A Bias-Variance-Privacy Trilemma for Statistical Estimation

by

Matthew Regehr

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Computer Science

Waterloo, Ontario, Canada, 2023

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

The material is based on a submission to the Journal of the American Statistical Society and is awaiting peer review.

Abstract

The canonical algorithm for differentially private mean estimation is to first clip the samples to a bounded range and then add noise to their empirical mean. Clipping controls the sensitivity and, hence, the variance of the noise that we add for privacy. But clipping also introduces statistical bias. We prove that this tradeoff is inherent: no algorithm can simultaneously have low bias, low variance, and low privacy loss for arbitrary distributions.

On the positive side, we show that unbiased mean estimation is possible under approximate differential privacy if we assume that the distribution is symmetric. Relaxing to approximate differential privacy is necessary. We show that, even when the data is sampled from a Gaussian, unbiased mean estimation is impossible under pure or concentrated differential privacy.

Acknowledgements

First, I would like to sincerely thank my advisors Gautam Kamath and Shai Ben-David for their invaluable encouragement and feedback. Gautam patiently introduced me to the world of differential privacy and as well as the problem studied in this thesis. I benefitted greatly from Shai's eagerness to discuss new problems and ideas as well as his excellent courses on the theory of machine learning. I would also like to thank my collaborators Thomas, Jon, Vikrant, and Argyris. It is such a pleasure working with all of you.

Thank you to my friends. I owe my sanity to board game night. Thank you also to my family for your love and support.

Finally, I am grateful to my committee members Shoja'eddin Chenouri and Yaoliang Yu for reviewing my thesis and providing constructive feedback.

Dedication

This is dedicated to my partner Erica.

Table of Contents

List of Algorithms

Chapter 1

Introduction

While the goal of statistical inference and machine learning is to learn about a population, most statistical and learning algorithms reveal lot of information that is specific to their sample, raising concerns about the *privacy* of the individuals who contribute their data. In response, *differential privacy* (DP) (Dwork et al. 2006) has emerged as the standard framework for addressing these privacy concerns. Informally, a differentially private algorithm guarantees that no attacker can infer much more about any one individual in the sample than they could have inferred in a hypothetical world where that person's data was never collected. There is a rich literature providing differentially private algorithms for various statistical inference and machine learning tasks, and many of these are now deployed.

Adding the constraint of differential privacy to a statistical inference or machine learning task can, and often does, incur an inherent cost (Bun et al. 2014, Dwork, Smith, Steinke, Ullman & Vadhan 2015, Karwa & Vadhan 2018, Kamath, Li, Singhal & Ullman 2019), and there has been a large body of work pinning down these costs for a variety of tasks. The costs are typically studied via the two-way tradeoff between privacy and error, as measured by some loss function. However, in many applications, we have multiple desiderata for the estimator, not all of which can be captured by a single loss function.

In this work, we study the *statistical bias* of differentially private mean estimators, which adds an extra dimension to the tradeoff betweeen privacy and error. More precisely, given n independent samples $X_1, \ldots, X_n \in \mathbb{R}$ from an unknown univariate distribution P, we estimate the mean $\mu(P) \coloneqq \mathbb{E}_{X \leftarrow P}[X]$, subject to the constraint that the estimator $\hat{\mu}$ be differentially private. Without a privacy constraint, the empirical mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is both unbiased and provides optimal error bounds for all the settings we consider. Research on private mean estimation has also pinned down the optimal mean squared error

(MSE) $\mathbb{E}[(\hat{\mu}(X) - \mu(P))^2]$ for a variety of families of distributions, such as sub-Gaussian distributions (Karwa & Vadhan 2018, Bun & Steinke 2019) and distributions satisfying bounded moment conditions (Barber & Duchi 2014, Kamath et al. 2020). Unfortunately, these estimators can be very biased. Estimators with little or no bias are desirable because error due to variance can be averaged out by combining multiple estimates, whereas error due to bias can be difficult to eliminate.

1.1 Differential Privacy

A *dataset* $x = (x_1, \ldots, x_n) \in \mathcal{X}^n$ is a sequence of elements from a *data universe* $\mathcal{X}$. Two datasets $x, x' \in \mathcal{X}^n$ are *neighboring* (denoted $x \sim x'$) if they differ in at most one element.

Definition 1.1.1 (Differential Privacy (DP) (Dwork et al. 2006))**.** For $\varepsilon, \delta \geq 0$, we say that a randomized algorithm $A : \mathcal{X}^n \to \mathcal{Y}$ satisfies (ε, δ)-*differential privacy* ((ε, δ)-DP) when, for every neighboring pair of datasets $x \sim x' \in \mathcal{X}^n$ and every measurable $Y \subseteq \mathcal{Y}$,

$$\mathbb{P}[A(x) \in Y] \leq e^{\varepsilon}\mathbb{P}[A(x') \in Y] + \delta.$$

This property is called *pure DP* (or ε-DP) when $\delta = 0$, and *approximate DP* when $\delta > 0$.

1.2 Overview of Results

Our main contribution is to show that privacy inherently leads to statistical bias, by establishing a *trilemma* between bias, variance, and privacy for the fundamental task of *mean estimation*. Estimating the mean of a distribution is both a ubiquitous task on its own, and a subroutine in algorithms for more sophisticated tasks such as optimization. We also identify *asymmetry* as the primary cause of bias in private mean estimation by constructing unbiased private estimators for symmetric distributions.

There are a variety of methods for private mean estimation, all of which introduce bias. To understand the source of bias, it is useful to review one common approach – the *noisy clipped mean* $M(X)$, which we define as follows. First, it clips the samples to some bounded range $[a, b]$, defined by

$$\mathsf{clip}_{[a,b]}(x) \coloneqq \min\{\max\{x, a\}, b\}.$$

Next, it computes the empirical mean of the clipped samples $\hat{\mu}_{[a,b]}(X) \coloneqq \frac{1}{n}\sum_{i=1}^{n} \mathsf{clip}_{[a,b]}(X_i)$. Finally, it perturbs the clipped mean with random noise whose variance is calibrated to

the *sensitivity* of the clipped mean – i.e., the width of the clipping interval. Specifically, to ensure ε-differential privacy (ε-DP), we have

$$M(X) := \frac{1}{n}\sum_{i=1}^{n} \mathsf{clip}_{[a,b]}(X_i) + \mathsf{Lap}\left(\frac{b-a}{\varepsilon n}\right),$$

where Lap denotes the Laplace distribution, which has mean 0 and variance $\frac{2(b-a)^2}{\varepsilon^2 n^2}$.

Since the Laplace distribution has mean 0, we have $\mathbb{E}[M(X)] = \mathbb{E}[\mathsf{clip}_{[a,b]}(X_i)]$, so the only step that can introduce bias is the clipping. If we choose a large enough interval so that the support of the distribution P is contained in $[a, b]$, then clipping has no effect, and the estimator is unbiased. However, in this case, $[a, b]$ might have to be very wide, resulting in a large variance. On the other hand, if we reduce the variance by choosing a small interval $[a, b]$, then we will have $\mathbb{E}[\mathsf{clip}_{[a,b]}(X_i)] \neq \mathbb{E}[X_i]$ and the estimator will be biased. Thus, we are faced with a non-trivial bias-variance-privacy tradeoff. The exact form of the bias and the variance depends on what assumptions we make about P. In particular, if we consider the class of distributions P with bounded variance $\mathbb{E}_{X\leftarrow P}[(X - \mu(P))^2] \leq 1$, then for any $\beta > 0$, one can instantiate the noisy-clipped-mean estimator M with an appropriate interval so that it satisfies ε-DP, has bias at most β, and has MSE

$$\mathbb{E}\big[(M(X) - \mu(P))^2\big] \leq O\left(\frac{1}{n} + \beta^2 + \frac{1}{n^2 \cdot \varepsilon^2 \cdot \beta^2}\right). \tag{1.1}$$

We show that no private estimator with bias bounded by β can achieve a smaller MSE.

We do so by proving an optimal lower bound on the MSE of any differentially private estimator for the mean of an arbitrary bounded-variance distribution.

Theorem 1.2.1 (Bias-Variance-Privacy Trilemma)**.** *Let $M : \mathbb{R}^n \to \mathbb{R}$ be an (ε, δ)-DP algorithm, for some ε, δ satisfying $0 < \delta \leq \varepsilon^2/200 \leq 1$. Suppose M satisfies the following bounds on its bias β and MSE α^2: for every distribution*[1] *P with $\mathop{\mathbb{E}}_{X\leftarrow P}[X] = \mu \in [0, 1]$*[2] *and $\mathbb{E}_{X\leftarrow P}[(X - \mu)^2] \leq 1$,*

$$\left|\mathop{\mathbb{E}}_{X\leftarrow P^n, M}[M(X) - \mu]\right| \leq \beta \leq \frac{1}{100} \quad \text{and} \quad \mathop{\mathbb{E}}_{X\leftarrow P^n, M}\big[(M(X) - \mu)^2\big] \leq \alpha^2.$$

Then

$$\alpha^2 \geq \Omega\left(\min\left\{\frac{1}{n^2 \cdot \varepsilon^2 \cdot \beta^2}\ ,\ \frac{1}{n^2 \cdot \varepsilon \cdot \delta^{1/2}}\right\}\right). \tag{1.2}$$

[1] We write $X \leftarrow P^n$ to mean that X is a sequence of n i.i.d. samples from a distribution P.

[2] Since this theorem proves a lower bound, restricting the mean only *strengthens* the result. In particular, even if our estimator is provided a coarse estimate of the mean, we still face the same bias-variance-privacy tradeoff. It is common to consider coarse and fine private mean estimation separately.

To interpret the lower bound in Theorem 1.2.1 and compare it to the upper bound (1.1), it helps to start by assuming δ is small – specifically, $\delta \ll \beta^4\varepsilon^2$ – so that the first term in the minimum dominates and the bound simplifies to $\alpha^2 \geq \Omega(1/n^2\varepsilon^2\beta^2)$. This is the most interesting case, because (ε, δ)-DP is only a meaningful privacy constraint when δ is quite small (see, e.g., Kasiviswanathan & Smith (2014)). Observe that the upper bound (1.1) has two other terms, which are not reflected in Theorem 1.2.1's lower bound (1.2). These terms are also inherent, but for reasons unrelated to the privacy constraint. First, we also know that $\alpha^2 \geq \Omega(1/n)$, which is a lower bound on the MSE of any mean estimator, even those that are not private (such as the unperturbed empirical mean). Second, by the standard bias-variance decomposition of MSE, we have that $\alpha(P)^2 \geq \beta(P)^2$ for each distribution P, where $\beta(P)$ and $\alpha(P)^2$ denote, respectively, the bias and MSE of the estimator on that distribution P. Thus, if we set[3] $\beta = \sup_P \beta(P)$ and combine the three lower bounds, we conclude that

$$\alpha^2 \geq \Omega\left(\frac{1}{n} + \beta^2 + \frac{1}{n^2 \cdot \varepsilon^2 \cdot \beta^2}\right), \tag{1.3}$$

which matches the upper bound (1.1) up to constant factors.

However, there is also a corner case when $\delta \gg \beta^4\varepsilon^2$, where the second term in the minimum of Theorem 1.2.1 dominates. For small enough β, this privacy guarantee is still meaningful. Thus, for completeness, we address this corner case by constructing an estimator that nearly matches the lower bound of Theorem 1.2.1 in most parameter regimes.

Theorem 1.2.2 (Tightness of Bias-Variance-Privacy Trilemma)**.** *For all $\varepsilon, \delta, \beta > 0$ and $n \in \mathbb{N}$, there exists an (ε, δ)-DP algorithm $M : \mathbb{R}^n \to \mathbb{R}$ satisfying the following bias and accuracy properties. For any distribution P on $\mathbb{R}$ with $\mathop{\mathbb{E}}_{X\leftarrow P}[X] = \mu \in [0,1]$, $\mathop{\mathbb{E}}_{X\leftarrow P}[(X-\mu)^2] \leq 1$, and $\mathop{\mathbb{E}}_{X\leftarrow P}[(X-\mu)^4] \leq \psi^4$, we have $\left|\mathop{\mathbb{E}}_{X\leftarrow P^n, M}[M(X) - \mu]\right| \leq \beta$ and*

$$\mathop{\mathbb{E}}_{X\leftarrow P^n, M}\left[(M(X)-\mu)^2\right] = O\left(\frac{1}{n} + \min\left\{\frac{1}{n^2\cdot\varepsilon^2\cdot\beta^2} + \beta^2, \frac{\psi^2}{n^{3/2}\cdot\varepsilon\cdot\delta^{1/2}} + \frac{1}{n^2\cdot\varepsilon^2}, \frac{1}{n\cdot\delta}\right\}\right).$$

The lower bound in Theorem 1.2.1 applies to private mean estimators that are accurate for the entire class of distributions with bounded variance. *Can we obtain unbiased private estimators by making stronger assumptions on the distribution?* Our subsequent results show that the answer depends on what assumptions we are willing to make. Namely, we show that a generalization of the lower bound in Theorem 1.2.1 holds even for the case of

[3]Theorem 1.2.1 permits us to set $\beta \gg \sup_P \beta(P)$. Thus, we cannot conclude $\alpha^2 \geq \beta^2$ in the theorem.

distributions with bounded higher moments, but we also show that unbiased private mean estimation is possible for symmetric distributions.

Generalization to Higher Moment Bounds. If the distribution P is supported on a bounded interval, then unbiased private mean estimation is possible, as clipping to this support becomes an identity operation. More generally, if P is more tightly concentrated, then we can clip more aggressively and obtain better bias-variance-privacy tradeoffs for the noisy clipped mean.

We consider the class of distributions that satisfy the stronger assumption $\mathbb{E}_{X\leftarrow P}[|X-\mu|^\lambda] \leq 1$ for some $\lambda > 2$. For bias β we can achieve MSE

$$\mathbb{E}\big[(M(X)-\mu(P))^2\big] \leq O\left(\frac{1}{n} + \beta^2 + \frac{1}{n^2\cdot\varepsilon^2\cdot\beta^{2/(\lambda-1)}}\right).$$

Note that, although we can achieve a lower MSE for the same bias, this tradeoff is still qualitatively similar to the case of Theorem 1.2.1 in that estimators with optimal MSE must have large bias. We prove an analogue of Theorem 1.2.1 showing that this tradeoff is tight for this class of distributions, for every $\lambda > 2$, and conclude that bias remains an essential feature of private estimation even under stronger concentration assumptions.

Symmetric Distributions. The reason the noisy clipped mean method leads to bias is because clipping to the interval $[a, b]$ might affect the distribution asymmetrically, introducing bias. Thus, it is natural to consider whether we can achieve unbiased private estimation when the distribution is *symmetric* around its mean, which holds for many families of distributions like Gaussians. If the distribution is symmetric and we could clip to an interval $[a, b] = [\mu - c, \mu + c]$, then the clipped mean would be unbiased, but this would require us to already know μ. Nonetheless, we construct a private, unbiased mean estimator for any symmetric distribution.

Theorem 1.2.3 (Unbiased Private Mean Estimation for Symmetric Distributions)**.** *For all $\varepsilon, \delta > 0$, $\lambda > 2$, and $n \geq O(\log(1/\delta)/\varepsilon)$ with $\delta \leq 1/n$, there exists an (ε, δ)-DP algorithm $M : \mathbb{R}^n \to \mathbb{R}$ satisfying the following: Let P be a symmetric distribution on $\mathbb{R}$ – i.e., there exists $\mu \in \mathbb{R}$ so that $X - \mu$ and $\mu - X$ are identically distributed – satisfying $\mathop{\mathbb{E}}_{X\leftarrow P}\big[|X-\mu|^\lambda\big] \leq 1$. If $X \leftarrow P^n$, then $\mathbb{E}[M(X)] = \mu$, and*

$$\mathbb{E}\big[(M(X)-\mu)^2\big] \leq O\left(\frac{1}{n} + \frac{1}{(n\cdot\varepsilon)^{2-2/\lambda}} + \frac{\delta\cdot\mu^2}{n}\right).$$

Note that the MSE in the theorem has a dependence on μ, which can be unbounded. However, if we assume that we know some r such that $|\mu| \leq r$, then we can remove this term

by setting $\delta \leq 1/O(nr^2)$. Furthermore, if the distribution P is Gaussian (or sub-Gaussian), then the central moments satisfy $\mathbb{E}_{X \leftarrow P}[|X - \mu|^\lambda] \leq O((\log \lambda)^{\lambda/2})$ for all λ. In particular, for the special case of Gaussians with bounded mean, we can set $\lambda = \Theta(\log n)$ and the guarantee of our algorithm simplifies to

$$\mu^2 \leq 1/\delta \quad \Longrightarrow \quad \mathop{\mathbb{E}}_{X \leftarrow \mathcal{N}(\mu,1)^n, M}\left[(M(X) - \mu)^2\right] \leq O\left(\frac{1}{n} + \frac{\log\log n}{n^2 \cdot \varepsilon^2}\right).$$

This matches what is possible without the unbiasedness constraint (Karwa & Vadhan 2018, Kamath, Li, Singhal & Ullman 2019, Kamath et al. 2020).

We note that, unlike the noisy clipped mean method, and many other methods for private mean estimation, the estimator of Theorem 1.2.3 only satisfies (ε, δ)-DP for $\delta > 0$. This is fundamental to the techniques we use; our estimator cannot be made to satisfy $(\varepsilon, 0)$-DP. We show that this is inherent by proving that every unbiased mean estimator even for restricted classes of distributions like Gaussians cannot satisfy $(\varepsilon, 0)$-DP.

Theorem 1.2.4 (Impossibilty of Unbiased Estimators under Pure DP)**.** *Let $M : \mathbb{R}^n \to \mathbb{R}$ be a randomized algorithm. Assume that M satisfies the following guarantee: there is a nonempty open interval (a, b) such that for every $\mu \in (a, b)$,*

$$\mathop{\mathbb{E}}_{X \leftarrow \mathcal{N}(\mu,1)^n, M}[M(X)] = \mu \quad \textit{and} \quad \mathop{\mathbb{E}}_{X \leftarrow \mathcal{N}(\mu,1)^n, M}[|M(X) - \mu|] < \infty.$$

Then M does not satisfy $(\varepsilon, 0)$-DP for any $\varepsilon < \infty$.

We also extend this impossibility result beyond Gaussians to exponential families and from pure DP to concentrated DP (Dwork & Rothblum 2016, Bun & Steinke 2016).

1.3 Related Work

Unbiased estimators have long been a topic of interest in statistics. For example, topics such as the minimum variance unbiased estimator (MVUE) and the best linear unbiased estimator (BLUE) are textbook. A number of celebrated results derive properties of estimators with low or no bias, often proving certain estimators are optimal within this class. Some examples include the Gauss-Markov theorem (Gauss 1823, Markov 1900), the Lehman-Scheffé theorem (Lehmann & Scheffé 1950), and the Cramèr-Rao bound (Rao 1945, Cramér 1946). These results often focus on unbiased estimators for mathematical

convenience: it is easier to prove optimality within this restricted class than for general estimators.

Within the context of differential privacy, relatively little work has considered the bias of private estimators separately from their overall mean squared error. A number of works (Duchi et al. 2013, Barber & Duchi 2014, Karwa & Vadhan 2018, Kamath, Li, Singhal & Ullman 2019, Kamath et al. 2020) bound the bias of the clipped mean, though only to the ends of trying to minimize the overall error of the estimator. Amin et al. (2019) examine bias-variance tradeoffs of a similar procedure in the context of private empirical risk minimization. Kamath, Liu & Zhang (2022) employ the mean estimation approach of Kamath et al. (2020) as an oracle for stochastic first-order optimization, but, due to specifics of their setting, employ a different balance between bias and noise. They raise the question of whether unbiased algorithms for mean estimation exist. Barrientos et al. (2021*b*,*a*) empirically measure the bias induced by various mean estimation algorithms. Works by Zhu et al. (2021, 2022, 2023) study bias induced by a variety of differentially private algorithms. Evans & King (2021), Evans et al. (2022), Covington et al. (2021) give methods for unbiased private estimation, though these rely upon strong assumptions or caveat their unbiasedness guarantees (e.g., guaranteeing a statistic is unbiased only with high probability). Ferrando et al. (2022) appeal to the parametric bootstrap to help reduce the bias introduced by data clipping in parameteric settings. Asi & Duchi (2020) study instance-specific error bounds for private mechanisms (on fixed datasets), and prove lower bounds on the error of the class of unbiased mechanisms. Concurrent and independent work by Nikolov & Tang (2023) shows that appropriate Gaussian noise addition is essentially an optimal unbiased private mechanism for mean estimation in certain cases. Our setting and theirs are different: while we focus on mean estimation with distributional moment assumptions and on an unbounded domain, they study mean estimation for arbitrary distributions (and fixed datasets) on a bounded domain.

Beyond considerations of bias, private statistical estimation has been a topic of much recent interest. Mean estimation is perhaps the most fundamental question in this space, enjoying significant attention (see, e.g., Barber & Duchi (2014), Karwa & Vadhan (2018), Bun & Steinke (2019), Kamath, Li, Singhal & Ullman (2019), Kamath et al. (2020), Wang et al. (2020), Du et al. (2020), Biswas et al. (2020), Cai et al. (2021), Brown et al. (2021), Huang et al. (2021), Liu et al. (2021, 2022), Kamath, Liu & Zhang (2022), Hopkins, Kamath & Majid (2022), Kothari et al. (2022), Tsfadia et al. (2022), Duchi et al. (2023)). Most relevant to our work are those which focus on estimation in settings with bounds on only the low-order central moments of the underlying distribution (Barber & Duchi 2014, Kamath et al. 2020, Hopkins, Kamath & Majid 2022), as the bias introduced due to clipping is more significant. Other related problems involve private covariance or density estimation (Bun

et al. 2019, Aden-Ali, Ashtiani & Kamath 2021, Kamath, Mouzakis, Singhal, Steinke & Ullman 2022, Ashtiani & Liaw 2022, Alabi et al. 2022, Hopkins, Kamath, Majid & Narayanan 2022). Beyond these settings, other works have examined statistical estimation under privacy constraints for mixtures of Gaussians (Kamath, Sheffet, Singhal & Ullman 2019, Aden-Ali, Ashtiani & Liaw 2021, Chen et al. 2023), graphical models (Zhang et al. 2020), discrete distributions (Diakonikolas et al. 2015), median estimation (Avella-Medina & Brunel 2019, Tzamos et al. 2020, Ramsay & Chenouri 2021, Ramsay et al. 2022, Ben-Eliezer et al. 2022), and more. Several recent works explore connections between private and robust estimation (Liu et al. 2021, Hopkins, Kamath & Majid 2022, Georgiev & Hopkins 2022, Liu et al. 2022, Kothari et al. 2022, Alabi et al. 2022, Hopkins, Kamath, Majid & Narayanan 2022, Chen et al. 2023) and between privacy and generalization (Hardt & Ullman 2014, Dwork, Feldman, Hardt, Pitassi, Reingold & Roth 2015, Steinke & Ullman 2015, Bassily et al. 2016, Rogers et al. 2016, Feldman & Steinke 2017). Emerging directions of interest include guaranteeing privacy when one person may contribute multiple samples (Liu et al. 2020, Levy et al. 2021, George et al. 2022), a combination of local and central DP for different users (Avent et al. 2019), or estimation with access to some public data (Bie et al. 2022). See Kamath & Ullman (2020) for more coverage of recent work on private statistical estimation.

Chapter 2

Main Bias-Variance-Privacy Trilemma

We now discuss our main negative result. Informally, we show that if an algorithm is differentially private and has low bias, then it must have high error. There are, of course, other parameters that arise in the analysis, such as bounds on the tails of the unknown distribution P.

We provide two different proofs, which give slightly different results. The first proof directly applies the fingerprinting technique for lower bounds on differentially private estimation (Bun et al. 2014), while the second proof is a "black-box" reduction.

2.1 Negative Result via Fingerprinting

Our primary approach to a lower bound leverages the fingerprinting method (alternatively called "tracing attacks" or "membership-inference attacks"), which emerged from the study of fingerprinting codes (Boneh & Shaw 1998) in the context of cryptographic traitor-tracing schemes. Tardos (2008) gave an optimal construction of fingerprinting codes. Bun et al. (2014), Dwork, Smith, Steinke, Ullman & Vadhan (2015) showed how the theory of fingerprinting codes could be used to prove optimal lower bounds on the error of differentially private estimation. Subsequently, many works have expanded this methodology (Steinke & Ullman 2015, Bun et al. 2017, Steinke & Ullman 2017*a*,*b*, Cai et al. 2020, 2021, Kamath, Mouzakis & Singhal 2022, Cai et al. 2023). Our proof of Theorem 1.2.1 is based on a

refinement of this method that separately accounts for the bias and mean squared error of the estimator, and thus allows for us to prove tradeoffs between these two parameters.

