On the (Im)Possibility of Estimating Various Notions of Differential Privacy (short paper)

Daniele Gorla[1], Louis Jalouzot[2], Federica Granese[1,3], Catuscia Palamidessi[3], and Pablo Piantanida[4]

[1] Dept. of Computer Science, Sapienza University of Rome
[2] ENS de Lyon (France)
[3] INRIA Saclay and LIX
[4] L2S, CentraleSupélec (France) and CNRS, Université Paris Saclay (France)

Abstract. We analyze to what extent final users can infer information about the level of protection of their data when the data obfuscation mechanism is a priori unknown to them (the so-called "black-box" scenario). In particular, we delve into the investigation of two notions of *local differential privacy* (LDP), namely ε-LDP and Rényi LDP. On one hand, we prove that, without any assumption on the underlying distributions, it is not possible to have an algorithm able to infer the level of data protection with provable guarantees. On the other hand, we demonstrate that, under reasonable assumptions (namely, Lipschitzness of the involved densities on a closed interval), such guarantees exist and can be achieved by a simple histogram-based estimator. We validate our results experimentally and we note that, on a particularly well-behaved distribution (namely, the Laplace noise), our method gives even better results than expected.

1 General setting

Differential privacy (DP) [5] is nowadays one of the best established and theoretically most solid tools to ensure data protection. Intuitively, given a set of databases, differential privacy requires that databases that only slightly differ one from the other (e.g. in one individual record) are mapped to the same obfuscated value with similar probabilities. The success of this privacy notion is witnessed by its wide application, both in academia and in industry (see e.g. [3, 6, 9, 14, 16]).

The first formulation of DP that we are going to consider is a distributed version of DP, called *local differential privacy* (LDP) [1]. Here, we do not work anymore with (adjacent) databases but directly on values from a set $\mathcal{X}$. In this setting, a randomized mechanism $\mathcal{K} : \mathcal{X} \to \mathcal{D}\mathcal{Z}$ (where $\mathcal{D}\mathcal{Z}$ denotes the set of probability distributions over $\mathcal{Z}$) is said to be *ϵ-LDP* if, for every $z \in \mathcal{Z}$ and $x_1, x_2 \in \mathcal{X}$, we have that

$$p_{Z|X}(z|x_1) \leq e^{\epsilon}\, p_{Z|X}(z|x_2) \tag{1}$$

where the notation $p_{Z|X}(\cdot|x_i)$ denotes either the density distribution of $\mathcal{K}(x_i)$ in the continuous case or the probability distribution $[z \mapsto \Pr(\mathcal{K}(x_i) = z)]$ in the discrete one. Here, ϵ controls the level of privacy: the smaller ϵ, the higher the level of privacy.

A second variant is the so-called *Rényi differential privacy* (RDP) [15], a relaxation of DP based on the notion of Rényi divergence. More formally, a randomized mechanism $\mathcal{K}$ is ϵ-*LRDP of order* $\alpha > 1$ if, for every x_1, x_2, we have that

$$D_\alpha(\mathcal{K}(x_1) \| \mathcal{K}(x_2)) \leq \epsilon \tag{2}$$

where $D_\alpha(\cdot \| \cdot)$ denotes Rényi divergence [17]. Also here the value of ϵ controls the level of privacy, in the sense that a smaller ϵ corresponds to a higher privacy.

However, both notions of DP are built into existing software products by the producing companies, and the final users have no way of testing the real level of security (i.e., the real value of ϵ). They can only trust the producers, sometimes leading to unexpected (and unwanted) behaviors. For this reason, we would like to study to what extent final users can infer information about the level of protection of their data when the data obfuscation mechanism is a priori unknown to them, and they can only sample from it (the so-called "black-box" scenario). A few black-box approaches to related problems have been presented in the literature:

