Private Hypothesis Selection

Mark Bun
Gautam Kamath
Thomas Steinke
Zhiwei (Steven) Wu

Boston University
University of Waterloo
IBM Research – Almaden
University of Minnesota
This Talk in One Slide

**Input:** Known collection of distributions \( H = \{h_1, \ldots, h_m\} \)

\[ D = \text{i.i.d. samples } x_1, \ldots, x_n \text{ from unknown } p \]

**Goal:** Find a hypothesis \( h \in H \) which is “close” to \( p \) in total variation distance while protecting privacy of \( D \)

**Our results:**
New algorithms with sample complexity competitive with the best *non-private* algorithms

**Applications:** Private distribution learning, complexity of private mean estimation under product vs. non-product distributions
Privacy-Preserving Data Analysis

Want curators that are:  
- Private  
- Statistically useful
Differential Privacy

[Dinur-Nissim03, Dwork-Nissim04, Blum-Dwork-McSherry-Nissim05]
[Dwork-McSherry-Nissim-Smith06]

\[D \rightarrow M \rightarrow \text{Outcome}\]

Outcome of \( M \) should not depend “too much” on any individual.
Differential Privacy

[Dinur-Nissim03, Dwork-Nissim04, Blum-Dwork-McSherry-Nissim05]
[Dwork-McSherry-Nissim-Smith06]

$D$ and $D'$ are neighbors if they differ on one row.

$M$ is differentially private if for all neighbors $D$, $D'$:

\[
\text{Distribution of } M(D) \approx \text{Distribution of } M(D')
\]
Differential Privacy

[Dinur-Nissim03, Dwork-Nissim04, Blum-Dwork-McSherry-Nissim05]
[Dwork-McSherry-Nissim-Smith06]

\[ D \text{ and } D' \text{ are neighbors if they differ on one row} \]

\[ D \]
\[ \begin{array}{c}
  x_1 \\
  x_2 \\
  \vdots \\
  x_n \\
\end{array} \]

\[ M \]

\[ D' \]
\[ \begin{array}{c}
  x_1 \\
  x_2' \\
  \vdots \\
  x_n \\
\end{array} \]

\[ M \]

\[ M \text{ is } \varepsilon\text{-differentially private if for all neighbors } D, D' \text{ and } T \subseteq \text{Range}(M): \]
\[ \Pr[M(D) \in T] \leq e^\varepsilon \Pr[M(D') \in T] \]

small constant, e.g. \( \varepsilon = 0.1 \)
\[ \Rightarrow e^\varepsilon \approx 1 + \varepsilon \]
Things to Love about Differential Privacy

Resilient to both known and unforeseen attacks

In particular, robust to post-processing

Group privacy

Automatic protection for small groups of individuals

Composition

– $m$-fold composition at worst $m\varepsilon$-DP

– Enables differentially private “programming”
Other Algorithmic Applications

- Privacy-preserving data analysis (duh)
- Algorithmic mechanism design [McSherry-Talwar07, Kearns-Pai-Roth-Ullman12, Nissim-Smorodinsky-Tennenholtz12]
- False discovery control in adaptive data analysis [Dwork-Feldman-Hardt-Pitassi-Reingold-Roth14, Hardt-Ullman14]
- Proofs of concentration inequalities [Steinke-Ullman17, Nissim-Stemmer17]
- Gentle measurement of quantum states [Aaronson-Rothblum19]
Variants of Differential Privacy

M satisfies **insert privacy definition** if for all neighbors $D, D'$

**$\epsilon$-“Pure DP”** [Dwork-McSherry-Nissim-Smith06]

For all $T \subseteq \text{Range}(M)$: $\Pr[M(D) \in T] \leq e^{\epsilon} \Pr[M(D') \in T]

Equivalently, “privacy loss” always $\leq \epsilon$

**$\epsilon$-“Concentrated DP”** [Dwork-Rothblum12, B.-Steinke16]

“Privacy loss” is subgaussian with standard dev. $\leq \epsilon$

**$(\epsilon, \delta)$-“Approximate DP”** [Dwork-Kenthapadi-McSherry-Mironov-Naor06]

For all $T \subseteq \text{Range}(M)$: $\Pr[M(D) \in T] \leq e^{\epsilon} \Pr[M(D') \in T] + \delta$

Equivalently, “privacy loss” $\leq \epsilon$ except with prob. $\leq \delta$
Variants of Differential Privacy

