JIRSS (2004)
Vol. 3, No. 2, pp 219-249

Measuring Post–Quickselect Disorder

Alois Panholzer[1], Helmut Prodinger[2*], Marko Riedel[3†]

[1]Institut für Diskrete Mathematik und Geometrie, Technische Universität Wien, Wiedner Hauptstraße 8–10, A–1040 Wien, Austria. (Alois.Panholzer@mail.tuwien.ac.at)
[2]The John Knopfmacher Centre for Applicable Analysis and Number Theory, School of Mathematics, University of the Witwatersrand, Private Bag 3, Wits, 2050 Johannesburg, South Africa. (helmut@maths.wits.ac.za)
[3]EDV, Neue Arbeit gGmbH, Gottfried-Keller-Str. 18c, 70435 Stuttgart, Germany. (mriedel@neuearbeit.de)

Abstract. This paper deals with the amount of disorder that is left in a permutation after one of its elements has been selected with quickselect with or without median-of-three pivoting. Five measures of disorder are considered: inversions, cycles of length less than or equal to some m, cycles of any length, expected cycle length, and the distance to the identity permutation. "Grand averages" for each measure of disorder for a permutation after one of its elements has been selected with quickselect, where $1, 2, \ldots, n$ are the elements being permuted, are computed, as well as more specific results.

*Supported by NRF Grant 2053748.
†I would like to dedicate this work to E. Sexauer, Neue Arbeit gGmbH. It would not have been possible without his support.
Received: October 2003, Revised: April 2004
Key words and phrases: Cycles, disorder, inversions, permutations, quickselect.

1 Introduction

Quickselect (sometimes called Hoare's FIND algorithm) is an algorithm that has been extensively studied and uses the principle behind the quicksort algorithm to select one or more elements from a permutation [4, 5, 7, 11]. The goal is to select an order statistic from a permutation. The algorithm selects a pivot (this is either the first element or the median of the first three elements in quickselect with median-of-three pivoting) and splits the data into those elements that are less than the pivot, those that are equal to the pivot and those that are larger than the pivot. If the pivot is the statistic that we wish to find, the algorithm halts. Otherwise it recursively selects the desired order statistic from those elements that are less than, or those that are larger than the pivot.

This paper treats the following question. What amount of disorder is left in a permutation after one of its elements has been selected with quickselect or quickselect with median-of-three pivoting? It seems clear that there should be less disorder than in a random permutation. Those elements that were pivots are in place, and the others are closer to their home position than before quickselect was applied to the permutation. We consider five measures of disorder, that seem to give a good impression of "what happens in the algorithm":

- **Inversions.** Two elements of a permutation such that the one at the lower position is larger than the one a the higher position constitute an inversion. The fewer inversions, the more ordered the permutation.

- **Cycles of length less than or equal to some m.** The value $m = 1$ is of particular interest, because it counts the number of fixed points. The more cycles, the more ordered the permutation.

- **Cycles of any length.** This is like the previous item, except that now all cycles are counted. Once again, the more cycles, the more ordered the permutation.

- **Expected cycle length.** Pick a random element of a random permutation. It belongs to a cycle of some length k. We study the expected value of k. This parameter should decrease after processing by quickselect.

- **Distance to the identity permutation.** Sum the absolute value of the distance of each element to its correct position, taken to some power p. We treat the case $p = 2$. The smaller the distance, the more ordered the permutation.

The goal of this paper is to compute the "grand average" for each measure of disorder for a permutation after one of its elements has been selected with quickselect, where $1, 2, \ldots, n$ are the elements being permuted. More specific results are obtained in the process of computing these "grand averages." The exact definitions of these measures follow in Section 2 resp. in the references given there.

2 Random permutations

As probability model, we always use the random permutation model, which means that all $n!$ permutations of $\{1, 2, \ldots, n\}$ are assumed to appear equally likely as input data for the quickselect algorithm. In order to compute the expected disorder (given by one of our measures considered) after quickselect has been performed, we have to compute first the expected value of each measure for random permutations of n elements, because these expectations enter the computations, as will become clear later. Then we can compare the value of this quantity with the corresponding one after quickselect has been performed; this will be summarized in Section 7. For the computations of the variance as done in Section 6 we also need the second factorial moments for these measures.

The random variable R_n will count—according to the measure of disorder—either the number of *inversions*, the number of *cycles of length less than or equal to some* m, the number of *cycles of any length*, the *expected cycle length*, or the *distance to the identity permutation* of a random permutation of length n.

Let $r_n(v)$ be the probability generating function of R_n, i. e.,

$$r_n(v) = \sum_{k \geq 0} \mathbb{P}\{R_n = k\} v^k. \tag{1}$$

For convenience, we define $r_0(v) = 1$. We further define the bivariate generating function

$$R(z, v) = \sum_{n \geq 0} r_n(v) z^n, \tag{2}$$

and since it always holds that $r_n(1) = 1$, we have $R(z,1) = \frac{1}{1-z}$. Using these functions, the expectation and the second factorial moment of R_n is given via $\mathbb{E}(X_n) = r_n'(1)$ resp. $\mathbb{E}\big(X_n(X_n-1)\big) = r_n''(1)$ and their generating functions via $\sum_{n\geq 0}\mathbb{E}(X_n)z^n = \frac{\partial}{\partial v}R(z,v)\big|_{v=1}$ resp. $\sum_{n\geq 0}\mathbb{E}\big(X_n(X_n-1)\big)z^n = \frac{\partial^2}{\partial v^2}R(z,v)\big|_{v=1}$.

In order to extract coefficients, we will use the following identities (see e. g. [3]):

$$[z^n]\frac{1}{(1-z)^{m+1}}\log\Big(\frac{1}{1-z}\Big) = \binom{n+m}{m}\Big(H_{n+m} - H_m\Big), \tag{3}$$

$$[z^n]\frac{1}{(1-z)^{m+1}}\log^2\Big(\frac{1}{1-z}\Big) = \binom{n+m}{m}\Big((H_{n+m} - H_m)^2 - (H_{n+m}^{(2)} - H_m^{(2)})\Big), \tag{4}$$

where $H_n := \sum_{k=1}^n \frac{1}{k}$ resp. $H_n^{(2)} := \sum_{k=1}^n \frac{1}{k^2}$ denote the first resp. second order harmonic numbers. Throughout this paper, 'log' always denotes the natural logarithm. For our asymptotic study of the respective parameters we will require the following expansions:

$$H_n = \log n + \gamma + \frac{1}{2n} - \frac{1}{12n^2} + \mathcal{O}\Big(\frac{1}{n^4}\Big) \quad \text{and}$$
$$H_n^{(2)} = \frac{\pi^2}{6} - \frac{1}{n} + \frac{1}{2n^2} + \mathcal{O}\Big(\frac{1}{n^3}\Big).$$

2.1 Measures for random permutations

We now list $r_n(v)$ resp. $R(z,v)$ for the five measures under consideration.

- **Inversions.** Here R_n measures the number of inversions of a random permutation. Since it holds for a random permutation that the number of inversions of element k is, independently of the other elements, either $0, 1, \ldots, n-k$ with equal probability (see e. g. [1]), we obtain the following formulæ for $r_n(v)$ resp. $R(z,v)$:

$$r_n(v) = \frac{1}{n!}\prod_{k=1}^{n}\sum_{l=0}^{n-k} v^l = \frac{1}{n!}\prod_{k=0}^{n-1}\sum_{l=0}^{k} v^l = \frac{1}{n!}\prod_{k=0}^{n-1}\frac{1-v^{k+1}}{1-v},$$

$$R(z,v) = \sum_{n\geq 0} r_n(v) z^n = \sum_{n\geq 0} \frac{z^n (v;v)_n}{n!(1-v)^n},$$

where we used the notation $(x;q)_n := (1-x)(1-xq)\cdots(1-xq^{n-1})$. The following relations

$$\mathbb{E}(R_n) = \frac{n(n-1)}{4},\ \mathbb{E}\big(R_n(R_n-1)\big) = \frac{n(n-1)(n-2)(9n+13)}{144}, \tag{5}$$

$$\frac{\partial}{\partial v} R(z,v)\Big|_{v=1} = \frac{1}{4} z^2 \sum_{n\geq 2} n(n-1) z^{n-2} = \frac{1}{2}\frac{z^2}{(1-z)^3},$$

$$\frac{\partial^2}{\partial v^2} R(z,v)\Big|_{v=1} = \frac{z^3(10-z)}{6(1-z)^5}$$

hold.

- **Cycles of length less than or equal to some m.** Here R_n measures the number of cycles of length $\leq m$ for $m \geq 1$, and one can use the decomposition of permutations into cycles to translate this combinatorial decomposition into the following equation for the bivariate generating function $R(z,v)$ (see e. g. [11, p. 353f]):

$$R(z,v) = \exp\bigg(v \sum_{k=1}^{m} \frac{z^k}{k} + \sum_{k=m+1}^{\infty} \frac{z^k}{k}\bigg) = \frac{1}{1-z}\exp\bigg((v-1)\sum_{k=1}^{m}\frac{z^k}{k}\bigg),$$

and thus

$$\frac{\partial}{\partial v} R(z,v)\Big|_{v=1} = \frac{z + \frac{z^2}{2} + \frac{z^3}{3} + \cdots + \frac{z^m}{m}}{1-z}. \tag{6}$$

We therefore obtain

$$\mathbb{E}(R_n) = \sum_{k=0}^{n} [z^k] \left(z + \frac{z^2}{2} + \cdots + \frac{z^m}{m}\right) = \begin{cases} H_m, & \text{for } n \geq m, \\ H_n, & \text{for } n < m. \end{cases} \tag{7}$$

We note the special instance $m = 1$ (counting fixed points), which gives

$$\frac{\partial}{\partial v} R(z,v)\Big|_{v=1} = \frac{z}{1-z}, \quad \text{and} \quad \frac{\partial^2}{\partial v^2} R(z,v)\Big|_{v=1} = \frac{z^2}{1-z}. \tag{8}$$

Here one gets

$$\mathbb{E}(R_n) = 1, \text{ for } n \geq 1, \quad \text{and} \quad \mathbb{E}\big(R_n(R_n-1)\big) = 1, \text{ for } n \geq 2. \tag{9}$$

- **Cycles of any length.** Now R_n measures the number of cycles of any length in a random permutation. The generating function $R(z,v)$ can be found e. g. in [11, p. 351]:

$$R(z,v) = \frac{1}{(1-z)^v}.$$

This immediately gives

$$\frac{\partial}{\partial v} R(z,v)\Big|_{v=1} = \frac{1}{1-z} \log \frac{1}{1-z},$$
$$\frac{\partial^2}{\partial v^2} R(z,v)\Big|_{v=1} = \frac{1}{1-z} \left(\log \frac{1}{1-z}\right)^2,$$

and further

$$\mathbb{E}(R_n) = H_n \quad \text{and} \quad \mathbb{E}\big(R_n(R_n-1)\big) = H_n^2 - H_n^{(2)}. \tag{10}$$

