Chapter 7

Privacy-preserving inference

Contents

The goal of this chapter is give a quick overview of privacy concerns in statistical inference, and to present a few standard statistical modellings designed to address the question. For further inspection of the topic, the interested reader shall refer to [Vad17], and to the lecture slides and practical sessions of Aurélien Bellet.

7.1 Introduction

7.1.1 Motivation

The classical statistical framework, based on data points, is usually referred to as *PAC-learning* [Val84] or *sample framework*. In this setting, the learner is given a set $\{x_1, \dots, x_n\}$ of n samples drawn, most commonly independently, from an unknown distribution P. From these samples, the learner then aims at estimating a parameter of interest $\theta(P)$ with high probability, and can use *any* technique (or algorithm) based on these samples.

Beyond this classical and almighty statistical setting, the modern practice of statistics raised concerns that naturally bring up to consider quantitative and qualitative estimation constraints. For instance, in many applications of learning methods, the studied data is contributed by individuals, and features represent their possibly private characteristics such as race, browsing history, geolocation, or

health history. The disclosure of such personal data can be harmful to the individuals. Hence, it is essential not to reveal too much information about any particular individual.

Keep in mind that even though one may think that they have nothing to hide, the most insignificant and ordinary facts may also be problematic if an individual is followed over time. For instance, if Alice buys bread every day for 20 years and then suddenly stops, then an analyst might conclude that Alice has been diagnosed with type 2 diabetes.

7.1.2 A few bad ideas

Anonimization is not safe A first idea towards privacy is to anonimize data by erasing the variables that yield obvious identification (name, address, age, etc), and then publish the resulting censored dataset. Unfortunately, this appears useless in practice, because the remaining data, still very rich, often suffices to recover which individual correspond to which person. For instance, it has been shown that in 2000, 87% of the U.S. population were be uniquely identifiable from the triple of their ZIP code, gender, and date of birth [NHF16]. Said otherwise, everything can turn out to quasi-identify individuals, especially in high-dimensional and sparse databases. Furthermore, combination of knowledge coming from another data source may also enable an adversary to de-anonimize both.

Aggregating statistics is not safe A seemingly more viable method would be to never publish data, and to mediate its access via a trusted interface (or oracle) that will only respond certain data queries. It is, however, very difficult to ensure that such a system does protect privacy: what rule should determine which query is valid (i.e. privacy-preserving) and which query is not?

For instance, such a system should forbid *differencing attacks*: combining aggregate queries to obtain precise information about specific individuals: "Average salary in a neighborhood before/after a family moving in". Here, we see that a combination of results from several queries can target an individual even though every single queries do not appear to do so. Hence, this can be hard to detect.

7.1.3 Toy examples in poll theory

Beyond ethical considerations leading to handle *existing* data privately, let us mention that ensuring some notion of privacy *when collecting* data can also benefit the learner, as we now illustrate.

Within a given population, say that you want to design a poll that measures the proportion $p \in [0,1]$ of an opinion (or trait), but for which you know that people may not answer honestly. For instance, think to poll questions such as "*Did you cheat at this exam?*", "*Do you fraud tax?*", or "*Do you support this sulfurous politician?*". For a given individual $i \in \{1,\dots,n\}$, write $X_i \in \{0,1\}$ for the *true* opinion (Yes/No) of person i, so that $(X_i)_{i\leq n}$ is an iid n-sample with common distribution Bernoulli(p).

In such a context, asking directly the question of interest to the individuals would lead to social biases that are difficult (if not impossible) to measure. That is, we would not observe $(X_i)_{i\leq n}$ directly, but (wildly) censored versions of it. To overcome this challenge and encourage individuals to answer honestly, the statistician shall hence use tricks when designing the poll. The idea is to have the individual feel[1] that their personal statement is kept private and protected.

Self-randomized response

A first strategy consists in introducing external randomness that the individual controls, so as to protect their privacy and hence induce honesty from them. More specifically, the instructions of the poll could be as follows:

[1] Purposely vague notion here! Privacy is mathematically formalized below.

- Flip two coins and keep the result of the flips secret.
- If you flipped:
  - Zero or one Tails, answer the question honestly.
  - Two Heads, answer the *opposite* of your actual opinion[2].

The fact that the individual keeps their coin flip secret in fundamental: it is the encryption key that leads to the privacy of their answer.

Write $\phi \in [0,1] \setminus \{1/2\}$ for the probability to be in the first option (i.e. not flipping two Heads). Then if Z_i is the answer of individual $i \in \{1,\dots,n\}$ to this poll, one easily checks that $Z_i \sim \text{Bernoulli}\big(p\phi + (1-p)(1-\phi)\big)$.

Exercise 7.1. *Build an estimator and an asymptotic confidence interval for p based on $(Z_i)_{i\le n}$ when $n \to \infty$.*

One may imagine other external randomness generators that coin flips, so that $\phi \in [0,1]$ could take any value. Intuitively, this parameter drives how private the poll keeps the individuals' opinion. The poll is not private at all for $\phi = 0$, and the "privacy" increases as ϕ increase. On the other hand, the larger ϕ, the more the information is lost about p: there seem to be a *privacy VS estimation* tradeoff.

Asking extra insignificant questions

To bypass the use of an additional randomization scheme while still encouraging honesty, a second strategy consists in "drowning" personal information X_i by adding other questions to the poll. That is, if Question 1 asks for the (disputable) opinion of interest, we can adjunct some unrelated and insignificant Question 2 to it. For instance, Question 2 could be "*Have you ever visited Brittany?*", or "*Do you practice cycling?*"? The poll would hence be as follows:

> Among Question 1 and Question 2, is your *true* answer "Yes" to at least one of them?

By its insignificance, it is Question 2 that effectively ensures privacy of the global answer.

