Parametric Bootstrap for Differentially Private Confidence Intervals

Cecilia Ferrando, Shufan Wang, Daniel Sheldon
Manning College of Information and Computer Sciences
University of Massachusetts Amherst
{cferrando, shufanwang, sheldon}@cs.umass.edu

Abstract

The goal of this paper is to develop a practical and general-purpose approach to construct confidence intervals for differentially private parametric estimation. We find that the parametric bootstrap is a simple and effective solution. It cleanly reasons about variability of both the data sample and the randomized privacy mechanism and applies "out of the box" to a wide class of private estimation routines. It can also help correct bias caused by clipping data to limit sensitivity. We prove that the parametric bootstrap gives consistent confidence intervals in two broadly relevant settings, including a novel adaptation to linear regression that avoids accessing the covariate data multiple times. We demonstrate its effectiveness for a variety of estimators, and find empirically that it provides confidence intervals with good coverage even at modest sample sizes and performs better than alternative approaches.

1 INTRODUCTION

Differential privacy provides a rubric for drawing inferences from data sets without compromising the privacy of individuals.

This paper is about privately constructing confidence intervals. In the non-private case, approximate methods based on asymptotic normality or the bootstrap (Efron, 1979) apply to a wide range of models and are very widely used in practice. In the private case, such "swiss army knife" methods are hard to find. The situation is complicated by the fact that private estimation procedures are necessarily randomized, which leads to a distinct source of randomness ("privacy noise") in addition to the usual random draw of a finite sample from a population ("sampling noise"). We find experimentally that asymptotic methods are significantly less effective in private settings, due to privacy noise that becomes negligible only for very large sample sizes (Section 7). Bootstrap approaches face the challenge of incurring privacy costs by accessing the data many times (Brawner and Honaker, 2018).

This paper advocates using the parametric bootstrap as a simple and effective method to construct confidence intervals for private statistical estimation. The parametric bootstrap resamples data sets from an estimated parametric model to approximate the distribution of the estimator. It is algorithmically simple, can be used with essentially any private estimator, and cleanly reasons about both sampling noise and privacy noise. Unlike the traditional bootstrap, it is based on post-processing and avoids accessing the data many times, so it often has little or no privacy burden. By reasoning about the distribution of a finite sample, it makes fewer assumptions than purely asymptotic methods and significantly mitigates the problem of non-negligible privacy noise. The parametric bootstrap can also help correct bias in private estimation caused by artificially bounding data to limit sensitivity.

We first introduce the parametric bootstrap and discuss its application to private estimation, including methods to construct confidence intervals and correct bias. We then review parametric bootstrap theory, and apply the parametric bootstrap to obtain provably consistent confidence intervals in two private estimation settings—exponential families and linear regression sufficient statistic perturbation (SSP)—as well as an empirical demonstration for the "one posterior sample" (OPS) method (Wang et al., 2015; Foulds et al., 2016; Zhang et al., 2016). These demonstrate the broad applicability the parametric bootstrap to private estimation.

One limitation of the parametric bootstrap is the restriction to fully parametric estimation. For example,

Proceedings of the 25th International Conference on Artificial Intelligence and Statistics (AISTATS) 2022, Valencia, Spain. PMLR: Volume 151. Copyright 2022 by the author(s).

it doesn't apply directly to regression problems that do not have a parametric model for covariates, and may not be appropriate very complex data. In our linear regression application, we contribute a novel hybrid bootstrap approach to circumvent this limitation; the resulting method is easy to use and simultaneously estimates regression coefficients and constructs confidence intervals with good coverage properties. A second limitation is computational cost, which scales with the data size. For small or medium data sets, the cost is likely manageable. For very large ones, cheap asymptotic methods will often be adequate (see Section 7; for exponential families and linear regression with sufficient statistic perturbation, the asymptotic distributions are a relatively simple byproduct of our bootstrap theory). However, it is unknown in general how large data must be for asymptotic methods to perform well.

The code to reproduce our experiments is available at `https://github.com/ceciliaferrando/PB-DP-CIs`.

2 BACKGROUND

Differential privacy is a formal definition to capture the notion that, to maintain privacy, the output of an algorithm should remain nearly unchanged if the data of one individual changes. Say that two data sets X and X' of size n are *neighbors* if they differ in exactly one data record.

Definition 1 (Differential privacy, Dwork et al. 2006)**.** *A randomized algorithm $\mathcal{A}$ satisfies ϵ-differential privacy (ϵ-DP) if, for neighboring data sets X and X', and any subset $O \subseteq$ Range($\mathcal{A}$),*

$$\Pr[\mathcal{A}(X) \in O] \leq \exp(\epsilon) \Pr[\mathcal{A}(X') \in O].$$

One common way to achieve differential privacy is by injecting calibrated noise onto the statistics computed from the data. Let f be any function that maps data sets to $\mathbb{R}^d$. The magnitude of noise required to privatize the computation of f depends on its *sensitivity*.

Definition 2 (Sensitivity, Dwork et al. 2006)**.** *The sensitivity of a function f is*

$$\Delta f = \max_{X,X'} \|f(X) - f(X^{'})\|_1$$

where $X, X^{'}$ are any two neighboring data sets.

When f is additive, it is straightforward to bound its sensitivity (proof in Appendix A):

Claim 1. *Suppose $X = (x_1, \ldots, x_n)$ and $f(X) = \sum_{i=1}^n g(x_i)$ where g maps data points to $\mathbb{R}^m$. Let* $\text{width}(g_j) = \max_x g_j(x) - \min_x g_j(x)$ *where x ranges over the data domain. Then $\Delta f \leq \sum_{j=1}^m \text{width}(g_j)$, which is a constant independent of n.*

Many algorithms satisfy differential privacy by using the Laplace mechanism.

Definition 3 (Laplace mechanism, Dwork et al. 2006)**.** *Given a function f that maps data sets to $\mathbb{R}^m$, the Laplace mechanism outputs the random variable $\mathcal{L}(X) \sim \text{Lap}(f(X), \Delta f/\epsilon)$ from the Laplace distribution, which has density* $\text{Lap}(z; u, b) = (2b)^{-m} \exp(-\|z-u\|_1 / b)$. *This corresponds to adding zero-mean independent noise $u_i \sim \text{Lap}(0, \Delta f/\epsilon)$ to each component of $f(X)$.*

3 PARAMETRIC BOOTSTRAP

Algorithm 1 Parametric Bootstrap

```
Require: $x_{1:n}$, $B$, estimator $\mathcal{A}$
$\hat{\theta}, \hat{\tau} \leftarrow \mathcal{A}(x_{1:n})$
for $b$ from 1 to $B$ do
    $x_1^*, \ldots, x_n^* \sim P_{\hat{\theta}}$
    $\hat{\theta}^{*b}, \hat{\tau}^{*b} \leftarrow \mathcal{A}(x_{1:n}^*)$
return $\hat{\tau}$, $(\hat{\tau}^{*1}, \ldots, \hat{\tau}^{*B})$
```

We consider the standard setup of parametric statistical inference, where a data sample $x_{1:n} = (x_1, \ldots, x_n)$ is observed and each x_i is assumed to be drawn independently from a distribution P_θ in the family $\{P_\theta : \theta \in \Theta\}$ with unknown θ.

The goal is to estimate some population parameter $\tau = \tau(\theta)$, the *estimation target*, via an estimator $\hat{\tau} = \hat{\tau}(x_{1:n})$.[1] We also seek a $1-\alpha$ confidence interval for τ, that is, an interval $[\hat{a}_n, \hat{b}_n]$ such that $\mathbb{P}_\theta(\hat{a}_n \leq \tau \leq \hat{b}_n) \approx 1 - \alpha$, where $\mathbb{P}_\theta$ is the probability measure over the full sample when the true parameter is θ. We will require $\hat{\tau}$ and $[\hat{a}, \hat{b}]$ to be differentially private. Our primary focus is not designing private estimators $\hat{\tau}$, but designing methods to construct private confidence intervals $[\hat{a}, \hat{b}]$ that can be used for many estimators and have little additional privacy burden.

The parametric bootstrap is a simple way to approximate the distribution of $\hat{\tau}$ for confidence intervals and other purposes. It is a variant of Efron's bootstrap (Efron, 1979, 1981a,b; Efron and Tibshirani, 1986), which runs an estimator many times on simulated data sets whose distribution approximates the original data. In the parametric bootstrap, data sets are simulated from $P_{\hat{\theta}}$, the parametric distribution with estimated parameter $\hat{\theta}$.[2] The procedure is shown in Algorithm 1, where $\mathcal{A}$ is an algorithm that computes the estimates $\hat{\theta}$ and $\hat{\tau}$ from the data. A simple case is when $\hat{\tau}(x_{1:n}) = \tau\big(\hat{\theta}(x_{1:n})\big)$ but in general these may be estimated separately.

[1] We use a hat on variables that are functions of the data and therefore random.

[2] In the non-parametric bootstrap, data sets are simulated from the empirical distribution of $x_{1:n}$.

The parametric bootstrap is highly compatible with differential privacy. The data is only accessed in Line 2, so the only requirement is that $\mathcal{A}$ be differentially private (which necessitates it is randomized). The remaining steps are post-processing and incur no additional privacy cost. The simulation cleanly handles reasoning about both data variability (Line 4) and randomness in the estimator (Line 5). When the estimation target is θ, we have $\hat{\tau} = \hat{\theta}$, and the procedure incurs no privacy cost beyond that of the private estimator $\hat{\tau}$. In other cases, additional privacy budget is required to estimate the full vector θ including nuisance parameters; an example is estimating the mean of a Gaussian with unknown variance (Du et al., 2020).

3.1 Confidence Intervals and Bias Correction

Table 1: Bootstrap confidence intervals. $\hat{\tau}$ and $\hat{\sigma}$ are parameter and standard error estimates of the main procedure, and $\hat{\tau}^*$ and $\hat{\sigma}^*$ are their bootstrapped counterparts. $\hat{\xi}_\gamma$ is the $1-\gamma$ quantile of a pivot, either $\hat{\tau}^* - \hat{\tau}$ or $(\hat{\tau}^* - \hat{\tau})/\hat{\sigma}^*$, and $\hat{\zeta}_\gamma$ is the $1-\gamma$ quantile of $\hat{\tau}^*$.

