RAIRO. THEORETICAL INFORMATICS AND APPLICATIONS

RUDOLF GRÜBEL

On the median-of-k version of Hoare's selection algorithm

RAIRO. Theoretical Informatics and Applications, tome 33, n° 2 (1999), p. 177-192

<http://www.numdam.org/item?id=ITA_1999__33_2_177_0>

Theoretical Informatics and Applications
Theoret. Informatics Appl. **33** (1999) 177–192

ON THE MEDIAN-OF-K VERSION OF HOARE'S SELECTION ALGORITHM

RUDOLF GRÜBEL[1]

Abstract. In Hoare's (1961) original version of the algorithm FIND the partitioning element in the central divide-and-conquer step is chosen uniformly at random from the set S in question. Here we consider a variant where this element is the median of a sample of size $2k+1$ from S. We investigate convergence in distribution of the number of comparisons required and obtain a simple explicit result for the limiting average performance of the median-of-three version.

AMS Subject Classification. 68Q25, 68P10.

1. INTRODUCTION

Given a set $S \subset \mathbb{R}$ with n elements the central step in Hoare's (1961) selection algorithm FIND is a reduction to a similar selection problem with a strictly smaller set $S' \subset S$. This reduction is based on the selection of a *partitioning element* x from S (according to some rule to be specified); let m be the size of $S_- := \{y \in S : y < x\}$. If the l^{th} smallest element of S is to be determined then the algorithm proceeds with (S, l) replaced by (S', l') where $S' := S_-$ and $l' := l$ if $m \geq l$ and $S' := S \setminus (S_- \cup \{x\})$ and $l' := l - 1 - m$ if $m < l - 1$; obviously the required value is x if $m = l - 1$. In Hoare's [9] original rule PARTITION x is chosen uniformly at random from S. Following a similar proposal for the closely related sorting algorithm QUICKSORT (Hoare [9]) we consider in this paper a variant of FIND where $2k+1$ elements are selected uniformly at random from S and the partitioning element x is taken to be the median of these.

The performance of such algorithms is essentially determined by the number of comparisons that have to be carried out, which is a random quantity in the above setup. Much of the early literature concentrates on the expected value of

Keywords and phrases: Randomized algorithms, average performance, asymptotic distribution.

[1] Institut für Mathematische Stochastik, Universität Hannover, Postfach 60 09, 30060 Hannover, Germany; e-mail: rgrubel@stochastik.uni-hannover.de

these random variables, *i.e.* on the average performance of the algorithm. While FIND and QUICKSORT are close to each other from an algorithmic point of view, their probabilistic behaviour is totally different. In particular, for QUICKSORT the average performance is easily obtained, and the resulting computation is perhaps the one part of computer science most likely to appear in a lecture course on probability. For FIND, the same problem is considerably more difficult; see Problem 32, rated M40, in Knuth ([11], p. 136) or the amusing discussion in Rawlins ([15], pp. 194-201). For both algorithms it turns out that there is a wide gap between average and worst case performance, which motivates a closer analysis. Tail bounds for the distribution of the number of comparisons can be used to judge the practical relevance of the worst case; such bounds for FIND were first obtained by Devroye [4]. Grübel and Rösler [7] proved convergence in distribution of the (standardized) number of comparisons if the rank $l = l_n$ of the required element varies with the size n of the basic set in such a way that $n^{-1}l_n$ tends to a limit; this was based on the idea of considering the selection problem as indexed by l and then investigating the resulting stochastic processes. Grübel [8] simplified the proof of convergence of the one-dimensional distributions (which is the main case of interest). Kodaj and Mori [12] obtained a similar result with different techniques, they also obtained results on the rate of convergence. Grübel [8] also contains a (non-asymptotic) bound on the tails of the distribution which improves upon Devroye's earlier result.

The results mentioned in the previous paragraph all refer to the classical variant based on PARTITION. It is well known that a more sophisticated choice of the partitioning element can improve the average or worst case performance of QUICKSORT and FIND; see *e.g.* Sedgewick and Flajolet ([16], p. 21). Indeed, in a seminal paper Blum *et al.* [3] showed that the number of comparisons required if their rule PICK is employed is at most $5.4305 \cdot n$; other authors have subsequently improved the constant in this linear bound. The existence of a deterministic worst case upper bound that grows only linearly in the size n of the input set is a considerable theoretical achievement, but PICK seems to be too complicated to enter the world of computational recipes. Floyd and Rivest (1975) constructed an algorithm with better average performance than the original FIND. In this paper we consider the following selection rule, which is very simple to implement: $2k + 1$ elements, the *presample*, are selected uniformly at random from the respective set, the partitioning element is then chosen to be the median of these values. Historically this was the first attempt to improve upon PARTITION. A detailed analysis of the number of comparisons for $k = 1$ is given in Anderson and Brown [1], who mention a renewed interest in median-of-three selection.

Our main results are limit theorems where the size n of the basic set S tends to infinity. In the following section we first consider the case of fixed k and then let $k = k_n$ tend to infinity with n. Proofs are collected in Section 3. In Section 4 we discuss the results and give some numerical examples. We write $\mathcal{L}(X)$ for the distribution of the random variable X and 1_A for the indicator function of the set A.

2. RESULTS

The selection strategies considered in this paper are distributionally invariant under permutations of the object under consideration, so we may regard FIND as operating on sets. The distribution of the number of comparisons depends on the underlying set S only through its size $\#S$. Let $C_{n,l}^{(k)}$ be the (random) number of comparisons required by FIND$(\{1,\dots,n\},l)$ if, in each recursion step, the partitioning element is the median of $2k+1$ elements chosen uniformly at random from the set in question. The rule PARTITION used in the classical version of FIND can be regarded as the special case $k=0$ of this procedure. Note that for $k>0$ we will have an overhead due to the determination of the presample medians. Our main aim is to show convergence in distribution of $n^{-1}C_{n,l_n}^{(k)}$ if l_n varies with n such that $n^{-1}l_n$ tends to a limit, say t, and to obtain information about the limit distribution. This distribution depends on t which means that a whole family $\{Q_t : 0 \le t \le 1\}$ of probability measures on the (Borel subsets of the) non-negative real halfline $\mathbb{R}_+$ arises. It turns out that the selection rule determines a probability distribution μ on the unit interval $[0,1]$, and that $\{Q_t : 0 \le t \le 1\}$ depends on this μ in a particular manner.

