7.1 Differential Privacy - Theory

From the lecture we know:

$$\frac{P[\mathcal{M}(X^n) \in Y]}{P[\mathcal{M}(\overline{X}^n) \in Y]} \le e^{\epsilon}$$

We can compute the probabilities of numerator and denumerator:

$$\begin{aligned} P[\mathcal{M}(X^n) = 0] &= \delta * P[x_1 = 0] + (1-\delta) * P[R = 0] \\ &= \delta * P[x_1 = 0] + \frac{1-\delta}{2} \end{aligned}$$

$$\begin{aligned} P[\mathcal{M}(\overline{X}^n) = 0] &= \delta * P[\overline{x}_1 = 0] + (1-\delta) * P[R = 0] \\ P[\overline{x}_1 = 0] &= \frac{n-1}{n} * P[x_1 = 0] + \frac{1}{n} * (1 - P[x_1 = 0]) \\ \Rightarrow P[\mathcal{M}(\overline{X}^n) = 0] &= \delta * (\frac{n-1}{n} * P[x_1 = 0] + \frac{1}{n} * (1 - P[x_1 = 0])) + \frac{1-\delta}{2} \end{aligned}$$

The fraction is greatest if the nominator is big and the denominator is small, therefore we can compute an upper bound and lower bound respectively :

$$\begin{aligned} P[\mathcal{M}(X^n) = 0] &\le \delta + \frac{1-\delta}{2} && \textit{for } P[x_1 = 0] = 1 \\ P[\mathcal{M}(\overline{X}^n) = 0] &\ge \frac{\delta}{n} + \frac{1-\delta}{2} && \textit{for } P[x_1 = 0] = 0 \end{aligned}$$

Therefore we can compute the ϵ:

$$\begin{aligned} e^{\epsilon} \ge \frac{P[\mathcal{M}(X^n) \in Y]}{P[\mathcal{M}(\overline{X}^n) \in Y]} &\le \frac{\frac{1-\delta}{2}}{\frac{\delta}{n} + \frac{1-\delta}{2}} \\ \Leftrightarrow \frac{P[\mathcal{M}(X^n) \in Y]}{P[\mathcal{M}(\overline{X}^n) \in Y]} &\le \frac{1-\delta}{\frac{2\delta}{n} + 1 - \delta} \\ \Leftrightarrow \quad &= \frac{1 - \delta - \frac{2\delta}{n} - 1 + \delta}{\frac{2\delta}{n} + 1 - \delta} + 1 \\ \Leftrightarrow \quad &= \frac{-\frac{2\delta}{n}}{\frac{2\delta}{n} + 1 - \delta} + 1 \\ \Leftrightarrow \quad &= \frac{-2\delta}{2\delta + n - \delta n} + 1 \\ \Leftrightarrow \quad &= \frac{2\delta}{2(n-1)\delta - n} + 1 \quad \sim 1 \textit{ for big } n \end{aligned}$$

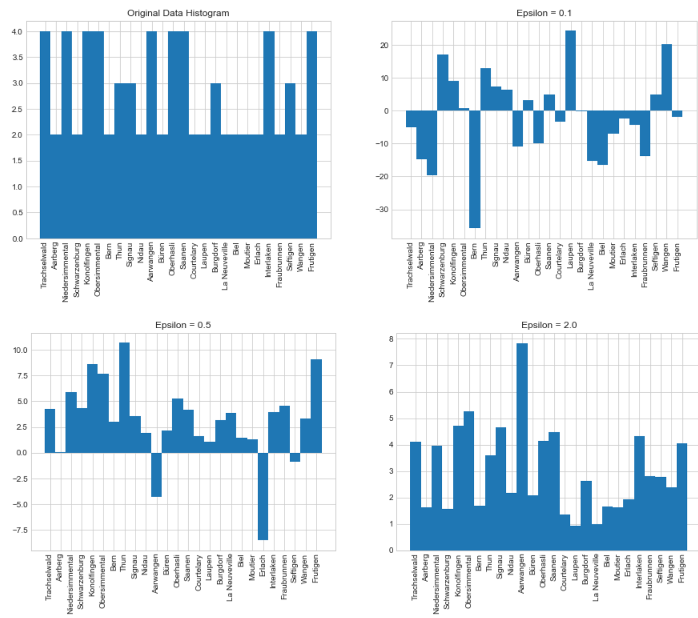
Because we get, that the formula is approximately around 1, we can approximate this with $1 + x \approx e^x$. Therefore:

$$\frac{2\delta}{2(n-1)\delta - n} + 1 \approx e^{\frac{2\delta}{2(n-1)\delta - n}} \textit{, therefore } \epsilon \approx \frac{2\delta}{2(n-1)\delta - n} \; (\approx \frac{1}{n}) \textit{ for big } n$$

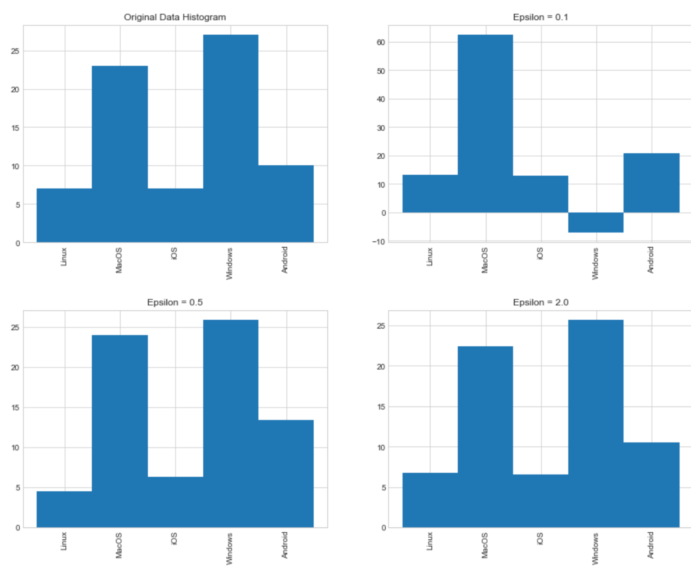
7.2 Differential Privacy - Practice

The corresponding Jupyter Notebook is also handed in.

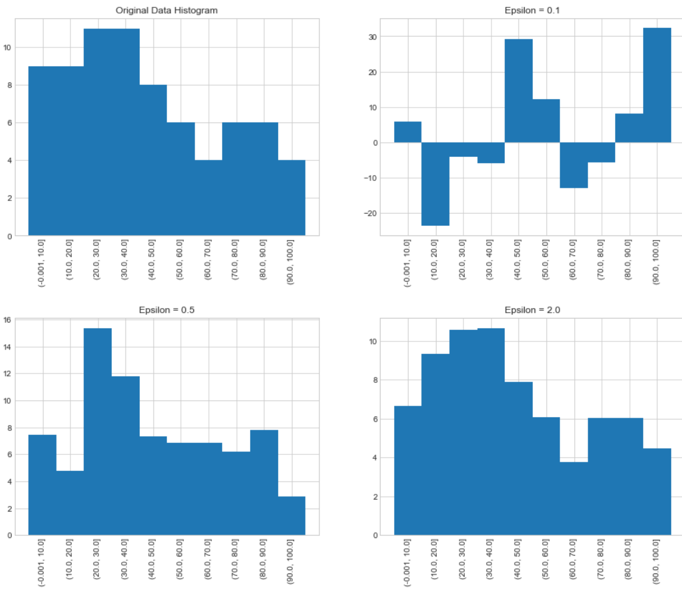
7.2.a ϵ-differential histogramm on attribute ORT



7.2.b ϵ-differential histogramm on attribute SYSTEM



7.2.c ϵ-differential histogramm on attribute POINTS



The higher the ϵ value is the more do the original and the newly calculated look alike. If the ϵ is very small the histogram looks very different but it can happen that the information one can get from the histogram is not usable at all.