The Asymptotic Behaviour of a Distributive Sorting Method W. B. van Dam, J. B. G. Frenk, and A. H. G. Rinnooy Kan

Received November 30, 1982

Abstract --- Zusammenfassung

The Asymptotic Behaviour of a Distributive Sorting Method. In the distributive sorting method of Dobosiewicz, both the interval between the minimum and the median of the numbers to be sorted and the interval between the median and the maximum are partitioned into n/2 subintervals of equal length; the procedure is then applied recursively on each subinterval containing more than three numbers. We refine and extend previous analyses of this method, e.g., by establishing its asymptotic linear behaviour under various probabilistic assumptions.

AMS Subject Classifications: 68 E 05. Key words: Sorting, probabilistic analysis.

Zum asymptotischen Verhalten eines distributiven Suchverfahrens. Bei dem distributiven Sortierverfahren von Dobosiewicz wird sowohl das Intervall zwischen Minimum und Median als auch das Intervall zwischen Median und Maximum in n/2 Teilintervalle gleicher Länge zerlegt; die Prozedur wird dann rekursiv in jedem, mindestens vier Zahlen enthaltenden Teilintervall angesetzt. In dieser Arbeit werden einige Aspekte des Verfahrens verfeinert und erweitert. Insbesondere wird das asymptotisch lineare Verhalten unter verschiedene Wahrscheinlichkeits-Annahmen untersucht.

1. Introduction

The distributive sorting method, proposed by W. Dobosiewicz in [5], has drawn considerable attention. The main reason for this is its attractive combination of worst case and average case properties. As shown by Dobosiewicz, the method combines an $O(n \log n)$ worst case running time (number of comparisons) on one hand, with an O(n) expected running time for the case that the numbers to be sorted are drawn from a *uniform* distribution on the other hand. Below, we refine and extend these results.

In Section 2, we briefly consider the worst case analysis of the method, primarily to correct a deficiency in Dobosiewicz's proof.

In Section 3, we briefly report on some computational experiments that led us to believe that linear expected running times are the rule rather than the exception for this sorting method and should be establishable for many distributions other than the uniform one. In Section 4, this intuition is confirmed. We show that linear expected running time can be demonstrated for any distribution satisfying two conditions: one to avoid excessively peaked distribution functions and one to avoid very thick tails. These conditions are complementary in the sense that if a more stringent version of one is satisfied, then a less stringent version of the other suffices.

In Section 5, we return to the uniform distribution. For a slightly different version of the algorithm, introduced only to simplify the notation, we show that the running time is not only asymptotically linear in expectation but also *in probability*. We conjecture that the result even holds with probability one i.e. *almost everywhere*, and establish a theorem that comes very close to proving this conjecture.

Section 6 contains some open problems and concluding remarks.

2. Worst Case Analysis

Let X be a set containing n numbers $x_1, ..., x_n$. The following distributive method can be used to sort X.

- 1. Find the minimum $x^{(1)}$, the maximum $x^{(n)}$ and the median (the $\lfloor n/2 \rfloor$ -th smallest number, where $\lfloor p \rfloor$ is the integer rounddown of p) $x^{(\lfloor n/2 \rfloor)}$ of X.
- 2. Partition the interval $[x^{(1)}, x^{([n/2])}]$ into [n/2] subintervals $I_1, \ldots, I_{[n/2]}$ of equal length and the interval $[x^{([n/2])}, x^{(n)}]$ into $\langle n/2 \rangle (\langle p \rangle)$ is the *integer roundup* of p) subintervals $I_{[n/2]+1}, \ldots, I_n$ of equal length.
- 3. Distribute the numbers over the subintervals to form groups G_1, \ldots, G_n .
- 4. Repeat the procedure for every group G_i whose cardinality g_i is larger than 3.

If we denote the *running time* (i.e., the number of comparisons) of the above procedure by T(X), the worst case running time is defined by

$$W(n) \stackrel{\triangle}{=} \max_{|X|=n} \{T(X)\}$$
(1)

The analysis of W(n) is based on the intuitive notion that the worst that can happen is for the n/2 elements smaller than the median as well as for the n/2 elements larger than the median to fall in a single group. Since the first three steps can be carried out in linear time [8], i.e. using at most *cn* comparisons, for some constant *c*, this leads to a recurrence relation of the form

$$W(n) \le cn + 2 W(n/2) \tag{2}$$

which provides intuitive justification of the first theorem.

Theorem 1:

$$W(n) = O(n \log n). \tag{3}$$

Proof: In providing a rigorous proof of (3) [1], Dobosiewicz uses the inequality

$$W(2m) \ge 2W(m) \tag{4}$$

which is not obviously true a priori. We start our analysis by correcting this

deficiency. To do so, we consider the worst case running time of the above procedure under the additional assumption that the first three steps require *exactly cn* comparisons, and show that this running time $\overline{T}(X)$ has a worst case behaviour defined by the *equation*

$$\bar{W}(n) = cn + 2 \,\bar{W}(n/2),\tag{5}$$

which can be solved to yield

$$\bar{W}(n) = C n \log n \tag{6}$$

with

$$C = c/\log 2. \tag{7}$$

Since obviously $W(n) \leq \overline{W}(n)$, (3) is an immediate consequence.

