A GENERATING FUNCTIONS APPROACH FOR THE ANALYSIS OF GRAND AVERAGES FOR MULTIPLE QUICKSELECT

ALOIS PANHOLZER AND HELMUT PRODINGER

Abstract. Hoare's Find algorithm can be used to select p specified order statistics $j_1, \dots, j_p$ from a file of n elements simultaneously. Averaging over all $\binom{n}{p}$ subsets of $\{1, \dots, n\}$ defines the grand averages of the number of passes and comparisons. We use a generating functions approach to compute averages and variances of these parameters.

Additionally we sketch analogous developments for the instance of median-of-three partition and binary (Catalan) trees.

1. Introduction

Hoare's Find algorithm ("Quickselect") [2] uses the idea of *Quicksort* [3, 7] in order to select the jth element of a file of n elements. In each partitioning step, a certain element will be brought into its correct position k. In the process, elements are moved—those brought to the left of it are smaller, and those brought to the right are larger. If $k = j$, the element is found; if $k < j$, one has to continue on the right side, searching for the $(j-k)$th element; if $k > j$, one has to continue on the left side, still searching for the jth element.

Two parameters are of interest: the (average) number of passes (recursive calls) $P[n; j]$ and the (average) number of comparisons $C[n; j]$. In Knuth's book [7] we already find the values

$$P[n; j] = H_j + H_{n+1-j} - 1$$

and

$$C[n; j] = 2\Big(n + 3 + (n+1)H_n - (j+2)H_j - (n+3-j)H_{n+1-j}\Big)$$

where $H_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n}$ denotes the nth harmonic number; compare also [6] and [10].

Hoare's idea can be extended as follows: Assume that we simultaneously search for the p elements with ranks $j_1, j_2, \dots, j_p$ where $1 \le j_1 < j_2 < \cdots < j_p \le n$ is a fixed set of p values. Then, according into which interval of the values $j_1, \dots, j_p$ the pivot element falls, we may have found one of the elements and/or have to continue on both sides, but with smaller sets. The idea is quite simple; a nice description of the algorithm can be found in [8].

In [13], explicit formulæ for both the average number of passes $P[n; j_1, \dots, j_p]$ and the average number of comparisons, $C[n; j_1, \dots, j_p]$ were given. As a corollary, the *grand* averages,

$$\mathcal{P}_{n,p} = \frac{1}{\binom{n}{p}} \sum_{1 \le j_1 < j_2 < \cdots < j_p \le n} P[n; j_1, \dots, j_p]$$

and

$$\mathcal{C}_{n,p} = \frac{1}{\binom{n}{p}} \sum_{1 \le j_1 < j_2 < \cdots < j_p \le n} C[n; j_1, \dots, j_p]$$

were also computed (compare also [8]). However, the approach used in this previous paper was a mixture of guessing and proofs by induction.

In the present paper, we use a generating functions approach to attack the analysis of the grand averages directly. This enables us to compute the respective variances, something that was out of reach with the old method.

In both instances (passes resp. comparisons) we will use a trivariate generating function $\Phi(z, u, v)$ where the coefficient of $[z^n/n!\, u^p v^m]$ is the number of permutations of $\{1, 2, \dots, n\}$ and the subsets of p elements of $\{1, 2, \dots, n\}$ such that the parameter of interest has value m.

Date: July 27, 1997.

Then, for fixed n and p with $0 \le p \le n$,

$$\frac{[z^n u^p]\Phi(z,u,v)}{\binom{n}{p}}$$

is the probability generating function of interest. Apart from the normalization factor $\binom{n}{p}$, the functions

$$M_1(z,u) = \frac{\partial}{\partial v}\Phi(z,u,v)\Big|_{v=1}$$

and

$$M_2(z,u) = \frac{\partial^2}{\partial v^2}\Phi(z,u,v)\Big|_{v=1}$$

are the generating functions of the expectations and the second factorial moments. To be more precise, the expectation $E_{n,p}$ and the variance $V_{n,p}$ are given by

$$\begin{aligned} E_{n,p} &= \frac{[z^n u^p]M_1(z,u)}{\binom{n}{p}}\,, \\ V_{n,p} &= \frac{[z^n u^p]M_2(z,u)}{\binom{n}{p}} + E_{n,p} - (E_{n,p})^2\,. \end{aligned} \tag{1}$$

(Since we decided not to overload the notation by an additional superscript $[P]$ (for passes) and $[C]$ (for comparisons), we have chosen the neutral $E_{n,p}$ which means $\mathcal{P}_{n,p}$ or $\mathcal{C}_{n,p}$, dependening on the section.)

For the following it is more suggestive to think about the binary search tree that is naturally associated to a permutation via the QUICKSORT procedure; the pivot element corresponds to the root, and the left (right) subfile corresponds to the left (right) subtree.

The last sections sketch analogous computations for median-of-three quickselect (see e.g. [4]); for the sake of shortness we confine ourselves to the instance of the expectations, since they are already quite complex. However, in principle, the variances could be obtained explicitly.

For reasons of comparisons we included also a section about binary (Catalan) trees. Things are a bit easier, since instead of differential equations we have to deal with algebraic equations. Again, we only give the expectations.

We also want to draw the reader's attention to the papers [10] and [11] which attack quickselect resp. multiple quickselect with probabilistic tools. The main achievements in these papers are the establishing of the limiting distributions. However, we choose the present approach (suggested to us by Philippe Flajolet) since we find it more promising in order to get *explicit* results.

2. Average number of passes

Let $\phi_p(z,v)$ be the following generating function: The coefficient of $z^n/n!v^k$ is the sum over all trees of size n and all labellings of exactly p nodes, such that the smallest subtree containing the root and these p nodes (which corresponds to the number of passes!) is k. Then we have the recurrence

$$\phi_p'(z,v) = v\sum_{i=0}^{p}\phi_i(z,v)\phi_{p-i}(z,v) + v\sum_{i=0}^{p-1}\phi_i(z,v)\phi_{p-1-i}(z,v) \tag{2}$$

for $p \ge 1$ and $\phi_0(z) = \frac{1}{1-z}$. (The first sum refers to a nonlabelled, and the second sum to labelled root.)

Typically in such analyses, a first derivative appears because of the (uniform) splitting probability $\frac{1}{n}$ (every element can be the root, with equal probability); the multiplication of the recursion then introduces a factor n on the left hand side, which means (in terms of generating functions) a derivative. This is quite classical (see for instance [7]) and we confine ourselves with this brief statement.

Observe that

$$\frac{1}{\binom{n}{p}}[z^n]\phi_p(z,v)$$

are the probability generating functions of interest.

Then the trivariate generating function from the introduction is given by

$$\Phi(z,u,v)=\sum_{p\ge0}\phi_p(z,v)u^p\ .$$

We get, by summing up the recursion (2),

$$\Phi'(z,u,v)=v(1+u)\Phi^2(z,u,v)+\frac{1-v}{(1-z)^2} \tag{3}$$

with $\Phi(0,u,v)=1$.

MAPLE is capable to solve this differential equation:

Lemma 1.

$$\Phi(z,u,v)=\frac{\Omega+1-2v+(1-z)^{\Omega}(\Omega-1+2v)}{\Big(\Omega+1-2v(1+u)+(1-z)^{\Omega}(\Omega-1+2v(1+u))\Big)(1-z)} \tag{4}$$

with

$$\Omega=\sqrt{1-4v(1+u)(1-v)}\ .\quad\square$$

Consequently we get

$$M_1(z,u)=\frac{1+2u}{1-z(1+u)}-\frac{2u}{\big(1-z(1+u)\big)^2}-\frac{1}{1-z}+\frac{2u(1+u)}{\big(1-z(1+u)\big)^2}\log\frac{1}{1-z}\ . \tag{5}$$

We have

$$[z^n]\left(\frac{1+2u}{1-z(1+u)}-\frac{2u}{\big(1-z(1+u)\big)^2}-\frac{1}{1-z}\right)=(1+2u)(1+u)^n-2(n+1)u(1+u)^n-1\ .$$

Let us assume that $p\ge1$. Reading off the coefficient of u^p in the last formula, we obtain

$$\binom{n}{p}-2n\binom{n}{p-1}\ .$$

Now we turn to the logarithmic term. The factor of $\log\frac{1}{1-z}$ is

$$\frac{2u(1+u)}{\big(1-z(1+u)\big)^2}=\frac{2u(1+u)}{(1-z)^2}\frac{1}{\big(1-\frac{uz}{1-z}\big)^2}\ ,$$

and the coefficient of u^p therein is

$$\frac{2pz^{p-1}}{(1-z)^{p+1}}+\frac{2(p-1)z^{p-2}}{(1-z)^p}\ .$$

We observe the general formula (see e.g [1])

$$[z^n]\frac{1}{(1-z)^{m+1}}\log\frac{1}{1-z}=\binom{n+m}{n}(H_{n+m}-H_m) \tag{6}$$

and find so, searching for the coefficient of z^n,

$$2p\binom{n+1}{p}(H_{n+1}-H_p)+2(p-1)\binom{n+1}{p-1}(H_{n+1}-H_{p-1})\ .$$

We collect both contributions, divide by $\binom{n}{p}$ and get

Theorem 2. *For $p\ge1$ the average number $E_{n,p}$ of passes to search for p random elements is given by*

$$E_{n,p}=(H_{n+1}-H_p)\,\frac{2p(n+1)^2}{(n+2-p)(n+1-p)}-\frac{n(2p-1)+p}{n+2-p}\ . \tag{7}$$

For $p\ge0$ we have $E_{n,0}=0$. $\square$

This result appeared already in [13].

