Federated Experiment Design under Distributed Differential Privacy

Wei-Ning Chen[1,2] **Graham Cormode**[2,3] **Akash Bharadwaj**[2,4] **Peter Romov**[2] **Ayfer Özgür**[1]

Stanford University[1] Meta[2] University of Warwick[3] ByteDance[4]

Abstract

Experiment design has a rich history and has found many critical applications across various fields since then. The use and collection of users' data in experiments often involve sensitive personal information, so additional measures to protect individual privacy are required. In this work, we focus on the rigorous protection of users' privacy (under the notion of differential privacy (DP)) while minimizing the trust toward service providers. Specifically, we consider the estimation of the average treatment effect (ATE) under DP, while only allowing the analyst to collect population-level statistics via secure aggregation, a distributed protocol enabling a service provider to aggregate information without accessing individual data. Although a vital component in modern A/B testing workflows, private distributed experimentation has not previously been studied. To achieve DP, we design local privatization mechanisms that are compatible with secure aggregation and analyze the utility in terms of the width of confidence intervals, both asymptotically and non-asymptotically. We show how these mechanisms can be scaled up to handle the very large number of participants commonly found in practice. In addition, when introducing DP noise, it is imperative to cleverly split privacy budgets to estimate both the mean and variance of the outcomes and carefully calibrate the confidence intervals according to the DP noise. Last, we present comprehensive experimental evaluations of our proposed schemes and show the privacy-utility trade-offs in experiment design.

Proceedings of the 27[th] International Conference on Artificial Intelligence and Statistics (AISTATS) 2024, Valencia, Spain. PMLR: Volume 238.

1 INTRODUCTION

Experimental design has a long history, tracing back to the early 1920s in the agricultural domain (Fisher, 1921, 1922, 1936), where statisticians used mathematical tools to design and analyze experiments. Since then, experimental design has found many applications, e.g., in chemistry, manufacturing, pharmaceuticals, and technology, etc. It not only enables the comparison of specific design alternatives, but also facilitates the production of generalizable knowledge to inform strategic decision-making. When designing experiments to estimate or test the effect of a treatment (for example, a tech company launching a new feature in an existing product), a standard procedure is to divide participants into test and control groups, introduce changes ("the treatment") to the test group, and collect feedback or outcomes from both groups to conduct further statistical analysis. When the test assignment is properly randomized and the estimators or tests for the outcomes are designed adequately, the analyst can infer the treatment effect and make decisions accordingly.

However, in many modern applications, such as pharmaceutical and online experimental designs, experimentation usually involves participants' private data, raising additional concerns about privacy and security. Thus, when designing and conducting experiments involving sensitive personal information, additional safeguards are desirable to protect it.

One way to enforce rigorous privacy for experiments is by restricting the final tests or estimators used to be differentially private (DP) (Dwork et al., 2006b). In brief, DP defines a formal notion of privacy that quantifies the amount of information leakage of an algorithm. It ensures that the output of a (randomized) algorithm $\mathcal{A}$ does not depend strongly on the contribution of any one individual. To achieve DP, a standard approach is to add carefully calibrated noise to the test statistics (e.g., the Laplace or Gaussian mechanisms (Dwork et al., 2006b, 2014)) and only using the perturbed results in downstream tasks. This approach is usually referred to as "Central DP", since an analyst collects all the

experimental data centrally before sanitizing the test statistics. While Central DP schemes control the view of downstream tasks and are relatively straightforward to design, the analyst stores and processes all the raw users' data in the clear. This not only requires the experiment participants to trust the analyst, but could make it challenging to comply with regulations on the storage of certain forms of personal data.

To address the above issues, an alternative approach is to aggregate test data in a "secure" way so that only necessary population-level statistics are collected and that analysts can never see individual data. Secure aggregation can be achieved by secure hardware or cryptographic multiparty computation (MPC) (Ben-Or et al.; Damgård et al., 2012) and is the focus of "federated learning and analytics" (Kairouz et al., 2019). Secure aggregation alone does not provide any formal differential privacy guarantees. To ensure DP, participants can locally randomize their data so that the securely aggregated outcome satisfies the standard DP requirement (Dwork et al., 2006a). This is referred to as Distributed DP (in contrast to Central DP) and is growing in prominence thanks to recent progress in practical aggregation protocols (Bonawitz et al., 2016; Bell et al., 2020). With secure aggregation and Distributed DP, one can minimize the level of trust in the data analysts and service providers.

In this work, we focus on experimental design with Distributed DP. Specifically, we consider estimating and testing the average treatment effect (ATE), subject to DP and secure aggregation constraints. In our framework, to construct private protocols, we make use of a secure aggregation primitive that we refer to as SecAgg, which can be instantiated by Bonawitz et al. (2016); Bell et al. (2020).

Our contributions are as follows:

- We present a decentralized framework that achieves a $(1-\alpha)$-confidence interval (CI) and a level-α test while ensuring distributed DP (defined in Section 2). We analyze the width of private confidence intervals and provide asymptotic and non-asymptotic guarantees. Our non-asymptotic bounds are based on a version of empirical Bernstein inequality, which guides how to allocate privacy budgets in estimating mean and variance.

- We incorporate the Poisson-binomial mechanism (PBM) (Chen et al., 2022) in our framework as the local randomizer, which offers several advantages, including unbiased estimation, efficient memory (or communication) usage, and bounded sensitivities, letting downstream parties develop privatization mechanisms.

- To use PBM for experimental design, we develop an improved privacy accounting tool based on a novel bound on the Rényi divergence. This enhancement greatly enhances efficiency in large sample scenarios. When the objective is to obtain CIs instead of point estimators, we must collect second-moment information such as sample variance. We show, via SecAgg and DP, that this can be done by judiciously allocating privacy budgets for estimating sample mean and variance.

- Last, our experimental study quantifies the trade-offs between privacy and utility.

1.1 Related Works

Private causal inference and testing. The design of experiments to identify causal relations and average treatment effects is crucial in various domains (Imbens and Rubin, 2015); when experiments involve sensitive data, additional privacy protection is needed such as differential privacy (DP). D'Orazio et al. (2015) proposes DP mechanisms for summary statistics in causal inference, and Lee et al. (2019); Niu et al. (2022); Ohnishi and Awan (2023) consider estimating conditional average treatment effects (CATE) and propose private estimation of inverse propensity scores. These works default to a Central DP setting where a central data curator collects and privatizes test statistics, while Ohnishi and Awan (2023) explore Local DP without a trusted curator. In contrast, we address the experimental design problem using Distributed DP via secure aggregation as a better compromise between privacy and security. Our experiment design problem is related to private hypothesis testing, which performs two-sample tests under DP when potential outcomes come from an unknown distribution. Previous work on two-sample tests has primarily focused on either Central DP (Rogers and Kifer, 2017; Cai et al., 2017; Raj et al., 2020) or Local DP (Raj et al., 2020). This work is the first to consider experimentation under Distributed DP with secure aggregation. We also analyze the distribution-free setting, where no distributional assumptions are imposed on potential outcomes.

Private mean estimation. The mechanisms in this paper are based on the difference-in-mean estimator, which relies on private mean estimation as a sub-routine. Differentially private mean estimation has been extensively studied under Central DP (Dwork et al., 2006b, 2014; Balle and Wang, 2018; Agarwal et al., 2018; Biswas et al., 2020) or Local DP (Duchi et al., 2013; Bhowmick et al., 2018; Chen et al., 2020; Feldman and Talwar, 2021). In addition to obtaining a point estimator for the mean, it is also desirable to obtain a $(1-\alpha)$-confidence interval (CI) for a level-α test. Existing methods either privately estimate both

sample means and variances separately (Du et al., 2020; Karwa and Vadhan, 2017; D'Orazio et al., 2015) or use a private bootstrap (Brawner and Honaker, 2018). Our approach resembles the former but is compatible with secure aggregation and does not require a central data curator. In addition, all of the previous methods, to our knowledge, study the *asymptotic* CIs, while in this work, we also characterize the non-asymptotic coverage guarantees with finite n. The only exception that also considers non-asymptotic bounds is the recent work Waudby-Smith et al. (2023). However, Waudby-Smith et al. (2023) considers a Local DP setting, so the analyst can directly estimate the mean and variance based on the locally private samples.

Secure aggregation and distributed DP. Our methods aggregate test data using secure aggregation protocols (specifically, single-server aggregation) to achieve distributed DP without introducing bias. Single-server secure aggregation is performed via additive masking over a finite group (Bonawitz et al., 2016; Bell et al., 2020). However, to achieve provable privacy guarantees, secure aggregation is insufficient as the sum of local model updates may still leak sensitive information (Melis et al., 2019; Song and Shmatikov, 2019; Carlini et al., 2019; Shokri et al., 2017). For DP, participants have to privatize their raw data with local noise before secure aggregation (Dwork et al., 2006a). This local noise has to be compatible with the secure aggregation protocol; candidate solutions include (Agarwal et al., 2018; Kairouz et al., 2021; Agarwal et al., 2021; Chen et al., 2022). Here, we aim to provide privacy guarantees in the form of Rényi DP (Mironov, 2017) because it allows for tracking the end-to-end privacy loss tightly. We distinguish our Distributed DP model from Local DP (Kasiviswanathan et al., 2011; Evfimievski et al., 2004; Warner, 1965), where data is perturbed on the client-side before the server collects it in the clear. Although simpler to implement, Local DP naturally suffers from poor privacy-utility trade-offs, as much more noise is introduced in total (Kasiviswanathan et al., 2011; Duchi et al., 2013).

2 PROBLEM SETUP AND PRELIMINARIES

We formulate the experiment design problem via the *Neyman-Roubin causal model*. When a service provider considers introducing a new feature to the public, it initiates a test phase by selecting a small group of users. This group is randomly divided into two: a test group where users are exposed to the new feature (referred to as the treatment), and a control group where users do not have access to the feature. The service provider collects responses from both groups, assesses the effects of the treatment, and, based on the evaluation, makes a decision regarding the launch of the new feature.

Formally, we define the experiment design problem as follows: for each test unit ("user") $i \in [n]$, we introduce the randomized treatment assignment variable $T_i \in \{c, t\}$ (for the control and test group, respectively), which indicates whether user i receives the treatment or not. Additionally, we consider the potential outcomes $y_i(t), y_i(c) \in \mathcal{Y}$ for user i when receiving or not receiving the treatment, respectively. For a test unit i, the service provider can only observe one of its potential outcomes: $X_i \triangleq y_i(T_i)$. The quantity of interest is the sample average treatment effect (SATE):

$$\Delta_{\mathsf{s}}(\mathbf{y}) \triangleq \frac{1}{n}\sum_{i=1}^{n} y_i(t) - y_i(c).$$

Notice that under Neymann's original framework, the potential outcomes

$$\mathbf{y} \triangleq \{(y_i(t), y_i(c)) | i = 1, ..., n\}$$

are deterministic; only the treatment variable T_i's are randomized. However, we can also impose distributional assumptions on the potential outcomes, i.e., $y_i(c) \overset{\text{i.i.d.}}{\sim} P_c$ and $y_i(t) \overset{\text{i.i.d.}}{\sim} P_t$, and the quantity of interest is the population average treatment effect (PATE):

$$\Delta_{\mathsf{p}}(P_c, P_t) \triangleq \mathbb{E}_{Y(t)\sim P_t,\, Y(c)\sim P_c}\left[Y(t) - Y(c)\right].$$

Our goal is to test if $\Delta_{\mathsf{s}} > 0$ (or $\Delta_{\mathsf{p}} > 0$) at a given confidence level α, which is equivalent to construct $(1-\alpha)$ confidence intervals of Δ_{s} (or Δ_{p}).

2.1 Secure aggregation and DP

When the service provider has access to all the observable data, it can estimate Δ_{s} via standard causal inference in statistics, social, and biomedical sciences (see, for instance, Imbens and Rubin (2015)), compute sample variances of $y_i(c)$'s and $y_i(t)$'s, and construct confidence intervals accordingly. However, when the samples X_i are treated as sensitive, they should be aggregated securely so that only necessary information is revealed to the service providers.

