Total path length for random recursive trees

[short title: Total path length]

BY ROBERT P. DOBROW AND JAMES ALLEN FILL[1]

Truman State University and
The Johns Hopkins University

September 27, 2004

Abstract

Total path length, or search cost, for a rooted tree is defined as the sum of all root-to-node distances. Let T_n be the total path length for a random recursive tree of order n. Mahmoud (1991) showed that $W_n := (T_n - E[T_n])/n$ converges almost surely and in L^2 to a nondegenerate limiting random variable W. Here we give recurrence relations for the moments of W_n and of W and show that W_n converges to W in L^p for each $0 < p < \infty$. We confirm the conjecture that the distribution of W is not normal. We also show that the distribution of W is characterized among all distributions having zero mean and finite variance by the distributional identity

$$W \stackrel{d}{=} U(1+W) + (1-U)W^* - \mathcal{E}(U),$$

where $\mathcal{E}(x) := -x \ln x - (1-x)\ln(1-x)$ is the binary entropy function, U is a uniform$(0,1)$ random variable, W^* and W have the same distribution, and U, W, and W^* are mutually independent. Finally, we derive an approximation for the distribution of W using a Pearson curve density estimator. Simulations exhibit a high degree of accuracy in the approximation.

[1]Research for the first author supported by NSF grant DMS-9626597. Research for the second author supported by NSF grant DMS-9626756.

[2]*AMS 1991 subject classifications.* Primary 05C05, 60C05; secondary 60F25.

[3]*Keywords and phrases.* Recursive trees, total path length, search cost, Pearson curve, cumulants.

1 Introduction and summary

A *recursive* (also called *increasing* or *ordered*) *tree* of order n is a rooted tree on n vertices (or nodes) labeled 1 through n, with the property that for each k such that $2 \leq k \leq n$, the labels of the vertices on the necessarily unique path from the root to the node labeled with k form an increasing sequence. We will refer to the node labeled with k as *node* k. We use familial terms, such as child, parent, and ancestor, to describe relations between nodes. Thus the children of node k are precisely the nodes incident to k with labels greater than k. We do *not* order the children of a given node; thus, for example, we consider there to be only two trees of order 3. When we draw a recursive tree in the plane, we place the root at the top and we arrange the children of each node in increasing order from left to right.

The most well-studied probability model on the space of recursive trees of order n is the uniform model, whereby we posit all $(n-1)!$ recursive trees to be equally likely. We refer the reader to the excellent survey article by Smythe and Mahmoud (1995) for numerous applications and properties of recursive trees.

The distance D_k between the root and node k in a random recursive tree has been studied by many authors, including Moon (1974), Szymański (1990), and Dobrow and Smythe (1996). In this paper we treat the *total path length* of a recursive tree, namely,

$$T_n := \sum_{k=1}^{n} D_k,$$

defined as the sum of all root-to-node distances. This random variable may serve as a global measure of the cost of constructing the tree. The strong dependence among the random variables D_k makes it nontrivial to obtain the exact distribution of T_n.

Knuth (1973) presents extensive material on total path length for (deterministic) binary trees. Takács (1992, 1994) has obtained the asymptotic distribution of total path length for random rooted trees and random rooted binary trees.

Returning to the case of recursive trees, it is apparent that the smallest and largest possible values of T_n are $n-1$ and $\sum_{i=1}^{n}(i-1) = \binom{n}{2}$. The expected values of root-to-node distances in recursive trees are well known and easily derived. Let $H_k := \sum_{i=1}^{k} i^{-1}$ be the kth harmonic number. Then

$E[D_k] = H_{k-1}$. Linearity of expectation gives

$$\mu_n := E[T_n] = \sum_{i=1}^{n-1} H_i = n(H_n - 1),$$

which is asymptotically equivalent to $n \ln n$.

Mahmoud (1991) proved that the sequence (W_n) of normalized random variables

$$W_n := \frac{T_n - \mu_n}{n}$$

is a martingale. He obtained the exact variance of T_n and by an application of the martingale convergence theorem showed that there exists a nondegenerate random variable W such that $W_n \to W$, almost surely and in L^2. Mahmoud showed that the normalized distances $(D_n - \ln n)/\sqrt{\ln n}$ are asymptotically standard normal. It has been conjectured that the distribution of W, however, is not normal.

In this paper we consider the random variables W_n and W. We obtain a recurrence relation for the moments of W [equation (1)] and for the factorial moments of W_n (Theorem 2). We show that W_n converges to W in L^p for each $0 < p < \infty$ (Theorem 1). We calculate the skewness and kurtosis of W and confirm the conjecture on the nonnormality of W (Section 3). We also characterize the distribution of W (Corollary 2.1). Specifically, letting U denote a uniform $(0, 1)$ random variable, we show that

$$W \stackrel{d}{=} U(1 + W) + (1 - U)W^* - \mathcal{E}(U),$$

where $\mathcal{E}(x) := -x \ln x - (1 - x) \ln(1 - x)$ is the binary entropy function, W and W^* have the same distribution, and U, W, and W^* are mutually independent. Finally we use the moments of W to obtain approximations for the distribution of W (Section 4). A Pearson curve density estimator appears to give a very good approximation, as indicated by numerical simulations.

2 Convergence in L^p

In working with random recursive trees, it is often useful to consider a dynamic construction of the tree evolving over discrete units of time. Let X_n denote a random recursive tree of size n. Then X_n can be built from X_{n-1} by adjoining node n as a child of node j, where j is chosen uniformly at random from $\{1, \ldots, n-1\}$. If $\pi(n)$ denotes the (random) parent of node n, observe that the random variables $\pi(1), \pi(2), \ldots$ are mutually independent. It thus follows that if one conditions on the size of the subtrees of X_n (that is, the induced trees, whose respective roots are the children of the root node of X_n), then each of the subtrees of X_n is a random recursive tree (conditioned on its size and with appropriately changed labels). Furthermore, these subtrees are mutually independent.

