Robust Testing and Estimation under Manipulation Attacks

Jayadev Acharya [*] [1] **Ziteng Sun** [*] [1] **Huanyu Zhang** [*] [1]

Abstract

We study robust testing and estimation of discrete distributions in the strong contamination model. We consider both the "centralized setting" and the "distributed setting with information constraints" including communication and local privacy (LDP) constraints. Our technique relates the strength of manipulation attacks to the earth-mover distance using Hamming distance as the metric between messages (samples) from the users. In the centralized setting, we provide optimal error bounds for both learning and testing. Our lower bounds under local information constraints build on the recent lower bound methods in distributed inference. In the communication constrained setting, we develop novel algorithms based on random hashing and an ℓ_1/ℓ_1 isometry.

1. Introduction

Data from users form the backbone of modern distributed learning systems such as federated learning (Kairouz et al., 2019). Two of the key aspects of such large-scale distributed systems that make inference tasks challenging are

(i) *information constraints* at the users (e.g., preserving privacy, bandwidth limitations), and
(ii) *γ-manipulation attacks* where an adversary has complete control over a γ fraction of the users.

An extreme example is when malicious users are deliberately injected to disrupt the system. Note that when there are only manipulation attacks but no information constraints, the setting is equivalent to the robust inference where a fraction of the samples can be adversarially corrupted.

(Cheu et al., 2021) initiated the study of manipulation attacks under local differential privacy (LDP), thereby considering the practically important setting where both of the challenges above exist simultaneously. Motivated by their work, we further study manipulation attacks for inference on discrete distributions both with and without information constraints.

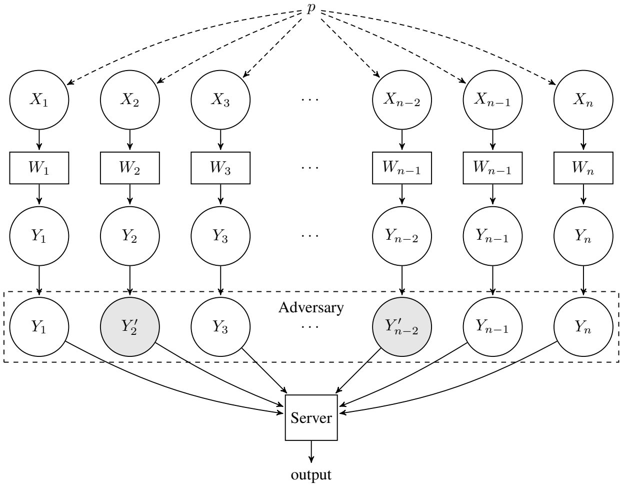

Figure 1. The information-constrained distributed model with manipulation attack. The shaded messages are manipulated.

Problem setup. Let $\triangle_k$ be the simplex of all distributions over a discrete domain $\mathcal{X}$ of size k (wlog let $\mathcal{X} = [k] := \{1, \dots, k\}$), and $X^n := (X_1, \dots, X_n)$ be n independent samples from an unknown $p \in \triangle_k$ which are distributed across n users. Each user i then sends a message $Y_i \in \mathcal{Y}$ based on X_i according to a pre-specified communication protocol Π to the server.

An adversary has access to Π and observes the *intended* messages $Y^n := (Y_1, \dots, Y_n)$. It *then* chooses a set $C \subset [n]$ with $|C| \le m{:=}\gamma n$ and performs an attack $M_C : \mathcal{Y}^n \to \mathcal{Y}^n$ as follows: for each $i \in C$, it can change Y_i to an arbitrary Y_i'. The output of this attack is $Z^n{:=}(Z_1, \dots, Z_n) = M_C(Y^n)$, which satisfies $Z_i = Y_i$ if $i \notin C$, and $Z_i = Y_i'$ if $i \in C$. We call this a *γ-manipulation attack*. A central server observes Z^n (and has no knowledge about C or M_C) and has to solve an inference task on p. See Figure 1 for an overview of the model.

Remark 1. A natural question to ask is what happens if the adversary, in addition to Y^n, can also observe original samples X^n. The algorithms proposed in this paper all work with the same guarantee under this setting as well.

[*] Equal contribution [1] Electrical and Computer Engineering, Cornell University, Ithaca, USA. Correspondence to: Jayadev Acharya <acharya@cornell.edu>, Ziteng Sun <zs335@cornell.edu>, Huanyu Zhang <hz388@cornell.edu>.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021. Copyright 2021 by the author(s).

Moreover, the proposed attacks only use Y^n, and therefore any optimality result naturally holds under this stronger threat model as well.

Remark 2. The threat model is closer to the strong contamination model (Diakonikolas & Kane, 2019) considered in recent literature on robust statistics. This adversary is stronger than the setting of (Cheu et al., 2021) where it is only allowed to select C and choose Y_i's based on the protocol Π instead of the messages Y^n.

Communication protocol. We consider public-coin non-interactive protocols in this work. The users have access to public randomness U, which is independent of X^n. Based on U, user i chooses a channel W_i, which is a (possibly randomized) mapping described by

$$W_i(y \mid x) = \Pr\left(Y_i = y \mid X_i = x\right).$$

When the input distribution is p and the channel is W_i, the distribution of Y_i is

$$\Pr\left(Y_i = y\right) = \sum_x p(x) W_i(y \mid x) = \mathbb{E}_{X\sim p}\left[W_i(y \mid X)\right].$$

For a given set of channels W^n and input distribution p, let p^{Y^n} denote the output distribution of messages, and by independence of X_i's,

$$p^{Y^n}(y^n) = \prod_{i=1}^{n} \mathbb{E}_{X\sim p}\left[W_i(y_i \mid X)\right]$$

All users then send their messages $Y_i = W_i(X_i), i \in [n]$ to the server simultaneously. The adversary also observes U and can use this information in its attack (e.g., choose C dependent on U as well).

Information constraints. We model information constraints at the users by a set of channels $\mathcal{W}$ with input domain $[k]$. We illustrate this with two canonical examples, local differential privacy and communication constraints.

Local differential privacy (LDP). A channel $W : [k] \to \mathcal{Y} = \{0,1\}^*$ is ε-LDP if $\forall y \in \mathcal{Y}, \forall x, x' \in \mathcal{X}$,

$$W(y \mid x) \le e^{\varepsilon} W(y \mid x'),$$

which requires that the output distributions are close no matter what the input is, hence protecting the identity of the input. We use $\mathcal{W}_\varepsilon$ to denote the set of all ε-LDP channels.

Communication constraints. Let $\ell < \log k$, and $\mathcal{W}_\ell := \{W : [k] \to \mathcal{Y} = \{0,1\}^\ell\}$ be the set of channels that output ℓ-bit messages.

Inference tasks. We consider the fundamental tasks of distribution estimation (learning) and goodness-of-fit (identity testing), described below.

Distribution learning (DL). The goal is to design messaging schemes and an estimator $\widehat{p} : \mathcal{Y}^n \to \triangle_k$ for the underlying distribution p. The loss is measured in the expected total variation (TV) distance between $\widehat{p}$ and p, i.e., $\mathbb{E}_p\left[d_{\mathrm{TV}}(\widehat{p}(Z^n), p)\right]$, where the expectation is over the randomness of samples $X^n \sim p$ and the scheme. We wish to characterize the following minimax loss (risk) under manipulation attacks, where we design the best messaging schemes $W^n := (W_1, \ldots W_n) \in \mathcal{W}^n$ and estimator $\hat{p}$ for the worst distribution[1]:

$$R_{\mathrm{DL}}(k, n, \mathcal{W}, \gamma) := \inf_{\hat{p}, W^n} \sup_p \sup_{M_C : |C| \le \gamma n} \mathbb{E}\left[d_{\mathrm{TV}}(\widehat{p}(Z^n), p)\right].$$

Without any information constraints and without any manipulation (i.e., $\gamma = 0$), when the server observes X^n, the risk is known to be $\Theta(\sqrt{k/n})$ achieved by the empirical histogram.

