Advanced Topics on Privacy Enhancing Technologies
CS-523
Privacy-preserving Data Publishing II Exercises

1 Survey responses

Carol	1
Bob	0
Peggy	1
Victor	0

D1

Carol	1
Bob	0
Peggy	1
Victor	0
Alice	**1**

D2

Consider the database $D1$ shown above. This database contains the results of a survey among students about password re-usage and the binary value of each record indicates whether this student has ever re-used a password for multiple websites (value 1) or not (value 0). You want to publish the results of the survey as a single query about this database: *How many students in this database have re-used a password?*. Answer the following questions:

1. Suppose you have published the result of your query over $D1$. After you have published your results, another student, Alice, answers the survey and her answer is added to the database (see database $D2$). You decide to update the survey result and publish the result of running your query over $D2$.
   Assume an attacker learns that Alice, and only Alice, has been added to the database and observes the survey results. What is the probability that this attacker infers Alice's true answer and learns whether she has ever re-used a password across websites?

2. Suppose you have realised that there might be a privacy risk for the students who answered the survey question if you publish your results in the clear. So you decide to use a differentially private mechanism to publish the results. This mechanism first computes the true count and then adds noise drawn from a *Laplace* distribution with scale $1/\epsilon$.
   You run this mechanism on the database $D2$ and the noisy answer is 5. Assume an attacker observes this result but already knows the answer of all other students in the database except for Alice's answer (i.e., the attacker knows $D1$). Discuss what the attacker can infer about Alice's survey answer from the noisy query result. How is the attacker's inference impacted by the noise addition? How does the attacker's inference power depend on ϵ?

3. Consider two different settings for running the differentially private mechanism described above. (1) You add noise coming from the Laplace distribution with scale $1/\epsilon$ for $\epsilon = 0.1$ (2) You add noise from the Laplace distribution with scale $1/\epsilon$ for $\epsilon = 0.01$. Which setting achieves better privacy guarantees for Alice and why?

2 Sensitivity

Recall that for two neighbouring datasets D and D' created by the addition or removal of a single record, the sensitivity of a mechanism f is the maximum change in the output of f over all possible inputs

$$\Delta f = \max_{D,D'} ||f(D) - f(D')|| \tag{1}$$

where $||\cdot||$ denotes a vector norm.

1. Assume that you have a database of n records where each record has exactly one attribute value. Find the sensitivity of the following computations f on this database under the L_1-norm:

   - f is a count query and each record takes a value in $\{0, 1\}$
   - f returns the sum over all values and each record takes a value in the range $[a, b]$ with $0 \leq a < b$
   - f return the arithmetic mean and each record takes a value in the range $[a, b]$ with $0 \leq a < b$
   - f returns the median and each record takes a value in the range $[a, b]$ with $0 \leq a < b$
   - f returns the maximum value across all records and and each record takes a value in the range $[a, b]$ with $a \leq 0 < b$
   - f returns the minimum value across all records and each record takes a value in the range $[a, b]$ with $a \leq 0 < b$

2. Suppose that you have a database where each record contains multiple binary attribute which take values in $\{0, 1\}$ and you perform counting queries on this database. Absent any further information, what is the worst-case sensitivity for a fixed but arbitrary list of k count queries over this database?
   Describe a noise addition mechanism that achieves $(\epsilon, 0)-$differential privacy for publishing the results of the k count queries. Which noise distribution would you choose? What would be the noise scale? How would you add the noise to the output vector of size k?

3 Composition

The composability property of differential privacy makes the real-life applications of differential privacy more practical. There are many different composition theorems to bound the total privacy budget depending on the algorithms. Below, we give definitions for two different composition theorems:

Sequential composition. Suppose that we have k algorithms $A_i(D, z_i)$ which are each independently differentially private and z_i denotes some auxiliary input. Suppose that each algorithm A_i is ϵ-differentially private for any auxiliary input z_i. Consider a sequence of computations $(z_1 = A_1(D), z_2 = A_2(D), ...)$ and suppose $A(D) = z_k$.

Theorem 1 *(Sequential Composition [1]): $A(D)$ is $k\epsilon$-differentially private.*

Parallel composition. Now consider the same setting where D_i denotes k disjoint subsets of one database D.

Theorem 2 *(Parallel Composition [1]): $A(D)$ is ϵ-differentially private.*

Note that each mechanism in parallel composition is applied on *independent subsets* of the database.

Looking at the definitions given above, come up with two separate scenarios in which you apply sequential and parallel composition, respectively. Explain your answer.

