Differential Privacy: a short tutorial

Presenter: WANG Yuxiang

Some slides/materials extracted from

- Aaron Roth's Lecture at CMU
  - The Algorithmic Foundations of Data Privacy [Website]
- Cynthia Dwork's FOCS'11 Tutorial
  - The Promise of Differential Privacy. A Tutorial on Algorithmic Techniques. [Website]
- Christine Task's seminar
  - A Practical Beginners' Guide to Differential Privacy [YouTube][Slides]

In the presentation

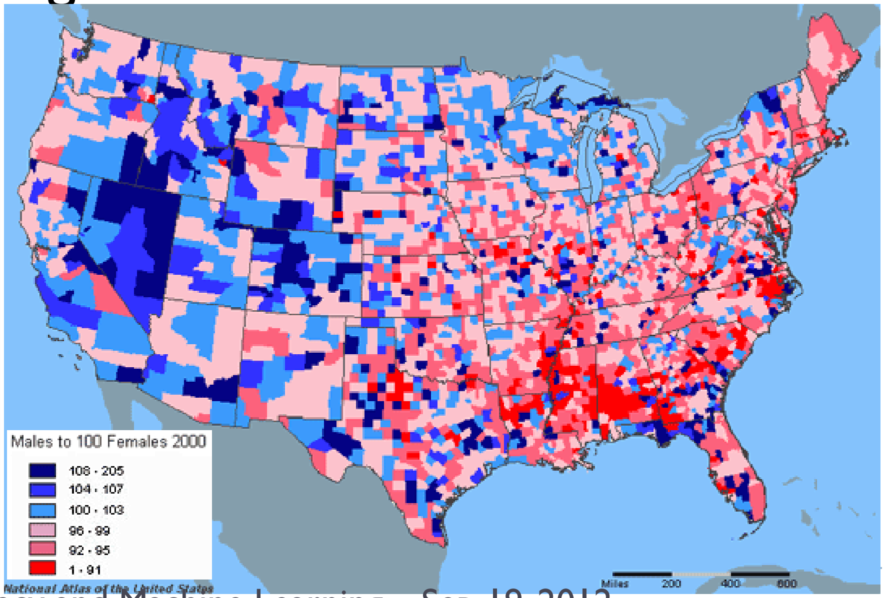
1. Intuitions
   - Anonymity means privacy?
   - A running example: Justin Bieber
   - What exactly does DP protects? Smoker Mary example.
2. What and how
   - ϵ-Differential Privacy
   - Global sensitivity and Laplace Mechanism
   - (ϵ, δ)-Differential Privacy and Composition Theorem
3. Many queries
   - A disappointing lower bound (Dinur-Nissim Attack 03)
   - Sparse-vector technique

In the presentation

4. **Advanced techniques**
   - Local sensitivity
   - Sample and Aggregate
   - Exponential mechanism and Net-mechanism
5. **Diff-Private in Machine Learning**
   - Diff-Private logistic regression (Perturb objective)
   - Diff-Private low-rank approximation
   - Diff-Private PCA (use Exponential Mechanism)
   - Diff-Private SVM

Privacy in information age

- Government, company, research centers collect personal information and analyze them.
- Social networks: Facebook, LinkedIn
- YouTube & Amazon use viewing/buying records for recommendations.
- Emails in Gmail are used for targeted Ads.



Privacy by information control

- Conventional measures for privacy :
  - Control access to information
  - Control the flow of information
  - Control the purposes information is used
- Typical approaches for private release
  - Anonymization (removing identifiers or k-anonymity)
  - (Conventional) sanitization (release a sampled subset)
- They basically do not guarantee privacy.

An example of privacy leak

- De-anonymize Netflix data
  - "Sparsity" of data: With large probability, no two profiles are similar up to ϵ. In Netflix data, not two records are similar more than 50%.
  - If the profile can be matched up to 50% similarity to a profile in IMDB , then the adversary knows with good chance the true identity of the profile.
  - This paper proposes efficient random algorithm to break privacy.

A. Narayanan and V. Shmatikov, "Robust de-anonymization of large sparse datasets (how to break anonymity of the netflix prize dataset)," in Proc. 29th IEEE Symposium on Security and Privacy, 2008.

Other cases

- Medical records of MA Governor in “anonymized” medical database
- Search history of Thelma Arnold in “anonymized” AOL query records
- DNA of a individual in genome-wide association study (getting scary…)
- 天涯人肉搜索引擎… Various ways to gather background information.

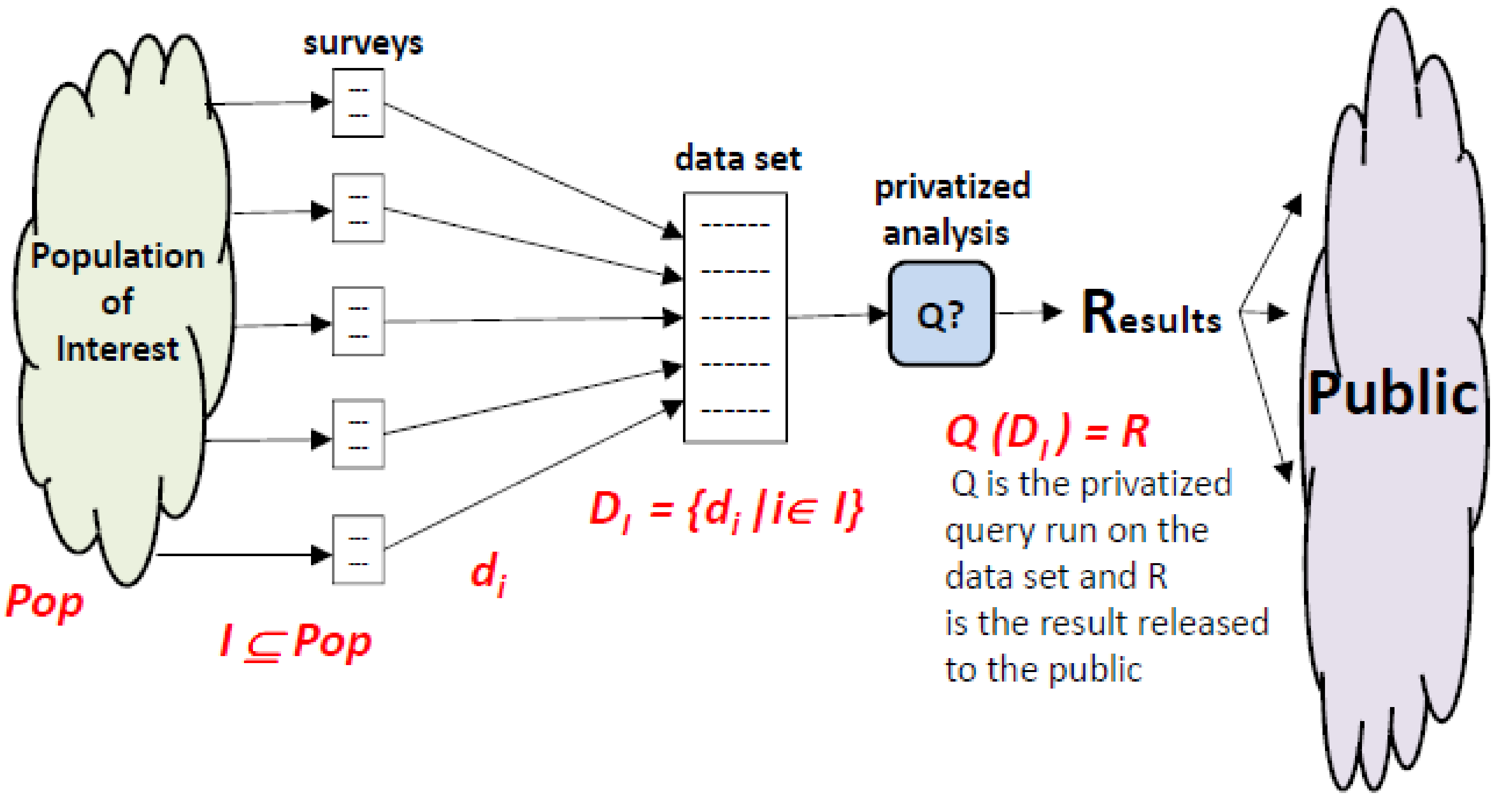
A running example: Justin Bieber

- To understand the guarantee and what it protects against.
- Suppose you are handed a survey:

> 1) Do you like listening to Justin Bieber?
>
> 2) How many Justin Bieber albums do you own?
>
> 3) What is your gender?
>
> 4) What is your age?

- If your music taste is sensitive information, what will make you feel safe? Anonymous?

Notations


surveys
Population of Interest
Pop
$I \subseteq Pop$
d_i
data set
$D_I = \{d_i \mid i \in I\}$
privatized analysis
Q?
Results
Public
$Q(D_I) = R$
Q is the privatized query run on the data set and R is the result released to the public

What do we want?

I would feel safe submitting a survey if…

- I knew that my answer had no impact on the released results.
- I knew that any attacker looking at the published results R couldn't learn (with any high probability) any new information about me personally.

- $Q(D_{(I\text{-}me)}) = Q(D_I)$
- Prob(*secret*(me) | R) = Prob(*secret*(me))

Why can't we have it?

