Plan

Logistics

Hashing Around the Clock

Concentration Inequalities

Load Balancing (Review)

Games!

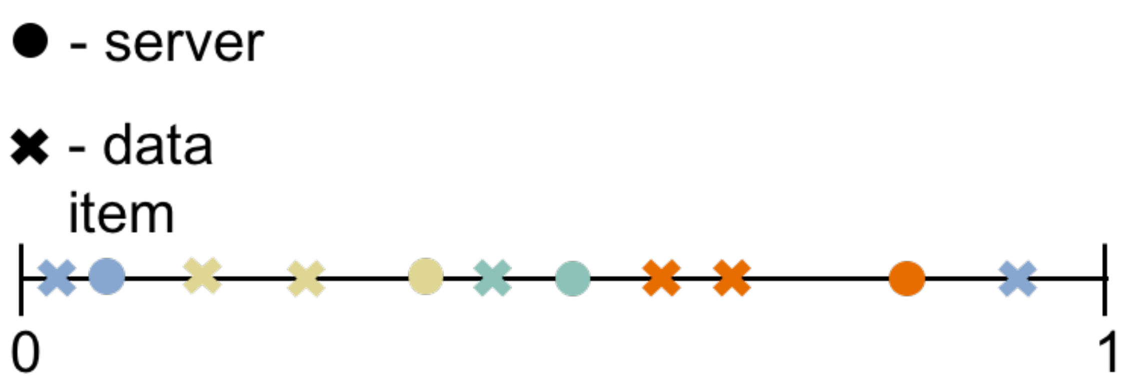
Problem set due tomorrow

Not available tomorrow,

↳ ask me today!

↳ post on canvas

Hashing Around the Clock



(1) $\mathbb{E}[\#\text{requests to more}]$

(2) $\Pr(\text{any server "owns"} \geq C \text{ fraction}) \leq 1/10$

union bound

$\Pr(\text{one server "owns"} \geq c) \leq \frac{1}{10n}$

$\Pr(\text{one server "owns"} \geq c)$

$= (1-c)^{n-1}$

(then a miracle occurs)

