Active Distribution Learning from Indirect Samples

Samarth Gupta
Dept. of ECE
Carnegie Mellon University
Pittsburgh, PA 15213
Email: samarthg@andrew.cmu.edu

Gauri Joshi
Dept. of ECE
Carnegie Mellon University
Pittsburgh, PA 15213
Email: gaurij@andrew.cmu.edu

Osman Yağan
Dept. of ECE
Carnegie Mellon University
Pittsburgh, PA 15213
Email: oyagan@ece.cmu.edu

Abstract—**This paper studies the problem of *learning* the probability distribution P_X of a discrete random variable X using indirect and sequential samples. At each time step, we choose one of the possible K functions, $g_1, \ldots, g_K$ and observe the corresponding sample $g_i(X)$. The goal is to estimate the probability distribution of X by using a minimum number of such sequential samples. This problem has several real-world applications including inference under non-precise information and privacy-preserving statistical estimation. We establish necessary and sufficient conditions on the functions $g_1, \ldots, g_K$ under which asymptotically consistent estimation is possible. We also derive lower bounds on the estimation error as a function of total samples and show that it is *order-wise* achievable. Leveraging these results, we propose an iterative algorithm that i) chooses the function to observe at each step based on past observations; and ii) combines the obtained samples to estimate p_X. The performance of this algorithm is investigated numerically under various scenarios, and shown to outperform baseline approaches.**

Index Terms—**distribution learning, hidden random variable, indirect samples, sequential decision-making**

I. INTRODUCTION

The modern world is rich with various types of data such as images, video, cloud job execution traces, social network data, and crowd-sourced survey data. These data can provide invaluable insights into the underlying random phenomenon which are generally not directly observable due to privacy concerns, or imprecise measurement mechanisms. For example, if we want to estimate the income distribution of a population, their salary data may not be public. However, it may be possible to estimate the income distribution using surveys about their spending on luxury goods, or whether their income is above or below some given thresholds.

In this work we seek to design techniques to use indirect and correlated samples to estimate the probability distribution of a hidden random phenomenon. We consider a stylized model, shown in Fig. 1, where a hidden variable X can be sampled through functions $g_1(X), \ldots, g_K(X)$, referred to as *arms*. Our objective is to accurately estimate the probability distribution of X with the minimum number of samples; see Section section II for a precise definition of the problem.

A. Related Prior Work

Learning the distribution of a random variable from its samples is a well-studied research problem [1]–[3] in information theory and theoretical computer science. Some
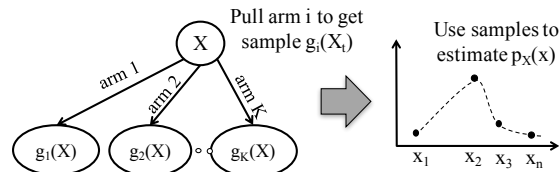

Fig. 1: *At step t we pull some arm i and observe $g_i(X_t)$, where X_t is an i.i.d. realization of the hidden variable X. Our objective is use the samples to estimate the distribution $p_X(x)$.*

works [4], [5] are interested in finding the min-max or worst-case loss for various loss functions; e.g., L2-loss and Kullback-Liebler (KL) divergence. Some other works study the properties of distribution from samples observed [6]–[9]. Unlike the majority of the literature on distribution learning, here we assume that only functions $g_i(X)$ of the samples can be observed instead of direct samples of X.

Inferring a hidden random variable from indirect samples is also related to works in estimation theory [10], where the objective is to estimate a set of parameters θ using observations $y_1, \ldots, y_T$ that follow a model $p(Y|\theta)$. In our problem, the unknown distribution P_X is analogous to the parameter θ while samples $g_i(X_t)$ correspond to the observations $y_1, \ldots, y_T$. A key difference between our model and typical parameter estimation problems is that we decide on the arm (say, arm k) to be pulled in each time slot t to obtain the corresponding sample $g_k(X_t)$. Our problem formulation falls under the class of sequential design of experiments [11], [12]. Such sequential/active learning frameworks have been considered for the purposes of hypothesis testing in [13]–[15]. The aspect of choosing arm in each time step is also closely related to the multi-armed bandit (MAB) sequential decision-making framework [16]–[18]. In the classical MAB framework [19], each arm gives a reward according to some unknown distribution that is independent across arms, and the objective is to maximize the total reward for a given number of pulls, or to identify the arm that has the largest mean reward with as few pulls as possible [20]–[23]. In contrast, the arms $g_1(X), \ldots, g_K(X)$ are correlated through the common hidden variable X in our formulation. In most sequential experiment design and multi-armed bandit problems, the main strategy is to identify the single "best" arm and then exploit it. What makes our formulation interesting is that there may not be a unique best arm for the purposes of

learning the distribution of X. Instead, the optimal strategy will often involve a combination of arms to be pulled, with each arm being pulled a specific number of times. It is also this aspect that makes our problem challenging since the optimal combination of arms to be pulled (to learn P_X) depends itself on the distribution P_X.

B. Main Contributions

To the best of our knowledge, this is the first work to consider the problem of using sequential, indirect samples to learn the distribution of a hidden random variable. Our main contributions include i) deriving conditions on the functions $g_1(X), g_2(X), \ldots, g_K(X)$ needed for *asymptotically consistent* estimation of the hidden distribution; ii) deriving a lower bound on the estimation error and showing that it is *order-wise* achievable; and iii) proposing algorithms that sequentially decide which arm to pull and return an estimation of P_X at each time step. Through simulations, our algorithms are also shown to outperform several baseline strategies in terms of error for a given number of pulls and the number of pulls needed to estimate P_X within a given error.

II. PROBLEM FORMULATION

Consider a discrete random variable X that can take values from a finite alphabet $\{x_1, x_2, \ldots, x_n\}$ with an *unknown* probability distribution $P_X = [p_1, p_2, \ldots, p_n]^\intercal$. Throughout this paper, we assume $p_i > 0$ for all i. Our objective is to estimate this probability distribution using a sequence of independent samples from K functions $\{g_1, g_2, \ldots, g_K\}$, where each g_i is a mapping from $\{x_1, x_2, \ldots, x_n\}$ to $\mathbb{R}$; throughout, we refer to these functions also as *arms*. More precisely, with $\{X_t : t = 1, 2, \ldots\}$ denoting a sequence of independent and identically distributed (i.i.d.) realizations of X, we can choose and observe only one of the K possible outcomes $g_1(X_t), \ldots, g_K(X_t)$, at each step $t \in \mathcal{N}$. Broadly speaking, for a given set of functions $\{g_1, g_2, \ldots, g_K\}$, our goal is to derive an efficient algorithm i) to decide which function will be observed at each iteration step t, and ii) to come up with an estimate $\tilde{P}_X(t) = [\tilde{p}_1(t), \tilde{p}_2(t), \ldots, \tilde{p}_n(t)]^\intercal$ of the true probability distribution based on the observations until step t. Ultimately, we aim to minimize the mean-squared error of this estimation, formally defined below.

Definition 1 (Estimation Error)**.** *The error in estimating* $P_X = [p_1, p_2, \ldots, p_n]^\intercal$ *at step* t *(i.e., after observing* t *samples) is defined as*

$$\varepsilon(t) = \mathbb{E}\left[\sum_{j=1}^n (\tilde{p}_j(t) - p_j)^2\right]. \tag{1}$$

Here, $\tilde{p}_i(t)$ denotes the estimation obtained after observing t samples $g_{c_1}(X_1), g_{c_2}(X_2), \ldots, g_{c_t}(X_t)$, where $c_\tau \in \{1, \ldots, K\}$ is the arm pulled at step τ. We now give two examples to illustrate and clarify the problem formulation.

