

# ON FINE FLUCTUATIONS OF THE COMPLEXITY OF THE QUICKSELECT ALGORITHM

JASPER ISCHEBECK, RALPH NEININGER

**ABSTRACT.** The Quickselect algorithm (also called FIND) is a fundamental algorithm for selecting ranks or quantiles within a set of data. Grübel and Rösler showed that the number of key comparisons required by Quickselect considered as a process of the quantiles  $\alpha \in [0, 1]$  converges within a natural probabilistic model after normalization in distribution within the càdlàg space  $D[0, 1]$  endowed with the Skorokhod metric.

We show that the residual process in the latter convergence after normalization converges in distribution towards a mixture of Gaussian processes in  $D[0, 1]$ . A similar result holds for the related algorithm QuickVal. Our method is applicable to other cost measures such as the number of swaps (key exchanges) required by Quickselect, or cost measures being based on key comparisons taking additionally into account that the cost of a comparison between two keys may depend on their values, an example being the number of bit comparisons needed to compare keys given by their bit expansions. For all the arising mixtures of Gaussian limit processes, we also discuss the Hölder continuity of their paths.

## 1. INTRODUCTION AND RESULTS

In 1961, Hoare [13] introduced the Quickselect algorithm, which he called FIND, to select a key (an element) of a given rank from a linearly ordered finite set of data. We assume that the data are distinct real numbers. To be definite, a simple version of the Quickselect algorithm is given as follows:  $\text{FIND}(S, k)$  has as input a set  $S = \{s_1, \dots, s_n\}$  of distinct real numbers of size  $n$  and an integer  $1 \leq k \leq n$ . The algorithm FIND operates recursively as follows: If  $n = 1$  we have  $k = 1$  and FIND returns the single element of  $S$ . If  $n \geq 2$  and  $S = \{s_1, \dots, s_n\}$  the algorithm first chooses an element from  $S$ , say  $s_j$ , called pivot, and generates the sets

$$S_{<} := \{s_i \mid s_i < s_j, i \in \{1, \dots, n\} \setminus \{j\}\}, \quad S_{\geq} := \{s_i \mid s_i \geq s_j, i \in \{1, \dots, n\} \setminus \{j\}\}.$$

If  $k = |S_{<}| + 1$ , the algorithm returns  $s_j$ . If  $k \leq |S_{<}|$ , recursively  $\text{FIND}(S_{<}, k)$  is applied. If  $k \geq |S_{<}| + 2$ , recursively  $\text{FIND}(S_{\geq}, k - |S_{<}| - 1)$  is applied. Note that  $\text{FIND}(S, k)$  returns the element of rank  $k$  from  $S$ . There are various variants of the algorithm, in particular regarding how the pivot element is chosen and how  $S$  is partitioned into the subsets  $S_{<}$  and  $S_{\geq}$ .

In a standard probabilistic model one assumes that the data are ordered, i.e. given as a vector  $(s_1, \dots, s_n)$ , and are randomly permuted, all permutations being equally likely. This can be achieved assuming that the data are given as  $(U_1, \dots, U_n)$  where  $(U_j)_{j \in \mathbb{N}}$  is a sequence of i.i.d. random variables with distribution  $\text{unif}[0, 1]$ , the uniform distribution over the unit interval  $[0, 1]$ . This is the probabilistic model considered below. Note that the randomness is within the data, while the algorithm is deterministic.

Various cost measures have been considered for Quickselect, mainly the number of key comparisons required, which we also analyse. Furthermore, we analyse the number of swaps (key exchanges) required and cost measures which are based on key comparisons, where the cost of a comparison may depend on the values of the keys  $s_i, s_j$ , and the number of bit comparisons required to decide whether  $s_i < s_j$  is a prominent example.

For analysis purposes a related process, called QuickVal, has been considered, see [24, 10]. Informally, QuickVal for an  $\alpha \in [0, 1]$  mimics QuickSelect to select (or to try to select) the value  $\alpha$  from the set of data, which, in our probabilistic model for large  $n$ , comes close to

QuickSelect selecting rank  $\lfloor \alpha n \rfloor$ . To be definite,  $\text{QuickVal}((U_1, \dots, U_n), \alpha)$  compares the  $U_i$  with  $U_1$  to generate sublists

$$S_{<} := (U_{j_1}, \dots, U_{j_{m-1}}), \quad S_{\geq} := (U_{j_{m+1}}, \dots, U_{j_n}),$$

with  $U_{j_i} < U_1$  for  $i = 1, \dots, m-1$  and  $2 \leq j_1 < \dots < j_{m-1}$  and  $U_{j_i} \geq U_1$  for  $i = m+1, \dots, n$  and  $2 \leq j_{m+1} < \dots < j_n$ . Hence,  $m-1 \in \{0, \dots, n-1\}$  is the number of the  $U_i$ ,  $2 \leq i \leq n$ , being smaller than  $U_1$ . The algorithm recursively calls  $\text{QuickVal}(S_{<}, \alpha)$  if  $\alpha < U_1$  and  $|S_{<}| > 0$ . If  $\alpha \geq U_1$  and  $|S_{\geq}| > 0$  recursively  $\text{QuickVal}(S_{\geq}, \alpha - U_1)$  is called. The number of key comparisons required by  $\text{QuickVal}((U_1, \dots, U_n), \alpha)$  is denoted by  $S_{\alpha, n}$ .

To describe the processes  $(S_{\alpha, n})_{\alpha \in [0, 1]}$  and their limit (after scaling) conveniently we also consider the binary search tree constructed from the data  $(U_i)_{i \in \mathbb{N}}$ . Part of the following definitions are depicted in Figure 1. The data are inserted into the rooted infinite binary tree, where we denote its nodes by the elements of  $\{0, 1\}^* := \cup_{n=0}^{\infty} \{0, 1\}^n$  as follows. Its root is denoted by the empty word  $\epsilon$  and for each node  $\phi \in \{0, 1\}^*$  we denote by  $\phi 0$  and  $\phi 1$  (the word  $\phi$  appended with a 0 resp. 1) its left and right child respectively. Moreover,  $|\phi|$  denotes the length of the word  $\phi$ , which is the depth of the corresponding node in the tree. To construct the binary search tree for  $(U_1, \dots, U_n)$  the first key  $U_1$  is inserted into the root and occupies the root. Then, successively the following keys are inserted, where each key traverses the already occupied nodes starting at the root as follows: Whenever the key traversing is less than the occupying key at a node it moves on to the left child of that node, otherwise to its right child. The first empty node found is occupied by the key.

To describe the costs of the algorithms we organize, using notation of Fill and Nakama [10], the sub-intervals  $([L_\phi, R_\phi])_{\phi \in \{0, 1\}^*}$  implicitly generated starting with  $[0, 1] =: [L_\epsilon, R_\epsilon]$  and recursively setting

$$\begin{aligned} \tau_\phi &:= \inf\{i \in \mathbb{N} \mid L_\phi < U_i < R_\phi\}, \\ L_{\phi 0} &:= L_\phi, \quad R_{\phi 0} := R_\phi, \quad L_{\phi 1} := R_{\phi 0} := U_{\tau_\phi}, \quad I_\phi := R_\phi - L_\phi. \end{aligned} \quad (1)$$

Now, if a sublist starting with pivot  $U_{\tau_\phi}$  has to be split by  $\text{QuickVal}$ , the keys which are inserted in the subtree rooted at  $U_{\tau_\phi}$  need to be compared with  $U_{\tau_\phi}$ . Hence, we get a contribution of key comparisons of

$$S_{\phi, n} = \sum_{\tau_\phi < k \leq n} \mathbf{1}_{[L_\phi, R_\phi]}(U_k). \quad (2)$$

Now, for  $\alpha \in [0, 1]$ ,  $\text{QuickVal}((U_1, \dots, U_n), \alpha)$  generates and splits sublists encoded by  $\phi(\alpha, k)$  for  $k = 0, 1, \dots$  for which we obtain by  $\phi(\alpha, 0) = \epsilon$  and

$$\phi(\alpha, k+1) = \begin{cases} \phi(\alpha, k)0, & \text{if } \alpha < U_{\tau_{\phi(\alpha, k)}}, \\ \phi(\alpha, k)1, & \text{if } \alpha \geq U_{\tau_{\phi(\alpha, k)}}. \end{cases} \quad (3)$$

When using the variables defined in (1) or (2), we abbreviate the notation  $\phi(\alpha, k)$  by  $\alpha, k$ , such as writing  $I_{\alpha, k} := I_{\phi(\alpha, k)}$  or  $S_{\alpha, k, n} := S_{\phi(\alpha, k), n}$ .

The number of key comparisons required by  $\text{QuickVal}((U_1, \dots, U_n), \alpha)$  is then given by the (finite) sum

$$S_{\alpha, n} = \sum_{k=1}^{\infty} S_{\alpha, k, n}.$$

Fill and Nakama [10, Theorem 3.2] showed (considering more general complexity measures) that for each  $\alpha \in [0, 1]$  almost surely

$$\frac{1}{n} S_{\alpha, n} \rightarrow S_\alpha := \sum_{k=0}^{\infty} I_{\alpha, k}, \quad (n \rightarrow \infty). \quad (4)$$

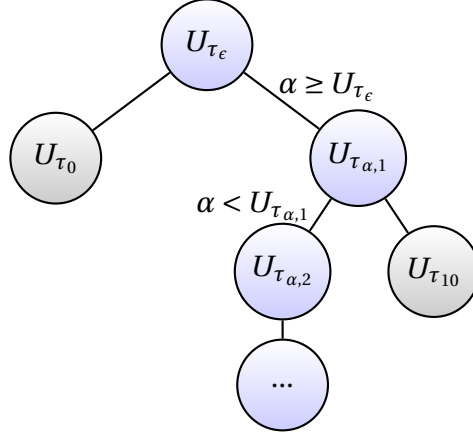


FIGURE 1. Part of the binary search tree. The pivots of sublists split by QuickVal $((U_1, \dots, U_n), \alpha)$  for some  $\alpha \in [0, 1]$  are on the path indicated. Note that we have  $\tau_\epsilon = \tau_{\phi(\alpha, 0)} = \tau_{\alpha, 0} = 1$  and in this example  $\alpha \geq U_1$  and  $\alpha < U_{\tau_{\alpha, 1}}$  so that  $\phi(\alpha, 2) = 10 \in \{0, 1\}^2$ .

The latter convergence also holds in  $L_p$ , see Fill and Matterer [8, Proposition 6.1].

We take the point of view that such an almost sure asymptotic result may be considered a strong law of large numbers (SLLN). The subject of the present paper is to study the fluctuations in such SLLN, sometimes called a central limit analogue. We study these fluctuations as processes in the metric space  $(D[0, 1], d_{\text{SK}})$  of càdlàg functions endowed with the Skorokhod metric; see Section 2.1 for the definitions and Billingsley [2] for background on weak convergence of probability measures on metric spaces in general and on  $(D[0, 1], d_{\text{SK}})$  in particular. Note that, by definition,  $(S_{\alpha, n})_{\alpha \in [0, 1]}$  and  $(S_\alpha)_{\alpha \in [0, 1]}$  have càdlàg paths almost surely.

The analysis of QuickVal is usually considered an intermediate step to analyze the original FIND algorithm. Grübel and Rösler [11] already pointed out that a version of FIND such as stated above with  $C_n^*(k)$  denoting the number of key comparisons for finding rank  $k$  within  $(U_1, \dots, U_n)$  does not lead to convergence within  $(D[0, 1], d_{\text{SK}})$  after the normalization  $\alpha \mapsto \frac{1}{n} C_n^*([\alpha n] + 1)$ , where here and below the convention  $C_n^*(n+1) := C_n^*(n)$  is used. To overcome this problem they propose a version that does not stop in case a pivot turns out to be the rank to be selected by including the pivot in the list  $S_<$  and proceeding until a list of size 1 is generated. Moreover, their pivots are chosen uniformly at random. The number of key comparisons  $C'_n(k)$  for Grübel and Rösler's FIND-version has the property that

$$\left( \frac{1}{n} C'_n([\alpha n] + 1) \right)_{\alpha \in [0, 1]} \xrightarrow{d} (S_\alpha)_{\alpha \in [0, 1]}, \quad \text{in } (D[0, 1], d_{\text{SK}}), \quad (5)$$

see [11, Theorem 4]. We may also obtain right-continuous limits with our deterministic choice of pivots by just recursively calling FIND $(S_\geq, 0)$  in case the pivot turns out to be the rank sought. We denote the number of key comparisons for this version by  $C_n(k)$ , which is close to Grübel and Rösler's FIND-version and also satisfies (5).

The convergence in (5) could only be stated weakly (not almost surely) since Grübel und Rösler's FIND-version due to randomization within the algorithm does not have a natural embedding on a probability space. Note that the formulation of the QuickVal complexity does have such an embedding which, e.g., makes the almost sure convergence in (4) possible. Also, it is easy to see that we have the distributional equality

$$(C_n(|\{U_i \leq \alpha : 1 \leq i \leq n\}|))_{\alpha \in [0, 1]} \stackrel{d}{=} (S_{\alpha, n})_{\alpha \in [0, 1]}. \quad (6)$$

This allows to naturally couple the complexities on one probability space, which we call its *natural coupling*. See [8, page 807] for a related discussion of natural couplings.

**1.1. Results.** In this section, we collect our results for three different complexity measures, the number of key comparisons, the number of key exchanges (also called swaps), and measures being based on key comparisons taking additionally into account that the cost of a comparison between two keys may depend on their values, an example being the number of bit comparisons needed to compare keys given by their bit expansions.

Part of our results has been announced in the extended abstract [14].

**1.1.1. Number of key comparisons.** As the normalized process of fluctuations we define

$$G_n := (G_{\alpha,n})_{\alpha \in [0,1]} := \left( \frac{S_{\alpha,n} - nS_\alpha}{\sqrt{n}} \right)_{\alpha \in [0,1]}. \quad (7)$$

Then we have the following result:

**Theorem 1.1.** *Let  $S_{\alpha,n}$  be the number of key comparisons required by  $\text{QuickVal}((U_1, \dots, U_n), \alpha)$  and  $(S_\alpha)_{\alpha \in [0,1]}$  as in (4). Then for the fluctuation process  $G_n$  defined in (7) we have*

$$G_n \xrightarrow{d} G_\infty \text{ in } (D[0,1], d_{\text{SK}}) \quad (n \rightarrow \infty),$$

where  $G_\infty$  is a mixture of centred Gaussian processes with random covariance function given by

$$\Sigma_{\infty,\alpha,\beta} := \sum_{k=0}^J \sum_{j=0}^{\infty} I_{\alpha,j \vee k} + \mathbf{1}_{\{\alpha \neq \beta\}} (J+1) \sum_{j=J+1}^{\infty} (I_{\beta,j}) - S_\alpha S_\beta, \quad \alpha, \beta \in [0,1], \quad (8)$$

where  $J = J(\alpha, \beta) := \max\{k \in \mathbb{N}_0 \mid \tau_{\alpha,k} = \tau_{\beta,k}\} \in \mathbb{N}_0 \cup \{\infty\}$ . We have the representation

$$G_\infty \stackrel{d}{=} \sum_{\phi \in \{0,1\}^*} Z_\phi \mathbf{1}_{[L_\phi, R_\phi)}, \quad (9)$$

with  $Z_\phi$  defined in Section 2.4.

