Sluch Dmitrii
dmitrybsluch@gmail.com

DPCH (Python Project)

Ariel, HDP
August 16, 2025

1 Introduction

Differential privacy addresses the challenge of providing analytical access to datasets while maintaining individual privacy guarantees. Traditional anonymization techniques, which involve removing identifying information and data shuffling, have proven inadequate. Statistical analysis methods, when applied to publicly available or leaked databases, can effectively de-anonymize such data, linking it back to specific individuals.

The differential privacy framework offers a robust solution to this challenge. It implements a controlled query execution mechanism where a server processes database queries while introducing carefully calibrated noise to the results. This approach maintains precise probabilistic invariants ensuring that aggregate query results remain statistically stable regardless of individual data points. The framework rejects queries that could potentially compromise privacy guarantees.

This project encompasses three main components:

1. Exploring the theoretical foundations of differential privacy, examining the underlying probability theory concepts and key mathematical results
2. Creating a Python-based implementation that provides differential privacy features for the ClickHouse database system (DPCH stands for Differential Privacy for ClickHouse)
3. Conducting experimental analysis by executing queries on paired datasets with single-record differences, and evaluating the results using statistical and machine learning methods

1.1 Notation

Throughout this paper, we employ the following notation:

- $\mathrm{F}[X]$ represents the cumulative distribution function (CDF) and $\mathrm{p}[X]$ represents the probability density function (PDF) of a random variable X.
- $\mathcal{D}$ denotes the set of all possible datasets that can be queried.
- D and D' represent two datasets that differ in exactly one row.
- f denotes a query function submitted by a user.
- $\mathcal{A}$ represents the randomized algorithm executed by the server.
- $S(f)$ denotes the sensitivity of function f.
- $K(f)$ for a Lipschitz function f denotes its Lipschitz constant.

- I_m stands for the $m \times m$ identity matrix.
- Lap denotes the Laplace distribution, $\mathcal{N}$ the Normal distribution; $\Phi(y)$ is the CDF and $\phi(y)$ the PDF of the standard normal distribution.
- We use the word *strategy* to denote both the analyst's decision rule and the server's response rule. When randomness is involved, we say *randomized strategy*. The term *algorithm* may appear synonymously; however, *strategy* is preferred throughout this text.

2 Theory behind Differential Privacy

The modern concept behind Differential Privacy was introduced in [1]. In this text we use a slightly modified definition:

Definition 1. *A randomized algorithm $\mathcal{A}$ is considered (ε, δ)-differentially private if for any set S in the Borel sigma-algebra over $\mathbb{R}^n$ and for any two datasets D, D' that differ in a single sample, the following holds:*

$$\Pr[\mathcal{A}(D) \in S] \leq e^{\varepsilon} \Pr[\mathcal{A}(D') \in S] + \delta$$

The algorithm is called pure ε-differentially private if the bound holds for $\delta = 0$.

Here is an example. Suppose we have a dataset with a single column of values from $\{0, 1\}$ and we want to compute the sum of these values. The server will respond using this algorithm:

$$\mathcal{A}(D) = \sum_i D_i + \mathrm{Lap}\left(0, \frac{1}{\varepsilon}\right),$$

where $\mathrm{Lap}(0, b)$ is Laplace noise with the following PDF:

$$\mathrm{p}_{\mathrm{Lap}(\mu,b)}(y) = \frac{1}{2b} \exp\left(-\frac{|y - \mu|}{b}\right).$$

As the Laplace distribution is absolutely continuous, it is enough to show a bound on the ratio of PDFs; the bound on the probability of all Borel sets will follow. Indeed, suppose for any $\tau \in S$ for some Borel S,

$$\frac{\mathrm{p}[\mathcal{A}(D)](\tau)}{\mathrm{p}[\mathcal{A}(D')](\tau)} \leq e^{\varepsilon},$$

then:

$$\Pr[\mathcal{A}(D) \in S] = \int_S \mathrm{p}[\mathcal{A}(D)](\tau)\, d\tau \leq \int_S e^{\varepsilon}\ \mathrm{p}[\mathcal{A}(D')](\tau)\, d\tau \leq e^{\varepsilon}\ \Pr[\mathcal{A}(D') \in S].$$

We are left to prove the bound on the ratio of PDFs.

$$\frac{\mathrm{p}[\mathcal{A}(D)](\tau)}{\mathrm{p}[\mathcal{A}(D')](\tau)} = \exp\left(-\frac{|\tau - \sum_i D_i|}{1/\varepsilon} + \frac{|\tau - \sum_i D_i'|}{1/\varepsilon}\right) =$$

$$\exp\left(\varepsilon\big(|\tau - \sum_i D_i + \sum_i D_i - \sum_i D_i'| - |\tau - \sum_i D_i|\big)\right) \overset{\textit{triangle inequality}}{\leq}$$

$$\exp\left(\varepsilon\big|\sum_i D_i' - \sum_i D_i\big|\right) \leq e^{\varepsilon}.$$

The last inequality follows because if a single item in the dataset changes (from 0 to 1 or vice versa), the sum changes by at most 1. This shows we have an $(\varepsilon, 0)$-differentially private algorithm for computing sums.

2.1 Algorithms for functions with bounded sensitivity

We use the notion of sensitivity to measure the amount of noise needed to make an algorithm differentially private.

Definition 2. *Given a norm $d : \mathbb{R}^n \to \mathbb{R}_+$, the sensitivity of a function f denoted $S(f)$ is defined as the supremum of $d(f(D), f(D'))$ over all datasets $D, D' \in \mathcal{D}$ differing in a single item. We refer to L1-sensitivity and L2-sensitivity when d is $\|\cdot\|_1$ and $\|\cdot\|_2$, respectively.*

Definition 3. *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz with constant $K(f)$, given norms d_1 on $\mathbb{R}^n$ and d_2 on $\mathbb{R}^m$, if for any $x, x' \in \mathbb{R}^n$ it holds that:*

$$d_2\big(f(x) - f(x')\big) \leq K(f)\, d_1(x - x').$$

Let us examine a fundamental property regarding sensitivity composition.

