Probabilistic analysis of bucket recursive trees*

Hosam M. Mahmoud*, Robert T. Smythe

Department of Statistics, George Washington University, 2201 G Street, NW #315, Washington, DC 20052, USA

In memory of M. Mahmoud, a remarkable man and a great father whose love defines the meaning of ideal parenthood.

Abstract

We introduce the bucket recursive tree, a generalization of recursive trees. The tree grows from a succession of integer labels that join the multitype nodes of the tree according to a stochastic rule. We study the multivariate structure of the tree and obtain a multivariate central limit theorem for the joint distribution of the number of nodes of different types for trees with bucket size $b < 26$. For trees with $b > 26$ a phase change in the distribution is detected and the central limit theorem does not hold. Recent results on the extended Pólya urn models (Smythe, 1995) provide the basic tool to reach these results. Two kinds of distances are also studied: the tree height and the depth of the $n$th label. Strong laws for the height are obtained via a technique based on introducing "ghost nodes", then removing them. The ghost nodes create a companion tree that relates our tree to Biggins' (1976) first- and last-birth problem in a multitype process. A weak law for the depth is obtained by formulating and asymptotically manipulating the depth probability generating function.

1. Introduction

In this paper we introduce bucket recursive trees, a generalization of recursive trees, which can model a variety of possible recruiting situations. In this model the nodes of a bucket recursive tree are buckets that can hold up to $b \geq 1$ labels. The case $b = 1$ reduces to that of the ordinary recursive tree. The bucket recursive tree grows by the progressive attraction of increasing integer labels. At the $(n + 1)$st stage the $n$ existing labels compete to attract the $(n + 1)$st label and all existing labels have equal chance of recruiting the next label. Thus a node with $i$ labels has "affinity" $i/n$ at stage $n + 1$, i.e., probability $i/n$ of attracting the $(n + 1)$st label. If the new label falls in an unfilled bucket, it simply joins the labels in that bucket; but if the new label has been attracted by a filled bucket, it is placed in a new bucket that is attached as a child of the

*Research has been supported in part by NSA grant No. MDA904-92-H3086.
*Corresponding author.
attracting node. Thus, the first $b$ labels $1, \ldots, b$ go into the root node. Label $b + 1$ is attached in a new bucket as a child to the root node. Label $b + 2$ may either join the same bucket as label $b + 1$, with probability $1/(b + 1)$, or start a new bucket, with probability $b/(b + 1)$, and so forth.

For the rest of this paper we shall simply use the term random tree to refer to the bucket recursive tree grown under the probability model just described. Fig. 1 illustrates all bucket recursive trees with $n = 5$ when the bucket capacity is 2. The top row of numbers gives the probabilities for these random trees.

There is extensive literature on the special case $b = 1$, the usual recursive trees. These have been proposed as models in a number of settings; a survey is given in [14]. There are also connections between the usual recursive trees and the UNION-FIND trees used in manipulation of data sets (the connection is explicit in [11]), and connections with the binary search tree, a fundamental structure of computer science (a bijection is given in [15]). A bucket recursive tree with capacity $b > 1$, by contrast, might model a growth or recruiting strategy for a business in the service sector; the presence of a franchise or facilities at a certain level in a given location offers logistical advantages in economics of scale, up to a point ($b$ units) of saturation, after which satellite facilities would be established in a new location.

The bucket recursive tree has a natural multivariate structure. Let $X_n^{(i)}$ denote the number of nodes of "type $i$" at stage $n$, where a node of type $i$ is one containing $i$ labels; the labels contained in a type $i$ node will be called type $i$ labels. The object of our probabilistic interest is then the vector

$$X_n = (X_n^{(1)}, \ldots, X_n^{(b)})^T,$$

the transpose of a vector is denoted by the superscript $T$ in this paper. Because $\sum_{i=1}^{b-1} i X_n^{(i)} = n$, there are only $b - 1$ linearly independent components of $X_n$, and for some results it is convenient to refer to the reduced vector

$$X_n^* = (X_n^{(1)}, \ldots, X_n^{(b-1)})^T.$$

To formulate our results we shall use the following standard notation throughout. The $k$th harmonic number $1 + 1/2 + \cdots + 1/k$ will be called $H_k$. The rising factorial $z(z + 1)\cdots(z + r - 1)$ will be denoted by $\langle z \rangle_r$, for any complex number $z$ and any integer $r \geq 0$. As usual, convergence in distribution, in probability, and almost surely
will be denoted by $x_i$, $y_i$, and $z_i$, respectively. A $b$-dimensional normal vector with mean 0 and covariance matrix $\Sigma$ will be denoted by $\mathcal{N}_b(0, \Sigma)$. An exponentially distributed random variable with parameter $\lambda$ will be denoted by $\text{EXP}(\lambda)$.

The paper is organized as follows. In Section 2, an exact formula for $E[X_n^{(i)}]$ will be found. Asymptotically, that formula gives

$$E[X_n^{(i)}] = \frac{1}{i(i + 1)H_b} n + O(n^a), \quad i = 1, \ldots, b - 1,$$

and

$$E[X_n^{(b)}] = \frac{1}{bH_b} n + O(n^a),$$

where $a_b < 1$. In Section 3, the results of Smythe [13] will be used to show

$$X_n^{(i)} \xrightarrow{p} \frac{1}{i(i + 1)H_b}, \quad i = 1, \ldots, b - 1;$$

and

$$X_n^{(b)} \xrightarrow{p} \frac{1}{bH_b}.$$

We will also obtain the multivariate central limit theorem

$$\frac{X_n^* - E[X_n^*]}{\sqrt{n}} \xrightarrow{d} \mathcal{N}_{b-1}(0, \Sigma)$$

for $b \leq 26$, for some covariance matrix $\Sigma$ that will be given for some small values of $b$. It will further be shown that for $b > 26$ there is, as in the case of search trees [8], a distinct phase change – the asymptotic behavior of $X_n$ is quite different for $b > 26$, and the central limit theorem does not hold. (The similarity to the phase change of $m$-ary trees is further explained in Section 3.)

Section 4 extends a technique used by Pittel [11] to show that

$$\frac{h_n}{\ln n} \xrightarrow{a_s} c_b,$$

where $h_n$ is the height of the tree with $n$ labels, and $c_b$ is a constant depending only on the bucket size $b$. The exact values of $c_2$ and $c_3$ are determined. Finally, in Section 5 the asymptotic expected depth of the $n$th label, and the order of magnitude of the variance of this depth are found. It is then shown that the depth of the $n$th insertion, normalized by $\ln n$, converges in probability to $1/H_b$. Appendix A contains a proof of the key Lemma 4. For the sake of completeness, Appendix B sketches some results from [13] adapted for direct application on bucket recursive trees. Appendix C has the details of the asymptotic calculations for the mean and variance of the depth of insertion.
2. Asymptotic analysis of the average number of nodes of different types

The growth of the random tree may be captured by considering a set of indicators. Let \( I^{(i)}_n \) be an indicator variable taking the value 1 if the \( n \)th insertion is made in a type \( i \) node, and 0 otherwise. Then

\[
X^{(i)}_{n+1} = X^{(i)}_n + I^{(i-1)}_{n+1} - I^{(i)}_{n+1}, \quad i = 2, 3, \ldots, b - 1. \tag{2.1}
\]

It is evident that

\[
E[I^{(i)}_{n+1} | \mathcal{F}_n] = \frac{iX^{(i)}_n}{n}, \quad i = 1, \ldots, b,
\]

where \( \mathcal{F}_n \) is the sigma-field generated by the first \( n \) stages; hence

\[
E[I^{(i)}_{n+1}] = \frac{i}{n} E[X^{(i)}_n], \quad i = 1, \ldots, b, \tag{2.2}
\]

and the expectation of (2.1), together with the last relation, gives

\[
E[X^{(i)}_{n+1}] = E[X^{(i)}_n] + E[I^{(i-1)}_{n+1}] - E[I^{(i)}_{n+1}]
\]

\[
= E[X^{(i)}_n] + \frac{i-1}{n} E[X^{(i-1)}_n] - \frac{i}{n} E[X^{(i)}_n].
\]

so that

\[
E[X^{(i)}_{n+1}] = \frac{n-i}{n} E[X^{(i)}_n] + \frac{i-1}{n} E[X^{(i-1)}_n], \quad i = 2, \ldots, b - 1. \tag{2.3}
\]

The recurrence is slightly different for the boundary values \( i = 1 \) and \( i = b \). For these boundary values we have

\[
X^{(1)}_{n+1} = X^{(1)}_n + I^{(b)}_{n+1} - I^{(1)}_{n+1}, \quad X^{(b)}_{n+1} = X^{(b)}_n + I^{(b-1)}_{n+1}.
\]