We prove (5) by induction on *n*. Suppose that (5) and hence (6) have been established for all $m \le n/2$, and consider a problem instance for which |X| = n:

$$\overline{T}(X) = cn + \sum_{i=1}^{n} \overline{T}(G_i) \leq$$

$$\leq cn + \sum_{i=1}^{n} \overline{W}(g_i) =$$

$$= cn + \sum_{i=1}^{n} C g_i \log g_i =$$

$$= cn + C \left(\sum_{i=1}^{\lfloor n/2 \rfloor} g_i \log g_i + \sum_{i=\lfloor n/2 \rfloor + 1}^{n} g_i \log g_i \right) \leq$$

$$\leq cn + 2 C n/2 \quad \log n/2 =$$

$$= cn + 2 \overline{W}(n/2).$$
(8)

Since the inequality (8) is satisfied for each X, it is easily verified that it is satisfied as an equality for

$$\bar{W}(n) = \max_{|X|=n} \{\bar{T}(X)\},$$

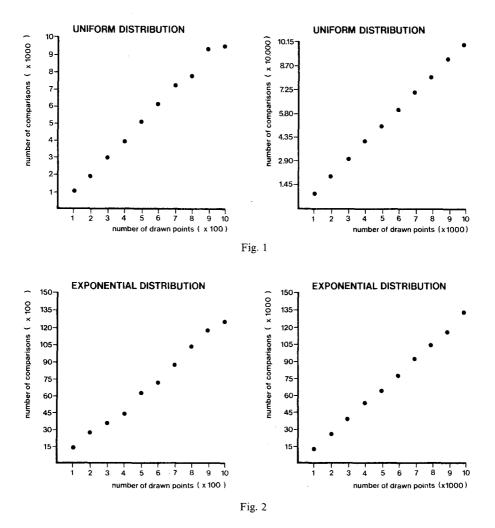
completing the inductive step.

3. Computational Experiments

In [5], Dobosiewicz also considers the *average case* running time of the distributive sorting method, and proves that the procedure runs in O(n) (linear) *expected* time if the numbers are drawn from a *uniform* distribution on [0, 1]. This result is intuitively not surprising, and indeed one suspects that, for many non-uniform distributions, the *recursive* nature of the method ensures that after only a few steps the numbers under consideration are evenly spread, so that the above result applies again.

To test this intuition, we programmed the method in ALGOL and ran two sets of experiments, in which the numbers were drawn from a uniform and an exponential

distribution respectively. The results are depicted in Figs. 1 and 2, and suggest that linear expected running time should occur for many distributions. The analysis in the next section confirms this impression.



4. Average Case Analysis

Suppose that $x_1, ..., x_n$ are drawn according to a density function f that is positive on every finite interval and continuous on $[0, \infty]$ and that satisfies the two following conditions:

(i) there are positive constants δ and D such that for all $|h| \leq \delta$

$$\limsup_{x \to \infty} \frac{f(x+h)}{f(x)} \le D;$$
(9)

(ii) there is a positive constant K such that

$$\lim \sup_{x \to \infty} x \log x \left(1 - F(x) \right) < K \tag{10}$$

where F is the distribution function corresponding to f.

Condition (i) is a *peakedness*-condition: it prevents the density function from being excessively steep.

Condition (ii) is a *tail* condition: it prevents the tail of the distribution from being too thick.

For an average case analysis under these conditions, we define

$$A(n) \stackrel{\Delta}{=} ET(\{\underline{x}_1, \dots, \underline{x}_n\}) \tag{11}$$

and prove the following theorem.

Theorem 2: If f satisfies (i) and (ii), then

$$\limsup_{n \to \infty} \frac{A(n)}{n} < \infty.$$
(12)

Proof: Our proof starts by separate treatment of the case that the maximum is very large. We define the event

$$L_n \stackrel{\triangle}{=} \left\{ \underline{x}^{(n)} \ge \frac{\delta}{4} n \right\}$$
(13)

and write

$$A(n) = E(T(\{\underline{x}_1, ..., \underline{x}_n\}) | L_n) Pr\{L_n\} + E(T(\{\underline{x}_1, ..., \underline{x}_n\}) | L_n^c) Pr\{L_n^c\}.$$
(14)

In view of the worst case analysis in Section 2,

$$E\left(T\left(\{\underline{x}_1,\ldots,\underline{x}_n\}\right)|L_n\right) \cdot Pr\left\{L_n\right\} \le C n \log n \Pr\left\{L_n\right\}.$$
(15)

Condition (ii) implies that

$$1 - F(x) < \frac{K}{x \log x}.$$
 (16)

Hence, for n sufficiently large,

$$Pr\left\{\underline{x}^{(n)} \ge \frac{\delta}{4}n\right\} \le \sum_{i=1}^{n} Pr\left\{\underline{x}_i \ge \frac{\delta}{4}n\right\} = O(1/\log n).$$
(17)

By substituting into (15), it follows immediately that

$$E\left(T\left(\{\underline{x}_{1},\ldots,\underline{x}_{n}\}\right)|L_{n}\right)\cdot Pr\left\{L_{n}\right\}=O\left(n\right).$$
(18)

We analyze the second expectation in (14) by conditioning on values $x^{(1)}$, $x^{([n/2])}$ and $x^{(n)}$ for the minimum, median and maximum respectively, with

$$0 \le x^{(1)} \le x^{([n/2])} \le x^{(n)} < \frac{\delta}{4} n.$$
(19)