Formally, if $(X_i, Y_i) \in \{0,1\} \times \{0,1\}$ writes for the couple of *true* opinions of individual $i \in \{1,\dots,n\}$ on Question 1 and Question 2 respectively, this poll only asks individuals to reveal $Z_i = \max\{X_i, Y_i\}$. If the (X_i, Y_i)'s are independent couples (i.e. uncorrelated questions) and that the Y_i's are independent copies with distribution Bernoulli(q) for some $q \in [0,1]$, then the $Z_i's$ are iid Bernoulli($p + q - pq$).

Exercise 7.2. *If $(Y'_j)_{j\le m}$ is a m-sample of* Bernoulli(q) *come from another (regular) poll with Question 2, build an asymptotic confidence interval for p based on $(Z_i)_{i\le n}$ and $(Y'_j)_{j\le m}$ when $m, n \to \infty$.*

Note that here, compared to the previous strategy where we controlled the randomness of the coin flip, the nuisance parameter q has to be estimated separately. This naturally induces more uncertainty.

7.2 Differential privacy

7.2.1 Datasets and histograms

In the framework to come, a trusted party holds a *dataset* on n individuals, represented by a tuple $D = (x_1,\dots,x_n) \in \mathcal{X}^n$ where $\mathcal{X}$ is the space of realizations.

Privacy can in fact formalize the need for hiding the very presence of an individual in dataset. This will naturally bring us to consider dataset with different sizes. To put all the datasets in the same

[2] A variant could be: answer "Yes" regardless of the person's actual opinion

space, it will hence be convenient to represent $D \in \mathscr{X}^n$ as a *histogram* $D \in \mathbb{N}^{\mathscr{X}} = \mathbb{N}^{|\mathscr{X}|}$ Namely, if $\mathscr{X} = \{v_1, \dots, v_K\}$, then for all $k \in \{1, \dots, K\}$,

$$D_k := |\{x_i \in D \mid x = v_k\}|.$$

In particular, the size of the dataset is $n = \|D\|_1 = \sum_{k=1}^K D_k$. With this notation, the ℓ^1 norm also allows to define neighboring datasets.

Definition 7.3 (Neighboring datasets)**.** *Two datasets $D, D' \in \mathbb{N}^{\mathscr{X}}$ are said to be* neighboring *if they differ by at most one element, i.e. $\|D - D'\|_1 \leq 1$.*

With this definition, two neighboring datasets only differ by either the addition or the removal of an individual's data. Similarly, changing value x_i into $x_i' \neq x_i$ yields datasets with $\|D - D'\|_1 = 2$.

7.2.2 Randomized algorithm and differential privacy

The seminal paper [KLN+11] on private learning introduces a learning framework inspired by differentially private algorithms [DMNS06]. Given samples $\{x_1, \dots, x_n\}$, this constraint imposes privacy to a learner by requiring it not to be significantly affected if a particular sample x_i is removed (see Definition 7.5).

Formally, to make sure that information about individuals are not disclosed, the statistician will only access information on this data through a so-called randomized algorithm.

Definition 7.4 (Randomized algorithm)**.** *A* randomized algorithm *is a map $\mathscr{A} : \mathbb{N}^{|\mathscr{X}|} \to \mathscr{O}$, where $\mathscr{O}$ is a probability space.*

Said otherwise, a randomized algorithm $\mathscr{A} : \mathbb{N}^{|\mathscr{X}|} \to \mathscr{O}$ defines a $\mathscr{O}$-valued random variable $\mathscr{A}(D)$ for all (fixed) $D \in \mathbb{N}^{|\mathscr{X}|}$. By definition, we measure the degree of differential privacy of an algorithm by is its stochastic sensitivity to its input data D[3].

Definition 7.5 (Differentially private (DP) algorithm)**.** *For $\varepsilon > 0$ and $0 < \delta < 1$, we say that $\mathscr{A}$ is (ε, δ)-differentially private if for all $D, D' \in \mathbb{N}^{|\mathscr{X}|}$ such that $\|D - D'\|_1 \leq 1$ and all measurable $S \subset \mathscr{O}$,*

$$\mathbb{P}(\mathscr{A}(D) \in S) \leq e^{\varepsilon} \mathbb{P}(\mathscr{A}(D') \in S) + \delta,$$

where the probability is taken with respect to randomness of $\mathscr{A}$.

In contrast to anonymization, we insist on the fact that differential privacy is a property *of the data analysis pipeline*, and not a property of a particular output. In the above definition, data are considered fixed and not random. In fact, the algorithm $\mathscr{A}$ can be made public, with only the randomness used to generate it needing to be kept secret. This underscores a critical aspect of contemporary security, rejecting the outdated concept of "security by obscurity". This feature also facilitates open discussions about the algorithms and their guarantees.

Remark 7.6. *$(\varepsilon, 0)$-DP is often called* pure *ε-DP. It guarantees that at each independent run of $\mathscr{A}(D)$, the output is almost equally likely to be observed than for any neighboring dataset D'. In practice, $\varepsilon = 1$ ($e^1 \simeq 2.7$) is considered reasonable, and $\varepsilon = 0.1$ ($e^{0.1} \simeq 1.1$) is considered to yield strong privacy guarantees.*

[3] For infinite space of realizations $\mathscr{X}$, defining differential privacy requires conditional distributions, through the notion of Markov transition kernel, which we chose not to cover for sake of simplicity

If $\mathscr{O}$ is finite, the log-likelihood ratio

$$L_{\mathscr{A}(D),\mathscr{A}(D')}(\theta) := \log\left(\frac{\mathbb{P}(\mathscr{A}(D)=\theta)}{\mathbb{P}(\mathscr{A}(D')=\theta)}\right)$$

is called *privacy loss*. To satisfy (ε,δ)-differential privacy, it is sufficient that for all $\theta \in \mathscr{O}$ and all $\|D - D'\|_1 \leq 1$,

$$\mathbb{P}_{\theta\sim\mathscr{A}(D)}\left(L_{\mathscr{A}(D),\mathscr{A}(D')}(\theta) \leq \varepsilon\right) \geq 1-\delta.$$

A priori, differential privacy does not provide an assurance that an individual's sensitive datum x_i will remain undisclosed. However, it safeguards — in a quantified manner — against the disclosure of one's participation in a survey and prevents the revelation of any specific contributions made to the survey.