- If individual answers had no impact on the released results… Then the results would have no utility
- If R shows there's a strong trend in my population (everyone is age 10-15 and likes Justin Bieber), with high probability, the trend is true of me too (even if I don't submit a survey).

- By induction, $Q(D_{(I\text{-}me)}) = Q(D_I) \Rightarrow$ $Q(D_I) = Q(D_{\varnothing})$
- Prob(*secret*(me) | *secret*(Pop)) > Prob(*secret*(me))

Why can't we have it?

- Even worse, if an attacker knows a function about me that's dependent on general facts about the population:

  - I'm twice the average age
  - I'm in the minority gender

  Then releasing just those general facts gives the attacker specific information about me. (Even if I don't submit a survey!)

- $(\mathit{age}(\text{me}) = 2^*\text{mean\_age}) \wedge$
  $(\mathit{gender}(\text{me}) \neq \text{mode\_gender}) \wedge$
  $(\text{mean\_age} = 14) \wedge$
  $(\text{mode\_gender} = \text{F}) \Rightarrow$

  $(\mathit{age}(\text{me}) = 28) \wedge$
  $(\mathit{gender}(\text{me}) = \text{M})$

Disappointing fact

- We can’t promise my data won’t affect the results

- We can’t promise that an attacker won’t be able to learn new information about me. Giving proper background information.

- What can we do?

One more try

I'd feel safe submitting a survey....

If I knew the chance that the privatized released result would be R was nearly the same, whether or not I submitted my information.

Differential Privacy

- Proposed by Cynthia Dwork in 2006.
- The chance that the noisy released result will be C is nearly the same, whether or not you submit your info.

Definition: ϵ-Differential Privacy

$$\frac{\Pr(M(D) = C)}{\Pr(M(D_{\pm i}) = C)} < \mathrm{e}^{\epsilon}$$

For any $\left|D_{\pm i} - D\right| \leq 1$ and any $C \in Range(M)$.

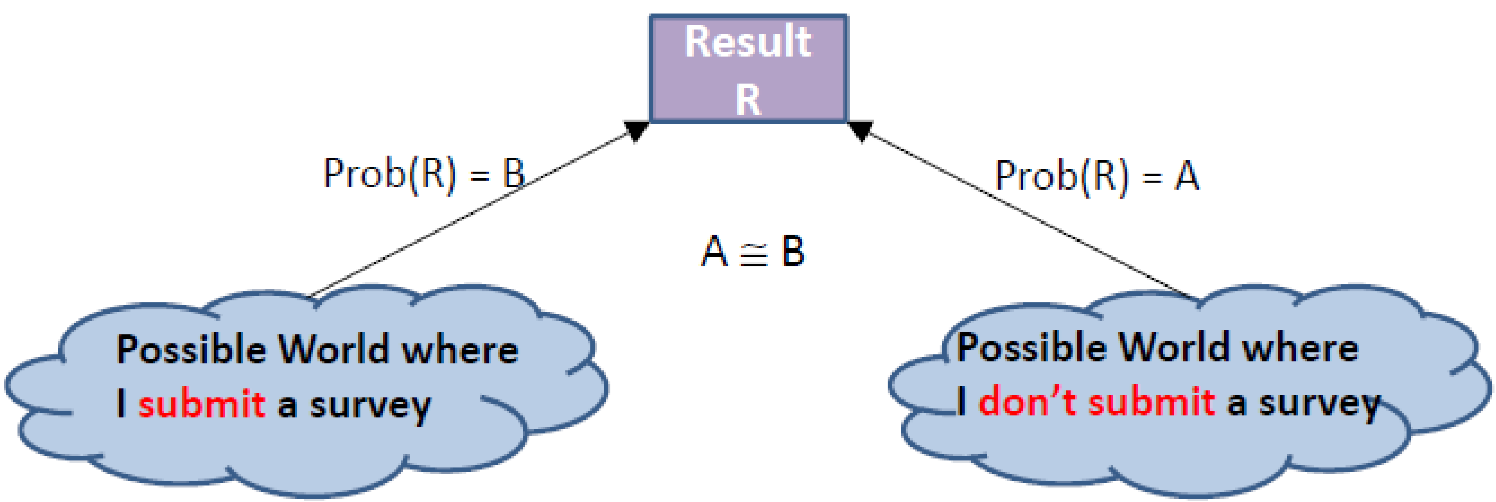
- The harm to you is "almost" the same regardless of your participation.

Differential Privacy

The chance that the noisy released result will be R is nearly the same, whether or not you submit your information.

$$\frac{Prob(R \mid true\ world\ = DI)}{Prob(R \mid true\ world = D_{I \pm i})} \leq e^{\varepsilon}, \qquad for\ all\ I, i, R\ and\ small\ \varepsilon > 0$$

Given R, how can anyone guess which possible world it came from?



Popular over-claims

- DP protects individual against ALL harms regardless of prior knowledge. Fun paper: "Is Terry Gross protected?"
  - Harm from the result itself cannot be eliminated.

- DP makes it impossible to guess whether one participated in a database with large probability.
  - Only true under assumption that there is no group structure.
  - Participants is giving information only about him/herself.

A short example: Smoking Mary

- Mary is a smoker. She is harmed by the outcome of a study that shows "smoking causes cancer":
  - Her insurance premium rises.
- Her insurance premium will rises regardless whether she participate in the study or not. (no way to avoid as this finding is the whole point of the study)
- There are benefits too:
  - Mary decided to quit smoking.

- Differential privacy: limit harms to the teachings, not participation
  - The outcome of any analysis is essentially equally likely, independent of whether any individual joins, or refrains from joining, the dataset.
  - Automatically immune to linkage attacks

Summary of Differential Privacy idea

- DP can:
  - Deconstructs harm and limit the harm to only from the results
  - Ensures the released results gives minimal evidence whether any individual contributed to the dataset
  - Individual only provide info about themselves, DP protects Personal Identifiable Information to the strictest possible level.

In the presentation

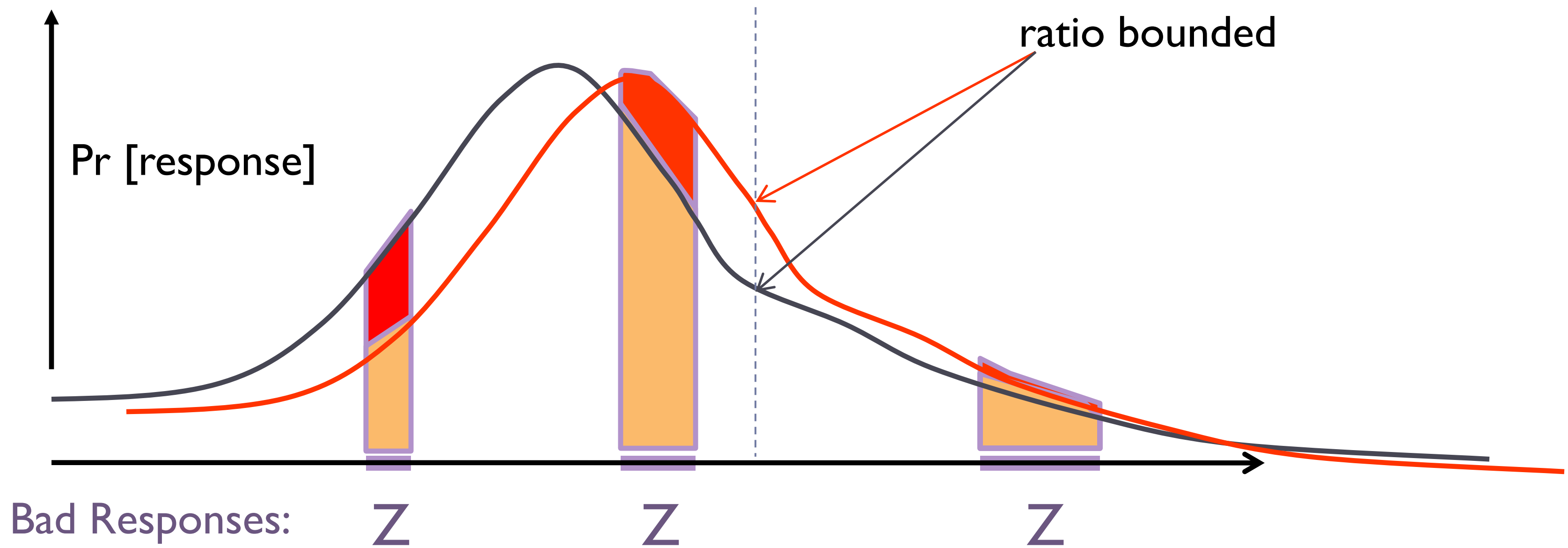
1. Intuitions
   - Anonymity means privacy?
   - A running example: Justin Bieber
   - What exactly does DP protects? Smoker Mary example.
2. What and how
   - ϵ-Differential Privacy
   - Global sensitivity and Laplace Mechanism
   - (ϵ, δ)-Differential Privacy and Composition Theorem
3. Many queries
   - A disappointing lower bound (Dinur-Nissim Attack 03)
   - Sparse-vector technique

Differential Privacy [D., McSherry, Nissim, Smith 06]

$\mathcal{M}$ gives ϵ- differential privacy if for all adjacent x and x', and all $C \subseteq range(\mathcal{M})$: Pr[$\mathcal{M}$ (x) ∈ C] ≤ e^{ε} Pr[$\mathcal{M}$ (x') ∈ C]

Neutralizes all linkage attacks.

