Optimal Query Complexity for Private Sequential Learning Against Eavesdropping: Supplementary Materials

1 Statements of Results

In this section we give the general statements of our results under both the noiseless response and noisy response models. When the responses are noisy, we give the precise constant multipliers.

Theorem 1 (Noiseless responses, Bayesian setting)**.** *If $2\epsilon \le \delta \le 1/L$, then*

$$\left\lfloor \log \frac{1}{L\delta} \right\rfloor + L\left(\log \frac{\delta}{\epsilon} - 2\right) - 1 \le N(\epsilon, \delta, L) \le \left\lfloor \log \frac{1}{L\delta} \right\rfloor + L\left(\left\lceil \log \frac{\delta}{\epsilon} \right\rceil + 2\right) - 1.$$

Theorem 2 (Noiseless responses, deterministic setting)**.** *If $2\epsilon \le \delta \le 1/L$, then*

$$\max\left\{\left\lceil \log \frac{1}{\epsilon} \right\rceil + L - 8, \left\lceil \log \frac{\delta}{\epsilon} \right\rceil + 2L - 4\right\} \le N(\epsilon, \delta, L) \le \max\left\{\left\lceil \log \frac{1}{\epsilon} \right\rceil + L, \left\lceil \log \frac{\delta}{\epsilon} \right\rceil + 2L\right\}.$$

Theorem 3 (Noisy responses, Bayesian setting)**.** *Let p be the probability that each responses is corrupted. If $4\epsilon \le \delta \le 1/L$, Then*

$$\frac{1}{2c_2(p)} \max\left\{L \log \frac{\delta}{16\epsilon},\ \log \frac{1}{8\epsilon}\right\} \le N_{\mathsf{avg}}(\epsilon, \delta, L) \le \left(\frac{14}{c_3(p)} + \frac{7}{c_4(p)}\right)\left(\log \frac{1}{\epsilon} + L \log \frac{64\delta}{\epsilon}\right), \tag{1}$$

and

$$\begin{aligned} &\max\left\{\frac{L}{2c_2(p)} \log \frac{\delta}{8\epsilon},\ \frac{1}{2c_2(p)} \log \frac{1}{4\epsilon},\ \frac{L}{2c_1(p)} \log \frac{M}{8}\right\} \\ &\le N_{\mathsf{whp}}(\epsilon, M, \delta, L) \le \left(\frac{8}{c_3(p)} + \frac{7}{c_4(p)}\right)\left(\log \frac{1}{\epsilon} + L \log \frac{12M\delta}{\epsilon}\right), \end{aligned} \tag{2}$$

where

$$\begin{aligned} c_1(p) =& D(Bernoulli(1-p)||Bernoulli(p)) = (1-p)\log\frac{1-p}{p} + p\log\frac{p}{1-p}, \\ c_2(p) =& h(1/2) - h(p), \quad with\ h(p) = H(Bernoulli(p)) = -p\log p - (1-p)\log(1-p), \\ c_3(p) =& (p-1/2)^2 \log e, \\ c_4(p) =& D(Bernoulli(1/2)||Bernoulli(p)) = \tfrac{1}{2}\left(\log\frac{1}{2p} + \log\frac{1}{2(1-p)}\right) \end{aligned}$$

are constants that only depend on p.

It follows from (1) and (2) and the basic inequality $\max\{a, b\} \ge (a+b)/2$ that

$$N_{\mathsf{avg}}(\epsilon, \delta, L) \asymp_p L \log \frac{\delta}{\epsilon} + \log \frac{1}{\epsilon},$$

$$N_{\mathsf{whp}}(\epsilon, M, \delta, L) \asymp_p L \log \frac{M\delta}{\epsilon} + \log \frac{1}{\epsilon},$$

where $\asymp_p$ denotes matching upper and lower bounds up to multiplicative constants that depend on p. Note that $c_1(p)$, $c_2(p)$, $c_3(p)$ and $c_4(p)$ are all on the order of $(p-1/2)^2$ for p close to $1/2$. Therefore all the multiplicative constants in the upper and lower bounds go to infinity at the rate of $(p-1/2)^{-2}$ as $p \to 1/2$.

2 Proofs of Theorem 1 and Theorem 2 (noiseless responses)

In this section we prove our results on the optimal query complexity for the sequential learning model, under the Bayesian setting (Theorems 1) and the deterministic setting (Theorem 2).

2.1 Analysis under the Bayesian setting

Proof of Theorem 1. **Upper bound:** To prove the upper bound of Theorem 1, we construct the following multistage querying strategy for the learner (precise description in Algorithm 1):

1. Run bisection search on $[0,1]$ for K_1 steps to locate X^* within an interval I of length $2^{-K_1} \approx L\delta$;
2. Divide I into L subintervals $I_1, ..., I_L$ of equal length (about δ). Query the $L-1$ endpoints of all the subintervals (the two endpoints of I were already queried in stage 1) to determine which subinterval contains X^*.
3. Say I_{i^*} is the true subinterval which contains X^*. Run bisection search on the I_{i^*} for K_2 steps until ϵ-accuracy is achieved, while submitting cloned queries in the other $L-1$ subintervals in parallel.

See Fig. 1 for a graphical illustration.

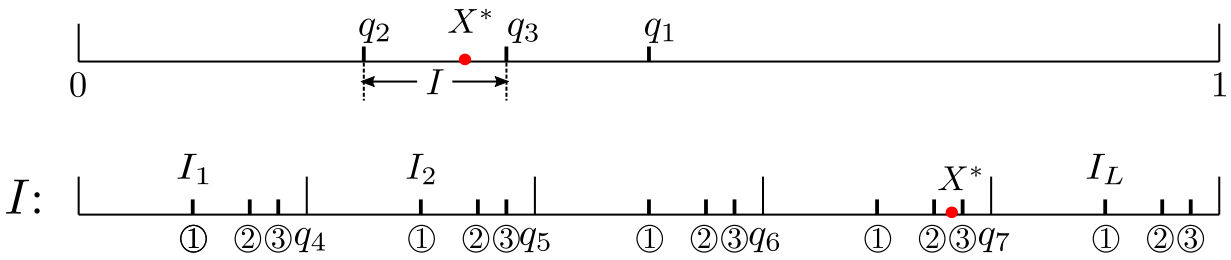

Figure 1: An example of the querying strategy with $L = 5$, $K_1 = 3$, $K_2 = 3$ under the Bayesian setting. The learner first runs K_1 steps of bisection to locate X^* within I. Divide I into L equal length subintervals $I_1, ..., I_L$. By querying the endpoints $q_4, ..., q_7$ of the subintervals, the learner locates the subinterval that contains X^*, in this case I_4. She then proceeds to submit K_2 batches of queries. The first, second and third batches of queries submitted are labeled ①, ②, ③ respectively. On I_4, the queries are submitted via bisection while clones are submitted on the other subintervals in parallel.

Next we show that Algorithm 1 achieves the upper bound in Theorem 1. Under algorithm 1, the total number of queries submitted is

$$K_1 + L - 1 + LK_2 = \left\lfloor \log \frac{1}{L\delta} \right\rfloor + L\left(\lceil \log \frac{\delta}{\epsilon} \rceil + 2 \right) - 1,$$

matching the desired upper bound. It suffices to show that algorithm 1 is both ϵ-accurate and (δ, L)-private. First we establish accuracy. From the responses to all the queries, the learner can narrow down the possible values of X^* to an interval $I^{(\text{final})}$ of length

$$\left| I^{(\text{final})} \right| = \frac{1}{L} 2^{-K_1} 2^{-K_2} = \frac{1}{L} 2^{-(\lfloor \log(1/L\delta) \rfloor + \lceil \log(\delta/\epsilon) \rceil + 1)} \leq \frac{1}{L} 2^{-\log(1/L\epsilon)} = \epsilon. \tag{3}$$

The learner can then take $\widehat{X}$ to be the midpoint of this interval so that $|\widehat{X} - X^*| \leq \epsilon/2$.

Next we show privacy. Recall that the learner performs parallel bisections on the L intervals $I_1, ..., I_L$. Since the adversary only observes the queries and the querying strategy ϕ, she learns that X^* is contained in one of L intervals $J_1, ..., J_L$ where $J_j = [a_j, b_j] \subseteq I_j$. But she cannot tell which one of them X^* is in. Therefore she cannot guess the location of X^* with probability higher than $1/L$. More precisely, the posterior distribution of

Algorithm 1: Our querying strategy under the Bayesian setting

$K_1 := \lfloor \log(1/(L\delta)) \rfloor$; $I := [0,1]$;
for *i = 1 to* K_1 **do** `// bisection to an interval` I `of length` 2^{-K_1}
  $q_i :=$ the midpoint of $I = [a,b]$;
  if $r_i = 1$ **then** $I := [q_i, b]$ **else** $I := [a, q_i]$;
for *i in 1 to L* **do** `// divide` I `into` L `equal-length subintervals` $I_1, ..., I_L$
  $I_i := [a + (i-1)(b-a)/L, a + i(b-a)/L]$, where $[a,b] = I$;
  $J_i := I_i$;
for *i in 1 to L-1* **do** `// query the endpoints of` $I_1, ..., I_L$
  $q_{K_1+i} :=$ the right endpoint of I_i;
Inspect the responses to find $i^* \in \{1, ..., L\}$ such that $X^* \in I_{i^*}$;
$K_2 := \lceil \log(\delta/\epsilon) \rceil + 1$;
for *i in 1 to* K_2 **do** `// replicated bisection on` $I_1, ..., I_L$
  for *j in 1 to L* **do**
    $q_{K_1+L-1+(i-1)L+j} :=$ the midpoint of J_j;
  if $r_{K_1+L-1+(i-1)L+i^*} = 1$ **then**
    for *j in 1 to L* **do** the left endpoint of $J_j := q_{K_1+L-1+(i-1)L+j}$;
  else
    for *j in 1 to L* **do** the right endpoint of $J_j := q_{K_1+L-1+(i-1)L+j}$;
$\widehat{X} :=$ the midpoint of J_{i^*};

X^* given all the query sequence is uniform over the union of $J_1, ..., J_L$. Use $|\cdot|$ to denote the Lebesgue measure of subsets of $[0,1]$. We have

$$\mathbb{P}\{|\widetilde{X} - X^*| \le \delta/2 \mid \text{the query sequence}\} = \frac{\left|(\cup_{i \le L} J_i) \cap \left[\widetilde{X} - \delta/2, \widetilde{X} + \delta/2\right]\right|}{|\cup_{i \le L} J_i|}. \tag{4}$$

Since the queries on $I_1, ..., I_L$ are exact copies of each other, $J_1, ..., J_L$ are also equidistant translations on the real line. The left endpoints $a_1, ..., a_L$ of $J_1, ..., J_L$ satisfy $a_{i+1} = a_i + |I_1|$ for all i where $|I_1| = 2^{-K_1}/L \ge \delta$. Moreover, note that the lengths of all J_i are equal, and because the adversary does not observe the response to the last batch of queries, $|J_i| = 2|I^{(\text{final})}|$. From (3) we have $|J_i| \le 2\epsilon$. Therefore under the assumption that $\delta \ge 2\epsilon$, any interval of length δ can only intersect with $\cup_i J_i$ on a set of Lebesgue measure at most $|J_1|$. Deduce that the right hand side of (4) is upper bounded by $|J_1|/|\cup_i J_i| = 1/L$. Therefore

$$\mathbb{P}\{|\widetilde{X} - X^*| \le \delta/2\} = \mathbb{E}\left(\mathbb{P}\{|\widetilde{X} - X^*| \le \delta/2 \mid \text{the queries}\}\right) \le 1/L.$$

Lower bound: Suppose ϕ is an ϵ-accurate and (δ, L)-private strategy that submits at most n queries. Denote $\mathbf{n}(X^*, Y)$ as the number of queries submitted when X^* is the truth and the random seed is Y, so $n = \sup_{X^*,Y} \mathbf{n}(X^*, Y)$. The goal is to bound n from below. Consider the querying strategy $\widetilde{\phi}$ that concatenates trivial queries at 0 to the query sequence so that the length of query sequence is always n, *i.e.*, $\widetilde{q}_i = q_i$ for $i \le \mathbf{n}(X^*, Y)$ and $\widetilde{q}_i = 0$ for $\mathbf{n}(X^*, Y) < i \le n$. Clearly $\widetilde{\phi}$ is also ϵ-accurate and (δ, L)-private, because the trivial queries at 0 do not provide the adversary with any extra information. Moreover the maximum number of queries submitted by $\widetilde{\phi}$ equals that submitted by ϕ. Hence for the rest of this proof, without loss of generality, we can assume that the learner always submits exactly n queries under ϕ.

Since ϕ is (δ, L)-private, we have $\mathbb{P}\{|\widetilde{X} - X^*| \le \delta/2\} \le 1/L$ for each adversary $\widetilde{X}$. Consider the adversary that adopts the *truncated proportional-sampling* strategy described in Section 4.1 in the main text: let $\widetilde{X} = q_J$ where $J \sim \text{Unif}\{K+1, ..., n\}$. Choose $K = \lfloor \log(1/(L\delta)) \rfloor$. Let us point out that n must be larger than K so truncated proportional-sampling can be run. We will show later in the proof that $n > K$ always holds for any strategy ϕ that is ϵ-accurate. By construction,

$$\mathbb{P}\left\{\left|\widetilde{X} - X^*\right| \le \delta/2\right\} = \mathbb{E}\frac{\sum_{i=K+1}^{n} \mathbb{1}\{|q_i - X^*| \le \delta/2\}}{n - K} \le \frac{1}{L}.$$

Deduce that

$$n \geq K + L\left(\sum_{i=1}^{n} \mathbb{P}\left\{|q_i - X^*| \leq \frac{\delta}{2}\right\} - \sum_{i=1}^{K} \mathbb{P}\left\{|q_i - X^*| \leq \frac{\delta}{2}\right\}\right).$$

We claim that

(i) $\sum_{i\leq n} \mathbb{P}\{|q_i - X^*| \leq \delta/2\} \geq \log(\delta/4\epsilon)$.

(ii) $\sum_{i\leq K} \mathbb{P}\{|q_i - X^*| \leq \delta/2\} \leq 1/L$.

(iii) $n > K$, so that the truncated proportional-sampling strategy is valid.

The desired lower bound immediately follows.

Proof of (i) and (iii): The statement (i) claims that on average, there are at least $\log(\delta/4\epsilon)$ queries in the interval $[X^* - \delta/2, X^* + \delta/2]$. One would expect this to be true because $[X^* - \delta/2, X^* + \delta/2]$ is an interval of length δ. In order for the learner to achieve ϵ-accuracy, it needs to submit at least $\log(\delta/\epsilon)$ queries by optimality of the bisection method. Next we make this argument rigorous. The randomness of the interval $[X^* - \delta/2, X^* + \delta/2]$ complicates the proof. We will instead show something stronger than (i). We claim that for each fixed interval $I \subseteq [0, 1]$, we have

$$\sum_{i\leq n} \mathbb{P}\{q_i \in I \mid X^* \in I\} \geq \log(|I|/2\epsilon). \tag{5}$$

To see why (i) follows from (5), note that for each interval I of length $\delta/2$,

$$\sum_{i\leq n} \mathbb{P}\left\{|q_i - X^*| \leq \delta/2 \mid X^* \in I\right\} \geq \sum_{i\leq n} \mathbb{P}\{q_i \in I \mid X^* \in I\} \geq \log(|I|/2\epsilon) = \log(\delta/4\epsilon).$$

Moreover, claim (iii) also follows from (5) by taking $I = [0, 1]$:

$$n = \sum_{i\leq n} \mathbb{P}\{q_i \in [0, 1]\} \geq \log(1/2\epsilon) > \lfloor \log(1/(L\delta)) \rfloor = K,$$

where the strict inequality holds because by assumption $2\epsilon \leq \delta$ and $L \geq 2$.

