Differential Privacy

Monica Rico
mrico@wisc.edu

Lucas Steplyk
lsteplyk@wisc.edu

1. Introduction

The increasing availability and use of data have enabled significant advances across domains including in medical research, social media, and the financial sector [1, 3, 6]. Data science, in particular, is incredibly reliant on large-scale data to enable effective modeling, decision-making and discovery. As researchers and companies try to leverage the advances in data science it often requires the release and coalescing of databases and statistics. When statistics are released to the public, even in good faith, they present privacy concerns. So as a natural first step companies often redact "sensitive information" or Personally Identifiable Information (PII). However, time and time again we have seen statistics that were thought to be successfully anonymized, through the removal of PII, were able to pinpoint individuals through coupling with other datasets. Whether it be through the combination of voting and medical records, or Netflix rentals and IMDB ratings [9]. This issue is further exacerbated by data leaks, where data that was never intended for public viewing can then be used to coupled with public data to isolate individuals. Differential privacy seeks to guarantee that the ability of an adversary to learn information about an individual is *almost* the same, independent of the individual's participation status in the dataset [13]. In the following sections, we give an overview of differential privacy, different model types and current challenges.

2 The Problem: Defining Differential Privacy

2.1 The path to Differential Privacy; *k*-Anonymity

One approach to privacy-preserving data release is k-anonymity, which seeks to ensure that each individual's record is indistinguishable from at least k-1 others based on quasi-identifiers. The goal is to be able to compute and release statistics on a database in such a way that the results do not depend *too* much on any single individual.

> "Sanitization is performed by applying techniques that preserve the truthfulness of the information of each data item. Such techniques include, for example, generalization (publishing more general values for the data) and suppression (removing some data)." [12]

In k-anonymity, the chance of linking a record back to an individual is at most is $\frac{1}{k}$. While k-anonymity is simple to implement and intuitive to understand, it has severe limitations. K-anonymity fails to protect individuals against attribute disclosure, especially in cases of undiversified data or undiversified groups within a dataset (e.g., all group members share the same sensitive attribute).

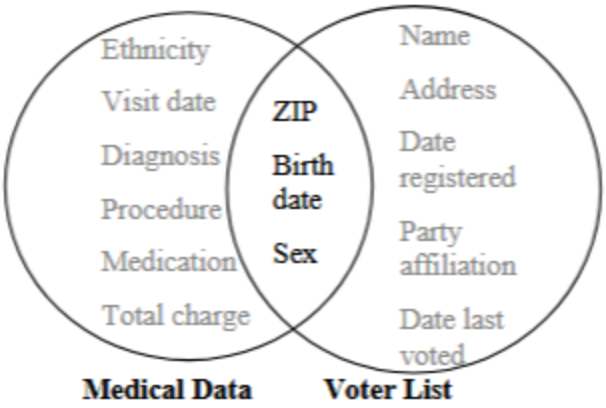

Figure 1: How a linkage attack may look

When those cases arise, even when sanitization through grouping has been achieved one can still gather information about individuals. Figure 1 illustrates a linkage attack, in which sanitized datasets can be linked with external sources to re-identify individuals. When groups lack diversity in sensitive attributes or when auxiliary datasets are available, the anonymity intended by grouping can break down. This linkage reveals more information about specific individuals than either dataset alone would have provided [13]. This motivates the need for stronger, formal privacy guarantees - specifically, differential privacy.

2.2 Defining Differential Privacy

Differential privacy provides a formal privacy guarantee by bounding the effect that any single individual's data can have on the output of a randomized algorithm. Let D and D' be two neighboring datasets that differ only by a single row (i.e., one individual's data). Let f be a randomized algorithm mapping datasets to some output space, and let S $\subseteq Range(f)$. Then f is said to be **(ϵ,δ)-differentially private** if:

$$P[f(D) \in S] \leq e^{\epsilon} \cdot P[f(D') \in S] + \delta$$

The parameter, ϵ, known as the privacy loss parameter, controls how much the output distributions can differ between neighboring datasets. The smaller ϵ the more private. The parameter, δ, is the failure probability which allows for a small chance that the bound does not hold. When $\delta = 0$, we say that the mechanism satisfies *pure* differential privacy. When $\delta > 0$, it provides *approximate* differential privacy.

2.3 Achieving Differential Privacy

Achieving differential privacy is brought upon by noise injection into our database through our function.

$$f(D) = g(D) + Noise$$

From the definition of differential privacy, it is not immediately obvious how a database could achieve the guarantee of differential privacy. To start we need to define the difference between the outputs of *f*.

$$\Delta f = max_{D,D'} ||f(D) - f(D')||_1$$

This is known as sensitivity [5]. The maximum here is pertinent as it provides our guarantees even in cases where extreme outliers were the ones omitted in the dataset. Sensitivity tells us what type of noise we can add to a dataset.

