Proceedings of Machine Learning Research vol 125:1–32, 2020

Locally Private Hypothesis Selection

Sivakanth Gopi SIGOPI@MICROSOFT.COM
Microsoft Research

Gautam Kamath G@CSAIL.MIT.EDU
University of Waterloo

Janardhan Kulkarni JAKUL@MICROSOFT.COM
Microsoft Research

Aleksandar Nikolov ANIKOLOV@CS.TORONTO.EDU
University of Toronto

Zhiwei Steven Wu ZSW@UMN.EDU
University of Minnesota

Huanyu Zhang HZ388@CORNELL.EDU
Cornell University.

Abstract

We initiate the study of hypothesis selection under local differential privacy. Given samples from an unknown probability distribution p and a set of k probability distributions $\mathcal{Q}$, we aim to output, under the constraints of ε-differential privacy, a distribution from $\mathcal{Q}$ whose total variation distance to p is comparable to the best such distribution. This is a generalization of the classic problem of k-wise simple hypothesis testing, which corresponds to when $p \in \mathcal{Q}$, and we wish to identify p. Absent privacy constraints, this problem requires $O(\log k)$ samples from p, and it was recently shown that the same complexity is achievable under (central) differential privacy. However, the naive approach to this problem under local differential privacy would require $\tilde{O}(k^2)$ samples.

We first show that the constraint of local differential privacy incurs an exponential increase in cost: any algorithm for this problem requires at least $\Omega(k)$ samples. Second, for the special case of k-wise simple hypothesis testing, we provide a non-interactive algorithm which nearly matches this bound, requiring $\tilde{O}(k)$ samples. Finally, we provide sequentially interactive algorithms for the general case, requiring $\tilde{O}(k)$ samples and only $O(\log \log k)$ rounds of interactivity. Our algorithms are achieved through a reduction to maximum selection with adversarial comparators, a problem of independent interest for which we initiate study in the parallel setting. For this problem, we provide a family of algorithms for each number of allowed rounds of interaction t, as well as lower bounds showing that they are near-optimal for every t. Notably, our algorithms result in exponential improvements on the round complexity of previous methods.

Keywords: Local differential privacy, hypothesis testing

1. Introduction

Perhaps the most fundamental question in statistics is that of simple hypothesis testing. Given two known distributions p and q, and a dataset generated according to one of these distributions, the goal is to determine which distribution the data came from. The optimal solution to this problem is the likelihood-ratio test, as shown by Neyman and Pearson (1933). This problem can be generalized in

two ways that we consider in this paper. First, rather than just two distributions, one can consider a setting where the goal is to select from a set of k distributions. We refer to this setting as *k-wise simple hypothesis testing*. Furthermore, the data may not have been generated according to *any* distribution from the set of known distributions – instead, the goal is to just select a distribution from the set which is competitive with the best possible (in an appropriate distance measure). This problem is the core object of our study, and we denote it as *hypothesis selection*.

The hypothesis selection problem appears naturally in a number of settings. For instance, we may have a collection of distribution learning algorithms that are effective under different assumptions on the data, but it is unknown which ones hold in advance. Hypothesis selection allows us to simply run all of these algorithms in parallel and pick a good output from these candidate distributions afterwards. More generally, a learning algorithm may first "guess" various parameters of the unknown distribution and for each guess produce a candidate output distribution. Hypothesis selection allows us to pick a final result from this set of candidates. Finally, near-optimal sample complexity bounds can often be derived by enumerating all possibilities within some parametric class of distributions (i.e., a cover) and then applying hypothesis selection with this enumeration as the set of hypotheses (Devroye and Lugosi, 2001).

Classical work (e.g., Yatracos (1985); Devroye and Lugosi (1996, 1997, 2001)) on these problems has shown that, even in the most general setting of hypothesis selection, there are effective algorithms with sample complexity scaling only *logarithmically* in the number of candidate hypotheses. Building on this, there has been significant study into hypothesis selection with additional desiderata, including computational efficiency, robustness, weaker access to hypotheses, and more.

One consideration which has not received significant attention in this setting is that of *data privacy*. The dataset may be comprised of personally sensitive data, including medical records, location history, or salary information, and classical hypothesis selection algorithms may violate the privacy of individuals who provided the data. Motivated by this issue, our goal is to perform our statistical analysis while ensuring that the output does not reveal significant information about any individual datapoint. We will be concerned with the formalization of this principle as *differential privacy* (Dwork et al., 2006), which can be seen as the gold standard for modern data privacy.

We first distinguish between two common definitions of differential privacy. The first is *central differential privacy* (Dwork et al., 2006), in which users transmit their data to a central server without any obfuscation, and the algorithm operates on this dataset with the restriction that its final output must be appropriately privatized. The second is *local differential privacy* (LDP) (Warner, 1965; Evfimievski et al., 2003; Kasiviswanathan et al., 2011), in which users trust no one: each individual privatizes their own data before sending it to the central server. In some sense, LDP places the privacy barrier closer to the users, and as a result, has seen adoption in practice by a number of companies that analyze sensitive user data, including Google (Erlingsson et al., 2014), Microsoft (Ding et al., 2017), and Apple (Differential Privacy Team, Apple, 2017).

Recently, Bun, Kamath, Steinke, and Wu (Bun et al., 2019) showed that in the central model, one can still perform hypothesis selection with sample complexity which scales logarithmically in the number of hypotheses. A priori, it was not clear that this would be possible. Non-privately, one can apply methods which essentially ask "Which of these two distributions fits the data better?" for all $O(k^2)$ pairs of hypotheses. Crucially, one can reuse the same set of $O(\log k)$ samples for all such comparisons (rather than drawing fresh samples for each one), and accuracy can be proved by a Chernoff and union bound style argument. A naive privatization of this method would result in a polynomial dependence on k, due to issues arising from sample reuse and the composition of privacy

losses. Bun et al. (2019) avoid this issue by a careful application of tools from the differential privacy literature (i.e., the exponential mechanism (McSherry and Talwar, 2007)), achieving an $O(\log k)$ sample complexity. However, their method relies upon techniques which are not available in the local model of differential privacy. Indeed, at first glance, it may not be clear how to improve upon an $\tilde{O}(k^2)$ sample complexity in the local model, achieved by simply using a fresh set of samples for each comparison, and using randomized response to privately perform the comparison. This raises the question: what is the sample complexity of hypothesis selection under local differential privacy? Can the problem be solved with a logarithmic dependence of the number of samples on the number of candidate hypotheses? Or do we require a polynomial number of samples?

1.1. Results, Techniques, and Discussion

We assume the reader is familiar with the notion of ε-local differential privacy (ε-LDP) and total variation distance $d_{\mathrm{TV}}(p, q)$; formal definitions appear in Section 2. To describe our results, we more formally define the problems of k-wise simple hypothesis testing and hypothesis selection.

Definition 1 *Suppose we are given a set of n data points $X_1, \ldots, X_n$, which are sampled i.i.d. from some (unknown) distribution p, and a set of k distributions $\mathcal{Q} = \{q_1, \ldots, q_k\}$. The goal is to output a distribution $\hat{q} \in \mathcal{Q}$ such that $d_{\mathrm{TV}}(p, \hat{q}) \leq c \min_{q^* \in \mathcal{Q}} d_{\mathrm{TV}}(p, q^*) + \alpha$, for some $c = c(\alpha, k)$.*

We refer to the value of $c(\alpha, k)$ as the agnostic approximation factor. *If $c(\alpha, k)$ is an absolute constant, then we denote this problem as* hypothesis selection. *If $c(\alpha, k)$ grows with k and $\frac{1}{\alpha}$, we refer to this problem as* weak hypothesis selection. *If we require that $p \in \mathcal{Q}$, that $\min_{i \neq j} d_{\mathrm{TV}}(q_i, q_j) \geq \alpha$, and that the algorithm must correctly identify p, then we denote this problem as k-*wise simple hypothesis testing.

We show k-wise simple hypothesis testing (and thus, hypothesis selection) requires $\Omega(k)$ samples.

Theorem 2 *Let $\varepsilon \in (0, 1)$. Suppose M is an ε-LDP protocol that solves the k-wise simple hypothesis testing problem with probability at least $1/3$ when given n samples from some distribution $p \in \mathcal{Q}$, for any set $\mathcal{Q} = \{q_1, \ldots, q_k\}$ such that $\min_{i \neq j} d_{\mathrm{TV}}(q_i, q_j) \geq \alpha$. Then $n = \Omega\left(\frac{k}{\alpha^2 \varepsilon^2}\right)$.*

Thus the cost of hypothesis testing is exponentially larger under LDP than under central differential privacy (i.e., $\Omega(k)$ versus $O(\log k)$), even when the LDP protocol is fully interactive. The construction used to prove this lower bound is the problem of 1-sparse mean estimation, previously identified as a problem of interest by Duchi, Jordan, and Wainwright (Duchi et al., 2013, 2017). The lower bound follows from results in Duchi and Rogers (2019). Given the construction, our result can be seen as a translation of existing results, so the details are given in Appendix C.

With a lower bound of $\Omega(k)$ samples, and a naive upper bound of $\tilde{O}(k^2)$ samples, the problem remains to identify the correct sample complexity. We give two algorithms which require $\tilde{O}(k)$ samples, nearly matching this lower bound. The first is for the special case of k-wise simple hypothesis testing, and is a non-interactive protocol – all users only send a message to the curator once, which does not depend on the messages sent by other users. The second solves the more general problem of hypothesis selection, but requires sequential interactivity (albeit only $O(\log \log k)$ rounds of interaction): users still only send a message to the curator once, but the curator may request different types of messages from later users based on the messages sent by earlier users. Less interaction in an protocol is generally preferred, and the role and power of interactivity in local differential privacy

is one of the most significant questions in the area (see, e.g. Kasiviswanathan et al. (2011); Joseph et al. (2019); Daniely and Feldman (2019); Duchi and Rogers (2019); Joseph et al. (2020)).

Our first algorithmic result gives a non-interactive mechanism with $\tilde{O}(k)$ sample complexity for sufficiently well separated instances. Define $\beta := \min_{q \in \mathcal{Q}} d_{\mathrm{TV}}(p, q)$.

Theorem 3 *For every $\varepsilon \in [0, 1)$, there is a non-interactive ε-LDP algorithm that with probability at least $1 - 1/k^2$ outputs a distribution $\hat{q} \in \mathcal{Q}$ such that $d_{\mathrm{TV}}(p, \hat{q}) \leq \alpha$, if the number of samples $n \gg k(\log k)^3/(\alpha^4\varepsilon^2)$ and $\beta \ll \alpha^2/\log k$.*[1]

We prove the theorem in Section 3. While somewhat more general, the above theorem immediately gives a non-interactive $\tilde{O}(k)$-sample algorithm for the important special case of LDP k-wise simple hypothesis testing.

Corollary 4 *Suppose our instance of hypothesis testing is such that $p \in \mathcal{Q}$ and all distributions in $\mathcal{Q}$ are $\Omega(\alpha)$-far from each other in total variation distance. For $\varepsilon \in (0, 1]$, there exists a non-interactive ε-LDP algorithm which identifies p with high probability, given $n = O\left(\frac{k \log^3 k}{\alpha^4 \varepsilon^2}\right)$ samples.*

Our algorithm is based on a noised log-likelihood test, though significant massaging and manipulation of the problem instance is required to achieve an acceptable sample complexity. In our algorithm, the users are divided into k groups. Each user in the i^{th} group sends the log-likelihood (with some Laplace noise added for privacy) of observing the sample given to the user if the true distribution was q_i. The log-likelihoods from all the users in the i^{th} group are aggregated and the most likely distribution is output. Alternatively, we can also think of our algorithm as using the samples from the i^{th} group to estimate KL-divergences between the unknown distribution and q_i and finally outputting the closest distribution. For this approach to work, we need all the log-likelihoods to be bounded. We achieve this by a *flattening lemma* which makes all the distributions close to uniform, while preserving their total variation distances. Moreover, this flattening can be implemented locally by the users transforming their samples from the original distribution. We believe that our flattening lemma may have applications in other DP problems.

Our second algorithmic result is a $O(\log \log k)$-round sequentially interactive $\tilde{O}(k)$-sample algorithm for LDP hypothesis selection.

Corollary 5 (Informal version of Corollary 25) *Suppose we are given n samples from an unknown distribution p and a set of descriptions of k distributions $\mathcal{Q}$. There exists an algorithm which identifies a distribution $\hat{q} \in \mathcal{Q}$, such that $d_{\mathrm{TV}}(p, \hat{q}) \leq 27 \min_{q^* \in \mathcal{Q}} d_{\mathrm{TV}}(p, q^*) + O(\alpha)$ with probability $9/10$. The algorithm is ε-LDP, requires $O(\log \log k)$ rounds of sequential interactivity, and $n = O\left(\frac{k \log k \log \log k}{\alpha^2 \varepsilon^2}\right)$ samples.*

The k-wise simple hypothesis testing and hypothesis selection problems can also be studied in the Statistical Queries (SQ) model of Kearns (1998). In this model, rather than being given samples from a distribution p, the algorithm can ask queries specified by bounded functions ϕ, and get a

1. We use $A \ll B$ to denote that $A \leq cB$ for some sufficiently small constant $c > 0$. Similarly we use $A \gg B$ to denote that $A \geq CB$ for some sufficiently large constant $C > 0$. $A \lesssim B$ is used interchangeably with $A = O(B)$. Similarly $A \gtrsim B$ is used interchangeably with $A = \Omega(B)$.

