Optimal Variance and Covariance Estimation Under Differential Privacy in the Add-Remove Model and Beyond

Shokichi Takakura
LY Corporation

Seng Pei Liew
LY Corporation

Satoshi Hasegawa
LY Corporation

Abstract

In this paper, we study the problem of estimating the variance and covariance of datasets under differential privacy in the add-remove model. While estimation in the swap model has been extensively studied in the literature, the add-remove model remains less explored and more challenging, as the dataset size must also be kept private. To address this issue, we develop efficient mechanisms for variance and covariance estimation based on the *Bézier mechanism*, a novel moment-release framework that leverages Bernstein bases. We prove that our proposed mechanisms are minimax optimal in the high-privacy regime by establishing new minimax lower bounds. Moreover, beyond worst-case scenarios, we analyze instance-wise utility and show that the Bézier-based estimator consistently achieves better utility compared to alternative mechanisms. Finally, we demonstrate the effectiveness of the Bézier mechanism beyond variance and covariance estimation, showcasing its applicability to other statistical tasks.

1 INTRODUCTION

Variance and covariance are fundamental statistics in data analysis. They are widely used in various applications, such as machine learning, statistics, and data mining. However, releasing statistics naively can lead to privacy issues, as they can reveal sensitive information about individuals in the dataset, for example, through difference attacks (Dalenius, 1977).

Differential privacy (DP) (Dwork et al., 2014a) is a widely used framework for protecting the privacy of individuals in a dataset. It provides a formal guarantee that the output of a computation does not reveal too much information about any individual in the dataset. In DP, there are two common settings for defining neighboring relations between datasets: *swap model* (Dwork et al., 2006) and *add-remove model* (Dwork et al., 2014a). Intuitively, in the swap model, two datasets are neighboring if one can be obtained from the other by changing one record, and in the add-remove model, two datasets are neighboring if one can be obtained from the other by adding or removing one record. Since the add-remove model is more conservative in the sense that it protects the size of the dataset, it is often preferred for statistical estimation in practice (McSherry, 2009; Wilson et al., 2020; Rogers et al., 2021; Amin et al., 2022).

Despite its practical importance, theoretical understanding of the add-remove model is less explored compared to the swap model. Recently, Kulesza et al. (2024) studied the problem of mean estimation in the add-remove model and provided a minimax lower bound and an optimal mechanism. However, estimation of variance and covariance is more complicated since it involves estimating the second moment of the dataset as well as the first moment. Indeed, naive approaches suffer from a large error compared to mean estimation.

Thus, we pose the following question: *Are variance and covariance estimation in the add-remove model harder than in the swap model or for mean estimation?*

In this paper, we answer this question negatively by developing efficient mechanisms for variance and covariance estimation which achieve the same optimal error rate in the high-privacy regime as in the swap model and mean estimation. Specifically, we establish a novel lower bound for variance and covariance estimation and propose minimax optimal mechanisms which achieve this lower bound. A key ingredient of these optimal mechanisms is the *Bézier mechanism*, a novel and versatile moment-release framework based on Bernstein bases. We show that the Bézier mechanism can estimate all moments up to order $k > 0$ simultaneously, without incurring any additional privacy cost compared

Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300. Copyright 2026 by the author(s).

to estimating only the k-th moment. Furthermore, beyond variance and covariance estimation, we show that the Bézier mechanism can be applied to the estimation of a wide range of statistics and improves utility, which is of independent interest.

Our contributions can be summarized as follows:

- We establish new minimax lower bounds for variance and covariance estimation under DP in both the swap and add-remove models. In the high-privacy regime, the optimal MSE is shown to be $\frac{2}{\varepsilon^2 n^2}$, matching the optimal MSE of mean estimation.
- We propose minimax optimal mechanisms for variance and covariance estimation in the add-remove model based on the *Bézier mechanism*, a novel moment-release framework based on Bernstein bases. Moreover, for variance estimation, we prove our Bézier-based estimator uniformly improves the leading constant compared to two alternative minimax optimal mechanisms.
- We show that our framework can be applied to higher-order moments estimation, achieving better utility for a wide class of moment-based quantities (e.g., skewness and correlation).

1.1 Other Related Work

Covariance Estimation Covariance estimation with DP has been intensively studied in various settings (Amin et al., 2019a; Biswas et al., 2020; Dong et al., 2022; Kalinin et al., 2025). However, they focus on the swap model and study the dependence of the error on the data dimension. Furthermore, they considered the estimation of uncentered covariance matrix. In contrast, we consider the estimation of *centered* covariance in the add-remove model and focus on the constant factor in the error rate.

Other Statistics Estimation Estimation of statistics other than variance and covariance such as quantile (Dwork et al., 2014b; Gillenwater et al., 2021; Imola et al., 2025), CDF (Rameshwar et al., 2025), F statistics (Swanberg et al., 2019), and confidence interval (Karwa and Vadhan, 2018) is also a field of active research. Several works (Gillenwater et al., 2021; Imola et al., 2025) have discussed the estimation in the add-remove model but the optimal mechanism for variance and covariance estimation remains unexplored.

1.2 Notation

$\mathrm{clip}(\cdot, [l, u]) : \mathbb{R} \to \mathbb{R}$ is the clipping function defined as $\mathrm{clip}(x, [l, u]) = \max\{l, \min\{u, x\}\}$. $\mathrm{Lap}(b)$ is the Laplace distribution with scale parameter b, i.e., the probability density function of $\mathrm{Lap}(b)$ is given by $\frac{1}{2b}\exp(-|x|/b)$. With a slight abuse of notation, we denote $\mathrm{Lap}(b)$ as a random variable following the Laplace distribution. We denote $[a_1, \ldots, a_d] \in \mathbb{R}^{n\times d}$ for $a_i \in \mathbb{R}^n$ as the matrix whose i-th column is a_i.

2 PROBLEM SETTING

In this paper, we consider datasets containing records, each of which is a d-dimensional vector in a bounded domain $[l, r]^d$ $(l < r)$. The case where the range of the records is not bounded can be transformed to the bounded case using the clipping technique (Amin et al., 2019b). Without loss of generality, we can assume $l = 0$ and $r = 1$ since any dataset can be transformed to this form by an affine transformation. Let $\mathcal{D}_n$ be the set of all datasets of size $n \geq 1$, $\mathcal{D}_{\geq n} := \cup_{k\geq n}\mathcal{D}_k$ be the set of all datasets of size at least n, and $\mathcal{D}_* := \cup_{k\geq 1}\mathcal{D}_k$ be the set of all datasets with any size. We identify the dataset $D \in \mathcal{D}_n$ with the data matrix $[x^{(1)}, \ldots, x^{(d)}] \in [0, 1]^{n\times d}$. In this paper, we consider the problem of estimating the variance and covariance defined as

$$\mathrm{var}(D) = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2,$$
$$\mathrm{cov}(D) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right),$$

where $D = [x] \in \mathcal{D}^*$, and $D = [x, y] \in \mathcal{D}^*$, respectively. We emphasize that our focus is on centered covariance estimation, whereas most previous work (Dwork et al., 2014b; Amin et al., 2019a; Biswas et al., 2020; Dong et al., 2022; Kalinin et al., 2025) has studied the estimation of uncentered covariance, $\frac{1}{n}\sum_{i=1}^{n} x_i y_i$. Centered covariance is more complicated since it depends on the first moment in a non-linear manner and we cannot directly apply the results for uncentered covariance estimation.

Here, we introduce the formal definition of DP.

Definition 2.1 (Differential Privacy)**.** A randomized algorithm $\mathcal{A}$ is said to be ϵ-DP if for all neighboring datasets $D, D' \in \mathcal{D}^*$ (we denote $D \sim D'$) and all measurable sets $S \subseteq \mathrm{Range}(\mathcal{A})$, it holds that

$$\Pr[\mathcal{A}(D) \in S] \leq e^{\epsilon}\Pr[\mathcal{A}(D') \in S].$$

For the definition of *neighboring* datasets, there are two common models:

- **Swap Model**: The neighboring datasets are defined by replacing an element in the dataset with another one. Formally, two datasets $D, D' \in \mathcal{D}^*$ are neighboring if and only if $|D\backslash D'| = 1$ and $|D'\backslash D| = 1$.

- **Add-remove Model**: The neighboring datasets are defined by adding or removing an element from the dataset. Formally, two datasets $D, D' \in \mathcal{D}^*$ are neighboring if and only if $|D \backslash D'| + |D' \backslash D| = 1$.

In this paper, we focus on the add-remove model, which is preferred in practice since it protects the size of the dataset while the swap model does not (McSherry, 2009; Wilson et al., 2020; Rogers et al., 2021; Amin et al., 2022).

For a statistic $f : \mathcal{D}^* \to \mathbb{R}$, the utility of an estimator $\hat{f} : \mathcal{D}^* \to \mathbb{R}$ for a dataset $D \in \mathcal{D}^*$ is measured by the mean squared error defined as

$$L(\hat{f}, f, D) := \mathbb{E}\left[(\hat{f}(D) - f(D))^2\right],$$

where the expectation is taken over the randomness of the estimator $\hat{f}$. Since the error of DP estimator typically decreases in the order of $1/|D|^2$ (Kulesza et al., 2024), we define the (worst-case) normalized mean squared error as

$$R(\hat{f}, f, D) = |D|^2 L(\hat{f}, f, D),$$
$$R(\hat{f}, f, n_0) = \sup_{D \in \mathcal{D}_{\geq n_0}} R(\hat{f}, f, D).$$

Our goal is to design an ε-DP estimator $\hat{f}$ that minimizes the above measures.

3 LOWER BOUND FOR VARIANCE AND COVARIANCE ESTIMATION

In this section, we provide novel minimax lower bounds for variance and covariance estimation. There have been several studies on the lower bound for covariance estimation (Dong et al., 2022; Portella and Harvey, 2024) but they focused on the swap model and the uncentered covariance estimation. Here, we show a novel lower bound for the centered covariance estimation in both the swap and add-remove models.

Theorem 3.1 (Lower bound)**.** *For $\varepsilon > 0$, any ε-DP estimators $\hat{f}_{\mathrm{var}}, \hat{f}_{\mathrm{cov}} : \mathcal{D}^* \to \mathbb{R}$ in the swap or add-remove model satisfies*

$$R(\hat{f}_{\mathrm{cov}}, \mathrm{cov}, n_0) \geq \sigma(\varepsilon)^2(1 - o(1)),$$
$$R(\hat{f}_{\mathrm{var}}, \mathrm{var}, n_0) \geq \sigma(\varepsilon)^2(1 - o(1)),$$

where $\sigma(\varepsilon)^2 = \frac{2^{-2/3}e^{-2\varepsilon/3}(1+e^{-\varepsilon})^{2/3}+e^{-\varepsilon}}{(1-e^{-\varepsilon})^2}$ and $o(1)$ term goes to 0 as $n_0 \to \infty$. In particular, if $\varepsilon \ll 1$, we have $\sigma(\varepsilon)^2 \simeq 2/\varepsilon^2$.

See Appendix B.1 for the proof. This can be shown by using a lower-bound for the sum estimation (Geng and Viswanath, 2014) and a novel reduction of the variance and covariance estimation problem to the sum estimation problem. In this paper, we focus on the high-privacy regime, where ε is small, since it is more challenging but provides meaningful privacy protection (Near et al., 2025). In the high privacy regime, the lower bound is asymptotically $2/\varepsilon^2$. As shown later, this lower bound can be achieved by the Laplace mechanism in the swap model and by our proposed mechanism in the add-remove model. That is, the leading term in the optimal mean squared error for variance and covariance estimation in high privacy regime is $2/(\varepsilon^2 n^2)$ for large n, where n is the size of the dataset. Interestingly, this error rate has the same constant in the leading term as that of mean estimation (Kulesza et al., 2024). Therefore, we can conclude that the variance and covariance estimation in the add-remove model is not harder than that in the swap model and the mean estimation.

4 OPTIMAL MECHANISM FOR SWAP MODEL

Estimation of variance and covariance in the swap model is rather straightforward since the number of records in the dataset is fixed. In this section, we briefly review the optimal mechanism for variance and covariance estimation in the swap model to compare the result with the add-remove model.

Here, we introduce the notion of l_1-sensitivity.

Definition 4.1. The l_1-sensitivity of a function $f : \mathcal{D}^* \to \mathbb{R}^d$ is defined as

$$\Delta f = \sup_{D \sim D'} \|f(D) - f(D')\|_1$$

It is well known that $f(D) + \mathrm{Lap}(\Delta f/\varepsilon)$ satisfies ε-DP (Dwork et al., 2014a). The following fact is standard in the literature (see, e.g., Aitsam (2022)).

Lemma 4.2. *The sensitivities of the empirical variance $\mathrm{var}(D)$ and covariance $\mathrm{cov}(D)$ are bounded by $1/n$ in the swap model.*

We provide the proof in Appendix C.1 for completeness. Considering the above fact, we can use the Laplace mechanism to privatize the empirical variance and covariance as follows:

$$\hat{v}_{\mathrm{swap}}(D) = \mathrm{var}(D) + \frac{1}{n} \cdot \mathrm{Lap}(1/\varepsilon),$$
$$\hat{c}_{\mathrm{swap}}(D) = \mathrm{cov}(D) + \frac{1}{n} \cdot \mathrm{Lap}(1/\varepsilon).$$

These mechanisms are ε-DP from Lemma 4.2. In addition, they achieve the lower bound in Theorem 3.1.