It is interesting to state the special cases $p=1$ (Hoare's FIND algorithm or QUICKSELECT) and $p=n$ (QUICKSORT) explicitly.

Corollary 3.

$$
\begin{aligned}
E_{n,1} &= 2\left(1+\frac{1}{n}\right)H_n - 3\,,\\
E_{n,n} &= n\,.
\end{aligned}
$$

By standard methods we obtain also the following asymptotic equivalents.

Corollary 4.

$$
\begin{aligned}
E_{n,p} &= 2p\log n + 2\gamma p - 2p - 2pH_p + 1 + \mathcal{O}\left(\frac{\log n}{n}\right) \qquad \text{for } p \text{ fixed}\,,\\
E_{n,p} &= n - \frac{1}{3}l + \mathcal{O}\left(\frac{\log n}{n}\right) \qquad \text{for } l = n-p \text{ fixed}\,,\\
E_{n,p} &= -\frac{2\rho(1-\rho+\log\rho)}{(1-\rho)^2}n + \frac{\rho(4\rho\log\rho + 2\log\rho - \rho^2 - 4\rho + 5)}{(1-\rho)^3} + \mathcal{O}\left(\frac{\log n}{n}\right)\\
&\qquad \text{for } p = \rho n \text{ and } 0 < \rho < 1.
\end{aligned}
\tag{8}
$$

3. Variance of the number of passes

From the function $\Phi(z,u,v)$ we find

$$
\begin{aligned}
M_2(z,u) = &-2\frac{1+4u+2u^2}{1-z(1+u)} - 2\frac{u(-3+2u)}{\left(1-z(1+u)\right)^2} + 8\frac{u^2}{\left(1-z(1+u)\right)^3} + 2\frac{1+u}{1-z}\\
&-\left(\frac{16u^2(1+u)}{\left(1-z(1+u)\right)^3} + \frac{4u(1-u)(1+u)}{\left(1-z(1+u)\right)^2}\right)\log\frac{1}{1-z}\\
&+\left(\frac{8u^2(1+u)^2}{\left(1-z(1+u)\right)^3} + \frac{4u(1+u)^2}{\left(1-z(1+u)\right)^2}\right)\log^2\frac{1}{1-z}\,.
\end{aligned}
\tag{9}
$$

Now we assume that $p \ge 2$ and read off the coefficient of $z^n u^p$ in the non–logarithmic term, and divide by $\binom{n}{p}$, which gives

$$
2\frac{(p+1)(2p-1)n^2 + (p^2+3p-3)n + p - 2}{(n+2-p)(n+1-p)}\,.
$$

Furthermore, we consider the term with $\log\frac{1}{1-z}$. The coefficient of u^p in the factor of $\log\frac{1}{1-z}$ is

$$
-8p(p-1)\frac{z^{p-2}}{(1-z)^{p+1}} - 8(p-1)(p-2)\frac{z^{p-3}}{(1-z)^p} - 4p\frac{z^{p-1}}{(1-z)^{p+1}} + 4(p-2)\frac{z^{p-3}}{(1-z)^{p-1}}\,.
$$

We are searching for the coefficient of z^n in this times $\log\frac{1}{1-z}$, use the general formula (6) and obtain

$$
\frac{-4p(n+1)^2\big((2p-1)n+2p\big)(H_{n+1}-H_p) - 16(p-1)^2n^2 - 4(8p^2-17p+10)n - 4(4p^2-9p+6)}{(n+3-p)(n+2-p)(n+1-p)}\,.
$$

For the term with the square of $\log\frac{1}{1-z}$ we get in a similar fashion

$$
\frac{X}{(n+4-p)(n+3-p)(n+2-p)(n+1-p)}
$$

with

$$
\begin{aligned}
X = \;& 4p(n+2)(n+1)^2(pn+p+2)\Big(H_{n+1}^2 - H_{n+1}^{(2)} + H_p^2 + H_p^{(2)}\Big)\\
&- 8(n+1)^2H_{n+1}\big(n^2p^2H_p + n(p(3p+2)H_p - (3p-2)(p-1)) + 2p(p+2)H_p - 5p^2 + 8p - 4\big)\\
&- 8(n+1)^2H_p\Big((3p-2)(p-1)n + 5p^2 - 8p + 4\Big)\\
&+ 8(n+1)(p-2)\Big(n(3p-4) + 4p - 7\Big)\,.
\end{aligned}
$$

For that we have used the general formula

$$[z^n]\frac{1}{(1-z)^{m+1}}\log^2\frac{1}{1-z} = \binom{n+m}{m}\Big((H_{n+m}-H_m)^2-(H^{(2)}_{n+m}-H^{(2)}_m)\Big)\ . \tag{10}$$

Dividing this by $\binom{n}{p}$ leads to the following theorem; for $p=1$ a similar but simpler computation, based on the formula

$$[u]M_2(z,u) = \frac{4z}{(1-z)^2} - \frac{4}{(1-z)^2}\log\frac{1}{1-z} + \frac{4}{(1-z)^2}\log^2\frac{1}{1-z}$$

was used.

Theorem 5. *The variance of the number of passes when searching for a random set of $p \geq 2$ elements with Hoare's* Find *algorithm is given by*

$$\begin{aligned} V_{n,p} = &-\frac{4p(n+1)^2\Psi_1}{(n+4-p)(n+3-p)(n+2-p)^2(n+1-p)^2}\Big(H_{n+1}-H_p\Big)^2 \\ &+\frac{2(n+1)^2\Psi_2}{(n+4-p)(n+3-p)(n+2-p)^2(n+1-p)}\Big(H_{n+1}-H_p\Big) \\ &-\frac{4p(n+2)(n+1)^2(pn+p+2)}{(n+4-p)(n+3-p)(n+2-p)(n+1-p)}\Big(H^{(2)}_{n+1}-H^{(2)}_p\Big) \\ &+\frac{2(n+1)^2\Psi_3}{(n+4-p)(n+3-p)(n+2-p)^2} \end{aligned} \tag{11}$$

with

$$\begin{aligned} \Psi_1 &= (3p-2)n^3-2(p^2-9p+5)n^2-(p^3+5p^2-33p+16)n-p^3-5p^2+20p-8\ , \\ \Psi_2 &= pn^3+(11p^2-13p+8)n^2-(9p^3-54p^2+62p-32)n-3p^4-p^3+38p^2-56p+32\ , \\ \Psi_3 &= (2p-1)n^2-2(5p^2-15p+8)n+8p^3-45p^2+72p-32\ . \end{aligned} \tag{12}$$

For $p=1$, the formula reads

$$V_{n,1} = -4\frac{n+1}{n^2}H^2_{n+1} + \frac{2(n+4)(n+1)}{n^2}H_{n+1} - 4\frac{n+1}{n}H^{(2)}_{n+1} + \frac{2(2n^2-n-2)}{n^2}\ ; \tag{13}$$

for $p=0$ we have $V_{n,0}=0$. □

Observe that for $p=n$ (Quicksort) this formula simplifies to $V_{n,n}=0$, as it should.