Example 1. *Fig. 2 shows an example in which X takes three possible values $\{x_1, x_2, x_3\}$, and there are three arms,*
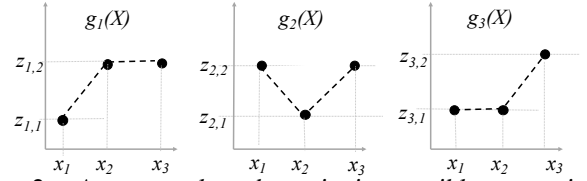

Fig. 2: *An example where it is possible to estimate $\{p_1, p_2, p_3\}$ asymptotically consistently (See Definition 2) although no arm is invertible.*
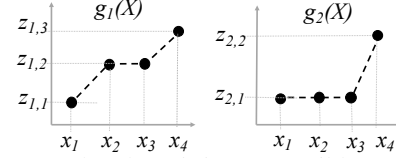

Fig. 3: *An example where it is not possible to get consistent estimation due to the ambiguity between p_2 and p_3.*

$g_1, g_2,$ *and g_3. The values of $g_1, g_2,$ and g_3 corresponding to x_1, x_2, x_3 are illustrated in Fig. 2. In arm 1, output $z_{1,2}$ can come from either x_2 or x_3. This ambiguity exists in output $z_{2,2}$ (between x_1 and x_3) in g_2 and in output $z_{3,1}$ (between x_1 and x_2) in g_3. Inspite of these ambiguities, it is possible to estimate p_1, p_2 and p_3 as we will show in Section III-A.*

Example 2. *Fig. 3 illustrates an example with two arms, with each arm showing outputs corresponding to $\{x_1, x_2, x_3, x_4\}$. Arm 1 has ambiguity coming from output of x_2 and x_3, whereas arm 2 exhibits ambiguity in the output of $x_1, x_2,$ and x_3. For this set of functions it is possible to estimate only $p_2 + p_3$ and nothing else can be known about p_2 and p_3, as we will prove in Section III-A.*

We note that if a function g_k is invertible, then every output sampled from g_k will be uniquely matched to a single value (say, x_j) that X can take without any ambiguity. In those cases, it would be optimal (in the sense of minimizing $\varepsilon(t)$ for each t) to pull g_k at every step. We formally prove a more general version of this result in Theorem 2.

III. STRUCTURAL PROPERTIES OF THE FUNCTIONS $g_k(X)$

A. Conditions for asymptotically consistent estimation

Definition 2 (Asymptotically consistent estimation)**.** *Given a random variable X and arms $\{g_1, g_2, \ldots, g_K\}$, the estimated probability distribution $\{\tilde{p}_1(t), \ldots, \tilde{p}_n(t)\}$ is said to be asymptotically consistent if $\lim_{t \to \infty} \varepsilon(t) = 0$.*

For each $k = 1, \ldots, K$, let $\{z_{k,1}, z_{k,2}, \ldots, z_{k,m_k}\}$ denote the set of possible outcomes (i.e., the range of g_k) of the function g_k; evidently, m_k is the number of distinct outputs of g_k. The information about g_k required to estimate P_X can be captured in matrix A_k with m_k rows and n columns, where

$$A_k(i,j) = \begin{cases} 1, & \text{if } g_k(x_j) = z_{k,i} \\ 0, & \text{otherwise}, \end{cases}$$

for each $i = 1, \ldots, m_k$ and $j = 1, \ldots, n$. Informally, $A_k(i,j) = 1$ if output $z_{k,i}$ could have been generated by x_j

in arm k. We refer A_k as the *Sample Generation Matrix* for arm k. Let the matrix A be given by $A = [A_1^\mathsf{T}, A_2^\mathsf{T}, \ldots, A_K^\mathsf{T}]^\mathsf{T}$; the size of A is $m \times n$, where $m = m_1 + \ldots + m_K$. The corresponding matrices $A^{\text{Example-1}}$ and $A^{\text{Example-2}}$ for Examples 1 and 2, respectively are shown below.

$$
A^{\text{Example-1}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad A^{\text{Example-2}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}
$$

Theorem 1. *It is possible to achieve asymptotically consistent estimation if and only if rank$(A) = n$.*

Proof of Theorem 1. Recall that $z_{k,i}$ represents the ith distinct output of arm k. Let $q_{k,i}$ denote the probability of observing $z_{k,i}$ each time arm k is pulled. Consider the system of linear equations below relating these probabilities to the probability distribution of X:

$$
q_{k,i} \triangleq \sum_{j=1}^{n} A_k(i,j)p_j, \qquad \begin{array}{l} k = 1, \ldots, K \\ i = 1, \ldots, m_k \end{array} \tag{2}
$$

These set of equations can be written as

$$
AP_X = Q,
$$

with Q denoting the vector $[q_{1,1}, \ldots, q_{1,m_1}, \ldots q_{K,m_K}]^\mathsf{T}$.

Suppose now that A is full rank. In order to construct an asymptotically consistent estimate of $P_X = [p_1, \ldots, p_n]^\mathsf{T}$, assume that arms are pulled in a round-robin manner. Thus, at step t we will have $\frac{t}{K}$ samples from each arm. With $t_{k,i}$ denoting the number of times $z_{k,i}$ is observed in t steps, we let $\hat{q}_{k,i}(t) = \frac{t_{k,i}}{t/K}$ be the estimate of $q_{k,i}$ at step t. By virtue of Strong Law of Large Numbers, we have $\hat{q}_{k,i}(t) \to q_{k,i}$ almost surely as t goes to infinity, that is, the estimates $\hat{q}_{k,i}(t)$ are asymptotically consistent. Given that A is full rank, the estimates $\hat{q}_{k,i}(t)$ can be used to obtain a unique solution of $P_X = [p_1, p_2, \ldots, p_n]^\mathsf{T}$ from the system of equations (2). Given that n is finite, this unique solution will constitute an asymptotically consistent estimation of P_X as well.

Conversely, if rank$(A) < n$, it is not possible to obtain a unique solution of the system of equations in (2). This implies that even if consistent estimation of each $q_{k,i}$ is possible, it is not possible to achieve asymptotically consistent estimation of the probability distribution, P_X. $\qquad \square$

Clearly, rank$(A^{\text{Example-1}}) = n$ while rank$(A^{\text{Example-2}}) < n$. Thus, asymptotically consistent estimation is possible for the set of functions in Example 1 but not in Example 2.

Remark 1. *Theorem 1 is not constrained to the Definitions 1 and 2 of error and asymptotically consistent estimation, respectively. In fact, the condition rank$(A) = n$ is necessary and sufficient to have (the possibility of achieving) for any $\epsilon > 0$ and $\forall i$ that $| \tilde{p}_i(t) - p_i | < \epsilon$ for all t sufficiently large.*

B. Redundant functions/arms

Recall the definition of sample generation matrix A_k for each arm k given in Section III-A.

Definition 3 (Redundant Arm). *An arm r is said to be a redundant if there exists another arm s such that the row space of A_r is a strict subset of the row space of A_s.*

Informally, this means that all information produced by arm r can be generated by arm s. For example, in Fig. 3 we see that arm 2 generates information about $p_1 + p_2 + p_3$, while arm 1 generates information about p_1 and $p_2 + p_3$ separately; also, both arms generate information about p_4 separately. Therefore, information produced by arm 2 can be generated by arm 1. This observation is made precise next.

Theorem 2. *If an arm r is redundant, then it is suboptimal to pull arm r at any step t for the purpose of minimizing $\varepsilon(t)$.*

Proof of Theorem 2. Since arm r is redundant, it implies that there exists an arm s such that the row space of A_r is a strict subset of the row space of A_s. Suppose we are given a set $\mathcal{Z}_s$ of the samples from arm s. For each observation $z_{s,i} \in \mathcal{Z}_s$, consider the set $\mathcal{X}_{s,i} = \{x : g_s(x) = z_{s,i}\}$. Since the row space of A_r is a subset of row space of A_s, each $x \in \mathcal{X}_{s,i}$ will be mapped to the same observation $g_r(x)$ in arm r. More formally, we have $g_r(x) = z_{r,i} \; \forall x \in \mathcal{X}_{s,i}$. Repeating this for all $z_{s,i} \in \mathcal{Z}_s$, we can construct a new sample set $\mathcal{Z}_r$. For the underlying set of realizations $\mathcal{X} = \{X_1, X_2, \ldots\}$ that lead to the samples $\mathcal{Z}_s$, we can see that $\mathcal{Z}_r$ is the exact sample set that one would have obtained if arm r was pulled each time instead of arm s. This shows that if arm r is redundant, then each time the arm s is pulled, we automatically know the sample that we would have obtained from arm r, thereby obviating the need to ever pull arm r. Thus, for the purposes of minimizing $\varepsilon(t)$, it is suboptimal to ever pull a redundant arm r; pulling the arm s at each time step will be at least as good. $\qquad \square$

By Theorem 2, if an *invertible* arm exists, all other arms will be redundant. This leads to the following corollary.

