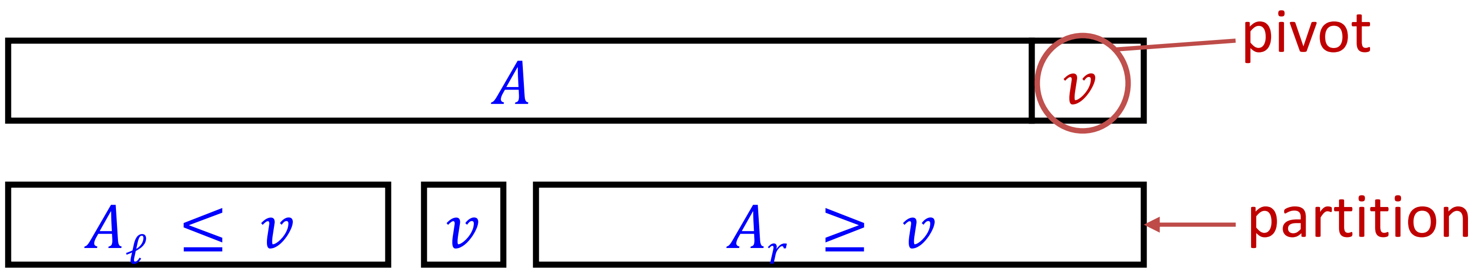
Algorithm Theory

Chapter 7
Randomized Algorithms

Part III:
Randomized Quicksort : Expected Time

Fabian Kuhn

Randomized Quicksort



function Quick (A: array): array

{returns the sorted array A}

begin

if $\text{size}(A) \leq 1$ then **return** A

else { choose pivot element v in A;

partition A into A_ℓ with elements $\geq v$,

and A_r with elements $\geq v$

return Quick(A_ℓ) | v | Quick(A_r)

end;

Randomized Quicksort:

pick pivot uniformly at random

Randomized Quicksort Analysis

Randomized Quicksort: pick uniform random element as pivot

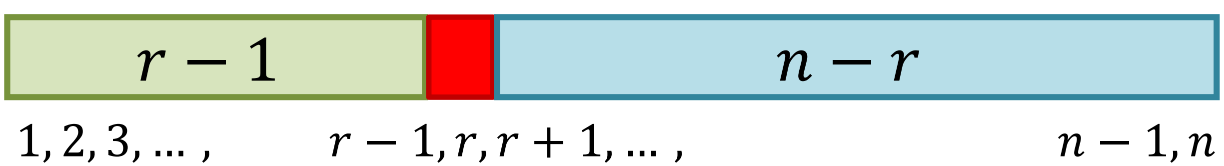
Running Time of sorting n elements:

- Let's just count the number of comparisons
- In the partitioning step, all $n-1$ non-pivot elements have to be compared to the pivot
- **Number of comparisons:**

depends on choice of pivot

$$\boldsymbol{n-1 \ + \ \#\text{comparisons in recursive calls}}$$

- **If rank of pivot is r:**
  recursive calls with $\boldsymbol{r-1}$ and $\boldsymbol{n-r}$ elements



Law of Total Expectation

- Given a random variable X and
- a set of events $A_1, \ldots, A_k$ that partition Ω
  - E.g., for a second random variable Y, we could have
    $$A_i \coloneqq \{\omega \in \Omega : Y(\omega) = i\}$$

Law of Total Expectation

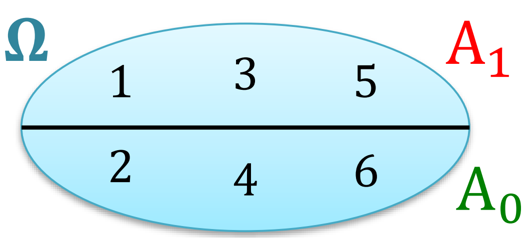
$$\mathbb{E}[X] = \sum_{i=1}^{k} \mathbb{P}(A_i) \cdot \mathbb{E}[X \mid A_i] = \sum_{y} \mathbb{P}(Y = y) \cdot \mathbb{E}[X \mid Y = y]$$

Example:

- X: outcome of rolling a die

Clearly: $\mathbb{E}[X] = \frac{1+2+3+4+5+6}{6} = 3.5$

- $A_0 = \{X \text{ is even}\}, \quad A_1 = \{X \text{ is odd}\}$



$$\mathbb{E}[X] = \underbrace{\underbrace{\mathbb{E}[X|A_0]}_{=4} \cdot \underbrace{\mathbb{P}(A_0)}_{=1/2} + \underbrace{\mathbb{E}[X|A_1]}_{=3} \cdot \underbrace{\mathbb{P}(A_1)}_{=1/2}}_{=3.5}$$

