Differential privacy – Basic notions and methods

Colin Rothgang

May 16, 2017

Contents

The idea of differential privacy

The idea of differential privacy

- The key idea behind differential privacy is to obfuscate one individual 's properties, but not the whole groups properties in a database

The idea of differential privacy

- The key idea behind differential privacy is to obfuscate one individual 's properties, but not the whole groups properties in a database
- So the probability for any individual in the database to have a property should barely differ from the base rate

The idea of differential privacy

- The key idea behind differential privacy is to obfuscate one individual 's properties, but not the whole groups properties in a database
- So the probability for any individual in the database to have a property should barely differ from the base rate
- Then, analyzing the database an attacker can't reliably learn anything new about any individual in the database, no matter how much additional information he has

Examples; Without differential privacy. . .

Examples; Without differential privacy...

- In 2007 Netflix offered a 1 million$ prize to improve its recommdation system and published a “anonymized” training dataset

Examples; Without differential privacy. . .

- In 2007 Netflix offered a 1 million$ prize to improve its recommdation system and published a “anonymized” training dataset
  - $\hookrightarrow$ Later that database was linked with the internet movie database IMdB, allowing identification of some users

Examples; Without differential privacy...

- In 2007 Netflix offered a 1 million$ prize to improve its recommdation system and published a "anonymized" training dataset
  - $\hookrightarrow$ Later that database was linked with the internet movie database IMdB, allowing identification of some users
- Latanya Sweeney from Carnegie Mellon University linked the anonymized Massachusetts Group Insurance Commission (GIC) medical encounter database with voter's registration records identifying the medical records of the Governor of Massachusetts.

We need surveys and databases, but also privacy

We need surveys and databases, but also privacy

- We want to be able to still use surveys and statistical studies, without compromising the privacy of our subjects.

We need surveys and databases, but also privacy

- We want to be able to still use surveys and statistical studies, without compromising the privacy of our subjects.
  - $\hookrightarrow$ A standard example are medical records, having an obvious use. However, many people want their medical data to be safe.

We need surveys and databases, but also privacy

- We want to be able to still use surveys and statistical studies, without compromising the privacy of our subjects.
  - $\hookrightarrow$ A standard example are medical records, having an obvious use. However, many people want their medical data to be safe.
- If we don't give people a proof of their privacy, they might not submit the surveys or lie

We need surveys and databases, but also privacy

- We want to be able to still use surveys and statistical studies, without compromising the privacy of our subjects.
  - $\hookrightarrow$ A standard example are medical records, having an obvious use. However, many people want their medical data to be safe.
- If we don't give people a proof of their privacy, they might not submit the surveys or lie
  - $\hookrightarrow$ This destroys the reliability of the obtained results.

The query-release problem

The query-release problem

- A very useful formalization of our problem

The query-release problem

- A very useful formalization of our problem
- Given a database, we want to run a query on

The query-release problem

- A very useful formalization of our problem
- Given a database, we want to run a query on
- Want this differentially private as follows:

The query-release problem

- A very useful formalization of our problem
- Given a database, we want to run a query on
- Want this differentially private as follows:
  - Modify database, such that its release is differentially private

The query-release problem

- A very useful formalization of our problem
- Given a database, we want to run a query on
- Want this differentially private as follows:
    - Modify database, such that its release is differentially private
    - Now, we can run the query on it and publish the result

The query-release problem

- A very useful formalization of our problem
- Given a database, we want to run a query on
- Want this differentially private as follows:
  - Modify database, such that its release is differentially private
  - Now, we can run the query on it and publish the result

Caveat: Depending on the query, the result of the query after the modification of the database, might not be very useful

What do we need, to feel safe submitting a survey?

What do we need, to feel safe submitting a survey?

- When do we feel safe submitting a survey?

What do we need, to feel safe submitting a survey?

- When do we feel safe submitting a survey?
- If attacker looking at result of survey cannot learn anything new about us (individually).

What do we need, to feel safe submitting a survey?

- When do we feel safe submitting a survey?
- If attacker looking at result of survey cannot learn anything new about us (individually).
- However this is not possible:

What do we need, to feel safe submitting a survey?

- When do we feel safe submitting a survey?
- If attacker looking at result of survey cannot learn anything new about us (individually).
- However this is not possible:

  $\hookrightarrow$ If the survey shows that any human a certain property, then we also have that property (example: smoking Mary)

What do we need, to feel safe submitting a survey?

- When do we feel safe submitting a survey?
- If attacker looking at result of survey cannot learn anything new about us (individually).
- However this is not possible:
  - $\hookrightarrow$ If the survey shows that any human a certain property, then we also have that property (example: smoking Mary)
  - This holds even if we don't submit the survey at all!

