Lecture 4
Intro to Differential Privacy, Part 2

Senstivity

$f: \mathcal{X}^n \to \mathbb{R}^k$, ℓ_1-sensitivity

$$\Delta^{(f)} = \max_{x,x' \text{ neighbour}} \|f(x) - f(x')\|_1$$

$\sqrt{k}$ diff (mult) w/ ℓ_1 vs ℓ_2

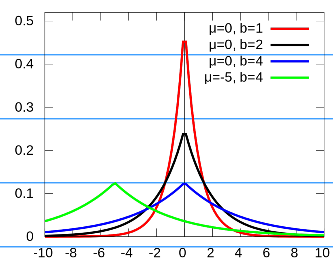
$f(X) = \frac{1}{n}\sum X_i$, $X_i \in \{0,1\}$ $\Delta = \frac{1}{n}$

Laplace Distribution = Two-sided exp. dist.

Params: Location $= 0$ $\sigma^2 = 2b^2$

Scale $= b$

$$p(x) = \frac{1}{2b}\exp\left(-\frac{|x|}{b}\right) \leftarrow \text{Lap}(b)$$



Exp: $x \in [0,\infty)$, $p(x) \propto \exp(-cx)$

Lap: $x \in \mathbb{R}$, $p(x) \propto \exp(-c|x|)$

Gaussian: $x \in \mathbb{R}$, $p(x) \propto \exp(-cx^2)$

Laplace Mechanism

$f: \mathcal{X}^n \to \mathbb{R}^k$. Laplace Mech:

$$M(X) = f(X) + (Y_1, \ldots, Y_k)$$

$Y_i \underset{iid}{\sim} \text{Lap}(\Delta/\varepsilon)$

Fact: If $Y \sim \text{Lap}(b)$

$\Pr[|Y| \geq tb] = \exp(-t)$

$\Pr[|Y| \geq b\log(1/\beta)] = \beta$

$f = \frac{1}{n}\sum X_i$. $\Delta = 1/n$. $k = 1$

$\tilde{p} = f(X) + \text{Lap}(1/n\varepsilon)$ $p = f(X)$

$E[\hat{p}] = p + E[\text{Lap}(1/n\varepsilon)] = p$

$\text{Var}[\hat{p}] = \text{Var}[\text{Lap}(1/n\varepsilon)] = \frac{2}{n^2\varepsilon^2}$

Chebyshev's $|p - \tilde{p}| \leq O\left(\frac{1}{n\varepsilon}\right)$ w/ reasonable prob

$|p - \hat{p}| \leq \frac{\log(1/\beta)}{n\varepsilon}$ w.p. $1-\beta$

$\text{Error} \leq O\left(\frac{1}{n\varepsilon}\right)$ ✓ vs RR $\leq O\left(\frac{1}{\sqrt{n}\varepsilon}\right)$

Privacy Proof

Lap. Mech is ε-DP.

X, Y be neighbouring. $M(X) \leftarrow P_X(z)$, $M(Y) \leftarrow P_Y(z)$ } PDF of Lap Mech on X, Y

$\forall X, Y$ neighbouring, $\forall z$, ratio p is bounded

$$\frac{P_X(z)}{P_Y(z)} = \frac{\prod_{i=1}^{k} \exp\left(-\frac{\varepsilon |f(X)_i - z_i|}{\Delta}\right)}{\prod \exp\left(-\frac{\varepsilon |f(Y)_i - z_i|}{\Delta}\right)}$$

$$= \prod_{i=1}^{k} \exp\left(\frac{\varepsilon}{\Delta}\left(|f(Y)_i - z_i| - |f(X)_i - z_i|\right)\right)$$

$$\leq \prod_{i=1}^{k} \exp\left(\frac{\varepsilon}{\Delta} |f(Y)_i - f(X)_i|\right)$$

$$= \exp\left(\frac{\varepsilon}{\Delta} \sum |f(X)_i - f(Y)_i|\right) = \exp\left(\frac{\varepsilon}{\Delta} \|f(X) - f(Y)\|_1\right)$$

$$\leq \exp(\varepsilon). \quad \square$$

Counting Queries

"How many people in X satisfy property P?"

$f(X) = \sum X_i$, $X_i = 1$ if P is true for i, 0 else

$\Delta = 1$,

$M(X) = \sum X_i + \mathrm{Lap}(1/\varepsilon)$.

Many q's?

k counting q's: $f = (f_1, \dots, f_k)$

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

$Y \sim$ vector of k $\mathrm{Lap}(\frac{k}{\varepsilon})$ $\frac{\Delta}{\varepsilon}$

$\Delta = k$,

$\|f(x) - f(y)\|_1 = \sum_{i=1}^{k} |1-0| = k$

$x \rightarrow y$

$f_1(x) = 0$ $f_1(y) = 1$

$f_2(x) = 0$ ...