- **Expected cycle length.** Here R_n measures the expected length of a cycle in a random permutation. In order to treat this parameter via generating functions, it is easier to consider the random variable "cycle length sum" $\tilde{R}_n := nR_n$ of a random permutation and the probability generating function

$$\tilde{r}_n(v) := \sum_{k\geq 0} \mathbb{P}\{\tilde{R}_n = k\} v^k.$$

Since every cycle of length equal to k gives exactly k^2 as contribution to the cycle length sum, we obtain for the bivariate generating function $\tilde{R}(z,v) := \sum_{n\geq 0} \tilde{r}_n(v) z^n$ the equation

$$\tilde{R}(z,v) = \exp\left(\sum_{k\geq 1} v^{k^2} \frac{z^k}{k}\right).$$

This leads to

$$\frac{\partial}{\partial v} \tilde{R}(z,v)\Big|_{v=1} = \frac{z}{(1-z)^3} \quad \text{and} \quad \frac{\partial^2}{\partial v^2} \tilde{R}(z,v)\Big|_{v=1} = \frac{7z^2}{(1-z)^5}. \tag{11}$$

Extracting coefficients gives

$$\mathbb{E}(\tilde{R}_n) = \binom{n+1}{2} \quad \text{and} \quad \mathbb{E}\big(\tilde{R}_n(\tilde{R}_n - 1)\big) = 7\binom{n+2}{4},$$

which leads finally to

$$\mathbb{E}(R_n) = \frac{n+1}{2} \text{ and } \mathbb{E}(R_n^2) = \frac{7(n+2)(n+1)(n-1)}{24n} + \frac{n+1}{2n}. \tag{12}$$

- **Distance to the identity permutation.** Now $R_{n,p}$ measures the distance to the identity permutation, where we define for $p \geq 1$ the distance $d_p(\pi)$ for a permutation $\pi_1\pi_2 \dots \pi_n \in S_n$ of size n by

$$d_p(\pi) := \sum_{k=1}^{n} |k - \pi_k|^p. \tag{13}$$

We have

$$\mathbb{E}(R_{n,p}) = \frac{1}{n} \sum_{1 \leq k,a \leq n} |k - a|^p \tag{14a}$$

and

$$\mathbb{E}(R_{n,p}^2) = \frac{1}{n} \sum_{1 \leq k,a \leq n} |k-a|^{2p} + \frac{1}{n(n-1)} \sum_{\substack{1 \leq k,l,a,b \leq n \\ k \neq l, a \neq b}} |k-a|^p |l-b|^p. \tag{14b}$$

These formulæ can be obtained by averaging equation (13):

$$\begin{aligned}
\mathbb{E}(R_{n,p}) &= \frac{1}{n!} \sum_{\pi \in S_n} d_p(\pi) = \frac{1}{n!} \sum_{\pi \in S_n} \sum_{k=1}^{n} |k - \pi_k|^p \\
&= \frac{1}{n!} \sum_{k=1}^{n} \sum_{a=1}^{n} \sum_{\pi \in S_n, \pi_k = a} |k - a|^p = \frac{1}{n} \sum_{1 \leq k,a \leq n} |k - a|^p
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}(R_{n,p}^2) &= \frac{1}{n!} \sum_{\pi \in S_n} d_p^2(\pi) \\
&= \frac{1}{n!} \sum_{\pi \in S_n} \left(\sum_{k=1}^{n} |k - \pi_k|^{2p} + \sum_{\substack{1 \leq k,l \leq n \\ k \neq l}} |k - \pi_k|^p |l - \pi_l|^p \right)
\end{aligned}$$

$$= \frac{1}{n} \sum_{1 \le k,a \le n} |k-a|^{2p} + \frac{1}{n!} \sum_{\substack{1 \le k,l \le n \\ k \ne l}} \sum_{\substack{\pi \in S_n \\ \pi_k = a, \pi_l = b}} |k-\pi_k|^p |l-\pi_l|^p$$
$$= \frac{1}{n} \sum_{1 \le k,a \le n} |k-a|^{2p} + \frac{1}{n(n-1)} \sum_{\substack{1 \le k,l,a,b \le n \\ k \ne l, a \ne b}} |k-a|^p |l-b|^p.$$

We denote by $r_{n,p}(v)$ the probability generating function of the random variable $R_{n,p}$ and obtain from (14a)

$$\mathbb{E}(R_{n,p}) = r'_{n,p}(1) = \frac{2}{n} \sum_{k=1}^{n-1} k^p \, (n-k). \tag{15}$$

Recall that the sums appearing can be expressed with the Bernoulli polynomials $B_p(n)$ (see e. g. [2]):

$$\sum_{k=0}^{n} k^p = \frac{1}{p+1} \big(B_{p+1}(n) - B_{p+1}(0) \big).$$

Throughout this paper, we restrict ourselves to the parameter $p = 2$ (thus $R_n := R_{n,2}$ and $r_n(v) := r_{n,2}(v)$), which gives

$$\mathbb{E}(R_n) = \frac{1}{6} n(n-1)(n+1) \quad \text{and} \quad \mathbb{E}(R_n^2) = \frac{1}{36} n^3 (n-1)(n+1)^2. \tag{16}$$

We also get

$$\frac{\partial}{\partial v} R(z,v) \Big|_{v=1} = \sum_{n \ge 0} \binom{n+1}{3} z^n = \frac{z^2}{(1-z)^4} \quad \text{and}$$
$$\frac{\partial^2}{\partial v^2} R(z,v) \Big|_{v=1} = \frac{z^2 (5z^3 + 13z^2 + 39z + 3)}{3(1-z)^7}.$$

3 Recurrence relations

Next we will study the random variables $Q_{n,j}$ that measure the disorder (measured by one of the five parameters considered in this paper) of the resulting permutation after the element j has been selected from a random permutation of size n via quickselect ($1 \le j \le n$). We introduce by

$$q_{n,j}(v) = \sum_{k \ge 0} \mathbb{P}\{Q_{n,j} = k\} v^k \tag{17}$$

the probability generating function of $Q_{n,j}$. In the instance where the expected cycle length is considered, it is advantageous for a generating functions approach to introduce the random variable $\tilde{Q}_{n,j} := nQ_{n,j}$ and the probability generating function $\tilde{q}_{n,j} := \sum_{k\geq 0} \mathbb{P}\{\tilde{Q}_{n,j} = k\}v^k$.

3.1 Ordinary quickselect algorithm

We will now translate the recursive nature of the quickselect algorithm into a recurrence for the functions $q_{n,j}(v)$. The probability that p with $1 \leq p \leq n$ is chosen as pivot element in the partitioning phase is $\frac{1}{n}$, independently of p. Since after the partitioning phase, p is at its correct position, the contribution of p to the measures inversions and distance to the identity permutation is 0, but its contribution is 1 to the measures fixed points and cycles, and it also contributes 1 to the sum of the cycle lengths. Distinguishing for $1 \leq j \leq n$ and $n \geq 1$ the cases $j = p$, $j < p$ and $j > p$, we immediately get the following recurrences.

- **Inversions and distance to the identity permutation:**

$$\begin{aligned} q_{n,j}(v) &= \frac{1}{n}\sum_{p=1}^{j-1} r_{p-1}(v)q_{n-p,j-p}(v) + \frac{1}{n}r_{j-1}(v)r_{n-j}(v) \\ &\quad + \frac{1}{n}\sum_{p=j+1}^{n} q_{p-1,j}(v)r_{n-p}(v), \quad \text{with } q_{1,1}(v) = 1. \end{aligned} \tag{18a}$$

- **Fixed points and cycles:**

$$\begin{aligned} q_{n,j}(v) &= \frac{v}{n}\sum_{p=1}^{j-1} r_{p-1}(v)q_{n-p,j-p}(v) + \frac{v}{n}r_{j-1}(v)r_{n-j}(v) \\ &\quad + \frac{v}{n}\sum_{p=j+1}^{n} q_{p-1,j}(v)r_{n-p}(v), \quad \text{with } q_{1,1}(v) = v. \end{aligned} \tag{18b}$$

- **Expected cycle length:**

$$\begin{aligned} \tilde{q}_{n,j}(v) &= \frac{v}{n}\sum_{p=1}^{j-1} \tilde{r}_{p-1}(v)\tilde{q}_{n-p,j-p}(v) + \frac{v}{n}\tilde{r}_{j-1}(v)\tilde{r}_{n-j}(v) \\ &\quad + \frac{v}{n}\sum_{p=j+1}^{n} \tilde{q}_{p-1,j}(v)\tilde{r}_{n-p}(v), \quad \text{with } \tilde{q}_{1,1}(v) = v. \end{aligned} \tag{18c}$$

3.2 Quickselect with median-of-three pivoting

The only difference to ordinary quickselect is, that for $n \geq 3$, the probability of p being selected as the pivot element is $\binom{n}{3}^{-1}(p-1)(n-p)$. Furthermore we assume, that for small subfiles $n \leq 2$, the ordinary quickselect algorithm is applied to get the required order statistic. This leads for $1 \leq j \leq n$ and $n \geq 3$, again by distinguishing the position p relative to j, to the following recurrences.

- **Inversions and distance to the identity permutation:**

$$\begin{aligned} q_{n,j}(v) = {} & \binom{n}{3}^{-1} \sum_{p=1}^{j-1}(p-1)r_{p-1}(v)(n-p)q_{n-p,j-p}(v) \\ & + \binom{n}{3}^{-1} (j-1)r_{j-1}(v)(n-j)r_{n-j}(v) \\ & + \binom{n}{3}^{-1} \sum_{p=j+1}^{n} (p-1)q_{p-1,j}(v)(n-p)r_{n-p}(v), \qquad \text{(19a)} \\ & \text{with } \ q_{1,1}(v) = 1, \ q_{2,1}(v) = q_{2,2}(v) = 1. \end{aligned}$$

- **Fixed points and cycles:**

$$\begin{aligned} q_{n,j}(v) = {} & v \binom{n}{3}^{-1} \sum_{p=1}^{j-1}(p-1)r_{p-1}(v)(n-p)q_{n-p,j-p}(v) \\ & + v \binom{n}{3}^{-1} (j-1)r_{j-1}(v)(n-j)r_{n-j}(v) \\ & + v \binom{n}{3}^{-1} \sum_{p=j+1}^{n} (p-1)q_{p-1,j}(v)(n-p)r_{n-p}(v), \qquad \text{(19b)} \\ & \text{with } \ q_{1,1}(v) = v, \ q_{2,1}(v) = q_{2,2}(v) = v^2. \end{aligned}$$

- **Expected cycle length:**

$$\begin{aligned} \tilde{q}_{n,j}(v) = {} & v \binom{n}{3}^{-1} \sum_{p=1}^{j-1}(p-1)\tilde{r}_{p-1}(v)(n-p)\tilde{q}_{n-p,j-p}(v) \\ & + v \binom{n}{3}^{-1} (j-1)\tilde{r}_{j-1}(v)(n-j)\tilde{r}_{n-j}(v) \\ & + v \binom{n}{3}^{-1} \sum_{p=j+1}^{n} (p-1)\tilde{q}_{p-1,j}(v)(n-p)\tilde{r}_{n-p}(v), \qquad \text{(19c)} \\ & \text{with } \ \tilde{q}_{1,1}(v) = v, \ \tilde{q}_{2,1}(v) = \tilde{q}_{2,2}(v) = v^2. \end{aligned}$$

4 Trivariate generating function

To get our results for the parameters $Q_{n,j}$ considered here, we will use a generating functions approach. We will introduce trivariate generating functions

$$F(z,u,v) = \sum_{n\geq 1} \sum_{1\leq j\leq n} q_{n,j}(v) u^j z^n \quad \text{resp.}$$
$$\tilde{F}(z,u,v) = \sum_{n\geq 1} \sum_{1\leq j\leq n} \tilde{q}_{n,j}(v) u^j z^n,$$

from which the recurrence relations (18) will translate into ordinary differential equations.