Following the economic view of [DR+14, Section 2.3.1], suppose that an individual $i \in \{1,\dots,n\}$ has a utility function $u : \mathscr{O} \to \mathbb{R}$ defined based on the outcome of an (ε,δ)-DP algorithm $\mathscr{A} : \mathbb{N}^{|\mathscr{X}|} \to \mathscr{O}$. If D is the dataset including their individual data and D_{-i} is the same dataset with their individual data removed, then

$$e^{-\varepsilon}\mathbb{E}[u(\mathscr{A}(D_{-i}))] - \delta\|u\|_\infty \leq \mathbb{E}[u(\mathscr{A}(D))] \leq e^{\varepsilon}\mathbb{E}[u(\mathscr{A}(D_{-i}))] + \delta\|u\|_\infty.$$

Hence, the expected utility of any user i is affected by a factor of at most $e^{\pm\varepsilon} \simeq 1 \pm \varepsilon$, when participating (or not) in a differentially private release. This reasoning also applies to censoring/not censoring one's datum, and it works regardless of the utility function $u : \mathscr{O} \to \mathbb{R}_+$.

7.2.3 Structure

Differential privacy fulfills desirable structural properties which we now detail. First, it is robust against post-processing: without additional knowledge about the private database, one cannot manipulate the output of a private algorithm $\mathscr{A}(D)$ to compromise its level of differential privacy.

Proposition 7.7 (Postprocessing)**.** *Let $\mathscr{A} : \mathbb{N}^{|\mathscr{X}|} \to \mathscr{O}$ be an (ε,δ)-differentially private algorithm, and $f : \mathscr{O} \to \mathscr{O}'$ be a (randomized) measurable function (independent from $\mathscr{A}$). Then $f \circ \mathscr{A} : \mathbb{N}^{|\mathscr{X}|} \to \mathscr{O}'$ is (ε,δ)-differentially private.*

Independence of f and $\mathscr{A}$ is crucial. Indeed, if $\mathscr{A}$ is one-to-one (says $\mathscr{A}(D) = D + Z$ with random Z), then $f = \mathscr{A}^{-1}$ yields a non differentially private composition $f \circ \mathscr{A} = \mathrm{Id}$.

Proof. Let $\|D - D'\|_1 \leq 1$ and $S' \subset \mathscr{O}'$ be measurable. Write $S := f^{-1}(S')$. Because $\mathscr{A}$ is (ε,δ)-DP and $f \perp\!\!\!\perp \mathscr{A}$ we have almost surely that

$$\begin{aligned}\mathbb{P}(f(\mathscr{A}(D)) \in S' \mid f) &= \mathbb{P}(\mathscr{A}(D) \in S \mid f)\\ &\leq e^{\varepsilon}\mathbb{P}(\mathscr{A}(D') \in S \mid f) + \delta\\ &= e^{\varepsilon}\mathbb{P}(f(\mathscr{A}(D')) \in S' \mid f) + \delta.\end{aligned}$$

Taking the expectation on both sides of the above bound yields the result. □

Similarly one may easily control the degree of differential privacy when several analyses of the same dataset are released. With the chosen notation, general sequences of algorithms accumulate privacy costs additively.

	Female	Male
Lives in Paris	13	46
Doesn't live in Paris	30	25

Table 7.1: Entry-wise differentially private contingency table

Proposition 7.8 (Simple composition)**.** *If $\mathscr{A}_1,\dots,\mathscr{A}_J$ are independent (ε_j,δ_j)-differentially private algorithms, then*

$$\mathscr{A}(D) := (\mathscr{A}_1(D),\dots,\mathscr{A}_J(D))$$

is (ε,δ)-differentially private, with

$$\varepsilon := \sum_{j=1}^J \varepsilon_j \quad \text{and} \quad \delta := \sum_{j=1}^J \delta_j.$$

Proof. The result is trivial for $\delta = 0$. See [DR+14, Theorem B.1] for the general case. □

For more advanced composition theorems, see [DR+14, Section 3.1]. Let us mention the following simple one.

Exercise 7.9 (Parallel composition)**.** *If $\mathscr{A}_1(D_1),\dots,\mathscr{A}_J(D_J)$ are independent (ε_j,δ_j)-differentially private algorithms on different datasets, show that*

$$\mathscr{A}(D_1,\dots,D_J) := (\mathscr{A}_1(D_1),\dots,\mathscr{A}_J(D_J))$$

is $(\max_{j\le J}\varepsilon_j, \max_{j\le J}\delta_j)$-differentially private. As an example, suppose that one surveys a population, and produces contingency Table 7.1. If each entry is (ε,δ)-DP, then any pair of entries is also (ε,δ)-DP. However, triples and the quadruplet might not be, as they are built on overlapping populations.

With the above definition of privacy given at the individual level, one may be interested in its consequences on groups of individuals. This question becomes particularly critical if N individuals have highly correlated data, or that a single individual contributes N times to the dataset.

Proposition 7.10 (Group-differential privacy)**.** *An (ε,δ)-differentially private algorithm is $(N\varepsilon, Ne^{N\varepsilon}\delta)$-diffentially private for groups of size N. That is, for all $D,D' \in \mathbb{N}^{|\mathscr{X}|}$ such that $\|D-D'\|_1 \le N$ and all measurable $S \subset \mathscr{O}$,*

$$\mathbb{P}\big(\mathscr{A}(D)\in S\big) \le e^{N\varepsilon}\mathbb{P}\big(\mathscr{A}(D')\in S\big) + Ne^{N\varepsilon}\delta.$$

Note that this is a completely distinct result from that on the stability under composition (Proposition 7.8).

