Combinatorics, Probability and Computing (1998) 7, 111–120. Printed in the United Kingdom
© 1998 Cambridge University Press

Comparisons in Hoare's Find Algorithm

PETER KIRSCHENHOFER and HELMUT PRODINGER

Institut für Algebra und Diskrete Mathematik, Technical University of Vienna,
Wiedner Hauptstrasse 8–10, A-1040 Vienna, Austria
(e-mail: {Peter.Kirschenhofer}{Helmut.Prodinger}@tuwien.ac.at)

Received 22 January 1996; revised 4 September 1996

We study the number of comparisons in Hoare's Find algorithm. Using trivariate generating functions, we get an explicit expression for the variance of the number of comparisons, if we search for the jth element in a random permutation of n elements. The variance is also asymptotically evaluated under the assumption that j is proportional to n. Similar results for the number of passes (recursive calls) are given, too.

1. Introduction

The traditional sorting method Quicksort [11, 6] uses a pivot element and a partitioning strategy to bring the pivot element into its correct position and having only smaller elements to the left and larger elements to the right, so that the procedure can be applied recursively to the smaller subfiles. There is a relation between Quicksort and binary search trees, as can be seen in the example in Figure 1. Subfiles are indicated by rectangular boxes, and the leftmost element will be the pivot element.

The corresponding binary search tree is depicted in Figure 2.

If one only wants to find the jth ranked element (which is j, since we assume that the elements in question are $\{1, 2, \ldots, n\}$), one uses the same partitioning strategy, but follows only the path which contains the sought element. This is the same as to say that one goes down recursively in only *one* subfile. This procedure is called Hoare's Find algorithm [5].

There is some renewed interest in Hoare's Find algorithm [12, 14, 8, 15]. For some general background on binary search trees, see Mahmoud [13] and Hwang [7].

Knuth [10] has already computed the average number of comparisons $C_{n,j}$. For this, it is assumed that every permutation of $\{1, 2, \ldots, n\}$ is equally likely, and that the partitioning phase needs $n-1$ comparisons. Then there is the recursion

$$C_{n,j} = n - 1 + \frac{1}{n} \sum_{1 \leqslant k < j} C_{n-k,j-k} + \frac{1}{n} \sum_{j < k \leqslant n} C_{k-1,j}. \tag{1.1}$$

https://doi.org/10.1017/S0963548397003325 Published online by Cambridge University Press

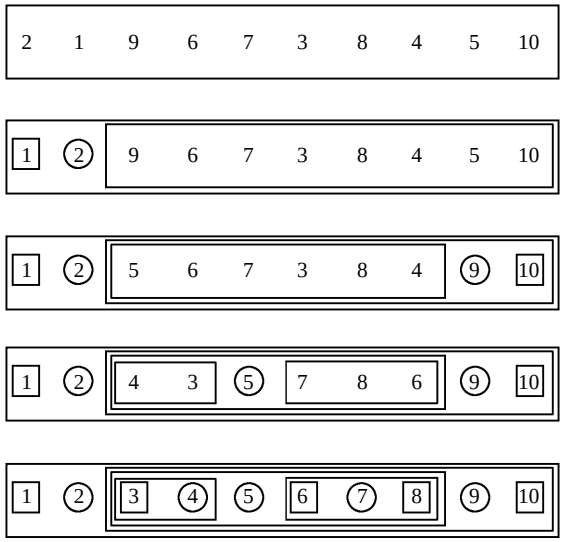


Figure 1 The partition process.

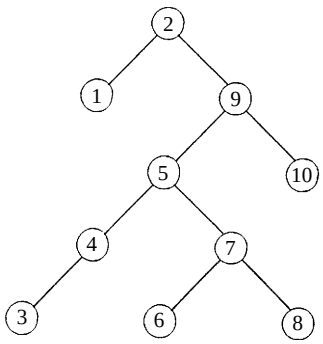


Figure 2 The corresponding binary search tree with 10 nodes.