Composes unconditionally and automatically: $\Sigma_i\ \varepsilon_i$



Sensitivity of a Function

$$\Delta f = \max_{\text{adjacent } x,x'} |f(x) - f(x')|$$

Adjacent databases differ in at most one row.

Counting queries have sensitivity 1.

Sensitivity captures how much one person's data can affect output.

Example

- How many survey takers are female?
  - Sensitivity =1
- In total, how many Justin Bieber albums are bought by survey takers?
  - Sensitivity = 4? Since he has only 4 different albums by far.

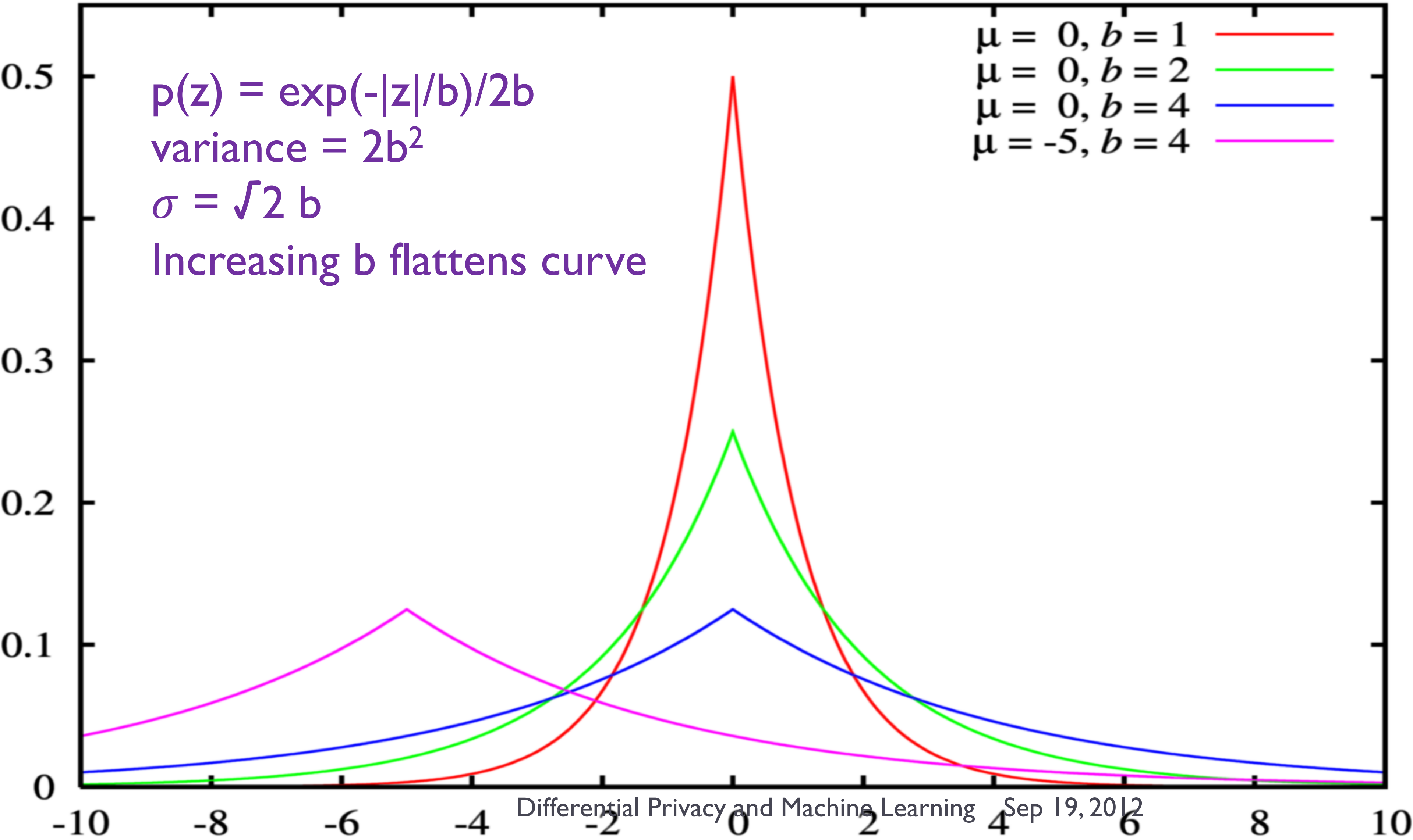
Laplace Distribution Lap(b)

p(z) = exp(-|z|/b)/2b

variance = $2b^2$

$\sigma = \sqrt{2}\, b$

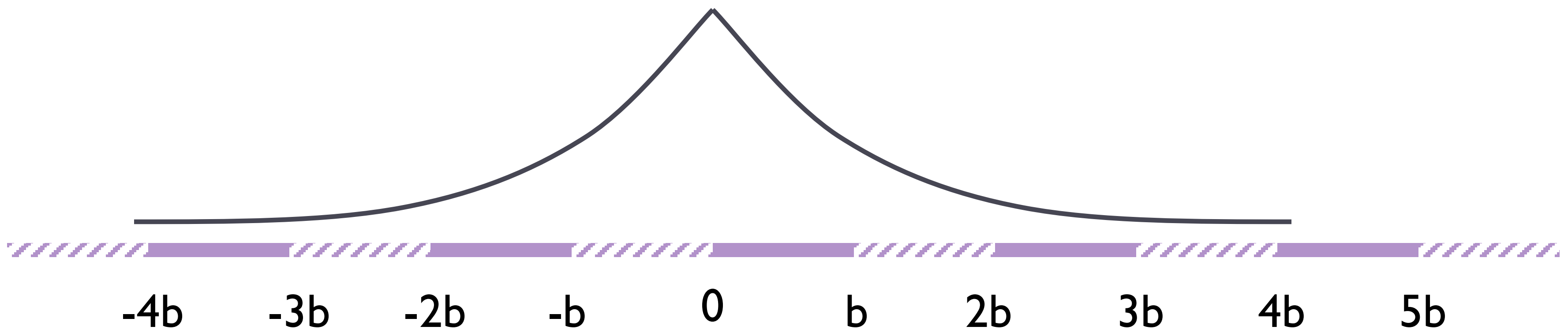
Increasing b flattens curve



Laplace Mechanism

$$\Delta f = \max_{\text{adj } x,x'} \; |f(x) - f(x')|$$

Theorem [DMNS06]: On query f, to achieve ε-differential privacy, use scaled symmetric noise [Lap(b)] with b = Δf/ε.



Noise depends on f and ε, not on the database

Smaller sensitivity (Δf) means less distortion

Proof of Laplace Mechanism

- $$\frac{Pr(f(x)+Lap(\Delta f/\epsilon)=y)}{Pr(f(x')+Lap(\Delta f/\epsilon)=y)} = \frac{\exp\left(-\frac{|y-f(x)|\,\epsilon}{\Delta f}\right)}{\exp\left(-\frac{|y-f(x')|\epsilon}{\Delta f}\right)}$$

$$= \exp\left(\frac{\epsilon}{\Delta f}\,(|y - f(x')| - |y - f(x)|)\right)$$

$$\leq \exp\left(\frac{\epsilon}{\Delta f}\,(|f(x) - f(x')|)\right) \leq e^{\epsilon}$$

- That's it!

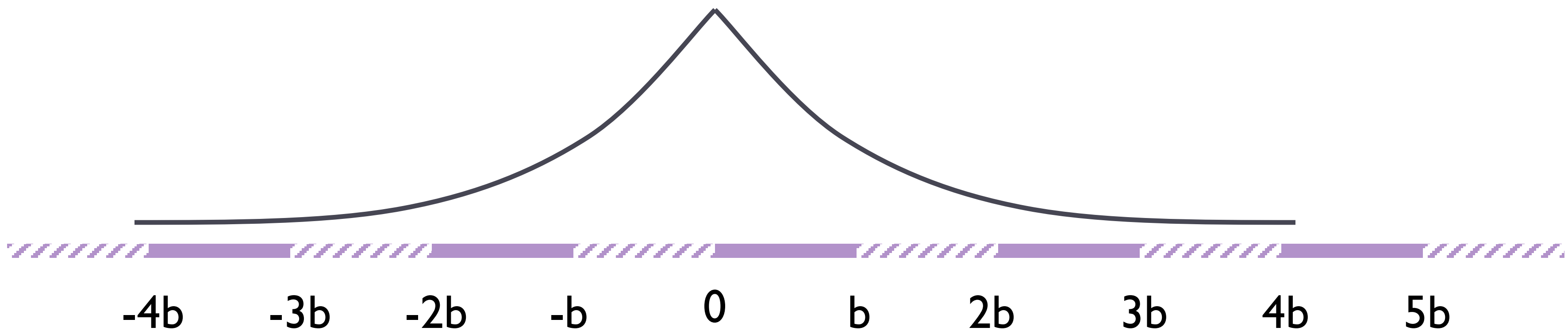
Example: Counting Queries

- How many people in the database are female?
  - Sensitivity = 1
  - Sufficient to add noise $\sim \mathrm{Lap}(1/\epsilon)$
- What about multiple counting queries?
  - It depends.

