Contraction of Locally Differentially Private Mechanisms

Shahab Asoodeh, Member, IEEE, and Huanyu Zhang

Abstract—We investigate the contraction properties of locally differentially private mechanisms. More specifically, we derive tight upper bounds on the divergence between PK and QK output distributions of an $\epsilon$-LDP mechanism K in terms of a divergence between the corresponding input distributions P and Q, respectively. Our first main technical result presents a sharp upper bound on the $\chi^2$-divergence $\chi^2(PK\|QK)$ in terms of $\chi^2(P\|Q)$ and $\epsilon$. We also show that the same result holds for a large family of divergences, including KL-divergence and squared Hellinger distance. The second main technical result gives an upper bound on $\chi^2(PK\|QK)$ in terms of total variation distance TV(P, Q) and $\epsilon$. We then utilize these bounds to establish locally private versions of the van Trees inequality, Le Cam’s, Assouad’s, and the mutual information methods—powerful tools for bounding minimax estimation risks. These results are shown to lead to tighter privacy analyses than the state-of-the-arts in several statistical problems such as entropy and discrete distribution estimation, non-parametric density estimation, and hypothesis testing.

Index Terms—Differential privacy, data processing inequality, contraction coefficient, minimax estimation risk, f-divergences.

I. INTRODUCTION

LOCAL differential privacy (LDP) has now become a standard definition for individual-level privacy in machine learning. Intuitively, a randomized mechanism (i.e., a channel) is said to be locally differentially private if its output does not vary significantly with arbitrary perturbation of the input. More precisely, a mechanism is $\epsilon$-LDP if the privacy loss random variable, defined as the log-likelihood ratio of the output for any two different inputs, is smaller than $\epsilon$.

Since its formal introduction [36], [45], LDP has been extensively incorporated into statistical problems, e.g., locally private mean estimation problem [2], [3], [4], [10], [14], [15], [20], [26], [32], [33], [34], [39], [40], [41], [51], [54], and locally private distribution estimation problem [2], [7], [17], [37], [38], [43], [53], [58]. The fundamental limits of such statistical problems under LDP are typically characterized using information-theoretic frameworks such as Le Cam’s, Assouad’s, and Fano’s methods [59]. A critical building block for sharp privacy analysis in such methods turns out to be the contraction coefficient of LDP mechanisms. Contraction coefficient $\eta_f(K)$ of a mechanism K under an f-divergence is a quantification of how much the data processing inequality can be strengthened: It is the smallest $\eta \leq 1$ such that $D_f(PK\|QK) \leq \eta D_f(P\|Q)$ for any distributions P and Q, where PK denotes the output distribution of K when its input is sampled from P.

Studying statistical problems under local privacy through the lens of contraction coefficients was initiated by Duchi et al. [33], [35] in which sharp minimax risks for locally private mean estimation problems were characterized for sufficiently small $\epsilon$. As the main technical result, they showed that any $\epsilon$-LDP mechanism K satisfies

$$D_{KL}(PK\|QK) \leq \min\{4, e^{2\epsilon}\} (e^\epsilon - 1)^2 TV^2(P, Q).$$

(1)

where $D_{KL}(\cdot\|\cdot)$ and TV(\cdot, \cdot) denote KL-divergence and total variation distance, respectively. In light of the Pinsker’s inequality $2TV^2(P, Q) \leq D_{KL}(P\|Q)$, this result gives an upper bound on $\eta_{KL}(K)$ the contraction coefficient under KL-divergence. However, thanks to the data processing inequality, this bound becomes vacuous if the coefficient in (1) is strictly bigger than 1 (i.e., $\epsilon$ is not sufficiently small). More recently, Duchi and Ruan [34, Proposition 8] showed a similar upper bound for $\chi^2$-divergence:

$$\chi^2(PK\|QK) \leq 4(e^{2\epsilon} - 1)^2 TV^2(P, Q).$$

(2)

According to Jensen’s inequality $4TV^2(P, Q) \leq \chi^2(P\|Q)$, and thus (2) implies an upper bound on $\eta_{\chi^2}(K)$ the contraction coefficient under $\chi^2$-divergence. Analogously, this bound is non-trivial only for sufficiently small $\epsilon$. Similar upper bounds on the contraction coefficients under total variation distance and hockey-stick divergence were determined in [44] and [15], respectively. Results of this nature are recurrent themes in privacy analysis in statistics and machine learning, see [2], [4], [5], [6] for other examples of such results.

In this work, we develop a framework for characterizing tight upper bounds on $D_{KL}(PK\|QK)$ and $\chi^2(PK\|QK)$ for any LDP mechanisms. We achieve this goal via two different approaches: (i) indirectly by bounding $\eta_{KL}(K)$ and $\eta_{\chi^2}(K)$, and (ii) directly by deriving inequalities of the form (1) and (2) that are considerably tighter for all $\epsilon \geq 0$. In particular, our main contributions are:

1) We obtain a sharp upper bound on $\eta_{\chi^2}(K)$ for any $\epsilon$-LDP mechanism K in Theorem 1, and show that this bound...
Summary of the Minimax Risks for ε-LDP Estimation, Where We Have Omitted Constants for All the Results. For the ℓ₀-Distribution Estimation, our Upper Bound Relies on Some Mild Condition Discussed in the Supplementary Material. Note That Our Result Is Order Optimal in n and d for the Dense Case Unless ε ≥ log d. For the Gaussian Location Model, We Assume That Θ Is the Unit ℓ₂-ball and V_d Is the Volume of the Unit ℓ∞-Ball (for Arbitrary Norm). Corollary 5 Concerns with the General Θ.

| Problem | UB | Previous LB | LB |
|---------|----|-------------|----|
| Entropy estimation | N.A. | N.A. | $\min\left\{1, \frac{1}{n}\left(\epsilon^2 + 1\right)^2\right\} \log^2 k$ (Corollary 2) |
| Distribution estimation, ℓ₀-loss | $\frac{e^{\epsilon(1-\epsilon^c)}}{\sqrt{n(\epsilon^c-1)^2}}$ (Theorem 5) | $\min\left\{1, \frac{e^{\epsilon^c/2d^1/h}}{\sqrt{n(\epsilon^c-1)^2}} \left(\frac{e^{\epsilon^c/2}}{\sqrt{n(\epsilon^c-1)^2}}\right)^{1-1/h}\right\}$ (Corollary 3), [2] |
| Density estimation, ℓ₁-loss, β-Hölder | N.A. | $\left(\frac{n\epsilon}{2}\right)^{\frac{1}{2+\beta}}$ for $\epsilon \leq 1$ [22] | $\frac{n\epsilon^{-\epsilon}(\epsilon^2 - 1)^{1/2}}{\sqrt{n(\epsilon^c-1)^2}}$ (Corollary 4) |
| Gaussian location model, arbitrary loss | N.A. | N.A. | $\frac{\sqrt{\epsilon}}{\epsilon^2(V_d^2(1+d)^{1/2})} \min\left\{1, \frac{\sqrt{\epsilon}}{\epsilon^2(V_d^2(1+d)^{1/2})}\right\}$ (Corollary 5) |
| Sample complexity of hypothesis testing | $\frac{\epsilon^{3\epsilon}}{(\epsilon^c-1)^2} \max\left\{\frac{1}{\nu(P,Q)}, \frac{1}{H^2(P,Q)}\right\}$ (Lemma 5) | $\epsilon^e$ (Corollary 5) |

holds for a large family of divergences, including KL-divergence and squared Hellinger distance.

2) We derive upper bounds for $D_{KL}(PK \| QK)$ and $\chi^2(PK \| QK)$ in terms of $TV(P,Q)$ and the privacy parameter ε in Theorem 2. While upper bounds in (1) and (2) scale as $O(e^{2\epsilon})$ and $O(e^{\epsilon^c})$, respectively, ours scales as $O(e^\epsilon)$, thus significantly improving over those bounds for practical range of $\epsilon$ (that is $\epsilon \geq \frac{1}{2}$).

