Dual pivot Quicksort

Vasileios Iliopoulos and David B. Penman

Department of Mathematical Sciences
University of Essex, Wivenhoe Park
Colchester, CO4 3SQ, U.K.

Abstract

In this paper, we analyse the dual pivot Quicksort, a variant of the standard Quicksort algorithm, in which two pivots are used for the partitioning of the array. We are solving recurrences of the expected number of key comparisons and exchanges performed by the algorithm, obtaining the exact and asymptotic total average values contributing to its time complexity. Further, we compute the average number of partitioning stages and the variance of the number of key comparisons. In terms of mean values, dual pivot Quicksort does not appear to be faster than ordinary algorithm.

1 Introduction

Quicksort is a sorting algorithm with an extensive literature regarding its mathematical analysis and its applications. Without loss of generality, suppose that we want to quick sort a random permutation of distinct keys $\{1, 2, \ldots, n\}$ with all the $n!$ permutations equally likely. A key is randomly chosen as pivot and by pairwise comparisons of the other elements with it, smaller keys are placed to left and greater to right. Now, the pivot j is at its final position and the algorithm is recursively invoked to sort independently the left and right subarrays of $(j-1)$ and $(n-j)$ elements, respectively. Letting C_n being the number of comparisons for sorting n keys, its average value is given by the following recursive relation

$$\mathbb{E}(C_n) = n - 1 + \frac{1}{n} \sum_{j=1}^{n} \big(\mathbb{E}(C_{j-1}) + \mathbb{E}(C_{n-j}) \big)$$

$$= n - 1 + \frac{2}{n} \sum_{j=1}^{n} \mathbb{E}(C_{j-1}),$$

1

with initial condition $C_0 := 0$. Subtracting $(n-1)\mathbb{E}(C_{n-1})$ from $n\mathbb{E}(C_n)$ and telescoping, the average number of comparisons is

$$\mathbb{E}(C_n) = 2(n+1)H_n - 4n \sim 2n\ln(n). \tag{1.1}$$

Similarly, it is a routine matter to compute that the average number of exchanges performed is

$$\frac{2(n+1)H_n - 3n}{6} \sim \frac{1}{3}n\ln(n), \tag{1.2}$$

when at the first partitioning stage, the expected number of exchanges is

$$\frac{n}{6} + \frac{5}{6n}, \ [2].$$

Recall that H_n is the n_{th} harmonic number, defined by $H_n := \sum_{j=1}^{n} \frac{1}{j}$ and $H_0 := 0$. Further, the sign $\sim$ denotes asymptotic equivalence between two functions $f(n)$ and $g(n)$. That is $f(n) \sim g(n)$ if and only if $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$. (In [2], [4], it is suggested that small segments of size less than some parameter m be sorted by a simpler algorithm, such as insertion sort, as this is in practice quicker: in order to simplify the calculations for the solutions of the recurrences, we assume that $m = 0$).

2 Partitioning on two pivots

The idea for this variant is that we can randomly select two elements as pivots for the partitioning of the array. The number of comparisons obeys the following recursive rule;

$C_n = $ *"Number of comparisons during first partitioning stage"* $+ C_{i-1} + C_{j-i-1} + C_{n-j}$,

since at the beginning, the pivots are compared each other and swapped if they are not in order. If elements $i < j$ are selected, the array is partitioned into three subarrays: one with $(i-1)$ keys smaller than i, a subarray of $(j-i-1)$ keys between two pivots and the part of $(n-j)$ elements greater than j. The algorithm then is recursively applied to each of these subarrays. The number of comparisons during the first stage is

$$A_n = 1 + \Big((i-1) + 2(j-i-1) + 2(n-j) \Big), \quad i = 1, \ldots, n-1 \ \text{ and } \ j = i+1, \ldots, n,$$

