Calibrating Noise to Sensitivity in Private Data Analysis

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith

Mamadou H. Diallo

Overview

- Issues: preserving privacy in statistical databases
- Assumption: trusted server
- General Approach:
  - Perturb the true answer to a query by adding random noise generated by a distribution
  - The query function f maps databases to reals
- Previous Work
  - 1. Revealing information while preserving privacy: – (minimum perturbation level, $\varepsilon=\sqrt{n}$, consider one column database containing a single binary attribute)
  - 2. Privacy-preserving datamining on vertically partitioned databases: – (generalize 1 to multi-attributes and vertically partitioned DB)

- In this paper
  - General functions f – map databases to ***vectors*** of reals
  - Privacy definition: ε-indistinguishability
  - Privacy preservation: calibration of the standard deviation of the noise according to the sensitivity of the function f
  - A simple method for adding noise based on Laplace distribution

Concepts

- **Statistical databases**: a database used for statistical analysis purposes, a query-response mechanism
- **Interactive scheme**: only the result of a query is perturbed
- **Non-interactive scheme**: the whole database in perturbed before any query
- **Adaptive interactive scheme**: queries based on answers given thus far
- **Probabilistic Turing machine**: randomly chooses between the available transitions at each point according to some probability distribution.
- **Laplace distribution**: It is the distribution of differences between two independent variates with identical exponential distributions
- **Probability density function**: a function that describes the relative likelihood for this random variable to occur at a given point in the observation space.
- **Hamming distance (2 strings)**: the number of positions at which the corresponding symbols are different
- **Transcript**: results of an interaction between a user and a privacy mechanism (query + response)

Definitions

- Adversary
  - Modeled as probabilistic interactive Turing machine with an advice tape
  - Probabilistic Turing machine: add randomness in transition choices
- Notations

  San: a database access protocol
  A: adversary
  x: a particular database
  $\mathbf{T_{San,A}(x)}$: a transcript, a random variable
  $\mathbf{x} = \langle x_1, x_2, \ldots, x_n \rangle$, where $x_i \in D = \{0,1\}^d$ or R^d, n = number entries
  $\mathbf{d_H(.,.)}$ over D_n: Hamming distance
  Pr[A=a]: probability density for continuous and discrete random variables

- **Definition**

  A mechanism is ε-indistinguishable if for all pairs x, x' in D^n, which differ in only one entry, for all adversaries A, and for all transcripts t:
  $| \ln(Pr[T_A(x)=t] / Pr[T_A(x')=t]) | \leq \varepsilon$, ε = leakage

  ε small ---> $\ln(1+\varepsilon) \approx \varepsilon$ and
  $(1-\varepsilon) \leq Pr[T_A(x)=t] / Pr[T_A(x')=t] \leq (1+\varepsilon)$, for all transcripts t

Definitions

- Laplace distribution: Lap(λ)
  - Density function: $h(y) = e^{(-|y|/\lambda)}$
  - Mean = 0
  - Standard deviation = λ
- Example: Noisy Sum
  - x in $\{0,1\}^n$
  - User wants to learn f(x), where $f(x) = \Sigma_i x_i$ (total number of 1's in DB)
  - Noise: Y~ Lap(1/ε) (Laplace distribution)
  - $T(x1, \ldots, xn) = \Sigma_i x_i + Y$
  - ε-indistinguishable mechanism

    $h(y)/h(y') \leq e^{\varepsilon|y-y'|}$ for any real numbers y, y'

    difference (x, x') = 1 ===> difference (f(x), f(x')) = 1

    $Pr(T(x) = t) / Pr(T(x') = t) = h(t-f(x)) / h(t-f(x')) \leq e^{\varepsilon|f(x)-f(x')|} \leq e^{\varepsilon}$

Definitions

- Non-negligible leakage and the choice of distance measure
  - Non-negligible leakage
    - Noisy Sum: $\varepsilon = \Omega(1/n)$ for a constant factor approximation to f(x)
    - Leakage is inherent for statistical utility
    - If $\varepsilon = o(1/n)$ for close databases, then $\varepsilon = o(1)$ for two databases ===> no utility
  - Average-case distance measures
    - Example: statistical distance
    - No meaningful privacy guarantee when $\varepsilon = \Omega(1/n)$
  - Example
    - $T(x_1, \ldots, x_n) = (i, x_i)$, where $i \in \{1, \ldots, n\}$
      If difference (x, x') =1, then difference (T(x), T(x')) = 1/n
      But, all transcripts reveal private information about some individual
    - Definition not satisfied:
      If x, x' differ in the (i, xi), Pr = 0 for x'

Sensitivity and Privacy

- Sensitivity
  - Sensitivity of query function f: S(f)
  - Noise: Lap(S(f)/ε)
  - Definition (L1 Sensitivity).
    The L1 sensitivity of a function f : D^n ---> R^d is the smallest number S(f) such that for all x, x' ∈ D^n which differ in a single entry,

    $$||f(x) - f(x')||_1 \leq S(f)$$

    Lipschitz condition on f: if $d_H(.,.)$ is Hamming metric on D^n, then
    $||f(x) - f(x')||_1 / d_H(x, x') \leq S(f)$

- Example: Sum and Histograms
  - (1)- If D = {0, 1} and $f(x) = \Sigma x_i$, i ∈ {1, …, n},
    then $S_{L1}(f) = 1$
  - (2)- If D = {B_1, …, B_d}, then f: D^n ---> Z^d is a histogram
    $S_{L1}(f) = 2$ - independent of d

