Asymptotic distribution of two-protected nodes in ternary search trees

Cecilia Holmgren* Svante Janson†

March 21, 2014; revised October 5, 2014

Abstract

We study protected nodes in m-ary search trees, by putting them in context of generalised Pólya urns. We show that the number of two-protected nodes (the nodes that are neither leaves nor parents of leaves) in a random ternary search tree is asymptotically normal. The methods apply in principle to m-ary search trees with larger m as well, although the size of the matrices used in the calculations grow rapidly with m; we conjecture that the method yields an asymptotically normal distribution for all $m \le 26$.

The one-protected nodes, and their complement, i.e., the leaves, are easier to analyze. By using a simpler Pólya urn (that is similar to the one that has earlier been used to study the total number of nodes in m-ary search trees), we prove normal limit laws for the number of one-protected nodes and the number of leaves for all $m \le 26$.

Keywords: Random trees, Pólya urns, Normal limit laws, M-ary search trees.
MSC 2010 subject classifications: Primary 60C05; secondary 05C05, 60F05, 68P05.

1 Introduction

There are many recent studies of so-called protected nodes in various classes of random trees, see e.g. [1, 3, 6, 8, 11, 18, 19]. A node is *protected* (more precisely, two-protected) if it is not a leaf and none of its children is a leaf.

In this paper we consider the number of protected nodes in m-ary search trees (see Section 1.1.2 for definitions), by putting them in context of generalised Pólya urns. The following result is our main theorem. We let $\xrightarrow{d}$ denote convergence in distribution and denote a normal distribution with mean μ and variance σ^2 by $\mathcal{N}(\mu, \sigma^2)$.

Theorem 1.1. *Let Z_n be the number of protected nodes in a ternary search tree with n keys. Then*

$$\frac{Z_n - \frac{57}{700}n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}\left(0, \frac{1692302314867}{43692253605000}\right).$$

*Department of Mathematics, Stockholm University, 114 18 Stockholm, Sweden. Supported in part by the Swedish Research Council.

†Department of Mathematics, Uppsala University, SE-75310 Uppsala, Sweden. Supported in part by the Knut and Alice Wallenberg Foundation.

For a binary search tree, we obtain by the same method a new proof of the following result, which earlier has been obtained by different methods, first by Mahmoud and Ward [18] (using generating functions), and later in [11] (using fringe trees).

Theorem 1.2. *Let Y_n be the number of protected nodes in a binary search tree with n keys. Then*

$$\frac{Y_n - \frac{11}{30}n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}\left(0, \frac{29}{225}\right).$$

Remark 1.3. Theorems 1.1 and 1.2 imply that $\frac{Z_n}{n} \xrightarrow{\text{p}} \frac{57}{700}$ and $\frac{Y_n}{n} \xrightarrow{\text{p}} \frac{11}{30}$. It follows from [12, Theorem 3.21] that moreover, for the sequence of search trees generated by an infinite sequence of i.i.d. keys, $\frac{Z_n}{n} \xrightarrow{\text{a.s.}} \frac{57}{700}$ and $\frac{Y_n}{n} \xrightarrow{\text{a.s.}} \frac{11}{30}$. (Similarly, convergence almost surely holds in the other limit theorems below too.) Since also $0 \le \frac{Z_n}{n} \le 1$ and $0 \le \frac{Y_n}{n} \le 1$, the dominated convergence theorem implies that $(\frac{Z_n}{n})$ and $(\frac{Y_n}{n})$ converge in L^1 to $\frac{57}{700}$ and $\frac{11}{30}$, respectively; in particular, $\frac{\mathbb{E}(Z_n)}{n} \to \frac{57}{700}$ and $\frac{\mathbb{E}(Y_n)}{n} \to \frac{11}{30}$. We conjecture that also the variances (and higher moments) converge in Theorems 1.1 and 1.2.

The methods apply to larger m too, at least in principle, see Sections 1.1.3 and 5.

Similarly, we may consider the one-protected nodes, i.e. the non-leaves. These are easier to analyze than the two-protected nodes and using a minor variation of a Pólya urn earlier used to study the total number of nodes [15, 12, 16], we prove in Sections 4 and 5.2 normal limit laws for the number of one-protected nodes and the number of leaves in an m-ary search tree for all $m \le 26$.

1.1 Protected nodes in m-ary search trees described as generalised Pólya urns

1.1.1 A generalised Pólya urn

A (generalised) Pólya urn process is defined as follows, see e.g. [12] or [16]. There are balls of q types (or colours) $1, \dots, q$, and for each n a random vector $X_n = (X_{n,1}, \dots, X_{n,q})$, where $X_{n,i}$ is the number of balls of type i in the urn at time n. The urn starts with a given vector X_0. For each type i, there is an activity (or weight) $a_i \ge 0$, where $a_i \in \mathbb{R}$, and a random vector $\xi_i = (\xi_{i1}, \dots, \xi_{iq})$, where $\xi_i \in \mathbb{Z}^q_{\ge 0}$. The urn evolves according to a discrete time Markov process. At each time $n \ge 1$, one ball is drawn at random from the urn, with the probability of any ball proportional to its activity. Thus, the drawn ball has type i with probability $\frac{a_i X_{n-1,i}}{\sum_j a_j X_{n-1,j}}$. If the drawn ball has type i, it is replaced together with $\Delta X^{(i)}_{n,j}$ balls of type j, $j = 1, \dots, n$, where the random vector $\Delta X^{(i)}_n = (\Delta X^{(i)}_{n,1}, \dots, \Delta X^{(i)}_{n,q})$ has the same distribution as ξ_i and is independent of everything else that has happened so far. (We allow $\Delta X^{(i)}_{n,i} = -1$, which means that the drawn ball is *not* replaced.) We let A denote the $q \times q$ matrix

$$A = (a_j \, \mathbb{E}\, \xi_{ji})^q_{i,j=1}. \tag{1.1}$$

The matrix A with its eigenvalues and eigenvectors is central for proving limit theorems.

The basic assumptions in [12] are the following. We say that a type i is *dominating*, if every other type j can be found with positive probability at some time in an urn started with a single ball of type i.

(A1) For each type i, there is an integer $d_i \ge 1$, such that X_{0i} and all ξ_{ji} a.s. are divisible by d_i, $\xi_{ij} \ge 0$ for $j \ne i$ (i.e., balls of other types than the drawn ball are never removed) and $\xi_{ii} \ge -d_i$.

(A2) $\mathbb{E}(\xi_{ij}^2) < \infty$ for all $i, j \in \{1, \dots, q\}$.

(A3) The largest real eigenvalue λ_1 of A is positive.

(A4) The largest real eigenvalue λ_1 is simple.

(A5) There exists a dominating type i with $X_{0,i} > 0$, i.e., we start with at least one ball of a dominating type.

(A6) λ_1 is an eigenvalue of the submatrix of A given by the dominating types.

Furthermore, [12] says that the process becomes *essentially extinct* if at some time there are no balls of any dominating type left. We will also use the following simplifying assumption.

(A7) With probability 1, the urn never becomes essentially extinct.

Condition (A1) is stated here somewhat more generally than in [12], where $d_i = 1$ is assumed, but the general case follows by replacing $X_{n,i}$ by $\frac{X_{n,i}}{d_i}$; see [12, Remark 4.2].

In the Pólya urns used in this paper, it is easily seen (from the definitions using trees) that every type with non-zero activity is dominating. If we remove rows and columns corresponding to the types with activity 0 from A, then the removed columns are identically 0, so the set of non-zero eigenvalues of A is not changed. The remaining matrix is irreducible, and using the Perron–Frobenius theorem, it is easy to verify all conditions (A1)–(A6), see [12, Lemma 2.1]. Furthermore, in our urns there will always be a ball of positive activity, so essential extinction is impossible.

Before stating the results that we use, we need some notation. With a vector v we mean a column vector, and we write v' for its transpose (a row vector). More generally, we denote the transpose of a matrix A by A'. By an eigenvector of A we mean a right eigenvector; a left eigenvector is the same as the transpose of an eigenvector of the matrix A'. If u and v are vectors then $u'v$ is a scalar while uv' is a $q \times q$ matrix of rank 1. We also use the notation $u \cdot v$ for $u'v$. We let λ_1 denote the largest real eigenvalue of A. (This exists by our assumptions and the Perron–Frobenius theorem.) Let $a = (a_1, \dots, a_q)$ denote the (column) vector of activities, and let u_1' and v_1 denote left and right eigenvectors of A corresponding to the largest eigenvalue λ_1, i.e., vectors satisfying

$$u_1'A = \lambda_1 u_1', \qquad\qquad Av_1 = \lambda_1 v_1.$$

We assume that v_1 and u_1 are normalized such that

$$a \cdot v_1 = a'v_1 = v_1'a = 1, \qquad\qquad u_1 \cdot v_1 = u_1'v_1 = v_1'u_1 = 1, \tag{1.2}$$

see [12, equations (2.2)–(2.3)]. We write $v_1 = (v_{11}, \dots, v_{1q})$.

We define

$$P_{\lambda_1} = v_1 u_1',$$

and $P_I = I_q - P_{\lambda_1}$, where I_q is the $q \times q$ identity matrix. (Thus P_{λ_1} is the one-dimensional projection onto the eigenspace corresponding to λ_1 such that P_{λ_1} commutes with the matrix A, see [12, equation (2.2)]; note that P_{λ_1} typically is not orthogonal). We define the matrices

$$B_i := \mathbb{E}(\xi_i \xi_i') \tag{1.3}$$

$$B := \sum_{i=1}^{q} v_{1i} a_i B_i \tag{1.4}$$

$$\Sigma_I := \int_0^\infty P_I e^{sA} B e^{sA'} P_I' e^{-\lambda_1 s} ds, \tag{1.5}$$

where we recall that $e^{tA} = \sum_{j=0}^{\infty} t^j A^j / j!$.

It is proved in [12] that, under assumptions (A1)–(A7), X_n is asymptotically normal if $\operatorname{Re}\lambda \le \lambda_1/2$ for each eigenvalue $\lambda \neq \lambda_1$; more precisely, if $\operatorname{Re}\lambda < \lambda_1/2$ for each such λ, then $n^{-1/2}(X_n - n\mu) \xrightarrow{d} \mathcal{N}(0,\Sigma)$ for some $\mu = (\mu_1,\dots,\mu_k)$ and $\Sigma = (\sigma_{i,j})_{i,j=1}^k$. (If $\lambda = \lambda_1/2$, then X_n is still asymptotically normal, however with another normalisation.) The asymptotic covariance matrix Σ may be calculated in different ways; we use the following results from [12], which apply under different additional assumptions.

