

CSE660

Differential Privacy

September 13, 2017

Marco Gaboardi

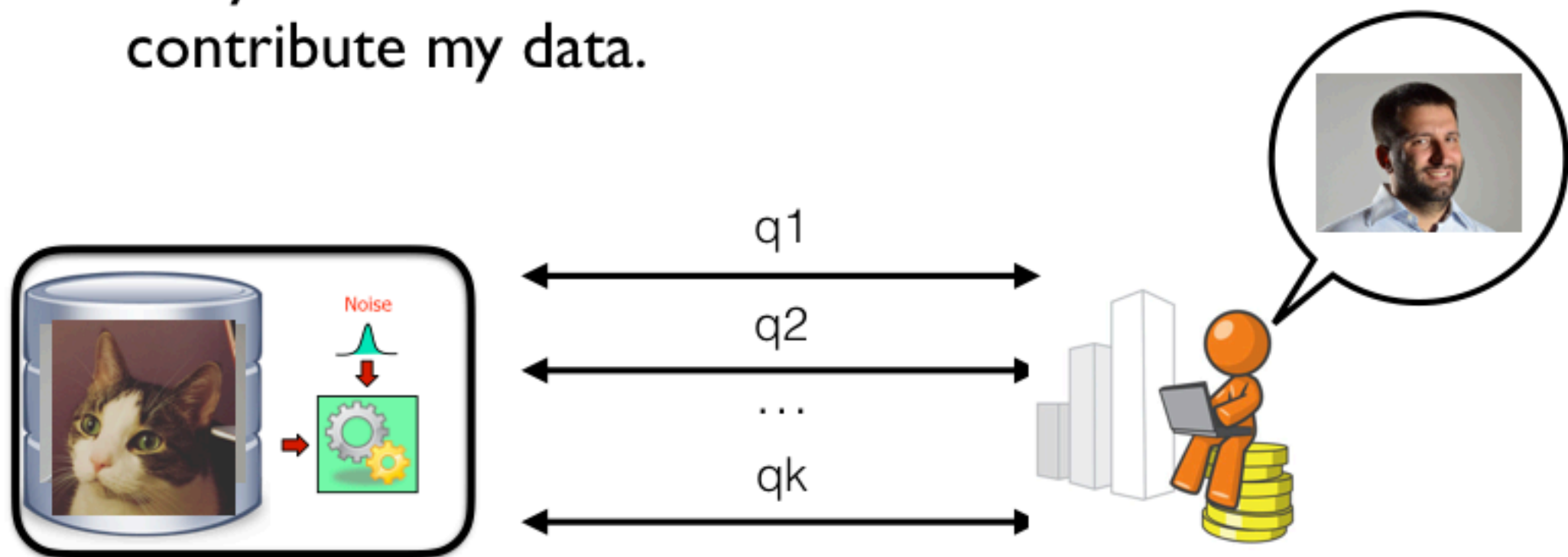
Room: 338-B

gaboardi@buffalo.edu

<http://www.buffalo.edu/~gaboardi>

Privacy-preserving data analysis?

- The analyst learn the same about me after the analysis as what she would have learnt if I didn't contribute my data.



Adjacent databases

- We can formalize the concept of contributing my data or not in terms of a notion of distance between datasets.

- Given two datasets $D, D' \in X^n$, their distance is defined as:

$$D \Delta D' = |\{k \leq n \mid D(k) \neq D'(k)\}|$$

- We will call two datasets adjacent when $D \Delta D' = 1$ and we will write $D \sim D'$.

