Distributional Convergence for the Number of Symbol Comparisons Used by QuickSort (Extended Abstract)

James Allen Fill[1†]

[1] *Department of Applied Mathematics and Statistics, The Johns Hopkins University, 34th and Charles Streets, Baltimore, MD 21218-2682 USA*

Most previous studies of the sorting algorithm `QuickSort` have used the number of key comparisons as a measure of the cost of executing the algorithm. Here we suppose that the n independent and identically distributed (iid) keys are each represented as a sequence of symbols from a probabilistic source and that `QuickSort` operates on individual symbols, and we measure the execution cost as the number of symbol comparisons. Assuming only a mild "tameness" condition on the source, we show that there is a limiting distribution for the number of symbol comparisons after normalization: first centering by the mean and then dividing by n. Additionally, under a condition that grows more restrictive as p increases, we have convergence of moments of orders p and smaller. In particular, we have convergence in distribution and convergence of moments of every order whenever the source is memoryless, i.e., whenever each key is generated as an infinite string of iid symbols. This is somewhat surprising: Even for the classical model that each key is an iid string of unbiased ("fair") bits, the mean exhibits periodic fluctuations of order n.

Keywords: QuickSort, symbol comparisons, distributional convergence, probabilistic source, tameness, coupling

1 Introduction, review of related literature, and summary

1.1 Introduction

We consider the `QuickSort` algorithm of Hoare (1962) applied to n distinct random keys $X_1, \dots, X_n$, each represented as a word (i.e., infinite string of symbols such as bits) from some specified finite or countably infinite alphabet. We will consider various probabilistic mechanisms [called (*probabilistic*) *sources*] for generating the symbols within a key, but we will always assume that the keys themselves are iid (independent and identically distributed), and we will later place conditions on the source that rule out the generation of equal keys.

`QuickSort`$(X_1, \dots, X_n)$ chooses one of the n keys $X_1, \dots, X_n$ (called the "pivot") uniformly at random, compares each of the other keys to it, and then proceeds recursively to sort both the keys smaller than the pivot and those larger than it.

[†] Research supported by the Acheson J. Duncan Fund for the Advancement of Research in Statistics.

Key observation (coupling): Because of the assumption that the keys are iid, we may take the pivot to be the *first* key in the sequence, X_1. Thus if $X_1, X_2, \dots$ is an infinite sequence of keys and C_n is any measure of the cost of sorting n random keys using any cost function c (for example, the number of key comparisons or the number of symbol comparisons), then we can place all the random variables C_n on a common probability space by using $C_n = c(X_1, \dots, X_n)$. Notice that then C_n is nondecreasing in n. We will assume throughout that this natural coupling of the random variables C_n has been used. The coupling opens up the possibility of establishing stronger forms of convergence than convergence in distribution, such as almost sure convergence and convergence in L^p, for suitably normalized C_n.

Many authors [Knuth (1998), Régnier (1989), Rösler (1991), Knessl and Szpankowski (1999), Fill and Janson (2000b), Fill and Janson (2002), Neininger and Rüschendorf (2002), and others] have studied K_n, the (random) number of key comparisons performed by the algorithm. This is an appropriate measure of the cost of the algorithm if each comparison has the same cost. On the other hand, if keys are represented as words and comparisons are done by scanning the words from left to right, comparing the symbols of matching index one by one, then the cost of comparing two keys is determined by the number of symbols compared until a difference is found. We call this number the number of *symbol comparisons* for the key comparison, and let S_n denote the total number of symbol comparisons when n keys are sorted by `QuickSort`. Symbol-complexity analysis allows us to compare key-based algorithms such as `QuickSort` with digital algorithms such as those utilizing digital search trees.

The goal of the present work is to establish a limiting distribution for the normalized sequence of random variables $(S_n - \mathbf{E}\, S_n)/n$. Both exact and limiting distributions of S_n will depend on the source, unlike for K_n.

1.2 Review of closely related literature (QuickSort and QuickSelect)

Until now, study of asymptotics for `QuickSort`'s S_n has been limited mainly to the expected value $\mathbf{E}\, S_n$. Fill and Janson (2004) were the pioneers in that regard, obtaining, *inter alia*, exact and asymptotic expressions for $\mathbf{E}\, S_n$ [consult their Theorem 1.1, and note that the asymptotic expansion extends through terms of order n with a $O(\log n)$ remainder] when the keys are infinite binary strings and the bits within a key result from iid fair coin tosses. (We will refer to this model for key-generation as "the standard binary source". Equivalently, a key is generated by sampling uniformly from the unit interval, representing the result in binary notation, and dropping the leading "binary point".) They found that the expected number of bit comparisons required by `QuickSort` to sort n keys is asymptotically equivalent to $\frac{1}{\ln 2}\, n \ln^2 n$, whereas the lead-order term of the expected number of *key* comparisons is $2\, n \ln n$, smaller by a factor of order $\log n$. Now suppose that $N = (N(t) : 0 \le t < \infty)$ is a Poisson process with rate 1 and is independent of the generation of the keys, and let $S(t) := S_{N(t)}$. The authors also found for each fixed $1 \le p < \infty$ an upper bound independent of $t \ge 1$ on the L^p-norm of

$$Y(t) := \frac{S(t) - \mathbf{E}\, S(t)}{t} \tag{1.1}$$

[see their Remark 5.1(a)], leading them to speculate that $Y(t)$ might have a limiting distribution as $t \to \infty$. We will see that a limiting distribution does indeed exist, not only for the standard binary source but for a wide range of sources, as well.

