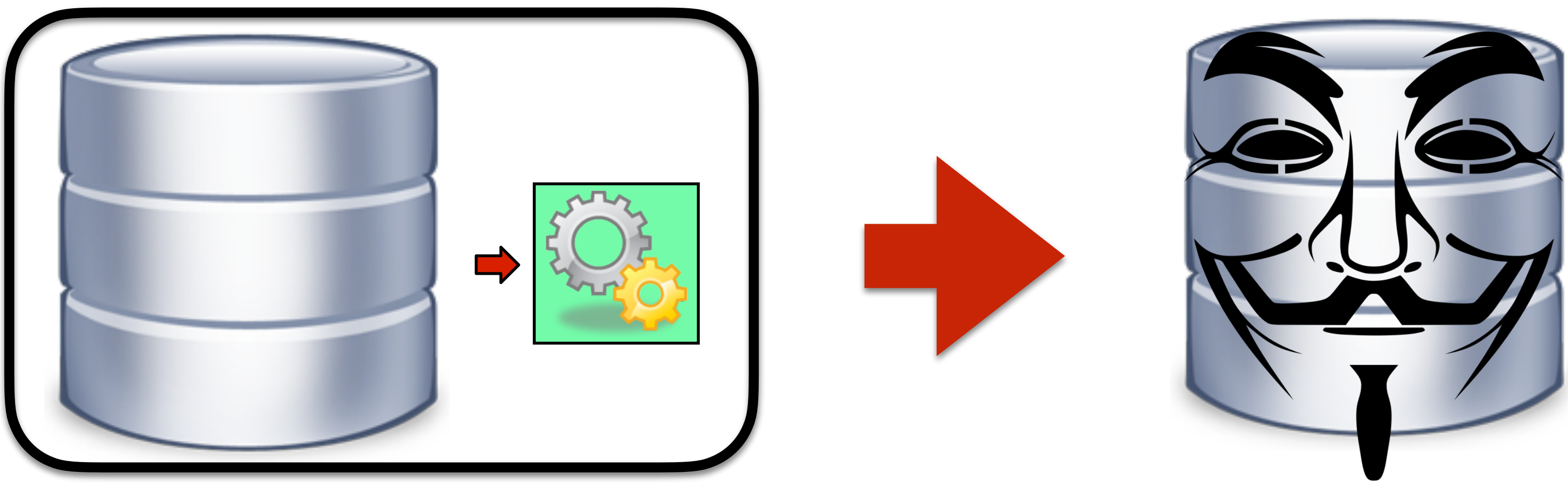
Differential Privacy

Basic mechanisms and some proofs.

Marco Gaboardi
Boston University

The opinions expressed in this course are mine and they do not not reflect those of the National Science Foundation or the US. Census Bureau.

A natural idea: anonymizing the data



- E.g. stripping PII, guaranteeing k-anonymity, swapping, etc.

Reconstruction attack with exponential adversary

Let $M:\{0,1\}^n \rightarrow R$ be a privacy mechanism adding noise within E perturbation. Then there is an adversary that can reconstruct the database within 4E positions.

[DinurNissim'02]

Attack

Query phase: For each $S \subseteq [n]$ let $a_S^* = q_S^*(D)$.

Rule out phase: For each $D' \in \{0,1\}^n$:
if there exists S such that $|q_S(D') - a_S^*| > E$ then rule out D'.

Output phase: Output a database D' that was not ruled out.

(ε,δ)-Differential Privacy

Definition

Given $\varepsilon,\delta \geq 0$, a probabilistic query $Q: X^n \to R$ is (ε,δ)-differentially private iff for all adjacent database b_1, b_2 and for every $S \subseteq R$:

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

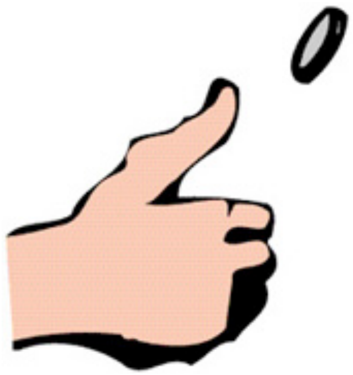
How can we build differentially private data analyses?

Randomized Response
[Warner65]

Suppose I ask a yes/no question.

