AN ANALYTIC APPROACH FOR THE ANALYSIS OF ROTATIONS IN FRINGE-BALANCED BINARY SEARCH TREES

ALOIS PANHOLZER AND HELMUT PRODINGER

ABSTRACT. This paper presents an analytic approach to the construction cost of fringe-balanced binary search trees. In [7], Mahmoud used a bottom-up approach and an urn model of Pólya. The present method is top-down and uses differential equations and Hwang's quasi-power theorem to derive the asymptotic normality of the number of rotations needed to constract such a *fringe balanced search tree.* We also obtain the exact expectation and variance with this method. Although Pólya's urn model is not needed anymore, we also present an elegant analysis of it, based of an operator calculus as in Greene and Knuth [4].

1. INTRODUCTION

One way to improve the speed of retrieval of data stored in the nodes of a binary search tree is a method called *fringe balancing*, which was introduced by Walker and Wood [12] and first analysed by Poblete and Munro [9]. The basic idea is the following: whenever a son is appended to a node that itself is a single son (its "brother" is an external node), a *rotation* of the three nodes is performed to place the median of the three elements as the root of the subtree and the other two elements as sons (see Figure 1). Therefore all subtrees of size 3 in the tree are complete. The distribution of the number of rotations that are made when constructing such a fringe-balanced binary tree (or by other authors called *locally balanced binary tree*) under the random permutation model was recently analysed by Mahmoud [7]. To establish a central limit theorem, he employed a Pólya urn model and used a theorem of Smythe [11]. (Strictly speaking, according to Mahmoud, *"Smythe's proof will go through without any modifications,"* but to work out the details seems to be cumbersome.)

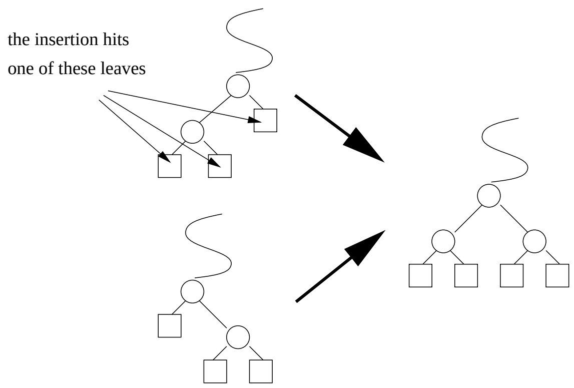


FIGURE 1. The fringe heuristic

The intention of this paper is to show an alternative way of proving the central limit theorem for the number of rotations, which is based on analytic methods using generating functions, singularity analysis [2] and Hwang's *quasi-power theorem* [5, 3].

Important for our analysis is the fact, that a recursive top-down randomness model for fringe balanced trees is also possible, as it was shown in [8]. In this sense the only difference to "ordinary" random search trees starting from a random permutation are the so called *splitting probabilities* $\pi_{n,k}$, which are the probabilities that in a tree of size n the root is the node k. Whereas in random search trees these splitting probabilities are $1/n$ independent from k, we have in fringe balanced search trees

$$\pi_{n,k} = \frac{(k-1)(n-k)}{\binom{n}{3}}$$

for $n \geq 3$ and $1 \leq k \leq n$, $\pi_{2,1} = \pi_{2,2} = \frac{1}{2}$ and $\pi_{1,1} = 1$.

2. Mathematical analysis

Considering the top-down randomness model we can immediately establish a recursion for the probabilities $F_{n,m}$, that the number of rotations is m, when generating a fringe balanced tree of size n ($= n$ data in the tree) starting with an empty tree. We use only the recursive definition of the trees and the fact, that with probability $1/3$ we need no and with probability $2/3$ exactly one rotation to construct the root of the tree for $n \geq 3$. We get then for $n \geq 3$ and $1 \leq m \leq n$

$$F_{n,m} = \frac{1}{3} \sum_{k=1}^{n} \pi_{n,k} \sum_{l=0}^{m} F_{k-1,l} F_{n-k,m-l} + \frac{2}{3} \sum_{k=1}^{n} \pi_{n,k} \sum_{l=0}^{m-1} F_{k-1,l} F_{n-k,m-1-l} \tag{1}$$

with initial values $F_{0,0} = 1$, $F_{1,0} = 1$, $F_{2,0} = 1$ and $F_{n,m} = 0$ otherwise.

