CSCI 5520: Foundations of Data Privacy **Lecture 1**
The Chinese University of Hong Kong, Spring 2015 6 January 2015

Suppose we have a database containing sensitive information (e.g. students' grades, patients' medical records) and we want to enable an outside user to query this database, while preserving the "privacy" of the data. Let's start with an example. Here is a database of students and their CSCI 5520 grades:

name	**gender**	**grade**
Aisha	female	fail
Benny	male	pass
Erica	female	fail
Fabio	male	fail
Johan	male	fail
Ming	male	pass
Orhan	male	pass
Vijay	male	pass
Vuk	male	pass
Yoshi	male	pass

Assume that the names and genders are public and the grades are private. Eve now asks us to provide her with the following information:

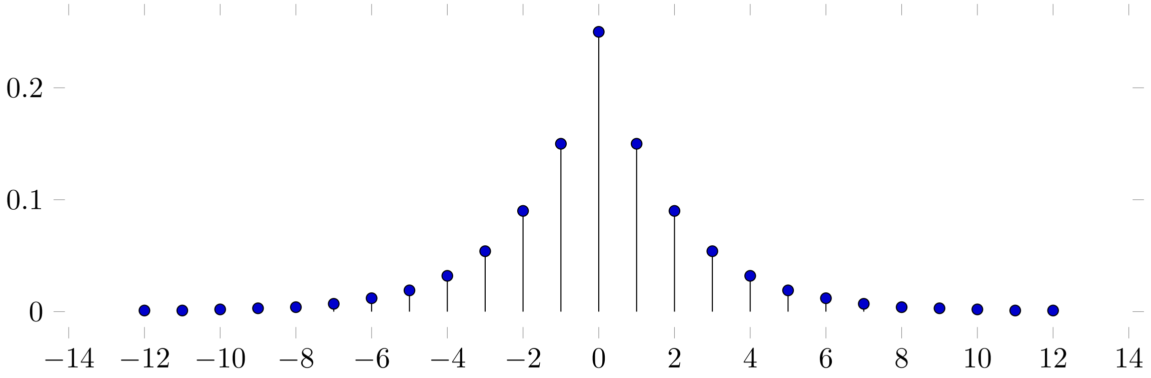
1. How many students passed the course?
2. Did Orhan pass the course?
3. How many female students failed the course?

We would like to provide Eve with the information she wants, but we don't really want her to know the individual grades of students in the class. If we tell her that 6 of students passed the class in response to her first question, she would be getting information about the group as a whole, but she couldn't tell much about how any of the individual students did in the class. But we may refuse to answer her second question as it concerns the privacy of a specific participant in the database. How about the third question? In this specific instance, if we told her that 2 female students failed the course, she would be able to deduce Aisha's and Erica's grades, violating individual privacy.

Hospitals commonly release their medical data to researchers who want to do various statistical analyses (e.g. are cancer rates unusually high among patients that are at least 200cm tall). To protect patients' privacy it is common to remove identifying information like names and ID numbers. However, based on the remaining data and some prior knowledge it is often possible to recover unintended information about individuals, especially if one has access to several databases.

Database privacy studies to what extent one can provide *useful* answers to certain types of users' queries, while preserving the database participants' *privacy.* There are very few settings in which one can provide completely accurate answers and fully preserve the privacy of the users. For example, after finding out that the failure rate of CSCI 5520 is 40%, Eve may conclude that the CSCI 5520 students are a bad lot and change her previously favorable opinion of Orhan.

We can achieve some interesting tradeoffs between the utility of query answers and the privacy of database participants by allowing randomized and approximate answers. Let's go back to the above example. When Eve asks a query q, instead of giving her the true answer A, consider the following *mechanism* that answers by $A + N$, where N is a random variable with the following probability mass function:



That is, if the actual answer to query q is A, we answer a exactly with probability 25%, we answer within the range $A \pm 1$ with probability 55%, we answer within $A \pm 2$ with probability 73%, and so on. This answer may still be useful for Eve as she finds out an approximation to her query.

