How to learn while respecting individual privacy?

An Introduction to Differential Privacy

Aurélien Garivier

École Normale Supérieure de Lyon, UMPA & LIP

November 15^{th}, 2021

UNIVERSITÉ DE LYON

ENS DE LYON

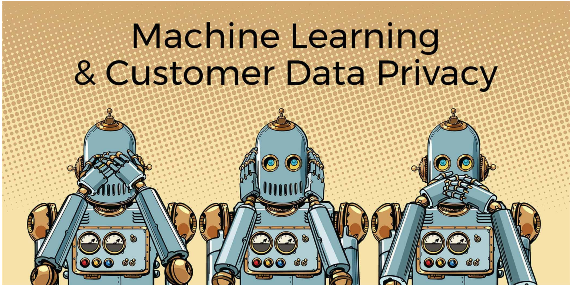
Outline

Machine Learning

Modeling Privacy

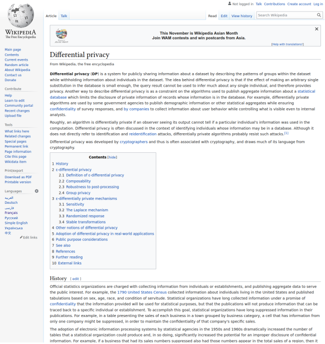
Exponential Mechanism

Laplace Mechanism



Src: https://www.actian.com/company/blog/laymans-guide-to-machine-learning-and-customer-data-privacy/

References



Not logged in Talk Contributions Create account Log in

Article Talk Read Edit View history Search Wikipedia

This November is Wikipedia Asian Month
Join WAM contests and win postcards from Asia.
[Help with translations!]

Main page
Contents
Current events
Random article
About Wikipedia
Contact us
Donate

Contribute
Help
Learn to edit
Community portal
Recent changes
Upload file

Tools
What links here
Related changes
Special pages
Permanent link
Page information
Cite this page
Wikidata item

Print/export
Download as PDF
Printable version

Languages
العربية
Deutsch
فارسی
Français
Русский
Simple English
Українська
中文
Edit links

Differential privacy

From Wikipedia, the free encyclopedia

Differential privacy (**DP**) is a system for publicly sharing information about a dataset by describing the patterns of groups within the dataset while withholding information about individuals in the dataset. The idea behind differential privacy is that if the effect of making an arbitrary single substitution in the database is small enough, the query result cannot be used to infer much about any single individual, and therefore provides privacy. Another way to describe differential privacy is as a constraint on the algorithms used to publish aggregate information about a statistical database which limits the disclosure of private information of records whose information is in the database. For example, differentially private algorithms are used by some government agencies to publish demographic information or other statistical aggregates while ensuring confidentiality of survey responses, and by companies to collect information about user behavior while controlling what is visible even to internal analysts.

Roughly, an algorithm is differentially private if an observer seeing its output cannot tell if a particular individual's information was used in the computation. Differential privacy is often discussed in the context of identifying individuals whose information may be in a database. Although it does not directly refer to identification and reidentification attacks, differentially private algorithms probably resist such attacks.[1]

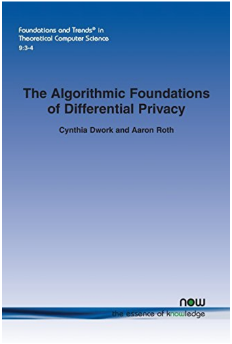
Differential privacy was developed by cryptographers and thus is often associated with cryptography, and draws much of its language from cryptography.

Contents [hide]

History [edit]

Official statistics organizations are charged with collecting information from individuals or establishments, and publishing aggregate data to serve the public interest. For example, the 1790 United States Census collected information about individuals living in the United States and published tabulations based on sex, age, race, and condition of servitude. Statistical organizations have long collected information under a promise of confidentiality that the information provided will be used for statistical purposes, but that the publications will not produce information that can be traced back to a specific individual or establishment. To accomplish this goal, statistical organizations have long suppressed information in their publications. For example, in a table presenting the sales of each business in a town grouped by business category, a cell that has information from only one company might be suppressed, in order to maintain the confidentiality of that company's specific sales.

The adoption of electronic information processing systems by statistical agencies in the 1950s and 1960s dramatically increased the number of tables that a statistical organization could produce and, in so doing, significantly increased the potential for an improper disclosure of confidential information. For example, if a business that had its sales numbers suppressed also had those numbers appear in the total sales of a region, then it



Near Instance-Optimality in Differential Privacy

Hilal Asi John C. Duchi
asi@stanford.edu jduchi@stanford.edu
Stanford University

Abstract

We develop two notions of instance optimality in differential privacy, inspired by classical statistical theory: one by defining a local minimax risk and the other by considering unbiased mechanisms and analogizing the Cramér-Rao bound, and we show that the local modulus of continuity of the estimand of interest completely determines these quantities. We also develop a complementary collection mechanisms, which we term the *inverse sensitivity* mechanisms, which are instance optimal (or nearly instance optimal) for a large class of estimands. Moreover, these mechanisms uniformly outperform the smooth sensitivity framework—on each instance—for several function classes of interest, including $\mathbb{R}$-valued continuous functions. We carefully present two instantiations of the mechanisms for median and robust regression estimation with corresponding experiments.

1 Introduction

We study instance-specific optimality for differentially private release of a function $f(\boldsymbol{x})$ of a dataset $\boldsymbol{x} \in \mathcal{X}^n$. In contrast to existing notions of optimality for private procedures, which measure mechanisms' worst case performance over all instances, we develop instance-specific notions to capture the difficulty of—and potential adaptivity of private mechanisms to—the given data x, rather than some potential worst case.