The solution is

$$C_{n,j} = 2\left(n+3+(n+1)H_n-(j+2)H_j-(n+3-j)H_{n+1-j}\right). \tag{1.2}$$

Here, $H_n = \sum_{1\leqslant k\leqslant n}\frac{1}{k}$ is the nth harmonic number; we will also need harmonic numbers of order two, $H_n^{(2)} = \sum_{1\leqslant k\leqslant n}(1/k^2)$.

More generally, one can also set up an analogous recursion for the *probability generating functions*, namely

$$\mathscr{C}_{n,j}(v) = \frac{v^{n-1}}{n}\left[\sum_{1\leqslant k<j}\mathscr{C}_{n-k,j-k}(v)+1+\sum_{j<k\leqslant n}\mathscr{C}_{k-1,j}(v)\right]. \tag{1.3}$$

In this paper, we want to compute the *variance* of the number of comparisons. This requires an incredible number of computations (by Maple, with careful human guidance). We cannot present all of the details, but rather present the cornerstones of the whole procedure. We have to deal a lot with harmonic numbers, and use many identities about

https://doi.org/10.1017/S0963548397003325 Published online by Cambridge University Press

them which can be found, for example, in Knuth [9], Graham, Knuth and Patashnik [3] and Greene and Knuth [4].

2. Analysis

Following a suggestion of Philippe Flajolet, we prefer to think about binary search trees (as explained above) and use a trivariate generating function.

The coefficient of $z^n u^j v^m$ in $F(z;u,v)$ is the probability that it takes m comparisons to find node j, under the usual assumption that all permutations of n elements are equally likely. For example, the tree in Figure 2 (obtained by permutation in Figure 1) contributes

$$\frac{z^{10}}{10!}\left(u^1v^9+u^2v^9+u^3v^{22}+u^4v^{22}+u^5v^{21}+u^6v^{23}+u^7v^{23}+u^8v^{23}+u^9v^{16}+u^{10}v^{16}\right). \quad (2.1)$$

Recursion (1.3) translates into the fundamental equation

$$\frac{\partial}{\partial z}F(z;u,v)=\frac{F(zv;u,v)}{1-zv}+\frac{u}{(1-zv)(1-zuv)}+\frac{u\,F(zv;u,v)}{1-zuv}. \quad (2.2)$$

We can see this by multiplying the recursion (1.3) by $nz^{n-1}u^j$ and summing over n and j. However, with a little experience, one can deduce it directly from the underlying binary search trees.

When speaking of differential equations, we will assume that z is the variable, whereas u and v are parameters. Note the special value

$$F(z;u,1)=\frac{zu}{(1-z)(1-zu)}=\sum_{1\leqslant j\leqslant n}u^jz^n. \quad (2.3)$$

If we differentiate the fundamental equation (2.2) w.r.t. v, and set $v=1$, we get a differential equation for the generating function $G(z,u)$ of the average values $C_{n,j}$, which Maple is able to solve†, subject to the initial condition $G(0,u)=0$. It is

$$G(z,u)=\frac{2}{(1-u)(1-z)(1-zu)}\left[(u-2)\log\frac{1}{1-zu}+u(2u-1)\log\frac{1}{1-z}\right]+\frac{2zu(z^2u-2zu+3-2z)}{(1-z)^2(1-zu)^2}, \quad (2.4)$$

which gives the coefficients $C_{n,j}$. (Knuth derived them using an elementary method.)

If we differentiate the fundamental equation (2.2) twice w.r.t. v, and set $v=1$, and use the already computed function $G(z,u)$, we get a differential equation for the generating function $D(z,u)$ of the second factorial moments $D_{n,j}$. Its solution is

$$D(z,u)=\frac{4(2-u)^2}{(1-z)(1-zu)(1-u)^2}\log^2\frac{1}{1-zu}+\frac{4u(1-2u)^2}{(1-z)(1-zu)(1-u)^2}\log^2\frac{1}{1-z}-\frac{z^2u(-10z^3u^2-5z^2u^2+6zu^2-15u-5z^2u+38zu-15+6z)}{(1-z)^3(1-zu)^3}$$

