# Randomised Algorithms

## Lecture 3: Concentration Inequalities, Application to Quick-Sort, Extensions

Thomas Sauerwald (tms41@cam.ac.uk)

Lent 2024

UNIVERSITY OF CAMBRIDGE

# Outline

Application 2: Randomised QuickSort

Extensions of Chernoff Bounds

Applications of Method of Bounded Differences

Appendix: More on Moment Generating Functions (non-examinable)

# QuickSort

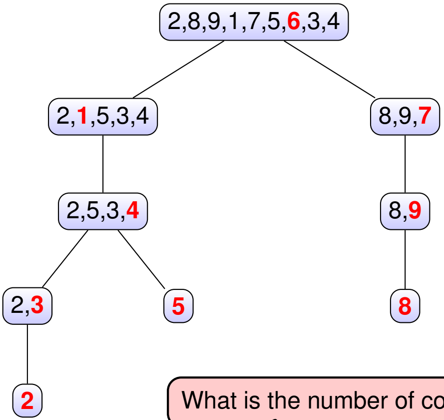
QUICKSORT (Input $A[1], A[2], \ldots, A[n]$)

Pick an element from the array, the so-called pivot
**If** $|A| = 0$ or $|A| = 1$ **then**
    **return** $A$
**else**
    Create two subarrays $A_1$ and $A_2$ (without the pivot) such that:
        $A_1$ contains the elements that are smaller than the pivot
        $A_2$ contains the elements that are greater (or equal) than the pivot
    QUICKSORT($A_1$)
    QUICKSORT($A_2$)
    **return** $A$

- Example: Let $A = (2, 8, 9, 1, 7, 5, 6, 3, 4)$ with $A[7] = 6$ as pivot.
  $\Rightarrow A_1 = (2, 1, 5, 3, 4)$ and $A_2 = (8, 9, 7)$
- Worst-Case Complexity (number of comparisons) is $\Theta(n^2)$, while Average-Case Complexity is $O(n \log n)$.

We will now give a proof of this "well-known" result!

# QuickSort: How to Count Comparisons




What is the number of comparisons?

Note that the number of comparison by QUICKSORT is equivalent to the sum of the depths of all nodes in the tree (why?). In this case:

$$0 + 1 + 1 + 2 + 2 + 3 + 3 + 3 + 4 = 19.$$

# Randomised QuickSort: Analysis (1/4)

How to pick a good pivot? We don't, **just pick one at random.**

This should be your standard answer in this course ☺

Let us analyse QUICKSORT with random pivots.

1. Assume $A$ consists of $n$ different numbers, w.l.o.g., $\{1, 2, \ldots, n\}$
2. Let $H_i$ be the deepest level where element $i$ appears in the tree.
   Then the number of comparison is $H = \sum_{i=1}^{n} H_i$
3. We will prove that there exists $C > 0$ such that

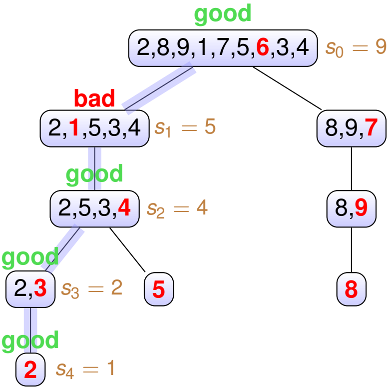
$$\mathbf{P}[H \leq Cn \log n] \geq 1 - n^{-1}.$$

4. Actually, we will prove sth slightly stronger:

$$\mathbf{P}\left[\bigcap_{i=1}^{n} \{H_i \leq C \log n\}\right] \geq 1 - n^{-1}.$$

## Randomised QuickSort: Analysis (2/4)

- Let $P$ be a path from the root to the deepest level of some element
  - A node in $P$ is called **good** if the corresponding pivot partitions the array into two subarrays each of size at most $2/3$ of the previous one
  - otherwise, the node is **bad**
- Further let $s_t$ be the size of the array at level $t$ in $P$.




