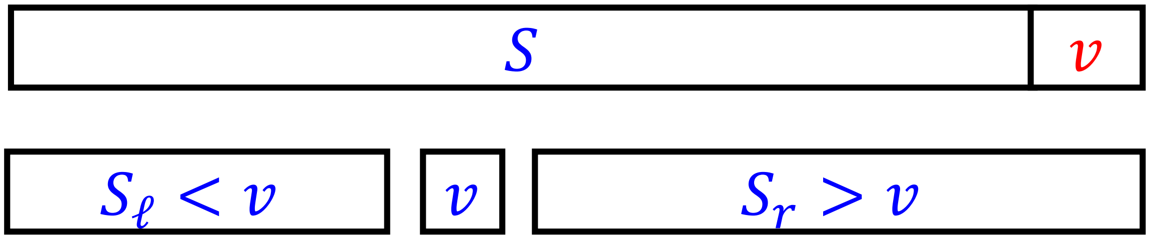
# Chapter 7
# Randomization

## Algorithm Theory
## WS 2019/20


## Philipp Bamberger

# Randomized Quicksort

**Quicksort:**




**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 $S_\ell$ with elements $< v$,

and $S_r$ with elements $> v$

**return** $\boxed{\text{Quick}(S_\ell) \mid v \mid \text{Quick}(S_r)}$

**end**;

# Randomized Quicksort Analysis

**Randomized Quicksort:** pick uniform random element as pivot

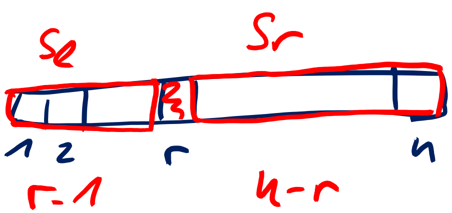
**Running Time** of sorting $n$ elements:

- 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:**

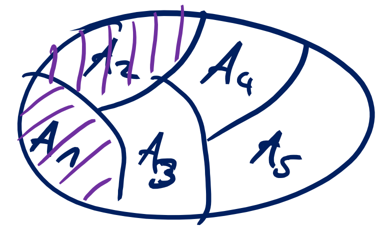
$$\boldsymbol{n-1} + \underbrace{\textbf{\#comparisons in recursive calls}}_{\text{depends on partition} \rightarrow \text{random variable}}$$

- **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 $X : \Omega \to \mathbb{R}$
- a set of events $A_1 \subseteq \Omega, \dots, A_k$ that partition $\Omega$
  - E.g., for a second random variable $Y$, we could have
    $A_i := \{\omega \in \Omega : Y(\omega) = i\}$



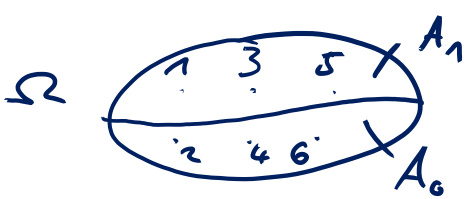

$Y : \Omega \to \mathbb{R}$

**Law of Total Expectation**

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

**Example:**

- $X$: outcome of rolling a die $\quad E[X] = 3.5$
- $A_0 = \{X \text{ is even}\}, \quad A_1 = \{X \text{ is odd}\}$




$$E[X] = P(A_0)\, E[X \mid A_0] + P(A_1)\, E[X \mid A_1]$$
$$\tfrac{1}{2} \cdot 4 + \tfrac{1}{2} \cdot 3 = 3.5$$

# Randomized Quicksort Analysis

**Random variables:**

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

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

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

**Law of Total Expectation:**

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

$$= \sum_{r=1}^{n} \mathbb{P}(R = r) \cdot \overbrace{(n - 1 + \mathbb{E}[\underbrace{C_\ell | R = r}_{\text{\# comparisons for an array of length } r-1}] + \mathbb{E}[\underbrace{C_r | R = r}_{\text{\# comparisons for an array of length } n-r}])}^{\mathbb{E}[C|R=r]}$$

