Differential Privacy

ε

$\mathcal{D}$ $\quad$ $\mathcal{D}'$

$\downarrow$

$X(\mathcal{D}) \sim X(\mathcal{D}')$

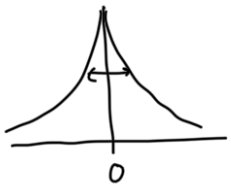
S $\quad$ $\Pr[X(\mathcal{D}) \in S] \leq e^{\varepsilon} \Pr[X(\mathcal{D}') \in S]$

$\Pr[X(\mathcal{D}') \in S] \leq e^{\varepsilon} \Pr[X(\mathcal{D}) \in S]$

Every cont. dist. has a density

$\mathcal{D} = \{0, 1, 0, 1, \ldots, 1\}$

$X(\mathcal{D}) = \underbrace{\#1's}_{T} + \underbrace{\text{Noise}}_{\approx \frac{1}{\varepsilon}}$ $\quad$ $f_{\mathcal{D}}$



$f_b(x) = \frac{1}{2b} e^{-|x|/b}$

$= \frac{\varepsilon}{2} e^{-\varepsilon|x|}$

D.P. vs Utility

$S = [a, b]$

$\Pr[X(\mathcal{D}) \in S] = \Pr[\text{Noise} + T \in [a, b]]$

$= \Pr[\text{Noise} \in [a-T, b-T]] = \int_{a-T}^{b-T} \frac{1}{2\lambda} e^{-|x|/\lambda} dx$

$\Pr[X(\mathcal{D}') \in S] = \int_{a-T'}^{b-T'} \frac{1}{2\lambda} e^{-|x|/\lambda} dx$

$\downarrow$

$\Delta = T - T'$

$y = x - \Delta$

$\downarrow$ $a - T' - T + T' = a - T$

$= \int_{a-T}^{b-T} \frac{1}{2\lambda} e^{-|x+\Delta|/\lambda} dx$

It is enough to show: $\left| \frac{e^{-|x|/\lambda}}{e^{-|x+\Delta|/\lambda}} \right| \leq e^{\varepsilon}$

$\lambda = 1/\varepsilon$

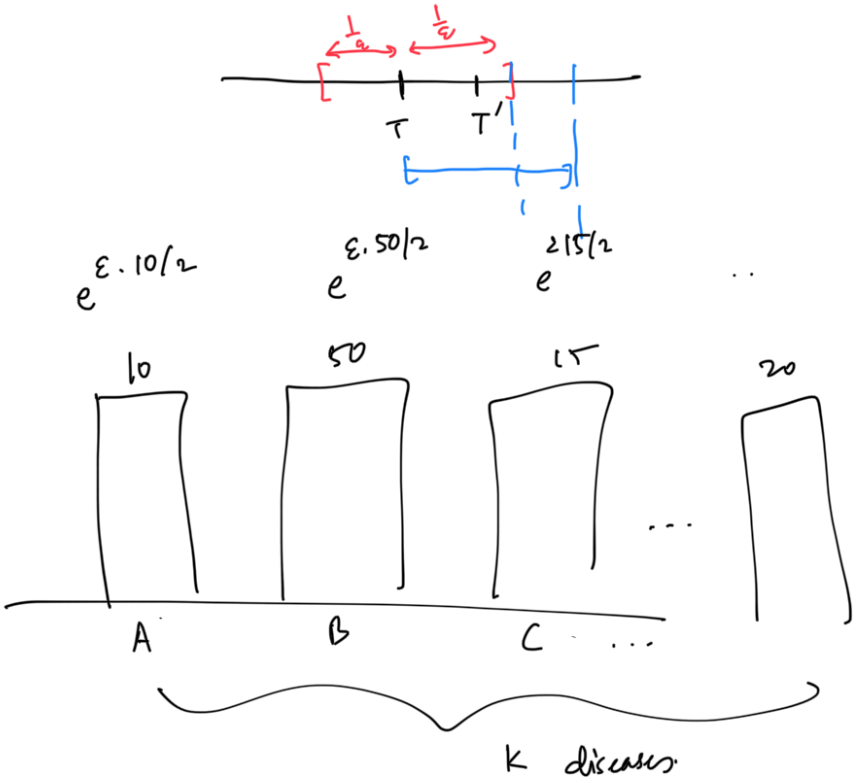
$\leq e^{+1/\lambda}$

$e^{-\frac{|x|}{\lambda} + \frac{|x+\Delta|}{\lambda}}$

In general, Noise $\sim$ Laplace $\left(\frac{\Delta}{\varepsilon}\right)$, where $\Delta = \max_{\mathcal{D}, \mathcal{D}'} |X(\mathcal{D}) - X(\mathcal{D}')|$

