

# DISTRIBUTIONAL CONVERGENCE FOR THE NUMBER OF SYMBOL COMPARISONS USED BY QUICKSORT<sup>1</sup>

BY JAMES ALLEN FILL

*Johns Hopkins University*

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 (i.i.d.) 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, that is, whenever each key is generated as an infinite string of i.i.d. symbols. This is somewhat surprising; even for the classical model that each key is an i.i.d. string of unbiased (“fair”) bits, the mean exhibits periodic fluctuations of order  $n$ .

## 1. Introduction, review of related literature and summary.

1.1. *Introduction.* We consider Hoare’s [13] `QuickSort` algorithm applied to  $n$  distinct random items (called *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 i.i.d. (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.

---

Received February 2012; revised March 2012.

<sup>1</sup>Supported by the Acheson J. Duncan Fund for the Advancement of Research in Statistics.  
*MSC2010 subject classifications.* Primary 60F20; secondary 68W40.

*Key words and phrases.* `QuickSort`, limit distribution,  $L^p$ -convergence, symbol comparisons, natural coupling, probabilistic source, tameness, key comparisons, de-Poissonization.

*Key observation (coupling).* Because of the assumption that the keys are i.i.d., 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  $c$  (e.g., 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 then that  $C_n$  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 [16], Régnier [19], Rösler [21], Knessl and Szpankowski [15], Fill and Janson [4, 6], Neininger and Ruschendorf [18] 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 key 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 [7] 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 i.i.d. 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  $2n \ln n$ , smaller by a factor of order  $\log n$ . Now suppose that  $N = (N(t) : 0 \leq t < \infty)$  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

$1 \leq p < \infty$  an upper bound independent of  $t \geq 1$  on the  $L^p$ -norm of

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

(see [7], Remark 5.1(a), and the corresponding [8], Proposition 5.7), leading them to speculate that  $Y(t)$  might have a limiting distribution as  $t \rightarrow \infty$ . 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. [23] greatly extended the scope of [7] 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 [9], who had in turn followed closely along the lines of [7], Vallée et al. [23] also studied the expected number of symbol comparisons required by the algorithm `QuickSelect`( $n, m$ ). This algorithm [aka `Find`( $n, m$ )], a close cousin of `QuickSort` also devised by Hoare [12], finds a key of specified rank  $m$  from a list of  $n$  keys. The authors of [23] considered the case where  $m = \alpha n + o(n)$  for general  $\alpha \in [0, 1]$  [note: we will sometimes refer to `QuickQuant`( $n, \alpha$ ), 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`( $n, \alpha$ ) 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 [10], Section 2.2; the number of symbol comparisons is discussed next.

Fill and Nakama [10] (see also [17]) were the first to establish a limiting distribution for the number of symbol comparisons for any sorting or searching algorithm. They considered `QuickQuant`( $n, \alpha$ ) for a broad class of sources and found a limiting distribution (depending on  $\alpha$ , 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 [10], Section 3.1, see (3.7), for an explicit description. In their paper they use the natural coupling discussed in Section 1.1 and prove, for each  $\alpha$ , 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  $\leq p$ . The approach taken in [10] 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 paper. Let  $K_n$  (resp.,  $S_n$ ) 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 [19]):

$$(1.2) \quad \frac{K_n - \mathbf{E}K_n}{n + 1} \xrightarrow{\mathcal{L}} T.$$

This was proved (i) by Régnier [19], 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 [21], who used the *contraction method* (see Rösler and Rüschemdorf [22] 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

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

with  $g(u) := 1 + 2u \ln u + 2(1 - u) \ln(1 - u)$  and where, on the right-hand side,  $T$ ,  $T^*$  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 [5] 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 paper 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 in [10] for `QuickQuant`. The approach used in this paper, 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 [19]. Like Fill and Janson [7, 8], we will find it much more convenient to work mainly in continuous time than in discrete time, but we will also “de-Poissonize” our result. 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 paper, assuming that the source is suitably “tame” (in a sense to be made precise), that

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

for some random variable  $Y$ . Following the lead of [19] and [10], 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  $L^p$  (see our main Theorem 3.1 for a precise statement). In particular, for any  $g$ -tamed source as defined in Remark 2.3(a) [e.g., for any (nondegenerate) memoryless source] we have convergence in  $L^p$  for every finite  $p$ . Nondegeneracy of the distribution of  $Y$  is proved by Bindjeme and Fill [1]; thus the denominator  $t$  used in (1.4) is not too large to get an interesting limiting distribution.

*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 continuous-time result about convergence in distribution for the number of symbol comparisons. We extend the result by de-Poissonization to discrete time in Section 4.

