Plan

Logistics

Review

Load Balancing

Games night!

6pm in Warner

Form

Problem Set

- Read typed notes
- Pre-lecture notes

Plan on me not being around Friday

Tools:

Markov's

Chebyshev's

Union Bound

Linearity of Expectation

Linearity of Variance

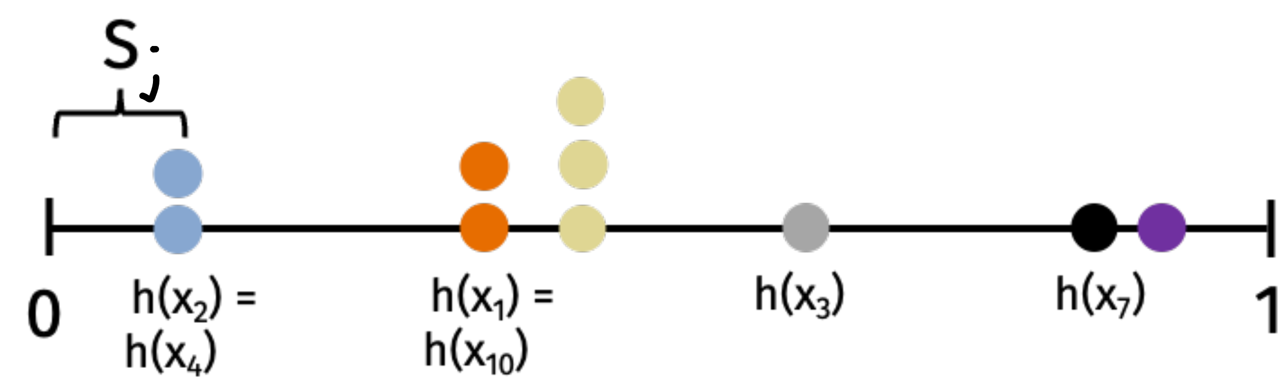
$\boxed{\text{Chebyshev's}}$

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

$\boxed{\text{Linearity of Variance}}$

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \underset{\text{indep}}{=} \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

Distinct Elements



natural numbers

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

continuous positive

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

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

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

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

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

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

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

$$D+1 \leq D+2$$

$$\frac{1}{D+2} \leq \frac{1}{D+1}$$

Variance Reduction

$$S = \frac{1}{\ell} \sum_{j=1}^{\ell} S_j$$

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

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

$$\leq \frac{1}{\ell^2} \sum_{j=1}^{\ell} \mu^2$$

$$= \frac{\mu^2}{\ell}$$

$$Var(cX)$$

$$= \mathbb{E}[(cX - c\mathbb{E}[X])^2]$$

$$= \mathbb{E}[c^2 (X - \mathbb{E}[X])^2] = c^2 \mathbb{E}[(X - \mathbb{E}[X])^2]$$

$$\sigma^2 = Var(S) \le \frac{\mu^2}{\ell}$$

$$\sigma \le \frac{\mu}{\sqrt{\ell}}$$

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

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

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

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

$$Pr(|S-\mu| \ge \mu \cdot \epsilon) \le \delta$$

$$\Leftrightarrow$$

$$Pr(S \ge \mu + \mu \cdot \epsilon \text{ or } S \le \mu - \mu\epsilon) \le \delta$$

$$Pr(A) \le \delta \qquad Pr(A) + Pr(\bar{A}) = 1$$

$$Pr(A) = 1 - Pr(\bar{A})$$

$$1 - Pr(\bar{A}) \le \delta$$

$$1 - \delta \le Pr(\bar{A})$$

$$1 - \delta \le Pr(\mu - \mu\epsilon \le S \le \mu + \mu \cdot \epsilon)$$

$$Pr((1-\epsilon) \cdot \mu \le S \le (1+\epsilon) \cdot \mu) \ge 1 - \delta$$

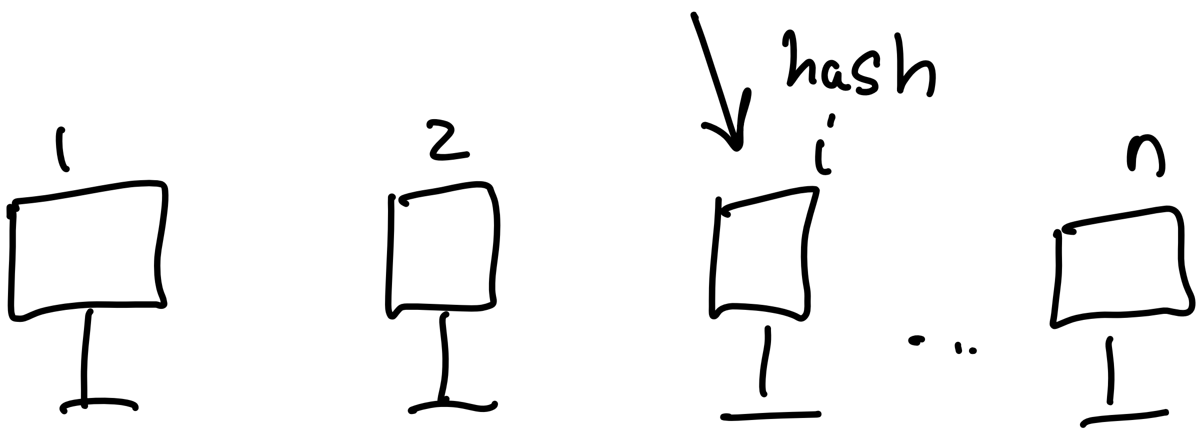
space $O(\ell) = O\left(\frac{1}{\epsilon^2 \cdot \delta} \log D\right)$

