Privacy-Preserving Parametric Inference: A Case for Robust Statistics

Marco Avella-Medina

Department of Statistics, Columbia University, New York, NY

ABSTRACT

Differential privacy is a cryptographically motivated approach to privacy that has become a very active field of research over the last decade in theoretical computer science and machine learning. In this paradigm, one assumes there is a trusted curator who holds the data of individuals in a database and the goal of privacy is to simultaneously protect individual data while allowing the release of global characteristics of the database. In this setting, we introduce a general framework for parametric inference with differential privacy guarantees. We first obtain differentially private estimators based on bounded influence M-estimators by leveraging their gross-error sensitivity in the calibration of a noise term added to them to ensure privacy. We then show how a similar construction can also be applied to construct differentially private test statistics analogous to the Wald, score, and likelihood ratio tests. We provide statistical guarantees for all our proposals via an asymptotic analysis. An interesting consequence of our results is to further clarify the connection between differential privacy and robust statistics. In particular, we demonstrate that differential privacy is a weaker stability requirement than infinitesimal robustness, and show that robust M-estimators can be easily randomized to guarantee both differential privacy and robustness toward the presence of contaminated data. We illustrate our results both on simulated and real data. Supplementary materials for this article are available online.

1. Introduction

Differential privacy is a cryptographically motivated approach to privacy which has become a very active field of research over the last decade in theoretical computer science and machine learning (Dwork and Roth 2014). In this paradigm, one assumes there is a trusted curator who holds the data of individuals in a database that might for instance be constituted by n individual rows. The goal of privacy is to simultaneously protect every individual row while releasing global characteristics of the database. Differential privacy provides such guarantees in the context of remote access query systems where the data analysts do not get to see the actual data, but can ask a server for the output of some statistical model. Here the trusted curator processes the queries of the user and releases noisy versions of the desired output to protect individual level data.

The interest in remote access systems was prompted by the recognition of fundamental failures of anonymization approaches. Indeed, it is now well acknowledged that releasing datasets without obvious individual identifiers such as names and home addresses are not sufficient to preserve privacy. The problem with such approaches is that an ill-intentioned user might be able to link the anonymized data with external non-anonymized data. Hence auxiliary information could help intruders break anonymization and learn sensitive information. One prominent example of privacy breach is the de-anonymization of a Massachusetts hospital discharge database by joining it with a public voter database in Sweeney (1997).

In fact combining anonymization with sanitization techniques such as adding noise to the dataset directly or removing certain entries of the data matrix are also fundamentally flawed (Narayanan and Smatikov 2008). On the other hand, differential privacy provides a rigorous mathematical framework to the notion of privacy by guaranteeing protection against identity attacks regardless of the auxiliary information that may be available to the attackers. This is achieved by requiring that the output of a query does not change too much if we add or remove any individual from the dataset. Therefore, the user cannot learn much about any individual data record from the output requested.

There is now a large body of literature in this topic and recent work has sought to link differential privacy to statistical problems by developing privacy-preserving algorithms for empirical risk minimization, point estimation and density estimation (Dwork and Lei 2009; Wasserman and Zhou 2010; Chaudhuri, Monteleoni, and Sarwate 2011; Smith 2011; Bassily, Smith, and Thakurta 2014). Despite the numerous developments made in the area of differential privacy since the seminal work of Dwork et al. (2006), one can argue that their practical utility in applied scientific work is very limited by the lack of broad guidelines for statistical inference. In particular, there are no generic procedures for performing statistical hypothesis testing for general parametric models which arguably constitutes one of the cornerstones of a statisticians data analysis toolbox.
1.1. Our Contribution

The basic idea of our work is to introduce differentially private algorithms leveraging tools from robust statistics. In particular, we use the Gaussian mechanism studied in the differential privacy literature in combination with robust statistics sensitivity measures. At a high level, this mechanism provides a generic way to release a noisy version of a statistical query, where the noise level is carefully calibrated to ensure privacy. For this purpose, appropriate notions of sensitivity have been studied in the computer science literature. By focusing on the class of parametric models, we show that the well-studied statistics notion of sensitivity given by the influence function can also be used to calibrate the Gaussian mechanism. This logic extends to tests derived from M-estimators since their sensitivity can also be understood via the influence function.

To the best of our knowledge, our work is the first one to provide a systematic treatment of estimation and hypothesis testing with differential privacy guarantees in the context of general parametric models. The main contributions of this article are the following:

(a) We introduce a general class of differentially private parametric estimators under mild conditions. Our estimators are computationally efficient and can be tuned to trade-off statistical efficiency and robustness.

(b) We propose differentially private counterparts of the Wald, score and likelihood ratio tests for parametric models. Our proposals are by construction robust in a contamination neighborhood of the assumed generative model and are easily constructed from readily available statistics.

(c) We further clarify the connections between differential privacy and robust statistics by showing that the influence function can be used to bound the smooth sensitivity of Nissim, Rashkodnikova, and Smith (2007). It follows that bounded-influence estimators can naturally be used to construct differentially private estimators. The converse is not true as our analysis shows that one can construct differentially private estimators that asymptotically do not have a bounded influence function.

1.2. Related Work

The notion of differential privacy is very similar to the intuitive one of robustness in statistics. The latter requires that no small portion of the data should influence too much a statistical analysis (Hampel et al. 1986; Belisle, Kuh, and Welsch 2005; Maronna, Martin, and Yohai 2006; Huber and Ronchetti 2009). This connection has been noticed in previous works that have shown how to construct differentially private robust estimators. In particular, the estimators of Dwork and Lei (2009), Smith (2011), Lei (2011), and Chaudhuri and Hsu (2012) are the most closely related to ours since they all provide differentially private parametric estimators building on M-estimators and establish statistical convergence rates. However, our construction compares favorably to previous proposals in many regards. Our estimators preserve the optimal parametric \( \sqrt{n} \)-consistency, and hence our privacy guarantees do not come at the expense of slower statistical rates of convergence as in Dwork and Lei (2009) and Lei (2011). Furthermore we do not assume a known diameter of the parameter space as in Smith (2011). Our construction is inspired by the univariate estimator of Chaudhuri and Hsu (2012) which is in general computationally inefficient as it requires the computation of the smooth sensitivity defined in Section 2.2. We broaden the scope of their technique to general multivariate M-estimators and more importantly, we overcome the computational barrier intrinsic to their method by showing that the empirical influence function can be used in the noise calibration of the Gaussian mechanism. There are however other possible approaches to construct differentially private estimators. Here we discuss three popular alternatives that have been explored in the literature.

The first approach seeks to design a mechanism to release differentially private data instead of constructing new estimators. This can be achieved by constructing a differentially private density estimator such as a perturbed histogram of the data. Once such a density estimator is available it can be used to either sample private data (Wasserman and Zhou 2010) or to construct a weighted differentially private objective function for empirical risk minimization (Lei 2011). Although the latter approach leads to better rates of convergence for parametric estimation, they remain slow and have a bad dimension dependence max\( \{1/\sqrt{n}, (\log n/n)^{2/(2+p)}\} \), where \( n \) is the sample size and \( p \) is the dimension of the estimated parameter. Indeed, this approach suffers from the curse of dimensionality since it relies on the computation of multivariate density estimators. Interestingly, a somehow related approach for releasing synthetic data existed in the statistics literature prior to the advent of differential privacy (Rubin 1993; Reiter 2002, 2005) and consequently also lacks formal theoretical privacy guarantees.

A second approach consists of releasing estimators that are defined as the minimizers of a perturbed objective function. Representative work in this direction includes Chaudhuri and Monteleoni (2008) in the context of penalized logistic regression, Chaudhuri, Monteleoni, and Sarwate (2011) in the general learning problem of empirical risk minimization and Kiefer, Smith, and Thakurta (2012) in a high-dimensional regression setting. A related idea to perturbing the objective function is to use a stochastic gradient descent algorithm where at each iteration update step an appropriately scaled noise term is added to the gradient to ensure privacy. This idea was used for example by Rajkumar and Argawal (2012) in the context of multiparty classification, Bassily, Smith, and Thakurta (2014) in the general learning setting of empirical risk minimization and Wang, Fienberg, and Smola (2015) for Bayesian learning. Although the potential applicability of these two perturbation approaches to a wide variety of models makes them appealing, it remains unclear how to construct test statistics in these settings.

A third alternative approach is to draw samples from a well suited probability distribution. The exponential mechanism of McSherry and Talwar (2007) is a main example of a general method for achieving \((\varepsilon, 0)\)-differential privacy via random sampling. This idea leads naturally to connections with posterior sampling in Bayesian statistics. Some articles exploring these ideas include Chaudhuri and Hsu (2012) and Dimitrakakis et al. (2014, 2017). See also Foulds et al. (2016) for a broader discussion of different mechanism for constructing...
privacy preserving Bayesian methods. Bayesian approaches that provide differentially private posterior distributions seem to be naturally amenable for the construction of confidence intervals and test statistics, as explored in Liu (2016). However, it does not seem obvious to us how to use Bayesian privacy preserving results such as Dimitrakakis et al. (2014, 2017) and Foulds et al. (2016) to provide analogue constructions to ours for estimation and testing. Interestingly, in this line of work the typical regularity conditions required on the likelihood and prior distribution are reminiscent of the regularity conditions required in frequentist setups as discussed below in Section 3.1.

The literature on hypothesis testing with differential privacy guarantees is much more recent and limited than the one focusing on estimation. A few articles tackling this problem are the work of (Uhler, Slavkovic, and Fienberg 2013; Wang, Lee, and Kifer 2015; Gaboardi et al. 2016) who consider differentially private chi-squared tests and (Sheffet 2017; Barrientos et al. 2015; Gaboardi et al. 2016) who consider differentially private inference in the work of Awan and Slavkovic (2018, 2019). After submitting the first version of this article, we have noticed some interesting new developments on differentially private hypothesis testing using M-estimators. In Section 5, we show how to further extend our construction to test functions to perform differentially private hypothesis testing using M-estimators. In Section 5, we illustrate the numerical performance of our methods in both synthetic and real data. We conclude our article in Section 6 with a discussion of our results and future research directions.

We relegated to the Appendix all the proofs and some auxiliary results and discussions.

