Randomized algorithms - Quick-Sort

March 13, 2002

1 Expected running time of Quick Sort

Let \( i_1, \ldots, i_n \) be the \( n \) given numbers.
Let \( a_1, \ldots, a_n \) be the \( n \) given numbers in their sorted order.

Observation 1.1 It is enough to bound the number of comparisons performed by quick sort to bound the running time of quick sort.

Observation 1.2 Two elements compared by quick sort exactly once (because comparisons happen against the pivot).

Definition 1.3 \( X_{ij} \) - random variable 1 if we compared \( a_i \) to \( a_j \), and zero otherwise.

Lemma 1.4 \( E \left[ X_{ij} \right] = \frac{2}{j-i+1} \).

Proof: By definition of expectation

\[
E \left[ X_{ij} \right] = 1 \ast Pr \left[ X_{ij} = 1 \right] + 0 \ast Pr \left[ X_{ij} = 0 \right] = Pr \left[ X_{ij} = 1 \right],
\]

namely, this is the probability that \( a_i \) is compared to \( a_j \). Namely, this is the probability that either \( a_i \) is picked as pivot and is being compared to \( a_j \) or the other way around (\( a_j \) is the pivot). Let us, as an exercise, run QS from the beginning, and track what happens to \( a_i \) and \( a_j \).

If QS pick a number smaller than \( a_i \) to be the pivot then \( a_i, a_j \) remain together in the next recursive call. And nothing interesting happens. Same if the number picked is larger than \( a_j \). If the pivot picked is between \( a_i \) and \( a_j \) then \( a_i \) goes into the left subproblem and \( a_j \) goes into the right subproblem. Namely, they would never be compared. Thus, in such a case \( X_{ij} = 0 \). Thus, \( X_{ij} = 1 \) only if among \( a_i, a_{i+1}, \ldots, a_j \) either \( a_i \) or \( a_j \) were the first to be picked as pivots. However, we pick the pivot uniformly from the range. Namely, the probability that \( a_i \) or \( a_j \) are the first to picked as pivots in this range, is \( \frac{2}{j-i+1} \). \( \blacksquare \)
We conclude:

\[
E \left[ \# \text{ of compares made by QS} \right] = E \left[ \sum_{i<j} X_{ij} \right] = \sum_{i<j} E \left[ X_{ij} \right] = \sum_{i<j} \frac{2}{j - i + 1} \\
= 2 \sum_{j=2}^{n} \sum_{i=1}^{j-1} \frac{1}{j - i + 1} = 2 \sum_{j=2}^{n} \sum_{l=2}^{j} \frac{1}{l} \leq 2 \sum_{j=2}^{n} \ln(j + 1) \\
\leq 2n \ln(n).
\]

We conclude that

\[
E \left[ \text{Running time QS} \right] = O(n \ln n).
\]

2 Better bounds

Theorem 2.1 [Markov Inequality] For a non-negative variable \( X \), and \( t > 0 \), we have:

\[
\Pr \left[ X \geq t \right] \leq \frac{E[X]}{t}.
\]

Proof: Assume that this is false, and there exists \( t_0 > 0 \) such that \( \Pr \left[ X \geq t_0 \right] > \frac{E[X]}{t_0} \). However,

\[
E[X] = \sum_{x} x \cdot \Pr \left[ X = x \right] \\
= \sum_{x < t_0} x \cdot \Pr \left[ X = x \right] + \sum_{x \geq t_0} x \cdot \Pr \left[ X = x \right] \geq 0 + t_0 \cdot \Pr \left[ X \geq t_0 \right] \\
> 0 + t_0 \cdot \frac{E[X]}{t_0} = E[X],
\]

a contradiction.

Definition 2.2 Variables \( X, Y \) are independent if for any \( x, y \) we have:

\[
\Pr \left[ (X = x) \cap (Y = y) \right] = \Pr \left[ X = x \right] \cdot \Pr \left[ Y = y \right].
\]

The following is easy to verify:

Claim 2.3 If \( X \) and \( Y \) are independent, then \( E[XY] = E[X]E[Y] \).

If \( X \) and \( Y \) are independent then \( Z = e^X, W = e^Y \) are also independent variables.