## 2. Background and preliminaries.

2.1. *Probabilistic source models for the keys.* In this subsection, extracted with only small modifications from [10], we describe what is meant by a probabilistic source (our model for how the i.i.d. keys are generated) using the terminology and notation of Vallée et al. [23].

Let  $\Sigma$  denote a finite totally ordered alphabet (set of symbols), therefore isomorphic to  $\{0, \dots, r - 1\}$ , with the natural order, for some finite  $r$ ; a *word* is then an element of  $\Sigma^\infty$ , that is, an infinite sequence (or “string”) of symbols. We will follow the customary practice of denoting a word  $w = (w_1, w_2, \dots)$  more simply by  $w_1w_2 \dots$ .

We will use the word “prefix” in two closely related ways. First, the symbol strings belonging to  $\Sigma^k$  are called *prefixes* of length  $k$ , and so  $\Sigma^* := \bigcup_{0 \leq k < \infty} \Sigma^k$  denotes the set of all prefixes of any nonnegative finite length. Second, if  $w = w_1w_2 \dots$  is a word, then we will call

$$(2.1) \quad w(k) := w_1w_2 \dots w_k \in \Sigma^k$$

its *prefix of length  $k$* .

*Lexicographic order* is the linear order (to be denoted in the strict sense by  $<$ ) on the set of words specified by declaring that  $w < w'$  if (and only if) for some  $0 \leq k < \infty$  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  $w < w'$  for such words is just  $k + 1$ , the number of symbol comparisons.

A *probabilistic source* is simply a stochastic process  $W = W_1 W_2 \dots$  with state space  $\Sigma$  (endowed with its total  $\sigma$ -field) or, equivalently, a random variable  $W$  taking values in  $\Sigma^\infty$  (with the product  $\sigma$ -field). According to Kolmogorov’s consistency criterion (e.g., [2], Theorem 3.3.6), the distributions  $\mu$  of such processes are in one-to-one correspondence with consistent specifications of finite-dimensional marginals, that is, of the probabilities

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

Here the *fundamental probability*  $p_w$  is the probability that a word drawn from  $\mu$  has  $w_1 \dots w_k$  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  $\mu$ . Expressed equivalently in terms of fundamental probabilities, our continuity assumption is that for any  $w = w_1 w_2 \dots \in \Sigma^\infty$  we have  $p_{w(k)} \rightarrow 0$  as  $k \rightarrow \infty$ , 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 [23], Section 3.

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

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

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

(c) An *intermittent source* (a model for long-range dependence) over the finite alphabet  $\Sigma = \{0, \dots, r - 1\}$  is defined by specifying the conditional distributions  $\mathcal{L}(W_j | W_1, \dots, W_{j-1})$  ( $j \geq 2$ ) 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  $\mathcal{L}(W_j | W_1, \dots, W_{j-1})$  depends only on the maximum value  $k$  such that the last  $k$  symbols in the prefix  $W_1 \dots W_{j-1}$  are all  $\underline{\sigma}$  and (i) is the uniform distribution on  $\Sigma$ , if  $k = 0$ ; and (ii) if  $1 \leq k \leq j - 1$ , assigns mass  $[k/(k + 1)]^\gamma$  to  $\underline{\sigma}$  and distributes the remaining mass uniformly over the remaining elements of  $\Sigma$ .

For our results, the quantity

$$(2.2) \quad \pi_k := \max\{p_w : w \in \Sigma^k\}$$

will play an important role, as it did in [23], equation (7), in connection with the generalized Dirichlet series  $\Pi(s) := \sum_{k \geq 0} \pi_k^{-s}$ . In particular, it will be sufficient to obtain  $L^p$  convergence in our main result (Theorem 3.1) that

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

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

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

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

Observe that a  $\Pi$ -tamed source is always continuous.

**REMARK 2.3.** (a) Many common sources have geometric decrease in  $\pi_k$  (call these “g-tamed”) and so for any  $\gamma$  are  $\Pi$ -tamed with parameters  $\gamma$  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. [3]) are also g-tamed.

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

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

Intermittent sources are therefore examples of  $\Pi$ -tamed sources for which  $\pi_k$  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-comparison results (the first two formulated in discrete time and the next two in continuous time) that will be useful in proving our main Theorem 3.1. 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., [8], (2.1)–(2.2).]

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

$$\begin{aligned} \mathbf{E}K_n &= 2(n + 1)H_n - 4n \\ (2.4) \quad &= 2n \ln n - (4 - 2\gamma)n + 2 \ln n + (2\gamma + 1) + O(1/n). \end{aligned}$$

The second result [mentioned previously at (1.2)] is due to Régnier [19], who also proved convergence in  $L^p$  for every finite  $p$ . Recall the *natural coupling* discussed in Section 1.1.

