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 $k = 0 \ldots n-1$

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 $k = 0 \ldots n-1$

$E[X_k] = 0 \cdot P(X_k = 0) + 1 \cdot P(X_k = 1)$ ← by definition

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 $k = 0 \dots n-1$

$$E[X_k] = 0 \cdot P(X_k = 0) + 1 \cdot \underbrace{P(X_k = 1)}$$ ← by definition

= probability of generating a "k-split"

Formal Analysis of Quicksort: Expected Time Complexity

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

for $k = 0 \dots n-1$

$E[X_k] = 0 \cdot P(X_k = 0) + 1 \cdot \underbrace{P(X_k = 1)}$ ← by definition

= probability of generating a "k-split" $= \frac{1}{n}$ } all pivots/splits are equally likely

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

$$T(n) = \Theta(n) + \underbrace{\left\{\begin{array}{l} T(0) + T(n-1) \\ T(1) + T(n-2) \\ T(2) + T(n-3) \\ \quad\vdots \\ T(n-1) + T(0) \end{array}\right\}}_{n \text{ cases}}$$

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

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

$$E[T(n)] = E\left[\sum\right]$$

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

$$E[T(n)] = \underbrace{E\left[\sum\right] = \sum E[(\cdot)]}_{\text{linearity of expectation}}$$

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

$$E[T(n)] = \underbrace{E\left[\sum\right] = \sum E[(\cdot)]}_{\text{linearity of expectation}} = E\left[\Theta(n)\right] + \sum_{k=0}^{n-1} E\left[X_k \cdot \left(T(k) + T(n-k-1) \right)\right] =$$

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

$$E[T(n)] = \underbrace{E\Big[\sum\Big] = \sum E[(\cdot)]}_{\text{linearity of expectation}} = E[\Theta(n)] + \sum_{k=0}^{n-1} E\Big[X_k \cdot \big(T(k) + T(n-k-1)\big)\Big] =$$

$$= \Theta(n) + \ldots$$

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

$$E[T(n)] = \underbrace{E\left[\sum\right] = \sum E[(\cdot)]}_{\text{linearity of expectation}} = E[\Theta(n)] + \sum_{k=0}^{n-1} E\Big[X_k \cdot \big(T(k) + T(n-k-1)\big)\Big] =$$

by independence of random variables

$$= \Theta(n) + \sum E[X_k] \cdot E\big[T(k) + T(n-k-1)\big]$$

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

$$\Theta(n) + \sum E[X_k] \cdot E\left[T(k) + T(n-k-1)\right]$$

$$\Theta(n) + \sum E[X_k] \cdot E\left[T(k) + T(n-k-1)\right]$$

$$= \Theta(n) + \sum_{k=0}^{n-1} \frac{1}{n} \cdot E\left[T(k) + T(n-k-1)\right]$$

$$\Theta(n) + \sum E[x_k] \cdot E\left[T(k) + T(n-k-1)\right]$$

$$= \Theta(n) + \sum_{k=0}^{n-1} \frac{1}{n} \cdot E\left[T(k) + T(n-k-1)\right]$$

$$= \Theta(n) + \underbrace{\frac{1}{n} \cdot \sum E\left[T(k)\right]} + \underbrace{\frac{1}{n} \sum E\left[T(n-k-1)\right]}$$

symmetric : equal

$$\Theta(n) + \sum E[X_k] \cdot E\left[T(k) + T(n-k-1)\right]$$

$$= \Theta(n) + \sum_{k=0}^{n-1} \frac{1}{n} \cdot E\left[T(k) + T(n-k-1)\right]$$

$$= \Theta(n) + \underbrace{\frac{1}{n} \cdot \sum E[T(k)]} + \underbrace{\frac{1}{n} \sum E[T(n-k-1)]}$$

symmetric : equal

$$= \Theta(n) + \frac{2}{n} \sum_{k=0}^{n-1} E[T(k)]$$

$$\Theta(n) + \sum E[X_k] \cdot E\left[T(k) + T(n-k-1)\right]$$

$$= \Theta(n) + \sum_{k=0}^{n-1} \frac{1}{n} \cdot E\left[T(k) + T(n-k-1)\right]$$