Theorem 1. *To any probability distribution μ on $[0,1]$ which is not concentrated on $\{0,1\}$ there corresponds a unique family $\{Q_t : 0 \le t \le 1\}$ of probability distributions on $\mathbb{R}_+$ with the following properties:* (a) $\sup_{0\le t\le 1}\int x^2\,Q_t(dx) < \infty$; (b) *if $\{X_t : 0 \le t \le 1\}$ is a family of random variables with $\mathcal{L}(X_t) = Q_t$ for all $t \in [0,1]$ and if ξ is another random variable, independent of the X-family and with distribution μ, then, for all $t \in [0,1]$,*

$$\mathcal{L}\left(1 + 1_{(t,1]}(\xi)\cdot\xi\cdot X\left(\frac{t}{\xi}\right) + 1_{[0,t]}(\xi)\cdot(1-\xi)\cdot X\left(\frac{t-\xi}{1-\xi}\right)\right) = Q_t.$$

As in the theorem we write $X(t)$ instead of X_t whenever typographically more convenient. We call $\{Q_t : 0 \le t \le 1\}$ the *μ-split*. The μ-splits needed in connection with FIND are built from a particular family: the beta distribution of the first kind with parameters $\alpha>0$, $\beta>0$, Beta(α,β) for short, is given by its density function

$$f(x\,|\,\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,x^{\alpha-1}(1-x)^{\beta-1}, \quad 0<x<1.$$

We consider convergence with respect to one of the Wasserstein (or Mallows) metrics: for any two probability measures μ, ν on $\mathbb{R}_+$ with finite second moments we define $d_2(\mu,\nu)$ by

$$d_2(\mu,\nu)^2 := \inf\left\{E(X-Y)^2 : \mathcal{L}(X)=\mu,\ \mathcal{L}(Y)=\nu\right\}.$$

Some properties of d_2 are given in Bickel and Freedman [2]; in particular, d_2-convergence is equivalent to weak convergence and convergence of the second moments. Recall that we partition at the median of a presample of size $2k+1$.

Theorem 2. *Let* $\{Q_t^{(k)} : 0 \le t \le 1\}$ *be the* Beta$(k+1, k+1)$*-split. Then, for all* $0 \le t \le 1$, $\mathcal{L}(n^{-1}C_{n,l_n}^{(k)})$ *converges to* $Q_t^{(k)}$ *with respect to* d_2 *if* l_n/n *converges to* t *as* $n \to \infty$.

Of particular interest is the average performance of the presample version of FIND, *i.e.* the expectation $EC_{n,l}^{(k)}$ of $C_{n,l}^{(k)}$. Theorem 2 implies that

$$m_k(t) := \lim_{n\to\infty} \frac{1}{n} EC_{n,l_n}^{(k)} = \int x\, Q_t^{(k)}(dx)$$

if $l_n/n \to t$. For the median-of-three version of FIND we have the following explicit result.

Theorem 3. $m_1(t) = 2 + 3t(1-t)$ *for all* $t \in [0,1]$.

It is known that $m_0(t) = 2 - 2t\log t - 2(1-t)\log(1-t)$ for the standard version of FIND (see *e.g.* Th. 11 in Grübel and Rösler [7], or the discussion following Lem. 1.1 in Kodaj and Mori [12]). Thus, for finding the median, the asymptotic average number of comparisons required drops from about $3.386 \cdot n$ to $2.75 \cdot n$. The following theorem contains some simple bounds on m_k for general k.

Theorem 4. *For all* $k \in \mathbb{N}_0$, $m_k(0) = m_k(1) = 2$ *and* $m_k(t) > 2$ *for* $0 < t < 1$. *Moreover, for all* $k \in \mathbb{N}_0$ *and* $t \in [0,1]$,

$$m_k(t) \le \frac{2}{1 - \binom{2k+1}{k+1} 2^{-2k-1}}.$$

A standard application of Stirling's formula shows that the upper bound is of the form $2 + c_k/\sqrt{k}$ with $\lim_{k\to\infty} c_k = 2/\sqrt{\pi}$; in particular, m_k tends to the constant function 2 if $k \to \infty$, uniformly on the unit interval. Note that this refers to the limiting behaviour of the limits, *i.e.* it does not apply to the asymptotic behaviour of the number of comparisons in the case where the presample size k varies with n.

Numerical experiments suggest that $m_{k+1}(t) < m_k(t)$ for all $k \in \mathbb{N}_0$, $0 < t < 1$ (see also Sect. 4), but I do not have a proof. The bounds in Theorem 4, together with the subsequent remark, yield the weaker result that for all $k \in \mathbb{N}_0$, $0 < t < 1$ there exists a $j \in \mathbb{N}$ such that $m_{k+j}(t) < m_k(t)$. For $k = 0$ we can of course use the above explicit formulae for m_0 and m_1, and it is indeed straightforward to check that $m_1(t) < m_0(t)$ for $0 < t < 1$.

As far as the total number of comparisons is concerned the worst case arises if in each recursion step with set S and partitioning element x from S the desired element is in the larger one of the sets S_-, $S_+ := S \setminus (S_- \cup \{x\})$. If $\#S = n$ we would then have the lower bound

$$\max\{\#S_-, \#S_+\} \ge \min_{k=0,\dots,n-1} \max\{k, n-k-1\} = \left\lfloor \frac{n}{2} \right\rfloor$$

on the number of comparisons needed by the next step. This implies

$$\liminf_{n\to\infty} \frac{1}{n} W_n \geq 2 \quad \text{for } W_n := \sup_{1\leq l\leq n} C_{n,l},$$

whatever rule we use to select the partitioning element. This gives a very pessimistic lower bound on the performance: if interpreted as a game between the algorithm designer and an adversary this would model a situation where the designer has to disclose his partitioning strategy and the adversary then determines which element of S the program has to find. Clearly, in this situation a selection rule that manages to split into sets of roughly equal size would be preferable, and this leads quite naturally to the investigation of versions of FIND with increasing presample size. Such a version is specified by a sequence $(k_n)_{n\in\mathbb{N}}$ of integers: if, in some recursion step, a set S of size n is under consideration, then the partitioning element is chosen to be the median of a sample of size $2k_n + 1$ from S. Again, we would have an overhead due to the additional number of comparisons needed within the presamples (on which we could use FIND recursively), and in the case of $k_n \to \infty$ this part of the total workload may well become dominant. For this reason we require an upper bound on the rate of increase of the presample size in the following result.