We define

$$P_{n}(x^{(1)}, x^{([n/2])}) \stackrel{\triangle}{=} Pr\left\{\underline{x}^{(1)} \le x^{(1)}, \underline{x}^{([n/2])} \le x^{([n/2])}\right\}$$
(20)

$$P'_{n}(x^{([n/2])}, x^{n}) \stackrel{\triangle}{=} Pr\left\{\underline{x}^{([n/2])} \le x^{([n/2])}, \underline{x}^{(n)} \le x^{(n)}\right\}$$
(21)

and observe from the discription of the procedure that

-

$$E\left(T\left(\{\underline{x}_{1},...,\underline{x}_{n}\}\right)|L_{n}^{*}\right) \le cn+$$

$$+\iint E\left(\sum_{i=1}^{\lfloor n/2 \rfloor} T(\underline{G}_{i})|\underline{x}^{(1)} = x^{(1)}, \underline{x}^{(\lfloor n/2 \rfloor)} = x^{(\lfloor n/2 \rfloor)}\right) dP_{n}(x^{(1)}, x^{(\lfloor n/2 \rfloor)}) +$$

$$+\iint E\left(\sum_{i=\lfloor n/2 \rfloor+1}^{n} T(\underline{G}_{i})|\underline{x}^{(\lfloor n/2 \rfloor)} = x^{(\lfloor n/2 \rfloor)}, \underline{x}^{(n)} = x^{(n)}\right) dP'_{n}(x^{(\lfloor n/2 \rfloor)}, x^{(n)}).$$
(22)

221 7 0

The first integral in (22) can be rewritten as follows:

$$E\left(\sum_{i=1}^{\lfloor n/2 \rfloor} T(\underline{G}_{i}) \mid \underline{x}^{(1)} = x^{(1)}, \ \underline{x}^{(\lfloor n/2 \rfloor)} = x^{(\lfloor n/2 \rfloor)}\right) = \sum_{i=1}^{\lfloor n/2 \rfloor} \sum_{g_{1} + \dots + g_{\lfloor n/2 \rfloor} = \lfloor n/2 \rfloor} E\left(T(\underline{G}_{i}) \mid \underline{g}_{i} = g_{i}\right) \cdot Pr\left\{\underline{g}_{1} = g_{1}, \dots, \underline{g}_{\lfloor n/2 \rfloor} = g_{\lfloor n/2 \rfloor}\right\}$$
(23)

where $(g_1, \ldots, g_{\ln/21})$ satisfy a multinomial distribution with cell probabilities

$$P_{i} = \frac{\int_{I_{i}} f(x) dx}{\int_{x^{(1n/2)}} f(x) dx} \qquad (i = 1, ..., [n/2]).$$
(24)

It follows that (23) is equal to

$$\sum_{i=1}^{[n/2]} \sum_{g_i=0}^{[n/2]} E(T(G_i) | \underline{g}_i = g_i) \cdot Pr\{\underline{g}_i = g_i\}$$
(25)

where \underline{g}_i satisfies a *binomial* distribution with parameters [n/2] and P_i (i=1,...,[n/2]). We now complete our analysis by proving that for all *i* there exists a constant M (independent of *i*, $x^{(1)}$ and $x^{(n/2)}$) such that

$$E(T(\underline{G}_i)|\underline{g}_i = \underline{g}_i) \le M \underline{g}_i.$$
⁽²⁶⁾

If we substitute this result in (25), we immediately obtain that the first integral in (22) is $0(\lfloor n/2 \rfloor)$. In a similar way, the second one is $0(\langle n/2 \rangle)$ and together with (18), this concludes the proof.

To prove (26) for all *i*, we map the interval $I_i = [y_i, y_{i+1}]$ onto [0, 1] by means of a transformation, which consists of a translation followed by a multiplication. Since the sorting method is invariant under such a transformation, we obtain immediately that

$$E(T(\underline{G}_{i})|g_{i}=g_{i}) = ET(\{\underline{x}_{1},...,\underline{x}_{g_{i}}\})$$
(27)

where \underline{x}_j are random variables on [0, 1] with distribution function

The Asymptotic Behaviour of a Distributive Sorting Method

$$F_{i}(x) = \frac{F(\gamma x + y_{i}) - F(y_{i})}{F(y_{i+1}) - F(y_{i})}$$
(28)

where

$$\gamma \stackrel{\triangle}{=} (x^{([n/2])} - x^{(1)}) / [n/2]$$
(29)

so that

$$y_i = x^{(1)} + (i-1)\gamma$$
 $(i=1,...,[n/2]).$ (30)

The density function corresponding to (28) is given by

$$f_i(x) = \frac{\gamma f(\gamma x + y_i)}{F(y_{i+1}) - F(y_i)}.$$
(31)

Since f is positive and continuous, the mean value theorem ([7; p. 23]) implies that there exists $\theta \in [0, 1]$ such that the denominator of (31) can be written as $\gamma f(\gamma \theta + y_i)$. By taking $z = \gamma x + y_i$, this implies that, for some θ' depending on x and i, with $|\theta'| \le 1$,

$$f_i(x) = \frac{f(z)}{f(z + \gamma \theta')}.$$
(32)

Now, condition (i) implies that for all i and all $x^{(1)}$, $x^{(ln/2)}$ satisfying (19), we have that

$$f_i(x) < M. \tag{33}$$

However, this implies that the conditions for application of Theorem 1 in [4] are satisfied. This theorem establishes that (33) implies expected linear running time and hence we may conclude that (33) implies the validity of (26) for every i.

Conditions (i) and (ii) are in a sense complementary: one can be relaxed at the expense of the other. More precisely, Theorem 2 can be established under the two conditions that for some k

(I) there are positive constants δ and D such that for $|x^{k-1}h| \leq \delta$

$$\lim \sup_{x \to \infty} \frac{f(x+h)}{f(x)} \le D; \tag{34}$$

(II) there is a positive constant K such that

$$\lim \sup_{x \to \infty} x^k \log x \left(1 - F(x) \right) < K.$$
(35)

The proof follows the same lines as above and is left to the reader.

5. Uniform Distribution

In this section, we return to the uniform distribution. As mentioned above, Dobosiewicz proved linear expected running time for this case in [5]. It is interesting to observe that his analysis hardly exploits the recursive nature of the method; indeed, a simple $O(g_i \log g_i)$ upper bound on the effort required to sort the groups G_i formed initially is all that is required for the proof. This feature has been made use of in several nonrecursive variations on distributive sorting ([1], [9]).

