E}$ as long as $\forall j \in \mathcal{C}(i) : \bar{\mu}_{i,t} - \bar{\mu}_{j,t} \geq 2(C_{i,t} + C_{j,t})$. Defining $\Delta_{i,j,t} := |\bar{\mu}_{i,t} - \bar{\mu}_{j,t}|$, we subsume both cases under the condition

$$\forall j \in \mathcal{C}(i) : \Delta_{i,j,t} \geq 2(C_{i,t} + C_{j,t}) \tag{114}$$

for which it is sufficient to show that

$$\forall j \in \mathcal{C}(i) : C_{i,t} \leq \Delta_{i,j,t}/4 \quad \text{and} \quad C_{j,t} \leq \Delta_{i,j,t}/4 \tag{115}$$

To this end define

$$\tau^{(1)}_{i,j,t} = \frac{256V_{i,t}}{\Delta^2_{i,j,t}} + \frac{56}{\Delta_{i,j,t}} \quad \text{and} \quad \tau^{(2)}_{i,j,t} := \frac{256V_{j,t}}{\Delta^2_{i,j,t}} + \frac{56}{\Delta_{i,j,t}} \tag{116}$$

We now show that $\tau_{i,t} = \max_{j\in\mathcal{C}(i)} \max\left\{\tau^{(1)}_{i,j,t}, \tau^{(2)}_{i,j,t}\right\}$, which by Lemmas D.3 and D.4 implies that Equation 115 will holds as long as

$$T \geq \tau_{i,t} \log\left(\frac{24n}{\delta} \log\left(\frac{12n\tau_{i,t}}{\delta}\right)\right) \tag{117}$$

Now, we bound $\tau_{i,t}$. Note that $\Delta_{i,j,t} \geq \Delta_{i,c_i,t} := \Delta_{i,t}$ for all $j \in \mathcal{C}(i)$. This implies that $\max_{j\in\mathcal{C}(i)} \tau^{(1)}_{i,j,t} \leq \frac{256V_{i,t}}{\Delta^2_{i,t}} + \frac{56}{\Delta_{i,t}}$. On the other hand, it holds that

$$\max_{j\in\mathcal{C}(i)} \tau^{(2)}_{i,j,t} \quad \leq \quad 256 \max_{j\in\mathcal{C}(i)} \left(\frac{V_{j,t}}{\Delta^2_{i,j,t}}\right) + \frac{56}{\Delta_{i,t}} \leq \frac{256V_{c_i,t}}{\Delta^2_i} + \frac{56}{\Delta_{i,t}} \tag{118}$$

where the second inequality invokes the following lemma.

Lemma D.5 *For $i \in \{1,2,3\}$, $Z_i \sim$ Bernoulli(p_i), where either $p_1 < p_2 < p_3$ or $p_3 > p_2 > p_1$. Then,*

$$\frac{\mathrm{Var}[Z_2]}{(\mathbb{E}[Z_1 - Z_2])^2} \geq \frac{\mathrm{Var}[Z_3]}{(\mathbb{E}[Z_1 - Z_3])^2} \tag{119}$$

Proof [Proof of Lemma D.5] The desired inequality and conditions are invariant under the transformation $p_i \mapsto 1 - p_i$ for $i \in \{1,2,3\}$, so we may assume without loss of generality that. $p_1 < p_2 < p_3 \in [0,1]$. Then $1 > p_1/p_2 > p_1/p_3$, which implies that

$$\begin{aligned} \frac{1}{1 - p_1/p_2} \geq \frac{1}{1 - p_1/p_3} \quad &\implies \quad \frac{p_2}{p_2 - p_1} \geq \frac{p_3}{p_3 - p_1} \\ &\implies \quad \frac{(1-p_2)p_2}{p_2 - p_1} \geq \frac{(1-p_3)p_3}{p_3 - p_1} \\ &\implies \quad \frac{(1-p_2)p_2}{(p_2 - p_1)^2} \geq \frac{(1-p_3)p_3}{(p_3 - p_1)^2} \end{aligned}$$

which is precisely the desired inequality. ■

D.3. Proof of Lemma B.4

Let e_t be denote the the "efficiency", so that, at round t, each call of uniform play for $s = 1, \dots, T(t)$ makes at most $e_t|U_t|$ queries. Furthermore, let τ_0 denote the first time such that $|U_t| < k$. By assumption, we have that $e_t \leq \frac{\alpha}{k}$ for $0 \leq t < \tau_0$, and that $e_t|U_t| \leq \alpha$ for $t \geq t_0$. Finally, let $\tau^*_i = \inf\{t : i \notin U_t\}$. Then, the total number of samples we collect is

$$\sum_{t=0}^{\infty} e_t|U_t|T(t) \quad = \quad \sum_{t=0}^{\tau_0 - 1} e_t|U_t|T(t) + \sum_{t=\tau_0}^{\infty} e_t|U_t|T(t) \tag{120}$$

$$\leq \quad \frac{\alpha}{k}\sum_{t=0}^{\tau_0 - 1} |U_t|T(t) + \alpha\sum_{t=\tau_0}^{\infty} \mathbb{I}(U_t \neq \emptyset)T(t) \tag{121}$$

The first sum can be re-arranged via

$$\sum_{t=0}^{\tau_0 - 1} |U_t|T(t) \quad = \quad \sum_{t=0}^{\tau_0 - 1}\left(\sum_{i=1}^{n} \mathbb{I}(i \in u_t)\right)T(t) = \sum_{i=1}^{n}\sum_{t=0}^{\tau_0 + 1} \mathbb{I}(i \in U_t)T(t) \tag{122}$$

$$\leq \quad \sum_{i=1}^{n}\sum_{t=0}^{\infty} \mathbb{I}(i \in U_t)T(t) \leq \sum_{i=1}^{n} 2^{\tau^*_i + 1} \tag{123}$$

whereas the second sum is bounded above by $\sum_{t=\tau_0}^{\infty} \mathbb{I}(U_t \neq \emptyset) T(t) \leq 2^{\max_j \tau_j^* + 1}$. Hence,

$$\sum_{t=0}^{\infty} e_t |U_t| T(t) \leq 2\alpha(2^{\max_j \tau_j^*} + \frac{1}{k}\sum_{i=1}^{n} 2^{\tau_i^*}) \tag{124}$$

Finally, let $T_i^* := 2^{\tau_i^*}$, and let $\sigma() : [n] \to n$ denote a permutation such that $T^*_{\sigma(1)} \geq T^*_{\sigma(2)} \dots T^*_{\sigma(n)}$. Then, a straight forward manipulation of the above display yields that

$$\sum_{t=0}^{\infty} e_t |U_t| T(t) \leq 2\alpha(2T^*_{\sigma(1)} + \frac{1}{k}\sum_{i=k+1}^{n} T^*_{\sigma(i)}) \tag{125}$$

since $\frac{1}{k}\sum_{i=1}^{k} T^*_{\sigma(i)} \leq T^*_{\sigma(1)}$.