Identity testing (IT). Let $q \in \triangle_k$ be a *known* reference distribution and $\alpha > 0$ be a distance parameter. The goal is to design $\mathcal{W}^n$ and a tester $\mathcal{T} : \mathcal{Y}^n \to \{\texttt{yes}, \texttt{no}\}$ such that under any γ-manipulation attack M_C,

$$\begin{aligned} &\Pr_p\left[\mathcal{T}(Z^n) = \texttt{yes}\right] > 0.9, \quad &&\text{if } p = q, \\ &\Pr_p\left[\mathcal{T}(Z^n) = \texttt{no}\right] > 0.9, \quad &&\text{if } d_{\mathrm{TV}}(p, q) \ge \alpha. \end{aligned} \tag{1}$$

In other words, with probability at least 0.9, we can test if $p = q$ or p is α-far in total variation distance from q. The minimax risk of IT under manipulation attacks is

$$\begin{aligned} &R_{\mathrm{IT}}(k, n, \mathcal{W}, \gamma) := \\ &\quad \inf\{\alpha \colon \forall q \in \triangle_k, \exists W^n, \mathcal{T}, \text{ s.t. (1) holds}\}, \end{aligned}$$

the smallest α for which we can test if a distribution is α-far from q. *Uniformity testing (UT)* refers to the testing problem where we restrict q to be $u[k]$, the uniform distribution over $[k]$, we denote the corresponding risk as $R_{\mathrm{UT}}(k, n, \mathcal{W}, \gamma)$. We also denote the smallest α such that (1) holds with probability β (instead of 0.9) by $R^{\beta}_{\mathrm{IT(UT)}}(k, n, \mathcal{W}, \gamma)$.

Without constraints and attacks, the risk is known to be $\Theta(k^{1/4}/\sqrt{n})$ (Paninski, 2008; Chan et al., 2014).

We now mention two special cases of our setting.

- When there are no information constraints (i.e., $\mathcal{W}$ contains any scheme), users can transmit $Y_i = X_i$, namely the samples can be sent as is. Then a γ-fraction of the samples are corrupted, reducing to the strong contamination model in robust estimation where a γ fraction of the samples are corrupted. We denote the rates as $R_{\mathrm{DL(IT)}}(k, n, \gamma)$, dropping $\mathcal{W}$ from the notation.

[1] Note here by definition, $\sup_{M_C : |C| \le \gamma n}$ should be inside the expectation since the attacker can observe the messages. However, since M_C is a function of Y^n, $\sup_{M_C}$ already covers all possible attacks. Hence both minimax formulations have the same quantity.

- When $\gamma = 0$, there is no manipulation and only information constraints are present, and we denote the rates by $R_{\text{DL(IT)}}(k, n, \mathcal{W})$.

Organization. We present our contributions and related work in Section 2 and 3 respectively. In Section 4, we establish our lower bound technique based on earth-mover distance (EMD). In Section 5 we establish tight risk bounds without manipulation attacks ($\gamma = 0$). In Section 6 we show bounds for manipulation attacks under information constraints.

2. Our contributions

Lower bounds from EMD. Since manipulation attacks can change a γ-fraction of the n messages, we characterize the difficulty of learning and testing under such attacks in terms of the earth-mover distance (EMD) between messages with Hamming distance as the metric, stated in Theorem 1. Using Le Cam's method, the lower bounds are provided in terms of EMD between distributions of messages from mixtures of sources (distributions), which is critical for obtaining tight bounds.

Robust learning and testing. Without information constraints, the server observes the true samples from p but with γ-fraction adversarially corrupted. For distribution learning the minimax risk is $\Theta(\sqrt{k/n} + \gamma)$. While this result is standard, we provide it for completeness in Corollary 3. For testing, the optimal risk is more involved. In Theorem 2 we show that when γ fraction of the samples are corrupted, the risk is $\Theta(k^{1/4}/\sqrt{n} + \gamma + \sqrt{k\gamma/n} + \sqrt{\gamma}\sqrt[4]{k/n})$ where the first term corresponds to the statistical rate proved in (Paninski, 2008; Valiant & Valiant, 2014; Diakonikolas et al., 2018). In particular, when $\gamma \gg \min\{1/\sqrt{k}, 1/\sqrt{n}\}$ the risk increases significantly compared to the uncorrupted case.

Manipulation attacks under information constraints. In Corollary 14, we provide a general lower bound for estimating and testing distributions under information constraints and a γ-fraction manipulation attack. The result builds on the recently developed framework for distributed inference in (Acharya et al., 2019b) and bounds the EMD between messages in terms of the trace norm of a channel information matrix (Definition 3).

Communication constraints. In Theorem 8 and 9, we establish risk bounds for distribution learning and testing under ℓ-bit communication constraints. We propose a protocol based on random hashing which matches the lower bound we prove up to logarithmic factors. Our bounds suggest that manipulation attacks are significantly stronger with communication constraints on the channels. More precisely, the error due to manipulation attack can be as large as $\tilde{\Theta}(\gamma\sqrt{k/2^{\ell}})$ compared to γ in the unconstrained setting. We also provide a robust testing algorithm under communication constraints based on an ℓ_1/ℓ_1 isometry in (Acharya et al., 2020a). However, the bounds only match the lower bounds for $\ell = O(1)$ or $\ell = \Theta(\log k)$. The testing bound in the unconstrained case suggests more effort is needed to study how communication constraints limit EMD. Closing this gap is an interesting future direction.

Privacy constraints. In Theorem 10 we prove a lower bound that matches the upper bounds provided in (Cheu et al., 2021) for both testing (up to a constant factor) and learning (up to logarithmic factor). We note that in (Cheu et al., 2021), a lower bound smaller by a logarithmic factor is proved under a weaker threat model, which is not directly comparable to our result.

The results are summarized in Table 1[2].

3. Related work

Without local information constraints, our work is related to the literature of robust statistical inference. Robust statistics has a long history (Huber, 2004). More recently, the interest focuses on designing computationally efficient robust algorithms in high-dimensional estimation (Diakonikolas et al., 2016; Lai et al., 2016). See (Diakonikolas & Kane, 2019) for a survey. In (Diakonikolas et al., 2016; Chen et al., 2016), it is proved that for estimating a single Gaussian distribution, the risk due to adversarial attack is $\Theta(\gamma)$. For discrete distributions, a line of work (Qiao & Valiant, 2018; Chen et al., 2020; Jain & Orlitsky, 2020a;b) consider robust estimation in the distributed setting where each user contributes $s \gg 1$ samples. Compared to the result in Corollary 3, they show that in this case the risk due to manipulation can be much smaller than γ.

(Valiant & Valiant, 2011; Daskalakis et al., 2018) study tolerant identity testing where the goal is to test between $d_{\text{TV}}(p, q) \le \alpha/10$ and $d_{\text{TV}}(p, q) \ge \alpha$ for a reference distribution q. The optimal sample complexity has been established as $\Theta(k/\alpha^2 \log k)$. Suppose $\gamma = \alpha$, then a γ-robust identity tester is also a tolerant tester since with γ fraction of the users controlled, the adversary can simply change the distribution to another distribution within TV distance $\frac{\alpha}{2}$ with high probability. Theorem 2 shows that the robust setting is strictly harder by showing that $\Theta(k)$ samples are needed when γ and α are both constants. This is due to the richer class of attacks that the adversary can perform compared to the tolerant testing where the samples are still independent.

Robust identity testing Gaussian distributions without information constraints has been studied in (Diakonikolas

[2] All stated risk bounds are upper bounded by 1, which is omitted throughout the paper for simplicity.