Proposition 4.3 (Optimality in Swap Model)**.** *The Laplace mechanisms achieve the following error bound:*

$$R(\hat{v}_{\text{swap}}, \text{var}, n_0) = R(\hat{c}_{\text{swap}}, \text{cov}, n_0) = \frac{2}{\varepsilon^2}.$$

The results follow from the fact that the variance of $\text{Lap}(1/\varepsilon)$ is $2/\varepsilon^2$. Therefore, for sufficiently small ε, we can conclude that the Laplace mechanism is optimal for variance and covariance estimation in the swap model. In the low privacy regime $\varepsilon \gg 1$, there exists a line of work on improving the utility using staircase mechanisms (Geng and Viswanath, 2014; Kulesza et al., 2025). Exploring such approaches lies beyond the scope of this work, and we leave it as an interesting direction for future research.

5 OPTIMAL MECHANISM FOR ADD-REMOVE MODEL

The situation in the add-remove model is more complicated than in the swap model since the number of records in the dataset can change and should be protected. We cannot apply the Laplace mechanism directly to the empirical variance and covariance since the sensitivity is dependent on the size of the dataset, which is not fixed in the add-remove model.

5.1 Naive Approaches do not Achieve the Lower Bound

A naive approach for the covariance estimation in the add-remove model is to use the Laplace mechanism to privatize n, $\sum_{i=1}^n x_i$, $\sum_{i=1}^n y_i$, and $\sum_{i=1}^n x_i y_i$ separately. That is, we use the privacy budget $\varepsilon/4$ to privatize each of these quantities, and obtain $\tilde{n} = n + \text{Lap}(4/\varepsilon)$, $\tilde{s}_x = \sum_{i=1}^n x_i + \text{Lap}(4/\varepsilon)$, $\tilde{s}_y = \sum_{i=1}^n y_i + \text{Lap}(4/\varepsilon)$, and $\tilde{s}_{xy} = \sum_{i=1}^n x_i y_i + \text{Lap}(4/\varepsilon)$. Then, we compute the empirical covariance as

$$\hat{c}(D) := \text{clip}\left(\frac{\tilde{s}_{xy}}{\tilde{n}} - \frac{\tilde{s}_x}{\tilde{n}} \cdot \frac{\tilde{s}_y}{\tilde{n}}, [-1/4, 1/4]\right).$$

Here, we apply clipping to improve the utility since the true empirical covariance lies in the range $[-1/4, 1/4]$. Similarly, we can estimate the variance as

$$\hat{v}(D) := \text{clip}\left(\frac{\tilde{s}_{xx}}{\tilde{n}} - \left(\frac{\tilde{s}_x}{\tilde{n}}\right)^2, [0, 1/4]\right),$$

where $\tilde{n} = n + \text{Lap}(3/\varepsilon)$, $\tilde{s}_x = \sum x_i + \text{Lap}(3/\varepsilon)$, and $\tilde{s}_{xx} = \sum x_i^2 + \text{Lap}(3/\varepsilon)$. For these naive approaches, we have the following utility bound:

Proposition 5.1. *For $\hat{c}$ and $\hat{v}$ defined above, we have*

$$R(\hat{c}, \text{cov}, n_0) \le \frac{128}{\varepsilon^2}(1 + o(1)),$$
$$R(\hat{v}, \text{var}, n_0) \le \frac{108}{\varepsilon^2}(1 + o(1)).$$

See Appendix D.1 for the proof. This does not match the lower bound in Theorem 3.1 and is not optimal.

An improved approach utilizes the following property of unnormalized variance and covariance:

Lemma 5.2. *The l_1-sensitivity of unnormalized covariance* $\text{ucov}([x, y]) := \sum_{i=1}^n (x_i - \frac{1}{n}\sum_{j=1}^n x_j)(y_i - \frac{1}{n}\sum_{j=1}^n y_j)$ *and unnormalized variance* $\text{uvar}([x]) := \sum_{i=1}^n (x_i - \frac{1}{n}\sum_{j=1}^n x_j)^2$ *is bounded by* 1 *in the add-remove model.*

See Appendix D.3 for the proof. Let $u = \text{ucov}([x, y])$ for covariance estimation and $u = \text{uvar}([x])$ for variance estimation. Then, the l_1-sensitivity of a vector $[n, u]$ is bounded by 2, and the Laplace mechanism with scale parameter $2/\varepsilon$ can be applied to privatize these quantities. That is, we can obtain $\tilde{n} = n + \text{Lap}(2/\varepsilon)$, $\tilde{u} = u + \text{Lap}(2/\varepsilon)$. Then, the empirical covariance and variance can be computed as

$$\hat{c}(D) := \text{clip}\left(\frac{\tilde{u}}{\tilde{n}}, [-1/4, 1/4]\right),$$
$$\hat{v}(D) := \text{clip}\left(\frac{\tilde{u}}{\tilde{n}}, [0, 1/4]\right).$$

For these improved approaches, we have the following utility bound:

Proposition 5.3. *For $\hat{c}$ and $\hat{v}$ defined above, we have*

$$R(\hat{c}, \text{cov}, n_0) \le \frac{17/2}{\varepsilon^2}(1 + o(1)),$$
$$R(\hat{v}, \text{var}, n_0) \le \frac{17/2}{\varepsilon^2}(1 + o(1)).$$

See Appendix D.2 for the proof. While this approach reduces the error compared to the naive method, it still suffers from large error and does not achieve the lower bound in Theorem 3.1.

5.2 Bézier Mechanism

To develop a mechanism which achieves the lower bound in Theorem 3.1, in this section, we propose a novel framework called *Bézier Mechanism*. First, we introduce the Bernstein basis and Bézier matrix (Lorentz, 2012).

Definition 5.4 (Bernstein Basis and Bézier Matrix)**.** The Bernstein basis of order $k \in \mathbb{N}$ is defined as

$$B_j^k(x) = \binom{k}{j} x^j (1 - x)^{k-j}, \quad 0 \le j \le k.$$

The Bézier matrix A is defined as

$$A_{j,l} = \begin{cases} (-1)^{l-j}\binom{k}{l}\binom{l}{j} & (j \le l) \\ 0 & (j > l) \end{cases}.$$

The Bernstein basis is widely used in computer graphics and known to form a basis for polynomials of degree at most k. In the context of DP, Aldà and Rubinstein (2017) have proposed Bernstein mechanism based on Bernstein polynomial for private function release. They utilize the property of Bernstein basis that it can approximate uniformly any continuous function. On the other hand, we utilize the following property of Bernstein basis known as a *partition of unity*:

- $B_j^k(x)$ is non-negative for all $x \in [0,1]$.
- $\sum_{j=0}^k B_j(x) = 1$.

Let $\mu(D) \in \mathbb{R}^{k+1}$ be the vector of (unnormalized) moments of $D = [x]$ up to order k, defined as

$$\mu(D) = \left(\sum_{i=1}^n x_i^0, \sum_{i=1}^n x_i^1, \dots, \sum_{i=1}^n x_i^k\right)^\top,$$

and $b(D) \in \mathbb{R}^{k+1}$ be the Bernstein representation of μ defined as

$$\begin{aligned} b(D) &= A\mu(D) \\ &= \left[\sum_{i=1}^n B_0^k(x_i), \sum_{i=1}^n B_1^k(x_i), \dots, \sum_{i=1}^n B_k^k(x_i)\right]^\top. \end{aligned}$$

From the partition of unity property of Bernstein basis, we have the following lemma:

Lemma 5.5. *The l_1-sensitivity of $b(D)$ is 1 in the add-remove model for any $k \in \mathbb{N}$.*

See Appendix D.4 for the proof. A key of the proof is the contribution of one record x to $b(D)$ is bounded by $\sum_{i=1}^k \left|B_i^k(x)\right| = 1$ from the partition of unity property. This lemma implies that $\tilde{b}(D) = b(D) + \mathrm{Lap}(1/\varepsilon)$ satisfy ε-DP. Since the Bézier matrix A is invertible, we can compute the noisy moments $\tilde{\mu}(D)$ from $\tilde{b}(D)$ as $\tilde{\mu}(D) = A^{-1}\tilde{b}(D)$. We provide the entire procedure in Algorithm 1. Bézier mechanism can be seen as a generalization of the transformed noise addition Kulesza et al. (2024), which is a special case when $k = 1$.

Algorithm 1 Bézier Mechanism

Require: Dataset $[x] \in \mathcal{D}^*$, privacy budget ε, degree k.

1: Let $b_j = \sum_{i=1}^{|x|} B_j^k(x_i)$ for $j = 0, \dots, k$.
2: Add Laplace noise: $\tilde{b}_i = b_i + \mathrm{Lap}(1/\varepsilon)$
3: Output $\hat{\mu}(D) = [\hat{\mu}_0, \dots, \hat{\mu}_k]^\top = A^{-1}\tilde{b}$.

For the utility of Algorithm 1, we have the following result:

Theorem 5.6. *For $0 \le j \le k$, let μ_j be the j-th order moment of a dataset $D \in \mathcal{D}^*$. Then, the mechanism that outputs $\hat{\mu}_j$ satisfies*

$$L(\hat{\mu}_j, \mu_j, D) = \frac{2}{\varepsilon^2 \cdot \binom{k}{j}} \sum_{l=j}^k \binom{l}{j}^2 \le \frac{2k}{\varepsilon^2}$$

In particular, $L(\hat{\mu}_k, \mu_k, D) = \frac{2}{\varepsilon^2}$, which matches the lower bound for k-th (unnormalized) moment estimation in the high-privacy regime.

See Appendix D.5 for the proof. For a naive mechanism which applies the Laplace mechanism to μ, the error is $\frac{2(k+1)^2}{\varepsilon^2}$ since the l_1-sensitivity of μ is $k+1$. Thus, the Bézier mechanism achieves a better utility-privacy trade-off. In addition, the Bézier mechanism achieves optimal error for the highest-order moment μ_k. Thus, we can estimate all moments up to order k without paying additional privacy budget.

5.3 Extension to Multivariate Cases

For cases with two or more variables in a record, the extension is natural using tensor products of Bernstein basis. For d variables, the multivariate Bernstein basis is defined as:

$$B_\alpha^k(z^{(1)}, \dots, z^{(d)}) = \prod_{j=1}^d B_{\alpha_j}^k(z^{(j)})$$

where $z = [z^{(1)}, \dots, z^{(d)}]^\top \in \mathbb{R}^d$ and $\alpha = (\alpha_1, \dots, \alpha_d)$ is a multi-index with $\alpha_j \in \{0, 1, \dots, k\}$. The corresponding Bézier matrix becomes $A_d = A \otimes A \otimes \cdots \otimes A$ (d tensor products). In a similar way as in the univariate case, let $\mu(D) \in \mathbb{R}^{(k+1)^d}$ be the vector of mixed moments of the dataset $D = [x^{(1)}, \dots, x^{(d)}]$ up to order k, defined as

$$\mu_\alpha(D) = \sum_{j=1}^n \left(x_j^{(1)}\right)^{\alpha_1} \cdots \left(x_j^{(d)}\right)^{\alpha_d},$$

and b be the Bernstein representation of μ defined as $b_\alpha(D) = \sum_{i=1}^n B_\alpha(x_i^{(1)}, \dots, x_i^{(d)})$. As in the univariate case, we have the following lemma:

Lemma 5.7. *The l_1-sensitivity of $b(D)$ is 1 in the add-remove model for any $k, d \in \mathbb{N}$.*

See Appendix D.6 for the proof. Thus, we can privatize $b(D)$ using the Laplace mechanism with scale parameter $1/\varepsilon$. Since A_d is invertible, we can recover all mixed moments up to order k from Bernstein representation.

For the utility of the multivariate Bézier mechanism, we have the following extension of Theorem 5.6:

Theorem 5.8 (Multivariate Bézier Utility). *For a multi-index $\alpha = (\alpha_1, \ldots, \alpha_d)$ with $0 \le \alpha_j \le k$, let μ_α be the mixed moment of order α of a dataset $D = [x^{(1)}, \ldots, x^{(d)}] \in \mathcal{D}^*$. Then, the multivariate Bézier mechanism that outputs $\hat{\mu}_\alpha$ satisfies*

$$L(\hat{\mu}_\alpha, \mu_\alpha, D) = \frac{2}{\varepsilon^2} \prod_{j=1}^{d} \frac{1}{\binom{k}{\alpha_j}} \sum_{l_j=\alpha_j}^{k} \binom{l_j}{\alpha_j}^2 \le \frac{2k^d}{\varepsilon^2}.$$

In particular, for the highest-order mixed moment $\mu_{(k,\ldots,k)}$, we have $L(\hat{\mu}_{(k,\ldots,k)}, \mu_{(k,\ldots,k)}, D) = \frac{2}{\varepsilon^2}$, which matches the lower bound for highest-order mixed moment estimation in the high-privacy regime.

See Appendix D.7 for the proof. This theorem shows that the multivariate Bézier mechanism can estimate all mixed moments up to order k simultaneously with controlled error, extending the univariate result to higher dimensions.

