Advancements in Gaussian and Local Differential Privacy

by

Yi Liu

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Statistical Machine Learning

Department of Mathematical and Statistical Sciences
University of Alberta

Abstract

This thesis presents a comprehensive study of Gaussian Differential Privacy (GDP) and Local Differential Privacy (LDP), exploring their properties, relationships, and applications in developing novel algorithms and optimization methods for efficient and accurate privacy-preserving data analysis. In the first paper, we examine asymptotic properties of privacy profiles, develop a criterion for identifying GDP algorithms, propose an efficient method for narrowing down optimal privacy measurement values, and introduce a post-processing procedure for non-GDP algorithms. We also compare single-parameter privacy notions and demonstrate the advantages of our measurement process and the composition theorem of GDP. The second paper focuses on estimating population quantiles under LDP using binary inquiries, developing a self-normalizing, online algorithm with valid inference and optimality results for median estimation. The third paper introduces a novel algorithm for estimating Cumulative Distribution Function (CDF) curves under LDP by combining constrained isotonic estimation and binary inquiries, uncovering an unexpected connection to the current status problem in survival data analysis. We establish error bounds and computational efficiency for our estimator. Collectively, these papers contribute to the understanding and development of efficient, privacy-preserving mechanisms in GDP and LDP, providing valuable insights and practical tools for data analysts and privacy researchers, and advancing the state of the art in differential privacy research.

Preface

The research conducted for this thesis is part of a series of collaborative projects under the supervision of Professor Kong, Linglong.

Chapter 2, titled ”Yi Liu, Ke Sun, Bei Jiang, & Linglong Kong (2022). Identification, Amplification and Measurement: A Bridge to Gaussian Differential Privacy,” has been published in Advances in Neural Information Processing Systems (Vol. 35, pp. 11410-11422). My contributions to this paper include the development of the general idea, writing, coding, numerical experiments, and theorem proofs. Ke Sun assisted with the paper's formatting, contributed to discussions during the review stages, and created the presentation poster.

Chapter 3, titled ”Yi Liu, Qirui Hu, Lei Ding, Bei Jiang, & Linglong Kong. (2023). Online Local Differential Private Quantile Inference via Self-normalization,” has been accepted by the International Conference on Machine Learning (2023). My involvement in this paper consists of writing the majority of the content, developing the binary inquiry approach, coding, numerical experiments, and some theorem proofs. Qirui Hu helped in the derivation of the weak convergence procedure, while Lei Ding provided proofreading support and offered suggestions for improving the paper's presentation to a broader audience.

Chapter 4, which was submitted to Neural Information Processing Systems 2023, is titled ”Yi Liu, Qirui Hu, & Linglong Kong (2023). Efficient CDF Estimation under Local Differential Privacy: A Constrained Isotonic Approach.” For this paper, I developed the concept of current-status style estimation and the design to separate LDP and statistical analysis. Additionally, I completed all coding and numerical experiments and wrote the majority of the paper. Qirui Hu aided in the modifications necessary to adapt isotonic regression to more lenient conditions.

Acknowledgements

I want to start by thanking my supervisor, Dr. Linglong Kong, for his continuous guidance, support, and motivation during my research. His expertise, insightful advice, and feedback have been instrumental in shaping my thesis and ensuring its completion. Dr. Kong has been an extraordinary mentor, inspiring me to strive for academic excellence.

I'm also appreciative of my collaborators, Qirui Hu and Lei Ding, for their crucial input to this project. Their knowledge, dedication, and innovative viewpoints have enriched our methodology and analysis. Working alongside them has been both an enriching and enjoyable experience.

A special thanks to Dr. Bei Jiang from the University of Alberta for her assistance in navigating through various scholarship applications. This has had a significant impact on my academic journey.

I would also like to acknowledge Yangdi Jiang and Xiaotian Chang for their valuable advice and constructive criticism. Their keen observations and willingness to share their knowledge have greatly contributed to the improvement of my work.

Moreover, my heartfelt appreciation goes to the faculty members, staff, and fellow students at the university for their unwavering support and fostering a nurturing academic environment that has made my experience here unforgettable.

Lastly, my appreciation extends beyond the walls of academia, towards the valiant soldiers fighting bravely against invasions. Their indomitable spirit, selfless courage, and relentless dedication serve as a continuous source of inspiration and a stark reminder of the tenacity and perseverance.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In today's digital age, modern AI systems and big models are becoming increasingly ubiquitous, offering powerful tools for data analysis. However, the potential dangers they pose to users' privacy have emerged as a major concern. This thesis presents a comprehensive study of Gaussian Differential Privacy (GDP) and Local Differential Privacy (LDP) by exploring their properties, relationships, and applications in developing novel algorithms and optimization methods for efficient and accurate privacy-preserving data analysis.

GDP is a single-parameter family of privacy notions that provides coherent guarantees to protect sensitive individual information. Although GDP offers greater interpretability and tighter bounds under composition, many mechanisms (e.g., the Laplace mechanism) inherently provide GDP guarantees but often fail to leverage this new framework due to their privacy guarantees being derived under a different context. In the first paper, we begin by examining the asymptotic properties of privacy profiles to develop a simple criterion for identifying algorithms with GDP properties. We propose an efficient method for narrowing down possible values of an optimal privacy measurement, μ, with an arbitrarily small and quantifiable margin of error. For non-GDP algorithms, we introduce a post-processing procedure that can amplify existing privacy guarantees to meet the GDP condition. Furthermore, we show that all pure-DP algorithms are intrinsically also GDP. We demonstrate that the combination of our measurement process and the composition theorem of GDP is a powerful and convenient tool for handling compositions compared to traditional standard and advanced composition theorems. Utilizing the procedures mentioned above, most existing DP algorithms can be incorporated into the GDP framework with minimal modifications.

In the usual DP setting, a trusted curator can acquire the actual sample quantiles and other summary statistics, with the only limitation being that the release of the output must conform to the DP condition, but the use of trusted curators undermines the spirit of the solid cryptographic level of privacy protection that DP provides. The concept of LDP was invented to address this issue as no such curator is needed under LDP. The second paper focuses on developing an algorithm for estimating population quantiles under LDP, based on binary inquiries. Our proposed method is self-normalizing and provides asymptotically normal estimation with valid inference, resulting in tight confidence intervals without the need for estimating nuisance parameters. The algorithm can be conducted fully online, leading to high computational efficiency and minimal storage requirements with $\mathcal{O}(1)$ space. Additionally, we prove an optimality result for median estimation through an elegant application of one central limit theorem of GDP.

Subsequently, in the third paper, we introduce a novel algorithm for estimating Cumulative Distribution Function (CDF) curves under LDP by utilizing a combination of constrained isotonic estimation and binary inquiries. These algorithms can be employed to estimate the distribution of sensitive attributes of people, such as income or drug use. We uncover an unexpected connection between LDP and the current status problem, a classical survival data problem in statistics. Through mathematical proofs and extensive numerical testing, we demonstrate that our method achieves a uniform error bound of $O(n^{-1/3} \log n)$ and an L_2 error bound of $O(n^{-1/3})$ when estimating the entire CDF curves. By concentrating on a finite grid, the error bound can be improved to $O(n^{-1/2})$, with an asymptotic normal distribution of error. Theoretically, we have shown that the error bound smoothly changes from $O(n^{-1/2})$ to $O(n^{-1/3})$ as the number of grids increases relative to the sample size n. Computationally, we demonstrate that our constrained isotonic estimator can be efficiently computed in a deterministic manner, without the need for any hyperparameters or random optimization.

Overall, these papers contribute to the understanding and development of efficient, privacy-preserving mechanisms in the fields of GDP and LDP. They provide valuable insights and practical tools for data analysts and privacy researchers. By addressing the challenges of privacy preservation and developing novel algorithms, this thesis advances the state of the art in differential privacy research.

Chapter 2

Paper 1: Identification, Amplification and Measurement: A bridge to Gaussian Differential Privacy

2.1 Abstract

Gaussian differential privacy (GDP) is a single-parameter family of privacy notions that provides coherent guarantees to avoid the exposure of sensitive individual information. Despite the extra interpretability and tighter bounds under composition GDP provides, many widely used mechanisms (e.g., the Laplace mechanism) inherently provide GDP guarantees but often fail to take advantage of this new framework because their privacy guarantees were derived under a different background. In this paper, we study the asymptotic properties of privacy profiles and develop a simple criterion to identify algorithms with GDP properties. We propose an efficient method for GDP algorithms to narrow down possible values of an optimal privacy measurement, μ, with an arbitrarily small and quantifiable margin of error. For non-GDP algorithms, we provide a post-processing procedure that can amplify existing privacy guarantees to meet the GDP condition. As applications, we compare two single-parameter families of privacy notions, ϵ-DP, and μ-GDP, and show that all ϵ-DP algorithms are intrinsically also GDP. Lastly, we show that the combination of our measurement process and the

composition theorem of GDP is a powerful and convenient tool to handle compositions compared to the traditional standard and advanced composition theorems.

2.2 Introduction

Recent years have seen explosive growth in the research and application of data-driven machine learning. While data fuels advancement in this unprecedented age of "big data", concern for individual privacy has deepened with the continued mining, transportation, and exchange of this new resource. While expressions of privacy concerns can be traced back as early as 1969 [1], the concept of privacy is often perceived as "vague and difficult to get into a right perspective" [2]. Through its alluring convenience and promise of societal prosperity, the use of aggregated data has long outstripped the capabilities of privacy protection measures. Indeed, early privacy protection protocols relied on the ad hoc enforcement of anonymization and offered little to no protection against the exposure of individual data, as evidenced by the AOL search log and Netflix Challenge dataset controversies [3–5].

Differential privacy (DP) first gained traction as it met the urgent need for rigour and quantifiability in privacy protection [6]. In short, DP bounds the change in the distribution of outputs of a query made on a dataset under an alteration of one data point. The following definition formalizes this notion.

Definition 2.1 *[6] A randomized algorithm $\mathcal{A}$, taking a dataset consisting of individuals as its input, is (ϵ, δ)-differentially private if, for any pair of datasets S and S' that differ in the record of a single individual and any event E,*

$$P[\mathcal{A}(S) \in E] \leq e^{\epsilon} P\left[\mathcal{A}\left(S'\right) \in E\right] + \delta.$$

When $\delta = 0$, $\mathcal{A}$ is called ϵ-differentially private (ϵ-DP).

While the notion of (ϵ, δ)-DP has wide applications [7–10], there are a few notable drawbacks to this framework. One is the poor interpretability of (ϵ, δ)-DP: unlike other concepts in machine learning, DP should not remain a black box. Privacy guarantees are intended for human interpretation and so must be understandable by the users it affects and by regulatory entities. A second drawback is (ϵ, δ)-DP's inferior composition properties and lack of versatility. Here, "composition" refers to the ability for DP properties to be inherited when DP algorithms are combined and used as building blocks. As an example, the training of deep learning models involves gradient evaluations and weight updates: each of these steps can be treated as a building block. It is natural to expect that a DP learning algorithm can be built using differentially-private versions of these components. However, the DP composition properties cannot generally be well characterized within the framework of (ϵ, δ)-DP, leading to very loose composition theorems.

To overcome the drawbacks of (ϵ, δ)-DP, numerous variants have been developed, including the hypothesis-testing-based f-DP [11, 12], the moments-accountant-based Rényi DP [13], as well as concentrated DP and its variants [14, 15]. Despite their very different perspectives, all of these DP variants can be fully characterized by an infinite union of (ϵ, δ)-DP guarantees. In particular, there is a two-way embedding between f-DP and the infinite union of (ϵ, δ)-DP guarantees: any guarantee provided by an infinite union of (ϵ, δ)-DP can be fully characterized by f-DP and vice visa [12]. Consequently, f-DP has the versatility to treat all of the above notions as special cases.

In addition to its versatility, f-DP is more interpretable than other DP paradigms because it considers privacy protection from an attacker's perspective. Under f-DP, an attacker is challenged with the hypothesis-testing problem

$$H_0: \text{ the underlying dataset is } S \qquad \text{versus} \qquad H_1: \text{ the underlying dataset is } S'$$

and given output of an algorithm $\mathcal{A}$, where S and S' are neighbouring datasets. The harder this testing problem is, the less privacy leakage $\mathcal{A}$ has. To see this, consider the dilemma that the attacker is facing. The attacker must reject either H_0 or H_1

based on the given output of $\mathcal{A}$: this means the attacker must select a subset R_0 of Range($\mathcal{A}$) and reject H_0 if the sampled output is in R_0 (or must otherwise reject H_1). The attacker is more likely to incorrectly reject H_0 (in a type I error) when R_0 is large. Conversely, if R_0 is small, the attacker is more likely to incorrectly reject H_1 (in a type II error). We say that an algorithm $\mathcal{A}$ is f-DP if, for any $\alpha \in [0, 1]$, no attacker can simultaneously bound the probability of type I error below α and bound the probability of type II error below $f(\alpha)$. Such f is called a trade-off function and controls the strength of the privacy protection.

The versatility afforded by f can be unwieldy in practice. Although f-DP is capable of handling composition and can embed other notions of differential privacy, it is not convenient for representing safety levels as a curve amenable to human interpretation. Gaussian differential privacy (GDP), as a parametric family of f-DP guarantees, provides a balance between interpretability and versatility. GDP guarantees are parameterized by a single value μ and use the trade-off function $f(\alpha) = \Phi\left(\Phi^{-1}(1 - \alpha) - \mu\right)$, where Φ is the cumulative distribution function of the standard normal distribution. With this choice of f, the hypothesis-testing problem faced by the attacker is as hard as distinguishing between $N(0, 1)$ and $N(\mu, 1)$ on the basis of a single observation. Aside from its visual interpretation, GDP also has unique composition theorems: the composition of a μ_1- and μ_2-GDP algorithm is, as expected, $\sqrt{\mu_1^2 + \mu_2^2}$-GDP. This property can be easily generalized to n-fold composition. GDP also has a special central limit theorem implying that all hypothesis-testing-based definitions of privacy converge to GDP in terms of a limit in the number of compositions. Readers are referred to [12] for more information.

2.2.1 Outline

The goal of this paper is to provide a bridge between GDP and algorithms developed under other DP frameworks. We start by presenting an often-overlooked partial order on (ϵ, δ)-DP conditions induced by logical implication. Ignoring this partial order will lead to problematic asymptotic analysis.

We then break down GDP into two parts: a head condition and a tail condition. We show that the latter, through a single limit of a mechanism's privacy profile, is sufficient to distinguish between GDP and non-GDP algorithms. For GDP algorithms, this criterion also provides a lower bound for the privacy protection parameter μ and can help researchers widen the set of available GDP algorithms. This criterion furthermore gives an interesting characterization of GDP without an explicit reference to the Gaussian distribution.

The next logical step is to measure the exact privacy performance. Interestingly, while the binary "GDP or not" question can be answered solely by the tail, the actual performance of a DP algorithm is determined by the head. We define and apply the Gaussian Differential Privacy Transformation (GDPT) to narrow the set of potential optimal values of μ with an arbitrarily small and quantifiable margin of error. We further provide procedure to adapt an algorithm to GDP or improve the privacy parameter when results from the GDP identification and measurement procedures are undesirable.

Lastly, we demonstrate additional applications of our newly developed tools. We first make a comparison between DP and GDP and show that any ϵ-DP algorithm is automatically GDP. We then show that the combination of our measurement process and the GDP composition theorem is a more powerful and convenient tool for handling compositions relative to traditional composition theorems.

2.3 Privacy profiles and an exact partial order on (ϵ, δ)-DP conditions

The benefits of DP come with a price. As outlined in the definition of DP, any DP algorithm must be randomized. This randomization is usually achieved by perturbing the intermediate step or the final output via the injection of random noise. Because of the noise, a DP algorithm cannot faithfully output the truth like its non-DP counterpart.

To provide a higher level of privacy protection, a stronger utility compromise should be made. This leads to the paramount problem of the "privacy–utility trade-off". Under the (ϵ, δ)-DP framework, this trade-off is often characterized in the form of $\sigma = f(\epsilon, \delta)$: to achieve (ϵ, δ)-DP, the utility parameter (usually the scale of noise) needs to be chosen as $f(\epsilon, \delta)$. Therefore, an algorithm can be (ϵ, δ)-DP for multiple pairs of ϵ and δ: the union of all such pairs provides a complete image of the algorithm under the (ϵ, δ)-DP framework. In particular, an (ϵ, δ)-DP mechanism $\mathcal{A}$ is also (ϵ', δ')-DP for any $\epsilon' \geq \epsilon$ and any $\delta' \geq \delta$. The infinite union of (ϵ, δ) pairs can thus be represented as the smallest δ associated with each ϵ. This intuition is formulated as a privacy profile in [16]. The privacy profile corresponding to a collection of (ϵ, δ)-DP guarantees Ω is defined as the curve in $[0, \infty) \times [0, 1]$ separating the space of privacy parameters into two regions, one of which contains exactly the pairs in Ω. The privacy profile provides as much information as Ω itself. Many privacy guarantees and privacy notions, including (ϵ, δ)-DP, Rényi DP, f-DP, GDP, and concentrated DP, can be embedded into a family of privacy profile curves and fully characterized [17]. A privacy profile can be provided or derived by an algorithm's designer or users.

Before proceeding with detailed discussions, we first give three examples of DP algorithms that are used throughout the paper. The first example we consider is the Laplace mechanism, a classical DP mechanism whose prototype is discussed in the paper that originally defined the concept of differential privacy [6]. The level of privacy that the Laplace mechanism can provide is determined by the scale b of the added Laplacian noise. Given a global sensitivity Δ, the value of b needs to be chosen as $f(\epsilon, 0) = \Delta/\epsilon$ in order to provide an $(\epsilon, 0)$-DP guarantee. Despite its long history, the Laplace mechanism has remained in use and study in recent years [18–21]. Our second example is a family of algorithms in which a noise parameter has the form $\sigma = A\epsilon^{-1}\sqrt{\log(B/\delta)}$. Examples include: the goodness of fit algorithm [22], noisy stochastic gradient descent and its variants [23–25] and the one-shot spectral method and the one-shot Laplace algorithm [26]. Our third example comes from the field of federated learning: given n users and the number of messages m, the invisibility cloak encoder algorithm (ICEA) from [27] is (ϵ, δ)-DP if $m > 10 \log(n/(\epsilon\delta))$ [28]. See also [29, 30] for other analysis of ICEA.

For figures and numerical demonstrations in this paper, we use $b = 2/\Delta$ for the Laplace mechanism; $A = 2$, $B = 1$, and $\sigma = 2$ for the second example, which we refer to as SGD; and $m = 20$ and $n = 4$ for the ICEA. We omit the internal details of these methods and focus on their privacy guarantees: other than for the classical Laplace mechanism, whose privacy profile is known [17], privacy guarantees are given in the form of privacy–utility trade-off equation $\sigma = g(\epsilon, \delta)$. Given σ, it is tempting to derive the privacy profile by inverting g (i.e., as $\delta_{\mathcal{A}}(\epsilon) = \min\{\delta \mid \sigma = g(\epsilon, \delta)\}$) because an (ϵ_0, δ_0)-DP algorithm is trivially (ϵ, δ)-DP for any $\epsilon \geq \epsilon_0$ and $\delta \geq \delta_0$. However, in most cases, a privacy profile naively derived in this way is not tight and will lead to a problematic asymptotic analysis, especially near the origin, because of a frequently overlooked partial order between (ϵ, δ)-DP conditions below.

Theorem 2.2 *Assume that $\epsilon_0 \geq 0$ and $0 \leq \delta_0 < 1$. The (ϵ_0, δ_0)-DP condition implies (ϵ, δ)-DP if and only if $\delta \geq \delta_0 + (1 - \delta_0)(e^{\epsilon_0} - e^{\epsilon})^+/(1 + e^{\epsilon_0})$.*

Theorem 2.2 states the exact partial order of logical implication on (ϵ, δ)-DP conditions. Though not explicitly discussed in this form in previous literature on DP, this partial order can be implicitly derived from other results (e.g. proposition 2.11 of [12]). Taking this partial order into account, the privacy profile derived from the naive inversion of the trade-off function can be refined into

$$\delta_{\mathcal{A}}(\epsilon) = \min\left(\left\{\delta \mid \sigma = g(\epsilon_0, \delta_0) \text{ and } \delta \geq \delta_0 + \frac{(1 - \delta_0)(e^{\epsilon_0} - e^{\epsilon})^+}{1 + e^{\epsilon_0}}\right\}\right).$$

Intuitively, the refined privacy profile not only considers (ϵ, δ)-DP provided directly by the trade-off function but also takes all pairs (ϵ, δ) inferred by corollary 2.2. See figure 2.1 for comparison before and after this refinement.

2.4 The identification of GDP algorithms

We next show the connection between GDP and the privacy profile: briefly, Gaussian differential privacy can be characterized as an infinite union of (ϵ, δ)-DP conditions.

Theorem 2.3 *([Corollary 2.13 [12]) A mechanism is μ-GDP if and only if it is $(\epsilon, \delta_\mu(\epsilon))$-DP for all $\epsilon \geq 0$, where*

$$\delta_\mu(\epsilon) = \Phi\left(-\frac{\epsilon}{\mu} + \frac{\mu}{2}\right) - \mathrm{e}^{\epsilon}\Phi\left(-\frac{\epsilon}{\mu} - \frac{\mu}{2}\right). \tag{2.1}$$

This result follows from properties of f-DP. Prior to this general form, an expression for a special case appeared in [16]. From the definition of the privacy profile, it follows immediately that an algorithm $\mathcal{A}$ with the privacy profile $\delta_{\mathcal{A}}$ is μ-GDP if and only if $\delta_\mu(\epsilon) \geq \delta_{\mathcal{A}}(\epsilon)$ for all non-negative ϵ. However, this observation does not automatically lead to a meaningful way to identify GDP algorithms.

Before proceeding with an analysis of privacy profiles, we give a few visual examples in Figure 2.1. The left side of2.1 illustrates the privacy profiles of our examples. That of the Laplace mechanism is derived in [17] as Theorem 3: given a noise parameter b and a global sensitivity Δ, the privacy profile of the Laplace mechanism is $\delta(\epsilon) = \max(1 - \exp\{\varepsilon/2 - \Delta/(2b)\},\ 0)$. For the second and the third examples, we compare the naive privacy profiles obtained by inverting the trade-off function with the refined privacy profiles. The refined and naive privacy profiles take on notably different values around $\epsilon = 0$. The inverted trade-off functions suggest that $(0, \delta)$ cannot be achieved by any choice of parameter σ. However, this is clearly not true, considering Theorem 2.2.

As shown in the right side of Figure 2.1, the Laplace mechanism's privacy profile is below the 2-GDP and 4-GDP curves but crosses the 1-GDP curve, indicating that the Laplace mechanism, in this case, is 2-GDP and 4-GDP but not 1-GDP. The ICEA curve intersects all of the displayed GDP curves, so the algorithm is not μ-GDP for $\mu \in \{1, 2, 4\}$. It is hard to tell whether or not the SGD curve crosses the 1-GDP curve, and we cannot say if it will cross the 2-GDP or even the 4-GDP curve at a large value of ϵ. These examples illustrate that we cannot draw conclusions simply by looking at a graph. A privacy profile is defined on $[0, \infty)$, so it is hard to tell if inequality is maintained as ϵ increases. Previous failures of ad hoc attempts at privacy have taught that privacy must be protected via tractable and objective means [3–5].

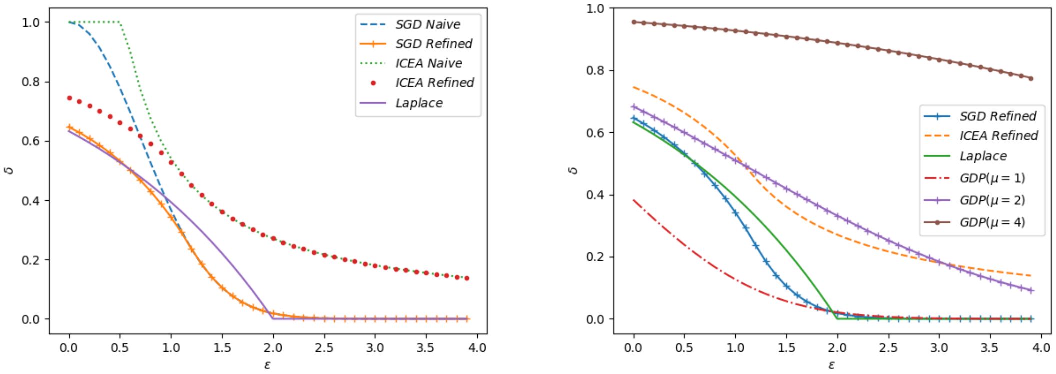


Figure 2.1: Left: Examples of privacy profiles obtained by inverting the trade-off function (naive) and by Theorem 2.2 (refined). Right: Comparison of 1-GDP and 2-GDP privacy profiles against those for our three examples.

Performing this check via numerical evaluation yields similar problems: we cannot consider all values of ϵ on an infinite interval (or even a finite one, for that matter). Turning to closed forms for privacy profiles and δ_μ is also difficult: even if a given privacy profile is easy to handle, δ_μ presents some technical hurdles. The profile δ_μ and Φ are transcendental with different asymptotic behaviors for different values of μ and ϵ. This is clear from the Figure 2.1: near $\epsilon = 0$, δ_μ is concave for $\mu = 4$ but convex for $\mu = 1$. As a further complication, both the first and second terms in the definition of δ_μ converge to 1 as $\epsilon \to \infty$, but the difference between them vanishes. Subtracting good approximations of two nearby numbers may cause a phenomenon called catastrophic cancellation and lead to very bad approximations [31, 32]. Due to the risk of catastrophic cancellation, a good approximation of Φ does not guarantee a good approximation of the GDP privacy profile. These problems make it difficult to tightly bound δ_μ by a function with a simple form.

To address the problem of differing asymptotic behaviours, we define the following two notions.

Definition 2.4 *(Head condition) An algorithm $\mathcal{A}$ with the privacy profile $\delta_\mathcal{A}$ is (ϵ_h, μ)-head GDP if and only if $\delta_\mathcal{A}(\epsilon) \leq \delta_\mu(\epsilon)$ when $\epsilon \leq \epsilon_h$.*

Definition 2.5 *(Tail condition) An algorithm $\mathcal{A}$ with the privacy profile $\delta_{\mathcal{A}}$ is (ϵ_t, μ)-tail GDP if and only if $\delta_{\mathcal{A}}(\epsilon) \leq \delta_\mu(\epsilon)$ when $\epsilon > \epsilon_t$.*

The head condition checks the μ-GDP condition for ϵ near zero and the tail condition checks the μ-GDP condition for ϵ far away from zero. As such, the combination of (ϵ, μ)-head GDP and (ϵ, μ)-tail GDP is equivalent to μ-GDP. For now, we put the exact value of μ aside and consider only the qualitative question of how to identify a GDP algorithm by its privacy profile. The following theorem answers this question.

Theorem 2.6 *An algorithm $\mathcal{A}$ is GDP if and only if $\mathcal{A}$ is (ϵ, μ)-tail GDP for any finite ϵ and μ.*

Interestingly, only the tail condition figures into the identification problem. The reason for this stems from theorem 2.2. Any nontrivial (ϵ, δ)-DP algorithm must be $(0, \delta)$-DP for some $\delta < 1$ and therefore must satisfy a head condition for some sufficiently large μ. The only problem left is the tail. However, it is not possible to check whether $\delta(\epsilon) < \delta_\mu(\epsilon)$ for all values of ϵ. To circumvent this issue, we present a key lemma that underlies much of the theoretical analysis in this section and may continue to be useful in future developments.

Lemma 2.7 *Define $\tilde{\delta}_\mu(\epsilon) = \frac{\mu e^{-a^2/2}}{\sqrt{2\pi a^2}}$, where $a = -\frac{\epsilon}{\mu} + \frac{\mu}{2}$. It follows that $\lim_{\epsilon \to +\infty} \frac{\delta_\mu(\epsilon)}{\tilde{\delta}_\mu(\epsilon)} = 1$.*

Using the key lemma above, a condition for identifying GDP algorithms is simple to formulate:

Theorem 2.8 *Let $\mu_t = \sqrt{\lim_{\epsilon \to +\infty} \frac{\epsilon^2}{-2 \log \delta_{\mathcal{A}}(\epsilon)}}$. An algorithm $\mathcal{A}$ with the privacy profile $\delta_{\mathcal{A}}(\epsilon)$ is μ-GDP if and only if $\mu_t < \infty$. Further, μ is no smaller than μ_t.*

Theorems 2.6 and 2.8 give a useful criterion characterizing GDP and deepen our understanding of GDP. Putting the exact value of μ aside, a GDP algorithm must provide an infinite union of (ϵ, δ)-DP conditions, where δ must be $O(e^{-c\epsilon^2})$ as $\epsilon \to \infty$. Refer to Appendices 2.8.1 for proofs of Theorems.

2.5 The Gaussian differential privacy transformation

While the binary "GDP or not" question can be answered solely by the tail condition, the actual performance of a DP algorithm is determined by the value of its privacy profile for small values of ϵ: intuitively, all (ϵ_t, μ)-tail conditions are weaker than the corresponding ϵ_t-DP condition, and the latter provides almost no privacy when $\epsilon_t > 10$. A more detailed discussion will be presented in 2.5.2. To solve the measurement problem, we first propose a new tool—the Gaussian differential privacy transformation (GDPT).

Definition 2.9 *(GDPT) Let f be a non-increasing, non-negative function defined on $[0, +\infty)$ satisfying $f(0) \leq 1$. The Gaussian differential privacy transformation (GDPT) of f is the function G_f mapping $[0, \infty)$ to $[0, \infty)$ such that $G_f(\epsilon) = \mu_{GDP}(\epsilon, f(\epsilon))$, where $\mu_{GDP}(x, y)$ is the implicit function defined by the equation $\delta_\mu(x) = y$.*

We highlight two critical features of the GDPT.

- The GDPT is order preserving: if $f(\epsilon) \geq g(\epsilon)$, then $G_f(\epsilon) \geq G_g(\epsilon)$.
- The GDPT of δ_μ is $G_{\delta_\mu}(\epsilon) = \mu$, a constant function.

The first of these two features derive from the monotonicity of $\delta_\mu(\epsilon)$. Given a fixed μ, $\delta_\mu(\epsilon)$ is a strictly decreasing continuous function of ϵ. Given a fixed ϵ, $\delta_\mu(\epsilon)$ is a strictly increasing continuous function of μ. Therefore, $\mu_{\text{GDP}}(x, y)$ is an increasing function of y: this leads to the order-preserving property. The second property follows immediately from the definition of μ_{GDP}.