because if an element is lower than i, then it is less than j automatically, so $i-1$ elements beneath i only need to be compared with one of the pivots. However if an element is greater than i then it needs to be compared with the other pivot as well, to determine whether or

not it is greater than j. We refer to Sedgewick [4] for code for a version of this scheme. The average value of A_n is

$$\frac{1}{\binom{n}{2}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(1 + (i-1) + 2(j-i-1) + 2(n-j) \right) = \frac{1}{\binom{n}{2}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(2n - i - 2 \right)$$

$$= \frac{2}{n(n-1)} \left(\frac{5}{6} n^3 - 2n^2 + \frac{7}{6} n \right) = \frac{5n-7}{3}.$$

Hence, the recurrence for the expected number of comparisons is

$$\mathbb{E}(C_n) = \frac{5n-7}{3} + \frac{2}{n(n-1)} \left(\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbb{E}(C_{i-1}) + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbb{E}(C_{j-i-1}) + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbb{E}(C_{n-j}) \right).$$

Note that the three double sums above are equal. Therefore, the recurrence becomes

$$\mathbb{E}(C_n) = \frac{5n-7}{3} + \frac{6}{n(n-1)} \sum_{i=1}^{n-1} (n-i) \mathbb{E}(C_{i-1}).$$

Letting $a_n = \mathbb{E}(C_n)$, we have

$$a_n = \frac{5n-7}{3} + \frac{6}{n(n-1)} \sum_{i=1}^{n-1} (n-i) a_{i-1}, \ n \geq 2.$$

Trivially, it holds that $a_0 = a_1 = 0$. Multiplying both sides by $\binom{n}{2}$, we obtain

$$\binom{n}{2} a_n = \binom{n}{2} \left(\frac{5n-7}{3} + \frac{6}{n(n-1)} \sum_{i=1}^{n-1} (n-i) a_{i-1} \right)$$

$$= \frac{n(n-1)(5n-7)}{6} + 3 \sum_{i=1}^{n-1} (n-i) a_{i-1}.$$

We introduce the difference operator Δ for the solution of this recurrence. The operator is defined by

$$\Delta F(n) := F(n+1) - F(n) \quad \text{and for higher orders}$$
$$\Delta^k F(n) := \Delta^{k-1} F(n+1) - \Delta^{k-1} F(n).$$

Thus, we have

$$\Delta \binom{n}{2} a_n = \binom{n+1}{2} a_{n+1} - \binom{n}{2} a_n = \frac{5n^2 - 3n}{2} + 3 \sum_{i=0}^{n-1} a_i$$

$$\Delta^2 \binom{n}{2} a_n = \Delta \binom{n+1}{2} a_{n+1} - \Delta \binom{n}{2} a_n = 5n + 1 + 3a_n.$$

3

By definition,

$$\Delta^2 \binom{n}{2} a_n = \Delta \binom{n+1}{2} a_{n+1} - \Delta \binom{n}{2} a_n$$
$$= \binom{n+2}{2} a_{n+2} - 2\binom{n+1}{2} a_{n+1} + \binom{n}{2} a_n$$

and the recurrence becomes

$$(n+1)(n+2)a_{n+2} - 2n(n+1)a_{n+1} + n(n-1)a_n = 2(5n+1+3a_n)$$
$$\Rightarrow (n+1)\big((n+2)a_{n+2} - (n-2)a_{n+1}\big) - (n+2)\big((n+1)a_{n+1} - (n-3)a_n\big) = 2(5n+1).$$

Dividing by $(n+1)(n+2)$, we obtain the telescoping recurrence

$$\frac{(n+2)a_{n+2} - (n-2)a_{n+1}}{n+2} = \frac{(n+1)a_{n+1} - (n-3)a_n}{n+1} + \frac{2(5n+1)}{(n+1)(n+2)},$$

which yields

$$\frac{(n+2)a_{n+2} - (n-2)a_{n+1}}{n+2} = 2\sum_{j=0}^{n} \frac{5j+1}{(j+1)(j+2)} = \frac{18}{n+2} + 10H_{n+1} - 18.$$

