Robustness Implies Privacy in Statistical Estimation

Spencer Cockerell, Anish Jayant

November 20, 2023

Topics

- Statistical Estimation and privacy
- Exponential Mechanism
- Robust-Private Reduction
- Application to Gaussian Mean Estimation

Topics

- Statistical Estimation and privacy
- Exponential Mechanism
- Robust-Private Reduction
- Application to Gaussian Mean Estimation

Statistical Estimation

- **Central Question**: Given that distribution $\mathcal{D}$ is parametrized by θ, can we produce $\hat{\theta}$ from samples such that $\|\hat{\theta} - \theta\|$ is small?

- **Strategy**: Find *unbiased* estimator F, draw $X_1, \dots, X_n \sim \mathcal{D}$ i.i.d., evaluate $F(\mathbf{X})$ and use concentration bounds
  - $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$ concentrates about μ.

Statistical Estimation

- **Central Question**: Given that distribution $\mathcal{D}$ is parametrized by θ, can we produce $\hat{\theta}$ from samples such that $\|\hat{\theta} - \theta\|$ is small?

- **Strategy**: Find *unbiased* estimator F, draw $X_1, ..., X_n \sim \mathcal{D}$ i.i.d., evaluate $F(\mathbf{X})$ and use concentration bounds
  - $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$ concentrates about μ.

Differential Privacy

- **Intuition**: "Learn nothing about the individual while learning useful information about the population"
  - No individual should change F by 'too much'

Definition (Differential Privacy)

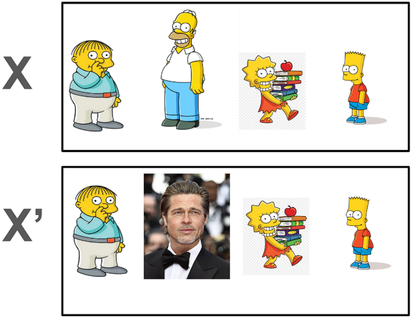
An algorithm $\mathcal{A}$: dataset $\to \Theta$ is (ε, δ)-differentially private if, for all *adjacent datasets* $\mathbf{X} = \{X_1, ..., X_n\}$ and $\mathbf{X}' = \{X_1, ..., X_i', ..., X_n\}$ and any $S \subset \Theta$,

$$\Pr[\mathcal{A}(\mathbf{X}') \in S] \leq e^{\varepsilon} \Pr[\mathcal{A}(\mathbf{X}) \in S] + \delta$$

Is it possible to produce *efficient* and *accurate* statistical estimates that uphold differential privacy?

Private Estimation

- **Intuition**: "Learn nothing about the individual while learning useful information about the population"
  - No individual should change $\mathcal{A}$ by 'too much'



Is it possible to produce *efficient* and *accurate* statistical estimates that uphold privacy?

Operationalizing DP

Task: Given samples $X_1, ..., X_n$ you predict an outcome $r \in \mathcal{R}$. A oracle $q : \mathcal{D}^n \times \mathcal{R} \to \mathbb{R}$ grades your response. Design a differentially private mechanism $\mathcal{M}$ that maximizes q.

Assumption: q is 'well-behaved,' adjacent datasets d, d' satisfy $|q(d', r) - q(d, r)| \leq \Delta q$ for all r.

Operationalizing DP

Task: Given samples $X_1, ..., X_n$ you predict an outcome $r \in \mathcal{R}$. An oracle $q : \mathcal{D}^n \times \mathcal{R} \to \mathbb{R}$ grades your response. Design a differentially private mechanism $\mathcal{M}$ that maximizes q.

Definition (Exponential Mechanism)

Randomized mechanism $\mathcal{M}$, which selects r given dataset d as

$$\Pr[\mathcal{M}(d) = r] \propto \exp(\varepsilon q(d, r))$$

where $\varepsilon \geq 0$.

- What happens with values assigned to adjacent d, d'?

Operationalizing DP

Lemma (Exponential Mechanism)

Randomized mechanism $\mathcal{M}$, which selects r given dataset d as

$$\Pr[\mathcal{M}(d) = r] \propto \exp(\varepsilon q(d, r))$$

satisfies $(2\varepsilon\Delta q, 0)$-DP.