$\varepsilon$-DP
Basic Composition
Laplace Noise
Randomized Response
Exponential Mechanism
Sparse Vector

$\varepsilon$-CDP
Advanced Composition
Gaussian Noise
Projection Mechanism

$(\varepsilon, \delta)$-DP
Truncated Laplace Noise
PTR/Stability
Smooth Sensitivity

Less stringent privacy
= more algorithmic techniques
= harder to prove lower bounds
(Privately) Answering Attribute Means

\[ d \text{ binary attributes} \]

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td></td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

\[ \frac{3}{4} + \text{Noise}(\quad) \]
(Privately) Answering Attribute Means

$d$ binary attributes

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

3/4 + Noise(1/εn)

(To get $\alpha$-error, need $n \geq 1/\alpha \varepsilon$)
(Privately) Answering Attribute Means

\[ d \text{ binary attributes} \]

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

\[ \frac{1}{2} + \text{Noise}(\ ) + \text{Noise}(\ ) + \text{Noise}(\ ) + \text{Noise}(\ ) \]

With pure differential privacy
(Privately) Answering Attribute Means

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

\[
\frac{1}{2} + \text{Noise}(\frac{d}{\varepsilon n}) + \text{Noise}(\frac{d}{\varepsilon n}) + \text{Noise}(\frac{d}{\varepsilon n}) + \text{Noise}(\frac{d}{\varepsilon n})
\]

(To get $\alpha$-error per query, need $n \geq \frac{d}{\alpha \varepsilon}$)

With pure differential privacy
(Privately) Answering Attribute Means

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

\[
d \text{ binary attributes}
\]

\[
\frac{1}{2} + \text{Noise}(\frac{d^{1/2}}{\varepsilon n}) + \text{Noise}(\frac{d^{1/2}}{\varepsilon n}) + \text{Noise}(\frac{d^{1/2}}{\varepsilon n}) + \text{Noise}(\frac{d^{1/2}}{\varepsilon n})
\]

(To get \( \alpha \)-error per query, need \( n \geq \frac{d^{1/2}}{\alpha \varepsilon} \))

With concentrated or approximate differential privacy
Outline of This Talk

• Problem: Differentially private hypothesis selection

• Algorithms
  • (The path to) a basic algorithm
  • A semi-agnostic algorithm
  • Exploiting combinatorial structure

• Applications
  • Privately learning Gaussians
  • Product vs. non-product distributions
The Problem: Hypothesis Selection

Input: Known collection of distributions $H = \{h_1, \ldots, h_m\}$

\[ D = \text{i.i.d. samples } x_1, \ldots, x_n \text{ from unknown } p \]

Goal: If there exists $h^* \in H$ such that $\text{TV}(p, h^*) \leq \alpha$, w.h.p. output $h \in H$ such that $\text{TV}(p, h) \leq O(\alpha)$
The Problem: Hypothesis Selection

**Input:** Known collection of distributions $H = \{h_1, \ldots, h_m\}$

$D =$ i.i.d. samples $x_1, \ldots, x_n$ from unknown $p$

**Goal:** If there exists $h^* \in H$ such that $\text{TV}(p, h^*) \leq \alpha$, w.h.p. output $h \in H$ such that $\text{TV}(p, h) \leq O(\alpha)$

**Theorem:** Achievable using $n = O(\log m / \alpha^2)$ samples (non-privately)
Non-Private Solution: “Scheffé Tournament”

[Iatracos85, Devroye-Lugosi01]

Idea: Set up $\binom{m}{2}$ pairwise contests between candidates, and output candidate which won the most contests

Contest subroutine: To compare distributions $h$, $h'$:

Define Scheffé set $S = \{x : h(x) > h'(x)\}$

Let $h(S) =$ probability mass $h$ places on $S$

$h'(S) =$ probability mass $h'$ places on $S$

$D(S) =$ fraction of $D$ which lands in $S$

$h$ wins if $|h(S) - D(S)| < |h'(S) - D(S)|$;

otherwise $h'$ wins
Scheffé Tournament Analysis
[Yatraco85, Devroye-Lugosi01]

**Theorem:** Achievable using \( n = O(\log \frac{m}{\alpha^2}) \) samples

**Lemma:** If \( h \) wins against \( h' \), then
\[
TV(h, p) \leq 3 \min\{TV(h, p), TV(h', p)\} + 4 |p(S) - D(S)|
= \text{err}
\]