Randomized Response
[Warner65]

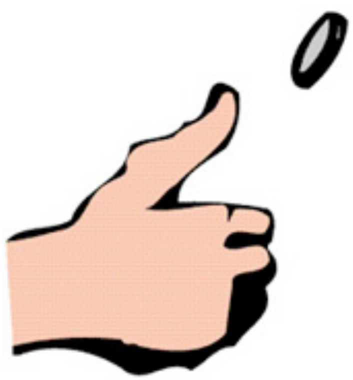
Suppose I ask a yes/no question.



biased
coin

Randomized Response
[Warner65]

Suppose I ask a yes/no question.

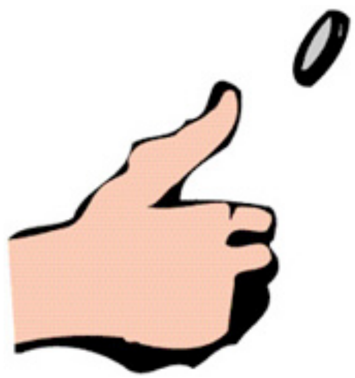


biased
coin

True
answer

Randomized Response
[Warner65]

Suppose I ask a yes/no question.



biased
coin

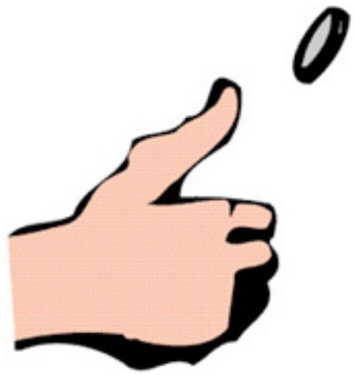
True
answer

Opposite
answer

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

Let's consider the case we have two adjacent data $\mathtt{b}$ and $\mathtt{b'}$. 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.

An example

Let's consider the case we have two adjacent data $\mathtt{b}$ and $\mathtt{b'}$. 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:

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

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

An example

Let's consider the case we have two adjacent data $\mathbf{b}$ and $\mathbf{b}'$. 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:

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

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

So for all b,b':

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

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:

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

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

So for all b,b':

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

Counting Queries

- A counting query $q : \mathcal{X}^n \to [0,1]$ is a function counting the fraction of people in a dataset satisfying the predicate $q : \mathcal{X} \to \{0,1\}$
- In symbols:

$$q(D) = \frac{1}{n} \sum_{i=1}^{n} q(d_i)$$

- Notice that we take a normalized count, which also corresponds to the average.

Randomized Response

Algorithm 1 Pseudo-code for Randomized Response

1: **function** RandomizedResponse(D, q, ϵ)

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** $\frac{(\textbf{sum } S)}{|D|}$

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.

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

A CS Example

Google Chrome RAPPOR project used 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 used 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 eps-differentially private.

Randomized Response

Privacy Theorem:
Randomized response is eps-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 $S \in \{0,1\}^n$.

Randomized Response

Privacy Theorem:
Randomized response is eps-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 eps-differentially private.

Continued Proof:
For every j≠k we have:

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

Randomized Response

Privacy Theorem:
Randomized response is eps-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 eps-differentially private.

Continued Proof:
In the case 1 we have

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

Randomized Response

Privacy Theorem:
Randomized response is eps-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 eps-differentially private.

Continued Proof:
In this case we have for example

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

Randomized Response

Privacy Theorem:
Randomized response is eps-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 eps-differentially private.

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

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

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

Randomized Response

Privacy Theorem:
Randomized response is eps-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

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

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

Accuracy Statements

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

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

Here X is a random variable representing some measure on the differentially private output.

Accuracy Statements

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

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

alpha 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

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

beta is the probability of failure

Accuracy Statements

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

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

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:

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

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.

Randomized Response

Accuracy Theorem:

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

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

Randomized Response

Accuracy Theorem:

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

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.

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 and then apply some form of Chebyshev inequality.

Additive Chernoff Bound

Theorem 1.2 (Additive Chernoff Bound)**.** Let $X_1, \ldots, X_n$ 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:

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

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
ε=1
β=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

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

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

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

So, we have:

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

Noise on Input vs Noise on Output

Input → Local Computation → Global Computation

$$q(d_1)$$
$$\vdots$$
$$q(d_n)$$

$$\frac{1}{n}\sum_{i=0}^{n} q(d_i)$$

Noise on Input vs Noise on Output

$$q(d_1)$$
$$\vdots$$
$$q(d_n)$$

$$\frac{1}{n}\sum_{i=0}^{n} q(d_i)$$

Noise on Input vs Noise on Output

$$q(d_1)$$
$$\vdots$$
$$q(d_n)$$

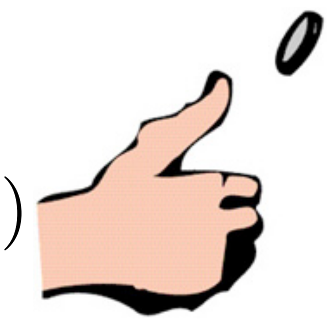
$$\frac{1}{n}\sum_{i=0}^{n} q(d_i)$$

Noise on Input vs Noise on Output

Input → Local Computation → Global Computation

$$q(d_1)$$
$$\vdots$$
$$q(d_n)$$

$$\frac{1}{n}\sum_{i=0}^{n} q(d_i)$$



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.

