Bandit-based Maximum Inner Product Search with Data-Dependent Confidence Intervals

Yoichi Sasaki*
NEC Corporation

Yuzuru Okajima*
NEC Corporation

Abstract

Maximum inner product search is a fundamental problem in recommender systems, information retrieval, and machine learning. Recently proposed bandit-based approaches have achieved high scalability with respect to dimensionality and offer favorable precision-speedup trade-offs. However, the lengths of their confidence intervals are determined independently of the actual reward distributions, which can lead to search inefficiency in practice. In this paper, we propose a data-dependent bandit-based algorithm in which the lengths of the confidence intervals are adaptively adjusted based on observed samples. Theoretical analysis demonstrates that our algorithm guarantees δ-correctness for a broad class of distributions, including all log-concave continuous distributions, and that the sample complexity can be reduced adaptively according to individual reward distributions. In experiments, our approach outperformed existing algorithms on both synthetic and real-world datasets.

1 INTRODUCTION

Maximum inner product search (MIPS) is a fundamental problem that is ubiquitous in many research areas such as recommender systems, information retrieval, and machine learning (Shrivastava and Li, 2014; Bachrach et al., 2014; Lewis et al., 2020). Given a query $\boldsymbol{q} \in \mathbb{R}^d$ and a set of n targets $V = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$, the goal of MIPS is to find the target with the maximum inner product:

$$i^* = \arg\max_{i \in [n]} \boldsymbol{q}^T \boldsymbol{v}_i,$$

where $[n]$ denotes $\{1, \ldots, n\}$. Since the naive computation requires $O(nd)$ time, approximate search algorithms with better query times have been developed for large and/or high-dimensional data. To reduce query-time computation, most approaches rely on prebuilt indexing structures, such as locality sensitive hashing (LSH) (Shrivastava and Li, 2014), quantization (Dai et al., 2020), trees (Song et al., 2021), and graphs (Chen et al., 2025).

Recently, two bandit-based MIPS approaches, BoundedME (Liu et al., 2019) and BanditMIPS (Tiwari et al., 2024), have been proposed to improve search efficiency on high-dimensional data. They formulate MIPS as a best arm identification (BAI) problem: the n targets are treated as n arms; when arm $i \in [n]$ is pulled, a coordinate j is randomly selected from $[d]$, and the product $q_j v_{i,j}$ is returned as the reward. Let Q and V_i for $i \in [n]$ denote the random variables representing q_j and $v_{i,j}$, respectively, for randomly chosen j. The expected reward μ_i of arm i is the normalized inner product, that is, $\mu_i = \mathbb{E}[QV_i] = \frac{1}{d}\sum_{j \in [d]} q_j v_{i,j}$. Thus, MIPS is reduced to the problem of identifying the best arm with the highest expected reward. In comparison with index-based MIPS algorithms, the bandit-based approaches offer high scalability with respect to dimensionality, require minimal preprocessing, and allow users to specify the error probability at query time to directly control the suboptimality of the search results with theoretical guarantees. Specifically, the bandit-based algorithms are δ-correct: for any $0 < \delta < 1$ and $0 \leq \epsilon$ chosen by the user, it is guaranteed that, with probability at least $1-\delta$, the returned solution $\hat{i}$ is ϵ-best, i.e., $\mu_{i^*} - \mu_{\hat{i}} \leq \epsilon$. The ability to provide such guarantees on suboptimality control is a unique strength of bandit-based algorithms.

A major limitation of the existing bandit-based MIPS approaches is that the lengths of the confidence intervals depend only on the sample size and are independent of the actual observed rewards. Because the same interval length is applied to all arms with the same sample size, the intervals must be wide enough to accommodate the worst case—that is, the highest

*Equal contribution. Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300. Copyright 2026 by the author(s).

variance of rewards among all arms. This leads to overly conservative estimates and reduces search efficiency. For example, BanditMIPS requires the user to specify a variance proxy σ^2, under the assumption that the rewards from any arm follow a sub-Gaussian distribution of parameter σ^2. Similarly, BoundedME requires a value range $[a, b]$ to be provided, assuming that all rewards fall within this range. While these dispersion parameters, σ^2 and $[a, b]$, must cover the worst case—that is, the highest variance among all arms—the actual reward distributions of individual arms may exhibit much smaller variances. For example, if $QV_1, \ldots, QV_n$ are normally distributed with variances $\sigma_1^2, \ldots, \sigma_n^2$, the parameter σ^2 must be chosen conservatively so that $\sigma^2 \geq \max_i \sigma_i^2$, which results in confidence intervals that are excessively wide relative to the actual distributions, thereby slowing down the search process. To address this inefficiency, we need a method to adaptively adjust the confidence interval lengths based on the actual reward distribution of individual arms.

To tackle the problem of fixed-length intervals, several *data-dependent* BAI algorithms have been developed in the bandit community. These algorithms adaptively adjust the lengths of confidence intervals under the assumption of boundedness (Lu et al., 2021) or normality (Jourdan et al., 2023). However, when applying BAI algorithms to MIPS, we must handle a broader class of reward distributions, since the product of Q and V_i is not normally distributed even when Q and V_i are independent normal random variables.

In this paper, we propose a bandit-based MIPS algorithm that adaptively changes the lengths of confidence intervals according to the reward distributions of individual arms. Leveraging the fact that maximizing $\mathbb{E}[QV_i]$ is equivalent to maximizing $\mathbb{E}[V_i^2] - \mathbb{E}[(Q - V_i)^2]$, we choose to maximize the latter by estimating $\mathbb{E}[V_i^2]$ and $\mathbb{E}[(Q - V_i)^2]$ separately, for three reasons. First, $\mathbb{E}[V_i^2]$ can be estimated from the targets alone as a preprocessing step before queries are given. Second, $\mathbb{E}[(Q - V_i)^2]$ can be small if Q and V_i are similarly distributed. Finally, because V_i^2 and $(Q - V_i)^2$ are always nonnegative, we can assume nonnegativity when constructing confidence intervals. In addition to nonnegativity, we assume that the moments of V_i^2 and $(Q - V_i)^2$ are bounded in terms of their expected values. This moment boundedness assumption holds for a broad range of distributions including normal, chi-squared, and any other log-concave continuous distributions (Schudy and Sviridenko, 2012, Lemma 7.3). Under this assumption, we derive an adaptive confidence interval that is multiplicative rather than additive, meaning that its length is proportional to the sample mean of the random variable. Theoretical analysis guarantees that our bandit-based MIPS algorithm with these intervals returns ϵ-best arms with probability at least $1 - \delta$ for given δ and ϵ, with a worst-case sample complexity that adapts to individual reward distributions. Performance evaluations on synthetic and real-world datasets reveal that our algorithm returned correct results more consistently with lower sample complexity and faster wall-clock time than the state-of-the-art bandit-based MIPS algorithms.

In summary, we make the following contributions:

- We derive a data-dependent confidence interval that can be applied to a broad range of distributions including normal, chi-squared, and any other log-concave continuous distributions, which could be of independent interest.
- We propose a bandit-based MIPS algorithm that utilizes the data-dependent confidence intervals to guarantee ϵ-best results with probability at least $1 - \delta$ and has a sample complexity that can be small adaptively according to individual reward distributions.
- Our experiments on both synthetic and real-world datasets showed that our algorithm outperformed the state-of-the-art bandit-based MIPS algorithms in terms of correctness, sample complexity, and wall-clock time.

2 RELATED WORK

Index-based MIPS algorithms MIPS is closely related to nearest neighbor search (NNS), but its main difficulty arises from the fact that the inner product does not satisfy the triangle inequality, which is often leveraged by NNS algorithms for efficient search. Most MIPS algorithms rely on preconstructed indexing structures, which can be categorized into four groups: locality sensitive hashing (LSH), quantization, trees, and graphs. LSH, a popular approach to NNS, is applied to MIPS with asymmetric transformation such as L2 (Shrivastava and Li, 2014), Correlation (Shrivastava and Li, 2015), XBOX (Bachrach et al., 2014; Huang et al., 2018), and random XBOX transformation (Zhao et al., 2023). Vector quantization is extended to MIPS by norm-explicit quantization (Dai et al., 2020) and multiple indices (Sun et al., 2023). Space partitioning trees are used with dimension reduction to avoid the curse of dimensionality (Bachrach et al., 2014; Song et al., 2021). Proximity graph-based methods such as NSW are also extended to the inner product metric (Malkov and Yashunin, 2018; Morozov and Babenko, 2018; Tan et al., 2021; Chen et al., 2025).

A disadvantage of index-based algorithms is that the suboptimality of the search, such as error probability δ, cannot be directly controlled at query time (Liu et al., 2019).

Bandit-based MIPS algorithms In contrast to index-based algorithms, bandit-based algorithms sample directly from the original data until the best target is found, which allows users to directly control suboptimality during querying. The first bandit-based MIPS algorithm, BoundedME (Liu et al., 2019), combines the median elimination strategy (Even-Dar et al., 2006) with confidence bounds for sampling without replacement from a bounded range. BanditMIPS (Tiwari et al., 2024) is based on successive elimination with the Hoeffding bounds (Even-Dar et al., 2006) and achieved high scalability with high dimensional data. A similar bandit-based MIPS algorithm under the sub-Gaussian assumption is also utilized to approximate the softmax function (Baharav et al., 2024).

BAI algorithms BAI algorithms have been studied in the bandit literature both in the fixed budget and fixed confidence settings. This paper focuses on the fixed confidence setting, where the algorithm is expected to return the best arm or an ϵ-best arm with probability at least $1-\delta$. Fixed confidence algorithms include median elimination, successive elimination, LUCB (Kalyanakrishnan et al., 2012) and lil'UCB (Jamieson et al., 2014). Most algorithms utilize confidence intervals of fixed length under the assumption of bounded or sub-Gaussian distributions with a known variance proxy. Despite the practical relevance, only a few BAI algorithms utilize confidence intervals with lengths depending on the empirical distributions. The empirical Bernstein inequality for bounded distributions (Maurer and Pontil, 2009) has been used for BAI (Gabillon et al., 2012; Lu et al., 2021). An empirical variance-aware BAI algorithm has been proposed for Gaussian distributions (Jourdan et al., 2023). Geist (2015) derived a multiplicative Hoeffding bound for gamma random variables under the assumption of a lower bound on the form factors is known. These approaches are restricted to bounded, Gaussian and gamma distributions, and, unlike our approach, cannot be applied to a broad class including different types of unbounded distributions.

3 CONFIDENCE INTERVALS FOR MOMENT BOUNDED VARIABLES

In this section, we derive a data-dependent confidence interval whose length adapts to the empirical distribution under an assumption about moment boundedness. First, we review the definition and basic properties of moment bounded random variables.

Definition 1. *In this paper, a random variable X is called L-moment bounded if*

$$\forall k \in \mathbb{Z}_{\geq 1},\ \mathbb{E}[|X|^k] \leq L^k k! \tag{1}$$

holds[1].

The moment boundedness condition (1) is satisfied by a wide range of distributions. For example, any nonnegative log-concave continuous random variable is moment bounded with $L = \mathbb{E}[X]$, and any (possibly negative) log-concave continuous random variable is moment bounded with $L = \frac{1}{\ln 2}\mathbb{E}[|X|]$ (Schudy and Sviridenko, 2012, Lemmas 7.2 and 7.3). Log-concave continuous distributions include normal, exponential, logistic, Laplace, chi, chi-squared, Weibull, gamma, and beta, where the last four are log-concave only for some parameter values.

One of the useful distributions that satisfies the moment boundedness is the noncentral chi-squared distribution. If a random variable X is normally distributed, X^2 follows the noncentral chi-squared distribution with $k = 1$, which implies X^2 is $2\mathbb{E}[X^2]$-moment bounded (see Section A of the Supplementary Material). This observation supports the plausibility of the moment boundedness assumption for our MIPS algorithm, which estimates confidence intervals for $\mathbb{E}[V_i^2]$ and $\mathbb{E}[(Q-V_i)^2]$.

Throughout this paper, we often assume the following:

Assumption 2. *For a random variable X, (i) X is $c\mathbb{E}[|X|]$-moment bounded for some constant $c \geq 1$, and (ii) $X - \mathbb{E}[X]$ is $b\mathbb{E}[|X|]$-moment bounded for some constant $b \geq 1$.*

The assumption (i) holds for any log-concave continuous distributions by choosing $c = 1$ (nonnegative case) or $c = \frac{1}{\ln 2}$ (general case), and for the noncentral chi-squared distribution with $k = 1$ by choosing $c = 2$. The following two lemmas suggest how to choose b given c (all lemmas and theorems are proved in the Supplementary Material):

Lemma 3. *If a random variable X is L-moment bounded, $X - \mathbb{E}[X]$ is $2L$-moment bounded.*

Lemma 4. *If a random variable X is $c\mathbb{E}[|X|]$-moment bounded and $Y = X - \mathbb{E}[X]$ is $a\mathbb{E}[|Y|]$-moment*

[1] A more strict condition

$$\forall k \in \mathbb{Z}_{\geq 1},\ \mathbb{E}\left[|X|^k\right] \leq k \cdot L \cdot \mathbb{E}\left[|X|^{k-1}\right] \tag{2}$$

is used as the definition of moment boundedness by Schudy and Sviridenko (2012, Definition 1.1). Any X satisfying (2) also satisfies (1).

bounded, Y is $\sqrt{2}ac\mathbb{E}[|X|]$-moment bounded. If X is nonnegative in addition, Y is $a\sqrt{2c^2-1}\mathbb{E}[X]$-moment bounded.

Example For any nonnegative log-concave continuous random variable X, $X - \mathbb{E}[X]$ is $\frac{1}{\ln 2}\mathbb{E}[X]$-moment bounded according to Lemma 4 with $(c, a, b) = (1, \frac{1}{\ln 2}, \frac{1}{\ln 2})$.

To obtain the confidence intervals, we give the following assumption and lemma:

Assumption 5. *For a nonnegative random variable X, (i) $X - \mathbb{E}[X]$ is sub-Gaussian on the left tail with parameter $\sqrt{2}c\mathbb{E}[X]$ for some constant $c \geq 1$, and (ii) $X - \mathbb{E}[X]$ is sub-gamma with parameters $(\sqrt{2}b\mathbb{E}[X], b\mathbb{E}[X])$ for some constant $b \geq 1$.*

Lemma 6. *If a nonnegative random variable X satisfies Assumption 2, X satisfies Assumption 5 for the same constants, b and c.*

Using this lemma, we obtain the confidence intervals for the nonnegative case:

Theorem 7. *(Nonnegative case) Let X be a nonnegative random variable satisfying Assumption 5. Let $X_1, X_2, \ldots, X_t$ be i.i.d. random variables following the same distribution as X. Let us denote $\overline{X} = (X_1 + \cdots + X_t)/t$. For any $\delta \in (0, 1)$, we have*

$$\mathbb{P}\left(\frac{\overline{X}}{1 + b_L(t, \delta)} > \mathbb{E}[X]\right) \leq \delta$$

with

$$b_L(t, \delta) \coloneqq 2b\sqrt{\frac{\ln(1/\delta)}{t}} + b\frac{\ln(1/\delta)}{t}$$

and

$$\mathbb{P}\left(\frac{\overline{X}}{(1 - b_U(t, \delta))_+} < \mathbb{E}[X]\right) \leq \delta$$

with

$$b_U(t, \delta) \coloneqq 2c\sqrt{\frac{\ln(1/\delta)}{t}},$$

where $(\cdot)_+ = max\{\cdot, 0\}$ and with the convention $\frac{1}{0} = \infty$. For $t \leq 4c^2 \ln(1/\delta)$, $1 - b_U(t, \delta) \leq 0$, and the confidence interval does not have a finite upper endpoint.