4 Geo-indistinguishability

Formal treatment of geo-indistinguishability from location privacy. Let us define the notion of *local differential privacy* (LDP). It differs from the standard notion of differential privacy by considering individual inputs as opposed to whole datasets. Let $\mathcal{X}$ be a discrete space of inputs, and $\mathcal{Y}$ a continuous space of outputs of some mechanism $M : \mathcal{X} \to \mathcal{Y}$. Consider a random variable X taking values in the space of inputs. The mechanism satisfies ε-LDP if for any two inputs $x, x' \in \mathcal{X}$, with the output location denoted as $Y = M(X)$, the following holds for any $S \subseteq \mathcal{Y}$:

$$P(Y \in S \mid X = x) \leq \exp(\varepsilon)\, P(Y \in S \mid X = x').$$

For convenience, this statement can also be written as follows:

$$\left| \ln \left(\frac{P(Y \in S \mid X = x)}{P(Y \in S \mid X = x')} \right) \right| \leq \varepsilon$$

Intuitively, if ε is small enough (e.g., 0.1), it means that the mechanism obfuscates inputs in such a way that guessing which exact input—x or x'—corresponds to the observed output y is hard.

Now let us formally define geo-indistinguishability (GeoInd). Let $(\mathcal{X}, d_{\mathcal{X}})$ be a metric space of locations with $d_{\mathcal{X}} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ being the metric. Consider a random variable X taking values in the space of locations. A probabilistic private location release mechanism $M : \mathcal{X} \to \mathcal{X}$ provides $(\varepsilon, d_{\mathcal{X}})$-GeoInd if for any two possible location inputs $x, x' \in \mathcal{X}$, with the output location denoted as $Y = M(X)$, the following holds for any $S \subseteq \mathcal{Y}$:

$$\left| \ln \left(\frac{P(Y \in S \mid X = x)}{P(Y \in S \mid X = x')} \right) \right| \leq \varepsilon\, d_{\mathcal{X}}(x, x').$$

Intuitively, if ε is small enough (e.g., 0.1), it means that the mechanism obfuscates locations in such a way that guessing which exact input location corresponds to the observed location y is hard, as long as the candidate locations are sufficiently close in the metric $d_{\mathcal{X}}$.

1. Let M satisfy $(\varepsilon, d_{\mathcal{X}})$-GeoInd. Characterize M in terms of LDP.

2. For a given input location $x \in \mathbb{R}^2$, the *Planar Laplace mechanism* releases an obfuscated location $y \in \mathbb{R}^2$ by randomly sampling from the probability distribution given by the following density function:

$$f_{Y \mid X = x}(t) = \frac{\varepsilon^2}{2\pi} e^{-\varepsilon\, d_{\mathcal{X}}(x, t)}.$$

   Show that this mechanism satisfies $(\varepsilon, d_{\mathcal{X}})$-GeoInd.

5 Adversarial gains

How to interpret the magic value ε? [Beyond the scope of the exam] The goal of a privacy adversary is, for a given mechanism output y, to tell which of any two given inputs x, x' was more likely to have produced this output. The "best possible" adversary uses the "best possible" classifier for this task, the Bayes-optimal classifier, which can be defined as follows:

$$g(y) = \arg \max_{z \in \{x, x'\}} P(X = z \mid Y = y).$$

That is, on observing y, the classifier simply chooses the input that is more likely according to the *a posteriori* distribution $P(X \mid Y)$.

Even though Bayes classifier is the best possible classifier in the probabilistic setting, it still makes mistakes. E.g., for an output y, if the actual input was x', the classifier makes a wrong prediction if $P(x \mid y) > P(x' \mid y)$.

1. For a fixed observation y, without loss of generality, assume that the actual input was x'. The probability of the classifier guessing incorrectly is therefore $P(x \mid y)$. Assuming (1) that adversary has no background knowledge, i.e., $P(x) = P(x') = \frac{1}{2}$, and (2) that the mechanism M satisfies ε-LDP, find the lower bound on the probability of the adversary making a mistake.

2. The expected error R^* of the Bayes classifier, called *Bayes error* is the expected probability of the classifier making a mistake, going over all possible observed outputs y:

   $$R^* = \mathbb{E}_Y[\min\{P(X = x \mid Y), P(X = x' \mid Y)\}]$$

   The *success rate* of the Bayes classifier is simply $1 - R^*$.

   Using the previous result, express the privacy risk parameter ε of LDP as the lower bound on the error rate (upper bound on the success rate) of the adversary equipped with the Bayes classifier, assuming the adversary has no background information.

3. Which LDP ε corresponds to maximum adversary success rates of 50%, 75%, 90%, 95%?

4. Express the maximum success rate of the adversary with no background knowledge in terms of ε parameter and radius $r = d_{\mathcal{X}}(x, x')$, if the mechanism M satisfies $(\varepsilon, d_{\mathcal{X}})$-GeoInd. Characterize the relationship between maximum success rate and the Euclidean distance between points in the case of $\varepsilon = 0.1$ meters^{-1}.

References

[1] F. McSherry, "Privacy integrated queries:an extensible platform for privacy-preserving data analysis," in *Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data (SIGMOD)*. Association for Computing Machinery, Inc., June 2009, for more information, visit the project page: http://research.microsoft.com/PINQ. [Online]. Available: https://www.microsoft.com/en-us/research/publication/privacy-integrated-queries/