- [4] that, given an oracle who has access to the probability density functions on the outputs, casts the problem of testing differential privacy on *typical datasets* (i.e., datasets with sufficiently high probability mass under a fixed data generating distribution) as a problem of testing the Lipschitz condition. Their result concerns variants of differential privacy called *probabilistic DP* and *approximate DP*.
- [10] prove both an impossibility result for DP and a possibility result for approximate DP. Their impossibility result shows that, for any $\epsilon > 0$ and proximity parameter $\alpha > 0$, no privacy property tester with finite query complexity exists for DP. Moreover, they achieve their possibility result using randomized algorithms.
- [12, 13], where the authors focus on (ϵ, δ)-DP and the estimation of the DP parameters of a given (unknown) mechanism. In [12] the authors aim to estimate the parameters for a fixed pair of adjacent databases by focusing on the relation between the number of samples required and the accuracy of the estimation, whereas they estimate the parameters of the mechanism by repeating their estimation on every possible pair. In [13] the authors aim at estimating, once ϵ is given, the δ of a certain (unknown) mechanism by focusing on *polynomial-time* approximate estimators on a given subset $\mathcal{T}$ of all the possible databases (thus defining and estimating the notion of *relative DP*).
- The paper that is most closely related to ours is [2], but there are some differences. First, they only consider central DP, whereas we focus on LDP and LRDP. Second, we both consider a pair of databases/values and evaluate the ϵ for this pair; to compute a better under-approximation of the overall ϵ, they iterate their method over a (somehow chosen) finite set of pairs without provable guarantees, whereas our method comes equipped with formal guarantees. Third, we use histograms and rely on the Lipschitzness of the noise function, whereas they use kernel density estimation and rely on Holder continuity (a generalization of Lipschitzness). Fourth, we upper bound the number of samples n needed to achieve a certain precision and confidence in the estimation of ϵ, whereas their theorem states that the estimation approximates ϵ asymptotically (within a certain confidence range) as n grows.

2 Our contributions

We first focus on (1) and try to estimate the (equivalent) quantity

$$\epsilon^\star(x_1, x_2) \stackrel{\text{def}}{=} \sup_{z \in \mathcal{Z}(x_1,x_2)} \log\left(\frac{p_{Z|X}(z|x_1)}{p_{Z|X}(z|x_2)}\right) \tag{3}$$

where $\mathcal{Z}(x_1, x_2) \stackrel{\text{def}}{=} \{z \in \mathcal{Z} \mid p_{Z|X}(z|x_1) > 0 \wedge p_{Z|X}(z|x_2) > 0\}$. In this setting, an estimator $\tilde{\epsilon}$ is an algorithm that takes in input a pair (x_1, x_2), a precision γ and a confidence δ (with $\gamma > 0$ and $0 < \delta < 1$), and returns a real that is supposed to approximate $\epsilon^\star(x_1, x_2)$ for at most the precision with probability at least the confidence. Our first main result states that such an estimator cannot exist:

Theorem 1 (Impossibility). *For every $\gamma > 0$ (precision), $0 < \delta < 1$ (confidence), $x_1, x_2 \in \mathcal{X}$, and probabilistic estimator algorithm $\tilde{\epsilon}$ that almost surely terminates, there exists a probability distribution $p_{Z|X}$ such that*

$$\Pr\left(\left|\epsilon^\star(x_1, x_2) - \tilde{\epsilon}(x_1, x_2, \gamma, \delta)\right| > \gamma\right) > 1 - \delta.$$

We remark that this impossibility result is very strong: it shows that no estimator exists, even if (1) we are not very demanding about the precision and the confidence (namely, even if γ is large and δ is small), (2) even if the number of samples is unbounded and (3) the estimator is adaptive (namely, it can decide on the fly whether to stop or to continue sampling, based on previous samples).

By contrast, if we confine ourselves to the continuous case and assume that the densities $p_{Z|X}(\cdot|x_1)$ and $p_{Z|X}(\cdot|x_2)$ over $\mathcal{Z} = [a, b]$ are C-Lipschitz with $C < \frac{2}{(b-a)^2}$, then a probabilistic histogram-based estimator exists, whose pseudocode is provided in Algorithm 1. For the desired precision γ, the estimator first divides $\mathcal{Z}$ into m sub-intervals, each of width $w \stackrel{\text{def}}{=} \frac{b-a}{m}$, where

$$m \stackrel{\text{def}}{=} \left\lceil \frac{6C(b-a)}{\tau\gamma} \right\rceil \qquad \text{and} \qquad \tau \stackrel{\text{def}}{=} \frac{1}{b-a} - \frac{C(b-a)}{2}. \tag{4}$$

