The Annals of Statistics
2020, Vol. 48, No. 5, 2646–2670
https://doi.org/10.1214/19-AOS1901

GEOMETRIZING RATES OF CONVERGENCE UNDER LOCAL DIFFERENTIAL PRIVACY CONSTRAINTS

BY ANGELIKA ROHDE[1] AND LUKAS STEINBERGER[2]

[1]*University of Freiburg, angelika.rohde@stochastik.uni-freiburg.de*
[2]*Department of Statistics and OR, University of Vienna, lukas.steinberger@univie.ac.at*

We study the problem of estimating a functional $\theta(\mathbb{P})$ of an unknown probability distribution $\mathbb{P} \in \mathcal{P}$ in which the original iid sample $X_1, \dots, X_n$ is kept private even from the statistician via an α-local differential privacy constraint. Let ω_{TV} denote the modulus of continuity of the functional θ over $\mathcal{P}$ with respect to total variation distance. For a large class of loss functions l and a fixed privacy level α, we prove that the privatized minimax risk is equivalent to $l(\omega_{\mathrm{TV}}(n^{-1/2}))$ to within constants, under regularity conditions that are satisfied, in particular, if θ is linear and $\mathcal{P}$ is convex. Our results complement the theory developed by Donoho and Liu (1991) with the nowadays highly relevant case of privatized data. Somewhat surprisingly, the difficulty of the estimation problem in the private case is characterized by ω_{TV}, whereas, it is characterized by the Hellinger modulus of continuity if the original data $X_1, \dots, X_n$ are available. We also find that for locally private estimation of linear functionals over a convex model a simple sample mean estimator, based on independently and binary privatized observations, always achieves the minimax rate. We further provide a general recipe for choosing the functional parameter in the optimal binary privatization mechanisms and illustrate the general theory in numerous examples. Our theory allows us to quantify the price to be paid for local differential privacy in a large class of estimation problems. This price appears to be highly problem specific.

1. Introduction. One of the many new challenges for statistical inference in the information age is the increasing concern of data privacy protection. Nowadays, massive amounts of data, such as medical records, smart phone user behavior or social media activity, are routinely being collected and stored. On the other side of this trend is an increasing reluctance and discomfort of individuals to share this sometimes sensitive information with companies or state officials. Over the last few decades, the problem of constructing privacy preserving data release mechanisms has produced a vast literature, predominantly in computer science. One particularly fruitful approach to data protection that is insusceptible to privacy breaches is the concept of differential privacy (see Dinur and Nissim (2003), Dwork (2008), Dwork and Nissim (2004), Dwork et al. (2006), Evfimievski, Gehrke and Srikant (2003)). In a nutshell, differential privacy is a form of randomization, where, instead of the original data, a perturbed version of the data is released, offering plausible deniability to the data providers who can always argue that their true answer was different from the one that was actually provided. Aside from the academic discussion, (local) differential privacy has also found its way into real world applications. For instance, the Apple Inc. privacy statement explains the notion quite succinctly as follows:

> It is a technique that enables Apple to learn about the user community without learning about individuals in the community. Differential privacy transforms the information shared with Apple before it ever leaves the user's device such that Apple can never reproduce the true data.[1]

Received May 2019; revised August 2019.
MSC2020 subject classifications. Primary 62G05; secondary 62C20.
Key words and phrases. Local differential privacy, minimax estimation, rate of convergence, moduli of continuity, nonparametric estimation.

[1]https://images.apple.com/privacy/docs/Differential_Privacy_Overview.pdf

The qualification of "local" differential privacy refers to a procedure which randomizes the original data already on the user's "local" machine and the original data is never released, whereas (central) differential privacy may also be employed to privatize and release an entire database that was previously compiled by a trusted curator. Here, we focus only on the local version of differential privacy.

More recently, differential privacy has also received some attention from a statistical inference perspective (see, e.g., Awan and Slavković (2018), Duchi, Jordan and Wainwright (2013a, 2013b, 2014, 2018), Dwork and Smith (2010), Smith (2008, 2011), Wasserman and Zhou (2010), Ye and Barg (2019)). In this line of research, the focus is more on the inherent trade-off between privacy protection and efficient statistical inference and on the question what optimal privacy preserving mechanisms may look like. Duchi, Jordan and Wainwright (2013a, 2013b, 2014, 2018) introduced new variants of the Le Cam, Fano and Assouad techniques to derive lower bounds on the privatized minimax risk. In this way, they were the first to provide minimax rates of convergence for specific estimation problems under privacy constraints in a very insightful case by case study. Here, we develop a general theory, in the spirit of Donoho and Liu (1991), to characterize the differentially private minimax rate of convergence. Characterizing the minimax rate of convergence under differential privacy, and comparing it to the minimax risk in the nonprivate case, is one way to quantify the price, in terms of statistical accuracy that has to be paid for privacy protection. The theory also allows us to develop (asymptotically) minimax optimal privatization schemes for a large class of estimation problems.

To be more precise, consider n individuals who possess data $X_1, \dots, X_n$, assumed to be i.i.d. from some probability distribution $\mathbb{P} \in \mathcal{P}$. However, the statistician does not get to see the original data $X_1, \dots, X_n$ but only a *privatized* version of observations Z. The conditional distribution of Z given $X = (X_1, \dots, X_n)$ is denoted by Q and referred to as a channel distribution or a privatization scheme, that is, $\Pr(Z \in A | X = x) = Q(A|x)$. For $\alpha \in (0, \infty)$, the channel Q is said to provide α-differential privacy if

$$\sup_A \sup_{x,x' : d_0(x,x')=1} \frac{\Pr(Z \in A | X = x)}{\Pr(Z \in A | X = x')} \le e^\alpha, \tag{1.1}$$

where the first supremum runs over all measurable sets and $d_0(x, x') := |\{i : x_i \neq x'_i\}|$ denotes the number of distinct entries of x and x'. This definition is due to Dwork et al. (2006) (see also Evfimievski, Gehrke and Srikant (2003)). It captures the idea that the distribution of the observation Z does not change too much if the data of any single individual in the database is changed, thereby protecting its privacy. The smaller $\alpha \in (0, \infty)$, the stronger is the privacy constraint (1.1). More formally, (if we consider the original data X as fixed) Wasserman and Zhou (2010), Theorem 2.4, show that under α-differential privacy, any level-γ test using Z to test $H_0 : X = x$ vs. $H_1 : X = x'$ has power bounded by γe^α. As mentioned above, in this paper we focus on a special case of differential privacy, namely, local differential privacy. Somewhat informally, a channel satisfying (1.1) is said to provide local differential privacy, if $Z = (Z_1, \dots, Z_n)$ is a random n-vector and if the ith individual can generate its privatized data Z_i using only its original data X_i and possibly other information, but without sharing X_i with anyone else. The point of this definition is that for such a protocol to be realized, we do not need a trusted third party to collect or process data. It is reminiscent of the idea of randomized response (Warner (1965)). We will see more concrete instances of such protocols below.

Suppose now that we want to estimate a real parameter $\theta(\mathbb{P})$, based on the privatized observation vector Z, whose unconditional distribution is equal to $Q\mathbb{P}^{\otimes n}(dz) := \int Q(dz|x)\, \mathbb{P}^{\otimes n}(dx)$, where $\mathbb{P}^{\otimes n}$ is the n-fold product measure of $\mathbb{P}$. The Q-privatized minimax risk of estimation under a loss function $l : \mathbb{R} \to \mathbb{R}$ is therefore given by

$$\mathcal{M}_n(Q, \mathcal{P}, \theta) := \inf_{\hat{\theta}_n} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{Q\mathbb{P}^{\otimes n}}\big[l\big(\big|\hat{\theta}_n - \theta(\mathbb{P})\big|\big)\big], \tag{1.2}$$

where the infimum runs over all estimators $\hat{\theta}_n$ taking Z as input data. Note that if the channel Q is given by $Q(A|x) = \Pr(Z \in A|X = x) = \mathbb{1}_A(x)$, then there is no privatization at all and the Q-privatized minimax risk reduces to the conventional minimax risk of estimating $\theta(\mathbb{P})$. If we want to guarantee α-differential privacy, then we may choose any channel Q that satisfies (1.1) and we will try to make (1.2) as small as possible. This leads us to the α-private minimax risk

$$\mathcal{M}_{n,\alpha}(\mathcal{P},\theta) := \inf_{Q\in\mathcal{Q}_\alpha} \mathcal{M}_n(Q,\mathcal{P},\theta),$$

where $\mathcal{Q}_\alpha$ is some set of α-differentially private channels. It is this additional infimum over $\mathcal{Q}_\alpha$ that makes the theory of private minimax estimation deviate fundamentally from the conventional minimax estimation approach. In particular, this situation is different from the statistical inverse problem setting, because the Markov kernel Q can be chosen in an optimal way and is not given a priori. A sequence of channels $Q^{(n)} \in \mathcal{Q}_\alpha$, for which $\mathcal{M}_n(Q^{(n)},\mathcal{P},\theta)$ is of the order of $\mathcal{M}_{n,\alpha}(\mathcal{P},\theta)$, is referred to as a minimax rate optimal channel and may depend on the specific estimation problem under consideration, that is, on θ and $\mathcal{P}$. We write $\mathcal{M}_{n,\infty}(\mathcal{P},\theta)$ for the classical (nonprivate) minimax risk.

The novel contribution of this article is to characterize the rate at which $\mathcal{M}_{n,\alpha}(\mathcal{P},\theta)$ converges to zero as $n\to\infty$, in high generality, and to provide concrete minimax rate optimal α-locally differentially private estimation procedures. To this end, we utilize the modulus of continuity of the functional $\theta:\mathcal{P}\to\mathbb{R}$ with respect to the total variation distance $d_{\mathrm{TV}}(\mathbb{P}_0,\mathbb{P}_1)$, that is,

$$\omega_{\mathrm{TV}}(\varepsilon) := \sup\bigl\{\bigl|\theta(\mathbb{P}_0)-\theta(\mathbb{P}_1)\bigr| : d_{\mathrm{TV}}(\mathbb{P}_0,\mathbb{P}_1)\le\varepsilon, \mathbb{P}_0,\mathbb{P}_1\in\mathcal{P}\bigr\},$$

and we show that for any fixed $\alpha\in(0,\infty)$,

$$\mathcal{M}_{n,\alpha}(\mathcal{P},\theta) \asymp l\bigl(\omega_{\mathrm{TV}}\bigl(n^{-1/2}\bigr)\bigr). \tag{1.3}$$

Here, $a_n \asymp b_n$ means that there exist constants $0<c_0<c_1<\infty$ and $n_0\in\mathbb{N}$, not depending on n, so that $c_0 b_n \le a_n \le c_1 b_n$, for all $n\ge n_0$. The lower bound on $\mathcal{M}_{n,\alpha}(\mathcal{P},\theta)$ that we establish holds in full generality, whereas, in order to obtain a matching upper bound, it is necessary to impose some regularity conditions on $\mathcal{P}$ and θ. These will be satisfied, in particular, if $\mathcal{P}$ is convex and dominated and θ is linear and bounded but also hold in some cases of nonconvex and potentially nondominated $\mathcal{P}$. It is important to compare (1.3) to the analogous result for the nonprivate minimax risk. This was established in the seminal paper by Donoho and Liu (1991), who, under regularity conditions similar to those imposed here, showed that

$$\mathcal{M}_{n,\infty}(\mathcal{P},\theta) \asymp l\bigl(\omega_{\mathrm{H}}\bigl(n^{-1/2}\bigr)\bigr), \tag{1.4}$$

where $\omega_{\mathrm{H}}(\varepsilon) = \sup\{|\theta(\mathbb{P}_0)-\theta(\mathbb{P}_1)| : d_{\mathrm{H}}(\mathbb{P}_0,\mathbb{P}_1)\le\varepsilon, \mathbb{P}_0,\mathbb{P}_1\in\mathcal{P}\}$ and d_{H} is the Hellinger distance. Comparing (1.4) to (1.3), we notice that the Hellinger modulus ω_{H} of θ is replaced by the total variation modulus ω_{TV}. This may, and typically will, lead to different rates of convergence in private and nonprivate problems. Note that even in cases where we do or can not compute the moduli ω_{TV} and ω_{H} explicitly, we always have the a priori information that

$$\omega_{\mathrm{H}}(\varepsilon) \le \omega_{\mathrm{TV}}(\varepsilon) \le \omega_{\mathrm{H}}(\sqrt{2\varepsilon}),$$

because $d_{\mathrm{TV}} \le d_{\mathrm{H}} \le \sqrt{2 d_{\mathrm{TV}}}$ (see, for instance, Tsybakov (2009)). This means that the private rate of estimation is never faster than the nonprivate rate and is never slower than the square root of the nonprivate rate. We shall see that both extremal cases can occur (see Section 6). We shall also see that the fastest possible private rate of convergence over a convex model $\mathcal{P}$ is $l(n^{-1/2})$ (see Lemma H.2 in Section H of the Supplementary Material Rohde and Steinberger (2020)). Also note that in (1.3) we have suppressed constants that depend on α. Our

results reveal that if α is small, the effective sample size reduces from n to $n(e^\alpha - 1)^2$ when α-differential privacy is required. That differential privacy leads to slower minimax rates of convergence was already observed by Duchi, Jordan and Wainwright (2013a, 2013b, 2018), for specific estimation problems. Here, we develop a unifying general theory to quantify the privatized minimax rates of convergence in a large class of different estimation problems, including (even irregular) parametric and nonparametric cases. This is also the first step towards a fundamental theory of adaptive estimation under differential privacy that will be pursued elsewhere.

We also provide a general construction of α-locally differentially private estimation procedures that is minimax rate optimal if $\mathcal{P}$ is convex and dominated and θ is linear and bounded. The construction relies on a functional parameter $\ell \in L_\infty$. Each individual generates Z_i independently and binary distributed on $\{-z_0, z_0\}$, with

$$\Pr(Z_i = z_0 | X_i = x_i) = \frac{1}{2}\left(1 + \frac{\ell(x_i)}{z_0}\right)$$

and $z_0 = \|\ell\|_\infty \frac{e^\alpha+1}{e^\alpha-1}$. The final estimator is then simply given by the sample mean $\bar{Z}_n = \frac{1}{n}\sum_{i=1}^n Z_i$. For appropriate $\ell = \ell_n$, this yields an α-locally differentially private procedure that attains the minimax rate in (1.3). The choice of functional parameter ℓ is problem specific but can often be guided by considering optimality of the estimator $\mathbb{E}[\bar{Z}_n | X_1, \dots, X_n] = \frac{1}{n}\sum_{i=1}^n \ell(X_i)$ in the problem with direct observations. We exemplify this choice in many classical moment or density estimation problems (cf. Section 6). We point out, however, that there are cases where a certain ℓ leads to rate optimal locally private estimation even though the estimator $\frac{1}{n}\sum_{i=1}^n \ell(X_i)$ is not rate optimal in the direct problem.

The paper is organized as follows. In the next section (Section 2) we formally introduce the private estimation problem, several classes of locally private channel distributions and a few tools required for the analysis of the α-private minimax risk $\mathcal{M}_{n,\alpha}(\mathcal{P}, \theta)$. Section 3 presents a general lower bound on $\mathcal{M}_{n,\alpha}(\mathcal{P}, \theta)$. That this lower bound is attainable, in surprisingly high generality and by the simple linear estimation procedure described above, is then established in Section 4. The results of that section, however, do not offer an explicit construction of the functional parameter ℓ. In Section 5 we then provide some guidance on choosing ℓ, as well as a high-level condition for optimality of ℓ that we verify in all our examples. We illustrate the general theory by a number of concrete examples that are presented in Section 6. Most of the technical arguments are deferred to the Supplementary Material.