Suppose \( \text{err} \leq \alpha \) for all \( \binom{m}{2} \) pairwise contests simultaneously

Divide \( H \) into 4 quality tiers:
- \( T_1: TV(h, p) \leq \alpha \)
- \( T_2: TV(h, p) \in (\alpha, 4\alpha] \)
- \( T_3: TV(h, p) \in (4\alpha, 12\alpha] \)
- \( T_4: TV(h, p) > 12\alpha \)

By Lemma,
- Every \( h \in T_1 \) has \( \geq |T_3| + |T_4| \) wins
- Every \( h \in T_4 \) has \( \geq |T_1| + |T_2| \) losses

Hence a \( T_4 \) hypothesis is never selected.
Towards a Private Tournament

A First Attempt: Noisy Pairwise Contests

To compare distributions \( h, h' \):

Define Scheffé set \( S = \{ x : h(x) > h'(x) \} \)

Let \( h(S) = \) probability mass \( h \) places on \( S \)

\( h'(S) = \) probability mass \( h' \) places on \( S \)

\( D(S) = \) fraction of \( D \) which lands in \( S \)

\[ \hat{D}(S) = D(S) + \text{Lap} \left( \frac{m}{2\varepsilon} \right) \]

\( h \) wins if \( |h(S) - \hat{D}(S)| < |h'(S) - \hat{D}(S)| \);
otherwise \( h' \) wins.
Analysis of First Attempt

Lemma: If \( h \) wins against \( h' \), then
\[
TV(h, p) \leq 3 \min\{TV(h, p), TV(h', p)\} + 4 |p(S) - D(S)| = err
\]

By previous analysis, select a good hypothesis as long as \( err \leq \alpha \) for all \( \binom{m}{2} \) pairwise contests simultaneously

\[
|p(S) - \hat{D}(S)| \leq |p(S) - D(S)| + |D(S) - \hat{D}(S)|
\]

Theorem: Private hypothesis selection is possible using
\[
n = O \left( \frac{\log m}{\alpha^2} + \frac{m^2 \log m}{\alpha \varepsilon} \right)
\]
Improving the First Attempt

Theorem: Private hypothesis selection is possible using

\[ n = O \left( \frac{\log m}{\alpha^2} + \frac{m^2 \log m}{\alpha \varepsilon} \right) \]
samples

- Relaxing to concentrated or approximate DP lets us use Gaussian noise and “advanced” composition, bringing the second term to \( \frac{m \sqrt{\log m}}{\alpha \varepsilon} \)

- Can possibly be further improved using more efficient tournaments making \( \tilde{O}(m) \) comparisons [Acharya-Jafarpour-Orlitsky-Suresh14, Daskalakis-Kamath14…] to something like \( \tilde{O} \left( \frac{\sqrt{m}}{\alpha \varepsilon} \right) \)

Still an exponential “price of privacy”
A Second (and Final) Attempt: Private Discrete Optimization

Given: An objective function $q : X^n \times H \rightarrow \mathbb{R}$

Private dataset $D = (x_1, \ldots, x_n)$

Output: $h \in H$ which approximately maximizes $q(D, h)$

Exponential Mechanism [McSherry-Talwar07]

Sample $h \in H$ with probability $\propto \exp \left( \frac{\varepsilon q(D, h)}{2\Delta} \right)$

where $\Delta = \sup_{h \in H, D \sim D'} |q(D, h) - q(D', h)|$

“Sensitivity” of the objective function $q$
Private Discrete Optimization

**Exponential Mechanism** [McSherry-Talwar07]

Sample $h \in H$ with probability $\propto \exp\left(\frac{\varepsilon q(D, h)}{2\Delta}\right)$

where $\Delta = \sup_{h \in H, D \sim D'} |q(D, h) - q(D', h)|$

Claim 1: Guarantees $\varepsilon$-differential privacy

Claim 2: W.h.p. produces $h \in H$ with

$$q(D, h) \geq \text{OPT} - O\left(\frac{\Delta \log |H|}{\varepsilon}\right)$$
Instantiating the Exponential Mechanism

Sample $h \in H$ w.p. $\propto \exp \left( \frac{\varepsilon q(D, h)}{2\Delta} \right)$

where $\Delta = \sup_{h \in H, D \sim D'} |q(D, h) - q(D', h)|$

- $\varepsilon$-DP
- Error $O(\Delta \log |H| / \varepsilon)$

How to choose $q$?