Taking expectations of the last two relations and using (2.2), we have

\[
E[X^{(1)}_{n+1}] = \frac{n-1}{n} E[X^{(1)}_n] + \frac{b}{n} E[X^{(b)}_n], \tag{2.4}
\]

\[
E[X^{(b)}_{n+1}] = E[X^{(b)}_n] + \frac{b-1}{n} E[X^{(b-1)}_n]. \tag{2.5}
\]
Organized in matrix form, Eq. (2.3)-(2.5) give the multivariate matrix recurrence relation

$$E[X_{n+1}] = \frac{1}{n} \begin{pmatrix} n-1 & 0 & 0 & 0 & \cdots & 0 & 0 & b \\ 1 & n-2 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 2 & n-3 & 0 & \cdots & 0 & 0 & 0 \\ & & & \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & \cdots & b-2 & n-(b-1) & 0 \\ 0 & 0 & \cdots & 0 & b-1 & n \end{pmatrix} E[X_n].$$

This system of simultaneous recurrence relations may be expressed in the form

$$E[X_{n+1}] = \left( I + \frac{1}{n} S \right) E[X_n]. \quad (2.6)$$

where $I$ is the $b \times b$ identity matrix and $S = [s_{ij}]$ is a $b \times b$ matrix of absolute numbers given by

$$S = \begin{pmatrix} -1 & 0 & 0 & 0 & \cdots & 0 & 0 & b \\ 1 & -2 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 2 & -3 & 0 & \cdots & 0 & 0 & 0 \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & \cdots & b-2 & -(b-1) & 0 \\ 0 & 0 & \cdots & 0 & b-1 & 0 \end{pmatrix}.$$ 

The type of matrix recurrence (2.6) appears in the fringe analysis technique [12] and can be solved exactly in terms of the eigenvalues and eigenvectors of $S$. For easy referencing, we index the eigenvalues according to the their real parts, starting from the largest and going down to the smallest. Namely, we assume the eigenvalues $\lambda_1, \ldots, \lambda_b$ are so that

$$\Re \lambda_1 \geq \Re \lambda_2 \geq \cdots \geq \Re \lambda_b.$$ 

The results are expressed in terms of a modal matrix of $S$. (A modal matrix of $S$ is one whose columns are eigenvectors of $S$.)

**Lemma 1.** Let $R$ be a modal matrix of $S$. If $R^{-1}$ exists, then the matrix recurrence (2.6) has the solution

$$E[X_n] = \frac{1}{(n-1)!} R \left( \prod_{j=1}^{n-1} (ji + D) \right) R^{-1} E[X_1],$$
where $D$ is the diagonal matrix

$$
\text{diag}(\lambda_1, \ldots, \lambda_b),
$$

and $\lambda_j$, for $j = 1, \ldots, b$, are the eigenvalues of $S$.

**Proof.** See [12]. \qed

In the following few lemmas we outline the required eigenvalue and eigenvector analysis of $S$.

**Lemma 2.** We have

$$
\prod_{j=1}^{n-1} (jI + D) = \text{diag}(\langle \lambda_1 + 1 \rangle_{n-1}, \ldots, \langle \lambda_b + 1 \rangle_{n-1}).
$$

**Proof.** This follows from an easy induction on $n$. \qed

**Lemma 3.** The characteristic equation of $S$ is

$$
\lambda(\lambda + 1) \cdots (\lambda + b - 1) = b!.
$$

(2.7)

**Proof.** Expand $\det(S - \lambda I)$ by its first row – the only nonzero elements in this row are $-\lambda - 1$ in the first column and $b$ in the last column. The former has a cofactor whose upper diagonal elements are all 0, hence the value of the cofactor is the product of its diagonal elements which are $-\lambda - 2, -\lambda - 3, \ldots, -\lambda - (b - 1), -\lambda$. The latter also has a cofactor whose lower diagonal elements are all 0, hence the value of the cofactor is the product of its diagonal elements which are $1, 2, \ldots, b - 1$. \qed

**Lemma 4.** The roots of the characteristic equation of $S$ are all simple and,

1. The value $\lambda_1 = 1$ is always the principal root\(^1\) with largest real part.
2. The integer $-b$ is a root iff $b$ is even.
3. Other than 1 and possibly $-b$, all the other eigenvalues are complex valued (with a nonzero imaginary part). These other roots occur in conjugate pairs. No two distinct eigenvalues have the same real part, except for conjugate pairs.
4. Apart from 1, all the other roots have real parts that are strictly less than 1. Thus 1 is the principal eigenvalue.
5. Let $\alpha_b = \Re \lambda_2$. For every $b \leq 26$, $\alpha_b < \frac{1}{2}$, whereas for all $b > 26$, $\alpha_b > \frac{1}{2}$.

A proof of Lemma 4 is included in Appendix A.

\(^1\)A principal root usually refers to the root with largest modulus. However, in this paper the essential role of a root $\lambda$ comes in the form $n^4$. Thus we are effectively exponentiating the roots, and the term principal root will refer in this paper to the root with largest real part.
Table 1
The eigenvalues of the characteristic equation for some small values of $b$

<table>
<thead>
<tr>
<th>$b$</th>
<th>$\lambda_b$</th>
<th>$b$</th>
<th>$\lambda_b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-2.00000</td>
<td>21</td>
<td>0.39621</td>
</tr>
<tr>
<td>3</td>
<td>-2.00000</td>
<td>22</td>
<td>0.41767</td>
</tr>
<tr>
<td>4</td>
<td>-1.50000</td>
<td>23</td>
<td>0.43729</td>
</tr>
<tr>
<td>5</td>
<td>-1.08527</td>
<td>24</td>
<td>0.45530</td>
</tr>
<tr>
<td>6</td>
<td>-0.77733</td>
<td>25</td>
<td>0.47189</td>
</tr>
<tr>
<td>7</td>
<td>-0.54749</td>
<td>26</td>
<td>0.48723</td>
</tr>
<tr>
<td>8</td>
<td>-0.37181</td>
<td>27</td>
<td>0.50146</td>
</tr>
<tr>
<td>9</td>
<td>-0.23407</td>
<td>28</td>
<td>0.51469</td>
</tr>
<tr>
<td>10</td>
<td>-0.12356</td>
<td>29</td>
<td>0.52704</td>
</tr>
<tr>
<td>11</td>
<td>-0.03310</td>
<td>30</td>
<td>0.53859</td>
</tr>
<tr>
<td>12</td>
<td>0.04223</td>
<td>35</td>
<td>0.58676</td>
</tr>
<tr>
<td>13</td>
<td>0.10590</td>
<td>40</td>
<td>0.62333</td>
</tr>
<tr>
<td>14</td>
<td>0.16041</td>
<td>45</td>
<td>0.65218</td>
</tr>
<tr>
<td>15</td>
<td>0.20716</td>
<td>50</td>
<td>0.67557</td>
</tr>
<tr>
<td>16</td>
<td>0.24887</td>
<td>100</td>
<td>0.78750</td>
</tr>
<tr>
<td>17</td>
<td>0.28527</td>
<td>500</td>
<td>0.90442</td>
</tr>
<tr>
<td>18</td>
<td>0.31761</td>
<td>1000</td>
<td>0.92822</td>
</tr>
<tr>
<td>19</td>
<td>0.34655</td>
<td>5000</td>
<td>0.95952</td>
</tr>
<tr>
<td>20</td>
<td>0.37261</td>
<td>10000</td>
<td>0.96738</td>
</tr>
</tbody>
</table>

Numerical solution of (2.7) shows that $\lambda_b < \frac{1}{2}$ for every $b \leq 26$; see Table 1. We have checked $\lambda_b$ for values of $b$ up to 10,000, and in the entire range of values $26 < b \leq 10,000$, $\lambda_b$ increases with $b$ and $\lambda_b > \frac{1}{2}$. Some selected values of $\lambda_b$ appear in Table 1.

Lemma 5. An eigenvector corresponding to the principal eigenvalue is

$$\left(\frac{1}{1}, \frac{1}{2 \times 2 \times 3 \times \cdots \times (b-1) b b} \right)^T.$$

Proof. Omitted. ☐

Let $R = [r_{ij}]$, for $1 \leq i, j \leq b$. According to Lemma 4, the roots of the characteristic equation are distinct. Hence the eigenvectors form a basis of $\mathbb{R}^b$, and $R^{-1}$ exists. Let $R^{-1} = [r_{ij}]^T$, for $1 \leq i, j \leq b$. We thus may proceed by multiplying out the matrices in Lemma 1, with $X_1 \equiv (1, 0, \ldots, 0)^T$, to obtain, with the aid of Lemma 2,

$$E[X_n^{(i)}] = \frac{1}{(n-1)!} \sum_{j=1}^{b} r_{ij} \langle \lambda_j + 1 \rangle_{n-1} r_{j1}.$$

(2.8)

As an illustration of this exact procedure, consider the case $b = 2$. In this case one finds

$$S = \begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix}.$$
having eigenvalues $\lambda_1 = 1$ and $\lambda_2 = -2$, with corresponding eigenvectors

$$v_1 = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} -2 \\ 1 \end{pmatrix}.$$ 

One then forms the modal matrix

$$R = \begin{pmatrix} \frac{1}{2} & -2 \\ \frac{1}{2} & 1 \end{pmatrix},$$

whose inverse is

$$R^{-1} = \frac{1}{3} \begin{pmatrix} 2 & 4 \\ -1 & 1 \end{pmatrix}.$$ 

Thus

$$E[X_n] = \frac{1}{(n - 1)!} \times 3 \begin{pmatrix} \frac{1}{2} & -2 \\ \frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} n! & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 4 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} n \\ \frac{1}{3} n \end{pmatrix}.$$ 

The exact averages as given in (2.8) may require very tedious linear algebra for high values of $b$. It is therefore desirable to obtain an approximation. The following theorem characterizes the asymptotic behavior of $E[X_n]$.

