Correlated bandits or: How to minimize mean-squared error online

Vinay Praneeth Boda [1] **Prashanth L.A.** [2]

Abstract

While the objective in traditional multi-armed bandit problems is to find the arm with the highest mean, in many settings, finding an arm that best captures information about other arms is of interest. This objective, however, requires learning the underlying correlation structure and not just the means of the arms. Sensors placement for industrial surveillance and cellular network monitoring are a few applications, where the underlying correlation structure plays an important role. Motivated by such applications, we formulate the *correlated bandit* problem, where the objective is to find the arm with the lowest mean-squared error (MSE) in estimating all the arms. To this end, we derive first an MSE estimator, based on sample variances and covariances, and show that our estimator exponentially concentrates around the true MSE. Under a best-arm identification framework, we propose a successive rejects type algorithm and provide bounds on the probability of error in identifying the best arm. Using minmax theory, we also derive fundamental performance limits for the correlated bandit problem.

1. Introduction

The traditional multi-armed bandit problem aims to find the arm with the highest payoff. This is often motivated by practical applications such as to identify an ad with highest payoff in showing to users, or identifying a strategy with maximum payoff. In this work, we consider a setting with the objective being the identification of an arm/node which best captures the entire information of a system, i.e., the identification of arm which can best estimate all the other arms. In contrast to the traditional multi-armed bandit problem, this objective involves an estimation of the correlation structure among the various arms. This is motivated by several practical applications. For instance, in internet-of-things, sensors are used to take measurements from multiple locations with the objective of estimating the underlying parameter, e.g., temperature, over a region. Resource constraints mean that it might not possible to place sensors at the desired level of granularity. However, an estimate of the underlying distribution enables one to form an estimate of the parameter at points not measured. This estimate of the statistics of the underlying randomness is often formed using limited measurements from multiple points, before choosing the final location of the sensors. Another application of interest is in identifying members who can best approximate the social network. Instances include sensors used for measuring temperature in a region (Guestrin et al., 2005), thermal sensors on microprocessors (Long et al., 2008), optimizing queries over a sensornet (Deshpande et al., 2004) and placing sensors to detect contaminants in a water distribution network (Krause et al., 2008). Problems of similar interest have also been studied in the realm of information theory in (Boda, 2019; Boda & Narayan, 2018). In all these applications, the underlying correlation structure plays an important role.

In this paper, we formulate a variant of the stochastic K-armed bandit problem, where the objective is to identify the arm that best estimates all the other correlated arms. We measure how good an arm $i \in \{1, \ldots, K\}$ can estimate other arms using the mean-squared error (MSE) criterion:

$$\mathcal{E}_i \triangleq \sum_{j=1}^{K} \mathbb{E}\left[\left(X_j - \mathbb{E}[X_j|X_i]\right)^2\right]. \tag{1}$$

We assume that the arms $X_1, \ldots, X_K$ are correlated sub-Gaussian random variables (r.v.s). (Paul et al., 2014) consider a celluar network application, where the goal is to monitor large communication networks with huge traffic. Since observing every node is computationally intensive, companies such as AT&T use measurements from various nodes to identify a subset which best captures the average behavior of the network. The requirement is for an algorithm that reduces the data acquisition cost by identifying the most-correlated subset of nodes, while using a minimum number of sample measurements. The authors in (Paul et al., 2014) show that a model approximating the

[1]LinkedIn Corp. [2]Department of Computer Science and Engineering, Indian Institute of Technology Madras. Majority of this work was done during the authors time at University of Maryland College Park. Correspondence to: Vinay Praneeth Boda <vpboda@gmail.com>, Prashanth L.A. <prashla@cse.iitm.ac.in>.

Proceedings of the 36^{th} International Conference on Machine Learning, Long Beach, California, PMLR 97, 2019.

underlying nodes as Gaussian r.v.s is useful and reliable.

Closely related problems in other application contexts include (i) selecting a few blogs that capture the information cascade (Leskovec et al., 2007); (ii) finding a subset of people who best represent the average behavior of a community; To put it differently, the notions of *centrality* in the context of document/news summarization (Erkan & Radev, 2004) and *prestige* in social networks (Heidemann et al., 2010) are closely related to the MSE objective in (1). In each of these applications, there is a cost associated with acquiring data and the challenge is to find the most correlated subset of blogs/people/etc using minimal observations about the community.

We study the basic problem of identifying the arm which has the best MSE in estimating the remaining arms in a multi-armed bandit framework. We consider the best arm identification setting (Audibert et al., 2010; Kaufmann et al., 2015), where a bandit algorithm is given a fixed sampling budget, and is evaluated based on the probability of incorrect identification. Challenges encountered for such a setup include:
(i) Any estimate for the MSE requires estimation of the underlying correlations, without assuming knowledge of the variances.
(ii) Estimate of the MSE of an arm i involves estimating the correlation of arm i with the remaining arms. This requires samples from all pairs of arms associated with i. In particular, sampling arm i alone would be insufficient towards estimating arm i's MSE; and hence
(iii) A bandit algorithm needs to optimize sampling across all pairs of arms and not just among arms. This requires intricate decisions over a larger set, in contrast to the classical mean-value optimizing algorithms in a best arm identification framework.

We summarize our contributions below.

First, we introduce a *new formulation* to study the identification of arm which best estimates all arms. We design an estimate and develop the concentration bound for the estimate of mean-squared error formed from available samples. Our estimator builds on the difference estimator introduced in (Liu & Bubeck, 2014), but estimation is technically more challenging in our setting as the underlying variances are not known and unlike (Liu & Bubeck, 2014), not necessarily assumed to be one.

Second, we analyze a nonadaptive uniform sampling strategy (i.e., a strategy that pulls each pair of arms an equal number of times) and propose an algorithm inspired by popular successive rejects (SR) (Audibert et al., 2010) for best-arm identification, but more intricate due to the nonlinearity of the objective function, the MSE objective function (1). A naive SR strategy that operates over phases, discarding all arm pairs associated with the arm having lowest empirical MSE is suboptimal. Instead, our SR algorithm maintains active sets for arms as well as pairs and discards a pair only if both constituent arms are out of the active arms set. We provide an upper bound on the probability of error in identifying the best arm for our SR algorithm and the bound involves a hardness measure that factors in the gaps in MSEs as well as the correlations, which are specific to the correlated bandit problem. As in the classic bandit setup, the upper bound shows that SR algorithm requires fewer samples to find the best arm in comparison to a uniform sampling strategy, especially, when K is large and the underlying gaps (difference between MSE of optimal and suboptimal arms) are uneven.