Corollary 6.

$$\begin{aligned} V_{n,p} &= 2p\log n + 4p^2H^{(2)}_p - 2pH_p + 2p\gamma - \frac{2}{3}p^2\pi^2 + \mathcal{O}\left(\frac{\log n}{n}\right) \quad \text{for } p \text{ fixed}\ , \\ V_{n,p} &= \frac{2}{9}l + \mathcal{O}\left(\frac{1}{n}\right) \quad \text{for } l=n-p \text{ fixed}\ , \\ V_{n,p} &= \left(-\frac{4\rho^2(\rho+3)\log^2\rho}{(1-\rho)^5} - \frac{2\rho(3\rho^2+12\rho+1)\log\rho}{(1-\rho)^4} - \frac{16\rho^2}{(1-\rho)^3}\right)n + \mathcal{O}(1) \\ &\quad \text{for } p=\rho n \text{ and } 0<\rho<1. \end{aligned} \tag{14}$$

4. Comparisons

If we have a tree with n nodes, there are costs of $n-1$ comparisons, and then the additional costs, resulting from the subtrees. Typically, this leads, apart from other things, to a substitution $z \mapsto zv$. Otherwise, we distinguish again whether or not the root is one of elements being sought.

The fundamental differential equation is thus

$$\Phi'(z,u,v) = (1+u)\Phi^2(zv,u,v) + \frac{1}{(1-z)^2} - \frac{1}{(1-zv)^2}\ ,$$

with $\Phi(0,u,v)=1$. Now, we cannot solve that, but we can differentiate it with respect to v (followed by $v:=1$) and solve the resulting differential equation for $M_1(z,u)$. (In the next section we will differentiate it twice, plug in the already known $M_1(z,u)$ and solve the equation for $M_2(z,u)$.)

$$
\begin{aligned}
M_1(z,u) = &-\frac{2(1+u)(1+3u)}{(1-z(1+u))^2}\log\frac{1}{1-z} + \frac{2}{(1-z(1+u))^2}\log\frac{1}{1-z(1+u)} \\
&+\frac{2(1+2u)}{u(1-z)} - \frac{1}{(1-z)^2} - \frac{2(1+3u+3u^2)}{u(1-z(1+u))} + \frac{3(1+2u)}{(1-z(1+u))^2}\,.
\end{aligned}
\tag{15}
$$

From this, we read off the coefficients much in the same style as before, divide by $\binom{n}{p}$ and obtain the expectations (known already from [13]).

Theorem 7. *For $p \geq 1$ the average number $E_{n,p}$ of comparisons to search for p random elements is given by*

$$
E_{n,p} = -\frac{2p(n+1)(4n-p+5)}{(n+2-p)(n+1-p)}H_{n+1} + \frac{2(n+1)^2(n+2p+2)}{(n+2-p)(n+1-p)}H_p + \frac{n^2+(5p-1)n+3p}{n+2-p}\,. \tag{16}
$$

For $p=0$, we have $E_{n,0}=0$. □

Again, we note the special cases $p=1$ and $p=n$.

Corollary 8.

$$
\begin{aligned}
E_{n,1} &= 3n - 8\left(1+\frac{1}{n}\right)H_n + 13\,,\\
E_{n,n} &= 2(n+1)H_n - 4n\,.
\end{aligned}
\tag{17}
$$

Corollary 9.

$$
\begin{aligned}
E_{n,p} &= (2H_p+1)n - 8p\log n + 2(4p+1)H_p + 3(2p-1) - 8p\gamma + \mathcal{O}\left(\frac{\log n}{n}\right) \quad \text{for } p \text{ fixed}\,,\\
E_{n,p} &= 2n\log n + 2(\gamma+2)n + 2\log n + 1 + 2\gamma + \mathcal{O}\left(\frac{\log n}{n}\right) \quad \text{for } l=n-p \text{ fixed}\,,\\
E_{n,p} &= 2n\log n + \frac{(4\rho+2)\log\rho + (2\rho^2-4\rho+2)\gamma - 5\rho^2+4\rho+1}{(1-\rho)^2}n + 2\log n + \mathcal{O}(1)\\
&\qquad \text{for } p=\rho n \text{ and } 0<\rho<1.
\end{aligned}
\tag{18}
$$

5. Variance of the number of comparisons

As indicated in the last section, we can set up a differential equation for $M_2(z,u)$, viz.

$$
\begin{aligned}
M_2'(z,u) = {}& 2\frac{(1+u)}{1-z(1+u)}M_2(z,u) + 8\frac{(3u+1)^2(1+u)^3}{(1-z(1+u))^4}\log^2\frac{1}{1-z}\\
&+8\frac{(1+u)}{(1-z(1+u))^4}\log^2\frac{1}{1-z(1+u)} - 16\frac{(3u+1)(1+u)^2}{(1-z(1+u))^4}\log\frac{1}{1-z}\log\frac{1}{1-z(1+u)}\\
&-8\frac{(3u+1)(1+u)^2(2z^2u^2-9zu^2+6u^2+5z^2u+9u-14zu+3-6z+3z^2)z}{(1-z(1+u))^4(1-z)^2}\log\frac{1}{1-z}\\
&+8\frac{(1+u)(2z^2u^2-9zu^2+6u^2+5z^2u+9u-14zu+3-6z+3z^2)z}{(1-z(1+u))^4(1-z)^2}\log\frac{1}{1-z(1+u)}\\
&+\frac{2z^2u\,\Psi}{(1-z)^4(1-z(1+u))^4}
\end{aligned}
\tag{19}
$$

with
$\Psi = 53 - 164z + 165u - 316z^3u^2 - 242z^3u + 223u^2 + 144u^3 - 68z^3 + 5z^4 - 438zu^3 + 174z^2 + 592z^2u + 766z^2u^2 - 700zu^2 + 19z^4u + 27z^4u^2 + 17z^4u^3 - 534zu + 105u^4z^2 + 36u^4 + 4z^4u^4 - 36z^3u^4 + 453z^2u^3 - 178z^3u^3 - 108u^4z$
and $M_2(0,u)=0$.

Now this time Maple does not do it automatically, since it cannot handle

$$
\int_0^z \frac{\log^2(1-t)}{(1-t(1+u))^2}dt\,.
$$

However, by integration by parts this integral can be evaluated, and consequently we get the following solution.

Lemma 10.

$$\begin{aligned}
M_2(z,u) &= 24\frac{(1+u)^3}{(1-z(1+u))^2}\int_0^z \frac{1}{1-t(1+u)}\log\frac{1}{1-t}\,dt \\
&+ 8\frac{(3u+1)(1+u)^2(-zu+3u+2-z)}{(1-z(1+u))^3}\log^2\frac{1}{1-z} \\
&+ 4\frac{1+z+zu}{(1-z(1+u))^3}\log^2\frac{1}{1-z(1+u)} \\
&- 16\frac{(1+u)(-zu+3u+2-z)}{(1-z(1+u))^3}\log\frac{1}{1-z}\log\frac{1}{1-z(1+u)} \\
&+ 4\frac{(1+u)\,\Psi_1}{u(1-z)(1-z(1+u))^3}\log\frac{1}{1-z} \\
&- 8\frac{8z^2u^3+9z^2u+z^2+16z^2u^2-20u^2z-9u^3z-2z-14zu+5u+3u^2+1}{u(1-z)(1-z(1+u))^3}\log\frac{1}{1-z(1+u)} \\
&+ \frac{2}{3}\frac{z\,\Psi_2}{(1-z)^3(1-z(1+u))^3}
\end{aligned} \tag{20}$$

with
$\Psi_1 = 10u^4z^2 - 16u^4z + 31z^2u^3 - 19u^3z - 20u^3 - 7u^2 + 35z^2u^2 - 30u^2z + 8u - 24zu + 16z^2u + 2z^2 - 4z + 2$
and
$\Psi_2 = 12 - 330u^3z + 39z^4u^3 - 12u^4z + 346z^2u^3 - 173u^3z^3 + 42z^4u^2 + 15z^4u - 519u^2z + 478z^2u^2 - 199u^2z^3 - 36z - 237zu + 36z^2 + 90u + 219z^2u - 87uz^3 - 12z^3 + 12z^4u^4 - 49u^4z^3 + 48u^4z^2 + 198u^2 + 120u^3$. □

We observe that

$$\begin{aligned}
&(1+u)\int_0^z \frac{1}{1-t(1+u)}\log\frac{1}{1-t}\,dt = \\
&\quad = \log(1-z(1+u))\log(u) + \operatorname{dilog}\Big(\frac{1+u}{u}\Big) - \operatorname{dilog}\Big(\frac{1-z+u-zu}{u}\Big) - \log(1+u)\log(1-z(1+u))\,,
\end{aligned}$$

which is more "closed form" but less useful, since for the coefficients the integrand is as good as the integral itself, since we can simply look for $1/n$ times the coefficient of $z^{n-1}u^p$.

