The space requirement of m-ary search trees: distributional asymptotics for $m \geq 27$

James Allen Fill[1⋆] and Nevin Kapur[2⋆⋆]

[1] Applied Mathematics and Statistics, The Johns Hopkins University, 3400 N. Charles St., Baltimore MD 21218-2682
[2] Computer Science, California Institute of Technology, MC 256-80, 1200 E.California Blvd., Pasadena CA 91125

Abstract. We study the space requirement of m-ary search trees under the random permutation model when $m \geq 27$ is fixed. Chauvin and Pouyanne have shown recently that X_n, the space requirement of an m-ary search tree on n keys, equals $\mu(n+1) + 2\mathrm{Re}\,[\Lambda n^{\lambda_2}] + \epsilon_n n^{\mathrm{Re}\,\lambda_2}$, where μ and λ_2 are certain constants, Λ is a complex-valued random variable, and $\epsilon_n \to 0$ a.s. and in L^2 as $n \to \infty$. Using the contraction method, we identify the distribution of Λ.

Keywords. m-ary search trees, space requirement, limiting distributions, contraction method.

1 Introduction

We start by giving a brief overview of search trees, which are fundamental data structures in computer science used in searching and sorting. For integer $m \geq 2$, the m-ary search tree, or multiway tree, generalizes the binary search tree. The quantity m is called the *branching factor*. According to [10], search trees of branching factors higher than 2 were first suggested by Muntz and Uzgalis [12] "to solve internal memory problems with large quantities of data." For more background we refer the reader to [7, 8] and [10].

An *m-ary tree* is a rooted tree with at most m "children" for each *node (vertex)*, each child of a node being distinguished as one of m possible types. Recursively expressed, an m-ary tree either is empty or consists of a distinguished node (called the *root*) together with an ordered m-tuple of *subtrees*, each of which is an m-ary tree.

An *m-ary search tree* is an m-ary tree in which each node has the capacity to contain $m-1$ elements of some linearly ordered set, called the set of *keys*. In typical implementations of m-ary search trees, the keys at each node are stored in increasing order and at each node one has m pointers to the subtrees. By spreading the input data in m directions instead of only 2, as is the case for a binary search tree, one seeks to have shorter path lengths and thus quicker searches.

We consider the space of m-ary search trees on n keys, and assume that the keys are linearly ordered. Hence, without loss of generality, we can take the set of keys to be $[n] := \{1, 2, \ldots, n\}$. We construct an m-ary search tree from a sequence s of n distinct keys in the following way:

(i) If $n < m$, then all the keys are stored in the root node in increasing order.

⋆ Research supported by NSF grant DMS–0104167 and by The Johns Hopkins University's Acheson J. Duncan Fund for the Advancement of Research in Statistics.
⋆⋆ Research partially supported by NSF grant 0049092.

(ii) If $n \geq m$, then the first $m-1$ keys in the sequence are stored in the root in increasing order, and the remaining $n-(m-1)$ keys are stored in the subtrees subject to the condition that if $\sigma_1 < \sigma_2 < \cdots < \sigma_{m-1}$ denotes the ordered sequence of keys in the root, then the keys in the jth subtree are those that lie between σ_{j-1} and σ_j, where $\sigma_0 := 0$ and $\sigma_m := n+1$, sequenced as in s.

(iii) All the subtrees are m-ary search trees that satisfy conditions (i), (ii), and (iii).

For example the m-ary search constructed from the sequence

$$(10, 7, 12, 4, 1, 8, 5, 6, 9, 14, 11, 2, 15, 13, 3)$$

is show in Figure 1. Note that empty nodes (also called *external nodes*) are represented as circles in the figure; m such nodes arise as children of a given node when that node becomes filled to its capacity of $m-1$ keys. In this paper the total number of nodes (empty and nonempty) in an m-ary search tree is called the *space requirement* of the tree.

