COUNTING ZEROS IN RANDOM WALKS ON THE INTEGERS AND ANALYSIS OF OPTIMAL DUAL-PIVOT QUICKSORT

MARTIN AUMÜLLER, MARTIN DIETZFELBINGER, CLEMENS HEUBERGER,
DANIEL KRENN, AND HELMUT PRODINGER

ABSTRACT. We present an average case analysis of two variants of dual-pivot quicksort, one with a non-algorithmic comparison-optimal partitioning strategy, the other with a closely related algorithmic strategy. For both we calculate the expected number of comparisons exactly as well as asymptotically, in particular, we provide exact expressions for the linear, logarithmic, and constant terms. An essential step is the analysis of zeros of lattice paths in a certain probability model. Along the way a combinatorial identity is proven.

1. INTRODUCTION

Dual-pivot quicksort [Sed75, WNN15, AD15] is a family of sorting algorithms related to the well-known quicksort algorithm. In order to sort an input sequence $(a_1, \ldots, a_n)$ of distinct elements, dual-pivot quicksort algorithms proceed as follows. (For simplicity we will forbid repeated elements in the input.) If $n \leq 1$, there is nothing to do. If $n \geq 2$, two of the input elements are selected as pivots. Let p be the smaller and q be the larger pivot. The next step is to partition the remaining elements into

- the elements smaller than p ("small elements"),
- the elements between p and q ("medium elements"), and
- the elements larger than q ("large elements").

Then the procedure is applied recursively to these three groups to complete the sorting.

The cost measure used in this work is the number of comparisons between elements. As is common, we will assume the input sequence is in random order, which means that each permutation of the n elements occurs with probability $1/n!$. With this assumption we may, without loss of generality, choose a_1 and a_n as the pivots. Even in this setting there are different dual-pivot quicksort algorithms; their difference lies in the way the partitioning is organized, which influences the partitioning cost. This is in contrast to standard quicksort with one pivot, where the partitioning strategy does not influence the cost—in partitioning always one comparison is needed per non-pivot element. In dual-pivot quicksort, the average

cost (over all permutations) of partitioning and of sorting can be analyzed only when the partitioning strategy is fixed.

Only in 2009, Yaroslavskiy, Bentley, and Bloch [Yar09] described a dual-pivot quicksort algorithm that makes $1.9n \log n + O(n)$ comparisons [WNN15].[1] This beats the classical quicksort algorithm [Hoa62], which needs $2n \log n + O(n)$ comparisons on average. In [AD15], the first two authors of this article described the full design space for dual-pivot quicksort algorithms with respect to counting element comparisons. Among others, they studied two special partitioning strategies. The first one—we call it "Clairvoyant" in this work—assumes that the number of small and large elements is given (by an "oracle") before partitioning starts. It cannot be implemented; however, it is optimal among all partioning strategies that have access to such an oracle, and hence its cost provides a lower bound for the cost of all algorithmic partitioning strategies. In [AD15] it was shown that dual-pivot quicksort carries out $1.8n \log n + O(n)$ comparisons on average when this partitioning strategy is used. Further a closely related algorithmic partitioning strategy—called "Count" here—was described, which makes only $O(\log n)$ more comparisons on average than "Clairvoyant" and hence leads to a dual-quicksort variant with only $O(n)$ more comparisons.

One purpose of this paper is to make the expected number of comparisons in both variants precise and to determine the exact difference of the cost of these two strategies, both for partitioning and for the resulting dual-pivot quicksort variants.

Already in [AD15] it was noted that the exact value of the expected partitioning cost (i.e., the number of comparisons) of both strategies depends on the expected number of the zeros of certain lattice paths (Parts I and II). A complete understanding of this situation is the basis for our analysis of dual-pivot quicksort, which appears in Part III.

Lattice path enumeration has a long tradition. An early reference is [Moh79]; a recent survey paper on the subject is [Kra15]. As space is limited, many proofs and some additional results can be found in the appendix.

2. Overview and Results

This work is split into three parts. We give a brief overview on the main results of each of these parts here. We use the Iversonian expression

$$[expr] = \begin{cases} 1 & \text{if } expr \text{ is true,} \\ 0 & \text{if } expr \text{ is false,} \end{cases}$$

popularized by Graham, Knuth, and Patashnik [GKP94].

The harmonic numbers and their variants will be denoted by

$$H_n = \sum_{m=1}^{n} \frac{1}{m}, \qquad H_n^{\text{odd}} = \sum_{m=1}^{n} \frac{[m \text{ odd}]}{m} \qquad \text{and} \qquad H_n^{\text{alt}} = \sum_{m=1}^{n} \frac{(-1)^m}{m}.$$

Of course, there are relations between these three definitions such as $H_n^{\text{alt}} = H_n - 2H_n^{\text{odd}}$ and $H_n^{\text{odd}} + H_{\lfloor n/2 \rfloor}/2 = H_n$, but it will turn out to be much more convenient to use all three notations.

---

[1] In this paper "log" denotes the natural logarithm to base e.

Part I: Lattice Paths. In the first part we analyze certain lattice paths of a fixed length n. To be precise, we start with a uniform distribution on the vertical axis, allow steps/increments $(1, +1)$ and $(1, -1)$ and end on the horizontal axis at $(n, 0)$. Once the starting point is chosen, all paths to $(n, 0)$ are equally likely. We are interested in the number of zeros, denoted by the random variable Z_n, of such paths.

An exact formula for the expected number $\mathbb{E}(Z_n)$ of zeros is derived in two different ways (see identity (2.1) for these formulæ): On the one hand, we use the symbolic method and generating functions (see Appendix A), which gives the result in form of a double sum. This machinery extends well to higher moments and also allows us to obtain the distribution. The exact distribution is given in Appendix E; its limiting behavior as $n \to \infty$ is the discrete distribution

$$\mathbb{P}(Z_n = r) \sim \frac{1}{r(r+1)}.$$

On the other hand, a more probabilistic approach gives the expectation $\mathbb{E}(Z_n)$ as the simple single sum

$$\mathbb{E}(Z_n) = \sum_{m=1}^{n+1} \frac{[m \text{ odd}]}{m} = H_{n+1}^{\text{odd}},$$

see Section 4 for more details. The asymptotic behavior $\mathbb{E}(Z_n) \sim \frac{1}{2} \log n$ can be extracted (Appendix D).

The two approaches above give rise to the identity

$$\sum_{m=1}^{n+1} \frac{[m \text{ odd}]}{m} = \frac{4}{n+1} \sum_{0 \le k < \ell < \lceil n/2 \rceil} \frac{\binom{n}{k}}{\binom{n}{\ell}} + [n \text{ even}] \frac{1}{n+1} \left(\frac{2^n}{\binom{n}{n/2}} - 1 \right) + 1; \quad (2.1)$$

the double sum above equals the single sum of Theorem 4.1 by combinatorial considerations. One might ask about a direct proof of this identity. This can be achieved by methods related to hypergeometric sums and the computational proof is presented in Appendix C. We also provide a completely elementary proof which is "purely human".

Part II: More Lattice Paths and Zeros. The second part acts as connecting link between the lattice paths of fixed length of Part I and the dual-pivot quicksort algorithms of Part III.

The probabilistic model introduced in Section 3 (in Part I) is extended, and lattice paths are allowed to vary in length. For a number n (the number of elements to sort) the length of a path is the number of elements remaining when the two pivots, given by a random set of elements of size two, and the elements between these pivots are cut out.

The number of zeros X_n in this full model is analyzed; we provide again exact as well as asymptotic formulæ for the expectation $\mathbb{E}(X_n)$. Details are given in Section 7. Moreover, more specialized zero-configurations (needed for the analysis of different partitioning strategies in Part III) are considered as well (Section 6).

Part III: Dual-Pivot Quicksort. The main result of this work analyzes comparisons in the dual-pivot quicksort algorithm that uses the optimal (but unrealistic) partitioning strategy "Clairvoyant". Aumüller and Dietzfelbinger showed in [AD15] that this algorithm requires $1.8n \log n + O(n)$ comparisons on average, which improves on the average number of comparisons in quicksort ($2n \log n + O(n)$) and the

recent dual-pivot algorithm of Yaroslavskiy et al. ($1.9n \log n + O(n)$, see [WNN15]). However, for real-world input sizes n the (usually negative) factor in the linear term has a great influence on the comparison count. Our asymptotic result is stated as the following theorem.

Theorem. *The average number of comparisons in the dual-pivot quicksort algorithm with a comparison-optimal partitioning strategy is*

$$\frac{9}{5}n \log n + An + B \log n + C + O(1/n)$$

as $n \to \infty$, with $A = \frac{9}{5}\gamma - \frac{1}{5}\log 2 - \frac{89}{25} = -2.659\ldots$.

The constants B and C are explicitly given, too, and more terms of the asymptotics are presented. The precise result is formulated as Corollary 10.2.

In fact, we even get an exact expression for the average comparison count. The precise result is formulated as Theorem 10.1. Moreover the same analysis is carried out for the partitioning strategy "Count", which is an algorithmic variant of the comparison-optimal strategy "Clairvoyant". Aumüller and Dietzfelbinger [AD15] could show that it requires $\frac{9}{5}n \log n + O(n)$ comparisons as well. In this paper we obtain the exact average comparison count (Theorem 10.3). The asymptotic result is again $\frac{9}{5}n \log n + An + O(\log n)$, but now with $A = -2.382\ldots$, so there is only a small gap between the average number of comparisons in the comparison-optimal strategy "Clairvoyant" and its algorithmic variant.

Part I. **Lattice Paths**

In this first part we analyze lattice paths of a fixed length n. These are introduced in Section 3 by a precise description of our probabilistic model. We will work with this model throughout Part I, and we analyze the number of zeros Z_n.

The outline is as follows: We derive an exact expression for the expected number $\mathbb{E}(Z_n)$ of zeros by the generating functions machinery in Appendix A; a more probabilistic approach can be found in Section 4. Appendix D deals with asymptotic considerations. Direct proofs of the obtained identity are given in Appendix C and the distribution of Z_n is tackled in Appendix E.

3. Probabilistic Model

We consider paths of a given length n on the lattice $\mathbb{Z}^2$, where only steps $(1, \pm 1)$ are allowed. These paths are chosen at random according to the rules below.

Let us fix a length $n \in \mathbb{N}_0$. We choose a path P_n ending in $(n, 0)$ (no choice for this end-point) according to the following rules.

(1) First, we choose a starting point $(0, S)$ where S is a random integer uniformly distributed in $\{-n, -n+2, \ldots, n-2, n\}$, i.e., $S = s$ occurs only for integers s with $|s| \leq n$ and $s \equiv n \pmod 2$.
(2) Second, a path is chosen uniformly at random among all paths from $(0, S)$ to $(n, 0)$.

The conditions on S characterize those starting points from which $(n, 0)$ is reachable.

We are interested in the number of intersections with the horizontal axis of a path. To make this precise, we define a *zero* of a path P_n as a point $(x, 0) \in P_n$.

Thus, let P_n be a path of length n which is chosen according to the probabilistic model above and define the random variable

$$Z_n = \text{number of zeros of } P_n.$$

In the following sections, we determine the value of $\mathbb{E}(Z_n)$ exactly (Appendix A and Section 4), as well as asymptotically (Appendix D). In Appendix A, we use the machinery of generating functions. This machinery turns out to be overkill if we are just interested in the expectation $\mathbb{E}(Z_n)$. However, it easily allows extension to higher moments and the limiting distribution.

In Section 4, we follow a probabilistic approach which first gives a surprising result on the probability model: the equidistribution at the initial values turns out to carry over to every fixed length of the remaining path. This result yields a simple expression for the expectation $\mathbb{E}(Z_n)$ in terms of harmonic numbers, and thus immediately yields a precise asymptotic expansion for $\mathbb{E}(Z_n)$. The generating function approach, however, gives the expectation in terms of a double sum of quotients of binomial coefficients (the right-hand side of (2.1)), see Appendix A.

Appendix C gives a direct computational proof that these two results coincide. The original expression in [AD15] (a double sum over a quotient of a product of binomial coefficients and a binomial coefficient) is also shown to be equal in Appendix C. Explicit as well as asymptotic expressions for the distribution $\mathbb{P}(Z_n = r)$ can be found in Appendix E.

4. A Probabilistic Approach

Theorem 4.1. *For a randomly (as described in Section 3) chosen path of length n, the expected number of zeros is*

$$\mathbb{E}(Z_n) = H_{n+1}^{\text{odd}}.$$

Before proving the theorem, we consider an equivalent probability model for our random paths formulated as an urn model. A number R from $\{0, \ldots, n\}$ is chosen uniformly at random. We place R red balls and $B = n - R$ black balls in an urn. Subsequently, in n rounds the balls are taken from the urn (without replacements), in each round choosing one uniformly at random. The color of the ball chosen in round i is denoted by U_i.

We construct a random walk $(W_i)_{0 \leq i \leq n}$ on $\{-n, \ldots, n\}$ from $U_1, \ldots, U_n$ by setting $W_0 = R - B = 2R - n$ and

$$W_i = \begin{cases} W_{i-1} + 1 & \text{if } U_i = \text{black}, \\ W_{i-1} - 1 & \text{if } U_i = \text{red} \end{cases}$$

for $1 \leq i \leq n$. In each step, W_i equals the difference of the number of remaining red and black balls in the urn. Clearly, then, $W_n = 0$.

One can look at the trajectories of this random walk, represented in the grid $\{0, \ldots, n\} \times \{-n, \ldots, n\}$ as sequences $((0, W_0), (1, W_1), \ldots, (n, W_n))$. Appendix B explains the equivalence between the two models.

In order to prove Theorem 4.1, we need the following property of our paths.

Lemma 4.2. *Let $m \in \mathbb{N}_0$ with $m \leq n$. The probability that a random path P_n (as defined in Section 3) runs through $(n - m, k)$ is*

$$\mathbb{P}((n - m, k) \in P_n) = \frac{1}{m + 1} \tag{4.1}$$

for all k with $|k| \le m$ and $k \equiv m \pmod{2}$, otherwise 0.

The proof of this lemma can be found in Appendix B; we continue with the actual proof of our theorem.

Proof of Theorem 4.1. By Lemma 4.2, the expected number of zeros of P_n is

$$\mathbb{E}(Z_n) = \sum_{m=0}^{n} \mathbb{P}((n-m,0) \in P_n) = \sum_{m=0}^{n} \frac{[m \text{ even}]}{m+1} = \sum_{m=1}^{n+1} \frac{[m \text{ odd}]}{m} = H_{n+1}^{\mathrm{odd}}.$$

$\square$

5. ADDITIONAL RESULTS

The expected number of zeros can be evaluated asymptotically. We obtain

Corollary 5.1.

$$\mathbb{E}(Z_n) = \frac{1}{2} \log n + \frac{\gamma + \log 2}{2} + \frac{1 + [n \text{ even}]}{2n} - \frac{2 + 9[n \text{ even}]}{12n^2} + O\left(\frac{1}{n^3}\right)$$

asymptotically as n tends to infinity.

The proof of this result uses the well-known asymptotic expansion of the harmonic numbers. The actual asymptotic computations[2] have been carried out using the asymptotic expansions module [HK15] of SageMath [Dev16], see Appendix D.

By combining the generating function and probabilistic approach we obtain the following identity.

