The Subtree Size Profile of Bucket Recursive Trees

Ramin Kazemi
Department of Statistics, Imam Khomeini International University, Qazvin, Iran.
E-mail: r.kazemi@sci.ikiu.ac.ir

Abstract. Kazemi (2014) introduced a new version of bucket recursive trees as another generalization of recursive trees where buckets have variable capacities. In this paper, we get the $p$-th factorial moments of the random variable $S_n$ which counts the number of subtrees size-1 profile (leaves) and shows a phase change of this random variable. These can be obtained by solving a first order partial differential equation for the generating function correspond to this quantity.

Keywords: Bucket recursive tree, Subtree size profile, Factorial moments.

2000 Mathematics subject classification: 05C05, 60F05.

1. Introduction

Trees are defined as connected graphs without cycles, and their properties are basics of graph theory. For example, a connected graph is a tree, if and only if the number of edges equals the number of nodes minus 1. Furthermore, each pair of nodes is connected by a unique path [1]. A rooted tree is a tree with a countable number of nodes, in which a particular node is distinguished from the others and called the root node [32].

The node profile is defined as the number of nodes at distance $k$ from the root in a tree. Several studies have been concerned on this quantity; for random binary search trees and recursive trees see [4, 5, 9, 10] and [19]; for random
plane-oriented recursive trees see [20]; for other types of random trees see [8, 11, 12, 30] and [26].

There is another kind of profile which is defined as the number of subtrees of size \( k \). This kind is called **subtree size profile** and has been investigated for random binary search trees, random recursive trees and random Catalan trees; see [3, 6, 13, 14, 15] and [18].

This kind of profile is an important tree characteristic carrying a lot of information on the shape of a tree. For instance, total path length (sum of distances of all nodes to the root) and Wiener index (sum of distances between all nodes) can be easily computed from the subtree size profile [17]. Also, studying patterns in random trees is an important issue with many applications in computer science (see [7] and [16]) and mathematical biology (see [3] and [31]).

Meir and Moon [28] defined recursive trees as the variety of non-plane increasing trees [2] such that all node degrees are allowed. In this model, the capacity of nodes is 1 [21]. Mahmoud and Smythe [29] introduced bucket recursive trees as a generalization of random recursive trees where the capacity of buckets is fixed. In this paper, we will consider another bucket recursive trees, i.e., bucket recursive trees with variable capacities of buckets that introduced by Kazemi (2014). He studied the following random variables in this model: the depth of the largest label [23], the first Zagreb index [22], the eccentric connectivity index [24] and the branches [25]. Also, Kazemi and Haji showed a phase change in the distribution in these models [27]. Our results for \( b = 1 \) reduce to the previous results for random recursive trees [13, 14]. We define the tree below for the reader’s convenience [23].

**Definition 1.1.** A size-\( n \) bucket recursive tree \( T_n \) with variable bucket capacities and maximal bucket size \( b \) starts with the root labeled by 1. The tree grows by progressive attraction of increasing integer labels: when inserting label \( j + 1 \) into an existing bucket recursive tree \( T_j \), except the labels in the non-leaf nodes with capacity \( < b \) all labels in the tree (containing label 1) compete to attract the label \( j + 1 \). For the root node and nodes with capacity \( b \), we always produce a new node \( j + 1 \). But for a leaf with capacity \( c < b \), either the label \( j + 1 \) is attached to this leaf as a new bucket containing only the label \( j + 1 \) or is added to that leaf and make a node with capacity \( c + 1 \). This process ends with inserting the label \( n \) (i.e., the largest label) in the tree.

By definition, a node \( v \) with capacity \( c(v) < b \) has the out-degree 0 or 1. In Figure 1, we diagrammatically show the step-by-step growth of a tree of size 11 with \( b = 2 \). We consider the random variable \( S_{n,k} \), which counts the number of buckets that are the root of a subtree of \( T_n \) with size \( k \). More precisely, we study the subtree size profile \( S_{n,1} \) in our model (=leaves).
The Subtree Size Profile of Bucket Recursive Trees