Appendix E. Dependent Lower Bound Proof

Recall that we query subsets of $S \subset \mathcal{S} := \binom{[n]}{k}$. Let T_S denote the number of times a given subset S is queried, and note that the expected sample complexity is simply:

$$\sum_{S \in \mathcal{S}} \mathbb{E}[T_S]$$

Further, let $d(x, y)$ denote the KL-divergence between two independent, Bernoulli random variables with means x and y, respectively. We first need a technical lemma, whose proof we defer the end of the section:

Lemma E.1 *Let $d(x, y) = x \log(\frac{x}{y}) + (1 - x) \log(\frac{1-x}{1-y})$. Then*

$$\frac{(y-x)^2/2}{\sup_{z \in [x,y]} z(1-z)} \leq d(x,y) \leq \frac{(y-x)^2/2}{x(1-x) - [(y-x)(2x-1)]_+} \leq \frac{(y-x)^2/2}{\min\{x(1-x), y(1-y)\}} \tag{126}$$

We break the proof up into steps. First we construct the dependent measure ν that is $(k-1)$-wise independent, meaning that for any subset $S \in \binom{[n]}{k}$, any subset of size $(k-1)$ of S behaves like independent arms. The construction makes it necessary to consider each set of k individually. To obtain the lower bounds we appeal to a change of measure argument (see Kaufmann et al. (2015) for details) that proposes an alterantive measure ν' in which a different subset is best than that subset that is best in ν, and then we calculate the number of measurements necessary to rule out ν'. The majority of the effort goes into 1) computing the gap between the best and all other subsets and 2) computing the KL divergences between ν and the alterantive measures nu' under the bandit and semi-bandit feedback mechanisms.

Step 1: Construct ν:

Fix $p \in [0, 1]$ and $\mu \in [0, 1/2]$. Let $X = (X_1, \dots, X_n)$ be distributed according to ν. Define the independent random variables Y as Bernoulli(p), Z_i as Bernoulli$(1/2)$, and U_i as Bernoulli(2μ) for all $i \in [n]$. For $i > 1$ let $X_i = Z_i U_i$ and let

$$X_1 = U_1 \widetilde{Z}_1 \qquad \text{where} \qquad \widetilde{Z}_1 = \begin{cases} 1 + \oplus_{i=2}^{k} Z_i & \text{if } Y = 1 \\ Z_1 & \text{if } Y = 0 \end{cases}$$

where $\oplus$ denotes modular-2 addition. Note that $\mathbb{E}_{\nu}[X] = \mu\mathbf{1}$ since

$$\mathbb{E}_{\nu_1}[X_1] = 2\mu\left[p\,\mathbb{P}_{\nu}\left(1 + \oplus_{i=2}^{k} Z_i = 1\right) + (1-p)\tfrac{1}{2}\right] = \mu$$

and the calculation for $\mathbb{E}[X_i]$ for $i > 1$ are immediate by independence. Henceforth, denote $S^* = \{1, \dots, k\}$.

Step 2: Relevant Properties of ν:

1. *Any subset of arms S which doesn't contain all of S^* are independent.* If $Y = 0$ then the claim is immediate so assume $Y = 1$. We may also assume that $1 \in S$, since otherwise the arms are independent by construction. Finally, we remark that even when $1 \in S$ and $Y = 1$, all arms in S are conditionally independent given $\{Z_i : i \in S \cap S^*\}$. Thus, it suffices to verify that $\{Z_i : i \in S \cap S^* - 1\} \cup \{\widetilde{Z}_1\}$ have a product distribution. To see this, note that $\{Z_i : i \in S \cap S^* - 1\}$ is a product distribution, so it suffices to show that $\widetilde{Z}_1$ is independent of $\{Z_i : i \in S \cap S^* - 1\}$. Write $\widetilde{Z}_1 = 1 + \oplus_{i \in S^* \setminus 1} Z_i = 1 \oplus_{i \in S^* \cap S} Z_i \oplus_{i \in S^* \setminus S} Z_i$. The sum over Z_i not in S^*, $\oplus_{i \in S^* \setminus S} Z_i$, is Bernoulli(1/2), and independent of all the Z_i for which $i \in S^* \cap S$. Thus, conditioned on any realization of $\{Z_i : i \in S \cap S^*\}$, $\widetilde{Z}_1$ is still Bernoulli(1/2), as needed.

2. *The distribution of ν is invariant under relabeling of arms in S^*, and under relabeling of arms $[n] \setminus S^*$.* The second part of the statement is clear. Moreover, since the arms in $[n] \setminus S^*$ are independent of those S^*, it suffices to show that the distribution of arms in S^* are invariant under relabeling. Using the same arguments as above, we may reduce to the case where $Y = 1$, and only verify that the distribution of $\{\widetilde{Z}_1\} \cup \{Z_i : i \in S^* - 1\}$ is invariant under relabeling.

   To more easily facilliate relabeling, we adjust our notation and set $\widetilde{Z}_i = Z_i$ for $i \in S^* \setminus 1$ (recall again that $Y = 1$, so there should be no ambiguity). Identify $S^* \equiv [k]$, fix $t \in \{0,1\}^k$, and consider any permutation $\pi : [k] \to [k]$. We have

   $$\begin{aligned} & \mathbb{P}((\widetilde{Z}_{\pi(1)}, \dots, \tilde{Z}_{\pi(k)}) = t) \\ = \;& \mathbb{P}((\widetilde{Z}_{\pi(1)} = t_1 \big| \widetilde{Z}_{\pi(2)}, \dots, \tilde{Z}_{\pi(k)}) = t_2, \dots, t_k)) \cdot \mathbb{P}(\widetilde{Z}_{\pi(2)}, \dots, \tilde{Z}_{\pi(k)}) = t_2, \dots, t_k) \end{aligned}$$

   Using our adjusted notation, the relation between between $\widetilde{Z}_i$'s becomes $\widetilde{Z}_1 = 1 \oplus_{i \in S^* - 1} \widetilde{Z}_i$. This constraint is deterministic (again, $Y = 1$) and can be rewritten as $\oplus_{i \in S^*} \widetilde{Z} = 1$, which is invariant under-relabeling. Hence, $\mathbb{P}(\widetilde{Z}_{\pi(1)} = t_1 \big| (\widetilde{Z}_{\pi(2)}, \dots, \tilde{Z}_{\pi(k)}) = (t_2, \dots, t_k)) = \mathbb{I}(\oplus_{i=1}^{k} t_i = 1)$. Moreover, we demonstrated above that, for any set S not containing S^*, $\{Z_i : i \in S \cap S^* - 1\} \cup \{\widetilde{Z}_1\}$ have a product distribution of $k-1$ Bernoulli(1/2) random variables. In our adjusted notation, this entails that $\mathbb{P}((\widetilde{Z}_{\pi(2)}, \dots, \tilde{Z}_{\pi(k)}) = t_2, \dots, t_k) = 2^{-(k-1)}$. Putting things together, we see that

   $$\mathbb{P}((\widetilde{Z}_{\pi(1)}, \dots, \tilde{Z}_{\pi(k)}) = t) = 2^{-(k-1)}\mathbb{I}(\oplus_i t_i = 1) \tag{127}$$

   which does not dependent on the permutation π.