Using the bounds for the nonnegative case in Theorem 7, we derive the bounds for the general (possibly negative) case:

Theorem 8. *(General case) Let X be a random variable such that both X and $|X|$ satisfy Assumption 2. Let $X_1, X_2, \ldots, X_t$ be i.i.d. random variables following the same distribution as X. Let us denote $\overline{X} = (X_1 + \cdots + X_t)/t$ and $\overline{|X|} = (|X_1| + \cdots + |X_t|)/t$. For any $\delta \in (0, 1)$, we have*

$$\mathbb{P}\Big(\overline{X} - \mathbb{E}[X] > b_G(t, \delta)\overline{|X|}\Big) \leq \delta$$

and

$$\mathbb{P}\Big(\mathbb{E}[X] - \overline{X} > b_G(t, \delta)\overline{|X|}\Big) \leq \delta$$

with

$$b_G(t, \delta) \coloneqq \frac{b_L(t, \delta/2)}{(1 - b_U(t, \delta/2))_+}.$$

For $t \leq 4c^2 \ln(2/\delta)$, $1 - b_U(t, \delta/2) \leq 0$, and the confidence interval has neither a finite upper endpoint nor a finite lower endpoint.

For notational simplicity, we omit b and c from the arguments of $b_L(t, \delta)$, $b_U(t, \delta)$, and $b_G(t, \delta)$ throughout the paper. While our MIPS algorithm only handles nonnegative random variables, Theorem 8 is presented for completeness and independent interest.

An advantage of our confidence intervals is that c and b only depend on the family of distributions (e.g. normal or chi-squared) and do not depend on the parameters of individual distributions. For example, if $QV_1, \ldots, QV_n$ are normally distributed with different variances $\sigma_1^2, \ldots, \sigma_n^2$, the sub-Gaussian parameter σ^2 of BanditMIPS must be set to be at least as high as the maximum of them. In contrast, $c = \frac{1}{\ln 2}$ and $b = 2c$ can be used for any log-concave continuous distributions including the normal distribution independently of the values of $\sigma_1^2, \ldots, \sigma_n^2$.

In our MIPS problem, V_i^2 and $(Q - V_i)^2$ are sampled without replacement from their corresponding finite-length vector elements, $\{v_{i,j}^2\}_{j=1}^d$ and $\{(q_j - v_{i,j})^2\}_{j=1}^d$. Theorems 7 and 8 also hold in the case of sampling without replacement, by applying Theorem 4 of Hoeffding (1963) (see Section F of the Supplementary Material for details). Furthermore, V_i^2 and $(Q - V_i)^2$ do not need to satisfy Assumption 2 directly, because Assumtion 5 is sufficient for Theorem 7 to hold. Based on this observation, in practice, we only need to assume that the vector elements are drawn from underlying distributions that satisfy Assumption 5, and that d is sufficiently large so that V_i^2 and $(Q - V_i)^2$ also satisfy Assumption 5. This assumption is justified by the relative deviation bound from Vapnik and Chervonenkis (1974), which indicates that the tail probability of an empirical distribution is upper bounded in terms of (the square root of) the tail probability of its underlying distribution with a probabilistic guarantee; for example, if the underlying distribution has a sub-Gaussian tail, the empirical distribution will also have

a similar sub-Gaussian tail (see Section G of the Supplementary Material for details). The assumption of large d is appropriate for the bandit-based approach, as sampling is more efficient than naive scanning only when the vectors are high-dimensional.

4 PROPOSED MIPS ALGORITHM

In this section, we propose our MIPS algorithm. We first present LUCB for MIPS, an abstract framework that extends LUCB, a commonly used BAI algorithm, to the MIPS setting and can be combined with different confidence intervals. We then introduce LUCB-MB, a concrete instance that employs the moment boundedness-based confidence intervals.

4.1 LUCB for MIPS

We now explain how LUCB for MIPS works step by step according to the pseudo code shown in Algorithm 1.

4.1.1 Input

Algorithm 1 receives $\boldsymbol{q}$ and $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ as input. To simplify the notation, we assume that the d indices of these vectors are randomly permuted before searching so that random sampling without replacement is represented as taking the first elements of the vectors sequentially in index order [2]. The algorithm also receives a dictionary D_0, which stores two kinds of sample statistics, the sample mean and the sample size, for random variables (in the case of LUCB-MB, V_i^2 and $(Q - V_i)^2$ for $i \in [n]$). The sample statistics are updated in every time step, and the dictionary at time T is denoted as D_T. For any random variable X, we denote the sample size stored in D_T as $D_T[X].\tau$ and the sample mean as $D_T[X].\hat{\mu}$. For any random variables depending on the query (i.e., $(Q-V_i)^2$ for LUCB-MB), the sample size and the sample mean stored in D_0 are initialized to zero. Because the other random variables (i.e., V_i^2 for LUCB-MB) only depend on the targets, we can collect their samples before querying and store their statistics in D_0. We can collect their samples as preprocessing or we can also use the samples collected through preceding querying.

4.1.2 Variables and Functions

The variables and functions appearing in Algorithm 1 are described here. Variable t_i represents that the first t_i elements of $\boldsymbol{v}_i$ are used to calculate the sample statistics in the dictionary. Function update_stats updates the sample statistics with the newly observed values, q_{t_i} and v_{i,t_i}. Using $(Q - V_i)^2$ as an example, the updates can be written as $D_T[(Q-V_i)^2].\tau \leftarrow \theta+1$ and $D_T[(Q - V_i)^2].\hat{\mu} \leftarrow (\theta D_{T-1}[(Q - V_i)^2].\hat{\mu} + (q_{t_i} - v_{i,t_i})^2)/(\theta + 1)$, where $\theta = D_{T-1}[(Q - V_i)^2].\tau$. Function update_estimate updates $\hat{\mu}_i, u_i, l_i$, which are the estimator, upper and lower bounds for μ_i, respectively.

[2]This is just for the notation, and there is no need to actually rearrange the data; instead, you can simply prepare a permutation array π containing randomly shuffled indices from 1 to d and interpret access to the i-th element as access to π_i-th element.

4.1.3 Initialization

The variables are initialized in Lines 1 to 4 of Algorithm 1. The first elements of all vectors, q_1 and $v_{1,1}, \ldots, v_{n,1}$, are used to update the statistics recorded in the dictionary, and subsequently $\hat{\mu}_i$, u_i and l_i.

4.1.4 Main Loop

The main loop starts at Line 6 in Algorithm 1. The target with the highest estimator is labeled as h_t, and the target with the highest upper bound excluding h_t is labeled as j_t (Lines 7-8). If the stopping condition $u_{j_t} < l_{h_y} + \epsilon$ is met, h_t is returned as an ϵ-best arm (Lines 9-10). If the stopping condition is not satisfied yet, the variables are updated for these two targets. If $t_i < d$, $\boldsymbol{v}_i$ has unobserved elements; t and t_i are incremented, and the sample statistics are updated with the newly observed values, q_{t_i} and v_{i,t_i} (Lines 13-15). If $\boldsymbol{v}_i$ still has unobserved elements, the estimator and the upper and lower bounds are calculated approximately (Lines 16-17). If all elements of $\boldsymbol{v}_i$ have been observed, the normalized inner product is calculated exactly and is set to $\hat{\mu}_i$,u_i, and l_i (Lines 18-19).

4.2 LUCB-MB

LUCB-MB is an instance of LUCB for MIPS that utilizes the moment boundedness-based confidence intervals. LUCB-MB is based on an additive decomposition: $\mathbb{E}[QV_i] = \frac{1}{2}\big(\mathbb{E}[Q^2] + \mathbb{E}[V_i^2] - \mathbb{E}[(Q - V_i)^2]\big)$. Because $\mathbb{E}[Q^2]$ does not depend on the targets, maximizing $\mathbb{E}[QV_i]$ is equivalent to maximizing $\omega_i := \frac{1}{2}\big(\mathbb{E}[V_i^2] - \mathbb{E}[(Q - V_i)^2]\big)$. As we discussed at the end of Section 3, we assume that V_i^2 and $(Q - V_i)^2$ satisfy Assumption 5 for all $i \in [n]$. We apply Theorem 7 to get their confidence bounds, and add them to get confidence bounds for ω_i. Specifically, update_estimate of LUCB-MB returns the following:

$$\hat{\mu}_{it} = \frac{1}{2}\big(D_T[V_i^2].\hat{\mu} - D_T[(Q - V_i)^2].\hat{\mu}\big),$$
$$l_{it} = \frac{1}{2}\big(l_N(V_i^2, t) - u_N((Q - V_i)^2, t)\big),$$
$$u_{it} = \frac{1}{2}\big(u_N(V_i^2, t) - l_N((Q - V_i)^2, t)\big),$$

Algorithm 1 LUCB for MIPS

Input: $\boldsymbol{q}, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$, D_0, δ, ϵ, b, c

$T \leftarrow 1$
for $i \in [n]$ **do** ▷ Initialization for each target
  $t_i \leftarrow 1$
  $D_1 \leftarrow \mathsf{update_stats}(D_0, q_1, v_{i1}, t_i)$
  $\hat{\mu}_i, u_i, l_i \leftarrow \mathsf{update_estimate}(D_1, t_i, \delta, b, c)$
while true **do** ▷ Main loop
  $h_t \leftarrow \arg\max_i \hat{\mu}_i$
  $j_t \leftarrow \arg\max_{i \neq h_t} u_i$
  if $u_{j_t} < l_{h_t} + \epsilon$ **then** ▷ Stopping condition
    return h_t
  else
    for $i \in \{h_t, j_t\}$ **do**
      if $t_i < d$ **then**
        $T \leftarrow T + 1; t_i \leftarrow t_i + 1$
        $D_T \leftarrow \mathsf{update_stats}(D_{T-1}, q_{t_i}, v_{it_i}, t_i)$
        if $t_i < d$ **then**
          $\hat{\mu}_i, u_i, l_i \leftarrow \mathsf{update_estimate}(D_T, t_i, \delta, b, c)$
        else ▷ All elements of $\boldsymbol{v}_i$ are observed
          $\hat{\mu}_i, u_i, l_i \leftarrow \mathsf{exact_calculation}(D_T)$

where

$$l_N(X,t) \coloneqq \frac{D_T[X].\hat{\mu}}{\rho_l(D_T[X].\tau)}, \quad u_N(X,t) \coloneqq \frac{D_T[X].\hat{\mu}}{\rho_u(D_T[X].\tau)},$$

$$\rho_l(\tau) \coloneqq 1 + b_L(\tau, \tilde{\delta}(\tau)), \quad \rho_u(\tau) \coloneqq \Big(1 - b_U(\tau, \tilde{\delta}(\tau))\Big)_+,$$

$$\tilde{\delta}(\tau) \coloneqq \frac{\delta}{4k\tau^2}.$$

Functions $u_N(X,t)$ and $l_N(X,t)$ represent the upper and lower bounds for $\mathbb{E}[X]$ of a nonnegative X. Variable k denotes the number of the random variables for which confidence intervals are calculated. LUCB-MB uses $k = 2n$ because it handles $2n$ variables, V_i^2 and $(Q - V_i)^2$ for $i \in [n]$. The error probability $\tilde{\delta}(\tau)$ depends on the sample size τ so that the sum of $\tilde{\delta}(\tau)$ for both the upper and lower bounds of all the k variables for $\tau \in [1, \infty)$ does not exceeds δ using the fact that $\sum_{\tau=1}^{\infty} \frac{1}{2\tau^2} = \frac{\pi^2}{12} < 1$. To simplify the notation, the definitions of these functions omit b, c, k, and δ as arguments.

4.3 Theoretical Analysis

We analyze the correctness and sample complexity of LUCB-MB. The computational complexity is linear in the sample complexity because only $O(1)$ computation is required per sampled element. Without loss of generality, we assume that the targets are sorted in the descending order of their inner products with the query, which indicates $i^* = 1$. Let us denote $\Delta_i = \mu_1 - \mu_i$ for $i \geq 2$ and $\Delta_1 = \Delta_2$.

Here is our main result:

Theorem 9. *Assume V_i^2 and $(Q - V_i)^2$ satisfy Assumption 5 for all $i \in [n]$. With probability at least $1 - \delta$, LUCB-MB returns an ϵ-best target correctly, and the sample complexity is*

$$O\left(\sum_{i=1}^{n} \left(\frac{m_i}{\Delta_i}\right)^2 \ln\left(\frac{n m_i}{\delta \Delta_i}\right)\right)$$

for $\epsilon = 0$, where $m_i \coloneqq \max\{\mathbb{E}[V_i^2], \mathbb{E}[(Q - V_i)^2]\}$.

By comparing this sample complexity with that of BanditMIPS,

$$O\left(\sum_{i=1}^{n} \frac{\sigma^2}{\Delta_i^2} \ln\left(\frac{n}{\delta \Delta_i}\right)\right),$$

we can see that the sample complexity in Theorem 9 can be more efficient than that of BanditMIPS when m_i is smaller enough than σ. While σ^2 must be at least as high as $\max_i \sigma_i^2$ for theoretical guarantee, where σ_i^2 is a variance proxy for QV_i, the value of m_i can be small according to the individual reward distribution of the i-th target. If V_i^2 has been sampled enough as preprocessing, m_i is reduced to $\mathbb{E}[(Q - V_i)^2]$. Although the targets with small i require many samples due to small Δ_i for BanditMIPS, this hardness can be mitigated in our approach if $\mathbb{E}[(Q - V_i)^2]$ are small too for the targets with small i; this situation is likely especially when the values of $\mathbb{E}[V_i^2]$ are not very different among the targets.

We conclude this section by discussing the three reasons we choose to estimate $\mathbb{E}[V_i^2]$ and $\mathbb{E}[(Q - V_i)^2]$ separately, instead of estimating $\mathbb{E}[QV_i]$ directly. The first reason is that $\mathbb{E}[V_i^2]$ can be estimated during preprocessing. If $\mathbb{E}[V_i^2]$ is estimated for targets in advance, LUCB-MB is expected to disregard targets with so small $\mathbb{E}[V_i^2]$ that they are not likely to have large $\mathbb{E}[QV_i]$. The second reason is that separating the estimation of $\mathbb{E}[V_i^2]$ would help the algorithm to find the difference among small $\mathbb{E}[(Q - V_i)^2]$ values. As discussed in the previous paragraph, $\mathbb{E}[(Q - V_i)^2]$ can be very small for small i, and the variance adaptivity of LUCB-MB is a good fit to find differences among such small values. The third reason is that, since V_i^2 and $(Q - V_i)^2$ are nonnegative, we can apply Theorem 7 to them. Even though Theorem 8 provides general-case intervals, Theorem 7 provides better intervals for nonnegative variables. While the confidence interval in Theorem 8 has neither finite upper nor lower endpoints for small values of t, the confidence interval in Theorem 7 always has a finite lower endpoint even for small t.