We begin by stating our general result and provide some remarks and corollaries to help interpret the result.

Theorem 2.1.1 (Bias-Variance-Privacy Tradeoff)**.** *Let $\varepsilon, \delta, \beta, \alpha, \tau \geq 0$ and $\lambda > 1$. Let $M : \mathbb{R}^n \to \mathbb{R}$ be an (ε, δ)-DP algorithm that satisfies the following bias and accuracy properties. For any distribution P on $\mathbb{R}$ with $\mu(P) \coloneqq \mathop{\mathbb{E}}_{X \leftarrow P}[X] \in [0,1]$ and $\mathop{\mathbb{E}}_{X \leftarrow P}\big[|X - \mu(P)|^\lambda\big] \leq 1$, we have the following:*

$$\left| \mathop{\mathbb{E}}_{X \leftarrow P^n, M}[M(X)] - \mu(P) \right| \leq \beta,$$
$$\mathop{\mathbb{E}}_{X \leftarrow P^n, M}[|M(X) - \mu(P)|] \leq \alpha,$$
$$\int_0^\infty \min\left\{ \delta, \mathop{\mathbb{P}}_{X \leftarrow P^n, M}[|M(X) - \mu(P)| > x] \right\} \mathrm{d}x \leq \alpha \cdot \tau.$$

If $16\beta \leq \gamma \leq 1/5$, then $\alpha \geq \left(32n \cdot \sinh(\varepsilon) \cdot \gamma^{1/(\lambda-1)} + 16n \cdot \tau \cdot \gamma^{-1}\right)^{-1}$.

Note that, for small values of ε, $\sinh(\varepsilon) \approx \varepsilon$, but, for large ε, $\sinh(\varepsilon) \approx \frac{1}{2}e^\varepsilon$. Since this is the "usual" dependence on ε in many such bounds under the constraint of DP, sinh allows us to capture behaviour in both regimes with a single function.

The first two accuracy conditions are not hard to interpret: The parameter β bounds the bias of the algorithm, while α bounds the mean absolute deviation. By Jensen's inequality,

$$\left| \mathop{\mathbb{E}}_{X \leftarrow P^n, M}[M(X)] - \mu(P) \right| \leq \mathop{\mathbb{E}}_{X \leftarrow P^n, M}[|M(X) - \mu(P)|] \leq \sqrt{\mathop{\mathbb{E}}_{X \leftarrow P^n, M}\big[(M(X) - \mu(P))^2\big]}.$$

Thus, we can assume that $\beta \leq \alpha$. Furthermore, if an estimator $M(X)$ has a bound on the mean squared error of α^2, it consequently also has a mean absolute error of at most α. Thus, our somewhat unconventional assumption controlling the mean absolute error only broadens the class of estimators against which our lower bound holds: for interpretability, one could instead replace this with α^2 being the mean squared error of $M(X)$.

The third property and the parameter τ is somewhat harder to interpret. We note that this condition is implied by a bound on the MSE of the estimator via the following lemma. It applies in more general circumstances as well, when we may have bounds on higher or lower moments of the estimator's error. See Appendix A.1 for the proof.

Lemma 2.1.2 (Setting $\tau = \delta^{1-1/\kappa}$ in Theorem 2.1.1)**.** *Let $\alpha, \delta \geq 0$ and $\kappa > 1$. Let Y be a random variable satisfying $\mathbb{E}[|Y|^\kappa] \leq \alpha^\kappa$. Then $\int_0^\infty \min\{\delta, \mathbb{P}[|Y| > x]\}\, \mathrm{d}x \leq \alpha \cdot \delta^{1-1/\kappa}$.*

In particular, if we have a mean squared error bound for the estimator $\mathbb{E}[(M(X) - \mu(P))^2] \leq \alpha^2$, then the third condition of Theorem 2.1.1 holds with $\tau = \sqrt{\delta}$. Larger values of κ entail sharper tail bounds on the estimator, allowing us to set τ larger (and thus implying stronger lower bounds), with $\tau \to \delta$ as $\kappa \to \infty$.

In general, Theorem 2.1.1's lower bound on the error α is maximized by setting

$$\gamma = \mathsf{clip}_{[16\beta, 1/5]}\left(\left(\frac{(\lambda-1)\tau}{2\sinh(\varepsilon)}\right)^{1-1/\lambda}\right). \tag{2.1}$$

Combining this parameter setting for γ, along with the bound of $\tau = \delta^{1-1/\kappa}$ given by Lemma 2.1.2, and focusing on the most natural case of $\kappa = 2$ (i.e., we assume only that the estimator has bounded variance), gives the following result.

Corollary 2.1.3 (Combining Theorem 2.1.1, Lemma 2.1.2 (with $\kappa = 2$), and Equation 2.1.)**.** *Let $M : \mathbb{R}^n \to \mathbb{R}$ be (ε, δ)-DP and satisfy the following bias and accuracy properties. For any distribution P on $\mathbb{R}$ with $\mu(P) \coloneqq \mathop{\mathbb{E}}_{X \leftarrow P}[X] \in [0,1]$ and $\mathop{\mathbb{E}}_{X \leftarrow P}\left[|X - \mu(P)|^\lambda\right] \leq 1$, we have*

$$\left|\mathop{\mathbb{E}}_{X \leftarrow P^n, M}[M(X)] - \mu(P)\right| \leq \beta \quad \text{and} \quad \mathop{\mathbb{E}}_{X \leftarrow P^n, M}\left[(M(X) - \mu(P))^2\right] \leq \alpha^2.$$

If $\beta \leq \frac{1}{80}$ and $\delta \leq \left(\frac{2\cdot\sinh(\varepsilon)}{5^{1+\frac{1}{\lambda-1}}\cdot(\lambda-1)}\right)^2$, then

$$\alpha \geq \left(32 \cdot n \cdot \sinh(\varepsilon) \cdot \frac{\lambda}{\lambda-1} \cdot \max\left\{(16\beta)^{\frac{1}{\lambda-1}}, \left(\frac{(\lambda-1)\sqrt{\delta}}{2\cdot\sinh(\varepsilon)}\right)^{1/\lambda}\right\}\right)^{-1}.$$

We illustrate the representative case where the underlying distribution has bounded variance by further fixing $\lambda = 2$. Combining the resulting lower bound with the non-private rate (Proposition A.2.1) gives the following result.

Theorem 2.1.4 (Setting $\lambda = 2$ in Corollary 2.1.3 to get Theorem 1.2.1)**.** *Let $M : \mathbb{R}^n \to \mathbb{R}$ be (ε, δ)-DP and satisfy the following bias and accuracy properties. For any distribution P on $\mathbb{R}$ with $\mu(P) \coloneqq \mathop{\mathbb{E}}_{X \leftarrow P}[X] \in [0,1]$ and $\mathop{\mathbb{E}}_{X \leftarrow P}[(X - \mu(P))^2] \leq 1$, we have*

$$\left|\mathop{\mathbb{E}}_{X \leftarrow P^n, M}[M(X)] - \mu(P)\right| \leq \beta \quad \text{and} \quad \mathop{\mathbb{E}}_{X \leftarrow P^n, M}\left[(M(X) - \mu(P))^2\right] \leq \alpha^2.$$

If $\beta \le 1/80$ *and* $\delta \le \left(\frac{2}{25}\sinh(\varepsilon)\right)^2$*, then*

$$\alpha \ge \max\left\{ \frac{1}{\sqrt{6(n+2)}}, \frac{1}{64 \cdot n \cdot \sinh(\varepsilon) \cdot \max\left\{16 \cdot \beta, \sqrt{\frac{\sqrt{\delta}}{2 \cdot \sinh(\varepsilon)}}\right\}} \right\} \ge \Omega\left(\frac{1}{\sqrt{n}} + \min\left\{ \frac{1}{n\varepsilon\beta}, \frac{1}{n\sqrt{\varepsilon\sqrt{\delta}}} \right\} \right).$$

A full proof can be found in Appendix A.3. In any case, to give intuition for the argument, we consider the case where M is an *unbiased* estimator, in which case the argument is similar to the proof of the Cramér-Rao bound. Assume that we have a suitable family of distributions P_μ with mean $\mathbb{E}_{X \leftarrow P_\mu}[X] = \mu$ and an unbiased estimator M such that $\mathbb{E}_{X \leftarrow P_\mu^n}[M(X)] = \mu$. As in the proof of the Cramér-Rao bound, we take the derivative of the unbiasedness constraint, which gives

$$1 = \frac{\mathrm{d}}{\mathrm{d}\mu}\left[\mathop{\mathbb{E}}_{X \leftarrow P_\mu^n, M}[M(X)]\right] = \sum_{i=1}^{n} \mathop{\mathbb{E}}_{X \leftarrow P_\mu^n, M}\left[M(X) \cdot \frac{\mathrm{d}}{\mathrm{d}\mu} \log P_\mu(X_i)\right],$$

where $P_\mu(x)$ denotes the probability mass or density function of P_μ evaluated at x. The (ε, δ)-differential privacy guarantee of M says that $M(X)$ and X_i are close to being independent where ε and δ quantify the distance from independence. Moreover, a straightforward calculation shows that $\mathbb{E}_{X \leftarrow P_\mu}[\frac{\mathrm{d}}{\mathrm{d}\mu} \log P_\mu(X)] = 0$. Thus, for all $i \in [n]$, we have

$$\mathop{\mathbb{E}}_{X \leftarrow P_\mu^n, M}\left[M(X) \cdot \frac{\mathrm{d}}{\mathrm{d}\mu} \log P_\mu(X_i)\right] \approx_{\varepsilon,\delta} \mathop{\mathbb{E}}_{X \leftarrow P_\mu^n, M}[M(X)] \cdot \mathop{\mathbb{E}}_{X \leftarrow P_\mu^n}\left[\frac{\mathrm{d}}{\mathrm{d}\mu} \log P_\mu(X_i)\right] = 0.$$

Intuitively, this leads to the contradiction

$$1 = \sum_{i=1}^{n} \mathop{\mathbb{E}}_{X \leftarrow P_\mu^n, M}\left[M(X) \cdot \frac{\mathrm{d}}{\mathrm{d}\mu} \log P_\mu(X_i)\right] \approx_{\varepsilon,\delta} \sum_{i=1}^{n} 0.$$

To make this argument precise, we must exactly quantify the approximation $\approx_{\varepsilon,\delta}$, which depends both on the privacy parameters ε and δ, as well as on the variances of $M(X)$ and of $\frac{\mathrm{d}}{\mathrm{d}\mu} \log P_\mu(X_i)$. The variance of $M(X)$ is the quantity that we are trying to bound. The variance of $\frac{\mathrm{d}}{\mathrm{d}\mu} \log P_\mu(X_i)$ (which is known as the Fisher information) is something we control by choosing the distribution P_μ to be a distribution supported on two points.

The above proof sketch applies to the unbiased case ($\beta = 0$). The general case ($\beta > 0$) introduces some additional complications to the proof. In particular, we cannot simply consider a single fixed value of the mean parameter μ, as we must rule out the pathological algorithm that ignores its input sample and outputs μ, which has somehow been hardcoded

into the algorithm. This pathological algorithm trivially satisfies privacy and is unbiased for the single distribution P_μ. To rule out this algorithm, we consider a distribution over the parameter μ and average over this distribution, where the distribution's support is wider than the allowable bias β. While we can no longer assume that $1 = \frac{\mathrm{d}}{\mathrm{d}\mu}\left[\mathop{\mathbb{E}}_{X\leftarrow P_\mu^n, M}[M(X)]\right]$, we can still argue that the derivative must be $\geq \Omega(1)$ *on average* over the choice of μ.

2.2 Negative Result via Amplification

In this section, we show that known MSE lower bounds (without bias constraints) (Kamath et al. 2020) combined with privacy amplification via shuffling (Erlingsson et al. 2019, Cheu et al. 2019, Balle et al. 2019, Feldman et al. 2022, 2023) can also be used to derive qualitatively similar lower bounds on MSE for private estimators with low bias as those yielded by fingerprinting in the previous section. Our reduction provides an alternative perspective on the bias-variance-privacy tradeoff, and could prove useful in future work as it is more "generic" than the fingerprinting approach. Specifically, we will use the following lower bound on the MSE of a private estimator in a black-box manner.

Theorem 2.2.1 ((Kamath et al. 2020, Theorem 3.8))**.** *Let $M : \mathbb{R}^n \to \mathbb{R}$ be (ε, δ)-DP. Then, for some distribution P with $\mu(P) \coloneqq \mathbb{E}_{X\leftarrow P}[X] \in [-1, 1]$ and $\mathbb{E}_{X\leftarrow P}[(X - \mu(P))^2] \leq 1$,*

$$\mathop{\mathbb{E}}_{X\leftarrow P^n}\left[(M(X) - \mu(P))^2\right] \geq \Omega\left(\frac{1}{n(\varepsilon + \delta)}\right).$$

The other ingredient in our proof is the following extension of the privacy amplification by subsampling result of Feldman et al. (2022). Specifically we extend from the setting of local differential privacy (where each algorithm has one input) to the setting where a dataset is randomly partitioned into blocks of fixed size $n > 1$, and these blocks are processed by a sequence of private mechanisms. A complete proof of the following result can be found in Appendix A.4.

Theorem 2.2.2 (Extension of Privacy Amplification by Shuffling (Feldman et al. 2022) to Larger Inputs)**.** *Suppose we have a randomized function $L_i : \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_{i-1} \times \mathcal{X}^n \to \mathcal{Y}_i$ for each $i \in [m]$ such that $L_i(y, x)$ is $(\varepsilon_0, \delta_0)$-DP in the parameter $x \in \mathcal{X}^n$ for every fixed y. Consider $L_m \otimes \cdots \otimes L_1 : (\mathcal{X}^n)^m \to \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_m$ defined by*

$$(L_m \otimes \cdots \otimes L_1)(x_1, \ldots, x_m) \coloneqq (y_1, \ldots, y_m)$$

where we recursively define $y_i := L_i(y_1, \ldots, y_{i-1}, x_i)$. *In addition, consider the shuffle operator* $\Pi : (\mathcal{X}^n)^m \to (\mathcal{X}^n)^m$ *given by*

$$\Pi((x_1^1, \ldots, x_1^n), \ldots, (x_m^1, \ldots, x_m^n)) := ((x_{\pi_1(1)}^1, \ldots, x_{\pi_n(1)}^n), \ldots, (x_{\pi_1(m)}^1, \ldots, x_{\pi_n(m)}^n))$$

where $\pi_1, \ldots, \pi_n$ *are uniform i.i.d. permutations of* $[m]$. *Then, for any* $\delta_1 \in [2\exp(-\frac{m}{16e^{\varepsilon_0}}), 1]$, $L_m \otimes \cdots \otimes L_1 \circ \Pi$ *is* $(\varepsilon_1, \delta_1 + (e^{\varepsilon_1} + 1)(e^{-\varepsilon_0}/2 + 1)m\delta_0)$*-DP, where*

$$\varepsilon_1 := \log\left(1 + 8\frac{e^{\varepsilon_0} - 1}{e^{\varepsilon_0} + 1}\left(\sqrt{\frac{e^{\varepsilon_0}\log(4/\delta_1)}{m}} + \frac{e^{\varepsilon_0}}{m}\right)\right). \tag{2.2}$$

Note that, when $\varepsilon_0 = O(1)$, we have $\varepsilon_1 = O(\varepsilon_0\sqrt{\log(1/\delta_1)/m})$.

We use this result to prove a slightly weaker version of Theorem 1.2.1 by reduction to Theorem 2.2.1. The proof is given in Appendix A.5.

Theorem 2.2.3 (Bias-Variance-Privacy Trilemma via Shuffling)**.** *Let* $M : \mathbb{R}^n \to \mathbb{R}$ *be* (ε, δ)*-DP and satisfy the following bias and accuracy properties. For any distribution* P *over* $\mathbb{R}$ *with* $\mu(P) := \mathop{\mathbb{E}}_{X \leftarrow P}[X] \in [-1, 1]$ *and* $\mathop{\mathbb{E}}_{X \leftarrow P}[(X - \mu(P))^2] \le 1$, *we have*

$$\left|\mathop{\mathbb{E}}_{X \leftarrow P^n, M}[M(X)] - \mu(P)\right| \le \beta \quad \text{and} \quad \mathop{\mathbb{E}}_{X \leftarrow P^n, M}\left[(M(X) - \mu(P))^2\right] \le \alpha^2.$$

If $\beta^2 \le \widetilde{\Omega}(\frac{1}{n\varepsilon})$ *and* $\delta \le O(n^3\varepsilon^4\beta^6)$, *then* $\alpha^2 \ge \Omega\left(\frac{1}{n^2\varepsilon^2\beta^2\log(1/n\varepsilon^2\beta^2)}\right)$.

Chapter 3

Low-Bias Estimators for General Distributions

In this chapter, we describe and analyze algorithms for private estimation with low or no bias. We give three algorithms: an $(\varepsilon, 0)$-DP algorithm based on the clipped mean (Proposition 3.0.1), a $(0, \delta)$-DP algorithm based on a variant of the "name-and-shame" algorithm (Proposition 3.0.2), and an (ε, δ)-DP algorithm obtained by combining the two (Proposition 3.0.3). By taking the best of the three resulting bounds, we get Theorem 1.2.2.

We first have a positive result based on clipping and adding noise, which satisfies pure DP. The clipped and noised mean is folklore in differential privacy. Analyzing such a procedure with bounded moments has been done in a few works (Duchi et al. 2013, Barber & Duchi 2014, Kamath et al. 2020). These works generally set algorithm parameters to achieve a prescribed bias, towards the goal of minimizing the overall error. As our goal is to explicitly quantify the bias, we leave it as a free variable.

Proposition 3.0.1 (ε-DP Algorithm)**.** *Fix any $\varepsilon, \beta > 0$, $a < b$, $\lambda \geq 2$, and $n \in \mathbb{N}$. Consider the following ε-DP mechanism $M : \mathbb{R}^n \to \mathbb{R}$:*

$$M(x) \coloneqq \left(\frac{1}{n} \sum_{i}^{n} \mathsf{clip}_{[\hat{a},\hat{b}]}(x_i) \right) + \mathsf{Lap}\left(\frac{\hat{b} - \hat{a}}{\varepsilon n} \right),$$

where $\hat{a} \coloneqq a - \beta^{-1/(\lambda-1)}$ and $\hat{b} \coloneqq b + \beta^{-1/(\lambda-1)}$. M satisfies the following bias and accuracy properties. For any distribution P on $\mathbb{R}$ with $\mu(P) \coloneqq \mathop{\mathbb{E}}_{X \leftarrow P}[X] \in [a, b]$ and

$\mathop{\mathbb{E}}_{X\leftarrow P}\left[|X-\mu(P)|^\lambda\right] \le 1$, *we have*

$$\left|\mathop{\mathbb{E}}_{X\leftarrow P^n,M}[M(X)] - \mu(P)\right| \le \beta,$$
$$\mathop{\mathbb{E}}_{X\leftarrow P^n,M}\left[(M(X)-\mu(P))^2\right] \le \frac{1}{n} + \beta^2 + \frac{2}{\varepsilon^2 n^2}\left(b-a+\frac{2}{\beta^{1/(\lambda-1)}}\right)^2.$$

Next, we give an algorithm based on the folklore "name-and-shame" procedure, which is $(0,\delta)$-DP. The name-and-shame procedure is generally phrased as randomly selecting a point from a dataset and outputting it, sans any further privacy protection. It is most commonly used as an illustration of which values of δ may or may not be meaningful when it comes to informal uses of the word "privacy", and not as a serious algorithm. However, we note that such a procedure gives an exactly unbiased estimate of the mean, which the previous $(\varepsilon,0)$-DP was unable to do. We thus use it to design an unbiased algorithm for mean estimation, albeit at a high price in the dependence on δ, which we recall is usually chosen to be very small.

Proposition 3.0.2 ($(0,\delta)$-DP Algorithm)**.** *Fix any $\delta \in (0,1]$ and $n \in \mathbb{N}$. Consider the following $(0,\delta)$-DP algorithm $M : \mathbb{R}^n \to \mathbb{R}$: $M(x) = \frac{1}{n}\sum_i^n A(x_i)$ where where each instantiation of $A(x_i)$ is independent and $A : \mathbb{R} \to \mathbb{R}$ is the randomized algorithm*

$$A(x) \coloneqq \begin{cases} 0 & \text{with probability } 1-\delta \\ \frac{x}{\delta} & \text{with probability } \delta \end{cases}.$$

M satisfies the following bias and accuracy properties. For any distribution P on $\mathbb{R}$,

$$\mathop{\mathbb{E}}_{X\leftarrow P^n,M}[M(X)] = \mu(P), \qquad \text{(i.e., } M \text{ is unbiased)}$$
$$\mathop{\mathbb{E}}_{X\leftarrow P^n,M}\left[(M(X)-\mu(P))^2\right] = \frac{\mathop{\mathbb{E}}_{X\leftarrow P}[(X-\mu(P))^2] + (1-\delta)\cdot\mu(P)^2}{\delta\cdot n}.$$

We can combine both of these methods to obtain a new algorithm for (ε,δ)-DP mean estimation. Essentially, it decomposes a sample into the non-tail and tail components, releasing the former via the $(\varepsilon,0)$-DP clip-and-noise method, and the latter via the $(0,\delta)$-DP name-and-shame approach. Note that we must consider a higher moment in our assumption about the unknown distribution P.

Proposition 3.0.3 ((ε, δ)-DP Algorithm)**.** *Fix any $\varepsilon > 0$, $\delta \in (0, 1]$, $\psi > 0$, $\lambda > 2$, $a < b$, and $n \in \mathbb{N}$. Consider the following (ε, δ)-DP algorithm $M : \mathbb{R}^n \to \mathbb{R}$:*

$$M(x) \coloneqq \underbrace{\left(\frac{1}{n}\sum_i^n \mathsf{clip}_{[\hat{a},\hat{b}]}(x_i)\right) + \mathsf{Lap}\left(\frac{\hat{b}-\hat{a}}{n\varepsilon}\right)}_{(\varepsilon,0)\text{-DP}} + \underbrace{\left(\frac{1}{n}\sum_i^n A\Big(x_i - \mathsf{clip}_{[\hat{a},\hat{b}]}(x_i)\Big)\right)}_{(0,\delta)\text{-DP}},$$

where $c \coloneqq \left(\frac{n\varepsilon^2\psi^\lambda(\lambda-2)}{4\lambda^2\delta}\right)^{1/\lambda}$, $\hat{a} \coloneqq a - c$, and $\hat{b} \coloneqq b + c$, and the Laplace noise and all instantiations of A (defined as in Proposition 3.0.2) are independent. M satisfies the following bias and accuracy properties. For any distribution P on $\mathbb{R}$ with $\mu(P) \coloneqq \mathop{\mathbb{E}}_{X\leftarrow P}[X] \in [a, b]$ and $\mathop{\mathbb{E}}_{X\leftarrow P}[(X - \mu(P))^2] \le 1$ and $\mathop{\mathbb{E}}_{X\leftarrow P}\big[|X - \mu(P)|^\lambda\big] \le \psi^\lambda$, we have

$$\mathop{\mathbb{E}}_{X\leftarrow P^n, M}[M(X)] = \mu(P), \qquad \text{(i.e., } M \text{ is unbiased)}$$

$$\mathop{\mathbb{E}}_{X\leftarrow P^n, M}\big[(M(X) - \mu(P))^2\big] \le \frac{2}{n} + \frac{4(b-a)^2}{n^2\varepsilon^2} + \frac{24\psi^2}{n^2\varepsilon^2} \cdot \left(\frac{n\varepsilon^2}{4\lambda\delta}\right)^{2/\lambda}.$$

Combining Propositions 3.0.1, 3.0.2, and 3.0.3 yields Theorem 1.2.2.

Chapter 4

Unbiased Estimators for Symmetric Distributions

We now present our unbiased private mean estimation algorithm for *symmetric* distributions over $\mathbb{R}$ that are weakly concentrated, i.e., those that have a bounded second moment.