Remark 1. *For any function $f : \mathcal{D} \to \mathbb{R}^m$ with sensitivity $S(f)$ under norm d_1, and any function $g : \mathbb{R}^m \to \mathbb{R}^k$ that is Lipschitz with constant $K(g)$ under norms d_1, d_2, their composition $g \circ f$ has sensitivity at most $K(g)S(f)$ under norm d_2.*

Proof. Consider any datasets D, D' that differ in one element. By definition of sensitivity, $d_1\big(f(D) - f(D')\big) \leq S(f)$. Then applying the Lipschitz property of g:

$$d_2\big(g(f(D)) - g(f(D'))\big) \leq K(g)\, d_1\big(f(D) - f(D')\big) \leq K(g)S(f).$$

□

Now we are ready to provide two noising algorithms and corresponding bounds.

2.1.1 Laplace Noise

Theorem 1. *Suppose $f : \mathcal{D} \to \mathbb{R}^m$ has L1-sensitivity $S(f)$. Then the following algorithm $\mathcal{A}$ is $(\varepsilon, 0)$-differentially private:*

$$\mathcal{A}(D)_i := f(D)_i + \mathrm{Lap}(0, \tfrac{S(f)}{\varepsilon}),$$

that is, we add independent Laplace noise scaled by $\frac{S(f)}{\varepsilon}$ to each element of the resulting vector (this is not the same as multidimensional Laplace noise).

Proof. The idea is the same as in the sum example. We study PDFs, and by absolute continuity, the bound translates to probabilities of arbitrary Borel sets. Fix arbitrary $\tau \in \mathbb{R}^m$. By independence of noise over vector elements:

$$\begin{aligned}\frac{\mathrm{p}[\mathcal{A}(D)](\tau)}{\mathrm{p}[\mathcal{A}(D')](\tau)} = &\prod_{i=1}^{m} \frac{\mathrm{p}[\mathrm{Lap}(0, S(f)/\varepsilon)](\tau_i - f(D)_i)}{\mathrm{p}[\mathrm{Lap}(0, S(f)/\varepsilon)](\tau_i - f(D')_i)} = \\ &\prod_{i=1}^{m} \exp\left(-\frac{|\tau_i - f(D)_i|}{S(f)/\varepsilon} + \frac{|\tau_i - f(D')_i|}{S(f)/\varepsilon}\right) = \\ &\prod_{i=1}^{m} \exp\left(\varepsilon\big(|\tau_i - f(D)_i + f(D)_i - f(D')_i| - |\tau_i - f(D)_i|\big)\right) \overset{\textit{triangle inequality}}{\leq} \\ &\prod_{i=1}^{m} \exp\left(\varepsilon\big|f(D)_i - f(D')_i\big|\right) = \exp\left(\frac{\varepsilon}{S(f)} \sum_{i=1}^{m} |f(D)_i - f(D')_i|\right) = \\ &\exp\left(\frac{\varepsilon \, \|f(D) - f(D')\|_1}{S(f)}\right) \leq e^{\varepsilon}.\end{aligned}$$

□

2.1.2 Gaussian Noise

The other option we have is applying Gaussian noise. Analysis in this case is a bit harder, so we start with the univariate case.

Note: The following bounds are known, but I provide proofs I came up with myself (although most likely it's the way these bounds are usually proved).

Lemma 1. *Suppose $f : \mathcal{D} \to \mathbb{R}$ has L2-sensitivity $S(f)$. Then for arbitrary $\alpha, \sigma > 0$ the following algorithm $\mathcal{A}$ is $\left(\frac{S(f)^2}{2\sigma^2} + \frac{\alpha S(f)}{\sigma},\ 2 - 2\Phi(\alpha)\right)$-differentially private:*

$$\mathcal{A}(D) := f(D) + \mathcal{N}(0, \sigma^2).$$

Proof. We start with a bound on the ratio of PDFs as before.

$$\begin{aligned}\frac{\mathrm{p}[\mathcal{A}(D)](\tau)}{\mathrm{p}[\mathcal{A}(D')](\tau)} =& \frac{\mathrm{p}[\mathcal{N}(0,\sigma^2)](\tau - f(D))}{\mathrm{p}[\mathcal{N}(0,\sigma^2)](\tau - f(D'))} = \\ & \exp\left(-\frac{|\tau - f(D)|^2}{2\sigma^2} + \frac{|\tau - f(D')|^2}{2\sigma^2}\right) = \\ & \exp\left(\frac{1}{2\sigma^2}\left(|\tau - f(D) + f(D) - f(D')|^2 - |\tau - f(D)|^2\right)\right) \\ & \leq \exp\left(\frac{1}{2\sigma^2}\left(|f(D) - f(D')|^2 + 2|f(D) - f(D')| \cdot |\tau - f(D)|\right)\right).\end{aligned}$$

Here we encounter a problem: $|\tau - f(D)|$ is a multiplier to the sensitivity. We use the following trick. For now, consider only Borel sets $S \subseteq [f(D) - \alpha\sigma, f(D) + \alpha\sigma]$. For any τ in such a set, $|\tau - f(D)| \leq \alpha\sigma$, therefore

$$\frac{\mathrm{p}[\mathcal{A}(D)](\tau)}{\mathrm{p}[\mathcal{A}(D')](\tau)} \leq \exp\left(\frac{S(f)^2}{2\sigma^2} + \frac{\alpha S(f)}{\sigma}\right).$$

Denote $\varepsilon := \frac{S(f)^2}{2\sigma^2} + \frac{\alpha S(f)}{\sigma}$. We have

$$\Pr[\mathcal{A}(D) \in S] \leq e^{\varepsilon} \Pr[\mathcal{A}(D') \in S].$$

Now consider an arbitrary Borel set $T \subseteq \mathbb{R}$. Let $S = T \cap [f(D) - \alpha\sigma, f(D) + \alpha\sigma]$, $S' = T \setminus [f(D) - \alpha\sigma, f(D) + \alpha\sigma]$. Then $T = S \sqcup S'$ and

$$\begin{aligned}\Pr[\mathcal{A}(D) \in T] = \Pr[\mathcal{A}(D) \in S] + \Pr[\mathcal{A}(D) \in S'] \leq& \\ e^{\varepsilon} \Pr[\mathcal{A}(D') \in S] + \Pr\left[\mathcal{A}(D) \notin [f(D) - \alpha\sigma, f(D) + \alpha\sigma]\right]& \\ \leq e^{\varepsilon} \Pr[\mathcal{A}(D') \in T] + (2 - 2\Phi(\alpha)).&\end{aligned}$$