Theorem 5. *If $k_n \to \infty$ and $n^{-1}(\log n)\, k_n \to 0$ then $n^{-1}W_n \to 2$ in probability as $n \to \infty$.*

Hyafil [10] gives a more detailed analysis of lower bounds for C_{nl}. The versions of FIND considered in this paper essentially discard the information contained in l. Note that the value of l changes with positive probability in each recursion step unless $l = 1$ or $l = n$; these extreme cases correspond to finding the minimum or maximum element of S which can easily be done with one pass through the data. Floyd and Rivest [6] invented a selection rule that makes use of the rank of the element to be found in each recursion step and consequently achieves a better average performance.

3. PROOFS

We recall the basic connection to Markov chains, discuss the number of recursions and then prove the theorems.

The pairs $(S, l), (S', l'), \ldots$ arising in the recursion steps can be regarded as the successive states of a Markov chain, we may even reduce the sets to their respective cardinality. This was the basic idea in Grübel [8] where the chains were rescaled to obtain a weak convergence result in the case $k = 0$. We can also use this in the present case; in particular, it enables us to analyse the overhead introduced by the presamples.

Let $I = \{(n, l) : n \in \mathbb{N}, l = 1, \ldots, n\}$. Starting with a sequence $(q_n)_{n\in\mathbb{N}}$ of probability distributions $q_n = (q_{n,1}, \ldots, q_{n,n})$ on $\{1, \ldots, n\}$ we define a transition mechanism on I as follows: from $(n, l) \in I$ move to $(1, 1)$ with probability $q_{n,l}$,

move to $(n-j, l-j)$ with probability $q_{n,j}$, $j = 1, \dots, l-1$, and move to $(j-1, l)$ with probability $q_{n,j}$, $j = l+1, \dots, n$. It is easy to check that this models FIND correctly; in particular, if $(Z_m)_{m \in \mathbb{N}_0}$, $Z_m = (X_m, Y_m)$, is a Markov chain with this transition mechanism and start at (n, l), then the total number of comparisons required by the partitioning steps is given by $\sum_{m=0}^{\infty}(X_m - 1)$. If presamples of size $2k+1$ are used then q_n is the distribution of the median of $2k+1$ independent random variables with the uniform distribution on $\{1, \dots, n\}$. Also, the number $R_{n,l}$ of recursion steps required by FIND$(\{1, \dots, n\}, l)$ is equal to the entry time

$$T := \inf\{m \in \mathbb{N}_0 : Z_m = (1,1)\}$$

into the state $(1,1)$ of the chain. We have the following inequality which provides us with the necessary tool to bound the overhead introduced by the presamples. We write $E_{(n,l)}$ for expectation with respect to start at $(n,l) \in I$; for the construction of Markov chains with given initial state and transition mechanism see *e.g.* Section 5.2 in Neveu [13]. Note that the upper bound does not depend on k.

Lemma 6.

$$E_{(n,l)}\left(\frac{4}{3}\right)^T \le n \qquad \textit{for all } (n,l) \in I.$$

Proof. We use induction over $n \in \mathbb{N}$; for $n = 1$ the assertion is trivial. Suppose that the inequality holds for all $(n', l') \in I$ with $n' < n$ for some $n \ge 2$. A decomposition with respect to the first step gives

$$E_{(n,l)}\left(\frac{4}{3}\right)^T = \sum_{j=1}^{l-1} q_{n,j}\, E_{(n-j,l-j)}\left(\frac{4}{3}\right)^{T+1} + q_{n,l}\,\frac{4}{3} + \sum_{j=l+1}^{n} q_{n,j}\, E_{(j-1,l)}\left(\frac{4}{3}\right)^{T+1} \le \frac{4}{3}\sum_{j=1}^{n} \max\{n-j, j-1\}\, q_{n,j}.$$

When applied to a random variable U which is uniformly distributed on the unit interval, the function $x \to \lceil nx \rceil$ yields a random variable $\tilde{U}$ which is uniformly distributed on the set $\{1, \dots, n\}$. As this function is monotone and as we consider presamples of odd size only, we can generate the median $\tilde{V}$ of a sample of size $2k+1$ from the uniform distribution on the set $\{1, \dots, n\}$ via $\tilde{V} := \lceil nV \rceil$, with V the median of a sample of size $2k+1$ from the uniform distribution on the unit interval. It is well known that such a V has distribution Beta$(k+1, k+1)$; let

$$f_k(x) = f(x|k+1, k+1) = \frac{(2k+1)!}{k!\,k!}\, x^k (1-x)^k, \quad 0 \le x \le 1,$$

be the associated density. By construction, $\tilde{V}$ has distribution q_n, and

$$\sum_{j=1}^{n} \max\{j-1, n-j\}\, q_{n,j} \le n \cdot E \max\{V, 1-V\}.$$

Standard calculations give

$$\begin{aligned} E\max\{V,1-V\} &= \int_{1/2}^{1} x f_k(x)\,dx + \int_0^{1/2} (1-x) f_k(x)\,dx \\ &= 2\int_{1/2}^{1} x f_k(x)\,dx \\ &= \frac{1}{2} + \nu_k \quad \text{with} \quad \nu_k := \binom{2k+1}{k+1} \frac{1}{2^{2k+2}}\,. \end{aligned}$$

We have $\nu_0 = 1/4$ and it is easy to check that $\nu_{k+1}/\nu_k \le 1$ for all $k \in \mathbb{N}_0$, hence $E\max\{V,1-V\} \le 3/4$. Putting pieces together we see that the formula holds for n too, which completes the induction step. □

The proof of Theorem 1 is built on a contraction argument and similar to the heuristic sketch given at the end of Section 4 in Grübel and Rösler [7] for a related situation; it is instructive to compare the details. In particular we neither need nor discuss the existence of a probability measure on some path space with marginals Q_t, $0 \le t \le 1$.

