Journal of Machine Learning Research 26 (2025) 1-57 Submitted 4/23; Revised 10/25; Published 10/25

Differentially Private Bootstrap: New Privacy Analysis and Inference Strategies

Zhanyu Wang ZHANYU.WANG.PURDUE@GMAIL.COM
Department of Statistics
Purdue University
West Lafayette, IN 47906, USA

Guang Cheng GUANGCHENG@UCLA.EDU
Department of Statistics
University of California, Los Angeles
Los Angeles, CA 90095, USA

Jordan Awan JAA557@PITT.EDU
Department of Statistics
Purdue University
West Lafayette, IN 47906, USA
and
Department of Statistics
University of Pittsburgh
Pittsburgh, PA 15260, USA

Editor: Po-Ling Loh

Abstract

Differentially private (DP) mechanisms protect individual-level information by introducing randomness into the statistical analysis procedure. Despite the availability of numerous DP tools, there remains a lack of general techniques for conducting statistical inference under DP. We examine a DP bootstrap procedure that releases multiple private bootstrap estimates to infer the sampling distribution and construct confidence intervals (CIs). Our privacy analysis presents new results on the privacy cost of a single DP bootstrap estimate, applicable to any DP mechanism, and identifies some misapplications of the bootstrap in the existing literature. For the composition of the DP bootstrap, we present a numerical method to compute the exact privacy cost of releasing multiple DP bootstrap estimates, and using the Gaussian-DP (GDP) framework (Dong et al., 2022), we show that the release of B DP bootstrap estimates from mechanisms satisfying $(\mu/\sqrt{(2-2/\mathrm{e})B})$-GDP asymptotically satisfies μ-GDP as B goes to infinity. Then, we perform private statistical inference by post-processing the DP bootstrap estimates. We prove that our point estimates are consistent, our standard CIs are asymptotically valid, and both enjoy optimal convergence rates. To further improve the finite performance, we use deconvolution with DP bootstrap estimates to accurately infer the sampling distribution. We derive CIs for tasks such as population mean estimation, logistic regression, and quantile regression, and we compare them to existing methods using simulations and real-world experiments on 2016 Canada Census

data. Our private CIs achieve the nominal coverage level and offer the first approach to private inference for quantile regression.

Keywords: Gaussian differential privacy, resampling, distribution-free inference, confidence interval, deconvolution.

1. Introduction

In the big data era, individual privacy protection becomes more critical than ever because personal information is collected and used in many different ways; while the intention of the data collection is usually to improve the user experience or, more generally, to benefit society, there have also been rising concerns about malicious applications of these data. To protect individuals against arbitrary attacks on their data, Dwork et al. (2006) proposed *differential privacy* (DP) which has become the state-of-the-art framework in privacy protection.

DP is a probabilistic framework that measures the level of privacy protection of a mechanism. A mechanism is a *randomized algorithm*, i.e., its output is a realization of a random variable following a distribution determined by the mechanism and its input. A mechanism satisfies DP if changing any individual in the input results in an output with a distribution similar to the original output distribution. Starting from the definition of (ε, δ)-DP (Dwork et al., 2010), there have been many variants of DP definitions serving different needs. For our results, we use f-DP (Dong et al., 2022), a hypothesis-testing perspective of DP formally defined in Definitions 1 and 2, as it is the most informative DP notion satisfying the post-processing property (Dong et al., 2022, Theorem 2). For statistical analysis under DP, a great deal of prior work focused on producing private point estimates of a parameter, e.g., the sample mean, sample median, and the maximum of the data. In contrast, while some prior work aims to quantify the uncertainty of a DP procedure, their techniques are usually restricted to specific settings, and there is still a lack of general-purpose methods (see related work for some notable exceptions.)

One of the most widely used methods to approximate a sampling distribution is the bootstrap method (Efron, 1979), which can be used to quantify the uncertainty of an estimator in many statistical ways, such as by producing a non-parametric confidence interval (CI). Although the bootstrap has been studied very well in Statistics, it is still an open question of how to build and analyze a DP bootstrap for private statistical inference. Brawner and Honaker (2018) were the first to propose and analyze a DP bootstrap procedure and used it to produce a CI based on the private bootstrap outputs. However, a key step in their privacy proof is incorrect,[1] and we show in Section 3 that their stated privacy guarantee does not hold. Furthermore, there is no theoretical analysis on the coverage and width of their CI. Balle et al. (2018) developed the state-of-the-art analysis of resampling for (ε, δ)-DP, both with and without replacement. As a particular case, their results can be used to analyze the privacy guarantees of mechanisms with their input being bootstrap samples. However, Balle et al. (2018) did not consider the cumulative privacy cost of multiple samples of the resampling methods, which restricts the usage of their results on the

1. They use a subsampling result for zCDP (Bun and Steinke, 2016) in Section 6.2, while there is no such result for zCDP as demonstrated on page 75 in Bun et al. (2018).

bootstrap. Moreover, we show that it is necessary to develop a new method for conducting statistical inference with the DP bootstrap samples while Balle et al. (2018) only focused on the privacy analysis.

Our Contributions In this paper, we obtain a tight privacy analysis of a DP bootstrap method and develop inference strategies on the sampling distribution. Specifically, we derive the privacy guarantee of the DP bootstrap, which generalizes the result by Balle et al. (2018) from (ε, δ)-DP to f-DP, and an aspect of our proof strategy applies to any mixture of DP mechanisms where the bootstrap is a special case. Our result also identifies misuses of resampling with replacement in the literature. For the cumulative privacy cost of releasing multiple DP bootstrap estimates, we present a numerical method for the finite composition and derive the asymptotic composition result via the central limit theorem (CLT) for f-DP (Dong et al., 2022). Then, we perform private statistical inference by post-processing the DP bootstrap estimates. While our privacy analysis of the DP bootstrap is applicable to arbitrary mechanisms, our statistical inference results are focused on the Gaussian mechanism. We prove that our point estimates are consistent, our DP CIs are asymptotically valid, and both enjoy optimal convergence rates. To improve finite-sample performance, we use deconvolution on the DP bootstrap results to obtain a private estimate of the non-private sampling distribution and develop CIs. Our simulations show the advantage of our deconvolution method in terms of the coverage of the CIs compared to Du et al. (2020), and the CI width compared to Brawner and Honaker (2018). We also conduct numerical experiments on the 2016 Canada Census Public Use Microdata, which reveals the dependence between individuals' income and shelter cost under DP guarantees by building CIs for the slope parameters of logistic regression and quantile regression. To the best of our knowledge, DP bootstrap is the first tool that can be used to perform valid private statistical inference for parameters in quantile regression.

Related Work For DP CIs, D'Orazio et al. (2015) presented algorithms for releasing DP estimates of causal effects, particularly the difference between means estimators along with their standard errors and CIs. Sheffet (2017) presented a DP estimator for the t-values in ordinary least squares and derived the CI based on the t-values. Karwa and Vadhan (2018) gave DP CIs for the population mean of a normal distribution along with finite sample coverage and lower bounds on the size of the DP CI. Brawner and Honaker (2018) gave the first attempt to use the DP bootstrap to obtain a private CI. Wang et al. (2019) developed algorithms for generating DP CIs for DP estimates from objective and output perturbation mechanisms in the empirical risk minimization framework. Awan and Slavković (2020) developed DP uniformly most powerful hypothesis tests and DP CIs for Bernoulli data. Du et al. (2020) proposed and compared different methods (including NoisyVar, which we discuss in our experimental sections) to build DP CIs for the mean of normally distributed data. Covington et al. (2025) used the bag of little bootstraps (BLB) and the CoinPress algorithm (Biswas et al., 2020) to privately estimate the parameters' sampling distribution through its mean and covariance. Chadha et al. (2024) also used BLB to produce private confidence sets with consistency results and asymptotic rate analysis. Drechsler et al. (2022) proposed several strategies to compute non-parametric DP CIs for the median. Awan and Wang (2024) used a simulation-based method to produce finite-sample CIs and hypothesis tests from DP summary statistics in parametric models.

Other literature involving the idea of bootstrap in DP includes Ferrando et al. (2022) and O'Keefe and Charest (2019). Ferrando et al. (2022) used parametric bootstrap through resampling data from a distribution parametrized by estimated parameters, while we use non-parametric bootstrap through resampling data from the empirical data distribution. O'Keefe and Charest (2019) proposed a relaxed definition of differential privacy based on bootstrap, but they did not use bootstrap to perform statistical inference as we do.

Organization The remainder of this paper is organized as follows. In Section 2, we review the definition of f-DP and results used in our DP bootstrap analysis. In Section 3, we provide our privacy guarantee for DP bootstrap along with a numerical method and a central limit theorem result for the cumulative privacy cost of multiple DP bootstrap outputs. In Section 4, we propose two methods for performing non-parametric statistical inference using the DP bootstrap: one uses the asymptotic distributions of two point estimates, and the other uses deconvolution. In Section 5, we use simulation to show that our CIs have better coverage than NoisyVar (Du et al., 2020), and better width than the method by Brawner and Honaker (2018). In Section 6, we analyze the dependence between market income and shelter cost in Ontario by building DP CIs for the slope parameters of logistic regression and quantile regression with the 2016 Canadian census data set. We compare DP bootstrap with DP-CI-ERM (Wang et al., 2019) only in logistic regression, as the DP-CI-ERM method is inapplicable to quantile regression. In Section 7, we discuss the implications of our work and highlight some directions for future research. Proofs and technical details are deferred to the appendix.

2. Background in Differential Privacy

In this section, we provide existing results related to our DP bootstrap method. First, we discuss the definitions of f-DP and (ε, δ)-DP, and introduce the Gaussian mechanism to guarantee DP. Then we show previous f-DP results on subsampling, composition, and group privacy. Finally, we provide the connection between (ε, δ)-DP and f-DP.

For a set $\mathcal{X}$, a data set $D \in \mathcal{X}^n$ with cardinality $|D| = n$ is a finite collection of elements from $\mathcal{X}$. We define $D_1 \simeq_k D_2$ if $|D_1| = |D_2|$ and they differ in k entries. We call D_1 and D_2 as *neighboring data sets* if $D_1 \simeq_1 D_2$, and we also write it as $D_1 \simeq D_2$. A *mechanism* $\mathcal{M} : \mathcal{X}^n \to \mathcal{Y}$ is a randomized algorithm taking a data set as input and outputting a value in $\mathcal{Y}$. *Differential privacy* measures how much the outputs of the mechanism differ when the inputs are two neighboring data sets.

In this paper, we use the f-DP framework, which measures the difference between two distributions by hypothesis testing. With an observation $\mathcal{M}(D_{\text{in}}) = X$, we consider a hypothesis test between $H_0 : X \sim \mathcal{M}(D)$ and $H_1 : X \sim \mathcal{M}(D')$. Intuitively, the harder this test is, the stronger the privacy guarantee that the mechanism $\mathcal{M}$ provides. Wasserman and Zhou (2010) first used this hypothesis testing framework in DP, and Dong et al. (2022) generalized it to the f-DP framework using the *tradeoff function* which characterizes the difficulty of this test (also known as the receiver operating characteristic (ROC) curve.)

Definition 1 (Tradeoff Function (Dong et al., 2022)) *Consider the hypothesis test $H_0 : X \sim P$ versus $H_1 : X \sim Q$. For any rejection rule $\phi(X)$, we use α_ϕ to denote the type*

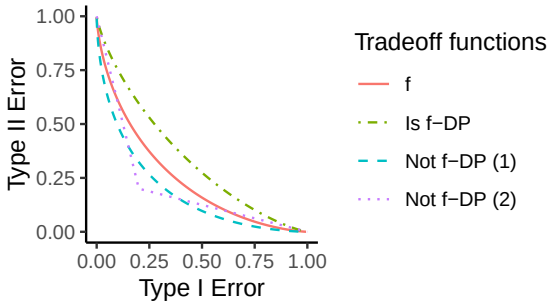

Figure 1: An illustration of tradeoff functions and f-DP (Dong et al., 2022).

I error and β_ϕ to denote the type II error. The tradeoff function $T_{P,Q}(\alpha) : [0,1] \to [0,1]$ is defined to be $T_{P,Q}(\alpha) := \inf_\phi\{\beta_\phi \mid \alpha_\phi \leq \alpha\}$. For a tradeoff function f, we denote its inverse by $f^{-1}(x) := \inf\{\alpha \in [0,1] : f(\alpha) \leq x\}$. We define that f is symmetric if $f = f^{-1}$.

Lower tradeoff functions indicate less privacy since an adversary can distinguish one distribution from the other with smaller type II error at a given type I error. The upper bound of the tradeoff function is $T(\alpha) = 1-\alpha$, since it corresponds to the family of random rejection rules which reject H_0 with probability α for any observation X.

Definition 2 (f-DP (Dong et al., 2022)) *Let f be a tradeoff function. We write $T_{\mathcal{M}(D),\mathcal{M}(D')} \geq f$ if $T_{\mathcal{M}(D),\mathcal{M}(D')}(\alpha) \geq f(\alpha)\ \forall \alpha \in [0,1]$. A mechanism M is said to be f-differentially private (f-DP) if $T_{\mathcal{M}(D),\mathcal{M}(D')} \geq f$ for any data sets D, D' with $D \simeq D'$.*

Intuitively, for a mechanism $\mathcal{M}$ satisfying f-DP where $f = T_{P,Q}$, testing $H_0 : X \sim \mathcal{M}(D)$ versus $H_1 : X \sim \mathcal{M}(D')$ is at least as hard as testing $H_0 : X \sim P$ versus $H_1 : X \sim Q$ when $D \simeq D'$. We visualize the definition of f-DP in Figure 1. Among all tradeoff functions, an important subclass is $G_\mu(\alpha) = T_{\mathcal{N}(0,1),\mathcal{N}(\mu,1)}(\alpha)$. G_μ-DP is also called μ-Gaussian DP (GDP), which is shown to be the limit of many DP procedures under composition by Dong et al. (2022). Another important subclass is $f_{\varepsilon,\delta}(\alpha) = \max\{0, 1-\delta-\mathrm{e}^\varepsilon\alpha, \mathrm{e}^{-\varepsilon}(1-\delta-\alpha)\}$ since $f_{\varepsilon,\delta}$-DP is equivalent to (ε,δ)-DP.

Definition 3 ((ε,δ)-DP: Dwork et al., 2006) *A mechanism $\mathcal{M} : \mathcal{X}^n \to \mathcal{Y}$ is (ε,δ)-differentially private ((ε,δ)-DP) if for any neighboring data sets $D, D' \in \mathcal{X}^n$, and every measurable set $S \subseteq \mathcal{Y}$, the following inequality holds: $\Pr[\mathcal{M}(D) \in S] \leq e^\varepsilon \Pr[\mathcal{M}(D') \in S] + \delta$. The privacy profile (Balle et al., 2018) of $\mathcal{M}$ is a map $\delta(\varepsilon)$ which represents the smallest δ such that $\mathcal{M}$ is (ε,δ)-DP.*

The Gaussian Mechanism Let the *ℓ_2-sensitivity* of a function $g : \mathcal{X}^n \to \mathbb{R}^d$ be $\Delta(g) = \sup_{D_1 \simeq D_2} \|g(D_1) - g(D_2)\|_2$. For any $g : \mathcal{X}^n \to \mathbb{R}^d$, the *Gaussian mechanism* on g adds Gaussian noises to the output of g: $\mathcal{M}_{\mathrm{G}}(D, g, \sigma) = g(D) + \xi$ where $\xi \sim \mathcal{N}(\mu = 0, \Sigma = \sigma^2 I_{d\times d})$. Dong et al. (2022) proved that $\mathcal{M}_{\mathrm{G}}(D, g, \sigma)$ satisfies μ-GDP if $\sigma^2 = (\Delta(g)/\mu)^2$.

Proposition 4 (Subsampling (Dong et al., 2022)) *Let $D \in \mathcal{X}^n$ be a data set and $D' := Sample_{m,n}(D) \in \mathcal{X}^m$ be chosen uniformly at random among all the subsets of D with size $m \leq n$ (sampling without replacement). For $\mathcal{M} : \mathcal{X}^m \to \mathcal{Y}$, $\mathcal{M} \circ Sample_{m,n} : \mathcal{X}^n \to \mathcal{Y}$ is the subsampled mechanism. For $0 \leq p \leq 1$, let $f_p := pf + (1-p)Id$ and $C_p(f) := \min\{f_p, f_p^{-1}\}^{**}$ where $Id(x) = 1 - x$, $f^*(y) = \sup_{-\infty<x<\infty} xy - f(x)$, and $f^{**} = (f^*)^*$. If $\mathcal{M}$ satisfies f-DP and $p = m/n$, then $\mathcal{M} \circ Sample_{m,n}$ satisfies $C_p(f)$-DP.*

Proposition 5 (Composition (Dong et al., 2022)) *The composition property of DP quantifies the cumulative privacy cost of several DP outputs. If $f = T_{P,Q}$ and $g = T_{P',Q'}$, their tensor product is defined as $f \otimes g := T_{P\times P',Q\times Q'}$ where $P \times P'$ is the product measure of P and P', and $Q \times Q'$ is the product measure of Q and Q'. If $\mathcal{M}_i$ satisfies f_i-DP for $i = 1, \ldots, k$, then $\mathcal{M} = (\mathcal{M}_1, \ldots, \mathcal{M}_k)$ satisfies $f_1 \otimes \ldots \otimes f_k$-DP. We define $f^{\otimes k} = f_1 \otimes \ldots \otimes f_k$ if $f_i = f$ for $i = 1, \ldots, k$. For the GDP guarantee, we have $G_{k\mu} = G_\mu^{\otimes k}$.*

Group Privacy While f-DP guarantees protection of the privacy of each individual, it can be generalized to give a privacy guarantee for groups of size k. We say a mechanism $\mathcal{M}$ satisfies f_k-DP for groups of size k if $T(D_1, D_2) \geq f_k$ for all D_1 and D_2 with $D_1 \simeq_k D_2$. If a mechanism is μ-GDP, then it is $k\mu$-GDP for groups of size k (Dong et al., 2022).

At the end of this section, we state the primal-dual result between (ε, δ)-DP and f-DP in Proposition 6. In the next section, we use it to transform the results by Balle et al. (2018) from (ε, δ)-DP to f-DP, and we use it in Proposition 14 for the numerical computation of composition, since Proposition 5 involves high-dimensional testing which is challenging.

Proposition 6 (Primal-Dual View for (ε, δ)-DP and f-DP: Dong et al., 2022) *A mechanism is $(\varepsilon_i, \delta_i)$-DP for all $i \in I$ if and only if it is f-DP with $f = \sup_{i\in I} f_{\varepsilon_i,\delta_i}$ where $f_{\varepsilon_i,\delta_i} = \max\{0, 1 - \delta - e^\varepsilon\alpha, e^{-\varepsilon}(1 - \delta - \alpha)\}$. For a symmetric tradeoff function f, a mechanism is f-DP if and only if it is (ε, δ)-DP for all $\varepsilon \geq 0$ with $\delta(\varepsilon) = 1 + f^*(-\mathrm{e}^\varepsilon)$ where $f^*(y) = \sup_{-\infty<x<\infty} xy - f(x)$, also known as the convex conjugate of f.*

3. Privacy Analysis of Bootstrap Resampling

In this section, we provide novel privacy guarantees for the release of DP bootstrap estimates by proving new results of resampling and composition. First, we apply the privacy guarantee for sampling with replacement by Balle et al. (2018) to bootstrap and convert it from (ε, δ)-DP to f-DP. However, the resulting formula is computationally intractable. We then give a new f-DP bound with a direct proof, which agrees with the result of Balle et al. (2018), but is more transparent and computationally friendly. For private statistical inference based on many DP bootstrap estimates, we provide a numerical method to compute the exact composition result in f-DP, along with a user-friendly asymptotic GDP guarantee. Proofs of the results in this section are provided in Appendix A.

3.1 f-DP Guarantee with One Bootstrap Sample as Input

Bootstrap sampling is denoted by a randomized mapping $\texttt{boot}_n : \mathcal{X}^n \to \mathcal{X}^n$, where $D = (x_1, \ldots, x_n)$ is a database, $\texttt{boot}_n(D) = (x_{i_1}, x_{i_2}, \ldots, x_{i_n})$ is a randomly generated data set

where $i_k \overset{\text{iid}}{\sim} \text{Uniform}(\{1, 2, \ldots, n\})$, $k = 1, 2, \ldots, n$. Let $p_{i,n} = \binom{n}{i}(1/n)^i(1-1/n)^{n-i}$ which is the probability that a given entry of D is included i times in $\texttt{boot}_n(D)$. We also write $\texttt{boot}_n$ and $p_{i,n}$ as $\texttt{boot}$ and p_i respectively when n is known from the context.

We obtain an f-DP result using the primal-dual conversion (Dong et al., 2022) on the result by Balle et al. (2018). The conversion formula and the original result from Balle et al. (2018) are included as Proposition 6 and Theorem 25 in the appendix.

Proposition 7 *For $i = 1, \ldots, n$, let f_i be symmetric tradeoff functions. For $\mathcal{M}$ satisfying group privacy f_i-DP with group size i, $\mathcal{M} \circ \texttt{boot}$ satisfies $f_{\mathcal{M}\circ\texttt{boot}}$-DP where $f_{\mathcal{M}\circ\texttt{boot}} = C_{1-p_0}\left(\left(\sum_{i=1}^n \frac{p_i}{1-p_0} f_i^*\right)^*\right)$, and $C_{1-p_0}(\cdot)$ and $f^*(\cdot)$ are introduced in Proposition 4.*

Although this representation of $f_{\mathcal{M}\circ\texttt{boot}}$ is seemingly simple, it is hard to compute or visualize this tradeoff function because evaluating $f_{\mathcal{M}\circ\texttt{boot}}(\alpha)$ requires solving over n optimization problems for each α. It is also hard to derive composition results from this $f_{\mathcal{M}\circ\texttt{boot}}(\alpha)$, which is crucial for using the bootstrap for statistical inference since multiple bootstrap samples will be used. Due to the intractability of Proposition 7, we prove a new f-DP result for $\mathcal{M} \circ \texttt{boot}$ using the *mixture of tradeoff functions*. The fundamental idea of our proof is to view the bootstrap as a mixture distribution and decompose the overall tradeoff function into a mixture of tradeoff functions where each one is easy to obtain.

Definition 8 (Mixture of Tradeoff Functions) *For $i = 1, 2, \ldots, k$, let f_i be tradeoff functions and $p_i \in (0, 1]$ satisfying $\sum_{i=1}^k p_i = 1$. We write $\underline{f} = (f_1, \ldots, f_k)$ and $\underline{p} = (p_1, \ldots, p_k)$. For a constant $C \in (-\infty, 0]$, define $A_i(C) := \{\alpha_i | C \in \partial f_i(\alpha_i)\}$ where $\partial f_i(\alpha_i)$ is the sub-differential of f_i at α_i, and $A(C) := \{\sum_{i=1}^k p_i\alpha_i | \alpha_i \in A_i(C)\}$. The mixture of tradeoff functions, $\text{mix}(\underline{p}, \underline{f})$, is defined as follows: For each $\alpha \in (0, 1)$, we find C such that $\alpha \in A(C)$ and $\alpha_i \in A_i(C)$ for $i = 1, 2, \ldots, k$ where $\sum_{i=1}^k p_i\alpha_i = \alpha$, then we define $\text{mix}(\underline{p}, \underline{f})(\alpha) = \sum_{i=1}^k p_i f_i(\alpha_i)$; we define $\text{mix}(\underline{p}, \underline{f})(0) = \sum_{i=1}^k p_i f_i(0)$ and $\text{mix}(\underline{p}, \underline{f})(1) = 0$.*

Remark 9 *If for $i = 1, \ldots, k$, f_i has derivative f_i' which is monotonically increasing for every α in $[0, 1]$, we can simplify Definition 8 as $\text{mix}(\underline{p}, \underline{f}) = (\sum_{i=1}^k (p_i f_i \circ (f_i')^{-1})) \circ (\sum_{i=1}^k p_i (f_i')^{-1})^{-1}$ since $\sum_{i=1}^k p_i (f_i')^{-1}$ maps the slope C to the type I error, and $\sum_{i=1}^k (p_i f_i \circ (f_i')^{-1})$ maps the slope C to the type II error.*

Intuitively, as illustrated in Figure 2, by matching the slopes of each tradeoff function f_i, we minimize the overall type II error given a fixed type I error, and $\text{mix}(\underline{p}, \underline{f})$ is well-defined; see Lemma 28 in the appendix. In Theorem 10, we show that $\text{mix}(\underline{p}, \underline{f})$ always gives a lower bound on the privacy cost of an arbitrary mixture mechanism. Note that this general result applies to any mixture of DP mechanisms.

Theorem 10 *For $i = 1, 2, \ldots, k$, let f_i be tradeoff functions and $\mathcal{M}_i : \mathcal{X}^n \to \mathcal{Y}_i$, $\mathcal{Y}_i \subset \mathcal{Y}$ be mechanisms satisfying f_i-DP. Let $\mathcal{M} : \mathcal{X}^n \to \mathcal{Y}$ be a mixture mechanism which randomly selects one mechanism from k mechanisms, $\{\mathcal{M}_i\}_{i=1}^k$, with corresponding probabilities $\{p_i\}_{i=1}^k$ where $\sum_{i=1}^k p_i = 1$, and the output of $\mathcal{M}$ will be the output of $\mathcal{M}_i$ if $\mathcal{M}_i$ is selected. Then $\mathcal{M}$ satisfies f-DP with $f = \text{mix}(\underline{p}, \underline{f})$ where $\underline{f} = (f_1, \ldots, f_k)$ and $\underline{p} = (p_1, \ldots, p_k)$.*

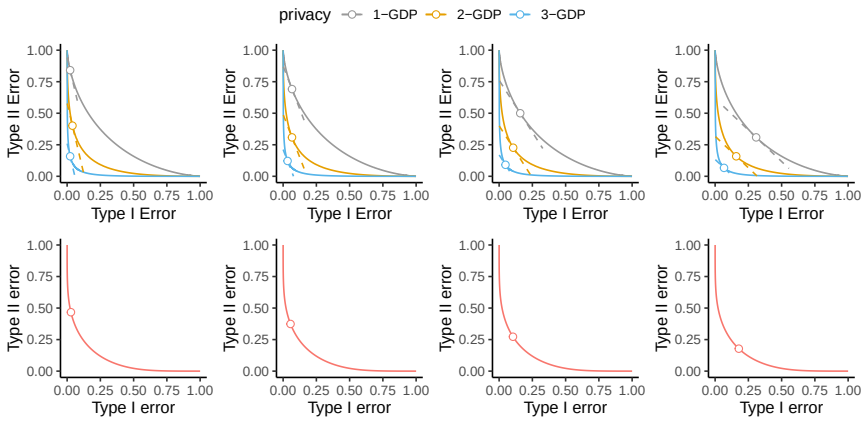

Figure 2: An illustration of the mixture of tradeoff functions. In the top row, the solid curves are the tradeoff functions f_1, f_2, f_3 corresponding to 1-GDP, 2-GDP, and 3-GDP, respectively, and the three dashed lines are the tangent lines with matched slopes. The mixture of $\underline{f} = (f_1, f_2, f_3)$ with corresponding weights $\underline{p} = (p_1, p_2, p_3) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ is $\text{mix}(\underline{p}, \underline{f})$ shown as the curves in the figures on the bottom row. Each circle dot on the mixture curve is the average of the circle dots on f_1, f_2, f_3 weighted by p_1, p_2, p_3. Note that $\text{mix}(\underline{p}, \underline{f})$ is neither $\frac{1}{3}(f_1 + f_2 + f_3)$, nor is it a member of the GDP family.

The proof of Theorem 10 is based on using different rejection rules when a different $\mathcal{M}_i$ is selected to allocate the overall type I error to each rejection rule to minimize the overall type II error. We can combine such rejection rules as one if $\mathcal{Y}_i$ are disjoint; therefore, Theorem 10 is not improvable. For the DP bootstrap setting, since each $\mathcal{M}_i$ maps to the same $\mathcal{Y}_i$, we can leverage this fact to strengthen the privacy guarantee in Theorem 11. We separately consider the case that a given entry of D is *not included* in $\texttt{boot}(D)$ to improve our bound.

Theorem 11 *Let f_i be symmetric tradeoff functions for $i = 1, 2, \ldots, n$. For a mechanism $\mathcal{M}$ satisfying group privacy f_i-DP with group size i, $\mathcal{M} \circ \texttt{boot}$ satisfies $f_{\mathcal{M}\circ\texttt{boot}}$-DP where $f_{\mathcal{M}\circ\texttt{boot}} = C_{1-p_0}(\text{mix}(\underline{p}, \underline{f}))$, $\underline{f} = (f_1, \ldots, f_n)$, and $\underline{p} = (\frac{p_1}{1-p_0}, \ldots, \frac{p_n}{1-p_0})$.*

Remark 12 *If $\mathcal{M}$ satisfies f_1-DP, Dong et al. (2022) proved that $\mathcal{M}$ also satisfies $[1-(1-f_1)^{\circ k}]$-DP for groups of size k where $f^{\circ k}$ denote f composed with itself for k times, e.g., $f^{\circ 3}(x) = f(f(f(x)))$. We can use this result in Theorem 11 if no tighter result is known.*

Remark 13 *The result by Balle et al. (2018) can be derived from our f-DP result in Theorem 11; see Proposition 32 in the appendix. Balle et al. (2018) showed that their results were based on a novel advanced joint convexity property used in (ε, δ)-DP; similarly, our Theorem 10 and 11 reveal the advanced joint convexity property for f-DP, and results related to Theorems 10 and 11 were independently proven by Wang et al. (2023, Lemma 3.1, Lemma 6.3). While Balle et al. (2018) provided specific settings attaining their bound, which also prove the optimality of our result, it is unknown for general settings how to construct rejection rules achieving each pair of the type I error and type II error given by*

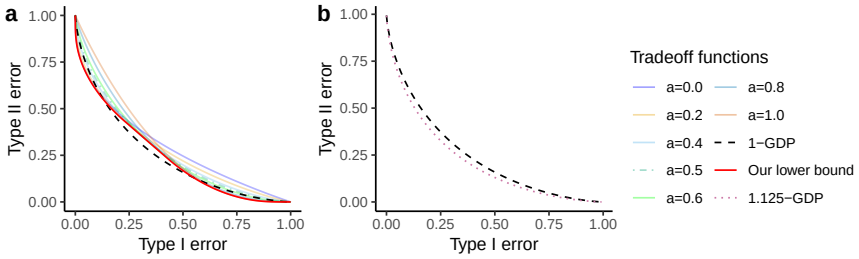

Figure 3: (a) An example showing the relationship between f, $f_{\mathcal{M}\circ\texttt{boot}}$, and tradeoff functions from specific neighboring data set pairs. We consider the Gaussian mechanism $\mathcal{M}$ satisfying 1-GDP shown as the dashed curve. The DP bootstrap mechanism $\mathcal{M}\circ\texttt{boot}$ satisfies $f_{\mathcal{M}\circ\texttt{boot}}$-DP by Theorem 11 shown as the solid opaque curve. The transparent curves are tradeoff functions $T_{\mathcal{M}\circ\texttt{boot}(D_1),\mathcal{M}\circ\texttt{boot}(D_2)}$ where $\mathcal{M}(D) = \frac{1}{n}\sum_{i=1}^{n} x_i + \xi, D = (x_1, x_2, \ldots, x_n), \xi \sim \mathcal{N}(0, \frac{1}{n^2})$, and $D_1 = (a, 0, \ldots, 0)$, $D_2 = (a-1, 0, \ldots, 0)$. The solid curve is tight as a lower bound of the transparent curves. The dashed line (1-GDP) and the dotted dashed line (corresponding to $a = 0.5$) are misused as lower bounds in Brawner and Honaker (2018) and Koskela et al. (2020), respectively (technically Brawner and Honaker (2018) worked in zCDP; see Appendix C for counter-examples to both).
(b) The asymptotic price of using the bootstrap. Running a 1-GDP mechanism on B different bootstrap samples has a similar privacy guarantee to running a $(\sqrt{2-2/\mathrm{e}})$-GDP mechanism on the original data set for B times if B is large enough ($\sqrt{2-2/\mathrm{e}} \approx 1.125$).

our tradeoff function. Therefore, it may be possible to further improve the privacy analysis for the DP bootstrap. Nevertheless, our bound is fairly tight for the specific settings that we show in Figure 3a, and it should suffice for most purposes.

