
Explore, Refine, then Commit: Nearly Optimal Multi-Group Mean Estimation with Active Learning

Shourya Pandey¹

Syamantak Kumar¹

¹ Department of Computer Science, UT Austin

Abstract

We present nearly optimal algorithms for simultaneously estimating the means of multiple distributions given a fixed total sample budget while ensuring uniform confidence intervals for the distributions. Our algorithms are based on the well-known *explore-then-commit* policy from the bandit literature along with an additional refinement phase between the exploration phase and the commit phase. Our algorithms achieve regret $\tilde{O}((\log \sigma_{\min}^{-1})^{0.5} T^{-1.5})$ when the rewards are subgaussian and $\tilde{O}(\sigma_{\min}^{-1} T^{-1.5})$ regret when the rewards have bounded fourth moment, where T is the horizon and $\sigma_{\min}^2$ is the minimum among the variances of the distributions of the G arms. We also provide lower bounds on the regret that nearly match the upper bounds for a constant G in both these cases. Importantly, we show that the dependence on $\sigma_{\min}$ in both cases is necessary, resolving a long-standing open problem of [Carpentier et al. \[2011\]](#). We also study the setting where the distributions are hypercontractive. This distribution class includes the family of Gaussian distributions but also includes several families of heavy-tailed distributions. We propose an algorithm for hypercontractive reward distributions that achieves a $\tilde{O}(T^{-1.5})$ regret, independent of $\sigma_{\min}$, and show a nearly matching lower bound for Gaussians.

1 INTRODUCTION

Active learning ([Settles \[2009\]](#), [Cohn et al. \[1996\]](#), [Bounieffouf et al. \[2014\]](#), [Kiran et al. \[2021\]](#), [Yu et al. \[2021\]](#), [Kuwata et al. \[2009\]](#)) is an efficient machine learning paradigm that aims to optimize model performance by selecting the most informative data points for labeling, thereby minimizing the need for large datasets. The model leverages an iterative process by identifying uncertain instances or

high-impact instances, which are then queried to accelerate the learning process. This targeted approach is valuable in domains such as natural language processing, image recognition, and medical diagnosis ([Sarkies et al. \[2015\]](#), [Acemoglu et al. \[2022\]](#), [De Leeuw \[2012\]](#)) where labeled data is often scarce or costly. If the identification procedure is incorrect, the decision can induce structural bias [Pessach and Shmueli \[2022\]](#), [Barbosa et al. \[2016\]](#), [Kleinberg et al. \[2016\]](#).

In this work, we consider the problem of estimating the mean of a group of G distributions using a fixed sampling budget T in the active learning framework. At each time step $t \in [T]$, the agent chooses from a distribution and obtains a sample. The agent wishes to minimize the largest confidence interval obtained for the sample means of the arms at the end of the T rounds. It is easy to show that the number of times the optimal clairvoyant policy samples from the i^{th} distribution is proportional to its variance σ_i^2 (see [Lemma 1](#)). However, this optimal policy is not realizable because the agent does not know the variances of the distributions.

This problem of multi-group mean estimation has been studied extensively in the literature. [Antos et al. \[2008\]](#) formulated the problem in the context of active learning in multi-armed bandits and proposed the GAFS-MAX algorithm, which achieves expected regret (see [\(1\)](#) for the definition of regret) $\tilde{O}(T^{-1.5})$ for bounded rewards, asymptotically in T . This was strengthened to non-asymptotic bounds by [Carpentier et al. \[2011\]](#) for subgaussian rewards. They proposed the B-AS and CH-AS algorithms based respectively on the Bernstein and Chernoff concentration bounds, obtaining a regret bound of $\tilde{O}(T^{-1.5})$. An interesting, and potentially counterintuitive, observation for B-AS and CH-AS highlighted in previous work is that the regret bound scales with $\text{poly}(\sigma_{\min}^{-1})$, where $\sigma_{\min}$ the smallest among the variance of the distributions. More recently, [Aznag et al. \[2024, 2025\]](#) proposed Variance-UCB, along with a generalization of B-AS to the ℓ_p -norm instead of the ℓ_∞ -norm considered in [\(1\)](#). They claim that for the case of bounded subgaussian

rewards¹, the asymptotic regret does not depend polynomially on $\sigma_{\min}^{-1}$, partially resolving the long-standing open question originally posed in [Carpentier et al. \[2011\]](#). All four of these algorithms are variants of the popular *Upper Confidence Bound* (UCB) policy for the standard multi-armed bandit setting [Auer \[2002\]](#), [Bubeck et al. \[2012\]](#) applied to the task of variance estimation.

In the context of the traditional multi-armed bandit problem ([Auer \[2002\]](#), [Lattimore and Szepesvári \[2020\]](#)), there has been an increased interest (see for e.g. [Jin et al. \[2021\]](#), [Chan et al. \[2021\]](#), [Karbasi et al. \[2021\]](#)) in parallelizing the implementations for computational efficiency. Existing algorithms such as Variance-UCB, B-AS cannot be easily parallelized, because they are based on the UCB algorithm which is fully sequential. That is, the algorithm must observe the rewards obtained at each of the previous time steps before deciding which arm to pull next.

Our contributions, motivated by these recent advances and the open questions therein, are summarized below:

1. We extend the previously known results to two popularly used heavy-tailed rewards distributions: (a) μ_k -bounded k^{th} moment (Def. 2), and (b) (k, C_k) -hypercontractivity (Def. 3) by providing highly parallelizable algorithms (Algorithm 1, 2) based on a variation of the ExploreThenCommit policy. Our analysis (Theorems 4 and 6) uses the median-of-means (MoM) estimator for the variance, along with an application of the Riesz–Thorin theorem to obtain tight bounds on the fourth moment given higher moment bounds which may be of independent interest. The algorithm for moment-bounded distributions (Alg 1) is also applicable to subgaussian distributions.
2. We provide a matching lower bound of $O((\log \sigma_{\min}^{-1})^{0.5} T^{-1.5})$ for subgaussian distributions and $O(\sigma_{\min}^{-1} T^{-1.5})$ for fourth-moment bounded distributions, showing that some dependence on $\sigma_{\min}^{-1}$ is indeed unavoidable. In contrast, for hypercontractive reward distributions, which includes the Gaussian distribution, we provide a matching lower bound, with no dependence on $\sigma_{\min}^{-1}$.
3. We also provide sample complexity lower bounds for variance estimation under subgaussianity and under bounded fourth-moment, which may be of independent interest.

Table 1 summarizes and compares our contributions with relevant prior work. The rest of the paper is organized as follows. Section 2 sets up the problem and contains some preliminary definitions and helper lemmas. Section 3 provides variance estimators with confidence intervals that we

use for our algorithms in the following sections. Section 4 describes our algorithms and Section 5 provides the lower bounds. Section 6 has a proof sketch for the upper bound for Gaussian distributions. Section 8 concludes the paper and discusses potential extensions of our methods and directions for future work.

2 PROBLEM SETUP AND PRELIMINARIES

Notation. Let $[n]$ denote the set $\{1, 2, \dots, n\}$. For any vector $v \in \mathbb{R}^G$ and any $p \geq 1$, let $\|v\|_p := (\sum_{i \in [G]} v_i^p)^{1/p}$ denote the ℓ_p -norm and $\|v\|_\infty := \max_{i \in [G]} |v_i|$ denote the ℓ_∞ -norm of v . We use the variable π to denote a policy.

Problem Setup. Consider a G -armed bandit with arms labeled $[G]$ and time horizon T . The reward of arm $i \in [G]$ is a random variable whose distribution $\mathcal{D}_i$ has unknown mean μ_i and standard deviation σ_i . Let $\Sigma := (\sigma_1 \ \sigma_2 \ \dots \ \sigma_G)$ be the vector of true standard deviations of the reward distributions, and let $\sigma_{\min} := \min_{i \in [k]} \sigma_i$ be the smallest standard deviation. We assume $\sigma_{\min} > 0$.

Let a_t and $r_t \sim \mathcal{D}_{a_t}$ be the index and reward of the arm sampled in round t . Let $n_{i,t}$ be the number of times the i^{th} arm is sampled by the end of round t . Let $\tilde{\mu}_{i,t}$ be the sample mean of these $n_{i,t}$ rewards and $\hat{\sigma}_{i,t}$ be an estimate of the variance σ_i^2 using these $n_{i,t}$ samples, with the convention that $\tilde{\mu}_{i,t} = 0$ if $n_{i,t} = 0$ and $\hat{\sigma}_{i,t} = 0$ if $n_{i,t} = 0$ or 1. Note that $\hat{\sigma}_{i,t}$ need not be the sample variance of the $n_{i,t}$ rewards; the exact definition of $\hat{\sigma}_{i,t}$ depends on the policy π and the distributional assumptions on the rewards.

For any fixed $n_{i,t}$, the variance of the random variable $\tilde{\mu}_{i,t}$ is $\sigma_i^2/n_{i,t}$. The vector of *true* sample variances of the arms by round t is denoted $v_t \in \mathbb{R}^G$ and defined as $v_t(i) = \sigma_i^2/n_{i,t}$ for all $i \in [G]$.

Objective. The objective is to minimize the expected ℓ_∞ -norm $R(\pi) := \mathbb{E}_\pi [\|v_T\|_\infty]$ of the vector of variances, called the **cost** of policy π or the **loss** incurred by π . The regret of π is defined by benchmarking against the cost of the optimal clairvoyant policy π^* :

$$R^* := \min_{\substack{n_{i,T} > 0 \ \forall i \in [G] \\ \sum_{i \in [G]} n_{i,T} = T}} \left\| \left\{ \frac{\sigma_i^2}{n_{i,T}} \right\}_{i \in [G]} \right\|_\infty.$$

For the optimal policy, we have relaxed the constraint that the number of pulls $n_{i,T}$ be a non-negative integer because the optimal solution with the additional integral constraint is difficult to characterize. The **regret** of policy π is defined as

$$\text{Reg}(\pi) := \|v_T\|_\infty - R^*, \quad (1)$$

so the objective is equivalent to finding a policy that minimizes the expected regret $\mathbb{E}_\pi [\text{Reg}(\pi)]$.

¹Assumption 5.1. of [Aznag et al. \[2025\]](#) and Lemma B.1. crucially uses Theorem 10 from [Maurer and Pontil \[2009\]](#), which only holds for bounded random variables.

Paper(s)	Algorithm	Reward Distribution	Regret	Lower Bound
Antos et al. [2008]	GAFS-MAX	Bounded Support	$\tilde{O}(\sigma_{\min}^{-1} T^{-1.5})$	—
Carpentier et al. [2011]	CH-AS	Subgaussian	$\tilde{O}(\sigma_{\min}^{-5} T^{-1.5})$	—
	B-AS	Subgaussian	$\tilde{O}(\sigma_{\min}^{-2} T^{-1.5})$	—
Aznag et al. [2024, 2025]	Variance-UCB	Subgaussian ²	$\tilde{O}(T^{-1.5})$	$\Omega(T^{-1.5})$
Our work	Algorithm 1	Subgaussian	$\tilde{O}(\sqrt{\log \sigma_{\min}^{-1} T^{-1.5}})$ (Th 6)	$\Omega(\sqrt{\log \sigma_{\min}^{-1} T^{-1.5}})$ (Th 12)
	Algorithm 2	Hypercontractive	$\tilde{O}(T^{-1.5})$ (Th 7)	$\Omega(T^{-1.5})$ (Th 9)
	Algorithm 1	Bounded moment	$\tilde{O}(\sigma_{\min}^{-1} T^{-1.5})$ (Th 4)	$\Omega(\sigma_{\min}^{-1} T^{-1.5})$ (Th 11)

Table 1: Comparison of different group mean algorithms. For clarity of presentation, we do not include dependence on other parameters such as the number of groups, G and norms of Σ . Although Aznag et al. [2024, 2025] present a lower bound with the correct dependence on T , their upper bounds use concentration result for the standard deviation which implicitly assumes bounded random variables; extending to general subgaussian distributions incurs an additional $\log(1/\sigma_{\min})$ factor which shows up in the leading term of the regret. In contrast, our lower bounds show that for subgaussian distributions and moment-bounded distribution, an dependence on $\sigma_{\min}^{-1}$ in some form is unavoidable. For hypercontractive distributions as well as Gaussian distributions, we achieve a $\sigma_{\min}$ -free bound and a matching lower bound for at least one bandit instance.

Lemma 1. *The cost of the optimal policy,*

$$R^* := \min_{\substack{n_{i,T} > 0 \forall i \in [G] \\ \sum_{i \in [G]} n_{i,T} = T}} \left\| \left\{ \frac{\sigma_i^2}{n_{i,T}} \right\}_{i \in [G]} \right\|_{\infty},$$

is $\frac{\|\Sigma\|_2^2}{T}$, achieved uniquely at $n_{i,T} = \frac{\sigma_i^2}{\|\Sigma\|_2^2} T \forall i \in [G]$.

Proof of Lemma 1. Let $\{n_{i,T}\}_{i \in [G]}$ be any positive reals such that $\sum_{i \in [G]} n_{i,T} = T$, and let $n_{i,T}^*$ be defined as in (2). Since $\sum_{i \in [G]} n_{i,T}^* = T = \sum_{i \in [G]} n_{i,T}$, there exists an index $j \in [G]$ such that $n_{j,T} \leq n_{j,T}^*$. Therefore,

$$\left\| \left\{ \frac{\sigma_i^2}{n_{i,T}} \right\}_{i \in [G]} \right\|_{\infty} \geq \frac{\sigma_j^2}{n_{j,T}} \geq \frac{\sigma_j^2}{n_{j,T}^*} = \frac{\|\Sigma\|_2^2}{T}.$$

Equality holds if and only if $n_{i,T} = n_{i,T}^*$ for all $i \in [G]$. $\square$

Lemma 1 shows that $n_{i,T}^*$, the number of times the optimal clairvoyant policy pulls arm i , is proportional to σ_i^2 :

$$n_{i,T}^* := \frac{\sigma_i^2}{\|\Sigma\|_2^2} T. \quad (2)$$

The value of R^* corresponding to this assignment is equal to $\frac{\|\Sigma\|_2^2}{T}$. Therefore, the regret of policy π can also be written as

$$\text{Reg}(\pi) := \|v_T\|_{\infty} - \frac{\|\Sigma\|_2^2}{T}. \quad (3)$$

Distributional Assumptions. In this paper, we consider three broad classes of nonparametric distributions. The first class of distributions is ν^2 -subgaussian distributions for some $\nu^2 > 0$.

Definition 1 (Subgaussianity). *A random variable $X \in \mathbb{R}$ is said to be ν^2 -subgaussian if for all $t \in \mathbb{R}$,*

$$\mathbb{E}[\exp(t(X - \mathbb{E}[X]))] \leq \exp\left(\frac{\nu^2 t^2}{2}\right).$$

Here ν^2 is called the subgaussian proxy or the variance proxy of the random variable X .

Previous works have analyzed the multi-group mean estimation problem with either subgaussian Carpentier et al. [2011], Aznag et al. [2024], exponential Aznag et al. [2025] or bounded Antos et al. [2008] rewards. We also extend our results to distributions with heavier tails.

Definition 2 (μ_k -bounded k^{th} moment). *A random variable $X \in \mathbb{R}$ is said to have a bounded k^{th} moment for some k if for some $\mu_k > 0$,*

$$\mathbb{E}[|X - \mathbb{E}[X]|^k] \leq \mu_k.$$

The third class of distributions we consider are a special type of heavy-tailed distributions, referred to as (k, C_k) -hypercontractive distributions for some $k > 2$, which allow for finer control on higher moments.

Definition 3 ((k, C_k) -Hypercontractivity). *A random variable $X \in \mathbb{R}$ with $\mu := \mathbb{E}[X]$ is said to be (k, C_k) -hypercontractive for some $k > 2$ and $C_k > 1$ if*

$$\mathbb{E}[|X - \mathbb{E}[X]|^k]^{\frac{1}{k}} \leq C_k \mathbb{E}[(X - \mathbb{E}[X])^2]^{\frac{1}{2}}$$

3 VARIANCE ESTIMATORS

In this section, we state helpful results on the concentration of variance estimators for the distributions which we consider in our policies. For distributions with bounded 4th moment, we restate Lemma 7.1 from Devroye et al. [2016] that provides high probability concentration bounds for a median-of-means estimator of the variance.

Proposition 1. Fix a $\delta \in (0, 1)$. Let $\{X_i\}_{i \in [n]}$ be iid random variables with variance σ^2 and fourth moment at most κ^4 (Definition 3). Let $B := \lfloor \min\{8 \log(e^{1/8}/\delta), n/2\} \rfloor$ and $N = \lfloor n/B \rfloor$. Define $\hat{\sigma}^2 := \text{Median} \{\hat{\sigma}_i^2\}_{i \in [B]}$, where

$$\hat{\sigma}_i^2 := \frac{\sum_{t=(i-1)N+1}^{iN} X_t^2}{N} - \left(\frac{\sum_{t=(i-1)N+1}^{iN} X_t}{N} \right)^2$$

Then, with probability at least $1 - \delta$,

$$|\hat{\sigma}^2 - \sigma^2| \leq 2e\sqrt{6(\kappa^4 + 3\sigma^4)} \sqrt{\frac{\log(1/\delta)}{n}}.$$

We use Proposition 1 to derive confidence intervals for our variance estimator for subgaussian distributions, μ_k -bounded k^{th} moment distributions with $k \geq 4$ and (k, C_k) -hypercontractive distributions with $k \geq 4$.

Remark 1. While Proposition 1 provides a generic method to get optimal rates on the variance estimates for heavy tails, other methods that achieve the same concentration bound exist. The median-of-means procedure we use in our algorithm can be replaced with any other equivalent mean-estimation method such as Lee and Valiant [2022] to improve the constant factors.

Using the median-of-means-based variance estimator and concentration bound of Proposition 1, we can obtain tighter concentration for the variance σ if a bound on the k^{th} moment for $k > 4$ is known or if it is known that the distribution is subgaussian. Theorems 1 and 2 record these guarantees. These confidence intervals depend only on the estimated variance $\hat{\sigma}$ and the moment bound μ_k (or the subgaussian parameter ν) and are therefore computable.

Theorem 1 (CI for bounded k^{th} moment where $k > 4$). Let X be a real-valued random variable with mean μ and variance $\sigma^2 > 0$. Let $Y := X - \mu$ and assume that for some $k > 4$ the centered k -th moment is bounded by μ_k :

$$\mu_k := \mathbb{E}[|Y|^k] < \infty.$$

Let $X_1, \dots, X_n$ be i.i.d. copies of X . Then, with probability at least $1 - \delta$,

$$|\hat{\sigma}^2 - \sigma^2| \leq 2e\sqrt{24}\hat{\sigma}^{\frac{k-4}{k-2}} \mu_k^{\frac{1}{k-2}} \sqrt{\frac{\log(1/\delta)}{n}} + (2e\sqrt{24})^{3/2} \mu_k^{\frac{2}{k}} \left(\frac{\log(1/\delta)}{n} \right)^{\frac{3k-8}{4(k-2)}}.$$

where $\hat{\sigma}^2$ is the median-of-means estimator of Proposition 1.

Theorem 2 (CI for subgaussians). Let X be a ν^2 -subgaussian real-valued random variable with mean μ and variance $\sigma^2 > 0$, and let $Y := X - \mu$.

$$\mu_k := \mathbb{E}[|Y|^k] < \infty.$$

Let $X_1, \dots, X_n$ be i.i.d. copies of X and let δ be such that

$$n \geq 32e^2 \log(1/\delta).$$

Then, with probability at least $1 - \delta$,

$$|\hat{\sigma}^2 - \sigma^2| \leq 32e\sqrt{6}(\hat{\sigma}^2 + \Delta)^{1/2} \nu \sqrt{\log \frac{e^2 \nu^2}{\hat{\sigma}^2 + \Delta} \cdot \frac{\log(1/\delta)}{n}},$$

where $\Delta = 12e\sqrt{2}\nu^2 \sqrt{\frac{\log(1/\delta)}{n}}$ and $\hat{\sigma}^2$ is the median-of-means estimator of Proposition 1.

For $2 < k < 4$, the fourth moment bound cannot be used and the dependence on n is weaker. The median-of-means algorithm with appropriate parameters achieves the optimal regret for mean estimation without a variance, as shown in Devroye et al. [2016] and Bubeck et al. [2013]. The proof of Theorem 3 follows a similar argument and is deferred to Appendix A.

Theorem 3 (CI for bounded k^{th} moment where $2 < k \leq 4$). Let X be a real-valued random variable with mean μ and variance $\sigma^2 > 0$. Let $Y = X - \mu$ and assume that for some $2 < k \leq 4$, the centered k -th moment is bounded by μ_k :

$$\mu_k := \mathbb{E}[|Y|^k] < \infty.$$

Let $X_1, \dots, X_n$ be i.i.d. copies of X . Then, for every $\delta \in (0, 1/2)$ there exists a median-of-means estimator $\hat{\sigma}^2$ with appropriate parameters such that with probability at least $1 - \delta$,

$$|\hat{\sigma}_n^2 - \sigma^2| \leq 64 \cdot 6^{2/k} \mu_k^{2/k} \left(\frac{\ln(1/\delta)}{n} \right)^{1-2/k}.$$

4 ALGORITHMS

We present two algorithms for the multi-group mean estimation problem. The algorithms are broadly based on the *explore-then-commit* policy from the standard bandit literature Lattimore and Szepesvári [2020], Garivier et al. [2016]. The basic skeleton of such a policy adapted to the multi-group mean estimation setting is as follows.

4.1 SUBGAUSSIAN DISTRIBUTIONS AND DISTRIBUTIONS WITH BOUNDED k -TH MOMENT

Algorithm 1 is one such explore-refine-then-commit policy for subgaussian rewards and for moment-bounded rewards.

In the **exploration** phase, the algorithm pulls the arms sequentially in a round-robin manner. We refer to each complete cycle of pulling all the arms once as an *iteration*. The policy explores for m_* iterations, which we call the *exploration parameter*.

The choice of the exploration parameter m_* is crucial. If m_* is too small, the variance estimates deviate significantly from the true variances Σ , and any estimates of $n_{i,T}^*$ based on these estimates may lead to high regret. On the other hand, if m_* is too large, say $m_* \gg \min_{i \in [G]} n_{i,T}^*$, the algorithm significantly over-pulls some of the arms, which

Algorithm 1: ExploreRefineThenCommit

```
1 Input: Number of arms  $G$ , Time horizon  $T$ 
2  $\delta \leftarrow 1/T^2$ 
3 Pull each arm  $\lceil 32e^2 \log(GT/\delta) \rceil$  times
4  $m \leftarrow \lceil 32e^2 \log(GT/\delta) \rceil$ 
5  $b \leftarrow \text{true}$ 
6 while  $b = \text{true}$  do
7   Pull each arm once
8    $m \leftarrow m + 1$ 
9   for  $i \in [G]$  do
10     $\tilde{n}_{i,m} \leftarrow \frac{\bar{\sigma}_{i,m}^2}{\sum_{j \in [G]} \bar{\sigma}_{j,m}^2} \cdot T$  where  $\bar{\sigma}_{j,m}^2$  are defined
    as in equation (5)
11    if  $\tilde{n}_{i,m} < 4m$  then
12       $b \leftarrow \text{false}$ 
13       $m_* \leftarrow m - 1$ 
14    end
15  end
16 end
17 for  $i \in [G]$  do
18    $n_i^{\text{coarse}} \leftarrow \tilde{n}_{i,m_*}$ 
19   Pull arm  $i$   $\lceil n_i^{\text{coarse}}/2 \rceil - (m_* + 1)$  times
20 end
21 for  $i \in [G]$  do
22    $t_i \leftarrow \lceil n_i^{\text{coarse}}/2 \rceil$ 
23    $n_i^{\text{fine}} \leftarrow \frac{\bar{\sigma}_{i,t_i}^2}{\sum_{j \in [G]} \bar{\sigma}_{j,t_j}^2} \cdot T$ 
24 end
25 if  $\lceil n_i^{\text{fine}} \rceil < t_i$  for some  $i$  then
26   for  $i \in [G]$  do
27     Pull arm  $i$   $\lceil n_i^{\text{coarse}} \rceil - \lceil n_i^{\text{coarse}}/2 \rceil$  times
28   end
29   return
30 else
31   for  $i \in [G]$  do
32     Pull arm  $i$   $\lceil n_i^{\text{fine}} \rceil - \lceil n_i^{\text{coarse}}/2 \rceil$  times
33   end
34   return
35 end
```

comes at the cost of under-pulling some other arm, leading to high regret. Another noteworthy point is that the exploration parameter m_* cannot be chosen in advance. Similar to the explore-then-commit policy in the standard bandit setting, for any data-oblivious choice of the exploration parameter m_* the regret of the explore-then-commit policy in the multi-group mean estimation setting is sub-optimal in T (precisely speaking, the regret can be shown to be $\tilde{O}(T^{-4/3})$). Therefore, m_* must be chosen in a data-dependent manner.

Once the algorithm finishes the exploration phase and the coarse estimates n_j^{coarse} are obtained, the algorithm pulls

each arm about $n_j^{\text{coarse}}/2$ times and **refines** the estimates of the optimal number of pulls $n_{i,T}^*$ based on these refined values n_j^{fine} . For moment-bounded and subgaussian distributions, this refinement step can potentially fail for small T , in which case the algorithm simply uses the coarse estimates n_j^{coarse} . Otherwise, the algorithm **commits** to the finer estimates n_j^{fine} .

Algorithm 1 has low regret as long as the variance estimates concentrate. More explicitly, we can define an event $\mathcal{E}$ conditioned on which the regret bounds of Theorems 6 and 4 hold without the expectation. For any event $\mathcal{E}$,

$$\mathbb{E}[\text{Reg}(\pi)] = \mathbb{E}[\text{Reg}(\pi)|\mathcal{E}] \mathbb{P}(\mathcal{E}) + \mathbb{E}[\text{Reg}(\pi)|\mathcal{E}^c] \mathbb{P}(\mathcal{E}^c).$$

If $\mathcal{E}$ occurs with sufficiently high probability, and the policy π almost surely pulls each arm at least once, then the expected regret can be bounded as

$$\mathbb{E}[\text{Reg}(\pi)] \leq \mathbb{E}[\text{Reg}(\pi)|\mathcal{E}] + \sigma_{\max}^2 \mathbb{P}(\mathcal{E}^c). \quad (4)$$

The proofs of Theorems 6 and 4 show an upper bound on the first term of equation (4) under the event $\mathcal{E}$. If $\mathcal{E}$ is such that $\mathbb{P}(\mathcal{E}^c) = T^{-2}$, the second term $\frac{\sigma_{\max}^2}{T^2}$ of (4) is smaller than the upper bound on the first term and can therefore be neglected. We define the event $\mathcal{E}$ of interest next, which is that the empirical variance estimates behave as intended.

Similar to the standard bandit setting, we can assume that T rewards for each of the G arms are generated in advance. Pulling an arm then reveals the next reward in the predetermined reward sequence for that arm, rather than generating a new reward at the time of the pull.

For each $i \in [G]$ and $2 \leq t \leq T$, we expect the variance estimate $\hat{\sigma}_{i,t}^2$ for σ_i^2 using the first t rewards to concentrate. For instance, if the rewards have centered fourth moment bounded by μ_4 , then Lemma 1 implies that with probability at least $1 - \frac{\delta}{GT}$, where $\delta = T^{-2}$,

$$|\hat{\sigma}_{i,t}^2 - \sigma^2| \leq 2e\sqrt{24}\mu_4^{1/2} \sqrt{\frac{\log(GT/\delta)}{t}}.$$

By a union bound, this holds for all $i \in [G]$ and $2 \leq t \leq T$ with probability at least $1 - \delta = 1 - T^{-2}$.

