Lecture 5
Approximate Differential Privacy
Definition

$$\Pr[M(x) \in S] \leq \Pr[M(x') \in S] + \varepsilon$$

small ε, bad accuracy
large ε, weak privacy

Definition 1 (Approximate Differential Privacy). *An algorithm $M : \mathcal{X}^n \to \mathcal{Y}$ is (ε, δ)-differentially private (i.e., it satisfies* approximate differential privacy*) if, for all neighbouring databases $X, X' \in \mathcal{X}^n$, and all $T \subseteq \mathcal{Y}$,*

$$\Pr[M(X) \in T] \leq e^{\varepsilon} \Pr[M(X') \in T] + \delta.$$

$\delta = 0 \Rightarrow$ pure

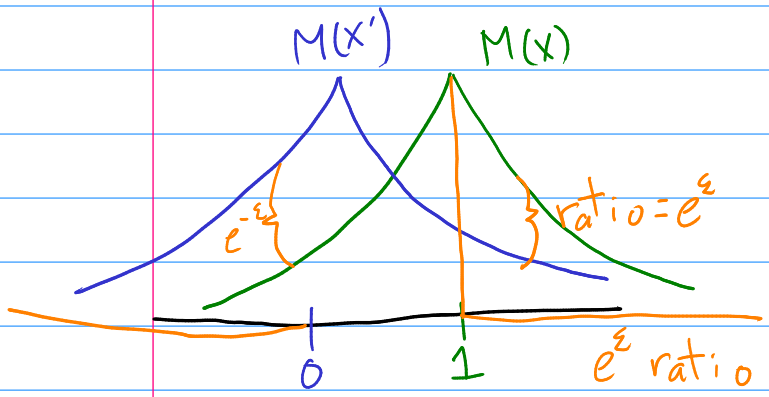
Privacy Loss Random Variable

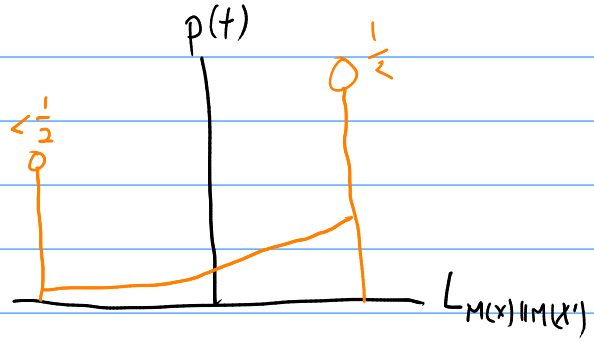
Definition 2. *Let Y and Z be two random variables. The* privacy loss random variable *$\mathcal{L}_{Y\|Z}$ is distributed by drawing $t \sim Y$, and outputting $\ln\left(\frac{\Pr[Y=t]}{\Pr[Z=t]}\right)$. If the supports of Y and Z are not equal, then the privacy loss random variable is undefined.*

- continuous
- undefined $\approx \infty$ priv loss
- $Y = M(x)$, $Z = M(x')$, x, x' nbrs

DP $\leftrightarrow$ Privacy Loss RV bounds

ε-DP of $M \leftrightarrow L_{M(x)\|M(x')} \leq \varepsilon$ w.p. 1 $\forall x, x'$ nbrs





(ε, δ)-DP $\leftrightarrow |L_{M(x)\|M(x')}| \leq \varepsilon$ w.p. $1-\delta$ $\forall x, x'$ nbrs

Lemma 3.17 [DR14]

Example 1: Release everything!

$M(X) = \begin{cases} 0 & \text{w.p } 1-\delta \\ X & \text{w.p. } \delta \end{cases}$

$M(X) = 0$ w.p $1-\delta$

$L_{M(X)\|M(X')} = 0$ when this happens

$(0, \delta)$-DP

δ should be small

Example 2: Name and Shame ($\delta > \frac{1}{n}$ bad!)