Recall that $q_{n,j}(1) = \tilde{q}_{n,j}(1) = 1$ and hence

$$\begin{aligned} F(z,u,1) &= \tilde{F}(z,u,1) \\ &= \sum_{1\leq j\leq n} u^j z^n = \frac{1}{1-z} \sum_{1\leq j} (zu)^j = \frac{zu}{(1-z)(1-zu)}. \end{aligned}$$

4.1 Ordinary quickselect

From the recurrence relations (18) we get the following first-order linear differential equations.

• **Inversions and distance to the identity permutation:**

$$\begin{aligned} \frac{\partial}{\partial z} F(z,u,v) &= uR(zu,v)F(z,u,v) \\ &\quad + uR(zu,v)R(z,v) + R(z,v)F(z,u,v). \end{aligned} \tag{20a}$$

• **Fixed points and cycles:**

$$\begin{aligned} \frac{\partial}{\partial z} F(z,u,v) &= uv\, R(zu,v)F(z,u,v) \\ &\quad + uv\, R(zu,v)R(z,v) + v\, R(z,v)F(z,u,v). \end{aligned} \tag{20b}$$

• **Expected cycle length:**

$$\begin{aligned} \frac{\partial}{\partial z} \tilde{F}(z,u,v) &= uv\, \tilde{R}(zu,v)\tilde{F}(z,u,v) \\ &\quad + uv\, \tilde{R}(zu,v)\tilde{R}(z,v) + v\, \tilde{R}(z,v)\tilde{F}(z,u,v). \end{aligned} \tag{20c}$$

In order to get the expected value of our parameters, we differentiate the functions $F(z,u,v)$ resp. $\tilde{F}(z,u,v)$ w. r. t. v and evaluate at $v=1$. We define the functions

$$G(z,u) = \Big(\frac{\partial}{\partial v} F(z,u,v)\Big)\Big|_{v=1} \text{ resp. } \tilde{G}(z,u) = \Big(\frac{\partial}{\partial v} \tilde{F}(z,u,v)\Big)\Big|_{v=1}, \tag{21}$$

and obtain $\mathbb{E}(Q_{n,j}) = [z^n u^j] G(z,u)$ for all parameters considered except the expected cycle length, where we obtain $\mathbb{E}(Q_{n,j}) = \frac{1}{n}\mathbb{E}(\tilde{Q}_{n,j}) = \frac{1}{n}[z^n u^j]\tilde{G}(z,u)$.

Using also the equations for $R(z,v)$ and $\tilde{R}(z,v)$ of Section 2, we obtain first-order differential equations for $G(z,u)$ resp. $\tilde{G}(z,u)$ with initial values $G(0,u)=0$ resp. $\tilde{G}(0,u)=0$, that are given in Table 1.

Table 1: Differential equations for $G(z,u)$ resp. $\tilde{G}(z,u)$.

Inversions	$\frac{\partial}{\partial z}G(z,u) = \frac{1}{2}uz^2\frac{A_1}{(1-zu)^4(1-z)^4} + \frac{1+u-2zu}{(1-zu)(1-z)}G(z,u)$
Fixed points	$\frac{\partial}{\partial z}G(z,u) = \frac{u(1+z+uz-3uz^2)}{(1-zu)^2(1-z)^2} + \frac{1+u-2zu}{(1-zu)(1-z)}G(z,u)$
Cycles of length $\leq m$	$\frac{\partial}{\partial z}G(z,u) = \frac{u}{(1-zu)(1-z)^2}\sum_{k=1}^{m}\frac{z^k}{k} + \frac{u}{(1-zu)^2(1-z)}\sum_{k=1}^{m}\frac{(uz)^k}{k} + \frac{u-u^2z^2}{(1-zu)^2(1-z)^2} + \frac{1+u-2zu}{(1-zu)(1-z)}G(z,u)$
Cycles	$\frac{\partial}{\partial z}G(z,u) = \frac{u}{(1-zu)(1-z)^2}\log\frac{1}{1-z} + \frac{u}{(1-zu)^2(1-z)}\log\frac{1}{1-zu} + \frac{u-u^2z^2}{(1-zu)^2(1-z)^2} + \frac{1+u-2zu}{(1-zu)(1-z)}G(z,u)$
Expected cycle length	$\frac{\partial}{\partial z}G(z,u) = -u\frac{A_2}{(1-z)^4(1-zu)^4} + \frac{1+u-2zu}{(1-z)(1-zu)}G(z,u)$
Distance to identity	$\frac{\partial}{\partial z}G(z,u) = \frac{1+u-2zu}{(1-zu)(1-z)}G(z,u) + uz^2\frac{A_3}{(1-zu)^5(1-z)^5}$

The quantities A_1, A_2, A_3 are defined as $A_1 = 1 - 3uz - 3u^2z + 6u^2z^2 - u^2z^3 - u^3z^3 + u^2$, $A_2 = u^3z^6 - 2u^3z^5 - 2u^2z^5 + 2u^3z^4 + 3u^2z^4 + 2uz^4 - 3u^2z^3 - 3uz^3 - u^2z^2 + 3uz^2 - z^2 + uz + z - 1$ and $A_3 = u^4z^4 + u^2z^4 - 4u^3z^3 - 4u^2z^3 + 12u^2z^2 - 4u^2z - 4uz + u^2 + 1$.

Solving these equations gives the expressions for $G(z,u)$ resp. $\tilde{G}(z,u)$, that are summarized in Table 2.

Table 2: Solutions for $G(z,u)$ resp. $\tilde{G}(z,u)$.

Inversions	$G(z,u) = \frac{1}{2}\frac{u}{(1-z)(1-zu)}\log\frac{1}{1-z} + \frac{1}{2}\frac{1}{(1-z)(1-zu)}\log\frac{1}{1-zu} + \frac{1}{4}zu\frac{A_4}{(1-zu)^3(1-z)^3}$
Fixed points	$G(z,u) = 2\,\frac{u}{(1-zu)(1-z)}\log\frac{1}{1-z} + 2\,\frac{1}{(1-zu)(1-z)}\log\frac{1}{1-zu} - 3\,\frac{zu}{(1-zu)(1-z)}$
Cycles of length $\leq m$	$G(z,u) = (1+H_m)\,\frac{u}{(1-zu)(1-z)}\log\frac{1}{1-z} + (1+H_m)\,\frac{1}{(1-zu)(1-z)}\log\frac{1}{1-zu} - \Big((1+2H_m)zu + \sum_{k=2}^{m}\frac{(H_m-H_{k-1})(uz^k+u^kz^k)}{k}\Big) \times \frac{1}{(1-zu)(1-z)}$
Cycles	$G(z,u) = \frac{u\log^2\frac{1}{1-z}+\log^2\frac{1}{1-zu}}{2(1-zu)(1-z)} - \frac{zu}{(1-zu)(1-z)} + \frac{u\ \log\frac{1}{1-z}+\log\frac{1}{1-zu}}{(1-zu)(1-z)}$

The abbreviations are $A_4 = 3z^3u^2 + 3z^3u - 2z^2u^2 - 12z^2u - 2z^2 + 7zu + 7z - 4$, $A_5 = -2z^4u^2 + 5z^3u^2 + 5z^3u - 2z^2u^2 - 12z^2u - 2z^2 + 5zu + 5z - 2$, $A_6 = u^3z^3 + u^2z^3 - 6u^2z^2 + 3u^2z + 3uz - u^2 - 1$.

With the approach presented here that leads to explicit solutions, we can treat parameters that satisfy the general recursive structure

$$na_{n,j} = T_{n,j} + \sum_{k=1}^{j-1} a_{n-k,j-k} + \sum_{k=j+1}^{n} a_{k-1,j},$$

Expected cycle length	$G(z,u)$	$= \frac{u}{(1-z)(1-zu)} \log \frac{1}{1-z} + \frac{1}{(1-z)(1-zu)} \log \frac{1}{1-zu} + \frac{1}{2} zu \frac{A_5}{(1-zu)^3(1-z)^3}$
Distance to identity	$G(z,u)$	$= -\frac{1}{3}\, uz^3 \, \frac{A_6}{(1-zu)^4(1-z)^4}$

with toll functions $T_{n,j} = f_1(n) + f_2(j) + f_3(n-j)$, where $f_i(m)$ are linear combinations of terms $m^p H_m$ and m^p with nonnegative integers p. This recurrence with such "harmonic" toll functions $f_1(n)$ (but $f_2(j) = f_3(n-j) = 0$) was studied in [10].

4.2 Quickselect with median-of-three pivoting

From the recurrence relations (19) we get the following third-order linear differential equations, where we use the abbreviations $R'(z,v) := \frac{\partial}{\partial z} R(z,v)$, resp. $\tilde{R}'(z,v) := \frac{\partial}{\partial z} \tilde{R}(z,v)$.

• **Inversions and distance to the identity permutation:**

$$\frac{\partial^3}{\partial z^3} F(z,u,v) = 6u^2 R'(zu,v) \frac{\partial}{\partial z} F(z,u,v) + 6u^2 R'(zu,v) R'(z,v) + 6R'(z,v) \frac{\partial}{\partial z} F(z,u,v). \tag{22a}$$

• **Fixed points and cycles:**

$$\frac{\partial^3}{\partial z^3} F(z,u,v) = 6u^2 v R'(zu,v) \frac{\partial}{\partial z} F(z,u,v) + 6u^2 v R'(zu,v) R'(z,v) + 6vR'(z,v) \frac{\partial}{\partial z} F(z,u,v). \tag{22b}$$

• **Expected cycle length:**

$$\frac{\partial^3}{\partial z^3} \tilde{F}(z,u,v) = 6u^2 v \tilde{R}'(zu,v) \frac{\partial}{\partial z} \tilde{F}(z,u,v) + 6u^2 v \tilde{R}'(zu,v) \tilde{R}'(z,v) + 6v\tilde{R}'(z,v) \frac{\partial}{\partial z} \tilde{F}(z,u,v). \tag{22c}$$

In order to get the expected value of our parameters, we differentiate the functions $F(z,u,v)$ resp. $\tilde{F}(z,u,v)$ w. r. t. v and evaluate at $v=1$. Here we define the functions

$$\Phi(z,u) = \frac{\partial}{\partial v} \frac{\partial}{\partial z} F(z,u,v) \Big|_{v=1}, \quad \text{resp. } \tilde{\Phi}(z,u) = \frac{\partial}{\partial v} \frac{\partial}{\partial z} \tilde{F}(z,u,v) \Big|_{v=1} \tag{23}$$

and obtain $\mathbb{E}(Q_{n,j}) = \frac{1}{n}[z^{n-1}u^j]\Phi(z,u)$ for all parameters except the expected cycle length, where we obtain $\mathbb{E}(Q_{n,j}) = \frac{1}{n}\mathbb{E}(\tilde{Q}_{n,j}) = \frac{1}{n^2}[z^{n-1}u^j]\tilde{\Phi}(z,u)$.