- Element 2: $(2, 8, 9, 1, 7, 5, \mathbf{6}, 3, 4) \to (2, \mathbf{1}, 5, 3, 4) \to (2, 5, 3, \mathbf{4}) \to (2, \mathbf{3}) \to (\mathbf{2})$

# Randomised QuickSort: Analysis (3/4)

- Consider now any element $i \in \{1, 2, \ldots, n\}$ and construct the path $P = P(i)$ one level by one
- For $P$ to proceed from level $k$ to $k+1$, the condition $s_k > 1$ is necessary

How far could such a path $P$ possibly run until we have $s_k = 1$?

- We start with $s_0 = n$
- **First Case**, **good** node: $s_{k+1} \leq \frac{2}{3} \cdot s_k$.
- **Second Case**, **bad** node: $s_{k+1} \leq s_k$.

This even holds always, i.e., deterministically!

$\Rightarrow$ There are at most $T = \frac{\log n}{\log(3/2)} < 3 \log n$ many **good** nodes on any path $P$.

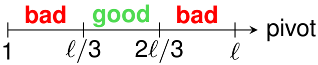
- Assume $|P| \geq C \log n$ for $C := 24$

  $\Rightarrow$ number of **bad** vertices in the first $24 \log n$ levels is more than $21 \log n$.

Let us now upper bound the probability that this “bad event” happens!

# Randomised QuickSort: Analysis (4/4)

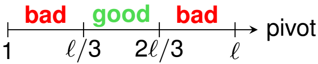
- Consider the first $24 \log n$ vertices of $P$ to the deepest level of element $i$.
- For any level $j \in \{0, 1, \ldots, 24 \log n - 1\}$, define an indicator variable $X_j$:
  - $X_j = 1$ if the node at level $j$ is **bad**,
  - $X_j = 0$ if the node at level $j$ is **good**.
- $\mathbf{P}\left[\, X_j = 1 \mid X_0 = x_0, \ldots, X_{j-1} = x_{j-1} \,\right] \leq \frac{2}{3}$
- $X := \sum_{j=0}^{24 \log n - 1} X_j$ satisfies relaxed independence assumption (Lecture 2)




**Question:** Edge Case: What if the path $P$ does not reach level $j$?

# Randomised QuickSort: Analysis (4/4)

- Consider the first $24 \log n$ vertices of $P$ to the deepest level of element $i$.
- For any level $j \in \{0, 1, \ldots, 24 \log n - 1\}$, define an indicator variable $X_j$:
  - $X_j = 1$ if the node at level $j$ is **bad**,
  - $X_j = 0$ if the node at level $j$ is **good**.
- $\mathbf{P}[\, X_j = 1 \mid X_0 = x_0, \ldots, X_{j-1} = x_{j-1} \,] \leq \frac{2}{3}$
- $X := \sum_{j=0}^{24 \log n - 1} X_j$ satisfies relaxed independence assumption (Lecture 2)

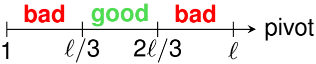



**Question:** Edge Case: What if the path $P$ does not reach level $j$?

**Answer:** We can then simply define $X_j$ as 0 (deterministically).

# Randomised QuickSort: Analysis (4/4)

- Consider the first $24 \log n$ vertices of $P$ to the deepest level of element $i$.
- For any level $j \in \{0, 1, \ldots, 24 \log n - 1\}$, define an indicator variable $X_j$:
  - $X_j = 1$ if the node at level $j$ is **bad**,
  - $X_j = 0$ if the node at level $j$ is **good**.




- $\mathbf{P}[X_j = 1 \mid X_0 = x_0, \ldots, X_{j-1} = x_{j-1}] \leq \frac{2}{3}$
- $X := \sum_{j=0}^{24 \log n - 1} X_j$ satisfies relaxed independence assumption (Lecture 2)

We can now apply the “nicer” **Chernoff Bound**!

