Computing 26, 1—7 (1981)

Computing

Average Time Behavior of Distributive Sorting Algorithms

L. Devroye and T. Klincsek, Montreal

Received February 20, 1980

Abstract — Zusammenfassung

Average Time Behavior of Distributive Sorting Algorithms. In this paper we investigate the expected complexity $E(C)$ of distributive ("bucket") sorting algorithms on a sample $X_1, \ldots, X_n$ drawn from a density f on R^1. Assuming constant time bucket membership determination and assuming the use of an average time $g(n)$ algorithm for subsequent sorting within each bucket (where g is convex, $g(n)/n \uparrow \infty$, $g(n)/n^2$ is nonincreasing and g is independent of f), the following is shown:

1) If f has compact support, then $\int g(f(x))\,dx < \infty$ if and only if $E(C) = O(n)$.

2) If f does not have compact support, then $E(C)/n \to \infty$.

No additional restrictions are put on f.

Key words and phrases: Distributive sorting, bucket sorting, average complexity, expected running time.

CR Categories: 5.31.

Mittleres Zeitverhalten von Fachsortier-Algorithmen. Wir untersuchen die mittlere Komplexität $E(C)$ von Fachsortier-Algorithmen, die auf eine Stichprobe $X_1, \ldots, X_n$ mit der Verteilungsdichte f auf R^1 angewendet werden. Wir nehmen an, daß die Zeit zur Bestimmung des Sortierfachs konstant ist, und daß für die Sortierung innerhalb jedes Fachs ein Algorithmus mit dem mittleren Zeitbedarf $g(n)$ zur Verfügung steht. Dabei ist g konvex, $g(n)/n \uparrow \infty$, $g(n)/n^2$ nichtsteigend und g unabhängig von f. Wir zeigen:

1) Wenn f kompakten Träger hat, dann gilt $\int g(f(x))\,dx < \infty$ genau dann, wenn $E(C) = O(n)$.

2) Wenn f keinen kompakten Träger hat, dann gilt $E(C)/n \to \infty$.

Über f benötigen wir keinerlei weitere Voraussetzungen.

1. Distributive Sorting

Consider a sample $X_1, X_2, \ldots, X_n$ of independent identically distributed random variables with density f. This sample is sorted by the method of distributive sorting ("bucket sorting"):

1. Find $Y = \min(X_1, \ldots, X_n)$ and $Z = \max(X_1, X_n)$.
2. Divide (Y, Z) into n equal parts ("buckets") of length $(Z - Y)/n$.
3. Assign each of the X_i's to one of the buckets using a linked list structure to keep track of the memberships.

0010-485X/81/0026/0001/$ 01.40

4. Sort the sample by scanning from bucket to bucket and using a sorting algorithm SORT for subsequent sorting in each bucket.

Sorting algorithms with this structure are well-known (see [1—2], [4]). In the expected time behavior analysis that follows we make the following assumptions:

(i) There is a constant k such that steps 1—3 can be executed in time less than kn; the constant k does not depend upon $X_1, \ldots, X_n$. We are thus tacitly assuming that the determination of membership in one of n buckets is a constant time operation.

(ii) The lower level algorithm SORT sorts n points $x_1, \ldots, x_n$ in time $h(r_1, \ldots, r_n)$ where $(r_1, \ldots, r_n)$ are the ranks of $x_1, \ldots, x_n$ and h is a given function. Its expected running time,

$$g(n) = \frac{1}{n!} \sum_{(r_1, \ldots, r_n)} h(r_1, \ldots, r_n)$$

satisfies: g is convex; $g(n)/n \uparrow \infty$ and $g(n)/n^2 \downarrow$ for all $n \in (0, \infty)$. For example, bubble sort ($g(n) = c n^2$) and most binary sorting algorithms ($g(n) = c(n+1) \log(n+1)$) satisfy this condition.

With this setup, the overall complexity C of the algorithm satisfies

$$E(C) = 0(n) + E\left(\sum_{i=1}^{n} g(N_i) \right) \tag{1}$$

where $E(.)$ denotes "expected value" (C itself is a random variable), and $N_1, \ldots, N_n$ are the number of points in the buckets numbered $1, \ldots, n$. Since the minimum and the maximum are already isolated, we have $N_1 + \ldots + N_n = n-2$. In this note we study the question: when is $E(C) = 0(n)$?

The catalyst for our work was Dobosiewicz's result on linear expected complexity sorting methods [2].

2. Main Results

We say a density f has *compact support* when there exists a finite K such that $P(|X| > K) = 0$ where X has density f.

Theorem 1: *When f has compact support, then $E(C) = 0(n)$ if and only if*

$$\int g(f(x))\, dx < \infty. \tag{2}$$

Proof:

We recall that the support S of a random variable X_1 is the smallest closed set such that $P(X_1 \in S) = 1$. Then ess inf $X_1 = \inf\{x \mid x \in S\}$, ess sup $X_1 = \sup\{x \mid x \in S\}$.

