CB1

DP Lecture

[based on lecture of Kamalika Chaudhuri]

Q: How do we run algorithms on private data w/out leaking sensitive information?

- Examples: Medical records (e.g. disease prevalence)
  Location data
  Census data
  Media preferences
  Training data for LLMs!
- If someone could reconstruct raw data from output of the alg, could result in catastrophic privacy breaches!
- But, data is very useful... so how can we privatize?
  - Let's look at some historical strategies
  - Moral: privacy is hard, & we should strive for formal/theoretical guarantees.

Idea 1: Data anonymization

- Say we have data base of form

name	feature 1	feature 2	...	sensitive bit (e.g. "has cancer")

- Let's just remove the name, then we don't know who participated in all & can safely run any statistical analysis/algorithm/etc!

- ~~In fact, this is what 1996's Mass governor Bill Weld fought~~
- Well, at least 1 person fought this would work, and unfortunately that person was Mass Governor Bill Weld

Case Study 1: Mass Group Insurance Commission

- Weld publicly released the medical records of all state employees w/ identifying info removed. Database looks like:

Name/SSN	Zip	Gender	D.O.B	Condition	...
↑	02540	M	1/9/45	Asthma	
⋮		⋮			
?					
!		;			

- ~~Fantastic~~ for open medical research! Valuable data and great to be able to use it for free.

- Small problem, CS grad student Latanya Sweeney realized two facts:

1) 87% of US residents are uniquely identified by Zip/Gen/D.O.B

2) Voter rolls could be bought for $20 & contain Name + Zip/Gender/DOB!

Uh-oh!

- This is known as a "Linkage attack"
  - Using Linkage, Sweeny reconstructed Governor Weld's full medical record & sent it to him.
  - Led to major priv laws in HIPAA & ended of data release ☺

- Note Linkage attacks are still widely used, e.g. in Netflix challenge... NYC Taxis ...

CB2 Attempt #2) Anonymize & aggregate

Case Study 2: 2010 U.S. Census. [Garfinkel, Abowd, Martindale '19]

- Census collects sensitive population data & releases anonymized data set aggregated by neighborhood

- Example

	Age		
	Count	median	mean
1. Total population	7	30	38
2. Female	4	30	33.5
3. Male	3	30	49

(race, marital status, bits...)

RB1

- Naively, looking at this seems like I can't reconstruct an individuals data by...
  - Nevertheless, let's see if we can infer anything about the individuals of the 3 males $A \le B \le C$

Fact 1) $A \le B \le C \le 125$ ← (upper bound on oldest person)

~~B~~ $A, B, C \in \mathbb{Z}^+$

Narrows it down to about 300,000 options.

Fact 2) $B = 30$

Fact 3) $\frac{1}{3}(A+B+C) = 44 \rightarrow \frac{1}{3}(A+C) = 34$

- I can easily enumerate all A, B, C satisfying these 3 constraints
  - only 31 options! e.g. $A=0, B=30, C=102$...
  - using additional fields, we can likely get this down to a unique A, B, C
  - Now I can use linkage to reconstruct name + personal info.

[GAO '19] recovers info of 50 million Americans using this strategy.

- In general, above amounts to an integer programming problem.
- NP-hard in worst-case, but often very efficiently solvable in any case (SAT-Solvers)...

Moral: We need a more formal theory of privacy in algorithms.

- Before I get into technical def, let's actually build a basic private algo to get some intuition.

Erase + CB1 Say

- Setting: teaching a big class of n students, I suspect many often cheat on the final. I'd like to estimate how many, but sensitive info (do students want tell me). How do I build a procedure s.t. I get an estimate w/out finding student privacy?

write Model:
- Formal: n individuals, each has sensitive bit $X_i \in \{0,1\}$
- Analyst would like to compute $\frac{1}{n}\sum X_i$
- Each individual will send me Y_i (depends on X_i + randomness)
- ~~I'd like to estimate $p = \frac{1}{n}\sum X_i$~~

Attempt 1: $Y_i = X_i$

$$Y_i = \begin{cases} X_i & \text{w.p. } 1 \\ 1 - X_i & \text{w.p. } 0 \end{cases}$$

- analyst computes $\hat{p} = \frac{1}{n}\sum Y_i = p$ ← perfect accuracy!
- But no privacy at all ...