We remark that Maple's dilog function is related to the more common polylog function

$$\mathrm{L}_2(z) = \sum_{n\ge1}\frac{z^n}{n^2}$$

via $\operatorname{dilog}(1-z) = \mathrm{L}_2(z)$ (compare [9]).

The relevant formula for $p \ge 1$ is this

$$\frac{1}{n}[z^{n-1}u^p]\frac{1}{1-z(1+u)}\log\frac{1}{1-z} = \frac{1}{n}\binom{n-1}{p-1}\Big(H_{n-1} - H_{p-1}\Big)\,.$$

For $p = 0$ it is slightly different;

$$\frac{1}{n}[z^{n-1}u^0]\frac{1}{1-z(1+u)}\log\frac{1}{1-z} = \frac{H_{n-1}}{n}\,.$$

On the other hand,

$$[z^nu^p]\frac{(1+u)^3}{(1-z(1+u))^2} = (n+1)\binom{n+3}{p}\,.$$

The convolution of these two quantities can be performed by (the by now) standard tools.

There is only one sum,

$$\lambda_{n,p} = \frac{1}{\binom{n}{p}} \sum_{k=1}^{n} \frac{1}{k^2} \binom{n-k}{p} = \frac{1}{\binom{n}{p}} [z^n] \frac{z^p}{(1-z)^{p+1}} \operatorname{L}_2(z) , \tag{21}$$

that has a "new" shape; we will relate it to more common quantities a bit later.

Observe that an analogous quantity has a well known evaluation:

$$\frac{1}{\binom{n}{p}} \sum_{k=1}^{n} \frac{1}{k} \binom{n-k}{p} = \frac{1}{\binom{n}{p}} [z^n] \frac{z^p}{(1-z)^{p+1}} \log \frac{1}{1-z} = H_n - H_p . \tag{22}$$

After a lengthy routine calculation we arrive at the following theorem.

Theorem 11. *For $p \ge 2$ the variance $V_{n,p}$ of the number of comparisons to search for p random elements is given by*

$$\begin{aligned} V_{n,p} = &- \frac{24(n+1)^2(n+2)}{(n+1-p)(n+2-p)} \lambda_{n+2,p} + \frac{4(n+1)^2 p\, \Psi_1}{(n+2-p)^{\underline{2}}\,(n+4-p)^{\underline{4}}} \Big(H_{n+1} - H_p \Big)^2 \\ &+ \frac{2(n+1)\Psi_2}{(n+4-p)^{\underline{4}}\,(n+2-p)} \Big(H_{n+1} - H_p \Big) + \frac{4(n+1)^2\Psi_3}{(n+4-p)^{\underline{4}}} \Big(H^{(2)}_{n+1} - H^{(2)}_p \Big) \\ &- 2(n+1)H_p - 4(n+1)^2 H^{(2)}_p + \frac{(n+1)\Psi_4}{3(p+1)^2(p+2)(n+2-p)\,(n+4-p)^{\underline{4}}} \end{aligned} \tag{23}$$

with

$$\begin{aligned} \Psi_1 =& -4n^4 + (-27p-13)n^3 + (30p^2 - 183p + 13)n^2 \\ &+ (p^3 + 108p^2 - 379p + 70)n - 2p^3 + 108p^2 - 250p + 48 , \\ \Psi_2 =& (54p-28)n^4 + (25p^2 + 317p - 172)n^3 + (-67p^3 + 198p^2 + 655p - 368)n^2 \\ &+ (-13p^4 - 107p^3 + 252p^2 + 618p - 320)n + p^5 - 20p^4 - 23p^3 + 62p^2 + 232p - 96 , \\ \Psi_3 =& 6n^3 + (-16p^2 - 8p + 54)n^2 \\ &+ (8p^3 - 63p^2 - 65p + 156)n - (p-1)(p^3 - 13p^2 + 36p + 144) , \\ \Psi_4 =& (21p^3 + 80p^2 + 81p - 2)n^6 + (-105p^4 - 76p^3 + 815p^2 + 1258p + 40)n^5 \\ &+ (210p^5 - 562p^4 - 2173p^3 + 2718p^2 + 8125p + 886)n^4 \\ &+ (-210p^6 + 1036p^5 - 223p^4 - 9232p^3 + 2283p^2 + 25662p + 4096)n^3 \\ &+ (105p^7 - 500p^6 + 2062p^5 + 1352p^4 - 18253p^3 - 4600p^2 + 43086p + 7972)n^2 \\ &+ (-21p^8 + 40p^7 - 512p^6 + 3390p^5 + 3029p^4 - 19638p^3 - 11176p^2 + 37736p + 7600)n \\ &- 18p^8 - 50p^7 - 348p^6 + 2232p^5 + 2298p^4 - 8886p^3 - 6540p^2 + 13616p + 2880 . \end{aligned} \tag{24}$$

For $p = 1$ we get

$$V_{n,1} = -\frac{16(n+1)(n+4)}{n^2} H_n^2 + \frac{12(9n+17)}{n} H_n - \frac{48(n+1)}{n} H^{(2)}_n + n^2 - 10n - 47 ; \tag{25}$$

and for $p = 0$ we have $V_{n,0} = 0$. □

Observe that for $p = n$ the formula simplifies to

$$V_{n,n} = -2(n+1)H_n - 4(n+1)^2 H^{(2)}_n + (7n+13)n . \tag{26}$$

We can rewrite the quantity $\lambda_{n,p}$ as follows:

Proposition 1.

$$\lambda_{n,p} = -\frac{1}{2}\Big[H_n^2 - H^{(2)}_n - H^2_{n-p} - H^{(2)}_{n-p} \Big] + \sum_{j=1}^{p-1} \frac{H_j}{j+n+1-p} . \tag{27}$$

PROOF. We write

$$\binom{n-k}{p} = \frac{n-k}{p}\binom{n-1-k}{p-1}$$

and get consequently the recursion

$$\lambda_{n,p} = \lambda_{n-1,p-1} - \frac{1}{n}\Big(H_{n-1} - H_{p-1}\Big) .$$

Unwinding this we get

$$\lambda_{n,p} = H^{(2)}_{n-p} - \sum_{k=n-p-1}^{n} \frac{1}{k}\Big(H_{k-1} - H_{k-1-n+p}\Big) .$$

Since

$$\sum_{k=1}^{n} \frac{H_{k-1}}{k} = \frac{1}{2}\Big(H_n^2 - H_n^{(2)}\Big) ,$$

the result readily follows. □

In order to handle the sum $\sum_{j=1}^{p-1} \frac{H_j}{n-p+1+j}$ asymptotically for either p or $l = n-p$ fixed or for a fixed ratio $\rho = \frac{p}{n}$ with $0 < \rho < 1$ as $n \to \infty$ we state a few relations that are easy to prove (but not so easy to find) by generating functions and/or can be found in [5].

Proposition 2.

$$\sum_{j=1}^{p-1} \frac{H_j}{n-p+1+j} = H_n H_p - \sum_{j=1}^{p} \frac{H_{n-p+j}}{j} , \tag{28a}$$

$$\begin{aligned}\sum_{j=1}^{p-1} \frac{H_j}{n-p+1+j} &= \sum_{j=1}^{p} \frac{H_{n-j}}{j} - \frac{1}{n+1-p}(H_n - H_p - H_{n+1-p}) - H_p H_{n+1-p} + H_n H_{n+1-p} \\ &\quad - \frac{1}{2}(H_n^2 - H_n^{(2)}) - \frac{1}{2}(H_{n+1-p}^2 + H_{n+1-p}^{(2)}) ,\end{aligned} \tag{28b}$$

$$\begin{aligned}\sum_{j=1}^{p-1} \frac{H_j}{n-p+1+j} &= -\sum_{j=p+1}^{n} \frac{H_{j-p}}{j} - \frac{1}{2}(H_{p-1}^2 + H_{p-1}^{(2)}) + H_n H_{p-1} + \frac{H_{n-p}}{p} \\ &\quad - \frac{1}{n+1-p}(H_n - H_p - H_{n+1-p}) - H_p H_{n+1-p} + H_n H_{n+1-p} \\ &\quad - \frac{1}{2}(H_{n+1-p}^2 + H_{n+1-p}^{(2)}) , \quad \square\end{aligned} \tag{28c}$$

