Federated Experiment Design under Distributed Differential Privacy

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

Abstract

Experiment design has a rich history dating back to the early 1920s and has found numerous critical applications across various fields since then. However, the use and collection of users' data in experiments often involve sensitive personal information, so additional measures to protect individual privacy are required during data collection, storage, and usage. 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 Neyman's potential outcome framework under DP and secure aggregation, a distributed protocol enabling a service provider to aggregate information without accessing individual data. To achieve DP, we design local privatization mechanisms that are compatible with secure aggregation. We show that when introducing DP noise, it is imperative to 1) cleverly split privacy budgets to estimate both the mean and variance of the outcomes and 2) carefully calibrate the confidence intervals according to the DP noise. Finally, we present comprehensive experimental evaluations of our proposed schemes and show the privacy-utility trade-offs in experiment design.

1. Introduction

Experimental design has a long history, tracing back to the early 1920s in the agricultural domain (Fisher, 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. 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 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 a nutshell, DP 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 crypotographic 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 black box secure aggregation primitive that we refer to as SecAgg, which can be instantiated by (Bonawitz et al., 2016; Bell et al., 2020).

Our contributions. We present a framework that achieves a $(1-\alpha)$-confidence interval (CI) and a level-α test while ensuring distributed DP (formally defined in Section 2). Our framework offers several advantages, including unbiased estimation, efficient memory (or communication) usage, and bounded sensitivities, enabling downstream parties to develop their own privatization mechanisms. We make use of the Poisson-binomial mechanism (PBM) (Chen et al., 2022) as a local randomizer. 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, it is necessary to collect second-moment information such as sample variance. We show that, via SecAgg and DP, this can be achieved by judiciously allocating privacy budgets for estimating the sample mean and variance. Last, our experimental study quantifies the trade-offs between privacy and utility.

1.1. Related Works

The design of experiments to identify causal relations and average treatment effects is crucial in various domains (Imbens & 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 & 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 & Awan, 2023) explores 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 & Kifer, 2017; Cai et al., 2017; Raj et al., 2020) or Local DP (Raj et al., 2020). This work is the first to consider Distributed DP with secure aggregation. We also analyze the distribution-free setting, where no distributional assumptions are imposed on potential outcomes. 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 & 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 & 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 estimate both sample means and variances separately privately (Du et al., 2020; Karwa & Vadhan, 2017; D'Orazio et al., 2015) or use a private bootstrap (Brawner & Honaker, 2018). Our approach resembles the former, but is compatible with secure aggregation and does not require a central data curator.

2. Problem Formulation and Preliminaries

We formulate the experiment design problem via the following *Neyman-Roubin causal model*. For each test unit ("user") $i \in [n]$, we introduce the randomized treatment assignment variable $T_i \in \{0, 1\}$, which indicates whether user i receives the treatment or not. Additionally, we consider the potential outcomes $y_i(1), y_i(0) \in \mathcal{Y}$ 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):

$$\Delta_{\mathsf{s}}(\boldsymbol{y}) \triangleq \frac{1}{n}\sum_{i=1}^{n} y_i(1) - y_i(0).$$

Notice that under the original Neyman-Roubin's framework, the potential outcomes $\boldsymbol{y} \triangleq \{(y_i(1), y_i(0)) | i = 1, ..., n\}$ are deterministic; only the treatment variable T_i's are randomized. However, we can also impose distributional assumptions on the potential outcomes, i.e., $Y_i(0) \overset{\text{i.i.d.}}{\sim} P_0$ and $Y_i(1) \overset{\text{i.i.d.}}{\sim} P_1$, and the quantity of interest is the population average treatment effect (PATE):

$$\Delta_{\mathsf{p}}(P_0, P_1) \triangleq \mathbb{E}_{Y(1)\sim P_1,\, Y(0)\sim P_0}\left[Y(1) - Y(0)\right].$$

In this work, our goals are 1) obtaining confidence intervals of estimated SATE (or PATE) from observed data $\hat{\Delta}_{\mathsf{s}}(X^n)$, and 2) testing if $\Delta_s > 0$ (or $\Delta_p > 0$).

