Chapter 7
Randomization

Algorithm Theory
WS 2015/16

Fabian Kuhn

Randomized Quicksort

Quicksort:

worst case cost of QS: $O(n^2)$

S	v

$S_\ell < v$	v	$S_r > v$

function Quick (S: sequence): sequence;

{returns the sorted sequence S}

begin

  if $\#S \leq 1$ then **return** S

  else { choose pivot element v in S;

    partition S into S_ℓ with elements $< v$,

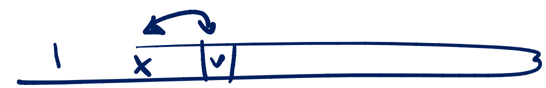
    and S_r with elements $> v$

    return | Quick(S_ℓ) | v | Quick(S_r) |

end;

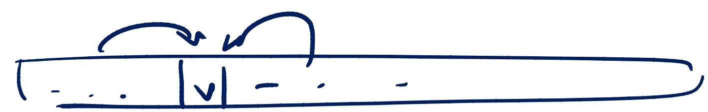
Randomized Quicksort Analysis

Randomized Quicksort: pick uniform random element as pivot



Running Time of sorting n elements:

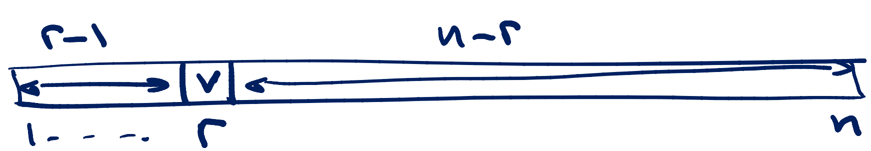
- Let's just count the number of comparisons
- In the partitioning step, all $n - 1$ non-pivot elements have to be compared to the pivot



- **Number of comparisons:**

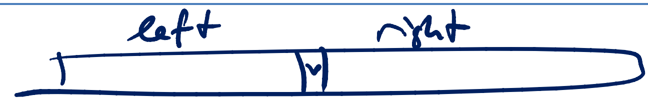
$$\boldsymbol{n - 1\ +\ \#\text{comparisons in recursive calls}}$$

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



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

$C = n-1 + C_\ell + C_r$

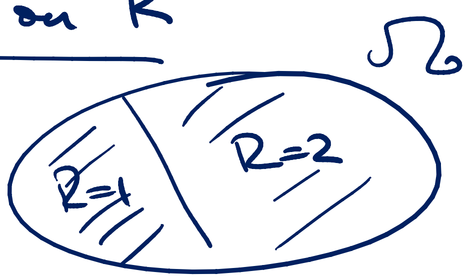
$\mathbb{E}[C] = \mathbb{E}[n-1+C_\ell+C_r]$

lin. of exp.

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

depend on R

Law of Total Expectation:



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

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

Randomized Quicksort Analysis

We have seen that:

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

$T(n) = \mathbb{E}[C]$

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:

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

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

Randomized Quicksort Analysis

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

Proof:

I.H.: $T(i) \leq 2 \cdot i \ln i$

$T(1) = 0$

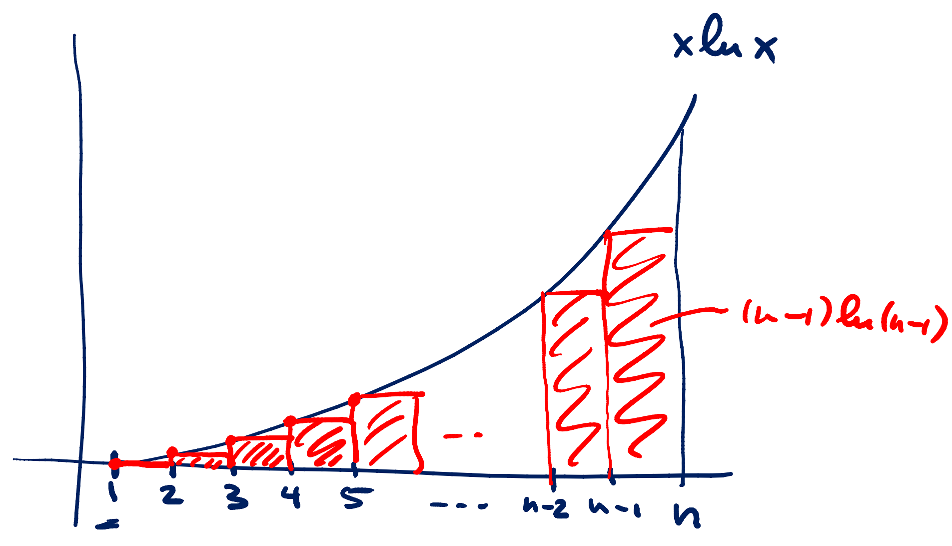
$$T(n) = \sum_{r=1}^{n} \frac{1}{n} \cdot \big(n - 1 + T(r-1) + T(n-r)\big), \qquad T(0) = 0$$

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

$$= n - 1 + \frac{2}{n} \cdot \sum_{i=0}^{n-1} T(i)$$

$$\overset{(IH)}{\leq} n - 1 + \frac{4}{n} \cdot \sum_{i=0}^{n-1} i \ln(i)$$

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



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

Proof:

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

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

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

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

$$= 2n \ln(n) + \underbrace{\frac{1}{n} - 1}_{<0} < \underline{\underline{2n \ln(n)}}$$

$$\hookrightarrow \underline{\underline{\mathbb{E}[C] < 2n \ln(n)}}$$

<u>also possible to show</u>

$C = O(n \log n)$ w.h.p.