$NS_\delta(X) \rightarrow$ For each $x \in X$
- w.p. δ output x
- else, do nothing

Claim: NS_δ is $(0, \delta)$-DP

Claim 2: w.p $\approx \delta n$, NS_δ outputs some x

$1-(1-\delta)^n \approx \delta n << 1$ req. ($\delta << \frac{1}{n}$)

$\delta \approx \frac{1}{n^{1.1}}$? "Cryptographically small"

X, X' differ in entry i

T is a set of datapoints

E is event that entry i is output

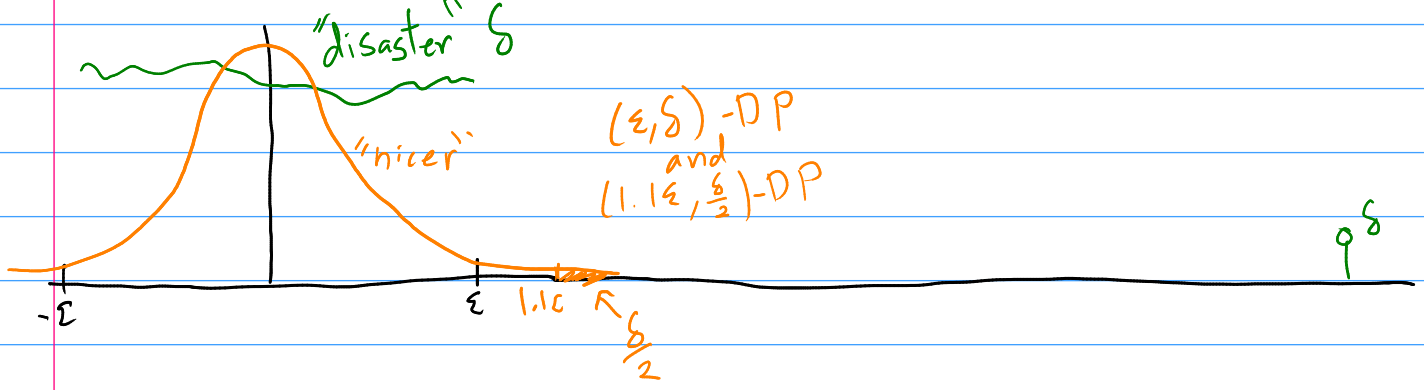
Conditioned on $\bar{E}$, $NS_\delta(X) = NS_\delta(X')$

$$\begin{aligned} Pr[NS_\delta(X) \in T] &= Pr[NS_\delta(X) \in T | E] \cdot Pr[E] \\ &\quad + Pr[NS_\delta(X) \in T | E] \cdot Pr[E] \\ &= Pr[NS_\delta(X') \in T | \bar{E}] \cdot Pr[\bar{E}] \\ &\quad + Pr[NS_\delta(X) \in T | E] \cdot Pr[E] \\ &\quad \leq 1 \qquad \leq \delta \\ &\leq Pr[NS_\delta(X') \in T] + \delta \end{aligned}$$

$(0, \delta)$-DP

Interpreting δ

- Pessimistic: disaster w.p. δ
- May not be so bad



Last comment on approximate DP

$\Pr[M(X) \in T] \leq e^{\varepsilon} \Pr[M(X') \in T] + \delta$

Suppose $M(X)$ outputs (X, s), s is uniform$\{1, \ldots, 1/\delta\}$

l_2-Sensitivity

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

l_2-sensitivity: $\Delta_2^{(f)} = \max_{\substack{x, x' \\ nbrs}} \|f(x) - f(x')\|_2$