For fixed p we can use the definition and find

$$\sum_{j=1}^{p-1} \frac{H_j}{n-p+1+j} = \frac{p(H_p - 1)}{n} + \mathcal{O}\left(\frac{1}{n^2}\right) \quad \text{for } n \to \infty . \tag{29}$$

For fixed $l = n-p$ we compute

$$\sum_{j=p+1}^{n} \frac{H_{j-p}}{j} = \sum_{j=1}^{l} \frac{H_j}{n+j-l} = \sum_{j=1}^{l} \frac{H_j}{n}\left(1 + \mathcal{O}\Big(\frac{1}{n}\Big)\right) = \frac{(l+1)H_l - l}{n} + \mathcal{O}\left(\frac{1}{n^2}\right) , \tag{30}$$

and use (28c) to find

$$\begin{aligned}\sum_{j=1}^{p-1} \frac{H_j}{n-p+1+j} &= \frac{1}{2}\log^2 n + \gamma \log n - \frac{\pi^2}{12} + \frac{H_{l+1}}{l+1} - \frac{1}{2}H_{l+1}^2 + \frac{\gamma^2}{2} - \frac{1}{2}H_{l+1}^{(2)} + \frac{\log n}{2n} + \frac{\gamma}{2n} \\ &\quad - \frac{lH_l}{n} - \frac{l}{(l+1)n} + \frac{1}{2n} + \frac{lH_{l+1}}{n} + \frac{l}{n} + \mathcal{O}\left(\frac{\log n}{n^2}\right) \quad \text{for } n \to \infty .\end{aligned} \tag{31}$$

For a fixed ratio $\rho = \frac{p}{n}$ with $0 < \rho < n$ we use the expansion

$$\sum_{j=1}^{p} \frac{H_{n-j}}{j} = \log^2 n + (2\gamma + \log\rho)\log n + \gamma^2 + \gamma\log\rho - \mathrm{L}_2(\rho) + o(1) \tag{32}$$

from [5]. With this and equation (28b) we get

$$\begin{aligned}\sum_{j=1}^{p-1} \frac{H_j}{n-p+1+j} &= -\log(1-\rho)\log n - \frac{1}{2}\log^2(1-\rho) - \gamma\log(1-\rho) - \log\rho\log(1-\rho)\\ &\quad - \mathrm{L}_2(\rho) + o(1) \quad \text{for } n \to \infty\,.\end{aligned} \tag{33}$$

Combining these formulæ with the standard expansions for harmonic numbers we get the following lemma.

Lemma 12. *The quantity $\lambda_{n,p}$ has the following asymptotic equivalents:*

$$\begin{aligned}\lambda_{n,p} &= \frac{\pi^2}{6} + \mathcal{O}\left(\frac{\log n}{n}\right) \quad \text{for } p \text{ fixed}\,,\\ \lambda_{n,p} &= \frac{l}{n} + \mathcal{O}\left(\frac{1}{n^2}\right) \quad \text{for } l = n - p \text{ fix}\,,\\ \lambda_{n,p} &= \mathrm{L}_2(1-\rho) + o(1) \quad \text{for } p = \rho n \text{ and } 0 < \rho < 1.\end{aligned} \tag{34}$$

With this we get the following corollary.

Corollary 13. *The variance $V_{n,p}$ of the number of comparisons is asymptotically given by*

$$\begin{aligned}V_{n,p} &= \left(\frac{21p^3 + 80p^2 + 81p - 2}{3(p+1)^2(p+2)} - 4H_p^{(2)}\right) n^2 + \mathcal{O}(n) \quad \text{for } p \text{ fixed}\,,\\ V_{n,p} &= \left(\frac{7l^4 + 70l^3 + 245l^2 + 368l + 168}{(l+4)^{\underline{4}}} - \frac{2}{3}\pi^2\right) n^2 + \mathcal{O}(n\log n) \quad \text{for } l = n - p \text{ fixed}\,,\\ V_{n,p} &= \left(7 - \frac{2}{3}\pi^2\right) n^2 + \mathcal{O}(n\log n) \quad \text{for } p = \rho n \text{ and } 0 < \rho < 1.\end{aligned} \tag{35}$$

Remark. The instance 'p fixed' of the above corollary appears in [11].

6. Median of three partition – passes

As it says, one takes a random sample of 3 elements and uses the median as the pivot element. See [7, 1] for that in the context of Quicksort and [4, 12] in the context of Quickselect.

The only thing that changes are the *splitting probabilities.* Instead of $\frac{1}{n}$ they are in this instance

$$\frac{(k-1)(n-k)}{\binom{n}{3}}.$$

Now $\binom{n}{3}$ moves as a factor to the left hand side and produces $\frac{1}{6}$ times a third derivative. In this way we find the following differential equation of interest:

$$\begin{aligned}&\frac{1}{6}\frac{\partial^3}{\partial z^3}\Phi(z,u,v) = v(1+u)\left(\frac{\partial}{\partial z}\Phi(z,u,v)\right)^2 + \frac{1-v}{(1-z)^4}\,,\\ &\Phi(0,u,v) = 1\,, \quad \frac{\partial}{\partial z}\Phi(z,u,v)\Big|_{z=0} = 1 + uv\,, \quad \frac{\partial^2}{\partial z^2}\Phi(z,u,v)\Big|_{z=0} = 2(1 + 2uv + u^2 v)\,.\end{aligned} \tag{36}$$

It is natural to consider the function

$$\Psi(z,u,v) = \frac{\partial}{\partial z}\Phi(z,u,v)\,.$$

For the evaluation of the expectated values we define

$$G(z,u) = \frac{\partial}{\partial v}\Psi(z,u,v)\Big|_{v=1}\;;$$

to get it, we have to solve the following differential equation

$$\begin{aligned} &\frac{1}{6}\frac{\partial^2}{\partial z^2}G(z,u) = \frac{2(1+u)^2}{(1-z(1+u))^2}G(z,u) + \frac{(1+u)^3}{(1-z(1+u))^4} - \frac{1}{(1-z)^4}\,, \\ &G(0,u) = u\,, \quad \frac{\partial}{\partial z}G(z,u)\bigg|_{z=0} = 4u + 2u^2\,. \end{aligned} \tag{37}$$

Therefore we obtain

$$\begin{aligned} G(z,u) = {}& \frac{12}{7}\frac{(u+1)^2\,p_1(z,u)}{(1-z(1+u))^3\,u^6}\log\frac{1}{1-z} \\ &+ \frac{60}{7}\frac{(u+1)^2\,(1-z(1+u))^4}{u^6}\log\frac{1}{1-z(1+u)} - \frac{1}{7}\frac{p_2(z,u)}{(1-z(1+u))^3\,(1-z)^2\,u^5}\,, \end{aligned} \tag{38}$$

with

$p_1(z,u) = 5z^7u^7 + 35z^7u^6 + 105z^7u^5 + 175z^7u^4 + 175z^7u^3 + 105z^7u^2 + 35z^7u + 5z^7 - 35z^6u^6 - 210z^6u^5 - 525z^6u^4 - 700z^6u^3 - 525z^6u^2 - 210z^6u - 35z^6 + 105z^5u^5 + 525z^5u^4 + 1050z^5u^3 + 1050z^5u^2 + 525z^5u + 105z^5 - 175z^4u^4 - 700z^4u^3 - 1050z^4u^2 - 700z^4u - 175z^4 + 175z^3u^3 + 525z^3u^2 + 525z^3u + 175z^3 - 105z^2u^2 - 210z^2u - 105z^2 + 35zu + 35z + 2u^7 - 5$,