Corollary 1. *If there is an invertible arm, then the optimal action (for the purpose of minimizing $\varepsilon(t)$) is to pull the invertible arm at every step.*

Remark 2. *The proofs of Theorem 2 and Corollary 1 are not specific to the error metric given in Definition 1. Thus, both results hold true under other error metrics as well; e.g., L1 norm, KL divergence etc.*

IV. BOUNDS ON THE ESTIMATION ERROR

A. Lower bounds on the estimation error

We first derive a *crude* lower bound on the estimation error which does not depend on the functions $\{g_1, g_2, \ldots g_K\}$.

Theorem 3 (Crude Lower Bound). *Estimation error of any unbiased estimator for the problem in Section II is lower bounded by $\sum_{j=1}^{n} \frac{p_j(1-p_j)}{t}$.*

Proof of Theorem 3. From Corollary 1, we know that it is optimal to always pull the invertible arm if there exists one. It is also clear that the optimal error can only decrease when an additional arm is included in the set of possible arms we can choose. Thus, for the purpose of deriving a lower bound on the estimation error, we can assume the existence of an invertible arm which is pulled in all t steps.

We define $\hat{p}_j(t) = \frac{t_{x_j}}{t}$, as the corresponding empirical estimator (which is also the maximum likelihood estimator), where t_{x_j} is the number of times the output corresponding to x_j was observed (from the invertible arm) in t steps. Under this scenario, the estimation error is given by

$$\varepsilon(t) = \sum_{j=1}^{n} \text{Var}\left[\hat{p}_j(t)\right] = \sum_{j=1}^{n} \frac{p_j(1-p_j)}{t}, \quad (3)$$

as t_{x_j} is a Binomial random variable, which has variance $tp_j(1-p_j)$. This also gives the minimum possible variance for any unbiased estimator (given the samples from the invertible arm). Using this fact and Corollary 1, we establish Theorem 3.

□

Remark 3. *The lower bound in Theorem 3 is achieved if an invertible arm exists.*

The final estimation error after a total of t steps depends on the number of times each arm is pulled till step t. Due to the sequential nature of the problem, we have control over which arm is pulled at each time step and hence on the number of times each arm k is pulled till step t, i.e., t_k. Next, we derive a lower bound on the error of any unbiased estimator given the number of times each arm is pulled.

Theorem 4 (Lower bound on error for a given number of pulls). *Let* $\mathbf{t} = [t_1, t_2, \ldots, t_K]^\mathsf{T}$ *be the number of times arms* $\{g_1, \ldots, g_K\}$ *are pulled, respectively. The estimation error of any unbiased estimator satisfies*

$$\varepsilon(\mathbf{t}) \geq tr(I(\theta, \mathbf{t})^{-1}) + \sum_{i=1}^{n-1}\sum_{j=1}^{n-1} I(\theta, \mathbf{t})^{-1}(i, j), \quad (4)$$

where $I(\theta, \mathbf{t})$ is the $n-1 \times n-1$ Fisher-Information matrix with entries

$$I_{i,j}(\theta, \mathbf{t}) = \sum_{k=1}^{K}\sum_{\ell=1}^{m_K} \frac{t_k A_k(\ell, i) A_k(\ell, j)(1 - A_k(\ell, n))}{q_{k,\ell}} + \frac{t_k(1 - A_k(\ell, i))(1 - A_k(\ell, j)) A_k(\ell, n)}{q_{k,\ell}}. \quad (5)$$

Proof. We use the Cramer-Rao bound [24], [25] that provides a lower bound on the covariance matrix of any unbiased estimator of an unknown deterministic parameter. Since $\sum_{i=1}^{n} p_i = 1$ it suffices to estimate any $n-1$ of the parameters $\{p_1, p_2, \ldots, p_n\}$. Let these parameters ($\theta = \{\theta_1, \theta_2, \ldots, \theta_{n-1}\}$) be $\{p_1, p_2, \ldots, p_{n-1}\}$. Let $\mathcal{D}_t$ be the event that after t steps, output $z_{k,i}$ from arm k is observed $t_{k,i}$ times, for all $k \in [1, K]$, and $i \in [1, m_k]$.

We evaluate the log likelihood $L(\mathcal{D}_t; \theta)$ of observed data $\mathcal{D}_t$ with respect to θ, We then compute the $n-1 \times n-1$ Fisher

information matrix, $I(\theta, \mathbf{t})$, whose $(i, j)^{th}$ entry is given by $-\mathbb{E}\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j} L(\mathcal{D}_t; \theta)|t_1, \ldots, t_K\right]$. For our problem, we obtain a closed form expression of $I_{i,j}(\theta, \mathbf{t})$ given in (5).

The Cramer-Rao lower bound on covariance matrix of θ for any unbiased estimator is then given by $I(\theta, \mathbf{t})^{-1}$. Our objective is to minimize $\sum_{i=1}^{n} \text{Var}[\tilde{p}_i]$, which can be bounded as

$$\sum_{j=1}^{n} \text{Var}[\tilde{p}_j] = \sum_{j=1}^{n-1} \text{Var}[\tilde{p}_j] + \text{Var}[\tilde{p}_n], \quad (6)$$

$$\geq tr(I(\theta, \mathbf{t})^{-1}) + \text{Var}\left[1 - \sum_{j=1}^{n-1} \tilde{p}_j\right], \quad (7)$$

$$= tr(I(\theta, \mathbf{t})^{-1}) + \sum_{i=1}^{n-1}\sum_{j=1}^{n-1} \text{Cov}(\tilde{p}_i, \tilde{p}_j), \quad (8)$$

$$\geq tr(I(\theta, \mathbf{t})^{-1}) + \sum_{i=1}^{n-1}\sum_{j=1}^{n-1} I(\theta, \mathbf{t})^{-1}(i, j). \quad (9)$$

□

Since the inverse Fisher information matrix, $I(\theta, \mathbf{t})^{-1}$, is also a lower bound on the covariance of any estimator that exhibits local asymptotic normality, therefore, when $t \to \infty$, the result in Theorem 4 also holds for any estimator which is asymptotically normal locally or exhibits asymptotical minimaxity. We now state the lower bound on estimation error for biased estimator with bias $b(\theta)$.

Theorem 5 (Lower bound for any estimator with given bias). *Let terms $\mathbf{t}$ and $I(\theta, \mathbf{t})$ be defined as in the statement of Theorem 4. The estimation error of any biased estimator with bias $b(\theta)$ satisfies*

$$\epsilon(\mathbf{t}) \geq tr(\bar{I}(\theta, \mathbf{t})^{-1}) + \sum_{i=1}^{n-1}\sum_{j=1}^{n-1} \bar{I}(\theta, \mathbf{t})^{-1}(i, j) + ||b(\theta)||_2^2,$$

where $\bar{I}(\theta, \mathbf{t}) = \frac{\partial(\theta - b(\theta))}{\partial\theta} I(\theta, \mathbf{t})^{-1}\left(\frac{\partial(\theta - b(\theta))}{\partial\theta}\right)^{-1}$.

Proof. The proof follows from the fact that Cramer-Rao lower bound on the covariance matrix of θ for any biased estimator with bias $b(\theta)$ is given by $\bar{I}(\theta, \mathbf{t})^{-1}$, with $\bar{I}(\theta, \mathbf{t}) = \frac{\partial(\theta - b(\theta))}{\partial\theta} I(\theta, \mathbf{t})^{-1}\left(\frac{\partial(\theta - b(\theta))}{\partial\theta}\right)^{-1}$. We can evaluate the $\text{Var}(\theta)$ as in proof of Theorem 4. Using this with the fact that $\text{MSE}(\theta) = \text{Var}(\theta) + ||b(\theta)||_2^2$ gives us Theorem 5. □

B. Orderwise Achievability

In section IV-A, we showed that $\epsilon(t) = \Omega\left(\frac{1}{t}\right)$. We now show that this lower bound is achievable if $rank(A) = n$ by analyzing the estimation error of RRPULL+PIEST algorithm. The RRPULL+PIEST pulls arms in a round-robin manner and uses the pseudo inverse of the matrix A to produce estimate $\tilde{P}_X(t)$ at each time step t. Formal description of RRPULL+PIEST is given in Algorithm 1.