Theorem 5.2. *For $n \ge 0$, we have*

$$\frac{4}{n+1} \sum_{0 \le k < \ell < \lceil n/2 \rceil} \frac{\binom{n}{k}}{\binom{n}{\ell}} + [n \text{ even}] \frac{1}{n+1} \left(\frac{2^n}{\binom{n}{n/2}} - 1\right) + 1$$

$$= \frac{1}{n+1} \sum_{m=0}^{\lfloor n/2 \rfloor} \sum_{\ell=m}^{n-m} \frac{\binom{2m}{m}\binom{n-2m}{\ell-m}}{\binom{n}{\ell}}$$

$$= H_{n+1}^{\mathrm{odd}}.$$

The second expression for the expected number of zeros, but without taking the zero at $(n,0)$ into account, has been given in [AD15, displayed equation after (14)]. In Appendix C we give two direct proofs of the identity above: One of them follows a computer generated proof ("creative telescoping") by extracting the essential recurrence. The second proof is "human" and completely elementary using not more than Vandermonde's convolution.

Furthermore, the generating function machinery allows us to determine the distribution of the number Z_n of zeros. Beside an exact formula (see Appendix E), we get the following asymptotic result.

Theorem 5.3. *Let $0 < \varepsilon \le \frac{1}{2}$. For positive integers r with $r = O\big(n^{1/2-\varepsilon}\big)$, we have asymptotically*

$$\mathbb{P}(Z_n = r) = \frac{1}{r(r+1)} \big(1 + O\big(1/n^{2\varepsilon}\big)\big)$$

as n tends to infinity.

---

[2]A worksheet containing the computations can be found at `http://www.danielkrenn.at/downloads/quicksort-paths/quicksort-paths.ipynb`.

Part II. More Lattice Paths and Zeros

This second part deals with an analysis of some special zero-configurations, which are needed for the analysis of the partitioning strategies in Part III. Moreover, in Section 7, we extend the model introduced in Section 3 to accommodate lattice paths of variable length. Again expectations are studied exactly and asymptotically.

6. Going to Zero and Coming From Zero

For the analysis of comparison-optimal dual-pivot quicksort algorithms (see Part III) we need the following two variants of zeros on the lattice path.

- An *up-to-zero situation* is a point $(x, 0) \in P_n$ such that $(x - 1, -1) \in P_n$.
- A *down-from-zero situation* is a point $(x, 0) \in P_n$ such that $(x + 1, -1) \in P_n$.

We show

$$\mathbb{E}(\text{number of up-to-zero situations on } P_n) = \frac{1}{2} \left(\mathbb{E}(Z_n) - \frac{[n \text{ even}]}{n + 1} \right) = \frac{1}{2} H_n^{\text{odd}}$$

and

$$\mathbb{E}(\text{number of down-from-zero situations on } P_n) = \frac{1}{2} \left(\mathbb{E}(Z_n) - 1 \right) = \frac{1}{2} \left(H_{n+1}^{\text{odd}} - 1 \right).$$

Proof idea. The factor $\frac{1}{2}$ stems from symmetry: Up-to-zero situations at $(x, 0)$ occur with the same probability as the symmetric "down-to-zero" situations at $(x, 0)$, similarly for down-from-zero situations. The correction terms $\frac{[n \text{ even}]}{n+1}$ and 1 are caused by the fact that there is a zero, but no up-to-zero situation, at $(0, 0)$, and a zero, but no down-from-zero situation, at $(n, 0)$. The full proofs are in Appendix F. □

7. Lattice Paths of Variable Length

In this section, we use a random variable N' instead of the fixed n above. Let us fix an $n \in \mathbb{N}$ with $n \geq 2$. We choose a path length N' according to the following rules.

(1) Choose (P, Q) with $1 \leq P < Q \leq n$ uniformly at random from all $\binom{n}{2}$ possibilities.
(2) Let $N' = n - 1 - (Q - P)$.
(3) Choose a path of length N' according to Section 3.

Let us denote the number of up-to-zero and down-from-zero situations on the path by $X_n^{\nearrow}$ and $X_n^{\searrow}$, respectively. We show

$$\mathbb{E}\left(X_n^{\nearrow}\right) = \frac{1}{2\binom{n}{2}} \sum_{n'=0}^{n-2} \sum_{m=1}^{n'} [m \text{ odd}] \frac{n' + 1}{m} = \frac{1}{2} H_{n-2}^{\text{odd}} - \frac{1}{8} + \frac{(-1)^n}{8(n - [n \text{ even}])}$$

and

$$\mathbb{E}\left(X_n^{\searrow}\right) = \frac{1}{2\binom{n}{2}} \sum_{n'=0}^{n-2} \sum_{m=3}^{n'+1} [m \text{ odd}] \frac{n' + 1}{m} = \mathbb{E}\left(X_n^{\nearrow}\right) - \frac{1}{2} + \frac{1}{2(n - [n \text{ even}])}.$$

in Appendix G.

Part III. **Dual-Pivot Quicksort**

In this third and last part of this work, we finally analyze two different partitioning strategies and the dual-pivot quicksort algorithm itself.

As mentioned in the introduction, the number of comparisons of dual-pivot quicksort depends on the concrete partitioning procedure. For example, if one wants to classify a large element, i.e., an element larger than the larger pivot, comparing it with the larger pivot is unavoidable, but it depends on the partitioning procedure whether a comparison with the smaller pivot occurs as well.

First, in Section 8, we make our set-up precise, fix notions, and start solving the dual-pivot quicksort recurrence (8.1). This recurrence relates the cost of the partitioning step to the total number of comparisons of dual-pivot quicksort.

In Section 9 two partitioning strategies, called "Clairvoyant" and "Count", are introduced and their respective cost is analyzed. It will turn out that the results on lattice paths obtained in Parts I and II are central in determining the partitioning cost exactly.

Everything is put together in Section 10: We obtain the exact comparison count for two versions of dual-pivot quicksort (Theorems 10.1 and 10.3). The asymptotic behavior is extracted out of the exact results (Corollaries 10.2 and 10.4).

8. Solving the Dual-Pivot Quicksort Recurrence

We consider versions of dual-pivot quicksort that act as follows on an input sequence $(a_1, \ldots, a_n)$ consisting of distinct numbers: If $n \leq 1$, do nothing, otherwise choose a_1 and a_n as pivots, and by one comparison determine $p = \min(a_1, a_n)$ and $q = \max(a_1, a_n)$. Use a partitioning procedure to partition the remaining $n - 2$ elements into the three classes *small*, *medium*, and *large*. Then call dual-pivot quicksort recursively on each of these three classes to finish the sorting, using the same partitioning procedure in all recursive calls.

Let P_n, a random variable, denote the *partitioning cost*. This is defined as the number of comparisons made by the partitioning procedure if the input $(a_1, \ldots, a_n)$ is assumed to be in random order. Further, let C_n be the random variable that denotes the number of comparisons carried out when sorting n elements with dual-pivot quicksort. The reader should be aware that both P_n and C_n are determined by the partitioning procedure used.

Since the input $(a_1, \ldots, a_n)$ is in random order and the partitioning procedure does nothing but compare elements with the two pivots, the inputs for the recursive calls are in random order as well, which implies that the distributions of P_n and C_n only depend on n. In particular we may assume that when the sorting algorithm is called on n elements during recursion, the input is a permutation of $\{1, \ldots, n\}$.

The recurrence

$$\mathbb{E}(C_n) = \mathbb{E}(P_n) + \frac{3}{\binom{n}{2}} \sum_{k=1}^{n-2} (n - 1 - k) \, \mathbb{E}(C_k) \tag{8.1}$$

for $n \geq 0$ describes the connection between the expected sorting cost $\mathbb{E}(C_n)$ and the expected partitioning cost $\mathbb{E}(P_n)$. It will be central for our analysis. Note that it is irrelevant for (8.1) how the partitioning cost $\mathbb{E}(P_n)$ is determined; it need not even be referring to comparisons. The recurrence is simple and well-known; a version of it occurs already in Sedgewick's thesis [Sed75]. For the convenience of the reader we give a brief justification in Appendix H. In Hennequin [Hen91]

recurrence (8.1) was solved exactly for $\mathbb{E}(P_n) = an + b$, where a and b are constants. For $\mathbb{E}(P_n) = an + O(n^{1-\varepsilon})$ the solution is $\mathbb{E}(C_n) = \frac{6}{5}an \log n + O(n)$, see [AD15, Theorem 1].

9. Partitioning Algorithms and Their Cost

In Section 8 we saw that in order to calculate the average number of comparisons $\mathbb{E}(C_n)$ of a dual-pivot quicksort algorithm we need the expected partitioning cost $\mathbb{E}(P_n)$ of the partitioning procedure used. The aim of this section is to determine $\mathbb{E}(P_n)$ for two such partitioning procedures, "Clairvoyant" and "Count", to be described below.

We use the set-up described at the beginning of Section 8. For partitioning we use comparisons to *classify* the $n-2$ elements a_2, ..., a_{n-1} as *small*, *medium*, or *large*. We will be using the term *classification* for this central aspect of partitioning. Details of a partitioning procedure that concern how the classes are represented or elements are moved around may and will be ignored. (Nonetheless, in Appendix M we provide pseudocode for the considered classification strategies turned into dual-pivot quicksort algorithms.) The cost P_n depends on the concrete classification strategy used, the only relevant difference between classification strategies being whether the next element to be classified is compared with the smaller pivot p or the larger pivot q first. This decision may depend on the whole history of outcomes of previous comparisons. (The resulting abstract classification strategies may conveniently be described as classification trees, see [AD15], but we do not need this model here.)

Two comparisons are necessary for each medium element. Furthermore, one comparison with p is necessary for small and one comparison with q is necessary for large elements. As the input consists of the elements 1, ..., n, there are $p-1$ small, $q-p-1$ medium, and $n-q$ large elements. Averaging over all $\binom{n}{2}$ positions of the pivots, we see that on average

$$\frac{4}{3}(n-2) + 1 \qquad (9.1)$$

necessary comparisons are required no matter how the classification procedure works, see [AD15, (5)]; the summand $+1$ corresponds to the comparison of a_1 and a_n when choosing the two pivots.

We call other comparisons occurring during classification *additional comparisons*. That means, an additional comparison arises when a small element is compared with q first or a large element is compared with p first. In order to obtain $\mathbb{E}(P_n)$ for some classification strategy, we have to calculate the expected number of additional comparisons.

Next we describe two (closely related) classification strategies from [AD15]. Let s_i and ℓ_i denote the number of elements that have been classified as small and large, respectively, in the first i classification rounds. Set $s_0 = \ell_0 = 0$.

Strategy "Clairvoyant". *Assume the input contains $s = p-1$ small and $\ell = n-q$ large elements. When classifying the ith element, for $1 \leq i \leq n-2$, proceed as follows: If $s - s_{i-1} \geq \ell - \ell_{i-1}$, compare with p first, otherwise compare with q first.*

The number of additional comparisons of this strategy is denoted by A_n^{cv}, its partitioning cost P_n^{cv}.

Note that the strategy "Clairvoyant" cannot be implemented algorithmically, since s and ℓ are not known until the classification is completed.

As shown in [AD15, Section 6], this strategy offers the smallest expected classification cost among all strategies that have oracle access to s and ℓ at the outset of a classification round. As such, its expected cost is a lower bound for the cost of all algorithmic classification procedures; hence we call it an *optimal strategy*.

The non-algorithmic strategy "Clairvoyant" can be turned into an algorithmic classification strategy, which is described next. It will turn out that its cost is only marginally larger than that of strategy "Clairvoyant".

Strategy "Count". *When classifying the ith element, for $1 \leq i \leq n-2$, proceed as follows: If $s_{i-1} \geq \ell_{i-1}$, compare with p first, otherwise compare with q first.*

The number of additional comparisons of this strategy is called A_n^{ct}, its cost P_n^{ct}.

No algorithmic solution for the classification problem can have cost smaller than "Clairvoyant". Strategy "Count" is algorithmic. Thus any cost-minimal algorithmic classification procedure has cost between $\mathbb{E}(P_n^{\mathrm{cv}})$ and $\mathbb{E}(P_n^{\mathrm{ct}})$, and a precise analysis of both will lead to good lower and upper bounds for the cost of such a procedure. It was shown in [AD15] that $\mathbb{E}(P_n^{\mathrm{ct}}) - \mathbb{E}(P_n^{\mathrm{cv}}) = O(\log n)$ and that, as a consequence, both strategies lead to dual-pivot quicksort algorithms that use $\frac{9}{5}n \log n + O(n)$ comparisons on average. In the following, we carry out a precise analysis of $\mathbb{E}(P_n^{\mathrm{cv}})$ and $\mathbb{E}(P_n^{\mathrm{ct}})$, which will make it possible to determine the expected comparison count of an optimal dual-pivot quicksort algorithm up to $0.28n$.

Lemma 9.1. *(a) The expected number of additional comparisons of strategy "Clairvoyant" is*

$$\mathbb{E}(A_n^{\mathrm{cv}}) = \frac{n}{6} - \frac{7}{12} + \frac{1}{4(n - [n \text{ even}])} - \mathbb{E}\big(X_n^{\searrow}\big).$$

(b) The expected number of additional comparisons of strategy "Count" is

$$\mathbb{E}\big(A_n^{\mathrm{ct}}\big) = \frac{n}{6} - \frac{7}{12} + \frac{1}{4(n - [n \text{ even}])} + \mathbb{E}\big(X_n^{\nearrow}\big).$$

Proof ideas. (The full proof can be found in Appendix J. A different proof of a related statement was given in [AD15].)

(a) Noticing that medium elements can be ignored, we consider a reduced input of size $n' = s + \ell$, consisting only of the s small and the ℓ large elements in the input. For $0 \leq i \leq n'$ let $s_i' = s - s_i$ and $\ell_i' = \ell - \ell_i$ denote the number of small respectively large elements left unclassified after step i. Then $\{(i, s_i' - \ell_i') \mid 0 \leq i \leq n'\}$ is a lattice path with distribution (including the distribution of n') exactly as in Section 7, so that the results on the expected number of zeros on such paths given there may be applied. We also note that the sign of $s_{i-1}' - \ell_{i-1}'$ decides whether the ith element to be classified is compared with p first or with q first, and that additional comparisons correspond to steps on the path that lead away from the horizontal axis, excepting down-from-zero steps (due to the asymmetry in treating the situation $s - s_i = \ell - \ell_i$ in strategy "Clairvoyant"). For the number of steps away from the horizontal axis one easily finds the expression $\min(s, \ell)$. Averaging over all choices for n' and the two pivots leads to the formula claimed in (a).

(b) Now assume strategy "Count" is applied to $n' = s + \ell$ elements. The set $\{(i, s_i - \ell_i) \mid 0 \leq i \leq n'\}$ forms a lattice path that starts at $(0, 0)$ and ends at $(n', s - \ell)$. It can be shown that reflection with respect to the vertical line at $n'/2$ maps these paths in a probability-preserving way to the paths from from (a) (and thus from our model), and it turns out that additional comparisons in this strategy

correspond to steps away from the horizontal axis and up-to-zero steps. As in (a), averaging leads to the formula claimed in (b). □

Lemma 9.1 allows us to give an exact expression for the average number of comparisons of "Clairvoyant" and "Count" in a single partitioning step. The expressions for $\mathbb{E}(P_n^{\mathrm{cv}})$ and $\mathbb{E}(P_n^{\mathrm{cv}})$ are obtained by adding the expected number of necessary comparisons $\frac{4}{3}(n-2)+1$ to the cost terms in Lemma 9.1 (see Appendix J).