Below, we present an analysis that is essentially recursive and that allows us to extend Dobosiewicz's initial result so as to prove convergence to linear running time *in probability*. To facilitate the exposition we prove this result for a simplified version of the method, in which the median is not used; rather, in Step 2, the interval between $x^{(1)}$ and $x^{(n)}$ is divided into *n* equal length subintervals I_1, \ldots, I_n , again corresponding to groups G_1, \ldots, G_n . All results, however, apply to the original version as well.

The first steps of our analysis are very similar to those in the previous section. We observe that in the case of a uniform distribution, the distribution of the order statistics $\underline{x}^{(2)}, \ldots, \underline{x}^{(n-1)}$ given $\underline{x}^{(1)} = x^{(1)}, \underline{x}^{(n)} = x^{(n)}$ is equal to the distribution of n-2 order statistics drawn from a uniform distribution on $[x^{(1)}, x^{(n)}]$. Hence, for all $x^{(1)}, x^{(n)}$ with $0 \le x^{(1)} \le x^{(n)} \le 1, (\underline{g}_1, \ldots, \underline{g}_n)$ satisfy a multinomial distribution of size n-2 with cell probabilities all equal to 1/n. If, as in Section 2, we analyze $\overline{T}(n)$ rather than T(n), we find (cf. (25)) that

$$\overline{A}(n) \stackrel{\Delta}{=} E\overline{T}(\{\underline{x}_1, \dots, \underline{x}_n\}) = cn +$$

$$+ \sum_{i=1}^n \sum_{g_i=0}^{n-2} \iint E(\overline{T}(\underline{G}_i) | \underline{g}_i = g_i) \cdot Pr(\underline{g}_i = g_i) dP''_n(x^{(1)}, x^{(n)})$$
(36)

with

$$P_n''(x^{(1)}, x^{(n)}) = \Pr\left\{\underline{x}^{(1)} \le x^{(1)}, \underline{x}^{(n)} \le x^{(n)}\right\}.$$
(37)

Again, we map each interval I_i onto [0, 1] to obtain that

$$E\left(\overline{T}(G_i) \mid \underline{g}_i = g_i\right) = E\overline{T}\left(\{\underline{x}_1, \dots, \underline{x}_{g_i}\}\right)$$
(38)

where in this case \underline{x}_i $(j = 1, ..., g_i)$ are independent uniformly distributed on [0, 1]. Since (38) does not depend on i, $x^{(1)}$ and $x^{(n)}$, we find from (36) that $\overline{A}(n)$ satisfies the following recurrence:

$$\overline{A}(n) = cn + n \, E \overline{A}(\underline{u}_{n-2}) \tag{39}$$

where u_{n-2} satisfies a binomial distribution with parameters n-2 and 1/n.

In a similar fashion, we now want to establish a recurrence for

$$\overline{V}(n) \stackrel{\triangle}{=} E \overline{T}^2(\{\underline{x}_1, \dots, \underline{x}_n\}).$$
(40)

We find that

 \overline{V}

$$(n) = c^{2} n^{2} + 2 cn \sum_{i=1}^{n} \iint E \overline{T}(\underline{G}_{i}) dP_{n}^{"}(x^{(1)}, x^{(n)}) + \\ + \iint E \left(\left(\sum_{i=1}^{n} \overline{T}(\underline{G}_{i}) \right)^{2} \right) dP_{n}^{"}(x^{(1)}, x^{(n)}) \\ = c^{2} n^{2} + 2 cn \sum_{i=1}^{n} \iint E \overline{T}(\underline{G}_{i}) dP_{n}^{"}(x^{(1)}, x^{(n)}) + \\ + \sum_{i=1}^{n} \sum_{\substack{j=1\\ j\neq i}}^{n} \iint E \left(\overline{T}(\underline{G}_{i}) \overline{T}(\underline{G}_{j}) \right) dP_{n}^{"}(x^{(1)}, x^{(n)}) + \\ + \sum_{i=1}^{n} \iint E \overline{T}^{2}(\underline{G}_{i}) dP_{n}^{"}(x^{(1)}, x^{(n)}).$$

$$(41)$$

Let us consider the term

$$\iint E\left(\bar{T}(\underline{G}_i)\,\bar{T}(\underline{G}_j)\right)dP_n''(x^{(1)},x^{(n)}).\tag{42}$$

Again we condition on possible values of g_1, \ldots, g_n to find that (42) is equal to

$$\iint \sum_{\substack{g_i + g_j = 0 \\ Pr \{\underline{g}_i = g_i, \underline{g}_j = g_j\}}}^{n-2} E\left(\bar{T}(\underline{G}_i) \, \bar{T}(\underline{G}_j) \, | \, \underline{g}_i = g_i, \underline{g}_j = g_j\right) \cdot Pr\left\{\underline{g}_i = g_i, \underline{g}_j = g_j\right\} \, dP''_n(x^{(1)}, x^{(n)})$$
(43)

where, for all *i*, *j* and $x^{(1)}$, $x^{(n)}$ such that $0 \le x^{(1)} \le x^{(n)} \le 1$, $(\underline{g}_i, \underline{g}_j)$ now satisfies a *trinomial* distribution with parameters n-2, $p_i = 1/n$, $p_j = 1/n$. Because of the mutual independence between I_i and I_j , (42) is therefore equal to

$$\sum_{g_i+g_j=0}^{n-2} \overline{A}(g_i) \overline{A}(g_j) \Pr\left\{\underline{g}_i = g_i, \underline{g}_j = g_j\right\}$$
(44)

and by summing over all i and j ($j \neq i$) we obtain that the third term in (41) is equal to

$$n(n-1)E(\bar{A}(\underline{v}_{n-2})\bar{A}(\underline{w}_{n-2})), \qquad (45)$$

where $(\underline{v}_{n-2}, \underline{w}_{n-2})$ is trinomially distributed with parameters n-2, 1/n and 1/n.

