365Days

English Stydy Plan

Privacy in ML

Randomized Response

Truth $\quad x_1, \ldots \quad x_n \qquad x_i \in \{0,1\}$

Revealed

$$y_i = \begin{cases} x_i & \text{w.p. } \dfrac{e^{\epsilon}}{1+e^{\epsilon}} \checkmark \Big\} \text{truth} \\ 1-x_i & \text{w.p. } \dfrac{1}{1+e^{\epsilon}} \checkmark \Big\} \text{falsehood} \end{cases}$$

Consider two datasets

Truth $\quad X = \{x_1, \ldots, x_n\} \qquad X' = \{x_1', \ldots, x_n'\}$

Assume $x_i = x_i' \quad \forall i = \{1, \ldots, n-1\}$

Randomized response

Revealed

$$RR(X) = Y = [y_1, \ldots, y_n]$$

$$RR(X') = Y' = [y_1', \ldots, y_n']$$

$$P\Big(\underset{\substack{\swarrow \\ [b_1 \cdots b_n]}}{RR(x) = b}\Big) = P\big([y_1, y_2 \cdots, y_n] = [b_1, b_2, \cdots b_n]\big)$$

$$= \left(\prod_{i=1}^{n-1} P(y_i = b_i)\right) \underbrace{P(y_n = b_n)}_{\to \; 1}$$

$$P\big(RR(x') = b\big) = \prod_{i=1}^{n} P(y_i' = b_i)$$

$$= \underbrace{\left(\prod_{i=1}^{n-1} P(y_i' = b_i\right)}\; P(y_n' = b_n)$$

$$\to \; = \left(\prod_{i=1}^{n-1} P(y_i = b_i)\right) \underbrace{P(y_n' = b_n)}_{\to \; 2}$$

$$\frac{P(y_n = b_n)}{P(y_n' = b_n)} = \begin{cases} \dfrac{e^{\epsilon} / \cancel{1+e^{\epsilon}}}{1 / \cancel{1+e^{\epsilon}}} = e^{\epsilon} & \text{if } \underline{b_n = x_n} \\[2ex] \dfrac{1 / 1+e^{\epsilon}}{e^{\epsilon} / 1+e^{\epsilon}} = e^{-\epsilon} & \text{if } b_n = 1 - x_n \end{cases}$$

$$\frac{P(y_n = b_n)}{P(y_n' = b_n)} \le e^{\epsilon}$$

$$\boxed{P(y_n = b_n) \le e^{\epsilon} P(y_n' = b_n)}$$ — substitute in ①

$$P(RR(x) = b) = \left(\prod_{i=1}^{n-1} P(y_i = b_i)\right) P(y_n = b_n)$$

$$\le \underbrace{\left(\prod_{i=1}^{n-1} P(y_i' = b_i)\right) e^{\epsilon} P(y_n' = b_n)}$$

$$= \underbrace{e^{\epsilon} \prod_{i=1}^{n} P(y_i' = b_i)}_{P(RR(x') = b)}$$

$$\Rightarrow \boxed{\frac{P(RR(x)=b)}{P(RR(x')=b)} \le e^{\epsilon}}$$

How good is this mechanism

Want $\frac{1}{n}\sum_{i=1}^{n} x_i$

$$\underset{\{0,1\}}{X_i} \xrightarrow{\text{Randomization}} Y_i$$

$\mathbb{E}[Y_i]$

$$\mathbb{E}[y_i] = \left(\frac{e^{\epsilon}}{1+e^{\epsilon}}\right)x_i + \left(\frac{1}{1+e^{\epsilon}}\right)^{(1-x_i)}$$

$$\frac{e^{\epsilon}x_i + 1 - x_i}{1+e^{\epsilon}}$$

$$\mathbb{E}[y_i] = x_i\left(\frac{e^{\epsilon}-1}{e^{\epsilon}+1}\right) + \frac{1}{1+e^{\epsilon}}$$

$$\begin{tikzcd} x_i \arrow[r] & y_i \arrow[r] & z_i \arrow[ll, bend right] \end{tikzcd}$$

$$z_i = \left(y_i - \frac{1}{1+e^{\epsilon}}\right)\left(\frac{e^{\epsilon}+1}{e^{\epsilon}-1}\right)$$

$$\mathbb{E}[z_i] \overset{?}{=} x_i$$ (exercise)