Notation: \( \|V\| \) denotes either Euclidean norm if \( V \in \mathbb{R}^n \) or its induced operator norm if \( V \in \mathbb{R}^{n \times n} \). The smallest and largest eigenvalues of a matrix \( A \) are denoted by \( \lambda_{\min}(A) \) and \( \lambda_{\max}(A) \). For two probability measures \( P \) and \( Q \), the notation \( d_{\infty}(P, Q) \) and \( d_{TV}(P, Q) \) stand for sup-norm (Kolmogorov–Smirnov) and total variation distance. We reserve calligraphic letters such as \( S \) for sets and denote their cardinality by \( |S| \). For two sets of \( S \) and \( S' \) of the same size, we denote their Hamming distance by \( d_H(S, S') := |S \setminus S'| = |S' \setminus S| \).

2. Preliminaries

Let us first review some important background concepts from differential privacy, robust statistics and the M-estimation framework for parametric models.

2.1. Differential Privacy

Consider a database consisting of a set of data points \( D = \{x_1, \ldots, x_n\} \subset \mathbb{R}^n \), where \( \mathbb{R} \subset \mathbb{R}^m \) is some data space. We also use the notation \( D(F_n) \) to emphasize that \( D \) can be viewed as a dataset associated with an empirical distribution \( F_n \) induced by \( \{x_1, \ldots, x_n\} \). Differential privacy seeks to release useful information from the dataset while protecting information about any individual data entry.

Definition 1. A randomized function \( A(D) \) is \((\varepsilon, \delta)\)-differentially private if for all pairs of databases \((D, D')\) with \( d_H(D, D') = 1 \) and all measurable subsets of outputs \( O \):

\[
\Pr(A(D) \in O) \leq e^\varepsilon \Pr(A(D') \in O) + \delta.
\]

Intuitively, \((\varepsilon, 0)\)-differential privacy ensures that for every run of algorithm \( A \) the output is almost equally likely to be observed on every neighboring database. This condition is relaxed by \((\varepsilon, \delta)\)-differential privacy since it allows that given a random output \( O \) drawn from \( A(D) \), it may be possible to find a database \( D' \) such that \( O \) is more likely to be produced on \( D' \) than in the database is \( D \). However, such an event will be extremely unlikely. In both cases the similarity is defined by the factor \( e^\varepsilon \) while the probability of deviating from this similarity is \( \delta \).

The magnitude of the privacy parameters \((\varepsilon, \delta)\) are typically considered to be quite different. We are particularly interested in negligible values of \( \delta \) that are smaller than the inverse of any polynomial in the size \( n \) of the database. The rational behind this requirement is that values of \( \delta \) of the order of \( \|x\|_1 \), for some vector values database \( x \), are problematic since they “preserve privacy” while allowing to publish the complete records of a small number of individuals in the database. On the other hand, the privacy parameter \( \varepsilon \) is typically thought of as a moderately small constant and in fact “the nature of privacy guarantees with differing but small epsilons are quite similar” (Dwork and Roth 2014, p. 25). Indeed, failing to be \((\varepsilon, 0)\)-differentially private for some large \( \varepsilon \) (i.e., \( \varepsilon = 10 \)) is just saying that there is at least a pair of neighboring datasets and an output \( O \) for which the ratio of probabilities of observing \( O \) conditioned on the database being \( D \) or \( D' \) is large.

1.3. Organization of the Article

In Section 2, we overview some key background notions from differential privacy and robust statistics that we use throughout the article. In Section 3, we introduce our technique for constructing differentially private estimators and study their theoretical properties. In Section 4, we show how to further extend our construction to test functions to perform differentially private hypothesis testing using M-estimators. In Section 5, we illustrate the numerical performance of our methods in both synthetic and real data. We conclude our article in Section 6 with a discussion of our results and future research directions.
One can naturally wonder how to compare two differentially private algorithms $A_1$ and $A_2$ with different associated privacy parameters $(\epsilon_1, \delta_1)$ and $(\epsilon_2, \delta_2)$. It seems natural to prefer the algorithm that ensures the smallest privacy loss incurred by observing some output, that is, log algorithm that ensures the smallest privacy loss incurred by observing some output, that is, log

\begin{equation}
\text{Algorithm that ensures the smallest privacy loss incurred by observing some output.}
\end{equation}

Since we only consider negligible $\delta_1$ and $\delta_2$, the privacy loss will be approximately proportional to the privacy parameter $\epsilon$. One could consequently prefer the algorithm with the smallest parameter $\epsilon$ even though we say that roughly speaking "all small epsilons are alike" (Dwork and Roth 2014, p. 24).

Differential privacy enjoys certain appealing properties that facilitate the design and analysis of complicated algorithms with privacy guarantees. Perhaps the two most important ones are that $(\epsilon, \delta)$-differential privacy is immune to post-processing and that combining two differentially private algorithms preserves differential privacy. More precisely, if $A$ is $(\epsilon, \delta)$-differentially private, then the composition of any data independent mapping $f$ with $A$ is also $(\epsilon, \delta)$-differentially private. In other words, releasing $f(A(D))$ for any $D$ still guarantees $(\epsilon, \delta)$-differential privacy. Furthermore, if we have two algorithms $A_1$ and $A_2$ with different associated privacy parameters $(\epsilon_1, \delta_1)$ and $(\epsilon_2, \delta_2)$, then releasing the outputs of $A_1(D)$ and $A_2(D)$ guarantees $(\epsilon_1 + \epsilon_2, \delta_1 + \delta_2)$-differential privacy. We refer interested readers to Dwork and Roth (2014, chaps. 2–3) for a more extensive discussion of the concepts presented in this subsection.

### 2.2. Constructing Differentially Private Algorithms

A general and very popular technique for constructing differentially private algorithms is the Laplace mechanism, which consists of adding some well calibrated noise to the output of a standard query (Dwork et al. 2006). This procedure relies on suitable notions of sensitivity of the function that is queried. All the following definitions of sensitivity are standard in the differential privacy literature and are typically defined with respect to the $L_1$ norm. We will instead use the Euclidean norm for the construction of our estimators as explained below.

**Definition 2.** The global sensitivity of a function $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^p$ is

\begin{equation}
\text{GS} (\varphi) := \sup_{D, D'} \left\{ \| \varphi(D) - \varphi(D') \| : d_H(D, D') = 1 \right\}.
\end{equation}

The local sensitivity of a function $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^p$ at a dataset $D \in \mathbb{R}^n$ is

\begin{equation}
\text{LS} (\varphi, D) := \sup_{D'} \left\{ \| \varphi(D) - \varphi(D') \| : d_H(D, D') = 1 \right\}.
\end{equation}

For $\xi > 0$, the $\xi$-smooth sensitivity of $\varphi$ at $D$ is

\begin{equation}
\text{SS}_{\xi} (\varphi, D) := \sup_{D'} e^{-\xi d_H(D, D')} \text{LS} (\varphi, D') \in \mathbb{R}^p.
\end{equation}

We are now ready to describe two versions of the Laplace mechanism using the above sensitivity notions defined with respect to the $L_1$ norm. Denote by Lap$(b)$ a scaled symmetric Laplace distribution with density function $h_b(x) = \frac{1}{2b} \exp(-|x|/b)$ and let Lap$_p(b)$ be the multivariate distribution obtained from $p$ independent and identically distributed $X_j \sim \text{Lap}(b)$ for $j = 1, \ldots, p$. A key idea introduced in the seminal article Dwork et al. (2006) is that for a function $f : \mathbb{R}^n \rightarrow \mathbb{R}^p$ and an input database $D$, one can simply compute $f(D)$ and then generate an independent noise term $U \sim \text{Lap}_p(\text{GS}(f)/\epsilon)$ to construct a $(\epsilon, 0)$-differentially private output $f(D) + U$. A related idea introduced by Nissim, Rashkodnikova, and Smith (2007) is to calibrate the noise using the smooth sensitivity instead of the local sensitivity. These authors showed that provided $\xi = \frac{\sqrt{2log(2/\delta)} + \epsilon}{4(p + 2log(2/\delta))}$ and $\hat{U} \sim \text{Lap}_p(\text{SS}_{\xi}(f)/\epsilon)$, then the output $f(D) + \hat{U}$ is $(\epsilon, \delta)$-differentially private. Our proposals will build on the latter idea for the construction of private estimation and inferential procedures for parametric models.

We would like to point out that the different notions of sensitivity introduced in Definition 2 are usually defined with respect to the $L_1$ norm. We chose to instead present these definitions in terms of the Euclidean distance as they are more naturally connected to well-studied concepts in robust statistics. In particular, it leads to connections with the standard way of presenting the notion of gross-error sensitivity in robust statistics and the related problem of optimal B-robust estimation (Hampel et al. 1986, chap. 4). Because we focus on sensitivities with respect to the Euclidean metric, our construction follows the same logic of the Laplace mechanism, but naturally replaces the noise distribution with an appropriately scaled normal random variable as proposed in Nissim, Rashkodnikova, and Smith (2007). In this case, the output $f(D) + \hat{U}$ is $(\epsilon, \delta)$-differentially private if $\hat{U} \sim N_p(0, \sigma^2I)$ where $\sigma = 5\sqrt{2log(2/\delta)}\text{SS}_{\xi}(f)/\epsilon$ and $\xi = \frac{\sqrt{2log(2/\delta)} + \epsilon}{4(p + 2log(2/\delta))}$. For obvious reasons the resulting procedure has been called the Gaussian mechanism in Dwork and Roth (2014).

As we were completing the revision of the current manuscript we noticed that Cai, Wang, and Zhang (2019) have also worked with this mechanism for the derivation of the optimal statistical minimax rates of convergence for parametric estimation under $(\epsilon, \delta)$-differential privacy.

### 2.3. Robust Statistics

Robust statistics provides a theoretical framework that allows to take into account that models are only idealized approximations of reality and develops methods that give results that are stable when slight deviations from the stochastic assumptions of the model occur. Book-length expositions on the topic include (Huber 1981; Hampel et al. 1986; Maronna, Martin, and Yohai 2006; Huber and Ronchetti 2009). We will focus on the infinitesimal robustness approach that considers the impact of moderate distributional deviations from ideal models on a statistical procedure (Hampel et al. 1986). In this setting the statistics of interest are viewed as functionals of the underlying distribution and the influence function is the key tool used to assess the robustness of a statistical functional.