Our result in Theorem 11 is easier to compute when ∂f_i are all known, since for any C, we can immediately obtain its corresponding α and β. For a given α or β, we can use the bisection method to search for C. From the following example and the composition result in Section 3.2, we see that our bound can be easily evaluated for the Gaussian mechanism.

Example: DP Bootstrap using the Gaussian Mechanism In Figure 3a, we show that if $\mathcal{M}$ satisfies 1-GDP, then the DP bootstrap mechanism $\mathcal{M}\circ\texttt{boot}$ satisfies $f_{\mathcal{M}\circ\texttt{boot}}$-DP where $f_{\mathcal{M}\circ\texttt{boot}}$, shown as the solid opaque curve, is our lower bound of all tradeoff functions $T_{\mathcal{M}\circ\texttt{boot}(D_1),\mathcal{M}\circ\texttt{boot}(D_2)}$ when $D_1 \simeq D_2$. We also show the tradeoff functions from specific neighboring data set pairs in Figure 3a (transparent curves) to illustrate misuses in the existing literature: 1) bootstrap cannot be used for free with the same privacy guarantee, i.e., $\mathcal{M}\circ\texttt{boot}$ no longer satisfies 1-GDP, as opposed to Brawner and Honaker (2018); 2) DP bootstrap cannot be analyzed using the privacy loss distribution (PLD) method as in Koskela et al. (2020). Details of this example are in Appendix C.

3.2 Finite and Asymptotic Composition of DP Bootstrap

In this section, we derive composition results for the DP bootstrap because we need many bootstrap estimates to estimate the sampling distribution which is then used to conduct

statistical inference. We first present a numerical method in Proposition 14 to calculate the exact privacy guarantee for the composition of several instances of the DP bootstrap. Then we prove an asymptotic GDP result for the composition of DP bootstrap when using the Gaussian mechanism.

By the definition of f-DP, there exist P and Q such that $f = T(P, Q)$. From the proof of (Dong et al., 2022, Proposition 1), a simple choice is to set the measurable space to be $[0, 1]$, P to be the uniform distribution, and Q to have density $-f'(1-x)$ on $[0, 1)$ and a point mass at 1 with $Q[\{1\}] = 1 - f(0)$. As a tradeoff function, f is convex (see Proposition 29) which means f is almost everywhere differentiable (Rockafellar, 1997, Theorem 25.5), and we set the density of Q to be 0 at the nondifferentiable points of $f(1-x)$. With this P and Q, we use Lemma 5.2 by Zheng et al. (2020) to calculate the privacy profile of the composition of $f_{\mathcal{M}\circ\texttt{boot}}$, which can be translated to f-DP by Proposition 6. We summarize the result in Proposition 14, where we assume that f_i' are strictly increasing to simplify the exposition. When f_i' are not strictly decreasing or f_i are not differentiable at some points, a general but more complex result can be derived from Theorem 11.

Proposition 14 *Let f_i be symmetric tradeoff functions with strictly increasing derivatives f_i' for $i = 1, 2, \ldots, n$. For a mechanism $\mathcal{M}$ satisfying group privacy f_i-DP with group size i, $\mathcal{M} \circ \texttt{boot}$ satisfies $f_{\mathcal{M}\circ\texttt{boot}}$-DP where $f_{\mathcal{M}\circ\texttt{boot}} = T(P, Q)$, P is* Uniform$(0, 1)$ *with density function $p(x) = 1$, and Q has the following density function $q(x)$, where q and x are parameterized by C, and $x^* = \sum_{i=1}^n \frac{p_i}{1-p_0}(f_i')^{-1}(-1)$,*

$$q(x) = \begin{cases} p_0 - (1-p_0)C & 1 \geq x \geq 1 - x^*, \quad C < -1 \\ x & 1 - x^* > x \geq 1 - p_0 - (1-2p_0)x^*, \quad C = -1 \\ 1/(p_0 - (1-p_0)/C) & 1 - p_0 - (1-2p_0)x^* > x \geq 0, \quad 0 > C > -1, \end{cases}$$

$$x = \begin{cases} 1 - \sum_{i=1}^n \dfrac{p_i}{1-p_0}(f_i')^{-1}(C) & C < -1 \\ 1 - \sum_{i=1}^n p_i (f_i')^{-1}(C) - p_0 \left(1 - \sum_{i=1}^n \dfrac{p_i}{1-p_0}(f_i')^{-1}(1/C)\right) & 0 > C > -1. \end{cases}$$

The privacy profile of $f_{\mathcal{M}\circ\texttt{boot}}^{\otimes B}$-DP is $\delta_{\otimes,B}(\varepsilon)$, which is recursively defined by $\delta_{\otimes,B}(\varepsilon) = \int_{\mathbb{R}} \delta_{\otimes,B-1}\left(\varepsilon - \log(\frac{q(x)}{p(x)})\right) q(x)\, \mathrm{d}x$ and $\delta_{\otimes,1}(\varepsilon) = \int_{\mathbb{R}} \max(0, q(x) - \mathrm{e}^{\varepsilon} p(x))\, \mathrm{d}x$.

Remark 15 *The C in Proposition 14 has the same meaning as the C in Definition 8: They are the matched slopes of the tradeoff functions $\{f_i\}_{i=1}^n$.*

Although Proposition 14 can be used to evaluate the privacy guarantee of composition, the iterative calculation is often burdensome for a large number of compositions (Zheng et al., 2020). We prove an asymptotic result to understand the behavior as the number of compositions goes to infinity. For simplicity, we assume the initial mechanism satisfies GDP, but it may be possible to extend our result to mechanisms satisfying f-DP for other tradeoff functions f. While Theorem 16 assumes that the base mechanism is GDP, it need not be

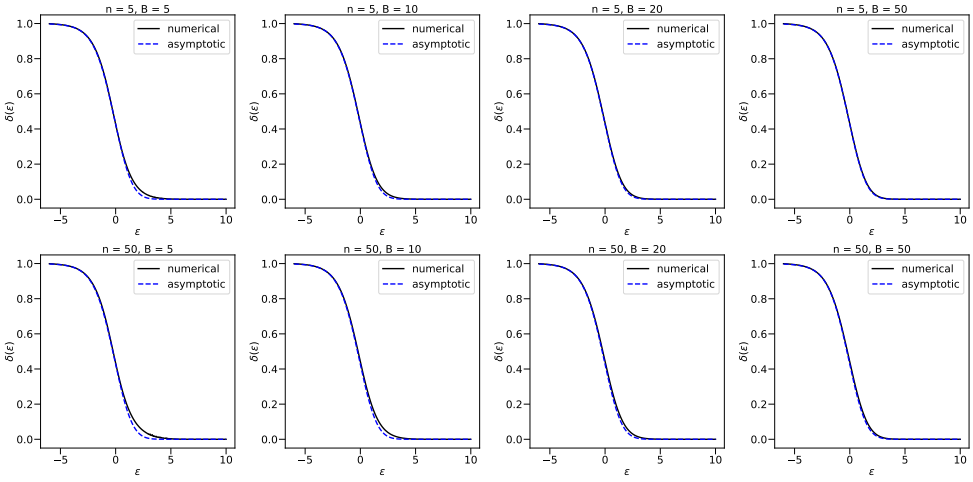

Figure 4: Comparison between the privacy profile $\delta(\varepsilon)$ of composition computed from asymptotics (Theorem 16) and numerical evaluation (Proposition 14). The composition is for releasing B DP bootstrap outputs where the original mechanism is $\sqrt{1/B}$-GDP.

the Gaussian mechanism. For example, Gopi et al. (2022) showed that the exponential mechanism satisfies GDP for many convex empirical risk minimization problems.

Theorem 16 *Let $\mu \in (0, \infty)$ be a given constant, and $\{\mu_B\}_{B=1}^{\infty}$ be a sequence such that $\mu_B \in (0, \infty)$ and $\lim_{B\to\infty} \mu_B \sqrt{(2-2/e)B} \to \mu$. For a mechanism $\mathcal{M}_B$ that satisfies μ_B -GDP, let $f_{\mathcal{M}_B,\texttt{boot}}$ be the f-DP guarantee of $\mathcal{M}_B \circ \texttt{boot}$ from Theorem 11. Then*

$$\lim_{B\to\infty} f_{\mathcal{M}_B,\texttt{boot}}^{\otimes B} = G_{\mu\sqrt{(2-1/n)(1-(1-1/n)^n)/(2-2/e)}} \geq G_\mu.$$

Figure 14 shows the comparison between the privacy profiles of the asymptotic result by Theorem 16 and the numerical result by Proposition 14 where $\mu = \sqrt{2-2/\mathrm{e}}$. When B is small, the asymptotic result tends to overestimate the privacy offered by the DP bootstrap, as the asymptotic $\delta(\varepsilon)$ is lower than the numerical $\delta(\varepsilon)$. However, as B increases, the asymptotic and numerical results converge. Furthermore, for larger values of n, a higher B is required for the asymptotic result to closely match the numerical result. More results are available in AppendixF.

Note that the asymptotic privacy guarantee of the above composition result is the same as running a $(\sqrt{2-2/\mathrm{e}})\mu_B$-GDP mechanism on the *original* data set (not on the bootstrap sample) for B times (Dong et al., 2022). Therefore, the factor $(\sqrt{2-2/\mathrm{e}}) = 1.12438\ldots <$ 1.125 is the price we pay for using the DP bootstrap (see Figure 3b). Although there is a small increase in the privacy cost, the bootstrap samples now contain the randomness from sampling as well as from the privacy mechanism. In the next section, we will see how we can use DP bootstrap estimates to perform statistical inference.

4. Statistical Inference with DP Bootstrap

In this section, we use the insights from the classic methods for statistical inference with bootstrap estimates to develop private statistical inference using DP bootstrap estimates. We analyze the asymptotic distributions of the sample mean and sample variance of the DP bootstrap estimates, and we use them to build asymptotic CIs. We prove that the sample mean of DP bootstrap estimates is a consistent estimator with optimal convergence rate, and our asymptotic CIs has coverage guarantee and optimal average width. Although Brawner and Honaker (2018) have explored similar ideas, their analysis is less rigorous, without proof, and restricted to the population mean inference problem, while we provide solid proofs under more general settings. To further improve the finite-sample performance of using DP bootstrap for private inference, we propose another approach using deconvolution to recover the sampling distribution from the DP bootstrap estimates. The deconvolved distribution satisfies the same f-DP guarantee because of the post-processing property (Dong et al., 2022, Proposition 4), and in our simulation, it is very close to the non-private sampling distribution. In the end of the section, we provide the algorithms of the DP bootstrap estimates, the asymptotic CIs, and the deconvolution method.

Throughout the rest of this paper, we focus on using the Gaussian mechanism in DP bootstrap and its corresponding asymptotic privacy guarantee, μ-GDP. We include a remark when other types of mechanism can be used. Note that the asymptotic privacy guarantee can always be replaced by numerical composition results in Proposition 14 if a strict privacy guarantee is needed.

4.1 Statistical Inference with Efron's Bootstrap

Before explaining how to conduct private statistical inference with the DP bootstrap, we briefly review two basic inference methods with the original Efron's bootstrap estimates (Efron and Tibshirani, 1994), and we develop our private inference methods based on these classic methods. We denote the original data set by $D \in \mathcal{X}^n$ and the estimator for a population parameter θ by $g(D)$. The sampling distribution of $g(D)$ can be estimated from bootstrap estimates, $\{g(D_j)\}_{j=1}^B$ where D_j is the jth bootstrap sample of D.

1. Standard interval: Estimate the sampling distribution using a normal distribution $\mathcal{N}(\theta, \hat{s}_{g,B}^2)$ where $\hat{s}_{g,B}^2 = \frac{1}{B-1}\sum_{j=1}^B (g(D_j) - \hat{m}_{g,B})^2$ and $\hat{m}_{g,B} = \frac{1}{B}\sum_{j=1}^B g(D_j)$. The $(1-\alpha)$-CI of θ is $\left[g(D) + \Phi^{-1}(\frac{\alpha}{2})\hat{s}_{g,B},\ g(D) + \Phi^{-1}(1-\frac{\alpha}{2})\hat{s}_{g,B}\right]$ where $\Phi(x)$ is the CDF (cumulative distribution function) of a standard normal distribution.

2. Percentile interval: Estimate the sampling distribution using the empirical CDF of $\{g(D_j)\}_{j=1}^B$ denoted by $\hat{F}_{g,B}$. The $(1-\alpha)$-CI of θ is $[\hat{F}_{g,B}^{-1}(\frac{\alpha}{2}),\ \hat{F}_{g,B}^{-1}(1-\frac{\alpha}{2})]$.

4.2 Convergence Rate of DP Bootstrap Point Estimates and Confidence Intervals

Similar to the standard interval based on the point estimates, $g(D)$ and $\hat{s}_{g,B}$, we denote $\tilde{m}_{g,B}$ and $\tilde{s}^2_{g,B}$ as the sample mean and sample variance of DP bootstrap estimates, respectively, and use their asymptotic distributions to build valid private CIs.

By analyzing the uncertainty from sampling, bootstrap resampling, and our DP mechanism, we prove Lemma 17 about the mean squared error of $\tilde{m}_{g,B}$ for the population mean inference, and the asymptotic distributions of $\tilde{m}_{g,B}$ and $\tilde{s}^2_{g,B}$ assuming finite support of D.

Lemma 17 (Consistency & Asymptotic Distribution) *Let $D = \{x_1, \cdots, x_n\} \in \mathcal{X}^n$ be a data set where $x_i \overset{\text{iid}}{\sim} F$ are random variables and θ is a parameter of interest determined by F, we let $g(D)$ be an estimator of θ, D_j be the jth bootstrap sample of D, and $\{\tilde{g}(D_j)\}_{j=1}^B$ be the DP bootstrap estimates using the Gaussian mechanism: $\tilde{g}(D_j) = g(D_j) + \xi_j$, $\xi_j \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_e^2)$. Let $\tilde{m}_{g,B} = \frac{1}{B}\sum_{j=1}^B \tilde{g}(D_j)$ and $\tilde{s}^2_{g,B} = \frac{1}{B-1}\sum_{j=1}^B (\tilde{g}(D_j) - \tilde{m}_{g,B})^2$.*

1. *For the population mean inference where $g(D) = \frac{1}{n}\sum_{i=1}^n x_i$ and $\theta = \mathbb{E}[x_i]$, we have $\mathbb{E}[\tilde{m}_{g,B} - \theta]^2 = \frac{1}{n}[(1 + \frac{1}{B} - \frac{1}{nB})\mathbb{E}(x_i - \theta)^2] + \frac{\sigma_e^2}{B}$.*
2. *Let Π denote the set of all distributions with finite support on $\mathcal{X}$. Assume that $F = F(\eta, s, d) = \sum_{k=1}^d \eta_k \delta(s_k) \in \Pi$ where $s = \{s_1, \cdots, s_d\} \subseteq \mathcal{X}$ is the support of F, $\delta(s_j)$ denotes the point mass at s_j with measure 1, the distribution parameter $\eta = (\eta_1, \cdots, \eta_{d-1}) \in H := \{\eta \in (0,1)^{d-1} \big| \sum_{k=1}^{d-1} \eta_k < 1\}$, and $\eta_d = 1 - \sum_{k=1}^{d-1} \eta_k$. The empirical distribution of D is $\hat{F}_n = F(\hat{\eta}, s, d)$ where $\hat{\eta}_k = \frac{1}{n}\sum_{i=1}^n I(x_i = s_k)$. For a continuously differentiable function $T : H \to \mathbb{R}$ such that $\theta = T(\eta)$ and $g(D) = T(\hat{\eta})$, we let $\sigma_g^2 = \frac{1}{n}\left(\frac{\partial T}{\partial \eta}\right)^\intercal \Sigma \frac{\partial T}{\partial \eta}$ and $\Sigma = \text{diag}(\eta) - \eta\eta^\intercal$, and we have*

$$\frac{\tilde{m}_{g,B} - \theta}{\sqrt{\sigma_g^2 + \frac{1}{B}(\sigma_g^2 + \sigma_e^2)}} \xrightarrow{d} \mathcal{N}(0,1) \quad \text{and} \quad (B-1)\frac{\tilde{s}^2_{g,B}}{\sigma_g^2 + \sigma_e^2} \xrightarrow{d} \chi^2_{B-1}.$$

Remark 18 *From Theorem 16, using $\sigma_e^2 = \frac{(\Delta(g))^2(2-2/e)B}{\mu^2}$ where $\Delta(g)$ is the ℓ_2-sensitivity of g on $\mathcal{X}^n$, we have that $(\tilde{m}_{g,B}, \tilde{s}^2_{g,B})$ satisfy μ-GDP asymptotically as $B \to \infty$. For bounded $\mathcal{X}$, we have $\Delta(g) = O(\frac{1}{n})$, therefore, part 1 of Lemma 17 has the rate $\mathbb{E}[\tilde{m}_{g,B} - \theta]^2 = O\left(\frac{1}{n} + \frac{1}{n^2\mu^2}\right)$ as $\frac{\sigma_e^2}{B} = O(\frac{1}{n^2\mu^2})$ and $(1 + \frac{1}{B} - \frac{1}{nB}) = O(1)$ for $B \geq 1$. This rate means that $\tilde{m}_{g,B}$ is a consistent estimate of θ, and it matches the minimax rate $\Omega\left(\frac{1}{n} + \frac{\log(1/\delta)}{n^2\varepsilon^2}\right)$ for the population mean estimation under (ε, δ)-DP (Cai et al., 2021, Theorem 3.1) since the Gaussian mechanism for μ-GDP satisfies (ε, δ)-DP when $\mu^2 \propto \frac{\varepsilon^2}{\log(1/\delta)}$ (Dwork and Roth, 2014).*

Remark 19 *Part 2 of Lemma 17 follows the proof by Beran (1997, Theorem 2.2), where they used the finite support assumption to prove that the non-parametric bootstrap distribu-*

tion converges to the sampling distribution. Although our proof is limited to discrete distributions, this should not cause any problems for real-world data sets, since all measurements are taken with finite precision.

In the remainder of this section, we build asymptotically valid private CIs for θ using the asymptotic distributions of $\tilde{m}_{g,B}$ and $\tilde{s}^2_{g,B}$ (part 2 of Lemma 17) in Theorem 20. The construction of this CI was inspired by the repro sample method (Xie and Wang, 2022); for the reader's convenience, we provide a simple and self-contained proof of Theorem 20 in the appendix, which does not require familiarity with the work of Xie and Wang (2022).

Theorem 20 (Asymptotically Valid CI) *Assume that* $\frac{X_n-\theta}{\sqrt{\sigma_g^2+\frac{1}{B}(\sigma_g^2+\sigma_e^2)}} \xrightarrow{d} \mathcal{N}(0,1)$ *and* $\frac{(B-1)Y_n}{\sigma_g^2+\sigma_e^2} \xrightarrow{d} \chi^2_{B-1}$*, respectively, as* $n \to \infty$*. Given* $\alpha \in [0,1]$*, for any* $0 < \omega < \alpha$*, we have*

$$\lim_{n\to\infty} \mathbb{P}\left(\theta \in [X_n - \hat{r}_n,\ X_n + \hat{r}_n]\right) \geq 1-\alpha,$$

where $\hat{r}_n = \Phi^{-1}(1-\frac{\omega}{2})\sqrt{\hat{\sigma}_g^2 + \frac{1}{B}(\hat{\sigma}_g^2+\sigma_e^2)}$*,* $\hat{\sigma}_g^2 = \max\left(0, \frac{B-1}{c}Y_n - \sigma_e^2\right)$*, and* c *is the* $(\alpha-\omega)$ *quantile of the* χ^2_{B-1} *distribution. That is,* $[X_n - \hat{r}_n,\ X_n + \hat{r}_n]$ *is an asymptotically valid CI for* θ *with level* $(1-\alpha)$*.*

When used in finite settings in part 2 of Lemma 17 where $X_n := \tilde{m}_{g,B}$ and $Y_n := \tilde{s}^2_{g,B}$, the private CIs in Theorem 20 are usually conservative. Note that the asymptotic coverage is for $n \to \infty$ with finite B. If we use a plugin estimator for σ_g^2 instead, i.e., $\acute{\sigma}_g^2 := Y_n - \sigma_e^2$ based on $\mathbb{E}[Y_n] = \sigma_g^2 + \sigma_e^2$, the CI will not have enough coverage when B is small since $\acute{\sigma}_g^2$ can be negative. Theorem 20 uses an upper bound where $\hat{\sigma}_g^2$ ensures good coverage of CIs for any fixed B.

The CI width $\hat{r}_n$ is a random variable depending on α, ω, B, σ_e^2, and σ_g^2, where α, B, and σ_e^2 are known, σ_g^2 is unknown, and ω is tunable while it must be chosen before observing (X_n, Y_n) to ensure the validity of the coverage guarantee.

Remark 21 *Brawner and Honaker (2018) built confidence intervals using a result similar to part 2 of Lemma 17 without providing a concrete proof. In their Equation (34), they estimated* σ_g^2 *by* $\hat{\sigma}_g^2 = \tilde{s}^2_{g,B} - \frac{c_{\alpha'}}{B-1}\sigma_e^2$ *where* $c_{\alpha'}$ *is a tunable parameter, and in Equation (35), they estimated the variance of* $\tilde{m}_{g,B}$ *by* $\hat{\sigma}^2 = \hat{\sigma}_g^2 + \frac{\sigma_e^2}{B}$*. Then they showed that the coverage of the naïve private version of the standard interval* $[\tilde{m}_{g,B} + \Phi^{-1}(\frac{\alpha}{2})\hat{\sigma},\ \tilde{m}_{g,B} + \Phi^{-1}(1-\frac{\alpha}{2})\hat{\sigma}]$ *is lower than the nominal confidence level when* $c_{\alpha'} = B-1$*, i.e.,* $\hat{\sigma}_g^2$ *is an unbiased estimator of* σ_g^2 *(Brawner and Honaker, 2018, Figure 7b). They noted that using an unbiased estimate could underestimate the uncertainty of* $\tilde{m}_{g,B}$ *(the worst case is that* $\hat{\sigma}_g^2$ *is negative), causing undercoverage. Therefore, they solved this problem using an ad hoc "conservative" (larger) estimate* $\hat{\sigma}_g^2$ *of* σ_g^2 *with a smaller* $c_{\alpha'}$*. However, they did not prove any guarantee of coverage for this approach. In contrast, we prove Theorem 20 on the coverage of our approach. In Section 5.1, we use simulation to compare our method with theirs.*

Proposition 22 shows that under certain choices of B and ω, the width of the CI in Theorem 20 is asymptotically optimal compared to the width derived from a normal approximation with variance computed by the Cramér-Rao lower bound.

Proposition 22 (Optimality) *Under the assumption of part 2 of Lemma 17, we use the CI in Theorem 20 where* $\sigma_e^2 = \frac{(2-2/e)B}{n^2\mu^2}$. *When* $n \to \infty$, *we assume* $B \to \infty$, $B = o(n^2)$, $\alpha - \omega = o(1)$, *and* $B^{-\frac{1}{2}} = o(\alpha - \omega)$, *then* $\frac{\hat{r}_n}{r_{\text{opt}}} \xrightarrow{p} 1$ *where* $r_{\text{opt}} = \Phi^{-1}(1 - \frac{\alpha}{2})\sigma_g$.

4.3 Deconvolution for Estimating the Sampling Distribution

The method developed in Section 4.2 is asymptotic and does not offer insight into the whole sampling distribution as the CI depends only on the sample mean and sample variance of the bootstrap estimates. In this section, we use deconvolution to build a non-parametric and non-asymptotic estimate for the sampling distribution of $g(D)$, and we derive CIs based on the estimated sampling distribution, which is similar to percentile intervals.

Deconvolution is the main tool to solve the contaminated measurement problem, i.e., we want to estimate the distribution of $\{X_i\}_{i=1}^B$ while our measurement is $\{Y_i\}_{i=1}^B$ where $Y_i = X_i + e_i$ and e_i is the measurement error. This is exactly the relationship between $g(D_j)$ and $\tilde{g}(D_j)$ since $\tilde{g}(D_j) = g(D_j) + \xi_j$ and $\xi_j \overset{\text{iid}}{\sim} \mathcal{N}(0, \frac{(\Delta(g))^2(2-2/\text{e})B}{\mu^2})$ in Lemma 17. As DP allows for all details of the privacy mechanism to be publicly revealed (except the private data set), the distribution of the added noise can be incorporated into our post-processing without raising any privacy concerns.

The deconvolution method is more flexible as it does not require a normal approximation to build CIs compared to the point estimates in Section 4.2. It also circumvents the problem caused by the possibly negative estimate of the variance of the sampling distribution. However, it is difficult to analyze the convergence rate of deconvolution in DP bootstrap with respect to B and n because the distribution of added noise e_i flattens when B increases, and the distribution to be recovered is from bootstrap estimates and varies for different n. At the end of this section, we propose a signal-noise-ratio (SNR) measure as a rule of thumb for the choice of B given n for use in deconvolution.

We compared multiple numerical deconvolution methods through preliminary simulations with different settings on n and B. Among different deconvolution methods, we choose to use `deconvolveR` (Efron, 2016) since it performs the best in our settings without tuning its hyper-parameters. We briefly summarize this method as follows. For the model $Y = X + e$, `deconvolveR` assumes that Y and X are distributed discretely with the sizes of their supports $|\mathcal{Y}| = k$ and $|\mathcal{X}| = m$. Then it models the distribution of X by $f(\alpha) = \text{e}^{Q\alpha}/c(\alpha)$ where Q is an $m \times p$ structure matrix with values from the natural spline basis of order p, $ns(\mathcal{X}, p)$, and α is the unknown p-dimensional parameter vector; $c(\alpha)$ is the divisor necessary to make f sum to 1. The estimation of the distribution of X is obtained through the estimation of α: We estimate α by maximizing a penalized log-likelihood $m(\alpha) = l(Y;\alpha) - s(\alpha)$ with respect to α where $s(\alpha) = c_0\|\alpha\|_2$ is the penalty term with a tunable parameter c_0 (default 1), and $l(Y;\alpha)$ is the log-likelihood function of Y derived from $f(\alpha)$ and the known distribution of e.

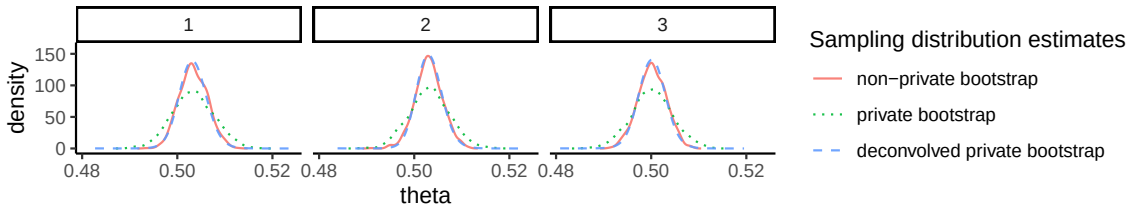

Figure 5: Comparison of the density functions from the non-private bootstrap result, the DP bootstrap result, and the deconvolved DP bootstrap result.

The output of deconvolution is an estimate of the sampling distribution of non-private bootstrap estimates $g(D_j)$. Then we construct $(1-\alpha)$-CI by the $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ percentiles of the sampling distribution estimate.

Example 1 *Let* $n = 10000$, $B = 1000$, $D = (x_1, \ldots, x_n)$, $x_i \in [0,1] = \mathcal{X}$, $g(D) = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\{D_j\}_{j=1}^{B}$ *are* B *bootstrap samples of* D, $\tilde{g}(D_j) = g(D_j) + \xi_j$ *where* $\xi_j \sim \mathcal{N}(0, \sigma_e^2 = \frac{B}{n^2})$. *We know that* $\{\tilde{g}(D_j)\}_{j=1}^{B}$ *asymptotically satisfies* $(\sqrt{2-2/e})$*-GDP. In Figure 5, we generate* D *by* $x_i \overset{\text{iid}}{\sim} \text{Unif}(0,1)$ *for three different replicates, and the 'non-private bootstrap' and 'private bootstrap' density functions are computed from* $\{g(D_j)\}_{j=1}^{B}$ *and* $\{\tilde{g}(D_j)\}_{j=1}^{B}$ *respectively. Since the distribution of 'private bootstrap' is much flatter than the 'non-private bootstrap' due to the additional noises added for privacy, percentile intervals directly from the 'private bootstrap' would be much more conservative (wider). In contrast, the 'deconvolved private bootstrap' has a distribution similar to the 'non-private bootstrap'.*

Remark 23 *Although the deconvolution method is specifically designed for additive noise mechanisms, we can potentially approximate more general mechanisms as additive ones asymptotically, e.g., exponential mechanism (Awan et al., 2019; Reimherr and Awan, 2019). Note that the covariance in the asymptotic distribution may depend on the confidential data set and would have to be estimated.*

Remark 24 *As it is difficult to analyze how to choose* B *theoretically, we provide a general rule of thumb* $B \in \Theta(\mu^2 n)$ *and the reason is as follows. We define the oracle signal-noise ratio (SNR) as the ratio between the variance of the sampling distribution and the variance of the noise added for DP: 1) If we have enough privacy budget, we choose the largest* B *satisfying* $\text{SNR} \geq 1$*; 2) If we only have a very limited privacy budget, e.g.,* $\text{SNR} \leq 1$ *for any* $B \geq 10$*, we choose the smallest* B *such that* $B \geq 2/\alpha$ *for* $1-\alpha$ *CIs. Note that for the case of the population mean inference in Section 4.2, the expectation of the variance of non-private bootstrap estimates is* $\frac{(n-1)\sigma_x^2}{n^2}$*, and for the DP bootstrap, the added noise* $\xi_j \sim \mathcal{N}(0, \frac{(2-2/e)B}{\mu^2 n^2})$*; therefore,* $\text{SNR} \geq 1$ *suggests* $B \in \Theta(\mu^2 n)$ *which is consistent with our rate analysis* $B = o(n^2)$ *in Proposition 22 for our asymptotic CIs. More empirical results are available in the Appendix E.*

4.4 DP Bootstrap Algorithm for the Gaussian Mechanism

In this section, we summarize our DP bootstrap method for the Gaussian mechanism in Algorithm 1, our method to build asymptotic CIs in Algorithm 2, and the deconvolution method to obtain the DP sampling distribution estimate and non-parametric CIs in Algorithm 3.

Note that our privacy analysis in this paper (for Algorithm 1) is valid for $g : \mathcal{X}^n \to \mathbb{R}^d$ for any $d \in \{1, 2, 3, \ldots\}$, but the `deconvolveR` used in Algorithm 3 can only be applied when $d = 1$. For $d \in \{2, 3, \ldots\}$, one may try other deconvolution methods in Algorithm 3 or use our current procedure in each dimension separately.