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\begin{array}{ccc} x_1, & \cdots & , x_n \\ \downarrow & & \downarrow \\ y_1, & \cdots & , y_n \\ \downarrow & & \downarrow \\ z_1 & & z_n \end{array}$$

Guess for $\bar{x}$ is $\underline{\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i}$

$$\mathbb{E}[\bar{z}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} z_i\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[z_i]$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

$$\left| \underbrace{\frac{1}{n}\sum_{i=1}^{n} x_i} - \underbrace{\frac{1}{n}\sum_{i=1}^{n} z_i} \right|$$

with high probability

$$\lesssim O\left(\frac{1}{\epsilon\sqrt{n}}\right)$$

utility

TRUSTED CURATOR MODEL

Cynthia Dwork [2010]

Differential Privacy

Let $M : \mathcal{X}^n \to \mathcal{Y}$. Consider two "neighbouring" datasets $x, x' \in \mathcal{X}^n$

M is ϵ-D.P if for all x, x' neighbouring and all $S \subseteq Y$,

$$\frac{\Pr(M(x) \in S)}{\Pr(M(x') \in S)} \leq e^{\epsilon}$$

LAPLACE MECHANISM

Sensitivity

$$f : X^n \to \mathbb{R} \quad \text{(average)}$$

$$\Delta = \max_{\substack{\text{neighbouring} \\ \text{datasets} \\ x, x'}} \left| f(x) - f(x') \right|$$

$$\left| \frac{1}{n}\sum_{i=1}^{n} x_i - \underbrace{\frac{1}{n}\sum_{i=1}^{n} x_i'}_{} \right|$$

$$\left| \frac{1}{n}\sum_{i=1}^{n-1} x_i + \frac{1}{n} x_n - \frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{n} x_n' \right|$$

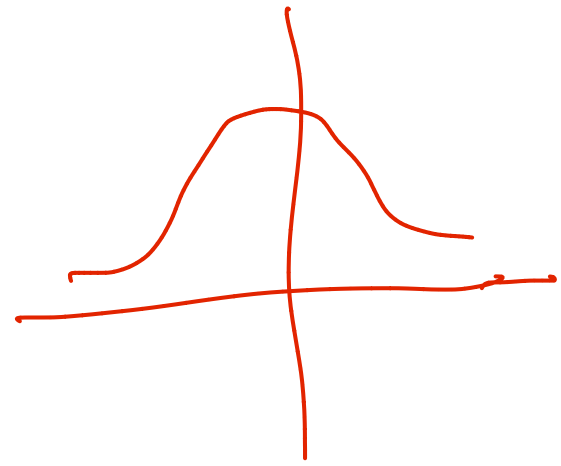
$$= \frac{1}{n} \left| x_n - x_n' \right|$$

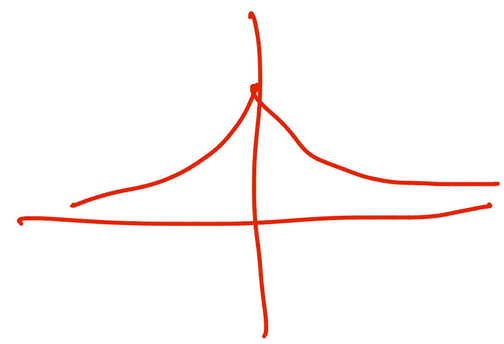
$$\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array}$$

