When Data Can't Meet: Estimating Correlation Across Privacy Barriers

Abhinav Chakraborty
Columbia University
New York, NY 10027
ac4662@columbia.edu

Arnab Auddy
The Ohio State University
Columbus, OH 43210
arnab.auddy@columbia.edu

T. Tony Cai
The Wharton School
University of Pennsylvania
Philadelphia, PA 19104
tcai@wharton.upenn.edu

Abstract

We consider the problem of estimating the correlation of two random variables X and Y, where the pairs (X, Y) are not observed together, but are instead separated co-ordinate-wise at two servers: server 1 contains all the X observations, and server 2 contains the corresponding Y observations. In this vertically distributed setting, we assume that each server has its own privacy constraints, owing to which they can only share suitably privatized statistics of their own component observations. We consider differing privacy budgets $(\varepsilon_1, \delta_1)$ and $(\varepsilon_2, \delta_2)$ for the two servers and determine the minimax optimal rates for correlation estimation allowing for both non-interactive and interactive mechanisms. We also provide correlation estimators that achieve these rates and further develop inference procedures, namely, confidence intervals, for the estimated correlations. Our results are characterized by an interesting rate in terms of the sample size n, ε_1, ε_2, which is strictly slower than the usual central privacy estimation rates. More interestingly, we find that the interactive mechanism is always better than its non-interactive counterpart whenever the two privacy budgets are different. Results from extensive numerical experiments support our theoretical findings.

1 Introduction

Federated learning is a popular and extensively studied framework in modern machine learning. In traditional federated learning, due to privacy concerns, the servers are not allowed to pool raw data, but are restricted to sharing only sufficiently privatized statistics derived from the local observations. This method is particularly beneficial when training on sensitive data, such as healthcare or finance. The federated scenario is very systematically studied when the separation occurs horizontally, i.e. observations of the same set of features are binned separately into different servers. See, for example, Kairouz et al. [2021], Li et al. [2020a,b], Zhang et al. [2021] and the references therein.

To encourage collaboration on proprietary data across different organizations, however, it is often more reasonable to assume that the federation occurs "vertically", or across features. For example, in healthcare data, a hospital and a pharmaceutical company might have different pieces of information on the same patient: the hospital does not share private clinical information such as patient demographics or test results with the company, which instead has its own private information on the same patient's response to certain drugs. This new framework called vertical federated learning has recently seen studied in Chen et al. [2020], Liu et al. [2024], Wu et al. [2020], Wei et al. [2022], Yang et al. [2019], but a theoretical understanding of estimation and inference has largely been missing. This motivates the current work. We study the correlation of bivariate data from n pairs of samples (X_i, Y_i)

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

which are not *observed* together, but are instead separated into two servers as $\{X_i : 1 \le i \le n\}$ and $\{Y_i : 1 \le i \le n\}$.

To distinguish our results from the influence of estimating the marginal distributions of X and Y, we assume that $\mathbb{E}(X) = \mathbb{E}(Y) = 0$ and $\mathrm{Var}(X) = \mathrm{Var}(Y) = 1$, and (X, Y) are sub-Gaussian. That is, we assume that our data are pre-normalized to have mean zero and variance one. We revisit the question of normalization in the supplementary material and show both theoretically and in numerical experiments that the rate of correlation estimation is not influenced by this step. In this situation, we consider estimating $\rho = \mathbb{E}(XY)$ from the statistics shared by the two servers: viz., Server 1 releases $T_1(X_1, \ldots, X_n)$, and Server 2 releases $T_2(Y_1, \ldots, Y_n)$. To protect user privacy, we impose the differential privacy framework (see, e.g., Abowd et al. [2020], Bassily et al. [2014], Dwork [2006], Karwa and Vadhan [2017]) on T_1 and T_2; both of which must satisfy $(\varepsilon_1, \delta_1)$ and $(\varepsilon_2, \delta_2)$-DP constraints. For ease of reference, we will somewhat loosely denote the above by a server-level $(\varepsilon_1, \varepsilon_2, \delta_1, \delta_2)$-DP constraint and introduce specific definitions later. Such distributed privacy requirements are frequently used in federated learning. See, e.g., Auddy et al. [2024], Cai et al. [2024a,b,c], Shen et al. [2022], Wei et al. [2020, 2021] and references therein.

1.1 Main results

The key finding in this work is that the complexity of the correlation estimation in the above setup fundamentally depends on whether or not the statistics T_1 and T_2 are allowed to depend on one another. We now present our main results. Throughout this paper, we assume $\varepsilon_1, \varepsilon_2 \le C$ for a constant $C > 0$.

1.1.1 Non-interactive protocol

In our first set of results, we consider estimating ρ in the non-interactive (NI) framework of stricter privacy requirements, where T_1 and T_2 are constructed independently, i.e., without any interaction or information about one another. In this case, the differential privacy requirements on T_1 and T_2 are as follows. With $\mathbf{X} = (X_1, \ldots, X_n)$, $\mathbf{Y} = (Y_1, \ldots, Y_n)$, and similarly $\mathbf{X}'$, $\mathbf{Y}'$ (with one data point replaced):

$$\begin{aligned}\mathbb{P}(T_1(\mathbf{X}) \in A|\mathbf{X}) &\le \exp(\varepsilon_1)\mathbb{P}(T_1(\mathbf{X}') \in A|\mathbf{X}') + \delta_1\\ \mathbb{P}(T_2(\mathbf{Y}) \in A|\mathbf{Y}) &\le \exp(\varepsilon_2)\mathbb{P}(T_2(\mathbf{Y}') \in A|\mathbf{Y}') + \delta_2.\end{aligned}$$

Let $\mathrm{NI}(\varepsilon_1, \varepsilon_2, \delta_1, \delta_2)$ to be the class of all correlation estimators constructed using $T_1(\mathbf{X})$ and $T_2(\mathbf{Y})$ satisfying the above privacy requirement. The following theorem states the minimax rate for estimating ρ in this scenario.

Theorem 1.1. *The minimax rate for estimating correlation ρ via a non-interactive procedure satisfying server level $(\varepsilon_1, \varepsilon_2, \delta_1, \delta_2)$-DP constraints is given by*

$$\inf_{\widehat{\rho} \in \mathrm{NI}(\varepsilon_1, \varepsilon_2, \delta_1, \delta_2)} \sup_{\rho \in [-1,1]} \mathbb{E}\left(\widehat{\rho} - \rho\right)^2 \asymp L_n \left(\frac{1}{n\varepsilon_1^2} + \frac{1}{n\varepsilon_2^2}\right)$$

for a factor L_n of order at most $O(\log(n))$, whenever $\delta_1, \delta_2 = o(n^{-1})$.

Note that the rate does not depend on δ's. This implies that our rate matching correlation estimator achieves $(\varepsilon_1, \varepsilon_2, 0, 0)$-DP, and is still rate optimal (up to logarithmic terms) even within $\mathrm{NI}(\varepsilon_1, \varepsilon_2, \delta_1, \delta_2)$, the class of all non-interactive estimators satisfying $(\varepsilon_1, \delta_1)$ and $(\varepsilon_2, \delta_2)$ DP constraints for δ_1, δ_2 are small positive numbers. The rate optimal estimator in this case is given by the correlation of privatized batch means from both servers.

It is useful to compare the above rate with the ones existing in the literature. Firstly, when (X, Y) are jointly observed, and we impose (ε, δ)-central DP constraints on (X_i, Y_i), the optimal correlation estimation rate is given by $\frac{1}{n^2\varepsilon^2}$. See, e.g., Biswas et al. [2020], Cai et al. [2021]. As expected, when $\varepsilon_1 = \varepsilon_2 = \varepsilon$, this is better than the rate we observe in the current feature separated case, thus highlighting the cost of vertical federation. A second comparison can be made with component-wise local privacy rates, studied in Amorino and Gloter [2023]. The authors there show that in the vertically separated scenario, if we impose $(\varepsilon_1, 0)$ and $(\varepsilon_2, 0)$ *local* DP constraints, the minimax estimation rate for correlation is given by $\frac{1}{n\varepsilon_1^2\varepsilon_2^2}$, which is again strictly worse than the rates we find under the server level DP constraints.

1.1.2 Interactive protocol

We next move on to a larger class of estimators in the interactive (INT) framework, where we still require server level privacy, but one of the servers is allowed access to the privatized statistic output from the other. In other words, we allow the functions T_1 and T_2 to have one way interaction with each other. This requires making exactly one out of two possible choices. The first possibility is that when constructing T_2, Server 2 has access to $T_1(\mathbf{X})$, in addition to its own data $\mathbf{Y}$. The second possibility arises by analogously interchanging the roles of servers 1 and 2. To fix ideas, if we are in the first case, i.e server 2 gets to observe the transcript T_1, before computing T_2, the privacy requirements become:

$$\begin{aligned}\mathbb{P}(T_1(\mathbf{X}) \in A|\mathbf{X}) &\leq \exp(\varepsilon_1)\mathbb{P}(T_1(\mathbf{X}') \in A|\mathbf{X}') + \delta_1 \\ \mathbb{P}(T_2(\mathbf{Y}, T_1(\mathbf{X})) \in A|\mathbf{X}, \mathbf{Y}) &\leq \exp(\varepsilon_2)\mathbb{P}(T_2(\mathbf{Y}', T_1(\mathbf{X})) \in A|\mathbf{X}, \mathbf{Y}') + \delta_2.\end{aligned}$$

Replacing $\mathbf{X}$ with $\mathbf{Y}$ and the index 1 with 2 allows one to write the analogous privacy constraint in the second case where Server 1 has access to $T_2(\mathbf{Y})$. Let $\mathrm{INT}(\varepsilon_1, \varepsilon_2, \delta_1, \delta_2)$ to be the class of all correlation estimators constructed using $T_1(\mathbf{X})$ and $T_2(\mathbf{Y}, T_1(\mathbf{X}))$ satisfying the above interactive privacy requirement. The following theorem states the minimax rate for estimating ρ in this scenario.

Theorem 1.2. *The minimax rate for estimating correlation ρ via a non-interactive procedure satisfying server level $(\varepsilon_1, \varepsilon_2, \delta_1, \delta_2)$-DP constraints is given by*

$$\inf_{\widehat{\rho} \in \mathrm{INT}(\varepsilon_1,\varepsilon_2,\delta_1,\delta_2)} \sup_{\rho \in [-1,1]} \mathbb{E}\left(\widehat{\rho} - \rho\right)^2 \asymp L_n \left(\frac{1}{n(\varepsilon_1 \vee \varepsilon_2)^2} + \frac{1}{n^2 \varepsilon_1^2 \varepsilon_2^2} \right).$$

for a factor L_n of order at most $O(\log(n))$, whenever $\delta_1, \delta_2 = o(n^{-1})$.

Note that unlike NI, in the INT rate, the dominating term depends on $\varepsilon_1 \vee \varepsilon_2$, i.e. the less stringent privacy requirement. The stronger privacy requirement i.e., $\varepsilon_1 \wedge \varepsilon_2$ appears in the second term, but its effect is mitigated by the better sample size factor n^{-2}. This leads to INT being a strictly better estimator than NI whenever $\varepsilon_1 \neq \varepsilon_2$. An interesting special case is when $\mathbf{X}$ are public, meaning ε_1 is a constant, in which case we find $(\varepsilon_2, \delta_2)$-central DP rates for correlation estimation.

The rate optimal estimator in the interactive case is borne out of a natural idea: the server with a less stringent privacy budget should share their statistics with the other server. That is, if $\varepsilon_1 > \varepsilon_2$, we should allow T_2 to depend on $T_1(\mathbf{X})$ and $\mathbf{Y}$. The situation is reversed if $\varepsilon_2 > \varepsilon_1$.

In addition to point estimates $\widehat{\rho}$, we also derive asymptotically valid confidence intervals in both the non-interactive (NI) and interactive (INT) scenarios. That is, we find $(\widehat{\rho}_{L,n}^{(\mathrm{NI})}, \widehat{\rho}_{U,n}^{(\mathrm{NI})})$ and $(\widehat{\rho}_{L,n}^{(\mathrm{INT})}, \widehat{\rho}_{U,n}^{(\mathrm{INT})})$ such that for fixed $\alpha \in (0,1)$

$$\mathbb{P}\left(\widehat{\rho}_{L,n}^{(k)} \leq \rho \leq \widehat{\rho}_{U,n}^{(k)}\right) \to 1 - \alpha \quad \text{as } n \to \infty, \qquad \text{for } k \in \{\mathrm{NI}, \mathrm{INT}\}.$$

We show that our estimation methods are minimax optimal by proving corresponding lower bounds, which to the best of our knowledge, has not been established previously under central differential privacy in a vertically distributed setting. While we follow the classical Le Cam framework, our main technical contribution is a direct control of KL divergence via Fisher information curvature bounds, yielding sharp lower bounds under both non-interactive and one-way interactive protocols. These bounds match our upper bounds up to constants in the Gaussian case and up to logarithmic factors in the sub-Gaussian case. Prior works, such as Hadar et al. [2019], bound KL via mutual information in communication constraint settings; we take a more direct route tailored to central DP. Unlike local DP lower bounds in Amorino and Gloter [2023], our approach handles the more delicate structure of central privacy with vertical data splitting.

The rest of the paper is organized as follows. In Sections 2 and 3 respectively, we describe non-interactive and interactive correlation estimators for bivariate Gaussian and bivariate sub-Gaussian distributions. Section 4 provides minimax lower bounds showing that our estimation procedures are nearly optimal in all cases. Finally, Section 5 shows numerical experiments to corroborate our theoretical results. All proofs are in the supplementary material.

2 Non-interactive estimation methods

We first demonstrate an estimation procedure in the non-interactive paradigm. Here Server 1 and Server 2 construct and share $T_1(\mathbf{X})$ and $T_2(\mathbf{X})$ without knowledge of one another. As mentioned in the introduction $T_1(\mathbf{X})$ must satisfy $(\varepsilon_1, \delta_1)$-DP and $T_2(\mathbf{X})$ must satisfy $(\varepsilon_2, \delta_2)$-DP constraints. Our estimator is based on sharing privatized batch means. Choosing $m \geq 1$ we separate the n observations in each server into batches of size m as follows:

$$B_j = \{m(j-1)+1, \ldots, mj\} \quad \text{for} \quad j = 1, \ldots, k \quad \text{where } k = \lfloor \tfrac{n}{m} \rfloor. \tag{1}$$

2.1 Non-interactive correlation estimation for Gaussian distribution

In this subsection, we assume that $(X, Y) \sim \mathcal{N}(\mathbf{0}, \Sigma(\rho))$ with $(\Sigma(\rho))_{11} = (\Sigma(\rho))_{22} = 1$ and $(\Sigma(\rho))_{12} = \rho$, the bivariate Gaussian distribution with $\mathbb{E}(X) = \mathbb{E}(Y) = 0$, $\mathrm{Var}(X) = \mathrm{Var}(Y) = 1$ and correlation $\mathbb{E}(XY) = \rho$.