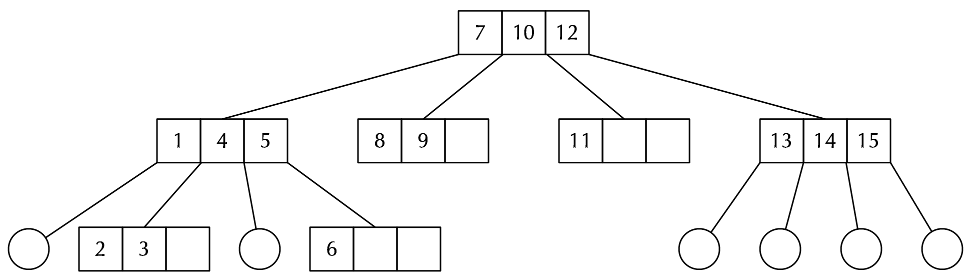

Fig. 1. An m-ary search tree with space requirement 13.

The uniform distribution on the space of permutations of $[n]$ induces a distribution of the space of m-ary search trees with n keys. This is known as the *random permutation model*.

Several authors have studied the limiting distribution of the space requirement under the random permutation model. Mahmoud and Pittel [11] showed that when $m \leq 15$, the limiting distribution is normal. The result was later extended to include $m \leq 26$ by Lew and Mahmoud [9]. Chern and Hwang [3] proved that when $m \geq 27$, the space requirement centered by its mean and scaled by its standard deviation does not have a limiting distribution. Our result, stated as Theorem 1, for the case $m \geq 27$ was inspired by a recent development (stated at the beginning of Section 2) of Chauvin and Pouyanne [2].

2 Summary

Let X_n denote the space requirement of an m-ary search tree on n keys chosen under the random permutation model. Recently, Chauvin and Pouyanne [2] have used martingale techniques to show that when $m \geq 27$, we have $X_n = \widehat{X}_n + n^{\sigma}\epsilon_n$, where

$$\widehat{X}_n := \frac{1}{H_m - 1}(n+1) + 2\mathrm{Re}\,[n^{\lambda_2}\Lambda], \tag{1}$$

with Λ some complex-valued random variable and $\epsilon_n \to 0$ a.s. and in L^2. [In fact, they derive the asymptotics of the random vector $(S_n^{(0)}, \ldots, S_n^{(m-1)})$, where $S_n^{(i)}$ denotes the number of nodes with i keys in a tree with n keys, but we shall be content here to study $X_n = \sum_{i=0}^{m-1} S_n^{(i)}$.] In this representation, $\lambda_2 = \sigma + i\tau$ is the root of the polynomial

$$\phi(z) \equiv \phi_m(z) := (z+1)\cdots(z+m-1) - m! \tag{2}$$

having second-largest real part and positive imaginary part. It is our goal to describe the distribution of the random variable Λ.

To begin, we define the following distributional transform T on $\mathcal{M}_2(\mu)$, the space of probability distributions with a certain mean μ defined at (7) and finite second absolute moment:

$$T: \ \mathcal{M}_2(\mu) \to \mathcal{M}_2(\mu), \quad \mathcal{L}(W) \mapsto \mathcal{L}\left(\sum_{k=1}^{m} S_k^{\lambda_2} W_k\right), \tag{3}$$

where $(W_k)_{k=1}^m$ are independent copies of W. Here $\mathbf{S} \equiv (S_1, \ldots, S_m)$ is the vector of spacings of $m-1$ independent Uniform$(0,1)$ random variables $U_1, \ldots, U_{m-1}$; i.e., if $U_{(1)}, \ldots, U_{(m-1)}$ are their order statistics and $U_{(0)} := 0$, $U_{(m)} := 1$, then

$$S_j := U_{(j)} - U_{(j-1)}, \quad j = 1, \ldots, m. \tag{4}$$

Furthermore, we take $\mathbf{S}$ to be independent of $(W_k)_{k=1}^m$. Next, define the metric d_2 on $\mathcal{M}_2(\mu)$ by

$$d_2(F,G) := \min\{\|X - Y\|_2 : \ \mathcal{L}(X) = F,\ \mathcal{L}(Y) = G\},$$

with $\|X\|_2 := (\mathbf{E}\,|X|^2)^{1/2}$ denoting the L^2-norm. In the sequel, for notational convenience we will write $d_2(X,Y)$ instead of $d_2(\mathcal{L}(X), \mathcal{L}(Y))$.

Our main result is the following. (See the remark below Lemma 7 for a strengthening.)