Task	Constraint	Manipulation risk (UB)	Manipulation risk (LB)
DL	None	$O\left(\sqrt{\frac{k}{n}}+\gamma\right)$ (Folklore)	$\Omega\left(\sqrt{\frac{k}{n}}+\gamma\right)$ (Folklore, Corollary 3)
	ε-LDP	$\tilde{O}\left(\sqrt{\frac{k^2}{\varepsilon^2 n}}+\frac{\sqrt{k}}{\varepsilon}\cdot\gamma\right)$ (Cheu et al., 2021)	$\Omega\left(\sqrt{\frac{k^2}{\varepsilon^2 n}}+\frac{\sqrt{k}}{\varepsilon}\cdot\gamma\right)$ (†) (Theorem 10)
	ℓ-bit	$\tilde{O}\left(\sqrt{\frac{k^2}{2^\ell n}}+\sqrt{\frac{k}{2^\ell}}\cdot\gamma\right)$ (Theorem 8)	$\Omega\left(\sqrt{\frac{k^2}{2^\ell n}}+\sqrt{\frac{k}{2^\ell}}\cdot\gamma\right)$ (Theorem 8)
IT	None	$O\left(\frac{k^{\frac{1}{4}}}{\sqrt{n}}+\gamma+\sqrt{\frac{k\gamma}{n}}+\sqrt[4]{\frac{k\gamma^2}{n}}\right)$ (Theorem 2)	$\Omega\left(\frac{k^{\frac{1}{4}}}{\sqrt{n}}+\gamma+\sqrt{\frac{k\gamma}{n}}+\sqrt[4]{\frac{k\gamma^2}{n}}\right)$ (Theorem 2)
	ε-LDP	$O\left(\sqrt{\frac{k}{\varepsilon^2 n}}+\frac{\sqrt{k}}{\varepsilon}\cdot\gamma\right)$ (Cheu et al., 2021)	$\Omega\left(\sqrt{\frac{k}{\varepsilon^2 n}}+\frac{\sqrt{k}}{\varepsilon}\cdot\gamma\right)$ (†) (Theorem 10)
	ℓ-bit	$O\left(\sqrt{\frac{k}{2^{\ell/2}n}}+\sqrt{\frac{k}{2^\ell}}\cdot\left(\gamma+\sqrt{\frac{2^\ell\gamma}{n}}+\sqrt[4]{\frac{2^\ell\gamma^2}{n}}\right)\right)$ (Theorem 9)	$\Omega\left(\sqrt{\frac{k}{2^{\ell/2}n}}+\sqrt{\frac{k}{2^\ell}}\cdot\left(\gamma+\sqrt{\frac{2^\ell\gamma}{n}}\right)\right)$ (Theorem 9)

Table 1. Summary of results. For problems marked by (†), (Cheu et al., 2021) also provides lower bounds lower than the stated bounds by a logarithmic factor under a weaker threat model.

& Kane, 2020), where the contamination model is slightly different. In (Acharya et al., 2020c), identity testing of Gaussians is studied under communication constraints without manipulation attacks.

Without manipulation attacks, there is significant recent work interest in studying discrete distribution learning and testing in the distributed setting under information constraints. Optimal risks have been established under communication constraints (Han et al., 2018; Han et al., 2018; Acharya et al., 2020d; 2019b; 2020b) and LDP constraints (Duchi et al., 2013; Kairouz et al., 2016; Sheffet, 2017; Acharya et al., 2019a; 2020b).

In distributed learning systems, especially federated learning (Kairouz et al., 2019), manipulation attack is related to the so-called model poisoning attack (Bhagoji et al., 2019; Bagdasaryan et al., 2020), where the attacker has full control of a fraction of the users and can change model updates arbitrarily which doesn't have to obey the local training and messaging protocol. In these works, it is shown empirically that manipulation attacks can significantly outperform the classic data poisoning attack where the attacker can only insert data points to local users whose messages still follow the local messaging protocol.

4. Moving the earth: the power of manipulation attacks

We now characterize the power of manipulation attacks in terms of the earth-mover (a.k.a. Wasserstein) distance between the distributions of the messages at the output of the channels. We first recall EMD with Hamming metric.

Definition 1. Let Q_1 and Q_2 be distributions over $\mathcal{Y}^n$ and $\pi(Q_1, Q_2)$ be the set of all couplings between Q_1 and Q_2. The earth-mover distance (EMD) between Q_1 and Q_2 is

$$d_{\mathrm{EM}}\left(Q_1, Q_2\right) := \inf_{Q\in\pi(Q_1,Q_2)} \mathbb{E}_{(Y^n_{(1)},Y^n_{(2)})\sim Q}\left[d_{\mathrm{Ham}}(Y^n_{(1)}, Y^n_{(2)})\right].$$

Note that a γ-manipulation attack can change $Y^n_{(1)} \in \mathcal{Y}^n$ to another sequence $Y^n_{(2)} \in \mathcal{Y}^n$ as long as $d_{\mathrm{Ham}}(Y^n_{(1)}, Y^n_{(2)}) \le \gamma n$. If Q_1 and Q_2 are distributions over length-n messages in $\mathcal{Y}^n$ with EMD at most $c\cdot\gamma n$, for some small constant c, the attack can effectively confuse sequences generated from Q_1 and Q_2. We formalize this intuition in Theorem 1 below. A key ingredient in the theorem below is to consider message distributions from a *mixture of input distributions*.

Let $q \in \triangle_k$ be some reference distribution and $\mathcal{P} \subset \triangle_k$ be a *finite* set such that for all $p \in \mathcal{P}$,

$$d_{\mathrm{TV}}(p, q) \ge \alpha. \tag{2}$$

Let $\tilde{p}$ be uniformly drawn from $\mathcal{P}$. Further, for a fixed $W^n \in \mathcal{W}^n$

$$\mathbb{E}\left[\tilde{p}^{Y^n}\right] = \frac{1}{|\mathcal{P}|}\left(\sum_{p\in\mathcal{P}} p^{Y^n}\right). \tag{3}$$

be the message distribution when the input distribution is uniformly chosen from $\mathcal{P}$.

Theorem 1. *Suppose $\mathcal{P}$ satisfies (2) for some q and $\mathbb{E}\left[\tilde{p}^{Y^n}\right]$ is as defined in (3). If for all $W^n \in \mathcal{W}^n$,*

$$d_{\mathrm{EM}}\left(\mathbb{E}\left[\tilde{p}^{Y^n}\right], q^{Y^n}\right) \le \frac{\gamma n}{2},$$

then for both distribution learning and testing,

$$R_{\mathrm{IT}}(k,n,\mathcal{W},\gamma) = \Omega(\alpha), \quad R_{\mathrm{DL}}(k,n,\mathcal{W},\gamma) = \Omega(\alpha).$$

Proof. We first show a reduction from testing to learning. Suppose there exists a distribution learning algorithm with risk $\alpha/20$ under any γ-manipulation attack. We can use this for testing as follows: (i) By Markov's inequality, we learn the distribution to output a $\widehat{p}$ such that with probability at least 0.9, $d_{\mathrm{TV}}(\widehat{p},p) \le \alpha/2$, and then (ii) test if $d_{\mathrm{TV}}(\widehat{p},q) \gtrless \alpha/2$ to perform testing with respect to q. This shows that $R_{\mathrm{DL}}(k,n,\mathcal{W},\gamma) \ge c \cdot R_{\mathrm{IT}}(k,n,\mathcal{W},\gamma)$, for some c. Therefore, we only need to prove that $R_{\mathrm{IT}}(k,n,\mathcal{W},\gamma) = \Omega(\alpha)$.