Vallée et al. (2009) greatly extended the scope of Fill and Janson (2004) by establishing for much more general sources both an exact expression for $\mathbf{E}\, S_n$ [consult their Proposition 3 and display (8)] and an asymptotic expansion (see their Theorem 1) through terms of order n with a $o(n)$ remainder. For the broad class of sources $\mathcal{S}$ considered, the expected number of symbol comparisons is of lead order $\frac{1}{h(\mathcal{S})} n \ln^2 n$, where $h(\mathcal{S})$ is the entropy of the source (see their Figure 1 for a definition).

Building on work of Fill and Nakama (2009), who had in turn followed closely along the lines of Fill and Janson (2004), Vallée et al. (2009) also studied the expected number of symbol comparisons required by the algorithm `QuickSelect`(n, m). This algorithm [also known as `Find`(n, m)], a close cousin of `QuickSort` also devised by Hoare (1961), finds a key of specified rank m from a list of n keys. Vallée et al. (2009) considered the case where $m = \alpha n + o(n)$ for general $\alpha \in [0, 1]$ [note: we will sometimes refer to `QuickQuant`(n, α), rather than `QuickSelect`(n, m), in this case] and a broad class of sources $\mathcal{S}$. They found that the expected number of symbol comparisons asymptotically has lead term $\rho_{\mathcal{S}}(\alpha) n$, where $\rho_{\mathcal{S}}(\alpha)$ is described in their Figure 1. Unlike in the case of `QuickSort`, this is only a constant times larger than the expected number of key comparisons, which is well known to be asymptotically $\kappa(\alpha) n$ with

$$\kappa(\alpha) := 2[1 - \alpha \ln \alpha - (1 - \alpha) \ln(1 - \alpha)].$$

For either `QuickSelect` or `QuickSort`, a deeper probabilistic analysis of the numbers of key comparisons and symbol comparisons is obtained by treating entire distributions and not just expectations—in particular, by finding limiting distributions for suitable normalizations of these counts and, if possible, establishing corresponding convergence of moments. Consider `QuickQuant`(n, α) first. For both key comparisons and symbol comparisons a suitable normalization is to divide by n, with no need to center first. For a literature review on the number of key comparisons, we refer the reader to Section 2.2 of Fill and Nakama (2010); the number of symbol comparisons is discussed next.

Fill and Nakama (2010) [see also Nakama (2009)] were the first to establish a limiting distribution for the number of symbol comparisons for any algorithm for sorting or searching. They considered `QuickQuant`(n, α) for a broad class of sources and found a limiting distribution (depending on α, and of course also on the source) for the number $S_n(\alpha)$ of symbol comparisons (after division by n). It would take us a bit too far afield to describe the limiting random variable $S(\alpha)$, so we refer the reader to Section 3.1 [see (3.7)] of Fill and Nakama (2010) for an explicit description. In their paper they use the natural coupling discussed in Section 1.1 and prove, for each α, that $S_n(\alpha)/n$ converges to $S(\alpha)$ both (i) almost surely and, under ever stronger conditions on the source as p increases, (ii) in L^p. Either conclusion implies convergence in distribution, and (ii) implies convergence of moments of order $\le p$. The approach taken in Fill and Nakama (2010) is sufficiently general that the authors were able to unify treatment of key comparisons and symbol comparisons and to consider various other cost functions: see their Example 2.1.

Now we turn our attention back to `QuickSort`, the focus of this extended abstract. Let K_n (respectively, S_n) denote the random number of key (resp., symbol) comparisons required by `QuickSort` to sort a list of n keys. We first consider K_n, for which we know the following convergence in law, for some random variable T [where the immaterial choice of scaling by $n + 1$, rather than n, matches with Régnier (1989)]:

$$\frac{K_n - \mathbf{E}\, K_n}{n + 1} \xrightarrow{\mathcal{L}} T. \tag{1.2}$$

This was proved (i) by Régnier (1989), who used the natural coupling and martingale techniques to establish convergence both almost surely and in L^p for every finite p; and (ii) by Rösler (1991), who used the *contraction method* [see Rösler and Rüschendorf (2001) for a general discussion] to prove convergence in the so-called *minimal L^p metric* for every finite p [from which (1.2), with convergence of all moments, again follows]. An advantage of Rösler's approach was identification of the distribution of the limiting T as the unique distribution of a zero-mean random variable with finite variance satisfying the distributional fixed-point equation

$$T \stackrel{\mathcal{L}}{=} UT + (1-U)T^* + g(U), \tag{1.3}$$

with $g(u) := 1 + 2u \ln u + 2(1-u)\ln(1-u)$ and where, on the right, T, T^*, and U are independent random variables; T^* has the same distribution as T; and U is distributed uniformly over $(0,1)$. Later, Fill and Janson (2000a) showed that uniqueness of the zero-mean solution $\mathcal{L}(T)$ to (1.3) continues to hold without the assumption of finite variance, or indeed any other assumption.