Secure aggregation. Recently, distributed protocols based on multi-party computation (MPC), such as secure aggregation (SecAgg, Bonawitz et al. (2016)), have emerged as powerful tools for securely aggregating population-level information from a group of users. Specifically, SecAgg enables a single server to compute the population sum and, consequently, the average of local variables while ensuring that no additional information, apart from the sum, is disclosed to the server or other participating entities. These properties make

SecAgg well-suited for aggregating experiment results from users to estimate or test Average Treatment Effects (ATE). This is because test statistics used for ATE estimation can often be expressed as a function of the average of users' potential outcomes. However, when applying SecAgg in experiment design, it is important to note that SecAgg typically operates on a finite field, like most cryptographic MPC protocols. Thus, each outcome X_i needs to be appropriately pre-processed (e.g., discretized) and mapped into a finite field.

Differential privacy. Secure aggregation alone does not provide any provable privacy guarantees. Sensitive information may still be revealed from the aggregated population statistics, causing potential privacy leakage. To address this issue, differential privacy (DP, Dwork et al. (2006b)) has been adopted as the gold standard that ensures individual information is not leaked. Specifically, it requires the ATE estimator (or a CI of ATE) released by the service provider to meet the following guarantee:

Definition 2.1 (Differential privacy) *We say an ATE estimator $\hat{\Delta}(X^n)$ is (ε, δ)-DP, if for any two possible outcome sets $\mathbf{y} \triangleq \{(y_i(c), y_i(t)) | i = 1, ..., n\}$ and $\mathbf{y}' \triangleq \{(y_i(c), y_i(t)) | i = 2, ..., n\} \cup \{(y_1'(0), y_1'(1))\}$ differing in one user, we have*

$$\Pr\left\{\hat{\Delta}(X^n|\mathbf{y}) \in \mathcal{S}\right\} \leq e^{\varepsilon} \Pr\left\{\hat{\Delta}(X^n|\mathbf{y}') \in \mathcal{S}\right\} + \delta,$$

for any measurable set $\mathcal{S}$.

A common approach to achieve DP is adding properly calibrated noise (such as Gaussian noise with appropriate variance) to standard (non-private) ATE estimators. However, this requires users to trust the service provider as the server can see the unprivatized aggregated information. To address this issue, one can instead *locally* perturb individual outcome X_i before secure aggregation via a local randomizer $\mathcal{M}(X_i)$. When the local noise mechanism $\mathcal{M}$ is designed in a way that the sum $\sum_i \mathcal{M}(X_i)$ satisfies DP, i.e.,

$$\Pr\left\{\sum_i \mathcal{M}(X_i) \in \mathcal{S} \middle| \mathbf{y}\right\} \leq e^{\varepsilon} \Pr\left\{\sum_i \mathcal{M}(X_i) \in \mathcal{S} \middle| \mathbf{y}'\right\} + \delta, \quad (1)$$

and when $\mathcal{M}(X_i)$'s are aggregated securely, one can ensure DP *even if the service provider is not trusted*. The idea of combining secure MPC with local noise dates back to Dwork et al. (2006a) and has been used extensively in private federated learning and analytics (Kairouz et al., 2021; Agarwal et al., 2018, 2021). The main challenge is that the local noise has to be discretized and compatible with secure aggregation; i.e., $\mathcal{M}$ has to map X_i into a space $\mathcal{Z}$ (a finite field, e.g., the integers modulo a prime p) for SecAgg to work in.

In addition to the above (ε, δ)-DP, we also use the following Rényi DP definition, which allows simpler and tighter privacy composition.

Definition 2.2 (Rényi differential privacy) *We say an ATE estimator $\hat{\Delta}(X^n)$ is $(\alpha, \varepsilon(\alpha))$-DP, if for any two neighboring sets of possible outcomes $\mathbf{y}$ and $\mathbf{y}'$ that differ in one user, it holds that*

$$D_\alpha\left(\hat{\Delta}(X^n|\mathbf{y}) \middle\| \hat{\Delta}(X^n|\mathbf{y}')\right) \triangleq \frac{1}{\alpha - 1} \log \mathbb{E}_{X \sim \hat{\Delta}(X^n|\mathbf{y})}\left[\left(\frac{f_{\hat{\Delta}(X^n|\mathbf{y})}(X)}{f_{\hat{\Delta}(X^n|\mathbf{y}')}(X)}\right)^{\alpha}\right] \leq \varepsilon(\alpha).$$

Similarly, for a local randomizer $\mathcal{M} : \mathcal{X} \to \mathcal{Z}$, we can define the following distributed Rényi DP.

Definition 2.3 (Distributed Renyi DP) *A local randomizer $\mathcal{M}$ is $(\alpha, \varepsilon(\alpha))$-DP, if, for any two neighboring outcome sets $\mathbf{y}$ and $\mathbf{y}'$ differing in one user:*

$$D_\alpha\left(\sum_i \mathcal{M}(X_i|\mathbf{y}) \middle\| \sum_i \mathcal{M}(X_i|\mathbf{y}')\right) \leq \varepsilon(\alpha).$$

Remark 2.4 *In the above framework, the treatment variables T_i's are randomly assigned by the server, and hence typically do not carry users' individual information. So by default, the server inherently knows all the treatment variables, and we do not view them as private information. Nevertheless, we can accommodate the case of private treatments in our work by having each client also send a "dummy" outcome on top of the actual one (i.e., setting $Y_i(t) = 0$ for clients in the control group). By introducing proper noise, we can still achieve the same level of DP.*

3 A DISTRIBUTED DP FRAMEWORK

Our objective is to construct a $(1 - \alpha)$-confidence interval for SATE and PATE (which can then be used to design a level-α test) while adhering to the distributed differential privacy (DP) constraint mentioned in equation (1). In Algorithm 1, we presented a general framework for causal inference using secure aggregation and distributed DP.

In this framework, the server securely aggregates necessary information from the control and test groups separately, along with local randomizers $\mathcal{M}_1$ and $\mathcal{M}_2$.

Algorithm 1 ATE Estimation with Distributed DP

Input: treatment variables $T_1, ..., T_n \in \{c, t\}$, outcomes $(y_1(T_1), ..., y_n(T_n))$, randomizers $\mathcal{M}_1, \mathcal{M}_2$, privacy budgets ε_1 and ε_2, $\mathsf{obj} \in \{\text{'SATE'}, \text{'PATE'}\}$.
Output: an $(1-\alpha)$-CI for ATE.

▷ **Local Randomization**
for each user i **do**
Obtains the observable outcome $X_i \triangleq y_i(T_i)$.
Computes $\mathcal{M}(X_i)$, and $\mathcal{M}(X_i^2)$.

▷ **Aggregation**
Server securely aggregates

$$\begin{cases} \sum_{i\in S_t} \mathcal{M}_1(X_i, n_t),\ \sum_{i\in S_t} \mathcal{M}_2(X_i^2, n_t); \\ \sum_{i\in S_c} \mathcal{M}_1(X_i, n_c),\ \sum_{i\in S_c} \mathcal{M}_2(X_i^2, n_c), \end{cases}$$

where $S_c \triangleq \{i : T_i = c\}$ and $S_t \triangleq \{i : T_i = t\}$.

▷ **Estimation**
Estimates sample means and variances:

$$\hat{\mu}_c\left(\sum_{i\in S_c} \mathcal{M}_1(X_i, n_c)\right) \text{ and } \hat{\mu}_t\left(\sum_{i\in S_t} \mathcal{M}_1(X_i, n_t)\right);$$

$$\hat{s}_c^2\left(\sum_{i\in S_c} \mathcal{M}_2(X_i^2, n_c), \hat{\mu}_c\right) \text{ and } \hat{s}_t^2\left(\sum_{i\in S_t} \mathcal{M}_2(X_i^2, n_t), \hat{\mu}_t\right).$$

Computes the diff-in-mean estimator $\hat{\Delta} \triangleq \hat{\mu}_t - \hat{\mu}_c$.
Computes the variance calibration term $\sigma_{\mathsf{pr}}^2(\varepsilon, n_c, n_t)$ according to (2) and let $z_{1-\alpha/2}$ be the $(1-\alpha/2)$-quantile of the standard normal.
if objective is 'SATE' **then**
Set $\hat{\sigma}_{\mathsf{s}}^2 \triangleq \frac{n_c n_t}{n}\left(\frac{\sqrt{\hat{s}_t^2}}{n_t} + \frac{\sqrt{\hat{s}_c^2}}{n_c}\right)^2$.
Return: $\hat{\Delta}_{\mathsf{s}} \pm z_{1-\alpha/2} \cdot (\hat{\sigma}_{\mathsf{s}} + \sigma_{pr})$.
if objective is 'PATE' **then**
Set $\hat{\sigma}_{\mathsf{p}}^2 \triangleq \frac{\hat{s}_t^2}{n_t} + \frac{\hat{s}_c^2}{n_c}$.
Return: $\hat{\Delta}_{\mathsf{p}} \pm z_{1-\alpha/2} \cdot (\hat{\sigma}_{\mathsf{p}} + \sigma_{\mathsf{pr}})$.

These randomizers satisfy the distributed DP conditions defined in Definition 2.3 and map individual observable outcomes X_i and their second moments X_i^2 to the finite field on which secure aggregation operates. Specifically, we have:

$$\begin{cases} \mathcal{M}_1 : \mathcal{X} \times [n] \to \mathcal{Z}; \\ \mathcal{M}_2 : \mathcal{X}_2 \times [n] \to \mathcal{Z}, \end{cases}$$

where we use $\mathcal{X}_2 \triangleq \{x^2 | x \in \mathcal{X}\}$ to denote the collection of all possible second moments of the samples. In the above notation, we allow the local randomizers to take the size of the control (or test) group, denoted as $n_c \triangleq \sum_{i=1}^n \left(1 - \mathbb{1}_{\{T_i=c\}}\right)$ (or $n_t \triangleq n - n_c$), as an input. This enables the local randomizers to calibrate the noise level based on the group size. As in Neyman-Robin's potential outcome framework, the test assignment variables $(T_1, ..., T_n)$ follow a uniform distribution across all sequences containing 'c' n_c times. We consider bounded observations and without loss of generality, we assume the outcome domain $\mathcal{X}$ is centered at 0:

Assumption 3.1 *Let $\mathcal{X} = [-R, R]$ (so we must have $\mathcal{X}_2 = [0, R^2]$).*

After receiving the aggregated information, the server constructs unbiased estimators for the sample means and variances of each group. The difference-in-means estimator is then used to estimate the ATEs. The second-moment information is used for variance estimation, which is needed for confidence intervals.

3.1 Privacy of Algorithm 1

The following theorem establishes privacy guarantees for the framework:

Theorem 3.2 *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be local randomizers for the first and second moments of X_i. Assume for all $n^* \in [n]$, $\mathcal{M}_j(\cdot, n^*)$ satisfies $(\alpha, \varepsilon_j(\alpha))$-distributed Rényi DP for $j \in \{1, 2\}$. Then, Algorithm 1 is $(\alpha, \varepsilon_1(\alpha) + \varepsilon_2(\alpha))$-Rényi DP.*

Proof. Since both $\hat{\Delta}_{\mathsf{s}}$ and $\hat{\sigma}_{\mathsf{s}}$ are functions of $\hat{\mu}_c, \hat{\mu}_t, \hat{s}_c^2$, and $\hat{s}_t^2$, we only need to ensure their Rényi DP due to the post-processing properties of DP. The Rényi DP follows from a simple application of the composition theorem for Rényi DP (Mironov, 2017). □

Note that although $\mathcal{M}_1$ and $\mathcal{M}_2$ are invoked twice in Algorithm 1, we only pay the privacy penalty once since one of the test or control groups remains the same for two neighboring datasets $\mathbf{y}$ and $\mathbf{y}'$.

