Unbiased Private Estimators by Postprocessing the Discrete Laplace Mechanism

Quentin Hillebrand
DI ENS, École Normale Suppérieure
Université PSL, CNRS

Jacob Imola
University of Waterloo

Rasmus Pagh
BARC
University of Copenhagen

Sia Sejer
BARC
University of Copenhagen

Abstract

We show that an "old dog," the classical discrete Laplace (aka. geometric) mechanism, can "perform new tricks":

1. It can be post-processed to yield a simple, unbiased estimator of any subexponential function f of the original data, giving a simple, discrete, multivariate version of the recent unbiasing result for the Laplace mechanism by Calmon et al. (FORC '25).
2. It can be post-processed to output the same distribution as the Laplace mechanism or the Staircase mechanism with identical privacy parameters.

Thus, the discrete Laplace mechanism is a versatile mechanism that should be preferred over the Laplace and Staircase mechanisms whenever the data is discrete.

We show bounds on the variance of our estimator, compared to the mean square error of the biased estimator that simply evaluates f on the output of the mechanism. Though our unbiased estimator has exponential running time for worst-case functions, we show that it can often be computed in linear or polynomial time for some common functions exhibiting structure. We showcase the properties of our methods empirically with several use cases including profile and entropy estimation, as well as distributed/federated data analysis applications in which unbiasedness is key to accuracy.

1 Introduction

Differential privacy has a powerful *post-processing* property: For any differentially private mechanism $\mathcal{M}$ and any function f, we can release $f(\mathcal{M}(x))$ while preserving the privacy guarantees of $\mathcal{M}$. As a consequence, we can release all values $f_i(\mathcal{M}(x))$ for any family of functions $(f_i)_{i\in I}$ with no loss of privacy. In this paper, we consider additive noise mechanisms where $\mathbb{E}\left[\mathcal{M}(x)\right] = x$. For non-linear functions f, the expected function value $\mathbb{E}\left[f(\mathcal{M}(x))\right]$, assuming it exists, is in general different from $f(x)$. Such bias can be a problem in some applications, which led Hillebrand, Suppakitpaisarn, and Shibuya [2023] and later Calmon, Du, Dwork, Finley, and Franguridi [2025] to study methods for debiasing via post-processing of the Laplace mechanism. First, Hillebrand et al. [2023] showed how to construct unbiased estimators for multivariate polynomials. Later, Calmon et al. [2025] gave estimators in the univariate setting for a wide class of twice differentiable functions f.

Debiasing in Post-Processing. In this paper, we propose a new post-processing method for the *discrete* Laplace mechanism $\mathcal{M}_{\text{DLap}} : \mathbb{Z}^n \to \mathbb{Z}^n$, first proposed and studied by Ghosh, Roughgarden,

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

and Sundararajan [2009]. For convenience we will use the notation $\tilde{x} = \mathcal{M}_{\text{DLap}}(x)$, that is, $\tilde{x}_i = x_i + \eta_i$ where η_i is sampled i.i.d. from the discrete Laplace distribution with parameter $p = e^{-\varepsilon}$. This mechanism is a canonical way of releasing histograms, and more generally integer-valued data with sensitivity 1, with an ε-differential privacy guarantee. Given $\tilde{x}$, our post-processing method returns an unbiased estimator of $f(x)$ for *any* univariate or multivariate function f where $\mathbb{E}\left[|f(\tilde{x})|\right]$ exists and is finite. A sufficient condition for this is that f is subexponential (or exponential with low enough rate). Our method thus supports a class of functions f much wider than that of Calmon et al. [2025], Hillebrand et al. [2023], both of which can only handle functions that are twice differentiable. Previously, unbiased postprocessing of the discrete Laplace distribution had only been described in the special case of univariate threshold functions, by Ghazi, Kamath, Kumar, and Manurangsi [2022].

We show that our post-processing method is unique in the sense that it is the only deterministic method that constructs unbiased estimates in the discrete setting. From this, it follows that it has the smallest variance of all unbiased estimators. Though our estimator takes exponential time to compute in the worst-case, we show that for functions which exhibit structure (e.g. decomposable functions such as entropy and order statistics), it is often possible to design linear or polynomial time algorithms for computing the unbiased estimation.

Transformations to other Additive Noise Mechanisms. We show how to post-process $\tilde{x}$ so that the resulting distribution is identical to that of the classical (real-valued) Laplace mechanism [Dwork, McSherry, Nissim, and Smith, 2006] or the Staircase mechanism [Geng and Viswanath, 2014, 2015] on input x with the same privacy parameter ε. This means the Discrete Laplace Mechanism is a versatile mechanism that should be preferred over the Laplace and Staircase Mechanisms whenever the data is discrete. Everything that can be done with these mechanisms can also be based on the Discrete Laplace Mechanism, and the latter may lead to better utility, especially for medium to large values of ε.

Applications to Federated/ Distributed Settings. Unbiased estimators are particularly useful in settings where the statistic of interest can be obtained by summing or averaging many individual contributions. For example, suppose that each of k parties (e.g. hospitals) holds a distribution given by a histogram, that each histogram has been released with the Discrete Laplace mechanism, and we wish to compute the entropy of the joint distribution over all parties. Assuming independence, the joint entropy is the sum of the entropy of each distribution, a non-linear function of the histogram. Summing an unbiased estimator for the entropy of each distribution we get an estimator for the total entropy whose error scales with $O(\sqrt{k})$. In contrast, summing biased estimators will result in error proportional to k, which will be worse for sufficiently large k. As another example, suppose we wish to gather statistics *across* histograms, for example the minimum, maximum, mean and variance of each column in the histograms. We can avoid bias in these estimates by using our estimator on the "transposed" histograms formed by combining information on each column across histograms.

1.1 Overview of Our Contributions.

In Section 2, we first show (Theorem 1) that in the univariate case ($n = 1$) a linear combination of $f(\tilde{x} - 1)$, $f(\tilde{x})$, and $f(\tilde{x} + 1)$ is an unbiased estimator of $f(x)$ for any function f and input x where $\mathbb{E}\left[|f(\tilde{x})|\right]$ exists. This estimator can be interpreted as the discrete second derivative of f evaluated at $\tilde{x}$, analogous to the approach by Calmon et al. [2025] for univariate functions. We then show (Theorem 2) that our method extends naturally to the multivariate case $x \in \mathbb{Z}^n$ where the estimator is a linear combination of the function values $f(\tilde{x} + \xi)$ for $\xi \in \{-1, 0, +1\}^n$.

Section 3 investigates properties of our unbiased estimator. We first show (Theorem 4) that it is unique among deterministic estimators and has minimum variance among all unbiased estimators. In particular, for a black-box function f, any deterministic estimator requires 3^n function evaluations. Then we consider the mean square error (MSE) and show (Theorem 5) that it is at most a factor $(3 + O(1/\varepsilon^4))^n$ worse than the MSE of the naive estimator $f(\tilde{x})$. This bound is tight: Proposition 6 shows that there are functions for which our estimator exhibits this rate of exponential variance growth, but for other functions we instead see that the unbiased estimator is exponentially *better*.

Section 4 shows that it is possible to post-process the Discrete Laplace Mechanism to be identical to either the classical Laplace Mechanism (Theorem 10) or to the Staircase mechanism (Theorem 17) while preserving the same privacy level ε. The insight is that it is possible to choose random variables

X_ε and $Y_{\gamma,\varepsilon}$ (independently of x), such that $\tilde{x} + X_\varepsilon$ and $\tilde{x} + Y_{\gamma,\varepsilon}$ have the same distribution as the classical Laplace mechanism and the Staircase Mechanism with privacy parameter ε, respectively.

Section 5 studies functions and classes of functions for which our estimator can be computed much faster than by direct summation of its 3^n terms. The key idea is to group terms with the same contribution, and compute how many terms are in each group. This allows us to compute order statistics (such as the minimum, median, and maximum), inference for decision trees, as well as entropy and KL divergence in polynomial time.

In Section 6 we present four case studies showcasing settings in which our estimator has an advantage over alternatives. We consider polynomials, for a use case in graph counting, for which we emphasize the advantage of the discrete Laplace mechanism over the classical Laplace mechanism for high values of ε. For empirical entropy, we see a significantly improved mean square error for low- and medium values of ε for sparse histograms. For estimating partition functions, a setting in which the naive estimator has considerable bias, we also see improved behavior near the convergence radius. Finally, we give the first practical demonstration of profile estimation by post-processing.

Societal Impact of Unbiasedness. So far, we have considered unbiasedness from the technical perspective of statistical bias. From a societal perspective, one can ask how unbiased estimation relates to algorithmic bias and fairness. There are many distinct and incompatible notions of algorithmic fairness (e.g. Kleinberg, Mullainathan, and Raghavan [2017]). A potential benefit of pointwise unbiased estimators is that they avoid directional statistical bias so that no group of points is systematically underestimated or overestimated in expectation. However, an unbiased estimator might have a high variance and over- or underestimate the function value most of the time, reducing its fairness in practice. For example, the estimator may overestimate the target value with high probability while remaining unbiased, because it occasionally produces underestimates that compensate in expectation. Another concern is that the estimator's variance may differ substantially across groups of input points. This raises the question of whether one can design unbiased estimators with uniform or controlled variance for all points, and what trade-offs such constraints would impose. To sum up, we stress that one should be careful not to conflate unbiasedness and fairness.

Limitations. Per Theorem 5, the gap between the MSE of the unbiased estimator and the naive estimator can sometimes be quite large, at $(1 + \frac{1}{\varepsilon^4})^{\Omega(n)}$. This fact, and the exponential (in n) runtime of our general estimator, may make it unsuitable for certain applications. Nevertheless, our work provides the first rigorous treatment of the unbiased estimator for the multidimensional discrete Laplace mechanism.

1.2 Related Work

Debiasing Functions of the Laplace Mechanism. Hillebrand et al. [2023] showed how to debias functions of the classical Laplace mechanism (an additive noise mechanism that yields privacy guarantees also for non-integer data) when f is a multivariate polynomial. Calmon et al. [2025] considered a much more general class of functions in the univariate case ($n = 1$). They showed that there is an essentially unique way of computing such unbiased estimators: for a suitable constant c_ε and any twice-differentiable, so-called tempered function f, the estimator

$$\hat{f}(\mathcal{M}(x)) = f(\mathcal{M}(x)) - c_\varepsilon f''(\mathcal{M}(x))$$

satisfies $\mathbb{E}\left[\hat{f}(\mathcal{M}(x))\right] = f(x)$. The result is derived via a deconvolution perspective in the Fourier domain and yields closed-form unbiased estimators for broad classes of nonlinear functions. It is not immediately clear from their results how to extend the result to the multivariate setting. For our estimator these strong assumptions on f are not necessary, so it is applicable to a wider class of functions for which the naive estimator of $|f(x)|$ has a well-defined, finite expectation.

Mechanisms based on Postprocessing the Discrete Laplace Mechanism. Ghazi et al. [2022] considered a related distributed setting in which a cryptographic primitive, a *shuffler* [Cheu, Smith, Ullman, Zeber, and Zhilyaev, 2019, Erlingsson, Feldman, Mironov, Raghunathan, Talwar, and Thakurta, 2019], is used to collect and noise histogram data. Similar to our results they consider Discrete Laplace noise, and in the univariate case they show how to debias a threshold indicator function applied to the output of the Discrete Laplace mechanism. This result is used as a stepping

stone to estimate any *symmetric* function of the histogram. For comparison we provide a method that can be used without shuffling and supports a wider class of functions that are not necessarily symmetric.

Wu and Pagh [2024] studied *profile estimation* which boils down to estimating for various k a sum of indicator function values $f(x)$ where $f(x) = 1$ if $x = k$ and $f(x) = 0$ otherwise. They present a way of "inverting" the noise added by the Discrete Laplace mechanism and show that it is asymptotically near-optimal. Our estimator gives a much simpler solution to this problem.

Privately Estimating Black-Box Statistics. Related to supporting privacy on a broad class of functions, Linder, Raskhodnikova, Smith, and Steinke [2025] designed a so-called down-local algorithm for estimating a function f with exponentially many evaluations of f on noisy data. Despite the superficial similarity, their setting is rather different from ours. Their method provides accurate estimations for *well-behaved* (finite-domain or c-Lipschitz) functions on the dataset and its subset. For monotone functions, the bias is always downwards, and for more general functions with finite range, they provided a guaranteed interval. Steinke and Steinke [2026] present a method for interpolating between computational efficiency and statistical accuracy when estimating a black box function f. Formally, they interpolate between high statistical accuracy and an exponential number of evaluations, and low statistical accuracy, where the number of evaluations depends only on the privacy parameters and the domain size.

1.3 Definitions

Definition 1 (Differential Privacy (DP) [Dwork et al., 2006])**.** For $\varepsilon > 0$ an algorithm $\mathcal{M}$ with domain $\mathcal{D}$ and range $\mathcal{S}$ satisfies ε-differential privacy (ε-DP) if, for every $S \subset \mathcal{S}$ and any $D^{(1)}, D^{(2)} \in \mathcal{D}$ that differ by only one element, we have

$$\Pr\left[\mathcal{M}\left(D^{(1)}\right) \in S\right] \leq e^{\varepsilon} \cdot \Pr\left[\mathcal{M}\left(D^{(2)}\right) \in S\right] \ .$$

In this paper we consider datasets that have been transformed to integer-valued vectors in $\mathbb{Z}^n$. A canonical example of such a dataset is a histogram of counts of elements in a discrete domain of size n. We consider two datasets x, x' to be neighboring if $\|x - x'\|_1 \leq 1$, i.e., at most one count can be changed by 1. For discrete-valued data, the canonical DP mechanism is based on adding noise drawn from the *discrete Laplace distribution*:

Definition 2 (Discrete Laplace distribution)**.** For $p \in (0, 1)$, $\texttt{DLap}(p)$ denotes the discrete Laplace distribution whose probability mass at $i \in \mathbb{Z}$ is $\frac{1-p}{1+p} \cdot p^{|i|}$.

We consider the setting where x has been published as a noisy histogram $\tilde{x} = x + \texttt{DLap}(p)^n$ by adding independent noise drawn from $\texttt{DLap}(p)$ to each coordinate. This is an example of the Discrete Laplace Mechanism, and it is known that if $p = e^{-\varepsilon}$, it satisfies ε-DP [Ghosh et al., 2009]. We are interested in estimating a function $f(x)$ of the sensitive dataset, where $f : \mathbb{Z}^n \to \mathbb{R}$, while only having access to $\tilde{x}$. To do so, we will use an estimator $g(\tilde{x})$. We measure the accuracy of an estimator using the mean square error.

Definition 3 (Mean Square Error)**.** The mean square error (MSE) of an estimator g on input x is $\mathrm{MSE}_g(x) = \mathbb{E}\left[(g(\tilde{x}) - f(x))^2\right]$. If the expected value does not exist, the MSE is *undefined*.

Note that this definition includes the dataset x as a parameter, rather than considering a maximum over all x. This pointwise definition will be useful when computing our error bounds later.

2 Debiasing the Discrete Laplace Mechanism

We now study how to design an unbiased estimator g that, given access to the dataset x published with Discrete Laplace Noise, has expected value $f(x)$ for a given function f.

Univariate Case. We first consider a single dimension $n = 1$. The basic idea is to "invert" the equation $\mathbb{E}\left[g(\tilde{x})\right] = f(x)$, which is generally possible because, when expanded out, expectation is a linear operator. In general, such an estimator would not have a closed form, but the Discrete Laplace Distribution has tails with a simple exponential shape, which enables cancellation. This makes it possible to debias $f(\tilde{x})$ by considering only $f(\tilde{x} + 1)$, $f(\tilde{x})$, and $f(\tilde{x} - 1)$. We show:

Theorem 1. *For $f : \mathbb{Z} \to \mathbb{R}$, $p \in (0,1)$, $x \in \mathbb{Z}$, and $y \in \mathbb{Z}$ define:*

$$g(y) = f(y) - \tfrac{p}{(1-p)^2}\left(f(y+1) - 2f(y) + f(y-1)\right). \tag{1}$$

Given $\tilde{x} = x + \texttt{DLap}(p)$, if $\mathbb{E}\left[|f(\tilde{x})|\right]$ exists and is finite, then $g(\tilde{x})$ is an unbiased estimator of $f(x)$, i.e., $\mathbb{E}\left[g(\tilde{x})\right] = f(x)$.