It remains to show (5). Since ϕ is ϵ-accurate, we have

$$\mathbb{P}\left\{\left|\widehat{X} - X^*\right| > \epsilon/2 \mid X^* \in I, Y = y\right\} = 0$$

for all but a negligible (zero-measure) set of the random seed Y, denoted as $\mathcal{N}^y$. For $y \notin \mathcal{N}^y$, conditioning on $Y = y$, the estimator $\widehat{X}$ is only a function of the responses $r_1, ..., r_n$. Further conditioning on $X^* \in I$, since X^* is independent from the random seed Y, X^* is distributed uniform in I. By the continuous version of Fano inequality [2, Proposition 2],

$$\mathbb{P}\left\{\left|\widehat{X} - X^*\right| > \epsilon/2 \mid X^* \in I, Y = y\right\} \geq 1 - \frac{I(X^*; r_1, ..., r_n \mid X^* \in I, Y = y) + 1}{\log(|I|/\epsilon)}.$$

Hence

$$H(r_1, ..., r_n \mid X^* \in I, Y = y) \geq I(X^*; r_1, ..., r_n \mid X^* \in I, Y = y) \geq \log(|I|/\epsilon) - 1 = \log(|I|/2\epsilon). \tag{6}$$

Using the entropy chain rule, the left hand side can also be written as

$$\begin{aligned} &H(r_1, ..., r_n \mid X^* \in I, Y = y) \\ =&H(r_1 | X^* \in I, Y = y) + \sum_{i=1}^{n-1} H(r_{i+1} \mid X^* \in I, Y = y, r_1, ..., r_i). \end{aligned} \tag{7}$$

Expand each summand:

$$\begin{aligned}&H\left(r_{i+1} \mid X^* \in I, Y=y, r_1, ..., r_i\right)\\ =&\sum_{\rho_1,..,\rho_i} \mathbb{P}\left\{r_1=\rho_1, ..., r_i=\rho_i \mid X^* \in I, Y=y\right\} H\left(r_{i+1} \mid X^* \in I, Y=y, r_1=\rho_1, ..., r_i=\rho_i\right). \end{aligned} \tag{8}$$

Write $I=[a,b]$. On the event $X^* \in I$, if $q_{i+1}=\phi_i(\rho_1, ..., \rho_i, y)$ is smaller than a, then $r_{i+1}=1$. Similarly if $q_{i+1}>b$, then $r_{i+1}=0$. In other words, the value of r_{i+1} is completely determined by $\rho_1, ..., \rho_i$ if $q_{i+1} \notin I$ and $X^* \in I$. Hence the summation (8) equals

$$\begin{aligned}&\sum_{\rho_1,..,\rho_i:\phi(\rho_1,...,\rho_i,y)\in I} \mathbb{P}\left\{r_1=\rho_1, ..., r_i=\rho_i \mid X^* \in I, Y=y\right\} H\left(r_{i+1} \mid X^* \in I, Y=y, r_1=\rho_1, ..., r_i=\rho_i\right)\\ \leq&\mathbb{P}\left\{q_{i+1} \in I \mid X^* \in I, Y=y\right\}.\end{aligned}$$

With $Y=y$ fixed, we have $q_1=f_0(y)$ is deterministic. Similarly argue that $H(r_1|X^*, Y=y) \leq \mathbb{1}\{q_1 \in I\}$. Combine with (6) and (27) to deduce that

$$\sum_{i\leq n} \mathbb{P}\{q_i \in I \mid X^* \in I, Y=y\} \geq H(r_1, ..., r_n \mid X^* \in I, Y=y) \geq \log(|I|/2\epsilon).$$

The above holds for all $y \notin \mathcal{N}^y$. Since $\mathcal{N}^y$ is a negligible set, we have

$$\sum_{i\leq n} \mathbb{P}\{q_i \in I \mid X^* \in I\} = \int_{[0,1]\setminus\mathcal{N}^y} \sum_{i\leq n} \mathbb{P}\{q_i \in I \mid X^* \in I, Y=y\} dy \geq \log(|I|/2\epsilon).$$

The proof of (5) is complete.

Proof of (ii): To show (ii) we introduce the notion of *learner intervals*, which stands for the sequence of intervals that the learner knows X^* is in, as the learner submits queries sequentially. Start from $I_0=[0,1]$. If $r_1=1$, then the learner learns that $X^* \in [q_1, 1]$ and I_1 is defined as $[q_1, 1]$. Otherwise $I_1=[0, q_1]$. For all i,

$$\begin{aligned}\mathbb{P}\left\{|q_i - X^*| \leq \frac{\delta}{2}\right\} =&\mathbb{E}\left[\mathbb{P}\left\{|q_i - X^*| \leq \frac{\delta}{2} \mid r_1, ..., r_{i-1}\right\}\right]\\ =&\mathbb{E}\left[\frac{|I_{i-1} \cap [q_i - \delta/2, q_i + \delta/2]|}{|I_{i-1}|}\right]\\ \leq&\delta\mathbb{E}(1/|I_{i-1}|).\end{aligned}$$

Next we show that $\mathbb{E}(1/|I_i|) \leq 2^i$ for all i by induction. Suppose it is true for $i=0, ..., k$. For $i=k+1$,

$$\mathbb{E}(1/|I_{k+1}|) = \mathbb{E}\left[\mathbb{E}\left(1/|I_{k+1}| \,\middle|\, r_1, ..., r_k\right)\right].$$

Conditioning on $r_1, ..., r_k$, the learner interval I_k is deterministic and so is q_{k+1}. Let $I_k=[a_k, b_k]$. There are three possibilities for I_{k+1}:

1. $q_{k+1} \notin I_k$. In this case the r_{k+1} provides no additional information on the location of X^*. Therefore $I_{k+1}=I_k$.
2. $q_{k+1} \in I_k$ and $r_{k+1}=1$. The learner learns that $X^* \geq q_{k+1}$ and $I_{k+1}=[q_{k+1}, b_k]$.
3. $q_{k+1} \in I_k$ and $r_{k+1}=0$. In this case $I_{k+1}=[a_k, q_{k+1}]$.

Therefore

$$\begin{aligned}\mathbb{E}\left(1/|I_{k+1}| \mid r_1, ..., r_k\right) =&\mathbb{1}\left\{q_{k+1} \notin I_k\right\} \frac{1}{|I_k|} + \mathbb{1}\left\{q_{k+1} \in I_k\right\} \mathbb{P}\left\{X^* \geq q_{k+1} \mid r_1, ..., r_k\right\} \frac{1}{b_k - q_{k+1}}\\ &+ \mathbb{1}\left\{q_{k+1} \in I_k\right\} \mathbb{P}\left\{X^* < q_{k+1} \mid r_1, ..., r_k\right\} \frac{1}{q_{k+1} - a_k}\\ =&\mathbb{1}\left\{q_{k+1} \notin I_k\right\} \frac{1}{|I_k|} + \mathbb{1}\left\{q_{k+1} \in I_k\right\} \frac{1}{|I_k|} + \mathbb{1}\left\{q_{k+1} \in I_k\right\} \frac{1}{|I_k|} \leq \frac{2}{|I_k|}.\end{aligned}$$

Hence $\mathbb{E}(1/|I_{k+1}|) \leq \mathbb{E}(2/|I_k|) \leq 2 \cdot 2^k = 2^{k+1}$. Deduce that $\mathbb{P}\{|q_i - X^*| \leq \delta/2\} \leq \delta \mathbb{E}(1/|I_{i-1}|) \leq \delta 2^{i-1}$ for all i. Therefore

$$\sum_{i=1}^{K} \mathbb{P}\left\{|q_i - X^*| \leq \frac{\delta}{2}\right\} \leq \delta \sum_{i=1}^{K} 2^{i-1} \leq \delta 2^K \leq 1/L,$$

where the last inequality is from $K = \lfloor \log(1/(L\delta)) \rfloor \leq \log(1/(L\delta))$. □

2.2 Analysis under the deterministic setting

Proof of Theorem 2. **Upper bound:** Recall that under the deterministic setting, a querying strategy is called (δ, L)-private if for each query sequence $\bar{q}$, the δ-covering number of the information set $\mathcal{I}(\bar{q})$ is at least L. To achieve (δ, L)-privacy, we design a strategy where the learner submits L guesses. Recall from Section 4 in the main text that a guess at q consists of a pair of queries at q and $q + \epsilon$. Below is the construction of our querying strategy. The precise description of the querying strategy is given in algorithm 2 (for $\delta \leq 2^{-L}$) and algorithm 3 (for $\delta > 2^{-L}$).

1. Submit L guesses that are at least δ apart. This further breaks down into two cases, depending on the value of δ:

   If $\delta \leq 2^{-L}$, submit the guesses via a bisection search. That is, the first guess is at $1/2$, the second guess is at $1/4$ if $X^* < 1/2$ and at $3/4$ otherwise, etc. However if at any point a guess turns out to be correct ($X^* \in [s, s + \epsilon)$ for a guess at s), then in order to hide this knowledge from the adversary, the learner keeps submitting guesses via a fake bisection search using random responses distributed *i.i.d.* Bernoulli(1/2).

   If $\delta > 2^{-L}$, submit the first guess at 0. The next K guesses are submitted via a bisection search, locating X^* in a interval I of length 2^{-K}. As in the previous case, transition into a fake bisection search whenever a guess is found to contain X^*. Submit the rest of the $(L - K - 1)$ guesses through a grid search on I. Here K is chosen to be an integer in $\{0, 1, ..., L - 1\}$ for which $2^{-K}/(L - K) \in [\delta, 2\delta]$. We will show that such K always exists. In fact the initial guess at 0 is to ensure existence of an integer solution for K. This way of submitting guesses ensures that the closest pair of guesses are made as close as possible, while still being at least δ apart.

2. If none of the L guesses made in stage 1 is correct, then through the L guesses the learner should locate X^* within an interval J of length about $\max\{2^{-L}, \delta\}$. Run a bisection search on J until ϵ-accuracy is reached. If any of the guesses is correct, replace this step with a fake bisection search on a simulated interval J. When $\delta \leq 2^{-L}$, J is obtained from the last step of the fake bisection search in stage 1; when $\delta > 2^{-L}$, J is selected from the $L - K$ subintervals of I uniformly at random.

Examples of the above querying strategy is illustrated in Fig. 2 (when $\delta \leq 2^{-L}$) and Fig. 3 (when $\delta > 2^{-L}$).

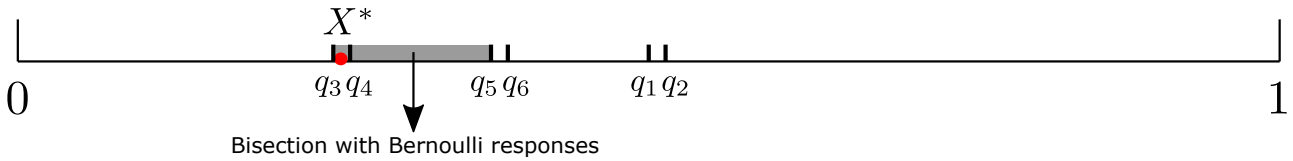

Figure 2: An example of the querying strategy under the deterministic setting when $\delta \leq 2^{-L}$, with $L = 3$. From the response to the first four queries the learner deduces that X^* is between q_3 and $q_4 = q_3 + \epsilon$. The learner proceeds to run a "fake" bisection in $[q_3, q_5)$ by generating Bernoulli responses to confuse the adversary. From the perspective of the adversary, X^* could be in any of the three length-ϵ subintervals.

Next we show that the querying strategy above achieves the upper bound in Theorem 2. First consider the case $\delta \leq 2^{-L}$. Under algorithm 2, the learner first submits L guesses ($2L$ queries). She then conducts a bisection search within an interval of length 2^{-L}, taking $\lceil 2^{-L}/\epsilon \rceil$ queries to achieve ϵ-accuracy. The total number of queries submitted is $L + \lceil \log(1/\epsilon) \rceil$.

Because the guesses $I_1, ..., I_L$ are all of length ϵ, algorithm 2 is ϵ-accurate. It suffices to show it is also (δ, L)-private. Firstly, note that the adversary cannot rule out the possibility that $X^* \in I_i$ for some $i = 1, .., L$, so the

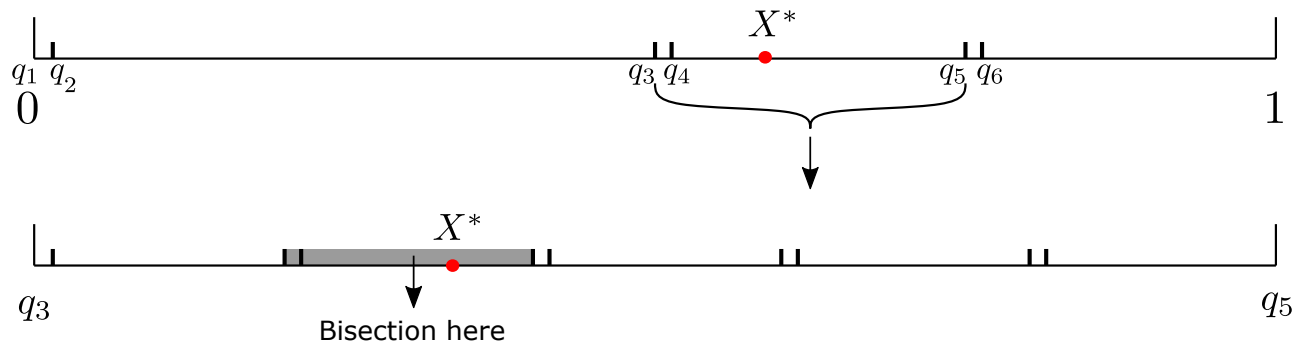

Figure 3: An example of the querying strategy under the deterministic setting when $\delta > 2^{-L}$, with $L = 7$ and $K = 2$. The first guess is at 0. The learner submits the next K guesses via bisection to locate X^* in $[q_3, q_5)$. She then partitions $[q_3, q_5)$ into $L - K$ equal length subintervals. Eventually through bisection the learner is able to approximate X^* up to accuracy ϵ. But from the perspective of the adversary, X^* could be in $[q_{2i-1}, q_{2i})$ for any $i \leq L$.

information set contains the union of $I_1, ..., I_L$. That is, for each query sequence q,

$$\mathcal{I}(q) \supseteq \cup_{i \leq L}[q_{2i-1}, q_{2i}).$$

When $\delta \leq 2^{-L}$, these L intervals do not overlap. Since their left endpoints are submitted via a bisection search, legitimate or fake, they are at least $2^{-L} \geq \delta$ apart from each other. Therefore the δ-covering number for $\mathcal{I}(q)$ is at least L. We have shown that algorithm 2 is (δ, L)-private.

Next consider the case $\delta > 2^{-L}$. Again algorithm 3 is clearly ϵ-accurate. To show that it is also (δ, L)-private, note that algorithm 3 is designed so that the closest pair of guesses are of distance $[\delta, 2\delta]$ apart. Hence

(i) The intervals $I_i = [q_{2i-1}, q_{2i})$, $i = 1, ..., L$ do not overlap, and their left endpoints are at least δ from each other;

(ii) After the L guesses are submitted, the learner can always narrow down the possibilities for X^* to an interval of length at most 2δ.

We claim that (i) ensures (δ, L)-privacy. As in the $\delta \leq 2^{-L}$ case, we have for each q, $\mathcal{I}(q) \supseteq \cup_{i \leq L}[q_{2i-1}, q_{2i})$. Assuming (i), the δ-covering number of $\cup_{i \leq L}[q_{2i-1}, q_{2i})$ is at least L.

Given (ii), the learner only needs to submit at most $\lceil \log(2\delta/\epsilon) \rceil$ queries to achieve ϵ-accuracy in stage 2. The total number of queries submitted under algorithm 3 is at most $2L + \lceil \log(\delta/\epsilon) \rceil + 1$. Moreover, as we can see from algorithm 3 the learner always submits $q_1 = 0$. Omit this first trivial query to obtain the desired query complexity upper bound $2L + \lceil \log(\delta/\epsilon) \rceil$.