10. Main Results and their Asymptotic Aspects

In this section we give precise formulations of our main results. We use the partitioning cost from the previous section to calculate the expected number of comparisons of the two dual-pivot quicksort variants obtained by using classification strategies "Clairvoyant" and "Count", respectively. We call these sorting algorithms "Clairvoyant" and "Count" again. Recall that "Clairvoyant" uses an oracle and is comparison-optimal, and that "Count" is its algorithmic version. We validated our main results in experiments which can be found in Appendix L.

Theorem 10.1. *For $n \geq 4$, the average number of comparisons in the comparison-optimal dual-pivot quicksort algorithm "Clairvoyant" (with oracle) is*

$$\mathbb{E}(C_n^{\mathrm{cv}}) = \frac{9}{5}nH_n + \frac{1}{5}nH_n^{\mathrm{alt}} - \frac{89}{25}n + \frac{77}{40}H_n + \frac{3}{40}H_n^{\mathrm{alt}} + \frac{67}{800} - \frac{(-1)^n}{10} + r_n$$

where

$$r_n = \frac{[n \text{ even}]}{320}\Big(\frac{1}{n-3} + \frac{3}{n-1}\Big) - \frac{[n \text{ odd}]}{320}\Big(\frac{3}{n-2} + \frac{1}{n}\Big).$$

Corollary 10.2. *The average number of comparisons in the dual-pivot quicksort algorithm "Clairvoyant" is*

$$\mathbb{E}(C_n^{\mathrm{cv}}) = \frac{9}{5}n\log n + An + B\log n + C + \frac{D}{n} + \frac{E}{n^2} + \frac{F[n \text{ even}] + G}{n^3} + O\Big(\frac{1}{n^4}\Big)$$

with

$$A = \frac{9}{5}\gamma - \frac{1}{5}\log 2 - \frac{89}{25} = -2.6596412392892\ldots, \qquad B = \frac{77}{40} = 1.925,$$

$$C = \frac{77}{40}\gamma - \frac{3}{40}\log 2 + \frac{787}{800} = 2.042904116393455\ldots, \quad D = \frac{13}{16} = 0.8125,$$

$$E = -\frac{77}{480} = -0.1604166\ldots, \qquad F = \frac{1}{8} = 0.125, \qquad G = -\frac{19}{400} = -0.0475,$$

asymptotically as n tends to infinity.

Before continuing with the second partitioning strategy, let us make a remark on the (non-)influence of the parity of n. It is noteworthy that in Corollary 10.2 no such influence is visible in the first six terms (down to $1/n^2$); only from $1/n^3$ on the parity of n appears. This is somewhat unexpected, since a term $(-1)^n$ appears in Theorem 10.1.

Theorem 10.3. *The average number of comparisons in the dual-pivot quicksort algorithm "Count" is*

$$\mathbb{E}\big(C_n^{\mathrm{ct}}\big) = \frac{9}{5}nH_n - \frac{1}{5}nH_n^{\mathrm{alt}} - \frac{89}{25}n + \frac{67}{40}H_n - \frac{3}{40}H_n^{\mathrm{alt}} - \frac{83}{800} + \frac{(-1)^n}{10} - r_n$$

where r_n is defined in Theorem 10.1.

Again, the asymptotic behavior follows from the exact result.

Corollary 10.4. *The average number of comparisons in the optimal dual-pivot quicksort algorithm "Count" is*

$$\mathbb{E}\big(C_n^{\mathrm{ct}}\big) = \frac{9}{5}n\log n + An + B\log n + C + \frac{D}{n} + \frac{E}{n^2} + \frac{F[n\text{ even}]+G}{n^3} + O\Big(\frac{1}{n^4}\Big)$$

with

$$A = \frac{9}{5}\gamma + \frac{1}{5}\log 2 - \frac{89}{25} = -2.3823823670652\ldots, \qquad B = \frac{67}{40} = 1.675,$$

$$C = \frac{67}{40}\gamma + \frac{3}{40}\log 2 + \frac{637}{800} = 1.81507227725206\ldots, \qquad D = \frac{11}{16} = 0.6875,$$

$$E = -\frac{67}{480} = -0.1395833\ldots, \qquad F = -\frac{1}{8} = -0.125, \qquad G = \frac{31}{400} = 0.0775,$$

asymptotically as n tends to infinity.

The idea of the proofs of Theorems 10.1 and 10.3 is to translate the recurrence (8.1) into a second order differential equation for the generating function $C(z)$ of $\mathbb{E}(C_n)$ in terms of the generating function $P(z)$ of $\mathbb{E}(P_n)$. Integrating twice yields $C(z)$. This generating function then allows extraction of the exact expressions for $\mathbb{E}(C_n)$. The asymptotic results follow. See Appendix K for details.

References

[Abl15]   Jakob Ablinger. *HarmonicSums V1.0*. RISC, 2015. `http://www.risc.jku.at/research/combinatorics/HarmonicSums/`.

[AD15]   Martin Aumüller and Martin Dietzfelbinger. Optimal partitioning for dual-pivot quicksort. *ACM Trans. Algorithms*, 12(2):18:1–18:36, November 2015.

[Dev16]   The SageMath Developers. *SageMath Mathematics Software (Version 7.0)*, 2016. `http://www.sagemath.org`.

[FS09]   Philippe Flajolet and Robert Sedgewick. *Analytic combinatorics*. Cambridge University Press, Cambridge, 2009.

[GKP94]   Ronald L. Graham, Donald E. Knuth, and Oren Patashnik. *Concrete mathematics. A foundation for computer science*. Addison-Wesley, second edition, 1994.

[Hen91]   P. Hennequin. *Analyse en moyenne d'algorithmes : tri rapide et arbres de recherche*. PhD thesis, Ecole Politechnique, Palaiseau, 1991.

[HK15]   Benjamin Hackl and Daniel Krenn. Asymptotic expansions in SageMath. `http://trac.sagemath.org/17601`, 2015. module in SageMath 6.10.beta2.

[Hoa62]   C. A. R. Hoare. Quicksort. *Comput. J.*, 5(1):10–15, 1962.

[Kra15]   Christian Krattenthaler. Lattice path enumeration. In *Handbook of enumerative combinatorics*, Discrete Math. Appl. (Boca Raton), pages 589–678. CRC Press, Boca Raton, FL, 2015.

[Moh79]   Sri Gopal Mohanty. *Lattice path counting and applications*. Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London-Toronto, Ont., 1979. Probability and Mathematical Statistics.

[Sch07]   Carsten Schneider. Symbolic summation assists combinatorics. *Sém. Lothar. Combin.*, 56:1–36, 2007.

[Sch15a] Carsten Schneider. *EvaluateMultiSums V0.96*. RISC, 2015. Unpublished.

[Sch15b] Carsten Schneider. *Sigma V1.81*. RISC, 2015. `http://www.risc.jku.at/research/combinatorics/Sigma/`.

[Sed75]   Robert Sedgewick. *Quicksort*. PhD thesis, Standford University, 1975.

[Wil13]   Sebastian Wild. Java 7's dual pivot quicksort. Master's thesis, University of Kaiserslautern, 2013. `https://kluedo.ub.uni-kl.de/files/3463/wild-master-thesis.pdf`.

[WNN15]   Sebastian Wild, Markus E. Nebel, and Ralph Neininger. Average case and distributional analysis of dual-pivot quicksort. *ACM Transactions on Algorithms*, 11(3):22, 2015.

[Yar09]   Vladimir Yaroslavskiy. Replacement of quicksort in java.util.arrays with new dual-pivot quicksort. `http://permalink.gmane.org/gmane.comp.java.openjdk.core-libs.devel/2628`, 2009. Archived version of the discussion in the OpenJDK mailing list.
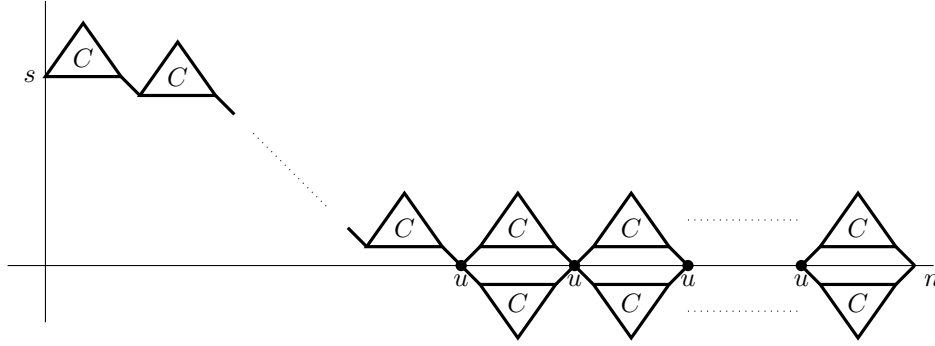
FIGURE A.1. Decomposition of a lattice path for $s \geq 0$ marking zeros.

APPENDIX A. USING THE GENERATING FUNCTION MACHINERY

Theorem A.1. *For a randomly (as described in Section 3) chosen path of length n, the expected number of zeros is*

$$\mathbb{E}(Z_n) = \frac{4}{n+1} \sum_{0 \leq k < \ell < \lceil n/2 \rceil} \frac{\binom{n}{k}}{\binom{n}{\ell}} + [n \text{ even}] \frac{1}{n+1} \left(\frac{2^n}{\binom{n}{n/2}} - 1 \right) + 1.$$

The remaining part of this section is devoted to the proof of this theorem. The main technique is to model our lattice paths by means of combinatorial classes and generating functions. For simplicity of notation, we denote a combinatorial class by its corresponding generating function.

Concerning the generating functions, we mark a step to the right by the variable z and a zero (except the last) by u. Note that we do not mark the zero at $(n, 0)$ for technical reasons; we'll take this into account at the end by adding a 1 to the final result. Thus, the coefficient of $z^n u^{r-1}$ of the function $Q_s(z, u)$ (the generating function of all paths starting in $(0, s)$ and ending in some $(n, 0)$) equals the number of paths of length n and exactly r zeros.

We also need the following auxiliary function. The generating function $C(z)$ counts all Catalan paths, i.e., paths starting and ending at the same height, but not going below it. This equals

$$C(z) = \frac{1 - \sqrt{1 - 4z^2}}{2z^2}.$$

In Figure A.1, we give a schematic decomposition of a path from $(0, s)$ to $(n, 0)$ for non-negative s. This decomposition translates to the following parts of the generating function (the path is read from the left to the right).

- We start by s consecutive blocks of $C(z)$, each followed by a single descent encoded as z. This gives the paths from $(0, s)$ to their first zero (i.e., where it touches the horizontal axis for the first time).
- We mark this zero by the symbol u.
- We either do a single ascent or a single decent (marked by a z), then continue with a $C(z)$-block and do a single decent or ascent respectively (marked by a z as well) again. Thus, we are back at a zero.
- We repeat such up/down blocks $z\,C(z)\,z$, each one preceded by a zero u, a finite number of times.

If $s < 0$, then the construction is the same, but everything is reflected at the horizontal axis.

Continuing using the symbolic method—the description above is already part of it, see, for example, Flajolet and Sedgewick [FS09]—the decomposition above translates to the generating function

$$Q_s(z, u) = \frac{C(z)^{|s|} z^{|s|}}{1 - 2uz^2 C(z)}, \tag{A.1}$$

which we will use from now on. Note that the coefficient 2 reflects the fact that there are two choices (up and down) for the blocks between zeros.

To obtain a nice explicit form, we perform a change of variables. The result is stated in the following lemma.

Lemma A.2. *With the transformation $z = v/(1 + v^2)$ we have*

$$Q_s(z, u) = \frac{v^{|s|}(1 + v^2)}{1 - v^2(2u - 1)}.$$

Proof. Transforming the counting generating function of Catalan paths yields

$$C(z) = 1 + v^2.$$

Thus (A.1) becomes

$$Q_s(z, u) = (1 + v^2)^{|s|} \left(\frac{v}{1 + v^2}\right)^{|s|} \frac{1}{1 - 2u\left(\frac{v}{1+v^2}\right)^2 (1 + v^2)}$$

and can be simplified to the expression stated in the lemma. □

The next step is to extract the coefficients out of the expressions obtained in the previous lemma. First we rewrite the extraction of the coefficients from the "z-world" to the "v-world", see Lemma A.3. Afterwards, in Lemma A.4, the coefficients can be determined quite easily.

Lemma A.3. *Let $F(z)$ be an analytic function in a neighborhood of the origin. Then we have*

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

Proof. We use Cauchy's formula to extract the coefficients of $F(z)$ as

$$[z^n]F(z) = \frac{1}{2\pi i} \oint_{\mathcal{D}} \frac{dz}{z^{n+1}} F(z)$$

where $\mathcal{D}$ is a positively oriented small circle around the origin. Under the transformation $z = v/(1 + v^2)$, the circle $\mathcal{D}$ is transformed to a contour $\mathcal{D}'$ which still winds exactly once around the origin. Using Cauchy's formula again, we obtain

$$[z^n]F(z) = \frac{1}{2\pi i} \oint_{\mathcal{D}'} \frac{dv(1 - v^2)}{(1 + v^2)^2} \frac{(1 + v^2)^{n+1}}{v^{n+1}} F\left(\frac{v}{1 + v^2}\right)$$

$$= [v^n](1 - v^2)(1 + v^2)^{n-1} F\left(\frac{v}{1 + v^2}\right).$$

□

Now we are ready to calculate the desired coefficients.

Lemma A.4. *Suppose $n \equiv s$ (mod 2). Then we have*

$$[z^n]Q_s(z,1) = \binom{n}{(n-s)/2}$$

and, moreover,

$$[z^n] \frac{\partial}{\partial u} Q_s(z,u)\Big|_{u=1} = 2 \sum_{k=0}^{(n-|s|)/2-1} \binom{n}{k}.$$

Proof. As $n \equiv s$ (mod 2), the number $n - s$ is even, and so we can set $\ell = \frac{1}{2}(n-s)$. Then $[z^n]Q_s(z,1)$ is the number of paths from $(0,s)$ to $(n,0)$. These paths have ℓ up steps and $n - \ell$ down steps; thus there are $\binom{n}{\ell}$ many such paths.

For the second part of this lemma, we restrict ourselves to $s \geq 0$ (otherwise use $-s$ and the symmetry in s of the generating function (A.1) instead). We start with the result of Lemma A.2. Taking the first derivative and setting $u = 1$ yields

$$\frac{\partial}{\partial u} Q_s(z,u)\Big|_{u=1} = \frac{2v^{s+2}(1+v^2)}{(1-v^2)^2}.$$

Thus, by using Lemma A.3, we get

$$[z^n] \frac{2v^{s+2}(1+v^2)}{(1-v^2)^2} = 2\,[v^{n-s-2}] \frac{(1+v^2)^n}{1-v^2}.$$

We use ℓ as above and get

$$[v^{n-s-2}] \frac{(1+v^2)^n}{1-v^2} = [v^{2\ell-2}] \frac{(1+v^2)^n}{1-v^2} = [v^{\ell-1}] \frac{(1+v)^n}{1-v} = \sum_{k=0}^{\ell-1} \binom{n}{k},$$

which was claimed to hold.   □

We are now ready to prove the main theorem (Theorem A.1) of this section, which provides an expression for the expected number of zeros. This exact expression is written as a double sum.