Fix $W^n \in \mathcal{W}^n$, we have $d_{\mathrm{EM}}\left(\mathbb{E}\left[\tilde{p}^{Y^n}\right], q^{Y^n}\right) \le \gamma n/2$.[3] By the existence of minimizer for optimal transport[4], there exists a randomized mapping F such that if $Y^n_{(1)} \sim \mathbb{E}\left[\tilde{p}^{Y^n}\right]$, then $Y^n_{(2)} \overset{D}{=} F(Y^n_{(1)}) \sim q^{Y^n}$ and $\mathbb{E}\left[d_{\mathrm{Ham}}(Y^n_1, F(Y^n_{(1)}))\right] \le \frac{\gamma n}{2}$. By Markov's inequality, we have

$$\Pr\left(d_{\mathrm{Ham}}(Y^n_{(1)}, F(Y^n_{(1)})) > \gamma n\right) \le \frac{1}{2}. \qquad (4)$$

Consider the following γ-manipulation attack M_C:

$$M_C(Y^n_{(1)}) = \begin{cases} F(Y^n_{(1)}), & \text{if } d_{\mathrm{Ham}}(Y^n_{(1)}, F(Y^n_{(1)})) \le \gamma n, \\ Y^n_{(1)}, & \text{if } d_{\mathrm{Ham}}(Y^n_{(1)}, F(Y^n_{(1)})) > \gamma n. \end{cases}$$

Then by (4), we have

$$d_{\mathrm{TV}}(M_C(Y^n_{(1)}), Y^n_{(2)}) = d_{\mathrm{TV}}(M_C(Y^n_{(1)}), F(Y^n_{(1)})) \le \frac{1}{2}.$$

Hence, if the attacker sends the true messages when the distribution is q, by Bayes risk, it is impossible to test between q and $\mathbb{E}\left[\tilde{p}^{Y^n}\right]$ with success probability at least 9/10. By applying Le Cam's two-point method, it is impossible to tell whether the unknown distribution p equals q, or comes from $\mathcal{P}$, which concludes the proof. □

With the main technique at hand, for each family of channels, we prove lower bounds by designing distributions that are separated in TV distances while the corresponding messages are close in earth-mover distance. We start with the unconstrained setting in Section 5 and turn to the constrained case in Section 6.

5. Robust identity testing and learning

We consider robust identity testing without information constraints, i.e., the server observes the raw samples, γ fraction of which are adversarially corrupted. We prove the following tight minimax rate.

Theorem 2.

$$R_{\mathrm{IT}}(k,n,\gamma) = \Theta\left(\frac{k^{1/4}}{\sqrt{n}} + \gamma + \sqrt{\frac{k\gamma}{n}} + \sqrt[4]{\frac{k\gamma^2}{n}}\right).$$

The first term is the statistical rate which is implied by the sample complexity bound in (Paninski, 2008). Our bound implies that when $\gamma > \min\{1/\sqrt{n}, 1/\sqrt{k}\}$, the risk due to manipulation can be significantly larger than the statistical risk. The upper bound is based on the ℓ_1-tester proposed in (Diakonikolas et al., 2018), which we present in Section 5.1. The lower bound is proved using the technique based on earth-mover distance developed in Section 4, provided in Section 5.2.

We get the next corollary for learning under γ-manipulation attacks without information constraints.

Corollary 3 (folklore)**.**

$$R_{\mathrm{DL}}(k,n,\gamma) = \Theta\left(\sqrt{\frac{k}{n}} + \gamma\right).$$

The upper bound is achieved by the empirical distribution. The first term is the standard risk without manipulation, and the second term follows from the second term in Theorem 2 and the reduction from testing to learning. We omit the details.

5.1. Upper bound for testing

Our upper bound proceeds in two stages. We will first reduce identity testing to uniformity testing, and then provide an algorithm for uniformity testing.

Reduction from identity to uniformity testing. For unconstrained distribution estimation, Theorem 1 of (Goldreich, 2016) showed that, up to constant factors, the risk for testing identity of any distribution is upper bounded by the risk of uniformity testing. We extend their argument to the γ-manipulation attack. In particular, we will show that

Lemma 4.

$$R_{\mathrm{IT}}(k,n,\gamma) \le 3R_{\mathrm{UT}}(6k,n,\gamma).$$

Proof. (Goldreich, 2016) showed that for any distribution q over $[k]$, there exists a randomized function $G_q : [k] \to [6k]$ such that if $X \sim q$, then $G_q(X) \sim u[6k]$, and if $X \sim p$ for a distribution with $d_{\mathrm{TV}}(p,q) \ge \alpha$, then $d_{\mathrm{TV}}(G_q(X), u[6k]) \ge \alpha/3$.

[3] Without loss of generality, we assume W^n is fixed since the adversary can observe public randomness U.

[4] Hamming distance is a lower semi-continuous cost function.

Let X^n be n samples from p, and Z^n be a γ-corrupted version of X^n. We then apply G_q independently to each of the Z_i to obtain a new sequence $G_q(Z^n){:=}(G_q(Z_1)\ldots G_q(Z_n))$. If $p=q$, $G_q(X^n)$ is distributed according to $u[6k]$ and $G_q(Z^n)$ is a γ-manipulated version of it. If $d_{\mathrm{TV}}(p,q)\geq\alpha$, $G_q(X^n)$ is distributed according to a distribution at least $\alpha/3$-far from $u[6k]$. Therefore, an algorithm for γ-robust uniformity testing with $\alpha'=\alpha/3$ can be used for α-identity testing with q. □

We now present an algorithm for γ-robust uniformity testing. Recall that Z^n is obtained upon perturbing γn samples from $X^n\sim p$. For $z\in[k]$, let $M_z(Z^n)$ be the number of appearances of z in Z^n. The TV distance between the empirical histogram of Z^n and uniform, which was used in (Diakonikolas et al., 2018),

$$S(Z^n){:=}\frac{1}{2}\cdot\sum_{z=1}^{k}\left|\frac{M_z(Z^n)}{n}-\frac{1}{k}\right| \tag{5}$$

will be used as our test statistic.

We now bound the difference in test statistic between Z^n and X^n when $d_{\mathrm{Ham}}(Z^n,X^n)\leq\gamma n$.

Lemma 5. *Suppose Z^n is obtained from X^n by manipulating at most γn samples, then*

$$|S(X^n)-S(Z^n)|\leq\min\left(\gamma,\frac{n\gamma}{k}\right).$$

Proof. For the first term, by the triangle inequality for any $a,b\in\mathbb{R}$, we have $||a|-|b||\leq|a-b|$. Using this in (5), we obtain,

$$|S(X^n)-S(Z^n)|\leq\frac{1}{2}\cdot\sum_{x=1}^{k}\left|\frac{M_x(X^n)-M_x(Z^n)}{n}\right|\leq\gamma,$$

where we used the fact that changing *one sample* from X^n changes $M_x(X^n)$ for at most two $x\in[k]$ each by at most one, and therefore $\sum_{x=1}^{k}|M_x(X^n)-M_x(Z^n)|\leq2\gamma n$.

We note that the second term is smaller than the first when $n<k$. When $n<k$,

$$\begin{aligned}&S(X^n)\\&=\frac{1}{2}\cdot\left(\sum_{x:M_x(Z^n)\geq1}\left(\frac{M_z(X^n)}{n}-\frac{1}{k}\right)+\sum_{x:M_x(X^n)=0}\frac{1}{k}\right)\\&=\frac{1}{2}\left(1-\left(1-\frac{\Phi_0(X^n)}{k}\right)+\frac{\Phi_0(X^n)}{k}\right)\\&=\frac{\Phi_0(X^n)}{k}.\end{aligned}$$

where $\Phi_0(X^n)$ is the number of symbols not appearing in X^n. Therefore,

$$|S(X^n)-S(Z^n)|\leq\frac{1}{k}\cdot|\Phi_0(X^n)-\Phi_0(Z^n)|\leq\frac{n\gamma}{k}.\ \square$$

Next we use the following result from (Diakonikolas et al., 2018), which shows a separation in the test statistic under $p=u[k]$ and $d_{\mathrm{TV}}(p,u[k])\geq\alpha$. Let $\mu(p){:=}\mathbb{E}_{X^n\sim p}\left[S(X^n)\right]$ be the expectation of the statistic of the original samples.