By conditioning on the size of the subtree rooted at node 2 we obtain our first lemma.

Lemma 2.1 *For $n \geq 2$,*

$$T_n \stackrel{d}{=} K + T_K + T^*_{n-K},$$

where $K \equiv K_n$ is distributed uniformly on $\{1, \ldots, n-1\}$ and the random variables $K, T_1, \ldots, T_{n-1}, T^*_1, \ldots, T^*_{n-1}$ are all mutually independent.

Proof Let K be the size of the subtree rooted at node 2. Then $K + T_K$ accounts for the contribution to total path length from all the nodes in the subtree rooted at node 2, and T^*_{n-K} accounts for the contribution to total path length from all the remaining nodes. The lemma will follow from the fact that in a random recursive tree of order n the size of the subtree rooted at node 2 is distributed uniformly on $\{1, \ldots, n-1\}$.

We give a simple combinatorial proof of the latter claim because it allows us to introduce the bijective correspondence between recursive trees and permutations. Stanley (1986) gives the following mapping. Let $\sigma = (\sigma_1, \ldots, \sigma_{n-1})$ be a permutation on $\{1, \ldots, n-1\}$. Construct a recursive tree with nodes $0, 1, \ldots, n-1$ by making 0 the root and defining the parent of node i to be the rightmost element j of σ which both precedes i and is less than i. If there is no such element j, then define the parent of i to be the root 0. Finally, to convert to a recursive tree on nodes $\{1, \ldots, n\}$, simply add 1 to each label.

For example, the permutation $(1, 2, 3)$ corresponds to the "linear" tree of size 4 where i is the parent of $i+1$ for $i = 1, 2, 3$; the permutation $(3, 2, 1)$ corresponds to the tree where nodes 2, 3, and 4 are each children of the root 1. This mapping is bijective between permutations of $\{1, \ldots, n-1\}$ and recursive trees with label set $\{1, \ldots, n\}$. Note that in this correspondence the size of the subtree rooted at node 2 is one greater than the number of elements in the corresponding permutation of size $n-1$ that succeed 1. This number, in turn, is just n minus the position of 1. The position of 1 is, of course, distributed uniformly on $\{1,\ldots,n-1\}$. ■

Remark: In the context of the above correspondence, total path length in trees corresponds to the following statistic on permutations. Let σ be a permutation on $\{1,\ldots,n\}$. For each $1 \le k \le n$, consider the "greedy" decreasing sequence, starting at σ_k and moving right to left, and count the number of elements in that sequence. The sum of these counts corresponds to total path length. Thus, for instance, $(1,2,3)$ gives a count of $1+2+3=6$ while $(1,3,2)$ gives a count of $1+2+2=5$. This statistic gives a measure of how "close" a permutation is to the identity permutation.

Theorem 1 *Let T_n denote total path length in a random recursive tree on n nodes, with $\mu_n := E[T_n]$. Let*

$$W_n := \frac{T_n - \mu_n}{n}$$

and let W be the almost sure limit of W_n as $n \to \infty$.

(i) For any (real) $0 < p < \infty$, $W_n \xrightarrow{L^p} W$. For integer $p \ge 1$, $E[W_n^p] \to E[W^p] \in (-\infty, \infty)$ as $n \to \infty$.

(ii) Letting $\nu(p) := E[W^p]$ for integer $p \ge 1$, we have the recurrence relation

$$\nu(p) = \int_0^1 \sum_{h,i,j,l} \binom{p}{h,i,j,l} x^{h+i} \nu(i)(1-x)^j \nu(j)(-\mathcal{E}(x))^l \, dx, \qquad (1)$$

where $\mathcal{E}(x) := -x\ln x - (1-x)\ln(1-x)$ is the binary entropy function and

$$\binom{p}{h,i,j,l} := \frac{p!}{h!\ i!\ j!\ l!},$$

where $h+i+j+l=p$.

Proof It follows immedately from Lemma 2.1 and simple algebra that for $n \ge 2$,

$$W_n \stackrel{d}{=} \frac{K}{n}(1+W_K) + \left(1 - \frac{K}{n}\right) W^*_{n-K} - z_{n,K}, \qquad (2)$$

where K is uniformly distributed on $\{1, \ldots, n-1\}$ and the random variables $K, W_1, \ldots, W_{n-1}, W_1^*, \ldots, W_{n-1}^*$ are all mutually independent and, with $\tilde{\mu}_n := \mu_n/n$ for $n \geq 1$,

$$z_{n,k} := \tilde{\mu}_n - \left[\frac{k}{n}\tilde{\mu}_k + \left(1 - \frac{k}{n}\right)\tilde{\mu}_{n-k}\right] = H_n - \left[\frac{k}{n}H_k + \left(1 - \frac{k}{n}\right)H_{n-k}\right].$$