**Attempt 2.1:** $q(D, h) = \#$ contests won by $h$

**Problem:** Very high sensitivity $\Delta = m - 1$

**Attempt 2.2:** $q(D, h) = \min \#$ of samples in $D$ that must be changed before $h$ loses at least one contest

Sensitivity 1! 😊 By how to ensure OPT is good? 😕
Instantiating the Exponential Mechanism

**Attempt 2.3:** (Really the final one, I swear)

\[ q(D, h) = \min \# \text{ of samples in } D \text{ that must be changed before } h \text{ loses at least one contest} \]

Pairwise contest with draws: To compare distributions \( h, h' \):

[Daskalakis-Diakonikolas-Servedio11, Daskalakis-Kamath14]

If \( h(S) - h'(S) < 6\alpha \):

Declare “Draw”

Else if \( D(S) > h(S) - 3\alpha \):

Declare \( h \) as winner

Else if \( D(S) < h'(S) + 3\alpha \):

Declare \( h' \) as winner

Else:

 Declare “Draw”
Instantiating the Exponential Mechanism

**Attempt 2.3:** (Really the final one, I swear)

\[ q(D, h) = \min \# \text{ of samples in } D \text{ that must be changed before } h \text{ loses at least one contest} \]

**Pairwise contest with draws**
[Daskalakis-Diakonikolas-Servedio11, Daskalakis-Kamath14]

**Main Lemma:** Suppose there exists \( h^* \in H \) with \( \text{TV}(p, h^*) \leq \alpha \). Let \( D = (x_1, \ldots, x_n) \) i.i.d. from \( p \) for \( n = O(\log m / \alpha^2) \). Then w.h.p.,

1) \( q(D, h^*) > \alpha n \) and \hspace{1cm} (completeness)

2) \( q(D, h) = 0 \) for every \( h \) where \( \text{TV}(p, h) > 7\alpha \) \hspace{1cm} (soundness)
Completing the Analysis

Exponential Mechanism with sensitivity-1 score

Sample $h \in H$ w.p. $\propto \exp \left( \frac{q(D, h)}{2\varepsilon} \right)$

- $\varepsilon$-DP

- W.h.p. outputs $h$ with $q(D, h) \geq \text{OPT} - O\left( \frac{\log m}{\varepsilon} \right)$

Main Lemma: Suppose there exists $h^* \in H$ with $\text{TV}(p, h^*) \leq \alpha$. Let $D = (x_1, \ldots, x_n)$ i.i.d. from $p$ for $n = O(\log m / \alpha^2)$. Then w.h.p.,

1) $q(D, h^*) > \alpha n$ and
2) $q(D, h) = 0$ for every $h$ where $\text{TV}(p, h) > 7\alpha$

(completeness) (soundness)

- $\text{OPT} = q(D, h^*) > \alpha n$ by 1), assuming $n \geq O(\log m / \alpha^2)$

- EM outputs $h$ with $q(D, h) > \alpha n - O(\log m / \varepsilon) > 0$
  assuming $n \geq O(\log m / \alpha \varepsilon)$

- Conclude $\text{TV}(p, h) \leq 7\alpha$ by 2), assuming $n \geq O(\log m / \alpha^2)$
Completing the Analysis

Exponential Mechanism with sensitivity-1 score
Sample $h \in H$ w.p. $\propto \exp \left( \frac{q(D, h)}{2\varepsilon} \right)$

- $\varepsilon$-DP
- W.h.p. outputs $h$ with $q(D, h) \geq \text{OPT} - O \left( \frac{\log m}{\varepsilon} \right)$

Main Lemma: Suppose there exists $h^* \in H$ with $\text{TV}(p, h^*) \leq \alpha$. Let $D = (x_1, \ldots, x_n)$ i.i.d. from $p$ for $n = O(\log m / \alpha^2)$. Then w.h.p.,

1) $q(D, h^*) > \alpha n$ and (completeness)
2) $q(D, h) = 0$ for every $h$ where $\text{TV}(p, h) > 7\alpha$ (soundness)