Definition 4.0.1 (Symmetric Distribution)**.** We say that a distribution P on $\mathbb{R}$ is *symmetric* if there exists some $\mu(P) \in \mathbb{R}$, such that $\forall x \in \mathbb{R}$, $\mathop{\mathbb{P}}_{X \leftarrow P}[X - \mu(P) \leq x] = \mathop{\mathbb{P}}_{X \leftarrow P}[\mu(P) - X \leq x]$. The value $\mu(P)$ is called the *center* of the distribution P.

Note that the center of the distribution is unique and coincides with the mean and the median (whenever these two quantities are well-defined).

Our algorithm is based on the approach of Karwa & Vadhan (2018), but with some modifications to ensure unbiasedness. First, we obtain a coarse estimate of the mean, and then we use this coarse estimate to perform clipping to obtain a precise estimate via noise addition. The key observation is that, if the coarse estimate we use for clipping is unbiased and symmetric (and also independent from the data used in the second step), then the clipping does not introduce bias. We obtain the coarse estimate via a DP histogram, where each bucket in the histogram is an interval on the real line. To ensure that this is unbiased and symmetric, we simply need to apply a random offset to the bucket intervals.

4.1 Coarse Unbiased Estimation

Our coarse estimator (Algorithm 1) is similar to that of Karwa & Vadhan (2018). The key modification to ensure unbiasedness is adding a random offset to the histogram bins. We define $\mathsf{round}_{\mathbb{Z}} : \mathbb{R} \to \mathbb{Z}$ to be the function that rounds real numbers to the nearest integer, i.e., for any $x \in \mathbb{R}$, we have $x \in [\mathsf{round}_{\mathbb{Z}}(x) - 1/2, \mathsf{round}_{\mathbb{Z}}(x) + 1/2)$.

Algorithm 1: Unbiased DP Coarse Estimator $\text{DPUCoarse}_{\varepsilon,\delta}(x)$

Input: Dataset $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$.
Output: Estimate $\widetilde{\mu} \in \mathbb{R} \cup \{\perp\}$.

Let T be uniform on the interval $[-1/2, +1/2]$.
Let $K = \{\mathsf{round}_{\mathbb{Z}}(x_i - T) : i \in [n]\} \subset \mathbb{Z}$.
For each $k \in K$, sample $\xi_k \leftarrow \mathsf{Lap}(2/\varepsilon)$ independently.
If $\max_{k \in K} |\{i \in [n] : \mathsf{round}_{\mathbb{Z}}(x_i - T) = k\}| + \xi_k \le 2 + \frac{2\log(1/\delta)}{\varepsilon}$, output $\perp$.
Otherwise, output $T + \arg\max_{k \in K} |\{i \in [n] : \mathsf{round}_{\mathbb{Z}}(x_i - T) = k\}| + \xi_k$.

First, it is easy to verify that Algorithm 1 is private. Similar to previous work (see, e.g., Vadhan (2017)), privacy follows from the privacy of the stable histogram algorithm, plus post-processing via argmax. A complete proof can be found in Appendix C.1.

Proposition 4.1.1. *Algorithm 1 (*$\text{DPUCoarse}_{\varepsilon,\delta}$*) satisfies* (ε, δ)*-DP.*

Now we turn to the utility analysis, which consists of two parts. First, conditioned on not outputting $\perp$, the estimate is symmetric and unbiased (Proposition 4.1.2). Second, we show that the probability of outputting $\perp$ is low for appropriately concentrated distributions, and that the MSE is bounded, as well (Proposition 4.1.3). To show that our estimator preserves symmetry, note that, unlike the static histogram bucket approach of prior work, the introduction of the uniformly random offset $T \in [\pm 1/2]$ in Algorithm 1 endows $\text{DPUCoarse}_{\varepsilon,\delta}$ with equivarience under translation. The random offsets also endow our estimator with equivariance under refleciton about the origin. See Appendix C.2 for the details.

Proposition 4.1.2 (Conditional Symmetry of DPUCoarse)**.** *Let P be a symmetric distribution with center $\mu(P)$. Let $X_1, \cdots, X_n \in \mathbb{R}$ be independent samples from P. Let $\widetilde{\mu} = \text{DPUCoarse}_{\varepsilon,\delta}(X_1, \cdots, X_n)$. Let Q be the distribution of $\widetilde{\mu}$ conditioned on $\widetilde{\mu} \neq \perp$. Then Q is symmetric with the same center as P – i.e., $\mu(P) = \mu(Q)$.*

The details of the MSE analysis can be found in Appendix C.3.

Proposition 4.1.3 (Accuracy of DPUCoarse)**.** *Let P be a distribution over $\mathbb{R}$ with mean $\mu(P)$ and variance $\mathop{\mathbb{E}}_{X\leftarrow P}[(X-\mu(P))^2] < 1/64$ and $\mathop{\mathbb{E}}_{X\leftarrow P}\big[|X-\mu(P)|^\lambda\big] \le \psi^\lambda$ for some $\lambda \ge 2$ and $\psi > 0$. Let $X = (X_1, \dots, X_n)$ be independent samples from P, and $\widetilde{\mu} \leftarrow \mathrm{DPUCoarse}_{\varepsilon,\delta}(X)$. If $n \ge 7 + \frac{7}{\varepsilon}\log(1/\delta)$, then*

$$\mathbb{P}[\widetilde{\mu} \neq \perp \wedge |\widetilde{\mu} - \mu(P)| \le 1] \ge 1 - e^{-n/128} - \frac{n}{2}e^{-n\varepsilon/16},$$
$$\mathbb{E}\big[\mathbb{I}[\widetilde{\mu} \neq \perp] \cdot |\widetilde{\mu} - \mu(P)|^\lambda\big] \le \frac{1}{2} + n \cdot 2^{\lambda-1} \cdot \psi^\lambda.$$

In particular, for $\gamma > 0$, to ensure $\mathbb{P}[\widetilde{\mu} \neq \perp \wedge |\widetilde{\mu} - \mu(P)| \le 1] \ge 1 - \gamma$, it suffices to set

$$n \ge \max\left\{7 + \frac{7}{\varepsilon}\log(1/\delta), 128\log(2/\gamma), \frac{16}{\varepsilon}\log(n/\gamma)\right\} = O(\log(n/\gamma\delta)/\varepsilon). \tag{4.1}$$

4.2 Final Algorithm

Now, we present our main algorithm (Algorithm 2) for unbiased mean estimation of symmetric distributions under (approximate) DP. The idea is straightforward: invoke our coarse estimator (Algorithm 1) to get a symmetric, unbiased, mildly accurate estimate of the mean privately; then apply the standard clip-average-noise technique on our dataset. The second step will not create any new bias because the clipping is performed around a symmetric, unbiased estimate that is independent of the data we are clipping and averaging, and the added noise has mean 0. There is an additional hiccup though: the coarse estimator may fail to produce an estimate. In this case, we fall back to a different algorithm that exploits $(0, \delta)$-DP and does not require a coarse estimate, as in Proposition 3.0.2.

The following privacy and utility gurantee is the more general version of Theorem 1.2.3.

Theorem 4.2.1 (Unbiased DP Estimator)**.** *Fix $\varepsilon, \delta \in (0, 1)$, $n_2 \in \mathbb{N}$, $\psi \ge 1$, and $\lambda \ge 2$. Set $\gamma = \delta^2$, $\sigma = 10$, $c = \sigma + \psi \cdot (n_2\varepsilon)^{1/\lambda}$, $n_1 = O(\log(n_1/\gamma\delta)/\varepsilon)$ (as in Proposition 4.1.3(4.1)), and $n = n_1 + n_2$. Algorithm 2 ($\mathrm{DPUMean}_{\varepsilon,\delta,c,\sigma,n_1,n_2}$) satisfies (ε, δ)-DP and the following. Let P be a symmetric distribution with center $\mu(P)$, $\mathop{\mathbb{E}}_{X\leftarrow P}[(X-\mu(P))^2] \le 1$, and $\mathop{\mathbb{E}}_{X\leftarrow P}\big[|X-\mu(P)|^\lambda\big] \le \psi^\lambda$. Let $X = (X_1, \cdots, X_n) \leftarrow P^n$ and*

Algorithm 2: Unbiased DP Estimator $\mathrm{DPUMean}_{\varepsilon,\delta,c,\sigma,n_1,n_2}(x)$

Input: Dataset $x = (x_1, \dots, x_{n_1}, x_{n_1+1}, \dots, x_{n_1+n_2}) \in \mathbb{R}^{n_1+n_2}$. Parameters: Privacy: $\varepsilon, \delta > 0$. Clipping & Scale: $c, \sigma > 0$. Dataset split: $n_1, n_2 \in \mathbb{N}$.

Output: Estimate $\widehat{\mu} \in \mathbb{R}$.

`// Get a coarse unbiased symmetric estimate of the mean privately.`

$\widetilde{\mu} \leftarrow \sigma \cdot \mathrm{DPUCoarse}_{\varepsilon,\delta}\left(\frac{x_1}{\sigma}, \dots, \frac{x_{n_1}}{\sigma}\right)$.

If $\widetilde{\mu} = \perp$ `// When the coarse estimator fails.`

Let $\xi_1, \xi_2, \cdots, \xi_{n_2} \in \{0,1\}$ be independent samples from $\mathsf{Bernoulli}(\delta)$.

Let $\widehat{\mu} = \frac{1}{n_2\delta} \sum_{i=1}^{n_2} x_{n_1+i} \cdot \xi_i$.

Else `// When the coarse estimator outputs` $\widetilde{\mu} \in \mathbb{R}$.

Let $\widehat{\mu} = \left(\frac{1}{n_2} \sum_{i=1}^{n_2} \mathsf{clip}_{[\widetilde{\mu}-c,\widetilde{\mu}+c]}(x_{n_1+i})\right) + \mathsf{Lap}\left(\frac{2c}{n_2\varepsilon}\right)$.

Return $(\widetilde{\mu}, \widehat{\mu})$.

$(\widetilde{\mu}, \widehat{\mu}) \leftarrow \mathrm{DPUMean}_{\varepsilon,\delta,c,\sigma,n_1,n_2}(X)$. *Then*

$$\begin{aligned}
\mathbb{E}[\widehat{\mu}] &= \mu(P), \\
\mathbb{E}\left[(\widehat{\mu} - \mu(P))^2\right] &\le \frac{1}{n_2} + O\left(\frac{\psi^2}{(n_2\varepsilon)^{2-2/\lambda}} + \delta \cdot \frac{\mu(P)^2}{n_2} + \delta^{2-4/\lambda} \cdot (n_1 + n_2\varepsilon)^{2/\lambda} \cdot \psi^2\right), \\
\mathbb{P}[\widetilde{\mu} \neq \perp] &\ge 1 - \gamma = 1 - \delta^2, \\
\mathbb{E}[\widehat{\mu} \mid \widetilde{\mu} \neq \perp] &= \mathbb{E}[\widetilde{\mu} \mid \widetilde{\mu} \neq \perp] = \mu(P), \\
\mathbb{E}\left[(\widehat{\mu} - \mu(P))^2 \mid \widetilde{\mu} \neq \perp\right] &\le \frac{1}{n_2} + O\left(\frac{\psi^2}{(n_2\varepsilon)^{2-2/\lambda}} + \delta^{2-4/\lambda} \cdot (n_1 + n_2\varepsilon)^{2/(\lambda-1)} \cdot \psi^2\right).
\end{aligned}$$

The proof is in Appendix C.4. In particular, we can apply Theorem 4.2.1 to (sub-)Gaussians. Using the bound $\mathbb{E}_{X \leftarrow \mathcal{N}(0,1)}\left[|X|^\lambda\right] = O(\sqrt{\log \lambda})^\lambda$ and setting $\lambda = \Theta(\log n)$ yields the following. Note that we restrict $|\mu| \le \delta^{-1/2}$ to remove the $\delta \cdot \mu^2/n$ term.

Corollary 4.2.2 (Unbiased Gaussian Mean Estimation)**.** *Let $\varepsilon \in (0,1)$, $\delta \in (0, 1/n)$, and $n \ge O(\log(1/\delta)/\varepsilon)$. Let $M = \mathrm{DPUMean}_{\varepsilon,\delta,c,\sigma,n_1,n_2}$ be as in Algorithm 2 with appropriate settings of parameters. Then, for all $\mu \in [-\delta^{-1/2}, \delta^{-1/2}]$,*

$$\mathbb{E}_{X \leftarrow \mathcal{N}(\mu,1)^n, M}[M(X)] = \mu \quad \text{and} \quad \mathbb{E}_{X \leftarrow \mathcal{N}(\mu,1)^n, M}\left[(M(X) - \mu)^2\right] \le O\left(\frac{1}{n} + \frac{\log\log n}{n^2\varepsilon^2}\right).$$

Chapter 5

Impossibility of Pure DP Unbiased Estimation

In Chapter 4, we showed that it is possible to perform unbiased mean estimation for symmetric distributions. However, this result only provides approximate DP (i.e., (ε, δ)-DP with $\delta > 0$). We now show that this is inherent. We show that unbiased estimation is impossible under pure DP (i.e., $(\varepsilon, 0)$-DP) when the data comes from an *exponential family*. Exponential families include a wide range of distributions, including Gaussians, exponential distributions, Laplace distributions with fixed mean, and Gamma distributions.

Definition 5.0.1 (Exponential Family)**.** Let $D \subseteq \mathbb{R}^n$, $h\colon D \to [0, \infty)$, and $T\colon D \to \mathbb{R}^k$. The exponential family with *carrier measure* and *sufficient statistic* h and T respectively is the collection of probability measures P_η with density $f_{T,h,\eta}(x) = h(x)e^{\eta^\top T(x) - Z(\eta)}$, where $Z(\eta) = \log\left(\int_S h(x)e^{\eta^\top T(x)}\,dx\right)$ is called the *log-partition function* of the family. The family is defined over all values $\eta \in \mathbb{R}^k$ for which $Z(\eta) < \infty$. The set of these values, which we denote U, is called the family's *range of natural parameters*.

Theorem 5.0.2 (Impossibility of Pure DP Unbiased Estimation for Exponential Families)**.** *Let $U \subseteq \mathbb{R}$ be an interval of infinite length, let $\{P_\eta : \eta \in U\}$ be an exponential family, and let $I \subseteq U$ be any interval of positive length. Then, for any $\varepsilon \geq 0$ and $n \geq 0$, there exists no $(\varepsilon, 0)$-DP algorithm $M : \mathbb{R}^n \to \mathbb{R}$ satisfying $\mathbb{E}_{X \leftarrow P_\eta^n, M}[M(X)] = \eta$ for all $\eta \in I$.*

We remark that the interval over which the algorithm is unbiased must have positive length. It is easy to construct a pathological estimator that is unbiased at a single point

$\eta_0 \in U$ but not anywhere else, e.g. by setting $M(x) = \eta_0$ for all $x \in \mathbb{R}^n$. On the other hand, the theorem implies that there can be no pure DP estimator that gives an unbiased estimate for the mean of Exponential(λ) for all $\lambda \in (0, 0.1)$. The range of natural parameters having infinite length is also essential to our analysis. We emphasize that this is a property of the distribution and not of the algorithm; that is, the algorithm does not need to "know" about U. Note that the family of distributions $\{\mathsf{Bernoulli}(p) : p \in [0, 1]\}$ is an exponential family[1] and it is possible to estimate the mean p under pure DP. In this case $U = [0, 1]$ has finite length, so we see that the assumption that U has infinite length is also necessary.

Proving this result requires tools from complex analysis and measure theory, a review of which can be found in Appendices D.1 and D.2, respectively. We first show that, for an estimator $\phi : \mathbb{R}^n \to \mathbb{R}$ and an exponential family $\{P_\eta : \eta\}$, the expected value of the estimator $\mathbb{E}_{X \leftarrow P_\eta^n}[\phi(X)]$ is an analytic function in η. We then apply the identity theorem for analytic functions to argue that if ϕ is locally unbiased, i.e., unbiased when η lies in some small set, then ϕ must also be globally unbiased, i.e., unbiased for all choices of η. On the other hand, we will argue that global unbiasedness over an infinite interval is impossible for pure DP estimators as a consequence of the strong group privacy properties of pure DP.

5.1 Locally Unbiased Estimators Are Globally Unbiased

The following result states that any estimator for the parameter of an exponential family that is locally unbiased is also globally unbiased. That is, if we have an unbiased estimator for a restricted range of parameter values (of nonzero length), then we have an unbiased estimator for the entire range. The proof can be found in Appendix D.3.

Proposition 5.1.1 (Locally Unbiased Implies Globally Unbiased)**.** *Let $\phi : \mathbb{R}^n \to \mathbb{R}$ be any well-defined estimator for an exponential family $\{P_\eta : \eta \in U\}$. Let $I \subseteq U$ be an interval of nonzero length. If $\mathbb{E}_{X \leftarrow P_\eta}[\phi(X)] = \eta$ for all $\eta \in I$, then $\mathbb{E}_{X \leftarrow P_\eta}[\phi(X)] = \eta$ for all $\eta \in U$.*

We prove this result by showing that the expectation of the estimator $\mathbb{E}_{X \leftarrow P_\eta}[\phi(X)]$ must be an analytic function of the parameter η. A function being analytic means that its Taylor series provides an exact representation of the function. Thus, if an analytic function is linear in some nontrivial interval, we can compute the Taylor series at an interior point of that interval to deduce that the function is linear globally, which yields the result.

[1] Definition 5.0.1 is stated in terms of densities, but it can be extended to discrete distributions.

5.2 Pure DP Estimators Are Uniformly Bounded

We now exploit the strong group privacy property of pure DP to show that a pure DP estimator that is bounded locally is uniformly bounded globally.

Proposition 5.2.1 (Pure DP Estimators Are Uniformly Bounded)**.** *Let $A : \mathcal{X}^n \to \mathbb{R}$ be a randomized algorithm. If A is $(\varepsilon, 0)$-DP, then for all $x, x^* \in \mathcal{X}^n$, $\left|\mathop{\mathbb{E}}_{A}[A(x)]\right| \leq e^{\varepsilon n}\mathop{\mathbb{E}}_{A}[|A(x^*)|]$.*

The proof is in Appendix D.4. We emphasize that the above result holds for any x, x^* and thus the bound on $\left|\mathop{\mathbb{E}}_{A}[A(x)]\right|$ is *uniform* – i.e., it does not depend on x.

Our impossibility result for exponential families now follows by stringing together the tools we have collected so far.

Proof of Theorem 5.0.2. Suppose, for the sake of contradiction, there exist $\varepsilon \geq 0$, $n \geq 0$, and an ε-DP algorithm $M : \mathbb{R}^n \to \mathbb{R}$ for which $\mathbb{E}_{X \leftarrow P_\eta^n, M}[M(X)] = \eta$ when $\eta \in I$. By Proposition D.3.2, $\{P_\eta^n : \eta\}$ is an exponential family, such that for every $P \in \{P_\eta : \eta\}$, the natural parameter of P^n is the same as that of P. Therefore, by Proposition 5.1.1, we have $\mathbb{E}_{X \leftarrow P_\eta^n, M}[M(X)] = \eta$ for all $\eta \in U$. In particular, since U is unbounded, $\mathbb{E}_{X \leftarrow P_\eta^n, M}[M(X)]$ must be an unbounded function of η, which contradicts Proposition 5.2.1. □

References

Aden-Ali, I., Ashtiani, H. & Kamath, G. (2021), On the sample complexity of privately learning unbounded high-dimensional gaussians, ALT.

Aden-Ali, I., Ashtiani, H. & Liaw, C. (2021), Privately learning mixtures of axis-aligned gaussians, NeurIPS.

Ahlfors, L. V. (1953), 'Complex analysis: an introduction to the theory of analytic functions of one complex variable', *New York, London* **177**.

Alabi, D., Kothari, P. K., Tankala, P., Venkat, P. & Zhang, F. (2022), 'Privately estimating a Gaussian: Efficient, robust and optimal', *arXiv preprint arXiv:2212.08018* .

Amin, K., Kulesza, A., Munoz, A. & Vassilvitskii, S. (2019), Bounding user contributions: A bias-variance trade-off in differential privacy, ICML.

Ashtiani, H. & Liaw, C. (2022), Private and polynomial time algorithms for learning Gaussians and beyond, COLT.

Asi, H. & Duchi, J. C. (2020), Instance-optimality in differential privacy via approximate inverse sensitivity mechanisms, NeurIPS.

Avella-Medina, M. & Brunel, V.-E. (2019), 'Differentially private sub-Gaussian location estimators', *arXiv preprint arXiv:1906.11923* .

Avent, B., Dubey, Y. & Korolova, A. (2019), 'The power of the hybrid model for mean estimation', *Proceedings on Privacy Enhancing Technologies* **2020**(4), 48–68.

Balle, B., Bell, J., Gascón, A. & Nissim, K. (2019), The privacy blanket of the shuffle model, CRYPTO.

Barber, R. F. & Duchi, J. C. (2014), 'Privacy and statistical risk: Formalisms and minimax bounds', *arXiv preprint arXiv:1412.4451* .

Barrientos, A. F., Williams, A. R., Snoke, J. & Bowen, C. M. (2021*a*), 'Differentially private methods for validation servers'.

Barrientos, A. F., Williams, A. R., Snoke, J. & Bowen, C. M. (2021*b*), 'A feasibility study of differentially private summary statistics and regression analyses for administrative tax data', *arXiv preprint arXiv:2110.12055* .

Bassily, R., Nissim, K., Smith, A., Steinke, T., Stemmer, U. & Ullman, J. (2016), Algorithmic stability for adaptive data analysis, STOC.

Ben-Eliezer, O., Mikulincer, D. & Zadik, I. (2022), Archimedes meets privacy: On privately estimating quantiles in high dimensions under minimal assumptions, NeurIPS.

Bie, A., Kamath, G. & Singhal, V. (2022), Private estimation with public data, NeurIPS.

Biswas, S., Dong, Y., Kamath, G. & Ullman, J. (2020), Coinpress: Practical private mean and covariance estimation, NeurIPS.

Boneh, D. & Shaw, J. (1998), 'Collusion-secure fingerprinting for digital data', *IEEE Transactions on Information Theory* **44**(5), 1897–1905.

Brown, G., Gaboardi, M., Smith, A., Ullman, J. & Zakynthinou, L. (2021), Covariance-aware private mean estimation without private covariance estimation, NeurIPS.

Bun, M., Kamath, G., Steinke, T. & Wu, Z. S. (2019), Private hypothesis selection, NeurIPS.

Bun, M. & Steinke, T. (2016), Concentrated differential privacy: Simplifications, extensions, and lower bounds, TCC-B.

Bun, M. & Steinke, T. (2019), Average-case averages: Private algorithms for smooth sensitivity and mean estimation, NeurIPS.

Bun, M., Steinke, T. & Ullman, J. (2017), Make up your mind: The price of online queries in differential privacy, SODA.

Bun, M., Ullman, J. & Vadhan, S. (2014), Fingerprinting codes and the price of approximate differential privacy, STOC.

Cai, T. T., Wang, Y. & Zhang, L. (2020), 'The cost of privacy in generalized linear models: Algorithms and minimax lower bounds', *arXiv preprint arXiv:2011.03900* .

Cai, T. T., Wang, Y. & Zhang, L. (2021), 'The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy', *The Annals of Statistics* **49**(5), 2825–2850.

Cai, T. T., Wang, Y. & Zhang, L. (2023), 'Score attack: A lower bound technique for optimal differentially private learning', *arXiv preprint arXiv:2303.07152* .

Chen, H., Cohen-Addad, V., d'Orsi, T., Epasto, A., Imola, J., Steurer, D. & Tiegel, S. (2023), 'Private estimation algorithms for stochastic block models and mixture models', *arXiv preprint arXiv:2301.04822* .

Cheu, A., Smith, A., Ullman, J., Zeber, D. & Zhilyaev, M. (2019), Distributed differential privacy via shuffling, EUROCRYPT.

Covington, C., He, X., Honaker, J. & Kamath, G. (2021), 'Unbiased statistical estimation and valid confidence intervals under differential privacy', *arXiv preprint 2110.14465* .

Cramér, H. (1946), 'Mathematical methods of statistics.'.

Diakonikolas, I., Hardt, M. & Schmidt, L. (2015), Differentially private learning of structured discrete distributions, NIPS.

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

Duchi, J. C., Jordan, M. I. & Wainwright, M. J. (2013), Local privacy and statistical minimax rates, FOCS.

Duchi, J., Haque, S. & Kuditipudi, R. (2023), 'A fast algorithm for adaptive private mean estimation', *arXiv preprint arXiv:2301.07078* .