We still need to show that (i) and (ii) are satisfied by algorithm 3. The first K guesses locate X^* within an interval I of length 2^{-K}. The remaining $L - K - 1$ odd queries then divide I into $L - K$ subintervals of equal length. Therefore the closest pair of odd queries among $q_1, q_3, ..., q_{2L-1}$ are at distance $2^{-K}/(L - K)$. In stage 2, the learner conducts a bisection search in one of the $L - K$ subintervals, which is also of length $2^{-K}/(L - K)$. Therefore (i) and (ii) translate to $\delta \leq 2^{-K}/(L - K) \leq 2\delta$. It remains to show that we can find at least one $K \in \{0, 1, ..., L - 1\}$ for which

$$\ell_K := \frac{2^{-K}}{L - K} \in [\delta, 2\delta]. \tag{9}$$

Observe that

1. $\ell_0 = 1/L \geq \delta$;
2. $\ell_{L-1} = 2^{-(L-1)} \leq 2\delta$;

Algorithm 2: Our querying strategy under the deterministic setting when $\delta \leq 2^{-L}$

GuessedIt := *False*;
$I := [0, 1]$;
for *i = 1 to L* **do** `// submit L guesses via (possibly partially fake) bisection`
  q_{2i-1} := the midpoint of $I = [a, b]$;
  $q_{2i} := q_{2i-1} + \epsilon$;
  if *not GuessedIt* **then**
    Inspect the responses r_{2i-1} and r_{2i};
    if $r_{2i-1} = 1$ **then** $I := [q_{2i-1}, b]$ **else** $I := [a, q_{2i-1}]$;
    if $r_{2i-1} = 1$ *and* $r_{2i} = 0$ **then** *GuessedIt* := *True*;
  else `// once guessed correctly, proceed with a fake bisection`
    Sample $R \sim$ Bernoulli(1/2);
    if $R = 1$ **then** $I := [q_{2i-1}, b]$ **else** $I := [a, q_{2i-1}]$;
$i := 2L + 1$, $J := I$;
while $|J| > \epsilon$ **do** `// run (possibly fake) bisection on J`
  q_i := the midpoint of $J = [a, b]$;
  if *not GuessedIt* **then**
    if $r_i = 1$ **then** $J := [q_i, b]$ **else** $J := [a, q_i]$;
  else
    Sample $R \sim$ Bernoulli(1/2);
    if $R = 1$ **then** $J := [q_i, b]$ **else** $J := [a, q_i]$;
  $i := i + 1$;
$\widehat{X}$:= the midpoint of J;

3. for all $K < L - 1$,

$$\frac{\ell_K}{\ell_{K+1}} = \frac{2^{-K}}{2^{-(K+1)}} \frac{L - K - 1}{L - K} \leq 2.$$

These facts above ensure that there is at least one solution to (9) in $\{0, 1, ..., L - 1\}$.

Lower bound: The lower bound $\lceil \log(\delta/\epsilon) \rceil + 2L - 4$ has already been proven in [16, Theorem 4.1]. As in the upper bound proof we separately consider the cases $\delta > 2^{-L}$ and $\delta \leq 2^{-L}$. When $\delta > 2^{-L}$, the term $\lceil \log(\delta/\epsilon) \rceil + 2L - 4$ is always larger than $\lceil \log(1/\epsilon) \rceil + L - 8$. Thus we only need to show that when $\delta > 2^{-L}$, optimal query complexity is lower bounded by $\lceil \log(1/\epsilon) \rceil + L - 8$.

It suffices to show the lower bound holds for all realizations of the random seed Y so the dependences on Y are suppressed for the rest of the proof. Fix any querying strategy ϕ that is both ϵ-accurate and (δ, L)-private. Let $\mathcal{Q}(X^*)$ denote the set of queries when the true value is X^*. We want to show there is at least one X^* for which $|\mathcal{Q}(X^*)| \geq L + \log(1/\epsilon) - 8$. To this end, we will prove the following claims:

(i) There exists an interval I of length 2δ and $\widetilde{\mathcal{Q}} = \{\widetilde{q}_1, ..., \widetilde{q}_K\}$ where $K \geq \log(1/\delta) - 3$ and $|\widetilde{q}_i - \widetilde{q}_j| > \delta$ for all $i \neq j$, such that for each $X^* \in I$, $\mathcal{Q}(X^*) \backslash I \supseteq \widetilde{\mathcal{Q}}$.

(ii) For each interval I of length 2δ and each $X^* \in I$, there exist at least $L - 5$ distinct pairs of queries $\{s_1, t_1\}, ...\{s_{L-5}, t_{L-5}\} \subseteq \mathcal{Q}(X^*) \backslash I$ for which $|s_i - t_i| \leq \epsilon$ for all i.

(iii) For each interval I of length 2δ, there exists $X^* \in I$ such that $\mathcal{Q}(X^*)$ contains at least $\log(\delta/\epsilon)$ queries in I.

Claims (i) and (ii) together imply that there exists an interval I of length 2δ such that for all $X^* \in I$,

$$\mathcal{Q}(X^*) \backslash I \supseteq \widetilde{\mathcal{Q}} \cup (\cup_{i \leq L-5} \{s_i, t_i\}) .$$

Since all members of $\widetilde{\mathcal{Q}}$ are at least δ-apart and $|s_i - t_i| \leq \epsilon$, at least one of s_i and t_i is outside of $\widetilde{\mathcal{Q}}$. To show that on top of $\widetilde{\mathcal{Q}}$, each pair $\{s_i, t_i\}$ contributes at least one extra member to $\mathcal{Q}(X^*) \backslash I$, we only need to rule out

Algorithm 3: Our querying strategy under the deterministic setting when $\delta > 2^{-L}$

$q_1 := 0; q_2 := \epsilon$; `// submit initial guess at 0`
$K :=$ an integer solution in $\{0, 1, ..., L-1\}$ to $\ell_K = 2^{-K}/(L-K) \in [\delta, 2\delta]$;
if $r_1 = 1$ *and* $r_2 = 0$ **then** $GuessedIt := True$ **else** $GuessedIt := False$;
$I := [0, 1]$;
for *i=2 to K+1* **do** `// submit the next` K `guesses via bisection`
  $q_{2i-1} :=$ the midpoint of $I = [a, b]$;
  $q_{2i} := q_{2i-1} + \epsilon$;
  if *not GuessedIt* **then**
    if $r_{2i-1} = 1$ **then** $I := [q_{2i-1}, b]$ **else** $I := [a, q_{2i-1}]$;
    if $r_{2i-1} = 1$ *and* $r_{2i} = 0$ **then** $GuessedIt := True$;
  else
    Sample $R \sim$ Bernoulli(1/2);
    if $R = 1$ **then** $I := [q_{2i-1}, b]$ **else** $I := [a, q_{2i-1}]$;
Divide I into $L - K$ equal length subintervals $I_1, ..., I_{L-K}$;
for *i = (K+2) to L* **do** `// submit the next` $L - K - 1$ `guesses via grid search`
  $q_{2i-1} :=$ the right endpoint of I_{i-K-1};
  $q_{2i} := q_{2i-1} + \epsilon$;
  if $r_{2i-1} = 1$ *and* $r_{2i} = 0$ **then** $GuessedIt := True$;
if *not GuessedIt* **then**
  $J :=$ the subinterval I_{i^*} that contains X^*;
else
  $J := I_{i^*}$ where i^* is sampled uniformly from $\{1, ..., L-K\}$;
$i := 2L + 1$;
while $|J| > \epsilon$ **do** `// run (possibly fake) bisection on` J
  $q_i :=$ the midpoint of $I = [a, b]$;
  if *not GuessedIt* **then**
    if $r_i = 1$ **then** $J := [q_i, b]$ **else** $J := [a, q_i]$;
  else
    Sample $R \sim$ Bernoulli(1/2);
    if $R = 1$ **then** $J := [q_i, b]$ **else** $J := [a, q_i]$;
  $i := i + 1$;
$\widehat{X} :=$ the midpoint of J;

the case where two pairs $\{s_i, t_i\}$ and $\{s_j, t_j\}$ are such that $s_i, s_j \in \widetilde{\mathcal{Q}}$ and $t_i = t_j$. This cannot happen because otherwise,

$$\delta < |s_i - s_j| \le |s_i - t_i| + |s_j - t_i| = |s_i - t_i| + |s_j - t_j| \le \epsilon + \epsilon,$$

contradicting the assumption $\delta \ge 2\epsilon$. Thus $\mathcal{Q}(X^*)\backslash I$ contains at least $K + L - 5$ distinct members. From claim (iii) there exists $X^* \in I$ for which $|\mathcal{Q}(X^*) \cap I| \ge \log(\delta/\epsilon)$. We have

$$|\mathcal{Q}(X^*)| = |\mathcal{Q}(X^*) \cap I| + |\mathcal{Q}(X^*)\backslash I| \ge \log(\delta/\epsilon) + K + L - 5 \ge L + \log\frac{1}{\epsilon} - 8,$$

which equals $2L + \log(\max\{2^{-L}, \delta\}/\epsilon) - 8$ when $\delta \le 2^{-L}$. It remains to prove the three claims.

Proof of (i): To prove this claim, we first construct for each $X^* \in [0, 1]$ a subsequence $\widetilde{q}$ of q where all the queries in $\widetilde{q}$ are at least δ apart from each other.

Let $\widetilde{q}_1 = q_1$. If $X^* \in [\widetilde{q}_1 - \delta, \widetilde{q} + \delta]$, then declare the construction finished, *i.e.* the subsequence $\widetilde{q} = (\widetilde{q}_1)$ is of length one. Otherwise look at $q_2 = \phi_1(r_1)$. If $q_2 \in [\widetilde{q}_1 - \delta, \widetilde{q}_1 + \delta]$, then $\widetilde{q}_1$ and q_2 must be on the same side of X^* and $r_2 = \mathbb{1}\{X^* \ge q_2\}$ must be equal to r_1. Proceed to look at $q_3 = \phi_2(r_1, r_2) = \phi_2(r_1, r_1)$, $q_4 = \phi_3(r_1, r_1, r_1)$ and so on, until $q_i \notin [\widetilde{q}_1 - \delta, \widetilde{q}_1 + \delta]$. Let $\widetilde{q}_2 = q_i$. Similarly define the rest of $\widetilde{q}$ as follows: For $k \ge 2$ if $\widetilde{q}_k$ is chosen to be q_{i_k}, then let $\widetilde{q}_{k+1} = q_{i_{k+1}}$, where

$$i_{k+1} = \min_j \left\{ j > i_k : q_j \notin \cup_{k' \le k} [\widetilde{q}_{k'} - \delta, \widetilde{q}_{k'} + \delta] \right\}.$$

Repeat this process until $[\widetilde{q}_k - \delta, \widetilde{q}_k + \delta]$ contains X^*. Note that such a k always exists, as ϕ is ϵ-accurate and hence there exists at least one query that is within ϵ distance to X^*.

Let $\widetilde{r}_i = \mathbb{1}\{X^* \ge \widetilde{q}_i\}$. Next we argue that $\widetilde{q}$ is completely determined by $\widetilde{r}$. Indeed, given $\widetilde{r} = (\widetilde{r}_1, ..., \widetilde{r}_k)$, we have $\widetilde{q}_j = q_{i_j}$ for all $j \le k$, where $i_1 = 1$ and

$$i_2 = \min_j \{ j > i_1 : \phi_{j-1}(\widetilde{r}_1, ..., \widetilde{r}_1) \notin [\widetilde{q}_1 - \delta, \widetilde{q}_1 + \delta] \}.$$

Thus $\tilde{q}_2 = q_{i_2} = \phi_{i_2 - 1}(\widetilde{r}_1, ..., \widetilde{r}_1)$. To determine i_3, inspect $q_{i_2+1} = \phi_{i_2}(r_1, ..., r_{i_2}) = \phi_{i_2}(\widetilde{r}_1, ..., \widetilde{r}_1, \widetilde{r}_2)$. If $q_{i_2+1} \notin \cup_{j=1,2}[\widetilde{q}_j - \delta, \widetilde{q}_j + \delta]$, the we have $i_3 = i_2 + 1$. Otherwise if $q_{i_2+1} \in [\widetilde{q}_1 - \delta, \widetilde{q}_1 + \delta]$, then we have $r_{i_2+1} = \widetilde{r}_1$ and $q_{i_2+2} = \phi_{i_2+1}(\widetilde{r}_1, ..., \widetilde{r}_1, \widetilde{r}_2, \widetilde{r}_1)$; similarly if $q_{i_2+1} \in [\widetilde{q}_2 - \delta, \widetilde{q}_2 + \delta]$, then $q_{i_2+2} = \phi_{i_2+1}(\widetilde{r}_1, ..., \widetilde{r}_1, \widetilde{r}_2, \widetilde{r}_2)$. As such we can reconstruct the queries q_{i_2+3}, q_{i_2+4} and so on until we find $j > i_2$ where $\phi_j \notin \cup_{j=1,2}[\widetilde{q}_j - \delta, \widetilde{q}_j + \delta]$. Then we have determined $i_3 = j$ and $\widetilde{q}_3 = q_j$, which is completely determined by $(\widetilde{r}_1, \widetilde{r}_2)$. Following the same argument, the entire $\widetilde{q}$ sequence can be reconstructed from $\widetilde{r}$. Consequently,

$$|\{\widetilde{q} : X^* \in [0, 1]\}| \le |\{\widetilde{r} : X^* \in [0, 1]\}|.$$

Suppose $K + 1$ is the maximum length of $\widetilde{q} \equiv \widetilde{q}(X^*)$ among all $X^* \in [0, 1]$. Then the total number of distinct binary $\widetilde{r}$ sequences is at most $\sum_{k \le K+1} 2^k < 2^{K+2}$. In addition if $\widetilde{q}$ is of length k, then $X^* \in [\widetilde{q}_k - \delta, \widetilde{q}_k + \delta]$ by construction. Hence

$$1 = |[0, 1]| \le \left| \cup_{X^* \in [0,1]} [\widetilde{q}_k - \delta, \widetilde{q}_k + \delta] \right| \le 2\delta \, |\{\widetilde{q} : X^* \in [0, 1]\}| \le 2\delta \cdot 2^{K+2}.$$

Deduce that $K \ge \log(1/\delta) - 3$. In other words, there exists $X^* \in [0, 1]$ for which $\widetilde{q}$ is of length k where $k \ge K + 1 \ge \log(1/\delta) - 2$. We choose $I = [\widetilde{q}_k - \delta, \widetilde{q}_k + \delta]$ for such $\widetilde{q}$ and show that it satisfies the statement in (i). By construction all the queries in $\widetilde{q}$ are more than δ apart; therefore, all the queries in $\widetilde{q}$ except $\widetilde{q}_k$ are all outside of I. As a result for all $X \in I$ and $i \le k - 1$, $\mathbb{1}\{X \ge \widetilde{q}_i\}$ yields the same response as $\mathbb{1}\{X^* \ge \widetilde{q}_i\}$. Deduce that $\widetilde{q}(X) = \widetilde{q}(X^*)$ for all $X \in I$. To complete the proof of (i), take $\widetilde{\mathcal{Q}} = \{\widetilde{q}_1, ..., \widetilde{q}_{k-1}\}$ to obtain a subset of $\mathcal{Q}(X^*)\backslash I$ of size at least $K \ge \log(1/\delta) - 3$.

Proof of (ii): For $q = q(X^*) = (q_1, ..., q_n)$, let $\overline{\mathcal{Q}}(X^*) = \{q_1, ..., q_n, 0, 1\}$. The key observation is that for each x in the information set $\mathcal{I}(q)$, there must be two queries $s, t \in \overline{\mathcal{Q}}(X^*)$ with $s \le x$, $t > x$ and $t - s \le \epsilon$. Otherwise when x is the truth, the learner could not have achieved ϵ-accuracy through the query sequence q. The inclusion of $0, 1$ in $\overline{\mathcal{Q}}(X^*)$ is because even if they are never queried, they could still serve in these (s, t) pairs.

Let

$$\mathcal{P} = \{(s, t) : s, t \in \overline{\mathcal{Q}}(X^*), 0 < t - s \le \epsilon\}$$

denote the set of all pairs of queries that are no more than ϵ-apart. We have

$$\mathcal{I}(q) \subseteq \cup_{(s,t)\in\mathcal{P}}[s,t].$$

From the definition of (δ, L)-privacy, the δ-covering number of $\cup_{(s,t)\in\mathcal{P}}[s,t]$ is at least L, which immediately implies $|\mathcal{P}| \geq L$. However since we want to lower bound the number of pairs (s,t) where both s and t are outside of I, the proof is slightly more complicated. Write $I = [a,b]$. If one of s,t is in I, then $[s,t] \subseteq [a-\epsilon, b+\epsilon] \subseteq [a-\delta/2, b+\delta/2]$. This is an interval of length 3δ. We also need to discount the pairs that use 0 or 1 as one of the endpoints. Let

$$\begin{aligned}\widetilde{\mathcal{P}} =&\{(s,t) : s,t \in \overline{\mathcal{Q}}(X^*)\backslash(I\cup\{0,1\}), 0 < t-s \leq \epsilon\}\\ \supseteq&\{(s,t)\in\mathcal{P} : [s,t]\subseteq[0,1]\backslash([a-\delta/2, b+\delta/2]\cup[0,\delta]\cup[1-\delta,1])\}.\end{aligned}$$

The δ-covering number for $[a-\delta/2, b+\delta/2]\cup[0,\delta]\cup[1-\delta,1])$ is at most 5. Deduce that the δ-covering number for $\cup_{(s,t)\in\widetilde{\mathcal{P}}}[s,t]$ is at least $L-5$. Thus $|\widetilde{\mathcal{P}}| \geq L-5$.

