AofA'12 DMTCS proc. **AQ**, 2012, 349–360

The Limiting Distribution for the Number of Symbol Comparisons Used by QuickSort is Nondegenerate (Extended Abstract)

Patrick Bindjeme[†] and James Allen Fill

Department of Applied Mathematics and Statistics, The Johns Hopkins University, Baltimore, MD 21218-2682 USA

In a continuous-time setting, Fill (2012) proved, for a large class of probabilistic sources, that the number of symbol comparisons used by `QuickSort`, when centered by subtracting the mean and scaled by dividing by time, has a limiting distribution, but proved little about that limiting random variable Y—not even that it is nondegenerate. We establish the nondegeneracy of Y. The proof is perhaps surprisingly difficult.

Keywords: QuickSort; limit distribution; L^p-convergence; symbol comparisons; probabilistic source; key comparisons; Chebyshev's other inequality

1 The number of symbol comparisons used by `QuickSort`: Brief review of a limiting-distribution result

In this section we briefly review the main theorem of Fill (2012). An infinite sequence of independent and identically distributed keys is generated; each key is a random word $(w_1, w_2, \ldots) = w_1 w_2 \cdots$, that is, an infinite sequence, or "string", of symbols w_i drawn from a totally ordered finite alphabet Σ. The common distribution μ of the keys (called a *probabilistic source*) is allowed to be any distribution over words, i.e., the distribution of any stochastic process with time parameter set $\{1, 2, \ldots\}$ and state space Σ . We know thanks to Kolmogorov's consistency criterion (e.g., Theorem 3.3.6 in Chung (2001)) that the possible distributions μ are in one-to-one correspondence with consistent specifications of finite-dimensional marginals, i.e., of the *fundamental probabilities*

$$p_w := \mu(\{w_1 w_2 \cdots w_k\} \times \Sigma^\infty) \text{ with } w = w_1 w_2 \cdots w_k \in \Sigma^*. \tag{1.1}$$

This p_w is the probability that a word drawn from μ has w as its length-k prefix.

For each n, the `QuickSort` algorithm of Hoare (1962) can be used to sort the first n keys to be generated. We may and do assume that the *first* key in the sequence is chosen as the pivot, and that the same is true recursively (in the sense, for example, that the pivot used to sort the keys smaller than the

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

original pivot is the *first* key to be generated that is smaller than the original pivot). A comparison of two keys is done by scanning the two words from left to right, comparing the symbols of matching index one by one until a difference is found. We let S_n denote the total number of symbol comparisons needed when n keys are sorted by `QuickSort`.

Theorem 1.1 (Fill (2012), 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*

$$Y(t) := \frac{S(t) - \mathbf{E}\, S(t)}{t}, \qquad 0 < t < \infty. \tag{1.2}$$

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

$$\sum_{k=0}^{\infty} \Big(\sum_{w \in \Sigma^k} p_w^2 \Big)^{1/p} < \infty. \tag{1.3}$$

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

We assume throughout this extended abstract that (1.3) holds with $p = 2$, which [as noted in Remark 3.2(b) of Fill (2012)] is the weakest instance of (1.3). From Theorem 1.1 we know that $\mathbf{Var}\, S(t) = O(t^2)$ as $t \to \infty$, but we don't know that $\mathbf{Var}\, S(t) = \Theta(t^2)$ because the theorem does not contain the important information that the limiting random variable Y is nondegenerate (i.e., does not almost surely vanish). **The purpose of the present extended abstract is to show that Y is nondegenerate**; this is stated as our main Theorem 2.1 below. The proof turns out to be surprisingly difficult; we do not know the value of $\mathbf{Var}\, Y$, and the proof of Theorem 2.1 does not provide it. The consequence $\mathbf{Var}\, S(t) = \Theta(t^2)$ of our Theorem 2.1 settles a question that has been open since the work of Fill and Janson (2004) even in the special case of the standard binary source with $\Sigma = \{0, 1\}$ and the fundamental probabilities of (1.1) equal to 2^{-k}.

2 Main results

The following is the main theorem of this extended abstract.

Theorem 2.1 *The limit distribution in Theorem 1.1 is nondegenerate.*

Throughout this extended abstract, we work in the setting of Theorem 1.1. Theorem 2.1 follows immediately from Propositions 2.3–2.4 in this section.

