

# $L_q$ Lower Bounds on Distributed Estimation via Fisher Information

Wei-Ning Chen and Ayfer Özgür

Department of Electrical Engineering, Stanford University

**Abstract**—Van Trees inequality, also known as the Bayesian Cramér-Rao lower bound, is a powerful tool for establishing lower bounds for minimax estimation through Fisher information. It easily adapts to different statistical models and often yields tight bounds. Recently, its application has been extended to distributed estimation with privacy and communication constraints where it yields order-wise optimal minimax lower bounds for various parametric tasks under squared  $L_2$  loss.

However, a widely perceived drawback of the van Trees inequality is that it is limited to squared  $L_2$  loss. The goal of this paper is to dispel that perception by introducing a strengthened version of the van Trees inequality that applies to general  $L_q$  loss functions by building on the Efroimovich’s inequality – a lesser-known entropic inequality dating back to the 1970s. We then apply the generalized van Trees inequality to lower bound  $L_q$  loss in distributed minimax estimation under communication and local differential privacy constraints. This leads to lower bounds for  $L_q$  loss that apply to sequentially interactive and blackboard communication protocols. Additionally, we show how the generalized van Trees inequality can be used to obtain *local* and *non-asymptotic* minimax results that capture the hardness of estimating each instance at finite sample sizes.

## I. INTRODUCTION

For a real-valued parameter  $\theta \in \Theta$  and an observation  $X \sim P_\theta$ , the basic question of parametric statistics is how well one can estimate  $\theta$  from  $X$  under a given loss function  $\ell(\hat{\theta} - \theta)$ . The Fisher information  $I_X(\theta)$  plays a crucial role in this context by capturing the *local* hardness of estimating  $\theta$  from  $X$ , with implications both asymptotically [17] and non-asymptotically [11, 16]. For example, the famous Cramér-Rao lower bound [11, 22] states that the squared  $L_2$  estimation error (a.k.a. the mean squared error) of any unbiased estimator is lower bounded by the inverse of the Fisher information. The unbiasedness assumption on the estimator can be removed by assuming that the parameter  $\theta$  is distributed according to a prior on  $\Theta$ . This is known as the Bayesian Cramér-Rao lower bound or the van Trees Inequality [26, 16]. By strategically constructing the prior, this approach can also be used to prove lower bounds for minimax estimation, where the goal is to minimize the worst-case error of the estimator over the parameter space  $\Theta$ .

The classic parametric estimation tasks have gained renewed popularity over recent years, driven by the prevalence of modern datasets often generated and stored on local devices, such as in federated learning (FL) and analytics (FA) [21].

In these settings, the collection and utilization of decentralized data encounter various resource constraints, including communication and privacy considerations. These constraints have raised the question of how well one can estimate the unknown parameter  $\theta \in \Theta$  from a processed observation  $Y$ , which corresponds to the output of a privatization and/or compression mechanism applied to  $X \sim P_\theta$ . In Barnes et al. [4, 5, 6], the classical Fisher information framework has been extended to the case of privacy and communication (or compression) constraints. This approach first upper bounds the Fisher information from a differentially private and/or compressed sample  $Y$  and then uses the van Trees inequality to lower bound the minimax squared  $L_2$  error of distributed estimation. This, in a unified fashion, leads to order-optimal lower bounds for various parametric tasks under privacy and/or communication constraints, including two tasks that have been of significant interest in the recent literature: distributed mean estimation [23] and discrete distribution (or frequency) estimation [28]. While order-optimal minimax lower bounds for distributed estimation under information constraints can also be derived by using other techniques, e.g., leveraging strong data-processing inequalities [13, 27, 15, 8, 12, 3] or employing methods based on Le Cam, Fano, or Assouad [1], the constrained Fisher information approach has several distinct advantages:

- It is relatively straightforward and applies to various parametric models in a unified fashion, e.g., mean estimation or distribution estimation. In contrast, methods based on Le Cam, Fano, or Assouad are more versatile in that they can be potentially adopted to diverse statistical problems beyond parametric settings, e.g., testing, but the construction of worst-case instances tends to be intricate and is heavily tied to the specific problem structure.
- The Fisher information approach elucidates a clear relationship between the tail behavior of the score function of the parametric model and error rates under privacy/communication constraints.
- Leveraging the chain rule of Fisher information, the Fisher information approach naturally extends to sequentially interactive communication protocols or even fully interactive blackboard protocols [19].
- Finally, Fisher information is inherently a local measure of hardness. For example, the influential Hájek-Le Cam’s local asymptotic minimax (LAM) theorem asserts that,

This work was partially supported by NSF Award # CCF-2213223.

for a general ball-shaped loss function  $\ell(\cdot)$ , the *asymptotic* estimation error at  $\theta \in \Theta$  is at least  $\mathbb{E}[\ell(Z)]$ , where  $Z \sim \mathcal{N}(0, I_X^{-1}(\theta))$  [17]. Using the van Trees inequality with a carefully constructed prior, Chen et al. [10] shows that Fisher information can be used to prove *local* and *non-asymptotic* minimax lower bounds that capture the hardness of estimating each instance of the problem at finite sample size. To the best of our knowledge, this is the only example of *non-asymptotic local* lower bounds for distributed estimation, combining the power of Hájek-Le Cam type results to capture *local* hardness with the *non-asymptotic* nature of global minimax lower bounds that have been almost exclusively the focus of the existing distributed estimation literature.

