Laplace Noise Mechanism

Reminder: Counting queries

Given a predicate $q : \mathcal{X} \to \{0, 1\}$ (e.g., “smoker?”), we define the corresponding (normalized) *counting query*

$$q(X) = \frac{1}{n} \sum_{i=1}^{n} q(x_i).$$

A *workload* Q of counting queries is given by predicates $q_1, \ldots, q_k$.

$$Q(X) = \begin{pmatrix} q_1(X) \\ \vdots \\ q_k(X) \end{pmatrix} \in [0, 1]^k.$$

E.g., “smoker?”, “smoker and over 30?”, “smoker and heart disease?”, etc.

"smoker, heart disease, over 30?"

Answering counting queries with Randomized Response

How can I answer k counting queries w/ ε-DP using RR?

k instances of RR (one per query) with $\left(\frac{\varepsilon}{k}\right)$-DP } whole mechanism is ε-DP by composition thm

All queries get answers w/ $\leq \pm\alpha$ error with prob. $1-\delta$

as long as $n \gg \frac{k^2 \log(k/\delta)}{\alpha^2 \varepsilon^2}$

Exercise: Do the details!

Sensitivity

e.g. $f(X) = Q(X) = \begin{pmatrix} q_1(X) \\ \vdots \\ q_k(X) \end{pmatrix}$

The ℓ_1 *sensitivity* of $f : \mathcal{X}^n \to \mathbb{R}^k$ is

$$\Delta_1 f = \max_{X \sim X'} \|f(X) - f(X')\|_1 = \max_{X \sim X'} \sum_{i=1}^{k} |f(X)_i - f(X')_i|$$

is neighbouring

Measure of how much a person can influence f.

e.g. if $f : \mathcal{X}^n \to \mathbb{R}$

then $\Delta_1 f = \max_{X \sim X'} |f(X) - f(X')|$

Sensitivity of a workload of counting queries

Suppose $f(X) = Q(X) = \begin{pmatrix} q_1(X) \\ \vdots \\ q_k(X) \end{pmatrix}$ for a workload of k counting queries

Give an upper bound on $\Delta_1 Q$

$q_i(X) = \frac{1}{n} \sum_{j=1}^{n} q_i(x_j)$ $\qquad q_i(x_j) \in \{0,1\}$ $\qquad \Delta_1 q_i = \frac{1}{n} \quad \forall i$

$$\Delta_1 Q = \max_{X \sim X'} \sum_{i=1}^{k} \underbrace{|q_i(X) - q_i(X')|}_{\leq 1/n} \leq \boxed{\frac{k}{n}}$$

Laplace noise mechanism

The Laplace noise mechanism $\mathcal{M}_{\mathrm{Lap}}$ (for a function $f : \mathcal{X}^n \to \mathbb{R}^k$) outputs

$$\mathcal{M}_{\mathrm{Lap}}(X) = f(X) + Z,$$

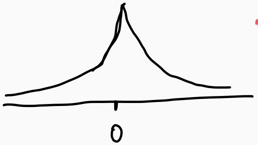
where $Z \in \mathbb{R}^k$ is sampled from $\mathrm{Lap}(0, \frac{\Delta_1 f}{\varepsilon})$.

$Z_1, \dots, Z_k$ are independent

Z_i is from a one-dimensional $\mathrm{Lap}(0, \frac{\Delta_1 f}{\varepsilon})$

$\mathrm{Lap}(\mu, b)$ is the *Laplace distribution* on $\mathbb{R}^k$ with expectation $\mu \in \mathbb{R}^k$ and scale $b > 0$, and has pdf

$$p(z) = \frac{1}{(2b)^k} e^{-\|z-\mu\|_1/b} = \frac{1}{(2b)^k} \exp\left(-\frac{1}{b}\sum_{i=1}^{k} |z_i - \mu_i|\right)$$



$\boxed{\mathcal{M}_{\mathrm{Lap}} \text{ is } \varepsilon\text{-DP}}$

Privacy of the Laplace noise mechanism

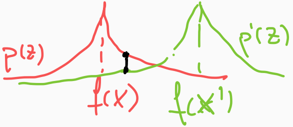
For any f, $\mathcal{M}_{\text{Lap}}$ is ε-DP.

