

# A Top-Down Analysis of Fringe-Balanced Binary Search Trees

*Helmut Prodinger*

Technical University of Vienna

May 25, 1998

[summary by Michèle Soria]

## Abstract

Fringe-balanced binary search trees are obtained by performing rotations only on subtrees of size three. The parameter “number of rotations” has recently been studied by Mahmoud [3], using a Pólya urn model. This talk, based on [5] proposes a top-down approach of the problem, that leads to a differential equation. The solution is related to the Weierstrass’  $\wp$ -function. This fact allows to derive the asymptotic normality of the parameter by means of Hwang’s quasi-power theorem [2]. An alternative way of obtaining the exact expectation and variance, which relies on operator calculus, is also presented.

It is well-known that in a random binary search tree constructed by insertion at the leaves, the average depth of a node is logarithmic in the size of the tree, so that retrieval of the data stored in the nodes can be done efficiently. One simple way to improve the speed of retrieval even more is to compress the subtrees near the leaves by doing a *fringe-balanced rotation*: 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. 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 under the random permutation model was recently analyzed by Mahmoud [3], using a Pólya urn model, and a central limit theorem for urn models by Smythe [6]. Here is presented an alternative way, based on analytic methods, of proving the Gaussian limiting distribution. This presentation follows [5].

## 1. Top-Down Approach

The recursive top-down analysis (see [4] for various uses of this approach) begins with a recurrence relation based on *splitting probabilities*. When constructing a fringe-balanced tree from a random permutation, the first three elements of the permutation determine the root of the final tree, as well as whether or not there is a rotation at the root (a rotation occurs in four cases out of six).

Hence the splitting probability  $\pi_{n,k}$ , which is the probability that in a tree of size  $n$  the root is the node  $k$ , is given by

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

for  $n \geq 3$  and  $1 \leq k \leq n$ ,  $\pi_{2,1} = \pi_{2,2} = 1/2$  and  $\pi_{1,1} = 1$ . And we also get a recurrence relation for the probability  $F_{n,m}$ , that the number of rotations is  $m$ , when generating a fringe balanced tree of size  $n$ , starting with an empty tree (the number of rotations to construct the root of the tree is 1

with probability  $2/3$ , and  $0$  with probability  $1/3$ ). Thus for  $n \geq 3$  and  $1 \leq m \leq n$ :

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

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  $F(z, v) = \sum_{n,m \geq 0} F_{n,m} z^n v^m$  this recurrence leads to the 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  $F(0, v) = 1$ ,  $\frac{\partial}{\partial z} F(z, v)|_{z=0} = 1$  and  $\frac{\partial^2}{\partial z^2} F(z, v)|_{z=0} = 2$ .

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

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

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

**1.1. Moments.** The moments of the distribution are obtained by differentiating equation (2) with respect to  $v$ , and evaluating at  $v = 1$ . Let

$$M_1(z) = \frac{\partial}{\partial v} G(z, v) \Big|_{v=1} = \sum_{n \geq 0} n M_n^{(1)} z^{n-1}, \quad \text{and} \quad M_2(z) = \frac{\partial^2}{\partial v^2} G(z, v) \Big|_{v=1} = \sum_{n \geq 0} n M_n^{(2)} z^{n-1},$$

where  $M_n^{(1)}$  and  $M_n^{(2)}$  denote the first and second factorial moments of the number of rotations.

$M_1(z)$  and  $M_2(z)$  both satisfy an Euler differential equation. Extracting the coefficients of the solutions  $M_1(z)$  and  $M_2(z)$  leads to exact values for the expectation  $M_n = M_n^{(1)}$  and the variance  $\text{Var}_n = M_n^{(2)} + M_n^{(1)} - \left(M_n^{(1)}\right)^2$ :

**Theorem 1.** *The expectation and the variance of the number of rotations when generating a fringe balanced binary search tree of size  $n$  are given by*

$$M_n = \frac{2}{7}n - \frac{8}{21} \quad (n \geq 6), \quad \text{Var}_n = \frac{66}{637}n - \frac{680}{5733} \quad (n \geq 12).$$

**1.2. Limiting Distribution.** Equation (2) transforms into

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

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

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

This form shows a close relation between  $G(z, v)$  and the Weierstrass'  $\wp$ -function, which can be characterized, within a simply connected domain of  $\mathbb{C}_\infty$ , which contains no zeros of the denominator, by the integral  $\zeta = \int_{\wp(\zeta)}^\infty dx / \sqrt{4x^3 - g_2x - g_3}$ . Constants  $g_2$  and  $g_3$  are called the *invariants* of  $\wp$ .

Indeed, making the substitution

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

it is shown in [5] that  $W(t, v) = G(z(t), v)$  is a Weierstrass'  $\wp$ -function with invariants  $g_2 = 0$  and  $g_3 = -8(1 - v)/(1 + 2v)$ .

Since  $\wp(\zeta)$  has a double pole at  $\zeta = 0$ , it follows that  $G(z, v)$  has a double pole at  $z = s(v)$ , where it admits the local expansion  $G(z, v) = \frac{3}{1+2v} (z - s(v))^{-2} + \mathcal{O}(z - s(v))^4$ .