Interval	Target	Target interval	τ interval
Pivotal	$\hat{\tau}-\tau$	$[\hat{\xi}_{1-\frac{\alpha}{2}}, \hat{\xi}_{\frac{\alpha}{2}}]$	$[\hat{\tau} - \hat{\xi}_{\frac{\alpha}{2}},\ \hat{\tau} - \hat{\xi}_{1-\frac{\alpha}{2}}]$
Studentized pivotal	$\frac{(\hat{\tau}-\tau)}{\hat{\sigma}}$	$[\hat{\xi}_{1-\frac{\alpha}{2}}, \hat{\xi}_{\frac{\alpha}{2}}]$	$[\hat{\tau} - \hat{\xi}_{\frac{\alpha}{2}}\hat{\sigma},\ \hat{\tau} - \hat{\xi}_{1-\frac{\alpha}{2}}\hat{\sigma}]$
Efron's percentile	$\hat{\tau}$	$[\hat{\zeta}_{1-\frac{\alpha}{2}},\ \hat{\zeta}_{\frac{\alpha}{2}}]$	$[\hat{\zeta}_{1-\frac{\alpha}{2}},\ \hat{\zeta}_{\frac{\alpha}{2}}]$

There are several well known methods to compute confidence intervals from bootstrap replicates. Three are listed in Table 1; note that names are inconsistent in the literature.[3] The general principle is to treat the pair $(\hat{\tau}^*, \hat{\tau})$ analogously to $(\hat{\tau}, \tau)$ to approximate the distribution of the latter. The intervals differ according to what function of $(\hat{\tau}, \tau)$ they target. A simple example is to approximate the "pivot" $\hat{\tau}-\tau$ by $\hat{\tau}^*-\hat{\tau}$, which leads to the pivotal interval. To construct it, we estimate the $1-\gamma$ quantile of $\hat{\tau}^* - \hat{\tau}$ as the $1-\gamma$ quantile of the bootstrap replicates $(\hat{\tau}^{*1} - \hat{\tau}, \ldots, \hat{\tau}^{*B} - \hat{\tau})$. The number of replicates controls the error introduced by this step. This error is usually ignored theoretically because, in principle, it can be reduced arbitrarily with enough computation, and it can be controlled well in practice. The studentized pivotal interval targets $(\hat{\tau} - \tau)/\hat{\sigma}$ instead, where $\hat{\sigma}$ is a standard error estimate of the main procedure; it can converge faster than the pivotal interval (Wasserman, 2006). Efron's percentile interval targets $\hat{\tau}$ directly and, while simple, its logic is less obvious; it can also be viewed as targeting the pivot $\hat{\tau} - \tau$ with a "reversed" interval, which is how theoretical properties are shown. By approximating $\hat{\tau} - \tau$ by $\hat{\tau}^* - \hat{\tau}$, we can also estimate the bias of $\hat{\tau}$. This leads to a simple bias corrected estimator $\hat{\tau}_{\text{bc}}$:

$$\widehat{\text{bias}} = \mathbb{E}[\hat{\tau}^* - \hat{\tau}], \quad \hat{\tau}_{\text{bc}} \leftarrow \hat{\tau} - \widehat{\text{bias}}.$$

Similar to the quantiles above, $\mathbb{E}[\hat{\tau}^* - \hat{\tau}]$ is estimated as the sample mean over bootstrap replicates.

3.2 Significance and Connection to Other Resampling Methods for Private Estimation

The parametric bootstrap can be applied to any parametric estimation problem, a wide range of private estimators, is very accurate in practice, and has little or no additional cost in terms of privacy budget or algorithm development. These make it an excellent default choice (to our knowledge, the best) for constructing private confidence intervals for any parametric estimation problem with small to medium data sets.

That such a simple and effective choice is available is not articulated in the literature. Two prior works use methods that can be viewed as the parametric bootstrap, but do not discuss the classical procedure and its wide ranging applications, or different facets of bootstrap methodology such as techniques for constructing confidence intervals and bootstrap theory. Specifically, the simulation approach of Du et al. (2020) for Gaussian mean confidence intervals is equivalent to the parametric bootstrap with a non-standard variant of Efron's percentile intervals, and performed very well empirically. In their application to independence testing, Gaboardi et al. (2016) approximate the distribution of a private test statistic by simulating data from a null model after privately estimating its parameters; this can be viewed as an application of the parametric bootstrap to the null model.

Several other works use resampling techniques that resemble the parametric bootstrap for a similar, but conceptually distinct, purpose (D'Orazio et al., 2015; Wang et al., 2019; Evans et al., 2019). A typical setup is when $\hat{\tau} = \tau' + \eta$, where τ' is a non-private estimator and η is noise added for privacy. Standard asymptotics are used to approximate $\sqrt{n}(\tau' - \tau)$ as $\mathcal{N}(0, \hat{\sigma})$, where $\hat{\sigma}$ is a (private) standard error estimate for τ'. For the private estimator, this gives $\sqrt{n}(\hat{\tau}-\tau) \approx \mathcal{N}(0, \hat{\sigma}) + \sqrt{n}\eta$. Because η has known distribution, Monte Carlo sampling can be used to draw samples from $\mathcal{N}(0, \hat{\sigma}) + \sqrt{n}\eta$

[3] Our mathematical presentation follows Van der Vaart (2000), but names follow Wasserman (2006). The names "pivotal" and "studentized pivotal" and are descriptive and avoid the confusion of "percentile interval" sometimes referring to the pivotal interval and other times to Efron's percentile interval. The possessive "Efron's" (Van der Vaart, 2000) clarifies that we use Efron's definition of "percentile interval" (e.g., Efron and Hastie, 2016).

for computing confidence intervals or standard errors. The key distinction is that *standard* asymptotics are used to approximate the distribution of τ', which captures all variability due to the data, and sampling is used only to combine that distribution with the privacy noise distribution. In contrast, the key feature of a bootstrap method is that it resamples data sets to reason about estimator variability due the random data, and thereby avoids standard asymptotics. This technique also does not apply when the privacy mechanism is more complicated than adding additive noise to a non-private estimate (cf. the OPS example of Sec. 5).

4 BOOTSTRAP THEORY

This section gives general results we can use to argue correctness of bootstrap confidence intervals in private settings. We give a general notion of "bootstrap" estimator that covers different resampling methods. Let $(\Omega, \mathcal{F}, \mathbb{P}_\theta)$ be the probability space for $x_1, x_2, \ldots \sim P_\theta$ and $\eta_1, \eta_2, \ldots$ where, for a given n, the data is $x_{1:n}$ and η_n captures any other randomness used in the privacy mechanism or estimator; we refer to this as the "outer" probability space. A bootstrap estimator is defined in terms of a random experiment over an "inner" probability space conditional on $\omega \in \Omega$ and n. Let $\mathbb{P}_n^*(\cdot|\omega)$ be a Markov kernel defining this space. The traditional bootstrap uses $\mathbb{P}_n^*(\cdot|\omega) = \hat{P}^n$ with $\hat{P}(dx) = \frac{1}{n}\sum_i \delta_{x_i}(dx)$; the parametric bootstrap uses $\hat{P} = P_{\hat{\theta}}$ instead. Our hybrid model in Section 5 uses a custom resampling method, which gives a custom measure $\mathbb{P}_n^*(\cdot|\omega)$.

For our purposes, a bootstrap estimator of a parameter $\tau(\theta)$ is a random variable $\hat{\tau}_n^*$ in the inner probability space that simulates the parameter estimate $\hat{\tau}_n$ of the "main" procedure. Typically, the bootstrap estimator arises from running the main procedure on resampled data. That is, if $\hat{\tau}_n = T_n(\omega)$, then $\hat{\tau}^* = T_n(\omega^*)$ with $\omega^* \sim \mathbb{P}_n^*(\cdot \mid \omega)$. Our hybrid OLS bootstrap will deviate slightly from this pattern.

4.1 Consistency

Bootstrap "success" has to do with the asymptotic distributions of the (approximate) pivot $\sqrt{n}(\hat{\tau}_n - \tau)$ and its bootstrapped counterpart $\sqrt{n}(\hat{\tau}_n^* - \hat{\tau}_n)$. For studentized intervals, the pivot $(\hat{\tau}_n - \tau)/\hat{\sigma}_n$ is used instead, where $\hat{\sigma}_n$ is a standard error estimate of the main procedure; theory for this case is a straightforward extension if $\hat{\sigma}_n \to \sigma(\theta)$ in $\mathbb{P}_\theta$-probability (Van der Vaart, 2000; Beran, 1997).

Definition 4. *The bootstrap estimator $\hat{\tau}_n^*$ is* consistent *if*

$$\sup_x \Big| \mathbb{P}_n^*\Big(\sqrt{n}(\hat{\tau}_n^* - \hat{\tau}_n) \le t \mid \omega\Big) - \\ -\mathbb{P}_\theta\Big(\sqrt{n}(\hat{\tau}_n - \hat{\tau}) \le t\Big)\Big| \xrightarrow{P} 0 \tag{1}$$

with convergence in $\mathbb{P}_\theta$-probability.

This says that the Kolmogorov-Smirnov distance between the distribution of the pivot and the conditional distribution of the bootstrapped pivot converges to zero, in probability over ω.

For many estimators $\sqrt{n}(\hat{\tau}_n - \tau) \rightsquigarrow T$ for a continuous random variable T. In this case it is enough for the bootstrapped pivot to converge to the correct limit distribution.

Lemma 1 (Van der Vaart 2000, Eq. (23.2))**.** *Suppose $\sqrt{n}(\hat{\tau}_n - \tau) \rightsquigarrow T$ for a random variable T with continuous distribution function F. Then, $\hat{\tau}_n^*$ is consistent if and only if, for all t,*

$$\mathbb{P}_n^*\Big(\sqrt{n}(\hat{\tau}_n^* - \hat{\tau}_n) \le t \mid \omega\Big) \xrightarrow{P} F(t).$$

Consistency is also preserved under continuous mappings: if $\hat{\tau}_n^*$ is consistent relative to $\sqrt{n}(\hat{\tau}_n - \tau) \rightsquigarrow T$ and g is continuous, then $g(\hat{\tau}_n^*)$ is consistent relative to $\sqrt{n}\big(g(\hat{\tau}_n) - g(\tau)\big)$ (Beran, 1997). In our applications we will show consistency of a bootstrap estimator $\hat{\theta}_n^*$ for the full parameter vector θ, which implies consistency for continuous functions of θ; a simple application is selecting one entry and constructing a confidence interval.

4.2 Confidence interval consistency

Bootstrap consistency implies consistent confidence intervals. The confidence interval $[\hat{a}_n, \hat{b}_n]$ for $\tau = \tau(\theta)$ is (conservatively) asymptotically consistent at level $1 - \alpha$ if, for all θ,

$$\liminf_{n\to\infty} \mathbb{P}_\theta\left(\hat{a}_n \le \tau \le \hat{b}_n\right) \ge 1 - \alpha. \tag{2}$$

Lemma 2 (Van der Vaart 2000, Lemma 23.3)**.** *Suppose $\sqrt{n}(\tau_n - \tau) \rightsquigarrow T$ for a random variable T with continuous distribution function and τ_n^* is consistent. Then the pivotal intervals are consistent, and, if T is symmetrically distributed around zero, then Efron's percentile intervals are consistent. When the analogous conditions hold for the studentized pivot $(\hat{\tau}_n - \tau)/\hat{\sigma}_n$, studentized intervals are consistent.*

4.3 Parametric bootstrap consistency

Beran (1997) showed that asymptotic equivariance of the main estimator guarantees consistency of the parametric bootstrap. Let $H_n(\theta)$ be the distribution of $\sqrt{n}(\hat{\tau}_n - \tau(\theta))$ under $\mathbb{P}_\theta$.

Definition 5 (Asymptotic equivariance, Beran 1997)**.** *The estimator $\hat{\tau}_n$ is asymptotically equivariant if $H_n(\theta + h_n/\sqrt{n})$ converges to a limiting distribution $H(\theta)$ for all convergent sequences h_n and all θ.*

Theorem 1 (Parametric bootstrap consistency)**.** *Suppose $\sqrt{n}(\hat{\theta}_n - \theta) \rightsquigarrow J(\theta)$ and $\hat{\tau}_n$ is asympotitcally equivariant with continuous limiting distribution $H(\theta)$. Then the parametric bootstrap estimator $\hat{\tau}^*_n$ is consistent.*

All proofs are provided in the appendix. Furthermore, under reasonably general conditions, the reverse implication is true, with bootstrap failures occurring precisely at those parameter values θ_0 for which asymptotic equivariance does not hold (Beran, 1997).

5 APPLICATIONS

We apply the parametric bootstrap to three private estimation settings: (i) exponential families with sufficient statistic perturbation (SSP), (ii) linear regression with SSP, (iii) the "one posterior sample" (OPS) estimator.

Exponential Families A family of distributions is an *exponential family* if P_θ has a density of the form:

$$p(x;\theta) = h(x)\exp(\theta^T T(x) - A(\theta))$$

where $h(x)$ is a base measure, θ is the natural parameter, T is the sufficient statistic function, and A is the log-partition function. Define the log-likelihood function of an exponential family as $\ell(\theta;x) = \log p(x;\theta) - \log h(x) = \theta^T T(x) - A(\theta)$. The constant term $\log h(x)$ does not affect parameter estimation and is subtracted for convenience. For a sample $x_{1:n}$, let $T(x_{1:n}) = \sum_{i=1}^n T(x_i)$. The log-likelihood of the sample is

$$\ell(\theta;x_{1:n}) = \theta^T T(x_{1:n}) - nA(\theta) := f\big(\theta; T(x_{1:n})\big),$$

which depends on the data only through the sufficient statistic $T(x_{1:n})$. The maximum-likelihood estimator (MLE) is $\hat{\theta} = \operatorname{argmax}_\theta f\big(\theta; T(x_{1:n})\big)$.

