Universally Instance-Optimal Mechanisms for Private Statistical Estimation

Hilal Asi HILAL.ASI94@GMAIL.COM
Apple

John C. Duchi JDUCHI@STANFORD.EDU
Departments of Statistics and Electrical Engineering, Stanford University

Saminul Haque SAMINULH@STANFORD.EDU
Department of Computer Science, Stanford University

Zewei Li ZEWEI.LI@U.NORTHWESTERN.EDU
Department of Statistics and Data Science, Northwestern University

Feng Ruan FENGRUAN@NORTHWESTERN.EDU
Department of Statistics and Data Science, Northwestern University

Editors: Shipra Agrawal and Aaron Roth

Abstract

We consider the problem of instance-optimal statistical estimation under the constraint of differential privacy where mechanisms must adapt to the difficulty of the input dataset. We prove a new instance specific lower bound using a new divergence and show it characterizes the local minimax optimal rates for private statistical estimation. We propose two new mechanisms that are universally instance-optimal for general estimation problems up to logarithmic factors. Our first mechanism, the total variation mechanism, builds on the exponential mechanism with stable approximations of the total variation distance, and is universally instance-optimal in the high privacy regime $\varepsilon \le 1/\sqrt{n}$. Our second mechanism, the T-mechanism, is based on the T-estimator framework (Birgé, 2006) using the clipped log likelihood ratio as a stable test: it attains instance-optimal rates for any $\varepsilon \le 1$ up to logarithmic factors. Finally, we study the implications of our results to robust statistical estimation, and show that our algorithms are universally optimal for this problem, characterizing the optimal minimax rates for robust statistical estimation.

1. Introduction

Private statistical estimation has become increasingly vital in contemporary data analysis, particularly in light of growing concerns surrounding individual privacy. In this problem, given n samples from a distribution $P \in \mathcal{P}$ over a space $\mathbb{S}$, an algorithm aims to estimate a parameter of the distribution $\theta(P)$ under the constraint of differential privacy where $\theta : \mathcal{P} \to \Theta \subset \mathbb{R}$.

A substantial amount of research has been dedicated to exploring differentially private statistical estimation (Smith, 2011; Duchi et al., 2018; Canonne et al., 2018), mostly focusing on the design of minimax optimal algorithms. This notion measures the performance of an algorithm for the family of distributions $\mathcal{P}$ through its performance on the hardest distribution $P \in \mathcal{P}$. As a result, the performance of minimax optimal algorithms need not adapt to the difficulty of distribution P, and hence might suffer from bad performance for easy distributions.

The concept of instance-optimality tackles these concerns: under this paradigm, an instance-optimal algorithm must simultaneously obtain the best possible performance for every distribution,

hence adapting to the difficulty of each distribution. Existing work in statistical estimation study instance-optimality through the notion of *local minimax risk* (Cai and Low, 2015), which allows to measure the optimal instance-specific risk for a distribution P by comparing to its hardest alternative distribution $Q \in \mathcal{P}$.

Driven by its significance, there have been a significant interest in instance-optimality for privacy-preserving algorithms (Asi and Duchi, 2020a,b; Huang et al., 2021; Dick et al., 2023; McMillan et al., 2022). Asi and Duchi (2020a) study instance-optimality for estimating empirical quantities of the input sample, and show that the inverse sensitivity mechanism is universally nearly instance-optimal for any 1-dimensional function of interest (up to logarithmic factors). However, the inverse sensitivity mechanism may not be instance-optimal for statistical estimation of parameters of the population, leading several recent papers to propose new algorithms for statistical estimation (Huang et al., 2021; Dick et al., 2023; McMillan et al., 2022).

In contrast to empirical estimation where universally nearly instance-optimal algorithms exist (Asi and Duchi, 2020a), existing work on instance-optimal statistical estimation is limited to specific problems (Huang et al., 2021; Dick et al., 2023; McMillan et al., 2022). Indeed, Huang et al. (2021) and Dick et al. (2023) propose instance-optimal procedures that work only for the problem of mean-estimation, while the procedures of McMillan et al. (2022) are limited to restricted families of distributions such as single-parameter exponential families with strict conditions on the moments of the distribution.

To address these issues, in this work we study instance-optimality for general private estimation problems, aiming to develop universally instance-optimal algorithms that work for any estimation task. We accomplish this in two steps: first, we prove new lower bounds on the local minimax risk for each distribution $P \in \mathcal{P}$, characterizing the instance-specific risk for each distribution. Then, we propose two new mechanisms that are universally (nearly) instance-optimal for general estimation problems, matching our local-minimax lower bounds up to logarithmic factors.

1.1. Our contributions

We study private statistical estimation where we are given a dataset $\mathcal{S} = (S_1, \ldots, S_n)$ drawn from a distribution $P \in \mathcal{P}$ on a space $\mathbb{S}$. The goal is to estimate a parameter $\theta(P)$ under the constraint of ε-differential privacy where $\theta : \mathcal{P} \to \Theta \subset \mathbb{R}$.

Tight characterization of local-minimax rates. Our main contribution in this work is establishing a tight characterization of the local minimax rates for private statistical estimation problems. To situate our work within the existing literature, it is instructive to review existing characterizations for the rates of statistical estimation problems with and without privacy constraints. These rates are typically based on the notion of local modulus ω_D with respect to a distance function $D : \mathcal{P} \times \mathcal{P} \to \mathbb{R}$ between two distributions. More precisely, we define the local modulus with respect to a distance:

$$\omega_D\left(\delta, P_0; \mathcal{P}\right) := \sup_{P_1 \in \mathcal{P}} \left\{|\theta(P_1) - \theta(P_0)| \mid D(P_1, P_0) \leq \delta\right\}. \tag{1}$$

The local modulus ω_D at P_0 essentially measures the amount of change in the parameter $\theta(P_0)$ when the distribution P_0 is allowed to shift by distance δ when the distance is measured by D.

This quantity proves extremely useful for determining the rates of convergence of estimation problems: indeed, without privacy, the local modulus $\omega_{d_{\rm hel}}(1/\sqrt{n}, P_0; \mathcal{P})$ with respect to the

Hellinger distance d_{hel} provides tight convergence rates for the local risk for a distribution P_0. Additionally, the local modulus have been used to give tight rates in the privacy literature as well: Asi and Duchi (2020a) show that the local modulus $\omega_{TV}(1/n\varepsilon, \hat{P}_0; \mathcal{P})$ with respect to the total variation distance provides tight rates for estimating functions of the empirical distribution $\hat{P}_0$. Moreover, Duchi and Ruan (2018a) show that the TV local modululs $\omega_{TV}(1/\sqrt{n\varepsilon^2}, P_0; \mathcal{P})$ gives tight rates for private statistical estimation in the local privacy model.

For private statistical estimation in the central model, McMillan et al. (2022) recently show that for generalized exponential families with certain moment assumptions, the optimal risk for a distribution P_0 is roughly

$$\omega_{d_{\mathrm{hel}}}(1/\sqrt{n}, P_0; \mathcal{P}) + \omega_{TV}(1/n\varepsilon, P_0; \mathcal{P}). \tag{2}$$

While the local modulus in (2) is a valid lower bound for general estimation problems, it remained unclear whether there exists a mechanisms that attains these bounds without additional conditions.

Our work resolves this and demonstrates that the rate (2) is not tight for general estimation problems. Instead, we prove new tight rates for private statistical estimation that require a new measure of distance D_ε between two distributions

$$D_\varepsilon(P, Q) := \int (P(x) - Q(x)) \left[\log \frac{P(x)}{Q(x)}\right]_{-\varepsilon}^{\varepsilon} dx,$$

where $[t]_a^b := \max\{a, \min\{t, b\}\}$ is the projection of t to the interval $[a, b]$. This distance can be viewed as a symmetrized version of the clipped KL divergence between two distributions. Our rates are then based on the local modulus with respect to D_ε:

$$\omega_{D_\varepsilon}(\delta, P_0; \mathcal{P}) = \sup_{P_1 \in \mathcal{P}} \{|\theta(P_1) - \theta(P_0)| \mid D_\varepsilon(P_1, P_0) \le \delta\}. \tag{3}$$

The central result of this paper is that the quantity $\omega_{D_\varepsilon}(\frac{1}{n}, P_0; \mathcal{P})$ accurately captures the optimal local minimax risk for general statistical estimation problems. We establish this result via improved lower bounds for private estimation and new mechanisms that attain these lower bounds. Throughout the paper, we will use the simpler notation ω_ε instead of ω_{D_ε}.

Instance-specific lower bound (Section 2). We prove tight instance-specific lower bounds for private statistical estimation using the notion of local minimax risk. In contrast from prior work which proved lower bounds that depend on the total variation local modulus (McMillan et al., 2022), our lower bounds are based on the local modulus with respect to D_ε. We show that this lower bound is always larger than the total-variation lower bound, and that the gap between them can be arbitrarily large for certain problems. We also show super-efficiency about this lower bound.

Universal instance-optimality via the TV mechanism (Section 3). We propose a new total-variation based framework for designing private algorithms for statistical estimation. This framework is based on the exponential mechanism with stable approximations of the total variation distance. By carefully developing stable approximations of the TV distance, this framework yields a mechanism that is universally nearly-instance-optimal (up to logarithmic factors) for any family of distributions in the high privacy regime $\varepsilon \le 1/\sqrt{n}$.

Universal instance-optimality via the T-mechanism (Section 4). Our main algorithm is the T-mechanism which is based on the T-estimation framework (Birgé, 2006). The mechanism uses T-estimators with clipped log-likelihood ratio as a stable score, and satisfies a strong universal instance-optimality guarantee (up to logarithmic factors) for any statistical estimation problem and all $\varepsilon \le 1$. As a result, we obtain near-optimal rates for multiple-hypothesis testing, extending prior work (Canonne et al., 2018; McMillan et al., 2022) which holds only for binary hypothesis testing.

Implication for robust 1-dimensional estimation (Section 5). Finally, building on recent connections between private and robust statistical estimation (Hopkins et al., 2022; Asi et al., 2023), our results for private estimation imply a tight characterization for the rates of convergence of τ-robust estimation problems where an adversary can corrupt $n\tau$ samples in the dataset. We show that our local modulus $\omega_{1/n\tau}$ (up to logarithmic factors) governs the rates of convergence for τ-robust estimation, demonstrating that the T-mechanism is universally nearly-optimal for robust estimation (up to logarithmic factors). To the best of our knowledge, our work is the first to provide a precise characterization of the rates for robust estimation problems. Previous studies (Donoho and Liu, 1988; Zhu et al., 2022) achieved optimality results exclusively in the infinite sample regime. However, as our results indicate, the total-variation rates presented in these papers do not capture the correct rates for the finite-sample regime.

1.2. Problem Setting

We assume we have access to i.i.d. data $\mathcal{S} = (S_1, \ldots, S_n)$ drawn from a distribution $P \in \mathcal{P}$ on a space $\mathbb{S}$. We let $\theta : \mathcal{P} \to \Theta \subset \mathbb{R}$ be a parameter of the sampling distribution P that we wish to estimate, so that $\theta(P)$ denotes the target parameter (e.g. $\theta(P) = E_{S\sim P}[S]$). We also require that $\mathsf{diam}(\Theta) \le R$ which is a necessary condition for pure ε-DP estimation.

Our goal is to design differentially private mechanisms that estimate $\theta(P)$ given $\mathcal{S}$. We recall the notion of differential privacy.

Definition 1 (Differential Privacy (Dwork et al., 2006)) *A (randomized) algorithm $\mathcal{A}\colon \mathbb{S}^n \to \Theta$ is* ε-differentially private *(DP) if for all neighboring datasets $\mathcal{S}, \mathcal{S}' \in \mathbb{S}^n$ that differ in one sample and any measurable output space $T \subseteq \Theta$ we have $\Pr[\mathcal{A}(\mathcal{S}) \in T] \le e^\varepsilon \Pr[\mathcal{A}(\mathcal{S}') \in T]$.*

Differentially private statistical estimation. In this problem, which is the main focus of this paper, we are interested in differentially private algorithms for estimating a parameter $\theta(P)$ of some distribution $P \in \mathcal{P}$, given n samples $\mathcal{S} = (S_1, \ldots, S_n)$ drawn from P. We require the following two properties for an algorithm $\mathcal{A} : \mathcal{S}^n \to \Theta$ for this problem:

1. (Privacy) $\mathcal{A}$ satisfies ε-DP for all input datasets $\mathcal{S} \in \mathbb{S}^n$ (regardless of P).
2. (Utility) We measure the error of the algorithm over inputs $\mathcal{S} \sim P^n$. The expected error of the algorithm for a distribution $P \in \mathcal{P}$ is $\mathsf{Err}(\mathcal{A}, P) := \mathbb{E}_{\mathcal{S}\sim P^n}[|\mathcal{A}(\mathcal{S}) - \theta(P)|]$.

   Throughout the paper, we will usually consider upper bounds on the error $|\mathcal{A}(\mathcal{S}) - \theta(P)|$ that hold with high probability for $\mathcal{S} \sim P^n$.

To measure the difficulty of a problem with respect the family of distributions, we use the standard notion of global minimax risk.

Definition 2 (Global Minimax Complexity) *The global minimax complexity for a family of distributions $\mathcal{P}$ is*

$$\mathfrak{M}_n(\mathcal{P}, \mathcal{A}_\varepsilon) = \inf_{\mathcal{A}\in\mathcal{A}_\varepsilon} \sup_{P\in\mathcal{P}} \mathbb{E}_{\mathcal{A},\mathcal{S}\sim P^n} \left[|\mathcal{A}(\mathcal{S}) - \theta(P)|\right],$$

where $\mathcal{A}_\varepsilon$ is the family of all ε-DP algorithms.

Finally, to evaluate the instance-optimality guarantees of our algorithm, we measure the risk of an algorithm over a specific distribution P_0 through the local minimax risk, following classical notions in statistics (Cai and Low, 2015), and existing papers on instance-optimality under privacy constraints (Asi and Duchi, 2020a; McMillan et al., 2022). Throughout the paper, we let $\mathcal{A}_\varepsilon$ be the family of ε-differentially private algorithms.

Definition 3 (Local Minimax Complexity) *The local minimax complexity for a distribution $P_0 \in \mathcal{P}$ is defined as*

$$\mathfrak{M}_n^{\text{loc}}(P_0; \mathcal{P}, \mathcal{A}_\varepsilon) := \sup_{P_1\in\mathcal{P}} \inf_{\mathcal{A}\in\mathcal{A}_\varepsilon} \max_{P\in\{P_0,P_1\}} \mathbb{E}_{\mathcal{A},\mathcal{S}\sim P^n} \left[|\mathcal{A}(\mathcal{S}) - \theta(P)|\right],$$

where $\mathcal{A}_\varepsilon$ is the family of all ε-DP algorithms.

This definition originates from an observation made by Stein (1956), suggesting that a problem's difficulty should be comparable to its "most challenging one-dimensional sub-problem". In this definition, the algorithm $\mathcal{A}'$ knows that the distribution can be either P_0 or P_1 for some $P_1 \in \mathcal{P}$, and its performance will be measured based on the hardest alternative P_1.

Finally, we say that an algorithm $\mathcal{A}$ is instance-optimal if it attains the local minimax risk for every distribution $P_0 \in \mathcal{P}$, that is, $\mathsf{Err}(\mathcal{A}, P_0) = O(\mathfrak{M}_n^{\text{loc}}(P_0; \mathcal{P}, \mathcal{A}_\varepsilon))$.

1.3. Preliminaries

T-estimators. Birgé (2006) introduced T-estimators as an alternative to MLE for estimating a distribution in a given distance d. The T-estimator takes a discretization $M \subset \mathcal{P}$ and for every pair of distributions $P, Q \in M$, defines a test $T_{P,Q} : \mathbb{S}^\star \to \{0, 1\}$: given a dataset $\mathcal{S} \overset{\text{iid}}{\sim} P_0$, the test either accepts P in favour of Q ($T_{P,Q}(\mathcal{S}) = 0$) or rejects P in favour of Q ($T_{P,Q}(\mathcal{S}) = 1$). The tests are required to satisfy the consistency condition that $T_{P,Q} = 1 - T_{Q,P}$ for all $P, Q \in M$. The T-estimator then outputs distribution $\hat{P} \in M$ if for all other $Q \in M$, the test $T_{\hat{P},Q}$ favours $\hat{P}$ to Q. In the case of Hellinger- and TV-distance estimation, (Birgé, 2006, Proposition 6) chooses the tests $T_{P,Q}$ to be of the form $T_{P,Q} = 1\{\frac{1}{n}\sum_{i=1}^n \psi_{P,Q}(S_i) \le 0\}$ for some function $\psi_{P,Q}$. This functional form for the tests motivates our private T-mechanism, as we discuss in Section 4.

Exponential mechanism. Given an input $\mathcal{S} \in \mathbb{S}^n$ and a distance function $g : \Theta \times \mathbb{S}^n \to \mathbb{R}$ where $\Theta \subset \mathbb{R}$, the exponential mechanism is a private mechanism that aims to return $t \in \Theta$ that approximately minimizes the distance $g(t; \mathcal{S})$ (McSherry and Talwar, 2007). It uses the *global sensitivity* of g, defines as $\Delta_g := \sup_{\theta,\mathcal{S},\mathcal{S}':d_{\text{ham}}(\mathcal{S},\mathcal{S}')\le 1} |g(\theta; \mathcal{S}) - g(\theta; \mathcal{S}')|$. Then, given an input $\mathcal{S} \in \mathbb{S}^n$, the exponential mechanism outputs $t \in \theta$ using the following density function

$$f_{\mathsf{exp}}(t; \mathcal{S}) \propto \exp\left(-\frac{\varepsilon}{2\Delta_g} \cdot g(\theta; \mathcal{S})\right).$$

The following lemma summarizes the guarantees of the exponential mechanism.

Lemma 4 (McSherry and Talwar (2007)) *The exponential mechanism f_{exp} is ε-DP.*

We also require a smooth version of the exponential mechanism which provides better utility bounds when Θ is not discrete. Following Asi and Duchi (2020a), given $\rho > 0$, we define the ρ-smooth score of g to be $g^\rho(t; \mathcal{S}) = \inf_{s \in \Theta : |s-t| \le \rho} g(t; \mathcal{S})$. The ρ-smooth exponential mechanism then applies the exponential mechanism with the smooth score g^ρ,

$$f_{\mathsf{sm-exp}}(t; \mathcal{S}) \propto \exp\left(-\frac{\varepsilon}{2\Delta_g} \cdot g^\rho(\theta; \mathcal{S})\right).$$

The smooth exponential mechanism also preserves privacy as the global sensitivity of g^ρ is also Δ_g.

