Date: 16/06/21

Black Board

Design and Analysis of Algorithms

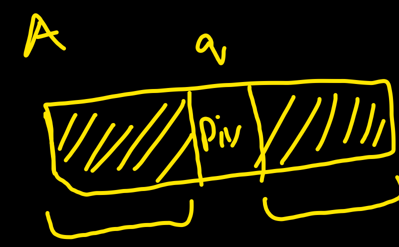
Topics:

-- **Randomized Quicksort Analysis**

-- The Expected Running time and its Variance

(i) Randomized Quick-Sort Revisited

```
Quicksort(A, p, r)
    IF p<r
        q ← partition(A, p, r)
        Quicksort(A, p, q-1)
        Quicksort(A, q+1, r)
```



Worst case:

$$T(n) = T(n-1) + O(n)$$
$$= O(n^2)$$

Claim: $\boxed{E[T(n)] = O(n \lg n)}$

$T(n) \rightarrow$ R.V

$$T(n) \begin{cases} 1 \\ \vdots \\ c\, n\lg n \\ \vdots \\ cn^2 \end{cases}$$

(ii) Proof that $E[T(n)] = O(n \lg n)$ for Rand QS

For any comp. based sort,

$$T(n) = O(\# \text{ of comparisons between data})$$

We wish to estimate the expected # of comparisons made by RQS.

Q What is the probability that the i^{th} smallest # a_i of the input & the j^{th} smallest element a_j of input are compared by RQS. $(i<j)$

$$a_1 \quad a_2 \quad a_3 \quad \ldots \quad a_i \quad \ldots \quad a_j \quad \ldots \quad a_n$$

Rand partition (consider #s in sorted order)

$$\frac{2}{j-i+1}$$

when will a_i & a_j be compared.

→ a_i & a_j will never be compared if p is between them → 7

$a_i = 5$

$a_j = 20$

→ a_i & a_j will only be compared if either a_i or a_j are picked as pivot.

$$\boxed{Pr\{a_i \,\&\, a_j \text{ are cmp}\} = \frac{2}{j-i+1}}$$

$$X_{ij} = \begin{cases} 0, & a_i \text{ \& } a_j \text{ not cmp.} \\ 1, & a_i \text{ \& } a_j \text{ are cmp.} \end{cases}$$

$$\boxed{E[X_{ij}] = \frac{2}{j-i+1}}$$

$X :=$ total number of cmp in RQs

X_{34} Same X_{43}

$$X = \sum_{i=1}^{n} \sum_{j=i+1}^{n} X_{ij}$$

$$E[X] = \sum_{i=1}^{n} \sum_{j=i+1}^{n} E[X_{ij}]$$

{ Linearity of Expectation }

$$E[X] = \sum_{i=1}^{n} \sum_{j=i+1}^{n} \frac{2}{j-i+1}$$

$$= 2 \sum_{i=1}^{n} \left[\sum_{j=i+1}^{n} \frac{1}{j-i+1} \right]$$

$$= 2 \sum_{i=1}^{n} \left[\frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \cdots + \frac{1}{n-i+1} \right]$$

$$\leq 2 \sum_{i=1}^{n} \left[1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \right]$$

$$\leq 2n \lg n \implies \boxed{T(n) = O(n \lg n)}$$

Expect Running time of RQS

worst

$T(n) = O(n^2)$ not realistic

(iii) How far does $E[T(n)]$ go from its mean $O(n \lg n)$

Are we likely to be close to $O(n \lg n)$ on most occasions?

Consider a loaded dice

$$\left\{ \begin{array}{l} Pr\{x=1\} = 1/2 \\ Pr\{x=6\} = 1/2 \end{array} \right\}$$

$Pr\{x=2\}$, $P\{x=3\}$, $P\{x=4\}$, $Pr\{x=5\} = 0$

$$E[x] = 1 \cdot 1/2 + 6 \cdot 1/2 = 3.5 = 7/2$$

Variance , $\sqrt{\text{Variance}} = \underline{\underline{s.d.}}$

$\cancel{E\{}$ $Var(X) = E\{(X - E[X])^2\}$

$= E\{(x - \frac{7}{2})^2\}$

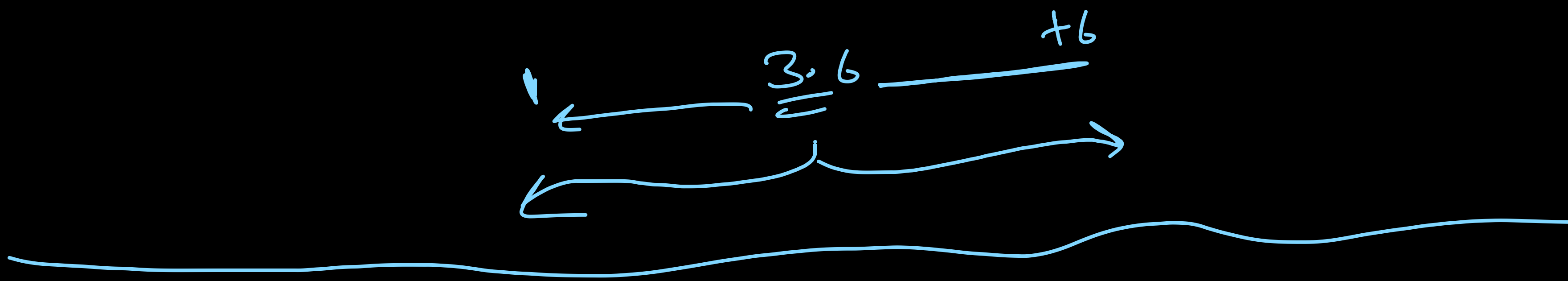
$= E[x^2 - 7x + \frac{49}{4}]$

$= E(x^2) - 7E(x) + \frac{49}{4}$

$\vdots$

≈ 6

$s.d = \underline{2.5}$



X

Variance of RQS

$s.d \simeq n$

$n\lg n - n$ $\qquad$ $n + n\lg n$

$n\lg n \simeq$ $\quad T(n) \leq n + n\lg n$

$T(n) \leq 2n\lg n$

with high prob.