Proof of Theorem A.1. By Lemma A.4, the average number of zeros (except the zero at the end point) of a path of length n which starts in $(0,s)$ is

$$\mu_{n,s} = \frac{[z^n] \frac{\partial}{\partial u} Q_s(z,u)\big|_{u=1}}{[z^n]Q_s(z,1)} = \frac{2}{\binom{n}{\ell}} \sum_{k=0}^{\ell-1} \binom{n}{k},$$

where we have set $\ell = \frac{1}{2}(n - |s|)$ as in the proof of Lemma A.4. If $s = 0$, this simplifies to

$$\mu_{n,0} = \frac{2}{\binom{n}{n/2}} \sum_{k=0}^{n/2-1} \binom{n}{k} = \frac{2^n}{\binom{n}{n/2}} - 1. \tag{A.2}$$

If $n \not\equiv s$ (mod 2), then we set $\mu_{n,s} = 0$.

Summing up yields

$$\sum_{s=-n}^{n} \mu_{n,s} = 2 \sum_{s=1}^{n} \mu_{n,s} + \mu_{n,0} = 4 \sum_{\ell=0}^{\lceil n/2 \rceil - 1} \frac{1}{\binom{n}{\ell}} \sum_{k=0}^{\ell-1} \binom{n}{k} + \mu_{n,0}$$

$$= 4 \sum_{0 \leq k < \ell < \lceil n/2 \rceil} \frac{\binom{n}{k}}{\binom{n}{\ell}} + [n \text{ even}] \left(\frac{2^n}{\binom{n}{n/2}} - 1 \right).$$

Dividing by the number $n+1$ of possible starting points and adding 1 for the zero at $(n,0)$ completes the proof of Theorem A.1. $\qquad\square$

Appendix B. Appendix to Section 4: A Probabilistic Approach

The following remark explains the equivalence between the urn model (Section 4) and the lattice path model (Section 3).

Remark B.1. Only sequences with $W_0 \equiv n \pmod 2$ and $W_i - W_{i-1} \in \{1, -1\}$ for $1 \le i \le n$ and $W_n = 0$ can have a probability bigger than 0. We denote the set of these paths by $\mathcal{P}_n$.

Now fix w_0 with $|w_0| \le n$ and $w_0 \equiv n \pmod 2$. Clearly, all sequences $p_n \in \mathcal{P}_n$ starting with $(0, w_0)$ together have probability $1/(n+1)$. What is the probability of a path $p_n = ((0, w_0), (1, w_1), \ldots, (n, w_n)) \in \mathcal{P}_n$ to appear as trajectory of the random walk?

The starting point $(0, w_0)$ fixes the number of black and red balls as $r = (n+w_0)/2$ and $b = (n - w_0)/2$, respectively. Let us number the n balls arbitrarily with $1, \ldots, n$. The random experiment in essence arranges the balls in one of the $n!$ possible orders, each of these orders having the same probability. A trajectory is determined not by this order, but by the pattern of red and black balls in the sequence. Thus $r!\,b!$ orders lead to the same trajectory p_n. Thus, given w_0, all $\binom{n}{r,b}$ paths starting in $(0, w_0)$ are equally likely. Therefore this urn model is equivalent to the model given in Section 3.

Proof of Lemma 4.2. We compute the transition probabilities by considering the urn model. The probabilty of the event $U_{i+1} = $ black is the number of remaining black balls divided by the number of remaining balls $n - i$. If $W_i = \ell$, then there are still $(n - i + \ell)/2$ red and $(n - i - \ell)/2$ black balls in the urn.

Thus

$$\begin{aligned}
\mathbb{P}(W_{i+1} = \ell + 1 \mid W_i = \ell) &= \frac{n - i - \ell}{2(n-i)}, \\
\mathbb{P}(W_{i+1} = \ell - 1 \mid W_i = \ell) &= \frac{n - i + \ell}{2(n-i)}.
\end{aligned} \tag{B.1}$$

We now prove the claim by backwards induction on m. For $m = n$, the assertion follows directly from the probability model.

For $m < n$, we obtain

$$\begin{aligned}
\mathbb{P}((n-m,k) \in P_n) &= \mathbb{P}(W_{n-m} = k) \\
&= \mathbb{P}(W_{n-m} = k \mid W_{n-m-1} = k - 1)\,\mathbb{P}(W_{n-m-1} = k - 1) \\
&\quad + \mathbb{P}(W_{n-m} = k \mid W_{n-m-1} = k + 1)\,\mathbb{P}(W_{n-m-1} = k + 1) \\
&= \frac{(m+1-(k-1)) + (m+1+(k+1))}{(m+2)\,2(m+1)} = \frac{1}{m+1}
\end{aligned}$$

by (B.1) and the induction hypothesis. $\qquad\square$

Appendix C. Identity

Using the two results of Section 4 and Appendix A we can show the following identity in a combinatorial way.

Corollary and Theorem C.1. For $n \geq 0$, we have the four equal expressions

$$\frac{4}{n+1} \sum_{0 \leq k < \ell < \lceil n/2 \rceil} \frac{\binom{n}{k}}{\binom{n}{\ell}} + [n \text{ even}] \frac{1}{n+1} \left(\frac{2^n}{\binom{n}{n/2}} - 1 \right) + 1 \tag{i}$$

$$= \frac{2}{\lfloor n/2 \rfloor + 1} \sum_{0 \leq k < \ell \leq \lfloor n/2 \rfloor} \frac{\binom{2\lfloor n/2 \rfloor + 1}{k}}{\binom{2\lfloor n/2 \rfloor + 1}{\ell}} + 1 \tag{ii}$$

$$= \frac{1}{n+1} \sum_{m=0}^{\lfloor n/2 \rfloor} \sum_{\ell=m}^{n-m} \frac{\binom{2m}{m}\binom{n-2m}{\ell-m}}{\binom{n}{\ell}} \tag{iii}$$

$$= H_{n+1}^{\text{odd}}. \tag{iv}$$

We note that the expressions (i) and (ii) are obviously equal for odd n. Furthermore, (ii) and (iv) only change when n increases by 2 (to be precise from odd n to even n). Once we prove that (i) equals (iv) for all $n \geq 0$, it follows that both expressions are equal to (ii) for all $n \geq 0$.

Combinatorial proof of Corollary C.1. Combine the results of Theorems A.1 and 4.1 to see that (i) and (iv) are equal.

Expression (iii) for the expected number of zeros, but without taking the zero at $(n, 0)$ into account, has been given in [AD15, displayed equation after (14)]: The number of paths from $(0, n - 2\ell)$ to $(n, 0)$ is $\binom{n}{\ell}$, whereas the number of paths from $(0, n - 2\ell)$ via $(n - 2m, 0)$ to $(n, 0)$ is $\binom{n-2m}{\ell-m}\binom{2m}{m}$. Summing over all possible m and all possible ℓ and dividing by $(n + 1)$ for the equidistribution of the starting point yields (iii). □

Beside this combinatorial proof, we intend to show Theorem C.1 in alternative ways. First, we prove that the identity between (iii) and (iv).

Proof of (iii) *equals* (iv). Consider

$$\frac{1}{n+1} \sum_{\ell=m}^{n-m} \frac{\binom{2m}{m}\binom{n-2m}{\ell-m}}{\binom{n}{\ell}} = \frac{1}{n+1} \sum_{\ell=m}^{n-m} \frac{(2m)! \, (n-2m)! \, \ell! \, (n-\ell)!}{m! \, m! \, (\ell-m)! \, (n-\ell-m)! \, n!}$$

$$= \frac{1}{(n+1)\binom{n}{2m}} \sum_{\ell=m}^{n-m} \binom{\ell}{m}\binom{n-\ell}{m}$$

$$= \frac{1}{(n+1)\binom{n}{2m}} \binom{n+1}{2m+1} = \frac{1}{2m+1},$$

where [GKP94, (5.26)] has been used in the penultimate step with $\ell = n$, $q = 0$, $k = \ell$, $m = m$ and $n = m$. Summing over m yields $\sum_{m=0}^{\lfloor n/2 \rfloor} 1/(2m+1) = H_{n+1}^{\text{odd}}$, thus the equality between (iv) and (iii). □

It remains to give a computational proof of the equality between (i) and (iv). We provide two proofs: one motivated by "creative telescoping" (Appendix C.1) and one completely elementary "human" proof (Appendix C.2) using not more than Vandermonde's convolution.

C.1. **Proof of the Identity Using Creative Telescoping.** A computational proof of the identity between (i) and (iv) can be generated by the summation package Sigma [Sch15b] (see also Schneider [Sch07]) together with the packages HarmonicSums [Abl15] and EvaluateMultiSums [Sch15a][3]. To succeed, we have to split the case of even and odd n. The obtained proof certificates are rather lengthy to verify.

Motivated by the previous observations, we also give a proof without additional machinery. Anyhow, the key step is, as with Sigma, creative telescoping. We prove an (easier) identity which Sigma comes up with in the following lemma.

Lemma C.2. *Let*

$$F(n, \ell) = \sum_{0 \leq k < \ell} \frac{\binom{n}{k}}{\binom{n}{\ell}},$$

$$G(n, \ell) = (\ell - 1) + (\ell - 1 - n)F(n, \ell)$$

for $0 \leq \ell \leq n$. Then

$$(n + 1)F(n + 1, \ell) - (n + 2)F(n, \ell) = G(n, \ell + 1) - G(n, \ell) \qquad (C.1)$$

holds for all $0 \leq \ell < n$.

Proof. For $0 \leq \ell < n$, we first compute

$$F(n + 1, \ell) = \frac{1}{\binom{n+1}{\ell}} \sum_{0 \leq k < \ell} \binom{n+1}{k} = \frac{1}{\binom{n+1}{\ell}} \sum_{0 \leq k < \ell} \left(\binom{n}{k-1} + \binom{n}{k} \right)$$

$$= -\frac{\binom{n}{\ell-1}}{\binom{n+1}{\ell}} + \frac{2\binom{n}{\ell}}{\binom{n+1}{\ell}} F(n, \ell) = -\frac{\ell}{n+1} + 2\frac{n+1-\ell}{n+1} F(n, \ell) \qquad (C.2)$$

and

$$F(n, \ell + 1) = \frac{1}{\binom{n}{\ell+1}} \sum_{0 \leq k < \ell + 1} \binom{n}{k} = \frac{\binom{n}{\ell}}{\binom{n}{\ell+1}} + \frac{\binom{n}{\ell}}{\binom{n}{\ell+1}} F(n, \ell)$$

$$= \frac{\ell + 1}{n - \ell} + \frac{\ell + 1}{n - \ell} F(n, \ell). \qquad (C.3)$$

By plugging (C.2) and (C.3) into (C.1), all occurrences of $F(n, \ell)$ cancel as well as all other terms, which proves (C.1). □

We are now able to prove the essential recurrence for the sum (i) in Theorem C.1.

Lemma C.3. *Let*

$$E_n = \frac{4}{n + 1} \sum_{0 \leq k < \ell < \lceil n/2 \rceil} \frac{\binom{n}{k}}{\binom{n}{\ell}} + [n \text{ even}] \frac{1}{n + 1} \left(\frac{2^n}{\binom{n}{n/2}} - 1 \right).$$

Then

$$E_{2N+1} - E_{2N} = 0 \qquad and \qquad E_{2N+2} - E_{2N+1} = \frac{1}{2N + 3} \qquad (C.4)$$

for $N \geq 0$.

Proof. Multiplying (C.1) with $4/((n+1)(n+2))$ and summing over $0 \le \ell < \lceil n/2 \rceil$ yields

$$\frac{4}{n+2} \sum_{0 \le k < \ell < \lceil n/2 \rceil} \frac{\binom{n+1}{k}}{\binom{n+1}{\ell}} - \frac{4}{n+1} \sum_{0 \le k < \ell < \lceil n/2 \rceil} \frac{\binom{n}{k}}{\binom{n}{\ell}}$$
$$= \frac{4}{(n+1)(n+2)} \Big(G(n, \lceil n/2 \rceil) - G(n, 0) \Big)$$

for $n \ge 0$. This is equivalent to

$$\frac{4}{n+2} \sum_{0 \le k < \ell < \lceil (n+1)/2 \rceil} \frac{\binom{n+1}{k}}{\binom{n+1}{\ell}} - \frac{4[n \text{ even}]}{n+2} F(n+1, n/2) - \frac{4}{n+1} \sum_{0 \le k < \ell < \lceil n/2 \rceil} \frac{\binom{n}{k}}{\binom{n}{\ell}}$$
$$= \frac{4}{(n+1)(n+2)} \Big(\lceil n/2 \rceil - 1 + (\lceil n/2 \rceil - 1 - n) F(n, \lceil n/2 \rceil) + 1 \Big). \quad \text{(C.5)}$$

We rewrite the double sums in terms of E_n and E_{n+1}, respectively, and use (A.2). We also replace $F(n+1, n/2)$ by (C.2). Then (C.5) is equivalent to

$$E_{n+1} - \frac{2[n \text{ odd}]}{n+2} F(n+1, (n+1)/2) - E_n$$
$$+ \frac{2n[n \text{ even}]}{(n+1)(n+2)} - \frac{2[n \text{ even}]}{n+1} F(n, n/2)$$
$$= \frac{4}{(n+1)(n+2)} \Big(\lceil n/2 \rceil - (\lfloor n/2 \rfloor + 1) F(n, \lceil n/2 \rceil) \Big). \quad \text{(C.6)}$$

If $n = 2N+1$, equation (C.6) is equivalent to

$$E_{2N+2} - E_{2N+1} - \frac{2}{2N+3} F(2N+2, N+1)$$
$$= \frac{2}{2N+3} - \frac{2}{2N+3} F(2N+1, N+1).$$

Using (C.2), this is equivalent to the second recurrence in (C.4).

If $n = 2N$, then (C.6) is equivalent to the first recurrence in (C.4). $\qquad \square$

Computational proof of Theorem C.1. The definition of E_0 in Lemma C.3 implies that $E_0 = 0$. Thus (i) and (iv) coincide for $n = 0$. This can be extended to all $n \ge 0$ by induction on n and Lemma C.3. $\qquad \square$

C.2. **Proof of the Identity Using Vandermonde's Convolution.** We now provide a completely elementary "human" proof of the identity between (i) and (iv).

We first prove an identity on binomial coefficients.

Lemma C.4. *The identity*

$$\sum_{0 \le k \le j} \binom{n}{k} = \sum_{0 \le k \le j} 2^k \binom{n-k-1}{j-k}$$

holds for all non-negative integers $j < n$.

Proof. We denote the right hand side by ρ. The binomial theorem and symmetry of the binomial coefficient yield

$$\rho = \sum_{0 \le i \le k \le j} \binom{k}{i} \binom{n-k-1}{n-1-j}.$$

The sum over k can be evaluated by [GKP94, (5.26)], resulting in

$$\rho = \sum_{0 \le i \le j} \binom{n}{n+i-j}.$$

Symmetry of the binomial coefficient and then replacing i by $j-i$ lead to

$$\rho = \sum_{0 \le i \le j} \binom{n}{j-i} = \sum_{0 \le i \le j} \binom{n}{i}.$$

$\square$

We are now able to establish a recurrence satisfied by the double sum in (i).