Proof of Theorem 1. Let M_2 be the set of all probability measures μ on $\mathbb{R}_+$ with the property $\int x^2 \mu(dx) < \infty$. Endowed with d_2, M_2 is a complete metric space. Let $\mathbb{M}$ denote the set of all functions $Q : [0,1] \to \mathrm{M}_2$, $t \to Q_t$, that satisfy $\sup_{0\le t\le 1} \int x^2\, Q_t(dx) < \infty$. Standard arguments from functional analysis such as given by Dunford and Schwarz ([5], p. 258) can be used to show that

$$d(Q,Q') := \sup_{0\le t\le 1} d_2(Q_t, Q'_t)$$

is a metric on $\mathbb{M}$, and that $(\mathbb{M}, d)$ is a complete metric space.

Now let μ be as in the theorem. We define an operator $T : \mathbb{M} \to \mathbb{M}$ as follows: $(T(Q))_t$ is the distribution of

$$Y_t := 1 + 1_{(t,1]}(\xi) \cdot \xi \cdot X\left(\frac{t}{\xi}\right) + 1_{[0,t]}(\xi) \cdot (1-\xi) \cdot X\left(\frac{t-\xi}{1-\xi}\right)$$

where $X = (X_t)_{0\le t\le 1}$ is such that $\mathcal{L}(X_t) = Q_t$, $0 \le t \le 1$, and ξ is independent of $(X_t)_{0\le t\le 1}$ with $\mathcal{L}(\xi) = \mu$ (note that there are no assumptions on the joint distribution of the variables X_t, $0 \le t \le 1$). It is easy to check that $\sup_{0\le t\le 1} EY_t^2 < \infty$. Given $Q, Q' \in \mathbb{M}$ the quantile transformation can be used to construct families X, X' such that

$$\mathcal{L}(X_t) = Q_t, \quad \mathcal{L}(X'_t) = Q'_t \quad \text{and} \quad E(X_t - X'_t)^2 = d_2(Q_t, Q'_t)^2$$

for all $t \in [0,1]$. We can further construct another random variable ξ, independent of both these families and with distribution μ. Then the families Y, Y' defined by

$$\begin{aligned} Y_t &:= 1 + 1_{(t,1]}(\xi) \cdot \xi \cdot X\left(\frac{t}{\xi}\right) + 1_{[0,t]}(\xi) \cdot (1-\xi) \cdot X\left(\frac{t-\xi}{1-\xi}\right), \\ Y'_t &:= 1 + 1_{(t,1]}(\xi) \cdot \xi \cdot X'\left(\frac{t}{\xi}\right) + 1_{[0,t]}(\xi) \cdot (1-\xi) \cdot X'\left(\frac{t-\xi}{1-\xi}\right), \end{aligned}$$

have the one-dimensional marginals $T(Q), T(Q')$ respectively; note that we use the same ξ in both lines. With this construction and $1_{[0,t]}1_{(t,1]} \equiv 0$ we obtain

$$\begin{aligned} \left(d_2\left(T(Q)_t, T(Q')_t\right)\right)^2 &\le E(Y_t - Y'_t)^2 = E\left(1_{(t,1]}(\xi) \cdot \xi \cdot \left(X\left(\frac{t}{\xi}\right) - X'\left(\frac{t}{\xi}\right)\right)\right)^2 \\ &\quad + E\left(1_{[0,t]}(\xi) \cdot (1-\xi) \cdot \left(X\left(\frac{t-\xi}{1-\xi}\right) - X'\left(\frac{t-\xi}{1-\xi}\right)\right)\right)^2 \\ &\le \left(E\xi^2 + E(1-\xi)^2\right) \sup_{0 \le s \le 1} E\left(X_s - X'_s\right)^2. \end{aligned}$$

By assumption, μ is not concentrated on the end points of the interval, which implies that $c(\mu) := E\xi^2 + E(1-\xi)^2$ is strictly smaller than 1. Taking the supremum over $t \in [0,1]$ in the above bound we obtain

$$d\left(T(Q), T(Q')\right) \le c(\mu)^{1/2}\, d(Q, Q') \quad \text{for all } Q, Q' \in \mathbb{M},$$

and all assertions of the theorem now follow on using Banach's fixed point theorem. □

Proof of Theorem 2. As k is held fixed here we may neglect the additional comparisons required by the determination of the medians of the presamples if $E(R^{(k)}_{n,l_n})^2 = o(n^2)$ as $n \to \infty$. This, however, is immediate from the above lemma. Constructing the partitioning elements via beta-distributed random variables as in the proof of the lemma we can now proceed exactly as in Grübel [8], if we replace the uniform variables U by beta variables V throughout: Let $I := \{(x,y) \in \mathbb{R}^2_+ : 0 \le y \le x\}$ and let $P((x,y), \cdot)$ be the distribution of ξ,

$$\xi := \begin{cases} (x - Vx, y - Vx), & \text{if } Vx \le y, \\ (Vx, y), & \text{if } Vx > y, \end{cases} \quad \text{with } V \sim \text{Beta}(k+1, k+1).$$

Consider the Markov chains $(Z^{(x,y)}_m)_{m \in \mathbb{N}_0}$ with state space I, transition kernel P and start at (x,y), $(x,y) \in I$. Then we obtain as in Grübel [8] that $n^{-1}C^{(k)}_{n,l_n}$ converges in distribution to $\sum_{m=0}^{\infty} X^{(1,t)}_m$, where $X^{(1,t)}_m$ denotes the first component of $Z^{(1,t)}_m$, $m \in \mathbb{N}_0$. These chains are stochastically self-similar in the sense that $(Z^{(x,tx)}_m)_{m \in \mathbb{N}_0}$ is identical in distribution to $(x \cdot Z^{(1,t)}_m)_{m \in \mathbb{N}_0}$. A decomposition with respect to the value of the first V-variable therefore shows that the limit

distribution, as a function of t, satisfies condition (b) of Theorem 1 with $\mu = \mathcal{L}(V)$. A simple upper bound similar to the one employed in the standard case $k = 0$ yields the other condition, so that Theorem 1 can be applied. □

