INFORMATIQUE THÉORIQUE ET APPLICATIONS

UWE RÖSLER

A limit theorem for "quicksort"

Informatique théorique et applications, tome 25, nº 1 (1991), p. 85-100

<http://www.numdam.org/item?id=ITA_1991__25_1_85_0>

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol. 25, n° 1, 1991, p. 85 à 100)

A LIMIT THEOREM FOR "QUICKSORT" (*)

by Uwe RÖSLER [1]

Communicated by P. FLAJOLET

Abstract. – *Let X_n be the number of comparisons needed by the sorting algorithm Quicksort to sort a list of n numbers into their natural ordering. We show that $(X_n - E(X_n))/n$ converges weakly to some random variable Y. The distribution of Y is characterized as the fixed point of some contraction. It satisfies a recursive equation, which is used to provide recursive relations for the moments. The random variable Y has exponential tails. Therefore the probability that Quicksort performs badly, e.g. that X_n is larger than $2E(X_n)$ converges polynomially fast of every order to zero.*

Résumé. – *Soit X_n le nombre de comparaisons utilisées par la procédure Quicksort pour trier une liste de nombres distincts. Nous démontrons que $(X_n - \mathrm{E}(X_n))/n$ converge faiblement vers une certaine variable aléatoire Y. La distribution de Y est le point fixe d'une contraction et peut être calculée numériquement par itération.*

0. INTRODUCTION

Probably the most widely used sorting algorithm is the algorithm "Quicksort" invented by C. A. R. Hoare in 1961, 1962. It is, for instance, the standard sorting procedure in Unix systems. The basic idea is as follows:

A list of n (different) real numbers is given. Select an element x from this list. Divide the remaining into sets of numbers smaller and larger than x. Next apply the same procedure to each of these two sets if they contain more than one element. Finally, we end up with a sorted list of the original numbers.

Our selection of the element x is by random choice with equal probability. The reason for this is given at the end of the introduction.

(*) Received September 1988, accepted May 1990.

([1]) Matematiska Institutionen, Thunbergsvägen 3, S-752 38 Uppsala, Sweden and Institut für Mathematische Stochastik, Lotzestrasse 13, 3400 Göttingen, Federal Republic of Germany.

Quicksort has many nice features, such as working *in situ* (using only a small auxiliary stack). For this and the importance of Quicksort we refer the reader to the general literature (Knuth, 1973; Sedgewick, 1988). Quicksort is the fastest known algorithm for sorting. This is mainly due to the conceptually very simple "inner loop".

Denote by X_n the total number of comparisons between numbers used to sort a list of length n. We will neglect all other aspects of the algorithm (Sedgewick, 1977) and deal exclusively with the random variable X_n. We think of the time used by the procedure as proportional to X_n.

In general the average of X_n for any sorting algorithm is greater than or equal to the entropy $\sum p_\pi \log_2 p_\pi$. Here the summation is over all permutations π of $\{1, \ldots, n\}$ and p_π denotes the probability of the permutation π. (Obviously one can identify any list of length n with a permutation of $1, \ldots, n$.) In case that all permutations have the same probability $1/n!$ we obtain the lower bound $n \log_2 n$. This bound is also called the information theoretical lower bound. The random variables themselves have no lower or upper bound. (Under fairly general assumptions n is a sharp lower bound for X_n.)

The average of X_n for Quicksort is known to be of the order $2n \ln n$ (Knuth, 1973). In the best case Quicksort uses approximately $2n \log_2 n \approx 1.4\ldots n \ln n$ comparisons, the best theoretical lower bound for the average. In the worst case Quicksort needs a horrendous number of $n^2/2$ comparisons. One of our purposes is to show that such bad behavior happens very seldom.

We are interested in the asymptotic behavior of X_n. Our main result is the convergence of $Y_n := (X_n - E(X_n))/n$ to a random variable called Y. This is done in the Wasserstein d_2-metric (d_p-metric) on the space of distribution functions. The Wasserstein d_p-metric is defined by (Cambanis *et al.* 1976; Major, 1978)

$$d_p(F, G) = \inf \| X - Y \|_p$$

where the infimum is over all X with distribution function F and all Y with distribution function G. Here $\|.\|_p$ denotes the L_p-norm, $1 \leqq p < \infty$. Notice d_p convergence is the same as weak convergence and the convergence of the absolute moments of order p (Denker and Rösler, 1985).

A consequence of our results are the following estimates of probabilities by the Markov inequality,

$$P(X_n \geqq 2E(X_n)) \leqq \text{Const.}(p) \ln^{-p} n, \qquad 1 \leqq p < \infty, \qquad n \in \mathbb{N}.$$

The order of $\ln^{-p}$ is pretty weak. However, we will show that Y_n has finite Laplace transforms. Using this Markov inequality gives for any positive λ and any $n \in \mathbb{N}$,

$$P(X_n \geqq 2E(X_n)) \leqq E(\exp(\lambda Y_n)) \exp(-\lambda E(X_n)/n) \leqq \text{Const.}(\lambda)/n^{2\lambda}.$$

The probability of bad behavior of Quicksort becomes extremely small for large n, so we may conclude that Quicksort is reliable.

