# PROBABILISTIC ANALYSIS OF OPTIMAL MULTI-PIVOT QUICKSORT

CECILIA HOLMGREN, JASPER ISCHEBECK, DANIEL KRENN, FLORIAN LESNY, AND RALPH NEININGER

Abstract. We consider a multi-pivot QuickSort algorithm using $K \in \mathbb{N}$ pivot elements to partition a nonsorted list into $K+1$ sublists in order to proceed recursively on these sublists. For the partitioning stage, various strategies are in use. We focus on the strategy that minimizes the expected number of key comparisons in the standard random model, where the list is given as a uniformly permuted list of distinct elements.

We derive asymptotic expansions for the expectation and variance of the number of key comparisons as well as a limit law for all $K \in \mathbb{N}$, where the convergence holds for all (exponential) moments. For $K \leq 4$ we also bound the rate of convergence within the Wasserstein and Kolmogorov–Smirnov distance.

Our analysis of the expectation is based on classical results for random $m$-ary search trees. For the remaining results, combinatorial considerations are used to make the contraction method applicable.

## 1. Introduction and results

The QuickSort algorithm, introduced by Hoare [12], is a divide-and-conquer sorting algorithm. It solves the problem of sorting a list of $n$ distinct elements from a totally ordered set by selecting an element as *pivot* and partitioning the other elements into two sublists according to whether they are smaller or larger than the pivot and acting recursively on these sublists. It is commonly used and variants of it are the standard sorting routines in many programming frameworks.

Multi-pivot QuickSort is a variant of QuickSort that uses $K \in \mathbb{N}$ pivots to partition into $K+1$ sublists. Java 7 switched from using the classical QuickSort to a dual-pivot QuickSort introduced by Yaroslavskiy [27], which is still used in Java SE as of Java 23 [13]. The advantages of Multi-pivot QuickSort algorithms include that they reduce the list sizes faster and need fewer passes over the list. In addition, they need fewer key comparisons. The Yaroslavskiy algorithm is analyzed in [25, 26] under the probabilistic model of uniformity of the order of the given items. In the present paper, we focus on the number $X_n$ of key comparisons as a cost measure to analyze the complexity of multi-pivot QuickSort and, to be explicit, assume that the items are numbers from $[0,1]$ being independent and identically, uniformly distributed.

The core of the QuickSort algorithm is formed by the partition step, where the list is divided into $K+1$ sublists. Aumüller, Dietzfelbinger and Klaue [1, 3] developed a partitioning strategy that is optimal (see [2]) with respect to the expected number of comparisons. In the present paper, we analyze multi-pivot QuickSort using this partition strategy considering the number $X_n$ of key comparisons when acting on $n$ items to be sorted.

To be definite , for $K \in \mathbb{N}$, a $K$-pivot QuickSort algorithm sorts a list by

(1) picking $K$ elements as *pivot elements* (according to some rule) and sorting these elements (by some elementary sorting algorithm) as $p_1 < \cdots < p_K$, then

(2) partitioning the list into $K+1$ sublists $S_0, \dots, S_K$, each containing only the elements below the smallest pivot element $p_1$, between two consecutive pivot elements or above the largest pivot element $p_K$, and finally
(3) proceeding recursively with the $K+1$ new lists.

If the list has $K$ or fewer elements, it uses directly some elementary sorting algorithm. Note that the unspecified components above (how to choose and sort the pivot elements and how the lists of at most $K$ elements are sorted) do not enter our analysis; our results are valid for any reasonable choice. However, there is one important aspect that needs further attention, namely the second step dealing with the partitioning of a list. The order in which an element is compared to the different pivot elements determines how many comparisons are needed. We call these orders *partition strategies*. For example, in the case of $K = 2$ pivot elements, we might first compare an element with the smaller pivot, $p_1$. If the element is smaller than $p_1$, we can skip the comparison with the larger pivot. *Optimal $K$-pivot QuickSort* always chooses the partition strategy that minimizes the expected number of key comparisons in the next step, based on how many elements it has already sorted into the sublists $S_0, \dots, S_K$. We will detail later, in Section 1.2, how exactly optimal $K$-pivot QuickSort determines the optimal partition strategy.

For the pivot elements $0 \le p_1 < \dots < p_k \le 1$ we define the *spacings*

$$D_0 := p_1, D_1 := p_2 - p_1, \dots, D_K := 1 - p_K.$$

Conditional on the pivots, $D_i$ is the probability to be sorted into the sublist $S_i$. Given $D = (D_0, \dots, D_K)$ and a partition strategy $t$, we define $l_t(D)$ as the expected number of comparisons of an element with the pivots and $l_{\mathrm{opt}}(D)$ as the minimum of $l_t(D)$ over all strategies. This optimal strategy is denoted $t_{\mathrm{opt}}$. As the relative sizes of the sublists converge towards $D$, this is also the strategy optimal $K$-pivot QuickSort will utilize for large $n$. We define

$$\gamma_K := \mathbb{E}\left[l_{\mathrm{opt}}(D)\right].$$

Furthermore, log denotes the natural (base $e$) logarithm.

1.1. **Results.** We now present our results on the asymptotic behavior of the number of key comparisons.

**Theorem 1.** *Let $K \in \mathbb{N}$. For the number $X_n$ of key comparisons we have*

$$\mathbb{E}\left[X_n\right] = \alpha_K n \log n + \beta_K n + o(n) \quad (n \to \infty), \tag{1}$$

*with $\alpha_K = \gamma_K/(H_{K+1} - 1)$ and a constant $\beta_K \in \mathbb{R}$. In the case of $K \le 4$, the error term $o(n)$ can be replaced by* $\mathrm{O}(\log n)$.

In the case of $K \le 4$, there are more explicit expansions of $\mathbb{E}\left[X_n\right]$ given through

$$\mathbb{E}\left[X_n\right] = \alpha_K n \log(n) + \beta_K n + \delta_K \log(n) + \epsilon_K + \mathrm{O}\left(\frac{1}{n}\right). \tag{2}$$

For $K = 3$ we have

$$\begin{aligned}
\beta_3 &= \frac{133}{78}\gamma - \frac{2}{117}\sqrt{3}\pi + \frac{4}{39}\log(3) + \frac{3}{26}\log(2) - \frac{6761}{2028},\\
\delta_3 &= \frac{707}{468},\\
\epsilon_3 &= \frac{707}{468}\gamma + \frac{1}{702}\sqrt{3}\pi + \frac{11}{234}\log(3) + \frac{5}{156}\log(2) + \frac{70315}{109512}.
\end{aligned}$$

These results coincide with the explicit expansions stated in [11]. We will need these more precise expansions of $\mathbb{E}\left[X_n\right]$ later only to bound the rate of convergence in Theorem 7. However, we conjecture that such expansions hold for all $K \in \mathbb{N}$ and that the restriction of $K \le 4$ can be removed in our Theorem 7.

| $K$ | $\gamma_K$ | $\alpha_K$ | | $\sigma_K^2$ |
|---|---|---|---|---|
| 1 | 1 | 2 | $= 2$ | 0.420 |
| 2 | $\frac{3}{2}$ | $\frac{9}{5}$ | $= 1.8$ | 0.242 |
| 3 | $\frac{133}{72}$ | $\frac{133}{78}$ | $\approx 1.7051$ | 0.135 |
| 4 | $\frac{2384}{1125}$ | $\frac{9536}{5775}$ | $\approx 1.6513$ | 0.083 |
| 5 | $\frac{36469}{15552}$ | $\frac{182345}{112752}$ | $\approx 1.6172$ | 0.056 |
| 6 | $\frac{31796145419183}{12522149640000}$ | $\frac{31796145419183}{19945995498000}$ | $\approx 1.5941$ | 0.040 |

**Table 1.** Mean and variance for small $K$. The variance is approximated using a Monte-Carlo method on (3) with 1 million samples of $D$.

Asymptotically, as $K \to \infty$, the multi-pivot QuickSort reaches in expectation the information-theoretic lower bound for the number of key comparisons required for sorting:

**Proposition 2.** *We have* $\lim_{K\to\infty} \alpha_K = \frac{1}{\log 2}$.

**Theorem 3.** *Let $K \in \mathbb{N}$. For the number $X_n$ of key comparisons we have*

$$\mathrm{Var}(X_n) \sim \sigma_K^2 n^2 \quad (n \to \infty),$$

*where*

$$\sigma_K^2 = \frac{K+2}{K}\,\mathbb{E}\left[\left(\alpha_K \sum_{i=0}^{K} D_i \log(D_i) + l_{\mathrm{opt}}(D)\right)^2\right]. \tag{3}$$

*For $K = 3$ an explicit value is given as*

$$\sigma_3^2 = \frac{3051169}{657072} - \frac{17689}{36504}\pi^2 + \frac{1463}{2808}\log(2) - \frac{665}{8424}\log(3) \approx 0.1354.$$

Note that $\sigma_1^2 = (21 - 2\pi^2)/3$ (an exact expression for $\mathrm{Var}(X_n)$ for $K = 1$ is given in Knuth [14]), and $\sigma_2^2$ has been derived in [21]. Numerical values for $\sigma_K^2$ for $K = 4, 5, 6$ are given in Table 1.