$x \in \mathbb{R}^k$

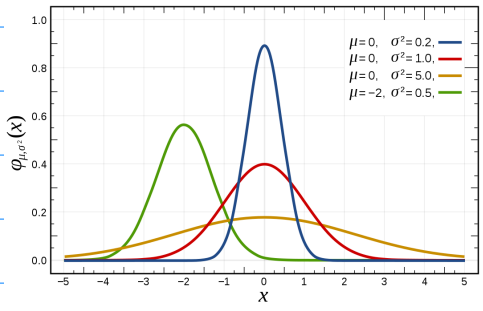
$\|x\|_2 \le \|x\|_1 \le \sqrt{k}\,\|x\|_2$

Gaussian Distribution

$\mathcal{N}(\mu, \sigma^2)$

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$\frac{1}{2b} \exp\left(-\frac{|x|}{b}\right)$$



Gaussian Mechanism

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

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

$$Y_i \sim N\left(0, \frac{\Delta_2^2}{\varepsilon^2} 2\ln(1.25/\delta)\right)$$

(ε, δ)-DP

$$\sigma^2 \geq \frac{\Delta_2^2}{\varepsilon^2} \cdot \log(1/\delta) \Rightarrow \sigma \simeq \frac{\Delta_2}{\varepsilon}\sqrt{\log(1/\delta)}$$

$$\text{Lap: } \sigma = \frac{\Delta_1}{\varepsilon}$$

Multivariate Estimation Example

$f(X) = \frac{1}{n}\sum X_i, \quad X \in \{0,1\}^{n \times d}$

Sensitivity: $X_1 = \vec{1} \Rightarrow X_1 = \vec{0}$

$\left\| \frac{1}{n}(\vec{1} - \vec{0}) \right\| = \left\| \frac{1}{n}\vec{1} \right\|$

$\Delta_1 = \frac{d}{n}, \quad \Delta_2 = \frac{\sqrt{d}}{n}$

Laplace Mechanism $f(X) + \text{Lap}\left(\frac{d}{\varepsilon n}\right)^{\otimes d}$

Error (l_2-norm) $= \left\| \text{Lap}\left(\frac{d}{\varepsilon n}\right)^{\otimes d} \right\|_2 \simeq \boxed{\frac{d^{3/2}}{\varepsilon n}}$

$(Y_1, \ldots, Y_d)$

$\uparrow \simeq \frac{d}{\varepsilon n}$

Gaussian Mech: $f(X) + N\left(0, \left(\frac{\sqrt{d}}{\varepsilon n}\right)^2\right)^{\otimes d}$

$\text{err}: \| \quad \|_2 \simeq \boxed{\frac{d}{\varepsilon n}}$

Gaussian Mechanism Privacy Proof

Theorem: $L_{M(X)\|M(X')} \leq \varepsilon$ w.p. $1-\delta$

$$M(x) = f(x) + N\left(0, \left(\frac{\Delta_2}{\varepsilon}\cdot\sqrt{\log(1/\delta)}\right)^2 \cdot I\right)$$

$Y \sim N(0,1)$, $Z \sim N(0,1)$ $\Rightarrow aY \sim N(0,a^2)$, $aY + bZ \sim N(0, a^2+b^2)$

Lemma 8. *Let $X, X' \in \mathcal{X}^n$ be neighbouring datasets, and let $M(Y) = f(Y) + N(0, \sigma^2 I)$ for some function $f : \mathcal{X}^n \to \mathbb{R}^k$. Then the privacy loss random variable between $M(X)$ and $M(X')$ is distributed as $N\left(\frac{\|f(X)-f(X')\|_2^2}{2\sigma^2}, \frac{\|f(X)-f(X')\|_2^2}{\sigma^2}\right)$.*

$f(X) - f(X') = v$, $x \sim N(0, \sigma^2 \cdot I)$

$$\ln\left(\frac{Pr[M(X) = f(X)+x]}{Pr[M(X') = f(X)+x]}\right) = \ln\left(\frac{\exp(-\|x\|_2^2/2\sigma^2)}{\exp(-\|x+v\|_2^2/2\sigma^2)}\right)$$

$$= \left(-\frac{1}{2\sigma^2}\right)\left(\|x\|_2^2 - \|x+v\|_2^2\right)$$

$$= \left(\frac{-1}{2\sigma^2}\right)\left(\sum_{j=1}^{k}\left(x_j^2 + (x_j+v_j)^2\right)\right)$$

$$= \left(\frac{1}{2\sigma^2}\right)\left(\sum_{j=1}^{k} 2x_j v_j + v_j^2\right)$$

$$\text{mean} = \frac{\sum v_j^2}{2\sigma^2} = \boxed{\frac{\|v\|_2^2}{2\sigma^2}}$$

k=1?

