Discussion: Measures of Empirical Privacy

PATCG | 15 June 23

Problem Statements

- **Comparability:** We don't have an aligned upon means for comparing
  - Privacy afforded by various approaches/proposals (e.g. IPA, Sandbox APIs)
  - Privacy mechanisms (Local DP, Central DP, thresholding, entropy)
- **Parameter Setting:** Difficult to credibly set parameters without an objective function
  - Parameters might include: epsilon, k, epoch, privacy unit, sensitivity
  - A measure of empirical privacy supports intelligent privacy-utility trade-offs
- **Understandability:** We need a way to talk about privacy
  - Within this group
  - With non-technical audiences (e.g. regulators, privacy advocates, etc)

Pros & Cons

Why Do This Work?

- Enable decision making (params, privacy units, etc)
- Enable innovation (encourage exploration of new approaches)
- Create alignment & understanding

Why Not?

- We don’t need it.
  - We can decide on techniques/params/units/budgets based on theoretical principles.
  - We should align on a utility bar, then maximize privacy instead
- Any approach requires assumptions which are difficult to make (e.g. prior information), so any method is inherently susceptible to criticism

Open questions

The work will likely require us to tackle:

- What are the threat models we need to consider?
  - Privacy from whom?
  - Against what prior knowledge?
- Is estimation based on the average case or worst case?
- How would we expect to use the score?
  - For decision making about methods?
  - Or is a score reported with any data release

These topics all remain out of scope:

- Any opinion or decision on the level of required privacy
- Any opinion on what privacy tools are acceptable
- Any opinion on how to measure the utility of results

Potential project plan

1. Determine scope of work and project goals
2. Align on requirements for a standardized measure of privacy
3. Literature review
4. Discussion of approaches to investigate
5. Investigation of several approaches
6. Create a written summary of learnings with recommended approach

Call for participants

(Actual project plan to be decided by sub-group members)

References

1. Carey, C. J., et al. "Measuring Re-identification Risk." *arXiv preprint arXiv:2304.07210* (2023).
2. Garfinkel, Simson. *De-identification of Personal Information:*. US Department of Commerce, National Institute of Standards and Technology, 2015.
3. Cormode, Graham, et al. "Empirical privacy and empirical utility of anonymized data." *2013 IEEE 29th International Conference on Data Engineering Workshops (ICDEW)*. IEEE, 2013.
4. Murakonda, Sasi Kumar, and Reza Shokri. "ML Privacy Meter: Aiding regulatory compliance by quantifying the privacy risks of machine learning." *arXiv preprint arXiv:2007.09339* (2020).
5. Institute of Medicine (US). *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk*. National Academies Press, 2015.
6. https://arxiv.org/pdf/2305.08846.pdf