(possibly adversarial) additive τ-approximation to the expectation of ϕ under p, where the parameter τ is usually called the tolerance. For distributional problems, Kasiviswanathan et al. (2011) showed that sample complexity in the LDP model is equivalent up to polynomial factors to complexity in the SQ model, measured in terms of the number of queries and the inverse tolerance $\frac{1}{\tau}$. In particular, this connection and our lower bound in Theorem 2 imply that k-wise simple hypothesis testing in the SQ model requires that either the number of queries or $\frac{1}{\tau}$ be polynomial in k. Because of the polynomial loss, however, our precise study of the sample complexity of these problems does not immediately translate to the SQ model. We remark that both the 1-round algorithm in Corollary 4, and the algorithm in Corollary 5 can be implemented in the SQ model, and require, respectively, 1 round and $O(\log \log k)$ rounds of adaptive queries. Understanding the precise relationship between the number of queries, the tolerance parameter, and the number of rounds of adaptivity for solving hypothesis selection in the SQ model is an interesting direction for future work.

Corollary 5 is derived as a consequence of a connection to maximum selection with adversarial comparators, a problem of independent interest. This connection was previously established in works by Acharya, Falahatgar, Jafarpour, Orlitsky, and Suresh (Acharya et al., 2014, 2018a). Prior work, however, has not exploited this connection under LDP constraints. Given the aforementioned importance of interactivity in the LDP setting, we initiate a study of the maximum selection with adversarial comparators problem from the perspective of understanding the trade-off between the number of rounds of parallel comparisons, and the total number of comparisons. The problem is as follows: we are given a set of items of unknown value, and we can perform comparisons between pairs of items. If the value of the items is significantly different, the comparison will correctly report the item with the larger value. If the values are similar, then the result of the comparison may be arbitrary. The goal is to output an item with value close to the maximum. We wish to minimize the total number of comparisons performed, as well as the number of rounds of interactivity.

Our main result for this setting gives a family of algorithms and lower bounds, parameterized by the number of rounds used (denoted by t). Setting $t = O(\log \log k)$ yields Corollary 5.

Theorem 6 (Restatement of Theorems 27 and 31) *For every $t \in \mathbb{Z}^+$, there exists a t-round protocol which, with probability $9/10$, approximately solves the problem of parallel approximate maximum selection with adversarial comparators from a set of k items. The algorithm requires $O(k^{1+\frac{1}{2^t-1}} t)$ comparison queries. Furthermore, any algorithm which provides these guarantees requires $\Omega(\frac{k^{1+\frac{1}{2^t-1}}}{3^t})$ comparison queries.*

For each number of rounds t, we prove an upper bound and an almost-matching lower bound. In order to get down to a near-linear number of comparisons, we require $O(\log \log k)$ rounds, which is exponentially better than the $O(\log k)$ rounds required by previous algorithms. Interestingly, in this setting, while maximum selection (with standard comparisons) with $\tilde{O}(k)$ queries is achievable in only 3 rounds, we show that $\Theta(\log \log k)$ rounds are both necessary and sufficient to achieve a near-linear number of comparisons when the results might be adversarial.

Our upper bounds follow by carefully applying a recursive tournament structure: in each round, we partition the input into appropriately-sized smaller groups, perform all pairwise-comparisons within each group, and send only the winners to the next round. Additional work is needed to prevent the quality of approximation from decaying as the number of rounds increases. For the lower bound, we restate the problem as a game, in which the adversary constructs a random complete directed graph with a unique sink, and the algorithm queries the directions of edges, and tries to

identify the sink in the smallest number of queries and rounds. We give a strategy in which the adversary constructs a layered graph with $t+1$ layers, where t is the number of rounds in the game. We can guarantee that, if the algorithm does not make enough queries, then even after conditioning on the answers to the queries in the first q rounds, the last $t+1-q$ layers of the graph remain sufficiently random, so that the algorithm cannot guess the sink with reasonable probability. In particular, after t rounds, there is still enough randomness in the $(t+1)$-st layer to make sure that algorithm cannot guess the sink correctly with high probability.

A self-contained description of the connection between hypothesis selection and maximum selection with adversarial comparators, as well as our upper and lower bounds, appear in Section 4.

1.2. Organization

We defer further discussion of related work to Appendix A. Preliminaries appear in Section 2, though we defer some of the more standard ones (i.e., the definition of differential privacy) to Appendix B. Our lower bounds for locally private hypothesis selection appear in Appendix C. Our non-interactive upper bound for locally private simple hypothesis testing appears in Section 3. Our sequentially interactive upper bound for locally private hypothesis selection, as well as our results on parallel maximum selection with adversarial comparators, appear in Section 4. Our full upper bound appears in Appendix G, and the lower bound for this setting appears in Appendix H.

2. Preliminaries

We note that there are many notions of interactivity in LDP, and we cover the two primary definitions which we will be concerned with: non-interactive and sequentially interactive protocols.

Definition 7 *An ε-LDP protocol is* non-interactive *if the number of rounds is $t=1$, and $U_1=[n]$, i.e., every individual i outputs a single message m_i, dependent only on their datapoint X_i.*

An ε-LDP protocol is sequentially interactive *with t rounds of interaction if the sets $U_1,\dots,U_t$ of active individuals in each round are disjoint.*

We recall the canonical ε-LDP algorithm, randomized response.

Lemma 8 Randomized response *is the protocol when each user has a bit $X_i\in\{0,1\}$ and outputs X_i with probability $\frac{e^\varepsilon}{1+e^\varepsilon}$ and $1-X_i$ with probability $\frac{1}{1+e^\varepsilon}$. It satisfies ε-local differential privacy.*

There exists a simple folklore algorithm for ε-LDP 2-wise simple hypothesis testing: use randomized response to privately count the number of samples which fall into the region where one distribution places more mass, and output the distribution which is more consistent with the resulting estimate. This gives the following guarantees.

Lemma 9 *There exists a non-interactive ε-LDP algorithm which solves 2-wise simple hypothesis testing with probability $1-\beta$, which requires $n=O(\log(1/\beta)/\alpha^2\varepsilon^2)$ samples.*

This can be extended to k-wise simple hypothesis testing by simply running said algorithm on pairs of distributions and picking the one which never loses a hypothesis test. This gives us an $\tilde{O}(k^2)$ baseline algorithm for locally private hypothesis selection.

Corollary 10 *There exists a non-interactive ε-LDP algorithm which solves k-wise simple hypothesis testing with high probability, which requires $n=O(k^2\log k/\alpha^2\varepsilon^2)$ samples.*

This algorithm also solves the more general problem of ε-LDP hypothesis selection (Section 4.2).

3. Non-Interactive LPHS

In this section, we prove Theorem 3. For simplicity of notation, we assume without loss of generality that $q_1, q_2, \ldots, q_k$ and p are discrete probability distributions on domain $[N]$, where $[N] := \{1, 2, \ldots, N\}$. See the discussion in Appendix E on how to deal with continuous distributions. Here we propose an algorithm which uses $n \lesssim k \operatorname{polylog}(k)/(\alpha^4\varepsilon^2)$ samples, and outputs a distribution $\hat{q} \in \mathcal{Q}$ which has TV distance of at most $O(\alpha)$ with p, when $\beta \ll \alpha^2/\log k$. Recall that $\beta := \min_{q\in\mathcal{Q}} d_{\mathrm{TV}}(p, q)$. In this mechanism, the users are divided into k groups $G_1, G_2, \ldots, G_k$ of size n/k each. Let $X_{ij} \sim p$ denote the sample with the j^{th} user in the group G_i. Our non-interactive mechanism is described in Algorithm 1.

Algorithm 1 Non-interactive ε-DP mechanism for LPHS

Input: Distributions $\mathcal{Q} = \{q_1, \ldots, q_k\}$, Samples $(X_{ij})_{i,j}$ from unknown distribution p, sensitivity parameter for Laplace noise L, privacy parameter ε, function $\gamma : [N] \to \mathbb{R}^+$ such that $|\log(\gamma(a)/q_i(a))| \le L$ for all $a \in [N], i \in [k]$.[2]

Output: $\hat{q} \in \mathcal{Q}$ such that $d_{\mathrm{TV}}(p, \hat{q}) \le \alpha$ with high probability.

for $i \in [k]$ **do**
  for $j \in [N/k]$ **do**
    The j^{th} user in group G_i sends $Z_{ij} := \log(\gamma(X_{ij})/q_i(X_{ij})) + \mathrm{Lap}(L/\varepsilon)$
  end for
  The central server computes $C_i = \frac{1}{(n/k)} \cdot \sum_{j\in[n/k]} Z_{ij}$.
end for
return $\operatorname{argmin}_i C_i$.

The proof of the following lemma appears in Appendix D.

Lemma 11 *Let $\varepsilon \in (0, 1)$ be some fixed privacy parameter. Suppose $\beta \ll \alpha^2/L$ and $n \gg \frac{k(\log k)L^2}{\alpha^4\varepsilon^2}$. Then Algorithm 1 is ε-LDP and outputs $\hat{q} \in \mathcal{Q}$ with probability at least $1 - 1/k^2$ such that $d_{\mathrm{TV}}(p, \hat{q}) \le \alpha$.*

We will now prove that we can take $L = O(\log k)$ in Algorithm 1 and Lemma 11. For this we will need the following lemma. Given a randomized map[3] $\phi : [N] \to [N']$ and a distribution q on $[N]$, the distribution $\phi \circ q$ on $[N']$ is defined as the distribution of $\phi(a)$ when a is sampled from q. For the remaining part of this section, let $U_{N'}$ denote the uniform distribution on $[N']$.

Lemma 12 (Flattening Lemma) *Let $q_1, q_2, \ldots, q_k$ be distributions over $[N]$. There exists a randomized map $\phi : [N] \to [N']$ (depending on $q_1, \ldots, q_k$) for some $N \le N' \le (k + 1)N$ s.t.*

1. *for every $a \in [N'], i \in [k]$, $\frac{1}{2N'} \le (\phi \circ q_i)(a) \le \frac{1}{N}$ and*
2. *$d_{\mathrm{TV}}(\phi \circ q_i, \phi \circ q_{i'}) = \frac{1}{2} \cdot d_{\mathrm{TV}}(q_i, q_{i'})$ for any two distributions $q_i, q_{i'}$.*

2. In other words, we require $D_\infty(\gamma||q_i), D_\infty(q_i||\gamma) \le L$ for all $i \in [k]$, i.e., all the distributions $q_1, q_2, \ldots, q_k$ are close to some distribution γ. To prove Theorem 3, we will instantiate Algorithm 1 with γ being the uniform distribution on $[N]$, but we state Algorithm 1 with arbitrary γ for generality.
3. i.e., $\phi(a)$ has a distribution over $[N']$ for each $a \in [N]$.

Proof Let $M(a) = \max_{i\in[k]} q_i(a)$ for $a \in [N]$. Let $N' = \sum_{a\in[N]} \lceil M(a)\cdot N\rceil$ and let $[N'] = \cup_{a\in[N']} S_a$ be a partition of $[N']$ with $|S_a| = \lceil M(a)\cdot N\rceil$. Define $\phi' : [N] \to [N']$ as follows: $\phi'(a)$ is uniformly distributed over S_a. Now it is clear that for every for every $b \in [N']$, $(\phi' \circ q_i)(b) \le \frac{1}{N}$. It is also clear that $||\phi' \circ q_i - \phi' \circ q_{i'}||_{\ell_1} = ||q_i - q_{i'}||_{\ell_1}$ for any two distributions $q_i, q_{i'}$. We now mix in the uniform distribution $U_{N'}$ into ϕ', i.e., we define $\phi : [N] \to [N']$ as follows: $\phi(a)$ is distributed as $\phi'(a)$ with probability $1/2$ and distributed as $U_{N'}$ with probability $1/2$. Now for every $b \in [N']$, $\frac{1}{2N'} \le (\phi \circ p_i)(b) \le \frac{1}{N}$. And $||\phi' \circ q_i - \phi' \circ q_{i'}||_{\ell_1} = \frac{1}{2}||q_i - q_{i'}||_{\ell_1}$ for any two distributions $q_i, q_{i'}$. We are now left with showing the upper bound on N. $N' = \sum_{a\in[N]} \lceil M(a)\cdot N\rceil \le \sum_{a\in[N]} (M(a)\cdot N + 1) = N + \sum_{a\in[N]} \left(\max_{i\in[k]} q_i(a)\right) N \le N + \sum_{a\in[N]} \left(\sum_{i\in[k]} q_i(a)\right) N = (k+1)N$. ■

Now we have all the ingredients to finish the proof of Theorem 3.

Proof (of Theorem 3) By using the randomized map ϕ as constructed in Lemma 12, the users first map their sample $a \sim p$ to a sample $\phi(a)$. Note that $\phi(a) \sim \phi \circ p$. Next we run the Algorithm 1 with distributions $\phi \circ q_1, \ldots, \phi \circ q_k$ and $\gamma : [N'] \to \mathbb{R}^+$ given by $\gamma(b) = 1/N'$ for all $b \in [N']$. From the first property mentioned in Lemma 12, we get $L = \log(k) + O(1)$. From the second property in Lemma 12, we know the TV distances are preserved by ϕ. This completes the proof. ■

4. Hypothesis Selection via Adversarial Comparators

In this section, we give upper bounds for locally private hypothesis selection via a reduction to adversarial comparators, as introduced by Acharya et al. (2014, 2018a). We begin by describing the reduction and how it can be implemented in the LDP setting in Section 4.1. This allows us to immediately obtain a non-interactive private algorithm which takes $\tilde{O}(k^2)$ samples and a sequentially-interactive algorithm which takes $\tilde{O}(k)$ samples (Section 4.2). However, this sequentially-interactive algorithm requires $O(\log k)$ rounds – we give an algorithm which improves upon this round-complexity by an exponential factor. We start in Section 4.3 by giving a simple $\tilde{O}(k^{4/3})$-sample algorithm which takes 2 rounds: with the addition of only a single additional round, the sample complexity becomes significantly subquadratic. This illustrates one of the main ideas behind our full upper bound, an $\tilde{O}(k)$-sample algorithm which takes only $O(\log\log k)$ rounds. This is acheived by generalizing our 2 round algorithm to general t: we give t-round algorithms for $1 \le t \le O(\log\log k)$, with sample complexities which interpolate between $\tilde{O}(k^2)$ and $\tilde{O}(k)$. Other ideas are required to achieve an approximation which does not increase with t, due to space restrictions, this is described in Appendix G. We complement these upper bounds with lower bounds which show that these algorithms in the adversarial comparator setting are essentially tight (for *every* choice of t) (Appendix H).