The trajectory of differential privacy research and private mechanisms reflects the desire to be adaptive both to the function f to be computed and dataset $\boldsymbol{x}$ at hand. Dwork et al.'s original perspective [21] targets the former, privatizing $f(x)$ by adding noise commensurate with the global sensitivity $\mathrm{GS}_f \coloneqq \sup_{\boldsymbol{x},\boldsymbol{x}': d_{\mathrm{ham}}(\boldsymbol{x},\boldsymbol{x}') \le 1} |f(\boldsymbol{x}) - f(\boldsymbol{x}')|$ of f and adapting to the function f at hand; more recent work expands this function adaptive approach [6]. As the classical approach can be conservative—it does not reflect the sensitivity of the underlying data $\boldsymbol{x}$—a natural idea is to add noise that scales with the *local sensitivity* (or local modulus of continuity) $\mathrm{LS}_f(\boldsymbol{x}) \coloneqq \sup_{\boldsymbol{x}': d_{\mathrm{ham}}(\boldsymbol{x},\boldsymbol{x}') \le 1} |f(\boldsymbol{x}) - f(\boldsymbol{x}')|$ of f at the dataset $\boldsymbol{x}$. Unfortunately, this fails to protect privacy, as the sensitivity itself may be compromising, leading Nissim et al. [35] to propose mechanisms that rely on smooth upper bounds to the local sensitivity. Yet these mechanisms are complex and, as our results show, may be conservative.

To understand these phenomena, we take a two-pronged approach, presenting both lower bounds on error and complementary (near) optimal mechanisms. We first consider the desiderata a lower bound should satisfy, following a program Cai and Low [11] develop (see also [16]):

Outline

Abstraction

Computer = machine able to make a few elementary operations on data, that can be combined arbitrarily
Problem = description of the desired output for any given input

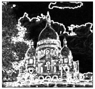
Input $\longrightarrow$  $\longrightarrow$ Output

Examples:

[3,2,5,1,4] $\longrightarrow$ [1,2,3,4,5]

le petit chat $\longrightarrow$ the little cat

 $\longrightarrow$ 5

 $\longrightarrow$ 

Two approaches

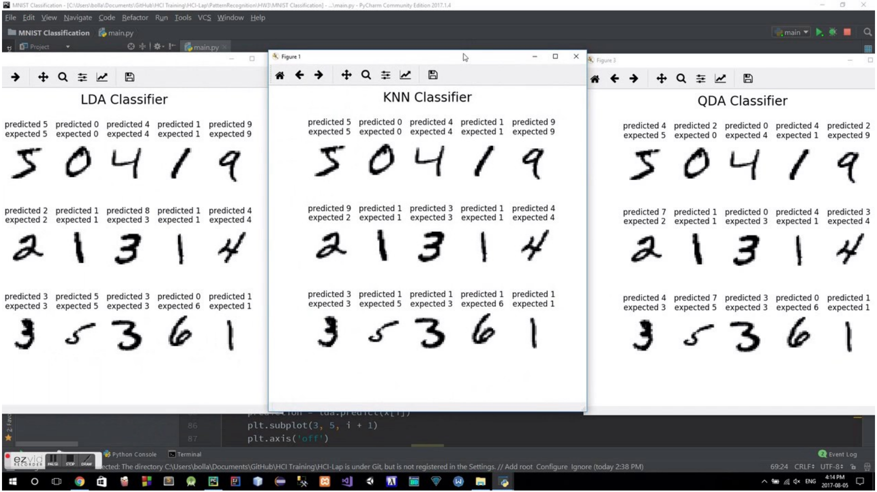
Classical approach : reduction = describe the sequence of elementary operations that permit to construct the output from the input
= computer programming (coding)

Artificial Intelligence = use a computer to *build itself* the program that will solve the task
= meta-programming

Machine Learning = feed the computer only with a (large number of) *example pairs* (input, output)
= search for the program that is supposed to work best on *new examples*

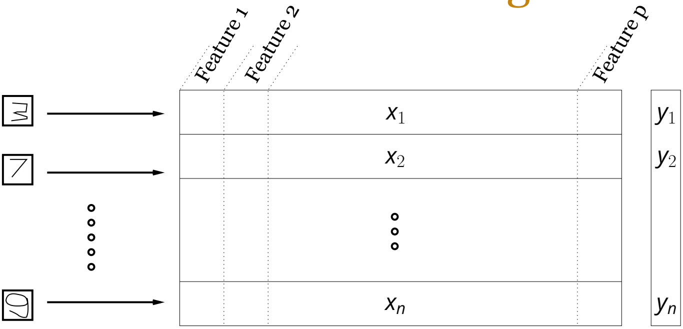
→ for each problem, both approaches are possible
→ but they are more or less efficient...

(Cliché) example: MNIST dataset



ENS DE LYON
UNIVERSITÉ DE LYON

Data in Machine Learning



$X \in \mathcal{M}_{n,p}(\mathbb{R})$ $\qquad$ $Y \in \mathcal{Y}^n$

Data: n-by-p matrix x

- n examples = points of observations
- p features = characteristics measured for each example

Classifier $\mathcal{A}_n$

$h_n : \mathcal{X} \to \mathcal{Y}$

$\mapsto 6$

Analysis

ML algorithm

- takes in input $x \in \mathfrak{M}_{n,p}(\mathbb{R})$
- performs some computations so as to optimize some criterion

  ex: minimizes the training error of some neural network
- returns the optimal predictor h_n

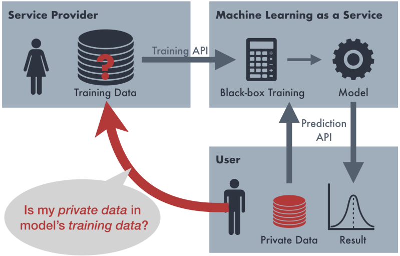
  for future use

- Simple example: return the average value, the proportion of votes for some candidate, etc.
- More difficult: 2-dim representation of the database, image recognition, automatic translation, etc.

Information leakage

Membership attack

Src: https://www.arxiv-vanity.com/papers/1904.05506/



Model inversion attack [Fredrikson et al. '2015]

Figure 1: An image recovered using a new model inversion attack (left) and a training set image of the victim (right). The attacker is given only the person's name and access to a facial recognition system that returns a class confidence score.

See https://arxiv.org/abs/1610.05820 for more information: *Membership Inference Attacks against Machine Learning Models* by Reza Shokri, Marco Stronati, Congzheng Song, Vitaly Shmatikov

Outline

Anonymization is not the solution

Linkage attack

[*Simple Demographics Often Identify People Uniquely*, by Latanya Sweeney] showed that gender, date of birth, and zip code are sufficient to uniquely identify the vast majority of Americans.