The proof of Theorem 1 is a straightforward computation of the expected value of $g(\tilde{x})$ and is deferred to Appendix A. The condition that $\mathbb{E}\left[|f(\tilde{x})|\right]$ exists and is finite is satisfied for every function whose growth rate is subexponential or, more generally, dominated by a certain exponential function depending on p, see Lemma 18 for details.

Comment. The value $f(y+1) - 2f(y) + f(y-1)$ is equal to $(f(y+1) - f(y)) - (f(y) - f(y-1))$ and can be interpreted as a discrete second derivative. This makes our estimator analogous to the unbiased estimator for real-valued functions $f(\tilde{x}) - c_\varepsilon f''(\tilde{x})$ derived in Calmon et al. [2025].

The special case where f is a threshold function, i.e., where for some value k, $f(x) = 1$ if $x \geq k$ and $f(x) = 0$ for $x < k$, was studied previously in Ghazi et al. [2022, Lemma 10]. The unbiased estimator given in their paper coincides with ours in this case. In the univariate case, our estimator in Equation (1) can in fact be recovered via indicator functions $f_k(x) = \mathbf{1}[x = k]$, combining the resulting linear estimators. Our approach operates directly at the level of the function f and extends naturally to arbitrary multivariate functions via tensorization as we discuss below.

Multivariate Case. We next discuss how to extend Theorem 1 to multivariate functions.

Theorem 2. *For $f : \mathbb{Z}^n \to \mathbb{R}$, $p \in (0,1)$, $y \in \mathbb{Z}^n$, define:*

$$g(y) = \sum_{\xi \in \{-1,0,1\}^n} \left(f(y+\xi) \prod_{j=1}^{n} \alpha_{\xi_j} \right) \text{ where } \alpha_{\xi_i} = \begin{cases} 1 + \frac{2p}{(1-p)^2}, & \text{for } \xi_i = 0 \\ \frac{-p}{(1-p)^2}, & \text{for } \xi_i \in \{-1,+1\} \end{cases} \tag{2}$$

For all $i \in [n]$ let $\tilde{x}_i = x_i + \texttt{DLap}(p)$ independently. If $\mathbb{E}\left[|f(\tilde{x})|\right]$ exists and is finite, $g(\tilde{x}_1, \ldots, \tilde{x}_n)$ is an unbiased estimator of f, i.e., $\mathbb{E}\left[g(\tilde{x}_1, \ldots, \tilde{x}_n)\right] = f(x_1, \ldots, x_n)$

The proof of Theorem 2, which can be found in Appendix A, essentially applies the reasoning of Theorem 1 to one variable at a time. A direct implementation of the estimator (2) sums 3^n terms, making it impractical for large n. The sum can be estimated by sampling, but that results in a larger variance. In Section 5 we show how to compute unbiased estimators for specific multivariate functions with much better computational complexity.

3 Properties of the Unbiased Estimator

In this section we show some basic properties of our estimator. First, we show that it is the unique deterministic, unbiased estimator (Section 3.1). Second, we study the mean square error (MSE) of our estimator. We show an upper bound in terms of $\mathrm{MSE}_f(x)$, the MSE of the naive estimator (Section 3.2), and give examples of functions for which the upper bound is essentially tight, as well as examples of functions where our estimator is much better than the naive estimator (Section 3.3).

3.1 Uniqueness of the Unbiased Estimator

We show that the estimator (2) is the only such unbiased, deterministic estimator for discrete Laplace noise. Moreover, if there exists another (potentially randomized) estimator, then its variance cannot be smaller than the one of ours. To obtain those results, we use the Rao-Blackwell Theorem:

Theorem 3 (Rao-Blackwell Theorem (Lehmann and Casella [1998], Theorem 7.8))**.** *Let Y be an observed random variable and let U be auxiliary randomness. Let $\delta(Y,U)$ be an estimator with finite variance, and define $\bar{\delta}(Y) \coloneqq \mathbb{E}\left[\delta(Y,U) \mid Y\right]$. Then*

$$\mathbb{E}\left[\bar{\delta}(Y)\right] = \mathbb{E}\left[\delta(Y,U)\right] \quad \textit{and} \quad \mathrm{Var}\left(\bar{\delta}(Y)\right) \leq \mathrm{Var}\left(\delta(Y,U)\right).$$

We are now ready to state the main claim of the section:

Theorem 4. *Let $n \in \mathbb{N}^*$, $f : \mathbb{Z}^n \to \mathbb{R}$, $p \in (0,1)$, and g the unbiased estimator of Equation* (2)*, and write $\tilde{x} = x + \texttt{DLap}(p)$. Call a deterministic estimator $h : \mathbb{Z}^n \to \mathbb{R}$* admissible *if $\mathbb{E}\left[|h(\tilde{x})|\right] < \infty$ and $\mathbb{E}\left[h(\tilde{x})\right] = f(x)$ for every $x \in \mathbb{Z}^n$. If g is admissible, then it is the only admissible estimator. Moreover, for every $x \in \mathbb{Z}^n$ at which $g(\tilde{x})$ has finite variance, $g(\tilde{x})$ has minimum variance among all estimators, including randomized ones, that are unbiased for $f(x')$ for all $x' \in \mathbb{Z}^n$.*

Note that unbiasedness is required simultaneously for all $x \in \mathbb{Z}^n$: the estimator does not know x, and uniqueness at a single fixed x would be false. The absolute summability condition $\mathbb{E}\left[|h(\tilde{x})|\right] < \infty$ is needed to rearrange the series in the proof; without it, the argument only shows that h solves a homogeneous linear recurrence, which has nonzero solutions such as $h(i) = p^{i_1} - p^{-i_1}$. To prove uniqueness, observe that any two admissible estimators g_1, g_2 for f produce an admissible estimator $g_1 - g_2$ for 0. We show that $g = 0$ is the only admissible estimator for $f = 0$ using cancellation properties. Optimality of g follows directly from Theorem 3. Details can be found in Appendix B.1.

3.2 Bounding MSE of the Unbiased Estimator

For a general function f, we bound $\text{MSE}_g(x)$ as follows:

Theorem 5. *For $f : \mathbb{Z}^n \to \mathbb{R}$, $p \in (0,1)$, $(x_1, \ldots, x_n) \in \mathbb{Z}^n$, and for all $i \in [n]$, $\tilde{x}_i = x_i + \texttt{DLap}(p)$, and g as in Equation* (2)*. Assuming $f(\tilde{x})$ has well defined expectation and variance,*

$$\textit{MSE}_g(x_1, \ldots, x_n) \leq \left(3\frac{\left(1+p^2\right)^2+2p}{(1-p)^4}\right)^n \textit{MSE}_f(x_1, \ldots, x_n).$$

To prove the theorem above, we start by showing a lemma that bounds $\mathbb{E}\left[(f(\tilde{x}+\xi) - f(x))^2\right]$ in terms of the MSE of f, and then applying the lemma. The details of the proof are included in Appendix B.2. This result affirms that $\text{MSE}_g(x)$ can be bounded in terms of $\text{MSE}_f(x)$, though the factor $O(1 + 1/\varepsilon^4)^n$ is exponential in terms of n, and can also be high with low ε. In the next section, we derive an example where this is asymptotically tight, but also show it is possible for the *other* direction to hold—i.e. for $\text{MSE}_g(x)$ to be exponentially *lower* than $\text{MSE}_f(x)$.

3.3 Exponential MSE Gaps Between Estimators

In this section, we first study a family of functions for which the debiasing operator has an especially simple closed form. Let $f_\gamma(y) = \gamma^{S(y)}$, with $S(y) = \sum_i y_i$. These functions are useful test cases because their unbiased estimator is obtained by multiplying the naive estimator by a scalar factor $A(\gamma)^n$. This makes the dependence on the dimension n explicit and shows that the exponential factor in the general multivariate bound is unavoidable.

Proposition 6. *For $p \in (0,1)$, $\gamma \in (-1/p, -p) \cup (p, 1/p)$, $x \in \mathbb{Z}^n$, and $\tilde{x} = x + \texttt{DLap}(p)$, f_γ has a well defined expected value $\mathbb{E}\left[|f_\gamma(\tilde{x})|\right]$ and the unbiased estimator defined in Equation* (2) *is $g_\gamma(y) = A(\gamma)^n f_\gamma(y)$ with*

$$A(\gamma) = 1 - \frac{p}{(1-p)^2}\frac{(1-\gamma)^2}{\gamma}.$$

If additionally $\sqrt{p} < |\gamma| < 1/\sqrt{p}$, then both variances exist and $\text{Var}\left(g_\gamma(\tilde{x})\right) = A(\gamma)^{2n}\,\text{Var}\left(f_\gamma(\tilde{x})\right)$.

This result follows by simple computation. The full proof can be found in Appendix B.3.

Note that for $\gamma \in (-1/p, -p)$ we have $A(\gamma) > 1$, while for $\gamma \in (p, 1/p)$ we have $0 < A(\gamma) \leq 1$, with $A(\gamma) = 1$ only for $\gamma = 1$, where $f_\gamma \equiv 1$ is constant and $g_\gamma = f_\gamma$. This means that unbiasing can both increase or reduce the variance of the estimator depending on the function:

Corollary 7. *For every $p \in (0,1)$, there are functions for which debiasing increases the MSE exponentially in n, and functions for which it decreases the MSE exponentially in n.*

Proof. First we notice that the MSE is equal to the variance for the unbiased estimator, but larger for the naive estimator. Proposition 6 with γ such that $\sqrt{p} < \gamma < 1/\sqrt{p}$, $\gamma \neq 1$ gives an exponential decrease in variance for the unbiased estimator, and thus at least an exponential decrease in the MSE.

For the exponential increase, we set γ to -1. In this case, Proposition 6 proves that the variance of the unbiased estimator is exponentially larger than the one of the naive estimator. Additionally, the value of the naive estimator is in $\{-1, 1\}$ and thus its bias is bounded by 2. This proves that the MSE of the unbiased estimator is also exponentially larger than the one of the naive estimator. □

A Monotone Function with Exponential Variance Growth. As a second example we show that exponential variance growth can occur even for a monotone indicator function where the naive estimator has constant MSE. To simplify calculations we focus on the regime $\varepsilon \le 1$.

Proposition 8. *Let n be even and $f(x) = \mathbf{1}\{\sum_i x_i > 0\}$. Let g be from Theorem 2. If $p = e^{-\varepsilon}$ with $\varepsilon \le 1$, then for every x with $\sum_i x_i = 0$, $\mathrm{Var}\left(g(\tilde{x})\right) = \exp(\Omega(n))$.*

The proof of Proposition 8 can be found in Appendix B.4. The idea is to show that for an all-zeros input the variance is lower bounded by the event that zero-sum noise is added by the Discrete Laplace mechanism, which happens with probability $\Omega(1/\sqrt{n})$, in which case the estimator has value $\exp(\Omega(n))$.

4 Postprocessing Discrete Laplace Noise into Laplace or Staircase Noise

We show that discrete Laplace noise can be converted, by input-independent postprocessing, into either continuous Laplace noise or staircase noise at the same privacy level. The constructions are local to each unit interval and work by adding an independent random variable supported on $[-1, 1]$. It suffices to consider the univariate setting, since the multivariate setting just repeats the same method to each coordinate.

Fix $\varepsilon > 0$ and write $p = e^{-\varepsilon}$. Let $\eta \sim \texttt{DLap}(p)$, so that $\tilde{x} = x + \eta$ is the discrete Laplace mechanism at privacy level ε for unit sensitivity. Since all transformations below are postprocessings of η, they preserve ε-DP automatically. Both our results use the following lemma.

Lemma 9. *Let q be a probability density function with support contained in $[-1, 1]$, symmetric around zero, and let $Y \sim q$ be independent of η. For $z \in \mathbb{R}$, write $|z| = n + u$ with $n \in \mathbb{Z}_{\ge 0}$ and $u \in [0, 1)$. Then $\eta + Y$ has density function*

$$f_{\eta+Y}(z) = \tfrac{1-p}{1+p}\, p^n \big(q(u) + p\, q(u-1)\big) \ .$$

Proof. First suppose $z \ge 0$. Since q is supported on $[-1, 1]$, only the atoms n and $n+1$ can contribute to $z = n + u$. Thus

$$f_{\eta+Y}(n+u) = \Pr[\eta = n]\, q(u) + \Pr[\eta = n+1]\, q(u-1) = \tfrac{1-p}{1+p}\, p^n \big(q(u) + p\, q(u-1)\big) \ .$$

Here we use that $q(u-1) = q(1-u)$ since q is symmetric around zero. For $z < 0$, by symmetry the same formula with $-z = n + u$ determines the density. □

Now we show a method for post-processing from the discrete Laplace mechanism to the continuous Laplace mechanism while preserving the same privacy level $\varepsilon = \ln(1/p)$.

Theorem 10. *Let Y be independent of η with density*

$$q_{\mathrm{Lap}}(y) = \begin{cases} \frac{\ln(1/p)}{2(1-p)^2}\left(p^{|y|} - p^{2-|y|}\right) & \text{for } |y| \le 1 \\ 0 & \text{otherwise} \end{cases} \ .$$

Then $\eta + Y$ has the Laplace density $f(z) = \frac{\ln(1/p)}{2} p^{|z|}$.

Proof. It is easy to verify that q_{Lap} is a density function, is symmetric around zero, and has support $[-1, 1]$. Fix z and write $|z| = n + u$ with $n \ge 0$ and $u \in [0, 1)$. We have

$$q_{\mathrm{Lap}}(u) + p\, q_{\mathrm{Lap}}(1-u) = \tfrac{\ln(1/p)}{2(1-p)^2}\Big((p^u - p^{2-u}) + p(p^{1-u} - p^{1+u})\Big) = \tfrac{(1+p)\ln(1/p)}{2(1-p)}\, p^u \ .$$

By Lemma 9 and using $q_{\mathrm{Lap}}(u-1) = q_{\mathrm{Lap}}(1-u)$:

$$f_{\eta+Y}(z) = \tfrac{1-p}{1+p} p^n \big(q_{\mathrm{Lap}}(u) + p\, q_{\mathrm{Lap}}(u-1)\big) = \tfrac{1-p}{1+p} p^n \cdot \tfrac{(1+p)\ln(1/p)}{2(1-p)}\, p^u = \tfrac{\ln(1/p)}{2} p^{|z|} \ . \quad \square$$

Figure 2 illustrates the transformation. We defer the method and the accompanying proofs for post-processing to the Staircase mechanism to Appendix C.2.

Discussion. We have shown in this section that both the continuous Laplace mechanism and the Staircase mechanism can be simulated by post-processing a draw from the discrete Laplace mechanism. Consequently, any unbiased estimator for either of these mechanisms induces an unbiased estimator based on discrete Laplace output. For example, if g is an unbiased estimator for the continuous Laplace mechanism and Y is independent post-processing noise from Theorem 10, then $y \mapsto g(y + Y)$, evaluated on discrete Laplace output, has the same distribution as g evaluated on continuous Laplace output. The same argument applies to the Staircase mechanism. By Theorem 4, the discrete Laplace unbiased estimator has minimum variance among all unbiased estimators, including randomized ones. Hence, no unbiased estimator based on the continuous Laplace or Staircase mechanism can have smaller variance than the corresponding unbiased estimator based directly on the discrete Laplace mechanism. This fact, along with the possibility to transform the output into a continuous Laplace or Staircase distribution if necessary, means that the discrete Laplace mechanism should be preferred over the Laplace and Staircase mechanisms for discretized quantities.