We can assume without loss of generality that ess inf $X_1 = 0$, ess sup $X_1 = 1$. We also introduce the following notation: $R = Z - Y$; $(x_1, x_2), \ldots, (x_n, x_{n+1})$ is the partition of $[Y, Z]$ into n intervals of equal length R/n. Furthermore, for $x \in (x_i, x_{i+1})$

$$f_n^+(x) = \int_x^{x_{i+1}} f(x)\,dx/(x_{i+1}-x),$$

$$f_n^-(x) = \int_{x_i}^{x} f(x)\,dx/(x-x_i),$$

$$f_n^*(x) = \min\,(f_n^+(x), f_n^-(x)).$$

For $x \notin (Y, Z)$, let $f_n^+(x) = f_n^-(x) = f_n^*(x) = 0$. Finally, let p_i be the probability contents of (x_i, x_{i+1}): $p_i = \int_{x_i}^{x_{i+1}} f(x)\,dx$.

Let us define p by

$$p = \sum_{i=1}^{n} p_i.$$

Clearly, p itself is a random variable, and by the probability integral transform, it is distributed as the range of n independent identically distributed uniform (0, 1) random variables. It has density

$$h(x) = n(n-1)\,x^{n-2}(1-x), \quad 0 < x < 1,$$

and it is easily seen that $E(p^{-1})$ and $E(p^{-2})$ tend to 1 as $n \to \infty$.

By lemma 1 (see Appendix), Jensen's inequality and the fact that $g(a\,x) \le a\,g(x)$, all a, $x > 0$, $a \le 1$, we have

$$\begin{aligned}
E(C \mid Y, Z) &= 0(n) + E\left(\sum_{i=1}^{n} g(N_i) \mid Y, Z\right) \\
&\le 0(n) + 3\sum_{i=1}^{n} g(n\,p_i/p) \\
&\le 0(n) + 3\sum_{i=1}^{n} \frac{1}{x_{i+1}-x_i} \int_{x_i}^{x_{i+1}} g\left(n\,(x_{i-1}-x_i)\,\frac{f(x)}{p}\right) dx \\
&= 0(n) + \frac{3\,n}{R} \int_Y^Z g\left(\frac{R}{p} f(x)\right) dx \\
&\le 0(n) + 3\,n \int_0^1 g(f(x))\,dx\,\max\,(p^{-1}, R\,p^{-2}).
\end{aligned}$$

Since $E(p^{-1})$ and $E(R\,p^{-2})$ remain bounded we have $E(C) = 0(n)$ as claimed whenever $g(f) \in L^1$.

Suppose next that $E(C) = 0(n)$. We have

$$\begin{aligned}
E(C \mid Y, Z) &\ge \frac{1}{2} \sum_{i=1}^{n} g(n\,p_i) \\
&\ge \frac{n}{2\,R} \sum_{i=1}^{n} \int_{x_i}^{x_{i+1}} g\left(n\,(x_{i+1}-x_i)\,f_n^*(x)\right) dx \\
&\ge \frac{n\,R}{2} \int_Y^Z g(f_n^*(x))\,dx
\end{aligned}$$

where we used the inequality $g(a\,x) \ge a^2\,g(x)$, all $a, x > 0$, $a \le 1$.

Let $E \subseteq [0,1]$ be the Lebesgue set of f, namely, the set of x's on which

$$\frac{1}{\delta} \int_x^{x+\delta} f(y)\,dy \to f(x) \text{ as } \delta \to 0$$

and

$$\frac{1}{\delta} \int_{x-\delta}^{x} f(y)\,dy \to f(x) \text{ as } \delta \to 0.$$

Let A be the set of $\omega \in \Omega$ ($(\Omega, \mathcal{F}, \mathcal{P})$ is our probability space for $X_1, X_2, \ldots$) for which $Y \to 0$, $Z \to 1$ as $n \to \infty$. Clearly $P(A)=1$, and $\int_E dx = 1$ (Lebesgue density theorem; see for instance, Stein [7]). On $A \times E$ we have

$$R\, I_{[Y<x<Z]}\, g(f_n^*(x)) \to g(f(x))$$

as $n \to \infty$. Also, $\int_A \int_E dx\, \mathcal{P}(d\omega) = 1$. Thus, by Fatou's Lemma and our hypothesis,

$$\begin{aligned} \infty > \liminf_{n\to\infty} \frac{2E(C)}{n} &\geq \int_A \int_E \liminf_{n\to\infty} R\, I_{[Y<x<Z]}\, g(f_n^*(x))\, dx\, \mathcal{P}(d\omega) \\ &= \int_A \int_E g(f(x))\, dx\, \mathcal{P}(d\omega) \\ &= \int g(f(x))\, dx. \end{aligned}$$

This concludes the proof of Theorem 1.