The other terms in (41) can be dealt with analogously, and we obtain

$$\overline{V}(n) = c^2 n^2 + 2 cn^2 E\left(\overline{A}(\underline{v}_{n-2})\right) + n(n-1) E\left(\overline{A}(\underline{v}_{n-2}) \overline{A}(\underline{w}_{n-2})\right) + n E \overline{V}(\underline{v}_{n-2}).$$
(46)

We shall now analyze the asymptotic form of recurrences (39) and (46).

We start with (39). It is well known ([3], [6]) that \underline{u}_{n-2} converges in distribution to a random variable \underline{u} that is Poisson distributed with parameter 1. Lemma A in the Appendix extablishes that $E\overline{A}(\underline{u}_{n-2})$ converges to $E\overline{A}(\underline{u})$ as well, and we have arrived at the following refinement of Dobosiewicz's original result.

Theorem 3: If the numbers \underline{x}_j are drawn from a uniform distribution on [0, 1], then

$$\lim_{n \to \infty} \frac{\bar{A}(n)}{n} = c + E\bar{A}(\underline{u}).$$
(47)

Recurrence (46) can be analyzed in a similar manner. It is easy to verify that $(\underline{v}_{n-2}, \underline{w}_{n-2})$ converges in distribution to $(\underline{v}, \underline{w})$ with

$$Pr\{\underline{v}=v, \underline{w}=w\}=e^{-2}\frac{1}{v!}\frac{1}{w!}.$$
(48)

(Note that \underline{v} and \underline{w} are independent.) Lemma A from the Appendix can again be used to prove that

$$E\left(\bar{A}\left(\underline{v}_{n-2}\right)\bar{A}\left(\underline{w}_{n-2}\right)\right)$$

converges to

$$E\left(\bar{A}\left(\underline{v}\right)\bar{A}\left(\underline{w}\right)\right) = E\bar{A}\left(\underline{v}\right)\cdot E\bar{A}\left(\underline{w}\right) \tag{49}$$

and we find that

$$\lim_{n \to \infty} \frac{V(n)}{n^2} = c^2 + 2 c E \overline{A} (\underline{u}) + (E \overline{A} (\underline{u}))^2.$$
(50)

We conclude from (47) and (50) that

$$\lim_{n \to \infty} \operatorname{var}\left(\frac{\bar{T}(\{\underline{x}_1, \dots, \underline{x}_n\})}{n}\right) = \lim_{n \to \infty} \frac{\bar{V}(n)}{n^2} - \left(\frac{\bar{A}(n)^2}{n}\right) = 0.$$
(51)

Through Chebyshev's inequality, we arrive at the desired result.

Theorem 4: If the numbers \underline{x}_i are drawn from a uniform distribution on [0, 1], then

$$\frac{T(\{\underline{x}_1,\ldots,\underline{x}_n\})}{n} \to c + E\overline{A}(\underline{u}) \quad in \ probability.$$
(52)

We now would like to prove that the convergence result established in Theorem 4. does not only hold in probability, but with probability 1 or *almost everywhere* (a.e.). We have not quite been able to prove this result, but have established the following slightly weaker version.

Theorem 5: If $\{a_n\}_{n \in \mathbb{N}}$ is a sequence of natural numbers such that

$$\frac{\overline{T}(\{\underline{x}_1, \dots, \underline{x}_{a_n}\})}{a_n} \to c + E\overline{A}(\underline{u}) \quad a.e.$$
(53)

then

To prove this theorem, we establish the speed of convergence of (47) and (51).

Lemma 1:

$$\lim \sup_{n \to \infty} n \left| \frac{\bar{A}(n)}{n} - c - E\bar{A}(\underline{u}) \right| < \infty.$$
(54)

Proof: In Lemma B of the Appendix, we establish the speed at which $E\overline{A}(\underline{u}_n)$ converges to $E\overline{A}(\underline{u})$:

$$\lim \sup_{n \to \infty} n |E\bar{A}(\underline{u}_n) - E\bar{A}(\underline{u})| < \infty.$$
(55)

The lemma is an immediate consequence of this result.

Lemma 2:

$$\lim \sup_{n \to \infty} n \operatorname{var}\left(\frac{\overline{T}(\{\underline{x}_1, \dots, \underline{x}_n\})}{n}\right) < \infty.$$
(56)

Proof: The proof is an immediate consequence of Lemma B in the Appendix and its generalization Lemma C. \Box

Lemma 2 and the Borel-Cantelli lemma [2] imply that, if $\sum_{n=1}^{\infty} \frac{1}{a_n} < \infty$,

$$\frac{\overline{T}(\{\underline{x}_1,\ldots,\underline{x}_{a_n}\})}{a_n} - \frac{E\overline{T}(\{\underline{x}_1,\ldots,\underline{x}_{a_n}\})}{a_n} \to 0 \quad (a.e).$$
(57)

Lemma 1 implies that

$$\frac{E\bar{T}(\{\underline{x}_1, \dots, \underline{x}_{a_n}\})}{a_n} \rightarrow c + E\bar{A}(\underline{u})$$
(58)

and hence

$$\frac{\overline{T}(\{\underline{x}_1,\ldots,\underline{x}_{a_n}\})}{a_n} \to c + E\overline{A}(\underline{u}) \quad (a.e.),$$
(59)

completing the proof of Theorem 5.