A simple way to create a private estimator is sufficient statistic perturbation (SSP); that is, to privatize the sufficient statistics using an elementary privacy mechanism such as the Laplace or Gaussian mechanism prior to solving the MLE problem. SSP is a natural choice because $T(x_{1:n})$ is a compact summary of the data and has sensitivity that is easy to analyze, and it often works well in practice (Bernstein and Sheldon, 2018; Foulds et al., 2016). Specifically, it means solving

$$\hat{\theta} = \operatorname*{argmax}_\theta f\big(\theta, T(x_{1:n}) + w\big) \tag{SSP-MLE}$$

where w is a suitable noise vector. This problem has closed form solutions for many exponential families and standard numerical routines apply to others. For the Laplace mechanism, $w_j \sim \text{Lap}(\frac{\Delta}{\epsilon})$ for all j, where $\Delta = \sum_j \text{width}(T_j)$ is an upper bound on the L_1 sensitivity of $T(x_{1:n})$ by Claim 1. If $\text{width}(T_j)$ is not known or is unbounded, the analyst must supply bounds and guarantee they are met, e.g., by discarding data points that don't meet the bounds, or clamping them to the bounded interval.

Theorem 2. *Let $\hat{\theta}_n$ be the solution to the (SSP-MLE) optimization problem for a sample $x_{1:n}$ from an exponential family model that satisfies the regularity conditions given in Davison (2003, Section 4.4.2). Then $\sqrt{n}(\hat{\theta}_n - \theta)$ is asymptotically equivariant with limiting distribution $\mathcal{N}(0, I(\theta)^{-1})$, where $I(\theta) = \nabla^2 A(\theta)$ is the Fisher information. This implies consistency of the parametric bootstrap estimator $\hat{\theta}^*_n$.*

Linear Regression We consider a linear regression model where we are given n pairs[4] $(\mathbf{x}_i, y_i)$ with $\mathbf{x}_i \in \mathbb{R}^p$ and $y_i \in \mathbb{R}$ assumed to be generated as $y_i = \beta^T \mathbf{x}_i + u_i$, where the errors u_i are i.i.d., independent of $\mathbf{x}_i$, zero-mean, and have finite variance σ^2, and the $\mathbf{x}_i$ are i.i.d. with $\mathbb{E}[\mathbf{x}\mathbf{x}^T] = Q$. We wish to estimate the regression coefficients $\beta \in \mathbb{R}^p$. Let $X \in \mathbb{R}^{n\times p}$ be the matrix with ith row equal to $\mathbf{x}_i^T$ and $\mathbf{y}, \mathbf{u} \in \mathbb{R}^N$ be the vectors with ith entries y_i and u_i, respectively. The *ordinary least squares* (OLS) estimator is:

$$\hat{\beta} = (X^T X)^{-1} X^T \mathbf{y}. \tag{3}$$

Like the MLE in exponential families, Eq. (3) depends on the data only through sufficient statistics $X^T X$ and $X^T \mathbf{y}$, and SSP is a simple way to privatize the estimator that works very well in practice (Wang, 2018). The privatized estimator is

$$\hat{\beta} = (X^T X + V)^{-1}(X^T \mathbf{y} + w), \tag{SSP-OLS}$$

where $V \in \mathbb{R}^{p\times p}$ and $w \in \mathbb{R}^p$ are additive noise variables drawn from distributions P_V and P_w to ensure privacy.

For the Laplace mechanism, we use

$$V_{jk} \sim \text{Lap}(0, \Delta_V/\epsilon_1) \text{ for } j \le k, \text{ and } V_{kj} = V_{jk}, \tag{4}$$

$$w_j \sim \text{Lap}(0, \Delta_w/\epsilon_2) \text{ for all } j, \tag{5}$$

where Δ_V and Δ_w bound the L_1 sensitivity of V and w, respectively. The result is $(\epsilon_1 + \epsilon_2)$-DP. Because $X^T X = \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T$ and $X^T \mathbf{y} = \sum_{i=1}^n \mathbf{x}_i y_i$ are additive, we can take $\Delta_V = \sum_{j\le k} \text{width}(x_j) \cdot \text{width}(x_k)$ and $\Delta_w = \sum_j \text{width}(x_j) \cdot \text{width}(y)$, where $\text{width}(x_j)$ and $\text{width}(y)$ are widths of the jth feature and response variable, respectively, which are enforced by the modeler.

[4] We use boldface for vectors as needed to distinguish from scalar quantities.

Algorithm 2 OLS hybrid parametric bootstrap

```
Require: $\hat{\beta}$, $\hat{Q}$, $\hat{\sigma}^2$, $P_V$, $P_w$, $B$, $n$, where $\hat{\beta} = (X^T X + V)^{-1}(X^T \mathbf{y} + w)$ and $\hat{Q} = \frac{1}{n} X^T X + \frac{1}{n} V$ for $V \sim P_V, w \sim P_w$, and $\hat{\sigma}^2$ is a private estimate of $\sigma^2$.
for $b$ from 1 to $B$ do
    Sample $V^* \sim P_V$ (Eq. (4) for Laplace noise)
    Sample $w^* \sim P_w$ (Eq. (5) for Laplace noise)
    Sample $Z^* \sim \mathcal{N}(0, \hat{\sigma}^2 \hat{Q})$
    Set $\hat{Q}^* = \hat{Q} + \frac{1}{n} V^*$
    Set $\beta^{*b} = (\hat{Q}^*)^{-1} \hat{Q} \hat{\beta} + (\hat{Q}^*)^{-1} \big(\frac{1}{\sqrt{n}} Z^* + \frac{1}{n} w^*\big)$.
return $\beta^{*1}, \ldots, \beta^{*B}$
```

For confidence intervals, we will also need a private estimate of σ^2: let $\hat{\sigma}^2 = (n-p)^{-1} \sum_{i=1}^n (y_i - \hat{\beta}^T \mathbf{x}_i)^2 + \text{Lap}(0, \Delta_z/\epsilon_3)$ where $\Delta_z = \text{width}((y - \hat{\beta}^T \mathbf{x})^2)$. The released values for SSP-OLS are then $(X^T X + V, X^T \mathbf{y} + w, \hat{\beta}, \hat{\sigma}^2)$, which satisfy $(\epsilon_1 + \epsilon_2 + \epsilon_3)$-DP.

Limitations of parametric bootstrap for private regression The parametric bootstrap is more difficult to apply to regression problems in a private setting due to the covariates. It is typical to bootstrap conditioned on X, which means simulating new response variables $\mathbf{y}$ from a parametric distribution $p(\mathbf{y}|X; \hat{\beta}, \hat{\sigma}^2)$, where $\hat{\beta}$ and $\hat{\sigma}^2$ are (privately) estimated parameters, and a fully parametric distribution $p(u; \sigma^2)$ is assumed for errors. A bootstrap replicate would look like $\hat{\beta}^* = (X^T X)^{-1} X^T \mathbf{y}^*$ with $\mathbf{y}^* = X\hat{\beta} + \mathbf{u}^*$ and $\mathbf{u}^*$ simulated form the error distribution. The challenge is that X is accessed to generate each replicate, so to make it differentially private would require additional randomization and consume privacy budget. An alternative would be to posit a model $p(\mathbf{x}; \theta)$ and perform the parametric bootstrap with respect to a joint model $p(\mathbf{x}, y; \theta, \beta, \sigma^2)$, but the additional demand to model covariates is unappealing in a regression context.

Hybrid parametric bootstrap for OLS We propose a novel hybrid approach that avoids the need to repeatedly access covariate data or to model the covariate or error distributions explicitly. Conceptually, we use the part of the standard asymptotic analysis that "works well" to approximate the relevant statistics of the covariate data, and use the parametric bootstrap to deal with the noise added for privacy. Following standard analysis for OLS, we can substitute $\mathbf{y} = X\beta + \mathbf{u}$ in (SSP-OLS) and scale terms to get:

$$
\begin{aligned}
\sqrt{n}\hat{\beta}_n &= \sqrt{n}\left(\check{Q}_n + \frac{1}{n}V\right)^{-1} \check{Q}_n \beta + \\
&\quad + \left(\check{Q}_n + \frac{1}{n}V\right)^{-1} \left(\frac{1}{\sqrt{n}} X^T \mathbf{u} + \frac{1}{\sqrt{n}} w\right), \\
\check{Q}_n &= \frac{1}{n} X^T X.
\end{aligned} \tag{6}
$$

This expression is instructive to see the different sources of randomness that contribute to the variability of $\hat{\beta}_n$: the terms $\check{Q}_n = \frac{1}{n} X^T X$ and $\frac{1}{\sqrt{n}} X^T \mathbf{u}$ are due to data variability, and $\frac{1}{n} V$ and $\frac{1}{\sqrt{n}} w$ are due to privacy. We form a bootstrap estimator $\hat{\beta}_n^*$ that treats $(\hat{\beta}_n, \beta)$ analogously to $(\hat{\beta}_n^*, \hat{\beta}_n)$ and simulates the different sources of variability using the best available information about their distributions:

$$
\begin{aligned}
\sqrt{n}\hat{\beta}_n^* &= \sqrt{n}\left(\hat{Q}_n + \frac{1}{n}V^*\right)^{-1} \hat{Q}_n \hat{\beta}_n + \\
&\quad + \left(\hat{Q}_n + \frac{1}{n}V^*\right)^{-1} \left(Z_n^* + \frac{1}{\sqrt{n}} w^*\right), \\
Z_n^* &\sim \mathcal{N}(0, \hat{\sigma}_n^2 \hat{Q}_n), \quad V^* \sim P_V, \quad w^* \sim P_w, \\
\hat{Q}_n &= \frac{1}{n} X^T X + \frac{1}{n} V.
\end{aligned} \tag{7}
$$

All privacy terms in Eq. (6) are simulated from their exact distributions in Eq. (7). The variables $\check{Q}_n$ and $\hat{Q}_n$ represent approximations of $Q = \mathbb{E}[\mathbf{x}\mathbf{x}^T]$ available to the corresponding estimator [5]. Both quantities converge in probability to Q. Our choice not to simulate variability in these estimates due to the covariates is analogous to the "fixed X" bootstrap strategy for regression problems (Fox, 2002); we *do* simulate the variability due to privacy noise added to the estimates. The blue terms represent contributions to estimator variability due to interactions between covariates and unobserved noise variables. In a traditional bootstrap, we might simulate this term in Eq. (7) as $\frac{1}{\sqrt{n}} X^T \mathbf{u}^*$ where $\mathbf{u}^*$ are simulated errors, but, as described above, we do not wish to access X within the bootstrap procedure. Instead, because we know $\frac{1}{\sqrt{n}} X^T \mathbf{u} \rightsquigarrow \mathcal{N}(0, \sigma^2 Q)$ by the central limit theorem,[6] and because σ^2 and Q are estimable, we simulate this term directly from the normal distribution with estimated parameters.

[5] In the implementation, when the private estimate of Q is not positive semidefinite (PSD), it is projected to a nearby PSD matrix. Specifically, we use the projection which sets all negative eigenvalues to a small positive number, which is equivalent to finding the closest PSD matrix in Frobenius norm.

[6] This is a standard result of OLS asymptotics and is expected to be accurate for modest sample sizes.

