QUICKSORT ALGORITHM AGAIN REVISITED

March 3, 1997

Charles Knessl*
Dept. Mathematics, Statistics & Computer Science
University of Illinois at Chicago
Chicago, Illinois 60607-7045
U.S.A
knessl@uic.edu

Wojciech Szpankowski†
Department of Computer Science
Purdue University
W. Lafayette, IN 47907
U.S.A.
spa@cs.purdue.edu

Abstract

We consider the standard Quicksort algorithm that sorts n distinct keys with all possible $n!$ orderings of keys being equally likely. Equivalently, we analyze the total path length $\mathcal{L}_n$ in a randomly built *binary search tree*. Obtaining the limiting distribution of $\mathcal{L}_n$ is still an outstanding open problem. In this paper, we establish an integral equation for the probability density of the number of comparisons $\mathcal{L}_n$. Then, we investigate the large deviations of $\mathcal{L}_n$. We shall show that the left tail of the limiting distribution is much "thinner" (i.e., double exponential) than the right tail (which is only exponential). Our results contain some constants that must be determined numerically. We use asymptotic methods of applied mathematics such as the WKB method and matched asymptotics.

*The work of this author was supported by NSF Grant DMS-93-00136 and DOE Grant DE-FG02-93ER25168.
†The work of this author was supported by NSF Grants NCR-9206315 and NCR-9415491, and NATO Collaborative Grant CRG.950060.

1 Introduction

Hoare's *Quicksort* algorithm [9] is the most popular sorting algorithm due to its good performance in practise. The basic algorithm can be briefly described as follows [9, 11, 13]:

> A partitioning key is selected at random from the unsorted list of keys, and used to partition the keys into two sublists to which the same algorithm is called recursively until the sublists have size one or zero.

To justify the algorithm's good performance in practise, a body of theory was built. First of all, every undergraduate learns in a data structures course that the algorithm sorts "on average" n keys in $\Theta(n \log n)$ steps. To be more precise, one assumes that all $n!$ possible orderings of keys are equally likely. It is, however, also known that in the worst case the algorithm needs $O(n^2)$ steps (e.g., think of an input that is given in a decreasing order when the output is printed in an increasing order). Thus, one needs a more detailed probabilistic analysis to understand better the Quicksort behavior. In particular, one would like to know how likely (or rather unlikely) it is for such pathological behavior to occur. Our goal is to answer precisely this question.

A large body of literature is devoted to analyzing the Quicksort algorithm [3, 4, 5, 7, 11, 12, 13, 14, 15, 17]. However, many aspects of this problem are still largely unsolved. To review what is known and what is still unsolved, we introduce some notation. Let $\mathcal{L}_n$ denote the number of comparisons needed to sort a random list of length n. It is known that after selecting randomly a key, the two sublists are still "random" (cf. [11]). Clearly, the sorting time depends only on the keys' ranking, so we assume that the input consists of the first n integers $\{1, 2, \ldots, n\}$, and key k is chosen with probability $1/n$. Then, the following recurrence holds

$$\mathcal{L}_n = n - 1 + \mathcal{L}_k + \mathcal{L}_{n-1-k} \ .$$

Now, let $L_n(u) = \mathbf{E}u^{\mathcal{L}_n} = \sum_{k \geq 0} \mathsf{Pr}[\mathcal{L}_n = k]u^k$ be the probability generating function of $\mathcal{L}_n$. The above recurrence implies that

$$L_n(u) = \frac{u^{n-1}}{n} \sum_{i=0}^{n-1} L_i(u) L_{n-1-i}(u) \tag{1}$$

with $L_0(u) = 1$. Observe that the same recurrences are obtained when analyzing the total path length $\mathcal{L}_n$ of a binary search tree built over a random set of n keys (cf. [11, 13]). Finally, let us define a bivariate generating function $L(z, u) = \sum_{n \geq 0} L_n(u) z^n$. Then (1) leads to the following partial-differential functional equation

$$\frac{\partial L(z,u)}{\partial z} = L^2(zu, u) \ , \qquad \frac{\partial L(0,u)}{\partial z} = 1. \tag{2}$$

Observe also that $L(z,1) = (1-z)^{-1}$.

The moments of $\mathcal{L}_n$ are relatively easy to compute since they are related to derivatives of $L_n(u)$ at $u = 1$. Hennequin [7] analyzed these carefully and computed the first five cumulants. He also conjectured an asymptotic formula for the cumulants as $n \to \infty$.

The main open problem is to find the limiting distribution of $\mathcal{L}_n$. Régnier [14] proved that the limiting distribution of $(\mathcal{L}_n - \mathbf{E}[\mathcal{L}_n])/n$ exists, while Rösler [15, 16] characterized this limiting distribution as a fixed point of a contraction satisfying a recurrence equation. A partial-differential functional equation seemingly similar to (2) was studied recently by Jacquet and Szpankowski [10]. They analyzed a *digital* search tree for which the bivariate generating function $L(z,u)$ (in the so–called symmetric case) satisfies

$$\frac{\partial L(z,u)}{\partial z} = L^2(\frac{1}{2}zu, u)$$

with $L(z,0) = 1$. The above equation was solved asymptotically in [10], and this led to a limiting normal distribution of the path length $\mathcal{L}_n$ in digital search trees. While the above equation and (2) look similar, there are crucial differences. Among them, the most important is the contracting factor $\frac{1}{2}$ in the right-hand side of the above. Needless to say, we know that (2) does *not* lead to a normal distribution since the third moment is not equal to zero (cf. [13]).

In view of the above discussion, a less ambitious goal was set, namely that of computing the large deviations of $\mathcal{L}_n$, i.e., $\mathsf{Pr}[|\mathcal{L}_n - \mathbf{E}[\mathcal{L}_n]| \geq \varepsilon \mathbf{E}[\mathcal{L}_n]]$ for $\varepsilon > 0$. Hennequin [7] used Chebyshev's inequality to show that the above probability is $O(1/(\varepsilon \log^2 n))$. Recently, Rösler [15] showed that this probability is in fact $O(n^{-k})$ for any fixed k, and soon after McDiarmid and Hayward [12] used the powerful method of bounded differences to obtain an even better estimate, namely that the tail is asymptotically equal to $n^{-2\varepsilon \log\log n}$ (see the comment after Theorem 1 of Section 2).

In this paper, we improve on some of the previous results. First of all, we establish an integral equation for the probability density of $\mathcal{L}_n$, and using this we derive a left tail and a right tail of the large deviations of $\mathcal{L}_n$. We demonstrate that the left tail is much thinner (i.e., double exponential) than the right tail, which is roughly exponential. We establish these results using asymptotic methods of applied mathematics such as the WKB method and matched asymptotics.

The paper is organized as follows. In the next section we describe our main findings and compare them with other known results. In Section 3 we derive the integral equation for the asymptotic probability density of $\mathcal{L}_n$. In Section 4 we obtain our large deviations results.

2 Formulation and Summary of Results

As before, we let $\mathcal{L}_n$ be the number of key comparisons made when Quicksort sorts n keys. The probability generating function of $\mathcal{L}_n$ becomes

$$L_n(u) = \sum_{k=0}^{\infty} \Pr[\mathcal{L}_n = k]u^k = \mathbf{E}[u^{\mathcal{L}_n}]. \tag{3}$$

The upper limit in this sum may be truncated at $k = \binom{n}{2}$, since this is clearly an upper bound on the number of comparisons needed to sort n keys.

The generating function $L_n(u)$ satisfies (1) which we repeat below (cf. also [5, 13, 14, 15])

$$L_{n+1}(u) = \frac{u^n}{n+1} \sum_{i=0}^{n} L_i(u)L_{n-i}(u), \quad L_0(u) = 1. \tag{4}$$

Note that $L_n(1) = 1$ for all $n \geq 0$, and that the probability $\Pr[\mathcal{L}_n = k]$ may be recovered from the Cauchy integral

$$\Pr[\mathcal{L}_n = k] = \frac{1}{2\pi i} \int_C u^{-k-1} L_n(u) du. \tag{5}$$

Here C is any closed loop about the origin.

In Section 3, we analyze (4) asymptotically for $n \to \infty$ and for various ranges of u. We use asymptotic methods of applied mathematics, such as the WKB method and matched asymptotics [2, 6]. The most important scale is where $n \to \infty$ with $u - 1 = O(n^{-1})$, which corresponds to $k = \mathbf{E}[\mathcal{L}_n] + O(n) = 2n \log n + O(n)$. Most of the probability mass is concentrated in this range of k. As mentioned before, the existence of a limiting distribution of $(\mathcal{L}_n - \mathbf{E}[\mathcal{L}_n])/n$ as $n \to \infty$ was established in [14, 15], though there seems to be little known about this distribution (cf. [3, 5, 7, 12, 17]). Numerical and simulation results in [3, 5] show that the distribution is highly asymmetric and that the right tail seems much thicker than the left tail. It is also of interest to estimate these tails (cf. [12, 15]), as they give the probabilities that the number of key comparisons will deviate significantly from $\mathbf{E}[\mathcal{L}_n]$, which is well known to be asymptotically equal to $2n \log n$ as $n \to \infty$ (cf. [7, 13]).