Lemma 5 (Asi and Duchi (2020a)) *The ρ-smooth exponential mechanism $f_{\mathsf{sm-exp}}$ is ε-DP.*

2. Local Minimax Lower Bounds

In this section we introduce our local modulus and prove local minimax lower bounds. We begin by defining the local modulus and proving a global minimax lower bound, and a local minimax lower bounds. Key to our lower bounds is the new functional D_ε

$$D_\varepsilon(P, Q) := \int (P(x) - Q(x)) \left[\log \frac{P(x)}{Q(x)}\right]_{-\varepsilon}^{\varepsilon} dx,$$

where $[t]_a^b := \max\{a, \min\{t, b\}\}$. We show several properties of D_ε in Appendix B such as non-negativity, finiteness, and connections to total variation and Hellinger distance. Our lower bounds are then based on the local modulus ω_ε (3) of continuity with respect to the functional D_ε:

$$\omega_\varepsilon(\delta, P_0; \mathcal{P}) = \sup_{P_1 \in \mathcal{P}} \left\{|\theta(P_1) - \theta(P_0)| \mid D_\varepsilon(P_1, P_0) \le \delta\right\}.$$

We also provide lower bounds for the global minimax risk of a family of distribution $\mathcal{P}$:

$$\mathfrak{M}_n(\mathcal{P}, \mathcal{A}_\varepsilon) = \inf_{\mathcal{A} \in \mathcal{A}_\varepsilon} \sup_{P \in \mathcal{P}} \mathbb{E}_{\mathcal{A}, \mathcal{S} \sim P^n}\left[|\mathcal{A}(\mathcal{S}) - \theta(P)|\right].$$

The following result presents our main lower bound, demonstrating that the local modulus ω_ε determines the rates for local and global miniamx risks. We defer the proof to Appendix D.1.

Theorem 6 *Let $\varepsilon \le 1$, $P_0 \in \mathcal{P}$, and $\mathcal{A}_\varepsilon$ be the collection of ε-differentially private estimators. Then for a universal constant $C < \infty$*

$$\mathfrak{M}_n^{\mathrm{loc}}(P_0; \mathcal{P}, \mathcal{A}_\varepsilon) \ge \frac{1}{6}\omega_\varepsilon\left(\frac{C}{n}, P_0; \mathcal{P}\right).$$

In particular, we also have

$$\mathfrak{M}_n(\mathcal{P}, \mathcal{A}_\varepsilon) \ge \frac{1}{6} \sup_{P_0 \in \mathcal{P}} \omega_\varepsilon\left(\frac{C}{n}, P_0; \mathcal{P}\right).$$

We show that this lower bound is always larger than the existing total variation lower bound by McMillan et al. (2022) and the standard hellinger lower bound for non-private estimation (see Proposition 39), showing that

$$\omega_\varepsilon\left(\frac{9}{n}, P_0; \mathcal{P}\right) \geq \omega_{\mathrm{TV}}\left(\frac{1}{n\varepsilon}, P_0; \mathcal{P}\right) \vee \omega_{d_{\mathrm{hel}}}\left(\frac{1}{\sqrt{n}}, P_0; \mathcal{P}\right).$$

Moreover, in Example 1, we show that the gap between these lower bound can be arbitrarily large: indeed, for certain distributions, we show that $\omega_\varepsilon(1/n, P_0; \mathcal{P}) = \Theta\left(\frac{1}{\varepsilon\sqrt{n}}\right)$ which is significantly larger than $\omega_{\mathrm{TV}}\left(\frac{1}{n\varepsilon}, P_0; \mathcal{P}\right) \vee \omega_{d_{\mathrm{hel}}}\left(\frac{1}{\sqrt{n}}, P_0; \mathcal{P}\right) = \Theta\left(\frac{1}{\varepsilon^2 n} + \frac{1}{\sqrt{n\varepsilon}}\right)$ in the regime $\varepsilon \ll 1$.

Finally, we also prove a super-efficiency result (see Appendix D.3): any improvement over our modulus of continuity lower bound at any distribution implies worse performance elsewhere.

3. Total Variation Framework

In this section, we begin our algorithmic contribution with a simple total-variation based framework for private estimation that obtains nearly instance-optimal guarantees in the high-privacy regime $\varepsilon \leq 1/\sqrt{n}$. In this section we assume that $\mathcal{P}$ is discrete and we show how to handle the continuous case in Section 4.2.

Our total variation framework builds on the observation that in order to match the optimal lower bounds for the high-privacy regime, that is, the total variation local modulus, one can run the exponential mechanism with the total variation score. However, to make this amenable to privacy, we must guarantee that our approximation for the total variation distance has low sensitivity.

Given $\mathcal{S} \overset{\text{iid}}{\sim} P^n$, our framework requires a distance function $\mathrm{dist}_{\mathrm{TV}}(t; \mathcal{S})$ with low sensitivity and a good approximation of the total variation distance between P and the closest Q such that $\theta(Q) = t$, that is $\inf_{Q\in\mathcal{P}:\theta(Q)=t} \|Q - P\|_{\mathrm{TV}}$. More precisely, we require the following two conditions:

(C_1) (Stability) The distance function $\mathrm{dist}_{\mathrm{TV}}(t; \mathcal{S})$ is $\frac{1}{n}$-sensitive.

(C_2) (Accuracy) For any $P \in \mathcal{P}$ and $\mathcal{S} \overset{\text{iid}}{\sim} P^n$, we have that for all $t \in \Theta$ with probability $1 - \beta$,

$$|\mathrm{dist}_{\mathrm{TV}}(t; \mathcal{S}) - \inf_{Q\in\mathcal{P}:\theta(Q)=t} \|Q - P\|_{\mathrm{TV}}| \leq \Delta.$$

Given a stable (C_1) and accurate (C_2) distance function $\mathrm{dist}_{\mathrm{TV}}$, we consider its ρ-smooth version $\mathrm{dist}^\rho_{\mathrm{TV}}(t) = \inf_{s:|s-t|\leq\rho} \mathrm{dist}_{\mathrm{TV}}(s)$. The total variation mechanism runs the ρ-smooth exponential mechanism with distance $\mathrm{dist}_{\mathrm{TV}}$, resulting in the following density function for $t \in \Theta$:

$$\mathcal{A}_{\mathsf{TV}}(t; \mathcal{S}) \propto e^{-\mathrm{dist}^\rho_{\mathrm{TV}}(t;\mathcal{S})\cdot n\varepsilon/2}. \tag{M.1}$$

The following theorem summarizes our guarantees for this mechanism.

Theorem 7 *Let $\mathcal{P}$ be discrete and P be a distribution such that $\|P - P_0\|_{\mathrm{TV}} \leq \eta$ for some $P_0 \in \mathcal{P}$. Let $\theta : \mathcal{P} \to \Theta$ such that $\mathsf{diam}(\Theta) \leq R$. Assume the distance function $\mathrm{dist}_{\mathrm{TV}}$ satisfies conditions (C_1) and (C_2). For $\rho > 0$ and given an input $\mathcal{S} \overset{\text{iid}}{\sim} P^n$, the mechanism $\mathcal{A}_{\mathsf{TV}}$ (M.1) is ε-DP and with probability $1 - 2\beta$ returns $\hat{\theta}$ such that*

$$|\hat{\theta} - \theta(P)| \leq \omega_{\mathrm{TV}}\left(\frac{2\log(R/\rho\beta)}{n\varepsilon} + 2\Delta + \eta, P; \mathcal{P}\right) + \rho.$$

Proof The privacy guarantee follows immediately from the guarantees of the smooth exponential mechanism (Lemma 5). Now we proceed to prove accuracy. We will upper bound the probability of returning $\hat{\theta}$ with large distance; let $W = \{t \in \Theta : \mathrm{dist}_{\mathrm{TV}}^{\rho}(t; \mathcal{S}) \geq K\}$ for some $K > 0$ to be chosen later.

To this end, let $P_0 \in \mathcal{P}$ be such that $\|P - P_0\|_{\mathrm{TV}} \leq \eta$. Note that $\mathrm{dist}_{\mathrm{TV}}(\theta(P_0); \mathcal{S}) \leq \Delta + \eta$ with probability at least $1 - \beta$ using the second property. This implies that $\mathrm{dist}_{\mathrm{TV}}^{\rho}(t; \mathcal{S}) \leq \Delta + \eta$ for $t \in \Theta$ such that $|t - \theta(P_0)| \leq \rho$. Now consider t such that $\mathrm{dist}_{\mathrm{TV}}^{\rho}(t; \mathcal{S}) \geq K$. The density for $\mathcal{A}_{\mathsf{TV}}(t; \mathcal{S})$ is upper bounded by $\mathcal{A}_{\mathsf{TV}}(t; \mathcal{S}) \leq \frac{e^{-Kn\varepsilon/2}}{\rho e^{-(\Delta+\eta)n\varepsilon/2}} = e^{-(K-\Delta-\eta)n\varepsilon/2}/\rho$. This implies that

$$\mathbb{P}(\hat{\theta} \in W) \leq \mathsf{diam}(\Theta) e^{-(K-\Delta-\eta)n\varepsilon/2}/\rho \leq \beta,$$

where the second inequality follows by setting $K = \Delta + \eta + 2\log(2R/\rho\beta)/n\varepsilon$.
Overall, this implies that with probability $1-2\beta$, the output $\hat{\theta} \notin W$. Thus we have that $\mathrm{dist}_{\mathrm{TV}}^{\rho}(\hat{\theta}; \mathcal{S}) \leq K$, which implies there is $t \in \Theta$ such that $|\hat{\theta} - t| \leq \rho$ and $\mathrm{dist}_{\mathrm{TV}}(t; \mathcal{S}) \leq K$. Condition ($C_1$) of $\mathrm{dist}_{\mathrm{TV}}$ now implies that

$$\inf_{Q \in \mathcal{P}: \theta(Q) = t} \|Q - P\|_{\mathrm{TV}} \leq \mathrm{dist}_{\mathrm{TV}}(t; \mathcal{S}) + \Delta \leq K + \Delta.$$

This implies that $|t - \theta(P)| \leq \omega_{\mathrm{TV}}(K + \Delta; P)$ and therefore the claim follows as

$$|\hat{\theta} - \theta(P)| \leq |\hat{\theta} - t| + |t - \theta(P)| \leq \rho + \omega_{\mathrm{TV}}(K + \Delta; P).$$

■

Having proved Theorem 7, the main challenge now is to design approximations for the total variation distance that are $1/n$-sensitive and accurate with $\Delta \ll 1$. Then, applying Theorem 7 with $\eta \ll 1/n$ and $\rho \ll 1/n$ results in instance-optimal rates up to logarithmic factors in n. In the next sections, we provide several techniques for constructing such approximations for a wide family of problems.

3.1. Approximations for distributions with monotone likelihood

In this section, we provide an approximation for families with monotone likelihood ratio.

Definition 8 *A family of distributions $\mathcal{P}$ over $\mathbb{R}$ satisfies the monotone likelihood ratio property (MLRP) if for any two distributions $P, Q \in \mathcal{P}$, we have that $\frac{P(x)}{Q(x)}$ is monotonically increasing or decreasing as a function of $x \in \mathbb{R}$.*

This property is well studied in statistics (Borges and Pfanzagl, 1963; Zacks, 1970) and many natural distributions satisfy it such as Gaussian, Binomial, or exponential families with a single parameter.

Given $\mathcal{S} = (S_1, \ldots, S_n) \overset{\text{iid}}{\sim} P$, we consider a set $\mathcal{A}$ that has subsets of $\mathbb{R}$ and approximate the total variation

$$\|Q - P\|_{\mathrm{TV}} = \sup_{A \subset \mathbb{R}} P(A) - Q(A) \approx \sup_{A \subset \mathcal{A}} \frac{1}{n} \sum_{i=1}^{n} 1\{S_i \in A\} - Q(A)$$

This approximation is $1/n$-sensitive, hence satisfying the stability condition (C_1). However, for accuracy, we need certain structural properties of the set $\mathcal{A}$ in order to guarantee uniform concentration. For MLRP families (Definition 8), the set $\mathcal{A}$ can be replaced by 1-dimensional intervals, hence resulting in the following approximation:

$$\mathrm{dist}^{\mathsf{mon}}(t;\mathcal{S}) = \inf_{Q\in\mathcal{P}:\theta(Q)=t} \sup_{s\in\mathbb{R}} \left| \frac{1}{n}\sum_{i=1}^{n} 1\{S_i \le s\} - \Pr_{S\sim Q}(S \le s) \right|. \tag{4}$$

We have the following guarantees for this approximation. We defer the proof to Appendix E.

Proposition 9 *Let $\mathcal{P}$ be discrete and P such that $\|P - P_0\|_{\mathrm{TV}} \le \eta$ for some $P_0 \in \mathcal{P}$. Let $\mathcal{P}\cup\{P\}$ satisfy the MLRP property (Definition 8), and $\mathcal{S} \overset{\text{iid}}{\sim} P^n$. Then $\mathrm{dist}^{\mathsf{mon}}$ is $1/n$-sensitive and for all $t \in \Theta$*

$$\left| \mathrm{dist}^{\mathsf{mon}}(t;\mathcal{S}) - \inf_{Q\in\mathcal{P}:\theta(Q)=t} \|Q-P\|_{\mathrm{TV}} \right| \le \sqrt{\frac{\log(2/\beta)}{n}}.$$

Using this approximation, Theorem 7 now implies the following guarantee for the total variation mechanism with $\mathrm{dist}^{\mathsf{mon}}$.

Corollary 10 *Let $\mathcal{P}$ be discrete and P such that $\|P - P_0\|_{\mathrm{TV}} \le \eta$ for some $P_0 \in \mathcal{P}$. Let $\mathcal{P}\cup\{P\}$ satisfy the MLRP property (Definition 8) Let $\theta : \mathcal{P} \to \Theta$ such that $\mathsf{diam}(\Theta) \le R$. Set $\rho = 1/n^c$ for $c \ge 1$. Given an input $\mathcal{S} \overset{\text{iid}}{\sim} P^n$, the total variation mechanism (M.1) with the distance function $\mathrm{dist}^{\mathsf{mon}}$ (4) is ε-DP and with probability $1-2\beta$ returns $\hat{\theta}$ such that*

$$|\hat{\theta} - \theta(P)| \le \omega_{\mathrm{TV}}\left(\sqrt{\frac{4\log(2/\beta)+1}{n}} + \frac{2c\cdot\log(Rn/\beta)}{n\varepsilon}, P; \mathcal{P} \right) + \frac{1}{n^c}.$$

When $\varepsilon \le 1/\sqrt{n}$, this theorem implies that the error is roughly $\omega_{\mathrm{TV}}\left(\frac{c\cdot\log(Rn/\beta)}{n\varepsilon}, P; \mathcal{P}\right) + \frac{1}{n^c}$, which matches the lower bounds in Section 2 up to logarithmic factors in n and a negligible additive term.

3.2. Approximation for general finite families

Moving beyond monotone likelihood distributions, in this section we propose another distance function for the total variation that works for any discrete family of distributions. This mechanism approximate the total variation distance via the notion of Scheffé sets, and is similar to the mechanism developed by Bun et al. (2019) for private hypothesis selection in total-variation distance. Their mechanism can be similarly used to obtain instance-optimal results for the high privacy regime $\varepsilon \le 1/\sqrt{n}$; however, it may not be optimal for the regime $\varepsilon \ge 1/\sqrt{n}$.

To this end, assume we are given a discrete family of distributions $\mathcal{P}$ over space $\mathbb{S}$. For any two distributions $P, Q \in \mathcal{P}$, we define the Scheffé set $A(P,Q) \subset \mathbb{S}$ to be the set such that

$$\|P - Q\|_{\mathrm{TV}} = P(A(P,Q)) - Q(A(P,Q)).$$

Then, we define the following set $\mathcal{A}(\mathcal{P}) = \{A(P,Q) : P, Q \in \mathcal{P}\}$. Finally, we define our distance approximation for general families to be

$$\mathrm{dist}^{\mathsf{gen}}(t;\mathcal{S}) = \inf_{Q\in\mathcal{P}:\theta(Q)=t} \sup_{A\in\mathcal{A}(\mathcal{P})} \left| \frac{1}{n}\sum_{i=1}^{n} 1\{S_i \in A\} - Q(A) \right|. \tag{5}$$

Proposition 11 *Let $\mathcal{P}$ be a discrete family of distributions and P such that $\|P - P_0\|_{\mathrm{TV}} \le \eta$ for some $P_0 \in \mathcal{P}$. Given $\mathcal{S} \overset{\mathrm{iid}}{\sim} P^n$, $\mathrm{dist}^{\mathsf{gen}}$ is $1/n$-sensitive and for all $t \in \Theta$*

$$\left|\mathrm{dist}^{\mathsf{gen}}(t; \mathcal{S}) - \inf_{Q \in \mathcal{P}: \theta(Q) = t} \|Q - P\|_{\mathrm{TV}}\right| \le \sqrt{\frac{4 \log(2|\mathcal{P}|/\beta)}{n}} + 2\eta.$$

We defer the proof to Appendix E. Note that Theorem 7 now immediately implies the following guarantees for the total variation mechanism with $\mathrm{dist}^{\mathsf{gen}}$.

Corollary 12 *Let $\mathcal{P}$ be a discrete family of distributions and P such that $\|P - P_0\|_{\mathrm{TV}} \le \eta$ for some $P_0 \in \mathcal{P}$ where $\eta \le 1/\sqrt{n}$. Let $\theta : \mathcal{P} \to \Theta$ such that $\mathsf{diam}(\Theta) \le R$. Set $\rho = 1/n^c$ for $c \ge 1$. Given an input $\mathcal{S} \overset{\mathrm{iid}}{\sim} P^n$, the mechanism $\mathcal{A}_{\mathsf{TV}}$ (M.1) with the distance function $\mathrm{dist}^{\mathsf{gen}}$ (5) is ε-DP and with probability $1 - 2\beta$ returns $\hat{\theta}$ such that*

$$|\hat{\theta} - \theta(P)| \le \omega_{\mathrm{TV}}\left(\sqrt{\frac{6 \log(2|\mathcal{P}|/\beta)}{n}} + \frac{2c \cdot \log(Rn/\beta)}{n\varepsilon}, P; \mathcal{P}\right) + \frac{1}{n^c}.$$

Note that when $\varepsilon \le 1/\sqrt{n}$ and $|\mathcal{P}| = O(\mathsf{poly}(n))$, this theorem implies that the error is roughly $\omega_{\mathrm{TV}}\left(\frac{\log(n/\beta)}{n\varepsilon}; P\right) + 1/n^c$, which matches the lower bounds in Section 2 up to logarithmic factors in n and a negligible additive term.

4. Private T-Mechanism

Moving beyond the high-privacy regime, in this section we present our main mechanism that obtains instance-optimality for all $\varepsilon \le 1$. Our mechanism, *the T-mechanism*, builds on the T-estimator framework for non-private estimation of distributions (Birgé, 2006). This framework follows the estimation-via-testing paradigm where it outputs a distribution $\hat{P}$ with an estimate $\theta(\hat{P})$. However, unlike standard approaches for distribution testing with respect to the total variation distance, our approach allows us to estimate distributions with respect to the new functional D_ε, allowing to obtain instance-optimal rates. Recall from Section 1.3 that the T-estimator outputs $\hat{P}$ if and only if $\psi_{\hat{P},Q}(\mathcal{S}) := \frac{1}{n}\sum_{i=1}^n \psi_{\hat{P},Q}(S_i) > 0$ for all $Q \ne \hat{P}$, for some predefined score functions $\psi_{P,Q}$: this function gives a score whether P or Q is more likely given the sample S.

Based on this, our T-mechanism instantiates the exponential mechanism (McSherry and Talwar, 2007) with the score function $F(P; \mathcal{S}) = \inf_{Q \in \mathcal{P}: Q \ne P} \psi_{P,Q}(\mathcal{S})$: this score will guarantee that the T-mechanism outputs $\hat{P}$ such that $\psi_{\hat{P},Q}(\mathcal{S})$ is large for all Q. To guarantee privacy, our chosen function $\psi_{P,Q}(S_i)$ have to be bounded. We first demonstrate our private T-mechanism in the case of finite families $\mathcal{P}$ in Section 4.1 and then extend it to general families in Section 4.2. Finally, in Section 4.3 we demonstrate the implications of our results to multiple hypothesis testing.