# Randomized Quicksort Analysis

We have seen that: $\mathbb{P}(R=r)=\frac{1}{n}$

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

**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:** Induction on $n$

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

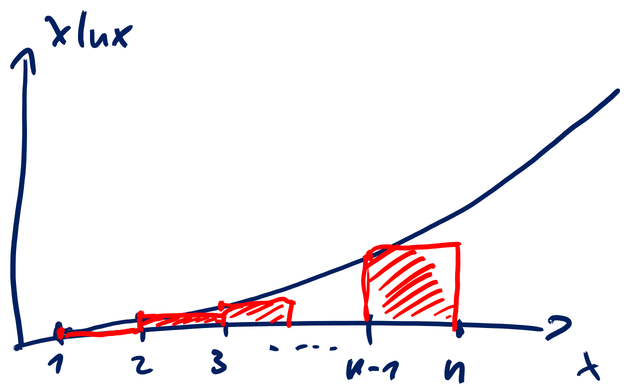
$$= n - 1 + \frac{1}{n}\sum_{i=0}^{n-1}\big(T(i) + T(n-i-1)\big)$$

$$\sum_{i=0}^{n-1} T(i) = \sum_{i=0}^{n-1} T(n-i-1)$$

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

$$\overset{\text{I.H.}}{\leq} n - 1 + \frac{4}{n}\underbrace{\sum_{i=1}^{n-1} i \ln(i)}$$

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



# Randomized Quicksort Analysis

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

**Proof:**

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

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

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

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

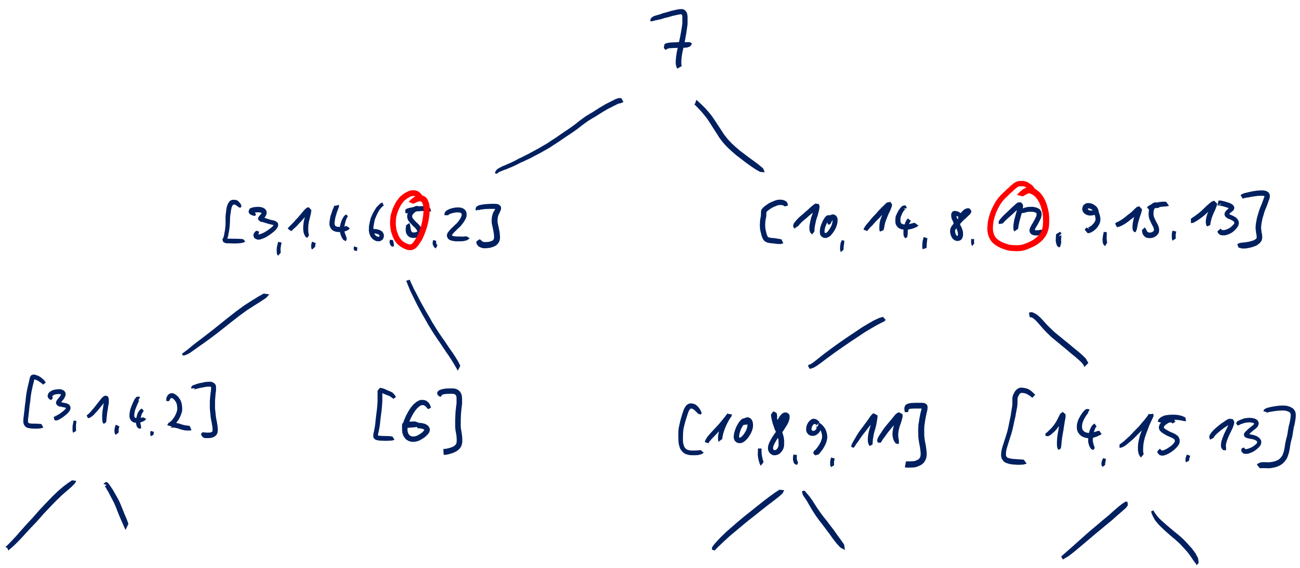
$$= 2n \ln n + \underbrace{\left(\frac{1}{n} - 1\right)}_{\leq 0} \leq 2n \ln n$$