5.4 Covariance Estimation

To compute the covariance, it suffices to estimate $n, \sum_{i=1}^n x_i, \sum_{i=1}^n y_i, \sum_{i=1}^n x_i y_i$. This can be done using the two-dimensional Bézier mechanism with degree $k = 1$. That is, we compute $\tilde{b}$ as

$$\tilde{b}_{0,0} = \sum_{i=1}^{n} (1 - x_i)(1 - y_i) + \mathrm{Lap}(1/\varepsilon),$$
$$\tilde{b}_{0,1} = \sum_{i=1}^{n} (1 - x_i) y_i + \mathrm{Lap}(1/\varepsilon),$$
$$\tilde{b}_{1,0} = \sum_{i=1}^{n} x_i (1 - y_i) + \mathrm{Lap}(1/\varepsilon),$$
$$\tilde{b}_{1,1} = \sum_{i=1}^{n} x_i y_i + \mathrm{Lap}(1/\varepsilon)$$

and then, reconstruct $\tilde{s}_{xy} = \tilde{b}_{1,1}$, $\tilde{s}_x = \tilde{b}_{1,0} + \tilde{b}_{1,1}$, $\tilde{s}_y = \tilde{b}_{0,1} + \tilde{b}_{1,1}$, and $\tilde{n} = \sum_{i=0}^1 \sum_{j=0}^1 \tilde{b}_{i,j}$. Finally, we compute the noisy covariance as $\hat{c} = \mathrm{clip}\left(\frac{\tilde{s}_{xy}}{\tilde{n}} - \left(\frac{\tilde{s}_x}{\tilde{n}}\right)\left(\frac{\tilde{s}_y}{\tilde{n}}\right), [-1/4, 1/4]\right)$. Note that $\hat{c}$ is ε-DP by the post-processing theorem (Dwork et al., 2014a). We provide the entire procedure in Algorithm 2.

The utility of Algorithm 2 is evaluated as follows:

Theorem 5.9 (Utility Analysis). *For privacy budget ε, Algorithm 2 achieves*

$$R(\hat{c}, \mathrm{cov}, n_0) \le \frac{2}{\varepsilon^2}(1 + o(1)).$$

See Appendix D.8 for the proof. While the Bézier mechanism can efficiently estimate the moments of the dataset, it is far from trivial to evaluate the error in the final estimator. Surprisingly, this upper bound matches the lower bound in Theorem 3.1 for small ε. Therefore, Algorithm 2 is minimax optimal for covariance estimation in the high-privacy regime. This is because the noise contained in each individual moment estimator is correlated with one another, and many of them eventually cancel out in the final estimator.

Algorithm 2 Optimal Covariance Estimation in the Add-Remove Model

Require: Dataset $[x, y] \in \mathcal{D}^*$, privacy budget ε

1: Let $b_{1,1} = \sum_{i=1}^{|x|} x_i y_i$
2: Let $b_{1,0} = \sum_{i=1}^{|x|} x_i (1 - y_i)$
3: Let $b_{0,1} = \sum_{i=1}^{|y|} y_i (1 - x_i)$
4: Let $b_{0,0} = \sum_{i=1}^{|x|} (1 - x_i)(1 - y_i)$
5: Add Laplace noise: $\tilde{b}_{i,j} = b_{i,j} + \mathrm{Lap}(1/\varepsilon)$.
6: Calculate $\tilde{n} = \sum_{i=0}^1 \sum_{j=0}^1 \tilde{b}_{i,j}$, $\tilde{s}_x = \tilde{b}_{1,0} + \tilde{b}_{1,1}$, $\tilde{s}_y = \tilde{b}_{0,1} + \tilde{b}_{1,1}$, and $\tilde{s}_{xy} = \tilde{b}_{1,1}$
7: Output $\hat{c} = \mathrm{clip}\left(\frac{\tilde{s}_{xy}}{\tilde{n}} - \left(\frac{\tilde{s}_x}{\tilde{n}}\right)\left(\frac{\tilde{s}_y}{\tilde{n}}\right), [-1/4, 1/4]\right)$

5.5 Variance Estimation

The empirical variance can be computed from the moments $n, \sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2$ up to order 2. This can be done using the Bézier mechanism with degree $k = 2$. Then, we compute the private variance as

$$\hat{v}_b(D) = \mathrm{clip}\left(\frac{\tilde{s}_{x^2}}{\tilde{n}} - \left(\frac{\tilde{s}_x}{\tilde{n}}\right)^2, [0, 1/4]\right),$$

where $\tilde{s}_{x^2} = \tilde{b}_2$, $\tilde{s}_x = \tilde{b}_1/2 + \tilde{b}_2$, and $\tilde{n} = \sum_{i=0}^2 \tilde{b}_i$. We provide the entire procedure in Algorithm 3.

Algorithm 3 Optimal Variance Estimation in the Add-Remove Model

Require: Dataset $[x] \in \mathcal{D}^*$, privacy budget ε

1: Let $b_0 = \sum_{i=1}^{|x|} (1 - x_i)^2$
2: Let $b_1 = \sum_{i=1}^{|x|} 2x_i(1 - x_i)$
3: Let $b_2 = \sum_{i=1}^{|x|} x_i^2$
4: Add Laplace noise: $\tilde{b}_i = b_i + \mathrm{Lap}(1/\varepsilon)$
5: Compute $\tilde{n} = \sum_{i=0}^2 \tilde{b}_i$, $\tilde{s}_x = \tilde{b}_1/2 + \tilde{b}_2$, and $\tilde{s}_{x^2} = \tilde{b}_2$
6: Output $\mathrm{clip}\left(\frac{\tilde{s}_{x^2}}{\tilde{n}} - \left(\frac{\tilde{s}_x}{\tilde{n}}\right)^2, [0, 1/4]\right)$

The utility of Algorithm 3 is evaluated as follows:

Theorem 5.10 (Utility analysis). *For privacy budget ε, Algorithm 3 achieves*

$$R(\hat{v}_b, \mathrm{var}, n_0) \le \frac{2}{\varepsilon^2}(1 + o(1)).$$

See Appendix D.9 for the proof. Since this upper bound matches the lower bound in Theorem 3.1 for small ε, Algorithm 3 is minimax optimal for variance estimation in the high privacy regime.

5.6 Comparison of Optimal Mechanisms for Variance Estimation

In the previous section, we proposed the optimal mechanism for variance estimation. In this section, we provide two alternative mechanisms which also achieve the lower-bound in Theorem 3.1. While these mechanisms are optimal in the worst-case sense, we further derive instance-dependent error bounds that explicitly depend on the dataset's mean and variance. This analysis demonstrates that Algorithm 3 always provides better utility for individual datasets.

The first alternative mechanism exploits the close relationship between variance and covariance: variance estimation is a special case of covariance estimation where we set $x = y$. Accordingly, we can apply Algorithm 2 to variance estimation. That is, we can use $\hat{v}_c([x]) = \hat{c}([x, x])$ as an estimator for variance. The utility bound for $\hat{v}_c$ can be obtained from Theorem 5.9.

Corollary 5.11. *For privacy budget ε, $\hat{v}_c$ achieves*

$$R(\hat{v}_c, \mathrm{var}, n_0) \le \frac{2}{\varepsilon^2}(1 + o(1)).$$

Since this upper bound matches the lower bound of variance estimation in Theorem 3.1, we can conclude that Algorithm 2 is also minimax optimal for variance estimation in the add-remove model.

To construct another optimal mechanism for variance estimation, we revisit the improved mechanism with unnormalized covariance. Let $u = \mathrm{ucov}(x)$, and $b_0 = n - u, b_1 = u$. Then, we have the following lemma:

Lemma 5.12. *The l_1-sensitivity of $b = [n - u, u]^\top$ is 1 in the add-remove model.*

See Appendix D.10 for the proof. Therefore, $\tilde{b} := b + \mathrm{Lap}(1/\varepsilon)$ is ε-DP and noisy variance can be computed as $\hat{v}_u(D) := \mathrm{clip}\big(\frac{\tilde{u}}{\tilde{n}}, [0, 1/4]\big)$, where $\tilde{n} = \tilde{b}_0 + \tilde{b}_1$ and $\tilde{u} = \tilde{b}_2$. See Algorithm 4 for the entire procedure. As shown in the following proposition, $\hat{v}_u$ also achieves the lower bound in Theorem 3.1.

Proposition 5.13. *For privacy budget ε, $\hat{v}_u$ achieves*

$$R(\hat{v}_u, \mathrm{var}, n_0) \le \frac{2}{\varepsilon^2}(1 + o(1)).$$

See Appendix D.11 for the proof. This mechanism can be seen as a generalization of transformed noise addition for mean estimation in Kulesza et al. (2024) to variance estimation. Note that this approach cannot be applied to covariance estimation since the sensitivity space of $\mathrm{ucov}(D)$ is different from that of $\mathrm{uvar}(D)$, and the sensitivity of $b = [n - \mathrm{ucov}([x, y]), \mathrm{ucov}([x, y])]$ is not bounded by 1.

Algorithm 4 Transformed Noise Addition for Variance Estimation

Require: Dataset $[x] \in \mathcal{D}^*$, privacy budget ε

1: Let $b_0 = n - \mathrm{uvar}(D)$ and $b_1 = \mathrm{uvar}(D)$
2: Add independent Laplace noise $Z_i \sim \mathrm{Lap}(1/\varepsilon)$ to obtain $\tilde{b}_i = b_i + Z_i$ for $i = 0, 1$
3: Compute $\tilde{n} = \tilde{b}_0 + \tilde{b}_1$, $\tilde{v}_x = \tilde{b}_1$.
4: Output $\mathrm{clip}\big(\frac{\tilde{v}_x}{\tilde{n}}, [0, 1/4]\big)$

While Algorithm 2, 3, and 4 achieve the same error rate in the worst case, they may perform differently in practice due to their distinct mechanisms and the specific characteristics of the data. To see this, we compare the constants in the leading term of the instance-wise error $R(\cdot, \mathrm{var}, D)$.

Theorem 5.14. *For the mean $r = \frac{1}{n}\sum_{i=1}^n x_i$ and the variance $v = \frac{1}{n}\sum_{i=1}^n (x_i - r)^2$, let us define*

$$\begin{aligned} C_b(r, v) &= 3v^2 - 2(3r^2 - 3r + 1)v \\ &\quad + 3r^4 - 6r^3 + 7r^2 - 4r + 1, \\ C_c(r, v) &= (1 - 2r + 2r^2)^2 - 2v(1 - 2r)^2 + 4v^2, \\ C_u(r, v) &= v^2 + (1 - v)^2. \end{aligned}$$

Then, normalized errors $R(\cdot, f, D)$ for Algorithm 2, 3, and 4 satisfies

$$\begin{aligned} R(\hat{v}_b, \mathrm{var}, D) &\le \frac{2}{\varepsilon^2}(C_b(r, v) + o(1)), \\ R(\hat{v}_c, \mathrm{var}, D) &\le \frac{2}{\varepsilon^2}(C_c(r, v) + o(1)), \\ R(\hat{v}_u, \mathrm{var}, D) &\le \frac{2}{\varepsilon^2}(C_u(r, v) + o(1)) \end{aligned}$$

for any $D = [x] \in \mathcal{D}_{\ge n_0}$. Furthermore, for any possible r and v, we have

$$C_b(r, v) \le C_c(r, v) \le C_u(r, v).$$

See Appendix D.12 for the proof. Note that r and v are not independent and should satisfy certain constraints for the existence of the dataset. Since $\frac{2}{\varepsilon^2}C_\cdot(r, v)$ is a first-order approximation of the error of each algorithm if we ignore the effect of clipping, we can use it to analyze the performance of the algorithms. We see that Algorithm 3 consistently outperforms the other algorithms for datasets with any possible mean and variance. We empirically confirm the superiority of Algorithm 3 in Section 7 with finite samples.

6 EXTENSION TO GENERAL MOMENT-BASED STATISTICS

While our main focus is on the variance and covariance estimation, our proposed Bézier mechanism is versatile and can be utilized to estimate other statistics beyond variance and covariance. In general, given a statistic $f(\mu(D))$ which depends on the (mixed) moments $\mu(D)$ up to order k, we can use the Bézier mechanism with degree k to estimate $\mu(D)$ by $\hat{\mu}(D)$ and compute an ε-DP estimator $\hat{f}(D) = f(\hat{\mu}(D))$ by the post-processing theorem. This formulation includes higher-order centered moments, skewness, kurtosis, and correlation. We provide a detailed procedure in Algorithm 5.

Algorithm 5 Bézier Mechanism for General Statistics

Require: Dataset $D \in \mathcal{D}^*$, privacy budget ε, order k, range $[l, u]$.

1: Let $b_\alpha(D) = \sum_{i=1}^n B_{\alpha_1}(x_{i,1}) \cdot \cdots \cdot B_{\alpha_d}(x_{i,d})$ for $\alpha_i = 0, \ldots, k$.

2: Add Laplace noise $\tilde{b}_\alpha(D) = b_\alpha(D) + \mathrm{Lap}(1/\varepsilon)$.

3: Compute $\tilde{\mu}(D) = (A_d)^{-1}\tilde{b}(D)$.

4: Output $\hat{f}(D) := \mathrm{clip}(f(\tilde{\mu}(D)), [l, u])$

Assume that f is differentiable. Then, if we ignore the clipping, the first order approximation of error $\hat{f}(D) - f(D)$ is given by $\mathcal{E}_{\mathrm{bezier}} := \nabla f(\mu(D))^\top(\hat{\mu}(D) - \mu(D)) = \nabla f(\mu(D))^\top(A_d^{-1}\mathrm{Lap}(1/\varepsilon))$. On the other hand, for a naive mechanism that applies the Laplace mechanism to each moment independently, the first order approximation of the error is given by $\mathcal{E}_{\mathrm{naive}} := \nabla f(\mu(D))^\top(\mathrm{Lap}((k+1)^d/\varepsilon))$ since the l_1-sensitivity of $\mu(D)$ is $(k+1)^d$. For these two errors, we have the following results:

Theorem 6.1. *For any differentiable function $f : \mathbb{R}^{(k+1)^d} \to \mathbb{R}$ and dataset $D \in \mathcal{D}^*$, we have*

$$\mathbb{E}\left[\mathcal{E}_{\mathrm{bezier}}(D)^2\right] \le \mathbb{E}\left[\mathcal{E}_{\mathrm{naive}}(D)^2\right].$$

See Appendix D.13 for the proof. This theorem indicates that the Bézier mechanism always outperforms the naive mechanism in terms of the first-order approximation of error.