$p_2(z,u) = 2z^9u^{13} + 13z^9u^{12} + 37z^9u^{11} + 58z^9u^{10} + 62z^9u^9 + 202z^9u^8 + 998z^9u^7 + 2872z^9u^6 + 5018z^9u^5 + 5637z^9u^4 + 4125z^9u^3 + 1910z^9u^2 + 510z^9u + 60z^9 - 4z^8u^{13} - 40z^8u^{12} - 151z^8u^{11} - 298z^8u^{10} - 348z^8u^9 - 554z^8u^8 - 2720z^8u^7 - 9846z^8u^6 - 20998z^8u^5 - 27878z^8u^4 - 23545z^8u^3 - 12360z^8u^2 - 3690z^8u - 480z^8 + 2z^7u^{13} + 41z^7u^{12} + 233z^7u^{11} + 611z^7u^{10} + 867z^7u^9 + 847z^7u^8 + 2611z^7u^7 + 12483z^7u^6 + 34481z^7u^5 + 56212z^7u^4 + 56240z^7u^3 + 34110z^7u^2 + 11550z^7u + 1680z^7 - 14z^6u^{12} - 161z^6u^{11} - 630z^6u^{10} - 1183z^6u^9 - 1170z^6u^8 - 1259z^6u^7 - 6966z^6u^6 - 27635z^6u^5 - 58808z^6u^4 - 72400z^6u^3 - 52060z^6u^2 - 20370z^6u - 3360z^6 + 42z^5u^{11} + 329z^5u^{10} + 917z^5u^9 + 1195z^5u^8 + 745z^5u^7 + 1557z^5u^6 + 10729z^5u^5 + 33377z^5u^4 + 53725z^5u^3 + 47550z^5u^2 + 22050z^5u + 4200z^5 - 70z^4u^{10} - 385z^4u^9 - 740z^4u^8 - 605z^4u^7 - 180z^4u^6 - 1595z^4u^5 - 9610z^4u^4 - 22625z^4u^3 - 26160z^4u^2 - 14910z^4u - 3360z^4 + 70z^3u^9 + 267z^3u^8 + 324z^3u^7 + 127z^3u^6 + 1070z^3u^4 + 4870z^3u^3 + 8210z^3u^2 + 6090z^3u + 1680z^3 - 78z^2u^8 - 149z^2u^7 - 71z^2u^6 - 390z^2u^3 - 1260z^2u^2 - 1350z^2u - 480z^2 + 24zu^8 + 55zu^7 + 31zu^6 + 60zu^2 + 120zu + 60z - 7u^6$.

Reading off the coefficients we find

Theorem 14. *The expected value $E_{n,p}$ of the number of passes in* Multiple Quickselect *with median-of-three partition is given by*

$$\begin{aligned} E_{n,p} = {}& \frac{12}{7}\frac{(n+1)^2\,p}{(n-p+1)\,(n-p+2)}\Big(H_n - H_p\Big) \\ &- \frac{1}{7}\frac{r(n,p)}{n^{\underline{6}}\,(p+6)^{\underline{6}}\,(n+1-p)\,(n+2-p)}\,, \quad \text{for } n \ge 6 \text{ and } 1 \le p \le n\,, \end{aligned} \tag{39}$$

$$\begin{aligned} &E_{1,1} = 1\,,\ \ E_{2,1} = 1\,,\ \ E_{2,2} = 1\,,\ \ E_{3,1} = \frac{5}{3}\,,\ \ E_{3,2} = \frac{7}{3}\,,\ \ E_{3,3} = 3\,, \\ &E_{4,1} = \frac{7}{4}\,,\ \ E_{4,2} = \frac{7}{3}\,,\ \ E_{4,3} = \frac{11}{4}\,,\ \ E_{4,4} = 3\,, \\ &E_{5,1} = \frac{51}{25}\,,\ \ E_{5,2} = \frac{141}{50}\,,\ \ E_{5,3} = \frac{17}{5}\,,\ \ E_{5,4} = \frac{96}{25}\,,\ \ E_{5,5} = \frac{21}{5}\,, \end{aligned} \tag{40}$$

with

$r(n,p) = 12n^8p^7 + 247n^8p^6 + 1995n^8p^5 + 7945n^8p^4 + 15813n^8p^3 + 13048n^8p^2 - 180n^8p - 5040n^8 - 12n^7p^8 - 422n^7p^7 - 5598n^7p^6 - 37058n^7p^5 - 131838n^7p^4 - 244328n^7p^3 - 191592n^7p^2 + 14688n^7p + 60480n^7 + 175n^6p^8 + 4536n^6p^7 + 48328n^6p^6 + 271362n^6p^5 + 850885n^6p^4 + 1429134n^6p^3 + 1000692n^6p^2 - 97272n^6p - 272160n^6 - 945n^5p^8 - 21560n^5p^7 - 200550n^5p^6 - 977060n^5p^5 - 2648205n^5p^4 - 3784340n^5p^3 - 2220540n^5p^2 + 236880n^5p + 544320n^5 + 2275n^4p^8 + 48048n^4p^7 + 403333n^4p^6 + 1708875n^4p^5 + 3752980n^4p^4 + 3919377n^4p^3 + 1317372n^4p^2 - 477540n^4p - 347760n^4 - 2163n^3p^8 - 42098n^3p^7 - 309372n^3p^6 - 991802n^3p^5 - 1085637n^3p^4 + 989548n^3p^3 + 2813412n^3p^2 + 1168272n^3p - 362880n^3 + 70n^2p^8 - 2196n^2p^7 - 73908n^2p^6 - 601272n^2p^5 - 2321130n^2p^4 - 4860324n^2p^3 - 5002312n^2p^2 - 1360368n^2p + 624960n^2 + 600np^8 + 15120np^7 + 127440np^6 + 574560np^5 + 1625400np^4 + 2756880np^3 + 2229600np^2 + 342720np - 241920n + 1440p^7 + 10080p^6 + 10080p^5 - 50400p^4 - 80640p^3 + 40320p^2 + 69120p$.

□

Again we note the special cases.

Corollary 15.

$$\begin{aligned} E_{n,1} &= \frac{12}{7}\frac{n+1}{n}H_n - \frac{131n-2}{49n}\,, \\ E_{n,n} &= \frac{6}{7}n - \frac{1}{7}\,. \end{aligned} \tag{41}$$

Related formulæ can be found in [1, 7, 4].

Corollary 16.

$$E_{n,p} = \frac{12}{7}p\log n - \frac{12}{7}pH_p + \frac{12}{7}p\gamma - \frac{12p^7+247p^6+1995p^5+7945p^4+15813p^3+13048p^2-180p-5040}{7\,(p+6)^{\underline{6}}} \tag{42}$$

$$+\,\mathcal{O}\left(\frac{\log n}{n}\right) \quad \text{for } p \text{ fixed}\,,$$

$$\begin{aligned} E_{n,p} &= \frac{6}{7}n - \frac{2}{7}l - \frac{1}{7} + \mathcal{O}\left(\frac{\log n}{n}\right) \ \text{ for } l = n-p \text{ fixed}\,, \\ E_{n,p} &= \frac{12}{7}\frac{\rho(\rho-1-\log\rho)}{(1-\rho)^2}n + \mathcal{O}(1) \ \text{ for fixed } \rho = \frac{p}{n} \text{ with } 0<\rho<1\,. \end{aligned} \tag{43}$$

7. Median of three – comparisons

In this instance the differential equation is

$$\begin{aligned} &\frac{1}{6}\frac{\partial^3}{\partial z^3}\Phi(z,u,v) = (1+u)\left(\frac{\partial}{\partial z}\Phi(zv,u,v)\right)^2 + \frac{1}{(1-z)^4} - \frac{v^2}{(1-zv)^4}\,, \\ &\Phi(0,u,v)=1\,, \quad \frac{\partial}{\partial z}\Phi(z,u,v)\Big|_{z=0} = 1+u\,, \quad \frac{\partial^2}{\partial z^2}\Phi(z,u,v)\Big|_{z=0} = 2(1+2uv+u^2v)\,. \end{aligned} \tag{44}$$

With

$$\Psi(z,u,v) = \frac{\partial}{\partial z}\Phi(z,u,v)$$

and

$$G(z,u) = \frac{\partial}{\partial v}\Psi(z,u,v)\Big|_{v=1}$$

we get

$$\begin{aligned} &\frac{1}{6}\frac{\partial^2}{\partial z^2}G(z,u) = \frac{2(1+u)^2}{(1-z(1+u))^2}G(z,u) + \frac{2(1+u)^3(1+z(1+u))}{(1-z(1+u))^5} - \frac{2(1+z)}{(1-z)^5}\,, \\ &G(0,u)=0\,, \quad \frac{\partial}{\partial z}G(z,u)\Big|_{z=0} = 4u+2u^2\,. \end{aligned} \tag{45}$$

and furthermore

$$\begin{aligned} G(z,u) = {} & \frac{24}{7}\frac{(1+u)^2\,p_1(z,u)}{(1-z(1+u))^3\,u^7}\log\frac{1}{1-z} - \frac{24}{7}\frac{(1+u)\,p_2(z,u)}{(1-z(1+u))^3\,u^7}\log\frac{1}{1-z(1+u)} \\ & - \frac{2}{49}\frac{z\,g_3(z,u)}{(1-z(1+u))^3\,(1-z)^3\,u^6}\,, \end{aligned} \tag{46}$$

with

$p_1(z,u) = 10z^7u^8 + 85z^7u^7 + 315z^7u^6 + 665z^7u^5 + 875z^7u^4 + 735z^7u^3 + 385z^7u^2 + 115z^7u + 15z^7 - 70z^6u^7 - 525z^6u^6 - 1680z^6u^5 - 2975z^6u^4 - 3150z^6u^3 - 1995z^6u^2 - 700z^6u - 105z^6 + 210z^5u^6 + 1365z^5u^5 + 3675z^5u^4 + 5250z^5u^3 + 4200z^5u^2 + 1785z^5u + 315z^5 - 350z^4u^5 - 1925z^4u^4 - 4200z^4u^3 - 4550z^4u^2 - 2450z^4u - 525z^4 + 350z^3u^4 + 1575z^3u^3 + 2625z^3u^2 + 1925z^3u + 525z^3 - 210z^2u^3 - 735z^2u^2 - 840z^2u - 315z^2 + 70zu^2 + 175zu + 105z - 3u^8 - u^7 - 10u - 15$,