The recurrence is equivalent to

$$na_n - (n-4)a_{n-1} = 18 + 10nH_{n-1} - 18n.$$

Multiplying by $\dfrac{(n-1)(n-2)(n-3)}{24}$, this recurrence is transformed to a telescoping one [4],

$$\binom{n}{4} a_n = \binom{n-1}{4} a_{n-1} + \frac{18(n-1)(n-2)(n-3)}{24} + 10\binom{n}{4} H_{n-1} - 18\binom{n}{4}.$$

Unwinding, we have

$$\binom{n}{4} a_n = 18 \sum_{j=1}^{n} \frac{(j-1)(j-2)(j-3)}{24} + 10 \sum_{j=1}^{n} \binom{j}{4} H_{j-1} - 18 \sum_{j=1}^{n} \binom{j}{4}. \qquad (2.1)$$

Using MAPLE, we found that

$$\sum_{j=1}^{n} (j-1)(j-2)(j-3) = 6\binom{n}{4}$$

and for the other sums in Eq. (2.1),

$$\sum_{j=1}^{n} \binom{j}{4} = \binom{n+1}{5} \quad \text{and} \quad \sum_{j=1}^{n} \binom{j}{4} H_j = \binom{n+1}{5}\left(H_{n+1} - \frac{1}{5}\right).$$

Therefore

$$\sum_{j=1}^{n} \binom{j}{4} H_{j-1} = \sum_{j=1}^{n} \left(\binom{j}{4} \left(H_j - \frac{1}{j} \right) \right) = \sum_{j=1}^{n} \binom{j}{4} H_j - \sum_{j=1}^{n} \binom{j}{4} \frac{1}{j}$$

$$= \binom{n+1}{5} \left(H_{n+1} - \frac{1}{5} \right) - \frac{1}{24} \sum_{j=1}^{n} (j-1)(j-2)(j-3)$$

$$= \binom{n+1}{5} \left(H_{n+1} - \frac{1}{5} \right) - \binom{n}{4} \frac{1}{4}.$$

Now Eq. (2.1) becomes

$$\binom{n}{4} a_n = \frac{9}{2} \binom{n}{4} + 10 \left(\binom{n+1}{5} \left(H_{n+1} - \frac{1}{5} \right) - \frac{1}{4} \binom{n}{4} \right) - 18 \binom{n+1}{5}.$$

$$\Rightarrow a_n = \frac{9}{2} + 10 \left(\frac{n+1}{5} \left(H_{n+1} - \frac{1}{5} \right) - \frac{1}{4} \right) - \frac{18(n+1)}{5}.$$

Finally, the expected number of comparisons, when two pivots are chosen is

$$a_n = 2(n+1)H_n - 4n \sim 2n \ln(n). \tag{2.2}$$

This is exactly the same as the expected number of comparisons in Eq. (1.1) computed earlier for ordinary Quicksort. The dual pivot variant is claimed to be faster in experimental measurements than the standard algorithm in [6]. A referee of this article commented to us that the variant gives a 30% performance boost on randomly permuted data.

We proceed to compute the average number of exchanges. Letting S_n denote the total number of exchanges we carry out when sorting n objects, we have that

$$S_n = \text{``Number of exchanges during first partitioning stage''} + S_{i-1} + S_{j-i-1} + S_{n-j}.$$

Now it is fairly clear that, again using that the pivots are chosen uniformly at random, that the average values of last three quantities are equal. So our main objective now is to determine the average number of exchanges during the first partitioning stage. At the end of the partition routine, $(i-1)$ elements are less than pivot i. Thus, the contribution to the number of exchanges is [4],

$$\frac{1}{\binom{n}{2}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (i-1) = \frac{1}{\binom{n}{2}} \sum_{i=1}^{n-1} (n-i)(i-1) = \frac{1}{\binom{n}{2}} \left(\sum_{i=1}^{n-1} (n-i)i - \sum_{i=1}^{n-1} (n-i) \right).$$