Noise on the output

Question: What is a good way to add noise to the output of a statistical query?

Intuitive answer: it must depend on ε or the accuracy we want to achieve, and on the scale that a change can have on the output.

Noise on the output

Question: What is a good way to add noise to the output of a statistical query?

Global Sensitivity

Definition 1.8 (Global sensitivity). The *global sensitivity* of a function $q : \mathcal{X}^n \to \mathbb{R}$ is:

$$\Delta q = \max \left\{ |q(D) - q(D')| \;\middle|\; D \sim_1 D' \in \mathcal{X}^n \right\}$$

Global Sensitivity

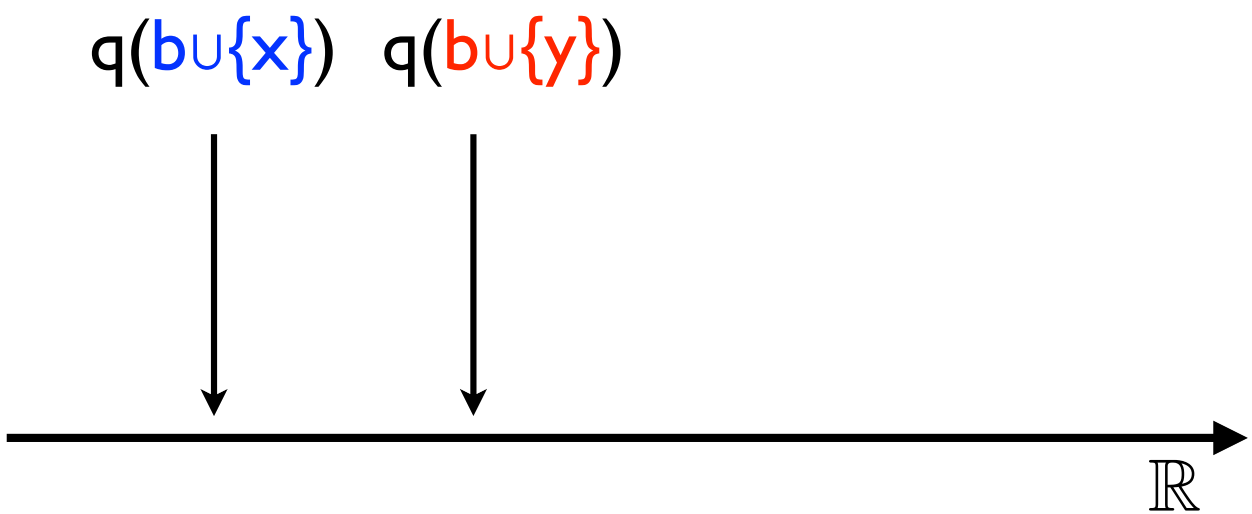
Definition 1.8 (Global sensitivity)**.** The *global sensitivity* of a function $q : \mathcal{X}^n \to \mathbb{R}$ is:

$$\Delta q = \max \left\{ |q(D) - q(D')| \;\middle|\; D \sim_1 D' \in \mathcal{X}^n \right\}$$

Global Sensitivity

Definition 1.8 (Global sensitivity). The *global sensitivity* of a function $q : \mathcal{X}^n \to \mathbb{R}$ is:

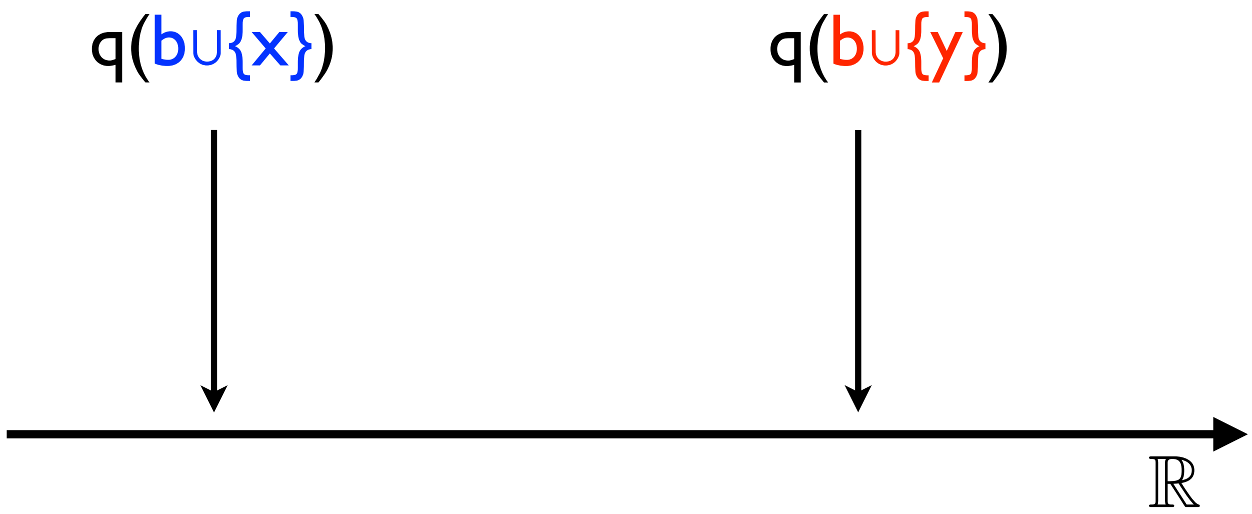
$$\Delta q = \max \left\{ |q(D) - q(D')| \;\middle|\; D \sim_1 D' \in \mathcal{X}^n \right\}$$



Global Sensitivity

Definition 1.8 (Global sensitivity). The *global sensitivity* of a function $q : \mathcal{X}^n \to \mathbb{R}$ is:

$$\Delta q = \max \left\{ |q(D) - q(D')| \;\middle|\; D \sim_1 D' \in \mathcal{X}^n \right\}$$



Laplace Mechanism

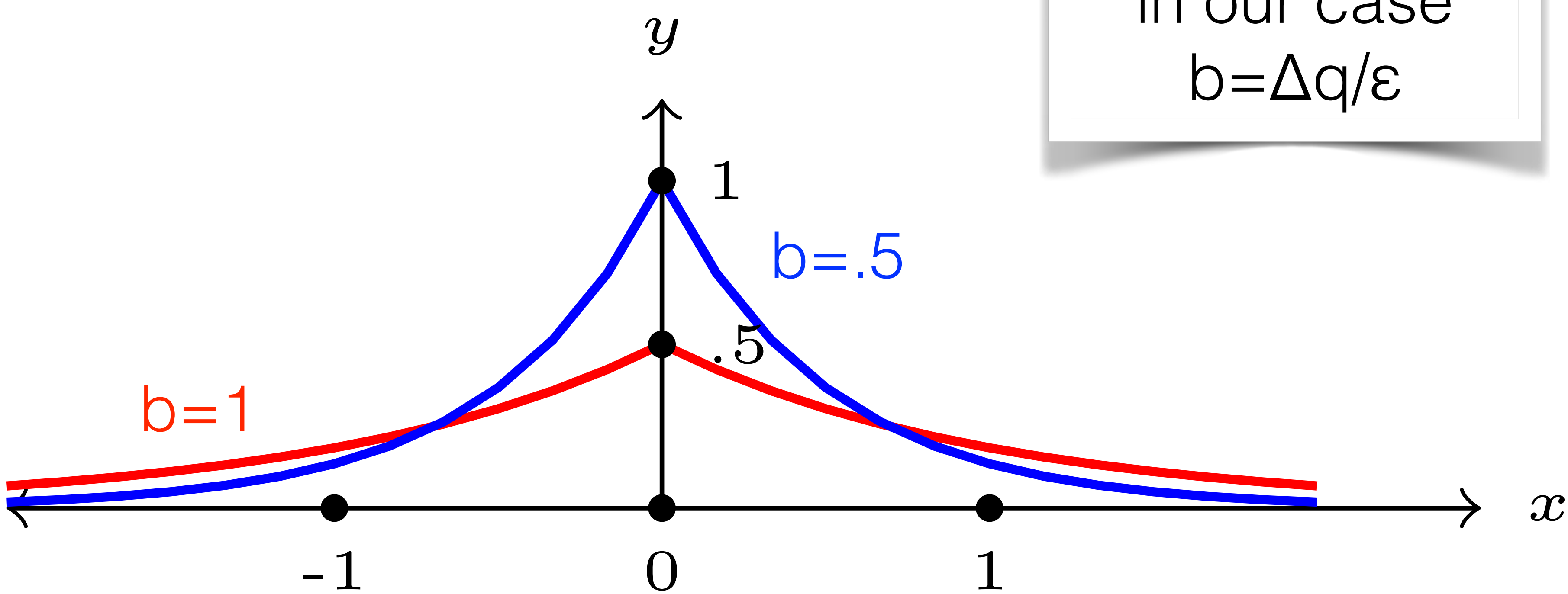
Algorithm 2 Pseudo-code for the Laplace Mechanism