4.1. Adversarial Comparators and Connections to Locally Private Hypothesis Selection

We describe the adversarial comparator setting of Acharya et al. (2014, 2018a), as well as their reduction to this model for the hypothesis selection problem. The input is a set of k items, with unknown values $x_1, \ldots, x_k \in \mathbb{R}$. An adversarial comparator is a function C, which takes two items x_i and x_j,[4], and outputs $\max\{x_i, x_j\}$ if $|x_i - x_j| > 1$ and x_i or x_j (adversarially) if $|x_i - x_j| \le 1$.

4. In a slight abuse of notation, we use x_i to refer to the item as well as its value.

We note that such a comparator can be either non-adaptive or adaptive. In the former case, the results of all comparisons must be fixed ahead of time, whereas in the latter case, results of comparisons may depend on previous comparisons. All of the mentioned algorithms will work in the (harder) adaptive case, and our lower bounds are for the (easier) non-adaptive case, and thus both have the same implications in the alternate setting for adaptivity.

We sometimes denote a comparison as a *query*. The goal is to output an item with value as close to the maximum as possible, with probability at least $2/3$.[5] More precisely, let $x^* = \max\{x_1, \ldots, x_k\}$. A number x is a τ-approximation of x^* if $x \geq x^* - \tau$. Simple examples (e.g., Lemma 2 of Acharya et al. (2018a)) show that it is impossible to output a τ-approximation with probability $\geq 2/3$ for any $\tau < 2$ when we have $k \geq 3$ items.

We initiate study of *parallel* approximate maximum selection under adversarial comparators. Parallel maximum selection has recently been studied in other settings (including the standard comparison setting and with noisy (but not adversarial) comparisons, see, e.g., Braverman et al. (2016b)). In this setting, the algorithm has t rounds: in round i, the algorithm simultaneously submits m_i pairs of items, and then simultaneously receives the results of the adversarial comparator applied to all m_i pairs. The total query complexity is $\sum_{i=1}^{t} m_i$.

We now discuss the connection between this problem and hypothesis selection, as presented in Section 6 of Acharya et al. (2018a). We will then show how this connection still applies when considering the same problem under LDP. First, we recall the Scheffé test of Devroye and Lugosi (Devroye and Lugosi, 2001), as described in Algorithm 4.[6] Given n samples from p, with probability at least $1 - \beta$, it will output a distribution $\hat{q}$ such that $d_{\mathrm{TV}}(p, \hat{q}) \leq 3 \min\{d_{\mathrm{TV}}(p, q_1), d_{\mathrm{TV}}(p, q_2)\} + \sqrt{\frac{2.5 \log(1/\beta)}{n}}$. In other words, if $\min\left(d_{\mathrm{TV}}(p, q_1), d_{\mathrm{TV}}(p, q_2)\right) \leq \alpha$, then $n = O\left(\frac{\log(1/\beta)}{\alpha^2}\right)$ samples suffice to output a $\hat{q} \in \{q_1, q_2\}$ such that $d_{\mathrm{TV}}(p, \hat{q}) \leq (3 + \gamma)\alpha$, where γ can be taken to be an arbitrarily small constant. Another way to phrase this is that the test returns q_1 if $d_{\mathrm{TV}}(p, q_1) < \frac{1}{3+\gamma} d_{\mathrm{TV}}(p, q_2)$, it returns q_2 if $d_{\mathrm{TV}}(p, q_2) < \frac{1}{3+\gamma} d_{\mathrm{TV}}(p, q_1)$, and it may return arbitrarily otherwise. If we let $x_i = -\log_{3+\gamma} d_{\mathrm{TV}}(p, q_i)$, then the test will output $\max\{x_i, x_j\}$ if $|x_i - x_j| > 1$, or arbitrarily otherwise. Note that this is precisely an implementation of the adversarial comparator function C as described above, and thus the hypothesis selection problem can be reduced to (approximate) maximum selection with adversarial comparators. In particular, a τ-approximation for the maximum selection problem becomes a $(3 + \gamma)^\tau$ agnostic approximation factor for hypothesis selection, which becomes $3^\tau + \gamma'$ if τ is a constant, for some other constant $\gamma' > 0$ which can be taken to be arbitrarily small. Each comparison is implemented using $O\left(\frac{\log(1/\beta)}{\alpha^2}\right)$ samples from p– in fact, by a union bound argument, if we wish to perform m comparisons and require the total failure probability under $1/3$, all of them can be done with the same set of $O\left(\frac{\log m}{\alpha^2}\right)$ samples.

It remains to justify that a similar reduction still holds under LDP constraints. Recall that each individual possesses a single X_i, and they wish for their messages sent to the curator to be ε-DP. Only Line 4 of Algorithm 4 depends on the private data, which is a statistical query, easily implemented under LDP. More precisely, rather than sending the bit $\mathbb{1}_{X_i \in S}$ to the curator, the user can send Y_i, which is a version of it privatized by Randomized response (Lemma 8). The curator can then form an ε-LDP estimate of $p(S)$ by computing $\hat{p}(S) = \frac{e^\varepsilon + 1}{e^\varepsilon - 1}\left(\frac{1}{n}\sum Y_i - \frac{1}{e^\varepsilon + 1}\right)$. Plugging this

5. Usual arguments allow us to boost this success probability to $1 - \beta$ at a cost of $O(\log(1/\beta))$ repetitions, which can be done in parallel.
6. We comment that this can be implemented in near-linear time, and $q_1(S)$ and $q_2(S)$ can be estimated to sufficient accuracy using Monte Carlo techniques.

estimate into Line 4, it is not hard to show the modified procedure satisfies the following accuracy guarantee: if $\min\left(d_{\mathrm{TV}}(p,q_1), d_{\mathrm{TV}}(p,q_2)\right) \leq \alpha$, then $n = O\left(\frac{\log(1/\beta)}{\varepsilon^2\alpha^2}\right)$ samples suffice to output an ε-LDP $\hat{q} \in \{q_1, q_2\}$ such that $d_{\mathrm{TV}}(p,\hat{q}) \leq (3+\gamma)\alpha$, where γ can be taken to be an arbitrarily small constant.

The above addresses the case of a single comparison. If we wish to make m comparisons (which are all correct with high probability), we partition users into m sets of size $O\left(\frac{\log m}{\varepsilon^2\alpha^2}\right)$ and use the data from each part to privately perform the appropriate comparison. This takes a total of $O\left(\frac{m\log m}{\varepsilon^2\alpha^2}\right)$ samples. In particular, we can not reuse the same set of $O(\log m)$ samples for all comparisons (as in the non-private case), since it violate the privacy constraint, and doing so would give rise to algorithms which violate our main lower bound for locally private hypothesis selection (Theorem 2). Finally, we note that a t-round algorithm in the maximum selection setting corresponds to a t-round sequentially interactive ε-LDP algorithm for hypothesis selection, as we never query the same individual twice.

To conclude this section, we state the guarantees of the (trivial) algorithm which performs maximum selection from a set of 2 elements. For space reasons, we omit the corollary for LDP hypothesis selection implied by the above reduction (as well as for further results in this section), these corollaries appear in the full version (as well as representative corollaries in Appendix G).

Claim 13 *There exists a* 1*-round algorithm which achieves a* 1*-approximation in the problem of parallel approximate maximum selection with adversarial comparators, in the special case where* $k = 2$*. The algorithm requires* 1 *query.*

For the following subsections, we will focus on the problem of parallel approximate maximum selection with adversarial comparators, stating corollaries to locally private hypothesis selection as appropriate. Our primary concerns will be to simultaneously minimize the query/sample complexity and the round complexity, while minimizing the approximation/agnostic approximation factor is a secondary concern. Nevertheless, our new algorithms for maximum selection will have an approximation constant of at most 3, very close to the information-theoretic optimum of 2.

4.2. Baseline Algorithms

In this section, we state some baseline results in this model, based on previously known algorithms. This includes a $O(k^2)$-query non-interactive algorithm, and a $O(k)$-query $O(\log k)$-round algorithm.

The first method is a "round-robin" tournament method, which, in a single round, performs all pairwise comparisons and outputs the item which is declared to be the maximum the largest number of times (Algorithm 2). This straightforward method is stated and analyzed in Acharya et al. (2014, 2018a), and the equivalent procedure for hypothesis selection (absent privacy constraints) was known prior (Devroye and Lugosi, 2001).

Claim 14 *There exists a* 1*-round algorithm which achieves a* 2*-approximation in the problem of parallel approximate maximum selection with adversarial comparators. The algorithm requires* $O(k^2)$ *queries.*

The clear drawback of this method is that the complexity of the resulting algorithms is quadratic in k. Unfortunately, a simple argument shows that this is tight for any 1-round protocol: roughly, if

Algorithm 2 1-Round Algorithm for Maximum Selection

Input: k items $x_1, \ldots, x_k$
Output: Approximate maximum x_i

procedure ROUND-ROBIN($x_1, \ldots, x_k$)
  for all pairs x_i, x_j **do**
    Compare x_i and x_j, record which one is reported to be the winner.
  end for
  return the x_i which is reported to be the winner the most times.
end procedure

we do not compare the smallest and second smallest items, we do not know which is smaller, and thus any algorithm which doesn't perform all $\binom{k}{2}$ comparisons in its 1 round will be wrong with probability $1/2$ (more formal lower bounds for more general settings appear in Appendix H). The natural questions are, if we expend more rounds, can we reduce the sample complexity? And how many rounds are needed to achieve the information-theoretic optimum of a linear query complexity? Many recent works have focused on this question without concern for the number of rounds expended (Ajtai et al., 2009; Daskalakis and Kamath, 2014; Suresh et al., 2014; Acharya et al., 2014, 2018a), culminating in algorithms with linear complexity. When the round complexity is analyzed, it can be shown that all these methods take $O(\log k)$ rounds. We state the implied results for our setting in the following claim, omitting details as we will shortly improve on the round complexity to be $O(\log \log k)$.

Claim 15 (Acharya et al. (2014, 2018a)) *There exists an $O(\log k)$-round algorithm which achieves a 2-approximation in the problem of parallel approximate maximum selection with adversarial comparators. The algorithm requires $O(k)$ queries.*

4.3. A Sub-Quadratic Algorithm with 2 Rounds

In this section, we give a simple 2-round algorithm which results in a significantly better query complexity of $O(k^{4/3})$. In Appendix G, we generalize this to t-round protocols, but provide this as a warm-up and to convey one of the main ideas.

Algorithm 3 2-Round Algorithm for Maximum Selection

Input: k items $x_1, \ldots, x_k$
Output: Approximate maximum x_i

procedure 2-ROUND($x_1, \ldots, x_k$)
  Partition x_1 through x_k into $k^{2/3}$ sets of size $k^{1/3}$.
  Run ROUND-ROBIN on each set to obtain $k^{2/3}$ winners.
  return the winner of ROUND-ROBIN on the set of $k^{2/3}$ winners.
end procedure

Algorithm 3 describes the procedure, whose guarantees are summarized in the following theorem.

Theorem 16 *There exists a 2-round algorithm which achieves a 4-approximation in the problem of parallel approximate maximum selection with adversarial comparators. The algorithm requires $O(k^{4/3})$ queries.*

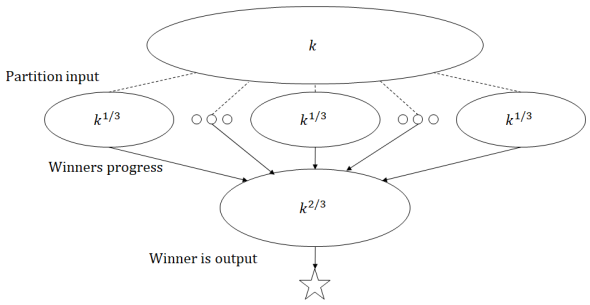

Figure 1: An illustration of Algorithm 3. In the first round, the input is partitioned into sets of size $k^{1/3}$ and a round-robin tournament is performed on each. In the second round, a single round-robin tournament is performed on the winners from the previous round.

Proof The number of rounds is easily seen to be 2: Lines 2 and 3 can be performed in one round, and Line 4, which depends on the results of the previous round, is performed in the second round.

We next analyze the number of queries. Line 3 performs the quadratic round-robin tournament of Claim 14 on sets of size $k^{1/3}$. The resulting number of queries for each set is $O(k^{2/3})$, and since there are $k^{2/3}$ sets, the total number of queries here is $O(k^{4/3})$. Line 4 performs the same quadratic round-robin tournament on one set of size $k^{2/3}$, which takes $O(k^{4/3})$ queries. Therefore, the total number of queries is $O(k^{4/3})$.