Dwork, C., Feldman, V., Hardt, M., Pitassi, T., Reingold, O. & Roth, A. (2015), 'The reusable holdout: Preserving validity in adaptive data analysis', *Science* **349**(6248).

Dwork, C., McSherry, F., Nissim, K. & Smith, A. (2006), Calibrating noise to sensitivity in private data analysis, TCC.

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

Dwork, C. & Rothblum, G. N. (2016), 'Concentrated differential privacy', *arXiv preprint arXiv:1603.01887* .

Dwork, C., Smith, A., Steinke, T., Ullman, J. & Vadhan, S. (2015), Robust traceability from trace amounts, FOCS.

Erlingsson, Ú., Feldman, V., Mironov, I., Raghunathan, A., Talwar, K. & Thakurta, A. (2019), Amplification by shuffling: From local to central differential privacy via anonymity, SODA.

Evans, G. & King, G. (2021), 'Statistically valid inferences from differentially private data releases, with application to the Facebook URLs dataset', *Political Analysis* **31**(1), 1–21.

Evans, G., King, G., Schwenzfeier, M. & Thakurta, A. (2022), 'Statistically valid inferences from privacy protected data'.

Feldman, V., McMillan, A. & Talwar, K. (2022), Hiding among the clones: A simple and nearly optimal analysis of privacy amplification by shuffling, FOCS.

Feldman, V., McMillan, A. & Talwar, K. (2023), *Stronger Privacy Amplification by Shuffling for Renyi and Approximate Differential Privacy*, pp. 4966–4981.

Feldman, V. & Steinke, T. (2017), Generalization for adaptively-chosen estimators via stable median, *in* 'Conference on Learning Theory'.

Ferrando, C., Wang, S. & Sheldon, D. (2022), Parametric bootstrap for differentially private confidence intervals, AISTATS.

Gauss, C.-F. (1823), *Theoria combinationis observationum erroribus minimis obnoxiae*, Henricus Dieterich.

George, A. J., Ramesh, L., Singh, A. V. & Tyagi, H. (2022), 'Continual mean estimation under user-level privacy', *arXiv preprint arXiv:2212.09980* .

Georgiev, K. & Hopkins, S. B. (2022), Privacy induces robustness: Information-computation gaps and sparse mean estimation, NeurIPS.

Hardt, M. & Ullman, J. (2014), Preventing false discovery in interactive data analysis is hard, FOCS.

Hodges, J. L. & Lehmann, E. L. (1950), 'Some Problems in Minimax Point Estimation', *The Annals of Mathematical Statistics* **21**(2), 182 – 197.

Hopkins, S. B., Kamath, G. & Majid, M. (2022), Efficient mean estimation with pure differential privacy via a sum-of-squares exponential mechanism, STOC.

Hopkins, S. B., Kamath, G., Majid, M. & Narayanan, S. (2022), 'Robustness implies privacy in statistical estimation', *arXiv preprint arXiv:2212.05015* .

Huang, Z., Liang, Y. & Yi, K. (2021), Instance-optimal mean estimation under differential privacy, NeurIPS.

Kamath, G., Li, J., Singhal, V. & Ullman, J. (2019), Privately learning high-dimensional distributions, COLT.

Kamath, G., Liu, X. & Zhang, H. (2022), Improved rates for differentially private stochastic convex optimization with heavy-tailed data, ICML.

Kamath, G., Mouzakis, A. & Singhal, V. (2022), New lower bounds for private estimation and a generalized fingerprinting lemma, NeurIPS.

Kamath, G., Mouzakis, A., Singhal, V., Steinke, T. & Ullman, J. (2022), A private and computationally-efficient estimator for unbounded gaussians, COLT.

Kamath, G., Sheffet, O., Singhal, V. & Ullman, J. (2019), Differentially private algorithms for learning mixtures of separated Gaussians, NeurIPS.

Kamath, G., Singhal, V. & Ullman, J. (2020), Private mean estimation of heavy-tailed distributions, COLT.

Kamath, G. & Ullman, J. (2020), 'A primer on private statistics', *arXiv preprint arXiv:2005.00010* .

Karwa, V. & Vadhan, S. (2018), Finite sample differentially private confidence intervals, ITCS.

Kasiviswanathan, S. P. & Smith, A. (2014), 'On the'semantics' of differential privacy: A bayesian formulation', *Journal of Privacy and Confidentiality* **6**(1).

Kothari, P. K., Manurangsi, P. & Velingker, A. (2022), Private robust estimation by stabilizing convex relaxations, COLT.

Lehmann, E. & Scheffé, H. (1950), 'Completeness, similar regions, and unbiased estimation. i. sankhy a 10, 305–340', *Proc. R. Soc. A* .

Levy, D., Sun, Z., Amin, K., Kale, S., Kulesza, A., Mohri, M. & Suresh, A. T. (2021), Learning with user-level privacy, NeurIPS.

Liu, X., Kong, W., Kakade, S. & Oh, S. (2021), Robust and differentially private mean estimation, NeurIPS.

Liu, X., Kong, W. & Oh, S. (2022), Differential privacy and robust statistics in high dimensions, COLT.

Liu, Y., Suresh, A. T., Yu, F., Kumar, S. & Riley, M. (2020), Learning discrete distributions: User vs item-level privacy, NeurIPS.

Markov, A. A. (1900), *Ischislenie veroiatnostei*, Tipografia Imperatorskoi Akademii nauk.

Nikolov, A. & Tang, H. (2023), 'Gaussian noise is nearly instance optimal for private unbiased mean estimation', *arXiv preprint arXiv:2301.13850* .

Ramsay, K. & Chenouri, S. (2021), 'Differentially private depth functions and their associated medians', *arXiv preprint arXiv:2101.02800* .

Ramsay, K., Jagannath, A. & Chenouri, S. (2022), 'Concentration of the exponential mechanism and differentially private multivariate medians', *arXiv preprint 2210.06459* .

Rao, C. R. (1945), 'Information and accuracy attainable in the estimation of statistical parameters', *Bulletin of the Calcutta Mathematical Society* **37**(3), 81–91.

Rogers, R., Roth, A., Smith, A. & Thakkar, O. (2016), Max-information, differential privacy, and post-selection hypothesis testing, FOCS.

Steinke, T. (2022), 'Composition of differential privacy & privacy amplification by subsampling'.
URL: *https://arxiv.org/abs/2210.00597*

Steinke, T. & Ullman, J. (2015), Interactive fingerprinting codes and the hardness of preventing false discovery, COLT.

Steinke, T. & Ullman, J. (2017*a*), 'Between pure and approximate differential privacy', *The Journal of Privacy and Confidentiality* **7**(2), 3–22.

Steinke, T. & Ullman, J. (2017*b*), Tight lower bounds for differentially private selection, FOCS.

Tardos, G. (2008), 'Optimal probabilistic fingerprint codes', *Journal of the ACM* **55**(2).

Tsfadia, E., Cohen, E., Kaplan, H., Mansour, Y. & Stemmer, U. (2022), Friendlycore: Practical differentially private aggregation, ICML.

Tzamos, C., Vlatakis-Gkaragkounis, E.-V. & Zadik, I. (2020), Optimal private median estimation under minimal distributional assumptions, NeurIPS.

Vadhan, S. (2017), The complexity of differential privacy.

Wang, D., Xiao, H., Devadas, S. & Xu, J. (2020), On differentially private stochastic convex optimization with heavy-tailed data, ICML.

Zhang, H., Kamath, G., Kulkarni, J. & Wu, Z. S. (2020), Privately learning Markov random fields, ICML.

Zhu, K., Fioretto, F. & Van Hentenryck, P. (2022), Post-processing of differentially private data: A fairness perspective, IJCAI.

Zhu, K., Fioretto, F., Van Hentenryck, P., Das, S. & Task, C. (2023), 'Privacy and bias analysis of disclosure avoidance systems', *arXiv preprint arXiv:2301.12204* .

Zhu, K., Van Hentenryck, P. & Fioretto, F. (2021), Bias and variance of post-processing in differential privacy, AAAI.

Appendix A

Proofs for Chapter 2

Throughout the proofs given here, we will use the symbol "$\stackrel{d}{=}$" to denote distributional equivalence. That is, if two random variables X and Y have the same distribution, then we write $X \stackrel{d}{=} Y$. Moreover, for distributions P and Q over a set $\mathcal{X}$, we say that P and Q are (ε, δ)-indistinguishable (denoted by $P \sim_{\varepsilon,\delta} Q$), if for all measurable $E \subseteq \mathcal{X}$,

$$e^{-\varepsilon}(Q(E) - \delta) \leq P(E) \leq e^{\varepsilon}Q(E) + \delta$$

A.1 Proof of Lemma 2.1.2

Proof of Lemma 2.1.2. We assume, without loss of generality, that $\delta < 1$. Suppose, for now, there exists $c > 0$ such that $\mathbb{P}[|Y| > c] = \delta$. If $x \geq c$, then $\min\{\delta, \mathbb{P}[|Y| > x]\} = \mathbb{P}[|Y| > x] = \mathbb{P}[|Y| \cdot \mathbb{I}[|Y| > c] > x]$. Likewise, if $x \leq c$, then $\min\{\delta, \mathbb{P}[|Y| > x]\} = \delta = \mathbb{P}[|Y| \cdot \mathbb{I}[|Y| > c] > x]$. Thus,

$$\begin{aligned}
\int_0^\infty \min\{\delta, \mathbb{P}[|Y| > x]\}\mathrm{d}x &= \int_0^\infty \mathbb{P}[|Y| \cdot \mathbb{I}[|Y| > c] > x] \\
&= \mathbb{E}[|Y| \cdot \mathbb{I}[|Y| > c]] \\
&\leq \mathbb{E}[|Y|^\kappa]^{\frac{1}{\kappa}} \cdot \mathbb{E}\Big[\mathbb{I}[|Y| > c]^{\frac{\kappa}{\kappa-1}}\Big]^{\frac{\kappa-1}{\kappa}} \qquad \text{(Hölder's Inequality)} \\
&= \mathbb{E}[|Y|^\kappa]^{\frac{1}{\kappa}} \cdot \mathbb{P}[|Y| > c]^{1-1/\kappa} \\
&\leq \alpha \cdot \delta^{1-1/\kappa}.
\end{aligned}$$

If the distribution of Y is continuous, then such a quantity c is guaranteed to exist. In general, there exists $c \geq 0$ such that $\mathbb{P}[|Y| > c] \leq \delta \leq \mathbb{P}[|Y| \geq c]$. We can define a random $I : \mathbb{R} \to \{0, 1\}$ such that $\mathbb{I}[|Y| > c] \leq I(|Y|) \leq \mathbb{I}[|Y| \geq c]$ with probability 1 and $\mathbb{E}[I(|Y|)] = \delta$. The above proof carries through in general if we replace $\mathbb{I}[|Y| > c]$ with $I(|Y|)$. □

A.2 Non-Private Error of Mean Estimation

Given independent samples $X_1, \cdots, X_n \in \mathbb{R}$ from an unknown distribution P, the empirical mean $\hat{\mu}(X) \coloneqq \frac{1}{n}\sum_i^n X_i$ is an unbiased estimator of the distribution mean $\mu(P) \coloneqq \mathop{\mathbb{E}}_{X \leftarrow P}[X]$ and its mean squared error is

$$\alpha^2 \coloneqq \mathop{\mathbb{E}}_{X \leftarrow P^n}\left[(\hat{\mu}(X) - \mu(P))^2\right] = \frac{\mathop{\mathbb{E}}_{X \leftarrow P}[(X - \mu(P))^2]}{n} = O(1/n).$$

This mean squared error is asymptotically optimal in a minimax sense and is optimal for the univariate Gaussian case $P = \mathcal{N}(\mu, \sigma^2)$.

We have the following well-known result which shows that the empirical mean is asymptotically optimal even for the simple case of Bernoulli data.

Proposition A.2.1. *Let $M : \{0, 1\}^n \to \mathbb{R}$ be an estimator satisfying*

$$\forall p \in [0, 1] \quad \mathop{\mathbb{E}}_{X \leftarrow \mathsf{Bernoulli}(p)^n}\left[(M(X) - p)^2\right] \leq \alpha^2.$$

Then $\alpha^2 \geq \frac{1}{6(n+2)}$.

The empirical mean attains MSE $\mathop{\mathbb{E}}_{X \leftarrow \mathsf{Bernoulli}(p)^n}[(\hat{\mu}(X) - p)^2] = \frac{p(1-p)}{n}$. However, this is not the minimax optimal estimator of the mean of a Bernoulli distribution, rather it is the biased estimator

$$\check{\mu}(X) \coloneqq \frac{1}{n + \sqrt{n}}\left(\frac{\sqrt{n}}{2} + \sum_i^n X_i\right),$$

which has MSE

$$\mathop{\mathbb{E}}_{X \leftarrow \mathsf{Bernoulli}(p)^n}\left[(\check{\mu}(X) - p)^2\right] = \frac{1}{4(\sqrt{n} + 1)^2}$$

for all $p \in [0, 1]$ (Hodges & Lehmann 1950).

Proof of Proposition A.2.1. Let $P \in [0,1]$ be uniform and, conditioned on P, let $X \leftarrow \mathsf{Bernoulli}(P)^n$ be n independent bits, each with conditional expectation P. Note that the marginal distribution of $\sum_i^n X_i$ is uniform on $\{0, 1, \cdots, n\}$.

Given $X = x$, the conditional distribution of P is

$$P|_{X=x} \stackrel{d}{=} \mathsf{Beta}\left(1 + \sum_i^n x_i, 1 + \sum_i^n (1 - x_i)\right).$$

In terms of mean squared error, the best estimator of P is simply the mean of this conditional distribution. That is, the function $f : \{0,1\}^n \to \mathbb{R}$ that minimizes $\mathop{\mathbb{E}}_{P,X}[(P - f(X))^2]$ is the conditional expectation $f(x) = \mathbb{E}[P \mid X = x]$. Indeed, this is the *definition* of conditional expectation in the general measure theoretic setting. Consequently, the best possible mean squared error of an estimator of P given X is the variance of this conditional distribution $P|X$.

The distribution $\mathsf{Beta}(a, b)$ has mean $\frac{a}{a+b}$ and variance $\frac{ab}{(a+b)^2(a+b+1)}$. Now we have

$$\begin{aligned}
\alpha^2 &\geq \mathop{\mathbb{E}}_{P \leftarrow [0,1], X \leftarrow \mathsf{Bernoulli}(P)^n} \left[(M(X) - P)^2\right] \\
&\geq \mathop{\mathbb{E}}_{X}\left[\mathop{\mathbb{E}}_{P}\left[\left(\mathop{\mathbb{E}}_{P}[P \mid X] - P\right)^2 \mid X\right]\right] \\
&= \mathop{\mathbb{E}}_{P \leftarrow [0,1], X \leftarrow \mathsf{Bernoulli}(P)^n}\left[\left(\frac{1 + \sum_i^n X_i}{2 + n} - P\right)^2\right] \\
&= \mathop{\mathbb{E}}_{P \leftarrow [0,1], X \leftarrow \mathsf{Bernoulli}(P)^n}\left[\frac{(1 + \sum_i^n X_i)(1 + \sum_i^n (1 - X_i))}{(n+2)^2(n+3)}\right] \\
&= \frac{1}{n+1}\sum_{k=0}^{n} \frac{(1+k)(1+n-k)}{(n+2)^2(n+3)} = \frac{1}{(n+1)(n+2)^2(n+3)}\sum_{k=0}^{n}(1+n) + n \cdot k - k^2 \\
&= \frac{1}{(n+1)(n+2)^2(n+3)}\left((1+n)\cdot(n+1) + n \cdot \frac{n(n+1)}{2} - \frac{n(n+1)(2n+1)}{6}\right) \\
&= \frac{6(n+1)^2 + 3n^2(n+1) - n(n+1)(2n+1)}{6(n+1)(n+2)^2(n+3)} = \frac{6(n+1) + 3n^2 - n(2n+1)}{6(n+2)^2(n+3)} \\
&= \frac{5n + 6 + n^2}{6(n+2)^2(n+3)} = \frac{(n+2)(n+3)}{6(n+2)^2(n+3)} = \frac{1}{6(n+2)}.
\end{aligned}$$

This completes the proof. □

If we change the distribution of $P \in [0,1]$ from uniform to $\mathsf{Beta}(\sqrt{n}/2, \sqrt{n}/2)$ in the above proof, then we obtain the stronger conclusion $\alpha^2 \geq \frac{1}{4(\sqrt{n}+1)^2}$, which is exactly optimal. However, this requires a more complicated calculation.

A.3 Proof of Theorem 2.1.1

We will use the following lemma in the proof.

Lemma A.3.1 (Fingerprinting Derivative Lemma (Steinke & Ullman 2017*b*, Lemma 9))**.** *Let $f : \{0,1\}^n \to \mathbb{R}$ be an arbitrary function. Define $g : [0,1] \to \mathbb{R}$ by $g(p) = \mathbb{E}_{X \leftarrow \mathsf{Bernoulli}(p)^n}[f(X)]$. Then, for all $p \in [0,1]$, we have*

$$g'(p) \cdot p(1-p) = \mathop{\mathbb{E}}_{X \leftarrow \mathsf{Bernoulli}(p)^n}\left[f(X) \cdot \sum_i^n (X_i - p)\right].$$

Proof of Theorem 2.1.1. For $p \in [0,1]$ and $v > 0$, define $\mathcal{D}_{v,p} = v \cdot \mathsf{Bernoulli}(p)$ – i.e., a sample from $\mathcal{D}_{v,p}$ is 0 with probability $1-p$ and v with probability p. Then $\mu(\mathcal{D}_{v,p}) = \mathop{\mathbb{E}}_{X \leftarrow \mathcal{D}_{v,p}}[X] = vp$ and

$$\mathop{\mathbb{E}}_{X \leftarrow \mathcal{D}_{v,p}}\left[|X - \mu(\mathcal{D}_{v,p})|^\lambda\right] = (1-p)(vp)^\lambda + p(v(1-p))^\lambda \leq 2pv^\lambda.$$

If we ensure $v \leq (2p)^{-1/\lambda} \leq 1/p$, then the λ-th absolute central moment is below 1, and the mean is in the interval $[0,1]$, so the bias and accuracy guarantees of M apply.

For $v > 0$, define $g_v : [0,1] \to \mathbb{R}$ by

$$g_v(p) \coloneqq \mathop{\mathbb{E}}_{X \leftarrow \mathcal{D}_{v,p}^n, M}[M(X)].$$

By Lemma A.3.1, for all $v > 0$ and $p \in [0,1]$, we have

$$\mathop{\mathbb{E}}_{X \leftarrow \mathcal{D}_{v,p}^n, M}\left[M(X) \cdot \sum_i^n \left(\frac{1}{v} X_i - p\right)\right] = p(1-p) g_v'(p). \tag{A.1}$$

Fix $0 < a < b \leq 1/2$ and $0 < v \leq (2b)^{-1/\lambda}$ (to be determined later). Now, let $P \in [a,b]$ be a random variable with density $\propto \frac{1}{P(1-P)}$ – i.e., $\forall t \in [a,b]$, $\mathbb{P}[P \leq t] = \frac{\int_a^t \frac{1}{x(1-x)} \mathrm{d}x}{\int_a^b \frac{1}{x(1-x)} \mathrm{d}x}$.

Conditioned on P, let $X_1, \ldots, X_n \in \mathbb{R}$ be independent samples from $\mathcal{D}_{v,P}$. Now,

$$\begin{aligned}
\mathop{\mathbb{E}}_{P,X,M}\left[M(X)\cdot\sum_i^n\left(\frac{1}{v}X_i - P\right)\right] &= \mathop{\mathbb{E}}_{P}[P(1-P)g_v'(P)] \qquad \text{(Equation A.1)}\\
&= \frac{\int_a^b g_v'(p)\mathrm{d}p}{\int_a^b \frac{1}{x(1-x)}\mathrm{d}x}\\
&= \frac{g_v(b)-g_v(a)}{\log(b/(1-b))-\log(a/(1-a))}.
\end{aligned}$$

By our bias assumption, $|g_v(b) - vb| \le \beta$ and $|g_v(a) - va| \le \beta$. Thus,

$$\mathop{\mathbb{E}}_{P,X,M}\left[M(X)\cdot\sum_i^n\left(\frac{1}{v}X_i - P\right)\right] \ge \frac{v\cdot(b-a)-2\beta}{\log\left(\frac{b\cdot(1-a)}{a\cdot(1-b)}\right)}.$$

Since $\mathbb{E}[X_i] = vP$ for all i, we can center $M(X)$ and rearrange slightly:

$$\sum_i^n \mathop{\mathbb{E}}_{P,X,M}\left[(M(X)-vP)\cdot\left(\frac{1}{v}X_i - P\right)\right] \ge \frac{v\cdot(b-a)-2\beta}{\log\left(\frac{b\cdot(1-a)}{a\cdot(1-b)}\right)}.$$

Next, we will use differential privacy to prove an upper bound on this quantity. Fix an arbitrary $i \in [n]$ and fix $P = p \in [a, b]$. Our goal is to upper bound $\mathop{\mathbb{E}}_{X\leftarrow\mathcal{D}_{v,p}^n,M}\left[(M(X)-vp)\cdot\left(\frac{1}{v}X_i - p\right)\right]$.

Since M satisfies (ε,δ)-DP, the distribution of the pair $(M(X), X_i)$ is (ε,δ)-indistinguishable from that of $\left(M(X_{-i},\tilde{X}_i), X_i\right)$, where $(X_{-i},\tilde{X}_i)$ denotes the dataset X with X_i replaced by $\tilde{X}_i$; here $\tilde{X}_i \leftarrow \mathcal{D}_{v,p}$ is a fresh sample from the distribution. Now $\tilde{X}_i$ and X_i are interchangeable, this means the distribution of $\left(M(X_{-i},\tilde{X}_i), X_i\right)$ is identical to that of $\left(M(X),\tilde{X}_i\right)$. By transitivity, the distribution of $(M(X), X_i)$ is (ε,δ)-indistinguishable from that of $\left(M(X),\tilde{X}_i\right)$. In particular,

$$(M(X)-vp)\cdot\left(\tfrac{1}{v}X_i - p\right) \sim_{\varepsilon,\delta} \left(M(X_{-i},\tilde{X}_i)-vp\right)\cdot\left(\tfrac{1}{v}X_i - p\right) \overset{d}{=} (M(X)-vp)\cdot\left(\tfrac{1}{v}\tilde{X}_i - p\right).$$

We also have $\left|(M(X)-vp)\cdot\left(\frac{1}{v}X_i - p\right)\right| \le |M(X)-vp|$ with probability 1.