Theorem 3. *The private estimator satisfies* $\sqrt{n}(\hat{\beta}_n - \beta) \rightsquigarrow \mathcal{N}(0, \sigma^2 Q^{-1})$ *and the bootstrap estimator* $\hat{\beta}_n^*$ *is consistent in the sense of Lemma 1.*

The proof of Theorem 3 is in the appendix, and the explicit hybrid bootstrap procedure is given in Algorithm 2; it is obtained from Eq. (7) by dividing both sides by $\sqrt{n}$.

OPS Dimitrakakis et al. (2014), Wang et al. (2015) and Foulds et al. (2016) used the idea of sampling from a Bayesian posterior distribution to obtain a differentially private point estimate. One Posterior Sampling (OPS), which releases one sample from the posterior, is a special case of the exponential mechanism, and the corresponding estimator is near-optimal for parametric learning (Wang et al., 2015). The parametric bootstrap applies easily to OPS estimators and produces well calibrated intervals (Figure 2). We expect the asymptotic analysis of Wang et al. (2015) can be adapted to prove asymptotic equivariance, and hence parametric bootstrap consistency, for OPS, but do not give a formal proof.

6 RELATED WORK

A number of prior works have studied private confidence intervals for different models (D'Orazio et al., 2015; Karwa and Vadhan, 2018; Sheffet, 2017; Barrientos et al., 2018; Gaboardi et al., 2019; Brawner and Honaker, 2018; Du et al., 2020). Smith (2011) showed that a broad class private estimators based on subsample & aggregate (Nissim et al., 2007) are asymptotically normal. D'Orazio et al. (2015) proposes an algorithm based on subsample & aggregate to approximate the variance of a private estimator (see Section 3.2). The topics of differentially private hypothesis testing (Vu and Slavkovic, 2009; Solea, 2014; Gaboardi et al., 2016; Couch et al., 2019) and Bayesian inference (Williams and McSherry, 2010; Dimitrakakis et al., 2014; Wang et al., 2015; Foulds et al., 2016; Zhang et al., 2016; Heikkilä et al., 2017; Bernstein and Sheldon, 2018, 2019) are also related, but the specific considerations differ somewhat from confidence interval construction. Finding practical and general-purpose algorithms for differentially private confidence intervals has been identified as an important open problem (King et al., 2020).

The confidence interval approach of Wang et al. (2019) applies to any model fit by empirical risk minimization with objective or output perturbation and is similar to the asymptotic methods we compare to in Section 7. Evans et al. (2019) also give a general-purpose procedure based on subsample & aggregate (S&A) (Nissim et al., 2007) with normal approximations. This method also uses S&A for the point estimates. We compare to a similar variant of S&A in Section 7. Wang et al. (2018) study statistical approximating distributions for differentially private statistics in a general setting.

Brawner and Honaker (2018) use the *non-parametric* bootstrap in a privacy context to estimate standard errors "for free" (at no additional cost beyond mean estimation) in some settings. Other methods most similar to the our work on the parametric bootstrap were discussed in more detail in Sec. 3.2.

Prior methods to construct confidence intervals for private linear regression include Sheffet (2017); Barrientos et al. (2018).

7 EXPERIMENTS

We design synthetic experiments to demonstrate our proposed methods for differentially private confidence interval (CI) estimation.

First, we evaluate the performance of private parametric bootstrap CIs vs a baseline method ("Fisher CIs") based on asymptotic normality of the private estimator and described in more detail below. Performance is measured by how well the coverage of the private CIs matches the nominal coverage. For all models, we also include Fisher CIs of the non-private estimators for comparison.

Second, we demonstrate the bias-correction procedure in Sec. 3.1 in the case of Gaussian and Poisson distributions with data points clamped to different thresholds, which introduces estimation bias. These results show the effectiveness of the parametric bootstrap at approximating and mitigating the bias of differentially private estimates when sensitivity is bounded by forcing data to take values within given bounds.

Third, we compare parametric bootstrap CIs to another general purpose method to construct confidence intervals based on subsample & aggregate (Nissim et al., 2007; Smith, 2011; D'Orazio et al., 2015).

Finally, the appendix includes additional experiments exploring a broader range of settings and performance metrics. These include: multivariate distributions, the effect of varying ϵ, and measurements of the upper- and lower-tail CI failures. We aim for private CIs to be as tight as possible while providing the correct coverage: in the appendix, we also compare the width of our intervals with that of intervals from existing methods for the specific case of Gaussian mean estimation of known variance.

Baseline: "Fisher CIs" As a byproduct of our consistency analysis we also derive asymptotic normal distributions of the private estimators for both expo-

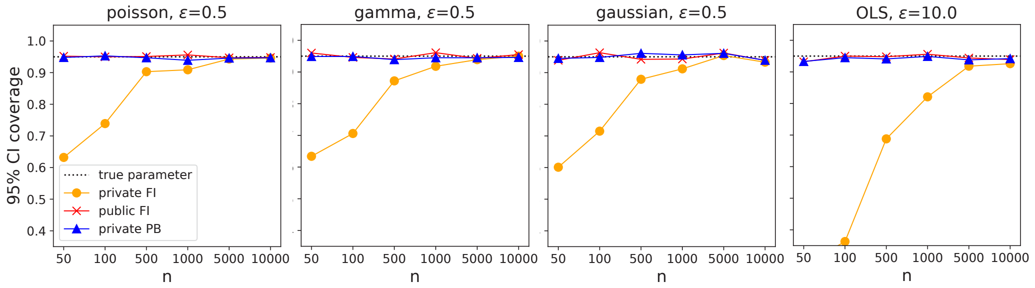


Figure 1: Observed vs nominal coverage of 95% CIs for different distributions for different n. ϵ is set to 0.5 for the exponential family distributions, and to 10 for OLS.

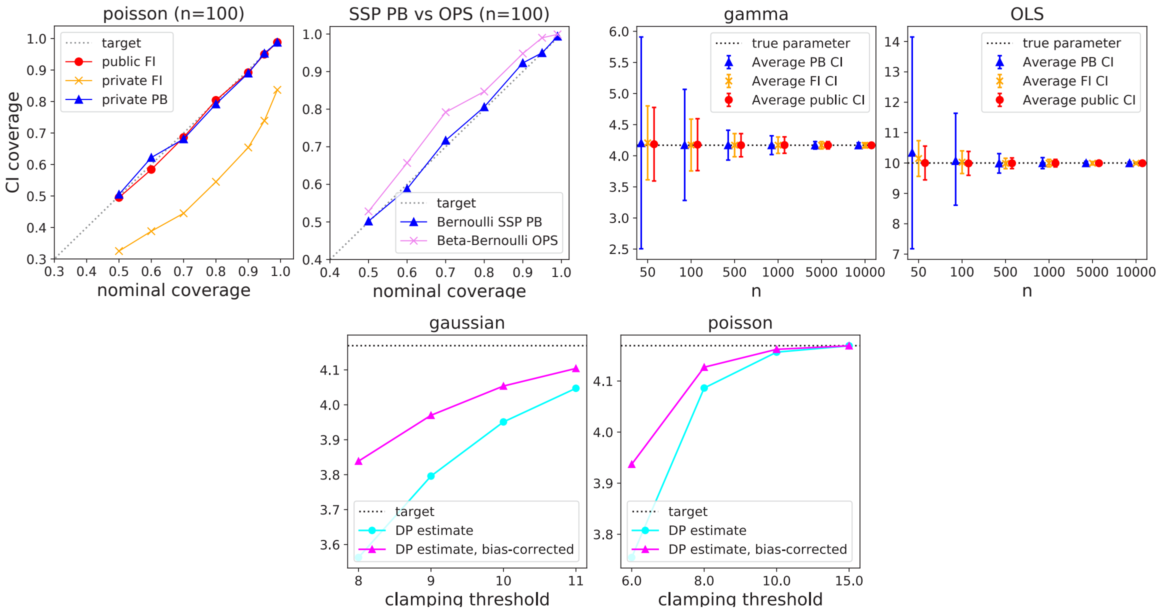


Figure 2: Top. (i) observed vs. nominal coverage for different coverage levels, for a Poisson, n=100, $\epsilon = 0.5$. (ii) same plot, comparing the OPS method (Foulds et al., 2016) and the parametric bootstrap for Bernoulli estimation. (iii) average CI widths for different n for Gamma, with $\epsilon = 0.5$, and (iv) for OLS, with $\epsilon = 10$ (other distributions give qualitatively similar results). Width of the private bootstrap CIs approaches that of the public CIs as $n \to \infty$. Bottom. (i) private and bias-corrected private estimates for a Gaussian clamped at -10 on the left tail and at varying thresholds on the right tail, and (ii) for a Poisson clamped at varying right-tail thresholds.

nential families (Theorem 2) and OLS (Theorem 3). In each case, we obtain a private, consistent estimate $\hat{\sigma}_j^2$ of the jth diagonal entry of the inverse Fisher information matrix of the private estimator $\hat{\theta}_n$, and then construct the confidence interval for θ_j as

$$C_n = \left[\hat{\theta}_{n,j} - z_{\alpha/2}\hat{\sigma}_j, \hat{\theta}_{n,j} + z_{\alpha/2}\hat{\sigma}_j\right], \tag{8}$$

where z_γ is the $(1-\gamma)$-quantile of the standard normal distribution. For exponential families, the Fisher information is estimated via plug-in estimation with the private estimator $\hat{\theta}_n$. For OLS, it is estimated via plugging in private estimates $\hat{Q}_n = \frac{1}{n}X^TX + \frac{1}{n}V$ and $\hat{\sigma}_n^2$, which are both released by the SSP mechanism. For non-private Fisher CIs, we follow similar (and very standard) procedures with non-private estimators.

Exponential families We use synthetic data sets drawn from different exponential family distributions. Given a family, true parameter θ, and data size n, a data set is drawn from P_θ. We release private statistics via SSP with the Laplace mechanism. To simulate the modeler's domain knowledge about the data bounds, we draw a *separate* surrogate data set of size 1000 drawn from the same distribution, compute the data range and use it to bound the width of each released statistic. For private estimation, sampled data is clamped to this range. Private $\hat{\theta}$ is computed from the privately released statistics using SSP-MLE. For the parametric bootstrap CIs, we implement Algorithm 1 and compute

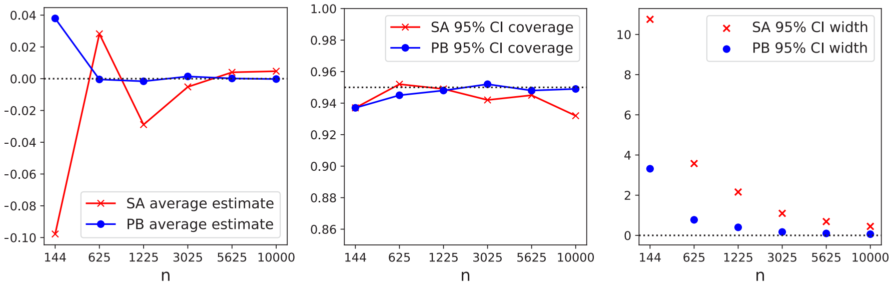


Figure 3: Point estimates (left), 95% CI coverage (center) and average CI width (right) of the S&A method (see Algorithm 3 in appendix) vs parametric bootstrap for the mean of a Gaussian of known variance. Settings: $\epsilon = 0.5$, $\theta = 0$, $\sigma = 1$, $(x_{min}, x_{max}) = (-20, +20)$, $(L_{min}, L_{max}) = (-10, +10)$, $var_{max} = 50$.

Efron's percentile intervals (see Table 1). The output coverage is computed over $T = 1000$ trials.

Results are shown in Figures 1 and 2. For the parametric bootstrap, actual coverage closely matches the nominal coverage, even for very small n. Coverage of private Fisher CIs is too low until n becomes large, due to the fact that it ignores the privacy noise. The bootstrap procedure correctly accounts for the increased uncertainty due to privacy by enlarging the CIs. The width of the bootstrap intervals approaches the width of the baseline Fisher intervals as $n \to \infty$. In the appendix, we show that the coverage failures are balanced between left and right tails and examine the effect of increasing ϵ (which reduces privacy noise and has the same qualitative effect as increasing n).

