Local Differential Privacy Is Equivalent to Contraction of an f-Divergence

Shahab Asoodeh[†], Maryam Aliakbarpour[*], and Flavio P. Calmon[†]
[†]Harvard University, [*]University of Massachusetts Amherst

***Abstract*—We investigate the local differential privacy (LDP) guarantees of a randomized privacy mechanism via its contraction properties. We first show that LDP constraints can be equivalently cast in terms of the contraction coefficient of the E_γ-divergence. We then use this equivalent formula to express LDP guarantees of privacy mechanisms in terms of contraction coefficients of arbitrary f-divergences. When combined with standard estimation-theoretic tools (such as Le Cam's and Fano's converse methods), this result allows us to study the trade-off between privacy and utility in several testing and minimax and Bayesian estimation problems.**

I. Introduction

A major challenge in modern machine learning applications is balancing statistical efficiency with the privacy of individuals from whom data is obtained. In such applications, privacy is often quantified in terms of Differential Privacy (DP) [1]. DP has several variants, including approximate DP [2], Rényi DP [3], and others [4–7]. Arguably, the most stringent flavor of DP is *local differential privacy* (LDP) [8, 9]. Intuitively, a randomized mechanism (or a Markov kernel) is said to be locally differentially private if its output does not vary significantly with arbitrary perturbations of the input.

More precisely, a mechanism is said to be ε-LDP (or *pure* LDP) if the privacy loss random variable, defined as the log-likelihood ratio of the output for any two different inputs, is smaller than ε with probability one. One can also consider an *approximate* variant of this constraint: K is said to be (ε,δ)-LDP if the privacy loss random variable does not exceed ε with probability at least $1-\delta$ (see Def. 1 for the formal definition).

The study of statistical efficiency under LDP constraints has gained considerable traction, e.g., [8–19]. Almost all of these works consider ε-LDP and provide meaningful bounds only for sufficiently small values of ε (i.e., the high privacy regime). For instance, Duchi et al. [10] studied minimax estimation problems under ε-LDP constraints and showed that for $\varepsilon \le 1$, the price of privacy is to reduce the effective sample size from n to $\varepsilon^2 n$. A slightly improved version of this result appeared in [13, 20]. More recently, Duchi and Rogers [21] developed a framework based on the *strong* data processing inequality (SDPI) [22] and derived lower bounds for minimax estimation risk under ε-LDP that hold for any $\varepsilon \ge 0$.

In this work, we develop an SDPI-based framework for studying hypothesis testing and estimation problems under (ε,δ)-LDP, extending the results of [21] to approximate LDP. In particular, we derive bounds for both the minimax and Bayesian estimation risks that hold for any $\varepsilon \ge 0$ and $\delta \ge 0$. Interestingly, when setting $\delta = 0$, our bounds can be slightly stronger than [10].

Our main mathematical tool is an equivalent expression for DP in terms of E_γ-divergence. Given $\gamma \ge 1$, the E_γ-divergence between two distributions P and Q is defined as

$$\mathsf{E}_\gamma(P\|Q) \coloneqq \frac{1}{2}\int |\mathrm{d}P - \gamma \mathrm{d}Q| - \frac{1}{2}(\gamma-1). \tag{1}$$

We show that a mechanism K is (ε,δ)-LDP *if and only if*

$$\mathsf{E}_\gamma(P\mathsf{K}\|Q\mathsf{K}) \le \delta \mathsf{E}_\gamma(P\|Q)$$

for $\gamma = e^\varepsilon$ and any pairs of distributions (P,Q) where $P\mathsf{K}$ represents the output distribution of K when the input distribution is P. Thus, the approximate LDP guarantee of a mechanism can be fully characterized by its contraction under E_γ-divergence. When combined with standard statistical techniques, including Le Cam's and Fano's methods [23, 24], E_γ-contraction leads to general lower bounds for the minimax and Bayesian risk under (ε,δ)-LDP for any $\varepsilon \ge 0$ and $\delta \in [0,1]$. In particular, we show that the price of privacy in this case is to reduce the sample size from n to $n[1-e^{-\varepsilon}(1-\delta)]$.

There exists several results connecting pure LDP to the contraction properties of KL divergence D_{KL} and total variation distance TV. For instance, for any ε-LDP mechanism K, it is shown in [10, Theorem 1] that $D_{\mathsf{KL}}(P\mathsf{K}\|Q\mathsf{K}) \le 2(e^\varepsilon-1)^2\mathsf{TV}^2(P,Q)$ and in [13, Theorem 6] that $\mathsf{TV}(P\mathsf{K}\|Q\mathsf{K}) \le \frac{e^\varepsilon-1}{e^\varepsilon+1}\mathsf{TV}(P,Q)$ for any pairs (P,Q). Inspired by these results, we further show that if K is (ε,δ)-LDP then $D_f(P\mathsf{K}\|Q\mathsf{K}) \le [1-e^{-\varepsilon}(1-\delta)]D_f(P\|Q)$ for any *arbitrary* f-divergences D_f and any pairs (P,Q).

Notation. For a random variable X, we write P_X and $\mathcal{X}$ for its distribution (i.e., $X \sim P_X$) and its alphabet, respectively. For any set A, we denote by $\mathcal{P}(A)$ the set of all probability distributions on A. Given two sets $\mathcal{X}$ and $\mathcal{Z}$, a Markov kernel (i.e., channel) K is a mapping from $\mathcal{X}$ to $\mathcal{P}(\mathcal{Z})$ given by $x \mapsto \mathsf{K}(\cdot|x)$. Given $P \in \mathcal{P}(\mathcal{X})$ and a Markov kernel $\mathsf{K}: \mathcal{X} \to \mathcal{P}(\mathcal{Z})$, we let $P\mathsf{K}$ denote the output distribution of K when the input distribution is P, i.e., $P\mathsf{K}(\cdot) = \int \mathsf{K}(\cdot|x)P(\mathrm{d}x)$. Also, we use $\mathsf{BSC}(\omega)$ to denote the binary symmetric channel with crossover probability ω. For sequences $\{a_n\}$ and $\{b_n\}$, we use $a_n \gtrsim b_n$ to indicate $a_n \ge Cb_n$ for some universal constant C.