Introducing the probability generating function $F(z,v) = \sum_{n \geq 0} \sum_{m \geq 0} F_{n,m} z^n v^m$ this recursion leads by multiplying with $z^n v^m$ and summing up to the following differential equation

$$\frac{1}{6} \frac{\partial^3}{\partial z^3} F(z,v) = \left(\frac{1}{3} + \frac{2}{3} v\right) \left(\frac{\partial}{\partial z} F(z,v)\right)^2$$

with initial conditions $F(0,v) = 1$, $\frac{\partial}{\partial z} F(z,v)\big|_{z=0} = 1$ and $\frac{\partial^2}{\partial z^2} F(z,v)\Big|_{z=0} = 2$.

Substituting $G(z,v) = \frac{\partial}{\partial z} F(z,v)$ this differential equation can be written as

$$\frac{\partial^2}{\partial z^2} G(z,v) = (2+4v) G(z,v)^2 \tag{2}$$

with $G(0,v) = 1$ and $\frac{\partial}{\partial z} G(z,v)\big|_{z=0} = 2$.

From this differential equation it is easy to establish the exact expectation and variance of the number of rotations generating a fringe balanced tree.

2.1. **The expectation.**

Differentiating (2) w.r.t. v and evaluating at $v = 1$ leads to the following second order differential equation (*Eulerian differential equation*) for $M_1(z) = \frac{\partial}{\partial v} G(z,v)\big|_{v=1} = \sum_{n \geq 0} n M_n^{(1)} z^{n-1}$, where $M_n^{(1)}$ denotes the expectation of the number of rotations generating a fringe balanced tree of size n:

$$M_1''(z) = \frac{12}{(1-z)^2} M_1(z) + \frac{4}{(1-z)^4}$$

with $M_1(0) = 0$ and $M_1'(0) = 0$.

Extracting the coefficients of the solution

$$M_1(z) = \frac{4}{7}\frac{1}{(1-z)^3} - \frac{2}{3}\frac{1}{(1-z)^2} + \frac{2}{21}(1-z)^4$$

leads to the following theorem.

Theorem 2.1. *The expectation $M_n^{(1)}$ of the number of rotations when generating a fringe balanced binary search tree of size n is for $n \geq 6$ given by*

$$M_n^{(1)} = \frac{2}{7}n - \frac{8}{21}.$$

2.2. **The variance.** In an analogous way we get by differentiating equation (2) twice w.r.t. v and evaluating at $v = 1$ the following differential equation for $M_2(z) = \frac{\partial^2}{\partial v^2} G(z,v)\Big|_{v=1} = \sum_{n\geq 0} n M_n^{(2)} z^{n-1}$, where $M_n^{(2)}$ denotes the second factorial moment of the number of rotations generating a fringe balanced tree:

$$M_2''(z) = \frac{12}{(1-z)^2} M_2(z) + 12 M_1^2(z) + \frac{16}{(1-z)^2} M_1(z).$$

Solving this equation leads to

$$\begin{aligned} M_2(z) = {} & \frac{24}{49}\frac{1}{(1-z)^4} - \frac{342}{273}\frac{1}{(1-z)^3} + \frac{8}{9}\frac{1}{(1-z)^2} \\ & - \frac{8}{5733}(1-z)^3(z^7 - 7z^6 + 21z^5 - 35z^4 + 35z^4 - 21z^2 + 98z + 64), \end{aligned}$$

and we get by extracting the coefficients the following theorem for the variance $V_n = M_n^{(2)} + M_n^{(1)} - \left(M_n^{(1)}\right)^2$:

Theorem 2.2. *The variance V_n of the number of rotations when generating a fringe balanced binary search tree of size n is for $n \geq 12$ given by*

$$V_n = \frac{66}{637}n - \frac{680}{5733}.$$

2.3. **The limiting distribution.** To solve the differential equation (2) we substitute $q(G,z) = \frac{\partial}{\partial z} G(z,v)$ and get

$$(2+4v)G(z,v)^2 = \frac{dq}{dz} = \frac{dq}{dG}\frac{dG}{dz} = \frac{dq}{dG} q(G,z)$$

or

$$(2+4v)G(z,v)^2 dG = q(G,z)dq.$$

Integration and adjusting the initial values leads to

$$\frac{4}{3}(1+2v)G^3(z,v) + \frac{8}{3}(1-v) = \left(\frac{\partial}{\partial z} G(z,v)\right)^2,$$

from which we get the implicitly given solution of $G(z,v)$

$$z = \int_1^{G(z,v)} \frac{dx}{\sqrt{\frac{4}{3}(1+2v)x^3 + \frac{8}{3}(1-v)}}. \tag{3}$$

In the following we use the fact, that the function $G(z,v)$ is closely related to the Weierstrass' $\wp$-function, which plays a fundamental rôle in the theory of elliptic functions [14]. The Weierstrass' $\wp$-function can be characterised within a simply connected domain of $\mathbb{C}_\infty$, which contains no zeros of the denominator, by the integral

$$z = \int_{\wp(z)}^{\infty} \frac{dx}{\sqrt{4x^3 - g_2 x - g_3}}.$$

The constants g_2 resp. g_3 are called the *invariants* of $\wp$.

The function $\wp(z)$ has around $z=0$ a Laurent expansion

$$\wp(z) = \frac{1}{z^2} + a_2 z^2 + a_4 z^4 + \ldots + a_{2k} z^{2k} + \ldots .$$

There are the following relations for the invariants $g_2 = 20a_2 = 60\sum_{\omega\in\Omega,\omega\neq 0}\frac{1}{\omega^4}$ and $g_3 = 28a_4 = 140\sum_{\omega\in\Omega,\omega\neq 0}\frac{1}{\omega^6}$. The elements $\omega = k_1\omega_1 + k_2\omega_2$ with $k_1, k_2 \in \mathbb{Z}$ and $\omega_1, \omega_2 \in \mathbb{C}$ are the elements of the period-group $\Omega(\omega_1, \omega_2)$.

Simple manipulations of equation (3) lead to

$$\sqrt{\frac{1+2v}{3}}\left(\int_1^\infty \frac{dx}{\sqrt{\frac{4}{3}(1+2v)x^3 + \frac{8}{3}(1-v)}} - z\right) = \int_{G(z,v)}^\infty \frac{dx}{\sqrt{4x^3 + \frac{8(1-v)}{1+2v}}},$$

from which we obtain with the substitutions $t = \sqrt{\frac{1+2v}{3}}\Big(\int\limits_1^\infty \frac{dx}{\sqrt{\frac{4}{3}(1+2v)x^3+\frac{8}{3}(1-v)}} - z\Big)$ and $W(t,v) = G(z,v)$ the equation

$$t = \int_{W(t,v)}^\infty \frac{dx}{\sqrt{4x^3 + \frac{8(1-v)}{1+2v}}},$$

so $W(t,v) = G(z(t),v)$ is a Weierstrass' $\wp$-function with invariants $g_2 = 0$ and $g_3 = -\frac{8(1-v)}{1+2v}$.

This leads to the following Laurent-expansion of $W(t,v)$ around $t=0$

$$W(t,v) = \frac{1}{t^2} - \frac{1}{28}\frac{8(1-v)}{1+2v}t^4 + \ldots$$

resp. to the following expansion of $G(z,v)$ around its singularity
$z = s(v) := \int\limits_1^\infty \frac{dx}{\sqrt{\frac{4}{3}(1+2v)x^3+\frac{8}{3}(1-v)}}$:

$$G(z,v) = \frac{1}{\left(\sqrt{\frac{1+2v}{3}}\left(\int\limits_1^\infty \frac{dx}{\sqrt{\frac{4}{3}(1+2v)x^3+\frac{8}{3}(1-v)}} - z\right)\right)^2} - \frac{2}{7}\frac{1-v}{1+2v}\left(\sqrt{\frac{1+2v}{3}}\left(\int\limits_1^\infty \frac{dx}{\sqrt{\frac{4}{3}(1+2v)x^3+\frac{8}{3}(1-v)}} - z\right)\right)^4 + \mathcal{O}\left(\left(\sqrt{\frac{1+2v}{3}}\left(\int\limits_1^\infty \frac{dx}{\sqrt{\frac{4}{3}(1+2v)x^3+\frac{8}{3}(1-v)}} - z\right)\right)^{10}\right). \tag{4}$$

