CS639: Data Management for Data Science

Lecture 26: Privacy
[slides from Vitaly Shmatikov]

Theodoros Rekatsinas
Reading

• Dwork. “Differential Privacy” (invited talk at ICALP 2006).
Basic Setting

Users (government, researchers, marketers, ...)

DB = \(x_1, x_2, x_3, \ldots, x_{n-1}, x_n\)

San

Random coins

Queries and answers: query 1, answer 1, \ldots, query T, answer T
Examples of Sanitization Methods

• Input perturbation
  • Add random noise to database, release

• Summary statistics
  • Means, variances
  • Marginal totals
  • Regression coefficients

• Output perturbation
  • Summary statistics with noise

• Interactive versions of the above methods
  • Auditor decides which queries are OK, type of noise
Strawman Definition

- Assume $x_1,\ldots,x_n$ are drawn i.i.d. from unknown distribution
- Candidate definition: sanitization is safe if it only reveals the distribution
- Implied approach:
  - Learn the distribution
  - Release description of distribution or re-sample points
- This definition is tautological!
  - Estimate of distribution depends on data... why is it safe?
Blending into a Crowd

• Intuition: “I am safe in a group of \( k \) or more”
  • \( k \) varies (3… 6… 100… 10,000?)

• Many variations on theme
  • Adversary wants predicate \( g \) such that \( 0 < \#\{i \mid g(x_i) = \text{true}\} < k \)

• Why?
  • Privacy is “protection from being brought to the attention of others” [Gavison]
  • Rare property helps re-identify someone
  • Implicit: information about a large group is public
    • E.g., liver problems more prevalent among diabetics
Clustering-Based Definitions

- Given sanitization $S$, look at all databases consistent with $S$
- Safe if no predicate is true for all consistent databases
- $k$-anonymity
  - Partition $D$ into bins
  - Safe if each bin is either empty, or contains at least $k$ elements
- Cell bound methods
  - Release marginal sums
Issues with Clustering

• Purely syntactic definition of privacy
• What adversary does this apply to?
  • Does not consider adversaries with side information
  • Does not consider probability
  • Does not consider adversarial algorithm for making decisions (inference)
“Bayesian” Adversaries

• Adversary outputs point $z \in D$
• Score = $1/f_z$ if $f_z > 0$, 0 otherwise
  • $f_z$ is the number of matching points in $D$
• Sanitization is safe if $E(\text{score}) \leq \varepsilon$
• Procedure:
  • Assume you know adversary’s prior distribution over databases
  • Given a candidate output, update prior conditioned on output (via Bayes’ rule)
  • If $\max_z E(\text{score} \mid \text{output}) < \varepsilon$, then safe to release
Issues with “Bayesian” Privacy

• Restricts the type of predicates adversary can choose
• Must know prior distribution
  • Can one scheme work for many distributions?
  • Sanitizer works harder than adversary
• Conditional probabilities don’t consider previous iterations
  • Remember simulatable auditing?
Classical Intuition for Privacy

• “If the release of statistics $S$ makes it possible to determine the value [of private information] more accurately than is possible without access to $S$, a disclosure has taken place.”  
  [Dalenius 1977]
  
  • Privacy means that anything that can be learned about a respondent from the statistical database can be learned without access to the database

• Similar to semantic security of encryption
  
  • Anything about the plaintext that can be learned from a ciphertext can be learned without the ciphertext
Problems with Classic Intuition

• Popular interpretation: prior and posterior views about an individual shouldn’t change “too much”
  • What if my (incorrect) prior is that every UTCS graduate student has three arms?

• How much is “too much?”
  • Can’t achieve cryptographically small levels of disclosure and keep the data useful
  • Adversarial user is supposed to learn unpredictable things about the database
Impossibility Result

• **Privacy:** for some definition of “privacy breach,”
  ∀ distribution on databases, ∀ adversaries A, ∃ A’
  such that \( \Pr(A(\text{San})=\text{breach}) - \Pr(A’()=\text{breach}) \leq \varepsilon \)
  • For reasonable “breach”, if San(DB) contains information about DB, then some adversary breaks this definition

• Example
  • Paris knows that Theo is 2 inches taller than the average Greek
  • DB allows computing average height of a Greek
  • This DB breaks Theos’s privacy according to this definition... even if his record is not in the database!

[Dwork]
(Very Informal) Proof Sketch

• Suppose DB is uniformly random
  • Entropy $I(\ DB \ ; \ San(DB) ) > 0$

• “Breach” is predicting a predicate $g(DB)$

• Adversary knows $r, H(r \ ; \ San(DB)) \oplus g(DB)$
  • $H$ is a suitable hash function, $r=H(DB)$

• By itself, does not leak anything about DB (why?)