In general, for bounded k th moment for $k > 2$ and for subgaussian distributions, Theorems 1, 2, and 3 imply that with probability at least $1 - \delta$, $\bar{\sigma}_{i,t}$ is an upper bound on σ_i for all $i \in [G]$ and $2 \leq t \leq T$ where

$$\sigma_i^2 \leq \bar{\sigma}_{i,t}^2 := \hat{\sigma}_{i,t}^2 + f(\hat{\sigma}_{i,t}, t) \quad (5)$$

where f also depends on δ, G, T , and also on k and μ_k for the moment-bounded case or ν for the subgaussian case. This is the event $\mathcal{E}$.

In the exploration phase, the algorithm pulls the arms in a round-robin fashion. The algorithm also maintains estimates

$\tilde{\Sigma}_m$ of the vector Σ of deviations of the reward distributions after iteration m , for each $2 \leq m \leq m_*$. Here, a single iteration involves pulling each of the G arms once (and updating any estimates). After m_* rounds of exploration, the algorithm finds some arm i for which the estimate $\tilde{n}_{i,m}$ is multiplicatively close to the number of times m it has already been pulled.

At this stage, the algorithm stops exploring and refines the estimates by pulling each arm i a total of about $n_i^{\text{coarse}}/2$ times. Finally, the algorithm commits to the estimates n_i^{fine} (line 32). As a fail-safe, if $n_i^{\text{coarse}}/2$ is already greater than any of the fine estimates n_i^{fine} then the arms are simply pulled according to the coarse estimates (line 27).

We present the regret bounds for large T ; a full derivation of the non-asymptotic regret for all T is provided in Appendix B.

Theorem 4. Suppose the distributions $\mathcal{D}_i$ have k th moment bounded by μ_k for some known $k > 4$ and $\mu_k > 0$. For sufficiently large T , the regret of Algorithm 1 satisfies

$$\text{Reg} = O \left(\frac{\sqrt{\log(GT)} \|\Sigma\|_2}{T^{1.5}} \cdot \mu_k^{\frac{1}{k-2}} \sum_{i \in [G]} \frac{1}{\sigma_i^{\frac{2}{k-2}}} \right). \quad (6)$$

Theorem 5. Suppose the distributions $\mathcal{D}_i$ have k th moment bounded by μ_k for some known $k \in (2, 4)$ and $\mu_k > 0$. For sufficiently large T , the regret of Algorithm 1 satisfies

$$\text{Reg} = O \left(\frac{(\log(GT))^\beta \|\Sigma\|_2^{2\beta} \mu_k^{1-\beta}}{T^{1+\beta}} \sum_{i \in [G]} \frac{1}{\sigma_i^{2\beta}} \right), \quad (7)$$

where $\beta = 1 - 2/k$.

Theorem 6. Suppose the distributions $\mathcal{D}_i$ are ν^2 -subgaussian. For sufficiently large T , the regret of Algorithm 1 satisfies

$$\text{Reg} = O \left(\frac{\sqrt{\log(GT)} \|\Sigma\|_2}{T^{1.5}} \cdot \nu \sum_{i \in [G]} \sqrt{\log \frac{e\nu}{\sigma_i}} \right). \quad (8)$$

4.2 HYPERCONTRACTIVE DISTRIBUTIONS

The following section presents Algorithm 2 for (k, C_k) -hypercontractive reward distributions (Definition 3) with $k \geq 4$, which includes Gaussian rewards as a special case but also includes distributions with heavier tails. Similar to Algorithm 1, this algorithm has an exploration phase after which the coarse estimates $\tilde{n}_i^{\text{coarse}}$ are multiplicatively close to the optimal $n_{i,T}^*$, followed by a refinement phase using these coarse estimates, after which the algorithm commits to these finer estimates. We show crucially that the algorithm does not incur a $1/\sigma_{\min}$ factor in the expected regret. It achieves this improved rate due to tighter concentration bounds on the variances of the rewards, which allows for a more accurate estimation of the optimal number of pulls $n_{i,T}^*$.

Algorithm 2: ExploreRefineThenCommit

- 1 **Input:** Number of arms G , Horizon T , Hypercontractivity parameters (k, C_k)
 - 2 $\delta \leftarrow GT^{-3}$
 - 3 $C \leftarrow 3 \log(1/\delta)$ // concentration parameter
 - 4 Pull each arm $\lceil 9C_k^4 C \rceil$ times
 - 5 $\tilde{\Sigma}_{\text{coarse}} \leftarrow$ estimates of the standard deviations
 - 6 $T' \leftarrow T - \lceil 9C_k^4 C \rceil G$
 - 7 $\tilde{n}_i^{\text{coarse}} \leftarrow \frac{\hat{\sigma}_{i,\text{coarse}}^2}{\|\tilde{\Sigma}_{\text{coarse}}\|_2^2} T'$
 - 8 Pull arm $i \lfloor \tilde{n}_i^{\text{coarse}}/6 \rfloor$ times
 - 9 $\tilde{\Sigma}_{\text{fine}} \leftarrow$ estimates of the standard deviations
 - 10 $\tilde{n}_i^{\text{fine}} \leftarrow \frac{\hat{\sigma}_{i,\text{fine}}^2 \left(1 + C_k^2 \sqrt{\frac{6C}{\tilde{n}_i^{\text{coarse}}}} \right)}{\sum_{j \in [G]} \hat{\sigma}_{j,\text{fine}}^2 \left(1 + C_k^2 \sqrt{\frac{6C}{\tilde{n}_j^{\text{coarse}}}} \right)} T'$
 - 11 Pull arm $i \lfloor \tilde{n}_i^{\text{fine}} \rfloor - \lfloor \tilde{n}_i^{\text{coarse}}/6 \rfloor$ times
-

Theorem 7. Suppose the distributions $\mathcal{D}_i$ are (k, C_k) -hypercontractive for some $k \geq 4$. Then, for $T = \tilde{\Omega}(G)$, the regret of Algorithm 2 satisfies

$$\mathbb{E}[\text{Reg}(\pi)] = \tilde{O} \left(\frac{\|\Sigma\|_1 \|\Sigma\|_2 C_k^2}{T^{1.5}} + \frac{\|\Sigma\|_2^2 G C_k^4}{T^2} \right). \quad (9)$$

Appendix C extends the algorithm to the full range of $k > 2$. We state the theorem here, which is restated and proved as Theorem 13.

Theorem 8. Suppose the distributions $\mathcal{D}_i$ are (k, C_k) -hypercontractive for some $k > 2$, and let $p_k = \min(1 - \frac{2}{k}, \frac{1}{2})$. Then, for $T = \tilde{\Omega}(G C_k^{2/p_k})$, the regret of Algorithm 3 satisfies

$$\mathbb{E}[\text{Reg}(\pi)] = \tilde{O} \left(\frac{\|\Sigma\|_1^{2(1-p_k)} \|\Sigma\|_2^{2p_k} C_k^2}{T^{1+p_k}} + \frac{\|\Sigma\|_2^2 G C_k^{2/p_k}}{T^2} \right). \quad (10)$$

5 LOWER BOUNDS

In this section, we state the lower bounds on the regret of any algorithm for the multi-group mean estimation problem.

For Gaussian distributions, the B-AS strategy achieves a regret bound of $\tilde{O}(G^2 \|\Sigma\|_2^2 T^{-1.5})$, independent of $\sigma_{\min}$. Our explore-refine-then-commit policy achieves a regret of $\tilde{O}(\|\Sigma\|_1 \|\Sigma\|_2 T^{-1.5})$ (see Theorem 7) for Gaussian distributions but also for (k, C_k) hypercontractive distributions with $k \geq 4$ and constant C_k . Our regret bound improves over the regret bound of B-AS by a factor of at least $\tilde{O}(G^{1.5})$, and matches the regret bound of Aznag et al. [2024] for Gaussian distributions while also generalizing it to hypercontractive distributions.

We also provide a matching lower bound for Gaussian distributions.

Theorem 9. *For any policy π , and positive integers G and T with $T > 2G$, there exists a bandit instance with G arms, time horizon T , and Gaussian rewards, such that*

$$\text{Reg}(\pi) \geq \frac{\|\Sigma\|_1 \|\Sigma\|_2}{T^{1.5}}.$$

The theorem is restated and proved in Theorem 14.

In the subgaussian setting, the B-AS strategy of Carpentier et al. [2011] (short for the Bernstein Allocation Strategy) achieves regret $\tilde{O}(\sigma_{\min}^{-2} T^{-1.5})$, which is optimal in T (Aznag et al. [2024]) and improves the dependence on $\sigma_{\min}$ compared to the CH-AS strategy in Carpentier et al. [2011] and the GAFS-MAX strategy in Antos et al. [2008]. Our ETC-based policy achieves regret $\tilde{O}(\sigma_{\min}^{-1} T^{-1.5})$ for distributions with bounded fourth moment and $\tilde{O}((\log \sigma_{\min}^{-1})^{0.5} T^{-1.5})$ for subgaussian distributions.

A natural question is whether the dependence on $\sigma_{\min}$ is unavoidable. As noted in prior work, this dependence may be necessary. The inverse dependence on $\sigma_{\min}$ is odd: decreasing the lowest variance makes the feedback from that arm, and one should expect the regret to decrease Aznag et al. [2024]. Perhaps surprisingly, **we show that such a dependence on $\sigma_{\min}$ is necessary.**

All of our lower bounds for subgaussian and bounded fourth moment distributions follow the same reduction. We work with anchored two-arm instances $\mathcal{I}_\theta = (D_\theta, \mathcal{R})$ where $\mathcal{R}$ is a fixed mean-zero reference arm with known variance $\text{Var}(\mathcal{R}) = 1$ (e.g. Rademacher), and the unknown arm is either D_0 or D_1 with $\text{Var}(D_0) = \sigma^2$ and $\text{Var}(D_1) = \sigma^2 + \Delta$. For this anchored problem, the “oracle” terminal allocation (the minimizer of the terminal variance proxy) depends only on the unknown variance level. Consequently, if a policy π achieves very small terminal regret, then its pull count N_T , must lie in a narrow interval around the oracle allocation corresponding to the true variance. A regret upper bound forces N_T to satisfy a *budget condition* of the form $N_T \leq m$ where m is the effective number of samples of D_θ that π can afford to collect while keeping regret small. Intuitively, when σ^2 is small the oracle pulls the unknown arm only about $T \cdot \frac{\sigma^2}{1+\sigma^2}$ times; hence a low-regret policy is forced to observe only $m \asymp \sigma^2 T$ samples from D_θ .

Given the above, we can define a test $\hat{C}_m$ that observes only the samples drawn from the unknown arm (up to time T) and tries to decide whether $\theta = 0$ or $\theta = 1$ based on the pull counts of each arm. If π had uniformly tiny regret on both $\mathcal{I}_0$ and $\mathcal{I}_1$, this would yield a test with small average error using only m i.i.d. samples from D_θ . Equivalently, the bandit policy would implicitly provide a *variance estimator* with accuracy $\asymp \Delta$ based on only m samples of the unknown arm.

The final ingredient is a *variance-testing lower bound*: construct distributions D_0, D_1 whose variances differ by Δ , but such that no procedure can reliably distinguish D_0^m from D_1^m (average error bounded below by δ). This is exactly the testing assumption (11) appearing in Theorem 10, and provided formally for subgaussian and bounded moment distributions in Theorems 15 and 16. Combining the above reduction with these bounds then yields a contradiction unless the regret is at least of order $\delta \Delta / T$, which is the content of Theorem 10.

Theorem 10 (Reduction to variance estimation). *Fix $T \in \mathbb{N}$ and $\delta \in (0, 1/4)$. Let D_0, D_1 be mean-zero distributions with*

$$\text{Var}(D_0) = \sigma^2, \quad \text{Var}(D_1) = \sigma^2 + \Delta,$$

for some $\sigma \in (0, 0.5)$ and $\Delta \in (0, \sigma^2/10)$. Let $\mathcal{R}$ be any mean-zero reference arm with $\text{Var}(\mathcal{R}) = 1$ (e.g. a Rademacher arm). Assume that for some integer m , over all functions $\hat{C}_m : \mathbb{R}^m \rightarrow \{0, 1\}$,

$$\inf_{\hat{C}_m} \frac{1}{2} \left(\Pr_{X \sim D_0^m} (\hat{C}_m(X) = 0) + \Pr_{X \sim D_1^m} (\hat{C}_m(X) = 1) \right) \leq 1 - \delta, \quad (11)$$

equivalently, every test has average error at least δ . Suppose in addition that m satisfies

$$m \geq T - \frac{T}{(\sigma^2 + \Delta) + 1 + \Delta/32}. \quad (12)$$

Then for every policy π run for horizon T on the instances $\mathcal{I}_0 = (D_0, \mathcal{R})$ and $\mathcal{I}_1 = (D_1, \mathcal{R})$, we have

$$\max_{\theta \in \{0,1\}} \mathbb{E}_\theta [\text{Reg}(\pi)] \geq c \cdot \frac{\delta \Delta}{T},$$

for a universal constant $c > 0$.

We now apply Theorem 10 to provide our lower bound results in Theorems 11 and 12.

Theorem 11 (Lower bound for fourth-moment bounded distributions). *Fix $T \in \mathbb{N}$, $\sigma \in (0, 0.5)$, and $\delta \in (0, 1/4)$. Let $\mathcal{R} = \text{Rademacher}$. Set $m := \lceil 2\sigma^2 T \rceil$ and let*

$$\Delta = \min \left\{ \frac{\sigma^2}{10}, \sqrt{\frac{\log(1/(4\delta))}{4m}} \right\} = \min \left\{ \frac{\sigma^2}{10}, \sqrt{\frac{\log(1/(4\delta))}{8\sigma^2 T}} \right\}.$$

Then there exist mean-zero distributions $D_0, D_1 \in \mathcal{F}_\sigma := \{\mathcal{D} : \text{Var}(\mathcal{D}) \leq 4\sigma^2, \kappa(\mathcal{D}) \leq 2\}$ satisfying $\text{Var}(D_0) = \sigma^2$ and $\text{Var}(D_1) = \sigma^2 + \Delta$ such that for every policy π ,

$$\max_{\theta \in \{0,1\}} \mathbb{E}_\theta [\text{Reg}(\pi)] \geq c \cdot \frac{\delta}{T} \min \left\{ \frac{\sigma^2}{10}, \sqrt{\frac{\log(1/(4\delta))}{8\sigma^2 T}} \right\},$$

for a universal constant $c > 0$.

Theorem 12 (Lower bound for subgaussian distributions). *Fix $T \in \mathbb{N}$, $\sigma \in (0, 0.5)$, and $\delta \in (0, 1/4)$. Let $\mathcal{R} =$*

Rademacher. Set $m := \lceil 2\sigma^2 T \rceil$ and let

$$\begin{aligned}\Delta &= \min \left\{ \frac{\sigma^2}{10}, \sigma \sqrt{\frac{\log(1/\sigma) \log(1/(4\delta))}{4m}} \right\} \\ &= \min \left\{ \frac{\sigma^2}{10}, \sqrt{\frac{\log(1/\sigma) \log(1/(4\delta))}{8T}} \right\}.\end{aligned}$$

Then there exist mean-zero distributions $D_0, D_1 \in \text{subG}(1)$ with $\text{Var}(D_0) = \sigma^2$ and $\text{Var}(D_1) = \sigma^2 + \Delta$ such that for every policy π ,

$$\max_{\theta \in \{0,1\}} \mathbb{E}_\theta [\text{Reg}(\pi)] \geq c \cdot \frac{\delta}{T} \min \left\{ \frac{\sigma^2}{10}, \sqrt{\frac{\log(1/\sigma) \log(1/(4\delta))}{8T}} \right\},$$

for a universal constant $c > 0$.

6 PROOF SKETCH OF THE UPPER BOUND FOR GAUSSIANS

For exposition, we provide a sketch of the argument for Gaussian distributions (Algorithm 2). The key idea of Algorithm 2 is to obtain as tight bounds on the variances of the distributions as possible. Any “good” policy π with low expected regret should pull arm i about $n_{i,T}^*$ times, the number of times the optimal policy π^* pulls arm i . Appealing to Lemma 8, a tight estimate on the variance of arm i using $n_{i,T}^*$ samples is $\sigma_i^2 \left(1 \pm \tilde{O}\left(\frac{1}{\sqrt{n_{i,T}^*}}\right)\right)$. So, it is reasonable to believe that the tightest concentration for $\tilde{n}_{i,T}$, the total number of times π pulls arm i , is

$$\tilde{n}_{i,T} \simeq \frac{\sigma_i^2 \left(1 \pm \frac{1}{\sqrt{n_{i,T}^*}}\right)}{\sum_{j \in [G]} \sigma_j^2 \left(1 \pm \frac{1}{\sqrt{n_{j,T}^*}}\right)} T.$$

Algorithm 2 achieves precisely this type of a concentration bound. The algorithm operates in three main phases. The first phase is **exploration**, where each arm is pulled $O(C)$ times to get a multiplicative approximation to the variances.

Lemma 2. For every $i \in [G]$, the estimate $\hat{\sigma}_{i,\text{coarse}}^2$ of σ_i^2 satisfies $\left| \frac{\hat{\sigma}_{i,\text{coarse}}^2}{\sigma_i^2} - 1 \right| \leq \frac{1}{3}$.

These multiplicative estimates of the variances suffice to get multiplicative approximations $\tilde{n}_{i,\text{coarse}}^*$ to the optimal number of pulls $n_{i,T'}^*$. Here, *optimal* is defined assuming the horizon is T' , that is, $n_{i,T'}^* := \frac{\sigma_i^2}{\|\Sigma\|_2^2} T'$ (we will account for the discrepancy later). These coarse estimates $\tilde{n}_{i,\text{coarse}}^*$ themselves are not good enough to achieve regret of equation (9), but are good enough to achieve the tightest (up to constants) concentration bounds on the variances σ_i^2 while not over-pulling any arms.

The **refinement** phase uses these tighter variance estimates to improve the estimates of $n_{i,T'}^*$. In this phase, arm i is

pulled $\frac{1}{6} \tilde{n}_{i,\text{coarse}}^* = \frac{\hat{\sigma}_{i,\text{coarse}}^2}{\|\Sigma_{\text{coarse}}\|_2^2} T'$ times, which is multiplicatively close to $n_{i,T'}^*$ due to Lemma 2.

Lemma 3. For all $i \in [G]$ the coarse estimate $\tilde{n}_{i,\text{coarse}}^*$ of $n_{i,T'}^*$ satisfies $\frac{1}{2} n_{i,T'}^* \leq \tilde{n}_{i,\text{coarse}}^* \leq 2n_{i,T'}^*$.

The factor $1/6$ is to ensure that no arm is over-pulled. Using the samples obtained by pulling these arms, the estimates of the variances can be refined.

Lemma 4. For all $i \in [G]$ the estimate $\hat{\sigma}_{i,\text{fine}}^2$ of the variance σ_i^2 satisfies

$$\left| \frac{\hat{\sigma}_{i,\text{fine}}^2}{\sigma_i^2} - 1 \right| \leq C_k^2 \sqrt{\frac{6C}{\tilde{n}_{i,\text{coarse}}^*}}.$$

These finer estimates are close to the best estimates we can get for the variances of the distributions. The algorithm **commits** to these estimates in the last phase and pull the arms roughly in proportion to the estimates $\text{sig} \hat{ma}_{i,\text{fine}}^2$. Lemma 5 states that $\tilde{n}_{i,\text{fine}}^*$ is close to $n_{i,T'}^*$ for all $i \in [G]$.

Lemma 5. For all $i \in [G]$ the fine estimate $\tilde{n}_{i,\text{fine}}^*$ of $n_{i,T'}^*$ satisfies

$$\tilde{n}_{i,\text{fine}}^* \geq \frac{\sigma_i^2}{\|\Sigma\|_2^2 + \frac{2\sqrt{12C}\|\Sigma\|_1\|\Sigma\|_2}{T^{0.5}}} T'.$$

By Lemma 5, the cost of algorithm 2 is

$$\max_{i \in [G]} \frac{\sigma_i^2}{\tilde{n}_{i,\text{fine}}^*} \leq \frac{\|\Sigma\|_2^2}{T'} + \frac{2\sqrt{12C}\|\Sigma\|_1\|\Sigma\|_2}{T'^{1.5}}.$$

Assuming $T = \Omega(GC)$ for a sufficiently large constant, the regret of algorithm 2 is bounded above as

$$\text{Reg} = \max_{i \in [G]} \frac{\sigma_i^2}{\tilde{n}_{i,\text{fine}}^*} - \frac{\|\Sigma\|_2^2}{T} = O\left(\frac{\sqrt{C}\|\Sigma\|_1\|\Sigma\|_2}{T^{1.5}} + \frac{CG\|\Sigma\|_2^2}{T^2}\right).$$

7 EXPERIMENTS

We compare our algorithms with three baselines on synthetic Gaussian bandits to validate the $T^{-3/2}$ rate predicted by our upper bounds and to illustrate the parallelizability advantage of our algorithms.

Setting. We consider $G \in \{5, 15\}$ arms with rewards $\mathcal{N}(0, \sigma_i^2)$, where $\sigma_i = 0.1 + (i - 1)/(G - 1) \cdot 0.9$ are linearly spaced in $[0.1, 1.0]$. Each (G, T) configuration is run for 50 independent trials with $T \in \{1000, 2000, 4000, 8000, 16000\}$; the regret $\text{Reg}(\pi)$ is as defined in (3).

Algorithms. We compare **Algorithm 1** and **Algorithm 2** from this paper against B-AS [Carpentier et al. \[2011\]](#), GAFS-MAX [Antos et al. \[2008\]](#), and Variance-UCB [Aznag et al. \[2024\]](#), all with subgaussian parameter $\nu = 1$ and UCB constants set by cross-validation.

Table 2: Mean regret $\mathbb{E} [\text{Reg}(\pi)]$ on Gaussian bandits, $\sigma_i \in [0.1, 1.0]$, 50 trials. **Bold:** our algorithms.

Algorithm	$T = 10^3$	$T = 4 \times 10^3$	$T = 8 \times 10^3$	$T = 1.6 \times 10^4$
$G = 5, \quad \ \Sigma\ _1 \ \Sigma\ _2 \approx 3.91, \quad \ \Sigma\ _2^2 / T = 2.02/T$				
Alg. 1 (ours)	5.09×10^{-4}	6.14×10^{-5}	2.16×10^{-5}	8.76×10^{-6}
Alg. 2 (ours)	5.56×10^{-4}	9.34×10^{-5}	4.49×10^{-5}	2.28×10^{-5}
B-AS	3.09×10^{-4}	4.64×10^{-5}	1.75×10^{-5}	6.03×10^{-6}
GAFS-MAX	4.99×10^{-4}	1.08×10^{-4}	5.15×10^{-5}	1.96×10^{-5}
Variance-UCB	4.15×10^{-4}	6.61×10^{-5}	2.44×10^{-5}	9.62×10^{-6}
$G = 15, \quad \ \Sigma\ _1 \ \Sigma\ _2 \approx 19.69, \quad \ \Sigma\ _2^2 / T = 2.02/T$				
Alg. 1 (ours)	3.06×10^{-3}	4.12×10^{-4}	1.49×10^{-4}	5.43×10^{-5}
Alg. 2 (ours)	6.17×10^{-3}	5.72×10^{-4}	2.44×10^{-4}	8.87×10^{-5}
B-AS	2.17×10^{-3}	2.67×10^{-4}	1.01×10^{-4}	3.72×10^{-5}
GAFS-MAX	1.45×10^{-2}	1.99×10^{-3}	5.57×10^{-4}	1.32×10^{-4}
Variance-UCB	2.53×10^{-3}	3.56×10^{-4}	1.31×10^{-4}	5.15×10^{-5}

Table 3: Rounds of adaptivity. One round is a maximal parallelisable batch of pulls. Algorithm 2 always uses 3; Algorithm 1 uses $O(\log T)$; UCB baselines are fully sequential.

Algorithm	10^3	10^4	10^5	10^6
Alg. 2 (ours)	3	3	3	3
Alg. 1 (ours)	~ 10	~ 14	~ 17	~ 20
B-AS	10^3	10^4	10^5	10^6
GAFS-MAX	10^3	10^4	10^5	10^6
Variance-UCB	10^3	10^4	10^5	10^6

Regret vs. T . Table 2 reports mean regret for $G \in \{5, 15\}$. Standard errors of the mean are $\leq 8\%$ of the reported values for all algorithms except GAFS-MAX ($\leq 31\%$).

Discussion. We make the following observations.

Rate. Algorithm 1, B-AS, and Variance-UCB all achieve a ratio of $50\text{--}60\times$ in mean regret from $T = 10^3$ to $T = 1.6 \times 10^4$ in both settings, consistent with the theoretical $(16)^{3/2} = 64\times$ of Theorems 4 and 6. Algorithm 2 achieves $70\times$ at $G = 15$ but only $24\times$ at $G = 5$. This is consistent with the lower-order term $\tilde{O}\left(\|\Sigma\|_2^2 GC_k^4 / T^2\right)$ in Theorem 7 dominating the leading $T^{-3/2}$ term at small G and the tested budgets.

Absolute regret. Algorithm 1 is within $1.3\text{--}1.5\times$ of B-AS across all configurations and uniformly below Variance-UCB at $G = 5$. Algorithm 2 incurs a larger constant due to the correction factor $1 + C_k^2 \sqrt{6C/\tilde{n}_i^{\text{coarse}}}$ in line 10, which over-weights small-variance arms when $\tilde{n}_i^{\text{coarse}}$ is small.

Parallelism. Algorithm 2 completes in exactly 3 rounds at every T ; Algorithm 1 uses at most ~ 20 rounds for all $T \leq 10^6$, matching the $O(\log T)$ bound of Lemma 13. In deployments where each round incurs latency τ (e.g.

A/B testing, clinical trials), the total wall-clock cost is $O(\tau \log T)$ and $O(\tau)$ for our algorithms versus $\Theta(\tau T)$ for the sequential baselines (Table 3).

8 CONCLUSION

We develop near-optimal algorithms for estimating the means of multiple distributions under a fixed sample budget, ensuring uniform confidence intervals. Utilizing an explore-then-commit policy, our parallelizable algorithms balance exploration and exploitation. As part of our analysis, we provide sample complexity bounds for variance estimation for moment-bounded distributions and subgaussian distributions. We prove that some dependence $\sigma_{\min}^{-1}$ for subgaussian distributions and moment-bounded distributions is necessary, and characterize a large class of distributions where this $\sigma_{\min}$ -dependence can be avoided, offering a matching lower bound for $G = 2$. Extending these bounds to $G \geq 2$ remains an open question. Our work advances multi-group mean estimation theory and offers practical solutions for scenarios with unknown variances, limited budgets, and parallelization potential, with promising extensions in statistical learning.

