JIRSS (2004)
Vol. 3, No. 2, pp 271-296

Quickselect Revisited

Uwe Rösler

Mathematisches Seminar, Christian-Albrecht-Universität zu Kiel, Ludewig-Meyn-Strasse 4, 24098 Kiel, Germany. (roesler@math.uni-kiel.de)

Abstract. We give an overview of the running time analysis of the random divide-and-conquer algorithm Find or Quickselect. The results concern moments, distribution of Find's running time, the limiting distribution, a stochastic bound and the key: a stochastic fixed point equation.

1 Introduction to Find

Some forty years ago, Hoare [13], [15] introduced the sorting algorithm 64: Quicksort, now a widely applied and well-studied sorting method. Quicksort is the first divide-and-conquer algorithm with a complete running time analysis, see Rösler [34]. This algorithm served during the last ten years as the major divide and conquer algorithm to develop and test new methods for the analysis, like the contraction method [36], [38], generating function equations [19], and stochastic fixed point equations [35].

Also some forty years ago Hoare [14], [27] introduced the algorithms 65: Find, sometimes also called Quickselect, for finding

Received: October 2003, Revised: March 2004

Key words and phrases: analysis of algorithms, asymptotic distribution, FIND, FIND process, QUICKSELECT, stochastic algorithm, stochastic fixed point equation.

the l-th smallest element out of n different numbers. (QUICKSELECT might be the more preferred term in view of the analogy to QUICKSORT.) Over the last 10 years many papers have been written on various aspects of FIND. Here we present major results, together with some of the problems inspired by it, and give an outlook of the work in progress.

The algorithm FIND: The input of the selection algorithm FIND(S, l) is a set S of n different real numbers and the output is the l-th smallest element of the input. If $n = 1$, the algorithm returns the single element of S. If $n > 1$, FIND proceeds as follows:

1. Choose a pivot element x uniformly at random from S.
2. Determine the two sets $S_< = \{s \in S : s < x\}$, and $S_> = \{s \in S : s > x\}$.
3. If $v := |S_<| + 1 = l$, return the pivot element x. If $v > l$, then continue with FIND$(S_<, l)$, and if $v < l$, continue with FIND$(S_>, l - v)$.

Obviously this algorithm terminates and returns the unique l-th smallest element of the data set S.

Our version of FIND is a random algorithm, the randomness is in the algorithm itself; we call this internal randomness. The running time of the algorithm is always a random variable.

There are deterministic versions of FIND selecting the pivot element by a deterministic rule. For example, if the input S is given in the form of a list, then always choose the first (or last or whatever ...) element from this list. The running time of FIND is determined on the order of the input list of S. If this order bears some randomness, then the running time is also random. If the input is a random permutation (with uniform distribution) of S, then the running time of the deterministic version is, again, a random variable with the same distribution as our (random) FIND.

One of the advantages of our FIND is that the given results depend only on the cardinality $|S| = n$ of the input set and not, for example, on the actual order (sequence in an array) in which S is stored on the computer. For this reason we stay exclusively with the random version as given.

Our version of FIND splits the set S into 3 sets $S_<, \{x\}, S_>$. (Here x is the pivot element. One of the sets may be empty.) The 2-version splits S into the sets $S_\leq = \{s \in S : s \leq x\}$, and $S_>$ (or equivalently

$S_<$, and $S_\geq = \{s \in S : s \geq x\}$). For the mathematical consequences and different asymptotic behaviour, this is to be compared with the discussion in the section on process convergence.

There are other good random rules to choose the pivot elements. The $2k+1$-median version of FIND, in Grübel [10], uses $2k+1$ random elements uniformly drawn from S and takes their median as the pivot element. Instead of taking the median of the $2k+1$, it may be better to take an adapted choice, the one with rank close to $\frac{(2k+1)l}{n}$, Martínez, Panario, Viola [29] pursue this line. An earlier similar approach is taken in Floyd and Rivest [7]. Shen and Chen [41] treat the parallel computing.

2 Running time analysis

Of main interest is an analysis of the running time of FIND, the time to select the l-th smallest in the input set S. This time depends on the computer, the number of operations, and so on. Sedgewick [40] considers these questions in his thesis on QUICKSORT but we will not look at this in this paper.

Basically, it suffices to consider only the number of comparisons performed by FIND. The average running time will be proportional to the number of keys, with a proportionality factor depending on the actual performing computer and the program code. For the (asymptotic) number of comparisons of FIND for various $n = |S|$, or for contrasting the running time number of comparisons to competing algorithms, we restrict ourselves to the random number $X_{n,l}$ of comparisons for FIND(S, l) in order to find the l-th smallest in a set S of n different numbers. (The distribution depends only on $|S| = n$, but not on the actual order of the input.)

The running time random variables $X_{\cdot,\cdot}$ satisfy the following recursion, which is basic for all the following analysis. After running one step of FIND, the random variable $X_{n,l}$ is the sum of $n-1$ comparisons in order to obtain the sets $S_<, S_>$, plus the running time for $S_<$ (respectively $S_>$). If the pivot element is the U-th smallest in the set S we add the running time of FIND$(S_<, l)$ in case $U > l$, or of FIND$(S_>, l-U)$ in case $U < l$. Notice the independence of the realisation of these random variables from the choice of the pivot element. Of course, their distribution depends on U, and the size of $S_<,\ S_>$.

Mathematically, the above description comes down to the recur-

sive equation

$$X_{n,l} \stackrel{\mathcal{D}}{=} 1_{U>l} X_{U-1,l} + 1\!\!1_{U<l} \overline{X}_{n-U,l-U} + n - 1, \tag{1}$$

for $1 \le l \le n$. Here $X_{i,j}, \overline{X}_{i,j}, U,\ 1 \le j \le i < n$, are independent random variables on some probability space $(\Omega, \mathcal{A}, P)$. The distribution of U is uniform on $\{1, \ldots, n\}$, the distribution of $X_{i,j}$ is the same as that of $\overline{X}_{i,j}$ and is recursively given by the formula (1), and the boundary condition $X_{1,1} \equiv 1$. The symbol $\stackrel{\mathcal{D}}{=}$ denotes equality of distribution (on $\mathbb{R}$).

For all our purposes the distribution of the random variables $X_{\cdot,\cdot}$ is more important than the realisation as random variables. For that reason we consider the equation (1) for distributions only. Notice that a formulation with random variables would require some more notation and processes on trees, the weighted branching processes, see Rösler[35], and Rösler and Rüschendorf [38].

Now we look at the running time analysis.

Best case: The sharp lower bound of $X_{n,l}$ is $n-1$. This happens if the pivot element is incidentally the l-th smallest.

Worst case: The sharp upper bound is $\frac{(n-1)n}{2}$. This happens if, for each recall of FIND, $U \neq l$.

Expectation: The recursive equation for the expectation $EX_{n,l}$ is

$$EX_{n,l} = \frac{1}{n} \sum_{l<u\le n} EX_{u-1,l} + \frac{1}{n} \sum_{1\le u<l} EX_{n-u,l-u} + n - 1.$$

The solution is already given in Knuth [22]:

$$EX_{n,l} = 2(n + 3 + (n+1)H_n - (l+2)H_l - (n+3-l)H_{n-l+1}),$$

where

$$H_n = \sum_{1\le i\le n} \frac{1}{i}$$

are the harmonic numbers.