$\Longrightarrow$ By linking these attributes in a supposedly anonymized healthcare database to public voter records, she was able to identify the individual health record of the Governor of Massachussetts.

Differencing attack

Imposing request on many lines is not the solution

Example from [Dwork & Roth]:

- How many people in the database have the sickle cell trait?
- How many people, not named X, in the database have the sickle cell trait?

Differential Privacy

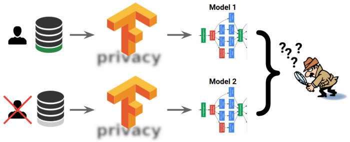
Differentially private algorithms make assurance that attackers can learn virtually nothing more about an individual than they would understand if that individual's record were absent from the dataset.

Smoker example

if an individual is openly ”smoking” but wants privacy on her medical status,

- a medical study will prove the risk associated with smoking (whether she participates or not)
- a *DP* study will make it impossible to know if she indeed participated or not, even to someone who would have all the remaining information

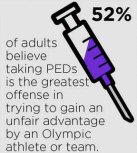
Fundamental Law of Information Recovery:
Need to *randomize* the output.



Src: `https://blog.tensorflow.org`

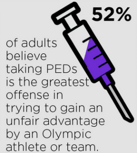
Survey on triathletes: "do you use doping?"

Triathletes doping status $X_i \overset{iid}{\sim} \mathcal{B}(p)$
but they may lie: answer $Y_i \in \{0, 1\}$



Survey on triathletes: ”do you use doping?”

Triathletes doping status $X_i \overset{iid}{\sim} \mathcal{B}(p)$
but they may lie: answer $Y_i \in \{0, 1\}$



RANDOMIZED RESPONSE: A SURVEY TECHNIQUE FOR ELIMINATING EVASIVE ANSWER BIAS

STANLEY L. WARNER
Claremont Graduate School

For various reasons individuals in a sample survey may prefer not to confide to the interviewer the correct answers to certain questions. In such cases the individuals may elect not to reply at all or to reply with incorrect answers. The resulting evasive answer bias is ordinarily difficult to assess. In this paper it is argued that such bias is potentially removable through allowing the interviewee to maintain privacy through the device of randomizing his response. A randomized response method for estimating a population proportion is presented as an example. Unbiased maximum likelihood estimates are obtained and their mean square errors are compared with the mean square errors of conventional estimates under various assumptions about the underlying population.

1. INTRODUCTION

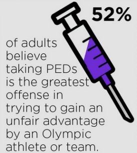
FOR reasons of modesty, fear of being thought bigoted, or merely a reluctance to confide secrets to strangers, many individuals attempt to evade certain questions put to them by interviewers. In survey vernacular, these people become the "non-cooperative" group [5, pp. 235–72], either refusing outright to be surveyed, or consenting to be surveyed but purposely providing wrong answers to the questions. In the one case there is the problem of refusal bias [1, pp. 355–61], [2, pp. 33–6], [5, pp. 261–9]; in the other case there is the problem of response bias [3, p. 89], [4, pp. 280–325].

Journal of the American Statistical Association, Mar. 1965, Vol.60, No.309, pp. 63-69

See also Chong, Chun Yin Andy & Chu, Amanda & So, Mike & Chung, Ray. (2019). *Asking Sensitive Questions Using the Randomized Response Approach in Public Health Research: An Empirical Study on the Factors of Illegal Waste Disposal*. International Journal of Environmental Research and Public Health.

Survey on triathletes: "do you use doping?"

Triathletes doping status $X_i \overset{iid}{\sim} \mathcal{B}(p)$
but they may lie: answer $Y_i \in \{0, 1\}$



Randomized Response [Warner'65]

Flip a coin, then:

$\rightarrow$ if tails, answer according to another coin flip

$\rightarrow$ if heads, give the right answer

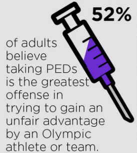
$$\mathbb{P}(Y = 1 | X = x) = 1/4 + x/2 \qquad \frac{\mathbb{P}(Y = 1 | X = 1)}{\mathbb{P}(Y = 1 | X = 0)} = 3$$

- No triathlete can be prosecuted one cannot condemn 1/4th of the innocent triathletes!
- But still permits to estimate the proportion of dopers by $2\bar{Y}_n - 1$.

Cost: for the same precision, requires $\approx 4x$ more data or even more if $x(1 - x) \ll 1$

Survey on triathletes: "do you use doping?"

Triathletes doping status $X_i \overset{iid}{\sim} \mathcal{B}(p)$
but they may lie: answer $Y_i \in \{0, 1\}$



Randomized Response [Warner'65]

Flip a coin, then:

$\rightarrow$ if tails, answer according to another coin flip

$\rightarrow$ if heads, give the right answer

$$\mathbb{P}(Y = 1|X = x) = 1/4 + x/2 \qquad \frac{\mathbb{P}(Y = 1|X = 1)}{\mathbb{P}(Y = 1|X = 0)} = 3$$

- No triathlete can be prosecuted one cannot condemn 1/4th of the innocent triathletes!
- But still permits to estimate the proportion of dopers by $2\overline{Y}_n - 1$.

Cost: for the same precision, requires $\approx 4x$ more data or even more if $x(1-x) \ll 1$

"smoker example": if $\hat{p} = 98\%$,
a lot of information on each triathlete
BUT no more than if she had not participated in the study

UNIVERSITÉ DE LYON

ENS DE LYON

Formal Definition

Randomized algorithm $\mathcal{A}(x)$ = random variable on $\mathcal{T}$

Def: Neighboring databases $x \sim x'$ if $\exists i \in \{1, \ldots, n\}, \forall j \neq i, x_{i,\cdot} = x'_{j,\cdot}$

Differential Privacy

[« Calibrating Noise to Sensitivity », TCC'2006, by Cynthia Dwork, Frank McSherry, Kobbi Nissim et Adam Smith $\Longrightarrow$ Gödel Prize 2017]

$\mathcal{A}$ is ϵ-DP if for all $x \sim x'$ and all $\mathcal{S} \subset \mathcal{T}$

$$\mathbb{P}\big(\mathcal{A}(x) \in S\big) \leq e^{\epsilon}\, \mathbb{P}\big(\mathcal{A}(x') \in S\big)$$