$$\boxed{\Delta = \frac{1}{n}}$$

Laplace distribution

$$f_{Lap}(x) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}$$



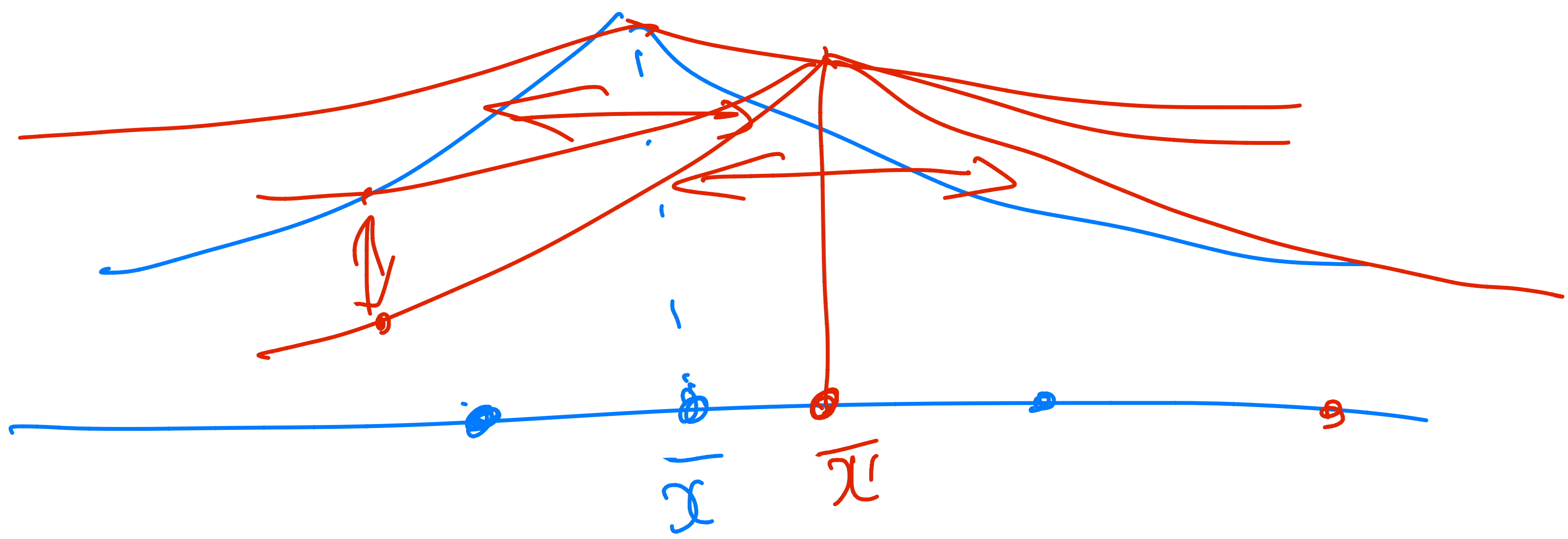


$$\hat{x} = \underbrace{\frac{1}{n}\sum_{i=1}^{n} x_i}\ + \eta \rightarrow \text{Laplace}\left(0, \frac{\Delta}{\epsilon}\right)$$