$$\mathbb{E}[C] \leq 2n \ln n$$

# Alternative Analysis

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

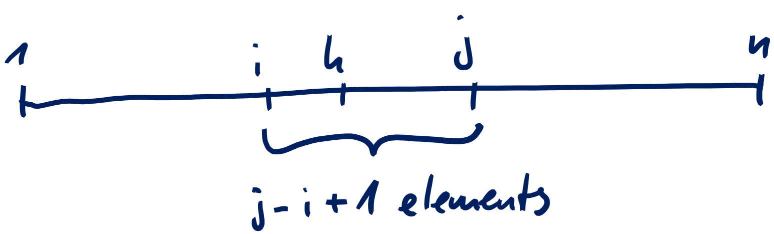
**Viewing quicksort run as a tree:**

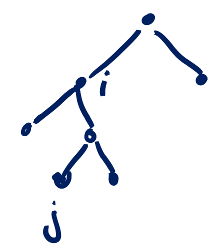


# Comparisons

- 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 $1,2,\ldots,n$
- 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 $(i, j)$:

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

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

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 $a_1, \dots, a_n \in \mathbb{R}$,

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

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

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

# Randomized Quicksort Analysis

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

**Proof:**

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

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

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

Harmonic series:

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

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

# Quicksort: High Probability Bound

- We have seen that the number of comparisons of randomized quicksort is $O(n \log n)$ in expectation.

- Can we also show that the number of comparisons is $O(n \log n)$ with high probability? $1 - \frac{1}{n^c}$

- **Recall:**

  On each recursion level, each pivot is compared once with each other element that is still in the same "part"

# Counting Number of Comparisons

- We looked at 2 ways to count the number of comparisons
  - recursive characterization of the expected number
  - number of different pairs of values that are compared

Let's consider yet another way:

- Each comparison is between a pivot and a non-pivot
- How many times is a specific array element $x$ compared as a non-pivot?

Value $x$ is compared as a non-pivot to a pivot once in every recursion level until one of the following two conditions apply:

1. $x$ is chosen as a pivot
2. $x$ is alone

# comparisons of x as non-pivot
= depth when x becomes pivot or alone

# Successful Recursion Level

$K_1 = n$

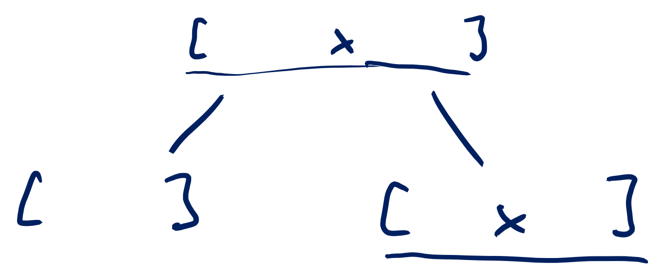
- Consider a specific recursion level $\ell$

$\underbrace{[\quad x \quad]}_{K_\ell}$

- Assume that at the beginning of recursion level $\ell$, element $x$ is in a sub-array of length $K_\ell$ that still needs to be sorted.
- If $x$ has been chosen as a pivot before level $\ell$, we set $K_\ell := 1$

**Definition:** We say that recursion level $\ell$ is successful for element $x$ iff the following is true:

$$K_{\ell+1} = 1 \quad \text{or} \quad K_{\ell+1} \leq \frac{2}{3} \cdot K_\ell$$



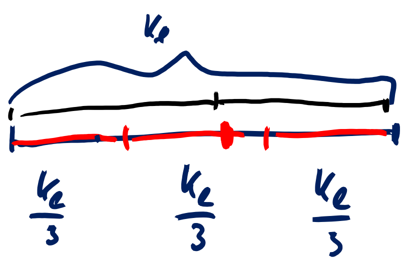
# Successful Recursion Level

**Lemma:** For every recursion level $\ell$ and every array element $x$, it holds that level $\ell$ is successful for $x$ with probability at least ${}^{1}/_{3}$, independently of what happens in other recursion levels.