2.3.1 Laplace Distribution

One standard way to achieve *pure* differential privacy ($\delta = 0$) is the Laplace mechanism, which adds random noise from the Laplace distribution:

$$f(D) = g(D) + Laplace(\mu = 0, b)$$

We see why that is through a rewrite of our definition of differential privacy.

$\frac{P[f(D)=y]}{P[f(D')=y]} \leq e^{\epsilon}$ assuming pure differential privacy ($\delta = 0$)

Using the PDF of a Laplace distribution,

$$\frac{\frac{1}{2b}e^{-\frac{|y-f(D)|}{b}}}{\frac{1}{2b}e^{-\frac{|y-f(D')|}{b}}} = \frac{e^{-\frac{|y-f(D)|}{b}}}{e^{-\frac{|y-f(D')|}{b}}} = e^{-\frac{|y-f(D)|-|y-f(D')|}{b}} \leq e^{\frac{|f(D)-f(D')|}{b}} \leq e^{\frac{\Delta f}{b}}$$

Where we need

$$e^{\frac{\Delta f}{b}} \leq e^{\epsilon} \Rightarrow b \geq \frac{\Delta f}{\epsilon}$$

Therefore by choosing our noise with $Laplace(\mu = 0, b = \frac{\Delta f}{\epsilon})$ we can guarantee *pure* differential privacy.

2.3.2 Gaussian Distribution

The Gaussian mechanism provides approximate differential privacy by adding from a normal/Gaussian distribution.

$$f(D) = g(D) + N(0, \sigma^2)$$

Roughly similar to the how we came to the Laplace distribution choose

$$\sigma \geq \frac{\Delta_2 f \sqrt{2ln(\frac{1.25}{\delta})}}{\epsilon}$$

And have to define our sensitivity based on the ℓ_2-norm, that is

$$\Delta_2 f = max_{D,D'} ||f(D) - f(D')||_2$$

However, because the Gaussian distribution has tails that do not decay as fast as the Laplace distribution we need to set a delta value, with Gaussian you get cases where it will inject a very large amount of noise into our result of *f*. Counterintuitively, large amounts of noise do not always correlate with stronger privacy. Again our privacy is based on the addition or omission of

a single individual. So in some cases of a very noisy result that result could be more likely with the addition/omission of single and would break pure differential privacy. To compensate we must add in a delta term.

An example of this in practice would be the US Census Bureau's approach for the decennial census data [7]. They have chosen to implement a Zero-Concentrated Differential Privacy (zCDP) based on a discrete Gaussian distribution instead of a Laplace mechanism. This shift means the same level of privacy-loss budget while zCDP has lower probability of injecting large amounts of noise than the pure differential privacy case.

3 Drawbacks

As data-driven methodologies continue to expand across domains, the demand for strong privacy protections grows in parallel. Differential privacy does offer rigorous guarantees, but its implementation still faces challenges. For instance, neural networks trained using differentially private stochastic gradient descent (DP-SGD) resulted in a reduction in the model's accuracy. The effect was observed in different cases such as gender classification, sentiment analysis, species classification and federated learning of language models. Notably, applying differential privacy showed to have a disparate negative impact on model accuracy, mainly toward underrepresented subgroups and those with relatively complex data [2]. These findings raise concerns when it comes to fairness and equity, suggesting that privacy guarantees come at a cost of amplified bias in model outcomes.

Beyond fairness, a more fundamental limitation lies in the mathematical trade-off between privacy and utility. In order to achieve stronger privacy guarantees - by selecting a lower privacy loss parameter ϵ - more noise must be added to the data or model outputs. While this protects individuals' information, it comes at the cost of reduced accuracy in statistical learning, especially in high-dimensional settings or when dealing with sparse data. Duchi, Jordan and Wainwright (2013) provide a rigorous analysis of this trade-off under local differential privacy constraints [4]. They establish minimax lower bounds that quantify how statistical utility degrades as privacy requirements become more stringent. The trade-off here raises the question of *how* to set the parameters, ϵ and δ, in practice.

In theoretical work, the parameters ϵ and δ are seen as tunable constants but there is little consensus on what appropriate values are in practice. Choosing values that are too large may offer weak protection, while overly strict settings can render the data unusable, leaving practitioners to navigate this uncertainty with limited guidance.

4 Critical Analysis

While there have been significant theoretical and practical contributions in the differential privacy space, several important themes emerge when the work is viewed from afar. One notable distinction is between central and local DP. In the central model, noise is added *after* data

aggregation, leading to higher data utility. This assumes a secure and trustworthy data collector, which is not feasible in many real-world settings. On the other hand, local DP eliminates the need for trust of a central source by adding noise at the source before it is shared. However, this stricter setting has shown losses in accuracy, particularly for high-dimensional or sparse data.

To bridge the gap between theoretical guarantees and practical utility, researchers have explored relaxations and extensions of differential privacy. These models include individual DP, the Renyi DP, zero-concentrated DP, and approximate DP [8, 11]. The extensions provide tighter composability and require less noise while maintaining formal privacy guarantees. These variants rely on mathematical assumptions and are less interpretable for those without a theoretical background.

