The Name of the Title Is Hope

Chih-Cheng Rex Yuan
hello@rexyuan.com
Institute of Information Science, Academia Sinica
Taipei, Taiwan

Bow-Yaw Wang
bywang@iis.sinica.edu.tw
Institute of Information Science, Academia Sinica
Taipei, Taiwan

Abstract

abstract

ACM Reference Format:
Chih-Cheng Rex Yuan and Bow-Yaw Wang. 2024. The Name of the Title Is Hope. In . ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

2 Related Work

3 Auditing Framework

3.1 Preliminaries

A *row* r_i is a lookup table or dictionary. A *database* $\mathcal{D} = \{r_1, r_2, ...\}$ is a collection of rows. The attributes of $\mathcal{D}$ is $\mathcal{A} = \{A_1, A_2, ...\}$. The domain of A_i is Ω_i.

Let $C \subseteq \mathcal{A}$. Let $\Omega_C = \Pi_{i \in C} \Omega_i$. The *marginal*$[1, 5]$ on C is a vector $\mu \in \mathbb{R}^{|\Omega_C|}$, indexed by domain element $t \in \Omega_C$, such that each entry is a count $\mu_t = \Sigma_{x \in \mathcal{D}} \mathbb{1}[x_C = t]$ where $\mathbb{1}$ is the indicator function. Let $M_C(\mathcal{D})$ be the function that computes the marginal on C, i.e., $\mu = M_C(\mathcal{D})$.

A *randomized mechanism* is a randomized algorithm M that takes a database $\mathcal{D}$ and, after, introducing noise, outputs some results in set R.

The p-norm is denoted by L_p and the p-norm of a vector x is denoted by $\|x\|_p$.

The normal distribution or Gaussian distribution with mean μ and standard deviation σ is denoted by $\mathcal{N}(\mu, \sigma^2)$.

The Kullback–Leibler divergence between probability distributions P and Q is denoted by $D_{KL}(P\|Q)$. The generalization of it, Rényi divergence[7], of order α is denoted by $D_\alpha(P\|Q)$.

3.2 Fairness Measures

3.3 Differential Privacy

Definition 3.1 (Sensitivity[3]). Let f be a function that takes a database $\mathcal{D}$ and outputs a vector $\mathbb{R}^p$. The L_2 sensitivity of f is for all databases $\mathcal{D}_1, \mathcal{D}_2$ that differ in exactly one row:

$$\Delta_f^2 = max_{\mathcal{D}_1, \mathcal{D}_2} \|f(\mathcal{D}_1) - f(\mathcal{D}_2)\|_p$$

Definition 3.2 (Gaussian Mechanism[3]). Let f be a function that takes a database $\mathcal{D}$ and outputs a vector $\mathbb{R}^p$. The Gaussian Mechanism M adds Gaussian noise with scale σ to each of the p outputs:

$$M(\mathcal{D}) = f(\mathcal{D}) + \mathcal{N}(0, \sigma^2 \mathbb{I})$$

Definition 3.3 (Differential Privacy (DP) [2, 3, 5]). A randomized mechanism M satisfies (ϵ, δ)-DP if, for all databases $\mathcal{D}_1, \mathcal{D}_2$ that differ in exactly one row and for all subsets S of R, we have

$$\Pr[M(\mathcal{D}_1) \in S] \leq e^\epsilon \Pr[M(\mathcal{D}_2) \in S] + \delta$$

Definition 3.4 (Rényi Differential Privacy (RDP)). A randomized mechanism M satisfies (α, γ)-RDP for $\alpha \geq 1$ and $\gamma \geq 1$ if, for all databases $\mathcal{D}_1, \mathcal{D}_2$ that differ in exactly one row, we have

$$D_\alpha(M(\mathcal{D}_1)\|M(\mathcal{D}_2)) \leq \gamma$$

Theorem 3.5 (RDP of the Gaussian Mechanism[4, 6]). *The Gaussian Mechanism satisfies* $(\alpha, \alpha \frac{\Delta_f^2}{2\sigma^2})$*-RDP.*