Third, we prove a lower bound over all bandit problems with a certain hardness measure and to the best of our knowledge, this is the first lower bound for the correlated bandit problem that involves adaptive sampling strategies. The lower bound involves constructing problem transformations, where the optimal arm is "swapped" with one of the sub-optimal ones, resulting in $K-1$ problem instances. Unlike in the classic setup, any local change in the distribution of an arm impacts the MSE of all the other arms. Moreover, pulling pairs of arms instead of individual arms makes the lower bound technically more challenging.

In (Liu & Bubeck, 2014), which is the closest related work, the authors consider a bandit problem, where the objective is to identify a subset of arms most correlated among themselves, i.e., to identify the local correlation structure within a subset of arms themselves. On the other hand, our problem is about forming global inference from samples of subsets of arms to identify the arm that is most correlated to the remaining arms. In (Liu & Bubeck, 2014), the authors consider a setting with positively correlated arms with unit variance, making the estimation task and hence, the overall best arm identification slightly easier. As we show later in Section 3, their estimation scheme does not extend to the more general non-unit variance setup that we consider. Finally, we also prove fundamental limits on the performance of any correlated bandit algorithm, through information-theoretic lower bounds, and to the best of our knowledge, no lower bounds exist for a correlated bandit problem.

The rest of the paper is organized as follows: In Section 2, we formalize the correlated bandit problem. In Section 3, we present the MSE estimation scheme and derive a concentration bound for our estimator. In Section 4, we examine uniform sampling strategy, while in Section 5, we present a successive-rejects type algorithm. In Section 6, we present a lower bound for the correlated bandit problem. We provide sketches of convergence proofs in Section 7. While not the thrust of this work, we provide a few illustrative examples in Section 8 showing the performance of our successive-rejects type algorithm. Finally, in Sec-

tion 9 we provide our concluding remarks.

2. Model

We consider a set $\mathcal{M} \triangleq \{1, \ldots, K\}$ of K correlated arms $X_1, \ldots, X_K$, whose samples are i.i.d. in time. For each arm i, let $\mathcal{E}_i$ denote the minimum mean-squared error (MMSE) of X_i estimating all the remaining arms, i.e.,

$$\mathcal{E}_i \triangleq \min_g \mathbb{E}\big[\big(X_{\mathcal{M}} - g(X_i)\big)^T \big(X_{\mathcal{M}} - g(X_i)\big)\big]. \quad (2)$$

Consider the special case of jointly Gaussian r.v.s $X_1, \ldots, X_K$, whose joint probability distribution is characterized by the mean (taken to be zero for the sake of expository simplicity), and *covariance matrix* $\Sigma \triangleq \mathbb{E}[X_{\mathcal{M}}^T X_{\mathcal{M}}]$:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \ldots & \rho_{1K}\sigma_1\sigma_K \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \ldots & \rho_{2K}\sigma_2\sigma_K \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1K}\sigma_1\sigma_K & \rho_{2K}\sigma_2\sigma_K & \ldots & \sigma_K^2 \end{bmatrix}. \quad (3)$$

In the above, σ_p^2, $p \in \mathcal{M}$ is the variance of arm p and ρ_{ij}, $i, j = 1, \ldots, K,\ i \neq j$, the correlation coefficient between arms i and j.

The best estimate g^*, which achieves the minimum in (2), is known to be the MMSE estimate. For zero-mean jointly Gaussian r.v.s, this is given by (cf. Chapter 3 of (Hajek, 2009))

$$g^*(X_i) = \mathbb{E}[X_{\mathcal{M}}|X_i] = [\mathbb{E}[X_1|X_i] \ldots \mathbb{E}[X_K|X_i]]^T,$$
$$\text{with } \mathbb{E}[X_j|X_i] = \frac{\mathbb{E}[X_j X_i]}{\mathbb{E}[X_i^2]} X_i = \frac{\rho_{ij}\sigma_j}{\sigma_i} X_i. \quad (4)$$

The corresponding MMSE for arm i is

$$\mathcal{E}_i = \sum_{j=1}^{K} \mathbb{E}\left[\big(X_j - \mathbb{E}[X_j|X_i]\big)^2\right] = \sum_{j \neq i} \sigma_j^2(1 - \rho_{ij}^2). \quad (5)$$

Note that there is no error in arm i estimating itself and the error in estimating the jth arm is characterized by the correlation between X_i and X_j and the relevant variances. Further, the MMSE estimate for the case of Gaussian r.v.s is linear. In the more general case of non-Gaussian r.v.s, the MMSE estimate is typically nonlinear and any online computation is typically a computationally intense task. In such cases, we restrict ourselves to employing an optimal linear estimator which is still defined as the right side of (4). Thus, the right-side of (5) holds for all optimal linear estimators, with it being optimal for Gaussian r.v.s.

We consider a setting where the arms $X_1, \ldots, X_K$ are sub-Gaussian, and focus on linear estimators. We recall the definition of sub-Gaussianity below.

Definition 1. *A r.v. X is said to be σ-sub-Gaussian if* $\mathbb{E}\left(e^{\lambda X}\right) \leq \exp\left(\frac{\lambda^2\sigma^2}{2}\right), \quad \forall \lambda \in \mathbb{R}$.

For equivalent characterizations of sub-Gaussianity, the reader is referred to Theorem 2.1 of (Wainwright, 2015).

We consider a fixed budget best-arm identification framework, and the interaction of our (bandit) algorithm with the environment is given below.

Correlated bandit algorithm

Input: set of pairs of arms $\mathcal{S}$, number of rounds n.
For all $t = 1, 2, \ldots, n$**, repeat**

1. Based on samples $\{(X_{i_l,l}, X_{j_l,l}),\ l = 1, \ldots, t-1\}$ seen so far, select a pair $(i_t, j_t) \in \mathcal{S} \triangleq \{(i,j) \mid i, j = 1, \ldots, K, i < j\}$.
2. Observe a sample from the bivariate distribution corresponding to the arms i_t, j_t.