Set $\delta := 2 - 2\Phi(\alpha)$. □

Corollary 1. *For $\delta \leq 0.1$ and $\varepsilon \leq 4$, the algorithm above provides (ε, δ)-differential privacy with parameters $\alpha = \sqrt{2 \ln \frac{1}{\delta}}$, $\sigma = \frac{2\alpha\, S(f)}{\varepsilon}$.*

Proof. First we check the bound on δ. We can bound Gaussian tails from above as

$$1 - \Phi(a) \leq \frac{\phi(a)}{a}\left(1 + \frac{1}{a^2}\right).$$

Substituting $a = \alpha$ and using $\phi(\alpha) = \frac{1}{\sqrt{2\pi}} \exp(-\alpha^2/2)$,

$$\begin{aligned}2(1 - \Phi(\alpha)) \leq \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{\alpha^2}{2}\right) \cdot \left(\frac{1 + \alpha^2}{\alpha^3}\right) = \frac{2}{\sqrt{2\pi}}\, e^{-\ln(1/\delta)} \left(\frac{1 + \alpha^2}{\alpha^3}\right)& \\ = \frac{2}{\sqrt{2\pi}}\, \delta \left(\frac{1 + \alpha^2}{\alpha^3}\right) \leq \delta,&\end{aligned}$$

where the last inequality holds since $\frac{2}{\sqrt{2\pi}} < 1$ and, for $\delta \le 0.1$, we have $\alpha = \sqrt{2\ln(1/\delta)} \ge 2$, implying $\frac{1+\alpha^2}{\alpha^3} \le \frac{5}{8} < 1$.
Now consider

$$\frac{S(f)^2}{2\sigma^2} + \frac{\alpha S(f)}{\sigma} = \frac{S(f)^2\varepsilon^2}{8\alpha^2 S(f)^2} + \frac{\varepsilon\,\alpha S(f)}{2\alpha S(f)} = \frac{\varepsilon^2}{8\alpha^2} + \frac{\varepsilon}{2}.$$

Since $\alpha^2 = 2\ln(1/\delta) \ge 2$ for $\delta \le 0.1$ and $\varepsilon \le 4$, we get $\frac{\varepsilon^2}{8\alpha^2} \le \frac{\varepsilon^2}{16} \le \frac{\varepsilon}{4}$. Hence the sum is at most $\frac{3}{4}\varepsilon \le \varepsilon$. □

Next, let's examine the multivariate case. In the univariate case, we defined a range where the multiplicative bound is valid and showed that the probability of being outside this range is small for the random variable $\mathcal{A}(D)$. While we could apply the same approach to the multivariate case, it would yield poor results due to the curse of dimensionality. Specifically, we would need to consider the box $f(D) + [-\alpha\sigma, \alpha\sigma]^m$ for the multiplicative bound to work for all coordinates. Applying a union bound over the probability of being outside this box in each dimension would result in $\delta = 1 - (2\Phi(\alpha) - 1)^m$. This bound approaches 1 exponentially fast with dimension m, making it impractical.

We take a different approach by modifying our bounds on probability density ratios. This modification allows us to redefine when the multiplicative bound fails in terms of a sum of weighted Gaussian variables. Since such a sum follows a Gaussian distribution, we can apply standard tail bounds to obtain the desired result.

Theorem 2. *Suppose $f : \mathcal{D} \to \mathbb{R}^m$ has L2-sensitivity $S(f)$. Then for arbitrary $\alpha, \sigma > 0$ the following algorithm $\mathcal{A}$ is $\left(\frac{S(f)^2}{2\sigma^2} + \frac{\alpha S(f)}{\sigma},\ 2 - 2\Phi(\alpha)\right)$-differentially private:*

$$\mathcal{A}(D)_i := f(D)_i + \mathcal{N}(0, \sigma^2)_i.$$

Proof. First we need a bound on PDFs. As PDFs are independent Gaussians, we can apply bound from Lemma 1 to each of marginals.

$$\begin{aligned}
\frac{\mathrm{p}[\mathcal{A}(D)](\tau)}{\mathrm{p}[\mathcal{A}(D')](\tau)} &= \prod_{i=1}^{m} \frac{\mathrm{p}[\mathcal{A}(D)_i](\tau_i)}{\mathrm{p}[\mathcal{A}(D')_i](\tau_i)} \\
&= \prod_{i=1}^{m} \exp\left(\frac{1}{2\sigma^2}\left(|\tau_i - f(D)_i + f(D)_i - f(D')_i|^2 - |\tau_i - f(D)_i|^2\right)\right) \\
&= \prod_{i=1}^{m} \exp\left(\frac{1}{2\sigma^2}\left(|f(D)_i - f(D')_i|^2 + 2(\tau_i - f(D)_i)(f(D)_i - f(D')_i)\right)\right) \\
&= \exp\left(\frac{1}{2\sigma^2}\left(\sum_{i=1}^{m} |f(D)_i - f(D')_i|^2 + 2\sum_{i=1}^{m}(\tau_i - f(D)_i)(f(D)_i - f(D')_i)\right)\right) \\
&\le \exp\left(\frac{1}{2\sigma^2}\left(\|f(D) - f(D')\|_2^2 + 2\big|\langle \tau - f(D), f(D) - f(D')\rangle\big|\right)\right).
\end{aligned}$$

Denote $\Delta = f(D) - f(D')$. Denote $C = \{\tau \in \mathbb{R}^m : |\langle \tau - f(D), \Delta\rangle| \leq \alpha\sigma S(f)\}$. Now as in Lemma 1 we will start just with the Borel sets $S \subseteq C$. For any point τ in such set S:

$$\frac{\mathrm{p}[\mathcal{A}(D)](\tau)}{\mathrm{p}[\mathcal{A}(D')](\tau)} \leq \exp(\frac{S(f)^2}{2\sigma^2} + \frac{\alpha S(f)}{\sigma}).$$

Denote $\varepsilon := \exp(\frac{S(f)^2}{2\sigma^2} + \frac{\alpha S(f)}{\sigma})$.

Next we have to bound probability of event $|\langle \mathcal{A}(D) - f(D), \Delta\rangle| > \alpha\sigma S(f)$. Random variable $\mathcal{A}(D) - f(D)$ has multivariate gaussian distribution $\mathcal{N}(0, \sigma^2 \mathrm{I}_m)$ by definition, therefore $\langle \mathcal{A}(D) - f(D), \Delta\rangle \sim \mathcal{N}(0, \sigma^2 \Delta^t \mathrm{I_n} \Delta) = \mathcal{N}(0, \sigma^2 S(f)^2)$.