On the other hand, when Eve asks "How many female students failed the course" and we happen to answer 1, Eve may have trouble telling whether the true answer to her query was 0, 1, or 2. Suppose that before she made her query, Eve believed that each student fails the class independently with probability 40%. How did the information that she found affect her belief that Aisha failed the class? Let AF and EF denote the events that Aisha and Erica failed the class, respectively. Working over the probability space induced by Eve's prior's beliefs and the randomness of the mechanism we obtain:

$$\Pr[AF \mid A + N = 1] = \frac{\Pr[A + N = 1 \mid AF] \Pr[AF]}{\Pr[A + N = 1]}$$

where

$$\begin{aligned}
\Pr[A + N = 1 \mid AF, EF] &= \Pr[N = -1] = 0.15 \\
\Pr[A + N = 1 \mid \overline{AF}, EF] &= \Pr[N = 0] = 0.25 \\
\Pr[A + N = 1 \mid AF, \overline{EF}] &= \Pr[N = 0] = 0.25 \\
\Pr[A + N = 1 \mid \overline{AF}, \overline{EF}] &= \Pr[N = 1] = 0.15.
\end{aligned}$$

By averaging, we obtain

$$\begin{aligned}
\Pr[A + N = 1 \mid AF] &= 0.4 \cdot 0.15 + 0.6 \cdot 0.25 = 0.21 \\
\Pr[A + N = 1] &= 0.4^2 \cdot 0.15 + 0.4 \cdot 0.6 \cdot 0.25 + 0.6 \cdot 0.4 \cdot 0.25 + 0.6^2 \cdot 0.15 = 0.198
\end{aligned}$$

and so

$$\Pr[AF \mid A + N = 1] = \frac{0.21 \cdot 0.4}{0.198} = 0.43.$$

Therefore Eve's belief in the event "Aisha failed the class" changed only from 40% to 43% after observing the answer to the query.

1 Definitions of privacy

Let's try to come up with a definitional framework that captures the above intuition. Given a database x and a query q, we want to design a (possibly randomized) answering mechanism $M(x)$ for q with the following properties:

- **Utility**: The value $M(x)$ is a good approximation to the actual answer that one would obtain when q is queried from x.
- **Privacy**: Seeing the answer $M(x)$ does not change one's beliefs about any specific row of x by much.

In a probabilistic (Bayesian) model of beliefs, the notion of privacy should capture the intuition that Eve's prior distribution X_i on any specific row i in the database was not much affected after seeing the mechanism's answer. However, Eve's prior view of the database does not only include the i-th row, but also all the other rows. If, say, Eve mistakenly believed that all of the CSCI 5520 students failed the course, and upon consulting the mechanism was surprised to find out that most of them actually passed, this would drastically change her belief that, in particular, Aisha failed. We want to disallow such distorted priors in our definition of privacy.

To achieve this, we will require that Eve's prior view of the database is accurate on all rows *except* (possibly) on the i-th row. In other words, we think of a mental experiment in which all the rows of the database have except for the i-th row are revealed to Eve. Then Eve comes up with an arbitrary prior distribution X_i about the i-th row. (For example, if after viewing the rest of the database, Eve believes that Aisha passed with probability 75%, then X_1 would assign 75% probability to the entry $(\text{Aisha}, \text{female}, \text{pass})$ and 25% probability to the entry $(\text{Aisha}, \text{female}, \text{fail})$.) After seeing the answer of the mechanism M, Eve's beliefs about the i-th row may change. Our privacy definition will require that Eve's posterior distribution X_i conditioned on seeing a given answer should not be very different from her prior distribution.

Formally, we think of each row in a database as taking values in some finite domain D. For instance, D could consist of all triples of the form (name, gender, grade). Let's fix the number n of rows in the database. A *database* x is an element of the power set D^n. A *query* is a function q from D^n to some set of values.

To begin with, let us take a special type of query that captures the above examples, as well as many other settings of interest. The *counting query* q_P associated to predicate $P \colon D \to \{\texttt{true}, \texttt{false}\}$ is an integer-valued query given by the formula

$$q_P(x) = \text{number of rows } i \text{ such that } P(x_i) \text{ is true.}$$

For example, the queries “How many students passed”, “Did Orhan pass”, and “How many female students passed” are all counting queries for the grades table.