This work was supported in part by NSF under grants CIF 1900750, CIF CAREER 1845852, and CCF 1934846.

2021 IEEE International Symposium on Information Theory (ISIT) | 978-1-5386-8209-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISIT45174.2021.9517999

II. Preliminaries

A. f-Divergences

Given a convex function $f:(0,\infty)\to\mathbb{R}$ such that $f(1)=0$, the f-divergence between two probability measures $P\ll Q$ is defined as [25, 26]

$$D_f(P\|Q) \coloneqq \mathbb{E}_Q\Big[f\big(\frac{\mathrm{d}P}{\mathrm{d}Q}\big)\Big]. \tag{2}$$

Due to convexity of f, we have $D_f(P\|Q)\geq f(1)=0$. If, furthermore, f is strictly convex at 1, then equality holds if and only $P=Q$. Popular examples of f-divergences include $f(t)=t\log t$ corresponding to KL divergence, $f(t)=|t-1|$ corresponding to total variation distance, and $f(t)=t^2-1$ corresponding to χ^2-divergence. In this paper, we mostly concern with an important sub-family of f-divergences associated with $f_\gamma(t)=\max\{t-\gamma,0\}$ for a parameter $\gamma\geq 1$. The corresponding f-divergence, denoted by $\mathsf{E}_\gamma(P\|Q)$, is called E_γ-divergence (or sometimes *hockey-stick divergence* [27]) and is explicitly defined in (1). It appeared in [28] for proving converse channel coding results and also used in [7, 29–31] for characterizing privacy guarantees of iterative algorithms in terms of other variants of DP.

B. Contraction Coefficient

All f-divergences satisfy data processing inequality, i.e., $D_f(P\mathsf{K}\|Q\mathsf{K})\leq D_f(P\|Q)$ for any pair of probability distributions (P,Q) and Markov kernel K [25]. However, in many cases, this inequality is strict. The *contraction coefficient* of Markov kernel K under D_f-divergence $\eta_f(\mathsf{K})$ is the smallest number $\eta\leq 1$ such that $D_f(P\mathsf{K}\|Q\mathsf{K})\leq \eta D_f(P\|Q)$ for any pair of probability distributions (P,Q). Formally, $\eta_f(\mathsf{K})$ is defined as

$$\eta_f(\mathsf{K}) \coloneqq \sup_{\substack{P,Q\in\mathcal{P}(\mathcal{X}):\\ D_f(P\|Q)\neq 0}} \frac{D_f(P\mathsf{K}\|Q\mathsf{K})}{D_f(P\|Q)}. \tag{3}$$

Contraction coefficients have been studied for several f-divergences, e.g., η_{TV} for total variation distance was studied in [32–34], η_{KL} for KL-divergence was studied in [35–40], and η_{χ^2} for χ^2-divergence was studied in [34, 40, 41]. In particular, Dobrushin [32] showed that η_{TV} has a remarkably simple two-point characterization $\eta_{\mathsf{TV}}(\mathsf{K})=\sup_{x_1,x_2\in\mathcal{X}}\mathsf{TV}(\mathsf{K}(\cdot|x_1),\mathsf{K}(\cdot|x_2))$.

Similarly, one can plug E_γ-divergence into (3) and define the contraction coefficient $\eta_\gamma(\mathsf{K})$ for a Markov kernel K under E_γ-divergence. This contraction coefficient has recently been studied in [31] for deriving approximate DP guarantees for online algorithms. In particular, it was shown [31, Theorem 3] that η_γ enjoys a simple two-point characterization, i.e., $\eta_\gamma(\mathsf{K})=\sup_{x_1,x_2\in\mathcal{X}}\mathsf{E}_\gamma(\mathsf{K}(\cdot|x_1)\|\mathsf{K}(\cdot|x_2))$. Since $\mathsf{E}_1(P\|Q)=\mathsf{TV}(P,Q)$, this is a natural extension of Dobrushin's result.

C. Local Differential Privacy

Suppose K is a randomized mechanism mapping each $x\in\mathcal{X}$ to a distribution $\mathsf{K}(\cdot|x)\in\mathcal{P}(\mathcal{Z})$. One could view K as a Markov kernel (i.e., channel) $\mathsf{K}:\mathcal{X}\to\mathcal{P}(Z)$.

Definition 1 ([8, 9]). *A mechanism $\mathsf{K}:\mathcal{X}\to\mathcal{P}(\mathcal{Z})$ is (ε,δ)-LDP for $\varepsilon\geq 0$ and $\delta\in[0,1]$ if*

$$\sup_{x,x'\in\mathcal{X}}\sup_{A\subset\mathcal{Z}}\big[\mathsf{K}(A|x)-e^\varepsilon\mathsf{K}(A|x')\big]\leq\delta. \tag{4}$$

K is said to be ε-LDP if it is $(\varepsilon,0)$-LDP. Let $\mathcal{Q}_{\varepsilon,\delta}$ be the collection of all Markov kernels K with the above property. When $\delta=0$, we use $\mathcal{Q}_\varepsilon$ to denote $\mathcal{Q}_{\varepsilon,0}$.

Interactivity in Privacy-Preserving Mechanisms: Suppose there are n users, each in possession of a datapoint X_i, $i\in[n]\coloneqq\{1,\dots,n\}$. The users wish to apply a mechanism K_i that generates a privatized version of X_i, denoted by Z_i. We say that the collection of mechanisms $\{\mathsf{K}_i\}$ is *non-interactive* if K_i is entirely determined by X_i and independent of (X_j,Z_j) for $j\neq i$. When all users apply the same mechanism K, we can view $Z^n\coloneqq(Z_1,\dots,Z_n)$ as independent applications of K to each X_i. We denote this overall mechanism by $\mathsf{K}^{\otimes n}$. If interactions between users are permitted, then K_i need not depend only on X_i. In this case, we denote the overall mechanism $\{\mathsf{K}_i\}_{i=1}^n$ by K^n. In particular, the *sequentially interactive* [10] 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. LDP As the Contraction of E_γ-Divergence

We show next that the (ε,δ)-LDP constraint, with δ not necessarily equal to zero, is *equivalent* to the contraction of E_γ-divergence.