For $u - 1 = w/n = O(n^{-1})$ and $n \to \infty$, we derive in Section 3 the asymptotic expansion

$$L_n(u) = \exp\left(A_n w/n\right) \left(G_0(w) + \frac{\log n}{n} G_1(w) + \frac{1}{n} G_2(w) + o(n^{-1}) \right) \tag{6}$$

where $A_n = \mathbf{E}[\mathcal{L}_n]$. The leading term $G_0(w)$ satisfies a non-linear integral equation. Indeed, in Section 3 we find that (cf. (46))

$$e^{-w} G_0(w) = \int_0^1 e^{2\phi(x)w} G_0(wx) G_0(w - wx) dx \tag{7}$$

$$G_0(0) = 1; \qquad G_0'(0) = 0 \tag{8}$$

where

$$\phi(x) = x \log x + (1-x)\log(1-x) \tag{9}$$

is the entropy of the Bernoulli(x) distribution. Furthermore, the correction terms $G_1(\cdot)$ and $G_2(\cdot)$ satisfy linear integral equations (cf. (47)–(48)).

By using (6) in (5) and asymptotically approximating the Cauchy integral we obtain

$$\Pr[\mathcal{L}_n - \mathbf{E}(\mathcal{L}_n) = ny] \sim \frac{1}{n} P(y) \tag{10}$$

where

$$\begin{aligned} P(y) &= \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{-yw} G_0(w)dw, & (11)\\ G_0(w) &= \int_{-\infty}^{\infty} e^{yw} P(y) dy & (12) \end{aligned}$$

and c is a constant. Hence, $G_0(w)$ is the moment generating function of the density $P(y)$.

Now, we can summarize our main findings:

Theorem 1 *Consider the Quicksort algorithm that sorts n randomly selected keys.*

(i) *The limiting density $P(y)$ satisfies*

$$P(y+1) = \int_0^1 \int_{-\infty}^{\infty} P\left(xt + \frac{y - 2\phi(x)}{2(1-x)}\right) P\left(-(1-x)t + \frac{y-2\phi(x)}{2x}\right) dt dx \tag{13}$$

and

$$\int_{-\infty}^{\infty} P(y) dy = 1, \quad \int_{-\infty}^{\infty} yP(y)dy = 0. \tag{14}$$

(ii) *The left tail of the distribution satisfies*

$$\Pr[\mathcal{L}_n - \mathbf{E}(\mathcal{L}_n) \le nz] \sim \frac{2}{\pi} \frac{1}{\sqrt{2\log 2 - 1}} \exp\left(-\alpha \exp\left(\frac{\beta - z}{2 - \log^{-1} 2}\right)\right) \tag{15}$$

for $n \to \infty$, $z \to -\infty$ where $\alpha = \frac{2\log 2 - 1}{e \log 2} = 0.205021\ldots$ and β is a constant.

(iii) *The right tail becomes*

$$\Pr[\mathcal{L}_n - \mathbf{E}(\mathcal{L}_n) \ge ny] \sim \frac{C_*}{2\sqrt{\pi}} \frac{1}{w_* \sqrt{w_* - 1}} e^{-w_*(w_* - \frac{1}{2} + 2\gamma + \log 2)} \exp\left(-yw_* + \int_1^{w_*} \frac{2e^u}{u} du\right) \tag{16}$$

for $n \to \infty$, $y \to +\infty$. Here C_* is a constant, γ is Euler's constant and $w_* = w_*(y)$ is the solution to

$$y = \frac{2}{w_*}\exp(w_*) \tag{17}$$

that asymptotically becomes

$$w_* = \log\left(\frac{y}{2}\right) + \log\log\left(\frac{y}{2}\right) + \frac{\log\log(y/2)}{\log(y/2)}(1+o(1)) \tag{18}$$

for $y \to \infty$ (cf. (84) in Section 4).

Finally, we relate our results for the tails to those of McDiarmid and Hayward [12]. These authors showed that

$$\Pr[|\mathcal{L}_n - \mathbf{E}(\mathcal{L}_n)| > \varepsilon\mathbf{E}(\mathcal{L}_n)] = \exp[-2\varepsilon\log n(\log\log n - \log(1/\varepsilon) + O(\log\log\log n))], \tag{19}$$

which holds for $n \to \infty$ and ε in the range

$$\frac{1}{\log n} < \varepsilon \le 1. \tag{20}$$

As pointed out in [12], this estimate is not very precise if, say, $\varepsilon = O(\log\log n/\log n)$. From Theorem 1 we conclude that (since the right tail is much thicker than the left tail)

$$\Pr[|\mathcal{L}_n - \mathbf{E}(\mathcal{L}_n)| \ge ny] \sim Ck(y)e^{\phi(y)}, \qquad y \to \infty \tag{21}$$

where C is a constant and

$$\begin{aligned} \phi(y) &= -yw_* + \int_1^{w_*} \frac{2e^u}{u} du, \qquad y = \frac{2e^{w_*}}{w_*}, \\ k(y) &= \frac{e^{-w_*^2} e^{w_*\left(\frac{1}{2} - 2\gamma - \log 2\right)}}{w_*\sqrt{w_* - 1}}. \end{aligned} \tag{22}$$

We have not been able to determine the upper limit on y for the validity of (21). However, it is easy to see that (21) reduces to (19) if we set $y = \varepsilon\mathbf{E}(\mathcal{L}_n)/n = 2\varepsilon\log n + {}$ $+\varepsilon(2\gamma - 4) + O(\varepsilon(\log n)/n)$ and use (18) to approximate w_* as $y \to \infty$. This yields

$$\begin{aligned} -yw_* + \int_1^{w_*} \frac{2e^u}{u} du &= y\left[-\log\left(\frac{y}{2}\right) - \log\log\left(\frac{y}{2}\right) + 1 + o(1)\right] \\ &= -2\varepsilon\log n\left[\log\log n - \log(1/\varepsilon) + \log\log(\varepsilon\log n) - 1\right] + o(\log n) \end{aligned} \tag{23}$$

which agrees precisely with the estimate (19), and also explicitly identifies the $O(\log\log\log n)$ error term. This suggests that (21) applies for y as large as $2\log n$, though it cannot hold for y as large as $n/2$ in view of the fact that $\Pr[\mathcal{L}_n = k] = 0$ for $k > \binom{n}{2}$. An important open problem is obtaining an accurate numerical approximation to the constant C. This would likely involve the numerical solution of the integral equation for $G_0(w)$.

3 Analysis of the Generating Function for $n \to \infty$

We study (4) asymptotically, for various ranges of n and u, namely: (i) $u \to 1$ with n fixed; (ii) $w \equiv n(u-1)$ fixed when $n \to \infty$ and $u \to 1$; (iii) $w \to \pm\infty$; and (iv) $u < 1$ or $u > 1$. We study these cases below.

A. CASE n FIXED AND $u \to 1$

First we consider the limit $u \to 1$ with n fixed. Then using the Taylor expansion

$$\begin{aligned} L_n(u) &= 1 + A_n(u-1) + B_n(u-1)^2 + O((u-1)^3) \\ &= e^{A_n(u-1)}[1 + (B_n - \frac{1}{2}A_n^2)(u-1)^2 + O((u-1)^3)] \end{aligned} \tag{24}$$

we find from (4) that $A_n = L_n'(1)$ and $B_n = L_n''(1)/2$ satisfy the linear recurrence equations

$$A_{n+1} = n + \frac{1}{n+1}\sum_{i=0}^{n}[A_i + A_{n-i}] = n + \frac{2}{n+1}\sum_{i=0}^{n} A_i; \quad A_0 = 0, \tag{25}$$

$$B_{n+1} = \binom{n}{2} + \frac{2n}{n+1}\sum_{i=0}^{n} A_i \frac{1}{n+1} + \sum_{i=0}^{n}[2B_i + A_i A_{n-i}]; \quad B_0 = 0. \tag{26}$$

These are easily solved using either generating functions, or by multiplying (25) and (26) by $n+1$ and then differencing with respect to n. The final result is (cf. [11, 13])

$$A_n = 2(n+1)H_n - 4n \tag{27}$$

$$B_n = 2(n+1)^2 H_n^2 - (8n+2)(n+1)H_n + \frac{n}{2}(23n+17) - 2(n+1)^2 H_n^{(2)}. \tag{28}$$