**Theorem 1.** As $n \to \infty$,

$$E[X_n^{(i)}] = \frac{1}{i(i + 1)H_b} n + O(n^\alpha), \quad i = 1, \ldots, b - 1,$n

$$E[X_n^{(b)}] = \frac{1}{bH_b} n + O(n^\alpha).$$

**Proof.** We first develop an asymptotic equivalent for the terms of the sum in (2.8). We have

$$\langle \lambda_j + 1 \rangle^{n-1} \begin{pmatrix} n \\ 1 \end{pmatrix} = \frac{\Gamma(n + \lambda_j)}{\Gamma(\lambda_j + 1) \Gamma(n)} = \frac{n^{\lambda_j}}{\Gamma(\lambda_j + 1)} \left( 1 + O\left(\frac{1}{n}\right) \right).$$

Hence from (2.8) we obtain

$$E[X_n^{(b)}] = r_{11}r_{11}' n + \sum_{j=2}^{b} \frac{r_{ij}r_{j1}n^{\lambda_j}}{\Gamma(\lambda_j + 1)} \left( 1 + O\left(\frac{1}{n}\right) \right) = r_{11}r_{11}' n + O(n^\alpha),$$

and according to Property 4 of Lemma 4, $\alpha_n < 1$. 


The values \( r_{ij} \) are the components of the eigenvector corresponding to \( \lambda_j \); this eigenvector is determined in Lemma 5 with a chosen scale. To determine \( r_{ij} \) for the same scale without doing any linear algebra, we may take advantage of the relation

\[
\sum_{i=1}^{b} iX^{(i)}_n = n.
\]

According to Lemma 5 and the asymptotic estimate (2.9), the last relation is equivalent to

\[
r_{11} \left[n + \sum_{i=1}^{b-1} \frac{n}{i+1} + O(n^2) \right] = n,
\]

valid for every \( n \) greater than some \( n_0 \). All the terms hidden in \( O \) are of the order \( n^{\lambda_j} \), \( j = 2, \ldots, b \). Thus all these terms are bounded by \( O(n^2) \), and are all of order less than \( n \) (see Property 4 of Lemma 4). The last equality is only possible if exact cancellations of all orders of magnitude less than \( n \) take place; for order \( n \) we must conclude

\[
r_{11} = \frac{1}{H_b}.
\]

Summing up the number of nodes of different types we get the following corollary.

**Corollary 1.** The asymptotic average number of buckets in a bucket recursive tree on \( n \) labels is \( n/H_b \).

**Remark 1.** One can refine type \( b \) nodes into leaf (childless) nodes and internal nodes. The computations are essentially the same to get the average proportion of internal nodes, only more detailed. These details have been checked by the authors but omitted here for brevity. One finds the asymptotic average number of internal nodes to be \( n/(b + 1) \). Summing up the average number of leaves of all types, one finds the asymptotic average total number of leaves to be \( bn/((b + 1)H_b) \); for large (but fixed) \( b \), the asymptotic average number of buckets in the tree is about \( n/\ln b \), and almost all the nodes of the tree are leaves.

3. **Probabilistic limits for the number of nodes of different types**

In formulating the results of this section, it is convenient to represent the growth of a bucket recursive tree by an urn model.

A generalized Pólya urn is an urn containing \( k \) types of balls with colors \( C_1, \ldots, C_k \). Initially, the urn contains a known number of each color. A ball is drawn at random (all choices of a ball being equally likely), the color of the ball is observed, and if its color is \( C_i \) then the ball is returned to the urn along with \( a_{ij} \) additional balls of color \( C_j \), for \( j = 1, \ldots, k \). The process is then repeated. For our purpose, it is convenient to
allow \( \alpha_{ij} \) to be negative. Technically speaking, these models are not generalized Pólya urns in the sense described in [1], for example; however they share most of their properties. In our case, as will be seen shortly, \( \sum_{j=1}^{i} \alpha_{ij} = 1 \), for each \( i \). In [13] these urns are called extended Pólya urns.

The equally likely objects in bucket recursive trees are the labels contained in the buckets. We may then make a correspondence between the labels and the balls of an extended Pólya urn with \( b \) colors. Each type \( i \) label in the tree corresponds to a ball of color \( C_i \), and the growth rules of the bucket recursive tree then correspond to the following ball addition scheme:

\[
\begin{pmatrix}
-1 & 2 & 0 & 0 & \cdots & 0 & 0 & 0 \\
0 & -2 & 3 & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & -3 & 4 & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & -(b-2) & b-1 & 0 \\
0 & 0 & \cdots & 0 & -(b-1) & b \\
1 & 0 & \cdots & 0 & 0 & 0 & 0
\end{pmatrix}
\]

Thus, \( iX_{n}^{(i)} \) corresponds to the number of balls of color \( C_i \) after \( n-1 \) draws from an urn with the initial composition of just 1 ball of color \( C_1 \). Let us denote this transition matrix by \( A = [a_{ij}] \). It is easy to see that \( A \) and the matrix \( S \) of Section 2 have the same characteristic polynomial, and hence the same eigenvalues. The results of Appendix B then enable us to conclude the following.

**Theorem 2.** We have

\[
\frac{X_{n}^{(i)}}{n} \xrightarrow{p} \frac{1}{i(i + 1)H_{b}}, \quad i = 1, \ldots, b - 1;
\]

\[
\frac{X_{n}^{(b)}}{n} \xrightarrow{p} \frac{1}{bH_{b}}.
\]

**Proof.** The left row eigenvector \( v \) of \( A \) corresponding to the eigenvalue 1 is easily seen to be

\[
v^T = \frac{1}{H_{b}} \left( \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{b}, 1 \right)^T.
\]

(3.1)

Since \( A \) governs the transition of individual balls (labels) and the number of type \( i \) balls at stage \( n \) is \( iX_{n}^{(i)} \), the theorem follows directly from (3.1) and Lemma B.1 of Appendix B. \( \square \)
To derive a central limit theorem for $X^*_n$, we again appeal to the results of [13], as given in Theorem B.1.

**Theorem 3.** If $b \leq 26$,

$$
\frac{X^*_n - E[X^*_n]}{\sqrt{n}} \Rightarrow \mathcal{N}_{b-1}(0, \Sigma).
$$

**Proof.** The associated extended Pólya urn has the same characteristic equation (2.7), and it turns out, by numerical solution of (2.7), that condition (c) of Theorem B.1 holds if $b \leq 26$. Also condition (d) of Theorem B.1 holds according to Lemma 4. Hence if $b \leq 26$, we have asymptotic multivariate normality. □

For $b = 27$, we do in fact have a "phase change," as in the case of $b$-ary search trees [8]. This latter family of trees is a class of search trees used in database systems. The $b$-ary search tree has nodes that can carry blocks of up to $b - 1$ data items, and consequently the nodes have branch factor $b$. It was shown in [8] that the normalized number of nodes allocated also experiences a phase change from being asymptotically normal for branch factors less than or equal to 26, to having no limit (under the same norming factor) for $b > 26$. This is also discussed in detail in [7]. A multivariate view for search trees is given in [5] with interpretations of the result as the search tree grows under the control of a memory management system such as fixed-allocation (paging), heap-allocation (available in the PASCAL programming language), and UNIX's buddy system.

The fact that the bucket recursive trees and $b$-ary search trees experience a distributional phase change at the same value (26) of their respective parameters suggests that there may be a bijection between the two classes. We do not pursue this any further in this paper. However, the interested reader may refer to [15] for an explicit construction of the bijection in the case $b = 1$.

**Theorem 4.** Let $Z^{(i)}_n$ be the number of labels of type $i$ in a bucket recursive tree with node capacity $b > 26$. Let $Z_n$ be the $b$-component vector whose components are $Z^{(i)}_n$. There exists a limiting random variable $W$ such that

$$
\frac{\xi Z_n - E[\xi Z_n]}{n^{1/2}} \Rightarrow W - \xi_1, \quad \text{almost surely and in } L^2,
$$

where $\xi = (\xi_1, \ldots, \xi_b)$ is the right eigenvector associated with $\lambda_2$.