Theorem 1.4 ([12, Theorem 3.22 and Lemma 5.4])**.** *Assume* (A1)–(A7) *and that we have normalized as in* (1.2)*. Also assume that* $\operatorname{Re}\lambda < \lambda_1/2$ *for each eigenvalue* $\lambda \neq \lambda_1$*. Suppose that* $a \cdot \mathbb{E}(\xi_i) = m$ *for some* $m > 0$ *and every* i*. Then, as* $n \to \infty$*,*

$$n^{-1/2}(X_n - n\mu) \xrightarrow{d} \mathcal{N}(0,\Sigma),$$

with $\mu = \lambda_1 v_1$ *and covariance matrix* Σ *equal to* $m\Sigma_I$*, with* Σ_I *as in* (1.5)*.* □

Theorem 1.5 ([12, Theorem 3.22 and Lemma 5.3])**.** *Assume* (A1)–(A7)*, and that we have normalized as in* (1.2)*. Also assume that* $\operatorname{Re}\lambda < \lambda_1/2$ *for each eigenvalue* $\lambda \neq \lambda_1$*. If the matrix* A *is diagonalisable, and* $\{u_i'\}_{i=1}^q$ *and* $\{v_i\}_{i=1}^q$ *are dual bases of left and right eigenvectors, respectively, i.e.,* $u_i' A = \lambda_i u_i'$*,* $Av_i = \lambda_i v_i$ *and* $u_i \cdot v_j = \delta_{ij}$ *(where* δ_{ij} *is the Kronecker delta). Then, as* $n \to \infty$*,*

$$n^{-1/2}(X_n - n\mu) \xrightarrow{d} \mathcal{N}(0,\Sigma),$$

with $\mu = \lambda_1 v_1$ *and covariance matrix* Σ *equal to*

$$\Sigma = \sum_{j,k=2}^{q} \frac{u_j' B u_k}{\lambda_1 - \lambda_j - \lambda_k} v_j v_k', \tag{1.6}$$

with the matrix B *as in* (1.4)*.* □

1.1.2 M-ary search trees

We recall the definition of m-ary search trees, see e.g. [14] or [7]. An m-ary search tree, where $m \ge 2$, is constructed recursively from a sequence of n *keys* (numbers). We assume that the keys are i.i.d. uniform random numbers in $[0,1]$. (Only the order of the keys matter, so alternatively, we may assume that the keys form a uniformly random permutation of $\{1,\dots,n\}$.) Each node may contain up to $m-1$ keys. We start with a tree containing just an empty root. The first $m-1$ keys are put in the root, and are placed in increasing order from left to right; they divide the set of real numbers into m intervals $J_1,\dots,J_m$. When the root is full (after the first $m-1$ keys are added), it gets m children that are initially empty, and each further key is passed to one of the children depending on which interval it belongs to; a key in J_i is passed to the i:th child. (The binary search tree is the simplest case where keys are passed to the left or right child depending on whether it is larger or smaller than the key in the root.) The procedure repeats recursively in the subtrees until all keys are added to the tree.

Nodes that contain at least one key are called *internal*, while empty nodes are called *external*. We regard the m-ary search tree as consisting only of the internal nodes; the external nodes are places for potential additions, and are useful when discussing the tree

(e.g. below), but are not really part of the tree. Thus, a *leaf* is an internal node that has no internal children, but it may have external children. (It will have external children if it is full, but not otherwise.) Similarly, a protected node is an internal node that is not a leaf, and has no child that is a leaf. (It may have external nodes as children.)

We say that a node with $i \le m-2$ keys has $i+1$ *gaps*, while a full node has no gaps. It is easily seen that a m-ary search tree with n keys has $n+1$ gaps; the gaps correspond to the intervals of real numbers between the keys (and $\pm\infty$), and a new key has the same probability $1/(n+1)$ of belonging to any of the gaps. Thus the evolution of the m-ary search tree may be described by choosing a gap uniformly at random at each step. Equivalently, the probability that the next key is added to a node is proportional to the number of gaps at that node.

Pólya urns have been used in some earlier studies, e.g. [15, 12], to describe the number of nodes in m-ary search trees containing i keys where $0 \le i \le m-1$; then a node containing i keys is called a node of type i and thus the generalised Pólya urn has m different types. It has been shown that for this process, when $m \le 26$ the number of different types has an asymptotic multivariate normal distribution, but this does not hold for larger m. (Since the condition $\operatorname{Re}\lambda < \lambda_1/2$ for $\lambda \neq \lambda_1$ on the eigenvalues of the matrix A in (1.1) holds only if $m \le 26$.) Since the number of nodes in the whole tree is a linear combination of these numbers, this implies in particular that the distribution of the random number of nodes in an m-ary search tree containing n keys is asymptotically normal for $m \le 26$. In this Pólya urn, with one ball representing each node, the activity of a ball is the number of gaps, i.e., $i+1$ for a ball of type $i \le m-2$, and 0 for a ball of type $m-1$.

Alternatively, see [12], we can use a Pólya urn where each ball represents a gap; thus a node with i keys corresponds to $i+1$ balls for $0 \le i \le m-2$, and these balls are all given type i. (Full nodes are ignored.) This is thus an urn with $m-1$ types, all with activities 1.

1.1.3 Protected nodes and generalised Pólya urns

We will see that it is possible to use a generalised Pólya urn also to study protected nodes in an m-ary search tree, although the urn consists of quite a few different types.

Description of the Types in the Pólya urn. Given an m-ary search tree T with n keys together with its external nodes, erase all edges that connect two internal non-leaves. This yields a forest of small trees, where (assuming $n \ge m$) each tree has a root that is a non-leaf in T while all other nodes are leaves or external nodes in T. We regard these small trees as the balls in our generalised Pólya urn. The type of a ball (tree) is the type of the tree as an unordered tree, i.e., up to permutations of the children. The type of a tree in the urn is thus described by the numbers k_i, $i = 0, \dots, m-1$, of children of the root with i keys; each of these children is an external node ($i = 0$) or a leaf ($i \ge 1$), and it has itself children only when $i = m-1$ when it has m external children; thus the type is uniquely determined by $k_0, \dots, k_{m-1}$, and we can label the type by $(k_0, \dots, k_{m-1})$. Since the root of any of the small trees has m children (including external ones) in the original tree T, we have $\sum_{i=0}^{m-1} k_i \le m$, (with the remainder $m - \sum_{i=0}^{m-1} k_i$ equal to the number of erased edges to children in the original tree T that are non-leaves). Furthermore, the case $k_0 = m$ is excluded, since the root of the small tree is a non-leaf in T. The total number of types is thus one less than the number of compositions of m into $m+1$ non-negative parts, i.e., $\binom{2m}{m} - 1$.

The activity in the Pólya urn of one of these types is the number of gaps that it contains.

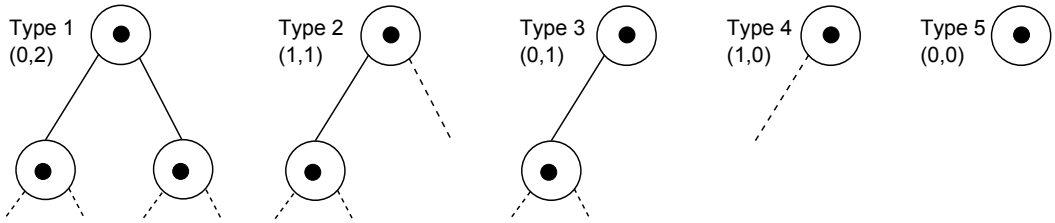

Figure 1: The different types characterizing protected and unprotected nodes in binary search trees. Type 4 and type 5 are the only ones that include protected nodes.

The root has no gaps, so a tree with type $(k_0, \dots, k_{m-1})$ has activity $\sum_{i=0}^{m-1}(i+1)k_i$. Moreover, if we add a new key to a leaf, it is still a leaf, so in the Pólya urn, this corresponds to replacing a tree by another tree where we have increased by 1 the number of keys of one of the children of the root. The same holds if we add a key to an external node that is a child of the root. However, if we add a key to an external node that is a child of a leaf, then that leaf becomes a non-leaf, so the edge from it to the root is erased and the tree is split into two (one of which always has the type $(m-1, 1, 0, \dots, 0)$). See Section 2 for examples. Note that in general, a small tree may be transformed in several different ways when we add a new key, depending on which gap it goes into. Hence, the additions ξ_i in the Pólya urn will be random.

A protected node in T is a non-leaf, and is therefore a root in one of the small trees. Moreover, it must not have any child that is a leaf, so all its children are external nodes. Thus, the number of protected nodes in T equals the number of balls in the urn that have types $(k_0, 0, \dots, 0)$ with $0 \le k_0 \le m-1$.

2 Protected nodes in binary search trees and Pólya urns

In this section we demonstrate the technique of using the Pólya urn defined above to study the number of protected nodes, by applying it to the simplest case $m = 2$, the binary search tree. This gives us a new proof of Theorem 1.2; for earlier proofs, see [18] and [11].

For a binary tree, the number of types in the Pólya urn defined above is $\binom{4}{2} - 1 = 5$. We show the different types in Figure 1, with a numbering that will be used below. (For convenience we omit the external nodes in the figures. We use dotted lines for edges attached to external nodes.) With our characterization of the types in Section 1.1.3, the types $i \in \{1, \dots, 5\}$ correspond to $(0,2)$, $(1,1)$, $(0,1)$, $(1,0)$ and $(0,0)$, respectively.

Let $X_n = (X_{n,1}, X_{n,2}, X_{n,3}, X_{n,4}, X_{n,5})$, where $X_{n,i}$ is the number of balls of type i in the urn corresponding to n keys (i.e., the number of trees that correspond to type i in our forest). Recall that we assume that $n \ge m = 2$; the initial conditions are $X_{2,2} = 1$ and $X_{2,i} = 0$ for $i \ne 2$. In a binary search tree, each leaf contains one key, so it has two external children, whereas other internal nodes have either 1 or 0 external children. There is one gap at each external node, and no gaps at any internal node. As explained in Section 1.1.2, each gap (i.e. external node) has activity 1.

When a ball is drawn from the urn (i.e., a new key is added to the tree), as explained in general in Section 1.1.3, a key is either added to an external node that is a child of the root (we return a ball of another type), or to an external node that is a child of a leaf (we return two balls). Figures 2–5 show the transitions in the Pólya urn when a ball of type i for

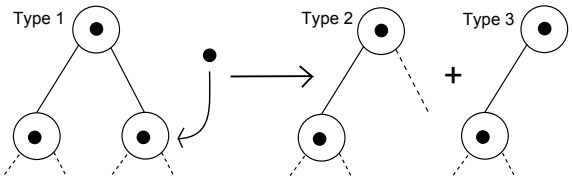

Figure 2: Adding a key to type 1.

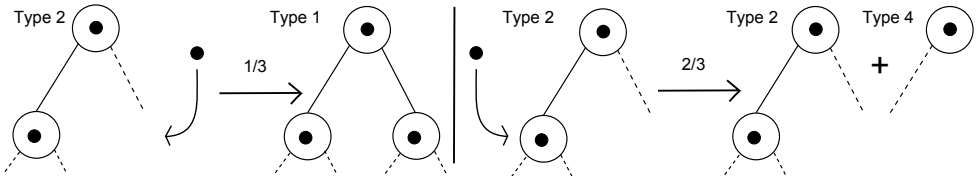

Figure 3: Adding a key to type 2.

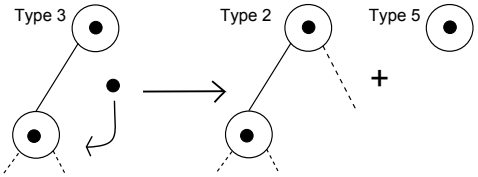

Figure 4: Adding a key to type 3

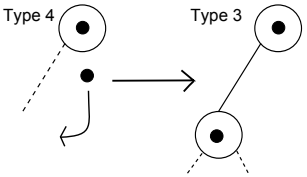

Figure 5: Adding a key to type 4

$i \in \{1,2,3,4\}$ is drawn (where the types are shown in Figure 1), so that the drawn ball is replaced by a new set of balls. (As said above, this set could depend on which of the nodes in the drawn type the key is added to, see Figure 3.) The activities of the different types depend on their number of gaps; the total activities for the types $1,2,3,4,5$ are $4,3,2,1,0$, respectively; thus $a = (4,3,2,1,0)'$.