Randomized Quicksort Analysis

Random variables:

- C: total number of comparisons (for a given array of length n)
- R: rank of first pivot
- C_ℓ, C_r: number of comparisons for the 2 recursive calls

$$\mathbb{E}[C] = \mathbb{E}[n - 1 + C_\ell + C_r] = n - 1 + \mathbb{E}[C_\ell] + \mathbb{E}[C_r]$$

Linearity of Expectation:

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

Law of Total Expectation:

$$\mathbb{E}[C] = \sum_{r=1}^{n} \mathbb{P}(R = r) \cdot \mathbb{E}[C | R = r]$$

$$= \sum_{r=1}^{n} \mathbb{P}(R = r) \cdot (n - 1 + \mathbb{E}[C_\ell | R = r] + \mathbb{E}[C_r | R = r])$$

Exp. #comp. to sort array of length n

Exp. #comp. to sort array of length $r - 1$

Exp. #comp. to sort array of length $n - r$

Randomized Quicksort Analysis

We have seen that:

$$\mathbb{E}[C] = \sum_{r=1}^{n} \mathbb{P}(R = r) \cdot (n - 1 + \mathbb{E}[C_\ell | R = r] + \mathbb{E}[C_r | R = r])$$

$= {}^1/_n$

Define:

- $\boldsymbol{T(n)}$: expected number of comparisons when sorting n elements

$$\mathbb{E}[C] = T(n)$$
$$\mathbb{E}[C_\ell | R = r] = T(r-1)$$
$$\mathbb{E}[C_r | R = r] = T(n-r)$$

Recursion:

$$\boldsymbol{T(n) = \sum_{r=1}^{n} \frac{1}{n} \cdot \big(n - 1 + T(r-1) + T(n-r)\big)}$$

$$\boldsymbol{T(0) = T(1) = 0}$$

Randomized Quicksort Analysis

Theorem: The expected number of comparisons when sorting n elements using randomized quicksort is $T(n) \leq 2n \ln n$.

Proof: *(by induction on n)*

$$T(n) = \sum_{r=1}^{n} \frac{1}{n} \cdot \big(n - 1 + T(r-1) + T(n-r)\big), \qquad T(0) = T(1) = 0$$

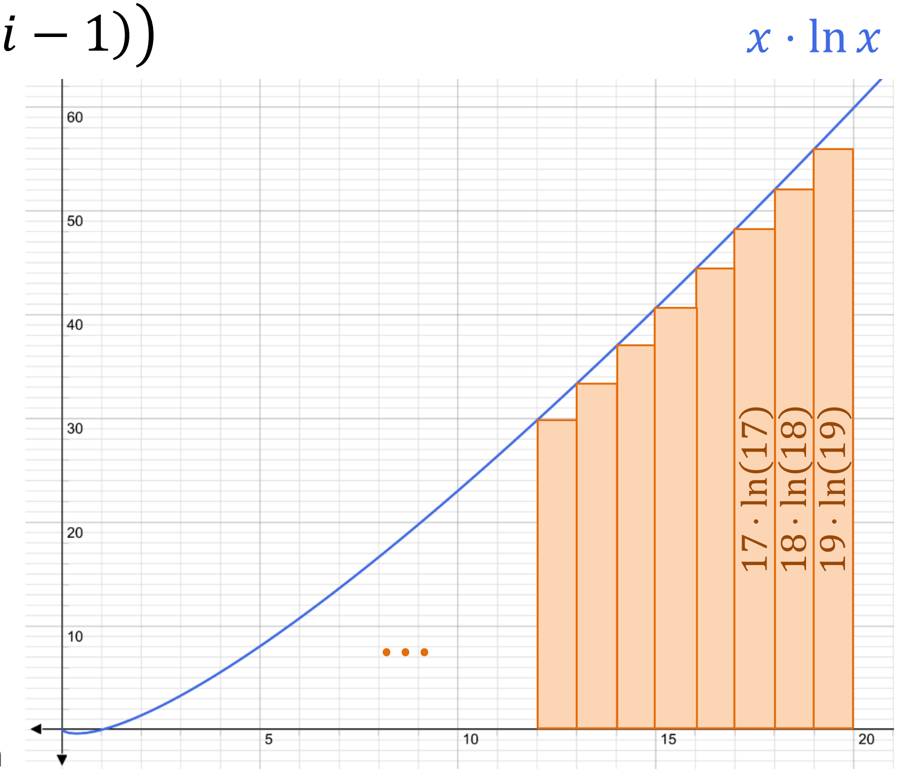
$$= n - 1 + \frac{1}{n} \cdot \sum_{i=0}^{n-1} \big(T(i) + T(n-i-1)\big)$$

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

induction hypothesis

$$\leq n - 1 + \frac{4}{n} \cdot \sum_{i=1}^{n-1} i \cdot \ln i$$

$$< n - 1 + \frac{4}{n} \cdot \int_1^n x \ln x \, dx$$



Randomized Quicksort Analysis

Theorem: The expected number of comparisons when sorting n elements using randomized quicksort is $T(n) \leq 2n \ln n$.