By taking advantage of the order-preserving property, direct comparisons between δ_μ and $\delta_{\mathcal{A}}$ are no longer necessary: instead, it is sufficient to compare their corresponding GDPTs. Furthermore, appealing to the second property above, we need only compare $G_{\mathcal{A}}$ to the constant function μ. The following theorems formalize this insight.

Corollary 2.10 *An algorithm $\mathcal{A}$ with the privacy profile $\delta_{\mathcal{A}}$ is μ-GDP if and only if $\mu \geq \sup(\{G_{\mathcal{A}}(\epsilon) \mid \epsilon \in [0, \infty)\})$.*

Theorem 2.11 *An algorithm $\mathcal{A}$ with the privacy profile $\delta_{\mathcal{A}}$ is (ϵ_h, μ)-head GDP or (ϵ_t, μ)-tail GDP if and only if $\mu \geq \sup(\{G_{\mathcal{A}}(\epsilon) \mid \epsilon \in [0, \epsilon_h])$ or $\mu \geq \sup(\{G_{\mathcal{A}}(\epsilon) \mid \epsilon \in (\epsilon_t, \infty))$, respectively.*

Without the above results, we would be forced to search through a large family of functions for a single δ_μ that never crosses $\delta_{\mathcal{A}}$ anywhere on $[0, \infty)$ and has μ as small as possible. Now, with Theorem 2.10, we need only consider one function: the GDPT of $\delta_{\mathcal{A}}$. The tightest value μ is $\sup_\epsilon\{G_{\mathcal{A}}(\epsilon)\}$. Now we revisit our previous three examples for which the limit in Theorem 2.8 is 0, $\sqrt{1/2}$, and $+\infty$, respectively. From these evaluations, we can conclude that the Laplace mechanism and SGD are GDP and that the privacy profile of the ICEA algorithm crosses every μ-GDP curve regardless of how large μ is, indicating that the ICEA algorithm is not GDP.

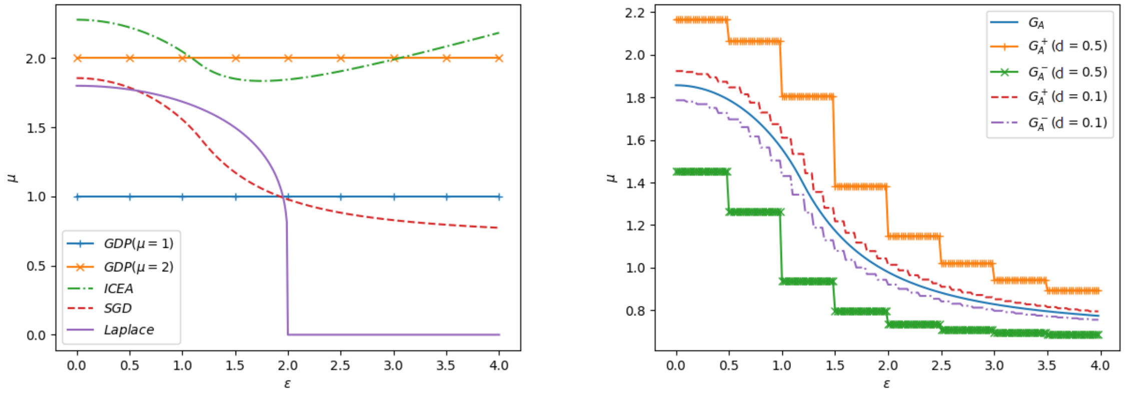


Figure 2.2: Left: Examples of GDPTs. Right: Plot of $G_{\mathcal{A}}^+$ and $G_{\mathcal{A}}^-$ with different values of d.

Left side of figure 2.2 shows the GDPTs of the three examples considered in this paper. All three GDPTs converge to a finite value as $\epsilon \to 0^+$. This can be attributed to the fact that any algorithm providing some non-trivial (ϵ, δ)-DP guarantee is $(0, \delta)$-DP for some $\delta \in [0, 1)$ (by theorem 2.2). For larger values of ϵ, the GDPT of the Laplace mechanism takes on a constant value of 0, the GDPT of SGD converges to a value that is approximately 0.7, and the GDPT of the ICEA seems to be diverging. These observations are consistent with the values of 0, $\sqrt{1/2}$, and ∞ obtained from Theorem 2.8. Once an algorithm is confirmed to be GDP via Theorems 2.6 and 2.8, it is natural to be interested in the exact level of privacy protection, quantified by μ. Nonetheless, plots are only good for visualization and are not sufficient proof when verifying GDP. We still need objective and tractable methods for obtaining bounds on GDPTs.

2.5.1 Measuring the head

Following the intuition outlined by definition 2.4 and 2.5, we decompose the GDP condition into head and tail conditions and first focus on finding μ such that $\mathcal{A}$ is (ϵ, μ)-head GDP. Without additional knowledge, finding $\sup\{G_{\mathcal{A}}(\epsilon) \mid \epsilon \in [0, \epsilon_h]\}$, even for a finite ϵ_h, seems computationally infeasible. To solve this problem, we take advantage of the fact that μ_{GDP} has a uniformly bounded partial derivative.

Theorem 2.12 $0 \leq \frac{\partial \mu_{GDP}(\epsilon,\delta)}{\partial \epsilon} \leq \frac{\sqrt{2}\pi}{2}$.

The first half of the inequality above is no surprise to us: the GDP privacy measurement μ is expected to be larger when ϵ is larger. However, the second half allow us to only conduct the search on a finite list of ϵ without the concern of spikes in between. We formulate this insight as the following theorem:

Theorem 2.13 *Given $\epsilon_h \geq 0$, let $d = \epsilon_h/n$ and $x_i = id$ for $i \in \{0, \ldots, n+1\}$. For $\epsilon \leq \epsilon_h$, the GDPT of $\mathcal{A}$, denoted by $G_{\mathcal{A}}(\epsilon)$, is bounded between the two staircase functions*

$$G^-_{\mathcal{A}}(\epsilon) = \sum_{i=0}^{n+1} \mu_{GDP}(x_i, \delta_{\mathcal{A}}(x_{i+1})) \times 1_{\epsilon \in [\mathrm{x_i}, \mathrm{x_{i+1}})} \quad \text{and } G^+_{\mathcal{A}}(\epsilon) = \sum_{i=0}^{n+1} \mu_{GDP}(x_{i+1}, \delta_{\mathcal{A}}(x_i)) \times 1_{\epsilon \in [\mathrm{x_i}, \mathrm{x_{i+1}})}.$$

Specifically,

$$\max_{i \in \{0,\ldots,n\}} G^-_{\mathcal{A}}(x_i) \leq \max_{\epsilon \in [0,\epsilon_h]} G_{\mathcal{A}}(\epsilon) \leq \max_{i \in \{0,\ldots,n+1\}} G^+_{\mathcal{A}}(x_i) \leq \max_{i \in \{0,\ldots,n\}} G^-_{\mathcal{A}}(x_i) + \sqrt{2}\pi d. \quad (2.2)$$

Refer to Appendix 2.8.1 and 2.8.1 for proofs of Theorem 2.12 and 2.13, respectively.

For any $\epsilon_h < +\infty$, we can now bound any GDPT $G_{\mathcal{A}}$ to any precision on $[0, \epsilon_h]$ without full pointwise evaluation because $G_{\mathcal{A}}$ is bounded between $G_{\mathcal{A}}^{+}$ and $G_{\mathcal{A}}^{-}$ and each staircase function takes on only finitely many values. For any $c > 0$, the inequalities in (2.2) provide a viable way to bound $\max_{\epsilon \in [0,\epsilon_h]} G_{\mathcal{A}}(\epsilon)$ in an interval with a length no greater than $1/c$.

First, a binary search algorithm (algorithm 2 in Appendix 2.8.3) can yield μ^{+} and μ^{-} such that $\mu^{-} \leq \mu_{\text{GDP}}(\epsilon, \delta) \leq \mu^{+}$ and $\mu^{+} - \mu^{-} < b$. For future references, we use $\mu_{\text{GDP}}^{+}(\epsilon, \delta, b)$ and $\mu_{\text{GDP}}^{-}(\epsilon, \delta, b)$ to represent such outputs of μ^{+} and μ^{-}, respectively. Therefore, we can naively go thorough all $G_{\mathcal{A}}^{-}(x_i)$ and $G_{\mathcal{A}}^{+}(x_i)$. By picking $n = \lceil \sqrt{8} c \pi \epsilon_h \rceil + 1$ and $b = \frac{1}{2c}$, the true gap between $\max G_{\mathcal{A}}^{-}(\epsilon)$ and $\max G_{\mathcal{A}}^{+}(\epsilon)$ is less than $\frac{1}{2c}$ and the error margin of the binary search estimate the μ_{GDP} is also $\frac{1}{2c}$. Therefore, the overall gap is bounded by $\frac{1}{c}$. As for complexity, each binary search has a time complexity of $O(\log(c))$ and the number of binary searches is $2n + 2 = O(\epsilon_h c)$. The overall time complexity of this naive approach is $O(\epsilon_h c \log(c))$. For a complete pseudocode of this naive approach, refer to algorithm 3 in Appendix 2.8.3.

By leveraging some properties of μ_{GDP} and shuffling, the expected number of binary searches needed can be reduced from linear ($2n + 2 \approx c\epsilon_h$) to logarithmic ($O(\log(c\epsilon_h))$). Such reduction will eliminate the logarithmic term in the time complexity from the naive algorithm. The improved algorithm is given as Algorithm 1 below.

Algorithm 1: Finding μ with privacy profiles (optimized).

```
**Input:** $\delta_{\mathcal{A}}$, $\epsilon_h$, $\mu_t$, $c$. (Privacy profile, searching range $\epsilon_h$, reciprocal of error margin)
$n \leftarrow \left\lceil \sqrt{8}c\pi\epsilon_h \right\rceil + 1$
$d \leftarrow \frac{\epsilon_h}{n-1}$
$\mu_- \leftarrow 0$
$\mu_+ \leftarrow 0$
$\mathcal{S} = [0, 1, \cdots, n+1]$
Shuffle $\mathcal{S}$
**for** $i = 0$ **to** $n+1$ **do**
    $x^- \leftarrow S[i]d$
    $x^+ \leftarrow (S[i]+1)d$
    **if** $\delta_{\mu^+}(x^-) < \delta_{\mathcal{A}}(x^+)$ **then**
        $\mu^+ \leftarrow \mu^+_{\text{GDP}}(x^-, \delta_{\mathcal{A}}(x^+), \frac{1}{2c}))$
    **end if**
    **if** $\delta_{\mu^-}(x^+) < \delta_{\mathcal{A}}(x^-)$ **then**
        $\mu^- \leftarrow \mu^-_{\text{GDP}}(x^+, \delta_{\mathcal{A}}(x^-), \frac{1}{2c}))$
    **end if**
**end for**
**Output:**$\mu_-$, $\mu_+$ (lower and upper bound of $\mu$).
```

We remark that this algorithm also has better accuracy than the naive algorithm because the lower and upper bounds will be closer while maintaining coverage. Refer to Appendix 2.8.3 for a detailed explanation of this algorithm.

2.5.2 Understanding the tail

With Theorem 2.13, one can verify (ϵ_h, μ)-head GDP conditions for arbitrarily large ϵ_h and an arbitrarily precise approximation of μ. While the error in μ can be quantified by D, one gap remains: ϵ_h can be arbitrarily large but can never truly be $+\infty$. In this subsection, we discuss the gap between (ϵ_h, μ)-head GDP and actual GDP (which is equivalent to $(+\infty, \mu)$-head GDP). Before giving a solution, we intuitively illustrate the gap between (ϵ_h, μ)-head GDP and actual GDP. Consider the following two cases:

- GDP with catastrophic failure, where with probability $1-p$, $\mathcal{A}_1$ functions properly as μ-GDP, with probability p, $\mathcal{A}_1$ malfunctions and discloses the entire dataset; and

- head-GDP with ϵ-DP, where $\mathcal{A}_2$ is both (ϵ_h, μ)-head GDP and $(\epsilon_h, 0)$-DP.

The head GDP privacy guarantee lies strictly between those of $\mathcal{A}_1$ and $\mathcal{A}_2$: specifically, $\delta_{\mathcal{A}_1}(\epsilon) < \delta(\epsilon) < \delta_{\mathcal{A}_2}(\epsilon)$. As an interpretation of this inequality, a head GDP privacy guarantee is safer than the original GDP guarantee but with a minuscule probability of failure, and when combined with a very weak ϵ-DP condition, the head GDP will be stronger than the actual GDP. In practice, μ is rarely above six in GDP, and ϵ is rarely above 10 in ϵ-DP because more extreme values provide almost no privacy protection [12]. If we verify the head condition up to $\epsilon_h = 100$ (which is not difficult because the time required for verification grows linearly) and take $\mu = 6$, then $p = \delta_\mu(\epsilon_h)$ will be on the order of 10^{-43}. Also, DP guarantee for ϵ this large is rarely considered to provide real protection. Hence, we conclude that the gap will not make any notable difference in practice with a proper choice of μ and ϵ_h.

2.5.3 Amplification

In some cases, one may wish to theoretically mend the gap discussed in the last subsection. This can be achieved by adding extra steps to perturb the output of the algorithm (i.e., via post-processing). We propose the following “clip and rectify” procedure that can turn any (ϵ_h, μ)-head GDP algorithm into a μ-GDP algorithm at some utility cost.

Theorem 2.14 *Let $\mathcal{A}$ be an (ϵ_h, μ)-head GDP algorithm with a numeric output. Assume that $-\infty < y^- < y^+ < +\infty$. Define $\mathcal{C}(y) = \max(\min(y, y^+), y^-)$ and $\mathcal{R}(z) = z + v$, where v is sampled from* Laplace(b) *with $b = (y^+ - y^-)/\epsilon_h$. Then $\mathcal{R} \circ \mathcal{C} \circ \mathcal{A}$ is μ-GDP.*

Refer to Appendix 2.8.1 for a proof of Theorem 2.14. We remark that, in order to minimize the utility loss, the bounds y^- and y^+ should be properly or dynamically chosen and the head condition should be verified to a value of ϵ_h that is as large as possible.

On the other hand, the performance (μ) of a GDP algorithm may be bottlenecked by the value of its privacy profile near the origin. This problem can be remedied by subsample pre-processing, the impact of which on privacy profiles has been thoroughly examined in [17]. The resulting privacy profile is explicitly given in Theorems 8–10 of [17]. With the help of the GDPT, we can select different subsample ratios and measure μ. For instance, the Laplace example in this paper was originally 1.80-GDP. If we introduce a 50%- or 10%- Poisson subsampling before the Laplace mechanism, μ will be reduced to 0.98 or 0.28, respectively. Refer to 2.8.6 for a complete graph of the new GDPTs.

While one could turn to other algorithms or design a new GDP mechanism in unfavourable cases where a candidate algorithm is incompatible with GDP from the start, rectifying these incompatibilities via pre- and post-processing may be more effective and efficient. This is especially true in cases where raw data is not easily accessible. In other cases, the DP mechanism might be inaccessible. This is particularly common for users of proprietary software. While they cannot identify and change the algorithm distributed in binary code, users can still control sensitive information by only approving a subset for release.

2.6 Applications

2.6.1 The Gaussian nature of ϵ-DP and the Laplace mechanism

By our previous analysis of the GDPT, we know that being GDP means that a privacy profile has a quickly vanishing tail (i.e., $\delta(\epsilon)$ must be $O(e^{-\epsilon^2})$). It is remarkable that another single parameter family of DP conditions, the ϵ-DP conditions, is also a property that pertains to the tail of privacy profiles. For any ϵ_0-DP algorithm, the privacy profile must be exactly 0 after ϵ_0. This suggests that ϵ-DP is stronger than GDP. Next, we will quantify this intuition using the tools we developed above.

By Theorem 2.2, we know if $\mathcal{A}$ is ϵ_0-DP, then in the worst case, $\delta_{\mathcal{A}}(\epsilon) = (e^{\epsilon_0} -$

$e^{\epsilon})^{+}/(1+e^{\epsilon_0})$.

We consider the GDPT of $\delta_{\mathcal{A}}$, denoted by $G_{\mathcal{A}}$. It is easy to see that, for $\epsilon \geq \epsilon_0$, $G_{\mathcal{A}}(\epsilon) = 0$: we need only consider $\epsilon \in [0, \epsilon_0)$. Let $G_{\delta_{\mathcal{A}}(\epsilon)}$ be denoted by μ_ϵ. Using the partial derivative of $G_{\mathcal{A}}$ derived in Appendix 2.8.1, we know that $\frac{\partial}{\partial \epsilon} G_{\delta_{\mathcal{A}}(\epsilon)} = \sqrt{2\pi} \exp\left\{ \left(\mu_\epsilon^2 + 2\epsilon\right)^2 / (8\mu_\epsilon^2) \right\} \left[\Phi(-\frac{\mu_\epsilon^2 + 2\epsilon}{2\mu_\epsilon}) - \Phi(\frac{-\mu_0}{2}) \right]$. Then $\text{sign}(\frac{\partial}{\partial \epsilon} G_{\delta_{\mathcal{A}}(\epsilon)}) = \text{sign}(\mu_\epsilon - \mu_0 - 2\epsilon/\mu_0)$. We can conclude that $\mu_\epsilon \leq \mu_0$ and, further, that $G_{\mathcal{A}}(\epsilon)$ is strictly decreasing on $[0, \epsilon_0)$. By Theorem 2.10, we know that $\mathcal{A}$ is μ_0-GDP. This finding can be more generally formulated as the following theorem.

Theorem 2.15 *Any $(\epsilon, 0)$-DP algorithm is also μ-GDP for $\mu = -2\Phi^{-1}(1/(1+e^{\epsilon})) \leq \sqrt{\pi/2}\epsilon$.*

[12] pointed out that the DP guarantees of the Laplace mechanism are stronger than those correspondingly provided by ϵ-DP. We reaffirm this difference by showing that it still exists under the GDP framework. The Laplace mechanism satisfies μ-GDP for μ smaller than the bound given in Theorem 2.15. The GDPTs presented in Appendix 2.8.5 illustrate this difference.

2.6.2 Handling composition with GDP

In practice, it is rare for a dataset to go through DP algorithms only once. Multiple statistics may be of interest or one statistic may require multiple inquiries to acquire. DP algorithms applied to the same dataset multiple times are usually still DP but with worse privacy parameters. Composition theorems quantitatively trace privacy loss and provide a privacy parameter for the ensemble. However, not only is exact composition an intrinsically (#P-)hard problem [33], but the conclusions of composition theorems are also often problematic. Take traditional (ϵ, δ)-DP as an example. [34] gives an optimal composition theorem, but the composition of two (ϵ, δ)-DP algorithms cannot be characterized under the (ϵ, δ)-DP framework. This result damages interpretability because the representation of a composition will no longer be in two parameters. This type of flaw is the major motivation for a GDP characterization of algorithms derived under other DP frameworks. The composition of GDP algorithms is easy, exact, and closed: the composition of a μ_1- and μ_2-GDP algorithm is simply $\sqrt{\mu_1^2 + \mu_2^2}$-GDP. GDP also has a special central limit theorem which implies that, for all privacy definitions that retain hypothesis testing with proper scaling, the privacy guarantee of a composition converges to GDP in the limit. In this subsection, we demonstrate that GDP is a powerful tool for composition by unifying other notions under the GDP framework and then using the GDP composition theorem. As baselines, we select basic composition [6], advanced composition [35] and Rényi-DP [13].

We consider the 50-fold composition of 0.2-DP algorithms. In this setting, the basic composition is pessimistic and says that the composition will be 10-DP, which means there is next to no privacy guarantee. According to corollary 1 of [13], the bound given by RDP is even looser. Refer to Figure 3 for the results of other theorems.

We next consider composition using the proposed measurement method. According to Theorem 2.15, a 0.2-DP algorithm is 0.2505-GDP. If the algorithm is the Laplace mechanism, then the algorithm in Appendix 2.8.3 can tighten μ to 0.2391. To compute μ for a 50-fold composition, we simply multiply the original μ by $\sqrt{50}$. The result is 1.771-GDP (1.691 for the Laplace mechanism). In this case, distinguishing two neighbouring datasets is as hard as distinguishing between $N(0, 1)$ and $N(1.771, 1)$ on the basis of a single observation.

In this particular case, the ground truth can be derived from the optimal composition theorem [34]. We present the results from the optimal composition theorem in Table 1 and Figure 3 for comparison, but we do not consider the optimal composition theorem to be generally superior because the ground truth is not easy to compute and because the former method is not as interpretable and only works for algorithms whose DP guarantees are fixed at (ϵ, δ). However, by applying the GDPT, the privacy guarantee of the optimal composition theorem can be summarised as 1.420-GDP . Compared to the central limit theorem in [12], which yields $\mu = \sqrt{2}$ (with an unknown asymptotic

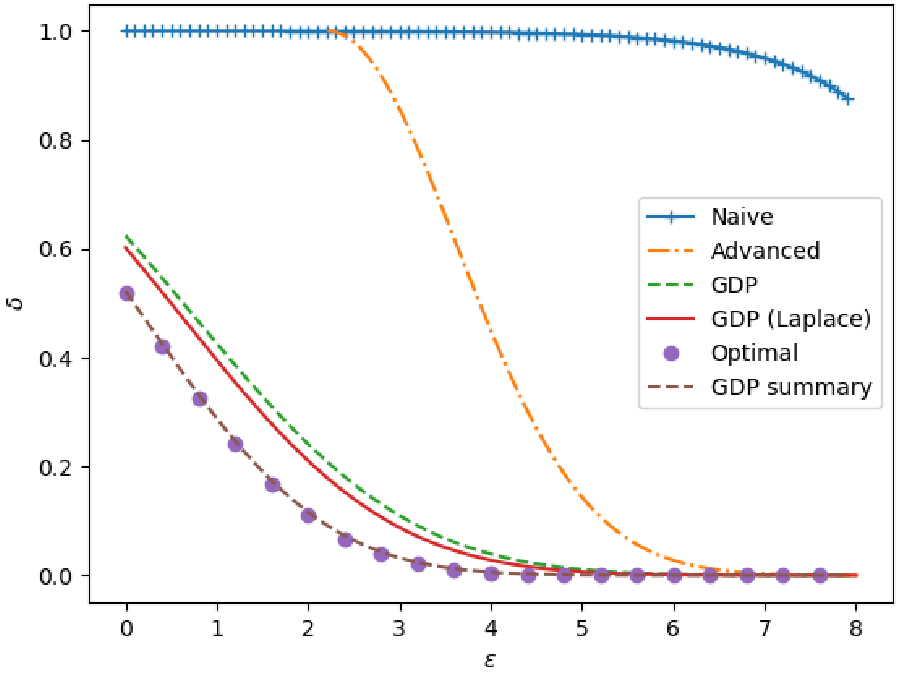


Figure 2.3: The plot of privacy guarantee under different methods.

Method \ δ	10^{-1}	10^{-2}	10^{-3}	10^{-4}
Basic	9.89	9.99	10	10
Advanced	5.25	6.51	7.47	8.28
RDP	12.14	17.17	21.03	24.28
GDP	3.1	5.06	6.47	7.62
GDP (Lap)	2.87	4.74	6.09	7.19
Optimal	2.12	3.64	4.76	5.28
GDP summary	2.14	3.73	4.87	5.80

Table 2.1: Minimum values of ϵ to achieve corresponding (ϵ, δ)-DP.

approximation error) in the same setting, the tractable numerical procedure of GDPT provides a satisfying result.

2.7 Conclusion and Future Work

In this paper, we provided both an analytic perspective of and engineering tools for the GDP framework. By using the new notions we proposed, we devised solutions to three aspects of GDP: identification, amplification, and measurement. The developments in this paper suggest numerous interesting directions for future work. First, more refined methods can be derived to expand the toolbox of rectification for more versatility. Second, the measurement procedure can be combined with the rectification procedure. Incrementally introducing more pre- and post-processing steps and dynamically checking whether privacy guarantees are already satisfactory can also be explored. Lastly, the idea underlying the GDPT can be generalized to other parameterized DP notions like CDP or RDP to enrich tractability and visualizability in the DP literature.

2.8 Appendix

2.8.1 Proofs

Proof.

Proof of Theorem 2.2:

Sufficiency:

When $\epsilon \geq \epsilon_0$, the sufficiency is trivial as $\delta = \delta_0$.

When $\epsilon < \epsilon_0$, given that $\mathcal{A}$ is (ϵ_0, δ_0)-DP, by the definition, for any pair of datasets S and S' that differ in the record of a single individual and any event E,

$$P[\mathcal{A}(S) \in E] \leq e^{\epsilon_0} P\left[\mathcal{A}\left(S'\right) \in E\right] + \delta_0.$$

When $P\left[\mathcal{A}\left(S'\right) \in E\right] \leq \frac{1-\delta_0}{1+e^{\epsilon_0}} := c_0$,

$$\begin{aligned}
P[\mathcal{A}(S) \in E] &\leq e^{\epsilon_0} P\left[\mathcal{A}\left(S'\right) \in E\right] + \delta_0 \\
&\leq (e^{\epsilon_0} + e^{\epsilon} - e^{\epsilon}) P\left[\mathcal{A}\left(S'\right) \in E\right] + \delta_0 + \delta - \delta \\
&\leq e^{\epsilon_0} P\left[\mathcal{A}\left(S'\right) \in E\right] + \delta + (e^{\epsilon_0} - e^{\epsilon}) c_0 + \delta_0 - \delta \\
&\leq e^{\epsilon} P\left[\mathcal{A}\left(S'\right) \in E\right] + \delta + (e^{\epsilon_0} - e^{\epsilon}) c_0 - \frac{(1-\delta_0)(e^{\epsilon_0} - e^{\epsilon})}{1+e^{\epsilon_0}} \\
&\leq e^{\epsilon} P\left[\mathcal{A}\left(S'\right) \in E\right] + \delta.
\end{aligned}$$

When $c_0 \leq P\left[\mathcal{A}\left(S'\right) \in E\right] \leq 1$,

$$\begin{aligned}
P[\mathcal{A}(S) \in E] &= 1 - P[\mathcal{A}(S) \in E^c] \\
&\leq 1 - e^{-\epsilon_0}(P\left[\mathcal{A}\left(S'\right) \in E^c\right] - \delta_0) \\
&= 1 - e^{-\epsilon_0}(1 - P\left[\mathcal{A}\left(S'\right) \in E\right] - \delta_0) \\
&= 1 - e^{-\epsilon_0} + e^{-\epsilon_0} P\left[\mathcal{A}\left(S'\right) \in E\right] + e^{-\epsilon_0}\delta_0 \\
&= 1 - e^{-\epsilon_0} + e^{-\epsilon_0}\delta_0 + \delta - \delta + (e^{-\epsilon_0} + e^{\epsilon} - e^{\epsilon}) P\left[\mathcal{A}\left(S'\right) \in E\right] \\
&= e^{\epsilon} P\left[\mathcal{A}\left(S'\right) \in E\right] + \delta + 1 - e^{-\epsilon_0} + e^{-\epsilon_0}\delta_0 - \delta + (e^{-\epsilon_0} - e^{\epsilon}) P\left[\mathcal{A}\left(S'\right) \in E\right] \\
&\leq e^{\epsilon} P\left[\mathcal{A}\left(S'\right) \in E\right] + \delta + 1 - e^{-\epsilon_0} + e^{-\epsilon_0}\delta_0 - \delta + (e^{-\epsilon_0} - e^{\epsilon}) c_0 \\
&= e^{\epsilon} P\left[\mathcal{A}\left(S'\right) \in E\right] + \delta + (1 - \delta_0)(\frac{e^{-\epsilon_0} - e^{\epsilon}}{1 + e^{\epsilon_0}} - e^{-\epsilon_0}) + 1 - \delta \\
&\leq e^{\epsilon} P\left[\mathcal{A}\left(S'\right) \in E\right] + \delta + (1 - \delta_0)(\frac{e^{-\epsilon_0} - e^{\epsilon}}{1 + e^{\epsilon_0}} - e^{-\epsilon_0} + 1 + \frac{e^{\epsilon} - e^{\epsilon_0}}{1 + e^{\epsilon_0}}) \\
&= e^{\epsilon} P\left[\mathcal{A}\left(S'\right) \in E\right] + \delta.
\end{aligned}$$

Necessity:

We prove the necessity by giving a specific (ϵ_0, δ_0)-DP algorithm $\mathcal{A}$ such that $\delta_{\mathcal{A}}(\epsilon)$ is exactly $\delta_0 + \frac{(1-\delta_0)(e^{\epsilon_0} - e^{\epsilon})^+}{1+e^{\epsilon_0}}$.