Equivalently,

- if $\mathcal{A}(x)$ is discrete, $\quad -\epsilon \leq \ln \frac{\mathbb{P}\left(\mathcal{A}(x)=t\right)}{\mathbb{P}\left(\mathcal{A}(x')=t\right)} \leq \epsilon \quad$ for all $t \in \mathcal{T}$
- if $\mathcal{A}(x)$ has density $f(\cdot|x)$, $\quad -\epsilon \leq \ln \frac{f(t|x)}{f(t|x')} \leq \epsilon \quad$ for all $t \in \mathcal{T}$

Formal Definition

Randomized algorithm $\mathcal{A}(x)$ = random variable on $\mathcal{T}$

Def: Neighboring databases $x \sim x'$ if $\exists i \in \{1, \ldots, n\}, \forall j \neq i, x_{i,\cdot} = x'_{j,\cdot}$

Differential Privacy

[« Calibrating Noise to Sensitivity », TCC'2006, by Cynthia Dwork, Frank McSherry, Kobbi Nissim et Adam Smith $\Longrightarrow$ Gödel Prize 2017]

$\mathcal{A}$ is ϵ-DP if for all $x \sim x'$ and all $\mathcal{S} \subset \mathcal{T}$

$$\mathbb{P}\big(\mathcal{A}(x) \in S\big) \leq e^{\epsilon}\, \mathbb{P}\big(\mathcal{A}(x') \in S\big)$$

In the previous example on the DP survey, algorithm $\mathcal{A}(x) = (Y_1, \ldots, Y_n)$ is $\ln(3)$-DP.

Note that it outputs an entire (differentially private), which is unusual: more often, we just want the answer to a query.

Formal Definition

Randomized algorithm $\mathcal{A}(x)$ = random variable on $\mathcal{T}$

Def: Neighboring databases $x \sim x'$ if $\exists i \in \{1, \ldots, n\}, \forall j \neq i, x_{i,\cdot} = x'_{j,\cdot}$

Differential Privacy

[« Calibrating Noise to Sensitivity », TCC'2006, by Cynthia Dwork, Frank McSherry, Kobbi Nissim et Adam Smith $\Longrightarrow$ Gödel Prize 2017]

$\mathcal{A}$ is ϵ-DP if for all $x \sim x'$ and all $\mathcal{S} \subset \mathcal{T}$

$$\mathbb{P}\big(\mathcal{A}(x) \in S\big) \leq e^{\epsilon}\, \mathbb{P}\big(\mathcal{A}(x') \in S\big)$$

A person's privacy cannot be compromised by a statistical release if their data are not in the database. Therefore, with differential privacy, the goal is to give each individual roughly the same privacy that would result from having their data removed. That is, the statistical functions run on the database should not overly depend on the data of any one individual.

Formal Definition

Randomized algorithm $\mathcal{A}(x)$ = random variable on $\mathcal{T}$

Def: Neighboring databases $x \sim x'$ if $\exists i \in \{1, \dots, n\}, \forall j \neq i, x_{i,\cdot} = x'_{j,\cdot}$

Differential Privacy

[« Calibrating Noise to Sensitivity », TCC'2006, by Cynthia Dwork, Frank McSherry, Kobbi Nissim et Adam Smith $\Longrightarrow$ Gödel Prize 2017]

$\mathcal{A}$ is ϵ-DP if for all $x \sim x'$ and all $\mathcal{S} \subset \mathcal{T}$

$$\mathbb{P}\big(\mathcal{A}(x) \in S\big) \leq e^{\epsilon}\, \mathbb{P}\big(\mathcal{A}(x') \in S\big)$$

An algorithm is said to be differentially private if by looking at the output, one cannot tell whether any individual's data was included in the original dataset or not.

Cryptographic origins (and vocabulary).

Formal Definition

Randomized algorithm $\mathcal{A}(x)$ = random variable on $\mathcal{T}$

Def: Neighboring databases $x \sim x'$ if $\exists i \in \{1, \ldots, n\}, \forall j \neq i, x_{i,\cdot} = x'_{j,\cdot}$

Differential Privacy

[« Calibrating Noise to Sensitivity », TCC'2006, by Cynthia Dwork, Frank McSherry, Kobbi Nissim et Adam Smith $\Longrightarrow$ Gödel Prize 2017]

$\mathcal{A}$ is ϵ-DP if for all $x \sim x'$ and all $\mathcal{S} \subset \mathcal{T}$

$$\mathbb{P}\big(\mathcal{A}(x) \in S\big) \leq e^{\epsilon}\, \mathbb{P}\big(\mathcal{A}(x') \in S\big)$$

Differential privacy mathematically guarantees that anyone seeing the result of a differentially private analysis will essentially make the same inference about any individual's private information, whether or not that individual's private information is included in the input to the analysis.

Properties

Post-processing

If $\mathcal{A} : \mathcal{M}_{n,p}(\mathbb{R}) \to \mathfrak{M}_1(\mathcal{T})$ is ϵ-DP, then for every $f : \mathcal{T} \to \mathcal{T}'$ algorithm $f \circ \mathcal{A}$ is also ϵ-DP.

Group privacy

If $x \sim x_2 \sim \cdots \sim x_k$, then for all $\mathcal{S} \subset \mathcal{T}$, $\quad \mathbb{P}\big(\mathcal{A}(x) \in S\big) \leq e^{k\epsilon}\, \mathbb{P}\big(\mathcal{A}(x_k) \in S\big)$.

”Composition”

If $\mathcal{A}_1 : \mathcal{M}_{n,p}(\mathbb{R}) \to \mathfrak{M}_1(\mathcal{T})$ is ϵ-DP and if $\mathcal{A}_2 : \mathcal{M}_{n,p}(\mathbb{R}) \to \mathfrak{M}_1(\mathcal{T}')$ is ϵ'-DP, then $x \mapsto \big(\mathcal{A}_1(x), \mathcal{A}_2(x)\big)$ is $(\epsilon + \epsilon')$-DP.

DP defines privacy not as a binary notion of ”was the data of individual exposed or not”, but rather a matter of accumulative risk.