Proof:

$$T(n) \leq n - 1 + \frac{4}{n} \cdot \int_1^n x \ln x \, dx$$

$$\int x \ln x \, dx = \frac{x^2 \ln x}{2} - \frac{x^2}{4}$$

$$T(n) \leq n - 1 + \frac{4}{n} \cdot \left[\frac{n^2 \ln n}{2} - \frac{n^2}{4} + \frac{1}{4}\right]$$

$$= n - 1 + 2n \ln n - n + 1$$

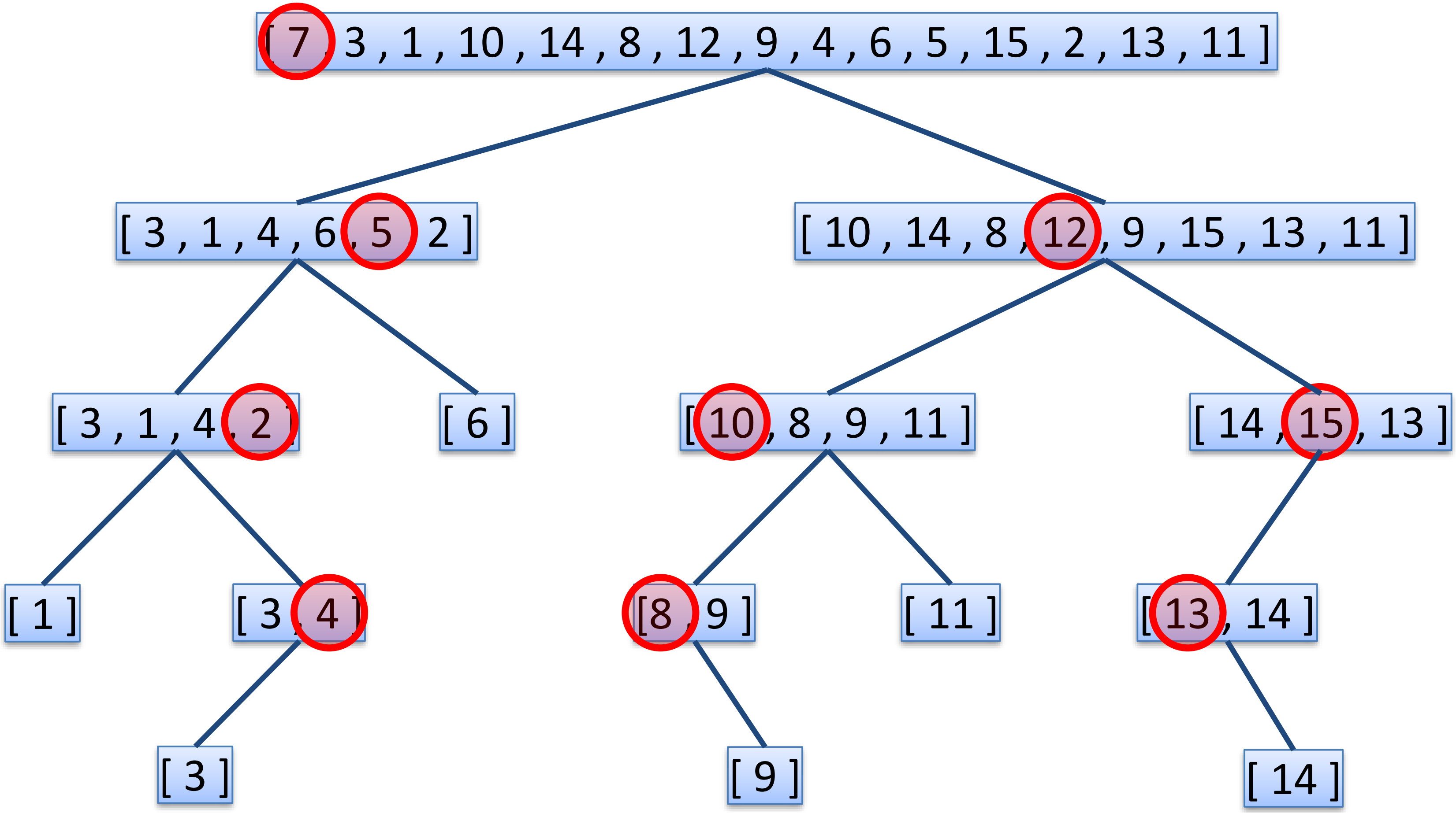
$$= 2n \ln n + \left(\frac{1}{n} - 1\right)$$

$$< 2n \ln n$$

Alternative Randomized Quicksort Analysis

Array to sort: [7 , 3 , 1 , 10 , 14 , 8 , 12 , 9 , 4 , 6 , 5 , 15 , 2 , 13 , 11]