Define $\Omega_e = \{1, 2, 3, 4\}$ and $\Omega_S = \{0, 1\}$. Let $\epsilon \geq 0$, $0 \leq \delta_0 \leq 1$ and denote $\frac{e^{\epsilon_0}}{1+e^{\epsilon_0}}$ as α_0. Let $\mathcal{A}$ be a randomized algorithm that take a single point from Ω_S and generate output as follows:

$$\begin{cases} P(\mathcal{A}(S) = 1 \mid S = 0) = \delta_0, \\ P(\mathcal{A}(S) = 2 \mid S = 0) = 0, \\ P(\mathcal{A}(S) = 3 \mid S = 0) = (1 - \delta_0)\alpha_0, \\ P(\mathcal{A}(S) = 4 \mid S = 0) = (1 - \delta_0)(1 - \alpha_0), \end{cases} \quad \begin{cases} P(\mathcal{A}(S) = 1 \mid S = 1) = 0, \\ P(\mathcal{A}(S) = 2 \mid S = 1) = \delta_0, \\ P(\mathcal{A}(S) = 3 \mid S = 1) = (1 - \delta_0)(1 - \alpha_0), \\ P(\mathcal{A}(S) = 4 \mid S = 1) = (1 - \delta_0)\alpha_0. \end{cases}$$

By definition, $\delta(\epsilon)$ is the smallest δ such that $P(\mathcal{A}(S) \subset E \mid S = s) \leq e^{\epsilon} P(\mathcal{A}(S) \subset E \mid S = 1 - s) + \delta$ holds true for all $E \subset \Omega_e$ and $s \in \Omega_S$. By checking all 64 combinations, we can conclude that $\delta_{\mathcal{A}}(\epsilon) = \delta_0 + \frac{(1-\delta_0)(e^{\epsilon_0} - e^{\epsilon})^+}{1+e^{\epsilon_0}}$. ■

Proof. Proof of Lemma 2.4:

It is well known that [36], for $t < 0$:

$$\frac{1}{-t+\sqrt{t^2+4}} < \sqrt{\frac{\pi}{2}} \exp\left(\frac{t^2}{2}\right) \Phi(t) < \frac{1}{-t+\sqrt{t^2+\frac{8}{\pi}}}.$$

Let $a = \left(-\frac{\varepsilon}{\mu} + \frac{\mu}{2}\right)$ and $b = \left(-\frac{\varepsilon}{\mu} - \frac{\mu}{2}\right)$,

$$\begin{aligned}
\overline{\lim_{\epsilon\to\infty}} \delta_\mu(\epsilon) &= \overline{\lim_{\epsilon\to\infty}} \Phi\left(a\right) - \mathrm{e}^{\epsilon}\Phi\left(b\right) \\
&\leq \sqrt{\frac{2}{\pi}} \overline{\lim_{\epsilon\to\infty}} \frac{\exp\left(\frac{-a^2}{2}\right)}{-a+\sqrt{a^2+\frac{8}{\pi}}} - \frac{\exp\left(\frac{-b^2}{2}+\epsilon\right)}{-b+\sqrt{b^2+4}}. \\
&= \sqrt{\frac{2}{\pi}} \overline{\lim_{\epsilon\to\infty}} \exp\left(\frac{-a^2}{2}\right) \left(\frac{1}{-a+\sqrt{a^2+\frac{8}{\pi}}} - \frac{1}{-b+\sqrt{b^2+4}}\right). \\
&\leq \sqrt{\frac{2}{\pi}} \overline{\lim_{\epsilon\to\infty}} \exp\left(\frac{-a^2}{2}\right) \left(\frac{-1}{a}\right). \\
&= 0.
\end{aligned}$$

$$\begin{aligned}
\underline{\lim}_{\epsilon\to\infty} \delta_\mu(\epsilon) &= \underline{\lim}_{\epsilon\to\infty} \Phi\left(a\right) - \mathrm{e}^{\epsilon}\Phi\left(b\right) \\
&\geq \sqrt{\frac{2}{\pi}} \underline{\lim}_{\epsilon\to\infty} \frac{\exp\left(\frac{-a^2}{2}\right)}{-a+\sqrt{a^2+4}} - \frac{\exp\left(\frac{-b^2}{2}+\epsilon\right)}{-b+\sqrt{b^2+\frac{8}{\pi}}}. \\
&= \sqrt{\frac{2}{\pi}} \underline{\lim}_{\epsilon\to\infty} \exp\left(\frac{-a^2}{2}\right) \left(\frac{1}{-a+\sqrt{a^2+4}} - \frac{1}{-b+\sqrt{b^2+\frac{8}{\pi}}}\right). \\
&\geq \sqrt{\frac{2}{\pi}} \underline{\lim}_{\epsilon\to\infty} \exp\left(\frac{-a^2}{2}\right) \left(\frac{-1}{b}\right). \\
&= 0.
\end{aligned}$$

Therefore,

$$\lim_{\epsilon\to\infty} \delta_\mu(\epsilon) = 0. \tag{2.3}$$

It is easy to see that,

$$\lim_{\epsilon\to\infty} \tilde{\delta}_\mu(\epsilon) = \lim_{\epsilon\to\infty} \frac{\mu e^{-a^2/2}}{\sqrt{2\pi}a^2} = 0 \tag{2.4}$$

By L'Hospital's rule:

$$\begin{aligned}
\lim_{\epsilon\to\infty}\frac{\tilde{\delta}_\mu(\epsilon)}{\delta_\mu(\epsilon)} &= \lim_{\epsilon\to\infty}\frac{\tilde{\delta}'_\mu(\epsilon)}{\delta'_\mu(\epsilon)}\\
&= \lim_{\epsilon\to\infty} -\frac{e^{-\frac{a^2}{2}}\,(a^2+2)}{\sqrt{2\pi}a^3}\Bigg/ e^{\epsilon}\Phi(b)\\
&= \lim_{\epsilon\to\infty}\frac{e^{-\frac{b^2}{2}}\Phi(b)}{\sqrt{2\pi}b}\\
&= \lim_{b\to-\infty}\frac{e^{-\frac{b^2}{2}}\Phi(b)}{\sqrt{2\pi}b}\\
&= 1.
\end{aligned}$$

■

Proof. Proof of Theorem 2.6:

Sufficiency:

If $\mathcal{A}$ is μ-GDP. Then $\overline{\lim_{\epsilon\to+\infty}} G_{\mathcal{A}}(\epsilon) \le \overline{\lim_{\epsilon\to+\infty}} G_{\delta_\mu}(\epsilon) = \mu$.

Necessity:

If $\overline{\lim_{\epsilon\to+\infty}} G_{\mathcal{A}}(\epsilon) = \mu < +\infty$, there must be a $\epsilon_t > 0$ such that $\mathcal{A}$ is $(\epsilon_t, \mu_0 + 1)$-tail GDP.

Notice that $\lim_{\mu\to\infty} \delta_\mu(\epsilon_t) = 1$, we can pick $\mu_1 > \mu_0$ large enough such that $\delta_{\mu_1}(\epsilon_t) > \delta_{\mathcal{A}}(0)$.

This is possible because by Theorem 2.2, $\delta_{\mathcal{A}}(0) < 1$. Then for $\epsilon \in [0, \epsilon_t)$, $\delta_{\mathcal{A}}(\epsilon) \le \delta_{\mathcal{A}}(0) \le \delta_{\mu_1}(\epsilon_t) \le \delta_{\mu_1}(\epsilon)$. $\mathcal{A}$ is both (ϵ_t, μ)-head and tail GDP for $\mu = \mu_0 + \mu_1 + 1$. $\mathcal{A}$ is GDP as desired. ■

Proof. Proof of Theorem 2.8:

Let $\overline{\lim_{\epsilon\to+\infty}} G_f(\epsilon) = \mu_t$.

First we show that $\overline{\lim_{\epsilon\to\infty}} \frac{\epsilon^2}{-2\log\delta_{\mathcal{A}}(\epsilon)} \le \mu_t^2$:

By the definition the limit, for any $\mu_0 > \mu_t$, for sufficient large ϵ, $G_f(\epsilon) < \mu_0$ and further $\delta_{\mathcal{A}}(\epsilon) \le \delta_{\mu_0}(\epsilon)$. Hence, $\overline{\lim_{\epsilon\to\infty}} \frac{\delta_{\mathcal{A}}(\epsilon)}{\delta_{\mu_0}(\epsilon)} \le 1$. By Lemma 2.4, $\overline{\lim_{\epsilon\to\infty}} \frac{\delta_{\mathcal{A}}(\epsilon)}{\tilde{\delta}_{\mu_0}(\epsilon)} \le 1$.

Then $\lim_{\epsilon\to\infty} \frac{\epsilon^2}{-2\log\delta_{\mathcal{A}}(\epsilon)} \le \lim_{\epsilon\to\infty} \frac{\epsilon^2}{-2\log\tilde{\delta}_{\mu_0}(\epsilon)} = \mu_0^2$.

$\lim_{\epsilon\to\infty} \frac{\epsilon^2}{-2\log\delta_{\mathcal{A}}(\epsilon)} \le \mu_t$ as desired as we take $\mu_0 \to \mu_t$.

Next we show that $\overline{\lim}_{\epsilon\to\infty} \frac{\epsilon^2}{-2\log\delta_{\mathcal{A}}(\epsilon)} \geq \mu_t^2$:

If $\overline{\lim}_{\epsilon\to\infty} \frac{\epsilon^2}{-2\log\delta_{\mathcal{A}}(\epsilon)} = \mu_0^2 < \mu_t^2$, then by Lemma 2.4,

$$\begin{aligned}\overline{\lim_{\epsilon\to\infty}} \frac{\epsilon^2}{-2\log\delta_{\mathcal{A}}(\epsilon)} - \frac{\epsilon^2}{-2\log\delta_{\mu_t}(\epsilon)} &= \overline{\lim_{\epsilon\to\infty}} \frac{\epsilon^2}{-2\log\delta_{\mathcal{A}}(\epsilon)} - \overline{\lim_{\epsilon\to\infty}} \frac{\epsilon^2}{-2\log\tilde{\delta}_{\mu_t}(\epsilon)} \\ &< \mu_0^2 - \mu_t^2\end{aligned}$$

Then for a sufficiently large ϵ_0,

$$\frac{\epsilon_0^2}{-2\log\delta_{\mathcal{A}}(\epsilon_0)} - \frac{\epsilon_0^2}{-2\log\delta_{\mu_0}(\epsilon_0)} < 0.$$

Since log is an increasing function, it follows that $\delta_{\mathcal{A}}(\epsilon_0) < \delta_{\mu_0}(\epsilon_0)$. Then $\underline{\lim}_{\epsilon\to+\infty} G_f(\epsilon) \leq \mu_0 < \mu_t$, which is a contradiction. ■

Proof. Proof of Theorem 2.12:

Let $G_\mu(\epsilon) = F(\epsilon, \delta_\mu(\epsilon))$ and $F(x,y) = \mu_{\text{GDP}}(x,y)$.

By definition of μ_{GDP}, $G_\mu(\epsilon) = \mu$.

On one hand, $\begin{cases} \dfrac{\partial G_\mu(\epsilon)}{\partial\epsilon} = \dfrac{\partial\mu}{\partial\epsilon} = 0, \\ \dfrac{\partial G_\mu(\epsilon)}{\partial\mu} = \dfrac{\partial\mu}{\partial\mu} = 1. \end{cases}$

On the other hand, by chain rule, $\begin{cases} \dfrac{\partial G_\mu(\epsilon)}{\partial\epsilon} = \dfrac{\partial F}{\partial x} + \dfrac{\partial F}{\partial y}\dfrac{\partial\delta_\mu(\epsilon)}{\partial\epsilon}, \\ \dfrac{\partial G_\mu(\epsilon)}{\partial\mu} = \dfrac{\partial F}{\partial y}\dfrac{\partial\delta_\mu(\epsilon)}{\partial\mu}. \end{cases}$

Therefore, $\begin{cases} \dfrac{\partial F}{\partial y} = (\dfrac{\partial\delta_\mu(\epsilon)}{\partial\mu})^{-1}, \\ \dfrac{\partial F}{\partial x} = -(\dfrac{\partial\delta_\mu(\epsilon)}{\partial\mu})^{-1}\dfrac{\partial\delta_\mu(\epsilon)}{\partial\epsilon}. \end{cases}$

Using the close forms, $\frac{\partial\delta_\mu(\epsilon)}{\partial\epsilon}$ and $\frac{\partial\delta_\mu(\epsilon)}{\partial\mu}$ can be directly computed:

$$\begin{cases} \dfrac{\partial\delta_\mu(\epsilon)}{\partial\epsilon} = -e^{\epsilon}\Phi(-\dfrac{\mu^2+2\epsilon}{2\mu}), \\ \dfrac{\partial\delta_\mu(\epsilon)}{\partial\mu} = \dfrac{e^{-\frac{\left(\mu^2-2\epsilon\right)^2}{8\mu^2}}}{\sqrt{2\pi}}. \end{cases}$$

Hence, $\begin{cases} \dfrac{\partial F}{\partial x} = \sqrt{2\pi}e^{\frac{\left(\mu^2+2\epsilon\right)^2}{8\mu^2}}\Phi(-\dfrac{\mu^2+2\epsilon}{2\mu}) \leq \sqrt{2\pi}e^{\frac{\mu^2}{8}}\Phi(-\dfrac{\mu}{2}) \leq \dfrac{\sqrt{2}\pi}{2}, \\ \dfrac{\partial F}{\partial y} = \sqrt{2\pi}e^{\frac{\left(\mu^2-2\epsilon\right)^2}{8\mu^2}} > 0. \end{cases}$

Notice that $\frac{\partial F}{\partial x} = \sqrt{2\pi} e^{\frac{\left(\mu^2+2\epsilon\right)^2}{8\mu^2}} \Phi(-\frac{\mu^2+2\epsilon}{2\mu}) > 0$, combined with the fact that $\frac{\partial F}{\partial x} \leq \frac{\sqrt{2}\pi}{2}$, we can conclude that $0 \leq \frac{\partial \mu_{\text{GDP}}(\epsilon,\delta)}{\partial \epsilon} \leq \frac{\sqrt{2}\pi}{2}$. By $\frac{\partial F}{\partial y} > 0$, we can see GDPT is order preserving. ■

Proof. Proof of Theorem 2.13:

We now consider the gap between $\max_{i\in\{0,\cdots,n\}}\{G^-_{\mathcal{A}}(x_i)\}$ and $\max_{i\in\{0,\cdots,n+1\}}\{G^+_{\mathcal{A}}(x_i)\}$ bound the length of $[\mu^-,\mu^+]$ in two cases.

Case 1: If $\max_{i\in\{0,\cdots,n+1\}}\{G^+_{\mathcal{A}}(x_i)\} = G^+_{\mathcal{A}}(x_0)$, then $\max_{i\in\{0,\cdots,n+1\}}\{G^+_{\mathcal{A}}(x_i)\} = G^+_{\mathcal{A}}(x_0) = \mu_{\text{GDP}}(D, \delta_{\mathcal{A}}(0)) \leq \mu_{\text{GDP}}(0, \delta_{\mathcal{A}}(0)) + \frac{\sqrt{2}\pi D}{2}$. Therefore,

$$\max_{\epsilon\in[0,\epsilon_h]} G(\epsilon) \leq G^+_{\mathcal{A}}(x_0) \leq \{G^-_{\mathcal{A}}(x_0)\} + \frac{\sqrt{2}\pi D}{2}.$$

Case 2: If $\max_{i\in\{0,\cdots,n+1\}}\{G^+_{\mathcal{A}}(x_i)\} \neq G^+_{\mathcal{A}}(x_0)$, then by the order preserving property, the optimal μ lies in $[\mu^-,\mu^+]$, where $\mu^- = \max(\mu_h, \max_{i\in\{0,\cdots,n\}}\{G^-_{\mathcal{A}}(x_i)\})$ and $\mu^+ = \max(\mu_h, \max_{i\in\{1,\cdots,n+1\}}\{G^+_{\mathcal{A}}(x_i)\})$. Notice that

$$\begin{aligned}\max_{i\in\{0,\cdots,n\}}\{G^-_{\mathcal{A}}(x_i)\} &= \max_{i\in\{0,\cdots,n\}}\{\mu_{\text{GDP}}(x_i, \delta_{\mathcal{A}}(x_{i+1}))\} = \max_{i\in\{1,\cdots,n+1\}}\{\mu_{\text{GDP}}(x_{i-1}, \delta_{\mathcal{A}}(x_i))\}\\ &\geq \max_{i\in\{1,\cdots,n+1\}}\{\mu_{\text{GDP}}(x_{i+1}, \delta_{\mathcal{A}}(x_i)) - \sqrt{2}\pi D\}\\ &\geq \max_{i\in\{1,\cdots,n+1\}}\{G^+_{\mathcal{A}}(x_i)\} - \sqrt{2}\pi D.\end{aligned}$$

In both cases the gap is no greater than $\sqrt{2}\pi D$ as desired. ■

Proof. Proof of Theorem 2.14:

By the definition of $\mathcal{C}$, $\mathcal{C} \circ \mathcal{A}$ is bounded in $[y^-, y+]$. Therefore the global sensitivity of $\mathcal{C} \circ \mathcal{A}$ is no greater than $y^+ - y^-$. Then $\mathcal{R} \circ \mathcal{C} \circ \mathcal{A}$ is a special case of the Laplace mechanism. By [17], $\mathcal{R} \circ \mathcal{C} \circ \mathcal{A}$ is ϵ_h-DP. Then $\delta_{\mathcal{R}\circ\mathcal{C}\circ\mathcal{A}}(\epsilon) = 0 < \delta_\mu(\epsilon)$ for any $\epsilon \geq \epsilon_h$.

In addition, because of the post-processing property, $\delta_{\mathcal{R}\circ\mathcal{C}\circ\mathcal{A}}(\epsilon) \leq \delta_{\mathcal{A}}(\epsilon) < \delta_\mu(\epsilon)$ for any $\epsilon < \epsilon_h$.

Therefore, $\mathcal{R} \circ \mathcal{C} \circ \mathcal{A}$ is μ-GDP. ■

2.8.2 Refining the privacy profile

Given a trade-off function $\sigma = f(\epsilon, \delta)$ and a fixed parameter σ. From definition of the trade-off function it is instant that the for any $(\epsilon, \delta) \in \Omega = \{(\epsilon, \delta) \mid \sigma = f(\epsilon, \delta)\}$, (ϵ, δ)-DP is guaranteed. Then, (ϵ, δ)-DP is also guaranteed if there is a $(\epsilon_0, \delta_0) \in \Omega$ such that (ϵ_0, δ_0)-DP implies (ϵ, δ)-DP. Therefore,

$$\delta_{\mathcal{A}}(\epsilon) = \min\left(\{\delta \mid \sigma = f(\epsilon_0, \delta_0) \text{ and } \delta \geq \delta_0 + \frac{(1-\delta_0)(e^{\epsilon_0} - e^{\epsilon})^+}{1 + e^{\epsilon_0}}\}\right).$$

Notice that by theorem 2.2, (ϵ_0, δ_0)-DP implies (ϵ, δ) with $\delta < \delta_0$ only if $\epsilon < \epsilon_0$, we rewrite the $\delta_{\mathcal{A}}(\epsilon)$ as:

$$\delta_{\mathcal{A}}(\epsilon) = \inf_{\epsilon_0 \in [\epsilon, \infty)} g(\epsilon, \epsilon_0),$$

where $g(\epsilon, \epsilon_0) := (1 - \hat{\delta_{\mathcal{A}}}(\epsilon_0))\frac{e^{\epsilon_0} - e^{\epsilon}}{e^{\epsilon_0}+1} + \hat{\delta_{\mathcal{A}}}(\epsilon_0)$ and $\hat{\delta_{\mathcal{A}}}$ is the naive privacy profile defined implicitly by $\sigma = f(\epsilon_0, \delta_0)$. For continuously differentiable f, the minimum value of the right-hand side can be found be take the derivative:

$$\frac{\partial g(\epsilon, \epsilon_0)}{\partial \epsilon_0} = \frac{1 + e^{\epsilon}}{(1 + e^{\epsilon_0})^2}\left[\hat{\delta_{\mathcal{A}}}'(\epsilon_0) + e^{\epsilon_0}(1 - \hat{\delta_{\mathcal{A}}}(\epsilon_0) + \hat{\delta_{\mathcal{A}}}'(\epsilon_0))\right].$$

We remark that the sign of $\frac{\partial g(\epsilon, \epsilon_0)}{\partial \epsilon_0}$ does not depend on ϵ when $\epsilon > \epsilon_0$. For both of our example 2 and 3, we both find a particular value ϵ^i such that $Sign(\frac{\partial g(\epsilon, \epsilon_0)}{\partial \epsilon_0}) = -Sign(\epsilon - \epsilon^i)$. This means for $\epsilon \geq \epsilon^i$, $\delta_{\mathcal{A}}(\epsilon) = \hat{\delta_{\mathcal{A}}}(\epsilon)$ and otherwise $\delta_{\mathcal{A}}(\epsilon)$ equals to the δ value derived from $(\epsilon^i, \hat{\delta_{\mathcal{A}}}(\epsilon^i))$.

There is an interesting byproduct or the privacy profile refinement. Theoretically, the privacy profile refinement can also be used to improve an algorithm's utility. For example, the projected noisy SGD algorithm in [25] is (ϵ, δ)-DP and the trade-off function is $\sigma = -C\log(\delta_0)/\epsilon_0$. To achieve $(0.2, e^{-2})$-DP, it appears that σ needs to be chosen as $-C\log(e^{-2})/0.2 = 10C$. (ϵ, δ)-DP implies $(0.2, e^{-2})$-DP when $\delta + (1-\delta)(e^{\epsilon} - e^{0.2})^+/(1+e^{\epsilon}) = e^{-2}$. Numerical methods suggest that, by choosing $\epsilon \approx 0.334$ and $\delta \approx 0.067$, (ϵ, δ)-DP implies $(0.2, e^{-2})$-DP but $\sigma = -C\log(\delta)/\epsilon \approx 8.086C < 10C$. Therefore, the desired level of DP can be achieved with a lower noise parameter. However, this type of refinement majorly affects privacy profile around the origin and therefore minor in practice.

2.8.3 Behind efficient head measurement algorithm

First we formalize the binary search algorithm to find μ_{GDP}:

Algorithm 2: Binary search

```
Input: $\epsilon$, $\delta$, $b$. (The $(\epsilon, \delta)$-pair, searching range, error margin)
$\mu_- \leftarrow 0$
$\mu_+ \leftarrow \mu_{\max}$
repeat
    $\mu = \frac{\mu^+ + \mu^-}{2}$
    if $\delta_\mu(\epsilon) > \delta$ then
        $\mu^+ \leftarrow \mu$
    else
        $\mu^- \leftarrow \mu$
    end if
until $\mu^+ - \mu^- < b$
Output: $\mu_-$, $\mu_+$ (lower and upper bound of $\mu$).
```

It is possible to drop the need for the searching range $\mu_{\max}$ for this algorithm (e.g., exponentially search for an upper bound first or conduct a binary search on $\arctan \mu$ instead). We keep this input for clarity and simplicity. $\mu_{\max}$ can be set to a large constant for convenience, for example, 10. If the outputted μ^+ equals the preset value (10), the privacy profile fails to imply 10-GDP. In practice, GDP with $\mu \geq 6$ already provides almost no privacy protection [12].

With the formal definition of binary search, an exhaustive iteration method to bound the staircase functions outlined in Theorem 2.13 can be formally written as follows:

Algorithm 3: Finding μ with privacy profiles (naive).

```
**Input:** $\delta_{\mathcal{A}}$, $\epsilon_h$, $c$. (Privacy profile, searching range, reciprocal of error margin)
$n \leftarrow \lceil \sqrt{8} c \pi \epsilon_h \rceil + 1$
$d \leftarrow \frac{\epsilon_h}{n-1}$
$\mu_- \leftarrow 0$
$\mu_+ \leftarrow 0$
**for** $i = 0$ **to** $n+1$ **do**
    $x^- \leftarrow id$
    $x^+ \leftarrow (i+1)d$
    $\mu_+ \leftarrow \max(\mu_+, \mu^+_{\text{GDP}}(x^-, \delta_{\mathcal{A}}(x^+), \frac{1}{2c}))$
    $\mu_- \leftarrow \max(\mu_-, \mu^-_{\text{GDP}}(x^+, \delta_{\mathcal{A}}(x^-), \frac{1}{2c}))$
    $i \leftarrow i+1$
**end for**
**Output:** $\mu^+$, $\mu^-$.
```

To transform this naive algorithm into the optimized one. The first key observation is that the reassignment of μ_+ and μ_- can be optimized.

We take $\mu_+ \leftarrow \max(\mu_+, \mu^+_{\text{GDP}}(x^-, \delta_{\mathcal{A}}(x^+), \frac{1}{2c}))$ for example, same optimization can be applied to $\mu_- \leftarrow \max(\mu_-, \mu^-_{\text{GDP}}(x^+, \delta_{\mathcal{A}}(x^-, \frac{1}{2c})))$ as well. The naive operation, $\mu_+ \leftarrow \max(\mu_+, \mu^+_{\text{GDP}}(x^-, \delta_{\mathcal{A}}(x^+), \frac{1}{2c}))$ can be optimized into "If $\delta_{\mu^+}(x^-) < \delta_{\mathcal{A}}(x^+)$, then $\mu^+ \leftarrow \mu^+_{\text{GDP}}(x^-, \delta_{\mathcal{A}}(x^+), \frac{1}{2c}))$" without lost of accuracy. To see this, we list all three possibilities as follows:

- Case 1: $\mu^+ < \mu_{\text{GDP}}(x^-, \delta_{\mathcal{A}}(x^+)) \leq \mu^+_{\text{GDP}}(x^-, \delta_{\mathcal{A}}(x^+), \frac{1}{2c}))$.
- Case 2: $\mu_{\text{GDP}}(x^-, \delta_{\mathcal{A}}(x^+)) \leq \mu^+ \leq \mu^+_{\text{GDP}}(x^-, \delta_{\mathcal{A}}(x^+), \frac{1}{2c}))$.
- Case 3: $\mu_{\text{GDP}}(x^-, \delta_{\mathcal{A}}(x^+)) \leq \mu^+_{\text{GDP}}(x^-, \delta_{\mathcal{A}}(x^+), \frac{1}{2c})) < \mu^+$.

In case 1, both of the naive operation and the optimized operation will update μ^+ to $\mu^+_{\text{GDP}}(x^-, \delta_{\mathcal{A}}(x^+), \frac{1}{2c}))$.

In case 2, the optimized operation will do nothing, because the test $\delta_{\mu^+}(x^-) < \delta_{\mathcal{A}}(x^+)$ will fail. The naive operation will update μ^+ due to the error of binary search, which should be avoided.

In case 3, the optimized operation will do nothing, because the test $\delta_{\mu^+}(x^-) < \delta_{\mathcal{A}}(x^+)$ will fail. The naive operation will also do nothing because the max operator will choose μ^+.

To sum up, the optimized operation always give a more accurate update.

The second insight is that we want to avoid case 1 because only in case 1 a binary search is needed. Notice that case 1 happens only if $\delta_{\mu^+}(x^-) < \delta_{\mathcal{A}}(x^+)$, which is equivalent to $\mu^+ < \mu_{\text{GDP}}(x^-, \delta_{\mathcal{A}}(x^+))$. In the $k+1$ round of loop, the condition $\mu^+ < \mu_{\text{GDP}}(x^-, \delta_{\mathcal{A}}(x^+))$ holds true only if for all $j \in \{0, \cdots, k\}$, $\mu_{\text{GDP}}(x_j^-, \delta_{\mathcal{A}}(x_j^+)) < \mu_{\text{GDP}}(x^-, \delta_{\mathcal{A}}(x^+))$, where x_j^- and x_j^+ are the values of x^- and x^+ in the round j. This inspire us to shuffle x_i before iteration because after shuffling, the probability of "$\mu_{\text{GDP}}(x_j^-, \delta_{\mathcal{A}}(x_j^+)) < \mu_{\text{GDP}}(x^-, \delta_{\mathcal{A}}(x^+))$ for all $j \in \{0, \cdots, k\}$" will be $\frac{1}{k+1}$. The expected occurrence of case 1 will be $\sum_{k=0}^{n+1} \frac{1}{k+1} = O(\log(n))$.

The time complexity of shuffling $\mathcal{S}$ is $O(n) = O(\epsilon_h c)$. Each binary search has a time complexity of $O(\log(c))$ and the expected number of binary searches is $O(\log(\epsilon_h c))$. The overall time complexity of the optimized algorithm is therefore $O(\epsilon_h c + \log(c)\log(c\epsilon_h))$=$O(\epsilon_h c)$.

2.8.4 Additional plots

2.8.5 The Laplace mechanism under GDP

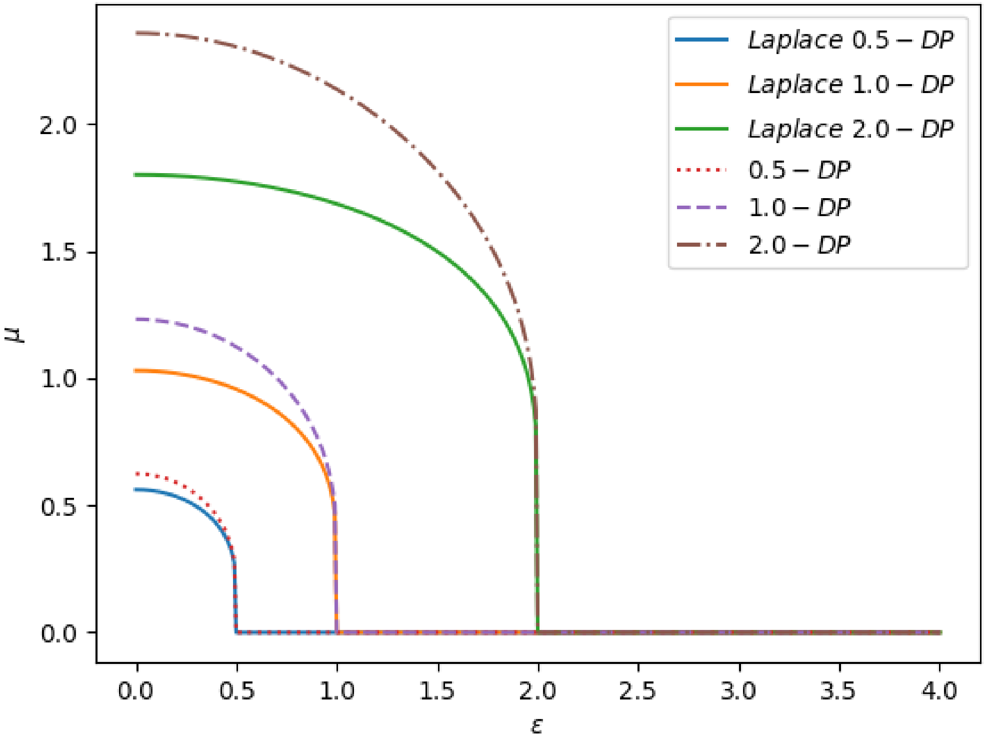


Figure 2.4: The plot of GDPT of ϵ-DP privacy profiles and the Laplace mechanisms with the same ϵ-DP guarantee.

From the figure we can see the privacy protection provided by the Laplace mechanisms is slightly better than ϵ-DP.

2.8.6 The effect of subsampling

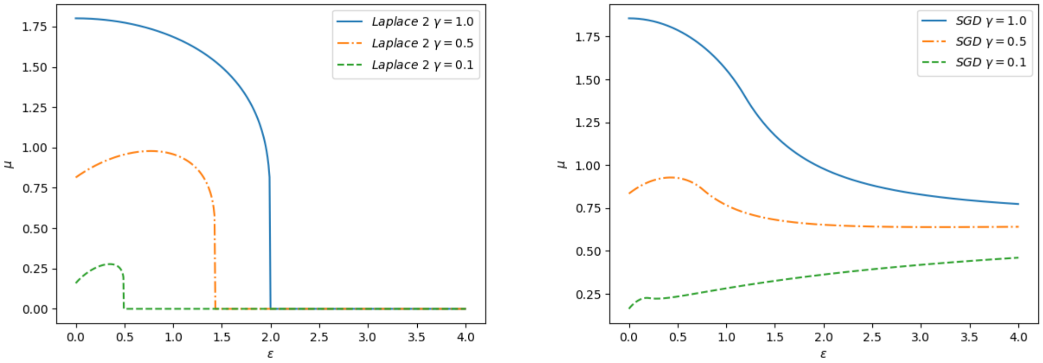
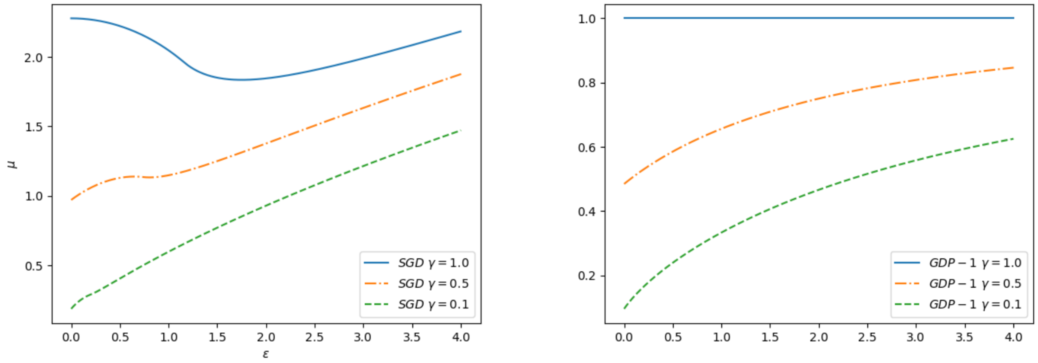


Figure 2.5: Left: A plot of the GDPT of the Laplace mechanism for various of γ. Right: A plot of the GDPT of the SGD for various of γ.