We took here $X_n \geqq 2E(X_n)$ as our standard of bad performance by Quicksort. If one compares the sorting algorithm Heapsort with Quicksort, then Heapsort has the advantage of always using less then $4n \ln n$ comparisons. Therefore we look at the event that Quicksort needs more than $2E(X_n) \approx 4n \ln n$ comparisons. The average for some version of Heapsort is about $n \log_2 n$, the best one can get. But simulations show that Heapsort uses more time than Quicksort on the average (Loeser, 1974).

We denote the distribution function of some random variable X by $L(X)$. We characterize $L(Y)$ as the fixed point of the function S. The function S maps the set of distribution functions, which have finite variance and zero expectation, to itself and is defined by

$$S(F) := L(\tau V + (1-\tau)\bar{V} + C(\tau)).$$

The random variables τ, V, $\bar{V}$ are independent, τ has a uniform distribution on [0, 1] and the distribution of V and $\bar{V}$ is F. Here C denotes some measurable function, *see* (1.4).

The function S is a contraction with respect to the Wasserstein d_2-metric. This is the main mathematical tool we use. Any sequence F, $S(F)$, $S^2(F)$, ... converges to the unique fixed point $L(Y)$ of S. The distribution of Y satisfies the fixed point relation

$$L(Y) = L(\tau Y + (1-\tau)\bar{Y} + C(\tau)).$$

From this relation we obtain recursive formulas for the higher moments of Y. For example, the variance of Y is three times the variance of $C(\tau)$,

$$\operatorname{Var} Y = \int_0^1 C(x)^2\, dx = 7 - 2/3\,\pi^2.$$

For higher moments the calculations become tedious. Hennequin calculated moments, cumulants and other values numerically to identify the distribution. Assuming always that Y exists, he obtained numerical results as well as a nice structural conjecture on the cumulants.

Not much more is known about the distribution of Y. For example, it is unknown whether $L(Y)$ has a density.

The existence of Y was independently shown by Régnier using martingale arguments. Our results provide more than just the existence, e. g. the fixed point relation, the representation by an infinite sum and finite Laplace transforms.

The Y_n converge to Y in any d_p-metric, $1 \leqq p < \infty$. Moreover, the Laplace transform of Y_n converges to that of Y. This implies an exponential tail of every order.

Sedgewick used a variant of Quicksort in this paper. For small files he switched from Quicksort to a different sorting algorithm, which seemed to speed up sorting. However, an analysis on the number X_n of comparisons for this variant shows that the expectation of X_n has an asymptotic behavior of $2 n \ln n + \mathrm{Const.} . n +$ smaller terms in n (and could be analysed much further). Notice that the leading term $2 n \ln n$ is the same as for standard Quicksort. The linear terms in n may be different. The asymptotic distribution of $(X_n - E(X_n))/n$ converges to Y, $L(Y)$ the fixed point for Quicksort as before, in any d_p-metric, $1 \leqq p < \infty$. Moreover all Laplace transforms converge, in particular Lemma 4.1 and its consequences remain true.

We do not prove this result. The proof is a variant of the given one and requires only obvious changes.

In section 5 we discuss the k-median version (Hoare, 1962).

Devroye (1985) used a variant "Find" to find the k'th largest element of a list. The main object of his study is also a fixed point, although he never uses this fact. He exploits the exponential moments of this fixed point to obtain probability estimates.

There is a well known connection of Quicksort to tree sorting (*see* Frazer and McKellar, 1970, for more details).

We would like to point out that our results are completely independent of the given list we start with. In particular the results are true even if the list is already sorted. The randomness in our approach is given purely by the equiprobable way we pick a random element x out of the list (Hoare, 1962).

Assume now that we do not pick a random number, but have a particular selection rule for the partitioning element. Then X_n depends deterministically on the specific list. Our probability space Ω is in this case the set of all permutations of n given numbers. If every permutation has the same probability, then the recursive structure of X_n remains valid and our results apply.

One way to overcome the difficulty of uniformity assumptions is to shuffle the given n numbers first. Assuming a perfect shuffling, all permutations have the same probability. Another way out is to close the eyes and hope for the best. A good choice, adopted by Unix is to choose a number in the middle position. If the list is already in natural or reverse order, then X_n takes its minimum, the lower theoretical bound $n \log_2 n$ of the average. For lists with preordered parts this version proceeds faster on the average than the random Quicksort version.

In order to avoid any complication in the discussion of the selection rule we have prefered the randomized version of Quicksort as presented.