They already adopted DP

Several uses of differential privacy in practice are known to date:

- 2008 U.S. Census Bureau, for showing commuting patterns.
- 2014 Google's RAPPOR, for telemetry such as learning statistics about unwanted software hijacking users' settings.
- 2015 Google, for sharing historical traffic statistics.
- 2016 Apple announced its intention to use differential privacy in iOS 10 to improve its Intelligent personal assistant technology.
- 2017 Microsoft, for telemetry in Windows.
- 2019 Privitar Lens is an API using differential privacy.
- 2019 Sarus provides ML with DP as a service.
- 2020 LinkedIn, for advertiser queries.

Outline

Example: Majority of Binary Observations

$n = 2k + 1$, target $f(x) = \mathbb{1}\left\{\sum x_i \geq n/2\right\} = \mathrm{median}(x)$.

- $\mathcal{A}(x)$ depends only on $s = \sum x_i \implies \mathbb{P}(\mathcal{A}(x) = 1) =: p(s)$
- By symmetry $p(n - s) = 1 - p(s)$
- $p(k+1) \leq e^{\epsilon} p(k) = e^{\epsilon}(1 - p(k+1)) \implies p(k+1) \leq \frac{1}{1+e^{-\epsilon}}$
- More generally, for all $s > n/2$, $p(s) \leq \dfrac{1}{1 + e^{-(2s-n)\epsilon}}$ and $p(n) \leq \dfrac{1}{1 + e^{-(n+2)\epsilon}}$
- In fact, $p(s) = \dfrac{1}{1 + e^{-(2s-n)\epsilon/2}}$ *is* ϵ-DP (see next slide)

  Better: $p(k+r) = \dfrac{1}{1 + e^{-r\epsilon}}$ *is* ϵ-DP: $\frac{p(k+r+1)}{p(k+r)} = e^{\epsilon}\,\frac{1+e^{-r\epsilon}}{e^{\epsilon}+e^{-r\epsilon}} \leq e^{\epsilon}$ and similarly for $\dfrac{p(k+1)}{p(k)}$ and $\dfrac{1-p(k+r+1)}{1-p(k+r)}$.
- Requires $n \gg 1/\epsilon$
- If $|s - n/2| \geq 3/\epsilon$, the answer is correct with probability $\geq 95\%$
- But if $|s - n/2| \leq \sqrt{n}$, the chances are high that the majority in the sample is not the majority in the population
- $\implies$ if $\epsilon \geq 3/\sqrt{n} \iff n \geq 9/\epsilon^2$, ϵ-DP does not really cost any reliability!

More generally: Exponential Mechanism

If $\mathcal{T}$ is discrete, one wants $\mathcal{A}$ to assign a probability to each possible outcome $t \in \mathcal{T}$ that depends on its utility $u(x,t)$ on the data x.
The sensibility of u is defined as $\Delta u = \max_{t \in \mathcal{T}} \max_{x \sim x'} \left| u(x,t) - u(x',t) \right|$.

Exponential Mechanism

The algorithm $\mathcal{A}$ defined by $\mathbb{P}(\mathcal{A}(x) = t) = \dfrac{\exp\left(\frac{\epsilon u(x,t)}{2\Delta u}\right)}{\sum_{t' \in \mathcal{T}} \exp\left(\frac{u(x,t')\epsilon}{2\Delta u}\right)}$ is ϵ-DP.

Previous example: for $u(x,t) = (2t-1)\left(s - \dfrac{n}{2}\right) = -u(x, 1-t)$,

$$\mathbb{P}(\mathcal{A}(x) = 1) = \frac{\exp\left(\frac{\left(s - \frac{n}{2}\right)\epsilon}{2}\right)}{\exp\left(\frac{\left(s - \frac{n}{2}\right)\epsilon}{2}\right) + \exp\left(-\frac{\left(s - \frac{n}{2}\right)\epsilon}{2}\right)} = \frac{1}{1 + \exp\left(-\left(s - \frac{n}{2}\right)\epsilon\right)}$$

Proof

For every $t \in \mathcal{T}$ and $x \sim x'$,

$$\frac{\mathbb{P}(\mathcal{A}(x) = t)}{\mathbb{P}(\mathcal{A}(x') = t)} = \frac{\exp\left(\frac{\epsilon u(x,t)}{2\Delta u}\right)}{\sum_{t' \in \mathcal{T}} \exp\left(\frac{u(x,t')\epsilon}{2\Delta u}\right)} \Bigg/ \frac{\exp\left(\frac{\epsilon u(x',t)}{2\Delta u}\right)}{\sum_{t' \in \mathcal{T}} \exp\left(\frac{u(x',t')\epsilon}{2\Delta u}\right)}$$

$$= \exp\left(\frac{\epsilon\big(u(x,t) - u(x',t)\big)}{2\Delta u}\right) \frac{\sum_{t' \in \mathcal{T}} \exp\left(\frac{\big(u(x',t') - u(x,t)\big)\epsilon}{2\Delta u}\right) \exp\left(\frac{u(x,t')\epsilon}{2\Delta u}\right)}{\sum_{t' \in \mathcal{T}} \exp\left(\frac{u(x,t')\epsilon}{2\Delta u}\right)}$$

$$\leq \exp\left(\frac{\epsilon}{2}\right) \frac{\sum_{t' \in \mathcal{T}} \exp\left(\frac{\epsilon}{2}\right) \exp\left(\frac{u(x,t')\epsilon}{2\Delta u}\right)}{\sum_{t' \in \mathcal{T}} \exp\left(\frac{u(x,t')\epsilon}{2\Delta u}\right)} = \exp(\epsilon)\ .$$

Bound on the resulting utility

Theorem

For every database x the exponential mechanism satisfies:

$$\mathbb{P}\left(u(x, \mathcal{A}(x)) \leq u(x, f(x)) - \frac{2\Delta u}{\epsilon} \ln \frac{|\mathcal{T}|}{\delta}\right) \leq \delta \, .$$

Proof: for any $t \in \mathcal{T}$ such that $u(x, t) \leq u(x, f(x)) - 2\Delta u \epsilon^{-1} \ln\left(|\mathcal{T}|/\delta\right)$,

$$\mathbb{P}(\mathcal{A}(x) = t) \leq \frac{\exp\left(\frac{\epsilon\left(u\left(x, f(x)\right) - \frac{2\Delta u}{\epsilon} \ln \frac{|\mathcal{T}|}{\delta}\right)}{2\Delta u}\right)}{\exp\left(\frac{\epsilon u(x, f(x))}{2\Delta u}\right)} = \frac{\delta}{|\mathcal{T}|}$$

and there are at most $\mathcal{T}$ of them.