Algorithm 1 `DP_bootstrap_estimates` (with Gaussian mechanism)

1: **Input** data set $D \in \mathcal{X}^n$, statistic $g : \mathcal{X}^n \to \mathbb{R}$ with ℓ_2 sensitivity $\Delta(g)$, privacy guarantee μ-GDP, number of bootstrap samples B.
2: **for** $j = 1, \ldots, B$ **do**
3: Obtain bootstrap sample (i.e., sample with replacement) $D_j \in \mathcal{X}^n$ from D.
4: DP bootstrap statistic: $\tilde{g}_j = g(D_j) + \xi_j$, $\xi_j \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_e^2)$, $\sigma_e = (\sqrt{(2 - 2/\mathrm{e})B})\Delta(g)/\mu$.
5: **end for**
6: **Return** $(\tilde{g}_1, \tilde{g}_2, \ldots, \tilde{g}_B, \sigma_e^2)$ which approximately satisfies μ-GDP.

Algorithm 2 `DP_bootstrap_Asymptotic_CI`

1: **Input** data set $D \in \mathcal{X}^n$, statistic $g : \mathcal{X}^n \to \mathbb{R}$, privacy guarantee μ-GDP, number of bootstrap samples B, confidence level $1 - \alpha$.
2: Let $(\tilde{g}_1, \tilde{g}_2, \ldots, \tilde{g}_B, \sigma_e^2) =$ `DP_bootstrap_estimates`(D, g, μ, B).
3: Compute the statistics $s_1 = \frac{1}{B}\sum_{j=1}^{B} \tilde{g}_j$ and $s_2 = \frac{1}{B-1}\sum_{j=1}^{B}(\tilde{g}_j - s_1)^2$.
4: Set $\omega \in (0, \alpha)$ and let c be the $(\alpha - \omega)$ quantile of the χ^2_{B-1} distribution.
5: Compute $\hat{\sigma}_g^2 = \max\left(0, \frac{B-1}{c}s_2 - \sigma_e^2\right)$ and $\hat{\sigma}^2_{\text{upper}} = \hat{\sigma}_g^2 + \frac{\hat{\sigma}_g^2 + \sigma_e^2}{B}$.
6: Compute the confidence interval radius $r = \Phi^{-1}(1 - \frac{\omega}{2})\hat{\sigma}_{\text{upper}}$.
7: **Return** $(s_1 - r,\ s_1 + r)$ which satisfies approximately μ-GDP.

Algorithm 3 `DP_bootstrap_deconvolution_sampling_distribution_and_CI`

1: **Input** data set $D \in \mathcal{X}^n$, statistic $g : \mathcal{X}^n \to \mathbb{R}$, privacy guarantee μ-GDP, number of bootstrap samples B.
2: Let $(\tilde{g}_1, \tilde{g}_2, \ldots, \tilde{g}_B, \sigma_e^2) =$ `DP_bootstrap_estimates`(D, g, μ, B).
3: DP estimate of sampling distribution of $g(D)$: $\tilde{f}_g =$ `deconvolveR`$((\tilde{g}_1, \tilde{g}_2, \ldots, \tilde{g}_B), \sigma_e^2)$.
4: Let $\tilde{F}_g$ be the CDF corresponding to $\tilde{f}_g$.
5: **Return** $(\tilde{f}_g,\ \tilde{F}_g^{-1}(\frac{\alpha}{2}),\ \tilde{F}_g^{-1}(1 - \frac{\alpha}{2}))$ which satisfies approximately μ-GDP.

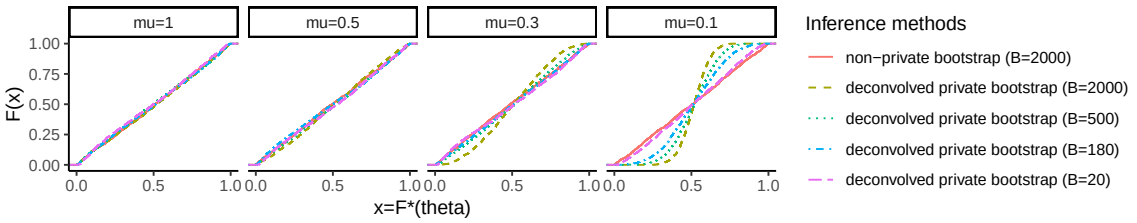

Figure 6: Coverage check of private CI with μ-GDP where $\mu = 1, 0.5, 0.3, 0.1$.

5. Simulations

In this section, we use the confidence interval construction as a showcase for statistical inference with our DP bootstrap algorithm. First, we compare the non-private CI from the original bootstrap to the private CI from the recovered sampling distribution by using deconvolution on our DP bootstrap estimates. Then we compare our deconvolution method with the asymptotic method, the method by Brawner and Honaker (2018), and NoisyVar (Du et al., 2020), and we discuss the difference between these methods.

5.1 Private CI Compared to Non-Private CI for the Population Mean

Consider $D = (x_1, x_2, \ldots, x_n)$ where $x_i \in [0, 1]$ and $x_i \overset{\text{iid}}{\sim} F_X$. We construct private CIs for $\mathbb{E}[x_i]$ with D. Let $g(D) = \frac{1}{n}\sum_{i=1}^{n} x_i$ be the non-private point estimate of $\mathbb{E}[x_i]$, and $\tilde{g}_B(D) = (\tilde{g}(D_1), \tilde{g}(D_2), \ldots, \tilde{g}(D_B))$ be DP bootstrap estimates where $\tilde{g}(D_j) = g(D_j) + \xi_j$, $\xi_j \overset{\text{iid}}{\sim} \mathcal{N}(0, \frac{(2-2/\mathrm{e})B}{n^2\mu^2})$ is Gaussian mechanism, and $D_j \overset{\text{iid}}{\sim} \texttt{boot}(D)$ is the jth bootstrap sample. From our privacy analysis, $\tilde{g}_B(D)$ asymptotically satisfies μ-GDP. We construct private CIs for $\mathbb{E}[X]$ based on $\tilde{g}_B(D)$. The result is run with 2000 replicates where $x_i = \max(0, \min(1, z_i)), z_i \overset{\text{iid}}{\sim} N(0.5, 1)$ and $n = 10000$.

We examine the coverage of CIs with all confidence levels in Figure 6: Since the CI is built with the quantiles of the recovered distribution, and the coverage is determined by whether the true parameter value is in between the two quantiles, we evaluate the CDF of the recovered sampling distribution, F^*, at the true parameter value θ, e.g., $0.05 \le F^*(\theta) \le 0.95$ is equivalent to the 90% CI covering θ. Therefore, the coverage at different confidence levels can be calculated by $\mathbb{E}[\mathbb{1}(p_{\text{lower}} \le F^*(\theta) \le p_{\text{upper}})]$, and we want it to be close to the nominal confidence level, $p_{\text{upper}} - p_{\text{lower}}$. This is achieved if the CDF of $u := F^*(\theta)$ is close to the line $F(u) = u,\ \forall u \in [0, 1]$. In Figure 6, we can see that our DP bootstrap result aligns with the $F(u) = u$ when $B = 2000\mu^2$, similar to the non-private bootstrap, for $\mu = 1, 0.5, 0.3, 0.1$, which corresponds to a constant SNR defined in Remark 24.

In this simulation, our choice of B varies from 20 to 2000. We use the DP bootstrap with a smaller B under a stronger privacy guarantee because larger B leads to smaller SNR, making deconvolution harder. If the coverage is satisfactory under many choices of B, e.g., $B = 20, 180, 500, 2000$ when $\mu = 1$, the largest B gives the shortest CI since the deconvolution accuracy is determined by B. Detailed comparisons of the width, coverage, and corresponding SNR for different choices of B are provided in Appendix E.

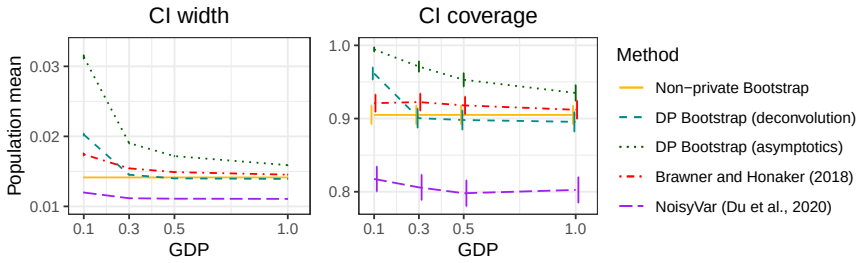

Figure 7: Coverage and width of CIs for the population mean with different privacy guarantees. The confidence level is 90%.

5.2 Comparison Between DP Bootstrap with Deconvolution and Other Methods

In this section, we compare our DP bootstrap with deconvolution to other methods, i.e., non-private bootstrap, DP bootstrap with asymptotic CI, the method by Brawner and Honaker (2018), and NoisyVar (Du et al., 2020), under the settings we used in Section 5.1.

In the non-private setting, the parametric CI for the population mean θ can be built with the t-statistic $t = \frac{\bar{X}-\theta}{\sqrt{s_X^2/n}}$ where $\bar{X}$ and s_X^2 are the sample mean and variance. Similarly, to construct a DP CI, one can obtain a DP t-statistic by replacing $\bar{X}$ and s_X^2 with their corresponding DP statistics, but this DP t-statistic may not follow the t-distribution (or an approximate normal distribution) because of the added noise for privacy. To tackle this issue, our asymptotic method (Theorem 20) and the method by Brawner and Honaker (2018) essentially use conservative, over-estimates of s_X^2, which result in the over-coverage and larger width of the correpsonding CIs. Du et al. (2020) adopted the idea of parametric bootstrap to construct the CI for the population mean: They plugged in the DP sample mean and the DP sample variance to generate normally distributed samples and compute the corresponding DP sample means to estimate the sampling distribution of the DP sample mean. We include NoisyVar in the appendix (Algorithm 4) for easier reference.

In Figure 7, the coverage and width of the 90% CIs from the deconvolution method based on DP bootstrap are similar to the non-private bootstrap, except that under the strongest privacy guarantee $\mu = 0.1$, DP bootstrap CI has over-coverage and larger width. In comparison, the asymptotic method and the method by Brawner and Honaker (2018) always have over-coverage and larger width even when μ is large, while NoisyVar has under-coverage and less width for all μ. The performance of NoisyVar is not as satisfactory as in (Du et al., 2020) in terms of the coverage, because the distribution of our data, the clamped normal random variables, is not in the normal distribution family used by NoisyVar. Our comparison highlights the importance of non-parametric inference: The bootstrap CI does not assume the family of the sampling distribution or data distribution; therefore, our deconvolution results do not suffer from the coverage issue. Note that NoisyVar is also limited to building DP CIs for the population mean, while our DP bootstrap can be used on any DP statistic with additive noise mechanisms, which we demonstrate in the real-world experiments below.

6. Real-World Experiments

In this section, we conduct experiments with the 2016 Census Public Use Microdata Files (PUMF), which provide data on the characteristics of the Canadian population (Canada, 2019). We analyze the dependence between market income and shelter cost in Ontario by the inference of logistic regression and quantile regression under DP guarantees. We use DP bootstrap with output perturbation mechanism (Chaudhuri et al., 2011) and compare our asymptotic and deconvolution method with Wang et al. (2019, Algorithm 5), Differentially Private Confidence Intervals for Empirical Risk Minimization, which we abbreviate as DP-CI-ERM.[2] Our main results are shown in Figure 8, and more detailed comparisons are available in Appendix E.

6.1 Experiment Settings

The PUMF data set contains 930,421 records of individuals, representing 2.7% of the Canadian population. Among the 123 variables in this data set, we choose three variables: the province or territory of current residence (named PR), the market income (named MRKINC), and the shelter cost (named SHELCO). We extract the records of MRKINC and SHELCO belonging to the people in Ontario (according to the values in PR). After removing the records with unavailable values, the sample size is 217,360. We define the extracted data as the original data set and analyze the relationship between MRKINC and SHELCO.

For the exploratory data analysis, we show the non-private empirical joint distribution between MRKINC and SHELCO in Figure 8a, and we assume that their maximum values are prior information since the data set is top-coded, as shown in Figure 8a. We preprocess the data by scaling MRKINC and SHELCO so their ranges are $[0, 1]$.

To evaluate the performance of different statistical inference methods, we calculate the coverage and width of the CIs from 2000 simulations for each setting where the input data sets are sampled from the original data set with replacement with size $n = 1000$, 3000, 10000, 30000, 100000. We also calculate the probability that the CI of the slope parameter covers 0, which indicates that there is not enough evidence to reject the independence between MRKINC and SHELCO. The privacy guarantee is set to be 1-GDP, the confidence level is 90%, and we use $B = 100$ for bootstrap and DP bootstrap.

6.2 Logistic Regression

We set the response $y_i = 1$ if SHELCO ≥ 0.5, otherwise $y_i = -1$. In logistic regression, the model is $P(Y|X) = \frac{1}{1+\exp(-\theta^\intercal X \cdot Y)}$, and the empirical risk minimizer (ERM), also the maximum likelihood estimate of θ, is $\hat{\theta} = \text{argmin}_\theta R(\theta)$ where $R(\theta) := \frac{1}{n}\sum_{i=1}^{n} -\log(P(y_i|x_i))$.

2. For the Line 5 in (Wang et al., 2019, Algorithm 5), we replace c with 0 since in their analysis, the eigenvalues of the Hessian matrix is no less than $2c > 0$ while the eigenvalues of the covariance matrix only need to be non-negative. This modification greatly improves the performance of DP-CI-ERM when n is large, as overestimating the covariance matrix leads to over-coverage and wider CIs.

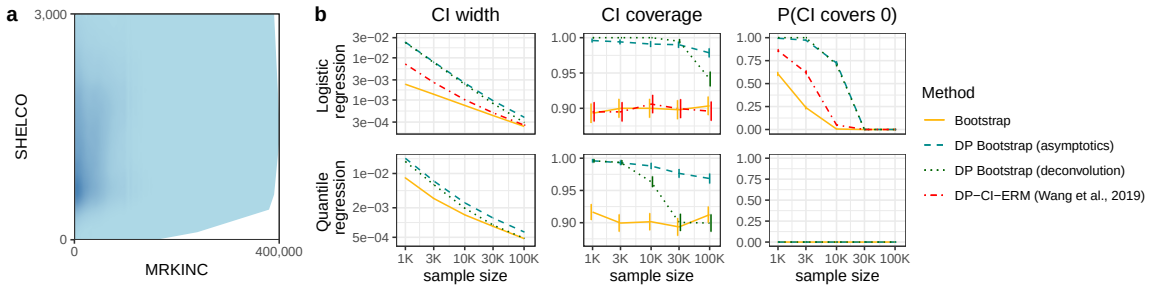

Figure 8: (a) Joint distribution between MRKINC and SHELCO in Ontario, Canada. The polygon region is the convex hull of all data points. (b) Results of 90% CIs for the slope parameters in logistic regression and quantile regression between MRKINC and SHELCO. Note that DP-CI-ERM cannot be used in the inference of quantile regression.

To obtain DP estimates, we implement the output perturbation mechanism following (Wang et al., 2019, Algorithm 5), which replaces $R(\theta)$ by a regularized empirical risk $R(\theta) + c\|\theta\|_2^2$ and adds noise to the output: $\tilde{\theta} = \hat{\theta} + \xi$ where $\hat{\theta} = \text{argmin}_\theta(R(\theta) + c\|\theta\|_2^2)$. As the sensitivity of the regularized ERM is $\Delta(\hat{\theta}) = \frac{1}{nc}$ (Wang et al., 2019), we use Gaussian mechanism, $\xi \sim \mathcal{N}(0, \frac{1}{(\mu nc)^2})$, then $\tilde{\theta}$ satisfies μ-GDP; to satisfy the constraint, $\|x_i\|_2^2 \leq 1$, in the sensitivity analysis, we let the covariate be $x_i = (1/\sqrt{2}, \text{MRKINC}/\sqrt{2})$. Following (Wang et al., 2019), we define the true parameter $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$ as the regularized ERM estimated with the original data set under the same c. We build CI for the slope parameter θ_2: If the 90% CI does not cover 0, we are confident that MRKINC is related to SHELCO.

The results are shown in the upper figures of Figure 8b where $c = 1$. The CIs by DP bootstrap from both deconvolution and asymptotic methods are wider than the ones by DP-CI-ERM when the sample size n is small: As a non-parametric method, DP bootstrap is applicable to more general settings but does not fully utilize the structure of the private ERM as opposed to DP-CI-ERM, so its CIs are often not as tight as the ones by DP-CI-ERM. Both types of private CIs are as wide as non-private bootstrap CI when $n = 100000$, do not suffer from the under-coverage issue, and have $P(\text{CI covers } 0) \approx 0$ when $n \geq 30000$.

6.3 Quantile Regression

We use quantile regression as an example to demonstrate the advantage of DP bootstrap as a non-parametric method being applicable to general settings where DP-CI-ERM is not. We set the response $y_i = \text{SHELCO}$ and the covariate $x_i = (1, \text{MRKINC})$. Following (Reimherr and Awan, 2019), we assume that the conditional quantile function of Y given X is $Q_{Y|X}(\tau) = X^\intercal \theta_\tau$, we estimate θ_τ also by ERM with the objective function $R(\theta) = \frac{1}{n}\sum_{i=1}^n l(z_i)$ where $l(z_i) = (\tau - \mathbb{1}(z_i \leq 0))z_i$ and $z_i = y_i - x_i^\intercal \theta$. Similar to our experiment with logistic regression, we define the true parameter $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$ as the regularized ERM estimated with the original data set under the same regularization parameter c. By (Chaudhuri et al., 2011, Lemma 7), the sensitivity of the regularized ERM,

$\hat{\theta} = \text{argmin}_\theta\{R(\theta) + c\|\theta\|_2^2\}$, is bounded by $\Delta(\hat{\theta}) = \frac{\max\{2\tau, 2(1-\tau), \sqrt{2}\}}{2nc}$; the derivation is in Appendix D. Then $\tilde{\theta} = \hat{\theta} + \xi$ satisfies μ-GDP when $\xi \sim \mathcal{N}(0, \frac{\Delta(\hat{\theta})^2}{\mu^2})$ (Gaussian mechanism.)

To the best of our knowledge, this is the first result of building private CIs for the coefficients of quantile regression. As DP-CI-ERM uses Taylor expansion of the gradient of $R(\theta)$ to characterize the difference between the estimate $\hat{\theta}$ and the true parameter θ^*, i.e., $\theta^* - \hat{\theta} \approx H[R(\hat{\theta})]^{-1}(\nabla R(\theta^*) - \nabla R(\hat{\theta}))$, it is not usable in quantile regression as the Hessian of $R(\theta)$ is always 0 when it exists. The results for DP bootstrap and non-private bootstrap are shown in the lower figures of Figure 8b where we set $c = 1$ and $\tau = 0.5$. Similar to our experiment in logistic regression, DP bootstrap never suffers from the under-coverage issue. The deconvolution CIs from DP bootstrap perform similarly to non-private bootstrap CIs when $n \geq 30000$. We can see that $\mathbb{P}(\text{CI covers } 0) \approx 0$ when $n \geq 1000$; therefore, we are confident that MRCINC and the median of corresponding SHELCO are not independent with 90% confidence.

7. Conclusion

Our analysis of the DP bootstrap provides a new perspective on resampling in DP by considering the output distribution as a mixture distribution which gives a tractable lower bound for the DP bootstrap in f-DP. Furthermore, our composition result for the DP bootstrap using the Gaussian mechanism gives a simple asymptotic privacy guarantee confirmed by our numerical evaluation results of the exact cumulative privacy cost, and it highlights the minimal cost of the DP bootstrap. For DP statistical inference, we show that the sample mean of the DP bootstrap estimates is a consistent point estimator, and we propose an asymptotic method and a deconvolution method to build CIs. We prove an asymptotic coverage guarantee of the asymptotic CIs, and show that their average width enjoys the optimal convergence rates. To improve the finite sample performance, we use deconvolution to construct CIs, and to the best of our knowledge, we are the first to use deconvolution[3] to recover the non-private sampling distribution from DP estimates and conduct statistical inference. Our simulations and experiments show that the CIs generated by the deconvolved distribution achieve the nominal coverage, and our results are not only comparable to existing methods like NoisyVar and DP-CI-ERM but also applicable to the inference problems such as quantile regression where existing methods cannot be used.

One direction of future work is on improving the privacy analysis: The lower bound in Theorem 11 could be tightened by considering all α_i jointly rather than individually.

For statistical inference using the DP bootstrap, the choice of the number of bootstrap samples B can potentially be further optimized for a given sample size n and privacy parameter μ. One may pursue finite sample utility results for the deconvolved distribution or the CIs obtained by the DP bootstrap deconvolution procedure. Furthermore, our statistical results are limited to the Gaussian mechanism, which may not be an ideal mechanism, due to its additive nature and the need to calculate ℓ_2-sensitivity. While our statistical approaches should be easily extended to allow for other additive noise mechanisms, non-additive noise

3. Farokhi (2020) applied deconvolution to estimate the distribution of the sensitive data under local DP guarantees, which is different from the DP guarantee discussed in this paper.

mechanisms may require new inference techniques. We also notice in Figure 7 and Figure 8b that with a strong privacy guarantee or small n, our CI may still be wider and have higher coverage than the non-private CI or the private CI from other methods, indicating that the deconvolution procedure can be further optimized to get a tighter estimate and improve the width of the CIs while maintaining the nominal coverage. Apart from the standard interval and the percentile interval, there are other inference methods based on the bootstrap estimates such as BC_a (bias-corrected and accelerated) (Efron, 1987) or ABC (approximate bootstrap confidence) (Diciccio and Efron, 1992), we leave it to future work to investigate the private versions of these methods.

We can also use the DP bootstrap framework for high-dimensional inference. Our privacy analysis applies to any mechanism, including those multivariate mechanisms, e.g., the Gaussian mechanism on each dimension of the estimate. Future work in this direction is to apply existing methods of multivariate deconvolution (Youndjé and Wells, 2008; Hazelton and Turlach, 2009; Sarkar et al., 2018) and bootstrap (Hall, 1987; Aelst and Willems, 2005) to our DP bootstrap framework.

Acknowledgments

This work was supported in part by the National Science Foundation [NSF grants no. SES-2150615, no. DMS-2134209, and no. CNS-2247795], the Office of Naval Research [ONR award no. N00014-22-1-2680], and Optum AI and CISCO research grants. The authors would like to thank Dr. Chendi Wang for the discussion about the relationship between the advanced joint convexity property of f-DP and our Theorem 10 and 11.

Appendix A. Proofs for Section 3

In this section, we provide the proofs for the theorems and propositions in Section 3.

A.1 Proofs for Section 3.1

We first restate the (ε, δ)-DP results in (Balle et al., 2018) and provide some useful results for the proof of Proposition 7.

Theorem 25 (Theorem 10 in (Balle et al., 2018)) *Given $\varepsilon \geq 0$, assume $\mathcal{M}$ satisfies $(\varepsilon, \delta_{\mathcal{M},i}(\varepsilon))$-DP with group size i. Let $p_i = \binom{n}{i}(\frac{1}{n})^i(1-\frac{1}{n})^{n-i}$ and $\varepsilon' = \log(1+(1-p_0)(e^\varepsilon-1))$, then $\mathcal{M} \circ \texttt{boot}$ satisfies $(\varepsilon', \delta_{\mathcal{M}\circ\texttt{boot}}(\varepsilon'))$-DP where $\delta_{\mathcal{M}\circ\texttt{boot}}(\varepsilon') = \sum_{i=1}^n p_i \delta_{\mathcal{M},i}(\varepsilon)$.*

Definition 26 *Let f be a tradeoff function, $\bar{x} = \inf\{x \in [0,1] : -1 \in \partial f(x)\}$. The symmetrization operator which maps a possibly asymmetric tradeoff function to a symmetric tradeoff function is defined as*

$$\mathrm{Symm}(f) := \begin{cases} \min\{f, f^{-1}\}^{**}, & \text{if } \bar{x} \leq f(\bar{x}), \\ \max\{f, f^{-1}\}, & \text{if } \bar{x} > f(\bar{x}). \end{cases}$$

Proposition 27 (Proposition E.1 in Dong et al., 2022) *Let f be a tradeoff function. Suppose a mechanism is $(\varepsilon, 1 + f^*(-e^\varepsilon))$-DP for all $\varepsilon \geq 0$, then it is* $\mathrm{Symm}(f)$*-DP.*

Proof [Proof of Proposition 7] We use f-DP to restate the $\delta_{\mathcal{M}\circ\texttt{boot}}(\varepsilon')$ in Theorem 25: Let $p = 1-p_0$, $\delta_{\mathcal{M}\circ\texttt{boot}}(\varepsilon') \leq \sum_{i=1}^n p_i \delta_{\mathcal{M},i}(\varepsilon) = \sum_{i=1}^n p_i(1+f^*_{\mathcal{M},i}(-\mathrm{e}^\varepsilon)) = p\left(1 + \sum_{i=1}^n \frac{p_i}{1-p_0} f^*_{\mathcal{M},i}(-\mathrm{e}^\varepsilon)\right)$ where $\mathcal{M}$ satisfies $f_{\mathcal{M},i}$-DP with group size i and $f_{\mathcal{M},i}$ is symmetric.

From the Supplement to (Dong et al., 2022) (Page 42), we know that $\varepsilon' = \log(1-p+p\mathrm{e}^\varepsilon)$ and $\delta' = p\big(1 + f^*(-\mathrm{e}^\varepsilon)\big)$ can be re-parameterized into $\delta' = 1 + f_p^*(-\mathrm{e}^{\varepsilon'})$ where $f_p = pf + (1-p)\mathrm{Id}$. For symmetric f, we have $\mathrm{Symm}(f_p) = C_p(f)$ since $\bar{x} \leq f_p(\bar{x})$ where $\bar{x} = \inf\{x \in [0,1] : -1 \in \partial f(x)\}$. Using Proposition 27, we have $f_{\mathcal{M}\circ\texttt{boot}} = C_p\left(\left(\sum_{i=1}^n \frac{p_i}{1-p_0} f^*_{\mathcal{M},i}\right)^*\right)$. ■

Lemma 28 *For $i = 1, 2, \ldots, k$, let f_i be tradeoff functions and $p_i \in (0,1]$ satisfying $\sum_{i=1}^k p_i = 1$. We write $\underline{f} = (f_1, \ldots, f_k)$ and $\underline{p} = (p_1, \ldots, p_k)$.*

1. *$\mathrm{mix}(\underline{p}, \underline{f}) : [0,1] \to [0,1]$ is a well-defined tradeoff function.*
2. *If the tradeoff functions f_i are all symmetric, then $\mathrm{mix}(\underline{p}, \underline{f})$ is symmetric.*

We first state some useful properties of tradeoff functions for proving Lemma 28.

Proposition 29 (Proposition 1 in Dong et al., 2022) *A function $f : [0,1] \to [0,1]$ is a tradeoff function if and only if f is convex, continuous, non-increasing, and $f(x) \leq 1 - x$.*

Proposition 30 *For any tradeoff function f,*

1. *$f(\alpha)$ is strictly decreasing for $\alpha \in \{\alpha : f(\alpha) > 0\}$.*
2. *$0 \notin \partial f(\alpha)$ if and only if $\alpha \in \{\alpha : f(\alpha) > 0\}$.*
3. *If $f^{-1}(f(y)) \neq y$, then $f(y) = 0$.*
4. *If $f = f^{-1}$ and there exists $C < 0$ such that $C \in \partial f(\alpha)|_{\alpha=f(\alpha_0)}$, then we have $f(f(\alpha_0)) = \alpha_0$ and $\frac{1}{C} \in \partial f(\alpha)|_{\alpha=f(\alpha_0)}$.*
5. *There is exactly one $\bar{\alpha}$ such that $f(\bar{\alpha}) = \bar{\alpha}$.*
6. *If $f = f^{-1}$ and $f(\bar{\alpha}) = \bar{\alpha}$, then $-1 \in \partial f(\alpha)|_{\alpha=\bar{\alpha}}$.*
7. *If $f = f^{-1}$ and $-1 \in \partial f(\alpha)|_{\alpha=\alpha_0}$, then $-1 \in \partial f(\alpha)|_{\alpha=f(\alpha_0)}$.*

Proof [Proof of Proposition 30] We provide the proofs for each property of f below.

1. If there are $0 \leq \alpha_1 < \alpha_2 \leq 1$ such that $f(\alpha_1) = f(\alpha_2) = \beta > 0$, since g is convex, we will have $f(\alpha) \geq \beta \ \forall \alpha \geq \alpha_1$. Therefore, $f(1) \geq \beta > 0$ which contradicts with $f(1) \leq 1 - 1 = 0$. Since $f(\alpha)$ is not increasing, it is strictly decreasing for $\alpha \in \{\alpha : f(\alpha) > 0\}$.

2. If $0 \in \partial f(\alpha)$ and $f(\alpha) > 0$, we have $f(y) \geq f(\alpha) > 0 \ \forall y \in [0, 1]$. This contradicts with the fact that $f(1) = 0$. If $0 \notin \partial f(\alpha)$ and $f(\alpha) = 0$, there exists $y \in [0, 1]$ such that $f(y) < 0 \cdot (y - \alpha) + f(\alpha) = 0$ which contradicts with the fact that $f : [0, 1] \to [0, 1]$.

3. If $f^{-1}(f(y)) = \inf\{t \in [0, 1] : f(t) \leq f(y)\} < y$, since f is not increasing, we have that $f(f^{-1}(f(y))) = f(y)$, which holds only when $f(y) = 0$ (otherwise $f(y)$ is strictly decreasing).

4. From the fact that C is a sub-differential value of f at α_0, we have $f(z) \geq C(z - \alpha_0) + f(\alpha_0) \ \forall z \in [0, 1]$. If there exists z such that $z < \alpha_0$ and $f(z) \leq f(\alpha_0)$, it contradicts with $f(z) \geq C(z - \alpha_0) + f(\alpha_0) > f(\alpha_0)$. Therefore,

$$f(f(\alpha_0)) = f^{-1}(f(\alpha_0)) = \inf\{t \in [0, 1] : f(t) \leq f(\alpha_0)\} = \alpha_0.$$

   We prove $\frac{1}{C} \in \partial f(\alpha)|_{\alpha=f(\alpha_0)}$ by showing $f(y) \geq \frac{1}{C}(y - f(\alpha_0)) + f(f(\alpha_0)) \ \forall y \in [0, 1]$. Since $C \in (-\infty, 0)$, if $f(y) \geq \frac{1}{C}(y - f(\alpha_0)) + f(f(\alpha_0))$ holds when $y = f(0)$, then it also holds for $y > f(0)$ as $f(y) = 0 = f(f(0)) \geq \frac{1}{C}(f(0) - f(\alpha_0)) + \alpha_0 \geq \frac{1}{C}(y - f(\alpha_0)) + \alpha_0$. For $y \in [0, f(0)]$, we define $z := f(y)$. Since $f = f^{-1}$ and f is strictly decreasing when $f > 0$, we know that $f(y) > 0$ and $f(z) = f^{-1}(f(y)) = y$ for $y \in [0, f(0))$. We also know that $f(z) = y$ when $y = f(0)$ since $z = f(y) = f(f(0)) = 0$. From the fact that C is a sub-differential value of f at α_i, we have $f(z) \geq C(z - \alpha) + f(\alpha)$. Now we show $f(y) \geq \frac{1}{C}(y - f(\alpha_0)) + f(f(\alpha_0)) \ \forall y \in [0, f(0)]$ through the analysis below.