Our estimation procedure for ρ is through the product of sample means across multiple batches. In order to bound the sensitivity directly, i.e., without clipping, we will use the signs of X_i and Y_i in place of (X_i, Y_i) themselves, to compute our correlation estimator.

$$\bar{X}^{(j)} = \frac{1}{m} \sum_{i \in B_j} \mathrm{sign}(X_i), \text{ and } \bar{Y}^{(j)} = \frac{1}{m} \sum_{i \in B_j} \mathrm{sign}(Y_i) \tag{2}$$

where B_j are as defined in (1) for $j = 1, \ldots, k$. To ensure $(\varepsilon_1, 0)$-DP and $(\varepsilon_2, 0)$-DP constraints each server adds Laplace noise to each batch mean and outputs the vectors $T_1(\mathbf{X}), T_2(\mathbf{Y}) \in \mathbb{R}^m$ with elements:

$$(T_1(\mathbf{X}))_j = \sqrt{m}(\bar{X}^{(j)} + Z_1^{(j)}) \quad \text{and} \quad (T_2(\mathbf{Y}))_j = \sqrt{m}(\bar{Y}^{(j)} + Z_2^{(j)}) \quad \text{for} \quad 1 \leq j \leq k,$$

where $Z_l^{(j)} \overset{indep}{\sim} \text{Laplace}\left(0, \frac{2}{m\varepsilon_l}\right)$ for $l = 1, 2$. We can then compute

$$\widehat{\eta}_{XY} = \frac{1}{k} \sum_{j=1}^{k} (T_1(\mathbf{X}))_j (T_2(\mathbf{Y}))_j. \tag{3}$$

Since (X, Y) are bivariate Gaussians, the covariance above satisfies

$$\mathbb{E}[\widehat{\eta}_{XY}] = 2\mathbb{P}(XY > 0) - 1 = 1 - \frac{2\arccos(\rho)}{\pi}, \tag{4}$$

which leads to the method-of-moments based private correlation estimator:

$$\widehat{\rho}_{\mathrm{NI}}^{(G)} := \cos\left(\frac{\pi}{2}(1 - \widehat{\eta}_{XY}^{(P)})\right) = \sin\left(\frac{\pi \widehat{\eta}_{XY}^{(P)}}{2}\right).$$

We would like to emphasize that (4) is precisely where we use the assumption of Gaussianity on (X, Y). Since the bivariate distribution is completely known once ρ is specified, we can explicitly write $\mathbb{P}(XY > 0)$ as a function of ρ, which in turn enables our sign-based estimation procedure. While this can be extended to other bivariate families which are specified by a single correlation parameter ρ, we do not discuss these details for brevity.

To create confidence intervals for ρ, let us define S_η^2 to be the sample variance of $\{(T_1(\mathbf{X}))_j (T_2(\mathbf{Y}))_j : 1 \leq j \leq k\}$. Then we can define the confidence interval:

$$\mathrm{CI}_{\mathrm{NI}}^{(G)}(\alpha) := \left(\widehat{\rho}_{\mathrm{NI}}^{(G)} - \frac{\pi S_\eta \sqrt{1 - (\widehat{\rho}_{\mathrm{NI}}^{(G)})^2}}{2\sqrt{k}} z_{1-\alpha/2}, \widehat{\rho}_{\mathrm{NI}}^{(G)} + \frac{\pi S_\eta \sqrt{1 - (\widehat{\rho}_{\mathrm{NI}}^{(G)})^2}}{2\sqrt{k}} z_{1-\alpha/2}\right) \tag{5}$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-th quantile of the standard Normal distribution.

2.2 Non-interactive correlation estimation for sub-Gaussian distributions

In general, we would deal with non-Gaussian data, and thus the sign-based procedure of the previous section would not be exact anymore. We will use a clipping based estimator for this case. For clipping parameters $\lambda_1, \lambda_2 > 0$ to be chosen later we replace (2) by

$$\bar{X}^{(j)} = \frac{1}{m} \sum_{i \in B_j} \mathrm{sign}(X_i)(|X_i| \wedge \lambda_1) \text{ and } \bar{Y}^{(j)} = \frac{1}{m} \sum_{i \in B_j} \mathrm{sign}(Y_i)(|Y_i| \wedge \lambda_2) \tag{6}$$

where B_j are as defined in (1) for $j = 1, \ldots, k$. As before, each server adds Laplace noise to each batch mean and shares:

$$(T_1(\mathbf{X}))_j = \sqrt{m}(\bar{X}^{(j)} + Z_1^{(j)}) \quad \text{and} \quad (T_2(\mathbf{Y}))_j = \sqrt{m}(\bar{Y}^{(j)} + Z_2^{(j)}) \quad \text{for} \quad 1 \le j \le k,$$

where $Z_l^{(j)} \overset{indep}{\sim} \text{Laplace}\left(0, \frac{2\lambda_l}{m\varepsilon_l}\right)$ for $l = 1, 2$. Then we will estimate ρ by the quantity:

$$\widehat{\rho}_{\text{NI}}^{(SG)} = \frac{1}{k}\sum_{j=1}^{k}(T_1(\mathbf{X}))_j(T_2(\mathbf{Y}))_j. \tag{7}$$

Once again defining S_ρ^2 to be the sample variance of $\{(T_1(\mathbf{X}))_j(T_2(\mathbf{Y}))_j : 1 \le j \le k\}$, we have the confidence interval:

$$\text{CI}_{\text{NI}}^{(SG)}(\alpha) := \left(\widehat{\rho}_{\text{NI}}^{(SG)} - \frac{S_\rho}{\sqrt{k}}z_{1-\alpha/2}, \widehat{\rho}_{\text{NI}}^{(SG)} + \frac{S_\rho}{\sqrt{k}}z_{1-\alpha/2}\right) \tag{8}$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-th quantile of the standard Normal distribution. The following theorem states the results for correlation estimator under non-interactive protocol.

Theorem 2.1. *The following results hold on the estimation error of ρ using a non-interactive componentwise privacy constrained estimator.*

1. *When $(X, Y) \sim \mathcal{N}(\mathbf{0}, \Sigma(\rho))$ with $(\Sigma(\rho))_{11} = (\Sigma(\rho))_{22} = 1$ and $(\Sigma(\rho))_{12} = \rho$, the estimator $\widehat{\rho}_{\text{NI}}^{(G)}$ described in Section 2.1 satisfies $\widehat{\rho}_{\text{NI}}^{(G)} \in \text{NI}(\varepsilon_1, \varepsilon_2, \delta_1, \delta_2)$ and*

$$\mathbb{E}(\widehat{\rho}_{\text{NI}}^{(G)} - \rho)^2 \lesssim \frac{1}{n}\left(\frac{1}{\varepsilon_1^2} + \frac{1}{\varepsilon_2^2}\right) \quad \text{if} \quad m = \left\lfloor \frac{8}{\varepsilon_1\varepsilon_2} \right\rfloor \vee 1.$$

2. *When (X, Y) have mean zero, variance one, X is η_1-sub-Gaussian, Y is η_2-sub-Gaussian, and $\mathbb{E}[XY] = \rho$, the estimator $\widehat{\rho}_{\text{NI}}^{(SG)}$ described in Section 2.2 satisfies $\widehat{\rho}_{\text{NI}}^{(SG)} \in \text{NI}(\varepsilon_1, \varepsilon_2, \delta_1, \delta_2)$ and*

$$\mathbb{E}(\widehat{\rho}_{\text{NI}}^{(SG)} - \rho)^2 \lesssim \frac{\log(n)}{n}\left(\frac{1}{\varepsilon_1^2} + \frac{1}{\varepsilon_2^2}\right) \quad \text{if} \quad m = \left\lfloor \frac{\lambda_1\lambda_2}{\varepsilon_1\varepsilon_2} \right\rfloor \vee 1,$$

   $\lambda_1 = 2\eta_1\sqrt{\log(n)}$, and $\lambda_2 = 2\eta_2\sqrt{\log(n)}$.

3. *For any fixed $\alpha \in (0, 1)$, the confidence intervals defined in (5) and (8) satisfy $\mathbb{P}(\rho \in \text{CI}_{\text{NI}}^{(k)}(\alpha)) \to 1 - \alpha$ as $n \to \infty$, for $k \in \{G, SG\}$.*

3 Interactive estimation methods

We now show that the rates in the previous section can be improved if we allow a one-step interactive scheme between the two servers. To fix ideas, suppose $\varepsilon_1 > \varepsilon_2$, i.e., the privacy requirement in the first server are less stringent than that in the second one. We will then share the private transcripts involving X to the second server containing the Y observations. This leads to an estimation error rate that improves over the non-interactive protocol.

3.1 Interactive correlation estimation for Gaussian distribution

In this case, our interactive estimator based on signs of (X, Y) is as follows. Server 1 first communicates to Server 2 the privatized sign vector $T_1(\mathbf{X})$ with elements:

$$(T_1(\mathbf{X}))_i = \frac{\exp(\varepsilon_1) + 1}{(\exp(\varepsilon_1) - 1)}(2S_i - 1)\,\text{sign}(X_i) \quad \text{for} \quad i = 1, \ldots, n$$

where $S_i \overset{iid}{\sim} \text{Bernoulli}\left(\frac{\exp(\varepsilon_1)}{\exp(\varepsilon_1)+1}\right)$ are independent sign flips introduced by Server 1 to protect the privacy of X_i. Given $T_1(\mathbf{X})$ the second server first computes the covariance

$$\widehat{\eta}_{XY,\text{int}} = \frac{1}{n}\sum_{i=1}^{n}(T_1(\mathbf{X}))_i\,\text{sign}(Y_i)$$

and then outputs the privatized version

$$T_2(\mathbf{Y}, T_1(\mathbf{X})) := \widehat{\eta}_{XY,\text{int}} + Z \quad \text{where } Z \sim \text{Laplace}\left(0, \frac{2(\exp(\varepsilon_1)+1)}{n(\exp(\varepsilon_1)-1)\varepsilon_2}\right). \tag{9}$$

As before we then have the private correlation estimator

$$\widehat{\rho}^{(G)}_{\text{INT}} = \sin\left(\frac{\pi \widehat{\eta}^{(\text{P})}_{XY,\text{int}}}{2}\right).$$

Similar to the non-interactive case, defining $\widehat{\sigma}^2_\eta := 1 - \left(\frac{\exp(\varepsilon_1)-1}{\exp(\varepsilon_1)+1}\right)^2 (\widehat{\eta}^{(P)}_{XY,\text{int}})^2$ allows the confidence interval given by the following.

1. If $c_* = \lim_{n\to\infty} \frac{2}{\sqrt{n}\sigma_\eta \varepsilon_2}$ is finite, then the CI is

$$\left(\widehat{\rho}^{(G)}_{\text{INT}} \mp \frac{\pi \widehat{\sigma}_\eta \sqrt{1-(\widehat{\rho}^{(\text{G})}_{\text{INT}})^2}}{2\sqrt{n}} \left(\frac{\exp(\varepsilon_1)+1}{\exp(\varepsilon_1)-1}\right) F_*^{-1}(1-\alpha/2)\right) \tag{10}$$

   where for any $x \in \mathbb{R}$ we define $F_*(x) := \mathbb{P}(Z_{XY} + \widehat{c}_* Z_{\text{Lap}} \le x)$ for $\widehat{c}_* = 2/(\sqrt{n}\widehat{\sigma}_\eta \varepsilon_2)$ and $Z_{\text{Lap}} \sim \text{Laplace}(0,1)$.

2. If $\frac{1}{\sqrt{n}\varepsilon_2}$ diverges as $n \to \infty$, then the CI is

$$\left(\widehat{\rho}^{(G)}_{\text{INT}} \pm \frac{\pi\sqrt{1-(\widehat{\rho}^{(\text{G})}_{\text{INT}})^2}}{n\varepsilon_2} \left(\frac{\exp(\varepsilon_1)+1}{\exp(\varepsilon_1)-1}\right) \log(\alpha)\right). \tag{11}$$

3.2 Interactive correlation estimation for sub-Gaussian distributions

Following previous sections, Server 1 will send to Server 2 the vector of privatized clipped observations $T_1(\mathbf{X}) \in \mathbb{R}^n$ with elements $(T_1(\mathbf{X}))_i = [X_i]_{\lambda_1} + Z_{1i}$ for a clipping parameter $\lambda_1 > 0$ and $Z_{1i} \overset{iid}{\sim} \text{Laplace}\,(2\lambda_1/\varepsilon_1)$ for $i = 1, \ldots, n$. Then Server 2 can construct

$$\widehat{\rho}^{(SG)}_{\text{INT}} = \frac{1}{n}\sum_{i=1}^{n} [(T_1(\mathbf{X}))_i Y_i]_{\lambda_2} + Z_2.$$

In the above $[x]_t := \text{sign}(x)(|x| \wedge t)$ for any $x \in \mathbb{R}$ and $t > 0$. Here $Z_2 \sim \text{Laplace}\,(2\lambda_2/n\varepsilon_2)$ is Laplace noise added to ensure DP requirements. In addition to $\widehat{\rho}^{(SG)}_{\text{INT}}$, Server 2 also outputs a privatized sample variance S^2_ρ of $[(T_1(\mathbf{X}))_i Y_i]_{\lambda_2}$ for $i = 1, \ldots, n$. Then we have the confidence interval constructed as follows:

1. If $c_* = \lim_{n\to\infty} \frac{2\lambda_2}{\sqrt{n}\sigma_\rho \varepsilon_2}$ is finite, then the CI is

$$\left(\widehat{\rho}^{(SG)}_{\text{INT}} - \frac{S_\rho}{\sqrt{n}} F_*^{-1}(1-\alpha/2), \widehat{\rho}^{(\text{SG})}_{\text{int}} + \frac{S_\rho}{\sqrt{n}} F_*^{-1}(1-\alpha/2)\right) \tag{12}$$

   where for any $x \in \mathbb{R}$ we define $F_*(x) := \mathbb{P}(Z_{XY} + \widehat{c}_* Z_{\text{Lap}} \le x)$ for $\widehat{c}_* = 2\lambda_2/(\sqrt{n} S_\rho \varepsilon_2)$, and $Z_{\text{Lap}} \sim \text{Laplace}(0,1)$.

2. If $\frac{\lambda_2}{\sqrt{n}\varepsilon_2}$ diverges as $n \to \infty$, then the CI is

$$\left(\widehat{\rho}^{(SG)}_{\text{INT}} + \frac{\lambda_2}{n\varepsilon_2}\log(\alpha), \widehat{\rho}^{(\text{SG})}_{\text{int}} - \frac{\lambda_2}{n\varepsilon_2}\log(\alpha)\right). \tag{13}$$

The following theorem states the results for correlation estimator under the interactive protocol.

Theorem 3.1. *The following results hold on the estimation error of ρ using the above privacy constrained interactive estimator.*

1. *When $(X,Y) \sim \mathcal{N}(\mathbf{0}, \Sigma(\rho))$ with $(\Sigma(\rho))_{11} = (\Sigma(\rho))_{22} = 1$ and $(\Sigma(\rho))_{12} = \rho$, the estimator $\widehat{\rho}^{(G)}_{\text{INT}}$ described in Section 3.1 satisfies $\widehat{\rho}^{(G)}_{\text{INT}} \in \text{INT}(\varepsilon_1, \varepsilon_2, \delta_1, \delta_2)$ and*

$$\mathbb{E}(\widehat{\rho}^{(G)}_{\text{INT}} - \rho)^2 \lesssim \frac{1}{n(\varepsilon_1 \vee \varepsilon_2)^2} + \frac{1}{n^2\varepsilon_1^2\varepsilon_2^2}.$$

2. *When (X, Y) have mean zero, variance one, X is η_1-sub-Gaussian, Y is η_2-sub-Gaussian, and $\mathbb{E}[XY] = \rho$, the estimator $\widehat{\rho}_{\mathrm{INT}}^{(SG)}$ described in Section 3.2 satisfies $\widehat{\rho}_{\mathrm{INT}}^{(SG)} \in \mathrm{INT}(\varepsilon_1, \varepsilon_2, \delta_1, \delta_2)$ and*

$$\mathbb{E}(\widehat{\rho}_{\mathrm{INT}}^{(SG)} - \rho)^2 \lesssim \frac{1}{n(\varepsilon_1 \vee \varepsilon_2)^2} + \frac{1}{n^2 \varepsilon_1^2 \varepsilon_2^2}$$

if $\lambda_1 = 2\eta_1\sqrt{\log(n)}$ and $\lambda_2 = 4(\eta_2 \vee 1)(\log(n))^2/(\varepsilon_1 \wedge 1)$.

3. *For any fixed $\alpha \in (0, 1)$, under their respective assumptions, the confidence intervals defined in (10), (11), (12), and (13) satisfy $\mathbb{P}(\rho \in \mathrm{CI}_{\mathrm{INT}}^{(k)}(\alpha)) \to 1 - \alpha$ as $n \to \infty$, for $k \in \{G, SG\}$.*

4 Minimax lower bounds

In this section, we show that the private correlation estimators derived in the previous section are in fact minimax optimal in many cases. Our proof strategy is based on bounding Fisher information of the private transcripts and then using Van Trees inequality. We recall some standard results from parameter estimation theory in the next subsection.

4.1 Fisher information and Van Trees inequality

Let θ be a real-valued parameter taking an unknown value in some interval $[a, b]$. We observe some random variable (or vector) X with distribution $P(x|\theta)$ parameterized by θ.

