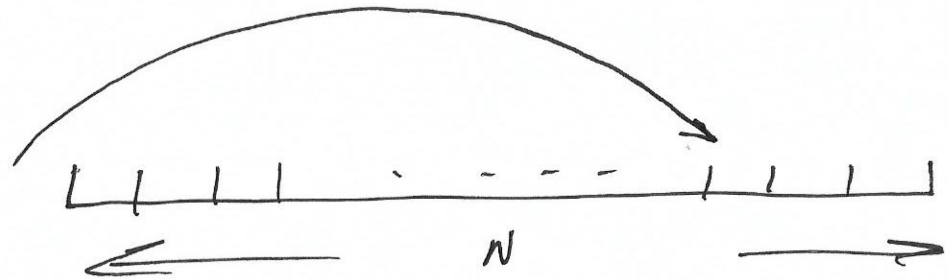
Hashing

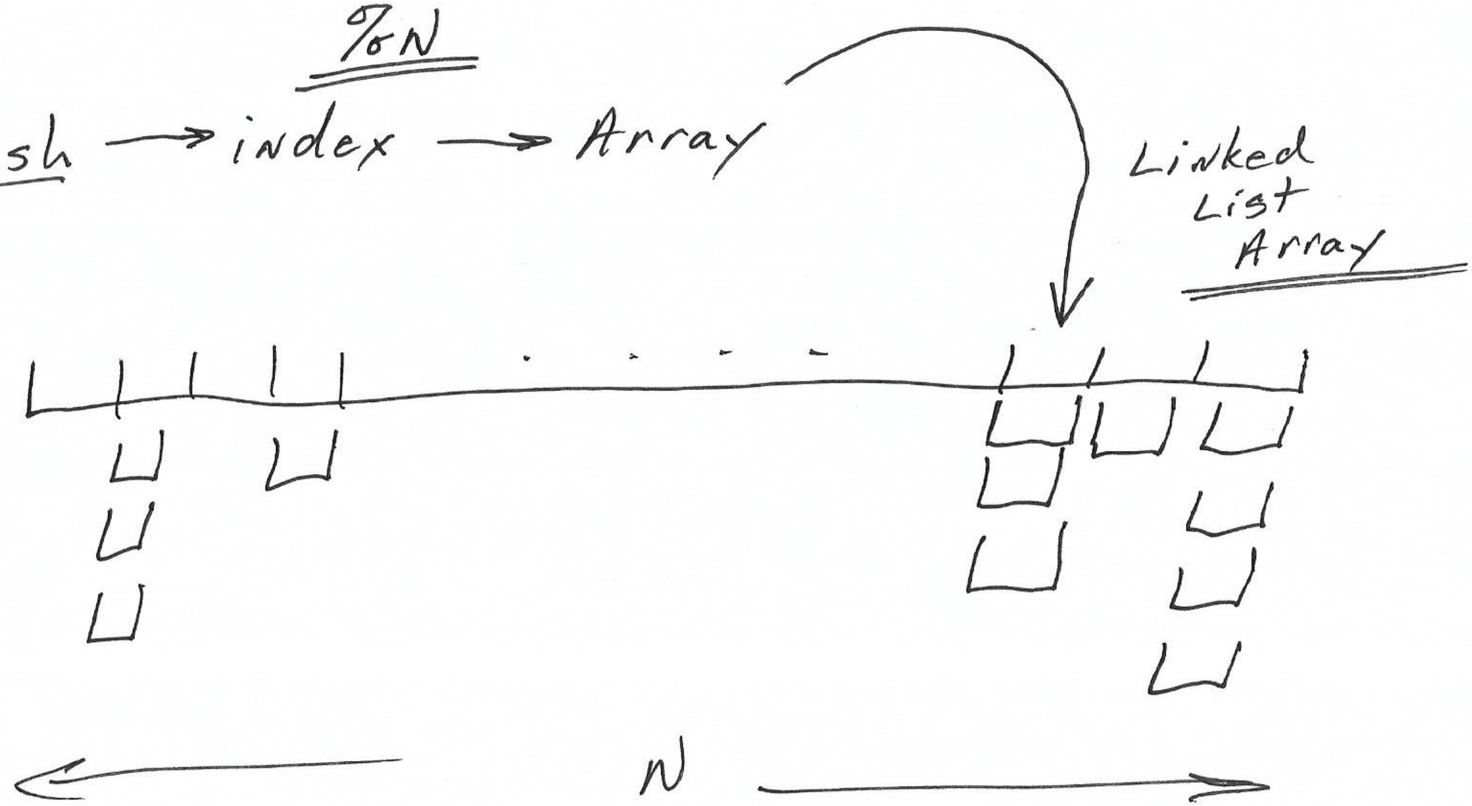
Like

index → Array →



$\%N$

Key String → Hash → index → Array



Hashing

Straight counting — Conditional Probabilities

$$P(c) = 1 - P(\bar{c})$$

$$P(\bar{c}) = \prod_{i=0}^{m-1} \left(1 - i/N\right) \longrightarrow \frac{P_{m,N}}{P^{R}_{m,N}}, \quad \prod_{i=1}^{M} P\left(\bar{C}_i / \bigcap_{j=1}^{i-1} \bar{C}_j\right)$$

$$\approx e^{-m^2/2N} \quad \text{using } \left(1 - i/N\right) \approx e^{-i/N}$$

