Notes 20: Differential privacy

1. Motivations

e.g. Robust De-anonymization of Large Sparse Datasets [Narayanan & Shmatikov '08]
  i.e. Breaking anonymity of Netflix Prize Dataset
e.g. Matching Known Patients to Health Records in Washington State Data [Sweeney '13]
  Breaking privacy with multiple overlapping datasets
e.g. Apple since '16, Google's RAPPOR, TensorFlow Privacy, etc, US 2020 Census

---

Suppose STAT in Statistical Query model answers average salary about a company
What if I query average salary of a company, and do so again right after you leave the company?

Randomized response [Warner 1965]
Suppose you are taking a survey on a sensitive topic (e.g. have you taken drug illegally)
  Flip a fair coin, with prob 1/2, you answer Yes
  With prob 1/2, you answer honestly
If p fraction of population belongs to "Yes" group, in expectation $(1+p)/2$ fraction will answer Yes
Survey researcher can deduce p from $(1+p)/2$
Even if you say Yes, you can plausibly deny

---

2. Definition

Dataset $S = \{x_1, \ldots, x_m\} \subseteq X$ and another dataset S' **differ in just one data point** if
  S' is obtained from S by replacing x_i with $x_i' \neq x_i$ for some $1 \leqslant i \leqslant m$
A randomized algorithm A reads a dataset S and outputs $y \in Y$
  Y is called the range of A
  If A is a learning algorithm, then Y = hypothesis class $\mathcal{H}$ of A
  But we also allow algorithms whose output isn't a hypothesis, e.g. STAT$(c, \mathcal{D})$

Definition 1. Randomized algorithm A satisfies **ε-differential privacy** if for any two datasets S, S' differing in just one data point, for any subset $Y' \subseteq Y$ of outcomes of A,

$$\mathbb{P}[A(S) \in Y']e^{-\varepsilon} \leqslant \mathbb{P}[A(S') \in Y'] \leqslant \mathbb{P}[A(S) \in Y']e^{\varepsilon}$$

Since $e^{\varepsilon} \approx 1 + \varepsilon$ and $e^{-\varepsilon} \approx 1 - \varepsilon$
  Above definition requires $\mathbb{P}[A(S) \in Y']/\,\mathbb{P}[A(S') \in Y']$ to be close to 1
If Y (range of A) is discrete, it's equivalent to requiring that for any outcome $y \in Y$ of A,

$$\mathbb{P}[A(S) = y]e^{-\varepsilon} \leqslant \mathbb{P}[A(S') = y] \leqslant \mathbb{P}[A(S) = y]e^{\varepsilon}$$

  Original definition also covers the case where Y is continuous (e.g. $Y = \mathbb{R}$)

---

3. Laplace mechanism

Suppose S consists of m points in $[0, b]$ and we want to estimate their average
Changing one data point in S changes the average by at most b/m
Laplace mechanism outputs the true average plus noise that is a Laplace random variable
Laplace distribution $\mathrm{Lap}(\mu, s)$ with mean μ and scale s has density $f(x \mid \mu, s) = \frac{1}{2s} \exp\left(-\frac{|x-\mu|}{s}\right)$

Laplace mechanism

Output $v = \mathrm{Lap}(a, b/\varepsilon m)$ where a is the true average

In other words, $v = a + x$ where x is the Laplace random variable $\mathrm{Lap}(0, b/\varepsilon m)$
Smaller ε requires larger $b/\varepsilon m$ i.e. more privacy requires larger noise

---

Theorem 2. *Laplace mechanism satisfies ε-differential privacy*

Proof. Fix two datasets S and S' differing in just one data point
If S has average a and S' has average a', then $|a-a'| \leqslant b/m$
Consider the ratio of densities $p_S(v)/p_{S'}(v)$ of outputting v given S (vs S')
Ratio is smallest when $a' = a + b/m$ (the means are furthest apart) and $v \geqslant a'$

$$\frac{p_S(v)}{p_{S'}(v)} \geqslant \frac{f\left(v \mid a, \frac{b}{\varepsilon m}\right)}{f\left(v \mid a + \frac{b}{m}, \frac{b}{\varepsilon m}\right)} = \frac{\exp\left(-\frac{v-a}{b/\varepsilon m}\right)}{\exp\left(-\frac{v-a-b/m}{b/\varepsilon m}\right)} \geqslant \exp(-\varepsilon)$$

Last inequality follows from dropping the denominator (which is at most 1) and taking $v = a'$
Likewise, ratio is largest when $a' = a + b/m$ and $v \leqslant a$

$$\frac{p_S(v)}{p_{S'}(v)} \leqslant \frac{f\left(v \mid a, \frac{b}{\varepsilon m}\right)}{f\left(v \mid a + \frac{b}{m}, \frac{b}{\varepsilon m}\right)} = \frac{\exp\left(-\frac{a-v}{b/\varepsilon m}\right)}{\exp\left(-\frac{a+b/m-v}{b/\varepsilon m}\right)} \leqslant \exp(\varepsilon)$$

Last inequality follows from dropping the numerator (which is at most 1) and taking $v = a$
Required inequality for event $Y \subseteq [0, b]$ follows by integrating over all $v \in Y$ $\square$

Proposition 3. *With prob $1-\delta$, Laplace mechanism adds an error of magnitude at most $\frac{b}{\varepsilon m} \ln \frac{1}{\delta}$*

Proof. For $\tau \geqslant 0$

$$\mathbb{P}[x \geqslant \tau] = \frac{\varepsilon m}{2b} \int_\tau^\infty e^{-x\varepsilon m/b} dx = \frac{1}{2} e^{-\tau\varepsilon m/b}$$

So $\mathbb{P}[x \geqslant \tau] = \delta/2$ when $\tau = \frac{b}{\varepsilon m} \ln \frac{1}{\delta}$
Identical analysis works for $\mathbb{P}[x \leqslant -\tau] = \delta/2$ $\square$

Generalization: To compute some real-valued function (e.g. statistics) g of dataset S
Let Δg = maximum change to g's output when just one data point changes
(General) Laplace mechanism outputs $v = \mathrm{Lap}(g(S), \Delta g/\varepsilon)$
This mechanism satisfies ε-differential privacy, by the same proof

---

Composition: Suppose independent mechanisms $A_1, \ldots, A_k$ answer k queries
Each satisfying ε-differential privacy
Then the vector of k responses $A = (A_1, \ldots, A_k)$ satisfies $k\varepsilon$-differential privacy, since

$$\begin{aligned}\mathbb{P}[A(S') = y] &= \mathbb{P}[A_1(S') = y_1] \cdots \mathbb{P}[A_k(S') = y_k] \leqslant e^\varepsilon\, \mathbb{P}[A_1(S) = y_1] \cdots e^\varepsilon\, \mathbb{P}[A_k(S) = y_k] \\ &= e^{k\varepsilon}\, \mathbb{P}[A(S) = y]\end{aligned}$$

The other inequality is analogous