What do we need, to feel safe submitting a survey?

- When do we feel safe submitting a survey?
- If attacker looking at result of survey cannot learn anything new about us (individually).
- However this is not possible:
  - ↪ If the survey shows that any human a certain property, then we also have that property (example: smoking Mary)
  - This holds even if we don't submit the survey at all!

↪ Can we at ensure that submitting the survey does not cause significant additional harm?

What do we need, to feel safe submitting a survey?

- When do we feel safe submitting a survey?
- If attacker looking at result of survey cannot learn anything new about us (individually).
- However this is not possible:
  - ↪ If the survey shows that any human a certain property, then we also have that property (example: smoking Mary)
  - This holds even if we don't submit the survey at all!

↪ Can we at ensure that submitting the survey does not cause significant additional harm?

- *This is the key idea of differential privacy*

Definition of ϵ-δ-differentially privacy

Definition of ϵ-δ-differentially privacy

- In 2006 Cynthia Dwork proposed the following definition:

Definition of ϵ-δ-differentially privacy

- In 2006 Cynthia Dwork proposed the following definition:
- Assume a database D consisting of n Vectors of m-components over some set $\mathcal{F}$ represented as a $m \times n$ matrix over $\mathcal{F}$.

Definition of ϵ-δ-differentially privacy

- In 2006 Cynthia Dwork proposed the following definition:
- Assume a database D consisting of n Vectors of m-components over some set $\mathcal{F}$ represented as a $m \times n$ matrix over $\mathcal{F}$.
- Define $\mathrm{dist}(D, D_2) := |\{i \in \{1, 2, \ldots, m\} \colon D_i \neq D_i\}|$ $\forall D,\ D_2 \in (\mathcal{F}^m)^n$ as the number of entries in which the databases D and D_2 differ.

Definition of ϵ-δ-differentially privacy

- In 2006 Cynthia Dwork proposed the following definition:
- Assume a database D consisting of n Vectors of m-components over some set $\mathcal{F}$ represented as a $m \times n$ matrix over $\mathcal{F}$.
- Define $\mathrm{dist}(D, D_2) := |\{i \in \{1, 2, \ldots, m\} \colon D_i \neq D_i\}|$ $\forall D,\ D_2 \in (\mathcal{F}^m)^n$ as the number of entries in which the databases D and D_2 differ.
- Let $\mathcal{A}$ be an algorithm processing D and $\mathrm{Range}(\mathcal{A})$ its image.

Definition of ϵ-δ-differentially privacy

- In 2006 Cynthia Dwork proposed the following definition:
- Assume a database D consisting of n Vectors of m-components over some set $\mathcal{F}$ represented as a $m \times n$ matrix over $\mathcal{F}$.
- Define $\mathrm{dist}(D, D_2) := |\{i \in \{1, 2, \ldots, m\} \colon D_i \neq D_i\}|$ $\forall D,\ D_2 \in (\mathcal{F}^m)^n$ as the number of entries in which the databases D and D_2 differ.
- Let $\mathcal{A}$ be an algorithm processing D and $\mathrm{Range}(\mathcal{A})$ its image.

Definition (ϵ-δ-differential privacy)

Now $\mathcal{A}$ is called ϵ-δ-differentially private if $\forall \mathcal{S} \subset \mathrm{Range}(\mathcal{A})$:

$$\forall D_2 \colon \mathrm{dist}(D, D_2) \leq 1 \Rightarrow \Pr\left[\mathcal{A}(D) \in \mathcal{S}\right] \leq e^{\epsilon} \cdot \Pr\left[\mathcal{A}(D_2) \in \mathcal{S}\right] + \delta$$

Intuitive meaning of ϵ-δ-differential privacy

What does ϵ-*delta*-differential privacy tell us?

Intuitive meaning of ϵ-δ-differential privacy

What does ϵ-*delta*-differential privacy tell us?

- Firstly, let us assume that $\delta = 0$.

Intuitive meaning of ϵ-δ-differential privacy

What does ϵ-*delta*-differential privacy tell us?

- Firstly, let us assume that $\delta = 0$.
  - $\hookrightarrow$ This is also called ϵ-differential privacy

Intuitive meaning of ϵ-δ-differential privacy

What does ϵ-*delta*-differential privacy tell us?

- Firstly, let us assume that $\delta = 0$.
  - $\hookrightarrow$ This is also called ϵ-differential privacy
  - $\hookrightarrow$ Then, given the result of the survey, an attacker cannot learn any new property about us with a significant probability

Randomized response

A very old technique developed by social scientists to collect data about embarrassing or incriminating behaviour.

