Algorithm Theory

Chapter 7
Randomized Algorithms

Part IV:
Rand. Quicksort : High Probability Bound

Fabian Kuhn

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?
- **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?

Element 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

Successful Recursion Level

- Consider a specific recursion level ℓ
  - Where the first recursion level is level 1

Define K_ℓ as follows:

- If x is contained in a subarray on recursion level ℓ, then K_ℓ is defined as the length of the subarray containing x on level ℓ.
  - We therefore have $K_1 = n$ and $K_{\ell+1} \leq K_\ell$ for all $\ell \geq 1$
- If x has been chosen as a pivot before level ℓ, we set $K_\ell \coloneqq 1$

#comparisons of x as non-pivot $\leq$ #levels ℓ for which $K_\ell > 1$

Definition: We say that recursion level ℓ 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 ℓ and every array element x, it holds that level ℓ is successful for x with probability at least ${}^{1}/_{3}$, independently of what happens in other recursion levels.

Proof:

- Assume that $K_\ell > 1$, otherwise level ℓ is trivially successful

$\geq \left\lfloor {}^{K_\ell}/_{3} \right\rfloor$	$= \left\lceil {}^{K_\ell}/_{3} \right\rceil$	$\geq \left\lfloor {}^{K_\ell}/_{3} \right\rfloor$

- If pivot is in the middle part, both remaining parts have size

$$\leq K_\ell - \left\lfloor {}^{K_\ell}/_{3} \right\rfloor - 1 \leq {}^{2}/_{3} \cdot K_\ell.$$

  - In this case, level ℓ is successful

- The probability that the pivot in in the middle part is $\geq {}^{1}/_{3}$.

Number of Successful Recursion Levels

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

Proof:

- We know that
$$K_1 = n, \qquad \forall i \geq 1 : K_{i+1} \leq K_i$$
- If level i is successful, then $K_{i+1} \leq {}^2/_3 \cdot K_i$ or $K_{i+1} = 1$
- If s among the first ℓ levels are successful, then
$$K_{\ell+1} \leq \max\left\{1, n \cdot \left({}^2/_3\right)^s\right\}$$
- If $s \geq \log_{3/2}(n)$, then $K_{\ell+1} \leq 1$.

Chernoff Bounds

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

If $p_i = p$ for all i:

$X \sim \mathrm{Bin}(n, p)$

Chernoff Bound (Lower Tail):

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

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

$$p_i = p = \frac{1}{2}, \qquad \mu := \mathbb{E}[X] = np = \frac{n}{2}$$

1. less than $n/3$ heads?

$$\mathbb{P}\left(X < \frac{n}{3}\right) = \mathbb{P}\left(X < \left(1 - \frac{1}{3}\right) \cdot \frac{n}{2}\right) < e^{-\frac{1}{2} \cdot \frac{1}{3^2} \cdot \frac{n}{2}} = e^{-n/36}$$

$$\mathbb{P}(X < (1-\delta)\mu) < e^{-\frac{\delta^2}{2}\mu}$$

2. more than $0.51n$ tails?

$$\mathbb{P}\left(X < (1+0.02) \cdot \frac{n}{2}\right) < e^{-\frac{0.02^2}{3} \cdot \frac{n}{2}} \approx e^{-0.0000667n}$$

$$\mathbb{P}(X > (1+\delta)\mu) < e^{-\frac{\delta^2}{3}\mu}$$

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

$$\mathbb{P}\left(X < \left(1 - \frac{2\sqrt{c \cdot n \ln n}}{n}\right) \cdot \frac{n}{2}\right) < e^{-\frac{4c \cdot n \ln n}{2n^2} \cdot \frac{n}{2}} = e^{-c \cdot \ln n} = \frac{1}{n^c}$$

With high probability, #heads/tails $= \frac{n}{2} \pm O\left(\sqrt{n \log n}\right)$

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:

- Consider some level $i \geq 1$, and let if level i not successful

$$q_i := \mathbb{P}(\text{level } i \text{ successful for } x \mid \text{history up to level } i)$$

- Previous lemma $\Longrightarrow q_i \geq 1/3$

- Define random variable

$$X_i := \begin{cases} 0 & \text{if level } i \text{ not successful for } x \\ 1 & \text{with probability } \frac{1/3}{q_i} \text{ if level } i \text{ successful for } x \end{cases}$$

- Then, $\mathbb{P}(X_i = 1) = 1/3$ and X_i are independent for different i

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:

- X_i independent, $\mathbb{P}(X_i = 1) = {}^1/_3$, $X_i = 1 \Longrightarrow$ level i successful
- Consider the first t levels and define $X := \sum_i^t X_i$
  - $\mathbb{E}[X] = {}^1/_3 \cdot t$
  - $X \leq$ successful levels for x among first t levels
- Hence, if $X \geq \log_{3/2}(n)$, then $K_{t+1} = 1$
- We thus need that for any const. $c > 0$ and some $t = O(\log n)$,

$$\mathbb{P}\left(X < \log_{3/2}(n)\right) \leq \frac{1}{n^c}$$

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:

- $\mu := \mathbb{E}[X] = {}^{1}/_{3} \cdot t$, for $c > 0$ and some $t = O(\log n)$, we need

$$\mathbb{P}\left(X < \log_{3/2}(n)\right) \leq \frac{1}{n^c}$$

- **Chernoff:** $\mathbb{P}(X < (1-\delta)\mu) \leq e^{-\frac{\delta^2}{2}\cdot\mu} \implies \mathbb{P}\left(X < {}^{\mu}/_{2}\right) \leq e^{-\frac{\mu}{8}}$

- We need $\mu \geq 2 \cdot \log_{3/2}(n)$ such that ${}^{\mu}/_{2} \geq \log_{3/2}(n)$

- We need $\mu \geq 8c \cdot \ln n$ such that $e^{-\mu/8} \leq n^{-c}$

- We can therefore choose $t = 3 \cdot \mu = O(\log n)$.

Number of Comparisons

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

Proof:

- For every const. $c > 0$, there exists const. $\alpha > 0$, s.t. for every element x, the number of comparisons for element x as a non-pivot is $\leq \alpha \ln n$ with probability at least $1 - {}^{1}/_{n^c}$.

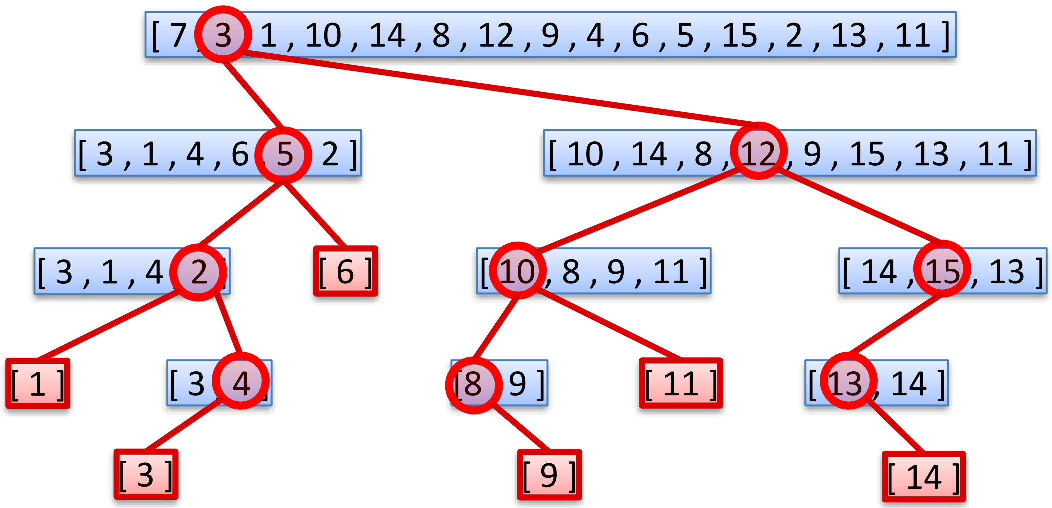
- Define event $\mathcal{E}_x \coloneqq \{\#\text{comparisons for } x \text{ as non–pivot} > \alpha \ln n\}$
  - $\mathbb{P}(\mathcal{E}_x) \leq n^{-c}$

- Union bound over all events $\mathcal{E}_x$:

$$\mathbb{P}\left(\bigcup_{x=1}^{n} \mathcal{E}_x\right) \leq \sum_{x=1}^{n} \mathbb{P}(\mathcal{E}_x) \leq n \cdot \frac{1}{n^c} = \frac{1}{n^{c-1}}$$

Relation to Random Binary Search Trees

Consider Recursion Tree: Label each subarray of size > 1 by the pivot and each subarray of size $= 1$ by the element in it.



- We get a binary search tree (BST) on the n elements
  - Corresponds to the BST with a random insertion order
- #comparisons of element x as non-pivot $=$ depth of x in tree
  - Our analysis shows that the height of a random BST is $O(\log n)$, w.h.p.
- #comp. of rand. quicksort $= n \cdot$ average depth in a random BST

Types of Randomized Algorithms

Las Vegas Algorithm:

- always a correct solution
- running time is a random variable

- **Example:** randomized quicksort, contention resolution

Monte Carlo Algorithm:

- probabilistic correctness guarantee (**m**ostly **c**orrect)
- fixed (deterministic) running time

- **Example:** primality test