Privacy in Statistics and Machine Learning — Spring 2025
In-class Exercises for Lecture 2 (Reconstruction Part 1)
January 23, 2025

Adam Smith (based on materials developed with Jonathan Ullman)

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

1. Do problem 1 from Lecture 1 exercises.
2. Do problem 2 from Lecture 1 exercises.
3. Consider the setting of Section 2.3 of the notes (Reconstruction from Many Queries) and the attack of Figure 4. Suppose the data set has size $n = 2,000$ and the attacker receives as input approximate answers to all possible linear queries on a secret vector s, each with error (at most) ± 100. What upper bound does Theorem 2.4 give on the attack's reconstruction error?
4. **(Baselines)** Suppose an attacker does not get access to the released statistics. They know only that $s \in \{0,1\}^n$ is uniformly random.
   (a) What is the *expected* error $\|\tilde{s} - s\|_1$ of every reconstruction attack that guesses a vector $\tilde{s} \in \{0,1\}^n$? Here $\|\tilde{s} - s\|_1$ is the number of bits in which $\tilde{s}$ and s differ.[1] This expected error gives us a *baseline* to evaluate when an attack that does use the released statistics is "interesting".
   (b) Suppose, as a different baseline, we want to understand the probability that an attacker without access to released statistic can guess $\tilde{s}$ such that $\|\tilde{s} - s\|_1 \leq n/4$. How could you argue that this probability is small? (Relevant tools include the Chebyshev inequality and Chernoff bounds.)
   In fact, the probability is exponentially small in n (that is, at most 2^{-cn} for a constant $c > 0$). Prove this using a Chernoff bound (see reminders at the end of Lecture 1 exercises.)

   (In general, finding a good "baseline" for evaluating reconstruction attacks is tricky.)
5. (Optimality of the attack in Section 2.3.)
   (a) Show that the guarantee of Theorem 2.4 is essentially tight. Specifically: Give an algorithm that takes as input a data set with a secret vector s of bits and an error rate α, and produces answers to all possible linear queries so that with (i) each approximate answer is within αn of the correct one, and (ii) the attack of Theorem 2.4 returns a vector $\tilde{s}$ with reconstruction error at least αn.
   (b) **(*)** Can you modify your procedure so it guarantees that *no attack algorithm* always returns a vector $\tilde{s}$ with reconstruction error less than αn—say $\alpha n/2$? What if the algorithm just has to work with high probability (say 0.95)? What additional assumptions does your result require? (For example, you might have to make assumptions about the distribution of s, and what the attack algorithm knows about it.)
6. Do the remaining problems from Lecture 1.

[1] More generally, $\|\cdot\|_1$ denotes the sum of the absolute values of the entries of a vector.