Definition 2.2 For an integer k and a prefix $w \in \Sigma^k$ we define (with little possibility of notational confusion), for comparisons among keys that have arrived by epoch t, the counts

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

The following two propositions combine to establish Theorem 2.1. We write $\Sigma^* := \cup_{0 \leq k < \infty} \Sigma^k$ for the set of all prefixes.

Proposition 2.3 *If the random variables $S_w(t)$, $w \in \Sigma^*$, are nonnegatively correlated for each fixed t, then the limit distribution in Theorem 1.1 is nondegenerate.*

A proof of Proposition 2.3 can be found in Section 3 (see Subsection 3.2).

Proposition 2.4 *For each fixed t, the random variables $S_w(t)$, $w \in \Sigma^*$, are nonnegatively correlated.*

A proof of Proposition 2.4 can be found in Section 4 (see Subsection 4.2).

3 Proof of Proposition 2.3

3.1 A lower bound for the variance of $K(t)$

Definition 3.1 If K_n is the number of key comparisons needed to sort the first n keys to arrive using `Quicksort`, and N is the Poisson process in Theorem 1.1 (independent of the generation of the keys), we define $K(t) := K_{N(t)}$.

In order to prove Proposition 2.3, we first establish the following lemma.

Lemma 3.2 *We have*

$$\mathbf{Var}\, K(t) \geq (1 + o(1))\sigma^2 t^2 \text{ as } t \to \infty$$

where $\sigma^2 := 7 - \frac{2}{3}\pi^2$.

Proof: By the law of total variance (namely, variance equals the sum of expectation of conditional variance and variance of conditional expectation) we have

$$\mathbf{Var}\, K(t) \geq \mathbf{E}\, \mathbf{Var}[K(t) \mid N(t)] = e^{-t} \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathbf{Var}\, K_n. \tag{3.1}$$

From (for example) (1.2) in Fill and Janson (2002) we have

$$\mathbf{Var}\, K_n = 7n^2 - 4(n+1)^2 H_n^{(2)} - 2(n+1)H_n + 13n$$

where $H_n^{(r)} := \sum_{i=1}^{n} i^{-r}$ is the nth harmonic number of order r and $H_n := H_n^{(1)}$ is the nth harmonic number, so

$$\lim_{n\to\infty} \frac{\mathbf{Var}\, K_n}{n^2} = \sigma^2.$$

It follows that, given $\alpha > 0$, there exists n_α such that

$$\mathbf{Var}\, K_n \geq (1 - \alpha)\sigma^2 n^2 \text{ for all } n \geq n_\alpha.$$

We therefore have from (3.1) that

$$\begin{aligned} \mathbf{Var}\, K(t) &\geq (1-\alpha)\sigma^2 e^{-t} \sum_{n=n_\alpha}^{\infty} \frac{t^n}{n!} n^2 \\ &= (1+o(1))(1-\alpha)\sigma^2 e^{-t} \sum_{n=0}^{\infty} \frac{t^n}{n!} n^2 \\ &= (1+o(1))(1-\alpha)\sigma^2 (t^2 + t) = (1+o(1))(1-\alpha)\sigma^2 t^2, \end{aligned}$$

where the asymptotics here are as $t \to \infty$. Since $\alpha > 0$ is arbitrary, the lemma follows. □

3.2 Proof of Proposition 2.3

Definition 3.3 For any nonnegative integer k, with $S_k(t)$ as in Definition 2.2 we define

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

Proof of Proposition 2.3: With $Y(t)$ as in Theorem 1.1, we have

$$Y(t) \quad = \quad \sum_{k=0}^{\infty} Y_k(t)$$

and, from Theorem 1.1,

$$\mathbf{Var}\, Y(t) \;\to\; \mathbf{Var}\, Y \;\text{ as } t \;\to\; \infty. \tag{3.2}$$