For integer $p \geq 0$ and $n \geq 1$, let $\nu_n(p) := E[W_n^p]$. Observe $\nu_1(p) = 0$ for all $p \geq 1$, and $\nu_n(0) = 1$ and $\nu_n(1) = 0$ for all $n \geq 1$. For integer $p \geq 0$ and $n \geq 2$ we have

$$\begin{aligned}
\nu_n(p) &= E[W_n^p] = E[E[W_n^p|K]] \\
&= \frac{1}{n-1}\sum_{k=1}^{n-1} E\left[\frac{k}{n}(1+W_k) + \left(1-\frac{k}{n}\right)W_{n-k}^* - z_{n,k}\right]^p \\
&= \frac{1}{n-1}\sum_{k=1}^{n-1}\sum_{h,i,j,l}\binom{p}{h,i,j,l}\left(\frac{k}{n}\right)^{h+i}\nu_k(i)\left(1-\frac{k}{n}\right)^j \nu_{n-k}(j)(-z_{n,k})^l. \quad (3)
\end{aligned}$$

We claim that for each integer $p \geq 1$, $\nu_n(p)$ converges to a finite limit $\nu(p)$ as $n \to \infty$. We prove the claim by induction on p. The base case $p = 1$ is trivial with $\nu(1) := 0$. Now by (3), for $n \geq 2$ we have

$$\begin{aligned}
\nu_n(p) &= \frac{1}{n-1}\sum_{k=1}^{n-1}\left(\frac{k}{n}\right)^p \nu_k(p) + \frac{1}{n-1}\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)^p \nu_{n-k}(p) \\
&\quad + \frac{1}{n-1}\sum_{k=1}^{n-1}\sum_{\substack{h,i,j,l\\ i,j\neq p}}\binom{p}{h,i,j,l}\left(\frac{k}{n}\right)^{h+i}\nu_k(i)\left(1-\frac{k}{n}\right)^j \nu_{n-k}(j)(-z_{n,k})^l \\
&= \frac{2}{n-1}\sum_{k=1}^{n-1}\left(\frac{k}{n}\right)^p \nu_k(p) \\
&\quad + \frac{1}{n-1}\sum_{k=1}^{n-1}\sum_{\substack{h,i,j,l\\ i,j\neq p}}\binom{p}{h,i,j,l}\left(\frac{k}{n}\right)^{h+i}\nu_k(i)\left(1-\frac{k}{n}\right)^j \nu_{n-k}(j)(-z_{n,k})^l \quad (4) \\
&=: \frac{2}{n-1}\sum_{k=1}^{n-1}\left(\frac{k}{n}\right)^p \nu_k(p) + B_n(p). \quad (5)
\end{aligned}$$

Letting

$$x_n(p) := (n+1)^p \nu_{n+1}(p), \quad n \geq 0,$$

and

$$a_n(p) := (n+1)^p B_{n+1}(p), \quad n \ge 1,$$

we transform (5) into the equivalent recurrence relation

$$x_n(p) = a_n(p) + \frac{2}{n} \sum_{k=0}^{n-1} x_k(p), \quad n \ge 1.$$

This simple and well-studied recurrence is solved explicitly in Lemma 4.3 in Fill (1996). The unique solution is, for arbitrarily defined $a_0(p)$,

$$x_n(p) = a_n(p) + (n+1) \left[x_0(p) - a_0(p) + 2 \sum_{k=0}^{n-1} \frac{a_k(p)}{(k+1)(k+2)} \right], \quad n \ge 0.$$

Using $x_0(p) = \nu_1(p) = 0$ and defining $a_0(p) := 0$, this gives (with $B_1(p) := 0$)

$$\begin{aligned} \nu_n(p) &= B_n(p) + 2n \sum_{k=1}^{n-1} \left(\frac{k}{n}\right)^p \frac{B_k(p)}{k(k+1)} \qquad (6) \\ &= B_n(p) + 2 \sum_{k=1}^{n-1} \frac{1}{n-1} \left(\frac{k}{n}\right)^{p-2} \frac{B_k(p)}{1+\frac{1}{k}}, \quad n \ge 1. \end{aligned}$$

We now argue by (strong) induction on $p \ge 0$ that each $(\nu_n(p))_{n \ge 0}$ is a convergent, and hence bounded, sequence. For the basis, $\nu_n(0) \equiv 1$ and $\nu_n(1) \equiv 0$. For the induction step, for $p \ge 2$ we note from elementary arguments together with the induction hypothesis that

$$\lim_{n \to \infty} B_n(p) = \int_0^1 \sum_{\substack{h,i,j,l \\ i,j \ne p}} \binom{p}{h,i,j,l} x^{h+i} \lambda(i) (1-x)^j \lambda(j) (-\mathcal{E}(x))^l \, dx =: B(p),$$

where $\lambda(i) := \lim_{n \to \infty} \nu_n(i)$ (for $i < p$). Now the explicit solution (6) yields the existence of $\lambda(p) := \lim_{n \to \infty} \nu_n(p)$, with value

$$\lambda(p) = B(p) \left[1 + 2 \int_0^1 x^{p-2} \, dx \right] = \frac{p+1}{p-1} B(p).$$

Rearranging gives (1), but with ν replaced by λ.

The above work demonstrates that $\sup_n E|W_n|^p$ is finite for any (real) $0 < p < \infty$. It follows from Exercise 4.5.8 in Chung (1974) that $(|W_n|^p)$ is uniformly integrable for any $0 < p < \infty$. The first assertion in part (i) of

Theorem 1 now follows from Theorem 4.5.4 in Chung. The second assertion follows immedately from Theorem 4.5.2 in Chung. Thus $\lambda(p) = \nu(p)$ for integer $p \geq 1$, completing the proof of part (ii). ■

Corollary 2.1 *In the notation above, the following identity characterizes the distribution of W among all distributions having zero mean and finite variance:*

$$W \stackrel{d}{=} U(1+W) + (1-U)W^* - \mathcal{E}(U), \tag{7}$$

where U is a random variable distributed uniformly on $(0,1)$, W^* has the same distribution as W, and U, W, and W^* are mutually independent. Furthermore, the distribution of W is absolutely continuous, possessing a density f that is positive everywhere and satisfies

$$f(t) = \int_0^1 \int_{-\infty}^{\infty} \frac{1}{u} f\left(\frac{t - u + \mathcal{E}(u) - (1-u)w}{u}\right) f(w)\, dw\, du \tag{8}$$

for Lebesgue almost every t.