Integrating term by term, we obtain the expansion of  $F(z, v)$ :

$$F(z, v) = \frac{1}{\frac{1+2v}{3}s(v) \left(1 - \frac{z}{s(v)}\right)} + \mathcal{O}(z - s(v))^5.$$

Singularity analysis leads immediately to the expansion of the coefficients:

$$(3) \quad [z^n]F(z, v) = \frac{1}{\frac{1+2v}{3}s(v)} s(v)^{-n} + \left(1 + \mathcal{O}\left(\frac{1}{n^4}\right)\right)$$

Performing a series expansion of the integrand, the integral  $s(v)$  can be expressed in terms of a hypergeometric function:

$$\sqrt{\frac{1+2v}{3}}s(v) = {}_2F_1\left(\frac{1}{2}, \frac{1}{6} \middle| -\frac{2(1-v)}{1+2v}\right).$$

In (3), the probability generating function of the number of rotations is given in a form that satisfies the hypothesis of Hwang's quasi-power theorem [2], so that we finally get the central limit theorem:

**Theorem 2.** *The distribution  $X_n$  of the number of rotations, when generating a fringe balanced binary search tree is asymptotically Gaussian:*

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

## 2. Urn Model and Operator Calculus

Insertions in fringe-balanced binary search trees can be translated into an urn model of Pólya. The urn contains balls of three different colors, corresponding to the leaves of the tree: binary subtrees of size 3 (which are always complete) have four leaves of color 1; in subtrees of size 2, the two deepest leaves are colored by 3, and the third one is colored by 2. We start with an urn containing two balls of color 1, corresponding to a starting tree with one internal node. Inserting at a leaf of color 1, i.e. picking a ball of color 1, results in replacing two leaves of color 1 by one leaf of color 2 and two leaves of color 3. Inserting at a leaf of color 2 results in replacing this leaf and its two associated leaves of color 3 by four leaves of color 1. In the same manner, inserting at a leaf of color 3 results in replacing two leaves of color 3 and one leaf of color 2 by four leaves of color 1. Thus the process of insertion translates into the ball addition matrix  $A$ , whose  $(i, j)$  entry is the number of balls of type  $j$  to be added when a ball of color  $i$  is picked:

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

Working on the addition matrix, Mahmoud [3] obtains the exact averages and covariances for the number of balls in the urn after  $n$  picks, as well as the exact average and variance of the number of rotations after  $n$  random insertions in an empty fringe balanced tree.

These results can also be obtained by operator calculus [1]. Define the random variables  $X_n^{(k)}$  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). 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  $X_n^{(1)} + X_n^{(2)} + X_n^{(3)} = n + 1$ . We denote by  $p_{n,k}$  the probability  $\mathbb{P}\{X_n^{(2)} = k\}$  and by  $P_n(u)$  the generating function  $\sum_{k \geq 0} p_{n,k} u^k$ . The urn picking process leads to the recursion

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

where  $D$  denotes the differential operator  $\frac{d}{du}$ . If evaluation at  $u = 1$  is denoted by operator  $U$ , we get then for  $n \geq 6$ :

$$\mathbb{E}\{X_{n+1}^{(2)}\} = U D P_{n+1}(u) = \left( U + \left(1 - \frac{6}{n+1}\right) U D \right) P_n(u) = 1 + \frac{n-5}{n+1} \mathbb{E}\{X_n^{(2)}\}$$

with  $\mathbb{E}\{X_6^{(2)}\} = 1$ . This linear recursion is easily solved and we get  $\mathbb{E}\{X_n^{(2)}\} = \frac{1}{7}(n+1)$  for  $n \geq 6$ . A similar computation holds for the second factorial moment  $\mathbb{E}\{X_{n+1}^{(2)}(X_{n+1}^{(2)} - 1)\} = U D^2 P_{n+1}(u)$ , from which the variance is obtained.

The average and variance of the number of rotations can also be treated in this way. Let the random variable  $R_n$  denote the number of rotations made when constructing a fringe balanced tree with  $n$  elements (or, in the urn model, the number of picks of elements with color 3 after  $n - 1$  random picks). Introduce the bivariate generating functions  $R_n(u, v) = \sum_{k,l} p_{n,k,l} u^k v^l$ , 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. Following the picking process, the recursion for  $R_n$  is

$$R_{n+1}(u, v) = \sum_{k,l} p_{n,k,l} \left( \frac{k}{n+1} u^{k-1} v^l + \frac{2k}{n+1} u^{k-1} v^{l+1} + \left(1 - \frac{3k}{n+1}\right) u^{k+1} v^l \right).$$

Denoting 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.  $v = 1$ , we get for the expectation  $\mathbb{E}\{R_{n+1}(u, v)\} = U_u U_v D_v R_{n+1}(u, v)$

$$\mathbb{E}\{R_{n+1}(u, v)\} = \left( \frac{2}{n+1} U_u U_v D_u + U_u U_v D_v \right) R_n(u, v) = \mathbb{E}\{R_n(u, v)\} + \frac{2}{n+1} \mathbb{E}\{X_n^{(2)}\}.$$

The average value of  $R_n$  is obtained by solving this recursion, and similar computations lead to the variance.

## Bibliography

- [1] Greene (D. H.) and Knuth (D. E.). – *Mathematics for the analysis of algorithms*. – Birkhäuser, Boston, 1982, second edition.
- [2] Hwang (H.-K.). – *Théorèmes limites pour les structures combinatoires et les fonctions arithmétiques*. – PhD thesis, École Polytechnique, December 1994.
- [3] Mahmoud (H.). – On rotations in fringe-balanced binary search trees. *Information Processing Letters*, vol. 65, 1998, pp. 41–46.
- [4] Martinez (C.), Panholzer (A.), and Prodinger (H.). – On the number of descendants and ascendants in random search trees. *Electronic Journal of Combinatorics*, vol. 5(1):#R20, 1998.
- [5] Panholzer (A.) and Prodinger (H.). – An analytic approach for the analysis of rotations in fringe-balanced binary search trees. – To appear in *Annals of Combinatorics*, 1998.
- [6] Smythe (R.). – Central limit theorems for urn models. *Stochastic Processes and their Applications*, vol. 65, 1996, pp. 115–137.