Higher Moments: How concentrated is the random variable $X_{n,l}$ around the expectation $a_{n,l} := EX_{n,l}$? Higher moments of $X_{n,l}$ provide some information, e.g. via the Markov inequality. The recursion (1) provides a recursion formula for all moments.

The recursion formula for the variance is

$$\begin{aligned}\mathrm{Var}X_{n,l} &= \frac{1}{n}\sum_{l<u\le n}\mathrm{Var}X_{u-1,l}+\frac{1}{n}\sum_{1\le u<l}\mathrm{Var}X_{n-u,l-u}\\ &\quad +E(n-1-a_{n,l}+\mathbb{1}_{l<U}a_{U-1,l}+\mathbb{1}_{U<l}a_{n-U,l-U})^2.\end{aligned}$$

The explicit solution "requires an incredible amount of calculations," mostly done by the symbolic manipulation system MAPLE under "careful human guidance," as stated in Kirschenhofer and Prodinger [17]:

$$\begin{aligned}\mathrm{Var}X_{n,l} &= -2(n+1)(3n+8)H_n^2\\ &\quad +8H_nH_l(n(l+2)+2)+8H_nH_{n+1-j}(n(n+3-l)+2)\\ &\quad +2\frac{-(2n^2+n-9)l^2+(n+1)(2n^2+n-9)l+8(n+1)}{l(n+1-l)}H_n\\ &\quad -2l(l-1)H_l^2-2(n+1-l)(n-l)H_{n+1-l}^2\\ &\quad +4(l^2-(n+1)l-n^2-5n-8)H_lH_{n+1-l}\\ &\quad -\frac{2}{l(n+1-l)}(-(2n+3)l^3+(2n^2+7n+24)l^2\\ &\quad -(n+1)(2n+21)l+8(n+1))H_l\\ &\quad -\frac{2}{l(n+1-l)}((2n+3)l^3-(4n^2+8n-15)l^2\\ &\quad\quad +(n+1)(2n^2+3n-18)l+8(n+1))H_{n+1-l}\\ &\quad +2(n+1)(n+6)H_n^{(2)}\\ &\quad -2(l^2+5l+8)H_l^{(2)}\\ &\quad -2(l^2-(2n+7)l+n^2+7n+14)H_{n+1-l}^{(2)}\\ &\quad +(10l^4-20(n+1)l^3+(9n^2+31n-6)l^2\\ &\quad\quad +(n^2-11n+16)(n+1)l+32)/(2l(n+1-l))\\ &\quad +4n\Big((n+1-l)\sum_{k=1}^{l}\frac{H_{n-k}}{k}+l\sum_{k=1}^{n+1-l}\frac{H_{n-k}}{k}\Big),\end{aligned}$$

where

$$H_n^{(2)}=\sum_{1\le k\le n}\frac{1}{k^2}$$

are the harmonic numbers of the second order.

There seems to be no point in giving more explicit formulas for even higher moments. After all, we are only interested in the asymptotic behaviour of the moments. Let $l=l_n$ be a function of n, and $\frac{l}{n}$

converge to some $t \in [0,1]$, as n approaches infinity. Then

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} EX_{n,l} &= 2 - 2t\ln t - 2(1-t)\ln(1-t) \\
\lim_{n\to\infty} \frac{\mathrm{Var}X_{n,l}}{n^2} &= -2t^2\ln^2 t + 4(t^2-t-1)\ln t \ln(1-t) \\
&\quad -2(1-t)^2\ln^2(1-t) - 4t\ln t \\
&\quad -4(1-t)\ln(1-t) - 4t\ \mathrm{dilog}(t) \\
&\quad -4(1-t)\ \mathrm{dilog}(1-t) \\
&\quad +\frac{20}{3}\pi^2 t(1-t) + 5t(1-t) + \frac{1}{2},
\end{aligned}$$

where $\mathrm{dilog}(t) = \int_1^t \frac{\ln u}{1-u}\, du$. The asymptotic formula given above for the variance is derived by Kirschenhofer and Prodinger [17] via combinatorial and generating function methods. Another explicit representation was given by Paulsen [32] via differential equations and probabilistic arguments.

Devroye [4] gives (more detailed) estimates for the tail of $\frac{X_{n,l}}{n} = C_{n,l},\ 1 \le l \le n$:

$$\begin{aligned}
EC_{n,l} &\le 4, \qquad E(C_{n,l}^k) \le \mathrm{const}(k) n^k, \\
P(C_{n,l} \ge u) &\le \left(\frac{3}{4}\right)^{u(1+o(1))},
\end{aligned}$$

for $k \in I\!N$. The idea is a stochastic upper bound of $C_{n,l}$. Compare with a similar idea used by Grübel and Rösler [8] in Lemma 9 for an upper exponential tail.

Now fix l, and let n tend to infinity. Then (as noted in [16] [24])

$$\frac{X_{n,l}}{n} - 1 \to_{n\to\infty} D;$$

the limit D is distributed like the Dickman random variable. One way of characterizing the Dickman distribution is via the solution of the fixed point equation

$$D \stackrel{\mathcal{D}}{=} UD + U,$$

where U, D are independent and U is uniformly distributed on $[0,1]$. (For more on the Dickman distribution see the references in [16].)

3 Miscellaneous

Recalls: Let $R_{n,l}$ be the number of recalls for FIND(n,l). This random variable satisfies the recursive equation

$$R_{n,l} \stackrel{\mathcal{D}}{=} \mathbb{1}_{U_n>l} R_{U_n-1,l} + \mathbb{1}_{U_n<l} \overline{R}_{n-U_n,l-U_n} + 1, \tag{2}$$

for $1 \le l \le n \in \mathbb{N}$. Here $R_{i,j}, \overline{R}_{i,j}, U_n,\ 1 \le j \le i < n$, are independent random variables on some probability space. The distribution of U_n is uniform on $\{1,\dots,n\}$, the distribution of $R_{i,j}$ is the same as that of $\overline{R}_{i,j}$, and is recursively given by the formula (2), and the boundary condition $R_{1,1} \equiv 0$.

The distributional equation (2) leads to recursive equations for the expectation $ER_{n,l}$ and the variance $\mathrm{Var}R_{n,l}$. The solution, derived by combinatorial methods using Maple, is given by Kirschenhofer and Prodinger [17]:

$$\begin{aligned} ER_{n,l} &= H_l + H_{n-l} - 1, \\ \mathrm{Var}R_{n,l} &= \frac{2(n+1)}{l(n+1-l)} H_n + (1 - \frac{2(n+1)}{l(n+1-l)})(H_l + H_{n+1-l}) \\ &\quad -(H_l^{(2)} + H_{n+1-l}^{(2)}) + \frac{2(n+1)}{l(n+1-l)} + 2. \end{aligned}$$

Joint distributions: Let J be a subset of $\{1,\dots,n\},\ n = |S|$ and let $X_{n,J}$ be the number of comparisons in order to find simultaneously all j-th smallest elements, $j \in J$, out of S. The algorithm FIND can be suitably adapted: split only a set if necessary. In other words, use all the splittings done by QUICKSORT, cf. Section 4, which are necessary for the given problem. This procedure is called MULTIPLE QUICKSELECT. Again the joint distribution of $X_{n,J}$ satisfies a recursive equation. We skip presenting this relation to avoid the introduction of new notation.

