Modernizing Disclosure Avoidance

Discussant Comments by

Andrew A. Samwick

Andrew.Samwick@Dartmouth.Edu

September 15, 2017

“Database reconstruction is a serious disclosure threat.”

- The Fundamental Law of Information Recovery: Overly accurate estimates of too many statistics can completely destroy privacy. (Abowd et al. 2017)
- So why publish “5.6 billion independent tabular summaries of a database of 308 million records” or allow similar quantities of (unmonitored) queries?
  - It cannot be that all of these tabulations are essential (or independent?) or of equal scientific merit.
  - How low would the 5.6 billion have to go before we wouldn’t have to worry about this Fundamental Law?

Empirical Disclosure Risk Assessment

- “Because the information environment is changing much faster than when traditional SDL techniques were developed, it may no longer be reasonable to assert that a product is empirically safe given best-practice disclosure limitation prior to its release.”
- Why equivocate? The Census Bureau must be doing its own simulated attacks to reconstruct its databases. Part of the motivation for alternative models should be the demonstrated inadequacy of the current SDL techniques (and inability or undesirability to limit the 5.6 billion tabular summaries).

More Generally

- Given heightened concerns about disclosure, let's make progress on all fronts simultaneously:
  - Designed protection of formal privacy models.
  - Traditional techniques of Statistical Disclosure Limitations.
  - Enhanced licensing requirements, more stringent access restrictions (e.g. RDCs), and more severe penalties for misuse of the most sensitive datasets.
- Roughly, the marginal cost of improving privacy in service of the Census Bureau's objectives should be equal across all of these fronts.

And Distinguish Census Products Based on the Potential for Harm

- What's the worst case scenario for (a de-identified) Census 2020?
  - At the Census Block level, why does any noise have to be introduced into tabulations of sex, age, race, Hispanic origin, and homeownership?
- I think the concerns are more evident in other datasets, particularly those that report dollar values like income with positive skewness.
- In any case, be specific about the harm these new methods are designed to prevent and which products are most vulnerable.

Some Readable Introductions to Differential Privacy

- Nissim, Kobbi, Thomas Steinke, Alexandra Wood et al. 2017. "Differential Privacy: A Primer for a Non-technical Audience." [link]
- Heffetz, Ori, and Katrina Liggett. 2014. "Privacy and Data-Based Research." *Journal of Economic Perspectives*. Vol. 28(2): 75 – 98. [link]
- Machanavajjhala, Ashwin, Xi He, and Michael Hay. 2016. "Differential Privacy in the Wild: A tutorial on current practices & open challenges." *Proceedings of the VLDB Endowment* 9(13): 1611-4. [link]

Defining Differential Privacy

- "An algorithm satisfies ε-differential privacy if its output on a database of individuals is statistically indistinguishable (measured by parameter ε) from the output of the algorithm if any one individual had opted out of the database."
- Statistically indistinguishable means:

$$\forall S \in Range(K) \qquad \frac{Prob[K(D) = S]}{Prob[K(D') = S]} \leq e^{\varepsilon}$$

For a randomized function K and all neighboring databases D and D'

Implementing Differential Privacy

- These algorithms work by infusing noise into query answers.
- To meet the criterion, the noise is drawn from a Laplace distribution with mean equal to zero and standard deviation = $\sqrt{2}\Delta f / \varepsilon$, where Δf is the maximum difference in the query answer between any neighboring databases.
- The higher is Δf, or the lower is ε (more privacy), the more noise that must be added.

Some Concerns

- For multiple queries, ε is additive, and more noise is required to preserve privacy.
- For count data, Δf = 1, but for highly skewed distributions, like income, Δf could be quite large.
- The noise is added to the query answer directly, not the potentially revealing underlying data. With large Δf, low ε, and small underlying values, the adjusted query responses could be obviously infeasible. How would users respond?