Lecture 11
Packing Lower Bounds

Settings

One-way marginals: $\mathcal{X} = \{0,1\}^d$
$f_j(X) = \frac{1}{n}\sum_{i=1}^{n} X_i^{(j)}$

Pure DP: Given $n \geq \Omega\left(\frac{d \log d}{\alpha \varepsilon}\right)$, $err \leq \alpha$ $\forall$ o.w.m q's w.p $\geq \frac{1}{2}$.
Approx: $n \geq \Omega\left(\frac{\sqrt{d \log(1/\delta)}}{\alpha \varepsilon}\right)$

Histograms: $\mathcal{X} = [k]$
$f_j(X) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_i = j\}$

Pure DP: $n \geq \Omega\left(\frac{\log k}{\alpha \varepsilon}\right)$
Approx DP: $n \geq \Omega\left(\frac{\log(1/\delta)}{\alpha \varepsilon}\right)$

"Packing" Lower bound

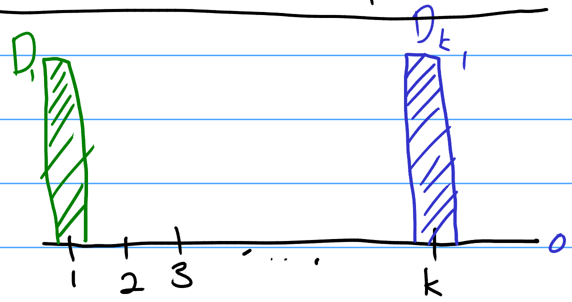
M: both <u>private</u> + <u>accurate</u>

Example 1: Mode of a dataset

$\mathcal{X} = [k]$

Want M: - ε-DP

- Output mode of X w.p. $\geq \frac{1}{2}$



$D_1, \ldots, D_k : D_j = \{j\}^n$

Fix j. Dist of $M(D_j)$.

$\exists \ell \in [k]$, s.t. $\Pr[M(D_j) = \ell] \leq \frac{1}{k}$

Acc. $\rightarrow$ $\Pr[M(D_\ell) = \ell] \geq \frac{1}{2}$

Group privacy

$$\frac{1}{2} \underset{\text{acc}}{\leq} \Pr[M(D_\ell) = \ell] \leq e^{\varepsilon n} \Pr[M(D_j) = \ell] \leq \frac{e^{\varepsilon n}}{k}$$

$$e^{\varepsilon n} \geq \frac{k}{2} \rightarrow \varepsilon n \geq \log(k/2)$$

$$\hookrightarrow n \geq \frac{\log(k/2)}{\varepsilon} = \Omega\left(\frac{\log k}{\varepsilon}\right)$$

$p = \frac{1}{2},\ t = n$
$m \leq 2e^{n\varepsilon}$

Key Packing Theorem

Theorem 1. *Let $D_1, \ldots, D_m \in \mathcal{X}^n$ be a set of m datasets, which are at Hamming distance at most t from some fixed dataset $D \in \mathcal{X}^n$. Let $Y_1, \ldots, Y_m \in \mathcal{Y}$ be a set of m disjoint subsets of the space $\mathcal{Y}$. If there is an ε-DP mechanism $M : \mathcal{X}^n \to \mathcal{Y}$ such that $\Pr[M(D_\ell) \in Y_\ell] \geq p$ for every $\ell \in [m]$, then*

$$\frac{1}{m} \geq pe^{-t\varepsilon}.$$

$\Pr[M(D_\ell) \in Y_\ell] \geq p.$

$\Pr[M(D_\ell) \in Y_\ell] \leq e^{t\varepsilon} \Pr[M(D) \in Y_\ell]$

Group privac

$\Pr[M(D) \in Y_\ell] \geq pe^{-t\varepsilon}.$

$mpe^{-t\varepsilon} \leq \sum_{\ell \in [m]} \Pr[M(D) \in Y_\ell] = \Pr\left[M(D) \in \bigcup_{\ell \in [m]} Y_\ell\right] \leq 1$ ▨

Disjoint

Approx
$n = O(\sqrt{d}/\varepsilon)$

Example 2: One-Way Marginals

Theorem 2. *Any ε-DP algorithm $M : \{0,1\}^{d \times n} \to [0,1]^d$ which simultaneously answers all one-way marginals to accuracy $< 1/2$ with probability $\geq 1/2$ requires $\boxed{n = \Omega(d/\varepsilon)}$.*

