CS2080: Applied Privacy for Data Science
Reconstruction Takeaways & Defenses

James Honaker, Priyanka Nanayakkara, Salil Vadhan

School of Engineering & Applied Sciences
Harvard University

February 3, 2025

Takeaway Message on Reconstruction Attacks

- Every statistic released yields a (hard or soft) constraint on the dataset.
  - Sometimes have nonlinear or logical constraints ⇒ use fancier solvers (e.g. SAT or SMT solvers)

- Releasing too many statistics with too much accuracy necessarily determines almost the entire dataset.

- This works in theory and in practice (see readings, ps2).

How to Defend Against Reconstruction

- **Q:** what is a way that we can release many pretty-accurate estimates of proportions (counts divided by n) on a dataset while ensuring that an adversary can only reconstruct a small fraction of our dataset?

- **A:** subsample $k \ll n$ rows at random, and answer all queries using just the k rows.
  - If k is large enough (e.g. $k = 1000$), each individual proportion should be approximately preserved whp
  - We're only giving the adversary information about k rows, no info to reconstruct the others.

Subsampling vs. Reconstruction

- **Q:** If the adversary is just trying to reconstruct a single sensitive bit per individual, what fraction of the dataset should we expect the adversary to reconstruct if we subsample k rows and answer arbitrarily many counts?
- **Guess 1:** $\approx \frac{k}{n}$
- **Guess 2:** $\approx \frac{k}{n} + \frac{n-k}{n} \cdot \frac{1}{2} = \frac{1}{2} + \frac{1}{2} \cdot \frac{k}{n}$.
- **A:** $\approx \frac{k}{n} + \frac{n-k}{n} \cdot \max\{p, 1-p\} = \max\{p, 1-p\} + \overbrace{\min\{p, 1-p\} \cdot k/n}^{\text{gain over baseline}}$

  if the sensitive attribute is 1 in a p fraction of the population and the adversary has no other prior information about whether individuals in the dataset are 1 or 0.
- **Q:** is subsampling a satisfactory privacy defense?

The Utility of Subsampling

Q: why doesn't the subsampling defense disprove the Dinur-Nissim reconstruction theorem?

A: the error is too large

- if $X = (X_1 + \cdots + X_k)/k$ for independent random variables X_i each with standard deviation σ, then the standard deviation of X is $\sigma/\sqrt{k}$.
- So the "sampling error" when we subsample k rows with replacement is $\Theta(1/\sqrt{k})$. Also true for subsampling without replacement if $k \leq n/2$.
- The Dinur-Nissim Theorem we covered requires error $o(1/\sqrt{n})$.

Q: are attacks still possible if we allow error larger than $1/\sqrt{n}$?

A: yes, if we answer more queries