Here $H_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$ is the harmonic number, and $H_n^{(2)} = \sum_{k=1}^{n} k^{-2}$ is the harmonic number of second order. In terms of A_n and B_n, the mean and variance of $\mathcal{L}_n$ are given by

$$\mathbf{E}[\mathcal{L}_n] = A_n = 2(n+1)H_n - 4n \tag{29}$$

$$\begin{aligned} \mathbf{Var}\ [\mathcal{L}_n] &= L_n''(1) + L_n'(1) - [L_n'(1)]^2 = 2B_n + A_n - A_n^2 \\ &= 7n^2 - 2(n+1)H_n + 13n - 4(n+1)^2 H_n^{(2)}. \end{aligned} \tag{30}$$

Asymptotically, for $n \to \infty$, we obtain

$$A_n = 2n\log n + (2\gamma - 4)n + 2\log n + 2\gamma + 1 + \frac{5}{6}n^{-1} + O(n^{-2}) \tag{31}$$

$$B_n - \frac{1}{2}A_n^2 \equiv C_n = \left(\frac{7}{2} - \frac{\pi^2}{3}\right)n^2 - 2n\log n + n\left(\frac{21}{2} - 2\gamma - \frac{2}{3}\pi^2\right) + o(n). \tag{32}$$

These expressions will be used in order to asymptotically match the expansion for $u \to 1$ and n fixed, to those that will apply for other ranges of n and u.

B. Case $w \equiv n(u-1)$ fixed when $n \to \infty$ and $u \to 1$

Next we consider the limit $u \to 1$, $n \to \infty$ with $w \equiv n(u-1)$ held fixed. We define $G(\cdot)$ by

$$L_n(u) = \exp\left(A_n w/n\right) G(w;n) = e^{A_n(u-1)} G(n(u-1);n). \tag{33}$$

With this change of variables, we rewrite (4) as

$$\begin{aligned}
\exp\left(A_{n+1}w/n\right) \quad &\times \quad G(w+\frac{w}{n};n+1) = \frac{(1+w/n)^n}{n+1} \\
&\times \quad \left\{\sum_{i=m}^{n-m} \exp\left(A_i \frac{w}{n} + A_{n-i}\frac{w}{n}\right) G\left(w\frac{i}{n};i\right) G\left(w\left(1-\frac{i}{n}\right);n-i\right)\right. \\
&+ \quad \left. 2\sum_{i=0}^{m-1} \exp\left(A_i \frac{w}{n} + A_{n-i}\frac{w}{n}\right)\left(1+C_i\frac{w^2}{n^2}+\cdots\right) G\left(w\left(1-\frac{i}{n}\right);n-i\right)\right\}
\end{aligned} \tag{34}$$

where $C_i = B_i - \frac{1}{2}A_i^2$. Here we have broken up the sum in (4) into the three ranges $0 \le i \le m-1$, $m \le i \le n-m$, and $n-m+1 \le i \le n$, and used the symmetry $(i \to n-i)$ of the summand. We expect (24) to also be valid for large values of n, as long as $n(u-1) \to 0$ as $n \to \infty$. Thus, for $0 \le i \le m-1$ we replaced $L_i(u)$ in the sum by the approximation (24). The integer m may be chosen arbitrarily, since the right side of (34) must ultimately be independent of m. For now we assume that $m \to \infty$ but $m/n \to 0$ as $n \to \infty$. For n large we have (cf. (31))

$$\begin{aligned}
\frac{A_i}{n} + \frac{A_{n-i}}{n} - \frac{A_{n+1}}{n} \quad &= \quad 2\left(\frac{i}{n}\log\left(\frac{i}{n}\right) + \left(1-\frac{i}{n}\right)\log\left(1-\frac{i}{n}\right)\right) \\
&+ \quad \frac{2}{n}\log\left(\frac{i}{n}\left(1-\frac{i}{n}\right)\right) + \frac{3}{n} + o(n^{-1})
\end{aligned} \tag{35}$$

with which we rewrite (34) as

$$\begin{aligned}
(n+1)e^{-w} \quad &\times \quad \left(1+\frac{w^2}{2n}+O(n^{-2})\right) G\left(w+\frac{w}{n};n+1\right) \\
&= \quad \sum_{i=m}^{n-m} e^{2\phi(i/n)w}\left(1+w\left(\frac{2}{n}\psi\left(\frac{i}{n}\right)+\frac{3}{n}\right)+O(n^{-2},n^{-1}m^{-1})\right) \\
&\times \quad G\left(w\frac{i}{n};i\right) G\left(w\left(1-\frac{i}{n}\right);n-i\right) \\
&+ \quad 2\sum_{i=0}^{m-1}\left(1+A_i\frac{w}{n}+O\left(B_i n^{-2}\right)\right)\exp\left(\left(A_{n-i}-A_{n+1}\right)\frac{w}{n}\right) \\
&\times \quad G\left(w\left(1-\frac{i}{n}\right);n-i\right)
\end{aligned} \tag{36}$$

where

$$\begin{aligned}
\phi(x) \quad &= \quad x\log x + (1-x)\log(1-x) \\
\psi(x) \quad &= \quad \log[x(1-x)].
\end{aligned} \tag{37}$$

We now evaluate the two sums in (36) asymptotically and show that when the two results are added, the dependence on m disappears.

From (33) and the identity $L'_n(1) = A_n = \mathbf{E}[\mathcal{L}_n]$ we find that for all n

$$G(0;n) = 1 \quad \text{and} \quad G'(0;n) = 0. \tag{38}$$

We assume that for $n \to \infty$, $G(w;n)$ has an asymptotic expansion of the form

$$G(w;n) = G_0(w) + a_1(n)G_1(w) + a_2(n)G_2(w) + \cdots \tag{39}$$

where $a_j(n)$ is an asymptotic sequence as $n \to \infty$, i.e., $a_{j+1}(n)/a_j(n) \to 0$ as $n \to \infty$. The appropriate sequence is determined by balancing terms in (36). This will eventually yield

$$a_1(n) = \frac{\log n}{n}; \qquad a_2(n) = \frac{1}{n} \tag{40}$$

so we use this form from the beginning. Note that $G_0(w)$ is the moment generating function of the limiting density of $(\mathcal{L}_n - \mathbf{E}[\mathcal{L}_n])/n$, which is discussed in [15]. The conditions (38) imply that

$$\begin{aligned} G_0(0) &= 1; \quad G_1(0) = G_2(0) = \cdots = 0 & (41)\\ G'_0(0) &= G'_1(0) = G'_2(0) = \cdots = 0. & (42) \end{aligned}$$

We consider the first sum in (36), which we denote by $S_1 = S_1(n;m)$. Using (39), (40) and the Euler-MacLaurin formula we obtain

$$\begin{aligned} S_1 &= \sum_{i=m}^{n-m} e^{2\phi(i/n)w} \left[1 + w\left(\frac{2}{n}\psi\left(\frac{i}{n}\right) + \frac{3}{n}\right) + O(n^{-2})\right] \qquad (43)\\ &\times \left\{G_0\left(w\frac{i}{n}\right) G_0\left(w\left(1-\frac{i}{n}\right)\right) + G_0\left(w\frac{i}{n}\right)\frac{\log(n-i)}{n-i}G_1\left(w\left(1-\frac{i}{n}\right)\right)\right.\\ &+ G_0\left(w\left(1-\frac{i}{n}\right)\right)\frac{\log i}{i}G_1\left(w\frac{i}{n}\right) + G_0\left(w\frac{i}{n}\right)\frac{1}{n-i}G_2\left(w\left(1-\frac{i}{n}\right)\right)\\ &+ \left. G_0\left(w\left(1-\frac{i}{n}\right)\right)\frac{1}{i}G_2\left(w\frac{i}{n}\right) + \cdots\right\}\\ &= n\int_{m/n}^{1-m/n} e^{2\phi(x)w}G_0(wx)G_0(w-wx)dx + e^{2\phi(m/n)w}G_0\left(w\frac{m}{n}\right)G_0\left(w\left(1-\frac{m}{n}\right)\right)\\ &+ w\int_{m/n}^{1-m/n}[2\psi(x)+3]e^{2\phi(x)w}G_0(wx)G_0(w-wx)dx\\ &+ \int_{m/n}^{1-m/n}\frac{\log n + \log(1-x)}{1-x}e^{2\phi(x)w}G_0(wx)G_1(w-wx)dx\\ &+ \int_{m/n}^{1-m/n}\frac{\log n + \log x}{x}e^{2\phi(x)w}G_0(w-wx)G_1(wx)dx \end{aligned}$$

$$
\begin{aligned}
&+ \int_{m/n}^{1-m/n} \frac{1}{1-x} e^{2\phi(x)w} G_0(wx) G_2(w-wx)\,dx \\
&+ \int_{m/n}^{1-m/n} \frac{1}{x} e^{2\phi(x)w} G_2(wx) G_0(w-wx)\,dx + o(1).
\end{aligned}
$$