5 Unbiased Estimators for Specific Multivariate Functions

The computational bottleneck of Theorem 2 is due to the exponential size of the set $B_y = \{y + \xi : \xi \in \{-1, 0, 1\}^n\}$ on which f needs to be evaluated. Fortunately, many common functions of interest exhibit structure which makes it tractable to compute $g(y)$. In this section, we give polynomial time algorithms for unbiasing the min, max, general order statistics, a decision tree evaluation, the entropy function, and multivariate polynomials. These functions all satisfy the conditions of Theorem 2 and have finite variance because they exhibit polynomial growth (formalized in Lemma 18). These results are unified by a general technique which can give efficient unbiased estimators for any function f that is *locally decomposable* as a sum of polynomially many "easy" functions on the domain B_y, for any $y \in \mathbb{Z}^n$. Easy functions refer to a basis class of functions where it is possible to evaluate (2) efficiently. Due to space constraints, we provide the details of this in Appendix D.

Remarks on Running Time and Error. We measure running time in the real-RAM arithmetic model; specifically, we assume that the input y is given as a sequence of n integers, we can evaluate f at any point in constant time (producing a real number), and we can do real-valued arithmetic in constant time in addition to the standard operations of a Turing machine. In preliminary experiments, we observed that some unbiased estimators in this section, namely order statistics and decision trees, can have high error that is possibly much higher than that of the naive estimator (consistent with the gap in Theorem 5). Thus, those results are primarily of theoretical interest, and further work is necessary to demonstrate their practical performance.

Order Statistics For an integer $1 \le i \le n$, the ith order statistic of an input $x_1, \ldots, x_n$ is the ith element in the sorted input. This encapsulates the minimum, median, and maximum. The unbiased estimators for order statistics have the following running times:

Theorem 11. *(Informal statement of Theorems 20, 21). The estimator (2) for min and max can be computed in $O(n)$ time. For the ith order statistic (e.g., median) it can be computed in $O(n^2)$ time.*

Decision Tree Evaluation A decision tree evaluation function can be described by a tree T. Each internal node of the tree is labeled with a predicate of the form $x_i \le c$ with c a real-valued constant. Internal nodes have two children corresponding to YES and NO. Finally, the leaves are labeled with real-valued constants. The function is evaluated at $x_1, \ldots, x_n$ by following the valid path from root to leaf, and then outputting the value at the leaf. We show:

Theorem 12. *(Informal statement of Theorem 19) The unbiased estimator of a decision tree evaluation function with s internal nodes can be computed in $O(ns)$ time.*

Entropy and KL Divergence If $x_1, \ldots, x_n$ is interpreted as a vector of counts over a domain of size n, then it is natural to consider the entropy of the empirical probability distribution $p_1, \ldots, p_n$ where $p_i = \frac{x_i}{\sum_j x_j}$. The entropy takes the form $H(x_1, \ldots, x_n) = \sum_{i=1}^n h(\frac{x_i}{\sum_j x_j})$, where $h(z) = -z \log_2(z)$. The estimator for the entropy has the following runtime:

Theorem 13. *(Informal statement of Theorem 23) The unbiased estimator for the entropy can be computed in $O(n^2)$ time.*

In Section D.5, we extend these results to KL divergence and obtain an unbiased estimator with $O(n^3)$ running time.

Multivariate Polynomials A multivariate monomial of the form $x_1^{a_1} \dots x_n^{a_n}$ does not fit easily into the local decomposition framework. Nonetheless, it is possible to compute the unbiased estimator efficiently because the variables in the monomial interact through a product, and this allows one to factor (2) directly. Specifically, products of functions can be unbiased as the product of the unbiased estimators as follows:

Theorem 14. *Suppose $f(y_1, \dots, y_n) = f_1(Y_1)f_2(Y_2)$ for two functions $f_1 : \mathbb{Z}^{n_1} \to \mathbb{R}$ and $f_2 : \mathbb{Z}^{n_2} \to \mathbb{R}$, where $Y_1 \sqcup Y_2$ is a partition of the variables $y_1, \dots, y_n$. Let $g_1(Y_1), g_2(Y_2)$ be the unbiased estimators given in Theorem 2. Then the unbiased estimator $g(y_1, \dots, y_n)$ for f is equal to $g_1(Y_1)g_2(Y_2)$.*

Thus, a multivariate monomial can be unbiased by multiplying the unbiased estimators for $x_i^{a_i}$ separately (which are given in Theorem 1). A general multivariate polynomial can be unbiased by adding the estimators for each monomial.

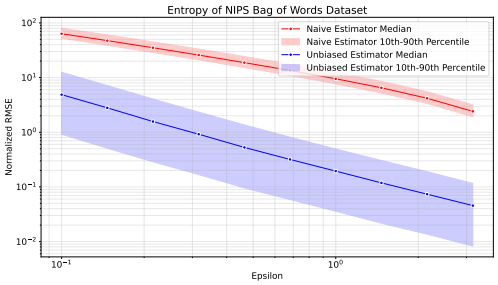
6 Empirical Results

Here we briefly discuss our empirical validation, for which further details can be found in Appendix E. Section E.1 considers polynomials, for a use case in graph counting (k-star counting), where we emphasize the advantage of the discrete Laplace mechanism over the classical Laplace mechanism for high values of ε. The k-stars experiments suggest that both the use of the discrete Laplace mechanism and the debiasing matter.

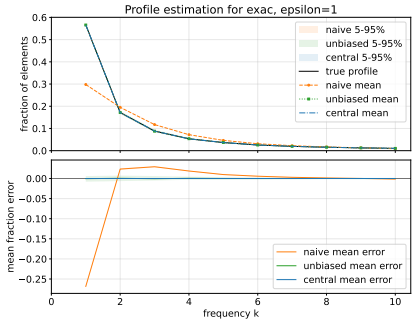
Section E.2 concerns empirical entropy, where we see a significantly improved root mean square error for low and medium values of ε. The entropy experiment, whose results are shown in Figure 1a, highlights that when the naive estimator consists of a sum of biased terms, debiasing yields a significant improvement when applied to a sparse bag-of-words dataset.

Other functions, such as partition functions, also fall into the category of functions written as a sum over disjoint variables. We consider empirical results on estimating partition functions in Section E.3, for a setting in which the naive estimator has considerable bias.

Finally, Section E.4 provides the first practical demonstration of post-processing-based profile estimation. One of the results is shown in Figure 1b. These results show that debiasing can help prevent the smoothing effect of the noise addition. Overall, debiasing is most compelling when bias would otherwise accumulate. Our empirical results show that even in non-distributed settings, there are functions for which our unbiased estimator leads to lower error than the naive plug-in estimator. In particular, they confirm the theoretical contributions and the applicability of the proposed method.



(a) Root mean square error of the entropy estimation on the bag of words of NIPS articles. Quantiles are shown over about 30,000 runs for each ε (20 repetitions over 1491 histograms).



(b) Profile estimation on ExAC histogram. Quantiles are shown for $\varepsilon = 1$ and for over 200 noise realizations for each value of k.

Figure 1: Examples where our estimator compares favorably to the naive estimator on real datasets.

7 Discussion and Limitations

Our proposed method is general and imposes minimal assumptions on the properties of f. Compared with previous work, this generality allows unbiased estimators to be constructed for a much wider class of functions. One cost of this generality is computational: in general computing our estimator requires exponential time. For many natural functions we show, however, that the estimator can be computed in polynomial time. A natural open problem is that of characterizing the classes of functions that admit polynomial-time unbiased estimation algorithms.

This is the case for some of the functions we have studied, such as order statistics. However, as we have seen in our theoretical and empirical results there are other cases of interest where the mean square error of our estimator is small or even significantly better than that of the naive estimator. Characterizing the functions and inputs for which such an improvement occurs is another interesting avenue for future work. Finally, unbiased estimators sometimes have worse error than estimators with a small, controlled bias — it would be interesting to understand if such biased estimators can lead to better estimation of sums in distributed/federated settings.

8 Acknowledgments

The authors were supported by a Data Science Distinguished Investigator grant from Novo Nordisk Fonden. The work was primarily carried out at BARC, supported by the VILLUM Foundation grant 54451.

References

Flavio Calmon, Elbert Du, Cynthia Dwork, Brian Finley, and Grigory Franguridi. Debiasing functions of private statistics in postprocessing. In Mark Bun, editor, *Symposium on Foundations of Responsible Computing (FORC)*, volume 329 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 17:1–17:18, Dagstuhl, Germany, 2025. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. ISBN 978-3-95977-367-6. doi: 10.4230/LIPIcs.FORC.2025.17.

Albert Cheu, Adam Smith, Jonathan Ullman, David Zeber, and Maxim Zhilyaev. Distributed differential privacy via shuffling. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques (EUROCRYPT)*, pages 375–403. Springer, 2019.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam D. Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography (TCC)*, volume 3876 of *Lecture Notes in Computer Science*, pages 265–284. Springer, 2006. doi: 10.1007/11681878_14.

Úlfar Erlingsson, Vitaly Feldman, Ilya Mironov, Ananth Raghunathan, Kunal Talwar, and Abhradeep Thakurta. Amplification by shuffling: From local to central differential privacy via anonymity. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2468–2479. SIAM, 2019.

Michael Friendly and Gavin Klorfine. *vcdExtra: 'vcd' Extensions and Additions*, 2026. URL `https://friendly.github.io/vcdExtra/`. R package version 0.9.9.

Quan Geng and Pramod Viswanath. The optimal mechanism in differential privacy. In *IEEE International Symposium on Information Theory (ISIT)*, pages 2371–2375. IEEE, 2014.

Quan Geng and Pramod Viswanath. Optimal noise adding mechanisms for approximate differential privacy. *IEEE Transactions on Information Theory*, 62(2):952–969, 2015.

Badih Ghazi, Pritish Kamath, Ravi Kumar, and Pasin Manurangsi. Anonymized histograms in intermediate privacy models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Arpita Ghosh, Tim Roughgarden, and Mukund Sundararajan. Universally utility-maximizing privacy mechanisms. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC)*, pages 351–360, 2009.

Quentin Hillebrand, Vorapong Suppakitpaisarn, and Tetsuo Shibuya. Unbiased locally private estimator for polynomials of Laplacian variables. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pages 741–751, 2023.

Jacob Imola, Takao Murakami, and Kamalika Chaudhuri. Locally differentially private analysis of graph statistics. In *USENIX Security Symposium (USENIX Security)*, pages 983–1000, 2021.

Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. Inherent trade-offs in the fair determination of risk scores. In *Innovations in Theoretical Computer Science Conference (ITCS)*, 2017.

E. L. Lehmann and George Casella. *Theory of Point Estimation*. Springer-Verlag, 1998. ISBN 0387985026. doi: 10.1007/b98854.

Monkol Lek, Konrad Karczewski, Eric Minikel, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature 536*, pages 285–291, 2016. doi: https://doi.org/10.1038/nature19057.

Jure Leskovec and Julian McAuley. Learning to discover social circles in ego networks. In F. Pereira, C.J. Burges, L. Bottou, and K. Weinberger, editors, *Advances in Neural Information Processing Systems (NeurIPS)*, volume 25. Curran Associates, Inc., 2012.

Ephraim Linder, Sofya Raskhodnikova, Adam Smith, and Thomas Steinke. Privately evaluating untrusted black-box functions. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing (STOC)*, 2025.

David Newman. Bag of Words. UCI Machine Learning Repository, 2008. DOI: https://doi.org/10.24432/C5ZG6P.

Günter F. Steinke and Thomas Steinke. Privately estimating black-box statistics. In *Conference on Learning Theory (COLT)*, volume 336 of *Proceedings of Machine Learning Research*, pages 6029–6074. PMLR, 2026. Preprint: arXiv:2510.00322.

Hao Wu and Rasmus Pagh. Profile reconstruction from private sketches. In *International Conference on Machine Learning (ICML)*, 2024.

A Omitted Proofs for Section 2

A.1 Proof of Theorem 1

Proof. Let $f : \mathbb{Z} \to \mathbb{R}$, $p \in (0,1)$, $x \in \mathbb{Z}$, $\tilde{x} = x + \texttt{DLap}(p)$, and g as defined in Equation (1).

$$
\begin{aligned}
\mathbb{E}\left[g(\tilde{x})\right] = &\sum_{i=-\infty}^{+\infty} \Pr\left[\tilde{x} - x = i\right] g(x+i) \\
=&\frac{1-p}{1+p} \sum_{i=-\infty}^{+\infty} p^{|i|} g(x+i) \\
=&\frac{1-p}{1+p} \sum_{i=-\infty}^{+\infty} p^{|i|} \left(f(x+i) - \frac{p}{(1-p)^2}(f(x+i+1) - 2f(x+i) + f(x+i-1)) \right) \\
=&\frac{(1-p)^2+2p}{(1+p)(1-p)} \sum_{i=-\infty}^{+\infty} p^{|i|} f(x+i) - \frac{p}{(1+p)(1-p)} \left[\sum_{i=-\infty}^{-1} p^{-i} f(x+i+1) \right. \\
&\left. + \sum_{i=0}^{+\infty} p^{i} f(x+i+1) + \sum_{i=-\infty}^{0} p^{-i} f(x+i-1) + \sum_{i=1}^{+\infty} p^{i} f(x+i-1) \right] \\
=&\frac{1+p^2}{(1+p)(1-p)} \sum_{i=-\infty}^{+\infty} p^{|i|} f(x+i) - \frac{p}{(1+p)(1-p)} \\
&\left[\sum_{i=-\infty}^{0} p^{-i+1} f(x+i) + \sum_{i=1}^{+\infty} p^{i-1} f(x+i) + \sum_{i=-\infty}^{-1} p^{-i-1} f(x+i) + \sum_{i=0}^{+\infty} p^{i+1} f(x+i) \right] \\
=&\frac{1+p^2}{(1+p)(1-p)} \sum_{i=-\infty}^{+\infty} p^{|i|} f(x+i) - \frac{p}{(1+p)(1-p)} \\
&\left[p \sum_{i=-\infty}^{+\infty} p^{|i|} f(x+i) + p f(x) + \frac{1}{p} \sum_{i=-\infty}^{+\infty} p^{|i|} f(x+i) - \frac{1}{p} f(x) \right] \\
=&f(x)
\end{aligned}
$$

□

A.2 Proof of Theorem 2

Proof. Let $f : \mathbb{Z}^n \to \mathbb{R}$, $p \in (0,1)$, $(x_1, \ldots, x_n) \in \mathbb{Z}^n$, for all $i \in [n]$, $\tilde{x}_i = x_i + \texttt{DLap}(p)$, and g as defined in Equation (2).

$$
\begin{aligned}
&\mathbb{E}_{\tilde{x}_1,\ldots,\tilde{x}_n}\left[g(\tilde{x}_1,\ldots,\tilde{x}_n)\right] \\
&= \mathbb{E}_{\tilde{x}_1,\ldots,\tilde{x}_n}\left[\sum_{(\xi_1,\ldots,\xi_n)\in\{-1,0,1\}^n} f(\tilde{x}_1+\xi_1,\ldots,\tilde{x}_n+\xi_n)\prod_{j=1}^{n}\alpha_{\xi_j}\right] \\
&= \mathbb{E}_{\tilde{x}_1,\ldots,\tilde{x}_{n-1}}\left[\mathbb{E}_{\tilde{x}_n}\left[\sum_{(\xi_1,\ldots,\xi_n)\in\{-1,0,1\}^n} f(\tilde{x}_1+\xi_1,\ldots,\tilde{x}_n+\xi_n)\prod_{j=1}^{n}\alpha_{\xi_j}\right]\right] \\
&= \mathbb{E}_{\tilde{x}_1,\ldots,\tilde{x}_{n-1}}\left[\sum_{(\xi_1,\ldots,\xi_{n-1})\in\{-1,0,1\}^{n-1}}\prod_{j=1}^{n-1}\alpha_{\xi_j}\cdot\mathbb{E}_{\tilde{x}_n}\left[-\frac{p}{(1-p)^2}f(\tilde{x}_1+\xi_1,\ldots,\tilde{x}_{n-1}+\xi_{n-1},\tilde{x}_n-1)\right.\right. \\
&\left.\left.+\left(1+\frac{2p}{(1-p)^2}\right)f(\tilde{x}_1+\xi_1,\ldots,\tilde{x}_{n-1}+\xi_{n-1},\tilde{x}_n)-\frac{p}{(1-p)^2}f(\tilde{x}_1+\xi_1,\ldots,\tilde{x}_{n-1}+\xi_{n-1},\tilde{x}_n+1)\right]\right] \\
&= \mathbb{E}_{\tilde{x}_1,\ldots,\tilde{x}_{n-1}}\left[\sum_{(\xi_1,\ldots,\xi_{n-1})\in\{-1,0,1\}^{n-1}} f(\tilde{x}_1+\xi_1,\ldots,\tilde{x}_{n-1}+\xi_{n-1},x_n)\prod_{j=1}^{n-1}\alpha_{\xi_j}\right] \\
&= f(x_1,\ldots,x_n)
\end{aligned}
$$