1. RECURSIVE EQUATION

Let X_n denote the random number of comparisons needed to sort a list of length n by Quicksort. Then the distribution $L(X_n)$ of X_n satisfies the recursive relation $X_0=0$, $X_1=0$, $X_2=1$,

$$L(X_n)=L(X_{Z_n-1}+\bar{X}_{n-Z_n}+n-1), \qquad n \geqq 2. \tag{1.1}$$

Observe that $n-1$ comparisons are used to compare every element of the list with the randomly chosen one. Then we have to sort a list of length Z_n-1, the list of smaller numbers, and a list of length $n-Z_n$, the list of larger numbers. The sortings of the lists are independent. The distribution of Z_n is a uniform distribution on $\{1, \ldots, n\}$. Furthermore the random variables Z_n, X_i, $\bar{X}_i$, $i=0, \ldots, n-1$, are obviously independent.

In order to avoid ambiguity we assume all numbers of the list are different.

It is easy to calculate the expectation of X_n. By (1.1) we obtain

$$E(X_n)=n-1+\sum_{i=1}^{n} P(Z_n=i)(E(X_{i-1})+E(X_{n-i}))$$

and

$$\frac{E(X_n)}{n+1}=\frac{E(X_{n-1})}{n}+\frac{2(n-1)}{n(n+1)}=\ldots=2\sum_{h=1}^{n+1}\frac{1}{h}+\frac{2}{n+1}-4.$$

Therefore $E(X_n)$ is approximately

$$E(X_n)=2n\ln n+n(2\gamma-4)+2\ln n+2\gamma+1+O(n^{-1}\ln n)$$

with $\gamma=0.57721\ldots$ being Euler's constant (*see* Knuth, 1973). We shall consider the random variables $Y_n := (X_n-E(X_n))/n$. Then immediately from (1.1) we obtain $Y_0=0$, $Y_1=0$,

$$L(Y_n)=L\left(Y_{Z_n-1}\frac{Z_n-1}{n}+\bar{Y}_{n-Z_n}\frac{n-Z_n}{n}+C_n(Z_n)\right),\qquad n\geqq 2. \qquad (1.2)$$

For any fixed n the random variables Z_n, Y_i, $\bar{Y}_i$, $1\leqq i\leqq n$ are independent, Z_n is uniformly distributed on $\{1, \ldots, n\}$ and C_n is a function defined by

$$C_n(i)=\frac{n-1}{n}+\frac{1}{n}(E(X_{i-1})+E(X_{n-i})-E(X_n)). \qquad (1.3)$$

As n goes to infinity Z_n/n converges in distribution to some random variable τ, which is uniformly distributed on [0, 1]. Furthermore $C_n(nZ_n/n)$ converges to $C(\tau)$,

$$C(x):=2x\ln x+2(1-x)\ln(1-x)+1,\qquad x\in[0, 1] \qquad (1.4)$$

(*see* Proposition 3.2). If we assume for a moment that Y_n converges in distribution to some Y, we expect from (1.2)

$$L(Y)=L(Y\tau+\bar{Y}(1-\tau)+C(\tau)) \qquad (1.5)$$

with τ, Y, $\bar{Y}$ independent, $L(Y)=L(\bar{Y})$. In section 2 we show the existence of some Y satisfying (1.5) by a fixed point argument. In section 3 we show that Y_n converges in fact to the fixed point Y.

2. FIXED POINT ARGUMENT

Let D be the space of distribution functions F with finite second moment $\int x^2\,dF(x)<\infty$ and the first moment $\int x\,dF(x)$ equal to zero. We use on D the Wasserstein (Mallow) metric

$$d(F, G)=\inf\|X-Y\|_2$$

where $\|.\|_2$ denotes the L_2 norm (*see* Cambanis *et al.*, 1976 or Major, 1978).

The infinimum is over all random variables X with distribution F and all Y with distribution function G.

The infimum is attained for a uniformly distributed random variable τ on $[0, 1]$,

$$d(F, G)=\|F^{-1}(\tau)-G^{-1}(\tau)\|_2=\left(\int_0^1 |F^{-1}(x)-G^{-1}(x)|^2\,dx\right)^{1/2}.$$

Here $F^{-1}(x)=\inf\{b \mid F(b)\geqq x\}$ denotes the left-continuous inverse of F.

The space D with the metric d is a complete separable metric space, *i.e.* a Polish space. It may be helpful to notice that $F_n \in D$ converges in d-metric to $F \in D$ if and only if F_n converges weakly to F and $\int x^2\,dF_n(x) \to \int x^2\,dF(x)<\infty$ (Denker, Rösler, 1985).

Define a map $S: D \to D$ by

$$S(F)=L(\tau X+(1-\tau)\bar{X}+C(\tau))$$

with X, $\bar{X}$, τ independent, $L(X)=L(\bar{X})=F$, τ uniformly distributed on $[0, 1]$ and $C:[0, 1]\to\mathbb{R}$ as in (1.4). S is well defined. Notice $E(C(\tau))=0$.

