Global and Stochastic Analysis
Vol. 8 No. 1 (January-June, 2021)

MUK PUBLICATIONS
Open Access Publisher

STOCHASTIC ANALYSIS OF THE SUBTREE SIZE PROFILE

R. KAZEMI

Abstract. We consider the random variable $S_{n,k}$, which counts the number of subtrees size-k profile in bucket recursive trees with variable bucket capacities. We get the p-th factorial moments of the random variable $S_{n,k}$ and show 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 with using a suitable substitution that satisfies a Riccati equation.

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. 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.

The *node profile* is defined as the number of nodes at distance k from the root in a tree. 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 [1, 2, 4, 5, 6] and. 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. Also, studying patterns in random trees is an important issue with many applications in computer science (see [3] and [7]) and mathematical biology [1].

Meir and Moon [11] defined recursive trees as the variety of non-plane increasing trees such that all node degrees are allowed. In this model, the capacity of nodes is 1. In this paper, we will consider the bucket recursive trees with variable capacities of buckets. Mahmoud and Smythe introduced bucket recursive trees as a generalization of random recursive trees [10]. In this model the bucket is a node that can hold up to $b \geq 1$ labels. The capacity of a bucket v $(c = c(v))$ is defined

Date: Date of Submission February 12, 2020 ; Date of Acceptance June 01, 2020 , Communicated by Yuri E. Gliklikh .

2000 *Mathematics Subject Classification.* Primary 60F05 (Central limit and other weak theorems); Secondary 05C05 (Trees).

Key words and phrases. Bucket recursive tree, Subtree size profile, Factorial moments, Limiting distribution.

by the number of its labels. They applied a probabilistic analysis for studying the height and depth of the largest label in these trees. A (probabilistic) description of random bucket recursive trees is given by a generalization of the stochastic growth rule for ordinary random recursive trees (which are the special instance $b = 1$), where a tree grows by progressive attraction of increasing integer labels: when inserting element $n+1$ into an existing bucket recursive tree containing n elements (i.e., containing the labels $\{1, 2, ..., n\}$) all n existing elements in the tree compete to attract the element $n + 1$, where all existing elements have equal chance to recruit the new element. If the element winning this competition is contained in a node with less than b elements (an unsaturated bucket or node), element $n+1$ is added to this node, otherwise if the winning element is contained in a node with already b elements (a saturated bucket or node), element $n + 1$ is attached to this node as a new bucket containing only the element $n + 1$. Starting with a single bucket as root node containing only element 1 leads after $n - 1$ insertion steps, where the labels $2, 3, ..., n$ are successively inserted according to this growth rule, to a so called random bucket recursive tree with n elements and maximal bucket-size b. In this paper we consider a model of bucket trees where the nodes are buckets with variable capacities labelled with integers $1, 2, \cdots, n$ (not the same capacities as bucket recursive trees).

Definition 1.1. [8] A size-n bucket recursive tree T_n with variable bucket capacities and maximal bucket size b starts with the root labelled 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 buckets with capacity $< b$ all labels in the tree (containing label 1) compete to attract the label $j + 1$. For the root node and buckets with capacity b, we always produce a new bucket $j + 1$. 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 bucket with capacity $c+1$. This process ends with inserting the label n (i.e., the largest label) in the tree.

Bucket recursive trees with variable capacities of buckets are appeared in chemistry, social science, in some computer science applications and furthermore. They are appeared as a model for the spread of epidemics, for pyramid schemes, for the family trees of preserved copies of ancient texts. In the family trees, suppose males with the same ethical traits come together in each generation. Suppose up to 3 people are matched with the same attributes. Then a bucket recursive trees with variable capacities of buckets with maximal bucket size 3 is formed. In this case, and in a genealogy of n people, the distance between two specific individuals is the quantity examined in this article. For another example, if n atoms in a branching molecular structure are stochastically labelled with integers 1, 2, ..., n, then atoms in different functional groups can be considered as the labels of different buckets of a bucket recursive tree (the size of the largest functional group is b).