Finally, we justify that this achieves a 4-approximation to the maximum. Consider the first round: a maximum element is placed into one of the $k^{2/3}$ sets, and by the guarantees of Claim 14, the winner for this set will be a 2-approximation to the maximum. Therefore, the maximum among the winners is a 2-approximation to the overall maximum, and again by the guarantees of Claim 14, the winner of this round will be a 4-approximation to the maximum, as desired. ■

Acknowledgments

The main algorithms of Section 4 are based off of (non-optimized) ideas from a previous version of Daskalakis and Kamath (2014) (see v1 on arXiv), and GK would like to thank Constantinos Daskalakis for his guidance and ideas in that project. GK would also like to thank Sepehr Assadi and Jieming Mao for helpful discussions related to the literature on parallel selection.

GK was supported by a University of Waterloo startup grant. Part of this work was done while he was supported as a Microsoft Research Fellow, as part of the Simons-Berkeley Research Fellowship program at the Simons Institute for the Theory of Computing, and while visiting Microsoft Research Redmond. AN was supported by Ontario ERA, and NSERC Discovery grants. Part of this work was done while visiting the Simons Institute for the Theory of Computing. ZSW was supported in part by the NSF FAI Award #1939606, a Google Faculty Research Award, a J.P. Morgan Faculty Award, a Facebook Research Award, and a Mozilla Research Grant. Part of this work was done while visiting the Simons Institute for the Theory of Computing. HZ was supported by NSF #1815893 and by NSF #1704443. This work was partially done while the author was an intern at Microsoft Research Redmond.

References

Jayadev Acharya, Ashkan Jafarpour, Alon Orlitsky, and Ananda Theertha Suresh. Sorting with adversarial comparators and application to density estimation. In *Proceedings of the 2014 IEEE International Symposium on Information Theory*, ISIT '14, pages 1682–1686, Washington, DC, USA, 2014. IEEE Computer Society.

Jayadev Acharya, Moein Falahatgar, Ashkan Jafarpour, Alon Orlitsky, and Ananda Theertha Suresh. Maximum selection and sorting with adversarial comparators. *Journal of Machine Learning Research*, 19(1):2427–2457, 2018a.

Jayadev Acharya, Gautam Kamath, Ziteng Sun, and Huanyu Zhang. Inspectre: Privately estimating the unseen. In *Proceedings of the 35th International Conference on Machine Learning*, ICML '18, pages 30–39. JMLR, Inc., 2018b.

Jayadev Acharya, Ziteng Sun, and Huanyu Zhang. Differentially private testing of identity and closeness of discrete distributions. In *Advances in Neural Information Processing Systems 31*, NeurIPS '18, pages 6878–6891. Curran Associates, Inc., 2018c.

Jayadev Acharya, Clément L. Canonne, Cody Freitag, and Himanshu Tyagi. Test without trust: Optimal locally private distribution testing. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics*, AISTATS '19, pages 2067–2076. JMLR, Inc., 2019a.

Jayadev Acharya, Clément L. Canonne, and Himanshu Tyagi. Inference under information constraints: Lower bounds from chi-square contraction. In *Proceedings of the 32nd Annual Conference on Learning Theory*, COLT '19, pages 1–15, 2019b.

Arpit Agarwal, Shivani Agarwal, Sepehr Assadi, and Sanjeev Khanna. Learning with limited rounds of adaptivity: Coin tossing, multi-armed bandits, and ranking from pairwise comparisons. In *Proceedings of the 30th Annual Conference on Learning Theory*, COLT '17, pages 39–75, 2017.

Miklós Ajtai, János Komlós, and Endre Szemerédi. An $o(n \log n)$ sorting network. In *Proceedings of the 15th Annual ACM Symposium on the Theory of Computing*, STOC '83, pages 1–9, New York, NY, USA, 1983. ACM.

Miklós Ajtai, Vitaly Feldman, Avinatan Hassidim, and Jelani Nelson. Sorting and selection with imprecise comparisons. In *Proceedings of the 36th International Colloquium on Automata, Languages, and Programming*, ICALP '09, pages 37–48, 2009.

Maryam Aliakbarpour, Ilias Diakonikolas, and Ronitt Rubinfeld. Differentially private identity and closeness testing of discrete distributions. In *Proceedings of the 35th International Conference on Machine Learning*, ICML '18, pages 169–178. JMLR, Inc., 2018.

Maryam Aliakbarpour, Ilias Diakonikolas, Daniel M. Kane, and Ronitt Rubinfeld. Private testing of distributions via sample permutations. In *Advances in Neural Information Processing Systems 32*, NeurIPS '19, pages 10877–10888. Curran Associates, Inc., 2019.

Noga Alon. Expanders, sorting in rounds and superconcentrators of limited depth. In *Proceedings of the 17th Annual ACM Symposium on the Theory of Computing*, STOC '85, pages 98–102, New York, NY, USA, 1985. ACM.

Noga Alon and Yossi Azar. The average complexity of deterministic and randomized parallel comparison-sorting algorithms. *SIAM Journal on Computing*, 17(6):1178–1192, 1988a.

Noga Alon and Yossi Azar. Sorting, approximate sorting, and searching in rounds. *SIAM Journal on Discrete Mathematics*, 1(3):269–280, 1988b.

Noga Alon, Yossi Azar, and Uzi Vishkin. Tight complexity bounds for parallel comparison sorting. In *Proceedings of the 27th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '86, pages 502–510, Washington, DC, USA, 1986. IEEE Computer Society.

Kareem Amin, Matthew Joseph, and Jieming Mao. Pan-private uniformity testing. *arXiv preprint arXiv:1911.01452*, 2019.

Jordan Awan and Aleksandra Slavković. Differentially private uniformly most powerful tests for binomial data. In *Advances in Neural Information Processing Systems 31*, NeurIPS '18, pages 4208–4218. Curran Associates, Inc., 2018.

Yossi Azar and Nicholas Pippenger. Parallel selection. *Discrete Applied Mathematics*, 27(1-2): 49–58, 1990.

Yossi Azar and Uzi Vishkin. Tight comparison bounds on the complexity of parallel sorting. *SIAM Journal on Computing*, 16(3):458–464, 1987.

Béla Bollobás and Graham Brightwell. Parallel selection with high probability. *SIAM Journal on Discrete Mathematics*, 3(1):21–31, 1990.

Béla Bollobás and Pavol Hell. Sorting and graphs. In Ivan Rival, editor, *Graphs and Order: The Role of Graphs in the Theory of Ordered Sets and Its Applications*, pages 169–184. Springer, 1985.

Béla Bollobás and Andrew Thomason. Parallel sorting. *Discrete Applied Mathematics*, 6(1):1–11, 1983.

Olivier Bousquet, Daniel M. Kane, and Shay Moran. The optimal approximation factor in density estimation. In *Proceedings of the 32nd Annual Conference on Learning Theory*, COLT '19, pages 318–341, 2019.

Mark Braverman, Ankit Garg, Tengyu Ma, Huy L. Nguyen, and David P. Woodruff. Communication lower bounds for statistical estimation problems via a distributed data processing inequality. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 1011–1020, 2016a. doi: 10.1145/2897518.2897582. URL https://doi.org/10.1145/2897518.2897582.

Mark Braverman, Jieming Mao, and S. Matthew Weinberg. Parallel algorithms for select and partition with noisy comparisons. In *Proceedings of the 48th Annual ACM Symposium on the Theory of Computing*, STOC '16, pages 851–862, New York, NY, USA, 2016b. ACM.

Mark Braverman, Jieming Mao, and Yuval Peres. Sorted top-k in rounds. In *Proceedings of the 32nd Annual Conference on Learning Theory*, COLT '19, pages 342–382, 2019.

Hai Brenner and Kobbi Nissim. Impossibility of differentially private universally optimal mechanisms. *SIAM Journal on Computing*, 43(5):1513–1540, 2014.

Mark Bun, Gautam Kamath, Thomas Steinke, and Zhiwei Steven Wu. Private hypothesis selection. In *Advances in Neural Information Processing Systems 32*, NeurIPS '19, pages 156–167. Curran Associates, Inc., 2019.

Bryan Cai, Constantinos Daskalakis, and Gautam Kamath. Priv'it: Private and sample efficient identity testing. In *Proceedings of the 34th International Conference on Machine Learning*, ICML '17, pages 635–644. JMLR, Inc., 2017.

Zachary Campbell, Andrew Bray, Anna Ritz, and Adam Groce. Differentially private ANOVA testing. In *Proceedings of the 2018 International Conference on Data Intelligence and Security*, ICDIS '18, pages 281–285, Washington, DC, USA, 2018. IEEE Computer Society.

Clément L. Canonne, Gautam Kamath, Audra McMillan, Adam Smith, and Jonathan Ullman. The structure of optimal private tests for simple hypotheses. In *Proceedings of the 51st Annual ACM Symposium on the Theory of Computing*, STOC '19, New York, NY, USA, 2019a. ACM.

Clément L. Canonne, Gautam Kamath, Audra McMillan, Jonathan Ullman, and Lydia Zakynthinou. Private identity testing for high-dimensional distributions. *arXiv preprint arXiv:1905.11947*, 2019b.

T-H Hubert Chan, Elaine Shi, and Dawn Song. Private and continual release of statistics. *ACM Transactions on Information and System Security (TISSEC)*, 14(3):1–24, 2011.

Vincent Cohen-Addad, Frederik Mallmann-Trenn, and Claire Mathieu. Instance-optimality in the noisy value-and comparison-model* accept, accept, strong accept: Which papers get in? In *Proceedings of the 31st Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '20, pages 2124–2143, Philadelphia, PA, USA, 2020. SIAM.

Simon Couch, Zeki Kazan, Kaiyan Shi, Andrew Bray, and Adam Groce. Differentially private nonparametric hypothesis testing. In *Proceedings of the 2019 ACM Conference on Computer and Communications Security*, CCS '19, New York, NY, USA, 2019. ACM.

Amit Daniely and Vitaly Feldman. Locally private learning without interaction requires separation. In *Advances in Neural Information Processing Systems 32*, NeurIPS '19, pages 14975–14986. Curran Associates, Inc., 2019.

Constantinos Daskalakis and Gautam Kamath. Faster and sample near-optimal algorithms for proper learning mixtures of Gaussians. In *Proceedings of the 27th Annual Conference on Learning Theory*, COLT '14, pages 1183–1213, 2014.

Constantinos Daskalakis, Ilias Diakonikolas, and Rocco A. Servedio. Learning Poisson binomial distributions. In *Proceedings of the 44th Annual ACM Symposium on the Theory of Computing*, STOC '12, pages 709–728, New York, NY, USA, 2012. ACM.

Luc Devroye and Gábor Lugosi. A universally acceptable smoothing factor for kernel density estimation. *The Annals of Statistics*, 24(6):2499–2512, 1996.

Luc Devroye and Gábor Lugosi. Nonasymptotic universal smoothing factors, kernel complexity and Yatracos classes. *The Annals of Statistics*, 25(6):2626–2637, 1997.

Luc Devroye and Gábor Lugosi. *Combinatorial methods in density estimation*. Springer, 2001.

Luc Devroye, Abbas Mehrabian, and Tommy Reddad. The total variation distance between high-dimensional Gaussians. *arXiv preprint arXiv:1810.08693*, 2018.

Ilias Diakonikolas, Gautam Kamath, Daniel M. Kane, Jerry Li, Ankur Moitra, and Alistair Stewart. Robust estimators in high dimensions without the computational intractability. In *Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '16, pages 655–664, Washington, DC, USA, 2016. IEEE Computer Society.

Differential Privacy Team, Apple. Learning with privacy at scale. https://machinelearning.apple.com/docs/learning-with-privacy-at-scale/appledifferentialprivacysystem.pdf, December 2017.

Bolin Ding, Janardhan Kulkarni, and Sergey Yekhanin. Collecting telemetry data privately. In *Advances in Neural Information Processing Systems 30*, NIPS '17, pages 3571–3580. Curran Associates, Inc., 2017.

John Duchi and Ryan Rogers. Lower bounds for locally private estimation via communication complexity. In *Proceedings of the 32nd Annual Conference on Learning Theory*, COLT '19, pages 1161–1191, 2019.

John C. Duchi, Michael I. Jordan, and Martin J. Wainwright. Local privacy and statistical minimax rates. In *Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '13, pages 429–438, Washington, DC, USA, 2013. IEEE Computer Society.

John C. Duchi, Michael I. Jordan, and Martin J. Wainwright. Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association*, 2017.

Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends® in Machine Learning*, 9(3–4):211–407, 2014.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the 3rd Conference on Theory of Cryptography*, TCC '06, pages 265–284, Berlin, Heidelberg, 2006. Springer.

Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. RAPPOR: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM Conference on Computer and Communications Security*, CCS '14, pages 1054–1067, New York, NY, USA, 2014. ACM.

Alexandre Evfimievski, Johannes Gehrke, and Ramakrishnan Srikant. Limiting privacy breaches in privacy preserving data mining. In *Proceedings of the 22nd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, PODS '03, pages 211–222, New York, NY, USA, 2003. ACM.

Uriel Feige, Prabhakar Raghavan, David Peleg, and Eli Upfal. Computing with noisy information. *SIAM Journal on Computing*, 23(5):1001–1018, 1994.

Marco Gaboardi and Ryan Rogers. Local private hypothesis testing: Chi-square tests. In *Proceedings of the 35th International Conference on Machine Learning*, ICML '18, pages 1626–1635. JMLR, Inc., 2018.

Marco Gaboardi, Hyun-Woo Lim, Ryan M. Rogers, and Salil P. Vadhan. Differentially private chi-squared hypothesis testing: Goodness of fit and independence testing. In *Proceedings of the 33rd International Conference on Machine Learning*, ICML '16, pages 1395–1403. JMLR, Inc., 2016.

Roland Häggkvist and Pavol Hell. Parallel sorting with constant time for comparisons. *SIAM Journal on Computing*, 10(3):465–472, 1981.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. In *The Collected Works of Wassily Hoeffding*, pages 409–426. Springer, 1994.