Vector-Valued Queries

$$\Delta f = \max_{\text{adj } x, x'} \|f(x) - f(x')\|_1$$

Theorem [DMNS06]: On query f, to achieve ε-differential privacy, use scaled symmetric noise $[\mathrm{Lap}(\Delta f/\varepsilon)]^d$.



Noise depends on f and ε, not on the database

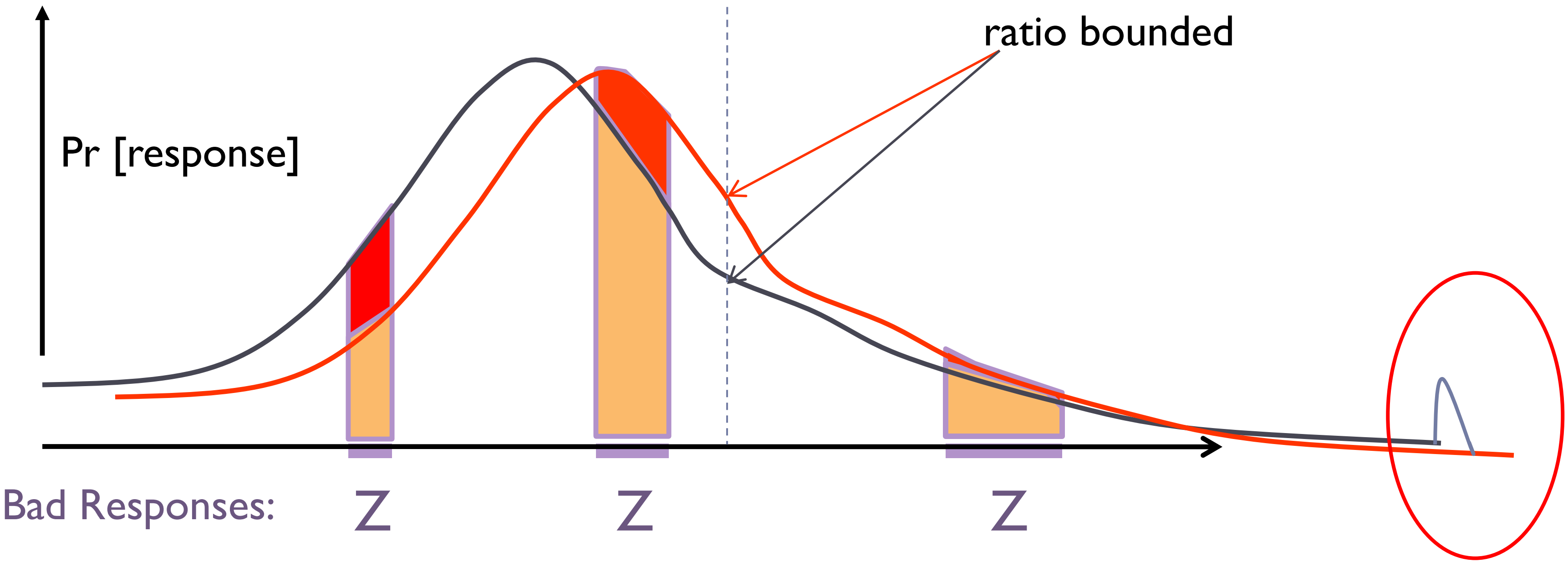
Smaller sensitivity (Δf) means less distortion

(ε, δ) - Differential Privacy

$\mathcal{M}$ gives (ϵ, δ)- differential privacy if for all adjacent x and x', and all $C \subseteq range(\mathcal{M})$: Pr[$\mathcal{M}$ (D)∈C] ≤ e^{ε} Pr[$\mathcal{M}$ (D') ∈C] + δ

Neutralizes all linkage attacks.

Composes unconditionally and automatically: $(\Sigma_i\, \varepsilon_i\, ,\, \Sigma_i\, \delta_i\,)$



This talk: δ negligible

From worst case to average case

- Trade off a lot of ϵ with only a little δ
- How?

$\forall C \in \mathrm{Range}(M)$:

$$\frac{\Pr[M(x) \in C]}{\Pr[M(x') \in C]} \leq e^{\epsilon}$$

Equivalently,

$$\ln\left[\frac{\Pr[M(x) \in C]}{\Pr[M(x') \in C]}\right] \leq \epsilon$$

"Privacy Loss"

Useful Lemma [D., Rothblum, Vadhan'10]:
Privacy loss bounded by $\epsilon \Rightarrow$ expected loss bounded by $2\epsilon^2$.

Max Divergence and KL-Divergence

- Max Divergence (exactly the definition of ϵ!) :

$$D_\infty(Y||Z) = \max_{S \subset Supp(Y)} \left[\ln \frac{\Pr[Y \in S]}{\Pr[Z \in S]} \right]$$

- KL Divergence (average divergence)

$$D(Y||Z) = \mathrm{E}_{y \sim Y} \left[\ln \frac{\Pr[Y = y]}{\Pr[Z = y]} \right]$$

- The Useful Lemma gives a bound on KL-divergence.

$$\mathrm{D}(\mathrm{Y} \parallel \mathrm{Z}) \leq \epsilon(e^\epsilon - 1)$$

$$\epsilon(e^\epsilon - 1) \leq 2\epsilon^2 \; when \; \epsilon < 1$$

“Simple” Composition

- k-fold composition of (ε, δ)-differentially private mechanisms is ($k\varepsilon$, $k\delta$)-differentially private.
  - If want to keep original guarantee, must inject k times the noise
  - When k is large, this destroys utility of the output

- Can we do better than that by again leveraging the tradeoff?
  - Trade-off a little δ with a lot of ϵ?

Composition [D., Rothblum,Vadhan'10]

- Qualitively: Formalize Composition
  - Multiple, adaptively and adversarially generated databases and mechanisms
  - What is Bob's lifetime exposure risk?
    - Eg, for a 1-dp lifetime in 10,000 ϵ-dp or (ϵ, δ)-dp databases
    - What should be the value of ϵ?

- Quantitatively
  - $\forall \epsilon, \delta, \delta'$: the k-fold composition of (ϵ, δ)-dp mechanisms is $\left(\sqrt{2k \ln 1/\delta'}\ \epsilon + k\epsilon(e^{\epsilon} - 1),\ k\delta + \delta'\right)$-dp
  - $\sqrt{k}\epsilon$ rather than $k\epsilon$

Flavor of Privacy Proof

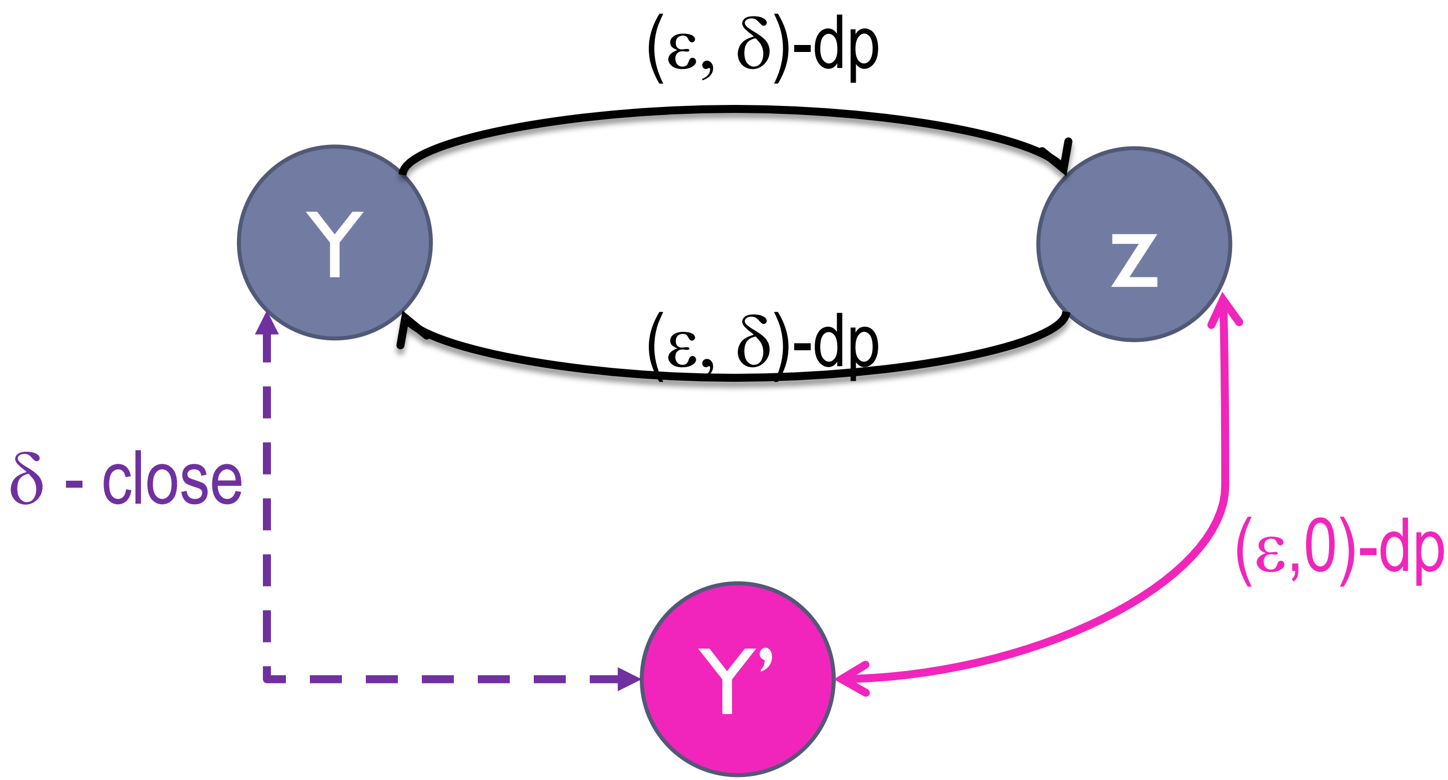
- Recall "Useful Lemma":

  Privacy loss bounded by $\epsilon \Rightarrow$ expected loss bounded by $2\epsilon^2$.