3.2 Asymptotic Coverage Guarantees

Next, we claim that Algorithm 1 gives a $(1-\alpha)$-CI asymptotically.

Assumption 3.3 *Assume the estimator $\hat{\mu}_j$, $j \in \{c, t\}$, are of an additive structure. That is, $\hat{\mu}_t = \sum_{i\in S_t} \hat{\mu}(\mathcal{M}_1(X_i))$ and $\hat{\mu}_c = \sum_{i\in S_c} \hat{\mu}(\mathcal{M}_1(X_i))$, where $\hat{\mu}(\mathcal{M}_1(x_i, n^*))$ gives an unbiased estimator, independent of T_i, on x_i with variance bounded by $\sigma_1^2(n^*, \varepsilon)$[1];*

Assumption 3.4 *Assume $\hat{s}_c^2$ and $\hat{s}_t^2$ defined in Algorithm 1 yield consistent estimation on the sample variances $s_c^2 \triangleq \frac{1}{n-1}\sum_{i\in[n]}(y_i(c) - \bar{y}(c))^2$ and*

[1] Indeed, we can relax the unbiasedness assumption and only require $\mathbb{E}[\hat{\mu}(\mathcal{M}_1(x_i, n^*))] = o(\frac{1}{n})$.

$s_t^2 \triangleq \frac{1}{n-1}\sum_{i\in[n]}(y_i(t)-\bar{y}(t))^2$, *respectively. That is,* $\hat{s}_c^2\left(\sum_{i\in S_c}\mathcal{M}_2(X_i^2)\right) \xrightarrow{p} s_c$ *as* $n\to\infty$ *(and so does* $\hat{s}_t$*).*

Theorem 3.5 *Let the calibration term (which depends on* $\mathcal{M}_1$*) be*

$$\sigma_{\mathsf{pr}}^2(n_c,n_t,\varepsilon) \triangleq \frac{n}{n_c}\sigma_1^2(n_c,\varepsilon)+\frac{n}{n_t}\sigma_1^2(n_t,\varepsilon). \quad (2)$$

Then, under assumptions 3.3 and 3.4, Algorithm 1 gives a $(1-\alpha)$*-confidence interval of SATE or PATE.*

Proof. We follow the standard analysis of the difference-in-mean estimator and incorporate the DP noise. To begin with, we analyze the unprivatized estimator. Let $\hat{\nu}_t \triangleq \frac{1}{n_t}\sum_i T_i y_i(0)$ and $\hat{\nu}_c \triangleq \frac{1}{n_c}\sum_i(1-T_i)y_i(1)$ be the unprivatized means of the test and control groups. In addition, let $s_c^2 \triangleq \frac{1}{n-1}\sum_i(y_i(0)-\bar{y}(0))^2$ and $s_t^2 \triangleq \frac{1}{n-1}\sum_i(y_i(1)-\bar{y}(1))^2$ be the sample variances; let $s_{tc} \triangleq \frac{1}{n-1}\sum_i(y_i(0)-\bar{y}(0))(y_i(1)-\bar{y}(1))$ be the sample covariance. Then, the variance of the (unprivatized) difference-in-mean estimator can be computed as

$$\mathsf{Var}\left(\hat{\nu}_t-\hat{\nu}_c|\mathbf{y}\right)=\frac{\sigma_s^2}{n}\triangleq\frac{1}{n}\left(\frac{n_c}{n_t}s_t^2+\frac{n_t}{n_c}s_c^2+s_{tc}\right).$$

The finite-sample central limit theorem (Hájek, 1961) (see also Li and Ding (2017); Li et al. (2018)) suggests that

$$\sqrt{n}\left((\hat{\nu}_t-\hat{\nu}_c)-\Delta_s\right)\xrightarrow{d}N(0,\sigma_s^2).$$

When there exists DP noise, we have, conditioned on $\mathbf{y}$ and T_i,

$$\sqrt{n}\left((\hat{\mu}_c-\hat{\mu}_t)-(\hat{\nu}_t-\hat{\nu}_c)\right)\xrightarrow{d}N\left(0,\sigma_{\mathsf{pr}}^2(n_c,n_t,\varepsilon)\right),$$

where $\sigma_{\mathsf{pr}}^2(n_c,n_t,\varepsilon)\triangleq\frac{n}{n_c}\sigma_1^2(n_c,\varepsilon)+\frac{n}{n_t}\sigma_1^2(n_t,\varepsilon)$ and the convergence is due to the (classical) central limit theorem and Assumption 3.3. Since the DP noise is independent with T_i, we conclude

$$\sqrt{n}\left((\hat{\mu}_c-\hat{\mu}_t)-\Delta_s\right)\xrightarrow{d}N\left(0,\sigma_{\mathsf{pr}}^2(n_c,n_t,\varepsilon)+\sigma_s^2\right),$$

Finally, since $\hat{\sigma}_{\mathsf{s}}^2$ defined in Algorithm 1 is a high probability upper bound on σ_{s}^2 from our assumptions, i.e.,

$$\lim_{n\to\infty}\Pr\left\{\hat{\sigma}_{\mathsf{s}}^2\geq\sigma_{\mathsf{s}}^2\right\}=1,$$

by Slutsky's theorem $\hat{\Delta}_{\mathsf{s}}\pm z_{1-\alpha/2}\cdot(\hat{\sigma}_{\mathsf{s}}+\sigma_{pr})$ gives an $(1-\alpha)$-CI asymptotically.

Next, we prove the coverage guarantee for estimating PATE. Observe that the conditional variance of the (unprivatized) difference-in-mean estimator, given the samples $y_i(0)\overset{\text{i.i.d.}}{\sim}P_0$ and $y_i(1)\overset{\text{i.i.d.}}{\sim}P_1$, can be computed as

$$\mathsf{Var}\left(\hat{\nu}_t-\hat{\nu}_c|\boldsymbol{y}\right)=\frac{1}{n}\left(\frac{n_c}{n_t}s_t^2+\frac{n_t}{n_c}s_c^2+s_{tc}\right).$$

Therefore, the unconditional variance is

$$\begin{aligned}&\mathbb{E}\left[\mathsf{Var}\left(\hat{\nu}_t-\hat{\nu}_c|\boldsymbol{y}\right)\right]+\mathsf{Var}\left(\mathbb{E}\left[\hat{\nu}_t-\hat{\nu}_c|\boldsymbol{y}\right]\right)\\&=\frac{1}{n}\left(\frac{n_c}{n_t}s_t^2+\frac{n_t}{n_c}s_c^2+2s_{tc}\right)+\frac{1}{n}\left(s_t^2+s_c^2-2s_{tc}\right)\\&=\frac{s_t^2}{n_t}+\frac{s_c^2}{n_c}.\end{aligned}$$

As a result, $\hat{\sigma}_{\mathsf{p}}$ in Algorithm 1 is a consistent estimator of the variance of the unprivatized estimator.

With the presence of DP noise, we follow the same analysis as SATE and add a calibration term $\sigma_{\mathsf{pr}}^2(n_c,n_t,\varepsilon)$. By the central limit theorem, the proof is complete. □

We make a few remarks. First, in Algorithm 1, the CIs of SATE and PATE take slightly different forms. This is because the variance of SATE σ_s^2 depends on the sample covariance s_{tc}, which is an unidentifiable quantity. Thus, we can obtain a conservative upper bound $\hat{\sigma}_{\mathsf{s}}$. On the other hand, when the objective is to estimate PATE, the variance of the estimator does not depend on the covariance term, and thus $\hat{\sigma}_{\mathsf{p}}^2$ yields an unbiased estimator on the variance.

Second, in order to determine a suitable treatment assignment size, denoted as n_c and n_t, we can observe that the average length of confidence intervals (CIs) is influenced by two main factors:

$$\begin{aligned}&\frac{\hat{\sigma}_s+\sigma_{\mathsf{pr}}^2(n_c,n_t,\varepsilon)}{n}\\&\approx\underbrace{\frac{n_cn_t}{n}\left(\frac{\hat{s}_c}{n_c}+\frac{\hat{s}_t}{n_t}\right)^2}_{\text{(a)}}+\underbrace{\frac{\sigma_1^2(n_c,\varepsilon)}{n_c}+\frac{\sigma_1^2(n_t,\varepsilon)}{n_t}}_{\text{(b)}}.\end{aligned}$$

The first term (a) depends on the sample variances. To minimize this term, we should set n_c and n_t proportional to the sample variances of the control and treatment groups. However, since the sample variances are often unknown, estimating them requires additional samples and a privacy budget. On the other hand, the second term (b) represents the impact of DP noise. It is important to note that for a given value of ε, the variance of DP noise typically scales as $O\left(\frac{1}{n\min(\varepsilon,\varepsilon^2)}\right)$ (as we will see in the next section). Therefore, if either n_c or n_t is set too small, this term may dominate the total variance.

Apart from determining n_c and n_t, another crucial question is how to allocate the privacy budget for estimating the first and second moments (i.e., the privacy

used in $\mathcal{M}_1$ and $\mathcal{M}_2$). Allocating a significant portion of the privacy budget to estimate the mean (or difference-in-means) estimator can result in a relatively confident estimate of the Average Treatment Effect (ATE). However, this allocation may lead to inaccuracies in estimating the variance, affecting the accuracy of plug-in estimators for constructing CIs. In such cases, a more conservative estimate may be required to compute the CIs. To address this issue, in the next section, we introduce non-asymptotic bounds that yield provable and more conservative coverage guarantees for specific mechanisms.

4 DISCRETE DP MECHANISMS FOR SECAGG

In this section, we introduce discrete mechanisms that can be combined with secure aggregation for causal inference, which fall into two classes.

1. Additive Noise Mechanisms: These mechanisms involve the addition of discrete noise approximating continuous Gaussian noise. In this approach, each local observable sample X_i is first quantized into a discrete domain and then perturbed by adding appropriate discrete random noise. Candidate noise distributions include Binomial (Agarwal et al., 2018), discrete Gaussian (Canonne et al., 2020; Kairouz et al., 2021), and Skellam (Agarwal et al., 2021).

2. Randomized Response Mechanisms: This class of mechanisms is based on the concept of randomized response introduced by Warner (1965). In these mechanisms, each sample X_i is locally quantized into a binary value, and randomized response is applied multiple times with an appropriate cross-over probability determined by ε. The results of the randomized responses are summed together. Equivalently, this scheme can be viewed as having each client encode its message as a parameter of a Binomial random variable, sending a sample of it to the server. The decoded output follows a Poisson-Binomial distribution, resulting in the Poisson-Binomial mechanism (PBM). Note that since the output space of PBM is finite, it is compatible with secure aggregation, and hence no modular-clipping is required. Therefore, the resulting estimator is unbiased, while all of the additive noise mechanisms inevitably have to introduce small biases.

For brevity, we only present the results of randomized response mechanisms here, and the analysis of additive noise mechanisms is similar.

4.1 Difference-in-mean estimator with the Poisson-Binomial mechanism

Algorithm 2 The Poisson Binomial Mechanism

Input: $c > 0$, $x_i \in [-R, R]$
Parameters: $\theta \in [0, \frac{1}{4}]$, $m \in \mathbb{N}$
Re-scaling x_i: $p_i \triangleq \frac{\theta}{R} x_i + \frac{1}{2}$.
Privatization: $Z_i \triangleq \mathsf{Binom}\,(m, p_i) \in \mathbb{Z}_m$.
Return: Z_i

Next, we describe and analyze our distributed DP scheme based on the Poisson-Binomial mechanism (PBM) (Chen et al., 2022). We make the same assumption that the potential outcome space $\mathcal{Y}$ is a bounded interval and is known ahead of time. Without loss of generality, we let $\mathcal{Y} = [-R, R]$ for some $R > 0$[2]. Per Theorem 3.2, our goal here is to specify the Rényi DP guarantees and the variance of the scheme.