The second equality is derived using the independence of the random variables $\tilde{x}_i$. For the fourth equality, we apply Theorem 1 to the functions $g(x) = f(\tilde{x}_1+\xi_1,\ldots,\tilde{x}_{n-1}+\xi_{n-1},x)$ for each assignment of $(\xi_1,\ldots,\xi_{n-1})$. Finally the last equality is obtained by induction on n by considering this time the function $g(x_1,\ldots,x_{n-1}) = f(x_1,\ldots,x_{n-1},x_n)$. □

B Omitted Proofs for Section 3

B.1 Proof of Theorem 4

For $y \in \mathbb{Z}^n$ we will prove that if g is an unbiased estimator of the function $y \mapsto 0$, then for all $y \in \mathbb{Z}^n, g(y) = 0$. From there, we can easily prove the theorem by noticing that if g_1 and g_2 are unbiased estimators of the same function f, then $g_1 - g_2$ is an unbiased estimator of the constant function $y \mapsto 0$. Let $p \in (0,1)$, $x \in \mathbb{Z}^n$, $\tilde{x} = x + \texttt{DLap}(p)$, and h be a deterministic unbiased estimator for $y \mapsto 0$. By the definition of the unbiased estimator, we have

$$\mathbb{E}\left[h(\tilde{x})\right] = \left(\tfrac{1-p}{1+p}\right)^n \sum_{i\in\mathbb{Z}^n} p^{\|i-x\|_1} h(i) = 0.$$

To obtain a cancellation in the sum, we will use the following equality for integer z:

$$p^{|z-1|} + p^{|z+1|} - \tfrac{p^2+1}{p}p^{|z|} = \begin{cases} p-\frac{1}{p}, & z = 0 \\ 0, & \text{otherwise} \end{cases}$$

Fix $x \in \mathbb{Z}^n$. Since $\tilde{x} \pm e_1$ is distributed as $(x \pm e_1) + \texttt{DLap}(p)$, and h is unbiased for 0 at the datasets $x - e_1$, $x + e_1$, and x, the following linear combination is also equal to 0. Here, rearranging the three series into a single iterated sum is justified by Fubini's theorem, because $\mathbb{E}\left[|h(\tilde{x}')|\right] < \infty$ for every $x' \in \mathbb{Z}^n$ makes each series absolutely convergent:

$$
\begin{aligned}
&\mathbb{E}\left[h(\tilde{x}-e_1)\right] + \mathbb{E}\left[h(\tilde{x}+e_1)\right] - \tfrac{p^2+1}{p}\mathbb{E}\left[h(\tilde{x})\right] \\
&= \left(\tfrac{1-p}{1+p}\right)^n \sum_{(i_2,\ldots,i_n)\in\mathbb{Z}^{n-1}} p^{\sum_{j=2}^n |i_j - x_j|} \sum_{i_1=-\infty}^{+\infty}\left[p^{|i_1-x_1-1|} + p^{|i_1-x_1+1|} - \tfrac{p^2+1}{p}p^{|i_1-x_1|}\right] h(i) \\
&= \left(\tfrac{1-p}{1+p}\right)^n \left(p - \tfrac{1}{p}\right) \sum_{(i_2,\ldots,i_n)\in\mathbb{Z}^{n-1}} p^{\sum_{j=2}^n |i_j - x_j|} h(x_1, i_2, \ldots, i_n) \\
&= \tfrac{1-p}{1+p}\left(p - \tfrac{1}{p}\right)\mathbb{E}\left[h(x_1, \tilde{x}_2, \ldots, \tilde{x}_n)\right]
\end{aligned}
$$

This shows that the zero expectation is preserved when replacing the random variable $\tilde{x}_1$ by x_1; note that the resulting $(n-1)$-dimensional series is again absolutely convergent. By performing the same cancellation for each coordinate j (each step again uses unbiasedness at the neighboring datasets $x \pm e_j$), we obtain $h(x) = 0$. This is true for every $x \in \mathbb{Z}^n$ and concludes the proof of the uniqueness of the deterministic estimator.

Concerning the optimality, let U be auxiliary randomness and $\delta(\tilde{x}, U)$ an unbiased estimator of $f(x)$. Using Theorem 3, we have that $\bar{\delta}(\tilde{x})$ is a deterministic estimator with $\mathbb{E}\left[\bar{\delta}(\tilde{x})\right] = f(x')$ for all $x' \in \mathbb{Z}^n$ and with variance at most that of $\delta(\tilde{x}, U)$ at every x' where the latter is finite. Moreover, $\mathbb{E}\left[|\bar{\delta}(\tilde{x})|\right] \leq \mathbb{E}\left[|\delta(\tilde{x}, U)|\right] < \infty$ by Jensen's inequality, so $\bar{\delta}$ is admissible. By the uniqueness of the admissible estimator, $\bar{\delta} = g$. □

B.2 Proof of Theorem 5

For simplicity, we first consider the $n = 1$ case: First, we state and prove the following lemma:

Lemma 15. *For $f : \mathbb{Z} \to X$, $p \in (0, 1)$, $x \in \mathbb{Z}$, $\tilde{x} = x + \texttt{DLap}(p)$, and $\xi \in \{-1, +1\}$,*

$$\mathbb{E}\left[(f(\tilde{x} + \xi) - f(x))^2\right] \leq \frac{1}{p} \cdot \mathit{MSE}_f(x).$$

Proof.

$$\begin{aligned}
\mathbb{E}\left[(f(\tilde{x} + \xi) - f(x))^2\right] &= \frac{1-p}{1+p} \sum_{i=-\infty}^{+\infty} p^{|i|} (f(x + i + \xi) - f(x))^2 \\
&= \frac{1-p}{1+p} \sum_{j=-\infty}^{+\infty} p^{|j-\xi|} (f(x + j) - f(x))^2 \\
&\leq \frac{1-p}{1+p} \sum_{j=-\infty}^{+\infty} \frac{p^{|j|}}{p} (f(x + j) - f(x))^2 \\
&= \frac{1}{p} \cdot \mathrm{MSE}_f(x)
\end{aligned}$$

□

Then, we prove the theorem for $n = 1$ by applying the lemma above:

Proof.

$$\begin{aligned}
\mathrm{MSE}_g(x) &= \mathbb{E}\left[\left(f(\tilde{x}) - \frac{p}{(1-p)^2}(f(\tilde{x}+1) - 2f(\tilde{x}) + f(\tilde{x}-1)) - f(x)\right)^2\right] \\
&= \frac{1}{(1-p)^4} \mathbb{E}\left[\left(\left(1+p^2\right)(f(\tilde{x}) - f(x)) - p\,(f(\tilde{x}+1) - f(x)) - p\,(f(\tilde{x}-1) - f(x))\right)^2\right] \\
&\leq \frac{3}{(1-p)^4} \mathbb{E}\left[\left(1+p^2\right)^2 (f(\tilde{x}) - f(x))^2 + p^2\,(f(\tilde{x}+1) - f(x))^2 + p^2\,(f(\tilde{x}-1) - f(x))^2\right] \\
&\leq \frac{3}{(1-p)^4}\left(\left(1+p^2\right)^2 \mathrm{MSE}_f(x) + p \cdot \mathrm{MSE}_f(x) + p \cdot \mathrm{MSE}_f(x)\right) \\
&= 3\frac{\left(1+p^2\right)^2 + 2p}{(1-p)^4} \cdot \mathrm{MSE}_f(x).
\end{aligned}$$

□

Now we consider the general case. Before proving the main theorem, we state and prove the following lemma (similar to the $n = 1$ case):

Lemma 16. *For $f : \mathbb{Z}^n \to X$, $p \in (0, 1)$, $(x_1, \ldots, x_n) \in \mathbb{Z}^n$, and for all $i \in [n]$, $\tilde{x}_i = x_i + \texttt{DLap}(p)$, and $(\xi_1, \ldots, \xi_n) \in \{-1, 0, 1\}^n$,*

$$\mathbb{E}\left[(f(\tilde{x}_1 + \xi_1, \ldots, \tilde{x}_n + \xi_n) - f(x_1, \ldots, x_n))^2\right] \leq \frac{1}{p^{\sum_{j=1}^n |\xi_j|}} \cdot \mathit{MSE}_f(x_1, \ldots, x_n).$$

Proof.

$$
\begin{aligned}
&\mathbb{E}\left[(f(\tilde{x}_1+\xi_1,\ldots,\tilde{x}_n+\xi_n)-f(x_1,\ldots,x_n))^2\right]\\
&=\left(\frac{1-p}{1+p}\right)^n \sum_{(i_1,\ldots,i_n)\in\mathbb{Z}^n} p^{\sum_{j=1}^n |i_j|}(f(x_1+\xi_1+i_1,\ldots,x_n+\xi_n+i_n)-f(x_1,\ldots,x_n))^2\\
&=\left(\frac{1-p}{1+p}\right)^n \sum_{(i_1,\ldots,i_n)\in\mathbb{Z}^n} p^{\sum_{j=1}^n |i_j-\xi_j|}(f(x_1+i_1,\ldots,x_n+i_n)-f(x_1,\ldots,x_n))^2\\
&\le\left(\frac{1-p}{1+p}\right)^n \sum_{(i_1,\ldots,i_n)\in\mathbb{Z}^n} p^{\sum_{j=1}^n |i_j|-|\xi_j|}(f(x_1+i_1,\ldots,x_n+i_n)-f(x_1,\ldots,x_n))^2\\
&=\frac{1}{p^{\sum_{j=1}^n|\xi_j|}}\cdot\mathbb{E}\left[(f(\tilde{x}_1,\ldots,\tilde{x}_n)-f(x_1,\ldots,x_n))^2\right].
\end{aligned}
$$

□

Then, we prove Theorem 5 by applying the lemma above:

Proof.

$$
\begin{aligned}
&\mathrm{MSE}_g(x_1,\ldots,x_n)\\
&=\mathbb{E}\left[\left(\sum_{(\xi_1,\ldots,\xi_n)\in\{-1,0,1\}^n} f(\tilde{x}_1+\xi_1,\ldots,\tilde{x}_n+\xi_n)\prod_{j=1}^n\alpha_{\xi_j}-f(x_1,\ldots,x_n)\right)^2\right]\\
&\le 3^n\sum_{(\xi_1,\ldots,\xi_n)\in\{-1,0,1\}^n}\mathbb{E}\left[(f(\tilde{x}_1+\xi_1,\ldots,\tilde{x}_n+\xi_n)-f(x_1,\ldots,x_n))^2\right]\prod_{j=1}^n\alpha_{\xi_j}^2\\
&\le 3^n\sum_{(\xi_1,\ldots,\xi_n)\in\{-1,0,1\}^n}\mathrm{MSE}_f(x_1,\ldots,x_n)\prod_{j=1}^n\frac{\alpha_{\xi_j}^2}{p^{|\xi_j|}}\\
&=\left(3\frac{\left(1+p^2\right)^2+2p}{(1-p)^4}\right)^n\mathrm{MSE}_f(x_1,\ldots,x_n).
\end{aligned}
$$

□

B.3 Proof of Proposition 6

Proof. Let $p\in(0,1)$, $\gamma\in(-1/p,-p)\cup(p,1/p)$, $x\in\mathbb{Z}^n$, and $\tilde{x}=x+\mathtt{DLap}(p)$. In one dimension,

$$
\mathbb{E}\left[|\gamma^{\tilde{x}}|\right]=\tfrac{1-p}{1+p}\sum_{i=-\infty}^{+\infty}p^{|i|}|\gamma|^{x+i}=\tfrac{1-p}{1+p}|\gamma|^x\left(\sum_{i=1}^{+\infty}\left(\frac{p}{|\gamma|}\right)^i+\sum_{i=0}^{+\infty}(p|\gamma|)^i\right).
$$

As $|\gamma|\in(p,1/p)$, both $p/|\gamma|<1$ and $p|\gamma|<1$. Thus the expected value is finite. This carries to the multidimensional case as the noise added to each component are independent and thus $\mathbb{E}\left[|f_\gamma(\tilde{x})|\right]=\prod_{i=1}^n\mathbb{E}\left[|\gamma^{\tilde{x}_i}|\right]$.

We use the same independence property to get the unbiased estimator. It is useful to express this using the operator form of the estimator. Let

$$
\Delta_i f(y)=f(y+e_i)-2f(y)+f(y-e_i)
$$

denote the discrete second derivative in coordinate i, and write $T_i=I-c\Delta_i$, $c=\frac{p}{(1-p)^2}$. Then Theorem 2 can be written compactly as

$$
g=\prod_{i=1}^n T_i f,
$$

where the operators commute because they act on different coordinates. As $\Delta_i(f_\gamma(y))=\frac{(1-\gamma)^2}{\gamma}f_\gamma(y)$, this gives $g_\gamma(y)=A(\gamma)^n f_\gamma(y)$.