Thus

$$\begin{aligned}\mathbb{E}\left[(M(X)-vp)\cdot\left(\frac{1}{v}X_i-p\right)\right] &= \mathbb{E}\left[\max\{(M(X)-vp)\cdot\left(\frac{1}{v}X_i-p\right),0\}\right] \\ &\quad -\mathbb{E}\left[\max\{-(M(X)-vp)\cdot\left(\frac{1}{v}X_i-p\right),0\}\right] \\ &= \int_0^\infty \mathbb{P}\left[(M(X)-vp)\cdot\left(\frac{1}{v}X_i-p\right)>x\right]\mathrm{d}x \\ &\quad -\int_0^\infty \mathbb{P}\left[(M(X)-vp)\cdot\left(\frac{1}{v}X_i-p\right)<-x\right]\mathrm{d}x.\end{aligned}$$

We have

$$\mathbb{P}\left[(M(X)-vp)\cdot\left(\frac{1}{v}X_i-p\right)>x\right]\leq\mathbb{P}[|M(X)-vp|>x]$$

and simultaneously,

$$\begin{aligned}\mathbb{P}\left[(M(X)-vp)\cdot\left(\frac{1}{v}X_i-p\right)>x\right] &\leq e^{\varepsilon}\cdot\mathbb{P}\left[(M(X)-vp)\cdot\left(\frac{1}{v}\tilde{X}_i-p\right)>x\right]+\delta \\ &= e^{\varepsilon}\cdot p\cdot\mathbb{P}[(M(X)-vp)\cdot(1-p)>x] \\ &\quad +e^{\varepsilon}\cdot(1-p)\cdot\mathbb{P}[(M(X)-vp)\cdot(0-p)>x]+\delta.\end{aligned}$$

Define $\delta(x) \coloneqq \min\{\delta, \mathbb{P}[|M(X) - vp| > x]\}$. Then

$$\begin{aligned}
\int_0^\infty \mathbb{P}\left[(M(X) - vp) \cdot \left(\frac{1}{v} X_i - p\right) > x\right] \mathrm{d}x &\leq \int_0^\infty e^\varepsilon \cdot p \cdot \mathbb{P}[(M(X) - vp) \cdot (1-p) > x] \mathrm{d}x \\
&\quad + \int_0^\infty e^\varepsilon \cdot (1-p) \cdot \mathbb{P}[(M(X) - vp) \cdot (0-p) > x] \mathrm{d}x \\
&\quad + \int_0^\infty \delta(x) \mathrm{d}x \\
&= \mathbb{E}[e^\varepsilon \cdot p \cdot \max\{(M(X) - vp) \cdot (1-p), 0\}] \\
&\quad + \mathbb{E}[e^\varepsilon \cdot (1-p) \cdot \max\{(M(X) - vp) \cdot (0-p), 0\}] \\
&\quad + \int_0^\infty \delta(x) \mathrm{d}x \\
&\leq e^\varepsilon \cdot p(1-p) \cdot \mathbb{E}[|M(X) - vp|] + \alpha \cdot \tau.
\end{aligned}$$

In the above, the final inequality holds because

$$\mathbb{E}[\max\{M(X) - vp, 0\}] + \mathbb{E}[\max\{-M(X) + vp, 0\}] = \mathbb{E}[|M(X) - vp|]$$

and due to the third utility assumption in our theorem statement. Similarly,

$$
\begin{aligned}
&\int_0^\infty \mathbb{P}\Big[(M(X)-vp)\cdot\Big(\frac{1}{v}X_i-p\Big)<-x\Big]\mathrm{d}x \\
&\qquad\geq \int_0^\infty \max\left\{\begin{array}{c} 0, \\ e^{-\varepsilon}\Big(\mathbb{P}\Big[(M(X)-vp)\cdot\Big(\frac{1}{v}\tilde{X}_i-p\Big)<-x\Big]-\delta\Big)\end{array}\right\}\mathrm{d}x \\
&\qquad= e^{-\varepsilon}\cdot\int_0^\infty \mathbb{P}\Big[(M(X)-vp)\cdot\Big(\frac{1}{v}\tilde{X}_i-p\Big)<-x\Big]\mathrm{d}x \\
&\qquad\quad + e^{-\varepsilon}\cdot\int_0^\infty \max\left\{\begin{array}{c} -\delta, \\ -\mathbb{P}\Big[(M(X)-vp)\cdot\Big(\frac{1}{v}\tilde{X}_i-p\Big)<-x\Big]\end{array}\right\}\mathrm{d}x \\
&\qquad\geq e^{-\varepsilon}\cdot\int_0^\infty \mathbb{P}\Big[(M(X)-vp)\cdot\Big(\frac{1}{v}\tilde{X}_i-p\Big)<-x\Big]-\delta(x)\mathrm{d}x \\
&\qquad= e^{-\varepsilon}\cdot\int_0^\infty p\cdot\mathbb{P}[(M(X)-vp)\cdot(1-p)<-x]\mathrm{d}x \\
&\qquad\quad + e^{-\varepsilon}\cdot\int_0^\infty (1-p)\cdot\mathbb{P}[(M(X)-vp)\cdot(0-p)<-x]\mathrm{d}x \\
&\qquad\quad - e^{-\varepsilon}\cdot\int_0^\infty \delta(x)\mathrm{d}x \\
&\qquad\geq e^{-\varepsilon}\cdot p(1-p)\cdot\mathbb{E}[|M(X)-vp|]-e^{-\varepsilon}\cdot\alpha\cdot\tau.
\end{aligned}
$$

Putting these two pieces together, we have:

$$\begin{aligned}
\mathop{\mathbb{E}}_{X_1,\cdots,X_n\leftarrow\mathcal{D}_{v,p}}\left[(M(X)-vp)\cdot\left(\frac{1}{v}X_i-p\right)\right] &\leq \int_0^\infty \mathbb{P}\left[(M(X)-vp)\cdot\left(\frac{1}{v}X_i-p\right)>x\right]\mathrm{d}x \\
&\quad - \int_0^\infty \mathbb{P}\left[(M(X)-vp)\cdot\left(\frac{1}{v}X_i-p\right)<-x\right]\mathrm{d}x \\
&\leq \left(e^{\varepsilon}-e^{-\varepsilon}\right)\cdot p(1-p)\cdot\mathbb{E}[|M(X)-vp|] \\
&\quad + (1+e^{-\varepsilon})\cdot\alpha\cdot\tau \\
&\leq \left(e^{\varepsilon}-e^{-\varepsilon}\right)\cdot b(1-b)\cdot\alpha+2\alpha\tau,
\end{aligned}$$

as $p\leq b\leq 1/2$. Then, combining this with our lower bound, we have

$$\begin{aligned}
\frac{v\cdot(b-a)-2\beta}{\log\left(\frac{b\cdot(1-a)}{a\cdot(1-b)}\right)} &\leq \sum_i^n \mathop{\mathbb{E}}_{P,X,M}\left[(M(X)-vP)\cdot\left(\frac{1}{v}X_i-P\right)\right] \\
&\leq n\cdot\left(\left(e^{\varepsilon}-e^{-\varepsilon}\right)\cdot b(1-b)\cdot\alpha+2\alpha\tau\right) \\
&\leq \alpha\cdot n\cdot 2\cdot(\sinh(\varepsilon)\cdot b+\tau),
\end{aligned}$$

which rearranges to

$$\alpha\geq\frac{v\cdot(b-a)-2\beta}{2n\cdot(\sinh(\varepsilon)\cdot b+\tau)\cdot\log\left(\frac{b\cdot(1-a)}{a\cdot(1-b)}\right)}.$$

It only remains to set the parameters subject to the constraints $0<a<b\leq 1/2$ and $0<v\leq(2b)^{-1/\lambda}$. First, we set $b=2a$, and $v=(2b)^{-1/\lambda}=(4a)^{-1/\lambda}$ and assume $(8\beta)^{\frac{\lambda}{\lambda-1}}\leq a\leq 1/5$, which simplifies the above expression to

$$\alpha\geq\frac{a^{1-1/\lambda}\cdot 4^{-1/\lambda}-2\beta}{2n\cdot(\sinh(\varepsilon)\cdot 2a+\tau)\cdot\log\left(\frac{2\cdot(1-a)}{1-2a}\right)}\geq\frac{a^{1-1/\lambda}-8\beta}{8n\cdot(\sinh(\varepsilon)\cdot 2a+\tau)}.$$

We reparameterize $a=\gamma^{\frac{\lambda}{\lambda-1}}$ for some $16\beta\leq\gamma\leq 1/5$ to obtain

$$\alpha\geq\frac{\gamma-8\beta}{8n\cdot(\sinh(\varepsilon)\cdot 2\cdot\gamma^{\frac{\lambda}{\lambda-1}}+\tau)}\geq\frac{\gamma/2}{8n\cdot(\sinh(\varepsilon)\cdot 2\cdot\gamma^{\frac{\lambda}{\lambda-1}}+\tau)}=\frac{1}{32n\sinh(\varepsilon)\gamma^{1/(\lambda-1)}+16n\tau\gamma^{-1}}.$$

This completes our proof. □

A.4 Proof of Theorem 2.2.2

We first prove the extension of privacy amplification by shuffling (Theorem 2.2.2). This proof is a direct reduction to the following result of Feldman et al. (2022).

Theorem A.4.1 (Local Privacy Amplification by Shuffling (Feldman et al. 2022, Theorem 3.8))**.** *Let $m \in \mathbb{Z}^+$, let $\mathcal{X}$ be the data universe, and let $\mathcal{Y}_1, \dots, \mathcal{Y}_m$ be image spaces. Suppose for each $i \in [m]$, we have a randomized function $R_i : \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_{i-1} \times \mathcal{X} \to \mathcal{Y}_i$ such that $R_i(y, a)$ is $(\varepsilon_0, \delta_0)$-DP in the parameter $a \in \mathcal{X}$ for every fixed $y \in \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_{i-1}$. Consider $R_m \otimes \cdots \otimes R_1 : \mathcal{X}^m \to \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_m$ defined by*

$$(R_m \otimes \cdots \otimes R_1)(x_1, \dots, x_m) \coloneqq (y_1, \dots, y_m)$$

where we recursively define $y_i \coloneqq R_i(y_1, \dots, y_{i-1}, x_i)$. In addition, consider the random shuffle operator $S : \mathcal{X}^m \to \mathcal{X}^m$ given by

$$S(x_1, \dots, x_m) \coloneqq (x_{\pi(1)}, \dots, x_{\pi(m)})$$

where π is a uniformly random permutation on $[m]$. Then, for any $\delta_1 \in [2\exp(-\frac{m}{16e^{\varepsilon_0}}), 1]$, the function $R_m \otimes \cdots \otimes R_1 \circ S : \mathcal{X}^m \to \mathcal{Y}_1 \times \cdots \mathcal{Y}_m$ is $(\varepsilon_1, \delta_1 + (e^{\varepsilon_1} + 1)(1 + e^{-\varepsilon_0}/2)m\delta_0)$-DP, where ε_1 is as in Equation 2.2.

Proof of Theorem 2.2.2. Consider neighboring datasets $x = ((x_1^1, \dots, x_1^n), \dots, (x_m^1, \dots, x_m^n))$ and $x' = ((x'^1_1, \dots, x'^n_1), \dots, (x'^1_m, \dots, x'^n_m))$ in $(\mathcal{X}^n)^m$ and assume, without loss of generality, that they differ in only the first entry of the first block. That is, $x_i^j = x'^j_i$ for all $(i, j) \neq (1, 1)$.

Now, decompose the operator $\Pi = \Pi_1 \circ \Pi_{-1}$ as follows.

$$\Pi_1(x)_i^j \coloneqq \begin{cases} \Pi(x)_i^j & \text{if } j = 1 \\ x_i^j & \text{otherwise} \end{cases} \quad \text{and} \quad \Pi_{-1}(x)_i^j \coloneqq \begin{cases} x_i^j & \text{if } j = 1 \\ \Pi(x)_i^j & \text{otherwise} \end{cases}$$

In other words, Π_1 applies the permutation π_1 to the first row and leaves the remaining $n - 1$ rows fixed, whereas Π_{-1} applies the permutations $\pi_2, \dots, \pi_n$ to every row except the first.

We claim that

$$(L_m \otimes \cdots \otimes L_1 \circ \Pi_1)(x) \sim_{\varepsilon', \delta'} (L_m \otimes \cdots \otimes L_1 \circ \Pi_1)(x') \tag{A.2}$$

with $\varepsilon' = O(\varepsilon_0 \sqrt{\log(1/\delta)/m})$ and $\delta' = \delta_1 + O(\delta_0 m)$, as in the conclusion of Theorem A.4.1.

To that end, consider the randomized function

$$R_i(y, a) \coloneqq L_i(y, (a, x_i^2, \ldots, x_i^n))$$

for $i \in [m]$. Since $(a, x_i^2, \ldots, x_i^n)$ and $(a', x_i^2, \ldots, x_i^n)$ are neighboring datasets for any $a, a' \in \mathcal{X}$, R_i must be $(\varepsilon_0, \delta_0)$-DP in the parameter a and hence $R_m \otimes \cdots \otimes R_1 \circ S$ is (ε', δ')-DP by Theorem A.4.1. In particular, since $\hat{x} \coloneqq (x_1^1, \ldots, x_m^1)$ and $\hat{x}' \coloneqq ({x'}_1^1, \ldots, {x'}_m^1)$ are neighbors, $(R_m \otimes \cdots \otimes R_1 \circ S)(\hat{x})$ must be (ε', δ')-indistinguishable from $(R_m \otimes \cdots \otimes R_1 \circ S)(\hat{x}')$.

Therefore, to prove our claim, it suffices to show that $(L_m \otimes \cdots \otimes L_1 \circ \Pi_1)(x)$ is identically distributed to $(R_m \otimes \cdots \otimes R_1 \circ S)(\hat{x})$, and likewise for $(L_m \otimes \cdots \otimes L_1 \circ \Pi_1)(x')$ and $(R_m \otimes \cdots \otimes R_1 \circ S)(\hat{x}')$. Indeed,

$$L_i(y, \Pi_1(x)_i) = L_i(y, (x_{\pi_1(i)}^1, x_i^2, \ldots, x_i^n)) = R_i(y, x_{\pi_1(i)}^1)$$

for all i. So, it follows by induction that

$$(L_m \otimes \cdots \otimes L_1 \circ \Pi_1)(x) = (R_m \otimes \cdots \otimes R_1)(x_{\pi_1(1)}^1, \ldots, x_{\pi_1(m)}^1) \stackrel{d}{=} (R_m \otimes \cdots \otimes R_1 \circ S)(\hat{x}).$$

Analogously, we get

$$(L_m \otimes \cdots \otimes L_1 \circ \Pi_1)(x') \stackrel{d}{=} (R_m \otimes \cdots \otimes R_1 \circ S)(\hat{x}'),$$

as desired.

We can now leverage the decomposition $\Pi = \Pi_1 \circ \Pi_{-1}$ to prove the theorem. Fixing Π_{-1}, $\Pi_{-1}(x)$ and $\Pi_{-1}(x')$ are neighboring datasets differing only on the first element of the first block. So, by the claim that we proved above (Equivalence A.2), which used only the fact that x and x' differ at $x_1^1 \neq {x'}_1^1$, we have that

$$(L_m \otimes \cdots \otimes L_1 \circ \Pi_1)(\Pi_{-1}(x)) \sim_{\varepsilon', \delta'} (L_m \otimes \cdots \otimes L_1 \circ \Pi_1)(\Pi_{-1}(x')).$$

But Π_{-1} depends only on $\pi_2, \ldots, \pi_n$ and is, thus, is independent of $L_m \otimes \cdots \otimes L_1 \circ \Pi_1$. Therefore, it follows that

$$\begin{aligned}
\mathbb{P}[(L_m \otimes \cdots \otimes L_1 \circ \Pi)(x) \in E] &= \mathop{\mathbb{E}}_{\Pi_{-1}} [\mathbb{P}[(L_m \otimes \cdots \otimes L_1 \circ \Pi_1)(\Pi_{-1}(x)) \in E \mid \Pi_{-1}]] \\
&\leq \mathop{\mathbb{E}}_{\Pi_{-1}} \left[\delta' + e^{\varepsilon'} \mathbb{P}[(L_m \otimes \cdots \otimes L_1 \circ \Pi_1)(\Pi_{-1}(x')) \in E \mid \Pi_{-1}]\right] \\
&= \delta' + e^{\varepsilon'} \mathbb{P}[(L_m \otimes \cdots \otimes L_1 \circ \Pi)(x') \in E]
\end{aligned}$$

for any measurable E. □

A.5 Proof of Theorem 2.2.3

Before we proceed, we recall the well-known post-processing property of differential privacy.

Lemma A.5.1 (Post-Processing (Dwork et al. 2006)). *If $M : \mathcal{X}^n \to \mathcal{Y}$ is (ε, δ)-DP and $P : \mathcal{Y} \to \mathcal{Z}$ is any randomized function, then the algorithm $P \circ M$ is (ε, δ)-DP.*

Proof of Theorem 2.2.3. Let $m \in \mathbb{N}$ (we delay our choice of m until later). Consider $A_m : (\mathbb{R}^n)^m \to \mathbb{R}$ defined by

$$\forall x_1, \ldots, x_m \in \mathbb{R}^n, \quad A_m((x_1, \ldots, x_m)) = \frac{1}{m} \sum_{i=1}^{m} M(x_i).$$

Fix some distribution P with mean and variance bounded by 1. This gives us the following guarantee about the mean of A_m.

$$\begin{aligned}
\breve{\mu} \coloneqq \mathop{\mathbb{E}}_{(X_1,\ldots,X_m)\leftarrow(P^n)^m, A_m}[A_m((X_1, \ldots, X_m))] &= \mathop{\mathbb{E}}_{(X_1,\ldots,X_m)\leftarrow(P^n)^m, M}\left[\frac{1}{m}\sum_{i=1}^{m} M(X_i)\right] \\
&= \frac{1}{m}\sum_{i=1}^{m} \mathop{\mathbb{E}}_{X_i\leftarrow P^n, M}[M(X_i)] \\
&= \mathop{\mathbb{E}}_{X\leftarrow P^n, M}[M(X)]
\end{aligned}$$

Thus, the bias of A_m is at most β, as we see from the following.

$$|\breve{\mu} - \mu(P)| = \left|\mathop{\mathbb{E}}_{(X_1,\ldots,X_m)\leftarrow(P^n)^m, A_m}[A_m((X_1, \ldots, X_m))] - \mu(P)\right| = \left|\mathop{\mathbb{E}}_{X\leftarrow P^n, M}[M(X)] - \mu(P)\right| \leq \beta$$

Similarly, the MSE of A_m is

$$\begin{aligned}
\mathop{\mathbb{E}}_{(X_1,\cdots,X_m)\leftarrow(P^n)^m, A_m}\left[(A_m(X_1, \cdots, X_m) - \mu(P))^2\right] &= (\breve{\mu} - \mu(P))^2 \\
&\quad + \mathop{\mathbb{E}}_{(X_1,\cdots,X_m)\leftarrow(P^n)^m, A_m}\left[(A_m(X_1, \cdots, X_m) - \breve{\mu})^2\right] \\
&= (\breve{\mu} - \mu(P))^2 \\
&\quad + \mathop{\mathbb{E}}_{(X_1,\cdots,X_m)\leftarrow(P^n)^m, M}\left[\left(\frac{1}{m}\sum_{i=1}^{m} M(X_i) - \breve{\mu}\right)^2\right] \\
&= (\breve{\mu} - \mu(P))^2 + \frac{1}{m} \cdot \mathop{\mathbb{E}}_{X\leftarrow P^n, M}\left[(M(X) - \breve{\mu})^2\right] \\
&\leq \beta^2 + \frac{\alpha^2}{m}.
\end{aligned}$$

Let $\Pi : (\mathcal{X}^n)^m \to (\mathcal{X}^n)^m$ be the shuffle operator described in Theorem 2.2.2. Since any set of samples drawn i.i.d. from a distribution is invariant under shuffling, $(A_m \circ \Pi)(X)$ has the same distribution as $A_m(X)$ when $X \leftarrow (P^n)^m$. In particular, $A_m \circ \Pi$ has the same bias and MSE as A_m on inputs from $(P^n)^m$. Privacy amplification by shuffling (Theorem 2.2.2) and post-processing (Lemma A.5.1), imply that $A_m \circ \Pi$ is $\left(\varepsilon' := O(\varepsilon\sqrt{\log(1/\delta_1)/m}), \delta' := \delta_1 + O(\delta m)\right)$-DP for all $\delta_1 \in [2\exp(-\frac{m}{16e^\varepsilon}), 1]$.

Now, we apply Theorem 2.2.1 (Kamath et al. 2020, Theorem 3.8) to A_m: There exists a distribution P with mean and variance bounded by 1, such that

$$\mathop{\mathbb{E}}_{(X_1,\cdots,X_m)\leftarrow(P^n)^m, A_m}\left[(A_m(X_1,\cdots,X_m) - \mu(P))^2\right] \geq \Omega\left(\frac{1}{nm(\varepsilon' + \delta')}\right).$$

Combining all these inequalities gives

$$\begin{aligned}
\beta^2 + \frac{\alpha^2}{m} &\geq \mathop{\mathbb{E}}_{(X_1,\cdots,X_m)\leftarrow(P^n)^m}\left[(A_m(X_1,\cdots,X_m) - \mu(P))^2\right] \\
&\geq \Omega\left(\frac{1}{nm(\varepsilon' + \delta')}\right) \\
&\geq \Omega\left(\frac{1}{nm(\varepsilon\sqrt{\log(1/\delta_1)/m} + \delta_1 + \delta m)}\right).
\end{aligned}$$

This rearranges to

$$\alpha^2 \geq \Omega\left(\frac{1}{n\varepsilon\sqrt{\log(1/\delta_1)/m} + n\delta_1 + nm\delta}\right) - m\beta^2.$$

It only remains to set $m \in \mathbb{N}$ and $\delta_1 \in [2\exp(-\frac{m}{16e^\varepsilon}), 1]$ to maximize this lower bound.

Now, we assume that $\delta_1 \leq O(\varepsilon/\sqrt{m})$ and $\delta \leq O(\varepsilon/m^{3/2})$. Then the first term in the denominator dominates and we have

$$\alpha^2 \geq \Omega\left(\frac{1}{n\varepsilon\sqrt{\log(1/\delta_1)/m}}\right) - m\beta^2.$$

Then setting $m = \Theta\left(\frac{1}{n^2\varepsilon^2\beta^4\log(1/\delta_1)}\right)$ optimizes the expression giving

$$\alpha^2 \geq \Omega\left(\frac{1}{n^2\varepsilon^2\beta^2\log(1/\delta_1)}\right).$$

We set $\delta_1 = n\varepsilon^2\beta^2 \leq O(\varepsilon/\sqrt{m})$. This satisfies $\delta_1 \in [2\exp(-\frac{m}{16e^\varepsilon}), 1]$ as long as $\beta^2 \leq 1/n\varepsilon^2$ and $m = \Theta\Big(\frac{1}{n^2\varepsilon^2\beta^4\log(1/n\varepsilon^2\beta^2)}\Big) \geq O(e^\varepsilon \log(1/n\varepsilon^2\beta^2))$. The latter constraint rearranges to $\beta^2\log(1/n\varepsilon^2\beta^2) \leq \Omega\Big(\frac{e^{-\varepsilon}}{n\varepsilon}\Big)$. To conclude, we note that the assumption $\delta \leq O(\varepsilon/m^{3/2})$ is implied by $\delta \leq O(n^3\varepsilon^4\beta^6)$. □

Appendix B

Proofs for Chapter 3

B.1 Proof of Proposition 3.0.1

We will require a couple of technical lemmata in our analysis.

Lemma B.1.1. $\forall \lambda > 1 \; \forall c > 0 \; \forall t \in \mathbb{R}, \quad \max\{0, t-c\} \leq \frac{(\lambda-1)^{\lambda-1}}{\lambda^{\lambda} \cdot c^{\lambda-1}} \cdot |t|^{\lambda}$.

Proof. If $t = 0$, the claim holds as an equality. Now, we assume that $t \in \mathbb{R} \setminus \{0\}$. Define $f : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ by $f(t) = |t|^{\lambda}$. Then $f'(t) = \lambda \cdot |t|^{\lambda-1} \cdot \mathsf{sign}(t)$ and $f''(t) = \lambda(\lambda-1) \cdot |t|^{\lambda-2} \geq 0$ for all $t \in \mathbb{R} \setminus \{0\}$. Since f is convex,

$$\forall a \geq 0 \, \forall t \in \mathbb{R} \setminus \{0\}, \;\; f(t) \geq f(a) + f'(a) \cdot (t-a) = a^{\lambda} + \lambda \cdot a^{\lambda-1} \cdot (t-a) = \lambda \cdot a^{\lambda-1} \cdot \left(t - \frac{\lambda-1}{\lambda} \cdot a \right).$$

Taking the maximum over $a = 0$ and $a = \frac{c\lambda}{\lambda-1}$ and rearranging yields the result. □

The following lemma decomposes the mean squared error of the clipped mean into the sum of the sampling error and the (squared) population bias introduced (which is further bounded).