Lemma 6 ((Diakonikolas et al., 2018), Lemma 4)**.** *Let X^n be i.i.d. samples from p over $[k]$. For every $\beta\in(0,1)$, there exist constants c_1,c_2 such that if $\alpha\geq c_1\cdot\frac{k^{\frac{1}{4}}}{\sqrt{n}}$, then with probability at least $1-\beta$,*

1. *when $p=u[k]$,*

$$S(X^n)-\mu(u[k])<\tfrac{9}{10}c_2\alpha^2\min\left\{\tfrac{n^2}{k^2},\sqrt{\tfrac{n}{k}},\tfrac{1}{\alpha}\right\},$$

2. *when $d_{\mathrm{TV}}(p,u[k])\geq\alpha$,*

$$S(X^n)-\mu(u[k])>\tfrac{11}{10}c_2\alpha^2\min\left\{\tfrac{n^2}{k^2},\sqrt{\tfrac{n}{k}},\tfrac{1}{\alpha}\right\}.$$

Our test for uniformity is the following:

$$\mathcal{T}=\begin{cases}\texttt{yes} & \text{if } S(Z^n)-\mu(u[k])\leq c_2\alpha^2\min\left\{\frac{n^2}{k^2},\sqrt{\frac{n}{k}},\frac{1}{\alpha}\right\}\\ \texttt{no} & \text{otherwise.}\end{cases} \tag{6}$$

By Lemma 5, when $\alpha\geq\frac{10}{c_2}\cdot\left(\gamma+\sqrt{\frac{k\gamma}{n}}+\sqrt[4]{\frac{k\gamma^2}{n}}\right)$,

$$|S(X^n)-S(Z^n)|\leq\frac{1}{10}\cdot c_2\alpha^2\min\left\{\frac{n^2}{k^2},\sqrt{\frac{n}{k}},\frac{1}{\alpha}\right\}.$$

Setting $\beta=1/10$ in Lemma 6 shows that the test in (6) solves the uniformity testing problem.

Remark 3. Note that the proof shows that for any constant failure probability β, the risk for robust identity testing is the same as that in Theorem 2 up to a constant factor, i.e., for any β, there is a constant $c(\beta)$, such that

$$R_{\mathrm{IT}}^{\beta}(k,n,\mathcal{W},\gamma)\leq c(\beta)\left(\frac{k^{\frac{1}{4}}}{\sqrt{n}}+\gamma+\sqrt{\frac{k\gamma}{n}}+\sqrt[4]{\frac{k\gamma^2}{n}}\right).$$

This will be important when we consider error boosting in the proof of Theorem 12.

5.2. Lower bound

In uniformity testing we have $q=u[k]$. We will use $\mathcal{P}$ to be the following class of $2^{k/2}$ distributions from (Paninski, 2008) indexed by $\mathbf{z}\in\{\pm1\}^{k/2}$, i.e., $\mathcal{P}=\{p_{\mathbf{z}}:\mathbf{z}\in\{\pm1\}^{k/2}\}$, where

$$p_{\mathbf{z}}(2i-1)=\frac{1+\mathbf{z}_i\cdot2\alpha}{k},\ p_{\mathbf{z}}(2i)=\frac{1-\mathbf{z}_i\cdot2\alpha}{k}. \tag{7}$$

Note that for any $\mathbf{z}\in\{\pm1\}^{k/2}$, $d_{\mathrm{TV}}(p_{\mathbf{z}},u[k])=\alpha$. The following lemma, proved in (Acharya et al., 2018) characterizes the earth-mover distance between $\mathbb{E}_{Z\sim u\{\pm1\}^{k/2}}\left[p_Z^n\right]$ and $u[k]^n$.

Lemma 7 ((Acharya et al., 2018) Lemma 7)**.**

$$
\begin{aligned}
&d_{\mathrm{EM}}\left(\mathbb{E}_{Z\sim u\{\pm 1\}^{k/2}}\left[p_Z^n\right], u[k]^n\right) \\
&= O\left(n \cdot \min\left\{\frac{n\alpha^2}{k}, \frac{\sqrt{n}\alpha^2}{\sqrt{k}}, \alpha\right\}\right).
\end{aligned}
$$

Note in the centralized case $Y^n = X^n$, hence $p^{Y^n} = p^n$. Plugging Lemma 7 in Theorem 1, we get the lower bound by setting the EMD to be $\gamma n/2$ and solving for α.

6. Robust constrained inference

In this section we consider learning and testing under communication and LDP constraints. We first state the results, then in Section 6.1 and 6.2 establish the upper bounds and finally close with lower bounds in Section 6.3.

Communication constraints. For distribution learning under ℓ-bit communication constraints, we establish the following risk bound, which is optimal up to logarithmic factors.

Theorem 8.

$$
R_{\mathrm{DL}}(k, n, \mathcal{W}_\ell, \gamma) = \tilde{\Theta}\left(\sqrt{\frac{k^2}{n(2^\ell \wedge k)}} + \gamma\sqrt{\frac{k}{2^\ell \wedge k}}\right).
$$

The first term is the risk under communication constraints without manipulation attacks (Han et al., 2018; Acharya et al., 2020d). The second term shows that if $\ell < \log k$, manipulation attack increases the risk by $\Theta\left(\gamma\sqrt{\frac{k}{2^\ell}}\right)$ compared to the the increase of $\Theta(\gamma)$ in the unconstrained setting (Corollary 3).

For identity testing, we obtain the following risk bounds.

Theorem 9. *Suppose $\ell \le \log k$,*

$$
\begin{aligned}
&R_{\mathrm{IT}}(k, n, \mathcal{W}_\ell, \gamma) = \\
&O\left(\sqrt{\frac{k}{2^{\ell/2} n}} + \sqrt{\frac{k}{2^\ell}} \cdot \left(\gamma + \sqrt{\frac{2^\ell \gamma}{n}} + \sqrt[4]{\frac{2^\ell \gamma^2}{n}}\right)\right),
\end{aligned}
$$

and

$$
R_{\mathrm{IT}}(k, n, \mathcal{W}_\ell, \gamma) = \Omega\left(\sqrt{\frac{k}{2^{\ell/2} n}} + \sqrt{\frac{k}{2^\ell}} \cdot \left(\gamma + \sqrt{\frac{2^\ell \gamma}{n}}\right)\right).
$$

The first term above is the risk of testing under $\mathcal{W}_\ell$ without information constraints. The risk due to manipulation attack in the upper bound is increased by a factor of $\sqrt{\frac{k}{2^\ell}}$ compared to the unconstrained case in Theorem 2 . We remark that the upper and lower bounds above match (up to constant factors) for $\ell = \Theta(1)$ and for $\ell = \Theta(\log k)$ (when $\ell = \log k$, it matches the risk for unconstrained testing in Theorem 2).

We believe that the upper bound is tight and proving a better lower bound is an interesting future work.

Local privacy constraints. We establish the following lower bounds for learning and testing under ε-LDP.

Theorem 10. *Suppose $\varepsilon = O(1)$,*

$$
R_{\mathrm{DL}}(k, n, \mathcal{W}_\varepsilon, \gamma) = \Omega\left(\sqrt{\frac{k^2}{n\varepsilon^2}} + \gamma\sqrt{\frac{k}{\varepsilon^2}}\right),
$$

$$
R_{\mathrm{IT}}(k, n, \mathcal{W}_\varepsilon, \gamma) = \Omega\left(\sqrt{\frac{k}{n\varepsilon^2}} + \gamma\sqrt{\frac{k}{\varepsilon^2}}\right).
$$

We note that (Cheu et al., 2021) designs algorithms that achieve the bounds above up to a constant factor for testing and up to logarithmic factors for learning. However, their lower bounds are a logarithmic factor smaller than Theorem 10 under a weaker threat model, making the bounds incomparable.