Step 3: Computation of the Gap under ν

Note that if $S \neq S^*$ then

$$\begin{aligned}
\mathbb{E}_\nu[\max_{i\in S} X_i] = \mathbb{E}_\nu[\max_{i\in S} Z_i U_i] &= \mathbb{P}\left(\cup_{i\in S}\{Z_i = 1, U_i = 1\}\right) = 1 - \mathbb{P}_\nu(\cap_{i\in S}\{Z_i = 1, U_i = 1\}^c) \\
&= 1 - \prod_{i\in S} \mathbb{P}_\nu(\{Z_i = 1, U_i = 1\}^c) = 1 - \prod_{i\in S}(1 - \mathbb{P}_\nu(Z_i = 1, U_i = 1)) \\
&= 1 - \prod_{i\in S}(1 - \mathbb{P}_\nu(Z_i = 1)\mathbb{P}(U_i = 1)) = 1 - (1-\mu)^k.
\end{aligned}$$

Otherwise,

$$\begin{aligned}
\mathbb{E}_\nu[\max_{i\in S^*} X_i] &= \mathbb{E}_\nu[\max_{i\in S^*} X_i | Y = 1]\, p + \mathbb{E}_\nu[\max_{i\in S^*} X_i | Y = 0]\, (1-p) \\
&= \mathbb{E}_\nu[\max_{i\in S^*} X_i | Y = 1]\, p + \left[1 - (1-\mu)^k\right]\, (1-p)
\end{aligned}$$

where

$$\begin{aligned}
\mathbb{E}_\nu[\max_{i\in S^*} X_i | Y = 1] &= 1 - \mathbb{P}(\max_{i\geq 1} U_i Z_i = 0) \\
&= 1 - \mathbb{P}(\max_{i\geq 1} U_i Z_i = 0, \oplus_{i>1} Z_i = 0) - \mathbb{P}(\max_{i\geq 1} U_i Z_i = 0, \oplus_{i>1} Z_i = 1) \\
&= 1 - (1-2\mu)\mathbb{P}(\max_{i>1} U_i Z_i = 0, \oplus_{i>1} Z_i = 0) - \mathbb{P}(\max_{i>1} U_i Z_i = 0, \oplus_{i>1} Z_i = 1) \\
&= 1 - \mathbb{P}(\max_{i>1} U_i Z_i = 0) + 2\mu\mathbb{P}(\max_{i>1} U_i Z_i = 0, \oplus_{i>1} Z_i = 0) \\
&= 1 - (1-\mu)^{k-1} + 2\mu\mathbb{P}(\max_{i>1} U_i Z_i = 0, \oplus_{i>1} Z_i = 0)
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{P}(\max_{i>1} U_i Z_i = 0, \oplus_{i>1} Z_i = 0) &= \sum_{\ell=0}^{\lfloor \frac{k-1}{2} \rfloor} \mathbb{P}\left(\max_{i>1} U_i Z_i = 0, \sum_{i>1} Z_i = 2\ell\right) \\
&= \sum_{\ell=0}^{\lfloor \frac{k-1}{2} \rfloor} \mathbb{P}\left(\max_{i>1} U_i Z_i = 0 \middle| \sum_{i>1} Z_i = 2\ell\right) \binom{k-1}{2\ell} 2^{-k+1} \\
&= \sum_{\ell=0}^{\lfloor \frac{k-1}{2} \rfloor} (1-2\mu)^{2\ell} \binom{k-1}{2\ell} 2^{-k+1} \\
&= \frac{2^{-(k-1)}}{2}\left(((1-2\mu)+1)^{k-1} + (-1)^{k-1}((1-2\mu)-1)^{k-1}\right) \\
&= \frac{1}{2}\left((1-\mu)^{k-1} + \mu^{k-1}\right)
\end{aligned}$$

since

$$\begin{aligned}
((1-2\mu)+1)^{k-1} &= \sum_{j=0}^{k-1}(1-2\mu)^j\binom{k-1}{j} \\
&= \sum_{\ell=0}^{\lfloor\frac{k-1}{2}\rfloor}(1-2\mu)^{2\ell}\binom{k-1}{2\ell} + \sum_{\ell=0}^{\lfloor\frac{k-1}{2}\rfloor}(1-2\mu)^{2\ell+1}\binom{k-1}{2\ell+1}
\end{aligned}$$

and

$$\begin{aligned}
((1-2\mu)-1)^{k-1} &= \sum_{j=0}^{k-1}(-1)^j(1-2\mu)^{k-1-j}\binom{k-1}{j} \\
&= \sum_{\ell=0}^{\lfloor\frac{k-1}{2}\rfloor}(1-2\mu)^{k-1-2\ell}\binom{k-1}{2\ell} - \sum_{\ell=0}^{\lfloor\frac{k-1}{2}\rfloor}(1-2\mu)^{k-2-2\ell}\binom{k-1}{2\ell+1} \\
&= (-1)^{k-1}\sum_{\ell=0}^{\lfloor\frac{k-1}{2}\rfloor}(1-2\mu)^{2\ell}\binom{k-1}{2\ell} - (-1)^{k-1}\sum_{\ell=0}^{\lfloor\frac{k-1}{2}\rfloor}(1-2\mu)^{2\ell+1}\binom{k-1}{2\ell+1}.
\end{aligned}$$

Putting it all together we have

$$\begin{aligned}
\mathbb{E}_\nu[\max_{i\in S^*} X_i] &= [1-(1-\mu)^{k-1}+\mu\left((1-\mu)^{k-1}+\mu^{k-1}\right)]\,p + \left[1-(1-\mu)^k\right]\,(1-p) \\
&= [1-(1-\mu)^k+\mu^k]\,p + \left[1-(1-\mu)^k\right]\,(1-p) \qquad (128) \\
&= \left[1-(1-\mu)^k\right] + \mu^k p
\end{aligned}$$

Thus, $\Delta = p\mu^k$ which is maximized at $\mu = \frac{1}{2}$ achieving $\Delta = p2^{-k}$.

Step 4: Change of measure: Consider the distribution ν that is constructed in Step 1 that is defined with respect to $S^* = \{1,\dots,k\}$. For all $S \in \mathcal{S}$ we will now construct a new distribution ν^S such that $\mathbb{E}_{\nu^S}[\max_{i\in S} X_i] > \mathbb{E}_{\nu^S}[\max_{i\in S^*} X_i] = \mathbb{E}_\nu[\max_{i\in S^*} X_i]$. We begin constructing ν^S identically to how we constructed ν but modify the distribution of X_{S^ℓ} where $S^\ell = \arg\min\{i : X_i, i \in S\}$. In essence X_{S^ℓ} with respect to $S \in \mathcal{S}$ will be constructed identically to the construction of X_1 with respect to $S^* = \{1,\dots,k\}$ with the one exception that in place of Y we will use a new random variable Y^S that is Bernoulli(p') where $p' > p$ (this is always possible as $p < 1$).