We note that all the integrals remain finite as $m/n \to 0$, in view of (41) and (42). However, if we were to consider higher order terms in the expansion (39), which would involve terms of order n^{-2} and n^{-3}, then the corresponding higher order terms in (43) would involve integrands not integrable over [0,1]. It then becomes essential that we integrate only over the range $[m/n, 1-m/n]$ and consider the contribution from the second sum $S_2 = S_2(n;m)$ in (36). We can further simplify S_1 by evaluating each term in the limit $m/n \to 0$, which we assumed to be true. We have

$$
\begin{aligned}
T_{11} &\equiv \int_{m/n}^{1-m/n} e^{2\phi(x)w} G_0(wx) G_0(w-wx)\,dx \\
&= \int_0^1 e^{2\phi(x)w} G_0(wx) G_0(w-wx)\,dx - 2\int_0^{m/n} e^{2\phi(x)w} G_0(wx) G_0(w-wx)\,dx \\
&= \int_0^1 e^{2\phi(x)w} G_0(wx) G_0(w-wx)\,dx - 2\int_0^{m/n} [1 + 2w(x\log x - x) + O(x^2)] \\
&\times\ [1+O(x^2)][G_0(w) - wxG_0'(w) + O(x^2)]\,dx \\
&\sim \int_0^1 e^{2\phi(x)w} G_0(wx) G_0(w-wx)\,dx \\
&- 2G_0(w)\left[\frac{m}{n} + 2w\left(\frac{m^2}{2n^2}\log\left(\frac{m}{n}\right) - \frac{3}{4}\frac{m^2}{n^2}\right)\right] + \frac{m^2}{n^2} wG_0'(w);
\end{aligned}
$$

$$
\begin{aligned}
T_{12} &\equiv 2\int_{m/n}^{1-m/n} \frac{\log n + \log x}{x} e^{2\phi(x)w} G_0(w-wx) G_1(wx)\,dx \\
&\sim 2\int_0^1 \frac{\log n + \log x}{x} e^{2\phi(x)w} G_0(w-wx) G_1(wx)\,dx - 2\frac{m}{n}(\log n) G_1(w);
\end{aligned}
$$

and

$$
\begin{aligned}
T_{13} &\equiv 2\int_{m/n}^{1-m/n} e^{2\phi(x)w} \frac{1}{x} G_2(wx) G_0(w-wx)\,dx \\
&\sim 2\int_0^1 e^{2\phi(x)w} \frac{1}{x} G_2(wx) G_0(w-wx)\,dx - 2\frac{m}{n} G_2(w).
\end{aligned}
$$

Thus S_1 simplifies to

$$
\begin{aligned}
S_1 &= n\int_0^1 e^{2\phi(x)w} G_0(wx) G_0(w-wx)\,dx + G_0(w) \qquad (44)\\
&+ w\int_0^1 [2\psi(x) + 3] e^{2\phi(x)w} G_0(wx) G_0(w-wx)\,dx
\end{aligned}
$$

$$
\begin{aligned}
&+ \quad 2\int_0^1 \frac{\log n + \log x}{x} e^{2\phi(x)w} G_0(w-wx)G_1(wx)dx \\
&+ \quad 2\int_0^1 \frac{1}{x} e^{2\phi(x)w} G_0(w-wx)G_2(wx)dx \\
&+ \quad \left\{ -2mG_0(w) + w\frac{m^2}{n}G_0'(w) - \frac{2m^2}{n}wG_0(w)\left[\log\left(\frac{m}{n}\right) - \frac{3}{2}\right] \right. \\
&- \quad \left. 2\frac{m}{n}(\log n)G_1(w) - 2\frac{m}{n}G_2(w) \right\} + \cdots
\end{aligned}
$$

where we have grouped the terms involving m inside the "{ }". The error term in (44) approaches zero as $n \to \infty$.

Now we consider the second sum S_2 in (36). Using $A(n-i) - A(n+1) \sim -(i+1)A'(n) \sim -(i+1)(2\log n + 2\gamma - 2)$, we obtain

$$
\begin{aligned}
S_2 \quad &\sim \quad 2\sum_{i=0}^{m-1}\left[1 + A_i\frac{w}{n}\right]\left[1 - \frac{w}{n}(i+1)(2\log n + 2\gamma - 2)\right] \\
&\times \quad \left[G_0(w) - \frac{i}{n}wG_0'(w) + \frac{\log n}{n}G_1(w) + \frac{1}{n}G_2(w)\right] \\
&\sim \quad 2G_0(w)\left[m + \frac{w}{n}\sum_{i=0}^{m-1}(A_i - 2(\log n)(i+1) - (2\gamma - 2)(i+1))\right] \\
&+ \quad 2\frac{m}{n}(\log n)G_1(w) + G_2(w)] - \frac{m^2}{n}wG_0'(w) \\
&\sim \quad 2G_0(w)m + 2\frac{m^2}{n}wG_0(w)\left[\log\left(\frac{m}{n}\right) - \frac{3}{2}\right] - \frac{m^2}{n}wG_0'(w) \\
&+ \quad 2\frac{m}{n}[(\log n)G_1(w) + G_2(w)].
\end{aligned}
\tag{45}
$$

Here we have used $2\sum_{i=0}^{n-1} A_i = n[A_n - (n-1)]$. Upon adding (44) to (45), we see that all the terms involving m cancel, and that the leading three terms in the expansion of $S_1 + S_2$ (i.e., the right side of (36)) are of order $O(n)$, $O(\log n)$ and $O(1)$, respectively. Using (39) and (40), the left side of (36) becomes

$$
(n+1)e^{-w}\left(1 + \frac{w^2}{2n}\right)\left(G_0(w) + \frac{\log n}{n}G_1(w) + \frac{1}{n}(wG_0'(w) + G_2(w)) + o(n^{-1})\right).
$$

Thus, comparing the above to $S_1 + S_2$, we find that

$$
e^{-w}G_0(w) = \int_0^1 e^{2\phi(x)w}G_0(wx)G_0(w-wx)dx, \tag{46}
$$

$$
e^{-w}G_1(w) = 2\int_0^1 \frac{1}{x}e^{2\phi(x)w}G_0(w-wx)G_1(wx)dx, \tag{47}
$$

$$
e^{-w}[G_0(w) + \frac{1}{2}w^2G_0(w) + wG_0'(w) + G_2(w)] = G_0(w) \tag{48}
$$

$$
\begin{aligned}
&+ \quad w \int_0^1 [2\psi(x)+3] e^{2\phi(x)w} G_0(wx) G_0(w-wx)\,dx \\
&+ \quad 2 \int_0^1 \frac{\log x}{x} e^{2\phi(x)w} G_0(w-wx) G_1(wx)\,dx \\
&+ \quad 2 \int_0^1 \frac{1}{x} e^{2\phi(x)w} G_0(w-wx) G_2(wx)\,dx.
\end{aligned}
$$

Equations (46)-(48), along with (37), (41) and (42) are integral equations for the first three terms in the series (39). Below we discuss some aspects of the solutions to these problems. The leading order equation (46) was previously obtained in [5], using more probabilistic arguments.

We observe that the solution to (46) is not unique: $G_0(w) = 0$ is one solution and if $G_0(w)$ is any solution, then so is $e^{cw} G_0(w)$ for any constant c. We can construct the solution as a Taylor series:

$$
G_0(w) = 1 + \sum_{j=1}^{\infty} g_j w^j. \tag{49}
$$

This eliminates the trivial solution $G_0(w) = 0$ and satisfies the normalization $G_0(0) = 1$. Using (49) in (46) and noting that $\int_0^1 [2\phi(x)+1]dx = 0$, we see that g_1 remains arbitrary, and then we can easily calculate g_j for $j \geq 2$ in terms of g_1. But, (42) forces $g_1 = 0$ and then all the Taylor coefficients in (49) are uniquely determined. They may be evaluated from the recursion

$$
\begin{aligned}
\left(1 - \frac{2}{n+1}\right) g_n \quad &= \quad \sum_{i=1}^{n-1} B(i, n-i, 0) g_i g_{n-i} \\
&+ \quad \sum_{\ell=0}^{n-1} \sum_{i=0}^{\ell} B(i, \ell-1, n-\ell) g_i g_{\ell-i}
\end{aligned} \tag{50}
$$

for $n \geq 2$ where

$$
B(i,j,k) = \int_0^1 x^i (1-x)^j \frac{1}{k!} [2\phi(x)+1]^k dx. \tag{51}
$$

In particular, $g_2 = \frac{7}{2} - \frac{\pi^2}{3}$.