A *mechanism* for query q is a possibly randomized algorithm that on input a database x outputs an answer $M(x)$. If M is randomized, then for every fixed x, $M(x)$ is a random variable. Intuitively, the mechanism $M(x)$ should be useful if the mechanism's answer $M(x)$ is typically close to the actual answer $q(x)$. I do not know of a definition of utility that captures all settings of interest so I won't attempt to give one. For numerical queries, one natural measure of utility could be the inverse of the standard deviation

$$\text{utility}(M) = \frac{1}{\max_x \sqrt{\mathrm{E}[(M(x) - q(x))^2]}}.$$

Let us now define privacy. Following the intuition we suggested, we want to say that for any row i and any prior distribution X_i on the contents of this row, the posterior distribution X_i conditioned on observing the mechanism's answer “looks like” the prior distribution. Here is a fairly strong quantitative definition that captures this:

Definition 1. We say mechanism M over D^n is ε-semantically private if for every $i \in [n]$, every distribution X over databases in which all rows but the i-th are fixed, every predicate P over D, and every possible output y of $M(X)$,

$$e^{-\varepsilon} \Pr[P(X_i)] \leq \Pr[P(X_i) \mid M(X) = y] \leq e^{\varepsilon} \Pr[P(X_i)]$$

where X_i is the i-th row of X.

To make sense of this definition, let us look at the extreme setting in which $\varepsilon = 0$. Then the posterior and prior probabilities on the i-th row must be the same, so observing the answer of the mechanism does not reveal any additional information about any specific row of the database. Although such a mechanism is extremely private, it is not useful at all: The value $M(X)$ does not teach us anything about the database. However, if we set ε to a small number like 0.05, then the probabilities would still be similar (recall that when ε is small, e^{ε} is about $1 + \varepsilon$), but now $M(X)$ might contain useful information.

One way to make ensure that the answer of the mechanism does not affect by much Eve's beliefs about any specific row of the database is to make the answer of the mechanism essentially independent of the contents of that row. This requirement is formalized by the notion of differential privacy:

Definition 2. We say mechanism M is ε-differentially private if for every pair of databases x, x' that differ only in one row, and every possible answer y of M,

$$\Pr[M(x) = y] \leq e^{\varepsilon} \Pr[M(x') = y]. \tag{1}$$

Again, let's look at the extreme setting $\varepsilon = 0$ so $e^{\varepsilon} = 1$. Then we must have $\Pr[M(x) = y] \leq \Pr[M(x') = y]$. By switching the roles of x and x' we obtain the same inequality in the other direction, and therefore it must be that $\Pr[M(x) = y] = \Pr[M(x') = y]$. Since we require that this equality holds for all y, it must be that $M(x)$ and $M(x')$ are identically distributed. This can only happen if M is independent of the database, in which case it is not useful. By setting ε to a small nonzero value, we can hope to get some tradeoff between the mechanism's utility and its privacy.

By switching the role of x and x' as we just did, we can replace (1) by the stronger condition

$$e^{-\varepsilon}\Pr[M(x') = y] \le \Pr[M(x) = y] \le \Pr[M(x') = y].$$

Differential privacy is easier to work with than semantic privacy. To verify that a mechanism is differentially private, we do not need to worry about prior and posterior beliefs at all, but merely need to show that the mechanism does not "distinguish" much between pairs of databases that differ in a single row.[1] Fortunately, differential privacy implies semantic privacy.

Theorem 3. *If M is ε-differentially private, then M is ε-semantically private.*

Proof. Assume M is ε-differentially private. Then for every pair of databases x, x' that differ in the i-th row and every value y,

$$e^{-\varepsilon}\Pr[M(x') = y] \le \Pr[M(x) = y] \le e^{\varepsilon}\Pr[M(x') = y].$$