Theorem 6 (Order-wise Achievability). *It is possible to achieve estimation error of $\text{O}\left(\frac{1}{t}\right)$ if $rank(A) = n$.*

Algorithm 1 RRPULL + PIEST

1: **Input:** $\{x_1, x_2, \ldots, x_n\}$, Functions $\{g_1, g_2 \ldots g_K\}$ where $g_i : \{x_1, x_2, \ldots, x_n\} \rightarrow \mathbb{R}$. Total number of steps, T.
2: **Initialize:** $t_k = 0 \; \forall k$. $t_{k,i} = 0, \forall i, k$. $\tilde{p}_j(0) = \frac{1}{n}, \forall j$.
3: **for** $t = 1 : T$ **do**
4:     $c_t = \mod(t, K) + 1$
5:     Pull arm c_t, observe output y_t
6:     $t_k = t_k + 1$
7:     **if** $y_t = z_{k,i}$ **then**
8:        $t_{k,i} = t_{k,i} + 1$
9:     **end if**
10:    $\hat{q}_{k,i} = \frac{t_{k,i}}{t_k} \; \forall i, k$.
11:    Obtain estimates $\tilde{p}_j(t)$ as $\tilde{P}_X = A^+ \hat{Q}$.
12: **end for**

Proof of Theorem 6. In order to show achievability we consider the RRPULL + PIEST algorithm that pulls arm in a round-robin manner due to which each arm is pulled $\frac{t}{K}$ times in t steps. For each $k = 1, \ldots, K$ and $i = 1, \ldots, m_k$ let $\hat{q}_{k,i}(t) = \frac{t_{k,i}}{t/K}$ be the estimate for $q_{k,i}$. From these estimates, we can generate estimates $\tilde{P}_X = [\tilde{p}_1(t), \ldots, \tilde{p}_n(t)]^\mathsf{T}$ by solving the system of equations described by (2). More precisely, with $\hat{Q} = [\hat{q}_{1,1}(t), \ldots, \hat{q}_{1,m_1}(t), \ldots \hat{q}_{K,m_K}(t)]^\mathsf{T}$, we can solve

$$A\tilde{P}_X = \hat{Q}. \tag{10}$$

First, we show that the estimates $\tilde{P}_X(t)$ are unbiased. Let Q be the list of true probabilities of observations, i.e., $Q = [q_{1,1}, \ldots, q_{1,m_1}, q_{2,1}, \ldots, q_{2,m_2} \ldots q_{K,m_K}]^\mathsf{T}$. Observe that the length of Q is $m_1 + m_2 + \cdots + m_K$. The solution of (10) is given by $\tilde{P}_X = A^+ \hat{Q}$, where A^+ is the pseudoinverse or the Moore-Penrose inverse of the matrix A. Thus, we get

$$\mathbb{E}\left[\tilde{P}_X\right] = \mathbb{E}\left[A^+ \hat{Q}\right] = A^+ \mathbb{E}\left[\hat{Q}\right] = A^+ Q,$$

upon using the fact that the estimates $\hat{q}_{k,i}(t) = \frac{t_{k,i}}{t/K}$ are unbiased. Here $t_{k,i}$ denotes the number of times i^{th} output of arm k, i.e $z_{k,i}$, is observed. The desired result $\mathbb{E}\left[\tilde{P}_X\right] = P_X$ is now established as we note that $A^+ Q = P_X$ in view of (2).

Next, we derive a bound on the estimation error $\varepsilon(t)$. It is easy to see that the variance of each empirical estimator $\hat{q}_{k,i}(t) = \frac{t_{k,i}}{t/K}$ is $\mathrm{O}\left(\frac{K}{t}\right)$. With $m = m_1 + \ldots + m_K$ denoting the number of rows in A, we then get

$$\varepsilon(t) = \sum_{j=1}^{n} \mathrm{Var}\left[\tilde{p}_j(t)\right] \tag{11}$$

$$= \sum_{j=1}^{n} \mathrm{Var}\left[\sum_{h=1}^{m} A^+(j,h)\hat{Q}(h)\right] \tag{12}$$

$$\leq \sum_{j=1}^{n} \sum_{h=1}^{m} \left(A^+(j,h)\right)^2 \mathrm{Var}\left[\hat{Q}(h)\right] \tag{13}$$
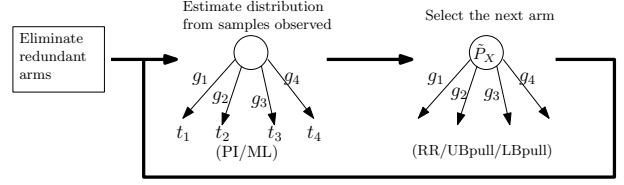

Fig. 4: *The design of algorithm has two key components: i) Estimating P_X from the samples observed which can be done by the maximum likelihood (ML) or pseudoinverse (PI) estimation schemes ii) Choosing the next arm which can be done in a Round-Robin(RR) manner or by using the* UBPULL, LBPULL *strategies.*

$$= \sum_{j=1}^{n} \sum_{k=1}^{K} \sum_{i=1}^{m_k} \left(A^+(j, s+i)\right)^2 \frac{q_{k,i}(1 - q_{k,i})}{t_k} \tag{14}$$

$$= \sum_{j=1}^{n} \sum_{k=1}^{K} \sum_{i=1}^{m_k} \left(A^+(j, s+i)\right)^2 \frac{K q_{k,i}(1 - q_{k,i})}{t} \tag{15}$$

$$= \mathrm{O}\left(\frac{1}{t}\right) \tag{16}$$

where $s = \sum_{\ell=1}^{k-1} m_\ell$. The inequality follows from the fact that elements in $\hat{Q}$ are negatively correlated since $\sum_{i=1}^{m_k} \hat{q}_{k,i} = 1$ for each $k = 1, \ldots, K$. $\qquad\square$

V. PROPOSED SEQUENTIAL DISTRIBUTION LEARNING ALGORITHMS

The design of an algorithm to minimize the estimation error can be divided into two parts: 1) producing the estimate of the distribution $\tilde{P}_X(t)$ based on the samples observed till step t, and 2) deciding which arm to pull at each time t. In Section V-A and Section V-B, we describe these two parts. Algorithm 2 and Algorithm 3 describes our proposed algorithms.

A. Combining observations to estimate P_X

We present a method to estimate P_X given $\mathbf{t} = [t_1, \ldots, t_K]^\mathsf{T}$, where t_k is the number of times arm g_k is pulled until time t.

In the RRPULL+PIEST Algorithm, the estimate of P_X was obtained using the Moore-Penrose inverse and the empirical probabilities of the observed output, $\hat{Q}$. A drawback of this estimation scheme is that it does not account for the number of times each arm is pulled. Motivated by this we propose the use of Maximum Likelihood Estimator for estimating P_X, which takes into account the number of times each arm is pulled to produce estimated probabilities.