Proof Take characteristic functions in (2). Now (7) follows routinely using the convergence and uniqueness theorems [e.g., Theorems 6.3.1 and 6.2.2 in Chung (1974)] for characteristic functions. The issue now is whether there could be more than one distribution with zero mean and finite variance that satisfies (7). To show that (7) characterizes the distribution of W we refer to analogous work in the analysis of the asymptotic run time distribution of the well-known Quicksort sorting algorithm invented by Hoare (1962).

Let X_n be the (random) number of comparisons needed to sort a list of length n by Quicksort. Régnier (1989) and Rösler (1991) showed that a normalized version of X_n converges in suitable senses to a limiting random variable X. Rösler also showed that the distribution of this limit satisfies

$$X \stackrel{d}{=} UX + (1-U)X^* - \mathcal{G}(U), \tag{9}$$

where $\mathcal{G}(x) := 2\mathcal{E}(x) - 1$, U is uniformly distributed on $(0,1)$, X^* and X have the same distribution, and U, X, and X^* are mutually independent. Note the similarity between (9) and (7). Rösler's arguments that there is a *unique* distribution with $EX = 0$ and $EX^2 < \infty$ satisfying (9) carry over to our (7). Tan and Hadjicostas (1995) used (9) to prove that X is absolutely continuous with an everywhere positive density; their calculations, too, are easily adapted to our (7). Thus our W has an everywhere positive density f. Now elementary arguments show that f satisfies (8) for Lebesgue almost every $t \in \mathbb{R}$. ■

Remark 1: Adapting Rösler's techniques, we can also show that

$$E\left|e^{\lambda W_n} - e^{\lambda W}\right| \to 0 \quad \text{as} \quad n \to \infty$$

for every $\lambda \in \mathbb{R}$, and in particular that $Ee^{\lambda W_n} \to Ee^{\lambda W} < \infty$. This enables Chernoff bounds for large deviations; cf. Rösler (1991). Also, just as Rösler does, we can obtain an infinite series representation of W; we omit the details.

Remark 2: Since the support of the distribution of W is the entire real line, it follows that (W_n) does *not* converge in L^∞.

Remark 3: It is well known that the analysis of the number of comparisons required by the Quicksort algorithm is equivalent to the analysis of total path length for a binary search tree [cf. Knuth (1973)].

3 Computing the moments

Formula (1) appears to be ill-suited for exact (as opposed to numerical) computation of the moments of W due both to the complexity of integrating powers of the entropy function and the rapid growth of the number of summands in (1) as a function of p. In this section we derive a new recurrence relation which we are able to use to compute the exact moments of W.

We use the falling factorial notation $x^{\underline{p}} := x(x-1)\cdots(x-p+1)$ for $p \geq 1$, writing $x^{\underline{0}} := 1$.

Theorem 2 *For integer $p \geq 1$, let*

$$R_p(w) := \sum_{n=1}^{\infty} E\left[T_n^{\underline{p}}\right] w^{n-1}$$

and

$$S_p(w) := \int_0^w R_p(x)\,dx = \sum_{n=1}^{\infty} E\left[T_n^{\underline{p}}\right] \frac{w^n}{n}.$$

Then

$$S_p(w) = \frac{1}{1-w} \sum \binom{p-1}{i} \binom{p-i}{j} \int_0^w S_i^{(1)}(x) x^j (1-x) S_{p-i-j}^{(j)}(x)\, dx,$$

where $S_i^{(j)}$ denotes the jth derivative of S_i and where the sum is over all pairs $(i,j) \neq (0,0)$ of nonnegative integers i and j satisfying $i+j \leq p$.

Proof Define

$$H(s,w) := \sum_{n=1}^{\infty} E[s^{T_n}] \frac{w^n}{n}. \tag{10}$$

Write $\varphi_n(s) := E[s^{T_n}]$. Then Lemma 2.1 says precisely

$$(n-1)\varphi_n(s) = \sum_{k=1}^{n-1} s^k \varphi_k(s) \varphi_{n-k}(s), \quad n \geq 2.$$

Multiply both sides by w^n and sum to get

$$w^2 \frac{\partial^2}{\partial w^2} H(s,w) = \left[(sw) \, \frac{\partial}{\partial v} H(s,v) \Big|_{v=sw} \right] \left[w \frac{\partial}{\partial w} H(s,w) \right],$$

which can be rearranged to

$$\frac{\partial}{\partial w} \log \left[\frac{\partial}{\partial w} H(s,w) \right] = \frac{\partial}{\partial w} H(s,sw).$$

Since both $\log[\frac{\partial}{\partial w} H(s,w)]$ and $H(s,sw)$ vanish at $w=0$, this is equivalent to

$$\log \left[\frac{\partial}{\partial w} H(s,w) \right] = H(s,sw)$$

or to

$$\frac{\partial}{\partial w} H(s,w) = \exp\left(H(s,sw)\right). \tag{11}$$

Differentiating (11) with respect to s gives

$$\frac{\partial^2}{\partial s \, \partial w} H(s,w) = \left[\frac{\partial}{\partial w} H(s,w) \right] \left[\frac{\partial}{\partial s} H(s,sw) \right]. \tag{12}$$