The local randomizer $\mathcal{M}_{\mathsf{PBM}}$ is described in Algorithm 2, which consists of two main steps: 1) first mapping x_i into $\left[\frac{1}{2} - \theta, \frac{1}{2} + \theta\right]$ by $p_i \triangleq \frac{1}{2} + \frac{\theta}{R} x_i$, and then 2) generating a Binomial random variable $Z_i \sim \mathsf{Binom}(m, p_i)$.

Upon securely aggregating $\sum_i Z_i$, the server can obtain an unbiased estimator on $\mu = \sum_i x_i$ as

$$\hat{\mu}\left(\sum_i Z_i\right) \triangleq \frac{R}{nm\theta}\left(\sum_i Z_i - \frac{m}{2}\right) \tag{3}$$

(recall that the server can only learn $\sum_i Z_i$ but not individual Z_i's). In the following theorem, we summarize the privacy and the variance of PBM for a given set of parameters (m, θ).

Theorem 4.1 (Chen et al. (2022)) *Let $\hat{\mu}$ be the estimator from* (3)*. Via Assumption 3.1, for any $\theta \in [0, \frac{1}{4}]$*

- *$\hat{\mu}$ yields an* unbiased *estimate on μ with variance at most $\frac{R^2}{4nm\theta^2}$.*
- *Algorithm 2, together with SecAgg (Bonawitz et al., 2016), satisfies $(\alpha, \varepsilon(\alpha))$-Rényi DP for any $\alpha > 1$ and*

$$\varepsilon(\alpha) \geq C\left(\frac{\theta^2}{(1-2\theta)^4}\right)\frac{\alpha m}{n}, \tag{4}$$

where $C > 0$ is a universal constant.

From this, we can re-write the MSE (i.e., the variance) as $\mathsf{Var}\,(\hat{\mu}) \leq \frac{R^2}{4nm\theta^2} = O\left(\frac{R^2\alpha}{n^2\varepsilon(\alpha)}\right)$.

Since $Z_i \leq m$ and thus $\sum_i Z_i \leq nm$, we set the modulo space $M = nm + 1$ to avoid overflow (recall that M

[2]Here we assume $R > 0$ is known beforehand, which is often the case. When R is unknown, we may need to estimate it through private range/quantile queries.

is the size of the finite group SecAgg operates on). Therefore, the communication cost of Algorithm 2 is $\log M \approx \log n + \log m$ bits per client. In addition, unlike in the additive mechanisms where the noise support is typically unbounded, there is no need to apply modular clipping, and thus $\hat{\mu}$ is unbiased.

A limitation of the PBM approach is that the mechanism was designed for federated learning tasks where local messages are high-dimensional vectors (i.e., model updates) and the number of per-round users is small (usually less than 10^3) (Chen et al., 2022). However, in the design of the experiments, the number of tests can easily exceed millions, and the privacy accounting algorithm in Chen et al. (2022) becomes infeasible. In this work, we develop new efficiently computable bounds on the Rényi DP of PBM that are within 1% (relatively) greater of the actual privacy loss, described in Appendix A.

Next, we construct the mechanisms $\mathcal{M}_1(\cdot, n^*)$ and $\mathcal{M}_2(\cdot, n^*)$ used in Algorithm 1. Let $(m_{1,c}, \theta_{1,c})$, $(m_{1,t}, \theta_{1,t})$ be the parameters of PBM used for estimating the mean of the control and test groups respectively. Similarly, let $(m_{2,c}, \theta_{2,c})$, $(m_{2,t}, \theta_{2,t})$ be the parameters used in estimating the second moments of the two groups. Then according to Theorem 4.1, the privacy losses of $\mathcal{M}_1(\cdot, n_c)$ and $\mathcal{M}_1(\cdot, n_t)$ are $O\left(\frac{\alpha\theta_{1,c}^2 m_{1,c}}{n_c}\right)$ and $O\left(\frac{\alpha\theta_{1,t}^2 m_{1,t}}{n_t}\right)$[3], and the privacy losses of $\mathcal{M}_2(\cdot, n_c)$ and $\mathcal{M}_2(\cdot, n_t)$ are $O\left(\frac{\alpha\theta_{2,c}^2 m_{2,c}}{n_c}\right)$ and $O\left(\frac{\alpha\theta_{2,t}^2 m_{2,t}}{n_t}\right)$. Therefore, combining Theorem 4.1 with Theorem 3.2, we summarize the guarantees of PBM in the following corollary:

Corollary 4.2 *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be implemented with PBM with parameters $(m_{1,c}, \theta_{1,c})$, $(m_{1,t}, \theta_{1,t})$, $(m_{2,c}, \theta_{2,c})$, and $(m_{2,t}, \theta_{2,t})$ respecitvely. Then*

1. *Alg. 1 is $(\alpha, \varepsilon(\alpha))$-Rényi DP for all $\alpha > 1$ and*

$$\varepsilon(\alpha) = O\left(\alpha\left(\frac{\theta_{1,c}^2 m_{1,c}}{n_c} + \frac{\theta_{1,t}^2 m_{1,t}}{n_t} + \frac{\theta_{2,c}^2 m_{2,c}}{n_c} + \frac{\theta_{2,c}^2 m_{2,c}}{n_c}\right)\right).$$

2. *The average width of the $(1-\alpha)$-CI is*

$$O\left(z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{s_c^2}{n_c} + \frac{s_t^2}{n_t} + \frac{c^2}{n_t m_{1,t} \theta_{1,c}^2} + \frac{c^2}{n_t m_{1,t} \theta_{1,t}^2}}\right)$$

for SATE, and for PATE it is

$$O\left(z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\mathsf{Var}(P_c)}{n_c} + \frac{\mathsf{Var}(P_t)}{n_t} + \frac{c^2}{n_t m_{1,t} \theta_{1,c}^2} + \frac{c^2}{n_t m_{1,t} \theta_{1,t}^2}}\right).$$

[3]Note that although here we present an asymptotic form of the privacy losses, in our experiments we can numerically compute the accurate privacy budgets.

Note that in the above expression, the parameters of $\mathcal{M}_2$ do not impact the (asymptotic) width of the confidence intervals (CIs). This is because as long as we can derive a consistent estimator for the sample variances, we can compute CIs accordingly. Therefore, one should allocate the maximum possible privacy budget to $\mathcal{M}_1$. In practice (Section 5), we set the privacy budget for $\mathcal{M}_1$ to be 0.99 of the total privacy allocation.

Parameter selection. In order to satisfy a (ε, δ)-DP, guarantee, we select

$$\frac{\theta_{1,c}^2 m_{1,c}}{n_c} \approx \frac{\theta_{1,t}^2 m_{1,t}}{n_t} = O_\delta\left(\varepsilon^2\right),$$

which means that the average width of the CIs is $O\left(z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{s_c^2}{n_c} + \frac{s_t^2}{n_t} + \frac{c^2}{\varepsilon^2}\left(\frac{1}{n_t^2} + \frac{1}{n_c^2}\right)}\right)$ for SATE, or $O\left(z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\mathsf{Var}(P_c)}{n_c} + \frac{\mathsf{Var}(P_t)}{n_t} + \frac{c^2}{\varepsilon^2}\left(\frac{1}{n_t^2} + \frac{1}{n_c^2}\right)}\right)$ for PATE.

4.2 Non-asymptotic coverage guarantees

In addition to the asymptotic CIs based on the central limit theorem, which are accurate only when n_c and n_t are large, we provide non-asymptotic CIs for estimating SATE and PATE based on variants of empirical Berstein inequalities. For ease of presentation, in the rest of this section, we assume $n_c = n_t = n/2$, but all of the results can be easily adapted to general cases. We first present the non-asymptotic bound for PATE.

Theorem 4.3 (Simplified) *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be PBM (Algorithm 2) with parameter (m_1, θ_1) and (m_2, θ_2). Let $\hat{\sigma}_{\mathsf{p}}^2$ be defined as in Algorithm 1. Then under Assumption 3.1, it holds that*

$$\Pr\left\{\Delta_{\mathsf{p}} \in \hat{\Delta}_{\mathsf{p}} \pm \left(\sqrt{2\hat{\sigma}_{\mathsf{p}}^2 \log(2.01/\delta)} + \gamma\right)\right\} \geq 1 - \delta,$$

where $\gamma = O(1/n)$ when $\varepsilon_1(\alpha) = C_1 \frac{\alpha m_1 \theta_1^2}{n}$ and $\varepsilon_2(\alpha) = C_2 \frac{\alpha m_2 \theta_2^2}{n}$ are constants[4].

The above theorem is proved via the empirical Berstein inequality (Maurer and Pontil, 2009) along with incorporating the tail bounds of the Poisson Binomial mechanism. The same analysis can be applied to other additive mechanisms (such as the Skellam or discrete Gaussian noise), though these mechanisms may not yield an unbiased estimator. The detailed proof can be found in Appendix C.1.

For a given privacy budget $\varepsilon(\alpha)$, Theorem 4.3 suggests a way to allocate privacy budgets (determined by (m_1, θ_1) and (m_2, θ_2)) to minimize the width of CIs (i.e., 2τ). Specifically, if we split the total privacy budget $\varepsilon(\alpha)$

[4]We provide the higher-order terms and constants of γ in Theorem C.1 in Appendix C.1

Table 1: Average widths and coverages of 90%-confidence intervals for PATE.

		$\varepsilon = 0.1$	$\varepsilon = 0.4$	$\varepsilon = 0.7$	$\varepsilon = 1.0$	$\varepsilon = 1.3$	$\varepsilon = 1.6$	$\varepsilon = 1.9$	$\varepsilon = \infty$
None private	Coverage (90% CI)	-	-	-	-	-	-	-	0.902
	Width (90% CI)	-	-	-	-	-	-	-	$2.08{\cdot}10^{-3}$
Central Gaussian	Coverage (90% CI)	0.899	0.901	0.902	0.899	0.897	0.897	0.899	-
	Width (90% CI)	0.771	0.189	0.110	0.078	0.063	0.053	0.044	-
PBM ($m = 256$)	Coverage (90% CI)	0.898	0.897	0.903	0.900	0.902	0.899	0.903	-
	Width (90% CI)	0.772	0.200	0.119	0.085	0.067	0.056	0.048	-
PBM ($m = 1024$)	Coverage (90% CI)	0.903	0.899	0.899	0.899	0.902	0.898	0.900	-
	Width (90% CI)	0.772	0.199	0.118	0.085	0.066	0.055	0.047	-

into $\varepsilon_1(\alpha)$ and $\varepsilon_2(\alpha)$ such that $\mathcal{M}_1$ and $\mathcal{M}_2$ satisfy $\varepsilon_1(\alpha)$ and $\varepsilon_2(\alpha)$ Rényi DP respectively, then we have the following proposition:

Proposition 4.4 *Make the same assumptions as Theorem 4.3. Assume $\mathcal{M}_1$ and $\mathcal{M}_2$ satisfies $(\alpha, \varepsilon_1(\alpha))$ and $(\alpha, \varepsilon_2(\alpha))$ Rényi DP. Let the sample variance $\hat{s}_c^2$ and $\hat{s}_t^2$ be constant and do not scale with n. Then the non-asymptotic CIs in Theorem 4.3 has width*

$$\Theta_\delta\left(\sqrt{\tfrac{\hat{s}_t^2+\hat{s}_c^2}{n}} + \tfrac{R}{n}\left(1+\sqrt{\tfrac{\alpha}{\varepsilon_1(\alpha)}}\right) + \tfrac{R^2\sqrt{\alpha/\varepsilon_2(\alpha)}+R\sqrt{\alpha/\varepsilon_1(\alpha)}}{n^{1.5}\sqrt{\hat{s}_t^2+\hat{s}_c^2}}\right).$$

We provide some insights regarding Proposition 4.4. First, it is important to note that the DP noise only impacts the smaller terms (i.e., $O(1/n)$). The first-order term $\sqrt{(\hat{s}_t^2 + \hat{s}_c^2)/n}$ remains consistent with the asymptotic confidence intervals. Additionally, when considering the allocation of privacy budget $\varepsilon_2(\alpha)$ for estimating sample variance, it exerts a comparatively lesser influence on the confidence intervals in contrast to $\varepsilon_1(\alpha)$ since $\varepsilon_2(\alpha)$ only plays a role in the $O(1/n^{1.5})$ term. This observation supports our intuition that allocating more privacy budget to $\mathcal{M}_1$ is advisable when dealing with sufficiently large values of n.