**Proof (sketch).** Let $\{c_n\}$ be a sequence of (complex) constants with $c_1 = 1$. Further, let $\mathcal{F}_n$ be the sigma-field generated by $Z_1, \ldots, Z_n$. We denote the difference $Z_n - Z_{n-1}$ by...
\[ \Delta Z_{n-1}. \] Then
\[
E \left[ \frac{\xi Z_n}{c_n} \mid \mathcal{F}_{n-1} \right] = \frac{1}{c_n} \left( \xi Z_{n-1} + E \left[ \xi \Delta Z_{n-1} \mid \mathcal{F}_{n-1} \right] \right)
\]
\[
= \frac{1}{c_n} \left( \xi Z_{n-1} + \sum_{i=1}^{b} \sum_{j=1}^{b} \xi_i u_{ij} \frac{Z_{n-1}^{(j)}}{n} \right)
\]
\[
= \frac{1}{c_n} \left( \xi Z_{n-1} + \lambda_2 \sum_{j=1}^{b} \xi_j \frac{Z_{n-1}^{(j)}}{n} \right)
\]
\[
= \frac{1}{c_n} \xi Z_{n-1} \left( 1 + \frac{\lambda_2}{n} \right).
\]

Thus \{\xi Z_n/c_n, \mathcal{F}_n\} is a (complex) martingale iff \( c_n = c_{n-1}(1 + \lambda_2/n) \). Clearly, \( c_n \sim n^{\lambda_2}/(\lambda_2 + 2) \), and since by (B.1) of Appendix B we have
\[
\sup_n E \left| \frac{\xi Z_n}{c_n} \right|^2 \leq M \quad \text{for some} \ M \in \mathbb{R},
\]
the martingale convergence theorem asserts that there exists a random variable \( W \) such that
\[
\frac{\xi Z_n}{n^{\lambda_2}} \overset{a.s.}{\to} W < \infty,
\]
and convergence also takes place in \( L^2 \). Because \( E[\xi Z_n/c_n] = \xi_1/c_1 = \xi_1 \) for our urn, we have \( E[W] = \xi_1 \), and (3.2) follows. \( \square \)

Remark 2. It can be shown using results of [13], that \( \text{Var}[W] > 0 \). Hence \( W \) is not a constant almost surely, and it follows from Theorem 4 that \( (\xi Z_n - E[\xi Z_n])/\sqrt{n} \) cannot have central limit behavior.

So far we have had nothing to say about the entries of the covariance matrix \( \Sigma = [\sigma_{ij}] \) that appears in Theorem 3. There are several approaches, none of them very satisfactory, to find \( \Sigma \). The problem is that \( \Sigma \) is a rather complicated function of the eigenvalues and eigenvectors, which may be difficult to find themselves. For \( b = 2 \), it is not difficult to show that
\[
\sigma_{11} = \frac{8}{45},
\]
for $b = 3$, it seems to require substantial effort to verify that the asymptotic covariance matrix of $X_n^*$ is

$$
\Sigma = \begin{pmatrix}
1214 & 160 \\
6655 & 6655 \\
160 & 964 \\
6655 & 6655
\end{pmatrix}.
$$

For larger values of $b$, an easier way of finding $\Sigma$ would be most desirable.

4. A strong law for the height of the tree

The height of the tree is the longest distance between the root node and any other node in the tree.

We consider the case $b = 2$; the extension to higher values of $b$ will be immediately apparent. Our method here is a simple generalization of that of Pittel [11].

Let $h_n$ denote the height of the bucket recursive tree $T_n$. (It is known (cf. [11] or [3]) that for $b = 1$, $h_n/\ln n$ converges almost surely to $e$.) We define a new companion tree $T_n^*$ which will be the "generation tree" of Crump–Mode process [9]. The generation tree is defined as follows. Start with one type 1 particle at time 0. A type 1 particle produces a single type 2 particle at time distributed like EXP(1). Type 2 particles produce only type 1 particles, according to a Poisson process with rate $2$. All particles reproduce independently of one another and of their own birth time.

In the tree $T_n^*$, the odd-numbered generation consist only of type 1 particles and the even-numbered only of type 2 particles. There is a one-to-one correspondence between a bucket recursive tree $T_n$ and its counterpart $T_n^*$; Fig. 2 illustrates the example with $n = 6$.

The tree $T_n$ is formed by collapsing $T_n^*$, i.e. removing all “ghost” particles: These are type 1 particles that have already reproduced; in the example of Fig. 2 the ghosts are nodes 1 and 3, and removing them from $T_n^*$ produces $T_n$.

Fig. 2. A bucket recursive tree (a), and its corresponding generating tree (b).
Let $h_n^*$ denote the height of the tree $T_n^*$, and let $t_n^*$ be the time when the $n$th particle of $T_n^*$ is produced. Let $B(k)$ denote the first-birth time for the $k$th generation of the Crump–Mode process. At time $t_n^*$ the tree $T_n^*$ has height $h_n^*$, and since $B(h_n^*)$ and $B(h_n^* + 1)$ are the first times the height becomes equal to $h_n^*$ and $h_n^* + 1$, respectively, we have

$$B(h_n^*) \leq t_n^* \leq B(h_n^* + 1).$$

Theorem 2 of [2] gives that

$$\frac{B(k)}{k} \xrightarrow{a.s.} \gamma_2$$

for the Crump–Mode process, where $\gamma_2$ is a constant. Thus

$$\frac{B(h_n^*)}{h_n^*} \leq \frac{t_n^*}{h_n^*} \leq \frac{B(h_n^* + 1)}{h_n^*},$$

so that

$$\frac{h_n^*}{t_n^*} \xrightarrow{a.s.} \frac{1}{\gamma_2}.$$

But $h_n^* = h_n + y_n$, where $y_n$ is the number of levels of $T_n$ that contain a type 2 bucket. By the rules for filling buckets, $y_n = h_n$ or $h_n - 1$, so,

$$\frac{2h_n}{t_n^*} \xrightarrow{a.s.} \frac{1}{\gamma_2}. \quad (4.1)$$

Next note that $t_n^*$ may be written as $\Sigma_{k=2}^n \tau_k$, where $\tau_2, \ldots, \tau_n$ are independent and $\tau_k$ is distributed like $\text{EXP}(k - 1)$; this follows because the inter-event times for a Poisson process are independent exponentially distributed, and the minimum of independent $\text{EXP}(a)$ and $\text{EXP}(c)$ is $\text{EXP}(a + c)$. We thus have $E[t_n^*] \sim \ln n$ and $\text{Var}[t_n^*] = O(1)$, so

$$\frac{t_n^*}{\ln n} \xrightarrow{a.s.} 1. \quad (4.2)$$

Thus from (4.1) and (4.2) we have

$$\frac{h_n}{\ln n} \xrightarrow{a.s.} \frac{1}{2\gamma_2}. \quad (4.3)$$

It remains to determine the constant $\gamma_2$. Biggins [2] defines the matrix $\Phi(\theta)$, in which the $(i, j)$ entry is

$$E\left[ \int e^{-\theta t} dz_j(t) \mid \text{initial ancestor is of type } i \right].$$
where $z_j(t)$ is the $j$th component of the vector that records the number of first generation births of type $j = 1, 2$, at or before time $t$. For our Crump–Mode process it is easy to see that

$$\Phi(\theta) = \begin{pmatrix} 0 & 1/(\theta + 1) \\ 2/\theta & 0 \end{pmatrix},$$

and the largest eigenvalue $\phi(\theta)$ is $\sqrt{2/[\theta(\theta + 1)]}$. Defining

$$\mu(a) \equiv \inf \{ \theta \geq 0 : e^{a\theta} \phi(\theta) < 1 \},$$

we find that

$$\mu(a) = \frac{2a \exp\left[\frac{1}{2}(1 - a + \sqrt{a^2 + 1})\right]}{\sqrt{1 + \sqrt{a^2 + 1}}}$$

and Biggins [2] shows that $\gamma_2$ is determined as

$$\gamma_2 \equiv \inf \{ a : \mu(a) > 1 \}.$$

Thus, for $b = 2$, numerical solution gives, from (4.3),

$$\frac{h_n}{\ln n} \xrightarrow{a.s.} 1.67384 \ldots \quad (4.4)$$

For general $b$, the argument extends easily. Here we get

$$\frac{h_n}{\ln n} \xrightarrow{a.s.} \frac{1}{b\gamma_b},$$

and again $t^*_n / \ln n \xrightarrow{a.s.} 1$. For general $b$, the generation tree consists of $b$ different types of particles, and a type $i$ particle reproduces a single type $i + 1$ particle after time distributed like EXP$(i)$, for $i = 1, \ldots, b - 1$. But a type $b$ particle reproduces a type 1 particle according to a Poisson process with rate $b$, so we get

$$\phi(\theta) = \left[ \frac{b}{\theta(1 + \theta)(1 + 2\theta)\ldots(1 + (b - 1)\theta)} \right]^{1/b}$$

and the value that minimizes $e^{a\theta} \phi(\theta)$ is the root of the equation

$$a = \frac{f'(\theta)}{bf(\theta)},$$

where $f(\theta) = \theta(1 + \theta)(1 + 2\theta)\ldots(1 + (b - 1)\theta)$. This gives a $b$th order polynomial for $\theta$; for example, when $b = 3$, $\theta_0$ is the positive root of

$$6a\theta^3 + (9a - 6)\theta^2 + (3a - 6)\theta - 1 = 0,$$

and

$$\mu(a) = e^{a\theta_0} \phi(\theta_0).$$
Solving numerically we get $y_3 = 0.38480...$, and the strong law for the height of bucket recursive trees with bucket capacity 3 is

\[
\frac{h_n}{\ln n} \xrightarrow{a.s.} \frac{1}{3y_3} = 0.86625...
\]  