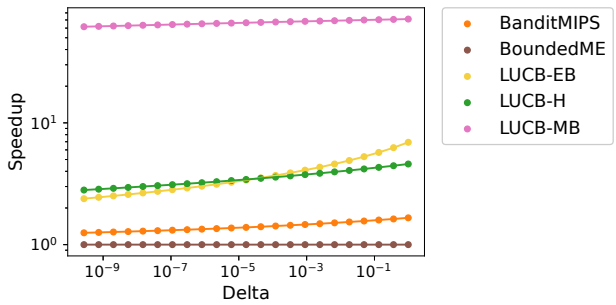

Figure 1: Speedup vs Delta: SIMILAR_TARGETS (n=1,000, d=10,000).

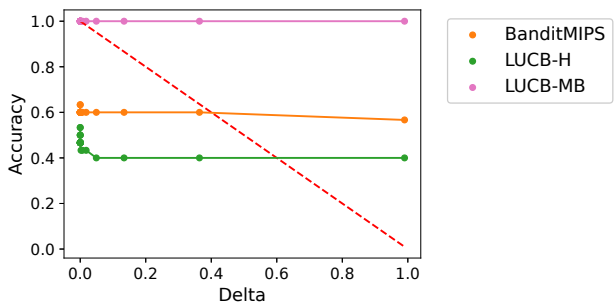

Figure 2: Accuracy vs Delta: DIFFERENT_VARIANCES (n=1,000, d=10,000).

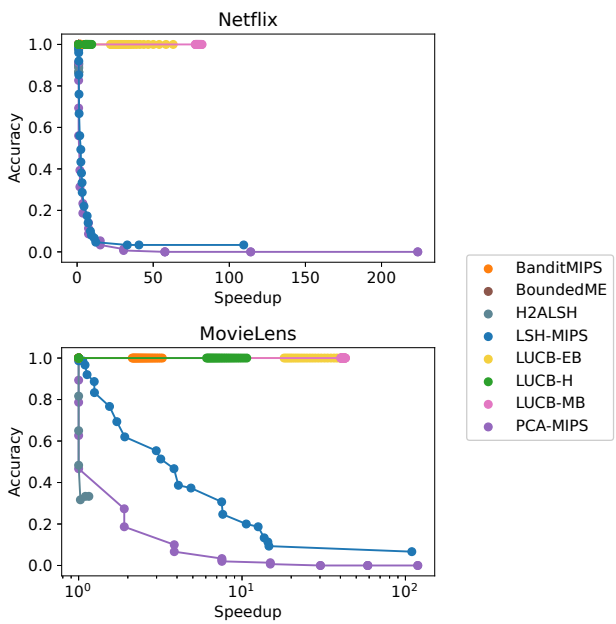

Figure 3: Accuracy vs Speedup: Netflix (n=1,000, d=10,000) and MovieLens (n=1,000, d=5,000).

Algorithm 2 SIMILAR_TARGETS

1: $\boldsymbol{q} \leftarrow [\mathrm{N}(0,1)$ for $j \in [d]]$
2: $\boldsymbol{p} \leftarrow [\mathrm{Beta}(5,5)$ for $i \in [n]]$
3: **for** $i \in [n]$ **do**
4: $\quad \boldsymbol{v}_i \leftarrow [p_i^2 q_j + (1-p_i)^2\,\mathrm{N}(0,1)$ for $j \in [d]]$

Algorithm 3 DIFFERENT_VARIANCES

1: $\sigma_q^2 \leftarrow \mathrm{InvGamma}(1,1)$; $\boldsymbol{q} \leftarrow [\mathrm{N}(0,\sigma_q^2)$ for $j \in [d]]$
2: **for** $i \in [n]$ **do**
3: $\quad \sigma_i^2 \leftarrow \mathrm{InvGamma}(1,1)$; $\boldsymbol{v}_i \leftarrow [\mathrm{N}(0,\sigma_i^2)$ for $j \in [d]]$

5 EXPERIMENTS

5.1 Experimental Setup

5.1.1 Bandit-based Algorithms

We compared LUCB-MB with BoundedME and BanditMIPS. As a baseline, we also evaluated *LUCB-H* and *LUCB-EB*. Both are instances of LUCB for MIPS, as is LUCB-MB. However, unlike LUCB-MB, they estimate expected rewards using confidence intervals based on the *Hoeffding* and *empirical Bernstein* bounds, respectively. For the datasets with bounded values, the sub-Gaussian parameter σ^2 was set to the minimum value for which the theoretical guarantee holds. For the datasets with unbounded values, the value of σ^2 was set to the empirical variance of rewards. For all datasets, LUCB-MB was run with $(c, b) = (1, \frac{1}{\ln 2})$. As preprocessing, the values of $\mathbb{E}[V_i^2]$ were exactly computed and stored in D_0. This setting reflects a practical scenario in which the targets are available prior to querying. The effect of δ was observed by varying it over $[10^{-10}, 0.99]$ with ϵ fixed at 0. All algorithms were implemented and evaluated by leveraging the code released by the original authors of BanditMIPS[3].

5.1.2 Index-based Algorithms

As baselines, we also evaluated three index-based MIPS algorithms: LSH-MIPS (Shrivastava and Li, 2014), H2-ALSH-MIPS (Huang et al., 2018), and PCA-MIPS (Bachrach et al., 2014).

5.1.3 Metrics

Query times were measured by sample complexity (defined as the number of coordinate-wise multiplications performed) and wall-clock time. The speedup is defined as $\frac{\text{query time of the naive scan}}{\text{query time of the compared algorithm}}$, where query time was measured by sample complexity unless oth-

[3]https://github.com/ThrunGroup/BanditMIPS

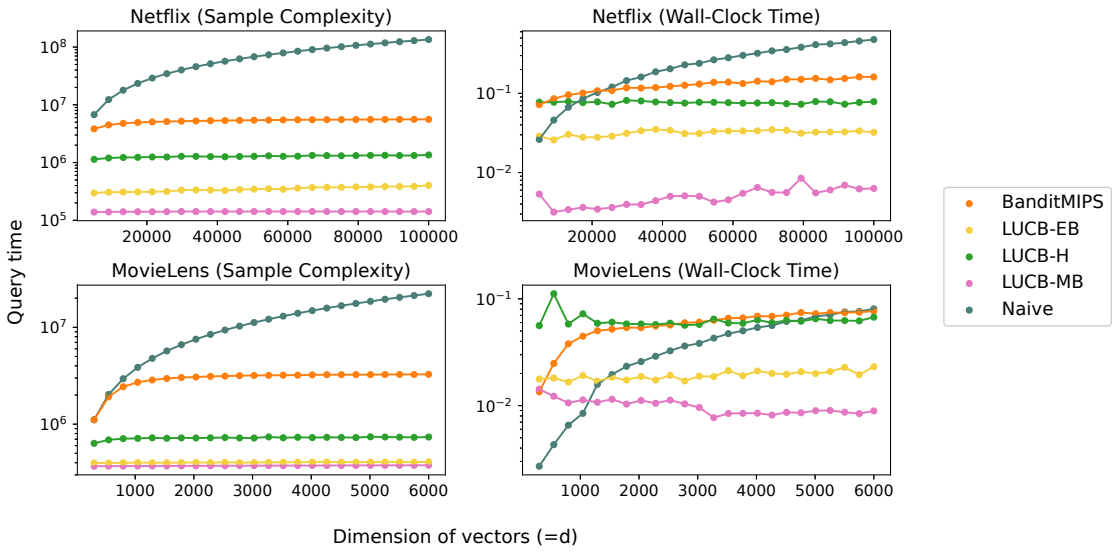

Figure 4: Query Time vs Dimension (d): Netflix (n=1,349, d=100,000) and MovieLens (n=3,705, d=6,000) for varying d values with $\delta = 0.01$.

erwise noted. In this paper, *naive scan* refers to computing the inner product between the query and all targets by a naive matrix-vector product. More experimental details are described in Section H of the Supplementary Material.

5.2 Synthetic Data Experiments

The goal of the synthetic data experiments was to answer two questions: (i) Is LUCB-MB more efficient than the existing bandit-based MIPS algorithms if there are several targets similar to the query? (ii) Can LUCB-MB adapt to targets with different variances? Based on the theoretical analysis in Section 4, we expected affirmative answers to both questions. To verify this in practice, we conducted experiments on two synthetic datasets: the *SIMILAR_TARGETS* dataset and the *DIFFERENT_VARIANCES* dataset.

To answer question (i), the SIMILAR_TARGETS dataset was generated using Algorithm 2. Both the query and the targets follow the standard normal distribution, but a small number of targets are very similar to the query while the others are not. This situation is common in practice; in movie recommendation systems, movies of the same genre are likely to receive similar scores from customers and are thus good candidates for recommendation. The experimental results are shown in Figure 1, where speedup is plotted against different δ values. As all bandit-based algorithms achieved an accuracy of 1.0 for all δ values, their performance should be compared in terms of speedup. LUCB-MB was several dozen times faster than the other algorithms; LUCB-EB was the second best, but the performance gap between them is evident. The only difference between LUCB-EB and LUCB-MB is the type of confidence intervals used: the confidence intervals based on the empirical Bernstein bound versus the proposed adaptively changing confidence intervals. These results demonstrate the effectiveness of the proposed confidence intervals.

To answer question (ii), the DIFFERENT_VARIANCES dataset was generated using Algorithm 3. The query and targets follow normal distributions with different variances, where the variances follow an inverse gamma distribution. This setting is particularly challenging: the variances can be arbitrarily large, and the sub-Gaussian assumption does not hold. The experimental results are shown in Figure 2, where the points below the red dashed line violate the δ-correctness. As seen from the figure, LUCB-MB satisfied the δ-correctness with all δ values, whereas LUCB-H and BanditMIPS failed for most δ values. BoundedME is omitted from the figure because it performed a naive scan of all target values and exhibited no speedup due to the challenging setting. This result demonstrates that algorithms based on the sub-Gaussian assumption cannot guarantee correctness if the assumption does not hold, whereas LUCB-MB can adapt to such data even when the variances are unbounded. Thanks to their adaptivity, both LUCB-MB and LUCB-EB achieved perfect accuracy for all δ values, and the latter is omitted from Figure 2 for visibility. However, since the empirical Bernstein bound assumes boundedness, LUCB-EB cannot guarantee accuracy of at least $1 - \delta$ when rewards are unbounded (see Supplementary

Material H.4 for additional experiments on this point).

5.3 Real-World Data Experiments

Our algorithm was also evaluated on two real-world datasets that are commonly used for MIPS evaluation: the Netflix dataset (Bennett and Lanning, 2007) and the MovieLens dataset (Harper and Konstan, 2015). Figure 3 shows that the five bandit-based algorithms outperformed the index-based methods in terms of the accuracy-speedup tradeoff. Since all bandit-based algorithms achieved perfect accuracy across all δ values, we compared their query times for different d values in Figure 4. BoundedME is omitted from the figure because it ended up scanning all data and was slower than naive scan in wall-clock time. With respect to sample complexity, LUCB-MB clearly outperformed all other bandit-based algorithms on both real-world datasets. In terms of wall-clock time, for small d, naive scan outperformed LUCB-MB because it is more efficient than random sampling for small data. However, for large d, LUCB-MB outperformed the naive scan and all other bandit-based algorithms in terms of wall-clock time due to its lower sample complexity.

6 LIMITATIONS

The first major limitation of our approach is its assumption of high data dimensionality, as is the case with existing bandit-based MIPS algorithms. When the data is low-dimensional, naive scanning is faster than sampling. Experimental results showed that LUCB-MB outperforms both naive scan and other bandit-based methods for high-dimensional data. Improving sample efficiency to broaden the regime in which bandit-based methods outperform naive scanning is an interesting direction for future research.

The second major limitation lies in Assumption 5. This assumption is valid for all distributions that satisfy Assumption 2, including all log-concave continuous distributions. Moreover, Assumption 5 holds asymptotically for empirical distributions drawn from these underlying distributions, as discussed at the end of Section 3. However, it is not yet clear how our approach might be adapted to cases where the underlying distributions are inherently discrete. Extending our results beyond Assumptions 2 or 5 remains an open challenge for further study.

LUCB-MB samples vector elements uniformly at random, whereas importance sampling has been explored in the literature (Baharav et al., 2024). Extending our approach to incorporate importance sampling is also an interesting direction for future research.

7 CONCLUSION

We proposed LUCB-MB, a bandit-based MIPS algorithm that utilizes confidence intervals whose lengths change adaptively to observed samples under the moment boundedness assumption that holds for a broad class of distributions. In the experiments, LUCB-MB outperformed the state-of-the-art bandit-based MIPS algorithms both on synthetic and real-world datasets.

References

Bachrach, Y., Finkelstein, Y., Gilad-Bachrach, R., Katzir, L., Koenigstein, N., Nice, N., and Paquet, U. (2014). Speeding up the xbox recommender system using a euclidean transformation for inner-product spaces. In *Proceedings of the 8th ACM Conference on Recommender Systems*, RecSys '14, page 257 – 264, New York, NY, USA. Association for Computing Machinery.

Baharav, T. Z., Kang, R., Sullivan, C., Tiwari, M., Luxenberg, E., Tse, D., and Pilanci, M. (2024). Adaptive sampling for efficient softmax approximation. In Globerson, A., Mackey, L., Belgrave, D., Fan, A., Paquet, U., Tomczak, J., and Zhang, C., editors, *Advances in Neural Information Processing Systems*, volume 37, pages 117580–117613. Curran Associates, Inc.

Bennett, J. and Lanning, S. (2007). The netflix prize. In *Proceedings of the KDD Cup Workshop 2007*, pages 3–6, New York. ACM.

Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration Inequalities - A Nonasymptotic Theory of Independence*. Oxford University Press.

Chen, T., Fu, C., Wang, K., Ke, X., Gao, Y., Zhou, W., Ni, Y., and Zeng, A. (2025). Maximum inner product is query-scaled nearest neighbor. *Proc. VLDB Endow.*, 18(6):1770–1783.

Dai, X., Yan, X., Ng, K. K. W., Liu, J., and Cheng, J. (2020). Norm-explicit quantization: Improving vector quantization for maximum inner product search. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(01):51–58.

Devroye, L. and Lugosi, G. (2001). *Combinatorial Methods in Density Estimation*. Springer Series in Statistics. Springer, 1st ed. edition.

Even-Dar, E., Mannor, S., and Mansour, Y. (2006). Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *J. Mach. Learn. Res.*, 7:1079–1105.

Gabillon, V., Ghavamzadeh, M., and Lazaric, A. (2012). Best arm identification: a unified approach to fixed budget and fixed confidence. In *Proceedings*

of the 26th International Conference on Neural Information Processing Systems - Volume 2, NIPS'12, page 3212 – 3220, Red Hook, NY, USA. Curran Associates Inc.

Geist, M. (2015). A multiplicative UCB strategy for Gamma rewards. In *European Workshop on Reinforcement Learning*, Lille, France.

Harper, F. M. and Konstan, J. A. (2015). The movielens datasets: History and context. *ACM Trans. Interact. Intell. Syst.*, 5(4).

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):pp. 13–30.

Huang, Q., Ma, G., Feng, J., Fang, Q., and Tung, A. K. H. (2018). Accurate and fast asymmetric locality-sensitive hashing scheme for maximum inner product search. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '18, page 1561 – 1570, New York, NY, USA. Association for Computing Machinery.