2. Preliminaries and notation. Let $\mathcal{P}$ be a set of probability measures on the measurable space $(\mathcal{X}, \mathcal{F})$. Let $\theta : \mathcal{P} \to \mathbb{R}$ be a functional of interest. In case $\mathcal{P}$ is convex, we say that the functional $\theta : \mathcal{P} \to \mathbb{R}$ is linear if for $\mathbb{P}_0, \mathbb{P}_1 \in \mathcal{P}$ and $\lambda \in [0, 1]$, we have $\theta(\lambda\mathbb{P}_0 + (1-\lambda)\mathbb{P}_1) = \lambda\theta(\mathbb{P}_0) + (1-\lambda)\theta(\mathbb{P}_1)$. We are given the privatized data $Z_1, \dots, Z_n$ on the measurable space $(\mathcal{Z}, \mathcal{B}(\mathcal{Z}))$, $\mathcal{Z} = \mathbb{R}^q$, where $\mathcal{B}(\mathcal{Z})$ denotes the Borel sets with respect to the usual topology. The conditional distribution of the observations $Z = (Z_1, \dots, Z_n)$ given the original sample $X = (X_1, \dots, X_n)$ is described by the *channel distribution* Q. That is, Q is a Markov probability kernel from $(\mathcal{X}^n, \mathcal{F}^{\otimes n})$ to $(\mathcal{Z}^n, \mathcal{B}(\mathcal{Z}^n))$. For ease of notation we suppress its dependence on n. Hence, if the X_i are distributed i.i.d., according to $\mathbb{P} \in \mathcal{P}$, and $\mathbb{P}^{\otimes n}$ denotes the corresponding product measure, then the joint distribution of the observation vector $Z = (Z_1, \dots, Z_n)$ on $\mathcal{Z}^n$ is given by $Q\mathbb{P}^{\otimes n}$, that is, the measure $A \mapsto \int_{\mathcal{X}^n} Q(A|x)\, d\mathbb{P}^{\otimes n}(x)$.

2.1. *Locally differentially private minimax risk.* Recall that for $\alpha \in (0, \infty)$, a channel distribution Q is called *α-differentially private*, if

$$\sup_{A \in \mathcal{B}(\mathcal{Z}^n)} \sup_{\substack{x, x' \in \mathcal{X}^n \\ d_0(x,x')=1}} \frac{Q(A|x)}{Q(A|x')} \le e^\alpha, \tag{2.1}$$

where $d_0(x,x') := |\{i : x_i \neq x_i'\}|$ is the number of distinct components of x and x'. Note that for this definition to make sense, the probability measures $Q(\cdot|x)$, for different $x \in \mathcal{X}^n$, have to be equivalent, and we interpret $\frac{0}{0}$ as equal to 1.

Next, we introduce two specific classes of locally differentially private channels. A channel distribution $Q : \mathcal{B}(\mathcal{Z}^n) \times \mathcal{X}^n \to [0,1]$ is said to be *α-sequentially interactive* (or provides α-sequentially interactive differential privacy) if the following two conditions are satisfied. First, we have for all $A \in \mathcal{B}(\mathcal{Z}^n)$ and $x_1, \dots, x_n \in \mathcal{X}$,

$$\begin{aligned} &Q(A|x_1,\dots,x_n) \\ &\quad = \int_{\mathcal{Z}} \cdots \int_{\mathcal{Z}} Q_n(A_{z_{1:n-1}}|x_n, z_{1:n-1}) \\ &\qquad \times Q_{n-1}(dz_{n-1}|x_{n-1}, z_{1:n-2}) \cdots Q_1(dz_1|x_1), \end{aligned} \tag{2.2}$$

where, for each $i = 1, \dots, n$, Q_i is a channel from $\mathcal{X} \times \mathcal{Z}^{i-1}$ to $\mathcal{Z}$. Here, $z_{1:n} = (z_1, \dots, z_n)^T$ and $A_{z_{1:n-1}} = \{z \in \mathcal{Z} : (z_1, \dots, z_{n-1}, z)^T \in A\}$ is the $z_{1:n-1}$-section of A. Second, we require that the conditional distributions Q_i satisfy

$$\sup_{A \in \mathcal{B}(\mathcal{Z})} \sup_{x_i, x_i', z_1, \dots, z_{i-1}} \frac{Q_i(A|x_i, z_1, \dots, z_{i-1})}{Q_i(A|x_i', z_1, \dots, z_{i-1})} \leq e^{\alpha} \quad \forall i = 1, \dots, n. \tag{2.3}$$

By the usual approximation of integrands by simple functions, it is easy to see that (2.2) and (2.3) imply (2.1). This notion coincides with the definition of sequentially interactive channels in Duchi, Jordan and Wainwright (2018), Definition 1. We note that (2.3) only makes sense if for all $x_i, x_i', z_1, \dots, z_{i-1}$, the probability measure $Q_i(\cdot|x_i, z_{1:i-1})$ is absolutely continuous with respect to $Q_i(\cdot|x_i', z_{1:i-1})$. Here, the idea is that individual i can only use X_i and previous Z_j, $j < i$ in its local privacy mechanism, thus leading to the sequential structure in the above definition. In the rest of the paper, we only consider α-sequentially interactive channels, to which we also refer simply as α-private channels.

An important subclass of sequentially interactive channels are the so-called *noninteractive* channels Q that are of product form

$$Q(A_1 \times \cdots \times A_n | x_1, \dots, x_n) = \prod_{i=1}^{n} Q_i(A_i|x_i) \quad \forall A_i \in \mathcal{B}(\mathcal{Z}), x_i \in \mathcal{X}. \tag{2.4}$$

Clearly, a noninteractive channel Q satisfies (2.1) if and only if,

$$\sup_{A \in \mathcal{B}(\mathcal{Z})} \sup_{x, x' \in \mathcal{X}} \frac{Q_i(A|x)}{Q_i(A|x')} \leq e^{\alpha} \quad \forall i = 1, \dots, n.$$

In that case it is also called α-noninteractive. Both, α-noninteractive and α-sequentially interactive channels satisfy the α-local differential privacy constraint, as defined in the Introduction. Of course, every α-noninteractive channel is also α-sequentially interactive.

If we measure the error of estimation by the measurable loss function $l : \mathbb{R}_+ \to \mathbb{R}_+$, where $\mathbb{R}_+ := [0, \infty)$, the minimax risk of the above estimation problem is given by

$$\mathcal{M}_n(Q, \mathcal{P}, \theta) = \inf_{\hat{\theta}_n} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{Q\mathbb{P}^{\otimes n}}\big[l\big(|\hat{\theta}_n - \theta(\mathbb{P})|\big)\big], \tag{2.5}$$

where the infimum runs over all estimators $\hat{\theta}_n : \mathcal{Z}^n \to \mathbb{R}$. Finally, define the set of α-private channels:

$$\mathcal{Q}_{\alpha} := \bigcup_{q \in \mathbb{N}} \{Q : Q \text{ is } \alpha\text{-sequentially interactive from } \mathcal{X}^n \text{ to } \mathcal{Z}^n = \mathbb{R}^{n \times q}\}. \tag{2.6}$$

Therefore, the α-private minimax risk is given by

$$\mathcal{M}_{n,\alpha}(\mathcal{P}, \theta) = \inf_{Q \in \mathcal{Q}_{\alpha}} \mathcal{M}_n(Q, \mathcal{P}, \theta). \tag{2.7}$$

Note that the above infimum includes all possible dimensions q of $\mathcal{Z} = \mathbb{R}^q$.

2.2. *Testing affinities and minimax identities.* Let $\mathcal{P}$, $\mathcal{P}_0$ and $\mathcal{P}_1$ be sets of probability measures on a measurable space $(\Omega, \mathcal{A})$ and for $\mathbb{P}_0 \in \mathcal{P}_0$, $\mathbb{P}_1 \in \mathcal{P}_1$, define the testing affinity

$$\pi(\mathbb{P}_0, \mathbb{P}_1) = \inf_{\text{tests } \phi} \mathbb{E}_{\mathbb{P}_0}[\phi] + \mathbb{E}_{\mathbb{P}_1}[1 - \phi], \tag{2.8}$$

where the infimum runs over all (randomized) tests $\phi : \Omega \to [0, 1]$. Moreover, we write

$$\pi(\mathcal{P}_0, \mathcal{P}_1) = \sup_{\mathbb{P}_j \in \mathcal{P}_j, j=0,1} \pi(\mathbb{P}_0, \mathbb{P}_1). \tag{2.9}$$

Throughout, we follow the usual conventions that $\sup \varnothing = -\infty$ and $\inf \varnothing = +\infty$. If $\theta : \mathcal{P} \to \mathbb{R}$ is a functional of interest, then for $t \in \mathbb{R}$ and $\Delta > 0$, denote $\mathcal{P}_{\le t} := \{\mathbb{P} \in \mathcal{P} : \theta(\mathbb{P}) \le t\}$ and $\mathcal{P}_{\ge t+\Delta} := \{\mathbb{P} \in \mathcal{P} : \theta(\mathbb{P}) \ge t + \Delta\}$ and let $\mathcal{P}_{\le t}^{(n)}$ and $\mathcal{P}_{\ge t+\Delta}^{(n)}$ be the sets of n-fold product measures with identical marginals from $\mathcal{P}_{\le t}$ and $\mathcal{P}_{\ge t+\Delta}$, respectively. If Q is a Markov probability kernel, then we write $Q\mathcal{P}^{(n)}$ for the set of all probability measures of the form $Q\mathbb{P}^{\otimes n}$, where $\mathbb{P} \in \mathcal{P}$. Recall that a family of measures on a common probability space is dominated if there exists a σ-finite measure μ such that every element of that family is absolutely continuous with respect to μ. We define the convex hull $\operatorname{conv}(\mathcal{P})$ in the usual way to be the set of all finite convex combinations $\sum_{i=1}^m \lambda_i \mathbb{P}_i$, for $m \in \mathbb{N}$, $\lambda_i \ge 0$, $\sum_{i=1}^m \lambda_i = 1$ and $\mathbb{P}_i \in \mathcal{P}$. For $\mathbb{P}_0, \mathbb{P}_1 \in \mathcal{P}$, we consider the Hellinger distance

$$d_{\mathrm{H}}(\mathbb{P}_0, \mathbb{P}_1) := \sqrt{\int_\Omega \bigl(\sqrt{p_0(x)} - \sqrt{p_1(x)}\bigr)^2 \, d\mu(x)},$$

where p_0 and p_1 are densities of $\mathbb{P}_0$ and $\mathbb{P}_1$ with respect to some dominating measure μ (e.g., $\mu = \mathbb{P}_0 + \mathbb{P}_1$) and the total variation distance is defined as $d_{\mathrm{TV}}(\mathbb{P}_0, \mathbb{P}_1) := \sup_{A \in \mathcal{A}} |\mathbb{P}_0(A) - \mathbb{P}_1(A)|$. Furthermore, for a monotone function $g : \mathbb{R} \to \mathbb{R}$, we write $g(x^-) = \lim_{y \uparrow x} g(y)$ and $g(x^+) = \lim_{y \downarrow x} g(y)$, for the left and right limits of g at $x \in \mathbb{R}$, respectively, and we write $g(\infty^-) = \lim_{x \to \infty} g(x)$ and $g([-\infty]^+) = \lim_{x \to -\infty} g(x)$. We also make use of the abbreviations $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$.

Next, we define the *upper affinity*

$$\eta_A^{(n)}(Q, \Delta) = \sup_{t \in \mathbb{R}} \pi\bigl(\operatorname{conv}\bigl(Q\mathcal{P}_{\le t}^{(n)}\bigr), \operatorname{conv}\bigl(Q\mathcal{P}_{\ge t+\Delta}^{(n)}\bigr)\bigr) \tag{2.10}$$

and its generalized inverse for $\eta \in [0, 1)$,

$$\Delta_A^{(n)}(Q, \eta) = \sup\bigl\{\Delta \ge 0 : \eta_A^{(n)}(Q, \Delta) > \eta\bigr\}. \tag{2.11}$$

Note that for $\eta < 1$ the set in the previous display is never empty, since $\eta_A^{(n)}(Q, 0) = 1$, and thus $\Delta_A^{(n)}(Q, \eta) \ge 0$. Also note that $\Delta \mapsto \eta_A^{(n)}(Q, \Delta)$ is nonincreasing.

In order to show that our subsequent lower bounds on $\mathcal{M}_{n,\alpha}(\mathcal{P}, \theta)$ are attained for convex and dominated models $\mathcal{P}$ and linear and bounded functionals $\theta : \mathcal{P} \to \mathbb{R}$, we will need the following consequence of a fundamental minimax theorem of Sion (1958), Corollary 3.3; see Section H.1 of the Supplementary Material for the proof.

PROPOSITION 2.1. *Fix constants $-\infty < a \le b < \infty$. Let $\mathbb{S}$ be a convex set of finite signed measures on a measurable space $(\Omega, \mathcal{A})$, so that $\mathbb{S}$ is dominated by a σ-finite measure μ. Furthermore, let $\mathbb{T} = \{\phi \in L_\infty(\Omega, \mathcal{A}, \mu) : a \le \int_\Omega \phi f \, d\mu \le b, \forall f \in L_1(\Omega, \mathcal{A}, \mu) : \|f\|_{L_1} \le 1\}$. Then,*

$$\sup_{\phi \in \mathbb{T}} \inf_{\sigma \in \mathbb{S}} \int_\Omega \phi \, d\sigma = \inf_{\sigma \in \mathbb{S}} \sup_{\phi \in \mathbb{T}} \int_\Omega \phi \, d\sigma.$$

Proposition 2.1 implies that for arbitrary subsets $\mathcal{P}_0$ and $\mathcal{P}_1$ of $\mathcal{P}$, and if the class $Q\mathcal{P}^{(n)}$ is dominated by some σ-finite measure (note that this is always the case if Q is α-private), we have the identity

$$\begin{aligned}
&\inf_{\text{tests } \phi} \sup_{\substack{\mathbb{P}_0 \in Q\mathcal{P}_0^{(n)} \\ \mathbb{P}_1 \in Q\mathcal{P}_1^{(n)}}} \mathbb{E}_{\mathbb{P}_0}[\phi] + \mathbb{E}_{\mathbb{P}_1}[1-\phi] \\
&\quad = \sup_{\substack{\mathbb{P}_0 \in \operatorname{conv}(Q\mathcal{P}_0^{(n)}) \\ \mathbb{P}_1 \in \operatorname{conv}(Q\mathcal{P}_1^{(n)})}} \inf_{\text{tests } \phi} \mathbb{E}_{\mathbb{P}_0}[\phi] + \mathbb{E}_{\mathbb{P}_1}[1-\phi] \qquad (2.12) \\
&\quad = \pi\big(\operatorname{conv}(Q\mathcal{P}_0^{(n)}), \operatorname{conv}(Q\mathcal{P}_1^{(n)})\big).
\end{aligned}$$

To see this, note that the left-hand side of (2.12) does not change if we replace $Q\mathcal{P}_r^{(n)}$ by its convex hull, for $r = 0, 1$, because for $\mathbb{P}_{r,i} \in Q\mathcal{P}_r^{(n)}$,

$$\begin{aligned}
\mathbb{E}_{\sum_{i=1}^k \alpha_i \mathbb{P}_{0,i}}[\phi] + \mathbb{E}_{\sum_{j=1}^l \beta_j \mathbb{P}_{1,j}}[1-\phi] &= \sum_{i,j} \alpha_i \beta_j \big(\mathbb{E}_{\mathbb{P}_{0,i}}[\phi] + \mathbb{E}_{\mathbb{P}_{1,j}}[1-\phi]\big) \\
&\le \sup_{\substack{\mathbb{P}_0 \in Q\mathcal{P}_0^{(n)} \\ \mathbb{P}_1 \in Q\mathcal{P}_1^{(n)}}} \mathbb{E}_{\mathbb{P}_0}[\phi] + \mathbb{E}_{\mathbb{P}_1}[1-\phi].
\end{aligned}$$

Now, apply Proposition 2.1 with $\mathbb{S} = \{\mathbb{P}_0 - \mathbb{P}_1 : \mathbb{P}_r \in \operatorname{conv}(Q\mathcal{P}_r^{(n)}), r = 0, 1\}$ and $a = 0$, $b = 1$.