It is also essential to emphasize that we do not advocate the use of non-asymptotic confidence bounds (as presented in Theorem 4.3) over the asymptotic ones (Theorem 3.5). This is because non-asymptotic bounds may still be overly conservative. Instead, Theorem 4.3 should be utilized as a guideline for allocating privacy budgets when dealing with finite sample sizes.

Finally, the same non-asymptotic CIs hold for SATE.

Theorem 4.5 *Theorem 4.3 holds for SATE by replacing $\hat{\sigma}_{\mathsf{p}}^2$, Δ_{p}, and $\hat{\Delta}_{\mathsf{p}}$ with $\hat{\sigma}_{\mathsf{s}}^2$, Δ_{s}, and $\hat{\Delta}_{\mathsf{s}}$.*

The proof is more involved as it requires a sample-without-replacement version of Bernstein inequality. We leave the details in Appendix C.3.

5 EXPERIMENTS

In this section, we provide empirical evaluations for our proposed framework.

Experiment Setup. We generate the potential outcomes according to truncated Gaussian distributions. Specifically, we set the (population) ATE to be 0.2 and generate $Y_i(c) \overset{\text{i.i.d.}}{\sim} N(-0.1, \sigma_p^2)$ and $Y_i(t) \overset{\text{i.i.d.}}{\sim} N(0.1, \sigma_p^2)$, with $\sigma_p = 0.05$. We truncate both $Y_i(c)$ and $Y_i(t)$ to $[-1, 1]$. We divide the sample size $n = 10^4$ equally into test and control groups (i.e., $n_c = n_t = 5 \cdot 10^3$). We set the confidence level to be 90%, simulate for $N = 10000$ rounds, and compute the empirical coverage ratio, i.e., the number of times that the true PATE lies in the estimated CIs.

Baselines. We compare the proposed distributed DP method, based on PBM (labeled as "PBM") with (1) the non-private difference-in-mean CIs and (2) the Central DP baseline (where we collect all observable samples and add Gaussian noise to the difference-in-mean estimator). For PBM, we compare different output sizes m (recall that m determines the per-user communication cost). We report the average widths of the 90%-CIs, as well as the empirical coverage rates.

From Table 1, we see that the widths of CIs are largely determined by the DP noise and the corresponding privacy levels. However, the CI widths of PBM are very close to the Central Gaussian mechanism, indicating that the price of adopting secure aggregation is small. Due to space limitations, we provide more detailed experimental results in the appendix, including the CIs for SATE and under different data distributions.

Acknowledgements

Wei-Ning Chen and Ayfer Özgür are supported by NSF Award # CCF-2213223. Graham Cormode is supported in part by a UKRI Prosperity Partnership Scheme (FAIR) under the EPSRC Grant EP/V056883/1, and The Alan Turing Institute.

References

Naman Agarwal, Ananda Theertha Suresh, Felix Xinnan X Yu, Sanjiv Kumar, and Brendan McMahan. cpsgd: Communication-efficient and differentially-private distributed sgd. In *Advances in Neural Information Processing Systems*, pages 7564–7575, 2018.

Naman Agarwal, Peter Kairouz, and Ziyu Liu. The skellam mechanism for differentially private federated learning. *Advances in Neural Information Processing Systems*, 34, 2021.

Borja Balle and Yu-Xiang Wang. Improving the gaussian mechanism for differential privacy: Analytical calibration and optimal denoising. In *International Conference on Machine Learning*, pages 394–403. PMLR, 2018.

Rémi Bardenet and Odalric-Ambrym Maillard. Concentration inequalities for sampling without replacement. 2015.

James Henry Bell, Kallista A Bonawitz, Adrià Gascón, Tancrède Lepoint, and Mariana Raykova. Secure single-server aggregation with (poly) logarithmic overhead. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, pages 1253–1269, 2020.

M Ben-Or, S Godwasser, and A Wigderson. Completeness theorems for non-cryptographic fault-tolerant distributed computation (1988). In *Proceedings of the 6th Annual ACM Symposium on Theory of Computing, pp379-386*.

Abhishek Bhowmick, John Duchi, Julien Freudiger, Gaurav Kapoor, and Ryan Rogers. Protection against reconstruction and its applications in private federated learning. *arXiv preprint arXiv:1812.00984*, 2018.

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

Keith Bonawitz, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth. Practical secure aggregation for federated learning on user-held data. *arXiv preprint arXiv:1611.04482*, 2016.

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

Bryan Cai, Constantinos Daskalakis, and Gautam Kamath. Priv'it: Private and sample efficient identity testing. In *International Conference on Machine Learning*, pages 635–644. PMLR, 2017.

Clément L Canonne, Gautam Kamath, and Thomas Steinke. The discrete gaussian for differential privacy. *arXiv preprint arXiv:2004.00010*, 2020.

Nicholas Carlini, Chang Liu, Úlfar Erlingsson, Jernej Kos, and Dawn Song. The secret sharer: Evaluating and testing unintended memorization in neural networks. In *28th {USENIX} Security Symposium ({USENIX} Security 19)*, pages 267–284, 2019.

Wei-Ning Chen, Peter Kairouz, and Ayfer Ozgur. Breaking the communication-privacy-accuracy trilemma. *Advances in Neural Information Processing Systems*, 33, 2020.

Wei-Ning Chen, Ayfer Ozgur, and Peter Kairouz. The poisson binomial mechanism for unbiased federated learning with secure aggregation. In *International Conference on Machine Learning*, pages 3490–3506. PMLR, 2022.

Ivan Damgård, Valerio Pastro, Nigel Smart, and Sarah Zakarias. Multiparty computation from somewhat homomorphic encryption. In *Advances in Cryptology–CRYPTO 2012: 32nd Annual Cryptology Conference, Santa Barbara, CA, USA, August 19-23, 2012. Proceedings*, pages 643–662. Springer, 2012.

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

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

John C Duchi, Michael I Jordan, and Martin J Wainwright. Local privacy and statistical minimax rates. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 429–438. IEEE, 2013.

Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 486–503. Springer, 2006a.

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

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

Alexandre Evfimievski, Ramakrishnan Srikant, Rakesh Agrawal, and Johannes Gehrke. Privacy preserving

mining of association rules. *Information Systems*, 29 (4):343–364, 2004.

Vitaly Feldman and Kunal Talwar. Lossless compression of efficient private local randomizers. *arXiv preprint arXiv:2102.12099*, 2021.

Ronald A Fisher. Studies in crop variation. i. an examination of the yield of dressed grain from broadbalk. *The Journal of Agricultural Science*, 11(2):107–135, 1921.

Ronald A Fisher. On the mathematical foundations of theoretical statistics. *Philosophical transactions of the Royal Society of London.*, 222(594-604):309–368, 1922.

Ronald Aylmer Fisher. Design of experiments. *British Medical Journal*, 1(3923):554, 1936.

Jaroslav Hájek. Some extensions of the wald-wolfowitz-noether theorem. *The Annals of Mathematical Statistics*, pages 506–523, 1961.

Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press, 2015.

Peter Kairouz, H Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Keith Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, et al. Advances and open problems in federated learning. *arXiv preprint arXiv:1912.04977*, 2019.

Peter Kairouz, Ziyu Liu, and Thomas Steinke. The distributed discrete gaussian mechanism for federated learning with secure aggregation. *arXiv preprint arXiv:2102.06387*, 2021.

Vishesh Karwa and Salil Vadhan. Finite sample differentially private confidence intervals. *arXiv preprint arXiv:1711.03908*, 2017.

Shiva Prasad Kasiviswanathan, Homin K Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. What can we learn privately? *SIAM Journal on Computing*, 40(3):793–826, 2011.

Si Kai Lee, Luigi Gresele, Mijung Park, and Krikamol Muandet. Privacy-preserving causal inference via inverse probability weighting. *arXiv preprint arXiv:1905.12592*, 2019.

Xinran Li and Peng Ding. General forms of finite population central limit theorems with applications to causal inference. *Journal of the American Statistical Association*, 112(520):1759–1769, 2017.

Xinran Li, Peng Ding, and Donald B Rubin. Asymptotic theory of rerandomization in treatment–control experiments. *Proceedings of the National Academy of Sciences*, 115(37):9157–9162, 2018.

Andreas Maurer and Massimiliano Pontil. Empirical bernstein bounds and sample variance penalization. *arXiv preprint arXiv:0907.3740*, 2009.

Luca Melis, Congzheng Song, Emiliano De Cristofaro, and Vitaly Shmatikov. Exploiting unintended feature leakage in collaborative learning. In *2019 IEEE Symposium on Security and Privacy (SP)*, pages 691–706. IEEE, 2019.

Ilya Mironov. Rényi differential privacy. In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, pages 263–275. IEEE, 2017.

Fengshi Niu, Harsha Nori, Brian Quistorff, Rich Caruana, Donald Ngwe, and Aadharsh Kannan. Differentially private estimation of heterogeneous causal effects. In *Conference on Causal Learning and Reasoning*, pages 618–633. PMLR, 2022.

Yuki Ohnishi and Jordan Awan. Locally private causal inference. *arXiv preprint arXiv:2301.01616*, 2023.

Anant Raj, Ho Chung Leon Law, Dino Sejdinovic, and Mijung Park. A differentially private kernel two-sample test. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2019, Würzburg, Germany, September 16–20, 2019, Proceedings, Part I*, pages 697–724. Springer, 2020.

Ryan Rogers and Daniel Kifer. A new class of private chi-square hypothesis tests. In *Artificial Intelligence and Statistics*, pages 991–1000. PMLR, 2017.

Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. Membership inference attacks against machine learning models. In *2017 IEEE Symposium on Security and Privacy (SP)*, pages 3–18. IEEE, 2017.

Congzheng Song and Vitaly Shmatikov. Auditing data provenance in text-generation models. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 196–206, 2019.

Stanley L Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309): 63–69, 1965.

Ian Waudby-Smith, Steven Wu, and Aaditya Ramdas. Nonparametric extensions of randomized response for private confidence sets. In *International Conference on Machine Learning*, pages 36748–36789. PMLR, 2023.

A Practical privacy accounting for PBM

In this section, we improve the efficiency of the privacy accounting mechanism Chen et al. (2022), which are originally designed for small sample and finite field sizes (usually when $n, m \leq 10^3$) due to the batch-SGD and the natural computation and communication constraints of using secure aggregation.

Following from the proof of Theorem 3.3 in Chen et al. (2022), for any set of parameters (m, n, θ, α), $\varepsilon(\alpha)$ can be expressed as

$$\max_{t_1,t_2\in[m\cdot n],\,|t_1-t_2|\leq m} D_\alpha\Big(P_{\mathsf{Binom}\left(t_1,\frac{1}{2}-\theta\right)+\mathsf{Binom}\left(mn-t_1,\frac{1}{2}+\theta\right)}\Big\|P_{\mathsf{Binom}\left(t_2,\frac{1}{2}-\theta\right)+\mathsf{Binom}\left(mn-t_2,\frac{1}{2}+\theta\right)}\Big). \tag{5}$$

In Chen et al. (2022), it is shown that the maximum of (5) occurs at $(t_1, t_2) = (0, m)$, which suggests the following (exact) privacy accounting mechanism in Algorithm 3.

Algorithm 3 Exact privacy accounting.

