Data Mining 2017: Privacy and Differential Privacy
Slides stolen from Kobbi Nissim (with permission)

Also used: The Algorithmic Foundations of
Differential Privacy by Dwork and Roth
Data Privacy – The Problem

• Given:
  • a dataset with sensitive information
    • Health records, census data, financial data, ...

• How to:
  • Compute and release functions of the dataset,
  • Without compromising individual privacy.
Data Privacy – The Problem

Individuals

$x_1$
$x_2$
$\vdots$
$x_n$

Server/agency

A

Users

Government, researchers, businesses (or) Malicious adversary

queries
answers
A Real Problem

Typical examples:
• Census
• Civic archives
• Medical records
• Search information
• Communication logs
• Social networks
• Genetic databases
• …

Benefits:
• New discoveries
• Improved medical care
• National security
The Anonymization Dream

- **Trusted curator:**
  - Removes identifying information (name, address, ssn, ...).
  - Replaces identities with random identifiers.
- **Idea hard wired into practices, regulations, ..., thought.**
  - Many uses.
- **Reality:** *series failures.*
  - Pronounced both in academic and public literature.
Data Privacy

• Studied (at least) from the 60s
• Approaches: De-identification, redaction, auditing, noise addition, synthetic data sets …
  • Focus on how to achieve privacy, not on what privacy protection is
  • May have been suitable for the pre-internet era
• Re-identification [Sweeney ‘00, …]
  • GIS data, health data, clinical trial data, DNA, Thesaurus data, text data, registry information, …
• Blatant non-privacy [Dinur, Nissim ‘03, …]
• Auditors [Kenthapadi, Mishra, Nissim ‘05]
• AOL Debacle ‘06
• Genetic wide association studies (GWAS) [Homer et al. ‘08]
• Netflix award [Narayanan, Shmatikov ‘09]
  • Netflix canceled second contest
• Social networks [Backstrom, Dwork, Kleinberg ‘11]
• Genetic research studies [Gymrek, McGuire, Golan, Halperin, Frich ‘11]
• Microtargeted advertising [Korolova ‘11]
• Recommendation Systems [Calandrino, Kiltzer, Narayanan, Kelten, Shmatikov ‘11]
• Israeli CBS [Mukatren, Nissim, Salman, Tromer ‘11]
• Attack on statistical aggregates [Homer et al. ‘08] [Dwork, Smith, Steinke, Vadhan ‘15]
• …

Slide idea stolen shamelessly from Or Sheffet
Linkage Attacks [Sweeney 2000]

Anonymized GIC data

- GIC Group Insurance Commission
- Ethnicity
- Visit date
- Diagnosis
- Procedure
- Medication
- Total Charge

Anonymized GIC data (135,000 patients)
- 100 attributes per encounter

Voter registration of Cambridge MA
- Name
- Address
- Date registered
- Party affiliation
- Date last voted

"Public records" open for inspection by anyone
Linkage Attacks [Sweeney 2000]

- Quasi identifiers $\rightarrow$ re-identification
  - Not a coincidence:
    - dob+5zip $\rightarrow$ 69%
    - dob+9zip $\rightarrow$ 97%

- William Weld (governor of Massachusetts at the time)
  - According to the Cambridge Voter list:
    - Six people had his particular birth date
    - Of which three were men
    - He was the only one in his 5-digit ZIP code!
AOL Data release (2006)

- AOL released search data
  - A sample of ~20M web queries collected from ~650k users over three months
- Goal: provide real query log data that is based on real users
  - “It could be used for personalization, query reformulation or other types of search research”
- The data set:

<table>
<thead>
<tr>
<th>AnonID</th>
<th>Query</th>
<th>QueryTime</th>
<th>ItemRank</th>
<th>ClickURL</th>
</tr>
</thead>
</table>

A Face Is Exposed for AOL Searcher No. 4417749

By MICHAEL BARBARO and TOM ZELLER Jr.
Published: August 9, 2006

Buried in a list of 20 million Web search queries collected by AOL and recently released on the Internet is user No. 4417749. The number was assigned by the company to protect the searcher’s anonymity, but it was not much of a shield.