Knowing that

$$\mathbf{E}\, Y_k(t) = 0 \text{ for any nonnegative integer } k \text{ and } t \in (0, \infty),$$

that

$$\mathbf{E}\, Y(t) = 0 \text{ for } t \in (0, \infty),$$

and finally [as shown in the proof in Fill (2012) of the above Theorem 1.1] that the random variables $Y_k(t)$ satisfy the hypotheses of the elementary probabilistic Lemma 2.8 of Fill (2012) with $t_0 = 1$ and $p = 2$ for some $p' \in (2, \infty)$, we have from conclusion (a) of that lemma for any $t \in (0, \infty)$ that

$$\mathbf{Var}\, \left(\sum_{k=0}^{n} Y_k(t) \right) \to \mathbf{Var}\, Y(t) \text{ as } n \to \infty. \tag{3.3}$$

Now, from the fact that

$$S_k(t) = \sum_{w \in \Sigma^k} S_w(t) \text{ for any } k,$$

we have

$$\begin{aligned} \mathbf{Var}\, \left(\sum_{k=0}^{n} Y_k(t) \right) \quad &= \quad \sum_{k=0}^{n} \mathbf{Var}\, Y_k(t) + 2 \sum_{0 \le i < j \le n} \mathbf{Cov}(Y_i(t), Y_j(t)) \\ &= \quad \sum_{k=0}^{n} \mathbf{Var}\, Y_k(t) + \frac{2}{t^2} \sum_{0 \le i < j \le n} \mathbf{Cov}(S_i(t), S_j(t)) \\ &= \quad \sum_{k=0}^{n} \mathbf{Var}\, Y_k(t) + \frac{2}{t^2} \sum_{0 \le i < j \le n} \sum_{\substack{w \in \Sigma^i \\ w' \in \Sigma^j}} \mathbf{Cov}(S_w(t), S_{w'}(t)). \end{aligned}$$

This allows us to conclude that if for each fixed t the random variables $S_w(t)$ with $w \in \Sigma^*$ are nonnegatively correlated, then

$$\mathbf{Var}\, \left(\sum_{k=0}^{n} Y_k(t) \right) \ge \sum_{k=0}^{n} \mathbf{Var}\, Y_k(t)$$

and therefore, considering (3.3), that

$$\mathbf{Var}\, Y(t) \geq \sum_{k=0}^{\infty} \mathbf{Var}\, Y_k(t) \geq \mathbf{Var}\, Y_0(t). \tag{3.4}$$

As noted in (Fill, 2012, (3.3)–(3.4)), for any fixed t and any $k \geq 0$ we have that

$$\text{the random variables } S_w(t) \text{ with } w \in \Sigma^k \text{ are independent, and } S_w(t) \stackrel{\mathcal{L}}{=} K(p_w t), \tag{3.5}$$

where $K(t)$ is defined in Definition 3.1. It follows from (3.4) and Lemma 3.2 that

$$\mathbf{Var}\, Y(t) \geq t^{-2}\mathbf{Var}\, K(t) \geq (1 + o(1))\sigma^2,$$

which implies from (3.2) that

$$\mathbf{Var}\, Y \geq \sigma^2 > 0.$$

□

4 The random variables $S_w(t)$, $w \in \Sigma^*$, are nonnegatively correlated

In this section, we first prove the following (in Subsection 4.1) and then complete the proof of Proposition 2.4 in Subsection 4.2.

Proposition 4.1 *Let $w \in \Sigma^*$. Then the random variables $S_\emptyset(t)$ and $S_w(t)$ are nonnegatively correlated.*

4.1 The random variables $S_\emptyset(t)$ and $S_w(t)$ for any $w \in \Sigma^*$ are nonnegatively correlated

In this Subsection 4.1 we prove Proposition 4.1, which states that

$$\mathbf{Cov}(S_\emptyset(t), S_w(t)) \geq 0 \text{ for any } w \in \Sigma^*, \tag{4.1}$$

with the understanding that $S_\emptyset(t) = K(t) = K_{N(t)}$.

Proof of Proposition 4.1: We have

$$\mathbf{Cov}(S_\emptyset(t), S_w(t)) = \mathbf{Cov}(K(t), S_w(t)) = T_w(t) + V_w(t) \tag{4.2}$$

where

$$T_w(t) := \mathbf{Cov}(\mathbf{E}[K(t) \,|\, N(t)], \mathbf{E}[S_w(t) \,|\, N(t)]) \tag{4.3}$$

and

$$V_w(t) := \mathbf{E}\,\mathbf{Cov}(K(t), S_w(t) \,|\, N(t)). \tag{4.4}$$

But Propositions 4.2–4.3 will demonstrate that the expressions $T_w(t)$ and $V_w(t)$ are each nonnegative. □

4.1.1 Nonnegativity of $T_w(t)$

Here we prove the following result.