Theorem 1. *Let X_n denote the space requirement of an m-ary search tree on n keys under the random permutation model with $m \geq 27$. Define*

$$V_n := X_n - \frac{1}{H_m - 1}(n+1)$$

and $\widehat{V}_n := 2\mathrm{Re}\,[n^{\lambda_2} Y]$. Here Y is a random variable with distribution equal to the unique fixed point $\mathcal{L}(Y)$ of the distributional transform (3)*. Then $d_2(V_n, \widehat{V}_n) = o(n^{\sigma})$ and consequently Λ has the same distribution as Y.*

The proof of Theorem 1 is presented in Section 3, with the existence of the unique fixed point established in Section 3.1 and bounds on the d_2-distance derived in Section 3.2.

Remark. As discussed in [2] and [6], the study of the random vector $(S_n^{(0)}, \ldots, S_n^{(m-1)})$ can be recast as a generalized Pólya urn scheme which in turn can be studied by embedding into a continuous-time Markov multitype branching process. Janson [6] obtains asymptotic distributional results for a very general class of urn schemes and multitype branching processes. These include results for m-ary search trees, with (1) as a notable example. We anticipate that our contraction-method technique for identifying $\mathcal{L}(\Lambda)$ in (1) will extend quite generally to oscillatory cases of Janson's results; this is the subject of ongoing research. □

In the sequel we will use $1 =: \lambda_1, \lambda_2, \ldots, \lambda_{m-1}$ to denote the $m-1$ roots of (2) in nonincreasing order of real parts and roots with positive imaginary parts listed before their conjugates. In [10, §3.3] and [5], the polynomial $\psi(\lambda) = \phi(\lambda - 1)$ is considered. The properties of the roots of ϕ that we employ follow immediately from those known for the roots of ψ.

3 Proofs

As preliminaries, note that the space requirement X_n has initial conditions $X_0 = X_1 = \cdots = X_{m-2} = 1$, and for $n \geq m-1$ that the number of keys *not* stored in the root is

$$n' := n - (m-1).$$

It is well known that, under the random permutation model, X_n satisfies the distributional recurrence

$$X_n \stackrel{\mathcal{L}}{=} \sum_{k=1}^{m} X^{(k)}_{J_k} + 1, \quad n \geq m-1, \tag{5}$$

where $\stackrel{\mathcal{L}}{=}$ denotes equality in law (i.e., in distribution), and where, on the right,

- the random vector $\mathbf{J} \equiv (J_1, \ldots, J_m)$ is uniformly distributed over all m-tuples $(j_1, \ldots, j_m)$ of nonnegative integers with $j_1 + \cdots + j_m = n'$;
- for each $k = 1, \ldots, m$, we have $X^{(k)}_j \stackrel{\mathcal{L}}{=} X_j$;
- the quantities $\mathbf{J}; X^{(1)}_0, \ldots, X^{(1)}_{n'}; X^{(2)}_0, \ldots, X^{(2)}_{n'}; \ldots; X^{(m)}_0, \ldots, X^{(m)}_{n'}$ are all independent.

Using (5), we get a distributional recurrence for V_n, with notation as for the X's:

$$V_n \stackrel{\mathcal{L}}{=} \sum_{k=1}^{m} V^{(k)}_{J_k}, \quad n \geq m-1. \tag{6}$$

The initial conditions here are $V_j = 1 - \frac{j+1}{H_m - 1}$ for $j = 0, 1, \ldots, m-2$. The asymptotics of the mean of V_n can be derived using [5, Equation (2.7)]:

$$\mathbf{E}\, V_n = \mu n^{\lambda_2} + \bar{\mu} n^{\lambda_3} + O(n^{\operatorname{Re} \lambda_4}), \tag{7}$$

where μ is a constant. Note that no two roots of (2) have the same real part unless they are mutually conjugate, so that $\operatorname{Re} \lambda_4 < \operatorname{Re} \lambda_3 = \operatorname{Re} \lambda_2 = \sigma$.

For the reader's convenience, we state here a part of the Asymptotic Transfer Theorem of [5]. We will use this result in Section 3.2. The constant K' can be expressed in terms of K, but we shall have no use here for such an expression.

Proposition 2. *For fixed $m \geq 2$, consider the recurrence*

$$a_n = b_n + \frac{m}{\binom{n}{m-1}} \sum_{j=0}^{n'} \binom{n-1-j}{m-2} a_j, \quad n \geq m-1,$$

with specified initial conditions $(a_j)_{j=0}^{m-2}$. If $b_n = K n^v + o(n^v)$ with $v > 1$ and K a constant, then

$$a_n = K' n^v + o(n^v)$$

where K' is a constant.