$\leq \frac{1}{10n}$

Concentration Inequalities

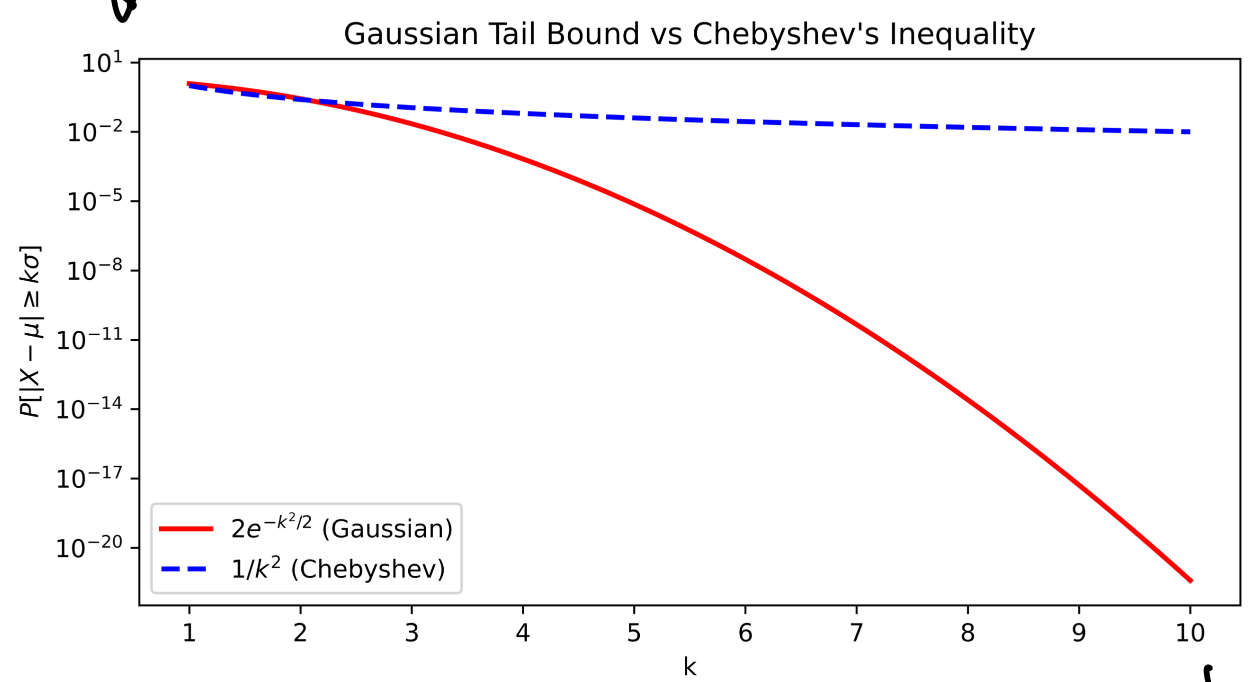
Goal: Develop "stronger" inequalities

$\mu \triangleq \mathbb{E}[X] \qquad \sigma^2 \triangleq \mathrm{Var}(X)$

$$\boxed{\begin{array}{l}\text{Chebyshev} \\ \Pr(|X-\mu| \geq k\cdot\sigma) \leq \dfrac{1}{k^2}\end{array}}$$

$$\boxed{\begin{array}{l}\text{Gaussian } X \\ \Pr(|X-\mu| \geq k\cdot\sigma) \leq 2e^{-k^2/2} = 2\cdot\dfrac{1}{e^{k^2/2}}\end{array}}$$

log scale



Q: Is Chebyshev just bad?

Chebyshev

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

$\Rightarrow$ $\mu = 0$, $\sigma = 1$

"tight"

$$\Pr(|X| \ge k) = \frac{1}{k^2}$$

$$\Pr(X \ge k \text{ or } X \le -k) = \frac{1}{k^2}$$

want equal

$$\Pr(X \ge k) + \Pr(X \le -k) = \frac{1}{k^2}$$

$$\Pr(X \ge k) = \frac{1}{2k^2}$$

$$X = \begin{cases} k & \text{wp } \frac{1}{2k^2} \\ 0 & \text{wp } 1 - \frac{1}{k^2} \\ -k & \text{wp } \frac{1}{2k^2} \end{cases}$$

$$\mathbb{E}[X] = k\cdot\frac{1}{2k^2} + 0\cdot f + -k\cdot\frac{1}{2k^2} = 0$$

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X^2] = (k)^2\cdot\frac{1}{2k^2} + 0\cdot + (-k)^2\cdot\frac{1}{2k^2}$$

$$= \frac{k^2}{2k^2} + \frac{k^2}{2k^2} = 1$$

⇒ We need assumptions for stronger concentration inequalities

Central Limit Theorem: Any sum of $\boxed{\textit{mutually independent}}$ and identically distributed random variables $X_1, \ldots, X_k$ with mean μ and finite variance σ^2 converges to a Gaussian random variable with mean $k \cdot \mu$ and variance $k \cdot \sigma^2$ as k goes to infinity. Formally, as $k \to \infty$!!

$$\cancel{\lim_{n\to\infty}} \sum_{i=1}^{k} X_i \sim \mathcal{N}(k\mu, k\sigma^2).$$

mean ✓ variance ✓

$$\Pr(X_1 = x_1, \ldots, X_k = x_k) \overset{*}{=} \Pr(X_1 = x_1) \cdots \Pr(X_k = x_k)$$

$$= \Pr(X_1 = x_1) \cdot \cdots$$

$$\overset{*}{=} \Pr(X_1 = x_1) \cdot \Pr(X_2 = x_2 \mid X_1 = x_1) \Pr(X_3 = x_3 \mid X_1 = x_1, X_2 = x_2) \cdots$$

