Combinatorics, Probability and Computing (2014) 23, 805–828.
doi:10.1017/S0963548314000121

QuickSelect Tree Process Convergence, With an Application to Distributional Convergence for the Number of Symbol Comparisons Used by Worst-Case Find

JAMES ALLEN FILL and JASON MATTERER†

Department of Applied Mathematics and Statistics, The Johns Hopkins University,
34th and Charles Streets, Baltimore, MD 21218-2682, USA
(e-mail: `jimfill@jhu.edu`, `jtmatterer@gmail.com`)

Received 20 March 2013; revised 5 August 2013; first published online 9 July 2014

We dedicate this paper to the memory of our colleague and friend Philippe Flajolet.

We define a sequence of tree-indexed processes closely related to the operation of the `QuickSelect` search algorithm (also known as `Find`) for all the various values of n (the number of input keys) and m (the rank of the desired order statistic among the keys). As a 'master theorem' we establish convergence of these processes in a certain Banach space, from which known distributional convergence results as $n \to \infty$ about

(1) the number of key comparisons required

are easily recovered

(a) when $m/n \to \alpha \in [0, 1]$, and
(b) in the worst case over the choice of m.

From the master theorem it is also easy, for distributional convergence of

(2) the number of symbol comparisons required,

both to recover the known result in the case (a) of fixed quantile α and to establish our main new result in the case (b) of worst-case `Find`.

Our techniques allow us to unify the treatment of cases (1) and (2) and indeed to consider many other cost functions as well. Further, all our results provide a stronger mode of convergence (namely, convergence in L^p or almost surely) than convergence in distribution. Extensions to `MultipleQuickSelect` are discussed briefly.

2010 *Mathematics subject classification*: Primary 60F25
Secondary 68W40, 60F15

† Research for both authors supported by the Acheson J. Duncan Fund for the Advancement of Research in Statistics.

1. Introduction

QuickSelect (also known as Find), introduced by Hoare [11], is a randomized algorithm for selecting a specified order statistic from an input sequence of objects, or rather their identifying labels usually known as *keys*. The keys can be numeric or symbol strings, or indeed any labels drawn from a given linearly ordered set. Suppose we are given keys $y_1, \ldots, y_n$ and we want to find the mth smallest among them. The algorithm first selects a key (called the pivot) uniformly at random. It then compares every other key to the pivot, thereby determining the rank (call it r) of the pivot among the n keys. If $r = m$, then the algorithm terminates, returning the pivot key as output. If $r > m$, then the algorithm is applied recursively to the keys smaller than the pivot to find the mth smallest among those; while if $r < m$, then the algorithm is applied recursively to the keys larger than the pivot to find the $(m-r)$th smallest among those. More formal descriptions of QuickSelect can be found in [11] and [13], for example.

Observe that, for fixed n and a given sequence $(y_1, \ldots, y_n)$ of keys, it is possible to build the randomness needed to run QuickSelect for *every* value of $m \in \{1, \ldots, n\}$ on a single probability space, as follows. Let π denote a uniformly random permutation of $\{1, \ldots, n\}$, and consider the sequence $(z_1, \ldots, z_n)$ with $z_i := y_{\pi_i}$ for $i = 1, \ldots, n$. Regardless of the value of m, choose z_1 as the initial pivot, and when the algorithm is applied recursively, apply it to the appropriate sequence of z_i-values listed in the *same* relative order as within $(z_1, \ldots, z_n)$.

The cost of running QuickSelect can be measured by assessing the cost of comparing keys. We assume that every comparison of two (distinct) keys costs some amount that is perhaps dependent on the values of the keys, and then the cost of the algorithm is the sum of the comparison costs.

Until recently, it has been customary to assign unit cost to each comparison of two keys, irrespective of their values. We denote the (random) key-comparisons-count cost for QuickSelect by $K_{n,m}$. As we have explained, for fixed n one can use a single uniformly random permutation of $\{1, \ldots, n\}$ to build a single probability space on which all of the random variables $K_{n,m}$ with $1 \leqslant m \leqslant n$ are defined. (Note also that the joint distribution of $K_{n,1}, \ldots, K_{n,n}$ does not depend on the initial sequence $(y_1, \ldots, y_n)$ of distinct keys.) Among other things, this opens up the possibility of studying the distribution of $\max_m K_{n,m}$, the cost of so-called 'worst-case Find', in which an adversary is allowed to choose the rank of the key sought by the QuickSelect algorithm. Our motivation for this paper was to investigate the large-n behaviour of worst-case Find for more general cost functions.

There have been many studies of the random variables $K_{n,m}$, including [2], [17], [10], [14], [9], [3], [12], [4], and [6], and several corresponding studies, including [19], [15], and [16], of the number(s) of key comparisons for an extension of QuickSelect called MultipleQuickSelect that searches simultaneously for multiple order statistics. Grübel and Rösler [10] analysed a modified version of QuickSelect that splits the collection of keys into two sets, those smaller than the pivot and those greater than or equal to the pivot, rather than into three sets (one of which has the pivot as its only element) as considered in this paper. They studied (see especially their Theorem 4) the limiting behaviour of this modified QuickSelect through the convergence (in distribution, in the

Skorokhod topology on the space $D[0,1]$ of *càdlàg* functions on the unit interval $[0,1]$) of a sequence $X_1, X_2, \ldots$ of stochastic processes defined by $X_n(\alpha) := n^{-1}K_{n,\lfloor n\alpha \rfloor+1}$ for $\alpha \in [0,1)$ and $X_n(1) := n^{-1}K_{n,n}$. Rüschendorf [14, Examples 4.1–4.2] utilized the contraction method to prove that the scale-normalized key-comparisons-count cost $n^{-1}\max_m K_{n,m}$ of worst-case `Find` (the version considered in this paper) converges in distribution. Devroye [3] presented an alternative proof of the latter result.

But unit cost is not always a reasonable model for comparing two keys. For example, if each key is a string of symbols, then a more realistic model for the cost of comparing two keys is the value of the first index at which the two symbol strings differ. To date, only a few papers ([21], [7], and [8]) have considered `QuickSelect` from this more realistic symbol-comparisons perspective. As in [8], in this paper we will treat a rather general class of cost functions that includes both key-comparisons cost and symbol-comparisons cost.

In our set-up (to be described in detail in Section 2) for this paper, we will consider a variety of probabilistic models (called *probabilistic sources*) for how a key is generated as an infinite-length string of symbols, but we will always assume that the keys form an infinite sequence of independent and identically distributed and almost surely distinct symbol strings. This gives us, on a single probability space, all the randomness needed to run `QuickSelect` for *every* value of n and *every* value of $m \in \{1,\ldots,n\}$ by always choosing the *first* key in the sequence as the pivot (and maintaining initial relative order of keys when the algorithm is applied recursively); this is what is meant by the *natural coupling* (see [5, Section 1]) of the runs of the algorithm for varying n and m. As explained in [5, Section 1], the coupling allows us to consider stronger forms of convergence than convergence in distribution, such as almost sure convergence and convergence in L^p.

Whatever cost function is used for comparisons of two keys, let $\texttt{FIND}(n,m)$ denote the corresponding total cost of `QuickSelect` (under the natural coupling) in selecting the mth-order statistic from the first n keys. Let $m_n \in \{1,\ldots,n\}$ for every n, and suppose that $m_n/n \to \alpha \in [0,1]$. Fill and Nakama [8] prove, under certain 'tameness' conditions (to be reviewed later) on the probabilistic source and the cost function, that $n^{-1}\texttt{FIND}(n,m_n)$ converges both in L^p and almost surely to a limiting random variable. We complement their result by proving analogous results for the cost of worst-case `Find`, namely, $\max_{1\leqslant m\leqslant n}\texttt{FIND}(n,m)$. Our new results and (under somewhat stronger hypotheses than assumed in [8]) the results of Fill and Nakama [8] are both obtained rather effortlessly from a 'master theorem', Theorem 4.1, which establishes convergence in a certain Banach space of a certain sequence of tree-indexed processes closely related to the operation of `QuickSelect` for all the various values of n and m.

An outline for this paper is as follows. First, in Section 2, we carefully describe our set-up and, in some detail, discuss probabilistic sources, cost functions, and tameness; we also discuss the idea of *seeds*, which allow us a unified treatment of all sources. In Section 3 we state and prove a number of useful lemmas. In Section 4 (specifically, our master Theorem 4.1) we prove that the '`QuickSelect` tree processes' to which we alluded in the preceding paragraph converge in a certain Banach space (described in Definition 3.10 and Proposition 3.11). Some consequences of Theorem 4.1 are provided in Section 5; the highlight is Corollary 5.4, which gives sufficient conditions for L^p-convergence of the cost of worst-case `Find`. In Section 6, we use Theorem 4.1 to provide (under an additional restriction) very simple proofs of Theorems 3.1 and 4.1 in [8]; the latter concerns L^p-convergence of the cost of `QuickSelect` for fixed α. Finally, in Section 7, we complement the L^p-convergence result of Corollary 5.4 for the cost of worst-case `Find` by providing a tameness condition under which the scale-normalized cost of worst-case `Find` converges almost surely.

Remark 1.1. As recalled from [8] at the end of our Section 2.1, many common sources, including memoryless and Markov sources, have the property that the source-specific cost function β corresponding to the symbol-comparisons cost for comparing keys is ϵ-tame for every $\epsilon > 0$. Thus, for such sources, the conclusions of our two main results, Theorem 4.1 and Corollary 5.4, hold for every $p \in [2, \infty)$, and the almost-sure convergence theorem (Theorem 7.1) for worst-case `Find` also applies to all such sources.

