Packing Lower Bounds

Jonathan Ullman, Northeastern University
Outline

- **Packing arguments** for DP lower bounds
  - Originated in [HT’10, BKN’10]
  - Intuitive, geometric approach to lower bounds
  - Applicable to a wide variety of problems
  - Often yields tight lower bounds for $\epsilon, 0$-dp
  - Separates $\epsilon, 0$-dp ("pure") from $\epsilon, \delta$-dp ("approx")
Main Idea

$X^n$: all datasets

$M(x_1)$

$R$: all outputs

$M(x_2)$

$f(x_1)$

$f(x_2)$

$f(x_3)$

“packing” $P \subseteq X^n$

- Find many datasets $P \subseteq X^n$ that are close, but whose answers are far
  - DP implies that $M(x), M(x')$ are close
  - Accuracy implies that $M(x), M(x')$ are far.
Main Idea

$X^n$: all datasets

$M(x_1)$

$M(x_2)$

$R$: all outputs

$G(x_1)$

$G(x_2)$

$G(x_3)$

“packing” $P \subseteq X^n$

• Find many datasets $P \subseteq X^n$ that are close, but whose answers are far
  • DP implies that $M(x), M(x')$ are close
  • Accuracy implies that $M(x), M(x')$ are far.

$f(x)$ is some function of interest
$G(x)$ are “good outputs for $x$”
Recall Group Privacy

- Two datasets \( x, x' \in X^n \) are neighbors if they differ on at most one row \((x \sim x')\).
- Two datasets \( x, x' \in X^n \) are \( m \)-neighbors if they differ on at most \( m \) rows \((x \sim_m x')\).
- Lemma: If \( M : X^n \rightarrow R \) is \((\varepsilon, 0)\)-differentially private then for every set of \( m \)-neighbors \( x \sim_m x' \), and every \( S \subseteq R \),

\[
\Pr[M(x) \in S] \leq e^{\varepsilon m} \Pr[M(x') \in S]
\]

- NB: \((\varepsilon, \delta)\)-dp doesn’t behave as nicely for large groups.
Example: Histograms

- Dataset: \( x = (x_1, \ldots, x_n) \in X^n \)
- Histogram: \( h(x)_j = \#\{i : x_i = j\} \)
- Accuracy: Release \( \widehat{h} \) such that \( \max_j |h(x)_j - \widehat{h}_j| \leq \frac{n}{3} \)}
Example: Histograms

- Dataset: $x = (x_1, \ldots, x_n) \in X^n$
- Histogram: $h(x)_j = \#\{i : x_i = j\}$
- Accuracy: $G(x) = \left\{ \hat{h} \mid \max_z |\hat{h}_z - h(x)_z| \leq \frac{n}{3} \right\}$

- Q1: Suppose we use Laplace, how much noise do we need?
  
- A1: Global $\ell_1$-sensitivity is 1, add $\text{Lap}\left(\frac{1}{\epsilon}\right)$ to each entry

- Q2: How big must $n$ be to satisfy accuracy?
  
- A2: Largest entry has error $\Theta\left(\frac{\ln|X|}{\epsilon}\right)$ whp. So $n = \Theta\left(\frac{\ln|X|}{\epsilon}\right)$ is sufficient for accuracy.
Example: Histograms

- Thm: If $M : X^n \rightarrow \mathbb{N}^{|X|}$ is $(\varepsilon, 0)$-differentially private and $\Pr[M(x) \text{ is an accurate histogram}] \geq \frac{1}{e}$, then $n \geq \frac{\ln|X| - 1}{\varepsilon}$.

- Proof: Define the following “packing” of $|X|$ datasets:

  - No histogram is “good” for both $p$ and $p'$
    - $G(p_1), \ldots, G(p_{|X|})$ are mutually disjoint
Example: Histograms

• Thm: If \( M : X^n \rightarrow \mathbb{N}^{\lvert X \rvert} \) is \((\varepsilon, 0)\)-differentially private and \( \Pr[M(x) \text{ is an accurate histogram}] \geq \frac{1}{e} \), then \( n \geq \frac{\ln |X| - 1}{\varepsilon} \)

• Proof:

\[
1 \geq \sum_z \Pr[M(p_1) \in G(p_z)] \quad \text{(disjointness)}
\]

\[
\geq \sum_z e^{-\varepsilon n} \Pr[M(p_1) \in G(p_z)] \quad \text{(group privacy, size } n)\]

\[
\geq \sum_z e^{-\varepsilon n} \frac{1}{e} \quad \text{(accuracy)}
\]

\[
= |X|e^{-\varepsilon n} - 1 \quad \text{(size of packing is } |X|)\]

\[
\Rightarrow n \geq \frac{\ln |X| - 1}{\varepsilon}
\]
General Packing Lemma

• Let \( \{G(x)\}_{x \in X^n} \) be a family of subsets of the output range \( R \)
  • These are the “good outputs for \( x \)”