Next we consider the equations (47) and (48) for the correction terms G_1 and G_2. These are linear, Fredholm integral equations of the second kind. Their solutions may also be constructed as Taylor series in w. In view of (41) and (42) we must have

$$
\begin{aligned}
G_0(w) \quad &\sim \quad 1 + \alpha_0 w^2; \\
G_1(w) \quad &\sim \quad \beta_0 w^2; \\
G_2(w) \quad &\sim \quad \gamma_0 w^2,
\end{aligned}
$$

as $w \to 0$, where we have already computed $\alpha_0 = g_2$. Given β_0, we can easily compute the higher order Taylor coefficients of $G_1(w)$ from (47), in terms of the (now uniquely determined) Taylor coefficients of $G_0(w)$. However, the constant β_0 cannot be determined solely from (47), (41) and (42). To fix β_0 we use the principle of asymptotic matching. We require that expansions (24) and (33) (with (39)) agree in some intermediate limit, where $u \to 1$, $n \to \infty$ and $n(u-1) \to 0$. Then the behavior of (24) as $n \to \infty$ must agree with the behavior of (33) (with (39)) as $w \to 0$. Writing (24) as $L_n(u) = \exp[A_n(u-1)][1 + C_n(u-1)^2 + O((u-1)^3)]$ the matching condition becomes

$$1 + C_n(u-1)^2 + \cdots \Big|_{n\to\infty} \sim G_0(w) + \frac{\log n}{n} G_1(w) + \frac{1}{n} G_2(w) + \cdots \Big|_{w\to 0}. \tag{52}$$

Setting $u - 1 = w/n$, using (32), and noting that the right side of (52) is $1 + n^{-2}[\alpha_0 n^2 + \beta_0(\log n)n + \gamma_0 n]w^2$, we obtain

$$\beta_0 = -2; \quad \gamma_0 = \frac{21}{2} - 2\gamma - \frac{2}{3}\pi^2. \tag{53}$$

Now consider equation (48) in the limit $w \to 0$. We write (47) abstractly as $TG_1 = 0$ where T is the linear integral operator in (47). Then (48) may be written as $TG_2 = f(G_0, G_1)$ where f is a known function of the first two terms. Since $TG_1 = 0$ has a non-zero solution (made unique by the condition $\beta_0 = -2$), we expect that $TG_2 = f$ will have a solution only if a solvability condition is met. To obtain this solvability condition we expand (48) as $w \to 0$. Obviously (48) is satisfied as $w \to 0$, and, since $\int_0^1 [2\psi(x) + 3]dx = -1$, (48) also holds to order $O(w)$. Comparing $O(w^2)$ terms in (48) using (52) gives

$$\begin{aligned} 1 + 2\alpha_0 &= -\frac{1}{2}\beta_0 - 3 + 4\int_0^1 \phi(x)\psi(x)dx \\ &= -\frac{1}{2}\beta_0 - 3 + 4\left(\frac{5}{2} - \frac{\pi^2}{6}\right) \end{aligned} \tag{54}$$

and thus $\beta_0 = -2$, which regains the result we obtained by using asymptotic matching. Note, however, that this argument required that we derive the equation satisfied by $G_2(w)$ in order to uniquely specify the previous term $G_1(w)$. In contrast, asymptotic matching made no use of (48). Presumably, by deriving the equation for the next term $G_3(w)$ in (39) and examining its solvability condition, we would have an alternate way of computing γ_0. Given β_0 and γ_0, we can easily obtain the Taylor expansions of G_1 and G_2 from (47) and (48).

To summarize, we have obtained the expansion (33), with (39)-(42), (46)-(48) and (53), for the scaling $n \to \infty$, $u \to 1$ with $n(u-1) = w$ fixed. We have not been able to explicitly solve these integral equations. However, we can derive some approximate formulas in the limits $w \to \pm\infty$, and these may be used to obtain approximations to the tail probabilities of the Quicksort distribution.

C. CASE $w \equiv n(u-1) \to \pm\infty$

We shall only examine the leading term G_0, and we first consider the limit $w \to -\infty$. As $w \to -\infty$, the "kernel" $\exp[2w\phi(x)]$ in (46) is sharply concentrated near $x = 1/2$, and behaves as a multiple of $|w|^{-1/2}\delta(x-1/2)$. Thus we treat (46) as a Laplace type integral (cf. [8]).

Assuming that $G_0(w)$ has a weak (e.g., algebraic) dependence on w, we approximate the right side of (46) by Laplace's method, which (to leading order) yields the functional equation

$$e^{-w}G_0(w) \approx \left[G_0\left(\frac{w}{2}\right)\right]^2 \sqrt{\frac{\pi}{-4w}}e^{-2w\log 2}, \qquad w \to -\infty. \tag{55}$$

But, if G_0 varies weakly with w, then the exponential orders of magnitude of the right and left sides of (55) do not agree. In order to get agreement, we would need $G_0(w)$ to vary much more rapidly as $w \to -\infty$, of the order $\exp[O(w\log(-w))]$. But this then contradicts the assumption used to obtain (55). Therefore, we return to (46) and allow more rapid variation of G_0. Specifically, we assume that as $w \to -\infty$

$$G_0(w) \sim e^{\alpha_1 w\log(-w)}e^{\beta w}(-w)^{\gamma_1}\delta_1. \tag{56}$$

Using (56) in (46) yields

$$e^{-w}(-w)^{\gamma_1}\delta_1 \sim \delta_1^2(-w)^{2\gamma_1}\int_0^1 x^{\gamma_1}(1-x)^{\gamma_1}e^{\phi(x)(2+\alpha_1)w}dx. \tag{57}$$

If $2+\alpha_1 > 0$ we again use Laplace's method to approximate the integral, thus obtaining

$$e^{-w} \sim \delta_1\left(\frac{1}{4}\right)^{\gamma_1}(-w)^{\gamma_1-\frac{1}{2}}\sqrt{\frac{\pi}{2(2+\alpha_1)}}e^{-(\alpha_1+2)(\log 2)w}. \tag{58}$$

Hence, we must have

$$\alpha_1 = -2+\frac{1}{\log 2}; \quad \gamma_1 = \frac{1}{2}; \quad \delta_1 = \frac{2\sqrt{2}}{\sqrt{\pi}\log 2} \tag{59}$$

and β remains arbitrary. Since the solution of (46) is not unique, we can never determine β using only this integral equation. We thus have

$$G_0(w) \sim \frac{2\sqrt{2}}{\sqrt{\pi}\log 2}\sqrt{-w}e^{\beta w}\exp\left[\left(\frac{1}{\log 2}-2\right)w\log(-w)\right], \qquad w \to -\infty. \tag{60}$$

By computing higher order terms in the expansion of the integral in (46), we can extend (56) and (60) to an asymptotic series in powers of w^{-1}. To fix β, we must use the condition in (41) and (42), and this probably requires a numerical solution of (46). By comparing this numerical solution to (46) for large, negative w to (60), we can obtain an approximation to β.

Now consider (46) in the limit $w \to +\infty$. Then $\exp[2w\phi(x)]$ is concentrated near the endpoints $x = 0$ and $x = 1$. We assume that $G_0(w)$ has an asymptotic expansion in the WKB form [6]:

$$G_0(w) \sim K(w)\exp[\Psi(w)]. \tag{61}$$

The major contribution to the integral will come from where $x, 1-x = o(w^{-1})$. Thus we use the Taylor series $G_0(wx) = 1 + O((wx)^2)$, and (61) to approximate $G_0(w - wx)$. This yields

$$\begin{aligned} e^{-w}K(w)e^{\Psi(w)} &\sim 2\int_0^{1/2}[1 + 2w\phi(x) + O(w^2\phi^2(x))][1 + O((wx)^2)]K(w - wx)e^{\Psi(w-wx)}dx \\ &\sim 2K(w)e^{\Psi(w)}\int_0^{1/2} e^{-w\Psi'(w)x}dx. \end{aligned} \tag{62}$$