6.1. Distribution estimation with ℓ bits under manipulation

Without loss of generality, we assume $\ell < \log k$. We now present a scheme based on random hasing that achieves the upper bound for learning with the rate in Theorem 8. Random hashing has been previously used for sparse estimation under communication constraints (Acharya et al., 2021).

Definition 2. A random mapping $h : [k] \to [2^\ell]$ is a random hashing function if $\forall x \in [k], y \in [2^\ell], \Pr\left(h(x) = y\right) = \frac{1}{2^\ell}$.

Let $\mathcal{T}$ be the set of all $k \times 2^\ell$ binary matrices that have exactly one '1' in each row. A random hashing function h is equivalent to a T_h drawn uniformly at random from $\mathcal{T}$ with the correspondence

$$
T_h(x, y) = \mathbf{1}\left\{h(x) = y\right\}.
$$

Now for any fixed $y \in [2^\ell]$, the yth column of T_h has each entry as an independent $\mathrm{Ber}(1/2^\ell)$ random variable.

We describe the protocol and the estimator below.

1. Using randomness U users obtain independent random hashing functions $h_1, \ldots, h_n$ and send
$$Y_i = h_i(X_i).$$
2. Upon receiving the manipulated samples $Z^n \in [2^\ell]^n$, the server outputs
$$\widehat{p}(Z^n) = \frac{2^\ell}{n(2^\ell - 1)}\left(\sum_{i=1}^{n} T_{h_i}(\cdot, Z_i) - \frac{n}{2^\ell}\right).$$

Without manipulation attacks, when the server receives Y^n, it has been shown in (Acharya et al., 2021) that

$$
\mathbb{E}\left[d_{\mathrm{TV}}(\widehat{p}(Y^n), p)\right] = O\left(\sqrt{\frac{k^2}{2^\ell n}}\right).
$$

Note that in Theorem 8, the second term becomes larger than one when $\gamma > c \cdot \sqrt{2^\ell/k}$. We therefore focus on $\gamma < \sqrt{2^\ell/k}$. By the triangle inequality, it suffices to bound $\mathbb{E}\left[d_{\mathrm{TV}}(\widehat{p}(Y^n), \widehat{p}(Z^n))\right]$. Let C be the set of samples that are manipulated, and hence $|C| \le \gamma n = m$. Therefore,

$$
\begin{aligned}
& d_{\mathrm{TV}}(\widehat{p}(Y^n), \widehat{p}(Z^n)) \\
&= \frac{2^\ell}{2n(2^\ell - 1)} \left\| \sum_{i \in C} (T_{h_i}(\cdot, Y_i) - T_{h_i}(\cdot, Z_i)) \right\|_1 \\
&\le \frac{1}{n} \max_{|C'|=m} \max_{y_i, z_i \in [2^\ell]} \left\| \sum_{i \in C'} (T_{h_i}(\cdot, y_i) - T_{h_i}(\cdot, z_i)) \right\|_1 \qquad (8) \\
&= \frac{1}{n} \max_{|C'|=m} \max_{y_i, z_i \in [2^\ell]} \max_{v \in \{\pm 1\}^k} v^T \sum_{i \in C'} (T_{h_i}(\cdot, y_i) - T_{h_i}(\cdot, z_i)), \qquad (9)
\end{aligned}
$$

where (8) follows by maximizing over C, and (9) holds since for $u \in \mathbb{R}^k$, $\|u\|_1 = \max_{v \in \{\pm 1\}^k} v^T u$.

Recall that for $\forall i \in [k]$ and $y_i, z_i \in \mathcal{Y}$, $T_{h_i}(\cdot, y_i)$ and $T_{h_i}(\cdot, z_i)$ are both k-dimensional binary vectors with each coordinate an independent $\mathrm{Ber}(1/2^\ell)$. Then for any $C' \subset [n]$, with $|C'| = m$, $x \in [k]$, $y_i, z_i \in [2^\ell]$, and $v \in \{\pm 1\}^k$, $v^T \sum_{i \in C'} T_{h_i}(\cdot, y_i) - T_{h_i}(\cdot, z_i))$ is distributed as the difference between two $\mathrm{Binom}(km, 1/2^\ell)$ random variables[5] since h_i's are independently generated.

Hence, by Chernoff bound (multiplicative form) and union bound, we have $\forall \eta \in (0, \sqrt{km/2^\ell})$,

$$
\Pr\left(v^T \sum_{i \in C'} (T_{h_i}(\cdot, y_i) - T_{h_i}(\cdot, z_i)) > 2\sqrt{\frac{km}{2^\ell}}\eta \right) \le 2e^{-\frac{\eta^2}{3}}. \qquad (10)
$$

Taking union bound over $\binom{n}{m}$ subsets C', $(2^\ell)^m \times (2^\ell)^m$ possible choices of $\{y_i, z_i\}_{i \in C'}$, and 2^k choices of $v \in \{\pm 1\}^k$, by (9) and (10), we have

$$
\Pr\left(d_{\mathrm{TV}}(\widehat{p}(Y^n), \widehat{p}(Z^n)) > \frac{2}{n}\sqrt{\frac{km}{2^\ell}}\eta \right) \le \frac{2\binom{n}{m}(2^\ell)^{2m} 2^k}{e^{\frac{\eta^2}{3}}}.
$$

For $\eta = \sqrt{6(k + 2m\ell \log 2 + m \log n + \log(2^\ell/\gamma^2 k))}$, we have[6],

$$
\Pr\left(d_{\mathrm{TV}}(\widehat{p}(Y^n), \widehat{p}(Z^n)) > \frac{2}{n}\sqrt{\frac{km}{2^\ell}}\eta \right) \le \gamma\sqrt{\frac{k}{2^\ell}}.
$$

Hence using $n = m/\gamma$, we have

$$
\mathbb{E}\left[d_{\mathrm{TV}}(\widehat{p}(Y^n), \widehat{p}(Z^n))\right] = \tilde{O}\left(\gamma\sqrt{\frac{k}{2^\ell}} + \sqrt{\frac{k^2}{2^\ell n}} \right).
$$

6.2. ℓ-bit identity testing under manipulation

We now establish the upper bound in Theorem 9. We first reduce the problem to an unconstrained testing problem over a domain of size $[2^\ell]$ that can be represented using ℓ bits and then invoke our bounds from Theorem 2 for robust identity testing without information constraints.

For the reduction, we use the domain compression technique proposed in (Acharya et al., 2020a) to compress the observed samples to a smaller domain of size 2^ℓ. Moreover, the protocol preserves the TV distance between any pair of distributions up to a factor of $\Omega(\sqrt{2^\ell/k})$ with a constant probability. More precisely, we will use the following lemma from (Acharya et al., 2020a).

Lemma 11 (Theorem 3.2 (Acharya et al., 2020a))**.** *For any $\ell < \log k$, there exists a mapping $\varphi : \{0,1\}^* \times [k] \to [2^\ell]$ and universal constants c_1 and c_2 such that $\forall p, q \in \triangle_k$ with $d_{\mathrm{TV}}(p, q) \ge \alpha$, we have*

$$
\Pr_U\left[d_{\mathrm{TV}}(\varphi(U, p), \varphi(U, q)) \ge c_1 \alpha \cdot \sqrt{\frac{2^\ell}{k}} \right] \ge c_2,
$$

where U is a public random string and with a slight abuse of notation we denote by $\varphi(U, p)$ the distribution of $\varphi(U, X)$ when $X \sim p$.

Using this lemma, our testing scheme is the following.