• \( m \)-Packing: Let \( P = \{x_0, x_1, \ldots \} \subseteq X^n \) be such that
  • every \( x, x' \in P \) are \( m \)-neighbors (datasets are close)
  • \( G(p_0), G(p_1), \ldots \) are mutually disjoint (answers are far)

Lemma: If \( P \) is an \( m \)-packing and \( M : X^n \to R \) is an \((\varepsilon, 0)\)-dp algorithm such that \( \Pr[M(x) \in G(x)] \geq \frac{1}{e} \),
then \( m \geq \frac{\ln|P| - 1}{\varepsilon} \)
Packing Lemma

• Lemma: If $M : X^n \rightarrow R$ is $(\varepsilon, 0)$-differentially private, and $\forall x \in P, \Pr[M(x) \in G(x)] \geq \frac{1}{e}$, then $m \geq \frac{\ln|P| - 1}{\varepsilon}$

• Proof:

\[ 1 \geq \sum_z \Pr[M(x_0) \in G(x_z)] \quad \text{(disjointness)} \]

\[ \geq \sum_z e^{-\varepsilon m} \Pr[M(x_z) \in G(x_z)] \quad \text{(group privacy, size } m) \]

\[ \geq \sum_z e^{-\varepsilon m} \frac{1}{e} \quad \text{(accuracy)} \]

\[ = |P|e^{-\varepsilon m} - 1 \quad \text{(size of packing)} \]

\[ \Rightarrow m \geq \frac{\ln|P| - 1}{\varepsilon} \]
Example: Dataset Mean

- Dataset: \( x = (x_1, \ldots, x_n) \in (\{0,1\}^d)^n \)
- Mean: \( \mu(x) = \frac{1}{n} \sum_i x_i \)
- Accuracy: \( G(x) = \left\{ \hat{\mu} \mid \max_c |\mu(x)_c - \hat{\mu}_c| \leq \alpha \right\} \)

Q1: Suppose we use Laplace, how much noise do we add?
A1: Global \( \ell_1 \)-sensitivity is \( \frac{d}{\epsilon n} \), add \( \text{Lap} \left( \frac{d}{\epsilon n} \right) \) to each entry

Q2: How accurate?
A2: \( \alpha = O \left( \frac{d \ln d}{\epsilon n} \right) \) whp. Can be improved to \( \alpha = O \left( \frac{d}{\epsilon n} \right) \).
Example: Dataset Mean

• Dataset: \( x = (x_1, \ldots, x_n) \in (\{0,1\}^d)^n \)

• Mean: \( \mu(x) = \frac{1}{n} \sum_i x_i \)

• Accuracy: \( G(x) = \left\{ \hat{\mu} \mid \max_c |\mu(x)_c - \hat{\mu}_c| \leq \alpha \right\} \)

\[
\begin{array}{cccccc}
\text{dataset } x & 1 & 1 & 0 & 0 & 0 \\
& 0 & 1 & 0 & 1 & \\
& 0 & 1 & 0 & 1 & \\
\hline
\mu(x) & 0.333 & 1.000 & 0.000 & 0.667 & \\
\hat{\mu} & 0.360 & 0.980 & 0.045 & 0.700 & \\
\end{array}
\]

\( \alpha = .045 \)
Example: Dataset Mean

- Dataset: \( x = (x_1, \ldots, x_n) \in (\{0,1\}^d)^n \)
- Mean: \( \mu(x) = \frac{1}{n} \sum_i x_i \)
- Accuracy: \( G(x) = \{ \hat{\mu} \mid \max_c |\mu(x)_c - \hat{\mu}_c| \leq \alpha \} \)
- Define the following packing:
  - \( P = \{p_z\}_{z \in \{0,1\}^d} \)
  - \(|P| = 2^d\)
  - \(p, p'\) are \( m = 3\alpha n \) neighbors
  - \( G(x) = \{ \hat{\mu} \mid \|\mu(x) - \hat{\mu}\|_\infty \leq \alpha \} \)
- Packing lemma \( \Rightarrow 3\alpha n \geq \frac{d-1}{\varepsilon} \)
Example: Dataset Mean

- Dataset: $x = (x_1, ..., x_n) \in (\{0,1\}^d)^n$
- Mean: $\mu(x) = \frac{1}{n} \sum_i x_i$
- Accuracy: $G(x) = \left\{ \hat{\mu} \mid \max_c |\mu(x)_c - \hat{\mu}_c| \leq \alpha \right\}$
- Define the following packing:
  - $P = \{p_z\}_{z \in \{0,1\}^d}$
  - $|P| = 2^d$
  - $p, p'$ are $m = 3\alpha n$ neighbors

Packing lemma $\Rightarrow \alpha \geq \frac{d-1}{3\varepsilon n}$

Ex: $p_{1001}$

\[\begin{array}{c|c}
\hline
1001 & 3\alpha n \text{ rows} \\
1001 & 1001 \\
0000 & 0000 \\
\hline
\end{array}\]