After n rounds, output an arm $\hat{A}_n$.

Notice that, in each round, the algorithm above pulls a pair of arms, and this is necessary to learn the underlying correlation structure.

In our setting, the performance metric associated with each arm i is its MSE $\mathcal{E}_i$, and the optimal arm, say i^*, has the lowest MSE, i.e.,

$$i^* = \arg\min_{i \in \mathcal{M}} \mathcal{E}_i.$$

The objective is to minimize the probability of error in identifying the best arm, i.e., $\mathbb{P}\left(\hat{i}_n \neq i^*\right)$, where $\hat{i}_n$ is the estimate of the best arm based on n samples.

For $i \neq i^*$, the suboptimality of the arm i is quantified by its gap in its MSE with respect to the optimal arm, i.e., $\Delta_i = \mathcal{E}_i - \mathcal{E}_{i^*}$. The notation (i) is used to refer to the i^{th} best arm (with ties broken arbitrarily), i.e., $\Delta_{(i)}$s are ordered gaps of the arms: $\Delta_{(1)} \triangleq \Delta_{(2)} \leq \Delta_{(3)} \leq \ldots \leq \Delta_{(K)}$.

Note that the problem with $K = 2$ reduces to identifying the arm with higher variance and has no dependence on the correlation between the arms. The analysis of this case would be similar (estimate variance instead of mean) to the classical bandit problems and differs considerably from the setting with $K \geq 3$ arms, which is the setting assumed hereafter.

3. MSE Estimation

Let $\{(X_{it}, X_{jt}),\ t = 1, \ldots, n\}$ denote the set of n i.i.d. samples obtained from the bivariate Gaussian distribution corresponding to the pair of arms (i, j). To identify the optimal arm, we form an estimate of $\mathcal{E}_i$ to which end we form estimates for the variances σ_i^2, σ_j^2 and the correlation coefficient ρ_{ij}. We employ the following estimators

for the aforementioned quantities: For any $(i,j) \in \mathcal{S} = \{(p,q), 1 \le p,q \le K,\ p \ne q\}$,

$$\hat{\rho}_{ij} \triangleq 1 - \frac{1}{2}\left(\frac{\overline{X}_i^2}{\hat{\sigma}_i^2} + \frac{\overline{X}_j^2}{\hat{\sigma}_j^2} - 2\frac{\overline{X_iX_j}}{\hat{\sigma}_i\hat{\sigma}_j}\right), \tag{6}$$
$$\hat{\sigma}_i^2 = \overline{X}_i^2,\ \hat{\sigma}_j^2 = \overline{X}_j^2, \text{ where}$$
$$\overline{X}_i^2 = \frac{1}{n}\sum_{t=1}^{n} X_{it}^2, \text{ and } \overline{X_iX_j} = \frac{1}{n}\sum_{t=1}^{n} X_{it}X_{jt}.$$

The estimate for ρ in (6) is akin to that proposed in (Liu & Bubeck, 2014), which considers a simpler setting where all the arms are known to have unit variance, i.e., $\sigma_i^2 = 1,\ i = 1,\ldots,K$. For the unit variance setup, (Liu & Bubeck, 2014) establish via a likelihood ratio test that the difference based estimator for ρ_{ij}

$$1 - \frac{1}{2}\left(\overline{X}_i^2 + \overline{X}_j^2 - 2\overline{X_iX_j}\right) \tag{7}$$

is advantageous over the natural estimator for ρ_{ij} : $\frac{\overline{X_iX_j}}{\hat{\sigma}_i\hat{\sigma}_j}$. This superiority depends explicitly on the *a priori* knowledge of the variances being one, which is not applicable to the general setting considered here, i.e., a setting where the variances are not necessarily one. However, to exploit the optimality of the likelihood ratio test, we express the estimator above in the spirit of (7) which depend on the estimates of the variances to scale the samples to obtain

$$\hat{\rho}_{ij} = 1 - \tfrac{1}{2}\left(\frac{\overline{X}_i^2}{\hat{\sigma}_i^2} + \frac{\overline{X}_j^2}{\hat{\sigma}_j^2} - 2\frac{\overline{X_iX_j}}{\hat{\sigma}_i\hat{\sigma}_j}\right),$$

Unlike the unit variance setup of (Liu & Bubeck, 2014), it is not possible to obtain a difference based estimator in our setting. Nevertheless, $\hat{\rho}_{ij}$ concentrates faster as ρ_{ij} approaches 1 and this can be argued as follows: On the high probability event $\mathcal{C} = \left\{\frac{\sigma_1^2}{2} \le \hat{\sigma}_1^2 \le 2\sigma_1^2,\ \frac{\sigma_2^2}{2} \le \hat{\sigma}_2^2 \le 2\sigma_2^2\right\}$, we have

$$\begin{aligned}
&\mathbb{P}\left((1-\hat{\rho}_{ij}) - (1-\rho_{ij}) \ge \epsilon, \mathcal{C}\right)\\
&= \mathbb{P}\left(\frac{Y_{ijn}}{2n} - 1 \ge \frac{\epsilon}{(1-\rho_{ij})}, \mathcal{C}\right)\\
&\le \mathbb{P}\left(\frac{\bar{Y}_{ijn}}{2n} - 1 \ge \frac{\epsilon}{2(1-\rho_{ij})}\right)\\
&\le \exp\left(-\frac{n}{8}\min\left(\frac{\epsilon}{2(1-\rho_{ij})}, \left(\frac{\epsilon}{2(1-\rho_{ij})}\right)^2\right)\right),
\end{aligned}$$

where $Y_{ijn} \triangleq \frac{1}{(1-\rho_{ij})}\left(\frac{\overline{X}_i^2}{\hat{\sigma}_i^2} + \frac{\overline{X}_j^2}{\hat{\sigma}_j^2} - 2\frac{\overline{X_iX_j}}{\hat{\sigma}_i\hat{\sigma}_j}\right)$, and

$\bar{Y}_{ijn} \triangleq \frac{1}{(1-\rho_{ij})}\left(\frac{\overline{X}_i^2}{\sigma_i^2} + \frac{\overline{X}_j^2}{\sigma_j^2} - 2\frac{\overline{X_iX_j}}{\sigma_i\sigma_j}\right)$.