$\mathcal{M}_{Lap}(X) = f(X) + Z \sim Lap\left(f(X), \frac{\Delta_1 f}{\varepsilon}\right)$

Let $X \sim X'$, and let p be the pdf of $\mathcal{M}_{Lap}(X)$, and p' the pdf of $\mathcal{M}_{Lap}(X')$. $= f(X') + Z \sim Lap\left(f(X'), \frac{\Delta_1 f}{\varepsilon}\right)$

$$p(z) = \left(\frac{\varepsilon}{2\Delta_1 f}\right)^k e^{-\varepsilon\|z-f(X)\|_1/\Delta_1 f} \qquad p'(z) = \left(\frac{\varepsilon}{2\Delta_1 f}\right)^k e^{-\varepsilon\|z-f(X')\|_1/\Delta_1 f}$$

Claim: enough to show $\max_{z\in\mathbb{R}^k} \frac{p(z)}{p'(z)} \leq e^{\varepsilon}$



$\forall S \subseteq \mathbb{R}^k$

$$\mathbb{P}(\mathcal{M}_{Lap}(X) \in S) = \int_S p(z)\,dz \leq \int_S e^{\varepsilon} p'(z)\,dz$$

$$= e^{\varepsilon} \int_S p'(z)\,dz = e^{\varepsilon}\, \mathbb{P}(\mathcal{M}(X') \in S).$$

Privacy of the Laplace noise mechanism

Need: $\forall z \quad \frac{p(z)}{p'(z)} \le e^{\varepsilon}$

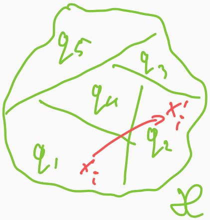
$$p(z) = \cancel{\left(\frac{\varepsilon}{2\Delta_1 f}\right)^k} e^{-\varepsilon\|z-f(X)\|_1/\Delta_1 f} \qquad p'(z) = \cancel{\left(\frac{\varepsilon}{2\Delta_1 f}\right)^k} e^{-\varepsilon\|z-f(X')\|_1/\Delta_1 f}$$

$$\frac{p(z)}{p'(z)} = \exp\left(\frac{\varepsilon}{\Delta_1 f}\left(\underbrace{\|z-f(X')\|_1 - \|z-f(X)\|_1}_{\le \|f(X)-f(X')\|_1}\right)\right) \le e^{\varepsilon}$$

$$\|z - f(X')\|_1 \le \|z - f(X)\|_1 + \|f(X) - f(X')\|_1$$ Triangle inequality

Histograms

Suppose the query workload $Q = \{\overbrace{q_1, \dots, q_k}^{\text{counting queries}}\}$ "partitions" the data.



- $\forall x \in \mathcal{X}$: at most one of $q_1(x), \dots, q_k(x)$ equals 1.
- e.g., "votes for the Liberal Party per riding"

What is the sensitivity?

$$X = \{x_1, \dots, x_n\}$$
$$\sim$$
$$X' = \{x_1, \dots, x_i', \dots, x_n\}$$

$$\max_{X \sim X'} \|Q(X) - Q(X')\|_1 = \max_{X \sim X'} \sum_{i=1}^{k} |q_i(X) - q_i(X')| \le \frac{2}{n}$$

because ≤ 2 query values change by $\le \frac{1}{n}$ each.

Accuracy of the Laplace noise mechanism

Generalizes to "k-norm mechanism"

If $Z \in \mathbb{R}^k$ is a Laplace random variable from $\mathrm{Lap}(\mu, b)$, then, for every i Hardt, Talwar

$$\mathbb{P}(|Z_i - \mu_i| \geq t) = e^{-t/b}.$$

$$\mathcal{M}_{Lap}(X) \sim Lap\left(f(X), \frac{\Delta_1 f}{\varepsilon}\right)$$

$$\mathbb{P}\left(\max_i |\mathcal{M}_{Lap}(X)_i - f(X)_i| \geq \alpha\right) \leq \sum_{i=1}^{k} \mathbb{P}(|\mathcal{M}_{Lap}(X)_i - f(X)_i| \geq \alpha)$$

$$\leq k \cdot e^{-\alpha\varepsilon/\Delta_1 f}$$

E.g. if $f(X) = Q(X)$, then $\Delta_1 f \leq \frac{k}{n}$,

i.e., answers to k counting queries

$$\mathbb{P}(\max \text{ error} \geq \alpha) \leq k \cdot e^{-\frac{\alpha\varepsilon n}{k}} \leq \beta$$

if $n \geq \dfrac{k \ln(k/\beta)}{\alpha\varepsilon}$