Remark 1: Theorem 1 does not impose *any* continuity conditions on f. In fact, *all* bounded densities satisfy (2). Only very peaked densities violate (2). For example, if $-1 < a < 0$ and

$$f(x) = \begin{cases} c\,x^a, & 0 < x < 1 \\ 0, & \text{elsewhere,} \end{cases}$$

then (2) holds for $g(n) = n^2$ if and only if $a > -1/2$. With $g(n) = n \log(n)$, (2) holds for all a in $(-1, 0)$. Notice however that some unbounded densities have such a weak peak that $\int f^k(x)\,dx < \infty$ for all $k > 0$. Take for instance

$$f(x) = \begin{cases} -\log(x), & 0 < x < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

The next result resolves the problem for all densities not having compact support.

Theorem 2: *If f does not have compact support, then $E(C)/n \to \infty$.*

Proof: By Lemma 1,

$$E(C) \geq \frac{1}{2} E\left(\sum_{i=1}^{n} g(n p_i) \right)$$

where we inherit the notation of the proof of Theorem 1. Let a and b be 1/3 and 2/3 quantiles of f, that is, points with the property that $P(X < a) = 1/3$ and $P(X < b) = 2/3$. Note that these points may not be unique but that in

any case $a<b$. If q is the number of intervals ("buckets") that have a nonempty intersection with (a, b), then

$$\frac{n(b-a)}{Z-Y}+2 \geq q \geq \frac{n(b-a)}{Z-Y} \quad \text{whenever } Y \leq a \text{ and } Z \geq b.$$

Let Q be the collection of these intervals and let p_i' be the probability contents of the i-th interval after intersection with (a, b). By Jensen's inequality, when I is the indicator function of the event $[Y \leq a,\ Z \geq b]$,

$$\begin{aligned}
&E\left(\sum_{i=1}^{n} g(n p_i)\right) \geq E\left(\sum_{i \in Q} g(n p_i)\right) \geq E\left(\sum_{i \in Q} g(n p_i')\right) \\
&\geq E\left(q\, g\left(\frac{n}{q} \sum_{i \in Q} p_i'\right)\right) = E\left(q\, g\left(\frac{n}{3q}\right)\right) \\
&\geq E\left(\left(\frac{n(b-a)}{Z-Y}\right) g\left(\frac{1}{6/n+3(b-a)/(Z-Y)}\right) I\right) \\
&\geq E\left(\left(\frac{n(b-a)}{Z-Y}\right) \min\left(g\left(\frac{n}{12}\right); g\left(\frac{Z-Y}{6(b-a)}\right)\right) I\right).
\end{aligned}$$

From this we see that we may assume that $(Z-Y) \leq n(b-a)/2$ because otherwise the random variable in $E(\cdot)$ is greater than or equal to $2\, g(n/12)\, I$. It suffices then to show that

$$E\left(\frac{I}{Z-Y} g\left(\frac{Z-Y}{6(b-a)}\right)\right) \to \infty.$$

Clearly, for all $K>0$,

$$\begin{aligned}
E\left(\frac{I}{Z-Y} g\left(\frac{Z-Y}{6(b-a)}\right)\right) &\geq \frac{g(K)}{6(b-a)K} P(Z-Y>6(b-a)K,\ Y \leq a,\ Z \geq b) \\
&= \frac{g(K)}{6(b-a)K}(1+o(1)).
\end{aligned}$$

Theorem 2 follows by the arbitrariness of K.

Remark 2: Theorem 2 applies to *all* densities with an infinite tail, such as the family of exponential densities, which includes the normal, gamma and chi-square densities.

Remark 3: Consider the problem of finding the convex hull of $X_1, \ldots, X_n$, a sample of independent identically distributed random vectors from R^2. From Theorem 1 we deduce that Graham's algorithm [3] when modified slightly to incorporate distributive sorting of the angles of the various points, has linear expected complexity whenever the density of X_1 has compact support and is bounded. Of course, this statement should be accompanied by the remarks (i) and (ii) given in the introduction.

3. Order Preserving Transformations

If h is strictly monotonically increasing mapping from R to a finite interval (a, b), then ordering $X_1, X_2, \ldots, X_n$ is essentially equivalent to ordering $h(X_1), \ldots, h(X_n)$. Linear expected complexity can be obtained whenever the density of $h(X_1)$ satisfies (2). Ideally, h should be such that the density of $h(X_1)$ is uniform on (a, b). If f is known beforehand, then the obvious choice for h is F (the distribution function corresponding to f) because $F(X_1)$ is uniformly distributed on (0, 1). In most situations, either F is unknown or the computation of the $F(X_i)$'s is too expensive. Fortunately, there are *simple* transformations h that yield densities for $h(X_1)$ that satisfy (2) for large classes of densities f with infinite tails. We will illustrate this for the important class of densities with exponentially dominated tail.

