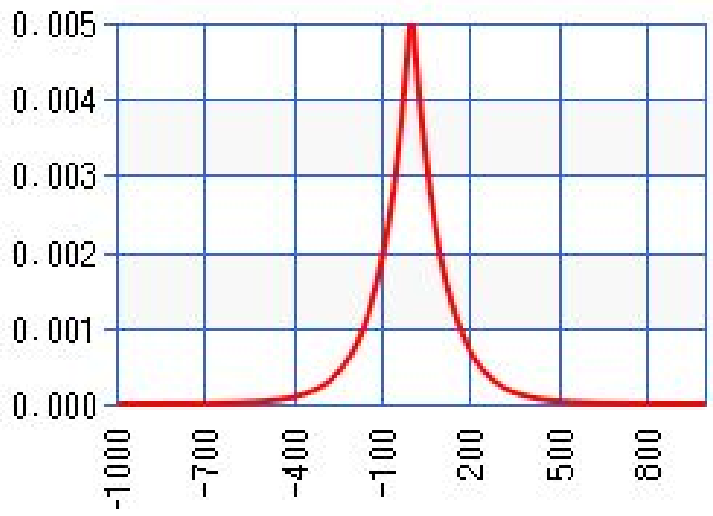
ROUET Gwenn
KIROV Mihail
LEVIFVE Gaspard

Privacy

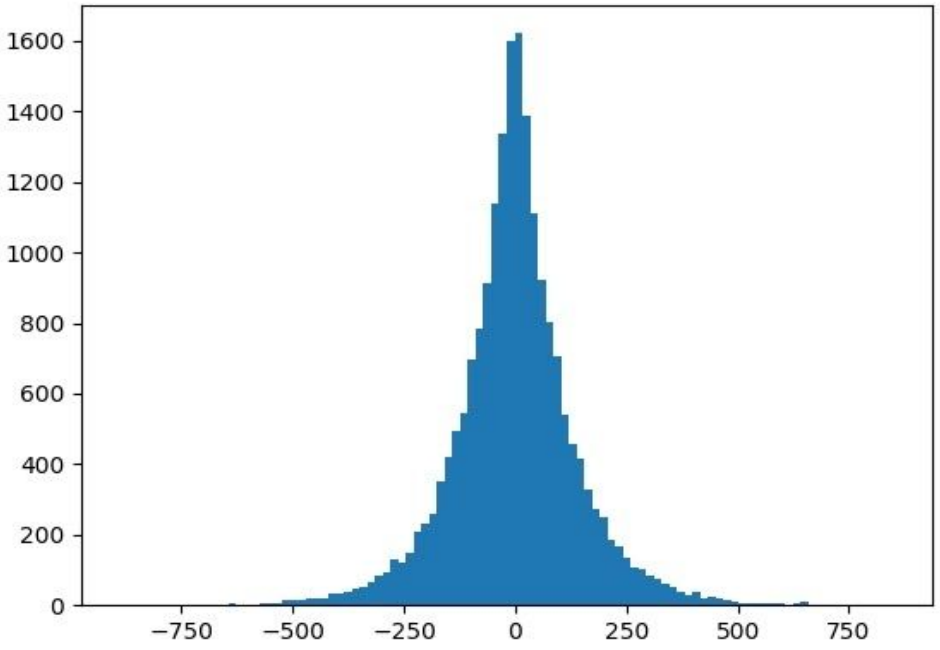
Implementing the Laplace mechanism

2 Test your implementation

- Compare your graph to the Laplace distribution that you should obtain (where b is sensitivity/) : http://keisan.casio.com/exec/system/1180573177