From (10) we have

$$\frac{\partial^{p+1}}{\partial s^p \, \partial w} H(s,w) \Big|_{s=1} = \sum_{n=1}^{\infty} E\left[T_n^p\right] w^{n-1} = R_p(w). \tag{13}$$

Also,

$$\frac{\partial^p}{\partial s^p} H(s, sw)\bigg|_{s=1} = \sum_{i=0}^{p} \binom{p}{i} w^i S_{p-i}^{(i)}(w),$$

noting that

$$S_k(w) = \frac{\partial^k}{\partial s^k} H(s, w)\bigg|_{s=1}.$$

Putting this together with (12) and (13) we have, for $p \geq 1$,

$$\begin{aligned} R_p(w) &= \frac{\partial^{p-1}}{\partial s^{p-1}} \left[\left(\frac{\partial}{\partial w} H(s, w) \right) \left(\frac{\partial}{\partial s} H(s, sw) \right) \right] \bigg|_{s=1} \\ &= \sum_{i=0}^{p-1} \binom{p-1}{i} S_i^{(1)}(w) \sum_{j=0}^{p-i} \binom{p-i}{j} w^j S_{p-i-j}^{(j)}(w), \end{aligned} \tag{14}$$

with

$$R_0(w) = \frac{1}{1-w} \quad \text{and} \quad S_0(w) = \ln\left(\frac{1}{1-w}\right).$$

We reexpress (14) as

$$\begin{aligned} S_p'(w) &= R_0(w) S_p(w) + R_0(w) \sum_{j=1}^{p} \binom{p}{j} w^j S_{p-j}^{(j)}(w) \\ &+ \sum_{i=1}^{p-1} \binom{p-1}{i} S_i^{(1)}(w) \sum_{j=0}^{p-i} \binom{p-i}{j} w^j S_{p-i-j}^{(j)}(w), \end{aligned}$$

which is a first order linear differential equation in the unknown function S_p. This admits a direct solution as given in the statement of the theorem. ■

Remark 1: All formal operations (such as interchange of derivative and sum) performed in the proof of Theorem 2 are easily justified using the finiteness of moment generating functions as discussed in Remark 1 following Corollary 2.1.

Remark 2: The derivation of the recurrence relation (14) is similar to that in Takács (1992), who obtains the asymptotic growth rate for the moments of total path length for random rooted trees. By the method of moments he obtains the asymptotic distribution function for total path length for these trees.

p	$S_p(w) \qquad \left(\text{with } L := \ln\left(\frac{1}{1-w}\right)\right)$
0	L
1	$(1-w)^{-1}(L-1)$ $+1$
2	$(1-w)^{-2}(L^2+1)$ $-(1-w)^{-1}(L^2+2L)$ -1
3	$(1-w)^{-3}(2L^3+3L^2+6L+(5/2))$ $-(1-w)^{-2}(3L^3+18L^2+12L+12)$ $+(1-w)^{-1}(L^3+9L^2+15L+(15/2))$ $+2$
4	$(1-w)^{-4}(6L^4+20L^3+48L^2+54L+(86/3))$ $-(1-w)^{-3}(6L^4+72L^3+156L^2+222L+114)$ $+(1-w)^{-2}(L^4+72L^3+216L^2+282L+204)$ $-(1-w)^{-1}(20L^3+108L^2+182L+(338/3))$ -6

Table 1. Generating function for S_p

The recurrence relation in Theorem 2 is easily implemented by a program such as `Mathematica`. Table 1 gives the solution to the recurrence for small values of p. As suggested by this table, the function S_p has the following form for $p \geq 1$:

Proposition 3.1 *For $p \geq 1$, there exist rational constants $b_p(\alpha,\beta)$, $1 \leq \alpha \leq p$ and $0 \leq \beta \leq p$, such that*

$$S_p(w) = (-1)^{p-1}(p-1)!f_{0,0}(w) + \sum_{\alpha=1}^{p}\sum_{\beta=0}^{p} b_p(\alpha,\beta)f_{\alpha,\beta}(w),$$

where, for integers α and β,

$$f_{\alpha,\beta}(w) := \frac{1}{(1-w)^{\alpha}}\left(\ln\frac{1}{1-w}\right)^{\beta}.$$

The proof, which we omit, is via straightforward, but rather laborious, (strong) induction on p. It is a direct consequence of the following two lemmas.

Lemma 3.1 *Let $\alpha, \beta \in \mathbb{Z}$ and $j \in \{0, 1, 2, \ldots\}$. Then*

$$f^{(j)}_{\alpha,\beta}(x) = \sum_{l=0}^{j} \sigma(\alpha, j, l)\beta^{\underline{l}} f_{\alpha+j,\beta-l}(x),$$

where

$$\sigma(\alpha, j, l) := [x^l]\{(x + \alpha + j - 1)^{\underline{j}}\}.$$

Lemma 3.1 follows by a simple induction on j.

Lemma 3.2 *Let $\alpha \in \mathbb{Z}$ and $\beta \in \{0, 1, 2, \ldots\}$ and define*

$$I_{\alpha,\beta}(w) := \int_0^w f_{\alpha,\beta}(x)\, dx.$$

(a) For $\alpha \neq 1$, we have

$$I_{\alpha,\beta}(w) = (\alpha - 1)^{-1}\beta! \left[\sum_{l=0}^{\beta} \frac{\left(-\frac{1}{\alpha-1}\right)^{\beta-l}}{l!} f_{\alpha-1,l}(w) - \left(-\frac{1}{\alpha - 1}\right)^{\beta} f_{0,0}(w) \right].$$

(b) For $\alpha = 1$, we have

$$I_{1,\beta}(w) = (\beta + 1)^{-1} f_{0,\beta+1}(w).$$

Lemma 3.2 is proved by fixing α and w and treating the exponential generating function of the sequence $I_{\alpha,\beta}(w)$ indexed by β.