We have further

$$\sqrt{\frac{1+2v}{3}}\int\limits_1^\infty \frac{dx}{\sqrt{\frac{4}{3}(1+2v)x^3+\frac{8}{3}(1-v)}} = {}_2F_1\left(\begin{matrix}\frac{1}{2},\frac{1}{6}\\ \frac{7}{6}\end{matrix}\,\middle|\, -\frac{2(1-v)}{1+2v}\right),$$

which can be seen by performing a series expansion of the integrand:

$$\int\limits_1^\infty \frac{dx}{\sqrt{\frac{4}{3}(1+2v)x^3+\frac{8}{3}(1-v)}} = \frac{1}{\sqrt{\frac{4}{3}(1+2v)}}\int\limits_1^\infty \sum_{k\ge 0}\binom{-\frac{1}{2}}{k}\left(\frac{2(1-v)}{1+2v}\right)^k \frac{dx}{x^{3k+\frac{3}{2}}}$$
$$= -\frac{1}{\sqrt{\frac{4}{3}(1+2v)}}\sum_{k\ge 0}\binom{-\frac{1}{2}}{k}\left(\frac{2(1-v)}{1+2v}\right)^k \frac{1}{3k+\frac{1}{2}}\frac{1}{x^{3k+\frac{1}{2}}}\Bigg|_{x=1}^{\infty}$$
$$= \frac{1}{\sqrt{\frac{4}{3}(1+2v)}}\sum_{k\ge 0}\binom{-\frac{1}{2}}{k}\left(\frac{2(1-v)}{1+2v}\right)^k \frac{1}{3k+\frac{1}{2}} = \frac{1}{\sqrt{\frac{1+2v}{3}}}\,{}_2F_1\left(\begin{matrix}\frac{1}{2},\frac{1}{6}\\ \frac{7}{6}\end{matrix}\,\middle|\, -\frac{2(1-v)}{1+2v}\right).$$

The function $F(z,v)$ is now obtained by integration of $G(z,v)$ and we get the following expansion of $F(z,v)$ around its singularity $s(v)$:

$$F(z,v) = \frac{2}{\sqrt{\frac{4}{3}(1+2v)}\,{}_2F_1\left(\begin{matrix}\frac{1}{2},\frac{1}{6}\\ \frac{7}{6}\end{matrix}\,\middle|\, -\frac{2(1-v)}{1+2v}\right)}\frac{1}{1-\frac{\sqrt{\frac{4}{3}(1+2v)}}{2\,{}_2F_1\left(\begin{matrix}\frac{1}{2},\frac{1}{6}\\ \frac{7}{6}\end{matrix}\,\middle|\, -\frac{2(1-v)}{1+2v}\right)}z} + h(v) + \mathcal{O}\left(\left({}_2F_1\left(\begin{matrix}\frac{1}{2},\frac{1}{6}\\ \frac{7}{6}\end{matrix}\,\middle|\, -\frac{2(1-v)}{1+2v}\right) - \frac{\sqrt{\frac{4}{3}(1+2v)}}{2}z\right)^5\right)$$

with a certain function $h(v)$. Singularity analysis leads immediately to

$$[z^n]F(z,v) = e^{n\log f(v)+\log g(v)}\left(1+\mathcal{O}\Big(\frac{1}{n^4}\Big)\right)$$

with

$$f(v) = \frac{\sqrt{\frac{1+2v}{3}}}{{}_2F_1\left(\begin{matrix}\frac12,\frac16\\ \frac76\end{matrix}\,\Big|\, -\frac{2(1-v)}{1+2v}\right)} \quad \text{and} \quad g(v) = \frac{1}{\sqrt{\frac{1+2v}{3}}\;{}_2F_1\left(\begin{matrix}\frac12,\frac16\\ \frac76\end{matrix}\,\Big|\, -\frac{2(1-v)}{1+2v}\right)}.$$

This representation allows an application of Hwang's *quasi-power theorem*. For the reader's convenience we include this important theorem here:

Theorem 2.3. *Assume that the Laplace transforms $\lambda_n(s) = \mathbb{E}\{e^{sX_n}\}$ of a sequence of random variables X_n are analytic in a disc $|s| < \rho$, for some $\rho > 0$, and satisfy there an expansion of the form*

$$\lambda_n(s) = e^{\beta_n U(s)+V(s)}\left(1 + \mathcal{O}\Big(\frac{1}{\kappa_n}\Big)\right),$$

with β_n, $\kappa_n \to +\infty$, and $U(s)$, $V(s)$ analytic in $|s| \le \rho$. Assume also the variability condition,

$$U''(0) \ne 0.$$

Under these assumptions, the mean and variance of X_n satisfy

$$\mathbb{E}\{X_n\} = \beta_n U'(0) + V'(0) + \mathcal{O}(\kappa_n^{-1}), \quad \mathbb{V}\{X_n\} = \beta_n U''(0) + V''(0) + \mathcal{O}(\kappa_n^{-1}).$$

The distribution of X_n is asymptotically Gaussian and the speed of convergence to the Gaussian limit is $\mathcal{O}(\kappa_n^{-1} + \beta_n^{-1/2})$:

$$\mathbb{P}\left\{\frac{X_n - \beta_n U'(0)}{\sqrt{\beta_n U''(0)}} \le x\right\} = \Phi(x) + \mathcal{O}\left(\frac{1}{\kappa_n} + \frac{1}{\sqrt{\beta_n}}\right).$$

$\Phi(x)$ denotes here the distribution function of the standard (Gaussian) normal distribution $\mathcal{N}(0,1)$.

Thus we have, with $U(s) = \log f(e^s)$ and $V(s) = \log g(e^s)$:

Theorem 2.4. *The expectation and the variance of the distribution X_n of number of rotations, when generating a fringe balanced binary search tree fulfill*

$$\mathbb{E}\{X_n\} = U'(0)n + V'(0) + \mathcal{O}\Big(\frac{1}{n^4}\Big) = \frac{2}{7}n - \frac{8}{21} + \mathcal{O}\Big(\frac{1}{n^4}\Big),$$
$$\mathbb{V}\{X_n\} = U''(0)n + V''(0) + \mathcal{O}\Big(\frac{1}{n^4}\Big) = \frac{66}{637}n - \frac{680}{5733} + \mathcal{O}\Big(\frac{1}{n^4}\Big)$$

and the distribution is asymptotically Gaussian:

$$\mathbb{P}\left\{\frac{X_n - \frac{2}{7}n}{\sqrt{\frac{66}{637}n}} \le x\right\} = \Phi(x) + \mathcal{O}\Big(\frac{1}{\sqrt{n}}\Big).$$

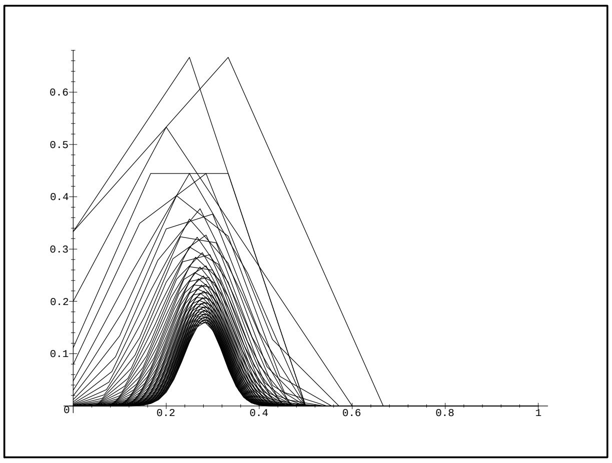


FIGURE 2. The histogram of the distribution of the number of rotations for a fringe-balanced tree of size $n = 3 \dots 60$