$$E\left(\sum_{i=1}^{N} x_i / H_k\right) \rightarrow E_k = C\,E_{k-1} + 1, \quad E_0 = 0,\ E_1 = 1$$

$$E_k = N\left(1 - \left(1 - 1/N\right)^k\right) \approx N\left(1 - e^{-k/N}\right)$$

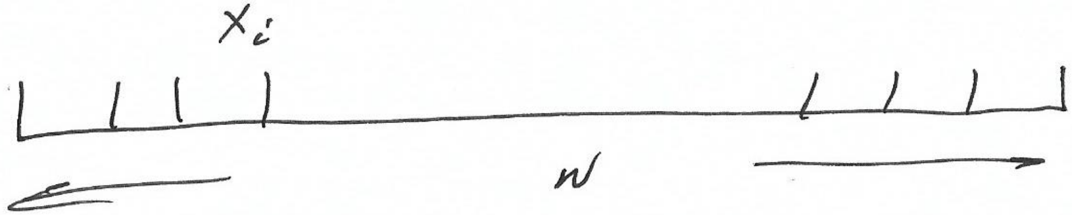
$$E_{km} \approx N\left(1 - e^{-km/N}\right)$$

$$P_{km} = E_{km}/N \approx \left(1 - e^{-km/N}\right)$$

Bloom Filter

$$P_{fa} = P_{fp} = P_{km}^{k} \rightarrow \text{Binomial} \approx \left(1 - e^{-km/N}\right)^k$$