   - If $f(0) \geq y > f(\alpha_0)$, we have $z = f(y) < \alpha_0$, $f(z) = y$, and $C \geq \frac{f(z) - f(\alpha_0)}{z - \alpha_0}$. Therefore, $\frac{1}{C}(y - f(\alpha_0)) + f(f(\alpha_0)) \leq \frac{z - \alpha_0}{f(z) - f(\alpha_0)}(y - f(\alpha_0)) + \alpha_0 = z = f(y)$.

- If $0 \leq y < f(\alpha_0)$, we have $z = f(y) > \alpha_0$, $f(z) = y$, and $C \leq \frac{f(z)-f(\alpha_0)}{z-\alpha_0}$. Therefore, $\frac{1}{C}(y - f(\alpha_0)) + f(f(\alpha_0)) \leq \frac{z-\alpha_0}{f(z)-f(\alpha_0)}(y - f(\alpha_0)) + \alpha_0 = z = f(y)$.
- If $y = f(\alpha_0)$, we have $\frac{1}{C}(y - f(\alpha_0)) + f(f(\alpha_0)) = f(y)$.

5. If there are $\bar{\alpha}_1 < \bar{\alpha}_2$ that $\bar{\alpha}_1 = f(\bar{\alpha}_1)$ and $\bar{\alpha}_2 = f(\bar{\alpha}_2)$, then since f is non-increasing, we have $f(\bar{\alpha}_1) \geq f(\bar{\alpha}_2)$ which contradicts with $\bar{\alpha}_1 < \bar{\alpha}_2$. Since $f(\alpha)$ is continuous and $f(0) - 0 \geq 0 - 0 = 0$, $f(1) - 1 = -1 < 0$, there exists $\bar{\alpha} \in [0, 1]$ such that $f(\bar{\alpha}) - \bar{\alpha} = 0$.

6. If $-1 \notin \partial f(\bar{\alpha})$, there exists $y \in [0, 1]$ such that $f(y) < -(y - \bar{\alpha}) + f(\bar{\alpha})$ and $f(f(y)) = y$ because $f = f^{-1}$ (if $f(f(y)) \neq y$, we have $f(y) = 0$, then we can replace y with $f(0) \leq y$ and we still have $f(y) < -(y - \bar{\alpha}) + f(\bar{\alpha})$.) Therefore, $(y, f(y))$ and $(f(y), y)$ are both on the curve of f. Since $y \neq \bar{\alpha}$, we know $y \neq f(y)$. Without of the loss of generality, we assume that $y > f(y)$. Then we know that $y > \bar{\alpha} > f(y)$ since otherwise we will have contradictions: $\bar{\alpha} \geq y > f(y) \geq f(\bar{\alpha}) = \bar{\alpha}$ or $\bar{\alpha} \leq f(y) < y = f(f(y)) \leq f(\bar{\alpha}) = \bar{\alpha}$. We denote $q = \frac{\bar{\alpha}-f(y)}{y-f(y)} > 0$ and $1 - q = \frac{y-\bar{\alpha}}{y-f(y)} > 0$. Then $\bar{\alpha} = qy + (1 - q)f(y)$, and by the convexity of f, we have $\bar{\alpha} = f(\bar{\alpha}) \leq qf(y) + (1 - q)f(f(y)) = qf(y) + (1 - q)y$. Therefore, $f(\bar{\alpha}) + \bar{\alpha} \leq (qy + (1 - q)f(y)) + (qf(y) + (1 - q)y) = y + f(y)$. But from $f(y) < -(y - \bar{\alpha}) + f(\bar{\alpha})$, we know $f(\bar{\alpha}) + \bar{\alpha} > y + f(y)$, which leads to a contradiction. Therefore, we have $-1 \in \partial f(\bar{\alpha})$.

7. As $-1 \in \partial f(\alpha)|_{\alpha=\alpha_0}$, we have $f(y) \geq -(y - \alpha_0) + f(\alpha_0)$ $\forall y \in [0, 1]$ and $f(f(\alpha_0)) = \alpha_0$. Therefore, $f(y) \geq -(y - f(\alpha_0)) + f(f(\alpha_0))$ $\forall y \in [0, 1]$ and we have $-1 \in \partial f(\alpha)|_{\alpha=f(\alpha_0)}$.

■

Proof [Proof of Lemma 28] For part 1, first we show that for every $\alpha \in (0, 1)$, there exists $C \in (-\infty, 0]$ such that $\alpha \in A(C)$. Since each f_i is convex and non-increasing, its sub-differential $\partial f_i(\alpha_i)$ is in $(-\infty, 0]$ and non-decreasing with respect to α_i. Therefore, for any $-\infty < C_1 < C_2 \leq 0$, we have $a_1 \leq a_2$ $\forall a_1 \in A_i(C_1), a_2 \in A_i(C_2)$, and for any $0 < a_1 < a_2 < 1$, we have $C_1 \leq C_2$ $\forall C_1 \in \partial f_i(a_1), C_2 \in \partial f_i(a_2)$. We name these two properties as the monotonicity of the sub-differential mapping.

Since each f_i is convex and continuous, its sub-differential $\partial f_i(\alpha_i)$ is also continuous in the sense that for any $\varepsilon > 0$, there exists $\delta > 0$ such that $\partial f_i(\alpha_i') \subset \partial f_i(\alpha_i) + (-\varepsilon, \varepsilon)$ whenever $\|\alpha_i' - \alpha_i\| < \delta$ (see Exercise 2.2.22(a) in (Borwein and Vanderwerff, 2010)).

From the continuity and monotonicity of the sub-differential mapping, we know that for any $C \in (-\infty, 0]$, $A_i(C)$ is a closed interval, e.g., $[a, b]$. Note that $A_i(C)$ is always nonempty: let $C_{i,\text{range}} := \bigcup_{\alpha \in (0,1)} \partial f_i(\alpha)$; if $C < C_i$ $\forall C_i \in C_{i,\text{range}}$, we have $A_i(C) = \{0\}$; if $C > C_i$ $\forall C_i \in C_{i,\text{range}}$, we have $A_i(C) = \{1\}$; if $C \in C_{i,\text{range}}$, there must be an $\alpha \in (0, 1)$ such that $C \in \partial f_i(\alpha)$. By the same reasoning, we also have $(0, 1) \subset \bigcup_{C \in (-\infty, 0]} A_i(C)$ for any i. As $A(C) = \{\sum_{i=1}^k p_i \alpha_i | \alpha_i \in A_i(C)\}$, we have $(0, 1) \subset \bigcup_{C \in (-\infty, 0]} A(C)$, and we also have the monotonicity of $A(C)$ with respect to C.

Next, we show that $\text{mix}(\underline{p}, \underline{f})$ is a well-defined function, i.e., for a given α, although there could be multiple choices of $\{\alpha_i\}_{i=1}^k$ such that $\sum_{i=1}^k p_i \alpha_i = \alpha$, we will obtain the

same value of $\sum_{i=1}^k p_i f_i(\alpha_i)$ for all choices. Consider two choices, $\{\alpha_i\}_{i=1}^k$ and $\{\alpha_i'\}_{i=1}^k$, that correspond to the same α. Let C and C' correspond to $\{\alpha_i\}_{i=1}^k$ and $\{\alpha_i'\}_{i=1}^k$ respectively. As there exist $i \neq j$ such that $\alpha_i < \alpha_i'$ and $\alpha_j > \alpha_j'$ (since $\sum_{i=1}^k p_i\alpha_i = \sum_{i=1}^k p_i\alpha_i'$ and $p_i > 0$ for $i = 1, 2, \ldots, k$), from the monotonicity of the sub-differential mapping, we know that $C = C'$ since $C \leq C'$ and $C' \leq C$. For $\{\alpha_i\}_{i=1}^k$ and $\{\alpha_i'\}_{i=1}^k$, since $\partial f_i(\alpha_i) = \partial f_i(\alpha_i') = C$, we have $f_i(\alpha_i) \geq C(\alpha_i - \alpha_i') + f_i(\alpha_i') \geq f_i(\alpha_i)$ for $i = 1, 2, \ldots, n$. Therefore, $f_i(\alpha_i) - f_i(\alpha_i') = C(\alpha_i - \alpha_i')$. As we know $\sum_{i=1}^k p_i\alpha_i = \sum_{i=1}^k p_i\alpha_i'$, we have $\sum_{i=1}^k p_i f_i(\alpha_i) - \sum_{i=1}^k p_i f_i(\alpha_i') = C\sum_{i=1}^k p_i(\alpha_i - \alpha_i') = 0$, which means that $\text{mix}(\underline{p}, \underline{f})$ is well-defined.

Finally, we show that $\text{mix}(\underline{p}, \underline{f})$ is a tradeoff function.

Let $f = \text{mix}(\underline{p}, \underline{f})$. We can see that $f(x) \in [0, 1]\ \forall x \in [0, 1]$ and $f(x) \leq 1 - x$. We also know f is non-increasing because of the monotonicity of $A(C)$, the monotonicity of the sub-differential mapping, and f_i being non-increasing.

Now we prove that f is continuous. For a fixed α and $\delta > 0$, we can find the $\{\alpha_i\}_{i=1}^k$ corresponding to this α, and find ε_i such that $|f_i(\alpha_i') - f_i(\alpha_i)| < \delta$ whenever $|\alpha_i' - \alpha_i| < \varepsilon_i$; then we let $\varepsilon = \min_{i\in\{1,2,\ldots,k\}} \varepsilon_i p_i$, and we have $|f(\alpha') - f(\alpha)| < \delta$ whenever $|\alpha' - \alpha| < \varepsilon$. To prove this, without loss of generality, we assume $\alpha < \alpha'$, and we can find $\{\alpha_i\}_{i=1}^k$ and $\{\alpha_i'\}_{i=1}^k$ corresponding to α and α' respectively, where $\alpha_i \leq \alpha_i'$ for $i = 1, 2, \ldots, k$. Then if $\alpha' - \alpha < \varepsilon$, we must have $\alpha_i' - \alpha_i < \varepsilon_i$ for $i = 1, 2, \ldots, k$, therefore $|f(\alpha') - f(\alpha)| < \delta$.

Now we prove that f is convex. By the definition of convexity, we only need to show that for any $\alpha, \alpha', t \in [0, 1]$, we have $tf(\alpha) + (1-t)f(\alpha') \geq f(t\alpha + (1-t)\alpha')$. From the construction of $\text{mix}(\underline{p}, \underline{f})$, we can find $\{\alpha_i\}_{i=1}^k$, $\{\alpha_i'\}_{i=1}^k$, and $\{\tilde{\alpha}_i\}_{i=1}^k$ with their matched sub-differential being C, C', and $\tilde{C}$ corresponding to α, α', and $\tilde{\alpha} = t\alpha + (1-t)\alpha'$ respectively. Then

$$\begin{aligned}
tf(\alpha) + (1-t)f(\alpha') &= \sum_{i=1}^k p_i(tf_i(\alpha_i) + (1-t)f_i(\alpha_i')) \\
&\geq \sum_{i=1}^k p_i f_i(t\alpha_i + (1-t)\alpha_i') \quad \text{(convexity)} \\
&\geq \sum_{i=1}^k p_i(\tilde{C}(t\alpha_i + (1-t)\alpha_i' - \tilde{\alpha}_i) + f_i(\tilde{\alpha}_i)) \quad (\text{since } \tilde{C} \in \partial f_i(\tilde{\alpha}_i)) \\
&= \tilde{C}(t\alpha + (1-t)\alpha' - \tilde{\alpha}) + f(t\alpha + (1-t)\alpha') = f(t\alpha + (1-t)\alpha').
\end{aligned}$$

Therefore, f is convex, and we have proved that f is a tradeoff function.

For part 2, let $g = \text{mix}(\underline{p}, \underline{f})$. We prove that g is symmetric by showing $g^{-1} = \text{mix}(\underline{p}, \underline{f})$.

By definition, $g^{-1}(\beta) = \inf\{\alpha \in [0, 1] : g(\alpha) \leq \beta\}$. By the construction of $\text{mix}(\underline{p}, \underline{f})$, for each $\alpha \in [0, 1]$, there exist a constant C and $\{\alpha_i\}_{i=1}^k$ such that $C \in \partial f_i(\alpha_i)$, $\alpha = \sum_{i=1}^k p_i\alpha_i$, and $g(\alpha) = \sum_{i=1}^k p_i f_i(\alpha_i)$.

If $\beta = 0$, then $g(\alpha) = 0$, and $f_i(\alpha_i) = 0$. Therefore, $g^{-1}(0) = \inf\{\alpha \in [0, 1] : \alpha = \sum_{i=1}^k p_i\alpha_i, f_i(\alpha_i) = 0\}$. Since f_i is symmetric, we have that $\inf\{\alpha_i \in [0, 1] : f_i(\alpha_i) = 0\} = f_i^{-1}(0)$. Therefore, $g^{-1}(0) = \sum_{i=1}^k p_i f_i^{-1}(0) = \sum_{i=1}^k p_i f_i(0) = g(0)$.

If $\beta \geq g(0)$, then from the definition of g^{-1}, we have $g^{-1}(\beta) = 0$, and we need to prove $g(\alpha) = 0$ for $\alpha \geq g(0)$. From the construction of $\text{mix}(\underline{p}, \underline{f})$, we have $g(\alpha) = \sum_{i=1}^k p_i f_i(\alpha_i)$ where $\alpha = \sum_{i=1}^k p_i \alpha_i$. Let $\alpha \geq g(0) = \sum_{i=1}^k p_i f_i(0)$. Then, if there exists i that $\alpha_i > f_i(0)$ which means that $f_i(\alpha_i) = 0$, then $\partial f_i(\alpha_i) = \{0\}$; therefore, $\partial f_j(\alpha_j) = \{0\}$ and $f_j(\alpha_j) = 0$ for $j = 1, 2, \ldots, k$. We have $g(\alpha) = 0$. If $\alpha_i \leq f_i(0)$ for $i = 1, 2, \ldots, k$, since $\sum_{i=1}^k p_i \alpha_i = \alpha \geq \sum_{i=1}^k p_i f_i(0)$, we have $\alpha_i = f_i(0)\ \forall i$, which means $f_i(\alpha_i) = 0\ \forall i$; therefore, $g(\alpha) = 0$.

If $g(0) > \beta > 0$, since g is a tradeoff function, there exists only one $\alpha \in [0, 1]$ such that $g(\alpha) = \beta$. From the construction of g, i.e., $g(\alpha) = \sum_{i=1}^k p_i f_i(\alpha_i) = \beta > 0$, for the $\{\alpha_i\}_{i=1}^k$ corresponding to α, there exists i_0 such that $f_{i_0}(\alpha_{i_0}) > 0$. Since $f_{i_0}(\alpha_{i_0}) > 0$, we have $0 \notin \partial f_{i_0}(\alpha_{i_0})$. Therefore, the corresponding constant C that $C \in \partial f_i(\alpha_i)$ for $i = 1, 2, \ldots, k$, and C is not 0. Therefore, $f_i(\alpha_i) \neq 0$ for $i = 1, 2, \ldots, k$.

From Proposition 30, we know that $\frac{1}{C} \in \partial f_i(\alpha)|_{\alpha = f_i(\alpha_i)}$. Let $\alpha = \sum_{i=1}^k p_i \alpha_i$, and $g(\alpha) = \sum_{i=1}^k p_i f_i(\alpha_i)$. As g is a tradeoff function, $g(\alpha)$ is strictly decreasing in $\{\alpha | g(\alpha) > 0\}$. Therefore, for any $g(\alpha) > 0$, there is a one-to-one mapping between α and $g(\alpha)$, and $g^{-1}(g(\alpha)) = \alpha$. Now we can view g^{-1} as a mixture of f_i at the values $f_i(\alpha_i)$: we let $\beta_i := f_i(\alpha_i)$; since $f_i(f_i(\alpha_i)) = \alpha_i$, we have $g^{-1}(\sum_{i=1}^k p_i \beta_i) = \sum_{i=1}^k p_i f_i(\beta_i)$, and there exists a constant $\frac{1}{C}$ such that $\frac{1}{C} \in \partial f_i(\alpha)|_{\alpha = \beta_i}$ for $i = 1, 2, \ldots, k$. Therefore, $g^{-1} = \text{mix}(\underline{p}, \underline{f})$. Since the mixture operation is well-defined, we know that $g = g^{-1}$. ■

Proof [Proof of Theorem 10] Consider the neighboring data sets D_1, D_2, and a rejection rule ψ giving the type I error $\alpha = \mathbb{E}_{\mathcal{M}(D_1)}\psi$, and type II error $\beta = \mathbb{E}_{\mathcal{M}(D_2)}(1 - \psi)$. We write $\alpha_i = \mathbb{E}_{\mathcal{M}_i(D_1)}\psi,\ \beta_i = \mathbb{E}_{\mathcal{M}_i(D_2)}(1 - \psi)$. Then $\alpha = \sum_{i=1}^k p_i \alpha_i,\ \beta = \sum_{i=1}^k p_i \beta_i$.

In order to obtain the lower bound of β given α, which we denote as $f_{\min}(\alpha)$, we not only need the tradeoff between α_i and β_i, which is f_i, but also consider the tradeoff between α_i and α_j because of the constraint $\sum_{i=1}^k p_i \alpha_i = \alpha$. Therefore, we consider $f_{\min}(\alpha) = \min\{\sum_{i=1}^k p_i \beta_i \mid \beta_i \geq f_i(\alpha_i), \alpha = \sum_{i=1}^k p_i \alpha_i, \alpha_i \in [0, 1]\}$, which is a convex optimization problem since the objective function is linear, and the constraints f_i are all convex (Proposition 29). Therefore, by Karush–Kuhn–Tucker theorem (Boyd and Vandenberghe, 2004), let the Lagrangian function be $L(\{\alpha_i, \beta_i, \mu_i, \nu_i, \kappa_i\}_{i=1}^k, \lambda) = \sum_{i=1}^k p_i \beta_i + \sum_{i=1}^k \big(\mu_i(f_i(\alpha_i) - \beta_i) + \nu_i(-\alpha_i) + \kappa_i(\alpha_i - 1)\big) + \lambda\left(\alpha - \sum_{i=1}^k p_i \alpha_i\right)$, and the minimum of $\sum_{i=1}^k p_i \beta_i$ is achieved at $\{\alpha_i, \beta_i\}_{i=1}^k$ if and only if the following conditions are satisfied

$$
\begin{aligned}
\text{Stationarity: } & p_i - \mu_i = 0,\ i = 1, 2, \ldots, k; \\
& 0 \in \mu_i \partial f_i(\alpha_i) - \nu_i + \kappa_i - \lambda p_i,\ i = 1, 2, \cdots, k. \\
\text{Primal feasibility: } & \sum_{i=1}^k \alpha_i = \alpha,\ \beta_i \geq f_i(\alpha_i),\ 0 \leq \alpha_i \leq 1,\ i = 1, 2, \ldots, k; \\
\text{Dual feasibility: } & \mu_i \geq 0,\ \nu_i \geq 0,\ \kappa_i \geq 0,\ i = 1, 2, \ldots, k. \\
\text{Complementary slackness: } & \sum_{i=1}^k (\mu_i(f_i(\alpha_i) - \beta_i) + \nu_i(-\alpha_i) + \kappa_i(\alpha_i - 1)) = 0.
\end{aligned}
$$

Therefore, we have $\mu_i = p_i \neq 0$ from the stationarity condition, and $f_i(\alpha_i) = \beta_i$ from the complementary slackness condition.

- If $0 < \alpha_i < 1$, we have $\nu_i = \kappa_i = 0$, therefore, $\lambda \in \partial f_i(\alpha_i)$;
- If $\alpha_i = 0$, we have $\kappa_i = 0$ and $\partial f_i(\alpha_i) \ni \frac{\nu_i}{\mu_i} + \lambda \geq \lambda$, therefore, for any $\alpha \in [0,1]$, we have $f_i(\alpha) \geq (\alpha - \alpha_i)(\frac{\nu_i}{\mu_i} + \lambda) + f_i(\alpha_i) \geq (\alpha - \alpha_i)\lambda + f_i(\alpha_i)$, i.e., $\lambda \in \partial f_i(\alpha_i)$;
- If $\alpha_i = 1$, we have $\nu_i = 0$ and $\partial f_i(\alpha_i) \ni -\frac{\kappa_i}{\mu_i} + \lambda \leq \lambda$; similarly, we have $\lambda \in \partial f_i(\alpha_i)$.

As f_i is a tradeoff functions, $\partial f_i(\alpha_i)$ is non-decreasing when α_i increases. We write $\partial f_i(\alpha_i) > \lambda$ if $a > \lambda \ \forall a \in \partial f_i(\alpha_i)$. For a given λ,

- if there exist two constants $0 < \alpha_{i,1} < \alpha_{i,2} < 1$ such that $\partial f_i(\alpha_{i,1}) \leq \lambda \leq \partial f_i(\alpha_{i,2})$, we define $\alpha_i^{\lambda,\text{lower}}$ and $\alpha_i^{\lambda,\text{upper}}$ such that $\alpha_i \in [\alpha_i^{\lambda,\text{lower}}, \alpha_i^{\lambda,\text{upper}}]$ if and only if $\lambda \in \partial f_i(\alpha_i)$;
- if $\partial f_i(\alpha_i) > \lambda$ for all $\alpha_i \in (0,1)$, we define $\alpha_i^{\lambda,\text{lower}} = \alpha_i^{\lambda,\text{upper}} = 0$; intuitively, we do not want to have any of the type I error of α on the f_i part since the corresponding type II error would be larger otherwise;
- if $\partial f_i(\alpha_i) < \lambda$ for all $\alpha_i \in (0,1)$, we define $\alpha_i^{\lambda,\text{lower}} = \alpha_i^{\lambda,\text{upper}} = 1$; intuitively, we want to have as much of the type I error of α on the f_i part as possible since the corresponding type II error would be larger otherwise.

Define $\alpha^{\lambda,\text{lower}} = \sum_{i=1}^k p_i \alpha_i^{\lambda,\text{lower}}$, and $\alpha^{\lambda,\text{upper}} = \sum_{i=1}^k p_i \alpha_i^{\lambda,\text{upper}}$. By definition, if $\lambda = -\infty$, we have $\alpha^{\lambda,\text{upper}} = \alpha^{\lambda,\text{lower}} = 0$, and if $\lambda = +\infty$, we have $\alpha^{\lambda,\text{upper}} = \alpha^{\lambda,\text{lower}} = 1$. For $\lambda \in (-\infty, +\infty)$, we have $[0,1] \subseteq \cup_{\lambda \in (-\infty,+\infty)} [\alpha^{\lambda,\text{lower}}, \alpha^{\lambda,\text{upper}}]$. Therefore, for any $\alpha \in [0,1]$, we can find λ such that $\alpha \in [\alpha^{\lambda,\text{lower}}, \alpha^{\lambda,\text{upper}}]$, and we can determine α_i by the λ.[4]

From the procedure above, we know $f_{\min} = \text{mix}(\underline{p}, \underline{f})$, and $\mathcal{M}$ satisfies $f_{\min}$-DP. ■

We will use Lemma 31 in the proof of Theorem 11.

Lemma 31 (Equation (13) in (Dong et al., 2022)) *For a symmetric tradeoff function f, define $f_p := pf + (1-p)Id$ for $0 \leq p \leq 1$, where $Id(x) = 1 - x$. Let x^* be the unique fixed point of f, that is $f(x^*) = x^*$, we have*

$$C_p(f)(x) = \begin{cases} f_p(x), & x \in [0, x^*] \\ x^* + f_p(x^*) - x, & x \in [x^*, f_p(x^*)] \\ f_p^{-1}(x), & x \in [f_p(x^*), 1]. \end{cases}$$

Proof [Proof of Theorem 11] We are going to find a lower bound of $T_{\mathcal{M}(\texttt{boot}(D_1)),\mathcal{M}(\texttt{boot}(D_2))}$ uniformly for any neighboring data sets $D_1 = (x_1, x_2, \cdots, x_n)$ and $D_2 = (x_1', x_2, \cdots, x_n)$ (without loss of generality, we let x_1 be the different data point in D_1 and D_2). We use $\texttt{boot}^i$ to denote the conditional bootstrap subsampling where x_1 or x_1' are drawn for exactly

4. If there are multiple choices of $\{\alpha_i\}_{i=1}^k$, all of them correspond to the same β.

i times from D_1 or D_2. Note that both $\texttt{boot}^i(D_1)$ and $\texttt{boot}^i(D_2)$ are random variables for any $i = 0, 1, 2, \cdots, n$. Furthermore, we define $\texttt{boot}^>$ to denote the conditional bootstrap subsampling where x_1 or x_1' is drawn for at least once from D_1 or D_2.

From Theorem 10, we know $\text{mix}(\{(q_i, g)\}_{i\in I}) = g$ for any g, I and q_i. Therefore, we have $T_{\mathcal{M}(\texttt{boot}^0(D_1)),\mathcal{M}(\texttt{boot}^0(D_2))}(\alpha) = f_0$ where $f_0(\alpha) = 1 - \alpha$ since $\texttt{boot}^0(D_1) = \texttt{boot}^0(D_2)$, and we also have $T_{\mathcal{M}(\texttt{boot}^k(D_1)),\mathcal{M}(\texttt{boot}^k(D_2))} \geq f_k$ since $\texttt{boot}^k(D_1)$ and $\texttt{boot}^k(D_2)$ are neighboring data sets with respect to group size k. Now we consider $\mathcal{M} \circ \texttt{boot}^>$ as a mixture of $\mathcal{M} \circ \texttt{boot}^i$. Using Theorem 10 again, we have $T_{\mathcal{M}(\texttt{boot}^>(D_1)),\mathcal{M}(\texttt{boot}^>(D_2))} \geq f_>$ where $f_> := \text{mix}(\{(\frac{p_i}{1-p_0}, f_i)\}_{i=1}^n)$ (here we use $\frac{p_i}{1-p_0}$ instead of p_i because $\sum_{i=1}^n \frac{p_i}{1-p_0} = 1$).

In order to obtain a better lower bound of $T_{\mathcal{M}(\texttt{boot}(D_1)),\mathcal{M}(\texttt{boot}(D_2))}$, we find the mixture of $f_>$ and f_0 with considering the bootstrap resampling context because $\mathcal{M} \circ \texttt{boot}$ is a mixture of $\mathcal{M} \circ \texttt{boot}^>$ and $\mathcal{M} \circ \texttt{boot}^0$.

Consider a rejection rule ψ. Let

$$\begin{aligned}
&\alpha := \mathbb{E}_{\mathcal{M}(\texttt{boot}(D_1))}\psi, && \beta := \mathbb{E}_{\mathcal{M}(\texttt{boot}(D_2))}(1-\psi),\\
&\alpha_0 := \mathbb{E}_{\mathcal{M}(\texttt{boot}^0(D_1))}\psi = \mathbb{E}_{\mathcal{M}(\texttt{boot}^0(D_2))}\psi, && \beta_0 := \mathbb{E}_{\mathcal{M}(\texttt{boot}^0(D_2))}(1-\psi) = 1-\alpha_0,\\
&\alpha_> := \mathbb{E}_{\mathcal{M}(\texttt{boot}^>(D_1))}\psi, && \beta_> := \mathbb{E}_{\mathcal{M}(\texttt{boot}^>(D_2))}(1-\psi).
\end{aligned}$$

We prove that when calculating the mixture of $f_>$ and f_0, there are two additional constraints, $\beta_0 \geq f_>(\alpha_>)$ and $\beta_> \geq f_>(\alpha_0)$, as the hypothesis testing between $\mathcal{M}(\texttt{boot}^0(D_1))$ and $\mathcal{M}(\texttt{boot}^>(D_1))$ is similar to the one between $\mathcal{M}(\texttt{boot}^>(D_2))$ and $\mathcal{M}(\texttt{boot}^>(D_1))$.

We consider the hypothesis testing between $\mathcal{M}(\texttt{boot}^0(D_1))$ and $\mathcal{M}(\texttt{boot}^>(D_1))$. Similar to the idea in (Balle et al., 2018), we consider replacing each x_1 in $\mathcal{M}(\texttt{boot}^>(D_1))$ with a data point independently and uniformly drawn from $(x_2, x_3, \ldots, x_n)$. We denote the distribution of $\texttt{boot}^>(D_1)$ as ω_1, the distribution of $\texttt{boot}^0(D_1)$ as ω_0, the distribution of $\texttt{boot}(D_1)$ as $\omega_{0\&1}$, and the replacement procedure as $\texttt{replace}$. Since Balle et al. (2018) did not provide proof for why this replacement procedure can transform ω_1 to ω_0, i.e., $\texttt{replace}(\omega_1) = \omega_0$, we prove it below.[5]

For one element in ω_1, let its histogram be $h^> = (h_1, h_2, \cdots, h_n)$ where h_i is the number of occurrences of x_i in this element, $h_1 \geq 1$, $h_i \geq 0$ for $i = 2, 3, \ldots, n$, $\sum_{i=1}^n h_i = n$. For one element in ω_0, we let its histogram be $h^0 = (h_1 = 0, h_2, \cdots, h_n)$. Then

$$\begin{aligned}
P_{\omega_{0\&1}}(H = h) &= \frac{1}{n^n}\binom{n}{h_1\ h_2\ \cdots\ h_n},\\
P_{\omega_0}(H = h^0) &= \frac{1}{(n-1)^n}\binom{n}{h_2\ h_3\ \cdots\ h_n},\\
P_{\omega_1}(H = h^>) &= \frac{1}{1-\left(1-\frac{1}{n}\right)^n}\frac{1}{n^n}\binom{n}{h_1\ h_2\ \cdots\ h_n} = \frac{1}{n^n-(n-1)^n}\binom{n}{h_1\ h_2\ \cdots\ h_n}.
\end{aligned}$$

For the replacement procedure, we replace the h_1 replicates of x_1 in $h^>$ with elements independently and uniformly drawing from $(x_2, x_3, \cdots, x_n)$, where the histogram of the replacement is $h' = (h_2', h_3', \cdots, h_n')$.

5. The replacement procedure is defined deliberately: if it is defined to be replacing all x_1 with the same data point uniformly randomly drawn from $(x_2, x_3, \ldots, x_n)$, e.g., x_1 are all replaced by x_2, one can verify that this procedure will not transform ω_1 to ω_0 even for $n = 3$.