The identity (2.12) was prominently used by Donoho and Liu (1991)—in the nonprivate case where $Q(A|x) = \mathbb{1}_A(x)$—in order to derive their lower bounds on the minimax risk. It is due to C. Kraft and L. Le Cam (Theorem 5 of Kraft (1955), see also page 40 of LeCam (1973)), who derived it more directly. We will also make use of (2.12) to derive lower bounds (see the proof of Theorem A.1 in the Supplementary Material). However, in order to show that there exist channel distributions $Q^{(n)}$ so that $\mathcal{M}_n(Q^{(n)}, \mathcal{P}, \theta)$ attains the rate of the lower bound, we need the generality of Proposition 2.1 (see Section 4.2 below).

3. A general lower bound on the α-private minimax risk. In this section we establish a lower bound on $\mathcal{M}_{n,\alpha}(\mathcal{P}, \theta) = \inf_{Q \in \mathcal{Q}_\alpha} \mathcal{M}_n(Q, \mathcal{P}, \theta)$, $\alpha \in (0, \infty)$, in terms of the total variation and Hellinger moduli of continuity ω_{TV} and ω_{H} of the functional $\theta : \mathcal{P} \to \mathbb{R}$. We also bridge the gap to the nonprivate case $\alpha = \infty$ in which the rate is characterized by ω_{H} only, and therefore, we extend results of Donoho and Liu (1991) to the case of privatized data. These extensions, however, do not constitute our main contribution. Therefore, we defer the technical details to Section A of the Supplementary Material. Our main conceptual innovation is to show that the lower bounds are rate optimal for a large class of possible estimation problems.

COROLLARY 3.1. *Fix $\eta_0, \varepsilon_0 \in (0, 1)$, $\alpha \in (0, \infty)$, and let $l : \mathbb{R}_+ \to \mathbb{R}_+$ be a nondecreasing loss function. Then, there exists a positive finite constant $c = c(\eta_0, \varepsilon_0)$, such that for all $\eta \in (0, \eta_0)$ and for all $n > |\log \eta|/\varepsilon_0$,*

$$\begin{aligned}
\mathcal{M}_{n,\alpha}(\mathcal{P}, \theta) &= \inf_{Q \in \mathcal{Q}_\alpha} \mathcal{M}_n(Q, \mathcal{P}, \theta) \\
&\ge l\bigg(\frac{1}{2}\bigg[\omega_{\mathrm{TV}}\bigg(\bigg[\frac{1-\eta}{\sqrt{2n(e^\alpha - 1)^2}}\bigg]^-\bigg) \vee \omega_{\mathrm{H}}\bigg(\bigg[c\sqrt{\frac{|\log \eta|}{n}}\bigg]^-\bigg)\bigg]^-\bigg)\frac{\eta}{2},
\end{aligned}$$

where $\mathcal{Q}_\alpha$ is the set of α-sequentially interactive channels Q as in (2.6).

Corollary 3.1 extends the lower bound of Donoho and Liu (1991) to privatized data. We point out that a similar lower bound with slightly worse constants can be easily derived from Proposition 1 of Duchi, Jordan and Wainwright (2018). In general, we have $\omega_{\mathrm{H}}(\varepsilon) \le \omega_{\mathrm{TV}}(\varepsilon)$, because $d_{\mathrm{TV}}(\mathbb{P}_0, \mathbb{P}_1) \le d_{\mathrm{H}}(\mathbb{P}_0, \mathbb{P}_1)$. Therefore, privatization leads to a larger lower bound compared to the direct case. This is hardly any surprise. Moreover, if α is sufficiently large, that is, the privatization constraint is weak, then the lower bound of Corollary 3.1 reduces to the classical lower bound in the case of direct estimation derived by Donoho and Liu (1991).

In our theory we consider the class $\mathcal{Q}_\alpha$ of α-sequentially interactive channels, because those admit a reasonably simple (cf. Duchi, Jordan and Wainwright (2018)) and attainable lower bound and they comprise a relevant class of *local* differential privatization mechanisms. In the next section we show that for estimation of linear functionals θ over convex parameter spaces $\mathcal{P}$ (and also for more general, but sufficiently regular θ and $\mathcal{P}$), the rate of our lower bound is attained even within the much smaller class of noninteractive channels. So within the class of sequentially interactive channels, the noninteractive channels already lead to rate optimal private estimation of linear functionals over convex parameter spaces.

REMARK 3.2. Corollary 3.1 does not restrict the values of $\alpha \in (0, \infty)$ and is formulated for any sample size n. In particular, it continues to hold if α is replaced by an arbitrary sequence $\alpha_n \in (0, \infty)$. The choice of this sequence has a fundamental impact on the private minimax rate of convergence. For example, if we consider the highly privatized case where $\alpha_n \asymp n^{-1/2}$, then $n(e^{\alpha_n} - 1)^2$ is bounded and the α_n-privatized minimax risk no longer converges to zero as $n \to \infty$.

4. Attainability of lower bounds. To establish upper bounds on the private minimax risk that match the rate of our lower bounds, some regularity conditions are needed. In the case where the channel Q is noninteractive and fixed, the main ingredients for a characterization of $\mathcal{M}_n(Q, \mathcal{P}, \theta)$ are a certain minimax identity and a type of second degree homogeneity of the privatized Hellinger modulus

$$\omega_{\mathrm{H}}^{(Q_1)}(\varepsilon) = \sup\bigl\{\bigl|\theta(\mathbb{P}_0) - \theta(\mathbb{P}_1)\bigr| : d_{\mathrm{H}}(Q_1\mathbb{P}_0, Q_1\mathbb{P}_1) \le \varepsilon, \mathbb{P}_0, \mathbb{P}_1 \in \mathcal{P}\bigr\},$$

where here $Q_1 : \mathcal{B}(\mathcal{Z}) \times \mathcal{X} \to [0, 1]$ is a *one*-dimensional marginal channel (see the discussion in Section B of the Supplementary Material for details). However, for the sake of readability, in the main article we only operate under the sufficient conditions that $\mathcal{P}$ is convex and dominated and that $\theta : \mathcal{P} \to \mathbb{R}$ is linear. Throughout this section we repeatedly make use of the following additional assumptions:

(A) The functional $\theta : \mathcal{P} \to \mathbb{R}$ of interest is bounded, that is, $\sup_{\mathbb{P} \in \mathcal{P}} |\theta(\mathbb{P})| < \infty$.

(B) The nondecreasing loss function $l : \mathbb{R}_+ \to \mathbb{R}_+$ is such that $l(0) = 0$ and $l(\frac{3}{2}t) \le al(t)$, for some $a \in (1, \infty)$ and for every $t \in \mathbb{R}_+$.

The boundedness Assumption A is also maintained in Donoho and Liu (1991). However, in their context, it is actually not necessary in some special cases such as the location model. On the other hand, the boundedness of θ appears to be much more fundamental in the case of private estimation. See, for example, Section G in Duchi, Jordan and Wainwright (2014), who show that in the privatized location model under squared error loss, Assumption A is necessary in order to obtain finite α-private minimax risk $\mathcal{M}_{n,\alpha}(\mathcal{P}, \theta)$. Assumption B is also taken from Donoho and Liu (1991). It is satisfied for many common loss functions, such as $l_\gamma(t) = t^\gamma$, with $\gamma > 0$, or the Huber loss $l_\gamma(t) = \mathbb{1}_{[0,\gamma)}(t)t^2/2 + \mathbb{1}_{[\gamma,\infty)}(t)\gamma(t - \gamma/2)$, which satisfies B with $a = 9/2$.

The following theorem (Theorem 4.1) provides sufficient conditions on the sequence of noninteractive channels $Q^{(n)} : \mathcal{B}(\mathcal{Z}^n) \times \mathcal{X}^n \to [0, 1]$ with identical marginals $Q_1^{(n)}$, the model

$\mathcal{P}$ and the functional θ, so that the privatized minimax risk $\mathcal{M}_n(Q^{(n)}, \mathcal{P}, \theta)$ is upper bounded by a constant multiple of

$$l \circ \omega_{\mathrm{H}}^{(Q_1^{(n)})}(n^{-1/2}).$$

A more general result is discussed and proved in Section B of the Supplementary Material. This even extends, and improves, the attainability result of Donoho and Liu (1991) in the nonprivate case. For the purpose of attainability under local differential privacy, the crucial point is the next one (cf. Theorem 4.2 below). Namely, to establish the existence of sequences of noninteractive α-private channels $Q^{(n)}$ that satisfy the imposed assumptions and are such that

$$\omega_{\mathrm{H}}^{(Q_1^{(n)})}(n^{-1/2}) \lesssim \omega_{\mathrm{TV}}(n^{-1/2}).$$

At this point our theory deviates conceptually from the one developed by Donoho and Liu (1991). We propose a class of α-private channels $Q_1^{(\alpha,\ell)}$ indexed by a functional parameter $\ell \in L_\infty$ and minimize the resulting Hellinger modulus of continuity

$$\omega_{\mathrm{H}}^{(Q_1^{(\alpha,\ell)})}(n^{-1/2})$$

with respect to ℓ. For this minimization to be successful we require another minimax identity to hold which is given by the conclusion of Proposition 2.1. Combining Theorem 4.1 and Theorem 4.2, which are stated below, then shows that the rate of the lower bound of the previous section can be attained.

4.1. *Upper bounds for given channel sequences.* An extended version of the following theorem (not assuming convexity, dominatedness and linearity), its proof and some further discussions are deferred to Section B of the Supplementary Material:

THEOREM 4.1. *Fix $n \in \mathbb{N}$; suppose that Conditions A and B hold and that Q is a noninteractive channel with identical marginals Q_1. Moreover, assume that $\mathcal{P}$ is dominated and convex and that $\theta : \mathcal{P} \to \mathbb{R}$ is linear. Fix $C \geq \sqrt{2\log 2a} + 1$, $\Delta = C^2 \omega_{\mathrm{H}}^{(Q_1)}(n^{-1/2})$ and $C_1 = [1 + \frac{8a^2}{2a-1}]a^{\lceil 2\log(C)/\log(3/2)\rceil}$, where $a > 1$ is the constant from Condition B. Then, there exists a binary search estimator $\hat{\theta}_n^{(\Delta)} : \mathcal{Z}^n \to \mathbb{R}$ with tuning parameter Δ (cf. Proposition 7.1 for details), such that*

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{Q\mathbb{P}^{\otimes n}}\big[l\big(\big|\hat{\theta}_n^{(\Delta)} - \theta(\mathbb{P})\big|\big)\big] \leq C_1 \cdot l\big(\omega_{\mathrm{H}}^{(Q_1)}(n^{-1/2})\big).$$

The general version of this theorem (see Section B of the Supplementary Material) is a strict generalization of results of Donoho and Liu (1991) to cover also the case where $Q(A|x)$ is an arbitrary noninteractive channel with identical marginals and not necessarily equal to $\mathbb{1}_A(x)$. Concerning its proof, we introduce a binary search estimator different from the one used by Donoho and Liu (1991) which, in particular, takes the privatized data as input data. Our new construction also has the advantage that it facilitates a detailed but highly nontrivial analysis in the case of the specific binary privatization scheme introduced in Section 4.2 below. The crucial point is that in conjunction with this privatization scheme, our minimax optimal binary search estimator can even be shown to be nearly linear (see Section 4.3). The details of our construction and an in-depth analysis of the estimator when based on this specific privatization scheme is presented in Section 7.

4.2. *A general attainability result.* The challenge in deriving rate optimal upper bounds on the α-private minimax risk $\mathcal{M}_{n,\alpha}(\mathcal{P},\theta)$ is now to find α-sequentially interactive channel distributions Q, such that the upper bound of the form $l\circ\omega_{\mathrm{H}}^{(Q_1)}(n^{-1/2})$ on $\mathcal{M}_n(Q,\mathcal{P},\theta)$, obtained in Theorem 4.1, matches the rate of the lower bound

$$l\biggl(\frac{1}{2}\,\omega_{\mathrm{TV}}\biggl(\biggl[\frac{1-\eta}{\sqrt{2n(e^\alpha-1)^2}}\biggr]^-\biggr)\biggr)$$

of Corollary 3.1. It turns out that noninteractive channels with identical binary marginals lead to rate optimal procedures for α-private estimation of a large class of functionals. More precisely, we suggest to use a channel with binary marginals

$$Q_1^{(\alpha,\ell)}\bigl(\{\pm z_0\}|x\bigr)=\frac{1}{2}\biggl(1\pm\frac{\ell(x)}{z_0}\biggr), \tag{4.1}$$

where $z_0:=\|\ell\|_\infty\frac{e^\alpha+1}{e^\alpha-1}$ and where $\ell:\mathcal{X}\to\mathbb{R}$ is an appropriate measurable and bounded function. Note that

$$\sup_{S\in\mathcal{B}(\mathbb{R})}\frac{Q_1^{(\alpha,\ell)}(S|x_1)}{Q_1^{(\alpha,\ell)}(S|x_2)}=\max\biggl(\frac{1+\frac{\ell(x_1)}{\|\ell\|_\infty}\frac{e^\alpha-1}{e^\alpha+1}}{1+\frac{\ell(x_2)}{\|\ell\|_\infty}\frac{e^\alpha-1}{e^\alpha+1}},\frac{1-\frac{\ell(x_1)}{\|\ell\|_\infty}\frac{e^\alpha-1}{e^\alpha+1}}{1-\frac{\ell(x_2)}{\|\ell\|_\infty}\frac{e^\alpha-1}{e^\alpha+1}}\biggr)\le\frac{1+\frac{e^\alpha-1}{e^\alpha+1}}{1-\frac{e^\alpha-1}{e^\alpha+1}}=e^\alpha,$$

so that a noninteractive channel distribution with identical marginals (4.1) is α-private. Actually, the support $\mathcal{Z}=\{-z_0,z_0\}$ of $Q_1^{(\alpha,\ell)}$ has no effect on its privacy provisions. However, with this specific choice of its support, the channel $Q_1^{(\alpha,\ell)}$ has the property that the conditional expectation of Z_i given $X_i=x$ under $Q_1^{(\alpha,\ell)}$ equals $\int_{\mathcal{Z}}z\,Q_1^{(\alpha,\ell)}(dz|x)=-z_0Q_1^{(\alpha,\ell)}(\{-z_0\}|x)+z_0Q_1^{(\alpha,\ell)}(\{z_0\}|x)=\ell(x)$.