The first sum being evaluated gives

$$\sum_{i=1}^{n-1} (n-i)i = \frac{n^3 - n}{6}$$

and the second is just $n(n-1)/2$. The average contribution is

$$\frac{2}{n(n-1)}\left(\frac{n^3-n}{6}-\frac{n(n-1)}{2}\right)=\frac{n-2}{3}.$$

Similarly, the average number of exchanges for the $(n-j)$ elements greater than the second pivot is $\left(\dfrac{n-2}{3}\right)$, since the double sums are equal. Adding the two final "exchanges" to get the pivots in place, the average number of exchanges during the partitioning routine is $\left(\dfrac{2(n+1)}{3}\right)$. Therefore, the recurrence for the mean number of exchanges in course of the algorithm is

$$\mathbb{E}(S_n)=\frac{2(n+1)}{3}+\frac{2}{n(n-1)}\left(\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\mathbb{E}(S_{i-1})+\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\mathbb{E}(S_{j-i-1})+\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\mathbb{E}(S_{n-j})\right)$$

$$=\frac{2(n+1)}{3}+\frac{6}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\mathbb{E}(S_{i-1})$$

$$=\frac{2(n+1)}{3}+\frac{6}{n(n-1)}\sum_{i=1}^{n-1}(n-i)\mathbb{E}(S_{i-1}).$$

This recurrence is solved in [4]: here we present a solution using generating functions. Letting $b_n=\mathbb{E}(S_n)$, we have

$$b_n=\frac{2(n+1)}{3}+\frac{6}{n(n-1)}\sum_{i=1}^{n}(n-i)b_{i-1}.$$

Multiplying by $\dbinom{n}{2}$ to clear fractions, we have

$$\binom{n}{2}b_n=\binom{n}{2}\left(\frac{2(n+1)}{3}+\frac{6}{n(n-1)}\sum_{i=1}^{n}(n-i)b_{i-1}\right)$$

$$=\frac{n(n-1)(n+1)}{3}+3\sum_{i=1}^{n}(n-i)b_{i-1}.$$

Multiplying by z^n and summing over n, in order to obtain the generating function $g(z)=$

$$\sum_{n=0}^{\infty} b_n z^n \text{ of the coefficients } b_n,$$

$$\sum_{n=0}^{\infty} \binom{n}{2} b_n z^n = \frac{1}{3} \sum_{n=0}^{\infty} n(n-1)(n+1)z^n + 3 \sum_{n=1}^{\infty} \sum_{i=1}^{n} (n-i)b_{i-1} z^n$$

$$\Rightarrow \frac{z^2}{2} \sum_{n=0}^{\infty} n(n-1)b_n z^{n-2} = \frac{z^2}{3} \sum_{n=0}^{\infty} n(n-1)(n+1)z^{n-2} + 3 \sum_{n=1}^{\infty} \sum_{i=1}^{n} (n-i)b_{i-1} z^n$$

$$\Rightarrow \frac{z^2}{2} \frac{\mathrm{d}^2 g(z)}{\mathrm{d} z^2} = \frac{z^2}{3} \frac{\mathrm{d}^3}{\mathrm{d} z^3} \left(\sum_{n=0}^{\infty} z^{n+1} \right) + 3 \sum_{n=1}^{\infty} \sum_{i=1}^{n} (n-i)b_{i-1} z^n.$$

The first term on the right-hand side of the previous equation is the third order derivative of the following geometric series

$$\sum_{n=0}^{\infty} z^{n+1} = \frac{z}{1-z} \quad \text{and} \quad \frac{\mathrm{d}^3}{\mathrm{d} z^3} \left(\frac{z}{1-z} \right) = \frac{6}{(1-z)^4}, \quad |z| < 1.$$