Proof of Theorem 3. Let $m(t) := \int x\, Q_t(dx)$. As $(Q_t)_{0\le t\le 1}$ arises as the fixed point of a strict contraction on $\mathbb{M}$, we obtain the μ-split in the limit if we apply the contraction repeatedly to any family from $\mathbb{M}$ (see the proof of Th. 1). In particular, we may start with $Q_t \equiv \delta_0$. The moment functions $t \to \int x\, Q_{n,t}(dx)$ of the successive families $(Q_{n,t})_{0\le t\le 1}$, $n \in \mathbb{N}$, will then all be measurable functions on the unit interval. As m is the pointwise limit of these it is measurable too, and we obtain from property (b) of the μ-split that m satisfies the following integral equation:

$$m(t) \;=\; 1 \;+\; \int_{(t,1]} x\, m\left(\frac{t}{x}\right)\, \mu(dx) \;+\; \int_{[0,t]} (1-x)\, m\left(\frac{t-x}{1-x}\right)\, \mu(dx).$$

Of course, property (a) from Theorem 1 implies that m is bounded. We may therefore regard m as the fixed point of an operator T, where $T : \mathbb{B} \to \mathbb{B}$ with $\mathbb{B}$ the space of bounded, measurable function $f : [0,1] \to \mathbb{R}$ is given by

$$T(f)(t) \;:=\; 1 \;+\; \int_{(t,1]} x\, f\left(\frac{t}{x}\right)\, \mu(dx) \;+\; \int_{[0,t]} (1-x)\, f\left(\frac{t-x}{1-x}\right)\, \mu(dx).$$

A straightforward generalization of the proof of Theorem 11 in Grübel and Rösler [7] shows that with $\|f\| := \sup_{0\le t\le 1} |f(t)|$ we have

$$\|T(f) - T(g)\| \;\le\; c(\mu)\, \|f - g\|$$

where

$$c(\mu) \;:=\; \sup_{0\le t\le 1} \left(\int_{(t,1]} x\, \mu(dx) \;+\; \int_{[0,t]} (1-x)\, \mu(dx) \right).$$

As $c(\mu) \le \int_{[0,1]} x \vee (1-x)\, \mu(dx)$ we obtain $c(\mu) < 1$ since the μ's considered here are not concentrated on $\{0,1\}$ (in fact, we know from the proof of the lemma that $c(\mu) \le 3/4$). Hence T is a strict contraction on the Banach space $(\mathbb{B}, \|\cdot\|)$, which implies that the solution to the above integral equation is unique. In particular, to prove Theorem 3, it remains to check that the integral equation

$$m(t) \;=\; 1 \;+\; 6 \int_t^1 x^2(1-x)\, m\left(\frac{t}{x}\right)\, dx \;+\; 6 \int_0^t x(1-x)^2\, m\left(\frac{t-x}{1-x}\right)\, dx$$

is satisfied with $m(t) := 2 + 3t(1-t)$. This is straightforward. □

This proof does not answer the question of how one arrives at the solution function. One possibility is to extrapolate from the result obtained by Anderson and Brown [1] for finite n and to guess the solution by letting n tend to infinity in their approximation formula (19). Here is another possibility, which could also be used, at least in principle, for $k > 1$: we look for the solution m_1 of

$$m_1(t) \;=\; 1 \;+\; 6\int_t^1 x^2(1-x)\,m_1\left(\frac{t}{x}\right)dx \;+\; 6\int_0^t x(1-x)^2\,m_1\left(\frac{t-x}{1-x}\right)dx,$$

$0 \le t \le 1$. A change of variables $\frac{t}{x} \hookrightarrow x$ in the first, $\frac{t-x}{1-x} \hookrightarrow 1-x$ in the second integral simplifies the dependence on t of the right hand side:

$$\begin{aligned} m_1(t) = 1 \;&+\; 6t^3\int_t^1 \frac{1}{x^4}\,m_1(x)\,dx \;-\; 6t^4\int_t^1 \frac{1}{x^5}\,m_1(x)\,dx \\ &+\; 6(1-t)^3\int_{1-t}^1 \frac{1}{x^4}\,m_1(1-x)\,dx \;-\; 6(1-t)^4\int_{1-t}^1 \frac{1}{x^5}\,m_1(1-x)\,dx. \end{aligned}$$

Differentiating both sides five times we arrive at the following differential equation for $\phi := m_1^{(3)}$:

$$\phi''(t) \;=\; 6\left(\frac{1}{t^2} + \frac{1}{(1-t)^2}\right)\phi(t)$$

(this is similar to Paulsen [14], but here we obtain a differential equation of higher order). The two-dimensional solution space consists of the linear combinations

$$\alpha\,\frac{(5t^3-20t^2+28t-14)t^3}{(1-t)^2} \;+\; \beta\,\frac{2t-1}{t^2(1-t)^2},$$

with $\alpha, \beta \in \mathbb{R}$ (interestingly, with $\alpha = 0$ and $\beta = -1$ this is $m_0^{(3)}$). It is easy to see that the operator T in the proof of Theorem 3 preserves symmetry about $1/2$ if the distribution μ is symmetric about $1/2$, hence uniqueness of the solution to the above integral equation implies that the third derivative must satisfy $m_1^{(3)}(t) = -m_1^{(3)}(1-t)$. This implies $\alpha = 0$. Integrating three times we obtain

$$m_1(t) \;=\; a_0 \;+\; a_1 t \;+\; a_2 t^2 \;+\; \beta\,(t\log t + (1-t)\log(1-t))$$

with suitable constants a_0, a_1, a_2 and β. Now insert into the above integral equation for m_1 and compare coefficients.

For $k = 2$ we can proceed in a similar fashion and obtain the differential equation

$$\phi'''(t) \;=\; 60\left(\frac{1}{(1-t)^3} - \frac{1}{t^3}\right)\phi(t)$$

for $\phi := m_2^{(4)}$. I have not been able to produce the relevant solution (one particular solution is given by $m_0^{(4)}$). Of course, if interest is primarily in the numerical

answer then one could simply discretize the right hand side of the original integral equation and iterate. As in Section 4.1 of Grübel [8] this strategy can also be used to obtain numerically the distribution functions associated with the distributions $Q_t^{(k)}$, $0 \le t \le 1$, $k \in \mathbb{N}_0$; see Section 4 below.

