CS 5114
Solutions to Homework Assignment 2
Meghendra Singh

February 10, 2017

[30] 1. CLRS Problem 7-5. Median-of-3 partition

One way to improve the RANDOMIZED-QUICKSORT procedure is to partition around a pivot that is chosen more carefully than by picking a random element from the subarray. One common approach is the median-of-3 method: choose the pivot as the median (middle element) of a set of 3 elements randomly selected from the subarray. (See Exercise 7.4-6.) For this problem, let us assume that the elements in the input array $A[1...n]$ are distinct and that $n \geq 3$. We denote the sorted output array by $A'[1..n]$. Using the median-of-3 method to choose the pivot element x, define $p_i = Pr\{x = A'[i]\}$.

a. Give an exact formula for p_i as a function of n and i for $i = 2, 3, ...n-1$. (Note that $p_1 = p_n = 0$.)

b. By what amount have we increased the likelihood of choosing the pivot as $x = A'[\lfloor (n+1)/2 \rfloor]$, the median of $A[1..n]$, compared with the ordinary implementation? Assume that $n \to \infty$, and give the limiting ratio of these probabilities.

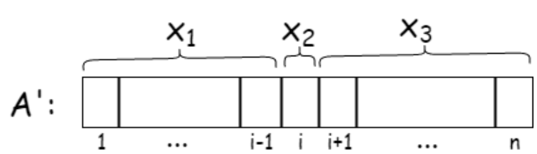
c. If we define a "good" split to mean choosing the pivot as $x = A'[i]$, where $n/3 \leq i \leq 2n/3$, by what amount have we increased the likelihood of getting a good split compared with the ordinary implementation? (*Hint:* Approximate the sum by an integral.)

d. Argue that in the $\Omega(n \lg n)$ running time of quicksort, the median-of-3 method affects only the constant factor.

---

a. Given: $A[1...n] \xrightarrow{sorted} A'[1...n]$, $n \geq 3$, x is the pivot, $p_i = Pr\{x = A'[i]\}$ and $p_1 = p_n = 0$.

Let $\langle x_1, x_2, x_3 \rangle$ be the three elements selected by median-of-3 method, hence x_2 will be chosen as the pivot, and its index in the sorted array A' would be i. We can choose $\langle x_1, x_2, x_3 \rangle$ in the following way:



Therefore,
$x_1 \in \{1...(i-1)\}$, $x_2 \in \{i\}$ & $x_3 \in \{(i+1)...n\}$

$\therefore Pr(x_1) = \frac{i-1}{n}, Pr(x_2) = \frac{1}{n-1}$ & $Pr(x_3) = \frac{n-i}{n-2}$

$$\begin{aligned} p_i &= Pr(\langle x_1, x_2, x_3 \rangle) \\ &= 3! * Pr(x_1) * Pr(x_2) * Pr(x_3) \\ &= 6 * (\frac{i-1}{n}) * (\frac{1}{n-1}) * (\frac{n-i}{n-2}) \\ p_i &= \frac{6(i-1)(n-i)}{n(n-1)(n-2)} \end{aligned} \tag{1}$$

b. Given: $x = A'[\lfloor \frac{(n+1)}{2} \rfloor] \therefore i = \frac{(n+1)}{2}$. Let $p_{ordinary}$ be the probability of selecting pivot as $x = A'[\lfloor \frac{(n+1)}{2} \rfloor]$ in the ordinary implementation, which is:

$$p_{ordinary} = \frac{1}{n} \tag{2}$$

Let $p_{median-of-3}$ be the probability of selecting pivot as $x = A'[\lfloor \frac{(n+1)}{2} \rfloor]$ in the median-of-3 method. Substituting $i = \frac{(n+1)}{2}$ in equation (1) above we have:

$$\begin{aligned} p_{median-of-3} &= \frac{6(\frac{n+1-2}{2})(\frac{2n-n-1}{2})}{n(n-1)(n-2)} \\ &= \frac{3}{2}\frac{(n-1)}{n(n-2)} \end{aligned}$$

Limiting ratio of the two probabilities is as follows:

$$\begin{aligned} \lim_{n\to\infty} \frac{p_{median-of-3}}{p_{ordinary}} &= \lim_{n\to\infty} \frac{\frac{3(n-1)}{2n(n-2)}}{\frac{1}{n}} \\ &= \lim_{n\to\infty} \frac{3(n-1)}{2(n-2)} \\ &= \frac{3}{2}(1) \\ &= \frac{3}{2} \end{aligned}$$

c. Let $p_{good-split}$ be the probability of choosing the pivot $x = A'[i]$ with respect to the definition of "good" split, i.e. $n/3 \le i \le 2n/3$. Using equation (1) we have:

$$\begin{aligned} p_{good-split} &= \sum_{i=\frac{n}{3}}^{\frac{2n}{3}} p_i \\ &= \sum_{i=\frac{n}{3}}^{\frac{2n}{3}} \frac{6(i-1)(n-i)}{n(n-1)(n-2)} \\ &= \frac{6}{n(n-1)(n-2)} \sum_{i=\frac{n}{3}}^{\frac{2n}{3}} (i-1)(n-i) \end{aligned}$$