3.1 Fixed point

The existence and uniqueness of the fixed point of the map T at (3) follows from the contraction method (see, e.g., [13]). Indeed a routine modification of the argument presented in [5, §6] yields that T is a contraction on $\mathcal{M}_2(\mu)$ with contraction factor

$$\rho = \left[m! \frac{\Gamma(2\sigma+1)}{\Gamma(2\sigma+m)} \right]^{1/2} = \left[\frac{m!}{(2\sigma+m-1)\cdots(2\sigma+1)} \right]^{1/2} < 1,$$

since for $m \geq 27$, we have $\sigma > 1/2$ [10, 5].

3.2 d_2 bounds

We begin by defining $d_n := d_2(V_n, \widehat{V}_n)$ and $f(t) := 2 \operatorname{Re} t = t + \bar{t}$. Unless otherwise noted we will henceforth assume $n \geq m-1$. Throughout $\sum_{\mathbf{j}}$ will denote a sum over all m-tuples $(j_1, \ldots, j_m)$ of nonnegative integers summing to n'.

By the triangle inequality,

$$d_n \leq a_n + b_n, \tag{8}$$

where, taking $(Y_k)_{k=1}^m$ to be independent copies of the random variable Y in Theorem 1 and $\mathbf{J}$ and $\mathbf{S}$ each independent of $(Y_k)_{k=1}^m$,

$$a_n := d_2\left(V_n, \sum_{k=1}^m f(J_k^{\lambda_2} Y_k) \right) \tag{9}$$

and

$$b_n := d_2\left(\sum_{k=1}^m f(J_k^{\lambda_2} Y_k), \sum_{k=1}^m f(n^{\lambda_2} S_k^{\lambda_2} Y_k) \right). \tag{10}$$

We proceed by deriving upper bounds for a_n and b_n separately. The bound on b_n is proved as Lemma 4.

For a_n a crude bound can be derived as follows. Even though this bound is not sufficient to show that $d_n = o(n^\sigma)$, it will be employed in Lemma 6, which in turn will be used to derive the estimate that we need.

Lemma 3. *With a_n defined at* (9)*,*

$$a_n = O(n^\sigma).$$

Proof. By the triangle inequality,

$$a_n \leq \|V_n\|_2 + \sum_{k=1}^m \|f(J_k^{\lambda_2} Y_k)\|_2 = \|V_n\|_2 + m \|f(J_1^{\lambda_2} Y_1)\|_2.$$

Since $J_1 \leq n'$ and $\|Y_1\|_2 < \infty$, we have $\|f(J_1^{\lambda_2} Y_1)\|_2 = O(n^\sigma)$. Using independence of the $V_{j_k}^{(k)}$'s, (6), and (7), we have

$$\begin{aligned} \|V_n\|_2^2 &= \sum_{\mathbf{j}} \mathbf{P}\left[\mathbf{J} = \mathbf{j}\right] \mathbf{E} \left| \sum_{k=1}^m V_{j_k}^{(k)} \right|^2 = \frac{1}{\binom{n}{m-1}} \sum_{\mathbf{j}} \sum_{k=1}^m \|V_{j_k}\|_2^2 + O(n^{2\sigma}) \\ &= \frac{m}{\binom{n}{m-1}} \sum_{j=0}^{n-(m-1)} \binom{n-1-j}{m-2} \|V_j\|_2^2 + O(n^{2\sigma}). \end{aligned}$$

It follows from Theorem 2 that $\|V_n\|_2^2 = O(n^{2\sigma})$, and the result follows. $\square$

To sharpen Lemma 3, we employ the following coupling between the distributions of V_n and of $\sum_{k=1}^m f(J_k^{\lambda_2} Y_k)$. The L^2 distance exhibited by this coupling serves as an upper bound on the d_2-distance. For $k = 1, \ldots, m$, let $(V_1^{(k)}, V_2^{(k)}, \ldots; Y_k)$ be independent copies of $(V_1, V_2, \ldots; Y)$ such that the coupling between V_j and Y is d_2-optimal for each j. [To construct such a coupling, first choose optimally-coupled V_1 and Y; having chosen $(V_1, \ldots, V_j; Y)$, choose V_{j+1} so that it is optimally-coupled with Y.] Then, with $\mathbf{J} \equiv (J_k)_{k=1}^m$ independent of everything else,

$$a_n^2 \leq \left\| \sum_{k=1}^m V_{J_k}^{(k)} - \sum_{k=1}^m f(J_k^{\lambda_2} Y_k) \right\|_2^2 = \sum_{\mathbf{j}} \mathbf{P}\left[\mathbf{J} = \mathbf{j}\right] \left\| \sum_{k=1}^m V_{j_k}^{(k)} - \sum_{k=1}^m f(j_k^{\lambda_2} Y_k) \right\|_2^2 . \tag{11}$$

