Poster: Novel strategies to calculate the privacy loss parameter in 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*—The privacy parameter, ϵ, of differential privacy is used to quantify the privacy risk posed by revealing statistics calculated on private and sensitive data. Though it has an intuitive theoretical explanation, choosing an appropriate value is non-trivial. We present a systematic and methodical way to calculate ϵ once the necessary constraints are given. In order to derive optimal values and an upper bound on epsilon, we use the confidence probability approach, Chebyshev's and McDiarmids inequalities.**

I. Introduction

Assigning the ideal and optimal value of ϵ is critical, as it allows us to strike the right balance between accuracy and privacy. Regrettably, there has been no clear and uniform indications in prior work as to the bounds or value for ϵ. [1] mentioned that selecting ϵ is a social issue and depends on different scenarios. They used random values like 0.1, ln2 etc. [2] used values running from 0.05 to 0.2. An economic method based on compensation was proposed by [3]. The individual contributing his/her data can expect compensation from the data analyst who uses the individual's information for performing experiments. The data analyst can't exceed the budget assigned to him. Though there have been studies related to the privacy-accuracy trade-off, evaluating the value for an optimalϵusing a mathematical and logical reasoning based approach is yet to be adopted. In this abstract, we make use of the probability distribution and standard deviation to come up with theorems to find the optimal value and upper bound of epsilon. For example, we make use of the confidence probability that a value lies between any two given points and for that bound we derive an optimal epsilon. Our approach offers a more intuitive way to calculate the optimal value ofϵbased on a myriad of parameters.

II. Our Contributions

We propose an approach of calculating the upper bound of ϵ given the standard deviation σ of the probability distribution $\mathcal{P}$, and two different approaches for calculating the optimal value of ϵ when different information, specifically, the probability density function $f(x)$, or the standard deviation σ of the probability distribution that $\mathcal{A}$ draws noise from are given.

III. Upper Bound of ϵ

In order to prove the upper bound of ϵ, we prove a lemma, Lemma 1, that will be used while proving Theorem 1.

Lemma 1. *For any ϵ or (ϵ,δ)-differential private mechanism $\mathcal{A}$ drawing noise from a probability distribution $\mathcal{P}$, the standard deviation, σ, of the probability distribution satisfies the relation,*

$$\sigma \propto \frac{1}{\epsilon}.$$

Theorem 1 (Upper bound of ϵ)**.** *Given an ϵ or (ϵ, δ)-differential private mechanism $\mathcal{A}$, where the random noise X is drawn from probability distribution $\mathcal{P}$ with standard deviation $\sigma = \frac{M}{\epsilon}$ for some real number M, and given a probability p then the following inequality holds:*

$$|\mathcal{A}(\mathcal{D}) - q(\mathcal{D})| \leq w \cdot q(\mathcal{D}),$$

where $1 > w > 0$, and

$$\epsilon \leq \frac{M}{w \cdot q(\mathcal{D})\sqrt{1-p}}.$$

The proof of Theorem 1 makes use of the well-known Chebyshev's inequality.

The upper bound of ϵ gives us the maximum theoretical privacy loss from applying an ϵ or (ϵ,δ)-differential private mechanism $\mathcal{A}$ using additive random noise X. We will calculate the privacy loss for Laplacian mechanism. For Laplacian mechanism, $\sigma = \frac{\Delta\sqrt{2}}{\epsilon}$, thus

$$\epsilon \leq \frac{\Delta\sqrt{2}}{w \cdot q(\mathcal{D})\sqrt{1-p}}$$

IV. Optimal ϵ

In this section, we present 2 approaches (Theorem 2 and Theorem 3) to find the optimal ϵ given the probability density function $f(x)$ or standard deviation σ of the probability distribution $\mathcal{P}$. The 'optimal' epsilon can be defined as the epsilon which bounds the noisy query output within a distance of c and with a probability p from the original output.

Theorem 2 (Optimal ϵ based on $f(x)$)**.** *Given an ϵ or (ϵ, δ)-differential private mechanism $\mathcal{A}$, where the random noise X*

is drawn from probability distribution $\mathcal{P}$ with probability density function $f(x)$, a real positive number c, and probability p such that

$$Pr\ [|\mathcal{A}(\mathcal{D}) - q(\mathcal{D})| < c] = p$$

Solving the following equation given a certain $f(x)$ yields an optimal ϵ:

$$p = \int_{-c}^{c} f(x)dx.$$

We use the idea of confidence probability in our proof.