3) We use our main results to develop a systemic framework for quantifying the cost of local privacy in several statistical problems under the “sequentially interactive” setting. Our framework enables us to improve and generalize several existing results, and also produce new results beyond the reach of existing techniques. In particular, we study the following problems:

a) Locally private Fisher information: We show that the Fisher information matrix $I_P(\theta)$ of parameter $\theta$ given a privatized sequence $Z^n := (Z_1, \ldots, Z_n)$ of $X^n \overset{iid}{\sim} P_\theta$ satisfies $I_P(\theta) \leq n\left[\epsilon^c \epsilon^c - 1\right]^2 I_N(\theta)$ (Lemma 1). This result then directly leads to a private version of the van Trees inequality (Corollary 1) that is a classical approach for lower bounding the minimax quadratic risk. Similarly, this result leads to a private version of the Cramér-Rao bound, provided that there exist unbiased private estimators (see [16] for more details). It is worth noting that Barnes et al. [17] recently investigated locally private Fisher information under certain assumptions regarding the regularity of $P_\theta$. More specifically, they derived various upper bounds on $Tr(I_P(\theta))$ for $\epsilon \geq 0$, when $\log f_\theta(X)$ is either sub-exponential or sub-Gaussian, or when $E[(u^T \log f_\theta(X))^2]$ is bounded for any unit vector $u \in \mathbb{R}^d$, where $f_\theta$ is the density of $P_\theta$ with respect to the Lebesgue measure. In contrast, Lemma 1 establishes a similar upper bound for small $\epsilon$ (i.e., $\epsilon \in [0, 1]$) but without imposing such regularity conditions.

b) Locally private Le Cam’s and Assouad’s methods: Following [33], we establish locally private versions of Le Cam’s and Assouad’s methods [46], [59] that are demonstrably stronger than those presented in [33] (Theorems 3 and 4). We then use our private Le Cam’s method to study the problem of entropy estimation under LDP where the underlying distribution is known to be supported over $\{1, \ldots, k\}$ (Corollary 2). As applications of our private Assouad’s method, we study two problems. First, we derive a lower bound for $\ell_\theta$ minimax risk in the locally private distribution estimation problem which improves the constants of the state-of-the-art lower bounds [58] in the special cases $h = 1$ and $h = 2$, and leads to the same order analysis for general $h \geq 1$ in [2]. We also provide an upper bound by generalizing the Hadamard response [7] to $\ell_h$-norm with $h \geq 2$ which matches the lower bound under some mild conditions. Second, we study private non-parametric density estimation when the underlying density is assumed to be Hölder continuous and derive a lower bound for $\ell_h$ minimax risk in Corollary 4. Unlike the best existing result [22], our lower bound holds for all $\epsilon \geq 0$.

c) Locally private mutual information method: Recently, mutual information method [56, Section 11] has been proposed as a more flexible information-theoretic technique for bounding the minimax risk. We invoke Theorem 1 to provide (for the first time) a locally private version of the mutual information bound in Theorem 6. To demonstrate the flexibility of this result, we consider the Gaussian location model where the goal is to privately estimate $\theta$ from $X^n \overset{iid}{\sim} N(\theta, \sigma^2 I_d)$. Most existing results (e.g., [17], [32], [33], [34]) assume $\ell_2$-norm as the loss and unit $\ell_\infty$-ball or unit $\ell_2$-ball as $\Theta$. However, our result presented in Corollary 5 holds for any arbitrary loss functions and any arbitrary set $\Theta$ (e.g., $\ell_h$-ball for any $h \geq 1$).

d) Locally private hypothesis testing: Given $n$ i.i.d. samples and two distributions $P$ and $Q$, we derive upper and lower bounds for $\mathcal{S}_C^{P,Q}$, the sample complexity of privately determining which distribution generates the samples. More precisely, we show in Lemma 2 that in the sequentially interactive (in fact, in the more general fully interactive)
setting $SC^ε_{P,Q} ≥ \frac{ε}{(ε-1)^2} \max\{\frac{1}{TV^2(P,Q)}, \frac{ε}{H^2(P,Q)}\}$ and $SC^ε_{P,Q} ≤ \frac{ε^2}{(ε-1)^2} \frac{1}{TV^2(P,Q)}$ for any $ε ≥ 0$, where $H^2(P,Q)$ is the squared Hellinger distance between $P$ and $Q$. These bounds subsume and generalize the best existing result in [33] which indicates $SC^ε_{P,Q} = θ\left(\frac{1}{εTV^2(P,Q)}\right)$ for sufficiently small $ε$. Furthermore, they have recently been shown in [49, Th. 1.6] to be optimal (up to a constant factor) for any $ε ≥ 0$ if $P$ and $Q$ are binary. This, in fact, implies that (sequential or full) interaction does not help in the locally private hypothesis testing problem if $P$ and $Q$ are binary or if $ε ≤ 1$. Therefore, our results extend [42, Th. 5.3] that indicates the optimal mechanism is non-interactive for $ε ≤ 1$.

A. Additional Related Work

Local privacy is arguably one of the oldest forms of privacy in statistics literature and dates back to Warner [55]. This definition resurfaced in [36] and was adopted in the context of differential privacy as its local version. The study of statistical efficiency under LDP was initiated in [33], [35] in the minimax setting and has since gained considerable attention. While the original bounds on the private minimax risk in [33], [35] were meaningful only in the high privacy regime (i.e., small $ε$), the order optimal bounds were recently given for several estimation problems in [32] for the general privacy regime. Interestingly, their technique relies on the decay rate of mutual information over a Markov chain, which is known to be equivalent to the contraction coefficient under KL-divergence [13]. However, their technique is quite different from ours in that it did not concern computing the contraction coefficient of an LDP mechanism.

Among locally private statistical problems studied in the literature, two examples have received considerably more attention, namely, mean estimation and discrete distribution estimation. For the first problem, Duchi et al. [35] used (1) to develop asymptotically optimal procedures for estimating the mean in the high privacy regime (i.e., $ε < 1$). For the high privacy regime (i.e., $ε > 1$), a new algorithm was proposed in [20] that is optimal and matches the lower bound derived in [32] for interactive mechanisms. There has been more work on locally private mean estimation that studies the problem under additional constraints [2], [3], [10], [14], [15], [17], [38], [40], [41], [51], [54]. For the second problem, Duchi et al. [33] studied (non-interactive) locally private distribution estimation problem under $ε_1$ and $ε_2$ loss functions and derived the first lower bound for the minimax risk, which was shown to be optimal [43] for high privacy regime. Follow-up works such as [2], [17], [37], [53], [58] characterized the optimal minimax rates for general $ε$. Recently, [2] derived a lower bound for $ℓ_h$ loss with $h ≥ 1$.

The problem of locally private entropy estimation has received significantly less attention in the literature, despite the vast line of research on the non-private counterpart. The only related work in this area seems to be [21], [23] which studied estimating Rényi entropy of order $λ$ and derived optimal rates only when $λ > 2$. Thus, the optimal private minimax rate seems to be still open. We remark that [8] explicitly considered the problem of entropy estimation, but in the setting of central differential privacy.

The closest work to ours are [15], [60] which extensively studied the contraction coefficient of LDP mechanisms under the hockey-stick divergence. More specifically, it was shown in [15] that $K$ is $ε$-LDP if and only if $E_{P,Q}[PK∥QK]$ the hockey-stick divergence between $PK$ and $QK$ is equal to zero for any distributions $P$ and $Q$, and thus if and only if the contraction coefficient of $K$ under the hockey-stick divergence is zero. By representing $χ^2$-divergence in terms of the hockey-stick divergence, this result leads to a conceptually similar, albeit weaker, result as Theorem 2.

In [9], Acharya et al. introduced an information-theoretic toolbox to establish lower bounds for private estimation problems. However, they considered the threat model of central differential privacy, a totally different model from the local differential privacy considered in this work.

B. Notation

We use upper-case letters (e.g., $X$) to denote random variables and calligraphic letters to represent their support. For the closest work to ours are [15], [60] which extensively studied the contraction coefficient of LDP mechanisms under the hockey-stick divergence. More specifically, it was shown in [15] that $K$ is $ε$-LDP if and only if $E_{P,Q}[PK∥QK]$ the hockey-stick divergence between $PK$ and $QK$ is equal to zero for any distributions $P$ and $Q$, and thus if and only if the contraction coefficient of $K$ under the hockey-stick divergence is zero. By representing $χ^2$-divergence in terms of the hockey-stick divergence, this result leads to a conceptually similar, albeit weaker, result as Theorem 2.

In [9], Acharya et al. introduced an information-theoretic toolbox to establish lower bounds for private estimation problems. However, they considered the threat model of central differential privacy, a totally different model from the local differential privacy considered in this work.

II. Preliminaries and Definitions

In this section, we give basic definitions of $f$-divergence, contraction coefficients, and LDP mechanisms.

$f$-Divergences and Contraction Coefficients: Given a convex function $f$ : $(0, ∞) → \mathbb{R}$ such that $f(1) = 0$, the $f$-divergence between two probability measures $P < Q$ is defined as [12], [31] $D_f(P∥Q) := E_{Q}[f(\frac{P}{Q})]$. Examples of $f$-divergences needed in the subsequent sections include:

- KL-divergence $D_{KL}(P∥Q) := D_f(P∥Q)$ for $f(t) = t \log t$,
- total-variation distance $TV(P,Q) := D_f(P∥Q)$ for $f(t) = \frac{t}{2}[t - 1]$,
- $χ^2$-divergence $χ^2(P∥Q) := D_f(P∥Q)$ for $f(t) = t^2 - 1$,
- squared Hellinger distance $H^2(P,Q) := D_f(P∥Q)$ for $f(t) = (1 - \sqrt{t})^2$, and
- hockey-stick divergence (aka $E_γ$-divergence [52]) $E_γ(P∥Q) := D_f(P∥Q)$ for $f(t) = (t - γ)^+$ for some $γ ≥ 1$, where $(t)_+ := \max(t, 0)$.

