Introduction to Differential Privacy

J.T. Cho

CIS700-003 — University of Pennsylvania

March 2017

TODO: Introduction and Motivation

Intuitively Formalizing Privacy

Desiderata. An individual's risk is not increased significantly by opting into a study.

In other words, individuals should have *plausible deniability*.

A Game of Plausible Deniability

Suppose we want to test for the percentage of smokers in a population of people.

Goal. Design a protocol for surveying people so they may claim plausible deniability of being a smoker.

Randomized Response

Protocol.

1. Flip fair coin.
2. If tails, respond truthfully.
3. If heads, flip second coin, respond `Yes` if heads, `No` if tails.

How does this give us plausible deniability?

Randomized Response, cont.

Plausibility deniability of any outcome gives us privacy - can't single out an individual.

Adding uncertainty to query output in the form of randomness/noise allows us to achieve this.

The issue is then to analyze the noisy data to derive an accurate result!

Intuitively Defining Differential Privacy

We are given a database, an individual, and a mechanism which processes queries.

This mechanism should with high probability output the same result whether or not the individual's information is in the database!

Model of Computation

Definition. (Probability Simplex) Given a discrete set B, the probability simplex over B denoted $\Delta(B)$ is

$$\Delta(B) = \left\{ x \in \mathbb{R}^{|B|} \ : \ x_i \geq 0 \ \forall i, \sum_{i=1}^{|B|} x_i = 1 \right\}$$

In plain English, $\Delta(B)$ is the set of all probability vectors of length $|B|$ that sum to 1.

Model of Computation, cont.

Definition. (Mechanism) A mechanism $\mathcal{M}$ with domain A and range B is associated with the mapping $\mathcal{M} : A \to \Delta(B)$. On input $a \in A$, the mechanism $\mathcal{M}$ outputs $\mathcal{M}(a) = b$ with probability $(\mathcal{M}(a))_b$ for each $b \in B$.

The probability is taken over the randomness of the mechanism (coin flips).

Our goal is to find a differentially private mechanism!

Model of Computation, cont.

A database will be represented as a 'histogram' vector $x \in \mathbb{N}^{|\chi|}$, counting the frequency of each element from the universe χ.

Definition. (Distance Between Databases) The ℓ_1 norm of a database x is denoted $||x||_1$, defined as

$$||x||_1 = \sum_{i=1}^{|\chi|} |\chi_i|$$

The ℓ_1 distance between 2 databases x, y is $||x - y||_1$, the number of records differing between x and y.

Differential Privacy

Definition. A mechanism $\mathcal{M}$ on a database with domain $\mathbb{N}^{|\mathcal{X}|}$ is (ϵ, δ)-differentially private if $\forall S \subseteq \text{Range}(\mathcal{M})$ and $\forall x, y \in \mathbb{N}^{|\mathcal{X}|}$ such that $||x - y||_1 \leq 1$,

$$\Pr(\mathcal{M}(x) \in S) \leq \exp(\epsilon) \Pr(\mathcal{M}(y) \in S) + \delta$$

with the probability space over the coin flips in the mechanism $\mathcal{M}$.

If $\delta = 0$, $\mathcal{M}$ is ϵ-differentially private.

Understanding the Definition

Consider the singleton set $\{s\} \subseteq \text{Range}(\mathcal{M})$ - s is an example output of $\mathcal{M}$.

If $\mathcal{M}$ is ϵ-diff. private, the probability of outputting s on x is at most e^{ϵ} times the probability of outputting s on any neighboring database y.

Understanding the Definition, cont.

In other words, the definition states that the probability of any output of $\mathcal{M}$ is within an e^{ϵ} factor of whether or not an individual is included in the database.

The smaller ϵ is, the stronger the 'privacy' guarantee!

Randomized Response, Revisited

Claim. Randomized response is $(\ln 3, 0)-$differentially private.