THEOREM 2.1: *The map $S: D\to D$ is a contraction on (D, d) and has a unique fixed point. Every sequence F, $S(F)$, $S^2(F)$, ..., $F\in D$, converges exponentially fast in the d-metric to the fixed point of S.*

Proof: Let F, G be in D,

$$S(F)=L(\tau X+(1-\tau)\bar{X}+C(\tau)), \qquad S(G)=L(\tau Y+(1-\tau)\bar{Y}+C(\tau)),$$
$$L(X)=L(\bar{X})=F, \qquad L(Y)=L(\bar{Y})=G,$$

the random variables τ, X, $\bar{X}$ be independent, also τ, X, $\bar{Y}$ be independent and τ be uniformly distributed on $[0, 1]$.

Then

$$\begin{aligned}d^2(S(F), S(G)) &\leqq \|\tau X+(1-\tau)\bar{X}+C(\tau)-\tau Y-(1-\tau)\bar{Y}-C(\tau)\|_2^2\\ &=\|\tau(X-Y)+(1-\tau)(\bar{X}-\bar{Y})\|_2^2\\ &=E((X-Y)^2)E(\tau^2)+E((\bar{X}-\bar{Y})^2)E((1-\tau)^2)=\frac{2}{3}E((X-Y)^2).\end{aligned}$$

Taking the infinum over all possible (X, Y) we obtain

$$d(S(F), S(G))\leqq\sqrt{\frac{2}{3}}\,d(F, G).$$

The sequence $S^n(F)$ is a Cauchy sequence since, for $m \leqq n$,

$$d(S^m(F), S^n(F)) \leqq \sum_{j=m}^{n-1} d(S^j(F), S^{j+1}(F))$$
$$\leqq \sum_{j=m}^{n-1} \left(\frac{2}{3}\right)^{j/2} d(F, S(F)) \leqq 3\, d(F, S(F)) \left(\frac{2}{3}\right)^{m/2}.$$

The Cauchy sequence $S^n(F)$ converges exponentially fast to some limit. This must be a fixed point. The fixed point is unique, as the contraction is strict.

Q.E.D.

3. CONVERGENCE TO THE FIXED POINT

In this section we shall show that Y_n of section 1 converges to Y in the Wasserstein d-metric, *i.e.* $d(L(Y_n), Y) \to 0$ with $L(Y)$ being the fixed point of S.

Define the mapping $T : \bigcup_{n=1}^{\infty} D^n \to D$ by

$$T(G_1, \ldots, G_{n-1}) = L\left(Y_{Z_n - 1} \frac{Z_n - 1}{n} + \bar{Y}_{n - Z_n} \frac{n - Z_n}{n} + C_n(Z_n) \right). \quad (3.1)$$

The random variables Z_n, Y_i, $\bar{Y}_i$, $i=0, \ldots, n-1$, are independent, $L(Y_i) = L(\bar{Y}_i) = G_i$, $i=1, \ldots, n-1$, $Y_0 = \bar{Y}_0 = 0$, Z_n is uniformly distributed on $\{1, \ldots, n\}$, C_n is as in (1.3). T is well defined. Notice $E(C_n(Z_n)) = 0$.

For every G use successively T to obtain a sequence

$$G_1 = G, \qquad G_2 = T(G_1), \qquad G_3 = T(G_1, G_2), \ \ldots$$

THEOREM 3.1: *Let $G \in D$ correspond to the point measure on* 0, *i.e.* $G(x) = \begin{cases} 1 : x \geqq 0 \\ 0 : x < 0 \end{cases}$. *Then G_n converges in the Wasserstein d-metric to the unique fixed point of S.*

We shall established two propositions before we give the proof. For $x \in \mathbb{R}$ let $\lceil x \rceil$ be the smallest integer larger equal to x.

PROPOSITION 3.2: *Let $C_n : \{1, \ldots, n\} \to \mathbb{R}$ and $C : [0, 1] \to \mathbb{R}$ be as in section* 1. *Then*

$$\sup_{x \in (0,\, 1]} |C_n(\ulcorner nx \urcorner) - C(x)| \leqq \frac{6}{n} \ln n + O(n^{-1}).$$

Proof: If $1 \leqq \lceil nx \rceil \leqq n-1$ we estimate

$$|C_n(\ulcorner nx \urcorner) - C(x)| \leqq \left| C\left(\frac{\ulcorner nx \urcorner}{n}\right) - C(x) \right| + \frac{4}{n} \ln \frac{n - \ulcorner nx \urcorner}{n} + O(n^{-1})$$

$$\leqq \sup_{|y-z| < 1/n} |C(y) - C(z)| + \frac{4}{n} \ln n + O(n^{-1})$$

$$\leqq \frac{6}{n} \ln n + O(n^{-1}).$$

Notice $O(n^{-1})$ is w.l.o.g. independent of x.

By

$$C_n(n) = (n-1)/n + 1/n\,(E(X_{n-1}) - E(X_n)) = 1 + O(n^{-1}) + \frac{2}{n} \ln n$$

we obtain

$$\sup_{x \in (1-(1/n),\, 1)} |C_n(n) - C(x)| \leqq \frac{2}{n} \ln n + O(n^{-1}) + \left| 1 - C\left(1 - \frac{1}{n}\right) \right|$$

$$\leqq \frac{4}{n} \ln n + O(n^{-1}).$$

Q.E.D.