Linear regression We follow a very similar procedure for OLS. Data is generated with $x_j \sim \text{Unif}([-5, 5])$ for all j and errors are $u_i \sim \text{Unif}[-10, 10]$; bounds on y are passed as inputs ($[-150, 150]$) and assumed known. Observed values of y exceeding the given bounds are dropped. These bounds are also used to compute widths for the sensitivity. Private coefficients are estimated with SSP-OLS and bootstrap CIs are constructed via Efron's percentile method. The results are shown in Fig. 1 and 2.

Bias correction In the case of distributions with infinite support, one option to bound the sensitivity is to clamp or truncate the data to given bounds. These procedures may induce estimation bias. As discussed in Sec 3.1, the parametric bootstrap can be used to approximate this bias and mitigate it. We demonstrate bias correction on the private estimates and CIs of a Poisson and Gaussian distribution where data is clamped on the right tail at different thresholds (Fig. 2).

Comparison with subsample & aggregate We compare the parametric bootstrap CIs with the intervals obtained via a subsample & aggregate (S&A) algorithm. We adapted the S&A procedure of D'Orazio et al. (2015) for privately estimating standard errors to compute confidence intervals; see Algorithm 3 in the appendix. We compare the accuracy of point estimates and 95% CIs for the mean of a Gaussian of known variance. We found that the parametric bootstrap provides more accurate point estimates and better calibrated, tighter CIs than S&A (Figure 3).

8 CONCLUSIONS

The parametric bootstrap is useful and effective to construct consistent differentially private confidence intervals for broad classes of private estimators, including private linear regression, for which we present a novel adaptation to avoid accessing the covariate data many times. The parametric bootstrap yields confidence intervals with good coverage even at modest sample sizes, and tighter than the ones based on subsample & aggregate or other general methods. It can be used with any privacy mechanism, and can help mitigate differentially private estimation bias.

Acknowledgements We would like to thank Aad van der Vaart for the helpful comments regarding Theorem 1. CF is grateful to Niccolò Dalmasso for discussion and feedback on an early draft. This material is based upon work supported by the National Science Foundation under Grant Nos. 1749854.

References

Andrés F. Barrientos, Jerome P. Reiter, Ashwin Machanavajjhalab, and Yan Chen. Differentially private significance tests for regression coefficients. *Journal of Computational and Graphical Statistics*, 2018.

Rudolf Beran. Diagnosing bootstrap success. *Annals of the Institute of Statistical Mathematics*, 49(1):1–24, 1997.

Garrett Bernstein and Daniel R. Sheldon. Differentially private bayesian inference for exponential families. In *Advances in Neural Information Processing Systems*, pages 2919–2929, 2018.

Garrett Bernstein and Daniel R. Sheldon. Differentially private bayesian linear regression. In *Advances in Neural Information Processing Systems*, pages 523–533, 2019.

Thomas Brawner and James Honaker. Bootstrap inference and differential privacy: Standard errors for free. *Unpublished Manuscript*, 2018.

Simon Couch, Zeki Kazan, Kaiyan Shi, Andrew Bray, and Adam Groce. Differentially private nonparametric hypothesis testing. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, pages 737–751, 2019.

Anthony C. Davison. *Statistical models*, volume 11. Cambridge University Press, 2003.

Christos Dimitrakakis, Blaine Nelson, Aikaterini Mitrokotsa, and Benjamin I.P. Rubinstein. Robust and private bayesian inference. In *International Conference on Algorithmic Learning Theory*, pages 291–305. Springer, 2014.

Vito D'Orazio, James Honaker, and Gary King. Differential privacy for social science inference. *Sloan Foundation Economics Research Paper*, (2676160), 2015.

Wenxin Du, Canyon Foot, Monica Moniot, Andrew Bray, and Adam Groce. Differentially private confidence intervals. *arXiv preprint arXiv:2001.02285*, 2020. URL `https://arxiv.org/abs/2001.02285`.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer, 2006.

Bradley Efron. Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7(1):1–26, 01 1979.

Bradley Efron. Nonparametric standard errors and confidence intervals. *Canadian Journal of Statistics*, 9(2):139–158, 1981a.

Bradley Efron. Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. *Biometrika*, 68(3):589–599, 1981b.

Bradley Efron and Trevor Hastie. *Computer age statistical inference*, volume 5. Cambridge University Press, 2016.

Bradley Efron and Robert Tibshirani. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical science*, pages 54–75, 1986.

Georgina Evans, Gary King, Margaret Schwenzfeier, and Abhradeep Thakurta. Statistically valid inferences from privacy protected data. *Working paper*, 2019. URL `https://gking.harvard.edu/files/gking/files/udpd.pdf`.

James Foulds, Joseph Geumlek, Max Welling, and Kamalika Chaudhuri. On the theory and practice of privacy-preserving bayesian data analysis. In *Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence*, UAI'16, page 192–201, 2016.

John Fox. *An R and S-Plus companion to applied regression.* Sage, 2002.

Marco Gaboardi, Hyun Lim, Ryan Rogers, and Salil Vadhan. Differentially private chi-squared hypothesis testing: Goodness of fit and independence testing. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48, pages 2111–2120, 2016.

Marco Gaboardi, Ryan Rogers, and Or Sheffet. Locally private mean estimation: z-test and tight confidence intervals. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *Proceedings of Machine Learning Research*, volume 89, pages 2545–2554, 2019.

Mikko Heikkilä, Eemil Lagerspetz, Samuel Kaski, Kana Shimizu, Sasu Tarkoma, and Antti Honkela. Differentially private bayesian learning on distributed data. In *Advances in neural information processing systems*, pages 3226–3235, 2017.

Vishesh Karwa and Salil Vadhan. Finite sample differentially private confidence intervals. *9th Innovations in Theoretical Computer Science Conference*, 2018.

Gary King et al. The OpenDP White Paper. Technical report, 2020.

Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. Smooth sensitivity and sampling in private data analysis. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 75–84, 2007.

Or Sheffet. Differentially private ordinary least squares. In *Proceedings of the 34th International Conference on Machine Learning*, 2017.

Adam Smith. Privacy-preserving statistical estimation with optimal convergence rates. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 813–822, 2011.

Eftychia Solea. Differentially private hypothesis testing for normal random variables. *Master's thesis, The Pennsylvania State University.* 2014.

Aad W. Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

Duy Vu and Aleksandra Slavkovic. Differential privacy for clinical trial data: Preliminary evaluations. In

2009 IEEE International Conference on Data Mining Workshops, pages 138–143. IEEE, 2009.

Yu-Xiang Wang. Revisiting differentially private linear regression: optimal and adaptive prediction & estimation in unbounded domain, 2018.

Yu-Xiang Wang, Stephen Fienberg, and Alex Smola. Privacy for free: Posterior sampling and stochastic gradient monte carlo. In *Proceedings of the 32nd International Conference on Machine Learning (ICML-15)*, pages 2493–2502, 2015.

Yue Wang, Daniel Kifer, Jaewoo Lee, and Vishesh Karwa. Statistical approximating distributions under differential privacy. *Journal of Privacy and Confidentiality*, 8(1), 2018.

Yue Wang, Daniel Kifer, and Jaewoo Lee. Differentially private confidence intervals for empirical risk minimization. *Journal of Privacy and Confidentiality*, 9 (1), 2019.

Larry Wasserman. *All of nonparametric statistics*. Springer Science & Business Media, 2006.

Oliver Williams and Frank McSherry. Probabilistic inference and differential privacy. In *Advances in Neural Information Processing Systems*, pages 2451–2459, 2010.

Zuhe Zhang, Benjamin I.P. Rubinstein, and Christos Dimitrakakis. On the differential privacy of bayesian inference. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

A PROOF OF CLAIM 1

Claim 1. *Suppose $X = (x_1, \dots, x_n)$ and $f(X) = \sum_{i=1}^{n} g(x_i)$ where g maps data points to $\mathbb{R}^m$. Let $\text{width}(g_j) = \max_x g_j(x) - \min_x g_j(x)$ where x ranges over the data domain. Then $\Delta f \le \sum_{j=1}^{m} \text{width}(g_j)$, which is a constant independent of n.*

Proof. Since X and X' differ in exactly one element and f is additive, $f(X) - f(X') = g(x) - g(x')$ for some elements x, x' in the data domain. The absolute value of the jth output $g_j(x) - g_j(x')$ is bounded by $\text{width}(g_j) = \max_{x^*} g_j(x^*) - \min_{x^*} g_j(x^*)$. The L_1 sensitivity $\|f(X) - f(X')\|_1 = \|g(x) - g(x)'\|_1$ is therefore at most the sum of the widths. □

B PROOFS FOR BOOTSTRAP THEORY

Theorem 1 (Parametric bootstrap consistency)**.** *Suppose $\sqrt{n}(\hat{\theta}_n - \theta) \rightsquigarrow J(\theta)$ and $\hat{\tau}_n$ is asympotitcally equivariant with continuous limiting distribution $H(\theta)$. Then the parametric bootstrap estimator $\hat{\tau}_n^*$ is consistent.*

This theorem is a simplified version of the result of Beran (1997). We give a self-contained proof. See also (Van der Vaart, 2000, Problem 23.5).

Proof. The distribution of $\sqrt{n}(\hat{\tau}_n - \tau)$ under $\mathbb{P}_\theta$ is $H_n(\theta)$, which, by asymptotic equivariance, converges to $H(\theta)$. In the parametric bootstrap, the distribution of $\sqrt{n}(\hat{\tau}_n^* - \hat{\tau}_n)$ conditional on $\hat{\theta}_n = \theta + h_n/\sqrt{n}$ is $H_n(\theta + h_n/\sqrt{n}))$, and, by asymptotic equivariance, $H_n(\theta + h_n/\sqrt{n})) \rightsquigarrow H(\theta)$ if h_n is convergent. Since $H(\theta)$ is continuous this is equivalent to saying that, for all convergent sequences h_n and all t

$$\mathbb{P}_n^* \left(\sqrt{n}(\hat{\tau}_n^* - \hat{\tau}_n) \le t \mid \hat{\theta}_n = \theta + h_n/\sqrt{n} \right) \to F_\theta(t). \tag{9}$$

where F_θ is the CDF of $H(\theta)$. Now, let $\hat{h}_n = \sqrt{n}(\hat{\theta}_n - \theta)$ so that $\hat{\theta}_n = \theta + \hat{h}_n/\sqrt{n}$. By assumption, $\hat{h}_n \rightsquigarrow J(\theta)$ and is therefore $O_P(1)$. Therefore, by Lemma 3, Eq. (9) implies

$$\mathbb{P}_n^* \left(\sqrt{n}(\hat{\tau}_n^* - \hat{\tau}_n) \le t \mid \hat{\theta}_n \right) \to F_\theta(t) \text{ in } \mathbb{P}_\theta\text{-probability}$$

and the result is proved. □

Lemma 3. *Suppose g_n is a sequence of functions such that $g_n(h_n) \to 0$ for any fixed sequence $h_n = O(1)$. Then $g_n(\hat{h}_n) \xrightarrow{P} 0$ for every random sequence $\hat{h}_n = O_P(1)$.*

Proof. Fix $\epsilon, \delta > 0$. We wish to show, for large enough n, that

$$\Pr\left[|g_n(\hat{h}_n)| > \epsilon \right] < \delta.$$

Since $\hat{h}_n$ is $O_P(1)$, there is some M such that, for all n,

$$\Pr\left[\|h_n\| > M \right] < \delta.$$

By our assumption on g_n, there is some N such that $|g_n(h)| < \epsilon$ for all $\|h\| \le M, n > N$ (take the sequence $h_n \equiv h$ for each such h). Then, for $n > N$,

$$\Pr\left[|g_n(\hat{h}_n)| > \epsilon \right] \le \Pr\left[\|h_n\| > M \right] < \delta.$$