Lemma C.5. *For $n \ge 0$, let*

$$S_n = \sum_{0 \le k < \ell < \lceil \frac{n}{2} \rceil} \frac{\binom{n}{k}}{\binom{n}{\ell}}.$$

Then the recurrence

$$S_n = \frac{n+1}{n-1} S_{n-2} + \frac{n+1}{4n} + [n \text{ even}] \left(\frac{2^{n-2}}{n \binom{n}{n/2}} - \frac{1}{2(n-1)} - \frac{1}{4n} \right)$$

holds for $n \ge 2$.

Proof. We replace the sum in the numerator of S_n using Lemma C.4 and obtain

$$\begin{aligned}
S_n &= \sum_{0 \le k \le \ell < \lceil \frac{n}{2} \rceil} \frac{\binom{n}{k}}{\binom{n}{\ell}} - \left\lceil \frac{n}{2} \right\rceil \\
&= \sum_{0 \le k \le \ell < \lceil \frac{n}{2} \rceil} \frac{2^k \binom{n-k-1}{\ell-k}}{\binom{n}{\ell}} - \left\lceil \frac{n}{2} \right\rceil \\
&= \sum_{0 \le k \le \ell < \lceil \frac{n}{2} \rceil} \frac{2^k (n-k-1)! \, k!}{n!} \binom{\ell}{k} \big((n-k) - (\ell-k) \big) - \left\lceil \frac{n}{2} \right\rceil \\
&= \sum_{0 \le k \le \ell < \lceil \frac{n}{2} \rceil} \frac{2^k (n-k)! \, k!}{n!} \binom{\ell}{k} \\
&\quad - \sum_{0 \le k \le \ell < \lceil \frac{n}{2} \rceil} \frac{2^k (n-k-1)! \, (k+1)!}{n!} \binom{\ell}{k+1} - \left\lceil \frac{n}{2} \right\rceil.
\end{aligned}$$

The sum over ℓ can be evaluated using upper summation, see [GKP94, Table 174]:

$$S_n = \sum_{0 \le k < \left\lceil \frac{n}{2} \right\rceil} \frac{2^k (n-k)! \, k!}{n!} \binom{\left\lceil \frac{n}{2} \right\rceil}{k+1}$$

$$- \sum_{0 \le k < \left\lceil \frac{n}{2} \right\rceil} \frac{2^k (n-k-1)! \, (k+1)!}{n!} \binom{\left\lceil \frac{n}{2} \right\rceil}{k+2} - \left\lceil \frac{n}{2} \right\rceil.$$

Shifting the summation index in the second sum leads to

$$S_n = \sum_{0 \le k < \left\lceil \frac{n}{2} \right\rceil} \frac{2^k (n-k)! k!}{n!} \binom{\left\lceil \frac{n}{2} \right\rceil}{k+1}$$

$$- \sum_{0 \le k < \left\lceil \frac{n}{2} \right\rceil} \frac{2^{k-1} (n-k)! k!}{n!} \binom{\left\lceil \frac{n}{2} \right\rceil}{k+1} - \left\lceil \frac{n}{2} \right\rceil + \frac{1}{2} \left\lceil \frac{n}{2} \right\rceil$$

$$= \frac{1}{2n!} \sum_{0 \le k < \left\lceil \frac{n}{2} \right\rceil} 2^k (n-k)! k! \binom{\left\lceil \frac{n}{2} \right\rceil}{k+1} - \frac{1}{2} \left\lceil \frac{n}{2} \right\rceil$$

$$= \frac{\left\lceil \frac{n}{2} \right\rceil!}{2n!} \sum_{0 \le k < \left\lceil \frac{n}{2} \right\rceil} 2^k \frac{(n-k)!}{(\left\lceil \frac{n}{2} \right\rceil - k - 1)!} \frac{1}{k+1} - \frac{1}{2} \left\lceil \frac{n}{2} \right\rceil. \tag{C.7}$$

We intend to derive a recurrence linking S_n with S_{n-2}. Therefore, for $n \ge 2$, we rewrite S_n as

$$S_n = \frac{\left\lceil \frac{n}{2} \right\rceil!}{2n!} \sum_{0 \le k < \left\lceil \frac{n}{2} \right\rceil - 1} 2^k \frac{(n-k-2)!}{(\left\lceil \frac{n}{2} \right\rceil - k - 2)!} \frac{(n-k)(n-k-1)}{(\left\lceil \frac{n}{2} \right\rceil - k - 1)(k+1)} - \frac{\left\lceil \frac{n}{2} \right\rceil}{2} + \frac{2^{\left\lceil \frac{n}{2} \right\rceil - 2}}{\binom{n}{\left\lceil \frac{n}{2} \right\rceil - 1}}.$$

Partial fraction decomposition in k yields

$$S_n = \frac{\left\lceil \frac{n}{2} \right\rceil!}{2n!} \sum_{0 \le k < \left\lceil \frac{n}{2} \right\rceil - 1} 2^k \frac{(n-k-2)!}{(\left\lceil \frac{n}{2} \right\rceil - k - 2)!} \left(-1 + \frac{(n - \left\lceil \frac{n}{2} \right\rceil)(n - \left\lceil \frac{n}{2} \right\rceil + 1)}{(\left\lceil \frac{n}{2} \right\rceil - k - 1)\left\lceil \frac{n}{2} \right\rceil} + \frac{n(n+1)}{(k+1)\left\lceil \frac{n}{2} \right\rceil} \right)$$

$$- \frac{\left\lceil \frac{n}{2} \right\rceil}{2} + \frac{2^{\left\lceil \frac{n}{2} \right\rceil - 2}}{\binom{n}{\left\lceil \frac{n}{2} \right\rceil - 1}}$$

$$= - \frac{\left\lceil \frac{n}{2} \right\rceil! (n - \left\lceil \frac{n}{2} \right\rceil)!}{2n!} \sum_{0 \le k < \left\lceil \frac{n}{2} \right\rceil - 1} 2^k \binom{n-k-2}{\left\lceil \frac{n}{2} \right\rceil - k - 2}$$

$$+ \frac{(\left\lceil \frac{n}{2} \right\rceil - 1)! (n - \left\lceil \frac{n}{2} \right\rceil + 1)!}{2n!} \sum_{0 \le k < \left\lceil \frac{n}{2} \right\rceil - 1} 2^k \binom{n-k-2}{\left\lceil \frac{n}{2} \right\rceil - k - 1}$$

$$+ \frac{(\left\lceil \frac{n}{2} \right\rceil - 1)! (n+1)n}{2n!} \sum_{0 \le k < \left\lceil \frac{n}{2} \right\rceil - 1} 2^k \frac{(n-k-2)!}{(\left\lceil \frac{n}{2} \right\rceil - k - 2)!} \frac{1}{(k+1)}$$

$$- \frac{\left\lceil \frac{n}{2} \right\rceil}{2} + \frac{2^{\left\lceil \frac{n}{2} \right\rceil - 2}}{\binom{n}{\left\lceil \frac{n}{2} \right\rceil - 1}}.$$

We again use Lemma C.4 for the first two summands and (C.7) with n replaced by $n-2$ for the third summand to obtain

$$S_n = -\frac{1}{2\binom{n}{\lceil\frac{n}{2}\rceil}} \sum_{0\leq k<\lceil\frac{n}{2}\rceil-1} \binom{n-1}{k} + \frac{1}{2\binom{n}{\lceil\frac{n}{2}\rceil-1}} \sum_{0\leq k<\lceil\frac{n}{2}\rceil} \binom{n-1}{k}$$
$$+ \frac{n+1}{n-1}\Big(S_{n-2} + \frac{\lceil\frac{n}{2}\rceil-1}{2}\Big) - \frac{\lceil\frac{n}{2}\rceil}{2}.$$

Adding another summand for $k = \lceil\frac{n}{2}\rceil - 1$ to the first sum leads to

$$S_n = \Big(\frac{1}{2\binom{n}{\lceil\frac{n}{2}\rceil-1}} - \frac{1}{2\binom{n}{\lceil\frac{n}{2}\rceil}}\Big) \sum_{0\leq k<\lceil\frac{n}{2}\rceil} \binom{n-1}{k}$$
$$+ \frac{\binom{n-1}{\lceil\frac{n}{2}\rceil-1}}{2\binom{n}{\lceil\frac{n}{2}\rceil}} + \frac{n+1}{n-1}S_{n-2} + \frac{\lceil\frac{n}{2}\rceil-1}{n-1} - \frac{1}{2}$$
$$= \frac{n-2\lceil\frac{n}{2}\rceil+1}{2\lceil\frac{n}{2}\rceil\binom{n}{\lceil\frac{n}{2}\rceil}} \sum_{0\leq k<\lceil\frac{n}{2}\rceil} \binom{n-1}{k}$$
$$+ \frac{n+1}{n-1}S_{n-2} + \frac{\lceil\frac{n}{2}\rceil}{2n} + \frac{\lceil\frac{n}{2}\rceil-1}{n-1} - \frac{1}{2}.$$

If n is odd, then $n - 2\lceil\frac{n}{2}\rceil + 1 = 0$ so that the first summand vanishes. The result follows in that case.

For even n, the remaining sum is 2^{n-2} and the result follows. $\qquad\square$

Computational proof of Theorem C.1. Denote the expression in (i) by E_n. From Lemma C.5, we obtain the recurrence

$$E_n = E_{n-2} + \frac{1}{n + [n \text{ even}]}$$

for $n \geq 2$. As $E_0 = E_1 = 1$, this implies that $E_n = H_{n+1}^{\text{odd}}$. Thus (i) and (iv) coincide. $\qquad\square$

Appendix D. Asymptotic Aspects

Lemma D.1. *We have*

$$H_n = \log n + \gamma + \frac{1}{2n} - \frac{1}{12n^2} + O\Big(\frac{1}{n^4}\Big), \tag{D.1a}$$

$$H_n^{\text{odd}} = \frac{1}{2}\log n + \frac{\gamma + \log 2}{2} + \frac{[n \text{ odd}]}{2n} + \frac{3[n \text{ even}]-2}{12n^2} + O\Big(\frac{1}{n^4}\Big), \tag{D.1b}$$

$$H_n^{\text{alt}} = -\log 2 + \frac{(-1)^n}{2n} - \frac{(-1)^n}{4n^2} + O\Big(\frac{1}{n^4}\Big) \tag{D.1c}$$

as n tends to ∞, and where $\gamma = 0.5772156649\ldots$ is the Euler–Mascheroni constant.

Proof. The asymptotic expansion (D.1a) is well-known, cf. for instance [GKP94, (9.88)].

We write $\alpha_n = [n \text{ even}]$ and thus $\lfloor n/2 \rfloor = (n - 1 + \alpha_n)/2$. As α_n is obviously bounded, we can simply plugin this expression into the asymptotic expansions and

simplify all occurring higher powers of α_n by the fact that $\alpha_n^2 = \alpha_n$. Using the relations $H_n^{\mathrm{odd}} = H_n - H_{\lfloor n/2 \rfloor}/2$ and $H_n^{\mathrm{alt}} = H_n - 2H_n^{\mathrm{odd}}$ leads to the expansions (D.1b) and (D.1c), respectively.

The actual asymptotic computations[4] have been carried out using the asymptotic expansions module [HK15] of SageMath [Dev16]. □

Corollary D.2. *The expected number of zeros is*

$$\mathbb{E}(Z_n) = \frac{1}{2}\log n + \frac{\gamma + \log 2}{2}$$
$$+ \frac{1 + [n \text{ even}]}{2n} - \frac{2 + 9[n \text{ even}]}{12n^2} + \frac{[n \text{ even}]}{n^3} + O\left(\frac{1}{n^4}\right)$$

asymptotically as n tends to infinity.

Proof. Combine Theorem 4.1 and Lemma D.1. □

Appendix E. Distribution

In this part of this article, we study the distribution of the number of zeros. As for the expectation $\mathbb{E}(Z_n)$, we get again an exact formula, as well as an asymptotic formula. We start with the former, which is stated in the following theorem; the latter is stated directly afterwards.

Theorem E.1. *Let $r \in \mathbb{N}_0$. For positive lengths $n \geq 2r - 2$, the probability that a randomly chosen path P_n has exactly r zeros is*

$$\mathbb{P}(Z_n = r) = \frac{2^r}{n+1} \frac{\binom{\lceil n/2 \rceil}{r}}{\binom{n}{r}} \left(\frac{2\lceil n/2 \rceil}{r(r+1)} + \frac{r-1}{r+1} + [n \text{ even}]\frac{1}{r} \right)$$
$$+ [n \text{ even}] \frac{2^{r-1}(r-1)}{(n+1)r} \frac{\binom{n/2}{r-1}}{\binom{n}{r}}$$

and we have $\mathbb{P}(Z_0 = r) = [r = 1]$. Otherwise $\mathbb{P}(Z_n = r) = 0$.

This exact formula admits a local limit theorem towards a discrete distribution. The details are as follows.

Corollary E.2. *Let $0 < \varepsilon \leq \frac{1}{2}$. For positive integers r with $r = O\left(n^{1/2-\varepsilon}\right)$, we have asymptotically*

$$\mathbb{P}(Z_n = r) = \frac{1}{r(r+1)}\left(1 + O\left(1/n^{2\varepsilon}\right)\right)$$

as n tends to infinity.

Proof of Theorem E.1. Again, we assume $s \geq 0$ (by symmetry of the generating function (A.1)). Note that Q_s counts the number of zeros by the variable u except for the last zero (at $(n,0)$). By starting with Lemma A.2 and some rewriting, we can extract the $(r-1)$st coefficient with respect to u as

$$[u^{r-1}]Q_s(z,u) = [u^{r-1}]\frac{v^s}{1 - u\frac{2v^2}{1+v^2}} = \frac{2^{r-1}v^{2(r-1)+s}}{(1+v^2)^{r-1}}.$$

---

[4]A worksheet containing the computations can be found at `http://www.danielkrenn.at/downloads/quicksort-paths/quicksort-paths.ipynb`.

Next, we extract the coefficient of z^n. We use Lemma A.3 to obtain

$$[z^n u^{r-1}]Q_s(z,u) = [v^n](1-v^2)(1+v^2)^{n-1}\frac{2^{r-1}v^{2(r-1)+s}}{(1+v^2)^{r-1}}$$
$$= 2^{r-1}[v^{n-s-2(r-1)}](1-v^2)(1+v^2)^{n-r}.$$

We set $\ell = \frac{1}{2}(n-s)$ and get

$$[z^n u^{r-1}]Q_s(z,u) = 2^{r-1}[v^{2\ell-2r+2}](1-v^2)(1+v^2)^{n-r}$$
$$= 2^{r-1}[v^{\ell-r+1}](1-v)(1+v)^{n-r}$$
$$= 2^{r-1}\binom{n-r}{\ell-r+1} - 2^{r-1}\binom{n-r}{\ell-r}$$
$$= 2^{r-1}\binom{n-r}{n-\ell-1} - 2^{r-1}\binom{n-r}{n-\ell}.$$

Note that we have to assume $n - r \geq 0$ to make this work. Otherwise, anyhow, there are no paths with exactly r zeros (and positive length n).