Assume that $P(\cdot|\theta)$ is absolutely continuous with respect to a reference measure μ, for each $\theta \in [a, b]$, and $\frac{dP(\cdot|\theta)}{d\mu}(x)$ is differentiable with respect to $\theta \in (a, b)$ for μ-almost all x. Then the *Fisher information* of θ w.r.t. X, denoted as $I_F(X; \theta)$, is defined as

$$I_F(X; \theta) \triangleq \int \left(\frac{\partial}{\partial \theta} \ln \frac{dP(\cdot|\theta)}{d\mu}(x) \right)^2 dP(x|\theta). \tag{14}$$

The following inequality is well-known. See for example Gill and Levit [1995].

Lemma 1 (Van Trees inequality)**.** *Let θ be a real parameter with prior density ζ supported on $[a, b] \subset \mathbb{R}$, and let $X \sim P(\cdot \mid \theta)$ with conditional density $p(x \mid \theta) = \frac{dP(\cdot|\theta)}{d\mu}(x)$. Under some regularity conditions we have that for every estimator $\widehat{\theta} = \widehat{\theta}(X)$ with $\mathbb{E}[(\widehat{\theta} - \theta)^2] < \infty$ under the joint law of (X, θ) satisfies*

$$\mathbb{E}\big[(\widehat{\theta} - \theta)^2\big] \;\geq\; \frac{1}{\mathbb{E}_\theta[I_F(X;\theta)] \;+\; I_F(\zeta)}, \qquad \mathbb{E}_\theta[I_F(X;\theta)] = \int_a^b I_F(X;\theta)\, \zeta(\theta)\, d\theta, \tag{15}$$

where $I_F(\zeta) := \int_a^b \frac{\big(\zeta'(\theta)\big)^2}{\zeta(\theta)}\, d\theta$ is the prior Fisher information.

The "regularity conditions" in Lemma 1 are to ensure that one can apply the dominated convergence theorem to exchange certain integrals and differentiations in the calculus. See for example Vaart [1998]. Additionally, assume that

$$I_F(X; \theta + \epsilon) = I_F(X; \theta)(1 + \eta(\epsilon)) \tag{16}$$

where $\eta(\epsilon) < C_\eta$ for all $|\epsilon| < c_0$ for some numerical constants $c_0 < 1$ and $C_\eta > 0$.

4.2 Non interactive

For the non-interactive protocols the servers output transcripts T_1 and T_2 which are $(\varepsilon_1, \delta_1)$ and $(\varepsilon_2, \delta_2)$-DP respectively. The transcripts are based on;y on the data from their own servers. An estimator $\widehat{\rho}$ is then calculated after combining T_1 and T_2.

Our lower bound is shown by the difficulty of correlation estimation when $\rho = 0$. Let us denote the transcripts by $T \equiv (T_1, T_2)$. As a first step, the next lemma shows that $I_F(T; 0)$ is smaller than a quantity involving the sample size n and the privacy parameters $\varepsilon_1, \varepsilon_2$.

Lemma 2. *Assume that for $k = 1, 2$, $\delta_k \log(1/\delta_k) = O(\varepsilon_k^2)$. Let us denote the Fisher information for the transcripts T by $I_F(T; \rho)$. We have that*

$$I_F(T; 0) \leq \frac{8}{\pi}(n\varepsilon_1^2 \wedge n\varepsilon_2^2).$$

The local regularity assumption in (16) at $\rho = 0$ ensures that up to a constant factor, the bound from the above lemma carries over to $I_F(T;\rho)$ for $|\rho| \le c_0$. For a suitable choice of prior density ζ this in turn implies an upper bound on $\mathbb{E}_0[I_F(T;0)]$ and allows us to complete the proof by using Van-Trees inequality (Lemma 1). We then have the following lower bound on the minimax risk for estimating ρ in the non-interactive setting.

Theorem 4.1. *Assume that $\delta_k = o(n^{-1-\omega})$ for $k = 1, 2$ and $n(\varepsilon_1^2 \wedge \varepsilon_2^2) \to \infty$. Then for non interactive protocols the minimax rate is lower bounded by*

$$\inf_{\widehat{\rho} \in \mathrm{NI}(\varepsilon_1,\varepsilon_2,\delta_1,\delta_2)} \sup_{\rho \in [-1,1]} |\widehat{\rho} - \rho|^2 \gtrsim \frac{1}{n} + \frac{1}{n\varepsilon_1^2} + \frac{1}{n\varepsilon_2^2}.$$

Remark 4.1. *The assumption $n(\varepsilon_1^2 \wedge \varepsilon_2^2) \to \infty$ assumes that the minimax rate is going to zero ensuring consistent estimation of ρ in the first place.*

4.3 Interactive

We next allow one way interaction among the servers where either of the server can share its transcripts with the other server. Let us denote the set of protocols which allow allow interaction from server 1 to 2 as $\Pi_{1\to 2}$, i.e server 2 gets to observe the transcript T_1, before computing T_2. We first show the following upper bound on $I_F(\Pi_{1\to 2}; 0)$.

Lemma 3. *Assume that $\delta_1 \log(1/\delta_1) = o(\varepsilon_1^2)$ and $\delta_2 \log(1/\delta_2)^2 = o(n\varepsilon_1^2\varepsilon_2^2)$. Let us denote the Fisher information for the transcripts $\Pi_{1\to 2}$ by $I_F(\Pi_{1\to 2};\rho)$. We have that*

$$I_F(\Pi_{1\to 2}; 0) \le n\varepsilon_1^2 \wedge n^2\varepsilon_1^2\varepsilon_2^2.$$

If we denote the protocol which allow interaction from server 2 to 1 we can show that $I_F(\Pi_{2\to 1}; 0) \le n\varepsilon_2^2 \wedge n^2\varepsilon_1^2\varepsilon_2^2$. Since we allow for either of the protocols $\Pi \equiv (\Pi_{1\to 2}, \Pi_{2\to 1})$ we have that

$$I_F(\Pi; 0) \le I_F(\Pi_{1\to 2}; 0) \vee I_F(\Pi_{2\to 1}; 0) \le (n\varepsilon_1^2 \wedge n^2\varepsilon_1^2\varepsilon_2^2) \vee (n\varepsilon_2^2 \wedge n^2\varepsilon_1^2\varepsilon_2^2). \tag{17}$$

Similar to the non-interactive case we can then use the local regularity assumption in (16) and a suitable prior density ζ with Van Trees inequality, leading to the following lower bound on the minimax risk in the interactive setting.

Theorem 4.2. *Assume that for $k = 1, 2$ $\delta_k = o(n^{-1-\omega})$ for $\omega > 0$, $n(\varepsilon_1^2 \vee \varepsilon_2^2) \to \infty$ and $n^2\varepsilon_1^2\varepsilon_2^2 \to \infty$. Then for interactive protocols the minimax rate is lower bounded by*

$$\inf_{\widehat{\rho} \in \mathrm{INT}(\varepsilon_1,\varepsilon_2,\delta_1,\delta_2)} \sup_{\rho \in [-1,1]} |\widehat{\rho} - \rho|^2 \gtrsim \frac{1}{n} + \frac{1}{n(\varepsilon_1^2 \vee \varepsilon_2^2)} + \frac{1}{n^2\varepsilon_1^2\varepsilon_2^2}.$$

Remark 4.2. *The assumption $n(\varepsilon_1^2 \vee \varepsilon_2^2) \to \infty$ and $n^2\varepsilon_1^2\varepsilon_2^2 \to \infty$ assumes that the minimax rate is going to zero, ensuring consistent estimation of ρ in the first place.*

5 Simulation study

We evaluate our **non–interactive sign–batch** (NI) and **interactive sign–flip** (INT) estimators across different parameter settings. All our codes can be found at https://github.com/abhinavc3/distributed-correlation.

5.1 Simulation Results

In our experiments we write non-normalized to mean that the mean and variances of the marginal distributions are known, and normalized to mean that they are unknown and estimated. We use two generative models.

- **Gaussian:** $(X, Y) \sim \mathcal{N}(\mu, 2\Sigma(\rho))$ with $\mu = (0.5, 0.5)^\top$, and $\Sigma(\rho)$ given by $\mathrm{Var}(X) = \mathrm{Var}(Y) = 1$ and $\mathrm{Corr}(X, Y) = \rho$. We run each estimator *with* and *without* the private normalization step.
- **Bounded-factor (sub-Gaussian):** $X = U + E_1$, $Y = U + E_2$ with $U \sim \mathrm{Unif}\big[-\sqrt{3\,\rho}, \sqrt{3\,\rho}\big]$ and $E_i \sim \mathrm{Unif}\big[-\sqrt{3(1-\rho)}, \sqrt{3(1-\rho)}\big]$, so each marginal is centred, variance–one, and bounded hence sub–Gaussian.

For every design point we record mean–squared error (MSE), average confidence–interval (CI) length, empirical coverage ($1-\alpha = 0.95$) and the mean CI offset band $\mathbb{E}[\mathrm{CI}_L - \rho] \to \mathbb{E}[\mathrm{CI}_U - \rho]$ where CI_L and CI_U are the upper and lower confidence bars. In practice it is sufficient to use the confidence intervals from (10) and (12) since (11) and (13) are respectively the limiting versions of the above two.

Parameter Grid. We vary our parameters as below, with 250 replications for each cell:

- Sample size: $n \in \{1000, 1500, 2500, 4000, 6000, 9000\}$.
- Correlation: $\rho \in \{0, 0.15, 0.3, 0.4, 0.5, 0.65, 0.8, 0.9\}$.
- Privacy budget: $(\varepsilon_1, \varepsilon_2) \in \{(0.5, 0.5), (1, 1), (1.5, 0.5)\}$.

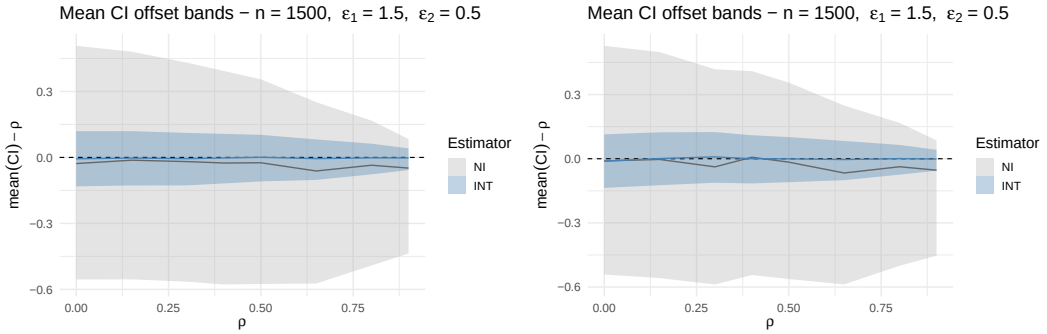

Figure 1: **Gaussian,** $n = 1500$**,** $(\varepsilon_1, \varepsilon_2) = (1.5, 0.5)$**.** Mean CI–offset bands for NI (grey) and INT (blue). Left: *without* normalization. Right: *with* private normalization. Curves overlap.

Figure 1 compares the mean CI offset bands for $n = 1500$ and the budget $(\varepsilon_1, \varepsilon_2) = (1.5, 0.5)$. With and without normalization the ribbons coincide, indicating that *private normalization is cost–free*. Figure 2 shows CI width and coverage versus n at $\rho = 0.5$; both variants adhere to the nominal 95% band. Figure 3 confirms that INT is uniformly more efficient than NI, while normalization leaves MSE unchanged (largest relative difference $< 2\%$).

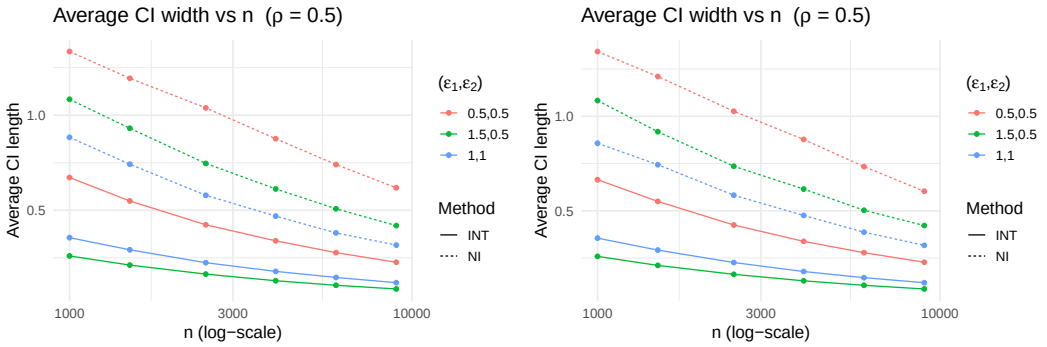

Figure 2: **Gaussian,** $\rho = 0.5$**.** Average CI length versus n. Left: *no* normalization; right: *with* normalization. Normalization has no discernible effect; INT yields shorter CIs. The coverage probabilities are above 0.91 for all CIs.

We repeat the study with the bounded-factor DGP. The qualitative picture is the same: INT enjoys narrower CIs and lower MSE , and both estimators achieve nominal coverage. For the sake of brevity we only show the MSE plots (Figure 3 right). The CI bands, coverage and width plots are deferred to the supplementary material.

5.2 Real Data Experiments

We illustrate our methods using data from the *Health and Retirement Study (HRS)*, a longitudinal survey of older adults in the United States. We focus on two variables—age and body mass index (BMI)—from Wave 2 (year 1993-94) corresponding to around 20k individuals. In this demographic, age and BMI are known to exhibit a mild negative correlation.

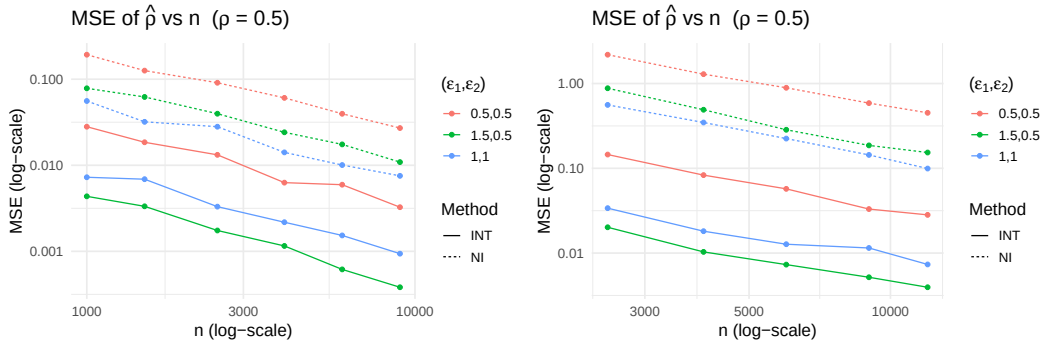

Figure 3: **Gaussian** (left) and **Bounded Factor** (right) MSE, $\rho = 0.5$. MSE versus n (log–log).

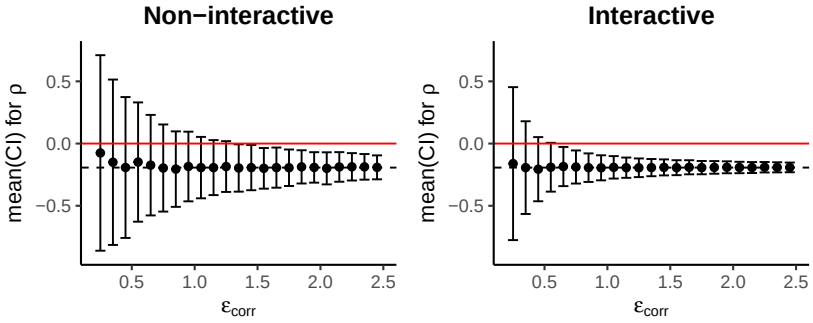

Figure 4: Mean confidence interval bands for non-interactive (left) and interactive (right) methods for estimating the correlation between age and BMI in the *Health and Retirement Study (HRS)* data. The black dotted line indicates the non-private estimator.

We consider a distributed scenario in which the two variables reside on separate servers, and the goal is to estimate their Pearson correlation coefficient ρ. Each server first applies a Central differentially private (CDP) normalization so that the privatized features have approximately zero mean and unit variance. Specifically, we allocate $\varepsilon = 0.1$ for each of the mean and standard deviation estimates. The clipping bounds are chosen based on domain knowledge—[45, 90] for age and [15, 35] for BMI—demonstrating a setting where the privacy mechanism leverages prior information rather than data-dependent thresholds.