2.1. Secure aggregation and distributed DP

When the service provider has access to all the observable data X_i, it can estimate Δ_{s} via standard difference-in-means estimator (Imbens & Rubin, 2015), compute sample variances of $Y_i(0)$'s and $Y_i(1)$'s, and construct confidence intervals accordingly. However, when the X_i values are treated as sensitive, they should be aggregated securely so that only necessary information is revealed to the service providers.

Secure aggregation. Secure aggregation (such as (Bonawitz et al., 2016)) 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. When applying SecAgg in experiment design, it is important to note that SecAgg typically operates on a finite field, like most cryptographic MPC protocols. Therefore, each outcome X_i needs to be appropriately preprocessed (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 $\hat{\Delta}\left(X^n\right)$ is (ε, δ)-DP, if for any two possible outcome sets $\boldsymbol{y} \triangleq \{(y_i(0), y_i(1))|i = 1, ..., n\}$ and $\boldsymbol{y}' \triangleq \{(y_i(0), y_i(1))|i = 2, ..., n\} \cup \{(y_1'(0), y_1'(1))\}$ differing in one user, we have $\Pr\left\{\hat{\Delta}\left(X^n|\boldsymbol{y}\right) \in \mathcal{S}\right\} \leq e^{\varepsilon}\Pr\left\{\hat{\Delta}\left(X^n|\boldsymbol{y}'\right) \in \mathcal{S}\right\} + \delta.$

A common approach to achieve DP is adding properly calibrated noise (such as zero-mean 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.,

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

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 properly discretized and compatible with secure aggregation; that is, $\mathcal{M}$ has to map X_i into a space $\mathcal{Z}$ (typically 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 use the Rényi DP definition, which allows simpler and tighter privacy composition. See Appendix A for a formal definition.

3. Causal inference with distributed DP

Recall that our objective is to obtain a $(1-\alpha)$-confidence interval for the Sample Average Treatment Effect (SATE) or the Population Average Treatment Effect (PATE), 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 performs secure aggregation to gather necessary information, along with local randomizers $\mathcal{M}_1$ and $\mathcal{M}_2$. These randomizers satisfy the distributed DP conditions defined in Definition A.2 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 $\mathcal{M}_1 : \mathcal{X} \times [n] \to \mathcal{Z}$ and $\mathcal{M}_2 : \mathcal{X} \cdot \mathcal{X} \times [n] \to \mathcal{Z}$, where we use $\mathcal{X} \cdot \mathcal{X} \triangleq \{x^2|x \in \mathcal{X}\}$ to denote the collection of all possible second moment of the samples. In the above notation, we allow the local randomizers to take the size of the control (or test) group, denoted as $n_c \triangleq \sum_{i=1}^n (1-T_i)$ (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. 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 needed for estimating the variance, which is used to construct the confidence intervals.

The following theorem establishes privacy guarantees.

Theorem 3.1. *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be local randomizers for the first and second moments of X_i. Assume $\mathcal{M}_j(\cdot, n^*)$ satisfies $(\alpha, \varepsilon_j(\alpha))$-distributed Rényi DP for $j \in \{1, 2\}$ and $n^* \in [n]$. Then, Algorithm 1 is $(\alpha, \varepsilon_1(\alpha) + \varepsilon_2(\alpha))$-Rényi DP.*

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 $\boldsymbol{y}$ and $\boldsymbol{y}'$.

The next theorem summarizes the performance guarantees, ensuring that Algorithm 1 gives a $(1-\alpha)$-CI asymptotically.

Theorem 3.2. *Under some assumptions on the local encoders and the estimator as listed in Appendix C and let the calibration term (which depends on $\mathcal{M}_1$) be*

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

Then Algorithm 1 gives a $(1-\alpha)$-confidence interval of SATE or PATE.