MAX, Expected, Distribution of Hashing Collisions



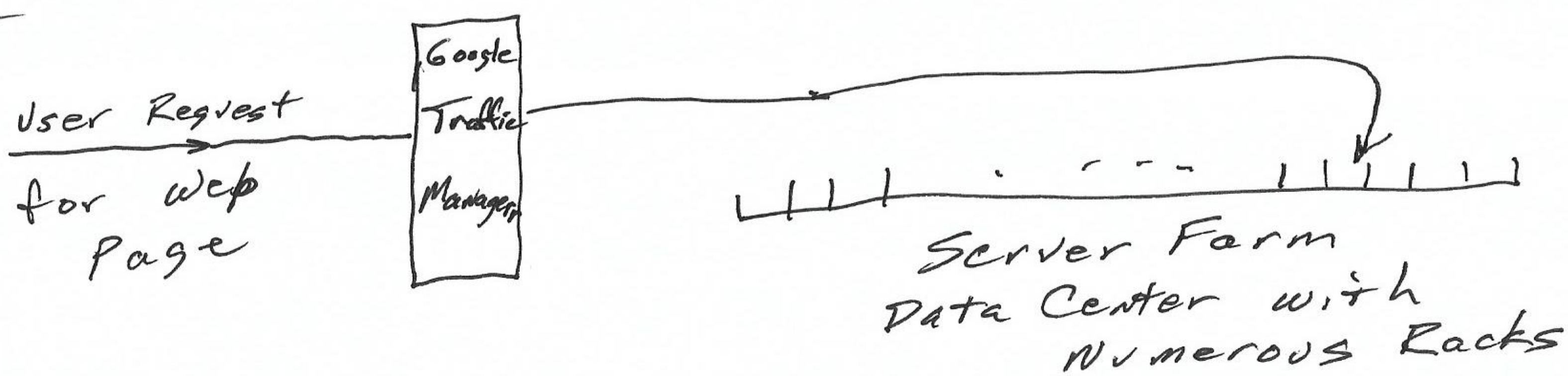
$P(X < k)$ where k in this case mean # of occupants in same cell $0 \le k \le N$

$P(X < k) = P$(fewer than K occupants in entire array)

$= 1 - P(X \ge k)$

$P(X \ge k)$ → represents a value of K we don't want to reach / exceed

Example



$$P(X \geq k) = P\left(\bigcup_{i=1}^{N}(X_i \geq k)\right)$$

Use Union Upper Bound

$$P\left(\bigcup_{i=1}^{N}(X_i \geq k)\right) \leq \sum_{i=1}^{N} P(X_i \geq k)$$

Example

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) + - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

$$\leq P(A) + P(B) + P(C)$$

So

$$P(X \geq k) \leq \sum_{i=1}^{N} P(X_i \geq k)$$

<u>Boundary Point / Inflection</u>

just like in $P(c) = 1 - P(\bar{c})$

Choose $P(c) = P(\bar{c}) = 1/2$

$$1/2 \overset{?}{\leq} N\, P(X_i \geq k) \Longrightarrow 1/2N$$

$P(X_i \geq k) \Rightarrow$ Binomial Distribution

$$= \sum_{j=k}^{N} \binom{N}{j} p^{j} q^{N-j}$$

where X_i is just 1 out of N

$p = 1/N \quad q = (1 - 1/N)$

$$= \sum_{j=k}^{N} \binom{N}{j} \left(1/N\right)^{j} \left(1 - 1/N\right)^{N-j}$$

Use the Bounds "Lower/Upper" of $\binom{N}{j} = \frac{N!}{(N-j)!\,j!}$

$$\underline{\underline{\left(\frac{N}{j}\right)^{j}}} \leq \binom{N}{j} \leq \frac{N^{j}}{j!} \leq \underline{\underline{\left(\frac{Ne}{j}\right)^{j}}}$$

Published | M.E.L. | Published

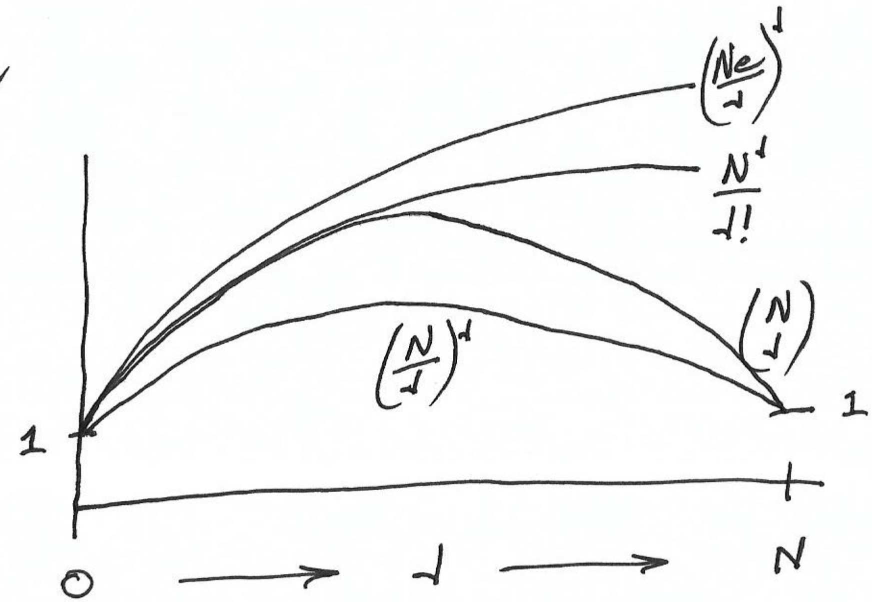
$$\binom{N}{d} = \frac{N!}{(N-d)!\,d!} = \frac{N\cdot(N-1)\cdot(N-2)\cdots(N-d+1)}{d\cdot(d-1)\cdot(d-2)\cdots 1} < \frac{N^d}{d!}$$