For any arm i, the corresponding MSE $\mathcal{E}_i$ is estimated using the quantities defined in (6) as follows:

$$\hat{\mathcal{E}}_i \triangleq \hat{\sigma}_j^2\left(1-\hat{\rho}_{ij}^2\right) + \sum_{p\ne i,j}\hat{\sigma}_p^2\left(1-\hat{\rho}_{ip}^2\right). \tag{8}$$

The main result concerning the concentration of the MSE estimate $\hat{\mathcal{E}}_i$ is given below.

Proposition 1. ***(MSE Concentration)*** *Assume $\sigma_i^2 \le 1, i = 1,\ldots,K$. Let $\hat{\mathcal{E}}_i$ be the MSE estimate given in* (8)*, for $i = 1,\ldots,K$. Then, for any $i = 1,\ldots,K$, and for any $\epsilon \in [0, 2K]$, we have*

$$\mathbb{P}\left(\left|\hat{\mathcal{E}}_i - \mathcal{E}_i\right| > \epsilon\right) \le 14K\exp\left(-\frac{nl^2\epsilon^2}{cK^5}\right),$$

where c is a universal constant, and $0 < l = \min_i\ \sigma_i^2$.

In the above, it suffices to look at $\epsilon \le 2K$, since $\mathcal{E}_i$ is less than $K-1$, owing to the assumption that $\sigma_i^2 \le 1,\ \forall i$.

Proof. See Section 7 for a sketch. The detailed proof is available in (Boda & Prashanth, 2019). □

The claim in Proposition 1 holds for the more general case of sub-Gaussian r.v.s $\{X_1,\ldots,X_K\}$. However, in this case, the MSE $\mathcal{E}_i$ is best in the class of linear estimators, and is not necessarily the minimum MSE estimator.

4. Uniform Sampling

A simple approach towards identifying the best arm is to select each pair $(i,j) \in \mathcal{S}$ equal number of times, estimate the MSE errors $\hat{\mathcal{E}}_p,\ p \in \mathcal{M}$ and recommend the arm with the lowest MSE estimate to be optimal, i.e., the samples used for estimation are $n_{ij} = \frac{n}{\binom{K}{2}} = \frac{2n}{K(K-1)},\ i \ne j$.

Theorem 1. *For uniform sampling, the probability of error in identifying the optimal arm is*

$$\mathbb{P}(\hat{A}_n \ne i^*) \le 84K^2\exp\left(-\frac{nl^2\Delta_{(1)}^2}{cK^7}\right),$$

where c is a universal constant.

Proof. Proof uses Proposition 1 along with an union bound and is available in (Boda & Prashanth, 2019). □

If the correlations between all pairs of arms and the variances of all the arms are similar, then the optimal strategy would involve sampling all pairs of arms an equal number of times. However, when this is not the case, uniform sampling is inferior. The elimination-based strategy that we present in the following section overcomes the shortcomings of uniform sampling.

Set $A_1 = \mathcal{S}, B_1 = \{1, \ldots, K\}$ and for $k = 1, \ldots, K-2$,

$$n_k = \left\lceil \frac{n - \binom{K}{2}}{C(K)\,(K+1-k)} \right\rceil, \text{ where } C(K) = \frac{K-1}{2} + \sum_{j=1}^{K-2} \frac{j}{K-j} \leq K \log K.$$

Phase 1: Sample each pair $(i, j) \in A_1$ for n_1 number of times, estimate the MSE differences using (8), remove the worst two arms from B_1 and the corresponding pair from A_1 to obtain B_2 and A_2 respectively.
Phase $k = 2, \ldots, K-1$:

1. Pull each pair in A_k $(n_k - n_{k-1})$ number of times. Estimate the MSEs using (8) and find the worst arm, say a_{k+1}, among the active arms in B_k.
2. Set $B_{k+1} = B_k \setminus a_{k+1}$ and $A_{k+1} = A_k \setminus \{(a_{k+1}, a_1), (a_{k+1}, a_2), \ldots, (a_{k+1}, a_k)\}$, where $B_k^c = \{a_1, \ldots, a_k\}$ is the set of arms that are out of contention by the end of phase $k-1$.

End of phase $K-1$: Recommend the arm in A_K.

Figure 1. Successive rejects algorithm for correlated bandits.

5. Successive Rejects

The successive rejects (SR) algorithm, which pulls pairs of arms[1] to identify the arm which minimizes MSE, operates over $K-2$ phases as described in Figure 1 . The idea is to maintain a set of active arms and pairs of arms (for phase k, these are denoted by A_k and B_k) and eliminate arms (and some of their corresponding pairs) that have high MSE. The elimination scheme employed in Figure 1 departs significantly from the approach adopted in the classic SR algorithm for finding the arm with highest mean. To illustrate this, consider a setting with 5 arms. If arms $4,\ 5$ are out of contention after phase 1, $A_2 = A_1 \setminus (4, 5)$. In the second phase, all the pairs in A_2 are pulled $(n_2 - n_1)$ number of times. Now, if arm 3 is out of contention at the end of this phase, the pairs $(3, 4)$ and $(3, 5)$ will be removed from A_2 and no longer be pulled in the later phases. From the foregoing, the total number of samples used by SR is

$$\begin{aligned}
&\binom{K}{2} n_2 + \left[\binom{K}{2} - \binom{2}{2}\right](n_3 - n_2) + \ldots \\
&+ \left[\binom{K}{2} - \binom{K-2}{2}\right](n_{K-1} - n_{K-2}) \\
&= \sum_{k=1}^{K-1} (k-1) n_k + (K-1) n_{K-1} < n,
\end{aligned}$$

where the final inequality follows by using the definition of n_k.

Notice that a strategy that finds the worst arm according to empirical MSE estimates and discards all pairs associated with that arm is clearly suboptimal, because samples from some of the discarded pairs of arms are essential to form estimate of MSE of arms which remain in contention. For e.g., in a 5-armed bandit setting, suppose that we discard all pairs associated with arm 5 in some round. This would impact the quality of MSE estimate of arm 1, since the pair $(1, 5)$ would be useful in training a better estimate of $\mathcal{E}_1$ via ρ_{15}.