No. 4417749 conducted hundreds of searches over a three-month period on topics ranging from “n umb fingers” to “60 single men” to “dog that urinates on

Name: Thelma Arnold
Age: 62
Widow
Residence: Lilburn, GA
Other Re-Identification Examples
[partial and unordered list]

• Netflix award [Narayanan, Shmatikov 08].
• Social networks [Backstrom, Dwork, Kleinberg 07, NS 09].
• Computer networks [Coull, Wright, Monrose, Collins, Reiter ’07, Ribeiro, Chen, Miklau, Townsley 08].
• Genetic data (GWAS) [Homer, Szelinger, Redman, Duggan, Tembe, Muehling, Pearson, Stephan, Nelson, Craig 08, ...].
• Microtargeted advertising [Korolova 11].
• Recommendation Systems [Calandrino, Kiltzer, Naryanan, Felten, Shmatikov 11].
• Israeli CBS [Mukatren, N, Salman, Tromer].
• ...
k-Anonymity [SS98,S02]

- Prevent re-identification:
  - Make every individual’s identity unidentifiable from other k-1 individuals

<table>
<thead>
<tr>
<th>ZIP</th>
<th>Age</th>
<th>sex</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>23456</td>
<td>55</td>
<td>Female</td>
<td>Heart</td>
</tr>
<tr>
<td>12345</td>
<td>30</td>
<td>Male</td>
<td>Heart</td>
</tr>
<tr>
<td>12346</td>
<td>33</td>
<td>Male</td>
<td>Heart</td>
</tr>
<tr>
<td>13144</td>
<td>45</td>
<td>Female</td>
<td>Breast Cancer</td>
</tr>
<tr>
<td>13155</td>
<td>42</td>
<td>Male</td>
<td>Hepatitis</td>
</tr>
<tr>
<td>23456</td>
<td>42</td>
<td>Male</td>
<td>Viral</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ZIP</th>
<th>Age</th>
<th>sex</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>23456</td>
<td>**</td>
<td>*</td>
<td>Heart</td>
</tr>
<tr>
<td>1234*</td>
<td>3*</td>
<td>Male</td>
<td>Heart</td>
</tr>
<tr>
<td>1234*</td>
<td>3*</td>
<td>Male</td>
<td>Heart</td>
</tr>
<tr>
<td>131**</td>
<td>4*</td>
<td>*</td>
<td>Breast Cancer</td>
</tr>
<tr>
<td>131**</td>
<td>4*</td>
<td>*</td>
<td>Hepatitis</td>
</tr>
<tr>
<td>23456</td>
<td>**</td>
<td>*</td>
<td>Viral</td>
</tr>
</tbody>
</table>

Both guys from zip 1234* that are in their thirties have heart problems

My (male) neighbor from zip 13155 has hepatitis!

Bugger! I Cannot tell which disease for the patients from zip 23456
Auditing

Here’s a new query: $q_{i+1}$

Here’s the answer

OR

Query denied (as the answer would cause privacy loss)
Example 1: Sum/Max auditing

dᵢ real, sum/max queries, privacy breached if some dᵢ learned

q1 = \text{sum}(d1,d2,d3)

\text{sum}(d1,d2,d3) = 15

q2 = \text{max}(d1,d2,d3)

Denied (the answer would cause privacy loss)

Oh well...
... After Two Minutes ...

\[ d_i \text{ real, sum/max queries, privacy breached if some } d_i \text{ learned} \]

\[ q_1 = \sum(d_1,d_2,d_3) \]

\[ q_2 = \max(d_1,d_2,d_3) \]

\[ \text{sum}(d_1,d_2,d_3) = 15 \]

Denied (the answer would cause privacy loss)

There must be a reason for the denial.

