CSE 201 2-4-20

Average case runtime of Quicksort.

- Assume all permutations of $A[1 \cdots n]$ are equally likely; i.e. Probability of any particular arrangement is $\frac{1}{n!}$

- let $t(n)$ = average case # of array comparisons by Q.S. on $A[1 \cdots n]$, then

$$t(n) = \frac{\displaystyle\sum_{\text{all permutations}} \left(\begin{array}{l}\text{\# of comp. performed} \\ \text{by Q.S. on that perm.}\end{array}\right)}{n!}$$

* Goal: find a recurrence for $t(n)$.

note: our assumption implies that the pivot element $A[q]$ is equally likely to end up in any position: $1 \dots n$, i.e.

$$\text{Prob.}(q = i) = \frac{1}{n} \quad (\text{for } 1 \le i \le n)$$

- also recall:

Partition(A, 1, n) does $n-1$ comparisons

$$A[\underbrace{1, \ldots, n-1}, \overset{\uparrow}{n}]$$

Pivot

$$\text{length}(A[1 \ldots (q-1)]) = (q-1) - 1 + 1 = q - 1$$

$$\text{length}(A[(q+1) \ldots n]) = n - (q+1) + 1 = n - q$$

Thus

$$t(n) = \frac{\sum_{q=1}^{n} \left((n-1) + t(q-1) + t(n-q) \right)}{n}$$

$$= \frac{\sum_{q=1}^{n} (n-1) + \sum_{q=1}^{n} \left(t(q-1) + t(n-q) \right)}{n}$$

$$t(n) = \frac{n(n-1) + \sum_{q=2}^{n} t(q-1) + \sum_{q=1}^{n-1} t(n-q)}{n}$$

$$= (n-1) + \frac{1}{n}\left(\sum_{q=1}^{n-1} t(q) + \sum_{q=1}^{n-1} t(n-q)\right)$$

$$t(n) = (n-1) + \frac{2}{n} \cdot \sum_{q=1}^{n-1} t(q)$$

trick: let $x_n = \sum_{q=1}^{n-1} t(q)$ $\left(\frac{\underline{\text{note:}}}{x_1 = 0}\right)$

so

$$t(n) = (n-1) + \frac{2}{n} \cdot x_n$$

notice

$$x_{n+1}-x_n = \sum_{q=1}^{n} t(q) - \sum_{q=1}^{n-1} t(q) = t(n)$$

so

$$x_{n+1}-x_n = (n-1) + \frac{2}{n}\cdot x_n$$

$$\therefore \quad x_{n+1} - \left(1+\frac{2}{n}\right)x_n = (n-1)$$

$$x_{n+1} - \left(\frac{n+2}{n}\right)x_n = (n-1)$$

multiply by magic number: $\frac{1}{(n+1)(n+2)}$

$$\therefore \quad \frac{x_{n+1}}{(n+1)(n+2)} - \frac{x_n}{n(n+1)} = \frac{n-1}{(n+1)(n+2)}$$

Replace n by k:

$$\frac{x_{k+1}}{(k+1)(k+2)} - \frac{x_k}{k(k+1)} = \frac{k-1}{(k+1)(k+2)}$$

$$= \frac{3}{k+2} - \frac{2}{k+1}$$

sum for $k = 1$ to $(n-1)$

$$\sum_{k=1}^{n-1}\left(\frac{x_{k+1}}{(k+1)(k+2)} - \frac{x_k}{k(k+1)}\right) = \sum_{k=1}^{n-1}\left(\frac{1}{k+2} + \frac{2}{k+2} - \frac{2}{k+1}\right)$$

$$\frac{x_n}{n(n+1)} - \cancel{\frac{x_1}{2}} = \sum_{k=1}^{n-1}\frac{1}{k+2} + \left(\frac{2}{n+1} - 1\right)$$

note $x_1 = 0$, so.

so

$$\frac{x_n}{n(n+1)} = \sum_{K=3}^{n+1} \frac{1}{K} + \frac{2}{n+1} - 1$$

$$= \sum_{K=1}^{n} \frac{1}{K} + \frac{1}{n+1} - 1 - \frac{1}{2} + \frac{2}{n+1} - 1$$

$$\frac{x_n}{n(n+1)} = \sum_{K=1}^{n} \frac{1}{K} + \frac{3}{n+1} - \frac{5}{2}$$