We note that all that would be required to convert Theorem 5 into the strongest possible result

$$\frac{\overline{T}(\{\underline{x}_1, \dots, \underline{x}_n\})}{n} \rightarrow c + E\overline{A}(\underline{u}) \quad (a.e.)$$
(60)

is the truth of the following conjecture:

(C) there is a constant $\alpha \in (0, 1)$ and a positive constant M such that

 $\overline{T}(\{\underline{x}_1, \dots, \underline{x}_n\}) \le \overline{T}(\{\underline{x}_1, \dots, \underline{x}_n, \underline{x}_{n+1}\}) + Mn^{\alpha} \quad (a.e.).$ (61)

To see why (C) implies (60), we take $a_n = n^{1+\varepsilon}$ with $0 < \varepsilon < 1 - \alpha$ and choose k(n) such that

$$a_{k(n)} \le n < a_{k(n)+1} \tag{62}$$

i.e.

$$k(n) = \left[n^{1/(1+\varepsilon)} \right] \tag{63}$$

so that

$$n - a_{k(n)} = o\left(n^{\varepsilon}\right) \tag{64}$$

$$a_{k(n)+1} - n = o(n^{\varepsilon}). \tag{65}$$

Hence, from (C)

$$\frac{\overline{T}(\{\underline{x}_{1},...,\underline{x}_{a_{k(n)}}\}) - o(n^{\varepsilon + \alpha})}{a_{k(n)+1}} \leq \frac{\overline{T}(\{\underline{x}_{1},...,\underline{x}_{n}\})}{n} \leq \frac{\overline{T}(\{\underline{x}_{1},...,\underline{x}_{a_{k(n)+1}}\}) - o(n^{\varepsilon + \alpha})}{a_{k(n)}}$$

$$(66)$$

and because $a_{k(n+1)}/a_{k(n)} \to 1$, (66) implies (60).

Condition (C) seems to be a very mild one: it says that $T(\{x_1, ..., x_n\})$ cannot decrease too fast as a function of n. We have been unable to convert our intuitive belief that this must be the case into a rigorous proof.

We conclude this section by observing that the case in which $\underline{x}_1, \dots, \underline{x}_n$ are sampled from an arbitrary distribution on [0, 1] with positive and continuous density function can be analyzed much along the same lines. In particular, we obtain a formula for the asymptotic behaviour of $\overline{A}(n)/n$ that is a direct generalization of (47). We omit the laborious proofs.

6. Concluding Remarks

The analysis of the preceding sections leaves two interesting questions unanswered.

The first one is whether conjecture (C) in the previous section can be proved. We believe that this should be possible; it would establish the linear running time of the distributive sorting method for the uniform case in the strongest possible way.

The second one is even more interesting. In spite of persistent efforts, we have been unable to construct a distribution for which the sorting method yields a superlinear expected running time. We know that such a distribution would have to violate the conditions (I) and (II) of Section 4, and indeed one would guess that such a distribution would be very peaked or would have a very thick tail, to achieve the worst possible configuration at the deepest possible level of the recursion. However, we have been unable to construct such a distribution; the ones that we considered moreover had the property that the numerical precision required to differentiate between the numbers drawn would grow very fast with n. If any finite precision is assumed, then linear expected running time can indeed be established without conditions (I) and (II).

We continue to feel, none the less, that even stronger results can be proved about this remarkable sorting method.

Acknowledgements

We gratefully acknowledge the useful comments by Luc Devroye and the computational assistance of A. A. van Beuzekom and R. Th. Wijmenga.

Appendix

In this appendix we provide a proof of some results, which are partly known from the literature.

The first lemma is a fairly general result on convergence of moments. The second and third lemma strengthen those results for some special cases.

Lemma A:

Suppose $\{F_n\}_{n \in \mathbb{N}}$ is a sequence of distributions on \mathbb{R}^k and that F_n converges in distribution to F. Let $h: \mathbb{R}^k \to \mathbb{R}$ and $p: \mathbb{R}^k \to \mathbb{R}$ be continuous with

$$\lim_{|x|\to\infty} |p(x)| = \infty \quad and \quad \lim_{|x|\to\infty} \left| \frac{h(x)}{p(x)} \right| = 0.$$

Then

$$\lim \sup_{n \to \infty} \int_{\mathbb{R}^{k}} |p(x)| dF_{n}(x) < \infty$$

implies that

$$\lim_{n \to \infty} \int_{\mathbb{R}^k} h(x) dF_n(x) = \int_{\mathbb{R}^k} h(x) dF(x).$$
(67)

Proof: By assumption

$$M \stackrel{\triangle}{=} \lim \sup_{n \to \infty} \int_{\mathbb{R}^{4}} |p(x)| dF_{n}(x) < \infty.$$
(68)

Define

$$A_p \stackrel{\triangle}{=} \{ x \in \mathbb{R}^k \mid |x_i| \le p(i=1,...,k) \}.$$

Since

$$\left|\frac{h(x)}{p(x)}\right| \to 0$$

there exists a positive constant M_1 such that if $x \in A_{M_1}^c$, then

$$\left|\frac{h(x)}{p(x)}\right| \le \frac{\varepsilon}{M}.$$
(69)

This implies that

$$\int_{\mathbb{R}^{k}} |h(x)| dF(x) \le \lim \inf_{n \to \infty} \int_{\mathbb{R}^{k}} |h(x)| dF_{n}(x) \le$$

$$\le C + \frac{\varepsilon}{M} \limsup_{n \to \infty} \int_{A_{M_{1}}^{c}} |p(x)| dF_{n}(x) < \infty.$$
(70)