On well...
Example 2: Interval Based Auditing

d_i \in [0,100], sum queries, error \epsilon = 1

q1 = \text{sum}(d1,d2)

q2 = \text{sum}(d2,d3)

\text{sum}(d2,d3) = 50

Sorry, denied

Auditor

\epsilon = 1
Max Auditing

q1 = max(d1,d2,d3,d4)

q2 = max(d1,d2,d3)

If denied: d4=M_{1234}

q2 = max(d1,d2)

If denied: d3=M_{123}

M_{1234}

M_{123} / denied

M_{12} / denied

Auditor
Adversary’s Success

q1 = \max(d1,d2,d3,d4)

q2 = \max(d1,d2,d3)

q2 = \max(d1,d2)

If denied: \(d4 = M_{1234}\)

Denied with probability \(1/4\)

If denied: \(d3 = M_{123}\)

Denied with probability \(1/3\)

Success probability: \(1/4 + (1 - 1/4) \cdot 1/3 = 1/2\)

Recover 1/8 of the database!
Thanks to Miri Regev
Thanks to Google
Randomization
The Scenario

• Users provide modified values of sensitive attributes

• Dataminer develops models about aggregated data

Can we develop accurate models without access to precise individual information?
Preserving Privacy

• Value distortion – return $x_i + r_i$
  • Uniform noise: $r_i \sim \mathcal{U}(-a,a)$
  • Gaussian noise: $r_i \sim \mathcal{N}(0,\text{stdev})$
  • Perturbation of an entry is fixed
    • So that repeated queries do not reduce noise

• Privacy quantification: interval of confidence [AS 2000]
  • With c% confidence $x_i$ is in the interval $[a_1, a_2]$. $a_2 - a_1$ define the amount of privacy at c% confidence level
  • Examples:

<table>
<thead>
<tr>
<th></th>
<th>50%</th>
<th>95%</th>
<th>99.9%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform</td>
<td>0.5 x 2a</td>
<td>0.95 x 2a</td>
<td>0.999 x 2a</td>
</tr>
<tr>
<td>Gaussian</td>
<td>1.34 stdev</td>
<td>3.92 stdev</td>
<td>6.8 stdev</td>
</tr>
</tbody>
</table>

• Intuition: the larger the interval is, the better privacy is preserved.
Prior knowledge affects privacy

• Let $X=\text{age } r_i \sim \mathcal{U}(-50,50)$
• [AS]: Privacy 100 at 100% confidence
• Seeing a measurement -10
  • Facts of life: Bob’s age is between 0 and 40
• Assume you also know Bob has two children
  • Bob’s age is between 15 and 40
• a-priori information may be used in attacking individual data
Privacy in Computation – The Problem

- Given a dataset with sensitive personal information
- How to compute and release functions of the dataset
- While protecting individual privacy

Health, social n/w, location, communication, ...

Academic research, informed policy, national security, …
How many Trump fans in class? Sensitive information ...
3 Trump fans in class
A few minutes later – Kobbi join the class.

A survey!

How many Trump fans?

Using SMPC!

What are you doing?

Me too!
3 Trump fans

(after @Kobbi joins): 4 Trump fans
Composition

• Differencing attack:
  • How is my privacy affected when an attacker sees analysis before and after I join/leave?

• More generally: Composition
  • How is my privacy affected when an attacker combines results from two or more privacy preserving analyses?

• Fundamental law of information: the more information we extract from our data, the more is learned about individuals!
  • So, privacy will deteriorate as we use our data more and more
  • Best desiderata:
    • Deterioration is quantifiable and controllable
    • Deterioration not abrupt
My Privacy Desiderata

Real world:

Data -> Analysis (Computation) -> Outcome

My ideal world:

Data w/my info removed -> Analysis (Computation) -> Outcome

same outcome
Things to Note

• We only consider the *intended outcome* of analyses
  • Security flaws, hacking, implementation errors, human factors, ...
    • Very important but very *different* questions

• My ideal world would hide whether I’m Trump fan
  • Resilient to differencing attacks

• *Does not mean* I’m fully protected
  • I’m only protected to the extent I’m protected in my ideal world
  • Some harm could happen to me even in my ideal world
    • Bob smokes in public
    • Study teaches that smoking causes cancer
    • Bob’s health insurer raises his premium
    • Bob is harmed even if he does not participate in the study!
Our Privacy Desiderata

Real world:

My ideal world:
Our Privacy Desiderata

Real world:

Gert's info removed

Analysis (Computation)

Outcome

Gert’s ideal world:

Data w/Gert’s info removed

Analysis (Computation)

Outcome

Should ignore Kobbi’s info and Gertrude’s!

same outcome
Our Privacy Desiderata

Real world:

Mark’s ideal world:

Should ignore Kobbi’s info and Gertrude’s! and Mark’s!
Our Privacy Desiderata

Real world:

Data → Analysis (Computation) → Outcome

J's info removed

Smiley's ideal world:

Data w/Smiley's info removed → Analysis (Computation) → Outcome

Should ignore Kobbi's info and Gertrude's! and Mark's! ... and everybody's!

same outcome
A Realistic Privacy Desiderata

Real world:

Data → Analysis (Computation) → Outcome

 skljk j[formula]

’s ideal world:

Data w/ j's info removed → Analysis (Computation) → Outcome

’s "data"
Differential Privacy [Dwork McSherry N Smith 06]

Real world:

Data → Analysis (Computation) → Outcome

(J’s info removed)

(J’s ideal world):

Data w/ J’s info removed → Analysis (Computation) → Outcome

(ε, δ)-"similar"
A (randomized) algorithm $M : X^n \to T$ satisfies $(\epsilon, \delta)$-differential privacy if for all datasets $x, x' \in X^n$ that differ on one entry, For all subsets $S$ of the outcome space $T$, 

$$\Pr_{M}[M(x) \in S] \leq e^{\epsilon} \Pr_{M}[M(x') \in S] + \delta.$$ 

- The parameter $\epsilon$ measures ‘leakage’ or ‘harm’
- For small $\epsilon$: $e^{\epsilon} \approx 1 + \epsilon \approx 1$. Think $\epsilon \approx \frac{1}{100}$ or $\epsilon \approx \frac{1}{10}$ or $\epsilon \approx \frac{1}{\sqrt{n}}$ but not $\epsilon < \frac{1}{n}$
- Meaningful $\delta$ should be small ($\delta \ll 1/n$). Think $\delta \approx 2^{-60}$, or cryptographically small, or zero
Why Differential Privacy?

• DP: Strong, quantifiable, composable mathematical privacy guarantee

• A definition/standard/requirement, not a specific algorithm
  • Opens door to algorithmic development

• Provably resilient to known and unknown attack modes!

• Has a natural interpretation: I am protected (almost) to the extent I’m protected in my privacy-ideal scenario

• Theoretically, DP enables many computations with personal data while preserving personal privacy
  • Practicality in first stages of validation
Interpreting Differential Privacy

• A naïve hope: Your beliefs about me are the same after you see the output as they were before.

• Suppose I smoke in public
  • A public health study could teach that I am at risk for cancer.
  • But it didn’t matter whether or not my data was part of it.
Interpreting Differential Privacy

• Compare $x = (x_1, x_2, \ldots, x_i, \ldots, x_n)$ to $x_{-i} = (x_1, x_2, \ldots, \perp, \ldots, x_n)$

• $A$ is $\varepsilon$-differentially private if for all vectors $x$ and for all $i$: $A(x) \approx \varepsilon \ A(x_{-i})$.

• No matter what you know ahead of time, you learn (almost) the same things about me whether or not my data are used.
Interpreting Differential Privacy

• Compare \( x = (x_1, x_2, \ldots, x_i, \ldots, x_n) \) to \( x_{-i} = (x_1, x_2, \ldots, \perp, \ldots, x_n) \)

\[ \begin{align*}
\text{• } A \text{ is } \varepsilon\text{-differentially private if for all vectors } x \text{ and for all } i: A(x) & \approx \varepsilon A(x_{-i}).
\end{align*} \]

• Bayesian interpretation [DM’06, KS’08]:
  • \( A \) is \( \varepsilon\text{-differentially private if for all distributions } X, \text{ for all } i: \)
    \[ \begin{align*}
    X_i \mid A(X) = t & \approx \varepsilon X_i \mid A(X_{-i}) = t
    \end{align*} \]

Under any reasonable distance metric
Are you HIV positive?

- **Toss Coin**
  - If heads tell the truth
  - If tails, toss another coin:
    - If heads say yes
    - If tails say no
Randomized response [Warner 65]

- \( x \in \{0,1\} \)
  \[ \alpha = \frac{1}{4} \]

- \( RR_\alpha(x) = f(x) = \begin{cases} 
  x & \text{w.p. } \frac{1}{2} + \alpha \\
  -x & \text{w.p. } \frac{1}{2} - \alpha 
\end{cases} \)
  \[ \text{w.p. } \frac{3}{4} \]
  \[ \text{w.p. } \frac{1}{4} \]

- **Claim:** setting \( \alpha = \frac{1}{2} \frac{e^\epsilon - 1}{e^\epsilon + 1} \), \( RR_\alpha(x) \) is \( \epsilon \)—differentially private.