From the transitions that are shown in Figures 2–5, we easily obtain the matrix $A = (a_j \, \mathbb{E}\, \xi_{ji})_{i,j=1}^{5}$ in (2.1).

$$A = \begin{pmatrix} -4 & 1 & 0 & 0 & 0 \\ 4 & -1 & 2 & 0 & 0 \\ 4 & 0 & -2 & 1 & 0 \\ 0 & 2 & 0 & -1 & 0 \\ 0 & 0 & 2 & 0 & 0 \end{pmatrix} \tag{2.1}$$

To do the matrix operations in this paper we use computer algebra (in our case Mathematica).

The eigenvalues of A are $1, 0, -2, -3, -4$. Corresponding right eigenvectors of A are:

$$\frac{1}{30}\begin{pmatrix}1\\5\\3\\5\\6\end{pmatrix}, \begin{pmatrix}0\\0\\0\\0\\1\end{pmatrix}, \frac{1}{3}\begin{pmatrix}1\\2\\-3\\-4\\3\end{pmatrix}, \frac{1}{2}\begin{pmatrix}1\\1\\-3\\-1\\2\end{pmatrix}, \frac{1}{5}\begin{pmatrix}1\\0\\-2\\0\\1\end{pmatrix}, \tag{2.2}$$

and corresponding left eigenvectors of A are:

$$\begin{pmatrix}4\\3\\2\\1\\0\end{pmatrix}', \begin{pmatrix}-1\\-1\\0\\0\\1\end{pmatrix}', \begin{pmatrix}2\\0\\1\\-1\\0\end{pmatrix}', \begin{pmatrix}-4\\1\\-2\\1\\0\end{pmatrix}', \begin{pmatrix}11\\-3\\3\\-1\\0\end{pmatrix}'. \tag{2.3}$$

Since the eigenvalues for the matrix A are distinct it follows automatically that $u_i \cdot v_j = 0$ for $i \neq j$ (recalling that $\{u_i'\}_{i=1}^q$ and $\{v_i\}_{i=1}^q$ are the left and right eigenvectors of A, respectively). Note that we have scaled the eigenvectors so that $u_i \cdot v_i = 1$ and (1.2) hold. Note also that u_1 is equal to the activity vector a. This is a consequence of the fact that the total activity always increases by 1 when we draw a ball from the urn, and thus $a \cdot \mathbb{E}\,\xi_i = 1$ for each i, see [12, Lemma 5.4].

It is easy to see that we can apply Theorem 1.5 for this generalised Pólya urn. Note that it is obvious that the matrix A is diagonalisable since all eigenvalues are simple. From Theorem 1.5 we obtain that $X_n = (X_{n,1}, X_{n,2}, X_{n,3}, X_{n,4}, X_{n,5})$ has asymptotically a multivariate normal distribution. Let Y_n be equal to the number of protected nodes in the binary search tree with n nodes. Since type 4 and type 5 each contains exactly one protected node, while the other types contain no protected nodes,

$$Y_n = X_{n,4} + X_{n,5}.$$

Thus, Theorem 1.5 implies that

$$n^{-1/2}(Y_n - n\mu_Y) \xrightarrow{d} \mathcal{N}(0, \sigma_Y^2) \tag{2.4}$$

with parameters $\mu_Y = \mu_4 + \mu_5$ and

$$\sigma_Y^2 = \sigma_{4,4} + \sigma_{4,5} + \sigma_{5,4} + \sigma_{5,5}. \tag{2.5}$$

Since $\lambda_1 = 1$, Theorem 1.5 implies, using v_1 in (2.2), that

$$\mu_Y = \mu_4 + \mu_5 = \frac{5}{30} + \frac{6}{30} = \frac{11}{30}. \tag{2.6}$$

Thus, to show Theorem 1.2 it remains to calculate the sum in (2.5).

To calculate the matrix B in (1.4) we need to calculate $B_i = \mathbb{E}(\xi_i \xi_i')$ in (1.3). In all cases except for B_2 these are deterministic and equal to $\xi_i \xi_i'$. We only show how to obtain B_2 (since the other cases are simpler). As shown in Figure 3 when adding a key to type 2 we can either add it to the leaf or to the external node. In case we add it to the external node (which happens with probability 1/3) a node of type 2 is replaced by a node of type 1; this change corresponds to the column vector $(1, -1, 0, 0, 0)'$. If the key is instead added to

the leaf (which happens with probability 2/3) a node of type 2 is replaced by another node of type 2 (the change of type 2 is 0) and a node of type 4; this change corresponds to the column vector $(0,0,0,1,0)'$. Hence

$$B_2 = \tfrac{1}{3}\cdot(1,-1,0,0,0)'(1,-1,0,0,0) + \tfrac{2}{3}\cdot(0,0,0,1,0)'(0,0,0,1,0)$$
$$= \begin{pmatrix} \frac{1}{3} & -\frac{1}{3} & 0 & 0 & 0 \\ -\frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{2}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{2.7}$$

By calculating the B_i's we obtain the matrix B in (1.4) as

$$B = \begin{pmatrix} \frac{3}{10} & -\frac{3}{10} & -\frac{2}{15} & 0 & 0 \\ -\frac{3}{10} & \frac{1}{2} & -\frac{1}{15} & 0 & \frac{1}{5} \\ -\frac{2}{15} & -\frac{1}{15} & \frac{1}{2} & -\frac{1}{6} & -\frac{1}{5} \\ 0 & 0 & -\frac{1}{6} & \frac{1}{2} & 0 \\ 0 & \frac{1}{5} & -\frac{1}{5} & 0 & \frac{1}{5} \end{pmatrix}. \tag{2.8}$$

From (1.6) in Theorem 1.5 it follows that the covariance matrix Σ for the asymptotic multivariate normal distribution of $X_n = (X_{n,1}, X_{n,2}, X_{n,3}, X_{n,4}, X_{n,5})$, is given by

$$\Sigma = \begin{pmatrix} \frac{43}{1575} & -\frac{67}{2520} & -\frac{113}{12600} & -\frac{29}{2520} & \frac{1}{1400} \\ -\frac{67}{2520} & \frac{23}{420} & -\frac{1}{42} & -\frac{13}{1260} & \frac{71}{2520} \\ -\frac{113}{12600} & -\frac{1}{42} & \frac{443}{6300} & -\frac{1}{30} & -\frac{59}{1800} \\ -\frac{29}{2520} & -\frac{13}{1260} & -\frac{1}{30} & \frac{181}{1260} & -\frac{11}{504} \\ \frac{1}{1400} & \frac{71}{2520} & -\frac{59}{1800} & -\frac{11}{504} & \frac{13}{450} \end{pmatrix}. \tag{2.9}$$

Thus, it follows that

$$\sigma_Y^2 = \sigma_{4,4} + \sigma_{4,5} + \sigma_{5,4} + \sigma_{5,5} = \frac{181}{1260} + \frac{13}{450} - 2\cdot\frac{11}{504} = \frac{29}{225}. \tag{2.10}$$

Thus, the proof of Theorem 1.2 is completed. □

3 Protected nodes in ternary search trees and Pólya urns

We now proceed by analyzing the number of protected nodes in ternary search trees, by using the Pólya urn in Section 1.1.3 (described for general m-ary search trees) when $m = 3$. The 19 different types we get are shown in Figure 6 (with a numbering that will be used below). From our characterization of the types in Section 1.1.3, for example type 2 corresponds to (0,1,2). Note that type 17, type 18 and type 19 contain one protected node each, while the other types contain no protected nodes.

To determine the matrix A we proceed (as for the binary search tree) to find the transitions when a ball (in our case one of the 19 trees in our forest) of type i is chosen. Figure

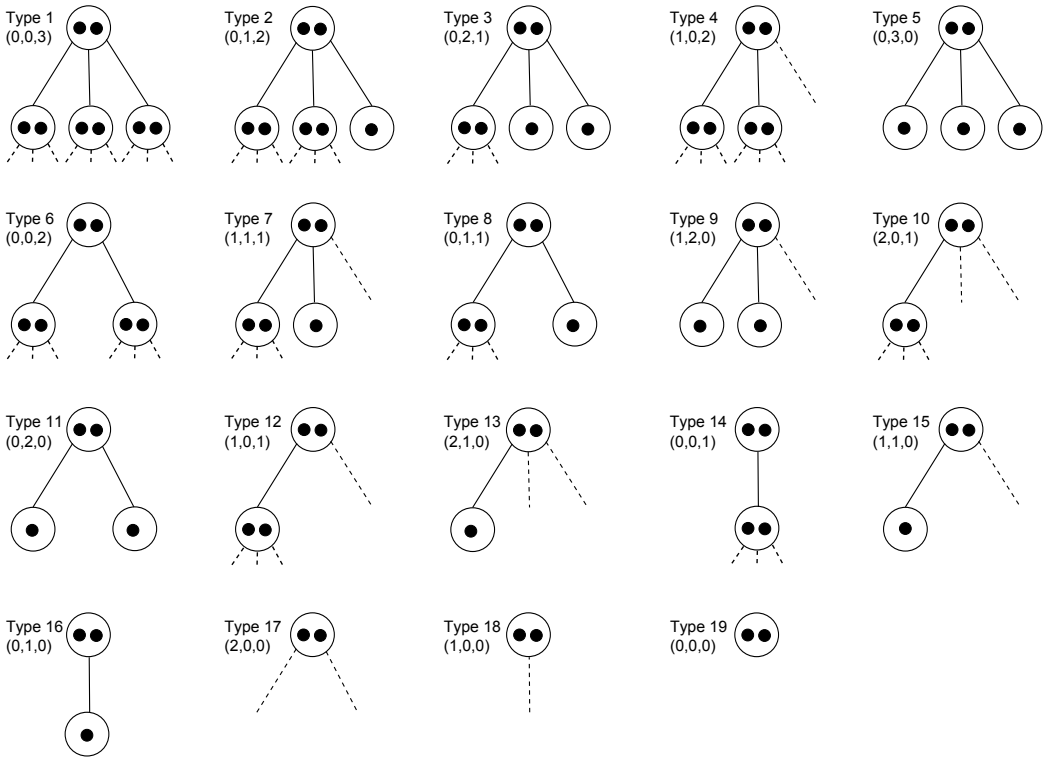

Figure 6: The different types characterizing protected and unprotected nodes in ternary search trees. Type 17, type 18 and type 19 are the only ones that include protected nodes.

7 illustrates the different situations for how a new key could be added to a ball (a tree) of type 2. All the other cases are similar, and we leave these cases as an exercise to the reader.

From the different transitions for changing a node of type i we get the matrix A for ternary search trees in Figure 8. The example in Figure 7 gives the second column of A.

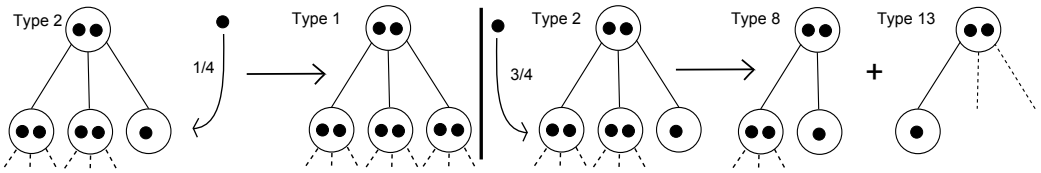

Figure 7: The two possibilities for adding a key to a node in a tree of type 2 of a ternary search tree.