1.3 Summary

This extended abstract establishes, for a broad class of sources, a limiting distribution for the number S_n of symbol comparisons for `QuickSort`. We tried without success to mimic the approach used by Fill and Nakama (2010) for `QuickQuant`. The approach used in this extended abstract, very broadly put, is to relate the count S_n of symbol comparisons to various counts of key comparisons and then rely (heavily) on the result of Régnier (1989). Like Fill and Janson (2004), we will find it much more convenient to work in continuous time than in discrete time. (We hope to "de-Poissonize" our result in the full-length paper.) In the continuous-time setting and notation established at (1.1) (but without limiting attention to the standard binary source), we will prove in this extended abstract, assuming that the source is suitably "tame" (in a sense to be made precise), that

$$Y(t) = \frac{S(t) - \mathbf{E}\, S(t)}{t} \xrightarrow{\mathcal{L}} Y \tag{1.4}$$

for some random variable Y. Following the lead of Régnier (1989) and Fill and Nakama (2010), we will use the natural coupling discussed in Section 1.1. Under a mild tameness condition that becomes more stringent as $p \in [2, \infty)$ increases we will in fact establish convergence in L^p. In particular, for any g-tamed source as defined in Remark 2.3(a)—for example, for any (nondegenerate) memoryless source—we have convergence in L^p for every finite p.

Outline of the paper. After carefully describing in Section 2.1 the probabilistic models used to govern the generation of keys, reviewing in Section 2.2 four known results about the number of key comparisons we will need in our analysis of symbol comparisons, and listing in Section 2.3 the other basic probability tools we will need, in Section 3 we state and prove our main results about convergence in distribution for the number of symbol comparisons.

2 Background and preliminaries

2.1 Probabilistic source models for the keys

In this subsection, extracted with only small modifications from Fill and Nakama (2010), we describe what is meant by a probabilistic source—our model for how the iid keys are generated—using the terminology and notation of Vallée et al. (2009).

Let Σ denote a totally ordered alphabet (i.e., set of symbols), assumed to be isomorphic either to $\{0, \ldots, r-1\}$ for some finite r or to the full set of nonnegative integers, in either case with the natural order; a *word* is then an element of Σ^∞, i.e., an infinite sequence (or "string") of symbols. We will follow the customary practice of denoting a word $w = (w_1, w_2, \ldots)$ more simply by $w_1 w_2 \cdots$.

We will use the word "prefix" in two closely related ways. First, the symbol strings belonging to Σ^k are called *prefixes* of length k, and so $\Sigma^* := \cup_{0 \le k < \infty} \Sigma^k$ denotes the set of all prefixes of any nonnegative finite length. Second, if $w = w_1 w_2 \cdots$ is a word, then we will call

$$w(k) := w_1 w_2 \cdots w_k \in \Sigma^k \tag{2.1}$$

its *prefix of length* k.

Lexicographic order is the linear order (to be denoted in the strict sense by $\prec$) on the set of words specified by declaring that $w \prec w'$ if (and only if) for some $0 \le k < \infty$ the prefixes of w and w' of length k are equal but $w_{k+1} < w'_{k+1}$. Then the symbol-comparisons cost of determining $w \prec w'$ for such words is just $k+1$, the number of symbol comparisons.

A *probabilistic source* is simply a stochastic process $W = W_1 W_2 \cdots$ with state space Σ (endowed with its total σ-field) or, equivalently, a random variable W taking values in Σ^∞ (with the product σ-field). According to Kolmogorov's consistency criterion [e.g., Theorem 3.3.6 of Chung (2001)], the distributions μ of such processes are in one-to-one correspondence with consistent specifications of finite-dimensional marginals, that is, of the probabilities

$$p_w := \mu(\{w_1 \cdots w_k\} \times \Sigma^\infty), \quad w = w_1 w_2 \cdots w_k \in \Sigma^*.$$

Here the *fundamental probability* p_w is the probability that a word drawn from μ has $w_1 \cdots w_k$ as its length-k prefix.

Because the analysis of `QuickSort` is significantly more complicated when its input keys are not all distinct, we will restrict attention to probabilistic sources with continuous distributions μ. Expressed equivalently in terms of fundamental probabilities, our continuity assumption is that for any word $w = w_1 w_2 \cdots \in \Sigma^\infty$ we have $p_{w(k)} \to 0$ as $k \to \infty$, recalling the prefix notation (2.1).

Example 2.1 We present a few classical examples of sources. For more examples, and for further discussion, see Section 3 of Vallée et al. (2009).

(a) In computer science jargon, a *memoryless source* is one with $W_1, W_2, \ldots$ iid. Then the fundamental probabilities p_w have the product form

$$p_w = p_{w_1} p_{w_2} \cdots p_{w_k}, \quad w = w_1 w_2 \cdots w_k \in \Sigma^*.$$

(b) A *Markov source* is one for which $W_1 W_2 \cdots$ is a Markov chain.