**Theorem 4.** *Let $K \in \mathbb{N}$ and $\lambda \in \mathbb{R}$. For the normalization $Y_n := (X_n - \mathbb{E}[X_n])/n$ as $n \to \infty$ we have*

$$Y_n \xrightarrow{d} Z_K,$$
$$\mathbb{E}[\exp(\lambda Y_n)] \longrightarrow \mathbb{E}[\exp(\lambda Z_K)] < \infty,$$

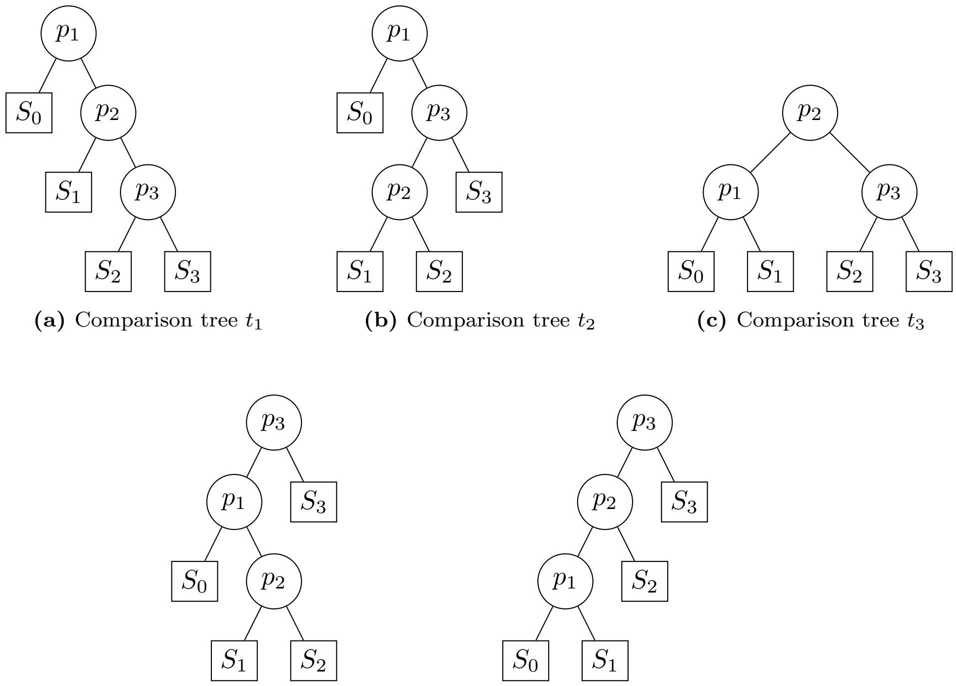
*where the distribution of $Z_K$ is determined as the unique centered, square integrable probability measure such that*

$$Z_K \stackrel{d}{=} \sum_{i=0}^{K} \left(D_i Z_K^{(i)} + \alpha_K D_i \log(D_i)\right) + l_{\mathrm{opt}}(D), \tag{4}$$

*where $D, Z_K^{(0)}, \ldots, Z_K^{(K)}$ are independent and $Z_K^{(i)}$ has the same distribution as $Z_K$ for all $i = 0, \ldots, K$.*

**Remark 5.** *We conjecture that the restriction to square integrability in Theorem 4 can be weakened to integrability in view of [7].*

It is well known that limit distributions as in Theorem 4 have smooth densities, first shown by Fill and Janson [8] for the case $K = 1$. A criterion of Leckey [15] implies:




**(a)** Comparison tree $t_1$ **(b)** Comparison tree $t_2$ **(c)** Comparison tree $t_3$

**(d)** Comparison tree $t_4$ **(e)** Comparison tree $t_5$

**Figure 1.** All comparison trees for $K = 3$

**Theorem 6.** *Let $K \in \mathbb{N}$. The limit $Z_K$ has a smooth Lebesgue density which, together with all its derivatives, is rapidly decreasing.*

Finally, we give bounds on the rate of convergence in Theorem 4 for $K = 2, 3, 4$. For the case $K = 1$ see [9], for $K \geq 5$ we do not know the asymptotic behavior of $\mathbb{E}[X_n]$ well enough. We use the minimal $\ell_p$ metrics for $p \geq 1$ given by

$$\ell_p(\mu, \nu) := \inf\left\{ \|X - Y\|_p : L(X) = \mu, L(Y) = \nu \right\}$$

on the space of real-valued probability measures with a finite absolute $p$th moment, where $\|\cdot\|_p$ denotes the $L_p$-norm. Moreover, we use the Kolmogorov–Smirnov metric defined as

$$\varrho(\mu, \nu) := \sup_{x \in \mathbb{R}} \left| F_\mu(x) - F_\nu(x) \right|,$$

where $F_\mu$ and $F_\nu$ denote the distribution functions of $\mu$ and $\nu$ respectively. When we plug random variables into $\ell_p$ and $\varrho$ we identify them with their probability distributions. We have:

**Theorem 7.** *For $K \in \{2, 3, 4\}$, all $p \geq 1$ and all $\varepsilon > 0$ we have, as $n \to \infty$, that*

$$\ell_p(Y_n, Z_K) = \mathrm{O}\left(n^{-1/2+\varepsilon}\right), \qquad \varrho(Y_n, Z_K) = \mathrm{O}\left(n^{-1/2+\varepsilon}\right).$$

1.2. **Partition Strategies.** Partition strategies can be associated with binary search trees with pivots as internal nodes and lists $S_i$ as external nodes. The depth of a list $S_i$ corresponds to the number of key comparisons to sort an element into this list. We call these trees *classification trees.* In Figure 1 we see the classification trees for $K = 3$ pivot elements. The set of all classification trees for a fixed $K$ is called $\mathcal{T}$. The number of such classification trees is given by the Catalan number $\binom{2K}{K}\frac{1}{K+1}$.

A partition strategy minimizing $\mathbb{E}[X_n]$ is called *optimal*. To analyze the optimality of partition strategies, see [11], we define the linear function

$$l_t(x_0,\dots,x_K) := \sum_{i=0}^{K} x_i h_t(i) \tag{5}$$

on $\mathbb{R}^{K+1}$, where $h_t(i)$ is the depth of the sublist $S_i$ in the tree $t$. So $l_t(D)$ is the expected number of comparisons given $D$, as mentioned above. Let $C_i^{(k)}$ be the number of elements sorted in the list $S_i$ prior to sorting the $k$-th element. [3] showed that given $C^{(k)}$ the probability that the $k$th element is sorted in the sublist $S_i$ is proportional to $C_i^{(k)}+1$ (see [1] for the special case of $K=2$). Therefore, the optimal QuickSort algorithm chooses the classification tree $t$ that minimizes $l_t\left(C^{(k)}+1\right)$. We fix an arbitrary order on $\mathcal{T}$ and choose the smallest tree in the case of a tie. Let $G_t$ be the set of elements sorted according to the classification tree $t$. For a tree $t\in\mathcal{T}$, we define its *asymptotic cone* as

$$C_t := \left\{(x_0,\dots,x_K)\in\mathbb{R}_{\geq 0}^{K+1} : l_t(x_0,\dots,x_K)\leq l_{t'}(x_0,\dots,x_K)\,\forall t'\in\mathcal{T}\right\}.$$

The number of key comparisons to divide a list with $n$ elements into $K+1$ sublists is denoted by $P_n$, with

$$P_n = \sum_{k=1}^{n-K}\sum_{t\in\mathcal{T}}\sum_{i=0}^{K}\mathbf{1}_{\{a_k\in G_t\cap S_i\}} h_t(i) + R_K, \tag{6}$$

where $a_1,\dots,a_n$ denote the elements of the unsorted list. The term $R_K$ is the number of comparisons required to sort the $K$ pivots. By choosing an appropriate sorting algorithm, we can ensure $R_K\leq K^2$. In particular, for fixed $K$ we have $R_K=\mathrm{O}(1)$.

Writing $I_i^{(n)} := |S_i|$ for the sublist sizes, $X_n$ can thus be defined by the recursion

$$X_n \stackrel{d}{=} \sum_{i=0}^{K} X^{(i)}_{I_i^{(n)}} + P_n, \tag{7}$$

for $n\geq K$, where $X_k^{(i)}$ for $k\in\mathbb{N}$ are copies of $X_k$ that are independent of each other and of the $(I_0^{(n)},\dots,I_K^{(n)},P_n)$, $n\geq K$.

## 2. Proofs

The proofs of our results are organized as follows. In Section 2.1 we show the expansion of the mean of $X_n$ from Theorem 1 with the help of classical results on the $m$-ary search trees. Theorem 4 is shown by an application of the contraction method in Section 2.3. The framework of the contraction method used is recalled in Section 2.3. Theorem 3 then follows as a corollary to Theorem 4. The properties of the densities of the limit distributions of Theorem 6 are shown in Section 2.4. The bounds on the rate of convergence from Theorem 7 are finally shown in Section 2.5. Here, we need to develop a couple of new estimates, since the problem for $K\in\{2,3,4\}$ is more involved compared to the case $K=1$.

2.1. **Asymptotic expansion of $\mathbb{E}[X_n]$.** If the algorithm knew $D$ beforehand, it could always use the asymptotically optimal classification tree $t_{\mathrm{opt}}$ where $l_t(D)$ is minimal, see (5). This "oracle strategy" uses only one tree and is easier to analyze. However, we show, see Lemma 8, that the difference in the number of comparisons is minor. Furthermore, the number of comparisons when using $t_{\mathrm{opt}}$ is closely related to the expected internal path length of a $K+1$-ary search tree, which shows the asymptotics of $\mathbb{E}[X_n]$ in Theorem 1.

**Lemma 8.** *The additional cost of the algorithm compared to always using $t_{\text{opt}}$ is of sub-polynomial order:*

$$\alpha_n := \mathbb{E}[P_n] - \mathbb{E}\left[\sum_{i=0}^{K} I_n^{(i)} h_{\text{opt}}(i)\right] = \mathrm{O}\big((\log n)^2\big) \tag{8}$$

*Proof.* The sum $\sum_{i=0}^{K} I_n^{(i)} h_{\text{opt}}(i)$ is the number of comparisons of elements with pivots if we were always using the optimal classification tree. Condition on $D$. Let $t_k$ be the classification tree used for the $k$-th element and $t_{\text{opt}}$ be the optimal classification tree. For every classification tree $t \neq t_{\text{opt}}$, consider

$$\psi_t(x_0, \dots, x_K) := \sum_{i=0}^{K} x_i (h_t(i) - h_{\text{opt}}(i)), \tag{9}$$

the expected number of additional comparisons when using $t$ instead of $t_{\text{opt}}$. By (6),

$$\alpha_n = \sum_{k=1}^{n-K} \sum_{t \in \mathcal{T}} \mathbb{E}\big[\mathbb{P}(a_k \in G_t \mid D)\psi_t(D)\big] + \mathrm{O}(1). \tag{10}$$

Because the algorithm always chooses the best classification tree, there exists $b_t \in \mathbb{R} \setminus \mathbb{Z}$ such that the algorithm chooses $t_{\text{opt}}$ over $t$ if and only if $\psi_t\big(C^{(k)}\big) > b_t$. Furthermore, $t_{\text{opt}}$ is optimal for $D$, so $\psi_t(D) > 0$ for all $t \neq t_{\text{opt}}$, and we can bound $\mathbb{P}(a_k \in G_t)$ by $\mathbb{P}\big(\psi_t\big(C^{(k)}\big) < b_t\big)$.

