1 Database privacy: an example

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 5440 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:

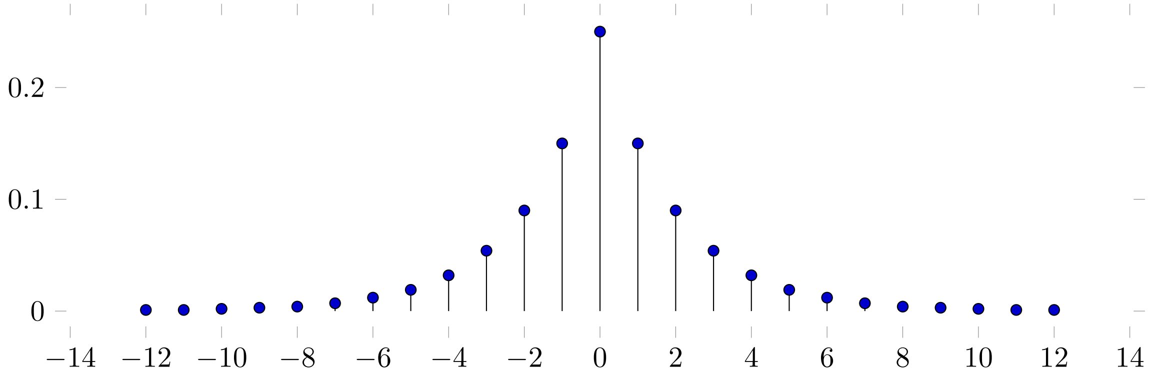
1. How many students passed the course?
2. Did Orhan pass the course?
3. How many female students failed the course?

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 5440 is 40%, Eve may conclude that the CSCI 5440 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 sampled from the following distribution:



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.

2 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 $M(x, q)$ with the following properties:

- **Utility**: The value $M(x, q)$ is a good approximation to the actual answer that one would obtain when q is queried from x.
- **Privacy**: Seeing the answer $M(x, q)$ does not change one's beliefs about any specific row of x by much. In a probabilistic (Bayesian) model of beliefs, we can formalize this requirement as follows. For any prior probability distribution X over the rows of the database modeling the beliefs about various individuals, the posterior distribution of the ith row X_i of X *conditioned* on seeing the answer $M(X, q)$ is "close" to the distribution X_i.

Formally, we think of each row in a database table as taking values in some finite set 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 table* x is an element of the power set D^n. A *query* is a function q from D^n to some set of values.

Among all queries, the following kind will play an important role. 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* is a possibly randomized algorithm that on input a database x and query q outputs an answer $M(x, q)$. If M is randomized, for every fixed x and q, $M(x, q)$ is a random variable. Intuitively, the mechanism $M(x, q)$ should be useful if the mechanism's answer $M(x, q)$ 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

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

Let us now define privacy. Following the intuition we suggested, we want to say that for any prior distribution X on n-row tables, no test can distinguish the ith row X_i from the posterior distribution on the ith row after observing the mechanism's answer $M(x, q)$. Here is a fairly strong quantitative definition that captures this:

Definition 1. Let q be a query over an n-row database $x \in D^n$. We say mechanism M is ε-semantically private for q if for every distribution $X \sim D^n$, every $i \in [n]$, every y such that $\Pr[M(X,q) = y] > 0$, and every test $A \colon D \to \{0,1\}$,

$$\big|\Pr[A(X_i) = 1] - \Pr[A(X_i) = 1 \mid M(X,q) = y]\big| \le \varepsilon$$

where X_i is the i-th row of X.

By analogy with cryptography, we might expect that there is also a definition that talks about indistingushability of tables. This is the notion of differential privacy.

Definition 2. We say mechanism M is ε-differentially private for q if for every $i \in [n]$, every pair of tables x, x' that differ only in row i, and every test $A \colon D \to \{0,1\}$,

$$\Pr[A(M(x,q)) = 1] \le e^{\varepsilon} \Pr[A(M(x',q)) = 1]. \tag{1}$$

To understand this definition let us look at the extreme setting where $\varepsilon = 0$ so $e^{\varepsilon} = 1$. Then we must have $\Pr[A(M(x,q)) = 1] \le \Pr[A(M(x',q)) = 1]$. By switching the roles of x and x' we obtain the same inequality in the other direction, and therefore it must be that $\Pr[A(M(x,q)) = 1] = \Pr[A(M(x',q)) = 1]$. Since we require that this equality holds for all tests A, it must be that $M(x,q)$ and $M(x',q)$ are identically distributed. This can only happen if M is independent of the database, in which case it doesn't appear to be very useful at all.