Proposition 4.2 *The expression $T_w(t)$ defined in* (4.3) *is nonnegative.*

Proof: We have

$$\mathbf{E}[K(t) \,|\, N(t) = n] = \kappa_n := \mathbf{E}\, K_n,$$

which is increasing with n; and

$$\mathbf{E}[S_w(t) \,|\, N(t) = n] = \sum_{j=0}^{n} \binom{n}{j} p_w^j (1-p_w)^{n-j} \kappa_j$$

is also increasing, following from the fact that the Binomial(n, p_w) distributions increase stochastically with n.

By "Chebyshev's other inequality" [Fink and Jodeit (1984)], we can conclude that

$$\mathbf{Cov}(\mathbf{E}[K(t) \,|\, N(t)], \mathbf{E}[S_w(t) \,|\, N(t)]) \geq 0,$$

which finishes the proof of the proposition. □

4.1.2 Nonnegativity of $V_w(t)$

In this subsection we prove the following proposition, thereby completing the proof of Proposition 4.1.

Proposition 4.3 *The expression $V_w(t)$ defined in* (4.4) *is nonnegative.*

This will be accomplished using the next two propositions, Propositions 4.5 and 4.8. But first, writing $\kappa_n := \mathbf{E}\, K_n$ for the expected number of key comparisons required to sort the first n keys to arrive, we need to record (a) the classical "divide and conquer" fact that

$$\kappa_d = \tfrac{1}{d} \sum_{j=1}^{d} (d - 1 + \kappa_{j-1} + \kappa_{d-j}) = \tfrac{1}{d} \sum_{j=e+1}^{e+d} (d - 1 + \kappa_{j-1-e} + \kappa_{e+d-j}) \tag{4.5}$$

for any two integers e and $d \geq 1$ and (b) the following lemma.

Lemma 4.4 *If*

$$\begin{aligned} \Sigma(n, a, b) \quad &:= \quad \sum_{j=a+1}^{a+b} (\kappa_{j-1} + \kappa_{n-j})(b - 1 + \kappa_{j-1-a} + \kappa_{b+a-j} - \kappa_b) \\ &= \quad \sum_{j=1}^{b} (\kappa_{j+a-1} + \kappa_{n-j-a})(b - 1 + \kappa_{j-1} + \kappa_{b-j} - \kappa_b) \end{aligned}$$

for any nonnegative integers a, b, and n with $a + b \leq n$, then

$$\Sigma(n, a, b) \geq 0.$$

Due to space limitations, the proof of of Lemma 4.4 is not included here, but will be included in the full-length paper.

Proposition 4.5 *If*

$$\begin{aligned}\psi(n,a,b) \\ := \quad n^{-1}\sum_{a<j\le a+b}(n-1+\kappa_{j-1}+\kappa_{n-j}-\kappa_n)(b-1+\kappa_{j-1-a}+\kappa_{b+a-j}-\kappa_b)\end{aligned} \tag{4.6}$$

for any nonnegative integers a, b, and n with $a+b\le n$, then

$$\psi(n,a,b)\ge 0.$$

Proof: We have

$$\begin{aligned}\psi(n,a,b) &= n^{-1}\sum_{a<j\le a+b}(\kappa_{j-1}+\kappa_{n-j})(b-1+\kappa_{j-1-a}+\kappa_{b+a-j}-\kappa_b)\\ &\quad + n^{-1}(n-1-\kappa_n)\sum_{a<j\le a+b}(b-1+\kappa_{j-1-a}+\kappa_{b+a-j}-\kappa_b)\\ &= n^{-1}\sum_{a<j\le a+b}(\kappa_{j-1}+\kappa_{n-j})(b-1+\kappa_{j-1-a}+\kappa_{b+a-j}-\kappa_b)\\ &= n^{-1}\Sigma(n,a,b)\ge 0,\end{aligned}$$

where the second equality follows from (4.5), and the inequality from Lemma 4.4. □

Definition 4.6 Let $w\in\sum^*$, and let n be any nonnegative integer. We define $S_{n,w}$ to be the number of key comparisons between those keys (from among the n first to arrive) with prefix w.