**Proof:**

$k_\ell = 1 \quad \rightarrow \quad k_{\ell+1} = 1$ ✓

$k_\ell > 1$




If the pivot is in the middle part, both remaining arrays have size $\leq \frac{2}{3} k_\ell \Rightarrow k_{\ell+1} \leq \frac{2}{3} k_\ell$

Probability for this is $\geq \frac{1}{3}$

# Number of Successful Recursion Levels

**Lemma:** If among the first $\ell$ recursion levels, at least $\underbrace{\log_{3/2}(n)}_{= O(\log n)}$ are successful for element $x$, we have $K_{\ell+1} = 1$.

**Proof:**

$\underline{K_1 = n}$ , $K_{i+1} \leq K_i$ If level $i$ is succ. : $K_{i+1} \leq \frac{2}{3} K_i$

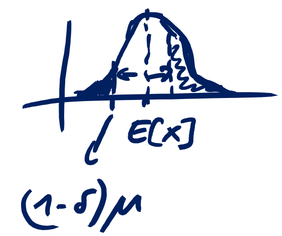
$$K_{\ell+1} \leq n \left(\frac{2}{3}\right)^{\#\text{succ. levels}} \leq n \underbrace{\left(\frac{2}{3}\right)^{\log_{\frac{3}{2}} n}}_{= \frac{1}{n}} = 1$$

$$\left(\frac{3}{2}\right)^{\log_{\frac{3}{2}} n} = n$$

$$\left(\left(\frac{3}{2}\right)^{-1}\right)^{\log_{\frac{3}{2}} n} = \frac{1}{n}$$

# Chernoff Bounds

- Let $X_1, \dots, X_n$ be independent 0-1 random variables and define $p_i := \mathbb{P}(X_i = 1)$. if $p_i = p$ for all $i$: $X \sim \mathrm{Bin}(n, p)$
- Consider the random variable $X = \sum_{i=1}^{n} X_i$
- We have $\mu := \mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} p_i$




**Chernoff Bound (Lower Tail):** $P(X < \mu/2) < e^{-\mu/8}$

$\delta < 1$

$$\forall \delta > 0: \; \mathbb{P}(X < (1-\delta)\mu) < e^{-\delta^2\mu/2}$$

**Chernoff Bound (Upper Tail):**

$$\forall \delta > 0: \; \mathbb{P}(X > (1+\delta)\mu) < \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu} < e^{-\delta^2\mu/3}$$

**holds for $\delta \leq 1$**

# Chernoff Bounds, Example

Assume that a fair coin is flipped $n$ times. What is the probability to have

1. less than $n/3$ heads?

2. more than $0.51n$ tails?

3. less than ${}^{n}/_{2} - \sqrt{c \cdot \ n \ln n}$ tails?

# Number of Comparisons for $x$

**Lemma:** For every array element $x$, with high probability, as a non-pivot, $x$ is compared to a pivot at most $O(\log n)$ times.

**Proof:**

show: w.h.p. there are $\log_{3/2} n$ succ. rec. levels among the first $O(\log n)$ levels.

$K$ rec. levels, $X_i = \begin{cases} 1 & \text{level } i \text{ is succ.} \\ 0 & \text{else} \end{cases}$, $X = \sum_{i=1}^{k} X_i$, $E[X] \geq \frac{k}{3}$

$P(X_i = 1) \geq \frac{1}{3}$

$$P\left(X < \frac{k}{6}\right) = P\left(X < \left(1 - \frac{1}{2}\right)\frac{k}{3}\right) < e^{-\frac{1}{4}\cdot\frac{k}{3\cdot 2}} = e^{-\frac{k}{36}} = \frac{1}{n^c}$$

$k = 36c\ln(n)$

# Number of Comparisons for $x$

**Lemma:** For every array element $x$, with high probability, as a non-pivot, $x$ is compared to a pivot at most $O(\log n)$ times.

**Proof:**

# Number of Comparisons

**Theorem:** With high probability, the total number of comparisons is at most $\boldsymbol{O(n \log n)}$.

**Proof:**

Last lemma + Union bound

probability that all elements are compared as non-pivot at most $O(\log n)$ times

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