DATE: / /
PAGE NO.:

Class 1: Jan 5, 2021

Advanced Algorithms

HW: 30%
Quiz: 30%
In-class Quiz: 25%
End Exam: 15%
Extra bonus: 5%

Ansi C — stdlib
└ qsort — quicksort

Runtime:
- depends on implementation & assumptions

if pivot = median, using the linear time median algorithm

so, median — $\boxed{O(n)}$
↓
pivot + partitioning
+ $\boxed{\log n}$ — recursions
$= \boxed{O(n \log n)}$

But if pivot ~~p~~icked is bad
↓
will take more time

$T(n)$ = time taken by Quick Sort

DATE: / /

PAGE NO.:

So, $T_n \leq O(n^2)$

$\because$ max value of $T(n)$ occurs when pivot is the largest / smallest element of remaining set during EACH recursive call.

In this case,

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

But the probability of this happening is low. How?

In first rec. call – prob. of choosing largest element is $\frac{1}{n}$

sec rec. ——— $\frac{1}{n-1}$

⋮

So, overall for entire algo –

$$\frac{1}{n} \times \frac{1}{n-1} \times \frac{1}{n-2} \dots \frac{1}{2} \times 1 = \boxed{\frac{1}{n!}}$$

& with smallest & largest $\longrightarrow \frac{2}{n} \times \dots = \frac{2^n}{n!}$ – still very small

So, if we assume independence of choice across these rec. calls, then the prob. that $T(n) = O(n^2)$ is $\left\{\frac{1}{n!}\right\}$ – very small.

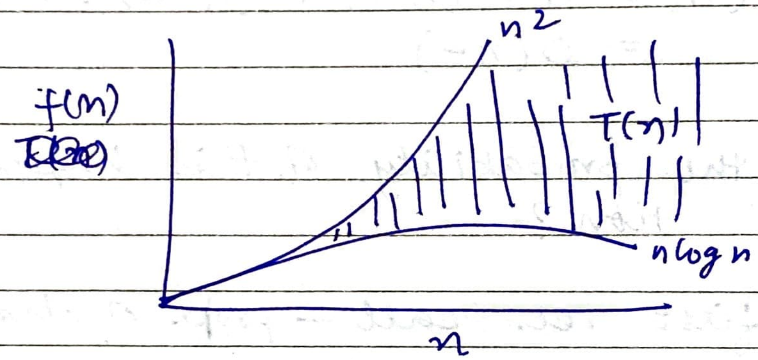
~~So, if we have uniformly random pick for pivot – we're not going to encounter the worst case very often.~~

Teacher's Signature

Equally, the best case scenario where pivot = median will also be very small

i.e. $T(n) = O(n \ln n)$ also small prob.

So, $O(n \ln n) \leq T(n) \leq O(n^2)$



So, what is expected value of $T(n)$.

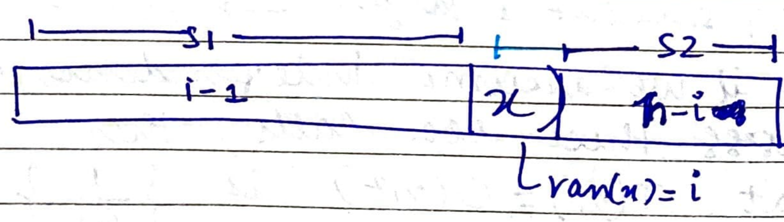
Let this be our arry —

and in the first iterat^n/recursive call:

pivot $= x$

& rank$(x) = i$

↳ no. of elements $\leq x = i$



~~pivot is added to S_2~~

So, this choice has prob. $= \frac{1}{n}$

☆ $\exists$ every element is likely to be pivot

~~So~~ if 'X' is the size of S_1

The recurrence relatn for $T(n)$ –

$$T(n) = T(X) + T(n-X-1) + \underbrace{n}_{\text{for the partition i.e separating array into } S_1 \text{ \& } S_2}$$

Now,

[event denoted by { }]

$$\{'X' = i\} \equiv \{|S_1| = i\} \equiv \{rank\ (pivot) = i+1\}$$

and $Pr(rank\ (pivot) = i+1) = 1/n$ ($\because$ ★)

So, $Pr('X' = i) = 1/n$ ⊙ (PTO)

$$\boxed{Pr(n-1-X = i) = 1/n}$$

why?

if $n-1-X = i \Rightarrow |S_2| = i$

$\Rightarrow |S_1| = n-1-i$

$\Rightarrow rank\ (pivot) = N-i-1$

& $Pr\ (rank\ (pivot) = \text{—}) = 1/n$

irrespective of what rank is

$[pr(pivot = \text{—}) = 1/n$

DATE: / /
PAGE NO.:

$$T(n) = n + T(X) + T(n-X-1)$$

Taking expectation on both sides

(using linearity of expectatⁿ on RHS)

~~expectation funᶜ is linear in its arguments~~

⇒ expectatⁿ of sum of LHS can be split over the sum of expectation of each individual term of RHS

⇒ $E(A+B) = E(A) + E(B)$

expectatⁿ of sum of r.v. = sum of their expectations

43:22

$$E[T(n)] = n + \left(\frac{1}{n}\right)\sum_{i=1}^{n-1} E[T(i)] + \left(\frac{1}{n}\right)\sum_{i=1}^{n-1} E[T(i)]$$

using the fact that for a Random variable Y with its support ~~in~~ partitioned into sets $A_1, A_2 \ldots A_n$:

$$E[Y] = \Sigma_i \Pr(A_i) \cdot E[Y \mid A_i]$$

So for 'n' we have n-1 sets. So for us:

$$E[X] = \sum_{i=1}^{n-1} \left(\frac{1}{n}\right) \cdot E[T(X) \mid |X| = i]$$

Teacher's Signature

All elements are unique. So if you choose one element you can divide the sets in only one way - So $\frac{1}{n}$

But if there were non-unique elements - i.e. - $\frac{d}{n}$ d = duplicates of the particular rank i. Or we can put an index on dup element and make them unique.