LEMMA 2.5 [19]. *Under the natural coupling, there exists a random variable  $T$  satisfying*

$$(2.5) \quad \frac{K_n - \mathbf{E}K_n}{n + 1} \rightarrow T \quad \text{almost surely.}$$

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 paper; for example, we will also write  $S_{N(t)}$  as  $S(t)$ .

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

LEMMA 2.6 ([8], Lemma 5.1). *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  $t \rightarrow \infty$  we have

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

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 ([8], Lemma 5.3). *For every real  $p \in [1, \infty)$ , there exists a constant  $c_p < \infty$  such that*

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

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

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



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  $k = 0, 1, 2, \dots$  and  $0 \leq t \leq \infty$ . Fix  $t_0 \in [0, \infty)$  and  $1 \leq p < p' < \infty$  and suppose for some sequences  $(b_k)$  and  $(b'_k)$  that:*

- (i) *for each  $k$  we have  $Y_k(t) \rightarrow Y_k(\infty)$  almost surely as  $t \rightarrow \infty$ ,*
- (ii) *for each  $k$  we have  $\|Y_k(t)\|_p \leq b_k$  for all  $t_0 \leq t < \infty$ ,*
- (ii') *for each  $k$  we have  $\|Y_k(t)\|_{p'} \leq b'_k < \infty$  for all  $t_0 \leq t < \infty$  and*
- (iii)  $\sum_{k=0}^{\infty} b_k < \infty$ .

Then:

- (a) *for each  $t_0 \leq t \leq \infty$  the series  $\sum_{k=0}^{\infty} Y_k(t)$  converges in  $L^p$  to some random variable  $Y(t)$ , and moreover,*
- (b)  $Y(t) \rightarrow Y(\infty)$  in  $L^p$  as  $t \rightarrow \infty$ .

PROOF. We assume without loss of generality that  $t_0 = 0$ . Note that hypotheses (ii) and (ii') extend to  $t = \infty$  by Fatou's lemma.

(a) From (ii) and (iii) it follows for each  $0 \leq t \leq \infty$  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  $L^p$  and so converges to some random variable  $Y(t)$ .

(b) We first claim for each  $k$  that  $Y_k(t) \rightarrow Y_k(\infty)$  in  $L^p$ , that is,  $|Y_k(t) - Y_k(\infty)|^p \rightarrow 0$  in  $L^1$  as  $t \rightarrow \infty$ . To see this, from (ii') it follows using [2], Exercise 4.5.8, 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., [2], Theorem 4.5.4).

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

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

Now let  $K \rightarrow \infty$ , using (iii), to complete the proof.  $\square$

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 in [14], Theorem 12.13.

LEMMA 2.9 (LIL for a Poisson process). *For a Poisson process  $N$  with unit rate,*

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

**3. Main results (in continuous time): Convergence in  $L^p$  (and therefore in distribution).** 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 (for continuous time).

THEOREM 3.1. *Consider the continuous-time setting in which independent and identically distributed keys are generated from a probabilistic source at the arrival times of an independent 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*

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

Let  $p \in [2, \infty)$  and assume that

$$(3.2) \quad \sum_{k=0}^{\infty} \left( \sum_{w \in \Sigma^k} p_w^2 \right)^{1/p} < \infty.$$

Then there exists a random variable  $Y$  such that  $Y(t) \rightarrow Y$  in  $L^p$ . Thus  $Y(t) \xrightarrow{\mathcal{L}} Y$ , with convergence of moments of orders  $\leq p$ ; in particular,  $\mathbf{E}Y = 0$ .

REMARK 3.2. (a) Observe that  $\sum_{w \in \Sigma^k} p_w = 1$  for each  $k$ . Thus  $\sum_{w \in \Sigma^k} p_w^2 \leq 1$ , and condition (3.2) grows increasingly stronger as  $p$  increases.

(b) Under the weakest instance  $p = 2$  of the assumption (3.2) we have  $Y(t) \rightarrow Y$  in  $L^2$ , and so  $Y(t) \rightarrow Y$  in law with convergence of means and variances. The random variable  $Y$  in Theorem 3.1 of course does not (more precisely, can be taken not to) depend on the value of  $p$  considered (because a limit in  $L^p$  for any  $p$  is also a limit in probability, and limits in probability are almost surely unique).

(c) 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  $t$  [in fact, it follows from calculations to be performed in the proof of Theorem 3.1 for  $p = 2$  that  $\mathbf{E}S^2(t) < \infty$  for every  $t$ ] and that with probability one  $S(t) < \infty$  for all  $t$ .