**Remark 1.2.** *An alternative representation of the random covariance function in (8) is as follows: With a random variable  $V$  uniformly distributed over  $[0,1]$  and independent of the  $(U_i)_{i \in \mathbb{N}}$ , we have*

$$\Sigma_{\infty,\alpha,\beta} = \text{Cov}(J(V, \alpha), J(V, \beta) \mid \mathcal{F}_\infty), \quad (10)$$

with the  $\sigma$ -algebra

$$\mathcal{F}_\infty := \sigma\{I_\phi \mid \phi \in \{0,1\}^*\}. \quad (11)$$

**Remark 1.3.** *A related functional limit law for the complexity of Radix Selection, an algorithm to select ranks based on the bit expansions of the data, with a limiting Gaussian process with a covariance function related to (10) can be found in [17, Theorem 1.2]. See [23, Theorem 1.1] for another related functional limit law.*

**Remark 1.4.** *In the recent preprint Fill and Matterer [9, Theorems 5.1 and 6.1.] convergence of the one-dimensional marginals for the functional limit law in Theorem 1.1 is shown, in distribution and with all moments. See also the PhD thesis, Matterer [20, Theorem 6.4].*

To transfer Theorem 1.1 to FIND we need to align jumps to come up with a suitable fluctuation process. The conventions  $C_n(0) := C_n(1)$  and  $C_n(n+1) := C_n(n)$  are used.

**Corollary 1.5.** Let  $C_n(k)$  be the number of key comparisons required to select rank  $1 \leq k \leq n$  within a set of  $n$  data by FIND with the natural coupling (6). Let  $\Lambda_n : [0, 1] \rightarrow [0, 1]$ ,  $n \in \mathbb{N}$ , be any (random) monotone increasing bijective function such that  $\Lambda_n(\frac{k}{n+1})$  is equal to the element of rank  $k$  within  $\{U_1, \dots, U_n\}$ . Then we have

$$\left( \frac{C_n(\lfloor t(n+1) \rfloor) - nS_{\Lambda_n(t)}}{\sqrt{n}} \right)_{t \in [0,1]} \xrightarrow{d} G_\infty \text{ in } (D[0,1], d_{\text{SK}}),$$

where  $G_\infty$  is the process defined in Theorem 1.1.

**1.1.2. Number of swaps.** Usually, QuickSelect is implemented in-place, which means that it only requires the memory for the list  $S$  of values and a bounded amount of additional memory. This is achieved by generating the sets  $S_<$  and  $S_\geq$  by swapping values within the ordered list that contains the elements of  $S$  so that the elements of  $S_<$  and  $S_\geq$  are moved to contiguous parts of the original list. Such a procedure is called *partition*. There are various *partition* procedures, we are discussing two of them.

The original *partition* procedure by Hoare [12] searches the list  $S$  from both ends at once: It repeatedly finds the index  $i = \min\{2 \leq i \leq n \mid U_i > U_1\}$  of the leftmost element bigger than the pivot and the index  $j := \max\{2 \leq j \leq n \mid U_j < U_1\}$  of the rightmost element smaller than the pivot. If  $i < j$ , it swaps  $U_i$  and  $U_j$ . Else the algorithm terminates.

A simpler, but less efficient implementation is the so-called Lomuto partition scheme [1, 3, 19] that only searches from one end of  $S$ . It keeps track of the amount  $i$  of elements at the start of the list it has already swapped. In every step, it finds the index  $j := \max\{2 \leq j \leq n \mid U_j < U_1\}$  of the rightmost element smaller than the pivot. If  $i + 1 < j$ , it swaps  $U_{i+1}$  and  $U_j$  and increases  $i$  by one. Otherwise, the algorithm terminates.

Both partition schemes only compare elements to the pivot, so the model of randomness is preserved within the sublists  $S_<$  and  $S_\geq$ . However, their original order is not preserved, so QuickSelect run on  $U_1, \dots, U_n$  will usually not select the same pivots as QuickSelect on  $U_1, \dots, U_{n+1}$ . For convenience, we assume that the pivot to split a sublist  $S'$  of  $S$  is the element of  $S'$  that came first in the original list  $S$ . We call this choice of the pivots a *suitable embedding*.

(i) *Hoare's partition.* For Hoare's partition procedure, via a hypergeometric distribution, for details see Section 2.5, the expected number of swaps in step  $k$  given  $\mathcal{F}_\infty$  is approximately  $nI_{\alpha,k+1}(I_\alpha - I_{\alpha,k})/I_{\alpha,k}$ , which leads to the limit process  $L = (L_\alpha)_{\alpha \in [0,1]}$  given by

$$L_\alpha := \sum_{k=0}^{\infty} \frac{I_{\alpha,k+1}(I_\alpha - I_{\alpha,k})}{I_{\alpha,k}}, \quad \alpha \in [0,1]. \quad (12)$$

With the  $Z_\phi$  appearing in Theorem 1.1 and  $\{Y_\phi \mid \phi \in \{0,1\}^*\}$  a set of i.i.d.  $\mathcal{N}(0,1)$  random variables being independent of  $\{Z_\phi \mid \phi \in \{0,1\}^*\}$  and of  $\mathcal{F}_\infty$  the limiting process  $G^{\text{swap}} = (G_\alpha^{\text{swap}})_{\alpha \in [0,1]}$  is given by

$$G_\alpha^{\text{swap}} := \sum_{\phi \in \{\phi(k,\alpha) \mid k \in \mathbb{N}_0\}} Y_\phi \frac{I_{\phi 0} I_{\phi 1}}{I_\phi^{3/2}} + Z_{\phi 0} \frac{I_{\phi 1}}{I_\phi} + Z_{\phi 1} \frac{I_{\phi 0}}{I_\phi} - Z_\phi \frac{I_{\phi 0} I_{\phi 1}}{I_\phi^2}, \quad \alpha \in [0,1]. \quad (13)$$

The  $Y_\phi$  represent fluctuations caused by the hypergeometric distribution related to the present partition procedure, while the terms containing  $Z_\phi$  represent fluctuations around the limit (12). Then we have the following result for key exchanges corresponding to Theorem 1.1:

**Theorem 1.6.** Let  $K_{\alpha,n}$  be the number of key exchanges required by  $\text{QuickVal}((U_1, \dots, U_n), \alpha)$  with Hoare's partition algorithm in a suitable embedding. Then, as  $n \rightarrow \infty$ , we have

$$\left( \frac{K_{\alpha,n} - nL_\alpha}{\sqrt{n}} \right)_{\alpha \in [0,1]} \xrightarrow{d} G^{\text{swap}} \quad \text{in } (D[0,1], d_{\text{SK}}).$$

Furthermore, the latter convergence also holds jointly with the convergence for the comparisons,

$$\left( \frac{S_{\alpha,n} - nS_\alpha}{\sqrt{n}} \right)_{\alpha \in [0,1]} \xrightarrow{d} G_\infty \quad \text{in } (D[0,1], d_{\text{SK}}),$$

from Theorem 1.1, when  $G^{\text{swap}}$  and  $G_\infty$  are constructed as in (13) and (50) respectively using the same  $Z_\phi$ .

(ii) *Lomuto's partition.*

The Lomuto partition procedure is simpler to implement and much easier to analyze. It swaps the pivot and every element smaller than the pivot, so the amount of swaps at some path  $\phi \in \{0,1\}^*$  is given by  $S_{\phi 0} + 1$ . Thus, the limiting process  $G^{\text{Lo}} = (G_\alpha^{\text{Lo}})_{\alpha \in [0,1]}$  is

$$G_\alpha^{\text{Lo}} = \sum_{k=0}^{\infty} Z_{\phi(\alpha,k)0}, \quad \alpha \in [0,1].$$

Our proof for the number of key comparisons in Theorem 1.1 can be straightforwardly transferred to yield:

**Theorem 1.7.** Let  $K_{\alpha,n}^{\text{Lo}}$  be the number of key exchanges required by  $\text{QuickVal}((U_1, \dots, U_n), \alpha)$  with Lomuto's partition procedure in a suitable embedding. Then, as  $n \rightarrow \infty$ , we have

$$\left( \frac{K_{\alpha,n}^{\text{Lo}} - n \sum_{k=0}^{\infty} I_{\phi(\alpha,k)0}}{\sqrt{n}} \right)_{\alpha \in [0,1]} \xrightarrow{d} G^{\text{Lo}} \quad \text{in } (D[0,1], d_{\text{SK}}). \quad (14)$$

1.1.3.  *$\varepsilon$ -tame cost functions.* We now consider the model where the cost to compare two keys depends on their values. These costs are described by a measurable *cost function*  $\beta : [0,1]^2 \rightarrow [0,\infty)$ , and we require that they have a polynomial tail, that is: There are constants  $c, \varepsilon > 0$  such that for all  $u \in [0,1]$ ,  $x \in \mathbb{N}$  and for  $V$  being  $\text{unif}[0,1]$  distributed

$$\mathbb{P}(\beta(u, V) \geq x) \leq cx^{-1/\varepsilon}.$$

This condition is called  $(c, \varepsilon)$ -*tameness*, see [10, 20, 9], and  $\beta$  is called  $\varepsilon$ -*tame* if it is  $(c, \varepsilon)$ -tame for some  $c > 0$ . Note that, e.g.,  $\beta$  counting the number of bit comparisons is  $\varepsilon$ -tame for all  $\varepsilon > 0$ . The costs of  $\text{QuickVal}((U_1, \dots, U_n), \alpha)$  in this model are given by

$$S_{\alpha,n}^\beta := \sum_{k=0}^{\infty} \sum_{\tau_{\alpha,k} < i \leq n} \mathbf{1}_{[L_{\alpha,k}, R_{\alpha,k})}(U_i) \beta(U_{\tau_{\alpha,k}}, U_i)$$

and the limit is, with  $V$  being  $\text{unif}[0,1]$  distributed and independent of the  $U_1, \dots, U_n$ , given as

$$S_{\alpha,\infty}^\beta := \sum_{k=0}^{\infty} \mathbb{E}[\mathbf{1}_{[L_{\alpha,k}, R_{\alpha,k})}(V) \beta(U_{\tau_{\alpha,k}}, V) \mid \mathcal{F}_\infty]$$

[9] show for  $\varepsilon < \frac{1}{2}$  that for fixed  $\alpha \in [0,1]$  the resulting residual

$$G_{\alpha,n}^\beta := \frac{S_{\alpha,n}^\beta - nS_{\alpha,\infty}^\beta}{\sqrt{n}}$$

converges to a mixed centred Gaussian random variable  $G_{\alpha,\infty}^\beta$  in distribution and with all moments. It is possible to combine them to a mixture of centred Gaussian processes

$$G_\infty^\beta = (G_{\alpha,\infty}^\beta)_{\alpha \in [0,1]}, \quad (15)$$

defined by the conditional covariance functions given, with

$$X_{\alpha,k}^\beta := \mathbf{1}_{[L_{\alpha,k}, R_{\alpha,k})}(V) \cdot \beta(U_{\tau_{\alpha,k,n}}, V), \quad (16)$$

by

$$\text{Cov}(G_{\alpha,\infty}^\beta, G_{\gamma,\infty}^\beta \mid \mathcal{F}_\infty) = \text{Cov}\left(\sum_{k=0}^{\infty} X_{\alpha,k}^\beta, \sum_{k=0}^{\infty} X_{\gamma,k}^\beta \mid \mathcal{F}_\infty\right), \quad \alpha, \gamma \in [0,1]. \quad (17)$$

We show later, in Lemma 2.6, that the latter covariance is well-defined for  $s < \varepsilon^{-1}$ .

Note that now the total cost for all key comparisons required by  $\text{QuickVal}((U_1, \dots, U_n), \alpha)$  or  $\text{FIND}((U_1, \dots, U_n), k)$  is no longer determined by the fact that the ranks of  $(U_1, \dots, U_n)$  form an uniformly random permutation. Here, the distribution of the  $U_i$  matters. We only consider the uniform distribution as for the other cost measures in the present paper and leave other distributions for future work. We have the following result corresponding to Theorem 1.1.

**Theorem 1.8** ( $\varepsilon$ -tame cost function). *Let  $S_{\alpha,n}^\beta$  be the cost of  $\text{QuickVal}((U_1, \dots, U_n), \alpha)$  for an  $\varepsilon$ -tame cost function  $\beta$  with  $\varepsilon < \frac{1}{4}$ . Then we have*

$$\left( \frac{S_{\alpha,n}^\beta - nS_{\alpha,\infty}^\beta}{\sqrt{n}} \right)_{\alpha \in [0,1]} \xrightarrow{d} G_\infty^\beta \quad \text{in } (D[0,1], d_{\text{SK}}),$$

where  $G_\infty^\beta$  is the mixture of centred Gaussian processes defined in (15).

**1.1.4. Path properties of the limit processes.** It is obvious from the covariance function that the limit process  $G_\infty$  is a.s. discontinuous at every  $U_i$ ,  $i \in \mathbb{N}$ , within the Euclidean metric. We equip  $[0,1]$  with two equivalent (random) metrics  $d_2$  and  $d_G$ , which are  $\mathcal{F}_\infty$ -measurable, and show that  $G_\infty$  is Hölder continuous in these metrics.

The first metric we consider is

$$d_2(\alpha, \beta) := 2^{-J(\alpha, \beta)} \quad (18)$$

with  $J(\alpha, \beta)$  defined in Theorem 1.1. Then,  $d_2$  is an ultrametric (i.e. the longest two sides of a triangle have the same length) and the balls in this metric are the intervals  $[L_\phi, R_\phi]$  resp.  $[L_\phi, R_\phi]$  in the case of  $R_\phi = 1$ . While we could use another base than 2, this choice is the natural choice in which the Hausdorff dimension of  $([0,1], d_2)$  is one. A function defined on  $([0,1], d_2)$  is continuous if and only if it is càdlàg and continuous in the Euclidean metric at all points outside the set  $\{U_\phi \mid \phi \in [0,1]^*\}$ . Consequently, the topology induced by  $d_2$  is coarser than the Euclidean topology on  $[0,1]$ .