Theorem 1. *A mechanism K is (ε,δ)-LDP if and only if $\eta_{e^\varepsilon}(\mathsf{K})\leq\delta$ or equivalently*

$$\mathsf{K}\in\mathcal{Q}_{\varepsilon,\delta}\iff \mathsf{E}_{e^\varepsilon}(P\mathsf{K}\|Q\mathsf{K})\leq\delta\mathsf{E}_{e^\varepsilon}(P\|Q),\quad\forall P,Q.$$

The proof of this theorem (and all other results) are given in [42]. We note that Duchi et al. [10] showed that if K is ε-LDP then $D_{\mathsf{KL}}(P\mathsf{K}\|Q\mathsf{K})\leq 2(e^\varepsilon-1)^2\mathsf{TV}^2(P,Q)$. They then informally concluded from this result that ε-LDP acts as a contraction on the space of probability measures. Theorem 1 makes this observation precise.

According to Theorem 1, a mechanism K is ε-LDP if and only if $\mathsf{E}_{e^\varepsilon}(P\mathsf{K}\|Q\mathsf{K})=0$ for any distributions P and Q. An example of such Markov kernels is given next.

Example 1. (Randomized response mechanism) Let $\mathcal{X}=\mathcal{Z}=\{0,1\}$ and consider the mechanism given by the binary symmetric channel $\mathsf{BSC}(\omega_\varepsilon)$ with $\omega_\varepsilon\coloneqq\frac{1}{1+e^\varepsilon}$. This is often called randomized response mechanism [43] and denoted by $\mathsf{K}^\varepsilon_{\mathsf{RR}}$. This simple mechanism is well-known to be ε-LDP which can now be verified via Theorem 1. Let $P=\mathsf{Bernoulli}(p)$ and $Q=\mathsf{Bernoulli}(q)$ with $p,q\in[0,1]$. Then $P\mathsf{K}^\varepsilon_{\mathsf{RR}}=\mathsf{Bernoulli}(p*\omega_\varepsilon)$ and $P\mathsf{K}^\varepsilon_{\mathsf{RR}}=\mathsf{Bernoulli}(q*\omega_\varepsilon)$ where $a*b\coloneqq a(1-b)+b(1-a)$. It is straightforward to verify that $|p*\omega_\varepsilon-e^\varepsilon q*\omega_\varepsilon|+|1-p*\omega_\varepsilon-e^\varepsilon(1-q*\omega_\varepsilon)|=0.5(e^\varepsilon-1)$ for any p,q, implying $\mathsf{E}_{e^\varepsilon}(P\mathsf{K}^\varepsilon_{\mathsf{RR}}\|Q\mathsf{K}^\varepsilon_{\mathsf{RR}})=0$. When $|\mathcal{X}|=k\geq 2$, a simple generalization of this mechanism, called k-ary randomized response, has been reported in literature (see, e.g., [13, 20]) and is defined by $\mathcal{Z}=\mathcal{X}$ and $\mathsf{K}_{\mathsf{kRR}}(x|x)=\frac{e^\varepsilon}{k-1+e^\varepsilon}$

and $\mathsf{K}_{\mathsf{kRR}}(z|x) = \frac{1}{k-1+e^{\varepsilon}}$ for $z \neq x$. Again, it can be verified that for this mechanism we have $\mathsf{E}_{e^{\varepsilon}}(P\mathsf{K}^{\varepsilon}_{\mathsf{kRR}} \| Q\mathsf{K}^{\varepsilon}_{\mathsf{kRR}}) = 0$, for all P and Q in $\mathcal{P}(\mathcal{X})$.

E_{γ}-divergence underlies all other f-divergences, in a sense that any arbitrary f-divergence can be represented by E_{γ}-divergence [44, Corollary 3.7]. Thus, an LDP constraint implies that a Markov kernel contracts for *all* f-divergences, in a similar spirit to E_{γ}-contraction in Theorem 1.

Lemma 1. *Let $\mathsf{K} \in \mathcal{Q}_{\varepsilon,\delta}$ and $\varphi(\varepsilon,\delta) \coloneqq 1-(1-\delta)e^{-\varepsilon}$. Then, $\eta_f(\mathsf{K}) \leq \varphi(\varepsilon,\delta)$ or, equivalently,*

$$D_f(P\mathsf{K}\|Q\mathsf{K}) \leq D_f(P\|Q)\varphi(\varepsilon,\delta) \qquad \forall P, Q \in \mathcal{P}(\mathcal{X}).$$

Notice that this lemma holds for any f-divergences and any *general* family of (ε,δ)-LDP mechanisms. However, it can be improved if one considers particular mechanisms or a certain f-divergence. For instance, it is known that $\eta_{\mathsf{KL}}(\mathsf{BSC}(\omega)) = (1-2\omega^2)$ [22]. Thus, we have $\eta_{\mathsf{KL}}(\mathsf{K}^{\varepsilon}_{\mathsf{RR}}) = (\frac{e^{\varepsilon}-1}{e^{\varepsilon}+1})^2$ for the randomized response mechanism $\mathsf{K}^{\varepsilon}_{\mathsf{RR}}$ (cf. Example 1), while Lemma 1 implies that $\eta_{\mathsf{KL}}(\mathsf{K}^{\varepsilon}_{\mathsf{RR}}) \leq 1-e^{-\varepsilon}$. Unfortunately, η_{KL} is difficult to compute in closed form for general Markov kernels, in which case Lemma 1 provides a useful alternative.

Next, we extend Lemma 1 for the non-interactive mechanism. Fix an (ε,δ)-LDP mechanism K and consider the corresponding non-interactive mechanism $\mathsf{K}^{\otimes n}$. To obtain upper bounds on $\eta_f(\mathsf{K}^{\otimes n})$ directly through Lemma 1, we would first need to derive privacy parameters of $\mathsf{K}^{\otimes n}$ in terms of ε and δ (e.g., by applying composition theorems). Instead, we can use the tensorization properties of contraction coefficients (see, e.g., [39, 40]) to relate $\eta_f(\mathsf{K}^{\otimes n})$ to $\eta_f(\mathsf{K})$ and then apply Lemma 1, as described next.

Lemma 2. *Let $\mathsf{K} \in \mathcal{Q}_{\varepsilon,\delta}$ and $\varphi_n(\varepsilon,\delta) \coloneqq 1-e^{-n\varepsilon}(1-\delta)^n$. Then $\eta_f(\mathsf{K}^{\otimes n}) \leq \varphi_n(\varepsilon,\delta)$ for $n \geq 1$.*

Each of the next three sections provide a different application of the contraction characterization of LDP.