LB 2

Attempt 2: ~~#~~

$$Y_i = \begin{cases} X_i & \text{w.p. } 1/2 \\ 1 - X_i & \text{w.p. } 1/2 \end{cases}$$

- Very private! But what's the issue here...? Ask audience
  - Right! Y_i's is totally independent of X_i, $Y_i \sim Ber\left(\frac{1}{2}\right)$
  - gives 0 info about $\frac{1}{n}\sum X_i$, so can't use to compute theorem.

LB 1 #2

Attempt 3: (γ-randomized response)

$$Y_i = \begin{cases} X_i & \text{w.p. } 1/2+\gamma \\ 1 - X_i & \text{w.p. } 1/2-\gamma \end{cases}$$

Ask audience for potential fix:

$\gamma = 1/2$ Attempt 1

$\gamma = 0$ Attempt 2

$\gamma = 1/4$: "plausible deniability"

- Ok, so intuitively this is somewhat private, but can I get an accurate estimate of $\frac{1}{n}\sum X_i$?
  - let's look at how Y_i & X_i are related
    - $\mathbb{E}[Y_i] = \left(\frac{1}{2}+\gamma\right)X_i + \left(\frac{1}{2}-\gamma\right)(1-X_i) = 2\gamma X_i + \frac{1}{2} - \gamma$

$$\rightarrow \mathbb{E}\left[\frac{1}{2\gamma}\left(Y_i - \frac{1}{2}+\gamma\right)\right] = X_i$$

- This suggests we use the unbiased estimator

$$\tilde{p} = \frac{1}{n}\sum \frac{1}{2\gamma}\left(Y_i + \frac{1}{2} + \gamma\right)$$

- By linearity of expectation:

$$\mathbb{E}[\hat{p}] = p \checkmark$$

- Great, but expectation is not enough, we'd like to get an answer near p w.h.p. (expectation could be pushed by $p+1000$ or $p-10000$ $\tilde{p}$)

ask author why this worth part?

→ Notice $\tilde{p}$ is sum of bounded i.i.d. random variables! this means $\tilde{p}$ is highly concentrated around its mean

- Chernoff/Hoeffding: Alpha!!