$$\Pr\left[\underbrace{X}_{\text{Lap}(\frac{1}{\varepsilon})} > \frac{t}{\varepsilon}\right] \leq e^{-t}$$

Output: True Value + Noise $\left[-\frac{1}{\varepsilon}, \frac{1}{\varepsilon}\right]$



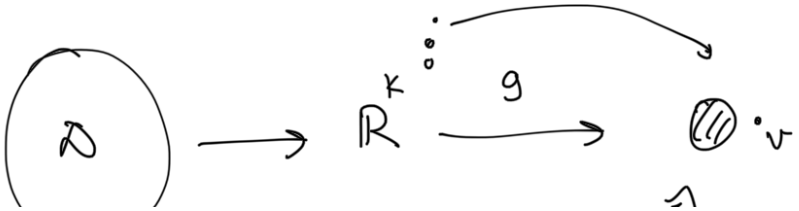
Thm: Suppose $X : \mathcal{D} \to \mathbb{R}^k$ $\mathcal{D}, \mathcal{D}'$: differ in only one "row"

$$\Delta := \sum_{i=1}^{k} \left| X_i(\mathcal{D}) - X_i(\mathcal{D}') \right| = \| X(\mathcal{D}) - X(\mathcal{D}') \|_1$$

$X(\mathcal{D}) + (\underbrace{N_1, \ldots, N_k}_{\text{each of them from } \text{Lap}(\frac{\Delta}{\varepsilon})})$

Disease which is most prevalent ?

Post processing cannot decrease D.P.



$$\Pr[Y(\mathcal{D}) = v] \quad \text{vs} \quad \Pr[Y(\mathcal{D}') = v]$$

$$\Pr[X(\mathcal{D}) \in g^{-1}(v)] \qquad \Pr[X(\mathcal{D}) \in g^{-1}(v)]$$

Auction: n people

$$v_1, \ldots, v_n \in [0,1]$$

$$\text{Revenue} = p \cdot [\#\text{ people whose } v_i \geq p]$$

p

How much noise should we add?

[McSherry & Talwar '07] Exponential Mechanism.

We have a set R of size r.

R ← set of diseases, ...

Each outcome r has a utility $u(r)$

Wanted to find $r \in R$ which has highest $u(r)$ ← r^*

Output $r \in R$ with probability proportional to $e^{\varepsilon u(r)/2}$

$$= \frac{e^{\varepsilon u(r)/2}}{\sum_{r'} e^{\varepsilon u(r')/2}}$$

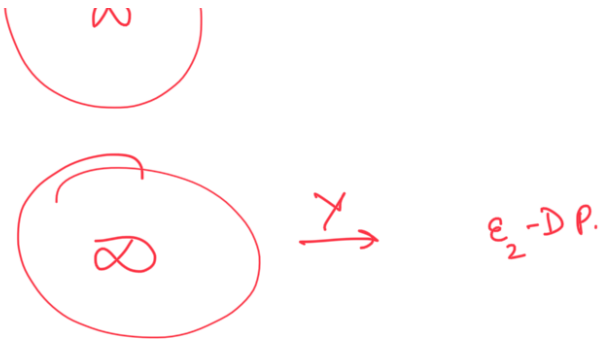
This mech. is ε-diff. private.

Error? It is very likely that $u(r) \geq u(r^*) - \frac{\ln|R|}{\varepsilon}$

Composition Theorem

ε_1-DP

$e^{\varepsilon_1} \cdot e^{\varepsilon_2} = e^{\varepsilon_1 + \varepsilon_2}$


ω
∞
γ
$\mathcal{E}_2$-D P.