Algorithm 1 Causal inference with distributed DP

```
Input: treatment variables $T_1, ..., T_n \in \{0, 1\}$, potential outcomes $\{(y_i(0), y_i(1)) | i = 1, ..., n\}$, local randomizer $\mathcal{M}_1 : \mathcal{X} \to \mathcal{Z}$, $\mathcal{M}_2 : \mathcal{X} \cdot \mathcal{X} \to \mathcal{Z}$, privacy budget $\varepsilon_t$ and $\varepsilon_2$.
Output: An unbiased SATE estimator $\hat{\Delta}_{\mathsf{s}}$.
for each user $i$ do
    Obtains the observable treatment outcome $X_i \triangleq T_i y_i(1) + (1 - T_i) y_i(0)$
    computes $\mathcal{M}(X_i)$, and $\mathcal{M}(X_i^2)$.
end for
▷ Aggregation
Server securely aggregates $\sum_i T_i \mathcal{M}_1(X_i, n_t)$, $\sum_i (1 - T_i) \mathcal{M}_1(X_i, n_c)$, $\sum_i T_i \mathcal{M}_2(X_i^2, n_t)$ and $\sum_i (1 - T_i) \mathcal{M}_2(X_i^2, n_c)$.
▷ Estimation
Constructs estimators for sample means and variances:
    $\hat{\mu}_c\left(\sum_i T_i \mathcal{M}_1(X_i)\right)$ and $\hat{\mu}_t\left(\sum_i (1 - T_i) \mathcal{M}_1(X_i)\right)$;
    $\hat{s}_c^2\left(\sum_i T_i \mathcal{M}_2(X_i^2)\right)$ and $\hat{s}_t^2\left(\sum_i (1 - T_i) \mathcal{M}_2(X_i^2)\right)$.
Compute the ATE estimator: $\hat{\Delta} \triangleq \frac{\hat{\mu}_t}{n_t} - \frac{\hat{\mu}_c}{n_c}$.
Compute the DP calibration $\sigma_{\mathsf{pr}}^2(\varepsilon, n_c, n_t)$ by (2).
Set $\hat{\sigma}_{\mathsf{s}}^2 \triangleq \frac{n_c n_t}{n}\left(\frac{\sqrt{\hat{s}_t^2}}{n_t} + \frac{\sqrt{\hat{s}_c^2}}{n_c}\right)^2$ and $\hat{\sigma}_{\mathsf{p}}^2 \triangleq \frac{\hat{s}_t^2}{n_t} + \frac{\hat{s}_c^2}{n_c}$.
Return: $\hat{\Delta} \pm z_{1-\alpha/2} \cdot (\hat{\sigma} + \sigma_{pr})$ for SATE and $\hat{\Delta} \pm z_{1-\alpha/2} \cdot (\hat{\sigma}_{\mathsf{p}} + \sigma_{\mathsf{pr}})$ for PATE.
```

In Algorithm 1, the CIs of SATE and PATE take slightly different forms 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 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.

3.1. Causal inference via Poisson-binomial mechanism

Next, we describe and analyze a particular 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 $\mathcal{Y} = [-c, c]$ for some $c > 0$.

The local randomizer $\mathcal{M}_{\mathsf{PBM}}$ consists of two main steps: 1) first mapping x_i into $\left[\frac{1}{2} - \theta, \frac{1}{2} + \theta\right]$ by $p_i \triangleq \frac{1}{2} + \frac{\theta}{c} x_i$, and then 2) generating a binomial random variable $Z_i \sim \mathsf{Binom}(m, p_i)$ (see Algorithm 2 in Appendix B for a formal statement). Upon securely aggregating $\sum_i Z_i$, the server can obtain an unbiased estimator on $\mu = \sum_i x_i$ as

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

Recall that the server can only learn $\sum_i Z_i$ but not individual Z_i's. Next, we construct $\mathcal{M}_1(\cdot, n^*)$ and $\mathcal{M}_2(\cdot, n^*)$ in Algorithm 1 via PBM and summarize the privacy and utility guarantee in the following theorem.

Corollary 3.3. *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be implemented with PBM with proper parameters (see Appendix B for the parameter selection). Then under a (ε, δ)-DP constraint, the average width of the CIs is $O_\delta\left(z_{1-\alpha/2} \cdot \sqrt{\frac{s_c^2}{n_c} + \frac{s_t^2}{n_t} + \frac{c^2}{\varepsilon^2}\left(\frac{1}{n_t^2} + \frac{1}{n_c^2}\right)}\right)$ for SATE, and $O_\delta\left(z_{1-\alpha/2} \cdot \sqrt{\frac{\mathsf{Var}(P_0)}{n_c} + \frac{\mathsf{Var}(P_1)}{n_t} + \frac{c^2}{\varepsilon^2}\left(\frac{1}{n_t^2} + \frac{1}{n_c^2}\right)}\right)$ for PATE.*

Note that the privacy parameters of $\mathcal{M}_2$ has little impact on the (asymptotic) width of CIs. This is due to the fact that as long as we can derive a consistent estimator for the sample variances, we can compute CIs accordingly. Therefore, we should allocate the maximum possible privacy budget to $\mathcal{M}_1$. In practice, we set the privacy budget for $\mathcal{M}_1$ to be 0.99 of the total privacy allocation.

