Plan

Logistics

Review

Load Balancing

Games!!

Form

Problem Set

Reading notes!

Chebyshev's

$\sigma^2 = Var(X)$

$$Pr(|X - \mathbb{E}[X]| \geq k \cdot \sigma) \leq \frac{1}{k^2}$$

choose k based on problem

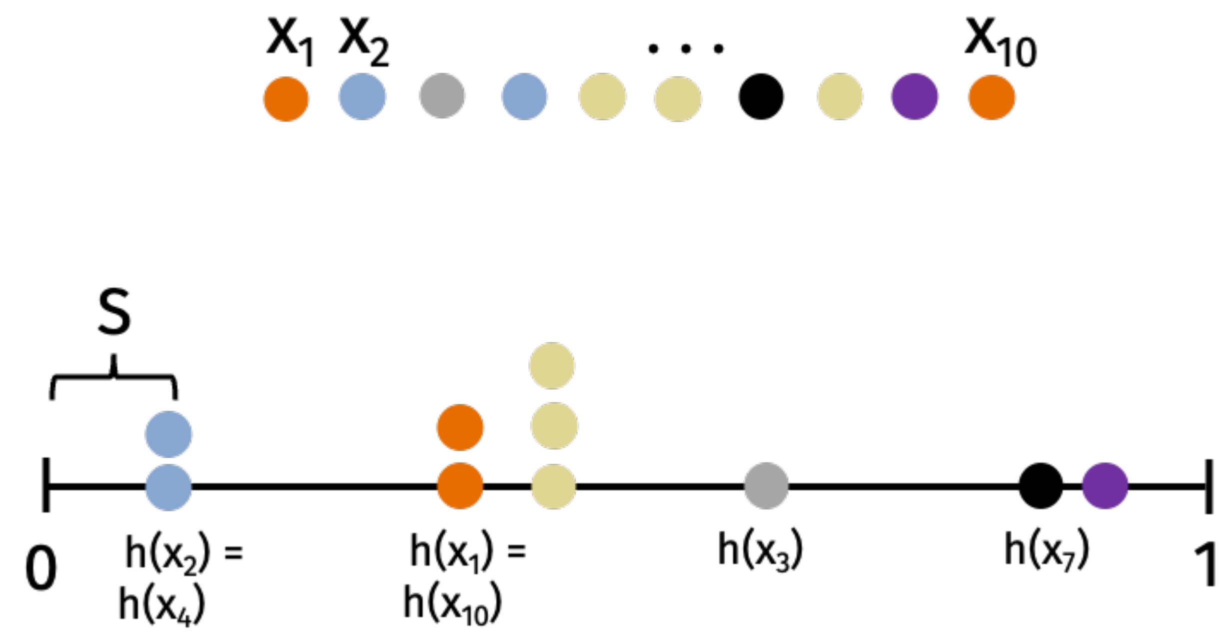
↳ covid, $k=1$ in Markov's

↳ Today, k chosen so $\frac{1}{k^2}$ small

Linearity of Variance

$$var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)$$

Distinct Elements



Tools

$$\mathbb{E}[X] = \int_{x=0}^{\infty} Pr(X \geq x)\, dx$$

Apply Chebyshev's?

$$\mathbb{E}[S] = \frac{1}{D+1}$$

$$\mathbb{E}[S^2] = \frac{2}{(D+1)(D+2)}$$

$$\mathrm{Var}(S) = \mathbb{E}[S^2] - \mathbb{E}[S]^2$$

$$= \frac{2}{(D+1)(D+2)} - \frac{1}{(D+1)^2}$$

$$\overset{*}{\leq} \frac{2}{(D+1)(D+1)} - \frac{1}{(D+1)^2} = \frac{1}{(D+1)^2}$$

$$\boxed{\begin{array}{c} D+1 \leq D+2 \; * \\ \frac{1}{D+2} \leq \frac{1}{D+1} \end{array}}$$

Variance Reduction

Repeat:

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

$$\mathbb{E}\left[\frac{1}{k}\sum_{j=1}^{k} S_j\right] = \frac{1}{k}\sum_{j=1}^{k} \mathbb{E}[S_j] = \mathbb{E}[S_j]$$

$$= \frac{1}{D+1}$$

$$\mathrm{Var}\left(\frac{1}{k}\sum_{j=1}^{k} S_j\right) = \frac{1}{k^2}\sum_{j=1}^{k} \mathrm{Var}(S_j)$$

$$= \frac{1}{k}\mathrm{Var}(S_j)$$

$$\leq \frac{1}{k}\left(\frac{1}{D+1}\right)^2$$

$$\sigma^2 = Var(S) \le \frac{1}{(D+1)^2} \cdot \frac{1}{k} = \frac{\mu^2}{k}$$

$$\sigma = \frac{\mu}{\sqrt{k}}$$

$$Pr(|S-\mu| \ge l \cdot \sigma) \le \frac{1}{l^2}$$

$$Pr\left(|S-\mu| \ge l \cdot \frac{\mu}{\sqrt{k}}\right) \le \frac{1}{l^2}$$

$$l = \frac{1}{\sqrt{\delta}} \qquad k = \frac{1}{\epsilon^2 \cdot \delta}$$

$$Pr\left(|S-\mu| \ge \frac{1}{\sqrt{\delta}} \cdot \frac{1 \cdot \mu}{\sqrt{1/\epsilon^2\delta}}\right) = *$$

$$* = Pr(|S-\mu| \ge \mu\epsilon) \le \frac{1}{(1/\sqrt{\delta})^2}$$

$$Pr(S \ge \mu - \mu\epsilon \text{ or}$$

$$S \le \mu + \mu\epsilon) \le \delta$$

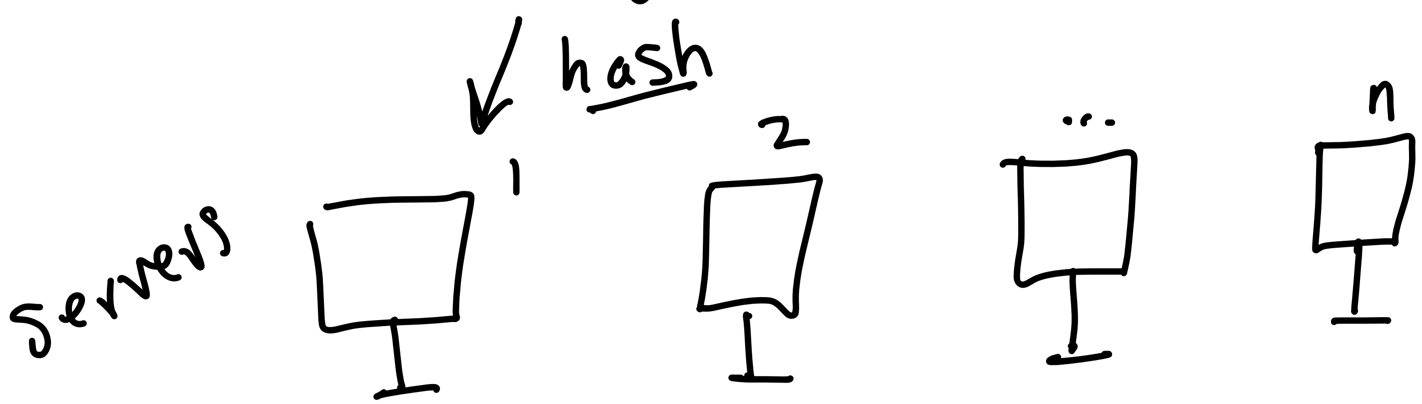
space $O(k) \cdot$ space for hash

$$= O\left(\frac{\log D}{\epsilon^2 \delta}\right)$$

Load Balancing

$x_1, \ldots, x_m$ requests

"Middlebury to NYC"



want:

↳ no server over loaded

↳ same request to same server to avoid recomputing

server load

$$S_i = \sum_{j=1}^{m} \overbrace{\mathbb{1}[h(x_j) = i]}^{S_{i,j}}$$

want to understand maximum load

$$S = \max_i S_i$$

$$\mathbb{E}[S_i] = \mathbb{E}\left[\sum_{j=1}^{m} S_{i,j}\right] = m \cdot \frac{1}{n}$$

$$S = \max_{i \in \{1, \dots, n\}} S_i$$

question

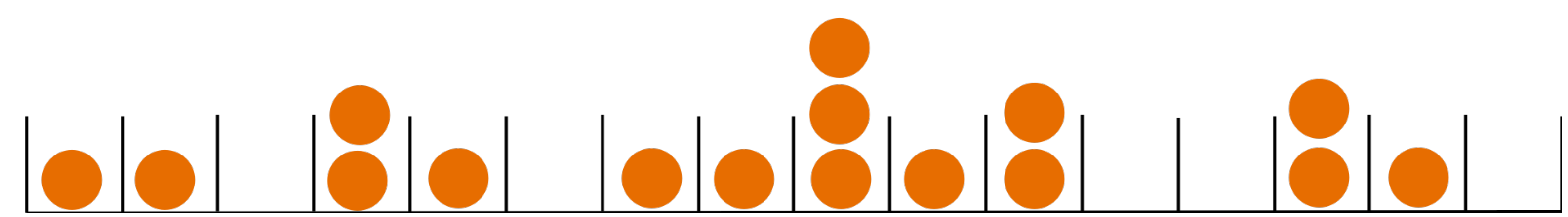
$$\mathbb{E}[S]^2 = \max_i \mathbb{E}[S_i]$$