(c) An intermittent source (a particular model for long-range dependence) over the finite alphabet $\Sigma = \{0, \ldots, r-1\}$ is defined by specifying the conditional distributions $\mathcal{L}(W_j \mid W_1, \ldots, W_{j-1})$ $(j \ge 2)$ in a way that pays special attention to a particular symbol $\underline{\sigma}$. The source is said to be *intermittent of exponent* $\gamma > 0$ *with respect to* $\underline{\sigma}$ if $\mathcal{L}(W_j \mid W_1, \ldots, W_{j-1})$ depends only on the maximum value k such that the last k symbols in the prefix $W_1 \cdots W_{j-1}$ are all $\underline{\sigma}$ and (i) is the uniform distribution on Σ, if $k = 0$; and (ii) if $1 \le k \le j-1$, assigns mass $[k/(k+1)]^\gamma$ to $\underline{\sigma}$ and distributes the remaining mass uniformly over the remaining elements of Σ.

For our results, the quantity

$$\pi_k := \max\{p_w : w \in \Sigma^k\} \tag{2.2}$$

will play an important role, as it did in equation (7) of Vallée et al. (2009) in connection with the generalized Dirichlet series $\Pi(s) := \sum_{k \geq 0} \pi_k^{-s}$. In particular, it will be sufficient for our main result (Theorem 3.1) that

$$\Pi(-\tfrac{1}{2}) = \sum_{k \geq 0} \pi_k^{1/2} < \infty; \tag{2.3}$$

a sufficient condition for this, in turn, is of course that the source is Π-tamed with $\gamma > 2$ in the sense of the following definition:

Definition 2.2 Let $0 < \gamma < \infty$ *and* $0 < A < \infty$. *We say that the source is* Π-tamed (with parameters γ and A) *if the sequence* (π_k) *at* (2.2) *satisfies*

$$\pi_k \leq A(k+1)^{-\gamma} \text{ for every } k \geq 0.$$

Observe that a Π-tamed source is always continuous.

Remark 2.3 (a) Many common sources have geometric decrease in π_k (call these "g-tamed") and so for *any* γ are Π-tamed with parameters γ and A for suitably chosen $A \equiv A_\gamma$.

For example, a memoryless source satisfies $\pi_k = p_{\max}^k$, where

$$p_{\max} := \sup_{w \in \Sigma^1} p_w$$

satisfies $p_{\max} < 1$ except in the highly degenerate case of an essentially single-symbol alphabet. We also have $\pi_k \leq p_{\max}^k$ for any Markov source, where now $p_{\max}$ is the supremum of all one-step transition probabilities, and so such a source is g-tamed provided $p_{\max} < 1$. Expanding dynamical sources [cf. Clément et al. (2001)] are also g-tamed.

(b) For an intermittent source as in Example 2.1, for all large k the maximum probability π_k is attained by the prefix $\underline{\sigma}^k$ and equals

$$\pi_k = r^{-1} k^{-\gamma}.$$

Intermittent sources are therefore examples of Π-tamed sources for which π_k decays at a truly inverse-polynomial rate, not an exponential rate as in the case of g-tamed sources.

2.2 Known results for the numbers of key comparisons for `QuickSort`

In this subsection we review four known `QuickSort` key-comparisons results—the first two formulated in discrete time and the next two in continuous time—that will be useful in proving our main results (Theorems 3.1 and 3.4). The first gives exact and asymptotic formulas for the expected number of key comparisons in discrete time and is extremely basic and well known. [See, e.g., (2.1)–(2.2) in Fill and Janson (2004).]

Lemma 2.4 *Let* K_n *denote the number of key comparisons required to sort a list of* n *distinct keys. Then*

$$\mathbf{E}\, K_n = 2(n+1)H_n - 4n = 2n \ln n - (4 - 2\gamma)n + 2 \ln n + (2\gamma + 1) + O(1/n). \tag{2.4}$$

The second result—mentioned previously at (1.2)—is due to Régnier (1989), who also proved convergence in L^p for every finite p. Recall the *natural coupling* discussed in Section 1.1.

Lemma 2.5 (Régnier (1989)) *Under the natural coupling, there exists a random variable T satisfying*

$$\frac{K_n - \mathbf{E}\,K_n}{n+1} \to T \text{ almost surely.} \tag{2.5}$$

We now shift to continuous time by assuming that the successive keys are generated at the arrival times of a Poisson process with unit rate. The number of key comparisons through epoch t is then $K_{N(t)}$, which we will abbreviate as $K(t)$; while the sequence (K_n) is thereby naturally embedded in the continuous-time process, the random variables $K(n)$ and K_n are not to be confused. We will use such abbreviations throughout this extended abstract; for example, we will also write $S_{N(t)}$ as $S(t)$.

The third result we review is the continuous-time analogue of Lemma 2.4. Note the difference in constant terms and the much smaller error term in continuous time.

