Lecture 20 Randomized Algorithm 2

Wednesday, April 8, 2020 3:00 PM

- quick selection

$a[\,] = \{7, 2, 5, 8, 3, 4, 1, 6\}$

$k = 6$

pivot = 2

$\boxed{\{1\}}$ $\underline{2}$ $\underbrace{\{7, 5, 8, 3, 4, 6\}}_{\text{4-th smallest}}$

pivot = 7

$\underbrace{\{5, 3, 4, 6\}}_{\text{4-th smallest}}$ 7 $\{8\}$

5-th

Quick Selection(a[ ], k)
base case
pick random pivot
Partition the array using pivot
recurse in the part that contains k-th smallest number

pivot = i $\Rightarrow$ pivot number is the i-th smallest

let X_n be the running time of quickselect on n numbers

$$E[X_n] = \sum_{i=1}^{n} E[X_n \mid \text{pivot}=i] \times \underbrace{Pr[\text{pivot}=i]}_{\frac{1}{n}}$$

$$E[X_n \mid \text{pivot}=i] = \begin{cases} \text{if } i<k & \text{recurse on right part } X_{n-i} \\ i=k & \text{output pivot, } 0 \\ i>k & \text{recurse on left part } X_{i-1} \end{cases} + n$$

$$E[X_n] = \sum_{i=1}^{k-1} \frac{1}{n} \cdot (X_{n-i} + n) + \underset{(i=k)}{\frac{1}{n} \cdot n} + \sum_{i=k+1}^{n} \frac{1}{n} (X_{i-1} + n)$$

$$= n + \sum_{i=1}^{k-1} \frac{X_{n-i}}{n} + \sum_{i=k+1}^{n} \frac{X_{i-1}}{n}$$

worst case: $k = \frac{n}{2}$

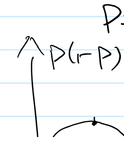
$$E[X_n] \leq n + \sum_{i=1}^{\frac{n}{2}-1} \frac{X_{n-i}}{n} + \sum_{i=\frac{n}{2}+1}^{n} \frac{X_{i-1}}{n}$$

- Monte Carlo example

X_i be random var $= \begin{cases} 1 & \text{if } (x_i, y_i) \text{ in circle} & P \\ 0 & \text{if not} & 1-P \end{cases}$

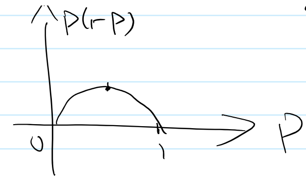
$$\text{Count} = \sum_{i=1}^{n} X_i \qquad P = \frac{\text{area of circle}}{4}$$

$$Var[X_i] = P(1-P) \leq \frac{1}{4}$$



$Count = \sum_{i=1} \cdots$ 4

$$Var[X_i] = p(1-p) \le \frac{1}{4}$$



$$Var[Count] \le \frac{n}{4}$$

Chebyshev. $\Pr\left[\,|Count - pn| \ge \underset{2\sqrt{Var}}{\sqrt{n}}\,\right] \le \frac{1}{4}$

when this does not happen $\left|\frac{Count}{n}\cdot 4 - p\cdot 4\right| \le \frac{4}{n}|Count - pn|$

$$\le \frac{4}{\sqrt{n}}$$

if we choose $n \ge \frac{16}{\varepsilon^2}$, with probability $\ge \frac{3}{4}$

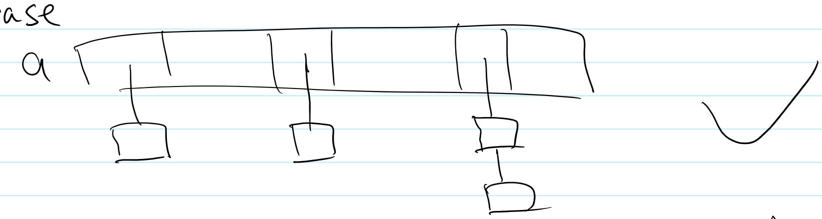
$$\left|\frac{Count}{n}\cdot 4 - p\cdot 4\right| \le \varepsilon$$

- Hashing
  - set problem: maintain dynamic subset of $\underbrace{\{0, 1, \cdots N-1\}}_{\text{universe}}$
  - supported operation: $\boxed{\begin{array}{l}\text{insert}\\ \text{delete}\end{array}}$ look-up
  - goal: 1. all 3 operations $O(1)$ (max)
    2. space is proportional to the size of the set (independent of size of universe)
  - design: allocate an array $a[0, \cdots m-1]$
    $\boxed{m = \Theta(\text{size of set})}$ size of set $= n =$ number of elements in the hash-table
    hash function: $f: \{0, \cdots, N-1\} \longrightarrow \{0, \cdots, m-1\}$
    - idea: put number i in location $f(i)$
    - problem: collision: x, y in the set where
      $$f(x) = f(y)$$
    - solution: maintain a linked list at every $a[\,]$ location
      add all numbers with same $f(x)$ to the linked list.
    - problem: running time: look-up $\Theta($length of list at$)$

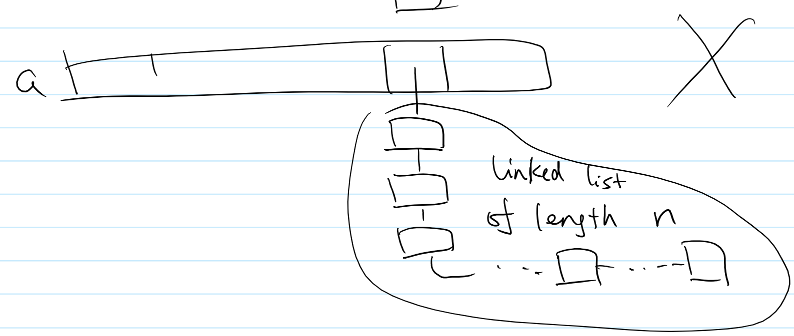
ust.

- problem: running time: look-up $\Theta\left(\begin{array}{c}\text{length of list at} \\ a[\underline{f(x)}]\end{array}\right)$

- best case



- worst-case



- hash families

  - totally random
    # functions $\{0, 1, \dots, N-1\} \to \{0, 1, \dots, m-1\}$

    for every number in $\{0, 1, \dots, N-1\}$, has m choices m^{N} number is too large

    space to store a random hash func: $\log_2 m^N = N \log_2 m > N$

  - pairwise independent / universal hash family

    say a family $\mathcal{F}$ of hash functions is pairwise independent/universal if for every $x \neq y \in \{0, \dots, N-1\}$

    $$\boxed{\Pr_{f \sim \mathcal{F}}[f(x) = f(y)] = \frac{1}{m}}$$

  - example: suppose hash table already contains n numbers $x_1, \dots, x_n$

    insert a new number $y \neq x_1, \dots, x_n$

    Q: what is the expected number of x_i that collides with y

    Let $X_i = \begin{cases} 1 & f(x_i) = f(y) \\ 0 & f(x_i) \neq f(y) \end{cases}$

    $$\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \underline{\mathbb{E}[X_i]} = \sum_{i=1}^{n} \Pr[X_i = 1]$$

    $$= \sum_{i=1}^{n} \underbrace{\Pr[f(x_i) = f(y)]}_{\frac{1}{m}}$$

    $n \cdot \frac{1}{m}$

$$= \sum_{i=1}^{n} \underbrace{\Pr[f(x_i) = f(y)]}_{\frac{1}{m}}$$

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

$$= \boxed{\frac{n}{m}}$$