Define the grand average $X_n^{(p)},\ p \in \mathbb{N}$, that is,

$$X_n^{(p)} := \frac{1}{\binom{n}{p}} \sum_J X_{n,J},$$

where the sum is over all $J \subset \{1,\dots,n\}$ with p elements. Analogously define the number of recalls $R_{n,J}$ and $R_n^{(p)}$.

The recursive equations admit an explicit solution for the expectation, given in Prodinger [33]:

$$EX_{n,J} = 2n + j_p - j_1 + 2(n+1)H_n - 2(j_1+2)H_{j_1}$$

$$
\begin{aligned}
& -2(n+3-j_p)H_{n+1-j_p} - \sum_{i=2}^{p}(j_i+4-j_{i-1})H_{j_i+1-j_{i-1}} \\
& +8p-2, \\
ER_{n,J} &= H_{j_1} + H_{n+1-j_p} + 2\sum_{j=1}^{p} H_{j_i+1-j_i} - 2p + 1, \\
EX_n^{(p)} &= 2n + j_p - j_1 + 2(n+1)H_n - 2(j_1+2)H_{j_1} \\
& -2(n+3-j_p)H_{n+1-j_p} \\
& -\sum_{i=2}^{p}(j_i+4-j_{i-1})H_{j_i+1-j_{i-1}} + 8p - 2, \\
ER_n^{(p)} &= (H_{n+1}-H_p)\frac{2p(n+1)^2}{(n+2-p)(n+1-p)} - \frac{n(2p-1)+p}{n+2-p},
\end{aligned}
$$

where $j_1 < j_2 < \cdots < j_p$ are the elements of J, $|J| = p \in \mathbb{N}$. Explicit solutions for the variance are in Panholzer and Prodinger [30].

Choice of the pivot element: Blum, Floyd, Pratt, Rivest, and Tarjan [3] consider the algorithm PICK instead of FIND. The point is a more careful choice of the pivot element. They were able to give an upper bound—no more than n 5.4305... comparisons are necessary to find the l-th smallest out of n. This bound was sharpened in a series of papers: Floyd and Rivest [7], Schöenhage, Peterson and Pippenger [39], and Yap [42].

We can improve the performance of FIND by a suitable choice of the pivot element. The first try would be the $2k+1$-median choice. Choose at random $2k+1$ elements of S and take the median as pivot element. For fixed k the additional amount of comparisons is negligible.

The expectations $ER_{n,l}$, and $EX_{n,l}$ for the median of 3-version are given in Panholzer and Prodinger [30], and in Kirschenhofer, Prodinger and Martínez [18]. The general $2k+1$-version was treated by Panholzer [31]. The techniques and final results for the asymptotics are similar to those in FIND (as given in the sequel), cf. Grübel [10].

An adaptive choice of the pivot out of $2k+1$ would be even better. The idea is a suitable choice of the pivot in the $2k+1$ sample, basically the one with a rank close to the relative l-th rank out of n, proposed in Martínez, Panario, Viola [29]. The advantage of an adapted choice is a continuation of FIND with small sets, at least with high probability. This leads to the idea of choosing two pivot elements, one larger and

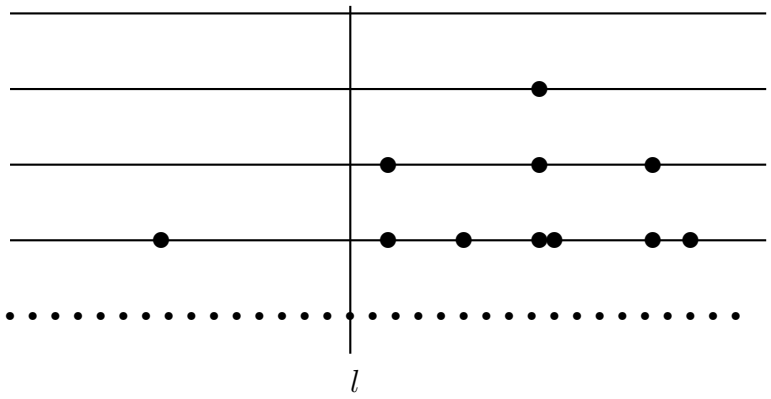
one smaller than the target (with high probability). Then recall the algorithm for the set of elements between the two pivots, see Floyd and Rivest [7]. For parallel selection see [41].

Choosing the pivot in a suitable way adds more comparisons to FIND. Martínez and Roura [28] studied the contribution of sampling to the moments, e.g. the variance, in some generality.

4 Convergence as a process

Asymptotic results are often easier to obtain than the exact ones in a discrete setting. One strategy is to first obtain (explicit) results on $X_{n,l}$ and then go to the asymptotics. We took that point of view in the previous sections for the first and second moment. Another strategy is to do first the asymptotics and then look for corresponding results. This approach often works better, as the limiting object often has preferable properties, which are not convoluted by discretisation. We come now to the process convergence of $(X_{n,l})_l$, as $n \to \infty$, to the FIND process Z as proved in Grübel, Rösler [8].

Why do we expect a limiting process? The sorting algorithm QUICKSORT splits randomly the input set S into the sets $S_<, \{ \text{ pivot } \}$, $S_>$ in this order. Then QUICKSORT is called recursively and each set is split until only one element or emplty sets are left. The output is an ordered list.



However, for FIND(n, l) we do not have to sort the whole input set. For finding the l-th smallest of n it suffices to split only the sets containing the l-th smallest element. In the picture this is indicated by l. The average running time of FIND is proportional to the length of all sets containing l.

For the mathematical description we use the tree structure for indexing the intervals. Compare with Rösler and Rüschendorf [38] for a more general embedding into weighted branching processes, and respectively random dynamical systems.

The Find process: Let V be the index set $\{0,1\}^* := \bigcup_{n=0}^{\infty}\{0,1\}^n$ of all finite 0–1 sequences. We use by convention $\{0,1\}^0 := \{\emptyset\}$. The empty set is called the root of V and denotes the 0-1 sequence with no 0's and no 1's. We use $v = v_1v_2\dots v_n \in \{0,1\}^n, |v| = n$ and $v_{|m} = v_1v_2\dots v_m \in \{0,1\}^m$, for $m \le n$. The index set is a natural tree by descendence, $v0$ and $v1$ are children of $v \in V$ and so on.

Let $(\Omega, \mathcal{A}, P)$ be a probablitlity space and $U(v) : \Omega \to [0,1],\ v \in V$, be independent uniformly distributed random variables. Let E be the set of half open intervals $[a,b),\ 0 \le a \le b \le 1$, of the unit interval, closed on the left and open on the right. The symbol $|I|$ denotes the length of the interval I. Define random maps $T(v) = (T_0(v), T_1(v)) : \Omega \to (E \times E)^E$ which split half open intervals into 2 half open intervals,

$$\begin{aligned}[a,b) \mapsto T(v)([a,b)) &= (T_0(v)([a,b)), T_1(v)([a,b))) \\ &= ([a, a+(b-a)U(v)), [a+(b-a)U(v), b)).\end{aligned}$$

(The formally correct writing including ω is

$$\begin{aligned}T(v)(\omega)([a,b)) &= (T_0(v)(\omega)([a,b)), T_1(v)(\omega)([a,b))) \\ &= ([a, a+(b-a)U(v)(\omega)), [a+(b-a)U(v)(\omega), b)).\end{aligned}$$

We drop here and in the sequel the stipulation $\omega \in \Omega$ for simplicitiy, as is common in probability theory.