The double sum is equal to

$$\sum_{n=1}^{\infty} \sum_{i=1}^{n} (n-i)b_{i-1} z^n = b_0 z^2 + (2b_0 + b_1)z^3 + (3b_0 + 2b_1 + b_2)z^4 + \dots$$

$$= z^2(b_0 + b_1 z + b_2 z^2 + \dots) + 2z^3(b_0 + b_1 z + b_2 z^2 + \dots) + \dots$$
$$= (z^2 + 2z^3 + 3z^4 + \dots)g(z)$$
$$= \left(\sum_{n=0}^{\infty} nz^{n+1} \right) g(z).$$

The sum which multiplies $g(z)$ on the last line is

$$\sum_{n=0}^{\infty} nz^{n+1} = \left(\frac{z}{1-z} \right)^2.$$

Therefore, our recurrence is transformed to the following differential equation

$$\frac{z^2}{2} \frac{\mathrm{d}^2 g(z)}{\mathrm{d} z^2} = \frac{2z^2}{(1-z)^4} + 3g(z) \left(\frac{z}{1-z} \right)^2.$$

Changing variables $v = 1 - z$, we have that $f(v) = g(1-v)$. Thus, it holds

$$\frac{\mathrm{d}^k f(v)}{\mathrm{d} v^k} = (-1)^k \frac{\mathrm{d}^k g(1-v)}{\mathrm{d} v^k}.$$

The differential equation becomes

$$\frac{(1-v)^2}{2} \frac{\mathrm{d}^2 f(v)}{\mathrm{d} v^2} = \frac{2(1-v)^2}{v^4} + 3f(v) \left(\frac{1-v}{v} \right)^2.$$

Using MAPLE, the general solution is

$$f(v) = c_1 v^3 + \frac{c_2}{v^2} - \frac{20\ln(v) + 4}{25v^2}, \quad c_1, \ c_2 \in \mathbb{R}.$$

For the computation of constants, we consider the fact that $f(1) = g(0) = 0$ and $f'(1) = -g'(0) = 0$ (as $b_0 = b_1 = 0$). The resulting system of linear equations in c_1 and c_2 has solution $(c_1, c_2) = \left(\frac{4}{25}, \ 0 \right)$. Therefore, the function is

$$f(v) = \frac{4}{25}v^3 - \frac{20\ln(v) + 4}{25v^2}.$$

But, since

$$(1 - z)^{-2} = \sum_{n=0}^{\infty}(n+1)z^n \quad \text{and} \quad \frac{\ln(v)}{v^2} = -\frac{1}{(1-z)^2}\ln\left(\frac{1}{1-z}\right),$$

this can be written as a product of the following two series;

$$
\begin{aligned}
\frac{1}{(1-z)^2}\ln\left(\frac{1}{1-z}\right) &= \left(\sum_{n=1}^{\infty} n z^{n-1} \right) \left(\sum_{n=1}^{\infty} \frac{1}{n} z^n \right) \\
&= \left(1 + 2z + 3z^2 + 4z^3 + \dots \right) \left(z + \frac{z^2}{2} + \frac{z^3}{3} + \frac{z^4}{4} + \dots \right) \\
&= z + \left(2 + \frac{1}{2} \right)z^2 + \left(3 + \frac{2}{2} + \frac{1}{3} \right)z^3 + \left(4 + \frac{3}{2} + \frac{2}{3} + \frac{1}{4} \right)z^4 + \dots \\
&= z + \left(H_1 + H_2 \right)z^2 + \left(H_1 + H_2 + H_3 \right)z^3 + \dots \\
&= \sum_{n=0}^{\infty}\left((n+1)H_n - n \right)z^n.
\end{aligned}
$$

Extracting the coefficients and discarding terms for $n \le 3$, the exact mean number of exchanges of dual pivot Quicksort is equal to,

$$
\begin{aligned}
b_n &= \frac{4}{5}\left((n+1)H_n - n \right) - \frac{4}{25}(n+1) \\
&= \frac{4}{5}(n+1)H_n - \frac{24n+4}{25} \sim \frac{4}{5}n\ln(n). \tag{2.3}
\end{aligned}
$$

Comparing the expected number of comparisons of this variant with the standard algorithm, we see that they are identical. However, the mean number of exchanges is 2.4 times greater than the figure of normal Quicksort.