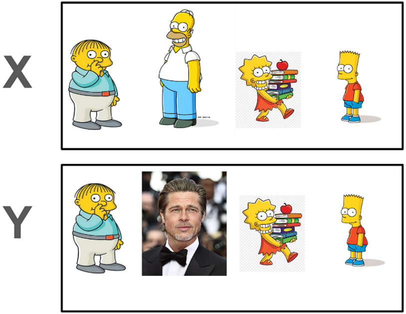
Proof Sketch.

Show $\Pr[\mathcal{M}(d) = r]$ and $\Pr[\mathcal{M}(d') = r]$ can only be $\exp(2\varepsilon\Delta q)$ apart. Result follows by definition. $\square$

Contamination

Definition (Strong Contamination Model)

The draw $X_1, \ldots, X_n$ first is given to an adversary, who *swaps* $\eta \cdot n$ samples, and we see the result $Y_1, \ldots, Y_n$.

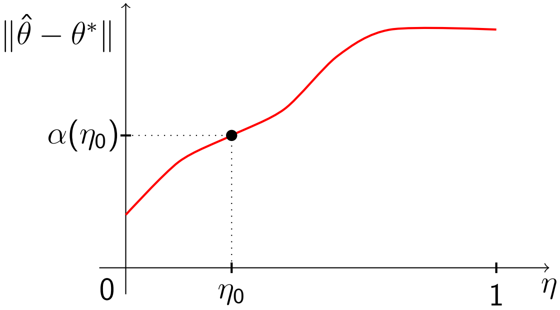


Robust Estimator

Definition (Strong Contamination Model)

The draw $X_1, \ldots, X_n$ first is given to an adversary, who *swaps* $\eta \cdot n$ samples, and we see the result $Y_1, \ldots, Y_n$.

- A **robust estimator** deals with $Y_1, \ldots, Y_n$ and produces $\hat{\theta}$ with property $\|\hat{\theta} - \theta^*\| \leq \alpha(\eta)$, whp.



Black-box Reduction

- **Goal**: Produce an *accurate* summary of $\mathcal{D}$ while respecting individual *privacy*.
  - Weapons: exponential mechanism, robust estimator

Black-box Reduction

- **Goal**: Produce an *accurate* summary of $\mathcal{D}$ while respecting individual *privacy*.
  - Weapons: exponential mechanism, robust estimator

- **Idea**: Give each summary statistic θ a 'score' using the robust estimator

$$s(\mathbf{X}, \theta) = \min\{d(\mathbf{X}, \mathbf{X}') : \|\hat{\theta}(\mathbf{X}') - \theta\| \le \alpha(\eta_0)\}$$

  - If $\theta = \theta^*$ is the true parameter, what is its score?
  - **Does a good θ have high or low score?**

Black-box Reduction

- **Idea**: Give parameter estimate θ a 'score' using the robust estimator

$$s(\mathbf{X}, \theta) = \min\{d(\mathbf{X}, \mathbf{X}') : \|\hat{\theta}(\mathbf{X}') - \theta\| \leq \alpha(\eta_0)\}$$

Lemma (Robust-Private Reduction)

Randomized mechanism $\mathcal{M}$, which selects θ given dataset $\mathbf{X}$ as

$$\Pr[\mathcal{M}(\mathbf{X}) = \theta] \propto \exp(-\varepsilon \cdot s(\mathbf{X}, \theta))$$

Privacy Guarantees

- **Question**: Does $\mathcal{M}$ satisfy notions of differential privacy? How strict?

$$s(\mathbf{X}, \theta) = \min\{d(\mathbf{X}, \mathbf{X}') : \|\hat{\theta}(\mathbf{X}') - \theta\| \leq \alpha(\eta_0)\}$$

  - How large can $\Delta s = |s(\mathbf{X}, \theta) - s(\mathbf{Y}, \theta)|$ be if $\mathbf{X}$ and $\mathbf{Y}$ are adjacent?

