Logica e Metodi Probabilistici per L'Informatica
Third session of Exercises

Academic year 2021/2022

Instructor: Prof. Stefano Leonardi   Teaching Assistant: Federico Fusco

April 5, 2022

Recap.

Algorithm 1 Randomized Quicksort Algorithm (as in Algorithm 2.1 of the textbook)

Input: A list $S = \{x_1, x_2, \ldots, x_n\}$ of n distinct numbers
Output: The elements of S in sorted order
If S has one or zero elements, return S, otherwise continue.
Choose an elements x of S uniformly at random. $\triangleright$ x is the pivot!
Compare every element in S with x and divide $S \setminus \{x\}$ in two sublists:

  (a) S_1 has all the elements of S that are less than x

  (b) S_2 has all those that are greater than x

Use Randomized Quicksort to sort S_1 and S_2.
Return the list S_1, x, S_2.

Theorem 1 (Multiplicative Chernoff bound)**.** Let $X_1, X_2, \ldots X_n$ be independent 0-1 random variables, let $X = \sum_{i=1}^{n} X_i$ and $\mu = \mathbb{E}[X]$. Then the following bound holds for all $\delta \in [0, 1]$:

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

Exercises.

Exercise 1. Suppose we keep flipping a fair coin until we see 100 heads. Show that the probability that we flip the coin more than 1000 times is smaller than e^{-160}?
Solution: We would like to use the Chernoff bounds we have derived, but it seems like we can't, because the number of flips is the sum of 100 geometric random variables, not $0-1$ random variables. The key observation that allows us to turn this into a Chernoff bounds problem (without having to derive Chernoff bounds for sums of geometric random variables) is that the probabilities of the following two events are equal:

- Event 1 (the event in the problem): We keep flipping a fair coin until we see 100 heads, and we flip the coin more than 1000 times.
- Event 2: We flip a fair coin exactly 1000 times, and the number of heads is less than 100.

The two events have the same probability and event 2 is something we can analyze with Chernoff bounds. Let X_i be the indicator of the event that the i^{th} toss is head, we have:

$$\mathbb{P}\left(\sum_{i=1}^{1000} X_i < 100\right) = \mathbb{P}\left(\sum_{i=1}^{1000} X_i - 500 < -\frac{4}{5}\cdot 500\right) \leq e^{-\left(\frac{4}{5}\right)^2\cdot 250} = e^{-160}.$$

Exercise 2. In this Exercise, we prove that the Randomized Quicksort Algorithm sorts a set of n numbers in time $O(n\log n)$ with high probability.

Good and bad nodes. For the sake of the analysis, we consider a view of the algorithm such that every point in the algorithm where it decides on a pivot element is called a node. Suppose the size of the set to be sorted at a particular node is s. The node is called *g*ood if the pivot element divides the set into two parts, each of size not exceeding $\frac{2}{3}s$. Otherwise the node is called *b*ad. The nodes can be thought of as forming a tree in which the root node has a whole set to be sorted and its children have two sets formed after the first pivot step and so on.

Lemma 1. The number of good nodes in any path from the root to a leaf in this tree is not greater than $2\cdot\log_2 n$.

Proof. Consider a path from root to leaf. Let it contain k good nodes. The number of elements left in each leaf is at most $n\cdot(2/3)^k$ which must be at least 1. Then we have that $k \leq \log_{3/2} n < 2\log n$. Then there cannot be more than $2\log(n)$ good nodes on any path from root to leaf. □

Lemma 2. With probability greater than $1-\frac{1}{n^2}$, the number of nodes in a given root to leaf path of the tree is not greater than $24\cdot\log_2 n$.

Proof. Consider how Exercise 1 relates to this problem. In the first Lemma, we showed that any path has to end after at most $2\log n$ good nodes –for now, let's just say it ends after exactly $2\log n$ good nodes. For a particular path, at each step, the set containing that element is divided in either a good way or a bad way, with a good division occurring with probability 1/3. In the coin flipping problem, suppose heads represents a good division, and we bias the coin so it comes up heads with probability 1/3. In the original problem, we keep dividing the set containing the particular element (cf. keep flipping biased coins) until we get $2\log n$ heads (cf. 100 heads). We want to find the probability that the path contains more than $24\log n$ nodes (cf. more than 1000 flips). By considering Event 1 and Event 2 in the coin flipping problem, we see that this equals the probability of getting less than $2\log n$ heads in exactly $24\log n$ flips. Let X_i be an indicator random variable for whether or not the i^{th} flip is heads. Recall that $\mathbb{E}[X_i] = \frac{1}{3}$. We get that:

$$\mathbb{P}\left(\sum_{i=1}^{24\log n} X_i < 2\log n\right) = \mathbb{P}\left(\sum_{i=1}^{24\log n} X_i - 8\log n < -\frac{3}{4}\cdot 8\log n\right) \leq e^{-\frac{9}{4}\log n} \leq \frac{1}{n^2}$$

Thus, with probability greater than $1-\frac{1}{n^2}$ there are at most $24\log n$ nodes on any path from root to leaf. □

Lemma 3. With probability greater than $1-\frac{1}{n}$, the number of nodes in the longest root to leaf path is not greater than $24\log_2 n$. (*Hint*: how many nodes are there in the tree?)

Proof. Let $\mathcal{E}_i$ be the event that the i^{th} path is longer than $24\log n$ nodes. Since there are at most n leaves, there are at most n paths from root to leaf. From Lemma 2 we know that $\mathbb{P}(\mathcal{E}_i) \leq n^{-2}$, so we can simply use the union bound:

$$\mathbb{P}(\cup_i \mathcal{E}_i) \leq \sum_i \mathbb{P}(\mathcal{E}_i) \leq \frac{n}{n^2} = \frac{1}{n}.$$

□

Theorem 2. Randomized Quicksort Algorithm sorts a set of n numbers in time $O(n \log n)$ with probability at least $1 - 1/n$.

Proof. The number of comparisons needed to place an element at the correct index is the length of the path from the root to the element's leaf in this Quicksort graph. There are n leaves, and with probability greater than $1 - \frac{1}{n}$, all paths to these leaves have length at most $24 \log n$, so the amount of comparisons is at most $n \cdot 24 \log n \in O(n \log n)$ with probability greater than $1 - \frac{1}{n}$ □