However, a widely perceived drawback of the Fisher information approach, both for classical parametric statistics as well as its extension to the distributed setting with privacy and communication constraints, is that its applicability is limited to squared  $L_2$  error. For example, Tsybakov [25] acknowledges several advantages of van Trees inequality, such as its relative simplicity in application and ability to establish sharp bounds, and also notes “*a limitation is that the van Trees inequality applies only to the squared loss function.*” The goal of this paper is to dispel this perception by showing how Fisher information can be used to prove lower bounds for any  $L_q$  loss for  $1 \leq q < \infty$ . We do this by building on a less-known entropic inequality called Efroimovich’s inequality [14] highlighted in recent works Aras et al. [2], Lee [20]. Our paper makes the following contributions:

- We prove a van Trees type inequality for  $L_q$  loss by combining Efroimovich’s inequality with a maximum entropy argument.<sup>1</sup>
- We leverage this generalized van Trees inequality to establish global minimax lower bounds for various distributed estimation tasks under  $L_q$  loss. As an immediate consequence, this yields lower bounds for  $L_q$  loss applicable to sequentially interactive and blackboard communication protocols. Our approach not only recovers previous lower bounds presented in Acharya et al. [1] for sequentially interactive models in a cleaner and more straightforward manner, but also extends to fully interactive blackboard communication protocols.
- We show how the generalized van Trees inequality can be used to derive *local non-asymptotic* minimax lower bounds for distribution estimation under  $L_q$  loss extending the approach of Chen et al. [10]. This *local* lower bounds match the performance of a scheme previously developed in Chen et al. [10] and establish its instance-optimality under  $L_1$  loss. The results extend to  $L_q$  loss.

*Organization.*: The rest of the paper is organized as follows. We recap van Trees inequality and extend it to  $L_q$

loss via Efroimovich inequality in Section II. In Section III, we leverage the generalized van Trees inequality to establish global minimax lower bounds for various distributed estimation tasks under  $L_q$  loss. In Section IV, we extend our focus to derive *local* minimax lower bounds for distribution estimation tasks and provide schemes that achieve the lower bounds pointwisely. Finally, we summarize the work in Section V.

## II. PRELIMINARIES

In this section, we recap van Trees inequality and Efroimovich inequality and refer interested readers to [2] and [20] for more details and recent advances. Throughout the paper, we consider the following parametric model: let  $\{P_\theta | \theta \in \Theta\}$  with  $\Theta \subseteq \mathbb{R}^d$  be a family of probability measures over  $\mathcal{X}$  with a dominating  $\sigma$ -finite measure  $\lambda$  such that the density:

$$dP_\theta(\cdot) = f(\cdot; \theta) d\lambda(\cdot)$$

exists. We make the following assumption on the model:

*Assumption 2.1:* The density function  $f(x; \theta)$  is differentiable for  $x \in \mathcal{X}$   $\lambda$ -a.e. In addition,  $f$  satisfies

$$\int_{\mathcal{X}} \nabla_\theta f(x; \theta) d\lambda(x) = 0,$$

for all  $\theta \in \Theta$ .

Assumption 2.1 is a common regularity condition in the Cramér-Rao type bounds, allowing for exchanging the differentiation and integration. For a prior distribution  $\pi$  over  $\Theta$ , the information theorists’ Fisher information is defined as

$$J(\pi) \triangleq \int_{\mathbb{R}^d} \frac{|\nabla_\theta \pi(\theta)|^2}{\pi(\theta)} d\theta, \quad (1)$$

if it exists. On the other hand, the Fisher information matrix of  $\{P_\theta | \theta \in \Theta\}$  is defined as

$$[I_X(\theta)]_{ij} \triangleq \int_{\mathcal{X}} \frac{\frac{\partial}{\partial \theta_i} f(x; \theta) \cdot \frac{\partial}{\partial \theta_j} f(x; \theta)}{f(x; \theta)} d\lambda(x). \quad (2)$$

Van Trees inequality states that the  $L_2$  estimation error for estimating  $\theta \in \Theta$ , given a prior distribution  $\pi$  over  $\Theta$ , is lower bound by the inverse Fisher information:

*Theorem 2.2 ([26, 16]):* Let  $X \sim P_\theta$  and  $\theta \sim \pi$  for some prior distribution  $\pi$ . Let Assumption 2.1 hold. Then, for any prior distribution  $\pi$  on  $\Theta$  for which the information theorist’s Fisher information  $J(\pi)$  exists, it holds that

$$\mathbb{E} \left[ \left\| \theta - \hat{\theta} \right\|_2^2 \right] \geq \frac{d}{\det(I_X(\theta) + J(\pi))^{1/d}}. \quad (3)$$

While van Trees inequality only applies to the  $L_2$  error, the following Efroimovich inequality can be used to establish lower bounds for more general loss functions.