3.2. Experiments

Table 1. Average widths and coverages of 90%-confidence intervals for PATE. We generate $Y_i(0) \overset{\text{i.i.d.}}{\sim} N(-0.1, \sigma_p^2)$ and $Y_i(1) \overset{\text{i.i.d.}}{\sim} N(0.1, \sigma_p^2)$, with $\sigma_p = 0.05$. We divide the sample size $n = 10^3$ equally into test and control groups and simulate for 10^4 rounds.

		$\varepsilon = 0.1$	$\varepsilon = 1.0$	$\varepsilon = 1.9$	$\varepsilon = \infty$
None private	Coverage (90% CI)	-	-	-	0.902
	Width (90% CI)	-	-	-	$2.08{\cdot}10^{-3}$
Gaussian	Coverage (90% CI)	0.899	0.899	0.899	-
	Width (90% CI)	0.771	0.078	0.044	-
PBM	Coverage (90% CI)	0.898	0.900	0.903	-
	Width (90% CI)	0.772	0.085	0.048	-

We compare the proposed distributed DP method, based on PBM with (1) the none-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). 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 relatively small. Mre detailed experimental results can be found in the Appendix F.

4. Acknowledgements

This work was mainly completed during Wei-Ning's internship at Meta, and was supported in part by NSF Award # CCF-2213223.

References

Agarwal, N., Suresh, A. T., Yu, F. X. X., Kumar, S., and McMahan, B. cpsgd: Communication-efficient and differentially-private distributed sgd. In *Advances in Neural Information Processing Systems*, pp. 7564–7575, 2018.

Agarwal, N., Kairouz, P., and Liu, Z. The skellam mechanism for differentially private federated learning. *Advances in Neural Information Processing Systems*, 34, 2021.

Balle, B. and Wang, Y.-X. Improving the gaussian mechanism for differential privacy: Analytical calibration and optimal denoising. In *International Conference on Machine Learning*, pp. 394–403. PMLR, 2018.

Bell, J. H., Bonawitz, K. A., Gascón, A., Lepoint, T., and Raykova, M. Secure single-server aggregation with (poly) logarithmic overhead. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, pp. 1253–1269, 2020.

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

Bhowmick, A., Duchi, J., Freudiger, J., Kapoor, G., and Rogers, R. Protection against reconstruction and its applications in private federated learning. *arXiv preprint arXiv:1812.00984*, 2018.

Biswas, S., Dong, Y., Kamath, G., and Ullman, J. Coinpress: Practical private mean and covariance estimation. *Advances in Neural Information Processing Systems*, 33: 14475–14485, 2020.

Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Ramage, D., Segal, A., and Seth, K. Practical secure aggregation for federated learning on user-held data. *arXiv preprint arXiv:1611.04482*, 2016.

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

Cai, B., Daskalakis, C., and Kamath, G. Priv'it: Private and sample efficient identity testing. In *International Conference on Machine Learning*, pp. 635–644. PMLR, 2017.

Canonne, C. L., Kamath, G., and Steinke, T. The discrete gaussian for differential privacy. *arXiv preprint arXiv:2004.00010*, 2020.

Chen, W.-N., Kairouz, P., and Ozgur, A. Breaking the communication-privacy-accuracy trilemma. *Advances in Neural Information Processing Systems*, 33, 2020.

Chen, W.-N., Ozgur, A., and Kairouz, P. The poisson binomial mechanism for unbiased federated learning with secure aggregation. In *International Conference on Machine Learning*, pp. 3490–3506. PMLR, 2022.

Damgård, I., Pastro, V., Smart, N., and Zakarias, S. 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*, pp. 643–662. Springer, 2012.

