

Investigation of the Elliptical Gaussian Noise in the case of multivariate normal data

Tim Sehested Poulsen

December 11, 2022

1 Preliminaries

1.1 Definitions

The dataset is denoted by X , where $X \in \mathbb{R}^{n \times d}$. I have that n denotes the number of entries in the dataset and d is the number of dimensions of the dataset. I will throughout the report refer to a single entry of the dataset as x_i and a single dimension of the dataset as $X^{(j)}$, and therefore $x_i^{(j)}$ denotes the j 'th dimension of the i 'th entry.

Differential privacy is the heuristic of releasing a database statistic whilst limiting the impact of any one entry. It builds on the intuition that computing a statistics on a private dataset should not reveal any sensitive information about any one individual as long as that individual has little to no effect on the outcome. Differential privacy has multiple slightly different formal definitions, one such is (ϵ, δ) -Differential Privacy referred to as (ϵ, δ) -DP which will be introduced later on. A prerequisite for almost all of the different differential privacy definitions relies on the concept of neighbouring dataset.

Definition 1.1 (Neighbouring dataset [4]). *Two dataset $X, X' \in \mathbb{R}^{n \times d}$ are said to be neighbouring if they differ in at most a single entry. Neighbouring dataset are denoted with the relation $X \sim X'$ and defined as followed*

$$X \sim X' \iff |\{i \in \mathbb{N} \mid i \leq n \wedge x_i \neq x'_i\}| \leq 1$$

Definition 1.2 (Sensitivity [10]). *Let $f(X) : \mathbb{R}^{n \times d} \rightarrow \mathbb{R}^d$ given by $f(X) = \sum_{i=1}^n x_i$ be the sum over all vectors in a dataset. The sensitivity is then the maximal possible difference in the output of our summation from two neighbouring dataset denoted as Δ . We denote the sensitivity of the j 'th dimension as*

$$\Delta_j = \max_{X \sim X'} |f(X)^{(j)} - f(X')^{(j)}|$$

and then the total l_2 -sensitivity is then

$$\|\Delta\| = \max_{X \sim X'} \|f(X) - f(X')\| \tag{1}$$

Definition 1.3 ((ϵ, δ)-Differential Privacy [4]). *A randomized algorithm $\mathcal{M} : \mathbb{R}^{n \times d} \rightarrow \mathcal{R}$ is (ϵ, δ)-differentially private if for all possible subsets of outputs $S \subseteq \mathcal{R}$ and all pairs of neighbouring dataset $X \sim X'$ we have that*

$$\Pr [M(X) \in S] \leq e^\epsilon \cdot \Pr [M(X') \in S] + \delta$$

Error Measure As I will be working exclusively with the sum of entries in a dataset, error will be defined as the expected squared l_2 -norm between the true sum and the output of a randomized algorithm. So let $X \in \mathbb{R}^{n \times d}$ be the dataset and $f(X) = \sum_i^n x_i$ be the true sum. The error of a randomized algorithm $M : \mathbb{R}^{n \times d} \rightarrow \mathbb{R}^d$ which estimates $f(X)$ is then

$$\text{Err}(M) := \mathbb{E} [\|M(X) - f(X)\|^2]$$

extra

If proof of elliptical ... is included then write about edp is preserved during transformation

1.2 Quadratic forms of random variables

Quadratics of random variables have been well studied [1, 9], specially in the case of multivariate gaussian variables [6, 9]. Even more research has been done in evaluating the CDF of these quadratic forms for Gaussian random vectors [3, 7].