By setting ε to a small nonzero value, we can hope to get some tradeoff between the mechanism's utility and its privacy. Since $e^{\varepsilon} = 1 + \varepsilon + O(\varepsilon^2)$, by the same reasoning we can interpret (1) as asking that

$$\begin{aligned}&\big|\Pr[A(M(x,q)) = 1] - \Pr[A(M(x',q)) = 1]\big| \\ &\qquad\qquad \le (\varepsilon + O(\varepsilon^2)) \max\big\{\Pr[A(M(x,q)) = 1], \Pr[A(M(x',q)) = 1]\big\}.\end{aligned}$$

Notice that this is *stronger* than the usual statistical distance requirement, in which the right hand side is ε. There are simple examples which show Definition 2 would become meaningless if inequality (1) was replaced with the requirement $|\Pr[A(M(x,q)) = 1] - \Pr[A(M(x',q)) = 1]| \le \varepsilon$.

Claim 3. *If M is ε-semantically private for q, then M is $(2\varepsilon + O(\varepsilon^2))$-differentially private for q.*

Proof. Assume M is ε-semantically private. Let x_0 and x_1 be any two databases that differ in row i. Let

$$X = \begin{cases} x_0, & \text{with prob. } 1/2 \\ x_1, & \text{with prob. } 1/2 \end{cases} \quad \text{and} \quad A(r) = \begin{cases} 0, & \text{if } r \text{ is the } i\text{'th row of } x_0 \\ 1, & \text{if } i \text{ is the } i\text{'th row of } x_1. \end{cases}$$

Since x_0 and x_1 differ in the i'th row, A is well-defined. Then $\Pr[A(X_i) = 1] = 1/2$ and so for every y such that $\Pr[M(X) = y] > 0$,

$$\big|\Pr[A(X_i) = 1 \mid M(X) = y] - 1/2\big| \le \varepsilon.$$

Looking at A, we have that

$$\Pr[X = x_0 \mid M(X) = y] = \tfrac{1}{2} - \gamma(y) \quad \text{and} \quad \Pr[X = x_1 \mid M(X) = y] = \tfrac{1}{2} + \gamma(y)$$

for some $\gamma(y)$ satisfying $-\varepsilon \le \gamma(y) \le \varepsilon$. On the other hand we have

$$\begin{aligned}\Pr[M(x_0) = y] &= \Pr[M(X) = y \mid X = x_0] \\ &= \frac{\Pr[M(X) = y \text{ and } X = x_0]}{\Pr[X = x_0]} \\ &= \Pr[X = x_0 \mid M(X) = y] \cdot 2 \Pr[M(X) = y]\end{aligned}$$

where we use $\Pr[X = x_0] = 1/2$. We get an analogous equation for x_1. Taking the ratio of the two we obtain that

$$\frac{\Pr[M(x_0) = y]}{\Pr[M(x_1) = y]} \le \frac{\Pr[X = x_0 \mid M(X) = y]}{\Pr[X = x_1 \mid M(X) = y]} = \frac{1/2 - \gamma(y)}{1/2 + \gamma(y)} = e^{2\varepsilon + O(\varepsilon^2)}$$

where the last step follows by Taylor expansion of the function $f(t) = \ln((1/2-t)/(1/2+t))$ in the range $-\varepsilon \le t \le \varepsilon$. (If $\Pr[M(X) = y] = 0$, then both probabilities are zero.)