function $\textsc{LapMech}(D, q, \epsilon)$
    $Y \xleftarrow{\$} \mathsf{Lap}(\frac{\Delta q}{\epsilon})(0)$
    return $q(D) + Y$
end function

Laplace Distribution

$$\text{Lap}(b, \mu)(X) = \frac{1}{2b} \exp\Big(- \frac{|\mu - X|}{b} \Big)$$

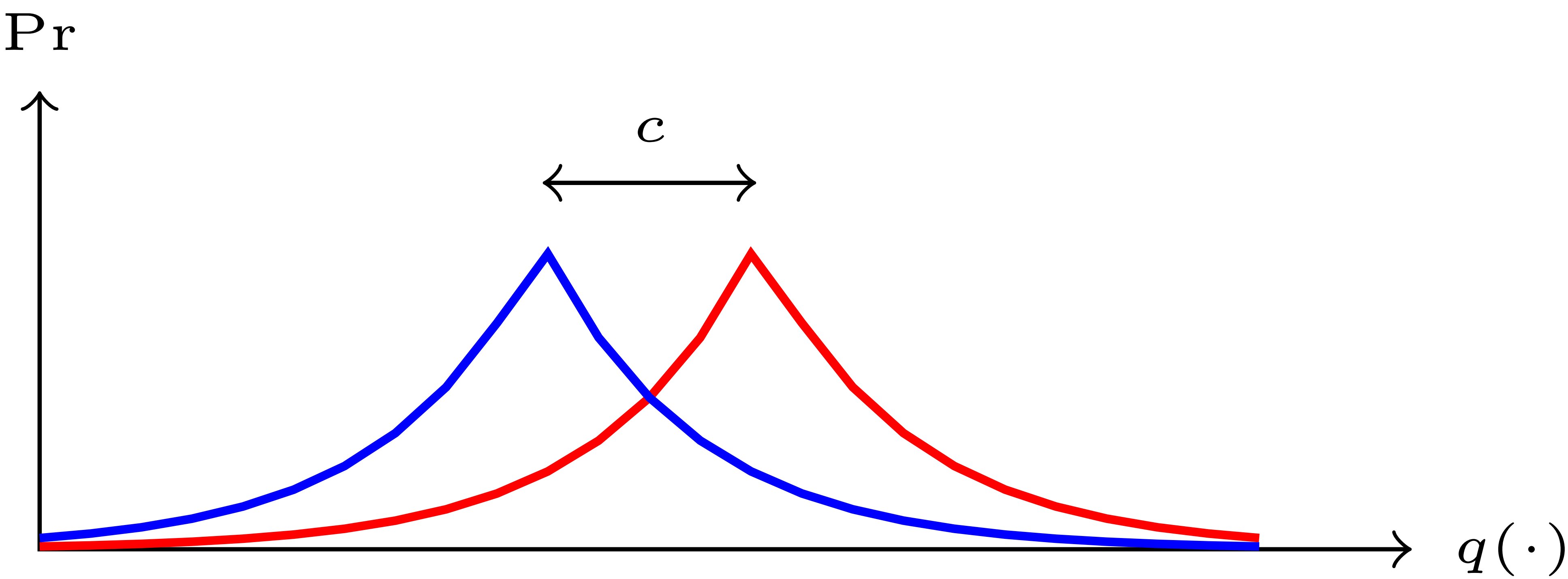
b regulates the skewness of the curve, in our case b=Δq/ε



Laplace Mechanism

Theorem (Privacy of the Laplace Mechanism)
The Laplace mechanism is ε-differentially private.

Proof: Intuitively



Laplace Mechanism

Theorem (Privacy of the Laplace Mechanism)
The Laplace mechanism is ε-differentially private.