Having established Proposition 3.1, we will now proceed to derive an expression for the moments of W. The main tool we will use to estimate the asymptotic growth of the coefficients in S_p is the following result from Flajolet and Odlyzko (1990), which we have narrowed somewhat for our purposes.

Lemma 3.3 (Flajolet and Odlyzko) *Let*

$$f_{\alpha,\beta}(w) \equiv f(w) = \frac{1}{(1 - w)^{\alpha}} \left(\ln \frac{1}{1 - w} \right)^{\beta},$$

where α is a positive integer and β is a nonnegative integer. The coefficient of w^n in $f(w)$, denoted $[w^n]f(w)$, admits the asymptotic expansion

$$[w^n]f(w) = \frac{n^{\alpha-1}}{(\alpha - 1)!}(\ln n)^{\beta} \left[1 + \sum_{k=1}^{\beta} \binom{\beta}{k} \frac{G_{\alpha,k}}{(\ln n)^k} + O\left(n^{\alpha-2}(\ln n)^{\beta}\right) \right], \quad (15)$$

where

$$G_{\alpha,k} = (\alpha-1)!\,\frac{d^k}{dx^k}\frac{1}{\Gamma(x)}\bigg|_{x=\alpha}$$

and $\Gamma(\cdot)$ *is the gamma function.*

Theorem 3 *Using the notation from Lemma 3.3, and setting* $G_{p,0} := 1$ *for all* p *and* $b_{p,j} := b_p(p,j)$,

$$E[W^p] = (1-\gamma)^p + \sum_{i=1}^{p}\binom{p}{i}(1-\gamma)^{p-i}\sum_{j=0}^{i} b_{i,j}\frac{G_{i,j}}{(i-1)!},$$

where γ *is Euler's constant.*

Proof The result is evidently correct for $p=0$. Fix $p \geq 1$. First observe that the pth falling factorial moment $E\left[T_n^{\underline{p}}\right]$ is just n times the coefficient of w^n in $S_p(w)$. From Proposition 3.1 and (15), for $n \geq 1$ we have

$$\begin{aligned}
E[T_n^{\underline{p}}] &= n[w^n]S_p(w)\\
&= n[w^n]\left\{(-1)^{p-1}(p-1)! + \sum_{\alpha=1}^{p}\sum_{\beta=0}^{p} b_p(\alpha,\beta)f_{\alpha,\beta}(w)\right\}\\
&= n\sum_{\alpha=1}^{p}\sum_{\beta=0}^{p} b_p(\alpha,\beta)\\
&\qquad \times\left\{\frac{n^{\alpha-1}}{(\alpha-1)!}(\ln n)^{\beta}\left[\sum_{k=0}^{\beta}\binom{\beta}{k}\frac{G_{\alpha,k}}{(\ln n)^k}\right] + O\left(n^{\alpha-2}(\ln n)^{\beta}\right)\right\}\\
&= \frac{n^p}{(p-1)!}\sum_{\beta=0}^{p} b_p(p,\beta)(\ln n)^{\beta}\left[\sum_{k=0}^{\beta}\binom{\beta}{k}\frac{G_{p,k}}{(\ln n)^k}\right] + O\left(n^{p-1}(\ln n)^p\right)\\
&= \frac{n^p}{(p-1)!}\sum_{i=0}^{p}(\ln n)^i\sum_{j=0}^{i} b_p(p,i+j)\binom{i+j}{j}G_{p,j} + O\left(n^{p-1}(\ln n)^p\right).
\end{aligned}$$

Since $E[T_n^p]$ is a fixed linear combination of $E[T_n^{\underline{l}}], l = 0,\ldots,p$, with coefficient of 1 for $E[T_n^{\underline{p}}]$, it now follows that

$$E[T_n^p] = \frac{n^p}{(p-1)!}\sum_{i=0}^{p}(\ln n)^i\sum_{j=0}^{i} b_p(p,i+j)\binom{i+j}{j}G_{p,j} + O\left(n^{p-1}(\ln n)^p\right).$$

Also,

$$\begin{aligned} E[(T_n - \mu_n)^p] &= E[(T_n - n(\ln n - (1-\gamma) + \epsilon_n))^p] \\ &= \sum_{t=0}^{p} \binom{p}{t} E[T_n^t](-1)^{p-t} n^{p-t} (\ln n - (1-\gamma) + \epsilon_n)^{p-t}, \end{aligned}$$

where $\epsilon_n = O(1/n)$. Now substitute the asymptotic expression for $E[T_n^t]$. Straightforward manipulation of sums gives that the asymptotic coefficient of $(\ln n)^z$ in $E[W_n^p]$ is given by

$$[(\ln n)^z]E[W_n^p] = (-1)^p \binom{p}{z} [-(1-\gamma)]^{p-z} \tag{16}$$

$$+ \sum_{t=1}^{p} \sum_{i=[z-(p-t)]\vee 0}^{t\wedge z} \sum_{j=0}^{t-i} \binom{p}{t} (-1)^{p-t} b_{t,i+j}$$

$$\times \binom{i+j}{j} \frac{G_{t,j}}{(t-1)!} \binom{p-t}{z-i} [-(1-\gamma)]^{p-t-z+i}. \tag{17}$$