Theorem: There is an $\varepsilon$-DP algorithm such that, if there exists $h^* \in H$ with $\text{TV}(p, h^*) \leq \alpha$, the algorithm outputs $h \in H$ with $\text{TV}(p, h) \leq 7\alpha$ w.h.p. as long as

$$n \geq O \left( \frac{\log m}{\alpha^2} + \frac{\log m}{\alpha \varepsilon} \right)$$
Outline of This Talk

• Problem: Differentially private hypothesis selection

• Algorithms
  • (The path to) a basic algorithm
  • A semi-agnostic algorithm
  • Exploiting combinatorial structure

• Applications
  • Privately learning Gaussians
  • Product vs. non-product distributions
Obtaining a Semi-Agnostic Algorithm

**Input:** Known collection of distributions \( H = \{ h_1, \ldots, h_m \} \)

\( D = \) i.i.d. samples \( x_1, \ldots, x_n \) from unknown \( p \)

**Goal:** If there exists \( h^* \in H \) such that \( TV(p, h^*) \leq \alpha \), w.h.p. output \( h \in H \) such that \( TV(p, h) \leq O(\alpha) \)
Obtaining a Semi-Agnostic Algorithm

Input: Known collection of distributions \( H = \{h_1, \ldots, h_m\} \)
\[ D = \text{i.i.d. samples } x_1, \ldots, x_n \text{ from unknown } p \]

Goal: Let \( \text{OPT} = \arg\min_h \text{TV}(p, h) \)
\[ \text{w.h.p. output } h \in H \text{ such that } \text{TV}(p, h) \leq O(\text{OPT}) + \alpha \]
Obtaining a Semi-Agnostic Algorithm

**Problem:** Even the non-private analysis of pairwise contest with draws seems to require \( \text{OPT} \leq \alpha \)

**Solution:**

1) Run algorithm of **Attempt 2.3** \( T = \log(1/\alpha) \) times, starting with \( \alpha_1 = \alpha, \alpha_2 = 2\alpha, \ldots, \alpha_T = 1 \) producing hypotheses \( h_1, \ldots, h_T \)

2) Use algorithm of **Attempt 1** to semi-agnostically select the best of \( h_1, \ldots, h_T \)

Final sample complexity bound is the same as Attempt 2.3, up to additive \( \log^2(1/\alpha) / \alpha \varepsilon \)
Outline of This Talk

• Problem: Differentially private hypothesis selection

• Algorithms
  • (The path to) a basic algorithm
  • A semi-agnostic algorithm
  • Exploiting combinatorial structure

• Applications
  • Privately learning Gaussians
  • Product vs. non-product distributions
Exploiting the Structure of $H$

For a set $H$ of distributions, define $\text{VC}(H)$ to be the VC dimension of the collection of Scheffé sets $S_{h,h'} = \{ x : h(x) > h'(x) \}$

For a distribution $p$, let $N_\alpha(p, H) = |\{ h \in H : \text{TV}(p, h) < \alpha \}|$

\underline{Theorem:} There is an $\varepsilon$–DP algorithm such that, if there exists $h^* \in H$ with $\text{TV}(p, h^*) \leq \alpha$, the algorithm outputs $h \in H$ with $\text{TV}(p, h) \leq 7\alpha$ w.h.p. as long as

$$n \geq O \left( \frac{\log m}{\alpha^2} + \frac{\log m}{\alpha \varepsilon} \right)$$
Exploiting the Structure of $\mathcal{H}$

For a set $\mathcal{H}$ of distributions, define $\text{VC}(\mathcal{H})$ to be the VC dimension of the collection of Scheffé sets $S_{h,h'} = \{ x : h(x) > h'(x) \}$

For a distribution $p$, let $N_\alpha(p, \mathcal{H}) = |\{ h \in \mathcal{H} : \text{TV}(p, h) < \alpha \}|$

**Theorem:** There is an $(\varepsilon, \delta)$-DP algorithm such that, if there exists $h^* \in \mathcal{H}$ with $\text{TV}(p, h^*) \leq \alpha$, the algorithm outputs $h \in \mathcal{H}$ with $\text{TV}(p, h) \leq 7\alpha$ w.h.p. as long as

$$n \geq O \left( \frac{\text{VC}(\mathcal{H})}{\alpha^2} + \frac{\log N_{7\alpha}(p, \mathcal{H}) + \log(1/\delta)}{\alpha\varepsilon} \right)$$