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

Du, W., Foot, C., Moniot, M., Bray, A., and Groce, A. Differentially private confidence intervals. *arXiv preprint arXiv:2001.02285*, 2020.

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

Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I., and Naor, M. Our data, ourselves: Privacy via distributed noise generation. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pp. 486–503. Springer, 2006a.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pp. 265–284. Springer, 2006b.

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

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

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

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

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

Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., et al. Advances and open problems in federated learning. *arXiv preprint arXiv:1912.04977*, 2019.

Kairouz, P., Liu, Z., and Steinke, T. The distributed discrete gaussian mechanism for federated learning with secure aggregation. *arXiv preprint arXiv:2102.06387*, 2021.

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

Lee, S. K., Gresele, L., Park, M., and Muandet, K. Privacy-preserving causal inference via inverse probability weighting. *arXiv preprint arXiv:1905.12592*, 2019.

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

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

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

Niu, F., Nori, H., Quistorff, B., Caruana, R., Ngwe, D., and Kannan, A. Differentially private estimation of heterogeneous causal effects. In *Conference on Causal Learning and Reasoning*, pp. 618–633. PMLR, 2022.

Ohnishi, Y. and Awan, J. Locally private causal inference. *arXiv preprint arXiv:2301.01616*, 2023.

Raj, A., Law, H. C. L., Sejdinovic, D., and Park, M. 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*, pp. 697–724. Springer, 2020.

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

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

A. Rényi DP for ATE

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

Definition A.1 (Rényi differential privacy)**.** We say an ATE estimator $\hat{\Delta}\left(X^n\right)$ is $(\alpha, \varepsilon(\alpha))$-DP, if for any two sets of possible outcomes $\boldsymbol{y} \triangleq \{(y_i(0), y_i(1)) | i = 1, ..., n\}$ and $\boldsymbol{y'} \triangleq \{(y_i(0), y_i(1)) | i = 2, ..., n\} \cup \{(y'_1(0), y'_1(1))\}$ that differ in one user, it holds that

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

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

Definition A.2 (Distributed Renyi DP)**.** A local randomizer $\mathcal{M}$ is $(\alpha, \varepsilon(\alpha))$-DP, if for any two possible outcome sets $\boldsymbol{y} \triangleq \{(y_i(0), y_i(1)) | i = 1, ..., n\}$ and $\boldsymbol{y'} \triangleq \{(y_i(0), y_i(1)) | i = 2, ..., n\} \cup \{(y'_1(0), y'_1(1))\}$ differing in one user, we have $D_\alpha\left(\sum_i \mathcal{M}(X_i|\boldsymbol{y}) \| \sum_i \mathcal{M}(X_i|\boldsymbol{y'})\right) \leq \varepsilon(\alpha)$.

B. Discrete DP Mechanisms for Secure Aggregation

In this section, we introduce discrete mechanisms that can be combined with secure aggregation for causal inference. We analyze their performance and provide empirical evaluations in Section 3.2. These discrete mechanisms can be roughly categorized into two classes:

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).

Randomized Response Mechanisms: This class of mechanisms is based on the concept of randomized response introduced by Warner (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 snce 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.

Due to the space limitation, we defer the analysis of additive noise mechanisms to supplemental material and only present the results of randomized response mechanisms here.

B.1. Difference-in-mean estimator with the Poisson-binomial mechanism

Algorithm 2 The Poisson Binomial Mechanism

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

Next, we describe and analyze another 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 $\mathcal{Y} = [-c, c]$ for some $c > 0$[1]. Per Theorem 3.1, 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 $p_i \triangleq \frac{1}{2} + \frac{\theta}{c} x_i$, 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

$$\hat{\mu}\left(\sum_i Z_i\right) \triangleq \frac{c}{nm\theta}\left(\sum_i Z_i - \frac{mn}{2}\right) \tag{4}$$

(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 B.1 (Cor. 3.3(Chen et al., 2022))**.** *Let $\hat{\mu}$ from (4) be the estimator. Then for any $\theta \in [0, 1/4]$ and $m, n \in \mathbb{N}$,*

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

$$\varepsilon(\alpha) \geq C\left(\frac{\theta^2}{(1-2\theta)^4}\right)\frac{\alpha m}{n}, \tag{5}$$

where $C > 0$ is a universal constant.