All $f$-divergences are known to satisfy the data-processing inequality. That is, for any channel $K : \mathcal{X} → \mathcal{Z}$, we have $D_f(PK∥QK) ≤ D_f(P∥Q)$ for any pair of distributions $(P,Q)$. However, this inequality is typically strict. One way to
strenthen this inequality is to consider \( \eta_f(K) \) the contraction coefficient of \( K \) under \( f \)-divergence [11] defined as

\[
\eta_f(K) := \sup_{P, Q \in \mathcal{P}(X)} \frac{D_f(Q^K || PK)}{D_f(Q || P)}.
\] (3)

With this definition at hand, we can write \( D_f(P^K || QK) \leq \eta_f(K) D_f(P || Q) \), which is typically referred to as the strong data processing inequality. We will study in details contraction coefficients under KL-divergence, \( \chi^2 \)-divergence, squared Hellinger distance, and total variation distance, denoted by \( \eta_{KL}(K) \), \( \eta_{\chi^2}(K) \), \( \eta_{TV}(K) \), respectively, in the next section. We also discuss the well-known fact about \( \eta_{KL}(K) \) [13]:

\[
\eta_{KL}(K) = \sup_{P, Q \in \mathcal{P}(X)} \frac{I(U; Z)}{I(U; X)}
\] (4)

where \( K \) is the channel specifying \( P_{Z|X} \), \( I(A; B) := D_{KL}(P_{AB} || P_A P_B) \) is the mutual information between two random variables \( A \) and \( B \), and \( U - X - Z \) denotes the Markov chain in that order. Another important property of \( \eta_{KL} \) essential for the proofs is its tensorization (see [16, Appendix A]).

Local Differential Privacy: A randomized mechanism \( K : \mathcal{X} \to \mathcal{P}(Z) \) is said to be \( \varepsilon \)-locally differentially private (\( \varepsilon \)-LDP for short) for \( \varepsilon \geq 0 \) if [36, 45] \( K(A|x) \leq e^\varepsilon K(A|x') \), for all \( A \subset Z \) and \( x, x' \in \mathcal{X} \). Let \( Q_\varepsilon \) be the collection of all \( \varepsilon \)-LDP mechanisms \( K \). It can be shown that LDP mechanisms can be equivalently defined in terms of the hockey-stick divergence:

\[
K \in Q_\varepsilon \iff E_{\varepsilon}(e^\varepsilon |K(x)| || K(x')) = 0, \forall x, x' \in \mathcal{X}.
\] (5)

Arguably, the most known LDP mechanism is the binary randomized-response mechanism, introduced by Warner [55]. For \( \mathcal{X} = \{0, 1\} \), the mechanism \( K \) be defined as \( K(1|x) = \text{Bernoulli}(\kappa) \) and \( K(0|x) = \text{Bernoulli}(1 - \kappa) \). It can be easily verified that this mechanism is \( \varepsilon \)-LDP if \( \kappa = \frac{e^\varepsilon}{e^\varepsilon + 1} \). In information theory parlance, the binary randomized response mechanism is a binary symmetric channel with crossover probability \( \frac{1}{e^\varepsilon + 1} \). A natural way to generalize this mechanism to the non-binary set is as follows.

**Example 1 (k-Ary Randomized Response):** Let \( \mathcal{X} = \mathcal{Z} = [k] \). Let the mechanism \( K \) be defined as

\[
K(z|x) = \begin{cases} 
\frac{e^\varepsilon}{e^\varepsilon + k - 1}, & \text{if } z = x, \\
\frac{e^\varepsilon + k - 1}{e^\varepsilon + k - 1}, & \text{otherwise}.
\end{cases}
\] (6)

It can be verified that \( E_{\varepsilon}(K(z|x) || K(z|x')) = 0 \) for all \( x, x' \in [n] \), implying this mechanism is \( \varepsilon \)-LDP.

Suppose there are \( n \) users, each in possession of a sample \( X_i, i \in [n] := \{1, \ldots, n\} \). User \( i \) applies a mechanism \( K_i \) to generate \( Z_i \), a privatized version of \( X_i \). The collection of such mechanisms is said to be non-interactive if \( K_i \) is entirely determined by \( X_i \) and independent of \( (X_j, Z_j) \) for \( j \neq i \). If, on the other hand, interactions between users are permitted, then \( K_i \) need not depend only on \( X_i \). In particular, the sequentially interactive [33] setting refers to the case when the input of \( K_i \) depends on both \( X_i \) and the outputs \( Z^{i-1} \) of the \((i-1)\) previous mechanisms.

**III. MAIN TECHNICAL RESULTS**

In this section, we present our main technical results. First, we establish a tight upper bound on \( \eta_{\chi^2}(K) \) for any \( \varepsilon \)-LDP mechanisms by deriving an upper bound for \( \chi^2(P^K || QK) \) in terms of \( \chi^2(P || Q) \) for any pair of distributions \( (P, Q) \). Interestingly, this upper bound is shown to hold for a large family of \( f \)-divergences, including KL-divergence and squared Hellinger distance. A similar result is known for total variation distance [44, Corollary 11]: for any \( K \in Q_\varepsilon \)

\[
\eta_{TV}(K) \leq \frac{e^\varepsilon - 1}{e^\varepsilon + 1}. \quad (7)
\]

It is known that \( \eta_f(K) \leq \eta_{TV}(K) \) for any channel \( K \) and any \( f \)-divergences (see, e.g., [28], [50]). Thus, it follows from (7) that \( \eta_{KL}(K) \leq \frac{e^\varepsilon - 1}{e^\varepsilon + 1} \) for any \( K \in Q_\varepsilon \). This upper bound holds for general \( f \)-divergences, thus it is necessarily loose. The following theorem shows that a significantly tighter bound can be obtained for specific \( f \)-divergences.

**Theorem 1:** If \( K \) is an \( \varepsilon \)-LDP mechanism, then we have for any \( \varepsilon \geq 0 \)

\[
\eta_{KL}(K) = \eta_{\chi^2}(K) = \eta_{TV}(K) \leq \left[ \frac{e^\varepsilon - 1}{e^\varepsilon + 1} \right]^2 =: \gamma_\varepsilon. \quad (8)
\]

The upper bound given in this theorem is in fact tight, that is, there exists an \( \varepsilon \)-LDP mechanism \( K \) and a pair of distributions \( (P, Q) \) such that \( \chi^2(P^K || QK) = \gamma_\varepsilon \chi^2(P || Q) \). To verify this, let \( K \) be the randomized response mechanism, and consider \( Q = \text{Bernoulli}(0.5) \) and \( P = \text{Bernoulli}(\alpha) \) for some \( \alpha \in [0, 1] \). In this case, it can be easily verified that \( \chi^2(P^K || QK) = 4\gamma \alpha^2 \) and \( \chi^2(P || Q) = 4\alpha^2 \). Therefore, binary randomized response mechanism satisfies the inequality in (8) with equality. In addition to the tightness of Theorem 1, this implies that the binary randomized response mechanism has the largest contraction coefficient among all \( \varepsilon \)-LDP mechanisms. Considering \( k \)-ary randomized response mechanisms, it is therefore expected that the contraction coefficients decrease as \( k \) increases, as proved next.

**Proposition 1 (Contraction of k-Ary Randomized Response):** Let \( K \) be the \( k \)-ary randomized response mechanism defined in (6). Then, we have

\[
\eta_{\chi^2}(K) = \frac{(e^\varepsilon - 1)^2}{(e^\varepsilon + 1)(e^\varepsilon + k - 1)}.
\]

**Proof:** Recall that

\[
\eta_{\chi^2}(K) = \sup_{P, Q \in \mathcal{P}(X)} \frac{\chi^2(P^K || QK)}{\chi^2(P || Q)}.
\]

It was recently shown in [48] that the above optimization can be restricted to pairs \( P \) and \( Q \) supported on two points in \( \mathcal{X} \) for all mechanisms \( K : \mathcal{X} \to \mathcal{P}(Z) \) with countable \( \mathcal{X} \). Therefore, due to the symmetry of \( k \)-ary randomized response mechanism, we can, without loss of generality, consider only \( P = (\alpha, 1 - \alpha, 0, \ldots, 0) \) and \( Q = (\beta, 1 - \beta, 0, \ldots, 0) \) in \( \mathcal{P}([k]) \) for \( \alpha, \beta \in (0, 1) \). In this case, we have

\[
\chi^2(P^K || QK) = \frac{(e^\varepsilon - 1)^2(e^\varepsilon + 1)(\alpha - \beta)^2}{(e^\varepsilon + k - 1)(\beta e^\varepsilon + \beta)(\beta + \beta e^\varepsilon)}.
\]
and
\[ \chi^2(P\|Q) = \frac{(\alpha - \beta)^2}{\beta \bar{\beta}}, \]
where \( \bar{\beta} = 1 - \beta \). Thus, we can write
\[ \eta_{\chi^2}(K) = \left( \frac{e^\epsilon - 1}{e^\epsilon + 1} \right)^2 \frac{\beta \bar{\beta}}{(\beta \bar{e} + \bar{\beta})(\bar{\beta} + \bar{\beta}e^\epsilon)}. \] (9)