IV. Private Minimax Risk

Let $X^n = (X_1, \ldots, X_n)$ be n independent and identically distributed (i.i.d.) samples drawn from a distribution P in a family $\mathcal{P} \subseteq \mathcal{P}(\mathcal{X})$. Let also $\theta : \mathcal{P} \to \mathcal{T}$ be a parameter of a distribution that we wish to estimate. Each user has a sample X_i and applies a privacy-preserving mechanism K_i to obtain Z_i. Generally, we can assume that K_i are sequentially interactive. Given the sequences $\{Z_i\}_{i=1}^n$, the goal is to estimate $\theta(P)$ through an estimator $\Psi : \mathcal{Z}^n \to \mathcal{T}$. The quality of such estimator is assessed by a semi-metric $\ell : \mathcal{T}\times\mathcal{T} \to \mathbb{R}_+$ and is used to define the minimax risk as:

$$\mathcal{R}_n(\mathcal{P},\ell,\varepsilon,\delta) \coloneqq \inf_{\mathsf{K}^n \subset \mathcal{Q}_{\varepsilon,\delta}} \inf_{\Psi} \sup_{P\in\mathcal{P}} \mathbb{E}[\ell(\Psi(Z^n),\theta(P))]. \quad (5)$$

The quantity $R_n(\mathcal{P},\ell,\varepsilon,\delta)$ uniformly characterizes the optimal rate of private statistical estimation over the family $\mathcal{P}$ using the best possible estimator and privacy-preserving mechanisms in $\mathcal{Q}_{\varepsilon,\delta}$. In the absence of privacy constraints (i.e., $Z^n = X^n$), we denote the minimax risk by $\mathcal{R}_n(\mathcal{P},\ell)$.

The first step in deriving information-theoretic lower bounds for minimax risk is to reduce the above estimation problem to a testing problem [23, 24, 45]. To do so, we need to construct an index set $\mathcal{V}$ with $|\mathcal{V}| < \infty$ and a family of distributions $\{P_v, v \in \mathcal{V}\} \subseteq \mathcal{P}$ such that $\ell(\theta(P_v),\theta(P_{v'})) \geq 2\tau$ for all $v \neq v'$ in $\mathcal{V}$ for some $\tau > 0$. The canonical testing problem is then defined as follows: Nature chooses a random variable V uniformly at random from $\mathcal{V}$, and then conditioned on $V = v$, the samples X^n are drawn i.i.d. from P_v, denoted by $X^n \sim P_v^{\otimes n}$. Each X_i is then fed to a mechanism K_i to generate Z_i. It is well-known [23, 24, 45] that $\mathcal{R}_n(\mathcal{P},\ell) \geq \tau\mathsf{P}_{\mathsf{e}}(V|X^n)$, where $\mathsf{P}_{\mathsf{e}}(V|X^n)$ denotes the probability of error in guessing V given X^n. Replacing X^n by its (ε,δ)-privatized samples Z^n in this result, one can obtain a lower bound on $R_n(\mathcal{P},\ell,\varepsilon,\delta)$ in terms of $\mathsf{P}_{\mathsf{e}}(V|Z^n)$. Hence, the remaining challenge is to lower-bound $\mathsf{P}_{\mathsf{e}}(V|Z^n)$ over the choice of mechanisms $\{\mathsf{K}_i\}$. There are numerous techniques for this objective depending on $\mathcal{V}$. We focus on two such approaches, namely Le Cam's and Fano's method, that bound $\mathsf{P}_{\mathsf{e}}(V|Z^n)$ in terms of total variation distance and mutual information and hence allow us to invoke Lemmas 1 and 2.

A. Locally Private Le Cam's Method

Le Cam's method is applicable when V is a binary set and contains, say, P_0 and P_1. In its simplest form, it relies on the inequality (see [23, Lemma 1] or [24, Theorem 2.2]) $\mathsf{P}_{\mathsf{e}}(V|X^n) \geq \frac{1}{2}\left[1-\mathsf{TV}(P_0^{\otimes n}, P_1^{\otimes n})\right]$. Thus, it yields the following lower bound for non-private minimax risk

$$\mathcal{R}_n(\mathcal{P},\ell) \geq \frac{\tau}{2}\left[1-\mathsf{TV}(P_0^{\otimes n}, P_1^{\otimes n})\right] \quad (6)$$

$$\geq \frac{\tau}{2}\left[1-\frac{1}{\sqrt{2}}\sqrt{nD_{\mathsf{KL}}(P_0\|P_1)}\right], \quad (7)$$

for any $P_0 \neq P_1$ in $\mathcal{P}$, where the second inequality follows from Pinsker's inequality and chain rule of KL divergence. In the presence of privacy, the estimator Ψ depends on Z^n instead of X^n, which is generated by a sequentially interactive mechanism K^n. To write the private counterpart of (6), we need to replace $P_0^{\otimes n}$ and $P_1^{\otimes n}$ with $P_0^{\otimes n}\mathsf{K}^n$ and $P_1^{\otimes n}\mathsf{K}^n$ the corresponding marginals of Z^n, respectively. A lower bound for $\mathcal{R}_n(\mathcal{P},\ell,\varepsilon,\delta)$ is therefore obtained by deriving an upper bound for $\mathsf{TV}(P_0^{\otimes n}\mathsf{K}^n, P_1^{\otimes n}\mathsf{K}^n)$ for all $\mathsf{K}^n \subset \mathcal{Q}_{\varepsilon,\delta}$.

Lemma 3. *Let $P_0, P_1 \in \mathcal{P}$ satisfy $\ell(\theta(P_0),\theta(P_1)) \geq 2\tau$. Then we have*

$$\mathcal{R}_n(\mathcal{P},\ell,\varepsilon,\delta) \geq \frac{\tau}{2}\left[1-\frac{1}{\sqrt{2}}\sqrt{n\varphi(\varepsilon,\delta)D_{\mathsf{KL}}(P_0\|P_1)}\right].$$

By comparing with the original non-private Le Cam's method (7), we observe that the effect of (ε,δ)-LDP is to reduce the effective sample size from n to $(1-e^{-\varepsilon}(1-\delta))n$. Setting $\delta = 0$, this result strengthens Duchi et al. [10, Corollary 2], where the effective sample size was shown to be $4\varepsilon^2 n$ for sufficiently small ε.