1) $\log m \rightarrow \text{VC}(\mathcal{H})$: Replace Chernoff + union with uniform convergence
2) $\log m \rightarrow N(p, \mathcal{H})$: Exploit *stability* to pay only for hypotheses with score $> 0$
Outline of This Talk

• Problem: Differentially private hypothesis selection

• Algorithms
  • (The path to) a basic algorithm
  • A semi-agnostic algorithm
  • Exploiting combinatorial structure

• Applications
  • Privately learning Gaussians
  • Product vs. non-product distributions
Problem: Let $p = \mathcal{N}(\mu, \Sigma)$ be an unknown $d$-dimensional Gaussian. Given i.i.d. samples $D = (x_1, \ldots, x_n)$ from $p$, privately find a Gaussian $h$ with $\text{TV}(p, h) \leq \alpha$. 
Privately Learning Gaussians

Problem: Let $p = \mathcal{N}(\mu, \Sigma)$ be an unknown $d$-dimensional Gaussian. Given i.i.d. samples $D = (x_1, \ldots, x_n)$ from $p$, privately find a Gaussian $h$ with $\text{TV}(p, h) \leq \alpha$.

Generic statistical estimation frameworks

[Dwork-Lei09, Smith11]

Private confidence intervals for univariate Gaussians

[Karwa-Vadhan18]

Private means/covariances of multivariate Gaussians

[Kamath-Li-Singhal-Ullman19]

Univariate mean estimation via smooth sensitivity

[Nissim-Raskhodnikova-Smith07, B.-Steinke19]
Privately Learning Gaussians

Problem: Let \( p = \mathcal{N}(\mu, \text{Id}) \) be an unknown \( d \)-dimensional Gaussian with \( \|\mu\|_2 \leq M \). Given i.i.d. samples \( D = (x_1, \ldots, x_n) \) from \( p \), privately find a Gaussian \( h \) with \( \text{TV}(p, h) \leq O(\alpha) \).

Solution via Private Hypothesis Selection

1) Construct a finite cover \( H \) of \( \{ \mathcal{N}(\mu, \text{Id}) : \|\mu\|_2 \leq M \} \) w.r.t. \( \text{TV} \) distance. I.e., construct set of Gaussians \( H \) such that for every \( \mu \) with \( \|\mu\|_2 \leq M \) there exists \( h \in H \) with \( \text{TV}(\mathcal{N}(\mu, \text{Id}), h) < \alpha \).

2) Apply Attempt 2.3 using the cover \( H \), incurring error \( O(\alpha) \) with sample complexity

\[
n = O \left( \frac{\text{VC}(H)}{\alpha^2} + \frac{\log |H|}{\alpha \varepsilon} \right)
\]

What’s the VC dimension? How big does the cover need to be?
Privately Learning Gaussians

**Problem:** Let \( p = \mathcal{N}(\mu, \text{Id}) \) be an unknown \( d \)-dimensional Gaussian with \( \|\mu\|_2 \leq M \). Given i.i.d. samples \( D = (x_1, \ldots, x_n) \) from \( p \), privately find a Gaussian \( h \) with \( \text{TV}(p, h) \leq O(\alpha) \).

**VC Dimension of Gaussians**

Scheffé sets are halfspaces, which have VC-dimension \( d+1 \)

**Covering Gaussians**

**Lemma:** \( \text{TV}(\mathcal{N}(\mu, \text{Id}), \mathcal{N}(\mu', \text{Id})) \leq \|\mu - \mu'\|_2 \)

\( \Rightarrow \) Suffices to cover \( l_2 \)-ball of radius \( M \) using balls of radius \( \alpha \)

Can be done using a cover of size \( \approx \left( \frac{\sqrt{dM}}{\alpha} \right)^d \)
Privately Learning Gaussians

Problem: Let \( p = \mathcal{N}(\mu, \text{Id}) \) be an unknown \( d\)-dimensional Gaussian with \( \|\mu\|_2 \leq M \). Given i.i.d. samples \( D = (x_1, \ldots, x_n) \) from \( p \), privately find a Gaussian \( h \) with \( \text{TV}(p, h) \leq O(\alpha) \).

Solution via Private Hypothesis Selection

1) Construct a size-\( \left( \frac{\sqrt{dM}}{\alpha} \right)^d \) cover \( H \) of \( \{ \mathcal{N}(\mu, \text{Id}) : \|\mu\|_2 \leq M \} \) w.r.t. TV distance. I.e., construct set of Gaussians \( H \) such that for every \( \mu \) with \( \|\mu\|_2 \leq M \) there exists \( h \in H \) with \( \text{TV}(\mathcal{N}(\mu, \text{Id}), h) < \alpha \). 