Now

$$\begin{aligned} &\left\| \sum_{k=1}^m V_{j_k}^{(k)} - \sum_{k=1}^m f(j_k^{\lambda_2} Y_k) \right\|_2^2 \\ &= \sum_{k=1}^m \left\| V_{j_k}^{(k)} - f(j_k^{\lambda_2} Y_k) \right\|_2^2 + \mathbf{E} \sum_{1 \leq k \neq l \leq m} \left[V_{j_k}^{(k)} - f(j_k^{\lambda_2} Y_k) \right] \overline{\left[V_{j_l}^{(l)} - f(j_l^{\lambda_2} Y_l) \right]} \\ &= \sum_{k=1}^m d_{j_k}^2 + \sum_{1 \leq k \neq l \leq m} \mathbf{E}\left[V_{j_k}^{(k)} - f(j_k^{\lambda_2} Y_k) \right] \mathbf{E}\, \overline{\left[V_{j_l}^{(l)} - f(j_l^{\lambda_2} Y_l) \right]} \end{aligned} \tag{12}$$

If we choose the mean $\mathbf{E}\, Y$ to be μ, it follows from (7) that $\mathbf{E}\left[V_n - f(n^{\lambda_2} Y)\right] = O(n^{\operatorname{Re} \lambda_4})$. It follows then that the second sum in (12) is $O(n^{2 \operatorname{Re} \lambda_4}) = o(n^{2\sigma})$ uniformly in $\mathbf{j}$. Thus, from (11) and (12),

$$a_n^2 \leq \mathbf{E} \sum_{k=1}^m d_{J_k}^2 + r_n, \tag{13}$$

where $r_n = o(n^{2\sigma})$.

Next, we proceed to bound b_n.

Lemma 4. *With b_n defined at* (10),

$$b_n = o(n^{\sigma}).$$

Proof. We take $Y_1, \ldots, Y_m$ to be independent copies of Y and $(\mathbf{J}, \mathbf{S})$ independent of $Y_1, \ldots, Y_m$. The conditional distribution of $\mathbf{J}$ given $\mathbf{S} = \mathbf{s} \equiv (s_1, \ldots, s_m)$ is taken to be Multinomial$(n', \mathbf{s})$. Indeed this yields the distribution of the vector of sizes of the subtrees rooted at the root of a random m-ary search tree [4]. Then

$$
\begin{aligned}
b_n &\le \left\| \sum_{k=1}^{m} f(J_k^{\lambda_2} Y_k) - \sum_{k=1}^{m} f(n^{\lambda_2} S_k^{\lambda_2} Y_k) \right\|_2 \\
&\le \sum_{k=1}^{m} \left\| f(J_k^{\lambda_2} Y_k) - f(n^{\lambda_2} S_k^{\lambda_2} Y_k) \right\|_2 \\
&\le 2 \sum_{k=1}^{m} \left\| \, [J_k^{\lambda_2} - (nS_k)^{\lambda_2}] Y_k \, \right\|_2 && \text{(by definition of } f\text{)} \\
&= 2\|Y\|_2 \sum_{k=1}^{m} \left\| J_k^{\lambda_2} - (nS_k)^{\lambda_2} \right\|_2 && \text{(by independence)} \\
&= 2m\|Y\|_2 \left\| J_1^{\lambda_2} - (nS_1)^{\lambda_2} \right\|_2. && \text{(by symmetry)}
\end{aligned}
$$

We know that $\|Y\|_2 < \infty$, and by Lemma 5 to follow the last factor above is $o(n^\sigma)$. ∎

Lemma 5. *With $\sigma > 1/2$ denoting $\operatorname{Re}\lambda_2$,*

$$\|J_1^{\lambda_2} - (nS_1)^{\lambda_2}\|_2 = o(n^\sigma).$$

Proof. Given $\epsilon > 0$ we will show that the L_2-norm in question is bounded by a constant times $\epsilon^{1/2} n^\sigma$. The lemma then follows by letting $\epsilon \downarrow 0$.