Last, whenever the variance of $f_\gamma(\tilde{x})$ exists, it follows immediately that the variance of $g_\gamma(\tilde{x})$ exists as well and that $\mathrm{Var}\left(g_\gamma(\tilde{x})\right) = A(\gamma)^{2n}\,\mathrm{Var}\left(f_\gamma(\tilde{x})\right)$. The existence of that variance is given by the previous claim on expected value by noticing that $\mathrm{Var}\left(f_\gamma(\tilde{x})\right)$ converges exactly when $\mathbb{E}\left[f_{\gamma^2}(\tilde{x})\right]$ converges. □

B.4 Proof of Proposition 8

Proof. Write $\alpha_0 = 1 + 2c$, $\alpha_{+1} = \alpha_{-1} = -c$, where $c = p/(1-p)^2$. For y with $\sum_i y_i = 0$,

$$g(y) = \sum_{\xi:\,\sum_i \xi_i > 0} \prod_i \alpha_{\xi_i}.$$

Let $\beta_k = \sum_{\xi:\,\sum_i \xi_i = k} \prod_i \alpha_{\xi_i}$. Then $\beta_k = \beta_{-k}$ and $\sum_k \beta_k = 1$, so $g(y) = \sum_{k>0} \beta_k = \frac{1-\beta_0}{2}$. Since all terms in the sum defining β_0 are non-negative we can lower bound β_0 using vectors ξ with $n/2$ entries $+1$ and $n/2$ entries -1:

$$\beta_0 \geq \binom{n}{n/2} c^n \geq \tfrac{(2c)^n}{n+1}.$$

For $\varepsilon \leq 1$ we have $p \geq e^{-1}$ and thus $c = p/(1-p)^2 > 0.92$. Consider the event E that $\sum_i \xi_i = 0$. By standard concentration bounds, $\Pr[E] = \Theta(n^{-1/2})$. Conditioned on E, we have $\sum_i \tilde{x}_i = 0$, so

$$|g(\tilde{x})| = \tfrac{|1-\beta_0|}{2} \geq \Omega\big((2c)^n/n\big).$$

Since $\mathbb{E}\left[g(\tilde{x})\right] = 0$,

$$\mathrm{Var}\left(g(\tilde{x})\right) \geq \Pr[E] \cdot \Omega\big((2c)^{2n}/n^2\big) = \Omega\big((2c)^{2n}/n^{5/2}\big),$$

which is exponential since $2c > 1$. □

C Details on Transformations to Other Mechanisms

C.1 Illustration of the Conversion to the Laplace Mechanism

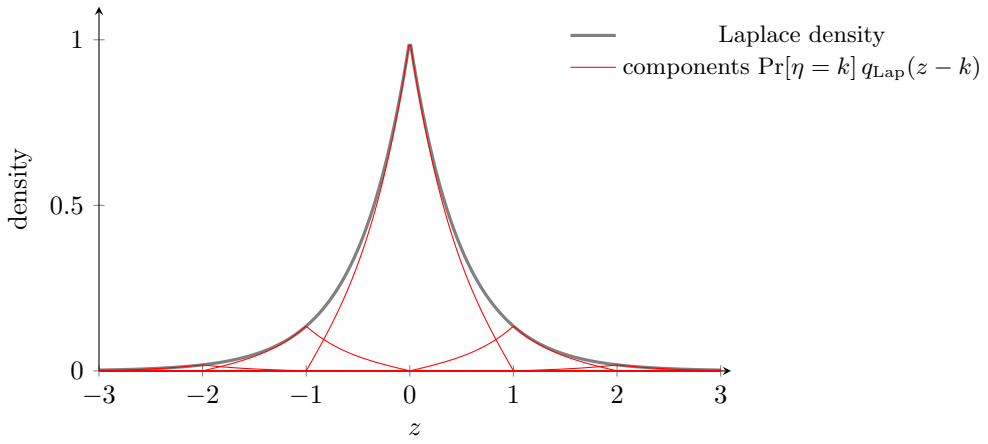


Figure 2: Illustration of Theorem 10 for $\varepsilon = 2$ (so $p = e^{-2}$). Each red curve is one translated component $\Pr[\eta = k]\, q_{\mathrm{Lap}}(z-k)$ corresponding to an atom of the discrete Laplace distribution $\eta \sim \mathrm{DLap}(e^{-\varepsilon})$. Summing all components yields exactly the continuous Laplace density $\frac{\varepsilon}{2} e^{-\varepsilon|z|}$ shown in black.

C.2 Conversion to the Staircase Mechanism

We next treat the Staircase noise distribution Geng and Viswanath [2014, 2015] in the standard form used in the differential privacy literature. Like for the Laplace mechanisms we use the parameter $p = e^{-\varepsilon}$ and further introduce a shape parameter $\gamma \in [0, 1/2]$. Define the symmetric density f_γ by

$$f_\gamma(x) = \begin{cases} a_\gamma\, p^k & \text{for } |x| \in [k, k+\gamma),\ k \in \mathbb{Z}_{\geq 0} \\ a_\gamma\, p^{k+1} & \text{for } |x| \in [k+\gamma, k+1),\ k \in \mathbb{Z}_{\geq 0} \end{cases}$$

where $a_\gamma = \frac{1-p}{2(\gamma + p(1-\gamma))}$. Additive noise using this density function is the staircase mechanism with shape parameter γ, satisfying ε-differential privacy Geng and Viswanath [2014, 2015].

Theorem 17. *Fix $\gamma \in [0, 1/2]$. Let η be sampled from the discrete Laplace distribution with parameter p, and let Y_γ be independent of η with density*

$$q_\gamma(y) = \begin{cases} \frac{1+p}{2(\gamma+p(1-\gamma))} & \textit{for } |y| < \gamma \\ \frac{p}{2(\gamma+p(1-\gamma))} & \textit{for } \gamma \leq |y| \leq 1-\gamma \\ 0 & \textit{otherwise} \end{cases}.$$

Then $\eta + Y_\gamma$ has density f_γ.

Proof. The function q_γ is symmetric and nonnegative. Its total mass (the integral over all reals) is

$$2\gamma \cdot \tfrac{1+p}{2(\gamma+p(1-\gamma))} + 2(1-2\gamma) \cdot \tfrac{p}{2(\gamma+p(1-\gamma))} = 1,$$

so it is indeed a density. For any z we write $|z| = n + u$ with $n \geq 0$ and $u \in [0, 1)$, and apply Lemma 9. If $u < \gamma$, then $q_\gamma(u) = \frac{1+p}{2(\gamma+p(1-\gamma))}$ and $q_\gamma(u-1) = 0$, because $|u-1| > 1-\gamma$. Therefore

$$f_{X+Y_\gamma}(z) = \tfrac{1-p}{1+p} p^n \cdot \tfrac{1+p}{2(\gamma+p(1-\gamma))} = a_\gamma p^n.$$

If $\gamma \leq u < 1-\gamma$, then $q_\gamma(u) = q_\gamma(u-1) = \frac{p}{2(\gamma+p(1-\gamma))}$, and hence

$$f_{X+Y_\gamma}(z) = \tfrac{1-p}{1+p} p^n \cdot \tfrac{p(1+p)}{2(\gamma+p(1-\gamma))} = a_\gamma p^{n+1}.$$

If $1-\gamma \leq u$, then $q_\gamma(u) = 0$ and $q_\gamma(u-1) = \frac{1+p}{2(\gamma+p(1-\gamma))}$, so again

$$f_{X+Y_\gamma}(z) = \tfrac{1-p}{1+p} p^n \cdot p \tfrac{1+p}{2(\gamma+p(1-\gamma))} = a_\gamma p^{n+1}.$$

This matches exactly the definition of f_γ. □

The transformation is illustrated in Figure 3.

D Omitted Details from Section 5

In Section D.1, we describe the general framework for obtaining faster unbiased formula. In the remaining sections, we use this framework to obtain faster estimators for various functions.

The functions in this section satisfy the condition of Theorem 2, and thus have unbiased estimators, because they all exhibit subexponential growth which ensures their expectation exists and is finite. We formalize this in the following lemma.

Lemma 18. *For $p \in (0, 1)$, suppose $|f(y)| \leq r(\|y\|_1)$, where $r(k) : \mathbb{R} \to \mathbb{R}$ is an increasing function such that $\lim_{k\to\infty} r(k)p^{k/4} = 0$. For a given $x \in \mathbb{Z}^d$, let $\tilde{x} = x + \mathtt{DLap}(p)^d$. Then, $\mathbb{E}[f(\tilde{x})]$ exists and is finite.*

Proof. We will show that the expectation converges to a finite value. To do so, observe the expectation can be written as a limit of the following partial sums:

$$\begin{aligned} \mathbb{E}[f(\tilde{x})] &= \sum_{z\in\mathbb{Z}^d} f(x+z) \left(\frac{1-p}{1+p}\right)^n p^{\|z\|_1} \\ &= \lim_{k\to\infty} \sum_{z:\|z\|_1 \leq k} f(x+z) \left(\frac{1-p}{1+p}\right)^n p^{\|z\|_1}. \end{aligned}$$

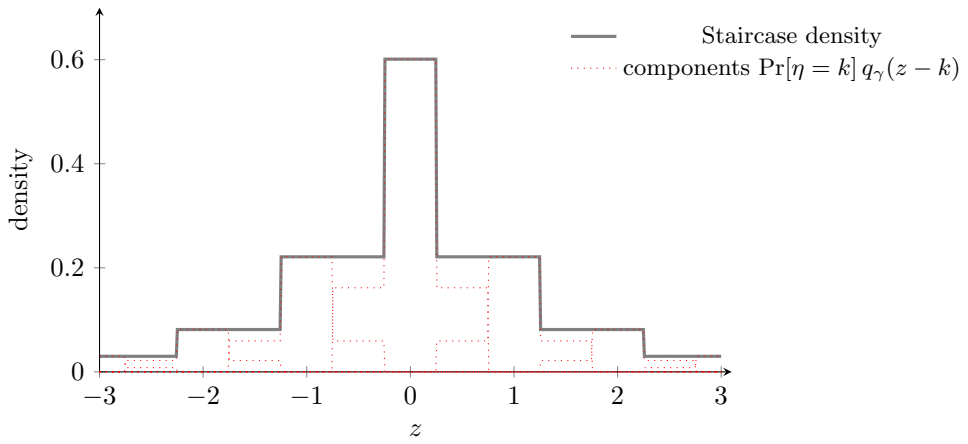


Figure 3: Illustration of Theorem 17 for $\varepsilon = 1$ and $\gamma = \frac{1}{4}$ (so $p = e^{-1}$). Each red dashed curve is one translated component $\Pr[\eta = k]\, q_\gamma(z-k)$ corresponding to an atom of the discrete Laplace distribution $\eta \sim \mathrm{DLap}(e^{-\varepsilon})$. Summing all components yields exactly the Staircase density shown in black.

Define the partial sum $P(k) = \sum_{z:\|z\|_1 \le k} f(x+z)p^{\|z\|_1}$. We will show that for any $\alpha > 0$, there exists k_0 such that for all $k_1, k_2 \ge k_0$ such that $k_1 < k_2$, we have $|P(k_2) - P(k_1)| \le \alpha$, and thus the above limit exists.

To do this, we bound

$$\begin{aligned}
|P(k_2) - P(k_1)| &= \left| \sum_{k=k_1}^{k_2} \sum_{z:\|z\|_1=k} f(x+z)p^k \right| \\
&\le \sum_{k=k_1}^{k_2} \sum_{z:\|z\|_1=k} |f(x+z)|p^k \\
&\le \sum_{k=k_1}^{k_2} \sum_{z:\|z\|_1=k} p^k r\left(\|x+z\|_1\right) \\
&\le \sum_{k=k_1}^{k_2} \sum_{z:\|z\|_1=k} p^k r\left(\|x\|_1 + \|z\|_1\right) \\
&= \sum_{k=k_1}^{k_2} \sum_{z:\|z\|_1=k} p^k r\left(\|x\|_1 + k\right) \\
&\le \sum_{k=k_1}^{k_2} 2^n (n+k)^n p^k r\left(\|x\|_1 + k\right),
\end{aligned}$$

where the last step follows because $|\{z : \|z\|_1 = k\}\| = 2^n \binom{n+k-1}{n-1} \le 2^n (n+k)^n$. k_0 can be made large enough so that $2^n(n+k)^n p^{k/4} < 1$ for all $k \ge k_0$. By the property of r, k_0 can be made large enough so that $r(\|x\|_1 + k)p^{k/2} \le r(2k)p^{k/2} \le \beta$ for all $k \ge k_0$. This means the above sum can be bounded as $\sum_{k=k_1}^{\infty} p^{k/4}\beta < \frac{\beta}{1-p^{1/4}}$, and existence is established by taking $\beta < \alpha(1-p^{1/4})$. Having shown the expectation exists, we can establish its finiteness by repeating the above argument on $P(k)$, showing $|P(k)| \le \sum_{i=0}^{k} 2^n(n+k)^n p^k r(\|x\|_1 + k) < \infty$. □

D.1 A General Debiasing Technique Via Local Decompositions

The unbiased estimator of Theorem 2 depends only on the values of f in the set $y + \{-1, 0, 1\}^n$. Suppose that f takes a constant value C on a region $y + R$ where $R \subseteq \{-1, 0, 1\}^d$. Then, the total contribution of the region $y + R$ to the unbiased estimator g is C times

$$\sum_{\xi \in R} \prod_{j=1}^{n} \alpha_{\xi_j} \triangleq \text{vol}(R),$$

which can be interpreted as a type of volume of R. Thus, if f can be decomposed into a polynomial number of regions on the entire space $y + \{-1, 0, 1\}^n$ *and* each region has an efficiently computable volume, then $g(y)$ can be efficiently computed. The precise notion of decomposability we need is:

Definition 4. Let $\mathcal{B}$ denote a set of subsets of $\{-1, 0, 1\}^n$. We will refer to such a $\mathcal{B}$ as a basis. We say a function $f : \mathbb{Z}^n \to \mathbb{R}$ is m-locally decomposable under a basis $\mathcal{B}$ if for any point $y \in \mathbb{Z}^n$, there exist sets $S_1, \ldots, S_m \in \mathcal{B}$ and coefficients $c_1, \ldots, c_m \in \mathbb{R}$ such that $f(y + \xi) = \sum_{i=1}^{m} c_i \mathbf{1}[\xi \in S_i]$ for all $\xi \in \{-1, 0, 1\}^n$.

Observe that the decomposition may be different for each point y and only needs to hold in a local neighborhood $y + \{-1, 0, 1\}^n$. We say that f is *efficiently* decomposable under $\mathcal{B}$ if the volume of each $S \in \mathcal{B}$ is efficiently computable, f is m-locally decomposable with $m = \text{poly}(n)$, and the decomposition $S_1, \ldots, S_m, c_1, \ldots, c_m$ can be efficiently computed for any y.

By combining Equation (2) with Definition 4, the unbiased estimator of f can be written as

$$g(y) = \sum_{i=1}^{m} c_i \text{vol}(S_i), \tag{3}$$

and this equation is efficiently computable as f is efficiently decomposable under $\mathcal{B}$.

A particularly nice basis is the set $\mathcal{R}_n$ of n-dimensional rectangles. A rectangle in $\{-1, 0, 1\}^n$ is a Cartesian product $R = R_1 \times \cdots \times R_n$, where each $R_i \subseteq \{-1, 0, 1\}$. We can compute the volume of rectangles easily: the volume of each R_i is $\text{vol}(R_i) = \sum_{r \in R_i} \alpha_r$, and volume of R is given by $\text{vol}(R) = \prod_{i=1}^{n} \text{vol}(R_i)$. Equation (3) gives an efficient method for computing the unbiased estimator of any efficiently decomposable function under $\mathcal{R}_n$. The next three sections will use this technique, with different choices of bases, to give efficient unbiased estimators for specific functions.

D.2 Decision Tree Regression

It is easy to see that a decision tree regression function with s internal nodes partitions the input space $\mathbb{Z}^n$ into $s + 1$ rectangles. This means it can be written as $f(x) = \sum_{i=1}^{s+1} c_i \mathbf{1}[x \in R_i]$ for rectangles $R_1, R_2, \ldots, R_{s+1} \subseteq \mathbb{Z}^n$. This global property is enough to imply local decomposability under rectangles, and allow us to obtain an efficient estimator:

Theorem 19. *A decision tree regression function f with s internal nodes is $s + 1$-decomposable under $\mathcal{R}_n$, and its corresponding estimator is*

$$g(x) = \sum_{i=1}^{s+1} c_i \text{vol}((R_i - x) \cap \{-1, 0, 1\}^n).$$

Since the volume of a rectangle may be computed in $O(n)$ time, we can compute the unbiased estimator in $O(ns)$ time.

Proof. For $\xi \in \{-1, 0, 1\}^n$, the decision tree regression is given by $f(x + \xi) = \sum_{i=1}^{s+1} c_i \mathbf{1}[x + \xi \in R_i] = \sum_{i=1}^{s+1} c_i \mathbf{1}[\xi \in R_i - x]$. As each $R_i - x$ is not necessarily a subset of $\{-1, 0, 1\}^n$, we express the function equivalently as $\sum_{i=1}^{s+1} c_i \mathbf{1}[\xi \in (R_i - x) \cap \{-1, 0, 1\}^n]$. The unbiased estimator then follows from (3). □

D.3 Order Statistics

The minima and maxima of x also admit efficient local decompositions under $\mathcal{R}_n$. This is because, for the points in $x + \{-1, 0, 1\}^n$, the max or min can take on three values, which can be written as a disjunction (sum) of three cases, each of which is a conjunction (rectangle). For example, in the case where $\min(x + \xi) = \min(x) + 1$, it holds that $\xi_i = +1$ for all i where $x_i = \min(x)$ and $\xi_i \in \{0, 1\}$ for all i such that $x_i = \min(x) + 1$. The other cases can be expressed similarly. This yields the following:

Theorem 20. *The minimum and maximum functions are* 3*-locally decomposable under* $\mathcal{R}_n$*, and the corresponding estimator is* $\min(x) - 1 + (\alpha_0 + \alpha_1)^{\underline{n}_0} + (\alpha_1)^{\underline{n}_0}(\alpha_0 + \alpha_1)^{\underline{n}_1}$ *for the minimum and* $\max(x) + 1 - (\alpha_0 + \alpha_{-1})^{\overline{n}_0} - (\alpha_{-1})^{\overline{n}_0}(\alpha_0 + \alpha_{-1})^{\overline{n}_1}$*, for the maximum, where* $\underline{n}_0$ *(resp.* $\overline{n}_0$*) is the number of times in* x *that* $\min(x)$ *(resp.* $\max(x)$*) appears, and* $\underline{n}_1$ *(resp.* $\overline{n}_1$*) is the number of times in* x *that* $\min(x) + 1$ *(resp.* $\max(x) - 1$*) appears.*

The above estimators can be computed in $O(n)$ time.