(ϵ, δ) -Differential Privacy

Definition

Given $\epsilon, \delta \geq 0$, a probabilistic query $Q: X^n \rightarrow R$ is (ϵ, δ) -differentially private iff

for all adjacent database b_1, b_2 and for every $S \subseteq R$:

$$\Pr[Q(b_1) \in S] \leq \exp(\epsilon) \Pr[Q(b_2) \in S] + \delta$$

Randomized Response

[Warner65]

Suppose I ask a yes/no question.



biased
coin



True
answer



Opposite
answer

The value of the bias is what determines the epsilon

An example

```
AlmostRandom (b : bool) : bool {  
  if coinToss()  
  then  
    return b;  
  else  
    return coinToss();  
}
```

An example

AlmostRandom is $(\ln 3, 0)$ -differentially private

Let's consider the case we have two adjacent data b and b' . By the fact that they are adjacent we know that one of them is 1 and one of them is 0. Without loss of generality let's assume $b=1$ and $b'=0$.

We have:

$$\Pr[\text{AR}(b) = 1] = 3/4 \quad \Pr[\text{AR}(b') = 1] = 1/4$$

$$\Pr[\text{AR}(b) = 0] = 1/4 \quad \Pr[\text{AR}(b') = 0] = 3/4$$

So for all b, b' :

$$\left| \frac{\Pr[\text{AR}(b) = R]}{\Pr[\text{AR}(b') = R]} \right| \leq 3$$

Randomized Response

Algorithm 1 Pseudo-code for Randomized Response

```
1: function RANDOMIZEDRESPONSE( $D, q, \epsilon$ )
2:   for  $k \leftarrow 1$  to  $|D|$  do
3:      $S_i \leftarrow \begin{cases} q(d_i) & \text{with probability } \frac{e^\epsilon}{1+e^\epsilon} \\ \neg q(d_i) & \text{with probability } \frac{1}{1+e^\epsilon} \end{cases}$ 
4:   end for
5:   return  $\frac{(\text{sum } S)}{|D|}$ 
6: end function
```

Example

Let's consider a medical dataset containing informations on whether each patient has a disease.

We can have $d_i=1$ if patient i has the disease and $d_i=0$ otherwise.

We can use randomized response to estimate the proportion of patient that have the disease.

The noise that each individual add protect his/her value.

A CS Example

Google Chrome RAPPOR project uses randomized response to collect statistics of opt-in users of the Chrome browser.

They collect statistics about the users home pages, the setting of the user browser, etc.

RAPPOR uses randomized response to estimate the proportion of users with specific settings.

The noise that each individual add protect his/her value.

Randomized Response

Privacy Theorem:

Randomized response is ϵ -differentially private.

Proof:

We need to consider two databases $D, D' \in X^n$ and every possible output.

Since $|D|=|D'|=n$ is fixed, the result will depend only on the value of $S \in \{0, 1\}^n$.

Since $D \sim D'$ there is only one element where they differ. Let's name this element k .

Randomized Response

Privacy Theorem:

Randomized response is ϵ -differentially private.

Continued Proof:

For every $j \neq k$ we have:

$$\Pr[S_j = q(d_j)] = \Pr[S_j = q(d'_j)]$$

For k we have two cases:

- 1 - either $q(d_k) = q(d'_k)$
- 2 - or $q(d_k) \neq q(d'_k)$

Randomized Response

Privacy Theorem:

Randomized response is ϵ -differentially private.

Continued Proof:

In the case 1 we have

$$\Pr[S_k = q(d_k)] = \Pr[S_k = q(d'_k)]$$

The case 2 where $q(d_k) \neq q(d'_k)$ is more interesting.

Randomized Response

Privacy Theorem:

Randomized response is ϵ -differentially private.

Continued Proof:

In this case we have for example

$$\frac{\Pr[S_k = q(d_k)]}{\Pr[S_k = q(d'_k)]} = \frac{\Pr[S_k = q(d_k)]}{\Pr[S_k = \neg q(d_k)]} = \frac{\left(\frac{e^\epsilon}{1+\epsilon}\right)}{\left(\frac{1}{1+\epsilon}\right)} = e^\epsilon$$

By a similar reasoning we can show

$$\frac{\Pr[S_k = q(d'_k)]}{\Pr[S_k = q(d_k)]} = \frac{\Pr[S_k = \neg q(d_k)]}{\Pr[S_k = q(d_k)]} = \frac{\left(\frac{1}{1+\epsilon}\right)}{\left(\frac{e^\epsilon}{1+\epsilon}\right)} = \frac{1}{e^\epsilon}$$

Randomized Response

Privacy Theorem:

Randomized response is ϵ -differentially private.