Laplace repartition as presented on the website with b = 100



Histogram of the number of individuals in bins of size 100

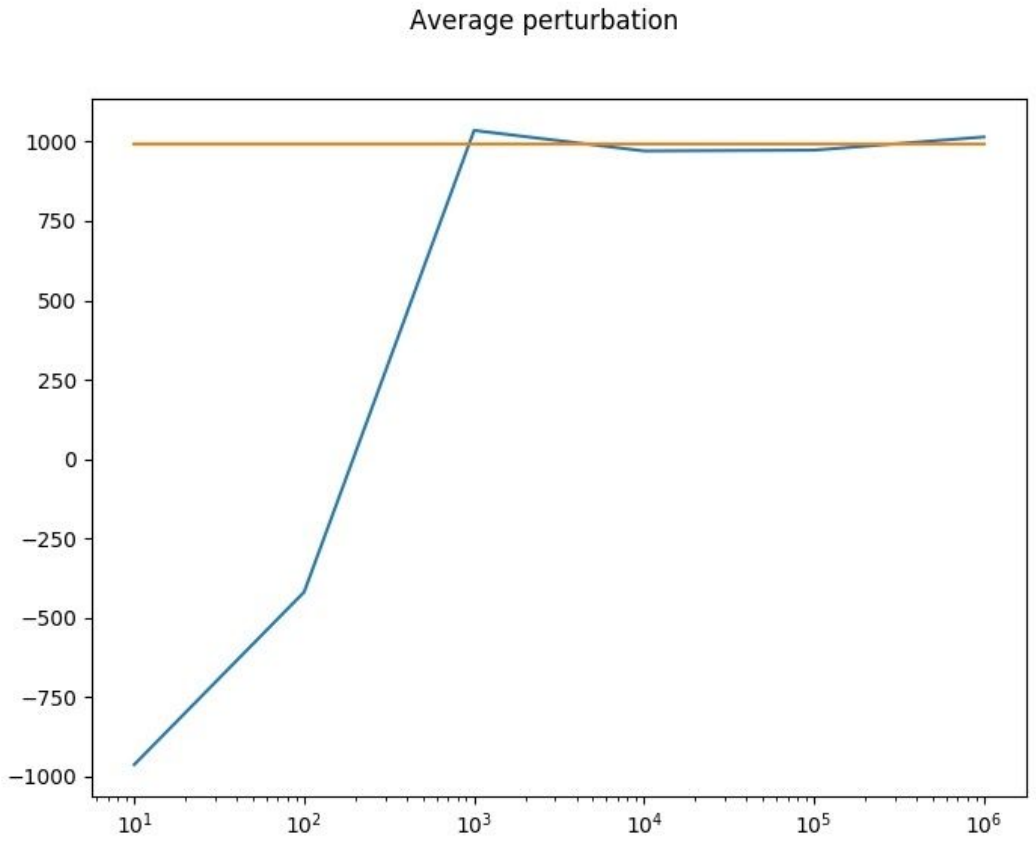
Our histogram with the sensitivity equal to 100 has similar shape and width as the theoretical Laplace graph as observed on the keisan.casio website with standard deviation equal to 100.

We can confirm that the distribution of the set of errors is laplacian and the ratio between the sensitivity and the budget to the standard deviation.

3 Analyze Laplace

5. Why must we limit the number of aggregates published ?

- Plot in a graph (e.g., Open Oce le) : on the x-axis the number of perturbations, and on the y-axis the averages. Plot also the true count



The graph of the averages in function of the number of perturbation, presented on a logarithmic scale

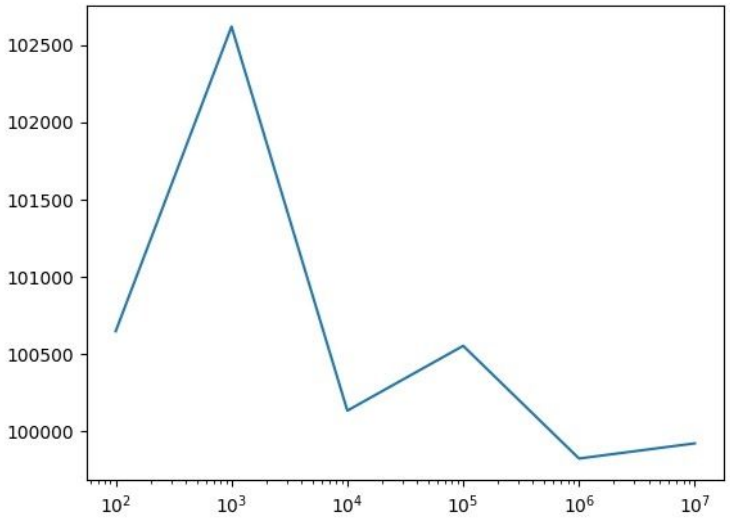
- How many perturbations are needed for being close to the true count (e.g., 10% difference)?

  On this graphic we can see that the average value of the perturbations converges to the real value of the count. The amount of perturbations needed is around 10^3 to be in 10% range of the true value.

  The more aggregates we publish, the more precise the estimate of the real unperturbed value is.That is due to the nature of the laplacian distribution of the perturbations which is 0 centered.That's why we have to limit the number of queries.

6. How big is the error due to the perturbation ?

- Does erravg depend on the dataset size ? On the dataset values ?