Matthew Joseph, Jieming Mao, Seth Neel, and Aaron Roth. The role of interactivity in local differential privacy. In *Proceedings of the 60th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '19, pages 94–105, Washington, DC, USA, 2019. IEEE Computer Society.

Matthew Joseph, Jieming Mao, and Aaron Roth. Exponential separations in local differential privacy through communication complexity. In *Proceedings of the 31st Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '20, pages 515–527, Philadelphia, PA, USA, 2020. SIAM.

Kazuya Kakizaki, Jun Sakuma, and Kazuto Fukuchi. Differentially private chi-squared test by unit circle mechanism. In *Proceedings of the 34th International Conference on Machine Learning*, ICML '17, pages 1761–1770. JMLR, Inc., 2017.

Gautam Kamath and Jonathan Ullman. A primer on private statistics. *arXiv preprint arXiv:2005.00010*, 2020.

Gautam Kamath, Jerry Li, Vikrant Singhal, and Jonathan Ullman. Privately learning high-dimensional distributions. In *Proceedings of the 32nd Annual Conference on Learning Theory*, COLT '19, pages 1853–1902, 2019.

Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. What can we learn privately? *SIAM Journal on Computing*, 40(3):793–826, 2011.

Michael J. Kearns. Efficient noise-tolerant learning from statistical queries. *J. ACM*, 45(6):983–1006, 1998.

Daniel Kifer and Ryan M. Rogers. A new class of private chi-square tests. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, AISTATS '17, pages 991–1000. JMLR, Inc., 2017.

Clyde P. Kruskal. Searching, merging, and sorting in parallel computation. *IEEE Transactions on Computers*, C-32(10):942–946, 1983.

Tom Leighton. Tight bounds on the complexity of parallel sorting. In *Proceedings of the 16th Annual ACM Symposium on the Theory of Computing*, STOC '84, pages 71–80, New York, NY, USA, 1984. ACM.

Satyaki Mahalanabis and Daniel Stefankovic. Density estimation in linear time. In *Proceedings of the 21st Annual Conference on Learning Theory*, COLT '08, pages 503–512, 2008.

Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '07, pages 94–103, Washington, DC, USA, 2007. IEEE Computer Society.

Jerzy Neyman and Egon Sharpe Pearson. Ix. on the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231(694-706):289–337, 1933.

Nicholas Pippenger. Sorting and selecting in rounds. *SIAM Journal on Computing*, 16(6):1032–1038, 1987.

Or Sheffet. Locally private hypothesis testing. In *Proceedings of the 35th International Conference on Machine Learning*, ICML '18, pages 4605–4614. JMLR, Inc., 2018.

Ananda Theertha Suresh, Alon Orlitsky, Jayadev Acharya, and Ashkan Jafarpour. Near-optimal-sample estimators for spherical Gaussian mixtures. In *Advances in Neural Information Processing Systems 27*, NIPS '14, pages 1395–1403. Curran Associates, Inc., 2014.

Marika Swanberg, Ira Globus-Harris, Iris Griffith, Anna Ritz, Adam Groce, and Andrew Bray. Improved differentially private analysis of variance. *Proceedings on Privacy Enhancing Technologies*, 2019(3), 2019.

Chao Tao, Qin Zhang, and Yuan Zhou. Collaborative learning with limited interaction: Tight bounds for distributed exploration in multi-armed bandits. In *Proceedings of the 60th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '19, pages 126–146, Washington, DC, USA, 2019. IEEE Computer Society.

Caroline Uhler, Aleksandra Slavković, and Stephen E. Fienberg. Privacy-preserving data sharing for genome-wide association studies. *The Journal of Privacy and Confidentiality*, 5(1):137–166, 2013.

Jonathan Ullman. Tight lower bounds for locally differentially private selection. *arXiv preprint arXiv:1802.02638*, 2018.

Leslie G. Valiant. Parallelism in comparison problems. *SIAM Journal on Computing*, 4(3):348–355, 1975.

Duy Vu and Aleksandra Slavković. Differential privacy for clinical trial data: Preliminary evaluations. In *2009 IEEE International Conference on Data Mining Workshops*, ICDMW '09, pages 138–143. IEEE, 2009.

Yue Wang, Jaewoo Lee, and Daniel Kifer. Revisiting differentially private hypothesis tests for categorical data. *arXiv preprint arXiv:1511.03376*, 2015.

Stanley L. Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965.

Yannis G. Yatracos. Rates of convergence of minimum distance estimators and Kolmogorov's entropy. *The Annals of Statistics*, 13(2):768–774, 1985.

Appendix A. Related Work

As mentioned before, our work builds on a long line of investigation on hypothesis selection. This style of approach was pioneered by Yatracos (Yatracos, 1985), and refined in subsequent work by Devroye and Lugosi (Devroye and Lugosi, 1996, 1997, 2001). After this, additional considerations have been taken into account, such as computation, approximation factor, robustness, and more (Mahalanabis and Stefankovic, 2008; Daskalakis et al., 2012; Daskalakis and Kamath, 2014; Suresh et al., 2014; Acharya et al., 2014; Diakonikolas et al., 2016; Acharya et al., 2018a; Bousquet et al., 2019; Bun et al., 2019). Most relevant is the recent work of Bun, Kamath, Steinke, and Wu (Bun et al., 2019), which studies hypothesis selection under central differential privacy. Our results are for the stronger constraint of local differential privacy.

Versions of our problem have been studied under both central and local differential privacy. In the local model, the most pertinent result is that of Duchi, Jordan, and Wainwright (Duchi et al., 2013, 2017), showing a lower bound on the sample complexity for simple hypothesis testing between two known distributions. This matches folklore upper bounds for the same problem. However, the straightforward way of extending said protocol to k-wise simple hypothesis testing would incur a cost of $\tilde{O}(k^2)$ samples. Other works on hypothesis testing under local privacy include Gaboardi and Rogers (2018); Sheffet (2018); Acharya et al. (2019a,b); Joseph et al. (2019). In the central model, some of the early work was done by the Statistics community (Vu and Slavković, 2009; Uhler et al., 2013). More recent work can roughly be divided into two lines – one attempts to provide private analogues of classical statistical tests (Wang et al., 2015; Gaboardi et al., 2016; Kifer and Rogers, 2017; Kakizaki et al., 2017; Campbell et al., 2018; Swanberg et al., 2019; Couch et al., 2019), while the other focuses more on achieving minimax sample complexities for testing problems (Cai et al., 2017; Acharya et al., 2018c; Aliakbarpour et al., 2018; Acharya et al., 2018b; Canonne et al., 2019b; Aliakbarpour et al., 2019; Amin et al., 2019). While most of these focus on composite hypothesis testing, we highlight Canonne et al. (2019a) which studies simple hypothesis testing. Work of Awan and Slavkovic (Awan and Slavković, 2018) gives a universally optimal test for binomial data, however Brenner and Nissim (Brenner and Nissim, 2014) give an impossibility result for distributions with domain larger than 2. For further coverage of differentially private statistics, see Kamath and Ullman (2020).

We are the first to study parallel maximum selection with adversarial comparators. Prior work has investigated (non-parallel) maximum selection and sorting with adversarial comparators (Ajtai et al., 2009; Acharya et al., 2014, 2018a). Works by Acharya, Falahatgar, Jafarpour, Orlitsky, and Suresh established the connection with hypothesis selection (Acharya et al., 2014, 2018a). The parallelism model we study here was introduced by Valiant (1975), for parallel comparison-based problems with non-adversarial comparators. This has inspired a significant literature on parallel sorting and selection (Häggkvist and Hell, 1981; Ajtai et al., 1983; Bollobás and Thomason, 1983; Kruskal, 1983; Leighton, 1984; Bollobás and Hell, 1985; Alon, 1985; Alon et al., 1986; Azar and Vishkin, 1987; Pippenger, 1987; Alon and Azar, 1988a,b; Azar and Pippenger, 1990; Bollobás and Brightwell, 1990; Feige et al., 1994; Braverman et al., 2016b, 2019; Cohen-Addad et al., 2020). Also, note that the noisy comparison models considered in some of these papers (where comparisons are incorrect with a certain probability) is different from the adversarial comparator model

we study. Thematically similar investigations on round complexity exist in the context of best arm identification for multi-armed bandits (Agarwal et al., 2017; Tao et al., 2019).

Appendix B. More Preliminaries

We recall the definition of total variation distance between probability distributions.

Definition 17 *The* total variation distance *between distributions p and q on Ω is defined as*

$$d_{\mathrm{TV}}(p,q) = \max_{S\subseteq\Omega} p(S) - q(S) = \frac{1}{2}\int_{x\in\Omega} |p(x) - q(x)|d = \frac{1}{2}\|p-q\|_1 \in [0,1].$$

We now define differential privacy, and the variants which we are concerned with.

Definition 18 (Dwork et al. (2006)) *An algorithm M with domain $\mathcal{X}^n$ is* ε-differentially private *if, for all $S \subseteq \mathrm{Range}(M)$ and all inputs $x, y \in \mathcal{X}^n$ which differ in exactly one point,*

$$\Pr[M(x) \in S] \leq e^{\varepsilon} \Pr[M(y) \in S].$$

This definition is also called pure differential privacy.

In the local setting of differential privacy, we imagine that each user has a single datapoint. We require that each individual's output is differentially private.

Definition 19 (Warner (1965); Evfimievski et al. (2003); Kasiviswanathan et al. (2011)) *Suppose there are n individuals, where the ith individual has datapoint X_i. In each round q of the protocol, there is a set $U_q \subseteq [n]$ of active individuals, and each individual i in U_q computes some (randomized) function of their datapoint X_i, and of all messages $\{m_{r,j} : r \leq q, j \in U_r\}$ output by all individuals in previous rounds, and outputs a message $m_{q,i}$. A protocol is* ε-locally differentially private *(LDP) if the set $\{m_{q,i} : q \in [t], i \in U_q\}$ of all messages output during the t rounds of the protocol is ε-differentially private with respect to the inputs $(X_1, \ldots, X_n)$.*

Appendix C. Lower Bounds for Locally Private Hypothesis Selection

In this section we state sample complexity lower bound results on locally private hypothesis selection. We will first focus on the lower bound for non-interactive protocols, and leverage a known lower bound on locally private selection due to Ullman (2018) (a similar statement appears in Duchi et al. (2017)), which also follows from the lower bound for sparse estimation in Duchi et al. (2017). Let $d \in \mathbb{N}$, $\alpha \in [0,1]$, and let U_d be a uniform distribution over $\{\pm 1\}^d$. For every $b \in \{\pm 1\}$ and $j \in [d]$, we define distribution $p_{b,j} = (1-\alpha)U_d + \alpha\,(U_d \mid x_j = b)$, that is, the distribution that is uniform over $\{\pm 1\}^d$ except that $X_j = b$ with probability $1/2 + \alpha$.

Theorem 20 (Theorem 3.1 of Ullman (2018)) *Let $\varepsilon \in (0,1)$. Let $d > 32$, B be distbuted uniformly over $\{\pm 1\}$, and let J be distributed uniformly over $[d]$. Suppose M is an non-interactive ε-LDP protocol and n is such that*

$$\Pr_{B,J,X_1,\ldots,X_n\sim(p_{B,J}|B,J)}[M(X_1,\ldots,X_n) = (B,J)] \geq 1/3.$$

Then

$$n = \Omega\left(\frac{d\log d}{\alpha^2\varepsilon^2}\right).$$

To obtain a lower bound on hypothesis selection, we will rely on the following fact that bounds the total variation distance between the distributions $p_{b,j}$ (see e.g., Lemma 6.4 in Kamath et al. (2019)).

Fact 21 *Let q and q' be two product distributions over $\{\pm 1\}^d$ with mean vectors μ and μ' respectively, such that $\mu_i \in [-1/3, 1/3]$ for all $j \in [d]$. Suppose that $\|\mu - \mu'\|_2 \geq \alpha$ for any $\alpha \leq \alpha_0$ with some absolute constant $0 < \alpha_0 \leq 1$. Then $d_{\mathrm{TV}}(q, q') \geq C\alpha$, for some absolute constant C.*

Theorem 22 (Non-interactive lower bound) *Let $\varepsilon \in (0, 1)$. Suppose M is a non-interactive an ε-LDP protocol that solves the k-wise simple hypothesis testing problem with probability at least $1/3$ when given n samples from some distribution $p \in \mathcal{Q}$, where $\mathcal{Q} = \{q_1, \ldots, q_k\}$ are distributions such that $\min_{i \neq j} d_{\mathrm{TV}}(q_i, q_j) \geq \alpha$. Then*

$$n \geq \Omega\left(\frac{k \log k}{\alpha^2 \varepsilon^2}\right).$$

Proof Let $\mathcal{Q} = \{p_{b,j} \mid b \in \{\pm 1\}, j \in [d]\}$ be a set of $k = 2d$ probability distributions. For any pair of distributions $q, q' \in \mathcal{Q}$, we know from Fact 21 that $d_{\mathrm{TV}}(q, q') \geq \alpha / C$ for some absolute constant C. Thus, our stated bound follows from Theorem 20. ■

Next we will derive a sample complexity lower bound for general locally private protocols. We will build on a result due to Duchi and Rogers (2019) and consider the set of 1-sparse Gaussian distributions $\{\mathcal{N}(\theta, I_d) \mid \theta \in \Theta\}$, where $\Theta = \{\theta \in \mathbb{R}^d \mid \|\theta\|_2 = \alpha, \|\theta\|_0 = 1\}$ is the set of vectors that have a single non-zero coordinate, equal to $-\alpha$ or $+\alpha$.