The second metric we consider is the canonical metric of the (mixed) Gaussian process  $G_\infty$  given by

$$d_G(\alpha, \beta) := \mathbb{E}[(G_{\alpha,\infty} - G_{\beta,\infty})^2 \mid \mathcal{F}_\infty]^{\frac{1}{2}}. \quad (19)$$

We show in Lemma 2.18 that the metrics  $d_2$  and  $d_G$  are a.s. equivalent.

**Theorem 1.9.** *For the limit processes  $G_\infty$ ,  $G_\infty^{\text{swap}}$  and  $G_\infty^\beta$  appearing in Theorems 1.1, 1.6 and 1.8 we have:*

- (1) *The processes  $G_\infty$  and  $G_\infty^{\text{swap}}$  are a.s. Hölder continuous with respect to  $d_2$  for any Hölder exponent  $\gamma < \frac{1}{4} \log_2 \frac{3}{2} \approx 0.1462$  and Hölder continuous with regard to  $d_G$  for any Hölder exponent  $\gamma < 1$ .*

- (2) For an  $\varepsilon$ -tame comparison cost function  $\beta$  with  $\varepsilon < 1/4$ , the residual process  $G_\infty^\beta$  is a.s.  $d_2$ -Hölder continuous with Hölder exponent  $\gamma < (1 - 2\varepsilon)\frac{1}{4}\log_2\frac{3}{2}$  and  $d_G$ -Hölder continuous with Hölder exponent  $\gamma < 1 - 2\varepsilon$ .

## 2. PROOFS

The present section is organized as follows: Section 2.1 contains an abstract criterion, Proposition 2.2, for weak convergence of probability measures on  $(D[0, 1], d_{\text{SK}})$ . All the functional limit theorems in the present paper are obtained via Proposition 2.2. To verify the conditions of Proposition 2.2 first in Section 2.2 a novel perturbation argument is introduced, which is the basis of our analysis. Then in Section 2.3 estimates are presented to use Proposition 2.2 to prove the main result, Theorem 1.1, on the number of key comparisons. Section 2.4 is an intermezzo where the explicit representation of the limit process appearing in Theorem 1.1 is added. Then, Sections 2.5 and 2.6 contain the estimates analogous to Section 2.3 to prove the main results on the number of swaps and on  $\varepsilon$ -tame cost functions from Theorems 1.6 and 1.8 respectively. The final Section 2.7 contains the proof of the path properties of the limit processes stated in Theorem 1.9.

**2.1. On Weak Convergence in  $D[0, 1]$ .** Within the space  $D[0, 1]$  of all càdlàg functions, closeness of functions  $f, g \in D[0, 1]$  is measured in the Skorokhod metric

$$d_{\text{SK}}(f, g) = \inf_{\lambda} \max\{\|f \circ \lambda - g\|_\infty, \|\lambda - \text{id}\|_\infty\}, \quad (20)$$

where the infimum is taken over all increasing bijections  $\lambda : [0, 1] \rightarrow [0, 1]$  and  $\text{id}$  denotes identity.

To obtain weak convergence on  $(D[0, 1], d_{\text{SK}})$  we first show tightness so that our sequences of processes in Theorems 1.1, 1.6 and 1.8 have weakly convergent subsequences respectively, and then show uniqueness of the limit. The tightness on  $(D[0, 1], d_{\text{SK}})$  can be shown using an equivalent of the Arzelà–Ascoli theorem, whereas the uniqueness follows from the fdd convergence.

For some  $x \in D[0, 1]$  and  $\delta > 0$ , we define a version of a modulus of continuity by

$$w'_x(\delta) = \inf_{\{t_i\}} \max_{0 < i \leq r} \sup_{t_{i-1} \leq s, u < t_i} |x(s) - x(u)|. \quad (21)$$

The infimum is taken over finite sets  $\{t_i\}_{i=1}^r$  with

$$\begin{cases} 0 = t_0 < \dots < t_r = 1 \text{ and} \\ t_i - t_{i-1} > \delta \quad \text{for } i = 1, 2, \dots, r. \end{cases}$$

We start recalling a well-known tightness criterion:

**Theorem 2.1** (Theorem 13.2 in [2]). *Let  $\{P_n\}_{n \in \mathbb{N}}$  be a sequence of probability measures on  $(D[0, 1], d_{\text{SK}})$ . The family  $\{P_n\}$  is tight if and only if the following two conditions are satisfied:*

- (1) For each  $\eta > 0$ , there exists an  $a > 0$  such that

$$\sup_{n \geq 1} P_n(\{x : \|x\|_\infty > a\}) \leq \eta, \quad (22)$$

- (2) For all  $\varepsilon, \eta > 0$ , there exist  $0 < \delta < 1$  and  $n_0 \in \mathbb{N}$  such that

$$\sup_{n \geq n_0} P_n(\{x : w'_x(\delta) \geq \varepsilon\}) \leq \eta. \quad (23)$$

We derive from the latter theorem our own criteria for convergence:



**Proposition 2.2** (Convergence Criterion). *Let  $X_1, X_2, \dots$  be a sequence of random variables in  $(D[0, 1], d_{\text{SK}})$ . Suppose that for every  $K \in \mathbb{N}$ , there exist random càdlàg step functions  $X_1^K, X_2^K, \dots$  such that their jumps are all contained in the set  $\{U_\phi \mid |\phi| < K\}$ . If*

- (1) *For all  $\alpha_1, \dots, \alpha_r \in [0, 1]$ , the marginals  $\mathcal{L}(X_n(\alpha_1), \dots, X_n(\alpha_r))$  converge weakly to some distribution  $\mu_{\alpha_1, \dots, \alpha_r}$ , and*
- (2) *For all  $\varepsilon > 0$ ,*

$$\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(\|X_n - X_n^K\|_\infty > \varepsilon) = 0,$$

*then  $X_n$  converges weakly to some random process  $X$  on  $(D[0, 1], d_{\text{SK}})$  as  $n \rightarrow \infty$ , and for  $\alpha_1, \dots, \alpha_r \in [0, 1]$ , we have*

$$(X(\alpha_1), \dots, X(\alpha_r)) \sim \mu_{\alpha_1, \dots, \alpha_r}. \quad (24)$$

*Proof.* We first show that the set of distributions  $\{\mathbb{P}_{X_1}, \mathbb{P}_{X_2}, \dots\}$  is tight using Theorem 2.1. Observe that since  $I_\phi > 0$  a.s. for all  $\phi \in \{0, 1\}^*$ , there exists for every  $K$  a random variable  $\delta_K$  such that all points in  $\{U_\phi \mid |\phi| < K\}$  are at least  $\delta_K$  away from 0, 1 and each other. Hence, for every  $\delta \leq \delta_K$ , the points  $\{0, 1\} \cup \{U_\phi \mid |\phi| < K\}$  are a valid choice for the grid  $\{t_0 < \dots < t_r\}$  in (21). But since the processes  $X_n^K$  are constant between these points,

$$w'_{X_n}(\delta) \leq \max_{0 < i \leq r} \sup_{t_{i-1} \leq s, u < t_i} |X_n(s) - X_n(u)| \leq 2 \|X_n - X_n^K\|_\infty.$$

To show (23), let  $\varepsilon > 0$ . Then, for every  $n, K \in \mathbb{N}$  and  $\delta > 0$ ,

$$\mathbb{P}(w'_{X_n}(\delta) > \varepsilon) \leq \mathbb{P}(\delta_K < \delta) + \mathbb{P}(\|X_n - X_n^K\|_\infty > \frac{\varepsilon}{2}).$$

For  $n$  and  $K$  sufficiently large, the last term is arbitrarily small; by then choosing  $\delta$  small, the first term is also small, showing (23). We now can use (23) to show (22): Let  $r \in \mathbb{N}$ . Every subinterval with length at least  $2^{-r}$  contains a dyadic number  $j2^{-r}$ ,  $j \in \mathbb{N}$ . Therefore,

$$\|X_n\|_\infty \leq w'_{X_n}(2^{-r}) + \max_{j=0, \dots, 2^r} |X_n(j2^{-r})|.$$

By (23), the first term fulfills

$$\mathbb{P}(w'_{X_n}(2^{-r}) > 1) < \eta/2 \quad (25)$$

for sufficiently large  $r$  and  $n$ . By our first condition,  $(X_n(j2^{-r}) \mid 0 \leq j \leq 2^r)$  converges weakly for  $n \rightarrow \infty$ , so it is a tight family on  $\mathbb{R}^{2^r}$ . Therefore there exists a  $\rho \in \mathbb{R}$  such that

$$\mathbb{P}\left(\max_{j=0, \dots, 2^r} |X_n(j2^{-r})| > \rho\right) < \eta/2. \quad (26)$$

Combining (25) and (26), we get (22) for all  $n$  bigger than some  $n_0$ , by possibly further enlarging  $\rho$ , this also holds for  $n < n_0$ .

By Theorem 2.1  $\{\mathbb{P}_{X_1}, \mathbb{P}_{X_2}, \dots\}$  is tight, so  $X_1, X_2, \dots$  has a subsequence converging in distribution to some random variable  $X$ . By Theorem 13.1 in [2] and the discussion preceding it, condition (1) implies that  $X$  is the unique limit and that

$$X(\alpha_1), \dots, X(\alpha_r) \sim \mu_{\alpha_1, \dots, \alpha_r}$$

for  $\alpha_1, \dots, \alpha_r \in [0, 1]$  such that  $X$  is a.s. continuous in  $\alpha_1, \dots, \alpha_r$ . For (24), it remains to show that for all  $a \in [0, 1]$  the function  $X$  is a.s. continuous in  $a$ .

On  $D[0, 1]$ , the function  $x \mapsto x(a) - x(a-)$  has discontinuities, but we may approximate it with

$$J_{\delta, a}(x) := \frac{1}{\delta} \int_0^\delta x(a+s) ds - \frac{1}{\delta} \int_{-\delta}^0 x(a+s) ds \longrightarrow x(a) - x(a-) \quad (\delta \downarrow 0).$$

Since Skorokhod convergence implies pointwise convergence almost everywhere and convergence of the maximum, we obtain that  $J_{\delta, a}$  is a continuous function.

We show that  $J_{\delta,a}(X) \rightarrow 0$  for  $\delta \rightarrow 0$  a.s. Let  $\varepsilon > 0$ , then by weak convergence

$$\lim_{\delta \rightarrow 0} \mathbb{P}(J_{\delta,a}(X) > \varepsilon) \leq \lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \mathbb{P}(J_{\delta,a}(X_n) > \varepsilon).$$

For any  $K \in \mathbb{N}$ , we have  $J_{\delta,a} \leq \|X_n - X_n^K\|_\infty$  if there is no  $U_\phi$ ,  $|\phi| < K$  closer than  $\delta$  to  $a$ , so

$$\mathbb{P}(J_{\delta,a}(X_n) > \varepsilon) \leq \inf_{K \in \mathbb{N}} \mathbb{P}(\|X_n - X_n^K\|_\infty > \varepsilon) + \mathbb{P}(\exists \phi: |\phi| < K, |U_\phi - a| < \delta).$$

The first term is small for  $n, K$  sufficiently large, the second term is small for  $\delta$  sufficiently small. With this,  $X$  is a.s. continuous in  $a$ . □

**2.2. Perturbation of the data.** QuickVal splits an interval  $[L_\phi, R_\phi)$  by the first value falling into  $[L_\phi, R_\phi)$  denoted by  $U_{\tau_\phi}$ . Obviously, this implies dependencies between the data  $U_i$  and the lengths  $I_\phi$  of the intervals  $[L_\phi, R_\phi)$ . In the present section we construct a perturbed sequence  $(\tilde{U}_i)_{i \in \mathbb{N}}$  to the data  $(U_i)_{i \in \mathbb{N}}$  such that we gain independence of  $(\tilde{U}_i)_{i \in \mathbb{N}}$  from the  $\sigma$ -algebra  $\mathcal{F}_\infty$  of the interval lengths defined in (11). In particular, we aim that conditional on  $\mathcal{F}_\infty$  the number of data  $(\tilde{U}_1, \dots, \tilde{U}_n)$  falling into an interval  $[L_\phi, R_\phi)$  is binomial  $B(n, I_\phi)$  distributed, see Lemma 2.4 below.

Every value  $U_i$ ,  $i \in \mathbb{N}$ , falls successively into subintervals generated by QuickVal until becoming a pivot element. These subintervals correspond to the path between the root of the corresponding binary search tree and the node where  $U_i$  is inserted. Let  $\phi_i \in \{0, 1\}^*$  denote the node where  $U_i$  is inserted. Hence, we have  $\tau_{\phi_i} = i$  and  $U_i = L_{\phi_i} + I_{\phi_i}0$ .

Let  $(V_i)_{i \in \mathbb{N}}$  be a sequence of i.i.d.  $\text{unif}[0, 1]$  random variables being independent of  $(U_i)_{i \in \mathbb{N}}$ . We define

$$\tilde{U}_i := L_{\phi_i} + I_{\phi_i} V_i. \tag{27}$$

**Lemma 2.3.** *The sequence  $(\tilde{U}_i)_{i \in \mathbb{N}}$  defined in (27) consists of i.i.d.  $\text{unif}[0, 1]$  distributed random variables and is independent of  $\mathcal{F}_\infty$ .*

*Proof.* It suffices to show that  $\tilde{U}_i$  conditional on  $\mathcal{F}_\infty$  and  $\tilde{U}_1, \dots, \tilde{U}_{i-1}$  is uniformly distributed on  $[0, 1]$  for all  $i \in \mathbb{N}$ . We use infinitesimal notation to denote this claim by

$$\mathbb{P}(\tilde{U}_i \in du \mid \mathcal{F}_\infty, \tilde{U}_1, \dots, \tilde{U}_{i-1}) = \mathbf{1}_{[0,1]}(u) du, \quad i \in \mathbb{N}.$$

For each  $i \in \mathbb{N}$  the random variables  $\tilde{U}_i$  and  $U_i$  fall into the same interval  $[L_{\phi_i}, R_{\phi_i})$ , hence  $\phi_1, \dots, \phi_{i-1}$  are determined by  $\tilde{U}_1, \dots, \tilde{U}_{i-1}$ . Let us additionally condition on  $\phi_i$ , then, by definition,

$$\mathbb{P}(\tilde{U}_i \in du \mid \mathcal{F}_\infty, \tilde{U}_1, \dots, \tilde{U}_{i-1}, \phi_i) = \frac{1}{I_{\phi_i}} \mathbf{1}_{[L_{\phi_i}, R_{\phi_i})}(u) du.$$