4.1. Private estimation for finite families

Beginning with the simpler setting where $\mathcal{P}$ is finite (P may not be in $\mathcal{P}$), our aim is to design a bounded score function $\psi_{P,Q}(S_i)$ that measures whether P is more likely than Q under S_i. To this end, we follow a useful tradition in robust statistics and privacy of clipping the log likelihood ratio, defining

$$\tilde{\psi}_\varepsilon(S; P, Q) := \left[\log \frac{P(S)}{Q(S)}\right]_{-\varepsilon}^{\varepsilon}.$$

This score was used by Huber (1965) to develop robust algorithms for binary hypothesis testing, and more recently by Canonne et al. (2018) and McMillan et al. (2022) to develop optimal algorithms for private binary hypothesis testing. We will not directly use that score, but instead consider the shifted version

$$\psi_\varepsilon(S; P, Q) := 2\tilde{\psi}_\varepsilon(S; P, Q) - \left(\mathbb{E}_{S\sim P}[\tilde{\psi}_\varepsilon(S; P, Q)] + \mathbb{E}_{S\sim Q}[\tilde{\psi}_\varepsilon(S; P, Q)]\right).$$

Then for a dataset $\mathcal{S} \in \mathbb{S}^n$, we let $\psi_\varepsilon(\mathcal{S}; P, Q) = \frac{1}{n}\sum_{i=1}^n \psi_\varepsilon(S_i; P, Q)$ denote the empirical mean of the sample pairwise scores. First observe that due to shifting, the score ψ_ε has the property that

$$\mathbb{E}_{S\sim P}[\psi_\varepsilon(S; P, Q)] = -\mathbb{E}_{S\sim P}[\psi_\varepsilon(S; Q, P)] = D_\varepsilon(P, Q).$$

Then, by concentration, given $\mathcal{S} \sim P_0^n$, we expect that $\psi_\varepsilon(\mathcal{S}; P_0, Q) \approx D_\varepsilon(P_0, Q) \geq 0$ for all Q, whereas $\psi_\varepsilon(\mathcal{S}; Q, P_0) \approx -D_\varepsilon(P_0, Q) \leq 0$. Based on this, we define the "distance" function $\mathsf{dist}_\varepsilon^{\mathcal{P}} : \mathcal{P} \to \mathbb{R}$,

$$\mathsf{dist}_\varepsilon^{\mathcal{P}}(P; \mathcal{S}) := -\inf_{Q\in\mathcal{P}} \psi_\varepsilon(\mathcal{S}; P, Q).$$

We expect that $\mathsf{dist}_\varepsilon^{\mathcal{P}}(P_0; \mathcal{S})$ is small for the correct distribution P_0 while $\mathsf{dist}_\varepsilon^{\mathcal{P}}(Q; \mathcal{S})$ is large for distributions Q where $D_\varepsilon(P_0, Q)$ is large.

Finally, we obtain our private T-mechanism by running the exponential mechanism (McSherry and Talwar, 2007) with the distance $\mathsf{dist}_\varepsilon^{\mathcal{P}}(P; \mathcal{S})$,

$$\mathcal{A}_{\mathsf{T}}(P; S) \propto \exp\left(-\frac{n}{2} \cdot \mathsf{dist}_\varepsilon^{\mathcal{P}}(P; \mathcal{S})\right). \tag{M.2}$$

The following theorem summarizes the guarantees of this mechanism.

Theorem 13 *There exists universal constants $c_1, c_2 > 0$ such that the following hold. Let $\mathcal{P}$ be a discrete family of distributions and P such that $\|P - P_0\|_{\mathrm{TV}} \leq \eta$ for some $P_0 \in \mathcal{P}$. Given an input $\mathcal{S} \overset{\text{iid}}{\sim} P^n$, Let $\varepsilon \leq 1$ and $\beta > 0$ such that $\log(9|\mathcal{P}|/\beta) > 6c_1\varepsilon\eta n$. The T-mechanism $\mathcal{A}_{\mathsf{T}}$ (M.2) is ε-DP and returns $\hat{P}$ such that with probability at least $1 - \beta$,*

$$|\theta(\hat{P}) - \theta(P)| \leq \omega_\varepsilon\Big(\frac{c_2 \log(|\mathcal{P}|/\beta)}{n}, P; \mathcal{P}\Big).$$

Before proving this upper bound, note that the error of the T-mechanism matches the lower bounds of Section 2 up to logarithmic factors.

Proof Observe that $\mathsf{dist}_\varepsilon^{\mathcal{P}}$ is ε/n-sensitive in $\mathcal{S}$, so the privacy guarantees of the exponential mechanism (Lemma 4) imply that the T-mechanism is ε-DP.

We now turn to proving the utility claim. To this end, we will use the guarantees of the exponential mechanism which state that the mechanism will output a distribution $\hat{P}$ such that $\mathsf{dist}_\varepsilon^{\mathcal{P}}(\hat{P}; \mathcal{S})$ is close to the minimum. Then, we will show that bad distributions Q (far away from P) have large $\mathsf{dist}_\varepsilon^{\mathcal{P}}(Q; \mathcal{S})$, while good distributions Q have small $\mathsf{dist}_\varepsilon^{\mathcal{P}}(Q; \mathcal{S})$.

We begin with the following two lemmas, both of which we prove in Appendix F.

Lemma 14 (Bad distributions have large distance) *There exists a universal constant $c > 0$ such that if $Q \in \mathcal{P}$ with $D_\varepsilon(Q, P) \geq 10\varepsilon\eta$, then we have that $\mathsf{dist}_\varepsilon^{\mathcal{P}}(Q; \mathcal{S}) \geq D_\varepsilon(P, Q)/10$ with probability at least $1 - 4\exp(-cn \cdot D_\varepsilon(Q, P))$.*

Lemma 15 (Good distributions have small distance) *There exists a universal constant $c > 0$ such that for $B \geq 4\varepsilon\eta$, we have $\mathsf{dist}_\varepsilon^{\mathcal{P}}(P_0; \mathcal{S}) \leq B$ with probability at least $1 - 4|\mathcal{P}| \exp(-cnB)$.*

Let $\gamma > 4\varepsilon\eta n$ be a quantity we decide later and define the set of bad distributions $\mathcal{P}_{bad} = \{Q \in \mathcal{P} : D_\varepsilon(Q, P) \geq 20\gamma/n\}$. Because $20\gamma/n > 10\eta\varepsilon$, we may apply Lemma 14 to get that $\mathsf{dist}_\varepsilon^{\mathcal{P}}(Q; \mathcal{S}) \geq 2\gamma/n$ with probability at least $1 - 4\exp(-c\gamma)$. Taking a union bound over all $Q \in \mathcal{P}_{bad}$, we have that the event $E_1 = \{\mathsf{dist}_\varepsilon^{\mathcal{P}}(Q; \mathcal{S}) \geq 2\gamma/n \text{ for all } Q \in \mathcal{P}_{bad}\}$ occurs with probability at least $1 - 4|\mathcal{P}_{bad}| \exp(-c\gamma)$.

On the other hand, Lemma 15 applied with $B = \gamma/n$ gives that $E_2 = \{\mathsf{dist}_\varepsilon^{\mathcal{P}}(P_0; \mathcal{S}) \leq \gamma/n\}$ holds with probability at least $1 - 4|\mathcal{P}| \exp(-c\gamma)$.

Under the event $E_1 \cap E_2$, we can bound the density $\mathcal{A}_{\mathsf{T}}(Q; S)$ for $Q \in \mathcal{P}_{bad}$ by

$$\mathcal{A}_{\mathsf{T}}(Q; S) = \frac{e^{-n\mathsf{dist}_\varepsilon^{\mathcal{P}}(Q;\mathcal{S})/2}}{\sum_{Q' \in \mathcal{P}} e^{-n\mathsf{dist}_\varepsilon^{\mathcal{P}}(Q;\mathcal{S})/2}} \leq \frac{e^{-n\mathsf{dist}_\varepsilon^{\mathcal{P}}(Q;\mathcal{S})/2}}{e^{-n\mathsf{dist}_\varepsilon^{\mathcal{P}}(P_0;\mathcal{S})/2}} \leq \exp(-\gamma/2).$$

Therefore, under the event $E_1 \cap E_2$, the probability that $\hat{P} \in \mathcal{P}_{bad}$ is at most $|\mathcal{P}_{bad}| \exp(-\gamma/2) \leq |\mathcal{P}| \exp(-c\gamma)$. By picking $\gamma = \log(9|\mathcal{P}|/\beta)/c$ and verifying that γ satisfies $\gamma > 4\varepsilon\eta n$, we have the desired result. ■

We briefly remark that this algorithm can be implemented in $O(n|\mathcal{P}|^2)$ time by computing all the pairwise scores between distributions.

4.2. Private estimation for unbounded families

In this section, we show how to extend our mechanisms to the non-discrete case where the size of $\mathcal{P}$ might be unbounded. To this end, we will use discretizations $\mathcal{P}_\eta$ of $\mathcal{P}$ that are η-covers in total variation distance: for all $P \in \mathcal{P}$ there is $P_\eta \in \mathcal{P}_\eta$ such that $\|P - P_\eta\|_{\mathrm{TV}} \leq \eta$. Given such a cover $\mathcal{P}_\eta$, we then run our mechanisms over the discrete family of distributions $\mathcal{P}_\eta$.

Proposition 16 *There exists a constant $c > 0$ such that the following holds. Let $\mathcal{P}$ be a family of distributions and $\mathcal{P}_\eta$ be an η-cover of $\mathcal{P}$ in total variation distance where $\eta \leq cn^{-2}$. Let $\varepsilon \leq 1$ and $\beta > 0$. Given an input $\mathcal{S} \overset{\text{iid}}{\sim} P^n$ for $P \in \mathcal{P}$, the mechanism $\mathcal{A}_{\mathsf{T}}$ applied over $\mathcal{P}_\eta$ returns $\hat{P}$ such that with probability at least $1 - \beta$,*

$$|\theta(\hat{P}) - \theta(P)| \leq \omega_\varepsilon\left(c \cdot \frac{\log(|\mathcal{P}_\eta|/\beta)}{n}, P; \mathcal{P}\right)$$

Proof We first argue that we can apply Theorem 13 to the sub-family $\mathcal{P}_\eta$. For a given η, by definition of a covering we have that there exists $P_0 \in \mathcal{P}_\eta$ such that $\|P - P_0\|_{\mathrm{TV}} \leq \eta$. Also, for $\eta = 6cn^{-2}$, $\log(9|\mathcal{P}_\eta|/\beta) > 6c\varepsilon\eta n$ because $\log(9|\mathcal{P}_\eta|/\beta) \geq \log 9 > 2$ while $6c\varepsilon\eta n \leq 1/n \leq 1$. Thus, all the requirements for applying Theorem 13 to P and $\mathcal{P}_\eta$ are met, and thus the T-estimator returns $\hat{P}$ such that with probability at least $1 - \beta$, $D_\varepsilon(P, \hat{P}) \leq c\frac{\log(|\mathcal{P}_\eta|/\beta)}{n}$. The claim follows. ■

To discuss these implications further, we say that $\mathcal{P}$ has *TV-dimension at most d* if there exists a constant $C > 0$ such that for all $\eta > 0$, there is an η-covering $\mathcal{P}_\eta$ in TV-distance with $|\mathcal{P}_\eta| \leq C\eta^{-d}$. Proposition 16 implies that the mechanism outputs $\hat{P}$ such that $D_\varepsilon(P, \hat{P}) \leq c\frac{d\log(n)+\log(1/\beta)}{n}$. Also,

by the final remark in Section 4.1, the runtime of the T-estimator in this case is $O(n^{2d+1})$, which is polynomial in n and exponential in d. In particular, when the TV-dimension d is of constant order, we see that $\mathcal{A}_{\mathsf{T}}$ obtains, in polynomial time, estimation rates matching the local minimax lower bounds for non-discrete families from Section 2 up to logarithmic factors.

Moreover, when $\varepsilon \le 1/\sqrt{n}$, Proposition 16 implies that $D_\varepsilon(P, \hat{P}) \le c\frac{d\log(n)+\log(1/\beta)}{n}$, which in turn implies that $\|\hat{P} - P\|_{\mathrm{TV}} \le c\frac{d\log(n)+1+\log(1/\beta)}{n\varepsilon}$ (see Lemma 25). For high-dimensional Gaussian estimation with unknown mean and covariance, Bun et al. (2019, Lemma 6.8) shows the TV-dimension of this family is $O(d^2)$ and so $\mathcal{A}_{\mathsf{T}}$ would have error of $O(\frac{d^2\log(n)+\log(1/\beta)}{n\varepsilon})$ in TV-distance. This is tight up to logarithmic factors in n by Alabi et al. (2023, Theorem 7.1).

4.3. Implications to multiple hypothesis selection

In this section, we briefly discuss the implications of our results to the setting of multiple hypothesis testing. In this setting, we have a finite family of distributions $\mathcal{P} = \{P_1, \dots, P_m\}$, and we have access to n i.i.d. samples from $P^\star \in \mathcal{P}$. The goal is to privately estimate $P^\star$. For the binary case ($m = 2$), Canonne et al. (2018) provide tight characterization of the sample complexity needed to guarantee that $\hat{P} = P^\star$ with constant probability. Combined with Theorem 29, their results essentially imply that

$$n = \widetilde{\Theta}\left(\frac{1}{\Delta_\varepsilon}\right), \qquad \Delta_\varepsilon = \min\{D_\varepsilon(P, P') \mid P, P' \in \mathcal{P}, P \ne P'\}.$$

For the case of multiple hypothesis $m > 2$, Bun et al. (2019) gives an algorithm that guarantees $\hat{P} = P^\star$ with constant probability when (see Lemma 3.4 in Bun et al. (2019) with minor modifications to handle the well-specified case).

$$n = \widetilde{\Omega}\left(\frac{\log m}{\Delta_{\mathrm{TV}}^2} + \frac{\log m}{\Delta_{\mathrm{TV}}\varepsilon}\right), \qquad \Delta_{\mathrm{TV}} = \min\{\left\|P - P'\right\|_{\mathrm{TV}} \mid P, P' \in \mathcal{P}, P \ne P'\}.$$

These rates are not optimal in general: indeed, they do not match the optimal rate of Canonne et al. (2018) for the binary case.

On the other hand, the T-mechanism $\mathcal{A}_{\mathsf{T}}$ (M.2) outputs an estimate with an error bound in D_ε rather than total variation. As a consequence, the T-mechanism guarantees $\hat{P} = P^\star$ with constant probability when

$$n = \widetilde{\Omega}\left(\frac{\log m}{\Delta_\varepsilon}\right), \qquad \Delta_\varepsilon = \min\{D_\varepsilon(P, P') \mid P, P' \in \mathcal{P}, P \ne P'\}.$$

It is instructive to note that our rates match the optimal rates for binary hypothesis testing when $m = 2$, implying that the sample complexity of the T-mechanism is optimal up to a factor of $\log m$: we conjecture that this rate is optimal and leave open the question of proving a better lower bound that depends on $\log m$. Moreover, in the high-privacy regime $\varepsilon \le 1/\sqrt{n}$, our rates and these of Bun et al. (2019) have the same sample complexity because $D_\varepsilon(P, P') = \Theta(\varepsilon \left\|P - P'\right\|_{\mathrm{TV}})$ (see Lemma 25). However, in the regime $\varepsilon \ge 1/\sqrt{n}$, the T-estimator gives better sample complexity for families $\mathcal{P}$ where $\Delta_{\mathrm{TV}}^2 < \Delta_\varepsilon$.

5. Implications to robust statistical estimation

We conclude the paper by discussing the implications of our results to robust statistical estimation of one dimensional functions which was recently shown to be equivalent to private statistical estimation (Asi et al., 2023). As a result, we show that the optimal rates for robust estimation are governed by our proposed local modulus $\omega_{1/n\tau}$ where τ is the robustness parameter, and that our T-Mechanism is an optimal robust mechanism. This should be compared to existing work on robust statistical estimation (Donoho and Liu, 1988; Zhu et al., 2022) which obtained optimality results only in the infinite sample regime. The rates in these works are based on the total variation local modulus and therefore do not provide a tight characterization for the finite-sample rates; indeed, our results below show that $\omega_{1/n\tau}$ represents the accurate rate of convergence.

The notion of robust estimation requires the algorithm to estimate the parameter accurately even when the input dataset has $n\tau$ corruptions. We formally define the notion of τ-robust estimation.

Definition 17 (τ-robust estimation) *Let $\mathcal{P}$ be a family of distributions and $\theta : \mathcal{P} \to \Theta \subset \mathbb{R}$ be a statistic of interest. Let $\mathcal{A}$ be a (randomized) algorithm for the estimation of statistic θ. We say that $\mathcal{A}$ is a τ-robust estimator for distribution P with error α if we have that*

$$\mathsf{Err}_\tau(\mathcal{A}, P) := \mathbb{E}_{\mathcal{S}\sim P^n}\left[\max_{\mathcal{S}' : d_{\rm ham}(\mathcal{S},\mathcal{S}')\le n\tau} \mathbb{E}_{\mathcal{A}}[|\mathcal{A}(\mathcal{S}') - \theta(P)|]\right] \le \alpha.$$

Asi et al. (2023) devise reductions between private and robust estimators which essentially show that the optimal error for ε-DP statistical estimation matches the error for τ-robust estimation where $\tau \approx 1/n\varepsilon$ under natural assumptions. This allows us to characterize the minimax optimal rates for τ-robust estimation, defined as

$$\mathfrak{M}_n(\mathcal{P}, \mathcal{A}_\tau) := \inf_{\mathcal{A}\in\mathcal{A}_\tau} \sup_{P\in\mathcal{P}} \mathsf{Err}_\tau(\mathcal{A}, P)$$

where $\mathcal{A}_\tau$ is the family of τ-robust estimators.

Using the fact that private algorithms are also robust, we have the following error for the T-mechanism for robust estimation, giving an upper bound on the minimax error.

Theorem 18 *There exists universal constants $c_1, c_2 > 0$ such that the following hold. Let $1/n \le \tau \le 1$ and suppose $\mathcal{S} \overset{\rm iid}{\sim} P^n \in \mathcal{P}$ and $\mathcal{S}'$ such that $d_{\rm ham}(\mathcal{S}, \mathcal{S}') \le n\tau$. Let $\mathcal{P}_\eta$ be an η-covering of $\mathcal{P}$ such that $\eta \le c_1/n$. The T-estimator mechanism $\mathcal{A}_T$ (M.2) with parameter $\varepsilon = 1/n\tau$ is τ-robust with error $\mathsf{Err}_\tau(\mathcal{A}_T, P) \le O\left(\omega_{1/n\tau}\left(\frac{c_2\log(|\mathcal{P}_\eta|/\beta)}{n}; P\right)\right) + \beta$. In particular,*

$$\mathfrak{M}_n(\mathcal{P}, \mathcal{A}_\tau) \le O\left(\sup_{P\in\mathcal{P}} \omega_{1/n\tau}\left(\frac{c_2\log(n|\mathcal{P}_\eta|)}{n}; P\right) + \frac{1}{n^2}\right).$$

Moreover, we have the following lower bound on the minimax risk for robust estimation. Our lower bound requires the mild assumption that the minimax is lower bounded by a polynomial which is usually the case in most statistical applications.