The variables $\psi_t(C^{(k)})$ are linear combinations of independent indicator variables with $\mathbb{E}[\psi_t(C^{(k)})] = \psi_t(D)$ and coefficients bound by $K$, and thus Hoeffding's inequality with $c = 2K$ can be applied to yield

$$\begin{aligned} \mathbb{P}\Big(\psi_t\Big(C^{(k)}\Big) < b_t \,\Big|\, D\Big) &\leq \exp\bigg(-\frac{2(n\psi_t(D) - b_t)^2}{4K^2 n}\bigg) \\ &= \exp\bigg(-\frac{1}{2K^2} n\psi_t^2(D) + \mathrm{O}\big(\psi_t(D) + n^{-1}\big)\bigg). \end{aligned}$$

Taking the expectation over $D$,

$$\mathbb{E}\big[\mathbb{P}(a_k \in G_t \mid D)\psi_t(D)\big] \leq \mathbb{E}\bigg[\psi_t(D) \exp\bigg(-\frac{1}{2K^2} n\psi_t^2(D) + \mathrm{O}\big(\psi_t(D) + n^{-1}\big)\bigg)\bigg].$$

Now decompose the expectation on whether $\psi_t^2(D)$ is smaller than $4K^2 \log n/n$ into

$$\begin{aligned} &\mathbb{E}\big[\mathbb{P}(a_k \in G_t \mid D)\psi_t(D)\big] \\ &\qquad \leq K\sqrt{\frac{4\log n}{n}}\,\mathbb{P}\left(\psi_t \leq K\sqrt{\frac{4\log(n)}{n}}\right) + \exp(-2\log n + \mathrm{O}(1)). \end{aligned}$$

Since $\psi_t(D)$ is a linear combination of $D$ and $D$ has bounded density (on a simplex), $\psi_t(D)$ itself has bounded density (on $t_{\text{opt}} \neq t$), so

$$\mathbb{E}\big[\mathbb{P}(a_k \in G_t \mid D)\psi_t(D)\big] = \mathrm{O}\left(\frac{\log n}{n}\right) + \mathrm{O}\big(n^{-2}\big) \tag{11}$$

and, considering (10), $\alpha_n$ is of order $\mathrm{O}\big((\log n)^2\big)$. □

**Lemma 9.** *Define the sequence $(\Psi_n)_{n \geq 0}$ recursively as*

$$\Psi_n := \mathbb{E}\left[\sum_{i=0}^{K} \Psi_{I_n^{(i)}} + I_n^{(i)} h_{\text{opt}}(i)\right], \tag{12}$$

*with* $\Psi_n := 0$ *for* $n \leq K$*. Then, there is a constant* $\widehat{\beta}_K \in \mathbb{R}$*, such that*

$$\Psi_n = \frac{\gamma_K}{H_{K+1} - 1} n \log n + \widehat{\beta}_K n + o(n). \tag{13}$$

*Proof.* Conditional on $D$, $\mathbb{E}\Big[I_n^{(i)} \,\Big|\, D\Big] = (n-K)^+ D_i$, so

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=0}^{K} I_n^{(i)} h_{\mathrm{opt}}(i)\right] &= \sum_{i=0}^{K} \mathbb{E}\Big[\mathbb{E}\Big[I_n^{(i)} h_{\mathrm{opt}}(i) \,\Big|\, D\Big]\Big] \\
&= (n-K)^+ \sum_{i=0}^{K} \mathbb{E}[D_i h_{\mathrm{opt}}(i)] \\
&= \gamma_K (n-K)^+. \qquad (14)
\end{aligned}$$

Therefore, $\frac{\Psi_n}{\gamma_K}$ satisfies the recursion

$$\frac{\Psi_n}{\gamma_K} = (n-K)^+ + \mathbb{E}\left[\sum_{i=0}^{K} \frac{\Psi_{I_n^{(i)}}}{\gamma_K}\right], \tag{15}$$

which is the same recursion as for the expected internal path length of a $K+1$-ary tree. This expectation has been identified by Mahmoud [16] as

$$\frac{\Psi_n}{\gamma_K} = \frac{1}{H_{K+1} - 1} n \log n + \frac{\widehat{\beta}_K}{\gamma_K} n + o(n) \tag{16}$$

for some explicitly given $\widehat{\beta}_K \in \mathbb{R}$, see also [4, 17]. □

*Proof of Theorem 1.* By Lemma 9 and (7),

$$\mathbb{E}[X_n] = \frac{\gamma_K}{H_{K+1} - 1} n \log n + \widehat{\beta}_K n + o(n) + \Upsilon_n,$$

where $\Upsilon_n$ is defined by the recursion

$$\Upsilon_n = \alpha_n + \mathbb{E}\left[\sum_{i=0}^{K} \Upsilon_{I_n^{(i)}}\right]. \tag{17}$$

Now, it suffices to show that $\Upsilon_n \sim c_\alpha n$ for some $c_\alpha$. Recursions such as (17) with small $\alpha_n$ were studied by Chern and Hwang [5], who show in their Proposition 7 that

$$\Upsilon_n \sim \frac{K_\alpha}{H_{K+1} - 1} n,$$

where

$$K_\alpha := \sum_{j=0}^{\infty} \frac{\alpha_j}{(j+1)(j+2)},$$

as long as $\alpha_n = o(n)$ and $\sum_j \alpha_j j^{-2} < \infty$, which we show in Lemma 8. See also Fill and Kapur [10] for transfer theorems for $m$-ary search trees. □

We finish this section with the proof of Proposition 2, which shows that the algorithm approaches the optimal first-order term for $K \to \infty$.

*Proof of Proposition 2.* When the classification tree is complete (that is, the height of the leaves differs by at most 1), the algorithm needs at most $\lceil \log_2(K) \rceil$ comparisons, so $\gamma_K \leq \log_2(K) + 1$ Since the expected number of comparisons is bounded from below by the expected binary entropy

$$\gamma_K \geq \mathbb{E}\left[\sum_{i=0}^{K} D_i \log_2(D_i)\right] = \frac{H_{K+1} - 1}{\log(2)} \sim \log_2(K), \tag{18}$$

this bound is sharp.

□

2.2. **The contraction method.** To show the convergence results, we will use the contraction method in the form of [25]; for a more general introduction also see [23] and [18]. Let $(X_n)_{n\geq 0}$ denote a sequence of real-valued random variables satisfying the distributional recurrence

$$X_n \stackrel{d}{=} \sum_{i=0}^{K} A_i^{(n)} X^{(i)}_{I_i^{(n)}} + b_n \tag{19}$$

for $n \geq n_0$, where $\left(X_n^{(0)}\right)_{n\geq 0}, \ldots, \left(X_n^{(K)}\right)_{n\geq 0}$ and $\left(A_0^{(n)}, \ldots, A_K^{(n)}, b_n\right)$ are independent and $X_j^{(i)}$ is distributed as $X_j$ for all $i = 0, \ldots, K$ and $j \geq 0$. The coefficients $A_i^{(n)}$ and $b_n$ are real random variables and $I^{(n)} = \left(I_0^{(n)}, \ldots, I_K^{(n)}\right)$ is a vector of random integers in $0, \ldots, n-K$, while $K$ and $n_0$ are fixed numbers. Furthermore, we assume that the coefficients are square-integrable and the following conditions hold:

(A) $\left(A_0^{(n)}, \ldots, A_K^{(n)}, b_n\right) \stackrel{\ell_2}{\to} (A_0, \ldots, A_K, b)$,
(B) $\sum_{i=0}^{K} \mathbb{E}\left[A_i^2\right] < 1$,
(C) $\sum_{i=0}^{K} \mathbb{E}\left[\mathbf{1}_{\left\{I_i^{(n)} \leq k\right\}} \left(A_i^{(n)}\right)^2\right] \to 0$ as $n \to \infty$ for all constants $k \geq 0$.

Then we have $X_n \stackrel{\ell_2}{\to} X$, where $X$ is the unique fixed point among all centered random variables with finite second moments of

$$X \stackrel{d}{=} \sum_{i=0}^{K} A_i X^{(i)} + b, \tag{20}$$

where $(A_0, \ldots, A_K, b), X^{(0)}, \ldots, X^{(K)}$ are independent and $X^{(i)}$ is distributed as $X$ for $i = 0, \ldots, K$.

2.3. **Proof of Theorem 4.** The normalized number of key comparisons $Y_n$ satisfy the recurrence

$$Y_n \stackrel{\mathrm{d}}{=} \sum_{i=0}^{K} \frac{I_i^{(n)}}{n} Y^{(i)}_{I_i^{(n)}} + \frac{1}{n}\left(P_n - \mathbb{E}[X_n] + \sum_{i=0}^{K} \mathbb{E}\left[X_{I_i^{(n)}} \mid I_i^{(n)}\right]\right). \tag{21}$$

To use the contraction method, we have to show that the following conditions hold:

(1) $\frac{I_i^{(n)}}{n} \stackrel{L_2}{\to} D_i$ for $i = 0, \ldots, K$,
(2) $\frac{P_n}{n} \stackrel{L_2}{\to} \sum_{t\in\mathcal{T}} \mathbf{1}_{\{D\in C_t\}} l_t(D)$,
(3) $\frac{1}{n}\left(\sum_{i=0}^{K} \mathbb{E}\left[X_{I_i^{(n)}} \mid I_i^{(n)}\right] - \mathbb{E}[X_n]\right) \stackrel{L_2}{\to} \sum_{i=0}^{K} \alpha_K D_i \log(D_i)$,
(4) $\sum_{i=0}^{K} \mathbb{E}\left[D_i^2\right] < 1$,
(5) $\mathbb{E}\left[\mathbf{1}_{\left\{I_i^{(n)} \leq k\right\} \cup \left\{I_i^{(n)} = n\right\}} \left(\frac{I_i^{(n)}}{n}\right)^2\right] \stackrel{n\to\infty}{\to} 0$ for all $k \in \mathbb{N}$ and $i = 0, \ldots, K$.