After normalization, we apply both the non-interactive (NI) and interactive (INT) protocols to obtain private confidence intervals for the estimated correlation $\widehat{\rho}$. We compare these to the non-private benchmark while varying the privacy budget $\varepsilon_{\text{corr}}$, keeping it equal across the two servers. Results are given in Figure 4. As $\varepsilon_{\text{corr}}$ increases, the private intervals contract and concentrate around the non-private ρ. Moreover, for a fixed $\varepsilon_{\text{corr}}$, the INT intervals are consistently shorter than their NI counterparts. Notably, at $\varepsilon_{\text{corr}} = 1$, the interactive CI excludes zero while the non-interactive CI includes it—illustrating that privacy noise can increase uncertainty and, in some cases, prevent rejection of the null hypothesis $\rho = 0$.

6 Discussion

Across both distributions and all privacy budgets explored, INT consistently outperforms NI, while the required private normalisation step incurs *no measurable loss* in bias, MSE or interval width. These findings support the theoretical claim that normalization's privacy cost is dominated by the subsequent correlation release.

We discuss two important directions of future work. First, allowing multiple features per server—rather than a single feature—introduces new challenges, particularly in handling inter-feature correlations and maintaining privacy in higher dimensions. Second, extending our methods to heavy-tailed distributions would broaden applicability, as such data often arise in practice and require more robust estimation techniques.

Acknowledgements

The research was supported in part by NSF grant NSF DMS-2413106 and NIH grants R01-GM123056 and R01-GM129781.

References

John M Abowd, Ian M Rodriguez, William N Sexton, Phyllis E Singer, and Lars Vilhuber. The modernization of statistical disclosure limitation at the us census bureau. *US Census Bureau*, 2020.

Chiara Amorino and Arnaud Gloter. Minimax rate for multivariate data under componentwise local differential privacy constraints. *arXiv preprint arXiv:2305.10416*, 2023.

Arnab Auddy, T Tony Cai, and Abhinav Chakraborty. Minimax and adaptive transfer learning for nonparametric classification under distributed differential privacy constraints. *arXiv preprint arXiv:2406.20088*, 2024.

Raef Bassily, Adam Smith, and Abhradeep Thakurta. Private empirical risk minimization: Efficient algorithms and tight error bounds. In *2014 IEEE 55th annual symposium on foundations of computer science*, pages 464–473. IEEE, 2014.

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.

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

T Tony Cai, Abhinav Chakraborty, and Lasse Vuursteen. Federated nonparametric hypothesis testing with differential privacy constraints: Optimal rates and adaptive tests. *arXiv preprint arXiv:2406.06749*, 2024a.

T Tony Cai, Abhinav Chakraborty, and Lasse Vuursteen. Optimal federated learning for nonparametric regression with heterogeneous distributed differential privacy constraints. *arXiv preprint arXiv:2406.06755*, 2024b.

Tony Cai, Abhinav Chakraborty, and Lasse Vuursteen. Optimal federated learning for functional mean estimation under heterogeneous privacy constraints. *arXiv preprint arXiv:2412.18992*, 2024c.

Tianyi Chen, Xiao Jin, Yuejiao Sun, and Wotao Yin. Vafl: a method of vertical asynchronous federated learning. *arXiv preprint arXiv:2007.06081*, 2020.

Cynthia Dwork. Differential privacy. In *International colloquium on automata, languages, and programming*, pages 1–12. Springer, 2006.

Richard D Gill and Boris Y Levit. Applications of the van trees inequality: a bayesian cramér-rao bound. *Bernoulli*, pages 59–79, 1995.

Uri Hadar, Jingbo Liu, Yury Polyanskiy, and Ofer Shayevitz. Communication complexity of estimating correlations. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 792–803, 2019.

Peter Kairouz, H Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Kallista Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, et al. Advances and open problems in federated learning. *Foundations and trends® in machine learning*, 14(1–2):1–210, 2021.

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

Li Li, Yuxi Fan, Mike Tse, and Kuo-Yi Lin. A review of applications in federated learning. *Computers & Industrial Engineering*, 149:106854, 2020a.

Tian Li, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith. Federated learning: Challenges, methods, and future directions. *IEEE signal processing magazine*, 37(3):50–60, 2020b.

Yang Liu, Yan Kang, Tianyuan Zou, Yanhong Pu, Yuanqin He, Xiaozhou Ye, Ye Ouyang, Ya-Qin Zhang, and Qiang Yang. Vertical federated learning: Concepts, advances, and challenges. *IEEE Transactions on Knowledge and Data Engineering*, 36(7):3615–3634, 2024.

Sheng Shen, Tianqing Zhu, Di Wu, Wei Wang, and Wanlei Zhou. From distributed machine learning to federated learning: In the view of data privacy and security. *Concurrency and Computation: Practice and Experience*, 34(16):e6002, 2022.

Alexandre B. Tsybakov. *Introduction to nonparametric estimation*. Springer series in statistics. Springer, New York ; London, 2009. ISBN 978-0-387-79051-0 978-0-387-79052-7. OCLC: ocn300399286.

A. W. van der Vaart. *Asymptotic statistics*. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, Cambridge, UK ; New York, NY, USA, 1998. ISBN 978-0-521-49603-2.

Kang Wei, Jun Li, Ming Ding, Chuan Ma, Howard H Yang, Farhad Farokhi, Shi Jin, Tony QS Quek, and H Vincent Poor. Federated learning with differential privacy: Algorithms and performance analysis. *IEEE transactions on information forensics and security*, 15:3454–3469, 2020.

Kang Wei, Jun Li, Ming Ding, Chuan Ma, Hang Su, Bo Zhang, and H Vincent Poor. User-level privacy-preserving federated learning: Analysis and performance optimization. *IEEE Transactions on Mobile Computing*, 21(9):3388–3401, 2021.

Kang Wei, Jun Li, Chuan Ma, Ming Ding, Sha Wei, Fan Wu, Guihai Chen, and Thilina Ranbaduge. Vertical federated learning: Challenges, methodologies and experiments. *arXiv preprint arXiv:2202.04309*, 2022.

Yuncheng Wu, Shaofeng Cai, Xiaokui Xiao, Gang Chen, and Beng Chin Ooi. Privacy preserving vertical federated learning for tree-based models. *arXiv preprint arXiv:2008.06170*, 2020.

Shengwen Yang, Bing Ren, Xuhui Zhou, and Liping Liu. Parallel distributed logistic regression for vertical federated learning without third-party coordinator. *arXiv preprint arXiv:1911.09824*, 2019.

Chen Zhang, Yu Xie, Hang Bai, Bin Yu, Weihong Li, and Yuan Gao. A survey on federated learning. *Knowledge-Based Systems*, 216:106775, 2021.

NeurIPS Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction clearly articulate the main contributions of the paper, including the problem setup, the proposed methodology, and the theoretical guarantees. They accurately reflect the scope of the work and are consistent with the results presented in both the theoretical analysis and the simulation study. Any assumptions and limitations are also stated appropriately, ensuring that the claims are well-aligned with the actual contributions of the paper.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Limitations and future directions are described in the Discussion section.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly

when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.

- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: The paper provides a complete and rigorous treatment of each theoretical result, with all necessary assumptions clearly stated alongside the corresponding theorems. Full proofs are included in the supplemental material, and the main paper provides intuitive explanations to aid understanding. All theorems and lemmas are properly numbered, referenced, and supported by either original arguments or citations to well-established results, ensuring the theoretical contributions are transparent and verifiable.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We have provided all experiment details in Section 5 and codes in an anonymized code repository.

   Guidelines:

   - The answer NA means that the paper does not include experiments.

- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: We have provided all codes in an anonymized code repository.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.

- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: All details are given in Section 5 and the anonymous code repository.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: We have reported 95% confidence intervals with all our estimates.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.

- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: All experiments were done on a desktop with 32 GB RAM, and were done over the course of 1 hour.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: There are no violations of the code of ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification: There is no societal impact of the work performed.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that

unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [NA]

    Justification: The paper does not use existing assets.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.

- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: The paper does not release new assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: This research has not used LLM as an important, original, or non-standard component of the core methods.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (https://neurips.cc/Conferences/2025/LLM) for what should or should not be described.

A Implementation details

Since the mean and variance of each server can be computed under the central differential privacy (CDP) framework, we adopt estimators similar to those proposed in Karwa and Vadhan [2017] for our simulation study. After obtaining these estimators, we standardize the data and use the resulting values for downstream analysis.

Additionally, to improve the stability of our estimators, we incorporate intermediate clipping steps in our simulation study. For example, in the Gaussian case, we clip the mean of the signs to the interval $[-1, 1]$ before applying the *sin* transformation. In the sub-Gaussian case, we clip the final estimator to $[-1, 1]$.

A.1 Additional simulation study

Here we collect the additional plots and results pertaining to our simulation study.

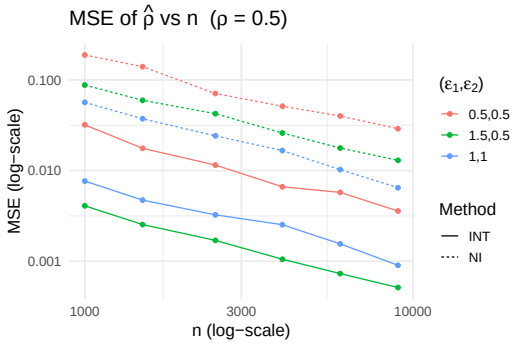

Figure 5: **Gaussian Normalized MSE, $\rho = 0.5$.** MSE versus n (log–log).

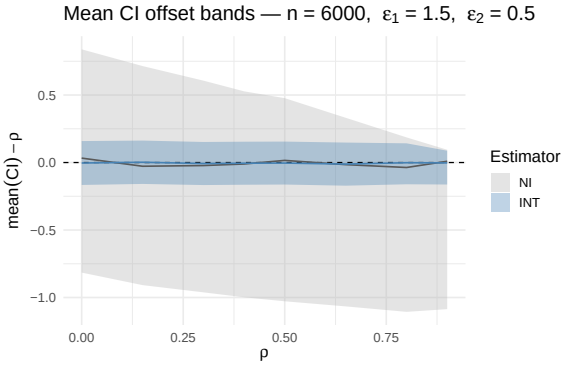

Figure 6: **Bounded-factor, $n = 6000$, $(\varepsilon_1, \varepsilon_2) = (1.5, 0.5)$.** Mean CI offset bands for NI and INT.

B Proofs

Proof of Theorem 1.1. The proof of this theorem follows from parts 1 and 2 of Theorem 2.1 and Theorem 4.1. □

Proof of Theorem 1.2. The proof of this theorem follows from parts 1 and 2 of Theorem 3.1 and Theorem 4.2. □

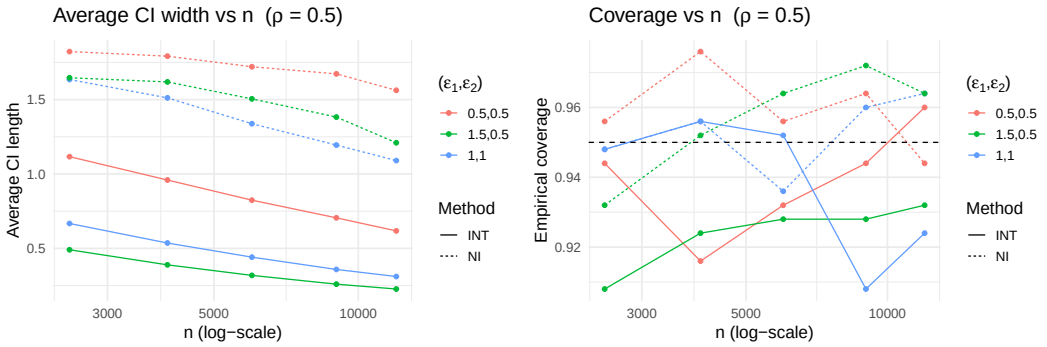

Figure 7: **Bounded-factor, $\rho = 0.5$.** CI length (left) and coverage (right) versus n.

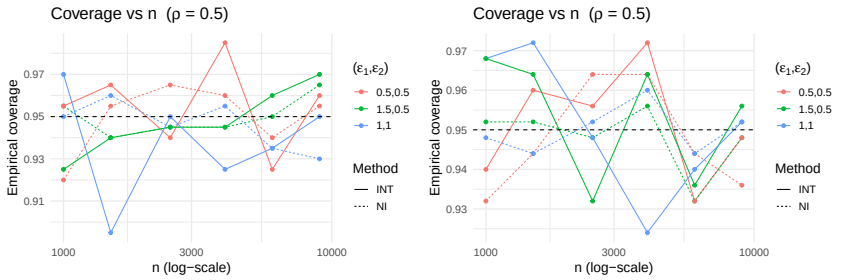

Figure 8: **Gaussian, $\rho = 0.5$.** Empirical coverage versus n. Left: *no* normalisation; right: *with* normalisation. Normalisation has no discernible effect; INT yields shorter CIs. The coverage probabilities are above 0.91 for all CIs.

B.1 Proofs of upper bound results

Proof of Theorem 2.1. We prove the two statements in the theorem separately.

1. It is straightforward to check that

$$\mathbb{E}[\widehat{\eta}_{XY}] = m\mathbb{E}[\bar{X}^{(1)}(\bar{Y}^{(1)})] = m \cdot \frac{1}{m^2}\sum_{i=1}^{m}\mathbb{E}[\operatorname{sign}(X_i)\operatorname{sign}(Y_i)]$$
$$= 2\mathbb{P}(XY > 0) - 1 = \frac{2\arccos(-\rho)}{\pi} - 1 = 1 - \frac{2\arccos(\rho)}{\pi}.$$

To bound the error in estimating ρ by $\widehat{\rho}^{(P)}$ we therefore note that

$$\left|\widehat{\rho}^{(G)}_{\mathrm{NI}} - \rho\right| = \left|\sin\left(\frac{\pi\widehat{\eta}_{XY}}{2}\right) - \sin\left(\frac{\pi\mathbb{E}[\widehat{\eta}_{XY}]}{2}\right)\right|$$
$$= \frac{\cos(\xi)\cdot\pi|\widehat{\eta}_{XY} - \mathbb{E}[\widehat{\eta}_{XY}]|}{2} \le \frac{\pi}{2}|\widehat{\eta}^{(P)}_{XY} - \mathbb{E}[\widehat{\eta}_{XY}]|. \tag{18}$$

where $\xi = t\widehat{\eta}_{XY} + (1-t)\mathbb{E}[\widehat{\eta}_{XY}]$ for some $t \in [0,1]$. Thus,

$$\mathbb{E}|\widehat{\rho}^{(G)}_{\mathrm{NI}} - \rho|^2 \le \frac{\pi^2}{4}\mathbb{E}|\widehat{\eta}_{XY} - \mathbb{E}[\widehat{\eta}_{XY}]|^2 = \frac{\pi^2}{4}\operatorname{Var}(\widehat{\eta}_{XY} - \mathbb{E}[\widehat{\eta}_{XY}])$$
$$\le \frac{\pi^2 m^2}{4k}\left[\mathbb{E}(\bar{X}^{(1)} + Z_1^{(1)})^4\right]^{1/2}\left[\mathbb{E}(\bar{Y}^{(1)} + Z_2^{(1)})^4\right]^{1/2}$$
$$\le \frac{\pi^2 m^2}{4k}\left(\frac{3}{m} + \frac{8}{m^2\varepsilon_1^2}\right)\left(\frac{3}{m} + \frac{8}{m^2\varepsilon_2^2}\right)$$
$$= \frac{\pi^2}{4}\left(\frac{9m}{n} + \frac{24}{n\varepsilon_1^2} + \frac{24}{n\varepsilon_1^2} + \frac{64}{mn\varepsilon_1^2\varepsilon_2^2}\right) \tag{19}$$
$$= \frac{\pi^2}{4}\left(\frac{24}{n\varepsilon_1^2} + \frac{24}{n\varepsilon_1^2} + \frac{80}{n\varepsilon_1\varepsilon_2}\right) \le \frac{10\pi^2}{n}\left(\frac{1}{\varepsilon_1} + \frac{1}{\varepsilon_2}\right)^2.$$

In the penultimate equality, we use the choice $m = \frac{8}{\varepsilon_1\varepsilon_2}$, which minimizes the expression in the previous line. The privacy constraints are satisfied by the Laplace mechanism and checking the sensitivity of the batch means.