Theorem 19 *Let $\tau \in (0, 1)$, $\mathcal{P}$ be a family of distributions, $|\theta(P)| \le 1$ for all $P \in \mathcal{P}$, and assume $\mathfrak{M}_n(\mathcal{P}, \mathcal{A}_\tau) \ge 1/n^{C_1}$ for some constant $C_1 < \infty$. Then there is a constant $C_2 < \infty$ such that the minimax error for τ-robust estimation is*

$$\mathfrak{M}_n(\mathcal{P}, \mathcal{A}_\tau) \ge \Omega\left(\sup_{P\in\mathcal{P}} \omega_{C_2\log(n)/n\tau}\left(\frac{1}{n}; P\right)\right).$$

References

Daniel Alabi, Pravesh K. Kothari, Pranay Tankala, Prayaag Venkat, and Fred Zhang. Privately estimating a gaussian: Efficient, robust and optimal. *arXiv:2212.08018 [cs.DS]*, 2023.

Hilal Asi and John Duchi. Near instance-optimality in differential privacy. *arXiv:2005.10630 [cs.CR]*, 2020a.

Hilal Asi and John C. Duchi. Instance-optimality in differential privacy via approximate inverse sensitivity mechanisms. In *Advances in Neural Information Processing Systems 33*, 2020b.

Hilal Asi, Jonathan Ullman, and Lydia Zakynthinou. From robustness to privacy and back. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 1121–1146, 2023.

Lucien Birgé. Model selection via testing: an alternative to (penalized) maximum likelihood estimators. In *Annales de l'IHP Probabilités et statistiques*, volume 42, pages 273–325, 2006.

Rudolf Borges and Johann Pfanzagl. A characterization of the one parameter exponential family of distributions by monotonicity of likelihood ratios. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 2(2):111–117, 1963.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: a Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

Lawrence D. Brown and Mark G. Low. A constrained risk inequality with applications to nonparametric functional estimation. *Annals of Statistics*, 24(6):2524–2535, 1996.

Mark Bun, Gautam Kamath, Thomas Steinke, and Zhiwei Steven Wu. Private hypothesis selection. In *Advances in Neural Information Processing Systems 32*, pages 156–167, 2019.

Tony Cai and Mark Low. A framework for estimating convex functions. *Statistica Sinica*, 25: 423–456, 2015.

Clément L. Canonne, Gautam Kamath, Audra McMillan, Adam D. Smith, and Jonathan R. Ullman. The structure of optimal private tests for simple hypotheses. *arXiv:1811.11148 [cs.DS]*, 2018.

Travis Dick, Alex Kulesza, Ziteng Sun, and Ananda Theertha Suresh. Subset-based instance optimality in private estimation. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pages 7992–8014, 2023.

David L. Donoho and Richard C. Liu. The" automatic" robustness of minimum distance functionals. *Annals of Statistics*, 16(2):552–586, 1988.

John C. Duchi. Introductory lectures on stochastic convex optimization. In *The Mathematics of Data*, IAS/Park City Mathematics Series. American Mathematical Society, 2018.

John C. Duchi and Feng Ruan. The right complexity measure in locally private estimation: It is not the Fisher information. *arXiv:1806.05756 [stat.TH]*, 2018a.

John C. Duchi and Feng Ruan. The right complexity measure in locally private estimation: It is not the Fisher information. *arXiv:1806.05756v1 [stat.TH]*, 2018b. URL https://arXiv.org/abs/1806.05756v1.

John C. Duchi and Feng Ruan. A constrained risk inequality for general losses. In *Proceedings of the 23nd International Conference on Artificial Intelligence and Statistics*, 2020.

John C. Duchi, Michael I. Jordan, and Martin J. Wainwright. Minimax optimal procedures for locally private estimation (with discussion). *Journal of the American Statistical Association*, 113 (521):182–215, 2018.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the Third Theory of Cryptography Conference*, pages 265–284, 2006.

Samuel B. Hopkins, Gautam Kamath, Mahbod Majid, and Shyam Narayanan. Robustness implies privacy in statistical estimation. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, 2022.

Ziyue Huang, Yuting Liang, and Ke Yi. Instance-optimal mean estimation under differential privacy. In *Advances in Neural Information Processing Systems 34*, volume 34, pages 25993–26004, 2021.

Peter J. Huber. A robust version of the probability ratio test. *Annals of Mathematical Statistics*, 36:1753–1758, 1965. URL https://api.semanticscholar.org/CorpusID:121454339.

Audra McMillan, Adam Smith, and Jon Ullman. Instance-optimal differentially private estimation. *arXiv:2210.15819 [math.ST]*, 2022.

Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *48th Annual Symposium on Foundations of Computer Science*, 2007.

Adam Smith. Privacy-preserving statistical estimation with optimal convergence rates. In *Proceedings of the Forty-Third Annual ACM Symposium on the Theory of Computing*, pages 813–822. ACM, 2011.

Charles Stein. Efficient nonparametric testing and estimation. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, pages 187–195, 1956.

S Zacks. Uniformly most accurate upper tolerance limits for monotone likelihood ratio families of discrete distributions. *Journal of the American Statistical Association*, 65(329):307–316, 1970.

Banghua Zhu, Jiantao Jiao, and Jacob Steinhardt. Generalized resilience and robust statistics. *Annals of Statistics*, 50(4):2256–2283, 2022.

Appendix A. Helper Lemmas

In our upper bounds, we use the DKW inequality.

Lemma 20 (DKW inequality) *Let $S_1, \ldots, S_n \overset{\text{iid}}{\sim} P$ with a cumulative distribution function $F(s) = \Pr_{S\sim P}(S \le s)$. Let $F_n(s) = \frac{1}{n}\sum_{i=1}^n 1\{S_i \le s\}$ be the empirical distribution function. Then for all $\alpha > 0$,*

$$\Pr\left(\sup_{s\in\mathbb{R}} |F_n(s) - F(s)| > \alpha\right) \le 2e^{-2n\alpha^2}.$$

We use the following form of Bernstein's inequality from (Boucheron et al., 2013, Corollary 2.11).

Lemma 21 *(Bernstein's inequality) Suppose $X_1, \ldots, X_n$ are independent random variables with mean 0 and variance σ^2 such that $|X_i| \le b$ almost surely. Then for all $t > 0$,*

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n X_i \ge t\right) \exp\left(-\frac{nt^2}{2(\sigma^2 + bt/3)}\right).$$

Lemma 22 *For distributions P, P' such that $\|P - P'\|_{\text{TV}} \le \eta$ and a function f such that $|f(t)| \le M$ for all M, we have $|\mathbb{E}_P[f] - \mathbb{E}_{P'}[f]| \le 2M\eta$.*

Proof Let μ be a dominating measure of P, P' and let p, p' be the densities of P and P', respectively, with respect to μ. Then,

$$\begin{aligned}
|\mathbb{E}_P[f] - \mathbb{E}_{P'}[f]| &= \left|\int f(x)(p(x) - p'(x))d\mu(x)\right| \\
&\le \int |f(x)||p(x) - p'(x)|d\mu(x) \\
&\le M\int |p(x) - p'(x)|d\mu(x) \\
&= 2M\eta,
\end{aligned}$$

proving the claim. ■

Appendix B. Properties of D_ε

In this section we study and prove several properties of the functional D_ε. Recall that

$$D_\varepsilon(P, Q) := \int (P(x) - Q(x)) \left[\log \frac{P(x)}{Q(x)}\right]_{-\varepsilon}^{\varepsilon} dx.$$

We begin by showing that D_ε can be expressed in a form similar to f-divergences.

Lemma 23 (Divergence form of D_ε) *For any distributions P and Q,*

$$D_\varepsilon(P, Q) = \int f_\varepsilon\left(\frac{P(x)}{Q(x)}\right) Q(x)dx.$$

In the above, the function $f_\varepsilon : \mathbb{R}_+ \to \mathbb{R}$ is defined piecewise by:

$$f_\varepsilon(t) = (t-1)[\log t]_{-\varepsilon}^{\varepsilon} = \begin{cases} -\varepsilon(t-1) & t \in [0, e^{-\varepsilon}) \\ \log(t)(t-1) & t \in [e^{-\varepsilon}, e^{+\varepsilon}] \\ +\varepsilon(t-1) & t \in (e^{+\varepsilon}, +\infty] \end{cases}.$$

Proof By definition of D_ε,

$$\begin{aligned} D_\varepsilon(P, Q) &= \int (P(x) - Q(x)) \left[\log \frac{P(x)}{Q(x)}\right]_{-\varepsilon}^{\varepsilon} dx \\ &= \int \left(\frac{P(x)}{Q(x)} - 1\right) \left[\log \frac{P(x)}{Q(x)}\right]_{-\varepsilon}^{\varepsilon} Q(x)dx \\ &= \int f_\varepsilon \left(\frac{P(x)}{Q(x)}\right) Q(x)dx. \end{aligned}$$

■

Moreover, D_ε enjoys many properties of distance metrics.

Lemma 24 (Properties of D_ε) *For any distributions P, Q:*

1. *(Non-negativity) $D_\varepsilon(P, Q) \geq 0$.*
2. *(Positivity) $D_\varepsilon(P, Q) = 0$ if and only if $P = Q$.*
3. *(Finiteness) $D_\varepsilon(P, Q) \leq 2\varepsilon \left\|P - Q\right\|_{\mathrm{TV}} < \infty$.*
4. *(Symmetry) $D_\varepsilon(P, Q) = D_\varepsilon(Q, P)$.*

Proof To prove non-negativity, note that $f_\varepsilon(t) \geq 0$ is nonnegative. Thus,

$$D_\varepsilon(P, Q) = \int f_\varepsilon \left(\frac{P(x)}{Q(x)}\right) Q(x)dx \geq 0.$$

The positivity follows because $f_\varepsilon(t) > 0$ for every $t \neq 1$.

To prove that $D_\varepsilon(P, Q) < \infty$, note that $|f_\varepsilon(t)| \leq \varepsilon|t - 1|$ for every t. Thus,

$$D_\varepsilon(P, Q) = \int f_\varepsilon \left(\frac{P(x)}{Q(x)}\right) Q(x)dx \leq \varepsilon \int |P(x) - Q(x)|dx = 2\varepsilon \left\|P - Q\right\|_{\mathrm{TV}} < \infty.$$

To prove symmetry, note that

$$\left[\log \frac{P(x)}{Q(x)}\right]_{-\varepsilon}^{\varepsilon} = -\left[\log \frac{Q(x)}{P(x)}\right]_{-\varepsilon}^{\varepsilon}.$$

Therefore,

$$\begin{aligned} D_\varepsilon(P, Q) &= \int (P(x) - Q(x)) \left[\log \frac{P(x)}{Q(x)}\right]_{-\varepsilon}^{\varepsilon} dx \\ &= \int (Q(x) - P(x)) \left[\log \frac{Q(x)}{P(x)}\right]_{-\varepsilon}^{\varepsilon} dx = D_\varepsilon(Q, P). \end{aligned}$$

■

Moreover, the functional D_ε is closely connected to the total variation distance.

Lemma 25 *For all distributions P, Q:*

$$2\varepsilon \left\|P - Q\right\|_{\mathrm{TV}} - \varepsilon(e^{\varepsilon} - 1) \leq D_{\varepsilon}(P, Q) \leq 2\varepsilon \left\|P - Q\right\|_{\mathrm{TV}}.$$

Proof The upper bound follows from Lemma 24. For the lower bound, consider the function $h_{\varepsilon}(t) = \varepsilon|t - 1| - \varepsilon|e^{\varepsilon} - 1|$. Note for all $t \in [e^{-\varepsilon}, e^{\varepsilon}]$, $h_{\varepsilon}(t) \leq 0 \leq f_{\varepsilon}(t)$. On the other hand, for $t \leq e^{-\varepsilon}$, $h'_{\varepsilon}(t) = f_{\varepsilon}(t) = -\varepsilon$ and for $t \geq e^{\varepsilon}$, $h'_{\varepsilon}(t) = f_{\varepsilon}(t) = \varepsilon$. Thus, we can conclude that for all $t > 0$, $h_{\varepsilon}(t) \leq f_{\varepsilon}(t)$. Therefore,

$$\begin{aligned} D_{\varepsilon}(P, Q) &= \int f_{\varepsilon}\left(\frac{P(x)}{Q(x)}\right) Q(x)dx \\ &\geq \int h_{\varepsilon}\left(\frac{P(x)}{Q(x)}\right) Q(x)dx = 2\varepsilon \left\|P - Q\right\|_{\mathrm{TV}} - \varepsilon(e^{\varepsilon} - 1). \end{aligned}$$

■

Additionally, D_{ε} is Lipschitz with respect to the total variation distance.

Lemma 26 (Lipschitzness w.r.t. TV Metric) *For any given distribution Q:*

$$|D_{\varepsilon}(P_1, Q) - D_{\varepsilon}(P_2, Q)| \leq 4(e^{\varepsilon} - 1) \left\|P_1 - P_2\right\|_{\mathrm{TV}}$$

holds for all distributions P_1, P_2.

Proof Note that f_{ε} is $2(e^{\varepsilon} - 1)$ Lipschitz. To see this, the derivative of f_{ε}, denoted as f'_{ε}, is piecewise defined with expressions:

$$f'_{\varepsilon}(t) = \begin{cases} -\varepsilon & t \in [0, e^{-\varepsilon}) \\ \log(t) + 1 - \frac{1}{t} & t \in (e^{-\varepsilon}, e^{+\varepsilon}) \\ +\varepsilon & t \in (e^{+\varepsilon}, +\infty) \end{cases}.$$

Using standard calculus, we can show that the derivative $|f'_{\varepsilon}(t)|$ is always bounded by $2(e^{\varepsilon} - 1)$ on its domain. Since f_{ε} is continuous, this proves that f_{ε} is $2(e^{\varepsilon} - 1)$ Lipschitz.

As its consequence, we deduce:

$$\begin{aligned} |D_{\varepsilon}(P_1, Q) - D_{\varepsilon}(P_2, Q)| &= \left| \int f_{\varepsilon}\left(\frac{P_1(x)}{Q(x)}\right) - f_{\varepsilon}\left(\frac{P_2(x)}{Q(x)}\right) Q(x)dx \right| \\ &\leq 2(e^{\varepsilon} - 1) \int \left| \frac{P_1(x)}{Q(x)} - \frac{P_2(x)}{Q(x)} \right| Q(x)dx = 4(e^{\varepsilon} - 1) \left\|P_1 - P_2\right\|_{\mathrm{TV}}. \end{aligned}$$

This completes the proof. ■

Lemma 27 (Triangle-like inequality of D_{ε}) *For $\varepsilon \leq 1$,*

$$D_{\varepsilon}(P, Q) \leq C(D_{\varepsilon}(P, R) + D_{\varepsilon}(R, Q))$$

Proof Here, I am taking $g_\varepsilon(t) = \int_1^t [\log s]_{-\varepsilon}^{\varepsilon} ds$, not the usual g_ε, but should be similar. This is so that $g_\varepsilon(0) = 1$ and $g_\varepsilon'(t) = [\log t]_{-\varepsilon}^{\varepsilon}$.

We can express the desired inequality as

$$\begin{aligned}\int f_\varepsilon\left(\frac{dQ}{dP}\right) dP &\le C\int f_\varepsilon\left(\frac{dR}{dP}\right) dP + C\int f_\varepsilon\left(\frac{dQ}{dR}\right) dR \\ &= \int C\left(f_\varepsilon\left(\frac{dR}{dP}\right) + f_\varepsilon\left(\frac{dQ}{dP}\cdot\frac{dP}{dR}\right)\frac{dR}{dP}\right) dP.\end{aligned}$$

Therefore it suffices to show that for all $x, y > 0$,

$$f_\varepsilon(x) \le C(f_\varepsilon(y) + f_\varepsilon(x/y)y).$$

We first define $g(t) = \int_1^t [\log s]_{-\varepsilon}^{\varepsilon} ds$ and show that there exists universal constants $c, c' > 0$ such that $cg(t) \le f_\varepsilon(t) \le c'g(t)$ for all $t > 0$. After showing this, it suffices to show that for all $x, y > 0$,

$$g(x) \le C(g(y) + g(x/y)y). \tag{6}$$

To show $g(t) = \Theta(f_\varepsilon(t))$, we first note that $g(1) = f_\varepsilon(1) = 0$, and so it suffices to show $c \le \frac{f'(t)}{g'(t)} \le c'$ for all $t > 0$. For $t < e^{-\varepsilon}$, $f'(t) = g'(t) = -\varepsilon$ and so $\frac{f'(t)}{g'(t)} = 1$. Similarly, for $t > e^{\varepsilon}$, $f'(t) = g'(t) = \varepsilon$ and so $\frac{f'(t)}{g'(t)} = 1$. For $t \in [e^{-\varepsilon}, e^{\varepsilon}]$, $f'(t) = \log t + 1 - \frac{1}{t}$ and $g'(t) = \log t$. This means that the ratio of the derivatives is

$$\frac{f'(t)}{g'(t)} = 1 + \frac{1}{\log t} - \frac{1}{t\log t},$$

which standard calculus shows is between 1 and 3 for all $\varepsilon \le 1$ and $t \in [e^{-\varepsilon}, e^{\varepsilon}]$. Thus, picking $c < 1$ and $c > 3$ gives that $cg(t) \le f(t) \le c'g(t)$, and now we turn to showing (6).

Fixing $y > 0$, we define $h(x) = g(y) + g(x/y)y$. First note that $h(y) = g(y)$ and we have derivatives

$$\begin{aligned}g'(x) &= [\log x]_{-\varepsilon}^{\varepsilon} \\ h'(x) &= [\log x - \log y]_{-\varepsilon}^{\varepsilon}.\end{aligned}$$

Consider the case $y \ge 1$. Then $\log y > 0$ and so $h'(x) \le g'(x)$ for all $x > 0$. This combined with $h(y) = g(y)$ shows that $h(x) \ge g(x)$ for all $x \le y$. For $x > y$, we will show that for some $\bar{y} > y$, $g(x) \le Ch(y)$ for $y \le x < \bar{y}$ and $g'(x) \le Ch'(y)$ for $x > \bar{y}$, proving that $g(x) \le Ch(x)$ for all $x \ge y$.

For $x \ge e^{\varepsilon}y$ we have $g'(x) = h'(x) = \varepsilon$ and for $x \ge y^2$ we have $g'(x) \le 2h'(x)$. Now take $\bar{y} = \min\{e^{\varepsilon}, y\}y$, so that $g'(x) \le Ch'(x)$ for $x > \bar{y}$ as long as $C > 2$. Because g and h are increasing away from 1 and $g(y) = h(y)$, if we show that $g(\bar{y}) \le Cg(y)$, then we will have for $y \le x \le \bar{y}$ that $g(x) \le g(\bar{y}) \le Cg(y) \le Ch(x)$. To show this, note that

$$\frac{g(\bar{y})}{g(y)} = \frac{(\min\{e^{\varepsilon}, y\}y - 1)[\min\{\varepsilon, \log y\} + \log y]_{-\varepsilon}^{\varepsilon}}{(y-1)[\log y]_{-\varepsilon}^{\varepsilon}}$$

which is bounded above by a constant for all $y > 1$ using standard calculus.