- We have $\mathbf{E}[X] \leq (2/3) \cdot 24 \log n = 16 \log n$
- Then, by the “nicer” Chernoff Bounds

$$\mathbf{P}[X \geq \mathbf{E}[X] + t] \leq e^{-2t^2/n}$$

$$\mathbf{P}[X > 21 \log n] \leq \mathbf{P}[X > \mathbf{E}[X] + 5 \log n]$$

- Hence $P$ has more than $24 \log n$ nodes with probability at most $n^{-2}$.
- As there are in total $n$ paths, by the union bound, the probability that at least one of them has more than $24 \log n$ nodes is at most $n^{-1}$.
- This implies $\mathbf{P}\left[\bigcap_{i=1}^{n} \{H_i \leq 24 \log n\}\right] \geq 1 - n^{-1}$, as needed. $\square$

# Randomised QuickSort: Final Remarks

- Well-known: any comparison-based sorting algorithm needs $\Omega(n \log n)$
- A classical result: expected number of comparison of randomised QuickSort is $2n \log n + O(n)$ (see, e.g., book by Mitzenmacher & Upfal)

**Exercise:** *[Ex 2-3.6]* Our upper bound of $O(n \log n)$ whp also immediately implies a $O(n \log n)$ bound on the expected number of comparisons!

- It is possible to deterministically find the best pivot element that divides the array into two subarrays of the same size.
- The latter requires to compute the median of the array in linear time, which is not easy...
- The presented randomised algorithm for QuickSort is much easier to implement!

# Outline

Application 2: Randomised QuickSort

Extensions of Chernoff Bounds

Applications of Method of Bounded Differences

Appendix: More on Moment Generating Functions (non-examinable)

# Hoeffding's Extension

- Besides **sums of independent Bernoulli** random variables, sums of independent and bounded random variables are very frequent in applications.
- Unfortunately the distribution of the $X_i$ may be unknown or hard to compute, thus it will be hard to compute the moment-generating function.
- Hoeffding's Lemma helps us here:

You can always consider
$X' = X - \mathbf{E}[X]$

**Hoeffding's Extension Lemma**

Let $X$ be a random variable with mean 0 such that $a \leq X \leq b$. Then for all $\lambda \in \mathbb{R}$,

$$\mathbf{E}\left[e^{\lambda X}\right] \leq \exp\left(\frac{(b-a)^2\lambda^2}{8}\right)$$

We omit the proof of this lemma!

## Hoeffding Bounds

**Hoeffding's Inequality**

Let $X_1, \ldots, X_n$ be independent random variable with mean $\mu_i$ such that $a_i \leq X_i \leq b_i$. Let $X = X_1 + \ldots + X_n$, and let $\mu = \mathbf{E}[X] = \sum_{i=1}^n \mu_i$. Then for any $t > 0$

$$\mathbf{P}[X \geq \mu + t] \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right),$$

and

$$\mathbf{P}[X \leq \mu - t] \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

### Proof Outline (skipped):

- Let $X_i' = X_i - \mu_i$ and $X' = X_1' + \ldots + X_n'$, then $\mathbf{P}[X \geq \mu + t] = \mathbf{P}[X' \geq t]$
- $\mathbf{P}[X' \geq t] \leq e^{-\lambda t} \prod_{i=1}^n \mathbf{E}\left[e^{\lambda X_i'}\right] \leq \exp\left[-\lambda t + \frac{\lambda^2}{8} \sum_{i=1}^n (b_i - a_i)^2\right]$
- Choose $\lambda = \frac{4t}{\sum_{i=1}^n (b_i - a_i)^2}$ to get the result.

This is not magic! you just need to optimise $\lambda$!

## Method of Bounded Differences

Framework

Suppose, we have independent random variables $X_1, \ldots, X_n$. We want to study the random variable:

$$f(X_1, \ldots, X_n)$$

Some examples:

1. $X = X_1 + \ldots + X_n$ (our setting earlier)
2. In balls into bins, $X_i$ indicates where ball $i$ is allocated, and $f(X_1, \ldots, X_m)$ is the number of empty bins
3. In a randomly generated graph, $X_i$ indicates if the $i$-th edge is present and $f(X_1, \ldots, X_m)$ represents the number of connected components of $G$

In all those cases (and more) we can easily prove concentration of $f(X_1, \ldots, X_n)$ around its mean by the so-called **Method of Bounded Differences**.

# Method of Bounded Differences

A function $f$ is called Lipschitz with parameters $\mathbf{c} = (c_1, \ldots, c_n)$ if for all $i = 1, 2, \ldots, n$,

$$|f(x_1, x_2, \ldots, x_{i-1}, \boldsymbol{x_i}, x_{i+1}, \ldots, x_n) - f(x_1, x_2, \ldots, x_{i-1}, \widetilde{\boldsymbol{x_i}}, x_{i+1}, \ldots, x_n)| \leq c_i,$$

where $x_i$ and $\widetilde{x_i}$ are in the domain of the $i$-th coordinate.