Theorem 1. *Let $D_1, \dots, D_m \in \mathcal{X}^n$ be a set of m datasets, which are at Hamming distance at most t from some fixed dataset $D \in \mathcal{X}^n$. Let $Y_1, \dots, Y_m \in \mathcal{Y}$ be a set of m disjoint subsets of the space $\mathcal{Y}$. If there is an ε-DP mechanism $M : \mathcal{X}^n \to \mathcal{Y}$ such that $\Pr[M(D_\ell) \in Y_\ell] \geq p$ for every $\ell \in [m]$, then*

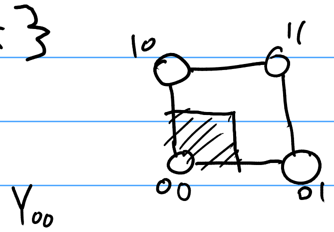
$$\frac{1}{m} \geq pe^{-t\varepsilon}.$$

$m = 2^d$ databases

$\forall\, w \in \{0,1\}^d$, $D_w = n$ copies of w.

$Y_w = \{x \in [0,1]^d : |x_j - w_j| < 1/2, \forall j \in k\}$

$Y_w = \ell_\infty$-ball of radius $1/2$ around w.



$\Pr[M(D_w) \in Y_w] \geq \frac{1}{2}$.

$p = 1/2$, $t = n$, $m = 2^d$

$2^{-d} \geq \frac{1}{2} e^{-n\varepsilon} \Rightarrow n = \Omega(d/\varepsilon)$ ∎

Theorem 1. *Let $D_1, \dots, D_m \in \mathcal{X}^n$ be a set of m datasets, which are at Hamming distance at most t from some fixed dataset $D \in \mathcal{X}^n$. Let $Y_1, \dots, Y_m \in \mathcal{Y}$ be a set of m disjoint subsets of the space $\mathcal{Y}$. If there is an ε-DP mechanism $M : \mathcal{X}^n \to \mathcal{Y}$ such that $\Pr[M(D_\ell) \in Y_\ell] \geq p$ for every $\ell \in [m]$, then*

$$\frac{1}{m} \geq pe^{-t\varepsilon}.$$

Example 3: Histograms

Theorem 3. *Any ε-DP algorithm $M : [k]^n \to [0,1]^k$ which estimates all histogram counts to accuracy $\leq \alpha$ with probability $\geq 1/2$ requires $n = \Omega\left(\frac{\log k}{\alpha\varepsilon}\right)$.*

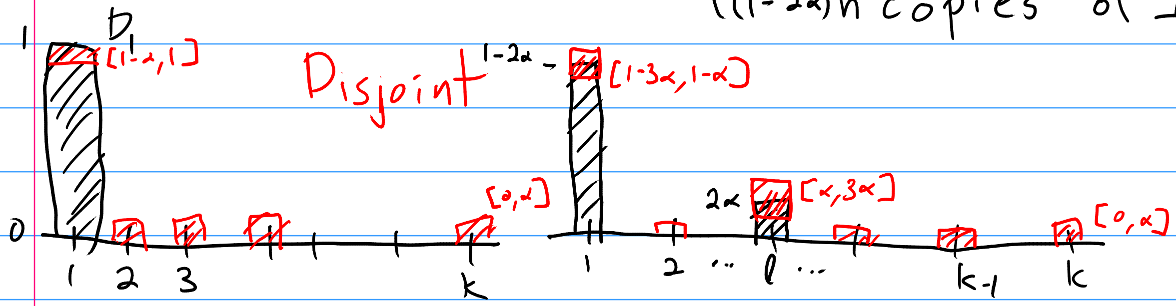
Proof: $X \in [k]^n \to h(X) \in [0,1]^k$. $\quad h_j(X) = \frac{1}{n} \sum \mathbb{1}\{X_i = j\}$

$y(X) \subseteq [0,1]^k$

$= \ell_\infty$-ball of radius α around $h(X)$

$y_j(X) = h_j(X) \pm \alpha \qquad D_1, \dots, D_k$ ($m=k$ databases)

$\Pr[M(X) \in y(X)] \geq \frac{1}{2} \qquad D_\ell = \begin{cases} t = 2\alpha n \text{ copies of } \ell \\ (1-2\alpha)n \text{ copies of } 1 \end{cases}$



$Y_\ell = y(D_\ell) \Rightarrow Y_\ell$'s are disjoint

$$\frac{1}{k} \geq \frac{1}{2} \exp(-2\alpha n \varepsilon)$$

$$k \leq 2 \exp(2\alpha n \varepsilon)$$

$$\log(k/2) \leq 2\alpha n \varepsilon$$

$$n \geq \frac{\log(k/2)}{2\alpha\varepsilon} = \Omega\left(\frac{\log k}{\alpha\varepsilon}\right) \quad \square$$