Now consider the case $y \le 1$. Then $\log y < 0$ and so $h'(x) \ge g'(x)$ for all $x > 0$ and hence $h(x) \ge g(x)$ for all $x \ge y$. For $x < y$, we will show that for some $\bar{y} < y$, $g(x) \le Ch(y)$ for $\bar{y} < x \le y$ and $g'(x) \ge Ch'(y)$ for $0 < x \le \bar{y}$, proving that $g(x) \le Ch(x)$ for all $x \le y$.

For $x \le e^{-\varepsilon}y$ we have $g'(x) = h'(x) = -\varepsilon$ and for $x \le y^2$ we have $g'(x) \ge 2h'(x)$. Now take $\bar{y} = \max\{e^{-\varepsilon}, y\}y$, so that $g'(x) \ge Ch'(x)$ for $x < \bar{y}$ as long as $C < 2$. Again because g and h are increasing away from 1 and $g(y) = h(y)$, if we show that $g(\bar{y}) \le Cg(y)$, then we will have for $\bar{y} \le x \le y$ that $g(x) \le g(\bar{y}) \le Cg(y) \le Ch(x)$. Then

$$\frac{g(\bar{y})}{g(y)} = \frac{(\max\{e^{-\varepsilon}, y\}y - 1)[\max\{-\varepsilon, \log y\} + \log y]_{-\varepsilon}^{\varepsilon}}{(y-1)[\log y]_{-\varepsilon}^{\varepsilon}}$$

which is bounded above by a constant for all $y < 1$ using standard calculus.

Therefore, we have that for all $y > 0$ that $g(x) \le Ch(x)$, which shows (6) and hence proves the claim. ■

Appendix C. Binary hypothesis testing

The functional D_ε is closely related to the minimax optimal sample size needed for binary hypothesis testing under differential privacy constraints.

Consider the setting where we have i.i.d. samples $S_1, S_2, \ldots, S_n$ from either P or Q. We are interested in ε-differentially private testing strategy $T : \mathbb{S}^n \to \{0, 1\}$ that minimizes the sum of the type I and type II error:

$$\mathfrak{E}_n(P, Q) = \inf_T P(T(\mathcal{S}) \ne 1) + Q(T(\mathcal{S}) \ne 0).$$

The minimax sample complexity for an ε-differentially private testing problem between two distributions is then determined by finding the smallest sample size n necessary to ensure the sum of the type I and type II errors is less than $1/3$ (here $1/3$ can be any absolute constant in $(0, 1)$):

$$\mathfrak{N}(P, Q) = \min\{n \in \mathbb{N} : \mathfrak{E}_n(P, Q) \le 1/3\}.$$

Canonne et al. (2018) determines the minimax sample complexity size $\mathfrak{N}$ up to constants for every private binary hypothesis problem. For any given distribution P and Q, they first define

$$\tau = \max\left\{\int \max\{P - e^{\varepsilon}Q, 0\}, \int \max\{Q - e^{\varepsilon}P, 0\}\right\}.$$

Suppose τ is attained by the second argument (without loss of generality) and let $\varepsilon' \ge 0$ be such that $\tau = \int \max\{P - e^{\varepsilon'}Q, 0\}$. They then further define two probability distributions P', Q' by setting $dP' \propto \min\{dP/dQ, e^{\varepsilon}\}dQ$ and $dQ' \propto \min\{dQ/dP, e^{\varepsilon}\}dP$. Finally, they set

$$U_\varepsilon(P, Q) = \tau\varepsilon + (1 - \tau)H^2(P', Q').$$

Theorem 28 *(Canonne et al. (2018)) For $\varepsilon = O(1)$:*

$$\mathfrak{N}(P, Q) = \Theta\left(1/U_\varepsilon(P, Q)\right).$$

Although these definitions are apparently different, we demonstrate that $D_\varepsilon(P, Q)$ and $U_\varepsilon(P, Q)$ are respectively upper and lower bounded by each other, with constants solely determined by ε. In particular, we have the following result.

Theorem 29 *For $\varepsilon = O(1)$,*

$$D_\varepsilon(P, Q) = \Theta(U_\varepsilon(P, Q)).$$

C.1. Proof of Theorem 29

To prove this result, we introduce an intermediate quantity: for any two probability distribution P and Q, we define

$$\begin{aligned} m_\varepsilon(P,Q) = \min_{P',P'',Q',Q'',\tau\in[0,1]} \; & \tau\varepsilon \left\|P''-Q''\right\|_{\mathrm{TV}} + (1-\tau)H^2(P',Q') \\ \text{subject to } \; & P=\tau P''+(1-\tau)P',\;\; Q=\tau Q''+(1-\tau)Q', \\ & \left\|dP'/dQ'\right\|_\infty \vee \left\|dQ'/dP'\right\|_\infty \le e^\varepsilon. \end{aligned}$$

We shall prove that $D_\varepsilon(P,Q), U_\varepsilon(P,Q), m_\varepsilon(P,Q)$ are respectively upper and lower bounded by each other, with constants solely determined by by ε.

Proposition 30 *For $\varepsilon = O(1)$,*

$$D_\varepsilon(P,Q) = \Theta(m_\varepsilon(P,Q)) = \Theta(U_\varepsilon(P,Q)).$$

Proof We prove in Lemma 31 that $D_\varepsilon(P,Q) = O(m_\varepsilon(P,Q))$, and in Lemma 32 that $m_\varepsilon(P,Q) \le U_\varepsilon(P,Q)$, and $U_\varepsilon(P,Q) = O(D_\varepsilon(P,Q))$ in 33. ■

Lemma 31 *For $\varepsilon = O(1)$,*

$$D_\varepsilon(P,Q) \le 2(e^\varepsilon+1)m_\varepsilon(P,Q).$$

Proof Consider any convex decomposition of the measure:

$$P=\tau P''+(1-\tau)P',\;\; Q=\tau Q''+(1-\tau)Q'$$

with $\|dP'/dQ'\|_\infty \vee \|dQ'/dP'\|_\infty \le e^\varepsilon$ and $\tau\in[0,1]$. We prove for any such convex decomposition:

$$D_\varepsilon(P,Q) \le 4(e^\varepsilon+1)\left(\tau\left\|P''-Q''\right\|_{\mathrm{TV}} + (1-\tau)d^2_{\mathrm{hel}}(P',Q')\right). \tag{7}$$

The function $f_\varepsilon(t)$ is piecewise defined, not convex, and is non-differentiable at points $t=e^{-\varepsilon}$ and $t=e^\varepsilon$. For this reason, we introduce a function g_ε which is defined by:

$$g_\varepsilon(t) = \begin{cases} f_\varepsilon(e^{+\varepsilon}) + f'_\varepsilon(e^{+\varepsilon})(t-e^{+\varepsilon}) & t\in(e^{+\varepsilon},+\infty) \\ f_\varepsilon(t) & t\in[e^{-\varepsilon},e^{+\varepsilon}] \\ f_\varepsilon(e^{-\varepsilon}) + f'_\varepsilon(e^{-\varepsilon})(t-e^{-\varepsilon}) & t\in(-\infty,e^{-\varepsilon}) \end{cases}$$

where we abuse the notation, and use

$$f'_\varepsilon(e^\varepsilon) = \lim_{t\to(e^\varepsilon)^-} f_\varepsilon(t), \quad f'_\varepsilon(e^{-\varepsilon}) = \lim_{t\to(e^{-\varepsilon})^+} f_\varepsilon(t)$$

to denote the left and right derivative of f_ε at $e^{+\varepsilon}, e^{-\varepsilon}$ respectively. This continuously differentiable function g_ε agrees with f_ε when $t\in(e^{-\varepsilon},e^\varepsilon)$, and is linear within $(e^{+\varepsilon},+\infty)$ and $(-\infty,e^{-\varepsilon})$. Since $f_\varepsilon(t) = (t-1)\log(t)$ is convex when $t\in(e^{-\varepsilon},e^{+\varepsilon})$, this then implies that the derivative of g_ε is monotonically increasing in t, and thus g_ε is convex in t. Additionally, one may verify that

- $f_\varepsilon(t) \le g_\varepsilon(t)$ for every $t \in \mathbb{R}$. This is because $g_\varepsilon(1) = f_\varepsilon(1) = 0$, and $g'_\varepsilon(t) \ge f'_\varepsilon(t) \ge 0$ for every $t \ge 1$, and $g'_\varepsilon(t) \le f'_\varepsilon(t) \le 0$ for every $t \le 1$.
- $g_\varepsilon(t) \le 2(e^\varepsilon - 1)|t-1|$ for every $t \in \mathbb{R}$. This is because $g_\varepsilon(1) = 0$, and $\sup_t |g'_\varepsilon(t)| \le 2(e^\varepsilon - 1)$.

Let us go back to the proof of equation (7). Since $f_\varepsilon \le g_\varepsilon$, we obtain for every distribution P, Q:

$$D_\varepsilon(P,Q) = D_{f_\varepsilon}(P||Q) \le D_{g_\varepsilon}(P||Q).$$

Given that g_ε is convex, the divergence $D_{g_\varepsilon}(P||Q)$ is jointly convex in (P,Q). Furthermore, since P and Q are convex combinations of P'', P' and Q'', Q' respectively, it then follows that

$$D_{g_\varepsilon}(P||Q) \le \tau D_{g_\varepsilon}(P''||Q'') + (1-\tau) D_{g_\varepsilon}(P'||Q').$$

We further estimate the divergence measures on the right-hand side.

For the first term, by leveraging the bound $g_\varepsilon(t) \le 2(e^\varepsilon - 1)\varepsilon|t-1|$, we deduce

$$D_{g_\varepsilon}(P''||Q'') \le 4(e^\varepsilon - 1) \left\|P'' - Q''\right\|_{\mathrm{TV}}.$$

For the second term, since $dP'(x)/dQ'(x) \in [e^{-\varepsilon}, e^\varepsilon]$, and $g_\varepsilon(t) = (t-1)\log(t)$ for $t \in [e^{-\varepsilon}, e^\varepsilon]$, we obtain that

$$D_{g_\varepsilon}(P'||Q') = D_{\mathrm{KL}}(P'||Q') + D_{\mathrm{KL}}(Q'||P') \le 4(e^\varepsilon + 1) d^2_{\mathrm{hel}}(P', Q').$$

where in the last inequality we utilize the bound $(t-1)\log(t) \le 2(e^\varepsilon+1)(\sqrt{t}-1)^2$ for $t \in [e^{-\varepsilon}, e^\varepsilon]$.

By integrating the established bounds, we successfully demonstrate equation (7) as intended. ■

Lemma 32 *For $\varepsilon = O(1)$,*

$$m_\varepsilon(P,Q) \le U_\varepsilon(P,Q).$$

Proof This immediately follows from the definition that $m_\varepsilon(P,Q)$ takes minimum over all possible decompositions, while $U_\varepsilon(P,Q)$ is functional value from one possible decomposition. ■

Lemma 33 *For $\varepsilon = O(1)$,*

$$U_\varepsilon(P,Q) = O(D_\varepsilon(P,Q)).$$

Proof Our proof adapts from the arguments in the proof of Theorem 2.5 (Canonne et.al. 2019').

Let us pick $\varepsilon' \le \varepsilon$ such that

$$\tau = \int \max\{P - e^\varepsilon Q, 0\} = \int \max\{Q - e^{\varepsilon'} P, 0\}.$$

We partition the space $\mathcal{X}$ by:

$$\mathcal{G} = \{x | P(x) - e^\varepsilon Q(x) > 0\}, \;\; \mathcal{L} = \{x | Q(x) - e^{\varepsilon'} P(x) > 0\}, \;\; \mathcal{J} = \mathbb{S} \setminus (\mathcal{G} \cup \mathcal{L}).$$

Since $\varepsilon' \le \varepsilon$, we have by definition:

$$\begin{aligned} D_\varepsilon(P,Q) &= \int (P(x)-Q(x)) \left[\log \frac{P(x)}{Q(x)}\right]_{-\varepsilon}^{\varepsilon} dx \\ &\ge \int (P(x)-Q(x)) \left[\log \frac{P(x)}{Q(x)}\right]_{-\varepsilon'}^{\varepsilon} dx \qquad (8) \\ &= (P(\mathcal{G})-Q(\mathcal{G}))\varepsilon + \int_{\mathcal{J}} (P(x)-Q(x)) \left[\log \frac{P(x)}{Q(x)}\right]_{-\varepsilon'}^{\varepsilon} dx + (Q(\mathcal{L})-P(\mathcal{L}))\varepsilon'. \end{aligned}$$

Since our two probability distributions P', Q' are defined by $dP' \propto \min\{dP/dQ, e^\varepsilon\} dQ$ and $dQ' \propto \min\{dQ/dP, e^\varepsilon\} dP$, we thereby have the identity:

$$(1-\tau)P' = \min\{P, e^\varepsilon Q\}, \quad (1-\tau)Q' = \min\{Q, e^{\varepsilon'} P\}.$$

In particular, we can see that

$$\begin{aligned} P(\mathcal{G}) &= (1-\tau)P'(\mathcal{G}) + \tau, \quad Q(\mathcal{G}) = (1-\tau)Q'(\mathcal{G}) \\ Q(\mathcal{L}) &= (1-\tau)Q'(\mathcal{L}) + \tau, \quad P(\mathcal{L}) = (1-\tau)P'(\mathcal{L}) \end{aligned}.$$

Also, for every $x \in \mathcal{J}$, $P(x)$ agrees with $(1-\tau)P'(x)$ and $Q(x)$ agrees with $(1-\tau)Q'(x)$. Furthermore, for every $x \in \mathcal{J}$, $\log(P(x)/Q(x)) \in [-\varepsilon', \varepsilon]$.

Thereby, we have

$$\begin{aligned} P(\mathcal{G}) - Q(\mathcal{G}) &= (1-\tau)(P'(\mathcal{G}) - Q'(\mathcal{G})) + \tau \\ Q(\mathcal{L}) - P(\mathcal{L}) &= (1-\tau)(Q'(\mathcal{L}) - P'(\mathcal{L})) + \tau \end{aligned},$$

and

$$\int_{\mathcal{J}} (P(x)-Q(x)) \left[\log \frac{P(x)}{Q(x)}\right]_{-\varepsilon'}^{\varepsilon} dx = \int_{\mathcal{J}} (P'(x)-Q'(x)) \log \frac{P'(x)}{Q'(x)} dx.$$

Substituting bounds into inequality (8) reveals that $D_\varepsilon(P,Q)$ is lower bounded by

$$\begin{aligned} &(1-\tau)(P'(\mathcal{G})-Q'(\mathcal{G})) + (1-\tau)(Q'(\mathcal{L})-P'(\mathcal{L})) + \int_{\mathcal{J}} (P'(x)-Q'(x)) \log \frac{P'(x)}{Q'(x)} dx + \tau(\varepsilon+\varepsilon') \\ &= (1-\tau)(D_{\mathrm{KL}}(P'||Q') + D_{\mathrm{KL}}(Q'||P')) + \tau(\varepsilon+\varepsilon'). \end{aligned}$$

Since $D_{\mathrm{KL}}(P'||Q') \ge d_{\mathrm{hel}}^2(P', Q')$, this then yields the bound as desired:

$$D_\varepsilon(P,Q) \ge (1-\tau) d_{\mathrm{hel}}^2(P', Q') + \tau\varepsilon = U_\varepsilon(P,Q).$$

This completes the proof of the theorem. ■

Appendix D. Missing proofs and details for Section 2

In this section, we provide proofs for our lower bounds. We begin in Section D.1 where we prove our local minimax lower bounds, and then we prove our global minimax lower bound. We also demonstrate the connection of this lower bound to the existing total variation and hellinger lower bounds in Section D.4.

D.1. Proof of Theorem 6

A standard argument lower bounds the minimax error of ε-DP estimation by the ε-DP testing error. More specifically, consider V to be a random variable that have equal (prior) probability to be from P_0 or P_1, T_ε to be a testing function that satisfies ε-DP, then

$$\inf_{\mathcal{A}\in\mathcal{A}_\varepsilon}\max_{P\in\{P_0,P_1\}}\mathbb{E}_{\mathcal{A},\mathcal{S}\sim P^n}\left[|\mathcal{A}(\mathcal{S})-\theta(P)|\right]\geq\left|\frac{\theta(P_0)-\theta(P_1)}{2}\right|\inf_{T_\varepsilon}P(T_\varepsilon(\mathcal{S})\neq V)$$

By the definition of sample complexity $\mathfrak{N}(P_0,P_1)$ in Cannone et al. (2019), we know that when $n\leq\mathfrak{N}(P_0,P_1)$, for all ε-DP tests T_ε, we have

$$P(T_\varepsilon(\mathcal{S})\neq V)=\frac{1}{2}[P_0(T_\varepsilon(\mathcal{S})=\theta(P_1))+P_1(T_\varepsilon(\mathcal{S})=\theta(P_0))]\geq\frac{1}{6}$$

By Theorem 28 and Theorem 29, we have $\mathfrak{N}(P_1,P_0)=\Theta(1/D_\varepsilon(P_1,P_0))$. Therefore, for every P_1 chosen from the set of distribution

$$\mathcal{P}_1=\left\{D_\varepsilon(P_1,P_0)=O\left(\frac{1}{n}\right)\right\}$$

we have

$$\inf_{\mathcal{A}\in\mathcal{A}_\varepsilon}\max_{P\in\{P_0,P_1\}}\mathbb{E}_{\mathcal{A},\mathcal{S}\sim P^n}\left[|\mathcal{A}(\mathcal{S})-\theta(P)|\right]\geq\frac{1}{6}\left|\frac{\theta(P_0)-\theta(P_1)}{2}\right|$$

Now, taking supremum over $P_1\in\mathcal{P}_1$ on both sides yields

$$\mathfrak{M}_n^{\mathrm{loc}}(P_0;\mathcal{P},\mathcal{A}_\varepsilon)\geq\frac{1}{6}\omega_\varepsilon\left(O\left(\frac{1}{n}\right),P_0;\mathcal{P}\right)$$

We can also prove an upper bound on the local minimax complexity.

Lemma 34

$$\mathfrak{M}_n^{\mathrm{loc}}(P_0;\mathcal{P},\mathcal{A}_\varepsilon)\leq\frac{2}{3}\omega_\varepsilon\left(O\left(\frac{1}{n}\right),P_0;\mathcal{P}\right)$$

Proof Fixing any pair of distribution $P_0,P_1\in\mathcal{P}$, any ε-DP estimation algorithm $\mathcal{A}\in\mathcal{A}_\varepsilon$ can be used to build an ε-DP testing function $\tilde{T}_\varepsilon$. For instance, define $\tilde{T}_\varepsilon(\mathcal{S})=P_0$ if $|\mathcal{A}(\mathcal{S})-\theta(P_0)|\leq|\mathcal{A}(\mathcal{S})-\theta(P_1)|$, and $T_\varepsilon(\mathcal{S})=P_1$ otherwise.

Similarly, any ε-DP testing function T_ε can be used to build an ε-DP estimation algorithm $\tilde{\mathcal{A}}$. Namely, we just set $\tilde{\mathcal{A}}(\mathcal{S})=\theta(P_0)$ if $T_\varepsilon(\mathcal{S})=P_0$ and $\tilde{\mathcal{A}}(\mathcal{S})=\theta(P_1)$ otherwise.

Therefore, we see that ε-DP testing function and ε-DP estimation algorithm are equivalent to each other in binary testing.