Jamieson, K., Malloy, M., Nowak, R., and Bubeck, S. (2014). lil' UCB : An optimal exploration algorithm for multi-armed bandits. In Balcan, M. F., Feldman, V., and Szepesvári, C., editors, *Proceedings of The 27th Conference on Learning Theory*, volume 35 of *Proceedings of Machine Learning Research*, pages 423–439, Barcelona, Spain. PMLR.

Jamieson, K. G. and Nowak, R. D. (2014). Best-arm identification algorithms for multi-armed bandits in the fixed confidence setting. In *48th Annual Conference on Information Sciences and Systems, CISS 2014, Princeton, NJ, USA, March 19-21, 2014*, pages 1–6. IEEE.

Jourdan, M., Degenne, R., and Kaufmann, E. (2023). Dealing with unknown variances in best-arm identification. In Agrawal, S. and Orabona, F., editors, *Proceedings of The 34th International Conference on Algorithmic Learning Theory*, volume 201 of *Proceedings of Machine Learning Research*, pages 776–849. PMLR.

Kalyanakrishnan, S., Tewari, A., Auer, P., and Stone, P. (2012). Pac subset selection in stochastic multi-armed bandits. In *Proceedings of the 29th International Coference on International Conference on Machine Learning*, ICML'12, page 227 – 234, Madison, WI, USA. Omnipress.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., and Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA. Curran Associates Inc.

Liu, R., Wu, T., and Mozafari, B. (2019). A bandit approach to maximum inner product search. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 4376–4383. AAAI Press.

Lu, P., Tao, C., and Zhang, X. (2021). Variance-dependent best arm identification. In de Campos, C. P., Maathuis, M. H., and Quaeghebeur, E., editors, *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence, UAI 2021, Virtual Event, 27-30 July 2021*, volume 161 of *Proceedings of Machine Learning Research*, pages 1120–1129. AUAI Press.

Malkov, Y. A. and Yashunin, D. A. (2018). Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs. *IEEE transactions on pattern analysis and machine intelligence*, 42(4):824–836.

Maurer, A. and Pontil, M. (2009). Empirical bernstein bounds and sample-variance penalization. In *COLT 2009 - The 22nd Conference on Learning Theory, Montreal, Quebec, Canada, June 18-21, 2009*.

Morozov, S. and Babenko, A. (2018). Non-metric similarity graphs for maximum inner product search. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, page 4726 – 4735, Red Hook, NY, USA. Curran Associates Inc.

Schudy, W. and Sviridenko, M. (2012). Concentration and moment inequalities for polynomials of independent random variables. In Rabani, Y., editor, *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2012, Kyoto, Japan, January 17-19, 2012*, pages 437–446. SIAM.

Shrivastava, A. and Li, P. (2014). Asymmetric lsh (alsh) for sublinear time maximum inner product search (mips). In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'14, page 2321 – 2329, Cambridge, MA, USA. MIT Press.

Shrivastava, A. and Li, P. (2015). Improved asymmetric locality sensitive hashing (alsh) for maximum inner product search (mips). In *Proceedings of the Thirty-First Conference on Uncertainty in Artifi-*

cial Intelligence, UAI'15, page 812 – 821, Arlington, Virginia, USA. AUAI Press.

Song, Y., Gu, Y., Zhang, R., and Yu, G. (2021). Promips: Efficient high-dimensional c-approximate maximum inner product search with a lightweight index. In *2021 IEEE 37th International Conference on Data Engineering (ICDE)*, pages 1619–1630.

Sun, P., Simcha, D., Dopson, D., Guo, R., and Kumar, S. (2023). Soar: improved indexing for approximate nearest neighbor search. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, Red Hook, NY, USA. Curran Associates Inc.

Tan, S., Xu, Z., Zhao, W., Fei, H., Zhou, Z., and Li, P. (2021). Norm adjusted proximity graph for fast inner product retrieval. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, KDD '21, page 1552 – 1560, New York, NY, USA. Association for Computing Machinery.

Tiwari, M., Kang, R., Lee, J., Lee, D., Piech, C., Thrun, S., Shomorony, I., and Zhang, M. J. (2024). Faster maximum inner product search in high dimensions. In *ICML 2024*.

Vapnik, V. and Chervonenkis, A. (1974). *Theory of Pattern Recognition [in Russian]*. Nauka, Moscow. (German Translation: W. Wapnik & A. Tscherwonenkis, *Theorie der Zeichenerkennung*, Akademie–Verlag, Berlin, 1979).

Zhao, X., Zheng, B., Yi, X., Luan, X., Xie, C., Zhou, X., and Jensen, C. S. (2023). Fargo: Fast maximum inner product search via global multi-probing. *Proc. VLDB Endow.*, 16(5):1100 – 1112.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes] Sections 3 and 4.
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes] Theorem 9.
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [No]

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes] Assumptions 2 and 5.
   (b) Complete proofs of all theoretical results. [Yes] The Supplementary Material.
   (c) Clear explanations of any assumptions. [Yes] Section 3.

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [No]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Not Applicable]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [No] Error bars are omitted from the graphs for visibility. Instead, we reported our algorithm consistently outperformed other algorithms for dozens of different δ and d values.
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Yes] The Supplementary Material.

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Yes] It is stated that we used the code released by the authors of BanditMIPS in the experimental section.
   (b) The license information of the assets, if applicable. [Yes] The license information for the BanditMIPS implementation is described on its site (MIT License).
   (c) New assets either in the supplemental material or as a URL, if applicable. [No]
   (d) Information about consent from data providers/curators. [Not Applicable]
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [Not Applicable]
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]

(c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

Supplementary Material for "Bandit-based Maximum Inner Product Search with Data-Dependent Confidence Intervals"

A THE MOMENT BOUNDEDNESS OF THE NONCENTRAL CHI-SQUARED DISTRIBUTION

In this section, we prove the moment boundedness of the noncentral chi-squared distribution, which we mentioned in Section 3 of the main text.

Lemma 10. *If a random variable X follows a noncentral chi-squared distribution, X is $2\mathbb{E}[X]$-moment bounded.*

Proof. If a random variable X follows a noncentral chi-squared distribution with non-centrality parameter $\lambda \geq 0$ and the degrees of freedom $k \geq 1$, we have $\mathbb{E}[X] = k + \lambda$ and for $p \geq 2$,

$$\mathbb{E}[X^p] = 2^{p-1}(p-1)!(k+p\lambda) + \sum_{j=1}^{p-1} \frac{(p-1)!2^{j-1}}{(p-j)!}(k+j\lambda)\mathbb{E}[X^{p-j}].$$

The goal is to prove that

$$\mathbb{E}[X^p] \leq 2^p \mathbb{E}[X]^p p! = 2^p (k+\lambda)^p p! \tag{3}$$

holds for any $p \geq 1$.

We prove it by induction. For $p = 1$, it is clear that (3) holds. Next, we assume that (3) is satisfied for all $i \leq p-1$, i.e.,

$$\mathbb{E}[X^i] \leq 2^i (k+\lambda)^i i!$$

holds for all $i \leq p-1$. Under this assumption, we have

$$\begin{aligned}
\mathbb{E}[X^p] &= 2^{p-1}(p-1)!(k+p\lambda) + \sum_{j=1}^{p-1} \frac{(p-1)!2^{j-1}}{(p-j)!}(k+j\lambda)\mathbb{E}[X^{p-j}] \\
&\leq 2^{p-1}(p-1)!(k+p\lambda) + \sum_{j=1}^{p-1} \frac{(p-1)!2^{j-1}}{(p-j)!}(k+j\lambda)2^{p-j}(k+\lambda)^{p-j}(p-j)! && \text{(by the assumption)} \\
&= 2^{p-1}(p-1)!(k+p\lambda) + (p-1)!2^{p-1}\sum_{j=1}^{p-1}(k+j\lambda)(k+\lambda)^{p-j} \\
&\leq 2^{p-1}(p-1)!(k+\lambda)^p + (p-1)!2^{p-1}\sum_{j=1}^{p-1}(k+\lambda)^j(k+\lambda)^{p-j} && \text{(because } k \geq 1 \text{ and } \lambda \geq 0) \\
&= 2^{p-1}(p-1)!(k+\lambda)^p + (p-1)!2^{p-1}\sum_{j=1}^{p-1}(k+\lambda)^p \\
&= 2^{p-1}(p-1)!(k+\lambda)^p + (p-1)!2^{p-1}(p-1)(k+\lambda)^p \\
&= 2^{p-1}(k+\lambda)^p p! \left(\frac{2^{p-1}(p-1)!(k+\lambda)^p}{2^{p-1}(k+\lambda)^p p!} + \frac{(p-1)!2^{p-1}(p-1)(k+\lambda)^p}{2^{p-1}(k+\lambda)^p p!} \right) \\
&= 2^{p-1}(k+\lambda)^p p! \left(\frac{1}{p} + \frac{p-1}{p} \right) \\
&= 2^{p-1}(k+\lambda)^p p!.
\end{aligned}$$

Thus, by induction, equation (1) has been proven. □

B The PROPERTIES OF MOMENT BOUNDED RANDOM VARIABLES

In this section, we prove Lemmas 3, 4, and 6, which are described in Section 3 of the main text and state properties of moment-bounded random variables that will be used later in the proofs of the main theorems.

First, we show that moment boundedness is maintained after centering.

Lemma 3. *If a random variable X is L-moment bounded, i.e.,*

$$\forall k \in \mathbb{Z}_{\geq 1}, \mathbb{E}[|X|^k] \leq L^k k!,$$

then $X - \mathbb{E}[X]$ is $2L$-moment bounded, i.e.,

$$\forall k \in \mathbb{Z}_{\geq 1}, \mathbb{E}[|X - \mathbb{E}[X]|^k] \leq (2L)^k k!.$$

Proof. For all integer $k \geq 1$, we have

$$\begin{aligned} |X - \mathbb{E}[X]|^k &\leq 2^{k-1}(|X|^k + |\mathbb{E}[X]|^k) && \text{(by the convexity of } f(x) = |x|^k) \\ &\leq 2^{k-1}(|X|^k + \mathbb{E}[|X|^k]) && \text{(by Jensen's inequality)} \end{aligned}$$

Taking expectations on both sides, we obtain

$$\begin{aligned} \mathbb{E}[|Y|^k] &\leq 2^{k-1}(\mathbb{E}[|X|^k + \mathbb{E}[|X|^k]]) \\ &= 2^{k-1}(\mathbb{E}[|X|^k] + \mathbb{E}[|X|^k]) \\ &= 2^k \mathbb{E}[|X|^k] \\ &\leq 2^k L^k k!. && \text{(by the assumption)} \end{aligned}$$

□

Lemma 4. *If a random variable X is $c\mathbb{E}[|X|]$-moment bounded and $Y = X - \mathbb{E}[X]$ is $a\mathbb{E}[|Y|]$-moment bounded, Y is $\sqrt{2}ac\mathbb{E}[|X|]$-moment bounded. If X is nonnegative in addition, Y is $a\sqrt{2c^2 - 1}\mathbb{E}[X]$-moment bounded.*

Proof. In general, we have

$$\begin{aligned} \mathbb{E}[|X - \mathbb{E}[X]|] &\leq \sqrt{\mathbb{E}[(X - \mathbb{E}[X])^2]} && \text{(by Jensen's inequality)} \\ &= \sqrt{\mathbb{E}[X^2] - \mathbb{E}[X]^2} && \text{(by a property of the variance)} \\ &\leq \sqrt{2c^2\mathbb{E}[|X|]^2 - \mathbb{E}[X]^2} && \text{(by the assumption)} \\ &\leq \sqrt{2}c\mathbb{E}[|X|]. \end{aligned}$$

If X is nonnegative, we obtain

$$\begin{aligned} \mathbb{E}[|X - \mathbb{E}[X]|] &\leq \sqrt{2c^2\mathbb{E}[X]^2 - \mathbb{E}[X]^2} \\ &= \sqrt{2c^2 - 1}\mathbb{E}[X]. \end{aligned}$$

□

Next, we will show that any centralized moment bounded variable is sub-gamma both on the right and left tails. A sub-gamma random variable is defined as follows.

Definition 11. *(Boucheron et al., 2013, Section 2.4) A centralized random variable X is called sub-gamma both on the right and left tails with parameters (σ, a) if, for all λ satisfying $|\lambda| < 1/a$, we have*

$$\mathbb{E}[\exp(\lambda X)] \leq \exp\left(\frac{\lambda^2\sigma^2}{2(1 - a|\lambda|)}\right).$$

Lemma 12. *Any centralized L-moment bounded random variable Y is sub-gamma both on the right and left tails with parameters $(\sqrt{2}L, L)$.*

Proof. For all $\lambda \in \mathbb{R}$, we have

$$\begin{aligned}
\mathbb{E}[\exp(\lambda Y)] &= 1 + \sum_{k=2}^{\infty} \lambda^k \frac{\mathbb{E}[Y^k]}{k!} && \text{(because } E[Y]=0\text{)} \\
&\leq 1 + \sum_{k=2}^{\infty} |\lambda|^k \frac{|\mathbb{E}[Y^k]|}{k!} \\
&\leq 1 + \sum_{k=2}^{\infty} |\lambda|^k \frac{\mathbb{E}[|Y^k|]}{k!} && \text{(by Jensen's inequality)} \\
&\leq 1 + \sum_{k=2}^{\infty} |\lambda|^k \frac{\mathbb{E}[|Y|^k]}{k!} \\
&\leq 1 + \sum_{k=2}^{\infty} (|\lambda| L)^k. && \text{(by the moment boundedness assumption)}
\end{aligned}$$

For $|\lambda| < \frac{1}{L}$, we obtain

$$\begin{aligned}
\mathbb{E}[\exp(\lambda Y)] &\leq 1 + \frac{L^2|\lambda|^2}{1 - L|\lambda|} \\
&\leq \exp\left(\frac{L^2|\lambda|^2}{1 - L|\lambda|}\right) \\
&= \exp\left(\frac{(\sqrt{2}L)^2|\lambda|^2}{2(1 - L|\lambda|)}\right).
\end{aligned}$$

□

We will show that $X - \mathbb{E}[X]$ is sub-Gaussian on the left tail for any nonnegative random variable X.

Lemma 13. *(Boucheron et al., 2013, Exercise 2.9) Let X be a nonnegative random variable with finite second moment. For any $\lambda > 0$, we have*

$$\mathbb{E}[\exp(-\lambda(X - \mathbb{E}[X]))] \leq \exp(\lambda^2 \mathbb{E}[X^2]/2)$$

i.e., $X - \mathbb{E}[X]$ is sub-Gaussian on the left tail with parameter $\sqrt{\mathbb{E}[X^2]}$.