- Model cumulative privacy loss as a Martingale [Dinur,D.,Nissim'03]
  - Bound on max loss (ε) A
  - Bound on expected loss ($2\varepsilon^2$) B
  - $\Pr_{M1,\ldots,Mk}[\ |\ \sum_i \text{loss from } M_i\ | > z\sqrt{k}\ A\ + kB] < \exp(-z^2/2)$

Extension to (ε,δ)-dp mechanisms

- Reduce to previous case via "dense model theorem" [MPRV09]



Composition Theorem

- $\forall \epsilon, \delta, \delta'$: the k-fold composition of (ϵ, δ)-dp mechanisms is $\left(\sqrt{2k \ln\left(\frac{1}{\delta'}\right)}\ \epsilon + k\epsilon(e^{\epsilon} - 1),\ k\delta + \delta'\right)$-dp

- What is Bob's lifetime exposure risk?
  - Eg, 10,000 ϵ-dp or (ϵ, δ)-dp databases, for lifetime cost of $(1, \delta')$-dp
  - What should be the value of ϵ?
  - 1/801
  - OMG, that is small! Can we do better?

In the presentation

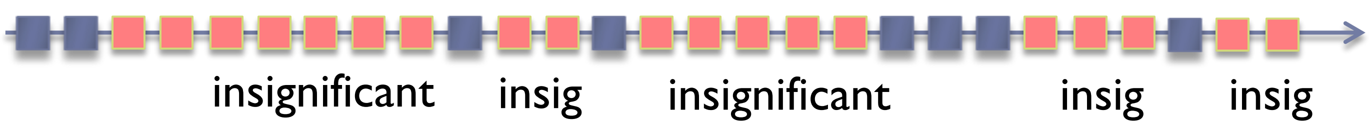
1. Intuitions
   - Anonymity means privacy?
   - A running example: Justin Bieber
   - What exactly does DP protects? Smoker Mary example.
2. What and how
   - ϵ-Differential Privacy
   - Global sensitivity and Laplace Mechanism
   - (ϵ, δ)-Differential Privacy and Composition Theorem
3. Many queries
   - A disappointing lower bound (Dinur-Nissim Attack 03)
   - Sparse-vector technique

Dinur-Nissim03 Attack

- Dinur and Nissim shows that the following negative results:
  - If adversary has exponential computational power (ask exp(n)) questions, a $O(n)$ perturbation is needed for privacy.
  - If adversary has poly(n) computation powers, at least $O(\sqrt{n})$ perturbation is needed.
- They also gave the following positive news:
  - If adversary can ask only less than $\mathrm{T}(n)$ questions, then a perturbation error of roughly $\sqrt{T(n)}$ is sufficient to guarantee privacy.

Sparse Vector Technique

- Database size n
- # Queries $m \gg n$, eg, m super-polynomial in n
- # "Significant" Queries $k \in O(n)$
  - For now: Counting queries only
  - Significant: count exceeds publicly known threshold T
- Goal: Find, and optionally release, counts for significant queries, *paying only for significant queries*



Algorithm and Privacy Analysis

[Hardt-Rothblum]

Caution:
Conditional branch leaks private information!

Need noisy threshold
$T + Lap(\sigma)$

Algorithm:

When given query f_t:

- If $f_t(x) \leq T$: [insignificant]
  - Output $\perp$
- Otherwise [significant]
  - Output $f_t(x) + Lap(\sigma)$

- First attempt: It's obvious, right?
  - Number of significant queries $k \Rightarrow \leq k$ invocations of Laplace mechanism
  - Can choose σ so as to get error $k^{1/2}$

Algorithm and Privacy Analysis

Caution:
Conditional branch leaks private information!

Algorithm:

When given query f_t:

- If $f_t(x) \leq T + Lap(\sigma)$: [insignificant]
  - Output $\perp$
- Otherwise [significant]
  - Output $f_t(x) + Lap(\sigma)$

- Intuition: counts far below T leak nothing
  - Only charge for noisy counts in this range:

Sparse Vector Technique

Expected total privacy loss $EX = O(\frac{k}{\sigma^2})$

- Probability of (significantly) exceeding expected number of borderline events is negligible (Chernoff)
- Assuming not exceeded: Use Azuma to argue that whp actual total loss does not significantly exceed expected total loss
- Utility: With probability at least $1 - \beta$ all errors are bounded by $\sigma(\ln m + \ln(\frac{1}{\beta}))$.
- Choose $\sigma = 8\sqrt{2\ln(\frac{2}{\delta})(4k + \ln\left(\frac{2}{\delta}\right))}/\epsilon$
  - Linear in k, and only log in m!

In the presentation

4. **Advanced Techniques**
   - Local sensitivity
   - Sample and Aggregate
   - Exponential mechanism and Net-mechanism
5. **Diff-Private in Machine Learning**
   - Diff-Private logistic regression (Perturb objective)
   - Diff-Private low-rank approximation
   - Diff-Private PCA (use Exponential Mechanism)
   - Diff-Private SVM

Large sensitivity queries

- Thus far, we've been consider counting queries or similarly low sensitivity queries.
- What if the query itself is of high sensitivity?
  - What is the age of the oldest person who like Justin Bieber?
  - Sensitivity is 50? 100? It's not even well defined.
  - Can use histogram: <10, 10-20, 20-30, 30-45, >45
- If data is unbounded, what is the sensitivity of PCA?
  - 1st Principal components can turn 90 degrees!

Local sensitivity/smooth sensitivity

- The sensitivity depends on f but not the range of data.
- Consider f= median income

$$D = \{\underbrace{0,0,\ldots,0}_{\frac{n-1}{2}},0,\underbrace{k,k,\ldots,k}_{\frac{n-1}{2}}\} \quad D' = \{\underbrace{0,0,\ldots,0}_{\frac{n-1}{2}},k,\underbrace{k,k,\ldots,k}_{\frac{n-1}{2}}\}$$

  - Global sensitivity is k! Perturb by Lap(k/ϵ) gives no utility at all.
- For typical data however we may do better:

$$D = \{1,2,2,\ldots,\frac{k}{2},\frac{k}{2},\frac{k}{2},\frac{k}{2}+1,\ldots,k,k\}$$

The local sensitivity of a function $f : 2^X \to \mathbf{R}$ *at a database* D *is:*

$$LS_f(D) = \max_{D' \in N(D)} |f(D) - f(D')|$$

Local sensitivity/smooth sensitivity

- Local sensitivity is defined to particular data D, but adding $Lap(LS_f(D)/\epsilon)$ doesn't guarantee ϵ-dp.
  - Because $LS_f(D)$ itself is sensitive to the values in D!

$$D = \{\underbrace{0,0,\ldots,0}_{\frac{n-1}{2}},0,0,\underbrace{k,k,\ldots,k}_{\frac{n-1}{2}-1}\} \quad D' = \{\underbrace{0,0,\ldots,0}_{\frac{n-1}{2}},0,\underbrace{k,k,\ldots,k}_{\frac{n-1}{2}}\}$$

$$LS_{\text{median}}(D) = 0, \text{ but } LS_{\text{median}}(D') = k.$$

$$\Pr[f(D) + \text{Lap}(LS_f(D)/\epsilon) = 0] = 1 \qquad \Pr[f(D') + \text{Lap}(LS_f(D')/\epsilon) = 0] = 0,$$

- Solution: Smooth the upper bound of LS_f

Smooth sensitivity

(Smooth Sensitivity) *For $\beta > 0$ the β-smooth sensitivity of f is:*

$$S^*_{f,\beta}(D) = \max_{D' \subset X} LS_f(D') \exp(-\beta d(D', D))$$

Theorem: Let S_f be an ϵ-smooth upper bound on f. Then an algorithm that output:

$$M(D) = f(D) + Lap(S_f(D)/\epsilon)$$

is 2ϵ-differentially private.

- Simple proof similar to original Laplace mechanism proof.