Let $\nu(S)$ describe the joint probability distribution of ν restricted to the set $i \in S$. And for any $S \in \mathcal{S}$ let τ denote the projection of $\nu(S)$ down to some smaller event space. For example, $\tau\nu(S)$ can represent the Bernoulli probability distribution describing $\max_{i\in S} X_i$ under distribution ν. By $(k-1)$-wise independence we have

$$KL(\nu(S')|\nu^S(S')) = 0 \quad \forall S' \in \mathcal{S}\setminus S$$

since S and S' differ by at least one element and $\nu(S^*) = \nu^S(S^*)$. Clearly, $KL(\tau\nu(S')|\tau\nu^S(S')) = 0$ as well for all $S' \in \mathcal{S}\setminus S$. By assumption, any valid algorithm correctly identifies S^* under ν, and S under ν^S, with probability at least $1-\delta$. Thus, by Lemma 1 of Kaufmann et al. (2015), for every $S \in \mathcal{S} \setminus S^*$

$$\log(\tfrac{1}{2\delta}) \leq \sum_{S' \in \mathcal{S}} \mathbb{E}_\nu[T_{S'}] KL(\tau\nu(S') | \tau\nu^S(S')) = KL(\tau\nu(S)|\tau\nu^S(S))\mathbb{E}_\nu[T_S] \, ,$$

where we recall that T_S is the number of times the set S is pulled. Hence,

$$\begin{aligned}\mathbb{E}_\nu\left[\sum_{S \in \mathcal{S}\setminus S^*} T_S\right] &\geq \sum_{S \in \mathcal{S}\setminus S^*} \frac{\log(\frac{1}{2\delta})}{KL(\tau\nu(S)|\tau\nu^S(S))} \\ &= \frac{\log(\frac{1}{2\delta})}{KL(\tau\nu(S)|\tau\nu^S(S))}\left[\binom{n}{k}-1\right] \geq \frac{\frac{2}{3}\log(\frac{1}{2\delta})}{KL(\tau\nu(S)|\tau\nu^S(S))}\binom{n}{k}\end{aligned}$$

where the equality holds for any fixed $S \in \mathcal{S}$ by the symmetry of the construction and the last inequality holds since $2 \leq k < n$, $\binom{n}{k} - 1 \geq \frac{2}{3}\binom{n}{k}$. It just remains to upper bound the KL divergence.

Bandit feedback: Let $\tau\nu(S)$ represent the Bernoulli probability distribution describing $\max_{i\in a} X_i$ under distribution ν. Then by the above calculations of the gap we have

$$\begin{aligned}KL(\tau\nu(S)|\tau\nu^S(S)) &= KL(1-(1-\mu)^k | 1-(1-\mu)^k + p'\mu^k) \\ &\leq \frac{p'^2\mu^{2k}/2}{(1-(1-\mu)^k)(1-\mu)^k - 2p'\mu^k[\frac{1}{2}-(1-\mu)^k]_+} \\ &\leq \frac{p'^2\mu^{2k}/2}{(1-(1-\mu)^k)((1-\mu)^k - p'\mu^k)}\end{aligned}$$

by applying Lemma E.1 and noting that

$$\begin{aligned}&(1-(1-\mu)^k)(1-\mu)^k - 2p'\mu^k[\tfrac{1}{2}-(1-\mu)^k]_+ \\ &\qquad \geq \min\{(1-(1-\mu)^k)(1-\mu)^k, (1-(1-\mu)^k)(1-\mu)^k - p'\mu^k(1-2(1-\mu)^k)\} \\ &\qquad \geq \min\{(1-(1-\mu)^k)(1-\mu)^k, (1-(1-\mu)^k)[(1-\mu)^k - p'\mu^k]\} \\ &\qquad \geq (1-(1-\mu)^k)((1-\mu)^k - p'\mu^k)\end{aligned}$$

Finally, let $p' \to p$. Setting $\mu = 1-2^{-1/k} \geq \frac{1}{2k}$ we have $(1-\mu)^k = 1/2$ and $\Delta \geq p(2k)^{-k}$ so that $\mathbb{E}_\nu\left[\sum_{S\in\mathcal{S}\setminus S^*} T_S\right] \geq \frac{1}{3}\binom{n}{k}\Delta^{-2}\log(\frac{1}{2\delta})$.

Marked-Bandit feedback: Let $\tau\nu(S)$ represent the distribution over $\perp \cup S$ under ν such that if $W \sim \tau\nu(S)$ then W is drawn uniformly at random from $\arg\max_{i\in S} X_i$ if $\max_{i\in S} X_i = 1$, and $W = \perp$ otherwise. By the permutation invariance property of ν described in Step 2, we have for any $S \in \binom{[n]}{k} - S_*$ and $i \in S$

$$\mathbb{P}_\nu(W = i | W \neq \perp) = \mathbb{P}_{\nu^S}(W = i | W \neq \perp) = \frac{1}{k}$$

so that

$$\begin{aligned}
KL(\tau\nu(S)|\tau\nu^S(S)) &= \sum_{w \in \perp \cup S} \mathbb{P}_\nu(W = w)\log(\frac{\mathbb{P}_\nu(W = w)}{\mathbb{P}_{\nu^S}(W = w)}) \\
&= \mathbb{P}_\nu(W = \perp)\log(\frac{\mathbb{P}_\nu(W = \perp)}{\mathbb{P}_{\nu^S}(W = \perp)}) + \sum_{i \in S}\frac{1}{k}\mathbb{P}_\nu(W \neq \perp)\log(\frac{\mathbb{P}_\nu(W \neq \perp)}{\mathbb{P}_{\nu^S}(W \neq \perp)}) \\
&= KL\big(\mathbb{P}_\nu(\max_{i \in S} X_i = 1)\big|\mathbb{P}_{\nu^S}(\max_{i \in S} X_i = 1)\big).
\end{aligned}$$

Thus, KL divergence for marked-bandit feedback is equal to that of simple bandit feedback.
Semi-Bandit feedback:

Let P denote the law of the entire construction for independent distribution, and Q the law of the construction for the distribution. The strategy is to upper bound the KL of X, together with the additional information from the hidden variables $Z_2, \dots, Z_k$. In this section, given $v \in \{0,1\}^k$, we use the compact notation $v^{(2;k)}$ to denote the vector $v_2, \dots, v_k$. We can upper bound the KL by

$$\begin{aligned}
KL(p(X), Q(X)) &\leq KL(P(X, Z^{(2;k)}), Q(X, Z^{(2;k)})) \\
&= \sum_{x \in \{0,1\}^k, z^{(2;k)} \in \{0,1\}^{k-1}} P\left(X = x, Z^{(2;k)} = z^{(2;k)}\right)\log\left(\frac{P(X = x, Z^{(2;k)} = z^{(2;k)})}{Q(X = x, Z^{(2;k)} = z^{(2;k)})}\right)
\end{aligned}$$

By the law of total probability, the above is just

$$\begin{aligned}
&\sum_{x^{(2;k)} \in \{0,1\}^{2;k}, z^{(2;k)} \in \{0,1\}^{k-1}} P\left(X^{(2;k)} = x^{(2;k)}, Z^{(2;k)} = z^{(2;k)}\right) \\
\times &\left(\sum_{x_1 \in \{0,1\}} P\left(X_1 = x_1\big|X^{(2;k)} = x^{(2;k)}, Z^{(2;k)} = z^{(2;k)}\right)\log\left(\frac{P(X = x, Z^{(2;k)} = z^{(2;k)})}{Q(X = x, Z^{(2;k)} = z^{(2;k)})}\right)\right)
\end{aligned}$$