Following the result of Duchi and Rogers (2019) (and the framework of Braverman et al. (2016a)), we can obtain a general lower bound analogous to Theorem 20.

Theorem 23 (Corollary 6 of Duchi and Rogers (2019), Theorem 4.5 of Braverman et al. (2016a)) *Let $\varepsilon \in (0, 1)$. Let U be a uniform distbution over Θ. Suppose M is an ε-LDP protocol, and n is such that*

$$\Pr_{\theta \sim U, X_1, \ldots, X_n \sim \mathcal{N}(\theta, I)}[M(X_1, \ldots, X_n) = \theta] \geq 1/3.$$

Then

$$n \geq \Omega\left(\frac{d}{\alpha^2 \varepsilon^2}\right).$$

Theorem 2 *Let $\varepsilon \in (0, 1)$. Suppose M is an ε-LDP protocol that solves the k-wise simple hypothesis testing problem with probability at least $1/3$ when given n samples from some distribution $p \in \mathcal{Q}$, for any set $\mathcal{Q} = \{q_1, \ldots, q_k\}$ such that $\min_{i \neq j} d_{\mathrm{TV}}(q_i, q_j) \geq \alpha$. Then $n = \Omega\left(\frac{k}{\alpha^2 \varepsilon^2}\right)$.*

Proof For any two 1-sparse vectors $\theta, \theta' \in \Theta$ such that $\theta \neq \theta'$, the total variation distance between their Gaussian distributions is given by $\|\theta - \theta'\|_2 = \sqrt{2}\alpha$ (see, e.g., Devroye et al. (2018)). Thus, our stated bound follows from Theorem 23. ■

Appendix D. Proof of Lemma 11

From our assumption, $|\log(\gamma(a)/q_i(a))| \leq L$ for $a \in [N], i \in [k]$. The algorithm adds noise sampled from $\mathrm{Lap}(L/\varepsilon)$, hence ε-LDP guarantee follows easily from the properties of the Laplace mechanism Dwork and Roth (2014). We will now prove correctness. Let $i^* \in [k]$ be such that $d_{\mathrm{TV}}(p, q_{i^*}) = \beta$. Fix a group G_i and consider,

$$\begin{aligned}
\mathbf{E}_{X_{ij}\sim p}[C_i] &= \frac{1}{(n/k)} \cdot \mathbf{E}\left[\sum_{j\in[n/k]} Z_{ij}\right] \\
&= \mathbf{E}_{a\sim p}\left[\log\left(\frac{\gamma(a)}{q_i(a)}\right)\right] \quad \text{(By the linearity of expectation and } \mathbf{E}[\mathrm{Lap}(L/\varepsilon)] = 0) \\
&= \sum_{a\in[N]} p(a)\log\left(\frac{\gamma(a)}{q_i(a)}\right) \\
&= \sum_{a\in[N]} q_{i^*}(a)\log\left(\frac{q_{i^*}(a)}{q_i(a)}\right) + \sum_{a\in[N]} (p(a) - q_{i^*}(a))\log\left(\frac{\gamma(a)}{q_i(a)}\right) + \sum_{a\in[N]} q_{i^*}(a)\log\left(\frac{\gamma(a)}{q_{i^*}(a)}\right) \\
&= D_{\mathrm{KL}}(q_{i^*}||q_i) + \sum_{a\in[N]} (p(a) - q_{i^*}(a))\log\left(\frac{\gamma(a)}{q_i(a)}\right) - D_{\mathrm{KL}}(q_{i^*}||\gamma).
\end{aligned}$$

Let $B = -D_{\mathrm{KL}}(q_{i^*}||\gamma)$. By re-arranging the above term we get

$$\begin{aligned}
\left|\mathbf{E}_{X_{ij}\sim q}[C_i] - D_{\mathrm{KL}}(q_{i^*}||q_i) - B\right| &\leq \sum_{a\in[N]} |p(a) - q_{i^*}(a)| \cdot |\log(\gamma(a)/q_i(a))| \\
&\leq \sup_{a\in[N]} \left|\log\left(\frac{\gamma(a)}{q_i(a)}\right)\right| \cdot \left(\sum_{a\in[N]} |p(a) - q_{i^*}(a)|\right) \\
&\leq L \cdot 2d_{\mathrm{TV}}(p, q_{i^*}) \\
&\leq 2L\beta \leq 0.1\alpha^2.
\end{aligned}$$

Now observe that each Z_{ij} can be expressed as $W_{ij} + Y_{ij}$, where $Y_{ij} \sim \mathrm{Lap}(L/\varepsilon)$, and the support of random variable W_{ij} is in the interval $[-L, L]$ from our assumption. Therefore, we can apply the standard Hoeffding's inequality and concentration of Laplace random variables (see Hoeffding (1994); Chan et al. (2011) for example) to obtain $\Pr\left[|C_i - \mathbf{E}[C_i]| \geq 0.1\alpha^2\right] \leq \exp\left(-\Omega(1) \cdot \frac{(n/k)\alpha^4}{(L/\varepsilon)^2}\right) \leq \frac{1}{k^3}$.

By taking the union bound, with probability at least $1 - 1/k^2$, $|C_i - D_{\mathrm{KL}}(q_{i^*}||q_i) - B| \leq 0.2\alpha^2$ for all $i \in [k]$. In particular, $C_{i^*} \leq B + 0.2\alpha^2$. This implies that if $i' = \mathrm{argmin}_i\, C_i$, then $C_{i'} \leq B + 0.2\alpha^2$. It remains to argue that $d_{\mathrm{TV}}(p, q_{i'}) < \alpha$. Suppose not. Consider any q_i such that $d_{\mathrm{TV}}(p, q_i) > \alpha$. This implies that $d_{\mathrm{TV}}(q_{i*}, q_i) > \alpha/2$ based on our assumption. Now consider $C_i \geq B + D_{\mathrm{KL}}(q_{i^*}||q_i) - 0.2\alpha^2 \geq B + 2d_{\mathrm{TV}}(q_{i^*}, q_i)^2 - 0.2\alpha^2 \geq B + 0.3\alpha^2$, where we used Pinsker's inequality.

Appendix E. Discussion on Handling Continuous Distributions

The arguments in our proof can be easily generalized to continuous probability distributions. However, as our results do not depend on the domain size, it is intuitive to think of the following simple mapping from continuous distributions to discrete distributions on the domain $[N]$. First, we can approximate (to any precision) a set of continuous distributions by a set of discrete distributions on a finite support such that TV distances are preserved. We can then map any set of discrete distributions on a finite support to a set of discrete distributions on the domain $[N]$, where N will depend on the desired precision.

Furthermore, if we are able to get $L = O(\alpha)$, then we get the nearly optimal sample complexity of $n = O\left(\frac{k\text{polylog}(k)}{\alpha^2\varepsilon^2}\right)$.

Appendix F. Algorithm 4

Algorithm 4 Scheffé Test

Input: n samples $X_1, \ldots, X_n$ from *unknown* p, distributions q_1 and q_2
Output: Distribution q_1 or q_2

procedure SCHEFFÉ(X, q_1, q_2)
  Let $S = \{x \,:\, q_1(x) > q_2(x)\}$.
  Let $q_1(S)$ and $q_2(S)$ be the probability mass that q_1 and q_2 assign to S .
  Let $\hat{p}(S) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i \in S}$ be the empirical mass assigned by $X_1, \ldots, X_n$ to S.
  if $|q_1(S) - \hat{p}(S)| < |q_2(S) - \hat{p}(S)|$ **then**
    return q_1.
  else
    return q_2.
  end if
end procedure

Appendix G. A Near-Linear-Sample Algorithm with $O(\log\log k)$ Rounds

In this section, we describe our main result in this setting, a family of algorithms for approximate maximum selection parameterized by t, which is the allowed number of rounds. By setting $t = O(\log\log k)$, we will get an $O(k\log\log k)$-query algorithm which requires only $O(\log\log k)$ rounds, improving exponentially on the round complexity of previous approaches. In particular, the following corollaries are obtained from Theorem 27 and Corollary 28 with an optimized setting of parameters.

Corollary 24 *There exists an $O(\log\log k)$-round algorithm which, with probability $9/10$, achieves a 3-approximation in the problem of parallel approximate maximum selection with adversarial comparators. The algorithm requires $O(k\log\log k)$ queries.*

Corollary 25 *There exists an $O(\log\log k)$-round algorithm which achieves a $(27+\gamma)$-agnostic approximation factor for locally private hypothesis selection with probability $9/10$, where $\gamma > 0$ is an arbitrarily small constant. The sample complexity of the algorithm is $O\left(\frac{k\log k\log\log k}{\varepsilon^2\alpha^2}\right)$.*

The method is a careful recursive application of the approach described in Algorithm 3. Specifically, given t allowed rounds of adaptivity, we partition the items into several smaller sets, perform the round-robin algorithm on each, and then feed the winners into the algorithm which is allowed $t-1$ rounds of adaptivity. A judicious setting of parameters will allow the number of comparisons to decay quite rapidly as the number of rounds is increased. This construction is described and analyzed in Appendix G.1. One challenge is that each round of the algorithm will potentially lose an additive 2 in the approximation, resulting in an overall $2t$-approximation. To avoid this, we employ ideas from Daskalakis and Kamath (2014): we simultaneously apply two algorithms, at least one of which will be effective depending on whether the density of elements close to the maximum is high or low. We describe the necessary modification and analyze the resulting approach in Appendix G.2.

G.1. A Recursive Application of the 2-Round Method

Our main result of this section will be the following lemma. While the round and query complexity are essentially optimal (see Appendix H), the quality of approximation is unsatisfactory – our approach to improving this approximation is described in G.2.

Lemma 26 *There exists a t-round algorithm which achieves a $2t$-approximation in the problem of parallel approximate maximum selection with adversarial comparators. The algorithm requires $O(k^{1+\frac{1}{2^t-1}}t)$ queries.*

The method is described in Algorithm 5. Note that for $t=1$ or $t=2$, this simplifies to Algorithms 2 and 3, respectively.

Algorithm 5 t-Round Algorithm for Maximum Selection

Input: k items $x_1,\ldots,x_k$, number of rounds t
Output: Approximate maximum x_i

1: **procedure** MULTI-ROUND($x_1,\ldots,x_k,t$)
2: **if** $t=1$ **then**
3: **return** the winner of ROUND-ROBIN on $x_1,\ldots,x_k$.
4: **end if**
5: Set $\eta_t=\frac{1}{2^t-1}$.
6: Partition x_1 through x_k into $k^{1-\eta_t}$ sets of size k^{η_t}.
7: Run ROUND-ROBIN on each set to obtain $k^{1-\eta_t}$ winners.
8: **return** the winner of MULTI-ROUND on the set of $k^{1-\eta_t}$ winners with $t-1$ rounds.
9: **end procedure**

We proceed with proving that this algorithm satisfies the guarantees stated in Lemma 26.

Proof We prove the guarantees by induction. The base case corresponds to $t=1$. As mentioned before, this is exactly equal to Algorithm 2, and thus by Claim 14, the lemma holds.

Now, we prove the lemma for a general $t>1$, assuming it holds for $t-1$. The number of rounds is trivial: 1 round is spent performing Lines 6 and 7, and $t-1$ rounds are spent on the recursive call in Line 8. The approximation is also easy to reason about: the maximum element in the input appears in one of the sets in the partition in Line 6, and therefore the winner of the corresponding set will be a 2-approximation of the maximum. Thus, the set of winners which are fed into the recursive call in Line 8 will have a 2-approximation of the maximum. The inductive

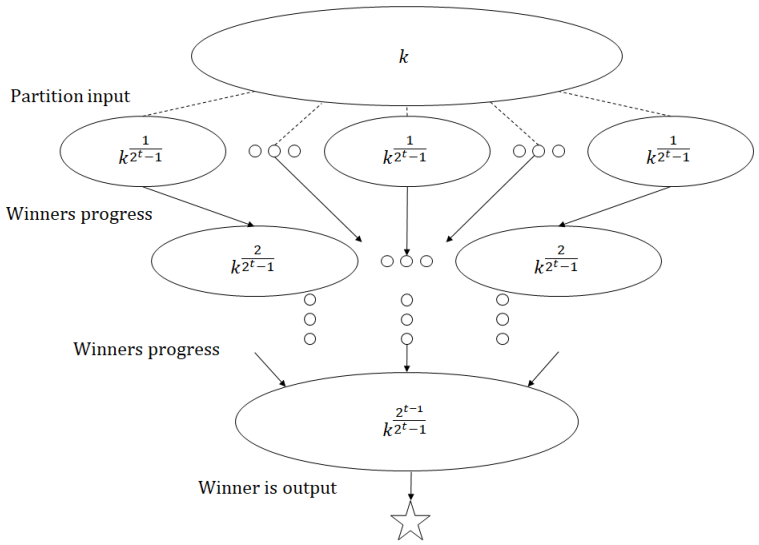

Figure 2: An illustration of Algorithm 5. The input is partitioned into several sets and a round-robin tournament is performed on each. In subsequent rounds, winners are merged into fewer but larger sets, until we have only a single winner.

hypothesis guarantees that the winner of the recursive call will be a $2(t-1)$-approximation to *this* item, making it a $2t$-approximation to the maximum.

