

# CS 281 Homework 3

Instructor: Carlos Guestrin

agataf@cs.stanford.edu

Available: 05/22/2023; Due: 6:00 PM PST, 05/31/2023

---

## 1 $\epsilon$ -DP mean estimation

Assume we are given a database of medical records  $X \in \{0, 1\}^{n \times d}$  where each record  $X_i$  is a vector  $(x_{i1}, \dots, x_{id})$ , where  $x_{ij} \in \{0, 1\}$  is a boolean denoting whether a person  $i$  has a medical condition  $j$  (diabetes, hypertension, chronic kidney disease etc.) or not.

We are interested in identifying an  $\epsilon$ -DP mechanism for calculating the prevalence of each of the  $d$  medical conditions in the dataset. In other words, we're interested in a mechanism  $M(X)$  which approximates  $f(X) = \mathbb{E}[X] = \frac{1}{n} \sum_{i=1}^n X_i$ , such that, for any two datasets  $X, X'$  which differ in exactly one entry, and all possible prevalence vectors  $T \in \mathbb{R}^d$ ,

$$\frac{P(M(X) \in T)}{P(M(X') \in T)} \leq \exp(\epsilon)$$

### 1.1 Intuition [2 points]

Explain, in plain language, what the  $\epsilon$ -DP mechanism would guarantee in our case. (1-2 sentences).

### 1.2 Univariate case [11 points]

Consider a simplified, univariate version of this problem, where you're trying to estimate the prevalence of diabetes ( $j = 0$ ), using the Laplace mechanism

$$M(X) = f(X) + W$$

where

$$W \sim \text{Laplace}\left(\frac{\Delta f}{\epsilon}\right)$$

(a). **Sensitivity and noise [2 points]**. Calculate  $\Delta f$  and state the distribution of  $W \sim \text{Laplace}()$ , using variables  $n, d, \epsilon$ .

(b). **Interpretation [2 points]**. Give a plain-language, one-sentence interpretation of the error value. Consider how it relates to the size of the dataset, and the parameter  $\epsilon$ , and what it means for the learning rate.

(c). **Implementation. [6 points]** For the simulated dataset in the starter code, implement the Laplace mechanism for calculating prevalence of diabetes for three different  $\epsilon$  values: 0.01, 0.1 and 1. For each  $\epsilon$ , calculate the mean 1000 times, and plot the resulting distributions of mean estimates (using a histogram with 30 bins). Hint: use the `np.random.laplace` function with appropriate parameters.

(d). **Interpretation. [1 point]** How do the plots change across the values of  $\epsilon$ ?

### 1.3 Multivariate case [10 points]

Consider a full version of this problem, where the Laplace mechanism

$$M(X) = f(X) + (W_1, \dots, W_d)$$

where  $W_j$  are independent Laplace random variables  $W_j \sim \text{Laplace}\left(\frac{\Delta f}{\varepsilon}\right)$

(a). **Sensitivity and noise [2 points]**. Calculate  $\Delta f$  and state the distribution of  $W_j \sim \text{Laplace}()$ , using variables  $n, d, \varepsilon$ .

(c). **Implementation. [6 points]** For the simulated dataset in the starter code, implement the Laplace mechanism for calculating prevalence of all 10 diseases for three different  $\varepsilon$  values: 0.01, 0.1 and 1. For each  $\varepsilon$ , calculate the mean 1000 times, and plot the resulting distributions of mean estimates (using a histogram with 30 bins).

(d). **Interpretation. [2 points]** How do the plots change across the values of  $\varepsilon$ ? Compare your plots to those generated in part 1.2. How do the plots for the first disease (diabetes) differ in the multivariate case from the univariate case?

## 2 $(\varepsilon, \delta)$ -DP mean estimation

Recall the Gaussian mechanism satisfies approximate DP.

Given  $f : \mathcal{X}^n \rightarrow \mathbb{R}^k$ , the Gaussian mechanism outputs

$$M(X) = f(X) + (Y_1, \dots, Y_d)$$

where  $Y_i \sim \mathcal{N}(0, \sigma^2)$ ,  $\sigma^2 = \frac{2 \ln(1.25/\delta) \cdot (\Delta_2 f)^2}{\varepsilon^2}$ , and  $\Delta_2 f$  is the  $\ell_2$  sensitivity of the function. The Gaussian mechanism is  $(\varepsilon, \delta)$ -DP.

### 2.1 The Gaussian mechanism [12 points]

(a). **Sensitivity and noise [4 points]**. Calculate  $\Delta_2 f$  and state the distribution of  $W \sim \mathcal{N}()$ , using variables  $n, d, \varepsilon, \delta$ .

(b). **Implementation. [6 points]** For the simulated dataset in the starter code, implement the Gaussian mechanism for calculating prevalence of all 10 diseases for three different  $\varepsilon$  values: 0.01, 0.1 and 1, and  $\delta = 0.01$ . For each  $\varepsilon$ , calculate the mean 1000 times, and plot the resulting distributions of mean estimates. Hint: use the `np.random.normal` function with appropriate parameters.

(d). **Interpretation. [2 points]** How do they change across the values of  $\varepsilon$ ? How is it different from the Laplace mechanism?