If $\ell > r - 1$ we can rewrite the previous formula to obtain

$$[z^n u^{r-1}]Q_s(z,u) = 2^{r-1}\binom{n-r}{n-\ell}\left(\frac{n-\ell}{\ell-r+1}-1\right) = 2^{r-1}\frac{s+r-1}{\ell-r+1}\binom{n-r}{n-\ell},$$

if $\ell = r - 1$, then we have $[z^n u^{r-1}]Q_s(z,u) = 2^{r-1}$ (independently of n), and if $\ell < r - 1$ we get $[z^n u^{r-1}]Q_s(z,u) = 0$.

To finish the proof, we have to normalize this number of paths with exactly r zeros and then sum up over all ℓ. So let us start with the normalization part. We set

$$\lambda_{n,r,s} = \mathbb{P}(Z_n = r \mid P_n \text{ starts in } (0,s)) = \frac{[z^n u^{r-1}]Q_s(z,u)}{[z^n]Q_s(z,1)}$$

for $n \equiv s \pmod 2$ and $\lambda_{n,r,s} = 0$ otherwise. The denominator $[z^n]Q_s(z,1)$ was already determined in Lemma A.4.

If $\ell > r - 1$, we have

$$\lambda_{n,r,s} = 2^{r-1}\frac{s+r-1}{\ell-r+1}\binom{n-r}{n-\ell}\bigg/\binom{n}{\ell}$$
$$= 2^{r-1}\frac{s+r-1}{\ell-r+1}\frac{(n-r)!\,\ell!\,(n-\ell)!}{(n-\ell)!\,(\ell-r)!\,n!}$$
$$= 2^{r-1}\frac{(n-r)!}{n!}\frac{\ell!\,(s+r-1)}{(\ell-r+1)!},$$

where the last line magically holds for $\ell = r - 1$ as well. In particular, we obtain

$$\lambda_{n,r,0} = 2^{r-1}(r-1)[n \geq 2r-2]\frac{(n/2)!}{n!}\frac{(n-r)!}{(n/2-r+1)!} = \frac{2^{r-1}(r-1)}{r}\binom{n/2}{r-1}\bigg/\binom{n}{r}.$$

We have arrived at the summation of the $\lambda_{n,r,s}$. The result follows as

$$\mathbb{P}(Z_n = r) = \sum_{s=-n}^{n} \mathbb{P}(Z_n = r \mid P_n \text{ starts in } (0, s)) \, \mathbb{P}(P_n \text{ starts in } (0, s))$$

$$= \frac{1}{n+1} \sum_{s=-n}^{n} \lambda_{n,r,s}$$

$$= \frac{2}{n+1} \sum_{\ell=0}^{\lceil n/2 \rceil - 1} \lambda_{n,r,n-2\ell} + \frac{1}{n+1} \lambda_{n,r,0},$$

and plugging in $\lambda_{n,r,s}$ gives the intermediate result

$$\mathbb{P}(Z_n = r) = 2^r \frac{(n-r)!}{(n+1)!} \sum_{\ell=r-1}^{\lceil n/2 \rceil - 1} \frac{\ell! \, (n - 2\ell + r - 1)}{(\ell - r + 1)!} \tag{E.1}$$

$$+ [n \text{ even}] \frac{2^{r-1}(r-1)}{(n+1)r} \binom{n/2}{r-1} \Big/ \binom{n}{r}$$

for $n \geq r$. At this point, we interrupt this proof to evaluate the sum over the ℓ and continue afterwards. $\qquad\square$

Lemma E.3. *We have*

$$\sum_{\ell=r-1}^{\lceil n/2 \rceil - 1} \frac{\ell! \, (n - 2\ell + r - 1)}{(\ell - r + 1)!} = r! \binom{\lceil n/2 \rceil}{r} \left(\frac{2\lceil \frac{n}{2} \rceil}{r(r+1)} + \frac{r-1}{r+1} + [n \text{ even}] \frac{1}{r} \right).$$

Of course, this lemma can be proved computationally by, for example, Sigma [Sch15b]. However, we give a direct proof here.

Proof. We obtain

$$\sum_{\ell=r-1}^{\lceil n/2 \rceil - 1} \frac{\ell! \, (n - 2\ell + r - 1)}{(\ell - r + 1)!} = (r-1)! \sum_{\ell=r-1}^{\lceil n/2 \rceil - 1} \binom{\ell}{r-1} (n - 2\ell + r - 1)$$

$$= (r-1)! \sum_{\ell=r-1}^{\lceil n/2 \rceil - 1} \binom{\ell}{r-1} (n + 1 + r - 2(\ell + 1))$$

$$= (n + 1 + r)(r-1)! \sum_{\ell=r-1}^{\lceil n/2 \rceil - 1} \binom{\ell}{r-1} - 2r! \sum_{\ell=r-1}^{\lceil n/2 \rceil - 1} \binom{\ell+1}{r}.$$

Using upper summation, cf. [GKP94, 5.10], yields

$$\sum_{\ell=r-1}^{\lceil n/2 \rceil - 1} \frac{\ell! \, (n - 2\ell + r - 1)}{(\ell - r + 1)!} = (n + 1 + r)(r-1)! \binom{\lceil n/2 \rceil}{r} - 2r! \binom{\lceil n/2 \rceil + 1}{r + 1}$$

$$= (r-1)! \binom{\lceil n/2 \rceil}{r} \left(n + 1 + r - 2 \frac{r}{r+1} (\lceil n/2 \rceil + 1) \right)$$

$$= (r-1)! \binom{\lceil n/2 \rceil}{r} \left(n + r - 1 - 2\lceil n/2 \rceil + \frac{2\lceil n/2 \rceil}{r + 1} + \frac{2}{r + 1} \right).$$

The lemma follows by replacing $n - 2\lceil n/2 \rceil$ with $[n \text{ even}] - 1$ and by collecting terms. $\qquad\square$

Proof of Theorem E.1 continued. To finish the proof, we simply plug in the result of Lemma E.3 into (E.1) and rewrite the factorials as binomial coefficients. $\quad\square$

As a next step, we want to prove Corollary E.2, which extracts the asymptotic behaviour of the distribution (Theorem E.1). To show that asymptotic formula, we will use the following auxiliary result.

Lemma E.4. *Let $0 < \varepsilon \leq \frac{1}{2}$. For integers c with $c = O\big(N^{1/2-\varepsilon}\big)$ we have*

$$c! \binom{N}{c} = N^c(1 + O\big(1/N^{2\varepsilon}\big)).$$

Proof. The inequality $N^c \geq c! \binom{N}{c}$ is trivial. We observe

$$c! \binom{N}{c} = N^c \cdot \prod_{0 \leq i < c} \left(1 - \frac{i}{N}\right) \geq N^c \cdot \left(1 - \sum_{0 \leq i < c} \frac{i}{N}\right) \geq N^c \left(1 - \frac{c^2}{2N}\right)$$
$$= N^c \left(1 + O\big(1/N^{2\varepsilon}\big)\right),$$

where the assumption on c has been used in the last step. $\quad\square$

Proof of Corollary E.2. By using Lemma E.4, the exact result of Theorem E.1 becomes

$$\mathbb{P}(Z_n = r) = \frac{2^r}{n} \frac{n^r}{2^r} \frac{1}{n^r} \left(\frac{n}{r(r+1)} + O(1)\right) \big(1 + O\big(1/n^{2\varepsilon}\big)\big)$$
$$+ [n \text{ even}] \frac{2^{r-1}(r-1)}{n} \frac{n^{r-1}}{2^{r-1}} \frac{1}{n^r} \big(1 + O\big(1/n^{2\varepsilon}\big)\big)$$
$$= \frac{1}{r(r+1)} \big(1 + O\big(1/n^{2\varepsilon}\big)\big)$$

as claimed. $\quad\square$

Appendix F. Appendix to Section 6: Going to Zero and Coming From Zero

For a path P_n of length n chosen according to the probabilistic model from Section 3 we define the random variables

$$Z_n^{\nearrow} = \text{number of up-to-zero situations on } P_n$$

and

$$Z_n^{\searrow} = \text{number of down-from-zero situations on } P_n.$$

Lemma F.1. *For a randomly (as described in Section 3) chosen path of length n,*

$$\mathbb{E}\big(Z_n^{\nearrow}\big) = \frac{1}{2}\left(\mathbb{E}(Z_n) - \frac{[n \text{ even}]}{n+1}\right) = \frac{1}{2} H_n^{\text{odd}}$$

and

$$\mathbb{E}\big(Z_n^{\searrow}\big) = \frac{1}{2}\left(\mathbb{E}(Z_n) - 1\right) = \frac{1}{2}\big(H_{n+1}^{\text{odd}} - 1\big).$$

Proof of Lemma F.1. In the proof of Theorem 4.1 linearity of the expectation gave that

$$\mathbb{E}(Z_n) = \sum_{m=0}^{n} \mathbb{P}(\text{path } P_n \text{ runs through } (n-m,0)). \tag{F.1}$$

Similarly we have

$$\mathbb{E}(Z_n^{\nearrow}) = \sum_{m=0}^{n-1} \mathbb{P}(\text{path } P_n \text{ runs through } (n-m-1,-1) \text{ and } (n-m,0)).$$

Note that the sum for $\mathbb{E}(Z_n)$ has a term for $m = n$ but the one for $\mathbb{E}(Z_n^{\nearrow})$ does not. Moreover, for $m < n$ the term

$$\mathbb{P}(\text{path } P_n \text{ runs through } (n-m-1,-1) \text{ and } (n-m,0))$$

is exactly half of $\mathbb{P}(\text{path } P_n \text{ runs through } (n-m,0))$, since for every path through $(n-m-1,-1)$ and $(n-m,0)$ there is one with the same probability through $(n-m-1,1)$ and $(n-m,0)$ by symmetry (or use the transition probabilities from (B.1)). So

$$\mathbb{E}(Z_n^{\nearrow}) = \frac{1}{2}\left(\mathbb{E}(Z_n) - \frac{[n \text{ even}]}{n+1}\right) = \frac{1}{2}\sum_{m=1}^{n}\frac{[m \text{ odd}]}{m} = \frac{1}{2}H_n^{\text{odd}}.$$

Similarly, comparing (F.1) with

$$\mathbb{E}(Z_n^{\searrow}) = \sum_{m=1}^{n} \mathbb{P}(\text{path } P_n \text{ runs through } (n-m,0) \text{ and } (n-m+1,-1))$$

we see that $\mathbb{E}(Z_n)$ has a term for $m = 0$ but $\mathbb{E}(Z_n^{\searrow})$ does not. For $m > 0$ exactly half of the paths that run through $(n-m,0)$ have a down-from-zero situation at $(n-m,0)$. So

$$\mathbb{E}(Z_n^{\searrow}) = \frac{1}{2}(\mathbb{E}(Z_n) - 1) = \frac{1}{2}(H_{n+1}^{\text{odd}} - 1).$$

$\square$

Applying Corollary D.2, we get the following asymptotic expansion.

Corollary F.2. *The expected number of up-to-zero situations is*

$$\mathbb{E}(Z_n^{\nearrow}) = \frac{1}{4}\log n + \frac{\gamma + \log 2}{4} + \frac{[n \text{ odd}]}{4n} + \frac{3[n \text{ even}] - 2}{24n^2} + O\left(\frac{1}{n^4}\right),$$

and the expected number of down-from-zero situations is

$$\mathbb{E}(Z_n^{\searrow}) = \frac{1}{4}\log n + \frac{\gamma + \log 2 - 2}{4}$$
$$+ \frac{[n \text{ even}] + 1}{4n} - \frac{9[n \text{ even}] + 2}{24n^2} + \frac{[n \text{ even}]}{2n^3} + O\left(\frac{1}{n^4}\right),$$

asymptotically as n tends to infinity.

APPENDIX G. APPENDIX TO SECTION 7: LATTICE PATHS OF VARIABLE LENGTH

The following proposition will be proven in this section.

Proposition G.1. *For a randomly chosen path (as described above), the expected number of up-to-zero situations is*

$$\mathbb{E}\big(X_n^{\nearrow}\big) = \frac{1}{2\binom{n}{2}} \sum_{n'=0}^{n-2} \sum_{m=1}^{n'} [m \text{ odd}] \frac{n'+1}{m} = \frac{1}{2} H_{n-2}^{\text{odd}} - \frac{1}{8} + \frac{(-1)^n}{8(n-[n \text{ even}])} \qquad (\text{G.1})$$

and the expected number of down-from-zero situations is

$$\mathbb{E}\big(X_n^{\searrow}\big) = \frac{1}{2\binom{n}{2}} \sum_{n'=0}^{n-2} \sum_{m=3}^{n'+1} [m \text{ odd}] \frac{n'+1}{m} = \mathbb{E}\big(X_n^{\nearrow}\big) - \frac{1}{2} + \frac{1}{2(n-[n \text{ even}])}. \qquad (\text{G.2})$$

The corresponding generating functions are

$$\sum_{n\geq 2} \mathbb{E}\big(X_n^{\nearrow}\big) z^n = \frac{\operatorname{artanh}(z)}{2(1-z)} - \frac{z^2}{8(1-z)} - \frac{3z+5}{8} \operatorname{artanh}(z) + \frac{1}{8}z,$$

$$\sum_{n\geq 2} \mathbb{E}\big(X_n^{\searrow}\big) z^n = \frac{\operatorname{artanh}(z)}{2(1-z)} - \frac{5z^2}{8(1-z)} + \frac{z-1}{8} \operatorname{artanh}(z) - \frac{3}{8}z.$$

Asymptotically, we have

$$\mathbb{E}\big(X_n^{\nearrow}\big) = \frac{1}{4}\log n + \frac{2\gamma + 2\log 2 - 1}{8} - \frac{3}{8n} - \frac{3[n \text{ even}]+1}{12n^2} - \frac{3[n \text{ even}]}{8n^3} + O\Big(\frac{1}{n^4}\Big),$$

$$\mathbb{E}\big(X_n^{\searrow}\big) = \frac{1}{4}\log n + \frac{2\gamma + 2\log 2 - 5}{8} + \frac{1}{8n} + \frac{3[n \text{ even}]-1}{12n^2} + \frac{[n \text{ even}]}{8n^3} + O\Big(\frac{1}{n^4}\Big),$$

as $n \to \infty$.

To prove this proposition, we first compute the distribution of N'. The following lemma is a simple consequence of the definition of N'.

Lemma G.2. *If $n' \in \{0, 1, \ldots, n-2\}$, then*

$$\mathbb{P}(N' = n') = \frac{n'+1}{\binom{n}{2}},$$

and otherwise $\mathbb{P}(N' = n') = 0$.

Proof. Fix $n' \in \{0, 1, \ldots, n-2\}$, and set $\Delta = n - 1 - n'$. There are $n - \Delta = n' + 1$ choices of the random variable P such that $Q = P + \Delta$ is still at most n. Thus, the lemma follows. $\qquad\square$

Proof of Proposition G.1. For any random variable Y_n depending on N', the law of total expectation yields

$$\mathbb{E}(Y_n) = \sum_{n'\in\mathbb{N}_0} \mathbb{P}(N' = n') \mathbb{E}(Y_n \mid N' = n'). \qquad (\text{G.3})$$

From (G.3) and the Lemmata F.1 and G.2, we immediately get that $\mathbb{E}\big(X_n^{\nearrow}\big)$ and $\mathbb{E}\big(X_n^{\searrow}\big)$ are equal to the respective double sum given in (G.1) and (G.2).