Given $D = (d_0, \ldots, d_K)$, $I_i^{(n)}$ is multinomially $M(n-K; d_0, \ldots, d_K)$ distributed. The strong law of large numbers gives us the almost sure convergence of $\frac{I_i^{(n)}}{n}$ towards $D_i$, and the dominated convergence theorem yields the convergence in $L_2$. Along these lines, the fifth condition also follows. We will now show condition (2).

**Lemma 10.** *We have*

$$\frac{P_n}{n} \xrightarrow{L_2} \sum_{t\in\mathcal{T}} \mathbf{1}_{\{D\in C_t\}} l_t(D).$$

*Proof.* Recall that $a_1, \dots, a_n$ are the elements of the unsorted list, $\{a_i \in G_t\}$ is the event that $a_i$ is sorted with the classification tree $t \in \mathcal{T}$ and $\{a_i \in S_j\}$ denotes that $a_i$ is sorted into sublist $S_j$. We now define for $t \in \mathcal{T}$ and $0 \le j \le K$ the random variable

$$A_{t,j}^{(n)} := \sum_{k=1}^{n-K} \mathbf{1}_{\{a_k \in G_t\} \cap \{a_k \in S_j\}}.$$

We claim $\frac{A_{t,j}^{(n)}}{n} \xrightarrow{L_2} \mathbf{1}_{\{D \in C_t\}} D_j$. In this paper, this is shown for $K = 3$ and $t = t_1$, but it works analogously for other trees.

To show that, we define random walks $W_1 = (W_{1,i})_{i \ge 0}$ and $W_2 = (W_{2,i})_{i \ge 0}$ by

$$W_{1,i} := \sum_{m=1}^{i} \mathbf{1}_{\{a_m \in S_1\}} - \mathbf{1}_{\{a_m \in S_3\}},$$
$$W_{2,i} := \sum_{m=1}^{i} \mathbf{1}_{\{a_m \in S_0\}} - \mathbf{1}_{\{a_m \in S_2\}} - \mathbf{1}_{\{a_m \in S_3\}}.$$

If $W_1$ is positive, the algorithm chooses $t_1$ over $t_2$ and if $W_2$ is nonnegative, it chooses $t_1$ over $t_3$.

Conditionally on $D = (d_0, \dots, d_3)$, the processes $W_1$ and $W_2$ are two simple walks on $\mathbb{Z}$ with constant probabilities to go one step up, one step down, or stay in the actual state. If $d_1 > d_3$ and $d_0 > d_2 + d_3$, $W_1$ and $W_2$ tend to infinity by the strong law of large numbers. Thus, there exists a random $n_0 \in \mathbb{N}$ such that both random walks are positive for every $i \ge n_0$, so the random walk $W := \min\{W_1, W_2\}$ is also positive for every $i \ge n_0$. This implies that starting from index $n_0$, the classification tree $t_1$ is always used. With $|S_j| = I_j^{(n)}$ we get

$$\frac{I_j^{(n)} - n_0}{n} \le \frac{A_{t_1,j}^{(n)}}{n} \le \frac{I_j^{(n)}}{n}$$

and therefore on $\{D_1 > D_3, D_0 > D_2 + D_3\} = \{D \in C_{t_1}\}$ we have $\frac{A_{t_1,j}^{(n)}}{n} \to D_j$ almost surely. Similarly, we can conclude $\frac{A_{t_1,j}^{(n)}}{n} \to 0$ almost surely on the complement, because $W_1$ or $W_2$ tends to $-\infty$ almost surely. Using the dominated convergence theorem, we find

$$\frac{A_{t_1,j}^{(n)}}{n} \xrightarrow{L_2} \mathbf{1}_{\{D \in C_{t_1}\}} D_j.$$

We now use the fact that

$$P_n = \sum_{t \in \mathcal{T}} \sum_{j=0}^{K} A_{t,j}^{(n)} h_j(t) + R_K$$

with $R_K = \mathrm{O}(1)$. So

$$P_n \xrightarrow{L_2} \sum_{t \in \mathcal{T}} \sum_{j=0}^{K} \mathbf{1}_{\{D \in C_t\}} D_j h_j(t) = \sum_{t \in \mathcal{T}} \mathbf{1}_{\{D \in C_t\}} l_t^{\infty}(D), \tag{22}$$

which concludes the proof. □

With the expansion $\mathbb{E}[X_n] = \alpha_K n \log n + \beta_K n + o(n)$ standard calculations imply

$$\frac{1}{n}\left(\sum_{i=0}^{K} \mathbb{E}\left[X_{I_i^{(n)}} \mid I_i^{(n)}\right] - \mathbb{E}[X_n]\right) = \alpha_K \sum_{i=0}^{K} \frac{I_i^{(n)}}{n} \log \frac{I_i^{(n)}}{n} + o(1). \tag{23}$$

The continuous mapping theorem now yields $\frac{I_i^{(n)}}{n} \log \frac{I_i^{(n)}}{n} \to D_i \log(D_i)$ almost surely and with the dominated convergence theorem also in $L_2$. This proves condition (3).

The spacings $D_i$, $i = 0, \ldots, K$ are identically beta$(1, K)$-distributed since $D_0$ is the minimum of $K$ independent, uniformly on $[0, 1]$ distributed random variables. Therefore, condition (4) also holds and the first part of Theorem 4 follows with the contraction method.

For the second part of Theorem 4 we use the following proposition:

**Proposition 11.** *If conditions (A)-(C) in Section 2.2 as well as*

(a) $\sup_{n \in \mathbb{N}} \|b_n\|_\infty < \infty$,

(b) $\sum_{i=0}^{K} \left(A_i^{(n)}\right)^2 < 1 \quad \forall n \in \mathbb{N}$

*hold, we have for all* $\lambda \in \mathbb{R}$

$$\mathbb{E}\left[e^{\lambda X_n}\right] \to \mathbb{E}\left[e^{\lambda X}\right] < \infty.$$

The proof of Proposition 11 is a straightforward extension of an argument of Rösler [22, Section 4]; see also Fill and Janson [9] for a quantified extension of [22, Section 4] and [20, Lemma 4.3]. The second part of Theorem 4 follows from Proposition 11, cf. also [20, Theorem 5.1].

2.4. **Proof of Theorem 6.** Leckey [15], building on [8], shows that $Y$ satisfying the recurrence $Y \stackrel{d}{=} \sum_{i=0}^{\infty} A_i Y^{(i)} + b$ has a smooth and bounded density function if the following conditions hold, where $\alpha^{\max}$ is the largest element and $\alpha^{\sec}$ the second largest element in $(A_i)_{i \geq 0}$:

(1) There exists a constant $a > 0$, such that $\mathbb{P}(\alpha^{\max} \geq a) = 1$,
(2) there are constants $\lambda, \nu > 0$, such that $\mathbb{P}(\alpha^{\sec} \leq x) \leq \lambda x^\nu$ for all $x > 0$,
(3) $\mathbb{P}(A_i \leq 1) = 1$,
(4) there does not exist a $c \in \mathbb{R}$ such that $\mathbb{P}(Y = c) = 1$,
(5) $\mathbb{P}\left(\sum_{i=0}^{\infty} \mathbf{1}_{\{A_i \in (0,1)\}} \geq 1\right) > 0$.

In our case, $A_i = D_i$ for $i = 0, \ldots, K$ and $A_i = 0$ for $i > K$. If we choose $a := \frac{1}{K+1}$, (1) holds. (3) and (5) hold since $D_i \in (0, 1)$ almost surely. Because $Y$ has a positive variance by Theorem 3, there could not be a $c \in \mathbb{R}$ with $Y = c$ almost surely. For condition (2), we need a little calculation:
Since $D_0 + \cdots + D_K = 1$, for $x \in \left(0, \frac{1}{K}\right)$ we have

$$\begin{aligned}
\mathbb{P}(\alpha^{\sec} \leq x) &\leq \mathbb{P}(\max\{D_0, \ldots, D_K\} \geq 1 - K \cdot x) \\
&= \mathbb{P}\left(\bigcup_{i=0}^{K} \{D_i \geq 1 - K \cdot x\}\right) \\
&\leq (K+1)\mathbb{P}(D_0 \geq 1 - K \cdot x) \\
&= (K+1)\mathbb{P}(\min\{U_1, \ldots, U_K\} \geq 1 - K \cdot x) \\
&= (K+1)(K \cdot x)^K = \left((K+1)K^K\right) x^K.
\end{aligned} \tag{24}$$

If we choose $\lambda := \left((K+1) K^K\right)$ and $\nu = K$, (2) holds for $x \in \left(0, \frac{1}{K}\right)$. The function $g(x) := \left((K+1)K^K\right) x^K$ fulfills $g\left(\frac{1}{K}\right) = K + 1 > 1$ and increases monotonically on $(0, 1)$. Therefore, (2) also holds for $x \geq \frac{1}{K}$.

2.5. **Rate of Convergence.** In the present section we are proving Theorem 7. We start with the bounds of the speed of convergence in the $\ell_p$ metrics. For later use, we have the following technical result.