Recall that we defined $t_{k,i}$ as the number of times i^{th} output from arm k, i.e., $z_{k,i}$, is observed. Let $\tilde{q}_{k,i}(t)$ be the probability of observing output $z_{k,i}$ under the probability distribution $\tilde{P}_X(t) = [\tilde{p}_1(t), \tilde{p}_2(t), \ldots, \tilde{p}_n(t)]^\mathsf{T}$. The log likelihood of $\mathcal{D}_t$ with respect to the probability distribution $\tilde{p}(t)$ is given by

$$L(\mathcal{D}_t; \tilde{P}_X(t)) = \sum_{k=1}^{K} \sum_{i=1}^{m_k} (t_{k,i} + 1) \log(\tilde{q}_{k,i}(t)). \tag{17}$$

where, $\tilde{q}_{k,i} = \sum_{j=1}^{n} A_k(i,j)\tilde{p}_j$. Note that we smooth the log-likelihood by using $t_{k,i} + 1$ instead of $t_{k,i}$. In order to obtain the maximum likelihood estimate of $\tilde{p}(t)$, we take the derivative of $L(\mathcal{D}_t; \tilde{p}(t))$ and equate it to zero under the constraint $\sum_{i=1}^{n} \tilde{p}_i(t) = 1$. This provides us a set of equations described by

$$\tilde{p}_j(t) = \frac{1}{t} \sum_{k=1}^{K} \sum_{i=1}^{m_k} (t_{k,i} + 1) \frac{A_k(i,j)\tilde{p}_j(t)}{\tilde{q}_{k,i}(t)}, \quad j = 1, 2, \ldots, n. \tag{18}$$

Observe that these set of equations are in the form of $x = f(x)$ and thus can be solved numerically by finding a fixed point using fixed point iteration method [26]. Since the log likelihood function is concave in $\tilde{p}(t)$, the solution from the set of equations described above maximizes the log likelihood function. It is known that the Maximum Likelihood Estimate $\hat{\theta}$ of a parameter θ behaves as $N(\theta, I(\theta)^{-1})$ asymptotically, where $I(\theta)$ is the Fisher Information matrix; here $N(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2. This means that MLE estimator is asymptotically consistent and belongs to the class of asymptotically normal estimator. Therefore, the lower bound in Theorem 4 holds for MLE estimator and it achieves the stated lower bound asymptotically.

B. Deciding which arm to pull

Section V-A described the Maximum Likelihood estimation approach to estimate $\tilde{P}_X(t)$ from the observations till time step t. In this section, we focus on the strategy to pull arm at step $t+1$ given observations till time step t. Although the round-robin arm-pulling strategy used in Algorithm 1 achieves order-wise optimal error (Theorem 6), it has two key drawbacks. Firstly, it is agnostic to the functions $g_k(X)$, and thus even redundant arms will be pulled t/K times. Secondly, it does not consider the distribution estimate $\tilde{P}_X(t)$ when deciding which arm to pull. We now propose an arm-pulling strategy that addresses these shortcomings. The first part of our algorithm involves removal of redundant arms. In the second part we define two strategies, namely UBPULL and LBPULL that can be used to choose an arm in each step.

Removing redundant arms. For each pair of arms r, s evaluate $rank(A_r), rank(A_s)$ and $rank(B)$, where $B = [A_r^\intercal, A_s^\intercal]^\intercal$. If $rank(B) = rank(A_r) > rank(A_s)$ remove arm s. If $rank(B) = rank(A_r) = rank(A_s)$ remove any one of r or s uniformly at random. This leaves us with a new matrix $\bar{A}$ with reduced number of rows (as some arms are removed). Without loss of generality, from now onwards we assume that matrix A does not contain any redundant arm.

The UBPULL strategy to choose the next arm. If we had an analytic expression for estimation error $\varepsilon(t)$ at each step t, we could find the arm that minimizes the estimation error. However, in the absence of an invertible arm, it is hard to obtain an analytic expression of ε, due to which we resort to a heuristic approach. In equation (14) we see an upper bound on estimation error for RRPULL+PIEST algorithm. An approach towards choosing arm could be to minimize this upper bound on the estimation error. However since true

---

Algorithm 2 UBPULL + MLEST

1: **Input:** $\{x_1, x_2, \ldots, x_n\}$, Functions $\{g_1, g_2 \ldots g_K\}$ where $g_i : \{x_1, x_2, \ldots, x_n\} \to \mathbb{R}$. Total number of steps, T.
2: **Initialize:** $t_{k,i} = 0, \forall i, k.\ \tilde{p}_j(0) = \frac{1}{n}, \forall j$.
3: Eliminate Redundant Arms
4: **for** $t = 1 : T$ **do**
5: $\quad c_t = \arg\max_k U(\tilde{P}_X(t), \mathbf{t}) - U(\tilde{P}_X(t), \bar{\mathbf{t}}^{(k)})$
6: $\quad$ Pull arm c_t, observe output y_t
7: $\quad$ **if** $y_t = z_{k,i}$ **then**
8: $\quad\quad t_{k,i} = t_{k,i} + 1$
9: $\quad$ **end if**
10: $\quad$ Obtain estimates $\tilde{p}_j(t)$ by obtaining fixed point solution of the set of equations described by

$$\tilde{p}_j(t) = \frac{1}{t} \sum_{k=1}^{K} \sum_{i=1}^{m_k} (t_{k,i} + 1) \frac{A_k(i,j)\tilde{p}_j(t)}{\tilde{q}_{k,i}(t)},$$

$\quad$ for $j = 1, 2, \ldots, n$.
11: **end for**

---

probability distribution P_X is unknown, we can obtain an estimate of this upper bound as

$$U(\tilde{P}_X(t), \mathbf{t}) = \sum_{j=1}^{n} \sum_{k=1}^{K} \sum_{i=1}^{m_k} \Bigg(\big(A^+(j, s+i)\big)^2 \times \frac{\tilde{q}_{k,i}(t)(1 - \tilde{q}_{k,i}(t))}{t_k} \Bigg), \tag{19}$$

where, $s = \sum_{\ell=1}^{k-1} m_\ell$.

Following this idea, we propose a UBPULL decision scheme which makes use of the observations made till time step t to select arm c_{t+1} at time step $t + 1$. The UBPULL scheme selects an arm c_{t+1}, if pulling c_{t+1} would result in maximum decrease of $U(\tilde{P}_X(t), \mathbf{t})$. More formally, we choose c_{t+1} that maximizes

$$c_{t+1} = \arg\max_k U(\tilde{P}_X(t), \mathbf{t}) - U(\tilde{P}_X(t), \bar{\mathbf{t}}^{(k)}), \tag{20}$$

with ties broken uniformly at random. Here $\bar{\mathbf{t}}^{(k)} = \mathbf{t} + \mathbf{e}^{(k)}$, with $\mathbf{e}^{(k)}$ representing a K length column vector with $e_i^{(k)} = 0 \ \forall i \neq k$ and $e_k^{(k)} = 1$. This results in $\mathbf{t}^{(k)} = [t_1, t_2, \ldots, t_k + 1, \ldots t_K]^\intercal$.

The UBPULL decision scheme along with the maximum likelihood estimation scheme proposed in Section V-A completes the design of UBPULL+MLEST algorithm. A formal description of UBPULL+MLEST is presented in Algorithm 2.

Theorem 7. *Algorithm 2 does asymptotically consistent estimation whenever $rank(A) = n$.*

Proof. In order to show Theorem 7, we first show that under Algorithm 2 each non-redundant arm is pulled infinitely many times as $t \to \infty$. More formally, for each non-redundant arm k, $t_k \to \infty$ as $t \to \infty$.

Observe that the next arm is selected as

$$c_{t+1} = \arg\max_k \zeta_k \left(\frac{1}{t_k} - \frac{1}{t_k+1} \right),$$

with,

$$\zeta_k = \sum_{i=1}^{m_k} \left(\sum_{j=1}^{n} (A^+(j,i+s))^2 \right) \tilde{q}_{k,i}(t)(1 - \tilde{q}_{k,i}(t)),$$

where $s = \sum_{\ell=1}^{k-1} m_\ell$. We have that $\sum_{i=1}^{n}(\tilde{A}^+(j,i))^2 > 0$ $\forall i$ and from Lemma 1 (See Appendix), $0 < \tilde{q}_{k,i} < 1$ $\forall k, i$. This results in $\zeta_k > 0$ $\forall k$.

Let us assume that a non-redundant arm k is pulled only $t_k = \text{O}(1)$ times in a total of t pulls, where $t \to \infty$. Any other arm s can only be pulled if $\zeta_s \left(\frac{1}{t_s} - \frac{1}{t_s+1} \right) > \zeta_k \left(\frac{1}{t_k} - \frac{1}{t_k+1} \right)$. Due to this $t_s < \sqrt{\frac{\zeta_s}{\zeta_k}}(t_k+1)$ $\forall s \neq k$. Therefore $t \leq \sum_{s=1}^{K} \sqrt{\frac{\zeta_s}{\zeta_k}}(t_k+1) = \text{O}(1)$, as we assumed $t_k = \text{O}(1)$. However, this contradicts our assumption that $t \to \infty$. Therefore as $t \to \infty$, $t_k \to \infty$.