Therefore $\Pr[|\langle \mathcal{A}(D) - f(D), \Delta\rangle| > \alpha\sigma S(f)] \leq 2 - 2\Phi(\alpha)$. Denote $\delta = 2 - 2\Phi(\alpha)$.

We finish the proof for arbitrary set T by splitting it into disjoin union $T = S \sqcup S'$, $S \subseteq C, s' \subseteq \overline{C}$ and using union bound:

$$\Pr[\mathcal{A}(D) \in T] = \Pr[\mathcal{A}(D) \in S] + \Pr[\mathcal{A}(D) \in S'] \leq \\ e^{\varepsilon} \Pr[\mathcal{A}(D') \in S] + \Pr[\mathcal{A}(D') \in \overline{C}] \leq e^{\varepsilon} \Pr[\mathcal{A}(D') \in T] + \delta$$

□

Remark 2. *The bounds from Corollary 1 literally translate to the multidimensional case.*

3 Adaptive algorithms

In real-world applications, analysts do not specify all queries in advance; querying is an iterative process. After observing the result of a query, the analyst chooses the next query. Moreover, the analyst is not required to make exactly n queries: the process may stop early, and the total number of steps is itself a random variable. Queries can also be multidimensional, and some of them may be rejected by the server. To handle all these cases, we introduce a more general model. We do not reuse the formalization from the original article; instead, we present a self-contained definition and the accompanying privacy relation.

Definition 4 (Generalized adaptive differential privacy)**.** *We consider an interaction between an adversary (analyst) Mallory and a server. Mallory runs a deterministic strategy $\mathcal{M}$, and the server runs a randomized strategy $\mathcal{A}$. At step i, given the transcript of previous answers $t_1, \ldots, t_{i-1}$, Mallory either proposes a new query together with its noise parameter, or stops and outputs a decision bit:*

$$\mathcal{M}_i : \ (\mathbb{R} \sqcup \{\perp\})^{i-1} \to (\mathcal{P} \times \mathbb{R}^{\mathcal{D}}) \sqcup \{0,1\},$$
$$\mathcal{M}_i(t_1, \ldots, t_{i-1}) = \begin{cases} (p_i, f_i) & \textit{continue with parameter } p_i \textit{ and query } f_i : \mathcal{D} \to \mathbb{R}, \\ b \in \{0,1\} & \textit{stop and output } b. \end{cases}$$

Here $\perp$ stands for query rejected by the Server. We assume the following properties:

1. *$\mathcal{M}$ stops in at most n steps for some fixed constant n.*
2. *For each $i \in [n]$, the sets $\mathcal{M}_i^{-1}(0)$ and $\mathcal{M}_i^{-1}(1)$ are Borel*[1]*; consequently, the event of continuing is also well-defined.*
3. *$\mathcal{P} = \mathbb{R}^q$ for some fixed q, and the map $(t_1, \dots, t_{i-1}) \mapsto p_i$ is measurable for all $i \in [n]$.*
4. *For every $D \in \mathcal{D}$ and every $i \in [n]$, the map $(t_1, \dots, t_{i-1}) \mapsto f_i(D)$ is measurable.*

The server strategy acts as

$$\mathcal{A}_i : \ \mathcal{D} \times \mathbb{R}^{i-1} \times \mathcal{P}^i \times (\mathbb{R}^{\mathcal{D}})^i \to \mathbb{R} \sqcup \{\perp\},$$

$$\mathcal{A}_i(D, t_1, \dots, t_{i-1}, p_1, \dots, p_i, f_1, \dots, f_i) = \begin{cases} t_i \in \mathbb{R} & \textit{answer is released,} \\ \perp & \textit{query is rejected.} \end{cases}$$

Let $\mathrm{Out}[\mathcal{M}, \mathcal{A}](D) \in \{0, 1\}$ *denote the final output of $\mathcal{M}$ when interacting with $\mathcal{A}$ on dataset D. We say that $\mathcal{A}$ provides (ε, δ)-generalized adaptive differential privacy if, for every $\mathcal{M}$ satisfying the properties above, the probabilities below are well-defined and*

$$\Pr[\mathrm{Out}[\mathcal{M}, \mathcal{A}](D) = 1] \leq e^{\varepsilon} \Pr[\mathrm{Out}[\mathcal{M}, \mathcal{A}](D') = 1] + \delta \tag{1}$$

for all neighboring $D, D' \in \mathcal{D}$.

The definition above is very general. Although Mallory has no access to the server's internal randomness, she can post-process the transcript in ways that force the induced distribution to be for example singular, which complicates analysis a lot. We adopt a restricted definition which considers only absolutely continuous mappings. It doesn't although cover all practical cases, as for example Mallory can branch on server output yielding discrete random variable.

Definition 5 (Adaptive differential privacy)**.** *We say that $\mathcal{A}$ provides (ε, δ)-adaptive differential privacy if* (1) *holds for all $\mathcal{M}$ satisfying items 1-4 of Definition 4 and, additionally:*

5. *For every $i \in [n]$, the map $(t_1, \dots, t_{i-1}) \mapsto p_i$ is absolutely continuous.*
6. *For every $i \in [n]$ and every $D \in \mathcal{D}$, the map $(t_1, \dots, t_{i-1}) \mapsto f_i(D)$ is absolutely continuous*

With this restriction, we can design strategies such that we only deal with absolutely continuous random variables, which substantially simplifies the analysis. We first record auxiliary notions and lemmas that apply to reasonable strategies.