Input: n, m, θ, α
Return: $\varepsilon(\alpha)$
$P_1 \leftarrow \mathsf{Binom}(mn, \frac{1}{2} - \theta)$ ▷ P_1 is a $mn + 1$-dim vector.
$P_2 \leftarrow \mathsf{Binom}(m(n-1), \frac{1}{2} - \theta)$
$P_2' \leftarrow \mathsf{Binom}(m, \frac{1}{2} + \theta)$
$P_2 \leftarrow P_2 * P_2'$ ▷ $*$ denotes the convolution operator.
$\varepsilon(\alpha) \leftarrow \frac{1}{\alpha-1}\log\left(\mathsf{sum}\left(\frac{P_1^\alpha}{P_2^{\alpha-1}}\right)\right)$ ▷ sum and $(\cdot)^\alpha$ are performed coordinate-wisely.

Note that the accounting involves binomial coefficients with large n, so in practice, all computations should be done in the log space to ensure computation stability, as described in Algorithm 4. The computation bottlenecks of Algorithm 3 and Algorithm 4 are at the convolution operation, which, when computed via fast Fourier transform, takes $\tilde{O}(mn)$ time.

Algorithm 4 Exact privacy accounting over the log space.

Input: n, m, θ, α
Return: $\varepsilon(\alpha)$
$\mathsf{logP_1} \leftarrow \log\left(\mathsf{Binom}(mn, \frac{1}{2} - \theta)\right)$
$\mathsf{logP_2} \leftarrow \log\left(\mathsf{Binom}(m(n-1), \frac{1}{2} - \theta)\right)$
$\mathsf{logP_2'} \leftarrow \mathsf{Binom}(m, \frac{1}{2} + \theta)$
$\mathsf{logP_2} \leftarrow \mathsf{logP_2} \,\tilde{*}\, \mathsf{logP_2'}$ ▷ $\tilde{*}$ denotes the convolution operator *over the log space*.
$\varepsilon(\alpha) \leftarrow \frac{1}{\alpha-1}\mathsf{logexpsum}\left(\alpha \cdot \mathsf{logP_1} + (1-\alpha)\cdot \mathsf{logP_2}\right)$

A.1 Approximation for large n and m

Unfortunately, in most private analytic or causal inference tasks, the number of samples n can be up to millions (and m may be up to thousands), making the $\tilde{O}(mn)$ time complexity of the above algorithms infeasible. To address this issue, we propose to account for the privacy loss via the following upper bound based on a data processing inequality:

$$(5) \leq \max_{k\in[n-1]} m \cdot D_\alpha\Big(P_{\mathsf{Binom}\left(1+k,\frac{1}{2}-\theta\right)+\mathsf{Binom}\left(n-k-1,\frac{1}{2}+\theta\right)}\Big\|P_{\mathsf{Binom}\left(k,\frac{1}{2}-\theta\right)+\mathsf{Binom}\left(n-k,\frac{1}{2}+\theta\right)}\Big). \tag{6}$$

Although (6) is always strictly greater than the exact privacy loss (5), when either m or n is large, the approximation error in $\varepsilon(\alpha)$ is negligible. For instance, when $n = 100$ and $\alpha = 2$, the approximation error is less than 0.1%. By leveraging (6), we arrive at the following approximate privacy accounting algorithm, which reduces the computational complexity to $O(n)$:

Algorithm 5 Efficient approximate privacy.

Input: n, m, θ, α
$\mathsf{logP}_1 \leftarrow \log\left(\mathsf{Binom}(n, \frac{1}{2} - \theta)\right)$
$\mathsf{logP}_2 \leftarrow \log\left(\mathsf{Binom}(n-1, \frac{1}{2} - \theta)\right)$
$\mathsf{logP}'_2 \leftarrow \mathsf{Ber}(\frac{1}{2} + \theta)$
$\mathsf{logP}_2 \leftarrow \mathsf{logP}_2 \tilde{*} \mathsf{logP}'_2$ ▷ $\tilde{*}$ denotes the convolution operator *over the log space*.
$\varepsilon(\alpha) \leftarrow \frac{1}{\alpha-1}\mathsf{logexpsum}\left(\alpha \cdot \mathsf{logP}_1 + (1-\alpha) \cdot \mathsf{logP}_2\right)$.
Return: $m\varepsilon(\alpha)$

In our experiments, we account the Rényi DP according to Algorithm 5 and convert the $(\alpha, \varepsilon(\alpha))$-Rényi DP to (ε, δ)-DP via the conversion lemma given by Canonne et al. (2020).

B Additional experiments

In this section, we provide more complete experimental results to demonstrate the utility of our proposed framework.

B.1 Gaussian potential outcomes

In the first set of examples, we consider random treatment effects, where the potential outcomes before and after the treatment are normally distributed: $Y_i(0) \overset{\text{i.i.d.}}{\sim} N(\mu_0, \sigma)$ and $Y_i(1) \overset{\text{i.i.d.}}{\sim} N(\mu_1, \sigma)$. Under this distributional assumption, the PATE is defined as $\Delta_{\mathsf{p}} \triangleq \mu_1 - \mu_0$, while the SATE is $\Delta_{\mathsf{s}} \triangleq \frac{1}{n_t}\sum_i Y_i(1) - \frac{1}{n_t}\sum_i Y_i(0)$, where n_c and n_t represent the numbers of the control and test groups.

In the experiments, we set $n_c = n_t = 10^3$, $\Delta_{\mathsf{p}} = 0.2$, and the noise level $\sigma = 0.01$. For each set of parameters of the privatization mechanisms, we set the confidence level to be 90%, simulate for $N = 10000$ rounds, and report the average widths of CIs and the empirical coverage ratios (i.e., the number of times that the true PATE lies within the estimated CIs).

In Table 2, we observe that without privacy constraints, we obtain tight CIs with a significantly higher coverage ratio than required. Specifically, we achieve a coverage ratio of 0.98 compared to the requested 0.9 coverage ratio under a 90% confidence constraint[5]. The issue of being overly conservative, however, vanishes under DP, since the DP noise dominates the total uncertainty and is much larger than the sampling variance.

Comparing the non-private setting, we found that the width of the private CIs is significantly larger than the non-private one, indicating that the DP noise is much larger than the sampling noise. Unfortunately, this is the price we need to pay. However, the CI widths of the centralized Gaussian mechanism are roughly the same as the width of PBM. The difference to the Gaussian mechanism is negligible when n and m are large enough. In Table 2, we can see that when $n = 1000$, setting $m = 256$ is sufficient to achieve the same performance as the centralized Gaussian mechanism. This implies that although the price for achieving DP is indispensable, the price for adopting secure aggregation to remove the trust toward the server can be made arbitrary small, as long as we are willing to slightly increase the communication costs (which are dictated by the finite field size m).

We can observe a similar trend when estimating the population level treatment effect (i.e., PATE). We see that when setting $m = 256$, the width of CIs is almost the same as the the centralized Gaussian. A major difference compared to estimating SATE, however, is that the average converge ratio of the non-private setting becomes aligned with our target confidence level (i.e., 90% in our setting). This is because the variance estimator of PATE given in Algorithm 1 becomes unbiased since the unidentifiable term (i.e., the covariance) is cancelled out (see the proof given in Section **??** for more details).

[5]Note that when estimating the confidence intervals of the difference-in-mean estimator for SATE, the true variance is unidentifiable. Therefore, we can only use an upper bound to obtain a conservative interval, as discussed in the proof of Theorem 3.2.

Table 2: Average width and coverage of 90%-confidence intervals for SATE. Gaussian potential outcomes with $n = 10^3$.

Non-private	0.980 0.002 ± 3.25e-05						
	$\varepsilon = 0.1$	$\varepsilon = 0.4$	$\varepsilon = 0.7$	$\varepsilon = 1.0$	$\varepsilon = 1.3$	$\varepsilon = 1.6$	$\varepsilon = 1.9$
Central Gaussian	0.899 0.771 ± 1.26e-07	0.897 0.199 ± 4.87e-07	0.899 0.118 ± 8.40e-07	0.900 0.084 ± 1.15e-06	0.901 0.066 ± 1.45e-06	0.897 0.055 ± 1.78e-06	0.899 0.047 ± 2.08e-06
PBM (m=256)	0.899 0.772 ± 1.26e-07	0.903 0.200 ± 4.85e-07	0.904 0.119 ± 8.34e-07	0.905 0.085 ± 1.13e-06	0.898 0.067 ± 1.42e-06	0.896 0.056 ± 1.73e-06	0.896 0.048 ± 2.00e-06
PBM (m=1024)	0.904 0.772 ± 1.26e-07	0.892 0.199 ± 4.83e-07	0.896 0.118 ± 8.23e-07	0.901 0.085 ± 1.15e-06	0.901 0.066 ± 1.47e-06	0.904 0.055 ± 1.76e-06	0.898 0.047 ± 2.07e-06
PBM (m=2048)	0.896 0.772 ± 1.27e-07	0.902 0.199 ± 4.81e-07	0.899 0.118 ± 8.16e-07	0.903 0.084 ± 1.15e-06	0.897 0.066 ± 1.45e-06	0.904 0.055 ± 1.77e-06	0.896 0.047 ± 2.08e-06

Table 3: Average width and coverage of 90%-confidence intervals for PATE. Gaussian potential outcomes with $n = 10^3$.

Non-private	0.901 0.002 ± 3.24e-05						
	$\varepsilon = 0.1$	$\varepsilon = 0.4$	$\varepsilon = 0.7$	$\varepsilon = 1.0$	$\varepsilon = 1.3$	$\varepsilon = 1.6$	$\varepsilon = 1.9$
Central Gaussian	0.905 0.771 ± 1.24e-07	0.895 0.199 ± 4.85e-07	0.899 0.118 ± 8.20e-07	0.902 0.084 ± 1.16e-06	0.904 0.066 ± 1.47e-06	0.899 0.055 ± 1.78e-06	0.899 0.047 ± 2.07e-06
PBM (m=256)	0.902 0.772 ± 1.25e-07	0.900 0.200 ± 4.84e-07	0.900 0.119 ± 8.15e-07	0.903 0.085 ± 1.15e-06	0.906 0.067 ± 1.43e-06	0.900 0.056 ± 1.72e-06	0.903 0.048 ± 2.02e-06
PBM (m=1024)	0.900 0.772 ± 1.26e-07	0.897 0.199 ± 4.85e-07	0.902 0.118 ± 8.28e-07	0.900 0.085 ± 1.17e-06	0.904 0.066 ± 1.46e-06	0.898 0.055 ± 1.77e-06	0.896 0.047 ± 2.05e-06
PBM (m=2048)	0.897 0.772 ± 1.24e-07	0.902 0.199 ± 4.85e-07	0.901 0.118 ± 8.19e-07	0.901 0.084 ± 1.16e-06	0.899 0.066 ± 1.47e-06	0.902 0.055 ± 1.77e-06	0.898 0.047 ± 2.06e-06

B.2 Constant treatment effects

In the second set of examples, we consider constant treatment effects. Specifically, we assume $Y_i(0) \overset{\text{i.i.d.}}{\sim} \mathsf{uniform}(a, b)$ and $Y_i(1) = Y_i(0) + \Delta_{\mathsf{s}}$, where Δ_{s} is a deterministic but unknown quantity that we want to estimate.

In the experiments, we set $n_c = n_t = 10^3$, $\Delta_{\mathsf{s}} = 0.2$, and $(a, b) = (-1, -0.8)$. For each set of parameters of the privatization mechanisms, we again set the confidence level to be 90%, simulate for $N = 10000$ rounds, and report the average widths of CIs and the empirical coverage ratios.

As shown in Table 4 and Table 5, under the assumption of a constant ATE, estimating SATE and PATE is essentially the same, both theoretically and empirically. The coverage ratios for both PATE and SATE are accurate, in contrast to SATE with random ATE. Furthermore, we observe a similar trend as in the Gaussian outcomes, where PBM achieves a negligible error compared to the central Gaussian.

Table 4: Average width and coverage of 90%-confidence intervals for SATE. Constant treatment effect with $n = 10^3$.