† We cannot refrain from mentioning a peculiarity of Maple. It frequently replaces $\log(1-z)$ by $\log(z-1)+i\pi$, *etc*. To avoid this unpleasant feature, we had to replace z by $-z$ and u by $-u$.

https://doi.org/10.1017/S0963548397003325 Published online by Cambridge University Press

$$
\begin{aligned}
&+\frac{8u(2-u)(1-2u)}{(1-z)(1-zu)(1-u)^2}\log\frac{1}{1-zu}\log\frac{1}{1-u}\\
&+\frac{8(2-u)(1-2u)}{(1-z)(1-zu)(1-u)^2}\log\frac{1}{1-z}\log\frac{1}{1-u}\\
&+\frac{8u(2-u)(1-2u)}{(1-z)(1-zu)(1-u)^2}\log\frac{1}{1-zu}\log(-u)\\
&-\frac{2u(-5z^2u+19z^2u^2-11zu^2-10zu+z+3u+3)}{(1-u)(1-z)^2(1-zu)^2}\log\frac{1}{1-z}\\
&-\frac{10z^2u^2-38z^2u-2zu^2+20zu+22z-6u-6}{(1-u)(1-z)^2(1-zu)^2}\log\frac{1}{1-zu}\\
&-\frac{8u(2-u)(1-2u)}{(1-z)(1-zu)(1-u)^2}\left(\operatorname{dilog}\left(-\frac{(1-z)u}{1-u}\right)-\operatorname{dilog}\left(-\frac{u}{1-u}\right)\right)\\
&-\frac{8(2-u)(1-2u)}{(1-z)(1-zu)(1-u)^2}\left(\operatorname{dilog}\left(\frac{1-zu}{1-u}\right)-\operatorname{dilog}\left(\frac{1}{1-u}\right)\right). \qquad (2.5)
\end{aligned}
$$

Let us recall the definition of the dilogarithm [1],

$$\operatorname{dilog} z=\int_1^z\frac{\log t}{1-t}\,dt. \qquad (2.6)$$

The appearance of a term $\log(-u)$ looks strange at first sight, but it compensates with the dilog terms.

To get the coefficients, we rewrite $D(z,u)$ as follows (with heavy use of Maple):

$$
\begin{aligned}
D(z,u)=&\frac{9}{(1-z)^3}\sum_{j\geqslant1}u^j\\
&+\log^2\frac{1}{1-z}\sum_{j\geqslant1}u^j\left\{\frac{4(j-4)}{(1-z)^2}+\frac{4(z^{j-3}-z)}{(1-z)^3}+\frac{4z^{j-3}(-1-2z+z^2)}{1-z}\right\}\\
&+\frac{1}{(1-z)^2}\sum_{j\geqslant1}u^j\{-8(j+2)H_j+21j-25\}\\
&+\frac{1}{(1-z)^2}\log\frac{1}{1-z}\sum_{j\geqslant2}u^j\Bigg\{(8j-16H_{j-1}-26)z^j\\
&\quad+\left(-16j+70+24H_{j-1}-\frac{8}{j-1}\right)z^{j-1}\\
&\quad+\sum_{i=2}^{j-1}\left(8H_{j-i}-\frac{8i}{j-i}+16-\frac{16}{i+1}+\frac{40}{i}-\frac{16}{i-1}\right)z^{j-i}\\
&\quad+\left(-8H_j+\frac{24}{j}-12-\frac{16}{j-1}\right)\Bigg\}\\
&+\frac{1}{(1-z)^2}\log\frac{1}{1-z}\cdot u(10-18z)
\end{aligned}
$$

https://doi.org/10.1017/S0963548397003325 Published online by Cambridge University Press

$$+\frac{1}{1-z}\sum_{j\geqslant 1}u^j\left\{\left(2j^2+18j+16\right)H_j^2+\left(-6j^2-14j+14-\frac{16}{j}\right)H_j\right.$$
$$\left.-\left(2j^2+10j+16\right)H_j^{(2)}-2j^2+25j+10\right\}. \quad (2.7)$$