7 NUMERICAL EXPERIMENTS

To see the empirical performance of our proposed mechanisms and compare it with naive approaches, we conduct numerical experiments on synthetic datasets. The results are averaged over 10,000 independent runs and MSE is normalized by $\frac{\varepsilon^2|D|^2}{2}$ to keep the results in a similar range for different parameters. See Appendix E for details of the experimental setup including dataset generation procedure.

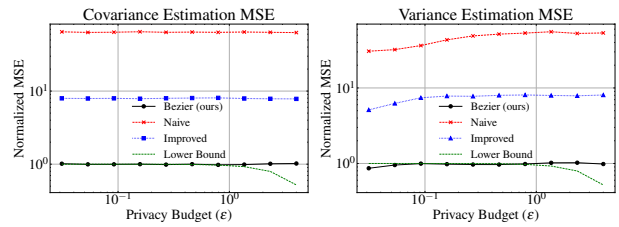

Figure 1: Comparison of worst-case performance for covariance and variance estimation.

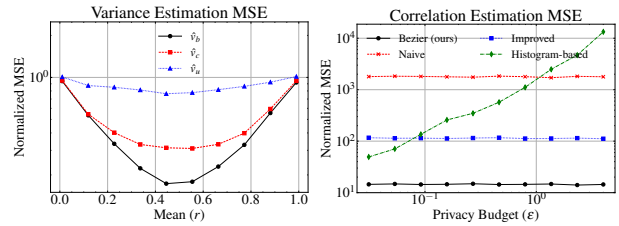

Figure 2: Comparison for variance estimation (left) and correlation estimation (right).

Bézier Mechanism Outperforms Baselines We compare our proposed mechanism in Algorithm 2 and Algorithm 3 with baselines including 1) naive approach: applying the Laplace mechanism to n, $\sum_{i=1}^n x_i$, $\sum_{i=1}^n y_i$, and $\sum_{i=1}^n x_i y_i$ (or n, $\sum x_i$, $\sum x_i^2$) separately, and 2) improved approach: applying Laplace mechanism to estimate n and $\mathrm{ucov}(D)$ (or $\mathrm{uvar}(D)$) separately. We generate a dataset of size $n = 10,000$ with different mean and variance (or covariance) and estimate the worst-case error for each method. Fig. 7 shows that our proposed mechanisms reduces the error to less than about 1/10 compared to baselines by achieving the performance close to the lower bound of Theorem 3.1 in the recommended range of privacy parameter $\varepsilon \le 1$ (Near et al., 2025).

Comparison of Optimal Mechanisms for Variance Estimation We compare our proposed mechanisms for variance estimation in Algorithm 2, Algorithm 3, and Algorithm 4 with different mean r. We generate a dataset of size $n = 10,000$ with mean r and variance $v = (r - r^2)/2$. We fix the privacy budget $\varepsilon = 1$ and compute the error of the three mechanisms. While both achieve the optimal error in the worst case, Fig. 7 (left) shows that Algorithm 3, based on Bézier mechanism, is more efficient than the other algorithms especially in the case $r \simeq 0.5$.

Bézier Mechanism is Effective Beyond Variance and Covariance Estimation As an example of moment-based statistics other than variance and covariance, we consider the Pearson correlation coefficient defined as $\mathrm{corr}(D) = \frac{\mathrm{cov}([x,y])}{\sqrt{\mathrm{var}([x])\cdot\mathrm{var}([y])}}$. This quantity

can be computed with mixed moments up to the $k = 2$ order. Thus, we can use the two-dimensional Bézier mechanism with degree $k = 2$ to estimate the correlation. Figure 7 (right) shows the results of comparing with baselines including 1) naive approach: applying Laplace mechanism to privatize the moments independently, 2) improved approach: applying Algorithm 2 and 3 to privatize $\mathrm{cov}([x, y])$ and $\mathrm{var}([x])$, $\mathrm{var}([y])$ separately, and 3) histogram-based approach: calculating the correlation from a privatized 2D histogram. We estimate 2D histogram using the Laplace mechanism and optimize the bin size for each ε. See Appendix E for details. We see that Algorithm 5 achieves a better utility-privacy trade-off with large margin.

8 CONCLUSION

In this paper, we study the variance and covariance estimation with DP in the add-remove model. We propose efficient mechanisms for variance and covariance estimation based on a novel framework called the Bézier mechanism, and prove that our methods are minimax optimal in the high-privacy regime. We further present two alternative optimal mechanisms and show that, beyond worst-case guarantees, the Bézier-based estimator achieves uniformly smaller leading constants for any data distribution. Finally, we demonstrate that the Bézier mechanism is also effective for the estimation of other statistics, such as correlation.

References

Aitsam, M. (2022). Differential Privacy Made Easy. In *2022 International Conference on Emerging Trends in Electrical, Control, and Telecommunication Engineering (ETECTE)*, pages 1–7. IEEE.

Aldà, F. and Rubinstein, B. (2017). The Bernstein Mechanism: Function Release under Differential Privacy. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1).

Amin, K., Dick, T., Kulesza, A., Munoz, A., and Vassilvitskii, S. (2019a). Differentially Private Covariance Estimation. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Amin, K., Gillenwater, J., Joseph, M., Kulesza, A., and Vassilvitskii, S. (2022). Plume: Differential privacy at scale. *arXiv preprint arXiv:2201.11603*.

Amin, K., Kulesza, A., Munoz, A., and Vassilvtiskii, S. (2019b). Bounding User Contributions: A Bias-Variance Trade-off in Differential Privacy. In Chaudhuri, K. and Salakhutdinov, R., editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 263–271. PMLR.

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

Dalenius, T. (1977). Towards a methodology for statistical disclosure control. *statistik Tidskrift*, 15(429-444):2–1.

Dong, W., Liang, Y., and Yi, K. (2022). Differentially private covariance revisited. *Advances in Neural Information Processing Systems*, 35:850–861.

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

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

Dwork, C., Talwar, K., Thakurta, A., and Zhang, L. (2014b). Analyze gauss: optimal bounds for privacy-preserving principal component analysis. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 11–20.

Geng, Q. and Viswanath, P. (2014). The optimal mechanism in differential privacy. In *2014 IEEE international symposium on information theory*, pages 2371–2375. IEEE.

Gillenwater, J., Joseph, M., and Kulesza, A. (2021). Differentially private quantiles. In *International Conference on Machine Learning*, pages 3713–3722. PMLR.

Imola, J., Boninsegna, F., Keller, H., Aamand, A., Chowdhury, A. R., and Pagh, R. (2025). Differentially private quantiles with smaller error. *arXiv preprint arXiv:2505.13662*.

Kalinin, N. P., Upadhyay, J., and Lampert, C. H. (2025). Continual release moment estimation with differential privacy. *arXiv preprint arXiv:2502.06597*.

Karwa, V. and Vadhan, S. (2018). Finite Sample Differentially Private Confidence Intervals. In *9th Innovations in Theoretical Computer Science Conference (ITCS 2018)*, pages 44–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

Kulesza, A., Suresh, A. T., and Wang, Y. (2024). Mean Estimation in the Add-Remove Model of Differential Privacy. In *Forty-first International Conference on Machine Learning*.

Kulesza, A., Suresh, A. T., and Wang, Y. (2025). General Staircase Mechanisms for Optimal Differential Privacy. In Li, Y., Mandt, S., Agrawal, S., and Khan,

E., editors, *Proceedings of The 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pages 4564–4572. PMLR.

Lorentz, G. G. (2012). *Bernstein polynomials*. American Mathematical Soc.

McSherry, F. D. (2009). Privacy integrated queries: an extensible platform for privacy-preserving data analysis. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of data*, pages 19–30.

Near, J. P., Darais, D., Lefkovitz, N., Howarth, G. S., et al. (2025). *Guidelines for evaluating differential privacy guarantees*. US Department of Commerce, National Institute of Standards and Technology.

Portella, V. S. and Harvey, N. (2024). Lower bounds for private estimation of Gaussian covariance matrices under all reasonable parameter regimes. *arXiv preprint arXiv:2404.17714*.

Rameshwar, V. A., Tandon, A., and Sharma, A. (2025). Optimal Tree-Based Mechanisms for Differentially Private Approximate CDFs. In *2025 17th International Conference on COMmunication Systems and NETworks (COMSNETS)*, pages 1269–1274.

Rogers, R., Subramaniam, S., Peng, S., Durfee, D., Lee, S., Kancha, S. K., Sahay, S., and Ahammad, P. (2021). Linkedin's audience engagements API: A privacy preserving data analytics system at scale. *Journal of Privacy and Confidentiality*, 11:3.

Simsek, Y. (2013). Functional equations from generating functions: a novel approach to deriving identities for the Bernstein basis functions. *Fixed Point Theory and Applications*, 2013(1):80.

Swanberg, M., Globus-Harris, I., Griffith, I., Ritz, A., Groce, A., and Bray, A. (2019). Improved Differentially Private Analysis of Variance. *Proceedings on Privacy Enhancing Technologies*.

Wilson, R. J., Zhang, C. Y., Lam, W., Desfontaines, D., Simmons-Marengo, D., and Gipson, B. (2020). Differentially Private SQL with Bounded User Contribution. *Proceedings on Privacy Enhancing Technologies*.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes] We provide a detailed description of the mathematical framework and the assumptions underlying our approach.
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes] We analyze the sample complexity of our proposed algorithms.
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes] We provide anonymized source code in the supplementary material.
2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes]
   (b) Complete proofs of all theoretical results. [Yes] We provide complete proofs in the appendix.
   (c) Clear explanations of any assumptions. [Yes]
3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes] We provide anonymized source code in the supplementary material.
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes] We provide detailed explanations in Appendix E.
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes]
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Yes] We provide a description in Appendix E.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Not Applicable]
   (b) The license information of the assets, if applicable. [Not Applicable]
   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]
   (d) Information about consent from data providers/curators. [Not Applicable]
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]
5. If you used crowdsourcing or conducted research with human subjects, check if you include:

(a) The full text of instructions given to participants and screenshots. [Not Applicable]

(b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]

(c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

Supplementary Materials

A Auxiliary Results

Lemma A.1. *For $x, y \in [0,1]^n$, we have*

$$\begin{aligned} -1/4 \le \mathrm{cov}([x,y]) \le 1/4, \\ 0 \le \mathrm{var}([x]) \le 1/4. \end{aligned}$$

Proof. From the Cauchy-Schwarz inequality, we have

$$\mathrm{cov}([x,y])^2 \le \mathrm{var}([x]) \cdot \mathrm{var}([y]).$$

Thus, it suffices to show that $0 \le \mathrm{var}([x]) \le 1/4$. From the definition of $\mathrm{var}([x])$, we have $0 \le \mathrm{var}([x])$. Let $r = \frac{1}{n}\sum_{i=1}^n x_i$. Since $x_i^2 \le x_i$ from $x_i \in [0,1]$, we have

$$\frac{1}{n}\sum_{i=1}^n x_i^2 \le r$$

Therefore, we obtain

$$\begin{aligned} \mathrm{var}([x]) &= \frac{1}{n}\sum_{i=1}^n x_i^2 - r^2 \\ &\le r - r^2 = 1/4 - (r - 1/2)^2 \le 1/4 \end{aligned}$$

since $r \in [0,1]$. This completes the proof. □

Lemma A.2. *Let X, Y be random variables with bounded second moments. Then, we have*

$$(1-p)\mathbb{E}\left[X^2\right] + (1-1/p)\mathbb{E}\left[Y^2\right] \le \mathbb{E}\left[(X+Y)^2\right] \le (1+p)\mathbb{E}\left[X^2\right] + (1+1/p)\mathbb{E}\left[Y^2\right]$$

for any $p > 0$.

Proof. For any $p > 0$, we have

$$\begin{aligned} \mathbb{E}\left[(X+Y)^2\right] &= \mathbb{E}\left[X^2\right] + 2\mathbb{E}\left[XY\right] + \mathbb{E}\left[Y^2\right] \\ &\le \mathbb{E}\left[X^2\right] + 2\sqrt{\mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right]} + \mathbb{E}\left[Y^2\right] \\ &\le (1+p)\mathbb{E}\left[X^2\right] + (1+1/p)\mathbb{E}\left[Y^2\right]. \end{aligned}$$

The first inequality follows from the Cauchy-Schwarz inequality, and the second inequality follows from the AM-GM inequality. On the other hand, we have

$$\begin{aligned} \mathbb{E}\left[(X+Y)^2\right] &= \mathbb{E}\left[X^2\right] + 2\mathbb{E}\left[XY\right] + \mathbb{E}\left[Y^2\right] \\ &\ge \mathbb{E}\left[X^2\right] - 2\sqrt{\mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right]} + \mathbb{E}\left[Y^2\right] \\ &\ge (1-p)\mathbb{E}\left[X^2\right] + (1-1/p)\mathbb{E}\left[Y^2\right]. \end{aligned}$$

The first inequality follows from the Cauchy-Schwarz inequality, and the second inequality follows from the AM-GM inequality. □

Lemma A.3. *For $x, y \in [0,1]^n$, we have*

$$\max(0, r_x + r_y - 1) \le r_{xy} \le \min(r_x, r_y),$$

where $r_x = \frac{1}{n}\sum_{i=1}^n x_i$, $r_y = \frac{1}{n}\sum_{i=1}^n y_i$, and $r_{xy} = \frac{1}{n}\sum_{i=1}^n x_i y_i$.