Again, by the law of total probability, we have

$$\begin{aligned}
&\frac{P(X = x, Z^{(2;k)} = z^{(2;k)})}{Q(X = x, Z^{(2;k)} = z^{(2;k)}} \\
= \quad &\frac{P(X_1 = x_1\big|X^{(2;k)} = x^{(2;k)}, Z^{(2;k)} = z^{(2;k)})}{Q(X_1 = x_1\big|X^{(2;k)} = x^{(2;k)}, Z^{(2;k)} = z^{(2;k)}))} \times \frac{P(X^{(2;k)} = x^{(2;k)}, Z^{(2;k)} = z^{(2;k)})}{Q(X^{(2;k)} = x^{(2;k)}, Z^{(2;k)} = z^{(2;k)}))}
\end{aligned}$$

Under our construction, $(X_2, \dots, X_k, Z_2, \dots, Z_k)$ have the same joint distribution under either P or Q, so the second multiplicand in the second line in the above display is just 1. Under the law P, X_1 is independent of $X_2, \dots, X_k, Z_2, \dots, Z_k$, so $P(X_1 = x_1\big|X^{(2;k)} = x^{(2;k)}, Z^{(2;k)} = z^{(2;k)}) = P(X_1 = x_1)$. Under the dependent law Q, X_1 only depends on $X_2, \dots, X_k, Z_2, \dots, Z_k$ through $W(Z^{(2;k)}) := 1 \oplus_{i=2}^k Z_i \in \{0,1\}$. Hence, if we define the conditional KL's:

$$\mathrm{KL}_1 := KL\left(P(X_1), Q(X_1)\big|W(z^{(2;k)}) = 1\right) = \sum_{x_1 \in \{0,1\}} p(X_1 = x_1)\log\left(\frac{P(X_1 = x_1)}{Q(X_1 = x_1\big|W(z^{(2;k)}) = 1)}\right)$$

and define $\mathrm{KL}_0 := KL\left(P(X_1), Q(X_1) \middle| W(z^{(2;k)}) = 0\right)$ analogously, then

$$\begin{aligned}
&\sum_{x_1 \in \{0,1\}} P(X_1 = x_1 \big| X^{(2;k)} = x^{(2;k)}, Z^{(2;k)} = z^{(2;k)}) \log(\frac{P(X = x, Z^{(2;k)} = z^{(2;k)})}{Q(X = x, Z^{(2;k)} = z^{(2;k)})}) \\
= \;& \mathbb{I}\left(W(z^{(2;k)}) = 1\right) \mathrm{KL}_1 + \mathbb{I}\left(W(z^{(2;k)}) = 0\right) \mathrm{KL}_0
\end{aligned}$$

Putting these pieces together,

$$\begin{aligned}
KL(P(X, Z^{(2;k)}), Q(X, Z^{(2lk)})) &= \sum_{(x^{(2;k)}, z^{(2;k)} \in \{0,1\}^{2(k-1)}} \mathbb{I}(W(Z^{(2;k)}) = 1)\mathrm{KL}_1 + \mathbb{I}(W(Z^{(2;k)}) = 0)\mathrm{KL}_0 \\
&= \mathbb{P}(W(Z^{(2;k)}) = 1)\mathrm{KL}_1 + \mathbb{P}(W(Z^{(2;k)}) = 0)\mathrm{KL}_0 \\
&= \frac{1}{2}(\mathrm{KL}_1 + \mathrm{KL}_0)
\end{aligned}$$

where the last line follows the parity $W(Z^{(2;k)})$ is Bernoulli $1/2$. A straightforward computation bounds KL_1 and KL_0.

Claim E.2 (Bound on $\mathrm{KL}_1, \mathrm{KL}_0$) *Let* KL_0 *and* KL_1 *be defined as above. Then* $\mathrm{KL}_0 \le \frac{p^2\mu/2}{(1-p)(1-\mu(1-p))}$ *and* $\mathrm{KL}_1 \le \frac{p^2\mu/2}{1-\mu(1+p)}$.

Proof Note $P(X_1 = 1) = \mu$,

$$\begin{aligned}
Q(X_1 = 1 \big| W(z^{(2;k)}) = 0) &= Q(X_1 = 1 \big| W(z^{(2;k)}) = 1, Y = 1)p + Q(X_1 = 1 \big| W(z^{(2;k)}) = 1, Y = 0)(1-p) \\
&= 0 \cdot p + \mu(1-p) = \mu(1-p)
\end{aligned}$$

and

$$\begin{aligned}
Q(X_1 = 1 \big| W(z^{(2;k)}) = 1) &= Q(X_1 = 1 \big| W(z^{(2;k)}) = 1, Y = 1)p + Q(X_1 = 1 \big| W(z^{(2;k)}) = 1, Y = 0)(1-p) \\
&= 2\mu p + \mu(1-p) = \mu(1+p).
\end{aligned}$$

Thus, by Lemma E.1 we have

$$\begin{aligned}
\mathrm{KL}_0 &= \sum_{x_1 \in \{0,1\}} P(X_1 = x_1) \log\left(\frac{P(X_1 = x_1)}{Q(X_1 = x_1 \big| W(z^{(2;k)}) = 0)}\right) \\
&= d(\mu, \mu(1-p)) \le \frac{(p\mu)^2/2}{\mu(1-p)(1-\mu(1-p))} = \frac{p^2\mu/2}{(1-p)(1-\mu(1-p))}.
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{KL}_1 &= \sum_{x_1 \in \{0,1\}} P(X_1 = x_1) \log\left(\frac{P(X_1 = x_1)}{Q(X_1 = x_1 \big| W(z^{(2;k)}) = 1)}\right) \\
&= d(\mu, \mu(1+p)) \le \frac{(p\mu)^2/2}{\min\{\mu(1-\mu), \mu(1+p)(1-\mu(1+p))\}} \le \frac{p^2\mu/2}{1-\mu(1+p)}.
\end{aligned}$$

■

Remark E.1 *Despite our seemingly arbitrary construction of random variables in Theorem 2.1 to produce the resulting measure ν, Theorem 2.2 states that the joint distribution is unique and would be arrived at using any other construction that satisfied the same properties.*

Remark E.2 (An Upper Bound When $\mu = 1/2$) *Suppose that $\mu = 1/2$. Then, our construction implies $Z_i = X_i$ for $i \geq 2$, and thus our bound on the KL is exact. In fact, we can use a simple parity estimator $W(S) = \oplus_{i \in S} X_i$ to distinguish between a subset S of correlated and uncorrelated arms. When S is an independent set, $W(S) \sim$ Bernoulli$(1/2)$. However, a simple computation reveals that $W(S^*) \sim$ Bernoulli$(1/2 + p/2)$. Thus, using a parity estimator reduces our problem to finding one coin with bias $p/2$ in a bag of $\binom{n}{k}$ unbiased coins, whose difficulty exactly matches our problem*

Surprisingly, Theorem 2.2 tells us that the construction outlined in this lower bound is the ***unique*** *construction which yields $k-1$-wise independent marginals of mean $\mu = 1/2$, with gap $p2^{-k}$; in other words, in any $k-1$-wise independent construction with $\mu = 1/2$, the parity estimator is optimal.*

E.1. Proof of Lemma E.1

Proof [Proof of Lemma E.1] If $f(z) = d(z,y)$ then $f'(z) = \log(\frac{z}{1-z}) - \log(\frac{y}{1-y})$, and $f''(z) = \frac{1}{z(1-z)}$ so

$$2(y-x)^2 \leq \frac{(y-x)^2/2}{\sup_{z \in [x,y]} z(1-z)} \leq d(x,y) \leq \frac{(y-x)^2/2}{\inf_{z \in [x,y]} z(1-z)}.$$