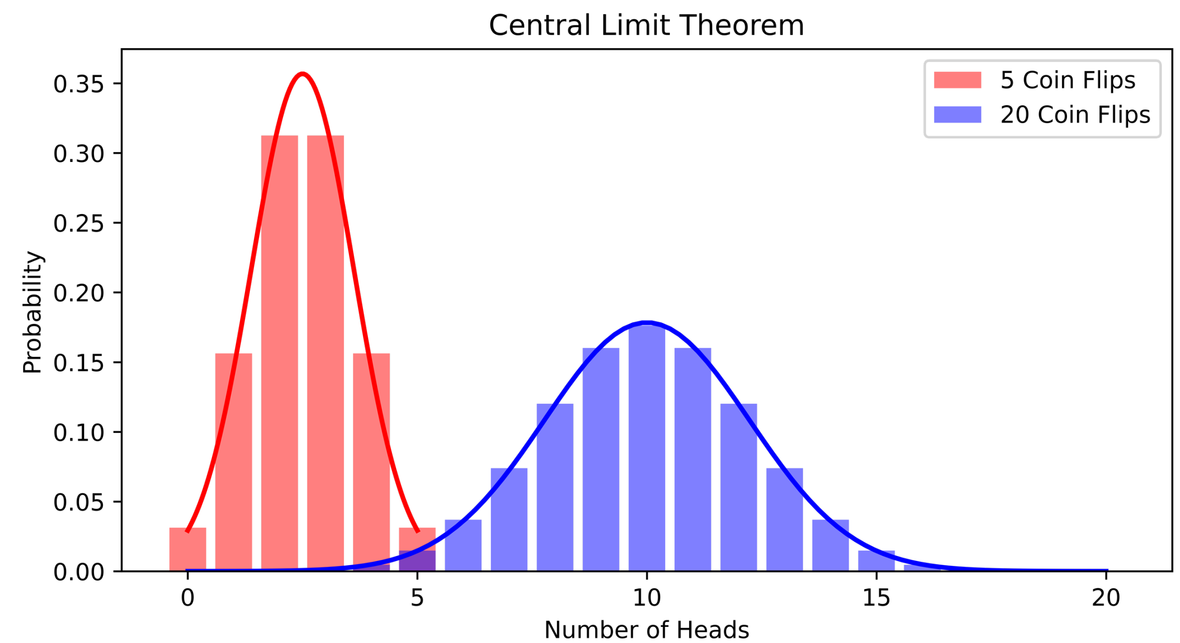
always

$$\mathbb{E}[X_i] = \mu$$

$$\mathbb{E}\left[\sum_{i=1}^{k} X_i\right] = \mu k$$

$$\mathrm{Var}(X_i) = \sigma^2$$

$$\mathrm{Var}\left(\sum_{i=1}^{k} X_i\right) = k \cdot \sigma^2$$



coin flip

$$H = \sum_{i=1}^{100} c_i$$

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

$$\mathbb{E}[H] = 50$$

$$Var(H) = 25 = \sigma^2 \qquad \sigma = 5$$

$$Var(c_i) = \mathbb{E}[c_i^2] - \mathbb{E}[c_i]^2$$
$$= 1/2 - 1/4 = 1/4$$

Markov's $Pr(X \geq 70) \leq 5/7$

Chebyshov's $Pr(|X - 50| \geq 20) \leq .0625$

Gaussian **If** CLT held exactly

$$Pr(|X - 50| \geq k \cdot 5) \leq 2e^{-k^2/2}$$

$$k = 4$$

$$Pr(|X - 50| \geq 20) \leq 2\exp\left(-\frac{16}{2}\right)$$
$$\approx .00067$$

Let's be formal!