Before presenting the main result that bounds the probability of error in identifying the best arm of the algorithm in Figure 1, we present the following problem complexities that capture the hardness of the learning task at hand (i.e., the order of number of samples required to find the best arm with reasonable probability):

$$H_2 = \max_i \frac{i}{\Delta^2_{(i)}} \text{ and } \overline{H} = \sum_{i \neq i^*} \frac{1}{\Delta_i^2}. \tag{9}$$

The quantities H_2 and $\overline{H}$, have a connotation similar to that in the classical bandit setup and using arguments similar to those employed in (Audibert et al., 2010), it can be shown that

$$H_2 \leq \overline{H} \leq \overline{\log}(K) H_2,$$

where $\overline{\log}(K) = \sum_{i=2}^{K-2} \frac{1}{i}$. Observe that the problem complexities depend both on the variances of the arms and the correlation between the arms through the gaps.

Theorem 2. *The probability of error in identifying the best arm of SR satisfies*

$$P(\hat{A}_n \neq i^*) \leq 84 K^3 \exp\left(-\frac{l^2}{cK^5} \frac{(n - \binom{K}{2})}{C(K) H_2}\right),$$

where c is a universal constant.

The detailed proof is available in (Boda & Prashanth, 2019). From Theorem 1, it is apparent that an uniform sampling strategy would require $O(\frac{K^7}{\Delta_2})$ samples to achieve a certain accuracy, while our SR variant for correlated bandits would require $O(K^6 \bar{H})$ number of samples. SR scores over uniform sampling w.r.t. dependence on the number of arms K because in our SR algorithm an increasing number of pairs of arms are removed from contention in successive

[1] With abuse of notation, (a_i, a_j) is used to denote the (unordered) pair of arms a_i, a_j.

phases. More importantly, SR has better dependence on the underlying gaps when compared to uniform sampling. In problem instances where the gaps are uneven, SR finds the best arm much faster than uniform sampling.

6. Lower Bound

To obtain the lower bound, we consider a K-armed Gaussian bandit problem with the underlying joint probability distribution governed by the following covariance matrix:

$$\Sigma = \begin{bmatrix} 1 & \rho & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho^2 & \rho^2 & \dots & \rho^2 \\ \rho & \rho^2 & 1 & \rho^3 & \dots & \rho^3 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho^2 & \rho^3 & \dots & \rho^{K-1} & 1 \end{bmatrix}. \tag{10}$$

Observe that Σ is a valid covariance matrix and is positive semi-definite. The MSEs corresponding to arms $1, \dots, K$ are $\mathcal{E}_1 = (K-1)(1-\rho^2)$, $\mathcal{E}_2 = (1-\rho^2)+(K-2)(1-\rho^4)$ and more generally

$$\mathcal{E}_i = (i-1) - \sum_{i=1}^{i-1} \rho^{2i} + (K-i)(1-\rho^{2i}), \quad i = 1, \dots, K.$$

Hence, we have the following order on the MSEs: $\mathcal{E}_1 \leq \mathcal{E}_2 \leq \ldots \leq \mathcal{E}_K$.

An approach in recent papers, cf. (Audibert et al., 2010; Kaufmann et al., 2015), for establishing lower bound for best-arm identification is to transform the bandit problem so that one of the sub-optimal arm is turned into an optimal one, while not affecting the rest of the arms. However, our setting involves correlated arms, with the correlation factors appearing in the mean-squared error objective and hence, one cannot make a sub-optimal arm optimal in a standalone fashion. We swap pairs of arms to interchange the MSE of a sub-optimal arm with that of the optimal arm and this introduces major deviations in the proof as compared classic K-armed case, as we shall soon see. We describe our problem transformations next.

We form $K-1$ transformations of the bandit problem formulated at the beginning of this section. For "problem m," $m = 2, \dots, K$, arm m is the best and for achieving this, we swap the first and mth rows in Σ. Let $\mathcal{G}$ be the pdf associated with the given problem as in (10), and $\mathcal{G}^m$ represent the pdf of the transformed bandit problem, where m represents the mth transformation. Since we consider arms whose samples are i.i.d. in time, the joint distribution of n samples is a product distribution of the underlying random variables $(\mathcal{G})^{\otimes n}$ and for the transformed problem by $(\mathcal{G}^m)^{\otimes n}$. For compactness, we use $\mathbb{P}_1 \triangleq \mathbb{P}_{(\mathcal{G})^{\otimes n}}$, $\mathbb{E}_1 \triangleq \mathbb{E}_{(\mathcal{G})^{\otimes n}}$ and $\mathbb{P}_m \triangleq \mathbb{P}_{(\mathcal{G}^m)^{\otimes n}}$, $\mathbb{E}_m \triangleq \mathbb{E}_{(\mathcal{G}^m)^{\otimes n}}$.

For any problem with $\rho^2 \leq UB_{\rho^2} \triangleq 1 - \frac{1}{\sqrt{K-2}}$, we define $c_1 = \frac{1}{1-UB_{\rho^2}}$ and $c_2 = \frac{\rho}{1-UB_{\rho^2}}$ and the min-max probability of error in identifying the optimal arm is given by the theorem below.

Theorem 3. *For any bandit strategy that returns the arm $\hat{A}_n$ after n rounds, there exists a transformation of the covariance matrix such that the probability of error on the transformed problem satisfies*

$$\max_{1 \leq m \leq K} \mathbb{P}_m(\hat{A}_n \neq m) \geq \frac{1}{6} \exp\Big(- \frac{6nK}{H_{lb}} - \binom{K}{2} n\tilde{\epsilon}_n \Big),$$

where $H_{lb} = \sum_{i \neq 1} \frac{1}{\Delta_i}$ is the problem complexity term,

$$\tilde{\epsilon}_n = \tilde{c}u \max\Big\{ \tfrac{8}{n} \log 12K(K-1)n, \sqrt{\tfrac{8}{n} \log 12K(K-1)n} \Big\}, \text{ and } \tilde{c} = \max(3c_1, 48c_2).$$

Proof. See Section 7 for a sketch. The detailed proof is available in (Boda & Prashanth, 2019). □

Note the gap between the upper and lower bounds on the probability of error in Theorems 2 and 3. The problem complexity term in the upper bound involved the square of the gaps, whereas the lower bound involves just the gaps. We believe the upper bound for SR algorithm is optimal in terms of gap dependence and it would be interesting future work to establish a lower bound that involves squares of the gaps. In the lower bound proof, the Kullback-Leibler divergence terms for the transformed problems were bounded above by the gaps (for e.g., see (12) in Section 7), leading to an overall lower bound with complexity H_{lb}. Nevertheless, the current proof is challenging owing to (i) pairs of arms being pulled in each round; (ii) the covariance matrix in (10) is non-trivial and its problem transformations are novel and finally, (iii) arriving at the bound for the aforementioned KL-divergence terms requires non-trivial algebraic effort.