Load Balancing

$x_1, x_2, \ldots, x_m$ requests

"

"Middlebury to NYC"



want:

↳ same request to same server

↳ no server overloaded

server load to the ith server

$$S_i = \sum_{j=1}^{m} \underbrace{\mathbb{1}[h(x_j)=i]}_{S_{i,j}} = \begin{cases} 1 \text{ if } h(x_j)=i \\ 0 \text{ else} \end{cases}$$

we want to understand

$$S = \max_i S_i$$

$$\mathbb{E}[S_i] = \mathbb{E}\Big[\sum_{j=1}^{m} \mathbb{1}[h(x_j)=i]\Big]$$

$$= \sum_{j=1}^{m} \mathbb{E}[\mathbb{1}[h(x_j)=i]]$$

$$= \sum_{j=1}^{m} \frac{1}{n} = \frac{1}{n}\sum_{j=1}^{m} 1 = \frac{1}{n} m$$

$S = \max_i S_i$

$\mathbb{E}[\max_i S_i] \overset{?}{\neq} \max_i \mathbb{E}[S_i]$

e.g.,

$A = \begin{cases} 1 & \text{wp } 1/2 \\ 0 & \text{wp } 1/2 \end{cases}$

$B = \begin{cases} 1 & \text{wp } 1/2 \\ 0 & \text{wp } 1/2 \end{cases}$

A	B	max{A,B}	
0	0	0	1/4
0	1	1	1/4
1	0	1	1/4
1	1	1	1/4

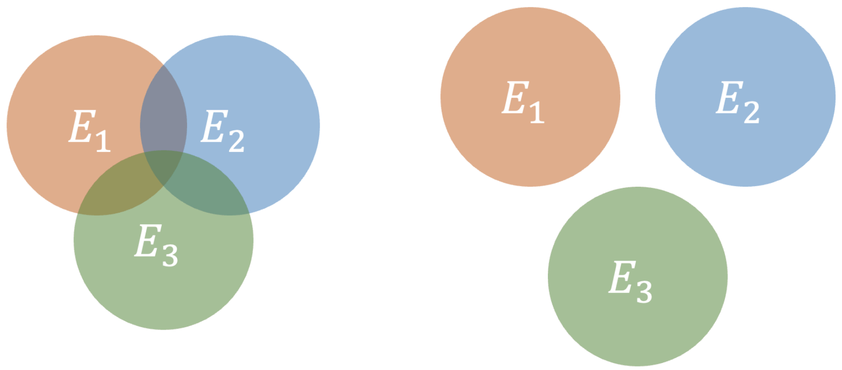
wp $\Pr(A=0 \cap B=0)$
$= \Pr(A=0)\Pr(B=0)$
$= \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

$\mathbb{E}[\max\{A,B\}] = \sum_x x \Pr(X=x)$

$= 0 \cdot \frac{1}{4} + 1 \cdot \frac{3}{4} = \frac{3}{4}$

$\mathbb{E}[A] = 1 \cdot 1/2 + 0 \cdot 1/2 = 1/2 = \mathbb{E}[B]$

Union Bound



$$\Pr(E_1 \cup E_2 \cup E_3 \cup \dots \cup E_n)$$
$$\le \Pr(E_1) + \dots + \Pr(E_n)$$

$$\Pr(X \ge t) \le \frac{\mathbb{E}[X]}{t}$$

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

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

$$\Pr\left(\sum_{i=1}^{n} \mathbb{1}[E_i] \ge t\right) \le \frac{\sum_{i=1}^{n} \Pr(E_i)}{t}$$

$t = 1$

$$\Pr(E_1 \cup E_2 \cup E_3 \cup \dots \cup E_n)$$
$$= \Pr\left(\sum_{i=1}^{n} \mathbb{1}[E_i] \ge 1\right) \le \sum_{i=1}^{n} \Pr(E_i)$$

$$\Pr\left(\max_i S_i \geq \overset{\swarrow \text{ some value}}{C}\right) \overset{\text{want}}{\leq} 1/10$$

$$\Longleftrightarrow$$

$$\Pr\left((S_1 \geq C) \cup (S_2 \geq C) \cup \ldots \cup (S_n \geq C)\right) \overset{\text{want}}{\leq} 1/10$$

$$\leq \Pr(S_1 \geq C) + \Pr(S_2 \geq C) + \ldots + \Pr(S_n \geq C) \overset{\text{want}}{\leq} 1/10$$

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

$m = n$ simplifying assumption

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

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

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

$$= m \cdot Var(S_{i,j}) \leq m \cdot \frac{1}{n} = 1$$

$\mathbb{E}[S_{i,j}] = Pr($ ith server gets jth item$) = \frac{1}{n}$

$$\mathbb{E}[S_{i,j}^2] = 1^2 \cdot \frac{1}{n} + 0^2 \cdot \left(1 - \frac{1}{n}\right) = \frac{1}{n}$$

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

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

$\sigma^2 = 1 \qquad \sigma = 1$

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

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

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

Union bound: analyze max
of random variables

Chebyshev: bound probability max rv is large

linearity of variance: compute variance for chebyshev