Sensitivity and Privacy

- Calibrating noise according to S(f)
  - If Y drawn from Laplace distribution, then $h(y)/h(y') \leq \exp(|y-y'|/\lambda)$
  - If $Y = \langle L1, \ldots, Ld \rangle$, then $d(y) \sim \exp(-\|y\|1/\lambda)$
  - Consequence
    $Pr(z+Y=t) / Pr(z'+Y=t) \in exp(\pm\|z-z'\|1/\lambda)$

    To release f(x) with privacy, add Laplace noise with sd=S(f)/ε in all coordinates

  - Proposition
    - (Non-interactive Output Perturbation)

      For all $f : D^n \dashrightarrow R^d$, the following mechanism is ε–indistinguishable:

      $San_f(x) = f(x) + (Y_1, \ldots, Y_d)$,

      where the Yi are drawn from Lap(S(f)/λ)

Sensitivity and Privacy

- Adaptive interactive scheme
  - t = [Q1, a1,Q2, a2 . . . ,Qd, ad]
    Q = f1, f2, …, fd, where fi: Dn ---> R (adaptive sequence of queries)
    San = None if S(ft) > treshold , otherwise
    San = fi(x) + Lap(λ)

- Theorem
  - *For an arbitrary adversary A, let $f_t(x) : D^n$ ---> R^d be its query function as parameterized by a transcript t. If $\lambda = max_t S(f_t)/\varepsilon$, the mechanism above is ε-indistinguishable.*

Sensitivity and Privacy

- Sensitivity in general metric spaces
  - Intuition: insensitive functions of a database can be released privately is not specific to L1
  - **Definition**:
    Let M be a metric space with a distance function $d_M(\cdot, \cdot)$. The sensitivity $S_M(f)$ of a function f: D^n ---> M is the amount that the function value varies when a single entry of the input is changed.
    $S_M(f)$ (def) = sup $d_M(f(x), f(x'))$, x,x': $d_H(x, x')=1$
  - Probability density function: z ∈ M

$$h_{z,\epsilon}(y) \propto \exp\left(\frac{\epsilon \cdot \mathrm{d}_{\mathcal{M}}(y,z)}{2 \cdot S_{\mathcal{M}}(f)}\right)$$

  - Sampling function: (approximating)

$$\Pr[\mathcal{T}(\mathbf{x}) = y] = \frac{\exp\left(\frac{\epsilon}{2S_{\mathcal{M}}(f)} \cdot \mathrm{d}_{\mathcal{M}}(y, f(\mathbf{x}))\right)}{\int_{y \in \mathcal{M}} \exp\left(\frac{\epsilon}{2S_{\mathcal{M}}(f)} \cdot \mathrm{d}_{\mathcal{M}}(y, f(\mathbf{x}))\right) dy}$$

  - **Theorem**. In a metric space where $h_{f(x),\varepsilon}()$ is well-defined, adding noise to f(x) as above yields an ε-indistinguishable scheme.

Interactive vs. Non-Interactive Mechanisms

- Notations
  - $D = \{0, 1\}^d$
    g: boolean function
    r = (r1, r2, ..., rn), ri in $\{0, 1\}^d$
    $g_r(i, x) = ri \Theta x$ (inner product, modulo 2, of ri and x)
- Theorem 3 (Non-interactive Schemes Require Large Databases)

  Suppose that San is an ε-indistinguishable non-interactive mechanism with domain $D= \{0, 1\}^d$. For at least 2/3 of the functions of the form $f_g(x) = \Sigma_i g(i, x_i)$, the following two distributions have statistical difference $O(n^{4/3} \varepsilon^{2/3} 2^{-d/3})$:

  Distribution 0: $T_{San}(x)$ where $x \in R \{x \in D^n : f_g(x) = 0\}$

  Distribution 1: $T_{San}(x)$ where $x \in R \{x \in D^n : f_g(x) = n\}$

  If $n = o(2^{d/4}/\sqrt{\varepsilon})$, for most $g_r(i, x) = r_i \Theta x$, it is impossible to learn the relative frequency of database items satisfying the predicates $g_r(i, x)$.

Interactive vs. Non-Interactive Mechanisms

- A Stronger Separation for Randomized Response Schemes
  - Randomized response – (data perturbed individually)
    - Randomization operator Z : D ---> {0, 1}* such that
      $T_{San}(x1, ..., xn) = Z(x1), ...,Z(xn)$
  - Strengthening previous theorem
    - Consider fg where the same predicate g: D ---> {0, 1} is applied to all the entries
    - gr(x) = r Θ x is difficult to learn from Z(x)
    - f(x) difficult to learn from Tsan(x) if n is not very large
  - Proposition (Randomized Response)
    Suppose that San is a ε-indistinguishable randomized response mechanism. For at least 2/3 of the values $r \in \{0, 1\}^d \setminus \{0^d\}$, the following two distributions have statistical difference $O(n\varepsilon^{2/3}2^{-d/3})$:

    Distribution 0: $T_{San}(x)$ where each $x_i \in \{x \in \{0, 1\}^d : r \Theta x = 0\}$
    Distribution 1: $T_{San}(x)$ where each $x_i \in \{x \in \{0, 1\}^d : r \Theta x = 1\}$

    if $n = o(2^{d/3}/\varepsilon^{2/3})$, no user can learn the relative frequency of database items satisfying the predicate gr(x) = r Θ x, for most values r.

Questions?