Now let B be an arbitrary test that maps outputs of M to 0 or 1. Then

$$\begin{aligned}\Pr[B(M(x_0)) = 1] &= \sum_{y:\, B(y)=1} \Pr[M(x_0) = y] \\ &\le \sum_{y:\, B(y)=1} e^{2\varepsilon + O(\varepsilon^2)} \Pr[M(x_1) = y] \\ &= e^{2\varepsilon + O(\varepsilon^2)} \Pr[B(M(x_1)) = 1].\end{aligned}$$

Since this holds for every B and every pair x_0, x_1 that differ in one row, M is $(2\varepsilon + O(\varepsilon^2))$-differentially private. □

Claim 4. *If M is ε-differentially private for q, then M is $(\varepsilon + O(\varepsilon^2))$-semantically private for q.*

Proof. Let $\varepsilon' = \varepsilon + O(\varepsilon^2)$. If M is not ε'-semantically private for q, then there exists a distribution X, an index i, an answer y and a test A such that

$$\Pr[A(X_i) = 1 \mid M(X) = y] - \Pr[A(X_i) = 1] > \varepsilon'.$$

Let X_{-i} denote the marginal distribution on all but the i'th row of X. By averaging we get

$$\mathrm{E}_{X_{-i}}\big[\Pr[A(X_i) = 1 \mid M(X_i, X_{-i}) = y, X_{-i}] - \Pr[A(X_i) = 1 \mid X_{-i}]\big] > \varepsilon'$$

Let x_{-i} be the value of X_{-i} that maximizes the probability on the left. Then

$$\Pr'[A(X_i) = 1 \mid M'(X_i) = y] - \Pr'[A(X_i) = 1] > \varepsilon'$$

where $\Pr'[\,\cdot\,]$ denotes $\Pr[\,\cdot \mid X_{-i} = x_{-i}]$ and $M'(x_i)$ denotes $M(x_i, x_{-i})$. We can rewrite the last inequality in the form

$$\Pr'[M'(X_i) = y \mid A(X_i) = 1] \Pr'[A(X_i) = 1] - \Pr'[M'(X_i) = y] \Pr'[A(X_i) = 1] > \varepsilon' \Pr'[M'(X_i) = y].$$

Since $\Pr'[A(X_i) = 1] \le 1$, we have

$$\Pr'[M'(X_i) = y \mid A(X_i) = 1] - \Pr'[M'(X_i) = y] > \varepsilon' \Pr'[M'(X_i) = y].$$

from where

$$\Pr'[M'(X_i) = y \mid A(X_i) = 1] > (1 + \varepsilon') \Pr'[M'(X_i) = y] \geq e^{\varepsilon} \Pr'[M'(X_i) = y]$$

by our choice of ε'. Now let x_0 be table that equals x_{-i} on all other rows that maximizes $\Pr'[M'(x_0) = y]$, and x_1 be the table that equals x_{-i} on all other rows that minimizes $\Pr'[M'(x_1) = y]$. Then we must have

$$\Pr[M(x_0) = y] \geq \Pr'[M'(X_i) = y \mid A(X_i) = 1] > e^{\varepsilon} \Pr'[M'(X_i) = y] \geq e^{\varepsilon} \Pr[M(x_1) = y]$$

and so M is not ε-differentially private for q. (The distinguisher on input x outputs 1 if $M(x) = y$ and 0 if not.) □

One nice property of differential privacy is that this notion is preserved (or rather, it degrades gracefully) if we allow more queries. Formally, given queries $q_1 \colon D^n \to R_1, \dots, q_t \colon D^n \to R_t$ their product query $q \colon D^n \to R_1 \times \cdots \times R_t$, $q = q_1 \times \cdots \times q_t$ is given by the formula

$$q(x) = (q_1(x), \dots, q_t(x)).$$

The following claim has an easy proof.

Claim 5. *If M is ε-differentially private for all of $q_1, \dots, q_t$, then it is $t\varepsilon$-differentially private for the product query $q_1 \times \cdots \times q_t$.*

3 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

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

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,

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

By the same argument as in the proof of Claim 3, we conclude that 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 with parameter ε, 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.

In cryptography we usually think of the indistinguishability parameter ε as taking extremely small values, e.g. $\varepsilon = 2^{-100}$. These kinds of parameters don't make much sense in the Laplace mechanism, since then the utility of the mechanism would be extremely small. To get reasonable utility we may want to use the Laplace mechanism with larger values of ε, say $\varepsilon = 0.1$. This could be a reasonable level of privacy if we allow only one query to the database. However, Claim 5 suggests that once we make $1/\varepsilon = 10$ queries, no privacy will be left!

So it seems that the Laplace mechanism is not terribly useful. However, in the next lecture we will see (time permitting) that the Laplace mechanism plays a role in the construction of a more complex mechanism with a better privacy-utility tradeoff.