Proof. Let $D =: D_0, D_1, \dots, D_N := D'$ be such that $\|D_{i+1} - D_i\|_1 \le 1$ for all $i \in \{0,\dots,N-1\}$. For all measurable $S \subset \mathscr{O}$, we apply the definition sequentially to get

$$\begin{aligned}
\mathbb{P}\big(\mathscr{A}(D_0)\in S\big) &\le e^{\varepsilon}\mathbb{P}\big(\mathscr{A}(D_1)\in S\big) + \delta \\
&\le e^{\varepsilon}\Big(e^{\varepsilon}\mathbb{P}\big(\mathscr{A}(D_2)\in S\big) + \delta\Big) + \delta \\
&\vdots \\
&\le e^{N\varepsilon}\mathbb{P}\big(\mathscr{A}(D_N)\in S\big) + \big(1 + e^{\varepsilon} + e^{2\varepsilon} + \dots + e^{(N-1)\varepsilon}\big)\delta \\
&\le e^{N\varepsilon}\mathbb{P}\big(\mathscr{A}(D')\in S\big) + Ne^{N\varepsilon}\delta,
\end{aligned}$$

where the last line comes from the AM–GM inequality $(u_0 + \dots + u_{N-1})/N \le (u_0 \dots u_{N-1})^{1/N}$. □

7.2.4 General privacy mechanisms

In this section, we will detail a few basic ways to *privatise* a function. That is, suppose that we want to "evaluate" a function of interest $f : \mathbb{N}^{|\mathscr{X}|} \to \mathscr{O}$ on confidential data D. We will exhibit explicit constructions of $\mathscr{A}(D)$ that are provably differentially private, while still close to the actual value $f(D)$.

Laplace mechanism

We first deal with the case where $f : \mathbb{N}^{|\mathscr{X}|} \to \mathbb{R}^d$ is vector-valued.

Definition 7.11 (Global ℓ^1 sensitivity)**.** *The global ℓ^1 sensitivity of $f : \mathbb{N}^{|\mathscr{X}|} \to \mathbb{R}^d$ is defined as*

$$\Delta_1(f) := \max_{\|D-D'\|_1 \leq 1} \|f(D) - f(D')\|_1.$$

To be sure, the first $\|\cdot\|_1$ norm is on histograms $\mathbb{N}^{|\mathscr{X}|}$, and the second one is on $\mathbb{R}^d$. The quantity $\Delta_1(f)$ measures how much f is affected by a change of a single value in the dataset. It naturally yields a minimal order of magnitude required to mask the contribution of an individual to $f(D)$.

Example 7.12.

- *If $f(D) :=$ #{inhabitants of Paris in D}, then $\Delta_1(f) = 1$;*
- *If $f(D) :=$ average salary, then $\Delta_1(f) \leq$ (maximum salary)$/n$.*

Given $b > 0$, we recall that the *Laplace distribution* Laplace(b) is the distribution on $\mathbb{R}$ that has density

$$p(y, b) := \frac{1}{2b} e^{-|y|/b}, \quad y \in \mathbb{R}$$

with respect to the Lebesgue measure. If $Y \sim \text{Laplace}(b)$, then

- $\mathbb{E}[Y] = 0$, $\mathbb{E}[|Y|] = b$, $\mathbb{E}[Y^2] = 2b^2$
- For all $t \geq 0$, $\mathbb{P}(|Y| \geq tb) \leq e^{-t}$.

The Laplace distribution is the building block of the eponymous privatisation mechanism.

Definition 7.13 (Laplace mechanism)**.** *For $f : \mathbb{N}^{|\mathscr{X}|} \to \mathbb{R}^d$ and $\varepsilon > 0$, we write*

$$\mathscr{A}_{\text{Laplace}}(D, f, \varepsilon) := f(D) + Y,$$

where $Y = (Y_1, \ldots, Y_d)$ is a iid sequence of random variables $Y_j \sim \text{Laplace}\big(\Delta_1(f)/\varepsilon\big)$

Here, the idea is simply to perturb the output $f(D)$ by adding random noise coordinate-wise, at a scale given by the ℓ^1 sensitivity and a target level of privacy. In practice, it requires to compute $\Delta_1(f)$, or an upper bound on it.

Proposition 7.14. *For all $\varepsilon > 0$, $\mathscr{A}_{\text{Laplace}}(\cdot, f, \varepsilon)$ is $(\varepsilon, 0)$-differentially private.*

Proof. Let $\|D - D'\|_1 \leq 1$ and $S \subset \mathbb{R}^d$ be measurable of non-empty interior, and write $b := \Delta_1(f)/\varepsilon$ for short. The random variables $\mathscr{A}_{\text{Laplace}}(D, f, \varepsilon)$ and $\mathscr{A}_{\text{Laplace}}(D', f, \varepsilon)$ have densities with respect to the Lebesgue measure in $\mathbb{R}^d$. Denoting them by g and g' respectively, we have

$$\begin{aligned} \frac{\mathbb{P}\big(\mathscr{A}_{\text{Laplace}}(D, f, \varepsilon) \in S\big)}{\mathbb{P}\big(\mathscr{A}_{\text{Laplace}}(D, f, \varepsilon) \in S\big)} &= \frac{\int_S g(y)\,dy}{\int_S g'(y)\,dy} \\ &\leq \sup_{y \in S} \frac{g(y)}{g'(y)}, \end{aligned}$$

where the last inequality if Hölder inequality. Furthermore, by construction of $\mathscr{A}_{\text{Laplace}}(\cdot, f, \varepsilon)$, for all $y \in \mathbb{R}^d$,

$$\begin{aligned} g(y) &= \prod_{j=1}^{d} \frac{1}{2b} e^{-|y_j - f_j(D)|/b} \\ &= \frac{1}{(2b)^d} e^{-\|y - f(D)\|_1 / b}, \end{aligned}$$

and similarly for $g'(y)$. As a result, triangle inequality for ℓ^1 norm yields

$$\begin{aligned} \frac{g(y)}{g'(y)} &= \exp\big(-(\|y - f(D)\|_1 - \|y - f(D')\|_1)/b\big) \\ &\leq \exp\big(\|f(D) - f(D')\|_1 / b\big). \end{aligned}$$

Since $b \geq \Delta_1(f)/\varepsilon$, the term in the exponential is further bounded by $\|f(D) - f(D')\|_1 / b \leq \varepsilon$, which yields the result. □

Naturally, pure random output or constant $\mathscr{A}$ would also lead to a differentially private algorithm. As opposed to such trivial examples, the Laplace mechanism has the extra property of staying close to f with high probability. In the field, such a property is called *utility*.