Define
\[ g_\epsilon(\beta) = \frac{\beta \bar{\beta}}{(\beta \bar{e} + \bar{\beta})(\bar{\beta} + \bar{\beta}e^\epsilon)}. \]

It can be verified that
\[ g_\epsilon'(\beta) = \frac{(1 - 2\beta)e^\epsilon}{(\beta \bar{e} + \bar{\beta})(\bar{\beta} + \bar{\beta}e^\epsilon)^2}. \]
thus \( \beta \mapsto g_\epsilon(\beta) \) is increasing on \((0, \frac{1}{2}]\) and decreasing on \([\frac{1}{2}, 1)\). In particular, \( g_\epsilon(\beta) \) attains its maximum at \( \beta = \frac{1}{2} \), that is,
\[ \sup_{\beta \in (0,1)} \frac{\beta \bar{\beta}}{(\beta \bar{e} + \bar{\beta})(\bar{\beta} + \bar{\beta}e^\epsilon)} = \frac{1}{(1 + e^\epsilon)^2}. \]

Plugging this in (9), the desired result follows.

According to this result, the contraction coefficient of k-ary randomized response mechanism decreases as \( k \) increases, see Fig. 1.

**Remark 1:** Proof of Theorem 1 reveals that the same result holds for a larger family of \( f \)-divergences. In fact, it can be shown that \( \eta_f(K) \leq \chi_f \) for \( K \in Q_e \) if \( f \) is a non-linear “operator-convex” function, see, e.g., [50, Sec. III.C] and [27, Th. 1] for the definition of operator convex. The reason behind this generalization is that \( \eta_f(K) = \eta_{\chi^2}(K) \) for all non-linear operator convex \( f \), see e.g., [47, Proposition 6], [29, Proposition II.6.13 and Corollary II.6.16].

Theorem 1 turns out to be instrumental in studying several statistical problems under local privacy as discussed in Section IV. Nevertheless, it falls short in yielding a well-known fact about \( \varepsilon \)-LDP mechanisms: \( \chi^2(P\|QK) < \infty \) even if \( \chi^2(P\|Q) = \infty \). We address this issue in the next theorem which presents an upper bound for \( \chi^2(P\|QK) \) in terms of \( TV(P, Q) \), thus implying that \( \chi^2(P\|QK) \) is always finite irrespective of \( \chi^2(P\|Q) \).

**Theorem 2:** If \( K \) is an \( \varepsilon \)-LDP mechanism, then
\[ \chi^2(PK\|QK) \leq \Psi_\varepsilon \min\{4TV^2(P, Q), TV(P, Q)\}, \]
for any pair of distributions \( (P, Q) \) and \( \varepsilon \geq 0 \), where
\[ \Psi_\varepsilon := e^{-\varepsilon}(e^\epsilon - 1)^2. \] (10)

**Proof:** First, we notice that it follows from [34, Proposition 8]
\[ \chi^2(PK\|QK) \leq 4 \sup_{x, x' \in X} \chi^2(K(\cdot|x)|K(\cdot|x'))TV^2(P, Q). \] (11)

We now derive an upper bound for \( \sup_{x, x' \in X} \chi^2(K(\cdot|x)|K(\cdot|x')) \), where \( K \in Q_e \). To this goal, first note that, by invoking (5) we can write
\[ \sup_{x, x' \in X} \chi^2(K(\cdot|x)|K(\cdot|x')) \leq \sup_{M, N \in P(X)} \frac{2TV(M, N)}{\Psi_\varepsilon}. \]

To solve the latter optimization problem, we resort to the integral representation of \( \chi^2 \)-divergence in term of \( E_\gamma \) (see, e.g., [52, eq. (430)]).
\[ \chi^2(M\|N) = 2 \int_1^\infty \left[ E_\gamma(M\|N) + \gamma^{-3} E_\gamma(N\|M) \right] d\gamma. \] (12)

Since \( E_\gamma(M\|N) = 0 \) and \( E_\gamma(N\|M) = 0 \), the monotonicity and convexity of \( \gamma \mapsto E_\gamma(M\|N) \) imply that \( E_\gamma(M\|N) = E_\gamma(N\|M) = 0 \) for all \( \gamma \geq e^\epsilon \) and \( E_\gamma(M\|N) \leq \frac{TV(M, N)(e^\epsilon - \gamma)}{e^\epsilon - 1} \) for all \( \gamma \leq e^\epsilon \). Thus, it follows from (12)
\[ \chi^2(M\|N) \leq 2TV(M, N) \int_1^{e^\epsilon} (e^\epsilon - 1) \left( 1 + \gamma^{-3} \right) d\gamma \]
\[ = e^{-\epsilon}(e^\epsilon - 1)(e^\epsilon + 1)TV(M, N). \] (13)

Next, we derive an upper bound for \( TV(M, N) \) for distributions \( M \) and \( N \) satisfying \( E_\gamma(N\|M) = 0 \) and \( E_\gamma(M\|N) = 0 \):
\[ \sup_{M, N} TV(M, N). \]
(15)

First, we show that this supremum is attained with binary distributions. To this goal, define \( \phi : X \rightarrow \{0, 1\} \) as
\[ \phi(x) = \begin{cases} 1, & \text{if } dM(x) \geq dN(x), \\ 0, & \text{if } dM(x) < dN(x). \end{cases} \] (16)

Let also \( M_b \) and \( N_b \) be the Bernoulli distributions induced by push-forward of \( M \) and \( N \) via \( \phi \). It can be verified that \( TV(M, N) = TV(M_b, N_b) \). Moreover, due to the data-processing inequality, we have \( E_\gamma(M_b\|N_b) = E_\gamma(N_b\|M_b) = 0 \). Hence, we can write
\[ \sup_{M, N} TV(M, N) = \sup_{p, q \in [0,1]} TV(B(p), B(q)). \]
(17)
where \( \mathcal{B}(q) \) denotes the Bernoulli distribution for \( q \in [0, 1] \) and the last equality comes from a basic linear programming problem. Plugging this into (14), we obtain that
\[
\chi^2(M||N) \leq e^{-\varepsilon} (e^\varepsilon - 1)^2,
\]
for any pair of distributions \( M \) and \( N \) satisfying \( E_{\varepsilon}(M||N) = E_{\varepsilon}(N||M) = 0 \). This inequality, together with (11), yields
\[
\chi^2(M||N) \leq 4e^{-\varepsilon} (e^\varepsilon - 1)^2 TV^2(P, Q).
\]
(19)

We now prove the second part. Note that, we can write from (13)
\[
\chi^2(PK||QK) \leq \frac{2TV(P, QK)}{e^\varepsilon - 1} \int_1^{e^\varepsilon} (e^\gamma - 1)(1 + e^{-\gamma}) d\gamma
\]
\[
= e^{-\varepsilon} (e^\varepsilon - 1)(e^\varepsilon + 1) TV(PK, QK)
\]
\[
\leq e^{-\varepsilon} (e^\varepsilon - 1)^2 TV^2(P, Q),
\]
(20)
where the last inequality follows from (7). Combining (19) and (20), we drive the desired result.

While the proof of this theorem relies partially on the proof of \([34, \text{Proposition 8}]\), which yields (2), it is substantially stronger than (2), especially for \( \varepsilon \geq 1 \). Notice that the upper bound in (2) is of order \( e^{-\varepsilon} \) for \( \varepsilon > 1 \) while Theorem 2 gives a bound that scales as \( e^\varepsilon \). Further, note that since \( D_{KL}(P||Q) \leq \chi^2(P||Q) \), Theorem 2 also gives an upper bound on \( D(P||Q) \) in terms of \( TV(P, Q) \) which is strictly stronger than (1).