To obtain the difference, one observes that

$$\mathbb{E}(Z_{n'}^{\nearrow}) - \mathbb{E}(Z_{n'}^{\searrow}) = \frac{1}{2}\Big(1 - \frac{[n' \text{ even}]}{n'+1}\Big),$$

from which we obtain

$$\mathbb{E}\big(X_n^{\nearrow}\big) - \mathbb{E}\big(X_n^{\searrow}\big) = \frac{1}{2\binom{n}{2}} \sum_{n'=0}^{n-2} \left(1 - \frac{[n' \text{ even}]}{n'+1}\right) \cdot (n'+1).$$

Simplifying this to $\frac{1}{2} - \frac{1}{2(n-[n \text{ even}])}$ is straightforward.

It remains to verify the second expression for $\mathbb{E}\big(X_n^{\nearrow}\big)$ given in (G.1). We have

$$\begin{aligned}
\mathbb{E}\big(X_n^{\nearrow}\big) &= \frac{1}{2\binom{n}{2}} \sum_{n'=0}^{n-2} \sum_{m=1}^{n'} [m \text{ odd}] \frac{n'+1}{m} \\
&= \frac{1}{2\binom{n}{2}} \sum_{m=1}^{n-2} \frac{[m \text{ odd}]}{m} \sum_{n'=m}^{n-2} (n'+1) \\
&= \frac{1}{2\binom{n}{2}} \sum_{m=1}^{n-2} [m \text{ odd}] \frac{(n-1)n - m(m+1)}{2m}.
\end{aligned}$$

(We just summed the arithmetic series.) So

$$\begin{aligned}
\mathbb{E}\big(X_n^{\nearrow}\big) &= \frac{1}{2\binom{n}{2}} \sum_{m=1}^{n-2} [m \text{ odd}] \left(\frac{\binom{n}{2}}{m} - \frac{m+1}{2}\right) \\
&= \frac{1}{2} H_{n-2}^{\text{odd}} - \frac{1}{4\binom{n}{2}} \sum_{m=2}^{n-1} [m \text{ even}] \, m \\
&= \frac{1}{2} H_{n-2}^{\text{odd}} - \frac{\lfloor (n-1)/2 \rfloor \lfloor (n+1)/2 \rfloor}{4\binom{n}{2}}.
\end{aligned}$$

This leads to (G.1).

We now compute the generating functions. We first note that for $k \in \mathbb{Z}$, we have

$$\sum_{n>k} \frac{[n-k \text{ odd}]}{n-k} z^n = z^k \operatorname{artanh}(z). \tag{G.4}$$

Taking the summatory function amounts to division by $(1-z)$ and a shift in the argument corresponds to multiplication by z, so we obtain

$$\sum_{n\geq 2} H_{n-2}^{\text{odd}} z^n = z^2 \frac{\operatorname{artanh}(z)}{1-z}.$$

The remaining summands lead to geometric series and another instance of (G.4), so we obtain

$$\sum_{n\geq 2} \mathbb{E}\big(X_n^{\nearrow}\big) z^n = \frac{z^2 \operatorname{artanh}(z)}{2(1-z)} - \frac{z^2}{8(1-z)} + \frac{z-1}{8} \operatorname{artanh}(z) + \frac{z}{8}$$

where the final summand $z/8$ has to be added as we sum over $n \geq 2$ only. Minor simplifications lead to the desired formula. The generating function for $\mathbb{E}\big(X_n^{\searrow}\big)$ follows by adding

$$-\frac{z^2}{2(1-z)} + \frac{1+z}{2} \operatorname{artanh}(z) - \frac{z}{2},$$

corresponding to (G.2).

The asymptotic expressions follow from the explicit formulæ via Lemma D.1. $\qquad\square$

APPENDIX H. APPENDIX TO SECTION 8

We justify (8.1). Assume the input is a random permutation of $\{1, \ldots, n\}$. The expected cost $\mathbb{E}(C_n)$ is the sum of the expected partitioning cost $\mathbb{E}(P_n)$ and the sum of the costs of the recursive calls for the small, the medium, and the large elements. A recursive call for a group of size k has expected cost $\mathbb{E}(C_k)$, for $0 \leq k \leq n - 2$. The probability that the number of small elements is exactly k is $(n - 1 - k)/\binom{n}{2}$, since there are $\binom{n}{2}$ possible pivot pairs $\{p, q\}$ in total and $n - 1 - k$ pairs $\{p, q\}$ with $p = k + 1$. So the expected contribution to $\mathbb{E}(C_n)$ from the set of small elements is $\sum_{1 \leq k \leq n-2}((n - 1 - k)/\binom{n}{2})\,\mathbb{E}(C_k)$. Since there are $n - 1 - k$ pivot pairs with $k = q - p - 1$ and $n - 1 - k$ pivot pairs with $q = n - k$, the contributions from the recursive calls for the set of medium and the set of large elements are the same. Adding the three contributions we obtain (8.1).

APPENDIX I. APPENDIX TO SECTION 8: SOLVING THE DUAL-PIVOT QUICKSORT RECURRENCE

We recall how to solve recurrence (8.1) using generating functions. We follow Wild [Wil13, § 4.2.2] who in turn follows Hennequin [Hen91]. The following lemma is contained in slightly different notation in [Wil13]; nevertheless, we include the proof here for the sake of self-containedness. This also allows us to make the integration bounds explicit.

Lemma I.1. *With C_n and P_n as above, $C(z) = \sum_{n \geq 0} \mathbb{E}(C_n)\,z^n$ and $P(z) = \sum_{n \geq 0} \mathbb{E}(P_n)\,z^n$, we have*

$$C(z) = (1 - z)^3 \int_0^z (1 - t)^{-6} \int_0^t (1 - s)^3 P''(s)\,ds\,dt.$$

Proof. Multiplying (8.1) by $n(n - 1)z^{n-2}$ and summing over all $n \geq 2$ yields

$$\sum_{n \geq 2} n(n - 1)\,\mathbb{E}(C_n)\,z^{n-2}$$

$$= \sum_{n \geq 2} n(n - 1)\,\mathbb{E}(P_n)\,z^{n-2} + 6 \sum_{n \geq 1} \sum_{k=0}^{n-1} (n - 1 - k)z^{n-k-2}\,\mathbb{E}(C_k)\,z^k.$$

Note that the range of the summations has been extended without any consequences because of $\mathbb{E}(C_0) = 0$. We replace $n - 1$ by n in the double sum and write it as a product of two generating functions:

$$\sum_{n \geq 1} \sum_{k=0}^{n-1} (n - 1 - k)z^{n-k-2}\,\mathbb{E}(C_k)\,z^k = \sum_{n \geq 0} \sum_{k=0}^{n} (n - k)z^{n-k-1}\,\mathbb{E}(C_k)\,z^k$$

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

Thus we obtained

$$C''(z) = P''(z) + \frac{6}{(1 - z)^2}C(z)$$

or, equivalently,

$$(1 - z)^2 C''(z) - 6C(z) = (1 - z)^2 P''(z).$$

Setting $(\theta f)(z) = (1 - z)f'(z)$ for a function f, this can be rewritten as

$$((\theta^2 + \theta - 6)C)(z) = (1 - z)^2 P''(z).$$

Factoring $\theta^2 + \theta - 6$ as $(\theta - 2)(\theta + 3)$ and setting $D = (\theta + 3)C$, we first have to solve

$$((\theta - 2)D)(z) = (1 - z)^2 P''(z),$$

i. e.,

$$(1 - z)D'(z) - 2D(z) = (1 - z)^2 P''(z).$$

Multiplication by $(1 - z)$ yields

$$\left((1 - z)^2 D(z)\right)' = (1 - z)^3 P''(z).$$

Integration and the fact that $D(0) = C'(0) + 3C(0) = \mathbb{E}(C_1 + 3C_0) = 0$ yields

$$D(z) = \frac{1}{(1 - z)^2} \int_0^z (1 - s)^3 P''(s)\, ds.$$

We still have to solve

$$(1 - z)C'(z) + 3C(z) = D(z).$$

We multiply by $(1 - z)^{-4}$ and obtain

$$\left((1 - z)^{-3}C(z)\right)' = (1 - z)^{-4}D(z).$$

As $C(0) = 0$, we obtain

$$C(z) = (1 - z)^3 \int_0^z (1 - t)^{-4}D(t)\, dt.$$

$\square$

Appendix J. Appendix to Section 9: Partitioning Algorithms and Their Cost

Proof. We start with Part (a). A different proof of a related statement was given in [AD15]. Since the pivots are p and q with $p < q$, there are $s = p - 1$ small, $\ell = n - q$ large, and $p - q - 1$ medium elements. Omit all medium elements (which require two comparisons) to obtain a reduced sequence $(b_1, \ldots, b_{n'})$ of elements to be classified, with $n' = s + \ell$. Note that the distribution of n' is exactly that of the random variable N' in Section 7. Further, if n' is given, s is uniform in $\{0, \ldots, n'\}$ and $s - \ell$ is uniform in $\{x \in \mathbb{Z} \mid |x| \leq n', x \equiv n' \pmod 2\}$. By the assumption that the input is in random order it is irrelevant in which order the elements are classified; so we may assume the order is $b_1, \ldots, b_{n'}$. For $0 \leq i \leq n'$ let s'_i and ℓ'_i denote the number of small respectively large elements in $(b_{i+1}, \ldots, b_{n'})$, and let $d_i = s'_i - \ell'_i$. Then the sequence

$$((0, d_0), (1, d_1), \ldots, (n', 0))$$

is a lattice path of length n' as defined in Section 3. (Figure J.1 depicts such a path.) When s and ℓ are given, the path is determined by the random order of the small and large elements in $(b_1, \ldots, b_{n'})$, and hence all these paths have the same probability $1/\binom{s+\ell}{s}$. We now note that the path contains the full information about the additional comparisons carried out by strategy "Clairvoyant": For $1 \leq i \leq n'$, an edge from $(i - 1, d_{i-1})$ to (i, d_i) will correspond to an additional comparison occurring when classifying b_i if and only if either

- $d_{i-1} = s'_{i-1} - \ell'_{i-1} \geq 0$ (so that b_i is compared with p first) and $d_i > d_{i-1}$ (which means that b_i is a large element), or
- $d_{i-1} = s'_{i-1} - \ell'_{i-1} < 0$ (so that b_i is compared with q first) and $d_i < d_{i-1}$ (which means that b_i is a small element).

Thus, if we let

$$\#\mathrm{up}_{\geq 0} = \#(\text{steps going up, starting on or above the horizontal axis}),$$

$$\#\mathrm{down}_{<0} = \#(\text{steps going down, starting strictly below the horizontal axis}),$$

then the additional cost equals $\#\mathrm{up}_{\geq 0} + \#\mathrm{down}_{<0}$.

Now assume $s \geq \ell$ (see Figure J.1), and let

$$\#\mathrm{down}_{>0} = \#(\text{steps going down, starting strictly above the horizontal axis}).$$

Since the path ends at $(n', 0)$ and up and down steps below the horizontal axis cancel, we have $s - \ell = d_0 = \#\mathrm{down}_{>0} - \#\mathrm{up}_{\geq 0}$, hence $\#\mathrm{up}_{\geq 0} = \#\mathrm{down}_{>0} - (s - \ell)$. Thus, the additional cost is

$$\#\mathrm{down}_{>0} + \#\mathrm{down}_{<0} - (s - \ell).$$

Let $z_{n'}^{\searrow}$ be the number of steps that go down-from-zero (cf. Appendix F). The total number of steps that go down is $s = \#\mathrm{down}_{>0} + \#\mathrm{down}_{<0} + z_{n'}^{\searrow}$, so that the additional cost turns out to be

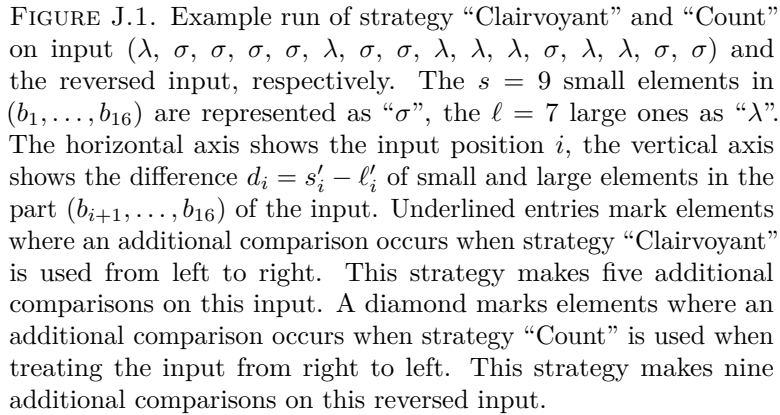$$s - z_{n'}^{\searrow} - (s - \ell) = \ell - z_{n'}^{\searrow}.$$

In a similar way one sees that the assumption $\ell > s$ leads to $s - z_{n'}^{\searrow}$ additional comparisons. Combining both cases, the number of additional comparisons is $\min(s, \ell) - z_{n'}^{\searrow}$. Recalling the definition of $X_n^{\searrow}$ from Section 7, averaging over all possible pivots $p < q$, and using $s = p - 1$ and $\ell = n - q$, we obtain

$$\mathbb{E}(A_n^{\mathrm{cv}}) = \frac{1}{\binom{n}{2}}\left(\sum_{1 \leq p < q \leq n} \min(p - 1, n - q)\right) - \mathbb{E}(X_n^{\searrow}).$$

The first summand evaluates to $\frac{n}{6} - \frac{7}{12} + \frac{1}{4(n - [n \text{ even}])}$.

We continue with Part (b). Assume that pivots p and q and a reduced input $(b_1, \ldots, b_{n'})$ is produced as in Part (a). For analyzing the classification strategy "Count", we wish to utilize our knowledge about "Clairvoyant". To this end, we use a reflection trick and assume the input is treated in the reverse order $b_{n'}, \ldots, b_1$. (This is harmless since the order in which the elements are treated is irrelevant anyway.) As in the specification of "Count", let s_i and ℓ_i denote the number of small and large elements that have been seen after the ith classification step. Here this is the number of small and large elements in $\{b_{n'-i+1}, \ldots, b_{n'}\}$; hence we have $s_i = s'_{n'-i}$ and $\ell_i = \ell'_{n'-i}$, and $s_i - \ell_i = d_{n'-i}$. Keeping track of $s_i - \ell_i$ for $i = 0, \ldots, n'$ gives rise to a lattice path, which we imagine as running in the opposite direction from our standard paths: It starts at $(n', 0)$ and ends at $(0, s - \ell)$. The point reached after step i is $(n' - i, s_i - \ell_i) = (n' - i, d_{n'-i})$. (An illustration is obtained by traversing the path in Figure J.1 from right to left.) Thus, the path induced by applying "Clairvoyant" to $(b_1, \ldots, b_{n'})$ and the path induced by applying "Count" to $(b_{n'}, \ldots, b_1)$ are the same, they are just traversed in opposite directions. The probability with which a path appears is the same in both situations.