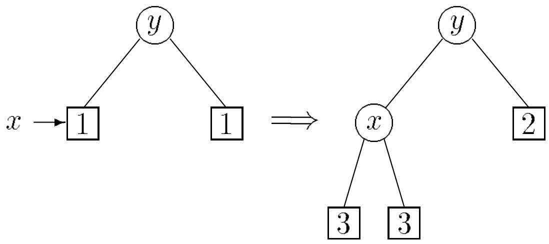
3. The associated urn model

We also want to discuss the generalized Pólya urn model, which was used in [7] to obtain exact formulæ for the expectation and variance of the number of rotations in fringe balanced trees. The method to describe the rotations in a fringe balanced tree is the following distinction of the leaves in the tree into three classes (or colors). In our context, the leaves of an (extended) binary tree with n internal nodes are the $n+1$ external nodes. Then we have the following partition of the nodes into 3 colors. Color 1 represents a local balancing at this point. This means that neither the subtree starting with the internal node associated to the leaf in question nor the subtree starting with the ascendant of this internal node is of size 2. A local imbalancing is coded by the leaves with colors 2 and 3: When the subtree starting with the internal node associated to the leaf in question is of size 2, then the leaf is of color 2, whereas the other leaves of this subtree are colored with color 3. When inserting a new element into the tree at the position of a leaf of color 1, the new tree loses two leaves of color 1 and gaines two leaves of color 3 and one leaf of color 2. Inserting a new element into the tree at the position of a leaf of color 2 or 3, the new tree loses two leaves of color 3 and one leaf of color 2, but has four more leaves of color 1. Inserting an element at the position of a leaf of color 3 requires exactly one rotation, to fulfill the balancing condition. Compare with Figure 3.

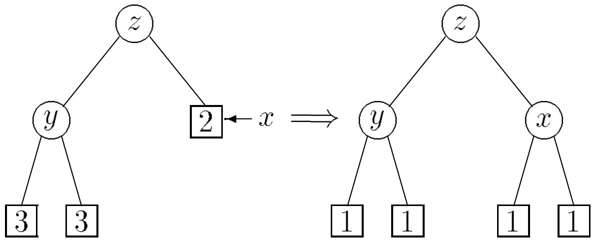
Now the associated Pólya urn model can be formulated, when thinking of the leaves of a fringe-balanced binary search tree as balls in an urn, which starts out with two balls of color 1 and the 3×3 ball addition matrix A, whose (i, j) entry is the number of balls of type j to add when a ball of color j is picked is given as

$$A = \begin{pmatrix} -2 & 1 & 2 \\ 4 & -1 & -2 \\ 4 & -1 & -2 \end{pmatrix}.$$

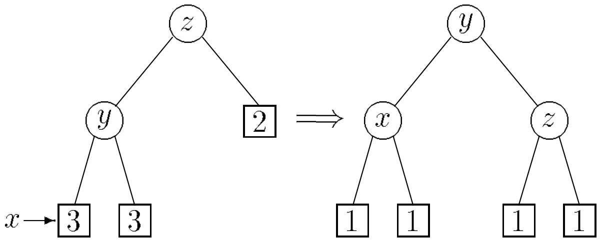
The random variable R_n "number of rotations made when constructing a fringe balanced tree with n elements" is in the urn model given as the number of picks of elements with color 3 after $n-1$ random picks. We define the random variables $X_n^{(k)}$



Inserting an element at a leaf of color 1.



Inserting an element at a leaf of color 2.



Inserting an element at a leaf of color 3 with one rotation to balance the tree locally.

FIGURE 3. Insertions in a fringe-balanced binary search tree.

to denote the number of balls of color k in the urn, when $n+1$ elements are in the urn altogether (after $n-1$ random picks).

In the following we want to give an alternative calculation of the following lemma, which was obtained in [7]:

Lemma 3.1. *The exact averages and covariances for the number of balls in the urn after n picks are given by:*

$$\begin{pmatrix} \mathbb{E}\{X_n^{(1)}\} \\ \mathbb{E}\{X_n^{(2)}\} \\ \mathbb{E}\{X_n^{(3)}\} \end{pmatrix} = \frac{n+1}{7} \begin{pmatrix} 4 \\ 1 \\ 2 \end{pmatrix}, \quad \text{for } n \geq 6,$$

$$Cov(X_n^{(1)}, X_n^{(2)}, X_n^{(3)}) = \frac{48(n+1)}{637} \begin{pmatrix} 9 & -3 & -6 \\ -3 & 1 & 2 \\ -6 & 2 & 4 \end{pmatrix}, \quad \text{for } n \geq 12.$$

Proof. At first we want to remark, that only the values $\mathbb{E}\{X_n^{(2)}\}$ and $\mathbb{V}\{X_n^{(2)}\}$ must be calculated, since all other values follow by the relations $X_n^{(3)} = 2X_n^{(2)}$ and $X_n^{(1)} + X_n^{(2)} + X_n^{(3)} = n+1$.