1. Divide users into $N = \lceil \log_{1-c_2/2}(1/10) \rceil$ disjoint batches of equal size, where c_2 is the constant in Lemma 11. For each batch $B_j, j \in [N]$, generate an independent public random string U_j.
2. Each user $i \in B_j$ sends the ℓ-bit message $Y_i = \varphi(U_j, X_i)$.
3. For $j \in [N]$, let $Z^{(B_j)}$ be the messages received from users in B_j. Perform the robust testing algorithm in Section 5.1 with alphabet size 2^ℓ and distance $R^{\beta}_{\mathrm{IT}}(2^\ell, n/N, N\gamma)$ where $\beta = \min\{c_2/2, 1 - \sqrt[N]{9/10}\}$.
4. If all tests output `yes`, output `yes`, else, output `no`.

We now analyze the algorithm.

In the null case, when $p = q$, $\varphi(U, p) = \varphi(U, q)$, the test in each batch outputs `yes` with probability at least $1 - \beta$ (see Remark 3). Since the batches are disjoint, all tests output `yes` with probability at least $(1-\beta)^N \ge 9/10$ since $\beta \le 1 - \sqrt[N]{9/10}$.

Now in the alternate case, suppose that

$$
d_{\mathrm{TV}}(p, q) \ge \frac{1}{c_1} \cdot \sqrt{\frac{k}{2^\ell}} \cdot R^{\beta}_{\mathrm{IT}}(2^\ell, n/N, N\gamma).
$$

Then with probability at least c_2 over the randomness of U, we have

$$
d_{\mathrm{TV}}(\varphi(U, p), \varphi(U, q)) \ge R^{\beta}_{\mathrm{IT}}\left(2^\ell, \frac{n}{N}, \gamma N\right).
$$

[5] It is possible that these two binomials are correlated. However, the union bound used after this still holds.

[6] $\ell < \log k$ implies the choice of $\eta \in (0, \sqrt{km/2^\ell})$.

Conditioned on this event, by Theorem 2, we have the test in this batch outputs no with probability at least $1 - c_2/2$ since $\beta < c_2/2$. By disjointness of the batches, and the union bound we have at least one of the tests output no with probability at least $1 - (1 - c_2/2)^N \geq 9/10$ by the choice of N.

From Remark 3 we know that $R^{\beta}_{\text{IT}}(2^{\ell}, n/N, N\gamma)$ is at most larger than $R_{\text{IT}}(2^{\ell}, n, \gamma)$ by a multiplicative constant, which gives the following reduction.

Lemma 12.

$$R_{\text{IT}}(k, n, \mathcal{W}_{\ell}, \gamma) = O\left(\sqrt{\frac{k}{2^{\ell} \wedge k}} \cdot R_{\text{IT}}(2^{\ell} \wedge k, n, \gamma)\right).$$

To obtain the upper bound in Theorem 9, we only need to plug in the bound of robust testing without information constraints from Theorem 2.

6.3. Lower bounds by χ^2-contraction

In order to establish the lower bounds, by Theorem 1, it is sufficient to establish upper bounds on the EMD of messages Y^n induced by a suitably chosen mixture of distributions from that induced by the uniform distribution.

In particular, we will relate the EMD under information constraints to the channel information matrices of the allowed channels, which describes how the channel can exploit the structure of the set of distributions in (7) to solve inference tasks.

Definition 3 (Channel Information Matrix)**.** For a channel $W : [k] \rightarrow \mathcal{Y}$, the channel information matrix of W, denoted by $H(W)$, is a $(k/2) \times (k/2)$ matrix and $\forall i_1, i_2 \in [k/2]$, $H(W)(i_1, i_2)$:=

$$\sum_{y \in \mathcal{Y}} \frac{(W(y \,|\, 2i_1 - 1) - W(y \,|\, 2i_1))(W(y \,|\, 2i_2 - 1) - W(y \,|\, 2i_2))}{\sum_{x \in [k]} W(y \mid x)}.$$

We will establish the following upper bounds on the EMD under local information constraints.

Lemma 13. *For $\mathcal{P}$ defined in (7), and $\tilde{p}$ be uniformly drawn from $\mathcal{P}$, and for any $W^n \in \mathcal{W}^n$, let $\mathbb{E}\left[\tilde{p}^{Y^n}\right] = \frac{1}{2^{k/2}}\left(\sum_{p \in \mathcal{P}} p^{Y^n}\right)$ be the message distribution under a uniform mixture and channels W^n. Then,*

$$d_{\text{EM}}\left(\mathbb{E}\left[p_Z^{Y^n}\right], u[k]^{Y^n}\right) \leq 2n\alpha\sqrt{\frac{\max_{W \in \mathcal{W}} \|H(W)\|_*}{k}}.$$

We will use this lemma with Theorem 1 to obtain the following lower bound for robust inference.

Corollary 14.

$$R_{\text{IT(DL)}}(k, n, \mathcal{W}, \gamma) = \Omega\left(\gamma\sqrt{\frac{k}{\max_{W \in \mathcal{W}} \|H(W)\|_*}}\right),$$

where $\|\cdot\|_*$ denotes the trace norm of a matrix.

Under privacy and communication constraints, we have $\max_{W \in \mathcal{W}_{\varepsilon}} \|H(W)\|_* \leq \varepsilon^2$ and $\max_{W \in \mathcal{W}_{\ell}} \|H(W)\|_* \leq 2^{\ell}$, which are proved in (Acharya et al., 2019b). Using Corollary 14, we obtain the corresponding terms in the lower bound part of Theorem 8, the 9, and 10. The first terms in these bounds are the lower bounds without manipulation attacks and are proved in (Acharya et al., 2019b; Han et al., 2018; Duchi et al., 2013) respectively.

Remark 4. Lower bounds without information constraints are automatically lower bounds of the constrained inference. Therefore, we get the lower bound of $\Omega\left(\sqrt{\frac{k\gamma}{n}}\right)$ in Theorem 9 from Theorem 2.

Now it is enough to prove Lemma 13.

Proof. We will use the same lower bound construction stated in (7). We bound the earth-mover distance using the naive coupling for length-n independent sequences which is equal to the sum of TV distances on each entry. Then we can relate the TV distances to the χ^2-divergence bounds proved in (Acharya et al., 2019b), stated below.

Lemma 15 (Theorem IV.11 (Acharya et al., 2019b))**.** *Let $\|\cdot\|_*$ denote the trace norm of a matrix,*

$$\mathbb{E}_{Z \sim u\{\pm 1\}^{k/2}}\left[d_{\chi^2}(W \cdot p_Z, W \cdot u[k])\right] = \frac{8\alpha^2}{k}\|H(W)\|_*,$$

where $\forall q$, $W \cdot q$ denotes the distribution of the message Y given the input $X \sim q$, and p_Z is defined in (7).

Let $Z \sim u\{\pm 1\}^{k/2}$. Then we have

$$\begin{aligned}
& d_{\text{EM}}\left(\mathbb{E}\left[p_Z^{Y^n}\right], u[k]^{Y^n}\right) \\
\leq\; & \mathbb{E}\left[d_{\text{EM}}\left(p_Z^{Y^n}, u[k]^{Y^n}\right)\right] && \text{(Convexity)} \\
\leq\; & \mathbb{E}\left[\sum_{i=1}^{n} d_{\text{TV}}(W_i \cdot p_Z, W_i \cdot u[k])\right] && \text{(Naive coupling)} \\
\leq\; & \sum_{i=1}^{n} \mathbb{E}\left[\sqrt{\frac{1}{2} d_{\chi^2}(W_i \cdot p_Z, W_i \cdot u[k])}\right] && \text{(Pinkser's Inequality)} \\
\leq\; & \sum_{i=1}^{n} \sqrt{\mathbb{E}\left[\frac{1}{2} d_{\chi^2}(W_i \cdot p_Z, W_i \cdot u[k])\right]} && \text{(Concavity)} \\
\leq\; & 2n\alpha\sqrt{\frac{\max_{W \in \mathcal{W}} \|H(W)\|_*}{k}}. && \text{(Lemma 15)}
\end{aligned}$$

□

7. Acknowledgement

This research is supported by the NSF through grants CCF-1846300 (CAREER), CCF-1815893.