In the lines that follow, we compute the average number of partitioning stages $\mathbb{E}(P_n)$ of dual pivot Quicksort. The recurrence is much simpler;

$$P_n = 1 + P_{i-1} + P_{j-i-1} + P_{n-j}.$$

Averaging over all $\binom{n}{2}$ pairs of pivots i and j we have

$$\mathbb{E}(P_n) = 1 + \frac{6}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbb{E}(P_{i-1})$$

$$= 1 + \frac{6}{n(n-1)} \sum_{i=1}^{n-1} (n-i) \mathbb{E}(P_{i-1}),$$

since the sums are equal. Again, we use generating functions for the solution of the recurrence. Letting $f(z) = \sum_{n=0}^{\infty} \mathbb{E}(P_n) z^n$ our recurrence is transformed to the following differential equation

$$f''(z) \frac{z^2}{2} = \frac{z^2}{(1-z)^3} + 3f(z) \left(\frac{z}{1-z} \right)^2.$$

Changing variables $x = 1 - z$ we have $h(x) = f(1-x)$ and the general solution is

$$h(x) = \frac{c_2}{x^2} + x^3 c_1 - \frac{1}{2x}.$$

Since $h(1) = h'(1) = 0$ the constants are $c_1 = \frac{1}{10}$ and $c_2 = \frac{2}{5}$. Consequently, the mean number of partitioning stages is found to be equal to

$$\mathbb{E}(P_n) = \frac{2}{5}(n+1) - \frac{1}{2}. \tag{2.4}$$

This is smaller than the expected number of stages of ordinary Quicksort, which is n, when there is no switch to straight insertion for the sorting of small subfiles, [4].

3   Variance

Finally, we set up a recurrence for the computation of the variance of the number of comparisons of dual pivot Quicksort. Recall that

$$A_n = 2n - i - 2 \quad \text{and} \quad \mathbb{E}(A_n) = \frac{5n-7}{3}.$$

By the recurrence relation for the number of comparisons, we have

$$P(C_n = t) = \frac{1}{\binom{n}{2}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} P(C_n = t) = \frac{1}{\binom{n}{2}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} P(A_n + C_{i-1} + C_{j-i-1} + C_{n-j} = t),$$

9

noting that the resulting subarrays are independently sorted, the above is

$$\frac{1}{\binom{n}{2}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \sum_{l,m} \left(P(C_{i-1} = l)P(C_{j-i-1} = m)P(C_{n-j} = t - m - l - 2n + i + 2) \right).$$

Letting $f_n(z) = \sum_{t=0}^{\infty} P(C_n = t)z^t$ be the ordinary probability generating function for the number of comparisons needed to sort n keys, we obtain

$$f_n(z) = \frac{1}{\binom{n}{2}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} z^{2n-i-2} f_{i-1}(z) f_{j-i-1}(z) f_{n-j}(z). \tag{3.1}$$

It holds that $f_n(1) = 1$ and $f_n'(1) = 2(n+1)H_n - 4n$. Moreover, the second order derivative of Eq. (3.1) evaluated at $z = 1$ is recursively given by

$$\begin{aligned}
f_n''(1) &= \frac{2}{n(n-1)} \Bigg(\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (2n-i-2)^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (2n-i-2) \\
&+ 2\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (2n-i-2)E(C_{i-1}) + 2\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (2n-i-2)E(C_{j-i-1}) \\
&+ 2\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (2n-i-2)E(C_{n-j}) + 2\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E(C_{i-1})E(C_{j-i-1}) \\
&+ 2\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E(C_{i-1})E(C_{n-j}) + 2\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E(C_{j-i-1})E(C_{n-j}) \\
&+ \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} f_{i-1}''(1) + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} f_{j-i-1}''(1) + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} f_{n-j}''(1) \Bigg).
\end{aligned}$$