Lemma B.1.2. *Fix $\lambda > 1$ and $a < b$. Let P be a distribution with mean $\mu(P) \in (a, b)$. Let $\mu_{[a,b]}(P) \coloneqq \mathop{\mathbb{E}}_{X \leftarrow P}\left[\mathsf{clip}_{[a,b]}(X)\right] \in [a, b]$. Let $X_1, \ldots, X_n$ be independent samples from P. Then*

$$\mathbb{E}\left[\left(\frac{1}{n} \sum_{i}^{n} \mathsf{clip}_{[a,b]}(X_i) - \mu(P) \right)^2 \right] \leq \frac{\mathop{\mathbb{E}}_{X \leftarrow P}[(X - \mu(P))^2]}{n} + (\mu_{[a,b]}(P) - \mu(P))^2$$

and

$$\left|\mu_{[a,b]}(P) - \mu(P)\right| \le \frac{(\lambda-1)^{\lambda-1}}{\lambda^\lambda} \cdot \frac{\mathop{\mathbb{E}}_{X\leftarrow P}\left[|X-\mu(P)|^\lambda\right]}{(\min\{\mu(P)-a, b-\mu(P)\})^{\lambda-1}} \le \frac{1}{\lambda} \cdot \frac{\mathop{\mathbb{E}}_{X\leftarrow P}\left[|X-\mu(P)|^\lambda\right]}{(\min\{\mu(P)-a, b-\mu(P)\})^{\lambda-1}}.$$

Proof. We have

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{1}{n}\sum_i^n \mathsf{clip}_{[a,b]}(X_i) - \mu(P)\right)^2\right] &= \mathbb{E}\left[\left(\frac{1}{n}\sum_i^n \mathsf{clip}_{[a,b]}(X_i) - \mu_{[a,b]}(P)\right)^2\right] + (\mu_{[a,b]}(P)-\mu(P))^2 \\
&= \frac{1}{n^2}\sum_i^n \mathbb{E}\left[\left(\mathsf{clip}_{[a,b]}(X_i) - \mu_{[a,b]}(P)\right)^2\right] + (\mu_{[a,b]}(P)-\mu(P))^2 \\
&\le \frac{1}{n^2}\sum_i^n \mathbb{E}\left[\left(\mathsf{clip}_{[a,b]}(X_i) - \mu(P)\right)^2\right] + (\mu_{[a,b]}(P)-\mu(P))^2 \\
&\le \frac{1}{n^2}\sum_i^n \mathbb{E}\left[(X_i-\mu(P))^2\right] + (\mu_{[a,b]}(P)-\mu(P))^2 \\
&= \frac{\mathop{\mathbb{E}}_{X\leftarrow P}[(X-\mu(P))^2]}{n} + (\mu_{[a,b]}(P)-\mu(P))^2.
\end{aligned}$$

The first inequality follows from the fact that $\mathop{\mathbb{E}}_{X\leftarrow P}[(X-\mu(P))^2] = \inf_{u\in\mathbb{R}} \mathop{\mathbb{E}}_{X\leftarrow P}[(X-u)^2]$. The second inequality follows from the fact that $\mu(P) \in [a,b]$ and, hence, $(\mathsf{clip}_{[a,b]}(x) - \mu(P))^2 \le (x-\mu(P))^2$ for all $x \in \mathbb{R}$.

It remains to bound $\mu_{[a,b]}(P) - \mu(P)$. We have

$$\begin{aligned}
\mu_{[a,b]}(P) - \mu(P) &= \mathop{\mathbb{E}}_{X\leftarrow P}\left[\mathsf{clip}_{[a,b]}(X) - X\right] \\
&= \mathop{\mathbb{E}}_{X\leftarrow P}[\mathbb{I}[X>b](b-X) + \mathbb{I}[X<a](a-X)] \\
&= \mathop{\mathbb{E}}_{X\leftarrow P}[\max\{a-X, 0\}] - \mathop{\mathbb{E}}_{X\leftarrow P}[\max\{X-b, 0\}].
\end{aligned}$$

By Lemma B.1.1,

$$\begin{aligned}
0 \le \mathop{\mathbb{E}}_{X\leftarrow P}[\max\{X-b,0\}] &= \mathop{\mathbb{E}}_{X\leftarrow P}[\max\{(X-\mu(P)) - (b-\mu(P)), 0\}] \\
&\le \mathop{\mathbb{E}}_{X\leftarrow P}\left[\frac{(\lambda-1)^{\lambda-1}}{\lambda^\lambda \cdot (b-\mu(P))^{\lambda-1}} \cdot |X-\mu(P)|^\lambda\right].
\end{aligned}$$

Similarly,

$$0 \le \mathop{\mathbb{E}}_{X\leftarrow P}[\max\{a-X,0\}] \le \frac{(\lambda-1)^{\lambda-1}}{\lambda^\lambda\cdot(\mu(P)-a)^{\lambda-1}}\cdot \mathop{\mathbb{E}}_{X\leftarrow P}\big[|X-\mu(P)|^\lambda\big].$$

Thus,

$$\frac{(\lambda-1)^{\lambda-1}}{\lambda^\lambda\cdot(b-\mu(P))^{\lambda-1}}\cdot \mathop{\mathbb{E}}_{X\leftarrow P}\big[|X-\mu(P)|^\lambda\big] \le \mu_{[a,b]}(P)-\mu(P) \le \frac{(\lambda-1)^{\lambda-1}}{\lambda^\lambda\cdot(\mu(P)-a)^{\lambda-1}}\cdot \mathop{\mathbb{E}}_{X\leftarrow P}\big[|X-\mu(P)|^\lambda\big].$$

Finally, note that $\frac{(\lambda-1)^{\lambda-1}}{\lambda^\lambda} = \left(1-\frac{1}{\lambda}\right)^{\lambda-1}\cdot\frac{1}{\lambda} \le \frac{e^{-1+1/\lambda}}{\lambda} \le \frac{1}{\lambda}$. □

Finally, we will also require standard properties of the well-known Laplace mechanism.

Definition B.1.3 (Sensitivity)**.** Let $f:\mathcal{X}^n\to\mathbb{R}$ be a function, its *sensitivity* is

$$\Delta_f \coloneqq \sup_{x\sim x'\in\mathcal{X}^n} |f(x)-f(x')|.$$

Lemma B.1.4 (Laplace Mechanism)**.** *Let $f:\mathcal{X}^n\to\mathbb{R}$ be a function with sensitivity Δ_f. Then, denoting by $\mathsf{Lap}(b)$ the Laplace distribution with location 0 and scale parameter b, the Laplace mechanism $M(x)\coloneqq f(x)+\mathsf{Lap}(\Delta_f/\varepsilon)$ satisfies ε-DP. Furthermore,*

$$\mathbb{P}\left[|M(x)-f(x)| \ge \frac{\Delta_f\cdot\log(1/\beta)}{\varepsilon}\right] \le \beta.$$

Proof of Proposition 3.0.1. The properties of the Laplace distribution ensure that M satisfies ε-DP, as the sensitivity of $\widehat{\mu}(x)\coloneqq\frac{1}{n}\sum_i^n \mathsf{clip}_{[\hat{a},\hat{b}]}(x_i)$ is $\frac{\hat{b}-\hat{a}}{n}$ (Lemma B.1.4).

It only remains to analyze the bias and accuracy. Fix an arbitrary distribution P which satisfies the conditions in the proposition statement. By Lemma B.1.2, the bias satisfies

$$\begin{aligned}
\mathop{\mathbb{E}}_{X\leftarrow P^n,M}[M(X)]-\mu(P) &= \mathop{\mathbb{E}}_{X\leftarrow P^n}[\widehat{\mu}(X)]-\mu(P)\\
&= \mathop{\mathbb{E}}_{X\leftarrow P}\Big[\mathsf{clip}_{[\hat{a},\hat{b}]}(X)-X\Big]\\
&\le \frac{(\lambda-1)^{\lambda-1}}{\lambda^\lambda}\cdot\frac{\mathop{\mathbb{E}}_{X\leftarrow P}\big[|X-\mu(P)|^\lambda\big]}{(\min\{\mu(P)-\hat{a},\hat{b}-\mu(P)\})^{\lambda-1}}\\
&\le 1\cdot\frac{1}{(\min\{a-\hat{a},\hat{b}-b\})^{\lambda-1}}\\
&= \beta,
\end{aligned}$$

and the mean squared error satisfies

$$\begin{aligned}
\mathop{\mathbb{E}}_{X_1,\cdots,X_n\leftarrow P,M}\left[(M(X)-\mu(P))^2\right] &= \mathop{\mathbb{E}}_{\substack{X\leftarrow P^n\\ \xi\leftarrow \mathsf{Lap}\left(\frac{b-a}{\varepsilon n}\right)}}\left[(\widehat{\mu}(X)+\xi-\mu(P))^2\right] \\
&= \mathop{\mathbb{E}}_{X\leftarrow P^n}\left[\left(\frac{1}{n}\sum_i^n \mathsf{clip}_{[\hat{a},\hat{b}]}(X_i)-\mu(P)\right)^2\right] + \mathop{\mathbb{E}}_{\xi\leftarrow \mathsf{Lap}\left(\frac{\hat{b}-\hat{a}}{\varepsilon n}\right)}\left[\xi^2\right] \\
&\le \frac{\mathop{\mathbb{E}}_{X\leftarrow P}[(X-\mu(P))^2]}{n} + \beta^2 + 2\left(\frac{\hat{b}-\hat{a}}{\varepsilon n}\right)^2 \\
&\le \frac{1}{n} + \beta^2 + \frac{2}{\varepsilon^2 n^2}\left(b-a+\frac{2}{\beta^{1/(\lambda-1)}}\right)^2.
\end{aligned}$$

Our proof is complete. □

B.2 Proof of Proposition 3.0.2

Proof of Proposition 3.0.2. Since A satisfies local $(0,\delta)$-DP, M satisfies $(0,\delta)$-DP. Since A is unbiased (i.e., $\forall x\in\mathbb{R}\ \mathop{\mathbb{E}}_{A}[A(x)]=x$), so is M. Finally, we calculate the mean squared error:

$$\begin{aligned}
\mathop{\mathbb{E}}_{X\leftarrow P^n,M}\left[(M(X)-\mu(P))^2\right] &= \frac{1}{n}\mathop{\mathbb{E}}_{X\leftarrow P,A}\left[(A(X)-\mu(P))^2\right] \\
&= \frac{\mathop{\mathbb{E}}_{X\leftarrow P,A}[A(X)^2]-\mu(P)^2}{n} \\
&= \frac{\mathop{\mathbb{E}}_{X\leftarrow P}[0+\delta\cdot(X/\delta)^2]-\mu(P)^2}{n} \\
&= \frac{\mathop{\mathbb{E}}_{X\leftarrow P}[X^2]-\delta\cdot\mu(P)^2}{\delta\cdot n} \\
&= \frac{\mathop{\mathbb{E}}_{X\leftarrow P}[(X-\mu(P))^2]+(1-\delta)\cdot\mu(P)^2}{\delta\cdot n}.
\end{aligned}$$

We have the required result. □

B.3 Proof of Proposition 3.0.3

Proof of Proposition 3.0.3. Since $\mathop{\mathbb{E}}_{A}[A(x)] = x$ for any $x \in \mathbb{R}$, the bias of the $(\varepsilon, 0)$-DP and the $(0, \delta)$-DP components cancel and thus M is unbiased – i.e. $\forall x \in \mathbb{R}^n \ \mathop{\mathbb{E}}_{M}[M(x)] = \frac{1}{n}\sum_i^n x_i$.

By composition and post-processing, M satisfies (ε, δ)-DP.

Now, we bound the mean squared error. Define $\mu_{[\hat{a},\hat{b}]}(P) \coloneqq \mathop{\mathbb{E}}_{X\leftarrow P}\Big[\mathsf{clip}_{[\hat{a},\hat{b}]}(X)\Big]$. We have:

$$\begin{aligned}
\mathop{\mathbb{E}}_{X\leftarrow P^n, M}\big[(M(X)-\mu(P))^2\big] &= \mathop{\mathbb{E}}_{\substack{X\leftarrow P^n \\ \xi\leftarrow \mathsf{Lap}\left(\frac{\hat{b}-\hat{a}}{n\varepsilon}\right), A}}\left[\left(\begin{array}{c} \frac{1}{n}\sum_i^n \mathsf{clip}_{[\hat{a},\hat{b}]}(X_i) + \xi \\ +\frac{1}{n}\sum_i^n A\Big(X_i - \mathsf{clip}_{[\hat{a},\hat{b}]}(X_i)\Big) - \mu(P) \end{array}\right)^2\right] \\
&= \mathop{\mathbb{E}}_{X\leftarrow P^n, A}\left[\left(\begin{array}{c} \frac{1}{n}\sum_i^n \mathsf{clip}_{[\hat{a},\hat{b}]}(X_i) - \mu_{[\hat{a},\hat{b}]}(P) \\ +\frac{1}{n}\sum_i^n A\Big(X_i - \mathsf{clip}_{[\hat{a},\hat{b}]}(X_i)\Big) - (\mu(P) - \mu_{[\hat{a},\hat{b}]}(P)) \end{array}\right)^2\right] \\
&\qquad + \mathop{\mathbb{E}}_{\xi\leftarrow \mathsf{Lap}\left(\frac{\hat{b}-\hat{a}}{n\varepsilon}\right)}\big[\xi^2\big] \\
&\le \frac{2}{n}\cdot \mathop{\mathbb{E}}_{X\leftarrow P}\Big[\Big(\mathsf{clip}_{[\hat{a},\hat{b}]}(X) - \mu_{[\hat{a},\hat{b}]}(P)\Big)^2\Big] + 2\left(\frac{\hat{b}-\hat{a}}{n\varepsilon}\right)^2 \\
&\qquad + \frac{2}{n}\cdot \mathop{\mathbb{E}}_{X\leftarrow P, A}\Big[\Big(A\Big(X - \mathsf{clip}_{[\hat{a},\hat{b}]}(X)\Big) - \Big(\mu(P) - \mu_{[\hat{a},\hat{b}]}(P)\Big)\Big)^2\Big].
\end{aligned} \tag{B.1}$$

The final inequality uses the fact that for independent mean-zero random variables U and V, we have $\mathbb{E}[(U+V)^2] = \mathbb{E}[U^2] + \mathbb{E}[V^2]$. For the terms that are not independent, we apply the inequality $\mathbb{E}[(U+V)^2] \le 2\mathbb{E}[U^2] + 2\mathbb{E}[V^2]$.

Since $\mu(P) \coloneqq \mathop{\mathbb{E}}_{X\leftarrow P}[X] \in [a, b] \subset [\hat{a}, \hat{b}]$, we have

$$\mathop{\mathbb{E}}_{X\leftarrow P}\Big[\Big(\mathsf{clip}_{[\hat{a},\hat{b}]}(X) - \mu_{[\hat{a},\hat{b}]}(P)\Big)^2\Big] \le \mathop{\mathbb{E}}_{X\leftarrow P}\Big[\Big(\mathsf{clip}_{[\hat{a},\hat{b}]}(X) - \mu(P)\Big)^2\Big] \le \mathop{\mathbb{E}}_{X\leftarrow P}\big[(X-\mu(P))^2\big] \le 1.$$

Finally we bound the last term:

$$\begin{aligned}
\mathop{\mathbb{E}}_{X\leftarrow P,A}\left[\left(A\big(X-\mathsf{clip}_{[a,b]}(X)\big)-\Big(\mu(P)-\mu_{[\hat{a},\hat{b}]}(P)\Big)\right)^2\right] &\le \mathop{\mathbb{E}}_{X\leftarrow P,A}\left[\left(A\Big(X-\mathsf{clip}_{[\hat{a},\hat{b}]}(X)\Big)\right)^2\right] \\
&= (1-\delta)\cdot 0+\delta\cdot\mathop{\mathbb{E}}_{X\leftarrow P}\left[\left(\frac{1}{\delta}\Big(X-\mathsf{clip}_{[\hat{a},\hat{b}]}(X)\Big)\right)^2\right] \\
&= \frac{1}{\delta}\cdot\mathop{\mathbb{E}}_{X\leftarrow P}\left[\Big(X-\mathsf{clip}_{[\hat{a},\hat{b}]}(X)\Big)^2\right] \\
&= \frac{1}{\delta}\cdot\mathop{\mathbb{E}}_{X\leftarrow P}\left[\begin{pmatrix}(X-\mu(P))-\\ \mathsf{clip}_{[\hat{a}-\mu(P),\hat{b}-\mu(P)]}(X-\mu(P))\end{pmatrix}^2\right] \\
&\le \frac{1}{\delta}\cdot\mathop{\mathbb{E}}_{X\leftarrow P}\left[\big((X-\mu(P))-\mathsf{clip}_{[-c,c]}(X-\mu(P))\big)^2\right] \\
&= \frac{1}{\delta}\cdot\mathop{\mathbb{E}}_{X\leftarrow P}\left[(\max\{0,|X-\mu(P)|-c\})^2\right],
\end{aligned}$$

where the final inequality holds because $c=\hat{b}-b\le\hat{b}-\mu(P)$ and $c=a-\hat{a}\le\mu(P)-a$. By Lemma B.1.1,

$$\max\{0,|X-\mu(P)|-c\}\le\frac{1}{\lambda/2}\cdot\frac{|X-\mu(P)|^{\lambda/2}}{c^{\lambda/2-1}}.$$

Thus,

$$\mathop{\mathbb{E}}_{X\leftarrow P}\left[(\max\{0,|X-\mu(P)|-c\})^2\right]\le\left(\frac{1}{(\lambda/2)\cdot c^{\lambda/2-1}}\right)^2\cdot\mathop{\mathbb{E}}_{X\leftarrow P}\left[|X-\mu(P)|^{\lambda}\right]\le\frac{4}{\lambda^2}\cdot\frac{\psi^\lambda}{c^{\lambda-2}}.$$

Now, we set parameters and assemble the bound from Inequality B.1:

$$\begin{aligned}
\mathop{\mathbb{E}}_{X\leftarrow P^n,M}\left[(M(X)-\mu(P))^2\right] &\le \frac{2}{n}+2\left(\frac{b-a+2c}{n\varepsilon}\right)^2+\left(\frac{2}{n}\cdot\frac{1}{\delta}\cdot\frac{4}{\lambda^2}\cdot\frac{\psi^\lambda}{c^{\lambda-2}}\right) \\
&\le \frac{2}{n}+\frac{4(b-a)^2}{n^2\varepsilon^2}+\frac{16}{n^2\varepsilon^2}\cdot c^2+\frac{8\psi^\lambda}{n\delta\lambda^2}\cdot(c^2)^{1-\lambda/2} \\
&= \frac{2}{n}+\frac{4(b-a)^2}{n^2\varepsilon^2}+\frac{16\psi^2}{n^2\varepsilon^2}\cdot\left(\frac{n\varepsilon^2}{4\lambda\delta}\right)^{2/\lambda}\cdot\left(\frac{\lambda}{\lambda-2}\right)^{1-2/\lambda} \\
&\le \frac{2}{n}+\frac{4(b-a)^2}{n^2\varepsilon^2}+\frac{24\psi^2}{n^2\varepsilon^2}\cdot\left(\frac{n\varepsilon^2}{4\lambda\delta}\right)^{2/\lambda},
\end{aligned}$$

where the final equality follows from setting $c = \left(\frac{n\varepsilon^2\psi^\lambda(\lambda-2)}{4\lambda^2\delta}\right)^{1/\lambda}$ to minimize the expression, and the final inequality follows from the fact that $\left(\frac{\lambda}{\lambda-2}\right)^{1-2/\lambda} = (1-2/\lambda)^{-1+2/\lambda} \leq e^{e^{-1}} < \frac{3}{2}$. □

Appendix C

Proofs for Chapter 4

Recall that, if two random variables X and Y have the same distribution, then we write $X \overset{d}{=} Y$.

C.1 Proof of Proposition 4.1.1

We recall the group privacy property of differential privacy, which quantifies the privacy guaranteed by a DP algorithm for a group of individuals within a dataset.

Lemma C.1.1 (Group Privacy (Dwork et al. 2006, Dwork & Roth 2014)). *Let $A : \mathcal{X}^n \to \mathcal{Y}$ be (ε, δ)-DP. Then for any integer $k \in \{0, \dots, n\}$, measurable subset $Y \subseteq \mathcal{Y}$, and pairs of datasets $x, x' \in \mathcal{X}^n$ differing in k elements,*

$$\mathbb{P}[A(x) \in Y] \leq e^{k\varepsilon} \cdot \mathbb{P}[A(x') \in Y] + \frac{e^{k\varepsilon} - 1}{e^{\varepsilon} - 1} \cdot \delta.$$

Proof of Proposition 4.1.1. We will prove that $\text{DPUCoarse}_{\varepsilon,\delta}$ satisfies $(\varepsilon/2, \delta/2e^{\varepsilon/2})$-DP with respect to addition or removal of one element of the dataset. By group privacy (Lemma C.1.1), this implies (ε, δ)-DP for replacement of an element.

Consider a fixed pair of datasets x and $x' = x_{-i_*}$, where x' is x with x_{i_*} removed for some $i_* \in [n]$. For the privacy analysis, we also consider the offset T to be fixed – i.e., T is not needed to ensure privacy.

By post-processing, we can consider an algorithm that outputs more information. Specifically, we can assume that for each $k \in \mathbb{Z}$, the algorithm outputs

$$\nu_k(x) \coloneqq \left\{ \begin{array}{cc} \max\left\{|\{i \in [n] : \mathsf{round}_{\mathbb{Z}}(x_i - T) = k\}| + \xi_k - 2 - \frac{2\log(1/\delta)}{\varepsilon}, 0\right\} & \text{if } k \in K \\ 0 & \text{if } k \in \mathbb{Z} \setminus K \end{array} \right\}.$$

Note that only finitely many of these $\nu_k(x)$ values will be nonzero, so the algorithm can output a compressed version of this infinite vector of values. We can obtain the true output of $\mathrm{DPUCoarse}_{\varepsilon,\delta}$ by taking the argmax of this vector or outputting $\perp$ if this vector is all zeros. The advantage of this perspective is that each $\nu_k(x)$ is independent, as it depends only on the noise ξ_k (the input x and offset T are fixed).

The output distributions on the neighboring inputs are the same except for one $\nu_k(x) \stackrel{d}{\neq} \nu_k(x')$, namely $k = \mathsf{round}_{\mathbb{Z}}(x_{i_*} - T)$. Thus, we must simply show that this value satisfies $(\varepsilon/2, \delta/2e^{\varepsilon/2})$-DP. That is, we must show $\nu_k(x) \sim_{\varepsilon/2,\delta/2e^{\varepsilon/2}} \nu_k(x')$, where $\nu_k(x)$ and $\nu_k(x')$ denote the relevant random variables on the two different inputs. There are two cases to consider: $|\{i \in [n] : \mathsf{round}_{\mathbb{Z}}(x_i - T) = k\}| = 1$ and $|\{i \in [n] : \mathsf{round}_{\mathbb{Z}}(x_i - T) = k\}| \geq 2$. (Note that $|\{i \in [n] : \mathsf{round}_{\mathbb{Z}}(x_i - T) = k\}| = 0$ is ruled out because $k = \mathsf{round}_{\mathbb{Z}}(x_{i_*} - T)$.)

Suppose $|\{i \in [n] : \mathsf{round}_{\mathbb{Z}}(x_i - T) = k\}| = 1$. Then $\nu_k(x') = 0$ deterministically. Therefore, it suffices to prove that $\mathbb{P}[\nu_k(x) = 0] \geq 1 - \delta/2e^{\varepsilon/2}$. We have

$$\begin{aligned} \mathbb{P}[\nu_k(x) \neq 0] &= \mathbb{P}\left[|\{i \in [n] : \mathsf{round}_{\mathbb{Z}}(x_i - T) = k\}| + \xi_k - 2 - \frac{2\log(1/\delta)}{\varepsilon} > 0\right] \\ &= \mathbb{P}\left[1 + \xi_k - 2 - \frac{2\log(1/\delta)}{\varepsilon} > 0\right] \\ &= \mathbb{P}\left[\xi_k > 1 + \frac{2\log(1/\delta)}{\varepsilon}\right] \\ &= \frac{1}{2}\exp\left(-\frac{\varepsilon}{2} \cdot \left(1 + \frac{2\log(1/\delta)}{\varepsilon}\right)\right) \\ &= \frac{\delta}{2e^{\varepsilon/2}}, \end{aligned}$$

where the penultimate equality follows from the fact that $\xi_k \leftarrow \mathsf{Lap}(2/\varepsilon)$ (Lemma B.1.4).

Now suppose $|\{i \in [n] : \mathsf{round}_{\mathbb{Z}}(x_i - T) = k\}| \geq 2$. Then $\nu_k(x)$ and $\nu_k(x')$ are post-processings of $|\{i \in [n] : \mathsf{round}_{\mathbb{Z}}(x_i - T) = k\}| + \xi_k$ and, respectively, $|\{i \in [n] : \mathsf{round}_{\mathbb{Z}}(x_i - T) = k\}| - 1 + \xi_k$. Thus, by the properties of Laplace noise, we have $e^{-\varepsilon/2}\mathbb{P}[\nu_k(x') \in S] \leq \mathbb{P}[\nu_k(x) \in S] \leq e^{\varepsilon/2}\mathbb{P}[\nu_k(x') \in S]$ for all S, as required. $\square$

C.2 Proof of Proposition 4.1.2

Lemma C.2.1. *For any $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ and $c \in \mathbb{R}$, we have*

$$\mathrm{DPUCoarse}_{\varepsilon,\delta}(x + c) \overset{d}{=} \mathrm{DPUCoarse}_{\varepsilon,\delta}(x) + c$$

where $\perp + c \coloneqq \perp$ and $x + c \coloneqq (x_1 + c, \dots, x_n + c)$.