Definition 4.7 For any $w\in\sum^*$, and nonnegative integer n, we define $N_{n,w}$ to be the number of keys (from among the n first to arrive) with prefix w, and

$$N_{n,w^-} := \sum_{\substack{w'\in\Sigma^{|w|}:\\ w'<w}} N_{n,w'}.$$

Proposition 4.8 *For any nonnegative integers a, b, and n with $a+b\le n$, we have*

$$\mathbf{Cov}(K_n, S_{n,w}\,|\,N_{n,w}=b, N_{n,w^-}=a)\ge 0. \tag{4.7}$$

Proof: We will prove the proposition by strong induction on n. For that, we further condition on $J_n :=$ (the rank of the root key among the first n keys). Applying the law of total covariance (namely, covariance equals the sum of expectation of conditional covariance and covariance of conditional expectations) to the conditional covariance in question, we find

$$\begin{aligned}
&\mathbf{Cov}(K_n, S_{n,w} \mid N_{n,w} = b, N_{n,w^-} = a) \qquad (4.8)\\
= &\sum_{j=1}^{n} \mathbf{P}[J_n = j | N_{n,w} = b, N_{n,w^-} = a]\\
&\quad \times (\mathbf{E}[K_n | N_{n,w} = b, N_{n,w^-} = a, J_n = j] - \mathbf{E}[K_n | N_{n,w} = b, N_{n,w^-} = a])\\
&\quad \times (\mathbf{E}[S_{n,w} | N_{n,w} = b, N_{n,w^-} = a, J_n = j] - \mathbf{E}[S_{n,w} | N_{n,w} = b, N_{n,w^-} = a])\\
&+ \sum_{j=1}^{n} \mathbf{P}[J_n = j | N_{n,w} = b, N_{n,w^-} = a]\\
&\quad \times \mathbf{Cov}(K_n, S_{n,w} | N_{n,w} = b, N_{n,w^-} = a, J_n = j).
\end{aligned}$$

In preparation for handling (4.8), we begin with three observations, mainly concerning the first of the two terms on the right in (4.8).

(i) (K_n, J_n) and $(N_{n,w}, N_{n,w^-})$ are independent, so for any $j = 1, \dots, n$ and any nonnegative integers a and b, we have

$$\mathbf{P}[J_n = j \mid N_{n,w} = b, N_{n,w^-} = a] = \mathbf{P}[J_n = j] = \tfrac{1}{n},$$

and

$$\mathbf{E}[K_n \mid N_{n,w} = b, N_{n,w^-} = a, J_n = j] = \mathbf{E}[K_n \mid J_n = j] = n - 1 + \kappa_{j-1} + \kappa_{n-j},$$

and

$$\mathbf{E}[K_n \mid N_{n,w} = b, N_{n,w^-} = a] = \mathbf{E}\, K_n = \kappa_n.$$

Also

$$\mathbf{E}[S_{n,w} \mid N_{n,w} = b, N_{n,w^-} = a] = \kappa_b.$$

Keep in mind in the observations to follow that a is the value of N_{n,w^-}, that b is the value of $N_{n,w}$, and that j is the value of J_n.

(ii) If $a < j \le a + b$, which happens in the case that the root key has its prefix of length $|w|$ equal to w, then there are $j - 1 - a$ keys among the $j - 1$ that fall to the left of the pivot key that have w as their prefix of length $|w|$, and $b + a - j$ keys among the $n - j$ that fall to the right of the pivot key that have w as their prefix of length $|w|$. So

$$\mathcal{L}(S_{n,w} \mid N_{n,w} = b, N_{n,w^-} = a, J_n = j) = \mathcal{L}(b - 1 + D'_{j-1,j-1-a} + D''_{n-j,b+a-j})$$

where $D'_{j-1,j-1-a}$ and $D''_{n-j,b+a-j}$ are independent, and

$$\mathcal{L}(D'_{j-1,j-1-a}) = \mathcal{L}(S_{j-1,w} | N_{j-1,w} = j - 1 - a, N_{j-1,w^-} = a) = \mathcal{L}(K_{j-1-a}),$$

and similarly

$$\mathcal{L}(D''_{n-j,b+a-j}) = \mathcal{L}(K_{b+a-j});$$

hence

$$\mathbf{E}[S_{n,w} \mid N_{n,w} = b, N_{n,w^-} = a, J_n = j] = b - 1 + \kappa_{j-1-a} + \kappa_{b+a-j}.$$