In view of (70) we can find a positive constant M_2 such that

$$\int_{A_{M_2}^c} |h(x)| \, dF(x) \le \varepsilon \tag{71}$$

Hence with $M_3 \stackrel{\triangle}{=} \max(M_1, M_2)$,

$$\begin{split} \lim \sup_{n \to \infty} \Big| \int_{\mathbb{R}^{4}} h(x) dF_{n}(x) - \int_{\mathbb{R}^{4}} h(x) dF(x) \Big| \leq \\ \leq \lim \sup_{n \to \infty} \Big| \int_{A_{M_{3}}} h(x) dF_{n}(x) - \int_{A_{M_{3}}} h(x) dF(x) \Big| + \\ + \lim \sup_{n \to \infty} \int_{A_{M_{3}}^{e}} |h(x)| dF_{n}(x) + \\ + \int_{A_{M_{3}}^{e}} |h(x)| dF(x). \end{split}$$
(72)

Since F_n converges in distribution to F and h is continuous we obtain that the first term of (72) tends to zero.

(For a proof of this result in the case that k = 1, see [3, p. 163].) By (68), (69) and (71) the second and third term of (72) are negligible and so we obtain the desired result.

Before proving the next two lemmas, we introduce some notation.

The binomial distribution will be denoted by

$$b(k; n, p) \stackrel{\Delta}{=} {\binom{n}{k}} p^k (1-p)^{n-k}, \tag{73}$$

the trinomial distribution by

$$b(k,l;n,p_1,p_2) \stackrel{\triangle}{=} \frac{n!}{k!\,l!\,(n-k-l)!} \, p_1^k \, p_2^l \, (1-p_1-p_2)^{n-k-l},\tag{74}$$

the Poisson distribution by

$$p(k;\lambda) \stackrel{\triangle}{=} e^{-\lambda} \frac{\lambda^k}{k!} \tag{75}$$

and the two-dimensional Poisson distribution by

$$p(k,l;\lambda_1,\lambda_2) \stackrel{\triangle}{=} e^{-\lambda_1} \frac{\lambda_1^k}{k!} e^{-\lambda_2} \frac{\lambda_2^l}{l!}.$$
(76)

Lemma B:

Suppose that $\{\underline{u}_n\}_{n \in \mathbb{N}}$ is a sequence of random variables with

$$Pr\left\{\underline{u}_{n}=k\right\}=b\left(k,n,\frac{1}{n+2}\right)$$

Then

(a) \underline{u}_n converges in distribution to \underline{u} with

$$Pr\{\underline{u} = k\} = p(k;1); \tag{77}$$

(b) $\limsup_{n\to\infty} n |Eh(\underline{u}_n) - Eh(\underline{u})| < \infty$ for every positive sequence $\{h(n)\}_{n\in\mathbb{N}}$ with

$$\sum_{p=1}^{\infty} \frac{h(p)p^2}{p!} < \infty.$$
(78)

Proof: We only state the proof of (b), since (a) is well known.

$$Eh(\underline{u}_{n}) - Eh(\underline{u}) = \sum_{k=0}^{n} h(k) \left(b\left(k; n, \frac{1}{n+2}\right) - p(k; 1) \right) - \sum_{k=n+1}^{\infty} h(k) p(k; 1)$$
(79)

Using an inequality for the binomial distribution ([6; exercise 34, p. 172]) we obtain from (79)

$$Eh(\underline{u}_{n}) - Eh(\underline{u}) \leq \sum_{k=0}^{n} h(k) \left(p\left(k; \frac{n}{n+2}\right) \exp\left(\frac{k}{n+2}\right) - p(k; 1) \right) \leq \\ \leq O\left(\sum_{k=0}^{n} \frac{h(k)}{k!} \left(\exp\left(\frac{k+2}{n+2}\right) - 1 \right) \right) = O\left(\frac{1}{n}\right).$$

$$(80)$$

We now establish a lower bound on $Eh(\underline{u}_n) - Eh(\underline{u})$.

Applying [6, exercise 34, p. 172] we find that

$$b\left(k;n,\frac{1}{n+2}\right) \ge p(k;1) \exp\left(-\frac{k^2}{n-k}\right) \exp\left(-\frac{2k}{n}\right)$$
 (81)

for *n* sufficiently large and $0 \le k \le n-1$.

This implies that

$$Eh(\underline{u}_{n}) - Eh(\underline{u}) \ge \sum_{k=0}^{n^{2}} h(k) p(k; 1) \left(\exp(-2k/n) \exp(-k^{2}/n - k) - 1 \right)$$

$$- \sum_{k=n^{4}+1}^{\infty} h(k) p(k; 1) \ge$$

$$\ge -O\left(\frac{1}{n}\right) + \sum_{k=2}^{n^{4}} h(k) p(k; 1) \left(\exp(-2k^{2}/n - k) - 1 \right)$$

$$\ge -O\left(\frac{1}{n} \sum_{k=0}^{\infty} \frac{h(k)k^{2}}{k!} \right) \ge -O\left(\frac{1}{n}\right).$$

(82)

Combining (80) and (82) yields the desired result.

Lemma C:

Suppose $\{\underline{v}_n, \underline{w}_n\}_{n \in \mathbb{N}}$ is a sequence of random vectors with

$$Pr\left\{\underline{v}=k, \underline{w}_n=l\right\}=b\left(k, l; n, \frac{n}{n+2}\right).$$

Then

(a) $(\underline{v}_n, \underline{w}_n)$ converges in distribution to $(\underline{v}, \underline{w})$ with

$$Pr\{\underline{u} = k, \underline{v} = l\} = p(k; 1) p(l; 1);$$
(83)

(b)
$$\lim \sup_{n \to \infty} n | E(g(\underline{v}_n) h(\underline{w}_n)) - E(g(v) h(w))| < \infty$$
(84)

for every pair of positive sequences $\{h(n)\}_{n \in \mathbb{N}}$ and $\{g(n)\}_{n \in \mathbb{N}}$ with

$$\sum_{p=1}^{\infty} \frac{h(p) p^2}{p!} < \infty$$

and

$$\sum_{p=1}^{\infty} \frac{g(p)p^2}{p!} < \infty.$$

Proof: As in Lemma B we only prove (b) since (a) is well known ([6, exercise 38, p. 172]).