• Together with $San(DB)$, reveals $g(DB)$ (why?)
Differential Privacy (1)

- Example with Greeks and Theo
  Adversary learns Theo’s height even if he is not in the database
- Intuition: “Whatever is learned would be learned regardless of whether or not Theo participates”
  Dual: Whatever is already known, situation won’t get worse
Define n+1 games

Game 0: Adv. interacts with San(DB)

Game i: Adv. interacts with San(DB_{-i}); DB_{-i} = (x_1, ..., x_{i-1}, 0, x_{i+1}, ..., x_n)

Given S and prior p() on DB, define n+1 posterior distrib's

\[ p_i(DB|S) = p(DB|S \text{ in Game } i) = \frac{p(San(DB_{-i}) = S) \times p(DB)}{p(S \text{ in Game } i)} \]
Differential Privacy (3)

Definition: San is safe if

\[ \forall \text{ prior distributions } p(\xi) \text{ on } DB, \]
\[ \forall \text{ transcripts } S, \forall \ i = 1, \ldots, n \]
\[ \text{StatDiff}( p_0(\xi | S) , p_i(\xi | S) ) \leq \epsilon \]
Indistinguishability

\[ x_1 \times x_2 \times x_{n-1} \times x_n \]

\[ y_3 \times x_2 \times x_{n-1} \times x_n \]

\[ \text{Distance between distributions is at most } \varepsilon \]
Which Distance to Use?

• Problem: $\varepsilon$ must be large
  • Any two databases induce transcripts at distance $\leq n\varepsilon$
  • To get utility, need $\varepsilon > 1/n$
• Statistical difference $1/n$ is not meaningful!
• Example: release random point in database
  • $\text{San}(x_1,\ldots,x_n) = (j, x_j)$ for random $j$
• For every $i$, changing $x_i$ induces statistical difference $1/n$
• But some $x_i$ is revealed with probability 1
Definition: San is \( \varepsilon \)-indistinguishable if

\[ \forall A, \forall DB, DB' \text{ which differ in 1 row}, \forall \text{ sets of transcripts } S \]

\[ p(\text{San}(DB) \in S) \in (1 \pm \varepsilon) p(\text{San}(DB') \in S) \]

Equivalently, \( \forall S: \)

\[ \frac{p(\text{San}(DB) = S)}{p(\text{San}(DB') = S)} \in 1 \pm \varepsilon \]
Indistinguishability ⇒ Diff. Privacy

Definition: San is safe if

∀ prior distributions \( p(\xi) \) on DB,
∀ transcripts \( S \), ∀ \( i = 1, \ldots, n \)

\[
\text{StatDiff}( p_0(\xi | S) , p_i(\xi | S) ) \leq \varepsilon
\]

\[
p_i(DB | S) = p(DB | S \text{ in Game } i) = \frac{p(San(DB_{-i}) = S) \times p(DB)}{p(S \text{ in Game } i)}
\]

For every \( S \) and DB, indistinguishability implies

\[
\frac{p_i(DB | S)}{p_0(DB | S)} = \frac{p(San(DB_{-i}) = S)}{p(San(DB) = S)} \times \frac{p(S \text{ in Game 0})}{p(S \text{ in Game } i)} \approx 1 \pm 2\varepsilon
\]

This implies StatDiff( \( p_0(\xi | S) , p_i(\xi | S) \) ) ≤ \( \varepsilon \)
Sensitivity with Laplace Noise

**Theorem**

If \( A(x) = f(x) + \text{Lap}\left(\frac{\text{GS}_f}{\varepsilon}\right) \) then \( A \) is \( \varepsilon \)-indistinguishable.

Laplace distribution \( \text{Lap}(\lambda) \) has density \( h(y) \propto e^{-\frac{\|y\|_1}{\lambda}} \)

Sliding property of \( \text{Lap}\left(\frac{\text{GS}_f}{\varepsilon}\right) \):

\[
\frac{h(y)}{h(y+\delta)} \leq e^{\varepsilon \cdot \frac{\|\delta\|}{\text{GS}_f}} \text{ for all } y, \delta
\]

**Proof idea:**

- \( A(x) \): blue curve
- \( A(x') \): red curve

\( \delta = f(x) - f(x') \leq \text{GS}_f \)
Differential Privacy: Summary

• San gives $\varepsilon$-differential privacy if for all values of DB and Me and all transcripts $t$:

\[
\frac{\Pr[ San (DB - Me) = t]}{\Pr[ San (DB + Me) = t]} \leq e^\varepsilon \approx 1 \pm \varepsilon
\]