Proof of (iii): The part of the proof is similar to the proof of (i). We take $\widetilde{q}(X^*)$ to be the subsequence of $q(X^*)$ that contains all the queries in $\mathcal{Q}(X^*)$ that are in I. Let $J(X^*)$ be the interval formed by the two queries in $q(X^*)$ to the left and right of X^* that are the closest to X^*. For all $X^* \in I$, $X^* \in J(X^*)$ and thus $I \subseteq \cup_{X^*\in I} J(X^*)$. Since $|I| = 2\delta$ and the querying strategy ϕ is ϵ-accurate so that $|J(X^*)| \leq \epsilon$, we have that $\{J(X^*) : X^* \in I\}$ contains at least $2\delta/\epsilon$ distinct members.

Let $\widetilde{r}_i(X^*) = \mathbb{1}\{X^* \geq \widetilde{q}_i(X^*)\}$. Next we show that for each $X^* \in I$, $J(X^*)$ is completely determined by $\widetilde{r}(X^*)$. Indeed given any $X^* \in I$, the responses to the queries outside of I can be deduced from their position relative to I. Therefore from only $\widetilde{r}(X^*)$, which only contains responses to the queries in I, one can reconstruct the entire query sequence $q(X^*)$, from which one can infer $J(X^*)$. Thus

$$|\{\widetilde{r}(X^*) : X^* \in I\}| \geq |\{J(X^*) : X^* \in I\}| \geq 2\delta/\epsilon.$$

Suppose T is the maximal length of $\widetilde{q}(X^*)$ among all $X^* \in I$. Then $\widetilde{r}(X^*)$ can take no more than $\sum_{t\leq T} 2^t < 2^{T+1}$ distinct values. Deduce that $T \geq \log(\delta/\epsilon)$. Recall that $\widetilde{q}(X^*)$ is a subsequence of $q(X^*)$ and only contains queries in I. Conclude that there exists $X^* \in I$ for which the querying strategy ϕ submits at least $\log(\delta/\epsilon)$ queries that are in I. □

3 Proof of Theorem 3 (noisy responses)

3.1 Proof of the upper bounds

The construction of our querying strategies with noisy responses relies heavily on an existing search algorithm known as the Burnashev-Zigangirov(BZ) algorithm [1]. For completeness, we give in Section 3.1.1 a brief description of the BZ algorithm and its statistical properties.

3.1.1 Background: the Burnashev-Zigangirov algorithm

Suppose $[0,1]$ is divided into $1/\Delta$ (assumed to be an integer) equal length subintervals, labeled $I_1, ..., I_{1/\Delta}$ from left to right. Let J denote the subinterval that contains the true value X^*. The BZ algorithm is a selection procedure that returns $\widehat{J}$, an estimator of J.

Since X^* is distributed uniformly on $[0,1]$, the algorithm starts from a uniform distribution μ_1 on $[0,1]$, which can be viewed as a priori belief distribution on the location of X^*. Each time the learner observes a response R_j to a query X_j, where $R_j \sim$ Bernoulli(p) if $X^* \geq X_j$ and $R_j \sim$ Bernoulli$(1-p)$ if $X^* < X_j$ for some $p \in (1/2, 1)$. Then the belief distribution is updated as follows:

$$\frac{d\mu_{j+1}}{d\mu_j}(x) = \begin{cases} \frac{2(1-\alpha)\mathbb{1}\{x\in[0,X_j)\}+2\alpha\mathbb{1}\{x\in[X_j,1]\}}{\int(2(1-\alpha)\mathbb{1}\{y\in[0,X_j)\}+2\alpha\mathbb{1}\{y\in[X_j,1]\})\mu_j(dy)} & \text{if } R_j = 1;\\ \frac{2\alpha\mathbb{1}\{x\in[0,X_j)\}+2(1-\alpha)\mathbb{1}\{x\in[X_j,1]\}}{\int(2(1-\alpha)\mathbb{1}\{y\in[0,X_j)\}+2\alpha\mathbb{1}\{y\in[X_j,1]\})\mu_j(dy)} & \text{if } R_j = 0,\end{cases}$$

where $\alpha \in (1/2, p)$ is a parameter whose value will be later specified. Note that if $\alpha = p$, the display above is exactly the posterior update rule for the distribution of X^* given the responses. The intuition behind choosing

$\alpha < p$ is to tilt the update rule in the more conservative direction, so that the effect of "incorrect" responses can be mitigated.

The query X_j is selected to be close to the median of μ_j. Specifically, if $I_j = [s,t)$ is the subinterval that contains the median of μ_j, then X_j is chosen to be the left endpoint s of I_j with probability $\pi_1 = (\mu_j[0,t) - \mu_j[t,1])/(2\mu_j(I_j))$, and $X_j = t$ with probability $\pi_2 = 1 - \pi_1 = (\mu_j[s,1] - \mu_j[0,s))/(2\mu_j(I_j))$. Here π_1 and π_2 are chosen so that the conditional mean of X_j is exactly the median of μ_j. Since the learner only queries the endpoints of the subintervals, the density of μ_j is a piecewise-constant function whose change points can only occur at the endpoints of the subintervals. Suppose n queries are submitted, the estimator $\widehat{J}$ is taken to be the subinterval with the highest μ_{n+1} density, ties broken arbitrarily.

For simplicity write $\overline{p} = 1-p$, $\overline{\alpha} = 1-\alpha$. It has been shown that the error probability of $\widehat{J}$ decreases exponentially in the number of queries [1, Eq (3.24)]:

$$\mathbb{P}\{X^* \notin \widehat{J}\} \le \frac{1-\Delta}{\Delta}\left[\frac{\overline{p}}{2\overline{\alpha}} + \frac{p}{2\alpha}\right]^n \le \frac{1}{\Delta}\left[\frac{\overline{p}}{2\overline{\alpha}} + \frac{p}{2\alpha}\right]^n. \tag{10}$$

The factor $\overline{p}/(2\overline{\alpha}) + p/(2\alpha)$ is minimized at

$$\alpha = \frac{\sqrt{p}}{\sqrt{p}+\sqrt{\overline{p}}}, \quad \text{with } \frac{\overline{p}}{2\overline{\alpha}} + \frac{p}{2\alpha} = \frac{1}{2} + \sqrt{p\overline{p}}.$$

It follows from (10) that

$$\mathbb{P}\{X^* \notin \widehat{J}\} \le \frac{1}{\Delta}\left(\frac{1}{2} + \sqrt{p\overline{p}}\right)^n \le \frac{1}{\Delta}\left(1-(p-1/2)^2\right)^n \le \frac{1}{\Delta}\exp\left(-(p-1/2)^2 n\right). \tag{11}$$

The last two inequalities are due to the basic inequalities $\sqrt{x(1-x)} \le 1/2 - (x-1/2)^2$ for $x \in [0,1]$ and $1 + x \le e^x$ for all $x \in \mathbb{R}$.

The lemma below follows from (11) via a simple scaling argument.

Lemma 1. *Suppose the BZ algorithm is run on an interval I divided into Δ-length subintervals. The output estimator $\widehat{J}$ satisfies*

$$\mathbb{P}\{X^* \notin \widehat{J}\} \le \frac{|I|}{\Delta} 2^{-c_3(p)n},$$

where $c_3(p) = (p-1/2)^2 \log e$.

3.1.2 Proof of the upper bound in (1)

The idea behind the construction of the querying strategies inherits from the construction under the noiseless response setting. Recall that when the responses are noiseless, the learner first runs bisection search to locate X^* within a length $L\delta$ interval. She then runs replicated bisection on the L length δ subintervals, submitting queries via the bisection search in the true subinterval containing X^* and cloning those queries in the other $L-1$ subintervals. When the responses are noisy, firstly we replace the bisection searches with the BZ algorithm. Moreover, the learner can no longer discern the true interval by querying the endpoints of the subinterval only once. Instead we need to query each endpoint enough times, so that via a maximum-likelihood type procedure, the learner can estimate the true subinterval with high enough certainty.

Under the requirement that the learner is accurate on average, as per definition (a), we construct the following multi-stage querying strategy.

1. Let $L' = 7L$. Divide $[0,1]$ into $(L'\delta)$-length subintervals[1] and run the BZ algorithm to estimate the subinterval that contains X^*. The BZ algorithm is run for $K_1 = \frac{1}{c_3(p)} \log \frac{8}{7\epsilon L\delta}$ iterations. Write I for the subinterval returned by the BZ algorithm.

[1] For simplicity, we assume $(L'\delta)^{-1}$ is an integer. If not, the analysis can be repeated by dividing $[0,1]$ into subintervals of length $(\lfloor (L'\delta)^{-1} \rfloor)^{-1}$.

2. Divide I into L' δ-length subintervals, labeled $J_1, ..., J_{L'}$ from left to right. Label the endpoints as $x_0, ..., x_{L'}$ so that $J_k = [x_{k-1}, x_k)$. Submit $m = \frac{1}{c_4(p)} \log \frac{64\delta}{\epsilon}$ queries at each of the $L'-1$ endpoints $x_1, ..., x_{L'-1}$, where $c_4(p) = D(\text{Bern}(1/2)||\text{Bern}(p)) = (\log \frac{1}{2p} + \log \frac{1}{2(1-p)})/2$.

   Write m_k for the sum of the m responses to the query at x_k. In other words, m_k denotes the number of times the learner receives the response to "$X^* \geq x_k$" being 1. Let

$$\widehat{k} = \arg \max_{1 \leq k \leq L'} \sum_{i=1}^{k-1} m_i + \sum_{i=k}^{L'-1} (m - m_i),$$

   and take $J_{\widehat{k}}$ as the estimator for the subinterval that contains X^*.

3. Divide $J_{\widehat{k}}$ into length $(\epsilon/4)$ subintervals and run the BZ algorithm, while submitting queries in parallel in the other $L'-1$ subintervals $\{J_k\}_{k \neq \widehat{k}}$, as one would do in the replicated bisection. Run the BZ algorithm for $K_2 = \frac{2}{c_3(p)} \log \frac{4\sqrt{2}\delta}{\epsilon}$ iterations and obtain the output $J \subseteq J_{\widehat{k}}$.

4. Define the estimator $\widehat{X}$ as the midpoint of J.

It suffices to show that under definition (a) of ϵ-accuracy, the multi-stage querying strategy above is ϵ-accurate, (δ, L)-private, and achieves the upper bound in (1).

Accuracy: Discuss the following events:

1. $\mathcal{E}_1$: the BZ algorithm returns the wrong subinterval in stage 1. In other words, $X^* \notin I$.
2. $\mathcal{E}_{2,j}$ for $j = 1, ..., L'-1$: stage 1 does not incur an error, but stage 2 returns a subinterval out of $J_1, ..., J_{L'}$ that does not contain X^*, with $\widehat{k}$ at distance j away from the correct index. In other words, $\mathcal{E}_{2,j} = \{X^* \in J_{k^*} \text{ for some } k^* \in [L'], \text{ and } |\widehat{k} - k^*| = j\}$.
3. $\mathcal{E}_3$: the BZ algorithm makes an error in stage 3: $X^* \in J_{\widehat{k}}$ but $X^* \notin J$.
4. $\mathcal{E}_4$: $X^* \in J$.

The events above are disjoint, and their union forms the entire probability space. It is also easy to see that $|\widehat{X} - X^*|$ is upper bounded by $1, (j+1)\delta, \delta, \epsilon/8$ on $\mathcal{E}_1, \mathcal{E}_{2,j}, \mathcal{E}_3, \mathcal{E}_4$ respectively. Hence

$$\mathbb{E}\left|\widehat{X} - X^*\right| \leq \mathbb{P}\{\mathcal{E}_1\} + \sum_{j=1}^{L'-1} (j+1)\delta \mathbb{P}\{\mathcal{E}_{2,j}\} + \delta \mathbb{P}\{\mathcal{E}_3\} + \tfrac{\epsilon}{8}\mathbb{P}\{\mathcal{E}_4\}. \tag{12}$$

We claim that all events but $\mathcal{E}_4$ occur with low probability. Firstly, it follows from Lemma 1 that

$$\mathbb{P}\{\mathcal{E}_1\} \leq \frac{1}{L'\delta} 2^{-c_3(p)K_1} \leq \frac{\epsilon}{8}, \tag{13}$$

$$\mathbb{P}\{\mathcal{E}_3\} \leq \frac{\delta}{\epsilon/4} 2^{-c_3(p)K_2} \leq \frac{\epsilon}{8\delta} \tag{14}$$

from the choice of K_1, K_2.

To handle $\mathcal{E}_{2,j}$, note that conditional on $X^* \in J_{k^*}$,

$$m_i \overset{indep}{\sim} \begin{cases} \text{Binomial}(m, p) & \text{for} \quad 1 \leq i \leq k^* - 1; \\ \text{Binomial}(m, 1-p) & \text{for} \quad k^* \leq i \leq L'. \end{cases}$$

Hence $\widehat{k} = \arg\max_{1 \leq k \leq L'} \sum_{i=1}^{k-1} m_i + \sum_{i=k}^{L'-1} (m - m_i)$ is the maximum likelihood for k^*, and for all $k \neq k^*$,

$$\begin{aligned} \mathbb{P}\left\{\widehat{k} = k | X^* \in J_{k^*}\right\} \geq & \mathbb{P}\left\{\sum_{i=1}^{k-1} m_i + \sum_{i=k}^{L'-1} (m - m_i) \leq \sum_{i=1}^{k^*-1} m_i + \sum_{i=k^*}^{L'-1} (m - m_i)\right\} \\ = & \mathbb{P}\left\{B \leq \frac{|k - k^*|m}{2}\right\}, \text{for some } B \sim \text{Binomial}(|k - k^*|m, p), \end{aligned}$$

which is further bounded by $2^{-c_4(p)|k-k^*|m}$ from the binomial tail bound [9, Theorem 2.1]. Deduce that

$$\mathbb{P}\left\{\mathcal{E}_{2,j}\right\} \leq \mathbb{P}\{X^* \in J_{k^*}, \widehat{k} = k^* - j\} + \mathbb{P}\{X^* \in J_{k^*}, \widehat{k} = k^* + j\} \leq 2 \cdot 2^{-c_4(p)jm}. \tag{15}$$

Thus

$$\begin{aligned}
\sum_{j=1}^{L'-1}(j+1)\delta\mathbb{P}\left\{\mathcal{E}_{2,j}\right\} \leq & 2\sum_{j=1}^{L'-1}(j+1)\delta 2^{-c_4(p)jm} \\
\leq & 2\delta\left(\sum_{j=1}^{\infty}2^{-c_4(p)jm} + \sum_{i=1}^{\infty}\sum_{j=i}^{\infty}2^{-c_4(p)jm}\right) \\
= & \frac{2\delta 2^{-c_4(p)m}}{1-2^{-c_4(p)m}}\left(1+\frac{1}{1-2^{-c_4(p)m}}\right) \qquad (16) \\
\leq & 8\delta 2^{-c_4(p)m} \leq \epsilon/8, \qquad (17)
\end{aligned}$$

where the last two inequalities are due to the choice $m = \frac{1}{c_4(p)}\log\frac{64\delta}{\epsilon}$ and $\delta \geq 2\epsilon$.