Proof of Lemma C.2.1. It will be easier to proceed by rewriting Algorithm 1 in a non-algorithmic form. To that end, we define the following notations.

- For $r \in \mathbb{R}$, set $\widehat{p}_r(x) \coloneqq \frac{1}{n}|i \in [n] : x_i \in [r \pm 1/2)|$ and sample $T \leftarrow \mathcal{U}[\pm 1/2]$, $\widetilde{p}_r(x) \leftarrow \widehat{p}_r(x) + \mathsf{Lap}(0, 2/(\varepsilon n))$ such that T and $\{\widetilde{p}_r(x)\}_{r\in\mathbb{R}}$ are all mutually independent.
- For $t \in \mathbb{R}$ and $S \subseteq \mathbb{R}$, we define $S + t \coloneqq \{s + t : s \in S\}$, and denote by $S + T$ the distribution over the set of sets $\{S + t : t \in [\pm 1/2]\}$ induced by the randomness of T.
- Set $R(x) \coloneqq \{r \in \mathbb{Z} + T : \widehat{p}_r(x) > 0\}$ and put $R^*(x) \coloneqq \arg\max_{r\in R(x)} \widetilde{p}_r(x)$, provided there is an $r \in R(x)$ for which $\widetilde{p}_r(x) > \frac{2\log(2/\delta)}{\varepsilon n} + \frac{2}{n} =: \eta$, otherwise $R^*(x) \coloneqq \perp$.

Essentially, we have reparameterized the terms of $\mathrm{DPUCoarse}_{\varepsilon,\delta}(x)$ so that $R(x) = K + T$ and $\widehat{p}_{k+T}(x) = \frac{1}{n}|\{i \in [n] : \mathsf{round}_{\mathbb{Z}}(x_i - T) = k\}|$ hold, so it follows that $R^*(x) \overset{d}{=} \mathrm{DPUCoarse}_{\varepsilon,\delta}(x)$.

Now, notice that

$$\widehat{p}_r(x + c) = \frac{1}{n}|\{i \in [n] : x_i + c \in [r \pm 1/2)\}| = \frac{1}{n}|\{i \in [n] : x_i \in [r - c \pm 1/2)\}| = \widehat{p}_{r-c}(x),$$

so in particular, we have that $\widetilde{p}_r(x + c)$ is identically distributed to $\widetilde{p}_{r-c}(x)$ for any $r \in \mathbb{R}$. Moreover, $\mathbb{Z} + T$ is identically distributed to $\mathbb{Z} + T - c$, so it follows that

$$\begin{aligned}
R(x + c) &= \{r \in \mathbb{Z} + T : \widehat{p}_r(x + c) > 0\} \\
&= \{r \in \mathbb{Z} + T : \widehat{p}_{r-c}(x) > 0\} \\
&= \{r \in \mathbb{Z} + T - c : \widehat{p}_r(x) > 0\} + c \\
&\overset{d}{=} \{r \in \mathbb{Z} + T : \widehat{p}_r(x) > 0\} + c \\
&= R(x) + c.
\end{aligned}$$

As T and the Laplace noise were all sampled in a mutually independent manner, these distributional equivalences hold jointly, i.e., $((\widetilde{p}_r(x+c))_{r\in\mathbb{R}}, R(x+c)) \stackrel{d}{=} ((\widetilde{p}_{r-c}(x))_{r\in\mathbb{R}}, R(x)+c)$. Hence,

$$\begin{aligned}
R^*(x+c) &= \begin{cases} \arg\max_{r\in R(x+c)} \widetilde{p}_r(x+c) & \text{if } \exists r \in R(x+c), \widetilde{p}_r(x+c) > \eta \\ \perp & \text{otherwise.} \end{cases} \\
&\stackrel{d}{=} \begin{cases} \arg\max_{r\in R(x)+c} \widetilde{p}_{r-c}(x) & \text{if } \exists r \in R(x)+c, \widetilde{p}_{r-c}(x) > \eta \\ \perp & \text{otherwise.} \end{cases} \\
&= \begin{cases} \arg\max_{r'\in R(x)} \widetilde{p}_{r'}(x)+c & \text{if } \exists r' \in R(x), \widetilde{p}_{r'}(x) > \eta \\ \perp & \text{otherwise.} \end{cases} && (r = r'+c) \\
&= R^*(x)+c. && (\perp = \perp + c)
\end{aligned}$$

The equivalence of $R^*(x)$ and $\mathrm{DPUCoarse}_{\varepsilon,\delta}(x)$ gives us the desired result. □

Lemma C.2.2. *For any $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, we have*

$$\mathrm{DPUCoarse}_{\varepsilon,\delta}(-x) \stackrel{d}{=} -\mathrm{DPUCoarse}_{\varepsilon,\delta}(x)$$

where $-\perp := \perp$ and $-x := (-x_1, \ldots, -x_n)$.

Proof of Lemma C.2.2. Recall the notation from the proof of Lemma C.2.1. Then, we have that

$$\widehat{p}_r(-x) = \frac{1}{n}|\{i \in [n] : -x_i \in [r \pm 1/2)\}| = \frac{1}{n}|\{i \in [n] : x_i \in [-r \pm 1/2)\}| = \widehat{p}_{-r}(x),$$

so it follows that for any $r \in \mathbb{R}$, $\widetilde{p}_r(-x)$ is identically distributed to $\widetilde{p}_{-r}(x)$. Moreover,

$$\begin{aligned}
R(-x) &= \{r \in \mathbb{Z}+T : \widehat{p}_r(-x) > 0\} \\
&= \{r \in \mathbb{Z}+T : \widehat{p}_{-r}(x) > 0\} \\
&= -\{r \in -(\mathbb{Z}+T) : \widehat{p}_r(x) > 0\} \\
&\stackrel{d}{=} -\{r \in \mathbb{Z}+T : \widehat{p}_r(x) > 0\} \\
&= -R(x).
\end{aligned}$$

As T and all of the Laplace noise is sampled independently, these distributional equivalences hold simultaneously, namely $((\widetilde{p}_r(-x))_{r\in\mathbb{R}}, R(-x)) \stackrel{d}{=} ((\widetilde{p}_{-r}(x))_{r\in\mathbb{R}}, -R(x))$. Combining these with $-\perp = \perp$, we obtain $R^*(-x) \stackrel{d}{=} -R^*(x)$ by the same argument as the one we used to prove Lemma C.2.1. □

Proof of Proposition 4.1.2. Due to Lemma C.2.1, we may assume without loss of generality that P has center 0. Then, since P is symmetric, $X \leftarrow P^n$ is identically distributed to $-X$. So, for any $a \geq 0$,

$$\begin{aligned}
\mathop{\mathbb{P}}_{X\leftarrow P^n}[\text{DPUCoarse}_{\varepsilon,\delta}(X) \in [a,\infty)] &= \mathop{\mathbb{E}}_{X\leftarrow P^n}[\mathbb{P}[\text{DPUCoarse}_{\varepsilon,\delta}(X) \in [a,\infty)|X]] \\
&= \mathop{\mathbb{E}}_{X\leftarrow P^n}[\mathbb{P}[\text{DPUCoarse}_{\varepsilon,\delta}(-X) \in (-\infty,-a]|X]] \\
&\qquad\qquad \text{(Lemma C.2.2)} \\
&= \mathop{\mathbb{P}}_{X\leftarrow P^n}[\text{DPUCoarse}_{\varepsilon,\delta}(-X) \in (-\infty,-a]] \\
&= \mathop{\mathbb{P}}_{X\leftarrow P^n}[\text{DPUCoarse}_{\varepsilon,\delta}(X) \in (-\infty,-a]]. \\
&\qquad\qquad (X \overset{d}{=} -X)
\end{aligned}$$

In particular,

$$Q([a,\infty)) = \frac{\mathbb{P}[\widetilde{\mu} \in [a,\infty)]}{\mathbb{P}[\widetilde{\mu} \neq \perp]} = \frac{\mathbb{P}[\widetilde{\mu} \in (-\infty,-a]]}{\mathbb{P}[\widetilde{\mu} \neq \perp]} = Q((-\infty,a])$$

for all $a \geq 0$, so Q must also be symmetric with center 0. □

C.3 Proof of Proposition 4.1.3

Proof of Proposition 4.1.3. By Lemma C.2.1, we assume, without loss of generality, that $\mu(P) = 0$.

Let $X \leftarrow P^n$ be the input to $\text{DPUCoarse}_{\varepsilon,\delta}$ and let $\widetilde{\mu} \in \mathbb{R} \cup \{\perp\}$ be the output. Let $T \in [\pm\frac{1}{2}]$, $K \subset \mathbb{Z}$, and $\xi_k \leftarrow \mathsf{Lap}(2/\varepsilon)$ be as in the algorithm (and define $\xi_k = 0$ for $k \notin K$). For $k \in \mathbb{Z}$, define

$$C_k := |\{i \in [n] : \mathsf{round}_{\mathbb{Z}}(X_i - T) = k\}|.$$

Recall, from Algorithm 1, that $k \in K \iff C_k \geq 1$ and that $\widetilde{\mu} = \perp \iff \max_{k\in K} C_k + \xi_k \leq 2 + \frac{2}{\varepsilon}\log(1/\delta)$ and, otherwise, $\widetilde{\mu} = T + \arg\max_{k\in K} C_k + \xi_k$.

We begin by showing $\widetilde{\mu} \in [\pm 1]$ with high probabiliy.

Define

$$k_+ = \mathsf{round}_{\mathbb{Z}}\left(\frac{1}{2} - T\right) \quad \text{and} \quad k_- = \mathsf{round}_{\mathbb{Z}}\left(-\frac{1}{2} - T\right).$$

Note that $k_+ = k_- + 1$ and $k_+, k_- \in (-T-1, -T+1]$. Thus, if $\arg\max_{k\in K} C_k + \xi_k \in \{k_+, k_-\}$ (and $\max_{k\in K} C_k + \xi_k > 2 + \frac{2}{\varepsilon}\log(1/\delta)$), then $\widetilde{\mu} \in \{k_+ + T, k_- + T\} \subset (-1, +1]$, as required.

In other words, it suffices for us to show that, with high probability, $C_{k_+} + \xi_{k_+}$ or $C_{k_-} + \xi_{k_-}$ are large and all other $C_k + \xi_k$ values are small.

For any $x \in \left(-\frac{1}{2}, +\frac{1}{2}\right)$, we have $\mathsf{round}_{\mathbb{Z}}(x - T) \in \{k_-, k_+\}$. Thus

$$C_{k_-} + C_{k_+} \geq \sum_i^n \mathbb{I}\left[X_i \in \left(-\frac{1}{2}, +\frac{1}{2}\right)\right].$$

Due to our assumption that $\mathop{\mathbb{E}}_{X \leftarrow P}[(X - \mu(P))^2] < 1/64$, Chebyshev's inequality yields $\mathop{\mathbb{P}}_{X \leftarrow P}\left[X \in \left(\pm\frac{1}{2}\right)\right] \geq \frac{15}{16}$. Furthermore, by Hoeffding's inequality, we have

$$\mathop{\mathbb{P}}_{X \leftarrow P^n}\left[\sum_i^n \mathbb{I}\left[X_i \in \left(-\frac{1}{2}, +\frac{1}{2}\right)\right] \geq \frac{15}{16}n - s\right] \geq 1 - e^{-2s^2/n}$$

for all $s \geq 0$. In particular,

$$\mathop{\mathbb{P}}_{X \leftarrow P^n}\left[\sum_i^n \mathbb{I}\left[X_i \in \left(-\frac{1}{2}, +\frac{1}{2}\right)\right] \geq \frac{7}{8}n\right] \geq 1 - e^{-n/128}.$$

This means $\mathbb{P}\left[C_{k_-} + C_{k_+} \geq \frac{7}{8}n\right] \geq 1 - e^{-n/128}$. Define $k_* \coloneqq \arg\max_{k \in \{k_+, k_-\}} C_k$ (breaking ties arbitrarily). If $C_{k_-} + C_{k_+} \geq \frac{7}{8}n$, then $C_{k_*} \geq \frac{7}{16}n$, while $C_k \leq \frac{1}{8}n$ for all $k \notin \{k_+, k_-\}$. Thus

$$\mathbb{P}\left[C_{k_*} \geq \frac{7}{16}n \wedge \max_{k \in K \setminus \{k_+, k_-\}} C_k \leq \frac{1}{8}n\right] \geq 1 - e^{-n/128}.$$

The next step is to analyze the noise. For all $k \in K$ and $r \geq 0$, $\mathbb{P}[\xi_k \geq r] = \mathbb{P}[\xi_k \leq -r] = \frac{1}{2}e^{-r\varepsilon/2}$. Note that $|K| \leq n$. Setting $r = n/8$ and taking a union bound over $k \in K$, we have

$$\mathbb{P}\left[\xi_{k_*} \geq \frac{-n}{8} \wedge \max_{k \in K \setminus \{k_+, k_-\}} \xi_k \leq \frac{n}{8}\right] \geq 1 - \frac{n}{2}e^{-n\varepsilon/16}.$$

Combining the high probability bounds on the noise bound and the data, we have

$$\mathbb{P}\left[C_{k_*} + \xi_{k_*} \geq \frac{5}{16}n \wedge \max_{k \in K \setminus \{k_+, k_-\}} C_k + \xi_k \leq \frac{1}{4}n\right] \geq 1 - e^{-n/128} - \frac{n}{2}e^{-n\varepsilon/16}.$$

Since $\frac{5}{16}n \geq 2 + \frac{2}{\varepsilon}\log(1/\delta)$, the event $C_{k_*} + \xi_{k_*} \geq \frac{5}{16}n \wedge \max_{k \in K \setminus \{k_+, k_-\}} C_k + \xi_k \leq \frac{1}{4}n$ implies $\widetilde{\mu} \in \{T + k_+, T + k_-\} \subset [-1, +1]$, as required.

Finally, we bound $\mathbb{E}\big[\mathbb{I}[\widetilde{\mu} \neq \perp] \cdot |\widetilde{\mu} - \mu(P)|^\lambda\big]$. Observe that $\widetilde{\mu} = T + k$ for some $T \in [-1/2, +1/2]$ and $k = \mathsf{round}_{\mathbb{Z}}(X_i - T)$ for some $i \in [n]$. Thus, $|\widetilde{\mu} - X_i| \leq 1/2$ for some $i \in [n]$ and, hence, $|\widetilde{\mu} - \mu(P)| \leq 1/2 + \max_{i\in[n]} |X_i - \mu(P)|$. It follows that

$$\begin{aligned}
\mathbb{E}\big[\mathbb{I}[\widetilde{\mu} \neq \perp] \cdot |\widetilde{\mu} - \mu(P)|^\lambda\big] &\leq \mathbb{E}\left[\left(\frac{1}{2} + \max_{i\in[n]} |X_i - \mu(P)|\right)^\lambda\right] \\
&\overset{\text{(A)}}{\leq} \mathbb{E}\left[\frac{1}{2} + 2^{\lambda-1} \cdot \max_{i\in[n]} |X_i - \mu(P)|^\lambda\right] \\
&\overset{\text{(B)}}{\leq} \frac{1}{2} + 2^{\lambda-1} \cdot \sum_{i\in[n]} \mathbb{E}\big[|X_i - \mu(P)|^\lambda\big] \\
&\leq \frac{1}{2} + n \cdot 2^{\lambda-1} \cdot \psi^\lambda,
\end{aligned}$$

where Inequality A follows from the fact that $\forall p \geq 1\ \forall x, y \geq 0,\ \ (x+y)^p \leq (x^p + y^p) \cdot 2^{p-1}$, and Inequality B holds because the maximum among a set of non-negative real numbers should be at most the sum of those numbers. This completes our proof. □

C.4 Proof of Theorem 4.2.1

To prove Theorem 4.2.1, we use the following lemma that characterizes the symmetry of a clipped random variable from a symmetric distribution under special circumstances.

Lemma C.4.1. *Let P and Q be symmetric distributions with the same center $\mu(P) = \mu(Q)$. Let $c > 0$. Define a distribution R to be $\mathsf{clip}_{[Y-c,Y+c]}(X)$ where $X \leftarrow P$ and $Y \leftarrow Q$ are independent. Then R is symmetric with the same center $\mu(R) = \mu(P) = \mu(Q)$.*

Proof. Assume, without loss of generality, that $\mu(P) = \mu(Q) = 0$. Let $X \leftarrow P$ and $Y \leftarrow Q$ be independent. Let $Z = \mathsf{clip}_{[Y-c,Y+c]}(X)$.

We claim that

$$\forall x, y \in \mathbb{R} \quad \mathsf{clip}_{[(-y)-c,(-y)+c]}(-x) = -\mathsf{clip}_{[y-c,y+c]}(x).$$

This can be verified by analyzing the following cases: (1) $x < y - c$; (2) $x \in [y - c, y + c]$; and (3) $x > y + c$.

Since P and Q are symmetric, $\mathsf{clip}_{[(-Y)-c,(-Y)+c]}(-X)$ has the same distribution as Z. By the claim, this is simply $-Z$. Ergo, the distribution of Z is symmetric and centered at 0. □

Proof of Theorem 4.2.1. The privacy of Algorithm 2 follows from parallel composition, as we split the dataset in two, and apply (ε, δ)-DP algorithms to each half. Computing $\widetilde{\mu}$ is (ε, δ)-DP by Proposition 4.1.1. If $\widetilde{\mu} = \perp$, then we compute $\widehat{\mu}$ in a $(0, \delta)$-DP manner by sampling a δ fraction of the data points. If $\widetilde{\mu} \neq \perp$, then we compute $\widehat{\mu}$ in a $(\varepsilon, 0)$-DP manner using clipping and Laplace noise addition (Lemma B.1.4).

Note that $\widetilde{\mu}$ is independent from $X_{n_1+1}, \ldots, X_{n_1+n_2}$, which are the data points used to compute $\widehat{\mu}$. If $\widetilde{\mu} = \perp$, then we compute $\widehat{\mu}$ in an unbiased manner:

$$\mathbb{E}[\widehat{\mu} \mid \widetilde{\mu} = \perp] = \mathbb{E}\left[\frac{1}{n_2\delta}\sum_{i=n_1+1}^{n_1+n_2} X_i\xi_i\right] = \frac{1}{n_2\delta}\sum_{i=n_1+1}^{n_1+n_2}\mathbb{E}[X_i]\mathbb{E}[\xi_i] = \frac{1}{n_2\delta}\sum_{i=n_1+1}^{n_1+n_2}\mu(P)\delta = \mu(P).$$

Now, condition on $\widetilde{\mu} \neq \perp$. By Proposition 4.1.2, $\widetilde{\mu}$ has a symmetric distribution with center $\mu(P)$. By Lemma C.4.1, $\mathbb{E}\big[\mathsf{clip}_{[\widetilde{\mu}-c,\widetilde{\mu}+c]}(X_i) \mid \widetilde{\mu} \neq \perp\big] = \mu(P)$, which implies that $\mathbb{E}[\widehat{\mu} \mid \widetilde{\mu} \neq \perp] = \mu(P)$ because the Laplace noise has expected value 0. Combining these two cases implies $\mathbb{E}[\widehat{\mu}] = \mu(P)$.

Finally, we analyze the variance:

$$\mathbb{E}\big[(\widehat{\mu} - \mu(P))^2\big] = \mathbb{P}[\widetilde{\mu} = \perp] \cdot \mathbb{E}\big[(\widehat{\mu} - \mu(P))^2 \mid \widetilde{\mu} = \perp\big] + \mathbb{P}[\widetilde{\mu} \neq \perp] \cdot \mathbb{E}\big[(\widehat{\mu} - \mu(P))^2 \mid \widetilde{\mu} \neq \perp\big].$$

We bound the two terms for $\widetilde{\mu} = \perp$ and $\widetilde{\mu} \neq \perp$ separately. For the first term, Proposition 4.1.3 gives us $\mathbb{P}[\widetilde{\mu} = \perp] \leq \gamma$. Then we have the following.

$$\begin{aligned}
\mathbb{P}[\widetilde{\mu} = \perp] \cdot \mathbb{E}\big[(\widehat{\mu} - \mu(P))^2 \mid \widetilde{\mu} = \perp\big] &= \mathbb{P}[\widetilde{\mu} = \perp] \cdot \mathbb{E}\left[\left(\frac{1}{n_2\delta}\sum_{i=n_1+1}^{n_1+n_2} X_i\xi_i - \mu(P)\right)^2\right] \\
&= \mathbb{P}[\widetilde{\mu} = \perp] \cdot \frac{1}{n_2^2\delta^2}\sum_{i=n_1+1}^{n_1+n_2}\mathbb{E}\big[(X_i\xi_i - \mu(P))^2\big] \\
&\leq \mathbb{P}[\widetilde{\mu} = \perp] \cdot \frac{1}{n_2^2\delta^2}\sum_{i=n_1+1}^{n_1+n_2}\mathbb{E}\big[(X_i\xi_i)^2\big] \\
&= \mathbb{P}[\widetilde{\mu} = \perp] \cdot \frac{\mu(P)^2 + \mathbb{E}_{X\leftarrow P}[(X - \mu(P))^2]}{n_2\delta} \\
&\leq \gamma \cdot \frac{\mu(P)^2 + 1}{n_2\delta}.
\end{aligned}$$

Now, we bound the second term: $\mathbb{P}[\widetilde{\mu} \neq \perp] \cdot \mathbb{E}[(\widehat{\mu} - \mu(P))^2 \mid \widetilde{\mu} \neq \perp] = \mathbb{E}[\mathbb{I}[\widetilde{\mu} \neq \perp] \cdot (\widehat{\mu} - \mu(P))^2]$. We split this into two cases: $A := [\widetilde{\mu} \in [\mu(P) - \sigma, \mu(P) + \sigma]]$ and $B := [\widetilde{\mu} \in \mathbb{R} \setminus [\mu(P) - \sigma, \mu(P) + \sigma]]$.