Non-private	0.897 0.108 ± 1.53e-03						
	$\varepsilon = 0.1$	$\varepsilon = 0.4$	$\varepsilon = 0.7$	$\varepsilon = 1.0$	$\varepsilon = 1.3$	$\varepsilon = 1.6$	$\varepsilon = 1.9$
Central Gaussian	0.904 0.779 ± 2.11e-04	0.902 0.227 ± 7.30e-04	0.901 0.160 ± 1.03e-03	0.899 0.137 ± 1.21e-03	0.895 0.127 ± 1.30e-03	0.899 0.121 ± 1.40e-03	0.896 0.118 ± 1.41e-03
PBM (m=256)	0.893 0.779 ± 2.12e-04	0.904 0.227 ± 7.40e-04	0.904 0.160 ± 1.03e-03	0.900 0.138 ± 1.20e-03	0.897 0.127 ± 1.31e-03	0.898 0.122 ± 1.36e-03	0.897 0.118 ± 1.40e-03
PBM (m=1024)	0.896 0.779 ± 2.13e-04	0.900 0.227 ± 7.36e-04	0.905 0.160 ± 1.03e-03	0.901 0.137 ± 1.19e-03	0.900 0.127 ± 1.29e-03	0.904 0.121 ± 1.36e-03	0.895 0.118 ± 1.40e-03
PBM (m=2048)	0.898 0.779 ± 2.11e-04	0.897 0.227 ± 7.30e-04	0.902 0.160 ± 1.03e-03	0.901 0.137 ± 1.20e-03	0.903 0.127 ± 1.30e-03	0.900 0.121 ± 1.40e-03	0.899 0.118 ± 1.41e-03

Table 5: Average width and coverage of 90%-confidence intervals for PATE. Constant treatment effect with $n = 10^3$.

Non-private	0.901 0.002 ± 3.24e-05						
	$\varepsilon = 0.1$	$\varepsilon = 0.4$	$\varepsilon = 0.7$	$\varepsilon = 1.0$	$\varepsilon = 1.3$	$\varepsilon = 1.6$	$\varepsilon = 1.9$
Central Gaussian	0.905 0.771 ± 1.24e-07	0.895 0.199 ± 4.85e-07	0.899 0.118 ± 8.20e-07	0.902 0.084 ± 1.16e-06	0.904 0.066 ± 1.47e-06	0.899 0.055 ± 1.78e-06	0.899 0.047 ± 2.07e-06
PBM (m=256)	0.902 0.772 ± 1.25e-07	0.900 0.200 ± 4.84e-07	0.900 0.119 ± 8.15e-07	0.903 0.085 ± 1.15e-06	0.906 0.067 ± 1.43e-06	0.900 0.056 ± 1.72e-06	0.903 0.048 ± 2.02e-06
PBM (m=1024)	0.900 0.772 ± 1.26e-07	0.897 0.199 ± 4.85e-07	0.902 0.118 ± 8.28e-07	0.900 0.085 ± 1.17e-06	0.904 0.066 ± 1.46e-06	0.898 0.055 ± 1.77e-06	0.896 0.047 ± 2.05e-06
PBM (m=2048)	0.897 0.772 ± 1.24e-07	0.902 0.199 ± 4.85e-07	0.901 0.118 ± 8.19e-07	0.901 0.084 ± 1.16e-06	0.899 0.066 ± 1.47e-06	0.902 0.055 ± 1.77e-06	0.898 0.047 ± 2.06e-06

B.3 Constant treatment effect with larger n

Finally, in the last set of experiments, we consider a larger sample size with Gaussian outcomes. We use the same set of parameters as in Section B.1, except that $n_t = n_c = 10^4$. From Table 4 and Table 5, we observe that when the privacy budget is large enough $\varepsilon > 1$, the CIs for both PBM and central Gaussian are very closed to the non-private one, indicating that the error is dominated by the sampling noise instead of the DP noise. Therefore, when n is large enough (depending on the sample variance), we can achieve DP with negligible effect on the utility.

Table 6: Average width and coverage of 90%-confidence intervals for SATE. Constant treatment effect with $n = 10^4$.

Non-private	0.896 0.034 ± 1.51e-04						
	$\varepsilon = 0.1$	$\varepsilon = 0.4$	$\varepsilon = 0.7$	$\varepsilon = 1.0$	$\varepsilon = 1.3$	$\varepsilon = 1.6$	$\varepsilon = 1.9$
Central Gaussian	0.905 0.084 ± 6.25e-05	0.903 0.040 ± 1.32e-04	0.896 0.036 ± 1.45e-04	0.903 0.035 ± 1.49e-04	0.899 0.035 ± 1.49e-04	0.899 0.035 ± 1.52e-04	0.903 0.035 ± 1.50e-04
PBM (m=256)	0.905 0.084 ± 6.15e-05	0.906 0.040 ± 1.32e-04	0.897 0.036 ± 1.42e-04	0.902 0.036 ± 1.48e-04	0.897 0.036 ± 1.47e-04	0.896 0.036 ± 1.46e-04	0.905 0.036 ± 1.47e-04
PBM (m=1024)	0.899 0.085 ± 6.16e-05	0.897 0.040 ± 1.32e-04	0.902 0.036 ± 1.45e-04	0.902 0.035 ± 1.48e-04	0.898 0.035 ± 1.49e-04	0.903 0.035 ± 1.51e-04	0.900 0.035 ± 1.52e-04
PBM (m=2048)	0.903 0.085 ± 6.22e-05	0.903 0.040 ± 1.31e-04	0.899 0.036 ± 1.45e-04	0.898 0.035 ± 1.49e-04	0.906 0.035 ± 1.49e-04	0.898 0.035 ± 1.51e-04	0.901 0.035 ± 1.50e-04

Table 7: Average width and coverage of 90%-confidence intervals for PATE. Constant treatment effect with $n = 10^4$.

Non-private	0.904 0.034 ± 1.53e-04						
	$\varepsilon = 0.1$	$\varepsilon = 0.4$	$\varepsilon = 0.7$	$\varepsilon = 1.0$	$\varepsilon = 1.3$	$\varepsilon = 1.6$	$\varepsilon = 1.9$
Central Gaussian	0.904 0.084 ± 6.23e-05	0.899 0.040 ± 1.33e-04	0.903 0.036 ± 1.42e-04	0.907 0.035 ± 1.50e-04	0.900 0.035 ± 1.51e-04	0.900 0.035 ± 1.51e-04	0.900 0.035 ± 1.51e-04
PBM (m=256)	0.903 0.084 ± 6.17e-05	0.897 0.040 ± 1.30e-04	0.907 0.036 ± 1.43e-04	0.911 0.036 ± 1.49e-04	0.901 0.036 ± 1.46e-04	0.900 0.036 ± 1.45e-04	0.899 0.036 ± 1.47e-04
PBM (m=1024)	0.899 0.085 ± 6.15e-05	0.898 0.040 ± 1.33e-04	0.905 0.036 ± 1.46e-04	0.903 0.035 ± 1.48e-04	0.904 0.035 ± 1.50e-04	0.901 0.035 ± 1.52e-04	0.896 0.035 ± 1.50e-04
PBM (m=2048)	0.905 0.085 ± 6.21e-05	0.900 0.040 ± 1.32e-04	0.901 0.036 ± 1.42e-04	0.903 0.035 ± 1.50e-04	0.903 0.035 ± 1.51e-04	0.896 0.035 ± 1.50e-04	0.901 0.035 ± 1.51e-04

C Omitted Proofs

C.1 Proof of Theorem 4.3

We first present the full version of the theorem with higher-order terms and constants.

Theorem C.1 (Detailed version of Theorem 4.3) *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be PBM (Algorithm 2) with parameter (m_1, θ_1) and (m_2, θ_2). Let $\hat{\sigma}_{\mathsf{p}}$ be defined as in Algorithm 1. Then under Assumption 3.1, it holds that*

$$\Pr\left\{\Delta_{\mathsf{p}} \in \left[\hat{\Delta}_{\mathsf{p}} - \left(\sqrt{2\hat{\sigma}_{\mathsf{p}}^2 \log(2.01/\delta)} + \gamma\right), \hat{\Delta}_{\mathsf{p}} + \left(\sqrt{2\hat{\sigma}_{\mathsf{p}}^2 \log(2.01/\delta)} + \gamma\right)\right]\right\} \geq 1 - \delta,$$

where

$$\gamma = \frac{56R\log(1200/\delta)}{3(n-1)} + \sqrt{\frac{R^2}{2m_1 n\theta_1^2}\log\left(\frac{1200}{\delta}\right)} + \sqrt{\frac{4\log(2.01/\delta_1)}{n}}$$
$$\cdot \min\left(\sqrt[4]{\log\left(\frac{1200}{\delta}\right)\frac{R^4}{4m_2 n\theta_2^2}} + \sqrt[4]{\log\left(\frac{1200}{\delta}\right)\frac{R^2}{2m_1 n\theta_1^2}}, \quad \frac{\sqrt{\log\left(\frac{1200}{\delta}\right)\frac{R^4}{4m_2 n\theta_2^2}} + \sqrt{\log\left(\frac{1200}{\delta}\right)\frac{R^2}{2m_1 n\theta_1^2}}}{\sqrt{\hat{s}_t^2 + \hat{s}_c^2}}\right).$$

Note that when $\varepsilon_1(\alpha) = C_1\frac{\alpha m_1\theta_1^2}{n}$ and $\varepsilon_2(\alpha) = C_2\frac{\alpha m_2\theta_2^2}{n}$ are constants, $\gamma = O(1/n)$.

Proof. Before entering the main proof, we will make use of the following (slightly adapted) empirical Berstein inequality:

Lemma C.2 (Theorem 11, Maurer and Pontil (2009)) *Let $\mathbf{X} = (X_1, ..., X_n)$ be a vector of independent random variables with values in $[-R, R]$. Let $\delta > 0$. Then for any $\delta_1, \delta_2 > 0$ and $\delta_1 + \delta_2 = \delta$, it implies*

$$\Pr\left\{\left|\frac{1}{n}\sum_i X_i - \frac{1}{n}\sum_i \mathbb{E}[X_i]\right| \leq \sqrt{\frac{2s^2(\mathbf{X})\log\left(\frac{2}{\delta_1}\right)}{n}} + \frac{14R\log\left(\frac{2}{\delta_2}\right)}{3(n-1)}\right\} \geq 1 - \delta, \tag{7}$$

where $s^2(\mathbf{X}) \triangleq \frac{1}{n(n-1)}\sum_{i\neq j}(X_i - X_j)^2$ denotes the sample variance.

Now, we apply the above lemma in our PATE estimation task. Under the PATE setting with $n_c = n_t = n/2$, it holds that $X_1, ..., X_{n/2} \overset{\text{i.i.d.}}{\sim} P_t$ and $X_{n/2+1}, ..., X_n \overset{\text{i.i.d.}}{\sim} P_c$. As a result, Lemma C.2 yields that, with probability $1 - \delta$,

$$\left|\frac{2}{n}\sum_{i\in[n/2]}\underbrace{(X_i - X_{i+n/2})}_{\triangleq W_i} - \Delta_{\mathsf{p}}\right| \leq \sqrt{\frac{4s^2(\boldsymbol{W})\log(1/\delta_1)}{n}} + \frac{56R\log(2/\delta_2)}{3(n-1)}$$
$$= \sqrt{\frac{4(s_t^2 + s_c^2)\log(1/\delta_1)}{n}} + \frac{56R\log(2/\delta_2)}{3(n-1)}, \tag{8}$$

where the first inequality holds since $W_i \in [-2R, 2R]$ and the second equality holds since

$$s^2(\boldsymbol{W}) = s^2\left(X_1, ..., X_{n/2}\right) + s^2\left(X_{n/2+1}, ..., X_n\right) = s_t^2 + s_c^2$$

by definition.

Next, it suffices to combine with the concentration bounds on $\hat{\Delta}_{\mathsf{p}} \triangleq \hat{\mu}_t - \hat{\mu}_c$ and $\hat{s}_t^2$ and $\hat{s}_c^2$ (recall that these are the private estimates of sample means and variance from PBM).