**Lemma 12.** *For $K \in \mathbb{N}$ and all $\varepsilon > 0$, there exists a $\xi > 1$ such that*

$$\sum_{j=0}^{n-K} \frac{(n-j-1)!}{(n-K-j)!} j^{1+\varepsilon} \leq \frac{n^\varepsilon}{\xi} \frac{n!}{K(K+1)(n-K-1)!}.$$

*Proof.* For a fixed $\varepsilon > 0$ we bound

$$\begin{aligned}\sum_{j=0}^{n-K} \frac{(n-j-1)!}{(n-K-j)!} j^{1+\varepsilon} &= \sum_{j=0}^{\lfloor \frac{n}{2} \rfloor} \frac{(n-j-1)!}{(n-K-j)!} j^{1+\varepsilon} + \sum_{j=\lfloor \frac{n}{2} \rfloor + 1}^{n-K} \frac{(n-j-1)!}{(n-K-j)!} j^{1+\varepsilon} \\ &\leq n^\varepsilon \left( \frac{1}{2^\varepsilon} \sum_{j=0}^{\lfloor \frac{n}{2} \rfloor} \frac{(n-j-1)!}{(n-K-j)!} j + \sum_{j=\lfloor \frac{n}{2} \rfloor + 1}^{n-K} \frac{(n-j-1)!}{(n-K-j)!} j \right).\end{aligned}$$

The first summand in the latter display contributes at least $\frac{1}{5}$ of the initial sum, while the second part is smaller than $\frac{4}{5}$ of the initial sum.
(The case $K = 1$ follows from the Gaussian sum formula, while in the case $K \geq 2$ the last term of the second sum is smaller than the first term of the first sum etc.).
Therefore we set $\frac{1}{\xi} := \frac{1}{5 \cdot 2^\varepsilon} + \frac{4}{5} < 1$ and with $x := \sum_{j=0}^{\lfloor \frac{n}{2} \rfloor} \frac{(n-j-1)!}{(n-K-j)!} j$, $y := \sum_{j=\lfloor \frac{n}{2} \rfloor + 1}^{n-K} \frac{(n-j-1)!}{(n-K-j)!} j$ and some $\eta \geq 0$ we obtain

$$\begin{aligned}\frac{1}{2^\varepsilon} x + y &= \frac{1}{2^\varepsilon} \left( \frac{1}{5}(x+y) + \eta \right) + \frac{4}{5}(x+y) - \eta \\ &\leq \left( \frac{1}{5 \cdot 2^\varepsilon} + \frac{4}{5} \right)(x+y) \\ &= \frac{1}{\xi}(x+y).\end{aligned} \tag{25}$$

The statement now follows with $\sum_{j=0}^{n-K} \frac{(n-j-1)!}{(n-K-j)!} j = \frac{n!}{K(K+1)(n-K-1)!}$. □

For bounding $\ell_p$ distances note that it is possible to define random variables $Y, (Y_n)_{n \geq 1}$ on a common probability space, the so-called optimal couplings, such that

$$\ell_p(Y_n, Y) := \ell_p(\mathcal{L}(Y_n), \mathcal{L}(Y)) = \|Y_n - Y\|_p.$$

Therefore, for fixed $2 \leq K \leq 4$, we can define $\left(Z_K, (Y_n)_{n \geq 0}\right)$ such that

$$\Delta(n) := \ell_2(Y_n, Z_K) = \|Y_n - Z_K\|_2.$$

They are also optimal $\ell_p$-couplings for every $p \geq 3$, see, e.g., [24]. Furthermore, we choose $\left(Z_K^{(i)}, \left(Y_n^{(i)}\right)_{n \geq 0}\right)$ as independent copies of $\left(Z_K, (Y_n)_{n \geq 0}\right)$. With the distributional recurrences for $Z_K$ and $Y_n$ we get

$$\begin{aligned}\Delta^2(n) &\leq \mathbb{E}\left[ \left| \sum_{i=0}^{K} \frac{I_i^{(n)}}{n} Y_{I_i^{(n)}}^{(i)} - D_i Z_K^{(i)} + b_n - b \right|^2 \right] \\ &=: \mathbb{E}\left[ \left| \sum_{i=0}^{K} W_i + W_{K+1} \right|^2 \right].\end{aligned} \tag{26}$$

Conditionally on $I^{(n)}$ and $D$, the terms $W_0, \ldots, W_{K+1}$ are independent. Furthermore, we have $\mathbb{E}[W_i] = 0$ for $i = 0, \ldots, K$ and therefore

$$\mathbb{E}\left[ \left( \sum_{i=0}^{K+1} W_i \right)^2 \middle| \left( I^{(n)}, D \right) \right] = \sum_{i=0}^{K+1} \mathbb{E}\left[ W_i^2 \middle| \left( I^{(n)}, D \right) \right] \tag{27}$$

and with (26) we obtain

$$\Delta^2(n) \le \sum_{i=0}^{K+1} \mathbb{E}\left[W_i^2\right] = \sum_{i=0}^{K} \mathbb{E}\left[\left(\frac{I_i^{(n)}}{n} Y^{(i)}_{I_i^{(n)}} - D_i Z_K^{(i)}\right)^2\right] + \mathbb{E}\left[(b_n - b)^2\right]. \tag{28}$$

Our basic strategy to obtain bounds on $\ell_p$ distances for all $p \ge 2$ is as in [9] for the case $K = 1$, i.e., we argue with induction over $p$ and start with the base case $p = 2$.

First, we bound the toll term $b_n - b$ which requires more effort and leads to different bounds in Lemma 13 compared to the case $K = 1$. Recall that

$$b_n = \frac{1}{n}\left(P_n - \mathbb{E}[X_n] + \sum_{i=0}^{K} \mathbb{E}\left[X_{I_i^{(n)}} \middle| I_i^{(n)}\right]\right),$$
$$b = \alpha_K \sum_{i=0}^{K} D_i \log(D_i) + \sum_{t\in T} \mathbf{1}_{\{D\in C_t\}} l_t(D).$$

**Lemma 13.** *For all $p \ge 1$ and $1 \le K \le 4$ we have*

$$\|b_n - b\|_p = \mathrm{O}\left(\frac{1}{\sqrt{n}}\right). \tag{29}$$

*Proof.* By triangle inequality we have

$$\begin{aligned}\|b_n - b\|_p \le{}& \left\| \frac{P_n}{n} - \sum_{t\in T} \mathbf{1}_{\{D\in C_t\}} l_t(D) \right\|_p \\ &+ \left\| \frac{1}{n}\left(\sum_{i=0}^{K} \mathbb{E}\left[X_{I_i^{(n)}} \middle| I_i^{(n)}\right] - \mathbb{E}[X_n]\right) - \alpha_K \sum_{i=0}^{K} D_i \log(D_i) \right\|_p.\end{aligned} \tag{30}$$

For the first summand in the latter display we obtain

$$\left\| \frac{P_n}{n} - \sum_{t\in T} \mathbf{1}_{\{D\in C_t\}} l_t(D) \right\|_p \le \sum_{t\in T}\sum_{i=0}^{K} |h_j(t)| \left\| \left(\frac{A_{t,i}^{(n)}}{n} - \mathbf{1}_{\{D\in C_t\}} D_i\right) \right\|_p + \mathrm{O}\left(\frac{1}{n}\right)$$

and similarly to the proof of Theorem 4 there exists an $n_0 \in \mathbb{N}$ such that

$$\begin{aligned}\left\| \frac{A_{t,i}^{(n)}}{n} - \mathbf{1}_{\{D\in C_t\}} D_i \right\|_p &= \mathbb{E}\left[\left|\frac{A_{t,i}^{(n)}}{n} - \mathbf{1}_{\{D\in C_t\}} D_i\right|^p\right]^{\frac{1}{p}} \\ &\le \left(\mathbb{P}(D \notin C_t)\, \mathbb{E}\left[\left|\frac{n_0}{n}\right|^p\right] + \mathbb{E}\left[\left|\frac{A_{t,j}^{(n)}}{n} - \mathbf{1}_{\{D\in C_t\}} D_j\right|^p \mathbf{1}_{\{D\in C_t\}}\right]\right)^{\frac{1}{p}} \\ &\le \frac{2n_0}{n} + \left\| \frac{I_i^{(n)}}{n} - D_i \right\|_p.\end{aligned}$$

Let $B_{n,u}$ denote a binomial-$(n,u)$-distributed and $\mathrm{Ber}_u$ a Bernoulli-$u$-distributed random variable. Further, let $B(p,q)$ denote the beta-function with parameters $p$ and $q$. In particular, we have $B(1,K) = \frac{1}{K}$. Using bounding ideas of [19], we condition on $D_i$, which is beta$(1,K)$-distributed, and obtain

$$\begin{aligned}
&\left\|\frac{I_i^{(n)}}{n-K} - D_i\right\|_p^p \\
&\quad= \frac{1}{B(1,K)}\int_0^1 (1-u)^{K-1}\mathbb{E}\left[\left|\frac{I_i^{(n)}}{n-K} - D_i\right|^p \;\middle|\; D_i = u\right]\mathrm{d}u \\
&\quad= K\int_0^1 (1-u)^{K-1}\,\frac{1}{(n-K)^p}\,\mathbb{E}\big[|B_{n-K,u} - (n-K)u|^p\big]\mathrm{d}u.
\end{aligned}$$

We now use the Marcinkiewicz–Zygmund inequality [6] to get

$$\begin{aligned}
&\left\|\frac{I_i^{(n)}}{n-K} - D_i\right\|_p \\
&\quad\le \left(K\int_0^1 (1-u)^{K-1}\,\frac{M_p}{(n-K)^p}\mathbb{E}\left[\left(\sum_{j=1}^{n-K}|\mathrm{Ber}_u|^2\right)^{\frac{p}{2}}\right]\mathrm{d}u\right)^{\frac{1}{p}} \\
&\quad\le \left(K\int_0^1 (1-u)^{K-1}\frac{M_p}{(n-K)^{\frac{p}{2}}}\mathrm{d}u\right)^{\frac{1}{p}} \\
&\quad\le \frac{M_p^{1/p}}{\sqrt{n-K}}
\end{aligned}$$

with a constant $M_p$ which only depends on $p$. Overall we obtain

$$\left\|\frac{I_i^{(n)}}{n} - D_i\right\|_p \le \left\|\frac{I_i^{(n)}}{n} - \frac{I_i^{(n)}}{n-K}\right\|_p + \left\|\frac{I_i^{(n)}}{n-K} - D_i\right\|_p = \mathrm{O}\left(\frac{1}{\sqrt{n}}\right)$$

and hence receive our bound for the first summand in (30)

$$\left\|\frac{P_n}{n} - \sum_{t\in T}\mathbf{1}_{\{D\in C_t\}}l_t(D)\right\|_p = \mathrm{O}\left(\frac{1}{\sqrt{n}}\right).$$