**Definition 3.** Given a measurable space $\mathfrak{X}$, a distribution space $\mathfrak{F}$, a parameter space $\Theta \subset \mathbb{R}^p$, and a functional $T : \mathfrak{X} \rightarrow \Theta$, the influence function of $T$ at a point $z \in \mathfrak{X}$ for a distribution $F$ is defined as

\begin{equation}
\text{IF}(z; T, F) := \lim_{t \rightarrow 0+} \frac{T(F_t) - T(F)}{t},
\end{equation}

where $F_t = (1 - t)F + t \Delta_z$ and $\Delta_z$ is a mass point at $z$. 
The influence function has the heuristic interpretation of describing the effect of an infinitesimal contamination at the point \( z \) on the estimate, standardized by the mass of contamination. Furthermore, if a statistical functional \( T(F) \) is sufficiently regular, a von Mises expansion (von Mises 1947; Hampel 1974; Hampel et al. 1986) yields

\[
T(G) = T(F) + \int \text{IF}(z; T, F) d(G - F)(z) + o(d_{\infty}(G, F)).
\]

(1)

Considering the approximation \( (1) \) over a neighborhood of the form \( \delta_G = \{p^{(1)} | p^{(1)} = (1-t) F + t G, G \text{ an arbitrary distribution} \} \), we see that the influence function can be used to linearize the asymptotic bias in a neighborhood of the idealized model \( F \). Therefore, a statistical functional with bounded influence function is robust in the sense that it will have a bounded approximate bias in a neighborhood of \( F \). A related notion of robustness is the gross-error sensitivity which measures the worst case value of the influence function.

**Definition 4.** The gross-error sensitivity of a functional \( T : \Theta \rightarrow \mathbb{X} \) at the distribution \( F \) is

\[
\gamma(T, F) := \sup_{x \in \mathbb{X}} \| \text{IF}(x; T, F) \|.
\]

Clearly if the space \( \mathbb{X} \) is unbounded, the gross-error sensitivity of \( T \) will be infinite unless its influence function is uniformly bounded. In Sections 3 and 4, we will show how to use the robust statistics tools described here in the construction of differentially private estimators and tests.

### 2.4. M-Estimators for Parametric Models

M-estimators are a simple class of estimators that is appealing from a robust statistics perspective and constitute a very general approach to parametric inference (Huber 1964; Huber and Ronchetti 2009). They will be the focus of the rest of this article. An M-estimator \( \hat{\theta} = T(F_n) \) of \( \theta_0 \in \Theta \subset \mathbb{R}^p \) is defined as a solution to

\[
\sum_{i=1}^{n} \Psi(x_i, T(F_n)) = 0,
\]

where \( \Psi : \mathbb{R}^m \times \Theta \rightarrow \mathbb{R}^p \) is independent identically distributed according to \( F \) and \( F_n \) denotes the empirical distribution function. This class of estimators is a strict generalization of the class of regular maximum likelihood estimators. Assuming that \( T(F) = \theta_0 \) and some mild conditions (Huber and Ronchetti 2009, chap. 6), as \( n \rightarrow \infty \) they are asymptotically normally distributed as

\[
\sqrt{n}(T(F_n) - \theta_0) \xrightarrow{d} N(0, V(T, F)),
\]

where \( V(T, F) = \mathbb{E}_F[\text{IF}(X; T, F)\text{IF}(X; T, F)^\top] \) and \( \mathbb{E}_F[\text{IF}(X; T, F)] = 0 \). Furthermore, their influence function is

\[
\text{IF}(x; T, F) = \left( M(T, F) \right)^{-1} \Psi(x, T(F)),
\]

(2)

where \( M(T, F) = -\mathbb{E}_F[\text{IF}(X, T(F))] = -\frac{d}{d\theta} \mathbb{E}_F[\Psi(X, \theta)] \big|_{\theta = \theta_0} \). Therefore, M-estimators defined by bounded functions \( \Psi \) are said to be infinitesimally robust since their influence function is bounded and by (1) their asymptotic bias will also be bounded for small amounts of contamination.

### 3. Differentially Private Estimation

#### 3.1. Assumptions

In the following we allow \( \Psi \) to depend on \( n \), but we do not stress it in the notation to make it less cumbersome. Here are the main conditions required in our analysis:

**Condition 1.** The function \( \Psi(x, \theta) \) is differentiable with respect to \( \theta \) almost everywhere for all \( x \in \mathbb{X} \), and we denote its derivative by \( \dot{\Psi}(x, \theta) \). Furthermore, for all \( \theta \in \Theta \) there exists constants \( K_n, L_n > 0 \) such that

\[
\sup_{x \in \mathbb{X}} \| \dot{\Psi}(x, \theta) \| \leq K_n \quad \text{and} \quad \sup_{x \in \mathbb{X}} \| \dot{\Psi}(x, \theta) \| \leq L_n.
\]

**Condition 2.** The matrix \( M_F = M(T, F) = -\mathbb{E}_F[\dot{\Psi}(X, T(F))] \) is positive definite at the generative distribution \( F = F_{\theta_0} \). Furthermore, the space of datasets \( \mathbb{X}^n \) is such that for all empirical distributions \( G_n \in \{G | D(G) \in \mathbb{X}^n \} \) with \( n \geq N_0 \) we have that \( 0 < b \leq \lambda_{\min}(M_{G_n}) \leq \lambda_{\max}(M_{G_n}) \leq B < \infty \).

**Condition 3.** There exist \( r_1 > 0, r_2 > 0, r_3 > 0, C_1 \) and \( C_2 > 0 \) such that

\[
\| \mathbb{E}_{G_n}[\dot{\Psi}(X, \theta)] - \mathbb{E}_{G_n}[\dot{\Psi}(X, \theta)] \| \leq C_1 d_{\infty}(F_n, G_n) \quad \text{and} \quad \| \mathbb{E}_{G_n}[\dot{\Psi}(X, \theta)] - \mathbb{E}_{F_n}[\dot{\Psi}(X, T(F_n))] \| \leq C_2 \| T(F_n) - \theta \|
\]

whenever \( d_{\infty}(F_n, G_n) \leq r_1 \), \( \| \theta - T(F_n) \| \leq r_2 \) and \( \| \theta - \theta_0 \| \leq r_3 \).

Condition 1 requires \( \Psi \) and \( \dot{\Psi} \) to be uniformly bounded in \( \mathbb{X} \) by some potentially diverging constants \( K_n \) and \( L_n \). The case \( K_n = K < \infty \) is particularly appealing from a robust statistics perspective as it guarantees that the resulting M-estimators has a bounded influence function. If additionally \( L_n = L < \infty \), then the resulting M-estimator will also be second order infinitesimally robust as defined in La Vecchia, Ronchetti, and Trojani (2012) and will have a bounded change of variance function; see Hampel, Ronchetti, and Ronchetti (1981) and our Appendix C in the supplementary materials for more details. Condition 2 restricts the space of datasets to one where some minimal regularity conditions on the Jacobian of \( \Psi \) hold. Similar assumptions are usually required to guarantee the asymptotic normality and Fréchet differentiability of M-estimators (see, e.g., Huber 1967; Clarke 1986; Huber and Ronchetti 2009, Corollary 6.7). Our assumptions are stronger to guarantee that \( M_{G_n} \) is invertible and hence that the empirical influence function is computable. Even though such requirements are not always explicitly stated, common statistical practice implicitly assumes them when computing estimated asymptotic variances with plug-in formulas. In a standard linear regression setting these conditions boil down to assuming that the design matrix is full rank. Even such a seemingly harmless condition seems stronger in the differential privacy context. Indeed, it might not be checkable by the users and one would like to have such a guarantee to hold over all possible configurations of the data. One possible way of tackling this problem is to let the algorithm halt with an output “No Reply” when this assumption fails (Dwork and Lei 2009; Avella-Medina and Brunel 2019). Condition 3 is a smoothness condition on \( \Psi \) at \( F_n \), similar to Condition 4 in Chaudhuri and Hsu (2012). It is
a technical assumption used when upper bounding the smooth sensitivity by the gross-error sensitivity. The constants $C_1$ and $C_2$ are effectively Lipschitz constants.

We would like to highlight that since the differential privacy paradigm assumes a remote access query framework where the user does not get to see the data, in principle it is not immediate that the user will be able to check basic features of the data, for example, whether the design matrix is full rank before performing an analysis. This is a serious limitation of this paradigm as it more generally prevents users from performing exploratory data analysis before fitting a model and it is also unclear how to do model checking and run diagnostics on fitted models. One would have to develop differentially private analogues of the whole data analysis pipeline to allow a data analyst to perform rigorous statistical analysis. An interesting recent development in this direction in a regression setting is the work of Chen et al. (2018).

### 3.2. A General Construction

Let us now introduce our mechanism for constructing differentially private M-estimators. Given a statistical M-functional $T$, we propose the randomized estimator

$$A_T(F_n) := T(F_n) + γ(T, F_n) 5 \sqrt{2 \log(n) \log(2/δ)} Z,$$

where $Z$ is a $p$ dimensional standard normal random variable. The intuition behind our proposal is simple: the gross-error sensitivity $γ(T, F_n)$ should be roughly of the same order as the smooth sensitivity. Therefore, multiplying it by $\sqrt{\log(n)}$ will guarantee that it upper bounds the smooth sensitivity. This in turn suffices to guarantee $(ε, δ)$-differential privacy. From a computational perspective, using the empirical gross-error-sensitivity is much more appealing than computing the exact smooth sensitivity. Indeed, the former can be further upper bounded in practice using the empirical influence function whereas the latter can be very difficult to compute in general as discussed in Nissim, Rashkodnikova, and Smith (2007).

**Theorem 1.** Let $n ≥ max[N_0, b^{C_4}m^{\log(2/δ)}[1 + 4/p + 2 \log(2/δ)] \log(\lambda_{max}(M_{F_n}))/b^2]$, $(C'C)^2 \log(2/δ)[\frac{2}{b} + \frac{1}{\lambda_{max}(M_{F_n})}(C_1 + C_2 K_{\sigma}/b^2)]^2$ and assume that Conditions 1–3 hold. Then hen $A_T$ is $(ε, δ)$-differentially private.