The tree of type 2 has activity 8. If it is drawn, and the new key is added to the node with only one key which happens with probability $\frac{2}{8}$, then a tree of type 2 is replaced with a tree of type 1. If the new key is instead added to one of the nodes containing two keys which happens with probability $\frac{6}{8}$, then the tree of type 2 is replaced by a tree of type 8 and one tree of type 13. Thus, the second column of the matrix A for the ternary search tree is given by

$$8 \cdot (\tfrac{2}{8}, -1, 0, 0, 0, 0, 0, \tfrac{6}{8}, 0, 0, 0, 0, \tfrac{6}{8}, 0, 0, 0, 0, 0, 0)'.$$

In this way we obtain A in Figure 8.

$$A = \begin{pmatrix}
-9 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -8 & 4 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -7 & 0 & 6 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -7 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -6 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
9 & 0 & 0 & 0 & 0 & -6 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -6 & 0 & 4 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 6 & 0 & 0 & 0 & 0 & 0 & -5 & 0 & 0 & 4 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -5 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -5 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -4 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -4 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\
9 & 6 & 3 & 6 & 0 & 6 & 3 & 3 & 0 & 3 & 0 & 3 & -4 & 3 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -3 & 0 & 2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -3 & 0 & 2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}$$

Figure 8: The transition matrix A for the Pólya urn defined in Section 1.1.3 in the case of the ternary search tree.

The activities of the different types are given by the vector

$$a = (9, 8, 7, 7, 6, 6, 6, 5, 5, 5, 4, 4, 4, 3, 3, 2, 2, 1, 0)'.$$

These correspond to the number of gaps for the different types. The eigenvalues of the matrix A are

$$1, 0, -2, -3, -3, -4, -4, -4, -4, -5, -5, -5, -6, -6, -6, -7, -7, -8, -9.$$

The eigenspace belonging to the eigenvalue -4 (which has algebraic multiplicity 4) has dimension 3. Since the dimension of the eigenspace belonging to the eigenvalue -4 is not equal to the algebraic multiplicity, the matrix A is not diagonalisable. (However, all other eigenspaces have full dimension.) Hence, we can not apply Theorem 1.5. However, Theorem 1.4 can be applied since $a \cdot \mathbb{E}(\xi_i) = 1$ for each i (this follows since we always add exactly one key when a tree of type i is chosen).

From Theorem 1.4 we obtain that the vector $X_n = (X_{n,1}, \ldots, X_{n,19})$, where $X_{n,i}$ are the number of balls of type i (in our case the number of trees that correspond to type i in our forest obtained from the ternary search tree), has asymptotically a multivariate normal distribution. Let Z_n be the number of protected nodes in the ternary search tree with n nodes. Since type 17, type 18 and type 19 each contains exactly one protected node, while the other types contain no protected nodes,

$$Z_n = X_{n,17} + X_{n,18} + X_{n,19}. \tag{3.1}$$

Thus, Theorem 1.4 implies that

$$n^{-1/2}(Z_n - n\mu_Z) \xrightarrow{d} \mathcal{N}(0, \sigma_Z^2), \tag{3.2}$$

with parameters

$$\mu_Z = \mu_{17} + \mu_{18} + \mu_{19}$$

and, writing $\Sigma = (\sigma_{i,j})_{i,j=1}^{19}$,

$$\sigma_Z^2 = \sum_{i=17}^{19} \sum_{j=17}^{19} \sigma_{i,j}. \tag{3.3}$$

Using the normalization in (1.2), we see that

$$v_1 = \frac{1}{2100} \cdot (1, 5, 9, 9, 6, 7, 36, 20, 42, 42, 15, 30, 126, 28, 48, 35, 42, 45, 84)' \tag{3.4}$$

and that

$$u_1 = (9, 8, 7, 7, 6, 6, 6, 5, 5, 5, 4, 4, 4, 3, 3, 2, 2, 1, 0)'.$$

(As in the binary case, $u_1 = a$ since $a \cdot \mathbb{E}\,\xi_i = 1$ for each i, see [12, Lemma 5.4].) Since $\lambda_1 = 1$, Theorem 1.4 and (3.4) yield

$$\mu_Z = \mu_{17} + \mu_{18} + \mu_{19} = \frac{42}{2100} + \frac{45}{2100} + \frac{84}{2100} = \frac{57}{700}. \tag{3.5}$$

Thus, to show Theorem 1.1 it remains to calculate the sum in (3.3).

Since we want to determine the matrix Σ_I in (1.5) we need to determine the matrices P_I and B. We have $P_I = I_{19} - v_1 u_1'$, which is a 19×19 matrix that is shown in (A.1) in the appendix. To calculate the matrix B in (1.4) we need to calculate $B_i = \mathbb{E}(\xi_i \xi_i')$ in (1.3). We only describe how to get B_2 since the other cases are analogous. From Figure 7 (and the explanation of that figure above) it is easy to see that

$$\begin{aligned} B_2 &= \tfrac{1}{4} \cdot b_1 b_1' + \tfrac{3}{4} \cdot b_2 b_2', \quad \text{where,} \\ b_1 &= (1, -1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)' \text{ and} \\ b_2 &= (0, -1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0)'. \end{aligned}$$

Note that B_2 is a 19×19 matrix. The matrix B is shown in (A.2) in the appendix. Now we can use Mathematica to evaluate the integral in (1.5), which yields Σ_1. Finally, $\Sigma = \Sigma_1$ by Theorem 1.4 with $m = 1$. This matrix is given last in the appendix.

By (3.1) and (3.3), we only need the submatrix

$$\Sigma_p = \begin{pmatrix} \sigma_{17,17} & \sigma_{17,18} & \sigma_{17,19} \\ \sigma_{18,17} & \sigma_{18,18} & \sigma_{18,19} \\ \sigma_{19,17} & \sigma_{19,18} & \sigma_{19,19} \end{pmatrix} = \begin{pmatrix} \frac{156031}{8085000} & -\frac{826069}{1387386000} & \frac{3453169}{15030015000} \\ -\frac{826069}{1387386000} & \frac{2222557}{118918800} & -\frac{439517549}{87603516000} \\ \frac{3453169}{15030015000} & -\frac{439517549}{87603516000} & \frac{142536826}{12384425625} \end{pmatrix}. \tag{3.6}$$

Summing the $\sigma_{i,j}$ in (3.6), which is equivalent to calculating $(1, 1, 1)\Sigma_p(1, 1, 1)'$, we find

$$\sigma_Z^2 = \sum_{i=17}^{19} \sum_{j=17}^{19} \sigma_{i,j} = \frac{1692302314867}{43692253605000},$$

which completes the proof of Theorem 1.1. □

a) An external node which is not a child of a leaf.

b) A leaf containing one key.

c) A leaf containing two keys and its three external children.

d) An internal node with two keys which is not a leaf.

Figure 9: The different types characterizing leaves and non-leaves in ternary search trees.

4 Leaves in ternary search trees

Recall that a leaf is an internal node without internal children, i.e., a node that contains at least one key and has no children except possibly external ones. The proof of Theorem 1.1 yields also the following theorem. (The corresponding result for a binary search tree was considered already by Devroye [5] using two different methods, one of them a Pólya urn as here.)

Theorem 4.1. *Let L_n be the number of leaves in a ternary search tree. Then,*

$$\frac{L_n - \frac{3}{10}n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}\Big(0, \frac{89}{2100}\Big).$$

First proof. Counting the number of leaves (of the original ternary search tree) in each type in Figure 6, we see that the number of leaves in a subtree of type i, $i = 1, \dots, 19$, is given by the vector

$$\ell = (3, 3, 3, 2, 3, 2, 2, 2, 2, 1, 2, 1, 1, 1, 1, 1, 0, 0, 0)'. \tag{4.1}$$

Hence, $L_n = \ell \cdot X_n$. By the proof of Theorem 1.1, the vector X_n has asymptotically a multivariate normal distribution, and it follows that

$$n^{-1/2}(L_n - n\mu_L) \xrightarrow{d} \mathcal{N}(0, \sigma_L^2) \tag{4.2}$$

with, using (3.4) and (4.1),

$$\mu_L = \ell \cdot v_1 = \frac{3}{10}, \tag{4.3}$$

and, using the covariance matrix Σ shown in the appendix,

$$\sigma_L^2 = \ell' \, \Sigma \, \ell = \frac{89}{2100}. \tag{4.4}$$

□

However, it is also possible to show Theorem 4.1 using a much simpler Pólya urn process, where we only need to consider four different types. We again chop up the ternary search tree into small trees, now using the following types of trees.

Type 1 is an external node which is not a child of a leaf. Type 2 is a node containing one key. Type 3 is a leaf containing two keys together with its three external children. Type 4 is an internal node containing two keys which is not a leaf (i.e., it has less than three external children). The types are shown in Figure 9. Note that all nodes in the ternary search tree belong to exactly one such subtree.

A ball of type 1 has activity 1; when it is drawn it is replaced by one ball of type 2. A ball of type 2 has activity 2; when it is drawn it is replaced by one ball of type 3. A ball of type 3 has activity 3; when it is drawn it is replaced by one ball of type 2, two balls of type 1 and one ball of type 4. A ball of type 4 has activity 0 and is thus never drawn. The types that contain leaves are type 2 and type 3.

To simplify we can study another urn using the gaps as balls. Type 1 has one gap, type 2 has two gaps, type 3 has three gaps and type 4 has 0 gaps. We label each gap with the type it belongs to; thus the gaps have only the three types 1–3. The gaps evolve as an urn with three types, with all activities 1 and the matrix A in (1.1) given by

$$\begin{pmatrix} -1 & 0 & 2 \\ 2 & -2 & 2 \\ 0 & 3 & -3 \end{pmatrix}. \tag{4.5}$$

Since we consider the gaps (with activity 1) it is obvious that all columns add to 1 (since we always add one ball to the urn). The eigenvalues of A are $1, -3, -4$. Theorem 1.5 shows that $(X_{n,1}, X_{n,2}, X_{n,3})$ has asymptotically a multivariate normal distribution, where $X_{n,i}$ is the number of balls of type i in the Pólya urn, i.e., the number of gaps of type i. Note that the number of subtrees of Types 1–3 thus is $(X_{n,1}, X_{n,2}/2, X_{n,3}/3)$, which thus also is asymptotically multivariate normal.

Since the number of leaves $L_n = X_{n,2}/2 + X_{n,3}/3$, it follows that L_n has asymptotically a normal distribution (4.2).

To find the parameters μ_L and σ_L^2, we note that right eigenvectors of A corresponding to the eigenvalues $1, -3, -4$ are:

$$\frac{1}{10}\begin{pmatrix} 3 \\ 4 \\ 3 \end{pmatrix}, \frac{1}{2}\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \frac{1}{5}\begin{pmatrix} -2 \\ -1 \\ 3 \end{pmatrix}, \tag{4.6}$$

and corresponding left eigenvectors of A are:

$$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}', \begin{pmatrix} -3 \\ 3 \\ -1 \end{pmatrix}', \begin{pmatrix} 2 \\ -3 \\ 2 \end{pmatrix}'. \tag{4.7}$$

Note that we have scaled the eigenvectors so that $u_i \cdot v_j = \delta_{ij}$ and (1.2) holds. We have $a = (1, 1, 1)'$. Since type 2 has two gaps and one leaf and type 3 has three gaps and one leaf, it follows that

$$\mu_L = \mu_2 + \mu_3 = \frac{1}{10}(3, 4, 3) \cdot \Big(0, \frac{1}{2}, \frac{1}{3}\Big) = \frac{3}{10},$$

corresponding to (4.3). By calculating B, we get from Theorem 1.5, that the covariance

matrix Σ is given by

$$\begin{pmatrix} \frac{479}{2100} & -\frac{7}{150} & -\frac{127}{700} \\ -\frac{7}{150} & \frac{32}{75} & -\frac{19}{50} \\ -\frac{127}{700} & -\frac{19}{50} & \frac{393}{700} \end{pmatrix}. \tag{4.8}$$

We thus obtain

$$\sigma_L^2 = \left(0, \tfrac{1}{2}, \tfrac{1}{3}\right) \begin{pmatrix} \frac{479}{2100} & -\frac{7}{150} & -\frac{127}{700} \\ -\frac{7}{150} & \frac{32}{75} & -\frac{19}{50} \\ -\frac{127}{700} & -\frac{19}{50} & \frac{393}{700} \end{pmatrix} \begin{pmatrix} 0 \\ \frac{1}{2} \\ \frac{1}{3} \end{pmatrix} = \frac{89}{2100} \tag{4.9}$$

(corresponding to (4.4)), which completes the proof of Theorem 4.1 with the simpler Pólya urn model.

