Poster: Improving the Adaptability of Differential Privacy

Vaikkunth Mugunthan
CSAIL
Massachusetts Institute of Technology
Cambridge, USA
vaik@mit.edu

Wanyi Xiao
CSAIL
Massachusetts Institute of Technology
Cambridge, USA
wanyi@mit.edu

Lalana Kagal
CSAIL
Massachusetts Institute of Technology
Cambridge, USA
lkagal@mit.edu

***Abstract*—Differential privacy is a mathematical technique that provides strong theoretical privacy guarantees by ensuring statistical indistinguishability of individuals in a dataset. It has become the de facto framework for providing privacy-preserving data analysis over statistical datasets. Differential privacy has garnered significant attention from researchers and privacy experts due to its strong privacy guarantees. In differential privacy, the standard approach is to add Laplacian noise to the output of queries. However, the lack of flexibility due to the dearth of configurable parameters in existing mechanisms, and the accuracy loss caused by the noise added have prevented its widespread adoption in the industry. We propose new probability distributions and noise adding mechanisms that preserve (ϵ)-differential privacy and (ϵ, δ)-differential privacy. The distributions can be observed as an asymmetric Laplacian distribution and a generalized truncated Laplacian distribution. We show that the proposed mechanisms add optimal noise in a global context, conditional upon technical lemmas. In addition, we also show that the proposed mechanisms have greater adaptability than Laplacian noise as there is more than one parameter to adjust. The presented mechanisms are highly useful as they enable data controllers to fine-tune the perturbation necessary to protect privacy to use case specific distortion requirements.**

***Index Terms*—Differential Privacy, Asymmetric Laplace Distribution, Truncated Laplace Distribution**

I. Introduction

Differentially private methods are used to publish or release statistics of a dataset as a whole while protecting the sensitive information of individuals in the dataset. Intuitively, for a given individual who is considering participating in a dataset, differential privacy requires that an analyst learns no more information from a dataset that contains this individuals information than one that does not. Essentially, differential privacy guarantees that the released results reveal little or no new information about an individual in the dataset. Differential privacy guarantees that if a particular individuals information was to be removed from the dataset, the released result would not be significantly different. As no individual sample can affect the output, attackers can thus not infer the private information corresponding to an individual sample. Though there has been a myriad of significant contributions in the field of differential privacy, it has not yet been adopted by many in the industry due to: i) lack of flexibility in the mechanisms due to the dearth of configurable parameters, and ii) concerns over reduced utility and privacy.

II. Our Contributions

We propose new probability distributions and noise adding mechanisms that preserve (ϵ)-differential privacy and (ϵ, δ)-differential privacy. The distributions can be observed as an asymmetric Laplacian distribution and a generalized truncated Laplacian distribution. We show that the proposed mechanisms add optimal noise in a global context, conditional upon technical lemmas. In addition, we also show that the proposed mechanisms have greater adaptability than Laplacian noise as there is more than one parameter to adjust. This also mitigates the problems pertaining to inaccuracy and provides better utility in bounding noise.

III. Generalized Truncated Laplace Mechanism

First, we talk about the probability distribution from which noise can be drawn from to preserve (ϵ, δ)-differential privacy. The probability distribution can be viewed as a generalized truncated Laplace distribution. Such a probability distribution is motivated by the symmetrically bounded Laplace distribution proposed by [1]. The proposed distribution is a more general version as it is asymmetrically bounded. This is shown in Fig. 1.

To construct such a distribution, we set the privacy parameter ϵ and δ. In contrast to most of the existing (ϵ, δ)-differential private mechanisms, where ϵ and δ are the only two variables in the algorithm design, the general truncated Laplace distribution allows another parameter to specify the upper or lower bound of the probability density function. Therefore, with the additional bounding parameter, not depending on the value of ϵ or δ, the proposed generalized truncated Laplace distribution provides more flexibility.

Definition 1. The zero-mean generalized truncated Laplace distribution has a probability density function $f(x)$ with scale λ, and is asymmetrically bounded by A and B where $A < 0 < B$, defined as:

$$f(x) = \begin{cases} Me^{-\frac{|x|}{\lambda}} \text{ for } x \in [A, B] \\ 0 \text{ otherwise} \end{cases}$$

where $M = \frac{1}{\lambda(2 - e^{\frac{A}{\lambda}} - e^{-\frac{B}{\lambda}})}$.

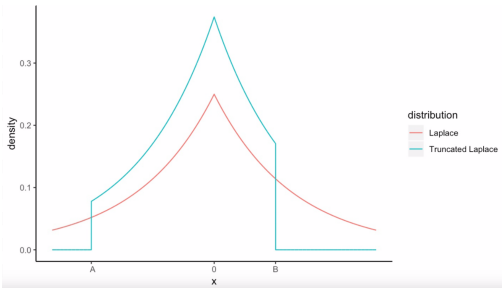

Fig. 1. Laplacian Mechanism vs Generalized Truncated Laplacian Mechanism

Given the global sensitivity, Δ, of the query function q, and the privacy parameters ϵ, δ, the *General Truncated Laplacian mechanism* $\mathcal{A}$ uses random noise X drawn from the General Truncated Laplace distribution in Definition 4 with the following parameters:

$$\lambda = \frac{\Delta}{\epsilon} \; and \; A + \Delta \leq 0 \leq B - \Delta$$