The Poisson subsampling procedure can significantly decrease the value of μ around $\epsilon = 0$ but has little effect on the GDPT's tail.

Figure 2.6: Left: A plot of the GDPT of the ICEA for various of γ. Right: A plot of the GDPT of the δ_μ for various of γ.

References

[1] A. R. Miller, “Personal privacy in the computer age: The challenge of a new technology in an information-oriented society,” *Michigan Law Review*, vol. 67, no. 6, pp. 1089–1246, 1969.

[2] E. Shils, “Privacy: Its constitution and vicissitudes,” *Law and Contemporary Problems*, vol. 31, no. 2, pp. 281–306, 1966.

[3] A. Narayanan and V. Shmatikov, “How to break anonymity of the Netflix prize dataset,” *arXiv preprint cs/0610105*, 2006.

[4] A. Narayanan and V. Shmatikov, “Robust de-anonymization of large sparse datasets,” in *2008 IEEE Symposium on Security and Privacy (sp 2008)*, IEEE, 2008, pp. 111–125.

[5] M. Barbaro and J. T. Zeller, “A face is exposed for aol searcher no. 4417749,” *New York Times (Aug, 9, 2006)*, 2006.

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

[7] F. K. Dankar and K. El Emam, “The application of differential privacy to health data,” in *Proceedings of the 2012 Joint EDBT/ICDT Workshops*, 2012, pp. 158–166.

[8] Ú. Erlingsson, V. Pihur, and A. Korolova, “Rappor: Randomized aggregatable privacy-preserving ordinal response,” in *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, 2014, pp. 1054–1067.

[9] G. Cormode, S. Jha, T. Kulkarni, N. Li, D. Srivastava, and T. Wang, “Privacy at scale: Local differential privacy in practice,” in *Proceedings of the 2018 International Conference on Management of Data*, 2018, pp. 1655–1658.

[10] M. U. Hassan, M. H. Rehmani, and J. Chen, “Differential privacy techniques for cyber physical systems: A survey,” *IEEE Communications Surveys & Tutorials*, vol. 22, no. 1, pp. 746–789, 2019.

[11] L. Wasserman and S. Zhou, “A statistical framework for differential privacy,” *Journal of the American Statistical Association*, vol. 105, no. 489, pp. 375–389, 2010.

[12] J. Dong, A. Roth, and W. J. Su, “Gaussian differential privacy,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2021.

[13] I. Mironov, “Rényi differential privacy,” in *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, IEEE, 2017, pp. 263–275.

[14] M. Bun and T. Steinke, “Concentrated differential privacy: Simplifications, extensions, and lower bounds,” vol. 9985, Nov. 2016, pp. 635–658, ISBN: 978-3-662-53640-7. DOI: 10.1007/978-3-662-53641-4_24.

[15] M. Bun, C. Dwork, G. N. Rothblum, and T. Steinke, “Composable and versatile privacy via truncated cdp,” in *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, 2018, pp. 74–86.

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

[17] B. Balle, G. Barthe, and M. Gaboardi, “Privacy profiles and amplification by subsampling,” *Journal of Privacy and Confidentiality*, vol. 10, no. 1, 2020.

[18] N. Phan, X. Wu, H. Hu, and D. Dou, “Adaptive laplace mechanism: Differential privacy preservation in deep learning,” in *2017 IEEE International Conference on Data Mining*, IEEE, 2017, pp. 385–394.

[19] Y. Hu, P. Liu, L. Kong, and D. Niu, “Learning privately over distributed features: An admm sharing approach,” *arXiv preprint arXiv:1907.07735*, 2019.

[20] X. Xu, Y. Yao, and L. Cheng, “Deep learning algorithms design and implementation based on differential privacy,” in *International Conference on Machine Learning for Cyber Security*, Springer, 2020, pp. 317–330.

[21] T. Li and C. Clifton, “Differentially private imaging via latent space manipulation,” *arXiv preprint arXiv:2103.05472*, 2021.

[22] M. Gaboardi, H. Lim, R. Rogers, and S. Vadhan, “Differentially private chi-squared hypothesis testing: Goodness of fit and independence testing,” in *International Conference on Machine Learning*, PMLR, 2016, pp. 2111–2120.

[23] R. Bassily, A. Smith, and A. Thakurta, “Private empirical risk minimization: Efficient algorithms and tight error bounds,” in *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, IEEE, 2014, pp. 464–473.

[24] M. Abadi *et al.*, “Deep learning with differential privacy,” in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016, pp. 308–318.

[25] V. Feldman, I. Mironov, K. Talwar, and A. Thakurta, “Privacy amplification by iteration,” in *2018 IEEE 59th Annual Symposium on Foundations of Computer Science*, IEEE, 2018, pp. 521–532.

[26] G. Qiao, W. Su, and L. Zhang, “Oneshot differentially private top-k selection,” in *Proceedings of the 38th International Conference on Machine Learning*, M. Meila and T. Zhang, Eds., ser. Proceedings of Machine Learning Research, vol. 139, PMLR, 2021, pp. 8672–8681.

[27] Y. Ishai, E. Kushilevitz, R. Ostrovsky, and A. Sahai, “Cryptography from anonymity,” in *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS’06)*, IEEE, 2006, pp. 239–248.

[28] B. Ghazi, R. Pagh, and A. Velingker, “Scalable and differentially private distributed aggregation in the shuffled model,” *arXiv preprint arXiv:1906.08320*, 2019.

[29] B. Balle, J. Bell, A. Gascón, and K. Nissim, “Private summation in the multi-message shuffle model,” in *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, 2020, pp. 657–676.

[30] B. Ghazi, P. Manurangsi, R. Pagh, and A. Velingker, “Private aggregation from fewer anonymous messages,” *Advances in Cryptology–EUROCRYPT 2020*, vol. 12106, p. 798, 2020.

[31] M. A. Malcolm, “On accurate floating-point summation,” *Communications of the ACM*, vol. 14, no. 11, pp. 731–736, 1971.

[32] A. Cuyt, B. Verdonk, S. Becuwe, and P. Kuterna, “A remarkable example of catastrophic cancellation unraveled,” *Computing*, vol. 66, no. 3, pp. 309–320, 2001.

[33] J. Murtagh and S. Vadhan, “The complexity of computing the optimal composition of differential privacy,” in *Theory of Cryptography Conference*, Springer, 2016, pp. 157–175.

[34] P. Kairouz, S. Oh, and P. Viswanath, “The composition theorem for differential privacy,” in *International conference on machine learning*, PMLR, 2015, pp. 1376–1385.

[35] C. Dwork, G. N. Rothblum, and S. Vadhan, “Boosting and differential privacy,” in *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, IEEE, 2010, pp. 51–60.

[36] M. Abramowitz, I. A. Stegun, and R. H. Romer, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, 1988.

Chapter 3

Paper 2: Online Local Differential Private Quantile Inference via Self-normalization

3.1 Abstract

Based on binary inquiries, we developed an algorithm to estimate population quantiles under Local Differential Privacy (LDP). By self-normalizing, our algorithm provides asymptotically normal estimation with valid inference, resulting in tight confidence intervals without the need for nuisance parameters to be estimated. Our proposed method can be conducted fully online, leading to high computational efficiency and minimal storage requirements with $\mathcal{O}(1)$ space. We also proved an optimality result by an elegant application of one central limit theorem of Gaussian Differential Privacy (GDP) when targeting the frequently encountered median estimation problem. With mathematical proof and extensive numerical testing, we demonstrate the validity of our algorithm both theoretically and experimentally.

3.2 Introduction

Personal data is currently widely used for various purposes, such as facial recognition, personalized advertising, medical trials, and recommendation systems to name a few. While there are potential benefits, it is important also to consider the risks associated with handling sensitive personal information. For instance, research on diabetes can provide valuable insights that may benefit society as a whole in the long term. However, it is crucial to keep in mind that participants may suffer direct consequences if their data is not properly protected through controlled disclosure, such as a rise in health insurance premiums.

The concept of Differential Privacy (DP; [6]) has been successful in providing a rigorous condition for controlled disclosure by bounding the change in the distribution of outputs of a query made on a dataset under the alteration of one data point. This has led to a vast amount of literature under the umbrella of DP, resulting in various generalizations, tools, and applications. However, while enjoying the mathematically solid guarantee of DP and its variants, concerns about a weak link in the process, the trusted curator, are beginning to arise.

The use of trusted curators undermines the spirit of the solid cryptographic level of privacy protection that DP provides. This risk is not limited to information security breaches and rogue researchers but also includes legal proceedings where researchers may be compelled to hand over the raw data, breaking the initial promise made to DP at the time of data collection. Two concepts, Local Differential Privacy (LDP) and pan-DPs, are proposed as solutions. The pan-DP directly counters this issue by solidifying the algorithm to withstand multiple announced intrusions (subpoenas) or one unannounced intrusion (hackers). The concept of LDP was first introduced formally by [37], but its early form can be traced back to [38] and [39], in the name of ”amplification” and ”randomized response survey,” respectively.

In LDP settings, the sensitive information never leaves the control of the users unprotected. The users encode and alter their data locally before sending them to an untrusted central data server for further analysis and computation. Recently, in [40] unveiled a connection between pan-DP and LDP by considering variants of pan-DP framework that can defend against multiple unannounced intrusions. Surprisingly, this requirement can only be fulfilled if the data is scrambled before it leaves the owner’s control, which goes back to the definition of LDP. For better privacy protection, many big tech companies have already implemented LDP into their products, such as Google [8] and Microsoft [41].

This discovery rekindled the research interest in LDP. Researchers have begun to consider fundamental statistical problems, such as estimating parameters, modeling, and hypothesis testing under this constraint. The quantiles, including the median, are basic summary statistics that have been widely studied within the framework of differential privacy. Early research in this area includes the estimation of quantiles under the central DP setting, as presented in [42] and [43]. More recent advancements, such as [44], have proposed a rate-optimal sample quantile estimator that does not rely on the evaluation of histograms. [45] further extended this research by estimating multiple quantiles simultaneously. Despite these advances, the quantile estimation under the central DP setting remains an active area of research, with new work in various applications such as [46] and [47].

In the central DP setting, a trusted curator can acquire the actual sample quantiles and other summary statistics, with the only limitation being that the release of the output must conform to the DP condition. However, under the local DP setting, the curator does not have access to the true data and can only see proxies generated by the users. This makes it more challenging to design local DP algorithms that can provide valid results leading to greater problems in developing corresponding theoretical properties and providing further statistical inference.

Researchers often propose consistent estimators for the parameters of interest and derive the asymptotic normality. However, these estimators often involve nuisance parameters that are not trivial to obtain or estimate, making them difficult to deploy in real-world scenarios. To address this issue, [48] developed the methodology of self-normalization for constructing confidence intervals. This method involves designing a statistic called the self-normalizer, which is proportional to the nuisance parameters, and making the original estimate a pivotal quantity by placing it and the self-normalizer in the numerator and the denominator, thereby canceling out the nuisance parameters and leading to an asymptotically pivotal distribution. This methodology provides a powerful tool for statistical inference under complex data, particularly in the context of LDP frameworks where obtaining accurate original data or consistently estimating nuisance parameters without an additional privacy budget is challenging.

Efficient computation is essential for the practicality of LDP algorithms, as large sample sizes are necessary to counteract the effects of local perturbations and achieve optimal performance. Meanwhile, online computation is another valuable attribute of LDP algorithms, as it reduces storage requirements and diminishes risks associated with information storage. Early attempts of introduce online computation to DP algorithms can be traced back to [49], where additive Gaussian noise was injected into the gradient to provide DP protection. Later, [50] gives an online linear optimization DP algorithm that with optimal regret bounds. The concept of online computation has also been incorporated into federated learning, as discussed by [51]. More recently, [52] has facilitated online computation for a random scaling quantity using only the trajectory of stochastic optimization, effectively eliminating the need for past state storage and enhancing computational efficiency. In contrast to traditional studies on DP online algorithms, our emphasis is on harnessing online computation for convenience. Our theoretical analysis concentrates on the statistical properties of the proposed estimators, encompassing aspects such as consistency, asymptotic normality, and more.

In this paper, our contributions are listed as follows.

- We propose a new LDP algorithm for population quantile estimation that does not require a trusted curator. Under some mild conditions, we derive the consistency and asymptotic normality of the proposed quantile estimator.

- We construct the confidence interval of the population quantiles via self-normalization, which eliminates the need for estimating the asymptotic variance in the limiting distribution. Furthermore, this procedure can be implemented online without storing all past statuses.

- We also discuss the optimality of the proposed algorithm. By combining it with the central limit theorem of GDP, we demonstrate that our algorithm for median estimation achieves the lower bound of asymptotic variance among all median estimators constructed by a binary random response-based sequential interactive mechanism under LDP.

The structure of this paper is as follows. We begin by providing an overview of the concepts of central DP and LDP. Then present our proposed methodology, detailing the algorithms and their corresponding theoretical results. Finally, we provide experimental results to demonstrate the effectiveness of our approach.

3.3 Preliminaries

3.3.1 Central Differential Privacy

Definition 3.1 *[6] A randomized algorithm $\mathcal{A}$, taking a dataset consisting of individuals as its input, is (ϵ, δ)-differentially private if, for any pair of datasets S and S' that differ in the record of a single individual and any event E, satisfies the below condition:*

$$\mathbb{P}[\mathcal{A}(S) \in E] \leq e^{\epsilon} \mathbb{P}\left[\mathcal{A}\left(S'\right) \in E\right] + \delta.$$

When $\delta = 0$, $\mathcal{A}$ is called ϵ-Differentially Private (ϵ-DP).

The concept of DP only imposes constraints on the output distribution of an algorithm $\mathcal{A}$, rather than placing restrictions on the credibility of the entity running the algorithm or protecting the internal states of $\mathcal{A}$. The existence of the curator who has access to the raw data set is why this approach is known as "Central" DP. The curator simplifies the algorithm design and often leads to an asymptotically negligible loss of accuracy from privacy protection [53].

3.3.2 Local Differential Privacy

Despite the varying definitions of LDP due to the level of interactions, all of them depend on the following concept called (ϵ, δ)-randomizer.

Definition 3.2 *[54] An (ϵ, δ)-randomizer $R : X \to Y$ is an (ϵ, δ)-differentially private function taking a single data point as input.*

The definition of randomizer is mathematically a special case of the central DP. The main difference between the central and local DP is the role of the curator, which is further determined by the level of interactions allowed. In LDP, the curator coordinates interactions between n users, each of whom holds their own private information X_i. In each round of interaction, the curator selects a user and assigns them a randomizer R_t. If the (ϵ, δ) parameters are allowed by the experiment setting, the user will run the randomizer on their private information and release the output to the curator.

The level of interactions can vary from full-interactive, where the curator can choose the randomizer and the next user based on all previous interactions, to sequential (also called one-shot) interactive, where the curator is not allowed to pick one user twice but is still able to adaptively picking the next the user-randomizer pairs based on all previous interactions, to non-interactive, where adaptivity is forbidden, and all user-randomizer pairs must be determined before any information is collected. If the curator is further forbidden from varying the randomizer R and tracking back outputs to a specific user, it will lead to another interesting setting called shuffle-DP [55].

3.3.3 Notations

In this paper, we employ the following notations. $\mathbf{1}_{\{.\}}$ is the indicator function and $[a]$ denotes the largest integer that does not exceed a. $\mathcal{O}$ (or $\mathcal{o}$) denotes a sequence of real numbers of a certain order. For instance, $\mathcal{o}(n^{-1/2})$ means a smaller order than $n^{-1/2}$, and by $\mathcal{O}_{a.s.}$ (or $\mathcal{o}_{a.s.}$) almost surely $\mathcal{O}$ (or $\mathcal{o}$). For sequences a_n and b_n, denote $a_n \asymp b_n$ if there exist positive contants c and C such that $cb_n \leq a_n \leq Cb_n$. The symbol $\xrightarrow{d}$

means weak convergence or converge in distribution.

3.4 Algorithm and Main Results

3.4.1 Algorithm

Let $x_1, \dots, x_n, \dots$ be independently and identically distributed(i.i.d.) random variables defined on $\mathbb{R}$ representing private information of each user, with target quantile τ and corresponding true value Q, i.e., $\mathbb{P}(x_i \leq Q) = \tau$. To ensure the uniqueness of quantiles, we assume the x_i's are continuous random variables, with positive density on the target quantile. In practice, we can perturb the data by a small amount of additive data-independent noise to remove atoms in the distribution as is in [45].

The design of the local randomizer is crucial for LDP mechanisms as it must properly choose the inquiry to the user in order to maximize the gathering of information related to the estimation of the target quantile without violating privacy conditions. The population quantiles can be considered as a minimizer of the check loss function:

$$l_\tau(x, \theta) = \begin{cases} \tau(x - \theta) & \text{if } x \geq \theta \\ (\tau - 1)(x - \theta), & \text{if } x < \theta \end{cases}.$$

In the non-DP case, a known solution is the use of stochastic gradient descent, as outlined in [56]. It is important to note that for each point, the gradient it contributes is purely determined by the binary variable representing whether the value is greater than θ or not. This motivates us to modify the stochastic gradient descent process by adding a local randomization process, resulting in the Algorithm 4 and 5 outlined below:

In Algorithm 4, generating randomness of v before the if-condition fork may seem wasteful, but it prevents side-channel attacks such as inferring the true value based on the timing of response [57, 58]. Algorithm 5 collects random responses and generates the next inquiry accordingly. Therefore, Algorithm 5 satisfies the definition of sequential interactive local DP.

Algorithm 4: Locally Randomized Compare (LRC)

```
Input: Inquiry $q$, response rate $r$, private data $x$
$u \sim Bernoulli(r)$
$v \sim Bernoulli(0.5)$
if $u = 1$ then
    return $\mathbf{1}_{x>q}$
else
    return $v$
end if
```

Algorithm 5: Main Algorithm

```
Input: Step sizes $d_n$, target quantile $\tau \in (0,1)$, truthful response rate $r$
Initialize: $n \leftarrow 0$, $q_0 \leftarrow 0$, $v_0^a \leftarrow 0$, $v_0^b \leftarrow 0$, $Q_0 \leftarrow 0$
repeat
    $n \leftarrow n + 1$
    Inquire: $s \leftarrow LRC(q_{n-1}, r, x_n)$
    if $s$ is 1 then
        $$q_n \leftarrow q_{n-1} + \frac{1 - r + 2\tau r}{2} d_n$$
    else
        $$q_n \leftarrow q_{n-1} - \frac{1 + r - 2\tau r}{2} d_n$$
    end if
    $Q_n = ((n-1)Q_n + q_n)/n$
    $v_n^a \leftarrow v_{n-1}^a + n^2 Q_n^2$
    $v_n^b \leftarrow v_{n-1}^b + n^2 Q_n$
    Destroy $v_{n-1}^a, v_{n-1}^b, Q_{n-1}, q_{n-1}$

until forever
```

The following algorithm can be used when estimations and confidence intervals are required. These values are not calculated at every step to minimize computational expenses.

```
Algorithm 6: Generate Confidence Interval
Input: Internal states of Algorithm 5: $n$, $Q_n$, $v_n^a$, $v_n^b$
$N_n \leftarrow n^{-1}\left(v_n^a - 2Q_n v_n^b + Q_n^2 n(n+1)(2n+1)/6\right)$
$W \leftarrow n^{-1}\mathcal{U}_{1-\alpha/2}\sqrt{N_n}$
Return: Confidence interval $(Q_n - W, Q_n + W)$
```

The use of dichotomous inquiry in data privacy brings multiple advantages. One benefit is the reduced communication cost, as it only takes one bit to respond. Additionally, the binary response can make full use of the DP budget, as opposed to methods such as the Laplace mechanism, which may provide unnecessary privacy guarantees beyond ϵ-DP, as outlined in Theorem 3 in [59] and Theorem 2.1 in [60].

Furthermore, people tend to be more comfortable answering dichotomous questions compared to open-ended ones [61] as they present a choice between two options and may be perceived as less threatening than open-ended questions, which require more detailed and nuanced responses. In addition, the binary approach is easy to understand for users. With the proper choice of truthful response rate r, the algorithm known as the random response can be easily simulated through coin flips or dice rolls, allowing users to understand it fully and are able to ”run” it without the help of electronic devices. This is in contrast to a DP mechanism involving the usage of random distribution on real numbers. Due to the finite nature of the computer, the imperfection of floating-point arithmetic leads to serious risks with effective exploits. For more information, please refer to [62–64].

Before discussing the specific characteristics of our estimator, we will first demonstrate its performance through a sample trajectory. The experiment is conducted with a truthful response rate with $r = 0.5$, which means half of the responses are purely random. The objective is to estimate the median from i.i.d. samples. The true underlying distribution is a standard normal distribution.

It can be seen that from Figure 3.1, the proposed estimator converges to the true value, and both infeasible and proposed confidence intervals, defined later, contain the true value at a slightly larger sample size. Also, the proposed confidence intervals are highly competitive with the infeasible one in width. Refer to Figure 3.5 and 3.6 for convergence trajectories under different initialization or target quantiles.

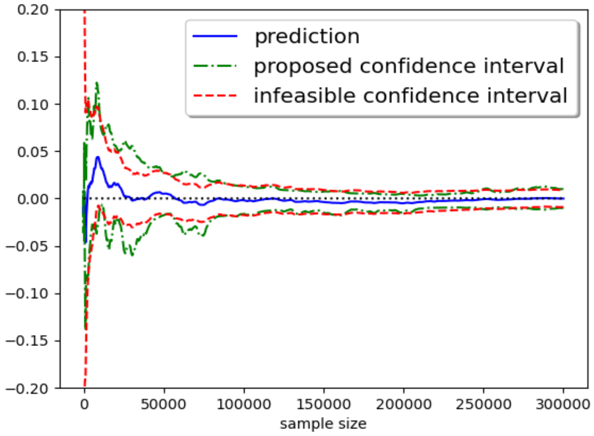


Figure 3.1: A sample trajectory of estimator Q_n, infeasible confidence interval (3.2) and proposed confidence interval (3.3). The horizon dotted line is the true value $Q = 0$.

Next, we show the LDP property of our algorithm:

Theorem 3.3 *Algorithm 4 is an* $(\epsilon, 0)$*-randomizer with* $\epsilon = \log((1 + r)/(1 - r))$*.*

Proof. see Appendix 3.7.3 ■

The algorithm presented in Algorithm 5 adaptively selects the next randomizer, determined by the parameter q in Algorithm 4, based on its internal state q_n. However, it never revisits previous users. As a result, Algorithm 5 satisfies sequential interactive $(\epsilon, 0)$-LDP, where $\epsilon = \log((1 + r)/(1 - r))$ (equivalently, $r = (e^\epsilon - 1)/(e^\epsilon + 1) = \tanh(\epsilon/2)$).

Throughout the remainder of this paper, we will use the truthful response rate r to represent the privacy budget, as opposed to the more standard ϵ. This choice is made for the following reasons:

In the context of LDP, it is crucial to ensure understanding and acceptance by end-users who may not possess expertise in the field. The truthful response rate, denoted by r, has a more intuitive interpretation. Additionally, r appears in multiple results presented in this paper, and maintaining this form allows for a more direct presentation. If necessary, the results can be easily converted by replacing all instances of r with $\tanh\left(\frac{\epsilon}{2}\right)$. For a conversion table, please refer to Table 3.5.

3.4.2 Consistency

To discuss the asymptotic properties of estimator Q_n, we rewrite it as a recursive equation. Let $\{U_n\}$ and $\{V_n\}$ be the i.i.d. Bernoulli sequences with

$$\mathbb{P}(U_n = 1) = r,\ \mathbb{P}(U_n = 0) = 1 - r,$$
$$\mathbb{P}(V_n = 1) = \mathbb{P}(V_n = 0) = 1/2.$$

For $q_0 \in \mathbb{R}$,

$$\begin{aligned} q_{n+1} = q_n &+ \frac{1 - r + 2r\tau}{2d_n}\left(\mathbf{1}_{x_{n+1}>q_n} U_n + (1 - U_n)V_n\right) \\ &- \frac{1 + r - 2r\tau}{2d_n}\left(\mathbf{1}_{x_{n+1}<q_n} U_n + (1 - U_n)(1 - V_n)\right), \end{aligned} \tag{3.1}$$

where the step size $\{d_n\}_{n=1}^{\infty}$, satisfies

$$\sum_{n=1}^{\infty} d_n = \infty, \qquad \sum_{n=1}^{\infty} d_n^2 < \infty.$$

The step size d_n is vital for the convergence of q_n, but it has a relatively minor effect on Q_n. The following theorem guarantees consistency:

Theorem 3.4 *For increasing positive number γ_n, satisfied*

$$\frac{\gamma_n}{\gamma_{n-1}} = 1 + o(d_n), \qquad \sum_{n=1}^{\infty} d_n^2 \gamma_n^2 < \infty,$$

the n-step output q_n enjoys that

$$\gamma_n \left| q_n - Q \right| = o_{a.s.}(1).$$

Proof. see Appendix 3.7.4. ■

In particular, if $d_n \asymp a/n^{\beta}$, for some constant $a > 0$ and $\beta \in (1/2, 1)$, then $\gamma_n \asymp n^{\gamma}$ for some $\gamma < \beta - 1/2$, and for the sake of simplicity, we will set the step sizes as $d_n \asymp a/n^{\beta}$.

3.4.3 Asymptotic Normality

Next, the asymptotic normality will be discussed.

Theorem 3.5 *If $\beta \in (0, 1)$, then*

$$\sqrt{n}\,(Q_n - Q) \xrightarrow{d} N\left(0, \frac{1 - r^2(1 - 2(1 - \tau))^2}{4r^2 f_X^2(Q)}\right),$$

where $f_X^2(Q)$ is the value on Q for density function of X.

Proof. see Appendix 3.7.4. ■

Noticed that the conditions on β in Theorem 3.4 and Theorem 3.5 are different. It is possible that q_n fails to converge to Q, but Q_n still enjoys asymptotic normality. Following Theorem 3.5, one constructs the confidence interval of Q, if $f_X(Q)$ can be obtained or estimated by $\widehat{f_X(Q)}$. Denote $z_{1-\alpha}$ as the upper $\alpha-$quantile of standard normal distribution. The infeasible confidence interval with significance level α is:

$$\begin{aligned}&\Big(Q_n - z_{1-\alpha}\sqrt{n(1 - r^2(1 - 2(1 - \tau))^2)}/(2r\widehat{f_X(Q)}),\\ &Q_n + z_{1-\alpha}\sqrt{n(1 - r^2(1 - 2(1 - \tau))^2)}/(2r\widehat{f_X(Q)})\Big).\end{aligned} \tag{3.2}$$

However, obtaining a consistent estimator $\widehat{f_X(Q)}$, such as using non-parametric methods under our differential privacy framework, is not straightforward, since we can only obtain the binary sequence $\mathbf{1}_{x_n > q_{n-1}}$ for protecting privacy, and the original data set $x_1, \ldots, x_n$ cannot be accessed directly.

An alternative approach to estimate the nuisance parameter $f_X(Q)$ is through the use of bootstrap methods to simulate the asymptotic distribution. Traditional bootstrap methods that rely on re-sampling are not suitable for the stochastic gradient descent method because of failing to recover the special dependence structure defined in (3.1).

Recently, [65] proposed online bootstrap confidence intervals for stochastic gradient descent, which involve recursively updating randomly perturbed stochastic estimates. Although this approach performs well when there are no constraints on DP, it requires multiple interactions with the users and will therefore blow up the privacy budget.

3.4.4 Inference via Self-normalization

To overcome the difficulties above, we propose a novel inference procedure of quantiles under the LDP framework via self-normalization, which will avoid estimating the nuisance parameter $f_X(Q)$. We hope to construct an estimator that is proportional to the nuisance parameters. To approach that, we will first establish further theoretical properties of the proposed estimator Q_n. Define the process $S_{[nt]} = \sum_{i=1}^{[nt]} q_i$, $t \in [0,1]$.

Theorem 3.6 *If $\beta \in (0,1)$, then*

$$n^{-1/2}(S_{[nt]} - nQ) \xrightarrow{d} \frac{\sqrt{(1 - r^2(1 - 2(1-\tau))^2)}}{2rf_X(Q)} W(t),$$

where $W(t)$ is the Brownian motion in $(C[0,1], \mathbb{R})$.

Proof. see Appendix 3.7.4. ■

Noticed that Theorem 3.5 is the special case in Theorem 3.6 when $t = 1$. Then, following Theorem 3.6, we define the self-normalizer:

$$N_n = \int_0^1 \left(S_{[nt]} - [nt]Q_n\right)^2 dt,$$

By the continuous mapping theorem, we can derive:

$$\frac{n^{-1/2}(S_n - nQ)}{\sqrt{n^{-1}N_n}} \xrightarrow{d} \mathcal{S} := \frac{W(1)}{\sqrt{\int_0^1 (W(t) - tW(1))^2 dt}},$$

where the asymptotical distribution $\mathcal{S}$ is not associated with any unknown parameters, and its quantile can be computed by Monte Carlo simulation. Therefore, we have constructed an asymptotical pivotal quantity. Denote $\mathcal{U}_{1-\alpha}$ the $1-\alpha$ quantile of $\mathcal{S}$, the $1-\alpha$ self-normalized confidence interval of Q is constructed by:

$$\left(Q_n - n^{-1}\mathcal{U}_{1-\alpha/2}\sqrt{N_n}, Q_n + n^{-1}\mathcal{U}_{1-\alpha/2}\sqrt{N_n}\right). \tag{3.3}$$

As noted by [66], the distribution of $\mathcal{S}$ has a heavier tail than that of the standard normal distribution, which is analogous to the heavier tail of $t-$distribution compared to the standard normal distribution, resulting in a wider but not conservative corresponding confidence interval. However, the average width of the confidence interval constructed through self-normalization is not excessively large when compared to the infeasible confidence interval, as demonstrated by numerical experiments in Figure 3.1. Furthermore, the construction of an asymptotic pivotal quantity is not unique. See Appendix 3.7.2 for examples of other possibilities.