5 Higher m

5.1 The Pólya urn defined in Section 1.1.3

The Pólya urn defined in Section 1.1.3 can be used for any given m, although the size of the matrices used in the calculations grow rapidly with m. (For $m = 4$ we have 69 types; for $m = 10$ we would have 184755.) However, the central condition $\operatorname{Re}\lambda < \lambda_1/2$ is not satisfied for large m. We do not know any general formula for the eigenvalues of the matrix A, but some of them are given as follows.

Lemma 5.1. *Let $m \geq 2$. Then every root of the polynomial*

$$\phi_m(\lambda) := \prod_{i=1}^{m-1} (\lambda + i) - m! \tag{5.1}$$

is an eigenvalue of the matrix A for the Pólya urn in Section 1.1.3.

Proof. Let $M := \binom{2m}{m} - 1$ be the number of types, and let as above $X_n \in \mathbb{Z}_{\geq 0}^M$ be the composition of the Pólya urn described in Section 1.1.3. Furthermore, let $V_{i,n}$ be the number of nodes containing exactly i keys (thus $V_{0,n}$ is the number of external nodes), and consider the vector $W_n = (W_{1,n}, \dots, W_{m-1,n})$ where $W_{i,n} = iV_{i-1,n}$; thus $W_{i,n}$ is the total number of gaps at nodes with i gaps. The random vector W_n can also be described by a Pólya urn, see e.g., [12, Example 7.8] and [16, Section 8.1.3]; we denote the activity vector and the matrix (1.1) for this urn by $a_W = (1, \dots, 1)'$ and A_W, where the $(m-1) \times (m-1)$ matrix A_W has elements $a_{i,i} = -i$ for $i \in \{1, \dots, m-1\}$, $a_{i,i-1} = i$ for $i \in \{2, \dots, m\}$, $a_{1,m-1} = m$ and all other elements $a_{i,j} = 0$, i.e.,

$$A_W = \begin{pmatrix} -1 & 0 & 0 & \dots & 0 & m \\ 2 & -2 & 0 & \dots & 0 & 0 \\ 0 & 3 & -3 & \dots & 0 & 0 \\ 0 & 0 & 4 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & m-1 & -(m-1) \end{pmatrix}. \tag{5.2}$$

As is well-known, the matrix A_W has characteristic polynomial $\phi_m(\lambda)$, see e.g., [12, Example 7.8] or [16, Section 8.1.3].

Since the vector X_n determines the number of nodes with different numbers of keys, there is a linear map $T : \mathbb{R}^M \to \mathbb{R}^{m-1}$ such that $W_n = TX_n$. T is determined by the description of the types in Section 1.1.3, and it is easily seen that T is onto. Furthermore, starting the urns with an arbitrary (deterministic) non-zero vector $X_0 \in \mathbb{Z}^M_{\geq 0}$ and $W_0 = TX_0$, the urn dynamics yield

$$\mathbb{E}(X_1 - X_0) = \frac{AX_0}{a \cdot X_0} \tag{5.3}$$

$$\mathbb{E}(W_1 - W_0) = \frac{A_W W_0}{a_W \cdot W_0}. \tag{5.4}$$

Consequently, since also $a \cdot X_0 = a_W \cdot W_0$,

$$TAX_0 = (a \cdot X_0)T\,\mathbb{E}(X_1 - X_0) = (a_W \cdot W_0)\,\mathbb{E}(W_1 - W_0) = A_W W_0 = A_W T X_0,$$

and thus $TA = A_W T$.

Suppose that λ is a root of $\phi_m(\lambda) = 0$. Then λ is an eigenvalue of A_W and thus there exists a left eigenvector u' with $u'A_W = \lambda u'$. Consequently,

$$u'TA = u'A_W T = \lambda u'T, \tag{5.5}$$

so $u'T = (T'u)'$ is a left eigenvector of A. Since T is onto, T' is injective and thus $T'u \neq 0$. This shows that λ is an eigenvalue of A too. □

The largest eigenvalue $\lambda_1 = 1$ for the matrix A, since the total activity increases by 1 at each step, see [12, Lemma 5.4]. Let $\lambda_1, \lambda_2, \dots, \lambda_{m-1}$ be the roots of (5.1) in order of decreasing real parts. It is well-known that $\lambda_1 = 1$ and, moreover, that $\operatorname{Re}\lambda_2 \leq 1/2$ if and only if $m \leq 26$, see [17] and [9]. Consequently, if $m \geq 27$, then Lemma 5.1 shows that A has an eigenvalue $\lambda = \lambda_2 \neq \lambda_1$ with $\operatorname{Re}\lambda_2 > 1/2$, and then X_n is *not* asymptotically normal. (See [12] for general results suggesting this, and [4] for a rigorous proof in the present case, showing that the total number of internal nodes is *not* asymptotically normal.) Furthermore, if $\alpha := \operatorname{Re}\lambda_2 > 1/2$, then $(X_n - \mathbb{E}\,X_n)/n^\alpha$ is stochastically bounded, but has no limit in distribution (the distribution oscillates), see [4, 2, 12].

Some exceptional linear combinations of the variables $X_{n,i}$ are asymptotically normal also in such cases [12], but we conjecture that for any $m \geq 27$, the number of protected nodes is not one of these exceptional cases and that it has the same non-normal behaviour as just described for the number of internal nodes.

On the other hand, if $m \leq 26$, although A has a much larger dimension that A_W, and thus presumably many more eigenvalues, we conjecture that all additional eigenvalues also have $\operatorname{Re}\lambda < 1/2$, so that Theorem 1.4 applies showing that the number of protected vertices is asymptotically normal, with asymptotic variance linear in n, just as for $m = 2$ and 3 in Theorems 1.2 and 1.1. (This conjecture has been verified for $m \leq 6$ by Heimbürger [10].)

5.2 One-protected nodes and leaves in m-ary search trees.

As mentioned in Section 1, the number of one-protected nodes and the number of leaves (the complement of the one-protected nodes) are easier to analyze than the two-protected

nodes, and we prove normal limit laws for all m-ary search trees where $m \le 26$. In these cases we can use a Pólya urn that is similar to the Pólya urn that has earlier been used to study the total number of internal nodes in an m-ary search tree, see e.g. Mahmoud [15] and [16, Section 8.1.3] or [12, Example 7.8].

We can generalise the study of the number of leaves in ternary search tree in Section 4 to arbitrary $m \ge 2$. (For $m = 2$, there are minor modifications in the formulas below; we leave these to the reader. As mentioned above, the case $m = 2$ was considered by Devroye [5].) We have in general $m+1$ types, defined in analogy with Figure 9: Type 1 is as before, Type i with $2 \le i \le m-1$ is a leaf with $i-1$ keys, Type m is a leaf with $m-1$ keys together with its m external children, and Type $m+1$ is an internal non-leaf.

Let $V'_{i,n} = V_{i,n}$ be the number of nodes containing exactly i keys for $i \in \{1, \dots, m-2\}$; let $V'_{0,n}$ be the number of nodes containing 0 keys (external nodes) that are not children of leaves; let $V'_{m-1,n}$ be the number of nodes containing $m-1$ keys that are leaves (i.e., they have only external children); finally, let $V'_{m,n}$ be the number of internal nodes that are not leaves (all containing $m-1$ keys). We consider again another, slightly simpler, urn with the balls representing the gaps, giving them types $1, \dots, m$, and consider the vector $W'_n = (W'_{1,n}, \dots, W'_{m,n})$ where $W'_{i,n} = iV'_{i-1,n}$ is the total number of gaps of type i. The random vector W'_n can be described by a Pólya urn, with all activities 1. We denote the $m \times m$ matrix (1.1) for this urn by A_L. It is a minor modification of the matrix A_W described in Section 5.1, see (5.2); the entries of A_L are given by $a_{i,i} = -i$ for $i \in \{1, \dots, m\}$, $a_{i,i-1} = i$ for $i \in \{2, \dots, m\}$, $a_{1,m} = m-1$, $a_{2,m} = 2$, and all other entries $a_{i,j} = 0$. I.e.,

$$A_L = \begin{pmatrix} -1 & 0 & 0 & \dots & 0 & 0 & m-1 \\ 2 & -2 & 0 & \dots & 0 & 0 & 2 \\ 0 & 3 & -3 & \dots & 0 & 0 & 0 \\ 0 & 0 & 4 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & m-1 & -(m-1) & 0 \\ 0 & 0 & 0 & \dots & 0 & m & -m \end{pmatrix}. \tag{5.6}$$

We can easily calculate the characteristic polynomial of A_L and find that it is

$$\phi^L_m(\lambda) = (m+\lambda)\phi_m(\lambda), \tag{5.7}$$

where $\phi_m(\lambda)$ is the characteristic polynomial of A_W in (5.1). Thus, A_L has the same eigenvalues as A, plus the additional eigenvalue $\lambda = -m$. Since ϕ_m has only simple roots [14, Section 3.3], and $-m$ is not one of them, also ϕ^L_m has only simple roots. Hence, A_L has m distinct eigenvalues, and is thus diagonalisable.

The largest eigenvalue of A_L is $\lambda_1 = 1$ (as for A) and this eigenvalue corresponds to the right and left eigenvectors

$$v_1 = \frac{1}{H_m - 1} \begin{pmatrix} \frac{m-1}{2(m+1)} \\ \frac{1}{3} \\ \frac{1}{4} \\ \vdots \\ \frac{1}{m-1} \\ \frac{1}{m} \\ \frac{1}{m+1} \end{pmatrix}, \qquad u'_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ 1 \\ 1 \end{pmatrix}', \tag{5.8}$$

where we have normalized so that (1.2) holds (H_m denotes the mth harmonic number).

Let L_n be the number of leaves in an m-ary search tree with n keys. Then

$$L_n = \sum_{i=1}^{m-1} V'_{i,n} = \sum_{k=2}^{m} \frac{1}{k} W'_{k,n}. \tag{5.9}$$

Theorem 5.2. *Suppose that $3 \le m \le 26$. Let L_n be the number of leaves in an m-ary search tree. Then,*

$$\frac{L_n - \mu_L n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}\bigl(0, \sigma_L^2\bigr), \tag{5.10}$$

where

$$\mu_L = \frac{1}{H_m - 1} \cdot \sum_{k=2}^{m} \frac{1}{k(k+1)} = \frac{1}{H_m - 1} \cdot \frac{m-1}{2(m+1)}, \tag{5.11}$$

and σ_L^2 can be evaluated as

$$\sigma_L^2 = \sum_{i,j=2}^{m} \frac{\sigma_{ij}}{ij} \tag{5.12}$$

where $(\sigma_{ij})_{i,j=1}^m$ is given by (1.6).

