EXCERCISE.

PROVE THE CORRECTNESS OF QUICKSORT BY INDUCTION ON THE LENGTH OF THE SUB-ARRAY $A[p..r]$: $n = r-p+1$.

THE RUN TIME OF QUICKSORT DEPENDS HEAVILY ON THE VALUE q RETURNED BY PARTITION. IF THE SUBARRAYS $A[p..(q-1)]$ AND $A[(q+1)..r]$ ARE NOT BALANCED (i.e. OF ROUGHLY EQUAL SIZE) THEN PERFORMANCE IS INHIBITED. IN THIS CASE OUR RECURSIVE CALL TO QUICKSORT IS ON A SUBARRAY WHICH IS INORDINATELY LONG.

THE WORST CASE OCCURS WHEN THE ARRAY IS ALREADY SORTED. THEN PARTITION RETURNS

$$\underbrace{A[p\ldots(r-1)]}_{\text{SORTED}} \le \underset{\uparrow\, q}{A[r]} \le \ldots \text{EMPTY} \ldots$$

LET $T(n)$ DENOTE THE WORST CASE RUN TIME OF QUICKSORT (i.e. WITH $A[1\ldots n]$ ALREADY SORTED.)

THEN $T(n)$ SATISFIES

$$T(n) = \begin{cases} \Theta(1) & n = 0, 1 \\ T(n-1) + \Theta(n) & n \geq 2 \end{cases}$$

SIMPLIFY THIS TO $T(n) = T(n-1) + cn$ FOR DEFINITENESS. BY THE ITERATION METHOD

$$\begin{aligned} T(n) &= cn + T(n-1) \\ &= cn + c(n-1) + T(n-2) \\ &= cn + c(n-1) + c(n-2) + T(n-3) \\ &= c\sum_{i=0}^{k-1}(n-i) + T(n-k) \\ &= cnk - \tfrac{1}{2}ck(k-1) + T(n-k). \end{aligned}$$

CHOOSE k SUCH THAT $n-k=1$, i.e. $k=n-1$. THEN

$$\begin{aligned} T(n) &= cn(n-1) - \tfrac{1}{2}c(n-1)(n-2) + \text{const.} \\ &= \Theta(n^2). \end{aligned}$$

SO WHAT'S SO QUICK ABOUT QUICKSORT?

THE ADVANTAGE OF QUICKSORT IS IN ITS AVERAGE CASE.

WE ASSUME THAT ALL $n!$ PERMUTATIONS OF THE INPUT ARRAY $A[1 \dots n]$ ARE EQUALLY LIKELY, I.E. THAT ANY GIVEN PERMUTATION OCCURS WITH PROBABILITY $\frac{1}{n!}$.

WE CHOOSE AS BASIC OPERATION (I.E. BAROMETER) THE COMPARISON OF NUMERICAL VALUES ON LINE 3 OF PARTITION. LET $t(n)$ DENOTE THE AVERAGE NUMBER OF COMPARISONS PERFORMED BY QUICKSORT ON AN INPUT ARRAY $A[1 \dots n]$ OF LENGTH n, I.E.

$$t(n) = \frac{\displaystyle\sum_{\text{All PERMUTATIONS}} (\text{\# OF COMPARISONS PERFORMED ON GIVEN PERMUTATION})}{n!}$$

WE WISH TO DETERMINE A RECURRENCE FOR $t(n)$. OUR ASSUMPTION IMPLIES THAT THE PIVOT $A[q]$ IS EQUALLY LIKELY TO BE PLACED IN ANY OF THE n LOCATIONS IN $A[1 \dots n]$.

THUS THE RETURN VALUE q OF PARTITION HAS PROBABILITY $\frac{1}{n}$ OF BEING ANY ONE OF THE n VALUES: $q = 1, 2, \dots, n$.

OBSERVE THAT PARTITION ITSELF DOES $(n-1)$ COMPARISONS ON $A[1\ldots n]$. ALSO

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

AND

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

THUS

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

SO

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

THE INITIAL VALUES $t(0)=0$, $t(1)=0$, $t(2)=1$ CAN BE SEEN BY INSPECTION. THUS

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

WHENCE

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

TO SOLVE THIS RECURRENCE WE RESORT TO SOME TRICKS. LET

$$x_n = \sum_{q=1}^{n-1} t(q) \quad , \quad x_1 = 0 .$$

THEN

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

AND SO

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

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

MULTIPLY BY THE MAGIC NUMBER $\frac{1}{(n+1)(n+2)}$

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

REPLACE n BY r:

$$\frac{x_{r+1}}{(r+1)(r+2)} - \frac{x_r}{r(r+1)} = \frac{3}{r+2} - \frac{2}{r+1}$$

SUM FOR $r=1$ TO $n-1$:

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

$$\therefore \sum_{r=2}^{n}\frac{x_r}{r(r+1)}-\sum_{r=1}^{n-1}\frac{x_r}{r(r+1)}=\sum_{r=3}^{n+1}\frac{1}{r}+\sum_{r=3}^{n+1}\frac{2}{r}-\sum_{r=2}^{n}\frac{2}{r}$$

$$\therefore \frac{x_n}{n(n+1)}-\frac{x_1}{2}=\sum_{r=3}^{n+1}\frac{1}{r}+\frac{2}{n+1}-1$$

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

$$=\sum_{r=1}^{n}\frac{1}{r}+\frac{3}{n+1}-\frac{5}{2}$$

RECALL $x_1=0$. DEFINE $H_n=\sum_{r=1}^{n}\frac{1}{r}$ (CALLED THE n^{TH} HARMONIC NUMBER.) THEN

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

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

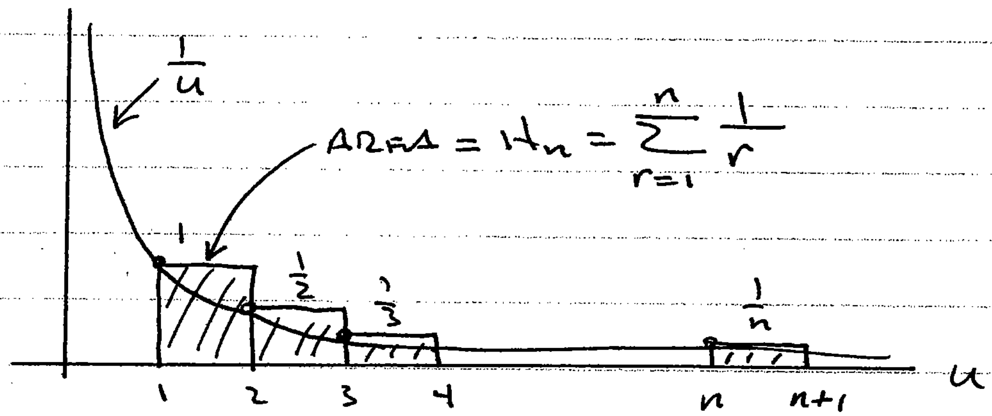
$$\therefore t(n)=(n-1)+\frac{2}{n}x_n$$

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

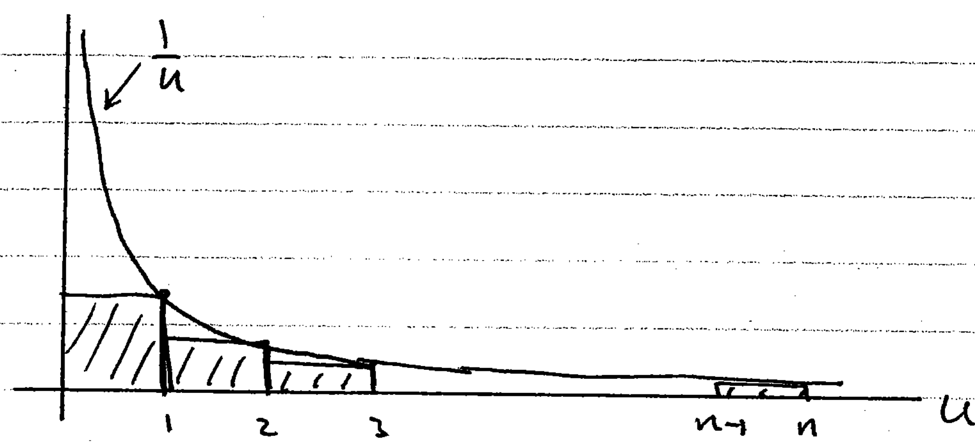
AND SO

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

WE MUST ESTIMATE THE SIZE OF H_n TO DETERMINE THE ASYMPTOTIC ORDER OF $t(n)$



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



THUS

$$\Omega(\ln(n)) = \ln(n+1) \le H_n \le 1+\ln(n) = O(\ln(n))$$

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

THUS

$$t(n) = \Theta(n \lg(n)),$$

WHICH IS BETTER THAN THE WORST CASE $\Theta(n^2)$ RUN TIME.

OUR ASSUMPTION THAT ALL PERMUTATIONS OF A[1...n] ARE EQUALLY LIKELY MAY NOT BE WELL FOUNDED IN PRACTICE. WE MAY WISH TO SORT A PRE-SORTED ARRAY, OR ONE THAT IS NEARLY SORTED MORE OFTEN THAN NOT.

THERE ARE SEVERAL WAYS TO RANDOMIZE QUICKSORT TO COMPENSATE FOR THIS.

ONE WAY IS TO SIMPLY APPLY A RANDOMLY CHOSEN PERMUTATION TO A[1...n] BEFORE CALLING QUICKSORT.