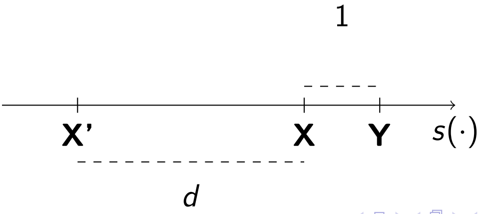
Privacy Guarantees

- **Question**: Does $\mathcal{M}$ satisfy notions of differential privacy? How strict?

$$s(\mathbf{X}, \theta) = \min\{d(\mathbf{X}, \mathbf{X}') : \|\hat{\theta}(\mathbf{X}') - \theta\| \leq \alpha(\eta_0)\}$$

  - How large can $\Delta s = |s(\mathbf{X}, \theta) - s(\mathbf{Y}, \theta)|$ be if $\mathbf{X}$ and $\mathbf{Y}$ are adjacent?

$$\Delta s \leq 1$$



Privacy Guarantees

Lemma (Robust-Private Reduction)

Randomized mechanism $\mathcal{M}$, which selects θ given dataset $\mathbf{X}$ as

$$\Pr[\mathcal{M}(\mathbf{X}) = \theta] \propto \exp(-\varepsilon \cdot s(\mathbf{X}, \theta))$$

satisfies $(2\varepsilon, 0)$-DP

Proof.

$\Delta s = |s(\mathbf{X}, \theta) - s(\mathbf{X}', \theta)| \leq 1$, Exponential Mechanism Lemma. $\square$

Accuracy

Question: Say we draw many samples using $\mathcal{M}(\mathbf{X})$, should we expect many points close to θ^*?

$$\Pr[\mathcal{M}(\mathbf{X}) = \theta] \propto \exp(-\varepsilon \cdot s(\mathbf{X}, \theta))$$

- We select with high probability from low-score regions, does low-score $\Rightarrow$ high-accuracy?

Accuracy

Consider $(\mathbf{X}, \theta)$ with score ηn, **wts** $\|\theta - \theta^*\| \leq$ something

- For the $\mathbf{X}'$ our s 'finds',

$$\|\hat{\theta}(\mathbf{X}') - \theta\| \leq \alpha(\eta_0)$$

Accuracy

Consider $(\mathbf{X}, \theta)$ with score ηn, **wts** $\|\theta - \theta^*\| \leq$ something

- For the $\mathbf{X}'$ our s 'finds',

$$\|\hat{\theta}(\mathbf{X}') - \theta\| \leq \alpha(\eta_0)$$

- By **robustness**

$$\|\theta^* - \hat{\theta}(\mathbf{X}')\| \leq \alpha(\eta)$$

Accuracy

Consider $(\mathbf{X}, \theta)$ with score ηn, **wts** $\|\theta - \theta^*\| \leq$ something

- For the $\mathbf{X}'$ our s 'finds',

$$\|\hat{\theta}(\mathbf{X}') - \theta\| \leq \alpha(\eta_0)$$

- By **robustness**

$$\|\theta^* - \hat{\theta}(\mathbf{X}')\| \leq \alpha(\eta)$$

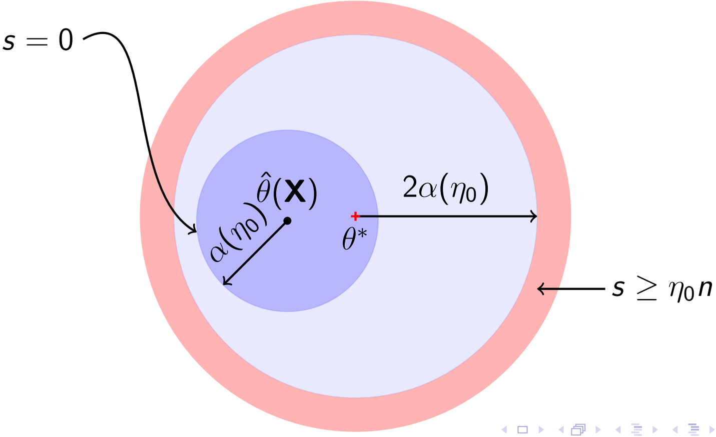
- So, by triangle inequality

$$\|\theta - \theta^*\| \leq \alpha(\eta_0) + \alpha(\eta) \leq \begin{cases} 2\alpha(\eta_0) & \text{if } \eta \leq \eta_0 \\ 2\alpha(\eta) & \text{else} \end{cases}$$

Since α is non-decreasing, small score $\Rightarrow$ small $\|\theta - \theta^*\|$

Accuracy, Concentration

$$\textit{Selecting candidates } \theta \textit{ with } \|\theta - \theta^*\| \le 2\alpha(\eta_0)\textit{?}$$
$$\Updownarrow$$
$$\textit{Avoid selecting } \theta \textit{ with score } \ge \eta_0 n$$



Accuracy, Concentration

Can we bound chances of selecting score $\geq \eta_0 n$?