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

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 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.

Remark B.2. 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% greater of the actual privacy loss, described in Appendix E.

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 B.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 $O\left(\frac{\alpha\theta_{1,t}^2 m_{1,t}}{n_t}\right)$[2], 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 B.1 with Theorem 3.1, we summarize the guarantees of PBM in the following corollary:

Corollary B.3. *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. *Algorithm 1 is $(\alpha, \varepsilon(\alpha))$-Rényi DP for all l$\alpha > 1$ and*

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

2. *The average width of the $(1-\alpha)$-CI is $O\left(z_{1-\alpha/2} \cdot \sqrt{\frac{s_c^2}{n_c} + \frac{s_t^2}{n_t} + \frac{c^2}{n_t m_{1,t}\theta_{1,c}^2} + \frac{c^2}{n_t m_{1,t}\theta_{1,t}^2}}\right)$ for SATE, and $O\left(z_{1-\alpha/2} \cdot \sqrt{\frac{\mathsf{Var}(P_0)}{n_c} + \frac{\mathsf{Var}(P_1)}{n_t} + \frac{c^2}{n_t m_{1,t}\theta_{1,c}^2} + \frac{c^2}{n_t m_{1,t}\theta_{1,t}^2}}\right)$ for PATE.*

[1] Note that here we assume $c > 0$ is known beforehand, which could be true in some cases. When c is unknown, we may need to estimate it through private range/quantile queries.

[2] 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 due to the fact that 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, as demonstrated in the next section, 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 implies that the average width of the CIs is $O\left(z_{1-\alpha/2} \cdot \sqrt{\frac{s_c^2}{n_c} + \frac{s_t^2}{n_t} + \frac{c^2}{\varepsilon^2}\left(\frac{1}{n_t^2} + \frac{1}{n_c^2}\right)}\right)$ for SATE (or $O\left(z_{1-\alpha/2} \cdot \sqrt{\frac{\mathsf{Var}(P_0)}{n_c} + \frac{\mathsf{Var}(P_1)}{n_t} + \frac{c^2}{\varepsilon^2}\left(\frac{1}{n_t^2} + \frac{1}{n_c^2}\right)}\right)$ for PATE).

C. Assumptions of Theorem 3.2

Assumption C.1. Assume the estimator $\hat{\mu}_j$, $j \in \{0, 1\}$, are of an additive structure. That is,

$$\begin{cases} \hat{\mu}_t(\sum_i T_i \mathcal{M}_1(X_i)) = \sum_i T_i \hat{\mu}(\mathcal{M}_1(X_i)); \\ \hat{\mu}_c(\sum_i (1-T_i)\mathcal{M}_1(X_i)) = \sum_i (1-T_i)\hat{\mu}(\mathcal{M}_1(X_i)), \end{cases}$$

where $\hat{\mu}\left(\mathcal{M}_1(x_i, n^*)\right)$ gives an unbiased estimator, independent with T_i, on x_i with variance bounded by $\sigma_1^2(n^*, \varepsilon)$[3];

Assumption C.2. Assume $\hat{s}_c^2$ and $\hat{s}_t^2$ defined in Algorithm 1 yield consistent estimation on the sample variances $s_c^2 \triangleq \frac{1}{n-1}\sum_i \left(y_i(0) - \bar{y}(0)\right)$ and $s_t^2 \triangleq \frac{1}{n-1}\sum_i \left(y_i(1) - \bar{y}(1)\right)$, respectively. That is, $\hat{s}_c^2\left(\sum_{i=1}^n (1-T_i)\mathcal{M}_2(X_i^2)\right) \xrightarrow{p} \hat{\mu}_c$ as $n \to \infty$ (and so does $\hat{s}_t$).