Theorem 1 shows that our proposal leads to differentially private estimation. It builds on two lemmas, relegated to the Appendix, that show that the smooth sensitivity of $T$ can indeed be upper bounded by twice its empirical gross error sensitivity. Note that the minimum sample size requirement depends on the values of $[N_0, b, K_{\sigma}, L_{\sigma}, C_1, C_2]$ defined in Conditions 1–3, as well as some constants $C$ and $C'$ resulting from our bounds on the error incurred by approximating the smooth sensitivity with the empirical gross-error-sensitivity. We provide a discussion about the evaluation of these constants in the Appendix.

### 3.3. Examples

Let us now present three important examples to illustrate how one can use readily available robust M-estimators and their influence functions to derive bounds on their empirical gross-error sensitivities. These quantities can in turn be used to release differentially private estimates $A_T(F_n)$ defined in (3).

#### 3.3.1. Example 1: Location-Scale Model

We consider the location-scale model discussed in Huber and Ronchetti (2009, chap. 6). Here we observe an iid random sample of univariate random variables $X_1, \ldots, X_n$ with density function of the form $\frac{1}{\sigma} f(\frac{x - μ}{σ})$, where $f$ is some known density function, $μ$ is an unknown location parameter and $σ$ is an unknown positive scale parameter. The problem of simultaneous location and scale parameter estimation is motivated by invariance considerations. In particular, to make an $M$-estimate of location scale invariant, we must couple it with an estimate of scale. If the underlying distribution $F$ is symmetric, location estimates $T$, and scale estimates $S$ typically are asymptotically independent, and the asymptotic behavior of $T$ depends on $S$ only through the asymptotic value $S(F)$. We can therefore afford to choose $S$ on criteria other than low statistical variability. Huber (1964) generalized the maximum likelihood system of equations by considering simultaneous M-estimates of location and scale any pair of statistics $(T_n, S_n)$ determined by two equations of the form

$$\sum_{i=1}^{n} \psi \left( \frac{x_i - T_n}{S_n} \right) = 0 \quad \text{and} \quad \sum_{i=1}^{n} \chi \left( \frac{x_i - T_n}{S_n} \right) = 0,$$

which lead $T_n = T(F_n)$ and $S_n = S(F_n)$ to be expressed in terms of functionals $T$ and $S$ defined by the population equations

$$\int \psi \left( \frac{x - T(F)}{S(F)} \right) dF(x) = 0 \quad \text{and} \quad \int \chi \left( \frac{x - T(F)}{S(F)} \right) dF(x) = 0.$$

From the latter equations one can show that, if $ψ$ is odd and $χ$ is even, the influence functions of $T$ and $S$ are

$$IF(x; T, F) = \frac{ψ \left( \frac{x - T(F)}{S(F)} \right) S(F)}{∫ ψ \left( \frac{x - T(F)}{S(F)} \right) dF(x)}$$

and

$$IF(x; S, F) = \frac{χ \left( \frac{x - T(F)}{S(F)} \right) S(F)}{∫ χ \left( \frac{x - T(F)}{S(F)} \right) dF(x)}.$$

The problem of robust joint estimation of location and scale was introduced in the seminal article of Huber (1964). In the important case of the normal model, where $F = Φ$ is the standard normal distribution, a prominent example of the above system of equations is Huber’s proposal 2. In this case, $ψ(τ) = ψ_ε(τ) = \text{min}(c, \text{max}(-c, τ))$ is the Huber function and $χ(τ) = χ_ε(τ) = ψ_ε(τ)^2 - κ$, where $κ = ∫ \text{min}(c^2, x^2) dΦ(x)$ is a constant that ensures Fisher consistency at the normal model, that is, $T(Φ) = μ$ and $S(Φ) = σ^2$. This particular choice of estimating equations and (4) show that the empirical gross-error sensitivities of $μ = T_n = T(F_n)$ and $σ = S_n = S(F_n)$ are

$$γ(T, F_n) = \frac{C_0}{n} \sum_{i=1}^{n} \left| \frac{x_i - \hat{μ}}{\hat{σ}} \right| < c$$

and

$$γ(S, F_n) = \frac{C_0}{n} \sum_{i=1}^{n} \left( \frac{x_i - \hat{μ}}{\hat{σ}} \right)^2 \left| \frac{x_i - \hat{μ}}{\hat{σ}} \right| < c,$$

where $\hat{μ}$ and $\hat{σ}$ are the empirical mean and standard deviation of the sample.
where the last equation used that \( \chi_i'(r) = \psi_i'(r)r \) almost everywhere and \( I_E \) is the indicator function taking the value 1 under the event \( E \) and is 0 otherwise. The formulas obtained in (5) can be used in the Gaussian mechanism (3) for obtaining private location and scale estimates. We refer the reader Huber and Ronchetti (2009, chap. 6) for more discussion and details on joint robust estimation of location and scale parameters.

3.3.2. Example 2: Linear Regression

One can naturally build on the construction of the previous example to obtain robust estimators for the linear regression model

\[ y_i = x_i^\top \beta + u_i, \quad \text{for } i = 1, \ldots, n, \tag{6} \]

where \( y_i \) is the response variable, \( x_i \in \mathbb{R}^p \) the covariates and the noise terms are \( u_i \overset{iid}{\sim} N(0, \sigma^2) \). The estimator discussed here is a Mallows’ type robust M-estimator defined as

\[ \hat{\beta}, \hat{\sigma} = \arg\min_{\beta, \sigma} \left\{ \sum_{i=1}^n \rho_i \left( \frac{y_i - x_i^\top \beta}{\sigma} \right) w(x_i) + \kappa n \sigma \right\}, \tag{7} \]

where \( \rho_i \) is the Huber loss function with tuning parameter \( c \), \( \kappa = \min \{ c^2, r^2 \} d\Phi(r) \) is a Fisher consistency constant for \( \sigma \) and \( w : \mathbb{R}^p \to \mathbb{R}_{\geq 0} \) is a downweighting function that controls the impact of outlying covariates on the estimators of \( \hat{\beta} \) and \( \hat{\sigma} \) (Hampel et al. 1986). This robust estimator uses Huber’s Proposal 2 for the estimation of the scale parameter. The influence function of the estimator \( \hat{\beta} = T(F_n) \) is

\[ IF(x, y, T, F) = M_F^{-1} f \left[ \frac{y - x^\top T(F)}{S(F)} \right] w(x), \]

where \( M_F = \int xx^\top w(x) \psi(r) dF \) and \( r = \frac{x^\top T(F)}{S(F)} \). Therefore, \( M_{F_n} = \frac{1}{n} \sum_{i=1}^n x_i x_i^\top w(x_i) I_{|y_i| \leq \hat{r}} \) with \( \hat{r} = (y_i - x_i^\top \hat{\beta})/\hat{\sigma} \), and assuming that \( \sup_n \| xw(x) \| \leq \bar{K} \), we see that \( \gamma(T, F_n) \leq \lambda_{\min}(M_{F_n})^{-1} \bar{K} \). This last bound can be used for the release of a differentially private estimates of \( \beta \). Note also that using the derivations from Example 1 we also have that the empirical gross-error sensitivity of \( \hat{\sigma} = S(F) \) is \( \gamma(S, F_n) = \frac{1}{n} \sum_{i=1}^n \hat{r}^2 I_{|y_i| \leq \hat{r}} \leq (c^2 - \kappa) \bar{K} \).

3.3.3. Example 3: Generalized Linear Models

Generalized linear models (McCullagh and Nelder 1989) assume that conditional on some covariates, the response variables belong to the exponential family, that is, the response variables \( Y_1, \ldots, Y_n \) are drawn independently from the densities of the form

\[ f(y_i; \theta_i) = \exp \left[ y_i \theta_i - b(\theta_i) \right] / \phi + c(y_i, \phi), \]

where \( a(\cdot), b(\cdot), \) and \( c(\cdot) \) are specific functions and \( \phi \) a nuisance parameter. Thus, \( E(Y_i) = \mu_i = b'(\theta_i) \) and \( \operatorname{var}(Y_i) = \psi(\mu_i) = \phi b''(\theta_i) \) and \( g(\mu_i) = \eta_i = x_i^\top \beta \), where \( \beta \in \mathbb{R}^p \) is the vector of parameters, \( x_i \in \mathbb{R}^p \) is the set of explanatory variables and \( g(\cdot) \) the link function.

Cantoni and Ronchetti (2001) proposed a class of M-estimators for GLM which can be viewed as a natural robustification of the quasi-likelihood estimators of Wedderburn (1974). Their robust quasilikelihood is

\[ \rho_n(\beta) = \frac{1}{n} \sum_{i=1}^n Q_M(y_i, x_i^\top \beta), \]

where the functions \( Q_M(y_i, x_i^\top \beta) \) can be written as

\[ Q_M(y_i, x_i^\top \beta) = \int_{y_i}^{\mu_i} v(y, t) w(x_i) dt - \frac{1}{n} \sum_{i=1}^n \int_{y_i}^{\mu_i} \left[ E[v(y, t)] w(x_i) dt \right] \]

with \( v(y, t) = \psi(y) - \psi(y - s)/\sqrt{\psi''(s)} \), \( s \) such that \( \psi(y - s)/\sqrt{\psi''(s)} = 0 \) and \( t \) such that \( E[\psi(y - s)/\sqrt{\psi''(s)}] = 0 \). The function \( \psi(\cdot) \) is bounded and protects against large outliers in the responses, and \( w(\cdot) \) downweights leverage points. The estimator of \( \beta \) derived from the minimization of this loss function is the solution of the estimating equation

\[ \psi^{(n)}(\beta) = \frac{1}{n} \sum_{i=1}^n \psi(y_i, x_i^\top \beta) = \frac{1}{n} \sum_{i=1}^n \left( \psi(t_i) - \frac{1}{\sqrt{\psi''(t_i)}} \frac{\partial \mu_i}{\partial \beta} \right) = 0, \tag{8} \]

where \( t_i = (y_i - \mu_i)/\sqrt{\psi''(\mu_i)} \) and \( a(\beta) = n^{-1} \sum_{i=1}^n E[\psi(t_i)/\sqrt{\psi''(\mu_i)}] w(x_i) \partial \mu_i / \partial \beta \) ensures Fisher consistency and can be computed using the formulas in Appendix A of Cantoni and Ronchetti (2001). We note that Appendix B of the same article show that \( M_F \) is of the form \( \frac{1}{n} \sum_{i=1}^n w(x_i) \psi(t_i) \) and that these estimators and formulas are implemented in the function \( \text{glmrob} \) of the \( \text{R} \) package \( \text{robustbase} \). They can be used to bound the empirical gross-error sensitivity with \( \gamma(T, F_n) \leq \lambda_{\min}(M_{F_n})^{-1} K_n \) where \( K_n \) is as in Condition 1 and will be depend on the choices of \( \psi \) and \( w \) as was the case in Example 2.