$$= \Theta(n) + \underbrace{\frac{1}{n} \cdot \sum E\left[T(k)\right]} + \underbrace{\frac{1}{n} \sum E\left[T(n-k-1)\right]}$$

symmetric : equal

$$= \Theta(n) + \frac{2}{n} \sum_{k=0}^{n-1} E\left[T(k)\right]$$

$$= \Theta(n) + \frac{2}{n}\left(\underbrace{E\left[T(0)\right] + E\left[T(1)\right]}\right) + \frac{2}{n} \sum_{k=2}^{n-1} E\left[T(k)\right]$$

to make math easier later

$$\Theta(n) + \sum E[X_k] \cdot E\left[T(k) + T(n-k-1)\right]$$

$$= \Theta(n) + \sum_{k=0}^{n-1} \frac{1}{n} \cdot E\left[T(k) + T(n-k-1)\right]$$

$$= \Theta(n) + \underbrace{\frac{1}{n} \cdot \sum E\left[T(k)\right] + \frac{1}{n} \sum E\left[T(n-k-1)\right]}_{\text{symmetric : equal}}$$

$$= \Theta(n) + \frac{2}{n} \sum_{k=0}^{n-1} E\left[T(k)\right]$$

$$= \underbrace{\Theta(n) + \frac{2}{n}\Big(\underbrace{E\left[T(0)\right] + E\left[T(1)\right]}_{\text{to make math easier later}}\Big)}_{\Theta(n)} + \frac{2}{n} \sum_{k=2}^{n-1} E\left[T(k)\right]$$

$$\Theta(n) + \sum E[X_k] \cdot E[T(k) + T(n-k-1)]$$

$$= \Theta(n) + \sum_{k=0}^{n-1} \frac{1}{n} \cdot E[T(k) + T(n-k-1)]$$

$$= \Theta(n) + \underbrace{\frac{1}{n} \cdot \sum E[T(k)]} + \underbrace{\frac{1}{n} \sum E[T(n-k-1)]}$$

symmetric: equal

$$= \Theta(n) + \frac{2}{n} \sum_{k=0}^{n-1} E[T(k)]$$

$$= \Theta(n) + \underbrace{\frac{2}{n}\Big(\underbrace{E[T(0)] + E[T(1)]}_{\text{to make math easier later}}\Big)}_{\Theta(n)} + \frac{2}{n} \sum_{k=2}^{n-1} E[T(k)]$$

guess $E[T(n)] \leq a \cdot n \log n$

assume true for $k < n$

$$\Theta(n) + \sum E[X_k] \cdot E\left[T(k) + T(n-k-1)\right]$$

$$= \Theta(n) + \sum_{k=0}^{n-1} \frac{1}{n} \cdot E\left[T(k) + T(n-k-1)\right]$$

$$= \Theta(n) + \underbrace{\frac{1}{n} \cdot \sum E[T(k)]}_{} + \underbrace{\frac{1}{n} \sum E[T(n-k-1)]}_{}$$

symmetric : equal

$$= \Theta(n) + \frac{2}{n} \sum_{k=0}^{n-1} E[T(k)]$$

$$= \underbrace{\Theta(n) + \frac{2}{n}\Big(\underbrace{E[T(0)] + E[T(1)]}_{\text{to make math easier later}}\Big)}_{\Theta(n)} + \frac{2}{n} \sum_{k=2}^{n-1} E[T(k)]$$

guess $E[T(n)] \leq a \cdot n \log n$

assume true for $k < n$

$$E[T(n)] = \underbrace{\Theta(n)}_{} + \frac{2}{n} \sum_{k=2}^{n-1} E[T(k)]$$

$$\Theta(n) + \sum E[X_k] \cdot E\left[T(k) + T(n-k-1)\right]$$

$$= \Theta(n) + \sum_{k=0}^{n-1} \frac{1}{n} \cdot E\left[T(k) + T(n-k-1)\right]$$

$$= \Theta(n) + \underbrace{\frac{1}{n} \cdot \sum E[T(k)]}_{} + \underbrace{\frac{1}{n} \sum E[T(n-k-1)]}_{}$$