If $|A| \geq |B|$,

$$\begin{cases} A = \lambda \ln \left[2 + (\frac{1-\delta}{\delta}) e^{-\frac{B}{\lambda}} - (\frac{1}{\delta}) e^{-\frac{B-\Delta}{\lambda}} \right] \\ B = \text{any positive real number satisfy} |A| \geq |B| \end{cases} ;$$

If $|A| < |B|$,

$$\begin{cases} A = \text{any negative real number satisfy } |A| < |B| \\ B = -\lambda \ln \left[2 + (\frac{1-\delta}{\delta}) e^{\frac{A}{\lambda}} - (\frac{1}{\delta}) e^{\frac{A+\Delta}{\lambda}} \right] \end{cases} .$$

Theorem 1. *The General Truncated Laplacian mechanism preserves (ϵ, δ)-differential privacy.*

We plan to prove Theorem 1 using the following two lemmas.

Lemma 1.

$$\max \left(\int_A^{A+\Delta} f(x)dx, \int_{B-\Delta}^{B} f(x)dx \right) = \delta$$

for the probability density function $f(x)$, Δ, A and B of the General Truncated Laplace distribution.

Lemma 2. *A mechanism $\mathcal{A}(\mathcal{D}) = q(\mathcal{D}) + X$ that adds a random noise X drawn from probability distribution $\mathcal{P}$ with probability density function $f(x)$, satisfies (ϵ, δ)-differential privacy when*

$$\mathcal{P}(\mathcal{S}) - e^{\epsilon} \mathcal{P}(\mathcal{S} + d) \leq \delta$$

holds for any $|d| \leq \Delta$, and any measurable set $\mathcal{S} \subseteq \mathbb{R}$, where Δ is the global sensitivity for the query function q.

Remark 1. *We claim that*

$$0 < \delta \leq \min \left(\int_A^0 f(x)dx, \int_0^B f(x)dx \right).$$

IV. Asymmetric Laplace Mechanism

The Asymmetric Laplacian mechanism is an ϵ-differentially private mechanism that offers better flexibility in terms of privacy and accuracy than the ubiquitously-used Laplacian mechanism. The asymmetric Laplacian mechanism uses random noise drawn from the asymmetric Laplace distribution. The asymmetric Laplace distribution is a generalization of Laplace distribution that consists of two exponential distributions of unequal scale back to back.

We set an asymmetry parameter k, which controls how unequal the two exponential distributions are. Therefore, with an extra asymmetry parameter k, the asymmetric Laplace probability distribution also provides more *flexibility* to the dataset curator in mechanism design. Fig. 2 represents the probability density functions with different values of k.

The asymmetric Laplace distribution has probability density function $f(x, \lambda, k)$ defined as:

$$f(x, \lambda, k) = \frac{\lambda}{k + \frac{1}{k}} \begin{cases} e^{\frac{\lambda x}{k}} & \text{for } x < 0 \\ e^{-\lambda k x} & \text{for } x \geq 0 \end{cases} .$$

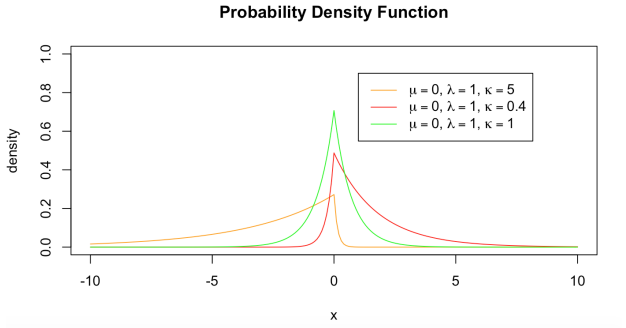

Fig. 2. Asymmetric Laplacian Distribution

Given the global sensitivity, Δ, of the query function q, and the privacy parameter ϵ, the *Asymmetric Laplacian mechanism* $\mathcal{A}$ uses random noise X drawn from the asymmetric Laplacian distribution with scale λ.

Theorem 2. *The Asymmetric Laplacian mechanism preserves $(\max(\frac{\epsilon}{k}, k\epsilon))$-differential privacy.*

To prove our mechanism preserves $(\max(\frac{\epsilon}{k}, k\epsilon))$ differential privacy, we need to show that for $\mathcal{D}_1 \sim \mathcal{D}_2$,

$$\Pr\left[\mathcal{A}(\mathcal{D}_1) \in \mathcal{T}\right] \leq e^{\max(\frac{\epsilon}{k}, k\epsilon)} \Pr\left[\mathcal{A}(\mathcal{D}_2) \in \mathcal{T}\right]$$

for any subset $\mathcal{T} \subseteq \mathcal{O}$, where $\mathcal{O}$ is the set of all outputs of the mechanism.

V. Conclusion and Future Work

Our initial results show that compared to the optimal Gaussian mechanism, the generalized truncated Laplacian mechanism reduces the noise power and noise amplitude across all privacy regimes. We plan to explore flexible and optimal differentially private mechanisms that merge more than one probability distribution.

References

[1] Quan Geng, Wei Ding, Ruiqi Guo, and Sanjiv Kumar. Truncated laplacian mechanism for approximate differential privacy. *CoRR*, abs/1810.00877, 2018.