2. Partial Differential Equation

A class $\mathcal{T}$ of a family of bucket-increasing trees can be defined in the following way (see [23, Section 2] for details). A sequence of non-negative numbers $(\alpha_k)_{k \geq 0}$ with $\alpha_0 > 0$ and a sequence of non-negative numbers $\beta_1, \beta_2, \ldots, \beta_{b-1}$ are used to define the weight $w(T)$ of any ordered tree $T$ by $w(T) = \Pi_v w(v)$, where

$$w(v) = \begin{cases} \alpha d(v), & v \text{ is the root or } c(v) = b \\ \beta c(v), & c(v) < b \end{cases} \quad (2.1)$$

and $d(v)$ denotes the out-degree of node $v$. Let $\mathcal{L}(T)$ be the set of different increasing labelings of the tree $T$ with distinct integers $\{1, 2, ..., |T|\}$ (|$|$ denotes the size of sets). Then the family $\mathcal{T}$ consists of all trees $T$ together with their weights $w(T)$ and the set of increasing labelings $\mathcal{L}(T)$. We define the exponential generating function

$$T_{n,b}(z) = \sum_{n=1}^{\infty} T_{n,b} \frac{z^n}{n!}, \quad (2.2)$$
where \( T_{n,b} := \sum_{|T|=n} w(T) \cdot L(T) \) is the total weights and \( L(T) := |\mathcal{L}(T)| \). If \( r \) is the out-degree of the root node, then

\[
T_{n,b} = \frac{(n-1)! (b)^{n-(\sum_{k=1}^{r} |P_{k}|)} \cdot n!}{n}, \quad n \geq 1, b \geq 1,
\]

(2.3)

where \( P_{k} \) is the set of all trees of size \( k \) and \( T_{n,b}(0) = 0 \) \[23]\). Let \( S_k(z, u) \) be the moment generating function

\[
S_k(z, u) = \sum_{n \geq 1} \sum_{m \geq 0} \mathbb{P}(S_{n,k} = m) T_{n,b}^{\frac{z^n}{n!} u^m}
\]

\[
= \sum_{n \geq 1} \sum_{m \geq 0} \mathbb{P}(S_{n,k} = m) (b)^{(n-(\sum_{k=1}^{r} |P_{k}|))} z^n u^m.
\]

(2.4)

According to the definition of the tree, the probabilities of \( \mathbb{P}(S_{n,k} = m) \) satisfy

\[
\mathbb{P}(S_{n,k} = m) = \sum_{r \geq 1} \frac{1}{n_1 + \cdots + n_r = n} \binom{n-1}{n_1, \ldots, n_r} T_{n_1, b}^{*} \cdots T_{n_r, b}^{*} \mathbb{P}(S_{n_1, k} = m_1) \cdots \mathbb{P}(S_{n_r, k} = m_r),
\]

(2.5)

with initial values \( \mathbb{P}(S_{k,k} = 1) = 1, \mathbb{P}(S_{n,k} = 0) = 1 \) for \( 1 \leq n < k \) where \( T_{n_r, b}^{*} \) is the total weights of the \( i \)th subtree. Thus recurrence (2.5) leads to the following functional equation \[23]\]

\[
\frac{\partial}{\partial z} S_k(z, u) = b^{l-\sum_{k=1}^{r} |P_{k}|} \left( e^{S_{k}(z,u)} + (u-1)z^{k-1} \right), \quad (k \geq 1)
\]

(2.6)

with initial condition \( S_k(0, u) = 0 \).

For \( b = 1 \), i.e., random recursive trees, Feng, et al. obtained a limit theorem for the subtree size profile by considering both \( k \) fixed and \( k = k(n) \) dependent on \( n \). Using analytic methods they characterized for the tree the phase change behavior of \( S_{n,k} \) \[14]\).

For \( b > 1 \), there is no unique solution of (2.6) for all \( k \). Suppose \( \beta(r, b) = b^{l-\sum_{k=1}^{r} |P_{k}|} \). Then

\[
S_1(z, u) = -\log \left( e^{-z(u-1)} \beta(r, b) \left( 1 + \frac{1}{u-1} \right) - \frac{1}{u-1} \right).
\]

(2.7)

3. Preliminaries

We can rewrite \( S_1(z, u) \) as follows:

\[
S_1(z, u) = \log \left( \frac{1}{1 - \int_0^1 e^{(u-1) \beta(r, b) \theta} \, d\theta} \right) + (u-1) z \beta(r, b).
\]