(d) The sum on  $w$  in (3.2) is bounded above by the max-prefix probability  $\pi_k$  defined at (2.2), and so (2.3) (namely,  $\sum_k \pi_k^{1/p} < \infty$ ) is sufficient for (3.2). Thus from the discussion in Section 2.1 we see that Theorem 3.1 gives  $L^p$ -convergence for  $Y(t)$  for all  $\Pi$ -tamed sources with parameter  $\gamma > p$ . In particular, for any  $g$ -tamed source, such as any (nondegenerate) memoryless source, we have  $Y(t) \rightarrow Y$  in  $L^p$  for every  $p < \infty$ .

(e) The standard binary source is a classical example of a periodic memoryless source; cf. [23]—specifically, Definition 3(d), Theorem 1(ii) and the discussion (ii) in Section 3. Fill and Janson [8], Proposition 5.4, show explicitly for the standard binary source that

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

where  $c_1, c_2$  are explicitly given constants and  $\pi_t$  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.

(f) We wonder (but have not yet considered): Under what conditions do we have almost sure convergence in Theorem 3.1 (or in the discrete-time Theorem 4.1)?

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*

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

PROOF. This is routine. According to Lemmas 2.5 and 2.4,

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

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

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

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

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

From the mean value theorem it follows that  $|y \ln y - x \ln x| \leq |y - x|(1 + \ln x + \ln y)$  for  $x, y \geq 1$ . 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

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

and so

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

The desired result now follows from (2.6) in Lemma 2.6.  $\square$

We are now ready for the proof.

**PROOF OF THEOREM 3.1.** We use an idea of Fill and Janson [8], Section 5, 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)

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

A major advantage of working in continuous time is that,

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

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

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

We define corresponding normalized variables as follows:

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

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

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

To complete the proof of  $L^p$ -convergence of  $Y(t)$  we then need only to find random variables  $Y_k(\infty)$  such that the hypotheses of Lemma 2.8 are satisfied for some  $p' \in (p, \infty)$ . [Once we have the main conclusion of the theorem that  $Y(t)$  converges to  $Y$  in  $L^p$ , convergence in law with convergence of moments of orders  $\leq p$  follows immediately; in particular, since  $\mathbf{E}Y(t) \equiv 0$  and  $\mathbf{E}Y(t) \rightarrow \mathbf{E}Y$ , we have  $\mathbf{E}Y = 0$ .]

But, for each  $w \in \Sigma^*$ , the existence of an almost-sure limit, call it  $Y_w(\infty)$ , for  $Y_w(t)$  as  $t \rightarrow \infty$  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.

To verify the remaining hypotheses we choose  $t_0 = 1$  and need to bound the  $L^q$ -norm of  $Y_k(t)$  for  $k$  a nonnegative integer,  $t \in [1, \infty)$  and  $q \in \{p, p'\}$ . According to Lemma 3.4 to follow, for any real  $q \in [2, \infty)$  there exists a constant  $c'_q$  such that

$$\|Y_k(t)\|_q \leq c'_q \left( \sum_{w \in \Sigma^k} p_w^2 \right)^{1/q}$$

for such  $k$  and  $t$ . Thus hypotheses (ii) and [for any  $p' \in (p, \infty)$ ] (ii') of Lemma 2.8 hold, and the assumption (3.2) implies that (iii) does as well.  $\square$

LEMMA 3.4. *Adopt the notation in the above proof of Theorem 3.1. Then for every real  $q \in [2, \infty)$ , there exists a constant  $c'_q < \infty$  such that*

$$\|Y_k(t)\|_q \leq c'_q \left( \sum_{w \in \Sigma^k} p_w^2 \right)^{1/q}$$

for every nonnegative integer  $k$  and every  $t \in [1, \infty)$ .

PROOF. Fix  $q \in [2, \infty)$ . The first step is to use (as did Fill and Janson [8], proof of Proposition 5.7) Rosenthal's inequality, relying on the fact [recall (3.3)] that  $S_k(t)$  is the independent sum of  $S_w(t)$  with  $w \in \Sigma^k$ . According to Rosenthal's inequality [20], Theorem 3 (see also, e.g., [11], Theorem 3.9.1), there exists a constant  $b_q$  (depending only on  $q$ ) such that

$$\begin{aligned} t^q \|Y_k(t)\|_q^q &= \|S_k(t) - \mathbf{E}S_k(t)\|_q^q \\ &\leq b_q \max \left\{ \sum_{w \in \Sigma^k} \|S_w(t) - \mathbf{E}S_w(t)\|_q^q, \left[ \sum_{w \in \Sigma^k} \|S_w(t) - \mathbf{E}S_w(t)\|_2^2 \right]^{q/2} \right\}. \end{aligned}$$