- **Proof:**
  - Neighboring databases: \( x = 0; x' = 1 \)
  - \( \Pr[RR(0)=0] = \frac{1}{2} \frac{1 + e^{\epsilon - 1}}{e^\epsilon + 1} \)
  - \( \Pr[RR(1)=0] = \frac{1}{2} \frac{1 - e^{\epsilon - 1}}{e^\epsilon + 1} \)
  - \( \epsilon = \ln 3 \)
Can we make use of randomized response?

- \((x_1, x_2, \ldots, x_n) \in \{0,1\}^n\); want to estimate \(\sum x_i\)
- \(Y_i = RR_\alpha(x_i)\)
- \(E[Y_i] = x_i \left(\frac{1}{2} + \alpha\right) + (1 - x_i) \left(\frac{1}{2} - \alpha\right) = \frac{1}{2} + \alpha(2x_i - 1)\)
- \(x_i = \frac{E[Y_i]-\frac{1}{2}+\alpha}{2\alpha}\) suggesting estimate \(\hat{x}_i = \frac{y_i\frac{1}{2}+\alpha}{2\alpha}\)
- \(E[\hat{x}_i] = x_i\) by construction but \(Var[\hat{x}_i] = \frac{\frac{1}{4}-\alpha^2}{4\alpha^2} \approx \frac{1}{\epsilon^2}\) high!
- \(E[\Sigma\hat{x}_i] = \Sigma x_i\) and \(Var[\Sigma\hat{x}_i] = n\frac{\frac{1}{4}-\alpha^2}{4\alpha^2} \approx \frac{n}{\epsilon^2}\); stdev \(\approx \frac{\sqrt{n}}{\epsilon}\)
- Useful when \(\frac{\sqrt{n}}{\epsilon} \ll n\) (i.e. \(n \gg \frac{1}{\epsilon^2}\))
- Lot of noise?
  - Compare with sampling noise \(\approx \sqrt{n}\)
Basic properties of differential privacy: post processing

- **Claim**: $M'$ is $(\epsilon, \delta)$-differentially private

- **Proof**:
  - Let $x, x'$ be neighboring databases and $S'$ a subset of $T'$
  - Let $S = \{z \in T: A(z) \in S'\}$ be the preimage of $S'$ under $A$

$$
\Pr[M'(x) \in S'] = \Pr[M(x) \in S] \\
\leq e^\epsilon \Pr[M(x') \in S] + \delta = e^\epsilon \Pr[M'(x') \in S'] + \delta
$$
Basic properties of differential privacy: Basic composition [DMNS06, DKMMN06, DL09]

• Claim: $M'$ is $(\epsilon_1 + \epsilon_2, \delta_1 + \delta_2)$-differentially private

• Proof (for the case $\delta_1 = \delta_2 = 0$)
  • Let $x, x'$ be neighboring databases and $S$ a subset of $(T_1 \times T_2)$

\[
\Pr[M'(x) \in S] = \sum_{(z_1, z_2) \in S} \Pr[M_1(x) = z_1 \land M_2(x) = z_2] \\
= \sum_{(z_1, z_2) \in S} \Pr[M_1(x) = z_1] \Pr[M_2(x) = z_2] \\
\leq \sum_{(z_1, z_2) \in S} e^{\epsilon_1} \Pr[M_1(x') = z_1] e^{\epsilon_2} \Pr[M_2(x') = z_2] = e^{\epsilon_1 + \epsilon_2} \Pr[M'(x') \in S]
\]
Basic properties of DP: Group privacy

• Let $M$ be $(\epsilon, \delta)$-differentially private:
  • For all datasets $x, x' \in X^n$ that differ on one entry, for all subsets $S$ of the outcome space $T$:
    $$\Pr[M(x) \in S] \leq e^\epsilon \Pr[M(x') \in S] + \delta.$$  

• **Claim:** for all databases $x, x' \in X^n$ that differ on $t$ entries, for all subsets $S$ of the outcome space $T$:
  $$\Pr[M(x) \in S] \leq e^{t\epsilon} \Pr[M(x') \in S] + t\delta e^{t\epsilon}.$$
Neighboring Inputs

[What Should Be Protected?]

