CS839: Differential Privacy and Learning April 7, 2026

Lecture 21: Packing Lower Bounds

Instructor: Gavin Brown *Scribe:* Akhil Vanukuri

Disclaimer: This document is intended as an informal supplement to in-class note-taking. It has not been given the level of scrutiny expected in polished lecture notes, let alone that reserved for peer-reviewed publications.

In this lecture, we will look at the techniques to derive lower bounds on the sample complexity of differentially private algorithms (i.e, for a given statistical task, what is the minimum number of samples required to construct a differentially private version of it with certain target accuracy). While our focus will be on deriving statistical/information-theoretic lower bounds, it is worth noting that computational lower bounds also exist and are interesting (although they are less well-studied).

Here are four approaches to show lower bounds in DP:

- Packing/Group Privacy
- Fingerprinting
- DP interior point
- Instance optimality

Over the next several classes, we will discuss the first three of these approaches. We previously talked informally about instance optimality, but won't see it in any greater detail in this class.

Packing (or) Group Privacy

Theorem. *Fix $\varepsilon, \delta, B, \alpha$. Suppose $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}$ is $(\varepsilon, \delta) - \mathsf{DP}$ and satisfies the following: $\forall \mu \in [-B, +B]$, if $X \sim \mathcal{N}(\mu, 1)^{\otimes n}$ then*

$$\Pr[|\mathcal{A}(X) - \mu| \leq \alpha] \geq \frac{2}{3}.$$

Then, suppressing an absolute constant,

$$n \gtrsim \frac{1}{\alpha^2} + \frac{1}{\alpha\varepsilon} + \min\left(\frac{\log(\frac{1}{\delta})}{\varepsilon}, \frac{\log(B)}{\varepsilon}\right)$$

Note: The lower bound above has three terms. The teal part is required without privacy, i.e, even the true empirical mean needs $n \approx \frac{1}{\alpha^2}$ samples to achieve α accuracy. The term is the tradeoff between the level of privacy (ε) and the accuracy (α), the error due to privacy. The purple part is the cost to get started (observe that this part is independent of α). Another way to interpret this result is that, once $n \geq \min\left(\frac{\log(\frac{1}{\delta})}{\varepsilon}, \frac{\log(B)}{\varepsilon}\right)$ then with constant probability we know that $|\mathcal{A}(X) - \mu| \geq \frac{1}{\sqrt{n}} + \frac{1}{n\varepsilon}$.

To recall one upper bound (there are many algorithms for one-dimensional data), the Friendly-Core algorithm that we discussed in the previous lectures achieves α accuracy once $n \approx \frac{1}{\alpha^2} + \frac{\sqrt{\log \frac{1}{\delta}}}{\alpha\varepsilon} + \frac{\log \frac{1}{\delta}}{\varepsilon}$, nearly matching the above when B is large.

Some observations:

1. If $B = +\infty$ then we need $\delta > 0$ otherwise sample complexity (i.e n) is ∞.
2. If $\delta = 0$ (i.e, pure DP) then error must grow with B, which quantifies the strength of our prior knowledge about where μ lies.

For the next few results, our focus will be on the cases where two datasets differ on multiple entries and for that purpose let's recall the following lemma about group privacy from previous lectures and use that to make observations about the lower bound guarantees in that case,

Lemma (Group Privacy)**.** *Suppose $\mathcal{A}$ is $(\varepsilon, \delta) - \mathsf{DP}$ and let x, x' differ in at most k entries then $\mathcal{A}(x) \approx_{(\varepsilon', \delta')} \mathcal{A}(x')$ for $\varepsilon' = k \cdot \varepsilon$ and $\delta' = k \cdot e^{\varepsilon \cdot k} \cdot \delta$.*

1 Packing Lower Bounds

In the rest of class, we will provide a partial proof of the purple term, the third term above. We will start by ignoring the fact that we want to prove lower bounds for data from Gaussian distributions.

The Setup Instead, we construct a set of datasets $\{X_{-B}, X_{-B+1}, \ldots, X_0, \ldots, X_B\}$ (assume for simplicity B is a natural number) where $X_j = (j, j, \ldots, j)$. So each dataset consists of n copies of the same number. We will ask for a very weak notion of accuracy: for each j, let $G_j = [j - 1/4, j + 1/4]$ be the set of "good answers" for dataset X_j. Note that G_i and G_j are disjoint when $i \neq j$.