PROPOSITION 3.3: *Let a_n, b_n, $n \in \mathbb{N}$ be two sequences of real numbers satisfying*

$$0 \leqq b_n \to 0 \quad \text{as } n \text{ tend to } \infty$$

$$0 \leqq a_n \leqq \frac{2}{n} \sum_{i=1}^{n-1} \frac{i^2}{n^2} a_i + b_n.$$

Then a_n converges to 0 as n tends to ∞.

Proof: Establish

$$a_n \leqq \sup_{0 \leqq i \leqq n-1} a_i \cdot \frac{2}{3} + b_n.$$

From this equation we can conclude a_n is uniformly bounded. Define $a = \limsup a_n < \infty$. For a given $\varepsilon > 0$ there exists a n_0 such that for $n \geqq n_0$ $a_n \leqq a + \varepsilon$. Then

$$a_n \leqq \frac{2}{n} \sum_{i=1}^{n_0} \frac{i^2}{n^2} a_i + \frac{2}{n} \sum_{i=n_0+1}^{n-1} \left(\frac{i}{n}\right)^2 (a+\varepsilon) + b_n \leqq \frac{2}{3}(a+\varepsilon) + O(1).$$

Therefore

$$0 \leqq a = \limsup_n a_n \leqq \frac{2}{3}(a+\varepsilon).$$

This is true for all $\varepsilon > 0$.

Q.E.D.

Proof of Theorem 3.1. – Fix Y, $\bar{Y}$ independent, $L(Y) = L(\bar{Y}) = F$ the distribution of the fixed point, $n \in \mathbb{N}$. Choose a version of Y_i, $\bar{Y}_i$ independent for each $0 \leqq i \leqq n-1$ with

$$\operatorname{Var}(Y_i - Y) = d^2(L(Y_i), F)$$
$$\operatorname{Var}(\bar{Y}_i - \bar{Y}) = d^2(L(\bar{Y}_i), F), \qquad 0 \leqq i \leqq n-1.$$

Notice $E(Y_i) = 0 = E(Y)$.

Define

$$V, \bar{V}: \quad \Omega \times [0, 1] \to \mathbb{R}, \qquad V(., x) = V_x, \ \bar{V}(., x) = \bar{V}_x$$
$$V_x = \sum_{i=1}^{n} 1_{(i-1)/n < x \leqq i/n} Y_{i-1}$$
$$\bar{V}_x = \sum_{i=1}^{n} 1_{(i-1)/n < x \leqq i/n} \bar{Y}_{n-i}.$$

Then for τ independent of Y_i, $\bar{Y}_i$, $0 \leqq i \leqq n-1$, uniformly distributed on $[0, 1]$,

$$G = L\left(\frac{\ulcorner n\tau \urcorner}{n} V_\tau + \frac{n - \ulcorner n\tau \urcorner}{n} \bar{V}_\tau + C_n(\ulcorner n\tau \urcorner)\right).$$

$$d^2(G_n, F) \leqq E\left(\left(\frac{\ulcorner n\tau \urcorner - 1}{n} V_\tau - \tau Y + \frac{n - \ulcorner n\tau \urcorner}{n} \bar{V}_\tau - (1-\tau)\bar{Y} + C_n(\ulcorner n\tau \urcorner) - C(\tau)\right)^2\right)$$

$$\leq E\left(\left(\frac{\ulcorner n\tau\urcorner-1}{n}V_r-\tau Y\right)^2\right)+E\left(\left(\frac{n-\ulcorner n\tau\urcorner}{n}\bar{V}_\tau-(1-\tau)\bar{Y}\right)^2\right)$$
$$+E((C_n(\ulcorner n\tau\urcorner)-C(\tau))^2)$$
$$\leq E\left(\sum_{i=1}^{n}1_{(i-1)/n<\tau\leq i/n}\cdot\left(\frac{i-1}{n}Y_{i-1}-\tau Y\right)^2\right)$$
$$+E\left(\sum_{i=1}^{n}1_{i-1/n<\tau\leq i/n}\left(\frac{n-i}{n}\bar{Y}_{n-i}-(1-\tau)\bar{Y}\right)^2\right)+O(n^{-1}\ln^{-2}n)$$
$$\leq\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left(\left(\frac{i-1}{n}Y_{i-1}-\frac{i-1}{n}Y\right)^2\right)+\frac{1}{n}\sum_{i=1}^{n}E\left(\left(\frac{n-i}{n}\bar{Y}_{n-i}-\frac{n-i}{n}\bar{Y}\right)^2\right)$$
$$+O(n^{-1}\ln^{-2}n)$$
$$\leq\frac{1}{n}\sum_{i=1}^{n}\left(\frac{i-1}{n}\right)^2E((Y_{i-1}-Y)^2)+\frac{1}{n}\sum_{i=1}^{n}\left(\frac{n-i}{n}\right)^2E((\bar{Y}_{n-i}-\bar{Y})^2)$$
$$+O(n^{-1}\ln^{-2}n)\leq\frac{2}{n}\sum_{i=1}^{n}\left(\frac{i-1}{n}\right)^2d^2(L(Y_{i-1}),F)+O(n^{-1}\ln^{-2}n).$$

Put $a_i=d^2(L(Y_i),F)$ and apply the previous proposition.