Before considering $E(g(\underline{v}_n)h(\underline{w}_n)) - E(g(\underline{v})h(\underline{w}))$ we need the following inequalities which can be proved in a similar way as in [6].

$$b\left(k,l;n,\frac{1}{n+2},\frac{1}{n+2}\right) \le p\left(k,l;\frac{n}{n+2},\frac{n}{n+2}\right) \exp\left(2(k+l)/n+2\right)$$
 (85)

$$b\left(k,l;n,\frac{1}{n+2},\frac{1}{n+2}\right) \ge p\left(k,l;\frac{n}{n+2},\frac{n}{n+2}\right) \exp\left(-(k+l)^2/n - (k+l)\right).$$

$$(86)$$

$$\cdot \exp\left(-4/n + 2\right).$$

By assumption g and h are nonnegative, and

$$G_1 \stackrel{\triangle}{=} \sum_{k=0}^{\infty} g(k) p(k; 1), \quad G_2 \stackrel{\triangle}{=} \sum_{l=0}^{\infty} h(l) p(l; 1)$$

are finite.

Hence we can construct two independent random variables v^* , w^* such that

$$Pr\{\underline{v}^* = k\} = g(k) p(k; 1) G_1^{-1}$$

and

$$Pr\{w^* = l\} = h(l) p(l; 1) G_2^{-1}.$$
(87)

In view of (85), (87) implies that

$$E\left(g\left(\underline{v}_{n}\right)h\left(\underline{u}_{n}\right)\right)-E\left(g\left(\underline{v}\right)h\left(\underline{u}\right)\right)\leq$$

$$\leq\sum_{k+l\leq n}g\left(k\right)h\left(l\right)p\left(k,l;1,1\right)\left(\exp\left(4+2\left(k+l\right)/n+2\right)-1\right)\leq$$

$$\leq O\left(\frac{1}{n}E\left(\underline{v}^{*}+\underline{w}^{*}\right)\right).$$
(88)

We now prove the required lower bound.

Using (86) we obtain

$$E\left(g\left(\underline{v}_{n}\right)h\left(\underline{w}_{n}\right)\right) - E\left(g\left(\underline{v}\right)h\left(\underline{w}\right)\right) \geq \\ \geq \sum_{k+l \leq n^{4}} g\left(k\right)h\left(l\right)p\left(k,l;1,1\right) \cdot \\ \cdot \left(\exp\left(-2\left(k+l\right)/n\right)\exp\left(-\left(k+l\right)^{2}/n-\left(k+l\right)\right)\exp\left(-4/n+2\right)-1\right) \\ - \sum_{k+l \geq n^{4}} g\left(k\right)h\left(l\right)p\left(k,l;1,1\right).$$
(89)

Consider the first term of (89).

It follows easily that this term is bounded by

$$\frac{1}{n}O(E((\underline{v}^* + \underline{w}^*)^2)).$$
(90)

The second term of (89) can be bounded by

$$G_1 G_2 \sum_{k+l \ge n^{\ddagger}} \Pr \{ \underline{v}^* = k \} \Pr \{ \underline{w}^* = l \} \le$$

$$\leq G_1 G_2 \sum_{k+l \ge n^{\ddagger}} \Pr \{ \underline{w}^* + \underline{v}^* = k + l \} = G_1 G_2 \Pr \{ \underline{w}^* + \underline{v}^* \ge n^{\frac{1}{2}} \}.$$

Hence by Chebyshev's inequality, it is bounded by

$$G_1 G_2 \frac{E((v^* + w^*)^2)}{n}.$$
(91)

Combining (89), (90), (91) yields a lower bound which implies the result. \Box

References

- [1] Akl, S. G., Meijer, H.: The design and analysis of a new hybrid sorting algorithm. Information Processing Lett. 10, 213-218 (1980).
- [2] Billingsley, P.: Probability and measure. New York: Wiley 1979.
- [3] Breiman, L.: Probability. Reading, Mass.: Addison-Wesley 1968.
- [4] Devroye, L., Klincsek, T.: Average time behavior of distributive sorting algorithms. Computing 26, 1-7 (1981).
- [5] Dobosiewicz, W.: Sorting by distributive partitioning. Information Processing Lett. 7, 1-6 (1978).
- [6] Feller, W.: An introduction to probability theory and its applications, Vol. I. New York: Wiley 1970.
- [7] Kawata, T.: Fourier analysis in probability theory. New York: Academic Press 1972.
- [8] Knuth, D. E.: The art of computer programming. Reading, Mass.: Addison-Wesley 1973.
- [9] Van der Nat, M.: A fast sorting algorithm, a hybrid of distributive and merge sorting. Information Processing Lett. 10, 163-167 (1980).

Dr. W. B. van Dam Department of Industrial Engineering and Management Science Eindhoven University of Technology P.O. Box 513 5600 MB Eindhoven The Netherlands

Prof. Dr. A. H. G. Rinnooy Kan Econometric Institute Erasmus University Rotterdam P.O. Box 1738 3000 DR Rotterdam The Netherlands Dr. J. B. G. Frenk Department of Industrial Engineering and Operations Research University of California Berkeley, CA 94720, U.S.A.