The most complicated term in this expansion is the $1/(1-z)$ term

$$8\frac{(2-u)(1-2u)}{(1-u)^2}\operatorname{dilog}\frac{1}{1-u}+4\frac{(2-u)^2}{(1-u)^3}\log^2\frac{1}{1-u}$$
$$+2\frac{11-u}{(1-u)^2}\log\frac{1}{1-u}+\frac{u\left(13-41u+10u^2\right)}{(1-u)^3}. \quad (2.8)$$

Reading off the coefficient of u^j is pretty much standard, except for the following:

$$\left[u^j\right]8\frac{(2-u)(1-2u)}{(1-u)^2}\operatorname{dilog}\frac{1}{1-u}=4jH_j^2-8\frac{j^2+j+2}{j}H_j+4jH_j^{(2)}+8j. \quad (2.9)$$

For this one we argue as follows. Since

$$\frac{d}{du}\operatorname{dilog}\frac{1}{1-u}=\left(-\frac{1}{1-u}-\frac{1}{u}\right)\log\frac{1}{1-u}, \quad (2.10)$$

we find

$$\left[u^j\right]\operatorname{dilog}\frac{1}{1-u}=-\frac{1}{j}\left(H_{j-1}+\frac{1}{j}\right). \quad (2.11)$$

The convolution, caused by the factor $\frac{1}{(1-u)^2}$, is then handled by formulæ in Knuth [9].

Now we read off the coefficient of u^jz^n in $D(z,u)$, as shown in (2.7). Using all kinds of identities for harmonic numbers (see Knuth [9]), we get as an intermediary result

$$\begin{aligned}
D_{n,j}=&-2(n+1)(n+8-2j)H_n^2-8(n+1)H_nH_j\\
&+\frac{\left(4-32n-12n^2\right)j^2+\left(44n^2+12n^3+28n-4\right)j+16n+16}{j(n+1-j)}H_n\\
&+2(j+8)(j+1)H_j^2+\left(-4j^2+(12+8n)j-4n^2+8-12n\right)H_jH_{n+1-j}\\
&-2\frac{(-6n-14)j^3+\left(14n+17+6n^2\right)j^2+\left(-3+6n^2+3n\right)j+8n+8}{j(n+1-j)}H_j\\
&+\left(2j^2+(-4n-18)j+2n^2+18n+32\right)H_{n+1-j}^2\\
&-2\frac{(6n+14)j^3+\left(-12n^2-25-46n\right)j^2+\left(37n+11+32n^2+6n^3\right)j+8n+8}{j(n+1-j)}H_{n+1-j}\\
&+2(n+1)(n+8-2j)H_n^{(2)}-\left(2j^2+16+10j\right)H_j^{(2)}\\
&+\left(-2n^2+4nj-18n-2j^2-32+18j\right)H_{n+1-j}^{(2)}\\
&+\frac{10j^4+(-20n-20)j^3+\left(n^2-66-13n\right)j^2+\left(76+9n^3+42n^2+101n\right)j+32}{2j(n+1-j)}\\
&+4(n+2-2j)(n+1)\sum_{k=1}^{j}\frac{H_{n-k}}{k}+8(n+1-j)\sum_{k=1}^{j}\frac{H_{n+k-j}}{k}. \quad (2.12)
\end{aligned}$$

https://doi.org/10.1017/S0963548397003325 Published online by Cambridge University Press

In this representation, there appear two sums, which (most likely) cannot be simplified any further. Also, from this representation, the symmetry $D_{n,j} = D_{n,n+1-j}$ is not at all obvious. Therefore, we derived the following lemmata.