Proof:

Consider $D \sim_1 D' \in \mathcal{X}^n$, $q : \mathcal{X}^n \to \mathbb{R}$, and let p and p' denote the probability density function of $\mathsf{LapMech}(D, q, \epsilon)$ and $\mathsf{LapMech}(D', q, \epsilon)$

We compare them at an arbitrary point $z \in \mathbb{R}$.

Laplace Mechanism

Theorem (Privacy of the Laplace Mechanism)
The Laplace mechanism is ε-differentially private.

Proof:

Consider $D \sim_1 D' \in \mathcal{X}^n$, $q : \mathcal{X}^n \to \mathbb{R}$, and let p and p' denote the probability density function of $\mathsf{LapMech}(D, q, \epsilon)$ and $\mathsf{LapMech}(D', q, \epsilon)$ We compare them at an arbitrary point $z \in \mathbb{R}$.

$$\frac{p(z)}{p'(z)} = \frac{\exp\left(-\frac{\epsilon|q(D)-z|}{\Delta q}\right)}{\exp\left(-\frac{\epsilon|q(D')-z|}{\Delta q}\right)}$$

Laplace Mechanism

Theorem (Privacy of the Laplace Mechanism)
The Laplace mechanism is ε-differentially private.

Continued proof:

$$\frac{p(z)}{p'(z)} = \frac{\exp\left(-\frac{\epsilon|q(D)-z|}{\Delta q}\right)}{\exp\left(-\frac{\epsilon|q(D')-z|}{\Delta q}\right)}$$

Laplace Mechanism

Theorem (Privacy of the Laplace Mechanism)
The Laplace mechanism is ε-differentially private.

Continued proof:

$$\frac{p(z)}{p'(z)} = \frac{\exp\left(-\frac{\epsilon|q(D)-z|}{\Delta q}\right)}{\exp\left(-\frac{\epsilon|q(D')-z|}{\Delta q}\right)}$$

$$= \exp\left(\frac{\epsilon(|q(D') - z| - |q(D) - z|)}{\Delta q}\right)$$

Laplace Mechanism

Theorem (Privacy of the Laplace Mechanism)
The Laplace mechanism is ε-differentially private.

Continued proof:

$$\frac{p(z)}{p'(z)} = \frac{\exp\left(-\frac{\epsilon|q(D)-z|}{\Delta q}\right)}{\exp\left(-\frac{\epsilon|q(D')-z|}{\Delta q}\right)}$$

$$= \exp\left(\frac{\epsilon(|q(D') - z| - |q(D) - z|)}{\Delta q}\right)$$

$$\leq \exp\left(\frac{\epsilon(|q(D') - q(D)|)}{\Delta q}\right)$$

Laplace Mechanism

Theorem (Privacy of the Laplace Mechanism)
The Laplace mechanism is ε-differentially private.

Continued proof:

$$\frac{p(z)}{p'(z)} = \frac{\exp\left(-\frac{\epsilon|q(D)-z|}{\Delta q}\right)}{\exp\left(-\frac{\epsilon|q(D')-z|}{\Delta q}\right)}$$

$$= \exp\left(\frac{\epsilon(|q(D') - z| - |q(D) - z|)}{\Delta q}\right)$$

$$\leq \exp\left(\frac{\epsilon(|q(D') - q(D)|)}{\Delta q}\right)$$

$$\leq \exp(\epsilon)$$

Laplace Mechanism

Theorem (Privacy of the Laplace Mechanism)
The Laplace mechanism is ε-differentially private.

Continued proof:

$$\frac{p(z)}{p'(z)} = \frac{\exp\left(-\frac{\epsilon|q(D)-z|}{\Delta q}\right)}{\exp\left(-\frac{\epsilon|q(D')-z|}{\Delta q}\right)}$$

$$= \exp\left(\frac{\epsilon(|q(D')-z| - |q(D)-z|)}{\Delta q}\right)$$

$$\leq \exp\left(\frac{\epsilon(|q(D')-q(D)|)}{\Delta q}\right)$$

$$\leq \exp(\epsilon)$$

Similarly, we can prove that $\exp(-\epsilon) \leq \frac{p(z)}{p'(z)}$

Example Revisited

Let’s consider again a medical dataset containing informations on whether each patient has a disease. We can have d_i=0 if patient i has the disease and d_i=1 otherwise.

Example Revisited

Let's consider again a medical dataset containing informations on whether each patient has a disease. We can have d_i=0 if patient i has the disease and d_i=1 otherwise.