In particular, we set $z_0 = a$ and $z_{j+1} = z_j + w$; one can readily check that $z_m = b$. Then, the estimator chooses n (the number of samples) such that

$$2m(1 - w\tau)^n + 4f(n, w\tau, \gamma/12) \leq 1 - \delta, \tag{5}$$

where f is defined as

$$f(x, y, z) \stackrel{\text{def}}{=} \frac{\exp\left(\frac{-xy(e^z-1)^2}{1+e^z}\right) + \exp\left(\frac{-xy(1-e^{-z})^2}{2}\right)}{1 - (1-y)^x} \tag{6}$$

(note that f is exponentially decreasing in x and y). The estimator then invokes the sampler n times both for x_1 and for x_2 (lines 4-7), counts the number of samples that appear in each sub-interval (lines 8-14), and considers these numbers as the approximations of $p_{Z|X}(\cdot|x_1)$ and $p_{Z|X}(\cdot|x_2)$ in that sub-interval; so, it computes their ratio and returns the highest value. The fact that this algorithm has provable guarantees is the second main result of our paper.

Algorithm 1 Histogram-based estimator for $\epsilon^\star(x_1, x_2)$

Input: $\mathcal{Z}(= [a, b]), \gamma, \delta, C$
Output: $\tilde{\epsilon}(x_1, x_2, \mathcal{Z}, \gamma, \delta, C)$ ▷ differing from $\epsilon^\star(x_1, x_2)$ for $\leq \gamma$ with prob. $\geq \delta$
Compute m and n as in eq. (4) and eq. (5), resp.
for $1 \leq i \leq n$ **do**
    $s_1[i] \leftarrow \mathcal{S}(x_1)$
    $s_2[i] \leftarrow \mathcal{S}(x_2)$
end for
for $1 \leq j \leq m$ **do**
    $N_j \leftarrow \sum_i \mathbb{1}\left(z_j \leq s_1[i] < z_{j+1}\right)$
    $M_j \leftarrow \sum_i \mathbb{1}\left(z_j \leq s_2[i] < z_{j+1}\right)$
    if $N_j = 0$ or $M_j = 0$ **then**
        fail
    end if
end for
return $\max_j \log\left(\frac{N_j}{M_j}\right)$

Theorem 2 (Correctness). *Let densities $p_{Z|X}(\cdot|x_1)$ and $p_{Z|X}(\cdot|x_2)$ over $\mathcal{Z} = [a, b]$ be C-Lipschitz, with $C < \frac{2}{(b-a)^2}$. For every $\gamma > 0$ (precision) and $0 < \delta < 1$ (confidence):*

$$\Pr\Big(\textit{Algorithm 1 succeeds and } |\epsilon^\star(x_1, x_2) - \tilde{\epsilon}(x_1, x_2, \mathcal{Z}, \gamma, \delta, C)| \leq \gamma\Big) \geq \delta.$$

Once we have this estimator for a single pair of values, we then aim at estimating the overall ϵ, i.e.

$$\epsilon^\star(p_{Z|X}) \stackrel{\text{def}}{=} \sup_{x_1, x_2 \in \mathcal{X}} \epsilon^\star(x_1, x_2). \tag{7}$$

To this aim, we assume $\mathcal{X}$ to be a closed interval as well, divide it in k buckets (for a proper k), take the mid-points of all the buckets, run the previous estimator for all pairs of mid-points, and return the maximum. The details are given in Algorithm 2. If we also assume $p_{Z|X}(z|\cdot)$ to be D-Lipschitz, for some D and for all $z \in \mathcal{Z}$ (so any doubly differentiable function satisfies this requirement), this new algorithm is able to estimate the overall ϵ with provable guarantees as established in the following result, where we say that the algorithm *succeeds* if at least one invocation of Algorithm 1 succeeds. This is our third main result.