If $\epsilon = y - x$ then

$$\inf_{z \in [x,y]} z(1-z) = \inf_{\epsilon \in [0, y-x]} x(1-x) + \epsilon(1-2x) = x(1-x) - [(y-x)(2x-1)]_+$$

■

Appendix F. Proof of Lower Bound Converse

To prove the above proposition, we need a convenient way of describing all feasible probability distributions over $\{0,1\}^k$ which are specified on their $k-1$ marginals. To this end, we introduce the following notation: We shall find it convenient to index the entries of vectors $w \in \mathbb{R}^{k-1}$ by binary strings $t \in \{0,1\}^{k-1}$. At times, we shall need to "insert" indices into strings of length $k-2$, as follows: For $u \in \{0,1\}^{k-2}$ and $j \in [k-1]$, denote by $u \oplus_j 0$ the string in $\{0,1\}^{k-1}$ obtained by inserting a 0 in the j-th position of u. We define $u \oplus_j 1$ similarly.

Lemma F.1 *Let $\mathbb{P}_0$ be any distribution over $\{0,1\}^k$. Then, a probability distribution $\mathbb{P}$ agrees with $\mathbb{P}_0$ on their $k-1$ marginals if and only if, for all binary strings $t \in \{0,1\}^{k-1}$, $\mathbb{P}$ is given by*

$$\mathbb{P}(X_{-k} = t, X_k = 0) = w(t) \tag{129}$$

where $w \in \mathbb{R}^{2^{k-1}}$ satisfies the following linear constraints:

$$\begin{array}{rl} \forall t \in \{0,1\}^{k-1}: & 0 \le w(t) \le \mathbb{P}_0(X_{-k} = t) \\ \forall j \in [k-1], u \in \{0,1\}^{k-2} & w(u \oplus_j 0) + w(u \oplus_j 1) = \mathbb{P}_0(X_{-\{j,k\}} = u_{-j}, X_k = 0) \end{array}$$

Remark F.1 *Note that the above lemma makes no assumptions about $k-1$ independence, only that the $k-1$ marginals are constrained*

Proof [Proof of Theorem 2.2] Let $\mathbb{P}_0$ denote the product measure on $X_1, \dots, X_k$, and $\mathbb{P}$ denote our coupled distribution. Fix $\mu \in [0,1]$. For $p \in \{0,1,\dots,k-1\}$, define the probability mass function

$$\psi(p) := \mu^p (1-\mu)^{k-1-p} \tag{130}$$

Further, for u and t in $\{0,1\}^{k-2}$ and $\{0,1\}^{k-1}$, respectively, define the hamming weights $H(t) = \sum_i t_i$ and $H(u) = \sum_i u_i$.

Since our distribution is $k-1$ wise independent, and each entry X_i has mean μ, we have $\mathbb{P}(X_{-k} = t) = \mathbb{P}_0(X_{-k} = t) = \psi(H(t))$. Moreover,

$$\begin{aligned} \mathbb{P}_0(X_{-\{j,k\}} = u_{-j}, X_k = 0) &= (1-\mu)\mathbb{P}_0(X_{-\{j,k\}} = u_{-j}) \\ &= (1-\mu)\mu^{H(u)}(1-\mu)^{k-2-H(u)} \\ &= \mu^{H(u)}(1-\mu)^{k-1-H(u)} = \psi(H(u)) \end{aligned}$$

Thus, our feasibility set is precisely

$$\begin{array}{rr} \forall t \in \{0,1\}^{k-1}: & 0 \le w(t) \le \psi(H(t)) \\ \forall j \in [k-1], u \in \{0,1\}^{k-2} & w(u \oplus_j 0) + w(u \oplus_j 1) = \psi(H(u)) \end{array} \tag{131}$$

The equality constraints show there is only one degree of freedom, which we encode into $w(\mathbf{0})$:

Claim F.2 *w satisfies the equality constraints of the LP if and only if, for all $t \in \{0,1\}^{k-1}$ of weight $H(t) = p$,*

$$w(t) = (-1)^p \, w(\mathbf{0}) + (-1)^{p-1} \, \Phi(p) \tag{132}$$

where $\Phi(p) = \sum_{i=0}^{p-1} (-1)^i \psi(i)$, so that $\Phi(0) = 0$. Note that Φ satisfies the identity

$$\Phi(p) = (-1)^{p-1} \, \psi(p-1) + \Phi(p-1) \tag{133}$$

Hence, we can replace the equality constraints by the explicit definitions of $w(t)$ in terms of $w(\mathbf{0})$ and $\Phi(p)$. This leads to the next claim:

Claim F.3 *w is feasible precisely when*

$$\max_{0 \le p \le k \text{ even}} \Phi(p) \le w(\mathbf{0}) \le \min_{1 \le p \le k \text{ odd}} \Phi(p) \tag{134}$$

We now establish a closed form solution for $\Phi(p)$ when $\mu < 1/2$, and parity-wise monotonicity when $\mu \ge 1/2$:

Claim F.4 *If $\mu < 1/2$, we have $\Phi(p) = (1-\mu)^k \left(1 - (\frac{-\mu}{1-\mu})^p\right)$, so $\Phi(p)$ is decreasing for odd p and increasing for even p. If $\mu \geq 1/2$, $\Phi(p)$ is nondecreasing for odd p and nonincreasing for even p*

To conclude, we note that when $\mu \geq 1/2$, the fact that $\Phi(p)$ is nondecreasing for odd p and nonincreasing for even p implies that

$$\begin{aligned} \max_{0 \leq p \leq k \text{ even}} \Phi(p) \leq w(\mathbf{0}) \leq \min_{1 \leq p \leq k \text{ odd}} \Phi(p) \quad &\iff \quad \Phi(0) \leq w(\mathbf{0}) \leq \Phi(1) \\ &\iff \quad 0 \leq w(\mathbf{0}) \leq \psi(0) \\ &\iff \quad 0 \leq w(\mathbf{0}) \leq (1-\mu)^{k-1} \end{aligned}$$

When $\mu < 1/2$, the fact that $\Phi(p)$ is decreasing for odd p and increasing for even p implies that

$$\begin{aligned} & \max_{0 \leq p \leq k \text{ even}} \Phi(p) \leq w(\mathbf{0}) \leq \min_{1 \leq p \leq k \text{ odd}} \Phi(p) \iff \Phi(k_{odd}) \leq w(\mathbf{0}) \leq \Phi(k_{even}) \\ \iff \quad & (1-\mu)^k \left(1 - \left(\frac{\mu}{1-\mu}\right)^{k_{even}}\right) \leq w(\mathbf{0}) \leq (1-\mu)^k \left(1 + \left(\frac{\mu}{1-\mu}\right)^{k_{odd}}\right) \end{aligned}$$

Since $w(\mathbf{0}) = \mathbb{P}(X_1, \dots, X_k = 0)$, we are done.