7. Analysis

Proof of Proposition 1 ***(Sketch)***

The MSE estimate in (8) involves sample variances and sample correlation coefficients, and hence, MSE concentration requires each of these quantities to concentrate. While one can use Bernstein's inequality for handling sample variance, a finite sample concentration bound for sample correlation coefficient does not exist, to the best of our knowledge. We fill this gap in the result below.

Lemma 1. *For independent Gaussian rvs X_i, $i = 1, \dots, n$, with mean zero and covariance matrix Σ as defined in* (3) *and with $\hat{\sigma}_i^2$, $\hat{\rho}_{ij}$ formed from n samples using* (6)*, for any $i, j = 1, \dots, K$, and for any $\epsilon \in [0, \eta]$, we have*

$$\mathbb{P}\left(|\hat{\rho}_{ij} - \rho_{ij}| > \epsilon\right)$$

$$\leq 26 \exp\left(-\frac{n}{8}\frac{1}{36(1+\eta)}\min\left(\frac{l\epsilon}{3}, \left(\frac{l\epsilon}{3}\right)^2\right)\right),$$

where l is a positive constant satisfying $l \leq \sigma_i^2 \leq 1$, $\forall i$.

The proof is available in (Boda & Prashanth, 2019). We now provide a sketch of the proof of Proposition 1.

Proof. ***(Sketch)***
We prove the proposition for $i = 1$, but the analysis below holds in general. Consider the event $\mathcal{B} = \left\{\sigma_i^2 - \epsilon \leq \hat{\sigma}_i^2 \leq \sigma_i^2 + \epsilon, i = 1, \ldots, K, \right.$ $\left.\text{and } \rho_{1j} - \epsilon \leq \hat{\rho}_{1j} \leq \rho_{1j} + \epsilon, \text{ for } j = 2, \ldots, K\right\}$. Then, from the lemma above on sample correlation coefficient concentration and sample variance concentration (cf. Proposition 2.2 in (Wainwright, 2015)), we have

$$\mathbb{P}\left(\mathcal{B}^c\right) \leq 28K \exp\left(-\tfrac{n}{8}\tfrac{1}{36(1+\eta)}\min\left(\tfrac{l\epsilon}{3}, \left(\tfrac{l\epsilon}{3}\right)^2\right)\right).$$

We shall bound the tail probability $\mathbb{P}\left(\hat{\mathcal{E}}_1 - \mathcal{E}_1 > \epsilon\right)$ on the event $\mathcal{B}$ and use the bounds on the probability of $\mathcal{B}^c$ to arrive at an unconditional bound on the aforementioned tail probability. Using an union bound, we have

$$\begin{aligned}
&\mathbb{P}\left(\hat{\mathcal{E}}_1 - \mathcal{E}_1 \geq \epsilon\right) \leq \mathbb{P}\left(\hat{\mathcal{E}}_1 - \mathcal{E}_1 \geq \epsilon, \mathcal{B}\right) + \mathbb{P}\left(\mathcal{B}^c\right) \\
&\leq \mathbb{P}\left(\hat{\sigma}_2^2\left(1 - \hat{\rho}_{12}^2\right) - \sigma_2^2\left(1 - \rho_{12}^2\right) \geq \frac{\epsilon}{(K-1)}, \mathcal{B}\right) \\
&+ \sum_{p=3}^{K} \mathbb{P}\left(\hat{\sigma}_p^2\left(1 - \hat{\rho}_{1p}^2\right) - \sigma_p^2\left(1 - \rho_{1p}^2\right) \geq \frac{\epsilon}{(K-1)}, \mathcal{B}\right) \\
&\quad + \mathbb{P}\left(\mathcal{B}^c\right). \qquad (11)
\end{aligned}$$

The first term on the RHS above can be bounded as follows:

$$\begin{aligned}
&\mathbb{P}\left(\hat{\sigma}_2^2\left(1 - \hat{\rho}_{12}^2\right) - \sigma_2^2\left(1 - \rho_{12}^2\right) \geq \frac{\epsilon}{(K-1)}, \mathcal{B}\right) \\
&\leq \mathbb{P}\left(\rho_{12} - \hat{\rho}_{12} \geq \frac{\epsilon}{4(K-1)(1+\eta)}, \mathcal{B}\right) \\
&\quad + \mathbb{P}\left(\hat{\sigma}_2^2 - \sigma_2^2 \geq \frac{\epsilon}{2(K-1)}\right).
\end{aligned}$$

The proof proceeds by applying concentration results for sample variance, sample standard deviation and sample correlation coefficients to bound the RHS above. The rest of the terms on the RHS of (11) are handled in a similar fashion. The reader is referred to Section 7.2 of (Boda & Prashanth, 2019) for further details. □

Proof of Theorem 3 ***(Sketch)***

Proof. Consider problem m with underlying covariance matrix Σ_m. For $(i,j) \in \mathcal{S}$, let $\nu_i\nu_j$ and $\nu_i'\nu_j'$ denote bivariate normal distributions with variance and correlations specified by Σ_1 and Σ_i, respectively. Let $\mathrm{KL}_{ij}^m \triangleq KL(\nu_i\nu_j||\nu_i'\nu_j')$ denote the Kullback-Leibler divergence between $\nu_i\nu_j$ and $\nu_i'\nu_j'$, where the latter distributions are derived from $\mathcal{G}_m$.