[1] It is a bit abuse of notation to use term Borel here, as $\mathcal{M}_i$ domain is not $\mathbb{R}^{i-1}$ but $(\mathbb{R} \cup \{\perp\})^{i-1}$ but it is straightforward to define σ-algebra on such a set. It will contain S and $\{\perp\} \cup S$ for all Borel S in $\mathbb{R}^{i-1}$. The term Borel and measurable should be understand in that sense throughout the text

Definition 6. *We call a server strategy $\mathcal{A}$ reasonable if, two following conditions hold:*

1. *Decision whether to accept query f_i with parameter p_i depends only on the previous successful queries and corresponding answers. It is independent of dataset as well as previous rejected queries. Formally consider two runs where server follows strategy $\mathcal{A}$: $(t_1, \ldots, t_{i-1}, p_1, \ldots, p_i, f_1, \ldots, f_i)$ and $(t'_1, \ldots, t'_{j-1}, p'_1, \ldots, p'_j, f'_1, \ldots, f'_j)$. Drop all of the queries which were rejected, let $\{q\}_1^k \subseteq [i-1]$ be the indices of successful queries in first run and $\{q\}_1^{k'} \subseteq [j-1]$ in the second. If*

$$\begin{cases} k = k', \\ (t_{q_1}, \ldots t_{q_k}) = (t'_{q'_1}, \ldots t'_{q'_k}), \\ (p_{q_1}, \ldots p_{q_k}) = (p'_{q'_1}, \ldots p'_{q'_k}), \\ (f_{q_1}, \ldots f_{q_k}) = (f'_{q'_1}, \ldots f'_{q'_k}), \\ p_i = p'_j, f_i = f'_j, \end{cases}$$

   then for all $D, D' \in \mathcal{D}$,

$$\mathcal{A}_i(D, t_1, \ldots, t_{i-1}, p_1, \ldots, p_i, f_1, \ldots, f_i) = \perp \Leftrightarrow \mathcal{A}_j(D', t'_1, \ldots, t'_{j-1}, p'_1, \ldots, p'_j, f'_1, \ldots, f'_j) = \perp.$$

   Note that we can build a family of predicates $r_k(t_1, \ldots, t_k, p_1, \ldots, p_k, f_1, \ldots, f_k, p_*, f_*)$ which tell if the query p_*, f_* gets rejected by only considering previous successful queries ($r_k = 0$ is query was rejected, 1 otherwise).

2. *Whenever a query is accepted at step i, $\mathcal{A}$ returns*

$$\mathcal{A}_i(D, t_1, \ldots, t_{i-1}, p_1, \ldots, p_i, f_1, \ldots, f_i) = f_i(D) + \phi(p_i, X_i),$$

   where $X_1, X_2, \ldots$ is a sequence of i.i.d. absolutely continuous random variables, and ϕ is a scaling function absolutely continuous in its arguments on the codomain of (p_i, X_i).

Lemma 2. *Suppose there exists a reasonable server strategy $\mathcal{A}$ such that (1) holds for all analyst strategies $\mathcal{M}$ satisfying items 1-6 above and making no rejected queries (i.e., all proposed queries are accepted). Then $\mathcal{A}$ is (ε, δ)-differentially private.*

Proof. Assume, toward a contradiction, that there exists some analyst strategy $\mathcal{M}$ (possibly with rejections) and neighboring D, D' such that

$$\Pr[\mathrm{Out}[\mathcal{M}, \mathcal{A}](D) = 1] > e^{\varepsilon} Pr[\mathrm{Out}[\mathcal{M}, \mathcal{A}](D') = 1] + \delta.$$

Construct $\mathcal{M}'$ that, at each step i, first computes $(p_i, f_i) = \mathcal{M}_i(t_1, \ldots, t_{i-1})$, then removes unsuccessful queries and evaluates the predicate r_k to decide acceptance. If the query would be accepted, it forwards it to the server; otherwise, it returns "rejected" to its internal rule and proceeds. Because $\mathcal{A}$ is reasonable the probability that we reach i-th step as well as the conditional distribution of inputs to $\mathcal{M}_i$ given we reached i-th step is equal both when we run $\mathcal{M}$ against the server and as a part of $\mathcal{M}'$. Consider the step i, and condition on the $t_1, \ldots, t_{i-1}$. Now the query and parameter (p_i, f_i) are fixed. Decision if the query would be rejected is deterministic and same both for $\mathcal{M}$ being run against the server and as a part of $\mathcal{M}'$. If the query is not rejected answer distribution depends only on the query (p_i, f_i) itself, so it is again the same for $\mathcal{M}$ being part of $\mathcal{M}'$ and communicating with server. □

Lemma 3. *Suppose there exists a reasonable server strategy $\mathcal{A}$ such that, for every n, (1) holds for all analyst strategies $\mathcal{M}$ satisfying items 1-6 above, with no rejections, and making exactly n queries. Then $\mathcal{A}$ is (ε, δ)-differentially private.*

Proof. By Lemma 2, it suffices to consider analyst strategies that experience no rejections. Let n be the maximal number of queries made with nonzero probability. If, with positive probability, $\mathcal{M}$ stops after $i < n$ queries, define $\mathcal{M}'$ that forces an additional $n - i$ dummy queries whose answers are ignored by the decision rule, and then outputs the same final bit as $\mathcal{M}$. The distribution of the final output remains unchanged, whereas $\mathcal{M}'$ now makes exactly n queries. It can be easily seen because the same values of underlying $\mathcal{A}$ random variables $X_1, X_2, \ldots$ yield same answers of both $\mathcal{M}$ and $\mathcal{M}'$. This contradicts the hypothesis unless (1) already holds. □

By using this two lemmas we can study only the strategies with the exact constant number of queries, that is we can consider transcript as a random vector taking values in $\mathbb{R}^n$ in ordinary sense. Moreover because of measurability conditions the probability of $\mathcal{M}$ answering 1 is exactly the probability of $t = (t_1, \ldots, t_n)$ belonging to some Borel set $S_{\mathcal{M}}$. That's basically the definition used in original article so the further proof is the same, but we have formally reasoned that rejected queries and non-uniform number of queries made doesn't alter the situation.

Remark 3. *For a reasonable algorithm, as $f_i(t_{\leq i-1}), p_i(t_{\leq i-1}), \phi$ are absolute continuous, and underlying noise r.v. X_i are also absolute continuous we can consider transcript t as an absolute continuous vector and operate with PDFs.*

3.1 Composition for Laplacian noise strategy

Theorem 3 (Composition for Laplacian noise)**.** *In the model above, let p_i be the Laplacian scale parameter and $S(f)$ L1-sensitivity. The server has a privacy budget ε and rejects the i-th query if $\sum_{j=1}^{i} r_i \frac{S(f_j)}{p_j} > \varepsilon$. Otherwise, it answers with the strategy:*

$$\mathcal{A}_i(D) = f_i(D) + \mathrm{Lap}(0, p_i).$$

The strategy above guarantees $(\varepsilon, 0)$-differential privacy.