2. Set-up

2.1. Probabilistic sources

Let us define the fundamental probabilistic structure underlying the analysis of `QuickSelect`. We assume that keys arrive independently and with the same distribution and that each key is composed of a sequence of symbols from some finite or countably infinite alphabet. Let Σ be this alphabet (which we assume is totally ordered by $\leqslant$). Then a key is an element of Σ^∞ (ordered by the lexicographic order, call it $\preceq$, corresponding to $(\Sigma, \leqslant)$) and a *probabilistic source* is a stochastic process $W = (W_1, W_2, W_3, \ldots)$ such that, for each i, the random variable W_i takes values in Σ. We will impose restrictions on the distribution of W that will have as a consequence that (with probability one) all keys are distinct.

We denote the cost (assumed to be non-negative) of comparing two keys w, w' by $\text{cost}(w, w')$. As two examples, the choice $\text{cost}(w, w') \equiv 1$ gives rise to a key-comparisons analysis, whereas if words are symbol strings then a symbol-comparisons analysis is obtained by letting $\text{cost}(w, w')$ be the first index at which w and w' disagree.

Since Σ^∞ is totally ordered, a probabilistic source W is governed by a distribution function F defined for $w \in \Sigma^\infty$ by

$$F(w) := \mathbb{P}(W \preceq w).$$

Then the corresponding inverse probability transform M, defined by

$$M(u) := \inf\{w \in \Sigma^\infty : u \leqslant F(w)\},$$

has the property that if $U \sim \text{uniform}(0, 1)$, then $M(U)$ has the same distribution as W. We refer to such uniform random variables U as *seeds*.

Using this technique we can define a *source-specific cost function*

$$\beta : (0, 1) \times (0, 1) \to [0, \infty)$$

by $\beta(u, v) := \text{cost}(M(u), M(v))$.

Definition 2.1. Let $0 < c < \infty$ and $0 < \epsilon < \infty$. A source-specific cost function β is said to be (c, ϵ)-*tame* if for $0 < u < t < 1$ we have

$$\beta(u, t) \leqslant c\,(t-u)^{-\epsilon},$$

and is said to be ϵ-*tame* if it is (c, ϵ)-tame for some c.

For further important background on sources, cost functions, and tameness, we refer the reader to Section 2.1 (see especially Definitions 2.3–2.4 and Remark 2.5) in Fill and Nakama [8]. Note in particular that many common sources, including memoryless and Markov sources, have the property that the source-specific cost function β corresponding to symbol-comparisons cost for comparing keys is ϵ-tame for every $\epsilon > 0$.

2.2. Tree of seeds and the `QuickSelect` tree processes

Let $\mathcal{T}$ be the collection of (finite or infinite) rooted ordered binary trees (whenever we refer to a binary tree we will assume it is of this variety) and let $\overline{T} \in \mathcal{T}$ be the complete infinite binary tree. We will label each node θ in such a tree by a binary sequence representing the path from the root to θ, where 0 corresponds to taking the left child and 1 to taking the right. We consider the set of real-valued stochastic processes each with index set equal to some $T \in \mathcal{T}$. For such a process, we extend the index set to $\overline{T}$ by defining $X_\theta = 0$ for $\theta \in \overline{T} \setminus T$. This convention allows us to define addition of any two such processes componentwise, as well as scalar multiplication componentwise. In doing so, we obtain a vector space $\mathcal{B}$ of such processes. We will have need for the following definition of levels of a binary tree.

Definition 2.2. For $0 \leqslant k < \infty$, we define the *kth level* Λ_k of a binary tree as the collection of vertices that are at distance k from the root.

Let

$$\Theta = \bigcup_{0 \leqslant k < \infty} \{0, 1\}^k$$

be the set of all finite-length binary strings, where $\{0, 1\}^0 = \{\varepsilon\}$ with ε denoting the empty string. Set $L_\varepsilon := 0$, $R_\varepsilon := 1$, and $\tau_\varepsilon := 1$. Then, for $\theta \in \Theta$, we define $|\theta|$ to be the length of the string θ, and $\upsilon_\theta(n)$ to be the size (through the arrival of the nth key) of the subtree rooted at node θ. Given a sequence of independent and identically distributed (i.i.d.) seeds $U_1, U_2, U_3, \ldots$, we recursively define

$$\begin{aligned} \tau_\theta &:= \inf\{i : L_\theta < U_i < R_\theta\}, \\ L_{\theta 0} &:= L_\theta, \; L_{\theta 1} := U_{\tau_\theta}, \\ R_{\theta 0} &:= U_{\tau_\theta}, \; R_{\theta 1} := R_\theta, \end{aligned}$$

where $\theta_1\theta_2$ denotes the concatenation of $\theta_1, \theta_2 \in \Theta$. For a source-specific cost function β and $0 \leqslant p < \infty$ we define

$$\begin{aligned} S_{n,\theta} &:= \sum_{\tau_\theta < i \leqslant n} \mathbf{1}(L_\theta < U_i < R_\theta)\beta(U_i, U_{\tau_\theta}), \\ I_p(x,a,b) &:= \int_a^b \beta^p(u,x)\,du, \\ I_{p,\theta} &:= I_p(U_{\tau_\theta}, L_\theta, R_\theta), \\ I_\theta &:= I_{1,\theta}, \\ C_\theta &:= (\tau_\theta, U_{\tau_\theta}, L_\theta, R_\theta). \end{aligned}$$

In some later definitions we will make use of the positive part function defined as usual by $x^+ := x\mathbf{1}(x > 0)$. Given a source-specific cost function β and the seeds $U_1, U_2, U_3, \ldots$, we define the nth `QuickSelect` seed process as the n-nodes binary tree indexed stochastic process obtained by successive insertions of $U_1, \ldots, U_n$ into an initially empty binary search tree.

Before we use these random variables, we supply some understanding of them for the reader. The arrival time τ_θ is the index of the seed that is slotted into node θ in the construction of the `QuickSelect` seed process. Note that for each $\theta \in \Theta$ we have $P(\tau_\theta < \infty) = 1$. The interval (L_θ, R_θ) provides sharp bounds for all seeds arriving after time τ_θ that interact with U_{τ_θ} in the sense of being placed in the subtree rooted at U_{τ_θ}. A crucial observation is that, conditioned on C_θ, the sequence of seeds $U_{\tau_\theta+1}, U_{\tau_\theta+2}, \ldots$ are i.i.d. uniform$(0,1)$; thus, again conditioned on C_θ, the sum $S_{n,\theta}$ is the sum of $(n-\tau_\theta)^+$ i.i.d. random variables. Note that when $n \leqslant \tau_\theta$ the sum defining $S_{n,\theta}$ is empty and so $S_{n,\theta} = 0$; in this case we shall conveniently interpret $S_{n,\theta}/(n-\tau_\theta)^+ = 0/0$ as 0. The random variable $S_{n,\theta}$ is the total cost of comparing the key with seed U_{τ_θ} with keys (among the first n to arrive) whose seeds fall in the interval (L_θ, R_θ), and $I_{p,\theta}$ is the conditional pth moment of one such comparison: If we let $U \sim$ uniform$(0,1)$ independent of C_θ, then

$$I_{p,\theta} = \mathbb{E}\big[\mathbf{1}(L_\theta < U < R_\theta)\beta^p(U, U_{\tau_\theta}) \mid C_\theta\big].$$

Conditioned on C_θ, the term $S_{n,\theta}$ is the sum of $(n-\tau_\theta)^+$ i.i.d. random variables with pth moment $I_{p,\theta}$.

We define the *nth* `QuickSelect` *tree process* as the binary-tree-indexed stochastic process $S_n = (S_{n,\theta})_{\theta\in\Theta}$ and the *limit* `QuickSelect` *tree process* (so called in light of Theorem 4.1) by $I = (I_\theta)_{\theta\in\Theta}$.

3. Preliminaries

We first prove some elementary lemmas that will be integral to the arguments used in the remainder of the paper. An important technique that will prove effective will be to bound moments of $I_{s,\theta}$ where $\theta \in \Lambda_k$ by an expression with geometric decrease in k. The following lemma provides such a bound in the case of an ϵ-tame source.

https://doi.org/10.1017/S0963548314000121 Published online by Cambridge University Press

Lemma 3.1. *If β is (c, ϵ)-tame with $0 \leqslant \epsilon < 1/s$, then for each fixed node $\theta \in \Lambda_k$ and $0 \leqslant r < \infty$ we have*

$$\mathbb{E} I_{s,\theta}^r \leqslant \left(\frac{2^{s\epsilon} c^s}{1 - s\epsilon} \right)^r \left(\frac{1}{r + 1 - rs\epsilon} \right)^k.$$

Proof. By ϵ-tameness and concavity of the $(1 - s\epsilon)$-power function,

$$\begin{aligned} I_{s,\theta} &\leqslant c^s \int_{L_\theta}^{R_\theta} |u - U_{\tau_\theta}|^{-s\epsilon} \, du \\ &= \frac{c^s}{1 - s\epsilon} \left[\left(R_\theta - U_{\tau_\theta} \right)^{1 - s\epsilon} + \left(U_{\tau_\theta} - L_\theta \right)^{1 - s\epsilon} \right] \\ &\leqslant \frac{2^{s\epsilon} c^s}{1 - s\epsilon} (R_\theta - L_\theta)^{1 - s\epsilon}. \end{aligned}$$

Since $R_\theta - L_\theta$ is distributed as the product of k independent uniform$(0, 1)$ random variables, taking rth moments gives the desired bound. □

