Privately Learning High-Dimensional Distributions

Gautam Kamath
Simons Institute → University of Waterloo
Data Privacy: From Foundations to Applications
March 8, 2019

With:
Jerry Li (Microsoft Research Redmond)
Vikrant Singhal (Northeastern University)
Jonathan Ullman (Northeastern University)
Algorithms vs. Statistics

Algorithms

\[
\begin{align*}
X_1 \\
X_2 \\
\vdots \\
X_n
\end{align*}
\]

\[ Y \rightarrow M \]

“utility”

\[ U(X, Y) \]

Statistics

\[
\begin{align*}
X_1 \\
X_2 \\
\vdots \\
X_n
\end{align*}
\]

\[ Y \rightarrow M \]

“utility”

\[ U(p, Y) \]

Distribution \( p \)

random sampling
Privacy in Statistics

Desiderata:

1. Algorithm is accurate (with high probability over $X \sim p$)
   - May require assumptions about $p$ to hold
   - Today: “Estimate” $p$

2. Algorithm is private (always)
   - Today: $\frac{\varepsilon^2}{2}$-concentrated differential privacy

What is the additional cost of privacy?
An Example

• Given female heights $X_1, \ldots, X_n$, compute the average height
  • $X_i \sim i.i.d. \ D$, compute $E[D]$
• Laplace Mechanism
  • $Z = \sum X_i + \text{Laplace} \left( \frac{\Delta}{\varepsilon} \right)$
An Example

- Given female heights $X_1, \ldots, X_n$, compute the average height
  - $X_i \sim i.i.d. \ D$, compute $E[D]$
- Laplace Mechanism
  - $Z = \sum X_i + \text{Laplace}\left(\frac{\Delta}{\varepsilon}\right)$
  - $\Delta = \text{realmax}$!
An Example

- Given female heights $X_1, \ldots, X_n$, compute the average height
  - $X_i \sim_{i.i.d.} D$, compute $E[D]$
- Laplace Mechanism
  - $Z = \sum X_i + \text{Laplace}\left(\frac{\Delta}{\varepsilon}\right)$
- A priori: most females between 120 cm and 200 cm
  - Clip/“Winsorize” data, $\Delta = 80$
  - $80/\varepsilon$ is still large...
- Things get worse in high dimensions
- Goal: Minimize cost due to uncertainty
Background: Univariate Private Statistics

• Theorem: There exists a $\frac{\varepsilon^2}{2}$-zCDP algorithm which estimates the mean of a Bernoulli distribution up to $\pm \alpha$, with $n = O \left( \frac{1}{\alpha^2} + \frac{1}{\alpha \varepsilon} \right)$ samples.
  • “Rate”: $|p - \hat{p}| \leq O \left( \frac{1}{\sqrt{n}} + \frac{1}{\varepsilon n} \right)$
  • Non-private cost: $O \left( \frac{1}{\alpha^2} \right)$ samples

• Low-dimensional problems are now (reasonably) well-understood
  • Univariate Gaussians [Karwa-Vadhan ’18]
  • Univariate discrete distributions
    • Kolmogorov distance [Bun-Nissim-Stemmer-Vadhan ’15]
    • Total variation distance [folklore, Diakonikolas-Hardt-Schmidt ’15]

• High dimensions?
Results: Multivariate Private Statistics

• Theorem: There exists a $\frac{\varepsilon^2}{2}$-zCDP algorithm
Results: Multivariate Private Statistics

• Theorem: There exists a $\frac{\varepsilon^2}{2}$-zCDP algorithm which learns a Gaussian $N(\mu, \Sigma)$ in $\mathbb{R}^d$
Results: Multivariate Private Statistics

• Theorem: There exists a $\frac{\varepsilon^2}{2}$-zCDP algorithm which learns a Gaussian $N(\mu, \Sigma)$ in $\mathbb{R}^d$ with $\|\mu\|_2 \leq R$ and $I \preceq \Sigma \preceq \kappa I$
Results: Multivariate Private Statistics

• Theorem: There exists a $\frac{\varepsilon^2}{2}$-zCDP algorithm which learns a Gaussian $N(\mu, \Sigma)$ in $\mathbb{R}^d$ with $\|\mu\|_2^2 \leq R$ and $I \preceq \Sigma \preceq \kappa I$ to $\alpha$ total variation distance with

$$n = \tilde{O} \left( \frac{d^2}{\alpha^2} + \frac{d^2}{\alpha \varepsilon} + \frac{d^{3/2} \log^{1/2} \kappa}{\varepsilon} + \frac{d^{1/2} \log^{1/2} R}{\varepsilon} \right)$$

samples.