Utilizing (3.4) and Lemma 2.7 together with the assumptions  $t \geq 1$  and  $q \geq 2$  we therefore find

$$\begin{aligned} \|Y_k(t)\|_q^q &\leq b_q \max \left\{ \sum_{w \in A_k(t)} c_q^q p_w^q + \sum_{w \in B_k(t)} (2c_q)^q p_w^2, \left( \sum_{w \in \Sigma^k} c_2^2 p_w^2 \right)^{q/2} \right\} \\ &\leq b_q \max \left\{ (2c_q)^q \sum_{w \in \Sigma^k} p_w^2, c_2^q \left( \sum_{w \in \Sigma^k} p_w^2 \right)^{q/2} \right\} \\ &\leq (c'_q)^q \sum_{w \in \Sigma^k} p_w^2, \end{aligned}$$

where  $A_k(t)$  and  $B_k(t)$  are the intersections of those  $\Sigma^k$  with  $\{w : p_w t \geq 1\}$  and  $\{w : p_w t < 1\}$ , respectively, and

$$c'_q := b_q^{1/q} \max\{2c_q, c_2\}.$$

The proof is complete.  $\square$

**4. Discrete time.** In this final section we de-Poissonize Theorem 3.1 in order to obtain an analogous result in discrete time, for which we need to strengthen the hypothesis slightly.

**THEOREM 4.1.** *Let  $S_n$  denote the number of symbol comparisons required by QuickSort to sort the first  $n$  keys generated. Let  $p \in [2, \infty)$  and assume that for some  $p' > p$  we have*

$$(4.1) \quad \sum_{k=0}^{\infty} \left( \sum_{w \in \Sigma^k} p_w^2 \right)^{1/p'} < \infty.$$

*If  $Y$  is the continuous-time limit from Theorem 3.1, then*

$$(4.2) \quad \frac{S_n - \mathbf{E}S_n}{n} \xrightarrow{L^p} Y \quad \text{as } n \rightarrow \infty.$$

*In particular, we have convergence in distribution, with convergence of moments of orders  $\leq p$ .*

We will derive Theorem 4.1 from Theorem 3.1, and our proof will need the following moderate deviation estimate for  $N(t)$ .

**LEMMA 4.2.** *For any  $0 < \epsilon < 1/6$ , we have*

$$\mathbf{P}(|N(t) - t| \geq t^{(1/2)+\epsilon}) \sim \sqrt{\frac{2}{\pi}} t^{-\epsilon} \exp\left(-\frac{1}{2}t^{2\epsilon}\right) \quad \text{as } t \rightarrow \infty.$$

**PROOF.** It is well known that the normal approximation gives correct lead-order asymptotics for right-tail deviations from the mean starting from a point that is, as here,  $o(t^{2/3})$ . Thus if  $Z$  is distributed standard normal, then

$$\mathbf{P}(|N(t) - t| \geq t^{(1/2)+\epsilon}) \sim \mathbf{P}(|Z| \geq t^\epsilon) \sim \sqrt{\frac{2}{\pi}} t^{-\epsilon} \exp\left(-\frac{1}{2}t^{2\epsilon}\right)$$

as claimed.  $\square$

In the following proof, given a sequence of events  $(B_n)$ , we say that  $B_n$  occurs “wvlp” (for “with very low probability”) if  $\mathbf{P}(B_n)$  is at most an amount exponentially small in a power of  $n$ ; we say that  $B_n$  occurs “wvhp” (for “with very high probability”) if the complement  $B_n^c$  occurs wvlp.

**PROOF OF THEOREM 4.1.** To prove (4.2) from the integer-time consequence

$$\frac{S(n) - \mathbf{E}S(n)}{n} \xrightarrow{L^p} Y$$

of Theorem 3.1, it is, of course, sufficient to prove

$$(4.3) \quad \frac{S(n) - S_n}{n} \xrightarrow{L^p} 0$$

and

$$(4.4) \quad \frac{\mathbf{E}S(n) - \mathbf{E}S_n}{n} \rightarrow 0.$$

Further, since (4.4) follows immediately from (4.3), it is sufficient to prove (4.3).