As a consequence of Lemma 3.1, we have

Lemma 3.2. *Let $1 \leqslant p < \infty$ and consider a fixed node θ. If the source-specific cost function β is ϵ-tame for some $0 \leqslant \epsilon < 1/p$, then as $n \to \infty$ we have*

$$\frac{S_{n,\theta}}{n} \xrightarrow{L^p} I_\theta.$$

Proof. The proof essentially repeats an argument within the proof of Theorem 3.1 in Fill and Nakama [8]. Conditioned on C_θ, the random variable $S_{n,\theta}$ is the sum of $(n - \tau_\theta)^+$ i.i.d. non-negative random variables with expectation I_θ and pth moment $I_{p,\theta}$. The L^p law of large numbers (L^pLLN) applies almost surely because $\mathbb{E} I_{p,\theta} < \infty$ by Lemma 3.1 and hence $I_{p,\theta} < \infty$ almost surely. The L^pLLN gives

$$\mathbb{E}\left[\left| \frac{S_{n,\theta}}{(n - \tau_\theta)^+} - I_\theta \right|^p \,\middle|\, C_\theta \right] \to 0 \quad \text{a.s.} \tag{3.1}$$

By convexity of the pth-power function,

$$\mathbb{E}\left[\left| \frac{S_{n,\theta}}{(n - \tau_\theta)^+} - I_\theta \right|^p \,\middle|\, C_\theta \right] \leqslant 2^{p-1} \left\{ \mathbb{E}\left[\left(\frac{S_{n,\theta}}{(n - \tau_\theta)^+} \right)^p \,\middle|\, C_\theta \right] + I_\theta^p \right\}$$

and also

$$\left(\frac{S_{n,\theta}}{(n - \tau_\theta)^+} \right)^p \leqslant \frac{1}{(n - \tau_\theta)^+} \sum_{i : \tau_\theta < i \leqslant n} \mathbf{1}(L_\theta < U_i < R_\theta) \beta^p(U_i, U_{\tau_\theta}),$$

which implies

$$\mathbb{E}\left[\left(\frac{S_{n,\theta}}{(n - \tau_\theta)^+} \right)^p \,\middle|\, C_\theta \right] \leqslant I_{p,\theta}.$$

https://doi.org/10.1017/S0963548314000121 Published online by Cambridge University Press

Therefore, we have the following bound:

$$\mathbb{E}\left[\left|\frac{S_{n,\theta}}{(n-\tau_\theta)^+}-I_\theta\right|^p \,\middle|\, C_\theta\right] \leqslant 2^{p-1}\big(I_{p,\theta}+I_\theta^p\big) \leqslant 2^p I_{p,\theta}. \tag{3.2}$$

Recall that Lemma 3.1 implies that $\mathbb{E}I_{p,\theta}<\infty$. Therefore, by (3.1)–(3.2) and the dominated convergence theorem,

$$\mathbb{E}\left|\frac{S_{n,\theta}}{(n-\tau_\theta)^+}-I_\theta\right|^p \to 0.$$

Now to complete the proof of the lemma we show that

$$\mathbb{E}\left|\frac{S_{n,\theta}}{n}-\frac{S_{n,\theta}}{(n-\tau_\theta)^+}\right|^p \to 0.$$

By our convention for $S_{n,\theta}/(n-\tau_\theta)^+$ when $n \leqslant \tau_\theta$, we have

$$\left|\frac{S_{n,\theta}}{n}-\frac{S_{n,\theta}}{(n-\tau_\theta)^+}\right|^p = \left|\frac{S_{n,\theta}}{n}-\frac{S_{n,\theta}}{(n-\tau_\theta)^+}\right|^p \mathbf{1}(\tau_\theta<n).$$

By a simple calculation,

$$\left|\frac{S_{n,\theta}}{n}-\frac{S_{n,\theta}}{(n-\tau_\theta)^+}\right|^p = \left|\frac{\tau_\theta}{n}\right|^p\left|\frac{S_{n,\theta}}{(n-\tau_\theta)^+}\right|^p.$$

Taking expectations conditioned on C_θ gives

$$\begin{aligned}\mathbb{E}\left[\left|\frac{S_{n,\theta}}{n}-\frac{S_{n,\theta}}{(n-\tau_\theta)^+}\right|^p \mathbf{1}(\tau_\theta<n)\,\middle|\, C_\theta\right] &= \mathbf{1}(\tau_\theta<n)\left(\frac{\tau_\theta}{n}\right)^p \mathbb{E}\left[\left|\frac{S_{n,\theta}}{(n-\tau_\theta)^+}\right|^p \,\middle|\, C_\theta\right]\\ &\leqslant \mathbf{1}(\tau_\theta<n)\left(\frac{\tau_\theta}{n}\right)^p I_{p,\theta} \leqslant I_{p,\theta},\end{aligned}$$

and, in particular,

$$\mathbb{E}\left[\left|\frac{S_{n,\theta}}{n}-\frac{S_{n,\theta}}{(n-\tau_\theta)^+}\right|^p \,\middle|\, C_\theta\right] \to 0 \quad \text{a.s.}$$

Thus, again by the dominated convergence theorem,

$$\mathbb{E}\left|\frac{S_{n,\theta}}{n}-\frac{S_{n,\theta}}{(n-\tau_\theta)^+}\right|^p \to 0. \qquad \square$$

A Poisson binomial sum is a generalization of a binomial distributed random variable. Let $X_i \sim \mathrm{Bern}(p_i)$, $i=1,2,\ldots,n$, be independent, where $\mathrm{Bern}(p)$ denotes the Bernoulli distribution with success probability p. Then we say that $X := \sum_i X_i$ is a *Poisson binomial sum*. The following lemma is a restatement of Theorem 4.4 (part 1) in [18]; we include a proof for completeness.

Lemma 3.3. *Let $X=\sum_i X_i$ be a Poisson binomial sum with*

$$X_i \sim \mathrm{Bern}(p_i) \quad \textit{and} \quad \mathbb{E}X=\sum_i p_i =: \mu.$$

https://doi.org/10.1017/S0963548314000121 Published online by Cambridge University Press

Then, for any $\beta \geqslant 0$ *we have*

$$\mathbb{P}(X \geqslant (1+\beta)\mu) \leqslant \left[\frac{e^{\beta}}{(1+\beta)^{1+\beta}}\right]^{\mu}.$$

Proof. The result is trivial for $\beta = 0$, so suppose $\beta > 0$. For any $t > 0$, by Markov's inequality

$$\mathbb{P}(X \geqslant (1+\beta)\mu) = \mathbb{P}(e^{tX} \geqslant e^{t(1+\beta)\mu}) \leqslant e^{-t(1+\beta)\mu}\mathbb{E}e^{tX}.$$

Since $X = \sum_i X_i$ and the X_i are independent,

$$\mathbb{E}e^{tX} = \prod_{i=1}^{n} \mathbb{E}e^{tX_i} = \prod_{i=1}^{n}\left[1 + p_i(e^t - 1)\right] \leqslant \prod_{i=1}^{n} \exp\left[p_i(e^t-1)\right] = \exp\left[\mu(e^t-1)\right].$$

Combining these two inequalities and choosing $t = \ln(1+\beta) > 0$ produces the desired bound. □

Lemma 3.4. *Let* $H_m := \sum_{i=1}^{m} i^{-1}$, *the mth harmonic number. If* $\theta \in \Lambda_k$ *with* $k > H_m$, *then*

$$\mathbb{P}(\tau_\theta \leqslant m) \leqslant \left(\frac{eH_m}{k}\right)^{k} e^{-H_m}.$$

Proof. By symmetry, it suffices to consider $\theta = 0^k$ (the leftmost node of the binary tree at level k). Then τ_θ is the arrival time of the kth record-smallest seed. If R_m is the number of seeds among the first m to be record-smallest upon arrival, then we have $P(\tau_\theta \leqslant m) = P(R_m \geqslant k)$. It is well known that R_m has the distribution of a Poisson binomial sum:

$$R_m \stackrel{\mathcal{L}}{=} \sum_{i=1}^{m} X_i,$$

where $X_i \sim \mathrm{Bern}(1/i)$, $i = 1, \ldots, m$, are independent and $\stackrel{\mathcal{L}}{=}$ denotes equality in law. (Consult, for example, [1, Problem 20.9].) This implies that

$$\mu := \mathbb{E}R_m = \sum_{i=1}^{m} i^{-1} = H_m.$$

Thus, for $\beta = (k/H_m) - 1 > 0$, applying Lemma 3.3 gives

$$\begin{aligned}\mathbb{P}(\tau_\theta \leqslant m) = \mathbb{P}(R_m \geqslant k) &= \mathbb{P}(R_m \geqslant (1+\beta)\mu) \\ &\leqslant \left[\frac{e^{\beta}}{(1+\beta)^{1+\beta}}\right]^{\mu} = \left(\frac{eH_m}{k}\right)^{k} e^{-H_m}.\end{aligned}$$ □

In the context of Lemma 3.4, we will use the following standard bound on harmonic numbers to approximate H_n by $\ln n$ up to a constant term.

Lemma 3.5. *Let γ be Euler's constant. Then, for $n = 1, 2, \ldots$ we have*

$$\gamma \leqslant H_n - \ln n \leqslant 1.$$

The next lemma uses Lemma 3.1 in its proof and can be seen to generalize Lemma 3.1 (by letting $a \to 0$ and then $w \to 1$).

Lemma 3.6. *Consider a fixed node $\theta \in \Lambda_k$, and let $0 \leqslant r < \infty$ and $0 < a < \infty$ be constants. Let the source-specific cost function β be (c, ϵ)-tame with $0 \leqslant \epsilon < 1/s$. Then, for any $v, w \in (1, \infty)$ such that $v^{-1} + w^{-1} = 1$, we have*

$$\mathbb{E}\left(\tau_\theta{}^{-a} I_{s,\theta}^r\right) \leqslant \left(\frac{2^{s\epsilon} c^s}{1 - s\epsilon}\right)^r \left(\frac{1}{wr + 1 - wrs\epsilon}\right)^{k/w} \times \left\{\left(2vak \int_1^{e-\delta_k} \left[\frac{e \ln \alpha + \delta_k}{\alpha^{1+va}}\right]^k d\alpha\right)^{1/v} + (e - \delta_k)^{-ak}\right\},$$

where $\delta_k = e/k$.