Theorem 3. *Let $\mathcal{Z} = [a, b]$, $\mathcal{X} = [c, d]$, and $p_{Z|X}$ be such that, for every $x \in \mathcal{X}$, $p_{Z|X}(\cdot|x)$ is C-Lipschitz, for $C < 2/(b-a)^2$, and that, for every $z \in \mathcal{Z}$, $p_{Z|X}(z|\cdot)$ is D-Lipschitz, for some D. For every $\gamma > 0$ (precision) and $0 < \delta < 1$ (confidence), we have that*

$$\Pr\Big(\textit{Algorithm 2 succeeds and } |\epsilon^\star(p_{Z|X}) - \tilde{\epsilon}(\mathcal{Z}, \mathcal{X}, \gamma, \delta, C, D)| \leq \gamma\Big) \geq \delta.$$

We note that the Lipschitzness assumptions required by our theorems are met by the two most widely used DP mechanisms, namely Laplacian and Gaussian [7, 8]. Then,

Algorithm 2 Estimator for $\epsilon^\star(p_{Z|X})$

1: **Input:** $\mathcal{Z}(=[a,b]), \mathcal{X}(=[c,d]), \gamma, \delta, C, D$
2: **Output:** $\tilde{\epsilon}(\mathcal{Z}, \mathcal{X}, \gamma, \delta, C, D)$
3: Let $k \geq \frac{3D(d-c)}{\tau\gamma}$, where τ is defined in eq. (4)
4: Divide $\mathcal{X}$ in k buckets, with x_i the mid-point of bucket i
5: **for all** $\{x_i, x_j\} \subseteq \{1, \ldots, k\}$ **do**
6: $\quad \tilde{\epsilon}_{ij} \leftarrow \tilde{\epsilon}(x_i, x_j, \mathcal{Z}, \frac{\gamma}{3}, \sqrt{\delta}, C)$, by invoking Alg.1
7: **end for**
8: **return** $\max_{ij} \tilde{\epsilon}_{ij}$

we validate all our results for the Laplace distribution. We first consider the number of samples the estimator does; this parameter depends on γ, δ, C and $|\mathcal{Z}|$, and we discover that the strongest dependency is on γ. Then, we compare the estimated ϵ against the real one and we discover that the number of samples required to have satisfactory results in practice is significantly lower than the theoretical one (i.e., that of (5)). Furthermore, we study the proportion of estimated ϵ that are close to ϵ within γ across 100 executions for different values of the number of samples. We discover that the lowest number of samples that yields a proportion greater than δ is around 400 times lower than the theoretical one in this case.

Finally, our last bunch of results is on LRDP (see eq. (2)), for which we mimic the steps outlined above, with similar outcomes. In this setting, the impossibility result is more surprising: indeed, if there is some output where the probabilities differ significantly but the probability of this output is low, then one would think that this would not violate the RDP guarantee since Rényi divergence averages over all outputs, instead of taking the pointwise maximum. However, we formally prove that this is not the case. Then, we adapt the two estimators by requiring more complex bounds both on the number of experiments and on the number of intervals required. In particular, for the estimator $\tilde{\epsilon}_\alpha(x_1, x_2, \mathcal{Z}, \gamma, \delta, C)$ (see Algorithm 1), the number m of sub-intervals and the number n of samples are such that

$$\frac{CK(b-a)(2\alpha-1)}{2m\tau_0 K'(\alpha-1)} \leq \frac{\gamma}{2} \qquad 1 - 2m(1-w\tau_0)^n - 2mf(n, w\tau_0, \gamma') \geq \delta$$

where f is defined in eq. (6) and $\tau_0 \stackrel{\text{def}}{=} \frac{1}{b-a} - \frac{C(b-a)}{2}$, $\tau_1 \stackrel{\text{def}}{=} \frac{1}{b-a} + \frac{C(b-a)}{2}$, $K \stackrel{\text{def}}{=} \frac{2\tau_1^\alpha}{\tau_0^{\alpha-1}}$, $K' \stackrel{\text{def}}{=} \frac{\tau_0^\alpha}{\tau_1^{\alpha-1}}$, and $\gamma' = \min\left(\frac{\gamma K'(\alpha-1)}{2K(2\alpha-1)}, \frac{\log 2}{2\alpha-1}\right)$. The returned value is $\frac{1}{\alpha-1}\log\sum_j \frac{1}{n}\left(\frac{N_j}{M_j}\right)^\alpha M_j$. For the estimator $\tilde{\epsilon}_\alpha(\mathcal{Z}, \mathcal{X}, \gamma, \delta, C, D)$ (see Algorithm 2), the new number of buckets is $k \geq \frac{3(2\alpha-1)KD(d-c)}{2(\alpha-1)K'\tau_0\gamma}$ and, of course, we invoke the estimator $\tilde{\epsilon}_\alpha(x_1, x_2, \mathcal{Z}, \gamma, \delta, C)$ modified as described above for LRDP.