Theorem 1 (Expectation of a quadratic random variable [1]). *Let X be a d -dimensional random vector with expected value $\mathbb{E}[X] = \boldsymbol{\mu}_X$ and covariance matrix $\text{Var}[X] = \boldsymbol{\Sigma}_X$. Let also A be a constant $k \times k$ symmetric matrix, then*

$$\mathbb{E} [X^T A X] = \text{tr} (A \boldsymbol{\Sigma}_X) + \boldsymbol{\mu}^T A \boldsymbol{\mu}$$

Proof

Blah blah

$$\begin{aligned} \mathbb{E} [X^T A X] &= \text{tr} (\mathbb{E} [X^T A X]) = \mathbb{E} [\text{tr} (X^T A X)] \\ &= \mathbb{E} [\text{tr} (A X X^T)] = \text{tr} (A \mathbb{E} [X X^T]) = \text{tr} (A (\text{Var} [X] + \boldsymbol{\mu} \boldsymbol{\mu}^T)) \\ &= \text{tr} (A \boldsymbol{\Sigma}) + \text{tr} (A \boldsymbol{\mu} \boldsymbol{\mu}^T) = \text{tr} (A \boldsymbol{\Sigma}) + \boldsymbol{\mu}^T A \boldsymbol{\mu} \end{aligned}$$

blah

■

2 Algorithms

2.1 The Gaussian Mechanism

One of the most foundational algorithms for achieving (ϵ, δ)-DP is the Gaussian Mechanism [5]. It computes the real value of a statistic, where the l_2 -sensitivity is known. That is it produces a (ϵ, δ)-DP estimate of a function $g : \mathbb{R}^{n \times d} \rightarrow \mathbb{R}^d$ where $\|\Delta\|$ for the function g is known. It does so by computing the value of $g(X)$ and then adding noise

Algorithm 1 The Gaussian Mechanism

Input

$\sigma_{\varepsilon,\delta}$ Standard deviation required to achieve (ε, δ) -DP
 $X \in \mathbb{R}^{n \times d}$ Dataset

Output

(ε, δ) -DP estimate of $g(X)$
 $\eta \leftarrow$ sample from $\mathcal{N}(\vec{0}, \sigma_{\varepsilon,\delta}^2 I)$
return $g(X) + \eta$

to each dimension from the normal distribution $\mathcal{N}(0, \sigma_{\varepsilon,\delta}^2)$. This can be seen as adding a noise vector η which is then distributed according to the multivariate normal distribution $\mathcal{N}(\vec{0}, \sigma_{\varepsilon,\delta}^2 I)$. The algorithm can be seen in Algorithm 1.

It is quite apparent that the main difficulty of the mechanism lies in determining a $\sigma_{\varepsilon,\delta}$ which achieves (ε, δ) -DP, and preferably the smallest such one.

The following theorem was initially proven

Theorem 2. [5] *Let $g : \mathbb{R}^{n \times d} \rightarrow \mathbb{R}^d$ be an arbitrary d -dimensional function with l_2 -sensitivity $\|\Delta\| = \max_{X \sim X'} \|g(X) - g(X')\|$, and let $\varepsilon \in (0, 1)$. The Gaussian Mechanism with $\sigma_{\varepsilon,\delta} = \|\Delta\| \sqrt{2 \ln(1.25/\delta)}/\varepsilon$ is (ε, δ) -DP.*

The proof is rather long and is therefore omitted here.

In the case where $g(X) = f(X)$, i.e. it is estimating the sum of entries we have quite intuitively that the error is given by the norm of the noise introduced as

$$\mathbb{E} [\|(g(X) + \eta) - f(X)\|^2] = \mathbb{E} [\|(f(X) + \eta) - f(X)\|^2] = \mathbb{E} [\|\eta\|^2]$$

Which by theorem ?? is

$$\mathbb{E} [\|\eta\|^2] = \sum_i^d \sigma_{\varepsilon,\delta}^2 = d \cdot \sigma_{\varepsilon,\delta}^2$$

-Does introduce the same variance for all dimensions regardless of individual variance

-Talk about finding the minimal σ s.t. privacy is held, and the error in that case

3 Problem setup

The problem consists of releasing the sum of vectors in a dataset under differential privacy. More formally we wish to release the value of $f : \mathbb{R}^{n \times d} \rightarrow \mathbb{R}^d$ given by

$$f(X) = \sum_{i=1}^n x_i$$

under (ε, δ) -DP.