We say that a density f has an *exponentially dominated tail* if there exist constants $a, b, c > 0$ such that for all $x \in R$,

$$f(x) \le a e^{-b|x|^c}.$$

Most well-known densities belong to this class: the normal, exponential gamma, beta, chi-square and rectangular densities, and all bounded densities with compact support.

We show that the distributive sorting algorithms discussed in this paper when used on $h(X_1), \ldots, h(X_n)$ satisfy $E(C) = 0(n)$ whenever f has an exponentially dominated tail,

$$h(x) = \begin{cases} x/(1+x), & x \ge 0 \\ x/(1-x), & x < 0, \end{cases} \tag{3}$$

and to conditions (i), (ii) we add

(iii) The computation of $h(x)$ takes time $t(x)$ where t is uniformly bounded on R.

Theorem 3: *If f is a density with exponentially dominated tail, if* (i), (ii), (iii) *hold, and if distributive sorting is used on $h(X_1), \ldots, h(X_n)$ where h is given by* (3), *then* $E(C) = 0(n)$.

Proof: We establish that the density of $h(X_1)$ is bounded and apply Theorem 1. The density of $h(X_1)$ is given by

$$f(h^{-1}(x)) \left| \frac{d}{dx} h^{-1}(x) \right|$$

which for $x > 0$ gives

$$f\left(\frac{x}{1-x}\right) \cdot \frac{1}{(1-x)^2} \le a (1-x)^{-2} \exp\left(-b \left| \frac{x}{1-x} \right|^c\right). \tag{4}$$

Using the inequality $e^{-u} < (s/e\,u)^s$, all $s, u > 0$, we see that (4) is not greater than $4a$ on $(0, 1/2)$ and not greater than $a x^{-2} (2/b\,c\,e)^{2/c}$ on $(1/2, 1)$. The case $x < 0$ is treated similarly.

Remark 4: We do not claim that (3) is the best possible transformation; we picked it because it was one of the simplest order preserving transformations for which Theorem 3 holds. It should be pointed out that for densities f with heavy tails (e. g., tails that are decreasing at a polynomial rate), Theorem 3 may no longer be true when (3) is used.

4. Appendix

Lemma 1: *If N is a binomial random variable with parameters n and p, $g(n)/n$ is nondecreasing and $g(n)/n^2$ is nonincreasing, then*

$$\frac{1}{2}\, g(n\,p) \le E\,(g(N)) \le 3\, g(n\,p) + 2\, g(1).$$

Proof: By a well-known binomial inequality (see [6]),

$$E\,(g(N)) \ge g(n\,p)\, P\,(N \ge n\,p) \ge g(n\,p)/2.$$

Furthermore,

$$g(N) = g(N)\, I_{[1\,,\,n\,p]}(N) + g(N)\, I_{[n\,p\,,\,n]}(N) + g(0)\, I_{[0]}(N)$$

$$\le \frac{N}{n\,p}\, g(n\,p) + \frac{N^2}{n^2\,p^2}\, g(n\,p) + g(0)$$

and

$$E\,(g(N)) \le g(0) + g(n\,p)\left(2 + \frac{1}{n\,p}\right) \le g(0) + 3\, g(n\,p)$$

when $n\,p \ge 1$. For $n\,p < 1$, we have

$$E\,(g(N)) \le E\,(N^2\, g(1)) \le 2\, g(1).$$

Note: The lower bound is valid for all nondecreasing g.

References

[1] Baase, S.: Computer algorithms. Introduction to design and analysis. Reading, Mass.: Addison-Wesley 1978.

[2] Dobosiewicz, W.: Sorting by distributive partitioning. Information Processing Letters *7*, 1—6 (1978).

[3] Graham, R. L.: An efficient algorithm for determining the convex hull of a planar set. Information Processing Letters *1*, 132—133 (1972).

[4] Knuth, D. E.: The art of computer programming 3. Reading, Mass.: Addison-Wesley 1973.

[5] Royden, H. L.: Real analysis. Toronto: Macmillan 1968.

[6] Slud, E. V.: Distribution inequalities for the binomial law. The Annals of Probability *5*, 404—412 (1977).

[7] Stein, E. M.: Singular integrals and differentiability properties of functions. Princeton, N. J.: Princeton University Press 1970.

L. Devroye
T. Klincsek
Mc Gill University
School of Computer Science
Burnside Hall
805, Sherbrooke St. West
Montreal, PQ, Canada H3A 2K6

Printed in Austria

Druck: Paul Gerin, A-1021 Wien