Proof. By Hölder's inequality,

$$\mathbb{E}\left(\tau_\theta^{-a} I_{s,\theta}^r\right) \leqslant \|\tau_\theta^{-a}\|_v \|I_{s,\theta}^r\|_w.$$

We can use Lemma 3.1 to bound the second factor. To treat the first factor, we express the expectation as an integral of tail probabilities, which we bound using Chernoff's inequality. Write

$$\mathbb{E}\tau_\theta^{-va} = \int_0^\infty \mathbb{P}(\tau_\theta^{-va} \geqslant t)\, dt = \int_0^\infty \mathbb{P}(\tau_\theta \leqslant \lfloor t^{-1/(va)} \rfloor)\, dt =: J.$$

Using the change-of-variables $\alpha^k = t^{-1/(va)}$, we get

$$J = \int_0^\infty \mathbb{P}(\tau_\theta \leqslant \lfloor \alpha^k \rfloor)(vak)\alpha^{-vak-1}\, d\alpha. \tag{3.3}$$

The next step is to use Lemma 3.4 to bound $\mathbb{P}(\tau_\theta \leqslant \lfloor \alpha^k \rfloor)$. However, we need to check that the hypothesis of Lemma 3.4 that

$$k \geqslant H_{\lfloor \alpha^k \rfloor} \tag{3.4}$$

is satisfied. By Lemma 3.5 we have the upper bound

$$H_{\lfloor \alpha^k \rfloor} \leqslant \ln \lfloor \alpha^k \rfloor + 1 \leqslant k \ln \alpha + 1,$$

so $\alpha \leqslant e^{1-(1/k)}$ is sufficient for (3.4), as therefore is $\alpha \leqslant e[1 - (1/k)]$. Writing $\delta_k = e/k$, we decompose the integral $\int_0^\infty$ in (3.3) for J into

$$\int_0^1 + \int_1^{e-\delta_k} + \int_{e-\delta_k}^\infty.$$

https://doi.org/10.1017/S0963548314000121 Published online by Cambridge University Press

When $0 \leqslant \alpha < 1$, we have $\mathbb{P}(\tau_\theta \leqslant \lfloor \alpha^k \rfloor) = 0$ since $\tau_\theta \geqslant 1$. For the second integral we bound this probability using Lemma 3.4, and for the third integral we use the bound 1:

$$\begin{aligned} J &\leqslant \int_1^{e-\delta_k} \left(\frac{eH_{\lfloor \alpha^k \rfloor}}{k}\right)^k e^{-H_{\lfloor \alpha^k \rfloor}} (vak)\alpha^{-vak-1}\, d\alpha + \int_{e-\delta_k}^{\infty} (vak)\alpha^{-vak-1}\, d\alpha \\ &= \int_1^{e-\delta_k} \left(\frac{eH_{\lfloor \alpha^k \rfloor}}{k}\right)^k e^{-H_{\lfloor \alpha^k \rfloor}} (vak)\alpha^{-vak-1}\, d\alpha + (e-\delta_k)^{-vak}. \end{aligned}$$

By Lemma 3.5,

$$\begin{aligned} \left(\frac{eH_{\lfloor \alpha^k \rfloor}}{k}\right)^k e^{-H_{\lfloor \alpha^k \rfloor}} &\leqslant \left(\frac{e(\ln\lfloor \alpha^k \rfloor + 1)}{k}\right)^k e^{-\ln\lfloor \alpha^k \rfloor} \\ &\leqslant \left[e\left(\ln \alpha + \frac{1}{k}\right)\right]^k \left(\frac{1}{\lfloor \alpha^k \rfloor}\right) \\ &\leqslant 2\left[\frac{e \ln \alpha + \delta_k}{\alpha}\right]^k, \end{aligned}$$

where in the last inequality we have used the fact that $2\lfloor x \rfloor \geqslant x$ for $x \geqslant 1$. Thus

$$\int_1^{e-\delta_k} \left(\frac{eH_{\lfloor \alpha^k \rfloor}}{k}\right)^k e^{-H_{\lfloor \alpha^k \rfloor}} (vak)\alpha^{-vak-1}\, d\alpha \leqslant 2vak \int_1^{e-\delta_k} \left[\frac{e \ln \alpha + \delta_k}{\alpha^{1+va}}\right]^k d\alpha. \tag{3.5}$$

Bounding $\|I^r_{s,\theta}\|_w$ using Lemma 3.1 gives

$$\|I^r_{s,\theta}\|_w \leqslant \left(\frac{2^{s\epsilon} c^s}{1 - s\epsilon}\right)^r \left(\frac{1}{wr + 1 - wrs\epsilon}\right)^{k/w}. \tag{3.6}$$

Combining (3.5) and (3.6) and using $(x+y)^{1/v} \leqslant x^{1/v} + y^{1/v}$ for $x, y \geqslant 0$ proves the lemma. □

The following elementary calculus lemma will prove useful in the proof of Theorem 4.1.

Lemma 3.7. *The function $f(x) := ex^{-1} \ln x$ has a unique maximum for $x \in (0, \infty)$, at $\hat{x} = e$ with value $f(\hat{x}) = 1$.*

The following simple consequence of the triangle inequality will be used in the proof of Corollary 5.3.

Lemma 3.8. *Let $(x_i)_{i \in I}$ and $(y_i)_{i \in I}$ be two collections of non-negative real numbers indexed by a common set I. Then*

$$\left|\sup_{i \in I} x_i - \sup_{i \in I} y_i\right| \leqslant \sup_{i \in I} |x_i - y_i|,$$

provided that at least one of the suprema on the left is finite (so that the difference of them is well defined).

Our next lemma is a simple bound, used in the proofs of Theorem 6.3 and Lemma 7.4, on the L^p-norm of a maximum of a finite collection of non-negative random variables.

https://doi.org/10.1017/S0963548314000121 Published online by Cambridge University Press

Lemma 3.9 (Max Lemma). *Consider random variables $X_m \geqslant 0$, $m = 1, \ldots, M$, and let $1 \leqslant p < \infty$. Then*

$$\|\max_m X_m\|_p \leqslant M^{1/p} \max_m \|X_m\|_p.$$

Proof. The key to the proof is to bound the maximum of the M random variables X_m^p by their sum and then the sum of their pth moments by M times the maximum of the pth moments:

$$\begin{aligned}\|\max_m X_m\|_p^p &= \mathbb{E}\big(\max_m X_m\big)^p = \mathbb{E}\max_m X_m^p \\ &\leqslant \sum_m \mathbb{E}X_m^p \leqslant M \max_m \mathbb{E}X_m^p \\ &= M \max_m \|X_m\|_p^p. \end{aligned}$$ □

Recall from the first paragraph of Section 2.2 the definition of the vector space $\mathcal{B}$. Given any sequence of positive real numbers (a_k) and any $1 \leqslant p \leqslant \infty$, we can define a very useful functional on $\mathcal{B}$ as follows.

Definition 3.10. We define a real-valued function $|||\cdot|||_p$ on the vector space $\mathcal{B}$ of binary tree indexed stochastic processes (X_θ) by setting

$$|||X|||_p := \sum_{k=0}^{\infty} a_k \max_{\theta \in \Lambda_k} \|X_\theta\|_p.$$

Let

$\mathcal{B}^{(p)} := \{X \in \mathcal{B} : |||X|||_p < \infty\}$; in typical fashion, we identify two processes X and Y in $\mathcal{B}^{(p)}$ if $|||X - Y|||_p = 0$.

Proposition 3.11. *$\mathcal{B}^{(p)}$ is a Banach space with the norm $|||\cdot|||_p$.* □

Proof. It is easily verified that $\mathcal{B}^{(p)}$ is a vector subspace of $\mathcal{B}$ and that $|||\cdot|||_p$ is a norm on $\mathcal{B}^{(p)}$. What remains is to establish that the vector space $\mathcal{B}^{(p)}$ is complete with respect to the metric induced by $|||\cdot|||_p$. Our proof of this is adapted from Section 6.10 in [22]. Suppose $X^{(n)} \in \mathcal{B}^{(p)}$ form a Cauchy sequence with respect to the norm $|||\cdot|||_p$. Choose a sequence $\ell_n \to \infty$ as $n \to \infty$ such that for $m_1, m_2 \geqslant \ell_n$ we have

$$|||X^{(m_1)} - X^{(m_2)}|||_p \leqslant 2^{-n}.$$

This implies that for each $\theta \in \Theta$, the random variables $X_\theta^{(n)}$ form a Cauchy sequence in L^p, and for $m \geqslant \ell_n$ and $t \geqslant n$ we have

$$|||X^{(m)} - X^{(\ell_t)}|||_p = \sum_{k=0}^{\infty} a_k \max_{\theta \in \Lambda_k} \|X_\theta^{(m)} - X_\theta^{(\ell_t)}\|_p \leqslant 2^{-n}.$$

https://doi.org/10.1017/S0963548314000121 Published online by Cambridge University Press

For each θ, let X_θ be the L^p limit of $X_\theta^{(n)}$. For m fixed and $k = 1, 2, \ldots$, define

$$f_t(k) := \max_{\theta\in\Lambda_k} \|X_\theta^{(m)} - X_\theta^{(\ell_t)}\|_p,$$

and

$$f(k) := \max_{\theta\in\Lambda_k} \|X_\theta^{(m)} - X_\theta\|_p.$$

Since for every θ we have $X_\theta^{(n)} \xrightarrow{L^p} X_\theta$, for each k we have $f_t(k) \to f(k)$ as $t \to \infty$. Let μ be the counting measure induced by (a_k) on $\mathbb{Z}_+$, *i.e.*, $\mu(\{k\}) = a_k$ for every $k \in \mathbb{Z}_+$. Then by Fatou's lemma,

$$\int \liminf_{t\to\infty} f_t \, d\mu \leqslant \liminf_{t\to\infty} \int f_t \, d\mu,$$

which, in the case of $m \geqslant \ell_n$, simplifies to

$$|||X^{(m)} - X|||_p \leqslant \liminf_{t\to\infty} |||X^{(m)} - X^{(k_t)}|||_p \leqslant 2^{-n},$$

where $X \in \mathcal{B}^{(p)}$ is the process of the X_θ. Since this inequality holds for all $m \geqslant \ell_n$, we get

$$|||X^{(m)} - X|||_p \to 0. \qquad \square$$

4. Process convergence

Our main result is to show that the normalized QuickSelect tree process $S_n/n = (S_{n,\theta}/n)$ converges to a limit, namely, $I = (I_\theta)$, in the Banach space $(\mathcal{B}^{(p)}, |||\cdot|||_p)$ described in Proposition 3.11.