$p_2(z,u) = 10z^7u^9 + 95z^7u^8 + 400z^7u^7 + 980z^7u^6 + 1540z^7u^5 + 1610z^7u^4 + 1120z^7u^3 + 500z^7u^2 + 130z^7u + 15z^7 - 70z^6u^8 - 595z^6u^7 - 2205z^6u^6 - 4655z^6u^5 - 6125z^6u^4 - 5145z^6u^3 - 2695z^6u^2 - 805z^6u - 105z^6 + 210z^5u^7 + 1575z^5u^6 + 5040z^5u^5 + 8925z^5u^4 + 9450z^5u^3 + 5985z^5u^2 + 2100z^5u + 315z^5 - 350z^4u^6 - 2275z^4u^5 - 6125z^4u^4 - 8750z^4u^3 - 7000z^4u^2 - 2975z^4u - 525z^4 + 350z^3u^5 + 1925z^3u^4 + 4200z^3u^3 + 4550z^3u^2 + 2450z^3u + 525z^3 - 210z^2u^4 - 945z^2u^3 - 1575z^2u^2 - 1155z^2u - 315z^2 + 70zu^3 + 245zu^2 + 280zu + 105z - u^7 - 10u^2 - 25u - 15$,

$p_3(z,u) = 2z^9u^{14} + 16z^9u^{13} + 49z^9u^{12} + 91z^9u^{11} + 14z^9u^{10} - 2450z^9u^9 - 17374z^9u^8 - 60926z^9u^7 - 130534z^9u^6 - 184296z^9u^5 - 176127z^9u^4 - 113365z^9u^3 - 47250z^9u^2 - 11550z^9u - 1260z^9 - 6z^8u^{14} - 62z^8u^{13} - 245z^8u^{12} - 518z^8u^{11} - 434z^8u^{10} + 6804z^8u^9 + 60998z^8u^8 + 253940z^8u^7 + 631002z^8u^6 + 1015546z^8u^5 + 1091195z^8u^4 + 780850z^8u^3 + 358470z^8u^2 + 95760z^8u + 11340z^8 + 6z^7u^{14} + 90z^7u^{13} + 483z^7u^{12} + 1260z^7u^{11} + 1701z^7u^{10} - 5439z^7u^9 - 78855z^7u^8 - 415647z^7u^7 - 1244265z^7u^6 - 2342319z^7u^5 - 2882880z^7u^4 - 2326065z^7u^3 - 1189230z^7u^2 - 350280z^7u - 45360z^7 - 2z^6u^{14} - 58z^6u^{13} - 469z^6u^{12} - 1652z^6u^{11} - 2989z^6u^{10} - 742z^6u^9 + 43617z^6u^8 + 332316z^6u^7 + 1274453z^6u^6 + 2914716z^6u^5 + 4216380z^6u^4 + 3907820z^6u^3 + 2256450z^6u^2 + 740880z^6u + 105840z^6 + 14z^5u^{13} + 224z^5u^{12} + 1211z^5u^{11} + 2961z^5u^{10} + 3640z^5u^9 - 7161z^5u^8 - 128646z^5u^7 - 709779z^5u^6 - 2096234z^5u^5 - 3701635z^5u^4 - 4043165z^5u^3 - 2685690z^5u^2 - 996660z^5u - 158760z^5 - 42z^4u^{12} - 462z^4u^{11} - 1743z^4u^{10} - 2898z^4u^9 - 2205z^4u^8 + 18648z^4u^7 + 201453z^4u^6 + 860622z^4u^5 + 1971627z^4u^4 +$

$2635290z^4u^3+2065350z^4u^2+882000z^4u+158760z^4+70z^3u^{11}+560z^3u^{10}+1484z^3u^9+1568z^3u^8+574z^3u^7-22330z^3u^6-183015z^3u^5-609210z^3u^4-1058575z^3u^3-1015770z^3u^2-511560z^3u-105840z^3-70z^2u^{10}-868z^2u^9-1526z^2u^8-728z^2u^7+14980z^2u^5+96110z^2u^4+243040z^2u^3+303030z^2u^2+186480z^2u+45360z^2+672zu^9+1491zu^8+819zu^7-5460zu^4-26670zu^3-48300zu^2-38430zu-11340z-252u^9-553u^8-350u^7+840u^3+2940u^2+3360u+1260.$

Reading off the coefficients we get

Theorem 17. *The expected value $E_{n,p}$ of the number of comparisons in* Multiple Quickselect *with median-of-three partition is given by*

$$\begin{aligned} E_{n,p} &= -\frac{12}{7}\frac{(n+1)\,(4n-p+5)\,p}{(n-p+1)\,(n-p+2)}H_n+\frac{12}{7}\frac{(n+1)^2\,(n+2p+2)}{(n-p+2)\,(n-p+1)}H_p \\ &\quad +\frac{1}{49}\frac{r(n,p)}{n^{\underline{6}}\,(p+7)^{\underline{7}}\,(n+1-p)\,(n+2-p)}\,, \quad \text{for } n\geq 6 \text{ and } 1\leq p\leq n\,, \end{aligned} \tag{47}$$

$$\begin{aligned} &E_{1,1}=0\,,\ E_{2,1}=1\,,\ E_{2,2}=1\,,\ E_{3,1}=2\,,\ E_{3,2}=2\,,\ E_{3,3}=2\,, \\ &E_{4,1}=\frac{7}{2}\,,\ E_{4,2}=\frac{23}{6}\,,\ E_{4,3}=4\,,\ E_{4,4}=4\,, \\ &E_{5,1}=\frac{126}{25}\,,\ E_{5,2}=\frac{141}{25}\,,\ E_{5,3}=\frac{148}{25}\,,\ E_{5,4}=6\,,\ E_{5,5}=6\,. \end{aligned} \tag{48}$$

with

$r(n,p)=37n^9p^7+1036n^9p^6+11914n^9p^5+72520n^9p^4+250453n^9p^3+485884n^9p^2+483516n^9p+246960n^9+178n^8p^8+4423n^8p^7+41608n^8p^6+168238n^8p^5+105322n^8p^4-1459913n^8p^3-5040948n^8p^2-6877548n^8p-3210480n^8-215n^7p^9-8651n^7p^8-139688n^7p^7-1178702n^7p^6-5578475n^7p^5-14341019n^7p^4-15631422n^7p^3+8099412n^7p^2+31066920n^7p+16299360n^7+3192n^6p^9+103600n^6p^8+1417654n^6p^7+10600072n^6p^6+46767364n^6p^5+121653280n^6p^4+168818566n^6p^3+84897288n^6p^2-38913336n^6p-40007520n^6-17780n^5p^9-529760n^5p^8-6609827n^5p^7-44872436n^5p^6-179988074n^5p^5-428908760n^5p^4-581222523n^5p^3-387896964n^5p^2-53621316n^5p+43711920n^5+45570n^4p^9+1288672n^4p^8+15030547n^4p^7+93440452n^4p^6+331748032n^4p^5+679515928n^4p^4+772514743n^4p^3+431025588n^4p^2+83709108n^4p+740880n^4-51485n^3p^9-1380029n^3p^8-14864922n^3p^7-80529498n^3p^6-228035325n^3p^5-283648701n^3p^4+46816532n^3p^3+407511188n^3p^2+191861280n^3p-48404160n^3+16758n^2p^9+376590n^2p^8+2608656n^2p^7+2675148n^2p^6-62468994n^2p^5-381548370n^2p^4-945477876n^2p^3-1039134408n^2p^2-367655184n^2p+42477120n^2+3960np^9+169560np^8+2311200np^7+18975600np^6+99068760np^5+310910040np^4+552004560np^3+497069280np^2+161881920np-11854080n+20160p^8+161280p^7+282240p^6-564480p^5-1834560p^4-564480p^3+1532160p^2+967680p.$ ☐

Again, we list the two prominent special cases.