If $\rho^2 \leq \frac{2K-5}{2K-4}$, we have the following bound for $j = 3, \ldots, K$:

$$\begin{aligned}
\mathrm{KL}_{1j}^2 &= \frac{1}{2}\left(\frac{2(1-\rho^3)}{1-\rho^4} - 2 + ln\frac{1-\rho^4}{1-\rho^2}\right) \\
&\leq \frac{\rho^2}{2} = \frac{\rho^2(1-\rho^2)(K-2)}{2(1-\rho^2)(K-2)} \leq \Delta_2.
\end{aligned}$$

Along similar lines, we can infer that

$$\max\{\mathrm{KL}_{1m}^m, \mathrm{KL}_{1j}^m, \mathrm{KL}_{mj}^m,\ j \neq 1, m\} \leq \Delta_m. \qquad (12)$$

For $(i,j) \in \mathcal{S}$, let N_{ij} denote the number of samples obtained from the joint distribution of (X_i, X_j). Let $n_{ij} = \mathbb{E}_1 N_{ij}$, for $(i,j) \in \mathcal{S}$. Observe that

$$\sum_{(i,j)\in\mathcal{S}} N_{ij} = \sum_{(i,j)\in\mathcal{S}} n_{ij} = n.$$

Notice that the problem transformations impact the distribution of each arm and hence, we cannot employ a change of measure identity similar to (Audibert et al., 2010). Instead, we factor in the KL-divergences KL_{ij}, $\forall\ (i,j) \in \mathcal{S}$ and derive a change of measure identity as follows: for any measurable event $\mathcal{E}$ based on the samples,

$$\begin{aligned}
\mathbb{P}_m(\mathcal{E}) &= \mathbb{E}_{\nu_1\cdots\nu_K}\left[\mathbf{1}\{\mathcal{E}\}\prod_{(i,j)\in\mathcal{S}}\prod_{s\in\mathcal{T}_{ij}(n)}\frac{d\nu_i'\nu_j'}{d\nu_i\nu_j}(X_{i,s}, X_{j,s})\right] \\
&= \mathbb{E}_{\nu_1\cdots\nu_K}\left[\mathbf{1}\{\mathcal{E}\}\exp\left(\sum_{(i,j)\in\mathcal{S}} -N_{ij}\widehat{\mathrm{KL}}_{ij,N_{ij}}\right)\right] \\
&= \mathbb{E}_1\left[\mathbf{1}\{\mathcal{E}\}\exp\left(\sum_{(i,j)\in\mathcal{S}} -N_{ij}\widehat{\mathrm{KL}}_{ij,N_{ij}}\right)\right],
\end{aligned}$$

where $\mathcal{T}_{ij}(n)$ is the set of time instants when the algorithm pulled the tuple (i,j). For $(i,j) \in \mathcal{S}, 1 \leq t \leq n$, let

$$\widehat{\mathrm{KL}}_{ij,t} = \frac{1}{t}\sum_{s=1}^{t}\log\frac{d\nu_i d\nu_j}{d\nu_i' d\nu_j'}(X_{i,s}, X_{j,s})$$

where $(X_{i,s}, X_{i,s})$ are i.i.d. $\sim \mathcal{G}$ for all $s \leq t$. Here, $\Sigma_{(i,j),m}$ is the covariance matrix of X_i, X_j for problem m, and is a submatrix of Σ_m. Let $\xi = \left\{\forall (i,j) \in \mathcal{S}, 1 \leq t \leq n, \widehat{\mathrm{KL}}_{(ij),t} - \mathrm{KL}_{ij} \leq \tilde{\epsilon}_t\right\}$, where $\tilde{\epsilon}_t$ is as defined in the theorem statement. Then, for $m = 1, \ldots, K$, $\mathbb{P}_m(\xi) \geq 5/6$, which implies that the empirical divergences concentrate.

Now, considering algorithms that satisfy $\mathbb{E}_1(\hat{A}_n \neq 1) \leq 1/2$, we obtain

$$\mathbb{P}_m(\hat{A}_n \neq m) \geq \frac{1}{6}\exp\Big(-\sum_{(i,j)\in\mathcal{S}} 6n_{ij}\mathrm{KL}_{ij}^m - \binom{K}{2}n\tilde{\epsilon}_n\Big)$$

$$\geq \frac{1}{6}\exp\Big(-\sum_{(i,j)\in\mathcal{S}} 6n_{ij}\Delta_m - \binom{K}{2}n\tilde{\epsilon}_n\Big),$$

where the final inequality follows from the bounds on KL_{ij}^m above.

Let $l_i = n_{12} + \ldots n_{1K} + \sum_{i\neq 1, i} n_{ij}$. Then, there exists an i such that $l_i \leq \frac{n(K-1)}{H_{lb}\Delta_i}$, where H_{lb} is as defined in the theorem statement. If not, then we obtain a contradiction. For the aforementioned i, we have

$$\mathbb{P}_m(\hat{A}_n \neq m) \geq \frac{1}{6}\exp\Big(-\frac{6nK}{H_{lb}} - \binom{K}{2}n\tilde{\epsilon}_n\Big). \quad (13)$$

The main claim follows. □

8. Numerical Experiments

We show a few simple experiments here to illustrate our theoretical analysis. Since, this line of work is new, we compare our successive rejects type algorithm and uniform sampling which is optimal in some settings. We show three experiments, in which all the arms are jointly Gaussian having mean zero and unit variance. Each experiment can be seen to consist of two clusters of arms with the arms in each cluster being independent of the arms in the other cluster. Arm 1, in the first cluster, is optimal in all the three experiments and the arms in the second cluster are typically less correlated among themselves than the arms in the first cluster.

In a setting with 35 arms, we employ the following covariance matrices for the three experimental setups:

$$\Sigma_1 = \begin{bmatrix} \mathbf{M}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{25\times 25} \end{bmatrix},\ \Sigma_2 = \begin{bmatrix} \mathbf{M}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{Tr}_{31\times 31} \end{bmatrix}, \quad (14)$$

$$\Sigma_3 = \begin{bmatrix} 1 & 0.5 & 0.45 & 0.5 & \mathbf{0} \\ 0.5 & 1 & 0.45 & 0.4 & \mathbf{0} \\ 0.45 & 0.45 & 1 & 0.4 & \mathbf{0} \\ 0.5 & 0.4 & 0.4 & 1 & \mathbf{0} \\ & & \mathbf{0} & & \mathbf{I}_{30\times 30} \end{bmatrix}, \quad (15)$$

where $\mathbf{M}_1 = [1\ 0.9\ 0.9\ 0.9;\quad 0.9\ 1\ 0.85\ 0.85;$
$0.9\ 0.85\ 1\ 0.85;\quad 0.9\ 0.85\ 0.85\ 1]$, and $\mathbf{Tr}_{K-5\times K-5}$ is a tridiagonal matrix with ones along the main diagonal, 0.2 in the diagonals above and below and zeros elsewhere. Notice that Σ_i is a block diagonal matrix for each $i = 1, 2, 3$ and hence, its eigenvalues are union of the nonzero diagonal submatrices. It is easy to verify that the individual blocks, i.e., M_1 and $\mathbf{Tr}_{K-5\times K-5}$, are positive semi-definite and hence, so is Σ_i for each i.

