

FORMAL ANALYSIS OF QUICKSORT: EXPECTED TIME COMPLEXITY

define X_k : $\begin{cases} 1 & \text{if a pivot partitions array into } \underbrace{k}_{\text{left}} \text{ \& } \underbrace{n-k-1}_{\text{right}} \\ 0 & \text{otherwise} \end{cases}$
for $k=0 \dots n-1$

$$E[X_k] = 0 \cdot P(X_k=0) + 1 \cdot \underbrace{P(X_k=1)}_{\text{probability of generating a "k-split"}} \leftarrow \text{by definition}$$

$$= \text{probability of generating a "k-split"} = \frac{1}{n} \left. \vphantom{\frac{1}{n}} \right\} \begin{array}{l} \text{all pivots/splits} \\ \text{are equally likely} \end{array}$$

(= Prob. picking the k^{th} smallest)

$$T(n) = \Theta(n) + \left\{ \begin{array}{c} T(0) + T(n-1) \\ T(1) + T(n-2) \\ T(2) + T(n-3) \\ \vdots \\ T(n-1) + T(0) \end{array} \right\} \text{ n cases} = \Theta(n) + \sum_{k=0}^{n-1} X_k \cdot (T(k) + T(n-k-1))$$

$$\begin{aligned} E[T(n)] &= E[\underbrace{\sum}_{\text{linearity of expectation}}] = \sum E[(\cdot)] = E[\Theta(n)] + \sum_{k=0}^{n-1} E\left[X_k \cdot (T(k) + T(n-k-1))\right] = \\ &\quad \downarrow \text{by independence of random variables} \\ &= \Theta(n) + \sum E[X_k] \cdot E[T(k) + T(n-k-1)] \end{aligned}$$

There are random variables (X_j) in recursive calls, but independent of X_k

$$\begin{aligned}
& \Theta(n) + \sum E[x_k] \cdot E[T(k) + T(n-k-1)] \\
&= \Theta(n) + \sum_{k=0}^{n-1} \frac{1}{n} \cdot E[T(k) + T(n-k-1)] \\
&= \Theta(n) + \underbrace{\frac{1}{n} \cdot \sum E[T(k)] + \frac{1}{n} \sum E[T(n-k-1)]}_{\text{symmetric: equal}} \\
&= \Theta(n) + \frac{2}{n} \sum_{k=0}^{n-1} E[T(k)] \\
&= \Theta(n) + \frac{2}{n} \left(\underbrace{E[T(0)] + E[T(1)]}_{\text{to make math easier later}} + \sum_{k=2}^{n-1} E[T(k)] \right)
\end{aligned}$$

$\Theta(n)$

guess $E[T(n)] \leq a \cdot n \log n$
 assume true for $k < n$

$$\begin{aligned}
E[T(n)] &= \Theta(n) + \frac{2}{n} \sum_{k=2}^{n-1} E[T(k)] \\
&\leq \Theta(n) + \frac{2}{n} \sum_{k=2}^{n-1} a \cdot k \cdot \log k
\end{aligned}$$

Use: $\sum_{k=2}^{n-1} k \log k \leq \frac{1}{2} n^2 \log n - \frac{1}{8} n^2$ [see CLRS]

$$\leq \Theta(n) + \frac{2a}{n} \cdot \left(\frac{1}{2} n^2 \log n - \frac{1}{8} n^2 \right)$$

$$= a n \log n - \frac{a n}{4} + \Theta(n)$$

$$= \underline{a n \log n} - \left(\frac{a n}{4} - c \cdot n \right)$$

QED positive if $a > 4c$

In practice, quicksort is 3x faster than Mergesort by using a few simple tricks and handling base cases