Bound on the resulting utility

Theorem

For every database x the exponential mechanism satisfies:

$$\mathbb{P}\left(u\big(x, \mathcal{A}(x)\big) \leq u\big(x, f(x)\big) - \frac{2\Delta u}{\epsilon} \ln \frac{|\mathcal{T}|}{\delta}\right) \leq \delta\ .$$

Equivalently, for every $v > 0$

$$\mathbb{P}\left(u\big(x, \mathcal{A}(x)\big) \leq u\big(x, f(x)\big) - v\right) \leq |\mathcal{T}| e^{-\frac{\epsilon v}{2\Delta_u}}\ .$$

In the example:

$\mathbb{P}\big(\mathcal{A}(x) \neq f(x)\big) = \mathbb{P}\Big(u\big(x, \mathcal{A}(x)\big) \leq u\big(x, f(x)\big) - 2u\big(x, f(x)\big)\Big) \leq 2 \exp\left(-\frac{u\big(x, f(x)\big)\epsilon}{\Delta u}\right) = 2e^{-\left|s - \frac{n}{2}\right|\epsilon}.$

Lower Bound for Discrete Mechanisms

$\mathcal{A}$ (discrete) is said to be unbiased if for x and $t \in \mathcal{T}$,

$$\mathbb{P}\big(\mathcal{A}(x) = t\big) \leq \mathbb{P}\big(\mathcal{A}(x) = f(x)\big) \ .$$

Inverse Sensibility

The inverse sensibility of function f on data x to output $t \in \mathcal{T}$ is defined as

$$\mathcal{D}_f(x, t) = \min \left\{ k : \exists x \sim x_1 \sim \ldots \sim x_k \text{ and } f(x_k) = t \right\} \ .$$

Lower bound

For every unbiased ϵ-DP mechanism $\mathcal{A}$, $\mathbb{P}\big(\mathcal{A}(x) = f(x)\big) \leq \dfrac{1}{\sum_{t \in \mathcal{T}} e^{-2\mathcal{D}_f(x,t)\epsilon}}$.

UNIVERSITÉ DE LYON

ENS DE LYON

Proof

Let $t \in \mathcal{T}$ and let x' be such that $x \sim x_1 \sim \ldots \sim x_{\mathcal{D}_f(x,t)-1} \sim x'$ and $f(x') = t$. Differential privacy implies

$$\frac{\mathbb{P}\big(\mathcal{A}(x) = t\big)}{\mathbb{P}\big(\mathcal{A}(x) = f(x)\big)} = \frac{\mathbb{P}\big(\mathcal{A}(x) = t\big)}{\mathbb{P}\big(\mathcal{A}(x') = t\big)} \frac{\mathbb{P}\big(\mathcal{A}(x') = t = f(x')\big)}{\mathbb{P}\big(\mathcal{A}(x') = f(x)\big)} \frac{\mathbb{P}\big(\mathcal{A}(x') = f(x)\big)}{\mathbb{P}\big(\mathcal{A}(x) = f(x)\big)} \geq e^{-\mathcal{D}_f(x,t)\epsilon} \times 1 \times e^{-\mathcal{D}_f(x,t)\epsilon}$$

and hence

$$1 = \sum_{t \in \mathcal{T}} \frac{\mathbb{P}\big(\mathcal{A}(x) = t\big)}{\mathbb{P}\big(\mathcal{A}(x) = f(x)\big)} \, \mathbb{P}\big(\mathcal{A}(x) = f(x)\big) \geq \sum_{t \in \mathcal{T}} e^{-2\mathcal{D}_f(x,t)\epsilon} \, \mathbb{P}\big(\mathcal{A}(x) = f(x)\big) .$$

In the previous example, $\mathcal{D}_f(x, 1 - f(x)) = \left| s - \frac{n}{2} \right| + \frac{1}{2}$ and this yields:

$$\mathbb{P}\big(\mathcal{A}(x) = f(x)\big) \leq \frac{1}{1 + e^{-\big(|2s-n|+1\big)\epsilon}} .$$

The Exponential Mechanism above is almost optimal: it has $\mathbb{P}\big(\mathcal{A}(x) = f(x)\big) = \frac{1}{1 + e^{-\frac{|2s-n|\epsilon}{2}}}$.

UNIVERSITÉ DE LYON

ENS DE LYON

Inverse Sensitivity Mechanism

The inverse sensibility $\mathcal{D}_f$ yields a good candidate utility function for an exponential mechanism! In fact, $\Delta\mathcal{D}_f = 1$, hence the ϵ-DP Inverse Sensitivity Mechanism (ISM)

$$\mathbb{P}\big(\mathcal{A}(x) = t\big) = \frac{e^{-\epsilon\mathcal{D}_f(x,t)/2}}{\sum_{t'\in\mathcal{T}} e^{-\mathcal{D}_f(x,t')\epsilon/2}} .$$

Remark: if $|\mathcal{T}| = 2$ the denominator 2 is not needed:

$$\mathbb{P}\big(\mathcal{A}(x) = t\big) = \frac{e^{-\epsilon\mathcal{D}_f(x,t)}}{\sum_{t'\in\mathcal{T}} e^{-\mathcal{D}_f(x,t')\epsilon}}$$

is ϵ-DP.

Previous example: here $\mathcal{D}_f(x, 1 - f(x)) = \left|s - \frac{n}{2}\right| + \frac{1}{2}$, the ISM $p(s) = \frac{1}{1 + e^{-\left(s-k-\mathbb{1}\{s\leq k\}\right)\epsilon}}$ is an $\epsilon-$DP mechanism slightly better than the exponential mechanism above.