In Example 1 corresponding to covariance matrix Σ_1, arms in the first cluster are highly correlated amongst themselves, and arms in the second cluster are independent of all the arms. On the other hand, in Example 2 corresponding to covariance matrix Σ_2, arms in the first cluster are highly correlated among themselves and the arms in the second cluster are weakly correlated amongst themselves. Finally, in Example 3 corresponding to covariance matrix Σ_3, arms in the first cluster are weakly correlated among themselves and arms in the second cluster are independent of all the arms. In all the three examples, multiple arms in the first cluster have MSE close to that of the optimal arm. Clearly, more samples of the pairs of arms corresponding to the first cluster of arms are required to identify the optimal arm accurately. As number of arms K increases, the proportion of samples used for pairs corresponding to the clearly sub-optimal arms increases at a faster rate for uniform sampling algorithm as compared to SR.

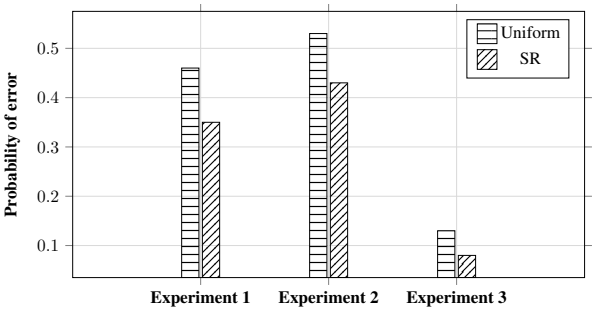

Figure 2. Probability of error for uniform sampling and SR algorithms on three different problems. The results are averages from 200 independent replications.

We conduct our experiments with the number of samples equaling $\cong \frac{\overline{H}}{32^2}$ for each experiment. Figure 2 compares the probability of error for the three settings with covariance matrices given in (14)–(15). In all three settings, SR recommends the optimal arm with higher probability, and this is because SR algorithm rejects the sub-optimal arms in the beginning phases using fewer samples and allocates more samples to the first cluster to distinguish between the arms in this cluster.

9. Conclusions

We presented a new formulation of the K-armed bandit problem where the goal, using the MSE criterion, is to find an arm that best captures information about all arms. Both estimation of MSE for individual arms, and exploration to find the best arm in a correlated bandit are challenging. We proposed an MSE estimator that uses samples from the distribution underlying any pair of arms, and showed that our estimator concentrates. We adapted the SR algorithm to successively eliminate arm pairs, and proved a bound on the probability of error in identifying the best arm. We also derived a lower bound for the correlated bandit problem.

ACKNOWLEDGMENTS

The authors would like to thank Akshay Reddy for inputs on the phase lengths in the SR algorithm. The first author would like to thank Prof. Prakash Narayan for guidance on Sampling rate distortion problem which led to the current work. This work was supported by the U.S. National Science Foundation under the Grant CCF-1319799.

References

Audibert, J., Bubeck, S., and Munos, R. Best arm identification in multi-armed bandits. In *Conference on Learning Theory*, pp. 41–53, 2010.

Boda, V. P. Reconstructing Gaussian sources by spatial sampling. *IEEE Transactions on Information Theory*, 2019.

Boda, V. P. and Narayan, P. Universal Sampling Rate Distortion. *IEEE Transactions on Information Theory*, 64 (12):7742 – 7758, 2018.

Boda, V. P. and Prashanth, L. Correlated bandits or: How to minimize mean-squared error online. *CoRR*, abs/1902.02953, 2019.

Deshpande, A., Guestrin, C., Madden, S. R., Hellerstein, J. M., and Hong, W. Model-driven data acquisition in sensor networks. In *International conference on Very large data bases*, pp. 588–599, 2004.

Erkan, G. and Radev, D. R. Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research*, 22:457–479, 2004.

Guestrin, C., Krause, A., and Singh, A. P. Near-optimal sensor placements in gaussian processes. In *International conference on Machine learning*, pp. 265–272, 2005.

Hajek, B. Notes for ECE 534: an exploration of random processes for engineers. *Univ. of Illinois at Urbana–Champaign*, 2009.

Heidemann, J., Klier, M., and Probst, F. Identifying key users in online social networks: A pagerank based approach. 2010.

Kaufmann, E., Cappé, O., and Garivier, A. On the complexity of best arm identification in multi-armed bandit models. *The Journal of Machine Learning Research*, 2015.

Krause, A., Leskovec, J., Guestrin, C., VanBriesen, J., and Faloutsos, C. Efficient Sensor Placement Optimization for Securing Large Water Distribution Networks. *Journal of Water Resources Planning and Management*, 134 (6):516–526, 2008.

Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., and Glance, N. Cost-effective outbreak detection in networks. In *ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 420–429, 2007.

Liu, C. Y. and Bubeck, S. Most correlated arms identification. In *Confernce on Learning Theory*, pp. 623–637, 2014.

Long, J., Memik, S. O., Memik, G., and Mukherjee, R. Thermal monitoring mechanisms for chip multiprocessors. *ACM Transactions on Architecture and Code Optimization*, 5(2):9, 2008.

Paul, U., Ortiz, L., Das, S. R., Fusco, G., , and Buddhikot, M. M. Learning probabilistic models of cellular network traffic with applications to resource management. In *IEEE International Symposium on Dynamic Spectrum Access Networks*, pp. 82–91, 2014.

Wainwright, M. High-dimensional statistics: A non-asymptotic viewpoint. `http://www.stat.berkeley.edu/~mjwain/stat210b/Chap2_TailBounds_Jan22_2015.pdf`, 2015.