This leads to second-order linear differential equations for $\Phi(z,u)$ (resp. $\tilde{\Phi}(z,u)$), where the functions $R'(z,1)\big|_{v=1}$ (resp. $\tilde{R}'(z,1)\big|_{v=1}$) appearing are obtained by differentiating the equations of Section 2. These differential equations are given as Table 3. The initial values are given by $\Phi(0,u) = 0$ and $\frac{\partial}{\partial z}\Phi(z,u)|_{z=0} = 0$ for the inversions and the distance to the identity permutation, by $\Phi(0,u) = u$ and $\frac{\partial}{\partial z}\Phi(z,u)|_{z=0} = 4u+4u^2$ for fixed points and cycles, and by $\tilde{\Phi}(0,u) = u$ and $\frac{\partial}{\partial z}\tilde{\Phi}(z,u)|_{z=0} = 4u + 4u^2$ for the expected cycle length.

Table 3: Differential equations for $\Phi(z,u)$ resp. $\tilde{\Phi}(z,u)$.

Inversions	$\frac{\partial^2}{\partial z^2}\Phi(z,u) = \frac{3zu^4(2+zu)(1-uz^2)}{(1-zu)^6(1-z)^2} + \frac{6u^2}{(1-zu)^2}\Phi(z,u) + \frac{3zu^3(2+zu)}{(1-zu)^4(1-z)^2} + \frac{3zu^2(2+z)}{(1-zu)^2(1-z)^4} + \frac{3zu(2+z)(1-z^2u)}{(1-zu)^2(1-z)^6} + \frac{6}{(1-z)^2}\Phi(z,u)$
Fixed points	$\frac{\partial^2}{\partial z^2}\Phi(z,u) = \frac{12u^3(1-z^2u)}{(1-zu)^4(1-z)^2} + \frac{6u^2}{(1-zu)^2}\Phi(z,u) + \frac{18u^2}{(1-zu)^2(1-z)^2} + \frac{12u(1-z^2u)}{(1-zu)^2(1-z)^4} + \frac{6}{(1-z)^2}\Phi(z,u)$
Cycles	$\frac{\partial^2}{\partial z^2}\Phi(z,u) = \frac{12u^3(1-z^2u)}{(1-zu)^4(1-z)^2} + \frac{6u^3(1-z^2u)}{(1-zu)^4(1-z)^2}\log\frac{1}{1-zu} + \frac{6u^2}{(1-zu)^2}\Phi(z,u) + \frac{18u^2}{(1-zu)^2(1-z)^2} + \frac{6u^2}{(1-zu)^2(1-z)^2}\log\frac{1}{1-zu} + \frac{6u^2}{(1-zu)^2(1-z)^2}\log\frac{1}{1-z} + \frac{12u(1-z^2u)}{(1-zu)^2(1-z)^4} + \frac{6u(1-z^2u)}{(1-zu)^2(1-z)^4}\log\frac{1}{1-z} + \frac{6}{(1-z)^2}\Phi(z,u)$

To solve the differential equation

$$\frac{\partial^2}{\partial z^2}\Phi(z,u) - 6\left(\frac{1}{(1-z)^2} + \frac{u^2}{(1-zu)^2}\right)\Phi(z,u) = g(z,u)$$

with different $g(z,u)$'s according to the parameter under consideration, we can transform it into a *hypergeometric* differential equation

Expected cycle length	$\frac{\partial^2}{\partial z^2}\tilde{\Phi}(z,u) = \frac{6u^2(u-z^2u^2)}{(1-zu)^4(1-z)^2} + \frac{6u^2(1+2zu)(u-z^2u^2)}{(1-zu)^6(1-z)^2} + \frac{6u^2}{(1-zu)^2}\tilde{\Phi}(z,u) + \frac{6u^2}{(1-zu)^2(1-z)^2} + \frac{6u^2(1+2zu)}{(1-zu)^4(1-z)^2} + \frac{6u^2(1+2z)}{(1-zu)^2(1-z)^4} + \frac{6(u-z^2u^2)}{(1-zu)^2(1-z)^4} + \frac{6(1+2z)(u-z^2u^2)}{(1-z)^6(1-zu)^2} + \frac{6}{(1-z)^2}\tilde{\Phi}(z,u)$
Distance to identity	$\frac{\partial^2}{\partial z^2}\Phi(z,u) = \frac{12zu^4(1+zu)(1-z^2u)}{(1-zu)^7(1-z)^2} + \frac{6u^2}{(1-zu)^2}\Phi(z,u) + \frac{12zu^4(1+zu)}{(1-zu)^5(1-z)^2} + \frac{12zu^2(1+z)}{(1-zu)^2(1-z)^5} + \frac{12zu(1+z)(1-z^2u)}{(1-zu)^2(1-z)^7} + \frac{6}{(1-z)^2}\Phi(z,u)$

$$t(1-t)\frac{\partial^2}{\partial t^2}G(t,u) - 4(1-2t)\frac{\partial}{\partial t}G(t,u) - 8G(t,u) = \frac{t^3(1-t)^3(1-u)^6}{u^4}g\big(1+\tfrac{1-u}{u}t,u\big).$$

There we used the substitutions

$$\Phi(z,u) = \frac{1}{(1-z)^2(1-zu)^2}E(z,u)\,,$$

$z = 1+\frac{1-u}{u}t$ and $G(t,u) = E\big(1+\frac{1-u}{u}t,u\big)$. This procedure was also used in [6].

The corresponding homogeneous differential equation has the solution

$$G^{\text{hom}}(t,u) = k_1(u)(1-2t) + k_2(u)t^5\Big(1-2t+\frac{10}{7}t^2 - \frac{5}{14}t^3\Big).$$

Solving the inhomogeneous differential equations by variation of the constants, we obtain after back substituting the solutions of the functions $\Phi(z,u)$ resp. $\tilde{\Phi}(z,u)$.[1] Since these solutions are very lengthy, we refrain from printing them here. We only give the structure of $\Phi(z)$ in the instance of cycles (the structure is very similar for the other parameters):

$$\Phi(z) = \frac{P_1(z,u)}{(1-z)^2(1-u)^7(1-uz)^2}\int_{t=0}^{z}\frac{\log\frac{1}{1-t}}{1-ut}dt$$

[1] This technique is described in more detail in [9]. The paper [10] is also of relevance here.

$$
\begin{aligned}
&+ \frac{P_2(z,u)}{(1-z)^2(1-u)^7(1-uz)^2}\log^2\frac{1}{1-z} \qquad (24)\\
&+ \frac{P_3(z,u)}{(1-z)^2(1-u)^7(1-uz)^2}\log\frac{1}{1-z}\\
&+ \frac{P_4(z,u)}{(1-z)^2(1-u)^7(1-uz)^2}\log^2\frac{1}{1-uz}\\
&+ \frac{P_5(z,u)}{(1-z)^2(1-u)^7(1-uz)^2}\log\frac{1}{1-uz}\\
&+ \frac{P_6(z,u)}{(1-u)^7(1-uz)^2}\log\frac{1}{1-z}\log\frac{1}{1-uz}\\
&+ \frac{P_7(z,u)}{(1-z)^2(1-u)^7(1-uz)^2},
\end{aligned}
$$

with certain polynomials $P_i(z,u)$ in z and u for $1 \le i \le 7$. Of course, these solutions are obtained with assistance of a computer algebra system.

5 Extracting coefficients

To obtain the desired expectations $\mathbb{E}(Q_{n,j})$ of the parameters considered, we have to extract coefficients from the generating functions computed in Section 4. We will use the following identities.

$$
[z^n u^j]\frac{1}{(1-z)^{m+1}(1-uz)^{m+1}} = \binom{m+j}{m}\binom{n-j+m}{m}, \quad \text{for } 0 \le j \le n \text{ and } n \ge 0,
$$

$$
[z^n u^j]\frac{u}{(1-uz)(1-z)}\log\frac{1}{1-z} = H_{n+1-j} \quad \text{and}
$$

$$
[z^n u^j]\frac{u}{(1-uz)(1-z)}\log^2\frac{1}{1-z} = H_{n+1-j}^2 - H_{n+1-j}^{(2)}, \quad \text{for } 1 \le j \le n \text{ and } n \ge 1,
$$

$$
[z^n u^j]\frac{1}{(1-uz)(1-z)}\log\frac{1}{1-uz} = H_j \quad \text{and}
$$

$$
[z^n u^j]\frac{1}{(1-uz)(1-z)}\log^2\frac{1}{1-uz} = H_j^2 - H_j^{(2)}, \quad \text{for } 1 \le j \le n \text{ and } n \ge 1.
$$

5.1 Ordinary quickselect

5.1.1 Explicit results

Extracting coefficients from $G(z,u)$ resp. $\tilde{G}(z,u)$ as given in Table 2, we obtain the following explicit results for $\mathbb{E}(Q_{n,j})$.

- Inversions:

$$\mathbb{E}(Q_{n,j}) = \frac{1}{8}(n+1-j)(n-4-j) + \frac{1}{8}j(j-5) + \frac{1}{2}H_{n+1-j} + \frac{1}{2}H_j. \tag{25a}$$

- Fixed points:

$$\mathbb{E}(Q_{n,j}) = 2H_{n+1-j} + 2H_j - 3. \tag{25b}$$

- Cycles of length up to m when $m-1 < j < n-m+2$:

$$\mathbb{E}(Q_{n,j}) = (1+H_m)(H_{n+1-j} + H_j) - \left(1 + H_m^2 + H_m^{(2)}\right). \tag{25c}$$

- Cycles:

$$\mathbb{E}(Q_{n,j}) = \frac{1}{2}H_{n+1-j}^2 - \frac{1}{2}H_{n+1-j}^{(2)} + \frac{1}{2}H_j^2 - \frac{1}{2}H_j^{(2)} - 1 + H_{n+1-j} + H_j. \tag{25d}$$

- Expected cycle length:

$$\mathbb{E}(Q_{n,j}) = \frac{1}{n}\Big(\frac{1}{4}(n+1-j)(n-j) + \frac{1}{4}j(j-1) - 1 + H_{n+1-j} + H_j\Big). \tag{25e}$$

- Distance to the identity permutation:

$$\mathbb{E}(Q_{n,j}) = \frac{1}{18}\left(n^3 - n\right) - \frac{1}{6}(n-1)j(n+1-j). \tag{25f}$$

5.1.2 Grand averages

The grand averages are defined by

$$E_n := \mathbb{E}\Big(\frac{1}{n}\sum_{j=1}^{n} Q_{n,j}\Big). \tag{26}$$

They can be obtained by

$$E_n = \frac{1}{n}[z^n]G(z,1) \tag{27a}$$

for all parameters except the expected cycle length, where we have

$$E_n = \frac{1}{n^2}[z^n]\tilde{G}(z,1). \tag{27b}$$

The required generating functions are given in Table 4.

Table 4: The functions $G(z,1)$ resp. $\tilde{G}(z,1)$.