3.4. Convergence Rates

We provide upper bounds for the convergence rates of \( A_T(F_n) \). Our result is an extension of Theorem 1 in Chaudhuri and Hsu (2012).

**Theorem 2.** Suppose Conditions 1–2 hold. Then, for \( \tau \in (0, 1) \), with probability at least \( 1 - \tau \)

\[ \| A_T(F_n) - T(F) \| \leq \| T(F_n) - T(F) \| \]

\[ + C \sqrt{\log(n) \log(2/\delta) K_n(\sqrt{\rho} + \sqrt{\log(1/\tau)})} \]

for some positive constant \( C \). If in addition \( K_n \sqrt{m \log(n) \log(1/\delta)} / \sqrt{n} \to 0 \) as \( n \to \infty \), then

\[ A_T(F_n) - T(F) = T(F_n) - T(F) + o_p(1/\sqrt{n}). \]

A direct consequence of the above result and Huber and Ronchetti (2009, Corollary 6.7) is that \( A_T(F_n) \) is asymptotically normally distributed as stated next.
Corollary 1. Assume that \( p \) is fixed and that Conditions 1–2 hold. Further assume that \( \mathbb{E} \frac{||\Psi(X, \theta_0)||^2}{\varepsilon \sqrt{n}} \) is nonzero and finite. If

\[
\frac{\log \log n \log(1/\delta)}{\varepsilon \sqrt{n}} \to 0 \quad \text{as} \quad n \to \infty,
\]

then we have that

\[
\sqrt{n}(A^T(F_n) - T(F)) \to_d N(0, V(T, F)).
\]

**Remark 1.** This asymptotic normality result can be easily extended to the case where \( p \) diverges as \( n \) increases. In particular, invoking the results of He and Shao (2000) asymptotic normality holds assuming \( \frac{\log \log n}{n} \to 0 \). Note also that when \( p \) diverges, \( K_n \) will be diverging even for robust estimators as componentwise boundedness of \( \Psi \) implies that \( K_n = O(\sqrt{p}) \).

### 3.5. Efficiency, Truncation, and Robustness Properties

Smith (2008, 2011) introduced a class of asymptotically efficient point estimators obtained by averaging subsampled estimators and adding well calibrated noise using the Laplace mechanism of Dwork et al. (2006). Unfortunately his construction relies heavily on the assumption that the diameter of the parameter space is known when calibrating the noise added to the output. Furthermore it is also assumed that we observe bounded random variables. Variants of this assumption are common in the differential privacy literature (Lei 2011; Smith 2011; Bassily, Smith, and Thakurta 2014). Our estimators can bypass these issues as long as the divergence level of truncation, but as a tool for achieving optimal nonasymptotic sub-Gaussian-type deviations for mean estimators.

Corollary 2. Let \( T_n \) denote the M-functional defined by the truncated score function \( s_t(x, \theta) = \frac{\partial \log f_0}{\partial \theta} w_c(x), \) where \( w_c(x) = \min\{1, c / ||\frac{\partial \log f_0}{\partial \theta}(x)||\}, \) is some positive constant and \( f_0 \) denotes the density of \( F. \) If \( c \to \infty \) and \( c \frac{\log n}{\sqrt{n}} \to 0 \) as \( n \to \infty, \) then we have that

\[
\sqrt{n}(A^T(F_n) - \theta_0) \to_d N(0, I^{-1}(\theta_0)),
\]

where \( I(\theta_0) \) denotes the Fisher information matrix.

The truncated maximum likelihood construction is reminiscent of the estimator of Catoni (2012). The latter also uses a diverging level of truncation, but as a tool for achieving optimal nonasymptotic sub-Gaussian-type deviations for mean estimators under heavy tailed assumptions.

From a robust statistics point of view a diverging level of truncation is not a fully satisfactory solution. Indeed, it is well known that maximum likelihood estimators can be highly sensitive to the presence of small fractions of contamination in the data. This remains true for the truncated maximum likelihood estimator if the truncation level is allowed to diverge as it entails that the estimator will fail to have a bounded influence function asymptotically and will therefore not be robust in this sense. Interestingly, Chaudhuri and Hsu (2012) showed that any differentially private algorithm needs to satisfy a somehow weaker degree of robustness. Our next Theorem provides a result in the same spirit for multivariate M-estimators.

**Theorem 3.** Let \( \varepsilon \in (0, \frac{\log 2}{3}) \) and \( \delta \in (0, \frac{1}{3^2}). \) Let \( \mathcal{F} \) be the family of all distributions over \( X \subset \mathbb{R}^p \) and let \( A \) be any \((\varepsilon, \delta)\)-differentially private algorithm of \( T(F). \) For all \( n \in \mathbb{N} \) and \( F \in \mathcal{F} \) there exists a radius \( \rho = \rho(n) = \frac{1}{n} \left( \frac{\log 2}{2\varepsilon} \right) \) and a distribution \( G \in \mathcal{F} \) with \( d_{TV}(F, G) \leq \rho, \) such that either

\[
\mathbb{E}_{F_n} \mathbb{E}_A \left[ ||A(D(F_n)) - T(F)|| \right] \geq \frac{\rho}{16} \gamma(T, F) + o(\rho)
\]

or

\[
\mathbb{E}_{G_n} \mathbb{E}_A \left[ ||A(D(G_n)) - T(G)|| \right] \geq \frac{\rho}{16} \gamma(T, F) + o(\rho),
\]

where \( F_n \) and \( G_n \) denote empirical distributions obtained from \( F \) and \( G \), respectively.

### 4. Differentially Private Inference

We now present our core results for privacy-preserving hypothesis testing building on the randomization scheme introduced in the previous section.

#### 4.1. Background

We denote the partition of a \( p \) dimensional vector \( \nu \) into \( p-k \) and \( k \) components by \( \nu = (\nu^{(1)}, \nu^{(2)})^\top. \) We are interested in testing hypothesis of the form

\[
H_0 : \theta = \theta_0, \quad H_1 : \theta \neq \theta_0
\]

and let \( \theta_{(1)} = \theta_{(2)} \neq 0 \) where \( \theta_{(1)} \) is unspecified against the alternative \( H_1 : \theta_{(2)} \neq 0 \) and \( \theta_{(2)} \) is unspecified. Assume throughout that the dimension \( k \) is fixed. A well-known result in statistics states that the Wald, score and likelihood ratio tests are asymptotically optimal and equivalent in the sense that they converge to the uniformly most powerful test (Lehmann and Romano 2006). The level functionals of these test statistics can be approximated by functionals of the form

\[
\alpha(F_n) := 1 - H_k(nU(F_n)^\top U(F_n)),
\]

where \( H_k(\cdot) \) is the cumulative distribution function of a \( \chi^2_k \) random variable, \( U(F_n) \) is a standardized functional such that under the null hypothesis \( U(F) = 0 \) and

\[
\sqrt{n}(U(F_n) - U(F)) \to_d N(0, I_k).
\]

Heritier and Ronchetti (1994) proposed robust tests based on M-estimators. Their main advantage over their classical counterparts is they have bounded level and power influence functions. Therefore, these tests are stable under small arbitrary contamination under both the null hypothesis and the alternative. Following Heritier and Ronchetti (1994), we therefore consider the three classes of tests described next.
1. A Wald-type test statistic is a quadratic statistic of the form
$$W(F_n) := T(F_n)^\top (V(T, F))^{-1} T(F_n).$$
(11)

2. A score (or Rao)-type test statistic has the form
$$R(F_n) := Z(T, F_n)^\top U(T, F)^{-1} Z(T, F_n),$$
(12)
where $Z(T, F_n) = \frac{1}{n} \sum_{i=1}^n \Psi(x_i, T, T_R(F_n))$, $T_R$ is the restricted $M$-functional defined as the solution of
$$\int \Psi(x, T_R(F)) \, dF = 0 \quad \text{and} \quad T_R(F) = 0,$$
$$U(T, F) = M_{(22.1)} V(T, F)_{(22)} M_{(22.1)}^\top,$$
and $M_{(22.1)} = M_{(22)} - M_{(21)} M_{(11)}^{-1} M_{(12)}$ with $M = M(T, F)$.

3. A likelihood ratio-type test has the form
$$S(F_n) := \frac{2}{n} \sum_{i=1}^n \left( \rho(x_i, T(F_n)) - \rho(x_i, T_R(F_n)) \right),$$
(13)
where $\rho(x, 0) = 0$, $\frac{\partial}{\partial \theta} \rho(x, \theta) = \Psi(x, \theta)$ and $T$ and $T_R$ are the $M$-functionals of the full and restricted models, respectively. As showed in Heritier and Ronchetti (1994) the likelihood ratio functional is asymptotically equivalent to the quadratic form $\tilde{S}(F) := U_{LR}(F)^\top U_{LR}(F)$ where $U_{LR}(F) = M_{(22.1)}^{1/2} T(F)^{(2)}$.

Note that in practice the matrices $M(T, F), U(T, F)$, and $V(T, F)$ need to be estimated. We discuss this point in Section 4.6.

### 4.2. Private Inference Based on the Level Gross-Error Sensitivity

We can use any of the robust test statistics described above to provide differential private $p$-values using an analogue construction to the one introduced for estimation in Section 3. Our proposal for differentially private testing is to build $p$-values of the form
$$A_\alpha(F_n) := \alpha(F_n) + \gamma(\alpha; F_n; \varepsilon) \frac{5 \sqrt{2 \log(n) \log(2/\delta)}}{n} Z,$$
where $Z$ is an independent standard normal random variable. The rationale behind our construction is that $\gamma(\alpha; F_n)$ is the right scaling factor for applying the Gaussian mechanism to $\alpha(F_n)$ since it should roughly be of the same order as its smooth sensitivity. Note also that one can use $A_\alpha(F_n)$ to construct randomized counterparts to the test statistics (11), (12), and (13) by simply computing
$$Q(F_n) := H^{-1}_k(A_\alpha(F_n)),$$
that is by evaluating the quantile function of a $\chi^2_k$ at $A_\alpha(F_n)$. Note that we can also apply the Gaussian mechanism to the Wald, score and likelihood ratio type statistics of Section 4.1 and construct differentially private $p$-values from them. Indeed postprocessing preserves differential privacy so computing the induced $p$-values preserves the privacy guarantees (Dwork and Roth 2014, Proposition 2.1). Our theoretical results extend straightforwardly to this alternative approach and the numerical performance is nearly identical to the one presented in this article in experiments. The following theorem establishes the differential privacy guarantee of our proposal.