Theorem 4.1 (Master Theorem). *For $2 \leqslant p < \infty$ and ϵ-tame source-specific cost function β with $0 \leqslant \epsilon < 1/p$, as $n \to \infty$ we have*

$$\left|\left|\left| \frac{S_n}{n} - I \right|\right|\right|_p \to 0,$$

with $a_k \equiv a_{k,p} \equiv 2^{k/p}$ in the definition of $|||\cdot|||_p$.

Proof. By Lemma 3.2, for each θ in the infinite binary tree, $\|(S_{n,\theta}/n) - I_\theta\|_p \to 0$, so that by the dominated convergence theorem it suffices to find a sequence (b_k) such that

(i) for $\theta \in \Lambda_k$ we have $\|(S_{n,\theta}/n) - I_\theta\|_p \leqslant b_k$, and
(ii) $\sum_{k=0}^{\infty} 2^{k/p} b_k < \infty$.

Decomposing I_θ into $I_\theta \mathbf{1}(\tau_\theta < n) + I_\theta \mathbf{1}(\tau_\theta \geqslant n)$ gives

$$\left|\frac{S_{n,\theta}}{n} - I_\theta\right| = \left|\left(\frac{S_{n,\theta}}{n} - \frac{n-\tau_\theta}{n} I_\theta\right)\mathbf{1}(\tau_\theta < n) - \frac{\tau_\theta}{n} I_\theta \mathbf{1}(\tau_\theta < n) - I_\theta \mathbf{1}(\tau_\theta \geqslant n)\right| \tag{4.1}$$

$$\leqslant \frac{1}{n}|S_{n,\theta} - (n-\tau_\theta) I_\theta| \mathbf{1}(\tau_\theta < n) + I_\theta, \tag{4.2}$$

https://doi.org/10.1017/S0963548314000121 Published online by Cambridge University Press

so that

$$\left\| \frac{S_{n,\theta}}{n} - I_\theta \right\|_p \leqslant \left\| \frac{1}{n} |S_{n,\theta} - (n - \tau_\theta) I_\theta | \mathbf{1}(\tau_\theta < n) \right\|_p + \| I_\theta \|_p. \tag{4.3}$$

Recall that, conditionally given $C_\theta = (\tau_\theta, L_\theta, R_\theta, U_{\tau_\theta})$, the random variable $S_{n,\theta}$ is the i.i.d. sum of $(n - \tau_\theta)^+$ random variables with mean I_θ. If we let $U \sim \text{uniform}(0,1)$ independent of C_θ and define $X_\theta := \mathbf{1}(L_\theta \leqslant U \leqslant R_\theta)\beta(U, U_{\tau_\theta})$, then by Rosenthal's inequality [20] there exists a constant c_p depending only on p such that

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{n^p} |S_{n,\theta} - (n - \tau_\theta) I_\theta|^p \mathbf{1}(\tau_\theta < n) \,\middle|\, C_\theta \right] \\
&\quad \leqslant c_p \frac{1}{n^p} \mathbf{1}(\tau_\theta < n) \{ (n - \tau_\theta) \mathbb{E}\big[|X_\theta - I_\theta|^p \mid C_\theta \big] + (n - \tau_\theta)^{p/2} \big(\mathrm{Var}\big[X_\theta \mid C_\theta\big] \big)^{p/2} \} \\
&\quad \leqslant c_p \mathbf{1}(\tau_\theta < n) \left\{ \frac{1}{n^{p-1}} \mathbb{E}\big[|X_\theta - I_\theta|^p \mid C_\theta \big] + \frac{1}{n^{p/2}} \big(\mathrm{Var}\big[X_\theta \mid C_\theta\big] \big)^{p/2} \right\}.
\end{aligned}$$

However, by convexity of the pth-power function,

$$\mathbb{E}\big[|X_\theta - I_\theta|^p \mid C_\theta \big] \leqslant 2^{p-1} \big(\mathbb{E}\big[X_\theta^p \mid C_\theta \big] + I_\theta^p \big) = 2^{p-1} \big(I_{p,\theta} + I_\theta^p \big) \leqslant 2^p I_{p,\theta},$$

and $\mathrm{Var}\big[X_\theta \mid C_\theta\big] \leqslant I_{2,\theta}$. Thus, utilizing the factor $\mathbf{1}(\tau_\theta < n)$ gives

$$\mathbb{E}\left[\frac{1}{n^p} |S_{n,\theta} - (n - \tau_\theta) I_\theta|^p \mathbf{1}(\tau_\theta < n) \,\middle|\, C_\theta \right] \leqslant c_p 2^p \big(\tau_\theta^{-(p-1)} I_{p,\theta} + \tau_\theta^{-p/2} I_{2,\theta}^{p/2} \big).$$

Taking expectations yields

$$\left\| \frac{1}{n} |S_{n,\theta} - (n - \tau_\theta) I_\theta | \mathbf{1}(\tau_\theta < n) \right\|_p^p \leqslant c_p 2^p \big(\mathbb{E}\big[\tau_\theta^{-(p-1)} I_{p,\theta} \big] + \mathbb{E}\big[\tau_\theta^{-p/2} I_{2,\theta}^{p/2} \big] \big). \tag{4.4}$$

From (4.3), (4.4), and Lemma 3.1 (with $s = 1$ and $r = p$), we get

$$\begin{aligned}
&\left\| \frac{S_{n,\theta}}{n} - I_\theta \right\|_p^p \\
&\quad \leqslant \left[\left\| \frac{1}{n} |S_{n,\theta} - (n - \tau_\theta) I_\theta | \mathbf{1}(\tau_\theta < n) \right\|_p + \| I_\theta \|_p \right]^p \\
&\quad \leqslant 2^{p-1} \left[\left\| \frac{1}{n} |S_{n,\theta} - (n - \tau_\theta) I_\theta | \mathbf{1}(\tau_\theta < n) \right\|_p^p + \| I_\theta \|_p^p \right] \\
&\quad \leqslant 2^{p-1} \left[c_p 2^p \big(\mathbb{E}\big[\tau_\theta^{-(p-1)} I_{p,\theta} \big] + \mathbb{E}\big[\tau_\theta^{-p/2} I_{2,\theta}^{p/2} \big] \big) + \left(\frac{2^\epsilon c}{1 - \epsilon} \right)^p (p + 1 - p\epsilon)^{-k} \right].
\end{aligned} \tag{4.5}$$

Because $p + 1 - p\epsilon > p \geqslant 2$, it suffices to prove geometric decay at a rate faster than 2^{-k} for each of the two expectations in (4.5). (Note that when $p = 2$, the two expectations are equal, so we can and do restrict to $p > 2$ in considering the first expectation but allow $p = 2$ in considering the second.)

https://doi.org/10.1017/S0963548314000121 Published online by Cambridge University Press

To establish this geometric decay, consider $\mathbb{E}\big[\tau_\theta^{-(p-1)} I_{p,\theta}\big]$ first. For any $v, w \in (1,\infty)$ satisfying $v^{-1} + w^{-1} = 1$, applying Lemma 3.6 gives

$$\mathbb{E}\big[\tau_\theta^{-(p-1)} I_{p,\theta}\big] \leqslant \left(\frac{2^{p\epsilon}c^p}{1-p\epsilon}\right)\left(\frac{1}{w+1-wp\epsilon}\right)^{k/w} \times \left\{\left(2v(p-1)k\int_1^{e-\delta_k}\left[\frac{e\ln\alpha+\delta_k}{\alpha^{1+v(p-1)}}\right]^k d\alpha\right)^{1/v} + (e-\delta_k)^{-(p-1)k}\right\}.$$

The factor $(w+1-wp\epsilon)^{k/w}$ here is bounded by unity, and the term involving $(e-\delta_k)^{-(p-1)k}$ in the bound on $\mathbb{E}\big[\tau_\theta^{-(p-1)} I_{p,\theta}\big]$ decays at a geometric rate faster than 2^{-k}. So it suffices to show that there exist $v > 1$ and $\eta > 0$ (each of which may depend on p, but not on k or $\alpha \in (1, e-\delta_k)$) such that

$$2^v\left[\frac{e\ln\alpha+\delta_k}{\alpha^{1+v(p-1)}}\right] \leqslant 1-\eta \tag{4.6}$$

for all large k. Since $\delta_k/\alpha^{1+v(p-1)} \leqslant \delta_k \to 0$ as $k \to \infty$, it suffices to establish (4.6) with δ_k replaced by 0. By Lemma 3.7, we have

$$2^v\left(\frac{e\ln\alpha}{\alpha^{1+v(p-1)}}\right) \leqslant \frac{2^v}{1+v(p-1)},$$

and the bound here is strictly less than 1 precisely when

$$p > 1 + \frac{2^v-1}{v}. \tag{4.7}$$

Since $1 + v^{-1}(2^v-1) \to 2$ as $v \to 1$, for any $p > 2$ a value of $v > 1$ can be found satisfying (4.7).