(iii) If $j \le a$ or $a + b < j$, which happens if the root key has its prefix of length $|w|$ different from w, then all of the keys that have w as their prefix of length $|w|$ fall on the same side of the pivot key. So

$$\mathcal{L}(S_{n,w} \mid N_{n,w} = b, N_{n,w^-} = a, J_n = j) = \mathcal{L}(K_b)$$

and

$$\mathbf{E}[S_{n,w} \mid N_{n,w} = b, N_{n,w^-} = a, J_n = j] = \kappa_b.$$

Equation (4.8) now yields

$$\begin{aligned}
&\mathbf{Cov}(K_n, S_{n,w} \mid N_{n,w} = b, N_{n,w^-} = a)\\
&= \frac{1}{n}\Bigg\{ \sum_{a<j\le a+b} (n-1+\kappa_{j-1}+\kappa_{n-j}-\kappa_n)(b-1+\kappa_{j-1-a}+\kappa_{b+a-j}-\kappa_b)\\
&\qquad + \sum_{1\le j\le a} (n-1+\kappa_{j-1}+\kappa_{n-j}-\kappa_n)(\kappa_b-\kappa_b)\\
&\qquad + \sum_{a+b<j\le n} (n-1+\kappa_{j-1}+\kappa_{n-j}-\kappa_n)(\kappa_b-\kappa_b)\Bigg\}\\
&\qquad + \frac{1}{n}\sum_{j=1}^{n} \mathbf{Cov}(K_n, S_{n,w} \mid N_{n,w} = b, N_{n,w^-} = a, J_n = j)\\
&= \frac{1}{n} \sum_{a<j\le a+b} (n-1+\kappa_{j-1}+\kappa_{n-j}-\kappa_n)(b-1+\kappa_{j-1-a}+\kappa_{b+a-j}-\kappa_b)\\
&\qquad + \frac{1}{n}\sum_{j=1}^{n} \mathbf{Cov}(K_n, S_{n,w} \mid N_{n,w} = b, N_{n,w^-} = a, J_n = j)\\
&= \psi(n,a,b) + \frac{1}{n}\sum_{j=1}^{n} \mathbf{Cov}(K_n, S_{n,w} \mid N_{n,w} = b, N_{n,w^-} = a, J_n = j)\\
&\ge \frac{1}{n}\sum_{j=1}^{n} \mathbf{Cov}(K_n, S_{n,w} \mid N_{n,w} = b, N_{n,w^-} = a, J_n = j),
\end{aligned}$$

where the last equality follows from (4.6), and the inequality from Proposition 4.5. So, to prove that (4.7) holds, we only need to prove that

$$\mathbf{Cov}(K_n, S_{n,w} \mid N_{n,w} = b, N_{n,w^-} = a, J_n = j) \ge 0 \text{ for any } 1 \le j \le n. \tag{4.9}$$

First note that if $n = 1$, then $K_n \equiv 0$ and hence (4.9) holds.

Now let's assume that (4.7) holds for any natural number smaller than a given natural number n. Then:

Case A. If $a < j \le a + b$ then there are $j - 1 - a$ keys among the $j - 1$ that fall to the left of the pivot key that have their prefix of length $|w|$ equal to w, and $b + a - j$ keys among the $n - j$ that fall to the right of the pivot key that have their prefix of length $|w|$ equal to w. So

$$\begin{aligned}
&\mathcal{L}(K_n, S_{n,w}|N_{n,w} = b, N_{n,w^-} = a, J_n = j) \\
&= \mathcal{L}(n-1+K'_{j-1}+K''_{n-j}, b-1+D'_{j-1,j-1-a}+D''_{n-j,b+a-j})
\end{aligned}$$

where

$$\mathcal{L}(K'_{j-1}, D'_{j-1,j-1-a}) = \mathcal{L}(K_{j-1}, S_{j-1,w}|N_{j-1,w} = j-1-a, N_{j-1,w^-} = a)$$

and

$$\mathcal{L}(K''_{n-j}, D''_{n-j,b+a-j}) = \mathcal{L}(K_{n-j}, S_{n-j,w}|N_{n-j,w} = b+a-j, N_{n-j,w^-} = 0)$$

and also

$$(K'_{j-1}, D'_{j-1,j-1-a}) \text{ and } (K''_{n-j}, D''_{n-j,b+a-j}) \text{ are independent.}$$