$$\mathbb{E}[\hat{x}] = \bar{x}$$

Can argue Laplace mechanism is

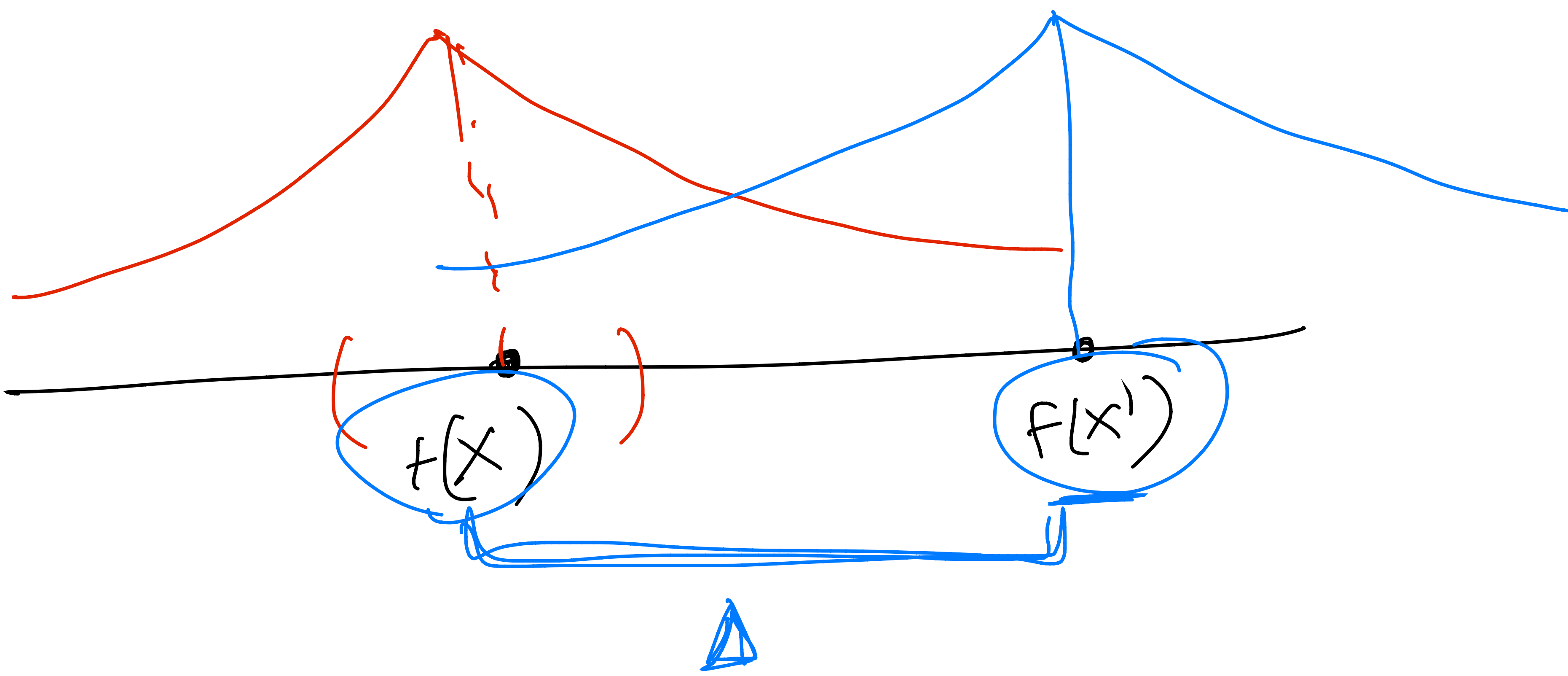
ϵ-DP ✓ [Exercise].