• Non-private: $O\left(\frac{d^2}{\alpha^2}\right)$ samples – exponent in $d$ unchanged

• Mild dependence on “uncertainty” parameters $R, \kappa$

• Some lower bounds

• Similar results for product distributions: $n = \tilde{\Theta} \left( \frac{d}{\alpha^2} + \frac{d}{\alpha \varepsilon} \right)$ samples
Today’s talk: Gaussian Covariance Estimation

• Theorem: There exists a \( \frac{\varepsilon^2}{2} \)-zCDP algorithm which learns a Gaussian \( N(0, \Sigma) \) in \( \mathbb{R}^d \) with \( I \preceq \Sigma \preceq \kappa I \) to \( \alpha \) total variation distance with

\[
    n = \tilde{O} \left( \frac{d^2}{\alpha^2} + \frac{d^2}{\alpha \varepsilon} + \frac{d^{3/2} \log^{1/2} \kappa}{\varepsilon} \right)
\]
samples.

• If \( \Sigma \) were well-conditioned (\( \kappa = O(1) \)), problem is easy

• A private recursive method to reduce the condition number
Learning a Multivariate Gaussian

Given samples from
\[ N(0, \Sigma), I \preceq \Sigma \preceq \kappa I, \]
output \( \hat{\Sigma} \), such that
\[ \| \Sigma - \hat{\Sigma} \|_{\Sigma} \leq \alpha \]
\[ \iff \]
\[ \| \Sigma^{-1/2} \hat{\Sigma} \Sigma^{-1/2} - I \|_{F} \leq \alpha. \]

Implies
\[ \text{TV} \left( N(0, \Sigma), N(0, \hat{\Sigma}) \right) = O(\alpha). \]
Non-Private Covariance Estimation

• Given: \( X_1, \ldots, X_n \sim N(0, \Sigma) \)

• Output: \( \hat{\Sigma} = \frac{1}{n} \sum_i X_i X_i^T \)

• Accuracy: \( \| \hat{\Sigma} - \Sigma \|_\Sigma = O \left( \sqrt{\frac{d^2}{n}} \right) \)
  
  • Learn in TV distance with \( n = O(d^2/\alpha^2) \)

• How to privatize?
Recap: Gaussian Mechanism

• $f : D^n \to \mathbb{R}$

• Sensitivity: $\Delta = \max_{X, X' : d_h(X, X') = 1} |f(X) - f(X')|$
  • Biggest difference on two neighboring datasets

• $\hat{f}(X) = f(X) + N\left(0, \left(\frac{\Delta}{\varepsilon}\right)^2\right)$

• Privacy: $\hat{f}$ is $\frac{\varepsilon^2}{2}$-zCDP

• Accuracy: $|\hat{f}(X) - f(X)| = O\left(\frac{\Delta}{\varepsilon}\right)$
Recap: Gaussian Mechanism

• \( f : D^n \rightarrow \mathbb{R}^{d \times d} \)

• Sensitivity: \( \Delta = \max_{X, X': d_h(X, X') = 1} \| f(X) - f(X') \|_F \)
  • Biggest difference on two neighboring datasets

• \( \hat{f}(X) = f(X) + N(0, \left( \frac{\Delta}{\varepsilon} \right)^2)^{d \times d} \)

• Privacy: \( \hat{f} \) is \( \frac{\varepsilon^2}{2} \)-zCDP

• Accuracy: \( \| \hat{f}(X) - f(X) \|_F = O \left( \frac{\Delta d}{\varepsilon} \right) \)
Private Covariance Estimation: Take 1

• Given: \( X_1, \ldots, X_n \sim N(0, \Sigma) \)

• Output:
\[
\hat{\Sigma} = \frac{1}{n} \sum_i X_i X_i^T + N \left( 0, \left( \frac{\Delta}{\epsilon} \right)^2 \right)_{d \times d}
\]

• Accuracy:
\[
\|\hat{\Sigma} - \Sigma\|_{\Sigma} = O \left( \sqrt{\frac{d^2}{n}} + \frac{\Delta d}{\epsilon} \right)
\]