$T(i)$ is the time taken to sort the set of size i elements

same thing applies on prev page ⓞ

So, $E[X] = \frac{1}{n} \sum_{i=1}^{n-1} E[T(i)]$

and for

$E[n-X-1] \rightarrow$ same thing } because for each $X = i$ & $X = n-1-i$ there can be n unique sets of ⓧ

why?

$1/n \sum_{i=1}^{n-1} E[T(i)]$ $i \rightarrow n-1-i$ } change summation order to get same terms

So, $E[T(n)] = n + \frac{1}{n} \sum_{i=1}^{n-1} E[T(i)] + \frac{1}{n} \sum_{i=1}^{n-1} E[T(i)]$

Let $f(i) = E[T(i)]$

So, $f(n) = n + \frac{2}{n} \sum_{i=1}^{n-1} f(i)$

<u>Further simplificatⁿ:</u>

$nf(n) = n^2 + 2 \sum_i f(i)$

$nf(n) = n^2 + 2(f(1) + f(2) + \dots + f(n-1))$ — ①

Replace n by $n-1$

$(n-1)f(n-1) = (n-1)^2 + 2(f(1) + \dots + f(n-2))$ — ②

DATE: / /

PAGE NO.:

Subtracting the two equat^{n}s

$$nf(n) - (n-1)f(n-1) = n^2 - (n-1)^2 + 2(f(n-1))$$

$$nf(n) = \underline{n^2 - (n-1)^2} + 2(f(n-1)) + (n-1)f(n-1)$$

$$nf(n) = 2n-1 + (2+n-1)f(n-1)$$

$$f(n) = \frac{2n-1}{n} + \frac{(n+1)}{n} f(n-1)$$

Next, we prove by induct^n that $f(n) \leq 2n\ln n$

But first,

we check $\bar{c}$ base case:

$$f(1) = \frac{2}{1} f(0) + \frac{1}{1}$$

$$f(1) = 2f(0) + 1$$

So, pick a big enough value for $f(0)$ so, that $f(1)$ value ~~accepts~~ satisfies the condition $f(n) \leq 2n \ln n$

$$\text{So, } f(1) \leq 2 \underbrace{\log(1)}_{0}$$

$$f(1) \leq 0$$

So we can have (?) $f(0) = 0$

~~Also~~ $f(0) = -1$ (?)

DATE. / /
PAGE NO.:

Induct^n: let's assume induct^n follows all the way upto $n-1$ & we will prove for n.

So,

$$f(n) = \frac{n+1}{n} f(n-1) + \frac{2n-1}{n}$$

by induction hypothesis

$$f(n) \leq \frac{n+1}{n} 2(n-1) \ln(n-1) + \frac{2n-1}{n}$$

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

$$= \frac{2(n^2-1)}{n} \ln\left(n\left(1-\frac{1}{n}\right)\right) + \frac{2n-1}{n}$$

$$= \frac{2(n^2-1)}{n} \underbrace{\left(\ln n + \ln\left(1-\frac{1}{n}\right)\right)} + \frac{2n-1}{n} \quad \text{--- (3)}$$

why? — because we want it to be of the form [$2n \ln n$ + rest of terms]

using the standard inequality
$1 + x \leq e^x$
to remove logarithm in the expression $\underline{\ln(1-\frac{1}{n})}$

So here if $x = -1/n$

$$\ln\left(1-\frac{1}{n}\right) \leq \ln\left(e^{-1/n}\right)$$

$$\ln\left(1-\frac{1}{n}\right) \leq \frac{-1}{n} \quad \text{--- (4)}$$

Teacher's Signature

Replacing (4) in (3)

$$f(n) \le \frac{2(n^2-1)}{n}\left(\ln n - \frac{1}{n}\right) + \frac{2n-1}{n}$$

$$f(n) \le 2n \ln n - \frac{2}{n}\ln n - 2 + \frac{2}{n^2} + 2 - \frac{1}{n}$$

$$\le 2n\ln n - \frac{2}{n}\ln n + \frac{2}{n^2} - \frac{1}{n}$$

$\Big[\underbrace{-\frac{2}{n}}_{a} < \underbrace{\frac{2}{n^2}}_{b}\Big]$ — a is much more negative than $\frac{2}{n^2}$ is +ve

So $-\frac{2}{n}\ln n + \frac{2}{n^2} - \frac{1}{n} < 0$

$f(n) \le 2n \ln n$

establishing the inductive step

So, randomised quick sort $O(n \ln n)$
averaged behavior

Expected Runtime $\le 2n \ln n$ i.e. $O(n \ln n)$
but ~~who~~ how much does it differ from the actual value.

Margin Error

lets say upto $12 n \ln n$ runtime is acceptable

(*) So can we find prob that runtime won't exceed $12 n \ln n$.

So, even tho we are told that this RV values stretch from $n \log n$ to n^2 — we are told that on average it takes $2n \ln n$. But more questn's can be asked to get more informatn — to know the actual view/standpoint. like (*)

(*) - $Pr(T(n) > 12 n \ln n)$ or
$Pr(T(n) > 20 n \ln n)$ or
$Pr(T(n) > n \ln^2 n)$

} we do this analyses using tail inequalities.

Taylor Series

$\log_e = \ln$

we took, $f(n) \le 2n \ln n$

because it works easily for $1+x \le e^x$

to figure out values we need to look at taylor series.

~~In Markov inequality~~

~~$P[X \geq c\mu] \le 1/c$~~

Teacher's Signature