(3.1)
Set $y = u - 1$ and $\beta(r, n, b) = b^{-1}(b!)^n(1 - \sum_{i=1}^{\infty} |P_i|)$. Thus
\[
S_1(z, 1 + y) = \sum_{n \geq 1} \sum_{m \geq 0} P(S_{n,1} = m) \beta(r, n, b) \frac{z^n}{n} (1 + y)^m
\]
\[
= \sum_{n \geq 1} \sum_{m \geq 0} \sum_{p=0}^{m} m \cdot \beta(r, n, b) \frac{z^n \cdot y^p}{p!},
\]
where $m^2 = m(m - 1) \cdots (m - p + 1)$. Hence
\[
\mathbb{E}(S_{n,1}^E) = \beta(r, n, b)^{-1} np! [z^n y^p] S_1(z, 1 + y),
\]
where $[z^n] f(z)$ denote the operation of extracting the coefficient of $z^n$ in the formal power series $f(z) = \sum f_n z^n$. Also
\[
\log \left( \frac{1}{1 - \int_0^1 e^y \beta(r, b)t dt} \right) = \log \left( \frac{1}{1 - \int_0^1 \sum_{j \geq 1} \frac{(y^j \beta(r, b))^j}{j!} dt} \right)
\]
\[
= \log \left( \frac{1}{1 - z - z \sum_{j \geq 1} \frac{y^j \beta(r, b)j}{(j+1)!}} \right)
\]
\[
= \log \left( \frac{1}{1 - \frac{z}{1-z} \sum_{j \geq 1} \frac{y^j \beta(r, b)j}{(j+1)!}} \right).
\]
For $p \geq 1$,
\[
[y^p] \log \left( \frac{1}{1 - \frac{z}{1-z} \sum_{j \geq 1} \frac{y^j \beta(r, b)j}{(j+1)!}} \right).
\]
\[
= [y^p] \sum_{i=1}^{p} \left( \frac{z}{1-z} \right)^i \sum_{j \geq 1} \frac{y^j \beta(r, b)j}{(j+1)!}^i
\]
\[
= \sum_{i=1}^{p} \frac{(z \beta(r, b))^p}{i} \left( \frac{z}{1-z} \right)^i \sum_{j_1 + \cdots + j_i = p, j_k \geq 1} \frac{1}{\prod_{k=1}^{i} (j_k + 1)!}
\]
\[
= \frac{1}{p!} \sum_{i=1}^{p} \frac{(z \beta(r, b))^p}{i} \left( \frac{z}{1-z} \right)^i \sum_{j_1 + \cdots + j_i = p, j_k \geq 1} \frac{1}{\prod_{k=1}^{i} (j_k + 1)(j_1, \ldots, j_i)}.
\]
Set
\[
I(A) := \begin{cases} 
1, & \text{if } A \text{ is true} \\
0, & \text{otherwise}.
\end{cases}
\]
From (3.1),
\[
[y^p]S_1(z, 1 + y) = \frac{1}{p!} \sum_{i=1}^{p} \left( z \beta(r, b) \right)^p \left( \frac{z}{1 - z} \right)^i \\
\times \sum_{j_1 + \cdots + j_i = p, j_k \geq 1} \frac{1}{\prod_{k=1}^{i}(j_k + 1)} \left( \frac{p}{j_1, \ldots, j_i} \right) \\
+ z \beta(r, b) I(p = 1). \tag{3.2}
\]

4. Main Results

**Theorem 4.1.** Let \( \beta(r, b) = b! - \sum_{i=1}^{r} |P_k| \) and \( \beta(r, n, b) = b^{-1} (b!)^{n/(1 - \sum_{i=1}^{r} |P_k|)} \).

Then
\[
E(S_{n, 1}) = \begin{cases} 
\frac{(b-1)!}{n}, & n = 1 \\
\frac{\beta(r, b)}{2 \beta(r, n, b)}, & n \geq 2
\end{cases} \tag{4.1}
\]

and for \( p \geq 2 \),
\[
E(S_{n, 1}^p) = n \beta(r, n, b)^{-1} \sum_{i=1}^{p} \frac{\beta(r, b)^i}{i} \left( \frac{n - p - 1}{i - 1} \right) I(n \geq p + 1) \\
\times \sum_{j_1 + \cdots + j_i = p, j_k \geq 1} \frac{1}{\prod_{k=1}^{i}(j_k + 1)} \left( \frac{p}{j_1, \ldots, j_i} \right). \tag{4.2}
\]

Let \( s(m, n) \) be the \( m \)-th Stirling number of order \( n \) (of the second kind).

Then in view of the classical relation
\[
E(S_{n, 1}^p) = \sum_{i=1}^{p} s(p, i) E(S_{n, 1}^i).
\]

Thus we can get closed formulas for ordinary \( p \)-th moments.