Since for any element in ω_0, the replacement does not change it, i.e., $\texttt{replace}(\omega_0) = \omega_0$, we can perform the replacement on $\omega_{0\&1}$ and show $\texttt{replace}(\omega_{0\&1}) = \omega_0$, then we also have

$$\texttt{replace}(\omega_1) = \texttt{replace}\left(\frac{\omega_{0\&1} - p_0\omega_0}{1-p_0}\right) = \frac{(\texttt{replace}(\omega_{0\&1}) - p_0\texttt{replace}(\omega_0))}{1-p_0} = \omega_0.$$

We prove $\texttt{replace}(\omega_{0\&1}) = \omega_0$ below.

$$\begin{aligned}
&P_{\texttt{replace}(\omega_{0\&1})}(H^0 = h^0) = \sum_{h'} P_{\texttt{replace}(h_1)}(H' = h') \cdot P_{\omega_{0\&1}}(H = (h_1, h^0 - h')) \\
&= \sum_{h_1=0}^{n} \frac{1}{(n-1)^{h_1}} \frac{1}{n^n} \sum_{\{h' | \sum_{i=2}^n h'_i = h_1, h'_i \le h_i \forall i\}} \frac{n!}{(h'_2!(h_2-h'_2)!)\cdots(h'_n!(h_n-h'_n)!)} \\
&= \frac{n!}{h_2!h_3!\cdots h_n!}\frac{1}{n^n} \sum_{h'_2=0}^{h_2} \cdots \sum_{h'_n=0}^{h_n} \frac{h_2!}{(n-1)^{h'_2}h'_2!(h_2-h'_2)!} \cdots \frac{h_n!}{(n-1)^{h'_n}h'_n!(h_n-h'_n)!} \\
&= \frac{n!}{h_2!h_3!\cdots h_n!}\frac{1}{n^n} \left(\sum_{h'_2=0}^{h_2} \frac{h_2!}{(n-1)^{h'_2}h'_2!(h_2-h'_2)!}\right) \cdots \left(\sum_{h'_n 0}^{h_n} \frac{h_n!}{(n-1)^{h'_n}h'_n!(h_n-h'_n)!}\right) \\
&= \frac{n!}{h_2!h_3!\cdots h_n!}\frac{1}{n^n} \left(\frac{n}{n-1}\right)^{h_2} \left(\frac{n}{n-1}\right)^{h_3} \cdots \left(\frac{n}{n-1}\right)^{h_n} = \frac{n!}{h_2!h_3!\cdots h_n!}\frac{1}{(n-1)^n} \\
&\Rightarrow \texttt{replace}(\omega_{0\&1}) = \omega_0.
\end{aligned}$$

We are ready to prove $\beta_0 \ge f_>(\alpha_>)$ and $\beta_> \ge f_>(\alpha_0)$. From the result $\texttt{replace}(\omega_1) = \omega_0$, if we consider the tradeoff function between the two distributions $\mathcal{M}(\texttt{boot}^0(D_1))$ and $\mathcal{M}(\texttt{boot}^>(D_1))$, we can break each of the two mixture distributions into parts following the replacement procedure so that there is a one-to-one mapping between the parts. When the number of occurrence of x_1 is h_1 in the outcome of $\texttt{boot}^>(D_1)$, two parts in such a mapping pair have distance h_1 (between $\texttt{boot}^{h_1}(D_1)$ and $\texttt{replace}(\texttt{boot}^{h_1}(D_1))$), so the tradeoff function between $\mathcal{M}(\texttt{boot}^{h_1}(D_1))$ and $\mathcal{M}(\texttt{replace}(\texttt{boot}^{h_1}(D_1)))$ is f_{h_1}. We use Theorem 10 to obtain the mixture of those f_{h_1}. Since f_{h_1} only depends on h_1, its corresponding probability in $\texttt{boot}^>(D_1)$ is $\frac{p_{h_1}}{1-p_0}$. Therefore, the tradeoff function between $\mathcal{M}(\texttt{boot}^>(D_1))$ and $\mathcal{M}(\texttt{boot}^0(D_1))$ is $\text{mix}(\{(\frac{p_i}{1-p_0}, f_i)\}_{i=1}^n) = f_>$. Recall that

$$\begin{aligned}
&\alpha_0 := \mathbb{E}_{\mathcal{M}(\texttt{boot}^0(D_1))}\psi = \mathbb{E}_{\mathcal{M}(\texttt{boot}^0(D_2))}\psi, \quad \beta_0 := \mathbb{E}_{\mathcal{M}(\texttt{boot}^0(D_1))}(1-\psi) = 1-\alpha_0, \\
&\alpha_> := \mathbb{E}_{\mathcal{M}(\texttt{boot}^>(D_1))}\psi, \qquad\qquad\qquad\quad \beta_> := \mathbb{E}_{\mathcal{M}(\texttt{boot}^>(D_2))}(1-\psi),
\end{aligned}$$

we have $\beta_0 \ge f_>(\alpha_>)$. Similarly, $\beta_> \ge f_>(\alpha_0)$.

Now we have established the additional constraints for the mixture of f_0 and $f_>$. We are ready to derive the final mixture tradeoff function. Notice that $\alpha = p_0\alpha_0 + (1-p_0)\alpha_>$ and $\alpha_0, \alpha_> \in [0,1]$: For $\alpha = 0$ and $\alpha = 1$, we have $\alpha_0 = \alpha_> = 0$ and $\alpha_0 = \alpha_> = 1$ respectively.

Now we consider the constrained optimization problem for $\alpha \in (0,1)$ where we replace $\alpha_>$ and β_0 with $\frac{\alpha - p_0\alpha_0}{1-p_0}$ and $1-\alpha_0$ respectively: $f_{\min}(\alpha) = \min\{p_0(1-\alpha_0) + (1-p_0)\beta_> \mid \beta_> \le$

$1, \alpha_0 \geq 0, \beta_> \geq f_>(\alpha_0), \beta_> \geq f_>(\frac{\alpha - p_0\alpha_0}{1-p_0}), 1-\alpha_0 \geq f_>(\frac{\alpha-p_0\alpha_0}{1-p_0})\}$. We ignore the constraint $\alpha_0 \leq 1$ and $\beta_> \geq 0$ because they can be derived from $1-\alpha_0 \geq f_>(\frac{\alpha-p_0\alpha_0}{1-p_0})$ and $\beta_> \geq f_>(\alpha_0)$ respectively. Since $f_> : \mathbb{R} \mapsto \mathbb{R}$ is a convex function, we use the Karush–Kuhn–Tucker theorem to solve to the convex optimization problem: let the Lagrangian function be $L(\alpha_0, \beta_>, \mu_1, \mu_2, \mu_3, \mu_4, \mu_5) = p_0(1-\alpha_0) + (1-p_0)\beta_> + \mu_1(\beta_> - 1) + \mu_2(-\alpha_0) + \mu_3(f_>(\alpha_0) - \beta_>) + \mu_4\left(f_>\left(\frac{\alpha-p_0\alpha_0}{1-p_0}\right) - \beta_>\right) + \mu_5\left(f_>\left(\frac{\alpha-p_0\alpha_0}{1-p_0}\right) - 1 + \alpha_0\right)$, and the minimum of $p_0(1-\alpha_0) + (1-p_0)\beta_>$ is achieved at $(\alpha_0, \beta_>)$ if and only if the following conditions are satisfied:
Stationarity: $0 \in -p_0 - \frac{\mu_2 p_0}{1-p_0} + \mu_3 \partial f_>(\alpha_0) - \frac{\mu_4 p_0}{1-p_0}\partial f_>(\frac{\alpha-p_0\alpha_0}{1-p_0}) + \mu_5(-\frac{p_0}{1-p_0}\partial f_>(\frac{\alpha-p_0\alpha_0}{1-p_0}) + 1)$ and $0 = (1-p_0) + \mu_1 - \mu_3 - \mu_4$.
Primal feasibility: $1 \geq \beta_> \geq f_>(\alpha_0),\ \beta_> \geq f_>(\frac{\alpha-p_0\alpha_0}{1-p_0}),\ 1 - f_>(\frac{\alpha-p_0\alpha_0}{1-p_0}) \geq \alpha_0 \geq 0$.
Dual feasibility: $\mu_i \geq 0,\ i = 1, 2, 3, 4, 5$.
Complementary slackness: $\mu_1(\beta_> - 1) + \mu_2(-\alpha_0) + \mu_3(f_>(\alpha_0) - \beta_>) + \mu_4(f_>(\frac{\alpha-p_0\alpha_0}{1-p_0}) - \beta_>) + \mu_5(f_>(\frac{\alpha-p_0\alpha_0}{1-p_0}) - 1 + \alpha_0) = 0$.
As our constrained optimization problem has both the convex objective function and convex constraints, we only need to find a solution of $\alpha_0, \beta_>$ satisfying the KKT conditions, and this solution will be a minimizer of the problem given fixed α.

To simplify our analysis, first, we show that the solution satisfies $\alpha_0 \geq \alpha$. This is because when $\alpha_0 < \alpha$, there is always another choice, $(\alpha_0' = \alpha, \beta_>' = f_>(\alpha))$, achieving a lower value of the objective function: Notice that $\alpha = p_0\alpha_0 + (1-p_0)\alpha_>$ where $p_0 \in (0,1)$. Therefore, if $\alpha_0 < \alpha$, we have $\alpha < \alpha_>$, and $\beta_> \geq f_>(\alpha_0) \geq f_>(\alpha)$ where the first inequality holds from previous analysis, and the second inequality holds due to $f_>$ being decreasing. Therefore,

$$\begin{aligned} p_0(1-\alpha_0) + (1-p_0)\beta_> &= p_0(1-\alpha) + (1-p_0)f_>(\alpha) + p_0(\alpha - \alpha_0) + (1-p_0)(\beta_> - f_>(\alpha)) \\ &> p_0(1-\alpha) + (1-p_0)f_>(\alpha), \end{aligned}$$

which means that the choice $(\alpha_0' = \alpha, \beta_>' = f_>(\alpha))$ satisfies all the constraints and also achieves a value of the objective function lower than the choice $(\alpha_0, \beta_>)$. This contradicts the optimality of $(\alpha_0, \beta_>)$. Therefore, $\alpha_0 \geq \alpha > 0$.

Similarly, we show that we only need to consider $\beta_> < 1$. If $\beta_> = 1$, since $\alpha_0 = \frac{\alpha - (1-p_0)\alpha_>}{p_0} \leq \frac{\alpha}{p_0}$, we have the objective function value $p_0(1-\alpha_0) + (1-p_0)\beta_> \geq 1-\alpha$. Now we consider another choice $(\alpha_0' = \alpha, \beta_>' = f_>(\alpha))$. We know that $\beta_>' = f_>(\alpha) \leq 1-\alpha < 1$, and the objective function value $p_0(1-\alpha) + (1-p_0)\beta_>' \leq p_0(1-\alpha) + (1-p_0)(1-\alpha) = 1-\alpha$ which means that the new choice $(\alpha_0' = \alpha, \beta_>' = f_>(\alpha))$ is never worse than the choice of $(\alpha_0, \beta_> = 1)$. Therefore, for $f_{\min}(\alpha)$ where $\alpha \in (0,1)$, we only need to consider $\beta_> < 1$.

With $\alpha_0 \geq \alpha > 0, \beta_> < 1$, from the complementary slackness, we know $\mu_1 = \mu_2 = 0$.

By Proposition 30, there is exactly one $\bar{\alpha}$ satisfying $\bar{\alpha} = f_>(\bar{\alpha})$ and $-1 \in \partial f_>(\bar{\alpha})$. We denote $C \in \partial f_>(\alpha)$, and $C' \in \partial f_>(\alpha)|_{\alpha = \frac{\alpha - p_0\alpha_0}{1-p_0}}$.

- If $0 < \alpha \leq \bar{\alpha}$, we verify that $(\alpha_0 = \alpha, \beta_> = f_>(\alpha), \mu_3 = (1-p_0)(p_0 + p_0/C), \mu_4 = (1-p_0)(1-p_0-p_0/C), \mu_5 = 0)$ satisfies the KKT conditions which means $(\alpha_0 = \alpha, \beta_> = f_>(\alpha))$ is a minimizer. First, we see that $\frac{\alpha - p_0\alpha_0}{1-p_0} = \alpha$.
  - The primal feasibility is satisfied because $f_>(\alpha) \leq 1-\alpha$.

- The complementary slackness is satisfied as $\mu_1 = \mu_2 = \mu_5 = 0$ and $f_>(\alpha) = \beta_>$.
- Since $0 < \alpha \le \bar{\alpha}$, we know that $C \le -1$. The dual feasibility conditions, $\mu_3 \ge 0$ and $\mu_4 > 0$, hold because $p_0 \in (0, 1)$.
- The stationarity conditions also hold when we plug in the value C for $\partial f_>(\alpha)$.

- If $\bar{\alpha} < \alpha \le (1-p_0)\bar{\alpha} + p_0(1-\bar{\alpha})$, we verify that $(\alpha_0 = \frac{\alpha-(1-p_0)\bar{\alpha}}{p_0}, \beta_> = \bar{\alpha}, \mu_3 = 0, \mu_4 = 1-p_0, \mu_5 = 0)$ satisfies the KKT conditions. First, we see that $\frac{\alpha-p_0\alpha_0}{1-p_0} = \bar{\alpha}$.
  - The primal feasibility is satisfied: since $\alpha > \bar{\alpha}$, we have $\alpha_0 > \bar{\alpha}$ and $f_>(\alpha_0) \le f_>(\bar{\alpha}) = \beta_>$; since $\alpha \le (1-p_0)\bar{\alpha} + p_0(1-\bar{\alpha})$, we have $\alpha_0 \le 1-\bar{\alpha}$, and $1 - f_>(\bar{\alpha}) = 1 - \bar{\alpha} \ge \alpha_0$.
  - The complementary slackness is satisfied as $\mu_1 = \mu_2 = \mu_3 = \mu_5 = 0, f_>(\bar{\alpha}) = \beta_>$.
  - The dual feasibility condition, $\mu_4 > 0$, holds because $p_0 \in (0, 1)$.
  - The stationarity conditions also hold: We plug in the value -1 for $\partial f_>(\bar{\alpha})$, then the right-hand sides of both conditions are 0.

- If $(1-p_0)\bar{\alpha} + p_0(1-\bar{\alpha}) < \alpha < p_0 + (1-p_0)f_>(0)$, we let α_0^* be the solution of $p_0\alpha_0 + (1-p_0)f_>(1-\alpha_0) = \alpha$. Since $f_>$ is continuous and not increasing, we know $g(\alpha_0) = p_0\alpha_0 + (1-p_0)f_>(1-\alpha_0)$ is continuous and strictly increasing with respect to α_0. We know $g(1) = p_0 + (1-p_0)f_>(0)$ and $g(1-\bar{\alpha}) = (1-p_0)\bar{\alpha} + p_0(1-\bar{\alpha})$. Therefore, $p_0\alpha_0 + (1-p_0)f_>(1-\alpha_0) = \alpha$ has only one solution which is α_0^* with $1 > \alpha_0^* > 1-\bar{\alpha}$. Now we verify that $(\alpha_0 = \alpha_0^*, \beta_> = 1-\alpha_0^*, \mu_3 = 0, \mu_4 = 1-p_0, \mu_5 = \frac{p_0(1+C')}{1-\frac{p_0C'}{1-p_0}})$ satisfies the KKT conditions. First, we see that $\frac{\alpha-p_0\alpha_0}{1-p_0} = f_>(1-\alpha_0) > f_>(\bar{\alpha}) = \bar{\alpha}$ and $\frac{\alpha-p_0\alpha_0}{1-p_0} = f_>(1-\alpha_0) < f_>(0)$. Therefore, $0 > C' \ge -1$.
  - The primal feasibility is satisfied: since $\alpha_0 \ge \alpha \ge \bar{\alpha}$, we have $1-\alpha_0 < \bar{\alpha}$. Therefore, by Proposition 30, since $f_>$ is convex and symmetric, there exists $C \le -1$ such that $C \in \partial f_>(1-\alpha_0)$ and $\beta_> = 1-\alpha_0 = f_>(f_>(1-\alpha_0)) = f_>(\frac{\alpha-p_0\alpha_0}{1-p_0})$. We also have $\beta_> = 1-\alpha_0 \ge f_>(\alpha_0)$.
  - The complementary slackness is also satisfied because $\mu_1 = \mu_2 = \mu_3 = 0$ and $\beta_> = 1-\alpha_0 = f_>(\frac{\alpha-p_0\alpha_0}{1-p_0})$.
  - The dual feasibility conditions hold as $p_0 \in (0,1), C' \in [-1, 0)$.
  - The stationarity conditions hold: We plug in the value C' for $\partial f_>(\alpha)|_{\alpha=\frac{\alpha-p_0\alpha_0}{1-p_0}}$, then the right hand side is 0.

- If $\alpha \ge p_0 + (1-p_0)f_>(0)$, we let $(\alpha_0 = 1, \beta_> = 0, \mu_5 = p_0, \mu_3 = (1-p_0)/2, \mu_4 = (1-p_0)/2)$. Notice that $\frac{\alpha-p_0\alpha_0}{1-p_0} \ge f_>(0)$. Therefore, $f_>(\frac{\alpha-p_0\alpha_0}{1-p_0}) = 0$.
  - The primal feasibility is satisfied.
  - The complementary slackness is satisfied because $f_>(\alpha_0) - \beta_> = 0,\ f_>(\frac{\alpha-p_0\alpha_0}{1-p_0}) - \beta_> = 0,\ f_>\left(\frac{\alpha-p_0\alpha_0}{1-p_0}\right) - 1 + \alpha_0 = 0.$
  - The dual feasibility conditions hold because $p_0 \in (0, 1)$.

– The stationarity conditions hold: We use $0 \in \partial f_>(\alpha_0)$ and $0 \in \partial f_>(\alpha)|_{\alpha=\frac{\alpha-p_0\alpha_0}{1-p_0}}$ to plug in the first condition, and it holds because $-p_0 + \mu_5 = 0$. The second condition holds because $(1-p_0) - \mu_3 - \mu_4 = 0$.

Therefore, we have the tradeoff function $f_{\min}$ from $f_>$ as follows

$$f_{\min}(\alpha) = \begin{cases} p_0 + (1-p_0)f_>(0), & \text{if } \alpha = 0 \\ p_0(1-\alpha) + (1-p_0)f_>(\alpha), & \text{if } 0 < \alpha \le \bar{\alpha} \\ p_0 - \alpha + 2(1-p_0)\bar{\alpha}, & \text{if } \bar{\alpha} < \alpha \le (1-p_0)\bar{\alpha} + p_0(1-\bar{\alpha}) \\ 1 - \alpha_0^*, & \text{if } (1-p_0)\bar{\alpha} + p_0(1-\bar{\alpha}) < \alpha < p_0 + (1-p_0)f_>(0) \\ 0, & \text{if } \alpha \ge p_0 + (1-p_0)f_>(0) \end{cases}$$

where $\bar{\alpha}$ satisfies $f_>(\bar{\alpha}) = \bar{\alpha}$ and α_0^* is the only solution of $p_0\alpha_0 + (1-p_0)f_>(1-\alpha_0) = \alpha$ with respect to α_0.

Now we verify the $f_{\min}$ above is the same as $C_{1-p_0}(f_>)$. In Lemma 31, we replace p and f with $1-p_0$ and $f_>$ respectively:

- For $\alpha \in [0, \bar{\alpha}]$, we have $f_{\min}(\alpha) = (f_>)_{1-p_0}(\alpha) = C_{1-p_0}(f_>)(\alpha)$.
- For $\alpha \in [\bar{\alpha}, (f_>)_{1-p_0}(\bar{\alpha})]$, we have $f_{\min}(\alpha) = p_0 - \alpha + 2(1-p_0)\bar{\alpha} = \bar{\alpha} + ((1-p_0)\bar{\alpha} + p_0(1-\bar{\alpha})) - \alpha = \bar{\alpha} + (f_>)_{1-p_0}(\bar{\alpha}) - \alpha = C_{1-p_0}(f_>)(\alpha)$.
- For $\alpha \in [(f_>)_{1-p_0}(\bar{\alpha}), p_0+(1-p_0)f_>(0)]$, since α_0^* satisfies $p_0\alpha_0+(1-p_0)f_>(1-\alpha_0) = \alpha$, we let $t = 1-\alpha_0^*$, and we have $(f_>)_{1-p_0}(t) = (1-p_0)f_>(t) + p_0(1-t) = \alpha$. Since $f_>$ is non-increasing, we have that $(f_>)_{1-p_0}(t)$ is strictly decreasing because $p_0 \in (0,1)$. Therefore, $(f_>)^{-1}_{1-p_0}(\alpha) = 1 - \alpha_0^* = f_{\min}(\alpha)$.
- For $\alpha \in [p_0+(1-p_0)f_>(0), 1]$, we know that $\alpha \ge (f_>)_{1-p_0}(0)$. Since $(f_>)_{1-p_0}$ is strictly decreasing, we have $(f_>)_{1-p_0}(t) \le \alpha \forall t \in [0,1]$. Therefore, $(f_>)^{-1}_{1-p_0}(\alpha) = \inf\{t \in [0,1] : (f_>)_{1-p_0}(t) \le \alpha\} = \inf\{t \in [0,1]\} = 0$. We have $(f_>)^{-1}_{1-p_0}(\alpha) = 0 = f_{\min}(\alpha)$.

■

A.2 Derive (ε, δ)-DP from f-DP for DP Bootstrap

In this section, we transform our Theorem 11 to (Balle et al., 2018, Theorem 10).

Proposition 32 *The (ε, δ)-DP result in Theorem 25 can be derived from Theorem 11.*

Proof

We use the primal-dual transformation (see Proposition 6) to obtain the (ε', δ')-DP result by our $f_{\mathcal{M}\circ\texttt{boot}}$-DP result: $\delta'(\varepsilon') = 1 + f^*_{\mathcal{M}\circ\texttt{boot}}(-e^{\varepsilon'})$.

Since f is convex function when f is a tradeoff function, we have $\alpha y - f(\alpha) \leq \alpha_0 y - f(\alpha_0)$ for any α and (α_0, y) satisfying $y \in \partial f(\alpha_0)$. Therefore, we have $f^*(y) = \sup_\alpha \alpha y - f(\alpha) = \alpha_0 y - f(\alpha_0)$ where $y \in \partial f(\alpha_0)$, and $1-\delta(\varepsilon) = -f^*(-\mathrm{e}^\varepsilon) = \mathrm{e}^\varepsilon \alpha_\varepsilon + f(\alpha_\varepsilon)$ where $-\mathrm{e}^\varepsilon \in \partial f(\alpha_\varepsilon)$.

We let $f_> = \mathrm{mix}(\{(\frac{p_i}{1-p_0}, f_i)\}_{i=1}^n)$, and our bootstrap privacy guarantee is $f_{\mathcal{M}\circ\mathtt{boot}} = C_{1-p_0}(f_>)$. We let $\bar{\alpha}$ be the solution of $\alpha = f_>(\alpha)$. By Proposition 30 and Lemma 31, we have $-1 \in \partial f_>(\bar{\alpha})$ and $-1 \in \partial f_{\mathcal{M}\circ\mathtt{boot}}(\bar{\alpha})$. Since $-\mathrm{e}^{\varepsilon'} \leq -1$ for any $\varepsilon' \geq 0$, we can find $\alpha_{\varepsilon'} \leq \bar{\alpha}$ that $-\mathrm{e}^{\varepsilon'} \in \partial f_{\mathcal{M}\circ\mathtt{boot}}(\alpha_{\varepsilon'})$. By Lemma 31, we also have $-\mathrm{e}^{\varepsilon'} \in \partial((f_>)_{1-p_0})(\alpha_{\varepsilon'})$ and $f_{\mathcal{M}\circ\mathtt{boot}}(\alpha_{\varepsilon'}) = C_{1-p_0}(f_>)(\alpha_{\varepsilon'}) = (f_>)_{1-p_0}(\alpha_{\varepsilon'}) = (1-p_0)f_>(\alpha_{\varepsilon'}) + p_0(1-\alpha_{\varepsilon'})$, and

$$\begin{aligned}\delta'(\varepsilon') &= 1 + f^*_{\mathcal{M}\circ\mathtt{boot}}(-\mathrm{e}^{\varepsilon'}) = -\mathrm{e}^{\varepsilon'}\alpha_{\varepsilon'} - f_{\mathcal{M}\circ\mathtt{boot}}(\alpha_{\varepsilon'}) \\ &= -\mathrm{e}^{\varepsilon'}\alpha_{\varepsilon'} - ((f_>)_{1-p_0})(\alpha_{\varepsilon'}) = 1 + ((f_>)_{1-p_0})^*(-\mathrm{e}^{\varepsilon'})\end{aligned}$$

From the Supplement to (Dong et al., 2022) (Page 42), we know the following two equations, $\varepsilon' = \log(1-p+p\mathrm{e}^\varepsilon), \delta' = p\big(1 + f^*(-\mathrm{e}^\varepsilon)\big)$ can be re-parameterized into $\delta' = 1 + f_p^*(-\mathrm{e}^{\varepsilon'})$ where $f_p = pf + (1-p)\mathrm{Id}$. We can re-parameterize $\delta'(\varepsilon') = 1 + ((f_>)_{1-p_0})^*(-\mathrm{e}^{\varepsilon'})$ and $\varepsilon' = \log(p_0 + (1-p_0)\mathrm{e}^\varepsilon)$ into $\delta'(\varepsilon') = (1-p_0)\big(1 + f_>^*(-\mathrm{e}^\varepsilon)\big)$. Since $\varepsilon' = \log(p_0 + (1-p_0)\mathrm{e}^\varepsilon)$ is also the relationship between ε' and ε in Theorem 25, in order to prove Theorem 25, we only need to show that $(1-p_0)\big(1 + f_>^*(-\mathrm{e}^\varepsilon)\big) = \sum_{i=1}^n p_i \delta_{\mathcal{M},i}(\varepsilon)$.

As we have $f_> = \mathrm{mix}(\{(\frac{p_i}{1-p_0}, f_i)\}_{i=1}^n)$, from the construction of the mixture tradeoff function, we let $C = -\mathrm{e}^\varepsilon$ and find $\{\alpha_i\}_{i=1}^n$ such that $C \in \partial f_i(\alpha_i)$, then we know that for $\alpha = \sum_{i=1}^n \frac{p_i}{1-p_0}\alpha_i$, we also have $C \in \partial f_>(\alpha)$: This is because for any $\alpha' = \sum_{i=1}^n \frac{p_i}{1-p_0}\alpha_i'$ and $f_>(\alpha') = \sum_{i=1}^n \frac{p_i}{1-p_0} f_i(\alpha_i')$, we have $f_>(\alpha') \geq f_>(\alpha) + C(\alpha' - \alpha)$ by the fact that $C \in \partial f_i(\alpha_i)$. Therefore, we can find $\alpha_\varepsilon = \sum_{i=1}^n \frac{p_i}{1-p_0}\alpha_{\varepsilon,i}$ such that $-\mathrm{e}^\varepsilon \in \partial f_>(\alpha_\varepsilon)$ and $-\mathrm{e}^\varepsilon \in \partial f_i(\alpha_{\varepsilon,i})$ for $i = 1, 2, \ldots, n$. Then we can prove $(1-p_0)\big(1 + f_>^*(-\mathrm{e}^\varepsilon)\big) = \sum_{i=1}^n p_i \delta_{\mathcal{M},i}(\varepsilon)$ using the primal-dual transformation of $\delta_{\mathcal{M},i}(\varepsilon)$ and the fact that $(1-p_0)\big(-\mathrm{e}^\varepsilon\alpha_\varepsilon - f_>(\alpha_\varepsilon)\big) = \sum_{i=1}^n p_i\big(-\mathrm{e}^\varepsilon\alpha_{\varepsilon,i} - f_i(\alpha_{\varepsilon,i})\big)$ due to $\alpha_\varepsilon = \sum_{i=1}^n \frac{p_i}{1-p_0}\alpha_{\varepsilon,i}$ and $f_>(\alpha_\varepsilon) = \sum_{i=1}^n \frac{p_i}{1-p_0} f_i(\alpha_{\varepsilon,i})$. ■

A.3 Proofs for Section 3.2

The composition computation follows the general result in (Zheng et al., 2020), and we only need to prove the $q(x)$ in our numerical composition result which is derived from Theorem 11, Remark 9 and Lemma 31.

Proposition 33 (Numerical Computation of Composition: Zheng et al., 2020) *Let $f_1 = T(P_1, Q_1)$ and $f_2 = T(P_2, Q_2)$, and we use δ_1, δ_2, and $\delta_\otimes$ to denote the dual view for f_1, f_2, and $f_\otimes := f_1 \otimes f_2$ correspondingly. If P_i and Q_i are distributions on $x \in \mathbb{R}$ with densities $p_i(x)$ and $q_i(x)$ for $i = 1, 2$ with respect to Lebesgue measure, then $\delta_\otimes(\varepsilon) = \int_{\mathbb{R}} \delta_1(\varepsilon - L_2(x)) q_2(x)\,\mathrm{d}x$ where $L_2(x) := \log(\frac{q_2(x)}{p_2(x)})$ and $\delta_1(\varepsilon) = \int_{\mathbb{R}} \max(0, q_1(x) - \mathrm{e}^\varepsilon p_1(x))\,\mathrm{d}x$.*

Proof [Proof of Proposition 14] In this proof, we replace C with λ to avoid confusion with the subsampling function C_p in Proposition 4.

We know that $f_{\mathcal{M}\circ\texttt{boot}} = C_{1-p_0}(\text{mix}(\underline{p},\underline{f}))$, $\text{mix}(\underline{p},\underline{f}) = (\sum_{i=1}^k (\frac{p_i}{1-p_0} f_i \circ (f_i')^{-1})) \circ (\sum_{i=1}^k \frac{p_i}{1-p_0}(f_i')^{-1})^{-1}$, and

$$C_p(f)(x) = \begin{cases} f_p(x), & x \in [0, x^*] \\ x^* + f_p(x^*) - x, & x \in [x^*, f_p(x^*)] \\ f_p^{-1}(x), & x \in [f_p(x^*), 1] \end{cases}$$

where x^* is the unique fixed point of f, $f_p := pf + (1-p)\text{Id}$ for $0 \le p \le 1$, and $\text{Id}(x) = 1-x$.

As f_i is symmetric, let $x_i = (f_i')^{-1}(-1)$, then we have $f_i(x_i^*) = x_i^*$. Therefore, the choice $x^* = \sum_{i=1}^n \frac{p_i}{1-p_0}(f_i')^{-1}(-1)$ satisfies $\text{mix}(\underline{p},\underline{f})(x^*) = x^*$. We have $\text{mix}(\underline{p},\underline{f})_{1-p_0}(x^*) = (1-p_0)x^* + p_0(1-x^*)$.

Let $y = 1-x$ and $g(y) = \text{mix}(\underline{p},\underline{f})(y)$.

When $x \ge 1-x^*$, $y \le x^*$, $q(x) = -f'_{\mathcal{M}\circ\texttt{boot}}(1-x) = -f'_{\mathcal{M}\circ\texttt{boot}}(y) = -((1-p_0)g + p_0\text{Id})'(y) = -(1-p_0)g'(y) + p_0$. We let $\lambda = g'(y)$ be the slope, then $y = \sum_{i=1}^k \frac{p_i}{1-p_0}(f_i')^{-1}(\lambda)$ is the corresponding type I error.

When $1-p_0-(1-2p_0)x^* \le x < 1-x^*$, $x^* < y \le p_0 + (1-2p_0)x^* = \text{mix}(\underline{p},\underline{f})_{1-p_0}(x^*)$. $q(x) = -f'_{\mathcal{M}\circ\texttt{boot}}(1-x) = -f'_{\mathcal{M}\circ\texttt{boot}}(y) = 1$.

When $1-p_0-(1-2p_0)x^* > x \ge 0$, $y > \text{mix}(\underline{p},\underline{f})_{1-p_0}(x^*)$. $q(x) = -f'_{\mathcal{M}\circ\texttt{boot}}(1-x) = -f'_{\mathcal{M}\circ\texttt{boot}}(y) = -(((1-p_0)g + p_0\text{Id})^{-1})'(y) = -\frac{1}{(1-p_0)g'(((1-p_0)g+p_0\text{Id})^{-1}(y))-p_0}$. We let $\lambda = 1/g'(((1-p_0)g+p_0\text{Id})^{-1}(y))$. Then $q(x) = 1/(p_0-(1-p_0)/\lambda)$. And by Proposition 30.4, we have $(g')^{-1}(1/\lambda) = g((g')^{-1}(\lambda))$ since g is a symmetric tradeoff function. Therefore,

$$\begin{aligned} y &= ((1-p_0)g + p_0\text{Id})((g')^{-1}(1/\lambda)) = (1-p_0)(g')^{-1}(\lambda) + p_0(1-(g')^{-1}(1/\lambda)) \\ &= (1-p_0)\sum_{i=1}^n \frac{p_i}{1-p_0}(f_i')^{-1}(\lambda) + p_0\left(1 - \sum_{i=1}^n \frac{p_i}{1-p_0}(f_i')^{-1}(1/\lambda)\right). \end{aligned}$$

The computation of the privacy profile of $f_1 \otimes \cdots \otimes f_B$ is based on Proposition 33. ■

To prove our asymptotic composition result, we first restate the results in (Dong et al., 2022) for comparing the results of $f_>$ and $C_{1-p_0}(f_>)$ which are used in our proof.