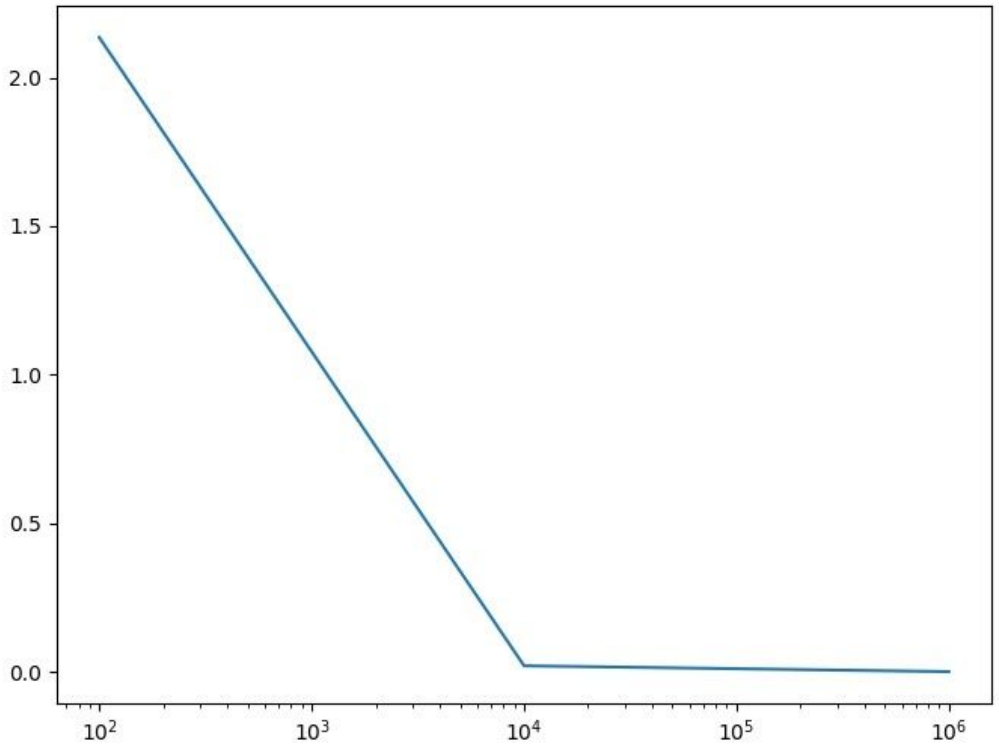
Graph of the err_avg in function of the dataset size

The error average is linked to the standard deviation of the Laplace law, which depends on the sensitivity. As the sensitivity depends on the maximum of all the values, **the error doesn't depend on the size but only on the values** : the greater the values are, the greater the sensitivity will be. As the values are generated in a fixed size interval with the same maximum value (in our case the sensitivity), the error converges always to the same value.

- What is the ratio between error average and the Laplace scale factor parameter (i.e., sensitivity/e) ?

  As the absolute value of the error average converges to the Laplace scale parameter, the ratio between the error and the scale factor parameter converges to 1.

- Plot in a graph the evolution of the ratio with respect to the various dataset sizes n.

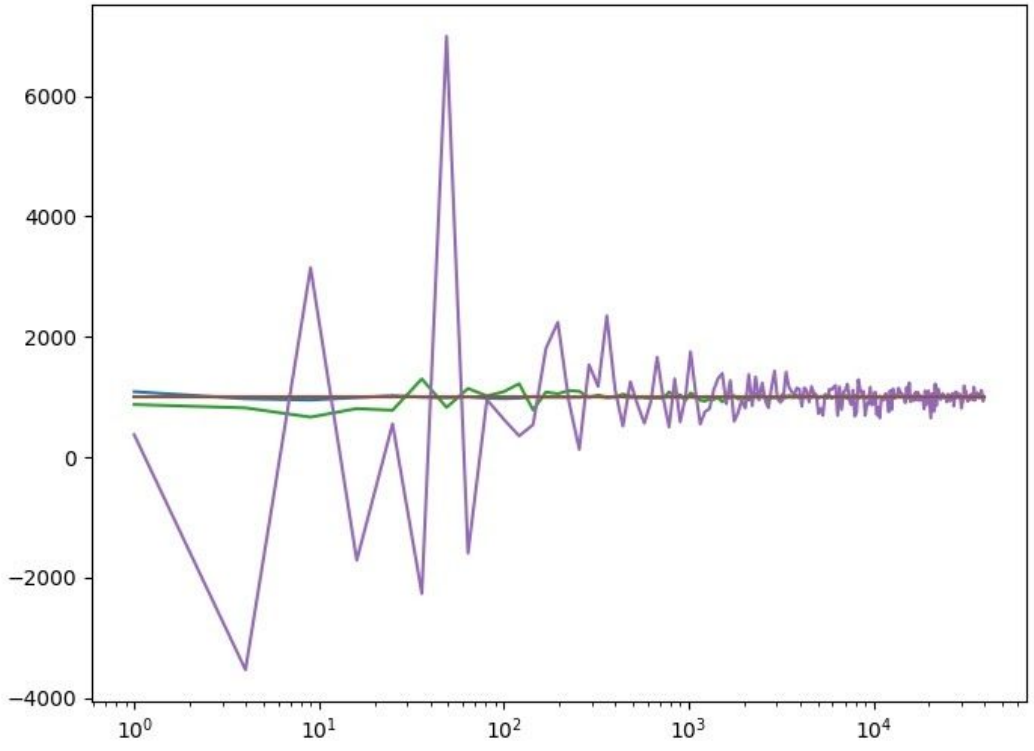


Graph of the ration in function of the dataset size

- How many tuples are “ needed “ for making this ratio small enough ? (e.g., 10%)

  The amount of tuples “needed” is between 10^3 and 10^4

Bonus :



budget = 10^(-6)

budget = 10^(-4)

budget = 10^(-2)

Graphic of the averaged perturbation ,in function of the number of aggregates, representing the convergence of the perturbed count to the real one with different budgets(scale factors).