Proof.

$$\begin{aligned}
\mathbb{E}[\exp(-\lambda(X - \mathbb{E}[X]))] &= \exp(\lambda \mathbb{E}[X]) \mathbb{E}[\exp(-\lambda X)] \\
&\leq \exp(\lambda \mathbb{E}[X]) \mathbb{E}[(1 - \lambda X + \lambda^2 X^2/2)] && \text{(since } e^x \leq 1 + x + x^2/2 \text{ for any } x \leq 0\text{)} \\
&= \exp(\lambda \mathbb{E}[X])(1 - \lambda \mathbb{E}[X] + \lambda^2 \mathbb{E}[X^2]/2) \\
&\leq \exp(\lambda \mathbb{E}[X]) \exp(-\lambda \mathbb{E}[X] + \lambda^2 \mathbb{E}[X^2]/2) && \text{(since } 1 + x \leq e^x \text{ for any } x\text{)} \\
&= \exp(\lambda^2 \mathbb{E}[X^2]/2).
\end{aligned}$$

□

Lemma 14. *Let X be a nonnegative $c\mathbb{E}[X]$-moment bounded random variable. For any $\lambda > 0$, we have*

$$\mathbb{E}[\exp(-\lambda(X - \mathbb{E}[X]))] \leq \exp(\lambda^2 c^2 \mathbb{E}[X]^2),$$

i.e., $X - \mathbb{E}[X]$ is sub-Gaussian on the left tail with parameter $\sqrt{2}c\mathbb{E}[X]$.

Proof. This is immediate from Lemma 13 and the fact that $\mathbb{E}[X^2] \leq 2c^2\mathbb{E}[X]^2$ follows from the assumption. □

Lemma 6, described in Section 3 of the main text, is an immediate consequence of Lemmas 12 and 14.

C A PROOF FOR THEOREM 7 (THE NONNEGATIVE CASE)

We now present the proof of Theorem 7. Our proof is based on the proof of Lemma 2 in the Γ-UCB paper (Geist, 2015), which only considers the case where the rewards from the i-th arm follow gamma distribution $\Gamma(a_i, b_i)$ for $i \in [n]$ and assumes that a lower bound $a_0 \leq \min_i a_i$ is known; however, it is hard to know such a lower bound in practice. We extend the lemma to the case of moment bounded random variables, which removes the requirement of such a lower bound.

We now begin the proof by considering the special case when $\mathbb{E}[X] = 0$. Since we have assumed X is nonnegative, $\mathbb{E}[X] = 0$ implies that X is almost surely zero, and thus Theorem 7 holds trivially in this case. Therefore, we may safely assume $\mathbb{E}[X] > 0$ in the remainder of the proof.

Under the assumption $\mathbb{E}[X] > 0$, we proceed to bound the cumulant generating functions (CGFs) of $X - \mathbb{E}[X]$ and $\mathbb{E}[X] - X$ in this order. Let $L_X(\lambda)$ denote the CGF of any random variable X. From Assumption 5, $X - \mathbb{E}[X]$ is sub-gamma with parameters $(\sqrt{2}b\mathbb{E}[X], b\mathbb{E}[X])$, which implies

$$\begin{aligned}\forall \lambda \in \left[0, \frac{1}{b\mathbb{E}[X]}\right), \quad L_{X-\mathbb{E}[X]}(\lambda) &= \ln \mathbb{E}[\exp(\lambda(X - \mathbb{E}[X]))] \\ &\leq \frac{(\sqrt{2}b\mathbb{E}[X])^2\lambda^2}{2(1 - b\mathbb{E}[X]\lambda)}.\end{aligned}$$

By replacing λ with $\lambda/\mathbb{E}[X]$, we obtain an upper bound on the right tail:

$$\forall \lambda \in \left[0, \frac{1}{b}\right), \quad L_{X-\mathbb{E}[X]}\left(\frac{\lambda}{\mathbb{E}[X]}\right) \leq \frac{b^2\lambda^2}{1 - b\lambda} =: \psi(\lambda).$$

Similarly, from Assumption 5, $X - \mathbb{E}[X]$ is sub-Gaussian on the left tail with parameter $\sqrt{2}c\mathbb{E}[X]$, which implies

$$\begin{aligned}\forall \lambda > 0, \quad L_{\mathbb{E}[X]-X}(\lambda) &= \ln \mathbb{E}[\exp(\lambda(\mathbb{E}[X] - X))] \\ &\leq \lambda^2 c^2 \mathbb{E}[X]^2.\end{aligned}$$

By replacing λ with $\lambda/\mathbb{E}[X]$, we obtain an upper bound on the left tail:

$$\forall \lambda > 0, \quad L_{\mathbb{E}[X]-X}\left(\frac{\lambda}{\mathbb{E}[X]}\right) \leq c^2\lambda^2 =: \varphi(\lambda).$$

We can now obtain Chernoff bounds. Let us start with the right tail. For any $\lambda \in \left[0, \frac{1}{b}\right)$, we have

$$\begin{aligned}\mathbb{P}\left(\frac{\overline{X} - \mathbb{E}[X]}{\mathbb{E}[X]} > \epsilon\right) &= \mathbb{P}\left(\exp\left(\lambda \sum_{i=1}^{t} \frac{X_i - \mathbb{E}[X]}{\mathbb{E}[X]}\right) > \exp(\lambda t \epsilon)\right) \\ &\leq \exp(-\lambda t\epsilon)\mathbb{E}\left[\exp\left(\lambda \sum_{i=1}^{t} \frac{X_i - \mathbb{E}[X]}{\mathbb{E}[X]}\right)\right] && \text{(by the Markov's inequality)} \\ &= \exp(-\lambda t\epsilon)\left(\mathbb{E}\left[\exp\left(\lambda \frac{X - \mathbb{E}[X]}{\mathbb{E}[X]}\right)\right]\right)^t && \text{(by the i.i.d. assumption)} \qquad (4) \\ &= \exp(-\lambda t\epsilon)\left(\exp\left(L_{X-\mathbb{E}[X]}\left(\frac{\lambda}{\mathbb{E}[X]}\right)\right)\right)^t \\ &= \exp(-\lambda t\epsilon)\exp\left(tL_{X-\mathbb{E}[X]}\left(\frac{\lambda}{\mathbb{E}[X]}\right)\right) \\ &= \exp\left(-t\left(\lambda\epsilon - L_{X-\mathbb{E}[X]}\left(\frac{\lambda}{\mathbb{E}[X]}\right)\right)\right) \\ &\leq \exp\left(-t\left(\lambda\epsilon - \psi(\lambda)\right)\right).\end{aligned}$$

Let $\psi_*(\epsilon)$ denote the Legendre-Fenchel transform of ψ:

$$\psi_*(\epsilon) = \sup_{0 \leq \lambda < 1/b} (\lambda\epsilon - \psi(\lambda)) = \frac{(\sqrt{b + \epsilon} - \sqrt{b})^2}{b}.$$

Using $\psi_*(\epsilon)$, we obtain

$$\mathbb{P}\left(\frac{\overline{X}-\mathbb{E}[X]}{\mathbb{E}[X]}>\epsilon\right)=\mathbb{P}\left(\frac{\overline{X}}{1+\epsilon}>\mathbb{E}[X]\right)$$
$$\leq \exp\left(-t\psi_*(\epsilon)\right).$$

Setting $\delta=\exp(-t\psi_*(\epsilon))$, for $\epsilon>0$, we have the inverse function of ψ_*:

$$\psi_*^{-1}(x)=2b\sqrt{x}+bx.$$

Using this, we obtain the bound on the right tail:

$$\mathbb{P}\left(\frac{\overline{X}}{1+\epsilon}>\mathbb{E}[X]\right)\leq\delta \quad \text{with} \quad \epsilon=\psi_*^{-1}(\frac{1}{t}\ln\frac{1}{\delta})=2b\sqrt{\frac{\ln(1/\delta)}{t}}+b\frac{\ln(1/\delta)}{t}.$$

Let us find the bound on the left tail. Similarly, we have

$$\mathbb{P}\left(\frac{\mathbb{E}[X]-\overline{X}}{\mathbb{E}[X]}>\epsilon\right)\leq\exp\left(-t\varphi_*(\epsilon)\right),$$

where $\varphi_*(\epsilon)$ denotes the Fenchel-Legendre transformation of φ:

$$\varphi_*(\epsilon)=\sup_{\lambda>0}(\lambda\epsilon-\varphi(\lambda))=\frac{\epsilon^2}{4c^2}.$$

On the other hand, we have

$$\mathbb{P}\left(\frac{\mathbb{E}[X]-\overline{X}}{\mathbb{E}[X]}>\epsilon\right)=\mathbb{P}\left(\overline{X}<\mathbb{E}[X](1-\epsilon)\right)$$
$$=\mathbb{P}\left(\overline{X}<\mathbb{E}[X](1-\epsilon)_+\right)$$
$$=\mathbb{P}\left(\frac{\overline{X}}{(1-\epsilon)_+}<\mathbb{E}[X]\right)$$

because $\mathbb{P}\left(\overline{X}<\mathbb{E}[X](1-\epsilon)\right)=0$ holds for $1-\epsilon<0$. Therefore, we have

$$\mathbb{P}\left(\frac{\overline{X}}{(1-\epsilon)_+}<\mathbb{E}[X]\right)\leq\exp\left(-t\varphi_*(\epsilon)\right).$$

Setting $\delta=\exp\left(-t\varphi_*(\epsilon)\right)$ and using $\phi_*^{-1}(x)=2c\sqrt{x}$ for $x\geq 0$, we obtain the bound on the left tail:

$$\mathbb{P}\left(\frac{\overline{X}}{(1-\epsilon)_+}<\mathbb{E}[X]\right)\leq\delta \quad \text{with} \quad \epsilon=\phi_*^{-1}\left(\frac{\ln(1/\delta)}{t}\right)=2c\sqrt{\frac{\ln(1/\delta)}{t}}.$$

D A PROOF FOR THEOREM 8 (THE GENERAL CASE)

We prove Theorem 8 (the general case) using Theorem 7 (the nonnegative case). The proof proceeds as follows. In Step 1, the upper deviation probability of $\mathbb{E}[|X|]/\overline{|X|}$ is bounded by $\delta/2$ using the theorem for the nonnegative case. In Step 2, the upper deviation probability of $(\overline{X}-\mathbb{E}[X])/\mathbb{E}[|X|]$ is bounded by $\delta/2$ using the sub-gamma property. In Step 3, combining the results of the previous two steps, the upper deviation probability of $(\overline{X}-\mathbb{E}[X])/\overline{|X|}$ is bounded by δ.

Step 1. Since $|X|$ satisfies Assumption 2, it also satisfies Assumption 5. By applying Theorem 7 to $|X|$, we have

$$\mathbb{P}\left(\frac{\overline{|X|}}{(1-b_U(t,\delta/2))_+}<\mathbb{E}[|X|]\right)\leq\frac{\delta}{2}. \tag{5}$$

Step 2. Let $Y_1, \ldots, Y_t$ be i.i.d. sub-gamma random variables with parameters (σ, a). The CGF of $S = Y_1 + \cdots + Y_t$ satisfies

$$\begin{aligned} Ł_S(\lambda) &\coloneqq \ln \mathbb{E} \exp(\lambda S) = t \ln \mathbb{E} \exp(\lambda Y_1) && \text{(by the i.i.d. assumption)} \\ &\leq t \frac{\lambda^2 \sigma^2}{2(1 - a\lambda)}, && \text{(by the sub-gamma property)} \end{aligned} \tag{6}$$

which implies S is $(\sqrt{t}\sigma, a)$-sub-gamma. From the sub-gamma property, we have

$$\mathbb{P}\Big(S \leq \sqrt{2t\sigma^2 s} + as\Big) \leq e^{-s}$$

for any $s > 0$. Setting $\delta/2 = e^{-s}$, we obtain $\ln(2/\delta) = s$ and

$$\mathbb{P}\Big(S_n > \sqrt{2t\sigma^2 \ln(2/\delta)} + a \ln(2/\delta)\Big) \leq \frac{\delta}{2}.$$

From Assumption 2, $X - \mathbb{E}[X]$ is $b\mathbb{E}[|X|]$-moment bounded. By applying Lemma 12, $X - \mathbb{E}[X]$ is sub-gamma with parameters $(\sqrt{2}b\mathbb{E}[|X|], b\mathbb{E}[|X|])$. By setting $S = t(\overline{X} - \mathbb{E}[X])$ and $(\sigma, a) = (\sqrt{2}b\mathbb{E}[|X|], b\mathbb{E}[|X|])$, we obtain

$$\begin{aligned} &\mathbb{P}\left(\overline{X} - \mathbb{E}[X] > \frac{2b\sqrt{t \ln(2/\delta)} + b \ln(2/\delta)}{t} \mathbb{E}[|X|]\right) \\ &= \mathbb{P}\big(\overline{X} - \mathbb{E}[X] > b_L(t, \delta/2)\mathbb{E}[|X|]\big) \leq \frac{\delta}{2}. \end{aligned} \tag{7}$$

Step 3. The union bound of (5) and (7) gives the bound on the right tail:

$$\mathbb{P}\left(\overline{X} - \mathbb{E}[X] > \frac{b_L(t, \delta/2)}{(1 - b_U(t, \delta/2))_+} \overline{|X|}\right) \leq \delta.$$

By symmetry, the same bound holds on the left tail.

E A PROOF FOR THEOREM 9 (PERFORMANCE ANALYSIS)

In this section, we prove Theorem 9, which establishes the correctness and sample complexity of our MIPS algorithm. The proof is based on the analysis of LUCB for BAI under the sub-Gaussian assumption in Jamieson and Nowak (2014), which assumes symmetric, additive confidence bounds. We extend it to asymmetric, multiplicative confidence bounds.

The proof proceeds as follows. First, let S denote the event where the expectations of all k random variables in K lie within their respective confidence intervals. Next, we show $\mathbb{P}(S) \geq 1 - \delta$; that is, with probability at least $1 - \delta$, event S occurs, and the algorithm returns a correct result, thereby establishing the δ-correctness. We then show that, under event S, the algorithm terminates once the sample size of each arm exceeds a certain threshold. The sum of these thresholds over all arms gives the worst-case sample complexity.

E.1 δ-Correctness

First, we prove the δ-correctness. Event S is defined as

$$S \coloneqq \bigcap_{X \in K} \bigcap_{t=1}^{\infty} \{\{l_N(X, t) \leq \mathbb{E}[X] \leq u_N(X, t)\}\},$$

where K is the set of k random variables for which confidence intervals are estimated; specifically, $K = \{V_i^2\}_{i=1}^n \cup \{(Q - V_i)^2\}_{i=1}^n$ and $k = 2n$. The event S occurs if the expectations of all k random variables in K lie within their respective confidence intervals.

We now show $\mathbb{P}(S) \geq 1-\delta$. To simplify the notation, we let $T_X(t) = D_T[X].\tau$ and $\hat{\mu}_{X,T_X(t)} = D_T[X].\hat{\mu}$ denote the sample size and the sample mean for a random variable X at time t, respectively. For a nonnegative $X \in K$, we have

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{t=1}^{\infty}\{\mathbb{E}[X] < l_N(X,t)\}\right) &= \mathbb{P}\left(\bigcup_{t=1}^{\infty}\left\{\mathbb{E}[X] < \frac{D_T[X].\hat{\mu}}{1+b_L(D_T[X].\tau, \tilde{\delta}(D_T[X].\tau))}\right\}\right) \\
&= \mathbb{P}\left(\bigcup_{t=1}^{\infty}\left\{\mathbb{E}[X] < \frac{\hat{\mu}_{X,T_X(t)}}{1+b_L(T_X(t), \tilde{\delta}(T_X(t)))}\right\}\right) \\
&\leq \sum_{t=1}^{\infty}\mathbb{P}\left(\left\{\mathbb{E}[X] < \frac{\hat{\mu}_{X,T_X(t)}}{1+b_L(T_X(t), \tilde{\delta}(T_X(t)))}\right\}\right) \\
&\leq \sum_{\tau=1}^{\infty}\mathbb{P}\left(\left\{\mathbb{E}[X] < \frac{\hat{\mu}_{X,\tau}}{1+b_L(\tau, \tilde{\delta}(\tau))}\right\}\right) \\
&\leq \sum_{\tau=1}^{\infty}\tilde{\delta}(\tau) \qquad (8) \\
&= \sum_{\tau=1}^{\infty}\frac{\delta}{4k\tau^2} = \frac{\delta\pi^2}{24k} \leq \frac{\delta}{2k}. \qquad (9)
\end{aligned}$$

The inequality (8) is based on Theorem 7. Although the sample mean $\hat{\mu}_{X,\tau}$ is computed from samples drawn without replacement, the same bound holds as in the case of sampling with replacement (see Section F of this supplementary material).