Note that  $\phi_i$  denotes one of the  $i$  external nodes of the binary search tree with internal nodes denoted by  $\phi_1, \dots, \phi_{i-1}$ . We denote by  $\text{Ext}_{i-1}$  the set of the labels of these external nodes. Hence, conditional on  $\mathcal{F}_\infty, \phi_1, \dots, \phi_{i-1}$  the label  $\phi_i$  is chosen from  $\text{Ext}_{i-1}$  with probability given by the length of the corresponding interval, i.e.,  $\mathbb{P}(\phi_i = \phi \mid \mathcal{F}_\infty, \phi_1, \dots, \phi_{i-1}) = I_\phi$  for all  $\phi \in \text{Ext}_{i-1}$ . Thus, by the law of total probability we obtain

$$\mathbb{P}(\tilde{U}_i \in du \mid \mathcal{F}_\infty, \tilde{U}_1, \dots, \tilde{U}_{i-1}) = \sum_{\phi \in \text{Ext}_{i-1}} I_\phi \frac{1}{I_\phi} \mathbf{1}_{[L_\phi, R_\phi)}(u) du = \mathbf{1}_{[0,1]}(u) du.$$

This implies the assertion. □

The  $\tilde{U}_i$  are now coupled with the  $U_i$  but independent of the  $I_\phi$ . To compare with the number of key comparisons required by QuickVal( $(U_1, \dots, U_n), \alpha$ ) we define

$$\tilde{S}_{\alpha,k,n} := \sum_{i=1}^n \mathbf{1}_{[L_{\alpha,k}, R_{\alpha,k})}(\tilde{U}_i).$$

**Lemma 2.4.** *Conditional on  $I_{\alpha,k}$  we have that  $\tilde{S}_{\alpha,k,n}$  has the binomial  $B(n, I_{\alpha,k})$  distribution. Moreover, for all  $\alpha \in [0, 1]$ ,  $n \in \mathbb{N}$  and  $0 \leq k \leq n$  we have*

$$S_{\alpha,k,n} \leq (\tilde{S}_{\alpha,k,n} - 1)^+ \leq S_{\alpha,k,n} + k - 1. \quad (28)$$

*Proof.* The conditional distribution of  $\tilde{S}_{\alpha,k,n}$  follows from Lemma 2.3. Recall that  $S_{\alpha,k,n}$  is defined as  $\sum_{i=\tau_{\alpha,k}}^n \mathbf{1}_{\{L_{\alpha,k-1} \leq U_i < R_{\alpha,k-1}\}}$ . By definition,  $U_i$  and  $\tilde{U}_i$  are in the interval  $[L_{\phi_i}, R_{\phi_i})$  for all  $i \in \mathbb{N}$ . If  $U_i \in (L_{\alpha,k}, R_{\alpha,k})$ , then  $U_i$  appears as a pivot after the  $k$ -th pivot. Hence, its interval  $[L_{\phi_i}, R_{\phi_i})$  and thus also  $\tilde{U}_i$  are contained in  $(L_{\alpha,k}, R_{\alpha,k})$ . The  $k$ -th pivot  $U_{\tau_{\alpha,k}}$  itself does not contribute to  $S_{\alpha,k,n}$ , which implies the left inequality stated in the present lemma.

For the right inequality, assume for some  $i \in \mathbb{N}$  that the perturbed value  $\tilde{U}_i$  is in  $(L_{\alpha,k}, R_{\alpha,k})$ , but  $U_i$  is not. Then the corresponding interval  $(L_{\phi_i}, R_{\phi_i})$  must contain  $(L_{\alpha,k}, R_{\alpha,k})$ , thus making  $U_i$  a pivot that appears before the  $k$ -th pivot. Since there are only  $k$  such pivots, the right inequality follows.  $\square$

**2.3. Proof of Theorem 1.1.** To split the contributions to the process  $G_n$  into costs resulting from above and below a level  $K \in \mathbb{N}$  we define

$$G_{\alpha,k,n} := \frac{S_{\alpha,k,n} - nI_{\alpha,k}}{\sqrt{n}} \quad (29)$$

as the normalized fluctuations of the contribution at level  $k$ , and set

$$G_{\alpha,n}^{\leq K} := \sum_{k=0}^K G_{\alpha,k,n}, \quad G_n^{\leq K} := (G_{\alpha,n}^{\leq K})_{\alpha \in [0,1]}, \quad G_{\alpha,n}^{>K} := \sum_{k=K+1}^{\infty} G_{\alpha,k,n}. \quad (30)$$

Hence,  $G_{\alpha,n} = G_{\alpha,n}^{\leq K} + G_{\alpha,n}^{>K}$ . Analogously, for the perturbed values  $\tilde{S}_{k,n}$  we define

$$W_{\alpha,k,n} := \frac{\tilde{S}_{\alpha,k,n} - nI_{\alpha,k}}{\sqrt{n}}, \quad W_{\alpha,n}^{\leq K} := \sum_{k=0}^K W_{\alpha,k,n}, \quad W_n^{\leq K} := (W_{\alpha,n}^{\leq K})_{\alpha \in [0,1]}. \quad (31)$$

**Lemma 2.5.** *For all  $K \in \mathbb{N}$  we have convergence in distribution of  $(G_n^{\leq K})_{n \in \mathbb{N}}$  towards a mixture  $G_\infty^{\leq K} = (G_{\alpha,\infty}^{\leq K})_{\alpha \in [0,1]}$  of centred Gaussian processes within  $\|\cdot\|_\infty$ . Conditional on  $\mathcal{F}_\infty$ , the limit  $G_\infty^{\leq K}$  is a centred Gaussian process with covariance function given, for  $\alpha, \beta \in [0, 1]$  by*

$$\text{Cov}(G_{\alpha,\infty}^{\leq K}, G_{\beta,\infty}^{\leq K} | \mathcal{F}_\infty) = \sum_{k=0}^{K \wedge J} \sum_{j=0}^K I_{\alpha,j \vee k} + (1 + (K \wedge J)) \sum_{j=J+1}^K I_{\beta,j} - S_\alpha^{\leq K} S_\beta^{\leq K}, \quad (32)$$

where  $J = J(\alpha, \beta)$  is as in Theorem 1.1 and  $S_\alpha^{\leq K} := \sum_{k=0}^K I_{\alpha,k}$ . The stated convergence in distribution also holds conditionally in  $\mathcal{F}_\infty$ , i.e., we have almost surely that  $\mathcal{L}(G_n^{\leq K} | \mathcal{F}_\infty)$  converges weakly towards  $\mathcal{L}(G_\infty^{\leq K} | \mathcal{F}_\infty)$ .

*Proof.* First note that by Lemma 2.4 we have  $\|G_n^{\leq K} - W_n^{\leq K}\|_\infty < K^2/\sqrt{n}$ , so it suffices to show the lemma for  $W_n^{\leq K}$ . Conditional on  $\mathcal{F}_\infty$ , the value of  $W_{\alpha,n}^{\leq K}$  is given by

$$W_{\alpha,n}^{\leq K} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{k=0}^K \mathbf{1}_{\{L_{\alpha,k} \leq \tilde{U}_i < R_{\alpha,k}\}} - (R_{\alpha,k} - L_{\alpha,k}), \quad (33)$$

thus the  $2^K$  different values of the process  $W_n^{\leq K}$  can be expressed as the sum of  $n$  centred, bounded i.i.d. random vectors, scaled by  $1/\sqrt{n}$ . By the multivariate central limit theorem, these converge towards a multivariate, centred normal variable. As the positions of the

jumps, still conditional on  $\mathcal{F}_\infty$ , are fixed, we have convergence of  $W_n^{\leq K}$  and thus also of  $G_n^{\leq K}$  towards a Gaussian process. Define  $X_{\alpha,k} := \mathbf{1}\{L_{\alpha,k-1} \leq \tilde{U}_1 < R_{\alpha,k-1}\}$ . The covariance function then is given by

$$\begin{aligned} \text{Cov}(G_{\alpha,\infty}^{\leq K}, G_{\beta,\infty}^{\leq K} | \mathcal{F}_\infty) &= \text{Cov}\left(\sum_{k=0}^K X_{\alpha,k}, \sum_{j=0}^K X_{\beta,j} \mid \mathcal{F}_\infty\right) \\ &= \sum_{k=0}^K \sum_{j=0}^K \mathbb{E}[X_{\alpha,k} X_{\beta,j} | \mathcal{F}_\infty] - S_\alpha^{\leq K} S_\beta^{\leq K} \\ &= \sum_{k=0}^{K \wedge J} \sum_{j=0}^K I_{\alpha,k \vee j} + (1 + (K \wedge J)) \sum_{j=J+1}^K I_{\beta,j} - S_\alpha^{\leq K} S_\beta^{\leq K}. \end{aligned} \quad (34)$$

The assertion follows.  $\square$

To see that the covariance functions in (34) converge towards the covariance function of  $G_\infty$  stated in Theorem 1.1 we restate a Lemma of Grübel and Rösler [11, Lemma 1] that the maximal length of the intervals at a level is decreasing geometrically with increasing levels. It is obtained observing that  $\mathbb{E}[\sum_{|\phi|=k} I_\phi^2] = \left(\frac{2}{3}\right)^k$  and states:

**Lemma 2.6.** *There exists an a.s. finite random variable  $K_1$  such that for all  $k \geq K_1$ :*

$$\max_{\alpha \in [0,1]} I_{\alpha,k} \leq k \left(\frac{2}{3}\right)^{k/2}. \quad (35)$$

Lemma 2.6 implies that the covariance functions of  $G_\infty^{\leq K}$  from (32) converge a.s. to the covariance function of  $G_\infty$  from (8).

For the costs from levels  $k > K$  we find:

**Proposition 2.7.** *For all  $\varepsilon, \eta > 0$  there are constants  $K, N \in \mathbb{N}$  such that for all  $n \geq N$*

$$\mathbb{P}(\|G_n^{>K}\|_\infty > \eta) < \varepsilon. \quad (36)$$

We postpone the proof of the above Proposition 2.7 and first use it and Lemma 2.5 to show convergence of the finite-dimensional distributions, denoted fdd-convergence.

**Lemma 2.8.** *We have fdd-convergence of  $G_n$  towards  $G_\infty$ .*

*Proof.* For any  $K$ , we can split  $G_n = G_n^{\leq K} + G_n^{>K}$ . By Lemma 2.5, we have

$$G_n^{\leq K} \xrightarrow{\text{fdd}} G_\infty^{\leq K} \quad (n \rightarrow \infty).$$

Furthermore, because the covariance functions of the  $G_\infty^{\leq K}$  converge a.s., we obtain

$$G_\infty^{\leq K} \xrightarrow{\text{fdd}} G_\infty \quad (K \rightarrow \infty)$$

by Lévy's continuity theorem. Hence, for all  $\alpha_1, \dots, \alpha_\ell \in [0, 1]$  and all  $t_1, \dots, t_\ell \in \mathbb{R}$  we find a sequence  $(K_n)_{n \in \mathbb{N}}$  in  $\mathbb{N}$  such that

$$\mathbb{P}(G_{\alpha_1, n}^{\leq K_n} < t_1, \dots, G_{\alpha_\ell, n}^{\leq K_n} < t_\ell) \longrightarrow \mathbb{P}(G_{\alpha_1, \infty} < t_1, \dots, G_{\alpha_\ell, \infty} < t_\ell) \quad (n \rightarrow \infty).$$

Now, since  $\|G_n^{>K_n}\|_\infty \rightarrow 0$  in probability by Proposition 2.7 the claim of Lemma 2.8 follows by Slutsky's theorem.  $\square$

To prepare for the proof of Proposition 2.7, we show that the fluctuations on each level are also at least geometrically decreasing. Recall  $K_1$  from Lemma 2.6.

**Lemma 2.9.** For  $a > 1$  sufficiently small, there exist constants  $b > 1$  and  $C_b > 0$  such that for all  $k, n \in \mathbb{N}$

$$\mathbb{P}\left(\max_{\alpha \in [0,1]} |W_{\alpha,k,n}| > a^{-k}, K_1 \leq k\right) \leq C_b \exp(-b^k) + c(k, n),$$

where

$$c(k, n) := 2^{k+1} \exp\left(-\frac{1}{2} a^{-k} \sqrt{n}\right)$$

In particular, we can choose  $a$  so small that for  $n \rightarrow \infty$

$$\sum_{k=1}^{\lfloor (9/2) \log n \rfloor} c(k, n) \rightarrow 0. \quad (37)$$

For the proof of Lemma 2.9 we require the following Chernoff bound:

**Lemma 2.10.** Let  $S_n$  be binomial  $B(n, p)$  distributed for some  $p \in [0, 1]$  and  $n \in \mathbb{N}$  and let  $\mu := \mathbb{E}[S_n]$ ,  $\varepsilon \geq 0$ . Then

$$\mathbb{P}\left(S_n \notin ((1 - \varepsilon)\mu, (1 + \varepsilon)\mu)\right) \leq 2 \exp\left(-\frac{\varepsilon^2 \mu}{2 + \varepsilon}\right).$$

*Proof.* Combine upper and lower bound in McDiarmid [21, Theorem 2.3].  $\square$

*Proof of Lemma 2.9.* Fix some  $\alpha \in [0, 1]$ . Conditionally on  $I_{\alpha,k}$ , the costs  $\tilde{S}_{\alpha,k,n}$  are  $B(n, I_{\alpha,k})$ -distributed by Lemma 2.3. The Chernoff bound in Lemma 2.10 implies

$$\begin{aligned} \mathbb{P}(|W_{\alpha,k,n}| > a^{-k} \mid I_{\alpha,k}) &= \mathbb{P}\left(|\tilde{S}_{\alpha,k,n} - nI_{\alpha,k}| > \sqrt{n}a^{-k} \mid I_{\alpha,k}\right) \\ &\leq 2 \exp\left(-\frac{na^{-2k}}{nI_{\alpha,k}(2 + \sqrt{n}a^{-k}/(nI_{\alpha,k}))}\right) \\ &= 2 \exp\left(-\left(2a^{2k}I_{\alpha,k} + a^k/\sqrt{n}\right)^{-1}\right). \end{aligned} \quad (38)$$

Because the function  $x \mapsto \exp(-x^{-1})$  is monotonically increasing and a sum is smaller than twice the maximum,

$$\begin{aligned} \mathbb{P}(|W_{\alpha,k,n}| > a^{-k} \mid I_{\alpha,k}) &\leq 2 \exp\left(-\left(4a^{2k}I_{\alpha,k} \vee 2a^k/\sqrt{n}\right)^{-1}\right) \\ &= 2 \exp\left(-\left(4a^{2k}I_{\alpha,k}\right)^{-1}\right) \vee 2 \exp\left(-\frac{1}{2}a^{-k}\sqrt{n}\right). \end{aligned} \quad (39)$$

To bound  $W_{\alpha,k,n}$  for all  $\alpha, k$ , we sum up over all  $\phi \in \{0, 1\}^k$ ,

$$\mathbb{P}\left(\max_{\alpha \in [0,1]} |W_{\alpha,k,n}| > a^{-k} \mid I_{\alpha,k}\right) \leq 2^{k+1} \exp\left(-\frac{1}{4}a^{-2k}I_{\alpha,k}^{-1}\right) + 2^{k+1} \exp\left(-\frac{1}{2}a^{-k}\sqrt{n}\right) \quad (40)$$

When furthermore  $k \geq K_1$ , we have  $I_{\alpha,k} < k(\frac{2}{3})^{k/2}$ , so

$$\mathbb{P}\left(\max_{\alpha \in [0,1]} |W_{\alpha,k,n}| > a^{-k}, K_1 \leq k\right) \leq 2^{k+1} \exp\left(-\frac{1}{4}a^{-2k}k^{-1}\left(\frac{3}{2}\right)^{k/2}\right) + 2^{k+1} \exp\left(-\frac{1}{2}a^{-k}\sqrt{n}\right). \quad (41)$$

For  $a^4 < 1.5$ , the term  $a^{-2k}(\frac{3}{2})^{k/2}$  is  $\Omega(b^k)$  for some  $b > 1$ , and we can choose the constant  $C_b > 0$  appropriately.  $\square$

We are now prepared for the proof of Proposition 2.7.