*Theorem 2.3 ([14]):* Under the assumptions of Theorem 2.2, we also have

$$\frac{1}{2\pi e} e^{\frac{2}{d} h(\theta|X)} \geq \frac{1}{\det(I_X(\theta) + J(\pi))^{1/d}}, \quad (4)$$

where  $h(\theta|X)$  is the conditional (differential) entropy.

<sup>1</sup>In Aras et al. [2], Lee [20], the authors comment that Efroimovich’s inequality can be potentially used to bound loss functions beyond  $L_2$ . However, we have not been able to find an explicit result in this direction.

To see how Theorem 2.3 implies the classical van Trees inequality, observe that

$$\begin{aligned} \frac{2}{d}h(\theta|X) &\leq \frac{2}{d}h(\theta|\hat{\theta}(X)) \leq \frac{2}{d}h(\theta - \hat{\theta}) \\ &\leq \log \left( 2\pi e \cdot \det \left( \mathbb{E} \left[ (\hat{\theta} - \theta)(\hat{\theta} - \theta)^\top \right] \right)^{1/d} \right) \\ &\leq \log \left( 2\pi e \cdot \frac{1}{d} \mathbb{E} \left[ \left\| \hat{\theta} - \theta \right\|_2^2 \right] \right). \end{aligned} \quad (5)$$

The last two inequalities follow from the fact that the Gaussian distribution maximizes entropy for fixed second moments and the AM-GM inequality. Rearranging yields Theorem 2.2.

Through Efroimovich's inequality, one can easily extend van Trees inequality to  $L_1$  loss: following (5), we obtain

$$\begin{aligned} \frac{2}{d}h(\hat{\theta} - \theta) &\leq \frac{2}{d} \sum_i h(\hat{\theta}_i - \theta_i) \leq \frac{2}{d} \sum_i \log \left( 2e \mathbb{E} \left[ \left| \theta_i - \hat{\theta}_i \right| \right] \right) \\ &\leq 2 \log \left( \frac{2e}{d} \sum_i \mathbb{E} \left[ \left| \theta_i - \hat{\theta}_i \right| \right] \right) \\ &= \log \left( \left( \frac{2e}{d} \mathbb{E} \left[ \left\| \hat{\theta} - \theta \right\|_1 \right] \right)^2 \right). \end{aligned}$$

The last two inequalities follow from (1) the fact that the Laplace distribution maximizes entropy for a fixed mean absolute error, and (2) the AM-GM inequality. Applying Theorem 2.3, we immediately obtain the following lower bound for  $L_1$  loss:

$$\mathbb{E} \left[ \left\| \hat{\theta} - \theta \right\|_1 \right] \geq \sqrt{\frac{\pi}{2e}} \frac{d}{\det(I_X(\theta) + J(\pi))^{1/2d}}. \quad (6)$$

The above lower bound is order-wise optimal. To see this, suppose that there exists an order-optimal estimator  $\hat{\theta}^*(X)$  such that

$$\mathbb{E} \left[ \left\| \theta - \hat{\theta}^* \right\|_2^2 \right] = C \cdot \frac{d}{\det(I_X(\theta) + J(\pi))^{1/d}}.$$

Then, the Cauchy-Schwartz inequality yields

$$\begin{aligned} \mathbb{E} \left[ \left\| \theta - \hat{\theta}^* \right\|_1 \right] &\leq \sqrt{d \cdot \mathbb{E} \left[ \left\| \theta - \hat{\theta}^* \right\|_2^2 \right]} \\ &= \sqrt{C} \cdot \frac{d}{\det(I_X(\theta) + J(\pi))^{1/2d}}. \end{aligned}$$

Since it is well-known that the  $L_2$  lower bound can be achieved, this argument implies the  $L_1$  lower bound is also tight up to a constant. More generally, we can prove the following van Trees inequality for any  $L_q$  loss:

**Theorem 2.4 (Generalized van Trees Inequality):** Let  $q \geq 1$  and let  $C_{\text{ME}}(q)$  be the partition function of the  $L_q$  max-entropy distribution, formally defined as<sup>2</sup>

$$C_{\text{ME}}(q) \triangleq 2e^{\frac{1}{q}} \Gamma \left( \frac{1}{q} \right) q^{\frac{1}{q}-1}.$$

<sup>2</sup>One can verify that  $C_{\text{ME}}(2) = \sqrt{2\pi e}$  and  $C_{\text{ME}}(1) = \sqrt{2e}$ .