Randomized response

A very old technique developed by social scientists to collect data about embarrassing or incriminating behaviour.
Participants are told to report as follows, whether or not they have property P

Randomized response

A very old technique developed by social scientists to collect data about embarrassing or incriminating behaviour.
Participants are told to report as follows, whether or not they have property P

- Flip a coin

Randomized response

A very old technique developed by social scientists to collect data about embarrassing or incriminating behaviour.
Participants are told to report as follows, whether or not they have property P

- Flip a coin
- If result is tails, report truthfully

Randomized response

A very old technique developed by social scientists to collect data about embarrassing or incriminating behaviour.
Participants are told to report as follows, whether or not they have property P

- Flip a coin
- If result is tails, report truthfully
- If result is heads, flip again and respond "yes" iff result is heads otherwise "no"

This way, participants are guaranteed plausible deniability,

Randomized response

A very old technique developed by social scientists to collect data about embarrassing or incriminating behaviour.
Participants are told to report as follows, whether or not they have property P

- Flip a coin
- If result is tails, report truthfully
- If result is heads, flip again and respond "yes" iff result is heads otherwise "no"

This way, participants are guaranteed plausible deniability,
Even if participant has property P and reports it, this is not incriminating.

Analysis of randomized responce surveys

Analysis of randomized responce surveys

- Assume participant X reports having property P, what can we learn about X?

Analysis of randomized responce surveys

- Assume participant X reports having property P, what can we learn about X?
- We don't really know whether X has the property

Analysis of randomized responce surveys

- Assume participant X reports having property P, what can we learn about X?
- We don't really know whether X has the property
  - $\hookrightarrow$ The probability, that a participant having property P will answer "yes" is only $\frac{1}{2} + \frac{1}{4} = \frac{3}{4}$, whereas the probability that participant not having property P answers "yes" is $\frac{1}{4}$

Analysis of randomized responce surveys

- Assume participant X reports having property P, what can we learn about X?
- We don't really know whether X has the property
  - $\hookrightarrow$ The probability, that a participant having property P will answer "yes" is only $\frac{1}{2} + \frac{1}{4} = \frac{3}{4}$, whereas the probability that participant not having property P answers "yes" is $\frac{1}{4}$
  - $\hookrightarrow$ Therefore, the probability to identify participant based on reply is 3-times as big as in the case, where question is not asked.

Analysis of randomized responce surveys

- Assume participant X reports having property P, what can we learn about X?
- We don't really know whether X has the property
  - $\hookrightarrow$ The probability, that a participant having property P will answer "yes" is only $\frac{1}{2} + \frac{1}{4} = \frac{3}{4}$, whereas the probability that participant not having property P answers "yes" is $\frac{1}{4}$
  - $\hookrightarrow$ Therefore, the probability to identify participant based on reply is 3-times as big as in the case, where question is not asked.
  - $\hookrightarrow$ Hence, this method is ln(3)-differentially private

Analysis of randomized responce surveys

- Assume participant X reports having property P, what can we learn about X?
- We don't really know whether X has the property
  - $\hookrightarrow$ The probability, that a participant having property P will answer "yes" is only $\frac{1}{2} + \frac{1}{4} = \frac{3}{4}$, whereas the probability that participant not having property P answers "yes" is $\frac{1}{4}$
  - $\hookrightarrow$ Therefore, the probability to identify participant based on reply is 3-times as big as in the case, where question is not asked.
  - $\hookrightarrow$ Hence, this method is $\ln(3)$-differentially private
  - $\hookrightarrow$ Since, the ϵ's for different sub-surveys add up, a survey of m such questions is $m \cdot \ln(3)$-differentially private

The l_1-sensitivity

The l_1-sensitivity

Definition (l_1-norm)

Define the l_1-norm $\|\cdot\|_1 : \mathbb{R}^p \to \mathbb{R}$ *by* $\|v\|_1 := \sum_{i=1}^{p} |v_i|$.

The l_1-sensitivity

Definition (l_1-norm)

Define the l_1-norm $\|\cdot\|_1 : \mathbb{R}^p \to \mathbb{R}$ by $\|v\|_1 := \sum_{i=1}^{p} |v_i|$.

- Let $A : (\mathbb{N}^m)^n \to \mathbb{R}^k$ be some numeric database query

The l_1-sensitivity

Definition (l_1-norm)

Define the l_1-norm $\|\cdot\|_1 : \mathbb{R}^p \to \mathbb{R}$ by $\|v\|_1 := \sum_{i=1}^p |v_i|$.

