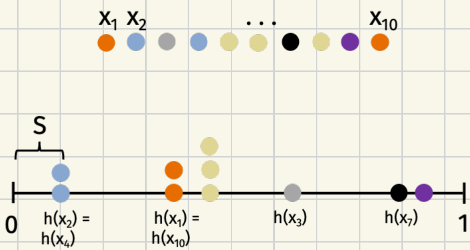
Tuesday, February 10

- Thank you for coming to GEMS/ my talk!
- CCMS Applied Math seminar Mondays : 4:15 pm Estella 1021
- CCMS Colloquium Fridays : 11am Davidson Lecture Hall
- No OH Thursday, DM to meet Friday!

Plan

- Clean up discrete elements
- Load balancing

Distinct Elements



$\bar{S} \leftarrow \sum_{j=1}^{k} \frac{S_j}{k}$

$D \leftarrow \frac{1}{\bar{S}} - 1$

wp $1-\delta$ in $O\left(\frac{\log D}{\epsilon^2 \delta}\right)$ space:

$$(1-\epsilon)\mu \le \bar{S} \le (1+\epsilon)\mu$$

$$\iff \frac{1}{(1+\epsilon)\mu} \le \frac{1}{\bar{S}} \le \frac{1}{(1-\epsilon)\mu}$$

$1-2\epsilon \le \frac{1}{1+\epsilon}$; $\frac{1}{1-\epsilon} \le 1+2\epsilon$ by Desmos

$$\iff (1-2\epsilon)\frac{1}{\mu} - 1 \le \frac{1}{\bar{S}} - 1 \le (1+2\epsilon)\frac{1}{\mu} - 1$$

$\mu \le \frac{1}{2}$

$$\iff (1-4\epsilon)\left(\frac{1}{\mu} - 1\right) \le \hat{D} \le (1+4\epsilon)\left(\frac{1}{\mu} - 1\right)$$

$$\iff (1-4\epsilon) D \le \hat{D} \le (1+4\epsilon) D$$

$$(1-\epsilon) D \le \hat{D} \le (1+\epsilon) D$$

where we hide 4 in big-O

Load Balancing

Goal: Distribute load evenly between servers using hash function

e.g., Google maps routing

Advantage of hash: cache redundant queries

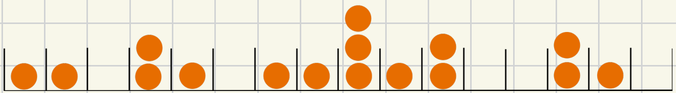
n servers and m (unique) queries

$$S_i = \#\text{ queries sent to server } i$$
$$= \sum_{j=1}^{m} \mathbb{1}[h(x_j) = i]$$

Bound $S = \max_{i \in [n]} S_i$

- $\mathbb{E}[S_i] =$
- $\mathbb{E}[S] \overset{?}{=} \max_{i \in [n]} \mathbb{E}[S_i]$

Let's assume $m = n$, so $\mathbb{E}[S_i] = 1$



Goal: Prove

$$\Pr\left(\max_i S_i \geq C\right) \leq 1/10$$

$$\Leftrightarrow \quad \Pr\left(S_1 \geq C \cup S_2 \geq C \cup \ldots \cup S_n \geq C\right) \leq 1/10$$

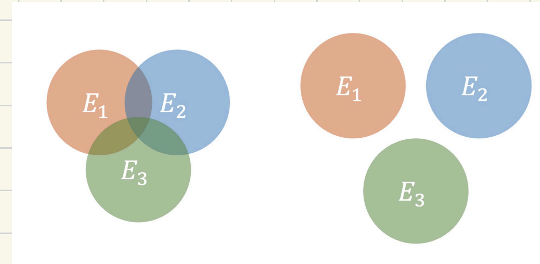
By union bound,

$$\Pr\left(S_1 \geq C \cup \ldots \cup S_n \geq C\right)$$

$$\leq \sum_{i=1}^{n} \Pr(S_i \geq C) = n \Pr(S_i \geq C) \overset{\text{want}}{\leq} n \cdot \frac{1}{10n}$$

Union Bound

Proof by picture:



Proof by Markov's:

$$\Pr(E_1 \cup \ldots \cup E_n) = \Pr\left(\sum_{i=1}^{n} \mathbb{1}[E_i] \geq 1\right)$$

$$\leq \mathbb{E} \sum_{i=1}^{n} \mathbb{1}[E_i] = \sum_{i=1}^{n} \Pr(E_i)$$

Prove that

$$\Pr(s_i \geq c) \leq \frac{1}{10n}$$

- Markov's ?
- Chebyshev's ?
- What c?