Lemma 2.4 The probability of throwing a fair coin \( N \) times and getting more than \((3/4)N\) heads is \( \leq 0.9^N \).
Proof: Let $X^i = 1$ if we got HEAD in the $i$-th flip, for $i = 1, \ldots, N$. Clearly, $X^1, \ldots, X^N$ are all independent variables. Let $X = \sum_{i=1}^N X^i$. Let $Y = \phi X$, and $Y_i = \phi X^i$, where $\phi$ is a constant to be determined shortly.

We have:

$$E[Y] = E[\phi X] = E[\phi \sum_{i=1}^N X^i] = E[\phi X^1 \cdot \phi X^2 \cdots \phi X^N]$$

$$= E[Y_1 \cdot Y_2 \cdots Y_N]$$

$$= E[Y_1] \cdots E[Y_N].$$

However, $Y_1, \ldots, Y_N$ have the same distribution (i.e., the same probability to be 1), thus:

$$E[Y] = \left( E[Y_1] \right)^N,$$

and $Y_1 = \phi X^1$

$$E[Y_1] = 1 \cdot \Pr[Y_1 = 1] + \phi \cdot \Pr[Y_1 = \phi] = 1 \cdot \frac{1}{2} + \phi \cdot \frac{1}{2} = \frac{1 + \phi}{2}.$$  

Namely, $Y = \phi X$

$$E[Y] = \left( \frac{1 + \phi}{2} \right)^N.$$  

Finally,

$$\Pr \left[ X \geq \frac{3}{4} N \right] = \Pr \left[ \phi X \geq \phi \left( \frac{3}{4} N \right) \right] = \Pr \left[ Y \geq \phi \left( \frac{3}{4} N \right) \right].$$

Set $t = \phi \left( \frac{3}{4} N \right)$, and use Markov’s inequality. We have:

$$\Pr \left[ X \geq \frac{3}{4} N \right] = \Pr \left[ Y \geq t \right] \leq \frac{E[Y]}{t} = \frac{\left( \frac{1 + \phi}{2} \right)^N}{t} = \left( \frac{1 + \phi}{2 \cdot \phi^{3/4}} \right)^N \leq 0.9^N,$$

since for $\phi = 3$, we have $\frac{1 + \phi}{2 \cdot 3^{3/4}} \leq 0.878$.  

This lemma is a special case of the Chernoff inequality.  

In particular, we now know that the probability of getting more than 300 heads when throwing a coin 400 times, is smaller than $0.9^{400} < 5 \cdot 10^{-19}$.

3 Nuts and Bolts with high probability

Let’s go back to the the nuts and bolts question. Consider a bolt $B$, and trace its history during the algorithm execution. Consider the first round of the algorithm, we split all the nuts and bolts into two groups, and continue recursively into each one of them. If the first pivot chosen by the algorithm was good, we set $X_1 = 1$, otherwise 0. Now, we look in
the recursive subproblem that contains $B$, and again we define $X_2 = 1$ if the pivot in this subproblem was good.

Namely, we trace $B$ in the recursion tree, and inside each node along this path, we define a variable $X_i$ if the pivot was good or not. This results in a sequence of random independent variables $X_1, X_2, \ldots, X_M$ (In the following, we assume that $M \geq 40 \log n$. If this is incorrect, we add fictitious variables $X_i = 1$ with probability $1/2$). We first note:

**Claim 3.1** A bolt might participate in at most $3.5 \log n$ successful rounds.

**Proof:** Indeed, every time $B$ passes through a node in the recursion with a good pivot, the size of the subproblem decreases by a factor of $3/4$. Since, $X_1, \ldots, X_M$ are independent, and the probability of each one of them to be 1 is $1/2$, it follows that we can apply the above Chernoff bound. Indeed, by the above lemma, the probability of $X = \sum_i X_i$ to be smaller than $M/4 = 10 \log n$ (which is equal to its probability to be larger than $(3/4)M$) is smaller than $0.9^M = 0.9^{40 \log n} \leq n^{-4}$.

What does this mean? It means that with probability $\geq 1 - n^{-4}$ a bolt participates in less than $M = O(\log n)$ rounds. In particular, the probability that any bolt participates in more than $M$ rounds, is smaller than $n^{-4} \cdot n = n^{-3}$. Namely, the overall running time of the algorithm is smaller than $O(n \log n)$ with probability $\geq 1 - n^{-3}$.

**Theorem 3.2** The algorithm for matching Nuts and Bolts runs in $O(n \log n)$ time, with probability $\geq 1 - 1/n^3$.

Note, that exactly the same analysis would work for QuickSort.