Observe that

$$\|J_1^{\lambda_2} - (nS_1)^{\lambda_2}\|_2^2 = \mathbf{E}\,|J_1^{\lambda_2} - (nS_1)^{\lambda_2}|^2 = \mathbf{E}\,\mathbf{E}\left[\,|J_1^{\lambda_2} - (nS_1)^{\lambda_2}|^2 \mid S_1\right]. \tag{14}$$

Until further notice assume $s > 2\epsilon$, and note that the conditional expectation $\mathbf{E}\,[\,|J_1^{\lambda_2} - (nS_1)^{\lambda_2}|^2 \mid S_1 = s]$ equals

$$\sum_{j=0}^{n'} \mathbf{P}\,[J_1 = j | S_1 = s] |j^{\lambda_2} - (ns)^{\lambda_2}|^2 = \sum_{0 \le j \le n(s-\epsilon)} + \sum_{n(s-\epsilon) < j < n(s+\epsilon)} + \sum_{n(s+\epsilon) \le j \le n}.$$

The conditional distribution of J_1 given $S_1 = s$ is Binomial(n', s). The last sum on the right is $o(1)$ uniformly in s since, by [7, Ex. 1.2.10-21],

$$\mathbf{P}\,[J_1 \ge n(s+\epsilon) \mid S_1 = s] \le \mathbf{P}\,[J_1 \ge n'(s+\epsilon) \mid S_1 = s] \le \exp(-\epsilon^2 n'/2).$$

For the first sum observe that, for n large enough (independently of s),

$$\mathbf{P}\,[J_1 \le n(s-\epsilon) \mid S_1 = s] \le \mathbf{P}\left[J_1 \le n'\left(s - \frac{\epsilon}{2}\right) \middle| S_1 = s\right] \le \exp(-\epsilon^2 n'/8),$$

the last inequality being a consequence of the aforementioned exercise. Thus the first sum is also $o(1)$ uniformly in s.

On the other hand, for the range of summation in the middle sum, by the mean value theorem and the assumed inequality $\epsilon < s/2$ we have

$$\left| \left(\frac{j}{n}\right)^{\lambda_2} - s^{\lambda_2} \right| \le \epsilon|\lambda_2| \max_{\zeta \in (s-\epsilon, s+\epsilon)} |\zeta|^{\sigma-1} \le \epsilon|\lambda_2| c_\sigma s^{\sigma-1},$$

where c_σ is $(3/2)^{\sigma-1}$ if $\sigma \geq 1$ and $(1/2)^{\sigma-1}$ if $\sigma < 1$. Thus

$$|j^{\lambda_2} - (ns)^{\lambda_2}|^2 = n^{2\sigma} \left| \left(\frac{j}{n}\right)^{\lambda_2} - s^{\lambda_2} \right|^2 \leq \epsilon^2 |\lambda_2|^2 c_\sigma^2 s^{2(\sigma-1)} n^{2\sigma}.$$

Hence the middle sum is at most $\epsilon^2|\lambda_2|^2 c_\sigma^2 s^{2(\sigma-1)} n^{2\sigma}$.

Note that S_1 has distribution Beta$(1, m)$ and that

$$\int_0^1 s^{2(\sigma-1)}(1-s)^{m-1}\, ds = \frac{\Gamma(m)\Gamma(2\sigma-1)}{\Gamma(m+2\sigma-1)} < \infty$$

since $\sigma > 1/2$. So

$$\int_{2\epsilon}^1 \mathbf{E}\left[\, |J_1^{\lambda_2} - (nS_1)^{\lambda_2}|^2 \mid S_1 = s\right] \mathbf{P}\left[S_1 \in ds\right] \leq \text{constant} \times \epsilon^2 n^{2\sigma}.$$

Finally,

$$\int_0^{2\epsilon} \mathbf{E}\left[\, |J_1^{\lambda_2} - (nS_1)^{\lambda_2}|^2 \mid S_1 = s\right] \mathbf{P}\left[S_1 \in ds\right]$$
$$\leq \text{constant} \times n^{2\sigma} \mathbf{P}\left[S_1 \leq 2\epsilon\right] \leq \text{constant} \times \epsilon n^{2\sigma}.$$