$x \sim N(0,\sigma^2)$

$\left(\frac{1}{2\sigma^2}\right)(2xv+v^2)$

$\text{mean} = \frac{v^2}{2\sigma^2}$

$\text{Var} = \left(\frac{2v}{2\sigma^2}\right)^2 \cdot \sigma^2 = \frac{v^2}{\sigma^2}$

$$\frac{1}{2\sigma^2}\left(\sum 2x_j v_j\right) = \boxed{\frac{1}{\sigma^2}\cdot y}, \quad y \sim N(0, \sigma^2\|v\|_2^2)$$

$$\sim N\left(0, \frac{\|v\|_2^2}{\sigma^2}\right)$$

$$\frac{1}{\sigma^2}\sum z_j \overset{?}{=}$$

$z_j \sim N(0, \sigma^2 v_j^2)$

$$L_{M(x)\|M(x')} = \frac{\|f(x)-f(x')\|_2}{\sigma} Z + \frac{\|f(x)-f(x')\|_2^2}{2\sigma^2}, \quad Z \sim N(0,1)$$

$$\sigma = \frac{\Delta_2 t}{\varepsilon}$$

$Pr[|Z| \geq \varepsilon] \leq \delta$

$$Pr\left[|Z| \geq \frac{\varepsilon\sigma}{\|f(x)-f(x')\|} - \frac{\|f(x)-f(x')\|_2}{2\sigma}\right] \leq Pr[|z| \geq t\cdot\tfrac{\varepsilon}{2t}] \approx Pr[|z| \geq t] \overset{?}{\leq} \delta$$

Useful fact

$\sigma = \frac{\Delta_2}{\varepsilon}\sqrt{\log(1/\delta)} \rightarrow (\varepsilon, \delta)\text{-DP}$

$Pr[Z \geq V] \leq \exp(-V^2/2)$

$t = \sqrt{2\log(2/\delta)} \rightarrow Pr[|Z| \geq t] \leq \delta$ ☒

Properties of Approximate DP

Post-Processing

Theorem 9. *Let $M : \mathcal{X}^n \to \mathcal{Y}$ be (ε, δ)-differentially private, and let $F : \mathcal{Y} \to \mathcal{Z}$ be an arbitrary randomized mapping. Then $F \circ M$ is (ε, δ)-differentially private.*

Group Privacy

Theorem 10. *Let $M : \mathcal{X}^n \to \mathcal{Y}$ be an (ε, δ)-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*

$$\Pr[M(X) \in T] \leq \exp(k\varepsilon) \Pr[M(X') \in T] + ke^{(k-1)\varepsilon}\delta.$$

Basic Composition

$\varepsilon_i = \varepsilon, \delta_i = \delta$

Theorem 11. *Suppose $M = (M_1, \ldots, M_k)$ is a sequence of algorithms, where M_i is $(\varepsilon_i, \delta_i)$-differentially private, and the M_i's are potentially chosen sequentially and adaptively.[2] Then M is $(\sum_{i=1}^k \varepsilon_i, \sum_{i=1}^k \delta_i)$-differentially private.*

$\Downarrow (k\varepsilon, k\delta)$-DP

Advanced $(\varepsilon\sqrt{8k\log(1/\delta')}, k\delta + \delta')$-DP

$\approx\approx (\varepsilon\sqrt{k}, k\delta)$-DP