*Proof of Proposition 2.7.* Let  $\varepsilon, \eta > 0$ . To  $K_1$  from Lemma 2.6 and  $a, b$  and  $C_b$  from Lemma 2.9 we choose  $K$  sufficiently large such that

$$\mathbb{P}(K_1 > K) \leq \frac{\varepsilon}{4}, \quad \sum_{k=K}^{\infty} a^{-k} \leq \frac{\eta}{4} \quad \text{and} \quad \sum_{k=K}^{\infty} C_b \exp(-b^k) \leq \frac{\varepsilon}{4}. \quad (42)$$

Let  $H_n$  be the maximum amount of steps needed by  $\text{QuickVal}((U_1, \dots, U_n), \alpha)$  for any  $\alpha$ . Thus,  $H_n$  is also the height of the binary search tree built from  $U_1, \dots, U_n$ . Devroye [5] showed that the height has expectation  $\mathbb{E}[H_n] = \gamma \log n + o(\log n)$  with  $\gamma = 4.311 \dots$  Reed [22] and Drmota [6, 7] further showed that  $\text{Var}(H_n) = O(1)$ . Hence, we can choose  $N$  sufficiently large such that

$$\mathbb{P}(H_n > \lfloor (9/2) \log n \rfloor) < \frac{\varepsilon}{4}, \quad \sum_{k=1}^{\lfloor (9/2) \log n \rfloor} c(k, n) \leq \frac{\varepsilon}{4}, \quad \text{and} \quad \frac{((9/2) \log n)^2}{\sqrt{n}} \leq \frac{\eta}{4} \quad (43)$$

for all  $n \geq N$ . Subsequently we use the decomposition

$$G_{\alpha, n}^{>K} = \sum_{k=K+1}^{\lfloor (9/2) \log n \rfloor} \frac{S_{\alpha, k, n} - nI_{\alpha, k}}{\sqrt{n}} + \sum_{\lfloor (9/2) \log n \rfloor + 1}^{\infty} \frac{S_{\alpha, k, n} - nI_{\alpha, k}}{\sqrt{n}} =: \Gamma_n + G_{\alpha, n}^{>\lfloor (9/2) \log n \rfloor}$$

and consider the event

$$A_n := \{K_1 > K\} \cup \{H_n > \lfloor (9/2) \log n \rfloor\}.$$

We have  $\mathbb{P}(A_n) < \varepsilon/2$  for all  $n \geq N$ . Note that on  $A_n^c$  (the complement of  $A_n$ ) we have  $S_{\alpha, k, n} = 0$  for all  $k > \lfloor (9/2) \log n \rfloor$  and also the bound on  $I_{\alpha, k}$  from Lemma 2.6 applies, hence

$$\begin{aligned} |G_{\alpha, n}^{>\lfloor (9/2) \log n \rfloor}| &\leq \sum_{\lfloor (9/2) \log n \rfloor + 1}^{\infty} \sqrt{n} I_{\alpha, k} \leq \sum_{\lfloor (9/2) \log n \rfloor + 1}^{\infty} \sqrt{n} k \left(\frac{2}{3}\right)^{k/2} \\ &= O\left(n^{1/2 - (9/4) \log(3/2)} \log n\right) = o(1) \end{aligned}$$

since  $(9/4) \log(3/2) = 0.912 \dots$ . Hence, we can enlarge  $N$  so that on  $A_n^c$  we have  $|G_{\alpha, n}^{>\lfloor (9/2) \log n \rfloor}| < \eta/2$  for all  $n \geq N$ . This implies the bound

$$\begin{aligned} \mathbb{P}(|G_{\alpha, n}^{>K}| > \eta) &\leq \mathbb{P}(A_n) + \mathbb{P}(\{|G_{\alpha, n}^{>K}| > \eta\} \cap A_n^c) \\ &\leq \frac{\varepsilon}{2} + \mathbb{P}\left(\left\{|\Gamma_n| > \frac{\eta}{2}\right\} \cap A_n^c\right) + \mathbb{P}\left(\left\{|G_{\alpha, n}^{>\lfloor (9/2) \log n \rfloor}| > \frac{\eta}{2}\right\} \cap A_n^c\right). \end{aligned} \quad (44)$$

Note that the third summand in (44) is 0. Hence, it remains to bound the second summand in (44). To this end note that

$$\begin{aligned} |\Gamma_n| &\leq \sup_{\alpha \in [0, 1]} \sum_{k=K+1}^{\lfloor (9/2) \log n \rfloor} \left( \left| \frac{\tilde{S}_{\alpha, k, n} - nI_{\alpha, k}}{\sqrt{n}} \right| + \left| \frac{S_{\alpha, k, n} - \tilde{S}_{\alpha, k, n}}{\sqrt{n}} \right| \right) \\ &\leq \left( \sum_{k=K+1}^{\lfloor (9/2) \log n \rfloor} \max_{\alpha \in [0, 1]} |W_{\alpha, k, n}| \right) + \frac{\lfloor (9/2) \log n \rfloor^2}{\sqrt{n}}, \end{aligned} \quad (45)$$

where Lemma 2.4 is used. The third relation in (43) assures that the second term in (45) is smaller than  $\eta/4$ . In view of the second relation in (42) and (45), we have

$$\left\{|\Gamma_n| > \frac{\eta}{2}\right\} \cap A_n^c \subset \bigcup_{k=K}^{\lfloor (9/2) \log n \rfloor} \left\{ \max_{\alpha \in [0, 1]} |W_{\alpha, k, n}| > a^{-k}, K_1 \leq k \right\}.$$

Thus, Lemma 2.9 together with (42) and (43) imply that the second summand in (44) is bounded by  $\varepsilon/2$ . This implies the assertion.  $\square$

*Proof of Theorem 1.1.* We apply Proposition 2.2 to  $G_n$  and  $G_n^{\leq K}$ . The first condition, fdd convergence, is Lemma 2.8, the second condition is Proposition 2.7. The representation of the limit process stated in Theorem 1.1 is the subject of Section 2.4.  $\square$

We now transfer the fluctuation result for QuickVal in Theorem 1.1 to the original Quickselect process in Corollary 1.5.

*Proof of Corollary 1.5.* Let  $\tilde{F}_n$  be the inverse of  $\Lambda_n$  in the statement of Corollary 1.5. By definition of  $\Lambda_n$ , the value of the element  $U_{(k)}$  of rank  $k$  within  $U_1, \dots, U_n$  is given by  $\frac{k}{n+1}$ , so

$$\lfloor (n+1)\tilde{F}_n(\alpha) \rfloor = |\{U_i \leq \alpha \mid 1 \leq i \leq n\}| \quad (46)$$

for all  $\alpha \in [0, 1]$ . Thus,  $C_n(\lfloor (n+1)\tilde{F}_n(\alpha) \rfloor) = S_{\alpha, n}$  a.s. for all  $\alpha \in [0, 1]$ , see (6). For  $\alpha = 1$  note that  $\tilde{F}_n(\alpha) = 1$  and  $C_n(n+1) = C_n(n)$  by definition. The Skorokhod distance  $d_{\text{SK}}$  is then bounded by

$$\begin{aligned} d_{\text{SK}}\left(G_n, \left(\frac{C_n(\lfloor t(n+1) \rfloor) - nS_{\Lambda_n(t)}}{\sqrt{n}}\right)_{t \in [0, 1]}\right) &= d_{\text{SK}}\left(G_n, \left(\frac{S_{\Lambda_n(t), n} - nS_{\Lambda_n(t)}}{\sqrt{n}}\right)_{t \in [0, 1]}\right) \\ &= d_{\text{SK}}(G_n, (G_{\Lambda_n(t), n})_{t \in [0, 1]}) \\ &\leq \|\tilde{F}_n - \text{id}\|_{\infty}. \end{aligned}$$

By (46),  $\tilde{F}_n$  is close to the empirical distribution function and thus converges a.s. uniformly to the identity  $\text{id}$  by the Glivenko–Cantelli theorem. The statement of Corollary 1.5 then follows from Slutsky's theorem.  $\square$

**2.4. Explicit construction of the limit process.** Note that on a step-wise level, for finitely many  $\phi_1, \dots, \phi_m \in \{0, 1\}^*$  of paths,  $G_{\phi_1, n}, \dots, G_{\phi_m, n}$  converge, conditional on interval sizes, jointly to normally distributed random variables  $Z_{\phi_1}, \dots, Z_{\phi_m}$ , with covariances given by

$$\mathbb{E}[Z_{\phi} Z_{\psi} \mid \mathcal{F}_{\infty}] = |[L_{\phi}, R_{\phi}] \cap [R_{\psi}, L_{\psi}]| = ((R_{\phi} \wedge R_{\psi}) - (L_{\phi} \vee L_{\psi}))^+ \quad (47)$$

for  $\phi, \psi \in \{0, 1\}^*$ , where  $|\cdot|$  denotes the length of an interval. Using the Kolmogorov extension theorem [15, Theorem 6.17], one can construct the entire family

**Lemma 2.11.** *Almost surely, for all but finitely many  $\phi \in \{0, 1\}^*$  we have*

$$|Z_{\phi}| < 2\sqrt{|\phi|I_{\phi}}. \quad (48)$$

*Proof.* Note that  $Z_{\phi}$  has the normal distribution with variance  $I_{\phi} - I_{\phi}^2 > I_{\phi}$ . Let  $Z$  be  $\mathcal{N}(0, 1)$ -distributed. Then

$$\mathbb{P}(|Z_{\phi}| > 2\sqrt{|\phi|I_{\phi}}) \leq \mathbb{P}(Z > 2\sqrt{|\phi|}) \leq \frac{2}{\sqrt{2\pi}} \int_{2\sqrt{|\phi|}}^{\infty} ye^{-y^2/2} dy = \frac{2}{\sqrt{2\pi}} e^{-2|\phi|}. \quad (49)$$

Since  $e^2 > 2$ , the sum of these probabilities over all  $\phi$  remains finite and hence the Borel–Cantelli lemma implies the assertion.  $\square$

With Lemma 2.11 and Lemma 2.6,  $\sum_{\phi} Z_{\phi} \mathbf{1}_{[L_{\phi}, R_{\phi}]}$  converges uniformly a.s. and

$$G_{\infty} \stackrel{d}{=} \sum_{\phi \in \{0, 1\}^*} Z_{\phi} \mathbf{1}_{[L_{\phi}, R_{\phi}]}, \quad (50)$$

since left- and right-hand sides of the latter display have the same covariance function and thus the same finite dimensional distributions (cf. [2, Theorem 13.1]). We can thus use the left- and right-hand sides of (50) as an explicit construction of  $G_{\infty}$  and henceforth assume pointwise equality in (50).

**2.5. Number of swaps.** In this section, we present a proof for Theorem 1.6.

In Hoare's partition algorithm applied to a uniform permutation, the number of swaps is well known to be, conditional on the rank of the pivot element, hypergeometrically distributed, see, e.g., [18, 19, 4]. The parameters of this hypergeometric distribution, in our notation, are  $n - 1$  (population size),  $S_0 + 1$  (number of trials) and  $n - S_0$  (number of successes). Let  $K_{\phi,n}$  denote the number of swaps at the path  $\phi \in \{0, 1\}^*$  when searching within  $n$  elements. On the  $|\phi|$ -th level, the swaps at previous levels may have changed the order of the elements. Let  $I_1, \dots, I_{S_{\phi,n}}$  denote the indices of elements that are compared to the pivot  $U_{\phi}$  in the order in which the procedure *partition* sees them. In this notation, the number of swaps at level  $k$  is given by

$$K_{\phi,n} = |\{U_{I_i} \mid 1 \leq i \leq S_{\phi,n}, U_{I_i} > U_{\phi}\}|. \quad (51)$$

Now we use a variation of the perturbation argument from Section 2.2 as follows. We extend  $I$  by the indices  $I_{S_{\phi,n}+1}, \dots, I_{\tilde{S}_{\phi,n}}$  of the elements that only contribute to  $\tilde{S}_{\phi,n}$  in arbitrary order and let

$$\tilde{K}_{\phi,n} := |\{\tilde{U}_{I_i} \mid 1 \leq i \leq \tilde{S}_{\phi,n}, \tilde{U}_{I_i} > U_{\phi}\}|. \quad (52)$$

By Lemma 2.4,  $S_{\phi,n} \leq \tilde{S}_{\phi,n} \leq S_{\phi,n} + |\phi| + 1$  and we therefore sum over at most  $|\phi| + 1$  more elements compared to  $K_{\phi,n}$ . Thus,  $K_{\phi,n} \leq \tilde{K}_{\phi,n} \leq K_{\phi,n} + |\phi| + 1$ . With (52) also the limit process can be interpreted: We have that  $\tilde{S}_{\phi,n}$  is close to  $nI_{\phi_0}$  and the probability of a value to be bigger than the pivot is  $I_{\phi_1}/I_{\phi}$ , so we expect that  $\tilde{K}_{\phi,n}$  is close to  $n \frac{I_{\phi_0} I_{\phi_1}}{I_{\phi}}$ . The fact that we draw without replacement does asymptotically not matter as  $n \rightarrow \infty$ . Hence, we define the limit process as

$$K_{\phi,\infty} := \frac{I_{\phi_0} I_{\phi_1}}{I_{\phi}}, \quad K_{\alpha,\infty} := \sum_{k=1}^{\infty} K_{\phi(\alpha,k),\infty}. \quad (53)$$

As for comparisons in (31), we define a process  $\overline{W}_n$  based on the perturbed  $\tilde{K}_{\alpha,n}$ :

$$\overline{W}_{\phi,n} := \frac{\tilde{K}_{\phi,n} - nK_{\phi,\infty}}{\sqrt{n}}, \quad \overline{W}_{\alpha,n}^{\leq K} := \sum_{k=0}^K \overline{W}_{\alpha,k,n}, \quad \overline{W}_n^{\leq K} := (\overline{W}_{\alpha,n}^{\leq K})_{\alpha \in [0,1]}. \quad (54)$$

and the contributions per step for  $G^{\text{swap}}$  as

$$\overline{G}_{\phi,n} := \frac{K_{\phi,n} - nK_{\phi,\infty}}{\sqrt{n}}, \quad \overline{G}_{\alpha,n}^{\leq K} := \sum_{k=0}^K \overline{G}_{\phi(\alpha,k),n}, \quad \overline{G}_n^{\leq K} := (\overline{G}_{\alpha,n}^{\leq K})_{\alpha \in [0,1]}. \quad (55)$$

Our result for swaps corresponding to Lemma 2.5 is as follows:

**Lemma 2.12.** *For the number of swaps in Hoare's partition procedure, above some level  $K \in \mathbb{N}$  we have, conditional on  $\mathcal{F}_k$ , weak convergence in  $\|\cdot\|_{\infty}$  of  $\overline{G}_n^{\leq K}$  to the process is given by*

$$\overline{G}_{\alpha,\infty}^{\leq K} := \sum_{|\phi| \leq K} \left( Y_{\phi} \frac{I_{\phi_0} I_{\phi_1}}{I_{\phi}^{3/2}} + Z_{\phi_0} \frac{I_{\phi_1}}{I_{\phi}} + Z_{\phi_1} \frac{I_{\phi_0}}{I_{\phi}} - Z_{\phi} \frac{I_{\phi_0} I_{\phi_1}}{I_{\phi}^2} \right) \mathbf{1}_{[L_{\phi}, R_{\phi}]}(\alpha), \quad \alpha \in [0, 1], \quad (56)$$

which conditioned on  $\mathcal{F}_k$  is a centred Gaussian process. This convergence also holds unconditionally. The convergence also holds jointly with the convergence of the corresponding quantity  $\overline{G}_n^{\leq K}$  for comparisons in Lemma 2.5.