Example 2. (One-dimensional mean estimation) For some $k >$

1, we assume $\mathcal{P}$ is given by

$$\mathcal{P} = \mathcal{P}_k := \{P \in \mathcal{P}(\mathcal{X}) : \ |\mathbb{E}_P[X]| \le 1, \mathbb{E}_P[|X|^k] \le 1\}.$$

The goal is to estimate $\theta(P) = \mathbb{E}_P[X]$ under $\ell = \ell_2^2$ the squared ℓ_2 metric. This problem was first studied in [10, Propsition 1] where it was shown $\mathcal{R}_n(\mathcal{P}_k, \ell_2^2, \varepsilon, 0) \ge (n\varepsilon^2)^{-(k-1)/k}$ *only* for $\varepsilon \le 1$. Applying our framework to this example, we obtain a similar lower bound that holds for all $\varepsilon \ge 0$ and $\delta \in [0, 1]$.

Corollary 1. *For all $k > 1$, $\varepsilon \ge 0$, and $\delta \in (0, 1)$, we have*

$$\mathcal{R}_n(\mathcal{P}_k, \ell_2^2, \varepsilon, \delta) \gtrsim \min\left\{1, \left[n\varphi^2(\varepsilon, \delta)\right]^{-\frac{(k-1)}{k}}\right\}. \tag{8}$$

It is worth instantiating this corollary for some special values of k. Consider first the usual setting of finite variance setting, i.e., $k = 2$. In the non-private case, it is known that the sample mean has mean-squared error that scales as $1/n$. According to Corollary 1, this rate worsens to $1/\varphi(\varepsilon, \delta)\sqrt{n}$ in the presence of (ε, δ)-LDP requirement. As $k \to \infty$, the moment condition $\mathbb{E}_p[|X|^k] \le 1$ implies the boundedness of X. In this case, Corollary 1 implies the more standard lower bound $(\varphi^2(\varepsilon, \delta)n)^{-1}$.

B. Locally Private Fano's Method

Le Cam's method involves a pair of distributions (P_0, P_1) in $\mathcal{P}$. However, it is possible to derive a stronger bound considering a larger subset of $\mathcal{P}$ by applying Fano's inequality (see, e.g., [23]). We follow this path to obtain a better minimax lower bound for the non-interactive setting.

Consider the index set $\mathcal{V} = \{1, \dots, |\mathcal{V}|\}$. The non-private Fano's method relies on the Fano's inequality to write a lower bound for $\mathsf{P_e}(V|X^n)$ in terms of mutual information as

$$\mathcal{R}_n(\mathcal{P}, \ell) \ge \tau\left[1 - \frac{I(X^n; V) + \log 2}{\log|\mathcal{V}|}\right]. \tag{9}$$

To incorporate privacy into this result, we need to derive an upper bound for $I(Z^n; V)$ over all choices of mechanisms $\{\mathsf{K}_i\}$. Focusing on the non-interactive mechanisms, the following lemma exploits Lemma 2 for such an upper bound.

Lemma 4. *Given X^n and V as described above, let Z^n be constructed by applying $\mathsf{K}^{\otimes n}$ on X^n. If K is (ε, δ)-LDP, then we have*

$$I(Z^n; V) \le \varphi_n(\varepsilon, \delta) I(X^n; V) \le \frac{n\varphi_n(\varepsilon, \delta)}{|\mathcal{V}|^2} \sum_{v, v' \in \mathcal{V}} D_{\mathsf{KL}}(P_v \| P_{v'})$$

This lemma can be compared with [10, Corollary 1], where it was shown

$$I(Z^n; V) \le 2(e^\varepsilon - 1)\frac{n}{|\mathcal{V}|^2} \sum_{v, v' \in \mathcal{V}} D_{\mathsf{KL}}(P_v \| P_{v'}). \tag{10}$$

This is a looser bound than Lemma 4 for any $n \ge 1$ and $\varepsilon \ge 0.4$ and only holds for $\delta = 0$.

Example 3. (High-dimensional mean estimation in an ℓ_2-ball) For a parameter $r < \infty$, define

$$\mathcal{P}_r := \{P \in \mathcal{P}(\mathsf{B}_2^d(r))\}, \tag{11}$$

where $\mathsf{B}_2^d(r) := \{x \in \mathbb{R}^d : \ \|x\|_2 \le r\}$ is the ℓ_2-ball of radius r in $\mathbb{R}^d$. The goal is to estimate the mean $\theta(P) = \mathbb{E}[X]$ given the private views Z^n. This example was first studied in [10, Proposition 3] that states $\mathcal{R}_n(\mathcal{P}, \ell_2^2, \varepsilon, 0) \gtrsim r^2 \min\left\{\frac{1}{\varepsilon\sqrt{n}}, \frac{d}{n\varepsilon^2}\right\}$ for $\varepsilon \in (0, 1)$. In the following, we use Lemma 4 to derive a similar lower bound for any $\varepsilon \ge 0$ and $\delta \in (0, 1)$, albeit slightly weaker than [10, Proposition 3].

Corollary 2. *For the non-interactive setting, we have*

$$\mathcal{R}_n(\mathcal{P}, \ell_2^2, \varepsilon, \delta) \gtrsim r^2 \min\left\{\frac{1}{n\varphi_n(\varepsilon, \delta)}, \frac{d}{n^2\varphi_n^2(\varepsilon, \delta)}\right\}. \tag{12}$$

V. Private Bayesian Risk

In the minimax setting, the worst-case parameter is considered which usually leads to over-pessimistic bounds. In practice, the parameter that incurs a worst-case risk may appear with very small probability. To capture this prior knowledge, it is reasonable to assume that the true parameter is sampled from an underlying prior distribution. In this case, we are interested in the *Bayes risk* of the problem.