References

- Daron Acemoglu, Ali Makhdoumi, Azarakhsh Malekian, and Asu Ozdaglar. Too much data: Prices and inefficiencies in data markets. *American Economic Journal: Microeconomics*, 14(4):218–256, 2022.
- András Antos, Varun Grover, and Csaba Szepesvári. Active learning in multi-armed bandits. In *Algorithmic Learning Theory: 19th International Conference, ALT 2008, Budapest, Hungary, October 13-16, 2008. Proceedings 19*, pages 287–302. Springer, 2008.
- P Auer. Finite-time analysis of the multiarmed bandit problem, 2002.
- Abdellah Aznag, Rachel Cummings, and Adam N Elmachtoub. An active learning framework for multi-group mean estimation. *Advances in Neural Information Processing Systems*, 36, 2024.
- Abdellah Aznag, Rachel Cummings, and Adam N. Elmachtoub. An active learning framework for multi-group mean estimation, 2025. URL <https://arxiv.org/abs/2505.14882>.
- Samuel Barbosa, Dan Cosley, Amit Sharma, and Roberto M Cesar Jr. Averaging gone wrong: Using time-aware analyses to better understand behavior. In *Proceedings of the 25th International Conference on World Wide Web*, pages 829–841, 2016.
- Djallel Bouneffouf, Romain Laroche, Tanguy Urvoy, Raphael Féraud, and Robin Allesiardo. Contextual bandit for active learning: Active thompson sampling. In *Neural Information Processing: 21st International Conference, ICONIP 2014, Kuching, Malaysia, November 3-6, 2014. Proceedings, Part I 21*, pages 405–412. Springer, 2014.
- Sébastien Bubeck, Nicolo Cesa-Bianchi, et al. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 5(1):1–122, 2012.
- Sébastien Bubeck, Nicolo Cesa-Bianchi, and Gábor Lugosi. Bandits with heavy tail. *IEEE Transactions on Information Theory*, 59(11):7711–7717, 2013.
- Alexandra Carpentier, Alessandro Lazaric, Mohammad Ghavamzadeh, Rémi Munos, and Peter Auer. Upper-confidence-bound algorithms for active learning in multi-armed bandits. In *International Conference on Algorithmic Learning Theory*, pages 189–203. Springer, 2011.
- Jeffrey Chan, Aldo Pacchiano, Nilesh Tripuraneni, Yun S Song, Peter Bartlett, and Michael I Jordan. Parallelizing contextual bandits. *arXiv preprint arXiv:2105.10590*, 2021.
- David A Cohn, Zoubin Ghahramani, and Michael I Jordan. Active learning with statistical models. *Journal of artificial intelligence research*, 4:129–145, 1996.
- Edith D De Leeuw. Choosing the method of data collection. In *International handbook of survey methodology*, pages 113–135. Routledge, 2012.
- Luc Devroye, Matthieu Lerasle, Gabor Lugosi, and Roberto I. Oliveira. Sub-Gaussian mean estimators. *The Annals of Statistics*, 44(6):2695 – 2725, 2016. doi: 10.1214/16-AOS1440. URL <https://doi.org/10.1214/16-AOS1440>.
- Luc Devroye, Abbas Mehrabian, and Tommy Reddad. The total variation distance between high-dimensional gaussians with the same mean. *arXiv preprint arXiv:1810.08693*, 2018.
- Aurélien Garivier, Tor Lattimore, and Emilie Kaufmann. On explore-then-commit strategies. *Advances in Neural Information Processing Systems*, 29, 2016.
- Tianyuan Jin, Pan Xu, Xiaokui Xiao, and Quanquan Gu. Double explore-then-commit: Asymptotic optimality and beyond. In *Conference on Learning Theory*, pages 2584–2633. PMLR, 2021.
- Amin Karbasi, Vahab Mirrokni, and Mohammad Shadravan. Parallelizing thompson sampling. *Advances in Neural Information Processing Systems*, 34:10535–10548, 2021.
- B Ravi Kiran, Ibrahim Sobh, Victor Talpaert, Patrick Mannion, Ahmad A Al Sallab, Senthil Yogamani, and Patrick Pérez. Deep reinforcement learning for autonomous driving: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 23(6):4909–4926, 2021.
- Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. Inherent trade-offs in the fair determination of risk scores. *arXiv preprint arXiv:1609.05807*, 2016.
- Yoshiaki Kuwata, Justin Teo, Gaston Fiore, Sertac Karaman, Emilio Frazzoli, and Jonathan P How. Real-time motion planning with applications to autonomous urban driving. *IEEE Transactions on control systems technology*, 17(5): 1105–1118, 2009.
- Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.
- Jasper CH Lee and Paul Valiant. Optimal sub-gaussian mean estimation in $\mathbb{R}$. In *Proceedings of the 2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 672–683. IEEE, 2022.
- Erich L. Lehmann and Joseph P. Romano. *Testing Statistical Hypotheses*. Springer, 3 edition, 2005.

- Andreas Maurer and Massimiliano Pontil. Empirical bernstein bounds and sample variance penalization. *arXiv preprint arXiv:0907.3740*, 2009.
- Dana Pessach and Erez Shmueli. A review on fairness in machine learning. *ACM Computing Surveys (CSUR)*, 55(3):1–44, 2022.
- Yao-Feng Ren and Han-Ying Liang. On the best constant in marcinkiewicz–zygmund inequality. *Statistics & probability letters*, 53(3):227–233, 2001.
- Mitchell Nicholas Sarkies, K-A Bowles, EH Skinner, D Mitchell, Romi Haas, M Ho, Karen Salter, Kerry May, Donna Markham, L O’Brien, et al. Data collection methods in health services research. *Applied clinical informatics*, 6(01):96–109, 2015.
- Burr Settles. Active learning literature survey. *University of Wisconsin-Madison Department of Computer Sciences*, 2009.
- Terence Tao. Lecture notes 1 for 247a. UCLA Math 247A lecture notes (Fourier analysis / harmonic analysis), 2006. URL <https://www.math.ucla.edu/~tao/247a.1.06f/notes1.pdf>. See Remark 5.7 for the log-convexity of L^p norms (inequality (6)).
- Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, 2009.
- Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.
- Liang Yu, Shuqi Qin, Meng Zhang, Chao Shen, Tao Jiang, and Xiaohong Guan. A review of deep reinforcement learning for smart building energy management. *IEEE Internet of Things Journal*, 8(15):12046–12063, 2021.

Supplementary Material : Explore, Refine, then Commit: Nearly Optimal Multi-Group Mean Estimation with Active Learning

Shourya Pandey¹

Syamantak Kumar¹

¹ Department of Computer Science, UT Austin

The Appendix is organized as follows,

- Section A contains deferred proofs from Section 3, with sections A.1, A.2 and A.3 discussing subgaussian, bounded moment and hypercontractive distributions respectively
- Section B contains proofs for moment-bounded and subgaussian distributions from Section 4.1
- Section C contains proofs for Hypercontractive Distributions from Section 4.2
- Section D contains proofs of lower bounds from Section 5
- Section E provides additional implementation details and information about code release

A VARIANCE ESTIMATION

A.1 VARIANCE ESTIMATION FOR SUBGAUSSIAN DISTRIBUTIONS

Lemma 6. Let $\mathcal{D} := \{X_i\}_{i \in [n]}$ be iid ν^2 -subgaussian random variables following Definition (1), with variance σ^2 . Let $\hat{\sigma}^2 := \frac{1}{n} \sum_{i=1}^n X_i^2 - \left(\frac{1}{n} \sum_{i=1}^n X_i\right)^2$ be the empirical variance estimated using $\mathcal{D}$. Then for $\delta \in (0, 1)$, with probability at least $1 - \delta$,

$$|\hat{\sigma}^2 - \sigma^2| \leq \nu^2 \left(\sqrt{\frac{4 \log(4/\delta)}{n}} + \frac{4 \log(4/\delta)}{n} \right).$$

Proof of Lemma 6. Let $\mu := \mathbb{E}[X]$. Then,

$$\hat{\sigma}^2 := \frac{1}{n} \sum_{i \in [n]} X_i^2 - \left(\frac{1}{n} \sum_{i \in [n]} X_i \right)^2 = \frac{1}{n} \sum_{i \in [n]} (X_i - \mu)^2 - \left(\frac{1}{n} \sum_{i \in [n]} (X_i - \mu) \right)^2$$

i.e., $\hat{\sigma}$ is invariant to an additive transformation in the random variable. Therefore, WLOG, we can assume $\{X_i\}_{i \in [n]}$ to be zero mean. We first recall the definition of the sub-gaussian norm (Definition 2.5.6 from Vershynin [2018]) for random variable $X \in \mathbb{R}$, as

$$\|X\|_{\psi_2} := \inf \{t > 0 : \mathbb{E}[X^2/t^2] \leq 2\} \quad (13)$$

Using Proposition 2.5.2 from Vershynin [2018], it is immediate that for a subgaussian random variable following Definition 1, the subgaussian norm is given by $\Theta(\nu)$. Similarly, the sub-exponential norm (Definition 2.7.5 in Vershynin [2018]) of a random variable, $X \in \mathbb{R}$, is given as

$$\|X\|_{\psi_1} := \inf \{t > 0 : \mathbb{E}[|X|/t] \leq 2\} \quad (14)$$

Using Lemma 2.7.6 from [Vershynin \[2018\]](#), we have that for subgaussian random variable, $X, Y := X^2$ is subexponential with $\|Y\|_{\psi_1} := \|X\|_{\psi_2}^2$. Then, $\{X_i\}_{i \in [n]}$ are subexponential with $\|X_i\|_{\psi_1} \lesssim \nu^2$. Using Theorem 2.8.1 from [Vershynin \[2018\]](#), we therefore have

$$\left| \frac{1}{n} \sum_{i \in [n]} X_i^2 - \sigma^2 \right| \leq 2\nu^2 \max \left\{ \sqrt{\frac{\log(\frac{4}{\delta})}{n}}, \frac{\log(\frac{4}{\delta})}{n} \right\} \quad (15)$$

Furthermore, using Theorem 2.6.2 from [Vershynin \[2018\]](#), we have with probability at least $1 - \delta$,

$$\left| \frac{1}{n} \sum_{i \in [n]} X_i \right| \leq \nu \sqrt{\frac{\log(\frac{4}{\delta})}{n}} \quad (16)$$

The result then follows by triangle inequality and a union bound using (15) and (16). $\square$

To obtain strong concentration bounds for subgaussian distributions, we state and prove an important lemma about subgaussian distributions (Definition 1) which bounds the fourth moment in terms of the variance proxy ν^2 and the variance σ^2 . This fact is central in our proofs for subsequent upper and lower bound proofs for the subgaussian case, and may be of independent interest to the reader.

Lemma 7 (Fourth moment of a subgaussian). *Let Y be a mean-zero ν^2 -subgaussian random variable (Definition 1) with variance $\text{Var}(Y) = \sigma^2$. Then*

$$\mathbb{E}[Y^4] \leq 32 \sigma^2 \nu^2 (1 + \log(\nu/\sigma)).$$

Moreover, there exist universal constants $c, C > 0$ such that for every $0 < \sigma \leq \nu$ there exists a mean-zero random variable Y with $\text{Var}(Y) = \sigma^2$ whose optimal subgaussian proxy $\nu_\star^2(Y) := \inf\{s^2 : \mathbb{E}[e^{tY}] \leq e^{s^2 t^2/2} \forall t \in \mathbb{R}\}$ satisfies

$$c\nu^2 \leq \nu_\star^2(Y) \leq C\nu^2, \text{ and simultaneously } \mathbb{E}[Y^4] \geq c\sigma^2 \nu^2 (1 + \log(\nu/\sigma)).$$

Proof of Lemma 7. Since Y is ν^2 -subgaussian, the standard Chernoff bound yields for all $u \geq 0$,

$$\Pr(|Y| > u) \leq 2 \exp\left(-\frac{u^2}{2\nu^2}\right).$$

Also Chebyshev gives $\Pr(|Y| > u) \leq \sigma^2/u^2$ for all $u > 0$. Hence

$$\Pr(|Y| > u) \leq \min\left\{\frac{\sigma^2}{u^2}, 2e^{-u^2/(2\nu^2)}\right\}.$$

Using the tail integral identity

$$\mathbb{E}[Y^4] = 4 \int_0^\infty u^3 \Pr(|Y| > u) du,$$

fix

$$u_0 := \nu \sqrt{2 \log\left(\frac{2\nu^2}{\sigma^2}\right)}.$$

(Note $\sigma^2 \leq \nu^2$ so the logarithm is $\geq \log 2 > 0$.) Split the integral:

$$\mathbb{E}[Y^4] \leq \underbrace{4 \int_0^{u_0} u^3 \frac{\sigma^2}{u^2} du}_{:= I_1} + \underbrace{4 \int_{u_0}^\infty u^3 \cdot 2e^{-u^2/(2\nu^2)} du}_{:= I_2}$$

For the first term,

$$I_1 = 4\sigma^2 \int_0^{u_0} u du = 2\sigma^2 u_0^2.$$

For the second term, a standard calculation (substitution $s = u^2$ or integration by parts) gives

$$\int_{u_0}^\infty u^3 e^{-u^2/(2\nu^2)} du = \nu^2 (u_0^2 + 2\nu^2) e^{-u_0^2/(2\nu^2)}.$$

Therefore

$$I_2 = 8\nu^2(u_0^2 + 2\nu^2)e^{-u_0^2/(2\nu^2)}.$$

By definition of u_0 , we have $e^{-u_0^2/(2\nu^2)} = \exp(-\log(2\nu^2/\sigma^2)) = \sigma^2/(2\nu^2)$, so

$$I_2 = 8\nu^2(u_0^2 + 2\nu^2) \cdot \frac{\sigma^2}{2\nu^2} = 4\sigma^2 u_0^2 + 8\sigma^2 \nu^2.$$

Thus

$$\mathbb{E}[Y^4] \leq 6\sigma^2 u_0^2 + 8\sigma^2 \nu^2 = 12\sigma^2 \nu^2 \log\left(\frac{2\nu^2}{\sigma^2}\right) + 8\sigma^2 \nu^2.$$

Finally,

$$\log\left(\frac{2\nu^2}{\sigma^2}\right) = \log 2 + 2\log(\nu/\sigma) \leq 2(1 + \log(\nu/\sigma)), \implies \mathbb{E}[Y^4] \leq 32\sigma^2 \nu^2 (1 + \log(\nu/\sigma)),$$

as claimed.

We next provide the matching lower bound. Fix $r := \nu/\sigma \geq 1$ and $L := 1 + \log r$. If $r \leq e$, take $Y \sim \mathcal{N}(0, \sigma^2)$. Then $\nu_\star^2(Y) = \sigma^2$, so $\nu_\star^2(Y) \in [\nu^2/e^2, \nu^2]$ and $\mathbb{E}[Y^4] = 3\sigma^4 \geq (3/e^2)\sigma^2 \nu^2 \geq c\sigma^2 \nu^2 (1 + \log r)$ since $1 + \log r \leq 2$ on $[1, e]$.

Now assume $r > e$ (so $L \geq 2$). Define a 3-point “rare spike” distribution:

$$Y = \begin{cases} +a & \text{w.p. } p/2, \\ -a & \text{w.p. } p/2, \\ 0 & \text{w.p. } 1 - p, \end{cases} \quad p := \frac{1}{8r^2 L} = \frac{\sigma^2}{8\nu^2 L}, \quad a^2 := \frac{\sigma^2}{p} = 8\nu^2 L.$$

Then $\mathbb{E}[Y] = 0$ and $\text{Var}(Y) = \mathbb{E}[Y^2] = pa^2 = \sigma^2$. Also

$$\mathbb{E}[Y^4] = pa^4 = (pa^2)a^2 = \sigma^2 \cdot 8\nu^2 L = 8\sigma^2 \nu^2 (1 + \log(\nu/\sigma)),$$

which gives the desired lower bound.

It now remains to show the *subgaussian proxy is truly of order ν^2* .

For the lower bound on the optimal proxy, we have, for any s^2 -subgaussian random variable, the mgf condition implies the tail bound $\Pr(|Y| > u) \leq 2e^{-u^2/(2s^2)}$. Applying this at $u = a$ yields

$$p = \Pr(|Y| = a) \leq 2e^{-a^2/(2s^2)} \implies s^2 \geq \frac{a^2}{2\log(2/p)}.$$

Here $\log(2/p) = \log(16r^2 L) = 2\log r + \log(16L) \leq 4L$ for $r > e$ (since $\log r \leq L$ and $\log(16L) \leq 2L$ for $L \geq 2$). Therefore

$$\nu_\star^2(Y) \geq \frac{a^2}{2\log(2/p)} \geq \frac{8\nu^2 L}{2 \cdot 4L} = \nu^2.$$

For the upper bound on the optimal proxy, we set $K^2 := a^2/\log(1/p)$. Then

$$\mathbb{E}\left[e^{Y^2/K^2}\right] = (1 - p) + p e^{a^2/K^2} = (1 - p) + p \cdot \frac{1}{p} = 2 - p \leq 2.$$

By Proposition 2.5.2 [Vershynin \[2018\]](#), this implies Y is $(4K^2)$ -subgaussian, Thus $\nu_\star^2(Y) \leq 4K^2$.

Finally, since $\log(1/p) = \log(8r^2 L) = 2\log r + \log(8L) \geq \log r$ and $L = 1 + \log r \leq 2\log r$ for $r > e$, we get $\log(1/p) \asymp L$, hence

$$K^2 = \frac{a^2}{\log(1/p)} = \frac{8\nu^2 L}{\log(1/p)} \leq C'\nu^2,$$

so $\nu_\star^2(Y) \leq 4K^2 \leq C\nu^2$. This completes the proof. $\square$

Proof of Theorem 2. Since X is ν^2 -subgaussian, $\mu_4 \leq 3\nu^4$. Lemma 7 implies

$$\mu_4 \leq 32\sigma^2\nu^2 \log \frac{e\nu}{\sigma}.$$

By Proposition 1, with probability $1 - \delta$,

$$|\hat{\sigma}^2 - \sigma^2| \leq 2e\sqrt{6(\mu_4 + 3\sigma^4)}\sqrt{\frac{\log(1/\delta)}{n}} \leq 2e\sqrt{24\mu_4}\sqrt{\frac{\log(1/\delta)}{n}},$$

which implies the two bounds

$$|\hat{\sigma}^2 - \sigma^2| \leq 12e\sqrt{2}\nu^2\sqrt{\frac{\log(1/\delta)}{n}} = \Delta, \quad |\hat{\sigma}^2 - \sigma^2| \leq 32e\sqrt{3}\sigma\nu\sqrt{\log(e\nu/\sigma)}\sqrt{\frac{\log(1/\delta)}{n}}.$$

The RHS of the second bound, when viewed as a function of σ , is strictly increasing in $\sigma \in (0, e\nu]$. Moreover, since $n \geq 32e^2 \log(1/\delta)$,

$$\Delta = 12e\sqrt{2}\nu^2\sqrt{\frac{\log(1/\delta)}{n}} \leq \frac{12e\sqrt{2}\nu^2}{4e\sqrt{2}} \leq 3\nu^2,$$

so $\sigma \leq \sqrt{\hat{\sigma}^2 + \Delta} \leq \sqrt{\sigma^2 + 2\Delta} \leq \sqrt{7}\nu < e\nu$.

It follows that under the same $1 - \delta$ probability event,

$$|\hat{\sigma}^2 - \sigma^2| \leq 32e\sqrt{3}(\hat{\sigma}^2 + \Delta)^{1/2}\nu\sqrt{2\log \frac{e^2\nu^2}{\hat{\sigma}^2 + \Delta}}\sqrt{\frac{\log(1/\delta)}{n}}.$$

□

A.2 VARIANCE ESTIMATION FOR DISTRIBUTIONS WITH BOUNDED MOMENTS

Proof of Theorem 1. Recall from Proposition 1 that the median-of-means estimator satisfies

$$|\hat{\sigma}^2 - \sigma^2| \leq 2e\sqrt{6(\mu_4 + 3\sigma^4)}\sqrt{\frac{\log(1/\delta)}{n}} \leq 2e\sqrt{24\mu_4}\sqrt{\frac{\log(1/\delta)}{n}} \quad (17)$$

with probability at least $1 - \delta$.

We now bound μ_4 in terms of σ^2 and μ_k . By Holder's inequality (see Riesz-Thorin or Remark 5.7. of Tao [2006]),

$$\mathbb{E}[|Y|^4]^{1/4} \leq (\mathbb{E}[|Y|^2]^{1/2})^\alpha (\mathbb{E}[|Y|^k]^{1/k})^{1-\alpha}, \quad \text{where } \frac{1}{4} = \frac{\alpha}{2} + \frac{1-\alpha}{k}.$$

Solving yields $\alpha = \frac{k-4}{2(k-2)}$. Therefore, $\mu_4^{1/4} \leq \sigma^{\frac{k-4}{2(k-2)}} \mu_k^{\frac{1}{2(k-2)}}$, or

$$\mu_4 \leq \sigma^{\frac{2(k-4)}{k-2}} \mu_k^{\frac{2}{k-2}}.$$

Substituting this bound into equation (17), we get with probability at least $1 - \delta$ that

$$|\hat{\sigma}^2 - \sigma^2| \leq 2e\sqrt{24}\sigma^{\frac{k-4}{k-2}} \mu_k^{\frac{1}{k-2}} \sqrt{\frac{\log(1/\delta)}{n}}. \quad (18)$$

The last step is to obtain the empirical variance $\hat{\sigma}^2$ on the right hand side instead of the true variance σ^2 . For this, we can use equation (17) again, which gives

$$\begin{aligned} \sigma^{\frac{k-4}{k-2}} &\leq \left(\hat{\sigma}^2 + 2e\sqrt{24\mu_4}\sqrt{\frac{\log(1/\delta)}{n}} \right)^{\frac{k-4}{2(k-2)}} \\ &\leq (\hat{\sigma}^2)^{\frac{k-4}{2(k-2)}} + \left(2e\sqrt{24\mu_4}\sqrt{\frac{\log(1/\delta)}{n}} \right)^{\frac{k-4}{2(k-2)}} \\ &\leq (\hat{\sigma}^2)^{\frac{k-4}{2(k-2)}} + \left(2e\sqrt{24}\mu_k^{2/k}\sqrt{\frac{\log(1/\delta)}{n}} \right)^{\frac{k-4}{2(k-2)}}, \end{aligned}$$

where we used $\mu_4^{1/4} \leq \mu_k^{1/k}$ and $(u+v)^t \leq u^t + v^t$ for all non-negative u, v and $0 \leq t \leq 1$. Plugging in this bound into equation (18), we get the desired bound

$$\begin{aligned} |\hat{\sigma}^2 - \sigma^2| &\leq 2e\sqrt{24}\hat{\sigma}^{\frac{k-4}{k-2}} \mu_k^{\frac{1}{k-2}} \sqrt{\frac{\log(1/\delta)}{n}} + 2e\sqrt{24} \left(2e\sqrt{24}\mu_k^{2/k} \sqrt{\frac{\log(1/\delta)}{n}} \right)^{\frac{k-4}{2(k-2)}} \mu_k^{\frac{1}{k-2}} \sqrt{\frac{\log(1/\delta)}{n}} \\ &= 2e\sqrt{24}\hat{\sigma}^{\frac{k-4}{k-2}} \mu_k^{\frac{1}{k-2}} \sqrt{\frac{\log(1/\delta)}{n}} + (2e\sqrt{24})^{\frac{3k-8}{2(k-2)}} \mu_k^{\frac{2}{k}} \left(\frac{\log(1/\delta)}{n} \right)^{\frac{3k-8}{4(k-2)}} \\ &\leq 2e\sqrt{24}\hat{\sigma}^{\frac{k-4}{k-2}} \mu_k^{\frac{1}{k-2}} \sqrt{\frac{\log(1/\delta)}{n}} + (2e\sqrt{24})^{3/2} \mu_k^{\frac{2}{k}} \left(\frac{\log(1/\delta)}{n} \right)^{\frac{3k-8}{4(k-2)}}. \end{aligned}$$

□

Proof of Theorem 3. Let $\alpha := k/2 \in (1, 2]$. We avoid the unknown mean by pairing samples. Let $X_1, \dots, X_n$ be i.i.d., set $m := \lfloor n/2 \rfloor$, and define for $j \in [m]$

$$Z_j := \frac{(X_{2j-1} - X_{2j})^2}{2}.$$

Then $\mathbb{E}[Z_j] = \text{Var}(X) = \sigma^2$ because for i.i.d. X, X' we have $\mathbb{E}[(X - X')^2] = 2\text{Var}(X)$. Next, using $|a - b|^k \leq 2^{k-1}(|a|^k + |b|^k)$ with $a = X_{2j-1} - \mu$ and $b = X_{2j} - \mu$,

$$\mathbb{E}[Z_j^\alpha] = \frac{1}{2^{k/2}} \mathbb{E}[|X_{2j-1} - X_{2j}|^k] \leq \frac{1}{2^{k/2}} \cdot 2^{k-1} (\mathbb{E}|X - \mu|^k + \mathbb{E}|X - \mu|^k) \leq 2^{k/2} \mu_k.$$

Also, by $|u - v|^\alpha \leq 2^{\alpha-1}(|u|^\alpha + |v|^\alpha)$ and Jensen $(\mathbb{E}Z_j)^\alpha \leq \mathbb{E}Z_j^\alpha$ (since $\alpha \geq 1$),

$$\mathbb{E}[|Z_j - \sigma^2|^\alpha] \leq 2^{\alpha-1}(\mathbb{E}Z_j^\alpha + (\sigma^2)^\alpha) \leq 2^{\alpha-1}(\mathbb{E}Z_j^\alpha + \mathbb{E}Z_j^\alpha) \leq 2^\alpha \cdot 2^{k/2} \mu_k = 2^k \mu_k.$$

Now define a median-of-means estimator for $\theta = \mathbb{E}[Z_j] = \sigma^2$. Choose

$$B := \min \left\{ m, \left\lceil \frac{9}{2} \ln \frac{2}{\delta} \right\rceil \right\}, \quad s := \left\lfloor \frac{m}{B} \right\rfloor,$$

discard any leftover points so that we use exactly Bs values $Z_1, \dots, Z_{Bs}$, split them into B blocks of size s , and let

$$\bar{Z}_b := \frac{1}{s} \sum_{j \in \text{block } b} Z_j, \quad \hat{\sigma}^2 := \text{median}\{\bar{Z}_1, \dots, \bar{Z}_B\}.$$

Let $U_j := Z_j - \sigma^2$, so $\mathbb{E}[U_j] = 0$. For $\alpha \in (1, 2]$, the von Bahr-Esseen inequality gives $\mathbb{E}|\sum_{i=1}^s U_i|^\alpha \leq \sum_{i=1}^s \mathbb{E}|U_i|^\alpha$. Hence

$$\mathbb{E}[|\bar{Z}_b - \sigma^2|^\alpha] = s^{-\alpha} \mathbb{E}\left|\sum_{i=1}^s U_i\right|^\alpha \leq s^{-\alpha} \sum_{i=1}^s \mathbb{E}|U_i|^\alpha \leq s^{1-\alpha} \cdot 2^k \mu_k.$$

By Markov's inequality, for any $t > 0$,

$$\Pr(|\bar{Z}_b - \sigma^2| > t) \leq \frac{\mathbb{E}|\bar{Z}_b - \sigma^2|^\alpha}{t^\alpha} \leq \frac{2^k \mu_k}{s^{\alpha-1} t^\alpha}.$$

Choose

$$t := (6 \cdot 2^k \mu_k)^{1/\alpha} s^{-(\alpha-1)/\alpha} = 4 \cdot 6^{2/k} \mu_k^{2/k} s^{-(1-2/k)}.$$

Then $\Pr(|\bar{Z}_b - \sigma^2| > t) \leq 1/6$.