The common factor for achieving (ε, δ) -DP in both the Gaussian Mechanism and the Elliptical Gaussian Mechanism is the requirement that data lie within some hyperrectangle. It is formally described as each dimension of the data must lie within some range

$x_i^{(j)} \in [-\Delta_j/2, \Delta_j/2]$. This requirement is needed to know the l_2 -sensitivity $\|\Delta\|$ as defined in equation 1 of the data. In this project I will change this assumption and instead look at the case where each dimension is normally distributed. This means that for each $j \in [d]$ we have that $X^{(j)} \sim \mathcal{N}(\mu_j, \sigma_j^2)$. An equivalent formulation is that the data is multivariately distributed but with no correlation between dimensions. This means that $X \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$, where Σ is a diagonal matrix with the variance of each dimension along its diagonal. It is quite apparent that determining a $\|\Delta\|$ is impossible in this setting as the Gaussian distribution is continuously defined on the range $(-\infty, \infty)$. What has been done in several recent papers is that data is *clipped* by some threshold C [2,8]. Clipping is the process of limiting the norm of any one entry to be at most C . This means that every vector is transformed as such

$$\hat{x}_i := \min \left\{ \frac{C}{\|x_i\|}, 1 \right\} \cdot x_i$$

Clipping entries by a factor C thus means that $\|\Delta\| = C$ as any one entry cannot have more impact on the summation than C . It can then be seen that if the summation $f(X)$ is instead performed on a clipped dataset \hat{X} then this is equivalent to defining the summation function $\hat{f} : \mathbb{R}^{n \times d} \rightarrow \mathbb{R}^d$ as

$$\hat{f}(X) = \sum_i^n \min \left\{ \frac{C}{\|x_i\|}, 1 \right\} \cdot x_i$$

Then by theorem 2 the gaussian mechanism with the function \hat{f} is (ε, δ) -DP with $\sigma_{\varepsilon, \delta} = C\sqrt{2 \ln(1.25/\delta)}/\varepsilon$. Though the mechanism is still (ε, δ) -DP it will now have a larger error when regarding the true sum $f(X) = \sum_i^n x_i$ as the actual answer.

3.1 Gaussian data

Let $X^{(j)} \sim \mathcal{N}(0, \sigma_j^2)$ As the expected l_2 -norm of x_i is given by

$$\mathbb{E} [\|x_i\|^2] = \sum_{j=1}^d \sigma_j^2$$

To achieve an expected norm of 1 I will scale each dimension by a factor $\frac{1}{b_j}$ which achieves this. If

$$\hat{x}_i = \left(\frac{x_i^{(0)}}{b_0}, \frac{x_i^{(1)}}{b_1}, \dots, \frac{x_i^{(d)}}{b_d} \right)$$

This means that $X^{(j)} \sim \mathcal{N}(0, \frac{\sigma_j^2}{b_j^2})$ and the expected norm is given by

$$\mathbb{E} [\|x_i\|^2] = \sum_{j=1}^d \frac{\sigma_j^2}{b_j^2}$$

and I can introduce that constraint that the expected norm after the transformation should be 1. In such a case when noise is added after the transformation $\hat{X} + \eta$ where $\eta \sim$

$N(0, t^2)$ and achieves (ε, δ) -DP in this space then due to linearity of transformation the noise introduced in the original space is then given by

$$\hat{\eta} = (b_0\eta_0, b_1\eta_1, \dots, b_d\eta_d)$$

then the error is

$$\mathbb{E} [\|\hat{\eta}\|^2] = \sum_{j=1}^d b_j^2 \cdot t^2 = t^2 \sum_{j=1}^d b_j^2$$