If furthermore $\Psi(\cdot)$ is such that $w\Psi'(w) \to \infty$ as $w \to \infty$, then (62) yields (after a slight rearrangement)

$$w\Psi'(w) = 2e^w$$

and hence

$$\Psi(w) = 2\int^w \frac{e^u}{u}du. \tag{63}$$

Note that (63) is consistent with $w\Psi'(w) \to \infty$ and shows that G_0 grows very rapidly (as a double exponential) as $w \to +\infty$. To obtain the next term $K(w)$, we refine the expansion (62) to

$$\begin{aligned} e^{-w}K(w)e^{\Psi(w)} &= 2\int_0^{1/2}[1 + 2w\phi(x) + O(w^2\phi^2(x))][1 + O((wx)^2)] \\ &\times [K(w) - wxK'(w) + O(w^2x^2K''(w))] \\ &\times \exp\left(\Psi(w) - w\Psi'(w)x + \frac{1}{2}w^2\Psi''(w)x^2 + O(w^3\Psi'''(w)x^3)\right)dx. \end{aligned} \tag{64}$$

For x small we have $2w\phi(x) = 2w(x\log x - x + O(x^2))$ and if $x = O(e^{-w})$ then $w^2x^2 = O(w^2e^{-2w})$ and $w^3x^3\Psi'''(w) = O(w^2e^{-2w})$. Setting $x = \eta e^{-w}/2$ we have

$$\exp\left[\frac{1}{2}x^2w^2\Psi''(w)\right] = 1 + \frac{1}{4}\eta^2(w-1)e^{-w}[1 + o(1)].$$

Thus (64) becomes, after cancelling the common factor $\exp[\Psi(w)]$,

$$\begin{aligned} e^{-w}K(w) &= e^{-w}\left\{K(w)\int_0^\infty e^{-\eta}d\eta\right. && (65)\\ &+ K(w)\int_0^\infty e^{-\eta}\left[\frac{1}{4}e^{-w}\eta^2(w-1)+we^{-w}\eta\left(-w+\log\left(\frac{\eta}{2}\right)-1\right)\right]d\eta \\ &- \left. K'(w)we^{-w}\int_0^\infty \frac{1}{2}\eta e^{-\eta}d\eta + (\text{error})\right\}. \end{aligned}$$

Here the error terms are roughly of the order $(e^{-2w}K(w))$. To leading order (65) is obviously satisfied and then collecting terms that are of order $O(e^{-w})$ as $w\to\infty$ (modulo some algebraic factors) we obtain the following differential equation

$$\frac{w}{2}\frac{K'(w)}{K(w)} = \int_0^\infty \left[\frac{1}{4}\eta^2 e^{-\eta}(w-1)+w\eta e^{-\eta}\left(-w+\log\left(\frac{\eta}{2}\right)-1\right)\right]d\eta \qquad (66)$$

and thus

$$\frac{K'(w)}{K(w)} = -2w+1-2\gamma-2\log 2-\frac{1}{w}. \qquad (67)$$

Solving (67) and using the result along with (63), we have

$$G_0(w)\sim C_*\exp\left(\int_1^w \frac{2e^u}{u}du\right)e^{-w^2}e^{(1-2\gamma-2\log 2)w}\frac{1}{w}. \qquad (68)$$

Here C_* is an undetermined constant and we have chosen the lower limit on the integral in (63) as one. An alternate choice would only change C_*, which we cannot determine using only the integral equation (46).

Our analysis shows that as $w\to+\infty$, the nonlinear equation (46) may be approximated by a linear one. To fix C_* it would seem that we will again need an accurate numerical solution to (46).

We have thus obtained formal asymptotic results as $w\to\pm\infty$ for the solution to (46). Using our procedure we can derive full asymptotic series in these limits, but the constants β and C_* will remain undetermined.

In Section 4, we will use our results for $G_0(w)$ to obtain asymptotic expansions for the limiting density $P(y)$ as $y\to\pm\infty$.

D. Case $u<1$ or $u>1$

We next study (4) for $n\to\infty$ but for fixed $u>1$ or $u<1$. First we assume that $0<u<1$. Note that u appears in (4) only as a parameter. We assume an expansion of the form

$$L_n(u)\sim e^{A(u)n\log n}e^{B(u)n}n^{C(u)}D(u) \qquad (69)$$

for $n \to \infty$ and $u < 1$. The major contribution to the sum in (4) will come from the midpoint $i \approx n/2$. Using (69) in (4) and noting that $(n+1)\log(n+1) = n\log n + \log n + 1 + O(n^{-1})$ and $i\log i + (n-i)\log(n-i) = n\log n - n\log 2 + 2n^{-1}(i-n/2)^2 + O(n^{-2}(i-n/2)^3)$, we obtain

$$
\begin{aligned}
e^{An\log n}e^{A\log n}e^{A+B}e^{Bn}n^C D &\sim D^2\left(\frac{n}{2}\right)^{2C} e^{Bn}\frac{u^n}{n}e^{An\log n}e^{-An\log 2} \\
&\times \sum_{i=0}^{n}\exp\left(-(2/n)|A|(i-n/2)^2\right)
\end{aligned}
$$

for $A < 0$. The sum is asymptotically equal to $\sqrt{n\pi/2|A|}$, and thus

$$
\begin{aligned}
A &= \frac{\log u}{\log 2}, \qquad C = \frac{1}{2} + \frac{\log u}{\log 2}, \\
D &= 2e^B u^2\sqrt{\frac{-2\log u}{\pi\log 2}}\exp\left(\frac{\log u}{\log 2}\right)
\end{aligned}
\tag{70}
$$

and $B = B(u)$ remains arbitrary.

Some further information may be obtained by asymptotically matching (69) to the expansion valid for $n \to \infty$, $u - 1 = O(n^{-1})$. In the intermediate limit where $u \to 1$, $n \to \infty$ and $n(u-1) \to -\infty$, we use (60) with $w = n(u-1)$ to get

$$
\begin{aligned}
&e^{A_n(u-1)}G_0(w) \sim e^{2(u-1)n\log n}e^{(2\gamma-4)n(u-1)}\frac{2\sqrt{2}}{\sqrt{\pi\log 2}}e^{\beta n(u-1)} \\
&\times\sqrt{n(1-u)}\exp\left\{\left(\frac{1}{\log 2}-2\right)n(u-1)[\log n + \log(1-u)]\right\}.
\end{aligned}
$$

As $u \uparrow 1$, (69) and (70) yields

$$
\exp\left\{\left[\frac{u-1}{\log 2} + O((u-1)^2)\right]n\log n + Bn\right\}n^{1/2}2e^B\sqrt{\frac{2(1-u)}{\pi\log 2}}
$$

and these two expressions agree provided that as $u \uparrow 1$

$$
B(u) \sim \left(\frac{1}{\log 2} - 2\right)(u-1)\log(1-u) + (2\gamma - 4 + \beta)(u-1), \qquad u \to 1. \tag{71}
$$

This relates the behavior of $B(u)$ as $u \uparrow 1$ to the constant β in (60).

Now consider $n \to \infty$ with $u > 1$. The dominant contribution in the sum (4) now comes from the terms with $i = 0$ and $i = n$. Thus

$$
L_{n+1}(u) = \frac{u^n}{n+1}[2L_0(u)L_n(u) + 2L_1(u)L_{n-1}(u) + \cdots] \sim \frac{2u^n}{n+1}L_n(u) \tag{72}
$$

and

$$
L_n(u) \sim u^{\binom{n}{2}}\frac{2^n}{n!}k_1(u); \qquad n \to \infty, u > 1 \tag{73}
$$

where $k_1(u)$ is an undetermined function. Since for $u \to \infty$ and all $n \geq 2$, $L_n(u) \sim u^{\binom{n}{2}} 2^n/[4n!]$, we have $k_1(u) \to 1/4$ as $u \to \infty$.

We examine asymptotic matching between the expansions for $u > 1$ and $u - 1 = O(n^{-1})$. If (73) matches to $e^{A_n(u-1)} G_0(w)$ then (68) would agree with the expansion of (73) as $u \downarrow 1$. However this cannot be true as the dominant exponential term in (68) is

$$O\left[\exp\left(\frac{2e^w}{w}\right)\right] = O\left[\exp\left(\frac{2e^{n(u-1)}}{n(u-1)}\right)\right] \tag{74}$$

while the dominant term in (73) as $u \to 1$ is

$$O\left[\exp\left(\frac{n^2}{2}(u-1)\right)\right]. \tag{75}$$

This suggests that yet another expansion is needed on some scale where $n \to \infty$ and $u \downarrow 1$ with $n(u-1) \to +\infty$. By comparing (74) to (75), this new scale is likely to be

$$w = n(u-1) = \log n + 2 \log\log n + O(1). \tag{76}$$

With this scaling both (74) and (75) are $\exp[O(n \log n)]$. We have not examined this intermediate scale in any detail. However, we would guess that with (76), the approximation of (4) will involve retaining an infinite number of terms in this sum (rather than just the 2 terms in (72)), but not approximating the sum by an integral, as was possible when $u - 1 = O(n^{-1})$.