Q.E.D.

4. LAPLACE TRANSFORMS

In this section we show that Y_n as in (1.2) has finite Laplace transform. Let G, $G_2=T(G)$, G_3, ... be the sequence as in (3.1). The distribution function G corresponds to the point measure on zero, $G(x)=\begin{cases}1: x\geq 0\\ 0: x<0\end{cases}$.

LEMMA 4.1: *For all $L>0$ there exists a K_L, such that for all $n\in\mathbb{N}$ and all $\lambda\in[-L,L]$*

$$Ee^{\lambda Y_n}\leq e^{\lambda^2K_L},\qquad L(Y_n)=G_n. \tag{4.1}$$

Proof: Let Z_n, $n\in\mathbb{N}$, be uniformly distributed on $\{1,\ldots,n\}$. Define

$$U_n=\left(\frac{Z_n-1}{n}\right)^2+\left(\frac{n-Z_n}{n}\right)^2-1\leq 0,$$

$$f(K,\lambda,n):=E\exp(\lambda C_n(Z_n)+\lambda^2KU_n),\qquad \lambda, K\in\mathbb{R}.$$

We show first that for all $L>0$ there exists a $K_L\in\mathbb{N}$, such that for all $n\in\mathbb{N}$ and all $\lambda\in(-L, L)$

$$f(K_L, \lambda, n)\leqq 1. \tag{4.2}$$

We show this in several steps. $f(K, \lambda, n)$ is well defined and at least twice continuously differentiable.

CLAIM 1. – *There exists a $K_1>0$ and $L_1>0$ such that for all $n\in\mathbb{N}$ and for all $\lambda\in[-L_1, L_1]$ we have $f(K, \lambda, n)\leqq 1$. Show*

$$\frac{d}{d\lambda}f(K, \lambda, n)\big|_{\lambda=0}=E(C_n(Z_n))=0$$

$$\frac{d^2}{d\lambda^2}f(K, \lambda, n)\big|_{\lambda=0}=E(C_n^2(Z_n)+2KU_n).$$

Choose a $K_1>0$ with $d^2/d\lambda^2 f(K_1, \lambda, n)|_{\lambda=0}<0$. Then $f(K_1, \lambda, n)$ has a strict maximum at $\lambda=0$ for every n. Therefore in some neighborhood of $\lambda=0$ the function is less then or equal to 1 for fixed n. We may choose the neighborhood of $\lambda=0$ uniformly in n, because $d^2/d\lambda^2 f(K, \lambda, n)$ is smaller than 0 uniformly in n for λ small.

CLAIM 2. – *For given $L>0$ there exists a $K_2>0$, such that for all $n\in\mathbb{N}$ we have $f(K_2, L, n)\leqq 1, f(K_2, -L, n)\leqq 1$. A simple calculation shows*

$$\lim_{K\to\infty}\sup_n f(K, \lambda, n)=0 \quad \textit{for all } \lambda\neq 0.$$

Further $f(K, \lambda, n)$ is monotone decreasing in K for fixed λ and n, as the first derivative

$$\frac{d}{d\lambda}f(K, \lambda, n)=E(\lambda^2 U_n \exp(\lambda C_n(Z_n)+\lambda^2 KU_n))\leqq 0$$

shows.

Using these two properties it is easy to verify claim 2.

Now we show (4.2). Let $L>0$ be given and define K_L as the maximum of K_1 as in claim 1 and $L^2/L_1^2 K_2$, K_2 as in claim 2.

If $|\lambda|\leqq L_1$ then $f(K_L, \lambda, n)\leqq 1$ is satisfied by claim 1 and the monotonicity in K. If $L_1\leqq|\lambda|\leqq L$ then estimate for $\alpha=L/|\lambda|\geqq 1$, $\lambda>0$ (analogously for

$\lambda<0$),

$$f(K_L,\lambda,n)\leqq(E(\exp\alpha(\lambda C_n(Z_n)+\lambda^2 K_L U_n)))^{1/\alpha}$$
$$\leqq\left(E\exp\left(LC_n(Z_n)+\frac{L_1^2 L^2}{L_1^2}K_2 U_n\right)\right)^{1/\alpha}\leqq 1 \text{ by claim 2.}$$

With the help of (4.2) the Lemma 4.1 follows easily by induction. The induction step is shown here,

$$E\exp\lambda Y_n=E\exp(\lambda(Y_{Z_{n-1}}(Z_{n-1}-1)/n$$
$$+\bar{Y}_{n-Z_n}(n-Z_n)/n+C_n(Z_n)))$$
$$\leqq\exp(\lambda^2 K_L).E\exp(\lambda^2 K_L U_n+\lambda C_n(Z_n))$$
$$=\exp(\lambda^2 K_L)f(K_L,\lambda,n)\leqq\exp(\lambda^2 K_L).$$

Q.E.D.