In passing, we give the combinatorial description of our model. Let $d(v)$ be the out-degree of node v. It will be convenient to define for trees the size $|T|$ of a tree T via $|T| = \sum_v c(v)$. An increasing labelling of an ordered tree T is then a labelling of T, where the labels $\{1, 2, ..., |T|\}$ are distributed amongst the nodes of

T. Then a class $\mathcal{T}$ of a new family of *bucket-increasing trees* can be defined in the following way: A sequence of non-negative numbers $(\alpha_k)_{k\geq 0}$ with $\alpha_0 > 0$ and a sequence of non-negative numbers $\beta_1, \beta_2, ..., \beta_{b-1}$ is used to define the weight $w(T)$ of any ordered tree T by $w(T) := \Pi_v w(v)$, where v ranges over all nodes of T. It is natural that $w(v)$ must be dependent on $c(v)$ and $d(v)$. Thus the weight $w(v)$ of a node v is given as follows:

$$w(v) := \begin{cases} \alpha_{d(v)}, & v \text{ is root or complete } (c(v) = b) \\ \beta_{c(v)}, & v \text{ is incomplete } (c(v) < b). \end{cases} \tag{1.1}$$

The above definition is reasonable because the root is the only incomplete node that has outdegree ≥ 1. Thus for complete nodes and root, the weight is dependent on the out-degree and described by the sequence α_k, whereas for incomplete nodes except of root the weights are dependent on the capacities.

Furthermore, $\mathcal{L}(T)$ denotes the set of different increasing labelings of the tree T with distinct integers $\{1, 2, ..., |T|\}$, where $L(T) := |\mathcal{L}(T)|$ denotes its cardinality. Then the family $\mathcal{T}$ consists of all trees T together with their weights $w(T)$ and the set of increasing labelings $\mathcal{L}(T)$. For a given degree-weight sequence $(\alpha_k)_{k\geq 0}$ with a degree-weight generating function $\varphi(t) := \sum_{k\geq 0} \alpha_k t^k$ and a bucket-weight sequence $\beta_1, \beta_2, ..., \beta_{b-1}$, we define the exponential generating function

$$T_{r,k_1,...k_r}(z) \quad := \quad \sum_{n=1}^{\infty} T_{n,b,r,k_1,...k_r} \frac{z^n}{n!}, \tag{1.2}$$

where $T_{n,b,r,k_1,...k_r} := \sum_{|T|=n} w(T) \cdot L(T)$ is the total weights. For this model,

$$\begin{aligned} T_{n,b,r,k_1,...k_r} &= \frac{(n-1)!(b!)^{n(1-\sum_{i=1}^{r}|\mathcal{P}_{k_i}|)}}{b}, \ n \geq 1 \\ \varphi(T_{r,k_1,...k_r}(z)) &= \frac{(b-1)!}{1 - b!^{1-\sum_{i=1}^{r}|\mathcal{P}_{k_i}|} z}, \end{aligned} \tag{1.3}$$

where $\mathcal{P}_{k_i}$ is the set of all trees of size k_i and r is the degree of root node [8]. For simplicity, we set $T_{n,b} := T_{n,b,r,k_1,...,k_r}$ and $T(z) := T_{r,k_1,...,k_r}(z)$.

The motivation of studying the random variables in random trees is multifold. Here, 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.

2. Main Results

Theorem 2.1. *Let $S_k(z,u)$ be the moment generating function*

$$\begin{aligned} S_k(z,u) &= \sum_{n\geq 1}\sum_{m\geq 0} \mathbb{P}(S_{n,k} = m) T_{n,b} \frac{z^n}{n!} u^m \\ &= \sum_{n\geq 1}\sum_{m\geq 0} \mathbb{P}(S_{n,k} = m) \frac{(b!)^{n(1-\sum_{i=1}^{r}|\mathcal{P}_{k_i}|)}}{b} \frac{z^n}{n} u^m. \end{aligned}$$

Then

$$S_k(z,u) = \frac{\beta(r,b)(u-1)z^k}{k} + \log\left(\frac{1}{1-\beta(r,b)\int_0^z e^{\frac{\beta(r,b)(u-1)t^k}{k}}dt}\right),\ k \geq 1 \quad (2.1)$$

where $\beta(r,b) = b!^{-\Sigma_{i=1}^{r}|\mathcal{P}_{k_i}|}$.