Lemma 2.6 (Fill and Janson (2004), Lemma 5.1; proved in Fill and Janson (2010))
In the continuous-time setting, the expected number of key comparisons is given by

$$\mathbf{E}\,K(t) = 2\int_0^t (t-y)(e^{-y} - 1 + y)y^{-2}\,dy.$$

Asymptotically, as $t \to \infty$ we have

$$\mathbf{E}\,K(t) = 2t\ln t - (4-2\gamma)t + 2\ln t + (2\gamma+2) + O(e^{-t}\,t^{-2}). \tag{2.6}$$

The fourth result gives bounds on the moments of $K(t)$. For real $p \in [1,\infty)$, we let $\|W\|_p := (\mathbf{E}\,|W|^p)^{1/p}$ denote L^p-norm.

Lemma 2.7 (Fill and Janson (2004), Lemma 5.2; proved in Fill and Janson (2010))
For every real $p \in [1,\infty)$, there exists a constant $c_p < \infty$ such that

$$\begin{aligned} &\|K(t) - \mathbf{E}\,K(t)\|_p \le c_p t && \textit{for } t \ge 1,\\ &\|K(t)\|_p \le c_p t^{2/p} && \textit{for } t \le 1. \end{aligned}$$

In the special case $p = 2$, it follows immediately from Lemma 2.7 that

$$\mathbf{Var}\,K(t) \le c_2^2\,t^2 \quad \text{for } 0 \le t < \infty. \tag{2.7}$$

2.3 Basic probability tools

The following elementary lemma is the basic tool we will use for L^p-convergence. For completeness and the reader's convenience, we supply a proof.

Lemma 2.8 *Let $Y_k(t)$ be random variables, all defined on a common probability space, for $k = 0, 1, 2, \ldots$ and $0 \le t \le \infty$. Fix $1 \le p_0 < \infty$ and suppose for some sequence (b_k) that*

(i) *for each k we have $Y_k(t) \to Y_k(\infty)$ almost surely as $t \to \infty$,*

(ii) *for each k we have $\|Y_k(t)\|_{p_0} \le b_k$ for all $0 \le t < \infty$, and*

(iii) $\sum_{k=0}^{\infty} b_k < \infty$.

Then

(a) *for each $0 \le t \le \infty$ the series $\sum_{k=0}^{\infty} Y_k(t)$ converges in L^{p_0} to some random variable $Y(t)$, and moreover*

(b) *$Y(t) \to Y(\infty)$ in L^p for every $p < p_0$.*

Proof: (a) First, hypothesis (ii) extends to $t = \infty$ by Fatou's lemma. From (ii) and (iii) it then follows for each $0 \le t \le \infty$ that the sequence of partial sums $\sum_{k=0}^{K} Y_k(t)$, $K = 0, 1, \dots$, is a Cauchy sequence in the Banach space L^{p_0} and so converges to some random variable $Y(t)$.

(b) Choose any $p < p_0$. We first claim for each k that $Y_k(t) \to Y_k(\infty)$ in L^p, i.e., $|Y_k(t) - Y_k(\infty)|^p \to 0$ in L^1 as $t \to \infty$. Indeed, from (ii) it follows using Exercise 4.5.8 of Chung (2001) that $|Y_k(t)|^p$ is uniformly integrable in t, as therefore is $|Y_k(t) - Y_k(\infty)|^p$. Our claim then follows from (i), since almost-sure convergence to 0 implies convergence in probability to 0, and that together with uniform integrability implies convergence in L^1 [e.g., Theorem 4.5.4 of Chung (2001)].

Using the triangle inequality for L^p-norm, the claim proved in the preceding paragraph, and the extended condition (ii), and bounding L^p-norm by L^{p_0}-norm, we find for any K that

$$\limsup_{t\to\infty} \|Y(t) - Y(\infty)\|_p \le \limsup_{t\to\infty} \sum_{k=K+1}^{\infty} \|Y_k(t) - Y_k(\infty)\|_p \le 2 \sum_{k=K+1}^{\infty} b_k.$$

Now let $K \to \infty$ to complete the proof. □

Later (Lemma 3.3) we will transfer Lemma 2.5 to continuous time. When we do so, the following result will prove useful. This *law of the iterated logarithm* (LIL) is well known, and for example can be found for general renewal processes as Theorem 12.13 in Kallenberg (1997).

Lemma 2.9 (LIL for a Poisson process) *Abbreviate the natural logarithm function as L. For a Poisson process N with unit rate,*

$$\mathbf{P}\left(\limsup_{t\to\infty} \frac{N(t) - t}{\sqrt{2tLLt}} = 1, \ \liminf_{t\to\infty} \frac{N(t) - t}{\sqrt{2tLLt}} = -1\right) = 1. \tag{2.8}$$

3 Main results

3.1 Convergence in distribution

For convergence in law the following theorem, which adopts the natural coupling discussed in Section 1.1 and utilizes the terminology and notation of Section 2.1 for probabilistic sources, is our main result.