By Canonne et al. (2018), we know that when $n=\Theta(1/U_\varepsilon)$, which by Theorem 29 is equivalent to $D_\varepsilon=\Theta(1/n)$, for all $P_1\in\mathcal{P}$, there exists ε-DP testing function T_ε that can distinguish between P_0 and P_1, namely,

$$P_0(T_\varepsilon(\mathcal{S})=\theta(P_1))+P_1(T_\varepsilon(\mathcal{S})=\theta(P_0))\leq\frac{1}{3}$$

Then by the equivalence between testing function and estimation algorithm discussed above, there exists ε-DP estimation algorithm $\tilde{\mathcal{A}}$ such that

$$\max_{P\in\{P_0,P_1\}} \mathbb{E}_{\tilde{\mathcal{A}},\mathcal{S}\sim P^n}\left[|\tilde{\mathcal{A}}(\mathcal{S})-\theta(P)|\right] \le \frac{1}{3}|\theta(P_0)-\theta(P_1)|$$

So then

$$\inf_{\mathcal{A}\in\mathcal{A}_\varepsilon}\max_{P\in\{P_0,P_1\}} \mathbb{E}_{\mathcal{A},\mathcal{S}\sim P^n}\left[|\mathcal{A}(\mathcal{S})-\theta(P)|\right] \le \frac{|\theta(P_0)-\theta(P_1)|}{3}$$

Next, by taking supremum over P_1 we have

$$\mathfrak{M}_n^{\rm loc}(P_0;\mathcal{P},\mathcal{A}_\varepsilon) \le \frac{2}{3}\omega_\varepsilon\left(O\left(\frac{1}{n}\right),P_0;\mathcal{P}\right)$$

■

D.2. Chi-squared Divergence Bound

Theorem 35 *Let P_0, P_1 be arbitrary distributions on a common space $\mathcal{X}$, $\mathcal{A}$ be any ε differential private channel (Def 1) and $M_\alpha(\cdot)=\int \mathcal{A}(\cdot\mid\mathcal{S})dP_\alpha(\mathcal{S})$ for $\alpha\in\{0,1\}$ be the marginal distribution on Θ. Consider any convex decomposition of the probability measure P_0, P_1:*

$$P_0=(1-\tau)P_0'+\tau P_0'', \quad P_1=(1-\tau)P_1'+\tau P_1''$$

obeying $\|dP_0''/dP_1''\|_\infty \vee \|dP_1''/dP_0''\| \le e^\varepsilon$. Here, P_0', P_0'', P_1', P_1'' are probability measures. $\tau\in[0,1]$. Then, when $\varepsilon=O(1)$, we have, for some constant C,

$$D_{\chi^2}(M_1||M_0) \le \exp(Cn(\varepsilon\tau+(1-\tau)d_{\rm hel}^2(P_1',P_0'))) - 1.$$

Our proof utilizes two different techniques in upper bounding the χ^2 divergence.

Lemma 36 *We have that*

$$D_{\chi^2}(M_1||M_0) \le \exp(nD_{\chi^2}(P_1||P_0)) - 1.$$

Proof By data processing inequality:

$$D_{\chi^2}(M_1||M_0) \le D_{\chi^2}(P_1^{\otimes n}||P_0^{\otimes n}).$$

The conclusion then follows as $D_{\chi^2}(P_1^{\otimes n}||P_0^{\otimes n})+1=(D_{\chi^2}(P_1||P_0)+1)^n \le \exp(nD_{\chi^2}(P_1||P_0))$. ■

D.2.1. PROOF OF THEOREM 35

For any pair of distribution $P_0, P_1 \in \mathcal{P}$, fixing ε, following Canonne et al. (2018)'s definition about $\tau_\varepsilon(P_0, P_1)$ (we write τ for abbreviation), we have

$$\tau = \max\left\{\int_{\mathcal{X}} \max\{P_0(x) - e^{\varepsilon}P_1(x), 0\}dx, \int_{\mathcal{X}} \max\{P_1(x) - e^{\varepsilon}P_0(x), 0\}dx\right\}.$$

Assuming without loss of generality that the first term above is the larger one, we can then let $0 < \varepsilon' < \varepsilon$ be the largest values such that

$$\int_{\mathcal{X}} \max\{P_0(x) - e^{\varepsilon}P_1(x), 0\}dx = \int_{\mathcal{X}} \max\{P_1(x) - e^{\varepsilon'}P_0(x), 0\}dx = \tau$$

Finally, we define

$$P_0^{'}(x) = \frac{\min\{P_0(x), e^{\varepsilon}P_1(x)\}}{1-\tau},\ P_0^{''}(x) = \frac{\max\{P_0(x) - e^{\varepsilon}P_1(x), 0\}}{\tau}$$

$$P_1^{'}(x) = \frac{\min\{P_1(x), e^{\varepsilon'}P_0(x)\}}{1-\tau},\ P_1^{''}(x) = \frac{\max\{P_1(x) - e^{\varepsilon'}P_0(x), 0\}}{\tau}.$$

So we can decompose

$$P_0 = (1-\tau)P_0' + \tau P_0'',\ P_1 = (1-\tau)P_1' + \tau P_1''$$

For $a \in \{0, 1\}$, if we generate $W_1, \ldots, W_n \overset{\text{iid}}{\sim} \text{Ber}(\tau)$, pick $S_i \sim P_a'$ when $W_i = 1$, and $S_i \sim P_a''$ when $W_i = 0$, then marginally we will have $S_1, \ldots, S_n \overset{\text{iid}}{\sim} P_a$. Basing on this construction, define set $N = \{i \in \{1, \ldots, n\} | W_i = 1\}$ and $N^c = [n] \setminus N$, then we can express the joint distribution as

$$P_a(X_{1:n}) = \sum_N \tau^{|N|}(1-\tau)^{|N^c|}P_a''(\mathcal{S}_N)P_a'(\mathcal{S}_{N^c})$$

For $\tau = 0$, the bound reduces to Lemma 36.

For $\tau > 0$, first note that

$$\begin{aligned} D_{\chi^2}(P_1'||P_0') &= \int \left(\frac{P_1'(x)}{P_0'(x)} - 1\right)^2 P_0'(x)dx \\ &= \int \left(\sqrt{\frac{P_1'(x)}{P_0'(x)}} - 1\right)^2 \left(\sqrt{\frac{P_1'(x)}{P_0'(x)}} + 1\right)^2 P_0'(x)dx. \end{aligned}$$

Then by our assumption that $\log(P_0'(x)/P_1'(x)) \in [-\varepsilon', \varepsilon]$, it follows that $\left(\sqrt{\frac{P_1'(x)}{P_0'(x)}}+1\right)^2 \le (e^{\varepsilon/2}+1)^2$ for all x and so $D_{\chi^2}(P_1'||P_0') \le 2(e^{\frac{\varepsilon}{2}}+1)^2 d_{\text{hel}}^2(P_1', P_0')$.

Then by Lemma 36,

$$D_{\chi^2}(M_1||M_0) \le \exp(2n(e^{\frac{\varepsilon}{2}}+1)^2 d_{\text{hel}}^2(P_1', P_0')) - 1 \le Cnd_{\text{hel}}^2(P_1', P_0').$$

For general case of $\tau \in [0,1]$. By decomposition of $P_a(\mathcal{S})$ discussed in the beginning of the section, we naturally have

$$M_a(z) = \sum_N \tau^{|N|}(1-\tau)^{|N^c|} M_{N,a}(z),$$

where $M_{N,a}$ is the marginal distribution of the channel that corresponds to a specific set N. As the Chi-squared divergence is convex in the pair of distribution, we then have

$$D_{\chi^2}(M_1||M_0) \le \sum_N \tau^{|N|}(1-\tau)^{|N^c|} D_{\chi^2}(M_{N,1}||M_{N,0})$$

Note that we have

$$D_{\chi^2}(M_{N,1}||M_{N,0}) = \int \frac{(M_{N,1}(z) - M_{N,0}(z))^2}{M_{N,0}(z)} dz \tag{9}$$

$$\le 2\int \frac{(M_{N,1}(z) - M_{N,2}(z))^2}{M_{N,0}(z)} dz + 2\int \frac{(M_{N,2}(z) - M_{N,0}(z))^2}{M_{N,0}(z)} dz, \tag{10}$$

where we define

$$M_{N,0}(z) = \int \mathcal{A}(z|\mathcal{S}) P_0''(\mathcal{S}_N) P_0'(\mathcal{S}_{N^c}) d\mathcal{S}$$

$$M_{N,1}(z) = \int \mathcal{A}(z|\mathcal{S}) P_1''(\mathcal{S}_N) P_1'(\mathcal{S}_{N^c}) d\mathcal{S}$$

$$M_{N,2}(z) = \int \mathcal{A}(z|\mathcal{S}) P_0''(\mathcal{S}_N) P_1'(\mathcal{S}_{N^c}) d\mathcal{S}$$

For the first term in (10), we first define

$$\tilde{\mathcal{A}}(z|\mathcal{S}_N) = \int \mathcal{A}(z|\mathcal{S}) P_1'(\mathcal{S}_{N^c}) d\mathcal{S}_{N^c}$$

Then by privacy of $\mathcal{A}$, $\tilde{\mathcal{A}}$ is naturally ε-DP with respect to $\mathcal{S}_N$. Therefore, for all $\mathcal{S}_N$ and $\mathcal{S}_N'$ and all z, we have

$$\frac{\mathcal{A}(z|\mathcal{S}_N)}{\mathcal{A}(z|\mathcal{S}_N')} \le e^{\varepsilon|N|}$$

Letting $\mathcal{S}_N''$ be a fixed dataset, we have

$$\frac{M_{N,1}(z)}{M_{N,2}(z)} = \frac{\int \mathcal{A}(z|\mathcal{S}) P_1''(\mathcal{S}_N) P_1'(\mathcal{S}_{N^c}) d\mathcal{S}}{\int \mathcal{A}(z|\mathcal{S}) P_0''(\mathcal{S}_N) P_1'(\mathcal{S}_{N^c}) d\mathcal{S}} \le e^{2|N|\varepsilon} \frac{\int \tilde{\mathcal{A}}(z|\mathcal{S}_N'') P_1''(\mathcal{S}_N) d\mathcal{S}_N}{\int \tilde{\mathcal{A}}(z|\mathcal{S}_N'') P_0''(\mathcal{S}_N) d\mathcal{S}_N} = e^{2|N|\varepsilon}.$$

So then

$$|M_{N,1}(z) - M_{N,2}(z)|^2 \le (\exp\{4|N|\varepsilon\} - 1) M_{N,2}(z)^2$$

For the second term in (10), we apply Lemma 36 to get

$$D_{\chi^2}(M_{N,2}||M_{N,0}) \le \exp\left\{|N^c|\, D_{\chi^2}(P_1'||P_0')\right\} - 1$$

Now plugging these bounds to (10), we have

$$D_{\chi^2}(M_{N,1}||M_{N,0}) \le (\exp\{4|N|\varepsilon\} - 1)\exp\left\{|N^c|\, D_{\chi^2}(P_1'||P_0')\right\} + (\exp\left\{|N^c|\, D_{\chi^2}(P_1'||P_0')\right\} - 1) \tag{11}$$

$$= \exp(4|N|\varepsilon + |N^c| D_{\chi^2}(P_1'||P_0')) - 1 \tag{12}$$

So then by Cauchy-Schwarz

$$\begin{aligned} D_{\chi^2}(M_1||M_0) &\le \mathbb{E}[\exp(4|N|\varepsilon + |N^c| D_{\chi^2}(P_1'||P_0'))] - 1 \\ &\le \sqrt{\mathbb{E}[\exp(8|N|\varepsilon)]\mathbb{E}[\exp(2|N^c| D_{\chi^2}(P_1'||P_0'))]} - 1 \end{aligned}$$

Because $|N|$ and $|N^c|$ are binomial random variables, referring to the moment generating function of binomial random variables, we have

$$\begin{aligned} E[\exp(8|N|\varepsilon)] &= (1 + \tau(e^{8\varepsilon} - 1))^n \\ &\le \exp(n\tau(e^{8\varepsilon} - 1)) \\ &\le \exp(2C' n\tau\varepsilon) \end{aligned}$$

$$\begin{aligned} E[\exp(2|N^c| D_{\chi^2}(P_1'||P_0'))] &= (1 + (1-\tau)(e^{2D_{\chi^2}(P_1'||P_0')} - 1))^n \\ &\le \exp(n(1-\tau)(e^{2D_{\chi^2}(P_1'||P_0')} - 1)) \\ &\le \exp(2C'' n(1-\tau) D_{\chi^2}(P_1'||P_0')) \end{aligned}$$

for some constant C', C'', where we leverage $e^x - 1 \le e^c x$ when $x \in [0, c]$ in the last inequality with $\varepsilon = O(1)$ and $D_{\chi^2}(P_1'||P_0') \le e^{2\varepsilon} - 1 = O(1)$.

Then we have

$$D_{\chi^2}(M_1||M_0) \le \exp(C' n\tau\varepsilon + C'' n(1-\tau) D_{\chi^2}(P_1'||P_0')) - 1$$

Now, by assumption, we have for all x,

$$\frac{dP_1'(x)}{dP_0'(x)} \in [e^{-\varepsilon}, e^{\varepsilon}],$$

and so $D_{\chi^2}(P_1'||P_0') \le 2(e^{\varepsilon/2} + 1)^2 d_{\text{hel}}^2(P_1', P_0')$. Thus

$$\begin{aligned} D_{\chi^2}(M_1||M_0) &\le \exp(C' n\tau\varepsilon + C'' n(1-\tau)(e^{\varepsilon/2} + 1)^2 d_{\text{hel}}^2(P_1', P_0')) - 1 \\ &\le \exp(Cn(\tau\varepsilon + (1-\tau) d_{\text{hel}}^2(P_1', P_0'))) - 1 \end{aligned}$$

for some constant C because $\varepsilon = O(1)$, which finishes the proof.

D.3. Super-efficiency

To provide our general super-efficiency result, we use constrained risk inequality (Brown and Low, 1996; Duchi and Ruan, 2020). The next proposition shows that improvement over modulus of continuity lower bound at a point P_0 implies worse performance elsewhere. The key technical

ingredient central to the development of the proposition is a new information contraction inequality which relates the χ^2-divergence between the private measures M_0, M_1 to the form optimal sample complexity (Canonne et al., 2018) defined in Theorem 28, whose presentation we defer to Theorem 35 in Appendix D.2.

Before going to the main theorem, we first denote some of the constants used in other theorem and lemma. Suppose Theorem 35 is satisfied with constant C_1, Theorem 29 is satisfied with constant C_2 and C_3 (e.g. $C_2 D_\varepsilon \le U_\varepsilon \le C_3 D_\varepsilon$) and Lemma 27 is satisfied with constant C_4.

To state the result, we also need to specify an upper bound on the sensitivity of local modulus ω_{L,D_ε}. The exact value of this bound can be computed under specific distribution family $\mathcal{P}$; for instance, it is satisfied with the parametric family introduced in Example 1 with $\lambda_k = \sqrt{k}$.

Condition 1 (Sensitivity Bound) *For all $k, \delta \in \mathbb{R}_+$, there exists $\lambda_k < \infty$ such that*

$$\omega_\varepsilon\left(k\delta, P_0; \mathcal{P}\right) \le \lambda_k \omega_\varepsilon\left(\delta, P_0; \mathcal{P}\right).$$

Theorem 37 *Suppose Condition 1 is satisfied with value λ when $k = 2C_3C_4/C_2$. Let $\mathcal{A}$ be any ε differentially private channel (Def 1) with marginal distributions $M_a^n(\cdot) = \int \mathcal{A}(\cdot|\mathcal{S}) dP_a^n(\mathcal{S})$. If for some $\eta \in [0,1]$ and some constant $\kappa > C_1C_3$, the estimator $\mathcal{A}$ satisfies*

$$R(\mathcal{A}, \theta_0, M_0) \le \eta \cdot \omega_\varepsilon\left(\frac{1}{n\kappa}, P_0; \mathcal{P}\right),$$

then for all $t \in [0,1]$, there exists a distribution $P_1 \in \mathcal{P}$ and constant $\zeta = C_1C_2$ such that,

$$R(\mathcal{A}, \theta_1, M_1) \ge \frac{1}{4\lambda}\left[1 - \eta^{\frac{1-t}{2}}\right]_+^2 \omega_\varepsilon\left(\frac{t \log \frac{1}{\eta}}{n\zeta}, P_1; \mathcal{P}\right).$$

Roughly speaking, Theorem 37 states that, given a constant ε, for any small enough η, that if an estimator $\mathcal{A}$ is super-efficient at P_0, in that its risk is considerably smaller than the local modulus at P_0, then that same estimator $\mathcal{A}$ pays nontrivial price elsewhere.

D.3.1. PROOF OF THEOREM 37

The proof leverages a constrained risk inequality that extends Brown and Low (1996, Thm. 1). For any two probability measures P_0 and P_1 we define the 2-affinity

$$\rho\left(P_1 \| P_0\right) := D_{\chi^2}\left(P_1 \| P_0\right) + 1 = \mathbb{E}_{P_0}\left[\frac{dP_1^2}{dP_0^2}\right] = \mathbb{E}_{P_1}\left[\frac{dP_1}{dP_0}\right],$$

which measures the similarity between distributions P_0 and P_1. Lemma 38 states the constrained risk inequality formally, which essentially says that, if an estimator has small risk under P_0, then its risk at P_1 must be nearly the size of the distance between the associated parameters θ_0 and θ_1

Lemma 38 (Duchi and Ruan (2020, Theorem 1)) *Let $\theta_0 = \theta(P_0)$, $\theta_1 = \theta(P_1)$, and define $\Delta = |\theta_0 - \theta_1|/2$. If the estimator $\widehat{\theta}$ satisfies $R(\widehat{\theta}, \theta_0, P_0) \le \delta$ for some $\delta \ge 0$, then*

$$R(\widehat{\theta}, \theta_1, P_1) \ge \left[\Delta^{1/2} - \left(\rho(P_1 \| P_0) \cdot \delta\right)^{1/2}\right]_+^2.$$

For notation shorthand, we introduce $R_a(\widehat{\theta}) = R(\widehat{\theta}, \theta_a, M_a)$. By Lemma 38, for any distributions P_0 and P_1, we have

$$R_1(\widehat{\theta}) \geq \left[\left(\frac{1}{2}|\theta_0 - \theta_1|\right)^{\frac{1}{2}} - \left(\rho\left(M_1 \| M_0\right) \cdot R_0(\widehat{\theta})\right)^{1/2}\right]_+^2.$$

Following Canonne et al. (2018), we define

$$\tau = \max\left\{\int_{\mathcal{X}} \max\{P_0(x) - e^{\varepsilon} P_1(x), 0\} dx, \int_{\mathcal{X}} \max\{P_1(x) - e^{\varepsilon} P_0(x), 0\} dx\right\}.$$

Assuming without loss of generality that the first term above is the larger one, we then let $0 < \varepsilon' < \varepsilon$ be the largest values such that

$$\int_{\mathcal{X}} \max\{P_0(x) - e^{\varepsilon} P_1(x), 0\} dx = \int_{\mathcal{X}} \max\{P_1(x) - e^{\varepsilon'} P_0(x), 0\} dx = \tau_{\varepsilon}$$

Finally, we define

$$P_0^{'}(x) = \frac{\min\{P_0(x), e^{\varepsilon} P_1(x)\}}{1 - \tau}, \; P_1^{'}(x) = \frac{\min\{P_1(x), e^{\varepsilon'} P_0(x)\}}{1 - \tau}$$

Then by Theorem 35, we further have the following information contraction inequality

$$\rho\left(M_1 \| M_0\right) = D_{\chi^2}\left(M_1 \| M_0\right) + 1 \leq \exp(C_1(n\tau + n(1-\tau) d_{\mathrm{hel}}^2(P_1', P_0'))). \tag{13}$$

This information contraction inequality is the key to the proof of Theorem 37. The rest follows the same argument in Duchi and Ruan (2018b, Proposition 3).