Combining (12)-(17) yields that

$$\mathbb{E}\left|\widehat{X} - X^*\right| \leq \epsilon/8 + \epsilon/8 + \epsilon/8 + \epsilon/8 = \epsilon/2.$$

Privacy: The goal is to show that for all adversary's estimators $\widetilde{X}$ that could depend on q,

$$\mathbb{P}\left\{|\widetilde{X} - X^*| \leq \delta/2\right\} \leq \frac{1}{L}.$$

In the multi-stage algorithm the learner first runs the BZ algorithm on a $L'\delta$-fine grid to obtain an interval estimator I, then runs replicated BZ on the L' subintervals $J_1, ..., J_{L'}$ of I. Write I^* for the true subinterval on the $L'\delta$-fine grid that contains X^*. When $X^* \in I$, *i.e.* $I^* = I$, we used k^* to index the true subinterval out of $J_1, ..., J_{L'}$ that contains X^*. For the proof of (δ, L)-privacy, we need to expand the definition of k^* to incorporate the case $X^* \notin I$ as well. Label the L' length-δ subintervals of I^* as $J_1^*, ..., J_{L'}^*$ and define k^* so that $X^* \in J_{k^*}^*$. Recall that $\widehat{k}$ is the learner's estimator of k^*. We have

$$\begin{aligned}
& \mathbb{P}\left\{|\widetilde{X} - X^*| \leq \delta/2\right\} \\
\leq & \mathbb{P}\{X^* \notin I\} \qquad (18) \\
& + \mathbb{P}\left\{|\widetilde{X} - X^*| \leq \delta/2, \widehat{k} < k^*, X^* \in I\right\} \qquad (19) \\
& + \mathbb{P}\left\{|\widetilde{X} - X^*| \leq \delta/2, \widehat{k} > k^*, X^* \in I\right\} \qquad (20) \\
& + \mathbb{P}\left\{|\widetilde{X} - X^*| \leq \delta/2, \widehat{k} = k^*, X^* \in I\right\}. \qquad (21)
\end{aligned}$$

Of the four terms above, the first term equals the probability that the algorithm makes a mistake in the BZ algorithm in the first stage:

$$(18) = \mathbb{P}\mathcal{E}_1 \leq \frac{1}{L'\delta}2^{-c_3(p)K_1} \leq \frac{1}{L'}$$

where the last inequality holds due to $K_1 = \frac{1}{c_3(p)}\log\frac{8}{L'\epsilon\delta} \geq \frac{1}{c_3(p)}\log(1/\delta)$ in view of $2\epsilon \leq 1/L$.

For the second term, use the Bayes rule to write

$$(19) = \mathbb{P}\left\{\widehat{k} < k^*, X^* \in I\right\}\mathbb{P}\left\{|\widetilde{X} - X^*| \leq \frac{\delta}{2} \mid \widehat{k} < k^*, X^* \in I\right\}. \tag{22}$$

Since $\widetilde{X}$ is a function of q,

$$\begin{aligned}
&\mathbb{P}\left\{|\widetilde{X}-X^*|\le\delta/2 \mid \widehat{k}<k^*, X^*\in I\right\}\\
\le&\mathbb{E}_q\left(\sup_{t\in[0,1]}\mathbb{P}\left\{X^*\in[t-\delta/2,t+\delta/2]\cap I \mid \widehat{k}<k^*, X^*\in I, q\right\}\right)\\
\le&2\mathbb{E}_q\left(\max_{k\le L'}\mathbb{P}\left\{k^*=k \mid \widehat{k}<k^*, X^*\in I, q\right\}\right). \qquad (23)
\end{aligned}$$

The last inequality is because all intervals of the form $[t-\delta/2, t+\delta/2]\cap I$ must be covered by the union of two consecutive subintervals $J_k\cap J_{k+1}$ for some k.

Next we show that for all q and k,

$$\mathbb{P}\left\{k^*=k \mid \widehat{k}<k^*, X^*\in I, q\right\}=\mathbb{P}\left\{k^*=k \mid \widehat{k}<k^*, X^*\in I\right\}, \textit{i.e.},$$

$$\mathcal{L}(q \mid k^*=k, \widehat{k}<k^*, X^*\in I)=\mathcal{L}(q \mid \widehat{k}<k^*, X^*\in I).$$

In other words, k^* is independent of q conditional on $\widehat{k}<k^*$ and $X^*\in I$. Denote the queries submitted in the three stages as $q^{(1)}$, $q^{(2)}$ and $q^{(2)}$. We will establish conditional independence in two steps:

1. Show that $(q^{(1)}, q^{(2)})$ is independent of k^* conditional on $\widehat{k}<k^*$ and $X^*\in I$:
   note that given I^*, the conditional distribution of X^* is uniform on I^*. Therefore k^* is distributed uniformly on $[L']$ and is independent of I^*. Since the BZ algorithm only queries the endpoints of the subintervals, the distribution of the responses in the first stage $r^{(1)}$ only depends on X^* through I^*. Hence k^* is independent of the tuple $(I^*, r^{(1)})$. Moreover, $r^{(1)}$ completely determines I, so that k^* is independent of $(I^*, I, r^{(1)})$. On the other hand, when $X^*\in I$, $\widehat{k}$ is can be written as $f(k^*, \mathbf{noise}^{(2)})$, a function of only k^* and the binary noise variables in the second stage. We have

   $$\begin{aligned}
   &\mathcal{L}\left(r^{(1)} \mid k^*=k, \widehat{k}<k^*, X^*\in I\right)\\
   =&\mathcal{L}\left(r^{(1)} \mid k^*=k, f(k^*, \mathbf{noise}^{(2)})<k^*, I^*=I\right)\\
   =&\mathcal{L}\left(r^{(1)} \mid \widehat{k}<k^*, I^*=I\right).
   \end{aligned}$$

   The second equality is because by the independence of $(k^*, \mathbf{noise}^{(2)})$ and $(I^*, I, r^{(1)})$, $(k^*, \mathbf{noise}^{(2)})$ and $r^{(1)}$ are conditionally independent given $I^*=I$.

   We have shown that $r^{(1)}$ and k^* are independent conditional on $\widehat{k}<k^*$ and $X^*\in I$. Notice that $(q^{(1)}, q^{(2)})$ is a deterministic function of $r^{(1)}$. Thus $(q^{(1)}, q^{(2)})$ and k^* are also conditionally independent.

2. Show that $q^{(3)}$ is independent of k^* conditional on $\widehat{k}<k^*$, $X^*\in I$, $q^{(1)}$ and $q^{(2)}$:
   conditional on $\widehat{k}<k^*$ and $X^*\in I$, all the queries submitted in $\widehat{J}=J_{\widehat{k}}$ are smaller than X^*. Therefore the joint distribution of the queries in the third stage that fall in $\widehat{J}$ does not depend on k^*. Since the queries in the other subintervals are only copies of those in $\widehat{J}$, we have that $q^{(3)}$ is independent of k^*, conditional on $\widehat{k}<k^*$, $X^*\in I$, $q^{(1)}$ and $q^{(2)}$.

We have shown that the entire query sequence q is independent of k^* conditional on $\widehat{k}<k^*$ and $X^*\in I$. Thus

$$\mathbb{P}\left\{k^*=k \mid \widehat{k}<k^*, X^*\in I, q\right\}=\mathbb{P}\left\{k^*=k \mid \widehat{k}<k^*, X^*\in I\right\}.$$

Combine with (22) and (23) to obtain

$$\begin{aligned}
(19)\le&\mathbb{P}\left\{\widehat{k}<k^*, X^*\in I\right\}\cdot 2\max_{k\le L'}\mathbb{P}\left\{k^*=k \mid \widehat{k}<k^*, X^*\in I\right\}\\
=&2\max_{k\le L'}\mathbb{P}\left\{k^*=k, \widehat{k}<k^*, X^*\in I\right\}\\
\le&2\max_{k\le L'}\mathbb{P}\left\{k^*=k\right\}=2/L'.
\end{aligned}$$

where the last inequality is because k^* is distributed uniformly on $[L']$.

Following the same arguments, $(20) \leq 2/L'$. Next we handle term (21).

Once again because $\widetilde{X}$ is a function of q, we have

$$(21) \leq 2\mathbb{E}\left(\max_{k\leq L'} \mathbb{P}\left\{k^* = k \mid \widehat{k} = k^*, X^* \in I, q\right\}\right).$$

We claim that k^* and q are independent conditional on $\widehat{k} = k^*$ and $X^* \in I$. By the same arguments as in the analysis of (19) we can show that k^* is independent of $(q^1, q^{(2)})$ conditional on $\widehat{k} = k^*$ and $X^* \in I$. It remains to show that k^* is independent of $q^{(3)}$ conditional on $\widehat{k} = k^*$, $X^* \in I$ and $(q^{(1)}, q^{(2)})$. In other words,

$$\mathcal{L}\left(q^{(3)} \mid k^* = k, \widehat{k} = k^*, X^* \in I, q^{(1)}, q^{(2)}\right) = \mathcal{L}\left(q^{(3)} \mid \widehat{k} = k^*, X^* \in I, q^{(1)}, q^{(2)}\right)$$

To show the above, first note that conditional on $k^* = k, \widehat{k} = k^*$, $X^* \in I$ and $(q^{(1)}, q^{(2)})$, X^* is distributed uniformly on J_k. The queries sequence $q^{(3)}$ are generated from running the BZ algorithm on J_k, and replicating in the other subintervals. The conditional distribution of the $q^{(3)}$ is therefore independent of the value of k. Thus

$$\begin{aligned}
(21) =& \mathbb{P}\left\{|\widetilde{X} - X^*| \leq \delta/2 \mid \widehat{k} = k^*, X^* \in I\right\} \mathbb{P}\left\{\widehat{k} = k^*, X^* \in I\right\} \\
\leq& 2 \max_{k\leq L'} \mathbb{P}\left\{k^* = k \mid \widehat{k} = k^*, X^* \in I\right\} \mathbb{P}\left\{\widehat{k} = k^*, X^* \in I\right\} \\
=& 2 \max_{k\leq L'} \mathbb{P}\left\{k^* = k, \widehat{k} = k^*, X^* \in I\right\} \\
\leq& 2 \max_{k\leq L'} \mathbb{P}\left\{k^* = k\right\} = 2/L'.
\end{aligned}$$

Collect all the terms to deduce that

$$\mathbb{P}\left\{|\widetilde{X} - X^*| \leq \delta/2\right\} \leq \frac{1}{L'} + \frac{2}{L'} + \frac{2}{L'} + \frac{2}{L'} = \frac{1}{L}$$

by picking $L' = 7L$.

Query complexity: the total number of queries submitted by the querying strategy is

$$\begin{aligned}
K_1 + (L'-1)m + L'K_2 =& \frac{1}{c_3(p)} \log \frac{8}{7\epsilon L\delta} + (7L-1)\frac{1}{c_4(p)} \log \frac{64\delta}{\epsilon} + \frac{14L}{c_3(p)} \log \frac{4\sqrt{2}\delta}{\epsilon} \\
\leq& \frac{1}{c_3(p)} \log \frac{8}{7\epsilon L\delta} + \left(\frac{14}{c_3(p)} + \frac{7}{c_4(p)}\right) L \log \frac{64\delta}{\epsilon} \\
\leq& \left(\frac{14}{c_3(p)} + \frac{7}{c_4(p)}\right)\left(\log \frac{1}{\epsilon} + L \log \frac{64\delta}{\epsilon}\right).
\end{aligned}$$

where the last inequality is because $L\delta \geq 2\delta \geq 2\epsilon$.

3.1.3 Proof of the upper bound in (2)

When the learner needs to be accurate with high probability, as per definition (b), we adopt the same multi-stage querying strategy as in the upper bound proof in (1), with a slightly modified set of parameters. Let

$$K_1 = \frac{1}{c_3(p)} \log \frac{3M}{\delta}, \quad m = \frac{1}{c_4(p)} \log(12M), \quad K_2 = \frac{1}{c_3(p)} \log \frac{12M\delta}{\epsilon}. \tag{24}$$

Next we show that the querying strategy is (ϵ, M)-accurate and (δ, L)-private with the desired query complexity. The proof is similar to that of the upper bound in (1).

Accuracy: Recall the events $\mathcal{E}_1$, $\mathcal{E}_{2,j}, \mathcal{E}_3, \mathcal{E}_4$ defined in the proof of (1). We have

$$\begin{aligned}\mathbb{P}\left\{|\widehat{X} - X^*| > \epsilon/2\right\} \leq & \mathbb{P}\mathcal{E}_4^c = \mathbb{P}\mathcal{E}_1 + \sum_{j=1}^{L'-1} \mathbb{P}\mathcal{E}_{2,j} + \mathbb{P}\mathcal{E}_3 \\ \leq & \frac{1}{L'\delta} 2^{-c_3(p)K_1} + 2\sum_{j=1}^{L'-1} 2^{-c_4(p)jm} + \frac{\delta}{\epsilon/4} 2^{-c_3(p)K_2} \\ \leq & \frac{1}{L'\delta} 2^{-c_3(p)K_1} + 4 \cdot 2^{-c_4(p)m} + \frac{\delta}{\epsilon/4} 2^{-c_3(p)K_2}\end{aligned}$$

where the second inequality follows from (13), (14) and (15). Plug in the the values of K_1, K_2 and m to conclude that $\mathbb{P}\{|\widehat{X} - X^*| > \epsilon/2\} \leq 1/(3M) + 1/(3M) + 1/(3M) = 1/M$.

Privacy: the proof for (δ, L)-privacy is almost identical to the proof of (1). The only part that differs is in the treatment of the term (18) due to a different choice of K_1. We have

$$(18) = \mathbb{P}\{X^* \notin I\} = \mathbb{P}\mathcal{E}_1 \leq \frac{1}{L'\delta} 2^{-c_3(p)K_1} \leq \frac{1}{L'}$$

for $K_1 = \frac{1}{c_3(p)} \log(3M/\delta)$.

Query complexity: the total number of queries submitted is

$$\begin{aligned}K_1 + (L'-1)m + L'K_2 = & \frac{1}{c_3(p)} \log \frac{3M}{\delta} + (7L-1)\frac{1}{c_4(p)} \log(12M) + \frac{7L}{c_3(p)} \log \frac{12M\delta}{\epsilon} \\ \leq & \left(\frac{8}{c_3(p)} + \frac{7}{c_4(p)}\right)\left(\log \frac{1}{\epsilon} + L \log \frac{12M\delta}{\epsilon}\right).\end{aligned}$$

3.2 Proof of the lower bounds

3.2.1 An auxiliary lemma

The lower bound proofs rely heavily on the following auxiliary lemma, which connects the expected number of queries near X^* with the learner's error probability.

Lemma 2. *For each deterministic interval J and $\eta > 0$,*

$$\mathbb{E}\left(\textit{the number of queries in } J \mid X^* \in J\right) \geq \frac{1}{c_2(p)}\left(\left(1 - \mathbb{P}\{|\widehat{X} - X^*| > \eta \mid X^* \in J\}\right) \log \frac{|J|}{2\eta} - 1\right).$$

Proof. **Step 1: discretize w.r.t. Y.**

By the independence between X^* and the random seed Y, we can write

$$\begin{aligned}& \mathbb{E}\left(\text{the number of queries in } J \mid X^* \in J\right) \\ = & \int_0^1 \mathbb{E}\left(\text{the number of queries in } J \mid X^* \in J, Y = y\right) dy \\ = & \int_0^1 \sum_{i \leq n} \mathbb{P}\left\{q_i \in J | X^* \in J, Y = y\right\} dy. \end{aligned} \tag{25}$$

Step 2: establish a rate of information transfer.