Pure DP Suppose A is ε-DP and somewhat accurate: for all j, $\Pr[A(X_j) \in G_j] \geq \frac{2}{3}$. Then, by group privacy, we have for all i

$$\Pr[A(X_i) \in G_j] \geq e^{-\varepsilon n} \Pr[A(X_j) \in G_j] \geq \frac{2}{3} e^{-\varepsilon n}.$$

This says that, for any i, j, on input X_i we have a small, but non-zero, probability of returning a good answer for dataset X_j.

We now use the fact that there are many datasets with disjoint good answers.

$$\begin{aligned} 1 &\geq \Pr\left[A(X_0) \in \bigcup_j G_j\right] \\ &= \sum_j \Pr[A(X_0) \in G_j] \\ &\geq 2B \cdot \frac{2}{3} \cdot e^{-\varepsilon n}. \end{aligned}$$

Rearranging, we see that $n \geq \frac{1}{\varepsilon} \log(4B/3)$, as desired.

Approximate DP Now, we will retrace the same steps for approx-DP, with the slightly more complex group privacy statement. Suppose $\mathcal{A}$ is $(\varepsilon, \delta) - \mathsf{DP}$ (i.e approx-DP) and, $\forall j \in \{-B \ldots, B\}, \Pr[\mathcal{A}(X_j) \in G_j] \geq \frac{2}{3}$ then,

$$\begin{aligned}\forall i \neq j, \Pr[\mathcal{A}(X_j) \in G_i] &\geq \big(\frac{2}{3} - ne^{\varepsilon n}\delta\big)e^{-\varepsilon n} \\ &\geq \big(\frac{2}{3} - \delta'\big)e^{-\varepsilon n} \text{ (simply rewriting } \delta' = n^{e^{\delta n}}\delta \text{ in the spirit of group privacy)}\end{aligned}$$

Assuming $\frac{2}{3} - \delta' \geq \frac{1}{10}$, we get $n \geq \frac{\log\left(\frac{4B}{3}\right)}{\varepsilon}$ as in the pure-DP case. However, please note that this is vacuous bound when $n \geq \log\left(\frac{1}{\delta}\right)$.

2 Extending the result to Gaussian distributions

Unlike the fingerprinting lower bounds we will see next lecture, packing arguments for DP aren't tailored to data coming from a specific family of distributions.[1] However, the approach extends quite naturally. The key fact is that we can substitute total variation distance between distributions P_i, P_j in place of the Hamming distance between datasets X_i, X_j.

Lemma (Group Privacy for Distributional Data)**.** *Let A be (ε, δ)-DP and let P, Q be distributions. Suppose $X \sim P^{\otimes n}$ and $X' \sim Q^{\otimes n}$. Then $A(X) \approx_{\varepsilon', \delta'} A(X')$ for*

$$\begin{aligned}\varepsilon' &= 6\varepsilon \cdot n d_{TV}(P, Q) \\ \delta' &= 4e^{\varepsilon'}\delta \cdot n d_{TV}(P, Q).\end{aligned}$$

This is, up to constants, exactly like standard group privacy with $k = nd_{TV}(P, Q)$. With this lemma, the rest of the argument is identical: we construct a set of Gaussian distributions (each of which is at total variation distance at most 1 from any other) with disjoint "good" answers.

We will not prove this lemma (it comes directly from Karwa and Vadhan [2017], Lemma 6.1). However, we will say a few words about the technique behind it, as this very useful.

Definition 1 (Coupling)**.** Let p, q be distributions over the same space $\mathcal{U}$. A *coupling of p and q* is a pair of random variables (X, Y) such that for all $u \in \mathcal{U}$, $\Pr[X = u] = p(u)$ and $\Pr[Y = u] = q(u)$.

One coupling is to take $X \sim p$ and $Y \sim q$ independently. In a general, X and Y will not be independent. We need a particular coupling which relates to total variation distance.

Lemma. *For any coupling (X, Y) of distributions p, q, we have $\Pr[X \neq Y] \geq d_{TV}(p, q)$. Moreover, there is a coupling (called a* maximal coupling*) which achieves equality.*

References

Vishesh Karwa and Salil Vadhan. Finite sample differentially private confidence intervals. *arXiv preprint arXiv:1711.03908*, 2017.

[1] Not everything called "fingerprinting" relies on such assumptions, but the technique we will see does.