Now consider $\mathbb{E}\big[\tau_\theta{}^{-p/2} I_{2,\theta}^{p/2}\big]$. Applying Lemma 3.6 gives

$$\mathbb{E}\tau_\theta{}^{-p/2} I_{2,\theta}^{p/2} \leqslant \left(\frac{2^{2\epsilon}c^2}{1-2\epsilon}\right)^{p/2}\left(\frac{1}{(wp/2)+1-wp\epsilon}\right)^{k/w} \times \left\{\left(vpk\int_1^{e-\delta_k}\left[\frac{e\ln\alpha+\delta_k}{\alpha^{1+(vp/2)}}\right]^k d\alpha\right)^{1/v} + (e-\delta_k)^{-pk/2}\right\}.$$

As before, the term involving $(e-\delta_k)^{-pk/2}$ poses no problem. It is therefore sufficient to show that we have geometric decay in k for the following expression:

$$2^{kv}\left(\frac{1}{(wp/2)+1-wp\epsilon}\right)^{kv/w}\int_1^{e-\delta_k}\left[\frac{e\ln\alpha+\delta_k}{\alpha^{1+(vp/2)}}\right]^k d\alpha.$$

When $p > 2$, arguing as before we need only find v and w satisfying $v^{-1}+w^{-1} = 1$ such that the following expression is bounded by $1-\eta$ for some $\eta > 0$:

$$\frac{2^v e\ln\alpha}{\alpha^{1+(vp/2)}}\frac{1}{[w((p/2)-1)+1]^{v/w}} \leqslant \frac{2^v}{(vp/2)+1}\frac{1}{[w((p/2)-1)+1]^{v/w}}.$$

However, as $v \to 1$ and $w \to \infty$ we have

$$\frac{2^v}{(vp/2)+1}\frac{1}{[w((p/2)-1)+1]^{v/w}} \to \frac{4}{p+2},$$

which for $p > 2$ is strictly less than 1. When $p = 2$ and $0 \leqslant \epsilon < 1/2$ is given, it suffices to find a value of $v > 1$ such that

$$\frac{2^v}{v+1} \frac{1}{\left[\frac{v}{v-1}(1-2\epsilon)+1\right]^{v-1}} < 1. \tag{4.8}$$

When $v = (3/2) - \epsilon$, the expression in (4.8) becomes

$$\frac{2^{(5/2)-\epsilon}}{5-2\epsilon}(4-2\epsilon)^{\epsilon-(1/2)} = \frac{4}{5-2\epsilon}(2-\epsilon)^{-[(1/2)-\epsilon]} =: f(\epsilon).$$

One can easily check that $f(1/2) = 1$ and that f is strictly increasing on $[0, 1/2]$; therefore, together they imply that (4.8) is satisfied when $v = (3/2) - \epsilon$. □

5. Consequences of process convergence

Let Γ be the collection of all non-empty paths (finite or infinite) from the root of the infinite binary tree. For any path $\gamma \in \Gamma$, define $|\gamma| \in \{1, 2, \ldots\} \cup \{\infty\}$ to be the number of levels visited by γ, and let

$$S_{n,\gamma} := \sum_{\theta \in \gamma} S_{n,\theta} \quad \text{and} \quad I_\gamma := \sum_{\theta \in \gamma} I_\theta.$$

Let $p \in [1, \infty)$, and let $a_k \equiv a_{k,p} \equiv 2^{k/p}$ be used in the definition of $|||\cdot|||_p$.

Proposition 5.1. *Let $1 \leqslant p < \infty$ and suppose that*

$$\left|\left|\left| \frac{S_n}{n} - I \right|\right|\right|_p \to 0.$$

Then as $n \to \infty$ we have the L^p-convergence

$$\left\| \sup_{\gamma \in \Gamma} \left| \frac{S_{n,\gamma}}{n} - I_\gamma \right| \right\|_p \to 0.$$

□

Proof. Consider a path $\gamma \in \Gamma$ and let $\gamma = (\theta_0, \theta_1, \ldots, \theta_{|\gamma|-1})$ if $|\gamma| < \infty$ and $\gamma = (\theta_0, \theta_1, \ldots)$ if $|\gamma| = \infty$. Then

$$\begin{aligned} \left| \frac{S_{n,\gamma}}{n} - I_\gamma \right| &= \left| \sum_{0 \leqslant k < |\gamma|} \left(\frac{S_{n,\theta_k}}{n} - I_{\theta_k} \right) \right| \\ &= \left| \sum_{0 \leqslant k < \infty} \mathbf{1}(k < |\gamma|) \left(\frac{S_{n,\theta_k}}{n} - I_{\theta_k} \right) \right| \\ &\leqslant \sum_{0 \leqslant k < \infty} \mathbf{1}(k < |\gamma|) \left| \frac{S_{n,\theta_k}}{n} - I_{\theta_k} \right|, \end{aligned} \tag{5.1}$$

where θ_k is defined arbitrarily for $k \geqslant |\gamma|$ if $|\gamma| < \infty$. Moreover, for any k we have

$$\left| \frac{S_{n,\theta_k}}{n} - I_{\theta_k} \right| \leqslant \left(\sum_{\theta \in \Lambda_k} \left| \frac{S_{n,\theta}}{n} - I_\theta \right|^p \right)^{1/p}. \tag{5.2}$$

https://doi.org/10.1017/S0963548314000121 Published online by Cambridge University Press

Note that the bound in (5.2) does not depend on which path $\gamma \in \Gamma$ was chosen. By combining (5.1)–(5.2), we get the following bound:

$$\begin{aligned}
\left\| \sup_{\gamma\in\Gamma} \left| \frac{S_{n,\gamma}}{n} - I_\gamma \right| \right\|_p &\leqslant \left\| \sum_{0\leqslant k<\infty} \left(\sum_{\theta\in\Lambda_k} \left| \frac{S_{n,\theta}}{n} - I_\theta \right|^p \right)^{1/p} \right\|_p \\
&\leqslant \sum_{0\leqslant k<\infty} \left\| \left(\sum_{\theta\in\Lambda_k} \left| \frac{S_{n,\theta}}{n} - I_\theta \right|^p \right)^{1/p} \right\|_p \\
&= \sum_{0\leqslant k<\infty} \left(\sum_{\theta\in\Lambda_k} \left\| \frac{S_{n,\theta}}{n} - I_\theta \right\|_p^p \right)^{1/p} \\
&\leqslant \sum_{0\leqslant k<\infty} 2^{k/p} \max_{\theta\in\Lambda_k} \left\| \frac{S_{n,\theta}}{n} - I_\theta \right\|_p \\
&= \left\| \left\| \frac{S_n}{n} - I \right\| \right\|_p .
\end{aligned}$$

□

Remark 5.2. The same proof shows that Proposition 5.1 continues to hold when Γ is enlarged to include all *random* paths defined on the same probability space as the seeds U_i.

We now focus our attention on the analysis of worst-case `QuickSelect`. Let

$$T_n := n^{-1} \max_{1\leqslant m\leqslant n} \mathtt{FIND}(n,m), \tag{5.3}$$

where $\mathtt{FIND}(n,m)$ is the cost for `QuickSelect` to find the mth smallest element of the first n keys. Note that T_n has the following representation:

$$T_n = \sup_{\gamma\in\Gamma} \frac{S_{n,\gamma}}{n} = \sup_{\gamma\in\Gamma} \sum_{\theta\in\gamma} \frac{S_{n,\theta}}{n}. \tag{5.4}$$

Define

$$T := \sup_{\gamma\in\Gamma} I_\gamma. \tag{5.5}$$

Corollary 5.3. *Suppose that*

$$\left\| \left\| \frac{S_n}{n} - I \right\| \right\|_p \to 0.$$

Then as $n \to \infty$ we have

$$T_n \xrightarrow{L^p} T.$$

Proof. By Lemma 3.8, we have

$$\| T_n - T \|_p = \left\| \sup_{\gamma\in\Gamma} \frac{S_{n,\gamma}}{n} - \sup_{\gamma\in\Gamma} I_\gamma \right\|_p \leqslant \left\| \sup_{\gamma\in\Gamma} \left| \frac{S_{n,\gamma}}{n} - I_\gamma \right| \right\|_p ,$$

and the result follows from Proposition 5.1. □

https://doi.org/10.1017/S0963548314000121 Published online by Cambridge University Press

Corollary 5.4 (L^p-convergence for worst-case `Find`). *Let $2 \leqslant p < \infty$, and suppose that the source-specific cost function β is ϵ-tame with $0 \leqslant \epsilon < 1/p$. Then, recalling* (5.3)–(5.5)*, the scale-normalized cost T_n of worst-case* `Find` *satisfies*

$$T_n \xrightarrow{L^p} T.$$

Proof. Combine Theorem 4.1 and Corollary 5.3. □

6. QuickSelect for fixed quantile(s)

6.1. QuickVal and QuickQuant

Fix $\alpha \in [0, 1]$ and let (m_n) be any sequence of integers satisfying $1 \leqslant m_n \leqslant n$ for every n and $m_n/n \to \alpha$ as $n \to \infty$. Consider the algorithm `QuickQuant`(n, α), defined as `Quick-Select`(n, m_n) applied to the first n keys. Vallée, Clément, Fill and Flajolet [21] introduced, and Fill and Nakama [8] further studied, an algorithm `QuickVal`(n, α), closely related to `QuickQuant`, described briefly as follows (see, *e.g.*, [8, Section 2.3] for a fuller description). While `QuickQuant`(n, α) searches successfully for the sample α-quantile among the first n seeds $U_1, \dots, U_n$, the algorithm `QuickVal`(n, α) searches (unsuccessfully, with probability 1) through the seeds $U_1, \dots, U_n$ for the value α.