In this case, therefore,

$$\begin{aligned}
&\mathbf{Cov}(K_n, S_{n,w} \,|\, N_{n,w} = b, N_{n,w^-} = a, J_n = j) \\
&= \mathbf{Cov}(n-1+K'_{j-1}+K''_{n-j}, b-1+D'_{j-1,j-1-a}+D''_{n-j,b+a-j}) \\
&= \mathbf{Cov}(K'_{j-1}, D'_{j-1,j-1-a}) + \mathbf{Cov}(K''_{n-j}, D''_{n-j,b+a-j}) \\
&= \mathbf{Cov}(K_{j-1}, S_{j-1,w} \,|\, N_{j-1,w} = j-1-a, N_{j-1,w^-} = a) \\
&\quad + \mathbf{Cov}(K_{n-j}, S_{n-j,w} \,|\, N_{n-j,w} = b+a-j, N_{n-j,w^-} = 0) \geq 0
\end{aligned}$$

by strong induction, since $j-1 < n$ and $n-j < n$.

CASE B. If $j \leq a$, which happens if the keys that have w as their prefix of length $|w|$ all fall to the right of the pivot key, then

$$\mathcal{L}(K_n, S_{n,w} \,|\, N_{n,w} = b, N_{n,w^-} = a, J_n = j) = \mathcal{L}(n-1+K'_{j-1}+K''_{n-j}, D''_{n-j,b})$$

where

$$\mathcal{L}(K'_{j-1}) = \mathcal{L}(K_{j-1})$$

and

$$\mathcal{L}(K''_{n-j}, D''_{n-j,b}) = \mathcal{L}(K_{n-j}, S_{n-j,w} \,|\, N_{n-j,w} = b, N_{n-j,w^-} = a-j)$$

and also

$$K'_{j-1} \text{ and } (K''_{n-j}, D''_{n-j,b}) \text{ are independent.}$$

In this case, therefore,

$$\begin{aligned}
&\mathbf{Cov}(K_n, S_{n,w} \,|\, N_{n,w} = b, N_{n,w^-} = a, J_n = j) \\
&= \mathbf{Cov}(n-1+K'_{j-1}+K''_{n-j}, D''_{n-j,b}) = \mathbf{Cov}(K''_{n-j}, D''_{n-j,b}) \\
&= \mathbf{Cov}(K_{n-j}, S_{n-j,w} \,|\, N_{n-j,w} = b, N_{n-j,w^-} = a-j) \geq 0
\end{aligned}$$

by strong induction, since $n-j < n$.

CASE C. If $a+b < j$, which happens if the keys that have w as their prefix of length $|w|$ all fall to the left of the pivot key, then

$$\mathcal{L}(K_n, S_{n,w}|N_{n,w} = b, N_{n,w^-} = a, J_n = j) = \mathcal{L}(n-1+K'_{j-1}+K''_{n-j}, D'_{j-1,b})$$

where

$$\mathcal{L}(K'_{j-1}, D'_{j-1,b}) = \mathcal{L}(K_{j-1}, S_{j-1,w} | N_{j-1,w} = b, N_{j-1,w^-} = a)$$

and

$$\mathcal{L}(K''_{n-j}) = \mathcal{L}(K_{n-j})$$

and also

$$(K'_{j-1}, D'_{j-1,b}) \text{ and } K''_{n-j} \text{ are independent.}$$

In this case, therefore,

$$\begin{aligned}
&\mathbf{Cov}(K_n, S_{n,w} \,|\, N_{n,w} = b, N_{n,w^-} = a, J_n = j) \\
&\quad = \mathbf{Cov}(n - 1 + K'_{j-1} + K''_{n-j}, D'_{j-1,b}) = \mathbf{Cov}(K'_{j-1}, D'_{j-1,b}) \\
&\quad = \mathbf{Cov}(K_{j-1}, S_{j-1,w} \,|\, N_{j-1,w} = b, N_{j-1,w^-} = a) \geq 0
\end{aligned}$$

by strong induction, since $j - 1 < n$.