References

Acharya, J., Sun, Z., and Zhang, H. Differentially private testing of identity and closeness of discrete distributions. In *Advances in Neural Information Processing Systems*, pp. 6879–6891, 2018.

Acharya, J., Canonne, C. L., Freitag, C., and Tyagi, H. Test without trust: Optimal locally private distribution testing. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 2067–2076, 2019a.

Acharya, J., Canonne, C. L., and Tyagi, H. Inference under information constraints: Lower bounds from chi-square contraction. *Proceedings of Machine Learning Research vol*, 99:1–15, 2019b.

Acharya, J., Canonne, C. L., Han, Y., Sun, Z., and Tyagi, H. Domain compression and its application to randomness-optimal distributed goodness-of-fit. In *Conference on Learning Theory*, pp. 3–40. PMLR, 2020a.

Acharya, J., Canonne, C. L., Liu, Y., Sun, Z., and Tyagi, H. Interactive inference under information constraints. *arXiv preprint arXiv:2007.10976*, 2020b.

Acharya, J., Canonne, C. L., and Tyagi, H. Distributed signal detection under communication constraints. In *Conference on Learning Theory*, pp. 41–63. PMLR, 2020c.

Acharya, J., Canonne, C. L., and Tyagi, H. Inference under information constraints ii: Communication constraints and shared randomness. *IEEE Transactions on Information Theory*, 66(12):7856–7877, 2020d.

Acharya, J., Kairouz, P., Liu, Y., and Sun, Z. Estimating sparse discrete distributions under privacy and communication constraints. In Feldman, V., Ligett, K., and Sabato, S. (eds.), *Proceedings of the 32nd International Conference on Algorithmic Learning Theory*, volume 132 of *Proceedings of Machine Learning Research*, pp. 79–98. PMLR, 16–19 Mar 2021. URL http://proceedings.mlr.press/v132/acharya21b.html.

Bagdasaryan, E., Veit, A., Hua, Y., Estrin, D., and Shmatikov, V. How to backdoor federated learning. In *International Conference on Artificial Intelligence and Statistics*, pp. 2938–2948. PMLR, 2020.

Bhagoji, A. N., Chakraborty, S., Mittal, P., and Calo, S. Analyzing federated learning through an adversarial lens. In *International Conference on Machine Learning*, pp. 634–643. PMLR, 2019.

Chan, S.-O., Diakonikolas, I., Valiant, G., and Valiant, P. Optimal algorithms for testing closeness of discrete distributions. In *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '14, pp. 1193–1203, Philadelphia, PA, USA, 2014. SIAM.

Chen, M., Gao, C., and Ren, Z. A general decision theory for huber's ϵ-contamination model. *Electronic Journal of Statistics*, 10(2):3752–3774, 2016.

Chen, S., Li, J., and Moitra, A. Learning structured distributions from untrusted batches: Faster and simpler. *Advances in Neural Information Processing Systems*, 2020.

Cheu, A., Smith, A., and Ullman, J. Manipulation attacks in local differential privacy. In *2021 2021 IEEE Symposium on Security and Privacy (SP)*, pp. 1–18, Los Alamitos, CA, USA, may 2021. doi: 10.1109/SP40001.2021.00001.

Daskalakis, C., Kamath, G., and Wright, J. Which distribution distances are sublinearly testable? In *Proceedings of the 29th Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '18, pp. 2747–2764, Philadelphia, PA, USA, 2018. SIAM.

Diakonikolas, I. and Kane, D. M. Recent advances in algorithmic high-dimensional robust statistics. *arXiv preprint arXiv:1911.05911*, 2019.

Diakonikolas, I. and Kane, D. M. The sample complexity of robust covariance testing. *arXiv e-prints*, pp. arXiv–2012, 2020.

Diakonikolas, I., Kamath, G., Kane, D. M., Li, J., Moitra, A., and Stewart, A. Robust estimators in high dimensions without the computational intractability. In *Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '16, pp. 655–664, Washington, DC, USA, 2016. IEEE Computer Society.

Diakonikolas, I., Gouleakis, T., Peebles, J., and Price, E. Sample-optimal identity testing with high probability. In *Proceedings of the 45th International Colloquium on Automata, Languages, and Programming*, ICALP '18, pp. 41:1–41:14, 2018.

Duchi, J. C., Jordan, M. I., and Wainwright, M. J. Local privacy and statistical minimax rates. In *Proceedings of the 54st Annual IEEE Symposium on Foundations of Computer Science*, FOCS '13, pp. 429–438. IEEE, 2013.

Goldreich, O. The uniform distribution is complete with respect to testing identity to a fixed distribution. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 23, 2016.

Han, Y., Mukherjee, P., Özgür, A., and Weissman, T. Distributed statistical estimation of high-dimensional and nonparametric distributions. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pp. 506–510, 2018. doi: 10.1109/ISIT.2018.8437818.

Han, Y., Özgür, A., and Weissman, T. Geometric lower bounds for distributed parameter estimation under communication constraints. In *Conference On Learning Theory*, pp. 3163–3188. PMLR, 2018.

Huber, P. J. *Robust statistics*, volume 523. John Wiley & Sons, 2004.

Jain, A. and Orlitsky, A. A general method for robust learning from batches. *Advances in Neural Information Processing Systems*, 33, 2020a.

Jain, A. and Orlitsky, A. Optimal robust learning of discrete distributions from batches. In *International Conference on Machine Learning*, pp. 4651–4660. PMLR, 2020b.

Kairouz, P., Bonawitz, K., and Ramage, D. Discrete distribution estimation under local privacy. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, ICML'16, pp. 2436–2444, 2016.

Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R. G. L., Rouayheb, S. E., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P. B., Gruteser, M., Harchaoui, Z., He, C., He, L., Huo, Z., Hutchinson, B., Hsu, J., Jaggi, M., Javidi, T., Joshi, G., Khodak, M., Konečný, J., Korolova, A., Koushanfar, F., Koyejo, S., Lepoint, T., Liu, Y., Mittal, P., Mohri, M., Nock, R., Özgür, A., Pagh, R., Raykova, M., Qi, H., Ramage, D., Raskar, R., Song, D., Song, W., Stich, S. U., Sun, Z., Suresh, A. T., Tramèr, F., Vepakomma, P., Wang, J., Xiong, L., Xu, Z., Yang, Q., Yu, F. X., Yu, H., and Zhao, S. Advances and open problems in federated learning, 2019.

Lai, K. A., Rao, A. B., and Vempala, S. Agnostic estimation of mean and covariance. In *Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '16, pp. 665–674. IEEE Computer Society, 2016.

Paninski, L. A coincidence-based test for uniformity given very sparsely sampled discrete data. *IEEE Transactions on Information Theory*, 54(10):4750–4755, 2008.

Qiao, M. and Valiant, G. Learning discrete distributions from untrusted batches. In *9th Innovations in Theoretical Computer Science Conference (ITCS 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

Sheffet, O. Differentially private ordinary least squares. In *Proceedings of the 34th International Conference on Machine Learning*, ICML '17, pp. 3105–3114. JMLR, Inc., 2017.

Valiant, G. and Valiant, P. Estimating the unseen: An $n/\log n$-sample estimator for entropy and support size, shown optimal via new CLTs. In *Proceedings of the 43rd Annual ACM Symposium on the Theory of Computing*, STOC '11, pp. 685–694, New York, NY, USA, 2011. ACM.

Valiant, G. and Valiant, P. An automatic inequality prover and instance optimal identity testing. In *Proceedings of the 55th Annual IEEE Symposium on Foundations of Computer Science*, pp. 51–60. IEEE, 2014.