Note that the event $A \wedge \widetilde{\mu} \neq \perp$ is equivalent to A because A cannot happen if $\widetilde{\mu} = \perp$, because $\perp \notin \mathbb{R}$. Similarly, $B \wedge \widetilde{\mu} \neq \perp$ is equivalent to B. Note that $\widetilde{\mu} \neq \perp \implies A \vee B$. Thus, we have

$$\begin{aligned}
\mathbb{P}[\widetilde{\mu} \neq \perp] \cdot \mathbb{E}\big[(\widehat{\mu} - \mu(P))^2 \mid \widetilde{\mu} \neq \perp\big] &= \mathbb{P}[\widetilde{\mu} \neq \perp] \cdot \mathbb{P}[A \mid \widetilde{\mu} \neq \perp] \cdot \mathbb{E}\big[(\widehat{\mu} - \mu(P))^2 \mid \widetilde{\mu} \neq \perp \wedge A\big] \\
&\quad + \mathbb{P}[\widetilde{\mu} \neq \perp] \cdot \mathbb{P}[B \mid \widetilde{\mu} \neq \perp] \cdot \mathbb{E}\big[(\widehat{\mu} - \mu(P))^2 \mid \widetilde{\mu} \neq \perp \wedge B\big] \\
&= \mathbb{P}[\widetilde{\mu} \neq \perp \wedge A] \cdot \mathbb{E}\big[(\widehat{\mu} - \mu(P))^2 \mid \widetilde{\mu} \neq \perp \wedge A\big] \\
&\quad + \mathbb{P}[\widetilde{\mu} \neq \perp \wedge B] \cdot \mathbb{E}\big[(\widehat{\mu} - \mu(P))^2 \mid \widetilde{\mu} \neq \perp \wedge B\big] \\
&= \mathbb{P}[\widetilde{\mu} \neq \perp \wedge A] \cdot \mathbb{E}\big[(\widehat{\mu} - \mu(P))^2 \mid \widetilde{\mu} \neq \perp \wedge A\big] \\
&\quad + \mathbb{E}\big[\mathbb{I}[\widetilde{\mu} \neq \perp] \cdot \mathbb{I}[B] \cdot (\widehat{\mu} - \mu(P))^2\big] \\
&\leq \mathbb{E}\big[(\widehat{\mu} - \mu(P))^2 \mid \widetilde{\mu} \neq \perp \wedge A\big] + \mathbb{E}\big[\mathbb{I}[\widetilde{\mu} \neq \perp] \cdot \mathbb{I}[B] \cdot (\widehat{\mu} - \mu(P))^2\big] \\
&= \mathbb{E}\big[(\widehat{\mu} - \mu(P))^2 \mid A\big] + \mathbb{E}\big[\mathbb{I}[B] \cdot (\widehat{\mu} - \mu(P))^2\big]. \quad \text{(C.1)}
\end{aligned}$$

If A holds (i.e., $\widetilde{\mu} \in [\mu(P) - \sigma, \mu(P) + \sigma]$), then $\mu(P) \in [\widetilde{\mu} - \sigma, \widetilde{\mu} + \sigma]$, so we can bound the

first term of the last line in Inequality C.1 as follows.

$$
\begin{aligned}
\mathbb{E}\big[(\widehat{\mu}-\mu(P))^2 \mid A\big] &= \mathbb{E}\left[\left(\frac{1}{n_2}\sum_{i=n_1+1}^{n_1+n_2}\mathsf{clip}_{[\widetilde{\mu}-c,\widetilde{\mu}+c]}(X_i)+\mathsf{Lap}\left(\frac{2c}{n_2\varepsilon}\right)-\mu(P)\right)^2 \mid A\right] \\
&\overset{(A)}{=} \mathbb{E}\left[\left(\frac{1}{n_2}\sum_{i=n_1+1}^{n_1+n_2}\mathsf{clip}_{[\widetilde{\mu}-c,\widetilde{\mu}+c]}(X_i)-\mu(P)\right)^2 \mid A\right]+2\left(\frac{2c}{n_2\varepsilon}\right)^2 \\
&\overset{(B)}{\leq} \mathbb{E}\left[\frac{\mathbb{E}_{X\leftarrow P}[(X-\mu(P))^2]}{n_2} \mid A\right] \\
&\qquad + \mathbb{E}\left[\left(\frac{\mathbb{E}_{X\leftarrow P}\big[|X-\mu(P)|^\lambda\big]}{\lambda\cdot(\min\{\mu(P)-(\widetilde{\mu}-c),(\widetilde{\mu}+c)-\mu(P)\})^{\lambda-1}}\right)^2 \mid A\right]+\frac{8c^2}{n_2^2\varepsilon^2} \\
&\leq \frac{\mathbb{E}_{X\leftarrow P}[(X-\mu(P))^2]}{n_2}+\left(\frac{1}{\lambda}\cdot\frac{\mathbb{E}_{X\leftarrow P}\big[|X-\mu(P)|^\lambda\big]}{(c-\sigma)^{\lambda-1}}\right)^2+\frac{8c^2}{n_2^2\varepsilon^2} \\
&\leq \frac{1}{n_2}+\frac{\psi^{2\lambda}}{\lambda^2\cdot(c-\sigma)^{2(\lambda-1)}}+\frac{8c^2}{n_2^2\varepsilon^2} \\
&\overset{(C)}{=} \frac{1}{n_2}+\frac{\psi^{2\lambda}}{\lambda^2\cdot\psi^{2(\lambda-1)}\cdot(n_2\varepsilon)^{2-2/\lambda}}+\frac{8c^2}{n_2^2\varepsilon^2} \\
&= \frac{1}{n_2}+\frac{8c^2+\psi^2\cdot(n_2\varepsilon)^{2/\lambda}\cdot\lambda^{-2}}{n_2^2\varepsilon^2} \\
&\leq \frac{1}{n_2}+\frac{8c^2+\psi^2\cdot(n_2\varepsilon)^{2/\lambda}}{n_2^2\varepsilon^2}.
\end{aligned}
$$

In the above: Equality A follows from the fact that the Laplace noise is independent from everything else. Inequality B follows from Lemma B.1.2 and linearity of expectations; and Equality C follows from the setting of $c=\sigma+\psi\cdot(n_2\varepsilon)^{1/\lambda}$.
Next, we bound the second term in the last line of Inequality C.1. We use the fact that

$$\forall x\in\mathbb{R}\quad |\mathsf{clip}_{[\widetilde{\mu}-c,\widetilde{\mu}+c]}(x)-\mu(P)|\leq|\widetilde{\mu}-\mu(P)|+c.$$

We have

$$\begin{aligned}
\mathbb{E}\big[\mathbb{I}[B]\cdot(\widehat{\mu}-\mu(P))^2\big] &= \mathbb{E}\left[\mathbb{I}[B]\cdot\left(\frac{1}{n_2}\sum_{i=n_1+1}^{n_1+n_2}\mathsf{clip}_{[\widetilde{\mu}-c,\widetilde{\mu}+c]}(X_i)+\mathsf{Lap}\left(\frac{2c}{n_2\varepsilon}\right)-\mu(P)\right)^2\right]\\
&\overset{\text{(D)}}{=} \mathbb{E}\left[\mathbb{I}[B]\cdot\left(\frac{1}{n_2}\sum_{i=n_1+1}^{n_1+n_2}\mathsf{clip}_{[\widetilde{\mu}-c,\widetilde{\mu}+c]}(X_i)-\mu(P)\right)^2\right]+2\left(\frac{2c}{n_2\varepsilon}\right)^2\cdot\mathbb{P}[B]\\
&\le \mathbb{E}\big[\mathbb{I}[B]\cdot(|\widetilde{\mu}-\mu(P)|+c)^2\big]+\frac{8c^2}{n_2^2\varepsilon^2}\cdot\mathbb{P}[B]\\
&\overset{\text{(E)}}{\le} \mathbb{E}\Big[\mathbb{I}[B]^{\frac{\lambda}{\lambda-2}}\Big]^{\frac{\lambda-2}{\lambda}}\cdot\mathbb{E}\Big[\mathbb{I}[\widetilde{\mu}\neq\perp]\cdot(|\widetilde{\mu}-\mu(P)|+c)^{\lambda}\Big]^{2/\lambda}+\frac{8c^2}{n_2^2\varepsilon^2}\cdot\mathbb{P}[B]\\
&\overset{\text{(F)}}{\le} \mathbb{E}[\mathbb{I}[B]]^{\frac{\lambda-2}{\lambda}}\cdot\mathbb{E}\big[\mathbb{I}[\widetilde{\mu}\neq\perp]\cdot\big(|\widetilde{\mu}-\mu(P)|^{\lambda}+c^{\lambda}\big)\cdot 2^{\lambda-1}\big]^{2/\lambda}+\frac{8c^2}{n_2^2\varepsilon^2}\cdot\mathbb{P}[B]\\
&\overset{\text{(G)}}{\le} \gamma^{\frac{\lambda-2}{\lambda}}\cdot\left(\frac{1}{2}+n_1\cdot 2^{\lambda-1}\cdot\psi^{\lambda}+c^{\lambda}\right)^{2/\lambda}\cdot 2^{2-2/\lambda}+\frac{8c^2}{n_2^2\varepsilon^2}\cdot\gamma\\
&\overset{\text{(H)}}{\le} \gamma^{\frac{\lambda-2}{\lambda}}\cdot\left(2^{-2/\lambda}+n_1^{2/\lambda}\cdot 2^{2-2/\lambda}\cdot\psi^2+c^2\right)\cdot 2^{2-2/\lambda}+\frac{8c^2}{n_2^2\varepsilon^2}\cdot\gamma\\
&\le 4\gamma^{\frac{\lambda-2}{\lambda}}\left(1+4n_1^{2/\lambda}\psi^2+c^2\right)+\frac{8c^2}{n_2^2\varepsilon^2}\cdot\gamma.
\end{aligned}$$

In the above: Equality D follows from the independence of the Laplace noise; Inequality E follows from Hölder's inequality; Inequality F holds because $\forall p\ge 1\ \forall x,y\ge 0\ \ (x+y)^p\le (x^p+y^p)\cdot 2^{p-1}$; Inequality G follows from Proposition 4.1.3; and Inequality H holds because $\forall p\in(0,1]\ \forall x,y\ge 0\ \ (x+y)^p\le x^p+y^p$.

Finally, we can combine all the pieces, and use our parameter settings $\gamma=\delta^2\le 1$ and $c^2=(10+\psi\cdot(n_2\varepsilon)^{1/\lambda})^2\le 2\psi^2(n_2\varepsilon)^{2/\lambda}+200$, to get the following.

$$\begin{aligned}
\mathbb{E}\big[(\widehat{\mu}-\mu(P))^2\big] &\le \gamma\left(\frac{\mu(P)^2+1}{n_2\delta}\right)+\left(\frac{1}{n_2}+\frac{8c^2+\psi^2(n_2\varepsilon)^{2/\lambda}}{n_2^2\varepsilon^2}\right)+\left(4\gamma^{\frac{\lambda-2}{\lambda}}\left(1+4n_1^{2/\lambda}\psi^2+c^2\right)+\frac{8c^2}{n_2^2\varepsilon^2}\cdot\gamma\right)\\
&\le \frac{1}{n_2}+\frac{16c^2+\psi^2\cdot(n_2\varepsilon)^{2/\lambda}}{n_2^2\varepsilon^2}+\delta\cdot\frac{1+\mu(P)^2}{n_2}+\delta^{2-4/\lambda}\cdot 4\left(4n_1^{2/\lambda}\cdot\psi^2+c^2+1\right)\\
&\le \frac{1}{n_2}+\frac{33\psi^2(n_2\varepsilon)^{2/\lambda}+3200}{n_2^2\varepsilon^2}+\delta\cdot\frac{1+\mu(P)^2}{n_2}+\delta^{2-4/\lambda}\cdot\left(16\psi^2n_1^{2/\lambda}+8\psi^2(n_2\varepsilon)^{2/\lambda}+804\right)\\
&= \frac{1}{n_2}+O\left(\frac{\psi^2}{(n_2\varepsilon)^{2-2/\lambda}}+\delta\cdot\frac{\mu(P)^2}{n_2}+\delta^{2-4/\lambda}\cdot(n_1+n_2\varepsilon)^{2/\lambda}\cdot\psi^2\right).
\end{aligned}$$

□

Appendix D

Proofs for Chapter 5

D.1 Background on Complex Analysis

The primary objects of interest in complex analysis are the *holomorphic* functions in the complex plane, namely those functions $f : U \to \mathbb{C}$ that are differentiable at every point $z \in U$. Many familiar functions, such as the polynomials, are in fact holomorphic or may be extended to a holomorphic function. Note that when $U = \mathbb{C}$, i.e., f is differentiable on the whole complex plane, we say that f is an *entire* function.

A basic result of complex analysis asserts that a function $f : U \to \mathbb{C}$ is holomorphic exactly when it is *analytic*, i.e., its Taylor series expansion around any point $z_0 \in U$ converges to f in some neighborhood of z_0. For this reason, holomorphic functions are typically referred to as analytic functions. We consider analyticity in our work as there exist useful mathematical tools to check when functions are analytic, and even more useful tools for constraining functions that we have established to be analytic.

For our purposes, we define a *closed contour* in a region $D \subseteq \mathbb{C}$ to be a continuously differentiable map $\gamma : [0, 1] \to D$ with $\gamma(0) = \gamma(1)$. Informally, we say that a region in the plane is *simply connected* if it contains no holes. For instance, the disk $\{z \in \mathbb{C} : |z| \leq 3\}$ is simply connected, whereas the "donut" $\{z \in \mathbb{C} : |z| \in [1, 3]\}$ is not.

A thorough review of the language of complex analysis with the precise definitions of the above (which are not necessary for the understanding of our application) is outside the scope of this work, so we recommend the textbook by Ahlfors (1953) for a more comprehensive background.

A useful property of analytic functions is that their closed contour integrals vanish in simply connected regions. The following theorem characterises this more formally.

Theorem D.1.1 (Cauchy's Theorem)**.** *Let U be an open, simply connected subset of $\mathbb{C}$ and let $f : U \to \mathbb{C}$ be analytic. Then, for any closed contour γ in U, we have $\oint_\gamma f(z)\,dz = 0$.*

The converse is true, as well, and is a convenient technique for establishing analyticity.

Theorem D.1.2 (Morera's Theorem)**.** *Let $U \subseteq \mathbb{C}$ be open and let $f : U \to \mathbb{C}$ be continuous. Suppose that, for all simply connected $D \subseteq U$ and any closed contour γ in D, we have $\oint_\gamma f(z)\,dz = 0$. Then f is analytic.*

Next, for functions $f_1, f_2 : U \to \mathbb{C}$ and any $L \subseteq U$, we write $f_1|L \equiv f_2|L$, if for all $x \in L$, $f_1(x) = f_2(x)$. Additionally, we write $f_1 \equiv f_2$, if $f_1|U \equiv f_2|U$. Finally, we define the *limit points* of a set.

Definition D.1.3 (Limit Point of a Set)**.** Given a topological space $\mathcal{X}$ and $S \subseteq \mathcal{X}$, we say that $x \in \mathcal{X}$ is a limit point of S, if for every neighbourhood $B \subseteq \mathcal{X}$ of x (with respect to the topology of $\mathcal{X}$), there exists a point $y \in B$, such that $y \in S$ and $y \neq x$.

In other words, a limit point x of S can be "approximated by points in S." The main property of analytic functions that we exploit is the fact that any two analytic functions that agree locally must, in fact, agree globally, as we show next.

Theorem D.1.4 (Identity Theorem)**.** *Let $U \subseteq \mathbb{C}$ be open, and $f_1, f_2 : U \to \mathbb{C}$ be analytic. Suppose there is a set $L \subseteq U$ with a limit point in U, such that $f_1|L \equiv f_2|L$. Then $f_1 \equiv f_2$.*

D.2 Background on Measure Theory

Recall that a *measure space* is the combination of a set $\mathcal{X}$ with a collection Σ of subsets of $\mathcal{X}$, which are closed under complement and countable unions, as well as a function $\mu : \Sigma \to [0, \infty]$ satisfying $\mu(\emptyset) = 0$ and $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$ for disjoint $A_1, A_2, \cdots \in \Sigma$. The subsets making up Σ are called the *measurable subsets* of $\mathcal{X}$ and μ is called a *measure* on $\mathcal{X}$. We say that $\mathcal{X}$ is *σ-finite* when it can be decomposed as $\mathcal{X} = \bigcup_{i=1}^{\infty} A_i$ where $A_1, A_2, \cdots \in \Sigma$ are all of finite measure $\mu(A_i) < \infty$. A function $f : \mathcal{X} \to \mathbb{C}$ is said to be *measurable* if $f^{-1}(U)$ is a measurable subset of $\mathcal{X}$ for any open $U \subseteq \mathbb{C}$. In this case, we say

that $f : \mathcal{X} \to \mathbb{R}$ is *μ-integrable* if $\int_{\mathcal{X}} |f| \, d\mu$, the Lebesgue integral of $|f|$ with respect to μ, exists and is finite.

Now, in order to apply Morera's theorem, we will require some standard integral-limit interchange theorems. The first is the dominated convergence theorem, which asserts that pointwise convergence of a sequence of functions may be interchanged with integration, provided that the sequence is uniformly bounded by an integrable function.

Theorem D.2.1 (Dominated Convergence Theorem)**.** *Let $\mathcal{X}$ be a measure space. Suppose that $(f_n)_{n \in \mathbb{N}}$ is a sequence of measurable functions $\mathcal{X} \to \mathbb{C}$ converging pointwise to some f, i.e., $f_n(x) \to f(x)$ for all $x \in \mathcal{X}$ as $n \to \infty$. Suppose further that there is some measurable $G : \mathcal{X} \to [0, \infty)$ such that $\int_{\mathcal{X}} G \, d\mu < \infty$ and $|f_n(x)| \leq G(x)$ for all $x \in \mathcal{X}$ and $n \in \mathbb{N}$. Then f is integrable such that*

$$\lim_{n \to \infty} \int_{\mathcal{X}} f_n \, d\mu = \int_{\mathcal{X}} f \, d\mu.$$

Switching the order of integration is a very useful operation that is permitted under fairly general measure-theoretic conditions. We describe it as follows.

Theorem D.2.2 (Fubini's Theorem)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be σ-finite measure spaces and suppose that $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is measurable such that*

$$\int_{\mathcal{X}} \int_{\mathcal{Y}} |f(x, y)| \, dy \, dx < \infty.$$

Then

$$\int_{\mathcal{X}} \int_{\mathcal{Y}} f(x, y) \, dy \, dx = \int_{\mathcal{Y}} \int_{\mathcal{X}} f(x, y) \, dx \, dy.$$

D.3 The Expectation of an Estimator on an Exponential Family is Analytic

In this section we prove the following result, which yields Proposition 5.1.1.

Theorem D.3.1 (Analyticity under Exponential Families)**.** *Let $\{P_\eta : \eta \in U\}$ be an exponential family on $\mathbb{R}^n$ in canonical form (recall Definition 5.0.1) and let $\phi : \mathcal{X}^n \to \mathbb{R}$ be any well-defined estimator for $\{P_\eta : \eta \in U\}$, i.e., $\mathbb{E}_{X \leftarrow P_\eta}[|\phi(X)|]$ is finite for all $\eta \in U$. Then $g : U \to \mathbb{R}$ defined by $g(\eta) \coloneqq \mathbb{E}_{X \leftarrow P_\eta}[\phi(X)]$ is an analytic function.*

Before delving into the proof of Theorem D.3.1, we need to show, for the sake of completeness, that the product distribution where each marginal has the same distribution from an exponential family is also an exponential family.

Proposition D.3.2. *Let $\{P_\eta : \eta \in U\}$ be an exponential family with support $D \subseteq \mathbb{R}$. Then for any $n \in \mathbb{N}$, the family of distributions $\left\{P_\eta^n : \eta \in U\right\}$ is an exponential family over D^n with the same natural parameters as well as carrier measure and sufficient statistic given, respectively, by*

$$h_n(x_1, \ldots, x_n) = \prod_{i=1}^n h(x_i) \quad \text{and} \quad T_n(x_1, \ldots, x_n) = \sum_{i=1}^n T(x_i).$$

Proof. Let $\eta \in U$ with density function $f_{T,h,\eta}$ as described in Definition 5.0.1. Suppose $f : D^n \to \mathbb{R}$ is the density function of P_η^n. Then for any $x = (x_1, \ldots, x_n) \in D^n$, we have the following.

$$\begin{aligned} f(x) &= \prod_{i=1}^n f_{T,h,\eta}(x_i) \\ &= \prod_{i=1}^n h(x_i) \exp(\eta \cdot T(x_i) - Z(\eta)) \\ &= \left(\prod_{i=1}^n h(x_i)\right) \exp\left(\eta \sum_{i=1}^n T(x_i) - nZ(\eta)\right) \end{aligned}$$

This gives us: $T_n(x) = \sum_{i=1}^n T(x_i)$; $h_n(x) = \prod_{i=1}^n h(x_i)$; and the natural parameter of P_η^n being $\eta_n = \eta \in U$. One can easily verify that the log-partition function $(Z_n(\eta_n))$ of P_η^n equals $nZ(\eta)$. □

With this detail out of the way, the main idea behind the proof of Theorem D.3.1 is that analyticity is preserved under integration under certain circumstances, which we show next. Although a proof for the real plane is possible, it will be technically convenient to pass to the complex plane where we can wield Morera's theorem (Theorem D.1.2).

Lemma D.3.3. *Let Ω be a σ-finite measure space with measure ν, let $V \subseteq \mathbb{C}$ be open, and let $f : \Omega \times V \to \mathbb{C}$. Assume that $f(\omega, \eta)$ is analytic in η for every fixed $\omega \in \Omega$ and that, for every compact $K \subseteq V$, there is a ν-integrable function (see Section D.2) $G : \Omega \to [0, \infty)$ for which $|f(\omega, \eta)| \leq G(\omega)$ for all $\eta \in K$. Then $g(\eta) \coloneqq \int_\Omega f(\omega, \eta)\, d\nu(\omega)$ is analytic, as well.*

Proof. Our plan is to apply Morera's theorem (Theorem D.1.2). To that end, we must first show that g is continuous, so let $(\eta_n)_{n\in\mathbb{N}}$ be any sequence with $\eta_n \to \eta$ as $n \to \infty$. By our assumption, there is a ν-integrable $G : \Omega \to [0, \infty)$ such that $|f(\omega, \eta_n)| \leq G(\omega)$ for all $n \in \mathbb{N}$ and $\omega \in \Omega$. So, by the dominated convergence theorem (Theorem D.2.1), $g(\eta_n) \to g(\eta)$ as $n \to \infty$.

Now, let $\gamma : [0, 1] \to \mathbb{C}$ be any closed contour lying in a simply connected (see Section D.1) subset of V, and let γ' denote its first derivative. Then, $|\gamma'|$ must be bounded by some $C > 0$, so

$$\int_0^1 \int_\Omega |f(\omega, \gamma(t))\gamma'(t)|\, d\nu(\omega)\, dt \leq \int_0^1 \int_\Omega G(\omega) C\, d\nu(\omega)\, dt = C \int_\Omega G\, d\nu < \infty$$

and thus Fubini's theorem (Theorem D.2.2) implies that

$$\begin{aligned}
\oint_\gamma g(\eta)\, d\eta &= \int_0^1 \int_\Omega f(\omega, \gamma(t))\gamma'(t)\, d\nu(\omega)\, dt \\
&= \int_\Omega \int_0^1 f(\omega, \gamma(t))\gamma'(t)\, dt\, d\nu(\omega) \\
&= \int_\Omega \oint_\gamma f(\omega, \eta)\, d\eta\, d\nu(\omega) \\
&= \int_\Omega 0\, d\nu(\omega) && \text{(Theorem D.1.1)} \\
&= 0.
\end{aligned}$$

As γ was arbitrary, g must be analytic by Morera's theorem. □

Proof of Theorem D.3.1. Our main goal is to show that $g(\eta) \coloneqq \mathbb{E}_{X \leftarrow P_\eta}[\phi(X)]$ is analytic. To that end, let h, T, and Z be the carrier measure, the sufficient statistic, and the log-partition function of $\{P_\eta : \eta \in U\}$, respectively.

We first show that $\exp(Z(\eta))$ is analytic by way of Lemma D.3.3. Indeed, $r(x,\eta) \coloneqq h(x)\exp(\eta T(x))$ is entire (see Section D.1) in $\eta \in \mathbb{C}$ for each fixed $x \in \mathbb{R}^n$. Let $K \subseteq \mathbb{C}$ be an arbitrary compact set, and let m and M be the minimum and the maximum real coordinates among the points within K, respectively. Then for any $x \in \mathbb{R}^n$ and $\eta \in K$,

$$T(x) < 0 \implies |r(x,\eta)| = h(x)\exp(\mathrm{Re}(\eta)T(x)) \le h(x)\exp(mT(x))$$

and

$$T(x) \ge 0 \implies |r(x,\eta)| \le h(x)\exp(MT(x)),$$

so we have

$$|r(x,\eta)| \le h(x)\exp(mT(x)) + h(x)\exp(MT(x)).$$

But $\int_{\mathbb{R}^n} h(x)\exp(mT(x)) + h(x)\exp(MT(x))\,dx = \exp(Z(m)) + \exp(Z(M)) < \infty$, so, since K was arbitrary, $\exp(Z(\eta)) = \int_{\mathbb{R}^n} r(x,\eta)\,dx$ must be entire by Lemma D.3.3.

As a consequence, $h(x)\exp(\eta T(x) - Z(\eta))$ is analytic in η for every fixed $x \in \mathbb{R}^n$, so we can apply nearly the same argument to $h(x)\exp(\eta T(x) - Z(\eta))$ in order to conclude that

$$g(\eta) = \mathop{\mathbb{E}}_{X \leftarrow P_\eta}[\phi(X)] = \int_\Omega \phi(x)h(x)\exp(\eta T(x) - Z(\eta))\,dx$$

is analytic, as well. □

D.4 Proof of Proposition 5.2.1

Recall that $\mathbb{E}[Y] = \int_0^\infty \mathbb{P}[Y \ge t]\,dt$ for any non-negative random variable Y. For any $x \in \mathcal{X}^n$, we have

$$\left|\mathop{\mathbb{E}}_A[A(x)]\right| \overset{\text{(a)}}{\le} \mathop{\mathbb{E}}_A[|A(x)|] = \int_0^\infty \mathop{\mathbb{P}}_A[|A(x)| \ge t]\,dt \overset{\text{(b)}}{\le} \exp(\varepsilon n)\int_0^\infty \mathop{\mathbb{P}}_A[|A(x^*)| \ge t]\,dt = \exp(\varepsilon n)\mathop{\mathbb{E}}_A[|A(x^*)|],$$

where inequalities (a) and (b) follow from Jensen's inequality and group privacy (Lemma C.1.1), respectively.