Then, it holds that

$$\mathbb{E}_{\theta, X} \left[ \left\| \hat{\theta}(X) - \theta \right\|_q^q \right] \geq \left( \frac{\sqrt{2\pi e}}{C_{\text{ME}}(q)} \right)^q \frac{d}{\det(I_X(\theta) + J(\pi))^{\frac{q}{2d}}}.$$

We can simplify the lower bound with the following AM-GM inequality

$$\det(A)^{1/d} \leq \frac{\text{Tr}(A)}{d} \text{ for any PSD } A \in \mathbb{R}^{d \times d},$$

obtaining the following simpler but slightly weaker form:

**Proposition 2.5:** Let  $q \geq 1$  and  $C_{\text{ME}}(q)$  be defined as in Theorem 2.4. Then, it holds that

$$\mathbb{E}_{\theta, X} \left[ \left\| \hat{\theta}(X) - \theta \right\|_q^q \right] \geq \left( \frac{\sqrt{2\pi e}}{C_{\text{ME}}(q)} \right)^q \cdot \frac{d^{1+\frac{q}{2}}}{\text{Tr}(I_X(\theta) + J(\pi))^{\frac{q}{2}}}.$$

### III. INFORMATION-CONSTRAINED ESTIMATION

The generalized van Trees inequality (in the form stated in Proposition 2.5), can be combined with an upper bound on the trace of the Fisher information matrix to obtain a lower bound on the minimax  $L_q$  loss achievable in a distributed setting under privacy and communication constraints. This approach is rather straightforward and is outlined in [5, 6, 4], where authors also develop the necessary upper bounds on the trace of the Fisher information matrix under privacy and communication constraints. In this section, we overview how privacy and communication constraints are mathematically modeled in a distributed estimation setting and state the corresponding lower bounds under  $L_q$  loss.

#### A. Problem Setup

The general distributed statistical task we consider in this paper can be formulated as follows. Each one of the  $n$  clients observes a local sample  $X_i \sim P_\theta$ , processes it via a local channel (i.e., a randomized mapping), and then sends a message  $Y_i \in \mathcal{Y}$  to the server, which, upon receiving  $Y^n$ , aims to estimate the unknown parameter  $\theta$ .

At client  $i$ , the message  $Y_i$  is generated via a *sequential* communication protocol; that is, samples are communicated sequentially by broadcasting the communication to all nodes in the system, including the server. Therefore, the encoding function  $W_i$  of the  $i$ -th client can depend on all previous messages  $Y_1, \dots, Y_{i-1} \in \mathcal{Y}$ . Formally, it can be written as a randomized mapping (possibly using shared randomness across participating clients and the server) of the form  $Y_i \sim W_i(\cdot | X_i, Y^{i-1})$ .

**$b$ -bit communication constraint:** The  $b$ -bit communication constraint restricts  $|\mathcal{Y}| \leq 2^b$ , ensuring the local message can be described in  $b$  bits.

**Remark 3.1:** While we only overview the  $b$ -bit sequentially interactive communication model, our results extend to the  $b$ -bit blackboard communication protocol [19], where each node is allowed to write  $b$ -bits in total on a publicly seen blackboard in a randomized order that can depend on the samples. The blackboard model allows for much more interaction between the nodes as compared to the sequential model (e.g. the

protocol can start with one of the nodes writing a single bit on the blackboard and the second node to write a bit can depend on the value of the first written bit etc.). The results in Corollary 3.2 trivially apply to these more powerful communication protocols simply because the Fisher information bounds in [5, 6] are proven under this more general model.

*$\varepsilon$ -local differential constraint:* The  $\varepsilon$ -LDP constraint requires that, for any  $i \in [n]$ ,  $x_i, x'_i \in \mathcal{X}$ ,  $y^{i-1} \in \mathcal{Y}^{i-1}$ ,  $\mathcal{S} \in \sigma(Y_i)$ , it holds that

$$\frac{W_i(\mathcal{S}|x_i, y^{i-1})}{W_i(\mathcal{S}|x'_i, y^{i-1})} \leq e^\varepsilon. \quad (7)$$

It ensures that any adversary who observes the output  $Y_i$  and the context  $Y^{i-1}$  cannot infer the local sample  $X_i$ .