The upper bound in Theorem 2 holds for all \( \varepsilon \)-LDP mechanisms. However, for specific \( \varepsilon \)-LDP mechanisms, one can achieve a slightly tighter upper bound. For instance, next proposition shows that \( \chi^2(PK||QK) \leq \Psi TV^2(P, Q) \) for the binary mechanism \( K : \mathcal{X} \to \mathcal{P}([0, 1]) \) defined as
\[
K(0|x) = \begin{cases} \frac{e^\varepsilon}{1 + e^\varepsilon}, & \text{if } P(x) \geq Q(x), \\ \frac{1}{1+e^\varepsilon}, & \text{if } P(x) < Q(x). \end{cases}
\]
(21)

Proposition 2: For the binary mechanism, we have for any \( \varepsilon \geq 0 \)
\[
\chi^2(PK||QK) \leq \Psi TV^2(P, Q).
\]

Proof: Note that for any \( \alpha, \beta \in [0, 1] \)
\[
\chi^2(\text{Bernoulli}(\alpha)||\text{Bernoulli}(\beta)) = \frac{(\alpha - \beta)^2}{\beta^2},
\]
where \( \hat{\beta} := 1 - \beta \). Let \( A = \{ x \in \mathcal{X} : P(x) \geq Q(x) \}. \) Since \( K \) is a binary mechanism, it can be shown that \( PK \sim \text{Bernoulli}(\xi P(A^c) + \xi P(A)) \) and similarly \( QK \sim \text{Bernoulli}(\xi Q(A^c) + \xi Q(A)) \), where \( \xi = \frac{e^\varepsilon}{1+e^\varepsilon} \) and \( A^c \) is the complement of \( A \). Thus, we have
\[
\chi^2(PK||QK) = \frac{(P(A) - Q(A))^2(2\xi - 1)^2}{(\xi Q(A^c) + \xi Q(A))(\xi Q(A) + \xi Q(A^c))}.
\]
Note that by definition \( P(A) - Q(A) = TV(P||Q) \). Also, it can be easily shown that the denominator is greater than \( \xi^2 \). Thus, we can write
\[
\chi^2(PK||QK) \leq \frac{(2\xi - 1)^2}{\xi^2} TV^2(P, Q),
\]
from which the desired result follows.

IV. APPLICATIONS

In this section, we use the results presented in the previous section to examine several statistical problems under LDP constraint, including minimax estimation risks in Sections IV-A–IV-D and sample complexity of hypothesis testing in Section IV-E. In all these applications, we allow our mechanisms to be sequentially interactive.

We first define private minimax estimation risk—the main quantity needed for most subsequent sections. Suppose \( \{P_\theta\}_{\theta \in \Theta} \) is a parametric family of probability measures on \( \mathcal{X} \). If they are absolutely continuous, we denote their densities by \( \{P_\theta\}_{\theta} \) as well. Let \( X^n := (X_1, \ldots, X_n) \) be \( n \) i.i.d. samples from \( P_\theta \) that are distributed among \( n \) users. User \( i \) chooses \( K_i \in Q_{\mathcal{E}} \) to generate \( Z_i \) in a sequentially interactive manner, i.e., the distribution of \( Z_i \) depends on \( Z_{i-1} := (Z_1, \ldots, Z_{i-1}) \). More specifically, \( K_i \) receives \( X_i \) and \( Z_{i-1} \), and generates \( Z_i \). Thus, \( Z_i \sim P_{\theta} K_i \) given a realization of \( Z_{i-1} = z_{i-1} \). The goal is to estimate a function of \( \theta \), denoted by \( T(\theta) \), given the observation \( Z^n \) via an estimator \( \psi \). Invoking the minimax estimation framework to formulate this goal, we define private minimax estimation risk as
\[
R^n(n, \Theta, \ell, \varepsilon) := \inf_{K_1, \ldots, K_n \in Q_{\mathcal{E}}} \inf_{\psi \in \Theta} \sup_{\theta \in \Theta} \mathbb{E}\left[ L(\psi(Z^n), T(\theta)) \right],
\]
(22)
where \( \ell : \Theta \times \Theta \to \mathbb{R}^+ \) is a loss function assessing the quality of an estimator. Note that \( R^n(n, \Theta, \ell, \infty) \) corresponds to the non-private minimax risk. In the following sections, if \( T \) is not explicitly specified, then it is assumed to be identity, i.e., \( T(\theta) = \theta \).

A. Locally Private Fisher Information

Let the loss function be quadratic, i.e., \( \ell = \ell_2 \), and \( f(\theta) \) be the Fisher information matrix of \( \theta \) given \( X \) defined as
\[
I_X(\theta) := \mathbb{E}\left[ (\nabla \log P_\theta(X))^T (\nabla \log P_\theta(X)) \right],
\]
(22)
where the gradient is taken with respect to \( \theta \). It is well-known that an upper bound on the trace of the Fisher information matrix amounts to a lower bound on the minimax estimation risk associated with quadratic loss. This typically follows from Cramér-Rao bound (for unbiased estimators) or its Bayesian version known as van Trees inequality. Thus, it is desirable to obtain a sharp upper bound on \( \text{Tr}(I_Z(\theta)) \).

This has recently been noted in \([17]\), wherein several upper bounds for \( \text{Tr}(I_Z(\theta)) \) were derived. However, those bounds only hold when \( P_\theta \) satisfy some regularity conditions, namely \( \mathbb{E}[(u^T \nabla \log f_0(X))^2] \) is bounded for any unit vector \( u \in \mathbb{R}^d \) or \( \nabla \log f_0(X) \) is sub-Gaussian, where \( f_0 \) is the density of \( P_\theta \) with respect to the Lebesgue measure. These conditions are restrictive as they may not hold for general distributions. The following lemma gives an upper bound on \( I_{Z^n}(\theta) \) that holds for any general \( P_\theta \).

Lemma 1: Let \( X^n \overset{iid}{\sim} P_\theta \) and \( Z^n \) be the output of sequentially interactive mechanisms \( K_1, \ldots, K_n \) with \( K_i \in Q_{\mathcal{E}} \) for \( i \in [n] \). Then, we have for every \( \varepsilon \geq 0 \)
\[
I_{Z^n}(\theta) \leq n \gamma_e I_X(\theta).
\]

This lemma can be proved directly from Theorem 1 as follows. Let \( \theta' = \theta + \xi u \) for a unit vector \( u \in \mathbb{R}^d \) and \( \xi \in \mathbb{R} \).
If $P_\theta$ and $P_{\theta'}$ are sufficiently close (i.e., $\zeta \to 0$), then it can be verified that for $n = 1$
\begin{equation}
\chi^2(P_\theta K \| P_{\theta'} K) = \chi^2 u^T I_{2}(\theta) u + o(\zeta^2).
\end{equation}
and
\begin{equation}
\chi^2(P_\theta \| P_{\theta'}) = \chi^2 u^T I_{2}(\theta) u + o(\zeta^2).
\end{equation}

These identities, together with Theorem 1, imply the desired upper bound on $I_{2}(\theta)$. The proof for $n > 1$ relies on the tensorization property of the contraction coefficient discussed in [16, Appendix A]. Next, we present a locally private version of the van Trees inequality.

**Corollary 1 (Private van Trees Inequality):** For any $\varepsilon > 0$ and $\Theta = [-B, B]^d$, we have

\begin{equation}
R^*(n, \Theta, \ell, \varepsilon) \geq \frac{d^2}{2n \gamma^2} \sum_{\theta \in \Theta} \text{Tr}(I_{2}(\theta)) + \varepsilon^2.
\end{equation}

**B. Private Le Cam’s Method: An Improved Version**

In this section, we propose a private version of the Le Cam’s method [46], which improves the existing one in the literature proved by Duchi et al. [33, 35]. Their result states that for two families of distributions $P_{\Theta_1} = \{P_{\theta}, \theta \in \Theta_1\}$ and $P_{\Theta_2} = \{P_{\theta}, \theta \in \Theta_2\}$, with $\Theta_1, \Theta_2 \subseteq \Theta$, such that $\min_{\theta \in \Theta_1, \theta \in \Theta_2} \ell(T(\theta_1), T(\theta_2)) \geq \alpha$, we have [35, Proposition 1]

\begin{equation}
R^*(n, \Theta, \ell, \varepsilon) \geq \frac{\alpha}{2} \frac{2^{\sqrt{\gamma^2 2}}}{\sqrt{\gamma^2}} \min \left\{ \sqrt{\gamma^2 D_{KL}(P_1 \| P_2)}, \sqrt{\gamma^2 TV(P_1, P_2)} \right\}.
\end{equation}

for any $P_1 \in P_{\Theta_1}$ and $P_2 \in P_{\Theta_2}$. Applying Theorem 1 and Theorem 2, we can obtain a strictly tighter lower bound on $R^*(n, \Theta, \ell, \varepsilon)$.

**Theorem 3 (Improved Private Le Cam’s Method):** Let $P_{\Theta_1} = \{P_{\theta}, \theta \in \Theta_1\}$ and $P_{\Theta_2} = \{P_{\theta}, \theta \in \Theta_2\}$, with $\Theta_1, \Theta_2 \subseteq \Theta$ such that $\min_{\theta \in \Theta_1, \theta \in \Theta_2} \ell(T(\theta_1), T(\theta_2)) \geq \alpha$. Then, we have for any $P_1 \in P_{\Theta_1}$ and $P_2 \in P_{\Theta_2}$

\begin{equation}
R^*(n, \Theta, \ell, \varepsilon) \geq \frac{\alpha}{2} \frac{2^{\sqrt{\gamma^2 2}}}{\sqrt{\gamma^2}} \min \left\{ \sqrt{\gamma^2 D_{KL}(P_1 \| P_2)}, \sqrt{\gamma^2 TV(P_1, P_2)} \right\}.
\end{equation}

Notice that since $\Psi_{\varepsilon} < (\varepsilon^2 - 1)^2$ for any $\varepsilon > 0$, this theorem yields a strictly better lower bound than (25). In particular, it improves the dependency on $\varepsilon$ from $\varepsilon^2$ to $\varepsilon^4$ for $\varepsilon > 1$.