Let $\gamma(\alpha) \in \Gamma$ be the infinite path from the root seed to seed α in the infinite binary search tree of seeds. By combining Theorem 4.1 and Remark 5.2, we obtain the following result.

Proposition 6.1 (L^p-convergence for `QuickVal`). *Let $2 \leqslant p < \infty$, and assume the source-specific cost function β is ϵ-tame with $0 \leqslant \epsilon < 1/p$. Then the cost V_n of* `QuickVal`(n, α) *satisfies*

$$\frac{V_n}{n} \xrightarrow{L^p} I_{\gamma(\alpha)} \text{ as } n \to \infty.$$

Remark 6.2. (a) This result is obtained effortlessly from our master Theorem 4.1 but is slightly weaker than Theorem 3.1 in [8] in two ways. First, we require $p \geqslant 2$, whereas only $p \geqslant 1$ is assumed in [8]. Second, our tameness hypothesis of $0 \leqslant \epsilon < 1/p$ is sufficient (but not necessary) for the hypothesis

$$\sum_{k \geqslant 1} (\mathbb{E} I_{p,\theta_k})^{1/p} < \infty$$

of [8, Theorem 3.1], $\gamma(\alpha) = (\theta_1, \theta_2, \dots)$ being the random path traversed by `QuickVal`(n, α). (Note the adjustment in indices needed to align notation with [8].) This sufficiency can be proved as follows. As argued at the start of the proof of Lemma 3.1,

$$I_{p,\theta_k} \leqslant \frac{2^{p\epsilon} c^p}{1 - p\epsilon} (R_{\theta_k} - L_{\theta_k})^{1-p\epsilon}.$$

But according to [8, Lemma 3.5] we have

$$\mathbb{E}(R_{\theta_k} - L_{\theta_k})^{1-p\epsilon} \leqslant \left(\frac{2-2^{-(1-p\epsilon)}}{2-p\epsilon}\right)^k,$$

and the fraction $[2-2^{-(1-p\epsilon)}]/(2-p\epsilon)$ here is strictly smaller than 1.

(b) Similarly, the next result is Theorem 4.1 in [8], but is proved there assuming only $p \geqslant 1$.

Theorem 6.3 (L^p-convergence for `QuickQuant`). *Let $2 \leqslant p < \infty$, and assume the source-specific cost function β is ϵ-tame with $0 \leqslant \epsilon < 1/p$. Then the cost Q_n of `QuickQuant`(n, α) satisfies*

$$\frac{Q_n}{n} \xrightarrow{L^p} I_{\gamma(\alpha)} \text{ as } n \to \infty.$$

Proof. For each n, let $\gamma_n \equiv \gamma_n(\alpha) = (\theta_{n,0}, \theta_{n,1}, \theta_{n,2}, \ldots, \theta_{n,|\gamma_n|})$ be the random path taken by `QuickQuant`(n, α) and let $\gamma \equiv \gamma(\alpha) = (\theta_0, \theta_1, \ldots)$ be the random path taken by `Quick-Val`(n, α). It follows from the random-paths extension of Proposition 5.1 mentioned in Remark 5.2 that

$$\left\|\frac{Q_n}{n} - I_{\gamma_n}\right\|_p \to 0.$$

Therefore, by the triangle inequality for L^p-norm, it suffices to show that

$$\|I_{\gamma_n} - I_\gamma\|_p \leqslant \sum_{k=0}^{\infty} \|\mathbf{1}(k < |\gamma_n|) I_{\theta_{n,k}} - I_{\theta_k}\|_p \to 0.$$

As an easy consequence of the strong law of large numbers, for each k we have almost surely that

$$\mathbf{1}(k < |\gamma_n|) I_{\theta_{n,k}} - I_{\theta_k} = 0 \quad \text{for all large } n.$$

Therefore by applying the dominated convergence theorem twice (first for expectation and then for counting measure), we finish the proof. It suffices to show that the upper bound in the inequality

$$\|\mathbf{1}(k < |\gamma_n|) I_{\theta_{n,k}} - I_{\theta_k}\|_p \leqslant 2\|\max_{\theta \in \Lambda_k} I_\theta\|_p$$

is summable over k. Indeed, by applying Lemmas 3.9 and 3.1 we get the bound

$$\|\max_{\theta \in \Lambda_k} I_\theta\|_p \leqslant 2^{k/p} \max_{\theta \in \Lambda_k} \|I_\theta\|_p \leqslant \left(\frac{c2^\epsilon}{1-\epsilon}\right)\left(\frac{p+1-p\epsilon}{2}\right)^{-k/p}.$$

Since $\epsilon < 1/p$ and $p \geqslant 2$, we have that $p + 1 - p\epsilon > 2$, which ensures summability over k. □

6.2. MultipleQuickQuant

We will be very brief in this subsection. Consult, for example, [16] for a description of the algorithm `MultipleQuickSelect`$(n; m_1, \dots, m_t)$ for finding simultaneously the keys of rank $m_1, \dots, m_t$ in an input sequence of length n.

Now fix a positive integer t and, for $i = 1, \dots, t$, values $\alpha^i \in [0,1]$ and sequences (m_n^i) of integers satisfying $1 \leqslant m_n^i \leqslant n$ and $m_n^i/n \to \alpha^i$ as $n \to \infty$. Consider the algorithm `MultipleQuickQuant`$(n; \alpha^1, \dots, \alpha^t)$, defined as `MultipleQuickSelect`$(n;\ m_n^1, \dots, m_n^t)$ applied to the first n keys. Let $\gamma \equiv \gamma(\alpha^1, \dots, \alpha^t)$ be the union of the sets $\gamma(\alpha^i)$ described just before Proposition 6.1, and let $I_\gamma := \sum_{\theta \in \gamma} I_\theta$. The next theorem generalizes Theorem 6.3.

Theorem 6.4 (L^p-convergence for `MultipleQuickQuant`). *Let $2 \leqslant p < \infty$, and assume the source-specific cost function β is ϵ-tame with $0 \leqslant \epsilon < 1/p$. Then the cost M_n of* `MultipleQuickQuant`$(n; \alpha^1, \dots, \alpha^t)$ *satisfies*

$$\frac{M_n}{n} \xrightarrow{L^p} I_\gamma \quad \text{as } n \to \infty.$$

In the interest of brevity, the proof is omitted. We also leave the statement (which by now should be rather obvious) and proof of L^p-convergence for worst-case `MultipleQuickSelect` to the reader.

7. Almost-sure convergence for worst-case `Find`

Theorem 7.1 (almost sure convergence for worst-case `Find`). *If the source-specific cost function β is ϵ-tame with $0 \leqslant \epsilon < 1/4$, then, recalling* (5.3)–(5.5)*, the scale-normalized cost T_n of worst-case* `Find` *satisfies*

$$T_n \to T \quad \text{a.s. as } n \to \infty.$$

Remark 7.2. Devroye [3] proved this almost sure convergence in the special case $\beta \equiv 1$ of key comparisons.

Proof of Theorem 7.1. For $0 \leqslant \ell < \infty$, let $\Gamma(\ell)$ be the set of 2^ℓ paths from the root to a node at level ℓ. For $0 \leqslant \ell < \infty$, define

$$T(\ell) := \max_{\gamma \in \Gamma(\ell)} I_\gamma, \qquad V(\ell) := \sum_{k > \ell} \max_{\theta \in \Lambda_k} I_\theta,$$
$$T_n(\ell) := \max_{\gamma \in \Gamma(\ell)} \frac{S_{n,\gamma}}{n}, \qquad V_n(\ell) := \sum_{k > \ell} \max_{\theta \in \Lambda_k} \frac{S_{n,\theta}}{n}.$$

Then we have the following inequalities for T and T_n:

$$T_n(\ell) \leqslant T_n \leqslant T_n(\ell) + V_n(\ell),$$
$$T(\ell) \leqslant T \leqslant T(\ell) + V(\ell),$$

and hence

$$\begin{aligned}|T_n - T| &\leqslant |T_n - T_n(\ell)| + |T_n(\ell) - T(\ell)| + |T(\ell) - T| \\ &\leqslant V_n(\ell) + |T_n(\ell) - T(\ell)| + V(\ell).\end{aligned}$$

We prove that $T_n \to T$ almost surely in three steps. First, we show for each fixed ℓ that $T_n(\ell) \to T(\ell)$ a.s. as $n \to \infty$, then we show that $V(\ell) \to 0$ a.s. as $\ell \to \infty$, and lastly we show that $\limsup_{\ell\to\infty} \limsup_{n\to\infty} V_n(\ell) = 0$ almost surely. Therefore the result follows from Lemmas 7.3–7.5. □

Lemma 7.3. *We have*

$$T_n(\ell) \to T(l) \quad \text{a.s. as } n \to \infty.$$

Proof. Let $\theta \in \Theta$. Condition on the random vector C_θ; recall that $S_{n,\theta}$ is then the sum of $n - \tau_\theta$ i.i.d. random variables and hence

$$\frac{S_{n,\theta}}{n} \to I_\theta \quad \text{a.s.} \tag{7.1}$$

Since (7.1) holds conditionally for (almost) every value of C_θ, it also holds unconditionally, by Tonelli's theorem. Therefore (summing over $\theta \in \gamma$), $S_{n,\gamma}/n \to I_\gamma$ a.s. for each path $\gamma \in \Gamma(\ell)$ because each such path is finite. Then, taking the maximum over the finite set of such paths, we conclude the desired result. □