Let $I_b = \mathbf{1}\{|\bar{Z}_b - \sigma^2| > t\}$. The blocks are independent and $\Pr(I_b = 1) \leq 1/6$. By Hoeffding's inequality,

$$\Pr\left(\sum_{b=1}^B I_b \geq B/2\right) \leq \exp\left(-2B(1/2 - 1/6)^2\right) = \exp\left(-\frac{2B}{9}\right) \leq \delta,$$

where the last inequality uses $B \geq \lceil \frac{9}{2} \ln \frac{2}{\delta} \rceil$ (unless $B = m$, in which case the bound is only easier since $\exp(-2B/9)$ decreases in B). On the complementary event, strictly more than half the blocks satisfy $|\bar{Z}_b - \sigma^2| \leq t$, hence their median does too: $|\hat{\sigma}^2 - \sigma^2| \leq t$.

Since $s = \lfloor m/B \rfloor$ and $m = \lfloor n/2 \rfloor$, we have $s \geq m/(2B) \geq n/(8B)$ (using $\lfloor x \rfloor \geq x/2$ for $x \geq 1$ and $B \leq m$ by construction). Therefore

$$t = 4 \cdot 6^{2/k} \mu_k^{2/k} s^{-(1-2/k)} \leq 4 \cdot 6^{2/k} \mu_k^{2/k} \left(\frac{8B}{n} \right)^{1-2/k}.$$

With $B \leq \lceil \frac{9}{2} \ln \frac{2}{\delta} \rceil \leq 6 \ln(1/\delta)$ for $\delta \in (0, 1/2)$, and using $1 - 2/k \leq 1/2$ (since $k \leq 4$), we get

$$t \leq 4 \cdot 6^{2/k} \mu_k^{2/k} \left(\frac{48 \ln(1/\delta)}{n} \right)^{1-2/k} \leq 64 \cdot 6^{2/k} \mu_k^{2/k} \left(\frac{\ln(1/\delta)}{n} \right)^{1-2/k}.$$

Combining the steps proves that with probability at least $1 - \delta$,

$$|\hat{\sigma}^2 - \sigma^2| \leq 64 \cdot 6^{2/k} \mu_k^{2/k} \left(\frac{\ln(1/\delta)}{n} \right)^{1-2/k}.$$

□

A.3 VARIANCE ESTIMATION FOR HYPERCONTRACTIVE DISTRIBUTIONS

Lemma 8. Fix a $\delta \in (0, 1)$. Let $\{X_i\}_{i \in [n]}$ be iid random variables with variance σ^2 and fourth moment at most κ^4 (Definition 3). Let $B := \lfloor \min\{8 \log(e^{\frac{1}{8}}/\delta), n/2\} \rfloor$ and $N = \lfloor n/B \rfloor$. Define

$$\hat{\sigma}^2 := \text{Median} \{ \hat{\sigma}_i^2 \}_{i \in [B]},$$

$$\text{where } \hat{\sigma}_i^2 := \frac{\sum_{t=(i-1)N+1}^{iN} X_t^2}{N} - \left(\frac{\sum_{t=(i-1)N+1}^{iN} X_t}{N} \right)^2$$

Then, for $\alpha_k := 100C_k^2$ and $p_k := \min\{\frac{k-2}{k}, \frac{1}{2}\}$, with probability at least $1 - \delta$,

$$|\hat{\sigma}^2 - \sigma^2| \leq \alpha_k \sigma^2 \left(\frac{\log(e^{\frac{1}{8}}/\delta)}{n} \right)^{p_k}$$

Proof of Lemma 8. Note that $C_k \geq 1$, since by Holder's inequality, for $k > 2$,

$$\mathbb{E}[|X - \mu|^k]^{\frac{1}{k}} \geq \mathbb{E}[|X - \mu|^2]^{\frac{1}{2}}$$

Therefore, combining the results for $k \in (2, 4]$ from Lemma 9 with that of $k \geq 4$ from Lemma 10, we have that with probability at least $1 - \delta$, we have

$$|\hat{\sigma}_i^2 - \sigma^2| \leq \frac{\alpha_k \sigma^2}{\delta^{1-p_k} N^{p_k}}$$

where $\alpha_k := 100C_k^2$, $p_k := \min\{\frac{k-2}{k}, \frac{1}{2}\}$. Define random variables $\forall i \in [B]$, $Y_i = \mathbb{1}\left(|\hat{\sigma}_i^2 - \sigma^2| \geq \frac{\alpha_k \sigma^2}{(4N)^{p_k}}\right)$. Then, $\{Y_i\}_{i \in [B]}$ are iid bernoulli random variables with parameter $p \leq \frac{1}{4}$. Using Hoeffding's inequality for the tail of binomial distribution, similar to Lemma 2 in Bubeck et al. [2013], we obtain

$$\mathbb{P}\left(\hat{\sigma}^2 \geq \sigma^2 + \frac{\alpha_k \sigma^2}{(4N)^{p_k}}\right) = \mathbb{P}\left(\sum_{i \in [B]} Y_i \geq \frac{k}{2}\right) \leq \exp\left(-2k \left(\frac{1}{2} - p\right)^2\right) \leq \exp\left(-\frac{k}{8}\right) = \delta$$

A similar bound holds for the left tail, yielding the result.

□

Remark 2 (Self-Normalizing Behaviour of Hypercontractive distributions). *Lemma 8 shows that it is possible to construct an estimator $\hat{\sigma}$ (the Median-of-means estimator) of the variance σ for a hypercontractive distribution such that with probability $(1 - \delta)$, $|\hat{\sigma}^2 - \sigma^2| \lesssim \sigma^2 \sqrt{\frac{\log(1/\delta)}{n}}$. In this case, the deviation in the variance estimate is proportional to the true variance σ^2 itself. However, this does not hold for moment-bounded distributions and ν^2 -subgaussian distributions, where a dependence on μ_k or ν is necessary. Finally, note that hypercontractive distributions include the family of Gaussian distributions because they satisfy Definition 3 with $k = 4$ and $C_k = 3$ Vershynin [2018].*

Lemma 9. *Let $\mathcal{D} := \{X_i\}_{i \in [n]}$ be iid (k, C_k) -hypercontractive random variables following Definition (3), with variance σ^2 and $k \in (2, 4]$. Let $\hat{\sigma}^2$ be the empirical variance estimate given as,*

$$\hat{\sigma}^2 := \frac{1}{n} \sum_{i \in [n]} X_i^2 - \left(\frac{1}{n} \sum_{i \in [n]} X_i \right)^2$$

Then, for any $\delta \in (0, 1)$, with probability atleast $1 - \delta$, we have

$$|\hat{\sigma}^2 - \sigma^2| \leq 100 (C_k^k + 1)^{\frac{2}{k}} \frac{\sigma^2}{\delta^{\frac{2}{k}} n^{\frac{k-2}{k}}}$$

Proof. Let $\mu := \mathbb{E}[X]$. Then,

$$\hat{\sigma}^2 := \frac{1}{n} \sum_{i \in [n]} X_i^2 - \left(\frac{1}{n} \sum_{i \in [n]} X_i \right)^2 = \frac{1}{n} \sum_{i \in [n]} (X_i - \mu)^2 - \left(\frac{1}{n} \sum_{i \in [n]} (X_i - \mu) \right)^2 \quad (19)$$

i.e, $\hat{\sigma}$ is invariant to an additive transformation in the random variable. Therefore WLOG, we assume $\{X_i\}_{i \in [n]}$ are mean zero. Define $\hat{\theta}^2 := \frac{1}{n} \sum_{i \in [n]} X_i^2$ and $\hat{\mu}^2 := \left(\frac{1}{n} \sum_{i \in [n]} X_i \right)^2$. Let $t, b > 0$. Then using a union-bound,

$$\mathbb{P}(\hat{\theta}^2 - \sigma^2 > b) \leq \mathbb{P}(\exists i \in [n] : |X_i^2 - \sigma^2| \geq t) + \mathbb{P}\left(\frac{1}{n} \sum_{i \in [n]} (X_i^2 - \sigma^2) \mathbb{1}(|X_i^2 - \sigma^2| \leq t) > b\right) \quad (20)$$

For the first term, we have using a union bound and Markov's inequality for the $\frac{k}{2}$ th exponent,

$$\mathbb{P}(\exists i \in [n] : |X_i^2 - \sigma^2| \geq t) \leq n \times \frac{\mathbb{E}[|X_i^2 - \sigma^2|^{\frac{k}{2}}]}{t^{\frac{k}{2}}} \quad (21)$$

For the second term, we have using Chebyshev's inequality,

$$\begin{aligned} & \mathbb{P}\left(\frac{1}{n} \sum_{i \in [n]} (X_i^2 - \sigma^2) \mathbb{1}(|X_i^2 - \sigma^2| \leq t) \geq b\right) \\ & \leq \frac{1}{b^2} \mathbb{E}\left[\left(\frac{1}{n} \sum_{i \in [n]} (X_i^2 - \sigma^2) \mathbb{1}(|X_i^2 - \sigma^2| \leq t)\right)^2\right] \\ & \leq \frac{\mathbb{E}[(X_i^2 - \sigma^2)^2 \mathbb{1}(|X_i^2 - \sigma^2| \leq t)]}{nb^2} + \frac{(\mathbb{E}[(X_i^2 - \sigma^2) \mathbb{1}(|X_i^2 - \sigma^2| \leq t)])^2}{b^2} \\ & = \frac{\mathbb{E}[(X_i^2 - \sigma^2)^2 \mathbb{1}(|X_i^2 - \sigma^2| \leq t)]}{nb^2} + \frac{(\mathbb{E}[(X_i^2 - \sigma^2) \mathbb{1}(|X_i^2 - \sigma^2| \geq t)])^2}{b^2} \end{aligned} \quad (22)$$

We have,

$$\begin{aligned} \mathbb{E}[(X_i^2 - \sigma^2)^2 \mathbb{1}(|X_i^2 - \sigma^2| \leq t)] &= \mathbb{E}\left[(X_i^2 - \sigma^2)^{\frac{k}{2}} (X_i^2 - \sigma^2)^{2-\frac{k}{2}} \mathbb{1}(|X_i^2 - \sigma^2| \leq t)\right] \\ &\leq t^{2-\frac{k}{2}} \mathbb{E}[|X_i^2 - \sigma^2|^{\frac{k}{2}}] \end{aligned} \quad (23)$$

Using Holder's inequality with $r = \frac{k}{2}, s = \frac{k}{k-2}$ such that $\frac{1}{r} + \frac{1}{s} = 1$,

$$\begin{aligned} \mathbb{E} [(X_i^2 - \sigma^2) \mathbb{1} (|X_i^2 - \sigma^2| \geq t)] &\leq \mathbb{E} [|X_i^2 - \sigma^2|^r]^{\frac{1}{r}} \mathbb{P} (|X_i^2 - \sigma^2| \geq t)^{\frac{1}{s}} \\ &\leq \mathbb{E} [|X_i^2 - \sigma^2|^{\frac{k}{2}}]^{\frac{2}{k}} \left(\frac{\mathbb{E} [|X_i^2 - \sigma^2|^{\frac{k}{2}}]}{t^{\frac{k}{2}}} \right)^{\frac{k-2}{k}} \\ &= \frac{\mathbb{E} [|X_i^2 - \sigma^2|^{\frac{k}{2}}]}{t^{\frac{k-2}{2}}} \end{aligned} \quad (25)$$

Using (24) and (25) in (23), we have

$$\mathbb{P} \left(\frac{1}{n} \sum_{i \in [n]} (X_i^2 - \sigma^2) \mathbb{1} (|X_i^2 - \sigma^2| \leq t) \geq b \right) \leq \frac{t^{2-\frac{k}{2}} \mathbb{E} [|X_i^2 - \sigma^2|^{\frac{k}{2}}]}{nb^2} + \frac{\mathbb{E} [|X_i^2 - \sigma^2|^{\frac{k}{2}}]}{b^2 t^{k-2}} \quad (26)$$

Substituting (21), (26) in (20), we have

$$\mathbb{P} (\hat{\theta}^2 - \sigma^2 > b) \leq \frac{n \mathbb{E} [|X_i^2 - \sigma^2|^{\frac{k}{2}}]}{t^{\frac{k}{2}}} + \frac{t^{2-\frac{k}{2}} \mathbb{E} [|X_i^2 - \sigma^2|^{\frac{k}{2}}]}{nb^2} + \frac{\mathbb{E} [|X_i^2 - \sigma^2|^{\frac{k}{2}}]}{b^2 t^{k-2}}$$

Set $t := nb$ to get

$$\mathbb{P} (\hat{\theta}^2 - \sigma^2 > b) \leq \frac{2 \mathbb{E} [|X_i^2 - \sigma^2|^{\frac{k}{2}}]}{n^{\frac{k}{2}-1} b^{\frac{k}{2}}} + \left(\frac{\mathbb{E} [|X_i^2 - \sigma^2|^{\frac{k}{2}}]}{n^{\frac{k}{2}-1} b^{\frac{k}{2}}} \right)^2$$

Therefore, if $\frac{\mathbb{E} [|X_i^2 - \sigma^2|^{\frac{k}{2}}]}{n^{\frac{k}{2}-1} b^{\frac{k}{2}}} < 1$, then

$$\mathbb{P} (\hat{\theta}^2 - \sigma^2 > b) \leq \frac{3 \mathbb{E} [|X_i^2 - \sigma^2|^{\frac{k}{2}}]}{n^{\frac{k}{2}-1} b^{\frac{k}{2}}}$$

Note that using Definition 3,

$$\mathbb{E} [|X_i^2 - \sigma^2|^{\frac{k}{2}}] \leq 2^{\frac{k}{2}-1} (\mathbb{E} [|X_i|^k] + \sigma^k) \leq 2^{\frac{k}{2}-1} (C_k^k + 1) \sigma^k$$

Therefore,

$$\mathbb{P} (\hat{\theta}^2 - \sigma^2 > b) \leq \frac{3 (C_k^k + 1) \sigma^k}{\left(\frac{n}{2}\right)^{\frac{k}{2}-1} b^{\frac{k}{2}}} \quad (27)$$

Next, using Ren and Liang [2001], we have using $k \geq 2$, from the Marcinkiewicz–Zygmund inequality,

$$\begin{aligned} \mathbb{P} (\hat{\mu}^2 > b) &= \mathbb{P} \left(\left| \sum_{i \in [n]} \frac{X_i}{n} \right| \geq \sqrt{b} \right) \leq \frac{\mathbb{E} \left[\left| \sum_{i \in [n]} X_i \right|^k \right]}{b^{\frac{k}{2}} n^k} \\ &\leq (3\sqrt{2})^k k^{\frac{k}{2}} n^{\frac{k}{2}-1} \frac{\sum_{i \in [n]} \mathbb{E} [|X_i|^k]}{b^{\frac{k}{2}} n^k} \\ &= (3\sqrt{2})^k k^{\frac{k}{2}} \frac{\mathbb{E} [|X|^k]}{b^{\frac{k}{2}} n^{\frac{k}{2}}} \\ &\leq (3\sqrt{2})^k k^{\frac{k}{2}} \frac{C_k^k \sigma^k}{b^{\frac{k}{2}} n^{\frac{k}{2}}} \end{aligned} \quad (28)$$

Therefore, with probability atleast $1 - \delta$, using (27) with a symmetrical argument for the left tail,

$$|\theta^2 - \sigma^2| \leq 2\sigma^2 \frac{(3(C_k^k + 1))^{\frac{2}{k}}}{\left(\frac{n}{2}\right)^{\frac{k-2}{k}}} \leq 10\sigma^2 \frac{(C_k^k + 1)^{\frac{2}{k}}}{\delta^{\frac{2}{k}} n^{\frac{k-2}{k}}} \quad (29)$$

and from (28), with probability atleast $1 - \delta$,

$$\hat{\mu}^2 \leq \frac{18kC_k^2\sigma^2}{n\delta^{\frac{2}{k}}} \leq \frac{72C_k^2\sigma^2}{n\delta^{\frac{2}{k}}} \leq 90(C_k^k + 1)^{\frac{2}{k}} \frac{\sigma^2}{\delta^{\frac{2}{k}} n^{\frac{k-2}{k}}} \quad (30)$$

The result follows using a union bound and triangle inequality over (29) and (30) along with (19). $\square$

Lemma 10. *Let $\mathcal{D} := \{X_i\}_{i \in [n]}$ be iid (k, C_k) -hypercontractive random variables following Definition (3), with variance σ^2 and $k \geq 4$. Let $\hat{\sigma}^2$ be the empirical variance estimate given as,*

$$\hat{\sigma}^2 := \frac{1}{n} \sum_{i \in [n]} X_i^2 - \left(\frac{1}{n} \sum_{i \in [n]} X_i \right)^2$$

Then, for any $\delta \in (0, 1)$, with probability atleast $1 - \delta$, we have

$$|\hat{\sigma}^2 - \sigma^2| \leq 2\sigma^2 \sqrt{\frac{C_k^4 + 1}{n}}$$

Proof. We proceed similar to Lemma 9. Let $\mu := \mathbb{E}[X]$. Then,

$$\hat{\sigma}^2 := \frac{1}{n} \sum_{i \in [n]} X_i^2 - \left(\frac{1}{n} \sum_{i \in [n]} X_i \right)^2 = \frac{1}{n} \sum_{i \in [n]} (X_i - \mu)^2 - \left(\frac{1}{n} \sum_{i \in [n]} (X_i - \mu) \right)^2 \quad (31)$$

i.e., $\hat{\sigma}$ is invariant to an additive transformation in the random variable. Therefore WLOG, we assume $\{X_i\}_{i \in [n]}$ are mean zero. Define $\hat{\theta}^2 := \frac{1}{n} \sum_{i \in [n]} X_i^2$ and $\hat{\mu}^2 := \left(\frac{1}{n} \sum_{i \in [n]} X_i \right)^2$. Let $b > 0$. Using Chebyshev's inequality, we have

$$\begin{aligned} \mathbb{P}(\hat{\theta}^2 - \sigma^2 \geq b) &= \mathbb{P}\left(\frac{1}{n} \sum_{i \in [n]} (X_i^2 - \sigma^2) \geq b\right) \\ &\leq \frac{1}{b^2 n^2} \sum_{i \in [n]} \mathbb{E}\left[(X_i^2 - \sigma^2)^2\right], \text{ since } \mathbb{E}[X_i^2 - \sigma^2] = 0 \\ &= \frac{n \mathbb{E}\left[(X^2 - \sigma^2)^2\right]}{b^2 n^2} \\ &= \frac{\mathbb{E}\left[(X^2 - \sigma^2)^2\right]}{b^2 n} \\ &\leq 2 \frac{\mathbb{E}[X^4] + \sigma^4}{nb^2} \\ &\leq 2 \frac{\sigma^4 (C_k^4 + 1)}{nb^2} \end{aligned} \quad (32)$$

where we used the fact that if X is (k, C_k) -hypercontractive for $k \geq 4$, then it is also $(4, C_k)$ -hypercontractive using

Holder's inequality. Similarly, we have using $\forall i \in [n], \mathbb{E}[X_i] = 0$,

$$\begin{aligned}
\mathbb{P}(\hat{\mu}^2 > b) &\leq \frac{\mathbb{E}[\hat{\mu}^4]}{b^2} = \frac{\mathbb{E}\left[\left(\sum_{i \in [n]} X_i\right)^4\right]}{b^2 n^4} = \frac{\sum_{i \in [n]} \mathbb{E}[X_i^4] + 3 \sum_{i \neq j \in [n]} \mathbb{E}[X_i^2] \mathbb{E}[X_j^2]}{b^2 n^4} \\
&\leq \frac{n \mathbb{E}[X^4] + 6n^2 \sigma^4}{b^2 n^4} \\
&\leq \frac{n C_k^4 \sigma^4 + 6n^2 \sigma^4}{b^2 n^4} \\
&= \left(\frac{C_k^4}{n} + 1\right) \frac{\sigma^4}{n^2 b^2} \leq \frac{\sigma^4 (C_k^4 + 1)}{n b^2}
\end{aligned} \tag{33}$$

The result then follows from (32), (33) along with a union bound. $\square$

B ALGORITHMS FOR MOMENT-BOUNDED DISTRIBUTIONS

B.1 REWARDS HAVE BOUNDED k -TH MOMENT FOR SOME $k > 4$

By Theorem 1, there is an event $\mathcal{E}$ such that with probability at least $1 - \delta$,

$$|\hat{\sigma}_{i,t}^2 - \sigma^2| \leq 27\sqrt{\mu_4} \sqrt{\frac{\log(GT/\delta)}{n}} \leq 27\mu_k^{2/k} \sqrt{\frac{\log(GT/\delta)}{n}} \quad (34)$$

and that

$$|\hat{\sigma}_{i,t}^2 - \sigma_i^2| \leq 27\hat{\sigma}_{i,t}^{\frac{k-4}{k-2}} \mu_k^{\frac{1}{k-2}} \sqrt{\frac{\log(GT/\delta)}{t}} + 137\mu_k^{\frac{2}{k}} \left(\frac{\log(GT/\delta)}{t} \right)^{\frac{3k-8}{4(k-2)}} \quad (35)$$

for all arms $i \in [G]$ and times $2 \leq t \leq T$, where $k > 4$ and μ_k is a bound on the k th moment of the reward distributions.

Let

$$\bar{\sigma}_{i,t}^2 = \hat{\sigma}_{i,t}^2 + 27\hat{\sigma}_{i,t}^{\frac{k-4}{k-2}} \mu_k^{\frac{1}{k-2}} \sqrt{\frac{\log(GT/\delta)}{t}} + 137\mu_k^{\frac{2}{k}} \left(\frac{\log(GT/\delta)}{t} \right)^{\frac{3k-8}{4(k-2)}} \quad (36)$$

and

$$\underline{\sigma}_{i,t}^2 = \max \left(\hat{\sigma}_{i,t}^2 - 27\hat{\sigma}_{i,t}^{\frac{k-4}{k-2}} \mu_k^{\frac{1}{k-2}} \sqrt{\frac{\log(GT/\delta)}{t}} - 137\mu_k^{\frac{2}{k}} \left(\frac{\log(GT/\delta)}{t} \right)^{\frac{3k-8}{4(k-2)}}, 0 \right), \quad (37)$$

which are lower and upper bounds on σ_i^2 (under the event $\mathcal{E}$).

For the rest of the proof, we operate under the event $\mathcal{E}$. We first show that the algorithm always terminates in finite number of rounds and the number of pulls being made line 19 of Algorithm 1 are non-negative and do not exceed T .

Lemma 11. *Under the event $\mathcal{E}$, for all $i \in [G]$ and $2 \leq t \leq T$,*

$$\hat{\sigma}_{i,t}^{\frac{k-4}{k-2}} \leq \sigma_i^{\frac{k-4}{k-2}} + \left(27\mu_k^{2/k} \sqrt{\frac{\log(1/\delta)}{n}} \right)^{\frac{k-4}{2(k-2)}}.$$

In particular,

$$\bar{\sigma}_{i,t}^2 \leq \sigma_i^2 + A_i t^{-\alpha} + B t^{-\beta}, \quad (38)$$

where $\alpha = 1/2$, $\beta = (3k-8)/(4k-8)$, and

$$A_i = 54\sigma_i^{\frac{k-4}{k-2}} \mu_k^{\frac{1}{k-2}} \sqrt{\log(GT/\delta)}, \quad B = 274\mu_k^{\frac{2}{k}} (\log(GT/\delta))^{\frac{3k-8}{4(k-2)}}. \quad (39)$$

Similarly, $\underline{\sigma}_{i,t}^2 \geq \min(\hat{\sigma}_{i,t}^2 - A_i t^{-\alpha} - B t^{-\beta}, 0)$.

Proof. The upper bound on $\hat{\sigma}_{i,t}$ follows from equation (34) and the fact that $(u+v)^\gamma \leq u^\gamma + v^\gamma$ for any $u, v \geq 0$ and $\gamma \in [0, 1]$. It follows that

$$\begin{aligned} 27\hat{\sigma}_{i,t}^{\frac{k-4}{k-2}} \mu_k^{\frac{1}{k-2}} \sqrt{\frac{\log(GT/\delta)}{t}} &\leq 27\sigma_i^{\frac{k-4}{k-2}} \mu_k^{\frac{1}{k-2}} \sqrt{\frac{\log(GT/\delta)}{t}} + 27^{\frac{3k-8}{2(k-2)}} \mu_k^{\frac{2}{k}} \left(\frac{\log(GT/\delta)}{n} \right)^{\frac{3k-8}{4(k-2)}} \\ &\leq 27\sigma_i^{\frac{k-4}{k-2}} \mu_k^{\frac{1}{k-2}} \sqrt{\frac{\log(GT/\delta)}{t}} + 137\mu_k^{\frac{2}{k}} \left(\frac{\log(GT/\delta)}{n} \right)^{\frac{3k-8}{4(k-2)}}. \end{aligned}$$

and we also know under the event $\mathcal{E}$ that

$$\hat{\sigma}_{i,t}^2 \leq \sigma_i^2 + 27\sigma_i^{\frac{k-4}{k-2}} \mu_k^{\frac{1}{k-2}} \sqrt{\frac{\log(GT/\delta)}{t}}.$$

The bounds on $\bar{\sigma}_{i,t}$ and $\underline{\sigma}_{i,t}$ follow by substituting this bound into equations (36) and (37). $\square$

Lemma 12. Under the event $\mathcal{E}$, for all $i \in [G]$ and $2 \leq t \leq T$, the estimate $\tilde{n}_{i,m}$ of line 10 of Algorithm 1 satisfies

$$\tilde{n}_{i,m} \geq \frac{\sigma_i^2}{\sum_{j \in [G]} (\sigma_j^2 + A_j m^{-\alpha} + B m^{-\beta})} T.$$

Proof. By Lemma 11,

$$\tilde{n}_{i,m} = \frac{\bar{\sigma}_{i,m}^2}{\sum_{j \in [G]} \bar{\sigma}_{j,m}^2} \cdot T \geq \frac{\sigma_i^2}{\sum_{j \in [G]} \bar{\sigma}_{j,m}^2} \cdot T \geq \frac{\sigma_i^2}{\sum_{j \in [G]} (\sigma_j^2 + A_j m^{-\alpha} + B m^{-\beta})} \cdot T.$$

□

Lemma 13. The value m_* of Algorithm 1 satisfies

$$m_* \leq \lfloor \frac{T}{4G} \rfloor$$

Proof. By assumption on T ,

$$T \geq 4G(1 + \lceil 32e^2 \log(GT/\delta) \rceil),$$

so the number of pulls before entering the while loop in Algorithm 1 is at most $\frac{T}{4G} - 1 < \lfloor (T/4G) \rfloor$. Since $\sum_{i \in [G]} n_{i,t} = T$ for all i , there exists index i such that

$$n_{i,t} \leq \frac{T}{G} \leq 4 \left(1 + \left\lfloor \frac{T}{4G} \right\rfloor \right).$$

Therefore, the condition in line 11 holds for some arm $i \in [G]$ when $m = 1 + \lfloor T/4G \rfloor$. It follows that the Algorithm enters the for-loop on line 17 with $m_* \leq \lfloor T/4G \rfloor$. □

In the next two Lemmas, we show that the coarse estimates n_j^{coarse} in line 18 are already a good approximation to the optimal values $n_j^* = \sigma_j^2 T / \|\Sigma\|_2^2$. Let i_* be any arm for which the condition in line 11 is true when $m = m_*$.