$$= \prod_{i=0}^{d-1}\left(\frac{N-i}{d-i}\right) < \frac{N\cdot N\cdot \cdots N}{d\cdot(d-1)\cdot \cdots 1}$$

$$= \prod_{i=0}^{d-1}\left(\frac{N-i}{d-i}\right) < \prod_{i=0}^{d-1}\left(\frac{N}{d-i}\right)$$

because $N-i \le N \quad \forall\ 0 \le i \le N$

Comparison



Include approximations for an estimate $N = f(k)$

$$P(x_i \geq k) = \sum_{d=k}^{N} \binom{N}{d} p^d q^{N-d}$$

x_i is just 1 out of N

$P(i) = 1/N$, $q = (1 - 1/N)$

$$\downarrow$$

$$e^{-1/N} \approx (1 - 1/N)$$

$$\leq \sum_{d=k}^{N} \frac{\cancel{N}^{d}}{d!} \left(\frac{1}{\cancel{N}}\right)^d \left(1 - \frac{1}{N}\right)^{N-d}$$

$$\leq \sum_{d=k}^{N} \frac{e^{-\frac{(N-d)}{N}}}{d!}$$

$$\leq e^{-1} \sum_{d=k}^{N} \frac{e^{d/N}}{d!}$$

At this point expand series and Simplify

Next page

Expanding

$$P(X_i \geq k) \leq e^{-1}\sum_{d=k}^{N} e^{d/N}/d! = e^{-1}\left(e^{k/N}/k! + e^{k+1/N}/k!(k+1) + e^{k+2/N}/k!(k+1)(k+2) + \dots\right.$$

$$\leq e^{-1} e^{k/N}/k! \underbrace{\left(1 + e^{1/N}/(k+1) + e^{2/N}/(k+1)(k+2) + \dots\right.}_{\substack{\text{What is max this} \\ \text{could sum to?} \\ < 7/4}}$$

also $e^{k/N}$ $k \ll N$ $= 1$

$$1/2N \leq \frac{7/4\, e^{-1}}{k!}$$

Almost Done

Invert the previous equation

$$2N \geq \frac{4}{7} e\, k!$$

$$N \geq \frac{2}{7} e\, k!$$

$$\underline{\underline{N \geq 0.78\, k!}}$$

The Big Take out of this is that for very large N the max number of collisions in any 1 cell of the array is very small. <u>The inverse factorial</u>

$$f(k_{max}) = O\left((N)^{-1}!\right)$$

The probabilty density function is approximately

$$P(x_i = k) = e^{-1}/k!$$

The expected value or average number of collisions or views to find the data would be

$$E(k) = \frac{\sum_{k=0}^{N} e^{-1}/k!\, k}{1 - \underset{\substack{\uparrow \\ P(k=0)}}{e^{-1}}} = \frac{e^{-1} \sum_{k=1}^{N} 1/k! = e^{-1} e^{1}}{1 - e^{-1}}$$

$$= \frac{1}{1 - e^{-1}} \approx \underline{\underline{1.582}}$$