Proof. First, we will start with the minimum. Let $k = \min(x)$. Observe that $\min(x + \xi)$ is equal to either $k + 1$, k or $k - 1$ when $\xi \in \{-1, 0, 1\}^n$. Let $A = \{i \in [n] : x_i = k\}$ and $B = \{i \in [n] : x_i = k + 1\}$. We have $\min(x + \xi) = k - 1$ if and only if at least one index in A satisfies $\xi_i = -1$. This set can be written as a difference of rectangles $R_1 \setminus R_2$, where $R_1 = \{-1, 0, 1\}^n$ and $R_2 = \{0, 1\}^A \times \{-1, 0, 1\}^{[n]\setminus A}$.

Next, $\min(x + \xi) = k$ if $\xi_i \geq 0$ for all indices in A, and $\xi_i = 0$ for at least one index in A, or $\xi_i = 1$ for all indices in A and $\xi_i = -1$ for at least one index in B. The first set of indices can be written as $R_2 \setminus R_3$, where $R_3 = \{1\}^A \times \{-1, 0, 1\}^{[n]\setminus A}$. The second set of indices can be written as $R_3 \setminus R_4$, where $R_4 = \{1\}^A \times \{0, 1\}^B \times \{-1, 0, 1\}^{[n]\setminus A\setminus B}$. Finally, $\min(x + \xi) = k + 1$ precisely when $\xi \in R_4$. This establishes that $\min$ can be written as

$$\begin{aligned}\min(x + \xi) &= (k - 1)\mathbf{1}[\xi \in R_1] - (k - 1)\mathbf{1}[\xi \in R_2] + k\mathbf{1}[\xi \in R_2] - k\mathbf{1}[\xi \in R_3] \\ &\quad + k\mathbf{1}[\xi \in R_3] - k\mathbf{1}[\xi \in R_4] + (k + 1)\mathbf{1}[\xi \in R_4] \\ &= (k - 1)\mathbf{1}[\xi \in R_1] + \mathbf{1}[\xi \in R_2] + \mathbf{1}[\xi \in R_4].\end{aligned}$$

Therefore, $\min$ is 3-decomposable. Thus, by (3), the unbiased estimator is

$$(k - 1)\text{vol}(R_1) + \text{vol}(R_2) + \text{vol}(R_4) = k - 1 + (\alpha_0 + \alpha_1)^{|A|} + (\alpha_1)^{|A|}(\alpha_0 + \alpha_1)^{|B|}.$$

We take a similar approach for the maximum. Let $k' = \max(x)$, $A' = \{i \in [n] : x_i = k'\}$, $B' = \{i \in [n] : x_i = k' - 1\}$, $R'_2 = \{-1, 0\}^{A'} \times \{-1, 0, 1\}^{[n]\setminus A'}$, $R'_3 = \{-1\}^{A'} \times \{-1, 0, 1\}^{[n]\setminus A'}$, and $R'_4 = \{-1\}^{A'} \times \{-1, 0\}^{B'} \times \{-1, 0, 1\}^{[n]\setminus A'\setminus B'}$. We can show that $\max$ is 3-locally decomposable as

$$\begin{aligned}\max(x + \xi) &= (k' + 1)\mathbf{1}[\xi \in R_1] - (k' + 1)\mathbf{1}[\xi \in R'_2] + k'\mathbf{1}[\xi \in R'_2] - k'\mathbf{1}[\xi \in R'_3] \\ &\quad + k'\mathbf{1}[\xi \in R'_3] - k'\mathbf{1}[\xi \in R'_4] + (k' - 1)\mathbf{1}[\xi \in R'_4] \\ &= (k' + 1)\mathbf{1}[\xi \in R_1] - \mathbf{1}[\xi \in R'_2] - \mathbf{1}[\xi \in R'_4].\end{aligned}$$

The unbiased formula is given by

$$k' + 1 - (\alpha_0 + \alpha_{-1})^{|A'|} - (\alpha_{-1})^{|A'|}(\alpha_0 + \alpha_{-1})^{|B'|}. \qquad \square$$

General order statistics, which return element of rank i in x for a given i, can also be computed efficiently. To do so, we require a more complex basis. For a given set $A \subseteq [n]$ and an integer $1 \leq a \leq |A|$, define the set

$$\begin{aligned}S^-_{A,a} &= \{\xi \in \{-1, 0, 1\}^A : \textstyle\sum_{i=1}^n \mathbf{1}[\xi_i = -1] = a\} \\ S^+_{A,a} &= \{\xi \in \{-1, 0, 1\}^A : \textstyle\sum_{i=1}^n \mathbf{1}[\xi_i = +1] = a\},\end{aligned}$$

or the vectors where exactly a of n coordinates are equal to 1. The volume of these sets are given by

$$\begin{aligned}\text{vol}(S^-_{A,a}) &= \binom{|A|}{a}\alpha_{-1}^a(\alpha_0 + \alpha_1)^{|A|-a} \\ \text{vol}(S^+_{A,a}) &= \binom{|A|}{a}\alpha_{+1}^a(\alpha_{-1} + \alpha_0)^{|A|-a}.\end{aligned}$$

Our basis will be the following collection of sets

$$\mathcal{S} = \left\{ S^+_{A,a} \times S^-_{B,b} \times \{-1,0,1\}^{[n]\setminus A\setminus B} : A, B \subseteq [n], A \cap B = \emptyset, 0 \le a \le |A|, 0 \le b \le |B| \right\}.$$

These are sets with both rectangular and symmetric structure, and their volume can be computed by multiplying the above volume formulae.

We now derive an efficient unbiased estimator for the ith order statistic using $\mathcal{S}$ as the basis.

Theorem 21. *The ith order statistic is $O(n^2)$-locally decomposable under $\mathcal{S}$, and the corresponding unbiased estimate is computable in $O(n^2)$ time.*

Proof. Define $k = x_{(i)}$, or the element in x of rank i. Observe that f will return either $k-1, k$, or $k+1$ in the set $x + \{-1,0,1\}^n$. Define the sets of indices $A = \{j \in [n] : x_j = k-1\}$, $B = \{j \in [n] : x_j = k\}$, and $C = \{j \in [n] : x_j = k+1\}$. Let ℓ_1, ℓ_2 denote the maximum index with rank $k-1, k$; i.e. $\ell_1 = \sum_{i=1}^n \mathbf{1}[x_i \le k-1]$ and $\ell_2 = \sum_{i=1}^n \mathbf{1}[x_i \le k]$. Define $d_1 = i - \ell_1$ and $d_2 = \ell_2 - i + 1$. These are the "distances" that the rank i element is from being equal to $k-1$ or $k+1$.

We will begin by decomposing the set where $f(x+\xi) = k-1$. For $\xi \in \{-1,0,+1\}^n$, let a^+ denote the number of indices $j \in A$ such that $\xi_j = +1$. Similarly, let b^- denote the number of indices $j \in B$ such that $\xi_j = -1$. Thus, the element of rank i will be $k-1$ if and only if $b^- - a^+ \ge d_1$. This means the ξ where $f(x+\xi) = k-1$ can be decomposed as the following disjoint union of sets

$$\bigsqcup_{b=0}^{|B|} \bigsqcup_{a=0}^{b-d_1} S^+_{A,a} \times S^-_{B,b} \times \{-1,0,1\}^{[n]\setminus A\setminus B}.$$

Denote the above region as $R_{a,b} = S^+_{A,a} \times S^-_{B,b} \times \{-1,0,1\}^{[n]\setminus A\setminus B}$. The volumes of the regions are

$$\begin{aligned}\text{vol}(R_{a,b}) &= \text{vol}(S^+_{A,a})\text{vol}(S^-_{B,b}) \\ &= \binom{|A|}{a}\binom{|B|}{b}\alpha_1^a(\alpha_{-1}+\alpha_0)^{|A|-a}\alpha_{-1}^b(\alpha_0+\alpha_1)^{|B|-b}.\end{aligned}$$

Similarly, let b^+ denote the number of indices $j \in B$ such that $x_j = 1$, and c^- denote the number of indices $j \in C$ such that $x_j = -1$. This means $f(x+\xi) = k+1$ if and only if $b^+ - c^- \ge d_2$, corresponding to the region

$$\bigsqcup_{b=0}^{|B|} \bigsqcup_{c=0}^{b-d_2} S^+_{B,b} \times S^-_{C,c} \times \{-1,0,1\}^{[n]\setminus B\setminus C}.$$

Denote the above regions as $R'_{b,c}$. Their volumes are

$$\begin{aligned}\text{vol}(R'_{b,c}) &= \text{vol}(S^+_{B,b})\text{vol}(S^-_{C,c}) \\ &= \binom{|B|}{b}\binom{|C|}{c}\alpha_1^b(\alpha_{-1}+\alpha_0)^{|B|-b}\alpha_{-1}^c(\alpha_0+\alpha_1)^{|C|-c}.\end{aligned}$$

Finally, let $R = \{-1,0,+1\}^n$. We can represent $f(x+\xi)$ on the whole domain by starting with $k\mathbf{1}[\xi \in R]$, and then using the two unions above to either add $+1$ or -1. Formally, the representation is

$$f(x+\xi) = k\mathbf{1}[\xi \in R] - \sum_{b=0}^{|B|}\sum_{a=0}^{b-d_1}\mathbf{1}[\xi \in R_{a,b}] + \sum_{b=0}^{|B|}\sum_{c=0}^{b-d_2}\mathbf{1}[\xi \in R'_{c,b}],$$

showing that f is $O(n^2)$-decomposable among symmetric rectangles. Using the volume formulae, the corresponding unbiased estimator is

$$\begin{aligned}g(x) = k &- \sum_{b=0}^{|B|}\sum_{a=0}^{b-d_1}\binom{|A|}{a}\binom{|B|}{b}\alpha_1^a(\alpha_{-1}+\alpha_0)^{|A|-a}\alpha_{-1}^b(\alpha_0+\alpha_1)^{|B|-b} \\ &+ \sum_{b=0}^{|B|}\sum_{c=0}^{b-d_2}\binom{|B|}{b}\binom{|C|}{c}\alpha_1^b(\alpha_{-1}+\alpha_0)^{|B|-b}\alpha_{-1}^c(\alpha_0+\alpha_1)^{|C|-c}.\end{aligned}$$

To prove the running time guarantee, we can precompute the binomial coefficients up to n in $O(n^2)$ time, as well as the exponents α_1^i, $(\alpha_{-1}+\alpha_0)^i$ up to n. Then, the above estimator is a sum of $O(n^2)$ terms, resulting in a total time of $O(n^2)$. □

D.4 Entropy

The entropy takes the form $H(x_1, \dots, x_n) = \sum_{i=1}^n h(\frac{x_i}{\sum_j x_j})$, where $h(z) = -z \log_2(z)$. As our results apply to the entire domain $\mathbb{Z}^n$, we need to extend the domain of the entropy function. To do so, we use an extension of the form

$$h(x, y) = \begin{cases} 0 & x \le 0 \text{ or } y \le 0 \\ h(\min\{\frac{x}{y}, 1\}) & \text{otherwise} \end{cases}$$

and write $H(x_1, \dots, x_n) = \sum_{i=1}^n h(x_i, \sum_j x_j)$. There are many possible extensions of entropy to negative counts, and as noted in Calmon et al. [2025], different extensions can produce different MSEs. We use the above extension because it is simple and bounded, which is sufficient for our purposes.

The structure of this function is particularly appealing for Definition 4 because, on the domain $x + \{-1, 0, 1\}^n$, each $h(x_i, \sum_j x_j)$ depends only on $x_i + \xi_i$ and $\sum_j x_j + \xi_j$. This means that for each i, $h(x_i, \sum_j x_j)$ is constant on the set $P^i_{a,b} = \{\xi \in \{-1, 0, 1\}^n : \xi_i = a, \sum_j \xi_j = b\}$. This suggests the following basis:

$$\mathcal{P}_n = \{P^i_{a,b} : 1 \le i \le n, a \in \{-1, 0, 1\}, b \in \{-n, -n+1, \dots, n\}\}.$$

Furthermore, each $P^i_{a,b}$ is a highly symmetric set, and its volume can be computed efficiently using trinomial coefficients as follows:

Lemma 22. *The volume of each $P^i_{a,b}$ satisfies*

$$\text{vol}(P^i_{a,b}) = \sum_{\substack{0 \le r \le n \\ b+r \text{ is even}}} \alpha_0^{n-r} \alpha_1^r \binom{n-1}{n-1-r+|a|, \frac{b-a+r-|a|}{2}, \frac{r-|a|-b+a}{2}}.$$

In the above formula, $\binom{n}{a,b,c}$ with integers $a + b + c = n$ denotes the trinomial coefficient given by $\frac{n!}{a!b!c!}$. If any of a, b, c are negative, we extend the notation to evaluate to 0.

Proof. The product $\prod_{j=1}^n \alpha_{\xi_j}$ simplifies to $\alpha_{-1}^{n_{-1}} \alpha_0^{n_0} \alpha_1^{n_1}$, where n_{-1}, n_0, n_1 is the number of -1, 0s, and 1s in ξ, respectively. Since $\alpha_{-1} = \alpha_1$, this simplifies to $\alpha_0^{n_0} \alpha_1^{n-n_0}$, or equivalently $\alpha_0^{n-\|\xi\|_1} \alpha_1^{\|\xi\|_1}$. Let $l_1(r) = \{\xi \in \{-1, 0, 1\}^n : \|\xi\|_1 = r\}$ denote the l_1 shell of radius r. We have

$$\begin{aligned} \text{vol}(P^i_{a,b}) &= \sum_{\xi \in P^i_{a,b}} \alpha_0^{n-\|\xi\|_1} \alpha_1^{\|\xi\|_1} \\ &= \sum_{r=0}^n \sum_{\xi \in P^i_{a,b} \cap l_1(r)} \alpha_0^{n-r} \alpha_1^r \\ &= \sum_{r=0}^n \alpha_0^{n-r} \alpha_1^r |P^i_{a,b} \cap l_1(r)|. \end{aligned}$$

Each vector $\xi \in P^i_{a,b} \cap l_1(r)$ satisfies $\xi_i = a$, $\sum_{j \ne i} \xi_j = b - a$, and $\sum_{j \ne i} |\xi_j| = r - |a|$. An equivalent property for membership in $P^i_{a,b} \cap l_1(r)$ is that $\xi_i = a$ and for coordinates $i \ne j$, there are $n - 1 - r + |a|$ equal to 0, $\frac{b-a+r-|a|}{2}$ equal to $+1$, and $\frac{-b+a+r-|a|}{2}$ equal to -1. Note that if r, b have different parity, then $P^i_{a,b} \cap l_1(r) = \emptyset$. Thus,

$$|P^i_{a,b} \cap l_1(r)| = \begin{cases} \binom{n-1}{n-1-r+|a|, \frac{b-a+r-|a|}{2}, \frac{-b+a+r-|a|}{2}} & b + r \text{ is even} \\ 0 & \text{otherwise} \end{cases},$$

and the result is shown. □

Having computed the volumes of each set in $\mathcal{P}_n$, we are able to give an efficient unbiased estimator for $H(x_1, \dots, x_n)$.

