CS208: Applied Privacy for Data Science
Reconstruction Attacks

School of Engineering & Applied Sciences
Harvard University

January 25, 2022

Announcements

- Introduce yourself to your neighbors
- Wanrong introduction
- Fill out background survey if you haven't
- (Partial) slides & live chat on Ed
- Auditors: we will send you a Perusall access code
- Section today & tomorrow
- Salil OH tomorrow 11-12
- PS1 posted, due next Wed
- Track highlights of your participation (we'll ask you to submit "participation portfolios")

Common Themes from Comments

- Reinforcing uniqueness, ease of identifying
- Is there positive value in the data?
- Pros and cons of a “tiered access model”
- What about collection and retention?
- Who “owns” the data, and can we quantify its value?
- What are the harms, and have they been realized?
- Ethical hacking
- How do users respond to privacy tech/threats?
- Defaults and consent
- Who decides what is safe?
- Obligations and consequences for companies?
- Government vs. industry actors

Attacks on Aggregate Statistics

- Stylized set-up:
  - Dataset $x \in \{0,1\}^n$.
  - (Known) person i has sensitive bit x_i.
  - Adversary gets $q_S(x) = \sum_{i \in S} x_i$ for various $S \subseteq [n]$.

- How to attack if adversary can query chosen sets S?

- What if we restrict to sets of size at least $n/10$?

ID	US?

This attack has been used on Israeli Census Bureau!
(see [Ziv `13])

Attacks on Exact Releases

- What if adversary cannot choose subsets, but $q_S(x)$ is released for "innocuous" sets S?

- Example: uniformly random $S_1, S_2, \ldots, S_m \subseteq [n]$ are chosen, and adversary receives:
$\left(S_1, a_1 = q_{S_1}(x)\right), \left(S_2, a_2 = q_{S_2}(x)\right), \ldots, (S_m, a_m = q_{S_m}(x))$

- Claim: for $m = n$, with prob. $1 - o(1)$ adversary can reconstruct entire dataset!

- Proof?

Example for $n = 5$

$S_1 = \{1,2,3\}$, $a_1 = 2$, $S_2 = \{1,3,4\}$, $a_2 = 1$, $S_3 = \{4,5\}$, $a_3 = 1$, $S_4 = \{2,3,4,5\}$, $a_4 = 3$, $S_5 = \{1,2,4,5\}$, $\mathrm{a}_5 = 2$

Attacks on Approximate Statistics

- What if we release statistics $a_i \approx q_{S_i}(x)$?

- Thm [Dinur-Nissim `03]: given $m = n$ uniformly random sets S_j and answers a_j s.t. $\left|a_j - q_{S_j}(x)\right| \leq E = o(\sqrt{n})$, whp adversary can reconstruct $1 - o(1)$ fraction of the bits x_i.

- Proof idea: $A(S_1, a_1, \ldots, S_m, a_n)$ = any $\hat{x} \in \{0,1\}^n$ s.t.
$$\forall j \ \left|a_j - q_{S_j}(\hat{x})\right| \leq E.$$

(Show that whp, for all $\hat{x}$ that differs from x in a constant fraction of bits, $\exists i$ such that $\left|q_{S_j}(\hat{x}) - q_{S_j}(x)\right| > 2E$.)

Integer Programming Implementation

$A(S_1, a_1, \dots, S_m, a_n)$:

1. Find a vector $\hat{x} \in \mathbb{Z}^n$ such that:
   - $0 \leq \hat{x}_i \leq 1$ for all $i = 1, \dots, n$
   - $-E \leq a_j - \sum_{i \in S_j} \hat{x}_i \leq E$ for all $j = 1, \dots, m$
2. Output $\hat{x}$.

Problem: Can be computationally expensive (“NP-hard”, exponential time in worst case)

Faster: Linear Programming Implementation

$A(S_1, a_1, \dots, S_m, a_n)$:

1. Find a vector $\hat{x} \in \mathbb{R}^n$ such that:
   - $0 \leq \hat{x}_i \leq 1$ for all $i = 1, \dots, n$
   - $-E \leq a_j - \sum_{i \in S_j} \hat{x}_i \leq E$ for all $j = 1, \dots, m$
2. Output $\hat{x}$

Linear Programming Implementation for Average Error

$A(S_1, a_1, \ldots, S_m, a_n)$:

1. Find vectors $\hat{x} \in \mathbb{R}^n$ and $E \in \mathbb{R}^m$
   - Minimizing $\sum_{j=1}^{m} E_j$ and such that
   - $0 \leq \hat{x}_i \leq 1$ for all $i = 1, \ldots, n$
   - $-E_j \leq a_j - \sum_{i \in S_j} \hat{x}_i \leq E_j$ for all $j = 1, \ldots, m$
2. Output $\mathrm{round}(\hat{x})$.

Least-Squares Implementation for MSE

•

$A(S_1, a_1, \dots, S_m, a_n)$:

1. Find vector $\hat{x} \in \mathbb{R}^n$ minimizing

$$\sum_{j=1}^{m} \left(a_j - \sum_{i \in S_j} \hat{x}_i \right)^2 = \|a - M_S \hat{x}\|^2$$

2. Output $\mathrm{round}(\hat{x})$.

Also works for random S_j's, and is much faster than LP!

On the Level of Accuracy

- The theorems require the error per statistic to be $o(\sqrt{n})$. This is necessary for reconstructing almost all of x.

- Q: What is significant about the threshold of $\sqrt{n}$?
  - If dataset is a random sample of size n from a larger population, the standard deviation of a count query is $O(\sqrt{n})$.
  - Reconstruction attacks $\Rightarrow$ if we want to release many ($> n$) arbitrary or random counts, then we need introduce error at least as large as the sampling error to protect privacy.

How to Make Subset Sum Queries?

- Stylized set-up:
  - Dataset $x \in \{0,1\}^n$.
  - (Known) person i has sensitive bit x_i.
  - Adversary gets $a_S \approx q_S(x) = \sum_{i \in S} x_i$ for various $S \subseteq [n]$.

- Q: How to attack if the subjects aren't numbered w/ ID's?
  - If we know the set of people but not their IDs? (e.g. current Harvard students)
  - If we only know the size n of the dataset?

ID	US?

Overall Message

- Every statistic released yields a (hard or soft) constraint on the dataset.
  - Sometimes have nonlinear or logical constraints $\Rightarrow$ 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).

- We need a quantitative theory that tells us "how much is too much" $\rightarrow$ differential privacy!