Let $\mathcal{P} = \{P_{X|\Theta}(\cdot|\theta) : \theta \in \mathcal{T}\}$ be a collection of parametric probability distributions on $\mathcal{X}$ and the parameter space $\mathcal{T}$ is endowed with a prior P_Θ, i.e., $\Theta \sim P_\Theta$. Given an i.i.d. sequence X^n drawn from $P_{X|\Theta}$, the goal is to estimate Θ from a privatized sequence Z^n via an estimator $\Psi : \mathcal{Z}^n \to \mathcal{T}$. Here, we focus on the non-interactive setting. Define the private Bayes risk as

$$R_n^{\mathsf{Bayes}}(P_\Theta, \ell, \varepsilon, \delta) := \inf_{\mathsf{K} \in \mathcal{Q}_{\varepsilon,\delta}} \inf_{\Psi} \mathbb{E}[\ell(\Theta, \Psi(Z^n))], \tag{13}$$

where the expectation is taken with respect to the randomness of both Θ and Z^n. It is evident that $R_n^{\mathsf{Bayes}}(P_\Theta, \ell, \varepsilon, \delta)$ must depend on the prior P_Θ. This dependence can be quantified by

$$\mathcal{L}(\zeta) := \sup_{t \in \mathcal{T}} \Pr(\ell(\Theta, t) \le \zeta), \tag{14}$$

for $\zeta < \sup_{\theta, \theta' \in \mathcal{T}} \ell(\theta, \theta')$. Xu and Raginsky [46] showed that the non-private Bayes risk (i.e., $X^n = Z^n$), denoted by $R_n^{\mathsf{Bayes}}(P_\Theta, \ell)$, is lower bounded as

$$R_n^{\mathsf{Bayes}}(P_\Theta, \ell) \ge \sup_{\zeta > 0} \zeta\left[1 - \frac{I(\Theta; X^n) + \log 2}{\log(1/\mathcal{L}(\zeta))}\right]. \tag{15}$$

Replacing $I(\Theta; X^n)$ with $I(\Theta; Z^n)$ in this result and applying Lemma 2 (similar to Lemma 4), we can directly convert (15) to a lower bound for $R_n^{\mathsf{Bayes}}(P_\Theta, \ell, \varepsilon, \delta)$.

Corollary 3. *In the non-interactive setting, we have*

$$R_n^{\mathsf{Bayes}}(P_\Theta, \ell, \varepsilon, \delta) \ge \sup_{\zeta > 0} \zeta\left[1 - \frac{\varphi_n(\varepsilon, \delta) I(\Theta; X^n) + \log 2}{\log(1/\mathcal{L}(\zeta))}\right].$$

In the following theorem, we provide a lower bound for $R_n^{\mathsf{Bayes}}(P_\Theta, \ell, \varepsilon, \delta)$ that directly involves E_γ-divergence, and thus leads to a tighter bounds than (3). For any pair of random

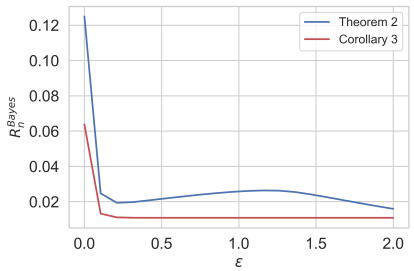

Fig. 1. Comparison of the lower bounds obtained from Theorem 2 and the private version of [46, Theorem 1] described in Corollary 3 for Example 4 assuming $\delta = 10^{-4}$ and $n = 20$.

variables $(A, B) \sim P_{AB}$ with marginals P_A and P_B and a constant $\gamma \geq 0$, we define their E_γ-information as

$$I_\gamma(A;B) \coloneqq \mathsf{E}_\gamma(P_{AB} \| P_A P_B).$$

Theorem 2. *Let K be an (ε, δ)-LDP mechanism. Then, for $n = 1$ we have*

$$R_1^{\text{Bayes}}(P_\Theta, \ell, \varepsilon, \delta) \geq \sup_{\zeta > 0} \zeta \left[1 - \delta I_{e^\varepsilon}(\Theta; X) - e^\varepsilon \mathcal{L}(\zeta)\right],$$

and for $n > 1$ in non-interactive setting we have

$$R_n^{\text{Bayes}}(P_\Theta, \ell, \varepsilon, \delta) \geq \sup_{\zeta > 0} \zeta \left[1 - \varphi_n(\varepsilon, \delta) I_{e^\varepsilon}(\Theta; X^n) - e^\varepsilon \mathcal{L}(\zeta)\right].$$

We compare Theorem 2 with Corollary 3 in the next example.

Example 4. Suppose Θ is uniformly distributed on $[0, 1]$, $P_{X|\Theta=\theta} = \text{Bernoulli}(\theta)$, and $\ell(\theta, \theta') = |\theta - \theta'|$. As mentioned earlier, $\mathcal{L}(\zeta) \leq \min\{2\zeta, 1\}$. We can write for $\gamma = e^\varepsilon$

$$I_\gamma(\Theta; X^n) = \int_0^1 \mathsf{E}_\gamma(P_{X^n|\theta} \| P_{X^n}) \mathrm{d}\theta. \tag{16}$$

A straightforward calculation shows that $P_{X^n|\theta}(x^n) = \theta^{s(x^n)}(1-\theta)^{n-s(x^n)}$, for any $\theta \in [0, 1]$, and $P_{X^n}(x^n) = \frac{s(x^n)!(n-s(x^n))!}{(n+1)!}$ where $s(x^n)$ is the number of 1's in x^n. Given these marginal and conditional distribution, one can obtain after algebraic manipulations

$$I_\gamma(\Theta; X^n) = \frac{1}{n+1} \sum_{s=0}^{n} \int_0^1 \left[\theta^s (1-\theta)^{n-s} \frac{(n+1)!}{s!(n-s)!} - \gamma\right]_+ \mathrm{d}\theta.$$

Plugging this into Theorem 2, we arrive at a maximization problem that can be numerically solved. Similarly, we compute $I(\Theta; X^n) = \int_0^1 D_{\mathsf{KL}}(P_{X^n|\theta} \| P_{X^n}) \mathrm{d}\theta$ and plug it into Corollary 3 and numerically solve the resulting optimization problem. In Fig. 1, we compare these two lower bounds for $\delta = 10^{-4}$ and $n = 20$, indicating the advantage of Theorem 2 for small ε.