2) Apply Attempt 2.3 using the cover \( H \), incurring error \( O(\alpha) \) with sample complexity

\[
n = O \left( \frac{d}{\alpha^2} + \frac{d}{\alpha \varepsilon} \log \left( \frac{dM}{\alpha} \right) \right)
\]
Other Applications
of “Cover-and-Select”

• Other variants of Gaussian estimation
  Unbounded means, unknown covariances, etc.
• Discrete product distributions
• Piecewise polynomials
• Sums of Independent Integer Random Variables (SIIRVs)
• Poisson Multinomial distributions
Product vs. Non-Product Distributions

**Definition:** A \((k, d)\)-product distribution is a product distribution over \([k]^d\).

**Lemma:** The set of \((k, d)\)-product distributions admits an \(\alpha\)-cover of size \(\approx \left(\frac{kd}{\alpha}\right)^{d(k-1)}\).

So using cover-and-select, we get an \(\varepsilon\)-DP algorithm for learning \((k, d)\)-product distributions to TV distance \(\alpha\) with sample complexity

\[
    n = \tilde{O}\left(\frac{kd}{\alpha^2} + \frac{kd}{\alpha \varepsilon}\right)
\]

Set \(k = 2, \alpha = 1/2\).
Product vs. Non-Product Distributions

**Definition:** A \((k, d)\)-product distribution is a product distribution over \([k]^d\).

**Lemma:** The set of \((k, d)\)-product distributions admits an \(\alpha\)-cover of size \(\approx \left(\frac{kd}{\alpha}\right)^{d(k-1)}\).

So using cover-and-select, we get an \(\varepsilon\)-DP algorithm for learning product distributions over \(\{0, 1\}^d\) to TV distance \(1/2\) with sample complexity

\[ n = \tilde{O} \left( \frac{d}{\varepsilon} \right) \]
Product vs. Non-Product Distributions

**Definition:** A \((k, d)\)-product distribution is a product distribution over \([k]^d\).

**Lemma:** The set of \((k, d)\)-product distributions admits an \(\alpha\)-cover of size \(\approx \left(\frac{kd}{\alpha} \right)^{d(k-1)}\).

So using cover-and-select, we get an \(\varepsilon\)-DP algorithm for learning the mean of a product distribution over \(\{0, 1\}^d\) to \(l_1\) distance \(2\sqrt{d}\) with sample complexity

\[
n = \tilde{O} \left( \frac{d}{\varepsilon} \right)
\]
### (Privately) Answering Attribute Means

#### $d$ binary attributes

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

$$\frac{1}{2} + \text{Noise}(\frac{d}{\epsilon n}) + \frac{1}{2} + \text{Noise}(\frac{d}{\epsilon n}) + \frac{3}{4} + \text{Noise}(\frac{d}{\epsilon n}) + \frac{1}{4} + \text{Noise}(\frac{d}{\epsilon n})$$

(To get $\alpha$-error per query, need $n \geq \frac{d}{\alpha \epsilon}$)  

[Hardt-Talwar10]

With pure differential privacy
(Privately) Answering Attribute Means

\[ d \text{ binary attributes} \]

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

\[ \frac{1}{2} + \text{Noise}(\frac{d}{\varepsilon n}) + \text{Noise}(\frac{d}{\varepsilon n}) + \frac{3}{4} + \text{Noise}(\frac{d}{\varepsilon n}) + \frac{1}{4} + \text{Noise}(\frac{d}{\varepsilon n}) \]

(To get \( L_1 \) distance \( 2\sqrt{d} \), \textbf{need} \( n \geq d^{3/2}/\varepsilon \))

[Hardt-Talwar10]

Compare to only \( d/\varepsilon \) for product distributions

With pure differential privacy
Conclusions

• New algorithms for private hypothesis selection with minimal “cost of privacy”
• Applications: Private distribution learning, complexity of privacy under product vs. non-product distributions

Open Questions:
• Combinatorial characterization of private (and non-private!) sample complexity
• Computationally efficient algorithms for sample-optimal Gaussian mean estimation
• Deeper understanding of complexity of estimation under product distributions

Thank you!