Continued Proof:

Putting the pieces together and using the fact that each coin is independent from each other we have:

$$\begin{aligned} \frac{\Pr[S = RR(D, q, \epsilon)]}{\Pr[S = RR(D', q, \epsilon)]} &= \frac{\Pr[S_k = q(d_k)] \prod_{j \neq k} \Pr[S_j = q(d_j)]}{\Pr[S_k = q(d'_k)] \prod_{j \neq k} \Pr[S_j = q(d'_j)]} \\ &= \frac{\Pr[S_k = q(d_k)]}{\Pr[S_k = q(d'_k)]} = e^\epsilon \end{aligned}$$

Randomized Response

Privacy Theorem:

Randomized response is ϵ -differentially private.

Continued Proof:

Similarly we can prove

$$\frac{\Pr[S = RR(D', q, \epsilon)]}{\Pr[S = RR(D, q, \epsilon)]} = \frac{1}{e^\epsilon}$$

and this concludes the proof.

Randomized Response

Question: How accurate is the answer that we get from randomized response?

Accuracy

- There are usually two main ways to measure accuracy:
- By comparing the noised result with the one that we would have without noise,
- By comparing the noised result with the one that we would obtain on the population.

Accuracy Statements

We can give statements about the accuracy of our algorithms by using formulas like

$$\Pr[X \geq \alpha] \leq \beta$$

beta is the probability of failure

Here X is a random variable representing some measurement on the differentially private output. α is a given value of accuracy we want to achieve.

Accuracy Statements

We can give statements about the accuracy of our algorithms by using formulas like

$$\Pr[X \geq \alpha] \leq \beta$$

For example, if we want to compare the noisy answer with the one without noise:

$$\Pr[|a - \hat{a}| \geq \alpha] \leq \beta$$

Randomized Response

Accuracy Theorem:

$$\Pr_{r \leftarrow RR(D, q, \epsilon)} \left[\left| \frac{1 + e^\epsilon}{e^\epsilon - 1} \left(r - \frac{1}{1 + e^\epsilon} \right) - q(D) \right| \geq \frac{1 + e^\epsilon}{(e^\epsilon - 1)} \sqrt{\frac{\log(2/\beta)}{2n}} \right] \leq \beta$$



This represents the variable measuring the difference between the noised answer and the non-noised one.



This is our alpha. Notice that we express it in terms of beta.

Randomized Response

Accuracy Theorem:

$$\Pr_{r \leftarrow RR(D, q, \epsilon)} \left[\left| \frac{1 + e^\epsilon}{e^\epsilon - 1} \left(r - \frac{1}{1 + e^\epsilon} \right) - q(D) \right| \geq \frac{1 + e^\epsilon}{(e^\epsilon - 1)} \sqrt{\frac{\log(2/\beta)}{2n}} \right] \leq \beta$$

Proof:

First we can compute the expected value of each S :

$$\begin{aligned} \mathbf{E}[S_j] &= q(d_j) \left(\frac{e^\epsilon}{1 + e^\epsilon} \right) + \neg q(d_j) \left(\frac{1}{1 + e^\epsilon} \right) \\ &= q(d_j) \left(\frac{e^\epsilon}{1 + e^\epsilon} \right) + (1 - q(d_j)) \left(\frac{1}{1 + e^\epsilon} \right) \\ &= \frac{q(d_j)(e^\epsilon - 1)}{1 + e^\epsilon} + \frac{1}{1 + e^\epsilon} \end{aligned}$$