By definition,

$$\Pr[\mathcal{M}(\mathbf{X}) = \theta] = \frac{\exp(-\varepsilon \cdot s(\mathbf{X}, \theta))}{\int_\Theta \exp(-\varepsilon \cdot s(\mathbf{X}, \theta))}$$

so, for any $\eta \geq \eta_0$,

$$\begin{aligned}\Pr[\mathcal{M}(\mathbf{X}) \text{ has score } \eta n] &= \frac{(\text{volume of } \eta n \text{ points}) \cdot e^{-\varepsilon \eta n}}{\sum_{0 \leq \gamma \leq 1} (\text{volume of } \gamma n \text{ points}) \cdot e^{-\varepsilon \gamma n}} \\ &\leq \frac{V_{2\alpha(\eta)} \cdot e^{-\varepsilon \eta n}}{V_{\alpha(\eta_0)} \cdot e^0}\end{aligned}$$

Accuracy, Concentration

Can we bound chances of selecting score $\geq \eta_0 n$?

$$\sum_{t=\eta_0 n}^{1\cdot n} \Pr[\mathcal{M}(\mathbf{X}) \text{ has score } t] \leq \sum_{t=\eta_0 n}^{1\cdot n} \frac{V_{2\alpha(t/n)} \cdot e^{-\varepsilon \eta n}}{V_{\alpha(\eta_0)}}$$

$\vdots$

$$\leq O(1) \cdot \max_{\eta_0 \leq \eta \leq 1} \left\{ (\eta n)^2 \cdot \frac{V_{2\alpha(\eta)}}{V_{\alpha(\eta_0)}} \cdot \exp(-\varepsilon \eta n) \right\} \leq \beta$$

Accuracy, Concentration

Theorem (Robust-Private Accuracy Guarantee)

Let $X_1, ..., X_n \sim p_{\theta^*}$ where $\theta^* \in \Theta \subseteq \mathbb{R}^d$, we have a random θ drawn by $\mathcal{M}(\mathbf{X})$ has $\|\theta - \theta^*\| \leq 2\alpha(\eta_0)$ given

$$n \geq \max_{\eta_0 \leq \eta \leq 1} \frac{d \log \frac{2\alpha(\eta)}{\alpha(\eta_0)} + \log(1/\beta) + O(\log \eta n)}{\eta \varepsilon}.$$

samples, with probability $1 - 2\beta$.

Accuracy, Concentration

Theorem (Robust-Private Accuracy Guarantee)

Let $X_1, \dots, X_n \sim p_{\theta^*}$ *where* $\theta^* \in \Theta \subseteq \mathbb{R}^d$*, we have a random* θ *drawn by* $\mathcal{M}(\mathbf{X})$ *has* $\|\theta - \theta^*\| \leq 2\alpha(\eta_0)$ *given*

$$n \geq \max_{\eta_0 \leq \eta \leq 1} \frac{d \log \frac{2\alpha(\eta)}{\alpha(\eta_0)} + \log(1/\beta) + O(\log \eta n)}{\eta \varepsilon}.$$

samples, with probability $1 - 2\beta$*.*

Fact: For Gaussians where $\|\mu\|_2 \leq R$, robust estimation produces $\|\hat{\mu} - \mu\|_2 \leq O(c + \eta)$ with $n = O(d/c^2)$ samples

Accuracy

Theorem (Private Gaussian Mean Estimation)

Let $\mu \in \mathbb{R}^d$ where $\|\mu\|_2 \le R$ is unknown. There is an ε-DP algorithm that takes n i.i.d. samples from $N(\mu, I)$ that with high probability outputs $\hat{\mu}$ such that $\|\mu - \hat{\mu}\| \le \alpha$ where

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

- standard cost for robustness and cost of privacy
- (ε, δ)-DP relaxes R to $1/\delta$

Takeaways

- Statistical Estimation and Desiderata
  - Individual privacy preserved in population estimates
  - Estimation despite strong corruption through robustness
- Privacy-preserving algorithms
  - Randomized exponential mechanism, 'maximize' q privately!
- A private algorithm that's *accurate* in context
  - Use some 'robust backbone' as q, why does this make sense?
  - Prove repeated sampling concentrates close to θ^*