2. It is straightforward to check that

$$\begin{aligned}\mathbb{E}[\widehat{\rho}_{\mathrm{NI}}^{(SG)}] &= \mathbb{E}[XY\mathbb{1}(|X| \le \lambda_1, |Y| \le \lambda_2)] \\ &= \rho - \mathbb{E}[XY\mathbb{1}(|X| > \lambda_1 \text{ or } |Y| > \lambda_2)].\end{aligned}$$

We can thus bound the bias of the estimator $\widehat{\rho}_{\mathrm{NI}}^{(SG)}$ as:

$$\begin{aligned}\left|\mathbb{E}[\widehat{\rho}_{\mathrm{NI}}^{(SG)}] - \rho\right| &\le \mathbb{E}[|XY|\mathbb{1}(|X| > \lambda_1 \text{ or } |Y| > \lambda_2)] \\ &\le \left(\mathbb{E}|X|^3\right)^{\frac{1}{3}} \left(\mathbb{E}|Y|^3\right)^{\frac{1}{3}} \left(\mathbb{P}(|X| > \lambda_1) + \mathbb{P}(|Y| > \lambda_2)\right)^{\frac{1}{3}} \\ &\lesssim \exp\left(-\frac{1}{3}\left\{\frac{\lambda_1^2}{\eta_1^2} \wedge \frac{\lambda_2^2}{\eta_2^2}\right\}\right) \qquad (20)\end{aligned}$$

where we use the fact that X is η_1-subgaussian and Y is η_2-subgaussian. At the same time,

$$\begin{aligned}\mathrm{Var}(\widehat{\rho}_{\mathrm{NI}}^{(SG)} &\le \frac{m^2}{k}\left[\mathbb{E}(\bar{X}^{(1)} + Z_1^{(1)})^4\right]^{1/2}\left[\mathbb{E}(\bar{Y}^{(1)} + Z_2^{(1)})^4\right]^{1/2} \\ &\lesssim \frac{m^2}{k}\left(\frac{1}{m} + \frac{\lambda_1^2}{m^2\varepsilon_1^2}\right)\left(\frac{1}{m} + \frac{\lambda_2^2}{m^2\varepsilon_2^2}\right) \lesssim \frac{1}{n}\left(\frac{\lambda_1}{\varepsilon_1} + \frac{\lambda_2}{\varepsilon_2}\right)^2\end{aligned}$$

where in the last step we use the choice of $m = \frac{\lambda_1\lambda_2}{\varepsilon_1\varepsilon_2}$, which minimizes the expression in the previous step. Thus the MSE of $\widehat{\rho}_{\lambda}^{(P)}$ in estimating ρ is given by:

$$\begin{aligned}\mathbb{E}|\widehat{\rho}_{\mathrm{NI}}^{(SG)} - \rho|^2 &= \left(\mathbb{E}[\widehat{\rho}_{\mathrm{NI}}^{(SG)}] - \rho\right)^2 + \mathrm{Var}(\widehat{\rho}_{\mathrm{NI}}^{(SG)}) \\ &\lesssim \exp\left(-\frac{2}{3}\left\{\frac{\lambda_1^2}{\eta_1^2} \wedge \frac{\lambda_2^2}{\eta_2^2}\right\}\right) + \frac{1}{n}\left(\frac{\lambda_1}{\varepsilon_1} + \frac{\lambda_2}{\varepsilon_2}\right)^2.\end{aligned}$$

We now choose $\lambda_1 = 2\eta_1\sqrt{\log(n)}$ and $\lambda_2 = 2\eta_2\sqrt{\log(n)}$ for some $\kappa > 0$. The bias bound from (20) then becomes:

$$\left|\mathbb{E}[\widehat{\rho}_{\mathrm{NI}}^{(SG)}] - \rho\right| \le \frac{1}{n} \qquad (21)$$

leading to the MSE bound

$$\mathbb{E}|\widehat{\rho}_{\mathrm{NI}}^{(SG)} - \rho|^2 \lesssim \exp\left(-2\log(n)\right) + \frac{\log(n)}{n\varepsilon_1^2} + \frac{\log(n)}{n\varepsilon_2^2} \lesssim \frac{\log(n)}{n}\left(\frac{\eta_1}{\varepsilon_1^2} + \frac{\eta_2}{\varepsilon_2^2}\right).$$

Once again the privacy constraints are satisfied by the Laplace mechanism and checking the sensitivity of the batch means.

3. We split the proofs for confidence interval coverage into the Gaussian and sub-Gaussian cases.

   (a) (Gaussian case) Note that $\widehat{\eta}_{XY}$ in (3) is an average of k iid observations T_j defined as follows:

$$\widehat{\eta}_{XY} = \frac{1}{k}\sum_{j=1}^{k} T_j \quad \text{where } T_j := m(\bar{X}^{(j)} + Z_1^{(j)})(\bar{Y}^{(j)} + Z_2^{(j)})$$

   and

$$\begin{aligned}\sigma_\eta^2 &:= \mathrm{Var}(T_j) \\ &= m^2\mathbb{E}[(\bar{X}^{(j)})^2(\bar{Y}^{(j)} + Z_2^{(j)})^2] + \frac{8}{\varepsilon_1^2}\mathrm{Var}(\bar{Y}^{(j)} + Z_2^{(j)}) - (\mathbb{E}[\widehat{\eta}_{XY}])^2 \\ &= m^2\mathbb{E}[(\bar{X}^{(j)})^2(\bar{Y}^{(j)})^2] - (\mathbb{E}[\widehat{\eta}_{XY}])^2 + m\,\mathrm{Var}(Z_2^{(j)}) + \frac{8}{\varepsilon_1^2}\left(\frac{1}{m} + \frac{8}{m^2\varepsilon_2^2}\right) \\ &= \left(\frac{m-1}{m}\right)^2[1 + (\mathbb{E}[\widehat{\eta}_{XY}])^2] + \frac{1}{m} + \frac{8}{m}\left(\frac{1}{\varepsilon_1^2} + \frac{1}{\varepsilon_2^2}\right) + \frac{64}{m^2\varepsilon_1^2\varepsilon_2^2}.\end{aligned}$$

where the last equality follows by expanding the squares of iid averages in the first term. Thus we have

$$\frac{\sqrt{k}(\widehat{\eta}_{XY} - \mathbb{E}(\widehat{\eta}_{XY}))}{\sigma_\eta} \xrightarrow{d} N(0,1) \quad \text{as } k \to \infty,$$

and thus by delta method, $\widehat{\rho}_{\mathrm{NI}}^{(G)} = \sin(\pi \widehat{\eta}_{XY}^{(P)}/2)$ satisfies:

$$\frac{\sqrt{k}(\widehat{\rho}_{\mathrm{NI}}^{(G)} - \rho)}{(\pi/2)\sigma_\eta \sqrt{1-\rho^2}} \xrightarrow{d} N(0,1) \quad \text{as } k \to \infty.$$

Here we used the fact that $\sin(\pi \mathbb{E}[\widehat{\eta}_{XY}]/2) = \rho$. To estimate σ_η^2 we use the sample variance of T_j:

$$S_\eta^2 := \frac{1}{k}\sum_{j=1}^{k} (T_j - \bar{T})^2.$$

Note that S_η^2 is constructed from $(\varepsilon_1, \varepsilon_2)$-DP statistics T_j, and thus S_η^2 is also differentially private. By standard calculations,

$$\mathbb{E}(S_\eta^2 - \sigma_\eta^2)^2 = O\left(\frac{1}{k}\right)$$

where we use our choice $= 8/(\varepsilon_1\varepsilon_2)$ and $k = n/m$, and thus by Slutsky's theorem, we then have

$$\frac{\sqrt{k}(\widehat{\rho}_{\mathrm{NI}}^{(G)} - \rho)}{(\pi/2) S_\eta \sqrt{1-(\widehat{\rho}_{\mathrm{NI}}^{(G)})^2}} \xrightarrow{d} N(0,1) \quad \text{as } k \to \infty.$$

We thus have asymptotically $(1-\alpha)$ coverage confidence intervals:

$$\left(\widehat{\rho}_{\mathrm{NI}}^{(G)} - \frac{\pi S_\eta \sqrt{1-(\widehat{\rho}^{(P)})^2}}{2\sqrt{k}} z_{1-\alpha/2}, \widehat{\rho}_{\mathrm{NI}}^{(G)} + \frac{\pi S_\eta \sqrt{1-(\widehat{\rho}^{(P)})^2}}{2\sqrt{k}} z_{1-\alpha/2}\right).$$

(b) (sub-Gaussian case) Identical to what we observed for the case of Gaussian data, note that $\widehat{\rho}_{\mathrm{NI}}^{(SG)}$ in (7) is an average of k iid observations T_j defined as follows:

$$\widehat{\rho}_{\mathrm{NI}}^{(SG)} = \frac{1}{k}\sum_{j=1}^{k} T_j \quad \text{where } T_j := m(\bar{X}^{(j)} + Z_1^{(j)})(\bar{Y}^{(j)} + Z_2^{(j)})$$

and

$$\begin{aligned}
\sigma_\rho^2 :=\ & \mathrm{Var}(T_j) \\
=\ & m^2 \mathbb{E}[(\bar{X}^{(j)})^2 (\bar{Y}^{(j)} + Z_2^{(j)})^2] + \frac{8}{\varepsilon_1^2}\mathrm{Var}(\bar{Y}^{(j)} + Z_2^{(j)}) - (\mathbb{E}[\widehat{\rho}_{\mathrm{NI}}^{(SG)}])^2 \\
=\ & m^2 \mathbb{E}[(\bar{X}^{(j)})^2 (\bar{Y}^{(j)})^2] - (\mathbb{E}[\widehat{\rho}_{\mathrm{NI}}^{(SG)}])^2 + m\,\mathrm{Var}(Z_2^{(j)}) + \frac{8}{\varepsilon_1^2}\left(\frac{\mathrm{Var}([Y]_{\lambda_2})}{m} + \frac{8}{m^2\varepsilon_2^2}\right) \\
=\ & \left(\frac{m-1}{m}\right)^2 (\mathrm{Var}([X]_{\lambda_1})\,\mathrm{Var}([Y]_{\lambda_2}) + (\mathbb{E}[\widehat{\rho}_{\mathrm{NI}}^{(SG)}])^2) + \frac{\mathbb{E}([X]_{\lambda_1}^4 [Y]_{\lambda_2}^4)}{m} \\
& + \frac{8}{m}\left(\frac{\mathrm{Var}([Y]_{\lambda_2})}{\varepsilon_1^2} + \frac{\mathrm{Var}([X]_{\lambda_1})}{\varepsilon_2^2}\right) + \frac{64}{m^2\varepsilon_1^2\varepsilon_2^2}.
\end{aligned}$$

where the last equality follows by expanding the squares of iid averages in the first term of the previous line. Thus we have

$$\frac{\sqrt{k}(\widehat{\rho}_{\mathrm{NI}}^{(SG)} - \mathbb{E}(\widehat{\rho}_{\mathrm{NI}}^{(SG)}))}{\sigma_\rho} \xrightarrow{d} N(0,1) \quad \text{as } k \to \infty,$$

To estimate σ_ρ^2 we use the sample variance of T_j:

$$S_\rho^2 := \frac{1}{k}\sum_{j=1}^{k}(T_j - \bar{T})^2.$$

Note that S_ρ^2 is constructed from $(\varepsilon_1, \varepsilon_2)$-DP statistics T_j, and thus S_ρ^2 is also differentially private. By standard calculations,

$$\mathbb{E}(S_\rho^2 - \sigma_\rho^2)^2 = O\left(\frac{1}{k}\right)$$

where we use our choice $m = 4\eta_1\eta_2(\log(n))/(\varepsilon_1\varepsilon_2)$ and $k = n/m$, and thus by Slutsky's theorem, along with the asymptotically vanishing bias from (21) we then have

$$\frac{\sqrt{k}(\widehat{\rho}_{\mathrm{NI}}^{(SG)} - \rho)}{S_\rho} \xrightarrow{d} N(0,1) \quad \text{as } k \to \infty.$$

We thus have an asymptotically $(1-\alpha)$ coverage confidence interval:

$$\left(\widehat{\rho}_{\mathrm{NI}}^{(SG)} - \frac{S_\rho}{\sqrt{k}} z_{1-\alpha/2}, \widehat{\rho}_{\mathrm{NI}}^{(SG)} + \frac{S_\rho}{\sqrt{k}} z_{1-\alpha/2}\right).$$

□

Proof of Theorem 3.1. We separate the proofs of the two statements as follows.

1. To derive the MSE of the interactive correlation estimator for Gaussian data, we first calculate from (9):

$$\begin{aligned}
&\mathbb{E}\left[\widehat{\eta}_{XY,\mathrm{int}} + Z - (2\mathbb{P}(XY>0) - 1)\right]^2 \\
&= \mathrm{Var}(\widehat{\eta}_{XY,\mathrm{int}} + Z) + (\mathbb{E}[\widehat{\eta}_{XY,\mathrm{int}}] - (2\mathbb{P}(XY>0) - 1))^2 \\
&= \frac{4e^{\varepsilon_1}}{n(\mathrm{e}^{\varepsilon_1} - 1)^2} + \frac{4(\mathrm{e}^{\varepsilon_1}+1)^2}{n^2(\mathrm{e}^{\varepsilon_1}-1)^2\varepsilon_2^2} + 0 \\
&\le \frac{4}{n(\varepsilon_1 \wedge 1)^2} + \frac{25}{n^2(\varepsilon_1 \wedge 1)^2\varepsilon_2^2}.
\end{aligned}$$

Consequently,

$$\begin{aligned}
&\mathbb{E}(\widehat{\rho}_{\mathrm{INT}}^{(G)} - \rho)^2 \\
&= \frac{\pi^2(1-\rho^2)}{4}\mathbb{E}\left[\widehat{\eta}_{XY,\mathrm{int}}^{(\mathrm{P})} - (2\mathbb{P}(XY>0) - 1)\right]^2 + \frac{\pi^4}{2}\mathbb{E}\left[\widehat{\eta}_{XY,\mathrm{int}}^{(\mathrm{P})} - (2\mathbb{P}(XY>0) - 1)\right]^4 \\
&\le \frac{\pi^2(1-\rho^2)+1}{n(\varepsilon_1 \wedge 1)^2} + \frac{25\pi^2(1-\rho^2)+1}{4n^2(\varepsilon_1 \wedge 1)^2\varepsilon_2^2}
\end{aligned}$$

whenever $n(\varepsilon_1 \wedge 1)$ is sufficiently large.