As each non-redundant arm is pulled infinitely many times, we see that each element in the Fisher information matrix (5) approaches infinity. Since the variance in the maximum likelihood estimator approaches the inverse Fisher information asymptotically, maximum likelihood estimator will asymptotically achieve the bound in Theorem 4. Since each element in Fisher information approaches infinity, the bound in Theorem 4 approaches zero and consequently $\epsilon(t) \to 0$ as $t \to \infty$ for the ML estimator. Therefore the statement in Theorem 7 holds true. □

The LBPULL strategy to choose next arm. In the UBPULL + MLEST algorithm, we used (19) as a metric for choosing arm at each step. An alternative metric could be the lower bound on estimation error stated in Theorem 4. Since true probability distribution P_X is unknown, we estimate the expression of Theorem 4 as

$$B(\tilde{P}_X(t), \mathbf{t}) = tr(\tilde{I}(\theta, \mathbf{t})^{-1}) + \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \tilde{I}(\theta, \mathbf{t})^{-1}(i,j),$$

where

$$\tilde{I}_{i,j}(\theta, \mathbf{t}) = \sum_{k=1}^{K} \sum_{\ell=1}^{m_K} \frac{t_k A_k(\ell,i) A_k(\ell,j)(1 - A_k(\ell,n))}{\tilde{q}_{k,\ell}(t)} + \frac{t_k(1 - A_k(\ell,i))(1 - A_k(\ell,j)) A_k(\ell,n)}{\tilde{q}_{k,\ell}(t)}. \quad (21)$$

Based on this idea, we propose a LBPULL decision scheme which chooses arm c_{t+1} at round $t+1$ if pulling c_{t+1} maximizes the decrease in $B(\tilde{P}_X(t), \mathbf{t})$. More formally, LBPULL chooses c_{t+1} that maximizes

$$c_{t+1} = \arg\max_k B(\tilde{P}_X(t), \mathbf{t}) - B(\tilde{P}_X(t), \bar{\mathbf{t}}^{(k)}), \quad (22)$$

with ties broken uniformly at random. As defined earlier, $\bar{\mathbf{t}}^{(k)} = \mathbf{t} + \mathbf{e}^{(k)}$.

Algorithm 3 LBPULL + MLEST

1: **Input:** $\{x_1, x_2, \ldots, x_n\}$, Functions $\{g_1, g_2 \ldots g_K\}$ where $g_i : \{x_1, x_2, \ldots, x_n\} \to \mathbb{R}$. Total number of steps, T.
2: **Initialize:** $t_{k,i} = 0, \forall i, k$. $\tilde{p}_j(0) = \frac{1}{n}, \forall j$.
3: Eliminate Redundant Arms
4: **for** $t = 1 : T$ **do**
5:   $c_t = \arg\max_k B(\tilde{P}_X(t-1), \mathbf{t}) - B(\tilde{P}_X(t-1), \bar{\mathbf{t}}^{(k)})$
6:   Pull arm c_t, observe output y_t
7:   **if** $y_t = z_{k,i}$ **then**
8:     $t_{k,i} = t_{k,i} + 1$
9:   **end if**
10:   Obtain estimates $\tilde{p}_j(t)$ by obtaining fixed point solution of the set of equations described by

$$\tilde{p}_j(t) = \frac{1}{t} \sum_{k=1}^{K} \sum_{i=1}^{m_k} (t_{k,i}+1) \frac{A_k(i,j)\tilde{p}_j(t)}{\tilde{q}_{k,i}(t)},$$

  for $j = 1, 2, \ldots, n$.
11: **end for**

This LBPULL decision scheme combined with the maximum likelihood estimation scheme completes the design of LBPULL+ MLEST algorithm. A Formal description is presented in Algorithm 3. We conjecture that the LBPULL+MLEST scheme also achieves asymptotically consistent estimation whenever possible, we leave the proof as a future work.

VI. SIMULATION RESULTS

In this section, we demonstrate the performance of our algorithm under different scenarios. We compare the estimation error of our algorithm with the Cramér-Rao lower bound evaluated in Section IV-A. Recall that Cramér-Rao bound gives a lower bound on the estimation error given the choice of $\{t_1, t_2, \ldots, t_K\}$. To evaluate the lower bound after a total of t time slots, we find the Cramér-Rao bound for all combinations of $\{\alpha_1 t, \alpha_2 t, \ldots, \alpha_K t\}$ where $\sum_{i=1}^{K} \alpha_i = 1$, and take the minimum over all such combinations. We iterate $\{\alpha_1, \alpha_2, \ldots, \alpha_K\}$ for all possible values between 0 to 1 with a precision of 0.001. Note that the existence of an algorithm that achieves the Cramér-Rao lower bound is not guaranteed.

Fig. 6 shows the results of our experiment for the example considered in Fig. 5. The experiment was repeated 1000 times and we report the average estimation error in the plot. In Table I we report the average number of pulls needed by each algorithm to achieve an error of 10^{-3}. For comparison purposes we included RRpull+MLest algorithm, which pulls arms in a round-robin manner and produces estimate using maximum likelihood estimation. As evident, the proposed UBPULL+MLEST and LBPULL+MLEST algorithms outperform the RRPULL+PIEST and RRPULL+MLEST algorithms in this scenario. While RRPULL+PIEST algorithm pulls each of the arms equal number of times, the proposed algorithms adapt according to shape of function and the probability

distribution estimates to pull each arm different number of times. This is one of the key reason behind the successful performance of UBPULL+MLEST and LBPULL+MLEST. This effect is illustrated in Fig. 7, where we report the average number of times each arm was pulled over 1000 experiments. The combination $\{\alpha_1 t, \alpha_2 t, \ldots, \alpha_K t\}$ resulting in the minimum Cramér-Rao bound for $t = 1000$ is displayed in Figures Fig. 6 and Fig. 5 as the "CRLB config." We see that the number of times each arm is pulled in LBPULL+MLEST algorithm is very close to these numbers. Given that maximum likelihood estimator is known to achieve Cramér-Rao bound asymptotically, this suggests that the asymptotic performance of the LBPULL+MLEST algorithm will be close to optimal.

We now demonstrate why an *active* learning framework, where the samples are obtained sequentially *based on the current estimate* of P_X, is necessary in order to achieve the best performance (e.g., to minimize the error) for the problem under consideration. This is primarily because of the fact that given a total number of available pulls, the optimal number of times that each arm needs to be pulled (in order to minimize the error) depends not only on the functions themselves (or, the sample generation matrix A), but also the probability distribution that the algorithm is trying to estimate. It is for this reason that we need an active learning approach where the current estimate of P_X (based on prior samples) is factored into deciding which of the available functions the next sample should come from.

In order to demonstrate the need for an active learning framework, we revisit the case considered in Fig. 5 with the functions $g_1(X), g_2(X)$ and $g_3(X)$ kept the same. This time, we assume that the underlying probability distribution is changed from $P_X = [0.05, 0.1, 0.1, 0.2, 0.2, 0.25, 0.1]$ to $P_X = [0.4, 0.25, 0.2, 0.05, 0.025, 0.025, 0.05]$. In the former case, we had seen that the lowest error is achieved when functions g_1, g_2, g_3 are sampled at a relative fraction of 0.104, 0.317, and 0.579, respectively. In other words, if it is indeed the case that $P_X = [0.05, 0.1, 0.1, 0.2, 0.2, 0.25, 0.1]$, an algorithm that chooses g_1, g_2, g_3 with probabilities 0.104, 0.317, and 0.579, respectively, at each step (independently) would be the optimal in learning this distribution. It might be tempting to think that this *baseline* algorithm would do well even if the underlying probability distribution is different, as long as the functions remain the same. However, under the modified probability distribution $P_X = [0.4, 0.25, 0.2, 0.05, 0.025, 0.025, 0.05]$, we observe that this baseline algorithm achieves an error which is 22.8% and 31.4% more than UBPULL+MLEST and LBPULL+MLEST, respectively. This highlights the fact that an algorithm considering only the shape of function may not perform well in all cases and indeed an active learning algorithm that uses the estimates $\hat{p}(t)$ at every step to make the next decision is necessary to tackle this problem (as done by UBPULL+MLEST and LBPULL+MLEST). Table II illustrates this insight in terms of number of samples required to achieve an error of 10^{-3} for UBPULL+MLEST, LBPULL+MLEST, Baseline and RRPULL+PIEST algorithms respectively.
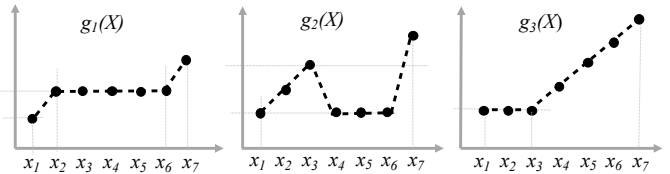