Define recursively the map $L : \Omega \to E^V$ via $L(\emptyset) = [0,1) \in E$, and

$$L(vi) = T_i(v)(L(v)),$$

$v \in V, i = 0, 1$. In a more elaborate form, $v = v_1v_2\dots v_n$,

$$L(v) = T_{v_n}(v_{|n-1})(T_{v_{n-1}}(v_{|n-2})(\dots(T_{v_1}(\emptyset)([0,1)))\dots).$$

Define the process $Z_n : \Omega \to \mathbb{R}^{[0,1]},\ n \in \mathbb{Z}^+$ via

$$Z_n(t) = \sum_{|v|\le n} \mathbb{1}_{t\in L(v)}|L(v)|,$$

for $t \in [0,1)$, and $Z_n(1) = \lim_{t\uparrow 1} Z_n(t)$. The Z_n random variables satisfy the (backward, see [38]) recursive equations

$$Z_n(t) = 1 + 1_{U\ge t}UZ_{n-1}^{(0)}\Big(\frac{t}{U}\Big) + 1_{U<t}(1-U)Z_{n-1}^{(1)}\Big(\frac{t-U}{1-U}\Big),$$

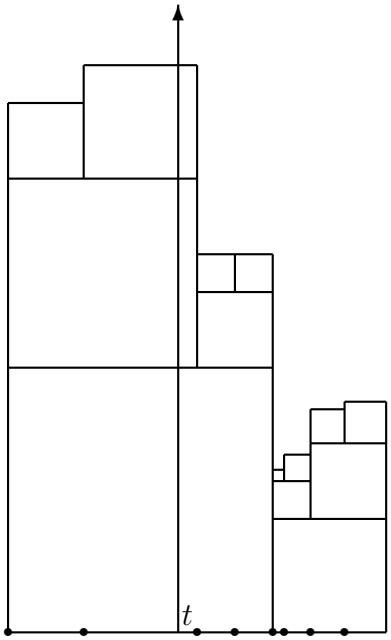
where $t \in [0,1]$, $U = U(\emptyset)$ and $Z_{n-1}^{(i)}$ denotes the Z_{n-1} process for the tree with the starting state $[0,1)$ on the root i and of height $n-1$, compare with the picture. (More formally,

$$Z_{n-1}^{(0)}(\frac{t}{U}) := \sum_{|v|\leq n-1} \mathbb{1}_{t\in L(0v)} \frac{|L(0v)|}{U},$$

and

$$Z_{n-1}^{(1)}(\frac{t-U}{1-U}) := \sum_{|v|\leq n-1} \mathbb{1}_{t\in L(1v)} \frac{|L(1v)|}{1-U}).$$

Here is a picture of Z_n, the graph $(t, Z_n(t))_t$ is given as the upper line.



The process $Z : \Omega \to I\!R^{[0,1]}$

$$Z(t) = \lim_{n\to\infty} Z_n(t) = \sum_{v\in V} \mathbb{1}_{t\in L(v)}|L(v)|$$

is the FIND process [8].

The FIND process is well defined since $Z_n(t)$ is monoton increasing in n. The FIND random variable Z takes values in $D = D([0,1])$, the set of right continuous functions $f : [0,1] \to I\!R$ with finite left hand

limits. All random variables $Z(t)$ have finite exponential moments uniformly in t. Consequently all moments of $Z(t)$ exist and are finite.

The FIND process satisfies the (backward [38]) recurrence relation

$$Z(t) = 1 + 1_{U>t} U Z^{(0)}\Big(\frac{t}{U}\Big) + 1_{U\le t}(1-U) Z^{(1)}\Big(\frac{t-U}{1-U}\Big), \qquad (3)$$

where $t \in [0,1]$, $U = U(\emptyset)$ and $Z^{(i)}(t)$ denotes the FIND process for the tree with starting state $[0,1)$ on the root $i \in \{0,1\}$. (More formally, let $Q^{(i)}(t) = \sum_{v\in V} 1\!\!1_{t\in L(iv)}|L(iv)|$ and define $Z^{(0)}(t) := \frac{1}{U(\emptyset)} Q^{(0)}(tU(\emptyset))$ and $Z^{(1)}(t) := \frac{1}{1-U(\emptyset)} Q^{(1)}(U(\emptyset) + t(1-U(\emptyset)))$.)

Process Convergence: Once we have identified the limiting process, we come now to the process convergence. Grübel and Rösler [8] take a quite probabilistic view showing Skorodhod convergence in D of the processes itself.

The Skorodhod distance $\rho(f,g)$ of two functions f, g in the space D is the minimal $a \in I\!R$ such that there is a bijective increasing function $\lambda : [0,1] \to [0,1]$ on the unit interval to itself satisfying

$$\frac{|\lambda(s) - \lambda(t)|}{|s-t|} \le a, \qquad |f(t) - g(\lambda(t))| \le a,$$

for all $s \ne t \in [0,1]$.

On some (new) probability space $(\Omega, \mathcal{A}, P)$ Grübel and Rösler [8] define a specific sequence $Y_n : \Omega \to D$ of D-valued processes such that $Y_n(\frac{l-1}{n})$ has the same distribution as $\frac{X_{n,l}}{n}$ for the 2-version of FIND. Theorem 4 of [8] shows Skorodhod convergence

$$Y_n \to_{n\to\infty} Z$$

of Y_n to the FIND process. (Taking the 2-version is essential, there is no Skorodhod convergence for the 3-version. The trouble is basically functions like $1\!\!1_{[1/n,2/n)} \in D$ which converge in n pointwise to 0, but do not converge in Skorodhod metric.)

Skorodhod convergence does not imply weak convergence $Y_n(t) \to Z(t)$ in general. However in our case, almost every path $[0,1] \ni t \mapsto Z(t)(\omega)$ is continuous at $t \in [0,1]$ (for almost all $\omega \in \Omega$). Theorem 8 in [8] states the weak convergence

$$\frac{X_{n,l_n}}{n} \to_n Z(t) \qquad (4)$$

for all sequences l_n satisfying $\frac{l_n}{n} \to_{n\to\infty} t \in [0,1]$. This result is true for the 2-version and the 3-version of FIND.

Since $\frac{X_{n,l_n}}{n}$ and $Z(t)$ have finite exponential moments, standard arguments on uniformly integrable random variables provide the convergence of any moment or exponential moment,

$$E\Big(\frac{X_{n,l_n}}{n}\Big)^a \to_{n\to\infty} EZ^a(t), \qquad E\Big(e^{a\frac{X_{n,l_n}}{n}}\Big) \to_{n\to\infty} Ee^{aZ(t)}.$$

5 Asymptotic moments

In this section we give some consequences of the process convergence for moments.

Expectation: Grübel [9] gives an alternative proof of (4) via Markov chain methods. Kodaj and Mori [23] give the rate

$$l_1\left(\frac{X^k_{n,l}}{n}, Z^k(\frac{l-1/2}{n})\right) \approx \ln\frac{l^5(n-l+1)^5}{n^2},$$

for the distance of $\frac{X^k_{n,l}}{n}$ to $Z^k(\frac{l-1/2}{n})$ in Wasserstein metric

$$l_1(\mu,\nu) = \inf \|X-Y\|_1,$$

where $\|.\|_1$ denotes the L_1-norm and the infimum is taken over all random variables X, Y on some probability space such that the distribution of X is μ, and that of Y is ν. We use here $l_1(X,Y)$ for the l_1-distance of the distribution of X to that of Y. (Therefore, $l_1(X,Y) = \int |F(x)-G(x)|\,dx$, where F, G are the distribution functions for X, Y.)