To prove (4.3), the key is to recall that  $S(t) = S_{N(t)}$  where  $N$  is a unit-rate Poisson process independent of  $(S_0, S_1, \dots)$  and to note that  $S_n$  increases with  $n$ . Let  $0 < \varepsilon < 1/3$ . Applying Lemma 4.2 with  $(t, \epsilon)$ , there taken to be  $(n + n^{(1/2)+\varepsilon}, \varepsilon/2)$ , wvhp we have

$$(4.5) \quad N(n + n^{(1/2)+\varepsilon}) \geq (n + n^{(1/2)+\varepsilon}) - (n + n^{(1/2)+\varepsilon})^{1/2+1/2\varepsilon} \geq n,$$

where the second inequality holds for large enough  $n$ . Similarly, wvhp we have

$$(4.6) \quad N(n - n^{(1/2)+\varepsilon}) \leq (n - n^{(1/2)+\varepsilon}) - (n - n^{(1/2)+\varepsilon})^{1/2+1/2\varepsilon} \leq n.$$

Because  $S$  is increasing, it follows from (4.5)–(4.6) that

$$S(n - n^{(1/2)+\varepsilon}) \leq S_n \leq S(n + n^{(1/2)+\varepsilon}) \quad \text{wvhp,}$$

and hence, wvhp

$$\begin{aligned} |S(n) - S_n| &\leq \max\{S(n) - S(n - n^{(1/2)+\varepsilon}), S(n + n^{(1/2)+\varepsilon}) - S(n)\} \\ &\leq [S(n) - S(n - n^{(1/2)+\varepsilon})] + [S(n + n^{(1/2)+\varepsilon}) - S(n)] \\ &= S(n + n^{(1/2)+\varepsilon}) - S(n - n^{(1/2)+\varepsilon}). \end{aligned}$$

So to complete the proof of Theorem 4.1 by proving (4.3), it is sufficient to show that

$$(4.7) \quad \frac{S(n + n^{(1/2)+\varepsilon}) - S(n - n^{(1/2)+\varepsilon})}{n} \xrightarrow{L^p} 0$$

and

$$(4.8) \quad \frac{S(n) - S_n}{n} \mathbf{1}(A_n) \xrightarrow{L^p} 0,$$

where  $A_n$  is any event wvlp and  $\mathbf{1}(A_n)$  is its indicator. We prove (a) (4.8) and then (b) (4.7).

(a) To bound the  $L^p$ -norm of the random variable on the left-hand side in (4.8), we use Hölder's inequality  $\|Z_1 Z_2\|_1 \leq \|Z_1\|_q \|Z_2\|_{q'}$  with

$$\begin{aligned} Z_1 &= \left| \frac{S(n) - S_n}{n} \right|^p, & Z_2 &= \mathbf{1}(A_n)^p = \mathbf{1}(A_n), \\ q &= \frac{p'}{p} > 1, & q' &= \frac{p'}{p' - p} > 1; \end{aligned}$$

note that  $(1/q) + (1/q') = 1$ , as required. Thus

$$\begin{aligned} \left\| \frac{S(n) - S_n}{n} \mathbf{1}(A_n) \right\|_p^p &= \mathbf{E} \left[ \left\| \frac{S(n) - S_n}{n} \mathbf{1}(A_n) \right\|_p^p \right] \\ &\leq \left\| \frac{S(n) - S_n}{n} \right\|_{p'}^p \times \mathbf{P}(A_n)^{1-(p/p')}. \end{aligned}$$

Because  $A_n$  occurs wvlp, it suffices to show that  $\|S(n)\|_{p'}$  and  $\|S_n\|_{p'}$  each grow at most polynomially in  $n$ .

The first of these two is very easy to handle. Using the hypothesis (4.1), we know from Theorem 3.1 that

$$\frac{S(t) - \mathbf{E}S(t)}{t} \xrightarrow{L^{p'}} Y,$$

and it follows that  $\|S(t) - \mathbf{E}S(t)\|_{p'}$  grows at most linearly in  $t$  as  $t \rightarrow \infty$ . But from the first sentence of Remark 3.2 we see that  $\mathbf{E}S(t)$  grows at most quadratically in  $t$ , so by the triangle inequality  $\|S(t)\|_{p'}$  grows at most quadratically in  $t$ .

Now we turn our attention to  $\|S_n\|_{p'}$ . Just as we observed in the preceding paragraph that  $\mathbf{E}S(t)$  grows at most quadratically in  $t$ , we observe here that

$$0 \leq S_n \leq \sum_{1 \leq i < j \leq n} C_{ij},$$

where  $C_{ij}$  is the cost of comparing the  $i$ th and  $j$ th keys, and hence (with  $C := C_{12}$ )

$$\|S_n\|_{p'} \leq \sum_{1 \leq i < j \leq n} \|C_{ij}\|_{p'} = \binom{n}{2} \|C\|_{p'}.$$

So, to conclude that  $\|S_n\|_{p'}$  grows at most quadratically in  $n$ , we need only show that  $\|C\|_{p'}$  is finite. Indeed, for any  $t < \infty$  we have

$$\begin{aligned} \infty > \mathbf{E}S(t)^{p'} &\geq \mathbf{E}[S(t)^{p'} \mathbf{1}(N(t) \geq 2)] \\ &\geq \mathbf{E}[C^{p'} \mathbf{1}(N(t) \geq 2)] = (\mathbf{E}C^{p'}) \mathbf{P}(N(t) \geq 2) \end{aligned}$$

and  $\mathbf{P}(N(t) \geq 2) > 0$ , so  $\mathbf{E}C^{p'} < \infty$ .