• Problem: What is the sensitivity?
Sensitivity of Empirical Covariance

\[ \| \hat{\Sigma} - \Sigma \|_{\Sigma} = O \left( \frac{d^2}{n} + \frac{\infty \cdot d}{\epsilon} \right) \]

\[ n = O(\infty) \text{ samples!} \]
Limiting Sensitivity via Truncation

\[ l \leq \Sigma \]

\[ \Sigma \leq \kappa l \]
Private Covariance Estimation: Take 2

• “Truncate-then-empirical” method
• Given: $X_1, \ldots, X_n \sim N(0, \Sigma), I \preceq \Sigma \preceq \kappa I$
• Remove points which don’t satisfy $\|X_i\|_2^2 \leq \tilde{O}(\kappa d)$
  • $\Delta = \tilde{O}(\kappa d)$

• Output: $\hat{\Sigma} = \frac{1}{n} \sum_i X_i X_i^T + N \left(0, \left(\frac{\tilde{O}(\kappa d)}{\varepsilon n}\right)^2\right)^{d \times d}$

• Accuracy: $\|\hat{\Sigma} - \Sigma\|_{\Sigma} = \tilde{O} \left(\sqrt{\frac{d^2}{n}} + \frac{\kappa d^2}{\varepsilon n}\right)$
  • $n = \tilde{O} \left(\frac{d^2}{\alpha^2} + \frac{\kappa d^2}{\alpha \varepsilon}\right)$ samples
Private Covariance Estimation, So Far...

- Theorem: There exists a $\frac{\varepsilon^2}{2}$-zCDP algorithm which learns a Gaussian $N(0, \Sigma)$ in $\mathbb{R}^d$ with $I \preceq \Sigma \preceq \kappa I$ to $\alpha$ TV distance with

$$n = \tilde{O} \left( \frac{d^2}{\alpha^2} + \frac{\kappa d^2}{\alpha \varepsilon} \right)$$

- Optimal for $\kappa = O(1)$
- But $\kappa$ can be very large...
What Went Wrong?

If \( n \gg \kappa d^2 \), noise \( Z \) is small...

\[
Z = N\left(0, \left(\frac{\tilde{O}(kd)}{\varepsilon n}\right)^2\right)^{d \times d}
\]

\[
\hat{\Sigma} = \Sigma + Z
\]
What Went Wrong?

If \( n \ll \kappa d^2 \), noise \( Z \) is large...

\[
Z = N \left( 0, \left( \frac{\tilde{O}(kd)}{\varepsilon n} \right)^2 \right)^{d \times d}
\]

\[
\hat{\Sigma} = \Sigma + Z
\]

Goal: Discover and add less noise in “short” directions!

Noised empirical covariance is wrong in “short” directions!
Private Recursive Preconditioning

- In directions where $\Sigma$ is small, our noise outweighed our signal!
- Solution: Approximately learn $\Sigma$ in all directions
- Theorem: There exists a $\frac{\epsilon^2}{2}$-zCDP algorithm which finds a matrix $\hat{A}$ such that $I \preceq \hat{A}\Sigma\hat{A} \preceq 100I$ with $n = \tilde{O}\left(\frac{d^{3/2} \log^{1/2} \kappa}{\epsilon}\right)$ samples.
Preconditioning: An Illustration

\[ \mathcal{N}(0, \Sigma) \quad \text{to} \quad \mathcal{N}(0, \hat{\Sigma} \hat{\Lambda} \hat{\Sigma}) \]

\[ \mathcal{N}(0, \Sigma) \quad \text{to} \quad \mathcal{N}(0, \hat{\Sigma} \hat{\Lambda} \hat{\Sigma}) \]
Private Covariance Estimation: Take 3

• Given: $X_1, \ldots, X_n \sim N(0, \Sigma)$, $I \preceq \Sigma \preceq \kappa I$
1. Learn $\hat{A}$ such that $I \preceq \hat{A}\Sigma\hat{A} \preceq 100I$
2. Let $\bar{\Sigma}$ be output of truncate-then-empirical method on $\hat{A}X_1, \ldots, \hat{A}X_n$
3. Output $\hat{\Sigma} = \hat{A}^{-1}\bar{\Sigma}\hat{A}^{-1}$

• Step 1: $n = \tilde{O} \left( \frac{d^{3/2}\log^{1/2}\kappa}{\varepsilon} \right)$ samples  

• Step 2: $n = \tilde{O} \left( \frac{d^2}{\alpha^2} + \frac{\kappa d^2}{\alpha \varepsilon} \right) = \tilde{O} \left( \frac{d^2}{\alpha^2} + \frac{d^2}{\alpha \varepsilon} \right)$ samples ✔
Recursive Private Preconditioning

• Reduce condition number by a factor of $O(\kappa)$
Recursive Private Preconditioning

• Reduce condition number by a factor of $O(1)$, $O(\log \kappa)$ times!