Whether there are theoretical advantages between the different constructions of self-normalizer is still open to discussion, but according to [52], the proposed self-normalizer can be computed in a fully online fashion and is computationally efficient, as outlined in Algorithm 5 and 6. The algorithm only needs to store a single integer n and four float numbers: v_n^a, v_n^b, q_n, Q_n and conduct only a dozen of arithmetic operations.

3.4.5 Discussion of Optimality

In this subsection, we will discuss the optimality of the proposed algorithm. To generalize the setting, we consider all binary random response-based sequential interactive mechanisms. The random response mechanism can be written as the following $K : \{0, 1\} \to \{0, 1\}$:

$$K(x) = \begin{cases} 0, & \text{w.p. } (1-r)/2, \\ 1, & \text{w.p. } (1-r)/2, \\ x, & \text{w.p. } r. \end{cases}$$

Let $\{T_1, \cdots, T_n\}$ be a collection of binary query functions, which means $T_i(x) = \mathbf{1}_{x \in C_i}$, for some subset $C \subset \mathbb{R}$. In the sequential interactive LDP setting, the curator will generate its output based on the transcript $\{\{K \circ T_1(x_1), \cdots, K \circ T_n(x_n)\}, \{C_1, \cdots, C_n\}\}$ and the choice of C_i may depend on the transcript up to this point:$\{\{K \circ T_1(x_1), \cdots, K \circ T_{i-1}(x_{i-1})\}, \{C_1, \cdots, C_{i-1}\}\}$. Notice that the Algorithm 4 is a special case where $C_i = \{z : z \geq q_{i-1}\}$, and q_{i-1} is given by

$$\sum_{j=1}^{i-1} T_j(x_j)\frac{1-r+2\tau r}{2}d_j - (1-T_j(x_j))\frac{1+r-2\tau r}{2}d_j.$$

We aim to determine a lower bound for the estimation variance. Therefore, any lower bounds derived under specific conditions also serve as a general lower bound for the estimation variance. To demonstrate this, we will present a pair of distributions with distinct medians that are, to the best of our knowledge, the most indistinguishable given randomized binary queries.

Define:

$$H_0 : x_i \sim Laplace(1) \text{ vs. } H_1 : x_i \sim Laplace(1) + \epsilon_n \tag{3.4}$$

Let $\epsilon = \log\left[(e^{\frac{1}{\sqrt{n}}}(r+1)+r-1)/(e^{\frac{1}{\sqrt{n}}}(r-1)+r+1)\right]$. Simple computation yields that for any $(a, b) \in \{0, 1\}^2$

$$\frac{\mathbb{P}(K \circ T_i(x_i) = a|H_b)}{\mathbb{P}(K \circ T_i(x_i) = a|H_{1-b})} \leq \frac{e^{\epsilon_n}(r+1)-r+1}{-e^{\epsilon_n}(r-1)+r+1} = e^{\sqrt{\frac{1}{n}}}. \tag{3.5}$$

Interestingly, if we consider the truth $H \in \{H_0, H_1\}$ as a data set containing only one data point, (3.5) shows that $K \circ T_i$ is $1/\sqrt{n}$-DP. Notice that the transcript is a n-fold adaptive composition [34] of $1/\sqrt{n}$-DP mechanisms. By Theorem 8 [12], the transcript and all post-processing of it (Proposition 4; [12]) asymptotically satisfies the Gaussian Differential Privacy condition with $\mu = 1$ (or briefly 1-GDP).

We will now examine the limit on the best possible variance imposed by the 1-GDP condition. Denote the estimator of median as $\hat{\theta}_n$. First, we will consider asymptotically normal, unbiased, shift-invariant estimators of the median. By restricting our discussion to unbiased, shift-invariant estimators, we ensure that no estimator has an unfair advantage by favoring specific values. Under the null hypothesis, for the standard deviation σ_n of $\hat{\theta}_n$, one has that

$$\frac{\hat{\theta}_n}{\sigma_n} \xrightarrow{d} N(0,1),$$

and under the alternative hypothesis,

$$\frac{\hat{\theta}_n - \epsilon_n}{\sigma_n} \xrightarrow{d} N(0,1).$$

The 1-GDP condition implies that for sufficiently large n, $\epsilon_n/\sigma_n \leq 1$ (see Appendix 3.7.5). By plugging in the values $\epsilon_n = (r\sqrt{n})^{-1} + \mathcal{O}(n^{-3/2})$ and $1/2 = f(F^{-1}(1/2))$, we deduce that:

$$\sigma_n \geq \frac{1}{2r\sqrt{n}f(F^{-1}(1/2))} + \mathcal{O}\left(n^{-1}\right),$$

which gives us an asymptotic lower bound of the variance: $(4r^2nf^2(F^{-1}(1/2)))^{-1}$. This lower bound matches the asymptotic variance obtained in Theorem 3.5, showing the optimality of our approach. Although most estimators we are interested in have an asymptotically normal distribution, we wish to generalize the minimal variance result to other families as the theorem below.

Theorem 3.7 *If $\hat{\theta}_n$ is a median estimator based on the random response of binary-based sequential interactive inquiries such that:*

$$\frac{\hat{\theta}_n - F^{-1}(1/2)}{\sigma_n} \xrightarrow{d} G$$

where G has a log-concave density $f_G(x) \propto \mathrm{e}^{-\varphi(x)}$ on $\mathbb{R}$ such that $\varphi(x) = \varphi(-x)$, $\mathbb{E}\left[(\varphi'(G))^2\right] < +\infty$, and $\mathbb{E}\left[G^2\right] = 1$.

Then,

$$\sigma_n \geq \frac{1}{2r\sqrt{n}f(F^{-1}(1/2))} + \mathcal{O}\left(n^{-1}\right).$$

The minimal variance result can be attributed to two factors. In Appendix 3.7.5, we demonstrate that asymptotic GDP imposes a condition on the variance of estimators that follow a normal distribution. This condition serves as a lower bound for 1-GDP estimators, without relying on any specific mechanism assumption. Secondly, the relaxation from the assumption of normality to milder conditions on the function G is a consequence of Theorem 1.2 in [67]. This theorem establishes that among all μ-GDP estimators satisfying the aforementioned conditions, the variance is lower bounded by $1/\mu^2$. This lower bound is attainable when the underlying distribution is normal.

3.5 Experiments

We evaluate the performance of our algorithms using a variety of distributions. The data come from four cases: standard Normal $N(0,1)$, Uniform $U(-1,1)$, standard Cauchy $C(0,1)$, and PERT distribution [68] with probability density function:

$$f(x) = 0.625(1-x)(1+x)^3, \quad x \in (-1,1).$$

These cases represent situations with heavy tails, compact or non-compact support, and asymmetric distributions commonly found in practice, as shown in Figure 3.2.

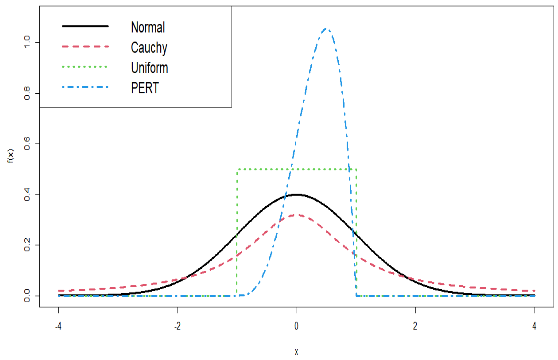


Figure 3.2: Plot of the density function, where the types of lines represent different distribution, solid: Normal, dashed: Cauchy, dotted: Uniform, dot-dash: PERT.

The target quantiles are $\tau = 0.3, 0.5, 0.8$, and the truthful response rate $r = 0.25, 0.5, 0.9$, which the privacy budget is $\epsilon = \log(1 + 2r/(1-r))$ corresponding to $0.51, 1.09, 2.94$ respectively. We use the step sizes $d_n = 2/(n^{0.51} + 100)$ for all experiments, which satisfies the assumptions of Theorem 3.5 and 3.6. The range of sample size n is $(10000, 400000)$, the initial value $q_0 = 0$, and the number of replication is 10000. The results from different sample sizes are independently conducted from scratch to eliminate the correlation among experiments.

To show the consistency of the proposed estimator Q_n, Figure 3.3 displays the box plots of estimator Q_n under Normal distribution with sample size $n = 10000, \dots, 50000$. As the sample size increases, the estimation becomes closer to the true values Q, the corresponding standard errors decay across all settings, and the truthful response rate leads to significantly better performance in small finite sample sizes but has diminishing effects afterward. Meanwhile, we can also see that the proximity between the true target value and the initialization 0 is beneficial to early performances. But in an asymptotic view, the proposed algorithm is insensitive to the initial value selection.

We also demonstrate the empirical coverage rate and mean absolute error of the developed method in Table 3.1. The empirical coverage rate of the proposed method becomes closer to the nominal confidence level as the sample size increases in most cases and the mean absolute error decreases to zero. The corresponding figures and tables of other distributions can be found in Appendix 3.7.1, which describes a similar phenomenon.

Figure 3.4 investigates the performance of the proposed confidence interval in other nominal levels. One can discover that the curves of the empirical coverage rate are getting closer to $y = x$ uniformly, as sample size increases in all privacy budget settings, which reveals the performance of the proposed method is irrelevant to the pre-determined significance level. It is worth noting that when $r = 0.25$, the effective sample size is $1/16$ of the original one, yet the performance of the proposed method remains excellent, which strongly supports the asymptotic theory.

3.6 Conclusion and Future Works

In this paper, we proposed a novel algorithm for estimating population quantiles under the settings of LDP. The core design idea of the algorithm is based on using

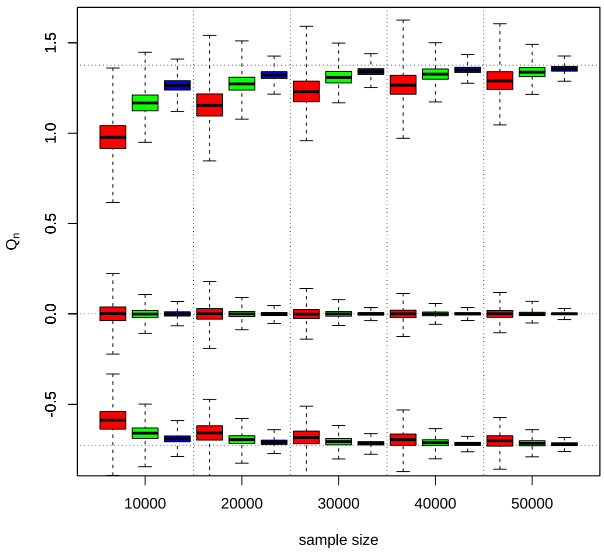
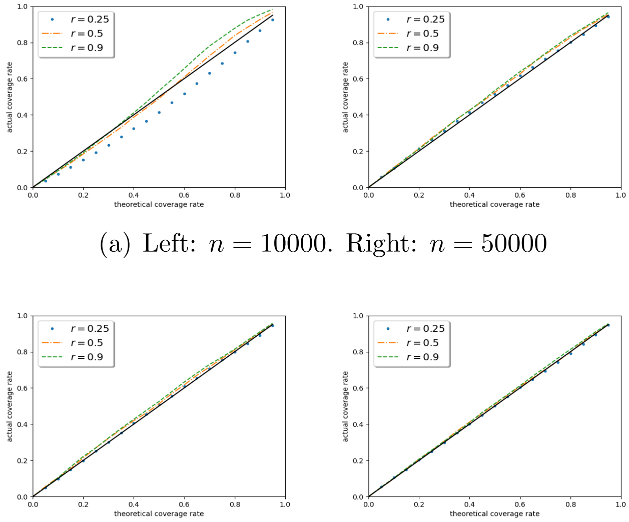


Figure 3.3: Box-plot of estimator Q_n for different target quantiles of Cauchy distribution. In each sample size divided by a vertical dotted line, the three boxes establish results with different privacy budgets by left: $r = 0.25$, middle $r = 0.5$, and right: $r = 0.9$. The horizontal dashed lines represent the true value Q in $\tau = 0.3$, 0.5, 0.8 from the bottom to the top.

dichotomous inquiry. The proposed estimator enjoys excellent theoretical properties, including consistency, asymptotic normality, and optimality in some special cases. Importantly, by applying the technique of self-normalization to cancel out the nuisance parameters, we can construct confidence intervals of population quantiles for statistical inference. Finally, our algorithm is designed in an online setting, making it suitable for handling large streaming data without the need for data storage. Extensive simulation studies reveal a positive confirmation of the asymptotic theory.

Despite the contributions above, this article still leaves many exciting questions unanswered, which opens many avenues for future research. A general tight lower bound for other quantiles under our setting is still undetermined, and we have yet to consider other variants of LDP (e.g. full-interactive). Other directions include exploring data that is not independently and identically distributed, such as time series or spatial series data. Additionally, the quantile of interest may be influenced by



(a) Left: $n = 10000$. Right: $n = 50000$

(b) Left: $n = 100000$. Right: $n = 200000$

Figure 3.4: The curve of the empirical coverage rate of proposed confidence interval (3.3) with nominal significance level, when the data are Normal and target quantile $\tau = 0.3$ under different privacy budget (dotted $r = 0.25$, dot-dash $r = 0.5$ and dashed $r = 0.9$).

other covariates, leading to the study of LDP quantile regression. This paper focuses on estimating quantiles for a specific sample size n, with the potential for developing consistent bounds, resulting in the transition from quantile confidence intervals to confidence sequences.

n	τ	$r = 0.25$	$r = 0.5$	$r = 0.9$
10000	0.3	0.926(0.069)	0.965(0.034)	0.982(0.018)
	0.5	0.834(0.037)	0.897(0.019)	0.911(0.011)
	0.8	0.962(0.121)	0.992(0.058)	0.999(0.031)
20000	0.3	0.936(0.041)	0.958(0.020)	0.971(0.011)
	0.5	0.888(0.027)	0.915(0.014)	0.936(0.008)
	0.8	0.965(0.063)	0.984(0.030)	0.994(0.016)
40000	0.3	0.943(0.025)	0.958(0.013)	0.967(0.007)
	0.5	0.910(0.020)	0.931(0.010)	0.937(0.006)
	0.8	0.966(0.035)	0.978(0.017)	0.984(0.009)
100000	0.3	0.946(0.015)	0.954(0.007)	0.958(0.004)
	0.5	0.929(0.013)	0.944(0.006)	0.941(0.004)
	0.8	0.954(0.019)	0.965(0.009)	0.973(0.005)
200000	0.3	0.947(0.010)	0.951(0.005)	0.956(0.003)
	0.5	0.942(0.009)	0.949(0.004)	0.947(0.002)
	0.8	0.956(0.013)	0.960(0.006)	0.964(0.003)
400000	0.3	0.945(0.007)	0.953(0.004)	0.948(0.002)
	0.5	0.942(0.006)	0.949(0.003)	0.944(0.002)
	0.8	0.952(0.009)	0.957(0.004)	0.958(0.002)

Table 3.1: Empirical results of coverage rate(mean absolute error) of proposed confidence interval (3.3) (estimator Q_n) with data collected from Normal.

3.7 Appendix

3.7.1 Additional figures and tables

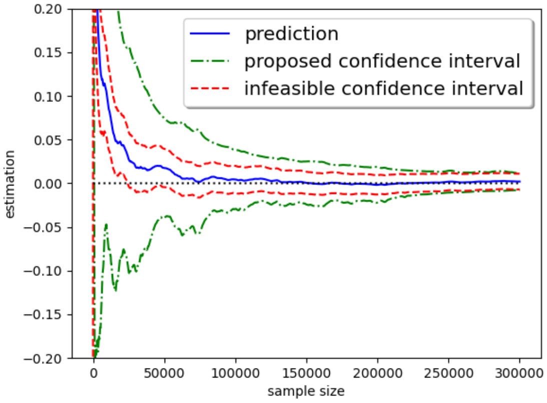


Figure 3.5: An alternative sample trajectory of estimator Q_n using a different initialization $q_0 = 1$.

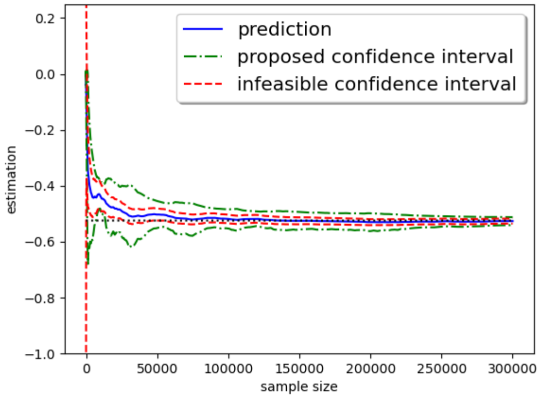


Figure 3.6: An alternative sample trajectory of estimator Q_n using a different target quantile $\tau = 0.3$.

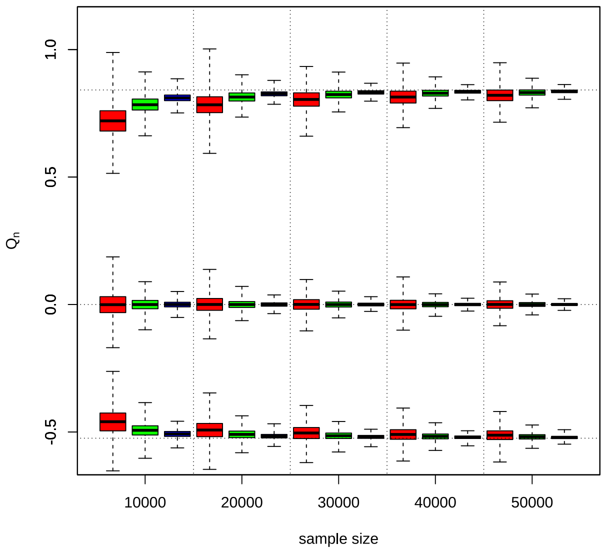


Figure 3.7: Box-plot of estimator Q_n for different target quantile of Cauchy distribution. In each sample size divided by a vertical dotted line, the three boxes establish results with different privacy budgets by left: $r = 0.25$, middle $r = 0.5$, and right: $r = 0.9$. The horizontal dashed lines represent the true value Q in $\tau = 0.3, 0.5, 0.8$ from the bottom to the top.

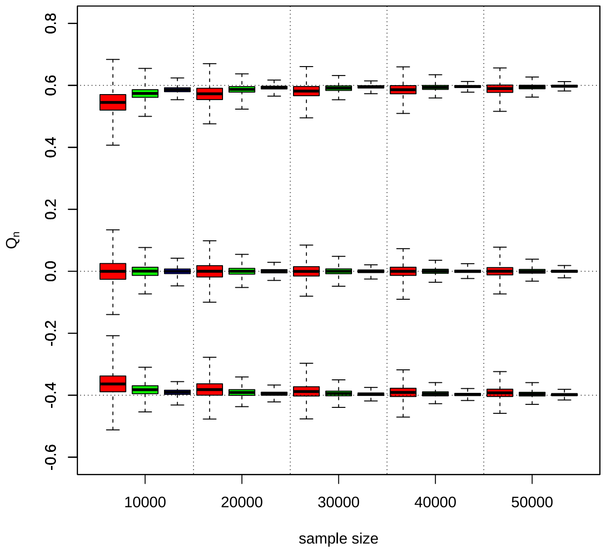


Figure 3.8: Box-plot of estimator Q_n for different target quantile of Uniform distribution. In each sample size divided by a vertical dotted line, the three boxes establish results with different privacy budgets by left: $r = 0.25$, middle $r = 0.5$, and right: $r = 0.9$. The horizontal dashed lines represent the true value Q in $\tau = 0.3, 0.5, 0.8$ from the bottom to the top.

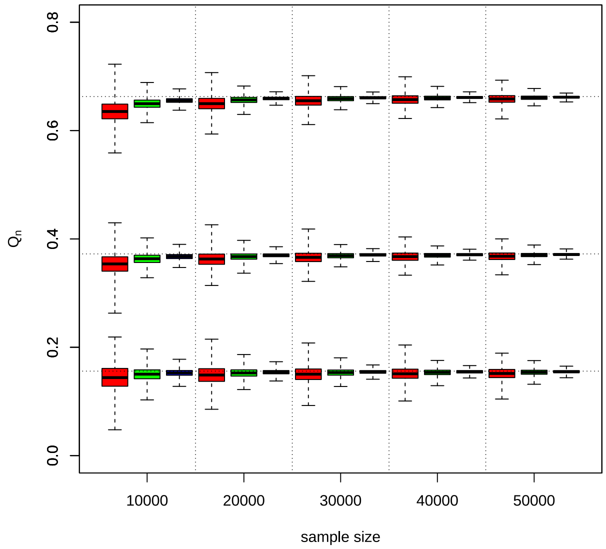


Figure 3.9: Box-plot of estimator Q_n for different target quantile of PERT distribution. In each sample size divided by a vertical dotted line, the three boxes establish results with different privacy budgets by left: $r = 0.25$, middle $r = 0.5$, and right: $r = 0.9$. The horizontal dashed lines represent the true value Q in $\tau = 0.3$, 0.5, 0.8 from the bottom to the top.

n	τ	$r = 0.25$	$r = 0.5$	$r = 0.9$
10000	0.3	0.894(0.140)	0.972(0.068)	0.987(0.037)
	0.5	0.807(0.045)	0.876(0.024)	0.906(0.014)
	0.8	0.853(0.399)	0.989(0.207)	1.000(0.112)
20000	0.3	0.928(0.076)	0.966(0.037)	0.982(0.020)
	0.5	0.872(0.034)	0.908(0.018)	0.927(0.010)
	0.8	0.950(0.219)	0.991(0.105)	0.998(0.055)
40000	0.3	0.944(0.044)	0.964(0.022)	0.974(0.012)
	0.5	0.900(0.025)	0.926(0.012)	0.939(0.007)
	0.8	0.965(0.114)	0.984(0.053)	0.993(0.028)
100000	0.3	0.944(0.025)	0.956(0.012)	0.963(0.007)
	0.5	0.927(0.016)	0.935(0.008)	0.945(0.004)
	0.8	0.956(0.054)	0.970(0.026)	0.980(0.013)
200000	0.3	0.948(0.017)	0.954(0.008)	0.958(0.004)
	0.5	0.936(0.011)	0.944(0.006)	0.945(0.003)
	0.8	0.952(0.034)	0.966(0.017)	0.971(0.008)
400000	0.3	0.942(0.012)	0.954(0.006)	0.952(0.003)
	0.5	0.944(0.008)	0.949(0.004)	0.946(0.002)
	0.8	0.948(0.023)	0.960(0.011)	0.961(0.005)

Table 3.2: Empirical results of coverage rate(mean absolute error) of proposed confidence interval (3.3) (estimator Q_n) with data collected from Cauchy.

n	τ	$r = 0.25$	$r = 0.5$	$r = 0.9$
10000	0.3	0.900(0.021)	0.927(0.011)	0.938(0.006)
	0.5	0.951(0.022)	0.970(0.011)	0.971(0.006)
	0.8	0.990(0.029)	0.997(0.014)	0.998(0.008)
20000	0.3	0.920(0.015)	0.932(0.007)	0.941(0.004)
	0.5	0.950(0.014)	0.957(0.007)	0.962(0.004)
	0.8	0.983(0.016)	0.990(0.008)	0.992(0.004)
40000	0.3	0.927(0.011)	0.937(0.005)	0.936(0.003)
	0.5	0.947(0.009)	0.951(0.004)	0.955(0.002)
	0.8	0.974(0.009)	0.978(0.005)	0.982(0.002)
100000	0.3	0.934(0.007)	0.936(0.003)	0.942(0.002)
	0.5	0.948(0.005)	0.948(0.003)	0.956(0.001)
	0.8	0.967(0.005)	0.969(0.003)	0.972(0.001)
200000	0.3	0.936(0.005)	0.935(0.002)	0.939(0.001)
	0.5	0.943(0.004)	0.952(0.002)	0.949(0.001)
	0.8	0.960(0.004)	0.963(0.002)	0.964(0.001)
400000	0.3	0.936(0.003)	0.935(0.002)	0.936(0.001)
	0.5	0.946(0.003)	0.946(0.001)	0.946(0.001)
	0.8	0.955(0.003)	0.956(0.001)	0.956(0.001)

Table 3.3: Empirical results of coverage rate(mean absolute error) of proposed confidence interval (3.3) (estimator Q_n) with data collected from PERT.

n	τ	$r = 0.25$	$r = 0.5$	$r = 0.9$
10000	0.3	0.922(0.043)	0.956(0.021)	0.972(0.011)
	0.5	0.853(0.030)	0.898(0.016)	0.928(0.009)
	0.8	0.965(0.057)	0.984(0.028)	0.994(0.015)
20000	0.3	0.930(0.027)	0.950(0.013)	0.963(0.007)
	0.5	0.896(0.022)	0.928(0.011)	0.934(0.006)
	0.8	0.960(0.032)	0.977(0.016)	0.984(0.008)
40000	0.3	0.939(0.017)	0.953(0.009)	0.959(0.004)
	0.5	0.921(0.016)	0.934(0.008)	0.943(0.004)
	0.8	0.959(0.019)	0.969(0.009)	0.974(0.005)
100000	0.3	0.942(0.010)	0.953(0.005)	0.955(0.003)
	0.5	0.939(0.010)	0.942(0.005)	0.943(0.003)
	0.8	0.954(0.011)	0.959(0.005)	0.960(0.003)
200000	0.3	0.944(0.007)	0.950(0.003)	0.950(0.002)
	0.5	0.938(0.007)	0.947(0.004)	0.946(0.002)
	0.8	0.950(0.007)	0.956(0.004)	0.957(0.002)
400000	0.3	0.945(0.005)	0.947(0.002)	0.950(0.001)
	0.5	0.944(0.005)	0.951(0.002)	0.950(0.001)
	0.8	0.946(0.005)	0.955(0.002)	0.948(0.001)

Table 3.4: Empirical results of coverage rate(mean absolute error) of proposed confidence interval (3.3) (estimator Q_n) with data collected from Uniform.

r	ϵ	r	ϵ
0	0	0.5	1.10
0.05	0.10	0.55	1.24
0.1	0.20	0.6	1.39
0.15	0.30	0.65	1.55
0.2	0.40	0.7	1.73
0.25	0.51	0.75	1.95
0.3	0.62	0.8	2.20
0.35	0.73	0.85	2.51
0.4	0.85	0.9	2.94
0.45	0.97	0.95	3.66

Table 3.5: Conversion table between r and ϵ

3.7.2 Alternative self-normalizes

The following self-normalizer can also be used to construct the asymptotically pivotal quantity,

$$N_n' = \sup_{t\in[0,1]} \left|S_{[nt]} - [nt]Q_n\right|,$$
$$N_n'' = \int_0^1 \left|S_{[nt]} - [nt]Q_n\right| dt,$$

and based on the continuous mapping theorem again, one has that,

$$\frac{n^{-1/2}(S_n - nQ)}{n^{-1/2}N_n'} \xrightarrow{d} \frac{W(1)}{\sup_{t\in[0,1]} |W(t) - tW(1)|},$$
$$\frac{n^{-1/2}(S_n - nQ)}{n^{-1/2}N_n''} \xrightarrow{d} \frac{W(1)}{\int_0^1 |W(t) - tW(1)|\, dt}.$$

3.7.3 Proof of Theorem 3.3

Exhaustive computation yields that for any $(a,b) \in \{0,1\}^2$

$$\frac{\mathbb{P}(LRC(q,r,x) = a|\mathbf{1}_{x>q} = b)}{\mathbb{P}(LRC(q,r,x) = a|\mathbf{1}_{x>q} = 1-b)} \in \{\frac{1+r}{1-r}, \frac{1-r}{1+r}\} \tag{3.6}$$

3.7.4 Proof of Theorem 3.4, 3.5 and 3.6

One can verify that the recursive equation (3.1) is asymptotically equivalent to

$$q_{n+1} = q_n + \frac{1}{d_n}\left(1 - \frac{2}{1-r+2r(1-\tau)} 1_{x_n^* > q_n}\right),$$

where $\mathbb{P}(x_n^* = x_n) = r$, $\mathbb{P}(x_n^* = -\infty) = \mathbb{P}(x_n^* = \infty) = (1-r)/2$. Let

$$H(z,X) = 1 - \frac{2}{1-r+2r(1-\tau)} 1_{X>z}$$
$$h(z,X) = \mathbb{E}H(z,X) = 1 - \frac{2(1-F(z))}{1-r+2r(1-\tau)}.$$

Hence $F(Q) = \tau$ is equivalent to $h(Q,X^*) = 0$. Then, one will find that the estimation of Q with sample $x_1, \ldots, x_n$ under LDP is equivalent to the estimation of Q^* with sample $x_1, \ldots, x_n$ without LDP constraints. The standard framework of the SGD method, such as Theorem 2 and 3 in [70], can be applied. Moreover, the statements in Theorems 3.4, 3.5, and 3.6 hold true.

3.7.5 Minimal variance under GDP

We prove this by contradiction. Assuming that for any $n_0 > 1$ there is a $n > n_0$ such that :

$$\epsilon_n/\sigma_n > k > 1.$$

Let

$$w = \Phi\left(-\frac{1}{2}\right) - \Phi\left(\frac{1}{2} - k\right) > 0.$$

We choose a sufficiently large n_0 such that for any $n > n_0$

$$\mathbb{P}(\hat{\theta}_n/\sigma_n < 1/2|H_0) \geq \Phi(1/2) - w/3$$

and

$$\mathbb{P}(\hat{\theta}_n/\sigma_n < 1/2|H_1) \leq \Phi(1/2 - \epsilon_n/\sigma_n) + w/3 \leq \Phi(1/2 - k) + w/3.$$

Then,

$$\begin{aligned}
\mathbb{P}(\hat{\theta}_n/\sigma_n < 1/2|H_0) - \mathbb{P}(\hat{\theta}_n/\sigma_n < 1/2|H_1) &\geq \Phi(1/2) - \Phi(1/2 - k) - 2w/3 \\
&= 2\Phi\left(\frac{1}{2}\right) - 1 + \Phi\left(-\frac{1}{2}\right) - \Phi\left(\frac{1}{2} - k\right) - 2w/3 \\
&= 2\Phi\left(\frac{1}{2}\right) - 1 + w/3 \\
&> 2\Phi\left(\frac{1}{2}\right) - 1 + w/6
\end{aligned}$$

Then $\hat{\theta}_n$ is not $(0, 2\Phi\left(\frac{1}{2}\right) - 1 + w/6)$-DP and therefore is not asymptotically 1-GDP leading to a contradiction.