To bound the second summand in (30) tightly we need to improve on the bounds used to prove Theorem 1. For $K \le 4$, see eq. (2), we have

$$\begin{aligned}
&\left\|\frac{1}{n}\left(\sum_{i=0}^K \mathbb{E}\left[X_{I_i^{(n)}} \mid I_i^{(n)}\right] - \mathbb{E}\left[X_n\right]\right) - \alpha_K\sum_{i=0}^K D_i\log(D_i)\right\|_p \\
&\quad= \left\|\frac{1}{n}\left(\sum_{i=0}^K \alpha_K I_i^{(n)}\log\left(I_i^{(n)}\right) + \beta_K I_i^{(n)} - (\alpha_K n\log(n) + \beta_K n)\right)\right. \\
&\qquad\quad \left. - \sum_{i=0}^K \alpha_K D_i\log(D_i)\right\|_p + \mathrm{O}\left(\frac{1}{\sqrt{n}}\right) \\
&\quad= \left\|-\frac{\beta_K K}{n} + \alpha_K\left(\sum_{i=0}^K \frac{I_i^{(n)}}{n}\log\frac{I_i^{(n)}}{n} - D_i\log(D_i)\right) - \alpha_K\frac{K}{n}\log(n)\right\|_p + \mathrm{O}\left(\frac{1}{\sqrt{n}}\right) \\
&\quad\le \alpha_K\sum_{i=0}^K\left\|\frac{I_i^{(n)}}{n}\log\frac{I_i^{(n)}}{n} - D_i\log(D_i)\right\|_p + \mathrm{O}\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

Using the same arguments as in the proof of [19, Proposition 2.2] for $p = 3$, we obtain

$$\left\| \frac{I_i^{(n)}}{n} \log \frac{I_i^{(n)}}{n} - D_i \log(D_i) \right\|_p = \mathrm{O}\left(\frac{1}{\sqrt{n}}\right),$$

hence the statement of Lemma 13 follows. $\square$

Let $\varepsilon > 0$ be fixed. We are proving

$$\Delta(n) \leq c n^{-1/2+\varepsilon} \tag{31}$$

for an appropriate constant $c > 0$ by induction over $n$. The induction start is clear. Recall that we have the bound (28) for $\Delta(n)$. To bound the first summand on the right hand side of (28) we start rewriting

$$\begin{aligned}\left(\frac{I_i^{(n)}}{n} Y^{(i)}_{I_i^{(n)}} - D_i Z_K^{(i)}\right)^2 &= \left(\frac{I_i^{(n)}}{n}\left(Y^{(i)}_{I_i^{(n)}} - Z_K^{(i)}\right)\right)^2 + \left(\left(\frac{I_i^{(n)}}{n} - D_i\right) Z_K^{(i)}\right)^2 \\ &\quad + 2\frac{I_i^{(n)}}{n}\left(Y^{(i)}_{I_i^{(n)}} - Z_K^{(i)}\right)\left(\frac{I_i^{(n)}}{n} - D_i\right) Z_K^{(i)}.\end{aligned}$$

For the final factor in the latter display we have

$$\mathbb{E}\left[\left(\left(\frac{I_i^{(n)}}{n} - D_i\right) Z_K^{(i)}\right)^2\right] \leq \left\|\frac{I_i^{(n)}}{n} - D_i\right\|_2^2 \|Z_K\|_2^2 = \mathrm{O}\left(\frac{1}{n}\right), \tag{32}$$

since $Z_K$ has a finite second moment, see Theorem 4. Conditioning on $I_i^{(n)} = j$ and $D_i = u$, we have

$$\begin{aligned}\mathbb{E}\left[\frac{j}{n}\left(\frac{j}{n} - u\right)\left(Y_j^{(i)} - Y^{(i)}\right) Z_K^{(i)}\right] &\leq \frac{j}{n}\left|\frac{j}{n} - u\right| \|Y_j - Z_K\|_2 \, \|Z_K\|_2 \\ &= \frac{j}{n}\left|\frac{j}{n} - u\right| \Delta(j)\sigma,\end{aligned} \tag{33}$$

using the Cauchy–Schwarz inequality and $\sigma := \|Z_K\|_2 < \infty$. Using the inductive hypothesis that $\Delta(j) \leq c j^{-1/2+\varepsilon}$ for $j < n$ we obtain

$$\begin{aligned}\mathbb{E}&\left[\frac{I_i^{(n)}}{n}\left(Y^{(i)}_{I_i^{(n)}} - Z_K^{(i)}\right)\left(\frac{I_i^{(n)}}{n} - D_i\right) Z_K^{(i)} \,\middle|\, I_i^{(n)} = j, D_i = u\right] \\ &\leq c\sigma \frac{j^{\frac{1}{2}+\varepsilon}}{n}\left|\frac{j}{n} - u\right| \leq \frac{c\sigma}{n^{\frac{1}{2}-\varepsilon}}\left|\frac{j}{n} - u\right|.\end{aligned}$$

and there exists an $w > 0$ such that

$$\begin{aligned}\mathbb{E}&\left[\frac{I_i^{(n)}}{n}\left(Y^{(i)}_{I_i^{(n)}} - Z_K^{(i)}\right)\left(\frac{I_i^{(n)}}{n} - D_i\right) Z_K^{(i)}\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\frac{I_i^{(n)}}{n}\left(Y^{(i)}_{I_i^{(n)}} - Z_K^{(i)}\right)\left(\frac{I_i^{(n)}}{n} - D_i\right) Z_K^{(i)} \,\middle|\, I_i^{(n)}, D_i\right]\right] \\ &\leq \frac{c\sigma}{n^{\frac{1}{2}-\varepsilon}}\left\|\frac{I_i^{(n)}}{n} - D_i\right\|_1 \leq \frac{w\sigma c}{n^{\frac{1}{2}-\varepsilon}}\frac{1}{\sqrt{n}} = \frac{w\sigma c}{n^{1-\varepsilon}}\end{aligned}$$

Note that the use of optimal couplings implies $\mathbb{E}[(j/n)^2(Y_j - Z_K)^2] = (j/n)^2\Delta^2(j)$, hence

$$\mathbb{E}\left[\left(\frac{I_i^{(n)}}{n}\left(Y^{(i)}_{I_i^{(n)}} - Z_K^{(i)}\right)\right)\right]^2 \leq \mathbb{E}\left[\left(\frac{I_0^{(n)}}{n}\right)^2 \Delta^2\left(I_i^{(n)}\right)\right].$$

Collecting our estimates, we obtain

$$\Delta^2(n) = (K+1)\left(\mathbb{E}\left[\left(\frac{I_0^{(n)}}{n}\right)^2 \Delta^2\left(I_0^{(n)}\right)\right] + \frac{w\sigma c}{n^{1-\varepsilon}}\right) + \mathrm{O}\left(\frac{1}{n}\right),$$

since the random variables $I_i^{(n)}$ are identically distributed for all $i = 0, \ldots, K$. Note that $D_i$ is $\mathrm{Beta}(1, K)$ distributed and given $D = (d_0, \ldots, d_K)$ the $I^{(n)}$ is multinominally$(n-K; d_0, \ldots, d_K)$ distributed. Using the inductive hypothesis, we have

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{I_0^{(n)}}{n}\right)^2 \Delta^2\left(I_0^{(n)}\right)\right] \\
&= \frac{1}{B(1,K)} \int_0^1 (1-u)^{K-1} \sum_{j=0}^{n-K} \binom{n-K}{j} u^j (1-u)^{n-K-j} \\
&\qquad \times \mathbb{E}\left[\left(\frac{I_0^{(n)}}{n}\right)^2 \Delta^2\left(I_0^{(n)}\right) \mid \left(I_0^{(n)} = j, D_0 = u\right)\right] \mathrm{d}u \\
&\le \frac{K}{n^2} \sum_{j=1}^{n-K} \frac{c^2 (n-K)!}{j!(n-K-j)!} \frac{j^2}{j^{1-2\varepsilon}} \int_0^1 u^j (1-u)^{n-j-1} \mathrm{d}u \\
&= \frac{c^2 K (n-K)!}{n^2} \sum_{j=1}^{n-K} \frac{1}{j!(n-K-j)!} \frac{j!(n-j-1)!}{n!} j^{1+2\varepsilon} \\
&= \frac{c^2 K (n-K)!}{n^2 n!} \sum_{j=1}^{n-K} \frac{(n-j-1)!}{(n-K-j)!} j^{1+2\varepsilon}.
\end{aligned} \tag{34}$$

By Lemma 12 there exists a $\xi > 1$ such that

$$\sum_{j=1}^{n-K} \frac{(n-j-1)!}{(n-K-j)!} j^{1+2\varepsilon} \le \frac{n^{2\varepsilon}}{\xi} \frac{n!}{K(K+1)(n-K-1)!}.$$

It follows

$$\mathbb{E}\left[\left(\frac{I_0^{(n)}}{n}\right)^2 \Delta^2\left(I_0^{(n)}\right)\right] \le \frac{c^2 (n-K) n^{2\varepsilon}}{\xi (K+1) n^2} \le \frac{c^2}{\xi(K+1)} \frac{1}{n^{1-2\varepsilon}}$$

and putting the estimates together, we obtain with an appropriate constant $d > 0$ that

$$\begin{aligned}
\Delta^2(n) &\le (K+1)\left(\frac{c^2}{\xi(K+1)} \frac{1}{n^{1-2\varepsilon}} + \frac{w\sigma c}{n^{1-2\varepsilon}}\right) + \frac{d}{n^{1-2\varepsilon}} \\
&= \left(\frac{1}{\xi} c^2 + (K+1) w\sigma c + d\right) \frac{1}{n^{1-2\varepsilon}} \\
&\le c^2 \frac{1}{n^{1-2\varepsilon}},
\end{aligned} \tag{35}$$

the last inequality being valid for sufficiently large $c$ in view of $\frac{1}{\xi} < 1$. This finishes the proof of the bound on the $\ell_2$ rate of convergence stated in (31).