Inversions	$G(z,1) = \frac{1}{(1-z)^2}\log\frac{1}{1-z} + \frac{1}{2}\frac{3z^2-2z}{(1-z)^4}$
Fixed points	$G(z,1) = \frac{4}{(1-z)^2}\log\frac{1}{1-z} - \frac{3z}{(1-z)^2}$
Cycles up to length m	$G(z,1) = 2(1+H_m)\frac{1}{(1-z)^2}\log\frac{1}{1-z} - \Big((1+2H_m)z + 2\sum_{k=2}^{m}\frac{1}{k}(H_m-H_{k-1})z^k\Big) \times \frac{1}{(1-z)^2}$
Cycles	$G(z,1) = \frac{1}{(1-z)^2}\log^2\frac{1}{1-z} - \frac{z}{(1-z)^2} + 2\frac{1}{(1-z)^2}\log\frac{1}{1-z}$
Expected cycle length	$\tilde{G}(z,1) = 2\frac{1}{(1-z)^2}\log\frac{1}{1-z} + \frac{-z(z^2-3z+1)}{(1-z)^4}$
Distance to identity	$G(z,1) = \frac{2}{3}\frac{z^3}{(1-z)^5}$

Extracting coefficients leads then to the following explicit and asymptotic results for the grand averages E_n.

- Inversions:

$$E_n = \Big(1+\frac{1}{n}\Big)H_n + \frac{n^2-6n-19}{12} \sim \frac{n^2}{12}. \tag{28a}$$

- Fixed points:

$$E_n = 4\Big(1+\frac{1}{n}\Big)H_n - 7 \sim 4\log n. \tag{28b}$$

- Cycles up to length m:

$$\begin{aligned} E_n &= 2(1+H_m)H_n\Big(1+\frac{1}{n}\Big) - 2(1+H_m) - \Big(1+\frac{1}{n}\Big)H_m^2 \\ &\quad - \Big(1+\frac{1}{n}\Big)H_m^{(2)} - 1 + 2\frac{m}{n} \qquad \text{for } n > m-2 \\ &\sim 2(1+H_m). \end{aligned} \tag{28c}$$

- Cycles:

$$E_n = \left(H_n^2 - H_n^{(2)}\right)\left(1 + \frac{1}{n}\right) + \frac{2}{n}H_n - 1 \sim \log^2 n. \tag{28d}$$

- Expected cycle length:

$$E_n = \frac{2}{n}\Big(1 + \frac{1}{n}\Big)H_n + \frac{1}{6}n\Big(1 - \frac{19}{n^2}\Big) \sim \frac{n}{6}. \tag{28e}$$

- Distance to the identity permutation:

$$E_n = \frac{1}{36}(n+1)(n-1)(n-2) \sim \frac{n^3}{36}. \tag{28f}$$

5.2 Quickselect with median-of-three pivoting

5.2.1 Explicit results

By extracting coefficients from $\Phi(z,u)$ resp. $\tilde{\Phi}(z,u)$ as computed in Subsection 4.2, one can obtain explicit results for $\mathbb{E}(Q_{n,j})$ also for quickselect with median-of-three pivoting. For the sake of completeness, we give these lengthy formulæ in the appendix.

They are obtained by using formulæ (3) and (4) under heavy usage of a computer algebra system to simplify and manipulate the resulting expressions. As an example, we consider the instance of cycles, where the formula for $\Phi(z)$ is given as (24), in a bit more detail. Picking for instance the summand

$$s_{n,j} = [z^n u^j]\frac{P_2(z,u)}{(1-z)^2(1-u)^7(1-uz)^2}\log^2\frac{1}{1-z},$$

where $P_2(z,u)$ is a polynomial in z and u, it is apparent that it is sufficient to compute

$$t_{n,j} = [z^n u^j]\frac{1}{(1-z)^2(1-u)^7(1-uz)^2}\log^2\frac{1}{1-z},$$

since the required coefficient $s_{n,j}$ is obtained by a linear combination of such shifted expressions: $s_{n,j} = \sum_{m\le M, l\le L}\alpha_{m,l}t_{n-m,j-l}$, with certain bounds L, M and coefficients $\alpha_{m,l}$.

Using (3), we obtain immediately

$$[z^n u^j]\frac{1}{(1-z)^2(1-u)^7(1-uz)^2}\log^2\frac{1}{1-z}$$
$$= \sum_{k=n-j}^{n}(n-k+1)\binom{j-n+k+6}{6}(k+1)\times$$
$$\times\big((H_{k+1}-1)^2-(H_{k+1}^{(2)}-1)\big).$$

To obtain a closed formula for this expression, one requires closed forms for sums $\sum_{k=1}^{n} k^p H_k^2$, $\sum_{k=1}^{n} k^p H_k$ and $\sum_{k=1}^{n} k^p H_k^{(2)}$ for non-negative integers $p \le 8$. Such identities like

$$\sum_{k=1}^{n} H_k^2 = (n+1)H_{n+1}^2 - (2n+3)H_{n+1} + 2(n+1),$$

can be computed by standard manipulations of harmonic numbers and are also generated "automatically" by computer algebra systems like MAPLE.

The given explicit expressions in the appendix were checked by the authors for a lot of "small" numbers j and n.

5.2.2 Grand averages

The grand averages E_n can now be obtained by

$$E_n = \frac{1}{n^2}[z^{n-1}]\Phi(z,1) \tag{29a}$$

for all parameters except the expected cycle length, where we have

$$E_n = \frac{1}{n^3}[z^{n-1}]\tilde{\Phi}(z,1). \tag{29b}$$

The explicit and asymptotic results obtained here are summarized in the following.

- Inversions:

$$E_n = \Big(\frac{6}{7}+\frac{6}{7n}\Big)H_n + \frac{1}{12}n^2 - \frac{1}{2}n - \frac{793}{588} + \frac{9}{98n} \sim \frac{n^2}{12} \text{ for } n \ge 6, \tag{30a}$$

$$E_1 = 0,\ E_2 = 0,\ E_3 = 0,\ E_4 = \frac{1}{4},\ E_5 = \frac{3}{5}.$$

- Fixed points:

$$E_n = \Big(\frac{24}{7} + \frac{24}{7n}\Big) H_n - \frac{38}{49n} - \frac{255}{49} \sim \frac{24}{7} \log n \quad \text{for } n \ge 6, \tag{30b}$$

$$E_1 = 1,\ E_2 = 2,\ E_3 = 3,\ E_4 = \frac{7}{2},\ E_5 = \frac{101}{25}.$$

- Cycles:

$$\begin{aligned} E_n &= \Big(\frac{6}{7} + \frac{6}{7n}\Big) H_n^2 + \Big(\frac{30}{49} + \frac{114}{49n}\Big) H_n - \Big(\frac{6}{7n} + \frac{6}{7}\Big) H_n^{(2)} \\ &\quad - \frac{65}{343n} - \frac{618}{343} \sim \frac{6}{7} \log^2 n \quad \text{for } n \ge 6, \end{aligned} \tag{30c}$$

$$E_1 = 1,\ E_2 = 2,\ E_3 = 3,\ E_4 = \frac{15}{4},\ E_5 = \frac{112}{25}.$$

- Expected cycle length:

$$E_n = \frac{1}{6} n - \frac{793}{294n} + \frac{9}{49}\frac{1}{n^2} + \frac{12}{7}\Big(\frac{1}{n^2} + \frac{1}{n}\Big) H_n \sim \frac{n}{6} \quad \text{for } n \ge 6, \tag{30d}$$

$$E_1 = 1,\ E_2 = 1,\ E_3 = 1,\ E_4 = \frac{9}{8},\ E_5 = \frac{31}{25}.$$

- Distance to the identity permutation:

$$E_n = \frac{(n+1)(14n^3 - 49n^2 + 14n + 36)}{525n} \sim \frac{2n^3}{75} \quad \text{for } n \ge 6, \tag{30e}$$

$$E_1 = 0,\ E_2 = 0,\ E_3 = 0,\ E_4 = \frac{1}{2},\ E_5 = \frac{36}{25}.$$

6 Variances

6.1 Explicit results

In principle, it is also possible to obtain explicit expressions for higher moments of $Q_{n,j}$ from equations (20), in particular for the variance $\mathbb{V}(Q_{n,j})$. It turns out that the explicit expressions that we obtain for the second factorial moment $\mathbb{E}\big(Q_{n,j}(Q_{n,j}-1)\big)$ for ordinary quickselect are already of daunting complexity. Defining

$$H(z,u) = \Big(\frac{\partial^2}{\partial v^2} F(z,u,v)\Big)\Big|_{v=1} \quad \text{resp.} \quad \tilde{H}(z,u) = \Big(\frac{\partial^2}{\partial v^2} \tilde{F}(z,u,v)\Big)\Big|_{v=1}, \tag{31}$$

the second factorial moments are given by $\mathbb{E}\big(Q_{n,j}(Q_{n,j}-1)\big) = [z^n u^j]H(z,u)$ for all parameters except the expected cycle length, where we obtain $\mathbb{E}\big(\tilde{Q}_{n,j}(\tilde{Q}_{n,j}-1)\big) = [z^n u^j]\tilde{H}(z,u)$. For the sake of completeness, we give these findings in the appendix. We refrain from doing such calculations for the more complicated instance of median-of-three partition.

6.2 The variance of the mean

Of particular interest is the variance of the mean

$$V_n := \mathbb{V}\Big(\frac{1}{n}\sum_{j=1}^{n} Q_{n,j}\Big) \tag{32}$$

for our parameters. This quantity can be obtained by evaluating at $u=1$ at the level of generating functions and extracting coefficients. One gets

$$V_n = \frac{1}{n^2}[z^n]H(z,1) + \frac{1}{n}E_n - E_n^2 \tag{33a}$$

for all parameters considered except the expected cycle length, where we get

$$V_n = \frac{1}{n^4}[z^n]\tilde{H}(z,1) + \frac{1}{n^2}E_n - E_n^2.$$

The appearing E_n are the grand averages as computed in Subsection 5.1.2.

Next we summarize the obtained explicit and asymptotic results.