Proof. For the upper bound, we have

$$\begin{aligned} r_{xy} &= \frac{1}{n}\sum_{i=1}^n x_i y_i \\ &= \frac{1}{n}\sum_{i=1}^n x_i \cdot y_i \\ &\le \frac{1}{n}\sum_{i=1}^n x_i \cdot 1 = r_x. \end{aligned}$$

Similarly, we can show that $r_{xy} \le r_y$. For the lower bound, we have

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n (1-x_i)(1-y_i) &= \frac{1}{n}\sum_{i=1}^n (1 - x_i - y_i + x_i y_i) \\ &= 1 - r_x - r_y + r_{xy} \ge 0, \end{aligned}$$

which implies that

$$r_{xy} \ge r_x + r_y - 1.$$

Thus, we conclude that

$$\max(0, r_x + r_y - 1) \le r_{xy} \le \min(r_x, r_y).$$

□

Lemma A.4. *For $f : \mathbb{R}^d \to \mathbb{R}$ and $A_n = \{\|Z\|_\infty \le n^{1/4}\} \subseteq \mathbb{R}^d$, assume that f is twice continuously differentiable on $\{Z/n \mid Z \in A_n\}$ and $\nabla f(Z/n), \nabla^2 f(Z/n) = O(1)$ for $Z \in A_n$, and a random variable Z satisfies $\mathrm{Prob}\,(Z \in A_n) = 1 - o(1/n^2)$ and $\mathbb{E}\left[\|Z\|^2\right] = O(1)$. Then, we have*

$$n^2 \cdot \mathbb{E}_Z\left[\|\mathrm{clip}(f_n(Z/n), [l,u]) - f(0)\|^2\right] \le \mathbb{E}_Z\left[\|\Delta(Z)\|^2\right] + o(1)$$

for any l, u such that $f(0) \in [l,u]$, where $\Delta(Z) = \langle \nabla f(0), Z\rangle$.

Proof. From the intermediate value theorem, we have

$$\begin{aligned} f(Z/n) - f(0) &= \langle \nabla f(0), Z/n\rangle + \frac{1}{2n^2} Z^\top \nabla^2 f(\xi) Z \\ &= \langle \nabla f(0), Z/n\rangle + O(\|Z\|_2^2/n^2) \\ &= \Delta(Z/n) + O(n^{-3/2}). \end{aligned}$$

for some ξ on the line segment between 0 and Z/n. Therefore, we obtain

$$\begin{aligned} n^2 \cdot \mathbb{E}\left[\|\mathrm{clip}(f(Z/n), [l,u]) - f(0)\|^2\right] &= n^2 \cdot \mathbb{E}\left[\|\mathrm{clip}(f(Z/n), [l,u]) - f(0)\|^2 \cdot 1_{Z \in A_n}\right] \\ &\quad + n^2 \cdot \mathbb{E}\left[\|\mathrm{clip}(f(Z/n), [l,u]) - f(0)\|^2 \cdot (1 - 1_{Z \in A_n})\right], \\ &\le n^2 \cdot \mathbb{E}\left[\|f(Z/n) - f(0)\|^2 \cdot 1_{Z \in A_n}\right] + n^2 \cdot |u - l| \cdot \mathrm{Prob}\,(Z \notin A_n), \\ &\le n^2 \cdot \mathbb{E}\left[\|f(Z/n) - f(0)\|^2 \cdot 1_{Z \in A_n}\right] + |u - l| \cdot o(1). \end{aligned}$$

The first term can be bounded as follows:

$$\begin{aligned} n^2 \cdot \mathbb{E}_{Z\sim\nu_n}\left[\|f(Z)-f(0)\|^2 \cdot 1_{Z\in A_n}\right] &= n^2 \cdot \mathbb{E}\left[\left(\Delta(Z/n)+O(n^{-3/2})\right)^2 \cdot 1_{Z\in A_n}\right] \\ &\le \mathbb{E}\left[\left(\Delta(Z)+O(n^{-1/2})\right)^2\right] \\ &\le (1+n^{-1/2})\mathbb{E}\left[\|\Delta(Z)\|^2\right] + (1+n^{1/2})O(n^{-1}) \\ &\le \mathbb{E}\left[\|\Delta(Z)\|^2\right] + o(1). \end{aligned}$$

The second inequality follows from Lemma A.2 and the last inequality follows from $\mathbb{E}\left[\|\Delta(Z)\|^2\right] = O(1)$. This completes the proof. □

B Proofs for Section 3

B.1 Proof of Theorem 3.1

Proof. We prove the theorem by reducing the problem to the sum estimation problem.

First, we consider the variance estimation. Given ε-DP mechanism $\hat{v} : \mathcal{D}^* \to \mathbb{R}$, we construct the following mechanism $\hat{s}_k : \mathcal{D}^* \to \mathbb{R}$ for some $k \in \mathbb{N}$:

$$\hat{s}_k(D) = k\hat{v}(D_k),$$

where $D = [x]$ and $D_k = [[\sqrt{x_1}, \sqrt{x_2}, \ldots, \sqrt{x_{|D|}}, \underbrace{0, \ldots, 0}_{k}]^\top]$. We show that $\hat{s}_k$ is an ϵ-DP estimator of the sum of the dataset x. For neighboring datasets $D, D' \in \mathcal{D}^*$ and $S \subset \mathbb{R}$, we have

$$\begin{aligned} \mathrm{Prob}\left(\hat{s}_k(D) \in S\right) &= \mathrm{Prob}\left(k\hat{v}(D_k) \in S\right) \\ &= \mathrm{Prob}\left(\hat{v}(D_k) \in \{s/k \mid s \in S\}\right) \\ &\le e^\epsilon \mathrm{Prob}\left(\hat{v}(D'_k) \in \{s/k \mid s \in S\}\right) \\ &= e^\epsilon \mathrm{Prob}\left(k\hat{v}(D'_k) \in S\right) \\ &= e^\epsilon \mathrm{Prob}\left(\hat{s}_k(D') \in S\right), \end{aligned}$$

where the inequality follows from the definition of differential privacy and the fact that D_k and D'_k are neighboring datasets if D and D' are neighboring datasets.

From Lemma 5.1 in Kulesza et al. (2024), we have

$$\sup_{D\in\mathcal{D}^*} \mathbb{E}\left[(\hat{s}_k(D) - \mathrm{sum}(D))^2\right] \ge \sigma(\varepsilon)^2(1-o(1)),$$

where $\mathrm{sum}(D)$ is the true sum of dataset D. Thus, there exists $D \in \mathcal{D}^*$ such that

$$\mathbb{E}\left[(\hat{s}_k(D) - \mathrm{sum}(D))^2\right] \ge \sigma(\varepsilon)^2(1-o(1)). \tag{1}$$

The error $|k \cdot \mathrm{var}(D_k) - \mathrm{sum}(D)|$ can be evaluated as follows:

$$\begin{aligned} |k \cdot \mathrm{var}(D_k) - \mathrm{sum}(D)| &= \left| \frac{k}{|D|+k}\sum_{i=1}^{|D|} x_i - \frac{k}{(|D|+k)^2}\left(\sum_{i=1}^{|D|}\sqrt{x_i}\right)^2 - \sum_{i=1}^{|D|} x_i \right| \\ &\le \frac{1}{k}\left(\sum_{i=1}^{|D|}\sqrt{x_i}\right)^2 + \left|\frac{|D|}{|D|+k}\sum_{i=1}^{|D|} x_i\right| \\ &\le \frac{2|D|^2}{k} \end{aligned}$$

Therefore, we obtain

$$\begin{aligned}
\mathbb{E}\left[|D_k|^2(\hat{v}(D_k)-\mathrm{var}(D_k))^2\right] &\geq \mathbb{E}\left[k^2(\hat{v}(D_k)-\mathrm{var}(D_k))^2\right] \\
&= \mathbb{E}\left[(\hat{s}_k(D)-\mathrm{sum}(D)+\mathrm{sum}(D)-k\mathrm{var}(D_k))^2\right] \\
&\geq (1-p)\mathbb{E}\left[(\hat{s}_k(D)-\mathrm{sum}(D))^2\right]+(1-1/p)\mathbb{E}\left[(\mathrm{sum}(D)-k\mathrm{var}(D_k))^2\right] \\
&\geq (1-p)\mathbb{E}\left[(\hat{s}_k(D)-\mathrm{sum}(D))^2\right]-\frac{4(1-1/p)|D|^4}{k^2} \\
&\geq (1-p)\sigma(\varepsilon)^2(1-o(1))-\frac{4(1-1/p)|D|^4}{k^2}
\end{aligned}$$

for any $p>0$. The second inequality follows from Lemma A.2 and the last inequality follows from Eq. (1). By setting $k=\Omega(|D|^4)$, $p=1/k$, and $n_0=k$, we have

$$R(\hat{v},n_0)\geq \mathbb{E}\left[|D_k|^2(\hat{v}(D_k)-\mathrm{var}(D_k))^2\right]\geq \sigma(\varepsilon)^2(1-o(1)).$$

Since variance estimation is a special case of covariance estimation where $D=[x,x]$, we can use the same argument to show the lower bound for covariance estimation. This completes the proof. □

C Proofs for Section 4

C.1 Proof for Lemma 4.2

C.1.1 Variance Case

Proof. Let $[x],[y]$ be neighboring datasets, i.e., $x_i=y_i$ for all $i\neq j$ and $x_j\neq y_j$. Without loss of generality, we assume $j=n$ and $x_n\geq y_n$. To simplify the notation, we denote $r=\frac{1}{n}\sum_{i=1}^{n-1}x_i$. Then, we have

$$\begin{aligned}
\mathrm{var}(x)-\mathrm{var}(y) &= \frac{1}{n}\sum_{i=1}^{n}x_i^2-\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)^2-\left(\frac{1}{n}\sum_{i=1}^{n}y_i^2-\left(\frac{1}{n}\sum_{i=1}^{n}y_i\right)^2\right) \\
&= \frac{1}{n}(x_n^2-y_n^2)-\left((r+x_n/n)^2-(r+y_n/n)^2\right) \\
&= \frac{1}{n}(x_n^2-y_n^2)-(2rx_n/n+x_n^2/n^2-2ry_n/n-y_n^2/n^2) \\
&= \left(\frac{1}{n}-\frac{1}{n^2}\right)(x_n^2-y_n^2)-\frac{2r}{n}(x_n-y_n).
\end{aligned}$$

Since $r\in[0,(n-1)/n]$ and $x_n\geq y_n$, we obtain

$$\begin{aligned}
-1/n &\leq -\frac{n-1}{n^2} \\
&\leq \frac{n-1}{n^2}\left((x_n-1)^2-(y_n-1)^2\right) \\
&= \left(\frac{1}{n}-\frac{1}{n^2}\right)(x_n^2-y_n^2)-\frac{2(n-1)}{n^2}(x_n-y_n) \\
&\leq \mathrm{var}(x)-\mathrm{var}(y) \\
&\leq \left(\frac{1}{n}-\frac{1}{n^2}\right) \\
&\leq \frac{1}{n}.
\end{aligned}$$

This completes the proof. □

C.1.2 Covariance Case

Proof. Let $[x, y], [z, w]$ be neighboring datasets, i.e., $(x_i, y_i) = (z_i, w_i)$ for all $i \neq j$ and $(x_j, y_j) \neq (z_j, w_j)$. Without loss of generality, we assume $j = n$ and $x_n \geq z_n$. To simplify the notation, we denote $r_x = \frac{1}{n}\sum_{i=1}^{n-1} x_i$ and $r_y = \frac{1}{n}\sum_{i=1}^{n-1} y_i$. Then, we have

$$\begin{aligned}
\mathrm{cov}(x, y) - \mathrm{cov}(z, w) &= \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) - \left(\frac{1}{n}\sum_{i=1}^{n} z_i w_i - \left(\frac{1}{n}\sum_{i=1}^{n} z_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} w_i\right)\right) \\
&= \frac{1}{n}(x_n y_n - z_n w_n) - ((r_x + x_n/n)(r_y + y_n/n) - (r_x + z_n/n)(r_y + w_n/n)) \\
&= \frac{1}{n}(x_n y_n - z_n w_n) - (r_x y_n/n + r_y x_n/n + x_n y_n/n^2 - r_x w_n/n - r_y z_n/n - z_n w_n/n^2) \\
&= \left(\frac{1}{n} - \frac{1}{n^2}\right)(x_n y_n - z_n w_n) - \frac{1}{n}(r_x(y_n - w_n) + r_y(x_n - z_n)).
\end{aligned}$$

If $y_n \geq w_n$, we have

$$\begin{aligned}
-1/n \leq& -\frac{n-1}{n^2} \\
&\leq \frac{n-1}{n^2}((x_n - 1)(y_n - 1) - (z_n - 1)(w_n - 1)) \\
&\leq \mathrm{cov}(x, y) - \mathrm{cov}(z, w) \\
&\leq \left(\frac{1}{n} - \frac{1}{n^2}\right)(x_n y_n - z_n w_n) \\
&\leq \left(\frac{1}{n} - \frac{1}{n^2}\right) \\
&\leq \frac{1}{n}.
\end{aligned}$$

On the other hand, if $y_n < w_n$, we have

$$\begin{aligned}
-1/n \leq& -\frac{n-1}{n^2} \\
&\leq \frac{n-1}{n^2}(x_n(y_n - 1) - z_n(w_n - 1)) \\
&\leq \mathrm{cov}(x, y) - \mathrm{cov}(z, w) \\
&\leq \left(\frac{1}{n} - \frac{1}{n^2}\right)((x_n - 1)y_n - (z_n - 1)w_n) \\
&\leq \left(\frac{1}{n} - \frac{1}{n^2}\right) \\
&\leq \frac{1}{n}.
\end{aligned}$$

Combining both cases, we obtain the result. □

D Proofs for Section 5

D.1 Proof for Proposition 5.1

D.1.1 Covariance Case

Let $Z = [Z_n, Z_x, Z_y, Z_{xy}]$ and $f(Z/n) = \frac{r_{xy} + Z_{xy}/n}{1 + Z_n/n} - \frac{r_x + Z_x/n}{1 + Z_n/n}\frac{r_y + Z_y/n}{1 + Z_n/n}$. Note that $f(0)$ is the true covariance and $\hat{c}(D) = \mathrm{clip}(f(Z/n), [-1/4, 1/4])$, where $Z \sim \mathrm{Lap}(4/\varepsilon)$. It is easy to see that $\|\nabla f(Z/n)\|, \left\|\nabla^2 f(Z/n)\right\| = O(1)$ for $Z \in A_n$. Then, from Lemma A.4, it suffices to show 1) $\mathrm{Prob}\,(Z \in A_n) = 1 - o(1/n^2)$, 2) $\mathbb{E}\left[\|\Delta(Z)\|^2\right] \leq \frac{128}{\varepsilon^2}$. We prove these two conditions in the following lemmas:

Lemma D.1. *We have* $\text{Prob}\,(Z \in A_n) = 1 - o(1/n^2)$.