We now extend the bound in (31) to $\ell_p$ for every $p \ge 1$. Because $\ell_p \le \ell_q$ for $p \le q$, it is sufficient to consider only $p \in \mathbb{N}$. The case $p = 2$ has just been shown above. We now consider $p \in \mathbb{N}$ with $p \geq 3$. Similar to Lemma 3.2 in [9] we have for every $m \in \mathbb{N}$, independent random variables $Q_1, \dots, Q_{m+1}$ and $p \in \mathbb{N}, p \geq 2$ that

$$\mathbb{E}\left[\left|\sum_{i=1}^{m+1} Q_i\right|^p\right] \leq \sum_{i=1}^{m} \mathbb{E}\left[|Q_i|^p\right] + \left(\sum_{i=1}^{m} \|Q_i\|_{p-1} + \|Q_{m+1}\|_p\right)^p. \tag{36}$$

We obtain

$$\Delta_p(n) := \ell_p\left(Y_n, Z_K\right) \leq \left\|\sum_{i=0}^{K+1} W_i\right\|_p$$

with the $W_i$ defined in (26). The Minkowski inequality yields

$$\mathbb{E}\Big[|W_i|^{p-1} \;\Big|\; I_i^{(n)} = j, D_i = u\Big]^{\frac{1}{p-1}} \leq \frac{j}{n}\ell_{p-1}\left(Y_j, Z_K\right) + \left|\frac{j}{n} - u\right| \|Z_K\|_{p-1}. \tag{37}$$

The second part of Theorem 4 implies $\tau := \|Z_K\|_{p-1} < \infty$ since a finite moment generating function yields $\|Z_K\|_p < \infty$ for all $p$. The inductive hypothesis for induction on $p$ is $\ell_{p-1}\left(Y_j, Z_K\right) \leq c_{p-1} j^{-\frac{1}{2}+\varepsilon}$ for all $j \geq 1$. Hence, we obtain

$$\mathbb{E}\left[|W_i|^{p-1} \mid \left(I_i^{(n)} = j, D_i = u\right)\right]^{\frac{1}{p-1}} \leq c_{p-1}\frac{j^{\frac{1}{2}+\varepsilon}}{n} + \tau\left|\frac{j}{n} - u\right| \leq \frac{c_{p-1}}{n^{\frac{1}{2}-\varepsilon}} + \tau\left|\frac{j}{n} - u\right|$$

and therefore, writing $\mathbb{E}_D[\cdot] := \mathbb{E}[\cdot \mid D, I^{(n)}]$,

$$\mathbb{E}_D\left[|W_i|^{p-1}\right]^{\frac{1}{p-1}} \leq \frac{c_{p-1}}{n^{\frac{1}{2}-\varepsilon}} + \tau\left|\frac{I_i^{(n)}}{n} - D_i\right|. \tag{38}$$

Let $k = (k_0, k_1, \dots, k_{K+1}) \in \mathbb{N}^{K+2}$ be a multiindex with $|k| := k_0 + \dots + k_{K+1} = p$. Expanding the power of the sum,

$$\begin{aligned}
&\left(\sum_{i=0}^{K} \mathbb{E}_D\left[|W_i|^{p-1}\right]^{\frac{1}{p-1}} + \mathbb{E}_D[|b_n - b|^p]^{\frac{1}{p}}\right)^p \\
&\quad= \sum_{|k|=p}\binom{p}{k}\prod_{i=0}^{K}\left(\mathbb{E}_D\left[|W_i|^{p-1}\right]^{\frac{1}{p-1}}\right)^{k_i} \mathbb{E}_D[|b_n - b|^p]^{\frac{k_{K+1}}{p}} \\
&\quad\leq \sum_{|k|=p}\binom{p}{k}\prod_{i=0}^{K}\left(\frac{c_{p-1}}{n^{\frac{1}{2}-\varepsilon}} + \tau\left|\frac{I_i^{(n)}}{n} - D_i\right|\right)^{k_i}\mathbb{E}_D[|b_n - b|^p]^{\frac{k_{K+1}}{p}},
\end{aligned}$$

and with (36) we have

$$\begin{aligned}
\mathbb{E}\left[\left|\sum_{i=0}^{K+1} W_i\right|^p\right] &\leq \sum_{i=0}^{K}\mathbb{E}\left[|W_i|^p\right] \\
&\qquad + \sum_{|k|=p}\binom{p}{k}\mathbb{E}\left[\prod_{i=0}^{K}\left(\frac{c_{p-1}}{n^{\frac{1}{2}-\varepsilon}} + \tau\left|\frac{I_i^{(n)}}{n} - D_i\right|\right)^{k_i}\mathbb{E}_D[|b_n - b|^p]^{\frac{k_{K+1}}{p}}\right].
\end{aligned}$$

To further analyze the latter term we use the Hölder's inequality, which implies

$$\begin{aligned}
&\mathbb{E}\left[\prod_{i=0}^{K}\left(\frac{c_{p-1}}{n^{\frac{1}{2}-\varepsilon}} + \tau\left|\frac{I_i^{(n)}}{n} - D_i\right|\right)^{k_i}\mathbb{E}_D[|b_n - b|^p]^{\frac{k_{K+1}}{p}}\right] \\
&\quad\leq \prod_{i=0}^{K}\left\|\left(\frac{c_{p-1}}{n^{\frac{1}{2}-\varepsilon}} + \tau\left|\frac{I_i^{(n)}}{n} - D_i\right|\right)^{k_i}\right\|_{K+2}\left\|\mathbb{E}_D[|b_n - b|^p]^{\frac{k_{K+1}}{p}}\right\|_{K+2}.
\end{aligned} \tag{39}$$

The second factor of the latter term is bounded with Lemma 13 by

$$\left\| \mathbb{E}_D[|b_n - b|^p]^{\frac{k_{K+1}}{p}} \right\|_{K+2} = \mathrm{O}\left(n^{-k_{K+1}/2}\right).$$

For the first term in (39), we have for $k_i \geq 1$ with some constant $w_i > 0$

$$\begin{aligned}
&\mathbb{E}\left[\left|\frac{c_{p-1}}{n^{\frac{1}{2}-\varepsilon}} + \tau\left|\frac{I_i^{(n)}}{n} - D_i\right|\right|^{(K+2)k_i}\right] \\
&\quad = \sum_{j=0}^{(K+2)k_i} \binom{(K+2)k_i}{j} \frac{c_{p-1}^j}{n^{\frac{j}{2}-j\varepsilon}} \tau^{(K+2)k_i-j} \mathbb{E}\left[\left|\frac{I_i^{(n)}}{n} - D_i\right|^{(K+2)k_i-j}\right] \\
&\quad = \sum_{j=0}^{(K+2)k_i} \binom{(K+2)k_i}{j} \frac{c_{p-1}^j}{n^{\frac{j}{2}-j\varepsilon}} \tau^{(K+2)k_i-j} \left\|\frac{I_i^{(n)}}{n} - D_i\right\|_{(K+2)k_i-j}^{(K+2)k_i-j} \\
&\quad \leq \sum_{j=0}^{(K+2)k_i} \binom{(K+2)k_i}{j} \frac{c_{p-1}^j}{n^{\frac{j}{2}-j\varepsilon}} \tau^{(K+2)k_i-j} \frac{w_i}{n^{\frac{(K+2)k_i-j}{2}}} \\
&\quad \leq \sum_{j=0}^{(K+2)k_i} \binom{(K+2)k_i}{j} w_i c_{p-1}^j \tau^{(K+2)k_i-j} \frac{1}{n^{\frac{(K+2)k_i}{2}-(K+2)k_i\varepsilon}} \\
&\quad = \mathrm{O}\left(\frac{1}{n^{\frac{(K+2)k_i}{2}-(K+2)k_i\varepsilon}}\right).
\end{aligned}$$

This yields

$$\left\|\left(\frac{c_{p-1}}{n^{\frac{1}{2}-\varepsilon}} + \tau\left|\frac{I_i^{(n)}}{n} - D_i\right|\right)^{k_i}\right\|_{K+2} = \mathrm{O}\left(\frac{1}{n^{k_i/2-k_i\varepsilon}}\right)$$

and overall we obtain

$$\mathbb{E}\left[\prod_{i=0}^{K}\left(\frac{c_{p-1}}{n^{\frac{1}{2}-\varepsilon}} + \tau\left|\frac{I_i^{(n)}}{n} - D_i\right|\right)^{k_i} \mathbb{E}_D[|b_n - b|^p]^{\frac{k_{K+1}}{p}}\right] = \mathrm{O}\left(\frac{1}{n^{\sum k_i/2-\sum k_i\varepsilon}}\right)$$

and thus

$$\sum_{|k|=p} \binom{p}{k} \mathbb{E}\left[\prod_{i=0}^{K}\left(\frac{c_{p-1}}{n^{\frac{1}{2}-\varepsilon}} + \tau\left|\frac{I_i^{(n)}}{n} - D_i\right|\right)^{k_i} \mathbb{E}_D[|b_n - b|^p]^{\frac{k_{K+1}}{p}}\right] = \mathrm{O}\left(n^{-\frac{p}{2}+p\varepsilon}\right).$$

Collecting the estimates, we obtain

$$\Delta_p^p(n) \leq (K+1)\mathbb{E}\left[|W_0|^p\right] + \mathrm{O}\left(\frac{1}{n^{p\left(\frac{1}{2}-\varepsilon\right)}}\right). \tag{40}$$

For the term $\mathbb{E}\left[|W_0|^p\right]$, analogously to the case $p = 2$, we have

$$\begin{aligned}
\mathbb{E}\left[|W_0|^p\right] &= \mathbb{E}\left[|W_0|^p \mid \left(I_0^{(n)}, D_0\right)\right] \\
&= \sum_{r=0}^{p-1} \binom{p}{r} \mathbb{E}\left[\left(\frac{I_0^{(n)}}{n}\right)^r \Delta_p^r\left(I_0^{(n)}\right) \tau^{p-r}\left|\frac{I_0^{(n)}}{n} - D_0\right|^{p-r}\right] \qquad (41) \\
&\quad + \mathbb{E}\left[\left(\frac{I_0^{(n)}}{n}\right)^p \Delta_p^p\left(I_0^{(n)}\right)\right].
\end{aligned}$$