Lemma 14. Let i_* be any arm for which the condition in line 11 is true when $m = m_*$. Let

$$C_\alpha = \frac{\sum_{j \in [G]} A_j}{\|\Sigma\|_2^2}, \quad C_\beta = \frac{GB}{\|\Sigma\|_2^2}, \quad K = \frac{\sigma_{i_*}^2}{4 \|\Sigma\|_2^2}. \quad (40)$$

Then, the exploration parameter m_* of Algorithm 1 satisfies

$$m_* + 1 \geq \begin{cases} KT - (C_\alpha(KT)^{1-\alpha} + C_\beta(KT)^{1-\beta}) \geq \frac{KT}{2} & \text{if } T \geq \max \left(\frac{(4C_\alpha)^{1/\alpha}}{K}, \frac{(4C_\beta)^{1/\beta}}{K} \right), \\ \min \left\{ \left(\frac{KT}{4C_\alpha} \right)^{\frac{1}{1-\alpha}}, \left(\frac{KT}{4C_\beta} \right)^{\frac{1}{1-\beta}} \right\} & \text{otherwise.} \end{cases}$$

Proof. Since the condition in line 11 is true when $m = m_* + 1$ and $i = i_*$,

$$4(m_* + 1) > \tilde{n}_{i_*, m_* + 1}.$$

For the remainder of this proof, let $m := m_* + 1$. By Lemma 12,

$$4m > \frac{\sigma_{i_*}^2}{\sum_{j \in [G]} (\sigma_j^2 + A_j m^{-\alpha} + B m^{-\beta})} T$$

which implies

$$m + \frac{\sum_{j \in [G]} A_j}{\|\Sigma\|_2^2} m^{1-\alpha} + \frac{GB}{\|\Sigma\|_2^2} m^{1-\beta} > \frac{\sigma_{i_*}^2}{4 \|\Sigma\|_2^2} T.$$

Write the above inequality as

$$\underbrace{m + C_\alpha m^{1-\alpha} + C_\beta m^{1-\beta}}_{f(m)} > KT.$$

Here, we take cases:

Case 1: $T \geq \max\left(\frac{(4C_\alpha)^{1/\alpha}}{K}, \frac{(4C_\beta)^{1/\beta}}{K}\right)$

Then, $KT \geq 4C_\alpha^{1/\alpha}(KT)^{1-\alpha}$ and $KT \geq 4C_\beta^{1/\beta}(KT)^{1-\beta}$. We show that

$$m > KT - (C_\alpha(KT)^{1-\alpha} + C_\beta(KT)^{1-\beta}) \geq \frac{KT}{2}.$$

Indeed,

$$\begin{aligned} f(KT - (C_\alpha(KT)^{1-\alpha} + C_\beta(KT)^{1-\beta})) &= KT - (C_\alpha(KT)^{1-\alpha} + C_\beta(KT)^{1-\beta}) + C_\alpha(KT - (C_\alpha(KT)^{1-\alpha} + C_\beta(KT)^{1-\beta}))^{1-\alpha} \\ &\quad + C_\beta(KT - (C_\alpha(KT)^{1-\alpha} + C_\beta(KT)^{1-\beta}))^{1-\beta} \\ &\leq KT - (C_\alpha(KT)^{1-\alpha} + C_\beta(KT)^{1-\beta}) + C_\alpha(KT)^{1-\alpha} + C_\beta(KT)^{1-\beta} \\ &= KT < f(m) \end{aligned}$$

and f is strictly increasing on the non-negative reals.

Case 2: $T < \max\left(\frac{(4C_\alpha)^{1/\alpha}}{K}, \frac{(4C_\beta)^{1/\beta}}{K}\right)$ Then, either $(KT)^\alpha < 4C_\alpha$ or $(KT)^\beta < 4C_\beta$.

We claim that

$$m > \underline{m} := \min\left\{\frac{KT}{4}, \left(\frac{KT}{4C_\alpha}\right)^{\frac{1}{1-\alpha}}, \left(\frac{KT}{4C_\beta}\right)^{\frac{1}{1-\beta}}\right\}.$$

Indeed,

$$C_\alpha \underline{m}^{1-\alpha} \leq C_\alpha \left(\frac{KT}{4C_\alpha}\right)^{\frac{1-\alpha}{1-\alpha}} = \frac{KT}{4}, \quad C_\beta \underline{m}^{1-\beta} \leq C_\beta \left(\frac{KT}{4C_\beta}\right)^{\frac{1-\beta}{1-\beta}} = \frac{KT}{4}.$$

Combining all inequalities, we get

$$f(\underline{m}) \leq \frac{3KT}{4} < f(m),$$

so $m > \underline{m}$.

The first term $\frac{KT}{4}$ is greater than at least one of the other two terms in the expression for $\underline{m}$. Indeed one of

$$T < \frac{(4C_\alpha)^{1/\alpha}}{K} \implies T < \frac{4C_\alpha^{1/\alpha}}{K} \implies \frac{KT}{4} > \left(\frac{KT}{4C_\alpha}\right)^{\frac{1}{1-\alpha}}$$

and

$$T < \frac{(4C_\beta)^{1/\beta}}{K} \implies T < \frac{4C_\beta^{1/\beta}}{K} \implies \frac{KT}{4} > \left(\frac{KT}{4C_\beta}\right)^{\frac{1}{1-\beta}}$$

is realized. □

Lemma 15. Let C_α, C_β, K be the constants in equation (40). The coarse estimates n_j^{coarse} satisfy the bound

$$n_j^{\text{coarse}} \geq \begin{cases} \frac{\sigma_j^2}{2\|\Sigma\|_2^2} \cdot T & \text{if } T \geq \max\left(\frac{(4C_\alpha)^{1/\alpha}}{K}, \frac{(4C_\beta)^{1/\beta}}{K}\right), \\ \frac{\sigma_j^2}{\|\Sigma\|_2^2(2C_\alpha \underline{m}^{1-\alpha} + 2C_\beta \underline{m}^{1-\beta})} \cdot T & \text{otherwise,} \end{cases}$$

where $\underline{m} = \min\left\{\left(\frac{KT}{4C_\alpha}\right)^{\frac{1}{1-\alpha}}, \left(\frac{KT}{4C_\beta}\right)^{\frac{1}{1-\beta}}\right\}$.

Proof. Since $\tilde{n}_{j,m_*} \geq 4m_*$ for all arms $j \in [G]$, Lemma 12 implies

$$n_j^{\text{coarse}} = \tilde{n}_{j,m_*} \geq \frac{\sigma_j^2}{\sum_{i \in [G]} (\sigma_i^2 + A_i m_*^{-\alpha} + B m_*^{-\beta})} \cdot T = \frac{\sigma_j^2}{\|\Sigma\|_2^2 (1 + C_\alpha m_*^{-\alpha} + C_\beta m_*^{-\beta})} \cdot T.$$

Here, we take cases:

Case 1: $T \geq \max\left(\frac{(4C_\alpha)^{1/\alpha}}{K}, \frac{(4C_\beta)^{1/\beta}}{K}\right)$

Then, Lemma 14 implies that $m_* + 1 \geq \frac{KT}{2}$. Since $m_* \geq 4$, $m_* \geq \frac{4(m_*+1)}{5}$. So,

$$m_* \geq \frac{2KT}{5}.$$

Therefore, for any $j \in [G]$,

$$\begin{aligned} \tilde{n}_{j,m_*} &\geq \frac{\sigma_j^2}{\|\Sigma\|_2^2 (1 + C_\alpha(2KT/5)^{-\alpha} + C_\beta(2KT/5)^{-\beta})} \cdot T \\ &\geq \frac{\sigma_j^2}{\|\Sigma\|_2^2 \left(1 + \frac{(5/2)^\alpha}{4} + \frac{(5/2)^\beta}{4}\right)} \cdot T \geq \frac{\sigma_j^2}{2\|\Sigma\|_2^2} \cdot T. \end{aligned}$$

Case 2: $T < \max\left(\frac{(4C_\alpha)^{1/\alpha}}{K}, \frac{(4C_\beta)^{1/\beta}}{K}\right)$.

Then, Lemma 14 and $m_* \geq \frac{4(m_*+1)}{5}$ imply

$$m_* \geq \frac{4}{5} \min \left\{ \left(\frac{KT}{4C_\alpha} \right)^{\frac{1}{1-\alpha}}, \left(\frac{KT}{4C_\beta} \right)^{\frac{1}{1-\beta}} \right\}.$$

Now,

$$C_\alpha m_*^{-\alpha} \geq C_\alpha \left(\frac{4}{5} \left(\frac{KT}{4C_\alpha} \right)^{\frac{1}{1-\alpha}} \right)^{-\alpha} = \left(\frac{5}{4} \right)^\alpha \cdot \left(\frac{KT}{4C_\alpha^{1/\alpha}} \right)^{\frac{-\alpha}{1-\alpha}}, \quad C_\beta m_*^{-\beta} \geq \left(\frac{5}{4} \right)^\beta \cdot \left(\frac{KT}{4C_\beta^{1/\beta}} \right)^{\frac{-\beta}{1-\beta}}.$$

At least one of these values is greater than 1, which implies

$$n_j^{\text{coarse}} \geq \frac{\sigma_j^2}{\|\Sigma\|_2^2 (2C_\alpha m_*^{-\alpha} + 2C_\beta m_*^{-\beta})} \cdot T \geq \frac{\sigma_j^2}{\|\Sigma\|_2^2 (2C_\alpha \underline{m}^{-\alpha} + 2C_\beta \underline{m}^{-\beta})} \cdot T,$$

where $\underline{m} = \min \left\{ \left(\frac{KT}{4C_\alpha} \right)^{\frac{1}{1-\alpha}}, \left(\frac{KT}{4C_\beta} \right)^{\frac{1}{1-\beta}} \right\}$. The last inequality follows Lemma 14. □

We are ready to show that the coarse estimates n_j^{coarse} already provide a good regret bound for Algorithm 1.

Lemma 16 (Coarse Estimates have low regret). *Suppose Algorithm 1 enters line 27. Then, the regret of any arm $j \in [G]$ is bounded as follows:*

1. If $T \geq \max\left(\frac{(4C_\alpha)^{1/\alpha}}{K}, \frac{(4C_\beta)^{1/\beta}}{K}\right)$ then

$$\text{Reg}_i \leq \frac{(8\|\Sigma\|_2^2)^\alpha}{\sigma_{i_*}^{2\alpha} T^{1+\alpha}} \sum_{i \in [G]} A_i + \frac{(8\|\Sigma\|_2^2)^\beta}{\sigma_{i_*}^{2\beta} T^{1+\beta}} GB + \frac{64\|\Sigma\|_2^4}{15\sigma_j^2 T^2}.$$

2. If $T < \max\left(\frac{(4C_\alpha)^{1/\alpha}}{K}, \frac{(4C_\beta)^{1/\beta}}{K}\right)$ then

$$\text{Reg}_i \leq \frac{\sum_{i \in [G]} (A_i \underline{m}^{-\alpha} + B \underline{m}^{-\beta})}{T} + \frac{64 \left(\sum_{i \in [G]} (A_i \underline{m}^{-\alpha} + B \underline{m}^{-\beta}) \right)^2}{15\sigma_j^2 T^2}.$$

Proof. If Algorithm 1 enters line 27 then pulls each arm exactly $\lfloor n_j^{\text{coarse}} \rfloor$ many times. Moreover, any intermediate values such that $\lfloor n_j^{\text{coarse}}/2 \rfloor - (m_* + 1)$ are non-negative, because

$$n_j^{\text{coarse}} \geq 4m_* \implies \lfloor n_j^{\text{coarse}}/2 \rfloor \geq 2m_* \geq m_* + 1.$$

We have

$$\begin{aligned} \text{Reg}_j &= \frac{\sigma_j^2}{\lfloor n_j^{\text{coarse}} \rfloor} - \frac{\|\Sigma\|_2^2}{T} \\ &\leq \frac{\sigma_j^2}{n_j^{\text{coarse}} - 1} - \frac{\|\Sigma\|_2^2}{T} = \left(\frac{\sigma_j^2}{n_j^{\text{coarse}}} - \frac{\|\Sigma\|_2^2}{T} \right) + \left(\frac{\sigma_j^2}{n_j^{\text{coarse}} - 1} - \frac{\sigma_j^2}{n_j^{\text{coarse}}} \right) \end{aligned}$$

By Lemma 12 and that $n_j^{\text{coarse}} \geq 32e^2 \log(GT^3) \geq 16$, it follows that

$$\text{Reg}_j \leq \frac{\sum_{i \in [G]} (A_i m_*^{-\alpha} + B m_*^{-\beta})}{T} + \frac{16\sigma_j^2}{15(n_j^{\text{coarse}})^2}.$$

Case 1: $T \geq \max\left(\frac{(4C_\alpha)^{1/\alpha}}{K}, \frac{(4C_\beta)^{1/\beta}}{K}\right)$

By Lemma 14,

$$\begin{aligned} \frac{\sum_{i \in [G]} (A_i m_*^{-\alpha} + B m_*^{-\beta})}{T} &\leq \frac{\sum_{i \in [G]} (A_i (KT/2)^{-\alpha} + B (KT/2)^{-\beta})}{T} \\ &= \frac{\sum_{i \in [G]} A_i}{T} \left(\frac{\sigma_{i_*}^2 T}{8 \|\Sigma\|_2^2} \right)^{-\alpha} + \frac{GB}{T} \left(\frac{\sigma_{i_*}^2 T}{8 \|\Sigma\|_2^2} \right)^{-\beta} \\ &= \frac{(8 \|\Sigma\|_2^2)^\alpha}{\sigma_{i_*}^{2\alpha} T^{1+\alpha}} \sum_{i \in [G]} A_i + \frac{(8 \|\Sigma\|_2^2)^\beta}{\sigma_{i_*}^{2\beta} T^{1+\beta}} GB. \end{aligned}$$

By Lemma 15,

$$\frac{16\sigma_j^2}{15(n_j^{\text{coarse}})^2} \leq \frac{16\sigma_j^2}{15 \left(\frac{\sigma_j^2 T}{2 \|\Sigma\|_2^2} \right)^2} = \frac{64 \|\Sigma\|_2^4}{15\sigma_j^2 T^2}$$

Combining the two bounds,

$$\text{Reg}_i \leq \frac{(8 \|\Sigma\|_2^2)^\alpha}{\sigma_{i_*}^{2\alpha} T^{1+\alpha}} \sum_{i \in [G]} A_i + \frac{(8 \|\Sigma\|_2^2)^\beta}{\sigma_{i_*}^{2\beta} T^{1+\beta}} GB + \frac{64 \|\Sigma\|_2^4}{15\sigma_j^2 T^2}.$$

Case 2: $T \geq \max\left(\frac{(4C_\alpha)^{1/\alpha}}{K}, \frac{(4C_\beta)^{1/\beta}}{K}\right)$

Then, recall from Lemma 14 and 15 that

$$m_* \geq \underline{m} = \min \left\{ \left(\frac{KT}{4C_\alpha} \right)^{\frac{1}{1-\alpha}}, \left(\frac{KT}{4C_\beta} \right)^{\frac{1}{1-\beta}} \right\}, \quad n_j^{\text{coarse}} \geq \frac{\sigma_j^2}{\|\Sigma\|_2^2 (2C_\alpha \underline{m}^{-\alpha} + 2C_\beta \underline{m}^{-\beta})} \cdot T.$$

Therefore,

$$\begin{aligned} \text{Reg}_i &\leq \frac{\sum_{i \in [G]} (A_i \underline{m}^{-\alpha} + B \underline{m}^{-\beta})}{T} + \frac{16\sigma_j^2}{15 \left(\frac{\sigma_j^2}{\|\Sigma\|_2^2 (2C_\alpha \underline{m}^{-\alpha} + 2C_\beta \underline{m}^{-\beta})} \cdot T \right)^2} \\ &= \frac{\sum_{i \in [G]} (A_i \underline{m}^{-\alpha} + B \underline{m}^{-\beta})}{T} + \frac{64 \left(\sum_{i \in [G]} (A_i \underline{m}^{-\alpha} + B \underline{m}^{-\beta}) \right)^2}{15\sigma_j^2 T^2}. \end{aligned}$$

□

In particular, if T is sufficiently large and $\alpha = 1/2$, then the leading term of the regret Reg_i is

$$\text{Reg}_i \leq \tilde{O}\left(\frac{1}{\sigma_{\min} T^{1.5}}\right).$$

This already achieves the correct $T^{-1.5}$ dependence in T . The dependence of the regret in the other parameters (especially $\sigma_{\min}$) can be improved by using the finer estimates n_j^{fine} .

First, we show that for sufficiently large T (and if event $\mathcal{E}$ holds), then Algorithm 1 enters line 32.

Lemma 17 (Fine Estimates are used for large T). *Suppose*

$$T \geq \max\left(\frac{10 \|\Sigma\|_2^2}{\sigma_{\min}^2} \cdot \left(\frac{432 \mu_k^{\frac{1}{k-2}} \sqrt{\log(GT/\delta)}}{\sigma_{\min}^{\frac{k}{k-2}}}\right)^{1/\alpha}, \frac{10 \|\Sigma\|_2^2}{\sigma_{\min}^2} \cdot \left(\frac{2192 \mu_k^{\frac{2}{k}} (\log(GT/\delta))^{\frac{3k-8}{4(k-2)}}}{\sigma_{\min}^2}\right)^{1/\beta}\right).$$

Then, $n_j^{\text{fine}} \geq \lfloor n_j^{\text{coarse}}/2 \rfloor$ for all $j \in [G]$ and Algorithm 1 enters line 32.

Proof. The given bound on T implies that for all $j \in [G]$,

$$A_j(2KT/5)^{-\alpha} \leq \frac{\sigma_j^2}{8}, \quad B(2KT/5)^{-\beta} \leq \frac{\sigma_j^2}{8}.$$

First, we upper bound n_j^{coarse} as follows:

$$n_j^{\text{coarse}} = \frac{\bar{\sigma}_{j,m}^2}{\sum_{i \in [G]} \bar{\sigma}_{i,m_*}^2} \cdot T \leq \frac{\sigma_j^2 + A_j m_*^{-\alpha} + B m_*^{-\beta}}{\sum_{i \in [G]} \sigma_i^2} \cdot T.$$

Since $m_* \geq 4$ and $m_* + 1 \geq KT/2$ by Lemma 14, it follows that $m_* \geq 2KT/5$. It follows that

$$n_j^{\text{coarse}} \leq \frac{\sigma_j^2 + A_j(2KT/5)^{-\alpha} + B(2KT/5)^{-\beta}}{\|\Sigma\|_2^2} \cdot T \leq \frac{5\sigma_j^2 T}{4 \|\Sigma\|_2^2}.$$

and since $n_j^{\text{coarse}} \geq 16$ for all j ,

$$\lfloor n_j^{\text{coarse}}/2 \rfloor \leq \frac{9n_j^{\text{coarse}}}{16} \leq \frac{45\sigma_j^2 T}{64 \|\Sigma\|_2^2}.$$

Also, similar to Lemma 12,

$$\begin{aligned} n_j^{\text{fine}} &= \frac{\bar{\sigma}_{j,t_j}^2}{\sum_{i \in [G]} \bar{\sigma}_{i,t_i}^2} \cdot T \geq \frac{\sigma_j^2 T}{\sum_{i \in [G]} (\sigma_i^2 + A_i t_i^{-\alpha} + B t_i^{-\beta})} \\ &\geq \frac{\sigma_j^2 T}{\sum_{i \in [G]} (\sigma_i^2 + A_i(2KT/5)^{-\alpha} + B(2KT/5)^{-\beta})} \\ &\geq \frac{\sigma_j^2 T}{\sum_{i \in [G]} (5\sigma_i^2/4)} = \frac{4\sigma_j^2 T}{5 \|\Sigma\|_2^2} > \frac{45\sigma_j^2 T}{64 \|\Sigma\|_2^2} - 1, \end{aligned}$$

where the last line follows because $T > \frac{10 \|\Sigma\|_2^2}{\sigma_{\min}^2} \cdot 216^{1/\alpha} \geq \left(\frac{4}{5} - \frac{45}{64}\right) \frac{\|\Sigma\|_2^2}{\sigma_{\min}^2} T$.

Therefore, $\lfloor n_j^{\text{fine}} \rfloor \geq \lfloor n_j^{\text{coarse}}/2 \rfloor$, as desired. $\square$

Lemma 18 (Fine Estimates have low regret). *Suppose Algorithm 1 enters line 32 and let $T \geq \max\left(\frac{(4C_\alpha)^{1/\alpha}}{K}, \frac{(4C_\beta)^{1/\beta}}{K}\right)$. Then, the regret of any arm $j \in [G]$ is bounded as*

$$\text{Reg}_j \leq \frac{54 \sqrt{\log(GT/\delta)} (4 \|\Sigma\|_2^2)^\alpha}{T^{1+\alpha}} \cdot \mu_k^{\frac{1}{k-2}} \sum_{i \in [G]} \frac{1}{\sigma_i^{2\alpha - \frac{k-4}{k-2}}} + \frac{274 \log(GT/\delta)^{\frac{3k-8}{4k-8}} (4 \|\Sigma\|_2^2)^\beta}{T^{1+\beta}} \mu_k^{\frac{2}{k}} \sum_{i \in [G]} \frac{1}{\sigma_i^{2\beta}} + \frac{512 \|\Sigma\|_2^4}{7\sigma_j^2 T^2}.$$

Proof. Similar to Lemma 12,

$$n_j^{\text{fine}} = \frac{\bar{\sigma}_{j,t_j}^2}{\sum_{i \in [G]} \bar{\sigma}_{i,t_i}^2} \cdot T \geq \frac{\sigma_j^2 T}{\sum_{i \in [G]} (\sigma_i^2 + A_i t_i^{-\alpha} + B t_i^{-\beta})}.$$

The regret can be written as

$$\begin{aligned} \text{Reg}_j &= \frac{\sigma_j^2}{\lfloor n_j^{\text{fine}} \rfloor} - \frac{\|\Sigma\|_2^2}{T} \\ &\leq \frac{\sigma_j^2}{n_j^{\text{fine}} - 1} - \frac{\|\Sigma\|_2^2}{T} = \left(\frac{\sigma_j^2}{n_j^{\text{fine}}} - \frac{\|\Sigma\|_2^2}{T} \right) + \left(\frac{\sigma_j^2}{n_j^{\text{fine}} - 1} - \frac{\sigma_j^2}{n_j^{\text{fine}}} \right) \\ &\leq \frac{\sum_{i \in [G]} (A_i t_i^{-\alpha} + B t_i^{-\beta})}{T} + \frac{4\sigma_j^2}{(n_j^{\text{coarse}}/2)(n_j^{\text{coarse}}/2 - 1)} \\ &\leq \frac{\sum_{i \in [G]} (A_i t_i^{-\alpha} + B t_i^{-\beta})}{T} + \frac{128\sigma_j^2}{7(n_j^{\text{coarse}})^2}, \end{aligned}$$

where the two inequalities follow from $n_j^{\text{fine}} \geq n_j^{\text{coarse}}/2$ and $n_j^{\text{coarse}} \geq 32e^2 \log(GT^3) \geq 16$. We bound both terms separately.

By Lemma 15 $t_i \geq n_i^{\text{coarse}}/2 \geq \sigma_i^2 T/4 \|\Sigma\|_2^2$, which implies

$$\begin{aligned} &\frac{\sum_{i \in [G]} (A_i t_i^{-\alpha} + B t_i^{-\beta})}{T} \\ &\leq \frac{\sum_{i \in [G]} \left(A_i \left(\frac{\sigma_i^2 T}{4 \|\Sigma\|_2^2} \right)^{-\alpha} + B \left(\frac{\sigma_i^2 T}{4 \|\Sigma\|_2^2} \right)^{-\beta} \right)}{T} \\ &\leq \sum_{i \in [G]} 54 \sigma_i^{\frac{k-4}{k-2}} \mu_k^{\frac{1}{k-2}} \sqrt{\log(GT/\delta)} \left(\frac{\sigma_i^2 T}{4 \|\Sigma\|_2^2} \right)^{-\alpha} T^{-1} + \sum_{i \in [G]} 274 \mu_k^{\frac{2}{k}} (\log(GT/\delta))^{\frac{3k-8}{4(k-2)}} \left(\frac{\sigma_i^2 T}{4 \|\Sigma\|_2^2} \right)^{-\beta} T^{-1} \\ &= \frac{54 \sqrt{\log(GT/\delta)} (4 \|\Sigma\|_2^2)^\alpha}{T^{1+\alpha}} \cdot \mu_k^{\frac{1}{k-2}} \sum_{i \in [G]} \frac{1}{\sigma_i^{2\alpha - \frac{k-4}{k-2}}} + \frac{274 \log(GT/\delta)^{\frac{3k-8}{4k-8}} (4 \|\Sigma\|_2^2)^\beta}{T^{1+\beta}} \mu_k^{\frac{2}{k}} \sum_{i \in [G]} \frac{1}{\sigma_i^{2\beta}}. \end{aligned}$$

The second term can be bounded as

$$\frac{128\sigma_j^2}{7(n_j^{\text{coarse}})^2} \leq \frac{128\sigma_j^2}{7 \left(\frac{\sigma_j^2 T}{2 \|\Sigma\|_2^2} \right)^2} \leq \frac{512 \|\Sigma\|_2^4}{7\sigma_j^2 T^2}.$$

Combining the two bounds gives the result. $\square$

If T is sufficiently large and $\alpha = 1/2$, the leading term of the regret of the algorithm is

$$\text{Reg} \leq \tilde{O} \left(\frac{1}{\sigma_{\min}^{\frac{2}{k-2}} T^{1.5}} \right),$$

which has an improved dependence on $\sigma_{\min}$ (compared to the coarse regret which depends on the regret as $\sigma_{\min}^{-1}$).

B.2 REWARDS ARE SUBGAUSSIAN

By Theorem 2, there is an event $\mathcal{E}$ such that with probability at least $1 - \delta$,

$$|\hat{\sigma}_{i,t}^2 - \sigma^2| \leq 12e\sqrt{2}\nu^2 \sqrt{\frac{\log(GT/\delta)}{n}} = \Delta \quad (41)$$

and that

$$|\hat{\sigma}_{i,t}^2 - \sigma_i^2| \leq 32e\sqrt{3}(\hat{\sigma}_{i,t}^2 + \Delta)^{1/2}\nu\sqrt{2\log\frac{e^2\nu^2}{\hat{\sigma}_{i,t}^2 + \Delta}}\sqrt{\frac{\log(GT/\delta)}{n}} \quad (42)$$

for all arms $i \in [G]$ and times $2 \leq t \leq T$.

Let

$$\bar{\sigma}_{i,t}^2 = \hat{\sigma}_{i,t}^2 + 32e\sqrt{3}(\hat{\sigma}_{i,t}^2 + \Delta)^{1/2}\nu\sqrt{2\log\frac{e^2\nu^2}{\hat{\sigma}_{i,t}^2 + \Delta}}\sqrt{\frac{\log(GT/\delta)}{n}} \quad (43)$$

and

$$\underline{\sigma}_{i,t}^2 = \max\left(\hat{\sigma}_{i,t}^2 - 32e\sqrt{3}(\hat{\sigma}_{i,t}^2 + \Delta)^{1/2}\nu\sqrt{2\log\frac{e^2\nu^2}{\hat{\sigma}_{i,t}^2 + \Delta}}\sqrt{\frac{\log(GT/\delta)}{n}}, 0\right), \quad (44)$$

which are lower and upper bounds on σ_i^2 (under the event $\mathcal{E}$). The expression for $\bar{\sigma}_{i,t}^2$ in equation (43) is used throughout in Algorithm 1 for the subgaussian case. The rest of the proof has exactly the same structure as subsection B.1; we state the lemmas that are all analogous to the lemmas in that subsection.