Proof. Firstly notice that server strategy is reasonable. Therefore we can operate with transcript as an absolutely continuous random vector, denote it $t(\mathcal{M}, D)$. Fix the dimension to be n. Moreover as no query was rejected $\sum_{i=1}^{n} \frac{S(f_j(t_1,\ldots,t_{i-1}))}{p_j(t_1,\ldots,p_{i-1})} \leq \varepsilon$. As in the Theorem 1 it is enough to prove that for any $\tau \in \mathbb{R}^n$ and neighboring $D, D' \in \mathcal{D}$ the PDF $\mathrm{p}[t(\mathcal{M}, D)](\tau) \leq e^{\varepsilon}\,\mathrm{p}[t(\mathcal{M}, D')](\tau)$. We do so by conditional probability law:

$$\frac{\mathrm{p}[t(\mathcal{M}, D)](\tau)}{\mathrm{p}[t(\mathcal{M}, D')](\tau)} =$$
$$\frac{\mathrm{p}[t_1(\mathcal{M}, D)](\tau_1)\,\mathrm{p}[t_2(\mathcal{M}, D)|t_1 = \tau_1](\tau_2)\ldots\mathrm{p}[t_n(\mathcal{M}, D)|t_1 = \tau_1, \ldots, t_{n-1} = \tau_{n-1}](\tau_n)}{\mathrm{p}[t_1(\mathcal{M}, D')](\tau_1)\,\mathrm{p}[t_2(\mathcal{M}, D')|t_1 = \tau_1](\tau_2)\ldots\mathrm{p}[t_n(\mathcal{M}, D')|t_1 = \tau_1, \ldots, t_{n-1} = \tau_{n-1}](\tau_n)} =$$
$$\prod_{i=1}^{n} \exp\left(-\frac{|f_i(D, \tau_1, \ldots, \tau_{i-1}) - \tau_i|}{p_i(\tau_1, \ldots, \tau_{i-1})} + \frac{|f_i(D', \tau_1, \ldots, \tau_{i-1}) - \tau_i|}{p_i(\tau_1, \ldots, \tau_{i-1})}\right) \overset{\textit{triangle inequality}}{\leq}$$
$$\prod_{i=1}^{n} \exp\left(\frac{|f_i(D, \tau_1, \ldots, \tau_{i-1}) - f_i(D', \tau_1, \ldots, \tau_{i-1})|}{p_i(\tau_1, \ldots, \tau_{i-1})}\right) \leq \exp(\sum_{i=1}^{n} \frac{S(f_i(\tau_1, \ldots, \tau_{i-1}))}{p_i(\tau_1, \ldots, \tau_{i-1})}) \leq e^{\varepsilon}$$

□

3.2 Advanced composition for Gaussian Noise

Theorem 4 (Azuma's inequality for subgaussians)**.** *Suppose $X_0, \ldots, X_n$ is a martingale and $\mathcal{F}_0 \ldots, \mathcal{F}_n$ coresponding filtration. Suppose additionally that almost surely $D_i = X_i - X_{i-1}$ is a subgaussian random variable, that is exist constants k_i such that is almost surely*

$$\mathrm{E}[\exp(\lambda D_i)\,|\mathcal{F}_{i-1}] \leq \exp(k_i^2 \lambda^2)$$

for all $\lambda \in \mathbb{R}, i \in 1 \ldots n$. Then

$$\Pr[|X_0 - X_n| \geq \epsilon] \leq 2 \exp\left(-\frac{\epsilon^2}{4 \sum_i k_i^2}\right)$$

Proof. We prove

$$* = \Pr[X_0 - X_n \geq \epsilon] \leq \exp\left(-\frac{\epsilon^2}{4 \sum_i k_i^2}\right),$$

the inequality in opposite direction is proved similarly.

$$\Pr[X_0 - X_n \geq \epsilon] = \Pr[\exp(\lambda(X_0 - X_n)) \geq e^{\lambda\epsilon}] \overset{\text{Markov}}{\leq}$$
$$\frac{\mathrm{E}[\exp(\lambda(X_0 - X_n))]}{e^{\lambda\epsilon}} = \frac{\mathrm{E}[\exp(\sum_{i=1}^{n} \lambda D_i)]}{e^{\lambda\epsilon}} =$$
$$\frac{\mathrm{E}[\mathrm{E}[\exp\left(\sum_{i=1}^{n-1} \lambda D_i\right) \exp(D_n)\,|\mathcal{F}_{n-1}]]}{e^{\lambda\epsilon}} = \frac{\mathrm{E}[\exp\left(\sum_{i=1}^{n-1} \lambda D_i\right) \mathrm{E}[\exp(\lambda D_n)\,|\mathcal{F}_{n-1}]]}{e^{\lambda\epsilon}}$$

The last equality follows because $D_i, i < n$ are $\mathcal{F}_{n-1}$ measurable. By theorem statement

$$\mathrm{E}[\exp(\lambda D_n)\,|\mathcal{F}_{n-1}] \leq \exp(k_i^2 \lambda^2),$$

continuing the derivation above inductively we arrive at

$$* \leq \frac{\exp\left(\lambda^2 \sum_i k_i^2\right)}{e^{\lambda\epsilon}}.$$

Select $\lambda = \frac{\epsilon}{2\sum_i k_i^2}$, then:

$$* \leq \exp\left(-\frac{\epsilon^2}{2\sum_i k_i^2} + \frac{\epsilon^2}{4\sum_i k_i^2}\right) = \exp\left(-\frac{\epsilon^2}{4\sum_i k_i^2}\right).$$

□

Now we are ready to prove the main theorem. It is the privacy bound for adaptive algorithm using the Gaussian noise. It uses subgaussians and martingale machinery and it is basically the reason I decided to choose this topic.

Theorem 5 (Composition for Gaussian noise)**.** *In the model above, let $p_i = \sigma_i$ be the parameter. Let $S(f)$ be L2-sensitivity. The server has a privacy budget (ε, δ) and a parameter γ defining the ratio between ε and δ used for a query. We further bound our model by allowing at most n queries for a constant n (previously n depended on $\mathcal{M}$). Server rejects the i-th query if $\sum_{j=1}^{i} \frac{S^2(f_j)}{\sigma_j^2} + \gamma > \varepsilon$ or $\frac{S(f_i)}{\sigma_i} > \frac{\gamma}{2\sqrt{n\ln(1/\delta)}} =: k$. Otherwise, it answers with the strategy:*

$$\mathcal{A}_i(D) = f_i(D) + \mathcal{N}(0, \sigma_i).$$

The strategy above guarantees (ε, δ)-differential privacy.