In all three cases (4.9) holds, which concludes the proof of the proposition. □

Proof of Proposition 4.3: To prove Proposition 4.3, which asserts that

$$\mathbf{E}\,\mathbf{Cov}(K(t), S_w(t) \,|\, N(t)) \geq 0,$$

it's enough to show that

$$\mathbf{Cov}(K(t), S_w(t) \,|\, N(t) = n) \geq 0 \quad \text{for all } n = 0, 1, 2, \ldots.$$

But

$$\mathbf{Cov}(K(t), S_w(t) \,|\, N(t) = n) = \mathbf{Cov}(K_n, S_{n,w}),$$

and conditioning on $N_{n,w}$ and N_{n,w^-} we have

$$\begin{aligned}
\mathbf{Cov}(K_n, S_{n,w}) \quad &= \quad \mathbf{Cov}(\mathbf{E}[K_n \,|\, N_{n,w}, N_{n,w^-}], \mathbf{E}[S_{n,w} \,|\, N_{n,w}, N_{n,w^-}]) \\
&\qquad + \mathbf{E}\,\mathbf{Cov}(K_n, S_{n,w} \,|\, N_{n,w}, N_{n,w^-}).
\end{aligned}$$

Knowing that K_n and $(N_{n,w}, N_{n,w^-})$ are independent, we have

$$\mathbf{Cov}(\mathbf{E}[K_n \,|\, N_{n,w}, N_{n,w^-}], \mathbf{E}[S_{n,w} \,|\, N_{n,w}, N_{n,w^-}]) = \mathbf{Cov}(\kappa_n, \kappa_{N_{n,w}}) = 0.$$

We have now reduced to proving

$$\mathbf{E}\,\mathbf{Cov}(K_n, S_{n,w} \,|\, N_{n,w}, N_{n,w^-}) \geq 0,$$

which is achieved by Proposition 4.8. □

4.2 The general case

Proof of Proposition 2.4:

Let w and w' be in Σ^*. On the one hand, if the prefixes w and w' are inconsistent in the sense that no word has both w and w' as prefixes (for example, if $w = 01$ and $w' = 1$), then $S_w(t)$ and $S_{w'}(t)$ are independent and therefore uncorrelated. On the other hand, if w and w' are not inconsistent, then either w' is a prefix of w or w is a prefix of w' (or both, which is precisely the case $w = w'$). Let's assume without loss of generality that w' is a prefix of w; then $w = w'w''$, the concatenation of w' with another prefix w''. Having begun with a probabilistic source μ, consider the source μ' obtained by conditioning on prefix w', and use notation S' for symbol-count variables for source μ' just as S is used for source μ. [Observe that μ', like μ, satisfies the condition (1.3).] Then

$$\mathcal{L}(S_{w'}(t), S_w(t)) = \mathcal{L}(S'_{\emptyset}(p_{w'}t), S'_{w''}(p_{w'}t)).$$

The result follows from Proposition 4.1. □

Acknowledgements

We thank an anonymous referee for helpful suggestions.

References

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

J. A. Fill. Distributional convergence for the number of symbol comparisons used by QuickSort. *Annals of Applied Probability*, 2012. Accepted subject to revision; preprint available from `http://www.ams.jhu.edu/~fill/`.

J. A. Fill and S. Janson. Quicksort asymptotics. *J. Algorithms*, 44(1):4–28, 2002. ISSN 0196-6774. doi: 10.1016/S0196-6774(02)00216-X. URL `http://dx.doi.org/10.1016/S0196-6774(02)00216-X`. 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.

A. M. Fink and M. Jodeit, Jr. On Chebyshev's other inequality. In *Inequalities in statistics and probability (Lincoln, Neb., 1982)*, volume 5 of *IMS Lecture Notes Monogr. Ser.*, pages 115–120. Inst. Math. Statist., Hayward, CA, 1984. doi: 10.1214/lnms/1215465637. URL `http://dx.doi.org/10.1214/lnms/1215465637`.

C. A. R. Hoare. Quicksort. *Comput. J.*, 5:10–15, 1962. ISSN 0010-4620.