Privacy in Statistics and Machine Learning Spring 2025
In-class Exercises for Lecture 3 (Reconstruction Part 2)
January 28, 2025

Adam Smith (based on materials developed with Jonathan Ullman)

Problems with marked with an asterisk (*) are more challenging or open-ended.

1. (Reconstruction via linear programming.) Consider the reconstruction attack that takes as input query vectors $F_1, \ldots, F_k \in \{0,1\}^n$ and noisy answers $a_1, \ldots, a_k \in \mathbb{R}$ and return the vector $\hat{s} \in [0,1]^n$ that minimizes

$$\max_{i=1,\ldots,k} |F_i \cdot \hat{s} - a_i| \tag{1}$$

Show how to write a linear program of the form introduced in the notes whose solution is the optimal vector $\hat{s}$.

2. (Preventing reconstructon with subsampling) Consider a dataset $\mathbf{x} = (x_1, \ldots, x_n)$. Now fix[1] $m = \frac{n}{5}$ and we will define the *subsampled dataset* $Y = (y_1, \ldots, y_m)$ as follows. For each $j \in [m]$, independently choose a random element $j' \in [n]$ and set $y_j = x_{j'}$. Note that the sampling is *independent* and *with replacement.* Suppose we now use Y to compute the statistics in place of $\mathbf{x}$. That is, using

$$5 \cdot f(Y) = 5 \cdot \sum_{j=1}^{m} \varphi(y_j) \tag{2}$$

in place of the true answer

$$f(\mathbf{x}) = \sum_{j=1}^{n} \varphi(x_j) \tag{3}$$

We multiply by 5 to account for the fact that $m = \frac{n}{5}$.

Prove that this random subsample will simultaneously give a good estimate of the answers to many statistics. Specifically, try to prove the following result

Claim 0.1. *Prove that for any set of statistics $f_1, \ldots, f_k$, with probability at least $\frac{99}{100}$,*

$$\forall i \in [k] \ : \ \left| 5 \cdot \sum_{j=1}^{m} \varphi_i(y_j) - \sum_{j=1}^{n} \varphi_i(x_j) \right| \leq O\left(\sqrt{n \log k}\right) \tag{4}$$

How good a reconstruction when queries are answered in this way?

Hint: To prove Claim 0.1 you will likely want to use the following form of "Chernoff Bound": if $Z_1, \ldots, Z_m$ are independent where each Z_j has expectation $\mathbb{E}(Z_j) = \mu$ and Z_j takes values in $[0,1]$ then for every $w > 0$,

$$\mathbb{P}\left(\left| \sum_{j=1}^{m} Z_j - m\mu \right| > w\sqrt{m} \right) \leq e^{-t^2/3} \tag{5}$$

[1] This setting of m just makes things more concrete. One can take m to be any size less than n; the statements just become more complicated.

3. * (More accurate reconstruction with more random queries.) In this question we'll explore how to interpolate between the two reconstruction theorems we've seen. Specifically, we will prove a version of Theorem 2.5 that gives a more accurate reconstruction when we have $k \gg n$ queries. Suppose we have the following version of Claim 2.6 from the lecture notes:

   Claim 0.2. *Let $t \in \{-1, 0, +1\}^n$ be a vector with at least m non-zero entries and let $u \in \{0, 1\}^n$ be a uniformly random vector. Then for every parameter $2 \le w \ll 2^m$*

   $$\mathbb{P}\left(|u \cdot t| \ge \frac{\sqrt{m \log w}}{10}\right) \ge \frac{1}{w} \tag{6}$$

   (a) Using this claim, prove the following theorem

   Theorem 0.3. *If we ask $n^2 \ll k \ll 2^n$ queries, and all queries have error at most αn, then with extremely high probability, the reconstruction error is at most $O(\frac{\alpha^2 n^2}{\log(k/n)})$.*

   (b) We can reformulate this as the following claim: the attacks gets nontrivial reconstruction error $o(n)$ when $\alpha = o(\frac{\cdots}{\cdots})$. Fill in the blanks.

   (c) How does this theorem compare to the reconstruction attacks we've seen for $k \approx n^2$? What about $k \approx 2^{\sqrt{n}}$? What about $k \approx 2^n$?

4. Now let's consider a slightly different setting, in which the attacker gets approximate answers to a highly structured set of queries.

   Specifically, suppose the secret data set s consists of n bits $s_1, \ldots, s_n$, and suppose the attacker receives approximate answers only to the n prefix sums of the form $\sum_{j=1}^{i} s_j$ (for i from 1 to n). These correspond to query vectors

   $$F_i = (\underbrace{1, 1, \ldots, 1}_{i \text{ ones}}, \underbrace{0, 0, \ldots, 0}_{n-i \text{ zeros}})$$

   (a) Suppose the curator answers all n questions *exactly*. How could the adversary recover s exactly?

   (b) Suppose that n is even (for simplicity) and s consists of alternating 0's and 1's, that is $s = (0101 \cdots 01)$. Show how you could give a sequence of answers $a_1, \ldots, a_n$ such that (i) each prefix sum query is answered to within 1, that is,

   $$|F_i \cdot s - a_i| \le 1 \quad \text{for all } i = 1, \ldots, n,$$

   and (ii) the algorithm of Figure 4 (in the lecture notes) would reconstruct a vector $\tilde{s}$ that is wrong in all n positions (that is, $\tilde{s}$ differs from s in every entry.

   (c) Try to generalize this as follows: suppose that s is uniformly random in $\{0, 1\}^n$. Give a procedure that takes s as input and returns a sequence of answers $a_1, \ldots, a_n$ such that (i) each prefix sum query is answered to within 1, that is,

   $$|F_i \cdot s - a_i| \le 1 \quad \text{for all } i = 1, \ldots, n,$$

   and (ii) the algorithm of Figure 4 would reconstruct a vector $\tilde{s}$ whose expected distance from s is $\Omega(n)$. (Here the expectation is taken over the choice of s; the attack of Figure 4 is deterministic and your algorithm can also be.)

   (d) **(*)** Can you come up with a version of this result that works against every attack algorithm (with high probability over the choice of s and any random choices made by your algorithm and the attack)?