Near-Optimality of the Inverse Senbility Mechanism

$1/4$-Optimality of the ISM

The ISM $\mathcal{A}$ is "more accurate" than any $\epsilon/4$-DP algorithm $\mathcal{A}'$:

$$\mathbb{P}\big(\mathcal{A}'(x) = f(x)\big) \leq \mathbb{P}\big(\mathcal{A}(x) = f(x)\big) \; .$$

Proof: Since $\mathcal{D}_f(x, f(x)) = 0, \quad \mathbb{P}\big(\mathcal{A}(x) = f(x)\big) = 1/\sum_{t\in\mathcal{T}} e^{-\mathcal{D}_f(x,t)\epsilon/2}$.

Recall the lower bound: for every unbiased ϵ-DP mechanism $\mathcal{A}'$, $\quad \mathbb{P}\big(\mathcal{A}'(x) = f(x)\big) \leq 1/\sum_{t\in\mathcal{T}} e^{-2\mathcal{D}_f(x,t)\epsilon}$.

Remark: if $|\mathcal{T}| = 2$, the ISM is $1/2$-optimal.

Continuous Exponential Mechanism

If continuous $\mathcal{T}$, taking $\mathcal{A}(x)$ with density

$$f_x(t) = \frac{\exp\left(\frac{\epsilon u(x,t)}{2\Delta u}\right)}{\int_{\mathcal{T}} \exp\left(\frac{u(x,t')\epsilon}{2\Delta u}\right) dt'}$$

also yields an ϵ-DP mechanism.

- The ISM is hence a very good candidate in theory.
- It is reminiscent of statistical physics "Gibbs law" (thermodynamics).
- It can be hard to sample from.
- In fact, the discrete case is already computationally challenging when the output space $\mathcal{T}$ is "big".
- Research question: does approximate sampling preserve differential privacy?

Outline

Example: estimate the mean

Here $x \in \mathfrak{M}_{n,1}(\{0,1\})$ and $f(x) = \bar{x}$. $\mathcal{T} = \{0, \frac{1}{n}, \ldots, \frac{n-1}{n}, 1\}$.
Inverse sensibility function: $\mathcal{D}_f(x,t) = n|t - \bar{x}|$.

ISM: $\mathbb{P}(\mathcal{A}(x) = t) = \dfrac{e^{-\epsilon n|t-\bar{x}|/2}}{\sum_{j=0}^{n} e^{-\epsilon n|j/n-\bar{x}|/2}}$.

$$\mathbb{P}\left(\mathcal{A}(x) \geq \bar{x} + \frac{k}{n}\right) = \frac{\sum_{j=k}^{n} e^{-\epsilon j/2}}{2\sum_{j=0}^{n} e^{-\epsilon j/2}} = \frac{e^{-\epsilon k/2}}{2} \quad \text{and} \quad \mathbb{P}\left(\mathcal{A}(x) \leq \bar{x} - \frac{k}{n}\right) = \frac{e^{-\epsilon k/2}}{2}$$

$\Longrightarrow$ up to the discretization, $\mathcal{A}(x)$ has the same distribution as

$$\mathcal{A}'(x) = \bar{x} + Y$$

where $Y \sim \mathrm{Lap}\left(\frac{2}{n\epsilon}\right)$ has density $\ell_{\frac{2}{n\epsilon}}(x) = \dfrac{n\epsilon}{4} e^{-\frac{n\epsilon|x|}{2}}$ since $\mathbb{P}(\mathcal{A}'(x) \geq \bar{x} + k/n) = \frac{e^{-\epsilon k/2}}{2}$.

$\Longrightarrow$ very simple mechanism: just add (well-calibrated) Laplace noise!

Laplace mechanism

The L^1-sensitivity of a function $f : \mathfrak{M}_{n,p}(\mathbb{R}) \to \mathbb{R}^k$ is defined as

$$\Delta f = \max_{x \sim x'} \|f(x) - f(x')\|_1 \ .$$

Example: if $f(x) = n^{-1} \sum x_i$ and $x_i \in [0, 1]$, then $\Delta(f) = 1/n$.

$\mathrm{Lap}(\sigma)$ has density $\ell_\sigma(x) = \dfrac{1}{2\sigma} e^{-\frac{|x|}{\sigma}}$.

Laplace Mechanism

The Laplace mechanism for $f : \mathfrak{M}_{n,p}(\mathbb{R}) \to \mathbb{R}^k$ defined by

$$\mathcal{A}(x) = f(x) + (Y_1, \ldots, Y_k), \qquad Y_i \overset{iid}{\sim} \mathrm{Lap}\left(\frac{\Delta f}{\epsilon}\right) \ .$$

is ϵ-differentially private.

Proof

For $t \in \mathbb{R}^k$ let

$$p_x(t) = \prod_{j=1}^{k} \frac{\epsilon}{2\Delta f} \exp\left(-\frac{\epsilon \left|t_j - f(x)_j\right|}{\Delta f}\right) = \left(\frac{\epsilon}{2\Delta f}\right)^k \exp\left(-\frac{\epsilon \left\|t - f(x)\right\|_1}{\Delta f}\right)$$

be the density of $\mathcal{A}(x)$. Then for every $x \sim x'$ and every $t \in \mathbb{R}^k$,

$$\frac{p_x(t)}{p_{x'}(t)} = \frac{\exp\left(-\frac{\epsilon \left\|t-f(x)\right\|_1}{\Delta f}\right)}{\exp\left(-\frac{\epsilon \left\|t-f(x')\right\|_1}{\Delta f}\right)} = \exp\left(-\frac{\epsilon\left(\left\|t - f(x)\right\|_1 - \left\|t - f(x')\right\|_1\right)}{\Delta f}\right)$$

$$\leq \exp\left(\frac{\epsilon \left\|f(x) - f(x')\right\|_1}{\Delta f}\right) \leq \exp(\epsilon)$$

since by definition $\|f(x) - f(x')\|_1 \leq \Delta f$.

Examples

- Counting queries — accuracy independent of *n*
- Histogram queries — accuracy independent of *n and of the number of bins!*
- Most common value of a variable
- Noisy max — accuracy depends on the number of values, but not on *n*
- Estimate average (or sum)

⚠ **Finite precision arithmetic $\Longrightarrow$ possible privacy leaks**

For example, one can have $\mathbb{P}\big(\mathcal{A}(x) = t\big) > 0$ while $\mathbb{P}\big(\mathcal{A}(x') = t\big) = 0$ for two neighbors $x \sim x'$ because of rounding.