2. To derive the MSE of the interactive correlation estimator for sub-Gaussian data, we take the following approach. It is straightforward to check that

$$\begin{aligned}
\mathbb{E}[\widehat{\rho}_{\mathrm{INT}}^{(SG)}] &= \mathbb{E}\left([([X]_{\lambda_1} + Z)Y]_{\lambda_2}\right) \\
&= \mathbb{E}\left(([X]_{\lambda_1} + Z)Y\right) - \mathbb{E}[([X]_{\lambda_1} + Z)Y\mathbb{1}(|([X]_{\lambda_1} + Z)Y| > \lambda_2)] \\
&= \rho - \mathbb{E}[XY\mathbb{1}(|X| > \lambda_1)] - \mathbb{E}[([X]_{\lambda_1} + Z)Y\mathbb{1}(|([X]_{\lambda_1} + Z)Y| > \lambda_2)].
\end{aligned}$$

We next have

$$\begin{aligned}
\mathbb{P}(|([X]_{\lambda_1} + Z)Y| \ge \lambda_2) &\le \mathbb{P}(|Y| \gtrsim \sqrt{\log(n)}) + \mathbb{P}(|([X]_{\lambda_1} + Z)| > \lambda_2/\sqrt{\log(n)}) \\
&\lesssim \frac{1}{n} + \exp\left(-\frac{1}{2}\left\{\frac{\lambda_2^2}{(\log(n))\eta_1^2} \wedge \frac{\varepsilon_1\lambda_2}{2\lambda_1\sqrt{\log(n)}}\right\}\right) \qquad (22)
\end{aligned}$$

the fact that X is η_1-subgaussian and Y is η_2-subgaussian. We can then bound the bias of the estimator $\widehat{\rho}^{(SG)}_{\mathrm{INT}}$ as:

$$\begin{aligned}
&\left|\mathbb{E}[\widehat{\rho}^{(SG)}_{\mathrm{INT}}] - \rho\right| \\
&\leq \mathbb{E}[|XY|\mathbb{1}(|X| > \lambda_1)] + \mathbb{E}[(|XY| + |ZY|)(\mathbb{1}(|([X]_{\lambda_1} + Z)Y| > \lambda_2))] \\
&\leq (\mathbb{E}[|XY|^2])^{1/2}[\mathbb{P}(|X| > \lambda_1)]^{1/2} \\
&\quad + (\mathbb{E}|XY|^2 + \mathbb{E}|ZY|^2)^{1/2}[\mathbb{P}(|Y| \geq 2\sqrt{\log(n)}) + \mathbb{P}(|([X]_{\lambda_1} + Z)| > \lambda_2/\sqrt{\log(n)})]^{1/2} \\
&\lesssim \frac{1}{n} \qquad (23)
\end{aligned}$$

where we use (22) with

$$\lambda_1 = 2\eta_1\sqrt{\log(n)} \quad \text{and} \quad \lambda_2 = 4(\eta_2 \vee 1)(\log(n))^2/(\varepsilon_1 \wedge 1)$$

and hence the variance becomes

$$\begin{aligned}
\mathrm{Var}(\widehat{\rho}^{(SG)}_{\mathrm{INT}}) &\leq \frac{\mathrm{Var}(([X]_{\lambda_1} + Z_1)Y)}{n} + \frac{4\lambda_2^2}{n^2\varepsilon_2^2} \\
&\leq \frac{\mathrm{Var}(XY)}{n} + \frac{4\lambda_1^2}{n\varepsilon_1^2} + \frac{4\lambda_2^2}{n^2\varepsilon_2^2} \\
&= \frac{\mathrm{Var}(XY)}{n} + \frac{16\eta_1^2(\log(n))}{n\varepsilon_1^2} + \frac{64(\eta_2^2 \vee 1)(\log(n))^4}{n^2\varepsilon_2^2(\varepsilon_1 \wedge 1)^2}.
\end{aligned}$$

3. We split the proofs for confidence interval coverage into the Gaussian and sub-Gaussian cases.

   (a) (Gaussian case) From (9) we write:

$$\widehat{\eta}_{XY,\mathrm{int}} + Z = \frac{\exp(\varepsilon_1)+1}{\exp(\varepsilon_1)-1} \times \frac{1}{n}\sum_{i=1}^n T_i + Z =: \frac{\exp(\varepsilon_1)+1}{\exp(\varepsilon_1)-1}\left(\bar{T} + Z_2\right)$$

   where $T_i = (2S_i - 1)\,\mathrm{sign}(X_i)\,\mathrm{sign}(Y_i)$ and $Z_2 \sim \mathrm{Laplace}\left(0, \frac{2}{n\varepsilon_2}\right)$. Let us define:

$$\sigma_\eta^2 := \mathrm{Var}(T_i) = 1 - \left(\frac{\exp(\varepsilon_1)-1}{\exp(\varepsilon_1)+1}\right)^2 (2\mathbb{P}(XY > 0) - 1)^2$$

   for which we have the consistent estimator:

$$\widehat{\sigma}_\eta^2 := 1 - \left(\frac{\exp(\varepsilon_1)-1}{\exp(\varepsilon_1)+1}\right)^2 (\widehat{\eta}_{XY,\mathrm{int}})^2.$$

   We recall that $\mathbb{E}[\widehat{\eta}_{XY,\mathrm{int}}] = 2\mathbb{P}(XY > 0) - 1$ and thus by the Berry Esseen limit theorem on T_i,

$$\begin{aligned}
\sup_x \Bigg|\mathbb{P}\left(\frac{\sqrt{n}}{\sigma_\eta}\left(\frac{\exp(\varepsilon_1)-1}{\exp(\varepsilon_1)+1}\right)\left(\widehat{\eta}^{(\mathrm{P})}_{XY,\mathrm{int}} - \mathbb{P}(|XY| > 0)\right) \leq x\right) \\
- \mathbb{P}\left(Z_{XY} + Z_2' \leq x\right)\Bigg| \leq \frac{C}{\sigma_\eta^3\sqrt{n}} \qquad (24)
\end{aligned}$$

   for a numerical constant $C > 0$. Here

$$Z_{XY} \sim N(0,1) \quad \text{and} \quad Z_2' \sim \mathrm{Laplace}\left(0, \frac{2}{\sqrt{n}\sigma_\eta\varepsilon_2}\right).$$

   and thus by the delta method,

$$\begin{aligned}
\sup_x \Bigg|\mathbb{P}\left(\frac{2\sqrt{n}}{\pi\sqrt{1-\rho^2}\sigma_\eta}\left(\frac{\exp(\varepsilon_1)-1}{\exp(\varepsilon_1)+1}\right)\left(\widehat{\rho}^{(G)}_{\mathrm{INT}} - \rho\right) \leq x\right) \\
- \mathbb{P}\left(Z_{XY} + Z_2' \leq x\right)\Bigg| \leq \frac{C}{\sigma_\eta^3\sqrt{n}}. \qquad (25)
\end{aligned}$$

To derive the confidence intervals we make two separate cases:

Case 1: ($(\sqrt{n}\varepsilon_2)^{-1} \to c$) In the first case we consider $(\sqrt{n}\varepsilon_2)^{-1} \to c$ for a finite constant $c \geq 0$. In this case we have the confidence interval

$$\left(\widehat{\rho}_{\text{INT}}^{(G)} \mp \frac{\pi\widehat{\sigma}_\eta\sqrt{1-(\widehat{\rho}_{\text{INT}}^{(G)})^2}}{2\sqrt{n}}\left(\frac{\exp(\varepsilon_1)+1}{\exp(\varepsilon_1)-1}\right)F^{-1}(1-\alpha/2)\right)$$

where for any $x \in \mathbb{R}$ we define

$$F(x) := \mathbb{P}(Z_{XY}+c_*Z_{\text{Lap}} \leq x) \text{ where } c_* = \lim_{n\to\infty}\frac{2}{\sqrt{n}\sigma_\eta\varepsilon_2} \text{ and } Z_{\text{Lap}} \sim \text{Laplace}(0,1).$$

The above is a valid confidence interval when $\lim_{n\to\infty}\frac{2}{\sqrt{n}\sigma_\eta\varepsilon_2} = c_*$ for some finite $c_* \geq 0$. This is no longer the case when $\sqrt{n}\varepsilon_2 \to 0$ as $n \to \infty$.

Case 2: ($\sqrt{n}\varepsilon_2 \to 0$) In this case, (24) and (25) imply that we have the asymptotic convergence:

$$\frac{n\varepsilon_2}{\pi\sqrt{1-\rho^2}}\left(\frac{\exp(\varepsilon_1)-1}{\exp(\varepsilon_1)+1}\right)\left(\widehat{\rho}_{\text{INT}}^{(G)}-\rho\right) \xrightarrow{d} \text{Laplace}(0,1)$$

leading to the asymptotically $(1-\alpha)$ coverage confidence interval

$$\left(\widehat{\rho}_{\text{INT}}^{(G)} \pm \frac{\pi\sqrt{1-(\widehat{\rho}_{\text{int}}^{(\text{P})})^2}}{n\varepsilon_2}\left(\frac{\exp(\varepsilon_1)+1}{\exp(\varepsilon_1)-1}\right)\log(\alpha)\right)$$

where the width of the CI is determined by the α-th quantiles of the Laplace$(0,1)$ distribution.

(b) (sub-Gaussian case) Note that

$$\widehat{\rho}_{\text{INT}}^{(SG)} = \frac{1}{n}\sum_{i=1}^{n}T_i + Z_2$$

where $T_i = [([X_i]_{\lambda_1}+Z_{1i})Y_i]_{\lambda_2}$ are iid random variables. Thus by the Berry Esseen theorem,

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\frac{\sqrt{n}(\widehat{\rho}_{\text{INT}}^{(SG)}-\mathbb{E}[\widehat{\rho}_{\text{INT}}^{(SG)}])}{\sigma_\rho} \leq x\right) - \mathbb{P}\left(Z_{XY}+Z_2' \leq x\right)\right| \leq \frac{C(\log(n))^{7.5}}{\sigma_\rho^3\sqrt{n}} \tag{26}$$

where $Z_{XY} \sim N(0,1)$, $Z_2' \sim \text{Laplace}\left(0, \frac{2\lambda_2}{\sqrt{n}\sigma_\rho\varepsilon_2}\right)$ and

$$\sigma_\rho^2 := \text{Var}(([X]_{\lambda_1}+Z_1)Y).$$

As before, we now make two cases to derive the confidence intervals.

Case 1: ($(\sqrt{n}\varepsilon_2/\lambda_2)^{-1} \to c$) In the first case we consider $(\sqrt{n}\varepsilon_2/\lambda_2)^{-1} \to c$ for a finite constant $c \geq 0$. In this case (23) and (26) imply that we have the confidence interval

$$\left(\widehat{\rho}_{\text{INT}}^{(SG)} - \frac{\widehat{\sigma}_\rho}{\sqrt{n}}F^{-1}(1-\alpha/2), \widehat{\rho}_{\text{INT}}^{(SG)} + \frac{\widehat{\sigma}_\rho}{\sqrt{n}}F^{-1}(1-\alpha/2)\right)$$

where

$$\widehat{\sigma}_\rho^2 = \frac{1}{n}\sum_{i=1}^{n}(T_i-\bar{T})^2,$$

the sample variance of T_i, is an ε_1-DP consistent estimator for σ_ρ^2. Moreover, as before for any $x \in \mathbb{R}$ we define

$$F(x) := \mathbb{P}(Z_{XY}+c_*Z_{\text{Lap}} \leq x) \text{ where } c_* = \lim_{n\to\infty}\frac{2\lambda_2}{\sqrt{n}\sigma_\rho\varepsilon_2} \text{ and } Z_{\text{Lap}} \sim \text{Laplace}(0,1).$$

The above is a valid confidence interval when $\lim_{n\to\infty} \frac{2\lambda_2}{\sqrt{n}\sigma_\rho \varepsilon_2} = c_*$ for some finite $c_* \geq 0$. This is no longer the case when $\sqrt{n}\varepsilon_2/\lambda_2 \to 0$ as $n \to \infty$.

Case 2: ($\sqrt{n}\varepsilon_2/\lambda_2 \to 0$) In this case, (23) and (26) imply that we have the asymptotic convergence:

$$\frac{n\varepsilon_2}{2\lambda_2}\left(\widehat{\rho}_{\mathrm{INT}}^{(SG)} - \rho\right) \xrightarrow{d} \mathrm{Laplace}(0,1)$$

leading to the asymptotically $(1-\alpha)$ coverage confidence interval

$$\left(\widehat{\rho}_{\mathrm{INT}}^{(SG)} + \frac{\lambda_2}{n\varepsilon_2}\log(\alpha), \widehat{\rho}_{\mathrm{INT}}^{(SG)} - \frac{\lambda_2}{n\varepsilon_2}\log(\alpha)\right)$$

where the width of the CI is determined by the α-th quantiles of the Laplace$(0,1)$ distribution.

□

B.2 Proofs of lower bound results

Proof of Theorem 4.1 . Fix any non–interactive $(\varepsilon_1, \varepsilon_2, \delta_1, \delta_2)$–DP protocol with transcript $T = (T_1, T_2)$, and let P_ρ denote the law of T when $(X_{i1}, X_{i2})_{i=1}^n \overset{\text{i.i.d.}}{\sim} \mathcal{N}\big(\mathbf{0}, \left(\begin{smallmatrix}1 & \rho \\ \rho & 1\end{smallmatrix}\right)\big)$. We check that $f(x) = x\log(1/x)$ is an increasing function of x whenever $x \in (0, \exp(-1))$. Thus $\delta_k = o(n^{-1-\omega})$ implies $\delta_k \log(1/\delta_k) = o(n^{-1}) = o(\varepsilon_k^2)$. The second inequality follows from the fact that $n\varepsilon_k^2 \to \infty$. Invoking Lemma 2 gives, at $\rho = 0$,

$$I_F(T;0) \;\leq\; \frac{8}{\pi}\big(n\varepsilon_1^2 \wedge n\varepsilon_2^2\big). \tag{27}$$

Step 1. Prior supported in a small neighborhood of 0. Let $\mathcal{J} = [-L/2, L/2]$ with $L \leq 2c_0$ and center $\rho_0 = 0$. Define the cosine–squared prior on $\mathcal{J}$:

$$\lambda(\rho) \;=\; \frac{2}{L}\,\lambda_0\!\left(\frac{2(\rho-\rho_0)}{L}\right), \qquad \lambda_0(x) = \begin{cases}\cos^2(\pi x/2), & |x| \leq 1,\\ 0, & \text{otherwise.}\end{cases}$$

This prior satisfies the well–known identity (see, e.g., Tsybakov [2009])

$$I_F(\lambda) \;=\; \int_{\mathcal{J}} \frac{\lambda'(\rho)^2}{\lambda(\rho)}\,d\rho \;=\; \left(\frac{2\pi}{L}\right)^2. \tag{28}$$

Step 2. Prior–averaged information of the transcript. By (16) with $\rho_0 = 0$ and $|\rho| \leq L/2 \leq c_0$,

$$I_F(T;\rho) \;\leq\; (1 + C_\eta)\, I_F(T;0).$$

Therefore

$$\mathbb{E}_\rho\big[I_F(T;\rho)\big] = \int_{\mathcal{J}} I_F(T;\rho)\,\lambda(\rho)\,d\rho \;\leq\; (1+C_\eta)\, I_F(T;0). \tag{29}$$

Step 3. Van Trees inequality. Applying Lemma 1 with parameter ρ, likelihood P_ρ, and prior λ, we obtain the Bayes risk lower bound

$$\mathbb{E}\big[(\widehat{\rho}(T) - \rho)^2\big] \;\geq\; \frac{1}{\mathbb{E}_\rho[I_F(T;\rho)] \;+\; I_F(\lambda)} \overset{(29),(28)}{\geq} \frac{1}{(1+C_\eta)\, I_F(T;0) \;+\; \big(2\pi/L\big)^2}.$$

Using (27) and writing $A := n\varepsilon_1^2 \wedge n\varepsilon_2^2$,

$$\mathcal{R}_{\mathrm{Bayes}}(\lambda) \;:=\; \inf_{\widehat{\rho}} \mathbb{E}[(\widehat{\rho} - \rho)^2] \;\geq\; \frac{1}{c_1\, A \;+\; (2\pi/L)^2}, \qquad c_1 := \frac{8}{\pi}(1 + C_\eta). \tag{30}$$

Step 4. Choice of L and consequence. To minimize the denominator in (30) we take the largest admissible support, $L = 2c_0$, yielding

$$\mathcal{R}_{\text{Bayes}}(\lambda) \;\geq\; \frac{1}{c_1\, A \;+\; (\pi/c_0)^2}.$$

Since the minimax risk dominates the Bayes risk for any prior,

$$\inf_{\widehat{\rho}} \sup_{\rho\in[-1,1]} \mathbb{E}_\rho\big[(\widehat{\rho}-\rho)^2\big] \;\geq\; \mathcal{R}_{\text{Bayes}}(\lambda) \;\geq\; \frac{1}{c_1\, A \;+\; (\pi/c_0)^2}.$$

In particular, whenever $A \to \infty$ (our standing regime), the constant prior term is negligible and we obtain

$$\inf_{\widehat{\rho}} \sup_{\rho\in[-1,1]} \mathbb{E}_\rho\big[(\widehat{\rho}-\rho)^2\big] \;\gtrsim\; \frac{1}{n\varepsilon_1^2 \wedge n\varepsilon_2^2}.$$

Step 5. Classical $(1/n)$ term. Additionally the non–private parametric difficulty contributes an additional $\Omega(1/n)$ term (e.g. by repeating the bound above without privacy constraints and with $A \asymp n$ near $\rho = 0$). Combining the terms yields

$$\inf_{\widehat{\rho}\in \mathrm{NI}(\varepsilon_1,\varepsilon_2,\delta)} \sup_{\rho\in[-1,1]} \mathbb{E}_\rho\big[(\widehat{\rho}-\rho)^2\big] \;\gtrsim\; \frac{1}{n} \;+\; \frac{1}{n\varepsilon_1^2} \;+\; \frac{1}{n\varepsilon_2^2}. \qquad \square$$

Proof of Theorem 4.2. Let (i,j) be such that $\varepsilon_i \geq \varepsilon_j$. Since $\delta_i = o(n^{-1-\omega})$, we have $\delta_i \log(1/\delta_i) = o(n^{-1}) = o(\varepsilon_i^2)$, where the last equality follows from $n\varepsilon_i^2 \to \infty$. Similarly, $\delta_j = o(n^{-1-\omega})$ implies $\delta_j \log^2(1/\delta_j) = o(n^{-1}) = o(n\varepsilon_1^2\varepsilon_2^2)$, using that $n^2\varepsilon_1^2\varepsilon_2^2 \to \infty$. Hence, the conditions of Lemma 3 are met, so that $C_{\Pi,n}$ in (17) indeed represents the Fisher–information bound for the (interactive, one–way) protocol.