As a special case, when  $W_i$  depends only on  $X_i$  and is independent of the other messages  $Y^{i-1}$  for all  $i$  (i.e.  $W_i(\cdot|X_i, Y^{i-1}) = W_i(\cdot|X_i)$ ), we say the corresponding protocol is *non-interactive*. Finally, we call the tuple  $(W^n, \hat{\theta}(Y^n))$  an estimation scheme, where  $\hat{\theta}(Y^n)$  is an estimator of  $\theta$ . We use  $\Pi_{\text{seq}}$  and  $\Pi_{\text{ind}}$  to denote the collections of all sequentially interactive and non-interactive schemes, respectively. The goal is to lower bound the global minimax  $L_q$  risk:

$$R(L_q, \Theta) \triangleq \inf_{(W^n, \hat{\theta}) \in \Pi_{\text{seq}}} \sup_{\theta \in \Theta} \mathbb{E} \left[ \left\| \theta - \hat{\theta}(Y^n) \right\|_q^q \right]$$

and construct schemes in  $\Pi_{\text{seq}}$  that match these lower bounds.

### B. Global Minimax Lower Bounds

Combining Proposition 2.5 with the Fisher information bounds in [5], we obtain the following lower bounds on the  $L_q$  risk of common statistical models, including the discrete distribution estimation and the Gaussian mean estimation.

*Corollary 3.2 (Estimation under  $b$ -bit constraint):* Let  $\kappa(q) \triangleq \left( \frac{\sqrt{2\pi e}}{C_{\text{ME}}(q)} \right)^q$ . Then the following lower bounds hold:

- Gaussian location model: let  $X \sim \mathcal{N}(\theta, \sigma^2 I_d)$  with  $[-B, B] \subset \Theta$ . For  $nB^2 \min(b, d) \geq d\sigma^2$ , we have

$$R(L_q, \Theta) \gtrsim d\kappa(q) \max \left\{ \left( \frac{d\sigma^2}{nb} \right)^{\frac{q}{2}}, \left( \frac{\sigma^2}{n} \right)^{\frac{q}{2}} \right\}. \quad (8)$$

- Gaussian covariance estimation: Suppose that  $X \sim \mathcal{N}(0, \text{diag}(\theta_1, \dots, \theta_d))$  with  $[\sigma_{\min}^2, \sigma_{\max}^2] \subset \Theta$ . Then for  $n(\sigma_{\max}^2 - \sigma_{\min}^2)^2 \min(b^2, d) \geq d\sigma_{\min}^4$ , we have

$$R(L_q, \Theta) \gtrsim d\kappa(q) \max \left\{ \left( \frac{d\sigma_{\min}^4}{nb^2} \right)^{\frac{q}{2}}, \left( \frac{\sigma_{\min}^4}{n} \right)^{\frac{q}{2}} \right\}. \quad (9)$$

- Distribution estimation: Suppose that  $\mathcal{X} = \{1, 2, \dots, d\}$  and that  $f(x|\theta) = \theta_x$ . Let  $\Theta = \Delta_d$  be the  $d$ -dim probability simplex. For  $n \min(2^b, d) \geq d^2$ , we have

$$R(L_q, \Theta) \gtrsim d\kappa(q) \max \left\{ \left( \frac{1}{n2^b} \right)^{\frac{q}{2}}, \left( \frac{1}{nd} \right)^{\frac{q}{2}} \right\}. \quad (10)$$

- Product Bernoulli model: suppose that  $X \sim \prod_{i=1}^d \text{Bern}(\theta_i)$ . If  $\Theta = [0, 1]^d$ , then for  $n \min\{b, d\} \geq d$ , we have

$$R(L_q, \Theta) \gtrsim d\kappa(q) \max \left\{ \left( \frac{d}{nb} \right)^{\frac{q}{2}}, \left( \frac{1}{n} \right)^{\frac{q}{2}} \right\}. \quad (11)$$

for some universal constant  $C$ . On the other hand,  $\Theta = \Delta_d$ , then for  $n \min\{2^b, d\} \geq d^2$ , we get instead

$$R(L_q, \Theta) \gtrsim d\kappa(q) \max \left\{ \left( \frac{1}{n2^b} \right)^{\frac{q}{2}}, \left( \frac{1}{nd} \right)^{\frac{q}{2}} \right\}. \quad (12)$$

Similarly, we can prove lower bounds under the  $\varepsilon$ -local DP model by using the Fisher information upper bounds in [6]. We defer this to Corollary A.1 in Appendix A. Corollary 3.2 and Corollary A.1 recover several existing lower bounds and extend them into the broader blackboard communication models. Specifically, (8) and (15) recover the non-sparse setting of in [1, Theorem 4] for  $q < \infty$ ; (10) and (17) recover [1, Corollary 3]; (11), (12), (17), and (18) recover [1, Theorem 3] for  $q < \infty$ .

### C. Achievability

We note that nearly all of the aforementioned lower bounds are order-wise tight<sup>3</sup>, meaning that they accurately characterize the correct dependence on parameters such as  $d$ ,  $n$ ,  $b$ , and  $\varepsilon$ .