For $t \in [0, 1]$, let $\mathcal{P}_t$ be the collection of distributions

$$\mathcal{P}_t := \left\{P \in \mathcal{P} \mid U_{\varepsilon} = \varepsilon\tau + (1-\tau) d_{\mathrm{hel}}^2(P', P_0') \leq \frac{t \log \frac{1}{\eta}}{2nC_1}\right\},$$

so that under the conditions of the proposition, any distribution $P_1 \in \mathcal{P}_t$ satisfies

$$R_1(\widehat{\theta}) \geq \left[\left(\frac{1}{2}|\theta_0 - \theta_1|\right)^{1/2} - \eta^{\frac{1-t}{2}} \omega_{\varepsilon}\left(\frac{1}{n\kappa}, P_0; \mathcal{P}\right)^{1/2}\right]_+^2. \tag{14}$$

Now, with inequalities (13) and (14), we see that for all $t \in [0, 1]$, there exists $P_1 \in \mathcal{P}_t$ such that

$$R_1(\widehat{\theta}) \geq \left[\omega_{U_{\varepsilon}}\left(\frac{t \log \frac{1}{\eta}}{nC_1}, P_0; \mathcal{P}\right)^{1/2} - \eta^{\frac{(1-t)}{2}} \omega_{\varepsilon}\left(\frac{1}{n\kappa}, P_0; \mathcal{P}\right)^{1/2}\right]_+^2.$$

Because $\delta \mapsto \omega(\delta, P_0; \mathcal{P})$ is non-decreasing, utilizing Theorem 29, if $t \in [0, 1]$ we may choose $P_1 \in \mathcal{P}_t$

$$\begin{aligned} R_1(\widehat{\theta}) &\geq \left[1 - \eta^{(1-t)/2}\right]_+^2 \omega_{\varepsilon}\left(\frac{t \log \frac{1}{\eta}}{nC_1C_3}, P_0; \mathcal{P}\right) \\ &\geq \frac{1}{\lambda}\left[1 - \eta^{(1-t)/2}\right]_+^2 \omega_{\varepsilon}\left(\frac{2C_4 t \log \frac{1}{\eta}}{nC_1C_2}, P_0; \mathcal{P}\right), \end{aligned} \tag{15}$$

where the last step is by Condition 1 and the definition of λ. Lastly, we lower bound the modulus of continuity at P_0 by a modulus at P_1. Following the same argument in the proof of Duchi and Ruan (2018b, Proposition 3), we claim that by Lemma 27, any P_1 with $D_\varepsilon(P_1, P_0) \le \delta$ satisfies

$$\omega_\varepsilon(2C_4\delta, P_0; \mathcal{P}) \ge \frac{1}{4}\omega_\varepsilon(\delta, P_1; \mathcal{P}). \tag{16}$$

Hence, if we take $\delta = t \log \frac{1}{\eta}/(n\zeta)$, all $P_1 \in \mathcal{P}_t$ satisfy $U_\varepsilon(P_0, P_1) \le C_2\delta$. By Theorem 29 they also satisfy $D_\varepsilon(P_0, P_1) \le \delta$. Then Eq. (15), inequality (16) imply for some $P_1 \in \mathcal{P}_t$

$$\begin{aligned} R_1(\widehat{\theta}) &\ge \frac{1}{\lambda}\left[1 - \eta^{(1-t)/2}\right]_+^2 \omega_\varepsilon\left(2C_4\delta, P_0; \mathcal{P}\right) \\ &\ge \frac{1}{4\lambda}\left[1 - \eta^{(1-t)/2}\right]_+^2 \omega_\varepsilon\left(\frac{t\log\frac{1}{\eta}}{n\zeta}, P_1; \mathcal{P}\right). \end{aligned}$$

Let us return to the claim (16). For distributions P_0, P_1, P_2 with parameters $\theta_a = \theta(P_a)$,

$$\left|\frac{\theta_1 - \theta_2}{2}\right| \le |\theta_0 - \theta_1| + |\theta_0 - \theta_2| \le 2\left|\frac{\theta_0 - \theta_1}{2}\right| + 2\left|\frac{\theta_0 - \theta_2}{2}\right|$$

Next for any $\delta \ge 0$, conditioned on $D_\varepsilon(P_0, P_1) \le \delta$, for any P that $D_\varepsilon(P_1, P) \le \delta$, by Lemma 27 we have $D_\varepsilon(P_0, P) \le C_4 D_\varepsilon(P_0, P_1) + C_4 D_\varepsilon(P_1, P) \le 2C_4\delta$, therefore

$$\begin{aligned} \omega_\varepsilon(2C_4\delta, P_0, \mathcal{P}) &= \sup_{D_\varepsilon(P_0,P)\le 2C_4\delta}\left|\frac{\theta_0 - \theta(P)}{2}\right| \ge \sup_{D_\varepsilon(P_1,P)\le\delta}\left|\frac{\theta_0 - \theta(P)}{2}\right| \\ &\ge \sup_{D_\varepsilon(P_1,P)\le\delta}\left\{\left|\frac{\theta_1 - \theta(P)}{4}\right| - \left|\frac{\theta_0 - \theta_1}{2}\right|\right\} \\ &\ge \frac{1}{2}\omega_\varepsilon(\delta, P_1, \mathcal{P}) - \omega_\varepsilon(\delta, P_0, \mathcal{P}) \end{aligned}$$

Rearranging, we have inequality (16), as for any distribution P_1 such that $D_\varepsilon(P_0, P_1) \le \delta$,

$$2\omega_\varepsilon(2C_4\delta, P_0, \mathcal{P}) \ge \omega_\varepsilon(\delta, P_0, \mathcal{P}) + \omega_\varepsilon(2C_4\delta, P_0, \mathcal{P}) \ge \frac{1}{2}\omega_\varepsilon(\delta, P_1, \mathcal{P})$$

D.4. Connection with Total Variation Bound and Hellinger Distance Bound

In this section, we compare our new D_ε-based lower bound to existing lower bounds based on the total variation metric and the Hellinger metric McMillan et al. (2022). First, we show in Proposition 39 that our lower bound is always larger than the one in McMillan et al. (2022). Morevoer,

we show our D_ε-based lower bound is always larger than existing lower bounds based on the total variation metric and the Hellinger metric McMillan et al. (2022). We provide an example (see Example 1) where our lower bound can be arbitrarily larger than existing lower bounds.

To this end, recall the local modulus of continuity with respect to the total variation metric and the Hellinger metric:

$$\omega_{\mathrm{TV}}(\delta, P_0; \mathcal{P}) = \sup_{P_1\in\mathcal{P}}\left\{|\theta(P_1) - \theta(P_0)| \mid \|P_1 - P_0\|_{\mathrm{TV}} \le \delta\right\}.$$

$$\omega_{d_{\mathrm{hel}}}(\delta, P_0; \mathcal{P}) = \sup_{P_1\in\mathcal{P}}\left\{|\theta(P_1) - \theta(P_0)| \mid d_{\mathrm{hel}}(P_1||P_0) \le \delta\right\}.$$

Proposition 39 *Let $\varepsilon \leq 1$ and $P_0 \in \mathcal{P}$. Then*

$$\omega_\varepsilon\left(\frac{9}{n}, P_0; \mathcal{P}\right) \geq \omega_{\mathrm{TV}}\left(\frac{1}{n\varepsilon}, P_0; \mathcal{P}\right) \vee \omega_{d_{\mathrm{hel}}}\left(\frac{1}{\sqrt{n}}, P_0; \mathcal{P}\right)$$

Proof Fixing P_0 and for any $P_1 \in \mathcal{P}$, by Lemma 24 we have

$$D_\varepsilon(P_0, P_1) \leq 2\varepsilon \left\|P_0 - P_1\right\|_{\mathrm{TV}}.$$

Then, $\|P_0 - P_1\|_{\mathrm{TV}} = \frac{1}{n\varepsilon}$ implies that $D_\varepsilon(P_0, P_1) = \frac{2}{n}$. Taking supremum over P_1's that satisfy the condition, we get

$$\omega_\varepsilon\left(\frac{2}{n}, P_0; \mathcal{P}\right) \geq \omega_{\mathrm{TV}}\left(\frac{1}{n\varepsilon}, P_0; \mathcal{P}\right)$$

For the other side, consider the divergence form of the function D_ε from (23) and the divergence from of d^2_{hel}. Then we have

$$D_\varepsilon(P, Q) = \int f_\varepsilon\left(\frac{P(x)}{Q(x)}\right) Q(x)dx.$$

$$d^2_{\mathrm{hel}}(P, Q) = \int f_{\mathrm{hel}}\left(\frac{P(x)}{Q(x)}\right) Q(x)dx.$$

where $f_\varepsilon(t) = (t-1)[\log t]^\varepsilon_{-\varepsilon}$ and $f_{\mathrm{hel}}(t) = \frac{1}{2}(\sqrt{t}-1)^2$.

When $t \geq 0$, we can piece-wisely compare the function values and deduct that

$$\frac{f_\varepsilon(t)}{f_{\mathrm{hel}}(t)} \leq \frac{2\varepsilon(e^{\frac{\varepsilon}{2}}+1)}{e^{\frac{\varepsilon}{2}}-1} \leq \frac{2(e^{\frac{1}{2}}+1)}{e^{\frac{1}{2}}-1} \leq 9,$$

for $\varepsilon \leq 1$. Thus we have $D_\varepsilon(P, Q) \leq 9d^2_{\mathrm{hel}}(P, Q)$. Therefore if $d_{\mathrm{hel}}(P, Q) = \frac{1}{\sqrt{n}}$ then we get that $D_\varepsilon(P, Q) \leq \frac{9}{n}$. Overall taking supremum over P_1 we have

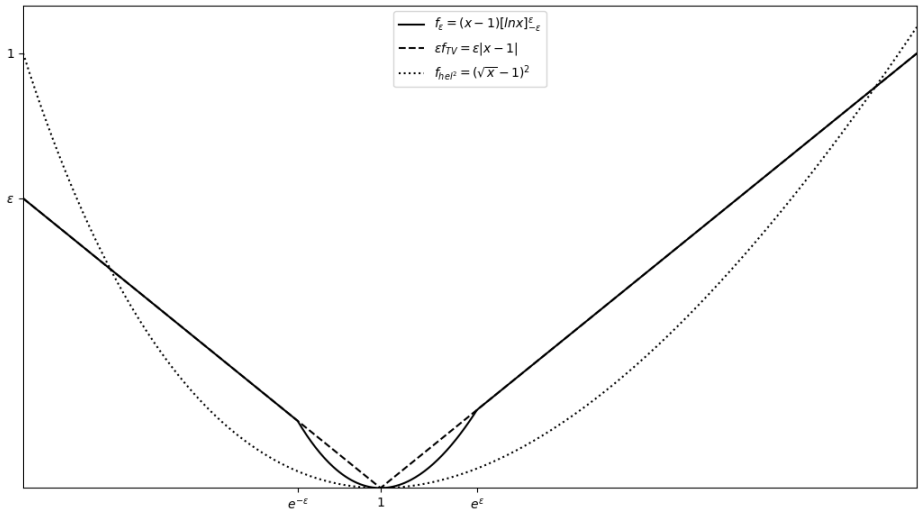
$$\omega_\varepsilon\left(\frac{9}{n}, P_0; \mathcal{P}\right) \geq \omega_{d_{\mathrm{hel}}}\left(\frac{1}{\sqrt{n}}, P_0; \mathcal{P}\right)$$

■

Next, we will provide an example where we can strictly separate the bounds in Proposition 39.

Example 1 *We first share some intuition about how to build this example. For the lower bounds introduced in Proposition 39, when fixing a bounding value of $\frac{1}{n}$ on the metrics, we are essentially comparing f-divergence metrics basing on the following functions (disregard the constants):*

$$f_\varepsilon(x) = (x-1)[\log x]^\varepsilon_{-\varepsilon},\ \varepsilon f_{TV}(x) = \varepsilon|x-1|,\ \textit{and}\ f_{hel^2}(x) = (\sqrt{x}-1)^2$$



From the plot of these functions we see that to make D_ε (which based on f_ε) much smaller than $\varepsilon\times$ TV distance and d^2_{hel}, we can
(1) pick some mass p_1 with likelihood ratio value $x_1 \in [e^{-\varepsilon}, e^{\varepsilon}]$ so then $f_\varepsilon(x_1)p_1 << \varepsilon f_{TV}(x_1)p_1$.
(2) pick some mass p_2 with likelihood ratio value $x_2 \approx 0$ or $x_2 >> e^{\varepsilon}$ so then $f_\varepsilon(x_2)p_2 << f_{hel^2}(x_2)p_2$.
(3) finally to ensure that the relative size comparison is still valid after considering the sum of separate parts above, we want $f_\varepsilon(x_1)p_1 \approx f_\varepsilon(x_2)p_2$.

Basing on these ideas, we consider a parametric family of distributions $\{P_\theta\}_{\theta\in\mathbb{R}}$, where $P_0(x) = I_{x\in[0,1]}$ and

$$P_\delta(x) = \begin{cases} \frac{1}{2}, & x \in \left[0, \frac{\varepsilon\delta^2}{1+2\varepsilon\delta^2}\right] \\ 1-\varepsilon|\delta|, & x \in \left(\frac{\varepsilon\delta^2}{1+2\varepsilon\delta^2}, \frac{1}{2}\right] \\ 1+\varepsilon|\delta|, & x \in \left(\frac{1}{2}, \frac{1+\varepsilon\delta^2}{1+2\varepsilon\delta^2}\right] \\ \frac{3}{2}, & x \in \left(\frac{1+\varepsilon\delta^2}{1+2\varepsilon\delta^2}, 1\right] \end{cases}$$

Then making use of approximation $\log(1+x) \approx x$ and $\sqrt{1+x} \approx 1 + \frac{x}{2}$ when $x \approx 0$, we have

$$\begin{aligned} D_\varepsilon(P_\delta, P_0) &= \varepsilon\frac{\varepsilon\delta^2}{1+2\varepsilon\delta^2} + \frac{1}{2}\varepsilon|\delta|(\log(1+\varepsilon|\delta|) - \log(1-\varepsilon|\delta|))\frac{1}{1+2\varepsilon\delta^2} = \Theta(\varepsilon^2\delta^2) \\ \|P_\delta - P_0\|_{\mathrm{TV}} &= \frac{1}{2}\left(\frac{\varepsilon\delta^2}{1+2\varepsilon\delta^2} + \varepsilon|\delta|\frac{1}{1+2\varepsilon\delta^2}\right) = \Theta(\varepsilon|\delta|) \\ d^2_{\mathrm{hel}}(P_\delta, P_0) &= \frac{1}{2}\left(\frac{((\sqrt{6}-1)^2 + (2-\sqrt{2})^2)\varepsilon\delta^2}{4(1+2\varepsilon\delta^2)} + \left((1-\sqrt{1-\varepsilon|\delta|})^2 + (\sqrt{1+\varepsilon|\delta|}-1)^2\right)\frac{1}{2+4\varepsilon\delta^2}\right) \\ &= \Theta(\varepsilon\delta^2). \end{aligned}$$

So then

$$\omega_\varepsilon\left(\frac{1}{n}, P_0; \mathcal{P}\right) = \Theta\left(\frac{1}{\varepsilon\sqrt{n}}\right)$$
$$\omega_{\mathrm{TV}}\left(\frac{1}{n\varepsilon}, P_0; \mathcal{P}\right) = \Theta\left(\frac{1}{\varepsilon^2 n}\right)$$
$$\omega_{d_{\mathrm{hel}}}\left(\frac{1}{\sqrt{n}}, P_0; \mathcal{P}\right) = \Theta\left(\frac{1}{\sqrt{n\varepsilon}}\right).$$

Now, consider a sequence of positive real numbers $\{\varepsilon_n\}$ such that $\varepsilon_n \to 0$ and $\varepsilon_n^2 n \to \infty$, we have

$$\lim_{n\to\infty} \frac{\omega_{\mathrm{TV}}\left(\frac{1}{n\varepsilon_n}, P_0; \mathcal{P}\right) \vee \omega_{d_{\mathrm{hel}}}\left(\frac{1}{\sqrt{n}}, P_0; \mathcal{P}\right)}{\omega_{\varepsilon_n}\left(\frac{1}{n}, P_0; \mathcal{P}\right)} = 0.$$

Appendix E. Missing proofs for Section 3

E.1. Proof of Proposition 9

First, it is clear from the definition that $\mathrm{dist}^{\mathsf{mon}}$ is $1/n$-sensitive. Thus, it remains to argue about the accuracy guarantees of it. First, note that for any two distributions $P, P' \in \mathcal{P}$

$$\left\|P - P'\right\|_{\mathrm{TV}} = \sup_{A\subset\mathbb{R}} P(A) - P'(A),$$

where $A \subset \mathbb{R}$ that maximizes $P(A) - P'(A)$ is $A = \{s \in \mathbb{R} : \frac{P(s)}{P'(s)} \geq 1\}$. Since $P(s)/P'(s)$ is monotone, this implies that either $A = [s_0, \infty)$ or $A = (-\infty, s_0]$ for some $s_0 \in \mathbb{R}$. Thus, we get that

$$\begin{aligned}
\left\|P - P'\right\|_{\mathrm{TV}} &= \max\left(\sup_{s\in\mathbb{R}} \Pr_{S\sim P}(S \leq s) - \Pr_{S\sim P'}(S \leq s), \sup_{s\in\mathbb{R}} \Pr_{S\sim P}(S \geq s) - \Pr_{S\sim P'}(S \geq s)\right), \\
&= \max\left(\sup_{s\in\mathbb{R}} \Pr_{S\sim P}(S \leq s) - \Pr_{S\sim P'}(S \leq s), \sup_{s\in\mathbb{R}} \Pr_{S\sim P'}(S \leq s) - \Pr_{S\sim P}(S \leq s)\right), \\
&= \sup_{s\in\mathbb{R}} |\Pr_{S\sim P}(S \leq s) - \Pr_{S\sim P'}(S \leq s)|.
\end{aligned}$$

Now, let $\rho_s = \frac{1}{n}\sum_{i=1}^n 1\{S_i \leq s\} - \Pr_{S\sim P}(S \leq s)$ and $\rho = \sup_{s\in\mathbb{R}} |\rho_s|$, and note that

$$\begin{aligned}
\mathrm{dist}^{\mathsf{mon}}(t; \mathcal{S}) &= \inf_{Q\in\mathcal{P}:\theta(Q)=t} \sup_{s\in\mathbb{R}} \left|\frac{1}{n}\sum_{i=1}^n 1\{S_i \leq s\} - \Pr_{S\sim Q}(S \leq s)\right| \\
&= \inf_{Q\in\mathcal{P}:\theta(Q)=t} \sup_{s\in\mathbb{R}} \left|\rho_s + \Pr_{S\sim P}(S \leq s) - \Pr_{S\sim Q}(S \leq s)\right| \\
&\leq \rho + \inf_{Q\in\mathcal{P}:\theta(Q)=t} \sup_{s\in\mathbb{R}} \left|\Pr_{S\sim P}(S \leq s) - \Pr_{S\sim Q}(S \leq s)\right| \\
&= \rho + \inf_{Q\in\mathcal{P}:\theta(Q)=t} \|Q - P\|_{\mathrm{TV}}
\end{aligned}$$

Similarly, we can show that $\mathrm{dist}^{\mathsf{mon}}(t; \mathcal{S}) \geq \inf_{Q \in \mathcal{P}: \theta(Q) = t} \|Q - P\|_{\mathrm{TV}} - \rho$, hence

$$\left| \mathrm{dist}^{\mathsf{mon}}(t; \mathcal{S}) - \inf_{Q \in \mathcal{P}: \theta(Q) = t} \|Q - P\|_{\mathrm{TV}} \right| \leq \rho.$$

Finally, the DKW inequality (Lemma 20) implies that $\rho \leq \sqrt{\log(2/\beta)/n}$ with probability $1 - \beta$. The claim now follows from Theorem 7 where we can set $\Delta = \sqrt{\log(2/\beta)/n}$.