To motivate the choice in (4.1), we make the following observation. The channel (4.1) has the nice feature that for $\mathbb{P}_0,\mathbb{P}_1\in\mathcal{P}$ with densities p_0 and p_1 with respect to $\mu=\mathbb{P}_0+\mathbb{P}_1$, we have

$$\begin{aligned}
&d_{\mathrm{TV}}\bigl(Q_1^{(\alpha,\ell)}\mathbb{P}_0,Q_1^{(\alpha,\ell)}\mathbb{P}_1\bigr)\\
&\quad=\sup_{A\in\mathcal{B}(\mathbb{R})}\biggl|\int_{\mathcal{X}}Q_1^{(\alpha,\ell)}(A|x)p_0(x)\,d\mu(x)-\int_{\mathcal{X}}Q_1^{(\alpha,\ell)}(A|x)p_1(x)\,d\mu(x)\biggr|\\
&\quad=\max\biggl\{\biggl|\int_{\mathcal{X}}\frac{1}{2}\biggl(1+\frac{\ell(x)}{z_0}\biggr)\bigl[p_0(x)-p_1(x)\bigr]\,d\mu(x)\biggr|,\\
&\qquad\biggl|\int_{\mathcal{X}}\frac{1}{2}\biggl(1-\frac{\ell(x)}{z_0}\biggr)\bigl[p_0(x)-p_1(x)\bigr]\,d\mu(x)\biggr|\biggr\}\\
&\quad=\biggl|\int_{\mathcal{X}}\frac{\ell(x)}{2z_0}\bigl[p_0(x)-p_1(x)\bigr]\,d\mu(x)\biggr|=\frac{1}{2z_0}\bigl|\mathbb{E}_{\mathbb{P}_0}[\ell]-\mathbb{E}_{\mathbb{P}_1}[\ell]\bigr|.
\end{aligned}\tag{4.2}$$

If the functional of interest is actually of the form $\theta(\mathbb{P})=\mathbb{E}_{\mathbb{P}}[\ell]$, then we can use the fact that $d_{\mathrm{TV}}\le d_{\mathrm{H}}$ to see that

$$\begin{aligned}
\omega_{\mathrm{H}}^{(Q_1^{(\alpha,\ell)})}(\varepsilon)&=\sup\bigl\{\bigl|\theta(\mathbb{P}_0)-\theta(\mathbb{P}_1)\bigr|:d_{\mathrm{H}}\bigl(Q_1^{(\alpha,\ell)}\mathbb{P}_0,Q_1^{(\alpha,\ell)}\mathbb{P}_1\bigr)\le\varepsilon,\mathbb{P}_0,\mathbb{P}_1\in\mathcal{P}\bigr\}\\
&\le\sup\bigl\{\bigl|\theta(\mathbb{P}_0)-\theta(\mathbb{P}_1)\bigr|:\bigl|\theta(\mathbb{P}_0)-\theta(\mathbb{P}_1)\bigr|\le 2z_0\varepsilon,\mathbb{P}_0,\mathbb{P}_1\in\mathcal{P}\bigr\}\\
&\le 2\|\ell\|_\infty\frac{e^\alpha+1}{e^\alpha-1}\varepsilon.
\end{aligned}$$

But at least for convex $\mathcal{P}$ and nonconstant and linear θ, Lemma H.2 in Section H of the Supplementary Material shows that $\omega_{\mathrm{TV}}(\varepsilon)\ge c_0\varepsilon$, for some positive constant c_0 and every

small $\varepsilon > 0$. Thus,

$$\omega_{\mathrm{H}}^{(Q_1^{(\alpha,\ell)})}(n^{-1/2}) \le 2\|\ell\|_\infty(e^\alpha+1)c_0^{-1}\cdot\omega_{\mathrm{TV}}\biggl(\sqrt{\frac{1}{n(e^\alpha-1)^2}}\biggr),$$

for all large n. In general, if the functional $\theta : \mathcal{P} \to \mathbb{R}$ is of a more complicated form, then we have to find a sequence (ℓ_n) in L_∞ for which

$$\omega_{\mathrm{H}}^{(Q_1^{(\alpha,\ell_n)})}(n^{-1/2}) \lesssim \omega_{\mathrm{TV}}\biggl(\sqrt{\frac{1}{n(e^\alpha-1)^2}}\biggr). \tag{4.3}$$

The following result realizes the claim of the previous display using Proposition 2.1. An extended version of it is stated and proved in Section C of the Supplementary Material.

THEOREM 4.2. *For $\alpha \in (0,\infty)$ and $\ell \in L_\infty(\mathcal{X})$, let $Q^{(\alpha,\ell)}$ be the noninteractive α-private channel with identical marginals $Q_1^{(\alpha,\ell)}$ as in (4.1). If $\mathcal{P}$ is convex and dominated and $\theta : \mathcal{P} \to \mathbb{R}$ is linear, then*

$$\inf_{\ell:\|\ell\|_\infty\le 1} \omega_{\mathrm{H}}^{(Q_1^{(\alpha,\ell)})}(\varepsilon) \le \omega_{\mathrm{TV}}\biggl(\biggl[\varepsilon\frac{e^\alpha+1}{e^\alpha-1}\biggr]^+\biggr) \quad \forall \varepsilon > 0.$$

PROOF. For $s \ge 0$, define

$$\Phi_\ell(s) := \sup\bigl\{\theta(\mathbb{P}_0)-\theta(\mathbb{P}_1) : \mathbb{P}_0,\mathbb{P}_1 \in \mathcal{P}, \bigl|\mathbb{E}_{\mathbb{P}_0}[\ell]-\mathbb{E}_{\mathbb{P}_1}[\ell]\bigr| \le s\bigr\},$$

and note that $d_{\mathrm{TV}} \le d_{\mathrm{H}}$, $\|\ell\|_\infty \le 1$ and (4.2), imply

$$\omega_{\mathrm{H}}^{(Q^{(\alpha,\ell)})}(\varepsilon) \le \Phi_\ell\biggl(2\varepsilon\frac{e^\alpha+1}{e^\alpha-1}\biggr).$$

Clearly, the function Φ_ℓ is nondecreasing. For $t \ge 0$, define $\Psi_\ell(t) := \inf\{s \ge 0 : \Phi_\ell(s) > t\}$. We claim that the functions Φ_ℓ and Ψ_ℓ have the following properties:

$$\begin{aligned} \Psi_\ell(t) > s \quad &\Rightarrow \quad \Phi_\ell(s) \le t, \\ \sup_{\ell:\|\ell\|_\infty\le 1} \Psi_\ell(t) > s \quad &\Rightarrow \quad \inf_{\ell:\|\ell\|_\infty\le 1} \Phi_\ell(s) \le t, \end{aligned} \tag{4.4}$$

$$\Psi_\ell(t) \ge \inf\bigl\{\bigl|\mathbb{E}_{\mathbb{P}_0}[\ell]-\mathbb{E}_{\mathbb{P}_1}[\ell]\bigr| : \theta(\mathbb{P}_0)-\theta(\mathbb{P}_1) \ge t, \mathbb{P}_0,\mathbb{P}_1 \in \mathcal{P}\bigr\}. \tag{4.5}$$

The first two are obvious. To establish (4.5), set $A_{\ell^*}(t) := \{s \ge 0 : \Phi_\ell(s) > t\}$ and $B_\ell(t) := \{|\mathbb{E}_{\mathbb{P}_0}[\ell]-\mathbb{E}_{\mathbb{P}_1}[\ell]| : \theta(\mathbb{P}_0)-\theta(\mathbb{P}_1) \ge t, \mathbb{P}_0,\mathbb{P}_1 \in \mathcal{P}\}$, and note that for $A_\ell(t) = \varnothing$ the claim is trivial. So, take $s \in A_\ell(t)$. Then, $\Phi_\ell(s) > t$, which implies that there are $\mathbb{P}_0,\mathbb{P}_1 \in \mathcal{P}$ with $|\mathbb{E}_{\mathbb{P}_0}[\ell]-\mathbb{E}_{\mathbb{P}_1}[\ell]| \le s$ and $\theta(\mathbb{P}_0)-\theta(\mathbb{P}_1) > t$. Thus, $\nu := |\mathbb{E}_{\mathbb{P}_0}[\ell]-\mathbb{E}_{\mathbb{P}_1}[\ell]| \le s$ and $\nu \in B_\ell(t)$. We have just shown that for every $s \in A_\ell(t)$ there exists a $\nu \in B_\ell(t)$ with $\nu \le s$. But this clearly means that $\Psi_\ell(t) = \inf A_\ell(t) \ge \inf B_\ell(t)$, as required.

Now, abbreviate $\eta := \varepsilon\frac{e^\alpha+1}{e^\alpha-1}$, $\delta := \omega_{\mathrm{TV}}(\eta+\xi_0)+\xi_1$, for $\xi_0,\xi_1 > 0$, and note that $\mathbb{T} := \{\phi \in L_\infty(\mathcal{X},\mathcal{F},\mu) : -1 \le \int_{\mathcal{X}} \phi f\,d\mu \le 1, \forall f \in L_1(\mathcal{X},\mathcal{F},\mu) : \|f\|_{L_1} \le 1\} = \{\phi \in L_\infty(\mathcal{X},\mathcal{F},\mu) : \|\phi\|_\infty \le 1\}$. Using convexity and dominatedness of $\mathcal{P}$ together with linearity of θ, we see that $\mathbb{S}_\delta := \{\mathbb{P}_0-\mathbb{P}_1 : \theta(\mathbb{P}_0)-\theta(\mathbb{P}_1) \ge \delta, \mathbb{P}_0,\mathbb{P}_1 \in \mathcal{P}\}$ is a dominated convex set of finite signed measures. Hence,

$$\sup_{\ell:\|\ell\|_\infty\le 1}\inf_{\sigma\in\mathbb{S}_\delta}\int_{\mathcal{X}} \ell\,d\sigma = \inf_{\sigma\in\mathbb{S}_\delta}\sup_{\ell:\|\ell\|_\infty\le 1}\int_{\mathcal{X}} \ell\,d\sigma,$$

follows from Proposition 2.1 with $a = -1$ and $b = 1$. Therefore, (4.5) yields

$$\begin{aligned}\sup_{\ell:\|\ell\|_\infty\le 1}\Psi_\ell(\delta) &\ge \sup_{\ell:\|\ell\|_\infty\le 1}\inf_{\sigma\in\mathbb{S}_\delta}\left|\int_{\mathcal{X}}\ell\,d\sigma\right| \ge \sup_{\ell:\|\ell\|_\infty\le 1}\inf_{\sigma\in\mathbb{S}_\delta}\int_{\mathcal{X}}\ell\,d\sigma\\ &= \inf_{\sigma\in\mathbb{S}_\delta}\sup_{\ell:\|\ell\|_\infty\le 1}\int_{\mathcal{X}}\ell\,d\sigma\\ &= 2\inf_{\sigma\in\mathbb{S}_\delta}\|\sigma\|_{\mathrm{TV}}\\ &= 2\inf\bigl\{d_{\mathrm{TV}}(\mathbb{P}_0,\mathbb{P}_1):\theta(\mathbb{P}_0)-\theta(\mathbb{P}_1)\ge\omega_{\mathrm{TV}}(\eta+\xi_0)+\xi_1\bigr\}\\ &\ge 2\inf\bigl\{d_{\mathrm{TV}}(\mathbb{P}_0,\mathbb{P}_1):\theta(\mathbb{P}_0)-\theta(\mathbb{P}_1)>\omega_{\mathrm{TV}}(\eta+\xi_0)\bigr\}\\ &\ge 2(\eta+\xi_0) > 2\eta = 2\varepsilon\frac{e^\alpha+1}{e^\alpha-1}.\end{aligned}$$

An application of (4.4) and letting $\xi_0 \to 0$ now finishes the proof. □

The next corollary now puts together Theorem 4.1 and Theorem 4.2. Its proof is deferred to Section D of the Supplementary Material. A somewhat more general version of this result that relaxes convexity of the model $\mathcal{P}$ and linearity of the functional θ is stated and proved in Section E of the Supplementary Material. We want to emphasize here that the assumptions of the more general result of Section E in the Supplementary Material can be verified, for instance, in the nonconvex case of estimating the endpoint of a uniform distribution (cf. Section G.4 in the Supplementary Material).

COROLLARY 4.3. *Fix $\alpha \in (0,\infty)$, $n \in \mathbb{N}$; suppose that Assumptions A and B hold, that $\mathcal{P}$ is convex and dominated and that $\theta : \mathcal{P} \to \mathbb{R}$ is linear. Then,*

$$\mathcal{M}_{n,\alpha}(\mathcal{P},\theta) := \inf_{Q\in\mathcal{Q}_\alpha}\mathcal{M}_n(Q,\mathcal{P},\theta) \le C_1\cdot l\biggl(\omega_{\mathrm{TV}}\biggl(\frac{4}{\sqrt{n}}\frac{e^\alpha+1}{e^\alpha-1}\biggr)\biggr),$$

where $\mathcal{Q}_\alpha$ is the collection of α-sequentially interactive channels as in (2.6). The constant C_1 is given by

$$C_1 = \biggl[1+\frac{8a^2}{2a-1}\biggr]a^{\lceil 2\log(C)/\log(3/2)\rceil+1},$$

where $C = \sqrt{2\log(2a)}+1$ and $a > 1$ is the constant from Condition B.

Summarizing, under the conditions of Corollary 4.3 and invoking our results of Section 3 (see also Section A in the Supplementary Material), we obtain the characterization (1.3) announced in the Introduction, that is, for any fixed $\alpha \in (0,\infty)$,

$$\mathcal{M}_{n,\alpha}(\mathcal{P},\theta) \asymp l\circ\omega_{\mathrm{TV}}\bigl(n^{-1/2}\bigr).$$

More precisely, we even find that for all $n \in \mathbb{N}$ and $\alpha \in (0,\infty)$,

$$\begin{aligned}\frac{1}{4}l\biggl(\frac{1}{2}\omega_{\mathrm{TV}}\biggl(\sqrt{\frac{1}{8n(e^\alpha-1)^2}}\biggr)\biggr) &\le \mathcal{M}_{n,\alpha}(\mathcal{P},\theta)\\ &\le C_1 l\biggl(\omega_{\mathrm{TV}}\biggl(\sqrt{\frac{16}{n}\frac{(e^\alpha+1)^2}{(e^\alpha-1)^2}}\biggr)\biggr).\end{aligned}$$

This also shows that, in the private case, the effective sample size reduces from n to $n\alpha^2$ for α small.

REMARK 4.4. Although our analysis was tailored to $\alpha \le 1$ and the bounds in the previous display are tight (up to universal constants) in that regime, they are not tight for large α, in the sense that $\frac{1}{e^\alpha - 1}$ and $\frac{e^\alpha+1}{e^\alpha-1}$ are vastly different for α large. An anonymous referee has pointed out the plausible conjecture that the correct scaling should actually be

$$\omega_{\mathrm{TV}}\biggl(\frac{1}{\sqrt{n(e^\alpha-1)^2}}\biggr) \vee \omega_{\mathrm{TV}}\biggl(\frac{1}{\sqrt{ne^{2\alpha/3}}}\biggr) \vee \omega_{\mathrm{H}}\biggl(\frac{1}{\sqrt{n}}\biggr),$$

for all $\alpha > 0$. This is motivated by the fact that the additive staircase mechanism $Z_i = X_i + G_i$, $G_i \sim f_{\gamma^*}$ of Geng and Viswanath (2016) has variance $\mathrm{Var}(Z_i) \lesssim T^2 e^{-2\alpha/3}$, provided that $|X_i| \le T$, thus leading to mean squared error for mean estimation of $\mathbb{E}[(\bar{Z}_n - \mathbb{E}[X_1])^2] \lesssim \frac{T^2}{ne^{2\alpha/3}}$. Inspired by this conjecture, we improved the α-dependence of our lower bounds for noninteractive channels to

$$\omega_{\mathrm{TV}}\biggl(\frac{1}{\sqrt{n(e^\alpha-1)^2}}\biggr) \vee \omega_{\mathrm{TV}}\biggl(\frac{1}{\sqrt{ne^{\alpha}}}\biggr) \vee \omega_{\mathrm{H}}\biggl(\frac{1}{\sqrt{n}}\biggr)$$

(see Section I of the Supplementary Material). The interesting question of whether one of these scalings is the correct one, or whether there exist even many more different regimes which describe the α-dependence in large generality, is left for future research. The fact that there is a regime change in the α-dependence of the private mean squared error was also observed by Duchi and Rogers (2019), and there is certainly more to say about the $\alpha > 1$ regime.

