Propose Test Release

Today

- Gaussian Mean Estimation
- Propose-Test-Release
- Gaussian Mean Estimation w/PTR

Task: given $x_1, \ldots, x_n \overset{iid}{\sim} N(\mu, I)$, estimate μ with a DP algorithm (as always, privacy is worst-case)

What should be our target error?

Need to hide difference between

$$\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) - \left(\frac{1}{n}\sum_{j=2}^{n+1} x_j\right) = \frac{1}{n}(x_1 - x_{n+1}) \sim N\left(0, \frac{2}{n^2} I\right)$$

Need Gaussian noise like $N\left(0, \sigma_{\varepsilon,\delta}^2 \cdot \frac{d}{n^2} I\right)$

Then might have:

$$\|\mu - \tilde{\mu}\|_2 \le \|\mu - \hat{\mu}\|_2 + \|\hat{\mu} - \tilde{\mu}\|_2$$

$$\lesssim \|N(0, \tfrac{1}{n} I)\|_2 + \|N(0, \sigma_{\varepsilon,\delta}^2 \tfrac{d}{n^2} I)\|_2$$

$$= \frac{1}{\sqrt{n}} \|N(0, I)\|_2 + \frac{\sigma_{\varepsilon,\delta}\sqrt{d}}{n} \|N(0, I)\|_2$$

$$\lesssim \sqrt{\frac{d}{n}} + \sigma_{\varepsilon,\delta} \frac{d}{n}$$

$\hookrightarrow \approx \frac{\sqrt{\log 1/\delta}}{\varepsilon}$

Happy: privacy noise $\to 0$ faster

Issue: global sensitivity of mean $= +\infty$

"Textbook Solution" Assume all i have $\|x_i\|_2 \le R$. or force $x_i \in B(0,R)$

Then $\Delta = \max_{x,x'} \|\hat{\mu}_x - \hat{\mu}_{x'}\|_2 \le \frac{2R}{n}$

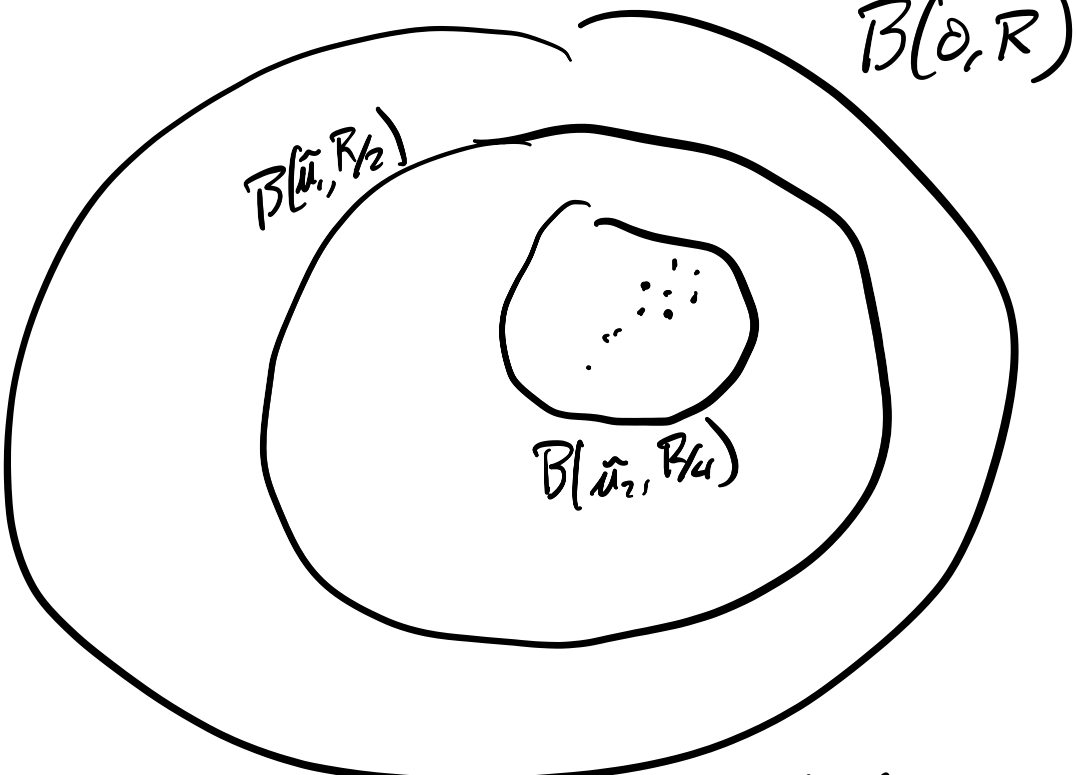
If all data live in $B(0,R)$, then

release $\hat{\mu} + N(0, \frac{R^2}{n^2}\sigma_{\varepsilon,\delta}^2 I)$

Error like $\|u-\tilde{u}\|_2 \lesssim \sqrt{\frac{d}{n}} + \frac{\sigma_{\varepsilon,\delta} R\sqrt{d}}{n}$

What if R is really big?