symmetric : equal

$$= \Theta(n) + \frac{2}{n} \sum_{k=0}^{n-1} E[T(k)]$$

$$= \underbrace{\Theta(n) + \frac{2}{n}\underbrace{\left(E[T(0)] + E[T(1)]\right)}_{\text{to make math easier later}}}_{\Theta(n)} + \frac{2}{n} \sum_{k=2}^{n-1} E[T(k)]$$

guess $E[T(n)] \leq a \cdot n \log n$

assume true for $k < n$

$$E[T(n)] = \Theta(n) + \frac{2}{n} \sum_{k=2}^{n-1} E[T(k)]$$

$$\leq \Theta(n) + \frac{2}{n} \sum_{k=2}^{n-1} a \cdot k \cdot \log k$$

$$\Theta(n) + \sum E[X_k] \cdot E\left[T(k) + T(n-k-1)\right]$$

$$= \Theta(n) + \sum_{k=0}^{n-1} \frac{1}{n} \cdot E\left[T(k) + T(n-k-1)\right]$$

$$= \Theta(n) + \underbrace{\frac{1}{n} \cdot \sum E[T(k)]} + \underbrace{\frac{1}{n} \sum E[T(n-k-1)]}$$

symmetric: equal

$$= \Theta(n) + \frac{2}{n} \sum_{k=0}^{n-1} E[T(k)]$$

$$= \underbrace{\Theta(n) + \frac{2}{n}\Big(\underbrace{E[T(0)] + E[T(1)]}_{\text{to make math easier later}}\Big)}_{\Theta(n)} + \frac{2}{n} \sum_{k=2}^{n-1} E[T(k)]$$

guess $E[T(n)] \le a \cdot n \log n$

assume true for $k < n$

$$E[T(n)] = \Theta(n) + \frac{2}{n} \sum_{k=2}^{n-1} E[T(k)]$$

$$\le \Theta(n) + \frac{2}{n} \sum_{k=2}^{n-1} a \cdot k \cdot \log k$$

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

$$\Theta(n) + \sum E[x_k] \cdot E[T(k) + T(n-k-1)]$$

$$= \Theta(n) + \sum_{k=0}^{n-1} \frac{1}{n} \cdot E[T(k) + T(n-k-1)]$$

$$= \Theta(n) + \underbrace{\frac{1}{n} \cdot \sum E[T(k)] + \frac{1}{n} \sum E[T(n-k-1)]}_{\text{symmetric: equal}}$$

$$= \Theta(n) + \frac{2}{n} \sum_{k=0}^{n-1} E[T(k)]$$

$$= \underbrace{\Theta(n) + \frac{2}{n}\Big(\underbrace{E[T(0)] + E[T(1)]}_{\text{to make math easier later}}\Big)}_{\Theta(n)} + \frac{2}{n} \sum_{k=2}^{n-1} E[T(k)]$$

guess $E[T(n)] \leq a \cdot n \log n$

assume true for $k < n$

$$E[T(n)] = \Theta(n) + \frac{2}{n} \sum_{k=2}^{n-1} E[T(k)]$$

$$\leq \Theta(n) + \frac{2}{n} \sum_{k=2}^{n-1} a \cdot k \cdot \log k$$

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

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

$$\Theta(n) + \sum E[X_k] \cdot E\left[T(k) + T(n-k-1)\right]$$

$$= \Theta(n) + \sum_{k=0}^{n-1} \frac{1}{n} \cdot E\left[T(k) + T(n-k-1)\right]$$

$$= \Theta(n) + \underbrace{\frac{1}{n} \cdot \sum E\left[T(k)\right]} + \underbrace{\frac{1}{n} \sum E\left[T(n-k-1)\right]}$$

symmetric : equal

$$= \Theta(n) + \frac{2}{n} \sum_{k=0}^{n-1} E\left[T(k)\right]$$

$$= \underbrace{\Theta(n) + \frac{2}{n}\Big(\underbrace{E\left[T(0)\right] + E\left[T(1)\right]}_{\text{to make math easier later}}\Big)}_{\Theta(n)} + \frac{2}{n} \sum_{k=2}^{n-1} E\left[T(k)\right]$$