To get the remaining cases we use the *operator method* as described in [4]. We denote by $p_{n,k}$ the probability $\mathbb{P}\{X_n^{(2)} = k\}$ and by $P_n(u)$ the generating function $\sum_{k\geq 0} p_{n,k}u^k$. This leads to the recursion

$$P_{n+1}(u) = \sum_{k\geq 0} p_{n,k}\left[\frac{3k}{n+1}u^{k-1} + \Big(1-\frac{3k}{n+1}\Big)u^{k+1}\right] = \sum_{k\geq 0}\Big(u + \frac{3}{n+1}(1-u^2)D\Big)p_{n,k}u^k$$

or when introducing the operator $\Phi_n = u + \frac{3}{n+1}(1-u^2)D$:

$$P_{n+1}(u) = \Phi_n P_n(u), \quad \text{for } n \geq 1, \quad P_1(u) = 1. \tag{5}$$

In the above, we denote the differential operator $\frac{d}{du}$ by D and the evaluation at $u=1$ by U. We get then for $n \geq 6$

$$\mathbb{E}\{X_{n+1}^{(2)}\} = UDP_{n+1}(u) = \Big(U + \big(1-\frac{6}{n+1}\big)UD\Big)P_n(u) = 1 + \frac{n-5}{n+1}\,\mathbb{E}\{X_n^{(2)}\}$$

with $\mathbb{E}\{X_6^{(2)}\} = 1$. This linear recursion is easily solved and we get

$$\mathbb{E}\{X_n^{(2)}\} = \frac{1}{7}(n+1) \quad \text{for } n \geq 6.$$

In a similar way we get for $n \geq 12$ a recursion for the second factorial moments via

$$\begin{aligned}\mathbb{E}\{X_{n+1}^{(2)}(X_{n+1}^{(2)}-1)\} &= UD^2P_{n+1}(u) = \Big(2\,\frac{n-2}{n+1}\,UD + \frac{n-11}{n+1}\,UD^2\Big)P_n(u)\\ &= \frac{n-11}{n+1}\,\mathbb{E}\{X_n^{(2)}(X_n^{(2)}-1)\} + 2\,\frac{n-2}{n+1}\,\mathbb{E}\{X_n^{(2)}\}\end{aligned}$$

with $\mathbb{E}\{X_{12}^{(2)}(X_{12}^{(2)}-1)\} = \frac{18}{7}$. Inserting the already computed values of $\mathbb{E}\{X_n^{(2)}\}$ and solving the appearing first order linear recursion leads to

$$\mathbb{E}\{X_n^{(2)}(X_n^{(2)}-1)\} = \frac{1}{637}(n+1)(13n-30)$$

resp. to

$$\mathbb{V}\{X_n^{(2)}\} = \frac{48}{637}(n+1) \quad \text{for } n \geq 12.$$

□

It is also possible to get the expectation and the variance of R_n with this method, but it seems to be necessary to introduce bivariate generating functions $R_n(u,v) = \sum_{k,l} p_{n,k,l}u^kv^l$, where $p_{n,k,l}$ denotes the probability, that after $n-1$ random picks k balls in the urn are of color 2 and l times a ball of color 3 was chosen. We get then the recursion

$$\begin{aligned}R_{n+1}(u,v) &= \sum_{k,l} p_{n,k,l}\left(\frac{k}{n+1}u^{k-1}v^l + \frac{2k}{n+1}u^{k-1}v^{l+1} + \Big(1-\frac{3k}{n+1}\Big)u^{k+1}v^l\right)\\ &= \Phi_n R_n(u,v)\end{aligned}$$

with the operator $\Phi_n = u + \frac{1+2v-3u^2}{n+1}D_u$ and $R_1(u,v) = 1$. Here we denote by D_u and D_v the differential operator w.r.t. u resp. v and by U_u and U_v the evaluations at $u=1$ resp. $v=1$.

