Prof. Christian Cachin
Orestis Alpos, Luca Zanolini

Exercise 7

7.1 Differential privacy – Theory (4pt)

Let $\mathcal{X}^n = \{0,1\}^n$ be the dataset, where we write $X^n = [x_1, \ldots, x_n]$ for $X^n \in \mathcal{X}^n$. Suppose a sanitization function $M : \mathcal{X} \to \mathcal{Y}$, with $\mathcal{Y} = \{0,1\}$, is as follows:

$$M(X^n) \ = \ \begin{cases} x_1 & \text{with probability } \delta \\ R & \text{with probability } 1-\delta, \end{cases}$$

where $R \xleftarrow{R} \{0,1\}$ is a uniformly random bit. In other words, M leaks the first entry in the dataset with probability δ and returns a random bit otherwise. How much differential privacy does M have?

7.2 Differential privacy – Practice (6pt)

Consider again the dataset `ex05-fake-registrations.csv`. Compute three differentially private histogram series using the Laplace mechanism for varying ε.

In particular, for each $\varepsilon \in \{0.1, 0.5, 2\}$, compute a ε-differentially private histogram of the dataset on

a) attribute *Ort*;

b) attribute *System*; and

c) attribute *Points*, where the histogram contains bins of width 10, that is, for the intervals 0–9, 10–19, . . . , 90–99.

Develop either your own implementation or exploit the material available in this online book, titled *Programming Differential Privacy*, in Chapter *Properties of Differential Privacy*:

`https://programming-dp.com/notebooks/ch4.html`