Remark 1. *The proof of Theorem 2 leads to the following lower bound for the non-private Bayes risk*

$$R_n^{\text{Bayes}}(P_\Theta, \ell) \geq \sup_{\substack{\zeta > 0, \\ \gamma \geq 0}} \zeta \left[1 - I_\gamma(\Theta; X^n) - \gamma \mathcal{L}(\zeta) - (1-\gamma)_+\right]. \tag{17}$$

For a comparison with (15), consider the following example. Suppose Θ is a uniform random variable on $[0, 1]$ and $P_{X|\Theta=\theta} = \text{Bernoulli}(\theta)$. We are interested in the Bayes risk with respect to the ℓ_1-loss function $\ell(\theta, \theta') = |\theta - \theta'|$. It can be shown that $I(\Theta; X) = 0.19$ nats while

$$I_\gamma(\Theta; X) = \begin{cases} 0.25\gamma^2 & \text{if } \gamma \in [0, 1] \\ 0.25(\gamma - 2)^2 & \text{if } \gamma \in [1, 2] \\ 0 & \text{otherwise.} \end{cases} \tag{18}$$

Moreover, $\mathcal{L}(\zeta) = \sup_{t \in [0,1]} \Pr(|\Theta - t| \leq \zeta) \leq \min\{2\zeta, 1\}$. It can be verified that (15) gives $R_1^{\text{Bayes}}(P_\Theta, \ell_1) \geq 0.03$, whereas our bound (17) yields $R_1^{\text{Bayes}}(P_\Theta, \ell_1) \geq 0.08$.

VI. Private Hypothesis Testing

We now turn our attention to the well-known problem of binary hypothesis testing under local differential privacy constraint. Suppose n i.i.d. samples X^n drawn from a distribution $Q \in \mathcal{P}(\mathcal{X})$ are observed. Let now each X_i be mapped to Z_i via a mechanism $\mathsf{K}_i \in \mathcal{Q}_{\varepsilon,\delta}$ (i.e., sequential interaction is permitted). The goal is to distinguish between the null hypothesis $H_0 : Q = P_0$ from the alternative $H_1 : Q = P_1$ given Z^n. Let T be a binary statistic, generated from a randomized decision rule $P_{T|Z^n} : \mathcal{Z}^n \to \mathcal{P}(\{0, 1\})$ where 1 indicates that H_0 is rejected. Type I and type II error probabilities corresponding to this statistic are given by $\Pr(T = 1|H_0)$ and $\Pr(T = 1|H_1)$, respectively. To capture the optimal tradeoff between type I and type II error probabilities, it is customary to define $\beta_n^{\varepsilon,\delta}(\alpha) \coloneqq \inf \Pr(T = 0|H_1)$ where the infimum is taken over all kernels $P_{T|Z^n}$ such that $\Pr(T = 1|H_0) \leq \alpha$ and non-interactive mechanisms $\mathsf{K}^{\otimes n}$ with $\mathsf{K} \in \mathcal{Q}_{\varepsilon,\delta}$. In the following lemma, we apply Lemma 1 to obtain an asymptotic lower bound for $\beta_n^{\varepsilon,\delta}(\alpha)$.

Corollary 4. *We have for any $\varepsilon \geq 0$ and $\delta \in [0, 1]$*

$$\lim_{n \to \infty} \frac{1}{n} \log \beta_n^{\varepsilon,\delta}(\alpha) \geq -\varphi(\varepsilon, \delta) D_{\mathsf{KL}}(P_0 \| P_1). \tag{19}$$

A similar result was proved by Kairouz et al. [13, Sec. 3] that holds only for sufficiently "small" (albeit unspecified) ε and $\delta = 0$. When compared to Chernoff-Stein lemma [47, Theorem 11.8.3], establishing $D_{\mathsf{KL}}(P_0 \| P_1)$ as the asymptotic exponential decay rate of $\beta_n^{\varepsilon,\delta}(\alpha)$, the above corollary, once again, justifies the reduction of effective sample size from n to $\varphi(\varepsilon, \delta)n$ in the presence of (ε, δ)-LDP requirement.

References

[1] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proc. Theory of Cryptography (TCC)*, Berlin, Heidelberg, 2006, pp. 265–284.

[2] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, "Our data, ourselves: Privacy via distributed noise generation," in *EUROCRYPT*, S. Vaudenay, Ed., 2006, pp. 486–503.

[3] I. Mironov, "Rényi differential privacy," in *Proc. Computer Security Found. (CSF)*, 2017, pp. 263–275.

[4] M. Bun and T. Steinke, "Concentrated differential privacy: Simplifications, extensions, and lower bounds," in *Theory of Cryptography*, 2016, pp. 635–658.

[5] C. Dwork and G. N. Rothblum, "Concentrated differential privacy," vol. abs/1603.01887, 2016. [Online]. Available: http://arxiv.org/abs/1603.01887

[6] J. Dong, A. Roth, and W. J. Su, "Gaussian differential privacy," *arXiv 1905.02383*, 2019.

[7] S. Asoodeh, J. Liao, F. P. Calmon, O. Kosut, and L. Sankar, "Three variants of differential privacy: Lossless conversion and applications," *To appear in Journal on Selected Areas in Information Theory (JSAIT)*, 2021.

[8] A. Evfimievski, J. Gehrke, and R. Srikant, "Limiting privacy breaches in privacy preserving data mining," in *Proc. ACM symp. Principles of Database Systems (PODS)*. ACM, 2003, pp. 211–222.

[9] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith, "What can we learn privately?" *SIAM J. Comput.*, vol. 40, no. 3, pp. 793–826, Jun. 2011.

[10] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Local privacy, data processing inequalities, and statistical minimax rates," in *Proc. Symp. Foundations of Computer Science*, 2013, p. 429–438. [Online]. Available: https://arxiv.org/abs/1302.3203

[11] M. Gaboardi, R. Rogers, and O. Sheffet, "Locally private mean estimation: z-test and tight confidence intervals," in *Proc. Machine Learning Research*, 2019, pp. 2545–2554.

[12] A. Bhowmick, J. Duchi, J. Freudiger, G. Kapoor, and R. Rogers, "Protection against reconstruction and its applications in private federated learning," *arXiv 1812.00984*, 2018.

