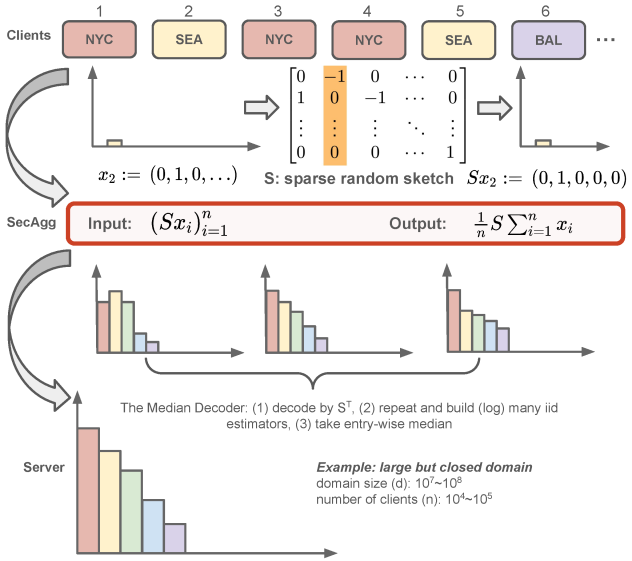
Private Federated Frequency Estimation: Adapting to the Hardness of the Instance

Jingfeng Wu[1], Wennan Zhu[2], Peter Kairouz[2], Vladimir Braverman[3]

[1]JHU => UC Berkeley (work done at Google), [2]Google Research, [3]Rice University and Google

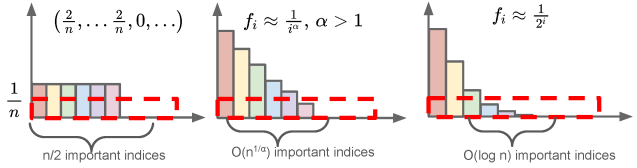
NEURAL INFORMATION PROCESSING SYSTEMS

Q. Secure frequency estimation with small communication?



Q1. How large should S be?

A1. Easier problems require smaller S

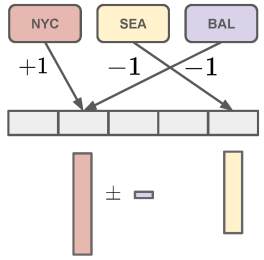


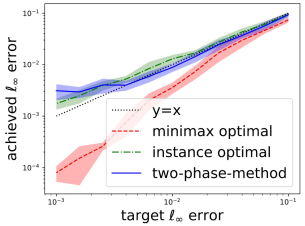
Q2. What S to use?

A2. Count-sketch

 # buckets is

$$\tilde{O}\Big(\underbrace{\#\{f_i \geq \tau\}}_{\text{\# important indices at error level } \tau} + \underbrace{\tfrac{1}{\tau^2}\textstyle\sum_{f_i<\tau} f_i^2}_{\text{additional buckets for dealing with collisions}}\Big)$$





C4 (Colossal Clean Crawled Corpus) dataset experiments

d = 150,868
n = 15,000 ≈ d/10
sketch length = 5

constant factor in $\tilde{O}(\cdot)$ is set to be 2

minimax optimal: sketch width w.r.t. the minimax-optimal bound
instance optimal: sketch width w.r.t. the instance-specific bound

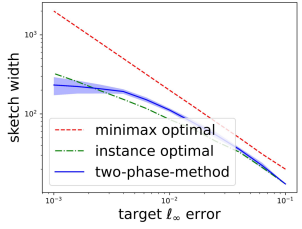
Q3. How to select S in practice?

A3. A two-phase method:

1. approximate # buckets
2. use sketch with approximated # buckets

C4 dataset experiments

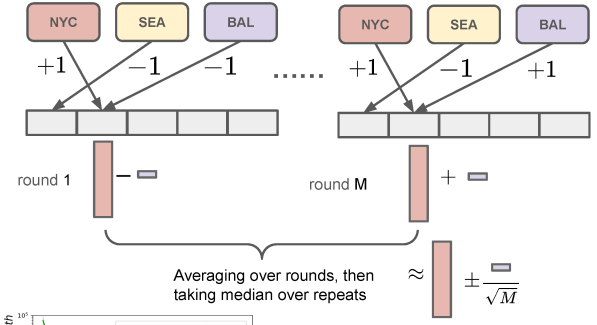
In phase 1, we model the freq. with a poly. We use n/10 data and a sketch of size 8x200 to estimate the top-20 index of the poly.

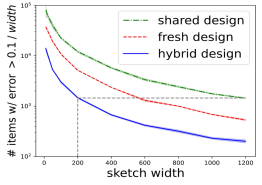


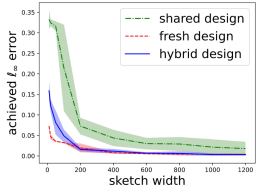
Q4. FA with multi-rounds? That is, server sees M partial sums instead of only 1.

A4. Count-sketch with **hybrid** design:

1. **shared** bucket hashes + **fresh** sign hashes
2. best of shared design and fresh design







	shared design	fresh design	hybrid design
bucket hashes across rounds	*shared*	*fresh*	*shared*
sign hashes across rounds	*shared*	*fresh*	*fresh*
variance?	*large*	*small*	***small***
confidence interval?	*large*	*small*	***large***

C4 dataset experiments.
d=150,868, total #clinets N = 150,000, #rounds M = 10, random split, #clinets per round n = N/M, sketch length L = 5.

Q5. (Central) differential privacy?

A5. Add proper noise to sketched matrices.