Here, the probabilities are taken over the randomness of the mechanism. Fix x and let X be an arbitrary distribution over those x' that differ from x in at most the i-th row. By averaging these inequalities, we get that

$$e^{-\varepsilon}\Pr[M(X) = y] \le \Pr[M(x) = y] \le e^{\varepsilon}\Pr[M(X) = y].$$

where now the probabilities are taken also over X. By Bayes' rule,

$$\Pr[X = x \mid M(X) = y] = \frac{\Pr[M(x) = y]}{\Pr[M(X) = y]} \cdot \Pr[X = x]$$

for every possible answer y of $M(X)$. Combining the last two formulas, we get that

$$e^{-\varepsilon}\Pr[X = x] \le \Pr[X = x \mid M(X) = y] \le e^{\varepsilon}\Pr[X = x].$$

These inequalities hold for all possible databases x. If we sum over all x such that $P(x_i)$ holds, we obtained the desired consequence

$$e^{-\varepsilon}\Pr[P(X_i)] \le \Pr[P(X_i) \mid M(X) = y] \le e^{\varepsilon}\Pr[P(X_i)]. \qquad \square$$

2 The Laplace mechanism

Inspired by our example, we construct and analyze a differentially private mechanism for counting queries. The *Laplace mechanism* with privacy parameter $\varepsilon > 0$ answers a counting query q by $M(x, q) = q(x) + N$, where N is chosen from the Laplace distribution $\mathrm{Lap}(1/\varepsilon)$

$$\Pr[N = t] = \frac{1}{Z}e^{-\varepsilon|t|}, \quad t \text{ is an integer.}$$

[1] If you have studied cryptography, you may have some experience with semantic and indistinguishability based notions of security.

Here $Z = \sum_{t=-\infty}^{\infty} e^{-\varepsilon|t|}$ is a normalization factor which ensures the above formula describes a probability distribution over the integers.

We now show that the Laplace mechanism is ε-differentially private for counting queries. Let x and x' be databases that differ in exactly one row. Because q is a counting query, we have $|q(x) - q(x')| \leq 1$. So for every value y,

$$\begin{aligned}\Pr[M(x) = y] = \Pr[q(x) + N = y] = \Pr[N = y - q(x)] &= \frac{1}{Z} e^{-\varepsilon|y-q(x)|} \\ &\leq \frac{1}{Z} e^{-\varepsilon|y-q(x')|+\varepsilon} = e^{\varepsilon} \cdot \frac{1}{Z} e^{-\varepsilon|y-q(x')|} = e^{\varepsilon} \Pr[M(x') = y].\end{aligned}$$

and so M is ε-differentially private.

What about the utility of the Laplace mechanism? If our notion of utility is the inverse of the standard deviation, we get that the utility of the mechanism is the inverse of the standard deviation σ of the Laplace distribution $\mathrm{Lap}(1/\varepsilon)$, which is $\sigma = \sqrt{2}/\varepsilon$. So the utility of this mechanism is $\varepsilon/\sqrt{2}$. The Laplace mechanism illustrates a general phenomenon: The more private we want our mechanism to be, the less useful it tends to be.

3 The product mechanism

One nice property of differential privacy is that this notion is preserved (or rather, it degrades gracefully) if we allow more queries. Given d mechanisms $M_1, \dots, M_d$ over D^n, the *product mechanism* M on input x outputs the vector of answers $(M_1(x), \dots, M_d(x))$, where each of the mechanisms is run with independent randomness. This mechanism can be used to answer several queries on the same database.

Claim 4. *Suppose M_i is ε-differentially private for every i. Then M is $d\varepsilon$-differentially private.*

Proof. Let x and x' be databases that differ in only one row. By independence,

$$\begin{aligned}\Pr[M(x) = (y_1, \dots, y_d)] &= \Pr[M_1(x) = y_1] \cdots \Pr[M_d(x) = y_d] \\ &\leq (e^{\varepsilon} \Pr[M_1(x') = y_1]) \cdots (e^{\varepsilon} \Pr[M_d(x) = y_d]) = e^{d\varepsilon} \Pr[M(x') = y]\end{aligned}$$

so M is $d\varepsilon$-differentially private. □

References

These notes are based on Chapters 2 and 3 of the survey *The Algorithmic Foundations of Differential Privacy* by Cynthia Dwork and Aaron Roth. Theorem 3 is from these lecture notes of Salil Vadhan. These notes also elaborate more of the notion of “closeness” of distributions (1) in the definition of differential privacy in contrast to the more common statistical distance between distributions.