■

F.1. Proofs

Proof [Proof Of Lemma F.1] We can consider the joint distribution of $(X_1, \dots, X_k)$ as a vector in the 2^k simplex. However, there are many constraints: in particular, the joint distribution of $X_1, \dots, X_{k-1}$ is entirely determined by the $k-1$-marginals of the distribution. In fact, if $\mathbb{P}$ is a distribution over $\{0,1\}^k$, then it must satisfy

$$\mathbb{P}(X_{-k} = t_{-k}, X_k = 1) + \mathbb{P}(X_{-k} = t_{-k}, X_k = 0) \quad = \quad \mathbb{P}(X_{-k} = t_{-k}).$$

Hence, without any loss of generality, we may encode any arbitrary probability distribution on $\{0,1\}^k$ by

$$\mathbb{P}(X = t) := \begin{cases} w(t_{-k}) & t_k = 0 \\ \mathbb{P}_0(X_{-k} = t_{-k}) - w(t_{-k}) & t_k = 1 \end{cases} \tag{135}$$

for a suitable $w \in \mathbb{R}^{2^{k-1}}$. This defines $\mathbb{P}$ on the atomic events $\{X = t\}$, and we extend $\mathbb{P}$ to all further events by additivity. We now show that the constraints on the Lemma hold if and only if w induces a proper probability distribution $\mathbb{P}$ whose $k-1$ marginals coincide with $\mathbb{P}$.

Recall that $\mathbb{P}$ is a proper distribution if and only if it is nonnegative, normalized to one, monotonic, and additive[4]. $\mathbb{P}$ satisfies additivity by construction. Moreover, by definition $\sum_{t \in \{0,1\}^k} \mathbb{P}(X = t) = \sum_{t_{-k} \in \{0,1\}^{k-1}} \mathbb{P}_0(X_{-k} = t_{-k}) = 1$, so $\mathbb{P}$ is normalized. Finally, monotonicity will follow

[4] As X has finite support, we don't need to worry about such technical conditions as σ-additivity

as long as we establish non-negativity of $\mathbb{P}$ on the atomic events $\{X = t\}$. But the constraint that $\mathbb{P}(X = t)$ is nonnegative holds if and only if

$$0 \le w(t_1, \dots, t_{k-1}) \le \mathbb{P}_0(X_{-k} = t_{-k}). \tag{136}$$

On the other hand, the constraint that $\mathbb{P}$'s $k-1$ marginals coincide with $\mathbb{P}_0$ is simply that

$$\begin{aligned} & w(t_1, \dots, t_{j-1}, 0, t_{j+1}, \dots, t_{k-1}) + w(t_1, \dots, t_{j-1}, 1, t_{j+1}, \dots, t_{k-1}) \\ = \;& \mathbb{P}_0(X_1 = t_1, \dots, X_{j-1} = t_{j-1}, X_{j+1} = t_{j+1}, \dots, X_{k-1} = t_{k-1}, X_k = 0) \end{aligned}$$

which can be expressed more succinctly using the concatenation notation $w(u \oplus_j 0) + w(u \oplus_j 1) = \mathbb{P}_0(X_{-\{j,k\}} = u_{-j}, X_k = 0)$. ■

Proof [Proof of Claim F.2] First, we prove "only if" by induction on $H(t)$. For $H(t) = 0$, the claim holds since $\Phi(0) = 0$. For a general $t \in \{0,1\}^{k-1}$ such that $H(t) = p \ge 1$, we can construct a sequence $t_0, \dots, t_p \in \{0,1\}^{k-1}$ such that $t_0 = 0$, $t_p = t$, and each string t_s is obtained by "flipping on" a zero in the string t_{s-1} to 1, that is, there is a string $u_s \in \{0,1\}^{k-2}$ such that $t_s = u_s \oplus_{j_s} 1$ and $t_{s-1} = u_s \oplus_{j_s} 0$. Thus, our equality constraints imply that

$$\begin{aligned} w(t_{p-1}) + w(t) &= w(t_{p-1}) + w(t_p) \\ &= w(u_s \oplus_{j_s} 1) + w(u_s \oplus_{j_s} 0) \\ &= \psi(H(u)) = \psi(p-1). \end{aligned}$$

Hence, we get the recursion $w(t_p) = \psi(p-1) - w(t_{p-1})$, which by the inductive hypothesis on t_{p-1} and Equation 133 imply that

$$\begin{aligned} w(t) &= \psi(p-1) - \big((-1)^{p-1} w(\mathbf{0}) + (-1)^{p-2}\Phi(p-1)\big) \\ &= (-1)^p w(\mathbf{0}) + (-1)^{p-1}\Phi(p-1) + \psi(p-1) \\ &= (-1)^p w(\mathbf{0}) + (-1)^{p-1}(\Phi(p-1) + (-1)^{p-1}\psi(p-1)) \\ &= (-1)^p w(\mathbf{0}) + (-1)^{p-1}\Phi(p) \end{aligned}$$

as needed. Next, we prove the "if" direction. Let $u \in \{0,1\}^{k-2}$ have weight p. Then

$$\begin{aligned} w(u \oplus 0) + w(u \oplus 1) &= (-1)^p w(\mathbf{0}) + (-1)^{p-1}\Phi(p) + (-1)^{p+1} w(\mathbf{0}) + (-1)^p \Phi(p+1) \\ &= (-1)^{p-1}\Phi(p) + (-1)^p\Phi(p+1) \\ &= (-1)^{p-1}\Phi(p) + (-1)^p \left(\Phi(p+1-1) + (-1)^p \psi(p+1-1)\right) \\ &= \left((-1)^{p-1} + (-1)^p\right)\Phi(p) + (-1)^{2p}\psi(p) = \psi(p) \end{aligned}$$

as needed. ■

Proof [Proof of Claim F.3] Our feasibility set is precisely is the set of $w(\mathbf{0})$ such that $0 \le w(\mathbf{0}) \le \psi(0)$, and for all $p \in \{1, 2, \dots, k-1\}$

$$0 \le (-1)^p w(\mathbf{0}) + (-1)^{p-1}\Phi(p) \le \psi(p). \tag{137}$$

Suppose first that p is even. If p is greater than 1, then the above constraint together with Claim F.2 imply

$$\Phi(p) \le w(\mathbf{0}) \le \psi(p) + \Phi(p) = (-1)^{(p+1)-1}\psi(p) + \Phi(p) = \Phi(p+1).$$

If p is 0, then $\Phi(0) = 0$ and $\Phi(1) = \psi(0)$, so the constraint $0 \le w(\mathbf{0}) \le \psi(0)$ is equivalent to $\Phi(p) \le w(\mathbf{0}) \le \Phi(p+1)$ for $p = 0$.