- Inversions:

$$\begin{aligned} V_n &= \frac{n^4}{720} + \frac{5n^3}{432} - \frac{11n^2}{216} + \frac{247n}{432} + \frac{4607}{2160} - \Big(\frac{1}{n} + \frac{1}{n^2}\Big)H_n^2 \\ &\quad + \Big(-\frac{n}{6} - \frac{1}{3} + \frac{11}{6n}\Big)H_n - \Big(1 + \frac{1}{n}\Big)H_n^{(2)}, \qquad n \ge 1 \\ &\sim \frac{n^4}{720}. \end{aligned} \tag{34a}$$

- Fixed points:

$$\begin{aligned} V_n &= -\Big(16 + \frac{16}{n}\Big)H_n^{(2)} - \Big(\frac{16}{n} + \frac{16}{n^2}\Big)H_n^2 + \Big(\frac{58}{n} + 10\Big)H_n \\ &\quad - \frac{19}{3n} + \frac{5}{3}, \qquad \text{for } n \ge 2, \quad V_1 = 0. \end{aligned} \tag{34b}$$

- Cycles:

$$\begin{aligned}
V_n &= \Big(-4 - \frac{2}{n} + \frac{2}{n^2}\Big) H_n^2 H_n^{(2)} + \Big(\frac{2}{3} - \frac{4}{n^2} + \frac{2}{3n}\Big) H_n^3 \\
&\quad - \Big(\frac{1}{n^2} + \frac{1}{n}\Big) H_n^4 + \Big(-\frac{1}{n^2} + 2 + \frac{1}{n}\Big) \big(H_n^{(2)}\big)^2 \\
&\quad + \Big(\frac{5}{n} + 5 - \frac{4}{n^2}\Big) H_n^2 + \Big(\frac{4}{3} + \frac{28}{3n}\Big) H_n^{(3)} + \Big(-8 - \frac{1}{6n}\Big) H_n \\
&\quad - \Big(5 + \frac{5}{n}\Big) H_n^{(2)} - \Big(6 + \frac{6}{n}\Big) H_n^{(4)} + \Big(\frac{4}{n^2} - 2 - \frac{10}{n}\Big) H_n H_n^{(2)} \\
&\quad + \Big(\frac{8}{n} + 8\Big) H_n H_n^{(3)} + 18, \qquad \text{for } n \geq 1 \qquad (34c) \\
&\sim \frac{2}{3} \log^3 n.
\end{aligned}$$

- Expected cycle length:

$$\begin{aligned}
V_n &= \frac{1}{720} \frac{7n^4 + 25n^3 - 185n^2 + 1895n + 6898}{n^2} - \frac{4(n+1)}{n^4} H_n^2 \\
&\quad - \frac{4(n+1)}{n^3} H_n^{(2)} - \frac{2}{3} \frac{(n+5)(n-2)}{n^3} H_n, \quad \text{for } n \geq 1 \qquad (34d) \\
&\sim \frac{7n^2}{720}.
\end{aligned}$$

- Distance to the identity permutation:

$$\begin{aligned}
V_n &= \frac{(n-1)(n-2)(n+1)(145\,n^3 + 836\,n^2 + 53\,n - 398)}{226800}, \\
&\qquad\qquad \text{for } n \geq 1 \\
&\sim \frac{29n^6}{45360}. \qquad (34e)
\end{aligned}$$

7 Conclusion

In Table 5 we collect our basic findings. Note for instance that the number of cycles is *increasing*, since we are “closer” to the identity permutation, which has the most number of cycles.

At first glance it might seem surprising that the parameters “number of fixed points” and “number of cycles” are smaller after the median-of-three algorithm than after the ordinary quickselect algorithm, which means that for these statistics the permutations are

Table 5: Averages and variances with/without one round of quickselect (leading term only).

	Random permutation	After quickselect		After quickselect (median–of–3)	Variance
		grand average	fixed $\rho = j/n$		
Inversions	$\frac{n^2}{4}$	$\frac{n^2}{12}$	$\frac{(\rho^2+(1-\rho)^2)n^2}{8}$	$\frac{n^2}{12}$	$\frac{n^4}{720}$
Fixed points	1	$4\log n$	$4\log n$	$\frac{24}{7}\log n$	$10\log n$
Cycles	$\log n$	$\log^2 n$	$\log^2 n$	$\frac{6}{7}\log^2 n$	$\frac{2}{3}\log^3 n$
Expected cycle length	$\frac{n}{2}$	$\frac{n}{6}$	$\frac{(\rho^2+(1-\rho)^2)n}{4}$	$\frac{n}{6}$	$\frac{7n^2}{720}$
Distance to identity permutation	$\frac{n^3}{6}$	$\frac{n^3}{36}$	$\frac{(1-3\rho(1-\rho))n^3}{18}$	$\frac{2n^3}{75}$	$\frac{29n^6}{45360}$

on average more disordered in the median-of-three case. The reason for this is that with median-of-three partition the number of recursive calls in the algorithm decreases and thus the requested element can be found "faster." Because every recursive call places one pivot element in the correct position and on average in every segment between the pivots we have one additional fixed point (these segments are random permutations), one expects that the average number of fixed points will asymptotically behave like twice the average number of recursive calls in Quickselect, a parameter that was studied in [6]. With the heuristic that for large n almost no pivots are neighbors, we get that asymptotically the number of fixed points is twice the number of pivots (or recursive calls). On average we make in the median-of-three case asymptotically 1/7 fewer recursive calls and thus we have about 1/7 fewer fixed points.

8 Appendix

8.1 Explicit results for median-of-three pivoting

We use here the abbreviation $E_{n,j} := \mathbb{E}(Q_{n,j})$.

- **Inversions:**

$$
\begin{aligned}
E_{n,j} &= -\frac{6}{35}H_n(3n^2-6jn-3n+4-6j+6j^2)+\frac{9}{35}H_j(2j^2-6j+5)\\
&+\frac{9}{35}H_{n+1-j}(2n^2-2n-4jn+2j^2+2j+1)-\frac{6}{7j}-\frac{6}{7(n+1-j)}\\
&+\frac{3}{28}n^2+\frac{18}{35}jn-\frac{167}{140}n+\frac{814}{245}-\frac{921}{2450}j^2-\frac{153}{70}j\\
&-\frac{3}{35}\frac{48j-25-36j^2+3j^3}{n}+\frac{3}{35}\frac{(j^2-2j+10)(j-1)^2}{n(n-1)}\\
&+\frac{1}{175}\frac{(j-1)(j-2)(2j-3)(3j^2-9j+50)}{n(n-1)(n-2)}\\
&+\frac{3}{175}\frac{(j-1)(j-3)(j^2-4j+25)(j-2)^2}{n(n-1)(n-2)(n-3)}\\
&-\frac{3}{245}\frac{(j-1)(j-2)(j-3)(j-4)(2j-5)(j^2-5j+35)}{n(n-1)(n-2)(n-3)(n-4)}\\
&+\frac{1}{490}\frac{(j-1)(j-2)(j-4)(j-5)(3j^2-18j+140)(j-3)^2}{n(n-1)(n-2)(n-3)(n-4)(n-5)}\ \text{for } 5\le j\le n-4,\\
E_{n,4} &= -\frac{87}{35}H_n+\frac{1}{4900}\frac{525n^5-4900n^4-131843n^2+48556n^3+104042n-13440}{n(n-1)(n-2)}\ \text{for } n\ge 8,\\
&E_{4,4}=\frac{1}{4},\ E_{5,4}=\frac{13}{20},\ E_{6,4}=\frac{17}{20},\ E_{7,4}=\frac{13}{10},\\
E_{n,3} &= -\frac{3}{7}H_n+\frac{1}{980}\frac{105n^4-770n^3-2090n+2755n^2-588}{n(n-1)}\\
&\text{for } n\ge 7,\ E_{3,3}=0,\ E_{4,3}=\frac{1}{4},\ E_{5,3}=\frac{2}{5},\ E_{6,3}=\frac{17}{20},\\
E_{n,2} &= \frac{3}{28}n^2-\frac{19}{28}n-\frac{1}{175}+\frac{3}{5}H_n\ \text{for } n\ge 6,\ E_{2,2}=0,\ E_{3,2}=0,\ E_{4,2}=\frac{1}{4},\ E_{5,2}=\frac{13}{20},\\
E_{n,1} &= \frac{3}{28}n^2-\frac{19}{28}n-\frac{1}{175}+\frac{3}{5}H_n\ \text{for } n\ge 5,\ E_{1,1}=0,\ E_{2,1}=0,\ E_{3,1}=0,\ E_{4,1}=\frac{1}{4}.
\end{aligned}
$$

- **Fixed points:**

$$
\begin{aligned}
E_{n,j} &= \frac{48}{35}H_n+\frac{36}{35}H_j+\frac{36}{35}H_{n+1-j}+\frac{24}{35j}+\frac{24}{35(n+1-j)}\\
&-\frac{583}{175}+\frac{12}{35}\frac{3j-5}{n}-\frac{24}{35}\frac{(j-1)^2}{n(n-1)}-\frac{8}{35}\frac{(j-1)(j-2)(2j-3)}{n(n-2)(n-1)}-\frac{12}{35}\frac{(j-1)(j-3)(j-2)^2}{n(n-2)(n-1)(n-3)}\\
&+\frac{12}{35}\frac{(2j-5)(j-1)(j-2)(j-3)(j-4)}{n(n-2)(n-1)(n-3)(n-4)}-\frac{8}{35}\frac{(j-5)(j-1)(j-2)(j-4)(j-3)^2}{n(n-2)(n-1)(n-3)(n-4)(n-5)}\\
&\text{for } 5\le j\le n-4,\\
E_{n,4} &= \frac{12}{5}H_n-\frac{1}{175}\frac{-537n^2+179n^3+1618n-1680}{n(n-2)(n-1)}\ \text{for } n\ge 8,\\
&E_{4,4}=\frac{7}{2},\ E_{5,4}=4,\ E_{6,4}=\frac{23}{5},\ E_{7,4}=5,\\
E_{n,3} &= \frac{12}{5}H_n-\frac{1}{175}\frac{209n^2-209n+420}{n(n-1)}\ \text{for } n\ge 7,\ E_{3,3}=3,\ E_{4,3}=\frac{7}{2},\ E_{5,3}=\frac{21}{5},\ E_{6,3}=\frac{23}{5},\\
E_{n,2} &= -\frac{37}{25}+\frac{12}{5}H_n\ \text{for } n\ge 6,\ E_{2,2}=2,\ E_{3,2}=3,\ E_{4,2}=\frac{7}{2},\ E_{5,2}=4,\\
E_{n,1} &= -\frac{37}{25}+\frac{12}{5}H_n\ \text{for } n\ge 5,\ E_{1,1}=1,\ E_{2,1}=2,\ E_{3,1}=3,\ E_{4,1}=\frac{7}{2}.
\end{aligned}
$$

- **Cycles:**

$$
E_{n,j}=\frac{12}{35}H_nH_j+\frac{12}{35}H_nH_{n+1-j}+\Big(\frac{34}{35}-\frac{12}{35j}-\frac{12}{35(n+1-j)}\Big)H_n+\frac{9}{35}H_j^2+\frac{9}{35}H_{n+1-j}^2
$$

$$
\begin{aligned}
&- \frac{12}{35} H_j H_{n+1-j} + \Big(\frac{12}{35j} + \frac{12}{35(n+1-j)} + \frac{103}{175} - \frac{12}{35}\frac{j}{n} - \frac{6}{35}\frac{j(j-1)}{n\big(n-1\big)} - \frac{4}{35}\frac{j(j-1)\big(j-2\big)}{n(n-1)(n-2)} \\
&- \frac{3}{35}\frac{j(j-1)(j-2)(j-3)}{n(n-1)(n-2)(n-3)} + \frac{6}{35}\frac{j(j-1)(j-2)(j-3)(j-4)}{n(n-1)(n-2)(n-3)(n-4)} \\
&- \frac{2}{35}\frac{j(j-1)(j-2)(j-3)(j-4)(j-5)}{n(n-1)(n-2)(n-3)(n-4)(n-5)}\Big) H_j \\
&+ \Big(\frac{12}{35j} + \frac{12}{35(n+1-j)} - \frac{2}{175} + \frac{6}{7}\frac{j-1}{n} - \frac{6}{35}\frac{(j-1)(j-2)}{n(n-1)} - \frac{4}{35}\frac{(j-1)(j-2)(j-3)}{n(n-1)(n-2)} \\
&\quad - \frac{3}{35}\frac{(j-1)(j-2)(j-3)(j-4)}{n(n-1)(n-2)(n-3)} + \frac{6}{35}\frac{(j-1)(j-2)(j-3)(j-4)(j-5)}{n(n-1)(n-2)(n-3)(n-4)} \\
&\quad - \frac{2}{35}\frac{(j-1)(j-2)(j-3)(j-4)(j-5)(j-6)}{n(n-1)(n-2)(n-3)(n-4)(n-5)}\Big) H_{n+1-j} \\
&- \frac{3}{5}H_j^{(2)} - \frac{3}{5}H_{n+1-j}^{(2)} + \frac{1}{35}\frac{17n+5}{(n+1)(n+1-j)} + \frac{1}{35}\frac{17n+5}{(n+1)j} - \frac{6221}{5250} - \frac{1}{175}\frac{2j+23}{n} \\
&- \frac{1}{35}\frac{(3j-7)(j-1)}{n(n-1)} + \frac{1}{105}\frac{(8j-9)(j-1)(j-2)}{n(n-1)(n-2)} + \frac{1}{35}\frac{(9j-28)(j-1)(j-2)(j-3)}{n(n-1)(n-2)(n-3)} \\
&- \frac{2}{175}\frac{(31j-90)(j-1)(j-2)(j-3)(j-4)}{n(n-1)(n-2)(n-3)(n-4)} \\
&+ \frac{62}{525}\frac{(j-1)(j-2)(j-3)^2(j-4)(j-5)}{n(n-1)(n-2)(n-3)(n-4)(n-5)} \quad \text{for } 5 \le j \le n-4,
\end{aligned}
$$