Proof. As said above, for $m \le 26$, $\operatorname{Re}\lambda < \lambda_1/2 = 1/2$ for all eigenvalues $\lambda \neq \lambda_1$ of A, and thus also of A_L. Furthermore, A is diagonalisable. Hence, Theorem 1.5 applies and shows asymptotic normality of W'_n. The result follows by (5.9), using v_1 in (5.8). □

Remark 5.3. Theorem 5.2 implies that $\frac{E(L_n)}{n} \to \mu_L$, by the same argument as in Remark 1.3.

For $m \ge 27$, we expect the same non-normal asymptotic behaviour as for the number of internal nodes [4, 2], see Section 5.1.

For the one-protected nodes we can use the first Pólya urn described above for the leaves, with $m+1$ types. For the leaves we could simplify by considering the gaps and use a Pólya urn with m types, with all activities 1. However, now we also need to consider type $m+1$, which has 0 gaps. So in the analysis of the one-protected nodes we use the urn with $m+1$ different types (as explained in the beginning of this subsection) where types $i \in \{1, \dots, m\}$ have activities $1, 2, \dots, m$ and type $m+1$ has activity 0. In this Pólya urn, the one-protected nodes correspond to type $m+1$. All other types correspond to leaves or external nodes. Theorem 1.5 implies the following result (the proof is analogous to the proof of Theorem 5.2).

Theorem 5.4. *Suppose that $3 \le m \le 26$. Let Q_n be the number of one-protected nodes in an m-ary search tree. Then,*

$$\frac{Q_n - \mu_Q n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}\bigl(0, \sigma_Q^2\bigr), \tag{5.13}$$

where

$$\mu_Q = \frac{1}{H_m - 1} \cdot \frac{1}{(m+1)}, \tag{5.14}$$

and σ_Q^2 can be evaluated as

$$\sigma_Q^2 = \sigma_{m+1,m+1} \tag{5.15}$$

where $(\sigma_{ij})_{i,j=1}^{m+1}$ is given by (1.6).

This urn can obviously also be used to study the number of leaves (the types $2 \le i \le m$ correspond to the leaves), giving another proof of Theorem 5.2. (Note that σ_{ij} refers to different urns and thus has different meanings in Theorems 5.2 and 5.4.) Moreover, we can study L_n and Q_n together and obtain joint asymptotic normality for $m \le 26$; the covariance σ_{LQ} of the limit variables in (5.10) and (5.13) equals $\sum_{i=1}^{m} \sigma_{i,m+1}$ with $(\sigma_{ij})_{i,j=1}^{m+1}$ as in Theorem 5.4. In particular, this implies the well-known asymptotic normality of the total number of internal nodes $I_n = L_n + Q_n$, see e.g. [17, 14, 13, 4, 15, 9, 16].

Example 5.5. For a binary search tree ($m = 2$), a straightforward calculation of the covariance matrix $\Sigma = (\sigma_{ij})_{i,j=1}^{3}$ in Theorem 5.4 yields

$$\Sigma = \begin{pmatrix} \frac{8}{45} & -\frac{4}{45} & \frac{4}{45} \\ -\frac{4}{45} & \frac{2}{45} & -\frac{2}{45} \\ \frac{4}{45} & -\frac{2}{45} & \frac{2}{45} \end{pmatrix}. \tag{5.16}$$

Hence

$$\sigma_{L,2}^2 = (0,1,0)\,\Sigma\,(0,1,0)' = \sigma_{22} = \frac{2}{45}, \tag{5.17}$$

as shown by Devroye [5]. Similarly, $\sigma_{Q,2}^2 = \sigma_{33} = \frac{2}{45}$ and $\sigma_{LQ,2} = \sigma_{23} = -\frac{2}{45}$. (We have $\sigma_{L,2}^2 = \sigma_{Q,2}^2 = -\sigma_{LQ,2}$ since the total number of internal nodes $L_n + Q_n = I_n = n$ is deterministic when $m = 2$.)

Example 5.6. For a ternary search tree ($m = 3$), similarly (cf. (4.8) for the corresponding urn using the gaps as in Theorem 5.2)

$$\Sigma = \begin{pmatrix} \frac{479}{2100} & -\frac{7}{300} & -\frac{127}{2100} & \frac{101}{1400} \\ -\frac{7}{300} & \frac{8}{75} & -\frac{19}{300} & \frac{1}{100} \\ -\frac{127}{2100} & -\frac{19}{300} & \frac{131}{2100} & -\frac{43}{1400} \\ \frac{101}{1400} & \frac{1}{100} & -\frac{43}{1400} & \frac{9}{350} \end{pmatrix}. \tag{5.18}$$

Hence, cf. (4.4) and (4.9),

$$\sigma_{L,3}^2 = (0,1,1,0)\,\Sigma\,(0,1,1,0)' = \frac{89}{2100}, \tag{5.19}$$

$$\sigma_{Q,3}^2 = (0,0,0,1)\,\Sigma\,(0,0,0,1)' = \frac{9}{350}, \tag{5.20}$$

$$\sigma_{LQ,3} = (0,1,1,0)\,\Sigma\,(0,0,0,1)' = -\frac{29}{1400}. \tag{5.21}$$

We also obtain the corresponding asymptotic variance $(0,1,1,1)\,\Sigma\,(0,1,1,1)' = \sigma_{L,3}^2 + \sigma_{Q,3}^2 + 2\sigma_{LQ,3} = \frac{2}{75}$ for the number of internal nodes $L_n + Q_n$, as found by Mahmoud and Pittel [17].

Acknowledgements: We would like to thank Hosam M. Mahmoud and Mark D. Ward for valuable discussions. We would also like to thank Johan Björklund for help with drawing of figures.

References

[1] Bóna M., k-protected nodes in binary search trees. *Adv. Appl. Math.* **53** (2014), 1–11.

[2] Chauvin B. and Pouyanne N., m-ary search trees when $m \geq 27$: a strong asymptotics for the space requirement. *Random Structures Algorithms* **24** (2004), 133–154.

[3] Cheon G.S. and Shapiro L., Protected points in ordered trees. *Appl. Math. Lett.* **21** (2008), no. 5, 516–520.

[4] Chern H.-H. and Hwang H.-K., Phase changes in random m-ary search trees and generalized quicksort. *Random Structures Algorithms* **19** (2001), no. 3-4, 316–358.

[5] Devroye L., Limit laws for local counters in random binary search trees. *Random Structures Algorithms* **2** (1991), no. 3, 303–315.

[6] Devroye L. and Janson S., Protected nodes and fringe subtrees in some random trees. *Electronic Communications in Probability* **19** (2014), no. 6, 1–10.

[7] Drmota M., *Random Trees*, Springer, Vienna, 2009.

[8] Du R. and Prodinger H., Notes on protected nodes in digital search trees. *Appl. Math. Lett.* **25** (2012), no. 6, 1025–1028.

[9] Fill J.A. and Kapur N., Transfer theorems and asymptotic distributional results for m-ary search trees. *Random Structures Algorithms* **26** (2005), no. 4, 359–391.

[10] Heimbürger A., Asymptotic distribution of two-protected nodes in m-ary search trees. Master thesis, Stockholm University and KTH, 2014.

[11] Holmgren, C. and Janson S., Limit laws for functions of fringe trees for binary search trees and recursive trees. Preprint, 2014. `arXiv:1406.6883`

[12] Janson S., Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stoch. Process. Appl.* **110** (2004), 177–245.

[13] Lew W. and Mahmoud H.M., The joint distribution of elastic buckets in multiway search trees. *SIAM J. Comput.* **23** (1994), no. 5, 1050–1074.

[14] Mahmoud H.M., *Evolution of Random Search Trees*. John Wiley & Sons, New York, 1992.

[15] Mahmoud H.M., The size of random bucket trees via urn models. *Acta Inform.* **38** (2002), no. 11-12, 813–838.

[16] Mahmoud H.M., *Pólya Urn Models*. CRC Press, Boca Raton, FL, 2009.

[17] Mahmoud H.M. and Pittel B., Analysis of the space of search trees under the random insertion algorithm. *J. Algorithms* **10** (1989), no. 1, 52–75.

[18] Mahmoud H.M. and Ward M.D., Asymptotic distribution of two-protected nodes in random binary search trees. *Appl. Math. Lett.* **25** (2012), no. 12, 2218–2222.

[19] Mansour T., Protected points in k-ary trees. *Appl. Math. Lett.* **24** (2011), no. 4, 478–480.