The reader should not be discouraged by this long expression, since many of the sums are equal. Specifically, the fourth and fifth turn out to be equal and by simple manipulation of indices, the sums of the products of expected values are equal. The double sum of the product of the mean number of comparisons can be simplified as follows:

$$\begin{aligned}
\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E(C_{i-1})E(C_{n-j}) &= \sum_{i=1}^{n-1} \left(E(C_{i-1}) \left(\sum_{j=0}^{n-i-1} E(C_j) \right) \right) \\
&= \sum_{i=1}^{n-1} \left(\left(2iH_{i-1} - 4(i-1) \right) \left(2\binom{n-i+1}{2} H_{n-i} + \frac{n-i-5(n-i)^2}{2} \right) \right).
\end{aligned}$$

10

The next sum was computed using results from a paper [5], which contains interesting identities and properties of sums involving harmonic numbers.

$$\sum_{i=1}^{n-1} i\binom{n-i+1}{2}H_{i-1}H_{n-i} = \sum_{i=1}^{n-1}\left((i-1)+1\right)\binom{n-i+1}{2}H_{i-1}H_{n-i}$$

$$= \sum_{i=1}^{n-1}(i-1)\binom{n-i}{2}H_{i-1}H_{n-i} + \sum_{i=1}^{n-1}\binom{n-i}{2}H_{i-1}H_{n-i}$$

$$+ \sum_{i=1}^{n-1}(i-1)(n-i)H_{i-1}H_{n-i} + \sum_{i=1}^{n-1}(n-i)H_{i-1}H_{n-i}.$$

The four sums can be evaluated using Corollary 3 in [5].

After some computations in our MAPLE worksheet[1], the recurrence is

$$f_n''(1) = 2(n+1)(n+2)(H_n^2 - H_n^{(2)}) - H_n\left(\frac{17}{3}n^2 + \frac{47}{3}n + 6\right) + \frac{209}{36}n^2 + \frac{731}{36}n + \frac{13}{6}$$

$$+ \frac{6}{n(n-1)}\sum_{i=1}^{n-1}(n-i)f_{i-1}''(1),$$

where $H_n^{(2)}$ is the second order harmonic number defined by $H_n^{(2)} := \sum_{k=1}^{n}\frac{1}{k^2}$. Letting $d_n = f_n''(1)$ and subtracting $\binom{n}{2}d_n$ from $\binom{n+1}{2}d_{n+1}$, we have

$$\Delta\binom{n}{2}d_n = 4n(n+1)(n+2)(H_n^2 - H_n^{(2)}) - \frac{nH_n}{9}(84n^2 + 198n + 42) + 3\sum_{i=1}^{n}d_{i-1}$$

$$+ \frac{n}{9}(79n^2 + 231n + 14),$$

using the identity [4]

$$H_{n+1}^2 - H_{n+1}^{(2)} = H_n^2 - H_n^{(2)} + \frac{2H_n}{n+1}.$$

Further, it holds that

$$\Delta^2\binom{n}{2}d_n = 12(n+1)(n+2)(H_n^2 - H_n^{(2)}) - H_n(20n^2 + 32n - 12) + 17n^2 + 37n + 3d_n.$$