Finally, it remains to reason about the query complexity. Comparisons are only performed in Lines 7 and 8. In the former, we perform the round-robin tournament on $k^{1-\eta_t}$ sets of size k^{η_t}, so the total number of comparisons is $k^{1-\eta_t} \cdot O(k^{2\eta_t}) = O(k^{1+\eta_t})$. In the latter, the recursive call has an input of size $k^{1-\eta_t}$, so by the inductive hypothesis, the number of comparisons done in the recursive call is $O\left(\left(k^{1-\eta_t}\right)^{1+\frac{1}{2^{t-1}-1}}(t-1)\right)$. Substituting in the value $\eta_t = \frac{1}{2^t-1}$, these two terms sum to $O(k^{1+\frac{1}{2^t-1}}t)$, as desired. ■

G.2. Bounding the Approximation Factor

While the guarantees of Lemma 26 are strong in terms of the round and query complexity, the approximation leaves something to be desired. We alleviate this issue in a similar way as Daskalakis and Kamath (2014), by running a very simple strategy in parallel to the main method of Algorithm 5. The intuition is as follows: if an item with maximum value x^* is never compared with an item with value x' such that $x^* > x' \geq 1$ (i.e., numbers which are 1-approximations to the maximum), it will never lose a comparison. If the fraction of such elements is low, then an item with value x^* will make it to the final round, thus guaranteeing that the overall winner will be a 2-approximation to the maximum. On the other hand, if the fraction of such elements is high, then we can sample a small number of items such that we select at least one 1-approximation to x^*, and running the round-robin algorithm on this set will guarantee a 3-approximation to the maximum.

Our method is described more precisely in Algorithm 6, and the guarantees are described in Theorem 27.

Algorithm 6 Better t-Round Algorithm for Maximum Selection

Input: K items $x_1, \ldots, x_k$, number of rounds t
Output: Approximate maximum x_i

1: **procedure** BETTER-MULTI-ROUND($x_1, \ldots, x_k, t$)
2: Run MULTI-ROUND on a random permutation of $x_1, \ldots x_k$ with t rounds, but halt when $t = 1$ and let L be the set of all remaining items.
3: Let H be a random subset of $\{x_1, \ldots, x_k\}$ of size $O\left(k^{\frac{2^t-1}{2^t-1}}\right)$.
4: Run ROUND-ROBIN on $L \cup H$ and **return** the winner.
5: **end procedure**

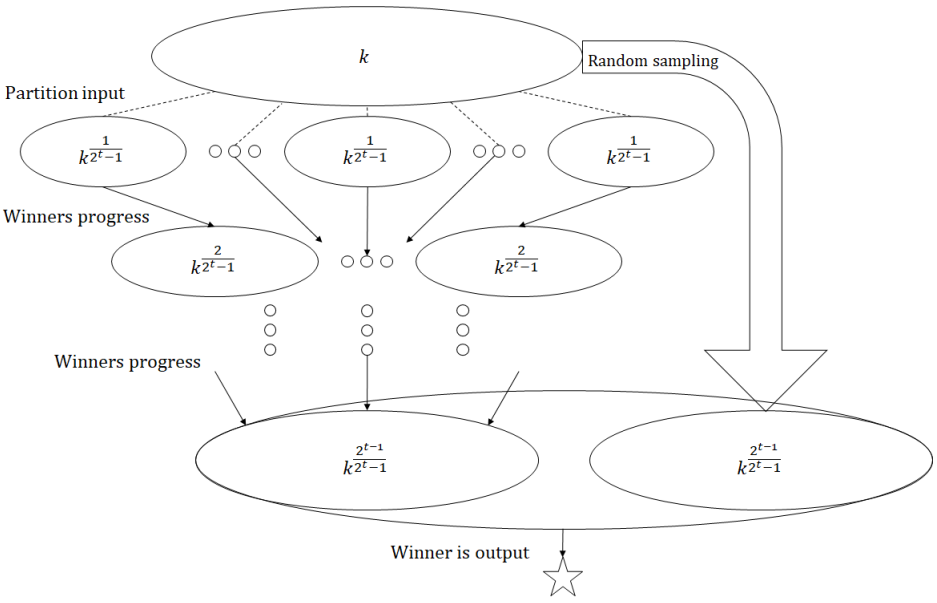

Figure 3: An illustration of Algorithm 6. Similar to Algorithm 5, but in the last round, we perform a round-robin tournament additionally involving a random sample of items from the input.

Theorem 27 *There exists a t-round algorithm which, with probability $9/10$, achieves a 3-approximation in the problem of parallel approximate maximum selection with adversarial comparators. The algorithm requires $O(k^{1+\frac{1}{2^t-1}}t)$ queries.*

This gives the following corollary for LDP hypothesis selection.

Corollary 28 *There exists a t-round algorithm which achieves a $(27+\gamma)$-agnostic approximation factor for locally private hypothesis selection with probability $9/10$, where $\gamma > 0$ is an arbitrarily small constant. The sample complexity of the algorithm is $O\left(\frac{k^{1+\frac{1}{2^t-1}}t\log k}{\varepsilon^2\alpha^2}\right)$.*

Corollaries 24 and 25 follow from these statements with an appropriate setting of t.

To conclude, we prove Theorem 27.

Proof The number of rounds is straightforward to analyze: Line 2 takes $t-1$ rounds (since we stop one round early), and Lines 3 and 4 can be done in 1 last round.

To analyze the number of comparisons, we require the following claim, which quantifies the number of items that make it to the last round of MULTI-ROUND.

Claim 29 $|L| = k^{\frac{2^{t-1}}{2^t-1}}$.

Proof We recall the η_t notation of Algorithm 5. The number of items which begin the first round of the algorithm is clearly k. Since these are partitioned into $k^{1-\eta_t}$ groups, each producing a single winner which progresses to the next round, we have $k^{1-\eta_t}$ items which begin the second round of the algorithm. A similar reasoning implies that the number of items entering the third round of the algorithm is $\left(k^{1-\eta_t}\right)^{1-\eta_{t-1}}$. Noting that $|L|$ is the number of items entering the t-th (i.e., final) round of the algorithm, the same logic shows that

$$\log_k |L| = \prod_{i=0}^{t-2}\left(1 - \frac{1}{2^{t-i}-1}\right) = \prod_{i=0}^{t-2}\left(\frac{2(2^{t-i-1}-1)}{2^{t-i}-1}\right) = \frac{2^{t-1}}{2^t-1},$$

as desired. The latter equality can be seen by a telescoping argument, as the numerators cancel the subsequent denominators. ■

With this in hand, the number of comparisons is the number of comparisons due to Line 2 (which is $O\left(k^{1+\frac{1}{2^t-1}}(t-1)\right)$ by the same argument as in the proof of Lemma 26) plus the number of comparisons due to Line 4, which is $O\left((|H|+|L|)^2\right) = O\left(\left(k^{\frac{2^{t-1}}{2^t-1}}\right)^2\right) = O\left(k^{1+\frac{1}{2^t-1}}\right)$. Combining both of these gives the desired number of comparisons.

Finally, we justify the accuracy guarantee. We split the analysis into two cases, based on the density of items which have value comparable to the maximum. Let $\zeta = \max_i x_i$ be the maximum value, let $B = \{i : \zeta > x_i \geq \zeta - 1\}$ be the set of items which are 1-approximations to (but not strictly equal to) the maximum value, and let $\gamma = \frac{|B|}{k}$ be their density.

First, suppose that $\gamma \geq 1/10k^{\frac{2^{t-1}}{2^t-1}}$. If we let the hidden constant in the size of H be 100 (i.e., $|H| = 100k^{\frac{2^{t-1}}{2^t-1}}$), then Markov's inequality says that at least one item in H will be a 1-approximation to the maximum value with probability at least $9/10$. By the guarantees of ROUND-ROBIN (quantified in Claim 14), the result of Line 4 will be a 3-approximation to the maximum value, as desired.

On the other hand, suppose that $\gamma \leq 1/10k^{\frac{2^{t-1}}{2^t-1}}$. We argue that an item with value ζ makes it to the final round of MULTI-ROUND and is included in L – if this happens, then by the guarantees of ROUND-ROBIN, the result of Line 4 will be a 2-approximation to ζ and the proof is complete. This happens if an item with value ζ is never compared to any element from B within the first $t-1$ rounds. Fix some such item: the probability it is compared with some element from B is upper bounded by the probability that any element of B appears in the same subtree of depth $t-1$ leading up to the final round. The number of elements contained in this subtree is $k/|L| = k^{\frac{2^{t-1}-1}{2^t-1}}$, by Claim 29. The expected number of items from B in this subtree is bounded as $\gamma \cdot k/|L| \leq \frac{1}{10k^{\frac{1}{2^t-1}}} \leq 1/10$, and the result follows again from Markov's inequality. ■

Appendix H. A Lower Bound for Selection via Adversarial Comparators

In this section, we provide a lower bound for adversarial maximum selection with constrained interactivity. In Appendix H.1, we consider a special case when $t = 2$ and prove that any 2-round algorithm requires $\Omega\left(k^{\frac{4}{3}}\right)$ comparisons to find an approximate maximum. In Appendix H.2, we generalize our result and technique to t rounds and prove that any t-round algorithm requires $\Omega\left(\frac{k^{1+\frac{1}{2^t-1}}}{3^t}\right)$ comparisons. These lower bounds hold even for a non-adaptive adversary.

H.1. A Lower Bound for 2-round Algorithms

We warm up with a simpler case which illustrates the main ideas, namely, a lower bound for 2-round algorithms. Specifically, we show that no matter how large the approximation factor τ is, any 2-round algorithm which solves the parallel approximate maximum selection problem requires $\Omega(k^{\frac{4}{3}})$ comparisons.

Theorem 30 *For any $\tau > 1$, any 2-round algorithm which achieves a τ-approximation in the problem of parallel approximate maximum selection with non-adaptive adversarial comparators requires $\Omega(k^{\frac{4}{3}})$ queries.*

We remark that, since our result is proved in the setting of non-adaptive adversarial comparators, it also automatically holds for adaptive comparators as well.

In our lower bound constructions, we reformulate the parallel approximate maximum selection problem as a game between an adversary and the algorithm. Before the game starts, the adversary commits to a random tournament (i.e., a complete directed graph)[7] on k nodes, each identified with one of the k items. We will require that the tournament has, with probability 1, a single sink node. Then, the algorithm player asks m_1 queries to the adversary, each query corresponding to a comparison between items x_i and x_j. If the corresponding edge between x_i and x_j in the tournament is directed from x_i to x_j, then the adversary answers that $x_j > x_i$, and, otherwise, the adversary answers that $x_i > x_j$. Equivalently, the algorithm asks for the directions of m_1 edges, which are revealed by the adversary. Afterwards, the player asks m_2 additional queries, based on the information gained from the initial m_1 queries, and the adversary answers them according to the directions of edges in the tournament. The game continues in this manner for t rounds, where in round q the algorithm asks m_q queries, possibly dependent on all the query answers so far. After the t-th round, the algorithm must declare the "winner", i.e., the sink in the tournament.

Note that we can always produce item values so that the query answers are valid for the adversarial comparators model, and the sink node is the unique τ-approximate maximum. Let $C_1, \dots C_\ell$ be the strongly connected components of the tournament, ordered so that, if $i < j$, then all edges between C_i and C_j are directed from C_i to C_j. Then we can set, for example, $x_j = 2i\tau$ for all x_j in C_i. This way all queries to two items in the same strongly connected component can be answered arbitrarily, and all queries to items in two different components can be answered according to the direction of edges in the tournament. Moreover, we want to mention two special components. First, since there is a unique sink node x_{i^*}, C_ℓ must be equal to $\{x_{i^*}\}$, and therefore, x_{i^*} is the unique τ-approximate maximum. Second, in order to "fool" the player, the adversary sets $C_{\ell-1} = \{x_{i'}\}$, where all edges incident on $x_{i'}$ are directed towards $x_{i'}$, except the edge from $x_{i'}$ to x_{i^*}. Thus, if the

7. Note that in this section we use "tournament" in the graph theoretic sense.

algorithm can achieve a τ-approximation in the parallel approximate maximum selection problem, it can identify the sink node in the game above, and especially, distinguish it from $x_{i'}$.

We are now ready to prove Theorem 30.

Proof We model the problem as the game described above, with $t = 2$. By Yao's minimax principle, we can assume, without loss of generality, that the algorithm player makes deterministic choices. We start with the construction of the random tournament. From now on, to make the notation more convenient, we will denote nodes/items by their indices, i.e., we will write i rather than x_i. Let U_0 denote the complete set of the nodes. Firstly, the adversary picks a uniformly random subset U_1 of $k^{\frac{2}{3}}$ nodes from U_0. Then from the adversary picks two nodes i^* and i' uniformly at random from U_1.

Now we describe the directions of the edges of the tournament. For convenience, we define $V_0 := U_0 \backslash U_1$ and $V_1 := U_1 \backslash \{i^*, i'\}$. All edges incident on i^* are directed towards i^*, i.e., i^* is our sink node. All edges incident on i' are directed towards i', except the edge from i' to i^*. All edges from V_0 to V_1 are directed towards the node in V_1. Finally, the direction of any edge between two nodes in V_0 or two nodes in V_1 is chosen uniformly and independently from all other random choices.