A Appendix

$$P_I = \begin{pmatrix}
\frac{697}{700} & -\frac{2}{525} & -\frac{1}{300} & -\frac{1}{300} & -\frac{1}{350} & -\frac{1}{350} & -\frac{1}{350} & -\frac{1}{420} & -\frac{1}{420} & -\frac{1}{420} & -\frac{1}{525} & -\frac{1}{525} & -\frac{1}{525} & -\frac{1}{700} & -\frac{1}{700} & -\frac{1}{1050} & -\frac{1}{1050} & -\frac{1}{2100} & 0 \\
-\frac{3}{140} & \frac{103}{105} & -\frac{1}{60} & -\frac{1}{60} & -\frac{1}{70} & -\frac{1}{70} & -\frac{1}{70} & -\frac{1}{84} & -\frac{1}{84} & -\frac{1}{84} & -\frac{1}{105} & -\frac{1}{105} & -\frac{1}{105} & -\frac{1}{140} & -\frac{1}{140} & -\frac{1}{210} & -\frac{1}{210} & -\frac{1}{420} & 0 \\
-\frac{27}{700} & -\frac{6}{175} & \frac{97}{100} & -\frac{3}{100} & -\frac{9}{350} & -\frac{9}{350} & -\frac{9}{350} & -\frac{3}{140} & -\frac{3}{140} & -\frac{3}{140} & -\frac{3}{175} & -\frac{3}{175} & -\frac{3}{175} & -\frac{9}{700} & -\frac{9}{700} & -\frac{3}{350} & -\frac{3}{350} & -\frac{3}{700} & 0 \\
-\frac{27}{700} & -\frac{6}{175} & -\frac{3}{100} & \frac{97}{100} & -\frac{9}{350} & -\frac{9}{350} & -\frac{9}{350} & -\frac{3}{140} & -\frac{3}{140} & -\frac{3}{140} & -\frac{3}{175} & -\frac{3}{175} & -\frac{3}{175} & -\frac{9}{700} & -\frac{9}{700} & -\frac{3}{350} & -\frac{3}{350} & -\frac{3}{700} & 0 \\
-\frac{9}{350} & -\frac{4}{175} & -\frac{1}{50} & -\frac{1}{50} & \frac{172}{175} & -\frac{3}{175} & -\frac{3}{175} & -\frac{1}{70} & -\frac{1}{70} & -\frac{1}{70} & -\frac{2}{175} & -\frac{2}{175} & -\frac{2}{175} & -\frac{3}{350} & -\frac{3}{350} & -\frac{1}{175} & -\frac{1}{175} & -\frac{1}{350} & 0 \\
-\frac{3}{100} & -\frac{2}{75} & -\frac{7}{300} & -\frac{7}{300} & -\frac{1}{50} & \frac{49}{50} & -\frac{1}{50} & -\frac{1}{60} & -\frac{1}{60} & -\frac{1}{60} & -\frac{1}{75} & -\frac{1}{75} & -\frac{1}{75} & -\frac{1}{100} & -\frac{1}{100} & -\frac{1}{150} & -\frac{1}{150} & -\frac{1}{300} & 0 \\
-\frac{27}{175} & -\frac{24}{175} & -\frac{3}{25} & -\frac{3}{25} & -\frac{18}{175} & -\frac{18}{175} & \frac{157}{175} & -\frac{3}{35} & -\frac{3}{35} & -\frac{3}{35} & -\frac{12}{175} & -\frac{12}{175} & -\frac{12}{175} & -\frac{9}{175} & -\frac{9}{175} & -\frac{6}{175} & -\frac{6}{175} & -\frac{3}{175} & 0 \\
-\frac{3}{35} & -\frac{8}{105} & -\frac{1}{15} & -\frac{1}{15} & -\frac{2}{35} & -\frac{2}{35} & -\frac{2}{35} & \frac{20}{21} & -\frac{1}{21} & -\frac{1}{21} & -\frac{4}{105} & -\frac{4}{105} & -\frac{4}{105} & -\frac{1}{35} & -\frac{1}{35} & -\frac{2}{105} & -\frac{2}{105} & -\frac{1}{105} & 0 \\
-\frac{9}{50} & -\frac{4}{25} & -\frac{7}{50} & -\frac{7}{50} & -\frac{3}{25} & -\frac{3}{25} & -\frac{3}{25} & -\frac{1}{10} & \frac{9}{10} & -\frac{1}{10} & -\frac{2}{25} & -\frac{2}{25} & -\frac{2}{25} & -\frac{3}{50} & -\frac{3}{50} & -\frac{1}{25} & -\frac{1}{25} & -\frac{1}{50} & 0 \\
-\frac{9}{50} & -\frac{4}{25} & -\frac{7}{50} & -\frac{7}{50} & -\frac{3}{25} & -\frac{3}{25} & -\frac{3}{25} & -\frac{1}{10} & -\frac{1}{10} & \frac{9}{10} & -\frac{2}{25} & -\frac{2}{25} & -\frac{2}{25} & -\frac{3}{50} & -\frac{3}{50} & -\frac{1}{25} & -\frac{1}{25} & -\frac{1}{50} & 0 \\
-\frac{9}{140} & -\frac{2}{35} & -\frac{1}{20} & -\frac{1}{20} & -\frac{3}{70} & -\frac{3}{70} & -\frac{3}{70} & -\frac{1}{28} & -\frac{1}{28} & -\frac{1}{28} & \frac{34}{35} & -\frac{1}{35} & -\frac{1}{35} & -\frac{3}{140} & -\frac{3}{140} & -\frac{1}{70} & -\frac{1}{70} & -\frac{1}{140} & 0 \\
-\frac{9}{70} & -\frac{4}{35} & -\frac{1}{10} & -\frac{1}{10} & -\frac{3}{35} & -\frac{3}{35} & -\frac{3}{35} & -\frac{1}{14} & -\frac{1}{14} & -\frac{1}{14} & -\frac{2}{35} & \frac{33}{35} & -\frac{2}{35} & -\frac{3}{70} & -\frac{3}{70} & -\frac{1}{35} & -\frac{1}{35} & -\frac{1}{70} & 0 \\
-\frac{27}{50} & -\frac{12}{25} & -\frac{21}{50} & -\frac{21}{50} & -\frac{9}{25} & -\frac{9}{25} & -\frac{9}{25} & -\frac{3}{10} & -\frac{3}{10} & -\frac{3}{10} & -\frac{6}{25} & -\frac{6}{25} & \frac{19}{25} & -\frac{9}{50} & -\frac{9}{50} & -\frac{3}{25} & -\frac{3}{25} & -\frac{3}{50} & 0 \\
-\frac{3}{25} & -\frac{8}{75} & -\frac{7}{75} & -\frac{7}{75} & -\frac{2}{25} & -\frac{2}{25} & -\frac{2}{25} & -\frac{1}{15} & -\frac{1}{15} & -\frac{1}{15} & -\frac{4}{75} & -\frac{4}{75} & -\frac{4}{75} & \frac{24}{25} & -\frac{1}{25} & -\frac{2}{75} & -\frac{2}{75} & -\frac{1}{75} & 0 \\
-\frac{36}{175} & -\frac{32}{175} & -\frac{4}{25} & -\frac{4}{25} & -\frac{24}{175} & -\frac{24}{175} & -\frac{24}{175} & -\frac{4}{35} & -\frac{4}{35} & -\frac{4}{35} & -\frac{16}{175} & -\frac{16}{175} & -\frac{16}{175} & -\frac{12}{175} & \frac{163}{175} & -\frac{8}{175} & -\frac{8}{175} & -\frac{4}{175} & 0 \\
-\frac{3}{20} & -\frac{2}{15} & -\frac{7}{60} & -\frac{7}{60} & -\frac{1}{10} & -\frac{1}{10} & -\frac{1}{10} & -\frac{1}{12} & -\frac{1}{12} & -\frac{1}{12} & -\frac{1}{15} & -\frac{1}{15} & -\frac{1}{15} & -\frac{1}{20} & -\frac{1}{20} & \frac{29}{30} & -\frac{1}{30} & -\frac{1}{60} & 0 \\
-\frac{9}{50} & -\frac{4}{25} & -\frac{7}{50} & -\frac{7}{50} & -\frac{3}{25} & -\frac{3}{25} & -\frac{3}{25} & -\frac{1}{10} & -\frac{1}{10} & -\frac{1}{10} & -\frac{2}{25} & -\frac{2}{25} & -\frac{2}{25} & -\frac{3}{50} & -\frac{3}{50} & -\frac{1}{25} & \frac{24}{25} & -\frac{1}{50} & 0 \\
-\frac{27}{140} & -\frac{6}{35} & -\frac{3}{20} & -\frac{3}{20} & -\frac{9}{70} & -\frac{9}{70} & -\frac{9}{70} & -\frac{3}{28} & -\frac{3}{28} & -\frac{3}{28} & -\frac{3}{35} & -\frac{3}{35} & -\frac{3}{35} & -\frac{9}{140} & -\frac{9}{140} & -\frac{3}{70} & -\frac{3}{70} & \frac{137}{140} & 0 \\
-\frac{9}{25} & -\frac{8}{25} & -\frac{7}{25} & -\frac{7}{25} & -\frac{6}{25} & -\frac{6}{25} & -\frac{6}{25} & -\frac{1}{5} & -\frac{1}{5} & -\frac{1}{5} & -\frac{4}{25} & -\frac{4}{25} & -\frac{4}{25} & -\frac{3}{25} & -\frac{3}{25} & -\frac{2}{25} & -\frac{2}{25} & -\frac{1}{25} & 1
\end{pmatrix} \tag{A.1}$$

$$B = \begin{pmatrix}
\frac{19}{2100} & -\frac{1}{210} & 0 & 0 & 0 & -\frac{3}{700} & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{3}{700} & 0 & 0 & 0 & 0 & 0 & 0 \\
-\frac{1}{210} & \frac{17}{420} & -\frac{3}{175} & -\frac{3}{700} & 0 & 0 & 0 & -\frac{1}{70} & 0 & 0 & 0 & 0 & -\frac{1}{70} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -\frac{3}{175} & \frac{9}{140} & 0 & -\frac{3}{175} & 0 & -\frac{3}{175} & 0 & 0 & 0 & -\frac{9}{700} & 0 & -\frac{9}{700} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -\frac{3}{700} & 0 & \frac{9}{140} & 0 & 0 & -\frac{6}{175} & 0 & 0 & 0 & 0 & -\frac{9}{350} & -\frac{9}{350} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -\frac{3}{175} & 0 & \frac{13}{350} & 0 & 0 & 0 & -\frac{1}{50} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-\frac{3}{700} & 0 & 0 & 0 & 0 & \frac{13}{300} & 0 & -\frac{2}{105} & 0 & 0 & 0 & 0 & -\frac{11}{700} & -\frac{1}{50} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -\frac{3}{175} & -\frac{6}{175} & 0 & 0 & \frac{39}{175} & 0 & -\frac{2}{25} & -\frac{1}{25} & 0 & 0 & -\frac{9}{175} & 0 & -\frac{9}{175} & 0 & 0 & 0 & 0 \\
0 & -\frac{1}{70} & 0 & 0 & 0 & -\frac{2}{105} & 0 & \frac{11}{105} & 0 & 0 & -\frac{1}{35} & -\frac{1}{70} & -\frac{1}{70} & 0 & 0 & -\frac{1}{35} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -\frac{1}{50} & 0 & -\frac{2}{25} & 0 & \frac{11}{50} & 0 & 0 & 0 & -\frac{3}{25} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{25} & 0 & 0 & \frac{11}{50} & 0 & 0 & -\frac{9}{50} & 0 & 0 & 0 & -\frac{3}{50} & 0 & 0 \\
0 & 0 & -\frac{9}{700} & 0 & 0 & 0 & 0 & -\frac{1}{35} & 0 & 0 & \frac{9}{140} & 0 & \frac{9}{700} & 0 & -\frac{4}{175} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -\frac{9}{350} & 0 & 0 & 0 & -\frac{1}{70} & 0 & 0 & 0 & \frac{9}{70} & -\frac{3}{175} & 0 & -\frac{8}{175} & 0 & 0 & -\frac{3}{70} & 0 \\
-\frac{3}{700} & -\frac{1}{70} & -\frac{9}{700} & -\frac{9}{350} & 0 & -\frac{11}{700} & -\frac{9}{175} & -\frac{1}{70} & -\frac{3}{25} & -\frac{9}{50} & \frac{9}{700} & -\frac{3}{175} & \frac{27}{50} & -\frac{1}{50} & \frac{9}{175} & \frac{1}{35} & \frac{3}{50} & \frac{3}{70} & \frac{1}{25} \\
0 & 0 & 0 & 0 & 0 & -\frac{1}{50} & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{50} & \frac{7}{75} & 0 & -\frac{1}{30} & 0 & 0 & -\frac{1}{25} \\
0 & 0 & 0 & 0 & 0 & 0 & -\frac{9}{175} & 0 & 0 & 0 & -\frac{4}{175} & -\frac{8}{175} & \frac{9}{175} & 0 & \frac{4}{25} & 0 & -\frac{1}{25} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{35} & 0 & 0 & 0 & 0 & \frac{1}{35} & -\frac{1}{30} & 0 & \frac{1}{12} & 0 & -\frac{3}{140} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{3}{50} & 0 & 0 & \frac{3}{50} & 0 & -\frac{1}{25} & 0 & \frac{1}{10} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{3}{70} & \frac{3}{70} & 0 & 0 & -\frac{3}{140} & 0 & \frac{9}{140} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{25} & -\frac{1}{25} & 0 & 0 & 0 & 0 & \frac{1}{25}
\end{pmatrix} \tag{A.2}$$