$$\Pr\left[|\tilde{p} - p| \geq \alpha\right] \leq e^{-2\alpha^2\gamma^2 n}$$

this is the standard deviation

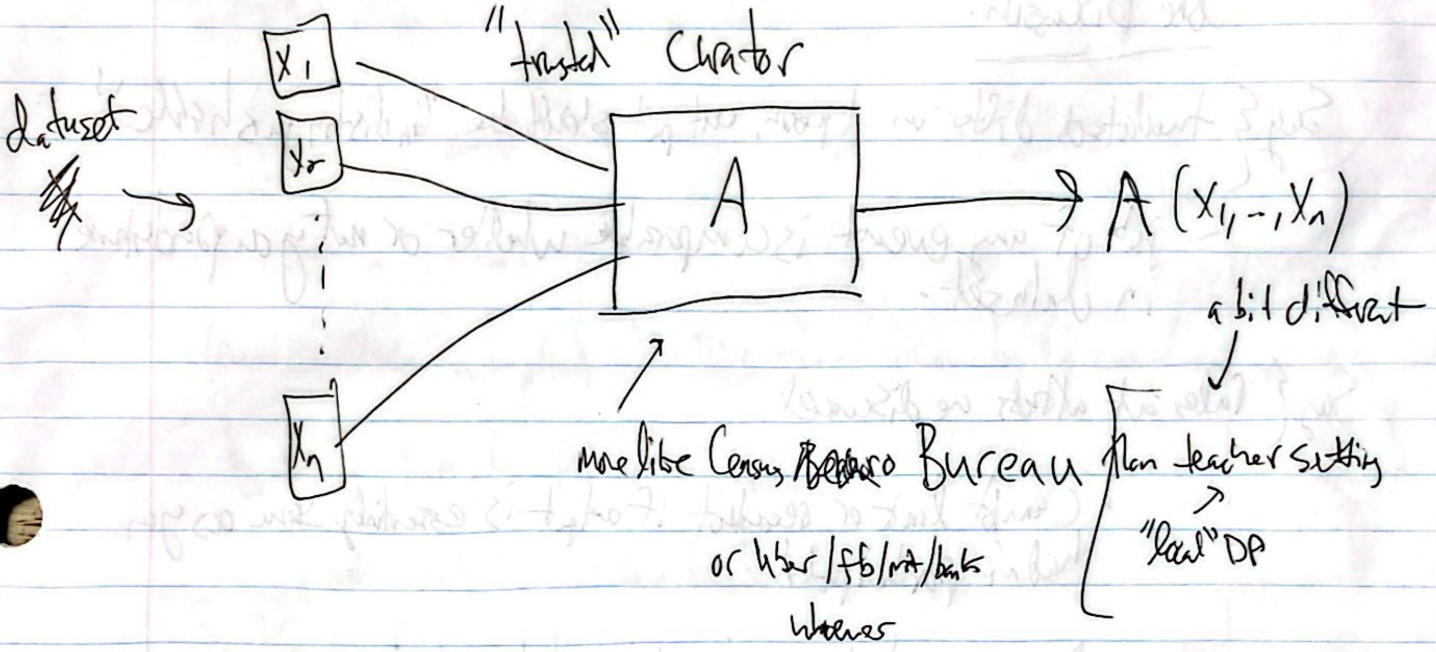
- in other words, error is $\leq \frac{1}{\gamma\sqrt{n}}$ w.h.p

- Said another way, to achieve α accuracy I need about $\frac{1}{\alpha^2\gamma^2}$ samples.

CB1

- Differential Privacy (central model)
  - We'll be careful in specifying the setup

- As before, we have n individuals w/ data X_i



Definition (ε-Differential Privacy) [Dwork, McSherry, Nissim, Smith '06]

↑ Start of based on older work of Dinur-Nissim-...

Say: "Morally, no 1 individual has much effect on the output"

Write: $A: X^n \to Y$ is ε-DP if $\forall$ "neighboring" datasets $x, x' \in X^n$ (differ in 1 coordinate)

$$\forall T \subseteq Y \quad \Pr[A(x) \in T] \leq e^{\varepsilon} \Pr[A(x') \in T]$$

Note important def in theory & practice! Used at all major companies, US Census, won Gödel & ToT

Why? 1) Very Strong privacy guarantee
2) Achievable!
3) User Friendly

← we'll see all 3 as we go

Some Discussion:

Say $\{$ two datasets differ on 1 person, output should be "indistinguishable"
- prob of any event is comparable whether or not you participate in dataset.

Say $\{$ rules at attacks we discussed
- Can't link or reconstruct if output is essentially same as you hadn't participated
- Basically protects against arbitrary risks

Some Technical Comments
- Any (non-trivial) DP algorithm is randomized
- Why e^{ε}? Part of user friendliness. Note this is roughly $(1 \pm \varepsilon)$-approx but $e^{\varepsilon_1} e^{\varepsilon_2} = e^{\varepsilon_1 + \varepsilon_2}$ which is very useful!