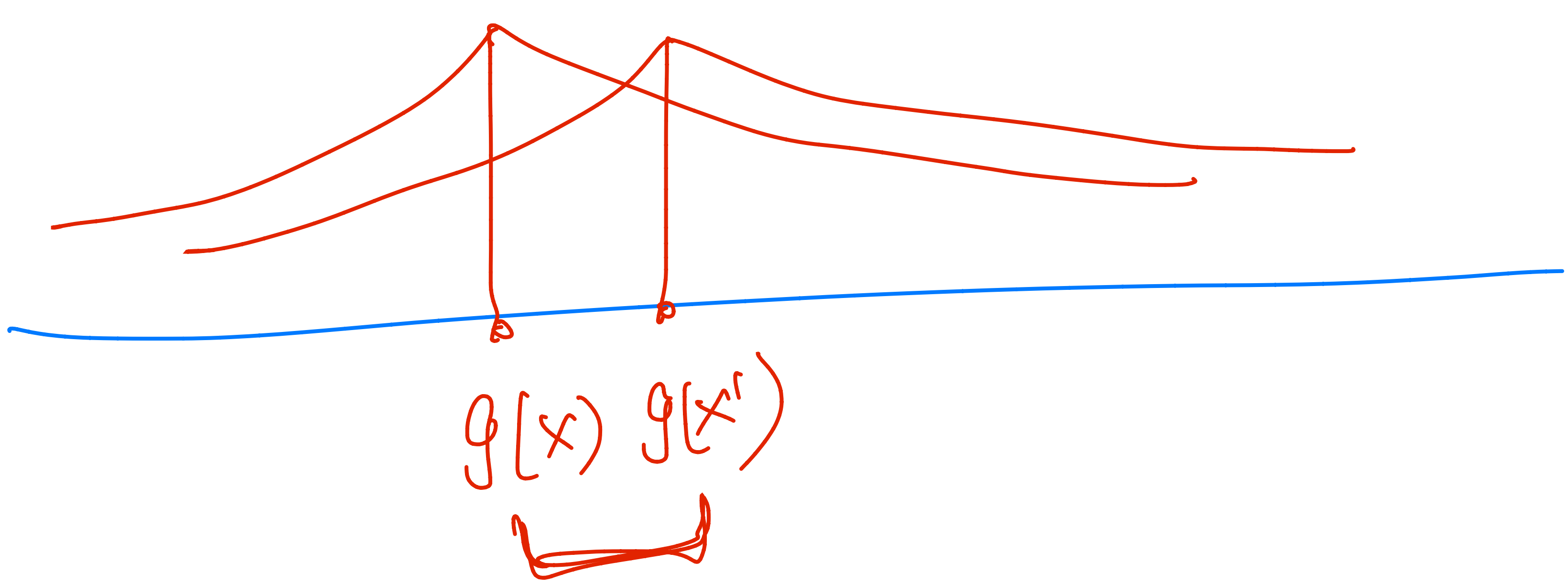


<u>utility</u>

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} x_i - \left(\frac{1}{n}\sum_{i=1}^{n} x_i + \eta\right)\right| \geq \epsilon\right)$$

LAP	RR
$O\left(\frac{1}{\epsilon n}\right)$	$O\left(\frac{1}{\epsilon\sqrt{n}}\right)$


f(x)
f(x')
Δ


g(x) g(x')

Approximate DP

$M: \mathcal{X}^n \to \mathcal{Y}$ is (ϵ, δ) D.P

if $\forall$ neighbouring $x, x' \in \mathcal{X}^n$

and all $S \subseteq \mathcal{Y}$

$$P[M(x) \subseteq S] \leq e^{\epsilon} P[M(x') \subseteq S] + \delta$$

Add Gaussian noise

L-2 sensitivity

$$N\left(0, \ln\left(\frac{1}{\delta}\right) \frac{\Delta_2^2}{\epsilon^2}\right)$$

privacy

Laplacian noise $\quad Lap\left(0, \frac{\Delta}{\epsilon}\ \boxed{\frac{d}{n\epsilon}}\right)$

Gaussian noise $\quad \mathcal{N}\left(0, \frac{\sqrt{d \log(1/\delta)}}{n\epsilon}\right)$

---

why does it work?