Lower bound for continuous mechanisms

We consider a target function $f : \mathfrak{M}_{n,p}(\mathbb{R}) \to \mathbb{R}$. For every $x \in \mathfrak{M}_{n,p}(\mathbb{R})$, let $\Delta_f(x,k) = \sup\left\{|f(x_k) - f(x)| : x \sim x_1 \sim \cdots \sim x_k\right\}$ and $\Delta_f(x) = \Delta_f(x,1)$.

Lower bound

$\mathcal{A}$ is unbiased if for every x, $\mathbb{E}[\mathcal{A}(x)] = f(x)$. Then, for every $x \in \mathfrak{M}_{n,p}(\mathbb{R})$,

$$\mathbb{E}\left[\left(\mathcal{A}(x) - f(x)\right)^2\right] \geq \sup_k \frac{\Delta_f(x,k)^2/4}{1 + e^{2k\epsilon}}$$

and in particular if $\Delta_f(x,k) = k\Delta_f(x)$, then for $\epsilon \leq 1/2$

$$\mathbb{E}\left[\left(\mathcal{A}(x) - f(x)\right)^2\right] \geq \frac{\Delta_f(x)^2}{68\epsilon^2}$$

For the average $f(x) = \bar{x}_n$, $\Delta_f(x,k) = k\Delta_f(x) = k/n$ and $\mathbb{E}\left[\left(\mathcal{A}(x) - f(x)\right)^2\right] \geq \dfrac{1}{68n^2\epsilon^2}$.

UNIVERSITÉ DE LYON

ENS DE LYON

Proof

Let $k \geq 1$ and let x_k be such that $x \sim x_1 \sim \cdots \sim x_k$ and $|f(x_k) - f(x)| = \Delta_f(x,k)$. By definition,

$$\mathbb{E}\left[(\mathcal{A}(x_k) - t)^2\right] = \int_{\mathcal{T}} (s-t)^2 d\mathbb{P}_{\mathcal{A}(x_k)}(s) \leq \int_{\mathcal{T}} (s-t)^2 e^{k\epsilon} d\mathbb{P}_{\mathcal{A}(x)}(s) = e^{k\epsilon}\, \mathbb{E}\left[(\mathcal{A}(x) - t)^2\right]$$

and hence

$$\frac{\mathbb{E}\left[(\mathcal{A}(x_k) - f(x_k))^2\right]}{\mathbb{E}\left[(\mathcal{A}(x) - f(x))^2\right]} = \frac{\mathbb{E}\left[(\mathcal{A}(x_k) - f(x_k))^2\right]}{\mathbb{E}\left[(\mathcal{A}(x_k) - f(x))^2\right]} \frac{\mathbb{E}\left[(\mathcal{A}(x_k) - f(x))^2\right]}{\mathbb{E}\left[(\mathcal{A}(x) - f(x))^2\right]} \leq 1 \times e^{k\epsilon}$$

since $\mathcal{A}$ is unbiased. Therefore, by the Bienaymé-Chebishev inequality

$$\mathbb{P}\left(|\mathcal{A}(x) - f(x)| \geq \frac{\Delta_f(x,k)}{2}\right) \leq \frac{4\, \mathbb{E}\left[(\mathcal{A}(x) - f(x))^2\right]}{\Delta_f(x,k)^2}$$ and

$$\mathbb{P}\left(|\mathcal{A}(x) - f(x_k)| \geq \frac{\Delta_f(x,k)}{2}\right) \leq e^{k\epsilon}\mathbb{P}\left(|\mathcal{A}(x_k) - f(x_k)| \geq \frac{\Delta_f(x,k)}{2}\right) \leq \frac{4e^{2k\epsilon}\, \mathbb{E}\left[(\mathcal{A}(x) - f(x))^2\right]}{\Delta_f(x,k)^2}.$$

But since $|f(x_k) - f(x)| = \Delta_f(x,k)$,

$$1 \leq \mathbb{P}\left(|\mathcal{A}(x) - f(x)| \geq \frac{\Delta_f(x,k)}{2}\right) + \mathbb{P}\left(|\mathcal{A}(x) - f(x_k)| \geq \frac{\Delta_f(x,k)}{2}\right) \leq \frac{4\left(1 + e^{2k\epsilon}\right)\mathbb{E}\left[(\mathcal{A}(x) - f(x))^2\right]}{\Delta_f(x,k)^2}.$$

The second statement is obtained by the choice $k = \lceil 1/(2\epsilon) \rceil$, noting that $1/(2\epsilon) \leq k \leq 1/\epsilon$.

UNIVERSITÉ DE LYON

ENS DE LYON

Working Research: Multi-quantile Estimation

Differentially Private Quantiles

Jennifer Gillenwater* Matthew Joseph† Alex Kulesza‡

September 21, 2021

Abstract

Quantiles are often used for summarizing and understanding data. If that data is sensitive, it may be necessary to compute quantiles in a way that is differentially private, providing theoretical guarantees that the result does not reveal private information. However, when multiple quantiles are needed, existing differentially private algorithms fare poorly: they either compute quantiles individually, splitting the privacy budget, or summarize the entire distribution, wasting effort. In either case the result is reduced accuracy. In this work we propose an instance of the exponential mechanism that simultaneously estimates exactly m quantiles from n data points while guaranteeing differential privacy. The utility function is carefully structured to allow for an efficient implementation that returns estimates of all m quantiles in time $O(mn\log(n) + m^2n)$. Experiments show that our method significantly outperforms the current state of the art on both real and synthetic data while remaining efficient enough to be practical.

1 Introduction

Quantiles are a widespread method for understanding real-world data, with example applications ranging from income [29] to birth weight [8] to standardized test scores [16]. At the same time, the individuals contributing data may require that these quantiles not reveal too much information about individual contributions. As a toy example, suppose that an individual joins a company that has exactly two salaries, and half of current employees have one salary and half have another. In this case, publishing the exact median company salary will reveal the new employee's salary.

Differential privacy [14] offers a solution to this problem. Informally, the distribution over a