(b) It remains to establish (4.7). From two applications of Theorem 3.1 it follows quickly that

$$\begin{aligned} \frac{S(n + n^{(1/2)+\varepsilon}) - \mathbf{E}S(n + n^{(1/2)+\varepsilon})}{n} &\xrightarrow{L^p} Y \quad \text{and} \\ \frac{S(n - n^{(1/2)+\varepsilon}) - \mathbf{E}S(n - n^{(1/2)+\varepsilon})}{n} &\xrightarrow{L^p} Y; \end{aligned}$$

thus it suffices to prove

$$(4.9) \quad \frac{\mathbf{E}S(n + n^{(1/2)+\varepsilon}) - \mathbf{E}S(n - n^{(1/2)+\varepsilon})}{n} \rightarrow 0.$$



Recall from the proof of Theorem 3.1 that

$$ES(t) = \sum_{k=0}^{\infty} \sum_{w \in \Sigma^k} EK(p_w t)$$

and from Lemma 2.6 that we know an explicit formula for  $EK(t)$ , namely,

$$EK(t) = 2 \int_0^t (t - y)(e^{-y} - 1 + y)y^{-2} dy.$$

This function and its increasing derivative, call it  $d(t)$ , are both easily studied. In particular,  $d(t) \sim t$  as  $t \downarrow 0$  and  $d(t) \sim 2 \ln t$  as  $t \rightarrow \infty$ . Hence, for any  $0 < \delta \leq 1$  there exists a finite constant  $a_\delta$  such that

$$d(t) \leq a_\delta t^\delta \quad \text{for all } t \in (0, \infty).$$

Then, for any  $0 < t < u < \infty$ , we have

$$\begin{aligned} 0 < ES(u) - ES(t) &= \sum_{k=0}^{\infty} \sum_{w \in \Sigma^k} [EK(p_w u) - EK(p_w t)] \\ &\leq (u - t) \sum_{k=0}^{\infty} \sum_{w \in \Sigma^k} p_w d(p_w u) \leq a_\delta b_\delta (u - t) u^\delta \end{aligned}$$

with

$$b_\delta := \sum_{k=0}^{\infty} \sum_{w \in \Sigma^k} p_w^{1+\delta}.$$

Therefore,

$$ES(n + n^{(1/2)+\varepsilon}) - ES(n - n^{(1/2)+\varepsilon}) \leq 2a_\delta b_\delta n^{(1/2)+\varepsilon} (n + n^{(1/2)+\varepsilon})^\delta = o(n)$$

as desired for (4.9), provided  $\frac{1}{2} + \varepsilon + \delta < 1$  and  $b_\delta < \infty$ . Our proof thus far has been valid for any  $0 < \varepsilon < 1/3$ , but we now restrict it to  $0 < \varepsilon < 1/4$  and choose  $\delta = \frac{1}{2} - 2\varepsilon \in (0, \frac{1}{2})$ . The proof of Theorem 4.1 will be complete once we see that  $\varepsilon$  and  $\delta$  can be chosen so that  $b_\delta$  is finite.

Fix  $k$  and recall that  $\sum_{w \in \Sigma^k} p_w = 1$ . Let  $V$  be a random variable with probability mass function  $(p_w, w \in \Sigma^k)$ , and let  $Z := p_V^\delta$ . Then

$$\sum_{w \in \Sigma^k} p_w^{1+\delta} = EZ = \|Z\|_1 \leq \|Z\|_{1/\delta} = (EZ^{1/\delta})^\delta = (E p_V)^\delta = \left( \sum_{w \in \Sigma^k} p_w^2 \right)^\delta.$$