Furthermore, Kodaj and Mori [23] give a limit process which provides for the 1-dimensional marginals an upper bound in stochastic order. Similar results were given for higher moments and for other (Orlicz) distances.

Higher moments: We are now able to give statements on higher moments of X via higher moments of the FIND process Z. Let $m_k : [0,1] \to I\!R,\ k \in I\!N$, be the moment generating function

$$m_k(t) = EZ^k(t)$$

of the FIND process. The previous section provides

$$\lim_{n\to\infty} E\frac{X^k_{n,l_n}}{n} = m_k(t),$$

for all $k \in I\!N$, and any sequence l_n satisfying $\frac{l_n}{n} \to_n t \in [0,1]$.

All moments exists since $\frac{X_{n,l}}{n}$ (cf. [5]), and $Z(t)$ (cf. [8]) have (uniformly in l,t) exponentially decreasing tails. Weak convergence and uniform integrability (consequences of exponential tails) imply the convergence of the moments.

We can identify m_k as a fixed point of some operator. Let B be the Banach space of all bounded measurable functions $f : [0,1] \to [0,1]$ endowed with the supremum norm. For given k and m_i, $i < k$ as above define the operator $K_k : B \to B$ via

$$K_k(f)(t) := \int_t^1 u^k f\Big(\frac{t}{u}\Big)\, du + \int_0^t (1-u)^k f\Big(\frac{t-u}{1-u}\Big)\, du + b_k(t),$$

where

$$b_k(t) = \sum_{i=1}^{k} (-1)^{i-1} \binom{k}{i} m_{k-i}(t).$$

The function m_k is the unique fixed point of K_k:

$$K_k(m_k) = m_k,$$

for $k \in I\!N$, see Paulsen [32].

From the fixed point property of m_k Paulsen [32] derived the differential equation

$$D^{k+2} m_k(t) = \left(\frac{1}{1-t} - \frac{1}{t}\right) D^{k+1} m_k(t) + D^{k+2} b_k(t),$$

where D denotes the differentiation with respect to t. Explicit solutions for m_1 and m_2 are given above. Explicit solutions for higher moments are not known and, although it seems manageable, they appear to be inattractive in view of the formula for the variance. The above two characterisations allow (good) numerical approximations.

Joint moments: We also obtain asymptotic results on joint distributions. Recall the definition of $X_{n,J}$ in Section 3. Let $j_{n,1} < \cdots < j_{n,p}$ be the elements of $J = J(n) \subset \{1,\ldots,n\}$. Assume $\lim_n \frac{j_{n,i}}{n} = t_i \in [0,1]$, for $1 \le i \le p$. Then we state as an exercise,

$$\frac{X_{n,J}}{n} \quad \xrightarrow{\mathcal{D}}_n \quad \sum_{v \in V} \mathbb{1}_{\exists i : t_i \in L(v)} L(v).$$

Further all moments converge as do all exponential moments.

Grand averages: For grand averages the asymptotic via FIND is not obvious, and although it has not been done, it will work in a straightforward way. However, grand averages satisfy an easier stochastic fixed point equation. The grand averages $\frac{X_n^{(p)}}{n} = Y_n^{(p)}$ converge (weakly and in Mallows l_2-metric) to some random variable $Y^{(p)}$. Lent and Mahmoud [24] follow up.

The limit satisfies the stochastic fixed point equation, cf. Mahmoud and Smythe [26],

$$\begin{aligned} Y^{(p)} \quad &\stackrel{\mathcal{D}}{=} \quad -2H_p + \sum_{i=0}^{p} \mathbb{1}_{U_{(i)}<U\le U_{(i+1)}}((Y^{(i)} + 2H_i + 1)U \\ &\qquad +(\overline{Y}^{(p-i)} + 2H_{p-r} + 1)(1-U)), \end{aligned}$$

where $U, U_1, \ldots, U_p, Y^{(i)}, \overline{Y}^{(i)}, 1 \le i \le p$, are independent random variables. The variables $U, U_1, \ldots, U_p$ are uniformly distributed on $[0,1]$, and $U_{(1)} \le \cdots \le U_{(p)}$ is the order statistic of $U_1, \ldots, U_p$. By definition $U_{(0)} := 0$ and $U_{(p+1)} := 1$. The random variables $Y^{(i)}$ are distributed as $\overline{Y}^{(i)}$, the distribution is recursively given.

The fixed point equation for the special case $p = 1$, shown in Mahmoud, Modarres, and Smythe [25], is

$$Y \stackrel{\mathcal{D}}{=} X(Y+3) - 2,$$

where X, Y are independent, and X has density $x \mapsto 2x\mathbb{1}_{[0,1]}(x)$. The characteristic function is

$$Ee^{itY} = e^{\int_0^t \frac{e^{iu}-1-iu}{u}du}.$$

Recalls: Recently Grübel and Stefanowski [11] studied recalls via the FIND process.

The supremum of FIND: Consider the supremum $M := \sup_t Z(t)$ of the FIND process. The fixed point equation (3), using the given representation via random variables, provides

$$M = 1 + (UM^{(0)}) \vee ((1-U)M^{(1)}),$$

with $M^{(i)} = \sup_t Z^{(i)}(t)$. Therefore M is a solution of the stochastic fixed point equation

$$X \stackrel{\mathcal{D}}{=} 1 + (UX) \vee ((1-U)\overline{X}),$$

where $U, X, \overline{X}$ are independent, U has a uniform distribution and $X, \overline{X}$ have the same distribution.

This fixed point equation has a unique solution [8] in the set of square integrable random variables. (It is to our knowledge the first stochastic fixed point equation considered involving the supremum (infimum). A systematic study has just started.)

Fixed point equations may help in deriving properties of random variables. Devroye [5] obtained lower and upper estimates

$$2.3862 < 1 + 2\ln 2 \le E(M-1) \le \frac{5}{\sqrt{2\pi}} + 12\frac{12e}{5} < 8.5185$$

for the expectation. For higher moments Devroye [5] obtains

$$E(M-1)^k \le 3^{k-1} k!\, E(M-1).$$

(A better constant is also given in [5].) M has a smooth density [5], which is supported on $[2, \infty)$, has a variation bounded by 2 and is Lipschitz with Lipschitz constant not exceeding 2. Thus, M has superexponential tails in the sense that, for all $\lambda > 0$, there exists a constant $C = C(\lambda)$ such that, for all $t \ge 0$,

$$P(M \ge t) \le Ce^{-\lambda t}.$$

We shall give a different proof highlightening the fixed point equation. The method is applicable to more general recurrences, e.g. the discrete random variables $\sup_l \frac{X_{n,l}}{n}$ have also superexponential tails.

Lemma 5.1. *Let $M = \sup_t Z(t)$, or $M = \sup_l \frac{X_{n,l}}{n}$. Then, for every $\lambda > 0$, there exists a constant $C = C(\lambda)$ (uniformly in n) such that*

$$P(M \ge t) \le Ce^{-\lambda t}.$$

$Ee^{\lambda M}$ is finite uniformly in n.

