Privacy in Statistics and Machine Learning — Spring 2025
In-class Exercises for Lecture 16 (Projection Mechanism)
March 20, 2025

Adam Smith (based on materials developed with Jonathan Ullman)

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

1. (Exploring Projection) It's a little magical why the projection mechanism is able to take answers that are so noisy as to seemingly contain no information about the true answers, and somehow produce very accurate answers. Here is an example that we think demystifies it a bit.

   Suppose we have k copies of the exact same count query $f_1, \ldots, f_k$, so that $\mathbf{F} \in \{0,1\}^{k \times m}$ is a matrix whose rows are all the same. Suppose you compute noisy answers $\hat{\mathbf{a}} \in \mathbb{R}^k$ with the Gaussian mechanism and project into the set of feasible answers C.

   Note: The idea of this question is to work through an example from first principles. You can come up with a much tighter analysis that what is is implied by Theorem 3.1 from the lecture notes.

   (a) What is the set of consistent answers C?

   (b) Given noisy answers $\hat{\mathbf{a}}$, what is the projection $\tilde{\mathbf{a}} = \Pi_C(\hat{\mathbf{a}})$?

   (c) What is the error of the noisy answers $\hat{\mathbf{a}}$, and the projected answers $\tilde{\mathbf{a}}$? Does it increase with k?

2. (Understanding the bounds of Theorem 3.1) Suppose our data comes from the universe $\mathcal{U} = \{0,1\}^d$ and we are interested in releasing all *3-way marginal queries*—namely, for every triple of indices $i_1, i_2, i_3 \in [d]$, and for every three bits b_1, b_2, b_3, we want to estimate the fraction of records x in the data set with $x_{i_1} = b_1$ and $x_{i_2} = b_2$ and $x_{i_3} = b_3$.

   What are m and k here? What error bounds do we get for the regular Gaussian mechanism and the projection mechanism respectively?

   How do these answer change if we look at 1-way marginals, 2-way marginals, or 4-way marginals (instead of 3-way marginals)?

3. (Projection for Monotone Queries) This question explores the special case of the projection mechanism that arises from threshold queries over the domain $\mathcal{U} = \{1, \ldots, D\}$. Here there are $D-1$ non-trivial queries[1]

   (a) When $D = 3$, there are 2 non-trivial queries, and the feasible set is

   $$C = \left\{(a_1, a_2) \in \mathbb{R}^2 : 0 \le a_1 \le a_2 \le 1\right\}$$

   Write (or illustrate) the projection operation $\Pi_C(a_1, a_2)$ for this specific set C.

   (b) **(*)** Suppose the data happens to be such that all the entries of the true answer $a_1, ..., a_{D-1}$ are the same. Projecting a noisy version of this vector onto the set C in (1) will eliminate much of noise. How good a bound can you give on the ℓ_2 error after projection?

[1] The query $f_D(\mathbf{x}) = \#\{i : x_i \le D\}$ evaluates to 1 for every dataset.

(c) **(*)** For the general case, the feasible set has the form

$$C = \{(a_1, \ldots, a_{D-1}) : 0 \leq a_1 \leq \cdots \leq a_{D-1} \leq 1\} \tag{1}$$

Design an explicit polynomial time algorithm for computing projection into the set C (i.e. without relying on general linear programming as a tool).
[***Hints:*** (1) Dynamic programming! (2) It's a bit easier if we further require that all the projected answers are integer multiples of some discretization parameter $\gamma > 0$.]

4. (Bias and Variance) Unlike the Gaussian mechanism which is *unbiased*, meaning $\mathbb{E}(\tilde{\mathbf{a}}) = \mathbf{a}$, the projection mechanism is not necessarily unbiased.

   (a) Consider the case of a single proportion query, for which the feasible set is just $C = \{a : 0 \leq a \leq 1\}$. Given a data set where the true answer is a, let $bias(a) = \mathbb{E}(\tilde{a} - a)$ denote the bias of the projection mechanism with Gaussian noise $N(0, \sigma^2)$. For what value of a is this bias maximized? In that case, what is the limit of $\frac{bias(a)}{\sigma}$ as σ tends to 0?
   (Useful fact: If $Z \sim N(0, \sigma^2)$, then $\mathbb{E}(|Z|) = \sigma\sqrt{2/\pi}$.)

   (b) **(*)** For a single count query, show that *every* (ε, δ)-differentially private mechanism whose output is always in the range $[0, 1]$ must have bias $\Omega(1/\varepsilon n)$, at least when δ is sufficiently small.

5. **(From Lecture 13 on the Matrix Mechanism)** Let $\mathbf{J}_T \in \{0, 1\}^{T \times T}$ be the workload matrix for threshold queries. This matrix is *Toeplitz*, which means that the values in the matrix are constant along each of the diagonals (parallel to the main diagonal). One factorization that often works well uses the Toeplitz matrix square root $\mathbf{R}_T$, which is a Toeplitz matrix such that $\mathbf{R}_T^2 = \mathbf{J}_T$. (The factorization is $L = R = \mathbf{R}_T$.)

   One way to find such a matrix is by diagonalizing $\mathbf{J}_T$ and taking the square roots of the eigenvalues (which might result in complex values). We will instead take a more abstract approach.

   A lower-triangular Toeplitz matrix can be described by the sequence of values $\alpha_0, \alpha_1, \alpha_2$ found on each of the diagonals, where entry (i, j) is given by α_{i-j} for $i \geq j$. So $\mathbf{J}_T$ is described by the sequence 1, 1, ..., 1.

   (a) Show that if two T-dimensional Toeplitz matrices are given by the sequences $\vec{\alpha}$ and $\vec{\beta}$, then their product is Toeplitz and described by $\vec{\gamma}$ where $\gamma_i = \sum_{j=0}^{i} \alpha_j \beta_{i-j}$.

   (b) We can associate each sequence with a real-valued polynomial. Specifically, let $f_{\vec{\alpha}}(x) = \sum_{i=0}^{T-1} \alpha_i x^i$ (where x in $\mathbb{R}$). Show that $\mathbf{J}_T$ is associated with the function $f_T(x) = \frac{1-x^{T-1}}{1-x}$.

   (c) Consider the matrix $\mathbf{R}_T$ described by the first T terms of the power series of $g(x) = \frac{1}{\sqrt{1-x}}$ (no need yet to write out the actual coefficients). Using just the fact that g is the square root of $\frac{1}{1-x}$, show that $\mathbf{R}_T^2 = \mathbf{J}_T$.

   (d) What sequence $\vec{\rho}$ describes the square root of $\mathbf{J}_T$? (You can use Internet tools for this.) How do the diagonal values ρ_i decrease with i?

   (e) (*) For finite T, how does this factorization compare to that implied by the binary tree mechanism? You could answer this numerically.

   (f) Given a sequence $\alpha \in \mathbb{R}^T$, show how we can multiply a vector by the Toeplitz matrix given by α in time $O(n \log n)$ by reducing to the problem of convolution (which is solved by the FFT).