Lemma 2.1 (Reciprocity law).

$$\sum_{k=1}^{j} \frac{H_{n-k}}{k} + \sum_{k=1}^{n+1-j} \frac{H_{n-k}}{k} = H_j H_{n+1-j} + H_n^2 - H_n^{(2)} - \frac{1}{j(n+1-j)}. \tag{2.13}$$

Lemma 2.2 (Reciprocity law).

$$\sum_{k=1}^{j} \frac{H_{n+k-j}}{k} + \sum_{k=1}^{n+1-j} \frac{H_{j+k-1}}{k} = \frac{1}{2}\left(H_j^2 + H_j^{(2)}\right) + \frac{1}{2}\left(H_{n+1-j}^2 + H_{n+1-j}^{(2)}\right) + H_j H_{n+1-j} + \frac{1}{j(n+1-j)} + \frac{n+1}{j(n+1-j)}\left(H_n - H_j - H_{n+1-j}\right). \tag{2.14}$$

And here is the formula relating the two types of sums.

Lemma 2.3.

$$\sum_{k=1}^{j} \frac{H_{n-k}}{k} + \sum_{k=1}^{j} \frac{H_{n+k-j}}{k} = \frac{1}{n+1-j}\left(H_n - H_j - H_{n+1-j}\right) + H_j H_n + H_j H_{n+1-j} - H_n H_{n+1-j} + \frac{1}{2}\left(H_n^2 - H_n^{(2)}\right) + \frac{1}{2}\left(H_{n+1-j}^2 + H_{n+1-j}^{(2)}\right). \tag{2.15}$$

All three lemmata can be easily proved by setting up bivariate generating functions on both sides. The formulæ are 'nice' and are easier to *prove* than to *find*. As an example, we give the proof of Lemma 2.1. In terms of generating functions we have to prove that

$$\begin{aligned}
&\frac{1}{(1-u)(1-z)} \log\frac{1}{1-uz} \log\frac{1}{1-z} - \frac{u}{(1-u)(1-uz)} \log^2\frac{1}{1-uz} \\
&\quad + \frac{u}{(1-1/u)(1-uz)} \log\frac{1}{1-z} \log\frac{1}{1-uz} - \frac{1}{(1-1/u)(1-z)} \log^2\frac{1}{1-z} \\
&= \frac{1}{z(1-uz)(1-z)} \log\frac{1}{1-uz} \log\frac{1}{1-z} + \frac{u}{(1-u)(1-z)} \log^2\frac{1}{1-z} \\
&\quad - \frac{u}{(1-u)(1-uz)} \log^2\frac{1}{1-uz} - \frac{1}{z} \log\frac{1}{1-z} \log\frac{1}{1-uz},
\end{aligned} \tag{2.16}$$

which is easily seen to be true.

Using the lemmata, we can do the following:

$$(n+1-2j)\sum_{k=1}^{j} \frac{H_{n-k}}{k} = (n+1-j)\sum_{k=1}^{j} \frac{H_{n-k}}{k} + j\sum_{k=1}^{n+1-j} \frac{H_{n-k}}{k} - j\left(H_j H_{n+1-j} + H_n^2 - H_n^{(2)} - \frac{1}{j(n+1-j)}\right). \tag{2.17}$$

https://doi.org/10.1017/S0963548397003325 Published online by Cambridge University Press

This gives us the final result on the second factorial moments.

Proposition 2.1.