We first compute the proportion of patients that have the disease. Notice that this has sensitivity 1/n.

Example Revisited

Let's consider again a medical dataset containing informations on whether each patient has a disease. We can have $d_i=0$ if patient i has the disease and $d_i=1$ otherwise.

We first compute the proportion of patients that have the disease. Notice that this has sensitivity 1/n.

Then we can add Laplace noise proportional to $1/\varepsilon n$.

Example Revisited

Let's consider again a medical dataset containing informations on whether each patient has a disease. We can have d_i=0 if patient i has the disease and d_i=1 otherwise.

We first compute the proportion of patients that have the disease. Notice that this has sensitivity 1/n.

Then we can add Laplace noise proportional to 1/εn.

The noise protect each individual value.

Laplace Mechanism

Question: How accurate is the answer that we get from the Laplace Mechanism?

Laplace Mechanism

Accuracy Theorem: $\text{let } r = \mathsf{LapMech}(D, q, \epsilon)$

$$\Pr\left[|q(D) - r| \geq \left(\frac{\Delta q}{\epsilon}\right) \ln\left(\frac{1}{\beta}\right)\right] = \beta$$

Laplace Mechanism

Accuracy Theorem: let $r = \mathsf{LapMech}(D, q, \epsilon)$

$$\Pr\left[|q(D) - r| \geq \left(\frac{\Delta q}{\epsilon}\right) \ln\left(\frac{1}{\beta}\right)\right] = \beta$$

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

Laplace Mechanism

Accuracy Theorem: let $r = \mathsf{LapMech}(D, q, \epsilon)$

$$\Pr\left[|q(D) - r| \geq \left(\frac{\Delta q}{\epsilon}\right) \ln\left(\frac{1}{\beta}\right)\right] = \beta$$

Laplace Mechanism

Accuracy Theorem: let $r = \mathsf{LapMech}(D, q, \epsilon)$

$$\Pr\left[|q(D) - r| \geq \left(\frac{\Delta q}{\epsilon}\right) \ln\left(\frac{1}{\beta}\right)\right] = \beta$$

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

Laplace Mechanism

Accuracy Theorem: let $r = \mathsf{LapMech}(D, q, \epsilon)$

$$\Pr\left[|q(D) - r| \geq \left(\frac{\Delta q}{\epsilon}\right) \ln\left(\frac{1}{\beta}\right)\right] = \beta$$

Laplace Mechanism

Accuracy Theorem: let $r = \mathsf{LapMech}(D, q, \epsilon)$

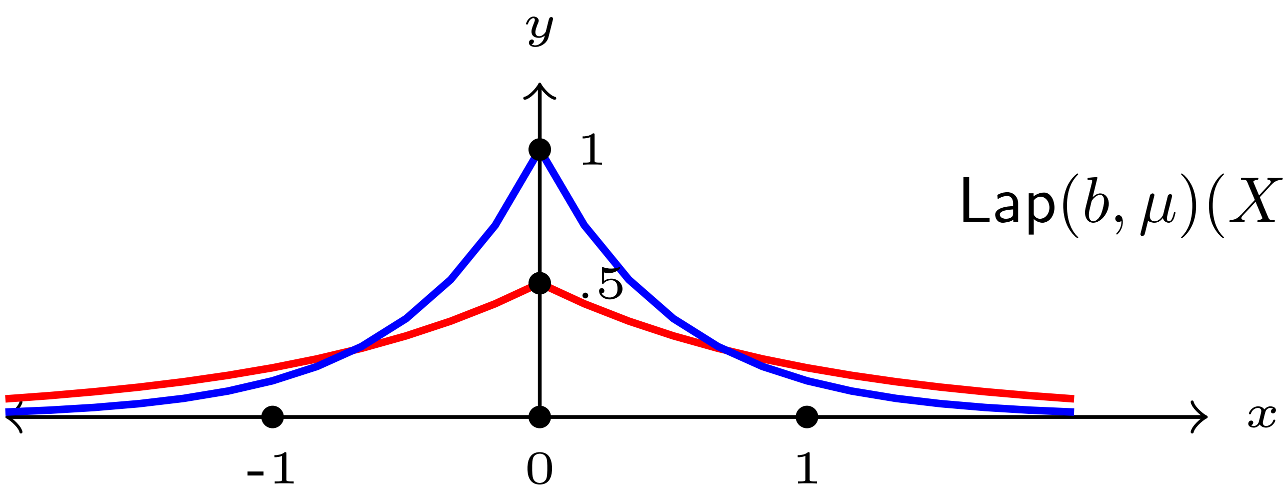
$$\Pr\left[|q(D) - r| \geq \left(\frac{\Delta q}{\epsilon}\right)\ln\left(\frac{1}{\beta}\right)\right] = \beta$$