Substituting $t = (i-1)$, and approximating the sum by integration, we have:

$$
\begin{aligned}
p_{good-split} &= \frac{6}{n(n-1)(n-2)} \int_{\frac{n}{3}-1}^{\frac{2n}{3}-1} t(n-t-1)\ dt \\
&= \frac{6}{n(n-1)(n-2)} \left[\frac{nt^2}{2} - \frac{t^3}{3} - \frac{t^2}{2} \right|_{\frac{n}{3}-1}^{\frac{2n}{3}-1} \\
&= \frac{6}{n(n-1)(n-2)} \left[t^2(\frac{n-1}{2}) - \frac{t^3}{3} \right|_{\frac{n}{3}-1}^{\frac{2n}{3}-1} \\
&= \frac{6}{n(n-1)(n-2)} \left[(\frac{2n}{3}-1)^2(\frac{n-1}{2}) - \frac{1}{3}(\frac{2n}{3}-1)^3 - \left((\frac{n}{3}-1)^2(\frac{n-1}{2}) - \frac{1}{3}(\frac{n}{3}-1)^3 \right) \right] \\
&= \frac{2}{3n(n-1)(n-2)} \left[\frac{13n^3 - 27n^2}{18} \right] \\
p_{good-split} &= \frac{1}{27} \left[\frac{13n^2 - 27n}{(n-1)(n-2)} \right] \qquad (3)
\end{aligned}
$$

Increase in likelihood of getting a good split as compared with the ordinary implementation can be given using the limiting difference between $p_{good-split}$ and $p_{ordinary}$. Therefore, using equations (2) and (3) we have:

$$
\begin{aligned}
\lim_{n\to\infty} p_{good-split} - p_{ordinary} &= \lim_{n\to\infty} \left\{ \frac{1}{27} \left[\frac{13n^2 - 27n}{(n-1)(n-2)} \right] - \frac{1}{n} \right\} \\
&= \lim_{n\to\infty} \left[\frac{13n^2 - 27n}{27n^2 - 81n + 54} \right] - \lim_{n\to\infty} \frac{1}{n} \\
&= \frac{13}{27} - 0 \\
&= \frac{13}{27}
\end{aligned}
$$

d. The median-of-3 operation involves random selection of 3 elements, and choosing the median of these 3 elements as the pivot for partition. This operation does not change with the number of inputs and hence only adds a constant factor c to the recurrence relation of quicksort, which can be defined as follows:

$$T(n) = 2T(\frac{n}{2}) + \Theta(n) + c$$

Here, the partition and median-of-3 operations add $\Theta(n) + c$ to the recurrence relation and do not impact the asymptotic running time complexity, which remains $\Theta(n \lg n)$. From ***Theorem 3.1*** of **CLRS**, we know that:

> For any two functions $f(n)$ and $g(n)$, we have $f(n) = \Theta(g(n))$ if and only if $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$.

Hence,

$$T(n) \in \Theta(n \lg n) \implies T(n) \in \Omega(n \lg n)$$

---

[30] 2. CLRS Exercise 9.3-8.
Let $X[1..n]$ and $Y[1..n]$ be two arrays, each containing n numbers already in sorted order. Give an $O(\lg n)$-time algorithm to find the median of all $2n$ elements in arrays X and Y.

The main idea to be used here is to divide the problem into two "almost" equal parts, by comparing the medians of the individual sorted arrays, and select one of these parts for further exploration (using recursion). If the median of $X[1..n]$ is less than the median of $Y[1..n]$ then the median of the combined $2n$ elements will be present in the later half of X and the initial half of Y when they are divided by the median (i.e. the median is present in $X[\frac{n}{2}..n]$ and $Y[1..\frac{n}{2}]$). Conversely if the median of $X[1..n]$ is greater than the median of $Y[1..n]$ then the median of the combined $2n$ elements will be present in $X[1..\frac{n}{2}]$ and $Y[\frac{n}{2}..n]$. The base case would be when the length of the two arrays reduces to 1, in which case we pick the smaller of the two elements present in the two arrays as the median. This is because we consider **Lower median** as the median in case of even number of elements in an array [Chapter 9, first paragraph of **CLRS**], and since we have two arrays here, we will always end up with $2n$ elements in the combined array (i.e. the total number of elements will always be even). The pseudo-code is as follows:

MEDIAN-OF-TWO(X, Y)

if $X.length == 1$
// Base case when $X.length = Y.length = 1$
    if $X[1] < Y[1]$
        return $X[1]$
    else
        return $Y[1]$
// i is the index of the medians of X and Y
$i = \left\lceil \frac{X.length}{2} \right\rceil$
if $X[i] < Y[i]$
    return MEDIAN-OF-TWO$(X[i+1 \,..\, X.length], Y[1 \,..\, i])$
else
    return MEDIAN-OF-TWO$(X[1..\, i], Y[i+1 \,..\, Y.length])$

The recurrence relation for MEDIAN-OF-TWO can be written as follows:

$$T(n) = T(\frac{n}{2}) + \Theta(c)$$

Here, a = 1, b = 2 and $f(n) = \Theta(c)$ for some constant c. Hence,

$$\begin{aligned} n^{\log_b a} &= n^{\log_2 1} \\ &= n^0 \approx c \end{aligned}$$

Since,

$$f(n) = \Theta(c)$$

Therefore, by Case 2 of Master Theorem, we obtain:

$$T(n) \in \Theta(c \lg n)$$

As, c is a constant we can write this as:

$$\begin{aligned} T(n) &\in \Theta(\lg n) \\ \implies T(n) &\in \Omega(\lg n) \\ also,\ T(n) &\in O(\lg n) \end{aligned}$$