Because the probability bound in (9) holds for the right and left tails of all k random variables in K, the union bound yields $\mathbb{P}(S^{\complement}) \leq \delta$, implying $\mathbb{P}(S) \geq 1-\delta$. Under S, $u_{it} \leq \omega_i \leq l_{it}$ holds for any i and t, and the algorithm never return wrong targets. This completes the proof of the δ-correctness.

E.2 Sample Complexity

We analyze the sample complexity under S. For simplicity, we analyze the worst-case sample complexity with $\epsilon = 0$, which corresponds to the strictest stopping condition.

Let $TERM^t$ denote the event that the stopping condition is satisfied at t:

$$TERM^t \coloneqq l_{h_t,t} > u_{j_t,t}.$$

We let $q = (\omega_1 + \omega_2)/2$. We say that target i_* is BAD if $l_{i_*,t} < q$ and that target $i \neq i_*$ is BAD if $u_{i,t} > q$. We claim that for all $t \geq 1$,

$$\neg TERM^t \implies (h_t \text{ is BAD}) \vee (j_t \text{ is BAD}).$$

By taking the contrapositive, we have

$$\neg(h_t \text{ is BAD}) \wedge \neg(j_t \text{ is BAD}) \implies TERM^t,$$

which implies the algorithm stops when h_t is not BAD and j_t is not BAD.

we prove this claim by contradiction. We assume the negation

$$\neg TERM^t \wedge \neg(h_t \text{ is BAD}) \wedge \neg(j_t \text{ is BAD})$$

is true. There are three cases, and every case leads to a contradiction.

$\boldsymbol{(h_t \neq i_*) \wedge (j_t \neq i_*)}$ The assumption says $(u_{h_t,t} \leq q) \wedge (u_{j_t,t} \leq q)$, but then $\omega_1 > q$ is higher than h_t.

$\boldsymbol{(h_t \neq i_*) \wedge (j_t = i_*)}$ The assumption says $(u_{h_t,t} \leq q) \wedge (l_{j_t,t} > q)$, but then $\hat{\mu}_{j_t} > \hat{\mu}_{h_t}$.

$\boldsymbol{(h_t = i_*) \wedge (j_t \neq i_*)}$ The assumption says $(l_{h_t,t} > q) \wedge (u_{j_t,t} \leq q)$, but then $TERM^t$ is true.

By contradiction, the algorithm stops when h_t is not BAD and j_t is not BAD.

Next, we show that, under event S, both h_t and j_t will eventually cease to be BAD after some finite time. Let $\tau_\dagger$ denote the least integer that satisfies $\rho_u(\tau_\dagger) > 0$. The threshold $\tau_\dagger$ represents the minimum sample size required for the upper end of a confidence interval to be finite, whereas $\rho_l(\tau) > 0$ holds for all $\tau \geq 1$. Since $\rho_u(\tau)$ is defined as

$$\rho_u(\tau) = \left(1 - 2c\sqrt{\frac{\ln(4k\tau^2/\delta)}{\tau}}\right)_+,$$

$\tau_\dagger$ is the least integer satisfying

$$\frac{1}{4c^2} > \frac{\ln(4k\tau^2/\delta)}{\tau}.$$

Applying Lemma 15, since $k = 2n$, we have

$$\tau_\dagger = O(\ln(n/\delta)).$$

For any $i \neq i_*$, let r_i and s_i denote the least integers that satisfy

$$r_i \geq \tau_\dagger \quad \wedge \quad \left(\frac{\rho_l(r_i)}{\rho_u(r_i)} - 1\right)\mathbb{E}[V_i^2] \leq \frac{\Delta_i}{2} \tag{10}$$

and

$$\left(1 - \frac{\rho_u(s_i)}{\rho_l(s_i)}\right)\mathbb{E}[(Q - V_i)^2] \leq \frac{\Delta_i}{2}, \tag{11}$$

respectively. The condition $r_i \geq \tau_\dagger$ is needed for $\rho_u(r_i) > 0$. For any t that satisfies $T_{V_i^2}(t) \geq r_i$ and $T_{(Q-V_i)^2}(t) \geq s_i$, we obtain $u_{i,t} \leq q$, i.e., i is not BAD, because, letting $r = T_{V_i^2}(t) \geq r_i$ and $s = T_{(Q-V_i)^2}(t) \geq s_i$, we obtain

$$\begin{aligned}
u_{i,t} &= \frac{1}{2}\left(\frac{\hat{\mu}_{V_i^2,r}}{\rho_u(r)} - \frac{\hat{\mu}_{(Q-V_i)^2,s}}{\rho_l(s)}\right) \\
&\leq \frac{1}{2}\left(\frac{\rho_l(r)}{\rho_u(r)}\mathbb{E}[V_i^2] - \frac{\rho_u(s)}{\rho_l(s)}\mathbb{E}[(Q - V_i)^2]\right) \\
&= \frac{1}{2}\left(\frac{\rho_l(r)}{\rho_u(r)}\mathbb{E}[V_i^2] - \frac{\rho_u(s)}{\rho_l(s)}\mathbb{E}[(Q - V_i)^2]\right) + q - \omega_i + \frac{(\omega_i - \omega_1) + (\omega_i - \omega_2)}{2} \\
&\leq \frac{1}{2}\left(\frac{\rho_l(r)}{\rho_u(r)}\mathbb{E}[V_i^2] - \frac{\rho_u(s)}{\rho_l(s)}\mathbb{E}[(Q - V_i)^2]\right) + q - \omega_i + \frac{(\omega_i - \omega_1)}{2} \\
&= q + \frac{1}{2}\left(\frac{\rho_l(r)}{\rho_u(r)} - 1\right)\mathbb{E}[V_i^2] - \frac{1}{2}\left(\frac{\rho_u(s)}{\rho_l(s)} - 1\right)\mathbb{E}[(Q - V_i)^2] + \frac{(\omega_i - \omega_1)}{2} \quad \left(\text{since } \omega_i = \frac{1}{2}\left(\mathbb{E}[V_i^2] - \mathbb{E}[(Q - V_i)^2]\right)\right) \\
&= q + \frac{1}{2}\left(\frac{\rho_l(r)}{\rho_u(r)} - 1\right)\mathbb{E}[V_i^2] + \frac{1}{2}\left(1 - \frac{\rho_u(s)}{\rho_l(s)}\right)\mathbb{E}[(Q - V_i)^2] - \frac{\Delta_i}{2} \quad (\text{since } \omega_i - \omega_1 = \mu_i - \mu_1 = -\Delta_i) \\
&\leq q.
\end{aligned}$$

The first inequality follows from the fact that under S, $X \in K$ satisfies

$$\hat{\mu}_{X,\tau} \leq \rho_l(\tau)\mathbb{E}[X] \quad \text{and} \quad \rho_u(\tau)\mathbb{E}[X] \leq \hat{\mu}_{X,\tau}$$

for all $\tau \geq 1$.

Similarly, for $i_* = 1$, let r_1 and s_1 denote the least integers that satisfy

$$\left(1 - \frac{\rho_u(r_1)}{\rho_l(r_1)}\right)\mathbb{E}[V_1^2] \leq \frac{\Delta_2}{2} \tag{12}$$

and

$$s_1 \geq \tau_\dagger \quad \wedge \quad \left(\frac{\rho_l(s_1)}{\rho_u(s_1)} - 1\right)\mathbb{E}[(Q - V_1)^2] \leq \frac{\Delta_2}{2}, \tag{13}$$

respectively. For any t that satisfies $T_{V_1^2}(t) \geq r_1$ and $T_{V_1^2}(t) \geq s_1$, we obtain $l_{1,t} \geq c$, i.e., i_* is not BAD, because, letting $r = T_{V_1^2}(t) \geq r_1$ and $s = T_{(Q-V_1)^2}(t) \geq s_1$, we obtain

$$\begin{aligned}
l_{1,t} &= \frac{1}{2}\left(\frac{\hat{\mu}_{V_1^2,r}}{\rho_l(r)} - \frac{\hat{\mu}_{(Q-V_1)^2,s}}{\rho_u(s)}\right) \\
&\geq \frac{1}{2}\left(\frac{\rho_u(r)}{\rho_l(r)}\mathbb{E}[V_1^2] - \frac{\rho_l(s)}{\rho_u(s)}\mathbb{E}[(Q - V_1)^2]\right) \\
&= \frac{1}{2}\left(\frac{\rho_u(r)}{\rho_l(r)}\mathbb{E}[V_1^2] - \frac{\rho_l(s)}{\rho_u(s)}\mathbb{E}[(Q - V_1)^2]\right) + q - \omega_1 + \frac{(\omega_1 - \omega_2)}{2} \\
&= q + \frac{1}{2}\left(\frac{\rho_u(r)}{\rho_l(r)} - 1\right)\mathbb{E}[V_1^2] - \frac{1}{2}\left(\frac{\rho_l(s)}{\rho_u(s)} - 1\right)\mathbb{E}[(Q - V_1)^2] + \frac{\Delta_2}{2} \\
&= q - \frac{1}{2}\left(1 - \frac{\rho_u(r)}{\rho_l(r)}\right)\mathbb{E}[V_1^2] - \frac{1}{2}\left(\frac{\rho_l(s)}{\rho_u(s)} - 1\right)\mathbb{E}[(Q - V_1)^2] + \frac{\Delta_2}{2} \\
&\geq q.
\end{aligned}$$

Therefore, under S, the total number of rounds does not exceed

$$\begin{aligned}
&\sum_{t=1}^{\infty} \mathbf{1}\{h_t \text{ is BAD or } j_t \text{ is BAD}\} \\
&= \sum_{t=1}^{\infty}\sum_{i=1}^{n} \mathbf{1}\{\{h_t = i \text{ or } j_t = i\} \cap \{i \text{ is BAD}\}\} \\
&\leq \sum_{t=1}^{\infty}\sum_{i=1}^{n} \mathbf{1}\{\{h_t = i \text{ or } j_t = i\} \cap \{\{T_{V_i^2}(t) \leq r_i\} \cup \{T_{(Q-V_i)^2}(t) \leq s_i\}\} \\
&\leq \sum_{i=1}^{n} \max\{r_i, s_i\},
\end{aligned}$$

where the last inequality follows from the fact that if $h_t = i$ or $j_t = i$, $T_X(t+1) = T_X(t) + 1$ and this can only occur r_i and s_i times for $X = V_i^2$ and $X = (Q_i - V_i)^2$, respectively.

We now estimate r_i and s_i, which are the least integers that satisfies (10), (11), (12) and (13). To analyze these conditions in a generalized form, we consider the following two questions: given a threshold $\theta > 0$,

(i) under what conditions does an integer τ satisfy $\frac{\rho_l(\tau)}{\rho_u(\tau)} - 1 \leq \theta$? and

(ii) under what conditions does an integer τ satisfy $1 - \frac{\rho_u(\tau)}{\rho_l(\tau)} \leq \theta$?

We start with question (i). Let us review that $\rho_l(\tau)$ and $\rho_u(\tau)$ are defined as

$$\begin{aligned}
\rho_l(\tau) &= 1 + 2b\sqrt{\frac{\ln(4k\tau^2/\delta)}{\tau}} + b\frac{\ln(4k\tau^2/\delta)}{\tau} \\
\rho_u(\tau) &= \left(1 - 2c\sqrt{\frac{\ln(4k\tau^2/\delta)}{\tau}}\right)_+.
\end{aligned}$$

For simplicity, let $f(\tau) = \sqrt{\frac{\ln(4k\tau^2/\delta)}{\tau}}$. If τ satisfies

$$f(\tau) < \frac{1}{2c}, \tag{14}$$

we can omit + and obtain

$$\rho_u(\tau) = 1 - 2cf(\tau) > 0.$$

Under (14), we obtain $f(\tau) \leq \frac{1}{2c} \leq \frac{1}{2}$ and then $f(\tau) \geq f(\tau)^2$, which implies

$$\rho_l(\tau) = 1 + 2bf(\tau) + bf(\tau)^2 \leq 1 + 3bf(\tau).$$

Therefore, $\frac{\rho_l(\tau)}{\rho_u(\tau)} - 1 \leq \theta$ holds if (14) and

$$\frac{\rho_l(\tau)}{\rho_u(\tau)} - 1 \leq \frac{1 + 3bf(\tau)}{1 - 2cf(\tau)} - 1 = \frac{(3b + 2c)f(\tau)}{1 - 2cf(\tau)} \leq \theta$$

hold. Solving this for $f(\tau)$ gives

$$f(\tau) \leq \frac{\theta}{3b + 2c + 2c\theta}.$$

Substituting $f(\tau)$ into this, we get

$$\frac{\ln(4k\tau^2/\delta)}{\tau} \leq \frac{\theta^2}{(3b + 2c + 2c\theta)^2}.$$

Applying Lemma 15, this inequality is satisfied for any

$$\tau > \frac{(3b + 2c + 2c\theta)^2}{\theta^2} \frac{e}{e - 2} \ln\left(\frac{4k}{\delta} \frac{(3b + 2c + 2c\theta)^4}{\theta^4}\right).$$

Expressed in Big-O notation, since k is $2n$, this becomes

$$\tau = O\left(\frac{1}{\theta^2} \ln\left(\frac{n}{\delta\theta}\right)\right) \quad (\theta \to 0). \tag{15}$$

Let us move on to question (ii). Similarly, for a threshold $\theta > 0$, $1 - \frac{\rho_u(\tau)}{\rho_l(\tau)} \leq \theta$ holds if

$$1 - \frac{\rho_u(\tau)}{\rho_l(\tau)} \leq 1 - \frac{1 - 2cf(\tau)}{1 + 3bf(\tau)} \leq \theta$$

holds. Solving this for $f(\tau)$ gives

$$f(\tau) \leq \frac{\theta}{2c + 3b(1 - \theta)}.$$

Substituting $f(\tau)$ into this, we have

$$\frac{\ln(4k\tau^2/\delta)}{\tau} \leq \frac{\theta^2}{(2c + 3b(1 - \theta))^2}$$

Applying Lemma 15, this inequality is satisfied for any

$$\tau > \frac{((2c + 3b(1 - \theta))^2}{\theta^2} \frac{e}{e - 2} \ln\left(\frac{4k}{\delta} \frac{(2c + 3b(1 - \theta))^4}{\theta^4}\right).$$

Expressed in Big-O notation, since $k = 2n$, this becomes

$$\tau = O\left(\frac{1}{\theta^2} \ln\left(\frac{n}{\delta\theta}\right)\right) \quad (\theta \to 0).$$

Thus, the answers to the two questions are expressed in the same order.