□

C PROOFS FOR EXPONENTIAL FAMILIES

Lemma 4. *Let w be any random variable with mean zero and finite variance. For any $r > 0$, $\frac{1}{N^r} w \xrightarrow{P} 0$.*

Proof. The variance of $N^{-r}w$ is equal to $N^{-2r}\operatorname{Var}(w)$, which goes to zero as $N \to \infty$. By Chebyshev's inequality, this implies that $N^{-r}w \xrightarrow{P} 0$. □

Theorem 2. *Let $\hat{\theta}_n$ be the solution to the (SSP-MLE) optimization problem for a sample $x_{1:n}$ from an exponential family model that satisfies the regularity conditions given in Davison (2003, Section 4.4.2). Then $\sqrt{n}(\hat{\theta}_n - \theta)$ is asymptotically equivariant with limiting distribution $\mathcal{N}(0, I(\theta)^{-1})$, where $I(\theta) = \nabla^2 A(\theta)$ is the Fisher information. This implies consistency of the parametric bootstrap estimator $\hat{\theta}_n^*$.*

Following standard practice, we will prove this for the case when θ is scalar; the generalization to vector θ is straightforward but cumbersome. We first state the required (standard) regularity conditions. Let

$$\begin{aligned}\ell_n(\theta) &= \sum_{i=1}^n \ell(\theta; x_i) = \sum_{i=1}^n \big(\log p(x_i;\theta) - \log h(x)\big) \\ &= \theta \sum_{i=1}^n T(x_i) - nA(\theta)\end{aligned}$$

be the log-likelihood of a sample $x_{1:n}$ from the exponential family model using the definition of log-likelihood from Sec. 1. Let $\ell(\theta) = \ell_1(\theta)$ be the log-likelihood of a single $x \sim p(x;\theta)$.

We assume the log-likelihood satisfies the conditions given in the book of Davison (2003, Section 4.4.2). If it does, then we have the following

(F1) $\mathbb{E}_\theta[\ell'(\theta)] = 0$.

(F2) $\operatorname{Var}_\theta[\ell'(\theta)] = -\mathbb{E}_\theta[\ell''(\theta)] = I(\theta)$.

(F3) Given a sequence of estimators $\hat{\theta}_n \xrightarrow{P} \theta$, for all $\tilde{\theta}_n \in [\theta, \hat{\theta}_n]$, $\frac{1}{2\sqrt{n}}\ell_n'''(\tilde{\theta}_n)(\hat{\theta}_n - \theta)^2 \xrightarrow{P} 0$

Facts (F1) and (F2) are well known exponential family properties. Also recall that, for an exponential family,

(F4) $I(\theta) = A''(\theta)$.

(F5) $-\ell''(\theta)$ is *deterministic* and equal to $I(\theta)$.

Proof. Let $\lambda_n(\theta) = f\big(\theta, w + \sum_{i=1}^n T(x_i)\big)$ be the objective of the SSP-MLE optimization problem. We have

$$\begin{aligned}\lambda_n(\theta) &= \theta\Big(w + \sum_{i=1}^n T(x_i)\Big) - nA(\theta) \\ &= \theta w + \theta\sum_{i=1}^n T(x_i) - nA(\theta) \\ &= \theta w + \ell_n(\theta)\end{aligned}$$

where $\ell_n(\theta)$ is the log-likelihood of the true sample. That is, the original objective $\ell_n(\theta)$ is perturbed by the linear function θw to obtain $\lambda_n(\theta)$. The derivatives are therefore related as:

$$\lambda_n'(\theta) = w + \ell_n'(\theta), \tag{10}$$

$$\lambda_n^{(k)}(\theta) = \ell^{(k)}(\theta), \quad k > 1. \tag{11}$$

At the optimum $\hat{\theta}_n$, the first derivative of λ_n is equal to zero. $\ell'(\theta)$ is a sum of i.i.d. terms with mean 0 and variance $I(\theta)$, more specifically:

$$\ell'(\theta) = \sum_{i=1}^n \big(T(x_i) - A'(\theta)\big)$$

For asymptotic equivariance, we are interested in the sequence of estimators $\hat{\theta}_n$ when the "true parameter" follows the sequence $\theta + h_n/\sqrt{n}$. We follow the standard approach of writing the Taylor expansion of the first derivative about the true parameter $\theta + h_n/\sqrt{n}$:

$$0 = w + \ell_n'(\theta + h_n/\sqrt{n}) + \ell_n''(\theta + h_n/\sqrt{n})(\hat{\theta}_n - \theta - h_n/\sqrt{n}) + Z_n \tag{12}$$

where we have used Eqs. (10) and (11) to replace the derivatives of λ on the right-hand side, and $Z_n = \frac{1}{2}\ell_n'''(\tilde{\theta}_n)(\hat{\theta}_n - \theta - h_n/\sqrt{n})^2$ is the second-order Taylor term, with $\tilde{\theta}_n$ some point in the interval $[\theta + h_n/\sqrt{n}, \hat{\theta}_n]$.

Multiply both sides of the equation by $\frac{1}{\sqrt{n}}$ and rearrange to get

$$\begin{aligned}\sqrt{n}(\hat{\theta}_n - \theta - h_n/\sqrt{n}) &= \frac{\frac{1}{\sqrt{n}}w + \frac{1}{\sqrt{n}}\ell_n'(\theta + h_n/\sqrt{n}) + \frac{1}{\sqrt{n}}Z_n}{-\frac{1}{n}\ell_n''(\theta + h_n/\sqrt{n})} \\ &= \frac{\frac{1}{\sqrt{n}}w + \frac{1}{\sqrt{n}}\ell_n'(\theta + h_n/\sqrt{n}) + \frac{1}{\sqrt{n}}Z_n}{I(\theta + h_n/\sqrt{n})}\end{aligned}$$

where in the second equality we used (F5). By Lemma 4, $\frac{1}{\sqrt{n}}w \xrightarrow{P} 0$ and by (F3) $\frac{1}{\sqrt{n}}Z_n \xrightarrow{P} 0$, so, by Slutsky,

$$\sqrt{n}(\hat{\theta}_n - \theta - h_n/\sqrt{n}) \xrightarrow{P} \frac{\frac{1}{\sqrt{n}}\ell'(\theta + \frac{h_n}{\sqrt{n}})}{I(\theta + \frac{h_n}{\sqrt{n}})} = \underbrace{\frac{\ell'(\theta + \frac{h_n}{\sqrt{n}})}{\sqrt{nI(\theta + \frac{h_n}{\sqrt{n}})}}}_{(B)} \frac{1}{\sqrt{I(\theta + \frac{h_n}{\sqrt{n}})}} \tag{13}$$

We know that under the regularity assumptions the Fisher information $I(\cdot)$ is a continuous function and so, since $\theta + \frac{h_n}{\sqrt{n}} \to \theta$, then by continuity ($I(\cdot)$ is deterministic):

$$\left(I\left(\theta + \frac{h_n}{\sqrt{n}}\right)\right)^{-1/2} \to \big(I(\theta)\big)^{-1/2}$$

We now focus on the asymptotic behavior of (B). We will use the fact that in exponential families, $\mathbb{E}_\theta\, T(x) = A'(\theta)$ and $\mathrm{Var}_\theta\, T(x) = A''(\theta) = I(\theta)$. For simplicity of notation, define:

$$\mu_n = A'\left(\theta + \frac{h_n}{\sqrt{n}}\right).$$

Define now the triangular array written in the following notation:

$$\begin{array}{ll} T(x_1) - \mu_1 & x_1 \sim \mathbb{P}_{\theta + \frac{h_1}{1}} \\ T(x_1) - \mu_2, T(x_2) - \mu_2 & x_{1:2} \overset{i.i.d.}{\sim} \mathbb{P}_{\theta + \frac{h_2}{\sqrt{2}}} \\ T(x_1) - \mu_3, T(x_2) - \mu_3, T(x_3) - \mu_3 & x_{1:3} \overset{i.i.d.}{\sim} \mathbb{P}_{\theta + \frac{h_3}{\sqrt{3}}} \\ ... & ... \\ T(x_1) - \mu_n, T(x_2) - \mu_n, ..., T(x_n) - \mu_n & x_{1:n} \overset{i.i.d.}{\sim} \mathbb{P}_{\theta + \frac{h_n}{\sqrt{n}}} \end{array}$$

Let's focus on the n-th row. By construction the sum over the n-th row is $S_n = \sum_{i=1}^n (T(x_i) - \mu_n) = \ell'(\theta + \frac{h_n}{\sqrt{n}})$, so the numerator of (A). Each term in the n-th row has mean zero and:

$$\sigma_n^2 = \sum_{i=1}^{n} \mathrm{Var}[T(x_i) - \mu_n] = nI\left(\theta + \frac{h_n}{\sqrt{n}}\right).$$

If for every $\epsilon > 0$ the following condition holds:

$$\lim_{n\to\infty} \frac{1}{\sigma_n^2} \sum_{i=1}^{n} \mathbb{E}\left[(T(x_i) - \mu_n)^2 \mathbf{1}\left(\left|T(x_i) - \mu_n\right| \geq \epsilon\sigma_n\right)\right] = 0,$$

then $S_n/\sigma_n \to \mathcal{N}(0,1)$ by the Lindeberg-Feller Central Limit Theorem. By plugging in the terms in the condition above we have that:

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{nI\left(\theta + \frac{h_n}{\sqrt{n}}\right)} \sum_{i=1}^{n} \mathbb{E}\left[(T(x_i) - \mu_n)^2 \mathbf{1}\left(\left|T(x_i) - \mu_n\right| \geq \epsilon\sqrt{n}\sqrt{I\left(\theta + \frac{h_n}{\sqrt{n}}\right)}\right)\right] \\
&\quad = \lim_{n\to\infty} I\left(\theta + \frac{h_n}{\sqrt{n}}\right)^{-1} \mathbb{E}\left[(T(x_1) - \mu_n)^2 \mathbf{1}\left(\left|T(x_1) - \mu_n\right| \geq \epsilon\sqrt{n}\sqrt{I\left(\theta + \frac{h_n}{\sqrt{n}}\right)}\right)\right]
\end{aligned}$$

with the equality due to i.i.d. sampling within the row of the triangular array. Note that for any x_1,

$$\lim_{n\to\infty} (T(x_1) - \mu_n)^2 \mathbf{1}\left(\left|T(x_1) - \mu_n\right| \geq \epsilon\sqrt{n}\sqrt{I\left(\theta + \frac{h_n}{\sqrt{n}}\right)}\right) = 0,$$

and that the integrand above is dominated by $(T(x_1) - \mu_n)^2$, which is integrable and finite, since $\mathbb{E}[(T(x_1) - \mu_n)^2]$ is finite. Hence by the dominated convergence theorem, the limit is zero and the condition is satisfied.

Going back to equation (13), we then have that

$$\frac{\frac{1}{\sqrt{n}}\ell'(\theta + \frac{h_n}{\sqrt{n}})}{I(\theta + \frac{h_n}{\sqrt{n}})} \rightsquigarrow \mathcal{N}(0,1) \cdot \big(I(\theta)\big)^{-1/2} = \mathcal{N}(0, I(\theta)^{-1}),$$

which proves that $\sqrt{n}(\hat{\theta}_n - \theta - h_n/\sqrt{n}) \rightsquigarrow \mathcal{N}(0, I(\theta)^{-1})$. Setting $h_n = 0$, it is straightforward to find that $\sqrt{n}(\hat{\theta}_n - \theta) \rightsquigarrow \mathcal{N}(0, I(\theta)^{-1})$. This proves that SSP-MLE is asymptotically equivariant. □

D PROOFS FOR OLS

Theorem 3. *The private estimator satisfies $\sqrt{n}(\hat{\beta}_n - \beta) \rightsquigarrow \mathcal{N}(0, \sigma^2 Q^{-1})$ and the bootstrap estimator $\hat{\beta}_n^*$ is consistent in the sense of Lemma 1.*

Before proving the theorem, we give two lemmas. The first is standard and describes the asymptotics of the dominant term.

