Differential Privacy

Privacy & Fairness in Data Science
CompSci 590.01 Fall 2018
Outline

• Problem

• Differential Privacy

• Algorithms
**Statistical Databases**

Individuals with sensitive data:
- Person 1
- Person 2
- Person 3
- ... (Person N)

Data Collectors:
- Hospital
- Census
- Google

Data Analysts:
- Doctors
- Medical Researchers
- Economists
- Ranking Algorithms
- Machine Learning Researchers
Statistical Database Privacy

Function provided by the analyst

Output can disclose sensitive information about individuals

$\text{Person 1} \quad r_1$

$\text{Person 2} \quad r_2$

$\text{Person 3} \quad r_3$

$\text{Person N} \quad r_N$

Server

$DB$

$f(DB)$
Statistical Database Privacy

Privacy for individuals (controlled by a parameter $\varepsilon$)

$f_{private}(DB, \varepsilon)$

Server

Person 1
$r_1$

Person 2
$r_2$

Person 3
$r_3$

Person N
$r_N$

...
Statistical Database Privacy

Utility for analyst
\( f_{\text{private}}(DB) \approx f(DB) \)

\( f_{\text{private}}(DB, \varepsilon) \)

Server

DB

Person 1
\( r_1 \)

Person 2
\( r_2 \)

Person 3
\( r_3 \)

\( \cdots \)

Person N
\( r_N \)
Statistical Database Privacy (untrusted collector)

Server wants to compute $f$

Individuals do not want server to infer their records
Statistical Database Privacy (untrusted collector)

Perturb records to ensure privacy for individuals and Utility for server

\[ f(\text{DB}^*) \]
## Statistical Databases in real-world applications

<table>
<thead>
<tr>
<th>Application</th>
<th>Data Collector</th>
<th>Private Information</th>
<th>Analyst</th>
<th>Function (utility)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical</td>
<td>Hospital</td>
<td>Disease</td>
<td>Epidemiologist</td>
<td>Correlation between disease and geography</td>
</tr>
<tr>
<td>Genome analysis</td>
<td>Hospital</td>
<td>Genome</td>
<td>Statistician/Researcher</td>
<td>Correlation between genome and disease</td>
</tr>
<tr>
<td>Advertising</td>
<td>Google/FB</td>
<td>Clicks/Browsing</td>
<td>Advertiser</td>
<td>Number of clicks on an ad by age/region/gender ...</td>
</tr>
<tr>
<td>Social Recommendations</td>
<td>Facebook</td>
<td>Friend links/profile</td>
<td>Another user</td>
<td>Recommend other users or ads to users based on social network</td>
</tr>
</tbody>
</table>
Statistical Databases in real-world applications

- Settings where data collector may not be trusted (or may not want the liability ...)

<table>
<thead>
<tr>
<th>Application</th>
<th>Data Collector</th>
<th>Private Information</th>
<th>Function (utility)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location Services</td>
<td>Verizon/AT&amp;T</td>
<td>Location</td>
<td>Traffic prediction</td>
</tr>
<tr>
<td>Recommendations</td>
<td>Amazon/Google</td>
<td>Purchase history</td>
<td>Recommendation model</td>
</tr>
<tr>
<td>Traffic Shaping</td>
<td>Internet Service Provider</td>
<td>Browsing history</td>
<td>Traffic pattern of groups of users</td>
</tr>
</tbody>
</table>
Privacy is not ...
Statistical Database Privacy is not …

• Encryption:
Statistical Database Privacy is not …

• Encryption:
  Alice sends a message to Bob such that Trudy (attacker) does not learn the message. Bob should get the correct message …

• Statistical Database Privacy:
  Bob (attacker) can access a database
  - Bob must learn aggregate statistics, but
  - Bob must not learn new information about individuals in database.
Statistical Database Privacy is not …

• Computation on Encrypted Data:
Statistical Database Privacy is not …

• Computation on Encrypted Data:
  - Alice stores encrypted data on a server controlled by Bob (attacker).
  - Server returns correct query answers to Alice, without Bob learning *anything* about the data.

• Statistical Database Privacy:
  - Bob is allowed to learn aggregate properties of the database.
Statistical Database Privacy is not …

- The Millionaires Problem:
Statistical Database Privacy is not …

• Secure Multiparty Computation:
  - A set of agents each having a private input $x_i$ …
  - … Want to compute a function $f(x_1, x_2, \ldots, x_k)$
  - Each agent can learn the true answer, but must learn no other information than what can be inferred from their private input and the answer.

• Statistical Database Privacy:
  - Function output must not disclose individual inputs.
Statistical Database Privacy is not …

• Access Control:
Statistical Database Privacy is not …

• Access Control:
  - A set of agents want to access a set of resources (could be files or records in a database)
  - Access control rules specify who is allowed to access (or not access) certain resources.
  - ‘Not access’ usually means no information must be disclosed

• Statistical Database:
  - A single database and a single agent
  - Want to release aggregate statistics about a set of records without allowing access to individual records
Privacy Problems

• In today's systems a number of privacy problems arise:
  – Encryption when communicating data across a unsecure channel
  – Secure Multiparty Computation when different parties want to compute on a function on their private data without using a centralized third party
  – Computing on encrypted data when one wants to use an unsecure cloud for computation
  – Access control when different users own different parts of the data

• Statistical Database Privacy: Quantifying (and bounding) the amount of information disclosed about individual records by the output of a valid computation.
What is privacy?
Desiderata for a Privacy Definition

1. **Resilience to background knowledge**
   - A privacy mechanism must be able to protect individuals’ privacy from attackers who may possess background knowledge.