□

Combining (8) and (13), we get

$$a_n^2 \leq \mathbf{E} \sum_{k=1}^m (a_{J_k} + b_{J_k})^2 + r_n = \mathbf{E} \sum_{k=1}^m a_{J_k}^2 + 2\mathbf{E} \sum_{k=1}^m a_{J_k} b_{J_k} + \mathbf{E} \sum_{k=1}^m b_{J_k}^2 + r_n. \quad (15)$$

Next we bound the terms on the right-hand side, so that (15) will yield a recursive inequality.

Lemma 6.

$$\mathbf{E} \sum_{k=1}^m b_{J_k}^2 = o(n^{2\sigma}).$$

Proof. By linearity of expectation and symmetry,

$$\mathbf{E} \sum_{k=1}^m b_{J_k}^2 = \sum_{k=1}^m \mathbf{E}\, b_{J_k}^2 = m\, \mathbf{E}\, b_{J_1}^2.$$

Now, the conditional distribution of J_1 given $S_1 = s$ is Binomial(n', s). We show that the conditional expectation $\mathbf{E}\,[b_{J_1}^2 \mid S_1 = s]$ is $o(n^{2\sigma})$. To that end, let X be distributed Binomial(n, s). For $\epsilon > 0$,

$$\mathbf{E}\, b_X^2 = \sum_{j=0}^n \mathbf{P}\left[X = j\right] b_j^2 = \sum_{0 \leq j \leq n(s-\epsilon)} + \sum_{n(s-\epsilon) \leq j \leq n}.$$

Now an argument similar to the one used in the proof of Lemma 5 can be employed. The first sum on the right is $o(n^{2\sigma})$. On the other hand, we use the fact that $b_n = o(n^\sigma)$ from Lemma 4 to conclude that the second sum is $o(n^{2\sigma})$. □

Lemma 7.

$$\mathbf{E}\sum_{k=1}^{m} a_{J_k} b_{J_k} = o(n^{2\sigma}).$$

Proof. The proof (using the crude bound on a_n established in Lemma 3) is very similar to that of Lemma 6. We omit the details. $\square$

We now complete the proof of Theorem 1. Using (15) and Lemmas 7 and 6 we find

$$\begin{aligned} a_n^2 &\leq \mathbf{E}\sum_{k=1}^{m} a_{J_k}^2 + g_n = \frac{1}{\binom{n}{m-1}} \sum_{\mathbf{j}} \sum_{k=1}^{m} a_{j_k}^2 + g_n = \frac{m}{\binom{n}{m-1}} \sum_{\mathbf{j}} a_{j_1}^2 + g_n \\ &= \frac{m}{\binom{n}{m-1}} \sum_{j=0}^{n-(m-1)} \binom{n-1-j}{m-2} a_j^2 + g_n, \end{aligned}$$

where $g_n = o(n^{2\sigma})$. It follows from Proposition 2 that $a_n^2 = o(n^{2\sigma})$, so that $d_n \leq a_n + b_n = o(n^{\sigma})$, as desired.

Remark. The o-estimates in Lemmas 4–7 can be improved to O-estimates. In the proof of Lemma 5, choosing ϵ as a function of n (specifically, taking ϵ_n to be a suitable constant multiple of $n^{-1/2}\log n$) sharpens the estimate $o(n^{\sigma})$ to $O(n^{\sigma-\frac{1}{4}}\sqrt{\log n})$, so that $b_n = O(n^{\sigma-\frac{1}{4}}\sqrt{\log n})$ in Lemma 4. In turn, Lemmas 6 and 7 are then immediately strengthened to $O(n^{2\sigma-\frac{1}{2}}\ln n)$ and $O(n^{2\sigma-\frac{1}{4}}\sqrt{\log n})$, respectively. This leads to $d_2(V_n, \widehat{V}_n) = O(n^{\operatorname{Re}\lambda_4}) + O(n^{\sigma-\frac{1}{8}}(\log n)^{\frac{1}{4}})$. Numerics strongly support the conjecture that $\sigma - \operatorname{Re}\lambda_4 \downarrow 0$ as $m \uparrow \infty$. If this is true, then $d_2(V_n, \widehat{V}_n)$ is $O(n^{\operatorname{Re}\lambda_4})$ whenever $m \geq 1044$. Due to the presence of $r_n = O(n^{2\operatorname{Re}\lambda_4})$ in (13), this large-m rate of convergence cannot be improved by the methods of this paper and presumably is the exact rate. $\square$