Subsample-and-Aggregate mechanism

- Subsample-and-Aggregate [Nissim, Raskhodnikova, Smith'07]

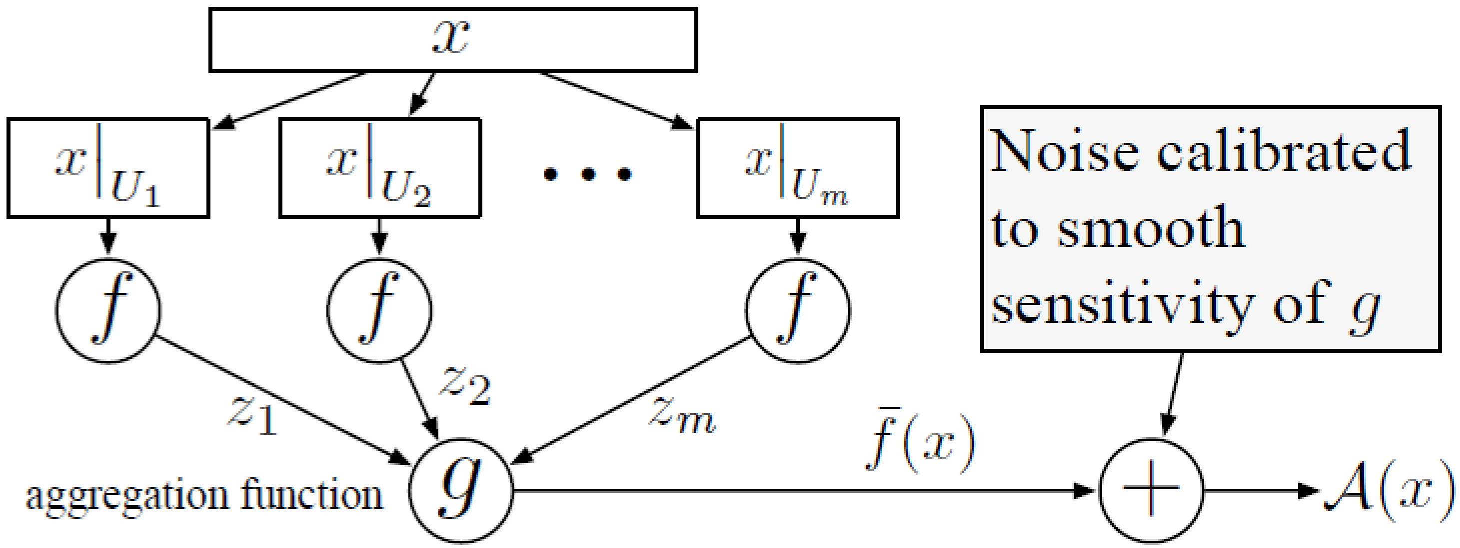
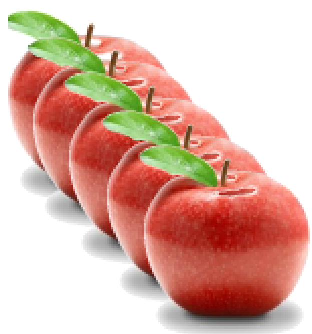


Figure 2: The Sample-Aggregate Framework

Beyond perturbation

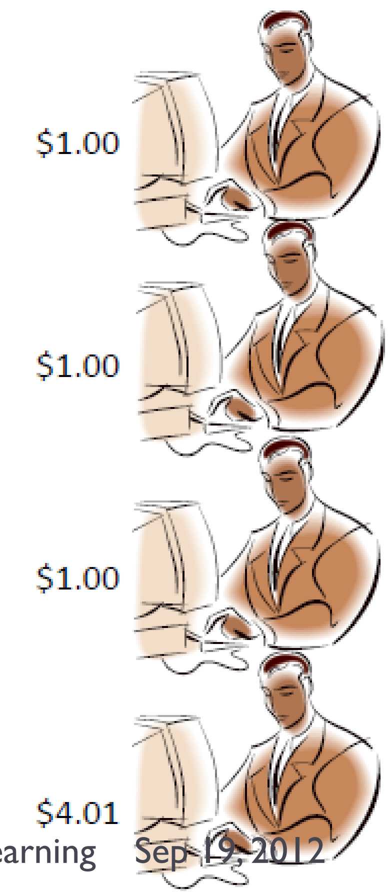
- Discrete-valued functions: $f(x) \in R = \{y_1, y_2, \dots, y_k\}$
  - Strings, experts, small databases, …
- Auction:



Could set the price of apples at \$1.00 for profit: \$4.00

Could set the price of apples at \$4.01 for profit \$4.01

Best price: \$4.01
2nd best price: \$1.00
Profit if you set the price at \$4.02: \$0
Profit if you set the price at \$1.01: \$1.01



Exponential Mechanism

- Define utility function:
  - Each $y \in \mathrm{R}$ has a utility for x, denoted $\mathrm{q}(x, y)$

- Exponential Mechanism [McSherry-Talwar'07]

  Output y with probability $\propto e^{\frac{\epsilon q(x,y)}{2\Delta \mathrm{q}}}$

  - Idea: Make high utility outputs exponentially more likely at a rate that depends on the sensitivity of $\mathrm{q}(x, y)$.

Exponential Mechanism

Theorem: The Exponential Mechanism preserves $(\epsilon, 0)$-differential privacy.

Proof: Fix any $D, D' \in \mathbb{N}^{|X|}$ with $\left\| D, D' \right\|_1 \leq 1$ and any $r \in R$...

$$\frac{\Pr[\text{Exponential}(D, R, q, \epsilon) = r]}{\Pr[\text{Exponential}(D', R, q, \epsilon) = r]} =$$

$$\frac{\left(\dfrac{\exp(\frac{\epsilon q(D,r)}{2\Delta})}{\sum \exp(\frac{\epsilon q(D,r')}{2\Delta})} \right)}{\left(\dfrac{\exp(\frac{\epsilon q(D',r)}{2\Delta})}{\sum \exp(\frac{\epsilon q(D',r')}{2\Delta})} \right)} = \left(\frac{\exp(\frac{\epsilon q(D,r)}{2\Delta})}{\exp(\frac{\epsilon q(D',r)}{2\Delta})} \right) \left(\frac{\sum_{r'} \exp(\frac{\epsilon q(D',r')}{2\Delta})}{\sum_{r'} \exp(\frac{\epsilon q(D,r')}{2\Delta})} \right)$$

Exponential Mechanism

$$\star = \left(\frac{\exp(\frac{\epsilon q(D,r)}{2\Delta})}{\exp(\frac{\epsilon q(D',r)}{2\Delta})}\right) =$$

$$\exp\left(\frac{\epsilon(q(D,r) - q(D',r))}{2\Delta}\right) \leq$$

$$\exp\left(\frac{\epsilon\Delta}{2\Delta}\right) = \exp\left(\frac{\epsilon}{2}\right)$$

$$\star\star = \left(\frac{\sum_{r'} \exp(\frac{\epsilon q(D',r')}{2\Delta})}{\sum_{r'} \exp(\frac{\epsilon q(D,r')}{2\Delta})}\right) \leq$$

$$\left(\frac{\sum_{r'} \exp(\frac{\epsilon(q(D,r') + \Delta)}{2\Delta})}{\sum_{r'} \exp(\frac{\epsilon q(D,r')}{2\Delta})}\right) =$$

$$= \left(\frac{\exp(\frac{\epsilon}{2}) \sum_{r'} \exp(\frac{\epsilon q(D,r')}{2\Delta})}{\sum_{r'} \exp(\frac{\epsilon q(D,r')}{2\Delta})}\right) = \exp(\frac{\epsilon}{2})$$

(α, β)-usefulness of a private algorithm

- A mechanism M is (α, β)-useful with respect to queries in class C if for every database $D \in N^{|X|}$ with probability at least $1 - \beta$, the output

$$\max_{Q_i \in C} |Q_i(D) - M(Q_i, D)| \leq \alpha$$

- So it is to compare the private algorithm with non-private algorithm in PAC setting.
- A remark here: The tradeoff privacy may be absorbed in the inherent noisy measurement! Ideally, there can be no impact scale-wise!

Usefulness of Exponential Mechanism

- How good is the output?

Define:

$$OPT_q(D) = \max_{r \in R} q(D, r)$$
$$R_{OPT} = \{r \in R : q(D, r) = OPT_q(D)\}$$
$$r^* = \text{Exponential}(D, R, q, \epsilon)$$

Theorem:

$$\Pr\left[q(r^*) \le OPT_q(D) - \frac{2\Delta}{\epsilon}\left(\log\left(\frac{|R|}{|R_{OPT}|}\right) + t\right)\right] \le e^{-t}$$

- The results depends ONLY on Δ (logarithm to $|R|$).
- Example: counting query. What is the majority gender that likes Justin Bieber? $|R| = 2$
  - Error is $\frac{2}{\epsilon}(\log(2) + 5)$ with probability $1 - e^{-5}$! Percent error $\rightarrow 0$, when number of data become large.

Net Mechanism

Many (fractional) counting queries [Blum, Ligett, Roth'08]:

Given n-row database x, set Q of properties, produce a synthetic database y giving good approx to "What fraction of rows of x satisfy property P?" $\forall P \in Q$.

- S is set of all databases of size $m \in \tilde{O}(\log|Q|/\alpha^2) \ll n$
- $u(x, y) = -\max_{q \in Q} |q(x) - q(y)|$

- The size of m is the α-net cover number of D with respect to query class Q.