Proof. We can interpret the probability p as the confidence probability. Hence,

$$\begin{aligned} p = &\ \Pr[|\mathcal{A}(\mathcal{D}) - q(\mathcal{D})| \le c] \\ = &\ \Pr[|X| \le c] \\ = &\ \Pr[-c < X < c] \\ = &\int_{-c}^{c} f(x)dx, \end{aligned}$$

and we can solve for ϵ once we know $f(x)$. □

For Laplacian mechanism, we know that

$$f(x) = \frac{1}{2\lambda}e^{-\frac{|x|}{\lambda}},$$

where $\lambda = \frac{\Delta}{\epsilon}$. Thus,

$$\begin{aligned} p &= \int_{-c}^{c} \frac{1}{2\lambda}e^{-\frac{|x|}{\lambda}} dx \\ &= 1 - e^{-\frac{c\epsilon}{\Delta}}. \end{aligned}$$

Solving for ϵ, we have

$$\epsilon = -\frac{\Delta \ln(1-p)}{c}.$$

Theorem 3 presents another way of calculating the optimal ϵ value.

Theorem 3 (Optimal ϵ based on σ)**.** *Given an ϵ or (ϵ, δ)-differential private mechanism $\mathcal{A}$, where the random noise X is drawn from probability distribution $\mathcal{P}$ with standard deviation $\sigma = \frac{M}{\epsilon}$ for some real number M and mean $\mu = 0$, a real positive number c, and the probability p such that*

$$Pr(|\mathcal{A}(\mathcal{D}) - q(\mathcal{D})| < c) > p,$$

then the optimal value of ϵ corresponds to

$$\epsilon = \frac{|M|}{c}\sqrt{\ln 2 - \ln(1-p)}.$$

The proof of Theorem 3 uses the McDiarmid's inequality, which is a result of AzumaHoeffding inequality.

This inequality provides another ideal ϵ for any ϵ or (ϵ,δ)-differential private mechanism $\mathcal{A}$ using additive random noise X. We will calculate the ideal ϵ for the Laplacian mechanism.

For the Laplacian mechanism, $\sigma = \frac{\Delta\sqrt{2}}{\epsilon}$, thus,

$$\epsilon = \frac{\Delta}{c}\sqrt{2(\ln 2 - \ln(1-p))}$$

Though the two approaches rely on knowing different information about the probability distribution $\mathcal{P}$, specifically, one is based on the probability density function $f(x)$ and the other is based on the standard deviation σ of $\mathcal{P}$, there are many instances in which we will be given all the information that we need about the probability distribution. In this case, it is natural to ponder whichϵwill have better performance. We solve this problem by taking the ratio of ϵs from Theorem 2 and Theorem 3. Thus

$$\begin{aligned} \epsilon_2 &= -\frac{\Delta \ln(1-p)}{c} \\ \epsilon_3 &= \frac{\Delta}{c}\sqrt{2(\ln 2 - \ln(1-p))} \end{aligned}$$

$$\begin{aligned} \frac{\epsilon_2}{\epsilon_3} &= \frac{-\frac{\Delta \ln(1-p)}{c}}{\frac{\Delta}{c}\sqrt{2(\ln 2 - \ln(1-p))}} \\ &= -\frac{\ln(1-p)}{\sqrt{2(\ln 2 - \ln(1-p))}} \end{aligned}$$

Fig. 1 is the graph of the ratio with respect to probability parameter p. After setting the ratio to be equal to 1, we can then calculate the threshold probability which determines the performance of the optimal ϵs calculated from the two different approaches.

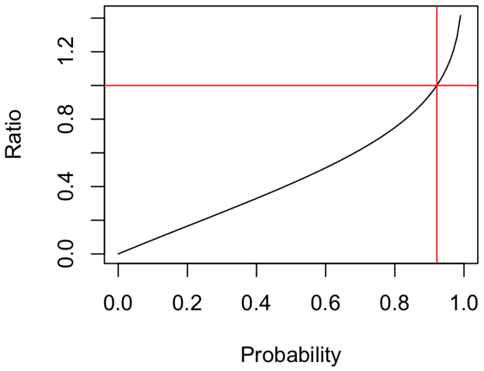

Fig. 1. Ratio: Laplacian Distribution

V. Conclusion and Future Work

We have presented different theorems to calculate ϵ and would like to explore other intuitive and mathematically-supported ways to derive ϵ.

References

[1] Cynthia Dwork. A firm foundation for private data analysis. *Communications of the ACM*, 54(1):86–95, 2011.

[2] Anand Sarwate, Claire Monteleoni, and Kamalika Chaudhuri. Differentially private support vector machines. Technical report, 2009.

[3] Justin Hsu, Marco Gaboardi, Andreas Haeberlen, Sanjeev Khanna, Arjun Narayan, Benjamin C Pierce, and Aaron Roth. Differential privacy: An economic method for choosing epsilon. In *2014 IEEE 27th Computer Security Foundations Symposium*, pages 398–410. IEEE, 2014.