Proof. Since $Z_n, Z_x, Z_y, Z_{xy} \sim \text{Lap}(4/\varepsilon)$, we have

$$\begin{aligned}\text{Prob}\,(Z \notin A_n) &\leq \text{Prob}\left(|Z_n| > n^{1/4}\right) + \text{Prob}\left(|Z_x| > n^{1/4}\right) + \text{Prob}\left(|Z_y| > n^{1/4}\right) + \text{Prob}\left(|Z_{xy}| > n^{1/4}\right) \\ &= O(\exp(-\varepsilon n^{1/4}/4)) = o(1/n^2).\end{aligned}$$

This completes the proof. □

Lemma D.2. *We have*

$$\mathbb{E}\left[\|\Delta(Z)\|^2\right] \leq \frac{128}{\varepsilon^2}.$$

Proof. From simple calculation, we have

$$\Delta(Z) = (-r_{xy} + 2r_x r_y)Z_n - r_y Z_x - r_x Z_y + Z_{xy}.$$

From Lemma A.3, we have

$$\begin{aligned}2r_x r_y - r_{xy} &\leq 2r_x r_y - \max(0, r_x + r_y - 1) \\ &= r_x r_y + \min(r_x r_y, (1 - r_x)(1 - r_y)) \\ &\leq r_x r_y + (1 - r_x)(1 - r_y) \\ &\leq 1, \\ 2r_x r_y - r_{xy} &\geq 2r_x r_y - \min(r_x, r_y) \geq -1.\end{aligned}$$

Since $Z \sim \text{Lap}(4/\varepsilon)$, we obtain

$$\begin{aligned}\mathbb{E}\left[\|\Delta(Z)\|^2\right] &= 32/\varepsilon^2\big((-r_{xy} + 2r_x r_y)^2 + r_y^2 + r_x^2 + 1\big) \\ &\leq 128/\varepsilon^2.\end{aligned}$$

The equality holds when $r_x = r_y = r_{xy} = 1$. □

D.1.2 Variance Case

Let $Z = [Z_n, Z_x, Z_{xx}]$ and $f(Z/n) = \frac{r_{xx} + Z_{xx}/n}{1 + Z_n/n} - \left(\frac{r_x + Z_x/n}{1 + Z_n/n}\right)^2$. As in the case of covariance, it suffices to show 1) $\text{Prob}\,(Z \in A_n) = 1 - o(1/n^2)$, 2) $\mathbb{E}\left[\|\Delta(Z)\|^2\right] = \frac{108}{\varepsilon^2}$. The proof of 1) can be done in a similar way as Lemma D.1. Thus, we prove 2) in the following lemma:

Lemma D.3. *We have*

$$\mathbb{E}\left[\|\Delta(Z)\|^2\right] \leq \frac{108}{\varepsilon^2}.$$

Proof. From simple algebra, we have

$$\Delta(Z) = (-r_{xx} + 2r_x^2)Z_n - 2r_x Z_x + Z_{xx}.$$

Since $Z \sim \text{Lap}(3/\varepsilon)$, we obtain

$$\begin{aligned}\mathbb{E}\left[\|\Delta(Z)\|^2\right] &= 18/\varepsilon^2\big((-r_{xx} + 2r_x^2)^2 + 4r_x^2 + 1\big) \\ &\leq 108/\varepsilon^2.\end{aligned}$$

Here, the equality holds when $r_x = r_{xx} = 1$. □

D.2 Proof for Proposition 5.3

Let $f([Z_n, Z_u]/n) = \frac{u+Z_u}{n+Z_n}$. As in the proof of Proposition 5.1, it suffices to show 1) $\text{Prob}\,(Z \notin A) = o(1/n^2)$, and 2) $\mathbb{E}\left[\|\Delta(Z)\|^2\right] = \frac{17/2}{\varepsilon^2}$. The proof of 1) can be done in a similar way as Lemma D.1. Thus, we prove 2) in the following lemma:

Lemma D.4. *We have*

$$\mathbb{E}\left[\|\Delta(Z)\|^2\right] \leq \frac{17/2}{\varepsilon^2}.$$

Proof. From simple algebra, we have

$$\Delta(Z) = -u \cdot Z_n + Z_u$$

Since $Z \sim \text{Lap}(2/\varepsilon)$, we obtain

$$\begin{aligned}\mathbb{E}\left[\|\Delta(Z)\|^2\right] &= 8/\varepsilon^2\big(u^2+1\big)\\ &\leq \frac{17/2}{\varepsilon^2}\end{aligned}$$

Here, the equality holds when $u = 1/4$. □

D.3 Proof of Lemma 5.2

D.3.1 Variance Case

Proof. Let $a \in \mathbb{R}^n, a' \in \mathbb{R}^{n+1}$ be neighboring datasets, i.e., $a_i = a'_i$ for all $i \leq n$ and $a'_{n+1} = x$. To simplify the notation, we denote $r = \frac{1}{n}\sum_{i=1}^n a_i$ Then, we have

$$\begin{aligned}\text{uvar}(a') - \text{uvar}(a) &= \sum_{i=1}^{n+1} a'_i - \frac{1}{n+1}\left(\sum_{i=1}^{n+1} a'_i\right)^2 - \sum_{i=1}^{n} a_i + \frac{1}{n}\left(\sum_{i=1}^{n} a_i\right)^2\\ &= x^2 - \frac{1}{n+1}\left(nr+x\right)^2 + \frac{1}{n}\left(nr\right)^2\\ &= x^2 - \frac{1}{n+1}\left(nr+x\right)^2 + \frac{1}{n+1}n(n+1)r^2\\ &= x^2 + \frac{1}{n+1}[n(n+1)r^2 - (nr+x)^2]\\ &= x^2 + \frac{1}{n+1}[n(n+1)r^2 - n^2r^2 - 2nrx - x^2]\\ &= \frac{1}{n+1}[nr^2 - 2nrx + nx^2]\\ &= \frac{n}{n+1}(r-x)^2.\end{aligned}$$

Thus, we have

$$0 \leq \text{uvar}(a') - \text{uvar}(a) \leq 1,$$

which completes the proof. □

D.3.2 Covariance Case

Proof. Let $(a, b) \in \mathbb{R}^n \times \mathbb{R}^n, (a', b') \in (\mathbb{R}^{n+1} \times \mathbb{R}^{n+1})$ be neighboring datasets, i.e., $a_i = a'_i$ for all $i \leq n, a'_{n+1} = x$ and $b_i = b'_i$ for all $i \leq n, b'_{n+1} = y$. To simplify the notation, we denote $r_a = \frac{1}{n}\sum_{i=1}^n a_i$ and $r_b = \frac{1}{n}\sum_{i=1}^n b_i$. Then,

we have

$$\begin{aligned}
\mathrm{ucov}(a', b') - \mathrm{ucov}(a, b) &= \sum_{i=1}^{n+1} a'_i b'_i - \frac{1}{n+1}\left(\sum_{i=1}^{n+1} a'_i\right)\left(\sum_{i=1}^{n+1} b'_i\right) - \sum_{i=1}^{n} a_i b_i + \frac{1}{n}\left(\sum_{i=1}^{n} a_i\right)\left(\sum_{i=1}^{n} b_i\right) \\
&= xy - \frac{1}{n+1}\left(nr_a + x\right)\left(nr_b + y\right) + \frac{1}{n}\left(nr_a\right)\left(nr_b\right) \\
&= xy - \frac{1}{n+1}\left(n^2 r_a r_b + nr_a y + nr_b x + xy\right) + \frac{1}{n+1} n(n+1) r_a r_b \\
&= -\frac{1}{n+1}\left(-nr_a r_b + nr_a y + nr_b x - nxy\right) \\
&= \frac{n}{n+1}(r_a - x)(r_b - y)
\end{aligned}$$

Thus, we have

$$-1 \leq \mathrm{ucov}(a', b') - \mathrm{ucov}(a, b) \leq 1,$$

which completes the proof. □

D.4 Proof of Lemma 5.5

Proof. For neighboring datasets D and D', we assume that $D = [x_1, \ldots, x_n]$ and $D' = [x_1, \ldots, x_n, x_{n+1}]$ without loss of generality. Then, we have

$$\|b(D') - b(D)\|_1 = \sum_{j=0}^{k} B_j(x_{n+1}) = 1,$$

where the first equality follows from the non-negativity of Bernstein basis and the last equality follows from the partition of unity property. □

D.5 Proof of Theorem 5.6

From the definition of $\hat{\mu}_j$, we have

$$\begin{aligned}
\hat{\mu}_j &= [A^{-1}(A\mu + \xi)]_j \\
&= \mu_j + [A^{-1}\xi]_j.
\end{aligned}$$

Since $\xi_j \sim \mathrm{Lap}(1/\varepsilon)$, we arrive at

$$\begin{aligned}
E[(\hat{\mu}_j - \mu_j)^2] &= \frac{2}{\varepsilon^2}\sum_{l=0}^{k}(A^{-1}_{jl})^2 \\
&= \frac{2}{\varepsilon^2}\sum_{l=j}^{k}\left(\frac{\binom{l}{j}}{\binom{k}{j}}\right)^2,
\end{aligned}$$

where, we used the fact that $A^{-1}_{jl} = \frac{\binom{l}{j}}{\binom{k}{j}}$ (Simsek, 2013). For $j = k$, we have

$$E[(\hat{\mu}_k - \mu_k)^2] = \frac{2}{\varepsilon^2}. \tag{2}$$

For any ε-DP estimator $\tilde{\mu}_k$, we can construct an ε-DP estimator $\hat{s}$ of μ_1 as follows:

$$\hat{s}(D) = \tilde{\mu}_k(\tilde{D}),$$

where $D = [x]$ and $\tilde{D} = [[x_1^{1/k}, \ldots, x_n^{1/k}]^\top]$. Note that $\mu_1(D) = \mu_k(\tilde{D})$. The utility of $\hat{s}$ can be evaluated as

$$E[(\hat{s}(D) - \mu_1(D))^2] = E[(\tilde{\mu}_k(\tilde{D}) - \mu_k(\tilde{D}))^2] \geq \sigma(\varepsilon)^2(1 - o(1))$$

from Lemma 6 in Kulesza et al. (2024). In the high-privacy regime, the lower bound becomes $2/\varepsilon^2(1 - o(1))$. Thus, Eq. (2) is asymptotically optimal for estimation of μ_k.

D.6 Proof of Lemma 5.7

Proof. For neighboring datasets D and D', we have

$$\begin{aligned}
\|b(D') - b(D)\|_1 &= \sum_{\alpha \in \{0,\dots,k\}^d} \left| B^k_{\alpha_1}(x^{(1)}_{n+1}) \cdots B^k_{\alpha_d}(x^{(d)}_{n+1}) \right| \\
&= \sum_{\alpha \in \{0,\dots,k\}^d} B^k_{\alpha_1}(x^{(1)}_{n+1}) \cdots B^k_{\alpha_d}(x^{(d)}_{n+1}) \\
&= \left(\sum_{\alpha_1=0}^{k} B^k_{\alpha_1}(x^{(1)}_{n+1}) \right) \cdots \left(\sum_{\alpha_d=0}^{k} B^k_{\alpha_d}(x^{(d)}_{n+1}) \right) \\
&= 1,
\end{aligned}$$

where the first equality follows from the non-negativity of Bernstein basis and the last equality follows from the partition of unity property. □

D.7 Proof of Theorem 5.8

Proof. From the definition of the multivariate Bézier mechanism, we have

$$\begin{aligned}
\hat{\mu}_\alpha &= [A_d^{-1}(A_d \mu + \xi)]_\alpha \\
&= \mu_\alpha + [A_d^{-1}\xi]_\alpha,
\end{aligned}$$

where $\xi_\beta \sim \mathrm{Lap}(1/\varepsilon)$ are independent Laplace random variables for each multi-index β.