Lemma 5. *Under the assumptions of the OLS model in Section 5, $\frac{1}{\sqrt{n}}X^T\mathbf{u} \rightsquigarrow \mathcal{N}(0, \sigma^2 Q)$.*

Proof. Observe that $X^T\mathbf{u} = \sum_{i=1}^{n} \mathbf{x}_i u_i$ is a sum of iid terms, and, using the assumptions of the model in Section 1, the mean and variance of the terms are $\mathbb{E}[\mathbf{x}_i u_i] = \mathbb{E}[\mathbf{x}_i]\,\mathbb{E}[u_i] = 0$ and $\mathrm{Var}(\mathbf{x}_i u_i) = \mathrm{Var}(\mathbf{x}u) = \mathbb{E}[\mathbf{x}uu\mathbf{x}^T] = \mathbb{E}[u^2\mathbf{x}\mathbf{x}^T] = \mathbb{E}[u^2]\,\mathbb{E}[\mathbf{x}\mathbf{x}^T] = \sigma^2 Q$. The result follows from the central limit theorem. □

The theorem involves asymptotic statements about $\hat{\beta}_n$ and $\hat{\beta}^*_n$. The following lemma is a general asymptotic result that will apply to both estimators using Eqs. (6) and (7).

Lemma 6. *Define the function*

$$\mathcal{B}_n\{\check{Q}, \check{\beta}, \check{Z}, \check{V}, \check{w}\} = \left(\check{Q} + \frac{1}{n}\check{V}\right)^{-1} \check{Q}\check{\beta} + \left(\check{Q} + \frac{1}{n}\check{V}\right)^{-1} \left(\frac{1}{\sqrt{n}}\check{Z} + \frac{1}{n}\check{w}\right)$$

and suppose the sequences $Q_n, \beta_n, Z_n, V_n, w_n$ are defined on a common probability space and satisfy

(i) $Z_n \rightsquigarrow \mathcal{N}(0, \sigma^2 Q)$,

(ii) $Q_n \xrightarrow{P} Q$,

(iii) β_n, V_n, w_n *are all* $O_P(1)$.

Then

$$\sqrt{n}\Big(\mathcal{B}_n\{Q_n, \beta_n, Z_n, V_n, w_n\} - \beta_n\Big) \rightsquigarrow \mathcal{N}(0, \sigma^2 Q^{-1}).$$

Proof. Substitute the sequences into $\mathcal{B}_n$ and rearrange to get

$$\sqrt{n}\left(\mathcal{B}_n - \beta_n\right) = \sqrt{n}\left(\left(Q_n + \frac{1}{n}V_n\right)^{-1} Q_n - I\right)\beta_n + \sqrt{n}\left(Q_n + \frac{1}{n}V_n\right)^{-1}\left(\frac{1}{\sqrt{n}}Z_n + \frac{1}{n}w_n\right) \tag{14}$$

First, note that the sequences $\frac{1}{n}V_n$, $\frac{1}{\sqrt{n}}V_n$ and $\frac{1}{\sqrt{n}}w_n$, which will appear below, are all $o_P(1)$, since V_n and w_n are $O_P(1)$.

The first term in Eq. (14) converges to zero in probability. Specifically, a manipulation shows:

$$\begin{aligned}\sqrt{n}\left(\left(Q_n + \frac{1}{n}V_n\right)^{-1} Q_n - I\right)\beta_n &= \left(Q_n + \frac{1}{n}V_n\right)^{-1}\left(-\frac{1}{\sqrt{n}}V_n\right)\beta_n \\ &= O_P(1)o_P(1)O_P(1) \\ &= o_P(1)\end{aligned}$$

For the first factor on the right side, $(Q_n + \frac{1}{n}V_n)^{-1} \xrightarrow{P} Q^{-1}$ (by Slutsky's theorem, since $Q_n \xrightarrow{P} Q$ and $\frac{1}{n}V_n \xrightarrow{P} 0$), and is therefore $O_P(1)$. For the second factor, we argued $-\frac{1}{\sqrt{n}}V_n = o_P(1)$. For the third factor, $\beta_n = O_P(1)$ by assumption.

The second term in Eq. (14) converges in distribution to $\mathcal{N}(0, \sigma^2 Q)$. Rewrite it as

$$\left(Q_n + \frac{1}{n}V_n\right)^{-1}\left(Z_n + \frac{1}{\sqrt{n}}w_n\right).$$

We already argued that $(Q_n + \frac{1}{n}V_n)^{-1} \xrightarrow{P} Q^{-1}$ and $\frac{1}{\sqrt{n}}w_n \xrightarrow{P} 0$. By assumption, $Z_n \rightsquigarrow \mathcal{N}(0, \sigma^2 Q)$. Therefore, by Slutsky's theorem, the entire term converges in distribution to $Q^{-1}\mathcal{N}(0, \sigma^2 Q) = \mathcal{N}(0, \sigma^2 Q^{-1})$. □

We are ready to prove the Theorem 3.

Proof of Theorem 3. We first wish to show that $\sqrt{n}\left(\hat{\beta}_n - \beta\right) \rightsquigarrow \mathcal{N}(0, \sigma^2 Q^{-1})$. To see this, write $\hat{\beta}_n = \mathcal{B}_n\left\{\frac{1}{n}X^T X, \beta, \frac{1}{\sqrt{n}}X^T\mathbf{u}, V, w\right\}$ and apply Lemma 6. It is easy to verify that the sequences satisfy the conditions of the lemma.

Next, we wish to show that the bootstrap estimator is consistent. By Lemma 1, it is enough to show that $\sqrt{n}(\hat{\beta}^*_n - \hat{\beta}_n) \rightsquigarrow \mathcal{N}(0, \sigma^2 Q^{-1})$ conditional on ω in $\mathbb{P}_\theta$-probability, where $\theta = (\beta, \sigma^2, Q)$ and $\mathbb{P}_\theta$ is the common probability space of the data and privacy random variables, represented by ω. The bootstrap variables Z_n^*, V^*, w^* correspond to the inner measure $\mathbb{P}_n^*$. Define $\hat{Q}_n = \frac{1}{n}X^TX + \frac{1}{n}V$. Observe that Eq. (7) is equivalent to

$$\hat{\beta}_n^* = \mathcal{B}_n\{\hat{Q}_n, \hat{\beta}_n, Z_n^*, V^*, w^*\} \text{ under } Z_n^* \sim \mathcal{N}(0, \hat{\sigma}_n^2\hat{Q}_n), V^* \sim F_V, w^* \sim F_w, \tag{15}$$

and $(\hat{Q}_n, \hat{\beta}_n, \hat{\sigma}^2)$ are consistent estimators and hence converge in $\mathbb{P}_\theta$-probability to (Q, β, σ^2).

We can't apply Lemma 6 directly to Eq.(15) because this expression mixes random variables from the outer space $(\hat{Q}_n, \hat{\beta}_n, \hat{\sigma}_n^2)$ and inner space (Z_n^*, V^*, w^*). Instead, we temporarily reason about a *deterministic* sequence $(Q_n, \beta_n, \sigma_n^2) \to (Q, \beta, \sigma^2)$. Then, by Lemma 6 applied to the inner probability space,

$$\begin{gathered}\sqrt{n}\Big(\mathcal{B}_n\{Q_n, \beta_n, Z_n^*, V^*, w^*\} - \beta_n\Big) \rightsquigarrow \mathcal{N}(0, \sigma^2 Q^{-1}) \\ \text{under } Z_n^* \sim \mathcal{N}(0, \sigma_n^2 Q_n), V^* \sim P_V, w^* \sim P_w\end{gathered} \tag{16}$$

The conditions of Lemma 6 can easily be checked. In particular, we have $Z_n^* \rightsquigarrow \mathcal{N}(0, \sigma^2 Q)$.

We can restate the result Eq. (16) as follows: for any fixed sequence $(Q_n, \beta_n, \sigma_n^2) \to (Q, \beta, \sigma^2)$ and all t,

$$\mathbb{P}_n^*\left(\sqrt{n}(\hat{\beta}_n^* - \hat{\beta}_n) \le t \mid \hat{Q}_n = Q_n, \hat{\beta}_n = \beta_n, \hat{\sigma}_n^2 = \sigma_n\right) \to F(t)$$

where F is the CDF of $\mathcal{N}(0, \sigma^2 Q^{-1})$. Lemma 7 below now implies that

$$\mathbb{P}_n^*\left(\sqrt{n}(\hat{\beta}_n^* - \hat{\beta}_n) \le t \mid \hat{Q}_n, \hat{\beta}_n, \hat{\sigma}_n^2\right) \to F(t) \text{ in } \mathbb{P}_\theta\text{-probability},$$

and the theorem is proved. □

Lemma 7. *Let $g_n : \mathbb{R}^k \to \mathbb{R}^\ell$ be a sequence of functions such that $g_n(h_n) \to c$ for any deterministic sequence $h_n \to h$. Then $g_n(\hat{h}_n) \xrightarrow{P} c$ for any random sequence $\hat{h}_n \xrightarrow{P} h$.*

Proof. Take $c = 0$ without loss of generality, let $\|\cdot\|$ be any norm and $d(x, y) = \|x - y\|$. Fix $\epsilon > 0$. It must be the case that

$$\exists\, \delta > 0, n_0 \in \mathbb{N} \text{ such that:} \quad d(h', h) < \delta \implies \|g_n(h')\| < \epsilon, \forall n \ge n_0. \tag{17}$$

Otherwise, we can construct a convergent sequence $h_n \to h$ with $\limsup_{n\to\infty} \|g_n(h_n)\| \ge \epsilon$, which violates the conditions of the Lemma.[7]

Now, suppose $\hat{h}_n \xrightarrow{P} h$. Then, for $n \ge n_0$, by Eq. (17),

$$\Pr\left[\|g_n(\hat{h}_n)\| > \epsilon\right] \le \Pr\left[d(\hat{h}_n, h) > \delta\right].$$

Therefore

$$\lim_{n\to\infty} \Pr\left[\|g_n(\hat{h}_n)\| > \epsilon\right] \le \lim_{n\to\infty} \Pr\left[d(\hat{h}_n, h) > \delta\right] = 0,$$

which proves the result. □

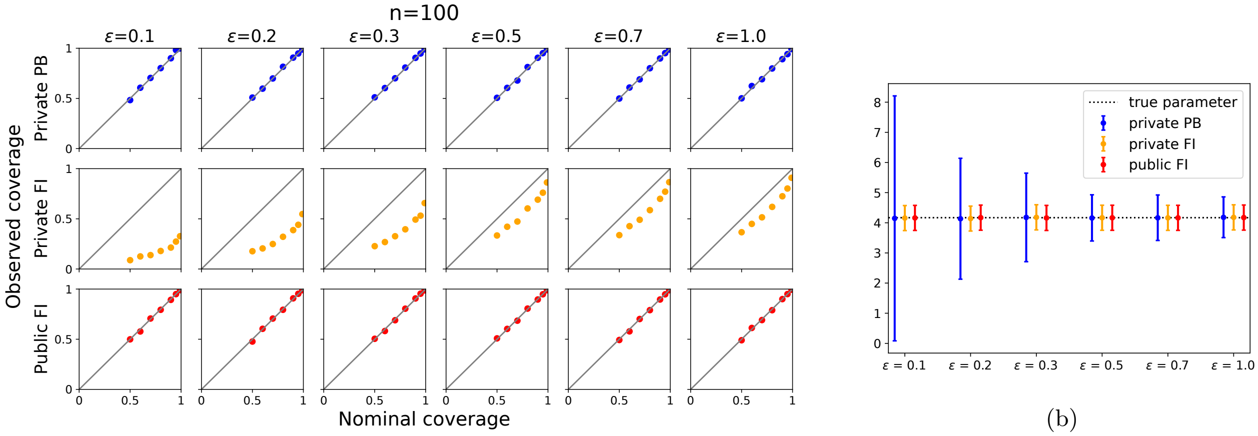
[7] If Eq. (17) is not true, then for all $\delta > 0$ and $n_0 \in \mathbb{N}$, there exists h' such that $d(h', h) < \delta$ and $\|g_n(h')\| \ge \epsilon$ for some $n \ge n_0$. Then we can construct a sequence $h_n \to h$ as follows. Let δ_k be any sequence such that $\delta_k \to 0$. Set $n_0 = 0$, and, for $k \ge 1$, select h' such that $d(h', h) < \delta_k$ and $\|g_{n'}(h')\| \ge \epsilon$ for some $n' \ge n_{k-1} + 1$. Set $h_n = h'$ for all $n \in \{n_{k-1} + 1, \ldots, n'\}$ and let $n_k = n'$. This sequence satisfies $h_n \to h$ but $g_{n_k}(h_{n_k}) \ge \epsilon$ for all k, so it is not true that $g_n(h_n) \to 0$. This contradicts the assumptions of the lemma, so Eq. (17) must be true.