• Inputs are neighboring if they differ on the data of a single individual
  • Record privacy: Databases $X$, $X'$ neighboring if differ on one record
Neighboring Inputs
[What Should Be Protected?]

• Inputs are neighboring if they differ on the data of a single individual
  • **Record privacy:** Databases $X$, $X'$ neighboring if differ on one record
  • **Edge privacy:** graphs $G$, $G'$ neighboring if differ on one edge

Image credit: www.perey.com
Neighboring Inputs

[What Should Be Protected?]

• Inputs are neighboring if they differ on the data of a single individual
  • **Record privacy:** Databases $X$, $X'$ neighboring if differ on one record
  • **Edge privacy:** graphs $G$, $G'$ neighboring if differ on one edge
  • **Node privacy:** graphs $G$, $G'$ neighboring if differ on one node and its adjacent edges

Image credit: www.perey.com
Laplacian: \( \text{Lap}(b) \)

- \( \text{Lap}(b) \) is a distribution over the reals with density function
  - \( h(x) = \frac{1}{2b} \exp(-\frac{|x|}{b}) \)
  - Expectation zero
  - Variance \( 2b^2 \)

\[ h(y) = \frac{\epsilon}{2} e^{-\epsilon|y|} \]
Laplacian: \( \text{Lap}(b) \)

- \( \text{Lap}(b) \) is a distribution over the reals with density function
  - \( h(x) = \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right) \)
  - Expectation zero
  - Variance \( 2b^2 \)
  - if \( Y \sim \text{Lap}(b) \) then \( \Pr(|Y| > t \cdot b) = \exp(-t) \)
Example: Counting Edges
[the basic technique]

• function $f(G) = \mathbb{E}_{ij}$ where $e_{ij} \in \{0,1\}$
Example: Counting Edges
[the basic technique]

• \( f(G) = \|e_{ij}\| \) where \( e_{ij} \in \{0,1\} \)

• Algorithm: On input \( G \) return \( f(G) + Y \), where \( Y \sim Lap\left( \frac{1}{\varepsilon} \right) \)

• \( E[Y] = 0; \sigma[Y] = \frac{\sqrt{2}}{\varepsilon} \)

\[ h(y) = \frac{\varepsilon}{2}e^{-\varepsilon|y|} \]
Example: Counting Edges
[the basic technique]

- \( f(G) = \mathbb{1}_{ij} \) where \( e_{ij} \in \{0,1\} \)
- **Algorithm:** On input \( G \) return \( f(G) + Y \), where \( Y \sim Lap(\frac{1}{\epsilon}) \)

- **Laplace Distribution:**
  - \( E[Y] = 0; \sigma[Y] = \sqrt{2}/\epsilon \)
  - Sliding property: \( \frac{h(y)}{h(y+1)} \leq e^{\epsilon} \)

\[ h(y) = \frac{\epsilon}{2} e^{-\frac{\epsilon}{2} |y|} \]
Example: Counting Edges
[the basic technique]

• $f(G) = \mathbb{I}_{ij}$ where $e_{ij} \in \{0,1\}$

• **Algorithm:** On input $G$ return $f(G) + Y$, where $Y \sim Lap\left(\frac{1}{\varepsilon}\right)$

• Laplace Distribution:
  • $E[Y] = 0; \sigma[Y] = \sqrt{2}/\varepsilon$
  • Sliding property: $\frac{h(y)}{h(y+1)} \leq e^\varepsilon$

• For $G, G'$ edge neighboring:
  $$|f(G) - f(G')| = \left| \sum_{ij} e_{ij} - \sum_{ij} e'_{ij} \right| \leq 1$$

Neighboring: $\ell_1$ distance at most one
differential privacy

\[
\Pr[M(D_1) \in C] \leq e^\varepsilon \Pr[M(D_2) \in C]
\]
Framework of Global Sensitivity

\[ GS_f = \max |f(G) - f(G')|_1 \] taken over neighboring \( G, G' \)

\[ A(G) = f(G) + \operatorname{Lap}^d \left( \frac{GS_f}{\epsilon} \right) \]