*Lemma 3.3:* Assume the same conditions in Corollary 3.2. Then (8), (10), (11), (12) from Corollary 3.2 are tight. The same results hold for Corollary A.1 in Appendix A.

For  $1 \leq q \leq 2$ , the upper bounds readily follow from Hölder's inequality. For  $q > 2$ , the order-optimal estimation schemes can be constructed via a sample splitting trick in a fashion similar to [18, 6], which yields optimal error performance. Notably, these upper bounds can be achieved through independent protocols, indicating that the global minimax error is not improved by interaction among clients.

## IV. LOCAL MINIMAX BOUNDS

In Section III, we establish minimax bounds for various distributed statistical estimation models. However, these *global* minimax lower bounds tend to be too conservative and may not accurately reflect the difficulty of estimating each instance  $\theta \in \Theta$ . This is especially notable in the case of discrete distribution estimation. For example, we would expect the estimation problem to be inherently easier when the underlying distribution we aim to estimate happens to be sparse. Ideally, it is desirable to have estimation schemes that can automatically adapt to the hardness of the underlying instance, i.e. achieve smaller error in easier instances of the problem, without knowing the instance ahead of time. In contrast, globally minimax optimal estimation schemes are typically designed and tuned for worst-case scenarios and can therefore be too pessimistic, i.e. not able to exploit structures that make the problem instance easier.

<sup>3</sup>The only exception is the sub-exponential case (9), in which the tightness of the lower bound (even under  $L_2$  loss) remains open.

In this section, we focus on the discrete distribution estimation problem and derive *local* bounds with respect to general  $L_q$  loss. We consider the same setting as in Section III; however to highlight that the parameter of interest is an instance of the  $d$ -dim probability simplex, we use  $p = (p_1, p_2, \dots, p_d) \in \Delta_d$  to denote the unknown parameter ( $\theta \in \Theta$ ). The goal here is to design a scheme  $(W^n, \hat{p}(Y^n))$  to minimize the  $L_q$  *local* statistical risk:

$$R(L_q, p, (W^n, \hat{p})) \triangleq \mathbb{E} \left[ \|p - \hat{p}(Y^n)\|_q^q \right] \text{ for all } p \in \Delta_d.$$

We mainly focus on the regime  $1 \ll d \ll n$  and aim to characterize the *pointwise* statistical convergence rates when  $n$  is sufficiently large.

#### A. $b$ -bit communication constraint

We start with the  $b$ -bit communication constraint. [10] provides a two-round scheme that achieves the following estimation error:

**Theorem 4.1** ( $L_2$  local minimax error [10]): Let  $b \leq \lfloor \log_2 d \rfloor$ .

- 1) There exists a sequentially interactive scheme (with a single round of interaction)  $(W^n, \check{p}) \in \Pi_{\text{seq}}$ , such that for all  $p \in \Delta_d$ ,

$$R(L_2, p, (W^n, \check{p})) \asymp \frac{\|p\|_{\frac{1}{2}} + o_n(1)}{n2^b}. \quad (13)$$

- 2) There exists a sequentially interactive scheme (with a single round of interaction)  $(\tilde{W}^n, \check{p}) \in \Pi_{\text{seq}}$ , such that for all  $p \in \Delta_d$ ,

$$R(L_1, p, (\tilde{W}^n, \check{p})) \lesssim \sqrt{\frac{\|p\|_{\frac{1}{3}} + o_n(1)}{n2^b}}. \quad (14)$$

Note that the upper bound on the risk is now a function of the unknown  $p$ . [9] develops the matching lower bound in (13) for  $L_2$  loss. The optimality of the  $L_1$  rate in (14) has remained open. In this paper, we close this gap by proving the following matching lower bound for general  $L_q$  loss.

**Theorem 4.2** ( $b$ -bit  $L_q$  local minimax lower bounds): Let  $p \in \Delta'_d \triangleq \{p \in \Delta_d \mid \frac{1}{2} < p_1 < \frac{2}{3}\}$ . Then for any  $\delta > 0, B \geq \sqrt{\frac{\|p\|_{1/2}}{d2^b}}$ , as long as  $n = \Omega\left(\frac{d^3 \log d}{\|p\|_{1/2}}\right)$ , it holds that<sup>4</sup>

$$\inf_{(W^n, \check{p}) \in \Pi_{\text{seq}}} \sup_{p' : \|p' - p\|_\infty \leq \frac{B}{\sqrt{n}}} \mathbb{E}_{p'} \left[ \|\hat{p}(W^n(X^n)) - p'\|_q^q \right] \gtrsim \max \left( \frac{C_\delta \|p\|_{\frac{q}{q+2} + \delta}}{(n2^b)^{\frac{q}{2}}}, \frac{\|p\|_{\frac{q}{q+2}}}{(n2^b)^{\frac{q}{2}} \log d}, \frac{C_\delta \|p\|_{\frac{q/2+\delta}{q/2+\delta}}}{n^{\frac{q}{2}}}, \frac{\|p\|_{\frac{q/2}{q/2}}}{n^{\frac{q}{2}} \log d} \right)$$

where  $C_\delta \triangleq (\delta/(1+\delta))^2$  is a  $\delta$ -dependent constant.