We proved, however, in Theorem 1 that $E[W_n^p]$ does in fact converge to $E[W^p]$. Thus, as $n \to \infty$, all of the coefficients above must vanish, except for the case $z = 0$. Substituting $z = 0$ gives the result. ∎

Using the recurrence relation for $S_p(w)$ and Theorem 3, we computed the moments $E[W^p]$ exactly for values of p up through 10. The values up through $p = 4$ are the fairly simple expressions displayed in Table 2. [Here $\zeta(k) := \sum_{i=1}^{\infty} i^{-k}$ denotes the Riemann zeta function.] However, the expressions grow very rapidly in size and complexity as p increases. For example,

$$\begin{aligned} &E[W^9] \\ &= (-2636007715410971/11113200000) + (2094155063\pi^2/108000) \\ &\quad -(10549\pi^4/72) + (45\pi^6/8) \\ &\quad +((37133299/450) - (23450\pi^2/3) + 84\pi^4 - 5\pi^6)\zeta(3) \\ &\quad -7560(\zeta(3))^2 + 2240(\zeta(3))^3 + (56280 - 6048\pi^2 + (252\pi^4/5))\zeta(5) \\ &\quad +(51840 - 4320\pi^2)\zeta(7) + 40320\zeta(9). \end{aligned}$$

p	$E[W^p]$ Exact	$E[W^p]$	$E[W^p]/(SD[W])^p$
1	0	0	0
2	$2-(\pi^2/6)$	.3550659	1
3	$(-9/4)+2\zeta(3)$	.1541138	.728414
4	$(335/18)-2\pi^2+(\pi^4/60)$	.4953872	3.929404

Table 2. Moments of W

Then we used standard formulas for the relations between moments and cumulants to compute the cumulants κ_p for W; these are listed in Table 3, with even powers of π converted to values of $\zeta(\cdot)$. The expressions for the cumulants are very much simpler than those for the moments. Since κ_3 and κ_4 do not vanish, we establish the conjecture that the distribution of W is not normal. Indeed, $\kappa_3 > 0$ indicates that the distribution of W is skewed to the right, and $\kappa_4 > 0$ suggests that the distribution of $W/(SD[W])$ is more peaked about the mode than is the standard normal.

One natural conjecture that arises immediately from Table 3 is that

$$\kappa_p = (-1)^p(c_p - (p-1)!\zeta(p)) \quad \text{for all } p \geq 2$$

where the constants c_p are all rational. This conjecture is correct, but the proof is by no means trivial; we checked the result using calculations much like those in Hennequin (1991) for Quicksort but omit the details here. We have not investigated any other natural conjectures, such as that for all $p \geq 2$ we have $c_p > 0$ and $\kappa_p > 0$.

p	κ_p Exact	κ_p
1	0	.0000000000
2	$2-\zeta(2)$	.3550659332
3	$-((9/4)-2\zeta(3))$	.1541138063
4	$(119/18)-6\zeta(4)$	.1171717088
5	$-((2675/108)-24\zeta(5))$	.1177476049
6	$(1320007/10800)-120\zeta(6)$	.1417029322
7	$-((470330063/648000)-720\zeta(7))$	.1934812582
8	$(1205187829669/238140000)-5040\zeta(8)$	.2875719321
9	$-((448979194501571/11113200000)-40320\zeta(9))$	.4461936608
10	$(9419145105819623/25930800000)-362880\zeta(10)$	.6818111319

Table 3. Cumulants of W

4 Approximating the distribution of W

We would like to utilize our knowledge of the moments to obtain an approximation to the distribution of W. In this section we obtain a Pearson curve density estimator for the standardized

$$W^* := \frac{W}{SD[W]} = \frac{W}{\sqrt{2 - (\pi^2/6)}},$$

based on the first four moments of W, to approximate the underlying distribution. Comparisons with numerical simulations indicate a good degree of accuracy in the estimation.

Pearson curves, introduced by Karl Pearson, are probability densities parametrized by the first four moments of the underlying distribution [cf. Kendall and Stuart (1963)]. We refer the reader to Solomon and Stephens (1980) for a modern treatment of the use of Pearson curves. They consider a variety of problems in geometric probability where the underlying distribution is intractable but the first few moments can be computed theoretically.

In classical notation, let μ_k denote the kth (central) moment of W. Then the key "shape" parameters in the Pearson curve construction are

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \left(\frac{(-9/4) + 2\zeta(3)}{(2 - (\pi^2/6))^{3/2}} \right)^2 = .530586\ldots$$

and

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{(335/18) - 2\pi^2 + (\pi^4/60)}{(2 - (\pi^2/6))^2} = 3.929404\ldots.$$

In the language of Pearson curve estimators, this gives a Type VI curve with density estimator

$$f(x) = N \left(1 + \frac{x}{A_1}\right)^{-q_1} \left(1 + \frac{x}{A_2}\right)^{q_2}, \qquad -3.41597 < x < \infty,$$

where $N = .400366$, $A_1 = 15.4849$, $A_2 = 3.41597$, $q_1 = 70.1506$, and $q_2 = 14.2547$. [See Elderton and Johnson (1969) for an exhaustive treatment on fitting Pearson curves.]

One obvious drawback of this estimator is that while the support of W is the entire real line, the support of f is not.