$$
\begin{aligned}
E_{n,4} &= \frac{3}{5}H_n^2 + \frac{6}{25}\frac{4\,n^3 - 12\,n^2 - 7\,n + 20}{n(n-1)(n-2)} H_n - \frac{3}{5}H_n^{(2)} \\
&\quad + \frac{3}{3500}\frac{53760\,n + 701\,n^6 + 4353\,n^4 - 11200 + 24348\,n^3 - 66356\,n^2 - 4206\,n^5}{n^2(n-1)^2(n-2)^2} \quad \text{for } n \ge 8, \\
&E_{4,4} = \frac{15}{4},\ E_{5,4} = \frac{89}{20},\ E_{6,4} = \frac{21}{4},\ E_{7,4} = \frac{59}{10}, \\
E_{n,3} &= \frac{3}{5}H_n^2 + \frac{6}{25}\frac{4\,n^2 - 4\,n - 5}{n(n-1)} H_n - \frac{3}{5}H_n^{(2)} + \frac{2}{875}\frac{-525 - 282\,n^3 - 454\,n^2 + 141\,n^4 + 1645\,n}{n^2(n-1)^2} \\
&\text{for } n \ge 7, E_{3,3} = 3,\ E_{4,3} = \frac{15}{4},\ E_{5,3} = \frac{23}{5},\ E_{6,3} = \frac{21}{4}, \\
E_{n,2} &= \frac{3}{5}H_n^2 + \frac{24}{25}H_n - \frac{3}{5}H_n^{(2)} + \frac{1}{125} \quad \text{for } n \ge 6,\ E_{2,2} = 2,\ E_{3,2} = 3,\ E_{4,2} = \frac{15}{4},\ E_{5,2} = \frac{89}{20}, \\
E_{n,1} &= \frac{3}{5}H_n^2 + \frac{24}{25}H_n - \frac{3}{5}H_n^{(2)} + \frac{1}{125} \quad \text{for } n \ge 5,\ E_{1,1} = 1,\ E_{2,1} = 2,\ E_{3,1} = 3,\ E_{4,1} = \frac{15}{4}.
\end{aligned}
$$

- **Expected cycle length:**

$$
\begin{aligned}
E_{n,j} &= \frac{18}{35}\frac{2j^2 - 6j + 5}{n} H_j + \frac{18}{35}\frac{2n^2 - 4jn - 2n + 2j^2 + 2j + 1}{n} H_{n+1-j} \\
&\quad - \frac{12}{35}\frac{3n^2 - 6jn - 3n + 6j^2 - 6j + 4}{n} H_n - \frac{12}{7}\frac{1}{jn} - \frac{12}{7}\frac{1}{(n+1-j)n} + \frac{3}{14}n + \frac{36}{35}j \\
&\quad - \frac{97}{70} + \frac{1628}{245n} - \frac{921}{1225}\frac{j^2}{n} - \frac{153}{35}\frac{j}{n} - \frac{6}{35}\frac{48\,j - 25 - 36\,j^2 + 3\,j^3}{n^2} \\
&\quad + \frac{6}{35}\frac{(j^2 - 2\,j + 10)(j-1)^2}{n^2(n-1)} + \frac{2}{175}\frac{(j-1)(j-2)(2j-3)(3j^2 - 9j + 50)}{n^2(n-1)(n-2)} \\
&\quad + \frac{6}{175}\frac{(j-1)(j-3)(j^2 - 4j + 25)(j-2)^2}{n^2(n-1)(n-2)(n-3)} \\
&\quad - \frac{6}{245}\frac{(j-1)(j-2)(j-3)(j-4)(2j-5)(j^2 - 5j + 35)}{n^2(n-1)(n-2)(n-3)(n-4)} \\
&\quad + \frac{1}{245}\frac{(j-1)(j-2)(j-4)(j-5)(3j^2 - 18j + 140)(j-3)^2}{n^2(n-1)(n-2)(n-3)(n-4)(n-5)} \quad \text{for } 5 \le j \le n-4, \\
E_{n,4} &= -\frac{174}{35}\frac{H_n}{n} + \frac{1}{2450}\frac{525n^5 - 2450n^4 + 41206n^3 - 126943n^2 + 104042n - 13440}{n^2(n-1)(n-2)} \quad \text{for } n \ge 8,
\end{aligned}
$$

$$E_{4,4} = \frac{9}{8},\ E_{5,4} = \frac{63}{50},\ E_{6,4} = \frac{77}{60},\ E_{7,4} = \frac{48}{35},$$

$$E_{n,3} = -\frac{6}{7}\frac{H_n}{n} + \frac{1}{490}\frac{105n^4 - 280n^3 + 2265n^2 - 2090n - 588}{n^2(n-1)} \text{ for } n \geq 7,$$

$$E_{3,3} = 1,\ E_{4,3} = \frac{9}{8},\ E_{5,3} = \frac{29}{25},\ E_{6,3} = \frac{77}{60},$$

$$E_{n,2} = \frac{6}{5}\frac{H_n}{n} + \frac{3n}{14} - \frac{5}{14} - \frac{2}{175n} \text{ for } n \geq 6,\ E_{2,2} = 1,\ E_{3,2} = 1,\ E_{4,2} = \frac{9}{8},\ E_{5,2} = \frac{63}{50},$$

$$E_{n,1} = \frac{6}{5}\frac{H_n}{n} + \frac{3n}{14} - \frac{5}{14} - \frac{2}{175n} \text{ for } n \geq 5,\ E_{1,1} = 1,\ E_{2,1} = 1,\ E_{3,1} = 1,\ E_{4,1} = \frac{9}{8}.$$

• **Distance to the identity permutation:**

$$\begin{aligned}
E_{n,j} = {} & \frac{6}{35}H_n(7n^3 - 21jn^2 - 9n^2 + 21j^2n + 2n + 18jn - 18 - 39j^2 + 39j) \\
& - \frac{6}{35}H_j(j-1)(j-2)(7j-9) - \frac{6}{35}H_{n+1-j}(7n - 2 - 7j)(n-j)(n-1-j) \\
& + \frac{1}{24}n^3 - \frac{6}{5}jn^2 + \frac{143}{140}n^2 - \frac{3419}{840}n + \frac{9}{4}j^2n + \frac{279}{140}jn - \frac{255}{28}j + \frac{70743}{4900} - \frac{21}{10}j^3 + \frac{57}{196}j^2 \\
& - \frac{108}{35j} - \frac{108}{35(n+1-j)} + \frac{3}{140}\frac{49j^4 - 218j^3 + 803j^2 - 850j + 360}{n} \\
& + \frac{6}{35}\frac{(5j^2 - 10j + 18)(j-1)^2}{n(n-1)} + \frac{6}{35}\frac{(j-1)(j-2)(2j-3)(j^2 - 3j + 6)}{n(n-1)(n-2)} \\
& + \frac{6}{35}\frac{(j-1)(j-3)(j^2 - 4j + 9)(j-2)^2}{n(n-1)(n-2)(n-3)} \\
& - \frac{6}{245}\frac{(j-1)(j-2)(j-3)(j-4)(2j-5)(5j^2 - 25j + 63)}{n(n-1)(n-2)(n-3)(n-4)} \\
& + \frac{3}{245}\frac{(j-1)(j-2)(j-4)(j-5)(5j^2 - 30j + 84)(j-3)^2}{n(n-1)(n-2)(n-3)(n-4)(n-5)} \text{ for } 5 \leq j \leq n-4,
\end{aligned}$$

$$E_{n,4} = \frac{684}{35}H_n + \frac{1}{29400}\frac{1225n^4 - 5250n^3 - 127225n^2 - 711282n + 302400}{n} \text{ for } n \geq 8,$$

$$E_{4,4} = \frac{1}{2},\ E_{5,4} = \frac{8}{5},\ E_{6,4} = 2,\ E_{7,4} = \frac{16}{5},$$

$$E_{n,3} = \frac{1}{24}n^3 - \frac{5}{28}n^2 - \frac{223}{168}n + \frac{144}{35}H_n - \frac{13177}{4900} \text{ for } n \geq 7,\ E_{3,3} = 0,\ E_{4,3} = \frac{1}{2},\ E_{5,3} = \frac{4}{5},\ E_{6,3} = 2,$$

$$E_{n,2} = -\frac{5}{28}n^2 + \frac{1}{24}n^3 + \frac{29}{168}n - \frac{1}{140} \text{ for } n \geq 6,\ E_{2,2} = 0,\ E_{3,2} = 0,\ E_{4,2} = \frac{1}{2},\ E_{5,2} = \frac{8}{5},$$

$$E_{n,1} = -\frac{5}{28}n^2 + \frac{1}{24}n^3 + \frac{29}{168}n - \frac{1}{140} \text{ for } n \geq 5,\ E_{1,1} = 0,\ E_{2,1} = 0,\ E_{3,1} = 0,\ E_{4,1} = \frac{1}{2}.$$

Due to the symmetry $E_{n,j} = E_{n+1-j,j}$, this parameter is fully described by the above values.

8.2 Explicit expressions for the second factorial moments for ordinary quickselect

We obtain the following results for the second factorial moments $M^{(2)}_{n,j} := \mathbb{E}\big(Q_{n,j}(Q_{n,j}-1)\big)$.