guess $E[T(n)] \leq a \cdot n \log n$

assume true for $k < n$

$$E[T(n)] = \Theta(n) + \frac{2}{n} \sum_{k=2}^{n-1} E\left[T(k)\right]$$

$$\leq \Theta(n) + \frac{2}{n} \sum_{k=2}^{n-1} a \cdot k \cdot \log k$$

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

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

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

$$\Theta(n) + \sum E[X_k] \cdot E\left[T(k) + T(n-k-1)\right]$$

$$= \Theta(n) + \sum_{k=0}^{n-1} \frac{1}{n} \cdot E\left[T(k) + T(n-k-1)\right]$$

$$= \Theta(n) + \underbrace{\frac{1}{n} \cdot \sum E[T(k)]}_{} + \underbrace{\frac{1}{n} \sum E[T(n-k-1)]}_{}$$

symmetric : equal

$$= \Theta(n) + \frac{2}{n} \sum_{k=0}^{n-1} E[T(k)]$$

$$= \underbrace{\Theta(n) + \frac{2}{n}\left(\underbrace{E[T(0)] + E[T(1)]}_{\text{to make math easier later}}\right)}_{\Theta(n)} + \frac{2}{n} \sum_{k=2}^{n-1} E[T(k)]$$

guess $E[T(n)] \leq a \cdot n \log n$

assume true for $k < n$

$$E[T(n)] = \Theta(n) + \frac{2}{n} \sum_{k=2}^{n-1} E[T(k)]$$

$$\leq \Theta(n) + \frac{2}{n} \sum_{k=2}^{n-1} a \cdot k \cdot \log k$$

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

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

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

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

$$\Theta(n) + \sum E[X_k] \cdot E\left[T(k) + T(n-k-1)\right]$$

$$= \Theta(n) + \sum_{k=0}^{n-1} \frac{1}{n} \cdot E\left[T(k) + T(n-k-1)\right]$$

$$= \Theta(n) + \underbrace{\frac{1}{n} \cdot \sum E\left[T(k)\right] + \frac{1}{n} \sum E\left[T(n-k-1)\right]}_{\text{symmetric : equal}}$$

$$= \Theta(n) + \frac{2}{n} \sum_{k=0}^{n-1} E\left[T(k)\right]$$

$$= \underbrace{\Theta(n) + \frac{2}{n}\underbrace{\left(E\left[T(0)\right] + E\left[T(1)\right]\right)}_{\text{to make math easier later}}}_{\Theta(n)} + \frac{2}{n} \sum_{k=2}^{n-1} E\left[T(k)\right]$$

guess $E\left[T(n)\right] \leq a \cdot n \log n$

assume true for $k < n$

$$E\left[T(n)\right] = \Theta(n) + \frac{2}{n} \sum_{k=2}^{n-1} E\left[T(k)\right]$$

$$\leq \Theta(n) + \frac{2}{n} \sum_{k=2}^{n-1} a \cdot k \cdot \log k$$

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

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

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

$$= a n \log n - \underbrace{\left(\frac{an}{4} - c \cdot n\right)}_{\text{positive if } a > 4c}$$

$$\Theta(n) + \sum E[X_k] \cdot E[T(k) + T(n-k-1)]$$

$$= \Theta(n) + \sum_{k=0}^{n-1} \frac{1}{n} \cdot E[T(k) + T(n-k-1)]$$

$$= \Theta(n) + \underbrace{\frac{1}{n} \cdot \sum E[T(k)]}_{} + \underbrace{\frac{1}{n} \sum E[T(n-k-1)]}_{}$$

symmetric: equal

$$= \Theta(n) + \frac{2}{n} \sum_{k=0}^{n-1} E[T(k)]$$

$$= \underbrace{\Theta(n) + \frac{2}{n}\Big(\underbrace{E[T(0)] + E[T(1)]}_{\text{to make math easier later}}\Big)}_{\Theta(n)} + \frac{2}{n} \sum_{k=2}^{n-1} E[T(k)]$$

guess $E[T(n)] \le a \cdot n\log n$

assume true for $k < n$

$$E[T(n)] = \Theta(n) + \frac{2}{n} \sum_{k=2}^{n-1} E[T(k)]$$

$$\le \Theta(n) + \frac{2}{n} \sum_{k=2}^{n-1} a \cdot k \cdot \log k$$

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

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

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

$$= \underbrace{an\log n}_{\text{QED}} - \Big(\underbrace{\frac{an}{4} - c \cdot n}_{\text{positive if } a > 4c}\Big)$$

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