L1-Δ $\quad \max_{x,x'} \| f(x) - f(x') \|_1$

L2-Δ $\quad \max_{x,x'} \| f(x) - f(x') \|_2$

$\forall x \quad \|x\|_2 \leq \|x_1\|_2 \leq \sqrt{d}\, \|x_2\|$

Properties of ADP

- ## Post-processing

- M is (ϵ, δ) D.P

  $F \circ M$ is also (ϵ, δ) D.P

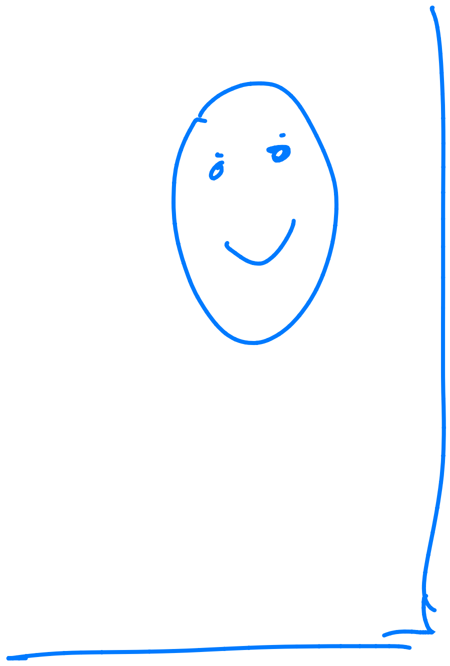
Composition

- $\{M_1, M_2, \ldots, M_k\}$ are all (ϵ, δ) D.P.

  Basic : $(k\epsilon, k\delta)$

- Advanced :

  $$\left(\epsilon\sqrt{k\log(1/\delta')} + k\epsilon(e^{\epsilon}-1),\ k\delta + \delta'\right) - \text{D.P}$$