Theorem 34 (Berry-Esseen CLT for Composition of f-DP: Dong et al., 2022) *Define* $\text{kl}(f) := -\int_0^1 \log|f'(x)|\,\text{d}x$, $\kappa_2(f) := \int_0^1 \log^2|f'(x)|\,\text{d}x$, $\kappa_3(f) := \int_0^1 \big|\log|f'(x)|\big|^3\,\text{d}x$, $\bar\kappa_3(f) := \int_0^1 \big|\log|f'(x)| + \text{kl}(f)\big|^3\,\text{d}x$. *Let* $f_1,\ldots,f_n$ *be symmetric tradeoff functions such that* $\kappa_3(f_i) < \infty$ *for all* $1 \le i \le n$. *Denote* $\mathbf{kl} := (\text{kl}(f_1),\ldots,\text{kl}(f_n))$, $\boldsymbol{\kappa_2} := (\kappa_2(f_1),\ldots,\kappa_2(f_n))$, $\bar{\boldsymbol{\kappa}}_{\mathbf{3}} := (\bar\kappa_3(f_1),\ldots,\bar\kappa_3(f_n))$, $\mu := \frac{2\|\mathbf{kl}\|_1}{\sqrt{\|\boldsymbol{\kappa_2}\|_1 - \|\mathbf{kl}\|_2^2}}$, $\gamma := \frac{0.56\|\bar{\boldsymbol{\kappa}}_{\mathbf{3}}\|_1}{\left(\|\boldsymbol{\kappa_2}\|_1 - \|\mathbf{kl}\|_2^2\right)^{3/2}}$ *and assume* $\gamma < \frac{1}{2}$. *Then, for all* $\alpha \in [\gamma, 1-\gamma]$, *we have*

$$G_\mu(\alpha+\gamma) - \gamma \leqslant f_1 \otimes f_2 \otimes \cdots \otimes f_n(\alpha) \leqslant G_\mu(\alpha-\gamma) + \gamma. \tag{1}$$

Let $\{f_{ni} : 1 \leqslant i \leqslant n\}_{n=1}^\infty$ *be a triangular array of symmetric tradeoff functions and assume for some* $K \ge 0$ *and* $s > 0$ *that* $\lim_{n\to\infty}\sum_{i=1}^n \text{kl}(f_{ni}) = K$, $\lim_{n\to\infty}\max_{1\leqslant i\leqslant n}\text{kl}(f_{ni}) = 0$,

$\lim_{n\to\infty}\sum_{i=1}^n \kappa_2(f_{ni}) = s^2$, $\lim_{n\to\infty}\sum_{i=1}^n \kappa_3(f_{ni}) = 0$. *Then, uniformly for all* $\alpha \in [0,1]$, *we have*

$$\lim_{n\to\infty} f_{n1} \otimes f_{n2} \otimes \cdots \otimes f_{nn}(\alpha) = G_{2K/s}(\alpha).$$

Lemma 35 (Lemma F.2 and F.3 in the Supplement to Dong et al., 2022) *Let* f *be a symmetric tradeoff function with* $f(0) = 1$ *and* x^* *be its unique fixed point. Then*

$$\begin{aligned}
\mathrm{kl}(f) &= \int_0^{x^*} \big(|f'(x)| - 1\big) \log|f'(x)|\,\mathrm{d}x \\
\kappa_2(f) &= \int_0^{x^*} \big(|f'(x)| + 1\big) \big(\log|f'(x)|\big)^2\,\mathrm{d}x \\
\bar{\kappa}_3(f) &= \int_0^{x^*} \Big(\big|\log|f'(x)| + \mathrm{kl}(f)\big|^3 + |f'(x)| \cdot \big|\log|f'(x)| - \mathrm{kl}(f)\big|^3\Big)\,\mathrm{d}x \\
\kappa_3(f) &= \int_0^{x^*} \big(|f'(x)| + 1\big) \big(\log|f'(x)|\big)^3\,\mathrm{d}x.
\end{aligned}$$

Let $g(x) = -f'(x) - 1 = |f'(x)| - 1$. *Then*

$$\begin{aligned}
\mathrm{kl}(C_p(f)) &= p\int_0^{x^*} g(x) \log\big(1 + pg(x)\big)\,\mathrm{d}x \\
\kappa_2(C_p(f)) &= \int_0^{x^*} \big(2 + pg(x)\big)\big[\log\big(1 + pg(x)\big)\big]^2\,\mathrm{d}x \\
\kappa_3(C_p(f)) &= \int_0^{x^*} \big(2 + pg(x)\big)\big[\log\big(1 + pg(x)\big)\big]^3\,\mathrm{d}x.
\end{aligned}$$

For DP bootstrap with Gaussian mechanism, we can apply our Theorem 11 as follows.

Corollary 36 *Let* $\mathcal{M} : \mathcal{X}^n \to \mathcal{Y}$ *satisfy* μ*-GDP. Then* $\mathcal{M} \circ \texttt{boot}$ *satisfies* $f_{\texttt{boot}}$*-DP where* $f_{\texttt{boot}} = C_{1-p_0}(f_>)$, $f_> = \mathrm{mix}(\{\frac{p_i}{1-p_0}, f_i\}_{i=1}^n)$, $p_i = \binom{n}{i}(1/n)^i(1-1/n)^{n-i}$, $f_i = G_{i\mu}$.

Now we provide the exact representation of $f_>$ in Corollary 36.

Lemma 37 *Let* $f_> = \mathrm{mix}(\{\frac{p_i}{1-p_0}, f_i\}_{i=1}^n)$ *where* $p_i = \binom{n}{i}(1/n)^i(1-1/n)^{n-i}$, $f_i = G_{i\mu}$. *Then* $f_>$ *is the tradeoff function between* $\sum_{i=1}^n \frac{p_i}{1-p_0}\mathcal{N}\left(-\frac{i^2\mu^2}{2}, i^2\mu^2\right)$ *and* $\sum_{i=1}^n \frac{p_i}{1-p_0}\mathcal{N}\left(\frac{i^2\mu^2}{2}, i^2\mu^2\right)$.

Proof [Proof of Lemma 37] First, we know that $f_i(\alpha) = G_{i\mu}(\alpha) = \Phi(\Phi^{-1}(1-\alpha) - i\mu)$ where Φ is the CDF of the standard normal distribution. We first obtain the subdifferential of f_i: $\frac{\mathrm{d}f_i(\alpha_i)}{\alpha_i} = -\exp(-i^2\mu^2/2 + i\mu\Phi^{-1}(1-\alpha_i))$. We let the type I error and type II error in f_i be α_i and β_i, and we have $f_i(\alpha_i) = \beta_i$. We let $\frac{\mathrm{d}f_i(\alpha_i)}{\alpha_i} = C$. Then we have $\alpha_i = 1 - \Phi\left(\log(-C)/(i\mu) + i\mu/2\right)$, $\beta_i = \Phi\left(\log(-C)/(i\mu) - i\mu/2\right)$. This setting of (α_i, β_i) can also be achieved by using the rejection rule $\phi(x) = I_{x \ge \log(-C)}$ for the hypothesis testing between $H_0 : x \sim \mathcal{N}\left(-\frac{i^2\mu^2}{2}, i^2\mu^2\right)$ and $H_1 : x \sim \mathcal{N}\left(\frac{i^2\mu^2}{2}, i^2\mu^2\right)$.

For $f_> = \min(\{\frac{p_i}{1-p_0}, f_i\}_{i=1}^n)$, we let $f_>(\alpha) = \beta$ where $\alpha = \sum_{i=1}^n \frac{p_i}{1-p_0}\alpha_i$, $\beta = \sum_{i=1}^n \frac{p_i}{1-p_0}\beta_i$, $\beta_i = f_i(\alpha_i)$, and $\frac{\mathrm{d}f_i(\alpha_i)}{\alpha_i} = C$. Then α and β as type I error and type II error can be achieved by using the rejection rule $\phi(x) = I_{\{x \geq \log(-C)\}}$ for the hypothesis testing between $H_0 : x \sim \sum_{i=1}^n \frac{p_i}{1-p_0}\mathcal{N}\left(-\frac{i^2\mu^2}{2}, i^2\mu^2\right)$ and $H_1 : x \sim \sum_{i=1}^n \frac{p_i}{1-p_0}\mathcal{N}\left(\frac{i^2\mu^2}{2}, i^2\mu^2\right)$.

We let $h_1(x) = \sum_{i=1}^n \frac{p_i}{1-p_0}\frac{1}{i\mu_B}\phi(\frac{x}{i\mu_B} + \frac{i\mu_B}{2})$ and $h_2(x) = \sum_{i=1}^n \frac{p_i}{1-p_0}\frac{1}{i\mu_B}\phi(\frac{x}{i\mu_B} - \frac{i\mu_B}{2})$ be the density functions of $\sum_{i=1}^n \frac{p_i}{1-p_0}\mathcal{N}\left(-\frac{i^2\mu^2}{2}, i^2\mu^2\right)$ and $\sum_{i=1}^n \frac{p_i}{1-p_0}\mathcal{N}\left(\frac{i^2\mu^2}{2}, i^2\mu^2\right)$ respectively. We have $\log\left(\frac{h_1(x)}{h_2(x)}\right) = \log(\mathrm{e}^x) = x$. Therefore, by the Neyman-Pearson Lemma, the most powerful rejection rule for the test between H_0 and H_1 is $\phi^*(x) = I_{\{x \geq \lambda\}}$ where λ is a constant. This aligns with our previous rejection rule $\phi(x) = I_{\{x \geq \log(-C)\}}$. Therefore, $f_>$ is the tradeoff function between $\sum_{i=1}^n \frac{p_i}{1-p_0}\mathcal{N}\left(-\frac{i^2\mu^2}{2}, i^2\mu^2\right)$ and $\sum_{i=1}^n \frac{p_i}{1-p_0}\mathcal{N}\left(\frac{i^2\mu^2}{2}, i^2\mu^2\right)$. ■

Proof [Proof of Theorem 16] Let n be the sample size for the bootstrap resampling. From the result of Corollary 36 and Lemma 37, the tradeoff function is $f_{Bi,\texttt{boot}} = C_{1-p_0}(f_>)$ where $p_0 = (1-1/n)^n$ and $f_>$ is the tradeoff function between two Gaussian mixtures, $\sum_{i=1}^n \frac{p_i}{1-p_0}\mathcal{N}(\mu = -\frac{i^2\mu_B^2}{2}, i^2\mu_B^2)$ vs $\sum_{i=1}^n \frac{p_i}{1-p_0}\mathcal{N}(\mu = \frac{i^2\mu_B^2}{2}, i^2\mu_B^2)$ where $p_i = \binom{n}{i}(1/n)^i(1-1/n)^{n-i}$. We let $h_1(x) = \sum_{i=1}^n \frac{p_i}{1-p_0}\frac{1}{i\mu_B}\phi(\frac{x}{i\mu_B} + \frac{i\mu_B}{2})$ and $h_2(x) = \sum_{i=1}^n \frac{p_i}{1-p_0}\frac{1}{i\mu_B}\phi(\frac{x}{i\mu_B} - \frac{i\mu_B}{2})$. From the proof of Lemma 37, we can parameterize the tradeoff function $f_>(\alpha_C) = \beta_C$ using $C \in (-\infty, \infty)$ with $\alpha_C = \int_C^\infty h_1(x)\,\mathrm{d}x$ and $\beta_C = \int_{-\infty}^C h_2(x)\,\mathrm{d}x$. We have $\frac{\mathrm{d}\alpha_C}{\mathrm{d}C} = -h_1(C)$, $\frac{\mathrm{d}\beta_C}{\mathrm{d}C} = h_2(C)$, $\frac{h_1(C)}{h_2(C)} = \mathrm{e}^{-C}$, $\frac{h_1(0)}{h_2(0)} = 1$, $\alpha_{-\infty} = \beta_\infty = 1$, $\alpha_\infty = \beta_{-\infty} = 0$, $\alpha_0 = \beta_0$, and $f'(\alpha_C) = \frac{\mathrm{d}\beta_C/\,\mathrm{d}C}{\mathrm{d}\alpha_C/\,\mathrm{d}C} = -\mathrm{e}^C$. we can transform the result in Lemma 35 to

$$
\begin{aligned}
\mathrm{kl}(f_>) &= \int_{\alpha_\infty}^{\alpha_0} (\mathrm{e}^C - 1)C\,\mathrm{d}\alpha_C = \int_0^\infty (\mathrm{e}^C - 1)Ch_1(C)\,\mathrm{d}C,\\
\kappa_2(f_>) &= \int_{\alpha_\infty}^{\alpha_0} (\mathrm{e}^C + 1)C^2\,\mathrm{d}\alpha_C = \int_0^\infty (\mathrm{e}^C + 1)C^2h_1(C)\,\mathrm{d}C,\\
\bar{\kappa}_3(f_>) &= \int_{\alpha_\infty}^{\alpha_0} (|C + \mathrm{kl}(f)|^3 + \mathrm{e}^C|C - \mathrm{kl}(f)|^3)\,\mathrm{d}\alpha_C = \int_0^\infty (|C + \mathrm{kl}(f)|^3 + \mathrm{e}^C|C - \mathrm{kl}(f)|^3)\,\mathrm{d}C,\\
\kappa_3(f_>) &= \int_{\alpha_\infty}^{\alpha_0} (\mathrm{e}^C + 1)C^3\,\mathrm{d}\alpha_C = \int_0^\infty (\mathrm{e}^C + 1)C^3h_1(C)\,\mathrm{d}C.\\
\mathrm{kl}(C_{1-p_0}(f_>)) &= (1-p_0)\int_0^\infty (\mathrm{e}^C - 1)\log\big(1 + (1-p_0)(\mathrm{e}^C - 1)\big)h_1(C)\,\mathrm{d}C,\\
\kappa_2(C_{1-p_0}(f_>)) &= \int_0^\infty \big(2 + (1-p_0)(\mathrm{e}^C - 1)\big)\big[\log\big(1 + (1-p_0)(\mathrm{e}^C - 1)\big)\big]^2h_1(C)\,\mathrm{d}C,\\
\kappa_3(C_{1-p_0}(f_>)) &= \int_0^\infty \big(2 + (1-p_0)(\mathrm{e}^C - 1)\big)\big[\log\big(1 + (1-p_0)(\mathrm{e}^C - 1)\big)\big]^3h_1(C)\,\mathrm{d}C.
\end{aligned}
$$

Now we consider $\mu_B \to 0$. Since $h_1(x) = \sum_{i=1}^n \frac{p_i}{1-p_0} \frac{1}{i\mu_B} \phi(\frac{x}{i\mu_B} + \frac{i\mu_B}{2})$,

$$\begin{aligned}
\mathrm{kl}(f_>) &= \sum_{i=1}^n \frac{p_i}{1-p_0} \int_0^\infty (\mathrm{e}^x - 1) \frac{x}{i\mu_B} \phi\left(\frac{x}{i\mu_B} + \frac{i\mu_B}{2}\right) \mathrm{d}x \\
&= \sum_{i=1}^n \frac{p_i}{1-p_0} \int_0^\infty (\mathrm{e}^x - 1) \frac{x}{i\mu_B} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(\frac{x}{i\mu_B} + \frac{i\mu_B}{2})^2}{2}} \mathrm{d}x \\
&= \sum_{i=1}^n \frac{p_i}{1-p_0} \int_0^\infty (\mathrm{e}^{i\mu_B y} - 1)(i\mu_B y) \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(y + \frac{i\mu_B}{2})^2}{2}} \mathrm{d}y \quad (\text{let } y = \frac{x}{i\mu_B}).
\end{aligned}$$

Similarly, we have

$$\mathrm{kl}(C_{1-p_0}(f_>)) = \sum_{i=1}^n p_i \int_0^\infty (\mathrm{e}^{i\mu_B y} - 1) \log\left(1 + (1-p_0)(\mathrm{e}^{i\mu_B y} - 1)\right) \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(\frac{y}{i} + \frac{i\mu_B}{2})^2}{2}} \mathrm{d}y.$$

Now we have $\lim_{\mu_B \to 0} \frac{\mathrm{kl}(C_{1-p_0}(f_>))}{\mathrm{kl}(f_>)} = (1-p_0)^2$ from the two following facts

$$\begin{aligned}
\lim_{\mu_B \to 0} \frac{\int_0^\infty \mathrm{e}^{i\mu_B y} \log\left(1 + (1-p_0)(\mathrm{e}^{i\mu_B y} - 1)\right) \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(y + \frac{i\mu_B}{2})^2}{2}} \mathrm{d}y}{\int_0^\infty \mathrm{e}^{i\mu_B y} (i\mu_B y) \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(y + \frac{i\mu_B}{2})^2}{2}} \mathrm{d}y} &= 1 - p_0, \\
\lim_{\mu_B \to 0} \frac{\int_0^\infty \log\left(1 + (1-p_0)(\mathrm{e}^{i\mu_B y} - 1)\right) \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(y + \frac{i\mu_B}{2})^2}{2}} \mathrm{d}y}{\int_0^\infty (i\mu_B y) \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(y + \frac{i\mu_B}{2})^2}{2}} \mathrm{d}y} &= 1 - p_0
\end{aligned}$$

which are obtained using L'Hospital's rule and the Leibniz integral rule (i.e., the interchange of the integral and partial differential operators).

Similarly, $\lim_{\mu_B \to 0} \frac{\kappa_2(C_{1-p_0}(f_>))}{\kappa_2(f_>)} = (1-p_0)^2$ and $\lim_{\mu_B \to 0} \frac{\kappa_3(C_{1-p_0}(f_>))}{\kappa_3(f_>)} = (1-p_0)^3$.

We re-parameterize the definition of $\mathrm{kl}(f_>)$ in Theorem 34 with C and calculate it using $h_1(x) = \sum_{i=1}^n \frac{p_i}{1-p_0} \frac{1}{i\mu_B} \phi(\frac{x}{i\mu_B} + \frac{i\mu_B}{2})$:

$$\begin{aligned}
\mathrm{kl}(f_>) &= -\int_{\alpha_\infty}^{\alpha_{-\infty}} C \, \mathrm{d}\alpha_C = \int_{-\infty}^{\infty} C h_1(C) \, \mathrm{d}C = \sum_{i=1}^n \frac{p_i}{1-p_0} \int_{-\infty}^{\infty} \frac{-x}{i\mu_B} \phi(\frac{x}{i\mu_B} + \frac{i\mu_B}{2}) \, \mathrm{d}x \\
&= \sum_{i=1}^n \frac{p_i}{1-p_0} \int_{-\infty}^{\infty} (-i\mu_B y) \phi(y + \frac{i\mu_B}{2}) \, \mathrm{d}y = \sum_{i=1}^n \frac{p_i}{1-p_0} \frac{i^2\mu_B^2}{2} = \frac{(2 - 1/n)\mu_B^2}{2(1 - (1-1/n)^n)}.
\end{aligned}$$

Similarly, for any $n = 1, 2, \ldots$, and $\mu_B \to 0$, we have

$$\begin{aligned}
\kappa_2(f_>) &= \sum_{i=1}^{n} \frac{p_i}{1-p_0} \int_{-\infty}^{\infty} \frac{x^2}{i\mu_B} \phi(\frac{x}{i\mu_B} + \frac{i\mu_B}{2}) \,\mathrm{d}x = \sum_{i=1}^{n} \frac{p_i}{1-p_0} \left(\frac{i^4\mu_B^4}{4} + i^2\mu_B^2 \right) \\
&= \frac{(2-1/n)\mu_B^2}{1-(1-1/n)^n} + \Theta(\mu_B^4), \\
\kappa_3(f_>) &= \sum_{i=1}^{n} \frac{p_i}{1-p_0} \int_{-\infty}^{\infty} \frac{|x|^3}{i\mu_B} \phi(\frac{x}{i\mu_B} + \frac{i\mu_B}{2}) \,\mathrm{d}x = \sum_{i=1}^{n} \frac{p_i}{1-p_0} (i\mu_B)^3 \int_{-\infty}^{\infty} \left| z - \frac{i\mu_B}{2} \right|^3 \phi(z) \,\mathrm{d}z \\
&\leq \sum_{i=1}^{n} \frac{p_i (i\mu_B)^3}{1-p_0} \int_{-\infty}^{\infty} \left(|z|^3 + 3\frac{i\mu_B}{2}|z|^2 + 3\left(\frac{i\mu_B}{2}\right)|z| + \left(\frac{i\mu_B}{2}\right)^3 \right) \phi(z) \,\mathrm{d}z \in \Theta(\mu_B^3).
\end{aligned}$$

By Theorem 34, we have $\lim_{B\to\infty} f_{\texttt{boot}}^{\otimes B} = G_{2K/s}$ if

$$\lim_{B\to\infty} \sum_{i=1}^{B} \mathrm{kl}(f_{Bi,\texttt{boot}}) = K, \quad \lim_{B\to\infty} \max_{1\leq i\leq B} \mathrm{kl}(f_{Bi,\texttt{boot}}) = 0,$$
$$\lim_{B\to\infty} \sum_{i=1}^{B} \kappa_2(f_{Bi,\texttt{boot}}) = s^2, \quad \lim_{B\to\infty} \sum_{i=1}^{B} \kappa_3(f_{Bi,\texttt{boot}}) = 0.$$

Since we assume that $\sqrt{B}\mu_B \to \mu'$, we have $\mu_B \in \Theta(B^{-\frac{1}{2}})$ and

$$\begin{aligned}
\lim_{B\to\infty} \sum_{i=1}^{B} \mathrm{kl}(f_{Bi,\texttt{boot}}) &= \lim_{B\to\infty} B(1-p_0)^2 \frac{(2-1/n)\mu_B^2}{2(1-(1-1/n)^n)} = (1-p_0)^2 \frac{(2-1/n)(\mu')^2}{2(1-(1-1/n)^n)}, \\
\lim_{B\to\infty} \max_{1\leq i\leq B} \mathrm{kl}(f_{Bi,\texttt{boot}}) &= \lim_{B\to\infty} (1-p_0)^2 \frac{(2-1/n)\mu_B^2}{2(1-(1-1/n)^n)} = 0, \\
\lim_{B\to\infty} \sum_{i=1}^{B} \kappa_2(f_{Bi,\texttt{boot}}) &= \lim_{B\to\infty} B(1-p_0)^2 \left(\frac{(2-1/n)\mu_B^2}{1-(1-1/n)^n} + \Theta(\mu_B^4) \right) \\
&= (1-p_0)^2 \frac{(2-1/n)(\mu')^2}{1-(1-1/n)^n}, \\
\lim_{B\to\infty} \sum_{i=1}^{B} \kappa_3(f_{Bi,\texttt{boot}}) &= \lim_{B\to\infty} B(1-p_0)^3 \Theta(\mu_B^3) = 0.
\end{aligned}$$

Therefore, $\frac{2K}{s} = \frac{2(1-p_0)^2 \frac{(2-1/n)(\mu')^2}{2(1-(1-1/n)^n)}}{\sqrt{(1-p_0)^2 \frac{(2-1/n)(\mu')^2}{1-(1-1/n)^n}}} = \sqrt{\left(2-\frac{1}{n}\right)\left(1-\left(1-\frac{1}{n}\right)^n\right)}\mu' < (\sqrt{2-2/\mathrm{e}})\mu'$.

This bound holds for $n \in \{1, 2, 3, \cdots\}$ since the function is monotonically increasing with respect to n and $\lim_{n\to\infty} \sqrt{\left(2-\frac{1}{n}\right)\left(1-\left(1-\frac{1}{n}\right)^n\right)} = \sqrt{2-2/\mathrm{e}} = 1.12438\ldots$ ■

Appendix B. Proofs for Section 4

In this section, we provide the proofs for the lemmas and theorems in Section 4.

B.1 Proofs for Section 4.2

To prove Lemma 17, we use Lemma 38 by Guillaume F.[6] and include their proof here.

Lemma 38 (Generalized Delta Method) *Let $\{X_n\}$ and $\{Y_n\}$ be sequences of random vectors taking values in $\mathbb{R}^k$ and let $f : \mathbb{R}^k \to \mathbb{R}^s$ be a transformation. Assume that*

(A1) $\{a_n\}$ is a real-valued sequence that $a_n > 0$, $a_n \to \infty$ and $a_n(\|X_n - Y_n\|_2) = O_p(1)$,

(A2) For any $\varepsilon > 0$, there exists a set S where $\limsup_{n\to\infty} P(Y_n \notin S) < \varepsilon$ and $\nabla f(y)$ is uniformly continuous for $y \in S$.

Then we have $\|a_n[f(X_n) - f(Y_n)] - a_n \nabla f(Y_n^*)^\intercal (X_n - Y_n)\|_2 \xrightarrow{p} 0$ where $Y_n^* = Y_n$ if f is differentiable at Y_n and Y_n^* is an arbitrary value in S otherwise.

Proof [Proof of Lemma 38] Define $A_n = a_n\|f(X_n) - f(Y_n) - \nabla f(Y_n^*)^\intercal (X_n - Y_n)\|_2$. Let $R(h; y) = \frac{\|f(y+h)-f(y)-\nabla f(y)^\intercal h\|_2}{\|h\|_2}$. By the mean value theorem, we have $R(h; y) = \frac{\|[\int_0^1 \nabla f(y+th)\mathrm{d}t-\nabla f(y)]^\intercal h\|_2}{\|h\|_2} \le \| \int_0^1 \nabla f(y+th)\mathrm{d}t-\nabla f(y)\|_2 \le \max_{t\in[0,1]} \|\nabla f(y+th)-\nabla f(y)\|_2$. From (A2) where ∇f is uniformly continuous, we have $\lim_{\|h\|_2\to 0}(\sup_{y\in S} R(h; y)) = 0$.

Given $\varepsilon > 0$, we can find S and N_y such that $P(Y_n \notin S) < \varepsilon/2$ for $n > N_y$ by (A2). For $Y_n \in S$, we can write $A_n = a_n\|X_n - Y_n\|_2 R(X_n - Y_n; Y_n)$. From (A1), we know $\|X_n - Y_n\|_2 \xrightarrow{p} 0$. Therefore, $a_n\|X_n - Y_n\|_2(\sup_{y\in S} R(X_n - Y_n; y)) \xrightarrow{p} 0$. Therefore, for any $\delta > 0$ and $\varepsilon > 0$, we can find N_b such that $P(a_n\|X_n-Y_n\|_2(\sup_{y\in S} R(X_n-Y_n; y)) > \delta) \le \varepsilon/2$ when $n > N_b$. Therefore, for any ε and δ, we can let $N = \max(N_y, N_b)$, and for $n > N$, we have $P(A_n > \delta) \le P(Y_n \notin S) + P(a_n\|X_n - Y_n\|_2(\sup_{y\in S} R(X_n - Y_n; y)) > \delta) \le \varepsilon$. ■

6. https://math.stackexchange.com/questions/2793833

Proof [Proof of Lemma 17] For part 1, we have

$$\mathbb{E}[\tilde{m}_{g,B}-\theta]^2 = \mathbb{E}\left[\mathbb{E}\left[\left(\frac{1}{B}\sum_{j=1}^{B}(g(D_j)+\xi_j)-g(D)+g(D)-\theta\right)^2\middle|D\right]\right] \tag{2}$$

$$= \mathbb{E}\left[\mathbb{E}\left[\left(\frac{1}{B}\sum_{j=1}^{B}(g(D_j)-g(D))+\frac{1}{B}\sum_{j=1}^{B}\xi_j\right)^2\middle|D\right]\right] + \mathbb{E}[\mathbb{E}[(g(D)-\theta)^2)|D]] \tag{3}$$

$$= \mathbb{E}\left[\mathbb{E}\left[\left(\frac{1}{B}\sum_{j=1}^{B}(g(D_j)-g(D))\right)^2+\left(\frac{1}{B}\sum_{j=1}^{B}\xi_j\right)^2\middle|D\right]\right] + \frac{1}{n}\mathbb{E}(x_i-\theta)^2 \tag{4}$$

$$= \mathbb{E}\left[\mathbb{E}\left[\frac{1}{B}\sum_{j=1}^{B}(g(D_j)-g(D))^2\middle|D\right]\right] + \frac{\sigma_e^2}{B} + \frac{1}{n}\mathbb{E}(x_i-\theta)^2 \tag{5}$$

$$= \frac{1}{B}\mathbb{E}\left[\mathbb{E}\left[(g(D_j)-g(D))^2\middle|D\right]\right] + \frac{\sigma_e^2}{B} + \frac{1}{n}\mathbb{E}(x_i-\theta)^2 \tag{6}$$

$$= \frac{1}{B}\mathbb{E}\left[\frac{1}{n}\left(\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\frac{1}{n}\sum_{i=1}^{n}x_i\right)^2\right)\right] + \frac{\sigma_e^2}{B} + \frac{1}{n}\mathbb{E}(x_i-\theta)^2 \tag{7}$$

$$= \frac{1}{B}\left(\frac{n-1}{n^2}\mathbb{E}(x_i-\theta)^2\right) + \frac{2-2/e}{\mu^2 n^2} + \frac{1}{n}\mathbb{E}(x_i-\theta)^2 \tag{8}$$

where Equation (3) is due to $\mathbb{E}\left[\frac{1}{B}\sum_{j=1}^{B}(g(D_j)-g(D)) + \frac{1}{B}\sum_{j=1}^{B}\xi_j\middle|D\right] = 0$ and $(g(D)-\theta|D) = g(D)-\theta$, Equation (4) is from $\mathbb{E}[\xi_j]=0$ and the independence between ξ_j and D_j, Equation (6) is because $D_1,\ldots,D_B$ are independent given D, and Equation (7) is by viewing D as the population and D_j as a sample from D to obtain $\mathbb{E}[(g(D_j)-g(D))^2|D] = \frac{1}{n}\mathrm{Var}(D)$.

For part 2, we use the techniques in the proofs of Beran (1997, Theorem 2.2) and Awan and Cai (2025, Theorem 2).

In the Lindeberg-Feller central limit theorem (Van der Vaart, 2000, Proposition 2.27), we let $Y_{n,i} = \frac{1}{\sqrt{n}}\sum_{k=1}^{d-1} I(x_i = s_k)e_k$ where e_k is the unit vector with k-th entry being 1, then $\sqrt{n}(\hat{\eta}-\eta) = \sum_{i=1}^{n}(Y_{n,i}-\mathbb{E}Y_{n,i}) \xrightarrow{d} \mathcal{N}(0,\Sigma)$ since $\sum_{i=1}^{n}\mathrm{Cov}(Y_{n,i}) = \Sigma$ where $\Sigma = \mathrm{diag}(\eta) - \eta\eta^\intercal$ (Severini, 2005, Example 12.7) and $\sum_{i=1}^{n}\mathbb{E}\|Y_{n,i}\|^2 I(\|Y_{n,i}\|\geq\epsilon)\to 0$ for every $\epsilon>0$ since $\|Y_{n,i}\|\leq\frac{1}{\sqrt{n}}$.