We run experiments similar to the ones for LDP that confirm the quality of our approach also for LRDP. In particular, for this second setting the gap between the number of samples sufficient for achieving the guarantees of the theorem and the theoretical one is even more dramatic than for LDP: here the practical one is around 10^5 times smaller.

For all details, we refer the reader to [11].

References

1. M. S. Alvim, K. Chatzikokolakis, C. Palamidessi, and A. Pazii. Local differential privacy on metric spaces: Optimizing the trade-off with utility. In *31st Computer Security Foundations Symposium*, pages 262–267. IEEE, 2018.
2. Ö. Askin, T. Kutta, and H. Dette. Statistical quantification of differential privacy: A local approach. In *43rd IEEE Symposium on Security and Privacy*, pages 402–421. IEEE, 2022.
3. Differential Privacy Team (Apple Inc.). Learning with privacy at scale. 2017. `https://machinelearning.apple.com/research/learning-with-privacy-at-scale`.
4. K. Dixit, M. Jha, S. Raskhodnikova, and A. Thakurta. Testing the lipschitz property over product distributions with applications to data privacy. In *10th Theory of Cryptography Conference (TCC)*, volume 7785 of *LNCS*, pages 418–436. Springer, 2013.
5. C. Dwork. Differential privacy. In *33rd International Colloquium on Automata, Languages and Programming (ICALP)*, volume 4052 of *LNCS*, pages 1–12. Springer, 2006.
6. C. Dwork. Differential privacy: A survey of results. In *5th International Conference on Theory and Applications of Models of Computation (TAMC)*, volume 4978 of *LNCS*, pages 1–19. Springer, 2008.
7. C. Dwork, F. McSherry, K. Nissim, and A. D. Smith. Calibrating noise to sensitivity in private data analysis. In S. Halevi and T. Rabin, editors, *Theory of Cryptography, Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006, Proceedings*, volume 3876 of *Lecture Notes in Computer Science*, pages 265–284. Springer, 2006.
8. C. Dwork and A. Roth. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3-4):211–407, 2014.
9. 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 (CCS)*, pages 1054–1067. ACM, 2014.
10. A. C. Gilbert and A. McMillan. Property testing for differential privacy. In *56th Annual Allerton Conference on Communication, Control, and Computing*, pages 249–258. IEEE, 2018.
11. D. Gorla, L. Jalouzot, F. Granese, C. Palamidessi, and P. Piantanida. On the (im)possibility of estimating various notions of differential privacy. *CoRR*, abs/2208.14414, 2022.
12. X. Liu and S. Oh. Minimax optimal estimation of approximate differential privacy on neighboring databases. In *Proc. of Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019.
13. Y. Lu, Y. Wei, M. Magdon-Ismail, and V. Zikas. Eureka: A general framework for black-box differential privacy estimators. *IACR Cryptol. ePrint Arch.*, page 1250, 2022.
14. A. Machanavajjhala, D. Kifer, J. Abowd, J. Gehrke, and L. Vilhuber. Privacy: Theory meets practice on the map. In *24th International Conference on Data Engineering*, pages 277–286. IEEE, 2008.
15. I. Mironov. Rényi differential privacy. In *30th Computer Security Foundations Symposium*, pages 263–275. IEEE, 2017.
16. A. Narayanan and V. Shmatikov. De-anonymizing social networks. In *30th Symposium on Security and Privacy*, pages 173–187. IEEE, 2009.
17. A. Rényi. On measures of entropy and information. In *4th Berkeley symposium on mathematical statistics and probability*, volume 1, pages 547–561, 1961.