$f_k(x) = 0$ $f_k(y) = 1$

1. Non-adaptive $\rightarrow$ Adaptive
2. Dinur-Nissim
   If $\Omega(n)$ q's, $O(\sqrt{n})$ noise, attack
   If $O(n)$ q's, $\theta(n/\varepsilon)$ noise, private
   $\theta(\sqrt{n}/\varepsilon)$ noise, private adv. comp.

Histograms

"How many people are X years old?" $\forall X$

$f(X) = (f_0(X), \dots, f_{k-1}(X))$

$f_i(x) =$ Is x i-years old?

$\Delta = 2$

$x \rightarrow y$

$f_1(x) = 1$ $f_1(y) = 0$

$f_2(x) = 0$ $= 0$

$\vdots$ $\vdots$

$= 0$ $= 1$

$\Delta = \max \|f(X) - f(Y)\|_1 = \sum |f_i(X) - f_i(Y)|$

$= |1-0| + |0| + \dots + |0-1|$

$= 2$

$M(X) = f(X) + (Y_1, \dots, Y_k)$ $\mathrm{Lap}(2/\varepsilon)$

$|f_i(X) - M(X)_i| \leq \frac{2}{\varepsilon} \log(1/\beta)$ w.p $1-\beta$

$\leq \frac{2}{\varepsilon} \log(k/\beta)$ w.p $1 - \frac{\beta}{k}$

$\forall i$ at same time $|f_i(X) - M(X)_i| \leq \frac{2\log(k/\beta)}{\varepsilon}$

Fact: If $Y \sim \mathrm{Lap}(b)$

$\Pr[|Y| \geq tb] = \exp(-t)$

$\Pr[|Y| \geq b\log(1/\beta)] = \beta$

Properties of Differential Privacy

Post Processing

Theorem 6. *Let $M : \mathcal{X}^n \to \mathcal{Y}$ be ε-differentially private, and let $F : \mathcal{Y} \to \mathcal{Z}$ be an arbitrary randomized mapping. Then $F \circ M$ is ε-differentially private.*

F is a dist over deterministic f's $\quad T \subseteq \mathcal{Z}$

$$\Pr[F(M(X)) \in T] = \mathop{E}_{f \sim F}\left[\Pr[M(X) \in f^{-1}(T)]\right]$$
$$\leq \mathop{E}_{f \sim F}\left[e^{\varepsilon} \Pr[M(X') \in f^{-1}(T)]\right]$$
$$= e^{\varepsilon} \Pr[F(M(X')) \in T] \quad \boxtimes$$

Group Privacy

X, X' differ in 1 entry.

Theorem 7. *Let $M : \mathcal{X}^n \to \mathcal{Y}$ be an ε-differentially private algorithm. Suppose X and X' are two datasets which differ in exactly k positions. Then for all $T \subseteq \mathcal{Y}$, we have*

$$\Pr[M(X) \in T] \leq \exp(k\varepsilon) \Pr[M(X') \in T].$$

$X^{(0)} = X, \; X^{(k)} = X', \; Y = X^{(0)} \sim X^{(1)} \sim \dots \sim X^{(k)} = X'$

$$\Pr[M(X^{(0)}) \in T] \leq \Pr[M(X^{(1)}) \in T] \cdot e^{\varepsilon}$$
$$\leq \Pr[M(X^{(2)}) \in T] e^{2\varepsilon}$$
$$\dots$$
$$\leq \Pr[M(X^{(k)}) \in T] e^{k\varepsilon} \quad \boxtimes$$

(Basic) Composition

Theorem 8. *Suppose $M = (M_1, \ldots, M_k)$ is a sequence of ε-differentially private algorithms, potentially chosen sequentially and adaptively. Then M is $k\varepsilon$-differentially private.*

X, X' n.brs $\quad y = (y_1, \ldots, y_k)$

$$\frac{\Pr[M(X) = y]}{\Pr[M(X') = y]} = \prod_{i=1}^{k} \frac{\Pr[M_i(X) = y_i \mid (M_1(X), \ldots, M_{i-1}(X)) = (y_1, \ldots, y_{i-1})]}{\Pr[M_i(X') = y_i \mid (M_1(X'), \ldots, M_{i-1}(X')) = (y_1, \ldots, y_{i-1})]}$$

$$\leq \prod_{i=1}^{k} e^{\varepsilon}$$

$$= e^{k\varepsilon} \qquad \square$$

Basic $\qquad$ Advanced.

$k\varepsilon \rightarrow O(\sqrt{k}\,\varepsilon)$