Now we switch to the side of the player. As noted above, any algorithm which achieves τ-approximation must correctly identify i^* as the sink, with probability higher than $\frac{2}{3}$. Given $m_1 = m_2 = \frac{1}{100}k^{\frac{4}{3}}$, we want to show that any algorithm which asks $m = m_1 + m_2$ queries can not find i^* with this probability. In the first round, the player asks $m_1 = \frac{1}{100}k^{\frac{4}{3}}$ number of queries. We use $e_j = \{\alpha_j, \beta_j\}$ to denote the j-th query, where $j \in [m_1]$. Let S denote the set of the nodes in U_1 which have ever competed with some other nodes from U_1, i.e., $S = \{i_1 \in U_1 : \exists i_2 \in U_1, \exists j \in [m_1], e_j = \{i_1, i_2\}\}$. Now we want to show that the following two "bad" events happen with a small probability:

$$A_1 = \{i' \in S \ \cup \ i^* \in S\}, \quad A_2 = \{|V_1 \cap S| \geq \frac{1}{2} \cdot k^{\frac{2}{3}}\}.$$

We bound the probability of event A_1 and A_2, respectively. For the rest of the proof, we will assume that k is a large enough constant. By a union bound,

$$\Pr[A_1] \leq \Pr\big[i' \in S\big] + \Pr[i^* \in S] \leq 2 \cdot m_1 \cdot \frac{k^{\frac{2}{3}} - 1}{\binom{k}{2}} \leq 0.05.$$

With respect to A_2, let $e_j = \{\alpha_j, \beta_j\}$, where $j \in [m_1]$. We note that $|V_1 \cap S| \leq 2 \cdot \sum_{j \in [m_1]} \mathbb{I}(\alpha_j \in V_1, \beta_j \in V_1)$, where $\forall j$, $\mathbf{E}(\mathbb{I}(\alpha_j \in V_1, \beta_j \in V_1)) = \binom{k^{\frac{2}{3}}}{2} \backslash \binom{k}{2} = \frac{k^{\frac{2}{3}}-1}{k-1} \cdot k^{-\frac{1}{3}}$. Furthermore, $\forall j_1 \neq j_2$, $\mathbb{I}(\alpha_{j_1} \in V_1, \beta_{j_1} \in V_1)$ and $\mathbb{I}(\alpha_{j_2} \in V_1, \beta_{j_2} \in V_1)$ are negatively correlated.

Therefore,

$$\mathbf{E}\left(\sum_{j \in [m_1]} \mathbb{I}(\alpha_j \in V_1, \beta_j \in V_1)\right) = m_1 \cdot \frac{k^{\frac{2}{3}} - 1}{k - 1} \cdot k^{-\frac{1}{3}} = \frac{1}{100} \cdot \frac{k}{k-1} \cdot (k^{\frac{2}{3}} - 1),$$

$$\mathbf{Var}\left(\sum_{j \in [m_1]} \mathbb{I}(\alpha_j \in V_1, \beta_j \in V_1)\right) \leq m_1 \cdot \frac{k^{\frac{2}{3}} - 1}{k-1} \cdot k^{-\frac{1}{3}} \cdot \left(1 - \frac{k^{\frac{2}{3}} - 1}{k-1} \cdot k^{-\frac{1}{3}}\right) < \frac{1}{100} \cdot \frac{k}{k-1} \cdot (k^{\frac{2}{3}} - 1).$$

By Chebyshev's inequality,

$$\Pr[A_2] \le \Pr\left[\sum_{j\in[m_1]} \mathbb{I}(\alpha_j \in V_1, \beta_j \in V_1) \ge \frac{1}{4}\cdot k^{\frac{2}{3}}\right] \le k^{-\frac{2}{3}} \le 0.05,$$

where in the last inequality, we assume $k \ge 100$.

Now we move to the second round. From now on, we condition on neither A_1 nor A_2 holding, which happens with probability at least 0.9. Then, conditional on A_1 and on the answers to the first m_1 queries, the pair $\{i^*, i'\}$ is distributed uniformly in the set $R = \{i^*, i'\} \cup (V_1 \setminus S)$. Moreover, if the algorithm does not query $\{i^*, i'\}$ in the second round, then i^* and i' will have the same distribution conditional on all m queries, and the algorithm will not be able to identify i^* with probability higher than 0.5. Then, conditional on A_1, A_2, and the queries from the first round, the probability that the algorithm queries $\{i^*, i'\}$ in the second round is at most

$$m_2 \cdot \frac{1}{\binom{|R|}{2}} \le \frac{k^{4/3}}{50\cdot\frac{1}{2}k^{2/3}\cdot(\frac{1}{2}k^{2/3}-1)} \le 0.1.$$

Therefore, the success rate of any deterministic 2-round algorithm making at most $\frac{k^{2/3}}{100}$ queries is at most $0.1 + 0.1 + 0.5 < \frac{9}{10}$. As already noted, by Yao's minimax principle this also implies the result for randomized algorithms.

■

H.2. A Lower Bound for t-round Algorithms

In this section, we extend our 2-round lower bound to t rounds. Specifically, we want to prove the following theorem.

Theorem 31 *For any $\tau > 1$, any t-round algorithm which achieves τ-approximation in the problem of parallel approximate maximum selection with non-adaptive adversarial comparators requires $\Omega\left(\frac{k^{1+\frac{1}{2^t-1}}}{3^t}\right)$ queries.*

We continue to model the problem as the game described in the previous subsection, but now with general t. We start with the construction of the random tournament, where a similar hierarchical structure to the 2-round construction is adopted. In the structure in Appendix H.1, we can view node i^* and i' as layer 2, nodes in set V_1 as layer 1, and all the other nodes as layer 0. We have thus designed a 3-layer hierarchical structure in the proof of the 2-round lower bound, where edges are directed from lower to higher layers, and edges in the same layer are directed randomly. In this section, we generalize this construction to the following $(t+1)$-layer hierarchical structure, which we denote as (k, t)-construction.

Let U_0 denote the complete set of the nodes. In the first round, the adversary uniformly at random picks $k^{\frac{2^t-2}{2^t-1}}$ different nodes from U_0, which are denoted as U_1; etc.; in the q-th round, the adversary uniformly randomly picks $k^{\frac{2^t-2^q}{2^t-1}}$ from U_{q-1}, denoted as U_q, where $q \in [t-1]$. Finally, the adversary uniformly at random picks two nodes from U_{t-1}, denoted as i^* and i', respectively,

and we let $U_t = \{i^*, i'\}$ for the purpose of consistency. For convenience, we define $V_0 = U_0 \backslash U_1$, $\cdots$, $V_q = U_q \backslash U_{q+1}$, where $0 \le q \le t-1$, and $V_t = U_t = \{i^*, i'\}$. For $i < j$, we direct all edges from V_i to V_j; for $q \neq t$, edges between two nodes in V_q are given a uniformly random direction; finally, the edge between i^* and i' is directed towards i^*. Thus, i^* is the unique sink in the graph.

The following is the core lemma in this section.

Lemma 32 *Given a (k,t)-construction, and $\forall \gamma < 1$, every deterministic t-round algorithm which finds i^* with probability higher than $\left(\frac{1}{2} + \frac{\gamma}{100} \cdot 3^t\right)$ requires $\Omega\left(\gamma\left(k^{1+\frac{1}{2^t-1}}\right)\right)$ queries.*

It is not hard to show that Theorem 31 can be viewed as a corollary of the lemma, since given the random (k,t)-construction, by setting $\gamma = \frac{1}{3^t}$, the lemma tells that every t-round algorithm which finds i^* with constant probability makes at least $\Omega\left(\frac{k^{1+\frac{1}{2^t-1}}}{3^t}\right)$ queries, and any algorithm which achieves τ-approximation should find i^* with constant probability. Finally, by Yao's minimax principle, this also holds for randomized algorithms. Therefore, our remaining task is to prove Lemma 32.

Proof We prove the lemma by induction. Throughout the proof we assume that k is large enough with respect to t and $\frac{1}{\gamma}$. We will assume that the algorithm makes at most $\frac{\gamma}{100}\left(k^{1+\frac{1}{2^t-1}}\right)$ queries, and show inductively that it succeeds in identifying i^* with probability at most $\frac{1}{2} + \frac{\gamma}{100} \cdot 3^t$.

For the base case when $t = 2$, the lemma holds from the argument in the previous section. For the inductive step, let t be any integer where $t \ge 3$. Recall that the number of queries asked by the algorithm in the first round is $m_1 \le \frac{\gamma}{100} k^{1+\frac{1}{2^t-1}}$. We use $e_j = (\alpha_j, \beta_j)$ to denote the j-th query, where $j \in [m_1]$. By analogy with the 2-round proof, let S denote the set of nodes in U_1 which have ever competed with some other nodes from U_1, i.e., $S = \{i_1 \in U_1 : \exists i_2 \in U_1, \exists j \in [m_1], e_j = (i_1, i_2)\}$. Now we want to show that the following t "bad" events happen with a small probability:

$$\forall q \in [t-1], A_q = \{|V_q \cap S| \ge \frac{1}{10} \cdot k^{\frac{2^t-2^q}{2^t-1}}\}, \quad A_t = \{i' \in S \ \cup \ i^* \in S\}.$$

We bound the probability of event A_t first. By a union bound,

$$\Pr[A_t] \le \Pr\left[i' \in S\right] + \Pr[i^* \in S] \le 2 \cdot m_1 \cdot \frac{k^{\frac{2^t-2}{2^t-1}} - 1}{\binom{k}{2}} \le 0.05\gamma.$$

With respect to $A_q, q \in [t-1]$, let $e_j = (\alpha_j, \beta_j)$, where $j \in [m_1]$. We note that $|V_q \cap S| \le 2 \cdot \sum_{j \in [m_1]} \mathbb{I}(\alpha_j \in V_1, \beta_j \in V_q)$, where $\forall j$, $\mathbf{E}(\mathbb{I}(\alpha_j \in V_1, \beta_j \in V_q)) = \frac{k^{\frac{2^t-2}{2^t-1}} \cdot k^{\frac{2^t-2^q}{2^t-1}}}{\binom{k}{2}}$, which is roughly $k^{-\frac{2^q}{2^t-1}}$. Furthermore, $\forall j_1 \neq j_2$, $\mathbb{I}(\alpha_{j_1} \in V_1, \beta_{j_1} \in V_q)$ and $\mathbb{I}(\alpha_{j_2} \in V_1, \beta_{j_2} \in V_q)$ are negatively correlated. Therefore,

$$\mathbf{E}\left(\sum_{j \in [m_1]} \mathbb{I}(\alpha_j \in V_1, \beta_j \in V_q)\right) = m_1 \cdot \frac{k^{\frac{2^t-2}{2^t-1}} \cdot k^{\frac{2^t-2^q}{2^t-1}}}{\binom{k}{2}} \le \frac{1}{50} \cdot k^{\frac{2^t-2^q}{2^t-1}},$$

$$\mathbf{Var}\left(\sum_{j \in [m_1]} \mathbb{I}(\alpha_j \in V_1, \beta_j \in V_q)\right) \le m_1 \cdot \frac{k^{\frac{2^t-2}{2^t-1}} \cdot k^{\frac{2^t-2^q}{2^t-1}}}{\binom{k}{2}} \cdot \left(1 - \frac{k^{\frac{2^t-2}{2^t-1}} \cdot k^{\frac{2^t-2^q}{2^t-1}}}{\binom{k}{2}}\right) \le \frac{1}{50} \cdot k^{\frac{2^t-2^q}{2^t-1}}.$$

By Chebyshev's inequality,

$$\Pr[A_q] \le \Pr\left[\sum_{j\in[m_1]} \mathbb{I}(\alpha_j \in V_1, \beta_j \in V_q) \ge \frac{1}{20}\cdot k^{\frac{2^t-2^q}{2^t-1}}\right] \le 25k^{-\frac{2^{t-1}}{2^t-1}} \le \frac{0.05\gamma}{t},$$

where in the last inequality, we assume $k \ge \frac{Ct^2}{\gamma^2}$ for a large enough constant C.

From now on, we condition on none of the bad events $A_1, \dots, A_t$ holding, which happens with probability at least $1 - 0.1\gamma$. We also condition on the answers to the first m_1 queries. We would like to say that the conditional distribution on the graph induced on $U_1 \setminus S$ is identical to that of a $(k', t-1)$-construction for $k' = U_1 \setminus S$. However, because of the random choice of S, the sizes of $U_q \setminus S$ are not exactly as prescribed in the definition of a $(k', t-1)$ construction. In order to finish the induction, we consider the following process. For $k' = \frac{1}{2}k^{\frac{2^t-2}{2^t-1}}$, we first denote $V'_t = \{i^*, i'\}$; then, we uniformly at random draw $(k')^{\frac{2^{t-2}}{2^{t-1}-1}} - 2$ nodes from from $V_{t-1}\setminus S$, and denote them as V'_{t-1}; from $V_q \setminus S$, $q \in [t-1]$, we uniformly at random draw $(k')^{\frac{2^{t-1}-2^{q-1}}{2^{t-1}-1}} - |V'_{q+1}| < \left(\frac{1}{2}\right)^{\frac{2^{t-1}-2^{q-1}}{2^{t-1}-1}} \cdot k^{\frac{2^t-2^q}{2^t-1}} < \frac{3}{4}k^{\frac{2^t-2^q}{2^t-1}}$ nodes, and denote them as V'_q. Conditonal on the bad events not holding, and on the query answers from the first round, the subgraph induced on the nodes from V'_1, V'_2, $\cdots$, and V'_t, is distributed identically to a $(k', t-1)$ construction. Clearly, for the algorithm to determine the sink i^* in the full tournament, it must also determine it in this subgraph. Ignoring queries in rounds $2, \dots, t$ to edges not in the subgraph, the algorithm is allowed to ask at most $m = \frac{\gamma}{100}\left(k^{1+\frac{1}{2^t-1}}\right) \le \frac{\gamma}{100}\cdot 2.7 \cdot \left((k')^{1+\frac{1}{2^{t-1}-1}}\right)$ queries, and, by the inductive assumption, any $(t-1)$-round algorithm can find i^* with probability at most $\frac{1}{2} + \frac{3^{t-1}}{100}\cdot 2.7\gamma$. Finally, by a union bound, the probability of success of the t-round algorithm is at most $\frac{1}{2} + \frac{3^{t-1}}{100}\cdot 2.7\gamma + 0.1\gamma \le \frac{1}{2} + \frac{3^t}{100}\gamma$. This finishes the inductive step. ■