$\hookrightarrow$ Different forms

$\hookrightarrow$ Use typed notes and/or wikipedia

$\hookrightarrow$ More assumptions $\rightarrow$ Better bounds

$2+\epsilon \leq 3$ when $\epsilon < 1$

$\frac{1}{3} \leq \frac{1}{2+\epsilon}$

$-\frac{1}{2+\epsilon} \leq \frac{-1}{3}$

indicator

Chernoff Bound: Let $X_1, \ldots, X_k$ be independent $\boxed{\text{binary}}$ random variables. That is, $X_i \in \{0,1\}$. Let $p_i = \mathbb{E}[X_i]$ where $0 < p_i < 1$. Choose a parameter $\epsilon > 0$. Then the sum $S = \sum_{i=1}^k X_i$, which has mean $\mu = \sum_{i=1}^k p_i$, satisfies

ϵ can be large

$$\Pr(S \geq (1+\epsilon)\mu) \leq \exp\left(\frac{-\epsilon^2 \mu}{2+\epsilon}\right)$$

and, if $0 < \epsilon < 1$,

$$\Pr(S \leq (1-\epsilon)\mu) \leq \exp\left(\frac{-\epsilon^2 \mu}{2}\right).$$

$$\Pr(|S-\mu| \geq \epsilon\mu)$$

$$\Pr(S \geq \mu + \epsilon\mu \text{ or } S \leq \mu - \epsilon\mu)$$

$$= \Pr(S \geq (1+\epsilon)\mu) + \Pr(S \leq (1-\epsilon)\mu)$$

$$\leq \exp\left(-\frac{\epsilon^2\mu}{2+\epsilon}\right) + \exp\left(\frac{-\epsilon^2\mu}{2}\right)$$

$$\leq \exp\left(-\frac{\epsilon^2\mu}{3}\right) + \exp\left(-\frac{\epsilon^2\mu}{3}\right)$$

$$= 2\exp\left(-\frac{\epsilon^2\mu}{3}\right)$$

$e^x = \exp(x)$

Less restrictive!

any value between -1 and 1

Bernstein Inequality: Let $X_1, \ldots, X_k$ be independent random variables with each $X_i \in [-1, 1]$. Let $\mu = \sum_{i=1}^{k} \mathbb{E}[X_i]$ and $\sigma^2 = \sum_{i=1}^{k} \mathrm{Var}[X_i]$. Then, for any $k \leq \frac{\sigma}{2}$, the sum $S = \sum_{i=1}^{k} X_i$ satisfies

$$\Pr(|S - \mu| > k\sigma) \leq 2 \exp\left(\frac{-k^2}{4}\right).$$

any value between a_i and b_i

Hoeffding's Inequality: Let $X_1, \ldots, X_k$ be independent random variables with each $X_i \in [a_i, b_i]$. Let $\mu = \sum_{i=1}^{k} \mathbb{E}[X_i]$. Then, for any $k > 0$, the sum $S = \sum_{i=1}^{k} X_i$ satisfies

$$\Pr(|S - \mu| > k) \leq 2 \exp\left(\frac{-k^2}{\sum_{i=1}^{k}(b_i - a_i)^2}\right).$$

Coin Flips

$$X_i = \begin{cases} 1 & \text{wp } b \\ 0 & \text{wp } 1-b \end{cases}$$

$$S = \sum_{i=1}^{k} X_i$$

Choose $k \geq \frac{3\log(2/\delta)}{\epsilon^2}$

(1) $\mathbb{E}[S] = bk = \mu$

(2) $\Pr(|S - bk| \geq \epsilon k) \leq \delta$

Chernoff Bound: Let $X_1, \ldots, X_k$ be independent binary random variables. That is, $X_i \in \{0, 1\}$. Let $p_i = \mathbb{E}[X_i]$ where $0 < p_i < 1$. Choose a parameter $\epsilon > 0$. Then the sum $S = \sum_{i=1}^{k} X_i$, which has mean $\mu = \sum_{i=1}^{k} p_i$, satisfies

$$\Pr(S \geq (1+\epsilon)\mu) \leq \exp\left(\frac{-\epsilon^2 \mu}{2+\epsilon}\right)$$