Fig. 5: *Example of a set of functions* $\{g_1, g_2, \ldots, g_K\}$. *Probability distribution* $P_X = [0.05, 0.1, 0.1, 0.2, 0.2, 0.25, 0.1]$.
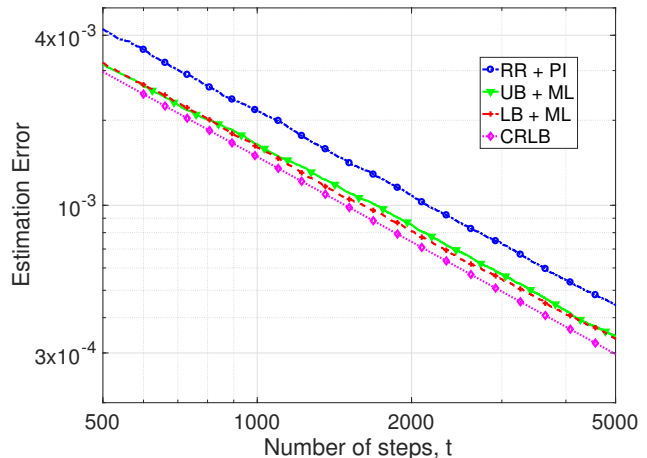

Fig. 6: *Comparison of our policies against the* RRPULL+PIEST *algorithm for the example in Fig. 5*

VII. CONCLUDING REMARKS

We consider the problem of learning the distribution P_X of a hidden random variable X, using indirect samples from the functions $g_1(X), g_2(X), \ldots g_K(X)$, referred to as *arms*. The samples are obtained in a sequential fashion, by choosing one of the K arms in each time slot. Several applications where we wish to infer properties of a hidden random phenomenon

Algorithm	Avg. pulls needed
LBPULL+MLEST	2019.1
UBPULL+MLEST	2126.8
RRPULL+MLEST	2610.7
RRPULL+PIEST	2808.4

TABLE I: *Average number of pulls needed to get an estimation error of* 10^{-3} *for the example in Fig. 5.*

Algorithm	Avg. pulls needed
LBPULL+MLEST	1945.2
UBPULL+MLEST	2019.2
Baseline	2579.4
RRPULL+PIEST	2347.4

TABLE II: *Average number of pulls needed to get an estimation error of* 10^{-3} *for the functions in Fig. 5 where* $[p_1, p_2, p_3, p_4, p_5, p_6, p_7]$ *is* $[0.4, 0.25, 0.2, 0.05, 0.025, 0.025, 0.05]$.
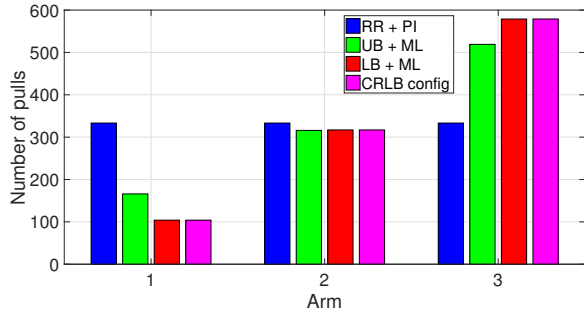
Fig. 7: *Average number of times each arm is pulled in a total of 1000 steps for the example in Fig. 5. Variance in the number of pulls across experiments is small for* LBPULL+MLEST *and* UBPULL+MLEST *algorithms.*
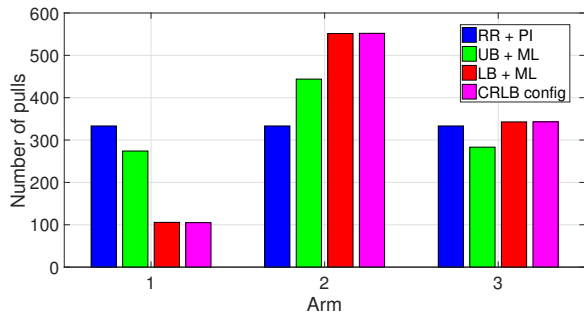

Fig. 8: *Average number of times each arm is pulled in a total of 1000 steps for the functions in Fig. 5 where* $[p_1, p_2, p_3, p_4, p_5, p_6, p_7]$ *is* $[0.4, 0.25, 0.2, 0.05, 0.025, 0.025, 0.05]$. *Variance in the number of pulls across experiments is small for* LBPULL+MLEST *and* UBPULL+MLEST *algorithms.*

using indirect or imprecise observations fit into our framework. We determine conditions for asymptotically consistent estimation of P_X and evaluate bounds on the estimation error. Using insights from this analysis, we propose algorithms to choose arms and combine their samples. Performance of these algorithms is is shown to outperform several intuitive baseline algorithms numerically.

Ongoing work includes obtaining result on asymptotic consistency for LBPULL+MLEST algorithm. Instead of the deterministic functions $g_i(X)$, we also plan to consider random observations Y_i, such that the conditional distribution $p(Y_i|X)$ is known.

REFERENCES

[1] L. G. Valiant, "A theory of the learnable," *Communications of the ACM*, vol. 27, pp. 1134–1142, Nov. 1984.

[2] M. Kearns, Y. Mansour, D. Ron, R. Rubinfeld, R. E. Schapire, and L. Sellie, "On the learnability of discrete distributions," in *Proceedings of the ACM Symposium on Theory of Computing (STOC)*, pp. 273–282, 1994.

[3] C. Daskalakis, I. Diakonikolas, and R. A. Servedio, "Learning k-modal distributions via testing," in *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*, pp. 1371–1385, Society for Industrial and Applied Mathematics, 2012.

[4] S. Kamath, A. Orlitsky, D. Pichapati, and A. T. Suresh, "On learning distributions from their samples," in *Conference on Learning Theory*, pp. 1066–1100, 2015.

[5] Y. Han, J. Jiao, and T. Weissman, "Minimax estimation of discrete distributions under ℓ_1 loss," *IEEE Transactions on Information Theory*, vol. 61, pp. 6343–6354, Nov 2015.

[6] J. Acharya, H. Das, A. Orlitsky, and A. T. Suresh, "A unified maximum likelihood approach for estimating symmetric properties of discrete distributions," in *International Conference on Machine Learning*, pp. 11–21, 2017.

[7] J. Acharya, A. Orlitsky, A. T. Suresh, and H. Tyagi, "The complexity of estimating rényi entropy," in *Proceedings of the twenty-sixth annual ACM-SIAM symposium on Discrete algorithms*, pp. 1855–1869, SIAM, 2014.

[8] Y. Han, J. Jiao, and T. Weissman, "Adaptive estimation of shannon entropy," in *IEEE International Symposium on Information Theory (ISIT)*, pp. 1372–1376, IEEE, 2015.

[9] J. Jiao, H. H. Permuter, L. Zhao, Y.-H. Kim, and T. Weissman, "Universal estimation of directed information," *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6220–6242, 2013.

[10] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

[11] H. Robbins, "Some aspects of the sequential design of experiments," in *Herbert Robbins Selected Papers*, pp. 169–177, Springer, 1985.

[12] H. Chernoff, "Sequential design of experiments," *The Annals of Mathematical Statistics*, vol. 30, no. 3, pp. 755–770, 1959.