Theorem 23. *The entropy function is $O(n^2)$-locally decomposable under $\mathcal{P}_n$, and its corresponding unbiased estimator is given by*

$$g(x_1, \ldots, x_n) = \sum_{i=1}^{n} \sum_{a \in \{-1,0,1\}} \sum_{b=-n}^{n} \text{vol}(P^i_{a,b}) h\left(x_i + a, b + \sum_{j=1}^{n} x_j\right).$$

Using memoization, the $O(n^2)$ trinomial coefficients ahead of time requires $O(n^2)$ time. Having access to each coefficient, each $\text{vol}(P^i_{a,b})$ can be computed in $O(n)$ time. Since $\text{vol}(P^i_{a,b})$ does not depend on i, the volumes for each possible a, b can all be computed in $O(n^2)$ time. The estimator in Theorem 23 is a sum over $O(n^2)$ terms, and thus the total running time is $O(n^2)$.

Proof. For all $\xi \in P^i_{a,b}$, $h(x + \xi_i, \sum_j x_j + \xi_j)$ takes the value $h(x_i + a, b + \sum_j x_j)$. Furthermore, the entire space $\{-1, 0, 1\}^n$ is partitioned by $P^i_{a,b}$ for $a \in \{-1, 0, 1\}$ and $b \in \{-n, \ldots, n\}$. Thus, we can write

$$h\left(x_i + \xi_i, \sum_{j=1}^{n} x_j + \xi_j\right) = \sum_{a \in \{-1,0,1\}} \sum_{b=-n}^{n} \mathbf{1}[\xi \in P^i_{a,b}] h\left(x_i + a, b + \sum_{j=1}^{n} x_j\right),$$

and this gives the decomposition

$$H(x_1 + \xi_1, \ldots, x_n + \xi_n) = \sum_{i=1}^{n} \sum_{a \in \{-1,0,1\}} \sum_{b=-n}^{n} \mathbf{1}[\xi \in P^i_{a,b}] h\left(x_i + a, b + \sum_{j=1}^{n} x_j\right).$$

The unbiased estimator follows by applying (3). □

D.5 KL Divergence

Assuming a domain of size n, one can compute the KL divergence between the empirical distributions of two different counts $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$. This has the following formula: $KL(x_1, \ldots, x_n \| y_1, \ldots, y_n) = \sum_{i=1}^{n} kl(\frac{x_i}{\sum_j x_j}, \frac{y_i}{\sum_j y_j})$, where $kl(x, y) = x \log(\frac{x}{y})$. Again, because our results apply to arbitrary $x, y \in \mathbb{Z}^n$, we must extend $kl(x, y)$ to the entire integral domain. We do so by defining

$$kl(x_1, x_2, y_1, y_2) = \begin{cases} 0 & x_1 \le 0, x_2 \le 0, y_1 \le 0, \text{ or } y_2 \le 0 \\ kl(\min\{1, \frac{x_1}{x_2}\}, \min\{1, \frac{y_1}{y_2}\}) & \end{cases}$$

which satisfies $KL(x_1, \ldots, x_n \| y_1, \ldots, y_n) = \sum_{i=1}^{n} kl(x_i, \sum_j x_j, y_i, \sum_j y_j)$. Each term $kl(x_i, \sum_j x_j, y_i, \sum_j y_j)$ is actually constant on the set $P^i_{a_1,b_1} \times P^i_{a_2,b_2}$ for any a_1, b_1, a_2, b_2. Thus, we will use the following basis:

$$\mathcal{P}^2_n = \{P^i_{a_1,b_1} \times P^i_{a_2,b_2} : 1 \le i \le n; a_1, a_2 \in \{-1, 0, 1\}; b_1, b_2 \in \{-n, \ldots, n\}\}.$$

Furthermore, because $\text{vol}(A \times B) = \text{vol}(A)\text{vol}(B)$, we have that $\text{vol}(P^i_{a_1,b_1} \times P^i_{a_2,b_2}) = \text{vol}(P^i_{a_1,b_1})\text{vol}(P^i_{a_2,b_2})$ and thus can be computed efficiently by Lemma 22. Using similar techniques as the entropy, we can obtain an unbiased estimator:

Theorem 24. *The KL divergence is $O(n^3)$-locally decomposable under $\mathcal{P}^2_n$, and its corresponding unbiased formula is*

$$g(x_1, \ldots, x_n, y_1, \ldots, y_n)$$
$$= \sum_{i=1}^{n} \sum_{a_1, a_2 \in \{-1,0,1\}} \sum_{b_1=-n}^{n} \sum_{b_2=-n}^{n} \text{vol}(P^i_{a_1,b_1})\text{vol}(P^i_{a_2,b_2}) kl\left(x_i + a_1, b_1 + \sum_{j=1}^{n} x_j, y_i + a_2, b_2 + \sum_{j=1}^{n} y_j\right).$$

By memoizing $\text{vol}(P^i_{a,b})$ for each possible a, b (taking $O(n^2)$ time) and then computing the final sum over $O(n^3)$ terms, the above estimator has running time $O(n^3)$.

Proof. For convenience, write $\xi'_i = \xi_{i+n}$. For all $\xi \in P^i_{a_1,b_1} \times P^i_{a_2,b_2}$, the value of $kl(x_i+\xi_i, \sum_j x_j + \xi_j, y_i+\xi'_i, \sum_j y_j+\xi'_j)$ is $kl(x_i+a_1, b_1+\sum_j x_j, y_i+a_2, b_2+\sum_j y_j)$. The entire space $\{-1,0,1\}^{2n}$ is partitioned by the $P^i_{a_1,b_1} \times P^i_{a_2,b_2}$ for $a_1, a_2 \in \{-1,0,1\}$ and $b_1, b_2 \in \{-n, \ldots, n\}$. Thus, we can write

$$
\begin{aligned}
&kl\left(x_i + \xi_i, \sum_j x_j + \xi_j, y_i + \xi'_i, \sum_j x'_j\right) \\
&= \sum_{a_1,a_2 \in \{-1,0,1\}} \sum_{b_1=-n}^{n} \sum_{b_2=-n}^{n} \mathbf{1}[\xi \in P^i_{a_1,b_1} \times P^i_{a_2,b_2}] kl\left(x_i + a_1, b_1 + \sum_j x_j, y_i + a_2, b_2 + \sum_j y_j\right),
\end{aligned}
$$

and the the KL divergence can be written as

$$
\begin{aligned}
&KL(x_1 + \xi_1, \ldots, x_n + \xi_n \| y_1 + \xi'_1, \ldots, y_n + \xi'_n) \\
&= \sum_{i=1}^{n} \sum_{a_1,a_2 \in \{-1,0,1\}} \sum_{b_1=-n}^{n} \sum_{b_2=-n}^{n} \mathbf{1}[\xi \in P^i_{a_1,b_1} \times P^i_{a_2,b_2}] kl\left(x_i + a_1, b_1 + \sum_j x_j, y_i + a_2, b_2 + \sum_j y_j\right),
\end{aligned}
$$

yielding the result. □

D.6 Proof of Theorem 14

Proof. Let I_1, I_2 partition $\{1, \ldots, n\}$ such that $Y_1 = (y_j)_{j \in I_1}$ and $Y_2 = (y_j)_{j \in I_2}$. Rearranging the definition of g,

$$
\begin{aligned}
g(y_1, \ldots, y_n) &= \sum_{(\xi_1,\ldots,\xi_n) \in \{-1,0,1\}^n} f(y_1 + \xi_1, \ldots, y_n + \xi_n) \prod_{j=1}^{n} \alpha_{\xi_j} \\
&= \sum_{(\xi_1,\ldots,\xi_n) \in \{-1,0,1\}^n} f_1(Y_1 + \xi^{(1)}) f_2(Y_2 + \xi^{(2)}) \prod_{j \in I_1} \alpha_{\xi_j} \prod_{j \in I_2} \alpha_{\xi_j},
\end{aligned}
$$

where $\xi^{(1)} = (\xi_j)_{j \in I_1}, \xi^{(2)} = (\xi_j)_{j \in I_2}$ is the partitioning of $(\xi_1, \ldots, \xi_n)$ into the coordinates corresponding to Y_1, Y_2. We can also partition the sum this way—observe that the sum over $(\xi_1, \ldots, \xi_n) \in \{-1,0,1\}^n$ is the same as a sum over pairs $(\xi^{(1)}, \xi^{(2)}) \in \{-1,0,1\}^{I_1} \times \{-1,0,1\}^{I_2}$. Thus, the above sum becomes

$$
\begin{aligned}
&\sum_{\xi^{(1)} \in \{-1,0,1\}^{I_1}, \xi^{(2)} \in \{-1,0,1\}^{I_2}} f_1(Y_1 + \xi^{(1)}) f_2(Y_2 + \xi^{(2)}) \prod_{j \in I_1} \alpha_\xi \prod_{j \in I_2} \alpha_\xi \\
&= \left(\sum_{\xi^{(1)} \in \{-1,0,1\}^{I_1}} f_1(Y_1 + \xi^{(1)}) \prod_{j \in I_1} \alpha_{\xi_j}\right) \left(\sum_{\vec{\xi}_2 \in \{-1,0,1\}^{I_2}} f_2(Y_2 + \xi^{(2)}) \prod_{j \in I_2} \alpha_P x i_j\right) \\
&= g_1(Y_1) g_2(Y_2).
\end{aligned}
$$

□

E Empirical Results

In this appendix we present details of our empirical validation of our estimator Equation (2). We study four different types of non-linear functions:

- Polynomials, for which we emphasize the advantage of the discrete Laplace mechanism over the classical Laplace mechanism for high values of ε.
- Empirical entropy, where we see a significantly improved mean square error for low- and medium values of ε.
- The partition function, for which we see improved estimation near the convergence radius boundary.
- Profiles, for which we give the first empirical results showing that accurate estimation is feasible in practice.

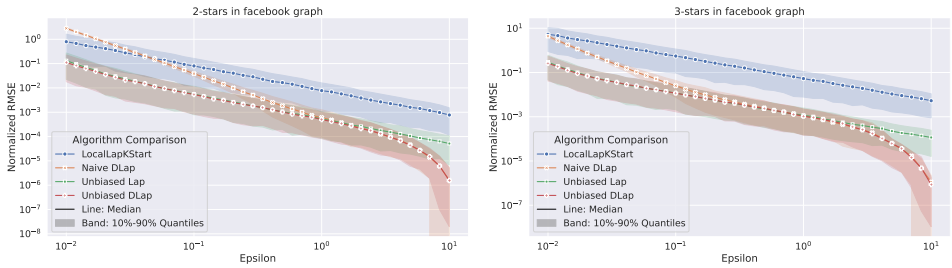


Figure 4: Root mean square error the number of k-stars (normalized) under varying privacy budgets.

E.1 Estimation of Polynomials

For the first application, we consider publishing the number of k-stars in a graph. A k-star is a subgraph consisting of a central node connected to k neighbors. As such, it can be expressed as $\sum_i \binom{d_i}{k}$ with d_i the degree of node i and $\binom{y}{k}$ a polynomial in y. We address the problem in the edge local differential privacy setting, where two adjacency vectors are considered neighbors if they have a Hamming distance of one. This enables us to publish each degree with Laplace or discrete Laplace with parameter $1/\varepsilon$. Note that this is the same setting as the one considered in Hillebrand et al. [2023] and Imola, Murakami, and Chaudhuri [2021].

We evaluate this task on the Facebook dataset from Leskovec and McAuley [2012]. This graph, where each node represents a user and edges represent friendships, contains 4039 nodes and 88,234 edges.

We compare four algorithms:

- Unbiased DLap: the number of k-stars is computed using our algorithm by post-processing the degrees published with discrete Laplace.
- Naive DLap: the number of k-stars is naively computed on the degrees published with discrete Laplace.
- Unbiased Lap: the number of k-stars is computed using the algorithm from Hillebrand et al. [2023] by post-processing the degrees published with continuous Laplace.
- LocalLapKStars: the method described in Imola et al. [2021] that does not rely on post-processing.

The results, presented on Figure 4, show that our method outperforms the three other methods across all privacy budgets. We see that the gap with LocalLapKStars is large (at least two orders of magnitude) for all budgets. Naive DLap has a larger error with a sharper slope than Unbiased DLap until a threshold around $\varepsilon = 1$ after which the two methods have similar performances. When compared to Unbiased Lap, the MSE follows the one of Unbiased DLap until $\varepsilon = 1$ after which the error of Unbiased DLap shifts downward with an exponential decay.

These results show that our method obtains the best performance across all budgets with the biggest improvement over Unbiased Lap being in the low privacy regime, and the one over Naive DLap being in the high privacy regime. This shows that both the Discrete Laplace distribution and the unbiasing are necessary to obtain optimal performances.

E.2 Estimation of Entropy

In the second application, we study the estimation of the entropy of the distribution given by a histogram. We focus on the *bounded setting*, i.e., we assume that the sum of all columns s is a publicly known value and that two histograms are neighbors if they have an ℓ_1-sensitivity of 2. The data consists of a histogram with $n \in \mathbb{N}^*$ columns, where the number of elements x_i in column $i \in [n]$ gets published with discrete Laplace calibrated with sensitivity 2. The goal is to estimate, from the published counts, the value of the entropy

$$\sum_{i=1}^{n} \frac{x_i}{s} \log \frac{s}{x_i}.$$

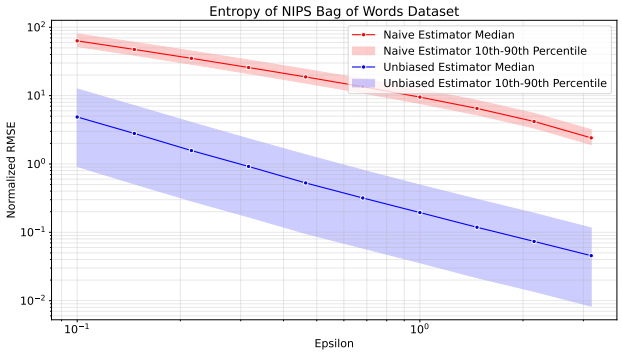


Figure 5: Mean square error of the entropy estimation on the bag of words of NIPS articles. Quantiles are shown over about 30,000 runs for each ε (20 repetitions over 1491 histograms).

We evaluate both the naive and the unbiased estimator on the NIPS Bag of Words dataset [Newman, 2008]. The technique based on unbiasing the continuous Laplace Calmon et al. [2025] distribution was excluded from this experiment as the entropy is not differentiable when one of the input is zero. To make the entropy function defined on $\mathbb{Z}^n$, as required for our estimator we replace the term $\frac{x_i}{s} \log \frac{s}{x_i}$ by zero when x_i is not positive. This dataset consists of a collection of 1,500 research papers available as a count histogram of words on a dictionary of 12,419 words. We excluded 9 documents that contained at most 5 words. The results are presented in Figure 5. We can see that the unbiased method always outperforms the naive estimator, with the greatest improvement being for the low ε regime.