Additive Chernoff Bound

Theorem 1.2 (Additive Chernoff Bound). Let X_1, \dots, X_n be i.i.d random variables such that $0 \leq X_i \leq 1$ for every $1 \leq i \leq n$. Let $S = \frac{1}{n} \sum_{i=1}^n X_i$ denote their mean and $E[S]$ their expected mean, where $E[S] = \frac{1}{n} \sum_{i=1}^n E[X_i]$ by linearity of expectation, then for every λ we have:

$$\Pr[|S - E[S]| \geq \lambda] \leq 2e^{-2\lambda^2 n}$$

Randomized Response

Accuracy Theorem:

$$\Pr_{r \leftarrow RR(D, q, \epsilon)} \left[\left| \frac{1 + e^\epsilon}{e^\epsilon - 1} \left(r - \frac{1}{1 + e^\epsilon} \right) - q(D) \right| \geq \frac{1 + e^\epsilon}{(e^\epsilon - 1)} \sqrt{\frac{\log(2/\beta)}{2n}} \right] \leq \beta$$

Continued Proof:

Applying the Chernoff bound and the fact that

$$\mathbf{E}[S_j] = \frac{q(d_j)(e^\epsilon - 1)}{1 + e^\epsilon} + \frac{1}{1 + e^\epsilon}$$

we can prove

$$\Pr \left[\left| \frac{1}{n} \sum_j S_j - \frac{1}{n} \sum_j \left(\frac{q(d_j)(e^\epsilon - 1)}{1 + e^\epsilon} + \frac{1}{1 + e^\epsilon} \right) \right| \geq \lambda \right] \leq 2e^{-2\lambda^2 n}$$

Randomized Response

Accuracy Theorem:

$$\Pr_{r \leftarrow RR(D, q, \epsilon)} \left[\left| \frac{1 + e^\epsilon}{e^\epsilon - 1} \left(r - \frac{1}{1 + e^\epsilon} \right) - q(D) \right| \geq \frac{1 + e^\epsilon}{(e^\epsilon - 1)} \sqrt{\frac{\log(2/\beta)}{2n}} \right] \leq \beta$$

Continued Proof:

This can be rewritten as

$$\Pr \left[\left| \frac{1 + e^\epsilon}{e^\epsilon - 1} \left(\frac{1}{n} \sum_j S_j - \frac{1}{1 + e^\epsilon} \right) - \frac{1}{n} \sum_j q(d_j) \right| \geq \frac{1 + e^\epsilon}{(e^\epsilon - 1)} \lambda \right] \leq 2e^{-2\lambda^2 n}$$

which by definition of counting queries is equivalent to

$$\Pr_{r \leftarrow RR(D, q, \epsilon)} \left[\left| \frac{1 + e^\epsilon}{e^\epsilon - 1} \left(r - \frac{1}{1 + e^\epsilon} \right) - q(D) \right| \geq \frac{1 + e^\epsilon}{(e^\epsilon - 1)} \lambda \right] \leq 2e^{-2\lambda^2 n}$$

Randomized Response

Accuracy Theorem:

$$\Pr_{r \leftarrow RR(D, q, \epsilon)} \left[\left| \frac{1 + e^\epsilon}{e^\epsilon - 1} \left(r - \frac{1}{1 + e^\epsilon} \right) - q(D) \right| \geq \frac{1 + e^\epsilon}{(e^\epsilon - 1)} \sqrt{\frac{\log(2/\beta)}{2n}} \right] \leq \beta$$

Continued Proof:

By setting

$$\lambda = \sqrt{\frac{\log(2/\beta)}{2n}}$$

we can conclude

$$\Pr_{r \leftarrow RR(D, q, \epsilon)} \left[\left| \frac{1 + e^\epsilon}{e^\epsilon - 1} \left(r - \frac{1}{1 + e^\epsilon} \right) - q(D) \right| \geq \frac{1 + e^\epsilon}{(e^\epsilon - 1)} \sqrt{\frac{\log(2/\beta)}{2n}} \right] \leq \beta$$

