Week 4 Exercises (ECE 598 DA)

Exercise (DP Counting Queries and Revealing Individual Data): We have a stateless ε-DP mechanism (Laplace mechanism) that answers counting queries by adding noise $\mathsf{Lap}(0, 1/\varepsilon)$ to each true count. Show that if an adversary strategically asks $k = 2n$ counting queries, this mechanism will almost surely reveal the exact data of at least one individual. (Hint: use queries that isolate each individual's data, e.g., the i-th query asks for the count of records in position i. Over many queries, some noise will be small compared to the difference in counts.)

Exercise (Laplace Mechanism Optimality with Variance Lower Bound): We have n binary data points (true mean μ). Consider any ε-DP mechanism A that outputs an estimate $\hat{\mu}$ of the true mean. For any two neighboring datasets $\mathbf{x}$ and $\mathbf{x}'$ differing in one entry (so their true means differ by $1/n$, say $\mu' = \mu + 1/n$), argue that the probability distributions $\mathsf{A}(\mathbf{x})$ and $\mathsf{A}(\mathbf{x}')$ must be ε-close, and use this to show $\mathsf{Var}(\hat{\mu}) = \Omega(1/(n^2\varepsilon^2))$. (Hint: If the variance were much smaller, the two output distributions would be nearly disjoint. One could distinguish $\mathbf{x}$ vs. $\mathbf{x}'$ with too high confidence, violating the DP likelihood ratio bound.

Exercise (Lower Bound for k Counting Queries via Group Privacy): Using group privacy, prove a lower bound on noise for answering multiple queries. Suppose we have k counting queries and a mechanism that adds noise with standard deviation σ to each query's true answer. Use the definition of group privacy to argue that if $\sigma \ll \frac{k}{\varepsilon}$ (noise much smaller than k/ε), then an adversary can distinguish (with too high confidence) between a database $\mathbf{x}$ and a database $\mathbf{x}'$ that differ in k individuals (each contributing to a different query). (Hint: Consider two databases that differ in k rows, with each differing individual affecting one query. Look at the likelihood ratio of the output vectors Y on $\mathbf{x}$ vs. $\mathbf{x}'$. Use a tail bound to show the probability of a certain output event under $\mathbf{x}$ is far larger than under $\mathbf{x}'$, violating the $e^{\varepsilon k}$ bound from group privacy.)