Theorem 3.1 *Consider the continuous-time setting in which keys are generated from a probabilistic source at the arrival times of a Poisson process N with unit rate. Let $S(t) = S_{N(t)}$ denote the number of symbol comparisons required by* `QuickSort` *to sort the keys generated through epoch t, and let*

$$Y(t) := \frac{S(t) - \mathbf{E}\, S(t)}{t}, \qquad 0 < t < \infty. \tag{3.1}$$

Assume that

$$\sum_{k=0}^{\infty} \Big(\sum_{w \in \Sigma^k} p_w^2 \Big)^{1/2} < \infty. \tag{3.2}$$

Then there exists a random variable Y such that $Y(t) \to Y$ in L^2. In particular, $Y(t) \xrightarrow{\mathcal{L}} Y$ and (because $\mathbf{E}\, Y(t) \to \mathbf{E}\, Y$) we have $\mathbf{E}\, Y = 0$, and $\mathbf{Var}\, Y(t) \to \mathbf{Var}\, Y$.

Remark 3.2 (a) The expected number of symbol comparisons in comparing two independent keys generated by the given source is $\sum_{w \in \Sigma^*} p_w^2 = \sum_{k=0}^{\infty} \sum_{w \in \Sigma^k} p_w^2$. So (3.2) is certainly sufficient to imply that $\mathbf{E}\, S(t) < \infty$ for every t and that with probability one $S(t) < \infty$ for all t.

(b) Observe that $\sum_{w \in \Sigma^k} p_w = 1$ for each k, and so (2.3) (namely, $\sum_k \sqrt{\pi_k} < \infty$) is sufficient for (3.2). Thus from the discussion in Section 2.1 we see that all Π-tamed sources with parameter $\gamma > 2$, including all (nondegenerate) memoryless sources, are covered by Theorem 3.1.

(c) The standard binary source is a classical example of a periodic memoryless source [cf. Vallée et al. (2009)—specifically, Definition 3(d), Theorem 1(ii), and the discussion (ii) in Section 3]. Equation (1.3) in Fill and Janson (2004) [proved as Proposition 5.4 in Fill and Janson (2010)] shows explicitly for the standard binary source that

$$\mathbf{E}\, S(t) = \tfrac{1}{\ln 2} t \ln^2 t - c_1 t \ln t + c_2 t + \pi_t t + O(\log t) \ \text{ as } t \to \infty,$$

where c_1, c_2 are explicitly given constants and π_t is a certain periodic function of $\log t$. Given the periodic term of order t in the mean for this periodic source, we find it surprising that Theorem 3.1 nevertheless applies.

To prepare for the proof of Theorem 3.1, we "Poissonize" Lemma 2.5.

Lemma 3.3 *In the continuous-time setting of Theorem 3.1, let $K(t) = K_{N(t)}$ denote the number of key comparisons required by* `QuickSort`*. Then for the same random variable T as in the discrete-time Lemma 2.5 we have*

$$\frac{K(t) - \mathbf{E}\, K(t)}{t} \to T \ \text{ almost surely as } t \to \infty.$$

Proof: This is routine. According to Lemmas 2.5 and 2.4,

$$\frac{K_n - [2n \ln n - (4 - 2\gamma) n]}{n+1} \to T \ \text{ almost surely as } n \to \infty.$$

Since $N(t) \to \infty$ almost surely as $t \to \infty$, it follows that

$$\frac{K(t) - [2N(t) \ln N(t) - (4 - 2\gamma) N(t)]}{N(t)+1} \to T \ \text{ almost surely as } t \to \infty.$$

Using the strong law of large numbers (SLLN) for N [namely, $N(t)/t \to 1$ almost surely, for which Lemma 2.9 is plenty sufficient], we deduce

$$\frac{K(t) - [2N(t) \ln N(t) - (4 - 2\gamma) t]}{t} \to T \ \text{ almost surely as } t \to \infty.$$

From the mean value theorem it follows that $|y \ln y - x \ln x| \le |y - x|(1 + \ln x + \ln y)$ for $x, y \ge 1$. Applying this with $x = t$ and $y = N(t)$ and invoking the SLLN and the LIL (Lemma 2.9), we find almost surely that for large t we have

$$\begin{aligned}|N(t)\ln N(t) - t\ln t| &\le |N(t) - t|\,[1 + \ln N(t) + \ln t] \le \sqrt{3t\ln\ln t}\,[2\ln t + 1 + o(1)]\\ &= O(\sqrt{t\ln\ln t} \times \ln t) = o(t),\end{aligned}$$

and so

$$\frac{K(t) - [2t\ln t - (4 - 2\gamma)t]}{t} \to T \text{ almost surely as } t \to \infty.$$

The desired result now follows from (2.6) in Lemma 2.6. □

We are now ready for the

Proof of Theorem 3.1: In this extended abstract we will prove only convergence in L^p for all $p < 2$; in the full paper we will use a strengthened version of Lemma 2.8 together with certain additional calculations to establish L^2-convergence.