*Proof.* Note that because  $|K_{\phi,n} - \tilde{K}_{\phi,n}| \leq |\phi| + 1$  for every path  $\phi$ , we again have  $\|\overline{G}_{\alpha,n}^{\leq K} - \overline{W}_{\alpha,n}^{\leq K}\|_{\infty} \rightarrow 0$  and it suffices to show convergence of  $\overline{W}_{\alpha,n}^{\leq K}$ . Now condition on  $\{\tilde{S}_{\phi,n} \mid \phi \in$



$\{0, 1\}^*, n \in \mathbb{N}\}$ . Then, the variables  $\tilde{K}_{\phi,n}$  for  $\phi \in \{0, 1\}^*$  are hypergeometrically distributed and independent from each other. By [16] we have conditional convergence

$$\left( \tilde{K}_{\phi,n} - \frac{\tilde{S}_{\phi 0,n} \tilde{S}_{\phi 1,n}}{\tilde{S}_{\phi,n}} \right) \left( \frac{\tilde{S}_{\phi,n}^{1.5}}{\tilde{S}_{\phi 0,n} \tilde{S}_{\phi 1,n}} \right) \xrightarrow{d} Y_{\phi} \quad (57)$$

with  $Y_{\phi} \sim N(0, 1)$ , as long as for all but a finitely many  $n$  both  $\tilde{S}_{\phi 0,n} > 0$  and  $\tilde{S}_{\phi 1,n} > 0$  hold and if  $\tilde{S}_{\phi,n}^{1.5} / \tilde{S}_{\phi 0,n} \tilde{S}_{\phi 1,n} \rightarrow 0$ . These conditions holds almost surely, so (57) also holds almost surely and thus also unconditionally. By the strong law of large numbers and Slutsky's lemma, we obtain

$$\left( \tilde{K}_{\phi,n} - \frac{\tilde{S}_{\phi 0,n} \tilde{S}_{\phi 1,n}}{\tilde{S}_{\phi,n}} \right) \frac{1}{\sqrt{n}} \xrightarrow{d} Y_{\phi} \frac{I_{\phi 0} I_{\phi 1}}{I_{\phi}^{1.5}}. \quad (58)$$

These  $Y_{\phi}$  are independent of each other, the interval sizes and  $(\tilde{S}_{\phi})_{\phi \in \{0,1\}^*}$ . The next component is the influence of variations in  $\tilde{S}_{\phi,n}$ . For this, define the centred sums  $\hat{S}_{\phi,n} := \tilde{S}_{\phi,n} - nI_{\phi}$ . By the law of the iterated logarithm, we have  $\hat{S}_{\phi,n} = O(\sqrt{n \log \log n})$  a.s. We expand

$$\begin{aligned} \frac{\tilde{S}_{\phi 0,n} \tilde{S}_{\phi 1,n}}{\tilde{S}_{\phi,n}} &= \frac{(nI_{\phi 0} + \hat{S}_{\phi 0,n})(nI_{\phi 1} + \hat{S}_{\phi 1,n})}{(nI_{\phi} + \hat{S}_{\phi,n})} \\ &= n \frac{I_{\phi 0} I_{\phi 1}}{I_{\phi}} + \hat{S}_{\phi 0,n} \frac{I_{\phi 1}}{I_{\phi}} + \hat{S}_{\phi 1,n} \frac{I_{\phi 0}}{I_{\phi}} - \hat{S}_{\phi,n} \frac{I_{\phi 0} I_{\phi 1}}{I_{\phi}^2} + o(\sqrt{n}) \end{aligned} \quad (59)$$

almost surely using the Taylor expansion  $\frac{1}{a+x} = \frac{1}{a} - \frac{x}{a^2} + O(x^2)$ . By the multivariate central limit theorem, the tuple  $(n^{-0.5} \hat{S}_{\phi} \mid |\phi| \leq k+1)$  conditioned on  $\mathcal{F}_{\infty}$  converges in distribution to  $(Z_{\phi} \mid |\phi| \leq k+1)$ . Putting (58) and (59) together, we obtain the convergence

$$\overline{W}_{\phi,n} = \sqrt{n} \left( \tilde{K}_{\phi,n} - n \frac{I_{\phi 0} I_{\phi 1}}{I_{\phi}} \right) \xrightarrow{d} Y_{\phi} \frac{I_{\phi 0} I_{\phi 1}}{I_{\phi}^{1.5}} + Z_{\phi 0} \frac{I_{\phi 1}}{I_{\phi}} + Z_{\phi 1} \frac{I_{\phi 0}}{I_{\phi}} - Z_{\phi} \frac{I_{\phi 0} I_{\phi 1}}{I_{\phi}^2} \quad (60)$$

for all  $|\phi| \leq k$  jointly. Since  $\overline{W}_n^{\leq K} = \sum_{|\phi| \leq K} \overline{W}_{\phi,n} \mathbf{1}_{[L_{\phi}, R_{\phi}]}$ , it thus converges towards a mixed centred Gaussian process formed by the right-hand side of (60). The convergence is jointly with the quantity for comparisons since we use the same limits  $Z_{\phi}$ .  $\square$

Since  $\text{Var}(Z_{\phi} \mid \mathcal{F}_{\infty}) = I_{\phi} - I_{\phi}^2$ , the variance of the random variable in (60) is falling geometrically with  $|\phi|$  by Lemma 2.6 and the sum of these terms along some value  $\alpha \in [0, 1]$  converges to a centred normally distributed random variable.  $Z_{\phi}$  and  $Y_{\phi}$  have the same distribution as described in Theorem 1.6, and the sum of the terms in (60) coincides with the definition of our limit process  $G_{\alpha}^{\text{swap}}$  from (13).

For tightness, we need an equivalent to Lemma 2.9:

**Lemma 2.13.** *For the number of swaps in Hoare's partition, there exist constants  $a, b > 1$  and  $C_b > 0$  such that one can decompose for all  $k, n \in \mathbb{N}$*

$$\mathbb{P}(\exists |\phi| = k : |\overline{W}_{\phi,n}| > a^{-k} + n^{-0.5}, K_1 \leq k) \leq C_b \exp(-b^k) + c(k, n) \quad (61)$$

where the double sequence  $c(k, n)$  satisfies

$$\sum_{k=1}^{\lfloor 4.5 \log n \rfloor} c(k, n) \rightarrow 0. \quad (62)$$

*Proof.* We fix some path  $|\phi| = k$  and condition on the interval sizes  $I_{\phi 0}$  and  $I_{\phi 1}$ . Note that the distribution of  $\tilde{K}_{\phi,n}$  is invariant under permutations of the elements, so we can, without loss

of generality, assume that the partition algorithm processes the elements in their original order.

For any  $1 \leq t \leq n$ , consider how many elements the indices  $i$  and  $j$  from Hoare's partition algorithm have encountered before reaching the element  $U_t$ . The number  $\tilde{S}_{\phi 1, t}$  of elements larger than  $U_\phi$  left of  $t$  is binomially  $B(t, I_{\phi 1})$ -distributed and independent of the number  $\tilde{S}_{\phi 0, n-t}$  of elements smaller than  $U_\phi$  right of  $t$ , which is  $B(\tilde{S}_\phi - t, I_{\phi 0})$ -distributed. Now if both quantities are smaller than some  $x \in \mathbb{N}$ , then both indices reach  $U_t$  before swapping  $x$  elements. When on the other hand both  $\tilde{S}_{\phi 1, t}$  and  $\tilde{S}_{\phi 0, n-t}$  are at least  $x$ , then  $i$  and  $j$  will swap at least  $x$  times before reaching  $U_t$ . Hence, for all  $x \in \mathbb{R}$  we can bound, where the symbol  $\leq$  can be replaced by either  $<$  (less than) at all three occurrences or by  $>$  (larger than),

$$\mathbb{P}(\tilde{K}_{\phi, n} \leq x) \leq \mathbb{P}(\tilde{S}_{\phi 1, t} \leq x) \mathbb{P}(\tilde{S}_{\phi 0, n-t} \leq x). \quad (63)$$

Choosing  $t = \lfloor n \frac{I_{\phi 0}}{I_\phi} \rfloor = n - \lceil n \frac{I_{\phi 0}}{I_\phi} \rceil$ , we obtain

$$\begin{aligned} \mathbb{P}\left(|\overline{W}_{\phi, n}| \geq a^{-k} + n^{-0.5}\right) &= \mathbb{P}\left(\left|\tilde{K}_{\phi, n} - n \frac{I_{\phi 0} I_{\phi 1}}{I_\phi}\right| \geq \sqrt{n} a^{-k} + 1\right) \\ &\leq \mathbb{P}\left(\left|\tilde{S}_{\phi 1, t} - n \frac{I_{\phi 0} I_{\phi 1}}{I_\phi}\right| \geq \sqrt{n} a^{-k} + 1\right) \\ &\quad \times \mathbb{P}\left(\left|\tilde{S}_{\phi 0, n-t} - n \frac{I_{\phi 0} I_{\phi 1}}{I_\phi}\right| \geq \sqrt{n} a^{-k} + 1\right). \end{aligned} \quad (64)$$

The latter difference is not centred because of rounding, but at most 1 of from centring. This offsets with the additional  $n^{-0.5}$ , and by Chernoff bounds we get

$$\begin{aligned} \mathbb{P}\left(|\overline{W}_{\phi, n}| \geq a^{-k} + \frac{1}{\sqrt{n}}\right) &\leq \exp\left(-\frac{\lfloor n \frac{I_{\phi 0}}{I_\phi} \rfloor a^{-2k}}{n(2I_{\phi 1} + a^{-k} n^{-0.5})} - \frac{\lceil n \frac{I_{\phi 1}}{I_\phi} \rceil a^{-2k}}{n(2I_{\phi 0} + a^{-k} n^{-0.5})}\right) \\ &\leq \exp\left(-\frac{a^{-2k}}{2I_\phi + a^{-k} n^{-0.5}}\right). \end{aligned} \quad (65)$$

This is the same bound as in (38) in the proof of Lemma 2.9 and the remaining of the present proof can then be done analogously.  $\square$

*Proof of Theorem 1.6.* By the same proof as for Proposition 2.7, we conclude that  $\|G_n^{\text{swap}} - \overline{G}_n^{\leq K}\|_\infty$  converges to 0 in probability. This guarantees tightness, and with the fdd convergence in Lemma 2.12 this implies process convergence.  $\square$

**2.6.  $\varepsilon$ -tame cost functions.** In the present section we present a proof for Theorem 1.8.

First note that the expression in (17) is well-defined in view of the following lemma using  $\varepsilon$ -tameness, cf. [8, Lemma 3.1]:

**Lemma 2.14.** *Let  $[L, R] \subseteq [0, 1]$  be an interval of length  $I = R - L > 0$  and  $u \in [L, R]$ . For  $V$  being  $\text{unif}[0, 1]$  distributed set  $X := \mathbf{1}_{[L, R]}(V) \beta(u, V)$ . Then, for every  $s \in (0, \varepsilon^{-1})$ , uniformly in  $L, R$ , we have*

$$\mathbb{E}[X^s] = I \cdot \mathbb{E}[X^s \mid V \in [L, R]] = O(I^{1-\varepsilon s}).$$

*Proof.* The tail distribution  $F(x) = \mathbb{P}(X \geq x)$  for  $x > 0$  is bounded by both  $\mathbb{P}(V \in [L, R]) = I$  and  $\mathbb{P}(\beta(u, V) > x) \leq C_\beta x^{-\varepsilon^{-1}}$  for some constant  $C$ . These two bounds are equal when  $x =$

$(I/C_\beta)^{-\varepsilon}$ , so

$$\begin{aligned}\mathbb{E}[X^s] &= \int_0^\infty s x^{s-1} F(x) dx \leq I \int_0^{(I/C_\beta)^{-\varepsilon}} s x^{s-1} dx + C_\beta \int_{(I/C_\beta)^{-\varepsilon}}^\infty s x^{s-1-\varepsilon^{-1}} dx \\ &= O(I^{1-s\varepsilon}) + O(I^{-s\varepsilon+1}),\end{aligned}\tag{66}$$

where  $s\varepsilon < 1$  is required for the second integral to exist.  $\square$

We use the same approach as for the other cost measures: First, we show that the costs at some level are close to a sum of independent random variables to which we apply the central limit theorem. Then we show that the residual below some level is converging to 0. Since our cost now depends on the actual values  $U_1, \dots, U_n$  instead of just whether the  $U_i$  are contained in the respective interval or not, we cannot define  $\tilde{S}_{\alpha,k,n}$  as before. Instead, we only replace  $U_i$  with  $\tilde{U}_i$  if it is below some level  $K \in \mathbb{N}$ , defining for  $i \in \mathbb{N}$

$$\tilde{U}_i^{(K)} := \begin{cases} U_i & |\phi_i| > K, \\ \tilde{U}_i & |\phi_i| \leq K \end{cases}\tag{67}$$

and then for  $k \leq K$

$$S_{\alpha,k,n}^\beta := \sum_{i=\tau_{\alpha,k}+1}^n \mathbf{1}_{[L_{\alpha,k}, R_{\alpha,k})}(U_i) \beta(U_{\tau_{\alpha,k}}, U_i)\tag{68}$$

$$\tilde{S}_{\alpha,k,n}^{\beta,K} := \sum_{i=1}^n \mathbf{1}_{[L_{\alpha,k}, R_{\alpha,k})}(\tilde{U}_i^{(K)}) \beta(U_{\tau_{\alpha,k}}, \tilde{U}_i^{(K)})\tag{69}$$

$$I_{\alpha,k}^\beta := \int_0^1 \mathbf{1}_{[L_{\alpha,k}, R_{\alpha,k})}(v) \beta(U_{\tau_{\alpha,k}}, v) dv.\tag{70}$$

Note that, conditional on  $\mathcal{F}_K$ , the values  $\tilde{U}_1^{(K)}, \tilde{U}_2^{(K)}, \dots$  are i.i.d.  $\text{unif}[0,1]$  and hence  $\tilde{S}_{\alpha,k,n}^{(K)}$  is a sum of i.i.d. random variables distributed as  $X_{\alpha,k}^\beta$ , with expectation  $I_{\alpha,k}^\beta$ . With these random variables, we define similar to before

$$G_{\alpha,k,n}^\beta := \frac{S_{\alpha,k,n}^\beta - n I_{\alpha,k,n}^\beta}{\sqrt{n}}, \quad W_{\alpha,k,n}^{\beta,K} := \frac{\tilde{S}_{\alpha,k,n}^{\beta,K} - n I_{\alpha,k,n}^\beta}{\sqrt{n}}.\tag{71}$$

The  $G_{\alpha,k,n}^\beta$  are the single steps within the residual process, and  $W_{\alpha,k,n}^{\beta,K}$  are the normalized sum of i.i.d. random variables used to approximate them.