In this step we show that for all $y \in [0,1]$,

$$I(X^*; r_1, ..., r_n | X^* \in J, Y = y) \leq c_2(p) \sum_{i \leq n} \mathbb{P}\left\{q_i \in J \mid X^* \in J, Y = y\right\}. \tag{26}$$

Recall that $c_2(p) = h(1/2) - h(p)$ where $h(t) = H(\text{Bern}(t)) = -t \log t - (1-t)\log(1-t)$. The intuition behind (26) is that since the observed responses are passed through a binary symmetric channel that flips the

noiseless responses with probability $1-p$, each query reveals at most $h(1/2)-h(p)$ bits of information about X^*. Next we prove (26). We abbreviate $\{X^* \in J, Y = y\}$ as $\mathcal{E}_{J,y}$. Start from the left-hand side:

$$\begin{aligned}
&I(X^*; r_1, ..., r_n | \mathcal{E}_{J,y}) \\
=&I(X^*; r_1 | \mathcal{E}_{J,y}) + \sum_{i=1}^{n-1} I(X^*; r_{i+1} | \mathcal{E}_{J,y}, r_1, ..., r_i) \\
=&H(r_1 | \mathcal{E}_{J,y}) - H(r_1 | X^*, \mathcal{E}_{J,y}) + \\
&\sum_{i=1}^{n-1} \left(H(r_{i+1} | \mathcal{E}_{J,y}, r_1, ..., r_i) - H(r_{i+1} | X^*, \mathcal{E}_{J,y}, r_1, ..., r_i) \right). \qquad (27)
\end{aligned}$$

To analyze the i'th summand in (27), write

$$H(r_{i+1} | \mathcal{E}_{J,y}, r_1, ..., r_i) = \sum_{\rho_1, ..., \rho_i} \mathbb{P}\{r_1 = \rho_1, ..., r_i = \rho_i | \mathcal{E}_{J,y}\} H(r_{i+1} | \mathcal{E}_{J,y}, r_1 = \rho_1, ..., r_i = \rho_i).$$

Recall that $q_{i+1} = \phi_i(r_1, ..., r_i, Y)$. Therefore with the values of $r_1, ..., r_i, Y$ fixed, the $i+1$'th query is deterministic. If it lands to the left of J, then conditional on $X^* \in J$, we have $X^* \geq q_{i+1}$ with (conditional) probability 1, implying that $r_{i+1} \sim \text{Bern}(p)$. We have

$$H(r_{i+1} | \mathcal{E}_{J,y}, r_1 = \rho_1, ..., r_i = \rho_i) = h(p), \quad \text{if } \phi_i(\rho_1, ..., \rho_i, y) \text{ is to the left of } J.$$

By the same logic, the equality also holds if $\phi_i(\rho_1, ..., \rho_i, y)$ is to the right of J. If $\phi_i(\rho_1, ..., \rho_i, y)$ lands inside of J, we can use $h(1/2)$ to bound the conditional entropy of r_{i+1} since it is a binary random variable. Deduce that

$$\begin{aligned}
H(r_{i+1} | \mathcal{E}_{j,y}, r_1, ..., r_i) \leq &h(1/2) \sum_{\rho_1, ..., \rho_i : \phi_i(\rho_1, ..., \rho_i, y) \in J} \mathbb{P}\{r_1 = \rho_1, ..., r_i = \rho_i | \mathcal{E}_{J,y}\} + \\
&h(p) \sum_{\rho_1, ..., \rho_i : \phi_i(\rho_1, ..., \rho_i, y) \notin J} \mathbb{P}\{r_1 = \rho_1, ..., r_i = \rho_i | \mathcal{E}_{J,y}\} \\
=&h(1/2) \mathbb{P}\{q_{i+1} \in J | \mathcal{E}_{J,y}\} + h(p) \mathbb{P}\{q_{i+1} \notin J | \mathcal{E}_{J,y}\}.
\end{aligned}$$

On the other hand, if we condition on $\mathcal{E}_{J,y}, r_1, ..., r_i$ and the value of X^*, then not only is q_{i+1} deterministic, so is $\mathbb{1}\{X^* \geq q_{i+1}\}$. As a result, r_{i+1} is distributed Bernoulli with success probability either p or $1-p$ depending on the relative position of q_{i+1} to X^*. Hence

$$H(r_{i+1} | X^*, \mathcal{E}_{J,y}, r_1, ..., r_i) = h(p).$$

We have shown that

$$\begin{aligned}
&H(r_{i+1} | \mathcal{E}_{J,y}, r_1, ..., r_i) - H(r_{i+1} | X^*, \mathcal{E}_{J,y}, r_1, ..., r_i) \\
\leq &h(1/2) \mathbb{P}\{q_{i+1} \in J | \mathcal{E}_{J,y}\} + h(p) \mathbb{P}\{q_{i+1} \notin J | \mathcal{E}_{J,y}\} - h(p) \\
= &(h(1/2) - h(p)) \mathbb{P}\{q_{i+1} \in J | \mathcal{E}_{J,y}\}. \qquad (28)
\end{aligned}$$

Similarly we obtain the following bound:

$$H(r_1 | \mathcal{E}_{J,y}) - H(r_1 | X^*, \mathcal{E}_{J,y}) \leq (h(1/2) - h(p)) \mathbb{P}\{q_1 \in J | \mathcal{E}_{J,y}\}. \qquad (29)$$

Combine (27), (28) and (29) to finish the proof of (26).

Step 3: apply Fano's inequality to connect the learner's probability of error with the expected number of queries in J.

From the continuous Fano's inequality [2, Proposition 2],

$$\mathbb{P}\left\{ |\widehat{X} - X^*| > \eta | \mathcal{E}_{J,y} \right\} \geq 1 - \frac{I(X^*; r_1, ..., r_n | \mathcal{E}_{J,y}) + 1}{\log \frac{|J|}{2\eta}}.$$

Combining the above with (26) yields

$$\sum_{i\le n}\mathbb{P}\left\{q_i\in J|\mathcal{E}_{J,y}\right\}\ge\frac{1}{c_2(p)}\left(\left(1-\mathbb{P}\left\{|\widehat{X}-X^*|>\eta|\mathcal{E}_{J,y}\right\}\right)\log\frac{|J|}{2\eta}-1\right). \tag{30}$$

Step 4: integrate w.r.t. Y.

Combine (32) with (30), and integrate w.r.t. Y to obtain

$$\begin{aligned}&\mathbb{E}\left(\text{the number of queries in } J\mid X^*\in J\right)\\ \ge&\int_0^1\frac{1}{c_2(p)}\left(\left(1-\mathbb{P}\left\{|\widehat{X}-X^*|>\eta|\mathcal{E}_{J,y}\right\}\right)\log\frac{|J|}{2\eta}-1\right)dy\\ =&\frac{1}{c_2(p)}\left(\left(1-\mathbb{P}\{|\widehat{X}-X^*|>\eta\mid X^*\in J\}\right)\log\frac{|J|}{2\eta}-1\right).\end{aligned}$$

□

3.2.2 Proof of the lower bound in (1)

Suppose ϕ is an ϵ-accurate and (δ,L)-private querying strategy that submits at most n queries. As in the noiseless case, we can assume WLOG that the learner always submits exactly n queries by concatenating trivial queries at 0 to the end of the query sequence.

We will split the lower bound in (1) into the following two inequalities, which we will prove in order.

(i) $n\ge\frac{1}{2c_2(p)}L\log\frac{\delta}{16\epsilon}$.

(ii) $n\ge\frac{1}{2c_2(p)}\log\frac{1}{8\epsilon}$.

Proof of (i). Consider the adversary who adopts the proportional-sampling strategy [18], *i.e.*, $\widetilde{X}$ is the sampled from the empirical distribution of all the queries. We have

$$\mathbb{P}\{|\widetilde{X}-X^*|\le\delta/2\}\ge\frac{1}{n}\mathbb{E}\left(\text{the number of queries in the interval }[X^*-\delta/2,X^*+\delta/2]\right).$$

For a querying strategy to be (δ,L)-private, we must have $\mathbb{P}\{|\widetilde{X}-X^*|\le\delta/2\}\le 1/L$. Hence

$$n\ge L\mathbb{E}\left(\text{the number of queries in the interval }[X^*-\delta/2,X^*+\delta/2]\right). \tag{31}$$

Divide $[0,1]$ into length $\delta/2$ subintervals labeled $J_1,...,J_{2/\delta}$ (again ignoring non-divisibility issues). Suppose J^* is the subinterval that contains X^*, then J^* is distributed uniformly on $\{J_1,...,J_{2/\delta}\}$, and it must be a subset of $[X^*-\delta/2,X^*+\delta/2]$. Therefore

$$\begin{aligned}&\mathbb{E}\left(\text{the number of queries in }[X^*-\delta/2,X^*+\delta/2]\right)\\ \ge&\mathbb{E}\left(\text{the number of queries in }J^*\right)\\ \ge&\frac{\delta}{2}\sum_{j\le 2/\delta}\mathbb{E}\left(\text{the number of queries in }J_j|X^*\in J_j\right).\end{aligned} \tag{32}$$

By applying Lemma 2 with $J=J_j$ and $\eta=\epsilon$, we have for all $j\in[2/\delta]$,

$$\mathbb{E}\left(\text{the number of queries in }J_j\mid X^*\in J_j\right)\ge\frac{1}{c_2(p)}\left(\left(1-\mathbb{P}\{|\widehat{X}-X^*|>\epsilon\mid X^*\in J_j\}\right)\log\frac{\delta}{4\epsilon}-1\right).$$

Plug into (32) to obtain

$$\begin{aligned}&\mathbb{E}\left(\text{the number of queries in } [X^*-\delta/2, X^*+\delta/2]\right)\\ \geq&\frac{\delta}{2}\sum_{j\leq 2/\delta}\frac{1}{c_2(p)}\left(\left(1-\mathbb{P}\{|\widehat{X}-X^*|>\epsilon \mid X^*\in J_j\}\right)\log\frac{\delta}{4\epsilon}-1\right)\\ =&\frac{1}{c_2(p)}\left(\left(1-\mathbb{P}\{|\widehat{X}-X^*|>\epsilon\}\right)\log\frac{\delta}{4\epsilon}-1\right)\geq\frac{1}{2c_2(p)}\log\frac{\delta}{16\epsilon},\end{aligned}$$

where the last inequality is from $\mathbb{P}\{|\widehat{X}-X^*|>\epsilon\}\leq\mathbb{E}|\widehat{X}-X^*|/\epsilon\leq 1/2$. We have arrived at the desired lower bound

$$n\geq L\mathbb{E}\left(\text{the number of queries in } [X^*-\delta/2, X^*+\delta/2]\right)\geq\frac{1}{2c_2(p)}L\log\frac{\delta}{16\epsilon}.$$

□

Proof of (ii). Apply lemma 2 with $J=[0,1]$ and $\eta=\epsilon$, we have

$$n\geq\frac{1}{c_2(p)}\left(\left(1-\mathbb{P}\{|\widehat{X}-X^*|>\epsilon\}\right)\log\frac{1}{2\epsilon}-1\right)\geq\frac{1}{c_2(p)}\log\frac{1}{8\epsilon},$$

where the second inequality is from $\mathbb{P}\{|\widehat{X}-X^*|>\epsilon\}\leq 1/2$. □

3.2.3 Proof of the lower bound in (2)

Suppose ϕ is an (ϵ, M)-accurate and (δ, L)-private querying strategy that submits at most n queries. Argue as in the proof of (1) that we can assume WLOG that the learner always submits exactly n queries. We will split the lower bound in (2) into the following three inequalities and prove them in order.

(i) $n\geq\frac{L}{c_2(p)}\log\frac{\delta}{8\epsilon}$.

(ii) $n\geq\frac{1}{c_2(p)}\log\frac{1}{4\epsilon}$.

(iii) $n\geq\frac{1}{c_1(p)}L\log M$.

Proof of (i). The proof is essentially identical to the proof of (i) in Section 3.2.2. Since the query complexities $N_{\mathsf{avg}}()$ and $N_{\mathsf{whp}}()$ are defined with the same notion of (δ, L)-privacy, the proportional sampling argument used in the proof of (1) remains valid here. As before, by considering a proportionally-sampling adversary and partitioning $[0,1]$ into length $\delta/2$ subintervals $J_1, ..., J_{2/\delta}$, we have

$$\begin{aligned}n\geq&L\mathbb{E}\left(\text{the number of queries in } [X^*-\delta/2, X^*+\delta/2]\right)\\ \geq&L\cdot\frac{\delta}{2}\sum_{j\leq 2/\delta}\mathbb{E}\left(\text{the number of queries in } J_j \mid X^*\in J_j\right).\end{aligned}$$

Applying Lemma 2 with $J=J_j$ and $\eta=\epsilon/2$ yields

$$\mathbb{E}\left(\text{the number of queries in } J_j \mid X^*\in J_j\right)\geq\frac{\left(1-\mathbb{P}\left\{|\widehat{X}-X^*|>\epsilon/2 \mid X^*\in J_j\right\}\right)\log\left(\frac{\delta}{2\epsilon}\right)-1}{c_2(p)}.$$

It follows that

$$\begin{aligned}n\geq&L\cdot\frac{\delta}{2}\sum_{j\leq 2/\delta}\frac{\left(1-\mathbb{P}\left\{|\widehat{X}-X^*|>\epsilon/2 \mid X^*\in J_j\right\}\right)\log\left(\frac{\delta}{2\epsilon}\right)-1}{c_2(p)}\\ =&\frac{L}{c_2(p)}\left(\left(1-\mathbb{P}\left\{|\widehat{X}-X^*|>\epsilon/2 \mid X^*\in J_j\right\}\right)\log\left(\frac{\delta}{2\epsilon}\right)-1\right).\end{aligned}$$

From the definition of (ϵ, M)-accuracy, $\mathbb{P}\{|\widehat{X}-X^*|>\epsilon/2\}\leq 1/M\leq 1/2$. Plug in to yield (i). □

Proof of (ii). Apply Lemma 2 with $J = [0, 1]$ and $\eta = \epsilon/2$. We have

$$n \geq \frac{1}{c_2(p)} \left(\left(1 - \mathbb{P}\left\{ |\widehat{X} - X^*| > \epsilon/2 \right\} \right) \log(1/\epsilon) - 1 \right).$$

Combine with $\mathbb{P}\{|\widehat{X} - X^*| > \epsilon/2\} \leq 1/2$ to yield (ii). □

Proof of (iii). As we have argued in the first two steps of the lower bound proof of (i) in Section 3.2.2,

$$n \geq L \cdot \frac{\delta}{2} \sum_{j \leq 2/\delta} \int_0^1 \mathbb{E}(\text{the number of queries in } J_j | \mathcal{E}_{j,y}) dy. \tag{33}$$

Fano's inequality is no longer sufficient to yield a $n = \Omega(\log M)$ lower bound on the expected number of queries in J_j. Instead our proof strategy is to reduce the estimation problem to a binary hypothesis test.

Denote $J_j = [a_j, b_j]$ with midpoint $m_j = (a_j + b_j)/2$. Let $I_0 = [a_j, m_j - \epsilon/2)$ and $I_1 = [m_j + \epsilon/2, b_j]$ be two subintervals of J_j that are ϵ apart. Then any learner that achieves $|\widehat{X} - X^*| \leq \epsilon/2$ must also be able to test between the two hypotheses $X^* \in I_0$ and $X^* \in I_1$. Indeed,

$$\begin{aligned}
&\mathbb{P}\left\{ |\widehat{X} - X^*| > \epsilon/2 | \mathcal{E}_{j,y} \right\} \\
\geq &\frac{|I_0|}{|J_j|} \mathbb{P}\left\{ \widehat{X} \geq m_j | X^* \in I_0, Y = y \right\} + \frac{|I_1|}{|J_j|} \mathbb{P}\left\{ \widehat{X} < m_j | X^* \in I_1, Y = y \right\} \\
= &\frac{2|I_0|}{|J_j|} \left(\tfrac{1}{2} \mathbb{P}\left\{ \widehat{X} \geq m_j | X^* \in I_0, Y = y \right\} + \tfrac{1}{2} \mathbb{P}\left\{ \widehat{X} < m_j | X^* \in I_1, Y = y \right\} \right),
\end{aligned} \tag{34}$$

where the last equality comes from $|I_0| = |I_1|$. The term in the parentheses is the average error probability of the test $\widehat{T} = \mathbb{1}\{\widehat{X} \geq m_j\}$ under the uniform prior on the hypotheses. Furthermore, it can be viewed as an average error probability of a family of simple tests by symmetry of I_0 and I_1:

$$\begin{aligned}
&\tfrac{1}{2} \mathbb{P}\left\{ \widehat{X} \geq m_j | X^* \in I_0, Y = y \right\} + \tfrac{1}{2} \mathbb{P}\left\{ \widehat{X} < m_j | X^* \in I_1, Y = y \right\} \\
= &\tfrac{1}{2} \int \frac{\mathbb{1}\{x \in I_0\}}{|I_0|} \mathbb{P}\left\{ \widehat{X} \geq m_j | X^* = x, Y = y \right\} dx + \tfrac{1}{2} \int \frac{\mathbb{1}\{x \in I_1\}}{|I_1|} \mathbb{P}\left\{ \widehat{X} < m_j | X^* = x, Y = y \right\} dx \\
= &\int \frac{\mathbb{1}\{x \in I_0\}}{|I_0|} \left(\tfrac{1}{2} \mathbb{P}\left\{ \widehat{X} \geq m_j | X^* = x, Y = y \right\} + \tfrac{1}{2} \mathbb{P}\left\{ \widehat{X} < m_j | X^* = 2m_j - x, Y = y \right\} \right) dx \\
\geq &\int \frac{\mathbb{1}\{x \in I_0\}}{|I_0|} \inf_{\widehat{T}} \left(\tfrac{1}{2} \mathbb{P}\left\{ \widehat{T}(r^{(n)}) = 1 | X^* = x, Y = y \right\} + \tfrac{1}{2} \mathbb{P}\left\{ \widehat{T}(r^{(n)}) = 0 | X^* = 2m_j - x, Y = y \right\} \right) dx.
\end{aligned} \tag{35}$$