4.3. *Optimality of affine estimators.* In the previous subsection we have seen that a simple noninteractive channel with binary marginals $Q_1^{(\alpha,\ell)}$ supported on $\mathcal{Z} = \{-z_0, z_0\}$ with $z_0 = \|\ell\|_\infty \frac{e^\alpha+1}{e^\alpha-1}$ and such that $Q_1^{(\alpha,\ell)}(\{z_0\}|x) = \frac{1}{2}(1 + \frac{\ell(x)}{z_0})$ leads to rate optimal locally private estimation, if $\ell \in L_\infty(\mathcal{X})$ is chosen appropriately. This means that the actual observations $Z_1, \dots, Z_n$ that are available for estimation of $\theta(\mathbb{P})$ are i.i.d., according to a binary distribution on $\mathcal{Z}$ with probability of outcome z_0 equal to $[Q_1^{(\alpha,\ell)}\mathbb{P}](\{z_0\}) = \frac{1}{2}(1 + \frac{\mathbb{E}_\mathbb{P}[\ell]}{z_0})$. But therefore, clearly, $\bar{Z}_n = \frac{1}{n}\sum_{i=1}^n Z_i$ is sufficient for $\mathbb{P}$. It is important to note that this only works because the chosen channel $Q_1^{(\alpha,\ell)}$ is binary. By Rao–Blackwellization,

$$\mathbb{E}_{[Q_1^{(\alpha,\ell)}\mathbb{P}]^{(n)}}\bigl[l\bigl(\bigl|\hat{\theta}_n(Z) - \theta(\mathbb{P})\bigr|\bigr)\bigr] \ge \mathbb{E}_{[Q_1^{(\alpha,\ell)}\mathbb{P}]^{(n)}}\bigl[l\bigl(\bigl|\mathbb{E}\bigl[\hat{\theta}_n(Z)|\bar{Z}_n\bigr] - \theta(\mathbb{P})\bigr|\bigr)\bigr],$$

we conclude that, at least for convex loss functions l, there must be a minimax optimal estimator that is a function of $\bar{Z}_n$ only. In fact, we will show more than that. Under the additional assumptions that $\mathcal{P}$ is convex and dominated and θ is linear, there is a choice of ℓ and a constant $b \in \mathbb{R}$ such that $\bar{Z}_n + b$ is minimax rate optimal (possibly after projection onto the range of θ). The proof of the following result is deferred to Section D in the Supplementary Material. It crucially relies on Proposition 7.1(ii) where we show that our binary search estimator in conjunction with the binary channel $Q_1^{(\alpha,\ell)}$ is approximately affine.

COROLLARY 4.5. *Fix $\alpha \in (0,\infty)$, $n \in \mathbb{N}$; suppose that Assumptions A and B hold, that $\mathcal{P}$ is dominated and convex and that $\theta : \mathcal{P} \to \mathbb{R}$ is linear. Then, there exists a function $\ell^* \in L_\infty(\mathcal{X})$ and a constant $b \in \mathbb{R}$, such that*

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{[Q_1^{(\alpha,\ell^*)}\mathbb{P}]^{(n)}}\bigl[l\bigl(\bigl|\Pi[\bar{Z}_n + b] - \theta(\mathbb{P})\bigr|\bigr)\bigr] \le C_2 \cdot l\biggl(\omega_{\mathrm{TV}}\biggl(\frac{4}{\sqrt{n}}\frac{e^\alpha+1}{e^\alpha-1}\biggr)\biggr),$$

where $\Pi : \mathbb{R} \to [M_-, M_+] := \mathrm{cl}\,[\theta(\mathcal{P})]$ is the projection onto the closure of the range of θ, which must be an interval, and $Q_1^{(\alpha,\ell^*)}$ is the binary channel of (4.1). The constant C_2 is

given by

$$C_2 = \left[2 + a^2 + \frac{8a^2}{2a-1}\right] a^{\lceil 2\log(C)/\log(3/2)\rceil + 3},$$

where $C = \sqrt{\frac{3}{2}}[\max\{8(e^\alpha + 1), \sqrt{2\log 2a} + 1\} + 1]$ *and* $a > 1$ *is the constant from Condition* B.

It is remarkable, and perhaps somewhat surprising, that a private estimation procedure as simple as the one described above, can be rate optimal in such a broad class of different estimation problems. In particular, the sample mean $\bar{Z}_n = \frac{1}{n}\sum_{i=1}^n Z_i$ can never achieve a faster rate than $l(n^{-1/2})$. Correspondingly, Lemma H.2 in Section H of the Supplementary Material, in conjunction with the lower bound of Corollary 3.1, reveals (at least for convex $\mathcal{P}$) that $l(n^{-1/2})$ is the best possible rate of convergence in locally differentially private estimation problems.

A prominent example of a (nonprivate) estimation problem with faster optimal convergence rate than $l(n^{-1/2})$ is estimation of the endpoint of a uniform distribution. In that case, the nonprivate optimal rate is $l(n^{-1})$, which can not be attained by a sample mean estimator. Even though the set of uniform distributions is not convex, we show in Section G.4 of the Supplementary Material that $\bar{Z}_n = \frac{1}{n}\sum_{i=1}^n Z_i$ with $\ell(x) = 2x$ is rate optimal in the corresponding private estimation problem.

Detailed inspection of the proof of Corollary 4.5 reveals that the function $\ell^* \in L_\infty(\mathcal{X})$ is a solution of a certain saddle-point problem. Solving this explicitly is not always straightforward, even though the assumptions of the corollary guarantee existence. Therefore, in the next section we explore another direction for finding an appropriate candidate ℓ^* in the construction of the locally private estimation procedure of Corollary 4.5.

5. Constructing rate optimal privatization mechanisms and estimators. We have seen so far (Corollary 4.5) that the noninteractive mechanism generating α-private observations as

$$Z_i|X_i = \begin{cases} z_0 & \text{with probability } \frac{1}{2}\left(1 + \frac{\ell(X_i)}{z_0}\right), \\ -z_0 & \text{with probability } \frac{1}{2}\left(1 - \frac{\ell(X_i)}{z_0}\right), \end{cases}$$

with $z_0 = \|\ell\|_\infty \frac{e^\alpha+1}{e^\alpha-1}$ and an appropriate choice of the function $\ell : \mathcal{X} \to \mathbb{R}$ leads to minimax rate optimality of the sample mean $\bar{Z}_n = \frac{1}{n}\sum_{i=1}^n Z_i$ for estimating $\theta(\mathbb{P})$ (possibly after an appropriate shift and projection). Since this result of the previous Section 4.3 was nonconstructive, it remains to determine ℓ for practical use. By construction of the above privacy mechanism, we have

$$\mathbb{E}[\bar{Z}_n|X_1, \ldots, X_n] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[Z_i|X_i] = \frac{1}{n}\sum_{i=1}^n \ell(X_i) =: \tilde{\theta}_n^{(\ell)}.$$

This means, in particular, that the bias of $\bar{Z}_n$ is the same as that of the linear estimator $\tilde{\theta}_n^{(\ell)}$ in the estimation problem with direct observations $X_1, \ldots, X_n$, with worst case absolute bias denoted by

$$B_{\mathcal{P},\theta}(\ell) := \sup_{\mathbb{P}\in\mathcal{P}} \left|\mathbb{E}_\mathbb{P}[\ell] - \theta(\mathbb{P})\right|.$$

We are thus lead to ask, "What is the optimal choice of ℓ in the direct estimation problem using $\tilde{\theta}_n^{(\ell)}$?"

Clearly, there is no universal answer, but the solution must depend on the estimand θ. For instance, if $\theta(\mathbb{P}) = p(x_0)$, where $p = d\mathbb{P}/d\lambda$ is a Lebesgue density of $\mathbb{P}$, then one might take $\ell(x) = \frac{1}{h}K(\frac{x-x_0}{h})$ for some kernel function K and an appropriate bandwidth $h = h_n > 0$. Or, if $\theta(\mathbb{P}) = \mathbb{E}_{\mathbb{P}}[f]$, for some measurable $f : \mathcal{X} \to \mathbb{R}$, then $\ell = f$ is natural. Note however, that ℓ has to be bounded in order for the privacy mechanism to be well defined. Thus, if f is unbounded, ℓ has to be taken as a truncated version of f, for example, $\ell(x) = f(x)\mathbb{1}_{|f(x)|\le\frac{1}{h}}$.

Now classically, we would trade off the bias $B_{\mathcal{P},\theta}(\ell)$ of the estimator $\tilde{\theta}_n^{(\ell)}$ against its variance (or standard deviation)

$$\operatorname{Var}[\tilde{\theta}_n^{(\ell)}] = \frac{1}{n}\big(\mathbb{E}_{\mathbb{P}}[\ell^2] - \mathbb{E}_{\mathbb{P}}[\ell]^2\big).$$

Instead, we must now trade off the bias $B_{\mathcal{P},\theta}(\ell)$ of the private estimator $\bar{Z}_n$ against its variance which is easily seen to equal

$$\operatorname{Var}[\bar{Z}_n] = \frac{1}{n}\big(z_0^2 - \mathbb{E}_{\mathbb{P}}[\ell]^2\big) = \frac{1}{n}\left(\|\ell\|_\infty^2\left(\frac{e^\alpha+1}{e^\alpha-1}\right)^2 - \mathbb{E}_{\mathbb{P}}[\ell]^2\right).$$

This trade-off is still hard to do over general $\ell \in L_\infty(\mathcal{X})$. Therefore, guided by our examples above, we here restrict to parametric classes $\{\ell_h : h \in \mathbb{R}^k\}$. As mentioned before, the choice of the class is problem specific, but in what follows we isolate a high-level sufficient condition (Condition C below) for the class $\{\ell_h : h \in \mathbb{R}^k\}$ that allows for solving the bias-variance trade-off such that the solution ℓ_{h^*} leads to rate optimal private estimation. Furthermore, we then show that Condition C can be checked in many classical examples. In particular, Condition C also holds in cases where the model $\mathcal{P}$ is not convex (cf. Section G.4 in the Supplementary Material).

In the case $k = 1$, $h \in \mathbb{R}$, the mentioned regularity condition C below states that the collection of measurable functions, $\ell_h : \mathcal{X} \to \mathbb{R}$, $h > 0$, satisfies $\|\ell_h\|_\infty \lesssim h^{-s}$, for some $s \ge 0$, and is such that the worst case absolute bias

$$B_{\mathcal{P},\theta}(\ell_h) := \sup_{\mathbb{P}\in\mathcal{P}}\big|\mathbb{E}_{\mathbb{P}}[\ell_h] - \theta(\mathbb{P})\big|$$

of the private estimator $\bar{Z}_n$ is bounded by an expression of the order h^t, as $h \to 0$, for some $t > 0$. We then show that for the choice of tuning parameter

$$h = h_n = \left(\frac{1}{\sqrt{n}}\frac{e^\alpha+1}{e^\alpha-1}\right)^{\frac{1}{s+t}},$$

the above privatization and estimation protocol based on ℓ_{h_n} is α-private minimax rate optimal if $\varepsilon^r \lesssim \omega_{\mathrm{TV}}(\varepsilon)$ for $r = t/(s+t)$. This consideration misleadingly suggests that the estimator $\frac{1}{n}\sum_{i=1}^n \ell_h(X_i)$ is minimax optimal in the nonprivate case for a possibly different choice of $h = \tilde{h}_n$. Although this appears to be correct in Examples G.1, G.2 and G.3 (see Section 6 below and the Supplementary Material for details), it is not true in general (see Example G.4 where the minimax rate optimal estimator in the direct problem is not even of linear form).

For some estimation problems, such as estimating a multivariate anisotropic density at a point (cf. Section G.3), the case $k = 1$ is not sufficient, and we need the full flexibility of Condition C:

(C) Suppose that $\mathcal{P}$ and θ are such that there exists $k \in \mathbb{N}$, $t \in (0,\infty)^k$, $s \in [0,\infty)^k$, $D_0 \in (0,\infty)$ and $h_0 \in (0,1]$ and a class of measurable functions $\ell_h : \mathcal{X} \to \mathbb{R}$ indexed by $h \in \mathbb{R}^k$, such that for all $h \in (0,h_0]^k$,

$$\|\ell_h\|_\infty \le D_0\prod_{j=1}^k h_j^{-s_j} \quad\text{and}\quad B_{\mathcal{P},\theta}(\ell_h) \le D_0\frac{1}{k}\sum_{j=1}^k h_j^{t_j}. \tag{5.1}$$

REMARK 5.1. Note that Condition C implies Condition A, because the converse of A implies that $B_{\mathcal{P},\theta}(\ell)$ is infinite whenever ℓ is bounded.

The proof of the following theorem is deferred to Section F in the Supplementary Material.

THEOREM 5.2. *Suppose that Conditions B and C hold, and set* $\bar{r} = \sum_{j=1}^{k} \frac{s_j}{t_j}$. *For* $\alpha \in (0, \infty)$, *let* $Q^{(\alpha,\ell)}$ *be the* α*-private channel with identical marginals* (4.1), *and set* $h_n = (h_{n,1}, \dots, h_{n,k})^T$ *and*

$$h_{n,j} = \left(\frac{1}{\sqrt{n}} \frac{e^\alpha + 1}{e^\alpha - 1} \right)^{\frac{1}{t_j(1+\bar{r})}}.$$

Then, the arithmetic mean $\bar{Z}_n(z) := \frac{1}{n} \sum_{i=1}^{n} z_i$, $z = (z_1, \dots, z_n)' \in \mathbb{R}^n$, *satisfies*

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{Q^{(\alpha,\ell_{h_n})} \mathbb{P}^{\otimes n}} \left[l\left(\left| \bar{Z}_n - \theta(\mathbb{P}) \right| \right) \right] \le C_0 \cdot l\left(\left(\frac{1}{\sqrt{n}} \frac{e^\alpha + 1}{e^\alpha - 1} \right)^{\frac{1}{1+\bar{r}}} \right), \tag{5.2}$$

for all $n \in \mathbb{N}$ *and a positive finite constant* C_0 *that depends only on* a *and* D_0.

The private estimator $\bar{Z}_n$ of Theorem 5.2 is α-private minimax rate optimal if the derived upper bound (5.2) on the worst case risk is of the same order as the lower bound of Corollary 3.1. The latter is true if $\varepsilon^{\frac{1}{1+\bar{r}}} \lesssim \omega_{\mathrm{TV}}(\varepsilon)$, for all small $\varepsilon > 0$. That this is often satisfied simultaneously with the conditions of Theorem 5.2 is demonstrated in the examples of Section 6. Lemma H.7 in the Supplementary Material shows that under Condition C the corresponding upper bound $\omega_{\mathrm{TV}}(\varepsilon) \lesssim \varepsilon^{\frac{1}{1+\bar{r}}}$ holds.