E SUBSAMPLE & AGGREGATE

Algorithm 3 Subsample&Aggregate

```
Input $X, M, x_{min}, x_{max}, L_{min}, L_{max}, var_{max}, \epsilon, \alpha$
procedure SubsampleAndAggregate
    $X_1, ..., X_M \leftarrow$ subsample$(X, M)$
    $L^*_{min}, L^*_{max} \leftarrow \frac{L_{min}}{\sqrt{N/M}}, \frac{L_{max}}{\sqrt{N/M}}$
    $var^*_{max} \leftarrow \frac{var_{max}}{N/M}$
    for $i = 1, ..., M$ do
        $\hat{c}_i \leftarrow$ clamp$(\mathcal{A}(X_i), L^*_{min}, L^*_{max})$
    $\Delta_1 \leftarrow \frac{|L^*_{max} - L^*_{min}|}{M}$
    $\hat{\theta}_{DP} \leftarrow \frac{1}{M} \sum_{i=1}^{M} \hat{c}_i + \text{Lap}(0, \frac{\Delta_1}{\epsilon/2})$
    for $i = 1, ..., M$ do
        for $b = 1, ..., B$ do
            $X_{i,b} \leftarrow$ resample$(X_i, \lfloor \frac{N}{M} \rfloor, \text{replace=True})$
            $\hat{c}_{i,b} \leftarrow$ clamp$(\mathcal{A}(X_{i,b}), L^*_{min}, L^*_{max})$
        $v\hat{a}r_{\hat{c}_i} \leftarrow$ clamp$(\text{Var}(\hat{c}_{i,1:B}), 10^{-6}, var^*_{max})$
    $\Delta_2 \leftarrow var^*_{max}/M$
    $v\hat{a}r_{\hat{c}} \leftarrow \frac{1}{M} \sum_{i=1}^{M} v\hat{a}r_{\hat{c}_i} + \text{Lap}(0, \frac{\Delta_2}{\epsilon/2})$
    $v\hat{a}r_{DP} \leftarrow \frac{1}{M} v\hat{a}r_{\hat{c}} + \text{Var}(\text{Lap}(0, \frac{\Delta_1}{\epsilon/2}))$
    $\text{CI}_{DP} \leftarrow [\hat{\theta}_{DP} - z_{\frac{\alpha}{2}} \sqrt{v\hat{a}r_{DP}}, \hat{\theta}_{DP} + z_{\frac{\alpha}{2}} \sqrt{v\hat{a}r_{DP}}]$
return $\hat{\theta}_{DP}, \text{CI}_{DP}$
```

F ADDITIONAL EXPERIMENTS



(a)

(b)

Figure 4: Effects of varying ϵ for a fixed $n = 100$. We selected a Gamma with inference on the scale parameter. The results are qualitatively equivalent for other distributions. (a) Observed coverage vs. nominal coverage of CIs. Coverage levels: $\{0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99\}$. From top to bottom: (i) differentially private parametric bootstrap; (ii) differentially private Fisher intervals; (iii) non-private Fisher CIs. Private methods use SSP via Laplace mechanism with varying values of ϵ. Note that the effect of increasing ϵ with n fixed is qualitatively similar to the effect of increasing n holding ϵ fixed. (b) Average CIs for the scale parameter for different ϵ. The width of the private bootstrap CIs approaches that of the public CIs as ϵ increases.

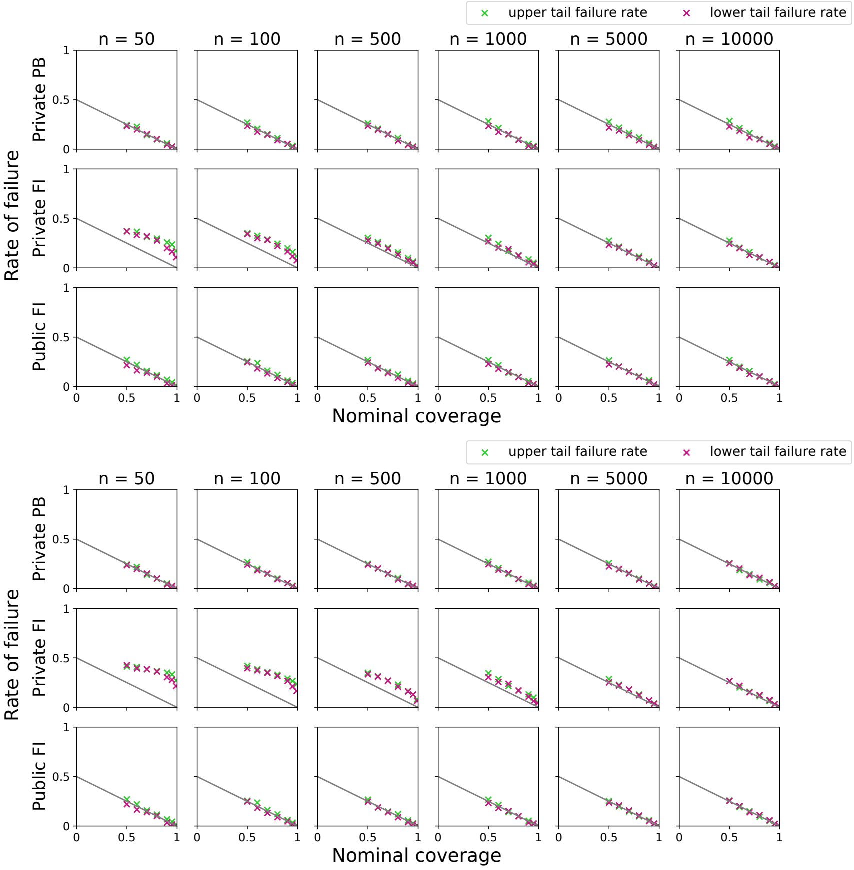
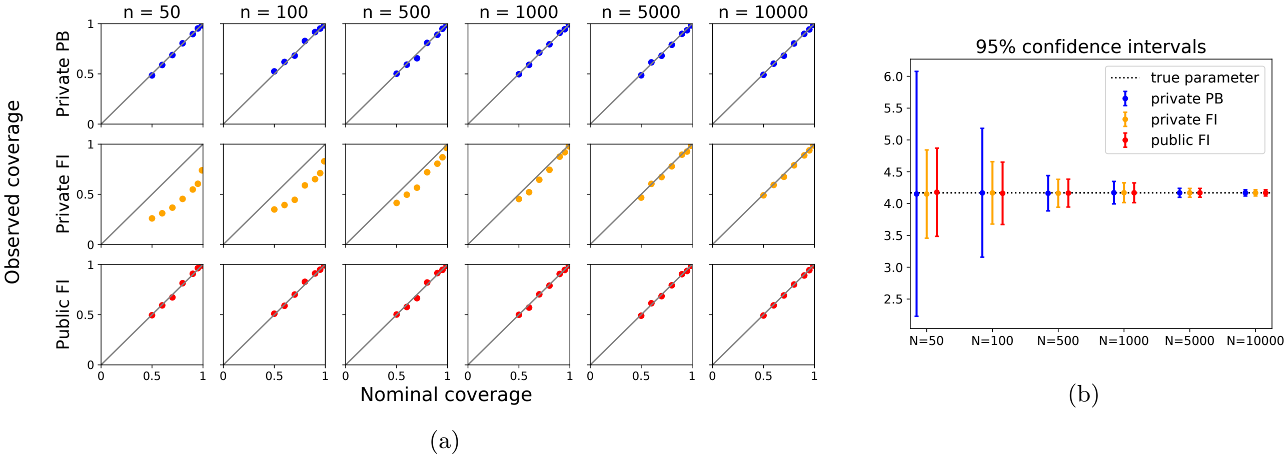


Figure 5: In this Figure, we look at the rate of failure of the confidence intervals on the upper vs lower tail. For each of the two plots, the rows represent (i) differentially private parametric bootstrap; (ii) differentially private Fisher intervals; (iii) non-private Fisher intervals. Top: data range and sensitivity computed as described in Section 7. Clamping the data to a range can introduce a bias if the range is not conservative enough. The bias becomes noticeable for large n, where the interval width is smaller. In our case, where the range is approximated from a data set of size 1000, a small bias becomes noticeable for $n \geq 5000$, where upper-tail failures systematically outnumber lower-tail failures by a small margin. Bottom: same as top plot, with double the range. Increasing the range mitigates the bias.



(a)

(b)

Figure 6: Observed vs nominal coverage (left) and average CI width (right) for a multivariate Gaussian in 5 dimensions, with $\epsilon = 0.5$. We compute CIs for each dimension separately and report results for the first dimension as an example.

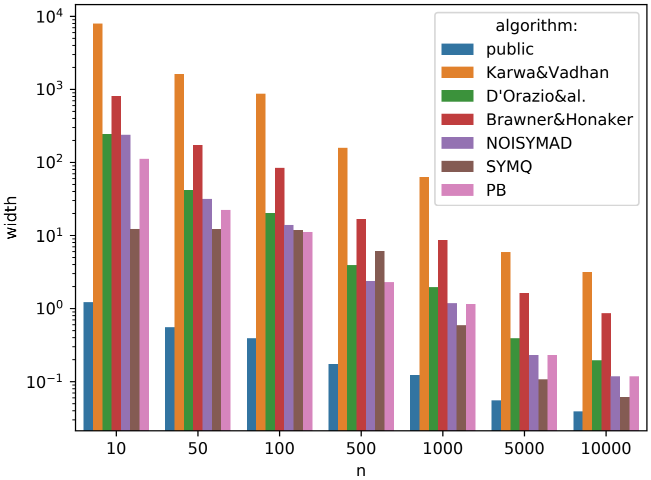


Figure 7: For different algorithms, average width (logscale) of differentially private confidence intervals for the mean of a standard normal, range $[-8, 8]$, $\epsilon = 0.1$, for different n levels. "public" is the confidence interval computed without differential privacy; "Karwa&Vadhan" refers to Karwa and Vadhan (2018); "D'Orazio&al." refers to D'Orazio et al. (2015); "Brawner&Honaker" refers to Brawner and Honaker (2018); "`NOISYMAD`" and "`SYMQ`" are methods from Du et al. (2020), and in particular "`NOISYMAD`" is very similar to our parametric bootstrap method ("`PB`"). We used the publicly available implementation by Du et al. (2020) to reproduce their methods as well as the other prior methods.