Proof heuristics: Assume, for any $a < 0$, there exists $b > 0$ such that for all, $x \ge 0$:

$$P(X \ge x) \le be^{ax}.$$

Further let $be^{ax_0} \ge 1$ and $\frac{2e^{-a}}{-a(x_0-1)} < 1$ for some $x_0 > 1$. Then the fixed point equation provides for $x \ge x_0$

$$\begin{aligned} P(X \ge x) &= P((UX) \vee ((1-U)\overline{X}) \ge x-1) \\ &\le P(UX \ge x-1) + P((1-U)\overline{X} \ge x-1) \end{aligned}$$

$$\begin{aligned} &= 2P(UX \geq x-1) \leq 2bEe^{a(x-1)\frac{1}{U}} \\ &\leq 2b\int_1^\infty \frac{1}{v^2} e^{a(x-1)v}\, dv \leq 2b\int_1^\infty e^{a(x-1)v} dv \\ &\leq 2b\frac{e^{a(x-1)}}{-a(x-1)} \leq be^{ax}\frac{2e^{-a}}{-a(x-1)} \leq be^{ax} \end{aligned}$$

The term I is strictly smaller than 1 for sufficiently large x.

This observation enables us to give a formal correct proof via the contraction method, [35], [38]. Let K be a map on probability measures defined by

$$K(\mu) = Ł(1 + (UX) \vee ((1-U)\overline{X})$$

where $U, X, \overline{X}$ are independent, U has a uniform distribution and $X, \overline{X}$ the same distribution μ. If μ satisfies the condition $\mu([x,\infty)) \leq be^{ax}$, for all $x \geq 0$ and some $a < 0 < b$, then $K(\mu)$ satisfies the condition with the same a and b.

Since K is a strict contraction, with respect to the Mallows l_2-metric, the sequence

$$\mu_0 := \delta_0, \quad \mu_1 := K(\mu_0), \quad \mu_2 := K(\mu_1), \ \ldots$$

converges in l_2-metric to a limit μ_∞, which is the distribution of M. If μ_0 satisfies the condition, then all $\mu_n, n \in I\!N$, satisfy the condition and also the l_2-limit μ_∞.

The same argument works for the discrete analog $\sup_l \frac{X_{n,l}}{n}$. We only have an additional index n, and have to use some uniformity in n for the induction step. (The choice of b may be different.) We shall not pursue the details.

6 Find revisited

From the above we have a good understanding of Find and all major questions seem to be reasonably answered. So, why return to Find? The (adhoc method of) Skorodhod convergence of the processes worked for the 2-version of Find, but it is not true for the 3-version. However, we still obtain all results for the 3-version as for the 2-version. There should be no major difference between the versions. It seems to be a purely technical difficulty.

The trouble is the Skorodhod convergence is too fine a convergence. We actually need to ask for process convergence in the sense

of convergence of the finite marginal distributions to some limit and a good representation of the limiting process. (This requires less than the convergence of special realisations of processes.) We will show in the sequel, that this program works for the 2-version and the 3-version simultaneously, since the limiting marginal distributions are the same.

Our point in doing this for FIND is to sharpen the tools and methods for more complex algorithms. FIND serves as a major toy example to try new ideas and techniques. We intend to develop a method, which works for both versions and also in a more general setting: the setting of stochastic fixed point equations for measures on function spaces.

Let X be a random variable with values in $\mathbb{R}^I, I \subset \mathbb{R}$. Let $\Phi_{J,K} : \mathbb{R}^J \to \mathbb{R}^K$, for $K \subset J \subset I$, be the projection from $\mathbb{R}^J$ to the subspace $\mathbb{R}^K$. We use Φ_J for $\Phi_{I,J}$. The distribution $P^{\Phi_J(X)}$ of the random variable $\Phi_J(X)$, $\Phi_J(X)(\omega) = \Phi_J(X(\omega))$, is called the marginal distribution of X for the set of coordinates J. The finite marginal distributions are those with a finite index set J.

Our program consists of 3 steps. Let $\varphi_J^{(n)}$ be the marginal distribution of Y_n to the coordinates J. The first step of our program is to show $\varphi_J^{(n)}$ converges weakly, as $n \to \infty$, to some limit called φ_J for every finite set $J \subset [0,1]$.

In the second step we show the consistency of the family $\varphi_J, J \subset [0,1]$ finite: For all finite $K \subset J \subset [0,1]$, with $|J| = |K| + 1 < \infty$, the following relation holds:

$$\varphi_J(\Phi_{J,K}^{-1}(\cdot)) = \varphi_K(\cdot).$$

The consistency is equivalent to the existence of a probability measure μ on $\mathbb{R}^{[0,1]}$ satisfying

$$\mu\Phi_J^{-1} = \varphi_J,$$

for all finite $J \subset [0,1]$. The function space $\mathbb{R}^{[0,1]}$ is endowed with the smallest σ-field $\mathcal{A}$ ensuring all one dimensional projections $\Phi_{\{i\}},\ i \in [0,1]$ are measurable.

In a third step we show a "nice" representation of a random variable X with the distribution μ as above. The meaning of "nice" is something like the paths $X(\omega) : I \to \mathbb{R}$ have desirable properties for almost all $\omega \in \Omega$. For the Brownian motion "nice" is continuity. In our setting "nice" requires the realisation of the limiting process with paths in D.

Step 1: This should purely be a consequence of the recursion on Y_n. The 1-dimensional distributions satisfy

$$\varphi_{\{t\}}^{(n)} \stackrel{l_2}{\to}_{n\to\infty} \varphi_{\{t\}},$$

the convergence, for all $t \in [0,1]$, is to some distribution [37]. The convergence is in Mallows l_2-metric. (Mallows l_2-convergence is equivalent to weak convergence and the convergence of the finite second moments. For general information on the Mallows metric see Bickel and Freedman [1].) Further the function

$$[0,1] \ni t \mapsto \varphi_{\{t\}}$$

is continuous in l_2-Mallows metric. The measures $\varphi_{\{t\}}$ were identified with the distribution of $Z(t)$. Recently Knof [20] showed steps 1 and 2 in full detail—the convergence of the finite dimensional marginals to some limit φ and the consistency.

Step 2: Once the convergence in step 1 is done, the consistency of the family $\varphi_J,\ J \subset [0,1]$ finite, is easy and an immediate consequence of the recursive structure.

Let μ be the corresponding probability measure. As a consequence there exists a process Z with distribution μ. Notice that the process Z has values in $\mathbb{R}^{[0,1]}$, which is endowed with the σ-field $\mathcal{A}$ generated by one-dimensional projections. This process satisfies the stochastic fixed point equation

$$(Z(t))_t \stackrel{\mathcal{D}}{=} \left(1 + 1_{U>t} U Z^{(0)}(\frac{t}{U}) + 1_{U\le t}(1-U) Z^{(1)}(\frac{t-U}{1-U})\right)_t, \quad (5)$$

where $t \in [0,1]$. The random variables $Z^{(0)}, Z^{(1)}, U$ are independent. The random variable U is uniformly distributed on $[0,1]$. The processes $Z^{(0)}, Z^{(1)}$ and Z all have the same distribution μ. The symbol $\stackrel{\mathcal{D}}{=}$ denotes the same distribution. We use the notation $X = (X(t))_t$ for a processes $X : \Omega \to \mathbb{R}^{[0,1]}$. Notice that all this involves the finite dimensional distributions only.