<sup>4</sup>Indeed, the lower bound holds for blackboard interactive schemes [19], a more general class of interactive schemes than  $\Pi_{\text{seq}}$ . See [5] for a discussion of blackboard schemes.

*Sketch of the proof:* The proof is based on the framework introduced in [5] (see also [10] for the analysis of  $\ell_2$  case), where a global upper bound on the quantized Fisher information is given and used to derive the minimax lower bound on the  $\ell_2$  error. We extend their results to the local regime and develop a *local* upper bound on the quantized Fisher information around a neighborhood of  $p$ .

To obtain a local upper bound, we construct an  $h$ -dimensional parametric sub-model  $\Theta_p^h$  that contains  $p$  and is a subset of  $\mathcal{P}_d$ , where  $h \in [d]$  is a tuning parameter and will be determined later. By considering the sub-model  $\Theta_p^h$ , we can control its Fisher information around  $p$  with a function of  $h$  and  $p$ . Optimizing over  $h \in [d]$  yield an upper bound that depends on  $\|p\|_{\frac{q}{q+2}}$ . Finally, the local upper bound on the quantized Fisher information is translated to a local minimax lower bound on the  $L_q$  error via the generalized van Trees inequality (Theorem 2.4). We defer the detailed proof to Appendix B.  $\square$

**Remark 4.3:** Note that Theorem 4.2 complements Theorem 4.1 and is nearly tight (up to a  $\log d$  factor) for  $q = 1$  and  $q = 2$ . We believe the upper bound can be easily adapted to match the lower bound for all  $q > 1$ .

#### B. $\varepsilon$ -local differential privacy

Similarly, under the  $\varepsilon$ -local DP constraint, we prove the following local lower bound:

**Theorem 4.4** ( $\varepsilon$ -LDP  $L_q$  local minimax lower bounds): Let  $p \in \Delta'_d \triangleq \{p \in \Delta_d \mid \frac{1}{2} < p_1 < \frac{2}{3}\}$ . Then for any  $\delta > 0, B \geq \sqrt{\frac{\|p\|_{1/2}}{d \min(e^\varepsilon, (e^\varepsilon - 1)^2)}}$ , as long as  $n = \Omega\left(\frac{d^3 \log d}{\|p\|_{1/2}}\right)$ , it holds that

$$\inf_{(W^n, \check{p}) \in \Pi_{\text{seq}}} \sup_{p' : \|p' - p\|_\infty \leq \frac{B}{\sqrt{n}}} \mathbb{E}_{p'} \left[ \|\hat{p}(W^n(X^n)) - p'\|_q^q \right] \gtrsim \max \left( \frac{C_\delta \|p\|_{\frac{q}{q+2} + \delta}}{\left(n \min(e^\varepsilon, (e^\varepsilon - 1)^2)\right)^{\frac{q}{2}}}, \frac{C_1 \|p\|_{\frac{q}{q+2}}}{\left(n \min(e^\varepsilon, (e^\varepsilon - 1)^2)\right)^{\frac{q}{2}} \log d}, \frac{C_\delta \cdot \|p\|_{\frac{q/2+\delta}{q/2+\delta}}}{n^{\frac{q}{2}}}, \frac{C_2 \|p\|_{\frac{q/2}{q/2}}}{n^{\frac{q}{2}} \log d} \right),$$

for some  $C_\delta, C_1, C_2 > 0$ .

However, unlike the  $b$ -bit communication constraint, the tightness of Theorem 4.4 is unknown, and a matching achievability scheme currently remains unsettled.

#### V. CONCLUSION

In this work, we first point out that the popular van Trees inequality can be generalized to accommodate general loss functions through Efroimovich's inequality. Subsequently, we extend the application of the generalized van Trees inequality to  $L_q$  loss in the context of distributed estimation under communication and local differential privacy constraints. Notably, combining with previous bounds on constrained Fisher information, our results offer a significantly simplified analysis over existing (global) lower bounds and, more importantly, can be used to derive local minimax results that capture the hardness of instances.