Proof. Let the databases be drawn from universe $\{0, 1\}$ and the mechanism range $\text{Range}(\mathcal{M}) = \{0, 1\}$.

$$\Pr(\text{Response} = \text{No} \mid \text{Truth} = \text{No}) = \Pr(M(0) \in \{0\}) = 3/4$$

$$\Pr(\text{Response} = \text{No} \mid \text{Truth} = \text{Yes}) = \Pr(M(1) \in \{0\}) = 1/4$$

If $\epsilon = \ln 3$,

$$\Pr(M(0) \in \{0\}) = 3/4 \leq \exp(\epsilon) \Pr(M(1) \in \{0\}) = 3/4$$
$$\Pr(M(1) \in \{0\}) = 1/4 \leq \exp(\epsilon) \Pr(M(0) \in \{0\}) = 9/4$$

Finding an ϵ-private Mechanism

Our intuition from before is that adding noise to original data gives 'privacy'.

Instead of coin flips, what if we chose a different probability distribution and added a dependence on ϵ?

We also want to be able to control how sensitive the mechanism is to changes in the database (i.e. should including a single individual result in a big change in the output?)

Laplace Distribution

Definition. (Laplace Distribution) The Laplace distribution centered at 0 with scale b has the pdf,

$$\mathrm{Lap}(x \mid b) = \frac{1}{2b}\exp\left(-\frac{|x|}{b}\right)$$

and variance,

$$\sigma^2 = 2b^2$$

Often written as $\mathrm{Lap}(b)$ for short.

ℓ_1 sensitivity

We define **numeric queries** to be functions $f : \mathbb{N}^{|\mathcal{X}|} \rightarrow \mathbb{R}^k$ (i.e. taking in a database and outputting a k-long real-valued vector).

Definition. (ℓ_1-sensitivity) The ℓ_1-sensitivity of a numeric query f is:

$$\Delta f : \max_{\substack{x,y \in \mathbb{N}^{|\mathcal{X}|} \\ ||x-y||_1 = 1}} ||f(x) - f(y)||_1$$

The ℓ_1-sensitivity captures the magnitude by which an individual's data can change the function f in the worst case.

Laplace Mechanism

Definition. (Laplace Mechanism) Given any function $f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^k$, the Laplace mechanism is defined,

$$\mathcal{M}_L(x, f(\cdot), \epsilon) = f(x) + (Y_1, Y_2, \ldots, Y_k)$$

where the Y_i are i.i.d. drawn from $\text{Lap}(\Delta f/\epsilon)$.

Laplace Mechanism, cont.

Theorem. The Laplace mechanism preserves $(\epsilon, 0)$-differential privacy.

Sketch of Proof. Consider any two databases x and y that differ in at most 1 record and a database function f.

Consider the probabilities of getting some arbitrary value z from evaluating the mechanism $\mathcal{M}_L(x, f, \epsilon)$ and $\mathcal{M}_L(y, f, \epsilon)$.

Taking the ratio and using the Laplace distribution pdf, use a series of inequality bounds to demonstrate that the ratio is bounded by $\exp(\epsilon)$.

Example

Input. Database x of medical information of N records.

Goal. Compute proportion of smokers in a differentially private way.

$g(x) = [\#\text{ of smokers in } x]/N$.

For any two databases differing in a single element, what is the largest amount that the proportion can change by?

Exponential Mechanism

Designed for non-numerical queries and cases where adding noise directly to the output is undesirable.

Utility function $u : \mathbb{N}^{|\mathcal{X}|} \times \mathcal{R} \to \mathbb{R}$, maps database/output pairs to utility scores.

Sensitivity of u:

$$\Delta u = \max_{r \in \mathcal{R}} \max_{x,y : ||x-y||_1 \leq 1} |u(x,r) - u(y,r)|$$

Intuition. Output element of $\mathcal{R}$ with maximum possible utility.

Exponential Mechanism, cont.

Definition. (Exponential Mechanism) The exponential mechanism $\mathcal{M}_E(x, u, \mathcal{R})$ selects and outputs an element $r \in \mathcal{R}$ with probability proportional to $\exp(\frac{\epsilon u(x,y)}{2\Delta u})$.

Exponential Mechanism, cont.