Now that we have answered the two generalized questions, we return to conditions (10), (11), (12) and (13). For any $i \in [n]$, from (10) and (12), if $\mathbb{E}[V_i^2] = 0$, $r_i = O(\ln(n/\delta))$, which is the order of $\tau_\dagger$. If $\mathbb{E}[V_i^2] > 0$, by dividing both side of the second conjunct in (10) and replacing θ in (15) with $\frac{\Delta_i}{2\mathbb{E}[V_i^2]}$, we obtain

$$r_i = O\left(\left(\frac{\mathbb{E}[V_i^2]}{\Delta_i}\right)^2 \ln\left(\frac{n\mathbb{E}[V_i^2]}{\delta\Delta_i}\right)\right),$$

where $\Delta_1 = \Delta_2$. Similarly, from (11) and (13), we obtain

$$s_i = O\left(\left(\frac{\mathbb{E}[(Q - V_i)^2]}{\Delta_i}\right)^2 \ln\left(\frac{n\mathbb{E}[(Q - V_i)^2]}{\delta\Delta_i}\right)\right).$$

In conclusion, the worst-case sample complexity of LUCB-MB is

$$\sum_{i=1}^{n} \max\{r_i, s_i\} = O\left(\sum_{i=1}^{n}\left(\frac{m_i}{\Delta_i}\right)^2 \ln\left(\frac{nm_i}{\delta\Delta_i}\right)\right),$$

where $m_i \coloneqq \max\{\mathbb{E}[V_i^2], \mathbb{E}[(Q - V_i)^2]\}$.

Finally, we prove the technical lemma used in the above arguments.

Lemma 15. *If $\theta > 0$, $c > 0$, and $n > 0$ satisfy*

$$n > \frac{1}{\theta}\frac{e}{e-2}\ln(c/\theta^2),$$

where e is the base of the natural logarithm, then

$$\theta > \frac{1}{n}\ln(cn^2).$$

Proof. Write $n = a \ln b$ with $a > 0$ and $b \geq 1$; this is always possible. Then

$$\begin{aligned}\theta - \frac{1}{n}\ln\left(cn^2\right) &= \theta - \frac{1}{a\ln b}\ln\left(ca^2(\ln b)^2\right)\\ &\geq \theta - \frac{1}{a\ln b}\ln\left(ca^2 b^{2/e}\right)\end{aligned}$$

Here, we used the fact that for $b \geq 1$, $\ln b \leq b^{1/e}$. Thus, showing

$$\theta a \ln b - \ln\left(ca^2 b^{2/e}\right) > 0$$

is sufficient. Now set $a = 1/\theta$ (this satisfies $b = e^{n/a} \geq 1$). We obtain

$$\ln b > \frac{e}{e-2}\ln(c/\theta^2).$$

Therefore,

$$n = a\ln b > \frac{1}{\theta}\frac{e}{e-2}\ln(c/\theta^2)$$

is a sufficient condition. □

F MULTIPLE SAMPLING WITHOUT REPLACEMENT

LUCB-MB samples rewards without replacement. In particular, if we estimate $\mathbb{E}[V_i^2]$ using a sample obtained through querying, the estimate can be based on a concatenation of multiple samples drawn without replacement.

In this section, we show that a sample collected through multiple rounds of sampling without replacement can be treated in the same way as a sample collected with replacement.

The following lemma serves as the basis for our discussion.

Lemma 16. *(Hoeffding, 1963, Theorem 4) For a population C, let $X_1, X_2, \ldots, X_n$ denote a random sample without replacement from C and $Y_1, Y_2, \ldots, Y_n$ a random sample with replacement from C. If the function $f(x)$ is continuous and convex then*

$$\mathbb{E} f\left(\sum_{i=1}^{n} X_i\right) \leq \mathbb{E} f\left(\sum_{i=1}^{n} Y_i\right)$$

From the previous lemma, the next naturally follows.

Lemma 17. *Consider k independent random samples without replacement are taken from population C. For j in $[k]$, let $X_{j,1}, X_{j,2}, \ldots, X_{j,n_j}$ denote the j-th random sample. Let Y denote a random variable sampled from C. For*

$$Z = \sum_{j=1}^{k} \sum_{i=1}^{n_j} X_{j,i},$$

its moment generating function satisfies

$$\mathbb{E} \exp(\lambda Z) \leq \mathbb{E}(\exp(\lambda Y))^n,$$

where $n = \sum_{j=1}^{k} n_j$.

Proof. Consider k independent random samples with replacement are taken from C, and the j-th sample has n_j random variables, $W_{j,1}, W_{j,2}, \ldots, W_{j,n_j}$. We have

$$\begin{aligned}
\mathbb{E} \exp(\lambda Z) &= \mathbb{E} \exp\left(\lambda \sum_{j=1}^{k} \sum_{i=1}^{n_j} X_{j,i}\right) \\
&= \mathbb{E} \prod_{j=1}^{k} \exp\left(\lambda \sum_{i=1}^{n_j} X_{j,i}\right) \\
&= \prod_{j=1}^{k} \mathbb{E} \exp\left(\lambda \sum_{i=1}^{n_j} X_{j,i}\right) && \text{(by the independence assumption)} \\
&\leq \prod_{j=1}^{k} \mathbb{E} \exp\left(\lambda \sum_{i=1}^{n_j} W_{j,i}\right) && \text{(by Lemma (16))} \\
&= \prod_{j=1}^{k} \prod_{i=1}^{n_j} \mathbb{E} \exp(\lambda W_{j,i}) && \text{(by the independence assumption)} \\
&= (\mathbb{E} \exp(\lambda Y))^n.
\end{aligned}$$

This implies the lemma. □

We can apply this lemma to the inequalities that are based on the i.i.d. assumption, (4) and (6), and obtain the same results as in the case of the i.i.d. assumption.

G NON-ASYMPTOTIC VALIDITY OF ASSUMPTION 5 UNDER THE EMPIRICAL DISTRIBUTION

In this section, we analyze when empirical distributions satisfy Assumption 5. We consider the setting discussed at the end of Section 3. Specifically, we assume the existence of an underlying continuous distribution that satisfies Assumption 5, and we consider a vector whose elements are sampled from this underlying distribution; that is, the vector constitutes a finite sample. The question is whether it is reasonable to further assume that the empirical distribution constructed from the elements of this vector also satisfies Assumption 5. In this section

only, following statistical conventions, we denote the sample size (i.e., the length of the vector) by n. In contrast, in the main body of the paper, d is used to denote the length of vectors, and n is used to denote the number of target vectors.

The strong law of large numbers states that the sample average converges almost surely to the expected value. This means that if the underlying distribution satisfies Assumption 5, then its empirical distribution (based on the vector elements, in our case) will also satisfy Assumption 5 almost surely as $n \to \infty$.

To provide further evidence, we here provide a non-asymptotic analysis for large but finite n. The key concept is *the preservation of the sub-gamma (or sub-Gaussian) propoerty.* Let X denote a nonnegative random variable drawn from the underlying distribution. Let $S_n = \{X_1, \ldots, X_n\}$ denote an *i.i.d.* sample of size n drawn from the underlying distribution. Let Y denote a nonnegative random variable uniformly drawn from S_n. We say the sub-gamma (sub-Gaussian) property is *preserved* if $X - \mathbb{E}[X]$ is sub-gamma (sub-Gaussian) with parameters proportional to $\mathbb{E}[X]$ and $Y - \mathbb{E}[X]$ is also sub-gamma (sub-Gaussian) with parameters proportional to $\mathbb{E}[X]$.

As long as the sub-gamma (sub-Gaussian) property is preserved for large n with high probability, it is reasonable to assume that Assumption 5 is satisfied by the empirical distribution, P_n. Even though the expected value of Y with respect to P_n, $\mathbb{E}_n[Y] = \overline{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$, may deviate from $\mathbb{E}[X]$, the deviation can be bounded using Theorem 7 in a multiplicative form; that is, for any $\delta \in (0, 1)$, with probability at least $1 - 2\delta$, $\frac{1}{1+b_L(n,\delta)}\mathbb{E}_n[Y] \leq \mathbb{E}[X] \leq \frac{1}{(1-b_U(n,\delta))_+}\mathbb{E}_n[Y]$. Thus, if X satisfies Assumption 5 and the sample size is large, $Y - \mathbb{E}_n[Y]$ is also sub-gamma with parameter proportional to $\mathbb{E}_n[Y]$ with high probability. This indicates that, even though the length of vectors is finite in the MIPS problem, Assumption 5 holds for sampling from the vectors with high probability, if the vectors are long enough.

In the remainder of this section, we will investigate the conditions for the preservation of the sub-gamma or sub-Gaussian properties to hold. In 7.1, we show that, with high probability, the tail probability of an empirical distribution can be upper bounded in terms of the tail probability of its underlying distribution, according to the relative deviation bound of Vapnik and Chervonenkis (1974). In 7.2, building on this fact, we prove that the sub-gamma property is preserved with high probability. In 7.3, we also show that the sub-Gaussian property is preserved with high probability under an additional assumption for the left tail case.

G.1 Tail Bounds for Empirical Distributions

We now discuss tail bounds for empirical distributions from any underlying distributions, according to the Vapnik–Chervonenkis (VC) inequality for relative deviation (Vapnik and Chervonenkis, 1974; Devroye and Lugosi, 2001). It states that for a class $\mathcal{A}$ of subsets of $\mathbb{R}^d$ and all $\varepsilon > 0$,

$$P\left(\sup_{A \in \mathcal{A}} \frac{P_n(A) - P(A)}{\sqrt{P(A)}} > \varepsilon\right) \leq 4S_{\mathcal{A}}(2n)e^{-n\varepsilon^2/4},$$

where P_n is the empirical distribution based on an i.i.d. sample of size n from the underlying distribution P, and $S_{\mathcal{A}}(n)$ is the Vapnik–Chervonenkis shatter coefficient, defined by

$$S_{\mathcal{A}}(n) = \max_{x_1,\ldots,x_n \in \mathbb{R}^d} \left|\{A \cap \{x_1, \ldots, x_n\} : A \in \mathcal{A}\}\right|.$$

Since we are interested in tail probabilities, we focus on the case $\mathcal{A} = \{(x, \infty) : x \in \mathbb{R}\}$. In this case, $S_{\mathcal{A}}(n) = n+1$, and the inequality can be rewritten as

$$P\left(\sup_{x \in \mathbb{R}} \frac{\bar{F}_n(x) - \bar{F}(x)}{\sqrt{\bar{F}(x)}} > \varepsilon\right) \leq 4(2n + 1)e^{-n\varepsilon^2/4},$$

where $\bar{F} = P((x, \infty))$ and $\bar{F}_n = P_n((x, \infty))$ denote the complementary cumulative distribution functions (CCDFs) of the underlying and empirical distributions, respectively.

Letting δ denote the right-hand side, we have

$$P\left(\sup_{x \in \mathbb{R}} \frac{\bar{F}_n(x) - \bar{F}(x)}{\sqrt{\bar{F}(x)}} > \varepsilon(n, \delta)\right) \leq \delta,$$

where

$$\varepsilon(n, \delta) \coloneqq 2\sqrt{\frac{\log(4(2n+1)/\delta)}{n}}.$$

As a result, we obtain the following lemma:

Lemma 18. *Let $\bar{F}_n$ be the CCDF of an empirical distribution of an i.i.d. sample of size n drawn from an underlying distribution with CCDF $\bar{F}$. For any $\delta > 0$, with probability at least $1-\delta$, the tail probability of the empirical distribution satisfies*

$$\begin{aligned}\bar{F}_n(x) &\leq \bar{F}(x) + \varepsilon(n,\delta)\sqrt{\bar{F}(x)} \\ &\leq (1+\varepsilon(n,\delta))\sqrt{\bar{F}(x)}.\end{aligned}$$

From this lemma, we can draw two conclusions. First, as $n \to \infty$, $\bar{F}_n(x)$ converges to $\bar{F}(x)$. Second, even for finite n, $\bar{F}_n(x)$ is upper bounded by a multiple of $\sqrt{\bar{F}(x)}$. For example, if the underlying distribution has an exponential tail bound

$$\bar{F}(t) \leq C\exp(-f(t))$$

for constant $C > 0$ and function f, its empirical distribution also has an exponential tail bound

$$\bar{F}_n(t) \leq (1+\varepsilon(n,\delta))\sqrt{C}\exp\left(-\frac{f(t)}{2}\right)$$

with probability at least $1-\delta$. This result justifies assuming that the empirical distribution has a tail bound similar to that of the underlying distribution. Next, we provide a more rigorous analysis, focusing on sub-gamma distributions.

G.2 The Preservation of the Sub-Gamma Property

By applying the result above, we now prove the preservation of the sub-gamma property; that is, if $X - \mathbb{E}[X]$ is sub-gamma with parameters proportional to $\mathbb{E}[X]$, then $Y - \mathbb{E}[X]$ is also sub-gamma with parameters proportional to $\mathbb{E}[X]$ with high probability when the sample size is large. Specifically, we prove the following lemma:

Lemma 19. *Let X be a centered sub-gamma random variable with parameters $(\sigma, a) = (\sqrt{2}b\mu, b\mu)$ for constants $b \geq 1$ and $\mu > 0$. Let S_n denote an i.i.d. sample of size n drawn from the same distribution as X. Let Y denote a random variable uniformly drawn from S_n. For any $\delta \in (0,1)$, with probability at least $1-\delta$, conditionally on S_n, Y is also sub-gamma with parameters $(8\sqrt{6(1+\varepsilon(n,\delta))}b\mu, 16\sqrt{(1+\varepsilon(n,\delta))}b\mu)$.*

In this lemma, we assume centering for simplicity. By substituting $X = X - \mathbb{E}[X]$ and $\mu = \mathbb{E}[X]$, the assumption of this lemma represents the sub-gamma assumption of Assumption 5, i.e., $X - \mathbb{E}[X]$ is sub-gamma with parameters $(\sqrt{2}b\mathbb{E}[X], b\mathbb{E}[X])$.