THEOREM 4.2: *Let Y_n be as in (1.2) and $L(Y)$ be the fixed point of S. Then for all $\lambda\in\mathbb{R}$*

$$E\exp(\lambda Y_n)\to E\exp(\lambda Y)<\infty \quad as\ n\to\infty.$$

Proof: We know Y_n converges in distribution to Y and $E\exp(\lambda Y_n)$ is uniformly bounded in n for fixed λ. For N large $P(|Y|=N)=0$ estimate

$$|E\exp(\lambda Y_n)-E\exp(\lambda Y)|$$
$$\leqq|E(\exp(\lambda Y_n)1_{|Y_n|<N})-E(\exp(\lambda Y)1_{|Y|<N})|$$
$$+E(\exp(\lambda Y_n)1_{|Y_n|\geqq N})+E(\exp(\lambda Y)1_{|Y|\geqq N}).$$

The first expression is small for n large and any fixed N. The third is small for N large. The second is small for N large by

$$E(\exp(\lambda Y_n)1_{|Y_n|\geqq N})\leqq\exp(-|\lambda|\mathrm{N})\mathrm{E}(\exp(|\lambda||Y_n|+|\lambda|N)1_{|Y_n|\geqq N})$$
$$\leqq\exp(-|\lambda|N)\sup_n E\exp(2|\lambda||Y_n|)$$
$$\leqq 2\exp(-|\lambda|N+4\lambda^2 K_L) \qquad \text{for} \quad 2\lambda\in[-L,+L].$$

Q.E.D.

COROLLARY 4.3: *We obtain for $\lambda>0$, $\varepsilon>0$*

$$P(|X_n-E(X_n)|\geqq\varepsilon E(X_n))\leqq E\exp(\lambda|Y_n|-\lambda\varepsilon E(X_n)/n)\leqq Const.(\lambda,\varepsilon)n^{-2\lambda\varepsilon}.$$

Proof: We shall use the Markov inequality

$$P(|U|>b)\leqq E|U|/b \qquad \text{for} \quad b>0.$$

Then

$$P(|X_n - E(X_n)| \geqq \varepsilon E(X_n)) = P(\exp\lambda |Y_n| \geqq \exp(\varepsilon\lambda E(X_n)/n))$$
$$\leqq E\exp(\lambda |Y_n| - \lambda\varepsilon E(X_n)/n) \approx E\exp(\lambda |Y|) n^{-2\lambda\varepsilon}.$$

Q.E.D.

The interpretation of this corollary is that Quicksort is reliable. For numerical estimates of the probability one has to calculate Const. (λ, ε) or the constant K_L of Lemma 4.1.

5. MISCELLANEA

In this section we consider higher moments of Y, the k-median variant of Quicksort and give a representation by an infinite sum.

The random variable Y has finite moments of any order p, $1 \leqq p < \infty$. This follows for example by a standard argument from the finiteness of the Laplace transform $E\exp(\lambda |Y|)$, for some $\lambda > 0$ (*see* section 4). It would also be possible to use refined methods exploiting the contractive behavior of S. (This will be published in a more general paper on these fixed point ideas.) The convergence of G, $S(G)$, $S^2(G)$, ... to $L(Y)$ is exponentially fast in the Wasserstein d_p-metric (replace $\|.\|_2$ by $\|.\|_p$) for any $1 \leqq p < \infty$. Also Y_n will converge to Y in the d_p-metric, $1 \leqq p < \infty$.

The higher moments satisfy, by the fixed point relation, the following recurrence relation, $n \geqq 2$,

$$\frac{n-1}{n+1} E(Y^n) = \sum_{j=0}^{n} \sum_{i} \binom{n}{i} E(Y^{i_1}) E(Y^{i_2}) E(\tau^{i_1}(1-\tau)^{i_2}(C(\tau))^{i_3}).$$

The summation is over all $i = (i_1, i_2, i_3)$

$$i_3 \in \{0, \ldots, n\}, \qquad i_1, i_2 \in \{0, 2, 3, \ldots, n-1\}, \qquad i_1 + i_2 + i_3 = j,$$
$$\binom{n}{i} = \binom{n}{i_1}\binom{n-i_1}{i_2}\binom{n-i_1-i_2}{i_3}.$$

The explicit calculation of the moments is tedious, but could be done for example with the help of a computer (Hennequin, 1989).

The function $\ln E\exp(\lambda Y)$ is an analytic function in λ. The coefficient K_i of λ^i in the powerseries $\sum_{i=1}^{n} (K_i/i!)\lambda^i$ is called the cumulant of order i.

Hennequin conjectured in this paper

$$K_i = (-1)^i (\alpha_i - 2^i (i-1)!\, \rho(i)), \qquad i \geqq 2,$$

α_i rational, $\rho(i)$ being the Riemann zeta function $\sum n^{-i}$. By the fixed point characterization we could obtain a recursive relation for the cumulants. However we were not able to prove or disprove Hennequin's conjecture.

