Generalizations (Chapter 5):

- What about nonlinear queries?
- What about different bounds for linear queries?

First generalize SmallDB to work on nonlinear queries.

Idea: Nets.

- Key from SmallDB: there is some small database that's accurate on queries in $\mathcal{Q}$.
- Let Exponential Mechanism handle the rest!
  - Need not many small databases.
- Can generalize!

Def: A class of data structures for class of linear queries $\mathcal{Q}$. Each $f \in \mathcal{Q}$ is of the form $f: \mathcal{D} \to \mathbb{R}$, but also can extend to $\mathcal{S}$: $f: \mathcal{D} \cup \mathcal{S} \to R$.

(could be synthetic databases, but not necessarily.

Def: An α-net for a class of queries Q is a set $N \subseteq \mathcal{S}$ s.t. for all $D \in \mathcal{D}$, $\exists S \in N$ s.t.

$$\max_{f \in Q} |f(D) - f(S)| \leq \alpha$$

Let $N_\alpha(Q)$ be an α-net of min size among all α-nets for Q.

SmallDB but for nets: net mechanism $(D, Q, \varepsilon, \alpha)$:

- Let $R = N_\alpha(Q)$
- Let $q: \mathcal{D} \times R \to \mathbb{R}$ be

$$q(D, S) = -\max_{f \in Q} |f(D) - f(S)|$$

- Use exponential mechanism $M_E(D, q, R)$

Thm: Net mechanism is ε-DP

Pf: exponential mechanism

Thm: Let Q be any class of sensitivity $\frac{1}{|D|}$ queries (like linear queries). Let $S = \text{NetMechanism}(D, Q, \varepsilon, \alpha)$.

Then w/prob $\geq 1 - \beta$

$$\max_{f \in \mathcal{Q}} |f(D) - f(S)| \leq \alpha + \frac{2(\log(|N_\alpha(\mathcal{Q})|) + \log\frac{1}{\beta})}{\varepsilon |D|}$$

Pf: utility of exp. mechanism, fact that $OPT_f(D) \leq \alpha$ by def of α-net.

So it suffices to answer $\mathcal{Q}$ construct small α-net for $\mathcal{Q}$!

Second: generalize private MWU

For now: stick with offline.

Offline Private MWU:

- Start with $\hat{D}_x = \frac{1}{|X|}$ $\forall x \in X$.
- while $\exists f \in \mathcal{Q}$ s.t. $|f(D) - f(\hat{D})| > \alpha$
  - find $f \in \mathcal{Q}$ s.t. $|f(D) - f(\hat{D})| > \alpha$ (distinguishing)
  - update $\hat{D}$ using MW
- Use final $\hat{D}$ to answer all queries (or just release it).

To generalize, need two things:

- Distinguisher
- Database update algorithm.

Can use EM as canonical distinguisher, MW for database update (linear queries).

Database update rule:

$$U : \mathcal{S} \times Q \times \mathbb{R} \to \mathcal{S}$$

- $\mathcal{S}$: current data structure
- Q: distinguishing query
- $\mathbb{R}$: estimate of $f(D)$

Database update sequence: Let $D \in \mathcal{D}$.

$\{D^t \in \mathcal{D}, (D^t, f^t, v^t)\}_{t \in [L]} \in (\mathcal{S} \times Q \times \mathbb{R})^L$ is a

(U, D, Q, α, L)-database update sequence if

1) $|f^t(D) - f^t(D^{t-1})| \geq \alpha \quad \forall t \in [L]$
2) $|f^t(D) - v^t| < \alpha \quad \forall t \in [L]$
3) $D^t = U(D^{t-1}, f^t, v^t) \quad \forall t \in [L]$

Def: U is a $T(\alpha)$-database update algorithm for Q if $\forall D \in \mathcal{D}$, every (U, D, Q, α, L)-database update sequence has $L \leq T(\alpha)$

So after $T(\alpha)$ updates, can't update anymore $\Rightarrow D^t$ memorizes for all $f \in Q$!

Ex: MW is $T(\alpha) = \frac{4 \log |X|}{\alpha^2}$-database update algorithm.

Note: Connection to online learning in mistake bound model.

MW ≃ ltt dsl

median mechanism ≃ halving alg.

Perceptron ≃ perceptron

Need to pair with a (private) distinguisher

Def: $(F(\varepsilon), \gamma)$-private distinguisher: ε-DP alg. that, given $D \in \mathcal{D}$, $S \in \mathcal{S}$, outputs $f^* \in Q$ s.t.

$$|f^*(D) - f^*(S)| \geq \max_{f \in Q} |f(D) - f(S)| - F(\varepsilon) \text{ w.p. } \geq 1 - \gamma$$

- Intuition: easier to solve distinguishing than full answering: don't need to find answer to all queries, just one w/ most error.
- Reduce private release to private distinguishing
- Combine w/ database update alg. to get iterative construction mechanism.