I desire a transformation of $x_i^{(j)}$ such that the expected norm is 1. Thus I must scale each dimension by $\frac{1}{b_j}$, and have that

$$\mathbb{E} [\|x_i\|^2] = 1$$

Minimize $\|\hat{\eta}\|$ under the constraint that $\mathbb{E} [\|x_i\|^2] = 1$

An alternative Again minimize $\|\hat{\eta}\|$, but instead the constraint comes from Chebyshev's inequality, I always have that

$$\Pr \left[\left| \|x_i\|^2 - \mathbb{E} [\|x_i\|^2] \right| \geq k \cdot \sqrt{\text{Var} [\|x_i\|^2]} \right] \leq \frac{1}{k^2}$$

Therefore I can set $\frac{1}{k^2} = 0.05$, and find a transformation where I decide how many standard deviations I must be away from the mean to have norm greater than 1, i.e.

$$\mathbb{E} [\|x_i\|^2] + k \cdot \sqrt{\text{Var} [\|x_i\|^2]} = 1 \implies k = \frac{1 - \mathbb{E} [\|x_i\|^2]}{\sqrt{\text{Var} [\|x_i\|^2]}}$$

I then have that

$$\frac{1}{k^2} = \frac{1}{\left(\frac{1 - \mathbb{E} [\|x_i\|^2]}{\sqrt{\text{Var} [\|x_i\|^2]}} \right)^2} = \left(\frac{\sqrt{\text{Var} [\|x_i\|^2]}}{1 - \mathbb{E} [\|x_i\|^2]} \right)^2 = \frac{\text{Var} [\|x_i\|^2]}{(1 - \mathbb{E} [\|x_i\|^2])^2}$$

I already know that

$$\begin{aligned} \mathbb{E} [\|x_i\|^2] &= \sum_{j=1}^d \frac{\sigma_j^2}{b_j^2} \\ \text{Var} [\|x_i\|^2] &= 2 \sum_{j=1}^d \frac{\sigma_j^4}{b_j^4} \end{aligned}$$

I therefore have the constraint

$$\frac{\text{Var} [\|x_i\|^2]}{(1 - \mathbb{E} [\|x_i\|^2])^2} = \frac{2 \sum_{j=1}^d \frac{\sigma_j^4}{b_j^4}}{\left(1 - \sum_{j=1}^d \frac{\sigma_j^2}{b_j^2} \right)^2} = 0.05$$

Be aware that this could find cases where expected norm is greater than 1 and the constraint then says that they are never less than 1 in norm. Another restriction could be that expected norm is ≤ 1 .

Chebyshev's inequality could also be used when expected norm should be 1 and then put a bound on number of std away one must be to have less than 0.05 fraction of data removed.

References

- [1] BATES, D. *Quadratic Forms of Random Variables*. University of Wisconsin-Madison: STAT 849 lectures, 2010.
- [2] BISWAS, S., DONG, Y., KAMATH, G., AND ULLMAN, J. Coinpress: Practical private mean and covariance estimation. *Advances in Neural Information Processing Systems 33* (2020), 14475–14485.
- [3] CHEN, T., AND LUMLEY, T. Numerical evaluation of methods approximating the distribution of a large quadratic form in normal variables. *Computational Statistics & Data Analysis 139* (2019), 75–81.
- [4] DWORK, C., MCSHERRY, F., NISSIM, K., AND SMITH, A. Calibrating noise to sensitivity in private data analysis. *Journal of Privacy and Confidentiality 7*, 3 (2016), 17–51.
- [5] DWORK, C., ROTH, A., ET AL. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science 9*, 3–4 (2014), 211–407.
- [6] HALLAM, A. *Some Theorems on Quadratic Forms and Normal Variables*. Iowa State University: Econ 671 lectures, 2004.
- [7] HANSON, D. L., AND WRIGHT, F. T. A bound on tail probabilities for quadratic forms in independent random variables. *The Annals of Mathematical Statistics 42*, 3 (1971), 1079–1083.
- [8] HUANG, Z., LIANG, Y., AND YI, K. Instance-optimal mean estimation under differential privacy.
- [9] MATHAI, A., AND PROVOST, S. Quadratic forms in random variables.
- [10] PAGH, R., AND LEBEDA, C. Private vector aggregation when coordinates have different sensitivity.