Proposition 7.15 (Utility of the Laplace mechanism)**.** *For all $\beta \in (0, 1]$,*

$$\mathbb{P}\left(\|\mathscr{A}_{\text{Laplace}}(D, f, \varepsilon) - f(D)\|_\infty \leq \log\big(d/\beta\big) \frac{\Delta_1(f)}{\varepsilon} \right) \geq 1 - \beta,$$

and

$$\mathbb{E}\left[\|\mathscr{A}_{\text{Laplace}}(D, f, \varepsilon) - f(D)\|_1 \right] = d \frac{\Delta_1(f)}{\varepsilon}$$

Proof. The first bound comes directly from the tail bound of Laplace distribution, and the bound in expectation from $\mathbb{E}[|Y|] = b$ when $Y \sim \text{Laplace}(b)$. □

Note the tradeoff between the level of privacy ε and the subsequent output precision of order $O(\Delta_1(f)/\varepsilon)$. Theory of differential privacy actually is all about finding the best privacy parameter ε given a target precision or conversely, finding the most precise mechanism under ε-DP constraint.

Gaussian mechanism

It is sometimes more convenient to handle Gaussian perturbations than Laplace ones. For instance, when data noise is itself Gaussian, one may use the additive stability of the Gaussian distribution to understand finely how privacy and statistical precision interact. Because of the form of its density, Gaussian perturbations blend well with a ℓ^2-type sensitivity.

Definition 7.16 (Global ℓ^2 sensitivity)**.** *The global ℓ^2 sensitivity of $f : \mathbb{N}^{|\mathscr{X}|} \to \mathbb{R}^d$ is defined as*

$$\Delta_2(f) := \max_{\|D - D'\|_1 \leq 1} \|f(D) - f(D')\|_2.$$

Paralleling the Laplace mechanism, the Gaussian privatisation mechanism has a straightforward definition, but with a twist.

Definition 7.17 (Gaussian mechanism)**.** *For $f : \mathbb{N}^{|\mathscr{X}|} \to \mathbb{R}^d$, $\varepsilon > 0$ and $\delta > 0$, we write*

$$\mathscr{A}_{\text{Gaussian}}(D, f, \varepsilon, \delta) := f(D) + Y,$$

where $Y = (Y_1, \ldots, Y_d)$ is a iid sequence of random variables $Y_j \sim \mathcal{N}\left(0, \sigma^2\right)$, with

$$\sigma := \sqrt{2\log(1.25/\delta)}\frac{\Delta_2(f)}{\varepsilon}.$$

Parameter $\delta > 0$ is the price to pay for handling Gaussian noise, and yields a provable (ε, δ)-differential privacy mechanism.

Proposition 7.18. *For all $\varepsilon > 0$ and $\delta > 0$, $\mathscr{A}_{\text{Gaussian}}(\cdot, f, \varepsilon, \delta)$ is (ε, δ)-differentially private.*

Proof. See [DR+14, Appendix A] □

When $\delta \ll 1$, the difference between ε-DP and (ε, δ)-DP is considered not significant in practice.

Proposition 7.19 (Utility of the Gaussian mechanism)**.** *For all $\beta \in (0, 1]$,*

$$\mathbb{P}\left(\|\mathscr{A}_{\text{Gaussian}}(D, f, \varepsilon, \delta) - f(D)\|_\infty \leq \sqrt{2\log(1.25/\delta)\log\left(d/\beta\right)}\frac{\Delta_1(f)}{\varepsilon}\right) \geq 1 - \beta,$$

and

$$\mathbb{E}\left[\|\mathscr{A}_{\text{Gaussian}}(D, f, \varepsilon, \delta) - f(D)\|_2\right] \leq \sqrt{2d\log(1.25/\delta)d}\frac{\Delta_2(f)}{\varepsilon}.$$

This result is the opportunity to note another interesting feature of the Gaussian mechanism: the Gaussian tails $O(\sqrt{\log(1/\delta)})$ instead of exponential ones in $O(\log(1/\delta))$ for the Laplace mechanism.

Remark 7.20 (Integer-valued mechanisms)**.** *The above mechanisms apply to any general function $f : \mathbb{N}^{|\mathscr{X}|} \to \mathbb{R}^d$, and produce random outputs having support in the entire continuous space $\mathbb{R}^d$. If the initial function is discrete, say $f : \mathbb{N}^{|\mathscr{X}|} \to \mathbb{N}$, then one may want to preserve this qualitative property after privatization. Specific mechanisms are designed for this, the main one using truncated geometric distributions [GRS09].*

Exponential mechanism

The Laplace and Gaussian mechanisms are specifically designed for numeric (vector-valued) functions, on which norms yield natural precision criteria. They are precise when $f(D)$ is regular enough in its variable D, in the sense that $\Delta_1(f)$ or $\Delta_2(f)$ are small enough.

When the output space $\mathscr{O}$ is finite and unstructured, the user needs to choose a *score function*

$$s : \mathbb{N}^{|\mathscr{X}|} \times \mathscr{O} \to \mathbb{R},$$

where $s(D, \theta)$ represents how satisfactory it is to return output θ if the true value is $f(D)$ is queried. Score function s should hence be thought of as depending on f, with $\theta = f(D)$ yielding a maximal value for $\theta \mapsto s(\theta, D)$.