We use an idea in Section 5 of Fill and Janson (2004) and decompose $S(t)$ as $\sum_{k=0}^{\infty} S_k(t)$, and each $S_k(t)$ further as $\sum_{w\in\Sigma^k} S_w(t)$, where for an integer k and a prefix $w \in \Sigma^k$ we define (with little possibility of notational confusion)

$$\begin{aligned}S_k(t) &:= \text{number of comparisons of } (k+1)\text{st symbols},\\ S_w(t) &:= \text{number of comparisons of } (k+1)\text{st symbols between keys with prefix } w.\end{aligned}$$

A major advantage of working in continuous time is that,

$$\text{for each fixed } k \text{ and } t, \text{ the variables } S_w(t) \text{ with } w \in \Sigma^k \text{ are independent.} \tag{3.3}$$

A further key observation, clear after a moment's thought, is this: For each $w \in \Sigma^*$, as stochastic processes,

$$(S_w(t) : t \in [0,\infty)) \text{ is a probabilistic replica of } (K(p_w t) : t \in [0,\infty)). \tag{3.4}$$

We define corresponding normalized variables as follows:

$$Y_k(t) := \frac{S_k(t) - \mathbf{E}\, S_k(t)}{t}, \qquad Y_w(t) := \frac{S_w(t) - \mathbf{E}\, S_w(t)}{t},$$

with the normalized variable $Y(t)$ corresponding to $S(t)$ defined at (3.1). Then

$$Y(t) = \sum_{k=0}^{\infty} Y_k(t), \qquad Y_k(t) = \sum_{w\in\Sigma^k} Y_w(t) \;\; (k = 0, 1, \dots).$$

To complete the proof we then need only find random variables $Y_k(\infty)$ such that hypotheses (i)–(iii) of Lemma 2.8 are satisfied for $p = 2$.

But, for each $w \in \Sigma^*$, the existence of an almost-sure limit, call it $Y_w(\infty)$, for $Y_w(t)$ as $t \to \infty$ follows from (3.4) and Lemma 3.3; indeed, we see that $Y_w(\infty)$ has the same distribution as $p_w T$, with T as in

Lemma 3.3. Taking the finite sum over $w \in \Sigma^k$, we see that $Y_k(\infty)$ can be defined as $\sum_{w\in\Sigma^k} Y_w(\infty)$ to meet hypothesis (i) of Lemma 2.8.

Finally, we verify hypothesis (ii) with $b_k := c_2\big(\sum_{w\in\Sigma^k} p_w^2\big)^{1/2}$ and c_2 as in Lemma 2.7, and then (3.2) gives hypothesis (iii). Here is the verification, using (3.3) at the second equality, (3.4) at the third equality, and the consequence (2.7) of Lemma 2.7 at the inequality:

$$\begin{aligned} t^2\|Y_k(t)\|_2^2 &= \|S_k(t) - \mathbf{E}\, S_k(t)\|_2^2 \\ &= \sum_{w\in\Sigma^k} \|S_w(t) - \mathbf{E}\, S_w(t)\|_2^2 \\ &= \sum_{w\in\Sigma^k} \|K(p_w t) - \mathbf{E}\, K(p_w t)\|_2^2 \\ &\le c_2^2\, t^2 \sum_{w\in\Sigma^k} p_w^2. \end{aligned}$$

This completes the proof of L^p-convergence for $p < 2$ in Theorem 3.1. □

3.2 Higher-order moments

Theorem 3.1 yields convergence of means and variances but not of higher-order moments. For those, the following theorem—to be proved in the full paper—may be used. We again adopt the natural coupling and consider the continuous-time setting, and we utilize the normalized-variable notation $Y(t)$ of (3.1).

Theorem 3.4 *Let $p \in [2,\infty)$. In the setting of Theorem 3.1, assume the strengthening*

$$\sum_{k=0}^{\infty}\Big(\sum_{w\in\Sigma^k} p_w^2\Big)^{1/p} < \infty \tag{3.5}$$

of (3.2)*. Then $Y(t) \to Y$ in L^p, with Y as in Theorem 3.1. In particular, $Y(t) \xrightarrow{\mathcal{L}} Y$, with convergence of moments of orders $\le p$.*

Remark 3.5 (a) Remark 3.2(b) extends to the observation that $\sum_k \pi_k^{1/p} < \infty$ is sufficient for (3.5)—for which, in turn, Π-tameness with parameter $\gamma > p$ is sufficient. In particular, for any g-tamed source, such as any (nondegenerate) memoryless source, we have $Y(t) \to Y$ in L^p for every $p < \infty$.

(b) We wonder (but have not yet considered): Under what conditions do we have $Y(t) \to Y$ almost surely?

3.3 Identification of the limit variable Y

In the full-length paper (or possibly elsewhere) we hope to expand carefully on the following ideas concerning explicit identification of the limit variable Y appearing in Theorems 3.1 and 3.4. We would like to gain enough understanding of Y, for example, that moments of any given order—or at least the variance—could be computed explicitly in terms of the fundamental probabilities p_w, $w \in \Sigma^*$, of the source.

Recall from the theorems and their proofs that the limiting variable Y satisfies $Y = \sum_{k=0}^{\infty} Y_k$, where $Y_k \equiv Y_k(\infty) = \sum_{w\in\Sigma^k} Y_w$ and $Y_w \equiv Y_w(\infty)$ is a probabilistic replica of $p_w T$. So it ought to be possible to identify Y by identifying explicitly the random variable T in the theorem of Régnier (1989) (recall our Lemma 2.5).