6. Examples. In this section we discuss several concrete estimation problems for which we derive bounds on the total variation modulus

$$\omega_{\mathrm{TV}}(\varepsilon) = \sup\left\{ \left| \theta(\mathbb{P}_0) - \theta(\mathbb{P}_1) \right| : d_{\mathrm{TV}}(\mathbb{P}_0, \mathbb{P}_1) \le \varepsilon, \mathbb{P}_j \in \mathcal{P} \right\},$$

which characterizes the rate of local differentially private estimation, and compare them to the Hellinger modulus $\omega_{\mathrm{H}}(\varepsilon)$ that determines the estimation rate under direct observations (see Table 1). Furthermore, we exhibit families of functions ℓ_h which, in conjunction with the

TABLE 1

Comparison of Hellinger (nonprivate) and total variation (private) moduli of continuity for several estimation problems. The minimax rate of convergence (for fixed α*) in each problem is given by* $l \circ \omega(n^{-1/2})$, *where* l *is the loss function*

$\mathcal{P}$	$\theta : \mathcal{P} \to \mathbb{R}$		$\omega_{\mathrm{H}}(\varepsilon)$	$\omega_{\mathrm{TV}}(\varepsilon)$
$\{\mathbb{P} : \mathbb{E}_{\mathbb{P}}[\lvert f \rvert^\kappa] \le L\}$ $L > 0, \kappa > 1$	$\mathbb{P} \mapsto \mathbb{E}_P[f]$	$\lVert f \rVert_\infty < \infty$ $\lvert f \rvert(\mathcal{X}) \supseteq (0, \infty)$	ε $\varepsilon^{(2\frac{\kappa-1}{\kappa}) \wedge 1}$	ε $\varepsilon^{\frac{\kappa-1}{\kappa}}$
$\mathcal{H}^{\ll\lambda}_{\beta,L}(\mathbb{R})$ $L > 0, \beta > 0$	$\mathbb{P} \mapsto p^{(m)}(x_0)$		$\varepsilon^{\frac{\beta-m}{\beta+1/2}}$	$\varepsilon^{\frac{\beta-m}{\beta+1}}$
$\mathcal{H}^{\ll\lambda}_{\beta,L}(\mathbb{R}^d)$ $L \in \mathbb{R}^d_+, \beta \in (0, 1]^d$	$\mathbb{P} \mapsto p(x_0)$	$\bar{r} = \sum_{j=1}^{d} \frac{1}{\beta_j}$	$\varepsilon^{\frac{1}{1+\bar{r}/2}}$	$\varepsilon^{\frac{1}{1+\bar{r}}}$
$\mathrm{Unif}[0, \vartheta]$ $\vartheta \in (0, M]$	$\mathbb{P} \mapsto \vartheta$		ε^2	ε

TABLE 2
Examples for the choice of class $\{\ell_h : h \in \mathbb{R}^k\}$ leading to rate optimal estimation in all the problems of Table 1

$\mathcal{P}$	$\theta : \mathcal{P} \to \mathbb{R}$		$\ell_h(x)$
$\{\mathbb{P} : \mathbb{E}_{\mathbb{P}}[\lvert f\rvert^\kappa] \le L\}$ $L > 0, \kappa > 1$	$\mathbb{P} \mapsto \mathbb{E}_{\mathbb{P}}[f]$	$\lVert f\rVert_\infty < \infty$ $\lvert f\rvert(\mathcal{X}) \supseteq (0, \infty)$	$f(x)$ $f(x)\mathbb{1}_{\lvert f(x)\rvert \le \frac{1}{h}}$
$\mathcal{H}^{\ll\lambda}_{\beta,L}(\mathbb{R})$ $L > 0, \beta > 0$	$\mathbb{P} \mapsto p^{(m)}(x_0)$		$\mathbb{1}_{[-1,1]}(\frac{x-x_0}{h}) \cdot \frac{d^m}{dx^m} \frac{1}{h} K(\frac{x-x_0}{h})$
$\mathcal{H}^{\ll\lambda}_{\beta,L}(\mathbb{R}^d)$ $L \in \mathbb{R}^d_+, \beta \in (0,1]^d$	$\mathbb{P} \mapsto p(x_0)$	$\bar{r} = \sum_{j=1}^d \frac{1}{\beta_j}$	$\prod_{j=1}^d \frac{1}{h_j} K(\frac{x_j - x_{0,j}}{h_j})$
$\mathrm{Unif}[0, \vartheta]$ $\vartheta \in (0, M]$	$\mathbb{P} \mapsto \vartheta$		$2x$

binary construction in (4.1) and an appropriate choice of tuning parameters, lead to minimax rate optimal locally private estimation procedures (see Table 2).

Even in cases where the moduli of continuity are hard to evaluate explicitly, the following relationship is always true:

$$\omega_{\mathrm{H}}(\varepsilon) \le \omega_{\mathrm{TV}}(\varepsilon) \le \omega_{\mathrm{H}}(\sqrt{2\varepsilon}) \quad \forall \varepsilon > 0,$$

because $d_{\mathrm{TV}} \le d_{\mathrm{H}} \le \sqrt{2 d_{\mathrm{TV}}}$ (cf. for instance Tsybakov (2009), Lemma 2.3). This shows that, in the worst case, the private minimax rate of estimation is the square root of the nonprivate minimax rate, whereas, the private rate can never be better than the nonprivate one. Both extremal cases can occur; see examples below.

The details and proofs of all claims made in Tables 1 and 2 are deferred to Section G of the Supplementary Material, except for the well-known facts about the Hellinger modulus. Our list of examples is far from being exhaustive, but due to space constraints we present only a few classical cases for which the nonprivate rates are well known. The first column in each of the two following tables describes the statistical model $\mathcal{P}$, that is, the set of (marginal) data generating distributions, and the second column displays the functional $\theta : \mathcal{P} \to \mathbb{R}$ that is to be estimated. In the first row we consider moment estimation. Here, $\lvert f\rvert(\mathcal{X}) := \{y \in \mathbb{R} : \exists x \in \mathcal{X} : \lvert f(x)\rvert = y\}$ denotes the range of $x \mapsto \lvert f(x)\rvert$. In the second row we consider estimation of the mth derivative of a density at a fixed point $x_0 \in \mathcal{X} = \mathbb{R}$ over the class $\mathcal{H}^{\ll\lambda}_{\beta,L}(\mathbb{R})$ of Lebesgue densities on $\mathbb{R}$ that are Hölder continuous with exponent $\beta > m$. That is, $p \in \mathcal{H}^{\ll\lambda}_{\beta,L}(\mathbb{R})$ is $b := \lfloor \beta \rfloor$ times differentiable with bth derivative $p^{(b)}$ satisfying

$$\lvert p^{(b)}(x) - p^{(b)}(y)\rvert \le L\lvert x - y\rvert^{\beta - b} \quad \forall x, y \in \mathbb{R}.$$

In the third row of our tables, we consider density estimation at a point $x_0 \in \mathcal{X} = \mathbb{R}^d$ over the anisotropic class $\mathcal{H}^{\ll\lambda}_{\beta,L}(\mathbb{R}^d)$ of Lebesgue densities on $\mathbb{R}^d$ such that for every $j \in \{1, \dots, d\}$ and every $x, x' \in \mathbb{R}^d$,

$$\lvert p(x_1, \dots, x_{j-1}, x'_j, x_{j+1}, \dots, x_d) - p(x)\rvert \le L_j \lvert x'_j - x_j\rvert^{\beta_j}.$$

Finally, in the last row we consider estimating the endpoint of a uniform distribution. The representations of the moduli of continuity in the last two columns of Table 1 are to be understood as upper and lower bounds up to constants and for small values of $\varepsilon > 0$.

We see that already for moment estimation, both extreme cases mentioned above can occur. If f is bounded, then $\theta(\mathbb{P}) = \mathbb{E}_{\mathbb{P}}[f]$ can be estimated at $n^{-1/2}$ rate in both the locally private as well as in the direct case. If, however, the range of $\lvert f\rvert$ contains the whole positive real line, and we have no more than a second moment being bounded (i.e., $\kappa \in (1, 2]$), then the locally

private rate is the square root of the rate under direct observations.[2] The density estimation problems are intermediate cases. There is some price to be paid for local differential privacy in terms of convergence rate, but it is not as bad as the square root of the direct rate. Finally, estimating the endpoint of a uniform distribution is another instance of a worst case situation under local differential privacy.

7. Studying the binary search estimator. Within this section, let $M_- := \inf_{\mathbb{P}\in\mathcal{P}} \theta(\mathbb{P})$, $M_+ := \sup_{\mathbb{P}\in\mathcal{P}} \theta(\mathbb{P})$, $M := M_+ - M_-$ and $\mathbb{S}_\Delta := \{\mathbb{P}_0 - \mathbb{P}_1 : \theta(\mathbb{P}_0) - \theta(\mathbb{P}_1) \geq \Delta, \mathbb{P}_0, \mathbb{P}_1 \in \mathcal{P}\}$, for $\Delta \geq 0$. Furthermore, recall the upper affinity

$$\eta_A^{(n)}(Q, \Delta) = \sup_{t\in\mathbb{R}} \pi\big(\operatorname{conv}(Q\mathcal{P}_{\leq t}^{(n)}), \operatorname{conv}(Q\mathcal{P}_{\geq t+\Delta}^{(n)})\big)$$

from equation (2.10).

PROPOSITION 7.1. *Fix a finite constant $\Delta > 0$, and suppose that $-\infty < M_- < M_+ < \infty$. Let $Q : \mathcal{B}(\mathcal{Z}^n) \times \mathcal{X}^n \to [0,1]$ be a noninteractive channel distribution with identical marginals Q_1. Moreover, let $N = N(M, \Delta)$ be the smallest integer such that $N\Delta > M > 0$. For $l \in \mathbb{N}_0$, set $\eta_l = (l+1)\Delta$:*

(i) *If $Q_1\mathcal{P}$ is dominated (by a σ-finite measure), then there exists an estimator $\hat{\theta}_n^{(\Delta)} : \mathcal{Z}^n \to \mathbb{R}$ with tuning parameter Δ, such that for every $l \in \mathbb{N}_0$,*

$$\sup_{\mathbb{P}\in\mathcal{P}} Q\mathbb{P}^{\otimes n}\big(z \in \mathcal{Z}^n : |\hat{\theta}_n^{(\Delta)}(z) - \theta(\mathbb{P})| > \eta_l\big) \leq 4 \sum_{k=l+1}^{N-2} \big[\eta_A^{(n)}(Q, k\Delta) \vee 0\big],$$

and an empty sum is interpreted as equal to zero. Moreover, $\hat{\theta}_n^{(\Delta)}$ takes values in the set $\{M_- + j\Delta : j \in \{1, \ldots, N-1\}\}$ if $N \geq 3$, and $\hat{\theta}_n^{(\Delta)} \equiv (M_- + M_+)/2$ else. We set $\hat{\theta}_n^{(0)}(z) = (M_- + M_+)/2$.

(ii) *Fix $\alpha \in (0, \infty)$. Suppose that $\mathcal{P}$ is convex and $\theta : \mathcal{P} \to \mathbb{R}$ is linear and there exists $\ell^* \in L_\infty(\mathcal{X})$ such that $\|\ell^*\|_\infty \leq 1$ and*

$$\inf_{\sigma\in\mathbb{S}_\Delta} \int_{\mathcal{X}} \ell^* \, d\sigma > 0.$$

If $Q_1 = Q_1^{(\alpha,\ell^*)}$ is the binary channel of (4.1) with $\mathcal{Z} = \{-z_0, z_0\}$, then there exists an affine function $g^{(aff)} : \mathbb{R} \to \mathbb{R}$ such that

$$\big|\Pi_{M_-,M_+}\big[g^{(aff)}(\bar{z}_n)\big] - \hat{\theta}_n^{(\Delta)}(z)\big| \leq 2\Delta \quad \forall z \in \mathcal{Z}^n,$$

for $\hat{\theta}_n^{(\Delta)}$ as in part (i) and where $\Pi_{M_-,M_+} : \mathbb{R} \to [M_-, M_+]$ is the projection onto $[M_-, M_+]$.

REMARK 7.2. Part (ii) of Proposition 7.1 is used in the proof of Corollary 4.5 on optimality of sample mean estimators (see Section D of the Supplementary Material). There, the existence of ℓ^* with $\|\ell^*\|_\infty \leq 1$ and $\inf_{\sigma\in\mathbb{S}_\Delta} \int_{\mathcal{X}} \ell^* \, d\sigma > 0$, for an appropriate choice of

$$\Delta = C^2 \omega_{\mathrm{H}}^{(Q_1^{(\alpha,\ell^*)})}(n^{-1/2})$$

(depending also on ℓ^*) is established using Proposition 2.1 and Lemma H.6 (see also Theorem C.1 in Section C of the Supplementary Material).

[2]Note that this private minimax rate of convergence was already discovered by Duchi, Jordan and Wainwright (2018) but with a rate optimal channel sequence and estimator different from ours.

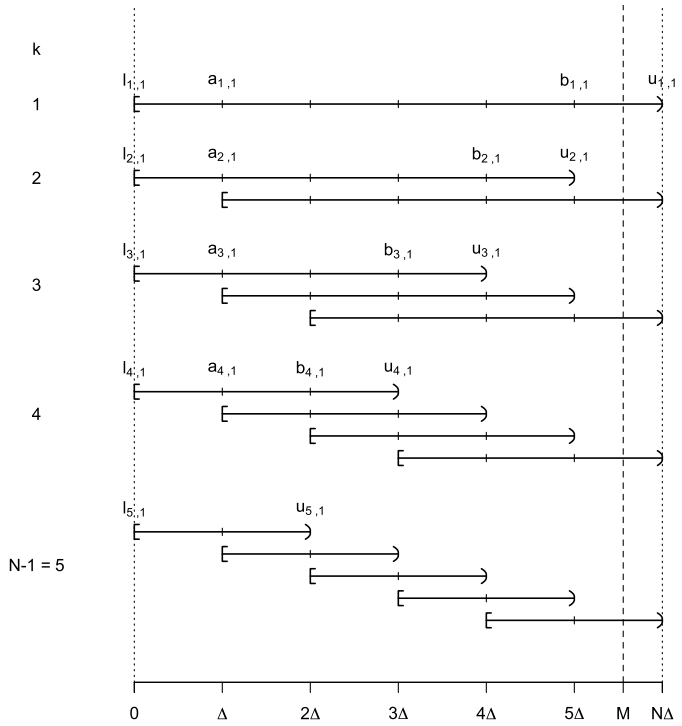

FIG. 1. *An example of the interval construction for the binary search estimator.*

PROOF OF PROPOSITION 7.1. Without loss of generality, we may assume that $0 = \inf_{\mathbb{P}\in\mathcal{P}} \theta(\mathbb{P}) < \sup_{\mathbb{P}\in\mathcal{P}} \theta(\mathbb{P}) = M$, by estimating $\theta(\mathbb{P}) - M_-$ instead of $\theta(\mathbb{P})$ and, in case of part (ii), by noting that $\Pi_{M_-,M_+}[x] - M_- = \Pi_{0,M}[x - M_-]$.

To rigorously introduce the binary search estimator, consider first the case where $\Delta > 0$ is such that $N = N(\Delta, M) \le 2$. In that case, we set $\hat{\theta}_n^{(\Delta)} \equiv M/2$, which satisfies the desired inequality trivially, because in this case $\Delta > M/2$ which implies $|\hat{\theta}_n^{(\Delta)}(z) - \theta(\mathbb{P})| = |M/2 - \theta(\mathbb{P})| \le M/2 < \Delta = \eta_0 \le \eta_l$. If $N \ge 3$, we shall construct the estimator $\hat{\theta}_n^{(\Delta)}$ such that it takes values in the set $\{j\Delta : j = 1, \dots, N-1\}$.

The construction is as follows: To select one of the values in the set $\{j\Delta : j = 1, \dots, N-1\}$ we first introduce a scheme to partition $[0, M)$ (cf. Figure 1). Start with the interval $[l_{1,1}, u_{1,1}) = [0, N\Delta)$, which contains $[0, M)$ by definition of N, and remove either the leftmost or the rightmost subinterval of length Δ, that is, $[0, \Delta)$ or $[(N-1)\Delta, N\Delta)$, to produce two new intervals $[l_{2,1}, u_{2,1}) = [0, (N-1)\Delta)$ and $[l_{2,2}, u_{2,2}) = [\Delta, N\Delta)$, each of length $(N-1)\Delta$. Then, proceed in the same way again to produce three (note that removing the leftmost subinterval of length Δ in the first step and then removing the rightmost in the second step results in the same interval as if we had removed them in the opposite order) new intervals $[l_{3,1}, u_{3,1})$, $[l_{3,2}, u_{3,2})$, $[l_{3,3}, u_{3,3})$, each of length $(N-2)\Delta$. Continue this process for $N-2$ steps to arrive at the intervals $[l_{N-1,j}, u_{N-1,j})$, $j = 1, \dots, N-1$, of length 2Δ whose midpoints are exactly the values $j\Delta$.