Lemma 19. *Under the event $\mathcal{E}$, for all $i \in [G]$ and $\lceil 32e^2 \log(GT/\delta) \rceil \leq t \leq T$,*

$$\bar{\sigma}_{i,t}^2 \leq \sigma_i^2 + A_i t^{-\alpha} + B t^{-\beta}, \quad (45)$$

where $\alpha = 1/2, \beta = 3/4$, and

$$A_i = 302\sigma_i\nu\sqrt{\log\frac{e\nu}{\sigma_i}}\sqrt{\log(GT/\delta)}, \quad B = 1024\nu^2\sqrt{\log\frac{e\nu}{\sigma_i}}(\log(GT/\delta))^{3/4}. \quad (46)$$

Proof. By equations (43) and (41),

$$\begin{aligned} \bar{\sigma}_{i,t}^2 &\leq \hat{\sigma}_{i,t}^2 + 32e\sqrt{3}(\sigma_i^2 + 2\Delta)^{1/2}\nu\sqrt{2\log\frac{e^2\nu^2}{\sigma_i^2 + 2\Delta}}\sqrt{\frac{\log(GT/\delta)}{t}} \\ &\leq \sigma_i^2 + 32e\sqrt{3}\sigma_i\nu\sqrt{2\log\frac{e^2\nu^2}{\sigma_i^2}}\sqrt{\frac{\log(GT/\delta)}{t}} + 32e\sqrt{3}(\sigma_i^2 + 2\Delta)^{1/2}\nu\sqrt{2\log\frac{e^2\nu^2}{\sigma_i^2 + 2\Delta}}\sqrt{\frac{\log(GT/\delta)}{t}} \end{aligned}$$

Using $(u + v)^{1/2} \leq u^{1/2} + v^{1/2}$ for non-negative reals u, v , it follows that

$$\begin{aligned} \bar{\sigma}_{i,t}^2 &\leq \sigma_i^2 + 64e\sqrt{3}\sigma_i\nu\sqrt{2\log\frac{e^2\nu^2}{\sigma_i^2}}\sqrt{\frac{\log(GT/\delta)}{t}} + 32e\sqrt{3}\Delta^{1/2}\nu\sqrt{2\log\frac{e^2\nu^2}{\sigma_i^2}}\sqrt{\frac{\log(GT/\delta)}{t}} \\ &\leq \sigma_i^2 + 64e\sqrt{3}\sigma_i\nu\sqrt{\log\frac{e\nu}{\sigma_i}}\sqrt{\frac{\log(GT/\delta)}{t}} + 32e\sqrt{3}\sqrt{12e\sqrt{2}\nu^2}\sqrt{\frac{\log(GT/\delta)}{t}}\nu\sqrt{\log\frac{e\nu}{\sigma_i}}\sqrt{\frac{\log(GT/\delta)}{t}} \\ &\leq \sigma_i^2 + 302\sigma_i\nu\sqrt{\log\frac{e\nu}{\sigma_i}}\sqrt{\frac{\log(GT/\delta)}{t}} + 1024\nu^2\sqrt{\log\frac{e\nu}{\sigma_i}}\left(\frac{\log(GT/\delta)}{t}\right)^{3/4}. \end{aligned}$$

□

Lemma 20. *Under the event $\mathcal{E}$, for all $i \in [G]$ and $\lceil 32e^2 \log(GT/\delta) \rceil \leq t \leq T$, the estimate $\tilde{n}_{i,m}$ of line 10 of Algorithm 1 satisfies*

$$\tilde{n}_{i,m} \geq \frac{\sigma_i^2}{\sum_{j \in [G]} (\sigma_j^2 + A_j m^{-\alpha} + B m^{-\beta})} T.$$

Proof. The proof is the same as that of Lemma 12. □

Lemma 21. The value m_* of Algorithm 1 satisfies

$$m_* \leq \lfloor \frac{T}{4G} \rfloor$$

Proof. The proof is the same as that of Lemma 13. □

Lemma 22. Let i_* be any arm for which the condition in line 11 is true when $m = m_*$. Let

$$C_\alpha = \frac{\sum_{j \in [G]} A_j}{\|\Sigma\|_2^2}, \quad C_\beta = \frac{GB}{\|\Sigma\|_2^2}, \quad K = \frac{\sigma_{i_*}^2}{4\|\Sigma\|_2^2}. \quad (47)$$

Then, the exploration parameter m_* of Algorithm 1 satisfies

$$m_* + 1 \geq \begin{cases} KT - (C_\alpha(KT)^{1-\alpha} + C_\beta(KT)^{1-\beta}) \geq \frac{KT}{2} & \text{if } T \geq \max\left(\frac{(4C_\alpha)^{1/\alpha}}{K}, \frac{(4C_\beta)^{1/\beta}}{K}\right), \\ \min\left\{\left(\frac{KT}{4C_\alpha}\right)^{\frac{1}{1-\alpha}}, \left(\frac{KT}{4C_\beta}\right)^{\frac{1}{1-\beta}}\right\} & \text{otherwise.} \end{cases}$$

Proof. The proof is the same as that of Lemma 14. □

Lemma 23. Let C_α, C_β, K be the constants in equation (47). The coarse estimates n_j^{coarse} satisfy the bound

$$n_j^{\text{coarse}} \geq \begin{cases} \frac{\sigma_j^2}{2\|\Sigma\|_2^2} \cdot T & \text{if } T \geq \max\left(\frac{(4C_\alpha)^{1/\alpha}}{K}, \frac{(4C_\beta)^{1/\beta}}{K}\right), \\ \frac{\sigma_j^2}{\|\Sigma\|_2^2(2C_\alpha \underline{m}^{-\alpha} + 2C_\beta \underline{m}^{-\beta})} \cdot T & \text{otherwise,} \end{cases}$$

$$\text{where } \underline{m} = \min\left\{\left(\frac{KT}{4C_\alpha}\right)^{\frac{1}{1-\alpha}}, \left(\frac{KT}{4C_\beta}\right)^{\frac{1}{1-\beta}}\right\}.$$

Proof. The proof is the same as that of Lemma 15. □

The next result shows that the coarse estimates n_j^{coarse} achieve low regret.

Lemma 24 (Coarse Estimates have low regret). Suppose Algorithm 1 enters line 27. Then, the regret of any arm $j \in [G]$ is bounded as follows:

1. If $T \geq \max\left(\frac{(4C_\alpha)^{1/\alpha}}{K}, \frac{(4C_\beta)^{1/\beta}}{K}\right)$ then

$$\text{Reg}_i \leq \frac{(8\|\Sigma\|_2^2)^\alpha}{\sigma_{i_*}^2 T^{1+\alpha}} \sum_{i \in [G]} A_i + \frac{(8\|\Sigma\|_2^2)^\beta}{\sigma_{i_*}^{2\beta} T^{1+\beta}} GB + \frac{64\|\Sigma\|_2^4}{15\sigma_j^2 T^2}.$$

2. If $T < \max\left(\frac{(4C_\alpha)^{1/\alpha}}{K}, \frac{(4C_\beta)^{1/\beta}}{K}\right)$ then

$$\text{Reg}_i \leq \frac{\sum_{i \in [G]} (A_i \underline{m}^{-\alpha} + B \underline{m}^{-\beta})}{T} + \frac{64 \left(\sum_{i \in [G]} (A_i \underline{m}^{-\alpha} + B \underline{m}^{-\beta})\right)^2}{15\sigma_j^2 T^2}.$$

Proof. The proof is the same as that of Lemma 16. □

If T is sufficiently large and $\alpha = 1/2$, then the leading term of the regret Reg_i is

$$\text{Reg}_i \leq \tilde{O}\left(\frac{1}{\sigma_{\min} T^{1.5}}\right).$$

This already achieves the correct $T^{-1.5}$ dependence in T . The dependence of the regret in the other parameters (especially $\sigma_{\min}$) can be improved by using the finer estimates n_j^{fine} , similar to the bounded moment case.

Lemma 25 (Fine Estimates are used for large T). *Suppose*

$$T \geq \max \left(\frac{10 \|\Sigma\|_2^2}{\sigma_{\min}^2} \cdot \left(\frac{2416\nu}{\sigma_{\min}} \sqrt{\log \frac{e\nu}{\sigma_{\min}}} \sqrt{\log(GT/\delta)} \right)^{1/\alpha}, \frac{10 \|\Sigma\|_2^2}{\sigma_{\min}^2} \cdot \left(\frac{8192\nu^2}{\sigma_{\min}^2} \sqrt{\log \frac{e\nu}{\sigma_{\min}}} (\log(GT/\delta))^{3/4} \right)^{1/\beta} \right).$$

Then, $n_j^{\text{fine}} \geq \lfloor n_j^{\text{coarse}}/2 \rfloor$ for all $j \in [G]$ and Algorithm 1 enters line 32.

Proof. The proof is the same as that of Lemma 17. $\square$

Finally, we show that the fine estimates have low regret, with an improved dependence on $\sigma_{\min}$ compared to the regret of the coarse estimates.

Lemma 26 (Fine Estimates have low regret). *Suppose Algorithm 1 enters line 32 and let $T \geq \max \left(\frac{(4C_\alpha)^{1/\alpha}}{K}, \frac{(4C_\beta)^{1/\beta}}{K} \right)$. Then, the regret of any arm $j \in [G]$ is bounded as*

$$\text{Reg}_j \leq \frac{604\nu \sqrt{\log(GT/\delta)}}{T^{1.5}} \|\Sigma\|_2 \sum_{i \in [G]} \sqrt{\log \frac{e\nu}{\sigma_i}} + \frac{2900\nu^2 (\log(GT/\delta))^{3/4}}{T^{1.75}} \|\Sigma\|_2^{3/2} \sum_{i \in [G]} \frac{1}{\sigma_i^{3/2}} \sqrt{\log \frac{e\nu}{\sigma_i}} + \frac{512 \|\Sigma\|_2^4}{7\sigma_j^2 T^2}.$$

Proof. Similar to Lemma 18,

$$\text{Reg}_j \leq \frac{\sum_{i \in [G]} (A_i t_i^{-\alpha} + B t_i^{-\beta})}{T} + \frac{128\sigma_j^2}{7(n_j^{\text{coarse}})^2}.$$

We bound both terms separately.

By Lemma 23 $t_i \geq n_i^{\text{coarse}}/2 \geq \sigma_i^2 T/4 \|\Sigma\|_2^2$, which implies

$$\begin{aligned} & \frac{\sum_{i \in [G]} (A_i t_i^{-\alpha} + B t_i^{-\beta})}{T} \\ & \leq \frac{\sum_{i \in [G]} \left(A_i \left(\frac{\sigma_i^2 T}{4 \|\Sigma\|_2^2} \right)^{-\alpha} + B \left(\frac{\sigma_i^2 T}{4 \|\Sigma\|_2^2} \right)^{-\beta} \right)}{T} \\ & \leq \sum_{i \in [G]} 302\sigma_i \nu \sqrt{\log \frac{e\nu}{\sigma_i}} \sqrt{\log(GT/\delta)} \cdot \sqrt{\frac{4 \|\Sigma\|_2^2}{\sigma_i^2 T}} T^{-1} + \sum_{i \in [G]} 1024\nu^2 \sqrt{\log \frac{e\nu}{\sigma_i}} (\log(GT/\delta))^{3/4} \left(\frac{4 \|\Sigma\|_2^2}{\sigma_i^2 T} \right)^{3/4} T^{-1} \\ & \leq \frac{604\nu \sqrt{\log(GT/\delta)}}{T^{1.5}} \|\Sigma\|_2 \sum_{i \in [G]} \sqrt{\log \frac{e\nu}{\sigma_i}} + \frac{2900\nu^2 (\log(GT/\delta))^{3/4}}{T^{1.75}} \|\Sigma\|_2^{3/2} \sum_{i \in [G]} \frac{1}{\sigma_i^{3/2}} \sqrt{\log \frac{e\nu}{\sigma_i}}. \end{aligned}$$

The second term can be bounded as

$$\frac{128\sigma_j^2}{7(n_j^{\text{coarse}})^2} \leq \frac{128\sigma_j^2}{7 \left(\frac{\sigma_j^2 T}{2 \|\Sigma\|_2^2} \right)^2} \leq \frac{512 \|\Sigma\|_2^4}{7\sigma_j^2 T^2}.$$

Combining the two bounds gives the result. $\square$

If T is sufficiently large and $\alpha = 1/2$, the leading term of the regret Reg of the algorithm is

$$\text{Reg} \leq \tilde{O} \left(\frac{\sqrt{\log(e\nu/\sigma_{\min})}}{T^{1.5}} \right),$$

which depends only sub-logarithmically in $1/\sigma_{\min}$ (compared to the coarse regret which depends on the regret as $\sigma_{\min}^{-1}$.)

B.3 BOUNDED k -TH MOMENT FOR $2 < k \leq 4$

By Theorem 3, there is an event such that with probability at least $1 - \delta$,

$$|\hat{\sigma}_{i,t}^2 - \sigma_i^2| \leq \frac{C_k \mu_k^{2/k}}{t^\beta}, \quad (48)$$

for all arms $i \in [G]$ and times $t \in [T]$, where $\beta = 1 - 2/k$ and $C_k = 64 \cdot 6^{2/k} (\ln GT/\delta)^{1-2/k}$. Call this event $\mathcal{E}$.

Let

$$\bar{\sigma}_{i,t}^2 = \hat{\sigma}_{i,t}^2 + \frac{C_k \mu_k^{2/k}}{t^\beta} \quad (49)$$

and

$$\underline{\sigma}_{i,t}^2 = \max \left(\hat{\sigma}_{i,t}^2 - \frac{C_k \mu_k^{2/k}}{t^\beta}, 0 \right), \quad (50)$$

which are lower and upper bounds on σ_i^2 (under the event $\mathcal{E}$). Clearly, under this same event $\mathcal{E}$,

$$\bar{\sigma}_{i,t}^2 \leq \sigma_i^2 + Bt^{-\beta}, \quad (51)$$

where $B = 2C_k \mu_k^{2/k} = 128 \cdot 6^{2/k} \mu_k^{2/k} (\ln(GT/\delta))^{1-2/k}$.

For the rest of the proof, we operate under the event $\mathcal{E}$. We first show that the algorithm always terminates in finite number of rounds and the number of pulls being made line 19 of Algorithm 1 are non-negative and do not exceed T .

Lemma 27. *Under the event $\mathcal{E}$, for all $i \in [G]$ and $2 \leq t \leq T$, the estimate $\tilde{n}_{i,m}$ of line 10 of Algorithm 1 satisfies*

$$\tilde{n}_{i,m} \geq \frac{\sigma_i^2}{\sum_{j \in [G]} (\sigma_j^2 + Bm^{-\beta})} T.$$

Proof. By Lemma 11,

$$\tilde{n}_{i,m} = \frac{\bar{\sigma}_{i,m}^2}{\sum_{j \in [G]} \bar{\sigma}_{j,m}^2} \cdot T \geq \frac{\sigma_i^2}{\sum_{j \in [G]} \bar{\sigma}_{j,m}^2} \cdot T \geq \frac{\sigma_i^2}{\sum_{j \in [G]} (\sigma_j^2 + Bm^{-\beta})} \cdot T.$$

□

Lemma 28. *The value m_* of Algorithm 1 satisfies*

$$m_* \leq \lfloor \frac{T}{4G} \rfloor$$

Proof. The proof is the same as that of Lemma 13. □

Lemma 29. *Let i_* be any arm for which the condition in line 11 is true when $m = m_*$. Let*

$$C_\beta = \frac{GB}{\|\Sigma\|_2^2}, \quad K = \frac{\sigma_{i_*}^2}{4\|\Sigma\|_2^2}. \quad (52)$$

Then, the exploration parameter m_* of Algorithm 1 satisfies

$$m_* + 1 \geq \begin{cases} KT - C_\beta (KT)^{1-\beta} \geq \frac{KT}{2} & \text{if } T \geq \frac{(4C_\beta)^{1/\beta}}{K}, \\ \left(\frac{KT}{4C_\beta} \right)^{\frac{1}{1-\beta}} & \text{otherwise.} \end{cases}$$

Proof. The proof is the same as that of Lemma 14. □

Lemma 30. Let C_β, K be the constants in equation (52). The coarse estimates n_j^{coarse} satisfy the bound

$$n_j^{\text{coarse}} \geq \begin{cases} \frac{\sigma_j^2}{2\|\Sigma\|_2^2} \cdot T & \text{if } T \geq \frac{(4C_\beta)^{1/\beta}}{K}, \\ \frac{\sigma_j^2}{\|\Sigma\|_2^2 (2C_\beta \underline{m}^{-\beta})} \cdot T & \text{otherwise,} \end{cases}$$

where $\underline{m} = \left(\frac{KT}{4C_\beta}\right)^{\frac{1}{1-\beta}}$.

Proof. The proof is the same as that of Lemma 15. □

The next result shows that the coarse estimates n_j^{coarse} achieve low regret.

Lemma 31 (Coarse Estimates have low regret). Suppose Algorithm 1 enters line 27. Then, the regret of any arm $j \in [G]$ is bounded as follows:

1. If $T \geq \frac{(4C_\beta)^{1/\beta}}{K}$, then

$$\text{Reg}_i \leq \frac{(8\|\Sigma\|_2^2)^\beta}{\sigma_{i_*}^{2\beta} T^{1+\beta}} GB + \frac{64\|\Sigma\|_2^4}{15\sigma_j^2 T^2}.$$

2. If $T < \frac{(4C_\beta)^{1/\beta}}{K}$ then

$$\text{Reg}_i \leq \frac{GBm^{-\beta}}{T} + \frac{64(GBm^{-\beta})^2}{15\sigma_j^2 T^2}.$$

Proof. The proof is the same as that of Lemma 16. □

In particular, if T is sufficiently large and $\alpha = 1/2$, then the leading term of the regret Reg_i is

$$\text{Reg}_i \leq \tilde{O}\left(\frac{G}{\sigma_{\min}^{2-4/k} T^{2-2/k}}\right).$$

This already achieves the correct $T^{-1.5}$ dependence in T . The dependence of the regret in the other parameters can be improved moderately by using the finer estimates n_j^{fine} .

First, we show that for sufficiently large T (and if event $\mathcal{E}$ holds), then Algorithm 1 enters line 32.

Lemma 32 (Fine Estimates are used for large T). Suppose

$$T \geq \frac{10\|\Sigma\|_2^2}{\sigma_{\min}^2} \left(\frac{1024 \cdot 6^{2/k} (\ln(GT/\delta))^{1-2/k}}{\sigma_{\min}^2} \right)^{1/\beta}.$$

Then, $n_j^{\text{fine}} \geq \lfloor n_j^{\text{coarse}}/2 \rfloor$ for all $j \in [G]$ and Algorithm 1 enters line 32.

Proof. The proof is the same as that of Lemma 17. □

Finally, we show that the fine estimates have low regret, with an improved dependence on $\sigma_{\min}$ compared to the regret of the coarse estimates.

Lemma 33 (Fine Estimates have low regret). Suppose Algorithm 1 enters line 32 and let $T \geq \frac{(4C_\beta)^{1/\beta}}{K}$. Then, the regret of any arm $j \in [G]$ is bounded as

$$\text{Reg}_j \leq \frac{768(\log(GT/\delta))^{1-2/k} \mu_k^{2/k}}{T^{2-2/k}} \|\Sigma\|_2^{2-4/k} \sum_{i \in [G]} \frac{1}{\sigma_i^{2-4/k}} + \frac{512\|\Sigma\|_2^4}{7\sigma_j^2 T^2}.$$

Proof. Similar to Lemma 18,

$$\text{Reg}_j \leq \frac{\sum_{i \in [G]} B t_i^{-\beta}}{T} + \frac{128 \sigma_j^2}{7(n_j^{\text{coarse}})^2}.$$

We bound both terms separately.

By Lemma 30, $t_i \geq n_i^{\text{coarse}}/2 \geq \sigma_i^2 T/4 \|\Sigma\|_2^2$, which implies

$$\begin{aligned} \frac{\sum_{i \in [G]} B t_i^{-\beta}}{T} &\leq \frac{\sum_{i \in [G]} \left(B \left(\frac{\sigma_i^2 T}{4 \|\Sigma\|_2^2} \right)^{-\beta} \right)}{T} \\ &\leq \sum_{i \in [G]} 128 \cdot 6^{2/k} \mu_k^{2/k} (\log(GT/\delta))^{1-2/k} \left(\frac{4 \|\Sigma\|_2^2}{\sigma_i^2 T} \right)^\beta T^{-1} \\ &\leq \frac{768 (\log(GT/\delta))^{1-2/k} \mu_k^{2/k}}{T^{2-2/k}} \|\Sigma\|_2^{2-4/k} \sum_{i \in [G]} \frac{1}{\sigma_i^{2-4/k}}. \end{aligned}$$

The second term can be bounded as

$$\frac{128 \sigma_j^2}{7(n_j^{\text{coarse}})^2} \leq \frac{128 \sigma_j^2}{7 \left(\frac{\sigma_j^2 T}{2 \|\Sigma\|_2^2} \right)^2} \leq \frac{512 \|\Sigma\|_2^4}{7 \sigma_j^2 T^2}.$$

Combining the two bounds gives the result. □

C ALGORITHMS FOR HYPERCONTRACTIVE DISTRIBUTIONS

C.1 ANALYSIS FOR HIGHER MOMENT BOUNDS

Recall the idea of a good event $\mathcal{E}$ from Section 4.1. Assume that T rewards from each of the G arms have been sampled beforehand and revealed to the agent one by one. For each $i \in [G]$ and $2 \leq t \in [T]$, let $\mathcal{E}_{i,t}$ denote the event that the median-of-means estimator $\hat{\sigma}_{i,t}^2$ of σ_i^2 defined in Lemma 8 satisfies the concentration guaranteed by the Lemma. To improve the constants in the regret bound of the algorithm, we can use the optimal mean estimation algorithm of Lee and Valiant [2022]. Using their estimator $\hat{\sigma}_{i,t}^2$ for the variance, the event

$$\mathcal{E}_{i,t} := \left\{ |\hat{\sigma}_{i,t}^2 - \sigma_i^2| \leq C_k^2 \sigma_i^2 \sqrt{\frac{(2 + o(1)) \log(1/\delta)}{t}} \right\}.$$

happens with probability at least $1 - \delta$. By a union bound over all $i \in [G]$ and $2 \leq t \in [T]$, the event

$$\mathcal{E} := \bigcap_{i \in [G]} \bigcap_{2 \leq t \leq T} \mathcal{E}_{i,t} \quad (53)$$

holds with probability at least $1 - GT\delta$. In the rest of the proof, we treat the $2 + o(1)$ term in the events $\mathcal{E}_{i,t}$ as the constant 3. Let $C = 3 \log(1/\delta)$.

First, we show that the coarse variance estimates $\hat{\sigma}_{i,\text{coarse}}^2$ are multiplicatively close to the true variances σ_i^2 .

Lemma 34. *Under event $\mathcal{E}$, for all $i \in [G]$ the coarse estimate $\hat{\sigma}_{i,\text{coarse}}^2$ of σ_i^2 satisfies*

$$\frac{2}{3} \sigma_i^2 \leq \hat{\sigma}_{i,\text{coarse}}^2 \leq \frac{4}{3} \sigma_i^2.$$

Proof. Each arm is pulled exactly $\lceil 9C_k^4 C \rceil$ times in the exploration phase, where $C = 3 \log(1/\delta)$. Since the event $\mathcal{E}$ holds, we have for all $i \in [G]$ that

$$\left| \frac{\hat{\sigma}_{i,\text{coarse}}^2}{\sigma_i^2} - 1 \right| = \left| \frac{\sigma_{i, \lceil 9C_k^4 C \rceil}^2}{\sigma_i^2} - 1 \right| \leq C_k^2 \sqrt{\frac{C}{\lceil 9C_k^4 C \rceil}} \leq \frac{1}{3}.$$

□

Next, we show that the coarse estimates $\tilde{n}_i^{\text{coarse}}$ are multiplicatively close to $n_{i,T'}^* = \frac{\sigma_i^2}{\|\Sigma\|_2^2} T'$.

Lemma 35. *Under event $\mathcal{E}$, for all $i \in [G]$ the estimate $\tilde{n}_i^{\text{coarse}}$ of $n_{i,T'}^*$ satisfies*

$$\frac{n_{i,T'}^*}{2} \leq \tilde{n}_i^{\text{coarse}} \leq 2n_{i,T'}^*$$

Proof. By Lemma 34, $\frac{2\sigma_j^2}{3} \leq \hat{\sigma}_{j,\text{coarse}}^2 \leq \frac{4\sigma_j^2}{3}$ for all $j \in [G]$, which also implies $\frac{2\|\Sigma\|_2^2}{3} \leq \|\tilde{\Sigma}_{\text{coarse}}\|_2^2 \leq \frac{4\|\Sigma\|_2^2}{3}$. Therefore,

$$\frac{n_{i,T'}^*}{2} = \frac{2\sigma_i^2/3}{4\|\Sigma\|_2^2/3} T' \leq \frac{\hat{\sigma}_{i,\text{coarse}}^2}{\|\tilde{\Sigma}_{\text{coarse}}\|_2^2} T' = \tilde{n}_i^{\text{coarse}} \leq \frac{4\sigma_i^2/3}{2\|\Sigma\|_2^2/3} T' = 2n_{i,T'}^*.$$

□

Since $T \geq 2G\lceil 9C_k^4 C \rceil$, the number of pulls T' remaining is at least $T/2$. In the refinement phase, arm i is pulled $\lfloor \tilde{n}_i^{\text{coarse}}/6 \rfloor$ times. Therefore, total number of pulls in this refinement phase is bounded above by $T/6 < T'$.

Lemma 36. *Under event $\mathcal{E}$, for all $i \in [G]$ the fine estimate $\hat{\sigma}_{i,\text{fine}}^2$ of σ_i^2 satisfies*

$$\left| \frac{\hat{\sigma}_{i,\text{fine}}^2}{\sigma_i^2} - 1 \right| \leq C_k^2 \sqrt{\frac{6C}{\tilde{n}_i^{\text{coarse}}}}.$$

Proof. The inequality holds because each arm has been pulled exactly $\lceil 9C_k^4 C \rceil + \lfloor \tilde{n}_i^{\text{coarse}}/6 \rfloor > \tilde{n}_i^{\text{coarse}}/6$ times. □

The algorithm commits to the estimates $\tilde{n}_i^{\text{fine}}$ using these finer estimates of the variances. We show next that $\tilde{n}_i^{\text{fine}}$ is a very good estimate of $n_{i,T'}^*$.