• **Inversions:**

$$M^{(2)}_{n,j} = -\frac{1}{4}H^{(2)}_j - \frac{1}{4}H^{(2)}_{n+1-j}$$

$$
\begin{aligned}
&- \frac{1}{24}\frac{H_n(n+1)(n^3j-4n^2j^2+3n^2j+6j^3n-8j^2n-10jn+9j^2-3j^4-12+6j^3-12j)}{j(n+1-j)}\\
&+ \frac{1}{4}H_j^2+\frac{1}{4}H_{n+1-j}^2+\frac{1}{2}H_{n+1-j}H_j\\
&+ \frac{1}{24}\frac{H_j}{j(n+1-j)}(3n^3j-12n^2j-9n^2j^2+12j^3n-4j^2n+j^4n-12n-43jn+15j^2-j^5\\
&- 28j-5j^4+19j^3-12)\\
&+ \frac{1}{24}\frac{H_{n+1-j}}{j(n+1-j)}(n^4j+7n^3j-4n^3j^2-21n^2j^2-13n^2j+6n^2j^3-12n-53jn+5j^2n\\
&\quad + 24j^3n-4j^4n-10j^4-34j+32j^2+11j^3+j^5-12)\\
&+ \frac{1}{3456}\frac{1}{j(n+1-j)}(1626j^4+702j^5-4422j^3-4296j^2+6624j-234j^6+1728+702j^5n\\
&\quad + 63jn^5+9382jn-805n^3j+3783n^2j^2+2191n^2j-5094j^3n-834j^4n+2468j^2n\\
&\quad - 175n^4j+304n^3j^2-954j^4n^2+66n^2j^3-315n^4j^2+738n^3j^3) \text{ for } n\ge j\ge 1.
\end{aligned}
$$

- **Fixed points:**

$$
\begin{aligned}
M_{n,j}^{(2)} &= -4H_j^{(2)}-4H_{n+1-j}^{(2)}+8\frac{H_n(n+1)}{(n+1-j)j}+4H_j^2+4H_{n+1-j}^2+8H_{n+1-j}H_j\\
&- \frac{H_j(-9j^2+9jn+9j+8n+8)}{(n+1-j)j}-\frac{H_{n+1-j}(-9j^2+9jn+9j+8n+8)}{(n+1-j)j}\\
&+ \frac{1}{3}\frac{118j+23n^2j^2-46j^3+47n^2j-46j^3n-j^2n+23j^4+60n-95j^2+72+165jn+12n^2}{j(n+2-j)(j+1)(n+1-j)}\\
&\quad \text{for } n-1\ge j\ge 2,\\
M_{n,1}^{(2)} &= 4H_n^2-4H_n^{(2)}-H_n+\frac{4}{n}-\frac{5}{2}\\
&\quad \text{for } n\ge 2,\ M_{1,1}^{(2)}=0.
\end{aligned}
$$

- **Cycles:**

$$
\begin{aligned}
M_{n,j}^{(2)} &= \Big(\frac{2}{(n+1-j)^2}-\frac{2}{n+1-j}-2\frac{H_{n+1-j}}{n+1-j}\Big)\sum_{k=1}^{j}\frac{H_{n-j+k}}{k}+\Big(\frac{2}{j^2}-\frac{2}{j}-2\frac{H_j}{j}\Big)\sum_{k=1}^{n+1-j}\frac{H_{k+j-1}}{k}\\
&+ 2H_{n+1-j}H_{n+1-j}^{(3)}+2H_jH_j^{(3)}-4H_jH_j^{(2)}-4H_{n+1-j}H_{n+1-j}^{(2)}-\frac{3}{2}H_{n+1-j}^{(4)}-\frac{3}{2}H_j^{(4)}\\
&+ \frac{8}{3}H_{n+1-j}^{(3)}+\frac{8}{3}H_j^{(3)}+\frac{3}{4}(H_j^{(2)})^2+\frac{3}{4}(H_{n+1-j}^{(2)})^2-\frac{3}{2}H_j^2H_j^{(2)}-\frac{3}{2}H_{n+1-j}^2H_{n+1-j}^{(2)}\\
&+ \frac{1}{4}H_j^4+\frac{1}{4}H_{n+1-j}^4+\frac{4}{3}H_j^3+\frac{4}{3}H_{n+1-j}^3+\frac{1}{2}H_j^2H_{n+1-j}^2+\frac{1}{2}H_j^{(2)}H_{n+1-j}^{(2)}\\
&- \frac{1}{2}H_j^2H_{n+1-j}^{(2)}-\frac{1}{2}H_{n+1-j}^2H_j^{(2)}+\Big(\frac{1}{j}-1\Big)H_jH_{n+1-j}^{(2)}+\Big(\frac{1}{n+1-j}-1\Big)H_{n+1-j}H_j^{(2)}\\
&+ \Big(-1+\frac{1}{j}-\frac{1}{j^2}\Big)H_{n+1-j}^{(2)}+\Big(-1+\frac{1}{n+1-j}-\frac{1}{(n+1-j)^2}\Big)H_j^{(2)}\\
&+ \Big(1-\frac{1}{j}\Big)H_jH_{n+1-j}^2+\Big(1-\frac{1}{n+1-j}\Big)H_{n+1-j}H_j^2\\
&+ \Big(-\frac{2}{j^2}-\frac{2}{j}+\frac{1}{(n+1-j)^2}-\frac{1}{n+1-j}+1\Big)H_j^2\\
&+ \Big(-\frac{2}{(n+1-j)^2}-\frac{2}{n+1-j}+\frac{1}{j^2}-\frac{1}{j}+1\Big)H_{n+1-j}^2\\
&+ \Big(-2-\frac{2(2n+1)}{(n+1)(n+1-j)}-\frac{2}{(n+1-j)^2(n+1)}+\frac{2}{(n+1-j)^3}-\frac{2(2n+1)}{(n+1)j}\\
&\quad + \frac{2}{(n+1)j^2}+\frac{2}{j^3}\Big)H_j\\
&+ \Big(-2-\frac{2(2n+1)}{(n+1)(n+1-j)}+\frac{2}{(n+1-j)^2(n+1)}+\frac{2}{(n+1-j)^3}-\frac{2(2n+1)}{(n+1)j}
\end{aligned}
$$

$$
\begin{aligned}
&- \frac{2}{(n+1)j^2} + \frac{2}{j^3})H_{n+1-j} \\
&+ (2 - \frac{2n}{(n+1)(n+1-j)} - \frac{2}{(n+1-j)^2} - \frac{2n}{(n+1)j} - \frac{2}{j^2})H_j H_{n+1-j} \\
&+ (\frac{2}{j^2} + \frac{2}{j} - \frac{2}{(n+1-j)^2} + \frac{2}{n+1-j})H_n H_j \\
&+ (\frac{2}{(n+1-j)^2} + \frac{2}{n+1-j} - \frac{2}{j^2} + \frac{2}{j})H_n H_{n+1-j} \\
&+ (\frac{2}{j} + \frac{2}{n+1-j})H_n H_j H_{n+1-j} + (-\frac{2}{j^3} + \frac{4}{j} - \frac{2}{(n+1-j)^3} + \frac{4}{n+1-j})H_n \\
&- \frac{2(2n-1)}{(n+1)j} - \frac{2(2n-1)}{(n+1)(n+1-j)} - \frac{2}{(n+1)j^3} - \frac{2}{(n+1)(n+1-j)^3} + 12 \text{ for } n \geq j \geq 1.
\end{aligned}
$$

- **Expected cycle length:**

$$
\begin{aligned}
M_{n,j}^{(2)} &= \frac{H_j^2}{n^2} + \frac{H_{n+1-j}^2}{n^2} - \frac{H_j^{(2)}}{n^2} - \frac{H_{n+1-j}^{(2)}}{n^2} + 2\frac{H_j H_{n+1-j}}{n^2} \\
&- (\frac{1}{6}\frac{n^3 - 3jn^2 + 3n^2 + 3j^2 n - 6jn - 10n + 3j^2 - 3j - 12}{n^2} - \frac{2}{jn^2} - \frac{2}{n^2(n+1-j)})H_n \\
&+ (\frac{1}{6}\frac{3n^2 - 6jn - 9n + j^3 + 6j^2 - 13j - 18}{n^2} - \frac{2}{jn^2} - \frac{2}{n^2(n+1-j)})H_j \\
&+ (\frac{1}{6}\frac{n^3 - 3jn^2 + 6n^2 + 3j^2 n - 12jn - 13n - j^3 + 9j^2 - 2j - 24}{n^2} \\
&\quad - \frac{2}{jn^2} - \frac{2}{n^2(n+1-j)})H_{n+1-j} \\
&+ \frac{1}{72n^2}(6n^4 - 24jn^3 - 2n^3 + 45j^2 n^2 - 9jn^2 - 72n^2 - 42j^3 n + 27j^2 n + 147jn + 68n + 21j^4 \\
&\quad - 42j^3 - 111j^2 + 132j + 432) \\
&+ \frac{2}{jn^2(n+1)} + \frac{2}{n^2(n+1-j)(n+1)} \text{ for } n \geq j \geq 1.
\end{aligned}
$$

- **Distance to the identity permutation:**

$$
\begin{aligned}
M_{n,j}^{(2)} &= \frac{2}{45}H_n(n+1)(n^4 - 5n^3 j + 4n^3 + 10n^2 j^2 + n^2 - 15n^2 j + 20j^2 n - 10j^3 n - 6n + 10j - 5j^2 \\
&\quad + 5j^4 - 10j^3) - \frac{2}{45}jH_j(j-1)(j-2)(j+2)(j+1) \\
&- \frac{2}{45}H_{n+1-j}(n+1-j)(n-j)(n-1-j)(n+3-j)(n+2-j) \\
&+ \frac{31n^6}{6480} - \frac{n^5(-27+62j)}{2160} + \frac{n^4(343 + 465j^2 - 693j)}{6480} - \frac{n^3(2588j + 580j^3 - 1986j^2 - 687)}{6480} \\
&+ \frac{n^2(202 + 345j^4 - 2466j^3 + 5562j^2 - 2349j)}{6480} - \frac{n(4628j^3 - 1053j^4 + 66j^5 - 236j + 1344 - 3909j^2)}{6480} \\
&+ \frac{1}{3240}j(j-1)(11j^4 - 22j^3 + 691j^2 - 680j - 252) \text{ for } n \geq j \geq 1.
\end{aligned}
$$

References

[1] Comtet, L. (1974), Advanced Combinatorics. D. Reidel Publishing Company.

[2] Graham, R. L., Knuth, D. E., and Patashnik, O. (1994), Concrete Mathematics. Second Edition, Reading, Massachussetts: Addison-Wesley.

[3] Greene, D. H. and Knuth. D. E. (1982), Mathematics for the Analysis of Algorithms. Second Edition, Berlin: Birkhäuser.

[4] Hoare, C. A. R. (1961), Find (Algorithm 65). Communications of the ACM, **4**, 321–322.

[5] Hoare, C. A. R. (1962), Quicksort. Computer Journal, **5**, 10–15.

[6] 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.

[7] Knuth, D. E. (1998), The Art of Computer Programming. Volume 3: Sorting and Searching, 2nd ed. Reading, Massachussetts: Addison-Wesley.

[8] Martínez, C., Panholzer, A., and Prodinger, H. (1998), Descendants and ascendants in random search trees. Electronic Journal of Combinatorics, **5**, R20.

[9] Panholzer, A. (1997), Untersuchungen zur durchschnittlichen Gestalt gewisser Baumfamilien (Mit besonderer Berücksichtigung von Anwendungen in der Informatik). Ph. D. Thesis, TU Vienna, 219 pages.

[10] Panholzer, A. and Prodinger, H. (2002), Binary search tree recursions with harmonic toll functions. Journal of Computational and Applied Mathematics, **142**, 211–225.

[11] Sedgewick, R. and Flajolet, P. (1996), An Introduction to the Analysis of Algorithms. Reading, Massachussetts: Addison-Wesley.