2. **Privacy without obscurity**
   - Attacker must be assumed to know the algorithm used as well as all parameters [MK15].

3. **Post-processing**
   - Post-processing the output of a privacy mechanism must not change the privacy guarantee [KL10, MK15].

4. **Composition over multiple releases**
   - Allow a graceful degradation of privacy with multiple invocations on the same data [DN03, GKS08].
Privacy Breach: Attempt 1

A privacy mechanism $M(D)$ that allows an unauthorized party to learn sensitive information about any individual in $D$, which could not have learnt without access to $M(D)$.
Alice + SMOKING & PASSIVE SMOKING CAUSES CANCER = Alice has Cancer

Is this a privacy breach? NO
Privacy Breach: Attempt 2

A privacy mechanism $M(D)$ that allows an unauthorized party to learn sensitive information about any individual Alice in $D$, which could not have been learnt even with access to $M(D)$ if Alice was not in the dataset.
Outline

• Problem

• Differential Privacy

• Algorithms
Differential Privacy

For every pair of inputs that differ in one row

$D_1$ $D_2$

Adversary should not be able to distinguish between any $D_1$ and $D_2$ based on any $O$

For every output ...

[ln $\left( \frac{\Pr[A(D_1) = o]}{\Pr[A(D_2) = o]} \right) \leq \varepsilon, \quad \varepsilon > 0$]
Why pairs of datasets *that differ in one row*?

For every pair of inputs that differ in one row

```
D1

D2
```

For every output ...

```
O
```

Simulate the presence or absence of a single record
Why *all* pairs of datasets …?

For every pair of inputs that differ in one row

For every output …

Guarantee holds no matter what the other records are.
Why *all* outputs?

\[ \text{Set of all outputs} \]

\[ A(D_1) = O_1 \]

\[ P [ A(D_1) = O_1 ] \]

\[ A(D_2) = O_k \]

\[ P [ A(D_2) = O_k ] \]
Should not be able to distinguish whether input was $D_1$ or $D_2$ no matter what the output.
Privacy Parameter $\varepsilon$

For every pair of inputs that differ in one row

For every output ...

$$\Pr[A(D_1) = o] \leq e^{\varepsilon} \Pr[A(D_2) = o]$$

Controls the degree to which $D_1$ and $D_2$ can be distinguished. Smaller the $\varepsilon$ more the privacy (and better the utility)
Desiderata for a Privacy Definition

1. Resilience to background knowledge
   – A privacy mechanism must be able to protect individuals’ privacy from attackers who may possess background knowledge

2. Privacy without obscurity
   – Attacker must be assumed to know the algorithm used as well as all parameters [MK15]

3. Post-processing
   – Post-processing the output of a privacy mechanism must not change the privacy guarantee [KL10, MK15]

4. Composition over multiple releases
   – Allow a graceful degradation of privacy with multiple invocations on the same data [DN03, GKS08]
Differential Privacy

• Two equivalent definitions:

\[
\Pr[A(D_1) \in \Omega] \leq e^\varepsilon \Pr[A(D_2) \in \Omega]
\]

Every subset of outputs

\[
\Pr[A(X) \in \Omega] \leq e^{\varepsilon \cdot d(X,Y)} \Pr[A(Y) \in \Omega]
\]

Number of row additions and deletions to change X to Y
Outline

• Problem

• Differential Privacy

• Algorithms
Non-trivial deterministic Algorithms do not satisfy differential privacy

Space of all inputs

Space of all outputs (at least 2 distinct outputs)
Non-trivial deterministic Algorithms do not satisfy differential privacy

Each input mapped to a distinct output.
There exist two inputs that differ in one entry mapped to different outputs.
Random Sampling …

… also does not satisfy differential privacy

\[ \Pr[D_2 \rightarrow O] = 0 \text{ implies } \log \left( \frac{\Pr[D_1 \rightarrow O]}{\Pr[D_2 \rightarrow O]} \right) = \infty \]
**Randomized Response** (a.k.a. local randomization)

<table>
<thead>
<tr>
<th>D</th>
<th>Disease (Y/N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
</tr>
</tbody>
</table>

**O**

<table>
<thead>
<tr>
<th>Disease (Y/N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>N</td>
</tr>
</tbody>
</table>

With probability $p$, Report true value

With probability $1-p$, Report flipped value
Differential Privacy Analysis

• Consider 2 databases $D, D'$ (of size $M$) that differ in the $j^{th}$ value
  – $D[j] \neq D'[j]$. But, $D[i] = D'[i]$, for all $i \neq j$

• Consider some output $O$

\[
\frac{P(D \rightarrow O)}{P(D' \rightarrow O)} \leq e^\varepsilon \iff \frac{1}{1 + e^\varepsilon} < p < \frac{e^\varepsilon}{1 + e^\varepsilon}
\]
Next class

• Basic Algorithmic Primitives

• Composition