• Theorem: There exists a $\frac{\varepsilon^2}{2}$-zCDP algorithm which finds a matrix $\hat{A}$ such that $I \preceq \hat{A}\Sigma\hat{A} \preceq \frac{3\kappa}{4} I$ with
  
  $$n = \tilde{O}\left(\frac{d^{3/2}}{\varepsilon}\right)$$
  
  samples.

• Composition of DP: use $O\left(\frac{\varepsilon^2}{\log \kappa}\right)$-zCDP for each round
Recursive Private Preconditioning

If $n \ll \kappa d^2$, noise $Z$ is large...

$$Z = N \left( 0, \left( \frac{O(\kappa d)}{\epsilon n} \right)^2 \right)^{d \times d}$$

$$\hat{\Sigma} = \Sigma + Z$$

Noised empirical covariance is wrong in “short” directions!
Recursive Private Preconditioning

- Recall: \( Z = N \left( 0, \left( \frac{\tilde{\sigma}(\kappa d)}{\epsilon n} \right)^2 \right)^{d \times d} \)
- If \( n = \tilde{\Omega}(d^{3/2}/\epsilon) \), \( \|Z\|_2 \leq \frac{\kappa}{100} \)
- In a given direction:
  - If noised variance is large \( \gg \frac{\kappa}{2} \), true variance is large
    - \( \kappa \) is a good estimate for variance in this direction
  - If noised variance is not large \( \ll \frac{\kappa}{2} \), true variance is not large
    - \( \kappa \) is too large an estimate for variance in this direction – reduce our estimate!
Recursive Private Preconditioning

• Given: $X_1, \ldots, X_n \sim N(0, \Sigma)$, $I \preceq \Sigma \preceq \kappa I$

1. Remove points which don’t satisfy $\|X_i\|_2^2 \leq \tilde{O}(\kappa d)$

2. Compute $\hat{\Sigma} = \frac{1}{n} \sum_i X_iX_i^T + N\left(0, \left(\frac{\tilde{O}(\kappa d)}{\varepsilon n}\right)^2\right)^{d \times d}$

3. Let $(\lambda_i, v_i)$ be eigenvalues/vectors of $\hat{\Sigma}$, $\hat{V} \leftarrow \text{span}\{v_i: \lambda_i \geq \frac{\kappa}{2}\}$

4. Output $\hat{A} \leftarrow \frac{1}{4} \Pi_{\hat{V}} + \Pi_V$

• If $n = \tilde{O}(d^{3/2}/\varepsilon)$, then $I \preceq \hat{A}\Sigma\hat{A} \preceq \frac{3\kappa}{4} I$

• $O(\log \kappa)$ reps: If $n = \tilde{O}(n^{3/2} \log^{1/2} \kappa / \varepsilon)$, then $I \preceq \hat{A}\Sigma\hat{A} \preceq O(1)I$
Results: Multivariate Private Statistics

- Theorem: There exists a $\frac{\varepsilon^2}{2}$-zCDP algorithm which learns a Gaussian $N(\mu, \Sigma)$ in $\mathbb{R}^d$ with $\|\mu\|_2 \leq R$ and $I \preceq \Sigma \preceq \kappa I$ to $\alpha$ TV distance with

$$n = \tilde{O} \left( \frac{d^2}{\alpha^2} + \frac{d^2}{\alpha \varepsilon} + \frac{d^{3/2} \log^{1/2} \kappa}{\varepsilon} + \frac{d^{1/2} \log^{1/2} R}{\varepsilon} \right)$$
samples.
Conclusions

• Algorithm for privately learning Gaussians and product distributions in high dimensions

• First high-dimensional algorithm with mild dependence on “uncertainty parameters”

• Privacy comes at small cost