RB1

User Friendliness of DP:

1. Group Privacy: if I take out k individuals ~~as~~ I still get

$$\Pr[M(x) \in T] \leq e^{\varepsilon k} \Pr[M(x') \in T]$$

x & x' differ in k-coords, so protects groups

2. Basic Composition: if $A_1, \ldots, A_k : X^n \to Y$ are ε-DP, then

$(A_1, \ldots, A_k) : X^n \to Y^k$ is $k\varepsilon$-DP

(even if chosen adaptively!) This means I can build complicated DP algs by using simple private mechanisms as "building blocks"

(well, really I'd output some function of $A_1, \ldots, A_k$, but...)

3. Post-Processing:

if $A : X^n \to Y$ is ε-DP & $f : Y \to Z$ is any fn:

$f(A) : X^n \to Z$ is ε-DP

"DP-alg cannot be unpublished if we don't release data"

RB2 Pf of Group)

Say $x = (x_1, \dots, x_n)$ & $x' = (y_1, \dots, y_k, x_{k+1}, \dots, x_n)$

- Define $x^{(i)} = (y_1, \dots, y_i, x_{i+1}, \dots, x_n)$
  - Observe $x^{(0)} = x$ & $x^{(k)} = x'$

$$\Pr[M(x^{(0)}) \in T] = \Pr[M(x^{(0)}) \in T]$$

$$\leq e^{\varepsilon} \Pr[M(x^{(1)}) \in T]$$

$$\leq e^{2\varepsilon} \Pr[M(x^{(2)}) \in T]$$

$$\vdots$$

$$\leq e^{k\varepsilon} \Pr[M(x^{(k)}) \in T]$$

$$\leq e^{k\varepsilon} \Pr[M(x') \in T] \quad \checkmark$$

Return to randomized response

← on LB1

Recall $Y_i = \begin{cases} X_i & \text{w.p. } \frac{1}{2}+\gamma \\ 1-X_i & \text{w.p. } \frac{1}{2}-\gamma \end{cases}$

- We showed $\tilde{p} = \frac{1}{n}\sum \frac{1}{2\gamma}\left(Y_i - \frac{1}{2} + \gamma\right)$ is $\frac{1}{\gamma\sqrt{n}}$-accurate
  - is it private? We argue this is $O(\gamma)$-DP!
  (rule $\tilde{p} = A(x_1, \ldots, x_n)$)
- By post-processing, enough to argue $(Y_1, \ldots, Y_n)$ is $O(\gamma)$-DP
- Fix some potential output $z \in \{0,1\}^n = Y^n$, we'll show for all $x \sim x'$

$$\frac{\Pr[M(x)=z]}{\Pr[M(x')=z]} \leq e^{\varepsilon}$$

enough, $\forall T \subseteq Y$

$$\Pr[M(x) \in T] = \sum_{z \in T} \Pr[M(x)=z]$$

$$\leq e^{\varepsilon} \sum_{z \in T} \Pr[M(x')=z]$$

$$\leq e^{\varepsilon} \Pr[M(x') \in T]$$

- Recall Y_i are independent, so

$$\frac{\Pr[M(x)=z]}{\Pr[M(x')=z]} = \frac{\prod_{i=1}^{k} \Pr[Y_i = z_i]}{\prod_{i=1}^{k} \Pr[Y_i' = z_i]}$$

Say x, x' differ in coord j

$$\frac{\Pr[Y_j = z_j]}{\Pr[Y_j' = z_j]} = \frac{\frac{1}{2}+\gamma}{\frac{1}{2}-\gamma} \leq e^{O(\gamma)} \quad (\text{for } \gamma \text{ say} < 1/4)$$

$e^{O(\gamma)}$ ← Taylor expansion

Summarize: We can estimate $\frac{1}{n}\sum X_i$ w/..

ε-DP & $\frac{1}{\varepsilon\sqrt{n}}$-accuracy.

OR: ε-DP, α-accuracy requires $1/\varepsilon^2\alpha^2$ data points

---

LB?