Proof of Theorem 4. We treat the upper bound first. Let again f_k be the density of Beta$(k+1, k+1)$. It follows from the proof of Theorem 3 that m_k satisfies the integral equation

$$m_k(t) \;=\; 1 \;+\; \int_t^1 x\, m_k\left(\frac{t}{x}\right) f_k(x)\, dx \;+\; \int_0^t (1-x)\, m_k\left(\frac{t-x}{1-x}\right) f_k(x)\, dx,$$

$0 \le t \le 1$. Property (a) of μ-splits implies that

$$\gamma_k \;:=\; \sup_{0 \le t \le 1} m_k(t) \;<\; \infty,$$

and the integral equation leads to the inequality

$$\gamma_k \;\le\; 1 + \gamma_k \rho_k \quad \text{with} \quad \rho_k := \sup_{0 \le t \le 1} \left(\int_t^1 x f_k(x)\, dx \;+\; \int_0^t (1-x) f_k(x)\, dx \right).$$

It is easy to see that $\rho_k = E \max\{V, 1-V\}$, which we computed in the proof of the lemma. Solving for γ_k we obtain the upper bound.

We know from the proof of Theorem 3 that the operator T_k defined by

$$T_k(m)(t) \;:=\; 1 \;+\; \int_t^1 x\, m\left(\frac{t}{x}\right) f_k(x)\, dx \;+\; \int_0^t (1-x)\, m\left(\frac{t-x}{1-x}\right) f_k(x)\, dx$$

is a strict contraction on the space of continuous functions $m : [0,1] \to \mathbb{R}$, and that m_k is the associated unique fixed point. In particular, m_k arises as the uniform pointwise limit of the sequence $(m_{k,l})_{l \in \mathbb{N}_0}$ defined by $m_{k,0} \equiv 2$ and $m_{k,l+1} := T_k(m_{k,l})$ for all $l \in \mathbb{N}_0$. As the mean associated with Beta$(k+1, k+1)$ is $1/2$ it is easy to see that $m_{k,l}(0) = m_{k,l}(1) = 2$ for all $l \in \mathbb{N}_0$, hence it follows that $m_k(0) = m_k(1) = 2$. We also have

$$m_{k,1}(t) \;=\; 1 \;+\; 2\left(\int_0^t (1-x)\, f_k(x)\, dx \;+\; \int_t^1 x\, f_k(x)\, dx \right) \;>\; 2$$

for $0 < t < 1$. To see this we first note that it is enough to consider $0 < t \le 1/2$ because of the same symmetry argument as used in the remarks following the proof of Theorem 3, and that for such t

$$(1-x)\, 1_{[0,t]}(x) \;+\; x\, 1_{(t,1]}(x) \;\ge\; x \qquad \text{for } 0 \le x \le 1,$$

with strict inequality for $0 \le x \le t$. As the Beta$(k+1, k+1)$-probability of the interval $[0, t]$ is strictly greater than 0 the above strict inequality follows.

It is easy to see that T_k is monotone, so that $m_{k,1} \geq m_{k,0}$ implies that $(m_{k,l})_{l \in \mathbb{N}_0}$ is pointwise monotone increasing. In particular, $m_k \geq m_{k,1}$. □

Proof of Theorem 5. Applying FIND recursively we see that the expected number of comparisons needed to determine all presample medians is bounded from above by

$$4 \cdot ER_{n,l}^{[k]} \cdot \max_{1 \leq m \leq n} k_m,$$

where $R_{n,l}^{[k]}$ denotes the number of recursions required by FIND$(\{1, \ldots, n\}, l)$ if the sequence $(k_m)_{m \in \mathbb{N}}$ is employed. Using Jensen's inequality with the concave function $x \to \log x$, $x > 0$, we obtain

$$\log E\left(\left(\frac{4}{3}\right)^{R_{n,l}^{[k]}}\right) \geq E\left(\log\left(\frac{4}{3}\right)^{R_{n,l}^{[k]}}\right) = \left(\log \frac{4}{3}\right) ER_{n,l}^{[k]},$$

which, together with the lemma, yields

$$\left(\frac{4}{3}\right)^{ER_{n,l}^{[k]}} \leq E\left(\left(\frac{4}{3}\right)^{R_{n,l}^{[k]}}\right) \leq n.$$

This shows that $ER_{n,l}^{[k]}$ is bounded by a constant multiple of $\log n$, where the constant depends neither on l nor on $(k_m)_{m \in \mathbb{N}}$. Hence, under the conditions of the theorem and by Markov's inequality, the total number of comparisons required outside the partition loops tends to 0 in probability if scaled by n^{-1} and may therefore be ignored for the rest of the proof.

Let $q_{n,j}$ be the probability that the larger one of the sets S_-, S_+ has j elements if $\#S = n$. For any fixed n these values determine a probability distribution on $\{\lfloor n/2 \rfloor, \ldots, n-1\}$. Let $\kappa_n := \sum_{j=\lfloor n/2 \rfloor}^{n-1} j q_{n,j}$ be the associated first moment. The basic recursion step of FIND leads to

$$EW_n = (n-1) + \sum_{j=\lfloor n/2 \rfloor}^{n-1} q_{n,j} EW_j.$$

With $a_n := n^{-1} EW_n$ this can be rewritten as

$$a_n \leq 1 + \frac{1}{n} \sum_{j=\lfloor n/2 \rfloor}^{n-1} (j q_{n,j})\, a_j.$$

Now we bootstrap: All a_n's are finite (indeed, $W_n \leq n^2$). If $\rho := \sup_{n \in \mathbb{N}} n^{-1} \kappa_n < 1$, then the above inequality implies

$$\sup_{1 \leq m \leq n+1} a_m \leq 1 + \rho \cdot \sup_{1 \leq m \leq n} a_m,$$

i.e. the sequence $(a_n)_{n\in\mathbb{N}}$ is bounded. The same inequality further implies

$$\limsup_{n\to\infty} a_n \;\le 1 \;+\; \limsup_{n\to\infty} a_n \cdot \limsup_{n\to\infty} \frac{1}{n}\,\kappa_n$$

as $q_{n,j} = 0$ for $j < \lfloor n/2 \rfloor$. For the partition rule based on the median of a presample of size $2k_n+1$ it is easy to show that $\lim_{n\to\infty} n^{-1}\kappa_n = 2$ if $k_n \to \infty$ (see the proofs of the lemma and of Th. 4), hence solving for $a := \limsup_{n\to\infty} a_n$ we obtain $a \le 2$. This asymptotic upper bound on the expectation of $n^{-1}W_n$ coincides with the lower bound on $\liminf_{n\to\infty} n^{-1}W_n$, hence the asserted convergence in probability follows on using Markov's inequality. □