We have reduced the problem to lower bounding the minimum error probability of the simple test

$$H_0 : X^* = x \quad \text{against} \quad H_1 : X^* = 2m_j - x.$$

From [4, Eq (49)] we have the lower bound

$$\inf_{\widehat{T}} \left(\tfrac{1}{2} \mathbb{P}\left\{ \widehat{T}(r^{(n)}) = 1 | H_0, Y = y \right\} + \tfrac{1}{2} \mathbb{P}\left\{ \widehat{T}(r^{(n)}) = 0 | H_1, Y = y \right\} \right) \geq \frac{\rho^2}{4}, \tag{36}$$

where ρ is the Bhattacharyya coefficient:

$$\rho = \sum_{r^{(n)} \in \{0,1\}^n} \sqrt{\mathbb{P}\{r^{(n)} | H_0, Y = y\} \mathbb{P}\{r^{(n)} | H_1, Y = y\}} = \mathbb{E}\left(2^{\Lambda/2} | H_0, Y = y \right)$$

with

$$\begin{aligned}
\Lambda = &\log \frac{\mathbb{P}\{r^{(n)} | H_1, Y = y\}}{\mathbb{P}\{r^{(n)} | H_0, Y = y\}} \\
= &\log \frac{\prod_{i \leq n} \left(\mathbb{1}\{2m_j - x \geq q_i\} p^{r_i} (1-p)^{1-r_i} + \mathbb{1}\{2m_j - x < q_i\} p^{1-r_i} (1-p)^{r_i} \right)}{\prod_{i \leq n} \left(\mathbb{1}\{x \geq q_i\} p^{r_i} (1-p)^{1-r_i} + \mathbb{1}\{x < q_i\} p^{1-r_i} (1-p)^{r_i} \right)} \\
= &\sum_{i=1}^n \mathbb{1}\{q_i \in (x, 2m_j - x]\} \left(r_i \log \frac{p}{1-p} + (1 - r_i) \log \frac{1-p}{p} \right).
\end{aligned}$$

By Jensen's inequality,

$$2\log\rho \geq \mathbb{E}\left(\Lambda|H_0, Y=y\right) = \sum_{i\leq N} \mathbb{P}\left\{q_i \in (x, 2m_j - x]|H_0, Y=y\right\}\left((1-p)\log\frac{p}{1-p} + p\log\frac{1-p}{p}\right)$$
$$\geq -c_1(p)\mathbb{E}(\text{the number of queries in } J_j|H_0, Y=y)$$

where $c_1(p) = p\log\frac{p}{1-p} + (1-p)\log\frac{1-p}{p}$ is always nonnegative. We have arrived at

$$\rho^2 \geq 2^{\mathbb{E}(\Lambda|H_0,Y=y)} \geq 2^{-c_1(p)\mathbb{E}(\text{the number of queries in } J_j|H_0,Y=y)}.$$

Together with (34), (35) and (36) we have

$$\begin{aligned}
&\mathbb{P}\left\{|\widehat{X} - X^*| > \epsilon/2|\mathcal{E}_{j,y}\right\}\\
\geq&\frac{|I_0|}{2|J_j|}\int \frac{\mathbb{1}\{x\in I_0\}}{|I_0|}2^{-c_1(p)\mathbb{E}(\text{the number of queries in } J_j|X^*=x,Y=y)}dx\\
\geq&\frac{|I_0|}{2|J_j|}2^{-c_1(p)\mathbb{E}(\text{the number of queries in } J_j|X^*\in I_0,Y=y)},
\end{aligned}$$

where the last inequality follows from Jensen's inequality. By symmetry the same inequality holds for I_1. Therefore

$$\begin{aligned}
&\mathbb{E}(\text{the number of queries in } J_j|\mathcal{E}_{j,y})\\
\geq&\mathbb{P}\{X^*\in I_0|\mathcal{E}_{j,y}\}\mathbb{E}(\text{the number of queries in } J_j|X^*\in I_0, Y=y)\\
&+\mathbb{P}\{X^*\in I_1|\mathcal{E}_{j,y}\}\mathbb{E}(\text{the number of queries in } J_j|X^*\in I_1, Y=y)\\
\geq&\frac{2|I_0|}{|J_j|}\cdot\frac{1}{c_1(p)}\left(-\log\mathbb{P}\left\{|\widehat{X}-X^*|>\frac{\epsilon}{2}|\mathcal{E}_{j,y}\right\} + \log\frac{|I_0|}{2|J_j|}\right).
\end{aligned}$$

Recall that $|J_j| = \delta/2$ and $|I_0| = |J_j|/2 - \epsilon/2$. We have $|I_0|/|J_j| \geq 1/4$ from the assumption $\delta \geq 4\epsilon$. Combine the inequality above with (33) to obtain

$$\begin{aligned}
n \geq& -\frac{L}{2c_1(p)}\cdot\frac{\delta}{2}\sum_{j\leq 2/\delta}\int_0^1 \log\left(8\mathbb{P}\{|\widehat{X}-X^*|>\epsilon/2|\mathcal{E}_{j,y}\}\right)\\
\geq& -\frac{L}{2c_1(p)}\log\left(\frac{\delta}{2}\sum_{j\leq 2/\delta}\int_0^1 8\mathbb{P}\{|\widehat{X}-X^*|>\epsilon/2|\mathcal{E}_{j,y}\}dy\right)\\
=& -\frac{L}{2c_1(p)}\log\left(8\mathbb{P}\{|\widehat{X}-X^*|>\epsilon/2\}\right) \geq \frac{L}{2c_1(p)}\log\frac{M}{8},
\end{aligned}$$

where the second inequality is due to Jensen's inequality and the last inequality follows from the definition of (ϵ, M)-accuracy. □

4 Extensions to multidimensions

In this section we extend our results for noiseless responses to d dimensions for $d > 1$, under both the Bayesian and the deterministic settings. Suppose the true value X^* is in $\mathbb{R}^d$. The closeness of estimators to X^* is measured with respect to the $\|\cdot\|_\infty$ norm, and the accuracy and privacy levels of a querying strategy are defined accordingly. By using the $\|\cdot\|_\infty$ norm to measure the adversary's accuracy, we are allowing the adversary to accurately estimate one or some of the coordinates of X^*. That is because in high dimensions, a single coordinate of the model parameter often does not provide meaningful predictive power. As a result, we only declare privacy breach when the adversary gets "close" to X^* in $\mathbb{R}^d$. Here we use the $\|\cdot\|_\infty$ norm to measure closeness. But we comment that as a consequence of our result, if the Euclidean norm were used, the complexity would only differ by a multiplicative constant.

We assume that the learner is only allowed to ask questions of the type "is $X_i^* \geq q$?" for some $i \in [d]$ and $q \in [0,1]$. Denote the optimal query complexity in d-dimensions as $N_d(\epsilon, \delta, L)$. Next we present our results for the Bayesian and deterministic settings.

Bayesian setting Suppose X^* is uniformly distributed on $[0,1]^d$. We say a querying strategy ϕ is

- ϵ-accurate for $\epsilon > 0$, if $\mathbb{P}\{\|\widehat{X} - X^*\|_\infty \le \epsilon/2\} = 1$;
- (δ, L)-private for $\delta > 0$ and an integer $L \ge 2$, if there is no adversary $\widetilde{X}$ such that $\mathbb{P}\{\|\widetilde{X} - X^*\|_\infty \le \delta/2\} > 1/L$.

We focus on the parameter regime $2\epsilon \le \delta \le 1/\lceil L^{1/d} \rceil$. Once gain $2\epsilon \le \delta$ is reasonable to assume, since the accuracy of the adversary is not expected to beat that of the learner. To justify the other end of the spectrum, note that if $\delta > 1/L^{1/d}$, then the naive estimator $\widetilde{X} = 1/2$ achieves $\mathbb{P}\{\|\widetilde{X} - X^*\|_\infty \le \delta/2\} = \delta^d > 1/L$, making it impossible to fulfill the privacy constraint.

Denote $\gamma = \gamma(L,d) = L^{1/d}$. Below is our main result on the multidimensional optimal query complexity in the Bayesian setting.

Theorem 4 (Bayesian setting)**.** *If $2\epsilon \le \delta \le 1/\lceil\gamma\rceil$, then*

$$N_d(\epsilon,\delta,L) \le d\left(\left\lfloor \log \frac{1}{\lceil\gamma\rceil\delta}\right\rfloor + \lceil\gamma\rceil\left(\left\lceil \log\frac{\delta}{\epsilon}\right\rceil + 2\right) - 1\right).$$

Furthermore, assuming that the queries on X_i^* depend only on the responses to the previous queries on X_i^* and some random seed Y_i, with $Y_1, ..., Y_d$ mutually independent, then

$$N_d(\epsilon,\delta,L) \ge d\left(\left\lfloor \log \frac{1}{\gamma\delta}\right\rfloor + \gamma\left(\log\frac{\delta}{\epsilon} - 2\right) - 1\right).$$

Deterministic setting Suppose $X^* \in [0,1]^d$ is deterministic, we say a querying strategy ϕ is

- ϵ-accurate for $\epsilon > 0$, if $\mathbb{P}\{\|\widehat{X} - X^*\|_\infty \le \epsilon/2\} = 1$ for all $X^* \in [0,1]^d$;
- (δ, L)-private for $\delta > 0$ and an integer $L \ge 2$, if for each query sequence $\bar{q}$, the δ-covering number of the information set $\mathcal{I}(\bar{q})$ is at least L. [2]

Theorem 5 (Deterministic setting)**.** *If $2\epsilon \le \delta \le 1/\lceil\gamma\rceil$, then*

$$d\left(2\gamma + \log\frac{\max\{2^{-\lceil\gamma\rceil},\delta\}}{\epsilon} - 8\right) \le N_d(\epsilon,\delta,L) \le d\left(2\lceil\gamma\rceil + \left\lceil \log\frac{\max\{2^{-\lceil\gamma\rceil},\delta\}}{\epsilon}\right\rceil + 1\right).$$

From the upper and lower bounds in the theorem statements, we see that in d-dimensions the optimal query complexity suffers from a multiplicative factor of d. This is consistent with the optimal query complexity $d\log(1/\epsilon)$ when there is no privacy constraint. The query complexity for each dimension depends on $\gamma = L^{1/d}$. As d grows, the price to pay for privacy per dimension decreases. In the extreme case where $d \to \infty$ with L fixed, the optimal query complexity behaves like $d\log(1/\epsilon)$ in both the Bayesian and the deterministic settings, making the privacy constraint obsolete in high dimensions.

One interesting direction to strengthen Theorem 4 and Theorem 5 is to allow the learner to query "is X^* in H?" where H is an arbitrary half-space in $\mathbb{R}^d$. The upper bounds are still valid since every comparison query corresponds to a half-space. However for both the Bayesian and the deterministic setting, our current lower bound proof strategies do not accommodate this wider class of queries. This variant of the problem was studied by [19, Theorem 2] under the Bayesian setting, where the author gives a lower bound of $c_1\delta^{d-1}L\log(\delta/\epsilon) - c_2 L$ for constants c_1, c_2 that depend on d. The lower bound is obtained via a hyperplane transversality argument. More specifically, divide $[0,1]^d$ into δ-wide cubes and consider an adversary who samples from the cubes that intersect with the queried hyperplanes. The maximum number of cubes each hyperplane can intersect with grows like $\delta^{-(d-1)}$, resulting in the δ^{d-1} factor in the lower bound. However this dependence on δ and the dimension is far from desirable. We conjecture that allowing the learner to query arbitrary half-spaces does not help lower the query complexity, and that the querying strategies we construct for the proofs of Theorem 4 and Theorem 5 remain optimal.

Next we prove Theorem 4 and Theorem 5. To avoid repetition we only outline the proofs and highlight the parts that differ from the one-dimensional case.

[2] Here the δ-covering number is defined in terms of the $\|\cdot\|_\infty$ norm in $\mathbb{R}^d$.

Proof of Theorem 4. **Upper bound:** Consider the following multistage querying strategy:

1. For each $i = 1, ..., d$, submit $K_1 = \lfloor \log(1/(\lceil\gamma\rceil\delta)) \rfloor$ queries on X_i^* via bisection. This stage locates X^* in a cube $J = [a_1, b_1] \times ... \times [a_d, b_d]$ of diameter $2^{-K_1} \geq \lceil\gamma\rceil\delta$. This stage involves dK_1 queries.
2. For each $i = 1, ..., d$, run replicated bisection on $[a_i, b_i]$: First evenly split $[a_i, b_i]$ into $\lceil\gamma\rceil$ subintervals. For all q that are endpoints of the subintervals, query the events $\{X_i^* \geq q\}$. From the responses determine the true subinterval X_i^* is in. Run bisection on this subinterval to find X_i^* up to ϵ-accuracy and submits cloned queries in all the other subintervals. This stage involves $d(\lceil\gamma\rceil - 1 + \lceil\gamma\rceil\lceil\log(2\delta/\epsilon)\rceil)$ queries.

To show this strategy is (δ, L)-private, notice that from the adversary's perspective, for each i there are $\lceil\gamma\rceil$ subintervals that contain X_i^* with equal probability. That creates $L' = \lceil\gamma\rceil^d$ cubes $J_1, ..., J_{L'}$ that are at least δ apart in $\|\cdot\|_\infty$ distance. Since $L' \geq L$, the adversary cannot achieve $\|\widetilde{X} - X^*\|_\infty \leq \delta$ with probability greater than $1/L$.

Lower bound: Suppose ϕ is an ϵ-accurate and (δ, L)-private strategy that submits at most n queries. By assumption the query sequence on X_i^* depends only on X_i^* and Y_i. Thus we can write $\mathbf{n}_i(X_i^*, Y_i)$ for the number of queries submitted on the i'th coordinate. Let $n_i = \sup_{X_i^*, Y_i} \mathbf{n}_i(X_i^*, Y_i)$, so that $n \geq \sum_{i\leq d} n_i$. As in the one-dimensional proof we can assume that the learner always submits exactly n_i queries on X_i^* by filling up the end of the query sequence with trivial queries on $\{X_i^* \geq 0\}$.

Consider an adversary that adopts the truncated proportional-sampling scheme on each coordinate. Let $q_i = (q_{i,1}, ..., q_{i,n_i})$ denote the sequence of queries submitted on X_i^*. The adversary obtains $\widetilde{X}_i$ by sampling from the empirical distribution of $q_{i,K_2+1}, ..., q_{i,n_i}$ with $K_2 = \lfloor\log(1/(\gamma\delta))\rfloor$. Since q_i only depends on X_i^* and Y_i with $\{(X_i^*, Y_i)\}_{i\leq d}$ mutually independent, we have that the sequences $q_1, ..., q_d$ are also mutually independent. Thus

$$\begin{aligned}\mathbb{P}\left\{\left\|\widetilde{X} - X^*\right\|_\infty \leq \delta/2\right\} =& \mathbb{E}\prod_{i\leq d}\left(\frac{\sum_{j=K_2+1}^{N_i} \mathbb{1}\{|X_i^* - q_{i,j}| \leq \delta/2\}}{n_i - K_2}\right) \\ =& \prod_{i\leq d}\left(\frac{\sum_{j=K_2+1}^{N_i} \mathbb{P}\{|X_i^* - q_{i,j}| \leq \delta/2\}}{n_i - K_2}\right) \leq \frac{1}{L}.\end{aligned}$$

Via the same analysis as in the one-dimensional proof,

$$\sum_{j=K_2+1}^{n_i} \mathbb{P}\{|X_i^* - q_{i,j}| \leq \delta/2\} \geq \log(\delta/4\epsilon) - \delta 2^{K_2}.$$

Deduce that

$$\prod_{i\leq d}(n_i - K_2) \geq L\left(\log(\delta/4\epsilon) - \delta 2^{K_2}\right)^d.$$

Given the lower bound on $\prod_{i\leq d}(n_i - K_2)$, the minimal value for $\sum_{i\leq d}(n_i - K_2)$ is achieved when all the summands are equal. Hence the total number of queries is at least

$$\sum_{i\leq d} n_i \geq dK_2 + d\gamma\left(\log(\delta/4\epsilon) - \delta 2^{K_2}\right) \geq d\left(K_2 + \gamma\left(\log\frac{\delta}{\epsilon} - 2\right) - 1\right),$$

where the second inequality is from the choice of K_2. □

Proof of Theorem 5. **Upper bound:** As in the one-dimensional proof, the upper bound is proved by constructing a querying strategy that first submits L guesses (intervals of length ϵ) that are at least δ apart. In $[0, 1]^d$, we submit $\lceil\gamma\rceil$ guesses on the location of X_i^* for each coordinate $i \leq d$. These guesses across the d coordinates form $L' = \lceil\gamma\rceil^d$ cubes of diameter ϵ, all of which are contained in the information set $\mathcal{I}(\bar{q})$. Moreover the centers of these L' cubes are at least δ away from each other in $\|\cdot\|_\infty$ norm. Since $L' \geq L$, the δ-covering number of $\mathcal{I}(\bar{q})$ is at least L.