**Theorem 4.** Let $n \geq \max\{N_0, \frac{1}{C_m \log(2/\delta)} \{1 + \frac{4}{5}(p + 2 \log(2/\delta))\} \log(C_{n,k,U})^3, C_0^2 n \log(1/\delta) \frac{K_k^2}{\max(M_{(n)})} (1 + \frac{4}{5}(p + 2 \log(2/\delta))\} \}$, $C_U$ and $C_{n,k,U}$ are constants depending on the test functional. If Conditions 1–3 hold, then $A_\alpha$ is $(\varepsilon, \delta)$-differentially private.

The minimum sample size required in Theorem 4 is similar to that of Theorem 1. In particular it also depends on the same $\{N_0, b, K_n, L_n, C_1, C_2, C_3\}$ as well as the test specific constants $C_U$ and $C_{n,k,U}$ resulting from our bounds on the error incurred by approximating the smooth sensitivity of the level functionals by their the empirical gross-error sensitivity. A discussion on these constants can be found in the Appendix.

### 4.3. Examples

The following two examples show how to upper bound empirical the level gross-error sensitivity $\gamma(\alpha; F_n)$ required for the construction of our differentially private $p$-values.

#### 4.3.1. Example 4: Testing and Confidence Intervals in Linear Regression

We consider the problem of hypothesis testing in the setting considered in Example 2. We focus on the same Mallow’s estimator in combination with the Wald statistics $W_n = W(F_n)$ defined in (11) for hypothesis testing. We first note that from the chain rule, the influence function of $W$ at the $F$ is
$$IF(x; W, F_n) = 2T(F_n)^\top (V(T, F)_{(22)})^{-1} T(x; F, F_n).$$
It follows that
$$\gamma(W, F_n) \leq 2 \lambda_{\min}(V(T, F)_{(22)})^{-1} \|T(F_n)\| \gamma(T(2), F_n)$$
and the respective level gross-error sensitivity can be bounded as $\gamma(\alpha_W; F_n) \leq n H_k(n W_n) \gamma(W, F_n)$. In the case of univariate null hypothesis of the form $H_0 : \beta_j = 0$ these expressions become
$$IF(x; W, F_n) = 2T(F_n)^\top |V(T, F)_{(22)}|^{-1} |T(F_n)| \text{ and } \gamma(\alpha_W; F_n) \leq 2n H_k(n W_n) |V(T, F)_{(22)}|^{-1} \|M_{(F_n)}\| \text{.}$$
where $(M_{(F_n)})^{-1}_{ik}$ denotes the $i$th row of $M_{(F_n)}^{-1}$. The above bound on $\gamma(\alpha_W; F_n)$ can be used in the Gaussian mechanism suggested in Section 4.2 for reporting differentially private $p$-values $A_{\alpha_W}(F_n)$ accompanying the regression slope estimates $A_T(F_n)$ of Example 2.

We further note that since $(\varepsilon, \delta)$-differential privacy is not affected by post-processing, one can also construct confidence intervals using the reported $p$-value $A_{\alpha_W}(F_n)$. Since the asymptotic distribution of the Wald test is a $\chi^2_k$ for the null hypothesis $H_0 : \beta_j = 0$, a natural way to construct a confidence interval is to map the value $A_{\alpha_W}(F_n)$ to the quantile of $\chi^2_k$ and output the interval defined by its squared root. More precisely, one can first compute $Q_{\alpha; \beta_j} = H^{-1}_k(A_{\alpha_W}(F_n))$ and then report the differentially private confidence interval $(-\sqrt{Q_{\alpha; \beta_j}}, \sqrt{Q_{\alpha; \beta_j}})$. 

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION 977
4.3.2. Example 5: Testing and Confidence Intervals in Logistic Regression

Let us now return to the robust quasi-likelihood estimator discussed in Example 3 and focus on the special case of binary regression with canonical link. Note that if one chooses \( \psi(r) = r \) and \( w(x) = 1 \) in (8), the resulting estimator is equivalent to logistic regression. In general (8) will take the form

\[
\frac{1}{n} \sum_{i=1}^{n} \left| \psi(r_i) \frac{e^{r_i}}{1 + e^{r_i}} w(x_i) x_i - a(\beta) \right| = 0,
\]

where \( r_i = (y_i - p_i)/\sqrt{p_i(1 - p_i)} \) and \( p_i = e^{r_i}/(1 + e^{r_i}) \). In this case, if \( \sup_x \| xw(x) \| \leq \bar{K} \) and \( |\psi(r_i)| \leq c_\psi \), then the gross-error sensitivity of \( \beta = T(F_n) \) can be bounded as \( \gamma(T, F_n) \leq 2\lambda_{\min}(M_F^{-1})^{-1}c_\psi/\bar{K} \). For example, if we consider the weight function \( w(x) = \{1, 1/\|x\|\} \) and the Huber function \( \psi(r) = \psi_\tau(r) \), then \( \bar{K} = 1 \) and \( c_\psi \) is the constant of the Huber function. Note also that Appendix B in Cantoni and Ronchetti (2001) provide formulas for \( M_F \) when \( \psi(r) = \psi_\tau(r) \) and this bound is readily obtained using standard functions in \( R \). Furthermore the computation of the gross-error sensitivity for the level functional of the Wald statistics follows from the same arguments discussed in Example 4. The extension of the proposed construction of confidence intervals is also immediate.

4.4. Validity of the Tests

In this subsection, we establish statistical consistency guarantees for our differentially private tests. The next theorem establishes rates of convergence and demonstrates the asymptotic equivalence between them and their non-private counterparts under both the null distribution and a local alternative.

**Theorem 5.** Assume Conditions 1 and 2 hold and let \( \alpha(\cdot) \) be the level functional of any of the tests (11)–(13). Then, for \( \tau \in (0, 1) \), with probability at least \( 1 - \tau \)

\[
|A_\alpha(F_n) - \alpha(F)| \leq |A_\alpha(F_n) - \alpha(F)| + C \sqrt{\log(n) \log(2/\delta) \log(2/\tau) K_n} \frac{\sqrt{n}}{\sqrt{k\epsilon}}
\]

for some positive constant \( C \). Furthermore, if \( K_n \sqrt{\log(n) \log(1/\delta)} \to 0 \) as \( n \to \infty \) then

\[
Q(F_n) = Q_0(F_n) + o_p(1),
\]

where \( Q_0(F_n) = H^{-1}_k(\alpha(F_n)) \).

A direct consequence of Theorem 5 is that the asymptotic distribution of \( Q(F_n) \) is the same as the one of its nonprivate counterpart \( Q_0(F_n) \) computed from the level functional of any of the tests (11)–(13). Therefore, the results of Heritier and Ronchetti (1994) also give the asymptotic distributions of \( Q(F_n) \) under both \( H_0 : \theta = \theta_0 \) and \( H_{1,n} : \theta = \theta_0 + \Delta/n \) for some \( \Delta > 0 \). In particular, Propositions 1 and 2 that establish that (11) and (12) are asymptotically equivalent as they both converge to \( \chi_k^2 \) under \( H_0 \) and to \( \chi_k^2(\delta) \) with \( \delta = \Delta^T V(T, F_n)^{-1} \Delta \) under \( H_{1,n} \). Proposition 3 of the same article shows that (13) converges instead to a weighted sum of \( k \) independent random variables distributed as \( \chi_k^2 \) under \( H_0 \) and to a weighted sum of \( k \) independent random variables \( \chi_k^2(\delta_i) \) for some \( \delta_1, \ldots, \delta_k > 0 \) under \( H_{1,n} \).

4.5. Robustness Properties of Differentially Private Tests

The tests associated with the differentially private \( p \)-values proposed in Section 4.2 enjoy some degree of robustness by construction. In particular, it is not difficult to extend the lower bound of Theorem 3 to the level functionals considered in this section.

**Theorem 6.** Assume the conditions of Theorem 3, but letting \( A \) be any \((\epsilon, \delta)\)-differentially private algorithm of the level functional \( \alpha(F) \) of either of the tests (11)–(13). Then either

\[
\mathbb{E}_{F_n} \mathbb{E}_A \left[ |A(D(F_n)) - \alpha(F)| \right] \geq \frac{\mu}{16} \left[ \frac{\log 2}{2e} \right] \gamma(U, F)^2 + o \left( \left[ \frac{\log 2}{2e} \right] \right)
\]

or

\[
\mathbb{E}_{G_n} \mathbb{E}_A \left[ |A(D(G_n)) - \alpha(G)| \right] \geq \frac{\mu}{16} \left[ \frac{\log 2}{2e} \right] \gamma(U, F)^2 + o \left( \left[ \frac{\log 2}{2e} \right] \right),
\]

where \( \mu = -\frac{\partial}{\partial \xi} H_k(q_{1-\alpha_0}; \xi) \big|_{\xi=0} \), \( H_k(\cdot, \xi) \) is the cumulative distribution function of a noncentral \( \chi_k^2(\xi) \) with noncentrality parameter \( \xi \geq 0, q_{1-\alpha_0} \) is the \( 1 - \alpha_0 \) quantile of a \( \chi_k^2 \) distribution and \( \alpha_0 = \alpha(F) \) is the nominal level of the test.

Similar to Theorem 3, Theorem 6 states that the convergence rates of any differentially private algorithm \( A \) estimating the level functional \( \alpha \) is lower bounded by the gross-error sensitivity of \( U(F) \) in a small neighborhood of \( F \), where \( U \) is defined in (9) and (10). Therefore, functionals \( U \) with diverging influence functions will lead to slower convergence rates for any algorithm \( A \) in all such neighborhoods. The result suggests that the influence function has to scale at most as \( \rho^{-1} = O(\epsilon \sqrt{n}) \).