Proof: By definition of the Laplace mechanism we have:

$$\Pr\left[|q(D) - r| \geq \left(\frac{\Delta q}{\epsilon}\right)\ln\left(\frac{1}{\beta}\right)\right] = \Pr\left[|Y| \geq \left(\frac{\Delta q}{\epsilon}\right)\ln\left(\frac{1}{\beta}\right)\right]$$

where Y is drawn from Lap(0,Δq/ε)

Tail bound for the Laplace Distribution



$$\mathsf{Lap}(b,\mu)(X) = \frac{1}{2b}\exp\left(-\frac{|\mu - X|}{b}\right)$$

$$\Pr\Big[|X| \geq b\,t\Big] = \exp(-t)$$

Laplace Mechanism

Accuracy Theorem: let $r = \mathsf{LapMech}(D, q, \epsilon)$

$$\Pr\left[|q(D) - r| \geq \left(\frac{\Delta q}{\epsilon}\right) \ln\left(\frac{1}{\beta}\right)\right] = \beta$$

Continued proof: applying this bound we get:

$$\Pr\left[|Y| \geq \left(\frac{\Delta q}{\epsilon}\right) \ln\left(\frac{1}{\beta}\right)\right] = \exp\left(-\ln\left(\frac{1}{\beta}\right)\right) = \beta$$

Laplace Mechanism

Accuracy Theorem: let $r = \mathsf{LapMech}(D, q, \epsilon)$

$$\Pr\left[|q(D) - r| \geq \left(\frac{\Delta q}{\epsilon}\right) \ln\left(\frac{1}{\beta}\right)\right] = \beta$$

Intuitive reading: with high probability we have:

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

Example revisited

Accuracy Theorem: let $r = \mathsf{LapMech}(D, q, \epsilon)$

$$\Pr\left[|q(D) - r| \geq \left(\frac{\Delta q}{\epsilon}\right) \ln\left(\frac{1}{\beta}\right)\right] = \beta$$

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 the Laplace Mechanism to estimate the proportion of patient that have the disease. We need to fix the parameters.

Example revisited

Accuracy Theorem: let $r = \mathsf{LapMech}(D, q, \epsilon)$

$$\Pr\left[|q(D) - r| \geq \left(\frac{\Delta q}{\epsilon}\right) \ln\left(\frac{1}{\beta}\right)\right] = \beta$$

Let's fix the following values for the parameters:

n=1,000,000
ε=1
β=0.05
$\ln(\frac{1}{\beta}) = 2.99$

Example revisited

Accuracy Theorem: let $r = \mathsf{LapMech}(D, q, \epsilon)$

$$\Pr\left[|q(D) - r| \geq \left(\frac{\Delta q}{\epsilon}\right) \ln\left(\frac{1}{\beta}\right)\right] = \beta$$

Let's fix the following values for the parameters:

n=1,000,000
ε=1
β=0.05
$\ln(\frac{1}{\beta}) = 2.99$

Question: What is the sensitivity?

Example revisited

Accuracy Theorem: let $r = \text{LapMech}(D, q, \epsilon)$

$$\Pr\left[|q(D) - r| \geq \left(\frac{\Delta q}{\epsilon}\right) \ln\left(\frac{1}{\beta}\right)\right] = \beta$$

Let's fix the following values for the parameters:

n=1,000,000
ε=1
β=0.05
$\ln(\frac{1}{\beta}) = 2.99$

Question: What is the sensitivity?

$$\Delta q = 10^{-6}$$

Example revisited

Accuracy Theorem: let $r = \mathsf{LapMech}(D, q, \epsilon)$

$$\Pr\left[|q(D) - r| \geq \left(\frac{\Delta q}{\epsilon}\right) \ln\left(\frac{1}{\beta}\right)\right] = \beta$$

With this set of parameters we have with 95% confidence

$$r - 0.0000299 \leq q(D) \leq r + 0.0000299.$$

Randomized Response vs Laplace

Accuracy for Randomize response: with high probability we have:

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

Accuracy for Laplace: with high probability we have:

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

Summary

- Definition of Differential Privacy
- Randomized Response
- Laplace Mechanism

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 $X_1, \ldots, X_n$ 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

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

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

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

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

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