Finally, to prove equality in distribution of Λ and Y, we show that $d_2(\Lambda, Y) = 0$. Indeed with $\Lambda = |\Lambda|e^{i\Theta}$ and $Y = |Y|e^{iT}$, we have

$$\begin{aligned} d_2\left(\operatorname{Re}(n^{\lambda_2}\Lambda), \operatorname{Re}(n^{\lambda_2}Y)\right) &= d_2\left(\operatorname{Re}(n^{\sigma+i\tau}|\Lambda|e^{i\Theta}), \operatorname{Re}(n^{\sigma+i\tau}|Y|e^{iT})\right) \\ &= d_2\left(n^{\sigma}|\Lambda|\cos(\tau\ln n + \Theta), n^{\sigma}|Y|\cos(\tau\ln n + T)\right). \end{aligned}$$

But $d_2\left(\operatorname{Re}(n^{\lambda_2}\Lambda), \operatorname{Re}(n^{\lambda_2}Y)\right) = o(n^{\sigma})$ so that, as $n \to \infty$,

$$d_2\left(|\Lambda|\cos(\tau\ln n + \Theta), |Y|\cos(\tau\ln n + T)\right) \to 0.$$

For any fixed $\phi \in [0, 2\pi)$ we can choose $n \to \infty$ such that $(\tau \ln n) \bmod 2\pi \to \phi$. Then $|\Lambda|\cos(\phi + \Theta)$ and $|Y|\cos(\phi + T)$ have the same distribution. It follows from the Cramer–Wold device [1, Theorem 29.4] that the random vectors $(|\Lambda|\cos\Theta, |\Lambda|\sin\Theta)$ and $(|Y|\cos T, |Y|\sin T)$ have the same distribution. In particular, $\Lambda = |\Lambda|e^{i\Theta}$ and $Y = |Y|e^{iT}$ have the same distribution, as claimed. This completes the proof of Theorem 1.

References

1. P. Billingsley. *Probability and measure*. John Wiley & Sons Inc., New York, third edition, 1995. A Wiley-Interscience Publication.
2. B. Chauvin and N. Pouyanne. m-ary search trees when $m \geq 27$: a strong asymptotics for the space requirements. *Random Structures Algorithms*, 24(2):133–154, 2004.
3. H.-H. Chern and H.-K. Hwang. Phase changes in random m-ary search trees and generalized quicksort. *Random Structures Algorithms*, 19(3-4):316–358, 2001. Analysis of algorithms (Krynica Morska, 2000).
4. L. Devroye. Universal limit laws for depths in random trees. *SIAM J. Comput.*, 28(2):409–432 (electronic), 1999.
5. J. A. Fill and N. Kapur. Transfer theorems and asymptotic distributional results for m-ary search trees, arXiv:math.PR/0306050. Submitted for publication.
6. S. Janson. Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stochastic Process. Appl.*, 110(2):177–245, 2004.
7. D. E. Knuth. *The art of computer programming. Volume 1.* Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 3rd edition, 1997.
8. D. E. Knuth. *The art of computer programming. Volume 3.* Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 2nd edition, 1998.
9. W. Lew and H. M. Mahmoud. The joint distribution of elastic buckets in multiway search trees. *SIAM J. Comput.*, 23(5):1050–1074, 1994.
10. H. M. Mahmoud. *Evolution of random search trees.* John Wiley & Sons Inc., New York, 1992. A Wiley-Interscience Publication.
11. H. M. Mahmoud and B. Pittel. Analysis of the space of search trees under the random insertion algorithm. *J. Algorithms*, 10(1):52–75, 1989.
12. R. R. Muntz and R. C. Uzgalis. Dynamic storage allocation for binary search trees in a two-level memory. In *Proceedings of Princeton Conference on Information Sciences and Systems*, volume 4, pages 345–349, 1971.
13. U. Rösler and L. Rüschendorf. The contraction method for recursive algorithms. *Algorithmica*, 29(1-2):3–33, 2001. Average-case analysis of algorithms (Princeton, NJ, 1998).