We should explain, why the equation (5) is a stochastic fixed point equation for distributions. Define the map K from probability measures on $(\mathbb{R}^{[0,1]}, \mathcal{A})$ to itself,

$$K(\nu) = Ł\left(\left(1 + 1_{U>t} U Z^{(0)}(\frac{t}{U}) + 1_{U\le t}(1-U) Z^{(1)}(\frac{t-U}{1-U})\right)_t\right).$$

The random variables $Z^{(0)}, Z^{(1)}, U$ are independent. The variable U is uniformly distributed on $[0,1]$. The processes $Z^{(0)}$ and $Z^{(1)}$ have the distribution ν. The symbol Ł denotes the distribution of a random variable. Equation (5) now reads $K(\mu) = \mu$, i.e. μ is a fixed point of K. Actually K is a strict contraction for a suitable metric, as shown in Knof [20]. This opens the field for the contraction method [38] for recursive equations of random variables now with values in a function space.

Step 3: Given a consistent family of distributions, when there exists a process with nice paths in D, the space of right continuous paths with left limits. The FIND process Z with path in D, seen as a process with values in $(\mathbb{R}^{[0,1]}, \mathcal{A})$, has distribution μ.

Lemma 6.1. *There exists a random variable with the distribution μ from step 2 and with values in D.*

Proof: According to Theorem 1 of Dubins, Hahn [6], it is equivalent that the family of marginals $(\varphi_J)_J$ on $\mathbb{R}^{[0,1]}$ is right continuous and the following condition is satisfied: For all $\epsilon > 0, \beta > 0$, there is a finite positive integer k such that, for all finite J, the condition $\varphi_J(\{f \in \mathbb{R}^{[0,1]} \mid N_{\epsilon,J}(f) > k\}) < \beta$ holds, where $N_{\epsilon,J}(f)$ is the supremum over all integers r such that there exist r nonoverlapping intervals $[s_i, t_i]$, with $s_i, t_i \in J$, and $|f(s_i) - f(t_i)| \geq \epsilon$.

The right continuity is easy to show, since even continuity is given, if the elements $j_{n,1} \leq j_{n,2} \leq \cdots \leq j_{n,p}$ of J_n converge to the elements $j_1 \leq j_2 \leq \cdots \leq j_p$ of J then φ_{J_n} converges weak to φ_J, see [20].

We next turn to showing the condition. Fix ϵ, J, β. Let n, k be natural numbers. Split the set S of n numbers by a uniform distributed random variable into two sets $S_<, S_\geq$. The left set gets the name 0, and the right the name 1. These are the sets of the first generation. Split every set of the first generation independently into two sets. Give them a name by adding 0 to the name of the mother for the left set and adding 1 for the right set. Let $L_n(v)$ be the set to the name $v \in V$. (Some of these sets might be empty.) The k-th generation consists of all sets $L_n(v), |v| = k$.

Let $L_{n,k} = \sup_{|v|=k} \frac{|L_n(v)|}{n}$ be the maximal size of a set in k-th generation divided by n. Let $X_{|L_n(v)|,J}(v)$ be the number of comparisons of FIND in order to find all l-th smallest elements, $l \in J$ of the v-interval. (This interval has $|L_n(v)|$ elements.) We continue the procedure up to the k-th generation. Then the distribution of X_n

satisfies

$$(X_{n,l})_l \stackrel{\mathcal{D}}{=} \left(\sum_{|v|=k} X_{|L_n(v)|,l(v)}(v) + C \right)_l ,$$

where $X_{.,.}(v), L_n(v), l(v), C, |v| = k$ are appropriate random variables. Consider now $[0,1] \ni t \mapsto Y_{n,t}$ as (linear or piecewise constant) continuation of $Y_{n,t} = \frac{X_{n,tn}}{n}$ for $t = \frac{1}{n}, \ldots, \frac{n}{n}$.

Let $M_n(v) = \sup_t Y_{|L_n(v)|,t}(v)$ be the maximum for the $X(v)$ process for the v-interval. Hence, $M_n(v) = \sup_l \frac{X_{|L_n(v)|,l}(v)}{|L_n(v)|}$. We use $M_n = \sup_l \frac{X_{n,l}}{n}$. We obtain an upper estimate for N via

$$N_{\epsilon,J}(Y_n) \le 2^k + \sum_{|v|=k} N_{\epsilon,J(v)}\Big(\frac{|L_n(v)|}{n} Y_{|L_n(v)}(v)\Big),$$

since the interval $[s_i, t_i]$ contains either a dividing point or is contained in an interval $L(v), |v| = k$.

Estimate for δ:
We have

$$\begin{aligned}
P(N_{\epsilon,J}(Y_n) \ge 2^k + 1) &\le P(L_{n,k} > n\delta) + P(L_{n,k} \le n\delta) P\Big(\exists |v| = k, \\
&\quad N_{\epsilon,J(v)}\Big(\frac{|L_n(v)|}{n} Y_{|L_n(v)|}(v)\Big) \ge 1 \mid L_{n,k} \le n\delta\Big) \\
&\le P(L_{n,k} > n\delta) + \sum_{|v|=k} P\Big(N_{\epsilon,J(v)}\Big(\frac{|L_n(v)|}{n} \\
&\qquad \times Y_{|L_n(v)|}(v)\Big) \ge 1 \mid L_{n,k} \le n\delta\Big) \\
&\le P(L_{n,k} > n\delta) \\
&\quad + \sum_{|v|=k} P\Big(M_n(v) \ge \frac{n}{|L_n(v)|} \mid L_{n,k} \le n\delta\Big) \\
&\le P(L_{n,k} > n\delta) \\
&\quad + \sum_{|v|=k} P\Big(M_n(v) \ge \frac{1}{\delta} \mid L_{n,k} \le n\delta\Big) \\
&\le P(L_{n,k} > n\delta) + \sum_{|v|=k} \sup_i P\Big(M_i \ge \frac{1}{\delta}\Big) \\
&\le P(L_{n,k} > n\delta) + 2^k \sup_i P\Big(M_i \ge \frac{1}{\delta}\Big) \\
&= I + II.
\end{aligned}$$

For fixed k, $\frac{L_{n,k}}{n}$ converges weakly to a random variable $L_{\infty,k}$. Lemma 1 of [8] states that there exists a random variable $K = K(\omega)$ such that, for all $k \geq K$,

$$M_k(\omega) \leq k \left(\frac{2}{3}\right)^{k/2}.$$

Choose k sufficiently large such that $P(\{\omega \in \Omega \mid k > K(\omega)\}) < \beta/2$. Choose $\delta = 2k \left(\frac{2}{3}\right)^{k/2}$. Then for $n = n(k)$ sufficiently large

$$I = P(L_{n,k} > \delta) \leq P(L_{\infty,k} > \delta/2) \leq \beta/2.$$

For an estimate of the second term, use the result of Section 5,

$$II = 2^k \sup_i P\Big(M_i \geq \frac{1}{\delta}\Big) \leq 2^k C(\lambda) e^{-\frac{\lambda}{\delta}},$$

for some (even for any) $\lambda > 0$. Fix $\lambda > 0$, and choose eventually k larger to ensure $II < \beta/2$. (Recall that δ depends on k.)

Altogether, we obtain for this fixed k and for n sufficiently large uniformly in J

$$P(N_{\epsilon,J}(Y_n) \geq 2^k + 1) < \beta.$$

Since $\varphi_J^{(n)}$ converges weakly to φ_J we obtain the condition for φ.

This proof works also for the 3-version.