We use the notations \( \overset{D}{\to} \) and \( \overset{P}{\to} \) to denote convergence in distribution and in probability, respectively. The standard random variable \( \text{Poi}(\lambda) \) and \( N(\mu, \sigma^2) \) appear in the following theorem for the Poisson distributed with parameter \( \lambda > 0 \) and the normal distributed with mean \( \mu \) and variance \( \sigma^2 \), respectively. These random variables appear in the results as limiting random variables.

**Theorem 4.2.** Let \( S_{n, 1} \) be the subtree size-1 profile in size-\( n \) bucket recursive trees with variable capacities of buckets. Then
\begin{enumerate} [i)]
\item \( S_{n, 1} \overset{P}{\to} 0, \) as \( \sqrt{\frac{\beta(r, n, b)}{n \beta(r, b)}} \to \infty. \)
\item \( S_{n, 1} \overset{D}{\to} \text{Poi}\left( \frac{1}{c^2} \right), \) as \( \sqrt{\frac{\beta(r, n, b)}{n \beta(r, b)}} \to c > 0, \)
\end{enumerate}
and otherwise no limiting distribution exists for $S_{n,1}$.

iii) $S_{n,1} - \frac{n \beta(r, b)}{2 \beta(r, n, b)} \xrightarrow{D} N(0, 1), \quad \text{as} \quad \sqrt{\frac{n \beta(r, n, b)}{\beta(r, b)}} \to 0.$

5. Proofs

Proof of Theorem 4.1. The formula (3.2) immediately gives

$$E(S_{n,1}) = n\beta(r, n, b) - \frac{1}{2} [z^n] \left( zr \beta(r, b) + \beta(r, b) \frac{z^2}{1 - z} \right)$$

$$= n\beta(r, n, b) - \frac{1}{2} \beta(r, n, b) I(n = 1) + n \beta(r, b) I(n \geq 2).$$

For $p \geq 2,$

$$E(S_{n,1}^p) = \beta(r, n, b)^{-1} np! [z^n y^p] S_1(z, 1 + y)$$

$$= n\beta(r, n, b)^{-1} [z^n] \sum_{i=1}^{p} \frac{\beta(r, b)^p}{i} \frac{z^{p+i}}{(1 - z)^i}$$

$$\times \sum_{j_1 + \ldots + j_i = p, j_k \geq 1} \prod_{k=1}^{i} (j_k + 1) \left( \frac{p}{j_1, \ldots, j_i} \right)$$

$$= n\beta(r, n, b)^{-1} \sum_{i=1}^{p} \frac{\beta(r, b)^p}{i} \left( \frac{n - p - 1}{i - 1} \right) I(n \geq p + 1)$$

$$\times \sum_{j_1 + \ldots + j_i = p, j_k \geq 1} \prod_{k=1}^{i} (j_k + 1) \left( \frac{p}{j_1, \ldots, j_i} \right).$$

□

Corollary 5.1. For $p = 2$,

$$E(S_{n,1}^2) = E(S_{n,1}(S_{n,1} - 1)) = n\beta(r, b)^2 \left( \frac{1}{3} + \frac{n - 3}{4} \right), \; n \geq 3.$$

For $b = 1,$

$$E(S_{n,1}) = \left\{ \begin{array}{ll}
1, & n = 1 \\
\frac{n}{2}, & n \geq 2
\end{array} \right.$$
Proof of Theorem 4.2. Suppose $\frac{n\beta(r,b)}{\beta(r,n,b)} \to \lambda > 0$. Thus from Theorem 4.1, $E(S_{n,1}) \to \lambda^2$. It is obvious that

$$\sum_{j_1 + \cdots + j_p \geq 1} \binom{p}{j_1, \ldots, j_p} = p! \left( e^{z} - 1 \right)^i \leq p! \left( e^{z} - 1 \right)^i = p^i.$$