On the other hand, when p is odd, we have $w(\mathbf{0}) \le \Phi(p)$, whilst

$$w(\mathbf{0}) \ge \Phi(p) - \psi(p) = \Phi(p) + (-1)^{(p+1)-1}\psi(p) = \Phi(p+1). \tag{138}$$

In other words, $w(\mathbf{0}) \le \Phi(p)$ for all p which are either odd and between 1 and $k-1$, or p of the form $p = q+1$ where q is even and between 1 and $k-1$. This is precisely the set of all odd p in $1, \dots, k$. By the same token, $w(\mathbf{0}) \ge \Phi(p)$ for all even p in $\{1, \dots, k\}$. Taking the intersection of these lower and upper bounds on $w(\mathbf{0})$ yields

$$\max_{0 \le p \le k \text{ even}} \Phi(p) \le w(\mathbf{0}) \le \min_{1 \le p \le k \text{ odd}} \Phi(p). \tag{139}$$

■

Proof [Proof of Claim F.4] Let $\rho = \frac{\mu}{1-\mu}$. We can write Φ yields as geometric series

$$\begin{aligned}
\Phi(p) &= \sum_{i=0}^{p-1}(-1)^i\psi(i) \\
&= \sum_{i=0}^{p-1}(-1)^i \cdot \mu^i(1-\mu)^{k-1-i} \\
&= (1-\mu)^{k-1}\sum_{i=0}^{p-1}(-1)^i(\frac{\mu}{1-\mu})^i \\
&= (1-\mu)^{k-1}\sum_{i=0}^{p-1}(-\rho)^i
\end{aligned}$$

When $\mu \ge 1/2$, $\rho \ge 1$, and thus this series is nondecreasing for odd p and nonincreasing for even p. When $\rho < 1/2$, the series is decreasing for odd p and increasing for even p and in fact we have

$$\begin{aligned}
\Phi(p) &= (1-\mu)^{k-1}\frac{1-(-\rho)^p}{1+\rho} \\
\Phi(p) &= (1-\mu)^{k-1}\frac{1-(-\frac{\mu}{1-\mu})^p}{1+\frac{\mu}{1-\mu}} \\
&= (1-\mu)^k\left(1-(\frac{-\mu}{1-\mu})^p\right)
\end{aligned}$$

■

Appendix G. Proof of Theorem 4.1: Lower Bound for Independent Arms

As in the proof of Theorem 2.1, let $\nu(a)$ describe the joint probability distribution of ν restricted to the set $i \in a$. Note that $\nu(a) = \prod_{i \in a} \nu_i$. And for any $a \in A$ let $\tau\nu(a)$ represent the Bernoulli probability distribution describing $\max_{i \in a} X_i$ under distribution ν. Let $\epsilon > 0$. For each $j \in [n]$ let ν^j be a product distirbution of Bernoullis fully defined by its marginals $\mu_i^j := \mathbb{E}_{\nu_i^j}[X_i]$ and

$$\mu_i^j = \begin{cases} \mu_k + \epsilon & \text{if } i = j \text{ and } i > k \\ \mu_{k+1} - \epsilon & \text{if } i = j \text{ and } i \leq k \\ \mu_i & \text{if } i \neq j. \end{cases}$$

By Lemma 1 of Kaufmann et al. (2015), for every $j \in [n]$

$$\sum_{a \in \binom{[n]}{p}} \mathbb{E}_\nu[T_a] KL(\tau\nu(a) | \tau\nu^j(a)) \geq \log(\tfrac{1}{2\delta}),$$

for arbitrarily small ϵ, so in what follows let $\epsilon = 0$. Then

$$KL(\tau\nu(a)|\tau\nu^j(a)) = \begin{cases} 0 & \text{if } j \notin a \\ d\left((1-\mu_j)\prod_{i \in a \setminus j}(1-\mu_i) | (1-\mu_j-\Delta_j)\prod_{i \in a \setminus j}(1-\mu_i)\right) & \text{if } j \in a \text{ and } j > k \\ d\left((1-\mu_j)\prod_{i \in a \setminus j}(1-\mu_i) | (1-\mu_j+\Delta_j)\prod_{i \in a \setminus j}(1-\mu_i)\right) & \text{if } j \in a \text{ and } j \leq k \end{cases}$$

where for $j > k$, by invoking Lemma E.1,

$$\begin{aligned} & d\left((1-\mu_j)\prod_{i \in a \setminus j}(1-\mu_i) | (1-\mu_j-\Delta_j)\prod_{i \in a \setminus j}(1-\mu_i)\right) \\ & \leq \frac{\Delta_j^2 \left(\prod_{i \in a \setminus j}(1-\mu_i)\right)^2}{2\left(1-(1-\mu_j)\prod_{i \in a \setminus j}(1-\mu_i)\right)\left((1-\mu_j-\Delta_j)\prod_{i \in a \setminus j}(1-\mu_i)\right)} \\ & \leq \frac{\Delta_j^2 \left(\prod_{i \in a \setminus j}(1-\mu_i)\right)}{2\left(1-(1-\mu_j)\prod_{i \in a \setminus j}(1-\mu_i)\right)(1-\mu_j-\Delta_j)} \end{aligned}$$

and a similar bounds holds for $j \leq k$. If $h_j = \max_{a \in \binom{[n]-j}{p-1}} \prod_{i \in a}(1-\mu_i)$ and

$$\tau_j = \begin{cases} \frac{(1-\mu_j-\Delta_j)}{\Delta_j^2}\frac{1-(1-\mu_j)h_j}{h_j} & \text{if } j > k \\ \frac{(1-\mu_j)}{\Delta_j^2}\frac{1-(1-\mu_j+\Delta_j)h_j}{h_j} & \text{if } j \leq k \end{cases}$$

$\forall j \in [n]$ then

$$\sum_{a \in \binom{[n]}{p} : j \in a} \mathbb{E}_\nu[T_a] \geq 2\tau_j \log(\tfrac{1}{2\delta}) \tag{140}$$

or, in words, arm j must be included in a number of bandit observations that is at least the right-hand-side of (140). Note that $\sum_{a\in\binom{[n]}{p}}\mathbb{E}_{\nu}[T_a] \ge \max\left\{\max_{j=1,\dots,n}\sum_{a\in\binom{[n]}{p}:j\in a}\mathbb{E}_{\nu}[T_a], \frac{1}{p}\sum_{j=1}^{n}\sum_{a\in\binom{[n]}{p}:j\in a}\mathbb{E}_{\nu}[T_a]\right\}$ where the first argument of the $\max$ follows from the fact that the number of rounds must exceed the number of bandit evaluations each arm must be included in, and the second term sums over all arms j and accounts for the fact that each p-set a appears p times. We conclude that

$$\sum_{a\in\binom{[n]}{p}}\mathbb{E}_{\nu}[T_a] \ge 2\log(\tfrac{1}{2\delta})\max\left\{\max_{j=1,\dots,n}\tau_j, \frac{1}{p}\sum_{j=1}^{n}\tau_j\right\} \ge \log(\tfrac{1}{2\delta})\left(\max_{j=1,\dots,n}\tau_j + \frac{1}{p}\sum_{j=1}^{n}\tau_j\right). \tag{141}$$

For semi-bandit feedback, we use the same ν^j construction but now realize that

$$KL(\nu(a)|\nu^j(a)) = \begin{cases} 0 & \text{if } j \notin a \\ d\left(1-\mu_j|1-\mu_j-\Delta_j\right) & \text{if } j \in a \text{ and } j > k \\ d\left(1-\mu_j|1-\mu_j+\Delta_j\right) & \text{if } j \in a \text{ and } j \le k. \end{cases}$$

Using the same series of steps as above, we find that if

$$\tau_j = \begin{cases} \frac{\mu_j(1-\mu_j-\Delta_j)}{\Delta_j^2} & \text{if } j > k \\ \frac{(\mu_j-\Delta_j)(1-\mu_j)}{\Delta_j^2} & \text{if } j \le k \end{cases}$$

then (141) holds with these defined values of τ_j for the semi-bandit case.