$$\begin{aligned}
D_{n,j} = & -2(n+1)(n+6)\,H_n^2 - 8H_n\left(jH_j + (n+1-j)\,H_{n+1-j}\right)\\
& + 4\frac{-\left(3n^2+8n+1\right)j^2 + (n+1)\left(3n^2+8n+1\right)j + 4(n+1)}{j(n+1-j)}\,H_n\\
& + 2(j+8)(j+1)H_j^2 + 2(n+9-j)(n+2-j)\,H_{n+1-j}^2\\
& + 4\left(-j^2+(n+1)j-n^2-n+4\right)H_j\,H_{n+1-j}\\
& - \frac{2}{j(n+1-j)}\left(-2(3n+7)j^3+\left(6n^2+14n+13\right)j^2+(n+1)(6n+1)j+8(n+1)\right)H_j\\
& - \frac{2}{j(n+1-j)}\Big(2(3n+7)j^3-\left(12n^2+46n+29\right)j^2\\
& \quad +(n+1)(6n^2+26n+15)j+8(n+1)\Big)\,H_{n+1-j}\\
& + 2(n+1)(n+6)\}\,H_n^{(2)}\\
& - 2\left(j^2+5j+8\right)\,H_j^{(2)} - 2\left(j^2-(2n+7)j+n^2+7n+14\right)\,H_{n+1-j}^{(2)}\\
& + \frac{10j^4-20(n+1)j^3+\left(n^2-13n-66\right)j^2+\left(9n^2+33n+76\right)(n+1)j+32}{2j(n+1-j)}\\
& + 4n\left((n+1-j)\sum_{k=1}^{j}\frac{H_{n-k}}{k} + j\sum_{k=1}^{n+1-j}\frac{H_{n-k}}{k}\right). \qquad (2.18)
\end{aligned}$$

Computing $\mathrm{Var}_{n,j} = D_{n,j} + C_{n,j} - (C_{n,j})^2$, we find the variance.

Theorem 2.1. *The variance of the number of comparisons needed to find the jth element out of n elements is given by the exact formula*

$$\begin{aligned}
\mathrm{Var}_{n,j} = & -2(n+1)(3n+8)\,H_n^2\\
& + 8\,H_n\,H_j\,(n(j+2)+2) + 8\,H_n\,H_{n+1-j}\,(n(n+3-j)+2)\\
& + 2\frac{-\left(2n^2+n-9\right)j^2+(n+1)\left(2n^2+n-9\right)j+8(n+1)}{j(n+1-j)}\,H_n\\
& - 2j(j-1)\,H_j^2 - 2(n+1-j)(n-j)\,H_{n+1-j}^2\\
& + 4\left(j^2-(n+1)j-n^2-5n-8\right)\,H_j\,H_{n+1-j}\\
& - \frac{2}{j(n+1-j)}\left(-(2n+3)j^3+\left(2n^2+7n+24\right)j^2-(n+1)(2n+21)j+8(n+1)\right)H_j\\
& - \frac{2}{j(n+1-j)}\Big((2n+3)j^3-\left(4n^2+8n-15\right)j^2\\
& \quad +(n+1)\left(2n^2+3n-18\right)j+8(n+1)\Big)\,H_{n+1-j}
\end{aligned}$$

https://doi.org/10.1017/S0963548397003325 Published online by Cambridge University Press

$$
\begin{aligned}
&+2(n+1)(n+6)\,H_n^{(2)}\\
&-2\left(j^2+5j+8\right)\,H_j^{(2)}-2\left(j^2-(2n+7)j+n^2+7n+14\right)\,H_{n+1-j}^{(2)}\\
&+\frac{10j^4-20(n+1)j^3+\left(9n^2+31n-6\right)j^2+\left(n^2-11n+16\right)(n+1)j+32}{2j(n+1-j)}\\
&+4n\left((n+1-j)\sum_{k=1}^{j}\frac{H_{n-k}}{k}+j\sum_{k=1}^{n+1-j}\frac{H_{n-k}}{k}\right).
\end{aligned}
\tag{2.19}
$$

In the instance $n=2N+1$, $j=N+1$ (searching for the *median*), we can use the first reciprocity law to simplify the variance. We get

$$
\begin{aligned}
\mathrm{Var}_{2N+1,N+1} \;=\; &-8(2N+5)(N+1)H_{2N+1}^2+16(2N+5)(N+1)H_{2N+1}H_{N+1}\\
&+4\frac{4N^3+9N^2+2N+5}{N+1}H_{2N+1}-8\left(2N^2+7N+7\right)H_{N+1}^2\\
&-4\frac{4N^3+9N^2-13N-2}{N+1}H_{N+1}+24(N+1)H_{2N+1}^{(2)}\\
&-4\left(N^2+7N+14\right)H_{N+1}^{(2)}+\frac{7N^4+15N^3+9N^2+5N+20}{(N+1)^2}.
\end{aligned}
\tag{2.20}
$$

Asymptotically, we get from this

$$
\mathrm{Var}_{2N+1,N+1}\sim N^2\left(16\log 2-16\log^2 2-\frac{2}{3}\pi^2+7\right).
\tag{2.21}
$$

The numerical constant is 3.823370396.