Proof. According to the definition of the tree, the probabilities of $\mathbb{P}(S_{n,k} = m)$ satisfy ($n_i \geq 1$, $m_i \geq 0$ and $n > k$)

$$\begin{aligned} \mathbb{P}(S_{n,k} = m) &= \sum_{r\geq 1}\frac{1}{r!}\sum_{n_1+\cdots+n_r=n-1}\binom{n-1}{n_1,...,n_r}\frac{T^*_{n_1,b}\cdots T^*_{n_r,b}}{T_{n,b}} \\ &\times \sum_{m_1+\cdots+m_r=m}\mathbb{P}(S_{n_1,k} = m_1)\cdots\mathbb{P}(S_{n_r,k} = m_r), \end{aligned} \quad (2.2)$$

with initial values $\mathbb{P}(S_{k,k} = 1) = 1$, $\mathbb{P}(S_{n,k} = 0) = 1$ for $1 \leq n < k$ where $T^*_{n_i,b}$ is the total weights of the ith subtree similar to [4]. Thus recurrence (2.2) leads to the following functional equation

$$\frac{\partial}{\partial z}S_k(z,u) = b!^{-\sum_{i=1}^{r}|\mathcal{P}_{k_i}|}\left(e^{S_k(z,u)} + (u-1)z^{k-1}\right),\ (k \geq 1) \quad (2.3)$$

with initial condition $S_k(0,u) = 0$. The solution of partial differential equation (2.3) can be found as follows:
By using the substitution

$$W(z,u) = \beta(r,b)e^{S_k(z,u)}$$

in (2.3), we can see that $W(z,u)$ satisfies a Riccati equation of the form

$$\begin{aligned} \frac{\partial}{\partial z}W(z,u) &= \beta(r,b)\times\frac{\partial}{\partial z}S_k(z,u)\times e^{S_k(z,u)} \\ &= \beta(r,b)\times\beta(r,b)(e^{S_k(z,u)} + (u-1)z^{k-1})\times e^{S_k(z,u)} \\ &= (\beta(r,b)e^{S_k(z,u)})^2 + \beta(r,b)(u-1)z^{k-1}\beta(r,b)e^{S_k(z,u)} \\ &= (W(z,u)^2 + \beta(r,b)(u-1)z^{k-1}W(z,u). \end{aligned}$$

Let $W(z,u) = -\frac{\frac{\partial}{\partial z}Q(z,u)}{Q(z,u)}$, it follows that $Q(z,u)$ satisfies the linear 2nd order partial differential equation

$$\frac{\partial^2}{\partial z^2}Q(z,u) = \beta(r,b)(u-1)z^{k-1}\frac{\partial}{\partial z}Q(z,u), \quad (2.4)$$

since

$$\begin{aligned} \frac{\partial}{\partial z}W(z,u) &= -\frac{\partial}{\partial z}\left(\frac{\frac{\partial}{\partial z}Q(z,u)}{Q(z,u)}\right) \\ &= -\frac{\frac{\partial^2}{\partial z^2}Q(z,u)}{Q(z,u)} + \left(\frac{\frac{\partial}{\partial z}Q(z,u)}{Q(z,u)}\right)^2 \\ &= -\frac{\frac{\partial^2}{\partial z^2}Q(z,u)}{Q(z,u)} + (W(z,u))^2. \end{aligned}$$

From (2.4) we have

$$\frac{\partial}{\partial z}Q(z,u) = c_1 e^{\frac{\beta(r,b)(u-1)z^k}{k}},$$

and hence

$$Q(z,u) = c_1 \int_0^z e^{\frac{\beta(r,b)(u-1)t^k}{k}} dt + c_2,$$

then we have

$$\begin{aligned} W(z,u) &= -\frac{\frac{\partial}{\partial z}Q(z,u)}{Q(z,u)} \\ &= -\frac{c_1 e^{\frac{\beta(r,b)(u-1)z^k}{k}}}{c_1 \int_0^z e^{\frac{\beta(r,b)(u-1)t^k}{k}} dt + c_2}. \end{aligned}$$

Since $W(z,u) = \beta(r,b)e^{S_k(z,u)}$, we have $S_k(z,u) = \log(\frac{W(z,u)}{\beta(r,b)})$. Then

$$\begin{aligned} S_k(z,u) &= \log\left(-\frac{c_1 e^{\frac{\beta(r,b)(u-1)z^k}{k}}}{c_1\beta(r,b) \int_0^z e^{\frac{\beta(r,b)(u-1)t^k}{k}} dt + c_2\beta(r,b)}\right) \\ &= \frac{\beta(r,b)(u-1)z^k}{k} + \log\left(-\frac{c_1}{c_1\beta(r,b) \int_0^z e^{\frac{\beta(r,b)(u-1)t^k}{k}} dt + c_2\beta(r,b)}\right). \end{aligned}$$