Since $A_d = A \otimes A \otimes \cdots \otimes A$ (d times), we have $A_d^{-1} = A^{-1} \otimes A^{-1} \otimes \cdots \otimes A^{-1}$ (d times). From the property of tensor products, the entries of A_d^{-1} can be expressed as

$$(A_d^{-1})_{\alpha,\beta} = \prod_{j=1}^{d} (A^{-1})_{\alpha_j,\beta_j}.$$

Therefore, the mean squared error of $\hat{\mu}_\alpha$ is

$$\begin{aligned}
L(\hat{\mu}_\alpha, \mu_\alpha, D) &= \mathbb{E}\left[(\hat{\mu}_\alpha - \mu_\alpha)^2\right] \\
&= \mathbb{E}\left[\left(\sum_\beta (A_d^{-1})_{\alpha,\beta} \xi_\beta \right)^2 \right] \\
&= \sum_\beta (A_d^{-1})^2_{\alpha,\beta} \mathbb{E}\left[\xi_\beta^2\right] \\
&= \frac{2}{\varepsilon^2} \sum_\beta \prod_{j=1}^{d} (A^{-1})^2_{\alpha_j,\beta_j} \\
&= \frac{2}{\varepsilon^2} \prod_{j=1}^{d} \sum_{l_j=0}^{k} (A^{-1})^2_{\alpha_j,l_j} \\
&= \frac{2}{\varepsilon^2} \prod_{j=1}^{d} \sum_{l_j=\alpha_j}^{k} (A^{-1})^2_{\alpha_j,l_j} \\
&= \frac{2}{\varepsilon^2} \prod_{j=1}^{d} \frac{1}{\binom{k}{\alpha_j}^2} \sum_{l_j=\alpha_j}^{k} \binom{l_j}{\alpha_j}^2,
\end{aligned}$$

where we used the independence of ξ_β in the third equality, the fact that $\mathbb{E}\left[\xi_\beta^2\right] = 2/\varepsilon^2$ in the fourth equality, the factorization property of tensor products in the fifth equality, the fact that $(A^{-1})_{\alpha_j,l_j} = 0$ for $l_j < \alpha_j$ in the sixth equality, and the formula $A^{-1}_{j,l} = \frac{\binom{l}{j}}{\binom{k}{j}}$ (Simsek, 2013) in the last equality.

To show the upper bound $L(\hat{\mu}_\alpha, \mu_\alpha, D) \leq \frac{2k^d}{\varepsilon^2}$, note that

$$\sum_{l_j=\alpha_j}^{k} \binom{l_j}{\alpha_j}^2 \leq \sum_{l_j=0}^{k} \binom{l_j}{\alpha_j}^2 \leq k \binom{k}{\alpha_j}^2,$$

where the last inequality follows from the fact that $\binom{l_j}{\alpha_j} \leq \binom{k}{\alpha_j}$ for $l_j \leq k$. Therefore,

$$L(\hat{\mu}_\alpha, \mu_\alpha, D) \leq \frac{2}{\varepsilon^2} \prod_{j=1}^{d} k = \frac{2k^d}{\varepsilon^2}.$$

For the highest-order mixed moment $\alpha = (k, \ldots, k)$, we have

$$\begin{aligned} L(\hat{\mu}_{(k,\ldots,k)}, \mu_{(k,\ldots,k)}, D) &= \frac{2}{\varepsilon^2} \prod_{j=1}^{d} \frac{1}{\binom{k}{k}^2} \sum_{l_j=k}^{k} \binom{l_j}{k}^2 \\ &= \frac{2}{\varepsilon^2} \prod_{j=1}^{d} 1 \\ &= \frac{2}{\varepsilon^2}. \end{aligned}$$

This matches the lower bound for highest-order mixed moment estimation in the high-privacy regime, following from a similar argument as in the univariate case (Theorem 5.6). □

D.8 Proof of Theorem 5.9

Let $Z = [Z_n, Z_x, Z_y, Z_{xy}]$ and $f(Z/n) = \frac{r_{xy}+Z_{xy}/n}{1+Z_n/n} - (\frac{r_x+Z_x/n}{1+Z_n/n})(\frac{r_y+Z_y/n}{1+Z_n/n})$. Note that $f(0)$ is the true covariance and $\hat{c} = \mathrm{clip}(f(Z/n), [-1/4, 1/4])$, where $Z_n = \sum_i Z_i, Z_x = Z_1 + Z_3, Z_y = Z_2 + Z_3, Z_{xy} = Z_3$ and $Z_i \sim \mathrm{Lap}(1/\varepsilon)$. From simple calculation, we have $\nabla f(Z/n), \nabla^2 f(Z/n) = O(1)$ for $Z \in A_n$. Thus, from Lemma A.4, it suffices to show 1) $\mathrm{Prob}\,(A_n) = 1 - o(1/n^2)$, and 2) $\mathbb{E}\left[\|\Delta(Z)\|^2\right] \leq 2/\varepsilon^2$. We prove these conditions in the following lemmas:

Lemma D.5. *We have*

$$\mathrm{Prob}\,(Z \in A_n) = 1 - o(1/n^2).$$

Proof. Since $|Z_i| \leq 1/4n^{1/4}$ for all i is sufficient for the event A_n to hold, we can bound the probability of $Z \in A_n$ as follows:

$$\begin{aligned} \mathrm{Prob}\,(Z \in A_n) &\geq 1 - \sum_{i=1}^{4} \mathrm{Prob}\left(|Z_i| > 1/4n^{1/4}\right) \\ &\geq 1 - 4\exp\left(-\frac{\varepsilon n^{1/4}}{4}\right) \end{aligned}$$

This completes the proof. □

Lemma D.6. *Let $r_x = \frac{1}{n}\sum_{i=1}^n x_i$, $r_y = \frac{1}{n}\sum_{i=1}^n y_i$, and $c = \frac{1}{n}\sum_{i=1}^n (x_i - r_x)(y_i - r_y)$. Then, we have*

$$\mathbb{E}\left[\Delta(Z)^2\right] = \frac{2}{\varepsilon^2} C_1(r_x, r_y, c) \leq \frac{2}{\varepsilon^2},$$

where $C_1(r_x, r_y, c) := (1 - 2r_x + 2r_x^2)(1 - 2r_y + 2r_y^2) - 2c(1 - 2r_x)(1 - 2r_y) + 4c^2$.

Proof. Let $r_{xy} = \frac{1}{n}\sum_{i=1}^n x_i y_i$. From simple calculation, we have

$$\Delta(Z) = Z_{xy} - r_{xy} Z_n + 2r_x r_y Z_n - r_y Z_x - r_x Z_y$$

To simplify the expression, we rewrite $\Delta(Z)$ as follows:

$$\Delta(Z) = \underbrace{(Z_{xy} - r_y Z_x - r_x Z_y + r_x r_y Z_n)}_{=:X} - cZ_n.$$

Substituting the definition of Z_x, Z_y and Z_{xy} into the expression for X, we have $X = \sum_{i=1}^4 x_i Z_i$, where

$$\begin{aligned} x_1 &= (1-r_x)(1-r_y) \\ x_2 &= -r_y(1-r_x) \\ x_3 &= -r_x(1-r_y) \\ x_4 &= r_x r_y. \end{aligned}$$

In addition, Y can be expressed as $Y = \sum_{i=1}^4 (x_i - c) Z_i$ since $Z_n = Z_1 + Z_2 + Z_3 + Z_4$. Thus, we have

$$\frac{\mathbb{E}\left[Y^2\right]}{2/\varepsilon^2} = \sum_{i=1}^4 (x_i - c)^2 = \sum_{i=1}^4 x_i^2 - 2c \sum_{i=1}^4 x_i + 4c^2.$$

Here, $\sum_{i=1}^4 x_i^2$ and $\sum_{i=1}^4 x_i$ can be computed as follows:

$$\begin{aligned} \sum_{i=1}^4 x_i^2 &= (1-r_x)^2(1-r_y)^2 + r_y^2(1-r_x)^2 + r_x^2(1-r_y)^2 + r_x^2 r_y^2 \\ &= \left((1-r_x)^2 + r_x^2\right)\left((1-r_y)^2 + r_y^2\right) \\ &= (1 - 2r_x + 2r_x^2)(1 - 2r_y + 2r_y^2) \\ \sum_{i=1}^4 x_i &= (1-r_x)(1-r_y) - r_y(1-r_x) - r_x(1-r_y) + r_x r_y \\ &= (1-2r_x)(1-2r_y). \end{aligned}$$

Therefore, it suffices to show that

$$\begin{aligned} f(c) := C_1(r_x, r_y, c) &:= \frac{\mathbb{E}\left[Y^2\right]}{2/\varepsilon^2} \\ &= (1 - 2r_x + 2r_x^2)(1 - 2r_y + 2r_y^2) - 2c(1-2r_x)(1-2r_y) + 4c^2 \le 1 \end{aligned}$$

Note that $f(c)$ is a convex quadratic function of c. The maximum value of a convex function on a closed interval is achieved at one of the endpoints of the interval. Therefore, it suffices to show that $f(c) \le 1$ at the endpoints of the interval of c. As shown in Lemma A.3, the domain of r_{xy} is given by $\max(0, r_x + r_y - 1) \le r_{xy} \le \min(r_x, r_y)$, which defines the interval for $c = r_{xy} - r_x r_y$.

Case 1: $r_{xy} = \min(r_x, r_y)$**.** Without loss of generality, assume $r_x \le r_y$, which implies $r_{xy} = r_x$. Then, $c = r_x - r_x r_y = r_x(1 - r_y)$. Substituting this into $f(c)$ yields

$$\begin{aligned} g(r_x, r_y) := f(r_x - r_x r_y) - 1 &= 16r_x^2 r_y^2 - 24r_x^2 r_y + 10r_x^2 - 8r_x y^2 + 10r_x y - 4r_x + 2r_y^2 - 2r_y \\ &= (16r_x^2 - 8r_x + 2)r_y^2 + (-24r_x^2 + 10r_x - 2)r_y + (10r_x^2 - 4r_x) \end{aligned}$$

The function $g(r_x, r_y)$ is a convex quadratic function in r_y since $(16r_x^2 - 8r_x + 2) \ge 0$ for any r_x. Thus, the maximum value of $g(r_x, r_y)$ occurs at one of the endpoints $r_y = r_x$ or 1. For $r_y = r_x$, we have

$$\begin{aligned} g(r_x, r_x) &= 16r_x^4 - 24r_x^3 + 10r_x^2 - 8r_x^3 + 10r_x^2 - 4r_x + 2r_x^2 - 2r_x \\ &= 16r_x^4 - 32r_x^3 + 22r_x^2 - 6r_x \\ &= 2r_x(8r_x^3 - 16r_x^2 + 11r_x - 3) \\ &= 2r_x(r_x - 1)(8r_x^2 - 8r_x + 3) \\ &\le 0 \end{aligned}$$

since $(8r_x^2 - 8r_x + 3) > 0$, $r_x \geq 0$, and $(1 - r_x) \leq 0$. For $r_y = 1$, we have

$$\begin{aligned} g(r_x, 1) &= (16r_x^2 - 8r_x + 2) - (24r_x^2 - 10r_x + 2) + (10r_x^2 - 4r_x) \\ &= (16 - 24 + 10)r_x^2 + (-8 + 10 - 4)r_x + (2 - 2) \\ &= 2r_x^2 - 2r_x \\ &= 2r_x(r_x - 1) \\ &\leq 0. \end{aligned}$$

Case 2: $r_{xy} = \max(0, r_x + r_y - 1)$**.** Assuming $r_x + r_y \leq 1$, we have $c = -r_x r_y$. Then, $f(c)$ can be expressed as follows:

$$\begin{aligned} g(r_x, r_y) &:= f(-r_x r_y) - 1 \\ &= -2r_x - 2r_y + 2r_x^2 + 2r_y^2 + 6r_x r_y - 8r_x^2 r_y - 8r_x r_y^2 + 16r_x^2 r_y^2. \end{aligned}$$

Here, $g(r_x, r_y)$ is a convex quadratic function in r_x since the coefficient of r_x^2 is $16r_y^2 - 8r_y + 2 \geq 0$ for any $r_y \in [0, 1]$. Since $r_x \in [0, 1 - r_y]$, the maximum value occurs at one of the endpoints $r_x = 0$ or $r_x = 1 - r_y$. For $r_x = 0$, we have

$$\begin{aligned} g(0, r_y) &= -2r_y + 2r_y^2 \\ &= 2r_y(r_y - 1) \\ &\leq 0. \end{aligned}$$

For $r_x = 1 - r_y$, let $p = r_x r_y$. Then, we have

$$\begin{aligned} g(1 - r_y, r_y) &= -2 + 2 + 2p - 8p + 16p^2 \\ &= 16p^2 - 6p \\ &= 16(p - 3/16)^2 - 9/16 \\ &\leq 0 \end{aligned}$$

since $p = r_x(1 - r_x) \in [0, 1/4]$.