Throughout, we abbreviate

$$\varepsilon_{\max}^2 = \varepsilon_1^2 \vee \varepsilon_2^2, \qquad \varepsilon_{\min}^2 = \varepsilon_1^2 \wedge \varepsilon_2^2, \qquad C_{\Pi,n} = n\varepsilon_{\max}^2 \;\wedge\; n^2\varepsilon_1^2\varepsilon_2^2.$$

Let $\Pi \in \{\Pi_{1\to2}, \Pi_{2\to1}\}$ be any fixed one–way interactive DP protocol, and denote by P_ρ the law of the full transcript under correlation ρ.

Step 1. Local regularity of information and the prior. By the standing regularity assumption (16), there are numerical constants $c_0 \in (0,1)$ and $C_\eta > 0$ such that

$$I_F(\Pi;\rho+\epsilon) = I_F(\Pi;\rho)\big(1+\eta(\epsilon)\big), \qquad |\epsilon| < c_0, \qquad \sup_{|\epsilon|<c_0} |\eta(\epsilon)| \leq C_\eta.$$

In particular, for $|\rho| \leq c_0$,

$$I_F(\Pi;\rho) \;\leq\; (1+C_\eta)\, I_F(\Pi;0). \tag{31}$$

We place on ρ the cosine–squared prior supported on $\mathcal{J} = [-L/2, L/2]$ with $L \leq 2c_0$ and center 0 that the prior Fisher information is (see proof of Theorem 4.1 for details)

$$I_F(\lambda) = \int_{\mathcal{J}} \frac{\lambda'(\rho)^2}{\lambda(\rho)}\, d\rho = \Big(\frac{2\pi}{L}\Big)^2. \tag{32}$$

Step 2. Prior–averaged information of the transcript. By (31) and Lemma 3 (or its analogue for $\Pi_{2\to1}$),

$$\mathbb{E}_\rho\big[I_F(\Pi;\rho)\big] = \int_{\mathcal{J}} I_F(\Pi;\rho)\,\lambda(\rho)\,d\rho \;\leq\; (1+C_\eta)\, I_F(\Pi;0) \;\leq\; (1+C_\eta)\, C_{\Pi,n}.$$

Step 3. Van Trees inequality. Applying Lemma 1 with parameter ρ, likelihood P_ρ, and prior λ, we obtain

$$\mathcal{R}_{\text{Bayes}}(\lambda) := \inf_{\widehat{\rho}} \mathbb{E}\big[(\widehat{\rho}-\rho)^2\big] \;\geq\; \frac{1}{\mathbb{E}_\rho[I_F(\Pi;\rho)] \;+\; I_F(\lambda)} \;\geq\; \frac{1}{c_1\, C_{\Pi,n} \;+\; (2\pi/L)^2},$$

where $c_1 := 1 + C_\eta$ is an absolute constant. Choosing the largest admissible support $L = 2c_0$ gives

$$\mathcal{R}_{\text{Bayes}}(\lambda) \;\geq\; \frac{1}{c_1\, C_{\Pi,n} \;+\; (\pi/c_0)^2}. \tag{33}$$

Since the minimax risk dominates the Bayes risk for every prior,

$$\inf_{\widehat{\rho}} \sup_{\rho\in[-1,1]} \mathbb{E}_\rho\big[(\widehat{\rho}-\rho)^2\big] \;\geq\; \mathcal{R}_{\text{Bayes}}(\lambda).$$

Step 4. Extracting the two interactive terms. By definition, $C_{\Pi,n} = \min\{A, B\}$ with

$$A := n\varepsilon_{\max}^2, \qquad B := n^2\varepsilon_1^2\varepsilon_2^2.$$

Hence $1/C_{\Pi,n} = \max\{1/A, 1/B\} \geq \frac{1}{2}\big(1/A + 1/B\big)$. Using (33) and the fact that the additive constant $(\pi/c_0)^2$ is negligible whenever $A \vee B \to \infty$, we obtain the privacy–induced contribution

$$\inf_{\widehat{\rho}} \sup_{\rho} \mathbb{E}_\rho\big[(\widehat{\rho} - \rho)^2\big] \gtrsim \frac{1}{n\varepsilon_{\max}^2} + \frac{1}{n^2\varepsilon_1^2\varepsilon_2^2}.$$

Step 5. Baseline parametric term and conclusion. Even without privacy constraints, estimating a correlation from n i.i.d. Gaussian samples incurs risk $\Theta(1/n)$; hence

$$\inf_{\widehat{\rho} \in \mathrm{INT}(\varepsilon_1,\varepsilon_2,\delta)} \sup_{\rho \in [-1,1]} \mathbb{E}_\rho\big[(\widehat{\rho} - \rho)^2\big] \gtrsim \frac{1}{n} + \frac{1}{n\varepsilon_{\max}^2} + \frac{1}{n^2\varepsilon_1^2\varepsilon_2^2},$$

which is the desired bound. □

B.3 Proofs of Lemmas

In this section we provide proofs of lemmas used to prove the lower bound theorems.

Proof of Lemma 2. The main technical ingredient that goes into proving the minimax lower bound is obtaining a upper bound on the Fisher Information under the null i.e $\rho = 0$. Denote $\boldsymbol{Z} = (X_i, Y_i)_{i=1}^n$, it can be shown that the score function $S_\rho(\boldsymbol{Z})$ for the parameter ρ under the null is given by

$$S_\rho(\boldsymbol{Z}) = \sum_{i=1}^n X_i Y_i \tag{34}$$

The fisher information under the null $I_F(T; 0)$ is given by

$$I_F(T; 0) = \mathbb{E}(\mathbb{E}(S_\rho(\boldsymbol{Z}) \mid T)^2) \tag{35}$$

The fisher info under null can be expressed as

$$\begin{aligned} I_F(T;0) &= \mathbb{E}\left[\left(\sum_{i=1}^n \mathbb{E}(X_i Y_i \mid T)\right)^2\right] \\ &= \mathbb{E}\left[\left(\sum_{i=1}^n \mathbb{E}(X_i \mid T_1)\mathbb{E}(Y_i \mid T_2)\right)^2\right] \\ &= \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}\left[\mathbb{E}(X_i \mid T_1)\mathbb{E}(Y_i \mid T_2)\mathbb{E}(X_j \mid T_1)\mathbb{E}(Y_j \mid T_2)\right] \end{aligned}$$

where we have used the fact that $X_i \perp Y_i \mid T$, $X_i \perp T_2 \mid T_1$ and $Y_i \perp T_1 \mid T_2$ in the second line. We now have that

$$I_F(T;0) = \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}\left[\mathbb{E}(X_i|T_1)\mathbb{E}(X_j|T_1)\right] \mathbb{E}\left[\mathbb{E}(Y_i|T_2)\mathbb{E}(Y_j|T_2)\right]$$

where in we use the fact that under the null $T_1 \perp T_2$. Define to matrices M_X and M_Y such that

$$(M_X)_{ij} = \mathbb{E}\left[\mathbb{E}(X_i \mid T_1)\mathbb{E}(X_j \mid T_1)\right] \text{ and } (M_Y)_{ij} = \mathbb{E}\left[\mathbb{E}(Y_i \mid T_2)\mathbb{E}(Y_j \mid T_2)\right]$$

then we have that $I_F(T;0)] = \mathrm{tr}(M_X^\top M_Y)$. Using Lemma 6 we have that $I_F(T;0)] \leq \mathrm{tr}(M_X)\|M_Y\|_2$ where $\|.\|_2$ is the spectral norm . Next let us bound $\mathrm{tr}(M_X)$. Note that we can rewrite M_X as

$$M_X = \mathbb{E}\left(\mathbb{E}(\boldsymbol{X} \mid T_1)\mathbb{E}(\boldsymbol{X} \mid T_1)^\top\right) \tag{36}$$

where X is the data vector $(X_i)_{i=1}^n$ and $\mathbb{E}(\boldsymbol{X} \mid T)$ is the vector $(\mathbb{E}(X_i \mid T))_{i=1}^n$. Hence

$$\begin{aligned}\operatorname{tr}(M_X) &\leq \operatorname{tr}\left(\mathbb{E}\left(\mathbb{E}(\boldsymbol{X} \mid T_1)\mathbb{E}(\boldsymbol{X} \mid T_1)^\top\right)\right) \\ &= \mathbb{E}\|\mathbb{E}(\boldsymbol{X} \mid T_1)\|_2^2 \\ &= \sum_{i=1}^n \mathbb{E}(\mathbb{E}(X_i \mid T_1))^2\end{aligned}$$

Using Lemma 5 we have that $\operatorname{tr}(M_X) \leq n\frac{2}{\pi}\left(\frac{e^{\varepsilon_1}-e^{-\varepsilon_1}}{2}\right)^2$. For bounding $\|M_Y\|_2$ we can either bound by $\operatorname{tr}(M_Y)$ which implies by the previous argument that $\|M_Y\|_2 \leq n\varepsilon_2^2$ or using contraction of the conditional expectation i.e.

$$M_X = \mathbb{E}(\mathbb{E}(\boldsymbol{X} \mid T)\mathbb{E}(\boldsymbol{X} \mid T)^\top) \preceq \mathbb{E}(\boldsymbol{X}\boldsymbol{X}^\top) = I_n$$

which implies $\|M_X\|_2 \leq 1$. Putting everything together we have that

$$\begin{aligned}I_F(T;0) &\leq \operatorname{tr}(M_X)\|M_Y\|_2 \wedge \operatorname{tr}(M_Y)\|M_X\|_2 \\ &\leq n\frac{2}{\pi}\left(\frac{e^{\varepsilon_1}-e^{-\varepsilon_1}}{2}\right)^2 \wedge n\frac{2}{\pi}\left(\frac{e^{\varepsilon_1}-e^{-\varepsilon_1}}{2}\right)^2.\end{aligned}$$

Using the fact that $e^x - 1 \leq 2x$ for $0 < x < 1$ we have that

$$I_F(T;0) \leq \frac{8}{\pi}\left(n\varepsilon_1^2 \wedge n\varepsilon_2^2\right).$$

□

Proof of Lemma 3. Denote $\boldsymbol{Z} = (X_i, Y_i)_{i=1}^n$, it can be shown that the score function $S_\rho(\boldsymbol{Z})$ for the parameter ρ under the null is given by

$$S_\rho(\boldsymbol{Z}) = \sum_{i=1}^n X_i Y_i \tag{37}$$

The fisher information under the null $I_F(T;0)$ is given by

$$I_F(T;0) = \mathbb{E}(\mathbb{E}(S_\rho(\boldsymbol{Z}) \mid T)^2) \tag{38}$$

The fisher info under null can be expressed as

$$\begin{aligned}I_F(T;0) &= \mathbb{E}\left[\left(\sum_{i=1}^n \mathbb{E}(X_i Y_i \mid T)\right)^2\right] \\ &= \mathbb{E}\left[\left(\sum_{i=1}^n \mathbb{E}(X_i \mid T_1)\mathbb{E}(Y_i \mid T_1, T_2)\right)^2\right] \\ &= \mathbb{E}\left[\left(\sum_{i=1}^n \mathbb{E}(\mathbb{E}(X_i \mid T_1)Y_i \mid T_1, T_2)\right)^2\right]\end{aligned} \tag{39}$$

where we have used the fact that $X_i \perp Y_i \mid T$, $X_i \perp T_2 \mid T_1$ in the second line. Using the fact that $\mathbb{E}(\mathbb{E}(A|B)^2) \leq \mathbb{E}A^2$ we have that

$$I_F(T;0) = \mathbb{E}\left[\left(\sum_{i=1}^n \mathbb{E}(\mathbb{E}(X_i \mid T_1)Y_i \mid T_1, T_2)\right)^2\right] \leq \mathbb{E}\left[\left(\sum_{i=1}^n \mathbb{E}(\mathbb{E}(X_i \mid T_1)Y_i)\right)^2\right]$$

Hence expanding the sum of squares we have that

$$\begin{aligned} I_F(T;0) &= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}\left[\mathbb{E}(X_i \mid T_1) Y_i \mathbb{E}(X_j \mid T_1) Y_j\right] \\ &= \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{E}(X_i \mid T_1)^2 Y_i^2\right] \\ &= \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{E}(X_i \mid T_1)^2\right] \leq n\left(\frac{e^{\varepsilon_1}-e^{-\varepsilon_1}}{2}\right)^2 \end{aligned}$$

where we used the fact that $Y_i \perp Y_j, T_1$ and $\mathbb{E}Y_i = 0$, $\mathbb{E}Y_i^2 = 1$ in the second and third line. The last inequality above follows from Lemma 5.