Once we have finished all 3 steps in a general setting including FIND, we are able to show process convergence in more complex situations, like the profile of binary trees. And we are able to use the stochastic fixed point equation for further asymptotic results, now for measures on function spaces.

The FIND process itself satisfies the stochastic fixed point equation

$$(Z(t))_t \stackrel{\mathcal{D}}{=} \left(1 + 1_{U>t} U Z^{(0)}(\frac{t}{U}) + 1_{U \leq t}(1-U) Z^{(1)}(\frac{t-U}{1-U})\right)_t,$$

now for processes on D endowed, for example, with the Borel σ-field induced by the Skorodhod metric.

Again, this is a fixed point equation, this time on the space D. It would be nice to have some good metric on D available, such that K is a strict contraction and converges by Banach's fixed point theorem. Of course, D has to be a complete metric space.

References

[1] Bickel, P. and Freedman, D. A. (1981), Some asymptotic theory for the bootstrap. Annals of Statistics, **9**, 1196–1217.

[2] Billingsley, P. (1999), Convergence of Probability Measures. New York: Wiley Series in Probability and Statistics.

[3] Blum, M., Floyd, R. W., Pratt, V., Rivest, R. L., and Tarjan, R. E. (1973), Time bounds for selection. Journal of Computer and System Sciences, **7**, 448–461.

[4] Devroye, L. (1984), Exponential bounds for the running time of a selection algorithm. Journal of Computer and System Sciences, **29**, 1–7.

[5] Devroye, L. (2001), On the probabilistic worst-case time of "Find". Algorithmica, **31**, 291–303.

[6] Dubins, L. and Hahn, M. (1978), A characterization of the family of finite-dimensional distributions associated with countably additive stochastic processes whose sample paths are in D. Z. Wahrscheinlichkeitstheorie verw. Gebiete, **43**, 97–1000.

[7] Floyd, R. W. and Rivest, R. L. (1975), Expected time bounds for selection. Commun. ACM, **18**, 165–172.

[8] Grübel, R. and Rösler, U. (1996), Asymptotic distribution theory for Hoare's selection algorithm. Adv. Appl. Prob., **28**, 252–269.

[9] Grübel, R. (1998), Hoare's selection algorithm: A Markov chain approach. J. Appl. Prob., **35**, 36–45.

[10] Grübel, R. (1999), On the median-of-K version of Hoare's selection algorithm. Theor. Inform. Appl., **33**, 177–192.

[11] Grübel, R. and Stefanoski, N. (2005), Mixed Poisson approximation of node depth distributions in random binary search tree. Annl. of Appl. Prob. (to appear).

[12] Hoare, C. A. R. (1961), Algorithm 63: PARTITION. Communications of the ACM, **4**, 321.

[13] Hoare, C. A. R. (1961), Algorithm 64: Quicksort. Communications in ACM, **4**, 321.

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

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

[16] Hwang, H-K. and Tsai, T-H. (2002), QUICKSELECT and the Dickman function. Combinatorics, Probability and Computing, **11**, 353–371.

[17] Kirschenhofer, P. and Prodinger, H. (1998), Comparisons in Hoare's Find Algorithms. Combinatorics, Probability and Computing, **7**, 111–120.

[18] Kirschenhofer, P., Prodinger, H., and Martínez, C. (1997), Analysis of Hoare's Find Algorithm with median-of-three partition. Random Structures and Algorithms, **10**, 143–156.

[19] Knessl, C. and Szpankowski, W. (1999), Quicksort algorithm again revisited. Discrete Math. Theor. Comput. Sci., **3**, 43–64, electronic only.

[20] Knof, D. (2004), Konvergenz von Find zu einem Maß auf $D[0,1]$. Diplomarbeit, University of Kiel, 2003, www.numerik.uni-kiel/reports/ Bericht 04–1.

[21] Knuth, D. (1973), The Art of Computer Programming. Vol 3: Sorting and Searching, 2nd ed. Reading, Massachessetts: Addison-Wesley.

[22] Knuth, D. E. (1972), Mathematical analysis of algorithms. Inform. Processing 71, Proc. IFIP Congr. 71, Ljubljana, Yugoslavia 1971, 19–27.

[23] Kodaj, B. and Mori, T. F. (1997), On the number of comparisons in Hoare's algorithm "Find". Studia Scientiarum mathematicarum Hungarica, **33**, 185–207.

[24] Lent, J. and Mahmoud, H. (1996), Average–case analysis of multiple Quickselect: An algorithm for finding order statistics. Statistics and Probability Letters, **28**, 299–310.

[25] Mahmoud, H., Modarres, R., and Smythe, R. (1995), Analysis of QUICKSELECT: An algorithm for order statistics. Theoretical Informatics and Its Applications, **29**, 255–276.

[26] Mahmoud, H. and Smythe, T. (1998), Probabilistic analysis of Multiple Quick Select. Algorithmica, **22**, 569–584.

[27] Mahmoud, H. M. (2000), Sorting. A distribution theory. Wiley-Interscience Series in Discrete Mathematics and Optimization. Chichester: Wiley xvi.

[28] Martínez, C. and Roura, S. (2001), Optimal sampling strategies in Quicksort and Quickselect. SIAM J. Comput., **31**, No.3, 683–705.

[29] Martínez, C., Panario, D., and Viola, A. (2004), Adaptive sampling for Quickselect. Proc. 15th ACM-SIAM Symp. on Discrete Algorithms (SODA), 447-455.

[30] Panholzer, A. and Prodinger, H. (1998), A generating functions approach for the analysis of grand averages for multiple Quickselect. Random Struct. Algorithms, **13**, No.3–4, 189–209.

[31] Panholzer, A. (2003), Analysis of multiple quickselect variants. Theoret. Comput. Sci., **302**, 45–91.

[32] Paulsen, V. (1997), The moment of Find. J. Appl. Prob., **34**, 1079–1082.

[33] Prodinger, H. (1995), Multiple Quickselect-Hoare's Find algorithm for several elements. Information Processing Letters, **56**, 123–129.

[34] Rösler, U. (1991), A limit theorem for 'Quicksort'. Informatique théorique et Applications/Theoretical Informatics and Applications, **25**, 85–100.

[35] Rösler, U. (1992), A fixed point theorem for distributions. Stochastic Processes and their Applications, **42**, 195–214.

[36] Rösler, U. (2001), On the analysis of stochastic divide and conquer algorithms. Algorithmica, **29**, 238–261, and Berichtsreihe des Mathematischen Seminars Kiel, Christian-Albrechts-Universitaet zu Kiel, www.numerik.uni-kiel/reports/ Bericht 99–1, 1999.

[37] Rösler, U. (2003), The backward view on the algorithm FIND. www.numerik.uni-kiel/reports/ Bericht 03–6.

[38] Rösler, U. and Rüschendorf, L. (2001), The contraction method for recursive algorithms. Algorithmica, **29**, 3–33.

[39] Schönhage, A., Paterson, M., and Pippenger, N. (1976), "Finding the median". J. Comput. System Sci, **13**, 184–199.

[40] Sedgewick, R. (1977), The analysis of Quicksort programs. Acta Inf., **7**, 327–355.

[41] Shen, H. and Chen, G. (1988), An optimal parallel selection algorithm. (Chinese. English summary) J. China Univ. Sci. Technol., **18**, 232–240.

[42] Yap, C. K. (1976), New upper bounds for selection. Comm. ACM, **19**, 501–508.