Note that the appearance of the quadratic term \( \gamma(U, F)^2 \) in the lower bound is intuitive from the definition of \( \alpha(F) \) and is in line with the robustness characterization of the level influence function of (Heritier and Ronchetti 1994; Ronchetti and Trojani 2001). In fact we can extend the robustness results of these articles to our setting and show that our tests have stable level and power functions in shrinking contamination neighborhoods of the model when \( \Psi \) is bounded.

We need to introduce additional notation to state the result. Consider the \((t, n)\)-contamination neighborhoods of \( F_{0} \) defined by

\[
\Lambda_{t,n}(F_{0}) = \left\{ F_{t,n,G} \right. \left| \left(1 - \frac{t}{\sqrt{n}}\right) F_{0} + \frac{t}{\sqrt{n}} G, G \text{ arbitrary} \right. \}
\]

and let \( U_n = U(F_n) \) be a statistical functional with bounded influence function and such that \( U(F) = 0 \) and

\[
\sqrt{n}(U(F_n) - U(F_{t,n,G})) \to_d N(0, I_k)
\]
uniformly over the sequence of \((t, n)-\)neighborhoods \(\mathcal{U}_{t,n}(F_{0|n})\). Further let
\[
\{F_{\text{alt},n}\}_{n \in \mathbb{N}} := \left\{ \left(1 - \frac{n}{\sqrt{n}}\right) F_{0|n} + \frac{n}{\sqrt{n}} F_{1|n}\right\}_{n \in \mathbb{N}}
\]
be a sequence of local alternatives to \(F_{0|n}\), and
\[
\mathcal{U}_{t,n}(F_{\text{alt},n}) := \left\{ F_{1|n,G} = \left(1 - \frac{t}{\sqrt{n}}\right) F_{\text{alt},n} + \frac{t}{\sqrt{n}} G, G \text{ arbitrary} \right\}
\]
be the corresponding neighborhood of \(F_{\text{alt},n}\) for a given \(n\). We denote by \(\{F_{1|n,G}\}_{n \in \mathbb{N}}\) a sequence of \((t, n, G)\)-contaminations of the underlying null distribution \(F_{0|n}\), each of them belonging to the neighborhood \(\mathcal{U}_{t,n}(F_{0|n})\). Similarly, we denote by \(\{F_{1|n,G}\}_{n \in \mathbb{N}}\) a sequence of \((t, n, G)\)-contaminations of the underlying local alternatives \(F_{\text{alt},n}\), each of them belonging to the neighborhood \(\mathcal{U}_{t,n}(F_{\text{alt},n})\). Finally, we denote by \(A_{\beta}\) and \(A_{\alpha}\) the power functionals of the tests based on \(\hat{\alpha}_{n}\) and \(\hat{\beta}_{n}\), respectively.

The following corollary follows from Ronchetti and Trojani (2001, Theorems 1–3) and Theorem 5. It shows that the level and power of our differentially private tests are stable in the contamination neighborhoods \(\mathcal{U}_{t,n}(F_{0|n})\) and \(\mathcal{U}_{t,n}(F_{\text{alt},n})\) when the influence function of the functional \(U\) is bounded.

**Corollary 3.** Our differentially private Wald, score and likelihood ratio type tests have stable level and power functionals when \(K_{n} < \infty\) in the sense that for all \(G\)
\[
\lim_{n \to \infty} A_{\alpha}(F_{1|n,G}) = \lim_{n \to \infty} \alpha(F_{1|n,G}) = \alpha_0 + t^2 \mu \left\| \int IF(x; U, F_{0|n})dG(x) \right\|^2 + o(t^2)
\]
and
\[
\lim_{n \to \infty} A_{\beta}(F_{1|n,G}) = \lim_{n \to \infty} \beta(F_{1|n,G}) = \beta_0 + 2\mu tn \int IF(x; U, F_{\text{alt},n})dG(x) \\
\times \int IF(x; U, F_{0|n})dG_0(x) + o(\eta),
\]
where \(\mu\) is as in Theorem 6.

### 4.6. Accounting for the Change of Variance Sensitivity

In practice the standardizing matrices \(M(T, F), U(T, F), U(T, F)\), and \(V(T, F)\) are estimated, so the actual form of the functional \(U\) defining the test functional is
\[
U(F_n) = S(F_n)^{-1/2} \hat{T}(F_n),
\]
where \(\hat{T}\) is such that \(\hat{T}(F) = 0\) and \(\sqrt{n} (\hat{T}(F_n) - \tilde{T}(F)) \to_d N(0, S(F))\). The general construction of Section 4.2 is still valid provided additional regularity conditions on \(\Psi\) hold. In particular, it remains true that \(\gamma(\alpha, F)\) can be used to upper bound \(\tilde{T}_n\) provided \(\sup_{\alpha} \|\dot{\Psi}\| < \infty\) for all \(j = 1, \ldots, p\). This condition implies third order infinitesimal robustness in the sense of La Vecchia, Ronchetti, and Trojani (2012). From a practical point of view an upper bound on \(\gamma(\alpha, F)\) can be computed in this case using both the influence function and the change of variance function of \(T\). The latter accounts for the fact the \(S(F)\) is also estimated. We refer the reader to the Appendix for the precise form of the change of variance function of general M-estimators and a more detailed discussion of the implications of estimating the variance in the noise calibration of our Gaussian mechanism.

### 5. Numerical Examples

We investigate the finite sample performance of our proposals with simulated and real data. We focus on a linear regression setting where we obtain consistent slope parameter estimates at the model and show that our differentially private tests reach the desired nominal level and has power under the alternative even in mildly contaminated scenarios. We first present a simulation experiments that shows the statistical performance of our methods in small samples before turning to a real data example with a large sample size. For the sake of space we relegated to the Appendix a more extended discussion about other existing methods, some complementary simulation results and a discussion of the evaluation of the constants of Theorems 1 and 4.

#### 5.1. Synthetic Data

We consider a simulation setting similar to the one of Salibian-Barrera, Van Aelst, and Yohai (2016) to explore the behavior of our consistent differentially private estimates and illustrate the efficiency loss incurred by them, relative to their non private counterparts. We generate the linear regression model (6) with \(\beta = (1, 1, 0, 0)\) and \(x_i \sim N(0, V)\) and \(V = \{0.5^{[j-k]}\}_{j,k=1}^{4}\). We illustrate the effect of small amounts of contaminated data by generating outliers in the responses as well as bad leverage points. This was done by replacing 1% of the values of \(y\) and \(x_2\) with observations following a \(N(12, 0.1^2)\) and a \(N(5, 0.1^2)\) distribution, respectively. All the results reported below were obtained over 5000 replications and samples sizes ranging from \(n = 100\) to \(n = 1000\).

The differentially private estimates considered here is the same Mallow’s type robust regression estimator of Example 2. In particular, we consider the robust estimators of \(\beta\) defined by
\[
(\hat{\beta}_0, \hat{\beta}, \hat{\sigma}) = \arg\min_{\beta, \sigma} \left\{ \sum_{i=1}^{n} \sigma_{\rho}(\{y_i - \hat{\beta}_0 + x_i^{\top} \hat{\beta}\}) w(x_i) + \kappa_i \right\},
\]
where \(\rho_{\sigma}\) is the Huber loss function with tuning parameter \(c\), \(w : \mathbb{R}^p \to \mathbb{R}_{>0}\) is a downweighting function and \(\kappa_i = \int \min(\sigma^2, c^2) d\Phi(x)\) is a constant ensuring that \(\sigma\) is consistent. In all our simulations we set \(c = 1.345\) and \(w(x) = \min(1, 2/\|x\|_2)\). This robust estimator uses Huber’s Proposal 2 for the estimation of the scale parameter (Huber and Ronchetti 2009). We computed it using the function \(z(1)m\) of the \texttt{R} package “MASS.” Figure 1 shows how the level of privacy affects the performance of estimation relative to that of the target robust estimator. In particular, it illustrates the slower convergence of our differentially private estimators for the range of privacy parameters \(\epsilon = \{0.2, 0.1, 0.05\}\) and \(\delta = 1/n^2\). Figure 2 shows the empirical level of the Wald statistics for testing the null hypothesis \(H_0 : \beta_3 = \beta_4 = 0\) with increasing sample sizes.
Figure 1. The plots show the componentwise estimation error of the parameter \( \beta \) for clean datasets ranging from size \( n = 100 \) to \( n = 1000 \). The dotted dark blue line shows the median estimated value of the target robust estimator while the light blue shaded area give pointwise quartiles of the same estimator. The larger shaded gray areas give the pointwise quartiles of the estimated differentially private estimators with privacy parameters \( \epsilon = \{0.2, 0.1, 0.05\} \).

Figure 2. (a) The convergence of our Wald statistic to the nominal level 0.05 at the model while (b) its behavior under 1% contamination. We report four empirical differentially private level curves: dotted lines, \( \epsilon = 1 \); dash-dotted lines, \( \epsilon = 0.1 \); dash-dotted lines, \( \epsilon = 0.01 \), two-dashed lines, \( \epsilon = 0.001 \).

and nominal level of 5%. We see that all the tests have good empirical coverage and that as expected the differentially private tests are not too sensitive to the presence of a small amount of contamination. Interestingly, the empirical levels of the robust test and the differentially private one are nearly identical when the privacy parameters \( \epsilon = \{1, 0.1\} \) and \( n \geq 200 \). When we choose the very stringent \( \epsilon = 0.001 \) the noise added to the target \( p \)-value is so large that the resulting test amounts to flipping a coin.

To explore the power of our tests we set the regression parameter \( \beta \) to \( (1, 1, \nu, 0)^T \), where \( \nu \) varied in the range \([-0.5, 0.5]\). As seen in Figure 3(a) the power function of the three tests considered is almost indistinguishable when the data follows the normal model (6). Figure 3(b) shows that the power functions of the robust Wald tests and the derived differentially private test remain almost identical to the one they have without contamination. This reflects the power function stability result established in Theorem 3. From the same figure, we clearly see that the power function of the Wald test constructed using least squares estimator is shifted as a result of a small amount of contamination.