Definition 7.21 (Sensitivity of score function)**.** *The sensitivity of a score function $s : \mathbb{N}^{|\mathscr{X}|} \times \mathscr{O} \to \mathbb{R}$ is*

$$\Delta(s) := \max_{\theta \in \mathscr{O}} \max_{\|D - D'\|_1 \leq 1} |s(D, \theta) - s(D, \theta')|.$$

This notion actually is a generalization of the ℓ^1 sensitivity seen above.

Example 7.22. *If* $f : \mathbb{N}^{|\mathscr{X}|} \to \mathbb{R}^d$ *and* $s(D,\theta) := -\|\theta - f(D)\|_p$ *(*$p \in \{1,2\}$*), then* $\Delta(s) = \Delta_p(f)$.

As announced, we are especially interested in the case where $\mathscr{O}$ is finite and unstructured. In this case, the idea of the *exponential mechanism* is to randomly output values $\theta \in \mathscr{O}$ non-uniformly, and to upweight those values θ with high associated scores $s(D,\theta)$.

Definition 7.23 (Exponential mechanism)**.** *For* $\varepsilon > 0$ *and score function* $s : \mathbb{N}^{|\mathscr{X}|} \times \mathscr{O} \to \mathbb{R}$*, we write* $\mathscr{A}_{\exp}(D, s, \varepsilon)$ *for a random variable such that for all* $\theta \in \mathscr{O}$*,*

$$\mathbb{P}(\mathscr{A}_{\exp}(D,s,\varepsilon) = \theta) = \frac{\exp\left(\frac{s(D,\theta)\varepsilon}{2\Delta(s)}\right)}{\sum_{\theta' \in \mathscr{O}} \exp\left(\frac{s(D,\theta')\varepsilon}{2\Delta(s)}\right)}.$$

Note that here, the dependence in $f : \mathbb{N}^{|\mathscr{X}|} \to \mathscr{O}$ is completely implicit: it only appears through the preliminary choice of the score function $s : \mathbb{N}^{|\mathscr{X}|} \times \mathscr{O} \to \mathbb{R}$.

Proposition 7.24. *For all* $\varepsilon > 0$*,* $\mathscr{A}_{\exp}(\cdot, s, \varepsilon)$ *is* $(\varepsilon, 0)$*-differentially private.*

Exercise 7.25. *Prove Proposition 7.24.*

Because of its design involving the exponential of the score function, the exponential mechanism has strong utility guarantees. Indeed, for $D \in \mathbb{N}^{\mathscr{X}}$ fixed, an output $\theta \in \mathscr{O}$ with score value $s(D,\theta)$ lower than

$$s^*(D) := \max_{\theta \in \mathscr{O}} s(D,\theta)$$

will be down-weighted exponentially compared to any output from

$$\mathscr{O}^*(D) := \operatorname*{argmax}_{\theta \in \mathscr{O}} s(D,\theta)$$

maximizing the score function with D.

Proposition 7.26 (Utility of the exponential mechanism)**.** *For all* $\beta \in (0,1]$*,*

$$\mathbb{P}\left(s\big(\mathscr{A}_{\exp}(D,s,\varepsilon)\big) \geq s^*(D) - 2\log\left(\frac{|\mathscr{O}|}{\beta|\mathscr{O}^*(D)|}\right)\frac{\Delta(s)}{\varepsilon}\right) \geq 1-\beta.$$

With high probability, this result asserts that $\mathscr{A}_{\exp}(D,s,\varepsilon)$ yields nearly best score up to $O(\Delta(s)/\varepsilon)$, with a prefactor if $|\mathscr{O}^*(D)|$ is large.

Proof. See [DR+14, Theorem 3.11]. □

7.3 Other types of privacy formalism

7.3.1 Local differential privacy

Definition

In scenarios where a trustworthy third party is not available for doing the data analysis, one cannot use the formalism of differential privacy. To guarantee confidentiality in such cases, we impose privacy at the level of the individuals themselves.

Definition 7.27 (Local randomizer)**.** *A* local randomizer *is a randomized function* $\mathscr{R} : \mathscr{X} \to \mathscr{Z}$*.*

From there, a local version of differential privacy comes naturally, as the $n = 1$ sample version of differential privacy.

Definition 7.28 (Locally differentially private (LDP) randomizer)**.** *A local randomizer $\mathscr{R} : \mathscr{X} \to \mathscr{Z}$ is said to be* (ε, δ)-locally differentially private *if for all $x, x' \in \mathscr{X}$ and all measurable $Z \subset \mathscr{Z}$,*

$$\mathbb{P}(\mathscr{R}(x) \in Z) \leq e^{\varepsilon}\mathbb{P}(\mathscr{R}(x') \in Z) + \delta$$

To come back to a class of global algorithms taking whole datasets into account, we will impose dependency in such LDP randomizers as input.

Definition 7.29 (Locally differentially private (LDP) randomizer)**.** *An algorithm $\mathscr{A} : \mathbb{N}^{\mathscr{Z}} \to \mathscr{O}$ is said to be* (ε, δ)-locally differentially private*, if it can be written as*

$$\mathscr{A}(z_1, \ldots, z_n) = \mathscr{A}(\mathscr{R}_1(x_1), \ldots, \mathscr{R}_n(x_n)),$$

where $\mathscr{R}_1, \ldots, \mathscr{R}_n : \mathscr{X} \to \mathscr{Z}$ are (ε, δ)-LDP randomizers.

Exercise 7.30. *Show that the mechanism of Exercise 7.1 is LDP. With what privacy parameter?*

Trivially, DP and LDP are equivalent for datasets of size $n = 1$. Furthermore, LDP algorithms are also DP.