KLSU17 show how to iterate:



error like $\|u-\tilde{u}\|_2 \le \sqrt{\frac{d}{n}} + \frac{\sigma_{\varepsilon,\delta}\sqrt{\log R}\, d}{n}$

What if R is <u>really</u> big? $R=\infty$?

PTR.

Algorithm 1

Propose-Test-Release

Input: $\varepsilon, \delta \in (0,1)$, $x \in \mathcal{X}^n$, $A : \mathcal{X}^n \to \mathcal{Y}$,
"safety oracle" $\mathcal{O}^A_{\varepsilon,\delta} : \mathcal{X}^n \to \mathbb{N}$

Output: $\hat{y} \in \mathcal{Y}$ or $\perp$

1) $k \leftarrow \mathcal{O}^A_{\varepsilon,\delta}(x)$
2) $\tilde{k} \leftarrow k + \mathrm{Lap}(1/\varepsilon)$
3) If $\tilde{k} \geq \frac{\log(1/\delta)}{\varepsilon}$; release $\hat{y} \sim A(x)$
4) Else; release $\perp$

Assumption 1 Fix A, ε, δ. Let

$$\mathrm{SAFE} = \mathrm{SAFE}^A_{\varepsilon,\delta} = \{x \in \mathcal{X}^n : \forall x' \sim x,\ A(x) \approx_{(\varepsilon,\delta)} A(x')\}.$$

There exists $S \subseteq \mathrm{SAFE}$ such that, for all $x \in \mathcal{X}^n$,

$$\mathcal{O}^A_{\varepsilon,\delta}(x) = \min_{x' \notin S} d(x, x')$$

Draw picture

Claim If Assumption 1 holds, then Algorithm 1 is $(2\varepsilon, e^{\varepsilon}\delta)$-DP.

Proof Fix x, x' and output $E \subseteq \mathcal{Y} \cup \{\perp\}$.
Write Alg 1 as $B(x)$.

Steps:

1) $\Pr[B(x) = \perp]$ & $\Pr[B(x) \neq \perp]$ are e^{ε}, b/c Laplace & low-sensitivity

2) Write

$$\Pr[B(x) \in E] = \Pr[B(x) \in E \mid B(x) \neq \perp] \Pr[B(x) \neq \perp] + \Pr[B(x) \in E \mid B(x) = \perp] \Pr[B(x) = \perp]$$

3) Two cases: $x \in S \subseteq SAFE$

$x \notin S$

$$\boxed{t>0,\ \Pr[Z > t] = \frac{1}{2}\exp\left(-\frac{t}{b}\right) \quad Z \sim Lap(b)}$$

$\boxtimes$

Back to mean estimation

Instantiate Alg 1 with . . . require $n \geq \frac{2\log 1/\delta}{\varepsilon}$

Alg 2

Input: $\varepsilon, \delta \in (0,1)$, $x \in \mathbb{R}^{n \times d}$

Return: $\tilde{\mu} \in \mathbb{R}^d$

1) $c_1 \leftarrow 3\sqrt{d}$; $c_2 \leftarrow c_2(\varepsilon, \delta, d, n)$ set this later
2) $\mu_0 \leftarrow \underset{\mu}{\operatorname{argmax}} \#\{i \in [n] : x_i \in B(\mu, c_1)\}$
3) Project all n points to $B(\mu_0, c_1)$
4) Release $\tilde{\mu} \leftarrow \mu_x + \mathcal{N}(0, c_2 \mathbb{I})$

Safety Oracle

keep for later

Define $S \subseteq \mathcal{X}^n$ as:

$x \in S$ iff $\exists \mu_0$ s.t. $\#\{i \in [n] : x_i \in B(\mu_0, c_1)\} \geq n - \frac{\log 1/\delta}{\varepsilon}$

Let $m(x) = \max_{\mu} \#\{i \in [n] : x_i \in B(\mu, c_1)\}$

$$O^A_{\varepsilon,\delta}(x) = \begin{cases} 0 & \text{if } m(x) \geq n - \frac{\log 1/\delta}{\varepsilon} \\ n - \frac{\log 1/\delta}{\varepsilon} - m(x) & \text{otherwise.} \end{cases}$$

What do we need to prove?

1) Does this oracle compute the distance to some fixed set S?

2) Is this $S \subseteq SAFE^{A}_{\varepsilon,\delta}$?

3) Accuracy: when $x \overset{iid}{\sim} \mathcal{N}(\mu, \mathbb{I})$, whp have $x \in S$, pass the check whp, and return $\hat{\mu} + \mathcal{N}(0, c_2^2 \mathbb{I})$

$$c_2 = c_2(\varepsilon, \delta, d, n) \approx \frac{\sqrt{d}\,(\log 1/\delta)^{3/2}}{\varepsilon^2 n}$$