# REFERENCES

- [1] Jayadev Acharya, Clément L Canonne, Ziteng Sun, and Himanshu Tyagi. Unified lower bounds for interactive high-dimensional estimation under information constraints. *Advances in Neural Information Processing Systems*, 36, 2023.
- [2] Efe Aras, Kuan-Yun Lee, Ashwin Pananjady, and Thomas A Courtade. A family of bayesian cramér-rao bounds, and consequences for log-concave priors. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 2699–2703. IEEE, 2019.
- [3] Shahab Asodeh and Huanyu Zhang. Contraction of locally differentially private mechanisms. *arXiv preprint arXiv:2210.13386*, 2022.
- [4] Leighton Pate Barnes, Yanjun Han, and Ayfer Özgür. Fisher information for distributed estimation under a blackboard communication protocol. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 2704–2708. IEEE, 2019.
- [5] Leighton Pate Barnes, Yanjun Han, and Ayfer Özgür. Lower bounds for learning distributions under communication constraints via fisher information, 2019.
- [6] Leighton Pate Barnes, Wei-Ning Chen, and Ayfer Özgür. Fisher information under local differential privacy. *IEEE Journal on Selected Areas in Information Theory*, 1(3): 645–659, 2020.
- [7] A. A. Borovkov. *Mathematical Statistics*. Gordon and Breach Science Publishers, 1998.
- [8] Mark Braverman, Ankit Garg, Tengyu Ma, Huy L Nguyen, and David P Woodruff. Communication lower bounds for statistical estimation problems via a distributed data processing inequality. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 1011–1020, 2016.
- [9] Wei-Ning Chen, Peter Kairouz, and Ayfer Ozgur. Breaking the dimension dependence in sparse distribution estimation under communication constraints. In *Conference on Learning Theory*, pages 1028–1059. PMLR, 2021.
- [10] Wei-Ning Chen, Peter Kairouz, and Ayfer Özgür. Pointwise bounds for distribution estimation under communication constraints. *Advances in Neural Information Processing Systems*, 34, 2021.
- [11] Harald Cramér. *Mathematical methods of statistics*, volume 43. Princeton university press, 1999.
- [12] John C. Duchi and Ryan Rogers. Lower bounds for locally private estimation via communication complexity. In *Conference on Learning Theory*, pages 1161–1191. PMLR, 2019.
- [13] John C. Duchi, Michael I Jordan, and Martin J Wainwright. Local privacy and statistical minimax rates. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 429–438. IEEE, 2013.
- [14] S Yu Efroimovich. Information contained in a sequence of observations. *Problems in Information Transmission*, 15:24–39, 1980.
- [15] Ankit Garg, Tengyu Ma, and Huy Nguyen. On communication cost of distributed statistical estimation and dimensionality. In *Advances in Neural Information Processing Systems*, pages 2726–2734, 2014.
- [16] Richard D Gill and Boris Y Levit. Applications of the van trees inequality: a bayesian cramér-rao bound. *Bernoulli*, 1(1-2):59–79, 1995.
- [17] Jaroslav Hájek. Some extensions of the wald-wolfowitz-noether theorem. *The Annals of Mathematical Statistics*, pages 506–523, 1961.
- [18] Yanjun Han, Pritam Mukherjee, Ayfer Özgür, and Tsachy Weissman. Distributed statistical estimation of high-dimensional and nonparametric distributions. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 506–510. IEEE, 2018.
- [19] Eyal Kushilevitz. Communication complexity. In *Advances in Computers*, volume 44, pages 331–360. Elsevier, 1997.
- [20] Kuan-Yun Lee. *New Information Inequalities with Applications to Statistics*. PhD thesis, UC Berkeley, 2022.
- [21] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR, 2017.
- [22] C Radhakrishna Rao. Information and the accuracy attainable in the estimation of statistical parameters. In *Breakthroughs in Statistics: Foundations and basic theory*, pages 235–247. Springer, 1992.
- [23] Ananda Theertha Suresh, Felix X. Yu, Sanjiv Kumar, and H. Brendan McMahan. Distributed mean estimation with limited communication. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70, ICML’17*, page 3329–3337. JMLR.org, 2017.
- [24] A. Tsybakov. *Introduction to Nonparametric Estimation*. Springer-Verlag, 2008.
- [25] Alexandre B Tsybakov. Introduction to nonparametric estimation, 2009. URL <https://doi.org/10.1007/b13794>. Revised and extended from the, 9(10), 2004.
- [26] Harry L Van Trees. *Detection, estimation, and modulation theory, part I: detection, estimation, and linear modulation theory*. John Wiley & Sons, 2004.
- [27] Yuchen Zhang, John C. Duchi, Michael I Jordan, and Martin J Wainwright. Information-theoretic lower bounds for distributed statistical estimation with communication constraints. In *Advances in Neural Information Processing Systems*, pages 2328–2336, 2013.
- [28] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 21st ACM Conference on Computer and Communications Security*, Scottsdale, Arizona, 2014. URL <https://arxiv.org/abs/1407.6981>.