Lemma 37. *For all $i \in [G]$ the fine estimate $\tilde{n}_i^{\text{fine}}$ of $n_{i,T'}^*$ satisfies*

$$\tilde{n}_i^{\text{fine}} \geq \frac{\sigma_i^2}{\|\Sigma\|_2^2 + \frac{2C_k^2 \sqrt{12C} \|\Sigma\|_1 \|\Sigma\|_2}{T'^{0.5}}} T' \geq \frac{n_{i,T'}^*}{2}.$$

Proof. By Lemma 36,

$$\tilde{n}_i^{\text{fine}} = \frac{\hat{\sigma}_{i,\text{fine}}^2 \left(1 + C_k^2 \sqrt{\frac{6C}{\tilde{n}_i^{\text{coarse}}}}\right)}{\sum_{j \in [G]} \hat{\sigma}_{j,\text{fine}}^2 \left(1 + C_k^2 \sqrt{\frac{6C}{\tilde{n}_j^{\text{coarse}}}}\right)} T' \geq \frac{\sigma_i^2}{\|\Sigma\|_2^2 + 2C_k^2 \sum_{j \in [G]} \sigma_j^2 \sqrt{\frac{6C}{\tilde{n}_j^{\text{coarse}}}}} T'$$

Using the bound $\tilde{n}_j^{\text{coarse}} \geq n_{j,T'}^*/2$ from Lemma 35,

$$\sum_{j \in [G]} \frac{\sigma_j^2}{\sqrt{\tilde{n}_j^{\text{coarse}}}} \leq \sum_{j \in [G]} \sqrt{\frac{2}{n_{j,T'}^*}} \sigma_j^2 = \sum_{j \in [G]} \frac{\sqrt{2} \sigma_j \|\Sigma\|_2}{T'^{0.5}} = \frac{\sqrt{2} \|\Sigma\|_1 \|\Sigma\|_2}{T'^{0.5}}.$$

The first inequality of the Lemma follows by combining the two bounds. To second inequality, it suffices to show that

$$\|\Sigma\|_2^2 \geq \frac{2C_k^2 \sqrt{12C} \|\Sigma\|_1 \|\Sigma\|_2}{T'^{0.5}}.$$

Rewriting the above inequality and using $\|\Sigma\|_1 \geq \|\Sigma\|_2$, it suffices to show that $T' \geq 48C_k^4 CG$, which is true because $T' \geq T/2$ and $T \geq 98C_k^4 CG$. $\square$

Finally, since each arm has been pulled at least $\lfloor \tilde{n}_i^{\text{fine}} \rfloor + \lceil 9C_k^4 C \rceil > \tilde{n}_i^{\text{fine}}$ times,

$$\begin{aligned} \text{Reg}(\pi) &\leq \max_{i \in [G]} \frac{\sigma_i^2}{\tilde{n}_i^{\text{fine}}} - \frac{\|\Sigma\|_2^2}{T} \\ &\leq \max_{i \in [G]} \left(\frac{\|\Sigma\|_2^2 + \frac{2C_k^2 \sqrt{12C} \|\Sigma\|_1 \|\Sigma\|_2}{T'^{0.5}}}{T'} - \frac{\|\Sigma\|_2^2}{T} \right) \\ &\leq \frac{2C_k^2 \sqrt{12C} \|\Sigma\|_1 \|\Sigma\|_2}{T'^{1.5}} + \frac{\|\Sigma\|_2^2 (T - T')}{TT'} \end{aligned}$$

Using the bounds $C = 2 \log(GT^3) \leq 6 \log(GT)$, $T' \geq T/2$, and $T - T' = \lceil 9C_k^4 C \rceil G \leq 10C_k^4 CG$,

$$\text{Reg}(\pi) \leq \frac{48C_k^2 \|\Sigma\|_1 \|\Sigma\|_2 \sqrt{\log(GT)}}{T^{1.5}} + \frac{120C_k^4 G \|\Sigma\|_2^2 \log(GT)}{T^2}.$$

C.2 ANALYSIS FOR GENERAL MOMENT BOUNDS

Theorem 13. *Suppose the distributions $\mathcal{D}_i$ are (k, C_k) -hypercontractive for some $k > 2$, and let $p_k = \min\left(\frac{k-2}{k}, \frac{1}{2}\right)$. Then, for $T = \tilde{\Omega}\left(GC_k^{2/p_k}\right)$, the regret of Algorithm 2 satisfies*

$$\mathbb{E} [\text{Reg}(\pi)] = \tilde{O} \left(\frac{\|\Sigma\|_2^{2(1-p_k)} \|\Sigma\|_2^{2p_k} C_k^2}{T^{1+p_k}} + \frac{\|\Sigma\|_2^2 GC_k^{2/p_k}}{T^2} \right) \quad (54)$$

Recall the idea of a good event $\mathcal{E}$ from 4.1. Assume that T draws from each of the G arms are made beforehand and revealed to the agent one by one. For each $i \in [G]$ and $t \in [T]$, let $\mathcal{E}_{i,t}$ denote the event that the median-of-means

Algorithm 3: ExploreRefineThenCommit

- 1 **Input:** Number of arms G , Horizon T , Parameters (k, C_k)
 - 2 $\delta \leftarrow GT^{-3}$
 - 3 $p_k \leftarrow \min\left(\frac{k-2}{k}, \frac{1}{2}\right)$
 - 4 $\alpha_k \leftarrow 100C_k^2$
 - 5 Pull each arm $\lceil (3\alpha_k)^{1/p_k} \log\left(\frac{2}{\delta}\right) \rceil$ times
 - 6 $\tilde{\Sigma}_{\text{coarse}} \leftarrow$ estimates of the standard deviations
 - 7 $T' \leftarrow T - \lceil (3\alpha_k)^{1/p_k} \log\left(\frac{2}{\delta}\right) \rceil G$
 - 8 $\tilde{n}_i^{\text{coarse}} \leftarrow \frac{\hat{\sigma}_{i,\text{coarse}}^2}{\|\tilde{\Sigma}_{\text{coarse}}\|_2^2} T'$
 - 9 Pull arm i $\lfloor \tilde{n}_i^{\text{coarse}}/6 \rfloor$ times
 - 10 $\tilde{\Sigma}_{\text{fine}} \leftarrow$ estimates of the standard deviations
 - 11 $\tilde{n}_i^{\text{fine}} \leftarrow \frac{\hat{\sigma}_{i,\text{fine}}^2 \left(1 + \alpha_k \left(\frac{6 \log(2/\delta)}{\tilde{n}_i^{\text{coarse}}}\right)^{p_k}\right)}{\sum_{j \in [G]} \hat{\sigma}_{j,\text{fine}}^2 \left(1 + \alpha_k \left(\frac{6 \log(2/\delta)}{\tilde{n}_j^{\text{coarse}}}\right)^{p_k}\right)} T'$
 - 12 Pull arm i $\lfloor \tilde{n}_i^{\text{fine}} \rfloor - \lfloor \tilde{n}_i^{\text{coarse}}/6 \rfloor$ times
-

estimator $\hat{\sigma}_{i,t}^2$ of σ_i^2 defined in Lemma 8 satisfies the concentration guaranteed by the Lemma. Precisely, for $i \in [G]$ and $\lceil (3\alpha_k)^{1/p_k} \log\left(\frac{2}{\delta}\right) \rceil \leq t \leq T$, define the event

$$\mathcal{E}_{i,t} := \left\{ |\hat{\sigma}^2 - \sigma^2| \leq \alpha_k \sigma^2 \left(\frac{\log(2/\delta)}{n} \right)^{p_k} \right\},$$

where $\alpha_k = 100C_k^2$ and $p_k = \min\left(\frac{k-2}{k}, \frac{1}{2}\right)$. This event holds with probability at least $1 - \delta$. By a union bound over all $i \in [G]$ and $\lceil (3\alpha_k)^{1/p_k} \log\left(\frac{2}{\delta}\right) \rceil \leq t \leq T$, the event

$$\mathcal{E} := \bigcap_{i \in [G]} \bigcap_{2 \leq t \leq T} \mathcal{E}_{i,t}$$

fails to hold with probability at most $GT\delta \leq T^{-2}$. Condition on this event $\mathcal{E}$.

First, we show that the coarse variance estimates $\hat{\sigma}_{i,\text{coarse}}^2$ are multiplicatively close to the true variances σ_i^2 .

Lemma 38. *Under event $\mathcal{E}$, for all $i \in [G]$ the coarse estimate $\hat{\sigma}_{i,\text{coarse}}^2$ of σ_i^2 satisfies*

$$\left| \frac{\hat{\sigma}_{i,\text{coarse}}^2}{\sigma_i^2} - 1 \right| \leq \frac{1}{3}.$$

Proof. Since $\mathcal{E}$ holds, we have for all $i \in [G]$ the inequality

$$|\hat{\sigma}_{i,\text{coarse}}^2 - \sigma_i^2| \leq \alpha_k \sigma_i^2 \left(\frac{\log(e^{1/8}/\delta)}{\ell} \right)^{p_k},$$

where $\ell = \lceil (3\alpha_k)^{1/p_k} \log\left(\frac{2}{\delta}\right) \rceil$ is the number of times each arm is pulled for the coarse estimate. Substituting the bound $\ell \geq (3\alpha_k)^{1/p_k} \log\left(\frac{2}{\delta}\right)$ into the inequality above leads to the desired result. $\square$

Next, we show that the coarse estimates $\tilde{n}_i^{\text{coarse}}$ are multiplicatively close to $n_{i,T'}^*$.

Lemma 39. *Under event $\mathcal{E}$, for all $i \in [G]$ the estimate $\tilde{n}_i^{\text{coarse}}$ of $n_{i,T'}^*$ satisfies*

$$\frac{n_{i,T'}^*}{2} \leq \tilde{n}_i^{\text{coarse}} \leq 2n_{i,T'}^*$$

Proof. By Lemma 38, $\frac{2\sigma_i^2}{3} \leq \hat{\sigma}_{i,\text{coarse}}^2 \leq \frac{4\sigma_i^2}{3}$, which also implies $\frac{2\|\Sigma\|_2^2}{3} \leq \|\tilde{\Sigma}_{\text{coarse}}\|_2^2 \leq \frac{4\|\Sigma\|_2^2}{3}$. Therefore,

$$\frac{n_{i,T'}^*}{2} = \frac{2\sigma_i^2/3}{4\|\Sigma\|_2^2/3} T' \leq \frac{\hat{\sigma}_{i,\text{coarse}}^2}{\|\tilde{\Sigma}_{\text{coarse}}\|_2^2} T' \leq \frac{4\sigma_i^2/3}{2\|\Sigma\|_2^2/3} T' = 2n_{i,T'}^*.$$

□

Since $T \geq 2G \lceil (3\alpha_k)^{1/p_k} \log(\frac{2}{\delta}) \rceil$, the number of pulls T' remaining is at least $T/2$. In the refinement phase, arm i is pulled $\lfloor \tilde{n}_i^{\text{coarse}}/6 \rfloor$ times. Therefore, the total number of pulls in this phase is bounded above by $T/6$, which is within the allowed total of T' . These pulls refine the variance estimates, as the next Lemma shows.

Lemma 40. *Under event $\mathcal{E}$, for all $i \in [G]$ the fine estimate $\hat{\sigma}_{i,\text{fine}}^2$ of σ_i^2 satisfies*

$$\left| \frac{\hat{\sigma}_{i,\text{fine}}^2}{\sigma_i^2} - 1 \right| \leq \alpha_k \sigma_i^2 \left(\frac{6 \log(2/\delta)}{\tilde{n}_i^{\text{coarse}}} \right)^{p_k} \leq \alpha_k \sigma_i^2 \left(\frac{12 \log(2/\delta)}{n_{i,T'}^*} \right)^{p_k}.$$

Proof. The first inequality holds because each arm has been pulled at least $\lceil (3\alpha_k)^{1/p_k} \log(\frac{2}{\delta}) \rceil + \lfloor \tilde{n}_i^{\text{coarse}}/6 \rfloor > \tilde{n}_i^{\text{coarse}}/6$ times. The second inequality follows from Lemma 39. □

The algorithm commits to the estimates $\tilde{n}_i^{\text{fine}}$ using these finer estimates of the variances. We show next that $\tilde{n}_i^{\text{fine}}$ is a very good estimate of $n_{i,T'}^*$.

Lemma 41. *For all $i \in [G]$ the fine estimate $\tilde{n}_i^{\text{fine}}$ of $n_{i,T'}^*$ satisfies*

$$\tilde{n}_i^{\text{fine}} \geq \frac{\sigma_i^2}{\|\Sigma\|_2^2 + \frac{2\alpha_k (12 \log(2/\delta))^{p_k} \|\Sigma\|_2^{2p_k} \|\Sigma\|_{2(1-p_k)}^{2(1-p_k)}}{T'^{p_k}} T' \geq \frac{n_{i,T'}^*}{2}.$$

Proof. By Lemma 40,

$$\tilde{n}_i^{\text{fine}} = \frac{\hat{\sigma}_{i,\text{fine}}^2 \left(1 + \alpha_k \left(\frac{6 \log(2/\delta)}{\tilde{n}_i^{\text{coarse}}} \right)^{p_k} \right)}{\sum_{j \in [G]} \hat{\sigma}_{j,\text{fine}}^2 \left(1 + \alpha_k \left(\frac{6 \log(2/\delta)}{\tilde{n}_j^{\text{coarse}}} \right)^{p_k} \right)} T' \geq \frac{\sigma_i^2}{\sum_{j \in [G]} \sigma_j^2 \left(1 + 2\alpha_k \left(\frac{6 \log(2/\delta)}{\tilde{n}_j^{\text{coarse}}} \right)^{p_k} \right)} T'$$

Using the bound $\tilde{n}_j^{\text{coarse}} \geq n_{j,T'}^*/2$ from Lemma 39,

$$\sum_{j \in [G]} 2\alpha_k \sigma_j^2 \left(\frac{6 \log(2/\delta)}{\tilde{n}_j^{\text{coarse}}} \right)^{p_k} \leq \sum_{j \in [G]} 2\alpha_k \sigma_j^2 \left(\frac{12 \|\Sigma\|_2^2 \log(2/\delta)}{\sigma_j^2 T'} \right)^{p_k} = \frac{2\alpha_k (12 \log(2/\delta))^{p_k} \|\Sigma\|_2^{2p_k} \|\Sigma\|_{2(1-p_k)}^{2(1-p_k)}}{T'^{p_k}}.$$

The first inequality of Lemma 40 follows by combining the two bounds above. To show the other inequality $\tilde{n}_i^{\text{fine}} \geq n_{i,T'}^*/2$, it suffices to show that

$$\|\Sigma\|_2^2 \geq \frac{2\alpha_k (12 \log(2/\delta))^{p_k} \|\Sigma\|_2^{2p_k} \|\Sigma\|_{2(1-p_k)}^{2(1-p_k)}}{T'^{p_k}}.$$

Rewrite the above inequality as

$$T' \geq 12(2\alpha_k)^{1/p_k} \log(2/\delta) \left(\frac{\|\Sigma\|_{2(1-p_k)}}{\|\Sigma\|_2} \right)^{\frac{2(1-p_k)}{p_k}}.$$

Using inequalities between norms (or via the Generalized Mean Inequality), $(\|\Sigma\|_{2(1-p_k)}/\|\Sigma\|_2)^{\frac{2(1-p_k)}{p_k}} \leq G$. Moreover, $T' \geq T/2$. Therefore, it suffices for $T \geq 24G(2\alpha_k)^{1/p_k} \log(2GT^3)$ to hold, which is true as long as $T = \Omega\left(G\alpha_k^{1/p_k} \log\left(G\alpha_k^{1/p_k}\right)\right)$ for a sufficiently large universal constant. □

Finally, since each arm has been pulled at least $\lceil (3\alpha_k)^{1/p_k} \log(\frac{2}{\delta}) \rceil + \lfloor \tilde{n}_i^{\text{fine}} \rfloor > \tilde{n}_i^{\text{fine}}$ times,

$$\begin{aligned} \text{Reg}(\pi) &\leq \max_{i \in [G]} \frac{\sigma_i^2}{\tilde{n}_i^{\text{fine}}} - \frac{\|\Sigma\|_2^2}{T} \\ &\leq \max_{i \in [G]} \left(\frac{\|\Sigma\|_2^2 + \frac{2\alpha_k (12 \log(2/\delta))^{p_k} \|\Sigma\|_2^{2p_k} \|\Sigma\|_{2(1-p_k)}^{2(1-p_k)}}{T'^{p_k}}}{T'} - \frac{\|\Sigma\|_2^2}{T} \right) \\ &\leq \frac{2\alpha_k (12 \log(2/\delta))^{p_k} \|\Sigma\|_2^{2p_k} \|\Sigma\|_{2(1-p_k)}^{2(1-p_k)}}{T'^{1+p_k}} + \frac{\|\Sigma\|_2^2 (T - T')}{TT'} \end{aligned}$$

Using $\log(2/\delta) \leq 3 \log(GT)$, $T' \geq T/2$, and $T - T' \leq 2(3\alpha_k)^{1/p_k} \log(\frac{2}{\delta}) G$, we obtain the regret bound

$$\text{Reg}(\pi) \leq \frac{4\alpha_k (72 \log(GT))^{p_k} \|\Sigma\|_2^{2p_k} \|\Sigma\|_{2(1-p_k)}^{2(1-p_k)}}{T^{1+p_k}} + \frac{4(3\alpha_k)^{1/p_k} \log(\frac{2}{\delta}) G \|\Sigma\|_2^2}{T^2}.$$

D LOWER BOUNDS

In this section, we describe our lower bounds for hypercontractive distributions in Section D.1 and subgaussian and bounded moment distributions in Section D.2.

D.1 LOWER BOUNDS FOR HYPERCONTRACTIVE DISTRIBUTIONS

Theorem 14. *For any policy π , and positive integers G and T with $T > 2G$, there exists a bandit instance with G arms, time horizon T , and Gaussian rewards, such that*

$$\text{Reg}(\pi) \geq \frac{\|\Sigma\|_1 \|\Sigma\|_2}{T^{1.5}}.$$

Proof. Consider two bandit instances $B^{(1)}$ and $B^{(2)}$ with G arms, time horizon T , and vectors of reward distributions $\mathcal{D}^{(1)}$ and $\mathcal{D}^{(2)}$ respectively, where

$$\mathcal{D}_1^{(1)} = \mathcal{N}\left(0, \sigma^2 \left(1 + \frac{C}{\sqrt{T}}\right)\right), \mathcal{D}_1^{(2)} = \mathcal{N}\left(0, \sigma^2 \left(1 - \frac{C}{\sqrt{T}}\right)\right), \mathcal{D}_k^{(1)} = \mathcal{N}(0, \epsilon^2) \quad \forall k \geq 3$$

and

$$\mathcal{D}_1^{(2)} = \mathcal{N}\left(0, \sigma^2 \left(1 - \frac{C}{\sqrt{T}}\right)\right), \mathcal{D}_2^{(2)} = \mathcal{N}\left(0, \sigma^2 \left(1 + \frac{C}{\sqrt{T}}\right)\right), \mathcal{D}_k^{(2)} = \mathcal{N}(0, \epsilon^2) \quad \forall k \geq 3$$

for some arbitrary real $\sigma > 0$, $\epsilon = \sqrt{2\sigma^2/(T - G + 2)}$, and $C = 0.03$. Since the time horizon is T , we can assume that the rewards for the arms are generated before the agent interacts with the bandit. Moreover, since the time horizon is T , it is information-theoretically equivalent to sample T rewards from each arm in advance and reveal the rewards one at a time as the agent pulls the arms.

Let $\mathcal{P}^{(1)} := (\mathcal{D}^{(1)})^T$ and $\mathcal{P}^{(2)} := (\mathcal{D}^{(2)})^T$ be the joint distribution of T rewards from each of the G distributions.

Lemma 42. $d_{\text{TV}}(\mathcal{P}^{(1)}, \mathcal{P}^{(2)}) \leq 0.1$.

Proof. Since the TV-distance is sub-additive,

$$d_{\text{TV}}(\mathcal{P}^{(1)}, \mathcal{P}^{(2)}) \leq 2d_{\text{TV}}\left(\mathcal{N}\left(0, \sigma^2(1 + C/\sqrt{T})\mathbf{I}_T\right), \mathcal{N}\left(0, \sigma^2(1 - C/\sqrt{T})\mathbf{I}_T\right)\right),$$

where $\mathbf{I}_T$ is the $T \times T$ identity matrix. By Theorem 1.1. of Devroye et al. [2018], the total variation distance between the two d -dimensional Gaussians $\mathcal{N}(0, \mathbf{M}_1)$ and $\mathcal{N}(0, \mathbf{M}_2)$ satisfies

$$\frac{1}{100} \min\left(1, \sqrt{\sum_{i=1}^d \lambda_i^2}\right) \leq d_{\text{TV}}(\mathcal{N}(0, \mathbf{M}_1), \mathcal{N}(0, \mathbf{M}_2)) \leq \frac{3}{2} \min\left(1, \sqrt{\sum_{i=1}^d \lambda_i^2}\right),$$

where $\lambda_1, \dots, \lambda_d$ is the sum of the squares of the eigenvalues of $\mathbf{M}_1^{-1}\mathbf{M}_2 - \mathbf{I}_d$.

Set $d = T$, $\mathbf{M}_1 = \sigma^2(1 + C/\sqrt{T})\mathbf{I}_T$ and $\mathbf{M}_2 = \sigma^2(1 - C/\sqrt{T})\mathbf{I}_T$. Every eigenvalue of $\mathbf{M}_1^{-1}\mathbf{M}_2 - \mathbf{I}_T = -\frac{2C/\sqrt{T}}{1+C/\sqrt{T}}\mathbf{I}_T$ is equal to $-\frac{2C/\sqrt{T}}{1+C/\sqrt{T}}$ and $\sum_{i=1}^T \lambda_i^2 = T \left(\frac{2C/\sqrt{T}}{1+C/\sqrt{T}}\right)^2 \leq 4C^2$.

Therefore, $d_{\text{TV}}(\mathcal{P}^{(1)}, \mathcal{P}^{(2)}) \leq 3C < 0.1$. $\square$

With the view that T rewards are generated from each arm before the agent interacts with the bandit, Lemma 42 establishes that the product distribution of these GT rewards for the two bandits are close in TV-distance. Therefore, for any policy π , the output distributions of $\mathbf{n}^{(1)} \sim \pi(B^{(1)})$ and $\mathbf{n}^{(2)} \sim \pi(B^{(2)})$ also have TV-distance at most 0.1. There exists a coupling between the two scenarios such that $\mathbf{n}^{(1)} = \mathbf{n}^{(2)}$ with probability at least 0.9. Condition on this event, and let $\tilde{\mathbf{n}}$ denote this common vector.

First, we calculate the vectors $\mathbf{n}^{(1*)}$ and $\mathbf{n}^{(2*)}$ of the optimal number of pulls for the two bandits. Let $\Sigma^{(1)}$ and $\Sigma^{(2)}$ be the vectors of standard deviations of the rewards of bandits B_1 and B_2 respectively. Then, $\|\Sigma^{(1)}\|_2^2 = \|\Sigma^{(2)}\|_2^2 = 2\sigma^2 + (G-2)\epsilon^2$, so

$$\begin{aligned} n_1^{(1*)} &= n_2^{(2*)} = \frac{\sigma^2(1 + C/\sqrt{T})}{2\sigma^2 + (G-2)\epsilon^2} T, \\ n_2^{(1*)} &= n_1^{(2*)} = \frac{\sigma^2(1 - C/\sqrt{T})}{2\sigma^2 + (G-2)\epsilon^2} T, \text{ and} \\ n_i^{(1*)} &= n_i^{(2*)} = \frac{\epsilon^2}{2\sigma^2 + (G-2)\epsilon^2} T = 1 \quad \forall i \geq 3. \end{aligned}$$

Any policy π with finite regret pulls each arm at least once a.s. So, $\tilde{n}_i \geq n_i^{(1*)} = n_i^{(2*)}$ for any $i \geq 3$, almost surely. Thus,

$$\begin{aligned} n_1^{(1*)} + n_2^{(2*)} - \tilde{n}_1 - \tilde{n}_2 &\geq \left(n_1^{(1*)} - \tilde{n}_1\right) + \left(n_2^{(2*)} - \tilde{n}_2\right) + \sum_{3 \leq i \leq G} \left(n_i^{(1*)} - \tilde{n}_i\right) \\ &= \left(n_2^{(2*)} - n_2^{(1*)}\right) + \sum_{i \in [G]} n_i^{(1*)} - \sum_{i \in [G]} \tilde{n}_i \\ &= n_2^{(2*)} - n_2^{(1*)} = \frac{2\sigma^2 C/\sqrt{T}}{2\sigma^2 + (G-2)\epsilon^2} T > \frac{C\sqrt{T}}{2}, \end{aligned}$$

where the last inequality follows from $G\epsilon^2 \leq 2\sigma^2$. It follows that either $n_1^{(1*)} - \tilde{n}_1 > \frac{C\sqrt{T}}{4}$ or $n_2^{(2*)} - \tilde{n}_2 > \frac{C\sqrt{T}}{4}$. Either way,

$$\begin{aligned} \max(\text{Reg}(\pi; B_1), \text{Reg}(\pi; B_2)) &\geq \max\left(\frac{\sigma^2(1 + C/\sqrt{T})}{\tilde{n}_1} - \frac{\|\Sigma^{(1)}\|_2^2}{T}, \frac{\sigma^2(1 + C/\sqrt{T})}{\tilde{n}_2} - \frac{\|\Sigma^{(2)}\|_2^2}{T}\right) \\ &\geq \frac{\sigma^2(1 + C/\sqrt{T})}{\frac{\sigma^2(1 + C/\sqrt{T})}{2\sigma^2 + (G-2)\epsilon^2} T - C\sqrt{T}/4} - \frac{2\sigma^2 + (G-2)\epsilon^2}{T} \\ &= \frac{\sigma^2(1 + C/\sqrt{T})}{\frac{\sigma^2(1 + C/\sqrt{T})}{2\sigma^2 + (G-2)\epsilon^2} T - C\sqrt{T}/4} - \frac{\sigma^2(1 + C/\sqrt{T})}{\frac{\sigma^2(1 + C/\sqrt{T})}{2\sigma^2 + (G-2)\epsilon^2} T} \\ &\geq \frac{\sigma^2(1 + C/\sqrt{T})}{\left(\frac{\sigma^2(1 + C/\sqrt{T})}{2\sigma^2 + (G-2)\epsilon^2} T\right)^2} \frac{C\sqrt{T}}{4} \geq \frac{C\sigma^2}{2T^{1.5}} = \Omega\left(\frac{\|\Sigma\|_1 \|\Sigma\|_2}{T^{1.5}}\right). \end{aligned}$$

Therefore, there exists an index $i \in \{1, 2\}$ such that policy π incurs regret $\Omega\left(\frac{\|\Sigma\|_1 \|\Sigma\|_2}{T^{1.5}}\right)$ on the bandit instance B_i with probability at least $0.9/2 = 0.45$. $\square$

D.2 LOWER BOUND FOR SUBGAUSSIAN AND BOUNDED MOMENT DISTRIBUTIONS

We now formalize the reduction of regret to variance estimation: if a policy attains very small terminal regret in a two-armed instance where one arm is a fixed reference distribution, then the *terminal pull counts* alone yield a variance test for the other arm. Theorems 15 and 16 provide lower bounds for variance estimation under subgaussianity and a bounded fourth-moment assumption respectively and may be of independent interest to the reader.

Theorem 15 (Variance testing under subGaussianity). *Let n be a positive integer, $\sigma \in (0, 0.5)$, and $\delta \in (0, 1/4)$. Consider the family*

$$\mathcal{F}_\sigma := \{\mathcal{D} \in \text{subG}(1) : \text{Var}(\mathcal{D}) \leq 4\sigma^2\}.$$

Then, there exist distributions $D_0, D_1 \in \mathcal{F}_\sigma$ with

$$|\text{Var}(D_0) - \text{Var}(D_1)| = \Delta, \quad \Delta := \min\left\{\frac{\sigma^2}{10}, \sigma\sqrt{\frac{\log(1/\sigma) \log(1/(4\delta))}{4n}}\right\},$$

such that for any test $\hat{C}_n : \mathbb{R}^n \rightarrow \{0, 1\}$,

$$\frac{1}{2} \left(\Pr_{X \sim D_0^n}(\hat{C}_n(X) = 0) + \Pr_{X \sim D_1^n}(\hat{C}_n(X) = 1) \right) \leq 1 - \delta.$$

Proof. Let $a^2 := \log(1/\sigma)$ and $a := \sqrt{\log(1/\sigma)}$. Define

$$p_0 := \frac{\sigma^2}{a^2}, \quad p_1 := \frac{\sigma^2 + \Delta}{a^2}.$$

Consider the symmetric three-point distributions

$$\begin{aligned} D_0 &:= \frac{p_0}{2} \text{Dir}(-a) + \frac{p_0}{2} \text{Dir}(+a) + (1 - p_0) \text{Dir}(0), \\ D_1 &:= \frac{p_1}{2} \text{Dir}(-a) + \frac{p_1}{2} \text{Dir}(+a) + (1 - p_1) \text{Dir}(0). \end{aligned}$$

where $\text{Dir}(x)$ denotes the dirac distribution at x . Both are mean-zero by symmetry. Moreover,

$$\text{Var}(D_0) = \mathbb{E}[X^2] = p_0 a^2 = \sigma^2, \quad \text{Var}(D_1) = \mathbb{E}[X^2] = p_1 a^2 = \sigma^2 + \Delta,$$

so $|\text{Var}(D_0) - \text{Var}(D_1)| = \Delta$. Since $\Delta \leq \sigma^2/10$, we have $\text{Var}(D_1) < 2\sigma^2 < 4\sigma^2$.