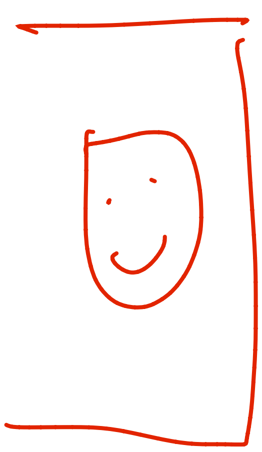


Buyer $= n$

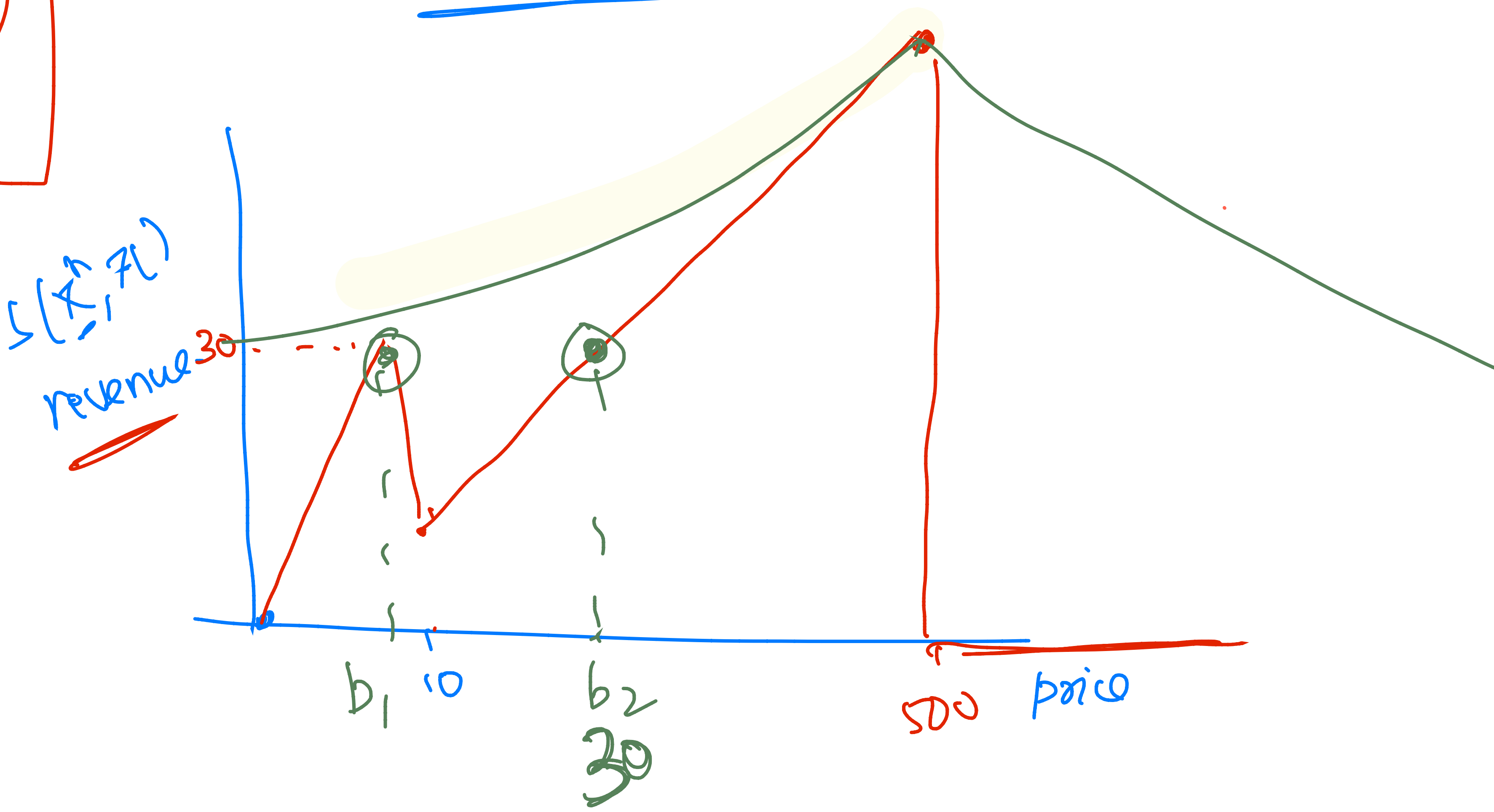
Valuation $v_1, \ldots, v_n.$

How to set price

$$p \longrightarrow p \, \left| \{ i : v_i \geq p \} \right|$$



10, 10, 500



$x \in \mathcal{X}^n$ (valuations) (10, 10, 500)

$\mathcal{H}$ (Prices)

$S : (\mathcal{X}^n \times \mathcal{H}) \rightarrow \mathbb{R}$

$$\Delta = \max_{h \in \mathcal{H}} \max_{x, x'} \left| S(x, h) - S(x', h) \right|$$

EXPONENTIAL MECHANISM (2018 Talwar et.al).