(4.5)

5. A weak law for the depth of the $n$th insertion

The height of a tree is a global property of the tree. But when the $n$th label joins the tree, it is attracted by some label that may be at any level in the tree varying between 0 and $h_n$. Thus, for $n > b$, the depth of insertion (distance from the root) of the $(n + 1)$st label is between 1 and $h_n + 1$. We therefore anticipate the average depth to be slightly less than the average height. In this section we derive the average and variance of $d_n$, the depth of the $n$th label when it joins the tree. This requires a further degree of detail: A classification of the nodes of type $i$ according to their level. Let $Y_n^{(i)}$ be the number of nodes of type $i$ at distance $k$ from the root in a tree containing $n$ labels. Define the indicators:

\[
I_{nj} = \begin{cases} 
1 & \text{if the } n\text{th node joins a type } j \text{ node,} \\
0 & \text{otherwise.}
\end{cases}
\]

For types $j = 2, 3, ..., b - 1$ we have

\[
E[Y_{n+1,k}^{(j)}|d_{n+1} = k, I_{n+1,j} = 1] = E[Y_{nk}^{(j)}] - 1,
\]

\[
E[Y_{n+1,k}^{(j)}|d_{n+1} = k, I_{n+1,j-1} = 1] = E[Y_{nk}^{(j)}] + 1,
\]

\[
E[Y_{n+1,k}^{(j)}|d_{n+1} \neq k \text{ or } \{I_{n+1,j} = 0 \text{ and } I_{n+1,j-1} = 0\}] = E[Y_{nk}^{(j)}].
\]

Unconditioning, we obtain

\[
E[Y_{n+1,k}^{(j)}] = E[Y_{nk}^{(j)}] + \text{Prob}\{d_{n+1} = k, I_{n+1,j-1} = 1\}
\]

\[
- \text{Prob}\{d_{n+1} = k, I_{n+1,j} = 1\}.\]

But clearly, for types $j = 2, 3, ..., b - 1$ we have

\[
\text{Prob}\{d_{n+1} = k, I_{n+1,j} = 1\} = \frac{jY_{nk}^{(j)}}{n}.
\]

Unconditionally, \( \text{Prob}\{d_{n+1} = k, I_{n+1,j} = 1\} = \frac{jE[Y_{nk}^{(j)}]}{n} \).

Hence, for $j = 2, 3, ..., b - 1$,

\[
E[Y_{n+1,k}^{(j)}] = E[Y_{nk}^{(j)}] + \frac{(j - 1)E[Y_{nk}^{(j-1)}] - jE[Y_{nk}^{(j)}]}{n}.\]  

(5.1)
To handle these multidimensional recurrence equations, introduce the generating functions

\[ B^{(j)}_{n,k}(z) = \sum_{k=0}^{\infty} E[Y^{(j)}_{nk}] z^k, \]

for \( j = 1, \ldots, b \) (the superscript \( j \) here does not indicate a derivative, it just emphasizes dependence on the node type). Multiplying both sides of (5.1) by \( z^k \) and summing over \( k \geq 1 \), then adjusting according to the boundary conditions, one obtains for \( j = 2, \ldots, b \):

\[ B^{(j)}_{n+1}(z) = \frac{n-j}{n} B^{(j)}_{n}(z) + \frac{j-1}{n} B^{(j-1)}_{n}(z), \quad (5.2) \]

valid for \( n \geq b \).

In a similar fashion, one obtains the following two recurrence equations for type 1 and type \( b \) nodes:

\[ E[Y^{(1)}_{n+1,k}] = \frac{n-1}{n} E[Y^{(1)}_{nk}] + \frac{b}{n} E[Y^{(b)}_{n,k-1}]; \]

\[ E[Y^{(b)}_{n+1,k}] = E[Y^{(b)}_{nk}] + \frac{(b-1)}{n} E[Y^{(b-1)}_{nk}]. \]

For types 1 and \( b \) the corresponding recurrence equations on the generating functions are:

\[ B^{(1)}_{n+1}(z) = \frac{n-1}{n} B^{(1)}_{n}(z) + \frac{b}{n} B^{(b)}_{n}(z), \quad (5.3) \]

\[ B^{(b)}_{n+1}(z) = B^{(b)}_{n}(z) + \frac{b-1}{n} B^{(b-1)}_{n}(z), \quad (5.4) \]

valid for \( n \geq b \). To solve the simultaneous recurrence relations, introduce

\[ B_n(z) = (B^{(1)}_{n}(z), \ldots, B^{(b)}_{n}(z))^T \]

and organize the recurrence equations (5.2)–(5.4) as a matrix recurrence in the form

\[ \frac{1}{n} \begin{pmatrix} n-1 & 0 & 0 & 0 & \cdots & 0 & 0 & bz \\ 1 & n-2 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 2 & n-3 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & b-2 & n-(b-1) & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & b-1 & n \end{pmatrix} B_n(z), \]
valid for \( n \geq b \). This functional recurrence equation can be solved by a method
similar to that of Lemma 1 (but we have to stop unwinding the matrix recurrence
when \( n = b \), because the recurrence is only valid for \( n \geq b \)). First write the matrix
recurrence in the form

\[
B_{n+1}(z) = \left( I + \frac{1}{n} S(z) \right) B_n(z).
\]

Let \( \lambda_1(z), \ldots, \lambda_b(z) \) be the eigenvalues of \( S(z) \), and let \( D(z) = \text{diag}(\lambda_1(z), \ldots, \lambda_b(z)) \).
Thus, \( \lambda_1(z), \ldots, \lambda_b(z) \) are the roots of the characteristic equation

\[
\lambda(\lambda + 1) \ldots (\lambda + b - 1) = b!z,
\]

which may be obtained in a manner similar to the proof of Lemma 3. If \( R(z) \) is
a modal matrix of \( S(z) \), the solution of the matrix recurrence can be expressed as

\[
B_n(z) = \frac{(b - 1)!}{(n - 1)!} R(z) \left( \prod_{j=b}^{n-1} (jI + D(z)) \right) R^{-1}(z) B_b(z).
\]

Remark 3. At \( z = 1 \), the characteristic equation (5.5) reduces to (2.7). Thus, each root
of (5.5), which is a continuous function of \( z \), approaches some root of (2.7), as \( z \to 1 \).
Therefore, we choose a natural indexing scheme: The root \( \lambda_j(z) \) of (5.5) is the one that
approaches the root \( \lambda_j \) of (2.7). In particular, \( \lambda_1(z) \to \lambda_1 = 1 \), as \( z \to 1 \). In addition \( R(1), R^{-1}(1), \) and \( D(1) \) are the same as \( R, R^{-1} \) and \( D \) of Section 2.

We next relate the depth \( d_n \) to the detailed profile of the multitype nodes at different
levels in the tree. The probability distribution of \( d_n \) is related to this profile as follows.
According to the growth rules

\[
\text{Prob}\{d_{n+1} = k | F_n \} = \frac{b Y_{n,k-1}^{(b)}}{n} + \frac{1}{n} \sum_{j=1}^{b-1} j Y_{nk}^{(j)}.
\]

Therefore, the unconditional probability distribution is

\[
\text{Prob}\{d_{n+1} = k \} = \frac{b}{n} E[ Y_{n,k-1}^{(b)} ] + \frac{1}{n} \sum_{j=1}^{b-1} j E[ Y_{nk}^{(j)} ].
\]

We can now find functional equations for the probability generating function of \( d_n \), i.e.
we will use (5.7) to find equations for

\[
P_n(z) \overset{\text{def}}{=} \sum_{k \geq 0} \text{Prob}\{d_n = k\} z^k.
\]

Multiply (5.7) by \( z^k \) and sum over \( k \); we obtain

\[
P_{n+1}(z) = \frac{z b}{n} B_n^{(b)}(z) + \frac{1}{n} \sum_{j=1}^{b-1} j B_n^{(j)}(z).
\]
Differentiating the last relation once at \( z = 1 \), we have

\[
P'_{n+1}(1) = E[d_{n+1}] = \frac{b}{n} B^{(b)}(1) + \frac{1}{n} (1, 2, \ldots, b) B'_n(1). \tag{5.8}
\]

A lengthy asymptotic computation, relegated to Appendix C, shows that

\[
E[d_n] \sim \frac{1}{H_b} \ln n. \tag{5.9}
\]

The second moment calculation is obtained from

\[
P''_{n+1}(1) = E[d_{n+1}(d_{n+1} - 1)] = \frac{2b}{n} \times \frac{dB^{(b)}(1)}{dz} + \frac{1}{n} (1, 2, \ldots, b) B'_n(1). \tag{5.10}
\]

The asymptotic calculations in Appendix C show that

\[
\text{Var}[d_n] = O(\ln n).
\]

By an application of Chebychev's inequality, we arrive at the main result of this section.