[13] M. Naghshvar, T. Javidi, *et al.*, "Active sequential hypothesis testing," *The Annals of Statistics*, vol. 41, no. 6, pp. 2703–2738, 2013.

[14] D. Golovin, A. Krause, and D. Ray, "Near-optimal bayesian active learning with noisy observations," in *Advances in Neural Information Processing Systems*, pp. 766–774, 2010.

[15] S. Javdani, Y. Chen, A. Karbasi, A. Krause, D. Bagnell, and S. S. Srinivasa, "Near optimal bayesian active learning for decision making.," in *AISTATS*, vol. 14, pp. 430–438, 2014.

[16] S. Bubeck, N. Cesa-Bianchi, *et al.*, "Regret analysis of stochastic and nonstochastic multi-armed bandit problems," *Foundations and Trends in Machine Learning*, vol. 5, no. 1, pp. 1–122, 2012.

[17] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine learning*, vol. 47, no. 2-3, pp. 235–256, 2002.

[18] S. Agrawal and N. Goyal, "Analysis of thompson sampling for the multi-armed bandit problem," in *Conference on Learning Theory*, pp. 39–1, 2012.

[19] T. L. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules," *Advances in applied mathematics*, vol. 6, no. 1, pp. 4–22, 1985.

[20] K. Jamieson and R. Nowak, "Best-arm identification algorithms for multi-armed bandits in the fixed confidence setting," in *Annual Conference on Information Sciences and Systems (CISS)*, pp. 1–6, IEEE, 2014.

[21] J.-Y. Audibert and S. Bubeck, "Best arm identification in multi-armed bandits," in *COLT-23th Conference on Learning Theory-2010*, pp. 13–p, 2010.

[22] E. Kaufmann, O. Cappé, and A. Garivier, "On the complexity of best-arm identification in multi-armed bandit models," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 1–42, 2016.

[23] V. Gabillon, M. Ghavamzadeh, and A. Lazaric, "Best arm identification: A unified approach to fixed budget and fixed confidence," in *Advances in Neural Information Processing Systems*, pp. 3212–3220, 2012.

[24] C. R. Rao, "Information and the accuracy attainable in the estimation of statistical parameters," *Bulletin of the Calcutta Mathematical Society*, vol. 37, pp. 81–91, 1945.

[25] H. Cramér, *Mathematical Methods of Statistics*. Princeton, NJ, USA: Princeton University Press, 1946.

[26] R. L. Burden and J. D. Faires, "2.2 fixed-point iteration," *Numerical Analysis (3rd ed.)*. PWS Publishers. ISBN 0-87150-857-5, 1985.

Lemma 1. *The estimates $\tilde{q}_{k,i}(t)$ of output probabilities $q_{k,i}$ produced by Algorithm 2 are bounded away from 0 and 1, i.e., we have $0 < \tilde{q}_{k,i}(t) < 1 \ \forall i, k$.*

Proof. Consider the following two cases

Case 1: ($t < M$ for some $M < \infty$.) Our proposed estimation scheme for $\tilde{P}_X$ maximizes the log likelihood in (17) at each time step t. Since the term $\log(\tilde{q}_{k,i}(t))$ has a positive coefficient $\forall k, i$, any estimate $\tilde{P}_X$ for which $q_{k,i} = 0$ for any k, i would result in $L(\mathcal{D}_t; \tilde{p}(t)) \to -\infty$. For any other estimate $\tilde{P}_X$, that has $q_{k,i} > 0 \ \forall k, i$ (say $\tilde{p}_j = 1/n \ \forall j$) would lead to $L(\mathcal{D}_t, \tilde{P}_X(t)) > -C$, for some $C < \infty$. This implies that $\tilde{q}_{k,i}(t) > 0 \ \forall k, i$. Moreover, since $\sum_{i=1}^{m_k} \tilde{q}_{k,i}(t) = 1$ $\forall k$, $\tilde{q}_{k,i}(t) < 1 \ \forall k, i$. This holds because $m_k > 1 \ \forall k$, as redundant arms are already removed by Algorithm 2. Hence we have $0 < \tilde{q}_{k,i}(t) < 1$ in this case.

Case 2: ($t \to \infty$). We denote $\mathcal{S}$ as the set of arms for which $t_k \to \infty$. Similarly we denote $\mathcal{S}^c$ as the set of arms for which $t_k = O(1)$. Suppose at time step t estimate $\tilde{P}_X(t)$ is such that $\tilde{q}_{k,i}(t) \to 0$ as $t \to \infty$, for some k, i. By the law of large numbers, we have that $t_{k,i} \to t_k q_{k,i}$ almost surely $\forall k \in S$. The log likelihoood expression of (17) for the estimated distribution $\tilde{P}_X(t)$ will be:

$$L(\mathcal{D}_t; \tilde{P}_X(t)) = \sum_{k=1}^{K} \sum_{i=1}^{m_k} (t_{k,i} + 1) \log(\tilde{q}_{k,i}(t)) \qquad (23)$$

$$= \sum_{k \in \mathcal{S}} t_k \sum_{i=1}^{m_k} q_{k,i} \log(\tilde{q}_{k,i}(t)) + \sum_{k \in \mathcal{S}} \sum_{i=1}^{m_k} \log(\tilde{q}_{k,i}(t)) +$$

$$\sum_{k \in \mathcal{S}^c} \sum_{i=1}^{m_k} (t_{k,i} + 1) \log(\tilde{q}_{k,i}(t)) \qquad (24)$$

We now evaluate the log likelihood expression for the distribution P_X.

$$L(\mathcal{D}_t; P_X) = \sum_{k=1}^{K} \sum_{i=1}^{m_k} (t_{k,i} + 1) \log(q_{k,i}(t)) \qquad (25)$$

$$= \sum_{k \in \mathcal{S}} t_k \sum_{i=1}^{m_k} q_{k,i} \log(q_{k,i}) + \sum_{k \in \mathcal{S}} \sum_{i=1}^{m_k} \log(q_{k,i}) +$$

$$\sum_{k \in \mathcal{S}^c} \sum_{i=1}^{m_k} (t_{k,i} + 1) \log(q_{k,i}) \qquad (26)$$

Now we evaluate the difference in $L(\mathcal{D}_t; P_X)$ and $L(\mathcal{D}_t, \tilde{P}_X(t))$

$$L(\mathcal{D}_t; P_X) - L(\mathcal{D}_t; \tilde{P}_X(t)) = \sum_{k \in \mathcal{S}} t_k \sum_{i=1}^{m_k} q_{k,i} \log\left(\frac{q_{k,i}}{\tilde{q}_{k,i}(t)}\right) +$$

$$\sum_{k \in \mathcal{S}} \sum_{i=1}^{m_k} \log\left(\frac{q_{k,i}}{\tilde{q}_{k,i}(t)}\right) + \sum_{k \in \mathcal{S}^c} \sum_{i=1}^{m_k} (t_{k,i} + 1) \log\left(\frac{q_{k,i}}{\tilde{q}_{k,i}(t)}\right)$$

$$(27)$$

The first term is non-negative as $\sum_{i=1}^{m_k} q_{k,i} \log\left(\frac{q_{k,i}}{\tilde{q}_{k,i}(t)}\right) \geq 0$, as this is the KL-Divergence between the output probability distribution of arm k and the estimated output probability

distribution of arm k. The sum of second and third term is positive as we assumed that at least one $\tilde{q}_{k,i}(t) \to 0$ as $t \to \infty$, whereas $q_{k,i} > 0 \ \forall k, i$ (as in our setup $p_j > 0 \ \forall j$). This gives us that

$$L(\mathcal{D}_t; P_X) - L(\mathcal{D}_t; \tilde{P}_X(t)) > 0. \qquad (28)$$

This implies that an estimated probability distribution $\tilde{P}_X$ having $\tilde{q}_{k,i}(t) \to 0$ as $t \to \infty$ cannot maximize the likelihood at time step t. This results in the fact that $\tilde{q}_{k,i}(t) > 0 \ \forall k, i$. Since $\sum_{i=1}^{m_k} \tilde{q}_{k,i}(t) = 1$, we have that $0 < \tilde{q}_{k,i}(t) < 1$. $\qquad \square$