i) For $p \geq 2$,

$$E(S_{n,1}^p) = n\beta(r,n,b)^{-1} \sum_{i=1}^{p} \frac{\beta(r,b)^p}{i} \binom{n-p-1}{i-1} \times \sum_{j_1 + \cdots + j_p \geq 1} \prod_{k=1}^{p} \frac{1}{(j_k+1)} \binom{p}{j_1, \ldots, j_p} \leq n\beta(r,n,b)^{-1} \sum_{i=1}^{p} \frac{\beta(r,b)^p}{i} \frac{n^i}{i!} i^p \leq \sum_{i=1}^{p} \frac{\beta(r,b)^p}{\beta(r,n,b)} \frac{n^i}{i!} (i-1)! i^{p-1} \leq p^{p-1} n\beta(r,b)^p \beta(r,n,b) \sum_{i=0}^{\infty} \frac{(n\beta(r,b))^{i}}{i!} \exp \left( \frac{n\beta(r,b)}{\beta(r,n,b)} \right),$$

since $p \geq i \geq 1$. By assumption $\frac{n\beta(r,b)}{\beta(r,n,b)} \to 0$. Then for all $p \geq 1$, $E(S_{n,1}^p) \to 0$. i.e., the random variable $S_{n,1}$ convergent to a degenerate distribution at point 0.

ii) From Theorem 4.1, $E(S_{n,1}^p) = A + B$, where

$$A = \sum_{i=1}^{p-1} n\beta(r,n,b)^p \binom{n-p-1}{i-1} \sum_{j_1 + \cdots + j_p \geq 1} \prod_{k=1}^{p} \frac{1}{(j_k+1)} \binom{p}{j_1, \ldots, j_p},$$

and

$$B = n\beta(r,b)^p \frac{p!}{p\beta(r,n,b)} \left( n-p-1 \right) \frac{p!}{2p^2} = \left( \frac{n\beta(r,b)}{2\beta(r,n,b)} \right)^p \left( 1 + O \left( \frac{1}{\sqrt{n}} \right) \right).$$

With the same technic of Part (i),

$$A \leq p^{p-1} \sqrt{\frac{\beta(r,b)}{\beta(r,n,b)}} n\beta(r,b) \exp \left( \frac{n\beta(r,b)}{\beta(r,n,b)} \right),$$

thus

$$A = O \left( \sqrt{\frac{\beta(r,b)}{\beta(r,n,b)}} \right) = O \left( \frac{1}{\sqrt{n}} \right).$$
since \( \frac{n\beta(r,b)}{\beta(r,n,b)} \to \lambda \). Finally for every \( p \geq 1 \),
\[
E(S_{n,1}^p) = \left( \frac{n\beta(r,b)}{2\beta(r,n,b)} \right)^p + O \left( \frac{1}{\sqrt{n}} \right) \to \left( \frac{\lambda}{2} \right)^p.
\]
Now, if we use the substitution \( c = \sqrt{\frac{2}{\lambda}} \), then \( \sqrt{\beta(r,n,b) n\beta(r,b)} \to c \) and this proves the Part (ii).

iii) Let \( S_{n,1} = S_{n,1} - E(S_{n,1}) \) and \( \overline{S}_1(z,s) = \sum_{n\geq1} E(e^{S_{n,1}}) \frac{zn}{n} \). Then
\[
\overline{S}_1(z,s) = \sum_{n\geq1} e^{-E(S_{n,1})} E(e^{S_{n,1}}) \frac{zn}{n}.
\]
From (2.4) and initial conditions of (2.5),
\[
S_1(e^{-\frac{\beta(r,b)}{2\beta(r,n,b)} z}, e^s) = \beta(r,n,b)e^{-\frac{\beta(r,b)}{2\beta(r,n,b)} z}e^s + \sum_{n\geq2} \sum_{m\geq0} P(S_{n,1} = m)(e^{-\frac{\beta(r,b)}{2\beta(r,n,b)} z}) \frac{zn}{n} e^{sm}.
\]
By (4.1),
\[
\overline{S}_1(z,s) = e^{-E(S_{1,1})} E(e^{S_{1,1}}) z + \sum_{n\geq2} e^{-E(S_{n,1})} E(e^{S_{n,1}}) \frac{zn}{n} = S_1(e^{-\frac{\beta(r,b)}{2\beta(r,n,b)} z}, e^s) + \left( e^{1 - \frac{1}{n-1}} - \beta(r,n,b) e^{-\frac{\beta(r,b)}{2\beta(r,n,b)} s} \right) z.
\]
Now by (3.1) and just similar to [13] proof is completed. \( \square \)

**Corollary 5.2.** The Theorem 4.2 for \( b = 1 \) reduce to the previous results for random recursive trees [13].

**Acknowledgments**

The author would like to expresses its sincere thanks to the referees for their valuable suggestions and comments.

**References**