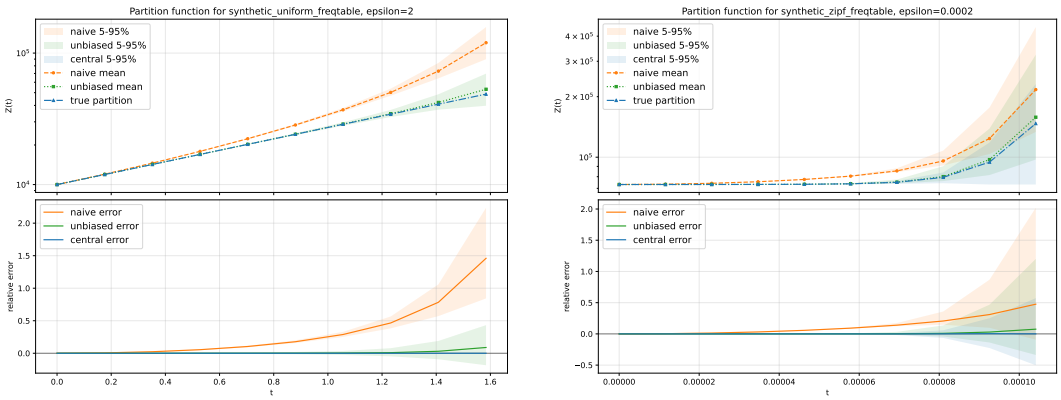
E.3 Estimation of Partition Functions

We evaluate the estimators on the task of computing the partition function $Z(t) = \sum_i \exp(tx_i)$ from a privatized histogram $\tilde{x} = x + \texttt{DLap}(e^{-\varepsilon})$. Partition functions are ubiquitous in machine learning: they normalize exponential-family models, energy-based models, and Gibbs distributions, and their logarithms encode quantities such as likelihoods, free energies, and cumulant-generating functions. Estimating $Z(t)$ from a private histogram therefore provides a simple test case for nonlinear post-processing of privatized data, while retaining the essential difficulty that exponentiation amplifies the bias introduced by additive noise.

To make estimation feasible we limit t to an interval in which the naive estimator $\tilde{Z}(t) = \sum_i \exp(t\tilde{x}_i)$ has a well-defined expectation. Note that even for those values of t, the partition functions are still not tempered distributions which makes them fall outsite of the application scope of the techniques of Calmon et al. [2025]. We compare $\tilde{Z}(t)$ to the unbiased estimator derived in Section 2, choosing parameter ε small enough that the former has a visible bias. To see how close the unbiased estimator is to what we might hope for, we also compare to a central model baseline that has access to the true histogram x and needs to output *a single value* of the partition function.

Datasets and results. Experiments are conducted on both synthetic and real datasets, varying the parameter t and observing the distribution of estimates over 200 noise realizations. The datasets cover two synthetic frequency tables and two real histograms. The synthetic uniform table has approximately equal frequencies across the domain, while the synthetic Zipf table has a heavy-tailed frequency profile. The ExAC histogram records low allele-count genetic variants, restricted to non-TCGA variants with allele count at most 10 [Lek, Karczewski, Minikel, et al., 2016]. The Shakespeare dataset is an expanded word-frequency profile derived from Shakespeare text, giving a real heavy-tailed distribution with a much larger dynamic range [Friendly and Klorfine, 2026].

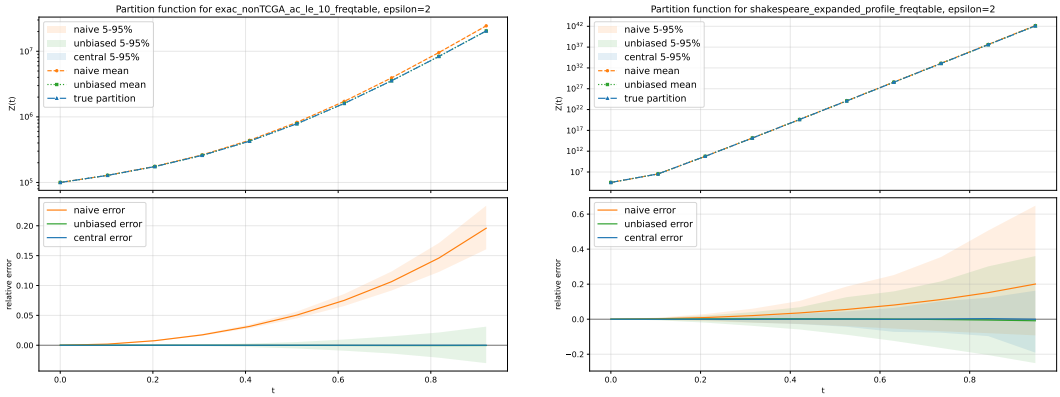
Figure 6a (synthetic uniform data) shows that the naive estimator exhibits a clear positive bias that grows rapidly with t, leading to large relative error for moderate values of t. In contrast, the unbiased estimator closely tracks the true partition function across the entire range, with substantially reduced bias. The variance, visualized by the 5–95% bands, is moderately larger for the unbiased estimator but



(a) Synthetic uniform data.

(b) Synthetic Zipf data.

Figure 6: Partition function estimation on synthetic datasets.



(a) ExAC data.

(b) Shakespeare data.

Figure 7: Partition function estimation on real datasets.

remains well-controlled. This demonstrates that for smooth distributions, debiasing can effectively remove systematic error without introducing prohibitive variance.

For heavy-tailed data (synthetic Zipf, Figure 6b), the qualitative picture is similar but more pronounced. The naive estimator significantly overestimates $Z(t)$ even for small t, while the unbiased estimator remains centered around the true value. However, the spread of the unbiased estimator increases for larger t, reflecting the sensitivity of the exponential function to noise in high-magnitude coordinates. This highlights the expected bias–variance trade-off: removing bias incurs increased variance, especially for skewed distributions.

Results on real datasets further confirm these trends. On the ExAC histogram (Figure 7a), the unbiased estimator achieves near-zero relative error across all tested values of t, while the naive estimator accumulates a steadily increasing bias. Similarly, on the Shakespeare dataset (Figure 7b), which has large dynamic range, the unbiased estimator remains accurate even when $Z(t)$ spans many orders of magnitude, whereas the naive estimator again shows systematic overestimation.

Overall, the experiments validate the theoretical claims: the unbiased estimator successfully eliminates bias for nonlinear functions, and although it may incur higher variance, its mean accuracy is consistently superior across distributions. In regimes where multiple independent estimates are aggregated, as for the partition function, this unbiasedness is particularly important, as it prevents the linear accumulation of error inherent to biased estimators.

E.4 Profile Estimation

We next evaluate the estimators on profile estimation. Given a histogram x, the profile records, for each frequency k, the fraction of domain elements with count exactly k. Equivalently, for each k we estimate

$$\phi_k(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}[x_i = k].$$

This is a natural discrete analogue of estimating the shape of a distribution from privatized counts, and it is also a basic primitive in distribution testing, species estimation, and frequency-statistics tasks. Since each coordinate contribution is an indicator function, profile estimation is a direct test of the local debiasing formula for nonsmooth functions where the techniques of Calmon et al. [2025] cannot be used.

Private profile estimation has been studied theoretically in several settings, including the sketch post-processing approach of Wu and Pagh [2024] and the shuffle model anonymized histogram estimators of Ghazi, Kamath, Kumar, and Manurangsi [2022]. An unbiased estimator for this setting can easily be obtained from the threshold estimator of [Ghazi et al., 2022], so our estimator is not algorithmically novel in this setting, but we include empirical results for completeness. (Prior works on private profile estimation did not, to our knowledge, include empirical results.)

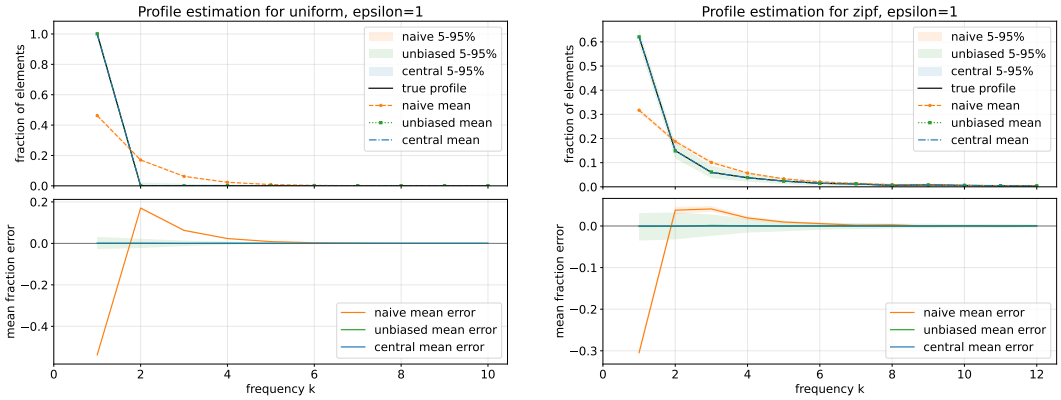
We compare the naive plug-in estimator, which computes the profile of the privatized histogram $\tilde{x} = x + \texttt{DLap}(e^{-\varepsilon})$, to the unbiased estimator obtained by applying the univariate debiasing operator to each indicator $\mathbf{1}[x_i = k]$ and then averaging over coordinates. As in the partition-function experiments, we include a central model baseline that has access to the true histogram and outputs a single profile estimate. Since the algorithms in [Wu and Pagh, 2024, Ghazi et al., 2022] are complex with unspecified implementation details we do not include them in the comparison. All experiments use $\varepsilon = 1$ and observe the distribution of estimates over 200 noise realizations.

Datasets and results. We use the same four datasets described in Section E.3. In the profile-estimation task, these datasets induce profiles with different shapes: the synthetic uniform table concentrates all mass at frequency 1, the synthetic Zipf and ExAC datasets have decreasing profiles over small frequencies, and the Shakespeare profile is more irregular and extends over a much wider range of frequencies.

Figure 8 shows the results on synthetic data. On the uniform table, the naive estimator spreads mass away from the true frequency, severely underestimating the dominant profile entry and creating artificial mass at neighboring frequencies. The unbiased estimator removes this smoothing bias and closely follows the true profile. On the Zipf table, the naive estimator again underestimates the mass at frequency 1 and overestimates several larger frequencies, whereas the unbiased estimator remains centered near the true profile.

Figure 9 shows the corresponding real-data experiments. On ExAC, the naive estimator has a large negative bias at frequency 1 and a positive bias at nearby frequencies, while the unbiased estimator is nearly indistinguishable from the true and central profiles. On Shakespeare, both estimators track the broad shape of the profile, but the naive estimator shows oscillatory bias across frequencies, whereas the unbiased estimator remains centered with smaller mean error.

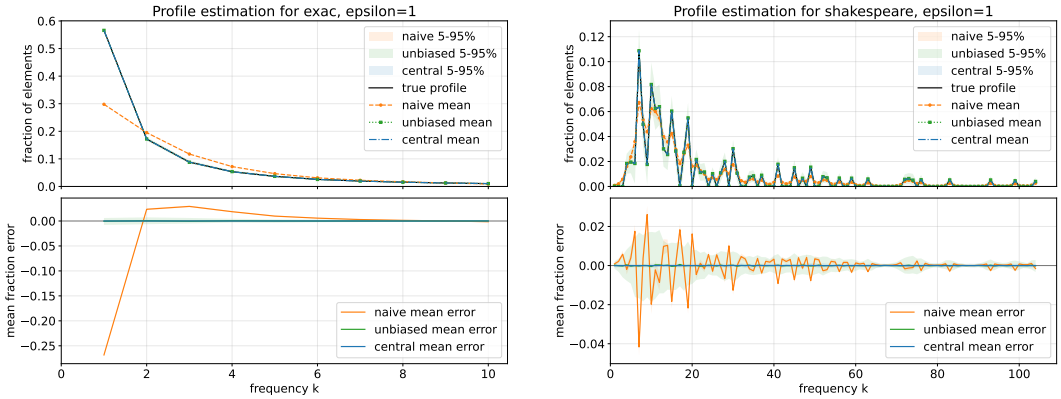
Overall, our empirical investigation of profile estimation demonstrates that our post-processing method can be effective for discontinuous statistics such as profiles. The improvement is most visible at low frequencies, where discrete Laplace noise tends to move counts between adjacent bins and therefore introduces systematic smoothing in the naive profile.



(a) Synthetic uniform data.

(b) Synthetic Zipf data.

Figure 8: Profile estimation on synthetic datasets.



(a) ExAC data.

(b) Shakespeare data.

Figure 9: Profile estimation on real datasets.

NeurIPS Paper Checklist

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 accurately describe the proposed method, its properties and limitations, and the empirical results of its application, which aligns with the content presented in the paper (Sections 2-5 and Appendix A-E). The technical overview in the introduction outlines which sections support which claims.

   Guidelines:

   - The answer [N/A] 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 [N/A] 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: Section 7 discusses the limitations of the method presented in the paper.

   Guidelines:

   - The answer [N/A] 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 states the necessary assumptions for all theorems, lemmas, and corollaries, and relevant proof sketches. Proofs are included for all novel claims in the Appendix with clear references both in the paper and in the Appendix.

Guidelines:

- The answer [N/A] 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: The paper clearly describes how to reproduce the algorithm presented. Additionally, the notebooks used for the empirical results include fixed seeds, publicly available datasets that the code ensures are downloaded, and do not use more computational power than regular laptops provide within a reasonable timeframe. As long as the datasets remain publicly available on the same URLs, the notebooks can be run to reproduce the results in the paper.

Guidelines:

- The answer [N/A] 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: The notebooks used to generate the results are self-contained, ensure all necessary dependencies (such as packages and downloaded datasets) are met, and are included in the supplementary material. No special instructions are necessary as the results can be reproduced by running said notebooks.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.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://neurips.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) necessary to understand the results?

   Answer: [Yes]

   Justification: Section 6 describes how the empirical work was obtained, and the code is provided open-sourced.

   Guidelines:

   - The answer [N/A] 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: All results in Section 6 (see Appendix Section E) include $5-95\%$ bands over several runs, clearly indicating statistical variability.

Guidelines:

- The answer [N/A] 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: The paper states that a standard laptop (CPU) can reproduce the results within a reasonable time frame.

   Guidelines:

   - The answer [N/A] 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: The research concerns algorithmic development of post-processing private estimates. It does not involve human participants and supports the development of privacy techniques, consistent with NeurIPS' code of ethics.

   Guidelines:

   - The answer [N/A] 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: [Yes]

Justification: The paper primarily focuses on algorithmic contributions but briefly discusses societal impacts in the introduction under Our Contributions.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] 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., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The paper releases or describes no data or model that poses a high risk of misuse.

Guidelines:

- The answer [N/A] 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: [Yes]

Justification: The paper properly cites the sources of existing assets (theory and datasets), and each notebook describes the license for all real-life datasets used. The latter, however, is not mentioned in the main paper or appendices. For completeness, we include the licenses here as well.

(a) **Shakespeare Words (ShakeWords)**: Provided via the VCDEXTRA R package. Distributed under the GPL-2 | GPL-3 license and is available at `https://CRAN.R-project.org/package=vcdExtra`.

(b) **ExAC (Exome Aggregation Consortium)**: The data is made available under the Database Contents License (DbCL) v1.0.

(c) **The Facebook graph**: The Dataset is part of the SNAP dataset library: J. McAuley and J. Leskovec. Learning to Discover Social Circles in Ego Networks. NIPS, 2012. It is available at the following address: `https://snap.stanford.edu/data/index.html`

(d) **Bag of Words from the NIPS dataset**: The NIPS dataset is part of the Bag of Words Dataset, David Newman (2008), licensed under Creative Commons Attribution 4.0 International (CC BY 4.0) and is available at `https://doi.org/10.24432/C5ZG6P`.

(e) **Synthetic Datasets (Zipf/Uniform)**: Generated programmatically within this notebook for research purposes.

Guidelines:

- The answer [N/A] 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: [Yes]

Justification: The paper's main contribution is theoretical, with empirical results demonstrating its practical use. All code details are structured in notebooks and provided and open-sourced under the CC BY 4.0 license. The data were either generated for this study or obtained from public datasets and used in accordance with their respective licenses. The theoretical results are clearly stated and rigorously proved in the main paper and appendix.

Guidelines:

- The answer [N/A] 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: [N/A]

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

Guidelines:

- The answer [N/A] 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: [N/A]

Justification: The research does not involve crowdsourcing nor research with human subjects, so IRB approvals are not applicable.

Guidelines:

- The answer [N/A] 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 rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: We have only used LLMs for drafting and for simple code generation (such as downloading datasets correctly), which we deem to be basic code assistance.

Guidelines:

- The answer [N/A] 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 in the NeurIPS handbook for what should or should not be described.