(Assume ℓ_1 sensitivity $\frac{1}{|D|}$ - queries)

$I(D, \alpha, \varepsilon_0)$

- Init $D^0 \in \mathcal{S}$
- For $t = 1$ to $T(\frac{\alpha}{2})$:
  - Let $f^t = \text{Distinguish}(D, D^{t-1})$
  - Let $\hat{v}^t = f^t(D) + \text{Lap}\left(\frac{1}{|D|\varepsilon_0}\right)$
  - if $|\hat{v}^t - f^t(D^{t-1})| < \frac{3\alpha}{4}$,
    output D^{t-1}
  - else $D^t = \mathcal{U}(D^{t-1}, f^t, \hat{v}^t)$
- Output $D^{T(\alpha/2)}$

Thm: ε-DP if $\varepsilon_0 \leq \frac{\varepsilon}{2T(\alpha/2)}$

(ε, δ)-DP if $\varepsilon_0 \leq \frac{\varepsilon}{4\sqrt{T(\frac{\alpha}{2})\ln\frac{1}{\delta}}}$

Pf: Basic/advanced composition: $2T(\frac{\alpha}{2})$ ε_0-DP steps $(\text{Distinguish}, \hat{v}^t)$.

Thm: Given $(F(\varepsilon), \gamma)$-private distinguisher, $T(\alpha)$-database update algorithm. Then w.p. $\geq (1-\beta)$, $I(\text{returns } S \in \mathcal{S} \text{ s.t.}$

$\max_{f \in Q} |f(D) - f(S)| \leq \alpha$ for any

$$\alpha \geq \max\left(\frac{8 \log\left(\frac{2T(\alpha/2)}{\beta}\right)}{\varepsilon |D|} , 8 F(\varepsilon_0) \right)$$

as long as $\gamma \leq \frac{\beta}{2T(\alpha/2)}$

Pf: w.p. $\geq 1 - \beta/2$, $|\hat{v}^t - f^t(D)| \leq \frac{\alpha}{8}$ $\forall t$ (Laplace)

w.p. $\geq 1 - \beta/2$, $|f^t(D) - f^t(D^{t-1})| \geq \max_{f \in Q} |f(D) - f^t(D^{t-1})| - \frac{\alpha}{8}$ (private distinguisher) $\forall t$

If return $D^{T(\alpha/2)}$, then each

$|\hat{v}^t - f^t(D^{t-1})| \geq \frac{3\alpha}{4}$ and

$|\hat{v}^t - f^t(D)| \leq \frac{\alpha}{8}$

$\Rightarrow |f^t(D) - f^t(D^{t-1})| \geq \frac{3\alpha}{4} - \frac{\alpha}{8} = \frac{5\alpha}{8} \geq \frac{\alpha}{2}$

$\Rightarrow$ each $T(\frac{\alpha}{2})$ distinguishing queries w/ param. $\frac{\alpha}{2}$

$\Rightarrow$ $\frac{\alpha}{2}$-accurate $\forall f \in Q$.

Else return D^{t-1} for some $t < T(\alpha/2)$

$\Rightarrow |\hat{v}^t - f^t(D^{t-1})| < \frac{3\alpha}{4}$

$\Rightarrow |f^t(D) - f^t(D^{t-1})| \leq |\hat{v}^t - f^t(D^{t-1})| + |f^t(D) - \hat{v}^t|$

$\leq \frac{3\alpha}{4} + \frac{\alpha}{8} = \frac{7\alpha}{8}$

$\Rightarrow$ by property of private distinguisher,

$$\max_{f \in Q} |f(x) - f(D^{+i})| \le |f^+(D) - f^+(D^{+i})| + F(\varepsilon_0)$$

$$\le \frac{2\alpha}{3} + F(\varepsilon_0) \le \alpha \quad \checkmark$$

use EM as private distinguisher: domain Q,

quality score $q(D, f) = |f(D) - f(D^+)|$

Thm: EM is an $(F(\varepsilon), \gamma)$-private distinguisher for

$$F(\varepsilon) = \frac{2}{|D|\varepsilon} \log \frac{|Q|}{\gamma}$$

Plug this thm into TC utility thm in/correct ε_0:

Thm: Given $T(\alpha)$-database update algorithm, TC w/ EM distinguisher is ε-DP, returns S s.t.

$$\max_{f \in Q} |f(D) - f(S)| \le \alpha \quad \text{for} \quad \text{w.p. } 2/3.$$

$$\alpha = \frac{8 T(\frac{\alpha}{2})}{|D|\varepsilon} \left(\log \frac{|Q|}{\gamma}\right)$$

is (ε, δ)-DP, returns S s.t. $\max_{f \in Q} |f(D) - f(S)| \le \alpha$ for

$$\alpha = \frac{16\sqrt{T(\alpha/2)\log(1/\delta)}}{|D|\varepsilon} \log \frac{|Q|}{\gamma}$$

s. long as $\gamma \le \beta / 2T(\alpha/2)$