- Many natural functions have low global sensitivity
  - e.g., histogram, mean,
  - Histogram: split input into bins, count how many in bin
The Laplace Mechanism [DMNS06]

\[ GS_f = \max |f(G) - f(G')|_1 \] taken over neighboring \( G, G' \)

- Theorem [DMNS06]:
  - \( A(G) = f(G) + \text{Lap}^d \left( \frac{GS_f}{\epsilon} \right) \) is differentially private
The Laplace Mechanism: Proof

• Neighboring inputs X and Y
• \( f(X) \in R^d, f(Y) \in R^d \)
• \( f(X) + \text{Lap}^d \left( \frac{GS_f}{\varepsilon} \right) \) gives probability density function \( p_X \)
• \( f(Y) + \text{Lap}^d \left( \frac{GS_f}{\varepsilon} \right) \) gives probability density function \( p_Y \)
• Choose any point \( z \in R^d \)
• \( p_X(z)/p_Y(z) = \prod_{i=1}^{k} \left( \exp \left( -\frac{\varepsilon |f(X)_i - z_i|}{GS(f)} \right) / \exp \left( -\frac{\varepsilon |f(Y)_i - z_i|}{GS(f)} \right) \right) = \prod_{i=1}^{k} \left( \exp \left( \varepsilon \left( \frac{|f(Y)_i - z_i| - |f(X)_i - z_i|}{GS(f)} \right) \right) \right) \)

Density for Lap(b): \( h(x) = \frac{1}{2b} \exp(-\frac{|x|}{b}) \)
The Laplace Mechanism: Proof

• Choose any point \( z \in R^d \)

\[
p_X(z)/p_Y(z) = \prod_{i=1}^{k} \left( \exp\left(-\frac{\varepsilon|f(X)_i-z_i|}{GS(f)}\right) / \exp\left(-\frac{\varepsilon|f(Y)_i-z_i|}{GS(f)}\right) \right) =
\]
\[
\prod_{i=1}^{k} \left( \exp\left(\frac{\varepsilon(|f(Y)_i-z_i|-|f(X)_i-z_i|)}{GS(f)}\right) \right) \leq
\]
\[
\prod_{i=1}^{k} \left( \exp\left(\frac{\varepsilon|f(X)_i-f(Y)_i|}{GS(f)}\right) \right) =
\]
\[
\exp\left(\frac{\varepsilon|f(X)-f(Y)|_1}{GS(f)}\right) \leq \exp(\varepsilon)
\]

Density for Lap(b): \( h(x) = \frac{1}{2b} \exp(-\frac{|x|}{b}) \)
Better Composition: Answering all threshold queries

• Data domain: $D = \{1, ..., T\}$ (ordered domain with $T$ elements)
• Database: $X \in D^n$

Want (approx.) answers to all queries of the form: $q_t(X) = |\{i : 1 \leq x_i \leq t\}|$

$GS(q_t) = 1$ (changing a data point in $X$ can increase/decrease $q_t(X)$ by at most one)

• Idea: answer all $T$ queries by adding noise $\text{Lap}(\frac{1}{\epsilon'})$ where $\epsilon' = \frac{\epsilon}{T}$
  • Using (simple) composition, this provides $\epsilon$-differential privacy
  • Problem: noise magnitude linear in $T$; can we do better?

$$n = 15 \quad q_7(d) = \frac{7}{15}$$
Answering all threshold queries

idea: compute $\log T$ histograms
Answering all threshold queries

\[ n = 15 \]
\[ q_7(D) = \frac{7}{15} \]
Answering all threshold queries

• What we get (using basic composition):
  • Computing \( \log T \) histograms, each with \( \epsilon' = \frac{\epsilon}{\log T} \)
    • E.g., add noise \( \text{Lap}(2\epsilon/\log T) \) to each count
    • Noise variance \( \sim \left( \frac{\log T}{\epsilon} \right)^2 \)
  • Each answer to threshold query is sum of (at most) \( \log T \) noisy estimates
    • Overall noise variance \( \sim \log T \left( \frac{\log T}{\epsilon} \right)^2 \)
    • Whp noise magnitude = \( \frac{\text{polylog}(T)}{\epsilon} \)
Edge vs. Node Privacy - Counting Edges

\[ GS_f = \max |f(G) - f(G')|_1 \text{ taken over neighboring } G, G' \]

\[ A(G) = f(G) + \text{Lap}^d\left(\frac{GS_f}{\epsilon}\right) \]