Corollary 18.

$$\begin{aligned} E_{n,1} &= \frac{245n^2+1111n-2}{98n}-\frac{48}{7}\frac{n+1}{n}H_n\,, \\ E_{n,n} &= \frac{12}{7}(n+1)H_n-\frac{173}{49}n+\frac{9}{49}\,. \end{aligned} \tag{49}$$

And here are the asymptotics.

Corollary 19.

$$E_{n,p}=\left(\frac{12}{7}H_p+\frac{37p^7+1036p^6+11914p^5+72520p^4+250453p^3+485884p^2+483516p+246960}{49\,(p+7)^{\underline{7}}}\right)n \tag{50}$$

$$+\,\mathcal{O}\left(\log n\right) \quad \text{for } p \text{ fixed}\,, \tag{51}$$

$$E_{n,p}=\frac{12}{7}n\log n+\frac{84\gamma-173}{49}n+\frac{12}{7}\log n+\frac{51}{49}+\frac{12}{7}\gamma+\mathcal{O}\left(\frac{\log n}{n}\right) \quad \text{for } l=n-p \text{ fixed},,$$

$$\begin{aligned} E_{n,p} &= \frac{12}{7}n\log n+\frac{168\rho\log\rho+84\log\rho+84\gamma\rho^2-168\gamma\rho+84\gamma-215\rho^2+178\rho+37}{49(1-\rho)^2}n \\ &\quad +\mathcal{O}(\log n) \ \text{ for fixed } \rho=\frac{p}{n} \text{ with } 0<\rho<1\,. \end{aligned} \tag{52}$$

8. Binary trees

In this section we briefly sketch – for the reason of comparison – the analogous questions for the instance of binary (Catalan) trees, enumerated by Catalan numbers

$$b_n=\frac{1}{n+1}\binom{2n}{n};$$

note that

$$B(z)=\sum_{n\ge0} b_n z^n=\frac{1-\sqrt{1-4z}}{2z}\,.$$

Then the equation for the number of passes is

$$F(z,u,v)=zv(1+u)F^2(z,u,v)+v+(1-v)B(z)\,. \tag{53}$$

The normalization factor is this time $\binom{n}{p}b_n$. This quadratic equation is of course readily solved; here, we only compute the average. We get

$$\left.\frac{\partial F(z,u,v)}{\partial v}\right|_{v=1}=-B\bigl(z(1+u)\bigr)+\frac{1}{2z}\frac{\sqrt{1-4z}-(1-4z)}{\sqrt{1-4z(1+u)}}\,. \tag{54}$$

A standard evaluation of this function gives after division by the normalization factor.

Theorem 20. *The average number of passes in the model of binary trees is given by*

$$E_{n,p}=\frac{p(n+1)\binom{2p}{p}4^{n+1-p}}{2(n+1-p)\binom{2n}{n}}-\frac{p(2n+1)}{n+1-p}\,. \tag{55}$$

Corollary 21.

$$\begin{aligned}
E_{n,p}&=\frac{2p\binom{2p}{p}}{4^p}\sqrt{\pi n}+\mathcal{O}(1)\quad\text{for } p \text{ fixed}\,,\\
E_{n,p}&=n+\mathcal{O}(1)\ \text{ for } l=n-p \text{ fixed}\,,\\
E_{n,p}&=\frac{2\sqrt{\rho}}{1+\sqrt{\rho}}n+\mathcal{O}(1)\quad\text{for } p=\rho n \text{ and } 0<\rho<1\,.
\end{aligned} \tag{56}$$

In the instance of the comparisons we get

$$F(z,u,v)=z(1+u)F^2(zv,u,v)+v^{-1}+B(z)-v^{-1}B(zv)\,. \tag{57}$$

The equation for the expectations (obtained by differentation w.r.t. v and $v=1$) has the solution

$$G(z,u)=\frac{4}{Y+1}+\frac{1}{Y^2}-\frac{4}{Y}+\frac{4}{Y(X+1)}-\frac{1}{XY}\,, \tag{58}$$

with

$$X=\sqrt{1-4z}\quad\text{and}\quad Y=\sqrt{1-4z(1+u)}\,.$$

After some routine computations we get from that:

Theorem 22. *The average number of comparisons in the model of binary trees is given by*

$$E_{n,p}=\frac{(n+1)4^n}{\binom{2n}{n}}-\frac{(n+1)(n+3p+1)4^{n-p}}{n+1-p}\frac{\binom{2p}{p}}{\binom{2n}{n}}+\frac{2p(2n+1)}{n+1-p}\quad\text{for } 1\le p\le n\,. \tag{59}$$

Corollary 23.

$$\begin{aligned}
E_{n,p}&=\left(1-\frac{\binom{2p}{p}}{4^p}\right)\sqrt{\pi}\,n^{3/2}+\mathcal{O}\left(\sqrt{n}\right)\ \text{ for } p \text{ fixed}\,,\\
E_{n,p}&=\sqrt{\pi}\,n^{3/2}-3n+\mathcal{O}\left(\sqrt{n}\right)\ \text{ for } l=n-p \text{ fixed}\,,\\
E_{n,p}&=\sqrt{\pi}\,n^{3/2}-\frac{1+\sqrt{\rho}+4\rho}{(1+\sqrt{\rho})\sqrt{\rho}}n+\mathcal{O}\left(\sqrt{n}\right)\ \text{ for fixed } \rho=\frac{p}{n} \text{ with } 0<\rho<1\,.
\end{aligned} \tag{60}$$

Acknowledgement. We thank Philippe Flajolet who insisted that such an approach should be feasible. As the paper shows, he was right (as usual).

References

[1] D. H. Greene and D. E. Knuth. *Mathematics for the analysis of algorithms.* Birkhauser, Boston, second edition, 1982.
[2] C. A. R. Hoare. Find (Algorithm 65). *Communications of the ACM*, 4:321–322, 1961.
[3] C. A. R. Hoare. Quicksort. *Computer Journal*, 5:10–15, 1962.
[4] P. Kirschenhofer, C. Martínez, and H. Prodinger. Analysis of Hoare's Find algorithm with median–of–three partition. *Random Structures and Algorithms*, 10:143–156, 1997.
[5] P. Kirschenhofer and H. Prodinger. Comparisons in Hoare's Find algorithm. *Combinatorics, Probability, and Computing*, 1996 (to appear).
[6] D. E. Knuth. Mathematical analysis of algorithms. In *Information Processing 71*, pages 19–27. North Holland Publishing Company, 1972. Proceedings of IFIP Congress, Ljubljana, 1971.
[7] D. E. Knuth. *The Art of Computer Programming*, volume 3: Sorting and Searching. Addison-Wesley, 1973.
[8] J. Lent and H. M. Mahmoud. Average–case analysis of multiple quickselect: An algorithm for finding order statistics. *Statistics and Probability Letters*, 28:299–310, 1996.
[9] L. Lewin, editor. *Structural Properties of Polylogarithms.* American Mathematical Society, 1991.
[10] H. M. Mahmoud, R. M. Modarres, and R. T. Smythe. Analysis of one-sided quicksort: An algorithm for order statistics. *RAIRO, Theoretical Informatics and Applications*, 29:255–276, 1995.
[11] H. M. Mahmoud and R. T. Smythe. Probabilistic analysis of multiple quickselect. *submitted*, 1996.
[12] C. Martínez and H. Prodinger. On the number of descendants and ascendants in random search trees. *Submitted*, 1996.
[13] H. Prodinger. Multiple quickselect — Hoare's Find algorithm for several elements. *Information Processing Letters*, 56:123–129, 1995.

Institut für Algebra und Diskrete Mathematik, Technical University of Vienna, Wiedner Hauptstrasse 8–10, A-1040 Vienna, Austria.
E-mail address: e9125354@@fbma.tuwien.ac.at (Alois Panholzer)
E-mail address: Helmut.Prodinger@@tuwien.ac.at
WWW-address: http://info.tuwien.ac.at/theoinf/proding.htm