The inductive hypothesis $\Delta_p(j) \le cj^{-\frac{1}{2}+\varepsilon}$ for $j < n$ yields

$$\begin{aligned}\mathbb{E}\left[\left(\frac{I_0^{(n)}}{n}\right)^r \Delta_p^r\left(I_0^{(n)}\right)\tau^{p-r}\left|\frac{I_0^{(n)}}{n} - D_0\right|^{p-r}\right] &\le \frac{\tau^{p-r}c^r}{n^{r/2-r\varepsilon}}\mathbb{E}\left[\left|\frac{I_0^{(n)}}{n} - D_0\right|^{p-r}\right]\\ &\le \frac{a_r c^r}{n^{p/2-p\varepsilon}}\end{aligned}$$

for some constants $a_r > 0$ for $r = 0, \dots, p-1$. The term $\mathbb{E}\left[\left(\frac{I_0^{(n)}}{n}\right)^p \Delta_p^p\left(I_0^{(n)}\right)\right]$ is bounded explicitly through

$$\begin{aligned}&\mathbb{E}\left[\left(\frac{I_0^{(n)}}{n}\right)^p \Delta_p^p\left(I_0^{(n)}\right)\right]\\ &= \frac{1}{B(1,K)}\int_0^1 (1-u)^{K-1}\sum_{j=0}^{n-K}\binom{n-K}{j}u^j(1-u)^{n-K-j}\\ &\qquad\qquad \times \mathbb{E}\left[\left(\frac{I_0^{(n)}}{n}\right)^p \Delta_p^p\left(I_0^{(n)}\right) \mid \left(I_0^{(n)} = j, D_0 = u\right)\right] du\\ &\le \frac{K}{n^p}\sum_{j=1}^{n-K}\frac{c^p(n-K)!}{j!(n-K-j)!}\frac{j^p}{j^{\frac{p}{2}-p\varepsilon}}\int_0^1 u^j(1-u)^{n-j-1}du\\ &\le \frac{c^pK(n-K)!n^{\frac{p}{2}-1}}{n^p}\sum_{j=1}^{n-K}\frac{1}{j!(n-K-j)!}\frac{j!(n-j-1)!}{n!}j^{1+p\varepsilon}\\ &= \frac{c^pK(n-K)!}{n^{\frac{p}{2}+1}n!}\sum_{j=1}^{n-K}\frac{(n-j-1)!}{(n-K-j)!}j^{1+p\varepsilon}.\end{aligned}$$

By Lemma 12 there exists a $\xi > 1$ (depending in $p$ and being different from the $\xi$ appearing above) such that

$$\sum_{j=1}^{n-K}\frac{(n-j-1)!}{(n-K-j)!}j^{1+p\varepsilon} \le \frac{n^{p\varepsilon}}{\xi}\frac{n!}{K(K+1)(n-K-1)!}.$$

Plugging in, we obtain

$$\mathbb{E}\left[\left(\frac{I_0^{(n)}}{n}\right)^p \Delta^p\left(I_0^{(n)}\right)\right] \le \frac{c^p(n-K)n^{p\varepsilon}}{\xi(K+1)n^{\frac{p}{2}+1}} \le \frac{c^p}{\xi(K+1)}\frac{1}{n^{\frac{p}{2}-p\varepsilon}}. \tag{42}$$

Overall we have

$$\Delta_p^p(n) \le \left(\frac{1}{\xi}c^p + \sum_{i=0}^{p-1}\tilde{a}_i c^i\right)\frac{1}{n^{\frac{p}{2}-p\varepsilon}} \le \frac{c^p}{n^{\frac{p}{2}-p\varepsilon}} \tag{43}$$

with some constants $\tilde{a}_0, \tilde{a}_1, \dots, \tilde{a}_{p-1} > 0$ and $c$ sufficiently large. This finishes the proof of the bounds on the $\ell_p$ metrics.

To bound the distance between $Y_n$ and $Z_K$ in the Kolmogorov-Smirnov metric, we use Lemma 5.1 in [9], which implies

$$\varrho\left(Y_n, Z_K\right) \le \left((1+p)\left\|f_{Z_K}\right\|_\infty^p\right)^{\frac{1}{p+1}}\left(\ell_p\left(Y_n, Z_K\right)\right)^{\frac{p}{p+1}}$$

since $Z_K$ has a bounded density function $f_{Z_K}$ with Theorem 6. We know that for all $p \ge 1$ and $\delta > 0$

$$\ell_p\left(Y_n, Z_K\right) \le \frac{c_p}{n^{\frac{1}{2}-\delta}}$$

with some constant $c_p$. For some fixed $\varepsilon$, we can choose $p$ large enough such that $p/(2(1+p)) > \frac{1}{2} - \varepsilon$. It is possible to choose $\delta > 0$ with $(p/(1+p))\left(\frac{1}{2} - \delta\right) > \frac{1}{2} - \varepsilon$ and thereby obtain

$$\varrho\,(Y_n, Z_K) \le c'_p \frac{1}{n^{\frac{p}{1+p}\left(\frac{1}{2}-\delta\right)}} \le c'_p \frac{1}{n^{\frac{1}{2}-\varepsilon}}$$

where the constant $c'_p$ depends on $\varepsilon$ but not on $n$. This finishes the proof of Theorem 7.

## References


1. Martin Aumüller and Martin Dietzfelbinger, *Optimal partitioning for dual-pivot quicksort*, ACM Trans. Algorithms **12** (2015), no. 2, Art. 18, 36. MR 3465941
2. Martin Aumüller, Martin Dietzfelbinger, Clemens Heuberger, Daniel Krenn, and Helmut Prodinger, *Dual-pivot quicksort: Optimality, analysis and zeros of associated lattice paths*, Combin. Probab. Comput. **28** (2019), no. 4, 485–518. MR 3984045
3. Martin Aumüller, Martin Dietzfelbinger, and Pascal Klaue, *How good is multi-pivot quicksort?*, ACM Trans. Algorithms **13** (2016), no. 1, Art. 8, 47. MR 3598113
4. Nicolas Broutin and Cecilia Holmgren, *The total path length of split trees*, The Annals of Applied Probability **22** (2012), no. 5, 1745 – 1777.
5. Hua-Huai Chern and Hsien-Kuei Hwang, *Phase changes in random $m$-ary search trees and generalized quicksort*, Random Structures & Algorithms **19** (2001), no. 3-4, 316–358.
6. Yuan Shih Chow and Henry Teicher, *Probability theory*, second ed., Springer Texts in Statistics, Springer-Verlag, New York, 1988, Independence, interchangeability, martingales. MR 953964
7. James Allen Fill and Svante Janson, *A characterization of the set of fixed points of the Quicksort transformation*, Electron. Comm. Probab. **5** (2000), 77–84. MR 1781841
8. ———, *Smoothness and decay properties of the limiting Quicksort density function*, Mathematics and computer science (Versailles, 2000), Trends Math., Birkhäuser, Basel, 2000, pp. 53–64. MR 1798287
9. ———, *Quicksort asymptotics*, J. Algorithms **44** (2002), no. 1, 4–28, Analysis of algorithms. MR 1932675
10. James Allen Fill and Nevin Kapur, *Transfer theorems and asymptotic distributional results for $m$-ary search trees*, Random Structures Algorithms **26** (2005), no. 4, 359–391. MR 2139868
11. Clemens Heuberger and Daniel Krenn, *Analysis and optimality of multi-pivot quicksort*, 2025, in preparation.
12. C. A. R. Hoare, *Quicksort*, The Computer Journal **5** (1962), no. 1, 10–16.
13. *Java SE 23 documentation*, https://docs.oracle.com/en/java/javase/23/docs/api/java.base/java Accessed: 2025-02-05.
14. Donald E. Knuth, *The art of computer programming. Volume 3*, Addison-Wesley Series in Computer Science and Information Processing, Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1973, Sorting and searching. MR 445948
15. Kevin Leckey, *On densities for solutions to stochastic fixed point equations*, Random Structures & Algorithms **54** (2019), no. 3, 528–558.
16. Hosam Mahmoud, *On the average internal path length of $m$-ary search trees*, Acta Informatica **23** (1986), 111–117.

17. Götz Olaf Munsonius, *On the asymptotic internal path length and the asymptotic Wiener index of random split trees*, Electron. J. Probab. **16** (2011), no. 35, 1020–1047. MR 2820068
18. Ralph Neininger, *On a multivariate contraction method for random recursive structures with applications to Quicksort*, Random Structures Algorithms **19** (2001), no. 3-4, 498–524, Analysis of algorithms (Krynica Morska, 2000). MR 1871564
19. ———, *Refined Quicksort asymptotics*, Random Structures Algorithms **46** (2015), no. 2, 346–361. MR 3302901
20. Ralph Neininger and Ludger Rüschendorf, *On the internal path length of d-dimensional quad trees*, Random Structures Algorithms **15** (1999), no. 1, 25–41. MR 1698407
21. Ralph Neininger and Jasmin Straub, *Probabilistic analysis of the dual-pivot quicksort "Count"*, 2018 Proceedings of the Fifteenth Workshop on Analytic Algorithmics and Combinatorics (ANALCO), SIAM, Philadelphia, PA, 2018, pp. 1–7. MR 3773630
22. Uwe Rösler, *A limit theorem for "Quicksort"*, RAIRO Inform. Théor. Appl. **25** (1991), no. 1, 85–100. MR 1104413
23. Uwe Rösler and Ludger Rüschendorf, *The contraction method for recursive algorithms*, Algorithmica **29** (2001), no. 1–2, 3–33.
24. Cédric Villani, *Optimal transport*, Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338, Springer-Verlag, Berlin, 2009, Old and new. MR 2459454
25. Sebastian Wild and Markus E. Nebel, *Average case analysis of Java 7's dual pivot Quicksort*, Algorithms—ESA 2012, Lecture Notes in Comput. Sci., vol. 7501, Springer, Heidelberg, 2012, pp. 825–836. MR 3032004
26. Sebastian Wild, Markus E. Nebel, and Ralph Neininger, *Average case and distributional analysis of dual-pivot quicksort*, ACM Trans. Algorithms **11** (2015), no. 3.
27. Vladimir Yaroslavskiy, *Replacement of quicksort in java.util.arrays with new dual-pivot quicksort*, 2009.

Department of Mathematics, Uppsala University, Sweden
*Email address*: cecilia.holmgren@math.uu.se

Institute of Mathematics, Goethe University Frankfurt, Frankfurt a.M., Germany
*Email address*: ischebec@math.uni-frankfurt.de

Fachbereich Mathematik, Paris Lodron University of Salzburg, Austria
*Email address*: daniel.krenn@plus.ac.at

Institute of Mathematics, Goethe University Frankfurt, Frankfurt a.M., Germany
*Email address*: lesny@math.uni-frankfurt.de

Institute of Mathematics, Goethe University Frankfurt, Frankfurt a.M., Germany
*Email address*: neininger@math.uni-frankfurt.de