Proposition 7.31 (LDP $\Rightarrow$ DP)**.** *An (ε, δ)-LDP algorithm is (ε, δ)-DP.*

Proof. Prove Proposition 7.31 using the structural results of Section 7.2.3. □

LDP vs DP

The converse of Proposition 7.31 is false. That is, limiting oneself to LDP algorithms only is strictly more restrictive that doing so with DP algorithms. When using privacy in statistical settings, this can lead to significant differences in terms of convergence rates. To exemplify this, let us first give a simple example of an LDP mechanism generalizing heads/tails draws for more than binary output. We recall that $\mathscr{X} = \{v_1, \ldots, v_K\}$ is finite.

Definition 7.32 (K-ary randomized response)**.** *Let $B \sim \text{Bernoulli}\left(\frac{K}{K+(e^{\varepsilon}-1)}\right)$ and $X \sim \text{Uniform}(\mathscr{X})$. For all $\varepsilon > 0$ and $x \in \mathscr{X}$,*

$$\mathscr{R}_{\text{RR}}(x, \varepsilon) := \begin{cases} x & \text{if } B = 0, \\ X & \text{if } B = 1. \end{cases}$$

As $\varepsilon \to 0$, K-ary randomized response outputs the true value with probability of order $\varepsilon/(K+\varepsilon)$, and a random value with probability of order $K/(K+\varepsilon)$.

Proposition 7.33. *$\mathscr{R}_{\text{RR}}(\cdot, \varepsilon)$ is ε-LDP.*

Proof. Let $x, x', v \in \mathscr{X}$. If either $x \neq v$ and $x' \neq v$, or $x = x' = v$, then $\mathbb{P}(\mathscr{R}_{\text{RR}}(x, \varepsilon) = v) = \mathbb{P}(\mathscr{R}_{\text{RR}}(x', \varepsilon) = v)$. Otherwise, assume that $x = v$ and $x' \neq v$ without loss of generality. Then we have

$$\begin{aligned} \mathbb{P}(\mathscr{R}_{\text{RR}}(x, \varepsilon) = v) &= \left(1 - \frac{K}{K + (e^{\varepsilon} - 1)}\right) + \frac{K}{K + (e^{\varepsilon} - 1)} \frac{1}{K} = \frac{e^{\varepsilon}}{K + (e^{\varepsilon} - 1)} \\ \mathbb{P}(\mathscr{R}_{\text{RR}}(x', \varepsilon) = v) &= \frac{K}{K + (e^{\varepsilon} - 1)} \frac{1}{K} = \frac{1}{K + (e^{\varepsilon} - 1)}. \end{aligned}$$

As the ratio of these probabilities is always between $e^{-\varepsilon}$ and e^{ε}, we get the result. □

Example 7.34 (LDP vs DP for histograms)**.** *Consider the problem of publishing the whole dataset $f(D) = h$ in the form of its renormalized histogram*

$$h_k = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{x_i = v_k} = \frac{D_k}{n},$$

for all $k \in \{1, \dots, K\}$.

- *(LDP) Writing $p_0 := 1/(K + (e^\varepsilon - 1))$, independent K-randomized responses $z_i = \mathscr{R}_{\mathrm{RR}}(x_i, \varepsilon)$ allow to construct an unbiased estimator $\widehat{h}^{(\mathrm{LDP})}$ of h. Indeed, for all $k \in \{1, \dots, K\}$, we have*

$$\mathbb{P}(z_i = v_k) = \begin{cases} p_0 e^\varepsilon & \text{if } x_i = v_k, \\ p_0 & \text{otherwise.} \end{cases}$$

  As a result,

$$N_k := \sum_{i=1}^{n} \mathbb{1}_{z_i = v_k} \sim \mathrm{Binomial}(n h_k, p_0 e^\varepsilon) * \mathrm{Binomial}(n(1 - h_k), p_0),$$

  which has :

  - *Mean* $\mathbb{E}[N_k] = n h_k p_0 e^\varepsilon + n(1 - h_k) p_0 = n p_0 (e^\varepsilon - 1) h_k + n p_0$*;*
  - *Variance* $\mathrm{Var}(N_k) = n h_k p_0 e^\varepsilon (1 - p_0 e^\varepsilon) + n(1 - h_k) p_0 (1 - p_0) \le n p_0 e^\varepsilon (1 - p_0)$.

  As a result,

$$\widehat{h}_k^{(\mathrm{LDP})} := \frac{\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{z_i = v_k} - p_0}{p_0(e^\varepsilon - 1)}$$

  is an unbiased estimator of h_k, with mean squared loss

$$\mathbb{E}[(h_k - \widehat{h}_k^{(\mathrm{LDP})})^2] \le \frac{(1 - p_0) e^\varepsilon}{n p_0 (e^\varepsilon - 1)^2} \underset{\varepsilon \to 0}{\sim} \frac{K(1 - 1/K)}{n\varepsilon^2}.$$

- *(DP) The ℓ^1 sensitivity of $f(x_1, \dots, x_n) = \left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{x_i = v_k}\right)_{1 \le k \le K}$ is $\Delta_1(f) = 1/n$. As a result, the Laplace mechanism ($b = 2/(n\varepsilon^2)$) yields an ε-DP algorithm with output $\widehat{h}_k^{(\mathrm{LDP})}$ such that*

$$\mathbb{E}[(h_K - \widehat{h}_k^{(\mathrm{DP})})^2] \lesssim \frac{1}{n^2 \varepsilon^2}.$$

We observe here a significant discrepancy in the utility, when comparing ε-LDP and ε-DP algorithms : a factor $1/n$ is lost. This gap is known to be unavoidable for functions involving averaging [CSS12], hence restricting the range of application of LDP to large samples.