To summarize this section, we have analyzed (4) in various asymptotic limits. These include (i) $u \to 1$, n fixed; (ii) $u \to 1$, $n \to \infty$, $n(u-1) = w$ fixed; (iii) $w \to \pm\infty$; (iv) $0 < u < 1$, $n \to \infty$; and (v) $u > 1$, $n \to \infty$. In the next section, we use these results to obtain information about the distribution $\mathsf{Pr}[\mathcal{L}_n = k]$.

4 Tails of the Limiting Distribution

Using the approximation (33) for $u - 1 = O(n^{-1})$, we obtain

$$\mathsf{Pr}[\mathcal{L}_n - \mathbf{E}(\mathcal{L}_n) = ny] = \frac{1}{2\pi i}\int_C z^{-[ny+A_n+1]} L_n(z)\,dz. \tag{77}$$

$$\begin{aligned} &\sim \frac{1}{n}\frac{1}{2\pi i}\int_C \left(1+\frac{w}{n}\right)^{-ny-1} e^{-A_n \log(1+w/n)} e^{A_n w/n} G_0(w)\,dw \\ &\sim \frac{1}{n}\frac{1}{2\pi i}\int_{Br} e^{-yw} G_0(w)\,dw = \frac{1}{n}P(y) \end{aligned}$$

where $Br = (c - i\infty, c + i\infty)$ for some constant c, is any vertical contour in the w-plane. Here we have set $z = 1 + w/n$ in the integral. It follows that

$$G_0(w) = \int_{-\infty}^{\infty} e^{wy} P(y) dy \tag{78}$$

so that $G_0(w)$ is the moment generating function of the density $P(y)$. In view of (41) and (42) we have $\int_{-\infty}^{\infty} P(y)dy = 1$ and $\int_{-\infty}^{\infty} yP(y)dy = 0$.

Observe that, using (33), (39), and (40), we can refine the approximation (77) to

$$\begin{aligned}\mathsf{Pr}[\mathcal{L}_n - \mathbf{E}(\mathcal{L}_n) = ny] &= \frac{1}{n}\Big(P(y) + \frac{\log n}{n}[P_1(y) + P''(y)] \\ &+ \frac{1}{n}[P_2(y) + P'(y) + \frac{1}{2}yP''(y) + (\gamma - 2)P''(y)] + o(n^{-1})\Big)\end{aligned}$$

where

$$P_k(y) = \frac{1}{n}\frac{1}{2\pi i}\int_{Br} e^{-yw} G_k(w) dw$$

for $k = 1, 2$.

An integral equation for $P(y)$ can easily be derived from (46). We multiply (46) by $e^{-wy}/(2\pi i)$ and integrate over a contour Br in the w-plane:

$$\begin{aligned} P(y+1) &= \frac{1}{2\pi i}\int_{Br} e^{-w(y+1)} G_0(w) dw \\ &= \frac{1}{2\pi i}\int_{Br}\int_0^1 e^{2w\phi(x)} G_0(wx) G_0(w - wx) e^{-wy} dx dw \\ &= \int_0^1 \frac{1}{2\pi i}\int_{Br} e^{-wy}\int_{-\infty}^{\infty} P(\xi) e^{wx\xi} d\xi \int_{-\infty}^{\infty} P(\eta) e^{(w - wx)\eta} d\eta\ e^{2w\phi(x)} dw dx \\ &= \int_0^1\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(\xi)P(\eta)\delta(y - x\xi - (1-x)\eta - 2\phi(x)) d\eta d\xi dx \\ &= \int_0^1 \frac{1}{x}\int_{-\infty}^{\infty} P(\eta) P\left(\frac{y}{x} - \frac{1-x}{x}\eta - \frac{2\phi(x)}{x}\right) d\eta dx \\ &= \int_0^1\int_{-\infty}^{\infty} P\left(xt + \frac{y/2 - \phi(x)}{1-x}\right) P\left(-(1-x)t + \frac{y/2 - \phi(x)}{x}\right) dt dx, \end{aligned} \tag{79}$$

where we used the well known identity (cf. [1])

$$\frac{1}{2\pi i}\int_{Br} e^{-yw} dw = \delta(y)$$

where $\delta(y)$ is Dirac's delta function. The last expression is precisely (13). The solution to this integral equation is not unique: if $P(y)$ is a solution, so is $P(y+c)$ for any c.

We study $P(y) = (2\pi i)^{-1}\int_{Br} e^{-yw} G_0(w) dw$ as $y \to \pm\infty$. We argue that the asymptotic expansion of the integral will be determined by a saddle point $w = s(y)$, which satisfies

$s(y) \to \pm\infty$ as $y \to \pm\infty$. Thus for $y \to -\infty$, we can use the approximation (60) for $G_0(w)$, which yields

$$P(y) \sim \frac{1}{2\pi i}\int_{Br} \frac{2\sqrt{2}}{\sqrt{\pi\log 2}}\sqrt{-w}e^{(\beta-y)w} \exp\left[\left(\frac{1}{\log 2}-2\right) w\log(-w)\right] dw. \tag{80}$$

This integrand has a saddle point where

$$\frac{d}{dw}\left[-(y-\beta)w + \left(\frac{1}{\log 2}-2\right) w\log(-w)\right] = 0$$

so that

$$w = -\frac{1}{e}\exp\left[-\frac{y-\beta}{2-1/\log 2}\right] \equiv \tilde{w}(y)$$

which satisfies $\tilde{w}(y) \to -\infty$ as $y \to -\infty$. Then the standard saddle point approximation to (80) yields

$$\begin{aligned} P(y) &\sim \frac{2\sqrt{2}}{\sqrt{\pi\log 2}}\sqrt{-\tilde{w}}e^{-(y-\beta)\tilde{w}} \exp\left[\left(\frac{1}{\log 2}-2\right)\tilde{w}\log(-\tilde{w})\right] \\ &\quad\times\frac{1}{2\pi i}\int_{Br} \exp\left[\frac{1}{2\tilde{w}}\left(\frac{1}{\log 2}-2\right)(w-\tilde{w})^2\right] dw \\ &= \frac{2}{\pi e}\frac{1}{\sqrt{2\log 2 - 1}} \exp\left[\frac{\beta-y}{2-1/\log 2} - \frac{2-1/\log 2}{e}\exp\left(\frac{\beta-y}{2-1/\log 2}\right)\right] \end{aligned} \tag{81}$$

for $y \to -\infty$. Thus, the left tail is very small and the behavior of $P(y)$ as $y \to -\infty$ is similar to that of an extreme value distribution (i.e., double-exponential distribution).

Now take $y \to +\infty$ and use (68) to get

$$P(y) \sim \frac{1}{2\pi i}\int_{Br} e^{-yw} C_* \exp\left(\int_1^w \frac{2e^u}{u}du\right)\frac{e^{-w^2}}{w}e^{(1-2\gamma-2\log 2)w}dw. \tag{82}$$

The saddle point now satisfies

$$\frac{d}{dw}\left[-yw + \int_1^w \frac{2e^u}{u}du\right] = 0$$

or $y = 2e^w/w$. Let $w_* = w_*(y)$ be the solution to (17) that satisfies $w_* \to \infty$ as $y \to \infty$. Then expanding the integrand in (82) about $w = w_*(y)$ and using the standard saddle point approximation yields

$$P(y) \sim \frac{C_*}{2\sqrt{2\pi}}\frac{\sqrt{y}}{\sqrt{1-1/w_*}} \exp\left[-yw_* + \int_1^{w_*}\frac{2e^u}{u}du\right]\exp[-w_*^2 - (2\gamma + 2\log 2)w_*] \tag{83}$$

as $y \to \infty$, from which (16) easily follows. Thus for $y \to \infty$ we have $P(y) = \exp[O(-y\log y)]$ and hence the right tail is *thinner* than that of an extreme value distribution. From (17) it is easy to show that

$$w_* = \log\left(\frac{y}{2}\right) + \log\log\left(\frac{y}{2}\right) + \frac{\log\log(y/2)}{\log(y/2)}(1+o(1)), \qquad y \to \infty. \tag{84}$$

For fixed z and y we have, as $n \to \infty$,

$$\begin{aligned} \Pr[\mathcal{L}_n - \mathbf{E}(\mathcal{L}_n) \le nz] &\sim \int_{-\infty}^{z} P(y)dy & (85) \\ \Pr[\mathcal{L}_n - \mathbf{E}(\mathcal{L}_n) \ge ny] &\sim \int_{y}^{\infty} P(u)du. & (86) \end{aligned}$$

If $z \to -\infty$ or $y \to +\infty$, then these integrals may be evaluated asymptotically using (81) and (83), and we obtain the results (15) and (16), respectively.