**Theorem 5.** In a bucket recursive tree, the average depth of the \( n \)th label satisfies the weak law

\[
\frac{d_n}{\ln n} \overset{p}{\to} \frac{1}{H_b}.
\]

In the usual random recursive tree \((b = 1)\), \( d_n/\ln n \overset{p}{\to} 1 \); see [10], and for additional distributional properties see [4] or [6]. On the other hand, in the usual random recursive tree \( h_n/\ln n \overset{p}{\to} e \); see [11]. This result is also implied in a subtle way in Devroye's study of the height of UNION-FIND trees [3]. The connection between these latter trees and recursive trees is made explicit by a construction in [11]. Thus, in the usual recursive tree the average depth is only about 37% of the height with high probability.

According to (5.9), for \( b = 2 \), we have

\[
E[d_n] \sim \frac{2}{3} \ln n.
\]

The matrices of Appendix C are 2 \( \times \) 2 and the calculation can be carried out exactly. The exact average depth for \( b = 2 \) is

\[
E[d_n] = \frac{2}{3} H_{n-1} - \frac{2}{9}, \quad n \geq 4.
\]

For \( b = 2 \) we can see from Theorem 5 and (4.4) that the depth is about 40% of the height with high probability. For \( b = 3 \), Theorem 5 and (4.5) assert that the ratio between the depth and the height goes up to 63% with high probability.
Acknowledgements

We are indebted to a referee whose help was essential in completing our proof of Lemma 4 in a rigorous form.

Appendix A

We prove Lemma 4 here. Verification of (1) and (2) is immediate. Other than 1 and possibly \(-b\), no real number can satisfy (2.7), and the first part of (3) follows because all the coefficients in the characteristic equation are real. It is easily seen that if \(a + ic\) is an eigenvalue, and \(c' > c\), then \(|a + ic'| > |a + 1 + ic'|\cdots|a + b - 1 + ic'| > b!\), verifying the second part of (3). Thus any root other than 1 and possibly \(-b\) must have a nonzero imaginary part and these other roots must occur in conjugate pairs.

If \(\lambda_j\) is a complex root (with nonzero imaginary part) of the characteristic equation (2.7) with \(\Re \lambda_j \geq 1\), then

\[|\lambda_j(\lambda_j + 1)\cdots(\lambda_j + b - 1)| > \Re \lambda_j \Re(\lambda_j + 1)\cdots\Re(\lambda_j + b - 1) \geq b!,\]

a contradiction that proves (4). To prove (5), we need to introduce new notation. Let

\[g_b(x) = \inf\{y > 0: (1 - x + iy)(2 - x + iy)\cdots(b - x + iy) \text{ is positive real}\}\]

and

\[f_b(x) = |(1 - x + ig_b(x))(2 - x + ig_b(x))\cdots(b - x + ig_b(x))|\]

A complex number \(z = 1 - x + iy\) solves the characteristic equation (2.7), if it satisfies:

(i) \(\arg z(z + 1)\cdots(z + b - 1) = 2\pi\).

(ii) \(f_b(x) = b!\).

For any \(x \in [0, \frac{1}{2}]\), standard Taylor expansion yields

\[\arg z(z + 1)\cdots(z + b - 1) = \tan^{-1}\frac{y}{1 - x} + \cdots + \tan^{-1}\frac{y}{b - x}\]

\[= y\left[\frac{1}{1 - x} + \cdots + \frac{1}{b - x}\right] - y^2 A_b(x, y),\]

where the function \(-y^2 A_b(x, y)\) is an error term that is continuous in \(x\), for \(x \in [0, \frac{1}{2}]\). Some little additional work investigating the derivatives of the error term (with respect to \(x\)) shows that it is bounded by 1. Property 3 states that no real root can be found in the interval \(x \in [0, \frac{1}{2}]\). Thus we must try to find roots with positive real part. That is, we can try to solve

\[y \sum_{k=1}^{b} \frac{1}{k - x} - y^2 A_b(x, y) = 2\pi,\]
or
\[ y^2 A_b(x, y) - y[\ln b + \theta_b(x)] + 2\pi = 0, \]
where \( \theta_b(x) \) is bounded from below uniformly in \( x \) by \(-1/(b+1)\). The graph of this curve lies below that of
\[ h_b(x, y) \triangleq y^2 - y[\ln b + \theta_b(x)] + 2\pi. \]
The quadratic equation \( h_b(x, y) = 0 \) has two real solutions, one of which can easily be shown to be
\[ \frac{2\pi}{\ln b + \theta_b(x)} \]
Thus
\[ 0 < g_b(x) < \frac{2\pi}{\ln b + \theta_b(x)} + O\left( \frac{1}{\ln^3 b} \right), \]
and the constant of \( O \) is less than 1. Now, observe that \( f_b(0) > b! \), and \( f_b(\frac{1}{2}) < b! \), if \( |\frac{1}{2} + ig_b(\frac{1}{2})| < 1 \).

It can now be easily shown that a sufficient condition for this to happen is \( \ln b + \theta_b(x) > 6 \). The uniform lower bound on \( \theta_b(x) \) renders \( b > \exp(\frac{1}{2}) \) a sufficient condition, which holds for all \( b > 666 \). By the continuity of \( f_b(x) \), the characteristic equation (2.7) has a solution with real part in \([0, \frac{1}{2}]\), i.e. \( \alpha_b = \Re \lambda_2 > \frac{1}{2} \), for all \( b > 666 \).

As discussed in Section 3 (cf. Table 1), we have checked by computer that \( \alpha_b > \frac{1}{2} \), for \( b = 27, \ldots, 10000 \). Indeed, for all \( b > 26 \), \( \alpha_b > \frac{1}{2} \).

It only remains to show that all the roots are simple. Let
\[ \eta(\lambda) \triangleq \langle \lambda \rangle_b - b! . \]
At \( \lambda_j \), a root of (2.7),
\[ \eta'(\lambda_j) = b! \left[ \frac{1}{\lambda_j} + \frac{1}{\lambda_j + 1} + \cdots + \frac{1}{\lambda_j + b - 1} \right]. \]
Clearly, neither \( \eta'(1) \) nor \( \eta'(-b) \) is equal to 0. Thus, \( \lambda_1 = 1 \) is a simple root, and for even values of \( b \), the root \(-b\) is also simple. For any other root \( \lambda_j \), we have
\[ \eta'(\lambda_j) = b! \left[ \frac{1}{\Re \lambda_j + i\Im \lambda_j} + \frac{1}{1 + \Re \lambda_j + i\Im \lambda_j} + \cdots + \frac{1}{b - 1 + \Re \lambda_j + i\Im \lambda_j} \right]. \]
Hence
\[ \Im \eta'(\lambda_j) = -b! \Im \lambda_j \left[ \frac{1}{(\Re \lambda_j)^2 + (\Im \lambda_j)^2} + \cdots + \frac{1}{(b - 1 + \Re \lambda_j)^2 + (\Im \lambda_j)^2} \right] \neq 0 \]
because the imaginary part of $\lambda_j$ is not 0. That is, $\eta'(\lambda_j) \neq 0$, and $\lambda_j$ is a simple root, too. □

Appendix B

Consider an extended Pólya urn model with $p$ types of balls where an integral number $\lambda_1$ of balls is added at each draw. For our purposes, we assume that for each type of ball drawn, the distribution of the balls added deterministic. We assume that the model is tenable; that is, the transition matrix $A = [a_{ij}]$ has the property that if $a_{ij} < 0$, then $a_{ij}$ is a divisor of $a_{ij}$ for $i \neq j$, and is also a divisor of the initial number of type $j$ balls in the urn. Assume further that $A$ is nondegenerate in the sense that not all rows of $A$ are multiples of the left row eigenvector $v_1$ of the principal eigenvalue (which is $\lambda_1$). Let $X_1^{(0)}, X_2$, and $X_3^*$ be defined as in the text.