The way the guesses are submitted following algorithm 2 when $\delta \leq 2^{-\lceil\gamma\rceil}$ and algorithm 3 when $\delta > 2^{-\lceil\gamma\rceil}$, except that L is replaced with $\lceil\gamma\rceil$ in the algorithms. Each guess consists of two queries ϵ away from each other. In total

it takes $2d\lceil\gamma\rceil$ queries to submit all the guesses. If none of the guesses is correct, the guesses help the learner narrow down the range of X_i^* to an interval J_i. Via similar analysis as in the one-dimensional proof, we have $|J_i| = 2^{-\lceil\gamma\rceil}$ when $\delta \le 2^{-\lceil\gamma\rceil}$ and $|J_i| \in [\delta, 2\delta]$ otherwise. The next stage of the strategy simply runs bisection in J_i to achieve ϵ-accuracy on X_i^*, which requires at most $\log(|J_i|/\epsilon) \le \log(\max\{2^{-\lceil\gamma\rceil}, \delta\}/\epsilon) + 1$ queries.

Lower bound: First consider the case $\delta \le 2^{-\lceil\gamma\rceil}$. From the lower bound proof of Theorem 2, we have for each $i \le d$,

(i) There exists an interval J_i of length 2δ such that if X_i^* is in J_i, then there are at least $\log(1/\delta) - 3$ queries on X_i^* that are outside of J_i and are separated from each other by at least δ;

(ii) For each interval J of length 2δ, there exists $x_i \in J$ such that if x_i is the true value for X_i^*, then there are at least $\log(\delta/\epsilon)$ queries on X_i^* that are in J.

The above guarantee that there are at least $d(\log(1/\epsilon) - 3)$ queries in total. The extra queries arise from the privacy requirement. For a point x to enter the information set $\mathcal{I}(\bar{q})$, there must be at least 2 queries on X_i^* surrounding x_i, that are ϵ-close to each other. Hence $\mathcal{I}(\bar{q})$ is contained in the union of d-dimensional hyperrectangles $\prod_{i\le d}[s_i, t_i]$, where s_i, t_i are pairs of queries on X_i^* with $0 < t_i - s_i \le \epsilon$. Suppose aside from the queries identified by (i), there are m_i extra queries on X_i^* outside of J_i. Approximately speaking, the queries outside of $\prod_{i\le d} J_i$ form at most $\prod_{i\le d} m_i$ hyperrectangles that are contained in $\mathcal{I}(\bar{q})$. Therefore the δ-covering number of $\mathcal{I}(\bar{q})$ is at most $\prod_{i\le d} m_i$. Deduce that $\prod_{i\le d} m_i \ge L$ and thus $\sum_{i\le d} m_i \ge d\gamma$. The queries identified in (i),(ii) plus $\sum_{i\le d} m_i$ is approximately the lower bound stated in Theorem 5. The extra constant -5 is to account for the hyperrectangles that are formed with queries that are either near the boundary of $[0,1]^d$ or near $\prod_{i\le d} J_i$.

The proof for the $\delta > 2^{-\lceil\gamma\rceil}$ case is much simpler. Fix any cube $J = \prod_{i\le d} J_i$ of diameter δ. For each $i \in [d]$, there are at least $\log(\delta/\epsilon)$ queries about X_i inside of J_i. Outside of J there are at least $2d\gamma$ queries to ensure that the information set contains at least L hyperrectangles. The proof is similar to the previous case and is therefore omitted. □

5 Discussion on Federated Learning

In the Introduction we briefly discussed the application of our work to Federated Learning. In this section we investigate this example in more depth.

Before explaining how the Private Sequential Learning model applies in this context, we briefly review the basic mechanisms of Federated Learning. In Federated Learning, a central learner trains a global model by interacting with a large pool of users $\mathcal{U}$. Suppose a central learner aims to estimate the optimal model parameter that minimizes the population risk, i.e., $\theta^* \in \arg\min_\theta L(\theta)$ where $L(\theta) = \mathbb{E}\ell(Z, \theta)$, ℓ denotes the loss function and the average is taken with respect to the underlying data distribution F of Z. Each user $u \in \mathcal{U}$ has access to a local dataset $\{Z_j, j \in S_u\}$ where $Z_j \overset{i.i.d.}{\sim} F$. A typical Federated Learning training process is sequential in nature and is outlined below (see e.g. [7, Algorithm 1]).

Algorithm 4: A Federated Learning framework known as *FederatedAveraging*

Initialize θ_0;
for *iteration* $i = 1, 2, ...$ **do**
  Sample a subset of users $\mathcal{U}_i \subset \mathcal{U}$;
  for *user* $u \in \mathcal{U}_i$ **do**
    Define local loss function $\ell_u(\theta_i) := \frac{1}{|S_u|}\sum_{j\in S_u} \ell(Z_j, \theta_i)$;
    Train local model update θ_i^u, for example by running one, or multiple steps of gradient descent on $\ell_u(\cdot)$ from θ_i;
  Aggregate θ_i^u across all $u \in \mathcal{U}_i$ to produce θ_{i+1};

When training with thousands of users, as the learner lacks enough administrative power over those external workers, the Federated Learning system is highly vulnerable to eavesdropping attacks [5]. An *honest-but-curious*

adversary can participate in the training stage by pretending to be an user, and eavesdrop on the sequence of broadcasted model parameters. Simply by taking the last set of model parameters, the adversary can approximate the learner's final model fairly well. Sophisticated models can be worth millions. The eavesdropper can use the stolen model to profit or even leverage them for illicit purposes [3]. It not only saves the eavesdropper from investing the tremendous amounts of funding into training the model, but it could also devalue the learner's model. Therefore, it is of paramount importance to protect the learner's privacy from eavesdropping attacks [3]. This consideration prompts us to investigate whether we can offer provable guarantees on the learner's privacy against the eavesdropping attack in Federated learning.

There are several potential techniques to conceal the model parameters from the users in Federated Learning, such as restricting each user to run the local computation inside a Trusted Executation Envrionments (TEE) [14] or encrypting the model parameters under a homomorphic encryption scheme before broadcasting it to the users [8]. Unfortunately, as pointed out by the recent survey [5, Section 4.3.3], TEEs may not be generally available across all workers especially when these workers represent end-devices such as smartphones. Moreover, TEEs and homomorphic encryption are often costly to implement and incurs large overhead. There is an emerging line of research on preventing model theft in the evaluation stage, where an adversary attempts to extract the deployed model by repeatedly querying the model and obtaining estimation on the input feature vectors [15, 12, 13, 17, 3, 11, 10, 6]. However, this line of work does not address the unique challenge of concealing the model parameters from the users during the training stage in Federated Learning.

By investigating the binary search model, we aim to address the following two natural but fundamental questions:

1. Can the learner arrive at an accurate model, while ensuring that the eavesdropping adversary cannot learn the same model with a high level of accuracy?

2. What is the minimal number of iterations needed in the training process, for accurate and private learning?

At a high level, when the model parameters are in one dimension, here is how we associate the Federated Learning framework with the binary search model. The optimal model parameter θ_* corresponds to the true unknown value the learner aims to learn. The model parameter θ_i is viewed as a query. The queries are broadcasted to the participating users, and is therefore assumed to be accessed by the adversary. The binary variable $\mathbb{1}\{\theta_* > \theta_i\}$, or its noisy variant, is associated with to the model updates the learner receives from the users. It is reasonable to assume that the adversary does not observe the responses. That is because in order to observe the responses, the adversary would have to access the updates generated by all users in the system, a formidable task that is not realistic for an adversary that only controls up to a small subset of the users. To be clear, we do not claim that the binary search model that we consider in the paper covers the general family of Federated Learning frameworks. But it is an idealized abstraction in one dimension, if we view the aggregated model updates as providing information on the direction of θ_* relative to the current broadcasted model parameter θ_i. To be more specific, consider the two special cases below.

1. Suppose the local model updates θ_i^u are obtained by running one step of gradient descent, i.e. $\theta_i^u = \theta_i - \eta \nabla \ell_u(\theta_i)$. This implementation of Federated Learning is known as *FederatedSGD*, or *FedSGD* [7]. By averaging over the model updates across all users in the i'th iteration, the learner obtains the weighted average

$$\hat{g}(\theta_i) := \sum_{u \in \mathcal{U}_i} w_u \nabla \ell_u(\theta_i) = \sum_{u \in \mathcal{U}_i} \frac{1}{|S_u|} \sum_{j \in S_u} \nabla \ell(Z_j, \theta_i) = \frac{1}{\sum_{u \in \mathcal{U}_i} |S_u|} \sum_{u \in \mathcal{U}_i} \sum_{j \in S_u} \nabla \ell(Z_j, \theta_i),$$

   where the weights $w_u = |S_u| / \sum_{u \in \mathcal{U}_i} |S_u|$ are defined proportional to the size of the users' local datasets. As the total number of data points $\sum_{u \in \mathcal{U}_i} |S_u|$ grows to infinity, with adequate regularity conditions, $\hat{g}(\theta_i)$ converges almost surely to $\mathbb{E}[\nabla \ell(Z, \theta_i)] = \nabla L(\theta_i)$ by the strong law of large numbers. If the loss function ℓ is convex, then L is also convex, and therefore

$$\operatorname{sign}\left(\hat{g}(\theta_i)\right) \to \mathbb{1}\left\{\nabla L(\theta_i) > 0\right\} = \mathbb{1}\left\{\theta_* < \theta_i\right\}.$$

[3] Although in Federated Learning the final trained model is usually released for all users to access, the learner often chooses to keep the model parameters in the central server and only allows users to perform evaluation tasks.

When the total number of data points $\sum_{u\in\mathcal{U}_i}|S_u|$ is small, sign $(\hat{g}(\theta_i))$ can be viewed a noisy version of the true directional information $\mathbb{1}\{\theta_* < \theta_i\}$. By viewing $\text{sign}(\hat{g}(\theta_i))$ as the response, the learner can apply the querying strategies developed in this paper and achieve private learning.

2. If the local model updates are obtained by running multiple steps of gradient descent, it is likely that the direction of the local updates indicate the direction of the optimal model parameter under the local loss function. To be more specific, suppose $\text{sign}(\theta_i^u - \theta_i) = \text{sign}(\theta_*^u - \theta_i)$, where $\theta_*^u \in \arg\min_\theta \ell_u(\theta)$.

   As a special case, suppose ℓ is the ℓ_1 loss $\ell(Z,\theta) = |Z-\theta|$. Then $\theta_* \in \arg\min_\theta L(\theta)$ is the (population) median of F, and θ_*^u is the sample median over user u's local dataset $\{Z_j\}_{j\in S_u}$. Ignoring the complication surround non-unique medians, we have

$$\mathbb{P}\left\{\theta_*^u \geq \theta_*\right\} = \mathbb{P}\left\{\sum_{j\in S_u}\mathbb{1}\{Z_j \geq \theta_*\} \geq \frac{|S_u|}{2}\right\} = \frac{1}{2},$$

   where the last equality is because $\sum_{j\in S_u}\mathbb{1}\{Z_j \geq \theta_*\} \sim \text{Bin}(|S_u|, 1/2)$. As a result, for all $\theta_i \leq \theta_*$,

$$\mathbb{P}\left\{\text{sign}\left(\theta_*^u - \theta_i\right) = 1\right\} = \mathbb{P}\left\{\theta_*^u \geq \theta_i\right\} \geq 1/2.$$

   Similarly, for all $\theta_i > \theta_*$, $\mathbb{P}\{\text{sign}(\theta_*^u - \theta_i) = 1\} \leq 1/2$. Therefore the majority vote of $\text{sign}(\theta_i^u - \theta_i)$ can be viewed as a deterministic, or noisy version of $\mathbb{1}\{\theta_* \geq \theta_i\}$, depending on the number of users $|\mathcal{U}_i|$. The majority vote corresponds to the response r_i under the binary search model.

   Following similar arguments, it is easy to check that under the ℓ_2 loss function, the majority vote of $\text{sign}(\theta_i^u - \theta_i)$ is also positively correlated with $\mathbb{1}\{\theta_* \geq \theta_i\}$, with the additional assumption that F is a symmetric distribution.

We remark that in Federated Learning, communication bandwidth is a scarce resource, as the data transmission between the external workers and the learner typically suffers from high latency and low throughput. Thus, determining the optimal query complexity (i.e. the minimum communication rounds) is of fundamental importance in both theory and practice. The rigorous study we carry out offers a complete understanding on the trade-off between accuracy, privacy and query complexity under the binary search model, and is likely to yield valuable insights and provide guidance for algorithm design under the general Federated Learning framework.

References

[1] Marat Valievich Burnashev and Kamil'Shamil'evich Zigangirov. An interval estimation problem for controlled observations. *Problemy Peredachi Informatsii*, 10(3):51–61, 1974.

[2] John C Duchi and Martin J Wainwright. Distance-based and continuum fano inequalities with applications to statistical estimation. *arXiv preprint arXiv:1311.2669*, 2013.

[3] Mika Juuti, Sebastian Szyller, Samuel Marchal, and N Asokan. Prada: protecting against dnn model stealing attacks. In *2019 IEEE European Symposium on Security and Privacy (EuroS&P)*, pages 512–527. IEEE, 2019.

[4] Thomas Kailath. The divergence and bhattacharyya distance measures in signal selection. *IEEE transactions on communication technology*, 15(1):52–60, 1967.

[5] Peter Kairouz, H Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Keith Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, et al. Advances and open problems in federated learning. *arXiv preprint arXiv:1912.04977*, 2019.

[6] Sanjay Kariyappa and Moinuddin K Qureshi. Defending against model stealing attacks with adaptive misinformation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2020.

[7] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. Communication-efficient learning of deep networks from decentralized data. *arXiv preprint arXiv:1602.05629*, 2016.

[8] Payman Mohassel and Yupeng Zhang. Secureml: A system for scalable privacy-preserving machine learning. In *2017 IEEE Symposium on Security and Privacy (SP)*, pages 19–38. IEEE, 2017.

[9] Wolfgang Mulzer. Five proofs of chernoff's bound with applications. *arXiv preprint arXiv:1801.03365*, 2018.

[10] Tribhuvanesh Orekondy, Bernt Schiele, and Mario Fritz. Knockoff nets: Stealing functionality of black-box models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4954–4963, 2019.

[11] Tribhuvanesh Orekondy, Bernt Schiele, and Mario Fritz. Prediction poisoning: Towards defenses against dnn model stealing attacks. In *International Conference on Learning Representations*, 2019.

[12] Nicolas Papernot, Patrick McDaniel, Ian Goodfellow, Somesh Jha, Z Berkay Celik, and Ananthram Swami. Practical black-box attacks against machine learning. In *Proceedings of the 2017 ACM on Asia conference on computer and communications security*, pages 506–519, 2017.

[13] Yi Shi, Yalin Sagduyu, and Alexander Grushin. How to steal a machine learning classifier with deep learning. In *2017 IEEE International Symposium on Technologies for Homeland Security (HST)*, pages 1–5. IEEE, 2017.

[14] Pramod Subramanyan, Rohit Sinha, Ilia Lebedev, Srinivas Devadas, and Sanjit A Seshia. A formal foundation for secure remote execution of enclaves. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 2435–2450, 2017.

[15] Florian Tramèr, Fan Zhang, Ari Juels, Michael K Reiter, and Thomas Ristenpart. Stealing machine learning models via prediction apis. In *25th {USENIX} Security Symposium ({USENIX} Security 16)*, pages 601–618, 2016.

[16] John N Tsitsiklis, Kuang Xu, and Zhi Xu. Private sequential learning. *Forthcoming in Operations Research*, 2020. arXiv:1805.02136.

[17] Binghui Wang and Neil Zhenqiang Gong. Stealing hyperparameters in machine learning. In *2018 IEEE Symposium on Security and Privacy (SP)*, pages 36–52. IEEE, 2018.

[18] Kuang Xu. Query complexity of Bayesian private learning. In *Advances in Neural Information Processing Systems*, pages 2431–2440, 2018.

[19] Kuang Xu. Query complexity of bayesian private learning. *arXiv preprint arXiv:1911.06903*, 2019.