Net Mechanism

- Usefulness

For any class of queries C the Net Mechanism is $(2\alpha, \beta)$-useful for any α

$$\alpha \geq \frac{2\Delta}{\epsilon} \log \frac{N_\alpha(C)}{\beta} \qquad \textit{Where } \Delta = \max_{Q \in C} GS(Q).$$

- For counting queries $|N_\alpha(C)| \leq |X|^{\frac{\log |C|}{\alpha^2}}$
- Logarithm to number of queries! Private to exponential number of queries!
- Well exceeds the fundamental limit of Dinur-Nissim03 for perturbation based privacy guarantee. (why?)

Other mechanisms

- Transform to Fourier domain then add Laplace noise.
  - Contingency table release.
- SuLQ mechanism use for any sublinear Statistical Query Model algorithms
  - Examples includes PCA, k-means and Gaussian Mixture Model
- Private PAC learning with Exponential Mechanism
  - for all Classes with Polynomial VC-dimension
  - Blum, A., Ligett, K., Roth, A.: A Learning Theory Approach to Non-Interactive Database Privacy (2008)

In the presentation

4. Advanced Techniques
   - Local sensitivity
   - Sample and Aggregate
   - Exponential mechanism and Net-mechanism
5. Diff-Private in Machine Learning
   - Diff-Private PCA (use Exponential Mechanism)
   - Diff-Private low-rank approximation
   - Diff-Private logistic regression (Perturb objective)
   - Diff-Private SVM

Structure of a private machine learning paper

- Introduction:
  - Why the data need to be learnt privately.
- Main contribution:
  - Propose a randomized algorithm.
  - Show this randomization indeed guarantees (ϵ, δ)-dp.
  - Show the sample complexity/usefulness under randomization.
- Evaluation:
  - Compare to standard Laplace Mechanism (usually Laplace mechanism is quite bad.)
  - Compare to non-private algorithm and say the deterioration in performance is not significant. (Argue it's the price of privacy.)

Differential Private-PCA

- K. Chaudhuri, A. D. Sarwate, K. Sinha, Near-optimal Differential Private PCA (NIPS'12): http://arxiv.org/abs/1207.2812
  - An instance of Exponential mechanism
  - Utility function is defined such that output close to ordinary PCA output is exponentially more likely.
  - Sample from Bingham distribution using Markov Chain Monte Carlo procedure.

- Adding privacy as a trait to RPCA?

Differential Private Low-Rank Approximation

- Moritz Hardt, Aaron Roth, Beating Randomized Response on Incoherent Matrices (STOC'12) http://arxiv.org/abs/1111.0623
  - Motivated by Netflix challenge, yet they don't assume missing data/matrix completion setting, but study general low rank approx.
  - Privatize Tropp's 2-step Low-rank approximation by adding noise. Very dense analysis, but not difficult.
  - Assume the sparse matrix itself is incoherent.

Differentially private logistic regression

- K. Chaudhuri, C. Monteleoni, Privacy-preserving logistic regression (NIPS'08) http://www1.ccls.columbia.edu/~cmontel/cmNIPS2008.pdf
- Journal version: Privacy-preserving Empirical Risk Minimization (JMLR 2011) http://jmlr.csail.mit.edu/papers/volume12/chaudhuri11a/chaudhuri11a.pdf

- I'd like to talk a bit more on this paper as a typical example of DP machine learning paper.
  - The structure is exactly what I described a few slides back.

Differentially private logistic regression

- Refresh on logistic regression
- Input:
  - $\{x_1, \dots, x_n\}$, each $x_i \in R^d$ and $\|x_i\| \le 1$
  - $\{y_1, \dots, y_n\}$, $y_i \in \{-1,1\}$ are class labels assigned to each x_i.
- Output:
  - Vector $w \in R^d$, $SGN(w^T x)$ gives the predicted classification of a point x.
- Algorithm for logistic regression:
  - $w^* = \underset{w}{\operatorname{argmin}} \frac{1}{2}\lambda w^T w + \frac{1}{n}\sum_{i=1}^{n} \log\left(1 + e^{-y_i w^T x_i}\right)$
  - $\hat{f}_\lambda(w) = \frac{1}{2}\lambda w^T w + \frac{1}{n}\sum_{i=1}^{n} \log\left(1 + e^{-y_i w^T x_i}\right) = \frac{1}{2}\lambda\|w\|^2 + \hat{L}(w)$

Differentially private logistic regression

- Results perturbation approach
  - Sensitivity:

$$\max_x |w^*(X,Y) - w^*([X,x],[Y,y]\,)| \leq \frac{2}{n\lambda}$$

  - Algorithm Laplace Mechanism: $h(\eta) \propto e^{-\frac{n\epsilon\lambda}{2}||\eta||}$.
    - 1. choose the norm of η from the $\Gamma(d, \frac{2}{n\epsilon\lambda})$ distribution
    - 2. direction of η uniformly at random.
    - 3. Output $w^* + \eta$.
  - Usefulness: with probability $1 - \delta$

$$\hat{f}_\lambda(w_2) \leq \hat{f}_\lambda(w_1) + \frac{2d^2(1+\lambda)\log^2(d/\delta)}{\lambda^2 n^2 \epsilon^2}$$

Differentially private logistic regression

- Objective perturbation approach
  - Algorithm: $h(b) \propto e^{-\frac{\epsilon}{2}||b||}$
    1. pick the norm of b from the $\Gamma(d, \frac{2}{\epsilon})$ distribution
       the direction of b uniformly random.
    2. Output

$$w^* = \text{argmin}_w \frac{1}{2}\lambda w^T w + \frac{b^T w}{n} + \frac{1}{n}\sum_{i=1}^n \log(1 + e^{-y_i w^T x_i})$$

- Observe:
  - Size of perturbation is independent to sensitivity! And independent to λ.
  - When $n \to \infty$, this optimization is consistent.

Differentially private logistic regression

- Theorem: The objective perturbation approach preserves ϵ-differential privacy.
- Proof:
  - Because both regularization and loss function are differentiable everywhere. There is a unique b for any output w^*.
  - Consider two adjacent databases differing only at one point, we have b_1 and b_2 that gives w^*.
  - Because b_1 and b_2 both gives zero derivative at w^*. We have an equation. Further with Triangular inequality,

$$-2 \leq \|b_1\| - \|b_2\| \leq 2$$

  - Lastly, by definition:

$$\frac{\Pr[w^*|x_1, \ldots, x_{n-1}, y_1, \ldots, y_{n-1}, x_n = a, y_n = y]}{\Pr[w^*|x_1, \ldots, x_{n-1}, y_1, \ldots, y_{n-1}, x_n = a', y_n = y']} = \frac{h(b_1)}{h(b_2)} = e^{-\frac{\epsilon}{2}(||b_1||-||b_2||)}$$

Differentially private logistic regression

- Generalization to a class of convex objective functions!

$$F(w) = G(w) + \sum_{i=1}^{n} l(w, x_i)$$

1. $G(w)$ and $l(w, x_i)$ *are differentiable everywhere, and have continuous derivatives*
2. $G(w)$ *is strongly convex and* $l(w, x_i)$ *are convex for all* i
3. $||\nabla_w l(w, x)|| \leq \kappa$, *for any* x.

- Proof is very similar to the special case in Logistic Regression.
- However wrong, corrected in their JMLR version…

Differentially private logistic regression

- Learning Guarantee:

$$\hat{f}_\lambda(w_2) \leq \hat{f}_\lambda(w_1) + \frac{8d^2 \log^2(d/\delta)}{\lambda n^2 \epsilon^2}$$

- Generalization bound (assume iid drawn from distribution)

$$if\, n > C \max(\frac{||w_0||^2}{\epsilon_g^2}, \frac{d \log(\frac{d}{\delta})||w_0||}{\epsilon_g \epsilon}),$$

  - With probability $1-\delta$, classification output is at most

$$most\ L + \epsilon_g\ over\ the\ data\ distribution$$

- Proof is standard by Nati Srebro's NIPS'08 paper about regularized objective functions.

Differentially private logistic regression

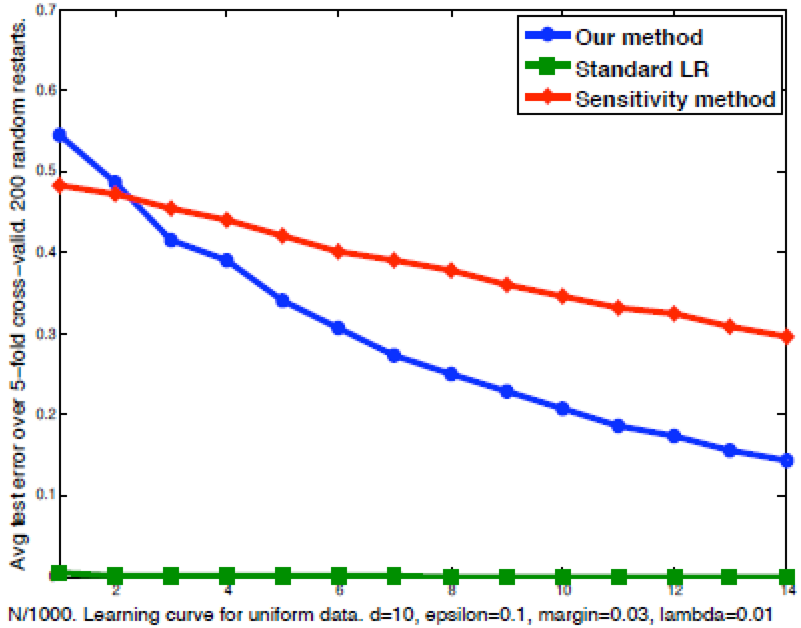
- Similar generalization bound is given for "results perturbation approach"
  - A: $$n > C \max\left(\frac{||w_0||^2}{\epsilon_g^2}, \frac{d \log(\frac{d}{\delta})||w_0||}{\epsilon_g \epsilon}, \frac{d \log(\frac{d}{\delta})||w_0||^2}{\epsilon_g^{3/2} \epsilon}\right)$$
- To compare with the bound for the proposed objective perturbation:
  - B: $$n > C \max\left(\frac{||w_0||^2}{\epsilon_g^2}, \frac{d \log(\frac{d}{\delta})||w_0||}{\epsilon_g \epsilon}\right),$$
  - A is always greater or equal to B
  - In low-loss (high accuracy) cases where $\|w_0\| > 1$, A is much worse than B