**Lemma 2.15.** *The residuals  $\sum_{k=0}^K G_{\alpha,k,n}$  above some level  $K \in \mathbb{N}$  converge conditional on  $\mathcal{F}_K$  to a mixture of centred Gaussian processes with random covariance function given by*

$$(\alpha, \gamma) \mapsto \text{Cov}\left(\sum_{k=0}^K X_{\alpha,k}^\beta, \sum_{k=0}^K X_{\gamma,k}^\beta \mid \mathcal{F}_K\right),\tag{72}$$

with  $X_{\alpha,k}^\beta$  as in (16).

*Proof.* Conditioned on  $\mathcal{F}_K$ ,  $\sum_{k=0}^K W_{\alpha,k,n}^{\beta,K}$  is a step function with  $2^K$  different values. These values are each sums of  $n$  independent,  $\sum_{k=0}^K X_{\alpha,k}^\beta$ -distributed variables. By the multivariate central limit theorem  $\sum_{k=0}^K W_{\alpha,k,n}^{\beta,K}$  thus converges to a mixed Gaussian process with covariances (72). The difference between  $\sum_{k=0}^K W_{\alpha,k,n}^{\beta,K}$  and  $\sum_{k=0}^K G_{\alpha,k,n}^\beta$  is at most the different costs of  $U_\phi$  and  $\tilde{U}_\phi$  at paths  $\phi \in \{0, 1\}^*$ ,  $|\phi| \leq k$ . Those are finitely many and are then divided by  $\sqrt{n}$ . By Slutsky's lemma,  $\sum_{k=0}^K G_{\alpha,k,n}^\beta$  hence also converges to the mixed Gaussian process.  $\square$

With the same technique as before (see Lemma 2.8 and the proof of Theorem 1.1), we can prove Theorem 1.8 with the following proposition:

**Proposition 2.16.** *For all  $0 < \varepsilon < \frac{1}{4}$  the residuals below level  $K \in \mathbb{N}$  converge as  $n, K \rightarrow \infty$  to 0:*

$$\sup_{\alpha \in [0,1]} \left| \sum_{k > K} G_{\alpha,k,n} \right| \xrightarrow{\mathbb{P}} 0 \quad (73)$$

This is a consequence of the following lemma:

**Lemma 2.17.** *For  $s \geq 2$ , we have uniformly in  $\alpha, k, n, I_{\alpha,k}$*

$$\mathbb{E} \left[ \left| G_{\alpha,k,n}^\beta \right|^s \middle| L_{\alpha,k}, R_{\alpha,k} \right] = O \left( I_{\alpha,k}^{2-s\varepsilon} n^{2-\frac{s}{2}} + k^{\frac{s}{2}} I_{\alpha,k}^{(\frac{1}{2}-\varepsilon)s} \right). \quad (74)$$

*Proof.* First, let  $s \in 2\mathbb{N}^+$  be even. Then decompose  $G_{\alpha,k,n}^\beta$  into

$$\begin{aligned} (G_{\alpha,k,n}^\beta)^s &\leq n^{-\frac{s}{2}} \left( S_{\alpha,k,n}^\beta - (n - \tau_{\alpha,k})^+ I_{\alpha,k}^\beta - (\tau_{\alpha,k} \wedge n) I_{\alpha,k}^\beta \right)^s \\ &\leq 2^s n^{-\frac{s}{2}} \left( \left( S_{\alpha,k,n}^\beta - (n - \tau_{\alpha,k})^+ I_{\alpha,k}^\beta \right)^s - (\tau_{\alpha,k} \wedge n)^s (I_{\alpha,k}^\beta)^s \right). \end{aligned} \quad (75)$$

The first part is a centred sum of  $(n - \tau_{\alpha,k})^+$  independent,  $X_{\alpha,k}^\beta$ -distributed random variables; the second part arises from  $n - (n - \tau_{\alpha,k})^+ = n \wedge \tau_{\alpha,k}$ . If we condition on  $\tau_k$ , expand the sum in  $S_{\alpha,k,n}^\beta$  and group the resulting terms by the number of occurrences of  $U_i$ , then terms where  $U_i$  appears only once vanish caused by centring, and we have

$$\begin{aligned} &\mathbb{E} \left[ \left( S_{\alpha,k,n}^\beta - (n - \tau_{\alpha,k})^+ I_{\alpha,k}^\beta \right)^s \middle| L_{\alpha,k}, R_{\alpha,k}, \tau_{\alpha,k} \right] \\ &= \mathbb{E} \left[ \left( \sum_{i=\tau_{\alpha,k}+1}^n (\mathbf{1}_{[L_{\alpha,k}, R_{\alpha,k})}(U_i) \beta(U_{\tau_{\alpha,k}}, U_i) - I_{\alpha,k}^\beta) \right)^s \middle| L_{\alpha,k}, R_{\alpha,k}, \tau_{\alpha,k} \right] \\ &= \sum_{l=1}^{s/2} \binom{(n - \tau_{\alpha,k})^+}{l} \sum_{\substack{i_1 + \dots + i_l = s \\ i_1, \dots, i_l \geq 2}} \prod_{j=1}^l \mathbb{E} \left[ (X_{\alpha,k}^\beta - I_{\alpha,k}^\beta)^{i_j} \middle| L_{\alpha,k}, R_{\alpha,k} \right]. \end{aligned} \quad (76)$$

By Lemma 2.14,  $I_{\alpha,k}^\beta = O(I_{\alpha,k}^{1-\varepsilon})$  and  $\mathbb{E}[(X_{\alpha,k}^\beta)^i] = O(I_{\alpha,k}^{1-\varepsilon i})$ , so the last product is of order  $O(I_{\alpha,k}^{l-\varepsilon s})$ , and the sum in (76) is equal to

$$= \sum_{l=1}^{s/2} O(n^l I_{\alpha,k}^{l-\varepsilon s}) = O(n I_{\alpha,k}^{1-\varepsilon s}) + O\left(n^{s/2} I_{\alpha,k}^{(\frac{1}{2}-\varepsilon)s}\right). \quad (77)$$

This holds uniformly over all  $\tau_{\alpha,k}$ . When  $\tau_{\alpha,k} \geq n$ , the latter sum vanishes, so by multiplying with  $\mathbb{P}(\tau_{\alpha,k} \leq n \mid I_{\alpha,k}) \leq 1 \wedge I_{\alpha,k} n$  we obtain

$$n^{-\frac{s}{2}} \mathbb{E}[S_{\alpha,k,n}^\beta - (n - \tau_{\alpha,k})^+ I_{\alpha,k}^\beta \mid L_{\alpha,k}, R_{\alpha,k}] = O\left(I_{\alpha,k}^{2-s\varepsilon} n^{2-\frac{s}{2}} + I_{\alpha,k}^{(\frac{1}{2}-\varepsilon)s}\right), \quad (78)$$

which has the order stated in (74).

To bound the second summand appearing in (75) we use an idea from [20, 9]: The estimate  $\tau_{\alpha,k} \wedge n \leq n^{\frac{1}{2}} \tau_{\alpha,k}^{\frac{1}{2}}$  implies

$$n^{-\frac{s}{2}} (\tau_{\alpha,k} \wedge n)^s (I_{\alpha,k}^\beta)^s = \tau_{\alpha,k}^{\frac{s}{2}} (I_{\alpha,k}^\beta)^s. \quad (79)$$

Conditional on  $\mathcal{F}_k$ ,  $\tau_{\alpha,k}$  is the sum of  $k$  independent geometrically distributed variables with success probabilities  $I_{\alpha,0}, \dots, I_{\alpha,k-1}$  and thus smaller in probability than a sum of  $k$  independent  $\text{Geom}(I_{\alpha,k})$ -distributed random variables  $\Gamma_1, \dots, \Gamma_k$ . Thus

$$\mathbb{E}\left[\tau_{\alpha,k}^{\frac{s}{2}} \mid L_{\alpha,k}, R_{\alpha,k}\right] \leq \mathbb{E}\left[\left(\sum_{j=1}^k \Gamma_j\right)^{\frac{s}{2}} \mid I_{\alpha,k}\right] \leq k^{\frac{s}{2}} \mathbb{E}\left[\Gamma_1^{\frac{s}{2}} \mid I_{\alpha,k}\right] = O(k^{s/2} I_{\alpha,k}^{-s/2}). \quad (80)$$

Together with  $(I_{\alpha,k}^\beta)^s = O(I_{\alpha,k}^{(1-\varepsilon)s})$ , this implies

$$\mathbb{E}\left[n^{-\frac{s}{2}} (\tau_{\alpha,k} \wedge n)^s (I_{\alpha,k}^\beta)^s \mid L_{\alpha,k}, R_{\alpha,k}\right] = O\left(k^{s/2} I_{\alpha,k}^{(\frac{1}{2}-\varepsilon)s}\right). \quad (81)$$

This completes the proof of the present lemma for  $s$  even. The general case  $s \geq 2$  is implied by Lyapunov's interpolation inequality.  $\square$

We are now prepared to prove Proposition 2.16:

*Proof of Proposition 2.16.* For some path  $\phi \in \{0, 1\}^k$  the interval length  $I_\phi$  is distributed as the product of  $k$  i.i.d.  $\text{unif}[0, 1]$  random variables, so

$$\mathbb{E}\left[|G_{\phi,n}^\beta|^s \mid L_\phi, R_\phi\right] = O\left(I_\phi^{2-s\varepsilon} n^{2-\frac{s}{2}} + k^{\frac{s}{2}} I_\phi^{(\frac{1}{2}-\varepsilon)s}\right). \quad (82)$$

Lemma 2.17 hence implies

$$\mathbb{E}\left[|G_{\phi,n}^\beta|^s\right] = O\left(\left(\frac{1}{3-s\varepsilon}\right)^k n^{2-\frac{s}{2}} + k^{\frac{s}{2}} \left(\frac{1}{1+(\frac{1}{2}-\varepsilon)s}\right)^k\right). \quad (83)$$

Summing over all  $\phi$  with  $|\phi| = k$  implies

$$\mathbb{E}\left[\max_{|\phi|=k} |G_{\phi,n}^\beta|^s\right] = O\left(\left(\frac{2}{3-s\varepsilon}\right)^k n^{2-\frac{s}{2}} + k^{\frac{s}{2}} \left(\frac{2}{1+(\frac{1}{2}-\varepsilon)s}\right)^k\right) \quad (84)$$

and the  $L_s$ -norm is thus bounded by

$$\left\|\max_{|\phi|=k} |G_{\phi,n}^\beta|\right\|_s = O\left(\left(\frac{2}{3-s\varepsilon}\right)^{\frac{k}{s}} n^{\frac{2}{s}-\frac{1}{2}} + k^{\frac{1}{2}} \left(\frac{2}{1+(\frac{1}{2}-\varepsilon)s}\right)^{\frac{k}{s}}\right). \quad (85)$$

Summing over  $k \in \mathbb{N}$  yields a convergent series if  $s < \varepsilon^{-1}$ ,  $s \geq 4$ , and  $s \geq \frac{2}{1-2\varepsilon}$ . If  $\varepsilon < \frac{1}{4}$ , then  $s$  can be chosen as 4, and

$$\left\|\max_{|\phi|=k} \left|\sum_{n \geq K} G_{\phi,n}^\beta\right|\right\|_s \leq \sum_{k \geq K} \left\|\max_{|\phi|=k} |G_{\phi,n}^\beta|\right\|_s \rightarrow 0. \quad (86)$$

$\square$

Theorem 1.8 now follows from Proposition 2.16 and Lemma 2.15 as for the previously discussed complexity measures.

**2.7. Path properties of the limit process.** In the present section we prove the claims from Section 1.1.4 on path properties of the limit processes.