Beran (1997) showed that a sufficient condition for the bootstrap distribution converging to the sampling distribution is the local asymptotic equivariance (LAE) of $\hat{\eta}$, i.e., for the distribution of $\sqrt{n}(\hat{\eta}-\eta)$ denoted by $H_n(\eta)$, there exists $H(\eta)$ such that $H_n(\eta+\frac{h_n}{\sqrt{n}})\to H(\eta)$ for every converging sequence $h_n\to h$ where $h_n, h\in\mathbb{R}^{d-1}$. For our setup of $\hat{\eta}$, we have $\|Y_{n,i}\|\leq\frac{1}{\sqrt{n}}$ and $\sum_{i=1}^{n}\mathrm{Cov}(Y_{n,i}) = \mathrm{diag}(\eta+\frac{h_n}{\sqrt{n}}) - (\eta+\frac{h_n}{\sqrt{n}})(\eta+\frac{h_n}{\sqrt{n}})^\intercal \to \Sigma$. Therefore, the LAE condition holds with $H(\eta)=\mathcal{N}(0,\Sigma)$ by the Lindeberg-Feller central limit theorem.

Using the Skorohod Representation Theorem (Serfling, 2009, Theorem 1.6.3), there exist $U_0 \sim \text{Uniform}([0,1])$ and measurable functions $A, A_n : [0,1] \to \mathbb{R}^d$ for $n \in \mathbb{N}$ such that $A_n(U_0) \sim H_n(\eta)$ for all $n \in \mathbb{N}$, $A(U_0) \sim \mathcal{N}(0,\Sigma)$, and $A_n(U_0) \overset{a.s.}{\to} A(U_0)$. In other words, $A_n(U_0) \overset{d}{=} \sqrt{n}(\hat{\eta} - \eta)$ and $\hat{\eta} \overset{d}{=} \eta + \frac{1}{\sqrt{n}} A_n(U_0)$. Let $S = \{u_0 \in [0,1] | A_n(u_0) \to A(u_0)\}$. We know $\mathbb{P}(U_0 \in S) = 1$ and we can use $A_n(u_0)$ and $A(u_0)$ to replace h_n and h for $u_0 \in S$, respectively, in the LAE condition: For $j = 1, \ldots, B$, let $\hat{\eta}_j^*$ be the empirical distribution parameter of the j-th bootstrap data set D_j, and we have $\sqrt{n}(\hat{\eta}_j^* - \hat{\eta}) \big| (\hat{\eta} = \eta + A_n(u_0)/\sqrt{n}) \sim H_n(\eta + \frac{1}{\sqrt{n}} A_n(u_0)) \to \mathcal{N}(0,\Sigma)$. Again, by the Skorohod Representation Theorem, we define $U_j \overset{\text{iid}}{\sim} \text{Uniform}([0,1])$ independent of U_0, and for all $u_0 \in S$, there exist measurable functions $C, C_{n,A_n(u_0)} : [0,1] \to \mathbb{R}^d$ for all $n \in \mathbb{N}$ such that $C_{n,A_n(u_0)}(U_j) \sim H_n(\eta + \frac{1}{\sqrt{n}} A_n(u_0))$ for all $n \in \mathbb{N}$, $C(U_j) \sim \mathcal{N}(0,\Sigma)$, and $C_{n,A_n(u_0)}(U_j) \overset{a.s.}{\to} C(U_j)$.

From the above construction of $A_n(U_0)$ and $C_{n,A_n(u_0)}(U_j)$, $j = 1, \ldots, B$, we have the joint distribution

$$\begin{pmatrix} \sqrt{n}(\hat{\eta} - \eta) \\ \sqrt{n}(\hat{\eta}_1^* - \hat{\eta}) \\ \cdots \\ \sqrt{n}(\hat{\eta}_B^* - \hat{\eta}) \end{pmatrix} \overset{d}{=} \begin{pmatrix} A_n(U_0) \\ C_{n,A_n(U_0)}(U_1) \\ \cdots \\ C_{n,A_n(U_0)}(U_B) \end{pmatrix} \overset{a.s.}{\to} \begin{pmatrix} A(U_0) \\ C(U_1) \\ \cdots \\ C(U_B) \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} \Sigma & & & \\ & \Sigma & & \\ & & \ddots & \\ & & & \Sigma \end{pmatrix}\right) \tag{9}$$

where we have $(B+1)$ sub-matrices $\Sigma \in \mathbb{R}^{(d-1)\times(d-1)}$ on the block diagonal of the covariance matrix for the multivariate normal distribution and zeros elsewhere.

We define a function $\tilde{T} : \mathbb{R}^{(d-1)(B+1)} \to \mathbb{R}^{(B+1)}$ that

$$\tilde{\eta} = \begin{pmatrix} \hat{\eta} \\ \hat{\eta}_1^* \\ \cdots \\ \hat{\eta}_B^* \end{pmatrix}, \quad \tilde{T}(\tilde{\eta}) = \begin{pmatrix} T(\hat{\eta}) \\ T(\hat{\eta}_1^*) \\ \cdots \\ T(\hat{\eta}_B^*) \end{pmatrix}, \quad \text{and} \quad \frac{\partial \tilde{T}}{\partial \tilde{\eta}} = \begin{pmatrix} \frac{\partial T}{\partial \eta}\big|_{\eta=\hat{\eta}} & & & \\ & \frac{\partial T}{\partial \eta}\big|_{\eta=\hat{\eta}_1^*} & & \\ & & \ddots & \\ & & & \frac{\partial T}{\partial \eta}\big|_{\eta=\hat{\eta}_B^*} \end{pmatrix}.$$

where we have $(B+1)$ sub-matrices of $\frac{\partial T}{\partial \eta} \in \mathbb{R}^{(d-1)\times 1}$ on the block diagonal of $\frac{\partial \tilde{T}}{\partial \tilde{\eta}}$ and zeros elsewhere.

In Lemma 38, we let $f = \tilde{T}$, $X_n = \tilde{\eta}$, $Y_n = (\eta, \hat{\eta}, \ldots, \hat{\eta})^\intercal \in \mathbb{R}^{(B+1)}$, $a_n = \sqrt{n}$, and $Y = (\eta, \eta \ldots, \eta)^\intercal \in \mathbb{R}^{(B+1)}$. We know that (A1) is satisfied by Equation (9). As $\|Y_n - Y\|_2 \overset{p}{\to} 0$ from Equation (9), for any $\varepsilon > 0$, we can find $N_y > 0$ and $\delta > 0$ such that $P(Y_n \notin B) < \varepsilon$ for $n > N_y$ where $B = \{Y_n | \|Y_n - Y\|_2 \le \delta\}$. We also know that ∇f is uniformly continuous on B because ∇f is continuous and B is compact. Therefore, (A2) is also satisfied. Using Lemma 38, we have $\|a_n[f(X_n) - f(Y_n)] - a_n \nabla f(Y_n)^\intercal (X_n - Y_n)\|_2 \overset{p}{\to} 0$. Furthermore, since ∇f is continuous and $\|Y_n - Y\|_2 \overset{p}{\to} 0$, we have $\|\nabla f(Y_n) - \nabla f(Y)\|_2 \overset{p}{\to} 0$, and $\|a_n[f(X_n) - f(Y_n)] - a_n \nabla f(Y)^\intercal (X_n - Y_n)\|_2 \overset{p}{\to} 0$. In other words, we have

$$\begin{pmatrix} \sqrt{n}(T(\hat{\eta}) - T(\eta)) \\ \sqrt{n}(T(\hat{\eta}_1^*) - T(\hat{\eta})) \\ \cdots \\ \sqrt{n}(T(\hat{\eta}_B^*) - T(\hat{\eta})) \end{pmatrix} \overset{d}{\to} \mathcal{N}\left(0,\ (n\sigma_g^2) I_{(B+1)}\right).$$

Since the ξ_j are independent of $T(\hat{\eta}_j^*)$ and $T(\hat{\eta})$, we have

$$\begin{pmatrix} \sqrt{n}(T(\hat{\eta}) - T(\eta)) \\ \sqrt{n}(T(\hat{\eta}_1^*) - T(\hat{\eta}) + \xi_1) \\ \cdots \\ \sqrt{n}(T(\hat{\eta}_B^*) - T(\hat{\eta}) + \xi_B) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{pmatrix} n\sigma_g^2 & & & \\ & (n\sigma_g^2 + n\sigma_e^2) & & \\ & & \ddots & \\ & & & (n\sigma_g^2 + n\sigma_e^2) \end{pmatrix}\right)$$

where the off-diagonal values in the covariance matrix are zeros.

Using $\tilde{m}_{g,B} = \frac{1}{B}\sum_{j=1}^{B} \tilde{g}(D_j) = \frac{1}{B}\sum_{j=1}^{B}(T(\hat{\eta}_j^*) + \xi_j)$ and $\theta = T(\eta)$, we know

$$\sqrt{n}(\tilde{m}_{g,B} - \theta) = \sqrt{n}(T(\hat{\eta}) - T(\eta)) + \frac{1}{B}\sum_{j=1}^{B} \sqrt{n}(T(\hat{\eta}_j^*) - T(\hat{\eta}) + \xi_j).$$

Then, using the continuous mapping theorem and the joint distribution above, we know that $\sqrt{n}(\tilde{m}_{g,B} - \theta)$ converges to a normal random variable with variance $(n\sigma_g^2 + \frac{n\sigma_g^2 + n\sigma_e^2}{B})$. Therefore, we have $\frac{\sqrt{n}(\tilde{m}_{g,B} - \theta)}{\sqrt{n\sigma_g^2 + \frac{n}{B}(\sigma_g^2 + \sigma_e^2)}} \xrightarrow{d} \mathcal{N}(0,1)$.

Similarly, from $\tilde{s}_{g,B}^2 = \frac{1}{B-1}\sum_{j=1}^{B}(\tilde{g}(D_j) - \tilde{m}_{g,B})^2$, we know

$$n\tilde{s}_{g,B}^2 = \frac{1}{B-1}\sum_{j=1}^{B}\left[\sqrt{n}(T(\hat{\eta}_j^*) - T(\hat{\eta}) + \xi_j) - \frac{1}{B}\sum_{j=1}^{B}\sqrt{n}(T(\hat{\eta}_j^*) - T(\hat{\eta}) + \xi_j)\right]^2.$$

Using the continuous mapping theorem and the joint distribution above, $n\tilde{s}_{g,B}^2$ is the sample variance of B random variables converging to B i.i.d. normal random variables with mean 0 and variance $(n\sigma_g^2 + n\sigma_e^2)$. Therefore, we have $(B-1)\frac{n\tilde{s}_{g,B}^2}{n\sigma_g^2 + n\sigma_e^2} \xrightarrow{d} \chi_{B-1}^2$. ■

Proof [Proof of Theorem 20] Let $r = \Phi^{-1}(1 - \frac{\omega}{2})\sqrt{\sigma_g^2 + \frac{\sigma_g^2 + \sigma_e^2}{B}}$.

We know $\mathbb{P}(\hat{r}_n \geq r) = \mathbb{P}\left(\hat{\sigma}_g^2 \geq \sigma_g^2\right) = \mathbb{P}\left(Y_n \geq \frac{c}{B-1}(\sigma_g^2 + \sigma_e^2)\right) \to (1 - \alpha + \omega)$ and $\mathbb{P}(\theta \in [X_n - r,\ X_n + r]) \to (1 - \omega)$. Therefore,

$$\begin{aligned} \lim_{n\to\infty} \mathbb{P}(\theta \in [X_n - \hat{r}_n,\ X_n + \hat{r}_n]) &\geq \left(\lim_{n\to\infty} \mathbb{P}(\theta \in [X_n - r,\ X_n + r])\right)\left(\lim_{n\to\infty} \mathbb{P}(\hat{r}_n \geq r)\right) \\ &= (1-\omega)(1-\alpha+\omega) = 1 - \alpha + \omega(\alpha - \omega) > 1 - \alpha. \end{aligned}$$

■

Proof [Proof of Proposition 22] First, we note that the Cramér-Rao lower bound (Shao, 2003, Theorem 3.3) indicates that $\mathrm{Var}(T(\hat{\eta})) \geq \sigma_g^2 = \frac{1}{n}\left(\frac{\partial T}{\partial \eta}\right)^{\intercal} \Sigma \frac{\partial T}{\partial \eta}$ where $\Sigma = \mathrm{diag}(\eta) - \eta\eta^{\intercal}$ is equal to the inverse of the Fisher information matrix for the multinomial distribution.

We will prove $\hat{r}_n \xrightarrow{a.s.} \Phi^{-1}(1 - \frac{\alpha}{2})\sigma_g$. As $B = o(n^2)$, we have $\sigma_e^2 \to 0$ and $\sigma_g^2 \to 0$ when $n \to \infty$. By the Berry-Esseen theorem and the fact that the Chi-square distribution is the

distribution of a sum of the squares of independent standard normal random variables, for $X \sim \chi^2_{B-1}$, we have $\left|\mathbb{P}\left(\frac{X-(B-1)}{\sqrt{2(B-1)}} \leq x\right) - \Phi(x)\right| = O(\frac{1}{\sqrt{B}})$ as $B \to \infty$.

Then we prove that $\frac{(B-1)-c_B}{B}$ converges to 0 where c_B is the $(\alpha-\omega)$ quantile of the χ^2_{B-1} distribution. As $(\alpha-\omega) \to 0$, if there is $\Delta > 0$ such that $\Delta < \frac{(B-1)-c_B}{B}$ for a subsequence of $\{c_B\}_{B=1}^{\infty}$, then $\alpha - \omega = \mathbb{P}(X \leq c_B) \leq \mathbb{P}(X < (B-1) - B\Delta) = \mathbb{P}(\frac{X-(B-1)}{\sqrt{2(B-1)}} < -\frac{B\Delta}{\sqrt{2(B-1)}}) \leq \Phi(-\frac{B\Delta}{\sqrt{2(B-1)}}) + O(\frac{1}{\sqrt{B}})$. From the Chernoff's bound, we know $\mathbb{P}(Z > c) \leq \mathrm{e}^{-c^2/2}$ for $c > 0$ and $Z \sim N(0,1)$, and we have $\Phi(-\frac{B\Delta}{\sqrt{2(B-1)}}) = o(\frac{1}{\sqrt{B}})$. This means $\alpha - \omega = O(B^{-\frac{1}{2}})$ which contradicts our assumption $B^{-\frac{1}{2}} = o(\alpha-\omega)$. Therefore, $\frac{(B-1)-c_B}{B}$ must converge to 0 which means $\frac{c_B}{B} \to 1$.

Let $\nu \sim \chi^2_{B-1}$. As χ^2_{B-1} is the sum of $(B-1)$ i.i.d. χ^2_1 random variables, by the strong law of large numbers, we have $\frac{1}{B-1}\nu \xrightarrow{a.s.} 1$. As we have $(B-1)\frac{Y_n}{\sigma_g^2+\sigma_e^2} \xrightarrow{d} \chi^2_{B-1}$, we also have $\left(\frac{Y_n-\sigma_e^2}{\sigma_g^2} - 1\right)\frac{\sigma_g^2}{\sigma_g^2+\sigma_e^2} = \frac{Y_n}{\sigma_g^2+\sigma_e^2} - 1 \xrightarrow{d} \left(\frac{1}{B-1}\nu - 1\right) \xrightarrow{a.s.} 0$ as $B \to \infty$ and $n \to \infty$, which means $\frac{Y_n-\sigma_e^2}{\sigma_g^2} \xrightarrow{p} 1$. Using $\frac{c_B}{B} \to 1$, we have $\frac{\hat{\sigma}_g^2}{\sigma_g^2} \xrightarrow{p} 1$.

Then, using the definition of σ_g^2 and σ_e^2, we have $\frac{\sigma_e^2}{B\sigma_g^2} = \frac{1}{n}\frac{(2-2/e)}{\mu^2\left(\frac{\partial T}{\partial \eta}\right)^\intercal \Sigma \frac{\partial T}{\partial \eta}} \to 0$. Therefore, $\frac{\sqrt{\hat{\sigma}_g^2 + \frac{1}{B}(\hat{\sigma}_g^2+\sigma_e^2)}}{\sigma_g} = \sqrt{(1+\frac{1}{B})\frac{\hat{\sigma}_g^2}{\sigma_g^2} + \frac{\sigma_e^2}{B\sigma_g^2}} \xrightarrow{p} 1$. As we know $\omega \to \alpha$, we have $\hat{r}_n \xrightarrow{a.s.} \Phi^{-1}(1-\frac{\alpha}{2})\sigma_g$ when $n \to \infty$ and $B \to \infty$. ■

Appendix C. Privacy Analysis of DP Bootstrap with Gaussian Mechanism

In this section, we first derive the curve for our lower bound on DP bootstrap with Gaussian mechanism in Figure 3a; then we show why the privacy analysis by Brawner and Honaker (2018) is incorrect; Finally, we show why the PLDs by Koskela et al. (2020) is incorrect.

C.1 Our Lower Bound

In this section, we explain how the curves in Figure 3a were derived.

We evaluate our lower bound $C_{1-p_0}(f_>)$ based on Corollary 36 and Lemma 37 and visualize it in Figure 3a. From the proof of Lemma 37, the tradeoff function $\beta = f_>(\alpha)$ is parameterized by $C \in (-\infty, 0)$ where $\alpha = \sum_{i=1}^n \frac{p_i}{1-p_0}\alpha_i$, $\alpha_i = 1 - \Phi\left(\log(-C)/(i\mu) + i\mu/2\right)$, $\beta = \sum_{i=1}^n \frac{p_i}{1-p_0}\beta_i$, $\beta_i = \Phi\left(\log(-C)/(i\mu) - i\mu/2\right)$; Then we use Lemma 31 to obtain $C_{1-p_0}(f_>)$.

In Figure 3a, we also visualize the tradeoff functions for specific neighboring data sets. We let $D = (x_1, x_2, \ldots, x_n)$, and $\mathcal{M}_G(D) = \frac{1}{n}\sum_{i=1}^n x_i + \xi$ where $\xi \sim \mathcal{N}(0, 1/(n\mu)^2)$. We study the tradeoff function between $\mathcal{M}_G \circ \texttt{boot}(D_1)$ and $\mathcal{M}_G \circ \texttt{boot}(D_2)$ where $D_1 =$

$(a, 0, 0, \ldots, 0)$, $D_2 = (a-1, 0, 0, \ldots, 0)$, $|D_1| = |D_2| = n$. The tradeoff function between $\mathcal{M}_G(D_1)$ and $\mathcal{M}_G(D_2)$ is G_μ. By the number of occurrences of a and $1-a$, we have

$$\mathcal{M}_G \circ \texttt{boot}(D_1) \sim \sum_{i=0}^{n} p_i \mathcal{N}\left(\frac{ia}{n}, \frac{1}{(n\mu)^2}\right), \quad \mathcal{M}_G \circ \texttt{boot}(D_2) \sim \sum_{i=0}^{n} p_i \mathcal{N}\left(\frac{i(a-1)}{n}, \frac{1}{(n\mu)^2}\right),$$

where we are referring to the distribution of the output of $\mathcal{M}_G$ applied to one bootstrap sample which includes the randomness of $\mathcal{M}_G$ as well as the randomness of `boot`. Since the tradeoff function is a lower bound for the curve between type I error and type II error from any rejection rule, we consider the hypothesis tests $H_0 : D = D_1,\ H_1 : D = D_2$ and the rejection rule $\{\mathcal{M}_G \circ \texttt{boot}(D) < C\}$ (which need not be the optimal rejection rule). The type I and type II errors are $\alpha = \sum_{i=0}^{n} \Phi(Cn\mu - ia\mu)$ and $\beta = \sum_{i=0}^{n} \Phi(i(a-1)\mu - Cn\mu)$. In Figure 3a, we show the curves of (α, β) for $\mu = 1,\ n = 1000,\ a \in 0, 0.2, 0.4, 0.6, 0.8, 1$. We can see that the curve for $a = 0$ is not lower bounded by 1-GDP which shows that the bootstrap cannot be used for free with the same f-DP guarantee.

C.2 Counterexample of the Privacy Analysis by Brawner and Honaker (2018)

In this section, we show an example disproving the statement 'bootstrap for free' by Brawner and Honaker (2018). Their result is under zCDP, a variant of DP.

Definition 39 (Zero-Concentrated DP (zCDP): Bun and Steinke, 2016) *$\mathcal{M}$ is ρ-zCDP if for all neighboring data sets D_1 and D_2 and all $\alpha \in (1, \infty)$, $D_\alpha(\mathcal{M}(D_1)||\mathcal{M}(D_2)) \leq \rho\alpha$ where $D_\alpha(P||Q)$ is the α-Rényi divergence, $D_\alpha(P||Q) = \frac{1}{\alpha-1} \log\left[\mathbb{E}_{x\sim Q}\left(\frac{\mathrm{d}P}{\mathrm{d}Q}(x)\right)^\alpha\right]$, and $\frac{\mathrm{d}P}{\mathrm{d}Q}$ is the Radon-Nikodym derivative of P with respect to Q.*

We consider $\mathcal{X} = [0, 1]$. For any data set D containing two individuals from $\mathcal{X}$, i.e., $D \in \mathcal{X}^2$, $D = (x_1, x_2)$, we define the Gaussian mechanism on the sample sum: $\mathcal{M}(D) := x_1 + x_2 + \xi$ where $\xi \sim \mathcal{N}(0, 1)$. From the results in (Bun and Steinke, 2016), we know that $\mathcal{M}$ satisfies $\frac{1}{2}$-zCDP, i.e., $\rho = \frac{1}{2}$. Now we show that $\mathcal{M} \circ \texttt{boot}$ does not satisfy $\frac{1}{2}$-zCDP.

To find a counterexample, we consider the neighboring data sets $D_1 = (1, 0)$, $D_2 = (0, 0)$. Under this case, we have $\mathcal{M} \circ \texttt{boot}(D_1) \sim \frac{1}{4}\mathcal{N}(0, 1) + \frac{1}{2}\mathcal{N}(1, 1) + \frac{1}{4}\mathcal{N}(2, 1)$ and $\mathcal{M} \circ \texttt{boot}(D_2) \sim \mathcal{N}(0, 1)$. We check the α-Rényi divergence when $\alpha = 2$:

$$D_2\left(\frac{1}{4}\mathcal{N}(0,1) + \frac{1}{2}\mathcal{N}(1,1) + \frac{1}{4}\mathcal{N}(2,1)\middle\|\mathcal{N}(0,1)\right) = \log\left(\frac{1 + 4\mathrm{e} + \mathrm{e}^4 + 4 + 2 + 4\mathrm{e}^2}{16}\right)$$
$$\approx 1.85265\ldots > \rho\alpha = 1.$$

Therefore, there exists a mechanism $\mathcal{M}$ satisfying $\frac{1}{2}$-zCDP such that $\mathcal{M} \circ \texttt{boot}$ does not satisfy $\frac{1}{2}$-zCDP which disproves the result in (Brawner and Honaker, 2018).

C.3 Counterexample of the Privacy Analysis by Koskela et al. (2020)

In this section, we show that the privacy loss distribution (PLD) of the DP bootstrap with Gaussian mechanism cannot be the one shown in Koskela et al. (2020).

The privacy loss function and privacy loss distribution are defined as below.

Definition 40 (Definition 3 and 4 (Koskela et al., 2020)) *Let $\mathcal{M} : \mathcal{X}^N \to \mathbb{R}$ be a randomised mechanism and let $X \simeq Y$. Let $f_X(t)$ denote the density function of $\mathcal{M}(X)$ and $f_Y(t)$ the density function of $\mathcal{M}(Y)$. Assume $f_X(t) > 0$ and $f_Y(t) > 0$ for all $t \in \mathbb{R}$. We define the privacy loss function of f_X over f_Y as $\mathcal{L}_{X/Y}(t) = \log \frac{f_X(t)}{f_Y(t)}$.*

Using the notation of Def. 40, suppose that $\mathcal{L}_{X/Y} : \mathbb{R} \to D$, $D \subset \mathbb{R}$ is a continuously differentiable bijective function. The privacy loss distribution (PLD) of $\mathcal{M}(X)$ over $\mathcal{M}(Y)$ is defined to be a random variable which has the density function

$$\omega_{X/Y}(s) = \begin{cases} f_X\big(\mathcal{L}_{X/Y}^{-1}(s)\big) \, \frac{\mathrm{d}\mathcal{L}_{X/Y}^{-1}(s)}{\mathrm{d}s}, & s \in \mathcal{L}_{X/Y}(\mathbb{R}), \\ 0, & \text{else.} \end{cases}$$

Although Koskela et al. (2020) did not mention in this definition of the privacy loss distribution, their density function requires $\mathcal{L}_{X/Y}(s)$ to be a monotonically increasing function with respect to s. When $\mathcal{L}_{X/Y}(s)$ monotonically decreases, we need to replace $\frac{\mathrm{d}\mathcal{L}_{X/Y}^{-1}(s)}{\mathrm{d}s}$ by $\left|\frac{\mathrm{d}\mathcal{L}_{X/Y}^{-1}(s)}{\mathrm{d}s}\right|$. Then they state the following result on privacy profile.

Lemma 41 (Lemma 5 (Koskela et al., 2020)) *Assume $(\varepsilon, \infty) \subset \mathcal{L}_{X/Y}(\mathbb{R})$. Then, $\mathcal{M}$ is tightly (ε, δ)-DP for $\delta(\varepsilon) = \max_{X \simeq Y}\{\delta_{X/Y}(\varepsilon)\}$, where $\delta_{X/Y}(\varepsilon) = \int_\varepsilon^\infty (1 - e^{\varepsilon - s}) \, \omega_{X/Y}(s) \, \mathrm{d}s$.*

Their result on subsampling with replacement is as below.

Proposition 42 (Section 6.3 in (Koskela et al., 2020)) *Consider the sampling with replacement and the $\simeq$-neighbouring relation. The number of contributions of each member of the data set is not limited. The size of the new sample is m. Then ℓ, the number of times the differing sample x' is in the batch, is binomially distributed, i.e., $\ell \sim \mathrm{Binomial}(1/n, m)$, where n is the size of the original sample. Then the subsampled Gaussian mechanism, $\mathcal{M}(D) = \sum_{x \in B} f(x) + \mathcal{N}(0, \sigma^2 I_d)$ where B is a uniformly randomly drawn subset (with replacement) of $D = \{x_1, \ldots, x_n\}$ and $\|f(x)\|_2 \leq 1$ for all $x \in X$, satisfies $(\varepsilon, \delta(\varepsilon))$-DP where $\delta(\varepsilon) = \delta_{Y/X}(\varepsilon) = \delta_{X/Y}(\varepsilon)$ which is derived from*

$$f_X(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \sum_{\ell=0}^{m} \left(\frac{1}{n}\right)^\ell \left(1 - \frac{1}{n}\right)^{m-\ell} \binom{m}{\ell} e^{\frac{-(t-\ell)^2}{2\sigma^2}},$$

$$f_Y(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \sum_{\ell=0}^{m} \left(\frac{1}{n}\right)^\ell \left(1 - \frac{1}{n}\right)^{m-\ell} \binom{m}{\ell} e^{\frac{-(t+\ell)^2}{2\sigma^2}}.$$

After restating the results by Koskela et al. (2020), we prove that the privacy profile $\delta(\varepsilon)$ from the above theorem is not a valid privacy guarantee for some neighboring data sets.

Consider the Gaussian mechanism $\mathcal{M}$ on $f(x) = x$ as $\mathcal{M}(D) = \sum_{x \in D} f(x) + \xi$ where $\xi \sim \mathcal{N}(0, 1)$, $D = (x_1, x_2)$, $x_1, x_2 \in \mathcal{X} := [-1, 1]$. We consider two neighboring data sets,

$D_1 = (1, 1)$ and $D_2 = (-1, 1)$, and bootstrap, i.e., sampling with replacement when sample size remaining the same. We have the two output distributions as $\mathcal{M}(\texttt{boot}(D_1)) \sim \mathcal{N}(2, 1)$ and $\mathcal{M}(\texttt{boot}(D_2)) \sim \frac{1}{4}\mathcal{N}(-2, 1) + \frac{1}{2}\mathcal{N}(0, 1) + \frac{1}{4}\mathcal{N}(2, 1)$ with their density functions being $f_{D_1,1}(t) = \frac{1}{\sqrt{2\pi}}\left(\mathrm{e}^{\frac{-(t-2)^2}{2}}\right), f_{D_2,1}(t) = \frac{1}{\sqrt{2\pi}}\left(\frac{1}{4}\mathrm{e}^{\frac{-(t-2)^2}{2}} + \frac{1}{2}\mathrm{e}^{\frac{-t^2}{2}} + \frac{1}{4}\mathrm{e}^{\frac{-(t+2)^2}{2}}\right)$.

However, Proposition 42 gives the privacy loss distribution from the following (which can also be derived from $D_1' = (1, 0)$ and $D_2' = (-1, 0)$):

$$f_{D_1,2}(t) = \frac{1}{\sqrt{2\pi}}\left(\frac{1}{4}\mathrm{e}^{\frac{-t^2}{2}} + \frac{1}{2}\mathrm{e}^{\frac{-(t-1)^2}{2}} + \frac{1}{4}\mathrm{e}^{\frac{-(t-2)^2}{2}}\right),$$
$$f_{D_2,2}(t) = \frac{1}{\sqrt{2\pi}}\left(\frac{1}{4}\mathrm{e}^{\frac{-t^2}{2}} + \frac{1}{2}\mathrm{e}^{\frac{-(t+1)^2}{2}} + \frac{1}{4}\mathrm{e}^{\frac{-(t+2)^2}{2}}\right).$$

Now, we can calculate the δ when $\varepsilon = 1$.

$$\begin{aligned}\delta_{X/Y}(\varepsilon) &= \int_{\varepsilon}^{\infty} (1 - \mathrm{e}^{\varepsilon - s})\, \omega_{X/Y}(s)\, \mathrm{d}s = \int_{\varepsilon}^{\infty} (1 - \mathrm{e}^{\varepsilon - s})\, f_X\big(\mathcal{L}_{X/Y}^{-1}(s)\big)\, \frac{\mathrm{d}\mathcal{L}_{X/Y}^{-1}(s)}{\mathrm{d}s}\, \mathrm{d}s \\ &= \int_{\varepsilon}^{\infty} (1 - \mathrm{e}^{\varepsilon - s})\, f_X\big(\mathcal{L}_{X/Y}^{-1}(s)\big)\, \mathrm{d}\mathcal{L}_{X/Y}^{-1}(s) = \int_{\mathcal{L}_{X/Y}^{-1}(\varepsilon)}^{\mathcal{L}_{X/Y}^{-1}(\infty)} (f_X(z) - \mathrm{e}^{\varepsilon} f_Y(z))\, \mathrm{d}z.\end{aligned}$$

For $f_{D_1,1}$ vs $f_{D_2,1}$, we know that $\frac{f_{D_2,1}}{f_{D_1,1}}$ is monotonically decreasing since the ratio between each of the three parts in $f_{D_2,1}$ and $f_{D_1,1}$ is decreasing. Therefore, $\log(\frac{f_{D_2,1}}{f_{D_1,1}})$ is decreasing, and we let it be $\mathcal{L}_{X/Y}$ with $f_X = f_{D_2,1}$ and $f_Y = f_{D_1,1}$. To calculate the δ, we find t_1 such that $t_1 = \mathcal{L}_{X/Y}^{-1}(1)$ and $t_2 = \mathcal{L}_{X/Y}^{-1}(\infty)$, i.e., $\frac{f_{D_2,1}(t_1)}{f_{D_1,1}(t_1)} = \mathrm{e}$ and $\frac{f_{D_2,1}(t_2)}{f_{D_1,1}(t_2)} = \infty$. Using `R`, we have $t_1 = 0.2228743\ldots$, $t_2 = -\infty$. Then $\delta_1 = F_{D_2,1}(t_1) - \mathrm{e} * F_{D_1,1}(t_1)$ where F is the CDF corresponding to f. Using `R`, we have $\delta_1 = 0.4475773$.