[redacted]

Simplify with $m = n$

Use hash function so

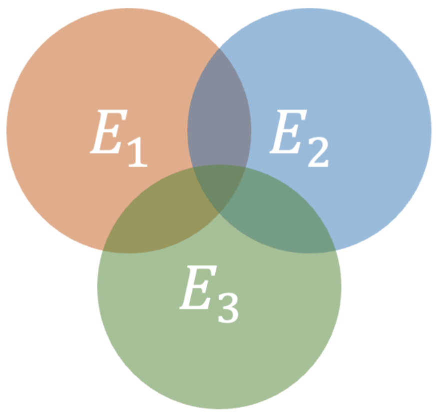
$$\Pr(h(x) = h(y)) = \frac{1}{n}$$

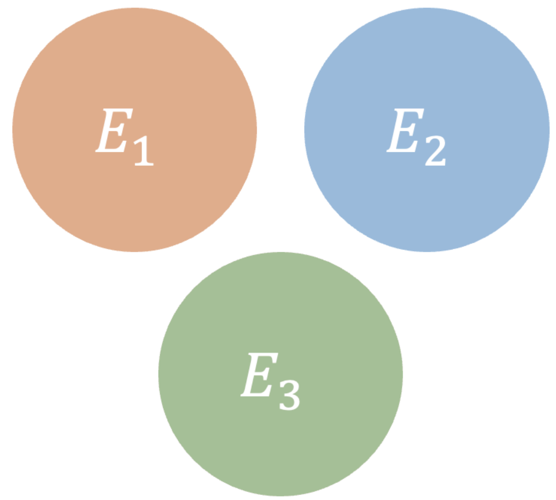


m = # balls

n = # bins

Union Bound





[redacted]

$$\mathbb{1}[E_i]$$

$$X = \sum_{i=1}^{n} \mathbb{1}[E_i]$$

$=$ [redacted]

Markov's

$$\Pr(X \geq t) \leq \frac{\mathbb{E}[X]}{t}$$

[redacted]

$$\Pr(\max_i s_i \geq c) \overset{\text{want}}{\leq} 1/10$$

some value

$\Leftrightarrow$

$$\Pr((s_1 \geq c) \cup (s_2 \geq c) \cup \ldots \cup (s_n \geq c)) \overset{\text{want}}{\leq} 1/10$$

One way to show this...

$$\Pr((s_1 \geq c) \cup (s_2 \geq c) \cup \ldots \cup (s_n \geq c))$$

$$\overset{\text{union}}{\leq} \Pr(s_1 \geq c) + \ldots + \Pr(s_n \geq c)$$

$$\leq n \Pr(s_i \geq c) \overset{\text{want}}{\leq} 1/10$$

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

hard to prove!

Chebyshev's!

$$\mathbb{E}[S_i] = \frac{m}{n} = 1$$

$$Var(S_i) = Var \sum_{j=1}^{m} S_{i,j}$$

$$= \sum_{j=1}^{m} Var(S_{i,j})$$

$$\mathbb{E}[S_{i,j}] = Pr(\text{ j goes to i})$$

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

$$\mathbb{E}[S_{i,j}^2] = \frac{1}{n}$$

$$Var(S_{i,j}) = \mathbb{E}[S_{i,j}^2] - \mathbb{E}[S_{i,j}]^2$$

$$= \frac{1}{n} - \frac{1}{n^2} \leq \frac{1}{n}$$

$$Var(S_i) \leq m \cdot \frac{1}{n} = \frac{m}{n} = 1$$

$$Pr(|S_i - 1| \geq k \cdot 1) \leq \frac{1}{k^2} \overset{\text{want}}{=} \frac{1}{10n}$$

$$k = \sqrt{10n}$$

$$Pr(|S_i - 1| \geq \sqrt{10n}) \leq \frac{1}{10n}$$