$$\Sigma = \frac{1}{2100} \cdot \begin{pmatrix}
\frac{41376542411}{41611670100} & -\frac{23408479}{876035160} & -\frac{24963443}{547521975} & -\frac{24963443}{547521975} & -\frac{338777}{11711700} & \frac{288002937}{9247037800} & -\frac{338777}{1951950} & -\frac{61814939}{730029300} & -\frac{2497361}{12882870} & -\frac{2497361}{12882870} & -\frac{6678998}{109504395} & -\frac{13357996}{109504395} & -\frac{72324583}{128828700} & -\frac{720772978}{10402917525} & -\frac{291188}{1533675} & -\frac{343662071}{4380175800} & -\frac{7052429}{42942900} & -\frac{10190867}{109504395} & -\frac{68043227}{3963016200} \\
-\frac{23408479}{876035160} & \frac{106759904}{21900879} & -\frac{10916617}{51531480} & -\frac{10916617}{51531480} & -\frac{244879}{1840410} & -\frac{1337363}{9733724} & -\frac{244879}{306735} & -\frac{53606911}{146005860} & -\frac{3248263}{3680820} & -\frac{3248263}{3680820} & -\frac{2680231}{10306296} & -\frac{2680231}{5153148} & -\frac{465730}{184041} & -\frac{59525563}{219008790} & -\frac{3427337}{4294290} & -\frac{126531917}{438017580} & -\frac{1469917}{2147145} & -\frac{3279013}{10306296} & \frac{3706671}{97337240} \\
-\frac{24963443}{547521975} & -\frac{10916617}{51531480} & \frac{12239749}{1415700} & -\frac{501551}{1415700} & -\frac{2178359}{9909900} & \frac{312293407}{1460058600} & -\frac{2178359}{1651650} & -\frac{3411301}{6134700} & -\frac{20368}{14157} & -\frac{20368}{14157} & -\frac{757667}{1981980} & -\frac{757667}{990990} & -\frac{443321}{108900} & -\frac{196315124}{547521975} & -\frac{941257}{825825} & -\frac{86297573}{257657400} & -\frac{3159973}{3303300} & -\frac{85739}{283140} & \frac{56654959}{208579800} \\
-\frac{24963443}{547521975} & -\frac{10916617}{51531480} & -\frac{501551}{1415700} & \frac{12239749}{1415700} & -\frac{2178359}{9909900} & \frac{312293407}{1460058600} & -\frac{2178359}{1651650} & -\frac{3411301}{6134700} & -\frac{20368}{14157} & -\frac{20368}{14157} & -\frac{757667}{1981980} & -\frac{757667}{990990} & -\frac{443321}{108900} & -\frac{196315124}{547521975} & -\frac{941257}{825825} & -\frac{86297573}{257657400} & -\frac{3159973}{3303300} & -\frac{85739}{283140} & \frac{56654959}{208579800} \\
-\frac{338777}{11711700} & -\frac{244879}{1840410} & -\frac{2178359}{9909900} & -\frac{2178359}{9909900} & \frac{1162547}{198198} & -\frac{1290076}{10735725} & -\frac{26641}{33033} & -\frac{3201412}{10735725} & -\frac{94213}{108900} & -\frac{94213}{108900} & -\frac{76651}{396396} & -\frac{76651}{198198} & -\frac{65447}{27225} & -\frac{456479}{2927925} & -\frac{889403}{1651650} & -\frac{3296674}{32207175} & -\frac{26839}{63525} & -\frac{39443}{1981980} & \frac{6456887}{21471450} \\
-\frac{288002937}{9247037800} & -\frac{1337363}{9733724} & \frac{312293407}{1460058600} & \frac{312293407}{1460058600} & \frac{1290076}{10735725} & \frac{1341943367}{198150810} & -\frac{2580152}{3578575} & \frac{433585133}{730029300} & \frac{29434157}{42942900} & \frac{29434157}{42942900} & \frac{117464953}{292011720} & \frac{117464953}{146005860} & -\frac{34515419}{21471450} & -\frac{366941303}{495377025} & -\frac{8194161}{7157150} & -\frac{676938}{790075} & -\frac{1247859}{1431430} & -\frac{59455463}{58402344} & \frac{3567263197}{3963016200} \\
-\frac{338777}{1951950} & -\frac{244879}{306735} & -\frac{2178359}{1651650} & -\frac{2178359}{1651650} & -\frac{26641}{33033} & -\frac{2580152}{3578575} & \frac{343114}{11011} & -\frac{6402824}{3578575} & -\frac{94213}{18150} & -\frac{94213}{18150} & -\frac{76651}{66066} & -\frac{76651}{33033} & -\frac{130894}{9075} & -\frac{912958}{975975} & -\frac{889403}{275275} & -\frac{6593348}{10735725} & -\frac{53678}{21175} & -\frac{39443}{330330} & \frac{6456887}{3578575} \\
-\frac{61814939}{730029300} & -\frac{53606911}{146005860} & -\frac{3411301}{6134700} & -\frac{3411301}{6134700} & \frac{3201412}{10735725} & -\frac{433585133}{730029300} & -\frac{6402824}{3578575} & \frac{96310504}{5214495} & \frac{33553843}{21471450} & -\frac{33553843}{21471450} & -\frac{8757649}{8588580} & -\frac{8757649}{4294290} & -\frac{33159457}{10735725} & -\frac{109833091}{60835775} & -\frac{10151767}{3578575} & -\frac{298480453}{146005860} & -\frac{7463983}{3578575} & -\frac{2909761}{1226940} & -\frac{95014589}{52144950} \\
-\frac{2497361}{12882870} & -\frac{3248263}{3680820} & -\frac{20368}{14157} & -\frac{20368}{14157} & -\frac{94213}{108900} & -\frac{29434157}{42942900} & -\frac{94213}{18150} & -\frac{33553843}{21471450} & \frac{90452}{2475} & -\frac{13498}{2475} & -\frac{439141}{495495} & -\frac{878282}{495495} & -\frac{13309}{900} & -\frac{712856}{1288287} & -\frac{25198}{12705} & \frac{12017233}{128828700} & -\frac{25111}{23100} & \frac{559703}{495495} & \frac{12078779}{4294290} \\
-\frac{2497361}{12882870} & -\frac{3248263}{3680820} & -\frac{20368}{14157} & -\frac{20368}{14157} & -\frac{94213}{108900} & -\frac{29434157}{42942900} & -\frac{94213}{18150} & -\frac{33553843}{21471450} & -\frac{13498}{2475} & \frac{90452}{2475} & -\frac{439141}{495495} & -\frac{878282}{495495} & -\frac{13309}{900} & -\frac{712856}{1288287} & -\frac{25198}{12705} & \frac{12017233}{128828700} & -\frac{25111}{23100} & \frac{559703}{495495} & \frac{12078779}{4294290} \\
-\frac{6678998}{109504395} & -\frac{2680231}{10306296} & -\frac{757667}{1981980} & -\frac{757667}{1981980} & -\frac{76651}{396396} & -\frac{117464953}{292011720} & -\frac{76651}{66066} & -\frac{8757649}{8588580} & -\frac{439141}{495495} & -\frac{439141}{495495} & \frac{5682427}{396396} & -\frac{263513}{198198} & -\frac{172523}{152460} & -\frac{122261819}{109504395} & -\frac{295822}{165165} & -\frac{519971}{425880} & -\frac{824749}{660660} & -\frac{538127}{396396} & -\frac{240667397}{292011720} \\
-\frac{13357996}{109504395} & -\frac{2680231}{5153148} & -\frac{757667}{990990} & -\frac{757667}{990990} & -\frac{76651}{198198} & -\frac{117464953}{146005860} & -\frac{76651}{33033} & -\frac{8757649}{4294290} & -\frac{878282}{495495} & -\frac{878282}{495495} & -\frac{263513}{198198} & \frac{2709457}{99099} & -\frac{172523}{76230} & -\frac{244523638}{109504395} & -\frac{591644}{165165} & -\frac{519971}{212940} & -\frac{824749}{330330} & -\frac{538127}{198198} & -\frac{240667397}{146005860} \\
-\frac{72324583}{128828700} & -\frac{465730}{184041} & -\frac{443321}{108900} & -\frac{443321}{108900} & -\frac{65447}{27225} & -\frac{34515419}{21471450} & -\frac{130894}{9075} & -\frac{33159457}{10735725} & -\frac{13309}{900} & -\frac{13309}{900} & -\frac{172523}{152460} & -\frac{172523}{76230} & \frac{39227}{450} & -\frac{7955746}{32207175} & \frac{54421}{127050} & \frac{161191087}{64414350} & \frac{893}{210} & \frac{215345}{30492} & \frac{217039103}{21471450} \\
-\frac{720772978}{10402917525} & -\frac{59525563}{219008790} & -\frac{196315124}{547521975} & -\frac{196315124}{547521975} & -\frac{456479}{2927925} & -\frac{366941303}{495377025} & -\frac{912958}{975975} & -\frac{109833091}{60835775} & -\frac{712856}{1288287} & -\frac{712856}{1288287} & -\frac{122261819}{109504395} & -\frac{244523638}{109504395} & -\frac{7955746}{32207175} & \frac{36088849118}{1486131075} & -\frac{30069938}{10735725} & -\frac{655132571}{156434850} & -\frac{19005188}{10735725} & -\frac{522952526}{109504395} & -\frac{4662351652}{495377025} \\
-\frac{291188}{1533675} & -\frac{3427337}{4294290} & -\frac{941257}{825825} & -\frac{941257}{825825} & -\frac{889403}{1651650} & -\frac{8194161}{7157150} & -\frac{889403}{275275} & -\frac{10151767}{3578575} & -\frac{25198}{12705} & -\frac{25198}{12705} & -\frac{295822}{165165} & -\frac{591644}{165165} & \frac{54421}{127050} & -\frac{30069938}{10735725} & \frac{11949972}{275275} & -\frac{61733263}{21471450} & -\frac{123237}{42350} & -\frac{483403}{165165} & -\frac{3529243}{3578575} \\
-\frac{343662071}{4380175800} & \frac{126531917}{438017580} & -\frac{86297573}{257657400} & -\frac{86297573}{257657400} & \frac{3296674}{32207175} & -\frac{676938}{790075} & \frac{6593348}{10735725} & \frac{298480453}{146005860} & \frac{12017233}{128828700} & \frac{12017233}{128828700} & -\frac{519971}{425880} & -\frac{519971}{212940} & \frac{161191087}{64414350} & -\frac{655132571}{156434850} & \frac{61733263}{21471450} & \frac{949120823}{31286970} & \frac{34044067}{21471450} & \frac{269791793}{51531480} & -\frac{694478313}{69526600} \\
-\frac{7052429}{42942900} & -\frac{1469917}{2147145} & -\frac{3159973}{3303300} & -\frac{3159973}{3303300} & -\frac{26839}{63525} & -\frac{1247859}{1431430} & -\frac{53678}{21175} & -\frac{7463983}{3578575} & -\frac{25111}{23100} & -\frac{25111}{23100} & -\frac{824749}{660660} & -\frac{824749}{330330} & \frac{893}{210} & -\frac{19005188}{10735725} & -\frac{123237}{42350} & -\frac{34044067}{21471450} & \frac{156031}{3850} & -\frac{826069}{660660} & \frac{3453169}{7157150} \\
-\frac{10190867}{109504395} & -\frac{3279013}{10306296} & -\frac{85739}{283140} & -\frac{85739}{283140} & -\frac{39443}{1981980} & -\frac{59455463}{58402344} & -\frac{39443}{330330} & -\frac{2909761}{1226940} & \frac{559703}{495495} & \frac{559703}{495495} & -\frac{538127}{396396} & -\frac{538127}{198198} & \frac{215345}{30492} & -\frac{522952526}{109504395} & -\frac{483403}{165165} & -\frac{269791793}{51531480} & -\frac{826069}{660660} & \frac{2222557}{56628} & -\frac{439517549}{41715960} \\
-\frac{68043227}{3963016200} & \frac{3706671}{97337240} & \frac{56654959}{208579800} & \frac{56654959}{208579800} & \frac{6456887}{21471450} & \frac{3567263197}{3963016200} & \frac{6456887}{3578575} & -\frac{95014589}{52144950} & \frac{12078779}{4294290} & \frac{12078779}{4294290} & -\frac{240667397}{292011720} & -\frac{240667397}{146005860} & \frac{217039103}{21471450} & -\frac{4662351652}{495377025} & -\frac{3529243}{3578575} & -\frac{694478313}{69526600} & \frac{3453169}{7157150} & \frac{439517549}{41715960} & \frac{3991031128}{165125675}
\end{pmatrix}$$