4. Discussion

The time required by a stochastic algorithm such as FIND is a random variable Z. Traditional analyses focus on the average performance, *i.e.* the associated expectation EZ, or the worst case behaviour, *i.e.* deterministic upper bounds on Z. Our main results deal with the distribution $\mathcal{L}(Z)$ of Z, which can be used to obtain more detailed information. In particular, the high quantiles of this distribution give the likelihood of excessively large running times, a performance aspect that lies between average behaviour and worst case bounds. Of course, Markov's inequality can be used because of $Z \ge 0$ to obtain upper bounds for the quantiles of $\mathcal{L}(Z)$ from the average quantity EZ *via*

$$P(Z \ge z) \;\le\; \frac{1}{z}\,EZ \qquad \text{for all } z \ge 0,$$

but these bounds are often conservative to the point of being useless; see Grübel [8] for a discussion and numerical results for FIND in the case $k = 0$.

In our analysis the time required by the algorithm is approximated by the total number $C_{n,l_n}^{(k)}$ of comparisons needed in the main recursion loops, a value that depends on the size n of the input set, the rank l_n of the required element and the presample size $2k+1$. The limit result given in Theorem 2 uses a suitable normalization, so that n disappears and the ratio l_n/n becomes a new parameter t. On the whole, we therefore end up with a family $\mathcal{L}(Z_t^{(k)})$ of distributions, with $0 \le t \le 1$ and $k \in \mathbb{N}_0$. While there seems to be little hope to obtain explicit formulae for the associated mean values, let alone the distribution functions, our results go beyond mere existence of the limit as they open up the possibility to obtain these quantities numerically. As explained in the proof of Theorem 3 the mean value function $t \to m_k(t) = EZ_t^{(k)}$ associated with the resulting μ-split satisfies the integral equation

$$m_k(t) \;=\; 1 \;+\; \int_t^1 x\, m_k\left(\frac{t}{x}\right) f_k(x)\,dx \;+\; \int_0^t (1-x)\, m_k\left(\frac{t-x}{1-x}\right) f_k(x)\,dx,$$

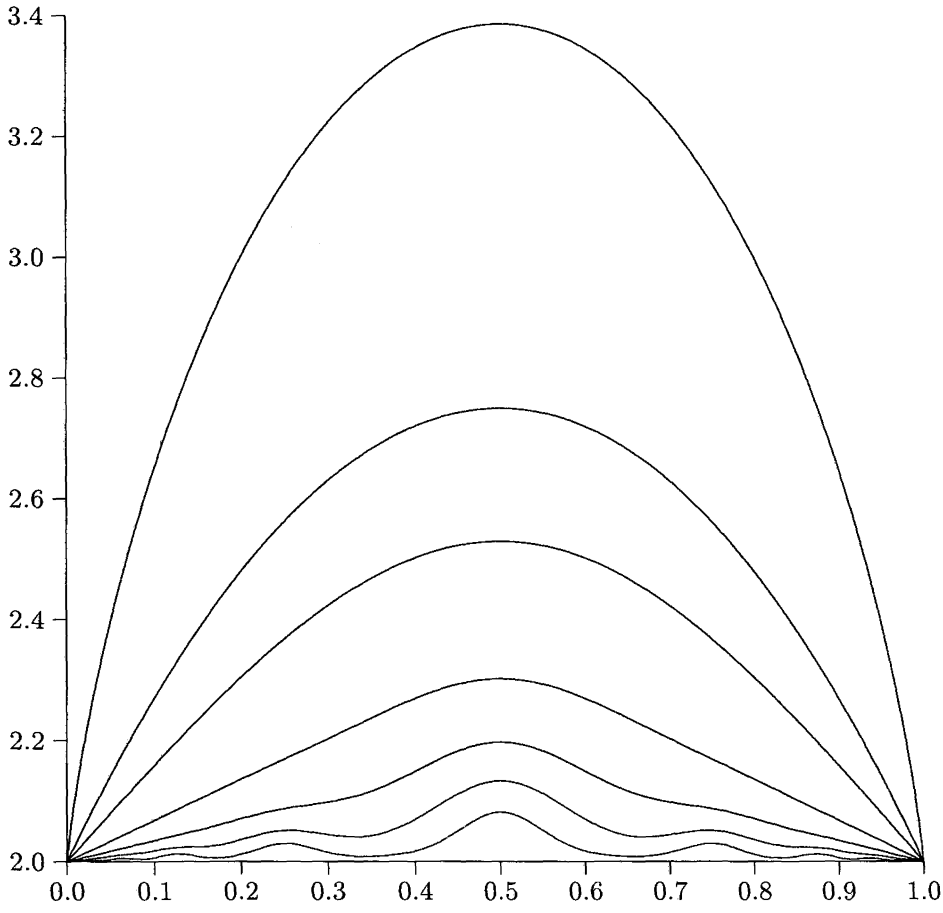

FIGURE 1. Limiting mean number of comparisons for $k = 0, 1, 2, 5, 10, 20, 50$ (top to bottom).

with $f_k(x)$ the density of $\mu = \text{Beta}(k+1, k+1)$. Similarly, we obtain an integral equation for the family of distribution functions $F_k(z|t) = P(Z_t^{(k)} \le z)$ associated with the Beta$(k+1, k+1)$-split,

$$\begin{aligned} F_k(z|t) &= \int_0^t F_k\left(\frac{(z-1)^+}{1-u} \,\middle|\, \frac{t-u}{1-u}\right) f_k(u)\,du \\ &\quad + \int_t^1 F_k\left(\frac{(z-1)^+}{u} \,\middle|\, \frac{t}{u}\right) f_k(u)\,du. \end{aligned}$$

In both cases discretization of the integrals leads to a system $\phi = I(\phi)$ which can easily be treated numerically by iteration $\phi_{m+1} := I(\phi_m)$. Interestingly, the fact that the underlying operators are strict contractions, which was most useful for