Even when trying to achieve theoretical guarantees, there are practical limitations around scalability, cost and feasibility. These limitations point to a need for models that are both robust to practical constraints and sensitive to privacy concerns.

5 Conclusion

Differential privacy has emerged as our world has become more reliant on data driven systems. The core concept - ensuring that the inclusion or exclusion of any single individual's data not significantly affecting the outcome of an analysis - provides a mathematically rigorous and widely applicable framework for privacy preservation. Differential privacy offers strong theoretical guarantees, clear quantification of privacy addition/loss, and adaptable mechanisms to a range of applications. However, there are still challenges to overcome. Practical deployments of differential privacy struggle with degraded model performance, fairness concerns, and difficulty in parameter selection. Continued research is essential to improve the usability, interpretability and scalability of differential privacy mechanisms. As data continues to be essential in every aspect of society, methodologies like differential privacy will need to be effective and accessible for all.

References

[1] Abkenar, S., Kashani, M., Mahdipour, E., & Jameii, S. (2021). “Big data analytics meets social media: A systematic review of techniques, open issues, and future directions.” *Telematics and Informatics, 57*, 101517. https://doi.org/10.1016/j.tele.2020.101517

[2] Bagdasaryan, Eugene, Omid Poursaeed, and Vitaly Shmatikov. 2019. “Differential Privacy Has Disparate Impact on Model Accuracy.” Advances in Neural Information Processing Systems. Vol. 32. Curran Associates, Inc. https://doi.org/10.48550/arXiv.1905.12101

[3] Batko, K., & Ślęzak, A. (2022). “The Use of Big Data Analytics in Healthcare.” *Journal of big data*, *9*(1), 3. https://doi.org/10.1186/s40537-021-00553-4

[4] Duchi,J. C., Jordan, M. I. and Wainwright, M. J., "Local Privacy and Statistical Minimax Rates," *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, Berkeley, CA, USA, 2013, pp. 429-438, doi: 10.1109/FOCS.2013.53.

[4] Dwork, C. (2006). “Differential Privacy.” In: Bugliesi, M., Preneel, B., Sassone, V., Wegener, I. (eds) Automata, Languages and Programming. ICALP 2006. Lecture Notes in Computer Science, vol 4052. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11787006_1

[5] Dwork, C., McSherry, F., Nissim, K., Smith, A. (2006). “Calibrating Noise to Sensitivity in Private Data Analysis.” In: Halevi, S., Rabin, T. (eds) Theory of Cryptography. TCC 2006. Lecture Notes in Computer Science, vol 3876. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11681878_14

[6] Hasan, M.M., Popp, J. & Oláh, J. “Current landscape and influence of big data on finance.” *J Big Data* 7, 21 (2020). https://doi.org/10.1186/s40537-020-00291-z

[7] U.S. Census Bureau. “Disclosure Avoidance for the 2020 Census: An Introduction”., U.S. Government Publishing Office. Washington, DC, November 2021. https://www2.census.gov/library/publications/decennial/2020/2020-census-disclosure-avoidance-handbook.pdf

[8] J. Soria-Comas, J. Domingo-Ferrer, D. Sánchez and D. Megías, "Individual Differential Privacy: A Utility-Preserving Formulation of Differential Privacy Guarantees," in IEEE Transactions on Information Forensics and Security, vol. 12, no. 6, pp. 1418-1429, June 2017, doi: 10.1109/TIFS.2017.2663337. https://ieeexplore-ieee-org.ezproxy.library.wisc.edu/document/8049725

[9] Kearns, M., & Roth, A. (2021). The Ethical Algorithm: The science of socially aware algorithm design. Algorithmic Privacy, From Anonymity to Noise, 22-56.

[10] Larry Wasserman & Shuheng Zhou (2010) A Statistical Framework for Differential Privacy, Journal of the American Statistical Association, 105:489, 375-389, DOI: 10.1198/jasa.2009.tm08651.
https://www.tandfonline.com/doi/abs/10.1198/jasa.2009.tm08651

[11] Mironov, Ilya. 2017. “Rényi Differential Privacy.” In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, 263–75. https://doi.org/10.1109/CSF.2017.11.

[12] Sabrina De Capitani di Vimercati, Sara Foresti, Giovanni Livraga, Pierangela Samarati, “k-Anonymity: From Theory to Applications” Computer Science Department, Universit `a degli Studi di Milano, Italy. https://www.tdp.cat/issues21/tdp.a460a22.pdf

[13] Sweeney. k-anonymity: a model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10 (5), 2002; 557-570.
https://doi.org/10.1142/S0218488502001648

[14] Wright, Catherine, Rumsey, Kellin. “The Strengths, Weaknesses and Promise of Differential Privacy as a Privacy-Protection Framework.” DifferentialPrivacy.pdf