For the expectations of R_n we get the recursion

$$\begin{aligned}\mathbb{E}\{R_{n+1}(u,v)\} &= U_u U_v D_v R_{n+1}(u,v) = \Big(\frac{2}{n+1} U_u U_v D_u + U_u U_v D_v\Big) R_n(u,v) \\ &= \mathbb{E}\{R_n(u,v)\} + \frac{2}{n+1}\mathbb{E}\{X_n^{(2)}\},\end{aligned}$$

which can be solved easily after insertion of the previously calculated expectation $\mathbb{E}\{X_n^{(2)}\}$, and we get

$$\mathbb{E}\{R_n\} = \frac{2}{7}n - \frac{8}{21} \quad \text{for } n \geq 6.$$

For the second factorial moments of R_n we get

$$\mathbb{E}\{R_{n+1}(R_{n+1}-1)\} = U_u U_v D_v^2 R_{n+1}(z,v) = \mathbb{E}\{R_n(R_n-1)\} + \frac{4}{n+1}\mathbb{E}\{R_n X_n^{(2)}\}.$$

Thus we also need $\mathbb{E}\{R_n X_n^{(2)}\}$ which we obtain via the recursion

$$\begin{aligned}\mathbb{E}\{R_{n+1}X_{n+1}^{(2)}\} &= U_u U_v D_u D_v R_{n+1}(z,v) \\ &= \Big(1 - \frac{6}{n+1}\Big)\mathbb{E}\{R_n X_n^{(2)}\} + \mathbb{E}\{R_n\} + \frac{2}{n+1}\mathbb{E}\{X_n^{(2)}(X_n^{(2)}-1)\}.\end{aligned}$$

Inserting the already obtained results and solving this recursions leads finally to

$$\mathbb{E}\{R_n(R_n-1)\} = \frac{4}{49}n^2 - \frac{764}{1911}n + \frac{2336}{5733} \quad \text{for } n \geq 12$$

from which the variance follows.

References

[1] M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions.* Dover Publications, Inc., New York, ninth edition, 1970.

[2] Ph. Flajolet and A. M. Odlyzko. Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics*, 3:216–240, 1990.

[3] Ph. Flajolet and R. Sedgewick. The average case analysis of algorithms: Multivariate asymptotics and limit distributions. Technical Report #3162, INRIA, 1997.

[4] D. H. Greene and D. E. Knuth. *Mathematics for the Analysis of Algorithms.* Birkhäuser, Boston, second edition, 1982.

[5] H.-K. Hwang. *Théorèmes limites pour les structures combinatoires et les fonctions arithmétiques.* PhD thesis, Ecole Polytechnique, 1994.

[6] E. Kamke. *Differentialgleichungen - Lösungsmethoden und Lösungen*, volume 1. Teubner, Stuttgart, tenth edition, 1983.

[7] H. Mahmoud. On rotations in fringe-balanced binary trees. *Information Processing Letters*, 65:41–46, 1998.

[8] C. Martínez, A. Panholzer, and H. Prodinger. On the number of descendants and ascendants in random search trees. *Electronic Journal of Combinatorics*, 5(1):#R20, 1998.

[9] P. Poblete and I. Munro. The analysis of a fringe heuristic for binary search trees. *Journal of Algorithms*, 6:336–350, 1985.

[10] R. Sedgewick and Ph. Flajolet. *Introduction to the Analysis of Algorithms.* Addison-Wesley, Reading, Massachusetts, 1996.

[11] R. Smythe. Central limit theorems for urn models. *Stochastic Processes and their Applications*, 65:115–137, 1996.

[12] W. Walker and D. Wood. Locally balanced binary search trees. *The Computer Journal*, 19:322–325, 1976.

[13] H. Wenzel and P. Meinhold. *Gewöhnliche Differentialgleichungen.* Teubner, Stuttgart, 1994.
[14] E. T. Whittaker and G. N. Watson. *A course of modern analysis.* University Press, Cambridge, fourth edition, 1927.

Institut für Algebra und Diskrete Mathematik, Technical University of Vienna, Wiedner Hauptstrasse 8-10, A-1040 Vienna, Austria.
E-mail address: Helmut.Prodinger@tuwien.ac.at
E-mail address: alois.panholzer+e1182@tuwien.ac.at