To get a better feeling about the sums $\sum_{k=1}^{j}\frac{H_{n-k}}{k}$, we compute their *average*:

$$
\frac{1}{n}\sum_{1\leqslant j\leqslant n}\sum_{k=1}^{j}\frac{H_{n-k}}{k}=\frac{n+1}{n}\left(H_n^2-H_n^{(2)}\right)-H_n+1.
\tag{2.22}
$$

It might be even more instructive to compute the average value

$$
\begin{aligned}
&\frac{1}{n}\sum_{1\leqslant j\leqslant n}\left[(n+1-j)\sum_{k=1}^{j}\frac{H_{n-k}}{k}+j\sum_{k=1}^{n+1-j}\frac{H_{n-k}}{k}\right]\\
&\qquad=\frac{(n+2)(n+1)}{n}\left(H_n^2-H_n^{(2)}\right)-\frac{3n+5}{2}H_n+\frac{5n+11}{4}.
\end{aligned}
\tag{2.23}
$$

Now we want to study the asymptotic behaviour of the variance, if $n\to\infty$ and $j=\rho n$, $0<\rho<1$. Using the well-known asymptotics for the harmonic numbers and Euler's summation formula, we find the asymptotic behaviour of the extra sum.

Lemma 2.4. *As $n\to\infty$ and $j=\rho n$ with $0<\rho<1$, we have*

$$
\sum_{k=1}^{j}\frac{H_{n-k}}{k}=\log^2 n+(2\gamma+\log\rho)\log n+\left(\gamma^2+\gamma\log\rho-\mathrm{dilog}(1-\rho)\right)+o(1).
\tag{2.24}
$$

https://doi.org/10.1017/S0963548397003325 Published online by Cambridge University Press

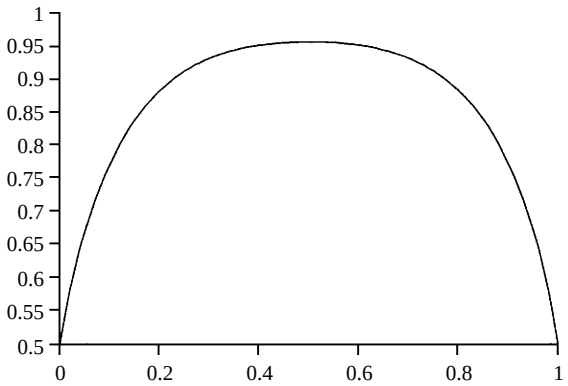


Figure 3 The coefficient of n^2 in Theorem 2, as a function of ρ.

Using this intermediary result in our formula for the variance, an elementary calculation gives us the asymptotic evaluation of the variance.

Theorem 2.2. *The variance of the number of comparisons in Hoare's Find algorithm to find the ρn-th out of n elements* $(0 < \rho < 1)$ *fulfills for* $n \to \infty$

$$\begin{aligned}\mathrm{Var}_{n,\rho n} \sim n^2 \Big(& -2\rho^2 \log^2 \rho + 4(\rho^2 - \rho - 1)\log\rho \log(1-\rho) - 2(1-\rho)^2 \log^2(1-\rho) \\ & - 4\rho \log\rho - 4(1-\rho)\log(1-\rho) - 4\rho\,\mathrm{dilog}\,\rho - 4(1-\rho)\,\mathrm{dilog}\,(1-\rho) \\ & + \frac{2}{3}\pi^2\rho(1-\rho) + 5\rho(1-\rho) + \frac{1}{2}\Big). \end{aligned} \tag{2.25}$$

3. The number of passes

A similar but somewhat easier treatment is possible for the number of passes (recursive calls) instead of the number of comparisons. Therefore we only cite the results.