$\mu(p_z) = 3\alpha z$
Statistical Queries (SQs)

• Recall statistical queries $q(x) = \frac{1}{n} \sum_i \phi(x_i)$

• The mean is $d$ statistical queries on $x \in (\{0,1\}^d)^n$

• Thm: Laplace noise is $(\varepsilon, 0)$-dp and answers $k$ arbitrary SQs up to error $\alpha = O\left(\frac{k \ln k}{\varepsilon n}\right)$

• Thm: No $(\varepsilon, 0)$-dp algorithm $M : X^n \rightarrow R$ can answer $k \leq \log |X|$ arbitrary SQs with $\alpha < \frac{k-1}{3 \varepsilon n}$
Statistical Queries (SQs)

• Recall statistical queries $q(x) = \frac{1}{n} \sum_i \phi(x_i)$

• The mean is $d$ statistical queries on $x \in (\{0,1\}^d)^n$

• Thm: Laplace noise is $(\varepsilon, \delta)$-dp and answers $k$ arbitrary SQs up to error $\alpha = \tilde{O}\left(\frac{\sqrt{k \ln(1/\delta)}}{\varepsilon n}\right)$

• Packing lower bound is false for approximate dp.

• Later on we’ll see how to show tight lower bounds for $(\varepsilon, \delta)$-dp using very different techniques
Example: Online Counting

• Data: stream of bits $x_1, \ldots, x_T \in \{0,1\}^T$, given one at a time
• Goal: after $x_t$, output $a_t$ approximating $c_t = \sum_{t' \leq t} x_{t'}$
• Accuracy: $\max_t |a_t - c_t| \leq \alpha$

• Fact: there is an $(\varepsilon, 0)$-dp algorithm with accuracy $\alpha = O(\varepsilon^{-1} \ln T)$. (Binary tree gives $\alpha = O(\varepsilon^{-1} \ln^2 T)$.)

• Theorem: for every $(\varepsilon, 0)$-dp algorithm $\alpha = \Omega(\varepsilon^{-1} \ln T)$. 
Example: Online Counting

• Theorem: for every \((\varepsilon, 0)\)-dp algorithm \(\alpha = \Omega(\varepsilon^{-1} \ln T)\).

\[
\text{Split the input into } B = \frac{T}{3\alpha} \text{ blocks of length } 3\alpha.
\]

<table>
<thead>
<tr>
<th>(p_j)</th>
<th>00000</th>
<th>11111</th>
<th>00000</th>
<th>00000</th>
<th>00000</th>
</tr>
</thead>
<tbody>
<tr>
<td>block (j)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

• \(P = \{ p_j : j = 1, \ldots, \frac{T}{3\alpha} \}; |P| = \frac{T}{3\alpha} \); distance \(m = 3\alpha\).

• \(G(x) = \{ (a_1, \ldots, a_T) : \max_t |c_t - a_t| \leq \alpha \}\)
  • \(G(p_j), G(p_j')\) are disjoint

• Packing Lem. \(\Rightarrow m \geq \frac{\ln|P|-1}{\varepsilon} \Rightarrow 3\alpha \geq \frac{\ln(T)-\ln(3\alpha)-1}{\varepsilon}\)
Example: Online Counting

• Theorem: for every $(\varepsilon, 0)$-dp algorithm $\alpha = \Omega(\varepsilon^{-1} \ln T)$.

• Also applies to answering threshold queries
  • Dataset $x \in [T]^n$
  • Queries $c_t(x) = \#\{i : x_i \geq t\}$
  • Goal: output $(a_1, \ldots, a_T)$ such that
    $\max_t |a_t - c_t(x)| \leq \alpha$
Packing vs. Covering

$X^n$: all datasets

$R$: all outputs

- packing: set of datasets; no output is "good" for two datasets
- covering: set of outputs; for every dataset, some output is "good"
Final Thought: Packing vs. Covering

• Suppose we have a function $f : X^n \to R$

• Suppose we have a covering $C$ such that for every $x$, there exists $c \in C$, such that $d(f(x), c) \leq \alpha$.
  • Some accuracy metric $d$.

• **Thm**: Exists an $(\varepsilon, 0)$-dp algorithm with error $\beta = \alpha + \frac{\ln|C|}{\varepsilon n}$.
  • If $n = \Omega \left( \frac{\ln|C|}{\varepsilon \alpha} \right)$, then we get error $\beta = O(\alpha)$

• **Thm**: size of minimum covering $\approx$ size of maximum packing
  • Implies LB of $n = \Omega \left( \frac{\ln|C|}{\varepsilon} \right)$; tight up to $O \left( \frac{1}{\alpha} \right)$ factor
Outline

- **Packing arguments** for DP lower bounds
  - Intuitive geometric approach to lower bounds
  - Applicable to a wide variety of problems
  - Often yields tight lower bounds for $(\varepsilon, 0)$-dp
  - Separates $(\varepsilon, 0)$-dp from $(\varepsilon, \delta)$-dp