As an example of Theorem 3, we next consider the locally private entropy estimation problem.

**Entropy Estimation under LDP:** Consider the following setting: Given a parameter $\theta \in \Theta = [0, 1]^{k-1}$ satisfying $\sum_{i=1}^{k-1} \theta_i \leq 1$, we define the parametric distribution by $P_{\theta} = (\theta_1, \ldots, \theta_{k-1}, \theta_k)$, where $\theta_k = 1 - \sum_{i=1}^{k-1} \theta_i$. Thus, $P_{\theta} \in P([k])$. We are interested in the estimation of $P_{\theta}$, i.e., $H(P_{\theta}) = -\sum_{i=1}^{k} \theta_i \log \theta_i$. We design the following hypothesis testing problem: let $P_1 = \left[\frac{\varepsilon}{k}, \frac{1-\varepsilon}{k}, \ldots, \frac{1-\varepsilon}{k} \right]$ and $P_2 = \left[\frac{\varepsilon}{k}, \ldots, \frac{1\varepsilon}{k}, \frac{1-\varepsilon}{k}, \ldots, \frac{1-\varepsilon}{k} \right]$ for some $\eta \in [0, 2]$ and $P_{\theta} \in [0, 1]^{k-1}$, and $P_2 = \left[\frac{\varepsilon}{k}, \ldots, \frac{1\varepsilon}{k}, \frac{1-\varepsilon}{k}, \ldots, \frac{1-\varepsilon}{k} \right]$. It can be verified that $(H(P_2) - H(P_1))^2 \geq \frac{1}{2} \eta^2 \log^2 (k - 1)$ and $D_{KL}(P_1 \| P_2) \leq \chi^2(P_1 \| P_2) \leq 2\eta^2$. Setting $\eta = \min\{1, \frac{1}{\log k + 1}\}$ and applying Theorem 3, we arrive at the following lower bound, which improves the non-private lower bound by $1/\gamma^2$, that is, at least by a constant even when $\varepsilon$ grows large.

**Corollary 2:** For the entropy estimation problem under LDP described above, we have for $k \geq 3$ and $\varepsilon \geq 0$

\begin{equation}
R^*(n, [0, 1]^{k-1}, \ell_2, \varepsilon) \geq \frac{1}{20} \frac{1}{n \varepsilon^2} \left( \frac{1}{\log k + 1} \right) \log^2 (k - 1).
\end{equation}

It is worth pointing out that Butucea and Issartel [23] have recently studied estimating Rényi entropy of order $\lambda$ for any $\lambda \in (0, 1) \cup (1, \infty)$ under LDP constraint. Specifically, they have established the minimax optimal rate $\Theta(1)$ for $\lambda \geq 2$. However, they fell short of providing optimal rate for estimating entropy (i.e., the case where $\lambda \rightarrow 1$).

**C. Private Assouad’s Method: An Improved Version**

Although the Le Cam’s method can provide sharp minimax rates for various problems, it is known to be constrained to applications that are reduced to binary hypothesis testing. In this section, we provide a private version of the Assouad’s method that is stronger than the existing one in [33, 35]. Let $\{P_\theta\}_{\theta \in \Theta}$ be a set of distributions indexed by $\mathcal{E}_k = \{\{1\}^k$ satisfying

\begin{equation}
\ell(T(\theta_1), T(\theta_2)) \geq 2\tau \sum_{j=1}^{k} \|u_j \neq v_j\|, \forall u, v \in \mathcal{E}_k.
\end{equation}

For each coordinate $i \in [k]$, we define the mixture of distributions obtained by averaging over distributions with a fixed value for the $j$-th position:

\begin{equation}
P_{\theta}^{+j} = \frac{1}{2^{k-1}} \sum_{\nu_j = 1}^{2^{k-1}} P_{\nu}^{+j} \text{ and } P_{\theta}^{-j} = \frac{1}{2^{k-1}} \sum_{\nu_j = -1}^{2^{k-1}} P_{\nu}^{-j},
\end{equation}

where $P_{\theta}^{+j}$ is the product distribution corresponding to $P_{\theta}$ when $\theta = \theta_1$ for $v \in \mathcal{E}_k$. The non-private Assouad’s method [59] yields

\begin{equation}
R^*(n, \Theta, \ell, \varepsilon) \geq \frac{1}{2} \sum_{j=1}^{k} \left( 1 - TV(P_{\theta}^{+j}, P_{\theta}^{-j}) \right).
\end{equation}

By applying Pinsker’s inequality and (1), Duchi et al. [35] extended this result to obtain a lower bound on the private minimax risk. Similarly, we apply Pinsker’s inequality and Theorem 2 to derive another bound for the private minimax risk which has a stronger dependence on $\varepsilon$.

**Theorem 4 (Improved Private Assouad’s Method):** Let the loss function $\ell$ satisfy (26), and define $P_{+j} = \frac{1}{2^{k-1}} \sum_{\nu_j = 1}^{2^{k-1}} P_{\nu}$ and $P_{-j} = \frac{1}{2^{k-1}} \sum_{\nu_j = -1}^{2^{k-1}} P_{\nu}$. Then, we have

\begin{equation}
R^*(n, \Theta, \ell, \varepsilon) \geq \frac{1}{2^k} \left[ 1 - \left( \frac{2n \Psi_{\varepsilon}}{k} \sum_{j=1}^{k} TV^2(P_{\theta}^{+j}, P_{\theta}^{-j}) \right)^2 \right].
\end{equation}

We apply this theorem to characterize lower bounds on the private minimax risk in the following two problems.

**Private Distribution Estimation:** Let $\Theta = \Delta_d = \{\theta \in [0, 1]^d : \sum_{i=1}^{d} \theta_i = 1\}$ and each $X_i$ is distributed according to the multinomial distribution with parameter $\theta$ on $\mathcal{X} = [d]$. We assume that the loss function is the $\ell_2$-norm for some $h \geq 1$, 

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
i.e., \( \ell(\hat{\theta}, \hat{\theta}) = \|\theta - \hat{\theta}\|_h \). The private minimax risk for this problem has been extensively studied for \( h = 1 \) and \( h = 2 \), see, e.g., [1], [6], [7], [17], [33], [43], [58]. The following corollary, built on Theorem 4, gives a lower bound on the private minimax risk for all \( h \geq 1 \).

Corollary 3: For any \( h \geq 1 \) and \( \varepsilon \geq 0 \), we have

\[
R^\alpha(n, \Delta_d, \| \cdot \|_h, \varepsilon) \geq \min \left\{ \frac{\sqrt{2h}}{h + 1} \left[ \frac{1}{2h + 2} \right]^{\frac{1}{2}} \frac{d^{1/h}}{\sqrt{n}\Psi_\varepsilon}, \frac{\sqrt{2h}}{h + 1} \left[ \frac{1}{\sqrt{2h}} \right]^{\frac{1}{2}} \frac{1}{\sqrt{n}\Psi_\varepsilon} \right\}.
\]

This lower bound matches (up to constant factors) with the upper bounds in [1], [7], [19], [33], [58] for both \( h = 1 \) and \( h = 2 \), and thus is order optimal in these cases. Furthermore, compared to the best existing lower bound [58], it improves the constants and applies to both non-interactive and sequentially interactive cases. We remark that a lower bound was recently derived by Acharya et al. [2, Th. 5] for general \( h \geq 1 \) which establishes the same order result as Corollary 3. While both results have the same order analysis, our approach is more amenable to deriving constants.

To further assess the quality of the lower bound in Corollary 3, we obtain an upper bound on \( R^\alpha(n, \Delta_d, \| \cdot \|_h, \varepsilon) \) by generalizing the Hadamard response [7] to \( \ell_h \)-norm with \( h \geq 2 \) in the following theorem. Under some mild conditions, the upper bound coincides with the second term in Corollary 3, with respect to the dependency on \( d \) and \( n \).