Theorem. The exponential mechanism preserves $(\epsilon, 0)$-differential privacy.

Differentially Private Online Learning

Context. You want to invest in the stock market and have assembled a panel of experts. Each day, you can pick one expert's choice of stock to invest in.

Goal. Each day, pick experts such that after a period of time you do almost as well as the best expert!

Differentially Private Online Learning, cont.

Scenario. Each day $t = 1, \ldots, T$.

(a) Choose expert $a_t \in \{1, \ldots, k\}$.

(b) Observe loss $\ell_i^t \in [0, 1]$ for each expert $i \in \{1, \ldots, k\}$ and experiences loss ℓ_a^t.

For sequence of losses $\ell^{\leq T} = \{\ell^k\}_{t=1}^T$,

$$L_i(\ell^{\leq T}) = \frac{1}{T} \sum_{t=1}^{T} \ell_i^t \text{ (total avg. loss of expert } i\text{)}$$

$$L_A(\ell^{\leq T}) = \frac{1}{T} \sum_{t=1}^{T} \ell_{a_t}^t \text{ (total avg. loss of algorithm)}$$

No Regret Learning

$$\text{Regret}(A, \ell^{\leq T}) = L_A(\ell^{\leq T}) - \min_i L_i(\ell^{\leq T})$$

Regret is the difference between the loss incurred by the algorithm and the loss of the best expert.

Goal. Design algorithms guaranteeing that *for all* possible loss sequences $\ell^{\leq T}$, even adversarilly chosen,

$$\text{Regret} \to 0 \text{ as } T \to \infty$$

Random Weighted Majority Algorithm

Input. Stream σ_ℓ of losses $\ell^1, \ell^2, \ldots$

Output. Stream of actions $a_1, a_2, \ldots$

---

procedure $\mathrm{RWM}(\eta)$
  for $i \in \{1, \ldots, k\}$, let $w_i \leftarrow 1$ **do**
  for $t = 1, \ldots,$ **do**
    Choose action $a_t = i$ with probability proportional to w_i.
    Observe ℓ^t and set $w_i \leftarrow w_i \cdot \exp(-\eta \ell_i^t), \forall i \in [k]$

Random Weighted Majority Algorithm, cont.

Theorem. for any adversarially chosen sequence of losses of length T, $\ell^{\leq T} = (\ell^1, \ldots, \ell^T)$, the R.W.M. algorithm with update parameter η has guarantee:

$$E[\text{Regret}(\text{RWM}(\eta), \ell^{\leq T})] \leq \eta + \frac{\ln(k)}{nT}$$

Choosing $\eta = \sqrt{\ln k / T}$ yields

$$E[\text{Regret}(\text{RWM}(\eta), \ell^{\leq T})] \leq 2\sqrt{\frac{\ln k}{T}}$$

which tends to 0 as T goes to ∞.

Differentially Private Online Learning, cont.

Can we do the same process but in a differentially private way?

What should our "input database" be? Our output?

Differentially Private Online Learning, cont.

Input Database. Collection of loss vectors $\ell^{\leq T} = (\ell^1, \ldots, \ell^T)$. Neighboring databases $\ell^{\hat{\leq} T}$ differs in entire vector in 1 timestep.

Output. Sequence of actions chosen by the algorithm, $a_1, \ldots, a_T$.

Random Weighted Majority Algorithm, cont.

We present the same algorithm from before, presented in a slightly different way.

---

procedure $\mathrm{RWM}(\eta)$

    for $t = 1, \ldots,$ **do**

        Choose action $a_t = i$ with probability proportional to $\exp(-\eta \sum_{j=1}^{t-1} \ell_i^j)$.

        Observe ℓ^t.

This is the exponential mechanism with quality score $q(i, \ell^{<T}) \sum_{j=1}^{t-1} \ell_i^j$.

Differential Privacy and RWM

Theorem. For a sequence of losses of length T, the algorithm RWM(η) with $\eta = \frac{\epsilon}{\sqrt{32T\ln(1/\delta)}}$ is (ϵ, δ)