3.4 Synthetic Data

4 Methodology

4.1 Differential Private Synthetic Data

4.2 Fairness Checker

4.3 Implementation

5 Results

5.1 Adult Income Dataset

5.2 COMPAS Dataset

5.3 One More Dataset

6 Discussion

6.1 Accuracy

6.2 Impossibility

7 Conclusion

Some examples. A paginated journal article [**?**]

References

[1] Boaz Barak, Kamalika Chaudhuri, Cynthia Dwork, Satyen Kale, Frank McSherry, and Kunal Talwar. 2007. Privacy, accuracy, and consistency too: a holistic solution to contingency table release. In *Proceedings of the twenty-sixth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*. 273–282.

[2] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3*. Springer, 265–284.

[3] Cynthia Dwork, Aaron Roth, et al. 2014. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science* 9, 3–4 (2014), 211–407.

[4] Vitaly Feldman, Ilya Mironov, Kunal Talwar, and Abhradeep Thakurta. 2018. Privacy amplification by iteration. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 521–532.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

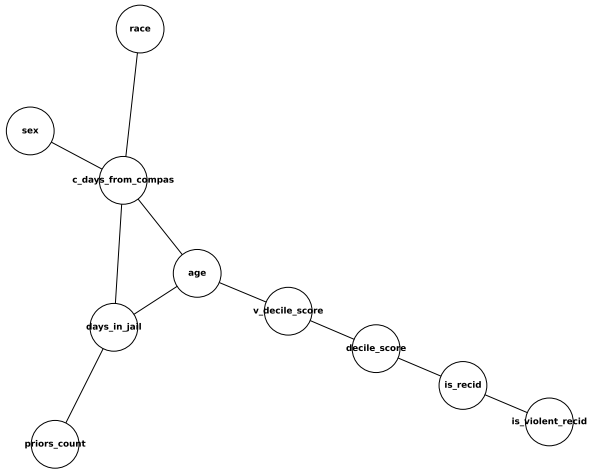

Figure 1: 1907 Franklin Model D roadster. Photograph by Harris & Ewing, Inc. [Public domain], via Wikimedia Commons. (https://goo.gl/VLCRBB).

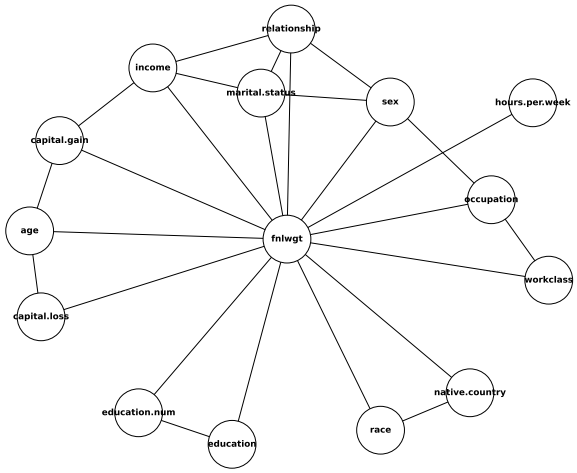

Figure 2: 1907 Franklin Model D roadster. Photograph by Harris & Ewing, Inc. [Public domain], via Wikimedia Commons. (https://goo.gl/VLCRBB).

[5] Ryan McKenna, Gerome Miklau, and Daniel Sheldon. 2021. Winning the NIST Contest: A scalable and general approach to differentially private synthetic data. *arXiv preprint arXiv:2108.04978* (2021).
[6] Ilya Mironov. 2017. Rényi differential privacy. In *2017 IEEE 30th computer security foundations symposium (CSF)*. IEEE, 263–275.
[7] Tim Van Erven and Peter Harremos. 2014. Rényi divergence and Kullback-Leibler divergence. *IEEE Transactions on Information Theory* 60, 7 (2014), 3797–3820.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009