Lemma B.1. Let $A$ be the transition matrix of a tenable extended Pólya model with $\lambda_1$ balls added at each draw. Assume that:

(a) The principal eigenvalue $\lambda_1 > 0$ is of multiplicity 1 and has a left row eigenvector $v_1$ with positive components;

(b) The eigenvectors of $A$ are linearly independent.

Then

$$\frac{X_i^{(n)}}{n} \xrightarrow{p} \lambda_1 v_1 \text{ for } i = 1, \ldots, p.$$ 

Proof (sketch). The right eigenvector $u$ of $\lambda_1$ is $u = (1, 1, \ldots, 1)^T$, so that $(n + 1)^{-1}uX_n = \lambda_1$, for any $n$. If $\zeta$ is a right eigenvector of a nonprincipal eigenvalue $\lambda_j$, it is shown in [13] that

$$E|\zeta X_n|^2 \leq \text{constant} \times \begin{cases} n & \text{if } 2\Re \lambda_j < \lambda_1, \\ n \ln n & \text{if } 2\Re \lambda_j = \lambda_1, \\ n^{2\Re \lambda_j / \lambda_1} & \text{if } 2\Re \lambda_j > \lambda_1. \end{cases} \quad (B.1)$$

Because $\Re \lambda_j < \lambda_1$, Chebychev's inequality gives $\frac{X_n}{n} \xrightarrow{p} 0$ in all three cases.

If $\mathcal{S}$ denotes the vector space spanned by the nonprincipal right eigenvectors,

$$\frac{\zeta X_n}{n} \xrightarrow{p} 0 \quad \text{for any } \zeta \in \mathcal{S}.$$ 

Representing a general vector $\eta$ as $\eta = c_1 u + c_2 \zeta$, where $\zeta \in \mathcal{S}$, it follows that $\eta v_1 = 0$ (if $\eta v_1$ is normalized to 1), and hence $n^{-1} \eta X_n \xrightarrow{p} \lambda_1 \eta v$, for any $\eta$, proving the lemma. □

Under additional assumptions, a central limit theorem for the urn contents is derived in [13].
Theorem B.1. Let \( A \) be the transition matrix of a nondegenerate tenable extended Pólya urn model with \( \lambda_1 \) balls added at each draw. Assume that (a) and (b) of Lemma B.1 hold, as well as

(c) For any nonprincipal eigenvalue \( \lambda_j \), \( 2\Re \lambda_j < \lambda_1 \);

(d) All eigenvalues are simple, and no two distinct nonconjugate eigenvalues have the same real part.

Then

\[
\frac{X_n^* - E[X_n^*]}{\sqrt{n}} \xrightarrow{d} \mathcal{N}_{n-1}(0, \Sigma).
\]

Appendix C

In this appendix we find asymptotic equivalents of the average and variance of the depth of the \( n \)th insertion.

The average depth is given by (5.8). The first factor in the average depth has already been computed in Section 2:

\[
B_n^{(b)}(1) = \sum_{k \geq 0} E[Y_{nk}^{(b)}],
\]

where the right-hand side is just the number of nodes of type \( b \) in the whole tree. According to Theorem 1,

\[
\frac{b}{n} B_n^{(b)}(1) \sim \frac{1}{H_n}.
\]

We can find \( B_n^b(1) \) by differentiating the solution (5.6) once with respect to \( z \) and evaluating the result at \( z = 1 \). This yields

\[
B_n^b(1) = \frac{(b - 1)!}{(n - 1)!} \times \left\{ \left[ \frac{d}{dz} R(1) \right] \prod_{j=b}^{n-1} (jI + D(1)) R^{-1}(1) B_b(1) \right. \\
+ R(1) \left[ \frac{d}{dz} \left( \prod_{j=b}^{n-1} (jI + D(z)) \right) \right]_{z=1} \left. R^{-1}(1) B_b(1) \right. \\
+ R(1) \prod_{j=b}^{n-1} (jI + D(1)) \left[ \frac{d}{dz} R^{-1}(1) \right] B_b(1) \\
+ R(1) \prod_{j=b}^{n-1} (jI + D(1)) R^{-1}(1) B_b^r(1) \right\} \\
= \frac{(b - 1)!}{(n - 1)!} \left( p_n + q_n + r_n + s_n \right).
\]
Asymptotic calculations are simplified by noticing that:

(i) $B_\ell(1) = 0$, which follows from the boundary conditions. Therefore, $s_n = 0$.

(ii) The term $[(b - 1)!/(n - 1)!]p_n$ is a diagonal matrix whose diagonal elements are all $O(n)$. In computing this term, we first find the diagonal matrix

$$
\frac{1}{(n - 1)!} \prod_{j-b}^{n-1} (jI + D(1)) = \text{diag}(y_{11}, \ldots, y_{bb}),
$$

where a typical diagonal element $y_{kk} = y_{kk}(n)$ is asymptotically

$$
y_{kk}(n) = \frac{(b + \lambda_k(1))(b + 1 + \lambda_k(1))\cdots(n - 1 + \lambda_k(1))}{(n - 1)!}
= \frac{\Gamma(n + \lambda_k(1))}{\Gamma(n)\Gamma(\lambda_k(1) + b)}
\sim \frac{n^{\lambda_k(1)}}{\Gamma(\lambda_k(1) + b)}.
$$

But, for $k = 1, \ldots, b$, we have $|\lambda_k(1)| \leq 1$, since at $\varepsilon = 1$ the characteristic equation (5.5) reduces to (2.7), whose principal eigenvalue is 1. Thus the diagonal elements are all $O(n)$ in the product, and subsequent premultiplication by $(b - 1)! R(1)$, then postmultiplication by $R^{-1}(1)B_\ell(1)$ (which are matrices of absolute constants) do not change the order of magnitude.

(iii) The term $[(b - 1)!/(n - 1)!]r_n$ is also a diagonal matrix whose diagonal elements are all $O(n)$ by a calculation similar to (ii).

(iv) The dominant term comes from the matrix $[(b - 1)!/(n - 1)!]q_n$. Here

$$
q_n = \sum_{j=b}^{n-1} R(1)((bI + D(1))((b + 1)I + D(1))\cdots((j - 1)I + D(1))
\times D'(1)\times((j + 1)I + D(1))((j + 2)I + D(1))\cdots((n - 1)I + D(1)))
\times R^{-1}(1)B_\ell(1).
$$

A matrix product of the form $(\ell I + D(1))\cdots(kI + D(1))$ can be expressed as a diagonal matrix

$$
\text{diag}\left(\langle\lambda_1(1)\rangle_k^{r+1}, \ldots, \langle\lambda_b(1)\rangle_k^{r+1}\right).
$$

So, using the fact that $B_\ell(1) = (0, 0, 0, \ldots, 0, 1)^T$, we can multiply out the matrices in the $j$th term of (C.1) to get a vector whose $k$th entry is

$$
\sum_{s=1}^{b} r_{ks}(1) f_n(\lambda_s(1)) r'_{ab}(1),
$$
where

\[ f_n(\lambda) \equiv \lambda'_1(1) \sum_{j=b}^{n-1} \frac{1}{\lambda_j(1) + j}. \]

And so,

\[
\frac{f_n(\lambda)}{(n-1)!} = \frac{\lambda'_1(1)}{\langle \lambda_1(1) \rangle_b} \times \frac{\lambda_1(1)(\lambda_1(1) + 1) \cdots (\lambda_1(1) + n - 1)}{(n - 1)!} \sum_{j=b}^{n-1} \frac{1}{\lambda_j(1) + j}
\]

\[
= \frac{\lambda'_1(1)}{\Gamma(\lambda_1(1))} \times \frac{\Gamma(n + \lambda_1(1))}{\Gamma(n)} \left[ \mathcal{H}_n(\lambda_1(1)) - \mathcal{H}_b(\lambda_1(1)) \right],
\]

where \( \mathcal{H}_m(\lambda) \) is Dirichlet's generalized harmonic number

\[ \mathcal{H}_m(\lambda) = \frac{1}{\lambda} + \frac{1}{\lambda + 1} + \cdots + \frac{1}{\lambda + m - 1}. \]

In particular

\[ \frac{f_n(\lambda_1)}{(n-1)!} \sim \frac{\lambda'_1(1)}{b!} n \ln n, \quad (C.2) \]

whereas, for \( s = 2, \ldots, b, \)

\[ \frac{f_n(\lambda_s)}{(n-1)!} \sim \frac{\lambda'_s(1)}{\Gamma(\lambda_s(1))} n^{\lambda_s(1)} \ln n = O(n^{\alpha_s} \ln n), \quad (C.3) \]

which is only \( O(n) \) because \( \alpha_b < 1 \), according to Lemma 4.