Proof. We start again by noticing that server strategy is reasonable, therefore all we have to do, is to prove that for any $n \in \mathbb{N}$, any Borel $S \in \mathbb{R}^n$ and any neighboring $D, D' \in \mathcal{D}$:

$$\Pr[(t_1(\mathcal{M}, D), \ldots, t_n(\mathcal{M}, D)) \in S] \leq e^{\varepsilon} \Pr[(t_1(\mathcal{M}, D'), \ldots, t_n(\mathcal{M}, D')) \in S] + \delta.$$

Next we notice following already common PDF relation, to simplify writing we use notation $\tau_{\leq i} = (\tau_1, \ldots, \tau_i)$. Denote $\Delta_i(\tau_{\leq i-1}) = f_i(D, \tau_{\leq i-1}) - f_i(D', \tau_{\leq i-1}), \Delta(\tau_{\leq n-1}) = (\Delta_1, \ldots, \Delta_n(\tau_{\leq n-1}))$.

$$\frac{\mathrm{p}[t(\mathcal{M}, D)](\tau)}{\mathrm{p}[t(\mathcal{M}, D')](\tau)} = \frac{\mathrm{p}[t_1(\mathcal{M}, D)](\tau_1) \ldots \mathrm{p}[t_n(\mathcal{M}, D)|t_1 = \tau_1, \ldots, t_{n-1} = \tau_{n-1}](\tau_n)}{\mathrm{p}[t_1(\mathcal{M}, D')](\tau_1) \ldots \mathrm{p}[t_n(\mathcal{M}, D')|t_1 = \tau_1, \ldots, t_{n-1} = \tau_{n-1}](\tau_n)} =$$

$$\prod_{i=1}^{n} \exp\left(\frac{|\tau_i - f_i(D, \tau_{\leq i-1}) + f_i(D, \tau_{\leq i-1}) - f_i(D', \tau_{\leq i-1})|^2}{\sigma_i^2(\tau_{\leq i-1})} - \frac{|\tau_i - f_i(D, \tau_{\leq i-1})|^2}{\sigma_i^2(\tau_{\leq i-1})}\right) \overset{\textit{triangle inequality}}{\leq}$$

$$\exp\left(\sum_i^n \frac{|f_i(D, \tau_{\leq i-1}) - f_i(D', \tau_{\leq i-1})|^2}{\sigma_i^2(\tau_{\leq i-1})} + \frac{2(\tau_i - f_i(D, \tau_{\leq i-1}))(f_i(D', \tau_{\leq i-1}) - f_i(D', \tau_{\leq i-1}))}{\sigma_i^2(\tau_{\leq i-1})}\right) \leq$$

$$\exp\left(\sum_i^n \frac{|\Delta_i|^2}{\sigma_i^2(\tau_{\leq i-1})} + \left|\sum_i^n \frac{(\tau_i - f_i(D, \tau_{\leq i-1}))\Delta_i}{\sigma_i^2(\tau_{\leq i-1})}\right|\right)$$

Denote $C = \{\tau : \left|\sum_i^n \frac{(\tau_i - f_i(D,\tau_{\le i-1}))\Delta_i}{\sigma_i^2(\tau_{\le i-1})}\right| \le \gamma\}$ (it is Borel as it is a preimage of measurable function). Consider arbitrary Borel set S inside C. As no queries are rejected $\sum_i^n \frac{S(f_i)}{\sigma_i^2}$???

Now we are to prove that the probability of $t(\mathcal{M}, D)$ being outside of C is small using Azuma's inequality 4.

We define filtration in following manner $\mathcal{F}_i = \sigma(t_1(\mathcal{M}, D), \dots t_i(\mathcal{M}, D))$, and consider a Doob martingale:

$$X_n = \sum_i^n \frac{(\tau_i - f_i(D, \tau_{\le i-1}))\Delta_i}{\sigma_i^2(\tau_{\le i-1})},$$

$$X_0 = E[X_n], X_i = E[X_n|\mathcal{F}_i].$$

Next we notice that $X_0 = 0$, indeed:

$$X_0 = \mathrm{E}\left[\sum_i^n \frac{(t_i - f_i(D, t_{\le i-1}))\Delta_i(t_{\le i-1})}{\sigma_i^2(t_{\le i-1})}\right] = \sum_i^n \mathrm{E}\left[\mathrm{E}\left[\frac{(t_i - f_i(D, t_{\le i-1}))\Delta_i(t_{\le i-1})}{\sigma_i^2(t_{\le i-1})}|\mathcal{F}_i\right]\right] = \\ \sum_i^n \mathrm{E}\left[\frac{\mathrm{E}\left[(t_i - f_i(D, t_{\le i-1}))|\mathcal{F}_i\right]\Delta_i(t_{\le i-1})}{\sigma_i^2(t_{\le i-1})}\right] = 0.$$

The last equality holds because when conditioning the value under internal expectation is distributed normally with expectation 0. Next we consider $D_i = X_i - X_{i-1}, i \in [n]$.

$$D_i = \mathrm{E}[X_n|\mathcal{F}_i] - \mathrm{E}[X_n|\mathcal{F}_{i-1}] = \mathrm{E}[X_n - \mathrm{E}[X_n|\mathcal{F}_{i-1}]|\mathcal{F}_i] = \\ \sum_j^n \mathrm{E}\left[\frac{(t_j - f_j(D, t_{\le j-1}))\Delta_j(t_{\le j-1})}{\sigma_j^2(t_{\le j-1})} - E\left[\frac{(t_j - f_j(D, t_{\le j-1}))\Delta_j(t_{\le j-1})}{\sigma_j^2(t_{\le j-1})}|\mathcal{F}_{i-1}\right]|\mathcal{F}_i\right].$$

We consider summands $j < i, j = i$ and $j > i$ separately. Denote $S_j(t_{\le j}) = \frac{(t_j - f_j(D, t_{\le j-1}))\Delta_j(t_{\le j-1})}{\sigma_j^2(t_{\le j-1})}$.