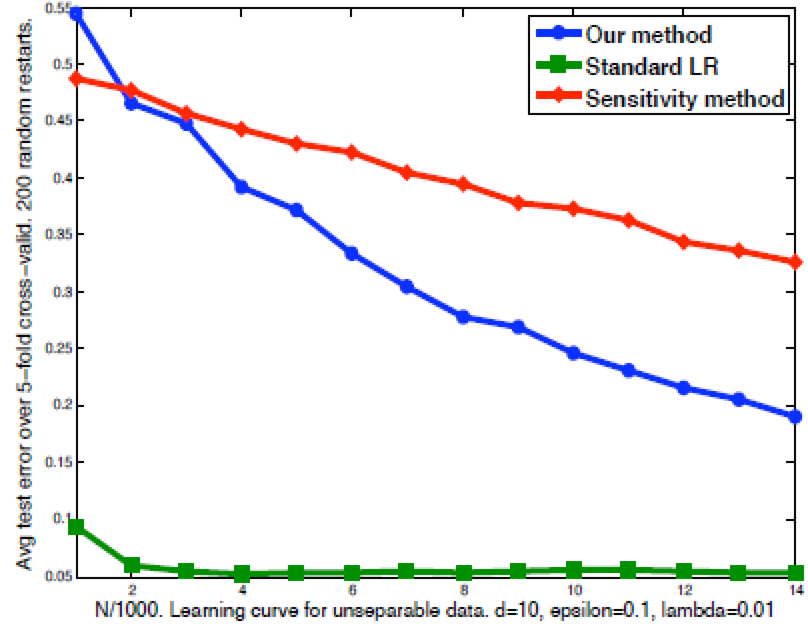
Differentially private logistic regression

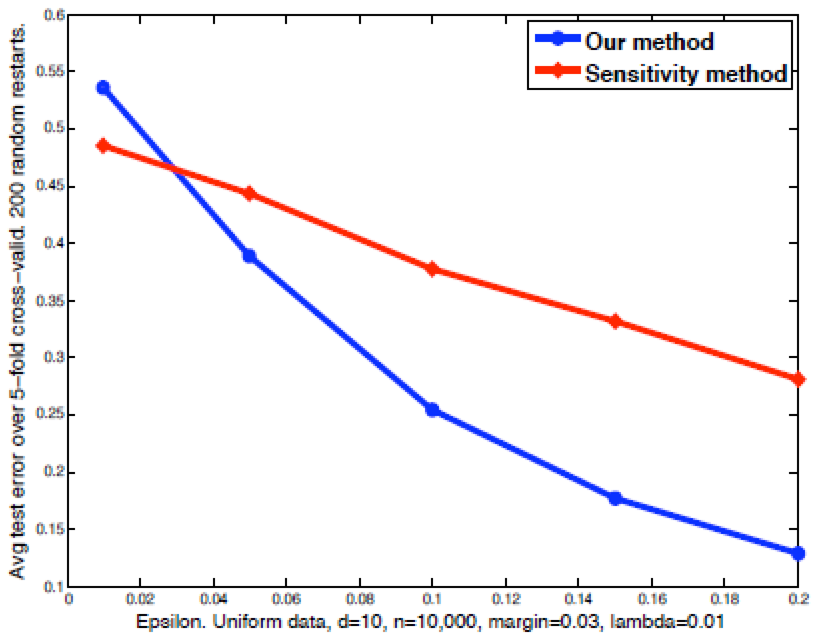
- Simulation results:
  - **Separable:** Random data on hypersphere with small 0.03 gap separating two labels.
  - **Unseparable:** Random data on hypersphere with 0.1 gap, then 0.2 Probability random flipping.

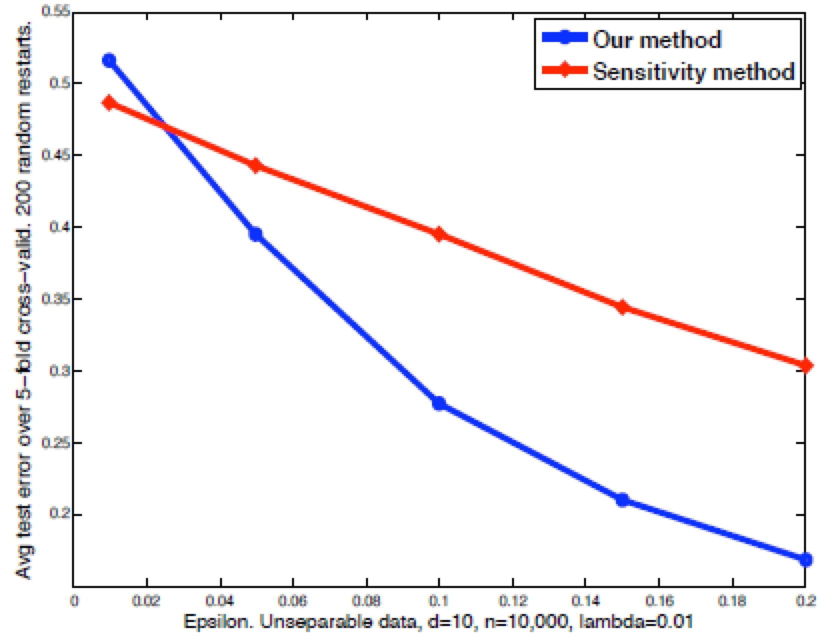
	Uniform, margin=0.03	Unseparable (uniform with noise 0.2 in margin 0.1)
Sensitivity method	0.2962 ± 0.0617	0.3257 ± 0.0536
New method	0.1426 ± 0.1284	0.1903 ± 0.1105
Standard LR	0 ± 0.0016	0.0530 ± 0.1105

Figure 1: Test error: mean $\pm$ standard deviation over five folds. N=17,500.


Our method
Standard LR
Sensitivity method
Avg test error over 5−fold cross−valid. 200 random restarts.
N/1000. Learning curve for uniform data. d=10, epsilon=0.1, margin=0.03, lambda=0.01


Our method
Standard LR
Sensitivity method
Avg test error over 5−fold cross−valid. 200 random restarts.
N/1000. Learning curve for unseparable data. d=10, epsilon=0.1, lambda=0.01


Our method
Sensitivity method
Avg test error over 5−fold cross−valid. 200 random restarts.
Epsilon. Uniform data, d=10, n=10,000, margin=0.03, lambda=0.01


Our method
Sensitivity method
Avg test error over 5−fold cross−valid. 200 random restarts.
Epsilon. Unseparable data, d=10, n=10,000, lambda=0.01

Differential Private-ERM in high dimension

- Follow-up:
  - D. Sarwate, K. Chaudhuri, C. Monteleoni, Differentially Private Support Vector Machines
    - Extend "objective perturbation" to larger class of convex method
    - Private non-linear kernel SVM
  - D. Kifer, A. Smith, A. Thakurta, Private Convex Empirical Risk Minimization and High-dimensional Regression (COLT'12)
    - Extend the "objective perturbation" to smaller added noise, and apply to problem with non-differentiable regularizer.
    - Best algorithm for private linear regression in low-dimensional setting.
    - First DP-sparse regression in high-dimensional setting.

Reiterate the key points

- What does Differential Privacy protect against?
  - Deconstruct harm. Minimize risk of joining a database.
  - Protect all personal identifiable information.
- Elements of (ϵ, δ)-dp
  - Global sensitivity (a function of f) and Smooth (local) sensitivity (a function of f and D)
  - Composition theorem (roughly $\sqrt{k}\epsilon$ for k queries)
- Algorithms
  - Laplace mechanism (perturbation)
  - Exponential mechanism (utility function, net-mechanism)
  - Sample and aggregate (for unknown sensitivity)
  - Objective perturbation (for many convex optimization based learning algorithms)

Take-away from this Tutorial

- Differential privacy as a new design parameter for algorithm
  - Provable (In fact, I've no idea how DP can be evaluated by simulation).
  - No complicated math. (Well, it can be complicated…)
  - Relevant to key strength of our group (noise, corruption robustness)
- This is a relatively new field (as in machine learning).
- Criticisms:
  - The bound is a bit paranoid (assume very strong adversary).
  - Hard/impossible to get practitioners to use it as accuracy is sacrificed. (Unless there's a legal requirement.)

Questions and Answers