E.2. Proof of Proposition 11

The claim will follow by showing that for $\mathcal{S} \stackrel{\mathrm{iid}}{\sim} P$, with probability $1 - \beta$ for all $A \in \mathcal{A}$ we have

$$\left| \frac{1}{n} \sum_{i=1}^{n} 1\{S_i \in A\} - P(A) \right| \leq \sqrt{\frac{4 \log(2|\mathcal{P}|/\beta)}{n}}. \tag{17}$$

Indeed, let $P_0 \in \mathcal{P}$ be such that $\|P - P_0\|_{\mathrm{TV}} \leq \eta$. We have that

$$\begin{aligned}
\mathrm{dist}^{\mathsf{gen}}(t; \mathcal{S}) &= \inf_{Q \in \mathcal{P}: \theta(Q) = t} \sup_{A \in \mathcal{A}(\mathcal{P})} \left| \frac{1}{n} \sum_{i=1}^{n} 1\{S_i \in A\} - Q(A) \right| \\
&\leq \sup_{A \in \mathcal{A}(\mathcal{P})} \left| \frac{1}{n} \sum_{i=1}^{n} 1\{S_i \in A\} - P(A) \right| + \inf_{Q \in \mathcal{P}: \theta(Q) = t} \sup_{A \in \mathcal{A}(\mathcal{P})} |P(A) - Q(A)| \\
&\leq \sqrt{\frac{4 \log(2|\mathcal{P}|/\beta)}{n}} + \|P - P_0\|_{\mathrm{TV}} + \inf_{Q \in \mathcal{P}: \theta(Q) = t} \sup_{A \in \mathcal{A}(\mathcal{P})} |P_0(A) - Q(A)| \\
&\leq \sqrt{\frac{4 \log(2|\mathcal{P}|/\beta)}{n}} + \eta + \inf_{Q \in \mathcal{P}: \theta(Q) = t} \|P_0 - Q\|_{\mathrm{TV}} \\
&\leq \sqrt{\frac{4 \log(2|\mathcal{P}|/\beta)}{n}} + 2\eta + \inf_{Q \in \mathcal{P}: \theta(Q) = t} \|P_0 - Q\|_{\mathrm{TV}}.
\end{aligned}$$

Similarly, we can show that $\mathrm{dist}^{\mathsf{gen}}(t; \mathcal{S}) \geq \inf_{Q \in \mathcal{P}: \theta(Q) = t} \|Q - P\|_{\mathrm{TV}} - \sqrt{\frac{4 \log(2|\mathcal{P}|/\beta)}{n}} - 2\eta$.

It remains to prove the concentration (17). Note that for $\mathcal{S} \stackrel{\mathrm{iid}}{\sim} P^n$ and any given $A \in \mathcal{A}(\mathcal{P})$, the variable $1\{S_i \in A\}$ is $1/4$-sub-gaussian, hence Hoeffding inequality (Duchi, 2018, Corollary 4.1.10) implies that

$$\mathrm{Pr}\left(\left| \frac{1}{n} \sum_{i=1}^{n} 1\{S_i \in A\} - P(A) \right| > \sqrt{\frac{4 \log(2|\mathcal{P}|/\beta)}{n}} \right) \leq \beta / |\mathcal{P}|^2.$$

Applying union bound over all $A \in \mathcal{A}(\mathcal{P})$ proves our desired concentration (17).

Appendix F. Proofs for Section 4

Lemma 40 *For any distributions P, P', Q and $\varepsilon > 0$ such that $\|P - P'\|_{\mathrm{TV}} \leq \eta$,*

$$\begin{aligned}
|\mathbb{E}_{X \sim P'}[\psi_\varepsilon(X; P, Q)] - D_\varepsilon(P, Q)| &\leq 2\varepsilon\eta \\
\mathrm{Var}_{X \sim P'}(\psi_\varepsilon(X; P, Q)) &\leq 4e^{\varepsilon} D_\varepsilon(P, Q) + 8\varepsilon^2 \eta.
\end{aligned}$$

Proof By definition of ψ_ε, it follows that $\mathbb{E}_{X\sim P}[\psi_\varepsilon(X;P,Q)] = D_\varepsilon(P,Q)$. The first claim then follows from Lemma 22 because $|\psi_\varepsilon(\cdot;P,Q)| \le \varepsilon$.

For the variance,

$$\begin{aligned}\mathrm{Var}_{X\sim P'}(\psi_\varepsilon(X;P,Q)) &= 4\mathrm{Var}_{X\sim P'}(\tilde{\psi}_\varepsilon(X;P,Q)) \\ &\le 4\mathbb{E}_{X\sim P'}[\tilde{\psi}_\varepsilon(X;P,Q)^2] \\ &\le 4\mathbb{E}_{X\sim P}[\tilde{\psi}_\varepsilon(X;P,Q)^2] + 8\varepsilon^2\eta,\end{aligned}$$

where the last inequality again follows from Lemma 22. Then,

$$\begin{aligned}\mathbb{E}_{X\sim P}[\tilde{\psi}_\varepsilon(X;P,Q)^2] &= \int P(x)\left(\left[\log\frac{P(x)}{Q(x)}\right]_{-\varepsilon}^{\varepsilon}\right)^2 dx. \\ &= \int P(x)\left(\left[\log\frac{Q(x)}{P(x)}\right]_{-\varepsilon}^{\varepsilon}\right)^2 dx. \\ &\le e^\varepsilon \int P(x)\left|\frac{Q(x)}{P(x)} - 1\right|\left|\left[\log\frac{Q(x)}{P(x)}\right]_{-\varepsilon}^{\varepsilon}\right| dx \\ &= e^\varepsilon \int |Q(x)-P(x)|\left|\left[\log\frac{Q(x)}{P(x)}\right]_{-\varepsilon}^{\varepsilon}\right| dx \\ &= e^\varepsilon \int (Q(x)-P(x))\left[\log\frac{Q(x)}{P(x)}\right]_{-\varepsilon}^{\varepsilon} dx \\ &= e^\varepsilon D_\varepsilon(Q,P),\end{aligned}$$

where the inequality follows by observing that $|\,[\log t]_{-\varepsilon}^{\varepsilon}\,| \le e^\varepsilon|t-1|$. The claim then follows because $D_\varepsilon(Q,P) = D_\varepsilon(P,Q)$ from Lemma 24. ■

Lemma 41 (Concentration of pairwise scores) *Suppose $S_i \overset{\text{iid}}{\sim} P_0 \in \mathcal{P}$ and $\varepsilon \le 1$. If $P, Q \in \mathcal{P}$ and $t \in \mathbb{R}$ are such that $\|P - P_0\|_{\mathrm{TV}} \le \eta$ and $t \ge \max\{4\varepsilon\eta, D_\varepsilon(P,Q)/2\}$, then*

$$\mathbb{P}\left(|\psi_\varepsilon(\mathcal{S};P,Q) - D_\varepsilon(P,Q)| > t\right) < 2\exp(-nt/100).$$

Proof Let $\mu = \mathbb{E}_{P_0}[\psi_\varepsilon(X;Q,P)]$ and $\sigma^2 = \mathrm{Var}_{P_0}(\psi_\varepsilon(X;Q,P))$. By Lemma 40, $|\mu - D_\varepsilon(P,Q)| \le 2\varepsilon\eta$ and $\sigma^2 \le 4e^\varepsilon D_\varepsilon(P,Q) + 8\varepsilon^2\eta$. Using Bernstein's inequality (Lemma 21),

$$\begin{aligned}\mathbb{P}\left(|\psi_\varepsilon(\mathcal{S};P,Q) - D_\varepsilon(P,Q)| > t\right) &\le \mathbb{P}\left(|\psi_\varepsilon(\mathcal{S};P,Q) - \mu| > t - 2\varepsilon\eta\right) \\ &= \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n \psi_\varepsilon(S_i;Q,P) - \mu\right| > t - 2\varepsilon\eta\right) \\ &\le 2\exp\left(-\frac{\frac{1}{2}n(t-2\varepsilon\eta)_+^2}{\sigma^2 + \frac{1}{3}\varepsilon(t-2\varepsilon\eta)_+}\right).\end{aligned}$$

Now using $t \ge \max\{4\varepsilon\eta, D_\varepsilon(P,Q)/2\}$, $\sigma^2 \le 4e^\varepsilon D_\varepsilon(P,Q) + 4\varepsilon^2\eta \le (8e^\varepsilon + 2\varepsilon)t$ and $\frac{1}{2}t \le (t-2\varepsilon\eta)_+ \le t$, which gives us that

$$\mathbb{P}\left(|\psi_\varepsilon(\mathcal{S};P,Q) - D_\varepsilon(P,Q)| > t\right) \le 2\exp\left(-\frac{\frac{1}{4}nt^2}{(8e^\varepsilon + 2\varepsilon + \frac{1}{3}\varepsilon)t}\right).$$

Because $8e^{\varepsilon} + 2\varepsilon + \frac{1}{3}\varepsilon \le 25$ for $\varepsilon \le 1$, the claim follows. ■

Proof [Proof of Lemma 14] Notice by definition that $\mathsf{dist}_{\varepsilon}^{\mathcal{P}}(Q; \mathcal{S}) \ge -\psi_{\varepsilon}(\mathcal{S}; Q, P_0)$. We apply Lemma 41 with $P \leftarrow Q$, $Q \leftarrow P_0$ and $t = D_{\varepsilon}(Q, P_0)/2$ to get that $|\psi_{\varepsilon}(\mathcal{S}; Q, P_0) - D_{\varepsilon}(Q, P_0)| \le D_{\varepsilon}(Q, P_0)/2$ with probability at least $1 - \exp(-cnD_{\varepsilon}(Q, P_0)/2)$.

Under this event, $\mathsf{dist}_{\varepsilon}^{\mathcal{P}}(Q; \mathcal{S}) \ge -\psi_{\varepsilon}(\mathcal{S}; Q, P_0) \ge D_{\varepsilon}(Q, P_0)/2$. Now, noting that Lemma 26 implies $|D_{\varepsilon}(Q, P_0) - D_{\varepsilon}(Q, P)| \le 4(e^{\varepsilon} - 1)\eta \le 8\varepsilon\eta$, the assumption $D_{\varepsilon}(Q, P) \ge 10\varepsilon\eta$ implies that $\frac{2}{10} \le \frac{D_{\varepsilon}(Q,P_0)}{D_{\varepsilon}(Q,P)} \le \frac{18}{10}$.

Therefore, by changing constants as necessary, we have with probability at least $1 - \exp(-cnD_{\varepsilon}(Q, P))$ that $\mathsf{dist}_{\varepsilon}^{\mathcal{P}}(Q; \mathcal{S}) \ge D_{\varepsilon}(Q, P_0)/2 \ge D_{\varepsilon}(Q, P)/10$. ■

Proof [Proof of Lemma 15] Consider $Q \in \mathcal{P}$ such that $D_{\varepsilon}(P_0, Q) \ge 2B$. Applying Lemma 41 with $t = D_{\varepsilon}(P_0, Q)/2$, we see that $|\psi_{\varepsilon}(\mathcal{S}; P_0, Q) - D_{\varepsilon}(P_0, Q)| \le D_{\varepsilon}(P_0, Q)/2$ with probability at least $1 - 4\exp(-cnD_{\varepsilon}(P_0, Q)) \ge 1 - 4\exp(-cnB)$. By assumption on $D_{\varepsilon}(P_0, Q)$, $|\psi_{\varepsilon}(\mathcal{S}; P_0, Q) - D_{\varepsilon}(P_0, Q)| \le D_{\varepsilon}(P_0)/2$ implies that $\psi_{\varepsilon}(\mathcal{S}; P_0, Q) \ge B \ge 0$.

Now consider $Q \in \mathcal{P}$ such that $D_{\varepsilon}(P_0, Q) < 2B$. By applying Lemma 41 with $t = B$, we see that $|\psi_{\varepsilon}(\mathcal{S}; P_0, Q) - D_{\varepsilon}(P_0, Q)| \le B$ with probability at least $1 - 4\exp(-cnB)$. Moreover, $|\psi_{\varepsilon}(\mathcal{S}; P_0, Q) - D_{\varepsilon}(P_0, Q)| \le B$ implies that $\psi_{\varepsilon}(\mathcal{S}; P_0, Q) > -B$.

By a union bound over all $Q \in \mathcal{P}$, we have with probability at least $1 - 4|\mathcal{P}|\exp(-cnB)$ that

$$\mathsf{dist}_{\varepsilon}^{\mathcal{P}}(P_0; \mathcal{S}) = -\inf_{Q \in \mathcal{P}} \psi_{\varepsilon}(\mathcal{S}; P_0, Q) \le B.$$

■

Appendix G. Missing results for Section 5

Lemma 42 *(Private to robust) Let $\varepsilon \le 1$. Assume $\mathcal{A}_p$ is a ε-DP algorithm such for any $P \in \mathcal{P}$, given $\mathcal{S} \overset{\text{iid}}{\sim} P^n$*

$$\mathbb{E}_{\mathcal{S} \overset{\text{iid}}{\sim} P^n, \mathcal{A}_p}[|\mathcal{A}_p(\mathcal{S}) - \theta(P)|] \le \alpha.$$

Then $\mathcal{A}_p$ is τ-robust with $\tau = 1/n\varepsilon$ and error

$$\mathbb{E}_{\mathcal{S} \sim P^n}\left[\max_{\mathcal{S}' : d_{\text{ham}}(\mathcal{S}, \mathcal{S}') \le n\tau} \mathbb{E}_{\mathcal{A}_p}[|\mathcal{A}_p(\mathcal{S}') - \theta(P)|]\right] \le e\alpha.$$

Proof Note that for $\mathcal{S}$ and $\mathcal{S}'$ such that $d_{\text{ham}}(\mathcal{S}, \mathcal{S}') \le n\tau = 1/\varepsilon$, we have that

$$\begin{aligned}
\mathbb{E}_{\mathcal{A}_p}[|\mathcal{A}_p(\mathcal{S}') - \theta(P)|] &= \int_{t \in \Theta} \Pr(\mathcal{A}_p(\mathcal{S}') = t)|t - \theta(P)| \\
&\le \int_{t \in \Theta} e \Pr(\mathcal{A}_p(\mathcal{S}) = t)|t - \theta(P)| \\
&\le e\mathbb{E}_{\mathcal{A}_p}[|\mathcal{A}_p(\mathcal{S}) - \theta(P)|].
\end{aligned}$$

Thus we have that

$$\mathbb{E}_{\mathcal{S} \sim P^n}\left[\max_{\mathcal{S}' : d_{\text{ham}}(\mathcal{S}, \mathcal{S}') \le n\tau} \mathbb{E}_{\mathcal{A}_p}[|\mathcal{A}_p(\mathcal{S}') - \theta(P)|]\right] \le e\mathbb{E}_{\mathcal{S} \sim P^n}\left[\mathbb{E}_{\mathcal{A}_p}[|\mathcal{A}_p(\mathcal{S}) - \theta(P)|]\right] \le e\alpha.$$

■

Lemma 43 *(Robust to private) Assume $\mathcal{A}_r$ is a τ-robust algorithm such for any P,*

$$\mathbb{E}_{\mathcal{S}\sim P^n}\left[\max_{\mathcal{S}':d_{\mathrm{ham}}(\mathcal{S},\mathcal{S}')\le n\tau}\mathbb{E}_{\mathcal{A}_r}[|\mathcal{A}_r(\mathcal{S}')-\theta(P)|]\right]\le\alpha.$$

Then there is an ε-DP algorithm $\mathcal{A}_p$ where $\varepsilon=\frac{C\cdot\log\left(\frac{R}{\alpha}\right)}{n\tau}$ for a constant $1\le C<\infty$ such that for $\mathcal{S}\overset{\text{iid}}{\sim}P^n$ we have

$$\mathbb{E}_{\mathcal{S}\overset{\text{iid}}{\sim}P^n,\mathcal{A}_p}[|\mathcal{A}_p(\mathcal{S})-\theta(P)|]\le 5\alpha.$$

Proof The proof follows the same arguments as the proof of Theorem 3.1 in Asi et al. (2023), except that we use expectations instead of high probability bounds. ■

G.1. Proof of Theorem 18

The proof follows directly from the guarantees of the T-mechanism (Theorem 13) and the fact that any ε-DP mechanism is also τ-robust with robust error $e\alpha$ for $\tau=1/n\varepsilon$ (Lemma 42).

G.2. Proof of Theorem 19

Assume $\mathfrak{M}_n(\mathcal{P},\mathcal{A}_\tau)=\alpha/2$. Otherwise assume $\alpha\le 1$. Then there is a τ-robust algorithm $\mathcal{A}_r$ with error α, that is, we have that for any $P\in\mathcal{P}$

$$\mathbb{E}_{\mathcal{S}\sim P^n}\left[\max_{\mathcal{S}':d_{\mathrm{ham}}(\mathcal{S},\mathcal{S}')\le n\tau}\mathbb{E}_{\mathcal{A}_r}[|\mathcal{A}_r(\mathcal{S}')-\theta(P)|]\right]\le\alpha.$$

Then, Lemma 43 implies that there is an ε-DP algorithm $\mathcal{A}_p$ where $\varepsilon=\frac{C\cdot\log\left(\frac{1}{\alpha}\right)}{n\tau}$ for $1\le C\le\infty$ such that for any $P\in\mathcal{P}$, given $\mathcal{S}\overset{\text{iid}}{\sim}P^n$,

$$\mathbb{E}_{\mathcal{S}\overset{\text{iid}}{\sim}P^n,\mathcal{A}_p}[|\mathcal{A}_p(\mathcal{S})-\theta(P)|]\le 5\alpha.$$

Lower bounds for ε-DP estimation now imply that $\alpha\ge\Omega\left(\sup_{P\in\mathcal{P}}\omega_{C\cdot\log\left(\frac{1}{\alpha}\right)/n\tau}\left(\frac{1}{n};P\right)\right)$. Finally, the claim follows as we know that $\omega_{\varepsilon_1}(1/n;P)\ge\omega_{\varepsilon_2}(1/n;P)$ whenever $\varepsilon_1\le\varepsilon_2$ and the assumption that $\alpha\ge 1/n^{C_1}$.