Similarly, we know $\frac{f_{D_2,2}}{f_{D_1,2}}$ is monotonically decreasing from the explanation in the Section 6.3 in (Koskela et al., 2020). We find t_3 such that $\frac{f_{D_2,2}(t_3)}{f_{D_1,2}(t_3)} = \mathrm{e}$, then $\delta_2 = F_{D_2,2}(t_3) - \mathrm{e} * F_{D_1,2}(t_3)$. Using `R`, we have $t_3 = -0.7830073$, and $\delta_2 = 0.369344$.

Since we have $\delta_1 > \delta_2$, the claim by Koskela et al. (2020) that $f_{D_1,2}$ and $f_{D_2,2}$ gives an upper bound of $\delta(\varepsilon)$ is not true.

Appendix D. Details of the Simulation and Real-World Experiment

In this section, we include NoisyVar (Algorithm 4) used in our simulation, and we derive the sensitivity of the regularized ERM of quantile regression for using DP bootstrap with output perturbation in quantile regression.

Algorithm 4 `NoisyVar` (DP CI for population mean (Du et al., 2020))

Input data set $D = \{x_1, x_2, \ldots, x_n\} \in \mathcal{X}^n$, $x_i \in [0,1]$ for $i = 1, \ldots, n$, $\mu > 0$ for μ-GDP, number of bootstrap samples B, confidence level $\alpha \in [0,1]$, simulation number $nsim$.
$\mathcal{M}_1(D) \leftarrow \frac{1}{n}\sum_{i=1}^n x_i + \xi_1, \quad \xi_1 \sim \mathcal{N}(0, \frac{2B}{(n\mu)^2})$.
$\mathcal{M}_2(D) \leftarrow \max\left\{0, \frac{1}{n-1}\sum_{i=1}^n \left(x_i - \frac{1}{n}\sum_{j=1}^n x_j\right)^2 + \xi_2\right\}, \quad \xi_2 \sim \mathcal{N}(0, \frac{2B}{(n\mu)^2})$.
for $i = 1, 2, \ldots, nsim$ **do**
  $D' \leftarrow \{x_1', x_2', \ldots, x_n'\}$ where $\tilde{x}_i' \sim \mathcal{N}(\mathcal{M}_1(D), \mathcal{M}_2(D))$ and $x_i' = \min(1, \max(0, \tilde{x}_i'))$.
  $\hat{\theta}_i \leftarrow \frac{1}{n}\sum_{i=1}^n x_i' + \xi_1, \quad \xi_1 \sim \mathcal{N}(0, \frac{2B}{(n\mu)^2})$
end for
$MoE \leftarrow \frac{q(\hat{\theta}, 1-\frac{\alpha}{2}) - q(\hat{\theta}, \frac{\alpha}{2})}{2}$ where $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_{nsim})$ and $q(x, \alpha)$ is a non-private empirical quantile function that outputs the α quantile of x.
Return $(\mathcal{M}_1(D) - MoE, \mathcal{M}_1(D) + MoE)$ which satisfies μ-GDP.

Let $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, $l(z_i) = (\tau - \mathbb{1}(z_i \le 0))z_i$ and $z_i = y_i - x_i^\intercal\theta$. The regularized empirical risk function $R_D^c(\theta) = \frac{1}{n}\sum_{i=1}^n l(z_i) + c\|\theta\|_2^2$ is $2c$-strongly convex. We consider a neighboring data set to D: $D = \{(x_1', y_1'), (x_2, y_2), \ldots, (x_n, y_n)\}$, where $x_i = (1, w_i)$, $x_1' = (1, w_1')$, and $w_i \in [0,1]$, $w_1' \in [0,1]$. Let $\hat{\theta}_D^c = \text{argmin}_\theta R_D^c(\theta)$. Denote $a = \mathbb{1}(y_1 \le \theta^\intercal x_1), b = \mathbb{1}(y_1' \le \theta^\intercal x_1')$. By (Chaudhuri et al., 2011, Lemma 7), $\|\hat{\theta}_D^c - \hat{\theta}_{D'}^c\| \le \frac{1}{2c}\max_\theta \|\nabla(l(z_1) - l(z_1'))/n\| = \frac{1}{2nc}\|-\tau(x_1 - x_1') + ax_1 - bx_1'\| = \frac{1}{2nc}\|\binom{a-b}{-\tau(w_1 - w_1') + aw_1 - bw_1'}\|$.
If $a = b = 1$, $\|\binom{a-b}{-\tau(w_1-w_1')+aw_1-bw_1'}\| = \|\binom{0}{(1-\tau)(w_1-w_1')}\| \le 2(1-\tau)$;
If $a = b = 0$, $\|\binom{a-b}{-\tau(w_1-w_1')+aw_1-bw_1'}\| = \|\binom{0}{-\tau(w_1-w_1')}\| \le 2\tau$;
If $a = 0, b = 1$, $\|\binom{a-b}{-\tau(w_1-w_1')+aw_1-bw_1'}\| = \|\binom{-1}{-\tau w_1 - (1-\tau)w_1'}\| \le \sqrt{2}$;
If $a = 1, b = 0$, $\|\binom{a-b}{-\tau(w_1-w_1')+aw_1-bw_1'}\| = \|\binom{1}{(1-\tau)w_1 + \tau w_1'}\| \le \sqrt{2}$.
Therefore, $\|\hat{\theta}_D^c - \hat{\theta}_{D'}^c\| \le \frac{1}{2nc}\max\{2\tau, 2(1-\tau), \sqrt{2}\}$ which is an upper bound of the sensitivity of the regularized ERM.

Appendix E. More Comparison Results Between Deconvolved DP Bootstrap and Other Methods

Figure 9 shows more simulation results of the population mean inference following the setting in Section 5.1 and Figure 7. In the legend, "Bootstrap (B=1000)" is the non-private baseline, and "B=..." is DP bootstrap with B bootstrap estimates. The first row of subfigures shows that the CI width decreases when the sample size n increases. The second row shows that DP bootstrap has over-coverage when n is small, while NoisyVar often has under-coverage. In the third row, we define Extra width := $\frac{\text{private CI width}}{\text{non}-\text{private CI width}} - 1$ and use a log scale for the y-axis in the subfigures (replacing Extra width with $\max\{10^{-4}, \text{Extra width}\}$) to better show the relationship between the extra width and the sample size. In the fourth row, we define $\sqrt{\text{SNR}} := \frac{\text{non}-\text{private CI width}/(2\Phi(0.95))}{\sigma_{\text{DP noise}}}$ as the confidence level is 90% and we assume the sampling distribution is approximately normal. We can see that if $\sqrt{\text{SNR}} \ge 1$,

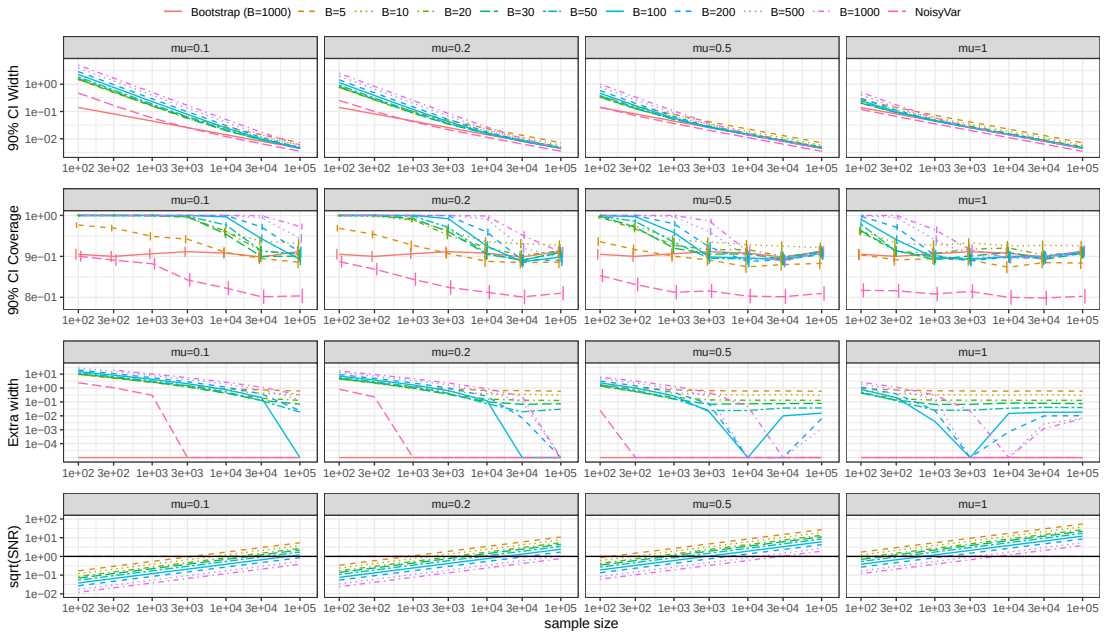

Figure 9: Results for various choices of B and μ-GDP for the population mean inference.

DP bootstrap does not have over-coverage, and the largest B satisfying $\sqrt{\text{SNR}} \geq 1$ gives the smallest extra width.

Figure 10 and 11 show more results of the slope parameter inference in logistic regression following the setting in Section 6.2 and Figure 8. The regularization parameter is $c = 1$ and $c = 0.01$ in Figure 10 and 11 respectively. In the figures, the third row is $P(\text{CI covers } 0)$, indicating the power of the private CI for rejecting the independence between MRKINC and SHELCO, and the other rows are the same as in Figure 9. We observe the same conclusion as from Figure 9. Additionally, with n being large enough, DP bootstrap has better performance (less extra width) than DP-CI-ERM if B is chosen to satisfy $\sqrt{\text{SNR}} \approx 1$. By comparing the second and third rows in Figure 10 and 11, we find that DP bootstrap has less over-coverage but also less power when the regularization parameter c is larger. Similarly, Figure 12 and 13 show more results of quantile regression. We observe that DP bootstrap has better SNR and more power for rejecting the independence between MRKINC and SHELCO in quantile regression than in logistic regression under the same choice of c. This resonates with the comparison between private linear regression and quantile regression in (Reimherr and Awan, 2019, Figure 1 and 2), which shows that robust methods like quantile regression are less affected by the DP constraints to some extent.

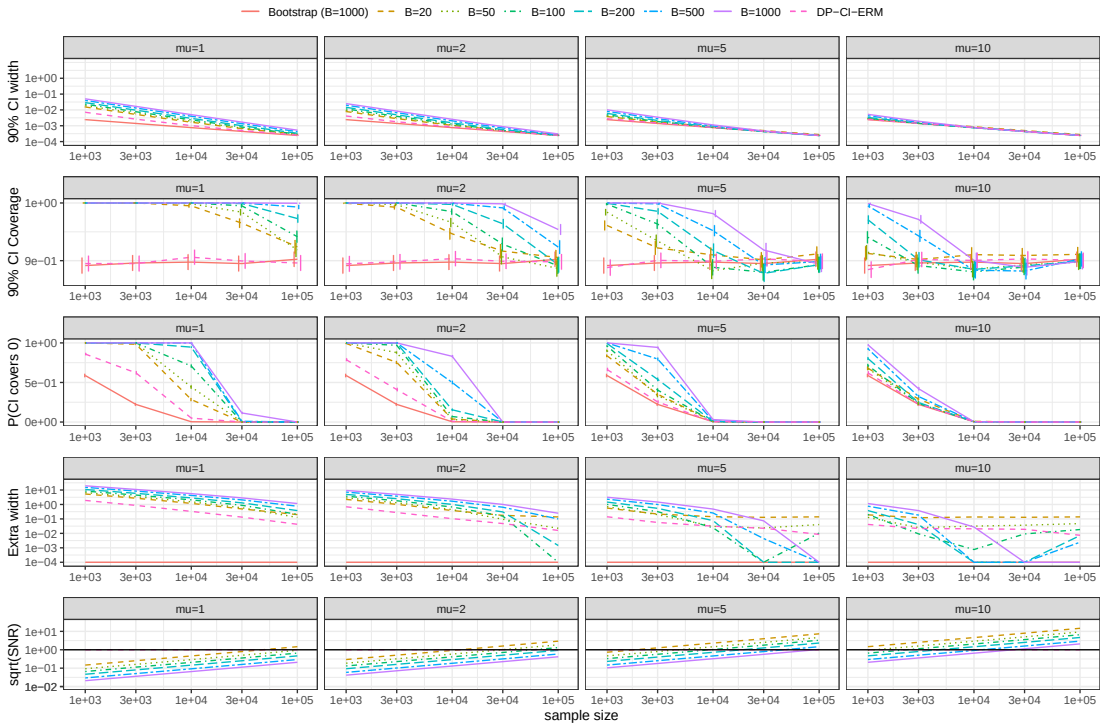

Figure 10: Results for various choices of B and μ-GDP for logistic regression with $c = 1$.

Appendix F. More Comparison Results Between Numerical and Asymptotic DP Guarantees of DP Bootstrap

As an supplement to Figure 4 in our main paper, Figure 14 shows the comparison between the numerical and asymptotic results for more choices of n.

References

Stefan Van Aelst and Gert Willems. Multivariate regression s-estimators for robust estimation and inference. Statistica Sinica, 15:981–1001, 2005.

Jordan Awan and Zhanrui Cai. One step to efficient synthetic data. Statistica Sinica, 35: 539–561, 2025.

Jordan Awan and Aleksandra Slavković. Differentially private inference for binomial data. Journal of Privacy and Confidentiality, 10(1):1–40, 2020.

Jordan Awan and Zhanyu Wang. Simulation-based, finite-sample inference for privatized data. Journal of the American Statistical Association, pages 1–14, 2024.

Jordan Awan, Ana Kenney, Matthew Reimherr, and Aleksandra Slavković. Benefits and pitfalls of the exponential mechanism with applications to Hilbert spaces and functional

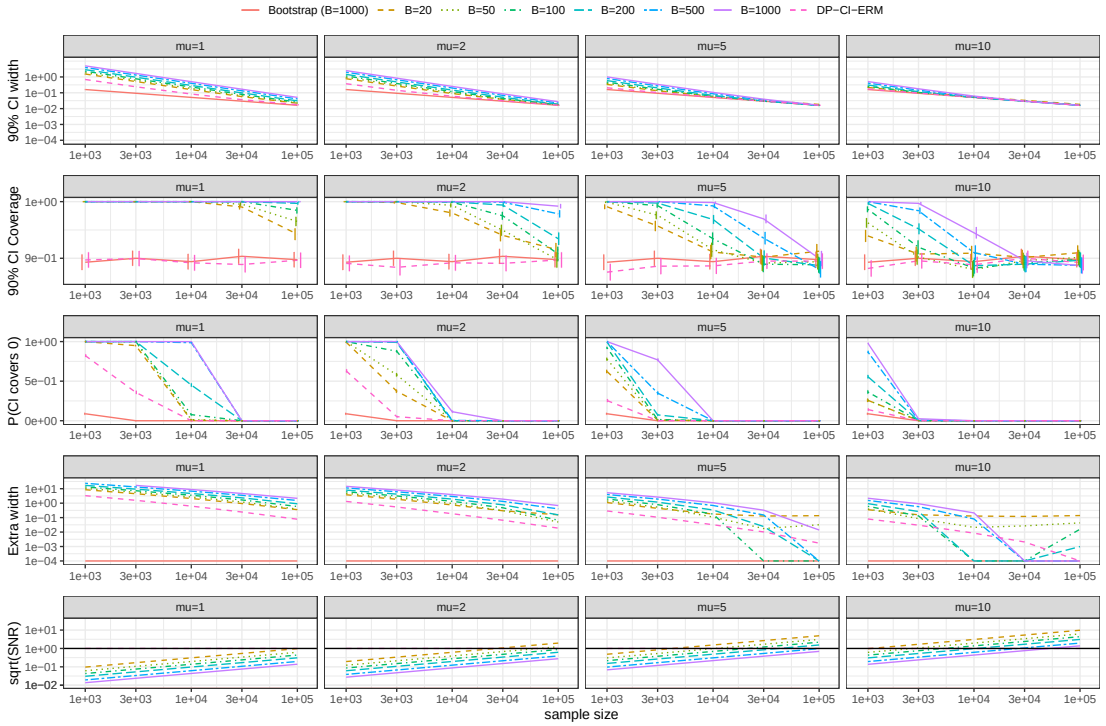

Figure 11: Results for various choices of B and μ-GDP for logistic regression with $c = 0.01$.

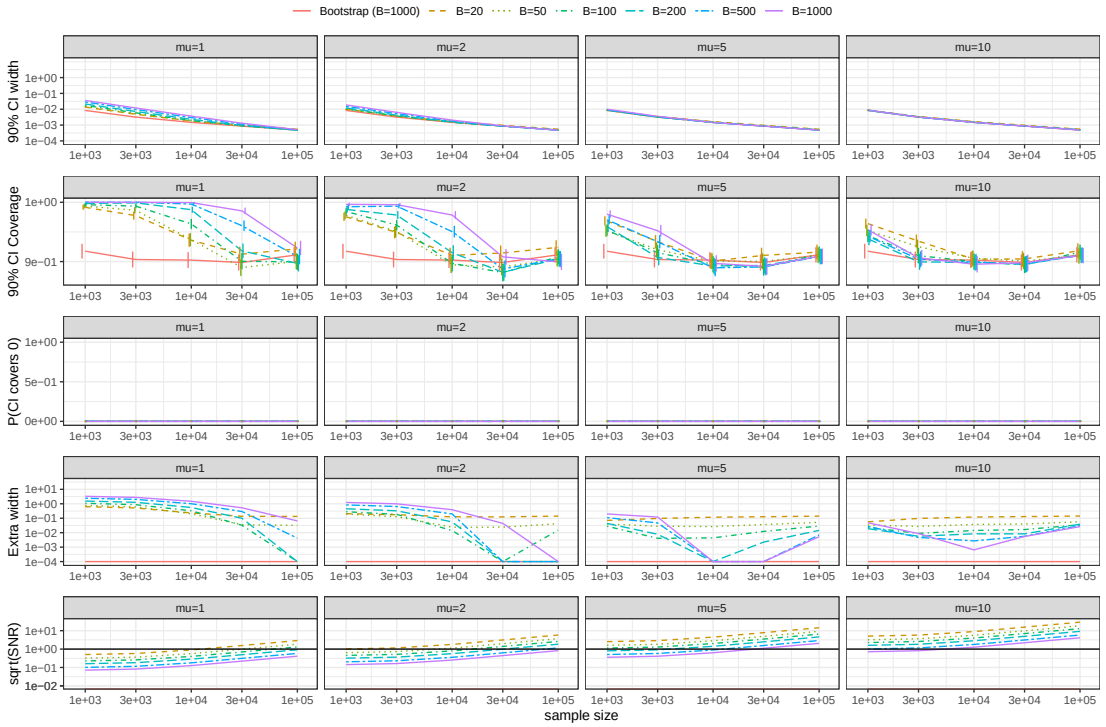

Figure 12: Results for various choices of B and μ-GDP for quantile regression with $c = 1$.

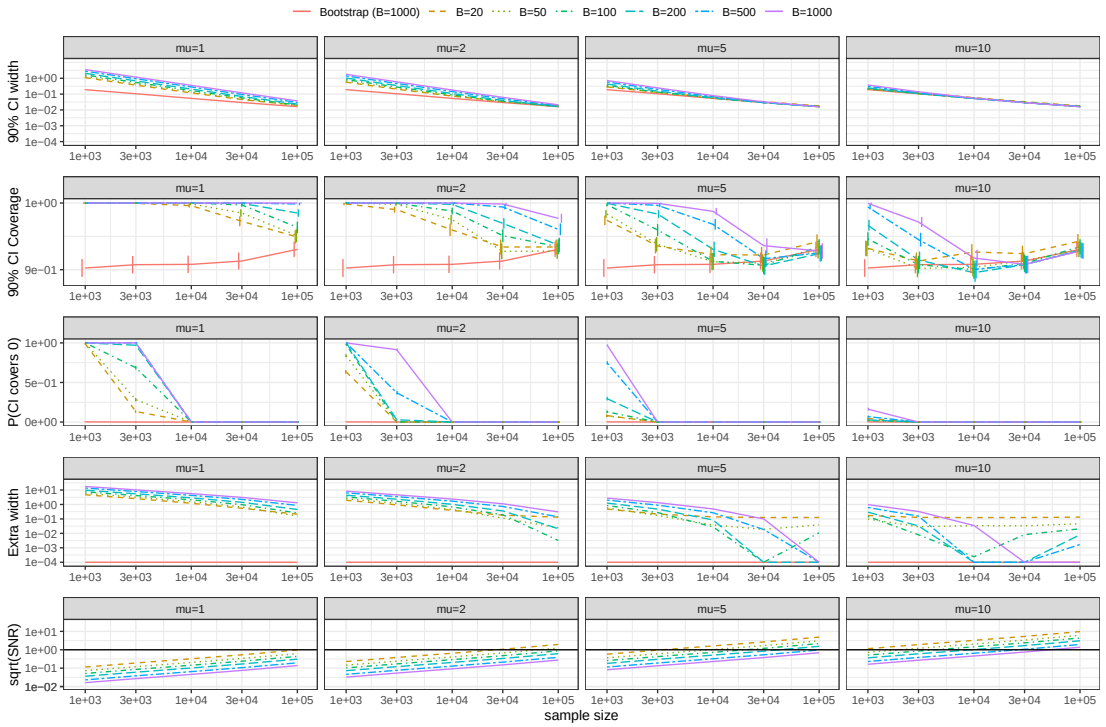

Figure 13: Results for various choices of B and μ-GDP for quantile regression with $c = 0.01$.

PCA. In Proceedings of the 36th International Conference on Machine Learning, volume 97, pages 374–384. PMLR, 2019.

Borja Balle, Gilles Barthe, and Marco Gaboardi. Privacy amplification by subsampling: Tight analyses via couplings and divergences. In Advances in Neural Information Processing Systems, volume 31, 2018.

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

Sourav Biswas, Yihe Dong, Gautam Kamath, and Jonathan Ullman. Coinpress: Practical private mean and covariance estimation. In Advances in Neural Information Processing Systems, volume 33, pages 14475–14485, 2020.

Jonathan M. Borwein and Jon D. Vanderwerff. Convex Functions: Constructions, Characterizations and Counterexamples. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2010.

Stephen Boyd and Lieven Vandenberghe. Convex Optimization. Cambridge University Press, 2004.

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

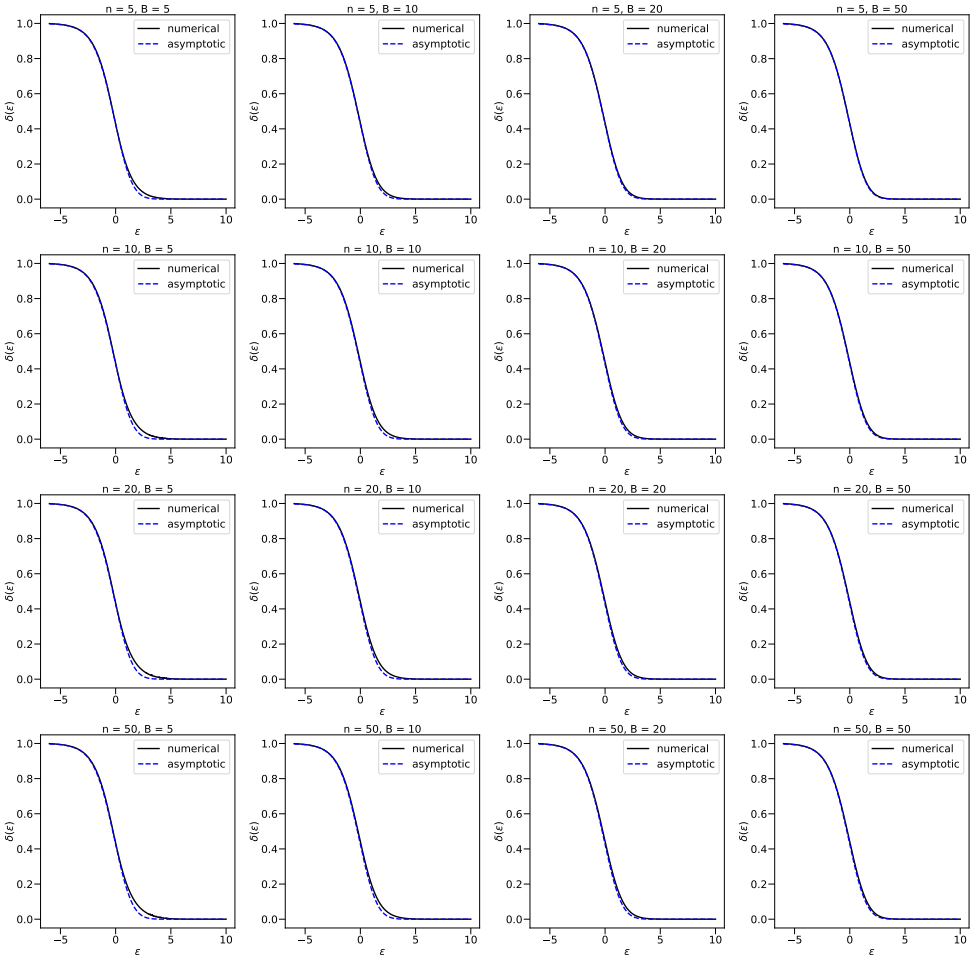

Figure 14: Comparison between the privacy profile $\delta(\varepsilon)$ of composition computed from asymptotics (Theorem 16) and numerical evaluation (Proposition 14).

Mark Bun and Thomas Steinke. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In Proceedings, Part I, of the 14th International Conference on Theory of Cryptography - Volume 9985, page 635–658, 2016.

Mark Bun, Cynthia Dwork, Guy N. Rothblum, and Thomas Steinke. Composable and versatile privacy via truncated cdp. In Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, page 74–86, 2018.

Tony Cai, Yichen Wang, and Linjun Zhang. The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy. The Annals of Statistics, 49(5): 2825–2850, 2021.

Statistics Canada. 2016 Census Public Use Microdata File (PUMF). Individuals File, 2019.

Karan Chadha, John Duchi, and Rohith Kuditipudi. Resampling methods for private statistical inference. arXiv preprint arXiv:2402.07131, 2024.

Kamalika Chaudhuri, Claire Monteleoni, and Anand D Sarwate. Differentially private empirical risk minimization. Journal of Machine Learning Research, 12(3), 2011.

Christian Covington, Xi He, James Honaker, and Gautam Kamath. Unbiased statistical estimation and valid confidence intervals under differential privacy. Statistica Sinica, 35: 651–670, 2025.

Thomas Diciccio and Bradley Efron. More accurate confidence intervals in exponential families. Biometrika, 79(2):231–245, 1992.

Jinshuo Dong, Aaron Roth, and Weijie J. Su. Gaussian differential privacy. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 84(1):3–37, 2022.

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

Jörg Drechsler, Ira Globus-Harris, Audra Mcmillan, Jayshree Sarathy, and Adam Smith. Nonparametric differentially private confidence intervals for the median. Journal of Survey Statistics and Methodology, 10(3):804–829, 2022.

Wenxin Du, Canyon Foot, Monica Moniot, Andrew Bray, and Adam Groce. Differentially private confidence intervals. arXiv preprint arXiv:2001.02285, 2020.

Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. Foundations and Trends® in Theoretical Computer Science, 9(3–4):211–407, 2014.

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

Cynthia Dwork, Guy N. Rothblum, and Salil Vadhan. Boosting and differential privacy. In IEEE 51st Annual Symposium on Foundations of Computer Science, pages 51–60, 2010.

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

Bradley Efron. Better bootstrap confidence intervals. Journal of the American Statistical Association, 82(397):171–185, 1987.

Bradley Efron. Empirical Bayes deconvolution estimates. Biometrika, 103(1):1–20, 2016.

Bradley Efron and Robert J Tibshirani. An Introduction to the Bootstrap. CRC press, 1994.

Farhad Farokhi. Deconvoluting kernel density estimation and regression for locally differentially private data. Scientific Reports, 10(1):1–11, 2020.

Cecilia Ferrando, Shufan Wang, and Daniel Sheldon. Parametric bootstrap for differentially private confidence intervals. In International Conference on Artificial Intelligence and Statistics, pages 1598–1618. PMLR, 2022.

Sivakanth Gopi, Yin Tat Lee, and Daogao Liu. Private convex optimization via exponential mechanism. In Conference on Learning Theory, pages 1948–1989. PMLR, 2022.

Peter Hall. On the bootstrap and likelihood-based confidence regions. Biometrika, 74(3): 481–493, 09 1987.

Martin L. Hazelton and Berwin A. Turlach. Nonparametric density deconvolution by weighted kernel estimators. Statistics and Computing, 19(3):217–228, Sep 2009.

Vishesh Karwa and Salil Vadhan. Finite sample differentially private confidence intervals. In 9th Innovations in Theoretical Computer Science Conference (ITCS 2018), volume 94, pages 44:1–44:9, 2018.

Antti Koskela, Joonas Jälkö, and Antti Honkela. Computing tight differential privacy guarantees using fft. In Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics, volume 108, pages 2560–2569. PMLR, 2020.

Christine M O'Keefe and Anne-Sophie Charest. Bootstrap differential privacy. Transactions on Data Privacy, 12(1):1–28, 2019.

Matthew Reimherr and Jordan Awan. Kng: The k-norm gradient mechanism. In Advances in Neural Information Processing Systems, volume 32, 2019.

R Tyrrell Rockafellar. Convex Analysis, volume 18. Princeton University Press, 1997.

Abhra Sarkar, Debdeep Pati, Antik Chakraborty, Bani K. Mallick, and Raymond J. Carroll. Bayesian semiparametric multivariate density deconvolution. Journal of the American Statistical Association, 113(521):401–416, 2018.

Robert J Serfling. Approximation Theorems of Mathematical Statistics. John Wiley & Sons, 2009.

Thomas A Severini. Elements of Distribution Theory. Number 17. Cambridge University Press, 2005.

Jun Shao. Mathematical Statistics. Springer Science & Business Media, 2003.

Or Sheffet. Differentially private ordinary least squares. In Proceedings of the 34th International Conference on Machine Learning, volume 70. PMLR, 2017.

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

Chendi Wang, Buxin Su, Jiayuan Ye, Reza Shokri, and Weijie Su. Unified enhancement of privacy bounds for mixture mechanisms via f-differential privacy. In Advances in Neural Information Processing Systems, volume 36, 2023.

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 and Shuheng Zhou. A statistical framework for differential privacy. Journal of the American Statistical Association, 105(489):375–389, 2010.

Min-ge Xie and Peng Wang. Repro samples method for finite-and large-sample inferences. arXiv preprint arXiv:2206.06421, 2022.

Élie Youndjé and Martin T. Wells. Optimal bandwidth selection for multivariate kernel deconvolution density estimation. TEST, 17(1):138–162, May 2008.

Qinqing Zheng, Jinshuo Dong, Qi Long, and Weijie Su. Sharp composition bounds for Gaussian differential privacy via edgeworth expansion. In Proceedings of the 37th International Conference on Machine Learning, volume 119, pages 11420–11435. PMLR, 2020.