But T can indeed be identified. Since we have now reduced to counting key comparisons, there is no loss of generality in assuming that the iid keys are uniformly distributed over $(0, 1)$. Construct an (almost surely complete) infinite rooted binary search tree in the usual way by starting with an empty tree and inserting each key as it is generated. Label the nodes in the natural binary way: the root gets an empty label, its left (respectively, right) child is labeled 0 (resp., 1), the left child of node 0 is labeled 00, etc. Let U_θ denote the key inserted at node θ. Let L_θ (resp., R_θ) denote the largest key smaller than U_θ (resp., smallest key larger than U_θ) inserted at any ancestor of θ, with the exceptions $L_\theta := 0$ and $R_\theta := 1$ if the specified ancestor keys don't exist. Then one can prove that T is the limit as $\ell \to \infty$, both almost surely and in L^p for any finite p, of

$$T_\ell := \sum_{|\theta| \le \ell} (R_\theta - L_\theta) g(U_\theta)$$

where $g(u) = 1 + 2u \ln u + 2(1-u)\ln(1-u)$ and $|\theta|$ is the length of the label θ. This result can be viewed as a marriage and extension of the `QuickSort` limit theorems of Régnier (1989) and Rösler (1991), since the former established the existence of T (but not its form) and the latter argued (cf. his Section 5) that $T_\ell \xrightarrow{\mathcal{L}} T$.

Acknowledgements

We thank Svante Janson for excellent suggestions that led to improvements to both our main theorems.

References

K. L. Chung. *A Course in Probability Theory*. Academic Press, London, 3rd edition, 2001.

J. Clément, P. Flajolet, and B. Vallée. Dynamical sources in information theory: a general analysis of trie structures. *Algorithmica*, 29(1-2):307–369, 2001. Average-case analysis of algorithms (Princeton, NJ, 1998).

J. A. Fill and S. Janson. A characterization of the set of fixed points of the Quicksort transformation. *Electron. Comm. Probab.*, 5:77–84 (electronic), 2000a.

J. A. Fill and S. Janson. Smoothness and decay properties of the limiting Quicksort density function. In *Mathematics and computer science (Versailles, 2000)*, Trends Math., pages 53–64. Birkhäuser, Basel, 2000b.

J. A. Fill and S. Janson. Quicksort asymptotics. *J. Algorithms*, 44(1):4–28, 2002. Analysis of algorithms.

J. A. Fill and S. Janson. The number of bit comparisons used by Quicksort: an average-case analysis. In *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 300–307 (electronic), New York, 2004. ACM.

J. A. Fill and S. Janson. The number of bit comparisons used by Quicksort: an average-case analysis (2010.03.24 draft of full paper). URL `http://www.ams.jhu.edu/~fill/papers/BitsQuickfulldraft.pdf`. 2010.

J. A. Fill and T. Nakama. Analysis of the expected number of bit comparisons required by Quickselect. To appear in *Algorithmica*, 2009.

J. A. Fill and T. Nakama. Distributional convergence for the number of symbol comparisons used by `QuickSelect` (draft).
URL `http://www.ams.jhu.edu/~fill/papers/QSelectdistndraft.pdf`. 2010.

C. A. R. Hoare. Quicksort. *Comput. J.*, 5:10–15, 1962.

C. R. Hoare. Find (algorithm 65). *Communications of the ACM*, 4:321–322, 1961.

O. Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer–Verlag, New York, 1997.

C. Knessl and W. Szpankowski. Quicksort algorithm again revisited. *Discrete Mathematics and Theoretical Computer Science*, 3:43–64, 1999.

D. E. Knuth. *The Art of Computer Programming. Volume 3: Sorting and Searching*. Addison-Wesley, Reading, Massachusetts, 1998.

T. Nakama. *Analysis of Execution Costs for QuickSelect*. Ph.D. dissertation, The Johns Hopkins University, Department of Applied Mathematics and Statistics, Aug. 2009. URL `http://www.ams.jhu.edu/~fill/papers/NakamaDissertation.pdf`.

R. Neininger and L. Rüschendorf. Rates of convergence for Quickselect. *Journal of Algorithms*, 44:51–62, 2002.

M. Régnier. A limiting distribution of Quicksort. *RAIRO Informatique Théorique et Applications*, 23:335–343, 1989.

U. Rösler. A limit theorem for Quicksort. *RAIRO Informatique Théorique et Applications*, 25:85–100, 1991.

U. Rösler and L. Rüschendorf. The contraction method for recursive algorithms. *Algorithmica*, 29(1):3–33, 2001.

B. Vallée, J. Clément, J. A. Fill, and P. Flajolet. The number of symbol comparisons in Quicksort and Quickselect. In S. A. et al., editor, *36th International Colloquium on Automata, Languages and Programming (ICALP 2009), Part I, LNCS 5555*, pages 750–763. Springer–Verlag, 2009.