5.2. Application to Housing Price Data

We revisit the housing price dataset considered in Lei (2011). The data consist of 348,189 houses sold in the San Francisco Bay Area between 2003 and 2006, for which we have the price, size, year of transaction, and county in which the house is located. The dataset has two continuous covariates (price and size), one ordinal variable with 4 levels (year), and one categorical variable (county) with 9 levels. We exclude the observations with missing entries and follow the preprocessing suggested in Lei (2011), that is, we filter out data points with price outside the range of \( 10^5 \sim 9 \times 10^5 \) or with size larger than 3000 squared feet. After preprocessing, we have 250,070 observations and the
Figure 3. (a) The power function of our Wald statistic at the model when $n = 200$ and $\beta_3 \in [-0.5, 0.5]$; (b) its behavior under 1% contamination. We report four empirical differentially private power curves: dotted lines, $\varepsilon = 1$; dash-dotted lines, $\varepsilon = 0.1$; dash-dotted lines, $\varepsilon = 0.01$, two-dashed lines, $\varepsilon = 0.001$.

Table 1. Linear regression coefficients using the Bay housing data after preprocessing.

| Method | OLS      | Rob | PHOLS | PHRob | DP        |
|--------|----------|-----|-------|-------|-----------|
| Intercept | 135,141 | 118,479 | 8.9 | 10.4 | 1.4 \times 10^{-4} |
| Size    | 209      | 216  | 4.0   | 5.1   | 7.3 \times 10^{-2} |
| Year    | 56,375   | 58,136 | 2.6 | 5.2   | 2.8 \times 10^{-4} |
| County 2 | -53,765  | -59,605 | 8.1 | 7.6   | 2.9 \times 10^{-4} |
| County 3 | 146,593  | 149,202 | 2.7 | 3.8   | 1.1 \times 10^{-4} |
| County 4 | -27,546  | -29,681 | 37.7 | 28.4 | 5.2 \times 10^{-4} |
| County 5 | 45,828   | 41,184 | 7.8 | 16.5  | 4.1 \times 10^{-4} |
| County 6 | -140,738 | -139,780 | 3.6 | 7.7   | 1.1 \times 10^{-4} |

NOTE: The second and third columns give the regression coefficients obtained by ordinary least squares and the robust Mallow’s estimator without privacy guarantees. We compare the performance of their differentially private counterparts using the perturbed histogram approach and our Gaussian mechanism for a fixed privacy level $\varepsilon = 0.1$. The reported number is the componentwise root mean square relative error over 1000 realizations.

Table 2. Linear regression coefficients using the raw Bay housing data without preprocessing.

| Method | OLS      | Rob | PHOLS | PHRob | DP        |
|--------|----------|-----|-------|-------|-----------|
| Intercept | 456,344 | 101,524 | 33.4 | 28.6  | 1.5 \times 10^{-4} |
| Size    | 0.5      | 229  | 247.1 | 229.3 | 6.2 \times 10^{-2} |
| Year    | 71,241   | 65,170 | 87.8 | 85.7  | 2.2 \times 10^{-4} |
| County 2 | -11,261  | -53,727 | 416.8 | 376.9 | 2.9 \times 10^{-4} |
| County 3 | 275,058  | 196,967 | 82.4 | 80.7  | 7.5 \times 10^{-5} |
| County 4 | -16,425  | -29,337 | 569.0 | 519.1 | 4.8 \times 10^{-4} |
| County 5 | 98,775   | 57524 | 101.9 | 95.9  | 2.6 \times 10^{-4} |
| County 6 | -149,027 | -152,499 | 143.3 | 141.2 | 9.2 \times 10^{-5} |

NOTE: The reported numbers are as in Table 1.

county variables has 6 levels after combination. We also consider the same data without filtering price and size, in which case we are left with 286,537 observations. We fitted a simple linear regression model to predict the housing price using ordinary least squares, a robust estimator and differentially private estimators. We computed the private estimator described in Section 5.1 as well as the differentially private M-estimators based on a perturbed histogram with enhanced thresholding as in Lei (2011). We assess the performance of the differentially private regression coefficients by comparing them with their nonprivate counterparts. More specifically, we look at the componentwise relative deviance from the nonprivate estimates $d_j = |\hat{\beta}_j^{DP}/\hat{\beta}_j - 1|$ where $\hat{\beta}_j$ stands for the $j$th regression coefficient of either the ordinary least squares or the robust estimator, and $\hat{\beta}_j^{DP}$ is its differentially private counterpart. To account for the randomness of the Gaussian mechanism, we report the mean square error of the deviations $d_j$ obtained over 500 realizations. The results are summarized in Tables 1 and 2.

It is interesting to notice that with the preprocessed data the least squares fit and the robust fit are very similar. However with the raw data, the large unfiltered values of price and size affect to a greater extent the estimator of Lei (2011). The accuracy of this estimator also deteriorates for the raw data as reflected by the larger mean squared deviations obtained in this case.
On the other hand, our differentially private estimators give similar results for both preprocessed and raw data, in terms of values of the fitted regression coefficients and mean squared deviations from the target robust estimates. This is a particularly desirable feature when privacy is an issue since researchers are likely to have limited access to the data and hence carrying out a careful preprocessing might not be possible. Note also that for the same level of privacy $\varepsilon = 10^{-3}$, our method provides much more accurate estimation. The poorer performance of the histogram estimator is to be expected as it suffers from the curse of dimensionality. In this particular example Lei’s estimator effectively reduces the sample size to only 2400 pseudo observations that can be sampled from the differentially private estimated histogram.

We see from the reported values in Tables 1 and 2 that the accuracy of our private estimator is comparable with that of the perturbed histogram if we impose the much stronger privacy requirement $\varepsilon = 10^{-5}$. This feature is also very appealing in practice and confirms what our theory predicts and what we observed in simulations: we can afford a fixed privacy budget with a smaller sample size or equivalently, for a fixed sample size we can ensure a higher level of privacy using our methods. Note that given the large sample size of this dataset, unsurprisingly all the covariates are significantly predictive for the nonprivate estimators. All univariate Wald statistics for the slope parameters in this example yield $p$-values smaller than $10^{-16}$ for the nonprivate estimators. Since our differentially private $p$-values give similar results we chose not to report them.

6. Concluding Remarks

We introduced a general framework for differentially private statistical inference for parametric models based on M-estimators. The central idea of our approach is to leverage tools from robust statistics in the design of a mechanism for the release of differentially private statistical outputs. In particular, we release noisy versions of statistics of interest that we view as functionals of the empirical distribution induced by the data. We use a bound of their influence function to scale the random perturbation added to the desired statistics to guarantee privacy. As a result, we propose a new class of consistently differentially private estimators that can be easily and efficiently computed, and provide a general framework for parametric hypothesis testing with privacy guarantees.

An interesting extension to be explored in the future is the construction of differentially private tests in the context of non-parametric and high-dimensional regression. In principle the idea of using the influence function to calibrate the noise added to test functionals also seems intuitive in these settings, but the technical challenge of these extensions is 2-fold. First, there are no general results regarding the level influence function of tests for these settings. Second, the influence function of non-parametric and high-dimensional penalized estimators has been formulated for a fixed tuning parameter (Christmann and Steinwart 2007; Avella-Medina 2017). Since in practice this parameter is usually chosen by some data driven criterion, it would be necessary to account for this selection step in the derivation of differentially private statistics following the approach of this work. Another interesting direction for future research is to explore whether information-standardized influence functions could be used to derive better or more general differentially private estimators (Hampel et al. 1986; He and Simpson 1992). It would also be interesting to explore the construction of tests based on alternative approaches to differential privacy such as objective function perturbation (Chaudhuri and Monteleoni 2008; Chaudhuri, Monteleoni, and Sarwate 2011; Kiefer, Smith, and Thakurta 2012) or stochastic gradient descent (Rajkumar and Argawal 2012; Bassily, Smith, and Thakurta 2014; Wang, Lee, and Kifer 2015).

Supplementary Materials

The supplementary materials include all the omitted proofs and some auxiliary results regarding the influence function. They also include more extended discussions of competing methods and details about the estimation of the variance in the noise calibration of our Gaussian mechanism.

Acknowledgment

The author would like to thank Roy Welsch for many helpful discussions.

Funding

The author is grateful for the financial support of the Swiss National Science Foundation.

References

Avella-Medina, M. (2017), "Influence Functions for Penalized M-Estimators," Bernoulli, 23, 3178–3196. [982]
Avella-Medina, M., and Brunel, V. E. (2019), "Differentially Private Sub-Gaussian Location Estimators," arXiv no. 1906.11923. [973]
Awan, J., and Slavkovic, A. (2018), "Differentially Private Uniformly Most Powerful Tests for Binomial Data," in Advances in Neural Information Processing Systems, pp. 4208–4218. [971]
Barrientos, A. F., Reiter, J. P., Machanavajjhala, A., and Chen, Y. (2019), "Differentially Private Significance Tests for Regression Coefficients," Journal of Computational and Graphical Statistics, 28, 440–453. [971]
Bassily, R., Smith, A., and Thakurta, A. (2014), "Private Empirical Risk Minimization: Efficient Algorithms and Tight Error Bounds," in IEEE 55th Annual Symposium on Foundations of Computer Science, pp. 464–473. [969,970,971,976,982]
Belksley, D. A., Kuh, E., and Welsch, R. E. (2005), Regression Diagnostics: Identifying Influential Data and Sources of Collinearity, New York: Wiley. [970]
Cai, T. T., Wang, Y., and Zhang, L. (2019), "The Cost of Differential Privacy: Optimal Rates of Convergence for Parameter Estimation With Differential Privacy," manuscript. [972]
Canonne, C. L., Kamath, G., McMillan, A., Smith, A., and Ullman, J. (2019), "The Structure of Optimal Private Tests for Simple Hypotheses," in Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, pp. 310–321. [971]
Canonne, C. L., Kamath, G., McMillan, A., Ullman, J., and Zakynthinou, L. (2019), "Private Identity Testing for High-Dimensional Distributions," arXiv no. 1905.11947. [971]
Cantoni, E., and Ronchetti, E. (2001), "Robust Inference for Generalized Linear Models," Journal of the American Statistical Association, 96, 1022–1030. [975,978]
Catonii, O. (2012), "Challenging the Empirical Mean and Empirical Variance: A Deviation Study," Annales de l’Institut Henri Poincaré, 48, 1148–1185. [976]
Chaudhuri, K., and Monteleoni, C. (2008), "Privacy-Preserving Logistic Regression," in Advances in Neural Information Processing Systems, pp. 289–296. [970,982]