- Counting edges: \( f(G) = \sum_{i,j} e_{ij} \) where \( e_{ij} \in \{0,1\} \)
- Edge privacy: \( GS_f = 1 \), noise \( \sim \frac{1}{\epsilon} \)
- Node privacy: \( GS_f = n \), noise \( \sim \frac{n}{\epsilon} \)
Exponential Sampling [MT07]

• $x_i = \{\text{books read by } i \text{ this year}\}$, $Y = \{\text{book names}\}$
• “Score” of $y \in Y$: $q(y, x) = \#\{i: y \in x_i\}$
• **Goal:** output book read by most

• **Mechanism:** given $x$, output book name $y$ with probability prop to
  • $\exp(\epsilon \cdot q(y, x))$

• **Claim:** Mechanism is $\epsilon$-differentially private

• **Claim:** If most popular website has score $T = \max_{y \in Y} q(y, x)$, then

$$E[q(y_0, x)] \geq T - O\left(\frac{\log|Y|}{\epsilon}\right)$$
Exponential Sampling

• Database X, function values in R, utility function from database and function value to the reals
• X=who reads what
• R=names of books
• u(X,r) – the closer to the real max the better
• \( \Delta u = \max_{r \in R} \max_{X,Y:|X-Y|_1 \leq 1} |u(X, r) - u(Y, r)| \)
• Output each possible \( r \in R \) with probability proportional to
  • \( \exp(\epsilon u(X, r) / \Delta u) \)
  • \( \exp\left(\frac{\epsilon u(X, r)}{\Delta u}\right) / \exp\left(\frac{\epsilon u(Y, r)}{\Delta u}\right) = \exp\left(\frac{\epsilon (u(X, r) - u(Y, r))}{\Delta u}\right) \)
The correct proof

\[
\frac{\Pr[M_E(x, u, R) = r]}{\Pr[M_E(y, u, R) = r]} = \frac{\left( \frac{\exp\left(\frac{\varepsilon u(x, r)}{2\Delta u} \right)}{\sum_{r' \in R} \exp\left(\frac{\varepsilon u(x, r')}{2\Delta u} \right)} \right)}{\left( \frac{\exp\left(\frac{\varepsilon u(y, r)}{2\Delta u} \right)}{\sum_{r' \in R} \exp\left(\frac{\varepsilon u(y, r')}{2\Delta u} \right)} \right)}
\]

\[
= \left( \frac{\exp\left(\frac{\varepsilon u(x, r)}{2\Delta u} \right)}{\exp\left(\frac{\varepsilon u(y, r)}{2\Delta u} \right)} \right) \cdot \left( \frac{\sum_{r' \in R} \exp\left(\frac{\varepsilon u(y, r')}{2\Delta u} \right)}{\sum_{r' \in R} \exp\left(\frac{\varepsilon u(x, r')}{2\Delta u} \right)} \right)
\]

\[
= \exp\left(\frac{\varepsilon (u(x, r) - u(y, r'))}{2\Delta u}\right)
\]

\[
\leq \exp\left(\frac{\varepsilon}{2}\right) \cdot \exp\left(\frac{\varepsilon}{2}\right) \cdot \left( \frac{\sum_{r' \in R} \exp\left(\frac{\varepsilon u(x, r')}{2\Delta u} \right)}{\sum_{r' \in R} \exp\left(\frac{\varepsilon u(x, r')}{2\Delta u} \right)} \right)
\]

\[
= \exp(\varepsilon).
\]
Exponential mechanism gives good utility

• $\text{Prob(utility of exp mechanism } \leq \text{ Opt utility } - \left(\frac{2\Delta u}{\epsilon}\right)(\ln(|R|) + t) \leq \exp(-t)$
Next week: Uri Stemmer

• Will talk about practical local differential privacy
• Local? “users retain their data and only send the server randomizations that are safe for publication.
• The random response could be viewed as local differential privacy
• Talk will be in the context of heavy hitters: items that appear many times.