Theorem 5: For any \( 2 \leq h \leq 100 \) and \( \varepsilon \geq 0 \), when \( n \geq \min \left( \frac{d^{1/2}}{\varepsilon}, (\varepsilon^2)^{\frac{1}{2}} \right) \), we have

\[
R^\alpha(n, \Delta_d, \| \cdot \|_h, \varepsilon) \geq \left( \frac{\beta}{\beta + d} \right)^{\frac{1}{2}} \frac{d^{1/2}}{(e^\varepsilon - 1)^{1/2}}.
\]

Private Non-Parametric Density Estimation: Suppose \( X^n \) is a sequence of i.i.d. samples from a probability distribution on \([0, 1]\) that has density \( f \) with respect to the Lebesgue measure. Assume that \( f \) is Hölder continuous with smoothness parameter \( \beta \in (0, 1] \) and constant \( L \), i.e.,

\[
|f(x) - f(y)| \leq L|x - y|^{\beta}, \quad \forall x, y \in [0, 1].
\]

Let \( \mathcal{H}_\beta \) be the set of all such densities. We are interested in characterizing the private minimax risk in the sequentially interactive setting denoted by

\[
R^\alpha(n, \mathcal{H}_\beta, \| \cdot \|_h, \varepsilon) := \inf_{K_1, \ldots, K_n \in \mathcal{Q}_\varepsilon} \sup_{f \in \mathcal{H}_\beta} \mathbb{E}\left[ \|f - \hat{f}\|_h^h \right],
\]

where the expectation is taken with respect to the density \( f \in \mathcal{H}_\beta \) and also the mechanisms \( K_1, \ldots, K_n \in \mathcal{Q}_\varepsilon \). The non-private minimax rate for this problem for \( h = 2 \) is known to be \( \Theta(n^{-2/3+}) \), see e.g., [18, Th. 4] for a more recent proof. Butucea et al. [22] established a lower bound on \( R^\alpha(n, \mathcal{H}_\beta, \| \cdot \|_h, \varepsilon) \) in the high privacy regime. In particular, it was shown [22, Proposition 2.1] that

\[
R^\alpha(n, \mathcal{H}_\beta, \| \cdot \|_h, \varepsilon) \geq (n\varepsilon^2)^{-\frac{h}{2h+2}}.
\]

This result indicates that the effect of local privacy for small \( \varepsilon \) concerns both the reduction of the effective sample size from \( n \) to \( n\varepsilon^2 \) and also change of the exponent of the convergence rate from \( -2h \) to \( -\frac{h}{2h+2} \). In the following corollary, we show that the same observation holds for all privacy regime by extending (27) to all \( \varepsilon \geq 0 \). More precisely, the privacy constraint causes the effective sample size to reduce from \( n \) to \( n\Psi_\varepsilon \) and also the convergence rate to reduce to \( -\frac{h}{2h+2} \) as before.

Corollary 4: We have for \( h \geq 1 \) and \( \varepsilon \geq 0 \)

\[
R^\alpha(n, \mathcal{H}_\beta, \| \cdot \|_h, \varepsilon) \geq (n\varepsilon^2)^{-\frac{h}{2h+2}}.
\]

This corollary is proved by incorporating Theorem 4 into the classical framework that reduces the density estimation to a parameter estimation over a hypercube of a suitable dimension. Note that \( \Psi_\varepsilon \approx \varepsilon^2 \) for \( \varepsilon \leq 1 \), thus Corollary 4 recovers Butucea et al.’s result shown in (27).

D. Locally Private Mutual Information Method

Mutual information method has recently been proposed in [56, Sec. 12] as a systemic tool for obtaining lower bounds for non-private minimax risks with better constants than what would be obtained by Le Cam’s and Assouad’s methods. Let, for simplicity, \( T \) be the identity function, i.e., \( T(\theta) = \theta \). Moreover, suppose \( \theta \) is distributed according to a prior \( \pi \in \mathcal{P}(\Theta) \) and the loss function is the \( h \)-th power of an arbitrary norm over \( \mathbb{R}^d \). Define the Bayesian private risk as

\[
R^\alpha(\pi, \Theta, \| \cdot \|_h, \varepsilon) := \inf_{K_1, \ldots, K_n \in \mathcal{Q}_\varepsilon} \inf_{\psi} \mathbb{E}_\pi \left[ \|\psi(Z^n) - \theta\|_h^h \right].
\]

Notice that \( R^\alpha(\pi, \Theta, \| \cdot \|_h, \varepsilon) \geq R^\alpha_\beta(n, \Theta, \| \cdot \|_h, \varepsilon) \) for any prior \( \pi \). In the sequel, we expound an approach to lower bound \( R^\alpha_\beta(n, \Theta, \| \cdot \|_h, \varepsilon) \), which in turn yields a lower bound on \( R^\alpha(\pi, \Theta, \| \cdot \|_h, \varepsilon) \). Fix \( n \) mechanisms \( K_1, \ldots, K_n \in \mathcal{Q}_\varepsilon \) that sequentially generate \( Z^n \) and let \( \hat{\theta} = \psi(Z^n) \) be an estimate of \( \theta \) with the corresponding risk \( \mathbb{E}_\pi[\|\psi(\hat{\theta})\|_h^h] \leq D \) for some \( D \geq 0 \). (We shall replace \( D \) with \( R^\alpha_\beta(n, \Theta, \| \cdot \|_h, \varepsilon) \) later.) We can clearly write

\[
I(\theta; \hat{\theta}) \geq \inf_{F_\psi} \mathbb{E}_\pi \left[ \|\theta - \hat{\theta}\|_h^h \right] \leq D \Rightarrow \text{RDF}(\pi, D).
\]

Notice that the lower-bound is the definition of the rate-distortion function (RDF) evaluated at the distortion function (RDF) evaluated at the distortion measure, where the distortion measure is given by \( \| \cdot \|_h^h \). On the other hand, the Markov chain \( \theta - Z^n - \hat{\theta} \) and the data processing inequality imply \( I(\theta; \hat{\theta}) \leq I(\theta; Z^n) \). Therefore, we have

\[
\text{RDF}(\pi, D) \leq I(\theta; Z^n). \tag{28}
\]

Combining (4) with the tensorization property of \( \eta_{KL} \), we can show that

\[
I(\theta; Z^n) \leq I(\theta; X^n) \max_{\pi \in \mathcal{P}(\Theta)} \eta_{KL}(K_i).
\]

see [16, Appendix A] for details. Therefore, in light of Theorem 1 we have

\[
\text{RDF}(\pi, D) \leq \gamma, I(\theta; X^n). \tag{29}
\]
If we could somehow analytically compute RDF(π, D) for a prior π, then (29) would enable us to forge a relationship between D and I(θ, X^n). This relationship is desirable as we can simply replace D with R^n_θ(n, θ, \| \cdot \|, ε) in this greatly simplified form. However, computing rate-distortion function is known to be notoriously difficult even for simple distortion measures. Nevertheless, we can invoke the Shannon Lower Bound (see, e.g., [57] or [30, Problem 10.6]) to find an asymptotically tight lower bound on RDF(π, D). This in turn leads to the following lower bound on R^n_θ(n, θ, \| \cdot \|, ε).

**Theorem 6 (Locally Private Mutual Information Method):** Let \( \theta \sim \pi \) for some \( \pi \in \mathcal{P}(\Theta) \) and \( X^n \overset{\text{iid}}{\sim} P_\theta \). For an arbitrary norm \( \| \cdot \| \), we have

\[
R^n_\theta(n, \Theta, \| \cdot \|, \varepsilon) \geq \frac{d}{re[V_d \Gamma(1 + d/r)]^2} e^{H(\theta) - I(\theta; X^n)},
\]

where \( V_d \) is the volume of the unit \( \| \cdot \| \)-ball, \( \Gamma(\cdot) \) is the Gamma function, and \( H(\cdot) \) is the entropy of \( \pi \).

To obtain the best lower bound for \( R^n_\theta(n, \Theta, \| \cdot \|, \varepsilon) \) from Theorem 6, we need to pick a prior \( \pi \) that maximizes \( I(\theta; X^n) \). This prior need not necessarily be supported on entire \( \Theta \). An example of such prior selection is given for the Gaussian location model described next.

**Private Gaussian Location Model:** Suppose \( P_\theta = \mathcal{N}(\theta, \sigma^2 I_d) \) for some \( \sigma > 0 \), where \( \Theta \subseteq \Theta \). Characterizing the minimax risk for estimating \( \theta \) under LDP has been extensively studied for particular choices of loss function and \( \Theta \). For instance, \( \| \cdot \| = \| \cdot \|_2 \) and \( \Theta = \text{unit} \ell_\infty\text{-ball} \) were adopted in [17], [32], [33], [34], \( \| \cdot \| = \| \cdot \|_2 \) and \( \Theta = \text{unit} \ell_2\text{-ball} \) in [20] and \( \| \cdot \| = \| \cdot \|_2 \) for some \( h > 1 \) and \( \Theta = \text{unit} \ell_\infty\text{-ball} \) in [2]. Theorem 6 allows us to construct lower bounds on \( R^n_\theta(n, \Theta, \| \cdot \|, \varepsilon) \) for arbitrary loss and arbitrary \( \Theta \). For any such arbitrary subset \( \Theta \subseteq \mathbb{R}^d \), we define \( \text{rad}(\Theta) := \inf_{x \in \mathbb{R}^d} \sup_{y \in \Theta} \| x - y \|_2 \).