References

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

[8] Ú. Erlingsson, V. Pihur, and A. Korolova, “Rappor: Randomized aggregatable privacy-preserving ordinal response,” in *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, 2014, pp. 1054–1067.

[12] J. Dong, A. Roth, and W. J. Su, “Gaussian differential privacy,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2021.

[34] P. Kairouz, S. Oh, and P. Viswanath, “The composition theorem for differential privacy,” in *International conference on machine learning*, PMLR, 2015, pp. 1376–1385.

[37] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith, “What can we learn privately?” *SIAM Journal on Computing*, vol. 40, no. 3, pp. 793–826, 2011.

[38] A. Evfimievski, J. Gehrke, and R. Srikant, “Limiting privacy breaches in privacy preserving data mining,” in *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, 2003, pp. 211–222.

[39] S. L. Warner, “Randomized response: A survey technique for eliminating evasive answer bias,” *Journal of the American Statistical Association*, vol. 60, no. 309, pp. 63–69, 1965.

[40] K. Amin, M. Joseph, and J. Mao, “Pan-private uniformity testing,” in *Proceedings of Thirty Third Conference on Learning Theory*, J. Abernethy and S. Agarwal, Eds., ser. Proceedings of Machine Learning Research, vol. 125, PMLR, 2020, pp. 183–218.

[41] B. Ding, J. Kulkarni, and S. Yekhanin, “Collecting telemetry data privately,” *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[42] C. Dwork and J. Lei, “Differential privacy and robust statistics,” in *Proceedings of the forty-first annual ACM symposium on Theory of computing*, 2009, pp. 371–380.

[43] J. Lei, “Differentially private m-estimators,” *Advances in Neural Information Processing Systems*, vol. 24, 2011.

[44] A. Smith, “Privacy-preserving statistical estimation with optimal convergence rates,” in *Proceedings of the forty-third annual ACM symposium on Theory of computing*, 2011, pp. 813–822.

[45] J. Gillenwater, M. Joseph, and A. Kulesza, “Differentially private quantiles,” in *Proceedings of the 38th International Conference on Machine Learning*, M. Meila and T. Zhang, Eds., ser. Proceedings of Machine Learning Research, vol. 139, PMLR, 2021, pp. 3713–3722.

[46] D. Alabi, O. Ben-Eliezer, and A. Chaturvedi, “Bounded space differentially private quantiles,” *arXiv preprint arXiv:2201.03380*, 2022.

[47] O. Ben-Eliezer, D. Mikulincer, and I. Zadik, “Archimedes meets privacy: On privately estimating quantiles in high dimensions under minimal assumptions,” *arXiv preprint arXiv:2208.07438*, 2022.

[48] X. Shao, “A self-normalized approach to confidence interval construction in time series,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 72, no. 3, pp. 343–366, 2010.

[49] P. Jain, P. Kothari, and A. Thakurta, “Differentially private online learning,” in *Conference on Learning Theory*, JMLR Workshop and Conference Proceedings, 2012, pp. 24–1.

[50] N. Agarwal and K. Singh, “The price of differential privacy for online learning,” in *Proceedings of the 34th International Conference on Machine Learning*, D. Precup and Y. W. Teh, Eds., ser. Proceedings of Machine Learning Research, vol. 70, PMLR, 2017, pp. 32–40.

[51] K. Wei *et al.*, “Federated learning with differential privacy: Algorithms and performance analysis,” *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3454–3469, 2020. DOI: 10.1109/TIFS.2020.2988575.

[52] S. Lee, Y. Liao, M. H. Seo, and Y. Shin, “Fast and robust online inference with stochastic gradient descent via random scaling,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, 2022, pp. 7381–7389.

[53] T. T. Cai, Y. Wang, and L. Zhang, “The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy,” *The Annals of Statistics*, vol. 49, no. 5, pp. 2825–2850, 2021.

[54] M. Joseph, J. Mao, S. Neel, and A. Roth, “The role of interactivity in local differential privacy,” in *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, 2019, pp. 94–105. DOI: 10.1109/FOCS.2019.00015.

[55] A. Cheu, A. Smith, J. Ullman, D. Zeber, and M. Zhilyaev, “Distributed differential privacy via shuffling,” in *Advances in Cryptology–EUROCRYPT 2019: 38th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Darmstadt, Germany, May 19–23, 2019, Proceedings, Part I 38*, Springer, 2019, pp. 375–403.

[56] A. G. Joseph and S. Bhatnagar, “A stochastic approximation algorithm for quantile estimation,” in *International Conference on Neural Information Processing*, Springer, 2015, pp. 311–319.

[57] B. Coppens, I. Verbauwhede, K. De Bosschere, and B. De Sutter, "Practical mitigations for timing-based side-channel attacks on modern x86 processors," in *2009 30th IEEE Symposium on Security and Privacy*, 2009, pp. 45–60. DOI: 10.1109/SP.2009.19.

[58] N. Lawson, "Side-channel attacks on cryptographic software," *IEEE Security & Privacy*, vol. 7, no. 6, pp. 65–68, 2009. DOI: 10.1109/MSP.2009.165.

[59] B. Balle, G. Barthe, and M. Gaboardi, "Privacy amplification by subsampling: Tight analyses via couplings and divergences," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[60] Y. Liu, K. Sun, L. Kong, and B. Jiang, "Identification, amplification and measurement: A bridge to gaussian differential privacy," *Advances in Neural Information Processing Systems*, 2022.

[61] T. C. Brown, P. A. Champ, R. C. Bishop, and D. W. McCollum, "Which response format reveals the truth about donations to a public good?" *Land Economics*, pp. 152–166, 1996.

[62] I. Mironov, "On significance of the least significant bits for differential privacy," in *Proceedings of the 2012 ACM conference on Computer and communications security*, 2012, pp. 650–661.

[63] J. Jin, E. McMurtry, B. I. Rubinstein, and O. Ohrimenko, "Are we there yet? timing and floating-point attacks on differential privacy systems," *arXiv preprint arXiv:2112.05307*, 2021.

[64] S. Haney, D. Desfontaines, L. Hartman, R. Shrestha, and M. Hay, "Precision-based attacks and interval refining: How to break, then fix, differential privacy on finite computers," *arXiv preprint arXiv:2207.13793*, 2022.

[65] Y. Fang, J. Xu, and L. Yang, "Online bootstrap confidence intervals for the stochastic gradient descent estimator," *The Journal of Machine Learning Research*, vol. 19, no. 1, pp. 3053–3073, 2018.

[66] X. Shao, "Self-normalization for time series: A review of recent developments," *Journal of the American Statistical Association*, vol. 110, no. 512, pp. 1797–1817, 2015.

[67] J. Dong, W. Su, and L. Zhang, "A central limit theorem for differentially private query answering," in *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds., vol. 34, Curran Associates, Inc., 2021, pp. 14 759–14 770.

[68] C. E. Clark, "The pert model for the distribution of an activity time," *Operations Research*, vol. 10, no. 3, pp. 405–406, 1962.

[69] P. Langley, "Crafting papers on machine learning," in *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, P. Langley, Ed., Stanford, CA: Morgan Kaufmann, 2000, pp. 1207–1216.

[70] J. Dippon, “Globally convergent stochastic optimization with optimal asymptotic distribution,” *Journal of applied probability*, vol. 35, no. 2, pp. 395–406, 1998.

Chapter 4

Paper 3: Efficient CDF Estimation under Local Differential Privacy: A Constrained Isotonic Approach

4.1 Abstract

We introduce a novel algorithm for estimating Cumulative Distribution Function (CDF) curves under Local Differential Privacy (LDP) by utilizing a combination of constrained isotonic estimation and binary inquiries. We uncover an unexpected connection between LDP and the current status problem, a classical survival data problem in statistics. Through mathematical proofs and extensive numerical testing, we demonstrate that our method achieves a uniform error bound of $\mathcal{O}_p(n^{-1/3}\log n)$ and an L_2 error bound of $\mathcal{O}_p(n^{-1/3})$ when estimating the entire CDF curves. By concentrating on a finite grid, the error bound can be improved to $\mathcal{O}_p(n^{-1/2})$, with an asymptotic normal distribution of error. Theoretically, we have shown that the error bound smoothly changes from $\mathcal{O}_p(n^{-1/2})$ to $\mathcal{O}_p(n^{-1/3})$ as the number of grids increases relative to the sample size n. Computationally, we demonstrate that our constrained isotonic estimator can be efficiently computed in a deterministic manner, without the need for any hyperparameters or random optimization.

4.2 Introduction

The cumulative distribution function (CDF) contains complete information about a random quantity, and claims about probability distributions are often made without sufficient evidence. Estimating CDFs has a long history, and for data from known parameter family distributions, such as the normal or exponential distributions, one can estimate the parameters using the method of moments or maximum likelihood estimation (MLE), and then obtain an estimation of the CDF. The empirical cumulative distribution function (ECDF) is the most commonly used CDF estimator and has good theoretical properties, including uniform consistency, weak convergence, see [71], and the invariance principle, see [72], leading to a series of statistical inference problems, such as constructing simultaneous confidence bands, goodness-of-fit tests and change point detection.

Advances in computational power and statistical methodologies have led to more sophisticated techniques for CDF estimation. Examples include kernel smoothing, which was proposed by [73] to overcome the discontinuity of ECDF, and constrained polynomial spline regression, proposed by [74] to ensure smoothness and monotone non-decreasing functions while reducing the computational burden. As data types become more complex, such as multivariate, time series, and spatial data, various CDF estimation methods have emerged, as seen in [75], [76], and [77].

On the other hand, as the power and utility of statistical methodologies continue to expand, concerns regarding the privacy of individuals behind the data become increasingly relevant, especially in a world where online services are omnipresent and sensor-rich devices such as smartphones consistently collect information. The advancement of data science has demonstrated that seemingly harmless data can often be exploited. Classic examples of such exploitation include the Netflix challenge [3, 4] and the AOL search log database [5], where users are re-identified from anonymized datasets. More recently, studies on mobile device sensor data have shown that driving patterns, location [78], identity, and even spoken words [79, 80] can be reconstructed from accelerometer data. Surprisingly, even a user's Internet activity can be detected from a wireless charger [81].

This need to protect users from current and future attacks has led to the development of privacy-preserving statistical techniques. Differential privacy (DP) [6] has garnered significant attention due to its capability to provide robust privacy guarantees while still enabling meaningful data analysis. However, despite the solid mathematical foundation of DP in safeguarding user information, concerns remain regarding the potential for data collectors to violate privacy guarantees. In the first few months of 2023, there have already been several notable data breach events, either intentional or accidental. In March 2023, an error in ChatGPT allowed users to view another active user's personal information before the service was taken offline. Earlier that month, a breach involving a Washington DC-based healthcare provider that handles sensitive data belonging to federal legislators and their families highlighted a long-existing issue in the data safety of the US healthcare industry [82]. For more information about faulty data curators, refer to [83, 84].

4.2.1 Related work

Incorporating LDP into CDF estimation poses a unique set of challenges, as traditional estimation techniques such as ECDF, kernel smoothing, or constrained polynomial spline regression may not inherently provide the required privacy guarantees. There have been several attempts to address these issues, which can be traced back to the development of the Frequency Oracle (FO) [8, 85, 86]. These mechanisms were primarily designed for discrete domains. By applying discretization to continuous domains, these methods can be used for estimating continuous distributions, but at the cost of losing some information about the continuous structure. Later, Li et al. [87] improved the FO algorithm by computing an MLE using the Expectation-Maximization (EM) algorithm through a square wave mechanism. However, such approach still relies on discretization (bucketing), and the quality of the output is highly sensitive to the stopping criteria hyperparameter. The finite number of bucketing not only prevents the estimator from being asymptotically consistent but also introduces additional challenges due to the extra parameter.

4.2.2 Outline

We begin by presenting a brief review of the relevant definitions and background related to CDF estimation and LDP followed by our data collection procedure, which employs a series of random-response randomizers to transform sensitive individual information into a private view of binary variables. We highlight that this data collection process results in a private view that resembles the structure of the current status problem, a well-studied issue in survival analysis. Subsequently, we construct an estimator based on the private view obtained in the previous step. Interestingly, the LDP treatment and statistical analysis technique can be disentangled by considering an alternative view of the collected data, where the randomized response can be treated as a truthful response originating from an alternative variable. We then refine the naive MLE method by imposing monotonic and bound constraints on the estimator and demonstrate that such an estimator can be computed in a fast, deterministic, and hyperparameter-free manner. Following this, we investigate the asymptotic properties of the proposed estimator, providing a comprehensive analysis of its performance under various conditions. Under different sampling strategies, introduced below, we first establish L_2 and uniform consistency up to order $\mathcal{O}_p(n^{-1/3})$ and $\mathcal{O}_p(n^{-1/3}\log n)$ respectively. Then, we derive the point-wise weak convergence results of the proposed estimator. In addition, the theoretical justifications show that the convergence rate varies continuously between these two samplings from $\mathcal{O}_p(n^{-1/3})$ to $\mathcal{O}_p(n^{-1/2})$. Especially, for the estimation on finite grids, we obtain the asymptotic normality for whole design points with a diagonal asymptotic covariance matrix, which can be applied for constructing confidence intervals and hypothesis testing.

Lastly, we demonstrate the effectiveness of our proposed protocol through numerical experiments, showcasing its practical utility and accuracy in CDF estimation under the LDP framework.

4.2.3 Notations

In this paper, we employ the following notations. $\mathbf{1}_{\{.\}}$ is the indicator function and $[a]$ denotes the largest integer that does not exceed a. $\mathcal{O}_p$ (or o_p) denotes a sequence of random variables of a certain order in probability. For instance, $o_p(n^{-1/2})$ means a smaller order than $n^{-1/2}$. For sequences a_n and b_n, denote $a_n \asymp b_n$ if there exist positive contants c and C such that $cb_n \leq a_n \leq Cb_n$. The symbol $\xrightarrow{d}$ means weak convergence or converge in distribution.

4.3 Preliminaries

4.3.1 Central Differential Privacy

Definition 4.1 *[6] A randomized algorithm $\mathcal{A}$, taking a dataset consisting of individuals as its input, is (ϵ, δ)-differentially private if, for any pair of datasets S and S' that differ in the record of a single individual and any event E, satisfies the below condition:*

$$\mathbb{P}[\mathcal{A}(S) \in E] \leq e^{\epsilon}\mathbb{P}\left[\mathcal{A}\left(S'\right) \in E\right] + \delta.$$

When $\delta = 0$, $\mathcal{A}$ is called ϵ-differentially private (ϵ-DP).

The concept of DP only imposes constraints on the output distribution of an algorithm $\mathcal{A}$, rather than placing restrictions on the credibility of the entity running the algorithm or protecting the internal states of $\mathcal{A}$. The existence of the curator who has access to the raw data set is why this approach is known as Central DP (CDP). The curator simplifies the algorithm design and often leads to an asymptotically negligible loss of accuracy from privacy protection [53].

4.3.2 Local Differential Privacy

The idea of LDP roots in the following concept is called (ϵ, δ)-randomizer.

Definition 4.2 *[54] An (ϵ, δ)-randomizer $R : X \to Y$ is an (ϵ, δ)-differentially private function taking a single data point as input.*

The randomizer definition is mathematically a special case of the CDP framework. While CDP focuses on protecting privacy by adding noise to the aggregated output of a query over a dataset, LDP emphasizes adding noise to each individual data point before any computation or aggregation is performed. Therefore, a mechanism is (ϵ, δ)-LDP if and only if it takes outputs of (ϵ, δ)-randomizer as its input [54]. This local approach to privacy ensures that the privacy of each data point is preserved, even if an adversary has access to the noisy data.

4.3.3 The current status problem

Current status data emerges in studies where the primary measurement is the occurrence time of a specific event, but observations are confined to indicators that reveal if the event has transpired at the time of data collection. This type of data is particularly relevant in survival analysis, such as in research investigating the survival of patients with cancer during an observation period, which is highly related to isotonic regression, see [88]. In these cases, researchers may passively acquire the patient's status through hospital visits (alive) or loss of contact (presumably dead). However, the exact time of death is unobtainable, especially if it lies in the future. Refer to [89–93] for more information about the research of this type.

4.4 Problem formulation and solution

Let $X = \{X_1, \ldots, X_n\}$ be independently and identically distributed (i.i.d.) random variables defined on $[0, 1]$ representing private information of each user. The goal is to estimate the underlying CDF of X_i (F) with inquiries to each user while conforming to the ϵ-LDP condition.

4.4.1 The LDP data collection

In contrast to the CDP setting, where user data is openly gathered for analysis, the development of an LDP protocol commences with the data collection process. This is due to the ϵ-LDP constraint, which presents a significant challenge for estimation problems. Initially, without the DP constraint, the CDF can be intuitively approximated by the ECDF, yielding a convergence rate of $\mathcal{O}_p(n^{-1/2})$. Nevertheless, in the LDP context, each data point's contribution is considerably restricted.

To illustrate this point, consider the canonical Laplace mechanism, which serves as the standard DP mechanism for bounded continuous variables. The noise variance (2.0) needed to achieve LDP with $\epsilon = 1$ is eight times larger than the highest possible variance (0.25) of the $[0, 1]$ bounded variable. In addition, reconstructing the original distribution from the Laplace noise perturbed data will lead to a notoriously hard deconvolution problem [94] with terrible sample efficiency [95]. This stringent condition compels us to constrain the inquiries directed toward end users, thereby reducing the scale of DP noise. To this end, we generate T_i from another distribution G and collect binary responses from users of the question below:

Is the T_i you see here greater or equal to than the private number you have (X_n)?

If the users are asked to answer this question truthfully, the users still faces a serious privacy concern. However, we can ask the user to perturb the answer locally leading to the following randomizer:

Definition 4.3 *Random response randomizer:*

$$\mathcal{E}_i(X_i) = \begin{cases} \mathbf{1}_{X_i \leq T_i}, & w.p. \quad r, \\ Bernoulli(0.5) & w.p. \quad 1\text{-}r. \end{cases} \tag{4.1}$$

As a special case of randomized response, $\mathcal{E}_i$ is a $(\epsilon, 0)$ randomizer for $\epsilon = \log((1 + r)/(1 - r))$ [96]. For certain values of r, the mechanism can be executed physically using a coin or dice. It is worth noting that in the definition provided, the value of T_i is generated by the curator and distributed to users. This is due to concerns that end-users may lack the knowledge or equipment to accurately produce the necessary randomness. However, if the entire process is automated on digital devices, the generation of T_i can be shifted to the user side. This approach reduces communication costs and offers additional privacy advantages since the datapoint be no longer trace by the assigned T_i. For the remainder of this paper, we will represent the privacy budget using $r = \tanh(\epsilon/2)$, as it affords a more intuitive interpretation and a simpler form in our results (refer to Table 4.2 for a conversion table).

Following the definitions above, the curator can collect a private data view of X namely $(\Delta_1, T_1), \ldots, (\Delta_i, T_i)$, where $\Delta_i = \mathcal{E}_i(X_i)$. By the post-processing property, any function of the private view or even the view itself can be safely released without any concern of privacy violation.

Compared to the current status problem, where control over T_i is limited, the T_i employed in our LDP mechanism can be fully tailored. Initially, by either sending i.i.d. T_i to users or requesting users to generate T independently, it is clear that T_i is independent of all X_i and all other T_j for $j \neq i$. This scenario is atypical in medical studies concerning current status. In practice, patients' visits often correlate with their own status and even the status of others (for instance, the weather may affect hospital visits, leading to correlated T_i. In another example, when observing patients' lifespans, since future deaths cannot be observed, the censoring is related to the current time and patients' birth data).

The ability to freely design G provides a significant advantage. Firstly, the independence between T_i and X_i simplifies the analysis of estimation. The known G also eliminates the need for estimation and the potential errors that may arise from it. However, the most crucial aspect we can design is the control of G to generate better estimations in areas of interest. Intuitively, the estimation of F will be more accurate when G samples more. Here, we introduce two types of sampling methods:

Density-based sampling: In this type of sampling, we let G correspond to a density g that is uniformly bounded away from 0, ensuring that every open set within the domain can be sampled with a non-zero probability. This approach provides a more comprehensive representation of the underlying distribution. The simplest G in this form is the CDF of the uniform distribution, given by $G(x) = x$. Such a choice will lead to uniform sampling. Other reasonable choices include weighted sampling, where G is selected to be denser in regions of particular interest, or if we possess prior knowledge about F, we can choose $G \approx F$ to achieve improved estimation outcomes.

Preselected sampling: In this sampling method, we let G be a discrete distribution on the interval $[0, 1]$. This means we pre-select a subset of values $\{x_1, \ldots, x_\kappa\} \subset [0, 1]$ and define $G(x) = \sum_{i=1}^{\kappa} p_i 1_{x>x_i}$. Preselected sampling focuses the estimation on the chosen nodes, making it particularly useful when there are specific exact x values of interest. For instance, when estimating income distribution, we might be especially interested in the proportion of people below the poverty line or above a certain income threshold. Alternatively, we might be interested in an evenly distributed grid of X_i to provide a plausible plot of the CDF curve.

4.4.2 The constrained isotonic estimator

The protocol in the last chapter provided an LDP view of the data. In this chapter we construct an estimator from the LDP view. Define:

$$X_i^\star = \begin{cases} X_i, & w.p. \text{ r,} \\ \text{Bernoulli}(0.5) & w.p. \text{ 1-r.} \end{cases} \tag{4.2}$$

The CDF of $X_i^\star$ can be derived from F as below:

$$F^\star(x) = \left(rF(x) + \frac{1-r}{2} \right) \mathbf{1}_{0<x<1} + \mathbf{1}_{x=1}. \tag{4.3}$$

It's easy to see the distribution of (Δ_i, T_i) is identical to $(\Delta_i^\star, T_i)$ where $\Delta_i^\star = \mathbf{1}_{X_i^\star \le T_i}$. Using the new notations of $F^\star$, we can consider the random response of T_i as a truthful response originating from an alternative variable, $x^\star$. This approach mitigates the interplay between LDP treatment and statistical analysis techniques. Employing the revised notations of $F^\star$, the log-likelihood can be expressed as

$$\begin{aligned} L(F, \Delta, T) :&= \sum_{i=1}^{n} \Delta_i^\star \log F^\star(T_i) + (1 - \Delta_i^\star) \log(1 - F^\star(T_i)) && (4.4) \\ &= \sum_{i=1}^{n} \Delta_i \log \left(rF(T_i) + \frac{1-r}{2} \right) + (1 - \Delta_i) \log \left(\frac{1+r}{2} - rF(T_i) \right). && (4.5) \end{aligned}$$

It may seem appealing to determine $F^\star$ and F through the naive maximization of the log-likelihood function. Nonetheless, several issues arise from this approach. For instance, the maximizing function $F^\star$ may not necessarily represent a CDF as there is no assurance that the estimation will exhibit monotonic behavior. Additionally, the error associated with $F^\star$ is likely to be significantly high around the values of 0 and 1 due to the lack of relevant samples. To address this issue, we define D as the function family of all non-decreasing functions mapping from $[0, 1]$ to $[0, 1]$, and propose our constrained isotonic estimator as follows:

$$\hat{F} \in \arg\max_{\hat{F} \in D} L(F, \Delta, T).$$

We remark that the $\hat{F}$ satisfying the definition is not unique. The right-hand side represents an equivalence class, wherein two distribution functions are considered equivalent if and only if they agree on all T_i. Consequently, the maximization process can be performed solely on the nodes T_i. The remaining part of the function can be arbitrarily monotonically interpolated. For instance, the function values can be determined by the nearest T_i to the left or right, or they can be linearly interpolated. Regardless of the interpolation technique, the properties presented in the following chapter remain applicable. In the numerical experiments, the function values are filled using the nearest T_i to the left, resulting in a left-continuous staircase function to avoid unfair advantages.

4.4.3 Algorithm

The disentanglement between the LDP treatment and the data analysis allows our to make use of the nonparametric likelihood estimation algorithm in the survival analysis

[97] with minor tweaking. A detailed description of the full algorithm is presented in Algorithm 7. The initial four steps adhere to the standard procedure of isotonic regression as delineated by Hill [98]. In the intermediate stage, the estimation $\widehat{F}^{\star}(x)$ optimizes the log-likelihood function, disregarding the constraint that $\hat{F}(x) \in [0, 1]$ (or equivalently $\widehat{F}^{\star}(x) \in [(1-r)/2, (1+r)/2]$). Notably, although the range constraint is not taken into account during the evaluation of $\widehat{F}^{\star}(x)$, it is effectively satisfied in step 6 through a clipping procedure. This process not only ensures compliance with the range constraint but also maintains optimality under this constraint (refer to Appendix 4.8.3 for a detailed proof). The GCM can be deterministically computed, eliminating the requirement for iterative optimization, as demonstrated by Robertson et al. [99] and referenced in [100]. This results in an overall deterministic algorithm free of hyperparameters for the analysis. In addition, the Algorithm 7 demonstrates excellent performance. For $n \leq 10^7$, it took less than 1s to execute on a single core of an AMD Threadripper PRO 3995WX CPU. For comprehensive details regarding computation times, refer to Table 4.6.

Algorithm 7: Constrained isotonic estimation

```
1: Compute the function $H_1(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{T_i \leq x\}}$, $H_2(x) = \frac{1}{n}\sum_{i=1}^{n} \Delta_i \mathbf{1}_{\{T_i \leq x\}}$.
2: Plot $M = (H_1(x), H_2(x))$, for $x \in [0, 1]$.
3: Compute Greatest Convex Minorant (GCM) of $M$ as $Z$.
4: Compute $\widehat{F}^{\star}(x) =$ left-derivative of $Z$ at $H_1(x)$ for $x \in [0, 1]$.
5: Invert the linear relationship by $\widetilde{F} = r^{-1}\left(\widehat{F}^{\star}(x) - \frac{1-r}{2}\right)$.
6: Give $\hat{F}(x) = 0 \bigvee (1 \bigwedge \widetilde{F}(x))$.
```

4.5 Asymtotic properties

In the following section, we will explore the mathematical underpinnings of the data processing algorithm proposed in the previous section. To facilitate our analysis, we will start by introducing consistency results under density-based sampling:

Theorem 4.4 *Consistency under density-based sampling:*

(i) L_2 consistency: If there exists $g(x) > 0$ that are the corresponding density function of $G(x)$, one has that

$$\left\| \widehat{F} - F \right\|_2 = \mathcal{O}_p(r^{-1}n^{-1/3}),$$

where $\|h\|_2^2 = \int_0^1 h^2(x)dx$.

(ii) Uniform consistency: With further assumptions that there exists $f(x) > 0$ that are the corresponding density function of $F(x)$, one has that

$$\sup_{x\in[0,1]} \left| \widehat{F}(x) - F(x) \right| = \mathcal{O}_p(r^{-1}n^{-1/3}\log n).$$

From Theorem 4.4, one finds that L_2 consistency requires the existence of the density function of $G(x)$ only, and the convergence rate will be slightly faster than uniform consistency. The asymptotic results are affected by the truthful response rate r, with smaller r requiring a larger sample size to obtain the same finite sample performance. Next, we will discuss the pointwise asymptotic distribution of the proposed estimator.

Theorem 4.5 *Under the assumption in Theorem 4.4, for any $x_0 \in (0,1)$,one obtains that*

$$\frac{4g(x_0)n^{1/3}(\widehat{F}(x_0) - F(x_0))}{\left(rF(x_0) + \frac{1-r}{2}\right)\left(\frac{1+r}{2} - rF(x_0)\right) f(x_0)} \Rightarrow \mathcal{Z} := \arg\max_{t\in\mathbb{R}} \left\{W(t) - t^2\right\}, \tag{4.6}$$

where $W(t)$ is standard two-sided Brownian motion, and $\mathcal{Z}$ is referred as Chernoff distribution.

The proposed CDF estimator exhibits significantly different asymptotic behaviour compared to non-DP cases. While the convergence rate of ECDF is up to order $\mathcal{O}_p(n^{-1/2})$, the convergence rate of the proposed estimator is much slower. Furthermore, we note that the sequence of stochastic processes $\{n^{1/3}(\widehat{F}(x) - F(x)), x \in [0,1]\}$ is not tight in $D[0,1]$ leading to a problematic topological structure caused by infinite dimensionality (see [97]). Such underlying difficulty limits us to point-wise asymptotic distribution, as opposed to the weak convergence results of $\widehat{F}$ and related goodness-of-fit statistics, such as KS statistics. However, in most practical scenarios, we are interested in estimating pre-design points on the CDF, rather than the entire curve, due to computational accuracy constraints.

To achieve this, we assume that the observation times T_i are i.i.d. random variables sampled from a discrete probability measure G_n supported on $[0,1]$. We denote the support of G_n by $\{x_{i,n}\}_{i=1}^{\kappa_n}$, where the ith grid point is given by $x_{i,n} = in^{-\gamma}$, $i = 1, \ldots \kappa_n = [n^\gamma]$, and $\gamma \in (0,1]$. We view the distribution G_n as a discretization of an absolutely continuous distribution G', with $G_n\{x_{i,n}\} = G'(x_{i,n}) - G'(x_{i-1,n})$ for $i = 2, 3, \ldots, \kappa_n - 1$, $G_n\{x_{1,n}\} = G'(x_{1,n})$, and $G_n\{x_{\kappa_n,n}\} = 1 - G'(x_{\kappa_n-1,n})$, which will allow us to unify the theoretical framework of density-based sampling and preselected sampling mentioned above while obtaining the relationship between their convergence rates. Our focus is on estimating F at a grid point. To this end, we choose a grid point with respect to a fixed time $x_0 \in (0,1)$ that does not depend on n and can be viewed as an anchor point. We define x_l as the largest grid point less than or equal to x_0, and x_r as the first grid point to the right of x_l.

Theorem 4.6 *Consistency and asymptotic distribution under preselected sampling:*

Under the assumptions in Theorem 4.4, and $f(x) \in C[0,1]$, for any $x_0 \in (0,1)$, if $\gamma \in (0, 1/3)$, as $n \to \infty$, one has that

$$n^{1/2-\gamma/2}\left(\widehat{F}(x_l) - F(x_l), \widehat{F}(x_r) - F(x_r)\right) \Rightarrow \sqrt{\frac{\left(rF(x_0) + \frac{1-r}{2}\right)\left(\frac{1+r}{2} - rF(x_0)\right)}{r^2 g'(x_0)}} N\left(0, I_2\right),$$

where $g'(x_0)$ is the density of G' on x_0.