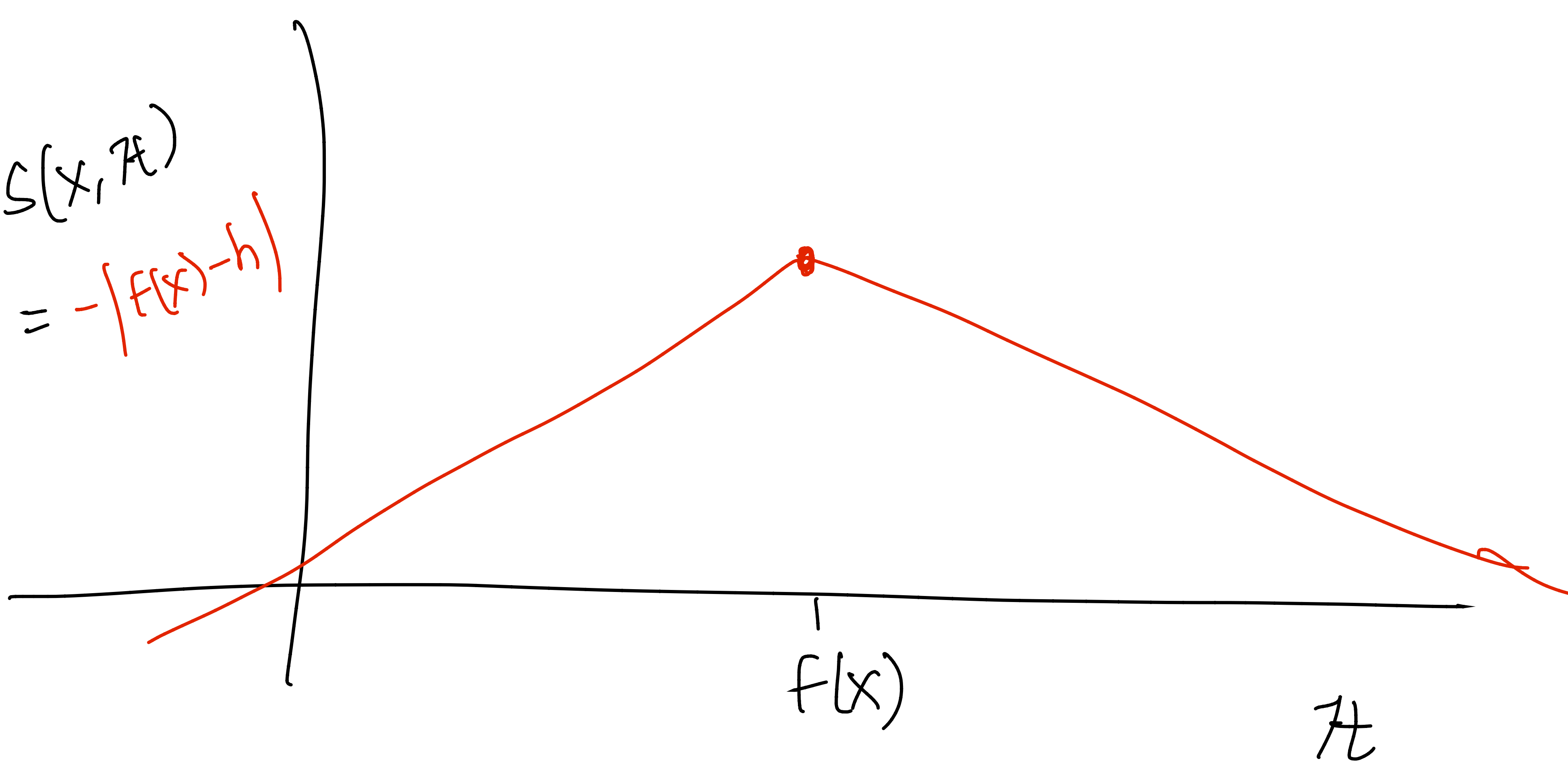
Select $h \in \mathcal{H}$ with probability proportional to $\boxed{e^{\frac{\epsilon}{2\Delta} S(x,h)}}$

EM is ϵ-DP

utility

$$P\Big(\underbrace{S(\underbrace{EM(\overset{\text{valuations}}{x})}_{\text{price}})}_{\text{revenue}} \le \underbrace{OPT(x)}_{\substack{\text{max money}\\ \text{I can make}}} - \Big(\frac{2\Delta \ln(|\mathcal{H}|)}{\epsilon} + t\Big)\Big) \le e^{-t}$$

Laplace Mechanism



Privacy in ML

$$L(w) = \frac{1}{n} \sum_{i=1}^{n} \underbrace{\ell(x_i, y_i; w)}_{\text{Loss function}} + \frac{R(w)}{n}$$

$$w^* = \arg\min_{w} L(w)$$

- **Output perturbation**

(ϵ, δ) $\quad O\left(\frac{d}{\epsilon\sqrt{n}}\right)$ Excess error

$$\hat{w} = w^* + \eta \leftarrow \text{noise.}$$

$\eta \in \mathbb{R}^d.$

- **Objective perturbation**

$$\hat{w} = \arg\min_{w} \left(L(w) + \underline{w^T \eta} \right)$$

$\rightarrow$ noise.

(ϵ, δ)-D.P $\qquad \tilde{O}\left(\frac{d}{\epsilon\sqrt{n}}\right)$

Gradient perturbation

Run SGD with "noisy gradients".

$(\epsilon, \delta) \qquad O\left(\frac{\sqrt{d}}{\epsilon n}\right)$