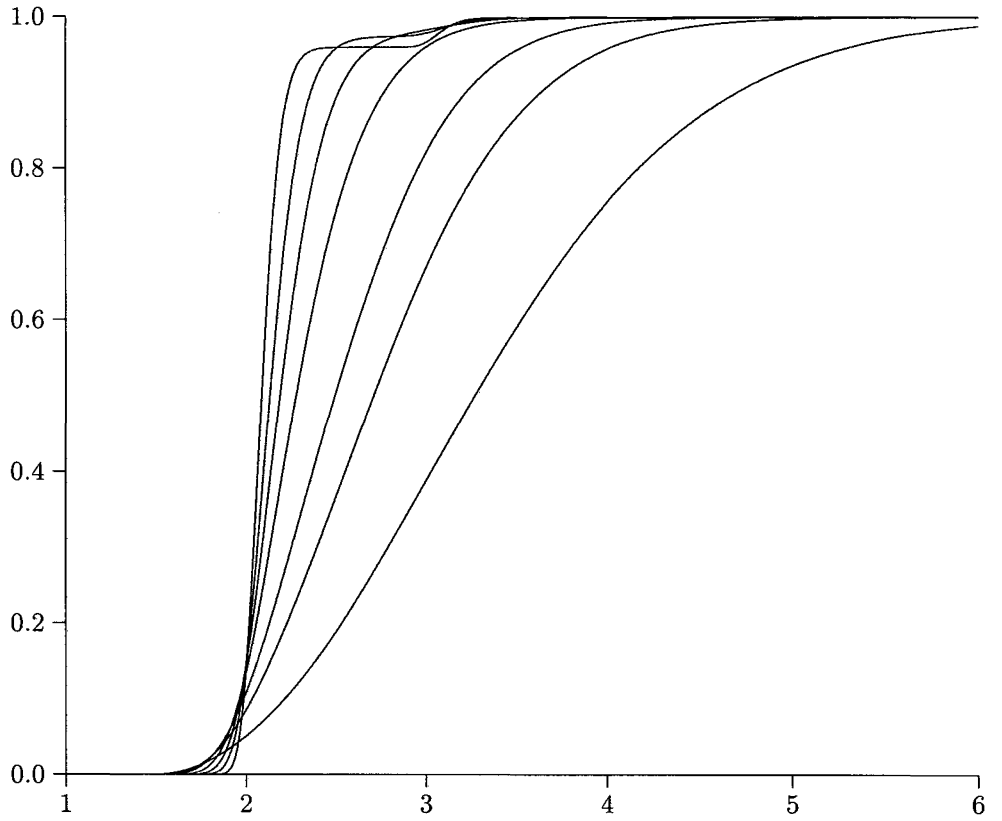

FIGURE 2. Distribution function of $Q_{1/2}^{(k)}$ for $k = 0, 1, 2, 5, 10, 20, 50$ (right to left).

the proofs in Section 3, is also responsible for the smooth working of this numerical approach.

Figures 1 and 2 show the numerical results obtained for various k-values. Figure 1 shows the average values, which are seen to decrease with increasing k. We see that a small value like $k = 5$ already reduces the excess over the lower bound 2 by a considerable amount. Some relatively large k-values have been included to show an initially unsuspected "ripple effect", which is due to the fact that beta distributions become very concentrated about $1/2$ for large k. In particular, for k large enough, $t \to m_k(t)$ is no longer increasing on $[0, 1/2]$ and decreasing on $[1/2, 1]$.

Figure 2 gives the distribution functions $F_k(\cdot\,|1/2)$, *i.e.* the distribution functions of the limiting normalized number of comparisons needed to find the median. Again, for large k an interesting effect occurs: $k \to P(Z_t^{(k)} \leq z)$ is not increasing for all $z > 2$. Nevertheless, it is easy to see that values larger than

average become more unlikely with increasing presample size. For the standard version of FIND we would have about one in ten runs needing more comparisons than 4.72 times the size of the input set, a value that decreases to 3.63, 3.22, 2.77 if a presample of size $2k+1$ with $k = 1$, 2 or 5 respectively is used.

Stimulating comments from both referees have led to an improved version of this paper.

REFERENCES

[1] D.H. Anderson and R. Brown, Combinatorial aspects of C.A.R. Hoare's FIND algorithm. *Australasian J. Combinatorics* **5** (1992) 109-119.
[2] P. Bickel and D.A. Freedman, Some asymptotic theory for the bootstrap. *Annals of Statistics* **9** (1981) 1196-1217.
[3] M. Blum, R.W. Floyd, V. Pratt, R.L. Rivest and R.E. Tarjan, Time bounds for selection. *J. Comput. System Sci.* **7** (1973) 448-461.
[4] L. Devroye, Exponential bounds for the running time of a selection algorithm. *J. Comput. System Sci.* **29** (1984) 1-7.
[5] N. Dunford and J.T. Schwartz, *Linear Operators, Part I: General Theory*. Wiley, New York (1958).
[6] R.W. Floyd and R.L. Rivest, Expected time bounds for selection. *Comm. ACM* **18** (1975) 165-172.
[7] R. Grübel and U. Rösler, Asymptotic distribution theory for Hoare's selection algorithm. *Adv. in Applied Probab.* **28** (1996) 252-269.
[8] R. Grübel, Hoare's selection algorithm: A Markov chain approach. *J. Appl. Probab.* **35** (1998) 36-45.
[9] C.A.R. Hoare, Algorithm 63: PARTITION, Algorithm 64: QUICKSORT, Algorithm 65: FIND. *Comm. ACM* **4** (1961) 321-322.
[10] L. Hyafil, Bounds for selection. *SIAM J. Comput.* **5** (1976) 109-114.
[11] D.E. Knuth, *The Art of Computer Programming* **3**, Sorting and Searching. Addison-Wesley, Reading (1973).
[12] B. Kodaj and T.F. Mori, On the number of comparisons in Hoare's algorithm "Find". *Studia Sci. Math. Hungar.* **33** (1997) 185-207.
[13] J. Neveu, *Mathematische Grundlagen der Wahrscheinlichkeitstheorie*. Oldenbourg, München (1969).
[14] V. Paulsen, The moments of FIND. *J. Appl. Probab.* **34** (1997) 1079-1082.
[15] G.J.E. Rawlins, *Compared to What? An Introduction to the Analysis of Algorithms*. Freedman, New York (1992).
[16] R. Sedgewick and P. Flajolet, *An Introduction to the Analysis of Algorithms*. Addison-Wesley, Reading (1996).

Communicated by I. Wegener.
Received May, 1998. Accepted March, 1999.