Randomized Response

Accuracy Theorem:

$$\Pr_{r \leftarrow RR(D, q, \epsilon)} \left[\left| \frac{1 + e^\epsilon}{e^\epsilon - 1} \left(r - \frac{1}{1 + e^\epsilon} \right) - q(D) \right| \geq \frac{1 + e^\epsilon}{(e^\epsilon - 1)} \sqrt{\frac{\log(2/\beta)}{2n}} \right] \leq \beta$$

Intuitive reading: with high probability we have:

$$\left| r - q(D) \right| \leq O\left(\frac{1}{\sqrt{n}}\right)$$

Randomized Response

Accuracy Theorem:

$$\Pr_{r \leftarrow RR(D, q, \epsilon)} \left[\left| \frac{1 + e^\epsilon}{e^\epsilon - 1} \left(r - \frac{1}{1 + e^\epsilon} \right) - q(D) \right| \geq \frac{1 + e^\epsilon}{(e^\epsilon - 1)} \sqrt{\frac{\log(2/\beta)}{2n}} \right] \leq \beta$$

Intuitive reading: with high probability we have:

$$\left| r - q(D) \right| \leq O\left(\frac{1}{\sqrt{n}}\right)$$

$$\frac{1}{\sqrt{n}} = \frac{\sqrt{n}}{n}$$

Notice that this is of the same order as the normalized sampling error.

Example revisited

Accuracy Theorem:

$$\Pr_{r \leftarrow RR(D, q, \epsilon)} \left[\left| \frac{1 + e^\epsilon}{e^\epsilon - 1} \left(r - \frac{1}{1 + e^\epsilon} \right) - q(D) \right| \geq \frac{1 + e^\epsilon}{(e^\epsilon - 1)} \sqrt{\frac{\log(2/\beta)}{2n}} \right] \leq \beta$$

Let's consider again the example of the medical dataset containing informations on whether each patient has a disease ($d_i=0$ or $d_i=1$).

We can use randomized response to estimate the proportion of patient that have the disease. **We need to fix the parameters.**

Example revisited

Accuracy Theorem:

$$\Pr_{r \leftarrow RR(D, q, \epsilon)} \left[\left| \frac{1 + e^\epsilon}{e^\epsilon - 1} \left(r - \frac{1}{1 + e^\epsilon} \right) - q(D) \right| \geq \frac{1 + e^\epsilon}{(e^\epsilon - 1)} \sqrt{\frac{\log(2/\beta)}{2n}} \right] \leq \beta$$

Let's fix the following values for the parameters:

$$n = 1,000,000$$

$$\epsilon = 1$$

$$\beta = 0.05$$

$$\frac{1 + e^\epsilon}{e^\epsilon - 1} \approx 2.16$$

$$\frac{1}{1 + e^\epsilon} \approx 0.26$$

Example revisited

Accuracy Theorem:

$$\Pr_{r \leftarrow RR(D, q, \epsilon)} \left[\left| \frac{1 + e^\epsilon}{e^\epsilon - 1} \left(r - \frac{1}{1 + e^\epsilon} \right) - q(D) \right| \geq \frac{1 + e^\epsilon}{(e^\epsilon - 1)} \sqrt{\frac{\log(2/\beta)}{2n}} \right] \leq \beta$$

With this set of parameters we have with 95% confidence

$$\left| 2.16 \left(r - 0.26 \right) - q(D) \right| \leq 2.16 \frac{0.89}{1000}$$

So, we have:

$$2.16r - 0.5591 \leq q(D) \leq 2.16r - 0.5619$$

Randomized Response

Summarizing:

- Randomized Response is a first simple example which is very useful in practice,
- It protect privacy at the local level - we will come back to this later in the class,
- Provides a theoretical accuracy guarantee that is of the same order as the sampling error.