We have the fundamental equation

$$\frac{\partial}{\partial z} F(z;u,v) = \frac{v\,F(z;u,v)}{1-z} + \frac{uv}{(1-z)(1-zu)} + \frac{uv\,F(z;u,v)}{1-zu}. \tag{3.1}$$

Here, $F(z;u,v)$ has as the coefficient of $z^n u^j v^m$ the probability that it takes m passes to find node j, under the assumption that all permutations of n elements are equally likely. The equation can easily be gained from referring to (2.2). Observing the fact that this time the partitioning phase 'costs' 1 (pass) yielding a factor v^1 instead of a 'cost' of $n-1$ (comparisons), which corresponded to the substitution $z \to zv$ on the right-hand side of (2.2), we find (3.1).

The expectation is an old result of Arora and Dent [2]:

$$E_{n,j} = H_j + H_{n+1-j} - 1. \tag{3.2}$$

https://doi.org/10.1017/S0963548397003325 Published online by Cambridge University Press

For the variance we get

$$\mathrm{Var}_{n,j} = \frac{2(n+1)}{j(n+1-j)} H_n + \left(1 - \frac{2(n+1)}{j(n+1-j)}\right)(H_j + H_{n+1-j}) - \left(H_j^{(2)} + H_{n+1-j}^{(2)}\right) + \frac{2}{j(n+1-j)} + 2. \tag{3.3}$$

Acknowledgements

One of us (H. P.) had the opportunity to use the powerful computers at Inria. Thanks are due to P. Flajolet for the generous invitation.

References

[1] Abramowitz, M. and Stegun, I. A. (1973) *Handbook of Mathematical Functions.* Dover. (Reprint of the Tenth National Bureau of Standards edition, 1964.)
[2] Arora, S. and Dent, W. (1969) Randomized binary search technique. *Comm. ACM* **12** 77–80.
[3] Graham, R. L., Knuth, D. E. and Patashnik, O. (1989) *Concrete Mathematics.* Addison-Wesley.
[4] Greene, D. H. and Knuth, D. E. (1982) *Mathematics for the Analysis of Algorithms.* Birkhäuser.
[5] Hoare, C. A. R. (1961) Find (Algorithm 65). *Comm. ACM* **4** 321–322.
[6] Hoare, C. A. R. (1962) Quicksort. *Computer J.* **5** 10–15.
[7] Hwang, H.-K. (1990) On the average shape of binary search trees. Unpublished manuscript.
[8] Kirschenhofer, P., Martínez, C. and Prodinger, H. (1997) Analysis of Hoare's Find algorithm with median-of-three partition. *Random Structures and Algorithms* **10** 143–156.
[9] Knuth, D. E. (1968) *The Art of Computer Programming: volume 1: Fundamental Algorithms.* Addison-Wesley. (2nd ed. 1973.)
[10] Knuth, D. E. (1972) Mathematical analysis of algorithms. In *Information Processing 71*, pp. 19–27. North-Holland.
[11] Knuth, D. E. (1973) *The Art of Computer Programming: volume 3: Sorting and Searching.* Addison-Wesley.
[12] Lent, J. and Mahmoud, H. M. (1996) Average-case analysis of multiple quickselect: An algorithm for finding order statistics. *Statistics and Probability Letters* **28** 299–310.
[13] Mahmoud, H. M. (1992) *Evolution of Random Search Trees.* John Wiley, New York.
[14] Mahmoud, H. M., Modarres, R. M. and Smythe, R. T. (1995) Analysis of one-sided quicksort: An algorithm for order statistics. *RAIRO, Theoretical Informatics and Applications* **29** 255–276.
[15] Prodinger, H. (1995) Multiple quickselect – Hoare's Find algorithm for several elements. *Information Processing Letters* **56** 123–129.

https://doi.org/10.1017/S0963548397003325 Published online by Cambridge University Press