- Let $A : (\mathbb{N}^m)^n \to \mathbb{R}^k$ be some numeric database query

Definition (l_1-sensitivity)

Define the l_1-sensitivity $\triangle A$ of A as

The l_1-sensitivity

Definition (l_1-norm)

Define the l_1-norm $\|\cdot\|_1 : \mathbb{R}^p \to \mathbb{R}$ by $\|v\|_1 := \sum_{i=1}^{p} |v_i|$.

- Let $A : (\mathbb{N}^m)^n \to \mathbb{R}^k$ be some numeric database query

Definition (l_1-sensitivity)

Define the l_1-sensitivity $\triangle A$ of A as

$$\triangle A := \max_{X,Y \in (\mathbb{N}^m)^n,\ \|X-Y\|_1=1} \|A(X) - A(Y)\|.$$

The l_1-sensitivity

Definition (l_1-norm)

Define the l_1-norm $\|\cdot\|_1 : \mathbb{R}^p \to \mathbb{R}$ by $\|v\|_1 := \sum_{i=1}^{p} |v_i|$.

- Let $A : (\mathbb{N}^m)^n \to \mathbb{R}^k$ be some numeric database query

Definition (l_1-sensitivity)

Define the l_1-sensitivity $\triangle A$ of A as

$$\triangle A := \max_{X,Y \in (\mathbb{N}^m)^n,\ \|X-Y\|_1 = 1} \|A(X) - A(Y)\| .$$

- The l_1-sensitivity intuitively tells us how much a single individual's data can affect the result of our query.

The l_1-sensitivity

Definition (l_1-norm)

Define the l_1-norm $\|\cdot\|_1 : \mathbb{R}^p \to \mathbb{R}$ by $\|v\|_1 := \sum_{i=1}^p |v_i|$.

- Let $A : (\mathbb{N}^m)^n \to \mathbb{R}^k$ be some numeric database query

Definition (l_1-sensitivity)

Define the l_1-sensitivity $\triangle A$ of A as

$$\triangle A := \max_{X,Y \in (\mathbb{N}^m)^n,\ \|X-Y\|_1 = 1} \|A(X) - A(Y)\| .$$

- The l_1-sensitivity intuitively tells us how much a single individual's data can affect the result of our query.
  - $\hookrightarrow$ This, gives upper bound, for amount of randomness we need to add to gain differential privacy

The Laplace-Distribution

The Laplace-Distribution

Definition

The probability density function of the Laplace-Distribution is defined as the function

$$Lap(x|b) := \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right).$$

The Laplace-Distribution

Definition

The probability density function of the Laplace-Distribution is defined as the function

$$Lap(x|b) := \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right).$$

Definition (Laplace-Distribution)

The Laplace-Distribution (centered at 0) and with scale b, is the distribution corresponding to $Lap(x|b)$*.*

The Laplace-Distribution

Definition

The probability density function of the Laplace-Distribution is defined as the function

$$Lap(x|b) := \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right).$$

Definition (Laplace-Distribution)

The Laplace-Distribution (centered at 0) and with scale b, is the distribution corresponding to $Lap(x|b)$.

Remark

We could also use the Gaussian-Distribution instead, but the Laplace-Distribution is a bit handier.

The Laplace mechanism

The Laplace mechanism

Let $A : (\mathbb{N}^m)^n \to \mathbb{R}^k$ be any numeric database query.

The Laplace mechanism

Let $A : (\mathbb{N}^m)^n \to \mathbb{R}^k$ be any numeric database query.

Definition (Laplace mechanism)

The Laplace mechanism $\mathcal{M}_{L,f,\epsilon}(x)$ for f and a given ϵ is defined as:

$$\mathcal{M}_{L,f,\epsilon}(x) := f(x) + (\mathcal{Y}_1, \mathcal{Y}_2, \ldots, \mathcal{Y}_k),$$

where the $\mathcal{Y}_j$ are random variables drawn from the Laplace-Distribution $Lap(\frac{\triangle f}{\epsilon})$.

The Laplace mechanism is ϵ-differentially private

The Laplace mechanism is ϵ-differentially private

Theorem

The Laplace mechanism is ϵ-differentially private.

The Laplace mechanism is ϵ-differentially private

Theorem

The Laplace mechanism is ϵ-differentially private.

Proving this theorem is beyond the scope of this talk.

References

- Cynthia Dwork and Aaron Roth, *The Algorithmic Foundation of Differential Privacy* https://www.cis.upenn.edu/~aaroth/Papers/privacybook.pdf, date = 2004,
- Wang Yuxiang *Differential Privacy: a short tutorial*, https://www.cs.cmu.edu/~yuxiangw/docs/Differential%20Privacy.pdf, 2012

Thank you for your attention