Specially, if $\kappa_n = \kappa < \infty$, then for increasing sequence $\{x_j\}_{j=1}^{\kappa}$, on has that

$$\sqrt{n}\left(\widehat{F}(x_j) - F(x_j)\right)_{j=1,\ldots,\kappa} \Rightarrow N\left(0, \operatorname{diag}\left(\left\{\frac{\left(rF(x_l) + \frac{1-r}{2}\right)\left(\frac{1+r}{2} - rF(x_l)\right)}{r^2(G'(x_j) - G'(x_{j-1}))}\right\}_{j=1,\ldots,\kappa}\right)\right), \tag{4.7}$$

It is noteworthy that the asymptotic results in Theorem 4.6 get rid of all nuisance parameters and can be evaluated directly. Theorem 4.6 establishes the connection between density-based sampling and preselected sampling. The convergence rate of the estimator $\widehat{F}(x)$ is determined by the density of grids (relative to sample size n), and the first statement in the theorem describes how the convergence rate varies continuously from $\mathcal{O}_p(n^{-1/3})$ to $\mathcal{O}_p(n^{-1/2})$ when γ varies from 1/3 to 0. When $\gamma \geq 1/3$, the convergence rate is still no slower than $\mathcal{O}_p(n^{-1/3})$, but the asymptotic distribution will be more complex, which is neither normal distribution in Theorem 4.6 or Chernoff distribution defined in Theorem 4.5. The second statement mainly focuses on estimating $F(x)$ on finite grids. The convergence rate approaches the order of $\mathcal{O}_p(n^{-1/2})$, as same as the convergence rate of ECDF in non-DP cases, and the asymptotic normality holds for the whole sequence $\{x_j\}_{j=1}^{\kappa}$. In the finite grids case, the observation grids are not necessarily uniform like infinite ones, which will be more fixable in practice. Also, the number of grids will not be increasing as the sample size in many scenarios, which will fall into the discussion of the second assertion. The covariance matrix in (4.7) is a diagonal matrix, which implies that the estimators $\{\widehat{F}(x_j)\}_{j=1}^{\kappa}$ are asymptotically independent. This may seem counter-intuitive, but [101] studied the local dependence structure of this type of process in a closely related problem, and one can construct i.i.d. random variables with the distribution of the estimators $\{\widehat{F}(x_j)\}_{j=1}^{\kappa}$ (see the Appendix), which simplifies the results. Therefore, we can conduct statistical inference on $\{F(x_j)\}_{j=1}^{\kappa}$ based on (4.7), such as constructing confidence intervals and hypothesis testing for $F(x)$ on the grids $\{x_j\}_{j=1}^{\kappa}$.

4.6 Experiments

In this chapter, we assess the performance of our proposed algorithms by employing various probability distributions. The datasets are derived from four distinct cases: Uniform distribution $U(0,1)$, Truncated normal distribution $N_c(0,1,\mu,\sigma^2)$, and Continuous Bernoulli distribution $CB(\lambda)$.

For the Truncated normal distribution, the parameters are set as $\mu = 1/2$ and $\sigma^2 = 1/4$. This results in a distribution equivalent to $x/2 + 1/2$, conditioned on the absolute value of x being less than 1, where x follows a standard normal distribution.

In the case of the Continuous Bernoulli distribution, the parameter λ is selected to be 1/4, which yields a non-symmetric density function. The density functions for the specified distributions are illustrated in Figure 4.2.

We consider the truthful response rate $r = 0.25, 0.5, 0.9$, which means the privacy budget is $\epsilon = \log(1 + 2r/(1 - r))$ corresponding to $0.51, 1.09, 2.94$ respectively. These values indicate varying levels of privacy protection, ranging from strong to moderate. For comparison, Apple's implementation of differential privacy employs privacy budgets of 8 for QuickType and auto-play intent, 4 for emoji usage and crash reports in Safari, and 2 for highly sensitive health data [102, 103].

The sample size ranges n spans from 10^3 to 10^7, with a total of $10,000$ replications (and reported means). To eliminate any correlation between experiments, the results from different sample sizes are independently conducted from scratch. Before delving into a more detailed presentation of the results, we first showcase a plot of our proposed estimator for the uniform distribution under density-based sampling. As depicted in Figure 4.3, our estimator, represented by the staircase functions, converges to the true CDF as n increases, resulting in diminishing absolute errors in the form of spikes.

4.6.1 Density based sampling performance

Next, we discuss the performance of our proposed estimator under density-based sampling. As for the sampling density, we consider two types of G. The first type is $G(x) = x$, which corresponds to uniform sampling. This is the preferred choice in situations where we do not have explicit preferences or knowledge about the distribution and domain. The second type of G is chosen as $G = F$. Although it is unlikely to occur in real practice, this represents the best possible case when we already have some prior knowledge about the distribution (as an extreme case of $G \approx F$). Results under the second type of G are marked with an additional $*$. We do not consider the uniform distribution in the second type of G as it overlaps with the uniform setting. The table below presents the empirical results for both uniform consistency (represented by the maximum absolute error) and L_2 consistency (represented by the L_2 error) of the estimator.

n	r	$U(0,1)$	$N_c(0,1,\mu,\sigma^2)$	$CB(\lambda)$	$N_c(0,1,\mu,\sigma^2)^\star$	$CB(\lambda)^\star$
10^3	0.25	0.262(0.118)	0.289(0.116)	0.270(0.120)	0.261(0.111)	0.265(0.116)
	0.5	0.183(0.076)	0.199(0.074)	0.185(0.075)	0.182(0.073)	0.183(0.076)
	0.9	0.127(0.050)	0.137(0.047)	0.129(0.049)	0.126(0.045)	0.127(0.049)
10^4	0.25	0.143(0.057)	0.156(0.057)	0.147(0.057)	0.142(0.055)	0.143(0.057)
	0.5	0.096(0.036)	0.104(0.035)	0.100(0.036)	0.096(0.035)	0.096(0.036)
	0.9	0.065(0.023)	0.073(0.022)	0.067(0.022)	0.065(0.021)	0.065(0.023)
10^5	0.25	0.074(0.027)	0.081(0.027)	0.077(0.027)	0.074(0.026)	0.074(0.027)
	0.5	0.048(0.017)	0.054(0.017)	0.050(0.017)	0.049(0.017)	0.049(0.017)
	0.9	0.033(0.011)	0.037(0.010)	0.034(0.010)	0.033(0.010)	0.033(0.010)
10^6	0.25	0.038(0.013)	0.041(0.013)	0.039(0.013)	0.038(0.013)	0.038(0.013)
	0.5	0.024(0.008)	0.027(0.008)	0.025(0.008)	0.024(0.008)	0.024(0.008)
	0.9	0.016(0.005)	0.019(0.005)	0.017(0.005)	0.016(0.005)	0.016(0.005)
10^7	0.25	0.019(0.006)	0.021(0.006)	0.020(0.006)	0.019(0.006)	0.019(0.006)
	0.5	0.012(0.004)	0.013(0.004)	0.013(0.004)	0.012(0.004)	0.012(0.004)
	0.9	0.008(0.002)	0.009(0.002)	0.008(0.002)	0.008(0.002)	0.008(0.002)

Table 4.1: Empirical results of uniform consistency (L_2 consistency) of the proposed estimator with G is uniform distribution or $G = F$.

As observed in the table, both the maximum absolute error and the L_2 error decrease as n increases and the privacy budget increase (larger r) as expected. The results for second type of G show a slight advantage over the first one, but the difference is negligible. This observation suggests that a uniform sample would be sufficient, and while sampling closer to the true distribution can be helpful, the improvement is only marginal. Therefore, we recommend using uniform sampling with the density-based approach, as it avoids the issues associated with acquiring prior knowledge. We remark that the maximum errors for the type 2 groups are nearly identical for larger samples;

this is because the effects of f and g cancel each other out in Theorem 4.5. To verify our claimed convergence rate we give a graphical illustration comparison between results from different sample sizes and the convergence rate and we showcase a term what we call standardized maximum absolute error (SMAE), which is defined as MAE multiplied by $rn^{1/3}/\log n$. Under Theorem 4.4, the SMAE should remain bounded as $n \to \infty$ and varying r. We show this in the plot of standardized and unstandardized maximum absolute error (Figure 4.4 in Appendix). The three bundles in the curve of SMAE representing the results from the same F tend to be similar (not r), suggesting that the effect of privacy budget r is also properly modelled the standardization factor $rn^{1/3}/\log n$. This supports our claim in Theorem 4.4.

4.6.2 Preselected sampling performance

Theorem 4.6 predicts a multivariate asymptotically normal distribution for the residual. To condense the results into interpretable numerical outcomes, we define the following standardized weighted L^2 error, which takes the sum of each square error divided by their corresponding predicted variance.

$$WMSE(\widehat{F}) := \sqrt{n} \sum_{j=1}^{\kappa} \frac{r^2 \left(G'(x_j) - G'(x_{j-1})\right) \left(\widehat{F}\left(x_j\right) - F\left(x_j\right)\right)^2}{\left(rF\left(x_l\right) + (1-r)/2\right)\left((1+r)/2 - rF\left(x_l\right)\right)} \tag{4.8}$$

According to Theorem 4.6, the $WMSE(\widehat{F})$ asymptotically follows a χ^2 distribution with a degree of freedom κ. We proceed to compare the empirical $WMSE(\widehat{F})$ with the theoretical $\chi^2(\kappa)$ distribution from two perspectives. First, we examine the relative χ^2 error (RCE), which we define as $WMSE(\widehat{F})/\kappa$. An RCE value greater than 1 indicates that the actual weighted error is larger than expected, and vice versa. Second, we consider the coverage rate, defined as $\mathbb{P}(WMSE(\widehat{F}) < \chi^2_{0.95,\kappa})$. If the distribution of the residuals aligns with expectations, the coverage rate should converge to 0.95. In the following, we present a plot illustrating the relative χ^2 error and coverage rate for the uniform distribution and sampling when $\kappa = 10$:

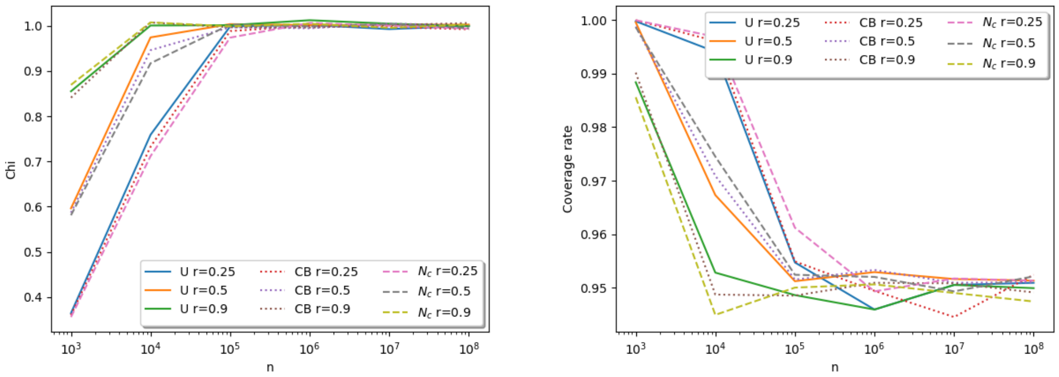


Figure 4.1: Left: The plot of relative χ^2 error compared to the true value Right: the plot of coverage rate

The results for large samples ($n \geq 10^5$) align with our claim in Theorem 4.6. Further, the small sample performance is actually better than our prediction. Consequently, our error bound proves to be numerically valid in this experiment, and for large samples, our asymptotic distribution permits statistical inference. For larger values of κ, the error bounds and asymptotic distribution remain valid, but a greater number of samples will be required to converge to the asymptotic results. We provide the RCE and coverage rates in tables located in the Appendix.

4.7 Conclusions and Future works

In this paper, we developed a data collection procedure and estimator for CDF estimation under the LDP framework, analyzed its asymptotic properties, and demonstrated its practical utility and accuracy through numerical experiments. Our work provides a comprehensive approach to CDF estimation while preserving privacy, offering valuable insights and applications for the field of privacy-preserving data analysis. Although this article makes significant contributions, there remain several intriguing unanswered questions, which pave the way for future research. Firstly, since the proposed estimator $\widehat{F}$ is non-differentiable, obtaining the density estimator directly becomes a challenge. Additionally, extending the estimation of the multivariate CDF to multivariate data poses further difficulties. Lastly, exploring the generation of bootstrap samples based on the proposed estimator for conducting further inference is an intriguing direction worth investigating.

4.8 Appendix

4.8.1 Figures and tables

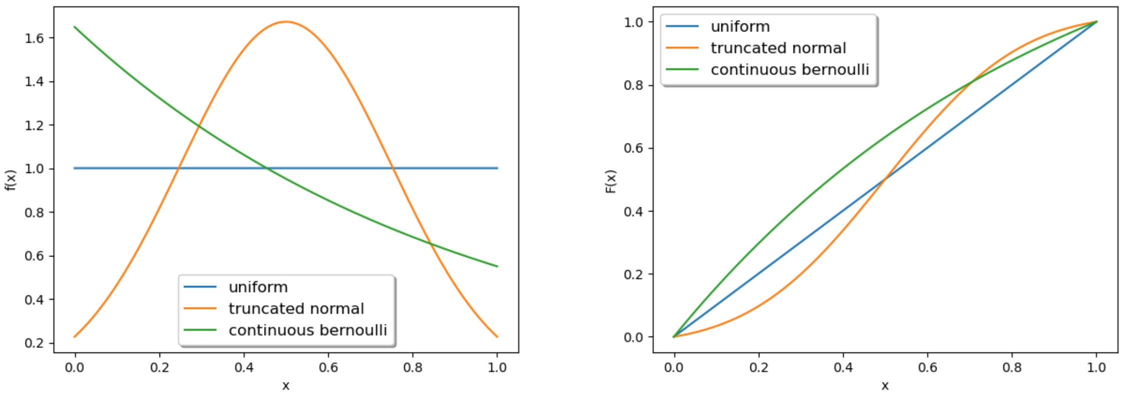


Figure 4.2: Left: Plot of CDF of the distributions Right: Plot of PDF of the distributions

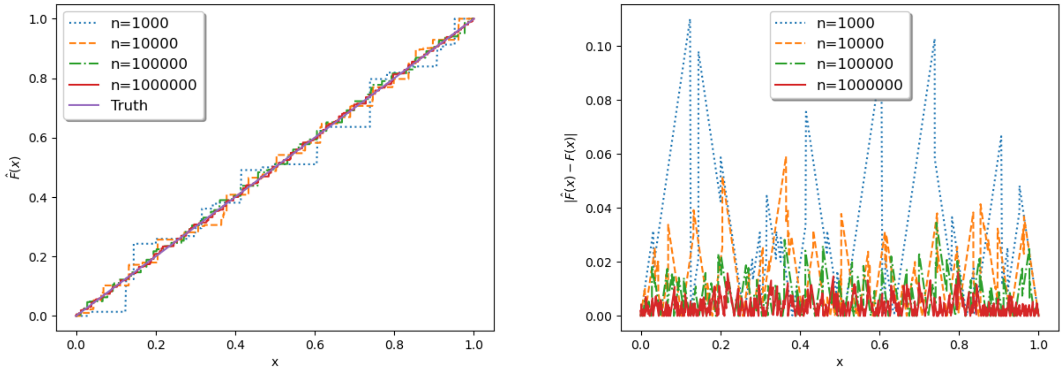


Figure 4.3: Left: The plot of the estimation and true value Right: The plot of absolute error

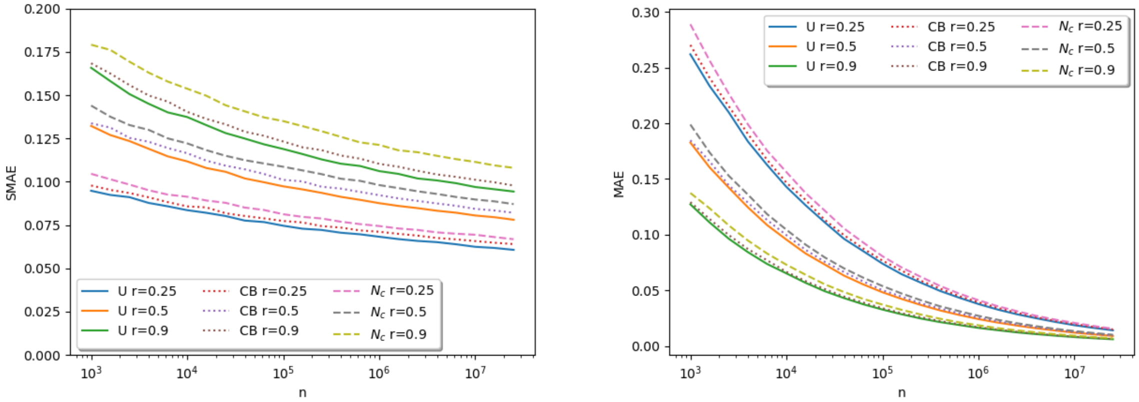


Figure 4.4: Plot of the maximum absolute error: standardized (left) and unstandardized(right)

r	ϵ	r	ϵ
0	0	0.5	1.10
0.05	0.10	0.55	1.24
0.1	0.20	0.6	1.39
0.15	0.30	0.65	1.55
0.2	0.40	0.7	1.73
0.25	0.51	0.75	1.95
0.3	0.62	0.8	2.20
0.35	0.73	0.85	2.51
0.4	0.85	0.9	2.94
0.45	0.97	0.95	3.66

Table 4.2: Conversion table between r and ϵ

n	r	$U(0,1)$	$N_c(0,1,\mu,\sigma^2)$	$CB(\lambda)$
10^3	0.25	1.000(0.365)	1.000(0.355)	1.000(0.360)
	0.5	0.999(0.598)	0.999(0.578)	0.999(0.586)
	0.9	0.986(0.859)	0.985(0.865)	0.988(0.839)
10^4	0.25	0.995(0.757)	0.997(0.718)	0.996(0.731)
	0.5	0.964(0.974)	0.974(0.916)	0.973(0.941)
	0.9	0.948(0.996)	0.949(0.998)	0.949(0.999)
10^5	0.25	0.950(1.004)	0.959(0.968)	0.955(0.988)
	0.5	0.951(1.002)	0.950(0.991)	0.946(1.003)
	0.9	0.952(1.001)	0.951(0.992)	0.947(1.010)
10^6	0.25	0.947(1.005)	0.949(0.999)	0.948(1.005)
	0.5	0.951(1.005)	0.948(1.005)	0.951(1.000)
	0.9	0.954(0.995)	0.949(1.005)	0.953(1.005)
10^7	0.25	0.950(1.003)	0.956(1.001)	0.951(1.009)
	0.5	0.953(0.996)	0.950(1.000)	0.952(1.002)
	0.9	0.947(1.002)	0.948(1.004)	0.949(1.002)

Table 4.3: Empirical coverage rate(RCE) of the proposed estimator with G is uniform distribution when $\kappa = 10$.

n	r	$U(0,1)$	$N_c(0,1,\mu,\sigma^2)$	$CB(\lambda)$
10^3	0.25	1.000(0.184)	1.000(0.182)	1.000(0.183)
	0.5	1.000(0.314)	1.000(0.311)	1.000(0.313)
	0.9	1.000(0.522)	1.000(0.525)	1.000(0.510)
10^4	0.25	1.000(0.420)	1.000(0.403)	1.000(0.411)
	0.5	1.000(0.662)	1.000(0.631)	1.000(0.644)
	0.9	0.987(0.896)	0.984(0.900)	0.986(0.891)
10^5	0.25	0.996(0.814)	0.997(0.764)	0.996(0.792)
	0.5	0.963(0.985)	0.976(0.933)	0.966(0.975)
	0.9	0.949(0.997)	0.953(0.997)	0.952(0.996)
10^6	0.25	0.950(1.000)	0.962(0.974)	0.949(0.998)
	0.5	0.947(1.002)	0.951(1.005)	0.953(0.998)
	0.9	0.951(1.000)	0.949(1.002)	0.949(0.995)
10^7	0.25	0.949(0.997)	0.950(1.002)	0.951(1.000)
	0.5	0.950(1.003)	0.948(1.001)	0.950(1.000)
	0.9	0.947(1.004)	0.952(0.999)	0.954(0.996)

Table 4.4: Empirical coverage rate(RCE) of the proposed estimator with G is uniform distribution when $\kappa = 20$.

n	r	$U(0,1)$	$N_c(0,1,\mu,\sigma^2)$	$CB(\lambda)$
10^3	0.25	1.000(0.124)	1.000(0.122)	1.000(0.123)
	0.5	1.000(0.213)	1.000(0.210)	1.000(0.210)
	0.9	1.000(0.361)	1.000(0.359)	1.000(0.356)
10^4	0.25	1.000(0.284)	1.000(0.274)	1.000(0.278)
	0.5	1.000(0.461)	1.000(0.447)	1.000(0.454)
	0.9	1.000(0.713)	0.999(0.713)	1.000(0.695)
10^5	0.25	1.000(0.601)	1.000(0.571)	1.000(0.587)
	0.5	0.994(0.866)	0.997(0.808)	0.996(0.836)
	0.9	0.956(0.995)	0.963(0.980)	0.958(0.987)
10^6	0.25	0.971(0.968)	0.986(0.894)	0.978(0.938)
	0.5	0.950(1.003)	0.956(0.985)	0.952(0.997)
	0.9	0.953(0.996)	0.950(1.002)	0.950(0.999)
10^7	0.25	0.949(1.001)	0.952(0.996)	0.948(1.001)
	0.5	0.955(0.996)	0.953(0.998)	0.950(0.997)
	0.9	0.953(0.999)	0.953(0.999)	0.948(1.005)

Table 4.5: Empirical coverage rate(RCE) of the proposed estimator with G is uniform distribution when $\kappa = 30$.

n	10^3	10^4	10^5	10^6	10^7	10^8
average time (ms)	0.081	0.345	4.011	42.28	527.5	5894
standard derivation	0.001	0.005	0.021	0.171	1.5899	40.12

Table 4.6: Computation times and standard derivation under different sample sizes

4.8.2 Proof of Theorem 4.4

For the L_2 consistency, applying the Lemma 4.1 in [104], one derives that

$$\int_0^1 (\sqrt{\widehat{F}^\star}(x) - \sqrt{F}(x))^2 dG(x) = \mathcal{O}_p(n^{-2/3}).$$

By the assumption of G, one has that

$$\int_0^1 (\sqrt{\widehat{F}^\star}(x) - \sqrt{F}(x))^2 dx = \mathcal{O}_p(n^{-2/3}).$$

Note that

$$\int_0^1 (\widehat{F}^\star(x) - F(x))^2 dx = \int_0^1 (\sqrt{\widehat{F}^\star}(x) - \sqrt{F}(x))^2 (\sqrt{\widehat{F}^\star}(x) + \sqrt{F}(x))^2 dx$$
$$< 4 \int_0^1 (\sqrt{\widehat{F}^\star}(x) - \sqrt{F}(x))^2 dx = \mathcal{O}_p(n^{-2/3}).$$

By the linear transformation between $(\widehat{F}(x), F(x))$ and $(\widehat{F}^\star(x), F^\star(x))$, the frist assertion of Theorem 4.5 holds

For the uniform consistency, we divided it into three parts, i.e.,

$$\sup_{x\in[0,1]} \left|\widehat{F}(x) - F(x)\right| \leq \sup_{x\in[0,2n^{-1/3}\log n]} \left|\widehat{F}(x) - F(x)\right|$$
$$+ \sup_{x\in[2n^{-1/3}\log n, 1-2n^{-1/3}\log n]} \left|\widehat{F}(x) - F(x)\right| + \sup_{x\in[1-2n^{-1/3}\log n, 1]} \left|\widehat{F}(x) - F\right|.$$

For any $x \in [2n^{-1/3}\log n, 1 - 2n^{-1/3}\log n]$, we first consider the estimator $\widehat{F}^*(x) \in [(1-r)/2, (1+r)/2]$. Since the CDF $F^\star(x)$ has a strictly positive density on $[x - n^{-1/3}\log n, x + n^{-1/3}\log n]$, one obtains that, for some positive constants c_1, c_2,

$$\mathbb{P}\left(\left|\widehat{F}^\star(x) - F^\star(x)\right| \geq n^{-1/3}\log n\right) \leq c_1 \exp\{-c_2(\log n)^2\}$$

based on the Lemma 5.9 in [105]. If there exists a sub-interval $\mathcal{I} \subset [2n^{-1/3}\log n, 1 - 2n^{-1/3}\log n]$ such that $\widehat{F}^*(x) \notin [(1-r)/2, (1+r)/2]$, the refinement of $\widehat{F}(x)$ does not lead to a worse convergence rate obviously. Hence, for any $r \in (0,1]$

$$\mathbb{P}\left(\left|\widehat{F}(x) - F(x)\right| \geq r^{-1}n^{-1/3}\log n\right) \leq c_1 \exp\{-c_2(\log n)^2\}$$

Now, for $x_i = in^{-1/3}\log n$, $i = 2, \ldots, [n^{1/3}\log n] - 1$, one has that

$$\mathbb{P}\left(\left|\widehat{F}(x_i) - F(x_i)\right| \geq r^{-1}n^{-1/3}\log n\right) \leq c_1 \exp\{-c_2(\log n)^2\},$$

and

$$\mathbb{P}\left(\max_{2\leq i\leq [n^{1/3}\log n]-1}\left|\widehat{F}(x_i) - F(x_i)\right| \geq r^{-1}n^{-1/3}\log n\right) \leq c_1 \exp\{-c_2(\log n)^2/2\}.$$

Due to the monotonic incrementality of $\widehat{F}(x)$ and $F(x)$, one obtains that

$$\mathbb{P}\left(\sup_{x\in[2n^{-1/3}\log n, 1-2n^{-1/3}\log n]}\left|\widehat{F}(x) - F(x)\right| \geq r^{-1}n^{-1/3}\log n\right) \leq c_1 \exp\{-c_2(\log n)^2/2\}.$$

For $x \in [0, 2n^{-1/3}\log n]$, for the same arguments, one has that

$$\sup_{x\in[0,2n^{-1/3}\log n]}\left|\widehat{F}(x) - F(x)\right| \leq \left|\widehat{F}(2n^{-1/3}\log n) - F(0)\right|$$
$$\leq \left|\widehat{F}(2n^{-1/3}\log n) - F(2n^{-1/3}\log n)\right| + \left|F(2n^{-1/3}\log n) - F(0)\right| = \mathcal{O}_p(r^{-1}n^{-1/3}\log n),$$

for the reason that $F(x)$ has a strictly positive density on $[0,1]$. The left part holds for the same derivation and the proof is complete.

Proof of Theorem 4.5

Under the assumption of Theorem 4.4, for any $x_0 \in [0,1]$, such that $0 < F(x_0), G(x_0) < 1$, the CDF $F^\star(x)$ has positive density $f^\star(x_0)$. Then, following Theorem 5.1 in [105], one obtains that

$$\frac{4g(x_0)n^{1/3}(\widehat{F}^\star(x_0) - F^\star(x_0))}{F^\star(x_0)\,(1 - F^\star(x))\,f^\star(x_0)} \Rightarrow \mathcal{Z} := \arg\max_{t\in\mathbb{R}}\left\{W(t) - t^2\right\}.$$

Therefore, by the linear transformation between $(\widehat{F}(x), F(x))$ and $(\widehat{F}^\star(x), F^\star(x))$, the assertion of Theorem 4.5 holds and the proof is complete.

4.8.3 Proof of Theorem 4.6

The first assertion will be confirmed by the careful check of the proof Theorem 3.1 in [106], and $F^\star(x_0)$ has positive density on $(x_0, x_0+n^{-\gamma})$, which satisfies the assumptions of Theorem 3.1 in [106]. Then, one obtains

$$n^{1/2-\gamma/2}\left(\widehat{F}^\star(x_l) - F^\star(x_l), \widehat{F}^\star(x_r) - F^\star(x_r)\right) \Rightarrow \sqrt{\frac{F^\star(x_0)\,(1-F^\star(x_0))}{g(x_0)}}N\left(0, I_2\right).$$

Therefore, by the linear transformation between $(\widehat{F}(x), F(x))$ and $(\widehat{F}^\star(x), F^\star(x))$, the first assertion of Theorem holds.

If $\kappa_n = \kappa < \infty$, let $Z_l = \sum_{i=1}^n \Delta_i^\star \mathbf{1}_{T_i=t_l}$, $N_l = \sum_{i=1}^n$ and $\bar{Z}_l = Z_l/N_l$, $l = 1, \ldots, \kappa$. Following proposition 3.4 in [106], one has that, as $n \to \infty$,

$$\mathbb{P}\left(\bar{Z}_1 \leq \cdots \leq \bar{Z}_\kappa\right) = 1. \tag{4.9}$$

Given $\{N_l\}_{l=1}^K$, for each i draw an i.i.d. sample $\{Y_{lj}\}_{j=1}^{N_l}$ from Bernoulli$\,(1, F^\star(t_l))$. Denote $\bar{Y}_l = N_l^{-1}\sum_{j=1}^{N_l} Y_{lj}$, for each l. The second model is as follows. Suppose $\{t_l\}_{i=1}^\kappa, \{X_i^\star\}_{i=1}^n$ and $\{N_l\}_{i=1}^\kappa$ are defined as before. Let $\{Y'_{li} : 1 \leq l \leq \kappa, 1 \leq i \leq n\}$ be a family of mutually independent random variables, distributed independently of the variables in the previous sentence, such that for each i, Y'_{ij} follows Bernoulli $(1, F^\star(t_l))$ for $1 \leq j \leq n$. Denote $\bar{Y}'_l = N_l^{-1}\sum_{j=1}^n Y'_{lj}\left\{X_j^\star = t_l\right\}$ for each l. Following Lemma 1.2 in [107], one has that

$$\left(\{N_l\}, \{\bar{Z}_l\}\right) \overset{d}{=} \left(\{N_l\}, \{\bar{Y}_l\}\right) \overset{d}{=} \left(\{N_l\}, \left\{\bar{Y}'_l\right\}\right).$$

Hence, combined with (4.9), we only need to prove the asymptotic properties of $\left(\{N_l\}, \left\{\bar{Y}'_l\right\}\right)$. Then, by a triangular array version of the multivariate central limit theorem, it is sufficient to check the Lindeberg condition, and following the argument about Proof of Proposition S.1 in [107], we will obtain the second assearations, after the linear transformation between $(\widehat{F}(x), F(x))$ and $(\widehat{F}^\star(x), F^\star(x))$.

Proof of the Algorithm 7

Firstly, $\widetilde{F}$ is the unconstrained maximizer of the log-likelihood function. If $\widetilde{F}(x) \in [0,1]$, then $\widehat{F} = \widetilde{F}$ is trivially the constrained maximizer, as it is the unconstrained maximizer that also satisfies the range constraint. If not, define x^- as $\inf x : \widetilde{F}(x) > 0$ and x^+ as $\sup x : \widetilde{F}(x) < 1$. Then $x^- > 0$ or $x^+ < 1$.