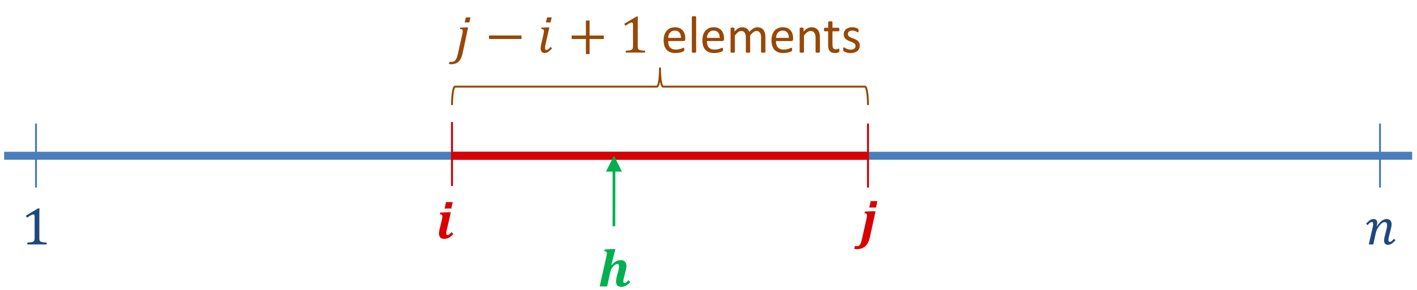
Viewing quicksort run as a tree:



Comparisons

- Comparisons are only between pivot and non-pivot elements
- Every element can only be the pivot once:
  → every 2 elements can only be compared once!

- W.l.o.g., assume that the elements to sort are $1, 2, \dots, n$
- Elements i and j are compared if and only if either i or j is a pivot before any element $h: i < h < j$ is chosen as pivot
  - i.e., iff i is an ancestor of j or j is an ancestor of i



$$\mathbb{P}(\text{comparison between } i \text{ and } j) = \frac{2}{j - i + 1}$$

Counting Comparisons

Random variable for every pair of elements (i, j), $i < j$:

$$\boldsymbol{X_{ij}} = \begin{cases} 1, & \text{if there is a comparison between } i \text{ and } j \\ 0, & \text{otherwise} \end{cases}$$

$$\mathbb{P}(X_{ij} = 1) = \frac{2}{j - i + 1}, \qquad \mathbb{E}[X_{ij}] = \frac{2}{j - i + 1}$$

Number of comparisons: $\boldsymbol{X}$

$$X = \sum_{i<j} X_{ij}$$

- What is $\mathbb{E}[X]$?

Randomized Quicksort Analysis

Theorem: The expected number of comparisons when sorting n elements using randomized quicksort is $T(n) \leq 2n \ln n$.

Proof:

- **Linearity of expectation:**
  For all random variables $X_1, \dots, X_n$ and all $a_1, \dots, a_n \in \mathbb{R}$,

$$\mathbb{E}\left[\sum_i^n a_i X_i\right] = \sum_i^n a_i \mathbb{E}[X_i].$$

$$X = \sum_{i<j} X_{ij} \quad \Longrightarrow \quad \mathbb{E}[X] = \mathbb{E}\left[\sum_{i<j} X_{ij}\right] = \sum_{i<j} \mathbb{E}\left[X_{ij}\right]$$

$$= \sum_{i<j} \frac{2}{j-i+1} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1}$$

Randomized Quicksort Analysis

Theorem: The expected number of comparisons when sorting n elements using randomized quicksort is $T(n) \leq 2n \ln n$.

Proof:

$$\mathbb{E}[X] = 2 \cdot \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{j-i+1} = 2 \cdot \sum_{i=1}^{n-1} \sum_{\ell=2}^{n-i+1} \frac{1}{\ell}$$

$$\leq 2 \cdot \sum_{i=1}^{n-1} \underbrace{\sum_{\ell=2}^{n} \frac{1}{\ell}}_{= H(n) - 1}$$

$$= 2 \cdot (n-1) \cdot (H(n) - 1)$$

$$\leq 2 \cdot (n-1) \cdot \ln n$$

Harmonic Series:

$$H(k) \coloneqq \sum_{i=1}^{k} \frac{1}{i}$$

$$H(k) \leq 1 + \ln k$$