[13] P. Kairouz, S. Oh, and P. Viswanath, "Extremal mechanisms for local differential privacy," *Journal of Machine Learning Research*, vol. 17, no. 17, pp. 1–51, 2016.

[14] L. P. Barnes, W. N. Chen, and A. Özgür, "Fisher information under local differential privacy," *IEEE Journal on Selected Areas in Information Theory*, vol. 1, no. 3, pp. 645–659, 2020.

[15] J. Acharya, C. L. Canonne, and H. Tyagi, "Inference under information constraints i: Lower bounds from chi-square contraction," *IEEE Transactions on Information Theory*, vol. 66, no. 12, pp. 7835–7855, 2020.

[16] 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, 2018.

[17] D. Wang and J. Xu, "On sparse linear regression in the local differential privacy model," *IEEE Trans. Inf. Theory*, pp. 1–1, 2020.

[18] A. Rohde and L. Steinberger, "Geometrizing rates of convergence under local differential privacy constraints," *Ann. Statist.*, vol. 48, no. 5, pp. 2646–2670, 10 2020.

[19] T. Berrett and C. Butucea, "Locally private non-asymptotic testing of discrete distributions is faster using interactive mechanisms," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 3164–3173.

[20] P. Kairouz, K. Bonawitz, and D. Ramage, "Discrete distribution estimation under local privacy," in *Proc. Int. Conf. Machine Learning*, vol. 48, 20–22 Jun 2016, pp. 2436–2444.

[21] J. Duchi and R. Rogers, "Lower bounds for locally private estimation via communication complexity," in *Proc. Conference on Learning Theory*, 2019, pp. 1161–1191.

[22] R. Ahlswede and P. Gács, "Spreading of sets in product spaces and hypercontraction of the markov operator," *Ann. Probab.*, vol. 4, no. 6, pp. 925–939, 12 1976.

[23] B. Yu, *Assouad, Fano, and Le Cam*. Springer New York, 1997, pp. 423–435.

[24] A. B. Tsybakov, *Introduction to Nonparametric Estimation*, 1st ed. Springer Publishing Company, Incorporated, 2008.

[25] I. Csiszár, "Information-type measures of difference of probability distributions and indirect observations," *Studia Sci. Math. Hungar.*, vol. 2, pp. 299–318, 1967.

[26] S. M. Ali and S. D. Silvey, "A general class of coefficients of divergence of one distribution from another," *Journal of Royal Statistics*, vol. 28, pp. 131–142, 1966.

[27] N. Sharma and N. A. Warsi, "Fundamental bound on the reliability of quantum information transmission," *CoRR*, vol. abs/1302.5281, 2013. [Online]. Available: http://arxiv.org/abs/1302.5281

[28] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[29] B. Balle, G. Barthe, and M. Gaboardi, "Privacy amplification by subsampling: Tight analyses via couplings and divergences," in *NeurIPS*, 2018, pp. 6280–6290.

[30] B. Balle, G. Barthe, M. Gaboardi, and J. Geumlek, "Privacy amplification by mixing and diffusion mechanisms," in *NeurIPS*, 2019, pp. 13 277–13 287.

[31] S. Asoodeh, M. Diaz, and F. P. Calmon, "Privacy analysis of online learning algorithms via contraction coefficients," *arXiv 2012.11035*, 2020.

[32] R. L. Dobrushin, "Central limit theorem for nonstationary markov chains. I," *Theory Probab. Appl.*, vol. 1, no. 1, pp. 65–80, 1956.

[33] P. Del Moral, M. Ledoux, and L. Miclo, "On contraction properties of markov kernels," *Probab. Theory Relat. Fields*, vol. 126, pp. 395–420, 2003.

[34] J. E. Cohen, Y. Iwasa, G. Rautu, M. Beth Ruskai, E. Seneta, and G. Zbaganu, "Relative entropy under mappings by stochastic matrices," *Linear Algebra and its Applications*, vol. 179, pp. 211 – 235, 1993.

[35] V. Anantharam, A. Gohari, S. Kamath, and C. Nair, "On hypercontractivity and a data processing inequality," in *2014 IEEE Int. Symp. Inf. Theory*, 2014, pp. 3022–3026.

[36] Y. Polyanskiy and Y. Wu, "Strong data-processing inequalities for channels and bayesian networks," in *Convexity and Concentration*, E. Carlen, M. Madiman, and E. M. Werner, Eds. New York, NY: Springer New York, 2017, pp. 211–249.

[37] Y. Polyanskiy and Y. Wu, "Dissipation of information in channels with input constraints," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 35–55, Jan 2016.

[38] F. P. Calmon, Y. Polyanskiy, and Y. Wu, "Strong data processing inequalities for input constrained additive noise channels," *IEEE Trans. Inf. Theory*, vol. 64, no. 3, pp. 1879–1892, 2018.

[39] A. Makur and L. Zheng, "Comparison of contraction coefficients for f-divergences," *Probl. Inf. Trans.*, vol. 56, pp. 103–156, 2020.

[40] M. Raginsky, "Strong data processing inequalities andϕ-sobolev inequalities for discrete channels," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3355–3389, June 2016.

[41] H. S. Witsenhausen, "On sequences of pairs of dependent random variables," *SIAM Journal on Applied Mathematics*, vol. 28, no. 1, pp. 100–113, 1975.

[42] S. Asoodeh, M. Aliakbarpour, and F. Calmon, "Local differential privacy is equivalent to contraction of E_γ-divergence," 2021. [Online]. Available: https://arxiv.org/abs/2102.01258

[43] S. L. Warner, "Randomized response: A survey technique for eliminating evasive answer bias," *Journal of the American Statistical Association*, vol. 60, no. 309, pp. 63–69, 1965.

[44] J. Cohen, J. Kemperman, and G. Zbăganu, *Comparisons of Stochastic Matrices, with Applications in Information Theory, Economics, and Population Sciences*. Birkhäuser, 1998.

[45] Y. Yang and A. Barron, "Information-theoretic determination of minimax rates of convergence," *Ann. Statist.*, vol. 27, no. 5, pp. 1564–1599, 10 1999.

[46] A. Xu and M. Raginsky, "Converses for distributed estimation via strong data processing inequalities," in *IEEE Int. Sympos. Inf. Theory (ISIT)*, 2015, pp. 2376–2380.

[47] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.