**McDiarmid's inequality**

Let $X_1, \ldots, X_n$ be independent random variables. Let $f$ be Lipschitz with parameters $\mathbf{c} = (c_1, \ldots, c_n)$. Let $X = f(X_1, \ldots, X_n)$. Then for any $t > 0$,

$$\mathbf{P}[\, X \geq \mu + t \,] \leq \exp\left( -\frac{2t^2}{\sum_{i=1}^{n} c_i^2} \right),$$

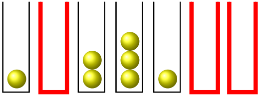
and

$$\mathbf{P}[\, X \leq \mu - t \,] \leq \exp\left( -\frac{2t^2}{\sum_{i=1}^{n} c_i^2} \right).$$

- Notice the similarity with Hoeffding's inequality! *[Exercise 2/3.14]*
- The proof is omitted here (it requires the concept of martingales).

# Outline

# Application 3: Balls into Bins (again...)

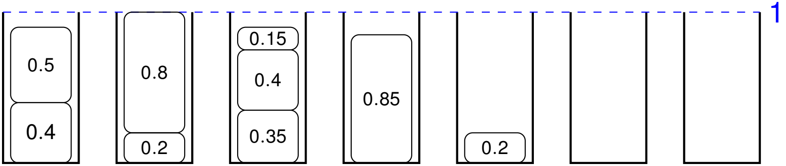


- Consider again $m$ balls assigned uniformly at random into $n$ bins.
- Enumerate the balls from 1 to $m$. Ball $i$ is assigned to a random bin $X_i$
- Let $Z$ be the number of empty bins (after assigning the $m$ balls)
- $Z = Z(X_1, \ldots, X_m)$ and $Z$ is Lipschitz with $\mathbf{c} = (1, \ldots, 1)$
  (If we move one ball to another bin, number of empty bins changes by $\leq 1$.)
- By McDiarmid's inequality, for any $t \geq 0$,

$$\mathbf{P}\left[\, |Z - \mathbf{E}[Z]| > t \,\right] \leq 2 \cdot e^{-2t^2/m}.$$

This is a decent bound, but for some values of $m$ it is far from tight and stronger bounds are possible through a refined analysis.

# Application 4: Bin Packing




- We are given $n$ items of sizes in the unit interval $[0, 1]$
- We want to pack those items into the fewest number of unit-capacity bins
- Suppose the item sizes $X_i$ are independent random variables in $[0, 1]$

- Let $B = B(X_1, \ldots, X_n)$ be the optimal number of bins
- The Lipschitz conditions holds with $\boldsymbol{c} = (1, \ldots, 1)$. **Why?**
- Therefore

$$\mathbf{P}\left[\, |B - \mathbf{E}[B]| \geq t \,\right] \leq 2 \cdot e^{-2t^2/n}.$$

This is a typical example where proving concentration is much easier than calculating (or estimating) the expectation!

# Outline

Application 2: Randomised QuickSort

Extensions of Chernoff Bounds

Applications of Method of Bounded Differences

Appendix: More on Moment Generating Functions (non-examinable)

## Moment Generating Functions (non-examinable)

**Moment-Generating Function**

The moment-generating function of a random variable $X$ is

$$M_X(t) = \mathbf{E}\left[e^{tX}\right], \qquad \text{where } t \in \mathbb{R}.$$

Using power series of $e$ and differentiating shows that $M_X(t)$ encapsulates all moments of $X$.

**Lemma**

1. If $X$ and $Y$ are two r.v.'s with $M_X(t) = M_Y(t)$ for all $t \in (-\delta, +\delta)$ for some $\delta > 0$, then the distributions $X$ and $Y$ are identical.
2. If $X$ and $Y$ are independent random variables, then

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t).$$

Proof of 2:

$$M_{X+Y}(t) = \mathbf{E}\left[e^{t(X+Y)}\right] = \mathbf{E}\left[e^{tX} \cdot e^{tY}\right] \stackrel{(!)}{=} \mathbf{E}\left[e^{tX}\right] \cdot \mathbf{E}\left[e^{tY}\right] = M_X(t)M_Y(t) \qquad \square$$