Another classical method for obtaining density estimators from moments is to use orthogonal polynomials. We fitted a Gram–Charlier curve using

Hermite polynomials and four moments. One drawback here is that the resulting curve need not be a density function, and in fact using more than four moments resulted in a poor estimate for a density. In Table 3 we include the results for the Gram–Charlier curve with four moments. Note that the Pearson curve estimator appears to give a much better fit. In fact, for most rows in the table, the Pearson curve agrees with the simulation to two significant digits, which is about the best one could expect in a simulation of 10,000. Our simulation was run with $n = 10,000$ and $100,000$ trials. Roughly, one would expect no better than agreement to 2 significant digits since $1/\sqrt{100,000} > .001$. We also include the standard normal distribution function, denoted Φ.

In Figure 1 and 2 we give the Pearson curve plot and a histogram of a simulation of normalized total path length with $n = 10,000$.

x	Gram-Charlier	Pearson curve	Simulation	$\Phi(x)$
$-.3.25$	.000152717	1.31532×10^{-14}	0	.000577025
-3.00	.000134841	4.27148×10^{-9}	0	.00134990
-2.75	.000153281	1.53857×10^{-6}	.00001	.00297976
-2.50	.000552937	.0000563928	.00015	.00620967
-2.25	.00227448	.00654813	.00093	.0122245
-2.00	.00726788	.00382581	.00460	.0227501
-1.75	.0188214	.0142212	.01509	.0400592
-1.50	.415100	.0385147	.03877	.0668072
-1.25	.0804674	.082779	.08214	.105650
-1.00	.139915	.149545	.14803	.158655
-0.75	.221303	.236476	.23630	.226627
-0.50	.321847	.337183	.33832	.308538
-0.25	.434306	.443352	.44489	.401294
0.00	.548432	.546974	.54834	.500000
0.25	.653711	.641866	.64319	.598706
0.50	.742265	.724261	.72461	.691462
0.75	.810686	.792679	.79314	.773373
1.00	.860085	.847383	.84748	.841345
1.25	.894587	.889738	.88882	.894350
1.50	.919181	.921644	.92059	.933193
1.75	.937972	.945121	.94466	.959941
2.00	.953404	.962049	.96156	.977250
2.25	.966418	.974046	.97327	.987776
2.50	.977103	.982421	.98182	.993790
2.75	.985357	.988194	.98783	.997020
3.00	.991257	.992128	.99191	.998650
3.25	.995136	.994784	.99438	.999423
3.50	.997481	.996462	.99641	.999767
3.75	.998786	.997744	.99757	.999912
4.00	.999455	.998526	.99839	.999968

Table 3. Estimate of $P(W^* \leq x)$

5 Acknowledgements

The authors thank Dan Jordan for assisting with the software for the simulations in Section 4, Marty Erickson and Svante Janson for helpful discussions on this problem, and Lajos Takács for sending us several papers on total path length.

6 References

Chung, K. L. (1974) *A Course in Probability Theory, 2nd Ed.* Academic Press, Orlando, Fl.

Dobrow, R. P. and Smythe, R. T. (1996). Poisson approximation for functionals of random trees. *Rand. Struct. & Alg.*. **9** 79–92.

Elderton, W. P. and Johnson, N. L. (1969). *Systems of Frequency Curves.* Cambridge University Press, Cambridge.

Fill, J. A. (1996). On the distribution of binary search trees under the random permutation model. *Rand. Struct. & Alg.*. **8** 1–25.

Flajolet, P. and Odlyzko, A. M. (1990) Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics* **3** 216–240.

Hennequin, P. (1991). Analyse en moyenne d'algorithmes, tri rapide et arbres de recherche. Ph. D. dissertation, L'École Polytechnique Palaiseau.

Hoare, C. A. (1962). Quicksort. *Computer J.* **5** 10–15.

Kendall, M. G. and Stuart, A. (1963). *The Advanced Theory of Statistics, Vol. 1*. Charles Griffin & Co., Ltd., London.

Knuth, D. (1973). *The Art of Computer Programming,* Vol 3: *Sorting and Searching,* 2nd ed. Addison–Wesley, Reading, Mass.

Mahmoud, H. (1991). Limiting distributions for path lengths in recursive trees. *Probab. Eng. Info. Sci.* **5** 53–59.

Moon, J. W. (1974). The distance between nodes in recursive trees. *London Mathematics Society Lecture Notes Series, No. 13* London: Cambridge University Press, 125–132.

Régnier, M. (1989). A limit distribution for Quicksort. *Inform. Theor. Appl.* **23** 335–343.

Rösler, U. (1991). A limit theorem for Quicksort. *Inform. Theor. Appl.* **25** 85–100.

Smythe, R. T. and Mahmoud, H. (1995). A survey of recursive trees. *Theo. Prob. and Math. Stat.*, No. 51, 1–27.

Solomon, H. and Stephens, M. A. (1980). Approximations to densities in geometric probability. *J. Appl. Prob.* **17** 145–153.

Stanley, R. P. (1986) *Enumerative Combinatorics, Vol. I.* Wadsworth & Brooks/Cole, Monterey, Calif.

Szymański, J. (1990). On the complexity of algorithms on recursive trees. *Theo. Comp. Sci.* **74** 355–361.

Takács, L. (1992). On the total heights of random rooted trees. *J. Appl. Prob.* **29** 543–556.

Takács, L. (1994) On the total heights of random rooted binary trees. *J. Comp. Theory, Series B* **61** 155-166.

Tan, K. H. and Hadjicostas, P. (1995). Some properties of a limiting distribution in Quicksort. *Statistics and Probability Letters* **25** 87–94.

Robert P. Dobrow
Division of Mathematics and Computer Science
Truman State University
Kirksville, MO 63501-4221
bdobrow@mathax.truman.edu

James Allen Fill
Department of Mathematical Sciences
The Johns Hopkins University
Baltimore, MD 21218-2682
jimfill@jhu.edu