Define: $H_n = \sum_{K=1}^{n} \frac{1}{K}$, called the n^{th} harmonic number.

$$\frac{x_n}{n(n+1)} = H_n + \frac{3}{n+1} - \frac{5}{2}$$

$$\therefore \quad x_n = n(n+1)H_n + 3n - \frac{5}{2}n(n+1)$$

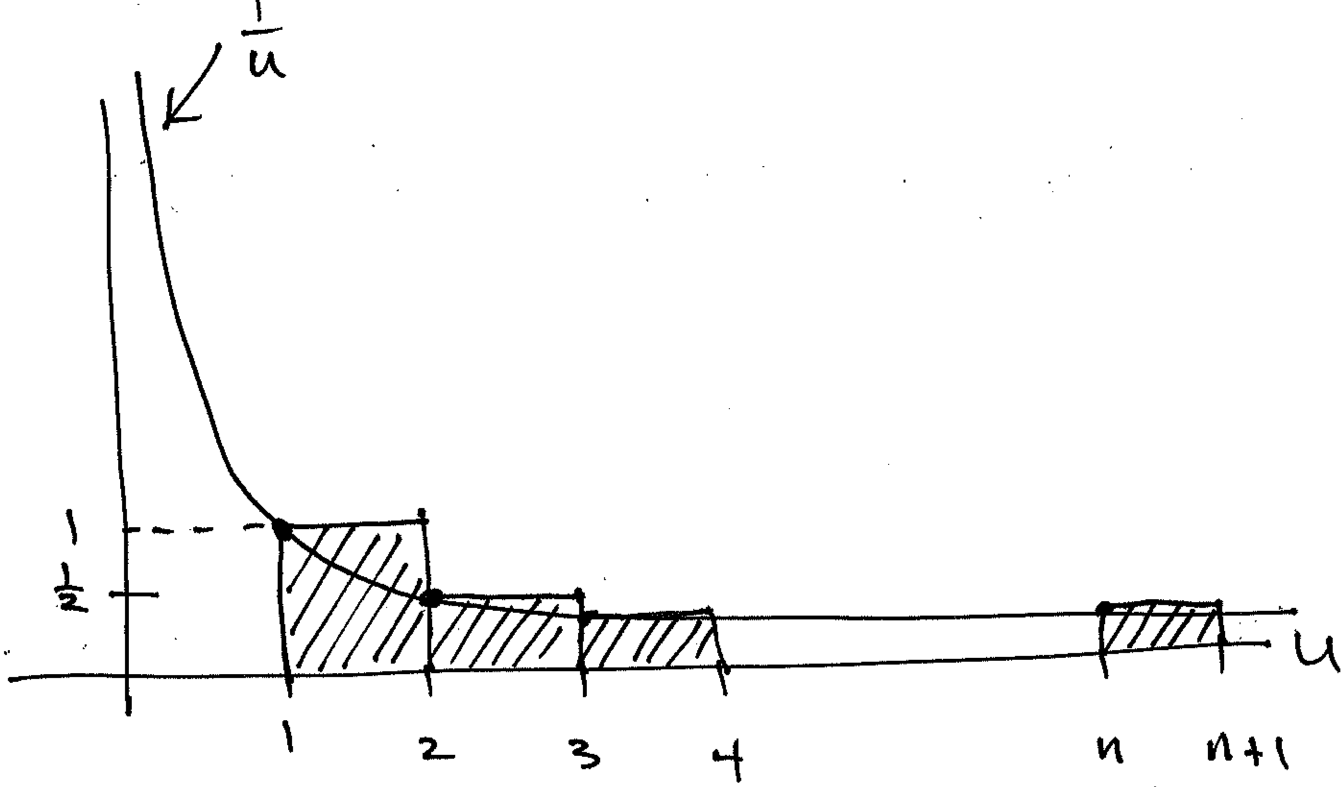
Hence

$$t(n) = (n-1) + \frac{2}{n}\left(n(n+1)H_n + 3n - \frac{5}{2}n(n+1)\right)$$

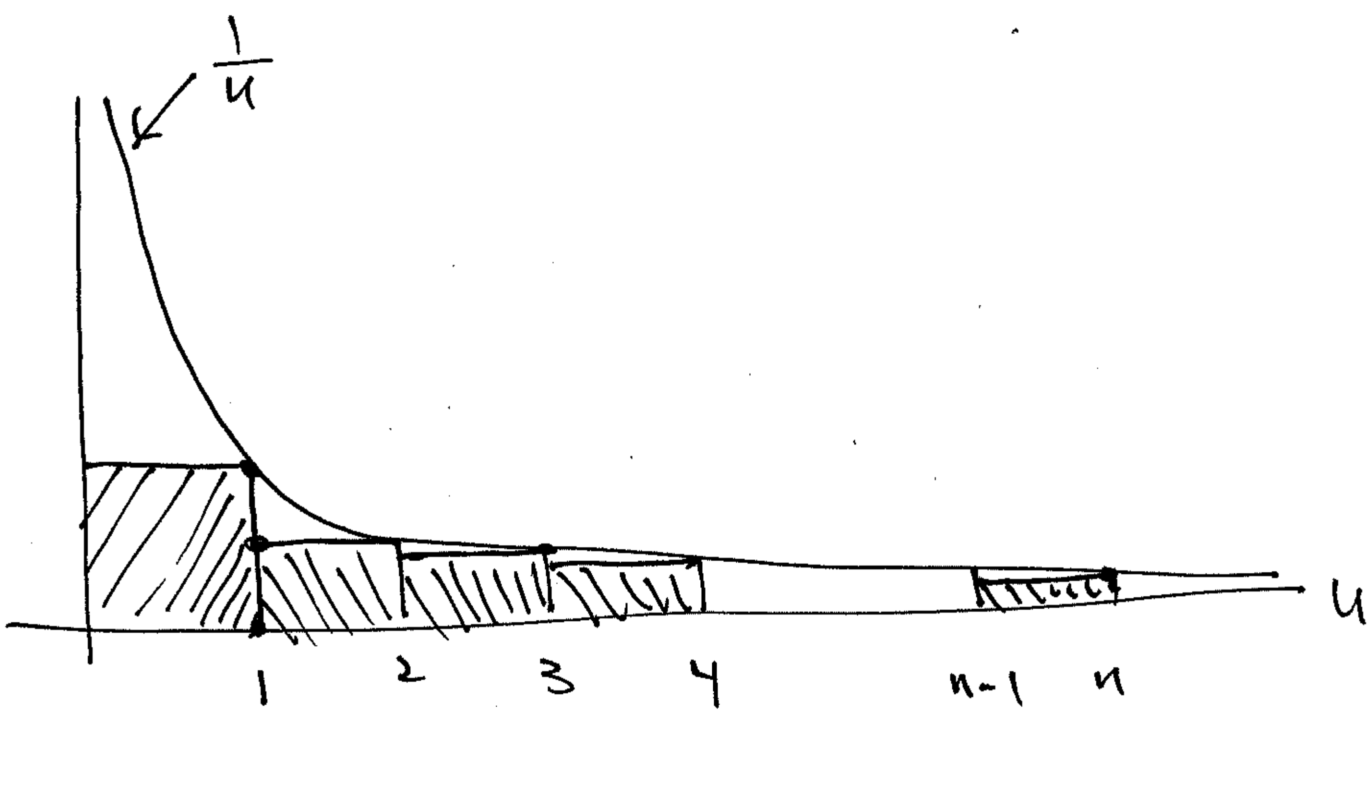
$$= (n-1) + 2(n+1)H_n + 6 \quad - 5(n+1)$$

$$\boxed{t(n) = 2(n+1)H_n - 4n}$$

must estimate growth rate of $H_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$



$$\int_1^{n+1} \frac{1}{u}\,du \le H_n \le 1 + \int_1^{n} \frac{1}{u}\,du$$



so

$$\ln(u)\Big|_1^{n+1} \le H_n \le 1 + \ln(u)\Big|_1^{n}$$

$$\underbrace{\ln(n+1)}_{\Omega(\ln n)} \le H_n \le \underbrace{1+\ln(n)}_{O(\ln n)}$$

$$\therefore \quad H_n = \Theta(\ln n)$$

Actually: $\lim_{n\to\infty}\left(\frac{H_n}{\ln(n)}\right) = 1$

Hence: $H_n \sim \ln(n)$

Therefore:

$$t(n) \sim 2n\ln(n) - 2\ln(n) - 4n$$

and

$$t(n) = \Theta(n \log n)$$

Randomize Quicksort:

<u>RandPartition(A, P, r)</u>

1. $i = \text{Rand}(P, r)$
2. $A[r] \leftrightarrow A[i]$ (swap)
3. return Partition(A, P, r)

RandQuicksort(A, p, r)

1. if $p < r$
2. $\quad q = \text{RandPartition}(A, p, r)$
3. $\quad \text{RandQuicksort}(A, p, q-1)$
4. $\quad \text{RandQuicksort}(A, q+1, r)$

Exercises

(a) show $\displaystyle\sum_{k=1}^{n} H_k = (n+1)H_n - n$

(b) show $\displaystyle\sum_{k=1}^{n} k \cdot H_k = \frac{1}{2}n(n+1)H_n - \frac{1}{4}n(n-1)$

use: $H_{n+1} = H_n + \dfrac{1}{n+1}$

(c) prove directly that

$$\boxed{t(n) = 2(n+1)H_n - 4n}$$

solves

$$\boxed{t(n) = (n-1) + \frac{2}{n}\sum_{k=1}^{n-1} t(k)}$$

Hint: use (a) & (b)

Chapter 9: Order statistics

Goal: find both max and min in $A[1 \ldots n]$

min(x, y)	**max(x, y)**
1. if $x < y$	1. if $x > y$
2. return x	2.
3. else	3.
4. return y	4.

$\# comp = 1$ $\qquad\qquad$ $\# comp = 1$

MinMax(A, p, r) Pre: $p \le r$

```
if p = r
    return (A[p], A[p])
else
    q = ⌊(p+r)/2⌋
    (m1, M1) = MinMax(A, p, q)
    (m2, M2) = MinMax(A, q+1, r)
    return (min(m1, m2), max(M1, M2))
```

let $T(n)$ = # comparisons by MinMax$(A, 1, n)$.

Then

$$T(n) = \begin{cases} 0 & n=1 \checkmark \\ T\left(\lceil \frac{n}{2} \rceil\right) + T\left(\lfloor \frac{n}{2} \rfloor\right) + 2 & n \geq 2 \checkmark \end{cases}$$

<u>master thm</u>: $T(n) = 2T\left(\frac{n}{2}\right) + 1$

compare: $1 = n^0$ to $n^{\log_2 2} = \underline{\underline{n^1}}$

case 1: $T(n) = \Theta(n)$

Actually $\boxed{T(n) = 2n - 2}$

check:

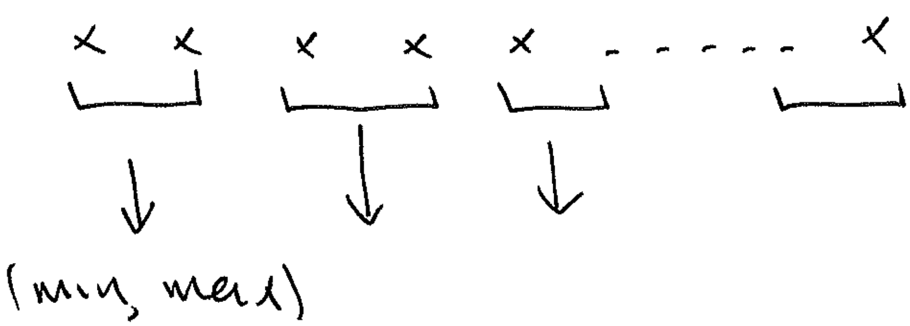
$RHS = T\left(\lceil \frac{n}{2} \rceil\right) + T\left(\lfloor \frac{n}{2} \rfloor\right) + 2$

$= \left(2\lceil \frac{n}{2} \rceil - 2\right) + \left(2\lfloor \frac{n}{2} \rfloor - \cancel{2}\right) + \cancel{2}$

$$= 2\left(\left\lceil \frac{n}{2} \right\rceil + \left\lfloor \frac{n}{2} \right\rfloor\right) - 2$$

$$= 2n - 2 \quad = \Theta(n)$$

<u>Better way</u>: sec. 9.1



<u>Exercise</u> write a recursive algorithm that solves using exactly: $\left\lceil \frac{3n}{2} \right\rceil - 2$ comparisons