- Can we do better? What about beyond counting? More general for data science?
- Previous strategy boiled down to adding noise that "hides" on particular pt, but is well-behaved over the sum.
- Similar, if slightly less intuitive idea is to just add mean 0 noise directly to the function we want to compute
  - But what form of noise, and how much?
  - e.g. output $\frac{1}{n}\sum X_i + \eta$ ← new "0" noise concentrated

1) Quantifying how much noise we need

Def) The l_1-sensitivity of $f: X^n \to \mathbb{R}^k$ is ← motivated DP

$$\Delta^{(f)} := \max_{x \sim x'} \|f(x) - f(x')\|_1$$ ← "Maximal Lipschitz Variable diff"

- measures maximum effect of any individual, high sensitivity = high noise

2) What type of noise? Take inspiration from DP def e^{ε} factor

add "Laplace" noise $Lap_b(x) = \frac{1}{2b} e^{-\frac{|x|}{b}}$

$\mathbb{E}(x) = 0$

$Var(x) = 2b^2$

$Pr[\|M - f\|_1 > \alpha] \leq k e^{-\frac{\alpha \varepsilon}{k \Delta^f}}$ ← general error fn

Def) Laplacian mechanism:

$$A(x) = f(x) + (\eta_1, \ldots, \eta_k), \quad \eta_i \sim Lap\left(\tfrac{\Delta}{\varepsilon}\right)$$

Claim 1) $A(x)$ is ε-DP

Claim 2) $A(x)$ is accurate

let's examine $f = \frac{1}{n} \sum X_i$, can show

$$Pr[|A(x) - f| > \alpha] \leq e^{-\varepsilon \alpha n}$$

error ~ $1/\varepsilon n$

or need $n = 1/\alpha\varepsilon$ samples!

much better.

R3 Pf) Analog immediate from tail bound of Laplacian (exp tail)

2) DP

fix output z, want to prove $\frac{P_x(z)}{P_{x'}(z)} \leq e^{\varepsilon}$

for P_x the pdf of $A(x)$.

$$\frac{P_x(z)}{P_{x'}(z)} \leq \frac{\prod_{i=1}^{k} \exp\left(-\frac{\varepsilon |f(x)_i - z_i|}{\Delta}\right)}{\prod_{i=1}^{k} \exp\left(-\frac{\varepsilon |f(x')_i - z_i|}{\Delta}\right)}$$

$$= \prod_{i=1}^{k} \exp\left(\frac{\varepsilon}{\Delta}|f(x')_i - z_i| - \frac{\varepsilon}{\Delta}|f(x)_i - z_i|\right).$$

$$\leq \prod_{i=1}^{k} \exp\left(\frac{\varepsilon}{\Delta}|f(x')_i - f(x)_i|\right) \quad (\Delta \text{ inequality})$$

$$= \exp\left(\frac{\varepsilon}{\Delta}\sum |f(x')_i - f(x)_i|\right)$$

$$\leq \exp(\varepsilon) \checkmark$$

• Post Processing Prof) $A: X \to Y$ & $f: Y \to Z$

• Fix neighboring $x \sim x'$ & event $T \subseteq Z$

$$Pr[f(M(x)) \in T] = Pr[M(x) \in f^{-1}(T)]$$

$$\leq e^{\varepsilon} Pr[M(x') \in f^{-1}(T)]$$

$$= e^{\varepsilon} Pr[f(M(x')) \in T]$$

Basic Composition Pf) $A = (A_1, \ldots, A_k)$ seq of ε-DP

• Fix nbr $x \sim x'$ & output $z = (z_1, \ldots, z_k)$

$$\frac{Pr[M(x) = z]}{Pr[M(x') = z]} = \prod_{i=1}^{k} \frac{Pr[M_i(x) = z_i \mid M_1(x) \ldots M_{i-1}(x) = (z_1, \ldots, z_{i-1})]}{Pr[M_i(x') = z_i \mid M_1(x') \ldots M_{i-1}(x') = (z_1, \ldots, z_{i-1})]}$$