Following (39) we can write

$$I_F(T;0) = \sum_{k=1}^{n} \mathbb{E}\left[\mathbb{E}(X_k \mid T_1) Y_k \left(\sum_{i=1}^{n} \mathbb{E}(\mathbb{E}(X_i \mid T_1) Y_i \mid T_1, T_2)\right)\right] \tag{40}$$

Let us call $G_k = \mathbb{E}(X_k \mid T_1) Y_k \left(\sum_{i=1}^{n} \mathbb{E}(\mathbb{E}(X_i \mid T_1) Y_i \mid T_1, T_2)\right)$ and $G_k' = \mathbb{E}(X_k \mid T_1) Y_k \left(\sum_{i=1}^{n} \mathbb{E}(\mathbb{E}(X_i \mid T_1) Y_i \mid T_1, T_2')\right)$. Also note that $\mathbb{E}G_k' = 0$ since $\mathbb{E}Y_k = 0$ and $Y_k \perp T_1, T_2'$. Now following a similar argument as in (46) we get that

$$\mathbb{E}G_k \leq \left(\frac{e^{\varepsilon_2}-e^{-\varepsilon_2}}{2}\right)\mathbb{E}|G_k'| + 2\delta_2 M + \int_M^\infty \mathbb{P}(|G_k| \geq t)dt + \int_M^\infty \mathbb{P}(|G_k'| \geq t)dt \tag{41}$$

Note that

$$\begin{aligned} \mathbb{E}|G_k'| &= \mathbb{E}\left|\mathbb{E}(X_k \mid T_1) Y_k \left(\sum_{i=1}^{n} \mathbb{E}(\mathbb{E}(X_i \mid T_1) Y_i \mid T_1, T_2')\right)\right| \\ &= \sqrt{\frac{2}{\pi}}\mathbb{E}\left|\mathbb{E}(X_k \mid T_1) \left(\sum_{i=1}^{n} \mathbb{E}(\mathbb{E}(X_i \mid T_1) Y_i \mid T_1, T_2')\right)\right| \\ &\leq \sqrt{\frac{2}{\pi}}\sqrt{\mathbb{E}(\mathbb{E}(X_k \mid T_1)^2)}\sqrt{\mathbb{E}\left[\left(\sum_{i=1}^{n} \mathbb{E}(\mathbb{E}(X_i \mid T_1) Y_i \mid T_1, T_2')\right)^2\right]} \\ &\leq \sqrt{\frac{2}{\pi}}\left(\frac{e^{\varepsilon_1}-e^{-\varepsilon_1}}{2} \wedge 1\right)\sqrt{I_F(T;0)}. \end{aligned}$$

The last line follows since $(T_1, T_2') \stackrel{d}{=} (T_1, T_2)$ and the fact that $\mathbb{E}(\mathbb{E}(X_k \mid T_1)^2) \leq EX_k^2 = 1$ Using the fact that $I_F(T;0) = \sum_k \mathbb{E}G_k$ and putting everything together we have that

$$I_F(T;0) \leq n\left(\frac{e^{\varepsilon_2}-e^{-\varepsilon_2}}{2}\right)\sqrt{\frac{2}{\pi}}\left(\frac{e^{\varepsilon_1}-e^{-\varepsilon_1}}{2} \wedge 1\right)\sqrt{I_F(T;0)} \tag{42}$$

$$+ 2n\delta_2 M + n\int_M^\infty \mathbb{P}(|G_k| \geq t)dt + n\int_M^\infty \mathbb{P}(|G_k'| \geq t)dt \tag{43}$$

Set

$$M = 64\left(\log\frac{8}{\delta_2}\right)^2$$

in Lemma 4, to obtain

$$\int_M^\infty \mathbb{P}(|G_k| \geq t)\, dt \leq 16\left(8\log(8/\delta_2) + 4\right)(\delta_2/8)^2 \leq \delta_2.$$

We can similarly show that

$$\int_M^\infty \mathbb{P}(|G_k'| \geq t)dt \leq \delta_2.$$

Putting everything together we have that

$$I_F(T;0) \leq n\left(\frac{e^{\varepsilon_2}-e^{-\varepsilon_2}}{2}\right)\sqrt{\frac{2}{\pi}}\left(\frac{e^{\varepsilon_1}-e^{-\varepsilon_1}}{2}\right)\sqrt{I_F(T;0)} \tag{44}$$

$$+ 2n\delta_2 64\left(\log\frac{8}{\delta_2}\right)^2 + 2n\delta_2 \tag{45}$$

If $\sqrt{I_F(T;0)} \leq n\sqrt{\frac{2}{\pi}}\left(\frac{e^{\varepsilon_2}-e^{-\varepsilon_2}}{2}\right)\left(\frac{e^{\varepsilon_1}-e^{-\varepsilon_1}}{2}\wedge 1\right)$ we are done else dividing both sides by $\sqrt{I_F(T;0)}$ we have

$$\begin{aligned}\sqrt{I_F(T;0)} \leq & n\sqrt{\frac{2}{\pi}}\left(\frac{e^{\varepsilon_2}-e^{-\varepsilon_2}}{2}\right)\left(\frac{e^{\varepsilon_1}-e^{-\varepsilon_1}}{2}\wedge 1\right)\\ &+\left(2n\delta_2 64\left(\log\frac{8}{\delta_2}\right)^2+2n\delta_2\right)n^{-1}\left(\frac{2}{\pi}\right)^{-1/2}\left(\frac{e^{\varepsilon_1}-e^{-\varepsilon_1}}{2}\wedge 1\right)^{-1}\left(\frac{e^{\varepsilon_2}-e^{-\varepsilon_2}}{2}\right)^{-1}\end{aligned}$$

The second term can be dropped if $\delta_2\log(1/\delta_2)^2 = o(n\varepsilon_1^2\varepsilon_2^2)$. The final form is achieved by using the fact that $\varepsilon_1, \varepsilon_2 \leq 1$. □

B.4 Auxiliary Lemmas

Lemma 4. *Define* $G_k = \mathbb{E}(X_k \mid T_1)Y_k\left(\sum_{i=1}^n \mathbb{E}(\mathbb{E}(X_i \mid T_1)Y_i \mid T_1, T_2)\right)$ *then we have that*

$$\int_M^\infty \mathbb{P}(|G_k| \geq t)dt \leq 16(\sqrt{M})e^{-\sqrt{M}/4}$$

Proof. Let us denote by $Z_i = \mathbb{E}(X_i \mid T_1)Y_i$. We begin by bounding $\mathbb{E}e^{t|G_i|^{1/2}}$. By AM-GM and Cauchy-Schwarz inequality, we have that

$$\begin{aligned}\mathbb{E}e^{t|G_i|^{1/2}} &= \mathbb{E}e^{t|Z_i|^{1/2}|\mathbb{E}(Z_i|T_1,T_2)|^{1/2}}\\ &\leq \mathbb{E}e^{\frac{1}{2}t(|Z_i|+|\mathbb{E}(Z_i|T_1,T_2)|)}\\ &\leq \sqrt{\mathbb{E}e^{t|Z_i|}}\sqrt{\mathbb{E}e^{t|\mathbb{E}(Z_i|T_1,T_2)|}}\end{aligned}$$

Using the conditional Jensen's Inequality with the function $x \to e^{tx^2}$ which is convex to obtain that

$$\mathbb{E}e^{t|\mathbb{E}(Z_i|T_1,T_2)|} \leq \mathbb{E}(\mathbb{E}(e^{t|Z_i|} \mid T_1, T_2)) = \mathbb{E}(e^{t|Z_i|})$$

Hence $\mathbb{E}e^{t|G_i|^{1/2}} \leq \mathbb{E}(e^{t|Z_i|})$. Bounding the RHS as follows

$$\begin{aligned}\mathbb{E}(e^{t|Z_i|}) &= \mathbb{E}e^{t|Y_i||\mathbb{E}(X_i|T_1)|}\\ &\leq \mathbb{E}e^{\frac{1}{2}t\left(Y_i^2+(\mathbb{E}X_i|T_1)^2\right)}\\ &\leq \mathbb{E}e^{\frac{1}{2}tY_i^2}\mathbb{E}e^{\frac{1}{2}t(\mathbb{E}X_i|T_1)^2}\end{aligned}$$

where we used the AM-GM inequality in the second line and the independence of Y_i and $\mathbb{E}(X_i \mid T_1)$ in the third line. Using conditional Jensen again, we would have $\mathbb{E}e^{\frac{1}{2}t(\mathbb{E}X_i|T_1)^2} \leq \mathbb{E}e^{\frac{1}{2}tX_i^2}$ which implies $\mathbb{E}(e^{t|Z_i|}) \leq \mathbb{E}e^{\frac{1}{2}tX_i^2}\mathbb{E}e^{\frac{1}{2}tY_i^2}$.

Putting everything together we have that $\mathbb{E}e^{t|G_i|^{1/2}} \leq \mathbb{E}e^{\frac{1}{2}tX_i^2}\mathbb{E}e^{\frac{1}{2}tY_i^2} \leq 2$ for $t \leq 1/2$ (since $X_i, Y_i \sim \chi_1^2$). This implies that

$$\mathbb{P}(|G_i| \geq t) \leq \mathbb{P}(e^{\frac{1}{4}|G_i|^{1/2}} \geq e^{\sqrt{t}/4}) \leq 2e^{-\sqrt{t}/4}.$$

The last inequality follows from Markov. Hence we have that

$$\int_M^\infty \mathbb{P}(|G_i| \geq t)\,dt \leq 2\int_M^\infty e^{-\sqrt{t}/4}\,dt = 16(\sqrt{M})e^{-\sqrt{M}/4}.$$

□

Lemma 5. *Assuming for $k = 1, 2$, $\delta_k \log(1/\delta_k) = o(\varepsilon_k^2)$, we have for any $1 \le i \le n$ $\mathbb{E}(\mathbb{E}(X_i \mid T_1))^2 \le \frac{2}{\pi}\left(\frac{e^{\varepsilon_1}-e^{-\varepsilon_1}}{2}\right)^2$, similarly we have $\mathbb{E}(\mathbb{E}(Y_i \mid T_2))^2 \le \frac{2}{\pi}\left(\frac{e^{\varepsilon_2}-e^{-\varepsilon_2}}{2}\right)^2$.*

Proof of Lemma 5. Note that $\mathbb{E}(\mathbb{E}(X_i \mid T_1))^2 = \mathbb{E}[X_i(\mathbb{E}(X_i \mid T_1))]$. Denote $A_i = X_i(\mathbb{E}(X_i \mid T_1))$ we can write $\mathbb{E}A_i = \mathbb{E}A_i^+ - \mathbb{E}A_i^-$. Also let us define $A_i' = X_i(\mathbb{E}(X_i \mid T_1'))$ where $T_1' = T_1(\boldsymbol{X}')$, where $\boldsymbol{X}'$ is the adjacent dataset with its ith data point replaced by X_i' which is an independent copy.

We can write $\mathbb{E}A_i^+$ as

$$\begin{aligned}
\mathbb{E}(A_i^+) &= \int_0^\infty \mathbb{P}(A_i^+ \ge t)dt \\
&= \int_0^M \mathbb{P}(A_i^+ \ge t)dt + \int_M^\infty \mathbb{P}(A_i^+ \ge t)dt \\
&\le \int_0^M e^{\varepsilon_1}\mathbb{P}((A_i')^+ \ge t)dt + \delta_1 M + \int_M^\infty \mathbb{P}(A_i^+ \ge t)dt \\
&= e^{\varepsilon_1}\mathbb{E}(A_i')^+ - e^{\varepsilon_1}\int_M^\infty \mathbb{P}((A_i')^+ \ge t)dt + \delta_1 M + \int_M^\infty \mathbb{P}(A_i^+ \ge t)dt \\
&\le e^{\varepsilon_1}\mathbb{E}(A_i')^+ + \delta_1 M + \int_M^\infty \mathbb{P}(|A_i| \ge t)dt
\end{aligned}$$

Similarly we have that

$$\begin{aligned}
\mathbb{E}(A_i^-) &= \int_0^\infty \mathbb{P}(A_i^- \ge t)dt \\
&= \int_0^M \mathbb{P}(A_i^- \ge t)dt + \int_M^\infty \mathbb{P}(A_i^- \ge t)dt \\
&\ge \int_0^M e^{-\varepsilon_1}\mathbb{P}((A_i')^- \ge t)dt - \delta_1 M + \int_M^\infty \mathbb{P}(A_i^- \ge t)dt \\
&= e^{-\varepsilon_1}\mathbb{E}(A_i')^- - e^{-\varepsilon_1}\int_M^\infty \mathbb{P}((A_i')^- \ge t)dt - \delta_1 M + \int_M^\infty \mathbb{P}(A_i^- \ge t)dt \\
&\ge e^{-\varepsilon_1}\mathbb{E}(A_i')^- - \int_M^\infty \mathbb{P}(|A_i'| \ge t)dt - \delta_1 M
\end{aligned}$$

Since $\mathbb{E}A_i = \mathbb{E}A_i^+ - \mathbb{E}A_i^-$ we have that

$$\begin{aligned}
\mathbb{E}A_i &\le e^{\varepsilon_1}\mathbb{E}(A_i')^+ - e^{-\varepsilon_1}\mathbb{E}(A_i')^- + 2\delta_1 M + \int_M^\infty \mathbb{P}(|A_i| \ge t)dt + \int_M^\infty \mathbb{P}(|A_i'| \ge t)dt \\
&= \left(\frac{e^{\varepsilon_1}+e^{-\varepsilon_1}}{2}\right)\mathbb{E}A_i' + \left(\frac{e^{\varepsilon_1}-e^{-\varepsilon_1}}{2}\right)\mathbb{E}|A_i'| + 2\delta_1 M + \int_M^\infty \mathbb{P}(|A_i| \ge t)dt + \int_M^\infty \mathbb{P}(|A_i'| \ge t)dt \\
&= \left(\frac{e^{\varepsilon_1}-e^{-\varepsilon_1}}{2}\right)\mathbb{E}|A_i'| + 2\delta_1 M + \int_M^\infty \mathbb{P}(|A_i| \ge t)dt + \int_M^\infty \mathbb{P}(|A_i'| \ge t)dt \qquad (46)
\end{aligned}$$

where we have used the fact that $\mathbb{E}A_i' = 0$. Observe that

$$\mathbb{E}|A_i'| = \mathbb{E}|X_i|\mathbb{E}|\mathbb{E}(X_i \mid T_1')| \le \sqrt{\frac{2}{\pi}}\sqrt{\mathbb{E}(\mathbb{E}(X_i \mid T_1'))^2} = \sqrt{\frac{2}{\pi}}\sqrt{\mathbb{E}A_i}$$

Next we upper bound $\int_M^\infty \mathbb{P}(|A_i| \ge t)dt$ in that direction we look at

$$\begin{aligned}
\mathbb{E}e^{t|A_i|} &= \mathbb{E}e^{t|X_i||\mathbb{E}(X_i|T_1)|} \\
&\le \mathbb{E}e^{\frac{1}{2}t(X_i^2 + (\mathbb{E}(X_i|T_1))^2)}
\end{aligned}$$

where we used the AM-GM inequality for the exponent. Next we can apply the Cauchy-Schwarz inequality to obtain that

$$\mathbb{E}e^{t|A_i|} \leq \sqrt{\mathbb{E}e^{tX_i^2}}\sqrt{\mathbb{E}e^{t\mathbb{E}(X_i|T_1))^2}}$$

the second term can further be bounded using the conditional Jensen's Inequality with the function $x \to e^{tx^2}$ which is convex to obtain that

$$\mathbb{E}e^{t\mathbb{E}(X_i|T_1))^2} \leq \mathbb{E}(\mathbb{E}(e^{tX_i^2} \mid T_1)) = \mathbb{E}(e^{tX_i^2})$$

Putting everything together we have that $\mathbb{E}e^{t|A_i|} \leq \mathbb{E}(e^{tX_i^2}) \leq \sqrt{2}$ for $t \leq 1/4$ (since $X_i \sim \chi_1^2$).This implies that

$$\mathbb{P}(|A_i| \geq t) \leq \mathbb{P}(e^{\frac{1}{4}|A_i|} \geq e^{t/4}) \leq \sqrt{2}e^{-t/4}.$$

The last inequality follows from Markov. Hence we have that $\int_M^\infty \mathbb{P}(|A_i| \geq t) \leq 4\sqrt{2}e^{-M/4}$, set $M = 4\log(1/\delta_1)$ to obtain $\int_M^\infty \mathbb{P}(|A_i| \geq t) \leq 4\sqrt{2}\delta_1$. we can similarly show that

$$\int_M^\infty \mathbb{P}(|A_i'| \geq t)dt \leq 4\sqrt{2}\delta_1.$$

Putting everything together we have that

$$\mathbb{E}A_i \leq \left(\frac{e^{\varepsilon_1} - e^{-\varepsilon_1}}{2}\right)\sqrt{\frac{2}{\pi}}\sqrt{\mathbb{E}A_i} + 8\delta_1\log(1/\delta_1) + 8\sqrt{2}\delta_1$$

If $\mathbb{E}A_i \leq \frac{2}{\pi}\left(\frac{e^{\varepsilon_1}-e^{-\varepsilon_1}}{2}\right)^2$ we are done else dividing both sides by $\sqrt{\mathbb{E}A_i}$ we have

$$\sqrt{\mathbb{E}A_i} \leq \left(\frac{e^{\varepsilon_1} - e^{-\varepsilon_1}}{2}\right)\sqrt{\frac{2}{\pi}} + (8\delta_1\log(1/\delta_1) + 8\sqrt{2}\delta_1)\left(\frac{2}{\pi}\right)^{-1/2}\left(\frac{e^{\varepsilon_1} - e^{-\varepsilon_1}}{2}\right)^{-1}$$

The second term can be dropped if $\delta_1\log(1/\delta_1) = o(\varepsilon_1^2)$. □

Lemma 6. *For square matrices A and B, if B is symmetric, we have*

$$\operatorname{tr}(AB) \leq \|A\|_2 \operatorname{tr}(B)$$

Proof of Lemma 6. The proof follows from von Neumann's trace inequality:

$$\operatorname{tr}(AB) \leq |\operatorname{tr}(AB)| \leq \sum_i \alpha_i\beta_i \leq \max(\alpha_i)\sum_i \beta_i = \max(\alpha_i) \times \operatorname{tr}(B)$$

where α_i and β_i are the singular values of A and B respectively. The proof follows by the definition of ℓ_2 operator norm used on matrix A. □