Suppose there is another function, $\widehat{F}_2$, such that $L(\widehat{F}_2, \Delta, T) > L(\widehat{F}, \Delta, T)$ and $\widehat{F}_2(x) \in [0,1]$. We then define a new function $\widehat{F}_3(x)$ as:

$$\widehat{F}_3(x) = (\mathbf{1}_{x<x^-} + \mathbf{1}_{x \geq x^+})\widetilde{F}(x) + \mathbf{1}_{x^- \leq x < x^+}\widehat{F}_2(x).$$

For simplicity, denote

$$L(F, \Delta, T, i) := \Delta_i \log\left(rF(T_i) + \frac{1-r}{2}\right) + (1-\Delta_i)\log\left(\frac{1+r}{2} - rF(T_i)\right).0$$

Now we compare $L(\widehat{F}_3, \Delta, T)$ and $L(\widetilde{F}, \Delta, T)$:

$$\begin{aligned}
L(\widehat{F}_3, \Delta, T) - L(\widetilde{F}, \Delta, T) &= \sum_{i=1}^{n} \left[L(\widehat{F}_3, \Delta, T, i) - L(\widetilde{F}, \Delta, T, i)\right] \\
&= \sum_{i=1}^{n} \mathbf{1}_{x^- \leq T_i < x^+} \left[L(\widehat{F}_3, \Delta, T, i) - L(\widetilde{F}, \Delta, T, i)\right] \\
&= \sum_{i=1}^{n} \mathbf{1}_{x^- \leq T_i < x^+} \left[L(\widehat{F}_2, \Delta, T, i) - L(\widehat{F}, \Delta, T, i)\right] \\
&\geq L(\widehat{F}_2, \Delta, T) - L(\widehat{F}, \Delta, T) \\
&> 0,
\end{aligned}$$

This implies that $\widehat{F}_3$ has a higher log-likelihood than $\widetilde{F}$, contradicting the assumption that $\widetilde{F}$ is the unconstrained maximizer. Therefore, the original claim holds.

References

[3] A. Narayanan and V. Shmatikov, “How to break anonymity of the Netflix prize dataset,” *arXiv preprint cs/0610105*, 2006.

[4] A. Narayanan and V. Shmatikov, “Robust de-anonymization of large sparse datasets,” in *2008 IEEE Symposium on Security and Privacy (sp 2008)*, IEEE, 2008, pp. 111–125.

[5] M. Barbaro and J. T. Zeller, “A face is exposed for aol searcher no. 4417749,” *New York Times (Aug, 9, 2006)*, 2006.

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

[8] Ú. Erlingsson, V. Pihur, and A. Korolova, “Rappor: Randomized aggregatable privacy-preserving ordinal response,” in *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, 2014, pp. 1054–1067.

[53] T. T. Cai, Y. Wang, and L. Zhang, “The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy,” *The Annals of Statistics*, vol. 49, no. 5, pp. 2825–2850, 2021.

[54] M. Joseph, J. Mao, S. Neel, and A. Roth, “The role of interactivity in local differential privacy,” in *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, 2019, pp. 94–105. DOI: 10.1109/FOCS.2019.00015.

[71] G. R. Shorack and J. A. Wellner, *Empirical processes with applications to statistics.* SIAM, 2009.

[72] J. Komlós, P. Major, and G. Tusnády, “An approximation of partial sums of independent rv’-s, and the sample df. i,” *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 32, pp. 111–131, 1975.

[73] H. Yamato, “Uniform convergence of an estimator of a distribution function,” 1973.

[74] L. Xue and J. Wang, “Distribution function estimation by constrained polynomial spline regression,” *Journal of Nonparametric Statistics*, vol. 22, no. 4, pp. 443–457, 2010.

[75] R. Liu and L. Yang, “Kernel estimation of multivariate cumulative distribution function,” *Journal of Nonparametric Statistics*, vol. 20, no. 8, pp. 661–677, 2008.

[76] H. Dehling and W. Philipp, *Empirical process techniques for dependent data.* Springer, 2002.

[77] S. N. Lahiri, M. S. Kaiser, N. Cressie, and N.-J. Hsu, “Prediction of spatial cumulative distribution functions using subsampling,” *Journal of the American Statistical Association*, vol. 94, no. 445, pp. 86–97, 1999.

[78] J. Hua, Z. Shen, and S. Zhong, "We can track you if you take the metro: Tracking metro riders using accelerometers on smartphones," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 2, pp. 286–297, 2016.

[79] L. Zhang, P. H. Pathak, M. Wu, Y. Zhao, and P. Mohapatra, "Accelword: Energy efficient hotword detection through accelerometer," in *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, 2015, pp. 301–315.

[80] S. A. Anand, C. Wang, J. Liu, N. Saxena, and Y. Chen, "Spearphone: A speech privacy exploit via accelerometer-sensed reverberations from smartphone loudspeakers," *arXiv preprint arXiv:1907.05972*, 2019.

[81] J. Liu *et al.*, "Privacy leakage in wireless charging," *IEEE Transactions on Dependable and Secure Computing*, 2022.

[82] I. Lee, "An analysis of data breaches in the us healthcare industry: Diversity, trends, and risk profiling," *Information Security Journal: A Global Perspective*, vol. 31, no. 3, pp. 346–358, 2022.

[83] R. Ayyagari, "An exploratory analysis of data breaches from 2005-2011: Trends and insights," *Journal of Information Privacy and Security*, vol. 8, no. 2, pp. 33–56, 2012.

[84] S. Quach, P. Thaichon, K. D. Martin, S. Weaven, and R. W. Palmatier, "Digital technologies: Tensions in privacy and data," *Journal of the Academy of Marketing Science*, vol. 50, no. 6, pp. 1299–1323, 2022.

[85] R. Bassily and A. Smith, "Local, private, efficient protocols for succinct histograms," in *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, 2015, pp. 127–135.

[86] J. Acharya, Z. Sun, and H. Zhang, "Hadamard response: Estimating distributions privately, efficiently, and with little communication," in *The 22nd International Conference on Artificial Intelligence and Statistics*, PMLR, 2019, pp. 1120–1129.

[87] Z. Li, T. Wang, M. Lopuhaä-Zwakenberg, N. Li, and B. Škoric, "Estimating numerical distributions under local differential privacy," in *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*, 2020, pp. 621–635.

[88] C. Durot and H. P. Lopuhaä, "Limit theory in monotone function estimation," *Statistical science*, vol. 33, no. 4, pp. 547–567, 2018.

[89] N. P. Jewell and M. van der Laan, "Generalizations of current status data with applications," *Lifetime data analysis*, vol. 1, pp. 101–109, 1995.

[90] A. Rossini and A. Tsiatis, "A semiparametric proportional odds regression model for the analysis of current status data," *Journal of the American Statistical Association*, vol. 91, no. 434, pp. 713–721, 1996.

[91] C. Aarts, E. Kylberg, A. Hörnell, Y. Hofvander, M. Gebre-Medhin, and T. Greiner, "How exclusive is exclusive breastfeeding? a comparison of data since birth with current status data," *International Journal of epidemiology*, vol. 29, no. 6, pp. 1041–1046, 2000.

[92] L. Wang, J. Sun, and X. Tong, "Efficient estimation for the proportional hazards model with bivariate current status data," *Lifetime Data Analysis*, vol. 14, pp. 134–153, 2008.

[93] V. G. Sal y Rosas and J. P. Hughes, "Nonparametric and semiparametric analysis of current status data subject to outcome misclassification," *Statistical Communications in Infectious Diseases*, vol. 3, no. 1, 2011.

[94] J. Fan, "On the optimal rates of convergence for nonparametric deconvolution problems," *The Annals of Statistics*, pp. 1257–1272, 1991.

[95] J. Fan, "Deconvolution with supersmooth distributions," *Canadian Journal of Statistics*, vol. 20, no. 2, pp. 155–169, 1992.

[96] C. Dwork, A. Roth, *et al.*, "The algorithmic foundations of differential privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014.

[97] J. Huang and J. A. Wellner, "Interval censored survival data: A review of recent progress," in *Proceedings of the first Seattle symposium in biostatistics: survival analysis*, Springer, 1997, pp. 123–169.

[98] C.-C. Hill, "Likelihood based inference for current status data on a grid: A boundary phenomenon and an adaptive inference procedure by runlong tang, moulinath banerjee michael r. kosorok,"

[99] T. Robertson, "Order restricted statistical inference," Tech. Rep., 1988.

[100] B. Klaus and K. Strimmer., *Fdrtool: Estimation of (local) false discovery rates and higher criticism*, R package version 1.2.15, 2015. [Online]. Available: https://CRAN.R-project.org/package=fdrtool.

[101] P. Groeneboom, "Estimating a monotone density," *Department of Mathematical Statistics*, no. R 8403, 1984.

[102] Apple, *Differential privacy overview - apple.* [Online]. Available: https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf.

[103] J. Tang, A. Korolova, X. Bai, X. Wang, and X. Wang, "Privacy loss in apple's implementation of differential privacy on macos 10.12," *arXiv preprint arXiv:1709.02753*, 2017.

[104] J. Huang and J. A. Wellner, "Asymptotic normality of the npmle of linear functionals for interval censored data, case 1," *Statistica Neerlandica*, vol. 49, no. 2, pp. 153–163, 1995.

[105] P. Groeneboom and J. A. Wellner, *Information bounds and nonparametric maximum likelihood estimation.* Springer Science & Business Media, 1992, vol. 19.

[106] R. Tang, M. Banerjee, and M. R. Kosorok, “Likelihood based inference for current status data on a grid: A boundary phenomenon and an adaptive inference procedure.,” *The Annals of Statistics*, vol. 40, no. 1, pp. 45–72, 2012.

[107] R. Tang, M. Banerjee, and M. R. Kosorok, “Supplement to “likelihood based inference for current status data on a grid: A boundary phenomenon and an adaptive inference procedure”.,” 2012.

Chapter 5

Conclusions and Future Work

In conclusion, this thesis has made significant strides in the fields of GDP and LDP by investigating their properties, relationships, and applications for efficient and accurate privacy-preserving data analysis. Through the development of identification, measurement, and amplification tools, legacy datasets and mechanisms can be easily integrated into the new GDP framework with little or no modification. By employing self-normalization techniques and binary inquiries, a method for LDP quantile estimation is established, complete with valid confidence intervals. Lastly, by discovering a link between LDP CDF estimation and the current status problem, a framework is introduced that guarantees convergence and derives the asymptotic distribution of error. Collectively, this research advances the state of the art in differential privacy, addressing pressing challenges in privacy preservation and paving the way for more secure and privacy-aware data analysis in today's digital age.

This thesis also opens up avenues for future research. The idea behind the GDP framework can be generalized to other parameterized DP notions like CDP or RDP to enhance tractability and visualizability in the DP literature. One-dimensional quantile and CDF estimation can be extended to multi-dimensional variables, leading to interesting discussions about balancing privacy budgets among distributions. The iterative and self-normalization techniques can also be extended to regression problems, building upon the established quantile estimation to develop corresponding quantile regression. Furthermore, the statistical inference methods presented have the potential to be generalized for multi-sample inference in relevant change detection problems.

Bibliography

[1] A. R. Miller, “Personal privacy in the computer age: The challenge of a new technology in an information-oriented society,” *Michigan Law Review*, vol. 67, no. 6, pp. 1089–1246, 1969.

[2] E. Shils, “Privacy: Its constitution and vicissitudes,” *Law and Contemporary Problems*, vol. 31, no. 2, pp. 281–306, 1966.

[3] A. Narayanan and V. Shmatikov, “How to break anonymity of the Netflix prize dataset,” *arXiv preprint cs/0610105*, 2006.

[4] A. Narayanan and V. Shmatikov, “Robust de-anonymization of large sparse datasets,” in *2008 IEEE Symposium on Security and Privacy (sp 2008)*, IEEE, 2008, pp. 111–125.

[5] M. Barbaro and J. T. Zeller, “A face is exposed for aol searcher no. 4417749,” *New York Times (Aug, 9, 2006)*, 2006.

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

[7] F. K. Dankar and K. El Emam, “The application of differential privacy to health data,” in *Proceedings of the 2012 Joint EDBT/ICDT Workshops*, 2012, pp. 158–166.

[8] Ú. Erlingsson, V. Pihur, and A. Korolova, “Rappor: Randomized aggregatable privacy-preserving ordinal response,” in *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, 2014, pp. 1054–1067.

[9] G. Cormode, S. Jha, T. Kulkarni, N. Li, D. Srivastava, and T. Wang, “Privacy at scale: Local differential privacy in practice,” in *Proceedings of the 2018 International Conference on Management of Data*, 2018, pp. 1655–1658.

[10] M. U. Hassan, M. H. Rehmani, and J. Chen, “Differential privacy techniques for cyber physical systems: A survey,” *IEEE Communications Surveys & Tutorials*, vol. 22, no. 1, pp. 746–789, 2019.

[11] L. Wasserman and S. Zhou, “A statistical framework for differential privacy,” *Journal of the American Statistical Association*, vol. 105, no. 489, pp. 375–389, 2010.

[12] J. Dong, A. Roth, and W. J. Su, "Gaussian differential privacy," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2021.

[13] I. Mironov, "Rényi differential privacy," in *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, IEEE, 2017, pp. 263–275.

[14] M. Bun and T. Steinke, "Concentrated differential privacy: Simplifications, extensions, and lower bounds," vol. 9985, Nov. 2016, pp. 635–658, ISBN: 978-3-662-53640-7. DOI: 10.1007/978-3-662-53641-4_24.

[15] M. Bun, C. Dwork, G. N. Rothblum, and T. Steinke, "Composable and versatile privacy via truncated cdp," in *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, 2018, pp. 74–86.

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

[17] B. Balle, G. Barthe, and M. Gaboardi, "Privacy profiles and amplification by subsampling," *Journal of Privacy and Confidentiality*, vol. 10, no. 1, 2020.

[18] N. Phan, X. Wu, H. Hu, and D. Dou, "Adaptive laplace mechanism: Differential privacy preservation in deep learning," in *2017 IEEE International Conference on Data Mining*, IEEE, 2017, pp. 385–394.

[19] Y. Hu, P. Liu, L. Kong, and D. Niu, "Learning privately over distributed features: An admm sharing approach," *arXiv preprint arXiv:1907.07735*, 2019.

[20] X. Xu, Y. Yao, and L. Cheng, "Deep learning algorithms design and implementation based on differential privacy," in *International Conference on Machine Learning for Cyber Security*, Springer, 2020, pp. 317–330.

[21] T. Li and C. Clifton, "Differentially private imaging via latent space manipulation," *arXiv preprint arXiv:2103.05472*, 2021.

[22] M. Gaboardi, H. Lim, R. Rogers, and S. Vadhan, "Differentially private chi-squared hypothesis testing: Goodness of fit and independence testing," in *International Conference on Machine Learning*, PMLR, 2016, pp. 2111–2120.

[23] R. Bassily, A. Smith, and A. Thakurta, "Private empirical risk minimization: Efficient algorithms and tight error bounds," in *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, IEEE, 2014, pp. 464–473.

[24] M. Abadi *et al.*, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016, pp. 308–318.

[25] V. Feldman, I. Mironov, K. Talwar, and A. Thakurta, "Privacy amplification by iteration," in *2018 IEEE 59th Annual Symposium on Foundations of Computer Science*, IEEE, 2018, pp. 521–532.

[26] G. Qiao, W. Su, and L. Zhang, "Oneshot differentially private top-k selection," in *Proceedings of the 38th International Conference on Machine Learning*, M. Meila and T. Zhang, Eds., ser. Proceedings of Machine Learning Research, vol. 139, PMLR, 2021, pp. 8672–8681.

[27] Y. Ishai, E. Kushilevitz, R. Ostrovsky, and A. Sahai, "Cryptography from anonymity," in *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, IEEE, 2006, pp. 239–248.

[28] B. Ghazi, R. Pagh, and A. Velingker, "Scalable and differentially private distributed aggregation in the shuffled model," *arXiv preprint arXiv:1906.08320*, 2019.

[29] B. Balle, J. Bell, A. Gascón, and K. Nissim, "Private summation in the multi-message shuffle model," in *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, 2020, pp. 657–676.

[30] B. Ghazi, P. Manurangsi, R. Pagh, and A. Velingker, "Private aggregation from fewer anonymous messages," *Advances in Cryptology–EUROCRYPT 2020*, vol. 12106, p. 798, 2020.

[31] M. A. Malcolm, "On accurate floating-point summation," *Communications of the ACM*, vol. 14, no. 11, pp. 731–736, 1971.

[32] A. Cuyt, B. Verdonk, S. Becuwe, and P. Kuterna, "A remarkable example of catastrophic cancellation unraveled," *Computing*, vol. 66, no. 3, pp. 309–320, 2001.

[33] J. Murtagh and S. Vadhan, "The complexity of computing the optimal composition of differential privacy," in *Theory of Cryptography Conference*, Springer, 2016, pp. 157–175.

[34] P. Kairouz, S. Oh, and P. Viswanath, "The composition theorem for differential privacy," in *International conference on machine learning*, PMLR, 2015, pp. 1376–1385.

[35] C. Dwork, G. N. Rothblum, and S. Vadhan, "Boosting and differential privacy," in *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, IEEE, 2010, pp. 51–60.

[36] M. Abramowitz, I. A. Stegun, and R. H. Romer, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, 1988.

[37] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith, "What can we learn privately?" *SIAM Journal on Computing*, vol. 40, no. 3, pp. 793–826, 2011.

[38] A. Evfimievski, J. Gehrke, and R. Srikant, "Limiting privacy breaches in privacy preserving data mining," in *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, 2003, pp. 211–222.

[39] S. L. Warner, "Randomized response: A survey technique for eliminating evasive answer bias," *Journal of the American Statistical Association*, vol. 60, no. 309, pp. 63–69, 1965.

[40] K. Amin, M. Joseph, and J. Mao, "Pan-private uniformity testing," in *Proceedings of Thirty Third Conference on Learning Theory*, J. Abernethy and S. Agarwal, Eds., ser. Proceedings of Machine Learning Research, vol. 125, PMLR, 2020, pp. 183–218.

[41] B. Ding, J. Kulkarni, and S. Yekhanin, "Collecting telemetry data privately," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[42] C. Dwork and J. Lei, "Differential privacy and robust statistics," in *Proceedings of the forty-first annual ACM symposium on Theory of computing*, 2009, pp. 371–380.

[43] J. Lei, "Differentially private m-estimators," *Advances in Neural Information Processing Systems*, vol. 24, 2011.

[44] A. Smith, "Privacy-preserving statistical estimation with optimal convergence rates," in *Proceedings of the forty-third annual ACM symposium on Theory of computing*, 2011, pp. 813–822.

[45] J. Gillenwater, M. Joseph, and A. Kulesza, "Differentially private quantiles," in *Proceedings of the 38th International Conference on Machine Learning*, M. Meila and T. Zhang, Eds., ser. Proceedings of Machine Learning Research, vol. 139, PMLR, 2021, pp. 3713–3722.

[46] D. Alabi, O. Ben-Eliezer, and A. Chaturvedi, "Bounded space differentially private quantiles," *arXiv preprint arXiv:2201.03380*, 2022.

[47] O. Ben-Eliezer, D. Mikulincer, and I. Zadik, "Archimedes meets privacy: On privately estimating quantiles in high dimensions under minimal assumptions," *arXiv preprint arXiv:2208.07438*, 2022.

[48] X. Shao, "A self-normalized approach to confidence interval construction in time series," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 72, no. 3, pp. 343–366, 2010.

[49] P. Jain, P. Kothari, and A. Thakurta, "Differentially private online learning," in *Conference on Learning Theory*, JMLR Workshop and Conference Proceedings, 2012, pp. 24–1.

[50] N. Agarwal and K. Singh, "The price of differential privacy for online learning," in *Proceedings of the 34th International Conference on Machine Learning*, D. Precup and Y. W. Teh, Eds., ser. Proceedings of Machine Learning Research, vol. 70, PMLR, 2017, pp. 32–40.

[51] K. Wei *et al.*, "Federated learning with differential privacy: Algorithms and performance analysis," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3454–3469, 2020. DOI: 10.1109/TIFS.2020.2988575.

[52] S. Lee, Y. Liao, M. H. Seo, and Y. Shin, "Fast and robust online inference with stochastic gradient descent via random scaling," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, 2022, pp. 7381–7389.

[53] T. T. Cai, Y. Wang, and L. Zhang, "The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy," *The Annals of Statistics*, vol. 49, no. 5, pp. 2825–2850, 2021.

[54] M. Joseph, J. Mao, S. Neel, and A. Roth, "The role of interactivity in local differential privacy," in *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, 2019, pp. 94–105. DOI: 10.1109/FOCS.2019.00015.

[55] A. Cheu, A. Smith, J. Ullman, D. Zeber, and M. Zhilyaev, "Distributed differential privacy via shuffling," in *Advances in Cryptology–EUROCRYPT 2019: 38th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Darmstadt, Germany, May 19–23, 2019, Proceedings, Part I 38*, Springer, 2019, pp. 375–403.

[56] A. G. Joseph and S. Bhatnagar, "A stochastic approximation algorithm for quantile estimation," in *International Conference on Neural Information Processing*, Springer, 2015, pp. 311–319.

[57] B. Coppens, I. Verbauwhede, K. De Bosschere, and B. De Sutter, "Practical mitigations for timing-based side-channel attacks on modern x86 processors," in *2009 30th IEEE Symposium on Security and Privacy*, 2009, pp. 45–60. DOI: 10.1109/SP.2009.19.

[58] N. Lawson, "Side-channel attacks on cryptographic software," *IEEE Security & Privacy*, vol. 7, no. 6, pp. 65–68, 2009. DOI: 10.1109/MSP.2009.165.

[59] B. Balle, G. Barthe, and M. Gaboardi, "Privacy amplification by subsampling: Tight analyses via couplings and divergences," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[60] Y. Liu, K. Sun, L. Kong, and B. Jiang, "Identification, amplification and measurement: A bridge to gaussian differential privacy," *Advances in Neural Information Processing Systems*, 2022.

[61] T. C. Brown, P. A. Champ, R. C. Bishop, and D. W. McCollum, "Which response format reveals the truth about donations to a public good?" *Land Economics*, pp. 152–166, 1996.

[62] I. Mironov, "On significance of the least significant bits for differential privacy," in *Proceedings of the 2012 ACM conference on Computer and communications security*, 2012, pp. 650–661.

[63] J. Jin, E. McMurtry, B. I. Rubinstein, and O. Ohrimenko, "Are we there yet? timing and floating-point attacks on differential privacy systems," *arXiv preprint arXiv:2112.05307*, 2021.

[64] S. Haney, D. Desfontaines, L. Hartman, R. Shrestha, and M. Hay, "Precision-based attacks and interval refining: How to break, then fix, differential privacy on finite computers," *arXiv preprint arXiv:2207.13793*, 2022.

[65] Y. Fang, J. Xu, and L. Yang, "Online bootstrap confidence intervals for the stochastic gradient descent estimator," *The Journal of Machine Learning Research*, vol. 19, no. 1, pp. 3053–3073, 2018.

[66] X. Shao, "Self-normalization for time series: A review of recent developments," *Journal of the American Statistical Association*, vol. 110, no. 512, pp. 1797–1817, 2015.

[67] J. Dong, W. Su, and L. Zhang, "A central limit theorem for differentially private query answering," in *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds., vol. 34, Curran Associates, Inc., 2021, pp. 14 759–14 770.

[68] C. E. Clark, "The pert model for the distribution of an activity time," *Operations Research*, vol. 10, no. 3, pp. 405–406, 1962.

[69] P. Langley, "Crafting papers on machine learning," in *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, P. Langley, Ed., Stanford, CA: Morgan Kaufmann, 2000, pp. 1207–1216.

[70] J. Dippon, "Globally convergent stochastic optimization with optimal asymptotic distribution," *Journal of applied probability*, vol. 35, no. 2, pp. 395–406, 1998.

[71] G. R. Shorack and J. A. Wellner, *Empirical processes with applications to statistics.* SIAM, 2009.

[72] J. Komlós, P. Major, and G. Tusnády, "An approximation of partial sums of independent rv'-s, and the sample df. i," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 32, pp. 111–131, 1975.

[73] H. Yamato, "Uniform convergence of an estimator of a distribution function," 1973.

[74] L. Xue and J. Wang, "Distribution function estimation by constrained polynomial spline regression," *Journal of Nonparametric Statistics*, vol. 22, no. 4, pp. 443–457, 2010.

[75] R. Liu and L. Yang, "Kernel estimation of multivariate cumulative distribution function," *Journal of Nonparametric Statistics*, vol. 20, no. 8, pp. 661–677, 2008.

[76] H. Dehling and W. Philipp, *Empirical process techniques for dependent data.* Springer, 2002.

[77] S. N. Lahiri, M. S. Kaiser, N. Cressie, and N.-J. Hsu, "Prediction of spatial cumulative distribution functions using subsampling," *Journal of the American Statistical Association*, vol. 94, no. 445, pp. 86–97, 1999.

[78] J. Hua, Z. Shen, and S. Zhong, "We can track you if you take the metro: Tracking metro riders using accelerometers on smartphones," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 2, pp. 286–297, 2016.

[79] L. Zhang, P. H. Pathak, M. Wu, Y. Zhao, and P. Mohapatra, "Accelword: Energy efficient hotword detection through accelerometer," in *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, 2015, pp. 301–315.

[80] S. A. Anand, C. Wang, J. Liu, N. Saxena, and Y. Chen, "Spearphone: A speech privacy exploit via accelerometer-sensed reverberations from smartphone loudspeakers," *arXiv preprint arXiv:1907.05972*, 2019.

[81] J. Liu *et al.*, "Privacy leakage in wireless charging," *IEEE Transactions on Dependable and Secure Computing*, 2022.

[82] I. Lee, "An analysis of data breaches in the us healthcare industry: Diversity, trends, and risk profiling," *Information Security Journal: A Global Perspective*, vol. 31, no. 3, pp. 346–358, 2022.

[83] R. Ayyagari, "An exploratory analysis of data breaches from 2005-2011: Trends and insights," *Journal of Information Privacy and Security*, vol. 8, no. 2, pp. 33–56, 2012.

[84] S. Quach, P. Thaichon, K. D. Martin, S. Weaven, and R. W. Palmatier, "Digital technologies: Tensions in privacy and data," *Journal of the Academy of Marketing Science*, vol. 50, no. 6, pp. 1299–1323, 2022.

[85] R. Bassily and A. Smith, "Local, private, efficient protocols for succinct histograms," in *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, 2015, pp. 127–135.

[86] J. Acharya, Z. Sun, and H. Zhang, "Hadamard response: Estimating distributions privately, efficiently, and with little communication," in *The 22nd International Conference on Artificial Intelligence and Statistics*, PMLR, 2019, pp. 1120–1129.

[87] Z. Li, T. Wang, M. Lopuhaä-Zwakenberg, N. Li, and B. Škoric, "Estimating numerical distributions under local differential privacy," in *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*, 2020, pp. 621–635.

[88] C. Durot and H. P. Lopuhaä, "Limit theory in monotone function estimation," *Statistical science*, vol. 33, no. 4, pp. 547–567, 2018.

[89] N. P. Jewell and M. van der Laan, "Generalizations of current status data with applications," *Lifetime data analysis*, vol. 1, pp. 101–109, 1995.

[90] A. Rossini and A. Tsiatis, "A semiparametric proportional odds regression model for the analysis of current status data," *Journal of the American Statistical Association*, vol. 91, no. 434, pp. 713–721, 1996.

[91] C. Aarts, E. Kylberg, A. Hörnell, Y. Hofvander, M. Gebre-Medhin, and T. Greiner, "How exclusive is exclusive breastfeeding? a comparison of data since birth with current status data," *International Journal of epidemiology*, vol. 29, no. 6, pp. 1041–1046, 2000.

[92] L. Wang, J. Sun, and X. Tong, "Efficient estimation for the proportional hazards model with bivariate current status data," *Lifetime Data Analysis*, vol. 14, pp. 134–153, 2008.

[93] V. G. Sal y Rosas and J. P. Hughes, "Nonparametric and semiparametric analysis of current status data subject to outcome misclassification," *Statistical Communications in Infectious Diseases*, vol. 3, no. 1, 2011.

[94] J. Fan, "On the optimal rates of convergence for nonparametric deconvolution problems," *The Annals of Statistics*, pp. 1257–1272, 1991.

[95] J. Fan, "Deconvolution with supersmooth distributions," *Canadian Journal of Statistics*, vol. 20, no. 2, pp. 155–169, 1992.

[96] C. Dwork, A. Roth, *et al.*, "The algorithmic foundations of differential privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014.

[97] J. Huang and J. A. Wellner, "Interval censored survival data: A review of recent progress," in *Proceedings of the first Seattle symposium in biostatistics: survival analysis*, Springer, 1997, pp. 123–169.

[98] C.-C. Hill, "Likelihood based inference for current status data on a grid: A boundary phenomenon and an adaptive inference procedure by runlong tang, moulinath banerjee michael r. kosorok,"

[99] T. Robertson, "Order restricted statistical inference," Tech. Rep., 1988.

[100] B. Klaus and K. Strimmer., *Fdrtool: Estimation of (local) false discovery rates and higher criticism*, R package version 1.2.15, 2015. [Online]. Available: https://CRAN.R-project.org/package=fdrtool.

[101] P. Groeneboom, "Estimating a monotone density," *Department of Mathematical Statistics*, no. R 8403, 1984.

[102] Apple, *Differential privacy overview - apple.* [Online]. Available: https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf.

[103] J. Tang, A. Korolova, X. Bai, X. Wang, and X. Wang, "Privacy loss in apple's implementation of differential privacy on macos 10.12," *arXiv preprint arXiv:1709.02753*, 2017.

[104] J. Huang and J. A. Wellner, "Asymptotic normality of the npmle of linear functionals for interval censored data, case 1," *Statistica Neerlandica*, vol. 49, no. 2, pp. 153–163, 1995.

[105] P. Groeneboom and J. A. Wellner, *Information bounds and nonparametric maximum likelihood estimation.* Springer Science & Business Media, 1992, vol. 19.

[106] R. Tang, M. Banerjee, and M. R. Kosorok, "Likelihood based inference for current status data on a grid: A boundary phenomenon and an adaptive inference procedure.," *The Annals of Statistics*, vol. 40, no. 1, pp. 45–72, 2012.

[107] R. Tang, M. Banerjee, and M. R. Kosorok, "Supplement to "likelihood based inference for current status data on a grid: A boundary phenomenon and an adaptive inference procedure".," 2012.