Collecting the different asymptotic contributions in (i)–(iv), we finally get from (5.8) the asymptotic estimate

\[ P_{n+1}(1) = E[d_{n+1}] = \frac{1}{n} \left[ \frac{(b - 1)!}{(n - 1)!} \sum_{k=1}^{b} \sum_{s=1}^{b} r_{ks} f_n(\lambda_s(1)) r_{sb} \right] + O(1). \]

According to the asymptotic developments (C.2) and (C.3), the term corresponding to the root \( \lambda_1(1) = 1 \) dominates and we get

\[ E[d_n] = \frac{\lambda'_1(1)r_{1b} \ln n}{b} \sum_{k=1}^{b} kr_{k1} + O(1). \]

The remaining sum can be readily computed from Lemma 5 yielding \( H_b \), and the asymptotic average depth reduces to the simple expression

\[ E[d_n] = \frac{\lambda'_1(1)r_{1b} H_b}{b} \ln n + O(1). \quad (C.4) \]

The derivative \( \lambda'_1(1) \) is easy to compute. The root \( \lambda_1(z) \) satisfies

\[ \lambda_1(z)(\lambda_1(z) + 1) \cdots (\lambda_1(z) + b - 1) = b! z, \]
whose derivative is
\[
\lambda_1'(1) = \frac{b^{-1}}{\lambda_1(1) + 1}\sum_{j=0}^{b-1} \frac{\lambda_1(1)(\lambda_1(1) + 1)\cdots(\lambda_1(1) + b - 1)}{\lambda_1(1) + j} = b!
\]
or
\[
\lambda_1'(1) = \frac{1}{H_b}. \quad (C.5)
\]

Lemma C.1. We have
\[
r_{1b}' = \frac{b}{H_b}.
\]

Proof. The modal matrix \( R \) satisfies
\[
R^{-1}S = DR^{-1}.
\]
Equating the \( k \)th entries of the first row on both sides, we have
\[
\sum_{j=1}^{b} r_{j}'_k s_{jk} = \lambda_1 r_{1k}' = r_{ik}'.
\]
We can now execute an iterative scheme starting at \( k = 1 \). At the \( k \)th step we obtain
\[
-k r_{1k}' + r_{1,k+1}' = r_{ik}'.
\]
Thus the components of the first row of \( R^{-1} \) occur in the proportion
\[
r_{11}': r_{12}': \ldots: r_{1b}' = 1:2: \ldots: b.
\]
We complete the proof by recalling that \( r_{11}' \) was found in the proof of Theorem 1 to be \( 1/H_b \). That is, the first row of \( R^{-1} \) is the row vector \( H_b^{-1}(1,2,\ldots,b) \). \( \square \)

From (C.4), (C.5) and Lemma C.1, we finally have
\[
E[d_n] = \frac{\ln n}{H_b} + O(1). \quad (C.6)
\]
Several computations for the second moment of \( d_n \) are similar to those in the average and will only be outlined. As a starting point for the second moment calculation, we use (5.10). The first factor is like the \( b \)th component of \( B_n'(1) \) derived by \( n \). In the course of calculating the average we have found this quantity to be only \( O(\ln n) \). For the second term in (5.10), we need to first compute \( B_n''(1) \). We can find this term by differentiating (5.6) twice at \( z = 1 \). This second derivative comprises ten terms, of which:

(i) Four are identically \( 0 \) (these four terms include \( B_0''(1) \) or \( B_0'(1) \), which are \( 0 \) owing to the boundary conditions).

(ii) Two other terms are matrices whose components are \( O(n \ln n) \): These terms correspond to differentiating the product in (5.6) once and differentiating either \( R(z) \)
or $R^{-1}(z)$ once, at $z = 1$. Clearly these matrices behave like $B_n(1)$, found in the calculation of the average, the only difference being that we premultiply or postmultiply by the derivative of $R(z)$ or $R^{-1}(z)$ at $z = 1$; that is, only the absolute numbers are changed, but not the order of magnitude.

(iii) Two terms involving the second derivatives of $R$ or $R^{-1}$ include terms that are only $O(n)$; they are like the terms $p_n$ and $r_n$ in the calculation of the mean, only differing in the constants, but not in the order of magnitude.

(iv) One term is a matrix whose components are $O(1)$: This is the term that includes the derivatives of $R(z)$ and $R^{-1}(z)$. At $z = 1$, $B_n(1)$ is a vector whose components are counts of the number of nodes of each type. Hence each component is $O(n)$. Comparing the matrix being considered here with $B_n(1)$ from (5.6), we see that our term has the same order of magnitude, and only differs in the constants.

(v) The main contribution comes from the term involving the second derivative of $\Pi_n(z)$ at $z = 1$, the product in (5.6). We can readily compute this diagonal matrix

$$
\Pi_n^2(1) = \sum_{j=b}^{n-1} \sum_{s=b}^{j-1} \frac{((bI + D(z))((b + 1)I + D(z))\cdots((s - 1)I + D(z))D'(z)}{j-b \choose s-b} \frac{(sI + D(z))}{s=j+1} \prod_{s=j+1}^{n-1} (sI + D(z)) \\
\times ((s + 1)I + D(z))\cdots((j - 1)I + D(z))D'(z) \\
+ \sum_{j=b}^{n-1} \left( \prod_{s=b}^{j-1} (sI + D(z)) \right) D''(s) \prod_{s=j+1}^{n-1} (sI + D(z)) \\
+ \sum_{j=b}^{n-1} \prod_{s=b}^{j-1} (sI + D(z)) \times D'(s) \\
\times \sum_{s=j+1}^{n-1} ((j + 1)I + D(z))\cdots((s - 1)I + D(z))D'(z) \\
\times ((s + 1)I + D(z))\cdots((n - 1)I + D(z)).
$$

At $z = 1$, the $k$th diagonal entry is

$$
\pi_{kk}(n) = (\lambda_k(1))^2 \sum_{j=b}^{n-1} \sum_{s=b}^{j-1} \frac{(b + \lambda_k(1))\cdots(n - 1 + \lambda_k(1))}{(s + \lambda_k(1))(j + \lambda_k(1))} \\
+ \lambda_k^*(1) \sum_{j=b}^{n-1} \frac{(b + \lambda_k(1))\cdots(n - 1 + \lambda_k(1))}{j + \lambda_k(1)} \\
= \left[ (\lambda_k(1))^2 \left\{ \left( \sum_{j=b}^{n-1} \frac{1}{j + \lambda_k(1)} \right)^2 - \frac{1}{(j + \lambda_k(1))^2} \right\} \\
+ \lambda_k^*(1) \sum_{j=b}^{n-1} \frac{1}{j + \lambda_k(1)} \right] \frac{\langle \lambda_k(1) \rangle_n}{\langle \lambda_k(1) \rangle_k}. \]
As in the asymptotic calculation of the mean, manipulating the term $\langle \lambda_k(1) \rangle_n$ by gamma functions yields the asymptotic relation

$$\frac{\pi_{kk}(n)}{(n-1)!} = \frac{(\lambda_k(1))^2}{\Gamma(\lambda_k(1)+b)} n^{\lambda_k(1)} \ln^2 n + O(n^{\lambda_k(1)} \ln n).$$

In particular, from (C.5),

$$\frac{\pi_{11}(n)}{(n-1)!} = \frac{1}{b! H_b^2} n \ln^2 n + O(n \ln n),$$

whereas, for $s = 2, \ldots, b$,

$$\frac{\pi_{ss}(n)}{(n-1)!} = O(n^s \ln^2 n) = O(n \ln n),$$

because $s_k < 1$ (cf. Lemma 4). Recalling that $B_s(1) = (0, 0, \ldots, 0, 1)^T$ and $\Pi_s(n)$ is a diagonal matrix, multiplication yields

$$\frac{(b-1)!}{(n-1)!} R(1) \Pi_s(n) R^{-1}(1) B_s(1) = \frac{(b-1)!}{(n-1)!} \left( \begin{array}{c} \sum_{s=1}^b r_{1s} \pi_{ss}(n) r'_{sb} \\ \vdots \\ \sum_{s=1}^b r_{bs} \pi_{ss}(n) r'_{sb} \end{array} \right) = \begin{pmatrix} r_{11} \ln^2 n + O(n \ln n) \\ \vdots \\ r_{b1} \ln^2 n + O(n \ln n) \end{pmatrix} \frac{r'_{1b}}{b H_b^2}.$$

Collecting all the asymptotic contributions to (5.10), we obtain

$$P_{n+1}(l) = O(\ln n) + \frac{r'_{1b}}{b H_b^2} \ln^2 n \sum_{k=1}^b k r_{k1}.$$

From Lemma 5, the remaining sum is $H_b$; the term $r'_{1b}$ is given in Lemma C.1. Hence,

$$P_n(l) = E[d_n(d_n - 1)] = \frac{1}{H_b^2} \ln^2 n + O(\ln n).$$

A bounding order of magnitude on the variance can now be found from this last relation and (C.6); exact cancellations take place in the order $\ln^2 n$ in the variance of $d_n$, leaving only the order $O(\ln n)$ in $\text{Var}[d_n]$.

References