Proof. From the properties of sub-gamma distributions with parameters (σ, a), we have

$$\bar{F}(t) = P(X > t) \leq \exp\left(-\frac{v}{a^2}h_1\left(\frac{at}{v}\right)\right),$$

where $v \coloneqq \sigma^2$ and

$$h_1(u) \coloneqq 1 + u - \sqrt{1+2u} \quad \text{for } u > 0.$$

From Lemma 18, we obtain that, with probability at least $1-\delta$, for all $t \geq 0$,

$$\begin{aligned}\bar{F}_n(t) &\leq (1+\varepsilon(n,\delta))\sqrt{\bar{F}(t)} \\ &\leq (1+\varepsilon(n,\delta))\exp\left(-\frac{v}{2a^2}h_1\left(\frac{at}{v}\right)\right) \\ &= \kappa\exp\left(-\frac{V}{C^2}h_1\left(\frac{Ct}{V}\right)\right),\end{aligned}$$

where we denote $(V, C) \coloneqq (2v, 2a)$ and $\kappa \coloneqq 1 + \varepsilon(n, \delta)$ for simplicity. Note that the exponential factor has the form of the right tail bound of a sub-gamma distribution with parameters $(\sqrt{V}, C)$. Thus, by considering the tail bounds on both sides, we obtain

$$\max\{P_n(Y > \sqrt{2Vt} + Ct), P_n(-Y > \sqrt{2Vt} + Ct)\} \leq \kappa e^{-t} \quad \text{for all } t > 0,$$

where Y denotes a random variable drawn from P_n. We will show that, if this bound holds, Y is sub-gamma, according to Theorem 2.3 from Boucheron et al. (2013). By integration by parts,

$$\mathbb{E}_n[Y^{2q}] = 2q \int_0^\infty y^{2q-1} P_n(|Y| > y) dy,$$

where $\mathbb{E}_n$ denotes the expectation with respect to P_n. Setting $y = \sqrt{2Vt} + Ct$,

$$\begin{aligned}
\mathbb{E}_n[Y^{2q}] &\leq 4\kappa q \int_0^\infty \left(\sqrt{2Vt} + Ct\right)^{2q-1} \left(\frac{\sqrt{2Vt} + 2Ct}{2t}\right) \frac{e^{-t}}{t} dt \\
&\leq 2\kappa q \int_0^\infty \left(\sqrt{2Vt} + 2Ct\right)^{2q} \frac{e^{-t}}{t} dt.
\end{aligned}$$

Since $(a+b)^{2q} \leq 2^{2q-1}(a^{2q} + b^{2q})$ holds by the convexity of x^{2q},

$$\begin{aligned}
\mathbb{E}_n[Y^{2q}] &\leq \kappa q 2^{2q} \int_0^\infty \left((2tV)^q + (2Ct)^{2q}\right) e^{-t} dt \\
&\leq \kappa 2^{2q} \left(q! 2^q V^q + \frac{(2q)!}{2} (2C)^{2q}\right) \\
&= q!(\kappa 2^{2q} 2^q V^q) + (2q)! \frac{\kappa 2^{2q} (2C)^{2q}}{2} \\
&= q!(8\kappa^{1/q} V)^q + (2q)! \frac{(4\kappa^{1/(2q)} C)^{2q}}{2} \\
&\leq q!(8\kappa V)^q + (2q)!(4\sqrt{\kappa} C)^{2q}.
\end{aligned}$$

Theorem 2.3 from Boucheron et al. (2013) states that, if for some positive constants A and B,

$$\mathbb{E}[X^{2q}] \leq q! A^q + (2q)! B^{2q},$$

then X is sub-gamma with parameters $(\sqrt{4(A + B^2)}, 2B)$. By substituting $A = 8\kappa V$ and $B = 4\sqrt{\kappa} C$, P_n is sub-gamma. Since $(\sigma, a) = (\sqrt{2} b\mu, b\mu)$, we have $(V, C) = (2v, 2a) = (2\sigma^2, 2a) = (4b^2\mu^2, 2b\mu)$. Thus, the sub-gamma parameters of P_n are

$$\begin{aligned}
(\sqrt{4(A + B^2)}, 2B) &= (\sqrt{4(8\kappa V + 16\kappa C^2)}, 8\sqrt{\kappa} C) \\
&= (4\sqrt{2\kappa V + 4\kappa C^2}, 8\sqrt{\kappa} C) \\
&= (4\sqrt{2\kappa(4b^2\mu^2) + 4\kappa(4b^2\mu^2)}, 16\sqrt{\kappa} b\mu) \\
&= (4\sqrt{24\kappa b^2\mu^2}, 16\sqrt{\kappa} b\mu) \\
&= (8\sqrt{6\kappa} b\mu, 16\sqrt{\kappa} b\mu).
\end{aligned}$$

This proves the lemma. □

G.3 The Preservation of the Sub-Gaussian Property

We now discuss the preservation of the sub-Gaussian property. If X is sub-Gaussian on both tails with parameters proportional to $\mathbb{E}[X]$, the preservation of the sub-Gaussian property can be proved in the same way as the sub-gamma property. The preservation of the sub-Gaussian property only on the left tail can be proved with an additional assumption by the following lemma. It is an open question for future research whether this assumption can be relaxed.

Lemma 20. *Let X be a nonnegative $c\mathbb{E}[X]$-moment bounded random variable. In addition, assume that X^2 is $f\mathbb{E}[X^2]$-moment bounded for some constant f. Let S_n be an i.i.d. sample of size n drawn from the same distribution as X. Let Y denote a nonnegative random variable uniformly drawn from S_n. Then, for any $\delta \in (0,1)$, with probability at least $1-\delta$, conditionally on S_n, $Y - \mathbb{E}_n[Y]$ is sub-Gaussian on the left tail with parameter $c\sqrt{2(1-2f\sqrt{\ln(1/\delta)/n})_+}\mathbb{E}[X]$.*

Proof. Since X^2 is $f\mathbb{E}[X^2]$-moment bounded, from Lemma 3, $X^2 - \mathbb{E}[X^2]$ is $2f\mathbb{E}[X^2]$-moment bounded, and thus satisfies Assumption 2.

Since $\mathbb{E}_n[Y^2] = (X_1^2 + \cdots + X_n^2)/n$, by applying Theorem 7, for any $\delta \in (0,1)$, with probability at least $1-\delta$, $\mathbb{E}_n[Y^2] \leq (1-2f\sqrt{\ln(1/\delta)/n})_+\mathbb{E}[X^2]$. By applying Lemma 13, conditionally on S_n, $Y - \mathbb{E}_n[Y]$ is sub-Gaussian on the left tail with parameter $\sqrt{\mathbb{E}_n[Y^2]}$. Finally, we obtain

$$\begin{aligned}\sqrt{\mathbb{E}_n[Y^2]} &\leq \sqrt{(1-2f\sqrt{\ln(1/\delta)/n})_+\mathbb{E}[X^2]} \\ &\leq \sqrt{(1-2f\sqrt{\ln(1/\delta)/n})_+(2c^2\mathbb{E}[X]^2)} \\ &\leq c\sqrt{2(1-2f\sqrt{\ln(1/\delta)/n})_+}\mathbb{E}[X].\end{aligned}$$

□

H EXPERIMENTAL DETAILS

H.1 Experimental Environment

All experiments were conducted with an Intel Xeon E5-2667v4 (16C/3.20GHz) processor, 64 GB of memory, and Python 3.10.

H.2 Preprocessing for the Real-World Datasets

The two real-world datasets, Netflix and MovieLens, were preprocessed in the manner described in the Bandit-MIPS paper (Tiwari et al., 2024). The movie vectors were used as the query and targets, and d was the number of subsampled customers.

Netflix We used a subset of the Netflix Prize dataset (Bennett and Lanning, 2007) that contained the ratings of $1,350$ movies by $143,458$ customers. Missing ratings were imputed by matrix approximation using singular value decomposition with 100 factors.

MovieLens We used a subset of the MovieLens-1M dataset (Harper and Konstan, 2015) that contained the ratings of $3,706$ movies by $6,040$ customers. Missing values were imputed by matrix approximation using nonnegative matrix factorization with 15 factors.

H.3 Experimental Settings

Method of Measurement The parameters of the algorithms and the number of dimensions were varied over the ranges detailed in the next two paragraphs to observe their effects on accuracy and speedup. The results were averaged over 30 trials with different random seeds.

Parameter Values For the five bandit-based algorithms (BanditMIPS, BoundedME, LUCB-H, LUCB-EB and LUCB-MB), δ was varied over $[10^{-10}, 0.99]$ with ϵ fixed at 0. For Figure 4, δ was set to 0.01. For the datasets with bounded values, the sub-Gaussian parameter σ^2 was set to $\frac{(b-a)^2}{4}$, where $[a,b]$ is the range of inner product values; this is the minimal value for which the theoretical guarantee holds (Boucheron et al., 2013). For the datasets with unbounded rewards, the value of σ^2 was set to the empirical variance of rewards, $\hat{\sigma}^2$. Although BoundedME and LUCB-EB are theoretically inapplicable to unbounded rewards, we normalized rewards using $\hat{\sigma}$ to make rewards approximately bounded, thereby enabling a forced empirical evaluation under this setting.

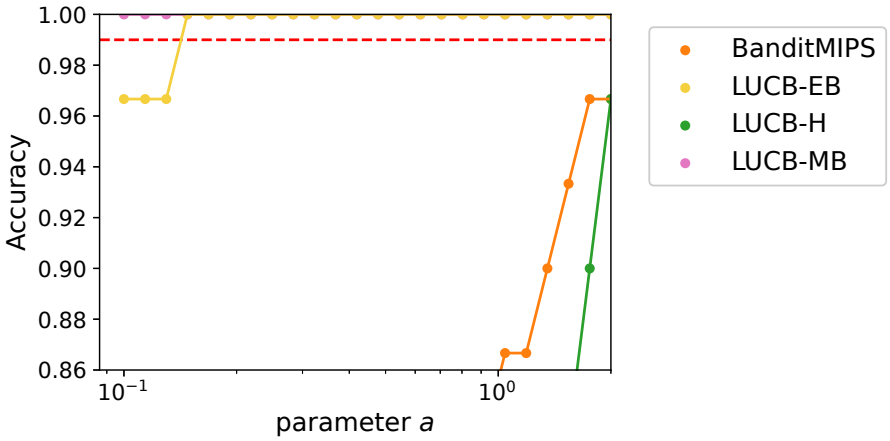

Figure H.1: Accuracy vs. the shape parameter a of the inverse gamma distribution ($\delta = 0.01$). The red dashed line indicates accuracy of $1 - \delta = 0.99$.

For LSH-MIPS, the number of hash functions and hash values were varied over $[1, 10]$. For H2ALSH, we varied δ over $[1/2^4, 1/2]$, c_0 over $[1.2, 5]$, and c over $[0.2, 0.9]$. For PCA-MIPS, the depth of the trees was varied over $[0, 10]$.

Batch Size In our experiments, all bandit-based algorithms were implemented using the BanditMIPS library provided by its original authors, which was designed to take samples in a batch manner to avoid computational overhead caused by one-by-one fetching. We followed their setting and set the same batch size, 99, for all bandit-based algorithms throughout the experiments described in the main text. The batch size was set to 3 for the additional experiment described in Section H.4 to make maintaining high accuracy more challenging.

Settings of n and d We set n and d as follows. For the two synthetic datasets (SIMILAR_TARGETS and DIFFERENT_VARIANCES), we used $n = 1,000$ and $d = 10,000$ (Figures 1 and 2). For the Netflix dataset, we used $n = 1,000$ and $d = 10,000$, and for the MovieLens datasets, we used $n = 1,000$ and $d = 5,000$ (Figure 3). Furthermore, to conduct experiments on larger-scale data, we used the maximum possible values for each dataset: for Netflix, we set $n = 1,349, d = 100,000$; for MovieLens, we set $n = 3,705, d = 6,040$ (Figure 5 in this supplementary material). To see the effect of d, we evaluated the query times on 21 different d values in $\{100, 5000, 10000, \ldots, 100000\}$ for Netflix, and in $\{100, 300, 600, \ldots, 6000\}$ for MovieLens (Figure 4).

H.4 Additional Experimental Results on the DIFFERENT_VARIANCES dataset

In Section 5.2 of the main text, both LUCB-EB and LUCB-MB achieved perfect accuracy for the DIFFERENT_VARIANCES dataset. However, as mentioned in Section 5.2, LUCB-EB cannot guarantee accuracy for unbounded rewards. To empirically demonstrate this limitation, we conducted an additional experiment under a more extreme setting in which queries exhibit unexpectedly large variances. In this setting, queries were sampled from $N(0, \sigma_q^2)$, where $\sigma_q^2 \sim \text{InvGamma}(a, 1)$, and the shape parameter a was varied from 0.1 to 2.0. The batch size was set to 3 in this experiment. The results are shown in Figure H.1. LUCB-EB failed to maintain accuracy at least $1-\delta$ for small values of a because query values had unexpectedly large variances. In contrast, LUCB-MB achieved perfect accuracy across all values of a because it guarantees accuracy regardless of the magnitude of variance, provided that the distribution family satisfies the moment boundedness assumption.

H.5 Additional Experimental Results on the Real-World Datasets

Additional experimental results with the two real-world datasets are shown in Figure H.2 in this supplementary material. Since all bandit-based algorithms achieved perfect accuracy across all δ values in Figure 3 of the main text, Figure H.2 compares their speedups for different δ values. For $\delta \leq 0.01$, LUCB-MB outperformed all other bandit-based algorithms in both sample complexity and wall-clock time on both real-world datasets. Although

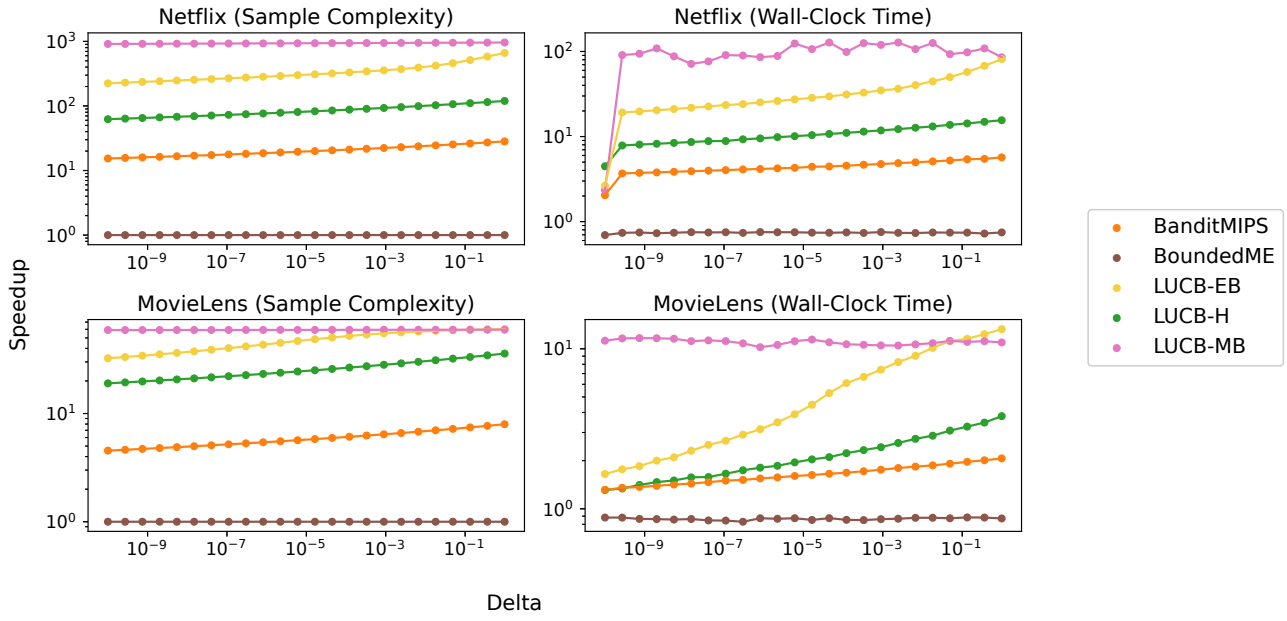

Figure H.2: Speedup vs Delta: Netflix (n=1,349, d=100,000) and MovieLens (n=3,705, d=6,040).

LUCB-EB performed on par with or better than LUCB-MB for $\delta \geq 0.1$, LUCB-MB was consistently better than LUCB-EB for smaller δ.