Lemma 7.4. *If β is ϵ-tame with $0 < \epsilon < 1$, then*

$$V(\ell) \to 0 \quad \text{a.s. as } \ell \to \infty.$$

Proof. Let $p \in [1, \infty)$. Using Markov's inequality, for any $\eta > 0$ we have $\mathbb{P}(V(\ell) \geqslant \eta) \leqslant \mathbb{E}V(\ell)^p/\eta^p$. Therefore, by the first Borel–Cantelli lemma, it suffices to show that

$$\sum_{\ell=0}^{\infty} \mathbb{E}V(\ell)^p < \infty.$$

However, applying the definition of $V(\ell)$, the triangle inequality for (and continuity of) $\|\cdot\|_p$, the Max Lemma (Lemma 3.9), and Lemma 3.1, we find

$$\|V(\ell)\|_p \leqslant C \sum_{k>\ell} \left(\frac{2}{p+1-p\epsilon}\right)^{k/p},$$

for a constant C not depending on ℓ. This implies that

$$\begin{aligned}\sum_{\ell=0}^{\infty} \mathbb{E}V(\ell)^p &\leqslant C^p \sum_{\ell=0}^{\infty} \left[\left(\frac{2}{p+1-p\epsilon}\right)^{\ell/p} \sum_{k=1}^{\infty} \left(\frac{2}{p+1-p\epsilon}\right)^{k/p}\right]^p \\ &= C^p \left[\sum_{\ell=0}^{\infty} \left(\frac{2}{p+1-p\epsilon}\right)^{\ell}\right] \left[\sum_{k=1}^{\infty} \left(\frac{2}{p+1-p\epsilon}\right)^{k/p}\right]^p.\end{aligned}$$

https://doi.org/10.1017/S0963548314000121 Published online by Cambridge University Press

Choosing p so that $\epsilon < (p-1)/p$, we have

$$\sum_{\ell} \mathbb{E} V(\ell)^p < \infty. \qquad \square$$

Lemma 7.5. *If β is ϵ-tame with $0 \leqslant \epsilon < 1/4$, then*

$$\limsup_{\ell\to\infty} \limsup_{n\to\infty} V_n(\ell) = 0 \quad \text{a.s.}$$

Proof. First observe that $V_n(\ell) \leqslant V(\ell) + \widetilde{V}_n(\ell)$, with

$$\widetilde{V}_n(\ell) := \sum_{k>\ell} \max_{\theta\in\Lambda_k} \left[\left| \frac{S_{n,\theta}}{n} - \left(\frac{n-\tau_\theta}{n}\right) I_\theta \right| \mathbf{1}(\tau_\theta < n) \right].$$

In light of Lemma 7.4, it is sufficient to prove that $\limsup_{\ell\to\infty} \limsup_{n\to\infty} \widetilde{V}_n(\ell)$ vanishes almost surely. In fact we will prove the stronger result that $\widetilde{V}_n(\ell) \to 0$ a.s. as $n \to \infty$ for each ℓ. Using the technique of Markov's inequality and first Borel–Cantelli lemma as in Lemma 7.4, it suffices to show that for some $2 \leqslant p < \infty$ we have

$$\sum_{n=1}^{\infty} \mathbb{E}\, \widetilde{V}_n(\ell)^p < \infty.$$

Arguing as in the proof of Lemma 7.4, we have the following bound for any $p \in [2,\infty)$:

$$\mathbb{E}\widetilde{V}_n(\ell)^p \leqslant \left[\sum_{k>\ell} 2^{k/p} \max_{\theta\in\Lambda_k} h(n,p,\theta) \right]^p,$$

where $h(n,p,\theta)$ is the first term on the right in inequality (4.3). By the arguments following (4.3) in the proof of Theorem 4.1, we have the following bound on $h(n,p,\theta)$:

$$h(n,p,\theta)^p \leqslant c_p\, \mathbb{E}\left[\mathbf{1}(\tau_\theta < n) \left(\frac{2^p I_{p,\theta}}{n^{p-1}} + \frac{I_{2,\theta}^{p/2}}{n^{p/2}} \right) \right].$$

So we have the following bound:

$$\mathbb{E}\widetilde{V}_n(\ell)^p \leqslant c_p \left[\sum_{k>\ell} \left(2^k\, \mathbb{E}\left[\mathbf{1}(\tau_{\theta_k} < n) \left(\frac{2^p I_{p,\theta_k}}{n^{p-1}} + \frac{I_{2,\theta_k}^{p/2}}{n^{p/2}} \right) \right] \right)^{1/p} \right]^p,$$

where $\theta_k \in \Lambda_k$ is chosen arbitrarily for each k. Choose p such that $\epsilon < 1/p < 1/4$ and then a such that $1 < a < (p/2) - 1$. After factoring out n^{-a}, we are left with the following bound for $\mathbb{E}\widetilde{V}_n(\ell)^p$:

$$\mathbb{E}\widetilde{V}_n(\ell)^p \leqslant c_p n^{-a} \left[\sum_{k>\ell} \left(2^k\, \mathbb{E}\left[\frac{2^p I_{p,\theta_k}}{\tau_{\theta_k}^{p-1-a}} + \frac{I_{2,\theta_k}^{p/2}}{\tau_{\theta_k}^{(p/2)-a}} \right] \right)^{1/p} \right]^p. \tag{7.2}$$

Note that the only dependence on n in this bound for $\mathbb{E}\widetilde{V}_n(\ell)^p$ is in the factor n^{-a}. Therefore, since $a > 1$ it suffices to prove that, for any ℓ, the sum over k in (7.2) is finite.

https://doi.org/10.1017/S0963548314000121 Published online by Cambridge University Press

This can be done by two applications of Lemma 3.6, just as in the proof of Theorem 4.1; the remaining details are routine and omitted. □

Note that the tameness condition $\epsilon < 1/4$ imposed on β for almost sure convergence of worst-case `QuickSelect` in Theorem 7.1 matches the condition imposed for almost sure convergence of `QuickVal` proved by Fill and Nakama in [8, Theorem 3.4] for fixed quantile α.

Acknowledgements

We thank two anonymous referees for their helpful comments.

References

[1] Billingsley, P. (1995) *Probability and Measure*, third edition, Wiley Interscience.
[2] Devroye, L. (1984) Exponential bounds for the running time of a selection algorithm. *J. Comput. System Sci.* **29** 1–7.
[3] Devroye, L. (2001) On the probabilistic worst-case time of 'Find'. *Algorithmica* **31** 291–303.
[4] Devroye, L. and Fawzi, O. (2010) Simulating the Dickman distribution. *Statist. Probab. Lett.* **80** 242–247.
[5] Fill, J. A. (2013) Distributional convergence for the number of symbol comparisons used by QuickSort. *Ann. Appl. Probab.* **23** 1129–1147.
[6] Fill, J. A. and Huber, M. L. (2010) Perfect simulation of Vervaat perpetuities. *Electron. J. Probab.* **15** 96–109.
[7] Fill, J. A. and Nakama, T. (2010) Analysis of the expected number of bit comparisons required by Quickselect. *Algorithmica* **58** 730–769.
[8] Fill, J. A. and Nakama, T. (2013) Distributional convergence for the number of symbol comparisons used by QuickSelect. *Adv. Appl. Probab.* **45**, 425–450.
[9] Grübel, R. (1998) Hoare's selection algorithm: A Markov chain approach. *J. Appl. Probab.* **35** 36–45.
[10] Grübel, R. and Rösler, U. (1996) Asymptotic distribution theory for Hoare's selection algorithm. *Adv. Appl. Probab.* **28** 252–269.
[11] Hoare, C. A. R. (1961) Find (algorithm 65). *Commun. Assoc. Comput. Mach.* **4** 321–322.
[12] Hwang, H. and Tsai, T. (2002) Quickselect and the Dickman function. *Combin. Probab. Comput.* **11** 353–371.
[13] Knuth, D. E. (1972) Mathematical analysis of algorithms. In *Information Processing 71: Proc. IFIP Congress, Ljubljana, 1971*, North-Holland, pp. 19–27.
[14] Kodaj, B. and Móri, T. F. (1997) On the number of comparisons in Hoare's algorithm 'FIND'. *Studia Sci. Math. Hungar.* **33** 185–207.
[15] Lent, J. and Mahmoud, H. M. (1996) Average-case analysis of multiple Quickselect: An algorithm for finding order statistics. *Statist. Probab. Lett.* **28** 299–310.
[16] Mahmoud, H. M. and Smythe, R. T. (1998) Probabilistic analysis of multiple Quickselect. *Algorithmica* **22** 569–584.
[17] Mahmoud, H. M., Modarres, R. and Smythe, R. T. (1995) Analysis of Quickselect: An algorithm for order statistics. *RAIRO Informatique Théorique et Applications* **29** 255–276.
[18] Mitzenmacher, M. and Upfal, E. (2005) *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*, Cambridge University Press.
[19] Prodinger, H. (1995) Multiple Quickselect: Hoare's Find algorithm for several elements. *Inform. Process. Lett.* **56** 123–129.

[20] Rosenthal, H. P. (1970) On the subspaces of L^p ($p > 2$) spanned by sequences of independent random variables. *Israel J. Math.* **8** 273–303.
[21] Vallée, B., Clément, J., Fill, J. A. and Flajolet, P. (2009) The number of symbol comparisons in QuickSort and QuickSelect. In *36th International Colloquium on Automata, Languages and Programming: ICALP 2009*, Part I (S. Albers *et al.*, ed.), Vol. 5555 of *Lecture Notes in Computer Science*, Springer, pp. 750–763.
[22] Williams, D. (1991) *Probability with Martingales*, Cambridge University Press.