Now we wish to determine the additional cost of the run of "Count" on $(b_{n'}, \ldots, b_1)$. For $1 \leq i \leq n'$, reading from right to left, an edge from $(n' - i + 1, s_{n'-i+1} - \ell_{n'-i+1})$

FIGURE J.1. Example run of strategy "Clairvoyant" and "Count" on input $(\lambda, \sigma, \sigma, \sigma, \sigma, \lambda, \sigma, \sigma, \lambda, \lambda, \lambda, \sigma, \lambda, \lambda, \sigma, \sigma)$ and the reversed input, respectively. The $s = 9$ small elements in $(b_1, \ldots, b_{16})$ are represented as "σ", the $\ell = 7$ large ones as "λ". The horizontal axis shows the input position i, the vertical axis shows the difference $d_i = s'_i - \ell'_i$ of small and large elements in the part $(b_{i+1}, \ldots, b_{16})$ of the input. Underlined entries mark elements where an additional comparison occurs when strategy "Clairvoyant" is used from left to right. This strategy makes five additional comparisons on this input. A diamond marks elements where an additional comparison occurs when strategy "Count" is used when treating the input from right to left. This strategy makes nine additional comparisons on this reversed input.

to $(n' - i, s_{n'-i} - \ell_{n'-i})$ will correspond to an additional comparison needed when classifying $b_{n'-i+1}$ if and only if either

- $0 \leq s_{i-1} - \ell_{i-1} = d_{n'-i+1}$ (so that $b_{n'-i+1}$ is compared with p first) and $d_{n'-i} < d_{n'-i+1}$ (which means that $b_{n'-i+1}$ is a large element), or
- $0 > s_{i-1} - \ell_{i-1} = d_{n'-i+1}$ (so that $b_{n'-i+1}$ is compared with q first) and $d_{n'-i} > d_{n'-i+1}$ (which means that $b_{n'-i+1}$ is a small element).

We now interpret these relations in the standard direction from left to right. Additional comparisons correspond to steps in the path

- from $(n' - i, d_{n'-i})$ to $(n' - i + 1, d_{n'-i+1})$ with $d_{n'-i+1} \geq 0$ and $d_{n'-i} < d_{n'-i+1}$ and
- from $(n' - i, d_{n'-i})$ to $(n' - i + 1, d_{n'-i+1})$ with $d_{n'-i+1} < 0$ and $d_{n'-i} > d_{n'-i+1}$.

If we let

$$\#\text{up}_{\geq -1} = \#(\text{steps going up, ending on or above the horizontal axis}),$$

$$\#\text{down}_{\leq 0} = \#(\text{steps going down, starting on or below the horizontal axis}),$$

then the additional cost equals $\#\text{up}_{\geq -1} + \#\text{down}_{<0}$, see Figure J.1 for an example.

Comparing with what we obtained in (a) for "Clairvoyant" we see that the additional cost of "Count" is the additional cost of "Clairvoyant" plus the number of steps from $(n' - i, -1)$ to $(n' - i + 1, 0)$ (in the notation of Sections 6 and 7 these are "up-to-zero steps", their number being denoted by $z_{n'}^{\nearrow}$) and the number of steps from $(n' - i, 0)$ to $(n' - i + 1, -1)$ (these are "down-from-zero steps", their number being denoted by $z_{n'}^{\searrow}$). Since the additional cost for "Clairvoyant" was $\min(p - 1, n - q) - z_{n'}^{\searrow}$, for "Count" we get a cost of

$$\min(p - 1, n - q) + z_{n'}^{\nearrow}.$$

Averaging over all $p < q$ as in (a) finally yields

$$\mathbb{E}(A_n^{\mathrm{ct}}) = \frac{1}{\binom{n}{2}} \left(\sum_{1 \le p < q \le n} \min(p - 1, n - q) \right) + \mathbb{E}(X_n^{\nearrow}),$$

and evaluating the sum as in Part (a) finishes the proof. $\square$

Lemma J.1. *Let $n \ge 2$.*

(a) The expected number of comparisons of strategy "Clairvoyant" is

$$\mathbb{E}(P_n^{\mathrm{cv}}) = \frac{3}{2}n - \frac{9}{4} + \frac{1}{4(n - [n \text{ even}])} - \mathbb{E}(X_n^{\searrow}).$$

(b) The expected number of comparisons of strategy "Count" is

$$\mathbb{E}(P_n^{\mathrm{ct}}) = \frac{3}{2}n - \frac{9}{4} + \frac{1}{4(n - [n \text{ even}])} + \mathbb{E}(X_n^{\nearrow}). \tag{J.1}$$

Proof. The expressions for $\mathbb{E}(P_n^{\mathrm{cv}})$ and $\mathbb{E}(P_n^{\mathrm{cv}})$ are obtained by adding the expected number of necessary comparisons $\frac{4}{3}(n - 2) + 1$ to the cost terms in Lemma 9.1. $\square$

Appendix K. Appendix to Section 10: Main Results and their Asymptotic Aspects

Proof of Theorems 10.1 and 10.3. The partitioning cost of strategy "Clairvoyant" is stated in Lemma J.1. The corresponding generating functions can be obtained by using Proposition G.1 and (G.4):

$$P^{\mathrm{cv}}(z) = \sum_{n \ge 2} \mathbb{E}(P_n^{\mathrm{cv}}) z^n = \frac{3}{2(1 - z)^2} - \frac{\operatorname{artanh}(z)}{2(1 - z)}$$
$$- \frac{25z^2}{8(1 - z)} + \frac{3 + z}{8} \operatorname{artanh}(z) - \frac{3}{2} - \frac{23z}{8}.$$

We calculate the comparison cost from the partitioning cost by means of Lemma I.1 and obtain

$$C^{\mathrm{cv}}(z) = -2\frac{\log(1 - z)}{(1 - z)^2} - \frac{2\operatorname{artanh}(z)}{5(1 - z)^2} - \frac{44}{25(1 - z)^2} + \frac{\operatorname{artanh}(z)}{4(1 - z)} + \frac{279}{160(1 - z)}$$
$$- \frac{(1 - z)^3}{320} \operatorname{artanh}(z) - \frac{2}{75}z^3 + \frac{123}{1600}z^2 - \frac{113}{1600}z + \frac{13}{800}.$$

Taking into account that $\operatorname{artanh}(z) = (\log(1+z) - \log(1-z))/2$,

$$\sum_{m \geq 1} H_m^{\text{alt}} z^m = -\frac{\log(1+z)}{(1-z)},$$

$$\sum_{m \geq 1} H_m z^m = -\frac{\log(1-z)}{(1-z)},$$

$$\sum_{m \geq 1} m H_m^{\text{alt}} z^m = z\left(-\frac{\log(1+z)}{(1-z)}\right)' = -\frac{\log(1+z)}{(1-z)^2} + \frac{\log(1+z)}{1-z} + \frac{1}{2(1+z)} - \frac{1}{2(1-z)},$$

$$\sum_{m \geq 1} m H_m z^m = z\left(-\frac{\log(1-z)}{(1-z)}\right)' = -\frac{\log(1-z)}{(1-z)^2} + \frac{1}{(1-z)^2} + \frac{\log(1-z)}{1-z} - \frac{1}{1-z},$$

as well as (G.4), we obtain the result.

The proof concerning strategy "Count" is analogous to the proof of Theorem 10.1 above. The corresponding generating functions are

$$P^{\text{ct}}(z) = \sum_{n \geq 2} \mathbb{E}\big(P_n^{\text{ct}}\big) z^n = \frac{3}{2(1-z)^2} + \frac{\operatorname{artanh}(z)}{2(1-z)}$$

$$- \frac{31z^2}{8(1-z)} - \frac{3+z}{8} \operatorname{artanh}(z) - \frac{3}{2} - \frac{25z}{8}$$

and

$$C^{\text{ct}}(z) = -\frac{8\log(1-z)}{5(1-z)^2} + \frac{2\operatorname{artanh}(z)}{5(1-z)^2} - \frac{44}{25(1-z)^2} - \frac{\operatorname{artanh}(z)}{4(1-z)} + \frac{281}{160(1-z)}$$

$$+ \frac{(1-z)^3}{320} \operatorname{artanh}(z) + \frac{1}{150} z^3 - \frac{27}{1600} z^2 + \frac{17}{1600} z + \frac{3}{800}.$$

in this case. □

Proof of Corollaries 10.2 and 10.4. Insert the expansions of Lemma D.1 into the explicit representations of Theorems 10.1 and 10.3. □

Appendix L. Empirical Validation

We implemented strategies "Clairvoyant" and "Count" as dual-pivot quicksort algorithms in a straight-forward way in C++. Pseudocode of the algorithms is presented in Appendix M.

For small input sizes of length $n \in \{2, \ldots, 12\}$ we enumerated all permutations of $\{1, \ldots, n\}$ and verified that the average number of comparisons (computed over all permutations) obtained from the experimental measurements equals the results from Theorem 10.1 and Theorem 10.3, respectively.

For larger inputs, we sorted random permutations of $\{1, \ldots, n\}$ for $n = 2^k, k \in \{11, \ldots, 28\}$, and counted the comparisons needed to sort the input. For each input size, we sorted 400 different inputs. Figure L.1 shows the measurements we got. From these measurements we conclude that the average comparison counts for sorting over a small number of inputs match the exact average counts from Theorem 10.1 and Theorem 10.3, resp., very well.
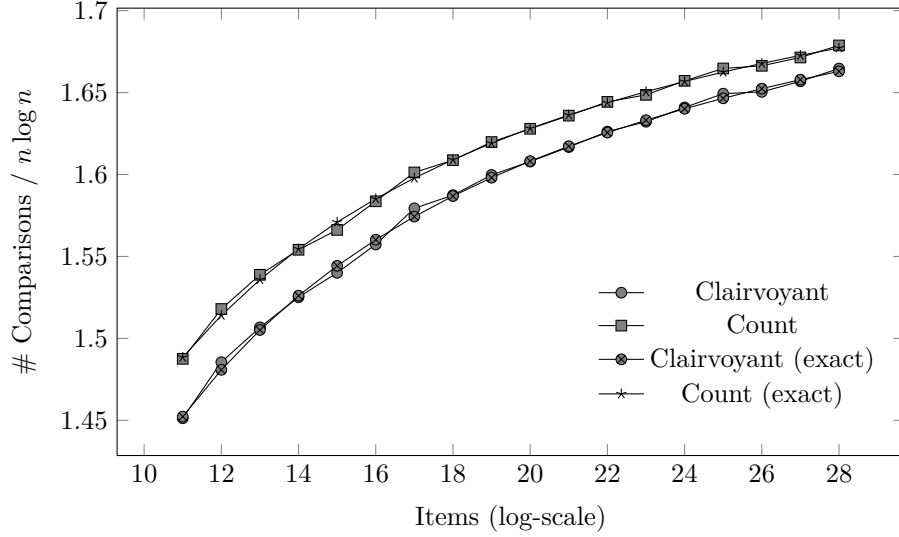
FIGURE L.1. Average comparison count (scaled by $n \log n$) needed to sort a random input of up to 2^{28} integers. We show the measurements we got for the comparison-optimal strategy "Clairvoyant" and its algorithmic variant "Count" together with the exact average comparison counts from Theorem 10.1 and Theorem 10.3. All data points from experiments are the average over 400 trials.

Appendix M. Pseudocode of Dual-Pivot Quicksort Algorithms

In this supplementary section, we give the full pseudocode for the strategies "Clairvoyant" (Algorithm 1) and "Count" (Algorithm 2) turned into dual-pivot quicksort algorithms.

Algorithm 1 Dual-Pivot Quicksort Algorithm "Clairvoyant"

---

procedure *Clairvoyant*(*A*, *left*, *right*)

1 **if** *right* ≤ *left* **then**
2   **return**
3 **if** $A[right] < A[left]$ **then**
4   swap $A[left]$ and $A[right]$
5 p ← $A[left]$
6 q ← $A[right]$
7 i ← *left* + 1; k ← *right* − 1; j ← i
8 d ← #(small elements) − #(large elements)   // d is given by an oracle.
9 **while** j ≤ k **do**
10   **if** d ≥ 0 **then**
11   **if** $A[j] < p$ **then**
12   swap $A[i]$ and $A[j]$
13   i ← i + 1; j ← j + 1; d ← d − 1
14   **else**
15   **if** $A[j] < q$ **then**
16   j ← j + 1
17   **else**
18   swap $A[j]$ and $A[k]$
19   k ← k − 1; d ← d + 1
20   **else**
21   **if** $A[k] > q$ **then**
22   k ← k − 1; d ← d + 1
23   **else**
24   **if** $A[k] < p$ **then**
25   // Perform a cyclic rotation to the left, i. e.,
26   // tmp ← $A[k]$; $A[k] ← A[j]$; $A[j] ← A[i]$; $A[i] ←$ tmp
27   *rotate3*($A[k], A[j], A[i]$)
28   i ← i + 1; d ← d − 1
29   **else**
30   swap $A[j]$ and $A[k]$
31   j ← j + 1
32 swap $A[left]$ and $A[i − 1]$
33 swap $A[right]$ and $A[k + 1]$
34 *Clairvoyant*(*A*, *left*, i − 2)
35 *Clairvoyant*(*A*, i, k)
36 *Clairvoyant*(*A*, k + 2, *right*)

Algorithm 2 Dual-Pivot Quicksort Algorithm "Count"

procedure *Count*(*A*, *left*, *right*)

1 **if** *right* ≤ *left* **then**
2   **return**
3 **if** $A[right] < A[left]$ **then**
4   swap $A[left]$ and $A[right]$
5 p ← $A[left]$
6 q ← $A[right]$
7 i ← *left* + 1; k ← *right* − 1; j ← i
8 d ← 0   // d holds the difference of the number of small and large elements.
9 **while** j ≤ k **do**
10   **if** d ≥ 0 **then**
11   **if** $A[j] < p$ **then**
12   swap $A[i]$ and $A[j]$
13   i ← i + 1; j ← j + 1; d ← d + 1
14   **else**
15   **if** $A[j] < q$ **then**
16   j ← j + 1
17   **else**
18   swap $A[j]$ and $A[k]$
19   k ← k − 1; d ← d − 1
20   **else**
21   **if** $A[k] > q$ **then**
22   k ← k − 1; d ← d − 1
23   **else**
24   **if** $A[k] < p$ **then**
25   *rotate3*($A[k], A[j], A[i]$)
26   i ← i + 1; d ← d + 1
27   **else**
28   swap $A[j]$ and $A[k]$
29   j ← j + 1
30 swap $A[left]$ and $A[i − 1]$
31 swap $A[right]$ and $A[k + 1]$
32 *Count*(*A*, *left*, i − 2)
33 *Count*(*A*, i, k)
34 *Count*(*A*, k + 2, *right*)

Martin Aumüller, IT University of Copenhagen, Rued Langgaards Vej 7, 2300 Copenhagen, Denmark
E-mail address: `maau@itu.dk`

Martin Dietzfelbinger, Institut für Theoretische Informatik, Fakultät für Informatik und Automatisierung, Technische Universität Ilmenau, Am Helmholtzplatz 5, 98694 Ilmenau, Germany
E-mail address: `martin.dietzfelbinger@tu-ilmenau.de`

Clemens Heuberger, Institut für Mathematik, Alpen-Adria-Universität Klagenfurt, Universitätsstrasse 65–67, 9020 Klagenfurt am Wörthersee, Austria
E-mail address: `clemens.heuberger@aau.at`

Daniel Krenn, Institut für Mathematik, Alpen-Adria-Universität Klagenfurt, Universitätsstrasse 65–67, 9020 Klagenfurt am Wörthersee, Austria
E-mail address: `math@danielkrenn.at` *or* `daniel.krenn@aau.at`

Helmut Prodinger, Department of Mathematical Sciences, Stellenbosch University, 7602 Stellenbosch, South Africa
E-mail address: `hproding@sun.ac.za`