We can arrange for  $\delta \geq \frac{1}{p'}$  by choosing  $0 < \varepsilon \leq \frac{1}{4} - \frac{1}{2p'}$ , which is possible because  $p' > p \geq 2$ . Then

$$b_\delta \leq \sum_{k=0}^{\infty} \left( \sum_{w \in \Sigma^k} p_w^2 \right)^{1/p'} < \infty$$

by assumption (4.1).  $\square$

**Acknowledgment.** We thank Svante Janson for excellent suggestions that led to improvements to Theorem 3.1.

## REFERENCES

- [1] BINDJEME, P. and FILL, J. A. (2012). The limiting distribution for the number of symbol comparisons used by QuickSort is nondegenerate. Available at <http://www.ams.jhu.edu/~fill/>.
- [2] CHUNG, K. L. (2001). *A Course in Probability Theory*, 3rd ed. Academic Press, San Diego, CA. MR1796326
- [3] CLÉMENT, J., FLAJOLET, P. and VALLÉE, B. (2001). Dynamical sources in information theory: A general analysis of trie structures. *Algorithmica* **29** 307–369. MR1887308
- [4] FILL, J. A. and JANSON, S. (2000). Smoothness and decay properties of the limiting Quicksort density function. In *Mathematics and Computer Science (Versailles, 2000)* 53–64. Birkhäuser, Basel. MR1798287
- [5] FILL, J. A. and JANSON, S. (2000). A characterization of the set of fixed points of the Quicksort transformation. *Electron. Commun. Probab.* **5** 77–84 (electronic). MR1781841
- [6] FILL, J. A. and JANSON, S. (2002). Quicksort asymptotics. *J. Algorithms* **44** 4–28. MR1932675
- [7] FILL, J. A. and JANSON, S. (2004). The number of bit comparisons used by Quicksort: An average-case analysis. In *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms* 300–307 (electronic). ACM, New York. MR2291065
- [8] FILL, J. A. and JANSON, S. (2012). The number of bit comparisons used by Quicksort: An average-case analysis. Available at <http://www.ams.jhu.edu/~fill/>.
- [9] FILL, J. A. and NAKAMA, T. (2010). Analysis of the expected number of bit comparisons required by Quickselect. *Algorithmica* **58** 730–769. MR2672478
- [10] FILL, J. A. and NAKAMA, T. (2012). Distributional convergence for the number of symbol comparisons used by QuickSelect. Available at <http://www.ams.jhu.edu/~fill/>.
- [11] GUT, A. (2005). *Probability: A Graduate Course*. Springer, New York. MR2125120
- [12] HOARE, C. A. R. (1961). Find (algorithm 65). *Communications of the ACM* **4** 321–322.
- [13] HOARE, C. A. R. (1962). Quicksort. *Comput. J.* **5** 10–15. MR0142216
- [14] KALLENBERG, O. (1997). *Foundations of Modern Probability*. Springer, New York. MR1464694
- [15] KNESSL, C. and SZPANKOWSKI, W. (1999). Quicksort algorithm again revisited. *Discrete Math. Theor. Comput. Sci.* **3** 43–64 (electronic). MR1695194
- [16] KNUTH, D. E. (1998). *The Art of Computer Programming. Volume 3: Sorting and Searching*. Addison-Wesley, Reading, MA.
- [17] NAKAMA, T. (2009). Analysis of execution costs for QuickSelect. Ph.D. thesis, Dept. Applied Mathematics and Statistics, Johns Hopkins Univ. Available at <http://www.ams.jhu.edu/~fill/papers/NakamaDissertation.pdf>.
- [18] NEININGER, R. and RÜSCHENDORF, L. (2002). Rates of convergence for Quicksort. *J. Algorithms* **44** 51–62. MR1932677
- [19] RÉGNIER, M. (1989). A limiting distribution for quicksort. *RAIRO Inform. Théor. Appl.* **23** 335–343. MR1020478
- [20] ROSENTHAL, H. P. (1970). On the subspaces of  $L^p$  ( $p > 2$ ) spanned by sequences of independent random variables. *Israel J. Math.* **8** 273–303. MR0271721
- [21] RÖSLER, U. (1991). A limit theorem for “Quicksort”. *RAIRO Inform. Théor. Appl.* **25** 85–100. MR1104413
- [22] RÖSLER, U. and RÜSCHENDORF, L. (2001). The contraction method for recursive algorithms. *Algorithmica* **29** 3–33. MR1887296

- [23] VALLÉE, B., CLÉMENT, J., FILL, J. A. and FLAJOLET, P. (2009). The number of symbol comparisons in QuickSort and QuickSelect. In *Automata, Languages and Programming. Part I* (S. Albers, A. Marchetti-Spaccamela, Y. Matias, S. Nikolettseas and W. Thomas, eds.). *Lecture Notes in Computer Science* **5555** 750–763. Springer, Berlin. [MR2544890](#)

DEPARTMENT OF APPLIED MATHEMATICS AND STATISTICS  
JOHNS HOPKINS UNIVERSITY  
34TH AND CHARLES STREETS  
BALTIMORE, MARYLAND 21218-2682  
USA  
E-MAIL: [jimfill@jhu.edu](mailto:jimfill@jhu.edu)