7.3.2 Statistical queries

First introduced by Kearns [Kea98], the statistical query (SQ) framework is a restriction of PAC-learning, where the learner is only allowed to obtain approximate averages of the unknown distribution P via an adversarial oracle, but cannot see any sample. That is, given a tolerance parameter $\tau > 0$, a $\mathrm{STAT}(\tau)$ oracle for the distribution P accepts functions $r : \mathbb{R}^d \to [-1, 1]$ as queries from the learner, and can

answer *any* value $a \in \mathbb{R}$ such that $|\mathbb{E}_{X\sim P}[r(X)] - a| \leq \tau$. Informally, the fact that the oracle is adversarial is the counterpart to the fact that differential privacy is a worst-case notion. We emphasize that in the statistical query framework, estimators (or learners) are only given access to such an oracle, and not to the data themselves. Limiting the learner's accessible information to adversarially perturbed averages both restricts the range of the usable algorithms, and effectively forces them to be robust and efficient.

We do not give a fully fledged definition of a statistical query algorithm (see [AK22, Definition 2.1]). Informally, a "statistical query algorithm with tolerance τ and making T queries" is an interactive algorithm that only requires answers to the functional queries defined above, which is robust to adversarial error τ on those answers, and that terminates after at most T queries.

Naturally, if actual sample is available, one may always simulate an oracle through empirical averages. As a result, the SQ framework is directly linked with LDP in following manner.

Proposition 7.35 (SQ $\Rightarrow$ LDP)**.** *If $\mathscr{A}_{\mathrm{SQ}}$ is a statistical query algorithm that makes at most T queries to a $\mathrm{STAT}(\tau)$ oracle, then there exists an ε-LDP algorithm simulating $\mathscr{A}_{\mathrm{SQ}}$ on $n \asymp T/(\varepsilon^2\tau^2)$ samples, which terminates with high probability.*

Proof. See [KLN+11, Theorem 5.7] for the full proof. Given a sample of size $n \asymp T/(\varepsilon^2\tau^2)$, the idea is to divide it into T sub-samples of size $1/(\varepsilon^2\tau^2)$. Then, for query $r_t : \mathbb{R}^d \to [-1,1]$ ($t \in \{1,\dots,T\}$), take answer $\hat{a}_t$ to be the empirical average of r_t on the samples of batch t to which the Laplace mechanism has been applied. By Hoeffding inequality, these empirical averages yield a valid $\mathrm{STAT}(\tau)$ oracle altogether, which yields the result. □

Let us conclude on an illustrative to example, to sum up all the privacy mechanisms on a specific inference problem.

Exercise 7.36 (Standard inference vs DP vs LDP vs SQ)**.** *Consider the statistical model composed of all the uniform distributions on intervals $[0,\theta]$, for $\theta \in (0,1]$ unknown. Assume classically that we want to estimate θ, but while making sure that the proposed estimator falls into the different scenarios described in this chapter. When sample is available, assume that it is iid, denoted by $X_1,\dots,X_n \sim P_\theta := \mathrm{Unif}([0,\theta])$.*

- *(PAC) A classical non-private estimator of θ is*

$$\hat{\theta}_{\mathrm{PAC}} := \max_{1\leq i\leq n} X_i.$$

  Twice the empirical mean would also lead to another (suboptimal) estimator).

- *(DP) Function $f : [0,1]^n \ni (x_1,\dots,x_n) \mapsto \max_{1\leq i\leq n} x_i$ ε-DP has ℓ^1 sensitivity $\Delta_1(f) = 1$. Hence, the Laplace mechanism yields a DP estimator*

$$\hat{\theta}_{\mathrm{DP}} := \left(\max_{1\leq i\leq n} X_i\right) + Y.$$

  A similar DP version of the empirical average can be written easily.

- *(LDP) Applied to each datum, the Laplace mechanism yields a LDP estimator*

$$\hat{\theta}_{\mathrm{LDP}} := \left(\max_{1\leq i\leq n} X_i + Y_i\right).$$

  A similar LDP version of the empirical average can be written easily.

- *(SQ) As $\theta = \min\{u \in [0,1] \mid P_\theta([u,1]) = 0\}$, one may estimate it with a divide and conquer strategy which queries indicator functions $r_u(x) = \mathbb{1}_{[u,1]}(x)$ at each step. Starting from $u = 1/2$, next query would be either $1/4$ or $3/4$ depending on whether the answer to $r_{1/2}$ is smaller or greater than the (known) tolerance parameter τ. The single query $r(x) = x$ would also lead to (a suboptimal) estimator.*

Bibliography

[AK22] Eddie Aamari and Alexander Knop. Adversarial manifold estimation. *Foundations of Computational Mathematics*, pages 1–97, 2022.

[CSS12] TH Hubert Chan, Elaine Shi, and Dawn Song. Optimal lower bound for differentially private multi-party aggregation. In *European Symposium on Algorithms*, pages 277–288. Springer, 2012.

[DMNS06] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography*, volume 3876 of *Lecture Notes in Comput. Sci.*, pages 265–284. Springer, Berlin, 2006.

[DR+14] Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

[GRS09] Arpita Ghosh, Tim Roughgarden, and Mukund Sundararajan. Universally utility-maximizing privacy mechanisms. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 351–360, 2009.

[Kea98] Michael J. Kearns. Efficient noise-tolerant learning from statistical queries. *J. ACM*, 45(6):983–1006, 1998.

[KLN+11] Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam D. Smith. What can we learn privately? *SIAM J. Comput.*, 40(3):793–826, 2011.

[NHF16] Arvind Narayanan, Joanna Huey, and Edward W. Felten. *A Precautionary Approach to Big Data Privacy*, pages 357–385. Springer Netherlands, Dordrecht, 2016.

[Vad17] Salil Vadhan. The complexity of differential privacy. In *Tutorials on the foundations of cryptography*, Inf. Secur. Cryptography, pages 347–450. Springer, Cham, 2017.

[Val84] Leslie G. Valiant. A theory of the learnable. *Commun. ACM*, 27(11):1134–1142, 1984.