Fix $i \in \{0, 1\}$ and let $X \sim D_i$ (so $\mathbb{E}[X] = 0$). For any $t \in \mathbb{R}$,

$$\mathbb{E} \exp(tX) = (1 - p_i) + \frac{p_i}{2} e^{at} + \frac{p_i}{2} e^{-at} = 1 - p_i + p_i \cosh(at).$$

It suffices to show that for all $t \geq 0$,

$$1 - p_i + p_i \cosh(at) \leq \exp(t^2/2),$$

since both sides are even in t . Define for $u \geq 0$,

$$F_i(u) := \exp(u^2/2) - (1 - p_i + p_i \cosh(au)).$$

Then $F_i(0) = 0$ and $F'_i(0) = 0$. Also,

$$F''_i(u) = (1 + u^2) \exp(u^2/2) - p_i a^2 \cosh(au) = (1 + u^2) \exp(u^2/2) - (\sigma^2 + \mathcal{I}_{i=1} \Delta) \cosh(au).$$

Using $\cosh(au) \leq e^{au}$ and $\sigma^2 + \Delta \leq 1.1\sigma^2 = 1.1e^{-2a^2}$,

$$F''_i(u) \geq \exp(u^2/2) - 1.1e^{-2a^2} e^{au}.$$

Since $(u - a)^2/2 \geq 0$, we have $u^2/2 \geq au - a^2/2$ and hence $\exp(u^2/2) \geq \exp(au - a^2/2)$, yielding

$$F''_i(u) \geq e^{au} \left(e^{-a^2/2} - 1.1e^{-2a^2} \right) = e^{au} e^{-2a^2} \left(e^{3a^2/2} - 1.1 \right).$$

Because $\sigma < 1/2$, $a^2 = \log(1/\sigma) \geq \log 2$, so $e^{3a^2/2} \geq 2^{3/2} > 1.1$ and thus $F''_i(u) \geq 0$ for all $u \geq 0$. Hence F_i is convex on $[0, \infty)$ and with $F_i(0) = F'_i(0) = 0$ it follows that $F_i(u) \geq 0$ for all $u \geq 0$. Therefore, for all $t \in \mathbb{R}$,

$$\mathbb{E} \exp(tX) \leq \exp(t^2/2),$$

i.e. X is 1-subgaussian in the sense of Definition 1. Thus $D_0, D_1 \in \text{subG}(1)$.

As in Theorem 16, the likelihood ratio depends only on whether $X = 0$ or $X \neq 0$, so

$$D_{\text{KL}}(D_0 \| D_1) = D_{\text{KL}}(\text{Ber}(p_0) \| \text{Ber}(p_1)), \quad D_{\text{KL}}(D_1 \| D_0) = D_{\text{KL}}(\text{Ber}(p_1) \| \text{Ber}(p_0)).$$

We first note $p_1 < 1/2$: indeed,

$$p_1 = \frac{\sigma^2 + \Delta}{\log(1/\sigma)} \leq \frac{1.1\sigma^2}{\log(1/\sigma)} \leq \frac{1.1 \cdot 0.25}{\log 2} < \frac{1}{2}.$$

Using $D_{\text{KL}}(\text{Ber}(p) \| \text{Ber}(q)) \leq \frac{(p-q)^2}{q(1-q)}$ (e.g. [Tsybakov, 2009, Lemma 2.6]), and $p_1 \geq p_0$, $1 - p_1 \geq 1/2$, we obtain for both directions

$$D_{\text{KL}}(D_0 \| D_1) \leq \frac{(p_1 - p_0)^2}{p_1(1 - p_1)} \leq \frac{2(p_1 - p_0)^2}{p_0} = 2 \cdot \frac{(\Delta/a^2)^2}{\sigma^2/a^2} = \frac{2\Delta^2}{\sigma^2 a^2}.$$

Therefore

$$D_{\text{KL}}(D_0^n \| D_1^n) \leq \frac{2n\Delta^2}{\sigma^2 a^2}.$$

By the definition of Δ , the RHS is at most $\log(1/(4\delta))$.

The same Bretagnolle-Huber and Neyman-Pearson argument as in Theorem 16 yields

$$\sup_{\hat{C}_n} \frac{1}{2} \left(\Pr_{D_0^n}(\hat{C}_n = 0) + \Pr_{D_1^n}(\hat{C}_n = 1) \right) \leq 1 - \delta,$$

as claimed. $\square$

Theorem 16 (Variance testing under bounded fourth moment). *Let n be a positive integer, $\sigma \in (0, 0.5)$, and $\delta \in (0, 1/4)$. Let $\kappa(\mathcal{D}) := \mathbb{E}_{X \sim \mathcal{D}}[(X - \mathbb{E}X)^4]$ denote the central fourth moment. Consider the family*

$$\mathcal{F}_\sigma := \{\mathcal{D} : \text{Var}(\mathcal{D}) \leq 4\sigma^2, \kappa(\mathcal{D}) \leq 2\}.$$

Then, there exist distributions $D_0, D_1 \in \mathcal{F}_\sigma$ with

$$|\text{Var}(D_0) - \text{Var}(D_1)| = \Delta, \quad \Delta := \min \left\{ \frac{\sigma^2}{10}, \sqrt{\frac{\log(1/(4\delta))}{4n}} \right\},$$

such that for any test $\hat{C}_n : \mathbb{R}^n \rightarrow \{0, 1\}$,

$$\frac{1}{2} \left(\Pr_{X \sim D_0^n}(\hat{C}_n(X) = 0) + \Pr_{X \sim D_1^n}(\hat{C}_n(X) = 1) \right) \leq 1 - \delta.$$

Proof. Define Δ as in the statement and consider the symmetric three-point distributions

$$\begin{aligned} D_0 &:= \frac{\sigma^4}{2} \text{Dir}(-1/\sigma) + \frac{\sigma^4}{2} \text{Dir}(+1/\sigma) + (1 - \sigma^4) \text{Dir}(0), \\ D_1 &:= \frac{\sigma^4 + \sigma^2 \Delta}{2} \text{Dir}(-1/\sigma) + \frac{\sigma^4 + \sigma^2 \Delta}{2} \text{Dir}(+1/\sigma) + (1 - \sigma^4 - \sigma^2 \Delta) \text{Dir}(0). \end{aligned}$$

where $\text{Dir}(x)$ denotes the dirac distribution at x . Both have mean 0 by symmetry. Moreover,

$$\text{Var}(D_0) = \mathbb{E}[X^2] = \sigma^4 \cdot \frac{1}{\sigma^2} = \sigma^2, \quad \text{Var}(D_1) = \mathbb{E}[X^2] = (\sigma^4 + \sigma^2 \Delta) \cdot \frac{1}{\sigma^2} = \sigma^2 + \Delta,$$

so $|\text{Var}(D_0) - \text{Var}(D_1)| = \Delta$. Since $\Delta \leq \sigma^2/10$, we have $\text{Var}(D_1) = \sigma^2 + \Delta < 2\sigma^2 < 4\sigma^2$.

Next, the (central) fourth moments are

$$\kappa(D_0) = \mathbb{E}[X^4] = \sigma^4 \cdot \frac{1}{\sigma^4} = 1, \quad \kappa(D_1) = \mathbb{E}[X^4] = (\sigma^4 + \sigma^2 \Delta) \cdot \frac{1}{\sigma^4} = 1 + \frac{\Delta}{\sigma^2} \leq 1.1 < 2,$$

so $D_0, D_1 \in \mathcal{F}_\sigma$.

Let $p_0 := \sigma^4$ and $p_1 := \sigma^4 + \sigma^2 \Delta$. Conditioned on $X \neq 0$, both D_0 and D_1 are uniform on $\{\pm 1/\sigma\}$, hence the likelihood ratio depends only on whether $X = 0$ or $X \neq 0$, giving

$$D_{\text{KL}}(D_0 \| D_1) = D_{\text{KL}}(\text{Ber}(p_0) \| \text{Ber}(p_1)), \quad D_{\text{KL}}(D_1 \| D_0) = D_{\text{KL}}(\text{Ber}(p_1) \| \text{Ber}(p_0)).$$

Using the standard inequality

$$D_{\text{KL}}(\text{Ber}(p) \| \text{Ber}(q)) \leq \frac{(p - q)^2}{q(1 - q)} \quad (\text{see e.g. [Tsybakov, 2009, Lemma 2.6]}),$$

and noting that $p_1 \leq 1.1\sigma^4 \leq 1.1 \cdot (1/16) < 1/2$, we have $1 - p_1 \geq 1/2$ and $p_1 \geq p_0$, so for both directions,

$$D_{\text{KL}}(D_0 \| D_1) \leq \frac{(p_1 - p_0)^2}{p_1(1 - p_1)} \leq \frac{2(\sigma^2 \Delta)^2}{\sigma^4} = 2\Delta^2,$$

and similarly $D_{\text{KL}}(D_1 \| D_0) \leq 2\Delta^2$. Therefore,

$$D_{\text{KL}}(D_0^n \| D_1^n) = n D_{\text{KL}}(D_0 \| D_1) \leq 2n\Delta^2.$$

By the definition of Δ , we have $2n\Delta^2 \leq \log(1/(4\delta))$.

By the Bretagnolle-Huber inequality (see e.g. [Tsybakov, 2009, Theorem 2.2]),

$$\text{TV}(D_0^n, D_1^n) \leq 1 - \frac{1}{2} \exp\left(-D_{\text{KL}}(D_0^n \| D_1^n)\right) \leq 1 - \frac{1}{2} \exp\left(-\log(1/(4\delta))\right) = 1 - 2\delta.$$

For testing two simple hypotheses, the Neyman-Pearson lemma (see e.g. [Lehmann and Romano, 2005, Chapter 3]) implies

$$\sup_{\hat{C}_n: \mathbb{R}^n \rightarrow \{0,1\}} \frac{1}{2} \left(\mathbb{P}_{D_0^n}(\hat{C}_n = 0) + \mathbb{P}_{D_1^n}(\hat{C}_n = 1) \right) = \frac{1 + \text{TV}(D_0^n, D_1^n)}{2} \leq 1 - \delta$$

□

Fix $T \in \mathbb{N}$. Let $\mathcal{R}$ be a *reference* distribution with mean 0 and variance 1 (e.g. $\mathcal{R} = \text{Rademacher}$). For any distribution $\mathcal{D}$ with mean 0 and variance $\text{Var}(\mathcal{D}) = s > 0$, consider the two-arm instance $\mathcal{I} = (\mathcal{D}, \mathcal{R})$. Let N_T be the number of pulls of arm 1 by time T (so arm 2 is pulled $T - N_T$ times). Recall that the terminal variance vector is $v_T(1) = s/N_T$ and $v_T(2) = 1/(T - N_T)$, with the convention $a/0 := +\infty$. In the two-armed case, Lemma 1 gives $R^* = (s + 1)/T$ and hence

$$\text{Reg}(\pi) = \|v_T\|_\infty - \frac{s + 1}{T} = \max\left(\frac{s}{N_T}, \frac{1}{T - N_T}\right) - \frac{s + 1}{T}.$$

Lemma 43. Fix $s > 0$ and $T \in \mathbb{N}$. On any realization on which $\text{Reg}(\pi) \leq r$ (for some $r \geq 0$), the pull-count ratio

$$Z_T := \frac{N_T}{T - N_T}$$

satisfies

$$\frac{s}{1 + rT} \leq Z_T \leq s + rT. \quad (55)$$

Proof. If $\text{Reg}(\pi) \leq r$, then $\|v_T\|_\infty \leq \frac{s+1}{T} + r$. Hence

$$\frac{s}{N_T} \leq \frac{s+1}{T} + r \Rightarrow N_T \geq \frac{s}{(s+1)/T + r} = \frac{sT}{s+1+rT},$$

and

$$\frac{1}{T - N_T} \leq \frac{s+1}{T} + r \Rightarrow T - N_T \geq \frac{1}{(s+1)/T + r} = \frac{T}{s+1+rT}.$$

From the second inequality,

$$N_T = T - (T - N_T) \leq T - \frac{T}{s+1+rT} = \frac{(s+rT)T}{s+1+rT},$$

so dividing by the lower bound on $T - N_T$ gives

$$Z_T = \frac{N_T}{T - N_T} \leq \frac{(s+rT)T/(s+1+rT)}{T/(s+1+rT)} = s + rT.$$

Similarly, from the first inequality,

$$T - N_T \leq T - \frac{sT}{s+1+rT} = \frac{(1+rT)T}{s+1+rT},$$

so dividing the lower bound on N_T by this upper bound on $T - N_T$ yields

$$Z_T = \frac{N_T}{T - N_T} \geq \frac{sT/(s+1+rT)}{(1+rT)T/(s+1+rT)} = \frac{s}{1+rT}.$$

□

Lemma 44. Fix $0 < s_0 < s_1$ and $T \in \mathbb{N}$. Suppose $r \geq 0$ satisfies

$$s_0 + rT < \frac{s_1}{1 + rT}. \quad (56)$$

Let τ be any threshold satisfying

$$s_0 + rT < \tau < \frac{s_1}{1 + rT}.$$

Then on the event $\{\text{Reg}(\pi) \leq r\}$, we have

$$Z_T \leq \tau \text{ whenever } \text{Var}(\mathcal{D}) = s_0, \quad Z_T > \tau \text{ whenever } \text{Var}(\mathcal{D}) = s_1.$$

Proof. Under $\text{Var}(\mathcal{D}) = s$, Lemma 43 implies $Z_T \in [s/(1 + rT), s + rT]$ on $\{\text{Reg}(\pi) \leq r\}$. The condition (56) makes these two intervals disjoint for $s = s_0$ versus $s = s_1$, so any τ in between separates the two cases. $\square$

Lemma 45. Fix $T \in \mathbb{N}$ and two mean-zero laws D_0, D_1 with variances $\text{Var}(D_0) = s_0 < s_1 = \text{Var}(D_1)$. Let $\mathcal{R}$ be a mean-zero reference arm with $\text{Var}(\mathcal{R}) = 1$. Fix $r \geq 0$ satisfying (56) and choose τ as in Lemma 44. Let $m \in \mathbb{N}$ satisfy

$$m \geq T - \frac{T}{s_1 + 1 + rT}. \quad (57)$$

Given m i.i.d. samples from D_θ (unknown $\theta \in \{0, 1\}$), define a (possibly randomized) test $\hat{\theta} : \mathbb{R}^m \rightarrow \{0, 1\}$ by simulating π on the instance $(D_\theta, \mathcal{R})$ for T rounds: feed the provided samples sequentially whenever π pulls arm 1, and generate fresh samples from $\mathcal{R}$ whenever it pulls arm 2. If π requests more than m samples from arm 1, output a fair coin; otherwise, output $\hat{\theta} = \mathcal{I}\{Z_T > \tau\}$, where $Z_T := \frac{N_T}{T - N_T}$. Then

$$\frac{1}{2} \left(\Pr_{D_0^m}(\hat{\theta} = 1) + \Pr_{D_1^m}(\hat{\theta} = 0) \right) \leq \frac{3}{2} \max_{\theta \in \{0, 1\}} \Pr_\theta(\text{Reg}(\pi) > r).$$

Proof. On the event $\{\text{Reg}(\pi) \leq r\}$, Lemma 44 implies the threshold rule $\mathcal{I}\{Z_T > \tau\}$ is correct, hence makes no error. Therefore, the only possible sources of error are: (i) $\text{Reg}(\pi) > r$; or (ii) the simulation aborts and we output a fair coin, contributing at most $1/2$ error on that event. Thus,

$$\Pr(\text{test error}) \leq \Pr(\text{Reg}(\pi) > r) + \frac{1}{2} \Pr(\text{abort}).$$

It remains to show $\{\text{abort}\} \subseteq \{\text{Reg}(\pi) > r\}$ under (57). Indeed, suppose $\text{Reg}(\pi) \leq r$. Then $\|v_T\|_\infty \leq (s + 1)/T + r$, where $s = \text{Var}(D_\theta) \leq s_1$. In particular,

$$\frac{1}{T - N_T} \leq \frac{s + 1}{T} + r \Rightarrow T - N_T \geq \frac{T}{s + 1 + rT} \Rightarrow N_T \leq T - \frac{T}{s + 1 + rT} \leq T - \frac{T}{s_1 + 1 + rT} \leq m,$$

where the last inequality is (57). Hence $N_T > m$ (i.e. abort) implies $\text{Reg}(\pi) > r$, and so $\Pr(\text{abort}) \leq \Pr(\text{Reg}(\pi) > r)$. Plugging this into the previous display yields

$$\Pr(\text{test error}) \leq \Pr(\text{Reg}(\pi) > r) + \frac{1}{2} \Pr(\text{Reg}(\pi) > r) = \frac{3}{2} \Pr(\text{Reg}(\pi) > r),$$

and taking the maximum over $\theta \in \{0, 1\}$ completes the proof. $\square$

Theorem 10 (Reduction to variance estimation). Fix $T \in \mathbb{N}$ and $\delta \in (0, 1/4)$. Let D_0, D_1 be mean-zero distributions with

$$\text{Var}(D_0) = \sigma^2, \quad \text{Var}(D_1) = \sigma^2 + \Delta,$$

for some $\sigma \in (0, 0.5)$ and $\Delta \in (0, \sigma^2/10)$. Let $\mathcal{R}$ be any mean-zero reference arm with $\text{Var}(\mathcal{R}) = 1$ (e.g. a Rademacher arm). Assume that for some integer m , over all functions $\hat{C}_m : \mathbb{R}^m \rightarrow \{0, 1\}$,

$$\inf_{\hat{C}_m} \frac{1}{2} \left(\Pr_{X \sim D_0^m}(\hat{C}_m(X) = 0) + \Pr_{X \sim D_1^m}(\hat{C}_m(X) = 1) \right) \leq 1 - \delta, \quad (11)$$

equivalently, every test has average error at least δ . Suppose in addition that m satisfies

$$m \geq T - \frac{T}{(\sigma^2 + \Delta) + 1 + \Delta/32}. \quad (12)$$

Then for every policy π run for horizon T on the instances $\mathcal{I}_0 = (D_0, \mathcal{R})$ and $\mathcal{I}_1 = (D_1, \mathcal{R})$, we have

$$\max_{\theta \in \{0,1\}} \mathbb{E}_\theta [\text{Reg}(\pi)] \geq c \cdot \frac{\delta \Delta}{T},$$

for a universal constant $c > 0$.

Proof. Set $r := \Delta/(32T)$, so $rT = \Delta/32$. We first verify the separation condition (56) with $(s_0, s_1) = (\sigma^2, \sigma^2 + \Delta)$: it suffices to show $(\sigma^2 + rT)(1 + rT) \leq \sigma^2 + \Delta$, i.e.

$$\begin{aligned} (\sigma^2 + rT)(1 + rT) &= \sigma^2 + (\sigma^2 + 1)rT + (rT)^2 \\ &\leq \sigma^2 + \frac{\sigma^2 + 1}{32}\Delta + \frac{1}{1024}\Delta^2 \\ &\leq \sigma^2 + \frac{1.25}{32}\Delta + \frac{1}{1024}\Delta^2 < \sigma^2 + \Delta, \end{aligned}$$

where we used $\sigma < 1/2$ and $\Delta \leq \sigma^2/10$. Thus Lemma 45 applies (with the budget condition (12)).

Assume for contradiction that $\max_\theta \mathbb{E}_\theta [\text{Reg}(\pi)] < c\delta \Delta/T$ for $c = 1/100$. By Markov's inequality, for all $\theta \in \{0, 1\}$,

$$\Pr_\theta (\text{Reg}(\pi) > r) \leq \frac{\mathbb{E}_\theta [\text{Reg}(\pi)]}{r} < \frac{(c\delta)(\Delta/T)}{\Delta/(32T)} = 32c\delta < \frac{2}{3}\delta.$$

Lemma 45 then constructs a test $\hat{\theta}$ using m samples whose average error is at most $\frac{3}{2} \cdot \frac{2}{3}\delta = \delta$, contradicting (11). Therefore $\max_\theta \mathbb{E}_\theta [\text{Reg}(\pi)] \geq c\delta \Delta/T$. $\square$

Theorem 12 (Lower bound for subgaussian distributions). *Fix $T \in \mathbb{N}$, $\sigma \in (0, 0.5)$, and $\delta \in (0, 1/4)$. Let $\mathcal{R} = \text{Rademacher}$. Set $m := \lceil 2\sigma^2 T \rceil$ and let*

$$\begin{aligned} \Delta &= \min \left\{ \frac{\sigma^2}{10}, \sigma \sqrt{\frac{\log(1/\sigma) \log(1/(4\delta))}{4m}} \right\} \\ &= \min \left\{ \frac{\sigma^2}{10}, \sqrt{\frac{\log(1/\sigma) \log(1/(4\delta))}{8T}} \right\}. \end{aligned}$$

Then there exist mean-zero distributions $D_0, D_1 \in \text{subG}(1)$ with $\text{Var}(D_0) = \sigma^2$ and $\text{Var}(D_1) = \sigma^2 + \Delta$ such that for every policy π ,

$$\max_{\theta \in \{0,1\}} \mathbb{E}_\theta [\text{Reg}(\pi)] \geq c \cdot \frac{\delta}{T} \min \left\{ \frac{\sigma^2}{10}, \sqrt{\frac{\log(1/\sigma) \log(1/(4\delta))}{8T}} \right\},$$

for a universal constant $c > 0$.

Proof. Apply Theorem 15 with $n = m$ to obtain $D_0, D_1 \in \text{subG}(1)$ with $\text{Var}(D_0) = \sigma^2$, $\text{Var}(D_1) = \sigma^2 + \Delta$, and the testing guarantee

$$\inf_{\hat{C}_m: \mathbb{R}^m \rightarrow \{0,1\}} \frac{1}{2} \left(\Pr_{X \sim D_0^m} (\hat{C}_m(X) = 0) + \Pr_{X \sim D_1^m} (\hat{C}_m(X) = 1) \right) \leq 1 - \delta,$$

i.e. every test has average error at least δ . Thus the assumption (11) in Theorem 10 holds.

It remains to verify the sample-budget condition needed to invoke Lemma 45 inside Theorem 10. In Theorem 10, we set $r = \Delta/(32T)$ and $s_1 = \sigma^2 + \Delta$, so the budget requirement is

$$m \geq T - \frac{T}{s_1 + 1 + rT} = T - \frac{T}{(\sigma^2 + \Delta) + 1 + \Delta/32}.$$

Let $a := (\sigma^2 + \Delta) + \Delta/32 = \sigma^2 + \frac{33}{32}\Delta$. Since $\frac{a}{1+a} \leq a$, we have

$$T - \frac{T}{1+a} = T \cdot \frac{a}{1+a} \leq aT.$$

Using $\Delta \leq \sigma^2/10$,

$$aT = \left(\sigma^2 + \frac{33}{32}\Delta\right)T \leq \left(1 + \frac{33}{320}\right)\sigma^2T < 2\sigma^2T \leq m,$$

where the last step uses $m = \lceil 2\sigma^2T \rceil \geq 2\sigma^2T$. Hence the budget condition holds, and we may invoke Theorem 10 to conclude $\max_{\theta} \mathbb{E}_{\theta}[\text{Reg}(\pi)] \geq c\delta\Delta/T$. $\square$

Theorem 11 (Lower bound for fourth-moment bounded distributions). *Fix $T \in \mathbb{N}$, $\sigma \in (0, 0.5)$, and $\delta \in (0, 1/4)$. Let $\mathcal{R} = \text{Rademacher}$. Set $m := \lceil 2\sigma^2T \rceil$ and let*

$$\Delta = \min \left\{ \frac{\sigma^2}{10}, \sqrt{\frac{\log(1/(4\delta))}{4m}} \right\} = \min \left\{ \frac{\sigma^2}{10}, \sqrt{\frac{\log(1/(4\delta))}{8\sigma^2T}} \right\}.$$

Then there exist mean-zero distributions $D_0, D_1 \in \mathcal{F}_{\sigma} := \{\mathcal{D} : \text{Var}(\mathcal{D}) \leq 4\sigma^2, \kappa(\mathcal{D}) \leq 2\}$ satisfying $\text{Var}(D_0) = \sigma^2$ and $\text{Var}(D_1) = \sigma^2 + \Delta$ such that for every policy π ,

$$\max_{\theta \in \{0,1\}} \mathbb{E}_{\theta}[\text{Reg}(\pi)] \geq c \cdot \frac{\delta}{T} \min \left\{ \frac{\sigma^2}{10}, \sqrt{\frac{\log(1/(4\delta))}{8\sigma^2T}} \right\},$$

for a universal constant $c > 0$.

Proof. Apply Theorem 16 with $n = m$ to obtain $D_0, D_1 \in \mathcal{F}_{\sigma}$ with $\text{Var}(D_0) = \sigma^2$, $\text{Var}(D_1) = \sigma^2 + \Delta$, and the testing guarantee

$$\inf_{\hat{C}_m: \mathbb{R}^m \rightarrow \{0,1\}} \frac{1}{2} \left(\Pr_{X \sim D_0^m}(\hat{C}_m(X) = 0) + \Pr_{X \sim D_1^m}(\hat{C}_m(X) = 1) \right) \leq 1 - \delta,$$

i.e. every test has average error at least δ . Thus (11) holds.

As in the proof of Theorem 12, it remains to check the budget condition needed by Theorem 10. With $r = \Delta/(32T)$ and $s_1 = \sigma^2 + \Delta$, we require

$$m \geq T - \frac{T}{(\sigma^2 + \Delta) + 1 + \Delta/32}.$$

Let $a := \sigma^2 + \frac{33}{32}\Delta$. Since $\frac{a}{1+a} \leq a$,

$$T - \frac{T}{1+a} = T \cdot \frac{a}{1+a} \leq aT.$$

Using $\Delta \leq \sigma^2/10$,

$$aT = \left(\sigma^2 + \frac{33}{32}\Delta\right)T \leq \left(1 + \frac{33}{320}\right)\sigma^2T < 2\sigma^2T \leq m,$$

where $m = \lceil 2\sigma^2T \rceil \geq 2\sigma^2T$. Thus the budget condition holds, and invoking Theorem 10 completes the proof. $\square$

E ADDITIONAL EXPERIMENTAL DETAILS

We implement practical variants of our algorithms with exploration $O(\log T)$ per arm, preserving the algorithmic structure faithfully (data-adaptive stopping rule of Algorithm 1; three batch phases and fine-allocation formula of Algorithm 2, line 10). Standard sample variance replaces the median-of-means estimator of Proposition 1; the two are equivalent for Gaussian distributions. We provide code for our experiments in the supplementary materials.