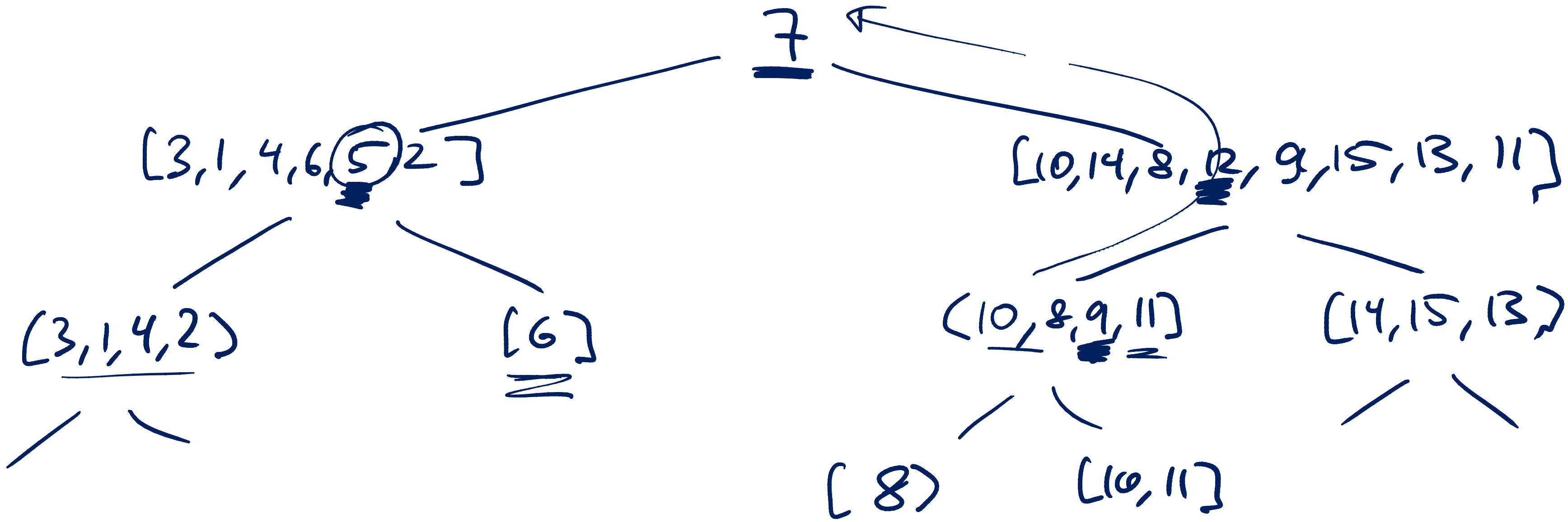
with high prob.

$$1 - \frac{1}{n^c}$$

Alternative Analysis

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

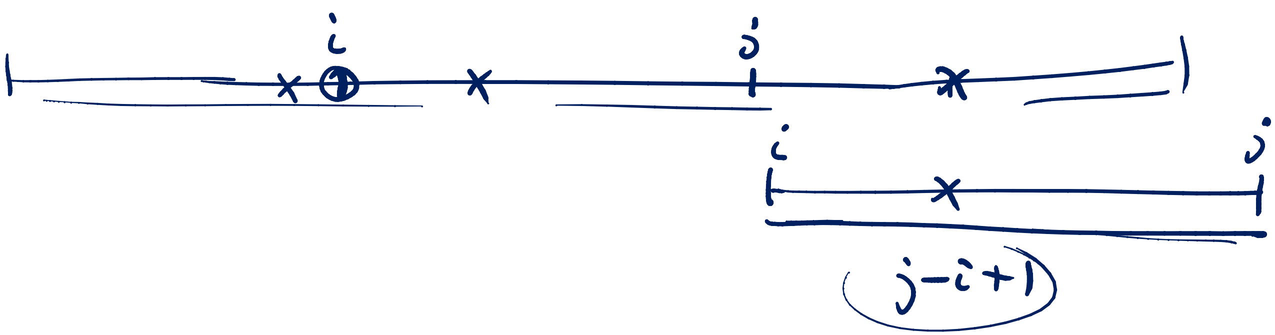
Viewing quicksort run as a tree:



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



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

Counting Comparisons

Random variable for every pair of elements (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 $a_1, \dots, a_n \in \mathbb{R}$,

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

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

$$\underline{E[X]} = E\Big[\sum_{i<j} X_{ij}\Big] = \sum_{i<j} E[X_{ij}] = \sum_{i<j} \frac{2}{j-i+1} = \underline{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1}}$$

Randomized Quicksort Analysis

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

Proof:

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

$$\leq 2 \sum_{i=1}^{n-1} \sum_{k=2}^{n} \frac{1}{k}$$

$$= 2 \sum_{i=1}^{n} \big(H(n) - 1 \big)$$

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

$$\leq \underline{2n \ln(n)}$$

Harmonic series

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

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