**Lemma 2.18.** *The metrics  $d_2$  and  $d_G$  are a.s. equivalent.*

*Proof.* To rewrite (19) in terms of the interval lengths, recall that the covariances can be expressed with a  $\text{unif}[0,1]$  variable  $V$  as in (10),

$$\begin{aligned} d_G(\alpha, \beta)^2 &= \mathbb{E}[(J(\alpha, V) - J(\beta, V))^2 | \mathcal{F}_\infty] - (S_\alpha - S_\beta)^2 \\ &= \sum_{l \geq 1} (2l - 1) \mathbb{P}(|J(\alpha, V) - J(\beta, V)| \geq l | \mathcal{F}_\infty) - (S_\alpha - S_\beta)^2. \end{aligned} \quad (87)$$

When  $J(\alpha, V) \geq l + J(\beta, V)$ , the value  $V$  must have followed the path of  $\alpha$  for at least  $l$  steps more than it has followed the path of  $\beta$ . That happens if and only if  $V$  is in the interval  $I_{\alpha, J(\alpha, \beta) + l}$ . Hence, writing  $J := J(\alpha, \beta)$ ,

$$\begin{aligned} d_G(\alpha, \beta)^2 &= \sum_{k > J} (2(k - J) - 1)(I_{\alpha, k} + I_{\beta, k}) - \left( \sum_{k > J} (I_{\alpha, k} - I_{\beta, k}) \right)^2 \\ &= \sum_{k > J} \left( (I_{\alpha, k} + I_{\beta, k})(2(k - J) - 1) - |I_{\alpha, k} - I_{\beta, k}| \sum_{l > J} |I_{\alpha, l} - I_{\beta, l}| \right). \end{aligned} \quad (88)$$

Since by Lemma 2.6 the sums converge a.s. uniformly for  $J \rightarrow \infty$ , every  $d_G$  ball contains a  $d_2$  ball. Conversely, by reordering the last sum such that  $l \leq k$ , and bounding  $|I_{\alpha, l} - I_{\beta, l}| < 1$ ,

$$\begin{aligned} d_G(\alpha, \beta)^2 &\geq \sum_{k > J} (2(k - J) - 1) [(I_{\alpha, k} + I_{\beta, k}) - |I_{\alpha, k} - I_{\beta, k}|] \\ &= \sum_{k > J} 2(2(k - J) - 1)(I_{\alpha, k} \wedge I_{\beta, k}). \end{aligned} \quad (89)$$

Since for all  $\alpha, \beta$  and  $k$  we have  $I_{\alpha, k}, I_{\beta, k} > 0$  a.s., every  $d_2$  ball thus also contains a  $d_G$  ball and the metrics are a.s. equivalent.  $\square$

To study Hölder continuity we prepare with bounds on the limit processes in terms of interval length of the underlying partition, cf. (1).

**Lemma 2.19.** *Almost surely for all  $\phi \in \{0, 1\}^*$  we have the following bounds:*

- (1)  $G_{\phi, \infty} := Z_\phi = O(\sqrt{|\phi| I_\phi})$ ,
- (2)  $G_{\phi, \infty}^{\text{swap}} := Y_\phi \frac{I_{\phi 0} I_{\phi 1}}{I_\phi^{3/2}} + Z_{\phi 0} \frac{I_{\phi 1}}{I_\phi} + Z_{\phi 1} \frac{I_{\phi 0}}{I_\phi} - Z_\phi \frac{I_{\phi 0} I_{\phi 1}}{I_\phi^2} = O(\sqrt{|\phi| I_\phi})$ ,
- (3)  $G_{\phi, \infty}^\beta = O(\sqrt{|\phi| I_\phi^{1-2\varepsilon}})$ .

In particular, we can subsequently only focus on  $G_{\phi, \infty}^\beta$ , since the Hölder continuity for comparisons and swaps from Theorem 1.9 is then covered by choosing  $\varepsilon = 0$ .

*Proof.* We already showed part (1) concerning  $G_{\phi, \infty}$  in Lemma 2.11. For part (2) note that the factors in front of  $Z_\phi$ ,  $Z_{\phi 0}$  and  $Z_{\phi 1}$  are smaller than 1, so we only require an additional bound for  $Y_\phi$ . For a sequence  $(a_k)$  in  $\mathbb{R}^+$ ,

$$\sum_{|\phi|=k} \mathbb{P}(|Y_\phi| > a_k) \leq \sum_{|\phi|=k} 2 \exp\left(-\frac{a_k^2}{2}\right) = 2 \exp\left(-\frac{a_k^2}{2} + k \log(2)\right). \quad (90)$$

Choosing  $a_k = 2\sqrt{k}$  is sufficient for the sequence of the latter terms to converge, thus  $Y_\phi = O(\sqrt{|\phi|})$  from the Borel–Cantelli lemma. With  $I_{\phi 0} I_{\phi 1} I_\phi^{1.5} \leq \sqrt{I_\phi}$ , this concludes the proof of part (2). For part (3), recall that  $G_{\phi, \infty}^\beta$  is mixed Gaussian with the same variance as  $X_\phi^\beta$ , as defined in (17). By Lemma 2.14 we have  $\mathbb{E}[(X_\phi^\beta)^s] = O(I_\phi^{1-\varepsilon s})$ , hence  $\text{Var}(X_\phi^\beta) = O(I_\phi^{1-2\varepsilon})$ . We may repeat the same argument as for  $Z_\phi$  and  $Y_\phi$  to get  $G_{\phi, \infty}^\beta = O(\sqrt{|\phi| I_\phi^{1-2\varepsilon}})$ .  $\square$

*Proof of Theorem 1.9 for  $d_2$ .* We prove the statement of Theorem 1.9 only for  $G^\beta$  with  $\varepsilon$ -tame  $\beta$ . Note that in view of Lemma 2.6 the same proof applies to  $G$  and  $G^{\text{swap}}$  with  $\varepsilon = 0$ . Lemma 2.19 implies that for any  $\alpha, \beta \in [0, 1]$  we have almost surely

$$|G_{\alpha,\infty}^\beta - G_{\beta,\infty}^\beta| \leq \sum_{k>J} |G_{\alpha,k,\infty}^\beta| + |G_{\beta,k,\infty}^\beta| \leq \sum_{k>J} \sqrt{k} O\left(I_{\alpha,k}^{0.5-\varepsilon} + I_{\beta,k}^{0.5-\varepsilon}\right). \quad (91)$$

By Lemma 2.6, these sizes are further bounded by

$$\begin{aligned} |G_{\alpha,\infty}^\beta - G_{\beta,\infty}^\beta| &\leq \sum_{k>J} k^{1-\varepsilon} O\left(\left(\frac{2}{3}\right)^{(1-2\varepsilon)k/4}\right) = O\left(J^{1-\varepsilon} 2^{-J(\frac{1}{4}\log_2 \frac{3}{2})(1-2\varepsilon)}\right), \\ \log_2 |G_{\alpha,\infty}^\beta - G_{\beta,\infty}^\beta| &\leq (1 + o(1)) \log_2(d_2(\alpha, \beta)) \left(\frac{1}{4}\log_2 \frac{3}{2}\right)(1-2\varepsilon). \end{aligned} \quad (92)$$

This implies  $d_2$ -Hölder continuity with all exponents smaller than  $\frac{1}{4}\log_2 \frac{3}{2}(1-\varepsilon)$ .  $\square$

Since in (88) the interval sizes are decreasing exponentially, we see that  $d_G(\alpha, \beta)^2$  is dominated by the term for  $k = J + 1$ , which is  $I_{\alpha,J}$ . On the other hand  $(G_{\alpha,\infty} - G_{\beta,\infty})^2$  is approximately  $\sum_{k>J} I_{\alpha,k} + I_{\beta,k}$ . So we may bound the intervals  $I_{\alpha,k}, k > J$ , with the length of their ancestor  $I_{\alpha,J}$ , again using a Borel–Cantelli argument.

**Lemma 2.20.** *For any  $\delta \in (0, 1)$  and  $s > \frac{1}{\delta}$  we have almost surely for all but finitely many  $\phi, \psi \in \{0, 1\}^*$  that*

$$\frac{I_{\phi\psi}}{I_\phi^{1-\delta}} \leq |\psi| \left(\frac{2}{1+s}\right)^{|\psi|/s}. \quad (93)$$

*In particular, for  $\gamma > 0$ , the following holds uniformly in  $\alpha$ :*

$$\sum_{k>J} \sqrt{k} I_{\alpha,k}^\gamma = \sum_{k>J} I_{\alpha,k}^{\gamma(1-\delta)} k^{1.5} \left(\frac{2}{1+s}\right)^{\gamma k/s} = O\left(J^{1.5} I_{\alpha,J}^{\gamma(1-\delta)}\right). \quad (94)$$

*Proof.* Note that  $I_{\phi\psi}/I_\phi$  is the product of  $|\psi|$  i.i.d.  $\text{unif}[0, 1]$  random variables and independent of  $I_\phi$ , which is also a product of  $|\phi|$  i.i.d.  $\text{unif}[0, 1]$  random variables. By Markov bounding,

$$\begin{aligned} \mathbb{P}\left(\left(\frac{I_{\phi\psi}}{I_\phi^{1-\delta}}\right)^s \leq |\psi|^s \left(\frac{2}{1+s}\right)^{|\psi|}\right) &\leq \mathbb{E}\left[\left(\frac{I_{\phi\psi}}{I_\phi}\right)^s I_\phi^{s\delta}\right] |\psi|^{-s} \left(\frac{1+s}{2}\right)^{|\psi|} \\ &= \left(\frac{1}{1+s}\right)^{|\psi|} \left(\frac{1}{1+s\delta}\right)^{|\phi|} |\psi|^{-s} \left(\frac{1+s}{2}\right)^{|\psi|} \\ &= 2^{-|\psi|-|\phi|} \left(\frac{2}{1+s\delta}\right)^{|\phi|} |\psi|^{-s}. \end{aligned} \quad (95)$$

Since  $s\delta > 1$  and  $s > 1$ , the sum of these probabilities over all  $\phi$  and  $\psi$  converges. Hence, as in previous arguments, the Borel–Cantelli lemma implies the assertion.  $\square$

*Proof of Theorem 1.9 for  $d_G$ .* With (91) and Lemma 2.20, we obtain

$$|G_{\alpha,\infty} - G_{\beta,\infty}| = O\left(\sum_{k>J} \sqrt{k} (|I_{\alpha,k}|^{0.5-\varepsilon} + |I_{\beta,k}|^{0.5-\varepsilon})\right) = O\left(J^{1.5} I_{\alpha,J}^{(0.5-\varepsilon)(1-\delta)}\right) \quad (96)$$

for every  $\delta \in (0, 1)$  almost surely and uniformly in  $\alpha$  and  $\beta$ . By the same reasoning as used to derive (89) from (88), we can bound  $d_G$  with only the term arising from  $k = l = J + 1$ ,

$$d_G(\alpha, \beta)^2 \geq (I_{\alpha,J+1} + I_{\beta,J+1}) - (I_{\alpha,J+1} - I_{\beta,J+1})^2 \geq I_{\alpha,J} - I_{\alpha,J}^2 \sim I_{\alpha,J}. \quad (97)$$

Now, (96) and (97) imply Hölder continuity with any exponent smaller than  $1 - 2\varepsilon$ .  $\square$

## REFERENCES

- [1] Jon Bentley, *Programming pearls*, Addison-Wesley, Reading, Mass., 2000.
- [2] Patrick Billingsley, *Convergence of probability measures*, Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons Inc., New York, 1999, A Wiley-Interscience Publication. MR MR1700749 (2000e:60008)
- [3] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein, *Introduction to algorithms*, third ed., MIT Press, Cambridge, MA, 2009. MR 2572804
- [4] Benjamin Dadoun and Ralph Neininger, *A statistical view on exchanges in Quickselect*, ANALCO14—Meeting on Analytic Algorithmics and Combinatorics, SIAM, Philadelphia, PA, 2014, pp. 40–51. MR 3248352
- [5] Luc Devroye, *A note on the height of binary search trees*, J. ACM **33** (1986), no. 3, 489–498.
- [6] Michael Drmota, *The variance of the height of binary search trees*, Theoret. Comput. Sci. **270** (2002), no. 1–2, 913–919. MR 1871103
- [7] ———, *An analytic approach to the height of binary search trees. II*, J. ACM **50** (2003), no. 3, 333–374. MR 2146358
- [8] James Allen Fill and Jason Matterer, *QuickSelect tree process convergence, with an application to distributional convergence for the number of symbol comparisons used by worst-case find*, Combin. Probab. Comput. **23** (2014), no. 5, 805–828. MR 3249225
- [9] James Allen Fill and Jason Matterer, *Convergence of the QuickVal Residual*, preprint, arXiv:2412.12599, 2024.
- [10] James Allen Fill and Takehiko Nakama, *Distributional convergence for the number of symbol comparisons used by quickselect*, Advances in Applied Probability **45** (2013), no. 2, 425–450.
- [11] Rudolf Grübel and Uwe Rösler, *Asymptotic distribution theory for Hoare’s selection algorithm*, Adv. in Appl. Probab. **28** (1996), no. 1, 252–269. MR 1372338
- [12] C. A. R. Hoare, *Algorithm 63: Partition*, Commun. ACM **4** (1961), no. 7, 321.
- [13] ———, *Algorithm 65: Find*, Commun. ACM **4** (1961), no. 7, 321–322.
- [14] Jasper Ischebeck and Ralph Neininger, *On fluctuations of complexity measures for the FIND algorithm*, 35th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms, LIPIcs. Leibniz Int. Proc. Inform., vol. 302, Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2024, pp. Art. No. 9, 15. MR 4774657
- [15] Olav Kallenberg, *Foundations of modern probability*, 2 ed., Springer International Publishing, 2002.
- [16] S. N. Lahiri and A. Chatterjee, *A berry-esseen theorem for hypergeometric probabilities under minimal conditions*, Proceedings of the American Mathematical Society **135** (2007), no. 5, 1535–1545.
- [17] Kevin Leckey, Ralph Neininger, and Henning Sulzbach, *Process convergence for the complexity of radix selection on markov sources*, Stochastic Processes and their Applications **129** (2019), no. 2, 507–538.
- [18] Hosam M. Mahmoud, *Sorting*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000, A distribution theory. MR 1784633
- [19] Hosam M. Mahmoud, *Distributional analysis of swaps in quick select*, Theoretical Computer Science **411** (2010), no. 16, 1763–1769.
- [20] Jason Matterer, *Quickselect process and quickval residual convergence*, Ph.D. thesis, The Johns Hopkins University, Baltimore, Maryland, 2015.
- [21] Colin McDiarmid, *Concentration*, Probabilistic Methods for Algorithmic Discrete Mathematics (Michel Habib, Colin McDiarmid, Jorge Ramirez-Alfonsin, and Bruce Reed, eds.), Springer Berlin Heidelberg, Berlin, Heidelberg, 1998, pp. 195–248.
- [22] Bruce Reed, *The height of a random binary search tree*, J. ACM **50** (2003), no. 3, 306–332.
- [23] Henning Sulzbach, Ralph Neininger, and Michael Drmota, *A Gaussian limit process for optimal FIND algorithms*, Electronic Journal of Probability **19** (2014), no. none, 1 – 28.
- [24] Brigitte Vallée, Julien Clément, James Allen Fill, and Philippe Flajolet, *The number of symbol comparisons in QuickSort and QuickSelect*, Automata, languages and programming. Part I, Lecture Notes in Comput. Sci., vol. 5555, Springer, Berlin, 2009, pp. 750–763. MR 2544890

JASPER ISCHEBECK, ischebec@math.uni-frankfurt.de, GOETHE UNIVERSITY FRANKFURT, INSTITUTE OF MATHEMATICS, 60629 FRANKFURT A.M., GERMANY

RALPH NEININGER, neininger@math.uni-frankfurt.de, GOETHE UNIVERSITY FRANKFURT, INSTITUTE OF MATHEMATICS, 60629 FRANKFURT A.M., GERMANY