The fixed point property of Y implies a representation by an infinite sum. Define $I=\varnothing \cup \bigcup_{n=1}^{\infty} \{0, 1\}^n$. Let τ_σ, $\sigma \in I$, be independent random variables with the uniform distribution on [0, 1]. Let $|\sigma|$ denote the length of σ, $\sigma|i$ the first i coordinates of σ, σ_i the i'th coordinate. Then

$$Y=C(\tau_\varnothing)+\sum_{i=1}^{\infty} \sum_{|\sigma|=i} \left(C(\tau_\sigma) \prod_{j=1}^{i} (\tau_{\sigma|(j-1)}(1-\sigma_j)+(1-\tau_{\sigma|(j-1)})\sigma_j)\right).$$

Another variant is the k-median Quicksort (Hoare, 1962). Here the selection rule is to choose randomly $2k+1$ elements and select the median of these (*see* Knuth, 1973, for a discussion). Our method gives the following results.

The recursive relation (1.1) remains true with Z_n having a different distribution described by the selection procedure. The expression Z_n/n converges weakly to some random variable τ. The distribution of τ is the same as of the $(k+1)$ largest random variable in the set of random variables $U_1, \ldots, U_{2k+1}$, which are independent and uniformly distributed on [0, 1]. The distribution of τ has the density

$$(k+1)\binom{2k+1}{k} x^{k+1}(1-x)^k.$$

The expectation $E(X_n)$ behaves asymptotically like

$$E(X_n)=cn\ln n+O(n)$$

with c such that $E(C_c(\tau))=0$,

$$C_c(x)=1+cx\ln x+c(1-x)\ln(1-x).$$

The random variables $(X_n-E(X_n))/n=Y_n$ converge to Y, $L(Y)$ the fixed point of the analogous mapping S,

$$L(Y)=L(Y\tau+\bar{Y}(1-\tau)+C_c(\tau))$$

Y, $\bar{Y}$, τ independent, $L(Y)=L(\bar{Y})$, τ, C_c as above.

The proof follows the contraction idea of this paper. There are quite a lot of details to be checked. This has been done by the author.

6. CONCLUSION

We showed the existence of a limiting distribution for the number of comparisons performed by Quicksort. This distribution is a fixed point of some map. This enables us to give a representation as an infinite sum. The moments satisfy some recursive relations. The exponential moments are finite. This gives good estimates of probabilities as for example the probability of a bad behavior of Quicksort. As a final conclusion, Quicksort is reliable.

REFERENCES

1. S. Cambanis, G. Simon and W. Stout, Inequalities for $Ek(X, Y)$ when the Marginals are Fixed, *Zeitschrift für Wahrscheinlichkeistheorie und verwandte Gebiete*, 1976, *36*, pp. 285-294.
2. M. Denker and U. Rösler, A Moment Estimate for Rank Statistics, *Journal of Statistical Planning and Interference*, 1985, *12*, pp. 269-284.
3. L. Devroye, Exponential Bounds for the Running Time of a Selection Algorithm, *Journal of Computer and System Sciences*, 1985, **29**, pp. 1-7.
4. N. Dunford and J. Schwartz, Linear Operators I, *John Wiley & Sons*, 1963.
5. W. D. Frazer and A. C. McKellar, Samplesort: A Sampling Approach to Minimal Storage Tree Sorting, *Journal of the Association for Computing Machinery*, 1970, *17*, pp. 496-507.
6. P. Hennequin, Combinatorial Analysis of Quicksort Algorithm, *Informatique théorique et Applications/Theoretical Informatics and Applications*, 1989, *23*, pp. 317-333.
7. C. A. R. Hoare, Algorithm 64: Quicksort, *Communications of the Association for Computing Machinery*, 1961, **4**, p. 321.
8. C. A. R. Hoare, Quicksort, *Computer Journal*, 1962, *5*, pp. 10-15.
9. D. E. Knuth, The art of computer programming, *3*, Sorting and searching, *M.A. Reading*, Addison-Wesley, 1973.
10. R. Loeser, Some Performance Tests of "Quicksort" and Descendants, *Communications of the Association for Computing Machinery*, 1974, *17*, pp. 143-152.
11. P. Major, On the invariance principle for sums of independent identically distributed random variables, *Journal of Multivariate Analysis*, 1978, *8*, pp. 487-517.
12. M. Régnier, A Limit Distribution for Quicksort, *Informatique théorique et Applications/Theoretical Informatics and Applications*, 1989, *23*, pp. 335-343.
13. R. Sedgewick, The analysis of Quicksort programs, *Acta Informatica*, 1977, *7*, pp. 327-355.
14. R. Sedgewick, Quicksort, Stanford Computer Science Report STAN-CS-75-492, *Ph. d. thesis*, 1975, Also published by Garland, Pub. Co., New York, 1980.
15. R. Sedgewick, Algorithms, Addison-Wesley, second edition, 1988.