This derivation of the expansions of $P(y)$ as $y \to \pm\infty$ has the disadvantage in that it assumes the existence of certain saddle points. However, we can obtain the identical results simply by using the integral equation for $P(y)$, which we now show.

Let us write (79) in the form

$$P(y+1) = \int_0^1 \frac{1}{x(1-x)} \int_{-\infty}^{\infty} P\left(\frac{\xi + y/2 - \phi(x)}{1-x}\right) P\left(\frac{-\xi + y/2 - \phi(x)}{x}\right) d\xi dx \qquad (87)$$

and assume that for $y \to -\infty$, $P(y)$ has the form

$$P(y) \sim de^{cy} \exp[-ae^{-by}] \qquad (88)$$

for some constants a, b, c, d. Using (88) in (87) we find that the major contribution to the double integral will come from $(x, \xi) = \left(\frac{1}{2}, 0\right)$. After scaling $x = \frac{1}{2}(1 + e^{by/2}u)$ and $\xi = e^{by/2}\eta$ we obtain

$$\begin{aligned} & de^c e^{cy} \exp[-ae^{-b}e^{-by}] \sim 2e^{by} d^2 \exp[-2ae^{-by}e^{-2b\log 2}] \\ & \times e^{2cy} e^{4c\log 2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\nu(u(y+2\log 2)+2\eta)^2 b^2} e^{-2\nu b(u^2(y+2\log 2-1)+2\eta u)} du d\eta \end{aligned}$$

where $\nu = ae^{-2b\log 2}$. The double integral is equal to $\pi(2\nu)^{-1} b^{-3/2} (2 - 1/b)^{-1/2}$ and thus the above yields

$$\begin{aligned} 2e^{-2b\log 2} &= e^{-b}, \qquad e^{cy} = e^{(b+2c)y} & (89) \\ de^c &= 2d^2 \frac{\pi}{2a} e^{4c\log 2} b^{-3/2} e^{2b\log 2} \left(2 - \frac{1}{b}\right)^{-1/2}. \end{aligned}$$

We thus have

$$b = \frac{1}{2 - 1/\log 2}, \qquad c = -b, \qquad d = \frac{2a}{\pi} b^{3/2} \sqrt{2 - \frac{1}{b}} \qquad (90)$$

and a remains arbitrary. But then (88) is equivalent to (81) if we identify a as

$$a = \frac{2 - 1/\log 2}{e} \exp\left[\frac{\beta}{2 - 1/\log 2}\right]. \qquad (91)$$

Next we consider $y \to +\infty$ and write the equation for $P(y)$ in the form

$$P(y+1) = 2\int_0^{1/2} \frac{1}{1-x} \int_{-\infty}^{\infty} P(\eta) P\left(y + \frac{x(y-\eta) - 2\phi(x)}{1-x}\right) d\eta dx. \tag{92}$$

We seek solutions in the WKB form

$$P(y) \sim G(y) e^{F(y)}, \qquad y \to +\infty \tag{93}$$

where $\log G$ varies less rapidly than F. Then (92) may be approximated by

$$e^{F(y)} e^{F'(y)} [G(y) + G'(y) + \cdots] \left[1 + \frac{1}{2} F''(y) + \cdots\right] = \tag{94}$$

$$\begin{aligned} &= 2\int_0^{1/2} \frac{1}{1-x} \int_{-\infty}^{\infty} P(\eta) e^{F(y)} \exp\left\{ \left[\frac{x(y-\eta)}{1-x} - \frac{2\phi(x)}{1-x}\right] F'(y) \right\} \\ &\times \left[G(y) + \frac{x(y-\eta) - 2\phi(x)}{1-x} G'(y) + \cdots\right] \left[1 + \frac{1}{2}\left(\frac{x(y-\eta) - 2\phi(x)}{1-x}\right)^2 F''(y) + \cdots\right] d\eta dx. \end{aligned}$$

After scaling $x = \omega/[-yF'(y)]$, the leading term in the right side of (94) is $e^{F(y)} G(y)(-2)/(yF'(y))$ and hence

$$e^{F'(y)} = -\frac{2}{yF'(y)}. \tag{95}$$

After some calculation, we obtain the following linear differential equation for G:

$$G' + \frac{1}{2} F'' G = \frac{G}{yF'} - \frac{2G}{y}[\gamma + \log(-yF')] - \frac{G'}{F'} + \frac{F''}{(F')^2} G. \tag{96}$$

Equation (95) is a first order non-linear ordinary differential equation for F, which is really just a transcendental equation for F'. Setting $F'(y) \equiv -w_*(y)$ we have

$$\begin{aligned} F(y) = -\int^y w_*(t) dt &= -\int^{w_*(y)} 2\zeta \left(\frac{e^\zeta}{\zeta} - \frac{e^\zeta}{\zeta^2}\right) d\zeta \\ &= -y w_* + 2\int^{w_*} \frac{e^\zeta}{\zeta} d\zeta \end{aligned} \tag{97}$$

where we have used (95) to change the variable of integration. But then the dominant exponential terms in (83) and (93) agree precisely. To solve (96) we change variables $y \leftrightarrow w_*$ and write (96) as

$$\frac{G'(y)}{G(y)} = -\frac{F''(y)}{G} \frac{dG}{dw_*} = \left(1 + \frac{1}{F'}\right)^{-1} \left[\frac{F''}{(F')^2} - \frac{1}{2} F'' + \frac{1}{yF'} + \frac{2F'}{y} - \frac{2(\gamma + \log 2)}{y}\right] \tag{98}$$

where we have used $\log(-yF') = \log(2e^{-F'}) = \log 2 - F'$. From (95) we get

$$e^{w_*}\left(\frac{1}{w_*} - \frac{1}{w_*^2}\right)(-F''(y)) = \frac{1}{2}. \tag{99}$$

Using (99) in (98) we obtain, after some calculation,

$$\frac{1}{G}\frac{dG}{dw_*} = \frac{1}{2} - \frac{1}{2}\frac{1}{w_* - 1} - 2w_* - 2(\gamma + \log 2). \tag{100}$$

Solving (100) we see that (93) is equivalent to (83), noting that $\sqrt{y} = \sqrt{2/w_*}\exp(w_*/2)$. Of course, G is determined only up to a multiplicative constant, which corresponds to $C_*/(2\sqrt{\pi})$ in (83).

References

[1] Abramowitz, M. and Stegun, I., eds. *Handbook of Mathematical Functions*. Wiley, New York, 1962.

[2] C. Bender and S. Orszag, *Advanced Mathematical Methods for Scientists and Engineers*, McGrew-Hill, New York 1978.

[3] M. Cramer, A Note Concerning the Limit Distribution of the Quicksort Algorithm, *Theoretical Informatics and Applications*, 30, 195–207, 1996.

[4] L. Devroye, A Note on the Height of binary Search Trees, *J. ACM*, 489–498, 1986.

[5] W. Eddy and M. Schervish, How Many Comparisons Does Quicksort Use, *J. Algorithms*, 19, 402–431, 1995.

[6] N. Froman and P. Froman, *JWKB Approximation*, North-Holland, Amsterdam 1965.

[7] P. Hennequin, Combinatorial Analysis of Quicksort Algorithm, *Theoretical Informatics and Applications*, 23, 317–333, 1989.

[8] P. Henrici, *Applied and Computational Complex Analysis*, John Wiley&Sons, New York 1977.

[9] C.A.R. Hoare, Quicksort, *Comput. J.*, 5, 10–15, 1962.

[10] P. Jacquet and W. Szpankowski, Asymptotic Behavior of the Lempel-Ziv Parsing Scheme and Digital Search Trees, *Theoretical Computer Science*, 144, 161-197, 1995.

[11] D. Knuth, *The Art of Computer Programming. Sorting and Searching,* Addison-Wesley (1973).

[12] C.J. McDiarmid and R. Hayward, Large Deviations for Quicksort, *J. Algorithms*, 21, 476–507, 1996.

[13] H. Mahmoud, *Evolution of Random Search Trees*, John Wiley & Sons, New York (1992).

[14] M. Régnier, A Limiting Distribution for Quicksort, *Theoretical Informatics and Applications*, 23, 335–343, 1989.

[15] U. Rösler, A Limit Theorem for Quicksort, *Theoretical Informatics and Applications*, 25, 85–100, 1991.

[16] U. Rösler, A Fixed Point Theorem for Distributions, *Stochastic Processes and Their Applications*, 42, 195–214, 1992.

[17] K.H. Tan and P. Hadjicostas, Some Properties of a Limiting Distribution in Quicksort, *Statistics & Probability Letters*, 25, 87–94, 1995.