Formally, for $k \in \{1, \dots, N-1\}$ and $j \in \{1, \dots, k\}$, we set $l_{k,j} = (j-1)\Delta$, $u_{k,j} = l_{k,j} + (N-k+1)\Delta$, and we also define $a_{k,j} = l_{k,j} + \Delta$ and $b_{k,j} = u_{k,j} - \Delta$, so that $b_{k,j} - a_{k,j} = (N-k-1)\Delta =: d_k$. With each pair (k, j) as before, we associate a (randomized) minimax test $\xi_{k,j} : \mathcal{Z}^n \to [0, 1]$ for $H_0 : Q_1\mathcal{P}_{\le a_{k,j}} = \{Q_1\mathbb{P} : \theta(\mathbb{P}) \le a_{k,j}, \mathbb{P} \in \mathcal{P}\}$ against $H_1 : Q_1\mathcal{P}_{\ge b_{k,j}} = \{Q_1\mathbb{P} : \theta(\mathbb{P}) \ge b_{k,j}, \mathbb{P} \in \mathcal{P}\}$. Recall that such a minimax test has the property

that

$$\sup_{\substack{\mathbb{P}_0\in[Q_1\mathcal{P}_{\le a_{k,j}}]^{(n)}\\ \mathbb{P}_1\in[Q_1\mathcal{P}_{\ge b_{k,j}}]^{(n)}}} \mathbb{E}_{\mathbb{P}_0}(\xi_{k,j})+\mathbb{E}_{\mathbb{P}_1}(1-\xi_{k,j})$$
$$=\inf_{\text{tests }\phi}\sup_{\substack{\mathbb{P}_0\in[Q_1\mathcal{P}_{\le a_{k,j}}]^{(n)}\\ \mathbb{P}_1\in[Q_1\mathcal{P}_{\ge b_{k,j}}]^{(n)}}} \mathbb{E}_{\mathbb{P}_0}(\phi)+\mathbb{E}_{\mathbb{P}_1}(1-\phi).$$

Existence is well known (see Lemma H.4 in Section H of the Supplementary Material which is a minor modification of a result by Krafft and Witting (1967); see also Lehmann and Romano (2005), Problem 8.1 and Theorem A.5.1). To obtain a nonrandomized test from $\xi_{k,j}$, we set $\xi^*_{k,j}=\mathbb{1}_{(1/2,1]}(\xi_{k,j})$. Since $\mathbb{E}_{\mathbb{P}}[\xi^*_{k,j}]=\mathbb{P}(\xi_{k,j}>1/2)\le 2\mathbb{E}_{\mathbb{P}}[\xi_{k,j}]$ and $\mathbb{E}_{\mathbb{P}}[1-\xi^*_{k,j}]=\mathbb{P}(\xi_{k,j}\le 1/2)=\mathbb{P}(1-\xi_{k,j}\ge 1/2)\le 2\mathbb{E}_{\mathbb{P}}[1-\xi_{k,j}]$, we get in view of (2.12), that

$$\sup_{\substack{\mathbb{P}_0\in[Q_1\mathcal{P}_{\le a_{k,j}}]^{(n)}\\ \mathbb{P}_1\in[Q_1\mathcal{P}_{\ge b_{k,j}}]^{(n)}}} \mathbb{E}_{\mathbb{P}_0}(\xi^*_{k,j})+\mathbb{E}_{\mathbb{P}_1}(1-\xi^*_{k,j})$$
$$\le 2\pi\big(\operatorname{conv}(Q\mathcal{P}^{(n)}_{\le a_{k,j}}),\operatorname{conv}(Q\mathcal{P}^{(n)}_{\ge b_{k,j}})\big) \tag{7.1}$$
$$\le 2\eta^{(n)}_A(Q,b_{k,j}-a_{k,j})$$
$$=2\eta^{(n)}_A(Q,d_k).$$

By definition of M_- and M_+, all the sets $\mathcal{P}_{\le a_{k,j}}$ and $\mathcal{P}_{\ge b_{k,j}}$ are nonempty, with the only exception of $\mathcal{P}_{\ge b_{1,1}}$, which is empty if and only if, $(N-1)\Delta=M$ and θ does not attain its supremum $M=\sup_{\mathbb{P}\in\mathcal{P}}\theta(P)$. In that case, we take $\xi_{1,1}\equiv 0$.

To define the value of the binary search estimator $\hat\theta^{(\Delta)}_n(z)$ for a given observation $z\in\mathcal{Z}^n$, we perform a stepwise testing procedure (cf. Figure 2). Starting at the full interval $[0,N\Delta)$ $(k=1)$, we remove the outermost subinterval of length Δ that was rejected by the test $\xi^*_{1,1}$. Depending on the outcome of $\xi^*_{1,1}$ and the corresponding subinterval of length $(N-1)\Delta$, we are left with $(k=2)$; we either perform the test $\xi^*_{2,1}$ or $\xi^*_{2,2}$ and, again, remove the rejected interval of length Δ. We proceed in this way until $k=N-2$. Finally, we set $\hat\theta^{(\Delta)}_n(z)$ equal to the midpoint of the remaining interval of length 2Δ that was selected by the test performed at level $k=N-2$. This procedure leads to the formal definition: Set $j_1(z)=1$, and for $k\in\{2,\dots,N-1\}$, set $j_k(z)=j_{k-1}(z)+\xi^*_{k-1,j_{k-1}(z)}(z)\in\{1,\dots,k\}$, that is, $j_k(z)$ is the

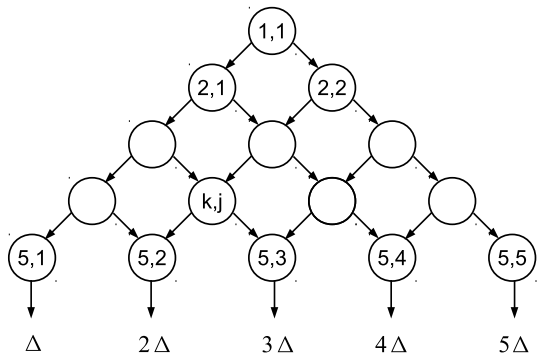

FIG. 2. *A graphical representation of the binary search estimator.*

index of the test to be performed on level $k \le N-2$ and $j_{N-1}(z)\Delta$ is the value of the estimator $\hat{\theta}_n^{(\Delta)}(z) = (u_{N-1,j_{N-1}(z)} + l_{N-1,j_{N-1}(z)})/2 = j_{N-1}(z)\Delta$.

We now analyze the estimation error of $\hat{\theta}_n^{(\Delta)}$. Fix $\mathbb{P} \in \mathcal{P}$ and $z \in \mathcal{Z}^n$. We say that the test $\xi^*_{k,j_k(z)}(z)$ decided incorrectly, if its decision lead to the removal of a length-Δ subinterval that actually contained $\theta(\mathbb{P})$. Formally, $\xi^*_{k,j_k(z)}(z)$ decided incorrectly if $\xi^*_{k,j_k(z)}(z) = 0$ and $\theta(\mathbb{P}) \in [b_{k,j_k(z)}, u_{k,j_k(z)})$, or $\xi^*_{k,j_k(z)}(z) = 1$ and $\theta(\mathbb{P}) \in [l_{k,j_k(z)}, a_{k,j_k(z)})$. Note that the test $\xi^*_{k,j_k(z)}(z)$ can not decide incorrectly if $\theta(\mathbb{P}) \notin [l_{k,j_k(z)}, u_{k,j_k(z)})$. If, for some $l \in \{0,\dots,N-3\}$, all the tests $\xi^*_{k,j_k(z)}(z)$, for $k = 1,\dots,N-2-l$, decide correctly, then $\theta(\mathbb{P}) \in [l_{N-1-l,j_{N-1-l}(z)}, u_{N-1-l,j_{N-1-l}(z)})]$. Since, by construction, we have $\hat{\theta}_n(z) \in [a_{N-1-l,j_{N-1-l}(z)}, b_{N-1-l,j_{N-1-l}(z)})]$ and the latter interval has length $d_{N-1-l} = l\Delta$, this means that $|\hat{\theta}_n^{(\Delta)}(z) - \theta(\mathbb{P})| \le (l+1)\Delta = \eta_l$. Therefore, if $|\hat{\theta}_n^{(\Delta)}(z) - \theta(\mathbb{P})| > \eta_l$, then there exists $k \in \{1,\dots,N-2-l\}$, so that $\xi^*_{k,j_k(z)}(z)$ decided incorrectly. If $\xi^*_{k,j_k(z)}(z)$ incorrectly decided for H_0, then $\theta(\mathbb{P}) \in [b_{k,j_k(z)}, u_{k,j_k(z)})$. By disjointness of $[b_{k,j}, u_{k,j})$, $j = 1,\dots,k$, there is at most one index $j_k^* = j_k^*(\mathbb{P}) \in \{1,\dots,k\}$, so that $\theta(\mathbb{P}) \in [b_{k,j_k^*}, u_{k,j_k^*})$. Thus, $j_k(z) = j_k^*$, $\xi^*_{k,j_k^*}(z) = 0$ and $\theta(\mathbb{P}) \in [b_{k,j_k^*}, u_{k,j_k^*})$. If, on the other hand, $\xi^*_{k,j_k(z)}(z)$ incorrectly decided for H_1, then $\theta(\mathbb{P}) \in [l_{k,j_k(z)}, a_{k,j_k(z)})$. But analogously there is at most one index $j_k^{**} = j_k^{**}(\mathbb{P}) \in \{1,\dots,k\}$, such that $\theta(\mathbb{P}) \in [l_{k,j_k^{**}}, a_{k,j_k^{**}})$. Thus, $\xi^*_{k,j_k^{**}}(z) = 1$ and $\theta(\mathbb{P}) \in [l_{k,j_k^{**}}, a_{k,j_k^{**}})$. This fact, that at any level k there are at most two tests that can decide incorrectly, is the crucial point of our construction. Consequently, for $l = 0,\dots,N-3$,

$$
\begin{aligned}
&Q\mathbb{P}^{\otimes n}\bigl(z \in \mathcal{Z}^n : |\hat{\theta}_n^{(\Delta)}(z) - \theta(\mathbb{P})| > \eta_l\bigr)\\
&\quad\le \sum_{k=1}^{N-2-l} Q\mathbb{P}^{\otimes n}\bigl(\xi^*_{k,j_k(z)}(z) \text{ decides incorrectly}\bigr)\\
&\quad\le \sum_{k=1}^{N-2-l} \bigl[Q\mathbb{P}^{\otimes n}\bigl(\xi^*_{k,j_k^*}(z) = 0, \theta(\mathbb{P}) \in [b_{k,j_k^*}, u_{k,j_k^*})\bigr)\\
&\qquad + Q\mathbb{P}^{\otimes n}\bigl(\xi^*_{k,j_k^{**}}(z) = 1, \theta(\mathbb{P}) \in [l_{k,j_k^{**}}, a_{k,j_k^{**}})\bigr)\bigr].
\end{aligned}
$$

But both

$$
Q\mathbb{P}^{\otimes n}\bigl(\xi^*_{k,j_k^*}(z) = 0, \theta(\mathbb{P}) \in [b_{k,j_k^*}, u_{k,j_k^*})\bigr)
$$

and

$$
Q\mathbb{P}^{\otimes n}\bigl(\xi^*_{k,j_k^{**}}(z) = 1, \theta(\mathbb{P}) \in [l_{k,j_k^{**}}, a_{k,j_k^{**}})\bigr)
$$

are bounded by the worst case risk of the respective test and thus, in view of (7.1), they are both bounded by $2\eta_A^{(n)}(Q, d_k) \vee 0$. We conclude that

$$
\begin{aligned}
Q\mathbb{P}^{\otimes n}\bigl(|\hat{\theta}_n^{(\Delta)} - \theta(\mathbb{P})| > \eta_l\bigr) &\le 4 \sum_{k=1}^{N-2-l} \bigl[\eta_A^{(n)}(Q, d_k) \vee 0\bigr]\\
&= 4 \sum_{k=l+1}^{N-2} \bigl[\eta_A^{(n)}(Q, k\Delta) \vee 0\bigr].
\end{aligned}
$$

If $l > N-3$, then the event $|\hat{\theta}_n^{(\Delta)} - \theta(\mathbb{P})| > \eta_l$ is impossible, because the range of $\hat{\theta}_n^{(\Delta)}$ is $\{j\Delta : j = 1,\dots,N-1\}$.

For part (ii), we further investigate the binary search estimator in the case $Q_1 = Q_1^{(\alpha,\ell^*)}$, as in (4.1). In particular, we have $\mathcal{Z} = \{-z_0, z_0\}$. If $N \le 3$, then any estimator taking values

in $[0, M]$ is at most 2Δ away from $\hat{\theta}_n^{(\Delta)}$. We continue with $N \geq 4$. The marginal data generating distribution $Q_1^{(\alpha,\ell^*)}\mathbb{P}$ is actually a binary distribution on $\{-z_0, z_0\}$, taking the value $z_0 = \|\ell^*\|_\infty \frac{e^\alpha+1}{e^\alpha-1}$ with probability $p(\mathbb{P}) := \frac{1}{2}(1 + \frac{\mathbb{E}_\mathbb{P}[\ell^*]}{z_0}) \in [\frac{1}{2}(1 - \frac{e^\alpha-1}{e^\alpha+1}), \frac{1}{2}(1 + \frac{e^\alpha-1}{e^\alpha+1})]$. The corresponding likelihood is given by

$$q(z_1, \ldots, z_n; p) := \prod_{i=1}^n p^{\frac{1}{2}(1+\frac{z_i}{z_0})}(1-p)^{\frac{1}{2}(1-\frac{z_i}{z_0})} = p^{nT(z)}(1-p)^{n(1-T(z)},$$

where $T(z) = \frac{1}{2}(1 + \frac{\bar{z}_n}{z_0})$ with $\bar{z}_n = \frac{1}{n}\sum_{i=1}^n z_i$. For $0 < a \leq b < M$, define $\bar{\mathcal{P}}_{\leq a} := \{p(\mathbb{P}) : \theta(\mathbb{P}) \leq a\}$, $\bar{\mathcal{P}}_{\geq b} := \{p(\mathbb{P}) : \theta(\mathbb{P}) \geq b\}$, $\bar{a} := \sup \bar{\mathcal{P}}_{\leq a}$ and $\bar{b} := \inf \bar{\mathcal{P}}_{\geq b}$. Clearly, $p_- := \inf_{\mathbb{P}\in\mathcal{P}} p(\mathbb{P}) \leq \bar{a}$ and $\bar{b} \leq \sup_{\mathbb{P}\in\mathcal{P}} p(\mathbb{P}) =: p_+$. Note that for any $\mathbb{P}_0, \mathbb{P}_1 \in \mathcal{P}$, with $\theta(\mathbb{P}_0) - \theta(\mathbb{P}_1) \geq \Delta$, we have

$$p(\mathbb{P}_0) - p(\mathbb{P}_1) = \big(\mathbb{E}_{\mathbb{P}_0}[\ell^*] - \mathbb{E}_{\mathbb{P}_1}[\ell^*]\big)/(2z_0) \geq \inf_{\sigma\in\mathbb{S}_\Delta} \int_\mathcal{X} \ell^* \, d\sigma/(2z_0) > 0 \tag{7.2}$$

by assumption. In particular, if $b - a \geq \Delta$, then $\bar{a} < \bar{b}$. Hence, a minimax test ξ for testing $H_0 : [Q_1\mathcal{P}_{\leq a}]^{(n)} \cong \bar{\mathcal{P}}_{\leq a}^{(n)}$ against $H_1 : Q_1\mathcal{P}_{\geq b} \cong \bar{\mathcal{P}}_{\geq b}$ based on an i.i.d. sample of size n is given by

$$\xi_{a,b}(z_1, \ldots, z_n) = \begin{cases} 1 & \text{if } \dfrac{q(z_1, \ldots, z_n; \bar{b})}{q(z_1, \ldots, z_n; \bar{a})} \geq 1, \\ 0 & \text{else.} \end{cases}$$

The test $\xi_{a,b}(z)$ decides for H_1 (i.e., $\xi_{a,b}(z) = 1$) iff $T(z) \geq G(\bar{a}, \bar{b})$, where

$$G(s,t) := \frac{\log(\frac{1-s}{1-t})}{\log(\frac{t}{s}\frac{1-s}{1-t})}, \quad \text{if } 0 < s < t < 1$$

and $G(s,s) := s$. In the following, we make repeated use of the facts that G is strictly increasing in both arguments and that $s < G(s,t) < t$ for $0 < s < t < 1$ (see Lemma H.5).