Concentration of private sample mean. To this end, observe that

$$\hat{\mu}_t - \hat{\mu}_c = \frac{2R}{nm_1\theta}\sum_{i=1}^{n}\left(Z_i - \frac{m_1}{2}\right),$$

where $Z_i \sim \mathsf{Binom}\left(m_1, \frac{1}{2} - \frac{\theta_1 X_{i+n/2}}{R}\right)$ for $i \in [n/2]$ and $Z_i \sim \mathsf{Binom}\left(m_1, \frac{1}{2} + \frac{\theta_1 X_{i+n/2}}{R}\right)$ for $i \in [n/2+1 : n]$. Conditioning on X_i's and applying Hoeffding's inequality yield

$$\Pr\left\{\left|(\hat{\mu}_t - \hat{\mu}_c) - \frac{2}{n}\sum_{i=1}^{n/2}(X_i - X_{i+n/2})\right| \geq \sqrt{\log\left(\frac{2}{\delta_\mu}\right)\frac{R^2}{2m_1 n\theta_1^2}}\right\}$$

$$= \Pr\left\{ \left| \underbrace{\sum_{i=1}^{n} Z_i - \sum_{i=1}^{n/2} \left(\frac{m_1\theta_1}{R}\left(X_i - X_{i+n/2}\right) - \frac{m_1}{2}\right)}_{\text{sum of } m_1 n/2 \text{ independent zero-mean bounded variables.}} \right| \geq \sqrt{\log\left(\frac{2}{\delta_\mu}\right)\frac{m_1 n}{2}} \right\}$$
$$\leq \delta_\mu. \tag{9}$$

Concentration of private sample variance. Next, we construct the estimator of the sample variance from PBM:

$$\hat{s}_t^2 = \frac{1}{n-1}\sum_{i=1}^{n/2}\left(\frac{R^2}{2m_2\theta_2}Z_i' - \frac{R^2}{4\theta_2} + \frac{R^2}{2}\right) - \frac{n}{n-1}\hat{\mu}_t^2,$$

where $Z_i' \sim \mathsf{Binom}\left(m_2, 2\theta\left(\frac{X_i^2}{R^2} - \frac{1}{2}\right) + \frac{1}{2}\right)$. We construct $\hat{s}_c$ in the same way. Notice that the above $\hat{s}_t^2$ is constructed such that $\mathbb{E}\left[\hat{s}_t^2 \middle| \boldsymbol{X}\right] = s_t^2$.

Since the sample variance estimator $\hat{s}_t^2$ and $\hat{s}_c^2$ are privatized by PBM, it is possible to obtain negative values, so we will replace them by its positive part, i.e., $\hat{s}_t^{2+} \triangleq \max\left(\hat{s}_t^2, 0\right)$ and $\hat{s}_c^{2+} \triangleq \max\left(\hat{s}_c^2, 0\right)$. For notational convenience, we abuse notation and let $\hat{s}_t^2$ and $\hat{s}_c^2$ be the positive parts so that $\hat{s}_t^2, \hat{s}_c^2 \geq 0$ always holds.

Since $\hat{s}_t^2$ is obtained by first estimating the second moment of samples $\sum_i X_i^2$ and then subtract the sample mean $n\hat{\mu}_t^2$, it holds that, conditioning on X_i's and the event

$$\left\{ \hat{\Delta}_{\mathsf{p}} - \frac{2}{n}\sum_{i=1}^{n/2}\left(X_i - X_{i+n/2}\right) < \sqrt{\log\left(\frac{2}{\delta_\mu}\right)\frac{R^2}{2m_1 n\theta_1^2}} \right\},$$

we have

$$\Pr\left\{ \left|\left(\hat{s}_t^2 + \hat{s}_c^2\right) - \left(s_t^2 + s_c^2\right)\right| \geq \sqrt{\log\left(\frac{2}{\delta_s}\right)\frac{R^4}{4m_2 n\theta_2^2}} + \sqrt{\log\left(\frac{2}{\delta_\mu}\right)\frac{R^2}{2m_1 n\theta_1^2}} \right\} \leq \delta_s. \tag{10}$$

This implies that with probability at least $1-\delta_s$, both of the following events hold:

- $$\sqrt{\hat{s}_t^2 + \hat{s}_c^2} \leq \sqrt{\left(s_t^2 + s_c^2\right) + \sqrt{\log\left(\frac{2}{\delta_s}\right)\frac{R^4}{4m_2 n\theta_2^2}} + \sqrt{\log\left(\frac{2}{\delta_\mu}\right)\frac{R^2}{2m_1 n\theta_1^2}}}$$
$$\leq \sqrt{s_t^2 + s_c^2} + \sqrt[4]{\log\left(\frac{2}{\delta_s}\right)\frac{R^4}{4m_2 n\theta_2^2}} + \sqrt[4]{\log\left(\frac{2}{\delta_\mu}\right)\frac{R^2}{2m_1 n\theta_1^2}}.$$

- $$\sqrt{\hat{s}_t^2 + \hat{s}_c^2} \leq \sqrt{s_t^2 + s_c^2} + \frac{\sqrt{\log\left(\frac{2}{\delta_s}\right)\frac{R^4}{4m_2 n\theta_2^2}} + \sqrt{\log\left(\frac{2}{\delta_\mu}\right)\frac{R^2}{2m_1 n\theta_1^2}}}{\sqrt{\hat{s}_t^2 + \hat{s}_c^2} + \sqrt{s_t^2 + s_c^2}}$$
$$\leq \sqrt{s_t^2 + s_c^2} + \frac{\sqrt{\log\left(\frac{2}{\delta_s}\right)\frac{R^4}{4m_2 n\theta_2^2}} + \sqrt{\log\left(\frac{2}{\delta_\mu}\right)\frac{R^2}{2m_1 n\theta_1^2}}}{\sqrt{\hat{s}_t^2 + \hat{s}_c^2}}.$$

Therefore, we arrive at the following bound on the private sample variance:

$$\Pr\left(\sqrt{\hat{s}_t^2 + \hat{s}_c^2} \geq \sqrt{s_t^2 + s_c^2} + \min\left(\sqrt[4]{\log\left(\frac{2}{\delta_s}\right)\frac{R^4}{4m_2 n\theta_2^2}} + \sqrt[4]{\log\left(\frac{2}{\delta_\mu}\right)\frac{R^2}{2m_1 n\theta_1^2}},\right.\right.$$

$$\frac{\sqrt{\log\left(\frac{2}{\delta_s}\right)\frac{R^4}{4m_2n\theta_2^2}}+\sqrt{\log\left(\frac{2}{\delta_\mu}\right)\frac{R^2}{2m_1n\theta_1^2}}}{\sqrt{\hat{s}_t^2+\hat{s}_c^2}}\Bigg)\Bigg) \le \delta_s. \tag{11}$$

Putting things together. Finally, by plugging (9) and (11) into (8), we obtain that, with probability at least $1-\delta_1-\delta_2-\delta_\mu-\delta_s$,

$$\left|\hat{\Delta}_{\mathsf{p}}-\Delta_{\mathsf{p}}\right| \le \sqrt{\frac{4\left(\hat{s}_t^2+\hat{s}_c^2\right)\log(2/\delta_1)}{n}}+\gamma = \sqrt{2\hat{\sigma}_{\mathsf{p}}^2\log(2/\delta_1)}+\gamma, \tag{12}$$

where $\gamma = o(1/\sqrt{n})$ and takes the following explicit expression:

$$\gamma = \frac{56R\log(2/\delta_2)}{3(n-1)}+\sqrt{\frac{R^2}{2m_1n\theta_1^2}\log\left(\frac{2}{\delta_\mu}\right)}+\sqrt{\frac{4\log(2/\delta_1)}{n}}\cdot$$
$$\min\left(\sqrt[4]{\log\left(\frac{2}{\delta_s}\right)\frac{R^4}{4m_2n\theta_2^2}}+\sqrt[4]{\log\left(\frac{2}{\delta_\mu}\right)\frac{R^2}{2m_1n\theta_1^2}}, \frac{\sqrt{\log\left(\frac{2}{\delta_s}\right)\frac{R^4}{4m_2n\theta_2^2}}+\sqrt{\log\left(\frac{2}{\delta_\mu}\right)\frac{R^2}{2m_1n\theta_1^2}}}{\sqrt{\hat{s}_t^2+\hat{s}_c^2}}\right).$$

Finally, we can pick $\delta_1 = 0.995\delta$ and $\delta_2 = \delta_\mu = \delta_s = \frac{\delta}{600}$, which yields

$$\left|\hat{\Delta}_{\mathsf{p}}-\Delta_{\mathsf{p}}\right| \le \sqrt{\frac{4\left(\hat{s}_t^2+\hat{s}_c^2\right)\log(2.01/\delta)}{n}}+\gamma = \sqrt{2\hat{\sigma}_{\mathsf{p}}^2\log(2.01/\delta)}+\gamma, \tag{13}$$

and

$$\gamma = \frac{56R\log(1200/\delta)}{3(n-1)}+\sqrt{\frac{R^2}{2m_1n\theta_1^2}\log\left(\frac{1200}{\delta}\right)}+\sqrt{\frac{4\log(2.01/\delta_1)}{n}}$$
$$\cdot\min\left(\sqrt[4]{\log\left(\frac{1200}{\delta}\right)\frac{R^4}{4m_2n\theta_2^2}}+\sqrt[4]{\log\left(\frac{1200}{\delta}\right)\frac{R^2}{2m_1n\theta_1^2}}, \quad \frac{\sqrt{\log\left(\frac{1200}{\delta}\right)\frac{R^4}{4m_2n\theta_2^2}}+\sqrt{\log\left(\frac{1200}{\delta}\right)\frac{R^2}{2m_1n\theta_1^2}}}{\sqrt{\hat{s}_t^2+\hat{s}_c^2}}\right).$$

□

C.2 Proof of Proposition 4.4

Since $\mathcal{M}_1$ and $\mathcal{M}_2$ satisfy $(\alpha, \varepsilon_1(\alpha))$ and $(\alpha, \varepsilon_2(\alpha))$ Rényi DP, it holds that $m_1\theta_1^2 \le \frac{Cn\varepsilon_1(\alpha)}{\alpha}$ and $m_2\theta_2^2 \le \frac{Cn\varepsilon_2(\alpha)}{\alpha}$ for some universal constant C. Plugging these into Theorem C.1 yields the desired result. □

C.3 Proof of Theorem 4.5

The proof follows from the same step as in the proof of Theorem 4.3, except for replacing the empirical Berstein inequality with the following finite sample (i.e., without-replacement) Berstein inequality:

Lemma C.3 (Proposition 1.4 of Bardenet and Maillard (2015)) *Let $\mathcal{X} = \{x_1, x_2, ..., x_n\}$ be a finite set of N points. Let*

$$a \triangleq \min_{i\in[n]} x_i, \text{ and } b \triangleq \max_{i\in[n]} x_i;$$
$$\mu \triangleq \frac{1}{n}\sum_{i\in[n]} x_i \text{ and } \sigma^2 \triangleq \frac{1}{n}\sum_{i\in[n]}(x_i-\mu)^2.$$

Let $X_1, X_2, ..., X_{n/2}$ denote a random sample drawn without replacement from $\mathcal{X}$. Then, for all $\varepsilon > 0$,

$$\Pr\left(\left|\frac{1}{n}\sum_{i\in[n]} X_i - \mu\right| \ge \varepsilon\right) \le 2\exp\left(-\frac{n\varepsilon^2}{2\sigma^2+(2/3)(b-a)\varepsilon}\right). \tag{14}$$

Notice that Lemma C.3 implies Lemma C.2 with $\mathbb{E}[X_i]$ being replaced by μ and $s^2(\mathbf{X})$ replaced by $\sigma(x_1, ..., x_n)$. As a result, we only need to apply concentration inequalities on the private estimate of $\hat{\mu}_{\mathsf{pr}}$ and $\hat{\sigma}^2_{\mathsf{s}}$, which follows from the proof of Theorem 4.3.

□