and, if $0 < \epsilon < 1$,

$$\Pr(S \leq (1-\epsilon)\mu) \leq \exp\left(\frac{-\epsilon^2 \mu}{2}\right).$$

$$\Pr(|S - \mu| \geq \epsilon\mu) \leq 2\exp\left(\frac{-\epsilon^2 \mu}{3}\right)$$

$$\Pr(|S-\mu| \geq \epsilon'\mu) \leq 2\exp\left(\frac{-\epsilon'^2\mu}{3}\right)$$

$$b \leq 1$$

$$\frac{1}{1} \leq \frac{1}{b}$$

$$-\frac{1}{b} \leq -1$$

$$\mu = bk$$

$$\Pr(|S-bk| \geq \epsilon' \cdot bk) \leq 2\exp\left(-\frac{\epsilon'^2 bk}{3}\right)$$

$$\epsilon k = \epsilon' bk$$

$$\epsilon' = \frac{\epsilon}{b}$$

$$k \geq \frac{3\log(2/\delta)}{\epsilon^2}$$

$$\Pr(|S-bk| \geq \epsilon k) \leq 2\exp\left(-\frac{\epsilon^2}{b^{\cancel{2}}} \cdot \frac{\cancel{b}k}{3}\right)$$

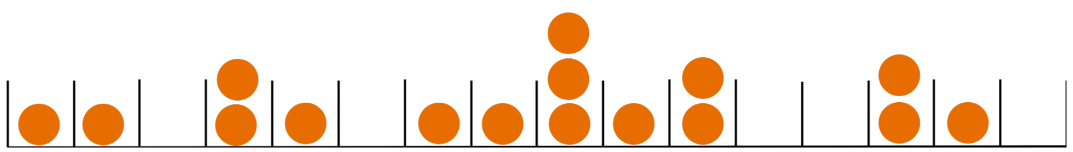
$$\leq 2\exp\left(-\frac{\cancel{\epsilon^2}}{b\,\cancel{3}} \cdot \frac{\cancel{3}\log(2/\delta)}{\cancel{\epsilon^2}}\right)$$

$$\leq 2\exp\left(-\frac{\log(2/\delta)}{b}\right)$$

$$\leq 2\exp(-\log(2/\delta)) = 2\,(2/\delta)^{-1} = 2\cdot\frac{\delta}{2}$$

Load Balancing

m requests to n servers



union

$$\Pr(\max_i S_i \ge c) \overset{\text{want}}{\le} 1/10$$

$$\Pr(S_i \ge c) \overset{\text{want}}{\le} \frac{1}{10n}$$

$$S_i = \sum_{j=1}^{m} \mathbb{1}[j \text{ goes to } i]$$

assumption $m = n$

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

Chernoff

$$\Pr(S_i \ge (1+\epsilon)\cdot\mu) \le \exp\left(-\frac{\epsilon^2\cdot\mu}{2+\epsilon}\right)$$

$$\Pr(S_i \ge (1+\epsilon)) \le \exp\left(-\frac{\epsilon^2}{2+\epsilon}\right)$$

$$\exp\left(-\frac{\epsilon^2}{2+\epsilon}\right) \overset{\text{want}}{\le} \frac{1}{10n}$$

$$\exp\left(-\frac{\epsilon^2}{2+\epsilon}\right) \overset{*}{\le} \exp\left(\frac{-\epsilon^2}{2\epsilon}\right) = \exp\left(-\frac{\epsilon}{2}\right) \overset{\text{want}}{=} 1/10n$$

$\epsilon \ge 2$

$2+\epsilon \overset{*}{\le} 2\epsilon$

$$-\frac{\epsilon}{2} = \log(1/10n)$$

$$\epsilon = 2\log(10n)$$

Practice: Hash request to 2 servers then choose the server with smaller load

$O(\log n)$ or $O(\log\log n)$ or $O(1)$ max load whp

atoms in universe $\approx 10^{82}$

$$\log_{10} \log_{10} 10^{82} = \log_{10} 82 \approx 1.91$$