Abbreviate $a_j := a_{N-2,j}$ and $b_j := b_{N-2,j}$, and set $a_{N-1} = (N-1)\Delta$ and $b_0 = \Delta$. Next, define the critical values for the tests $\xi_{k,j} := \xi_{a_{k,j},b_{k,j}}$ by $c_{k,j} := G(\bar{a}_{k,j}, \bar{b}_{k,j})$, and abbreviate $c_j := c_{N-2,j}$. Since $\bar{a}_{k,j} \leq \bar{a}_{k,j+1}$ and $\bar{b}_{k,j} \leq \bar{b}_{k,j+1}$, this defines a partition of $[0,1]$ (note that range$(T) \subseteq [0,1]$), that is, $C_1 := [0, c_1)$, $C_{N-1} := [c_{N-2}, 1]$ and $C_j := [c_{j-1}, c_j)$, for $j = 2, \ldots, N-2$, where we interpret $[c,c) = \varnothing$. We now show that $T(z) \in C_j$ implies that $\hat{\theta}_n^{(\Delta)}(z) = j\Delta$. If $T(z) \in C_1$, then all the tests along the binary search path must have decided for H_0, because $c_1 = G(\bar{a}_1, \bar{b}_1)$ is the smallest critical value among all critical values $c_{k,j}$. Thus, $\hat{\theta}_n^{(\Delta)}(z) = \Delta$. For $j \in \{2, \ldots, N-2\}$, suppose that $T(z) \in C_j$. At some level k_0 along the binary search path, either $a_{k_0,j_{k_0}(z)} = j\Delta = a_j$ or $b_{k_0,j_{k_0}(z)} = j\Delta = b_{j-1}$ occurs first. In the former case all tests at levels $k \geq k_0$ must decide for H_0, because $T(z) < c_j$ and c_j is the smallest critical value of all tests with $k \geq k_0$ and $a_{k,j} = j\Delta$. In the latter case all tests at levels $k \geq k_0$ must decide for H_1, because $T(z) \geq c_{j-1}$ and c_{j-1} is the largest critical value among all tests for which $k \geq k_0$ and $b_{k,l} = j\Delta$. Thus, in either case, $\hat{\theta}_n^{(\Delta)}(z) = j\Delta$. Finally, if $T(z) \in C_{N-1}$, all tests must decide for H_1 and $\hat{\theta}_n^{(\Delta)}(z) = (N-1)\Delta$.

Next, by convexity of $\mathcal{P}$ and linearity of $p : \mathcal{P} \to [0,1]$, we see that the range $p(\mathcal{P}) \subseteq [p_-, p_+] \subseteq [0,1]$ of p is an interval. Moreover, we have $p_- < c_1 \leq c_{N-2} < p_+$, because $\bar{a}_1 < \bar{b}_1$, $\bar{a}_{N-2} < \bar{b}_{N-2}$ and the properties of G. In this paragraph we investigate the correspondence between the two functionals $\theta : \mathcal{P} \to [0, M]$ and $p : \mathcal{P} \to [p_-, p_+]$. For

$t \in (p_-, p_+)$, define $\varphi(t) := \sup\{\theta(\mathbb{P}) : p(\mathbb{P}) = t, \mathbb{P} \in \mathcal{P}\}$. If $t \le p_-$, set $\varphi(t) := 0$, and if $t \ge p_+$, set $\varphi(t) := M$. For $p_- < t < c_1$, we see that $0 \le \varphi(t) \le 2\Delta$ because

$$\begin{aligned}\varphi(t) &\le \sup\bigl\{\theta(\mathbb{P}) : p(\mathbb{P}) \le c_1, \mathbb{P} \in \mathcal{P}\bigr\} \le \sup\bigl\{\theta(\mathbb{P}) : p(\mathbb{P}) < \bar{b}_1, \mathbb{P} \in \mathcal{P}\bigr\} \\ &\le b_1 = 2\Delta.\end{aligned}$$

If $t \in C_j$, for some $j \in \{2, \dots, N-2\}$, then

$$\begin{aligned}\varphi(t) &\le \sup\bigl\{\theta(\mathbb{P}) : p(\mathbb{P}) \le c_j, \mathbb{P} \in \mathcal{P}\bigr\} \le \sup\bigl\{\theta(\mathbb{P}) : p(\mathbb{P}) < \bar{b}_j, \mathbb{P} \in \mathcal{P}\bigr\} \\ &\le b_j = (j+1)\Delta,\end{aligned}$$

and

$$\begin{aligned}\varphi(t) &\ge \inf\bigl\{\theta(\mathbb{P}) : p(\mathbb{P}) \ge c_{j-1}, \mathbb{P} \in \mathcal{P}\bigr\} \ge \inf\bigl\{\theta(\mathbb{P}) : p(\mathbb{P}) > \bar{a}_{j-1}, \mathbb{P} \in \mathcal{P}\bigr\} \\ &\ge a_{j-1} = (j-1)\Delta.\end{aligned}$$

Finally, for $c_{N-1} \le t < p_+$, one obtains $(N-2)\Delta \le \varphi(t) \le M$. Thus, we have defined φ on all of $[0, 1]$ in such a way that $|\varphi(T(z)) - \hat{\theta}_n^{(\Delta)}(z)| \le \Delta$.

Next, we show that φ can be approximated on (p_-, p_+) by an affine function. First, note that for $t \in (p_-, p_+)$,

$$\begin{aligned}E(t) &:= \sup\bigl\{\theta(\mathbb{P}) : p(\mathbb{P}) = t, \mathbb{P} \in \mathcal{P}\bigr\} - \inf\bigl\{\theta(\mathbb{P}) : p(\mathbb{P}) = t, \mathbb{P} \in \mathcal{P}\bigr\} \\ &\le \sup\bigl\{\theta(\mathbb{P}_0) - \theta(\mathbb{P}_1) : p(\mathbb{P}_0) - p(\mathbb{P}_1) = 0, \mathbb{P}_0, \mathbb{P}_1 \in \mathcal{P}\bigr\}.\end{aligned}$$

But if $\theta(\mathbb{P}_0) - \theta(\mathbb{P}_1) \ge \Delta$, then $p(\mathbb{P}_0) - p(\mathbb{P}_1) > 0$ (cf. (7.2)). Thus, $E(t) \le \Delta$. Now, for $\lambda \in [0, 1]$ and $s, t \in (p_-, p_+)$, choose $\mathbb{P}_0, \mathbb{P}_1 \in \mathcal{P}$ such that $p(\mathbb{P}_0) = s$ and $p(\mathbb{P}_1) = t$, set $\bar{\mathbb{P}} = \lambda\mathbb{P}_0 + (1-\lambda)\mathbb{P}_1 \in \mathcal{P}$, by convexity, and note that

$$\begin{aligned}\varphi\bigl(\lambda s + (1-\lambda)t\bigr) &\le \inf\bigl\{\theta(\mathbb{P}) : p(\mathbb{P}) = \lambda s + (1-\lambda)t, \mathbb{P} \in \mathcal{P}\bigr\} + \Delta \\ &\le \theta(\bar{\mathbb{P}}) + \Delta = \lambda\theta(\mathbb{P}_0) + (1-\lambda)\theta(\mathbb{P}_1) + \Delta \\ &\le \lambda\varphi(s) + (1-\lambda)\varphi(t) + \Delta,\end{aligned}$$

where we have used linearity of θ and the previously derived bound on E. Similarly, we obtain

$$\begin{aligned}\varphi\bigl(\lambda s + (1-\lambda)t\bigr) &\ge \theta(\bar{\mathbb{P}}) = \lambda\theta(\mathbb{P}_0) + (1-\lambda)\theta(\mathbb{P}_1) \\ &\ge \lambda\varphi(s) + (1-\lambda)\varphi(t) - \Delta.\end{aligned}$$

We conclude that $|\varphi(\lambda s + (1-\lambda)t) - [\lambda\varphi(s) + (1-\lambda)\varphi(t)]| \le \Delta$. Now, fix $s_0 \in (p_-, c_1)$ and $t_0 \in (c_{N-1}, p_+)$ and, for $t \in \mathbb{R}$, define

$$\psi(t) := \frac{t_0 - t}{t_0 - s_0}\varphi(s_0) + \frac{t - s_0}{t_0 - s_0}\varphi(t_0).$$

Thus, for $t \in [s_0, t_0]$ and $\lambda_t := (t_0 - t)/(t_0 - s_0) \in [0, 1]$, we have

$$\begin{aligned}\bigl|\psi(t) - \varphi(t)\bigr| &= \bigl|\psi\bigl(\lambda_t s_0 + (1-\lambda_t)t_0\bigr) - \varphi\bigl(\lambda_t s_0 + (1-\lambda_t)t_0\bigr)\bigr| \\ &\le \bigl|\lambda_t\psi(s_0) + (1-\lambda_t)\psi(t_0) - \bigl[\lambda_t\varphi(s_0) + (1-\lambda_t)\varphi(t_0)\bigr]\bigr| + \Delta \\ &= \Delta.\end{aligned}$$

We have therefore found an affine function ψ, such that for $T(z) \in [c_1, c_{N-1}]$, we have $|\psi(T(z)) - \hat{\theta}_n^{(\Delta)}(z)| \le 2\Delta$. Recall that by construction of φ, we have $\varphi(s_0) \le 2\Delta \le (N-2)\Delta \le \varphi(t_0)$, because $N \ge 4$. Thus, ψ is nondecreasing. If $T(z) < c_1$, then $\hat{\theta}_n^{(\Delta)}(z) = \Delta$ and $0 \le \Pi[\psi(T(z))] \le \Pi[\varphi(T(z)) + \Delta] \le \Pi[3\Delta] = 3\Delta$, where $\Pi : \mathbb{R} \to [0, M]$ is the projection onto $[0, M]$. Thus, $|\Pi[\psi(T(z))] - \hat{\theta}_n^{(\Delta)}(z)| \le 2\Delta$. An analogous argument shows that the same bound holds if $T(z) > c_{N-1}$. Since $T(z)$ is affine in $\bar{z}_n$, the claim follows. $\square$

Acknowledgements. We want to thank an anonymous referee for raising the interesting conjecture that in the private case all linear functionals can be estimated by simple sample averages of appropriately privatized observations. We were able to confirm this conjecture after a substantial revision of an earlier version of the paper.

This work was supported by the DFG Research Grant RO 3766/4-1.

SUPPLEMENTARY MATERIAL

Supplement to "Geometrizing rates of convergence under local differential privacy constraints" (DOI: 10.1214/19-AOS1901SUPP; .pdf). The supplementary material contains most of the proofs, additional technical lemmas and some more technical discussions.

REFERENCES

AWAN, J. and SLAVKOVIĆ, A. (2018). Differentially private uniformly most powerful tests for binomial data. In *Advances in Neural Information Processing Systems* 4212–4222.

DINUR, I. and NISSIM, K. (2003). Revealing information while preserving privacy. In *Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems* 202–210. ACM, New York.

DONOHO, D. L. and LIU, R. C. (1991). Geometrizing rates of convergence. II. *Ann. Statist.* **19** 633–667. MR1105839 https://doi.org/10.1214/aos/1176348114

DUCHI, J. C., JORDAN, M. I. and WAINWRIGHT, M. J. (2013a). Local privacy and statistical minimax rates. In 2013 *IEEE 54th Annual Symposium on Foundations of Computer Science—FOCS* 2013 429–438. IEEE Computer Soc., Los Alamitos, CA. MR3246246 https://doi.org/10.1109/FOCS.2013.53

DUCHI, J., JORDAN, M. I. and WAINWRIGHT, M. J. (2013b). Local privacy and minimax bounds: Sharp rates for probability estimation. In *Adv. Neural Inf. Process. Syst.* 1529–1537.

DUCHI, J. C., JORDAN, M. I. and WAINWRIGHT, M. J. (2014). Local privacy, data processing inequalities, and statistical minimax rates. Preprint. Available at arXiv:1302.3203.

DUCHI, J. C., JORDAN, M. I. and WAINWRIGHT, M. J. (2018). Minimax optimal procedures for locally private estimation. *J. Amer. Statist. Assoc.* **113** 182–201. MR3803452 https://doi.org/10.1080/01621459.2017.1389735

DUCHI, J. and ROGERS, R. (2019). Lower bounds for locally private estimation via communication complexity. Preprint. Available at arXiv:1902.00582.

DWORK, C. (2008). Differential privacy: A survey of results. In *Theory and Applications of Models of Computation. Lecture Notes in Computer Science* **4978** 1–19. Springer, Berlin. MR2472670 https://doi.org/10.1007/978-3-540-79228-4_1

DWORK, C. and NISSIM, K. (2004). Privacy-preserving datamining on vertically partitioned databases. In *Advances in Cryptology—CRYPTO* 2004. *Lecture Notes in Computer Science* **3152** 528–544. Springer, Berlin. MR2147523 https://doi.org/10.1007/978-3-540-28628-8_32

DWORK, C. and SMITH, A. (2010). Differential privacy for statistics: What we know and what we want to learn. *J. Priv. Confid.* **1** 2.

DWORK, C., MCSHERRY, F., NISSIM, K. and SMITH, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography. Lecture Notes in Computer Science* **3876** 265–284. Springer, Berlin. MR2241676 https://doi.org/10.1007/11681878_14

EVFIMIEVSKI, A., GEHRKE, J. and SRIKANT, R. (2003). Limiting privacy breaches in privacy preserving data mining. In *Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems* 211–222. ACM, New York.

GENG, Q. and VISWANATH, P. (2016). The optimal noise-adding mechanism in differential privacy. *IEEE Trans. Inform. Theory* **62** 925–951. MR3455907 https://doi.org/10.1109/TIT.2015.2504967

KRAFFT, O. and WITTING, H. (1967). Optimale Tests und ungünstigste Verteilungen. *Z. Wahrsch. Verw. Gebiete* **7** 289–302. MR0217929 https://doi.org/10.1007/BF01844447

KRAFT, C. (1955). Some conditions for consistency and uniform consistency of statistical procedures. *Univ. California Publ. Statist.* **2** 125–141. MR0073896

LECAM, L. (1973). Convergence of estimates under dimensionality restrictions. *Ann. Statist.* **1** 38–53. MR0334381

LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*, 3rd ed. *Springer Texts in Statistics*. Springer, New York. MR2135927

ROHDE, A. and STEINBERGER, L. (2020). Supplement to "Geometrizing rates of convergence under local differential privacy constraints." https://doi.org/10.1214/19-AOS1901SUPP.

SION, M. (1958). On general minimax theorems. *Pacific J. Math*. **8** 171–176. MR0097026

SMITH, A. (2008). Efficient, differentially private point estimators. Preprint. Available at arXiv:0809.4794.

SMITH, A. (2011). Privacy-preserving statistical estimation with optimal convergence rates [extended abstract]. In *STOC'*11*—Proceedings of the* 43*rd ACM Symposium on Theory of Computing* 813–821. ACM, New York. MR2932032 https://doi.org/10.1145/1993636.1993743

TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer, New York. MR2724359 https://doi.org/10.1007/b13794

WARNER, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *J. Amer. Statist. Assoc*. **60** 63–69.

WASSERMAN, L. and ZHOU, S. (2010). A statistical framework for differential privacy. *J. Amer. Statist. Assoc*. **105** 375–389. MR2656057 https://doi.org/10.1198/jasa.2009.tm08651

YE, M. and BARG, A. (2017). Asymptotically optimal private estimation under mean square loss. Preprint. Available at arXiv:1708.00059.