**Corollary 5 (Private Gaussian Location Model):** Let \( P_\theta = \mathcal{N}(\theta, \sigma^2 I_d) \) with \( \sigma > 0 \) and \( \Theta \subseteq \Theta \). Moreover, let \( \| \cdot \| \) be an arbitrary norm over \( \mathbb{R}^d \) and \( \Theta \) be an arbitrary subset of \( \mathbb{R}^d \) with a non-empty interior. Then, we have

\[
R^n_\theta(n, \Theta, \| \cdot \|, \varepsilon) \geq \frac{d^{1-r/2} \text{rad}(\Theta)^r}{r e^{2[V_d \Gamma(1 + d/r)]^2} [V(\Theta)]^{r/d} [V_2(\Theta)]^{r/d}} \min \left\{ \frac{\sigma^2 d^{r/2}}{n \text{rad}(\Theta)^r} \right\},
\]

where \( V(\Theta) \) is the volume of \( \Theta \) and \( V_2(\Theta) \) is the volume of the \( \ell_2\)-ball of radius \( \text{rad}(\Theta) \).

Instantiating this corollary, we may recover or generalize some existing lower bounds for Gaussian location models. For instance, for \( \| \cdot \| = \| \cdot \|_2 \), \( r = 2 \), and \( \Theta = \text{unit} \ell_2\text{-ball} \), we have \( V_2(\Theta) = 1/\sqrt{d} \), \( V(\Theta) = V_2(\Theta) = 1 \), and \( \Gamma(1 + d/2)^{1/2} \). It then follows from Corollary 5 that \( R^n_\theta(n, \Theta, \| \cdot \|_2, \varepsilon) \geq \min \left\{ \frac{\sigma^2 d^{r/2}}{n \text{rad}(\Theta)^r} \right\} \) which is optimal for \( \varepsilon \leq 1 \), as it matches the upper bounds in [20]. Also, for \( \| \cdot \| = \| \cdot \|_h \) with \( h \geq 1 \), \( r = 1 \), and \( \Theta = \text{unit} \ell_\infty\text{-ball} \), we have \( V_2(\Theta) = \text{rad}(\Theta) = \sqrt{d} \) and \( \Gamma(1 + d/2)^{1/2} \). Then, it follows from Corollary 5 that \( R^n_\theta(n, \Theta, \| \cdot \|_h, \varepsilon) \geq \min \left\{ \frac{\sigma^2 d^{r/2}}{n \text{rad}(\Theta)^r} \right\} \), which generalizes [2, Th. 4] from \( \varepsilon \leq 1 \) to all \( \varepsilon \geq 0 \).

**E. Binary Hypothesis Testing Under LDP**

Consider the following typical setting of binary hypothesis testing: Given \( n \) i.i.d. samples \( X^n \) and two distributions \( P \) and \( Q \), we seek to determine which distributions generated \( X^n \). That is, we wish to test the null hypothesis \( H_0 = P \) against the alternative hypothesis \( H_1 = Q \). To address the privacy concern, we take sequentially interactive mechanisms \( K_1, \ldots, K_n \) that generate \( Z^n \). The goal is now to perform the above test given \( Z^n \). Let \( \phi : Z^n \rightarrow \{0, 1\} \) be a test that accepts the null hypothesis if it is equal to zero. For any such test \( \phi \), define \( A_n(\phi) = \{ z^n \in Z^n : \phi(z^n) = 1 \} \). There are two error probabilities associated with \( \phi \), namely, \( P(A_n(\phi)) \) and \( 1 - Q(A_n(\phi)) \). We say that this test privately distinguishes \( P \) from \( Q \) with sample complexity \( n^*(\phi) \) if both \( P(A_n(\phi)) \) and \( 1 - Q(A_n(\phi)) \) are smaller than 1/10 for every \( n \geq n^*(\phi) \). We then define the sample complexity of privately distinguishing \( P \) from \( Q \) as

\[
SC^n_{P, Q} := \inf \left\{ k \in \mathbb{N} : \inf_{\phi : Z^n \rightarrow \{0, 1\}} n^*(\phi) \geq k \right\}.
\]

The characterization of sample complexity of hypothesis testing is well-understood in the non-private setting: The number of samples needed to distinguish \( P \) from \( Q \) is \( \Theta(1/H^2(P, Q)) \).

Under local privacy, it has been shown in [33] that \( SC^n_{P, Q} = \Theta(1/\varepsilon^2 TV^2(P, Q)) \) for sufficiently small \( \varepsilon \). In the following lemma, we extend this result to any \( \varepsilon \geq 0 \).

**Lemma 2:** Given \( \varepsilon \geq 0 \) and two distributions \( P \) and \( Q \), we have

\[
SC^n_{P, Q} \geq \frac{4}{35} \max \left\{ \frac{1}{\gamma}, \frac{1}{2\gamma}, \frac{1}{TV^2(P, Q)} \right\},
\]

and

\[
SC^n_{P, Q} \leq \frac{2 \log(5)}{\gamma^2 TV^2(P, Q)}.
\]

Our lower bound reveals an interesting phase transition: the sample complexity of the binary hypothesis testing appears to be dependent on the Hellinger distance instead of the total variation distance as \( \varepsilon^2 \geq \Omega\left(\frac{H^2(P, Q)}{TV^2(P, Q)}\right) \). Furthermore, when \( \varepsilon^2 \) is large, our result has made a constant-factor (1/\( \varepsilon^2 \)) improvement compared to the non-private lower bound. Recently, Pensia et al. [49] formalized these observations and demonstrated that the lower bound in Lemma 2 is in fact optimal (up to a constant factor) for any \( \varepsilon \geq 0 \) if \( P \) and \( Q \) are binary, that is,

\[
SC^n_{P, Q} \approx \begin{cases} \frac{\epsilon^2 TV^2(P, Q)}{H^2(P, Q)} & \text{if } \epsilon \in (0, 1), \\ \frac{1}{TV^2(P, Q)} & \text{if } \epsilon \in [\varepsilon^2, H^2(P, Q)/TV^2(P, Q)], \\ \frac{1}{H^2(P, Q)} & \text{if } \epsilon^2 > H^2(P, Q)/TV^2(P, Q). \end{cases}
\]

1This statement is folklore, but see, e.g., [24] for a simple proof.

2It can be shown that Lemma 2 holds in the most general setting, i.e., in the fully interactive setting (aka the so-called blackboard model); see [32], [42] for the formal definition.
A. Pensia, A. R. Asadi, V. Jog, and P.-L. Loh, “Simple binary hypothesis testing under local differential privacy and communication constraints,” 2023, arXiv:2301.03566.

M. Raginsky, “Strong data processing inequalities and φ-Sobolev inequalities for discrete channels,” IEEE Trans. Inf. Theory, vol. 62, no. 6, pp. 3355–3389, Jun. 2016.

A. Rohde and L. Steinberger, “Geometrizing rates of convergence under local differential privacy constraints,” Ann. Stat., vol. 48, no. 5, pp. 2646–2670, Oct. 2020.

I. Sason and S. Verdú, “f-divergence inequalities,” IEEE Trans. Inf. Theory, vol. 62, no. 11, pp. 5973–6006, Nov. 2016.

A. Shah, W.-N. Chen, J. Balle, P. Kairouz, and L. Theis, “Optimal compression of locally differentially private mechanisms,” 2021, arXiv:2111.00092.

Ú. Erlingsson, V. Feldman, I. Mironov, A. Raghunathan, K. Talwar, and A. Thakurta, “Amplification by shuffling: From local to central differential privacy via anonymity,” in Proc. ACM-SIAM Symp. Discr. Algorithms (SODA), 2019, pp. 2468–2479.

S. L. Warner, “Randomized response: A survey technique for eliminating evasive answer bias,” J. Amer. Stat. Assoc., vol. 60, no. 309, pp. 63–69, 1965.

Y. Wu, “Lecture notes for information-theoretic methods for high-dimensional statistics,” Lecture Notes ECE598YW2020, Univ. Illinois Urbana-Champaign, Champaign, IL, USA.

Y. Yamada, S. Tazaki, and R. Gray, “Asymptotic performance of block quantizers with difference distortion measures,” IEEE Trans. Inf. Theory, vol. 26, no. 1, pp. 6–14, Jan. 1980.

M. Ye and A. Barg, “Optimal schemes for discrete distribution estimation under locally differential privacy,” IEEE Trans. Inf. Theory, vol. 64, no. 8, pp. 5662–5676, Aug. 2018.

B. Yu, Assouad, Fano, and Le Cam. New York, NY, USA: Springer, 1997, pp. 423–435.

B. Zamanlooy and S. Asoodeh, “Strong data processing inequalities for locally differentially private mechanisms,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), 2023, pp. 1794–1799.