Assuming $r_x + r_y \geq 1$, we have $c = r_x + r_y - 1 - r_x r_y = -(1 - r_x)(1 - r_y)$. Then, we can express $f(c)$ as follows:

$$\begin{aligned} g(r_x, r_y) &:= f(-(1 - r_x)(1 - r_y)) - 1 \\ &= 16r_x^2 r_y^2 - 24r_x^2 r_y + 10r_x^2 - 24r_x r_y^2 + 38r_x r_y - 16r_x + 10r_y^2 - 16r_y + 6. \end{aligned}$$

Here, $g(r_x, r_y)$ is a convex quadratic function in r_x since the coefficient of r_x^2 is $16r_y^2 - 24r_y + 10 \geq 0$ for any $r_y \in [0, 1]$. Since $r_x \in [1 - r_y, 1]$, the maximum value of $g(r_x, r_y)$ occurs at one of the endpoints $r_x = 1 - r_y$ or 1. For $r_x = 1$, we have

$$\begin{aligned} g(1, r_y) &= 16r_y^2 - 24r_y + 10 - 24r_y^2 + 38r_y - 16 + 10r_y^2 - 16r_y + 6 \\ &= 2r_y^2 - 2r_y \leq 0. \end{aligned}$$

For $r_x = 1 - r_y$, let $p = r_x r_y$. Then, we have

$$\begin{aligned} g(1 - r_y, r_y) &= 16p^2 - 24p + 10 + 18p - 10 \\ &= 16p^2 - 6p \\ &= 16(p - 3/16)^2 - 9/16 \\ &\leq 0. \end{aligned}$$

For the last inequality, we use the fact that $p = r_x r_y \leq (1 - r_y) r_y \in [0, 1/4]$.

Combining Case 1 and Case 2, we obtain the result. □

D.9 Proof of Theorem 5.10

The proof follows the same structure as the proof of Theorem 5.9. Let $Z = [Z_n, Z_x, Z_{xx}]$ and $f(Z/n) = \frac{r_{xx}+Z_{xx}/n}{1+Z_n/n} - (\frac{r_x+Z_x/n}{1+Z_n/n})^2$. Note that $f(0)$ is the true variance and $\hat{v} = \text{clip}(f(Z/n), [0, 1/4])$, where $Z_n = Z_0+Z_1+Z_2$, $Z_x = Z_1/2+Z_2$, and $Z_{xx} = Z_2$ for $Z_i \sim \text{Lap}(1/\varepsilon)$. From Lemma A.4, it suffices to show 1) $\text{Prob}\,(A_n) = 1-o(1/n^2)$, and 2) $\mathbb{E}\left[\|\Delta(Z)\|^2\right] \le 2/\varepsilon^2$.

Lemma D.7. *The probability of the event A_n is bounded as follows:*

$$\text{Prob}\,(Z \in A_n) = 1 - o(1/n^2).$$

Proof. Since $|Z_i| \le 1/4n^{1/4}$ for all i is sufficient for the event A_n to hold, we can bound the probability of A_n as follows:

$$\begin{aligned}\text{Prob}\,(Z \in A_n) &\ge 1 - \sum_{i=1}^{3} \text{Prob}\left(|Z_i| > 1/4n^{1/4}\right) \\ &\ge 1 - 4\exp\left(-\frac{\varepsilon n^{1/4}}{4}\right) = 1 - o(1/n^2).\end{aligned}$$

This completes the proof. □

Lemma D.8. *Let $r = \frac{1}{n}\sum_{i=1}^n x_i$ and $v = \frac{1}{n}\sum_{i=1}^n (x_i - r)^2$. Then, we have*

$$\mathbb{E}\left[\Delta(Z)^2\right] = \frac{2}{\varepsilon^2} C_b(r, v) \le \frac{2}{\varepsilon^2},$$

where $C_b(r, v) := 3v^2 - 2(3r^2 - 3r + 1)v + 3r^4 - 6r^3 + 7r^2 - 4r + 1$.

Proof. Let $s = \frac{1}{n}\sum_{i=1}^n x_i^2$. From simple calculation, we have

$$\Delta(Z) = Z_{xx} - sZ_n + 2r^2 Z_n - 2rZ_x.$$

Let $Y = Z_{xx} - sZ_n + 2r^2 Z_n - 2rZ_x$. Then, we have

$$Y = (2r^2 - s)Z_1 + (2r^2 - r - s)Z_2 + (1 - 2r + 2r^2 - s)Z_3.$$

Since the Z_i are independent and have variance $2/\varepsilon^2$, $\mathbb{E}\left[Y^2/(2/\varepsilon^2)\right]$ is given by:

$$\begin{aligned}\mathbb{E}\left[Y^2/(2/\varepsilon^2)\right] &= (2r^2 - s)^2 + (2r^2 - r - s)^2 + (1 - 2r + 2r^2 - s)^2 \\ &= (r^2 - v)^2 + (r^2 - r - v)^2 + ((1 - r)^2 - v)^2 \\ &= 3v^2 - 2(3r^2 - 3r + 1)v + 3r^4 - 6r^3 + 7r^2 - 4r + 1 \\ &=: C_3(r, v).\end{aligned}$$

We compare C_3 with $C_1(r, r, v)$ in the proof of Theorem 5.9. We have

$$C_1(r, r, v) - C_3(r, v) = v^2 - 2(r^2 - r)v + r^2(r - 1)^2 = (v - (r^2 - r))^2 \ge 0.$$

Thus, we have

$$C_3(r, v) \le C_1(r, r, v) \le 1.$$

This completes the proof. □

D.10 Proof of Lemma 5.12

Proof. Let $D = [x]$ and $D' = [x']$ be neighboring datasets, i.e., $x_i = x'_i$ for all $i \le n$ and $x'_{n+1} = a$. Then, we have

$$\|b(D') - b(D)\|_1 = |1 - (\text{uvar}(x') - \text{uvar}(x))| + |\text{uvar}(x') - \text{uvar}(x)| \le 1.$$

Here, we used the fact $0 \le \text{uvar}(x') - \text{uvar}(x) \le 1$ from Lemma 5.2. □

D.11 Proof of Proposition 5.13

The proof can be done in a similar way to the proof of Theorem 5.9. Thus, it suffices to show Lemmas corresponding to Lemma D.5, and D.6.

Lemma D.9. *The probability of the event A_n is bounded as follows:*

$$\text{Prob}\,(A) = 1 - o(1/n^2).$$

Proof. Since $|Z_i| \leq 1/4n^{1/4}$ for all i is sufficient for the event A to hold, we can bound the probability of A as follows:

$$\begin{aligned}\text{Prob}\,(A) &\geq 1 - \sum_{i=1}^{4} \text{Prob}\left(|Z_i| > 1/4n^{1/4}\right) \\ &\geq 1 - 4\exp\left(-\frac{\varepsilon n^{1/4}}{4}\right) = 1 - o(1/n^2).\end{aligned}$$

This completes the proof. □

Lemma D.10. *Let $r = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $v = \frac{1}{n}\sum_{i=1}^{n}(x_i - r)^2$. Then, we have*

$$\mathbb{E}\left[\Delta(Z)^2\right] = \frac{2}{\varepsilon^2} C_u(r, v) \leq \frac{2}{\varepsilon^2},$$

where $C_u(r, v) = v^2 + (1 - v)^2$.

Proof. From simple calculation, we have

$$\Delta(Z) = Z_u - vZ_n = -vZ_1 + (1 - v)Z_2.$$

Since the Z_i are independent and have variance $2/\varepsilon^2$, we obtain

$$\mathbb{E}\left[\Delta(Z)^2/(2/\varepsilon^2)\right] = C_u(r, v) \leq 1$$

since $v \in [0, 1/4]$. This completes the proof. □

D.12 Proof of Theorem 5.14

The upper bound on $R(\cdot, \text{var}, D)$ is already obtained in the proof of Theorem 5.9, 5.10, and Proposition 5.13. In addition, we have already show that $C_b(r, v) \leq C_c(r, v) = C(r, r, v)$ in the proof of Lemma D.8. Thus, it suffices to show $C_c(r, v) \leq C_u(r, v)$. Let $f(r, v) = (C_c(r, v) - C_u(r, v))/2$. Then, we have

$$f(r, v) = v^2 - (4r^2 - 4r)v + (2r^4 - 4r^3 + 4r^2 - 2r).$$

This is a convex quadratic function in v. Thus, the maximum value of $f(r, v)$ occurs at one of the endpoints $v = 0$ or $r - r^2$. For $v = 0$, we have

$$\begin{aligned}f(r, 0) &= 2r^4 - 4r^3 + 4r^2 - 2r \\ &= 2r(r^3 - 2r^2 + 2r - 1) \\ &= 2r(r - 1)(r^2 - r + 1) \leq 0\end{aligned}$$

for any $r \in [0, 1]$. For $v = r - r^2$, we have

$$\begin{aligned}f(r, r - r^2) &= (r - r^2)^2 - (4r^2 - 4r)(r - r^2) + (2r^4 - 4r^3 + 4r^2 - 2r) \\ &= 5(r - r^2)^2 + (r - r^2)(-2 + 2r - 2r^2) \\ &= (r - r^2)(-7r^2 + 7r - 2) \\ &= -r(1 - r)(7(r - 1/2)^2 + 1/4) \leq 0.\end{aligned}$$

Combining the above two cases, we have $f(r, v) \leq 0$ for any $r \in [0, 1]$ and $v \in [0, r - r^2]$. Thus, we obtain the result.

D.13 Proof of Theorem 6.1

Let $g := \nabla f(D)$. Then, the variance of the errors can be expressed as

$$\mathbb{E}\left[\mathcal{E}_{\text{bezier}}(D)^2\right] = \mathbb{E}\left[(g^\top A_d^{-1}\xi)^2\right] = \frac{2}{\varepsilon^2} g^\top (A_d A_d^\top)^{-1} g,$$
$$\mathbb{E}\left[\mathcal{E}_{\text{naive}}(D)^2\right] = \mathbb{E}\left[((k+1)^d \cdot g^\top \xi)^2\right] = \frac{2}{\varepsilon^2}(k+1)^{2d}\|g\|_2^2.$$

Thus, it suffices to show that the maximum eigenvalue $\lambda_{\max}((A_d A_d^\top)^{-1})$ satisfies $\lambda_{\max}((A_d A_d^\top)^{-1}) \le (k+1)^{2d}$, which is equivalent to $\sigma_{\max}(A_d^{-1}) \le (k+1)^d$, where $\sigma_{\max}(\cdot)$ denotes the maximum singular value. Since $A_d = A^{\otimes d}$, we have $\sigma_{\max}(A_d^{-1}) = \sigma_{\max}(A^{-1})^d$. Therefore, it remains only to show $\sigma_{\max}(A^{-1}) \le (k+1)$. From Simsek (2013), we have

$$A_{ij}^{-1} = \frac{\binom{j}{i}}{\binom{k}{i}} \quad \text{for } i \le j, \text{ and 0 otherwise.}$$

To evaluate $\left\|A^{-1}\right\|_2$, we evaluate $\left\|A^{-1}\right\|_1, \left\|A^{-1}\right\|_\infty$. The l_∞ norm of A^{-1} as follows:

$$\begin{aligned}\left\|A^{-1}\right\|_\infty &= \max_{i\in\{0,\dots,k\}} \sum_{j=0}^{k} \left|A_{ij}^{-1}\right| = \max_{i\in\{0,\dots,k\}} \sum_{j=i}^{k} \frac{\binom{j}{i}}{\binom{k}{i}} = \max_{i\in\{0,\dots,k\}} \frac{1}{\binom{k}{i}} \sum_{j=i}^{k} \binom{j}{i} \\ &= \max_{i\in\{0,\dots,k\}} \frac{1}{\binom{k}{i}} \binom{k+1}{i+1} = \max_{i\in\{0,\dots,k\}} \frac{k+1}{i+1} = k+1.\end{aligned}$$

The l_1 norm can be evaluated as

$$\left\|A^{-1}\right\|_1 = \max_{j\in\{0,\dots,k\}} \sum_{i=0}^{k} \left|A_{ij}^{-1}\right| = \max_{j\in\{0,\dots,k\}} \sum_{i=0}^{j} \frac{\binom{j}{i}}{\binom{k}{i}} \le \sum_{i=0}^{k} \frac{\binom{k}{i}}{\binom{k}{i}} = k+1.$$

Combining the above results, we arrive at

$$\left\|A^{-1}\right\|_2 \le \sqrt{\|A^{-1}\|_\infty \cdot \|A^{-1}\|_1} \le k+1.$$

This completes the proof.

E Detailed Experimental Setup and Additional Results

E.1 Experimental Setup

Computing Infrastructure All experiments were conducted on a machine with an Apple M3 Pro chip and 36GB of RAM.

Hyperparameters Settings For the histogram-based approach, we optimize the number of bins from $\{4, 8, 16, 32, 64\}$ based on the lowest MSE for each dataset and privacy budget.

E.2 Datasets

Variance Estimation Given mean r and variance v, we generate n samples from Beta distribution with shape parameters $\alpha = r \cdot k$ and $\beta = (1-r) \cdot k$, where $k = \frac{r(1-r)-v}{v}$. We can easily verify that the mean and variance of the generated samples are r and v, respectively. To evaluate the worst-case performance, we first select 5 equally spaced values of r from the interval $[0, 1]$. For each chosen r, we then select 5 equally spaced values of v from its feasible range.

Covariance Estimation Given means r_x, r_y and covariance c, we generate n samples from a multinomial distribution over $\{(0,0),(0,1),(1,0),(1,1)\}$ with probabilities $1 - r_x - r_y + c + r_x r_y, r_y - (c + r_x r_y), r_x - (c + r_x r_y), c + r_x r_y$. We can easily verify that the means and covariance of the generated samples are r_x, r_y, and c, respectively. To evaluate the worst-case performance, we first select 5 equally spaced values of r_x, r_y from the interval $[0, 1]$. For each pair of chosen r_x, r_y, we then select 5 equally spaced values of c from its feasible range.

Correlation Estimation First, we generate n samples from a bivariate Gaussian distribution with mean vector $[1/2, 1/2]^\top$ and covariance matrix $\begin{bmatrix} 1/2, r/2 \\ r/2, 1/2 \end{bmatrix}$. Then, we clip the samples to the range $[0, 1]$. Specifically, we set $r = 0.7$ in our experiments.