1. $j \le i-1$: $S_j(t_{\le j})$ is $\mathcal{F}_{i-1}$ measurable, therefore

   $$\mathrm{E}\left[S_j(t_{\le j})|\mathcal{F}_{i-1}\right] = S_j(t_{\le j}) \Rightarrow \mathrm{E}\left[S_j(t_{\le j}) - \mathrm{E}\left[S_j(t_{\le j})|\mathcal{F}_{i-1}\right]|\mathcal{F}_i\right] = 0.$$

2. $j \ge i+1$: Firstly we notice:

   $$\mathrm{E}[S_j(t_{\le j})|\mathcal{F}_{j-1}] = \mathrm{E}\left[\frac{(t_j - f_j(D, t_{\le j-1}))\Delta_j(t_{\le j-1})}{\sigma_j^2(t_{\le j-1})}|\mathcal{F}_{j-1}\right] = \\ \frac{\mathrm{E}\left[t_j - f_j(D, t_{\le j-1})|\mathcal{F}_{j-1}\right]\Delta_j(t_{\le j-1})}{\sigma_j^2(t_{\le j-1})} = 0.$$

   The last line holds because when conditioned to $\mathcal{F}_{j-1}$, random variable $t_j - f_j(D, t_{\le j-1})$ has normal distribution with zero mean. Therefore:

   $$\mathrm{E}\left[S_j(t_{\le j})|\mathcal{F}_i\right] - \mathrm{E}\left[S_j(t_{\le j})|\mathcal{F}_{i-1}\right] = \mathrm{E}\left[\mathrm{E}\left[S_j(t_{\le j})|\mathcal{F}_{j-1}\right]|\mathcal{F}_i\right] - \mathrm{E}\left[\mathrm{E}\left[S_j(t_{\le j})|\mathcal{F}_{j-1}\right]|\mathcal{F}_{i-1}\right] = 0$$

3. $j = i$: Notice, that as $\mathrm{E}\left[S_i(t_{\le i})|\mathcal{F}_{i-1}\right] = 0$,

$$\mathrm{E}\left[S_j(t_{\le j})|\mathcal{F}_i\right] - \mathrm{E}\left[S_j(t_{\le j})|\mathcal{F}_{i-1}\right] = \mathrm{E}\left[S_i(t_{\le i})|\mathcal{F}_i\right] = S_i(t_{\le i}).$$

   Last line again follows as S_i is $\mathcal{F}_i$ measurable.

Therefore $D_i = \mathrm{E}\left[S_i(t_{\le i})|\mathcal{F}_i\right]$. We need a subgaussian bound for D_i to apply Theorem 4. To be more precise we have to bound

$$\mathrm{E}\left[\exp(\lambda D_i)|\mathcal{F}_{i-1}\right] = \mathrm{E}\left[\exp(\lambda S_i(t_{\le i}))|\mathcal{F}_{i-1}\right] = \\ \mathrm{E}\left[\exp\left(\lambda\frac{(t_i - f_i(t_{\le i-1}))\Delta_i(t_{\le i-1})}{\sigma_i^2(t_{\le i-1})}\right)|\mathcal{F}_{i-1}\right] = *.$$

Notice that under such conditioning $\frac{(t_i - f_i(t_{\le i-1}))\Delta_i(t_{\le i-1})}{\sigma_i^2(t_{\le i-1})}$ has distribution $\mathcal{N}(0, \frac{|\Delta_i^2(t_{\le i-1})|}{\sigma^2(t_{\le i-1})})$. By a moment generating function of Gaussian: $* = \frac{\lambda^2\Delta^2(t_{\le i-1})}{2\sigma^2(t_{\le i-1})} \le \lambda^2 k^2$.
Finally we apply Theorem 4 which yields that:

$$\Pr[|X_n| \ge \gamma] = \Pr[|X_0 - X_n| \ge \gamma] \le \exp\left(-\frac{\gamma^2}{4nk^2}\right) = \delta.$$

Now for arbitrary Borel set T we can decompose it into $S \sqcup S', S \subseteq C, S' \cap C = \varnothing$,

$$\Pr[t(\mathcal{M}, D) \in T] = \Pr[t(\mathcal{M}, D) \in S] + \Pr[t(\mathcal{M}, D) \in S'] \le \\ e^{\varepsilon}\Pr[t(\mathcal{M}, D') \in S] + \Pr[t(\mathcal{M}, D) \notin C] \le e^{\varepsilon}\Pr[t(\mathcal{M}, D') \in T] + \delta.$$

□

Now we select optimal parameters.

Corollary 2. *Suppose we want to build $(\varepsilon_0, \delta_0)$ private strategy for $\varepsilon_0 \le 4, \delta_0 \le 0.1$. Select $\gamma = \frac{1}{2}\varepsilon_0$, $k = \frac{\varepsilon_0}{4\sqrt{n\ln(1/\delta_0)}}$. Then strategy $\mathcal{A}$ is $(\varepsilon_0, \delta_0)$-differentially private.*

Proof. By Theorem 5,

$$\varepsilon \le \sum_{i=1}^{n}\frac{S(f_i)^2}{\sigma_i^2} + \gamma \le nk^2 + \gamma = n\frac{\gamma^2}{2n\ln(1/\delta_0)} + \gamma = \frac{\varepsilon_0^2}{8\ln(1/\delta_0)} + \frac{\varepsilon_0}{2} \le \varepsilon_0,$$

as $\varepsilon_0 \le 4, \ln(1/\delta_0) > 1$. $\delta \le \delta_0$ follows because $\frac{S(f_i)}{\sigma_i} \le k$ a.s. for each query. □

4 Further work

1. We have restricted Mallory strategies only to those operating with absolutely continuous functions to simplify analysis. In fact even when f_i or p_i is discrete the resulting transcript can be absolutely continuous for a reasonable server strategy. For example let $t_1 \sim \mathcal{N}(0, 1)$, and $t_2 = f_2 + \mathcal{N}(0, 1) = \mathbf{I}[t_1 \leq 0] + \mathcal{N}(0, 1)$. Then CDF of $\mathbf{F}_{t_2}(x) = \frac{1}{2}\Phi(x) + \frac{1}{2}\Phi(x - 1)$ which is obviously absolute continuous. So more broad criterion for allowed Mallory strategies would benefit the paper.

2. In the Theorem 5 we basically force Mallory to use specific sigma for query, by rejection condition, while in 3 scaling factor for the noise can be arbitrary as far as sum condition holds. The bound for fixed n in gaussian case is required for the same reason, we have to know how to distribute privacy budget between queries. We could overcome this if the k_i in the bound in the Azuma's inequality were random variables themselves with their second norm limited almost surely with a constant. Yet I haven't found a proof to the such inequality.

References

[1] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In Shai Halevi and Tal Rabin, editors, *Theory of Cryptography*, pages 265–284, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.