D. Proof of Theorem 3.1

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). □

E. 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

$$\max_{t_1, t_2 \in [m \cdot n],\, |t_1 - t_2| \leq m} D_\alpha\left(P_{\mathsf{Binom}\left(t_1, \frac{1}{2}-\theta\right) + \mathsf{Binom}\left(mn - t_1, \frac{1}{2}+\theta\right)} \middle\| P_{\mathsf{Binom}\left(t_2, \frac{1}{2}-\theta\right) + \mathsf{Binom}\left(mn - t_2, \frac{1}{2}+\theta\right)}\right). \tag{6}$$

In (Chen et al., 2022), it is shown that the maximum of (6) occurs at $(t_1, t_2) = (0, m)$, which suggests the following (exact) privacy accounting mechanism in Algorithm 3.

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

Algorithm 3 Exact privacy accounting.

```
**Input:** $n, m, \theta, \alpha$
**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.}
$\varepsilon(\alpha) \leftarrow \frac{1}{\alpha-1} \log \left( \mathsf{sum} \left( \frac{P_1^\alpha}{P_2^{\alpha-1}} \right) \right)$ {$\mathsf{sum}$ 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:** $n, m, \theta, \alpha$
**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*.}
$\varepsilon(\alpha) \leftarrow \frac{1}{\alpha-1} \mathsf{logexpsum} \left( \alpha \cdot \mathsf{logP}_1 + (1-\alpha) \cdot \mathsf{logP}_2 \right)$
```

E.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 process inequality:

$$(6) \leq \max_{k \in [n-1]} m \cdot D_\alpha \left(P_{\mathsf{Binom}\left(1+k, \frac{1}{2}-\theta\right) + \mathsf{Binom}\left(n-k-1, \frac{1}{2}+\theta\right)} \middle\| P_{\mathsf{Binom}\left(k, \frac{1}{2}-\theta\right) + \mathsf{Binom}\left(n-k, \frac{1}{2}+\theta\right)} \right). \tag{7}$$

Although (7) is always strictly greater than the exact privacy loss (6), 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 (7), we arrive at the following approximate privacy accounting algorithm, which reduce the computational complexity to $O(n)$:

Algorithm 5 Efficient approximate privacy.

```
**Input:** $n, m, \theta, \alpha$
$\mathsf{logP}_1 \leftarrow \log \left( \mathsf{Binom}(n, \frac{1}{2} - \theta) \right)$
$\mathsf{logP}_2 \leftarrow \log \left( \mathsf{Binom}(n-1, \frac{1}{2} - \theta) \right)$
$\mathsf{logP}_2' \leftarrow \mathsf{Ber}(\frac{1}{2} + \theta)$
$\mathsf{logP}_2 \leftarrow \mathsf{logP}_2 \,\tilde{*}\, \mathsf{logP}_2'$ {$\tilde{*}$ 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 in (Canonne et al., 2020).

F. Additional experiments

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

F.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[4]. 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).

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

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) are cancelled out (see the proof given in Section G for more details).

[4] 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 3.1.

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

F.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

F.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 F.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

G. Proof of Theorem 3.1

In this section, we provide a formal proof of the asymptotic confidence level of estimating PATE in Theorem 3.1.

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

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

The finite-sample central limit theorem (Hájek, 1961) (see also (Li & 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 $\boldsymbol{y}$ and T_i,

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

where $\sigma^2_{\mathsf{pr}}(n_c, n_t, \varepsilon) \triangleq \frac{n}{n_c}\sigma_1^2(n_c, \varepsilon) + \frac{n}{n_t}\sigma_1^2(n_t, \varepsilon)$ and the convergence is due to the (classical) central limit theorem and Assumption C.1. Since the DP noise is independent with T_i, we conclude

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

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

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

by Slutsky's theorem $\left[\hat{\Delta}_{\mathsf{s}} - z_{1-\alpha/2}\cdot\left(\hat{\sigma}_{\mathsf{s}} + \sigma_{pr}\right), \hat{\Delta}_{\mathsf{s}} + z_{1-\alpha/2}\cdot\left(\hat{\sigma}_{\mathsf{s}} + \sigma_{\mathsf{pr}}\right)\right]$ gives an $1-\alpha$ CI asymptotically. The analysis for PATE is similar to the above and we leave the details to the supplementary materials.

Next, we prove the confidence level of estimating Proof. 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

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

Therefore, $\hat{\sigma}_{\mathsf{p}}$ 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 $\sigma^2_{\mathsf{pr}}(n_c, n_t, \varepsilon)$. By the central limit theorem, the proof is complete. □