The previous equation is the same as

$$\binom{n+2}{2}d_{n+2} - 2\binom{n+1}{2}d_{n+1} + \binom{n}{2}d_n$$

---

[1]The MAPLE worksheet containing the computations for the variance can be found at the web page: http://www.essex.ac.uk/maths/staff/profile.aspx?ID=1326

and our recurrence becomes

$$(n+1)(n+2)d_{n+2} - 2n(n+1)d_{n+1} + n(n-1)d_n$$
$$= 2\left(12(n+1)(n+2)(H_n^2 - H_n^{(2)}) - H_n(20n^2 + 32n - 12) + 17n^2 + 37n + 3d_n\right).$$

Dividing by $(n+1)(n+2)$, we obtain the telescoping recurrence

$$\frac{(n+2)d_{n+2} - (n-2)d_{n+1}}{n+2} = \frac{(n+1)d_{n+1} - (n-3)d_n}{n+1}$$
$$+ 2\left(12(H_n^2 - H_n^{(2)}) - \frac{H_n(20n^2 + 32n - 12)}{(n+1)(n+2)} + \frac{17n^2 + 37n}{(n+1)(n+2)}\right)$$

with solution

$$(n+2)d_{n+2} - (n-2)d_{n+1} = (24n^2 + 100n + 104)(H_{n+1}^2 - H_{n+1}^{(2)})$$
$$- H_{n+1}(88n^2 + 292n + 224) + 122n^2 + 346n + 224,$$

which is equivalent to

$$nd_n - (n-4)d_{n-1} = (24n^2 + 4n)(H_{n-1}^2 - H_{n-1}^{(2)})$$
$$- H_{n-1}(88n^2 - 60n - 8) + 122n^2 - 142n + 20.$$

Again as before, multiplying both sides by $\dfrac{(n-1)(n-2)(n-3)}{24}$, the recurrence telescopes with solution

$$f_n''(1) = 4(n+1)^2(H_{n+1}^2 - H_{n+1}^{(2)}) - 4H_{n+1}(n+1)(4n+3) + 23n^2 + 33n + 12.$$

Using the well known fact that

$$\mathrm{Var}(C_n) = f_n''(1) + f_n'(1) - \left(f_n'(1)\right)^2,$$

the variance of the number of key comparisons of dual pivot Quicksort is

$$7n^2 - 4(n+1)^2 H_n^{(2)} - 2(n+1)H_n + 13n. \tag{3.2}$$

The asymptotic figure is

$$\left(7 - \frac{2}{3}\pi^2\right)n^2 - 2n\ln(n) + O(n). \tag{3.3}$$

Note that the variance of dual pivot Quicksort is identical with the variance of the ordinary algorithm - see Eq. (32) in [3]. Also, in this paper we showed that the dual

pivot Quicksort variant has the same expected number of key comparisons as the standard algorithm and as one might expect, the mean number of stages is smaller than the respective figure of the ordinary algorithm. However, the expected number of exchanges is notably large. An efficient partitioning procedure is described in a paper written by Frazer and McKellar [1], where they present and analyse the Samplesort algorithm. It is shown [1] that the expected number of comparisons of Samplesort slowly approaches the Information - theoretic lower bound.

References

[1] W. D. Frazer and A. C. McKellar, Samplesort: A Sampling Approach to Minimal Storage Tree Sorting, *J. ACM.* **17**, 3 (1970) pp. 496-507.

[2] C. A. R. Hoare, Quicksort, *Comp. J.* **5**, 1 (1962) pp. 10-16.

[3] C. Knessl and W. Szpankowski, Quicksort algorithm again revisited, *Disc. Math. & Theor. Comp. Sci.* **3** (1999) pp. 43-64.

[4] R. Sedgewick, Quicksort, *Ph.D thesis* (1980) Garland Publishing.

[5] J. Spieß, Some Identities involving Harmonic numbers, *Math. Comp.* **55**, 192 (1990) pp. 839-863.

[6] V. Yaroslavskiy, Replacement of Quicksort in java.util.Arrays with new Dual-Pivot Quicksort (2009), web page:
`http://permalink.gmane.org/gmane.comp.java.openjdk.core-libs.devel/2628`