From initial condition $S_k(0,u) = 0$ we have $c_2\beta(r,b) = -c_1$. Finally the function

$$S_k(z,u) = \frac{\beta(r,b)(u-1)z^k}{k} + \log\left(\frac{1}{1-\beta(r,b)\int_0^z e^{\frac{\beta(r,b)(u-1)t^k}{k}} dt}\right),$$

is the solution of partial differential equation (2.3) for any $k \geq 1$ which satisfies the given initial condition. $\square$

Set $I(A) = 1$ if A is true and $I(A) = 0$ otherwise. Assume

$$\beta(r,n,b) = b^{-1}(b!)^{n(1-\sum_{i=1}^r |\mathcal{P}_{k_i}|)}.$$

Theorem 2.2. *Let $\mu_{n,k,p} = \mathbb{E}(S^p_{n,k})$ for $p \geq 1$. Then $\mu_{n,k,1} = 0$ for $1 \leq n < k$ and*

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

Also, for $p \geq 2$,

$$\begin{aligned} \mu_{n,k,p} &= \sum_{\ell=1}^{p} s(p,\ell)n\beta(r,n,b)^{-1}\sum_{i=1}^{\ell}\frac{\beta(r,b)^{n+\ell-k\ell}}{i\ k^\ell}\binom{n-k\ell-1}{i-1}I(n \geq k\ell+1) \\ &\times \sum_{j_1+\cdots+j_i=\ell, j_q\geq 1}\frac{1}{\prod_{m=1}^{i}(kj_m+1)}\binom{\ell}{j_1,\ldots,j_i}, \end{aligned}$$

where $s(m,n)$ is the mth Stirling number of order n of the second kind.

Proof. Set $y=u-1$. Thus

$$\begin{aligned} S_k(z,1+y) &= \sum_{n\geq 1}\sum_{m\geq 0}\mathbb{P}(S_{n,k}=m)\beta(r,n,b)\frac{z^n}{n}(1+y)^m \\ &= \sum_{n\geq 1}\sum_{m\geq 0}\sum_{p=0}^{m} m^{\underline{p}}\mathbb{P}(S_{n,k}=m)\beta(r,n,b)\frac{z^n}{n}\frac{y^p}{p!}, \end{aligned}$$

where $m^{\underline{p}}=m(m-1)\cdots(m-p+1)$. Hence

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

where $[z^n]f(z)$ denote the operation of extracting the coefficient of z^n in the formal power series $f(z)=\sum f_nz^n$. Suppose

$$\alpha:=\sum_{j\geq 1}\frac{y^j(\frac{z^k}{k}\beta(r,b))^j}{(kj+1)j!}.$$

Then

$$\begin{aligned} 1-\beta(r,b)\int_0^z \exp\Big(\frac{\beta(r,b)yt^k}{k}\Big)dt &= 1-\beta(r,b)\int_0^z\sum_{j\geq 0}\frac{(\frac{\beta(r,b)yt^k}{k})^j}{j!}dt \\ &= 1-\beta(r,b)z(1-\alpha). \end{aligned}$$

and

$$\log\left(\frac{1}{1-\beta(r,b)\int_0^z\exp\Big(\frac{\beta(r,b)yt^k}{k}\Big)dt}\right)=\log\frac{1}{1-\beta(r,b)z}+\log\left(\frac{1}{1-\frac{\beta(r,b)z}{1-\beta(r,b)z}\alpha}\right).$$

We have $\log\frac{1}{1-z}=\sum_{i\geq 1}\frac{z^i}{i}$. Thus, for $p\geq 1$,

$$\begin{aligned} &[y^p]\log\left(\frac{1}{1-\frac{\beta(r,b)z}{1-\beta(r,b)z}\alpha}\right) \\ =\ &[y^p]\sum_{i\geq 1}\frac{1}{i}\left(\frac{\beta(r,b)z}{1-\beta(r,b)z}\right)^i\left(\sum_{j\geq 1}\frac{y^j(\frac{z^k}{k}\beta(r,b))^j}{(kj+1)j!}\right)^i \\ =\ &\frac{1}{p!}\sum_{i=1}^{p}\frac{(\frac{z^k}{k}\beta(r,b))^p}{i}\left(\frac{\beta(r,b)z}{1-\beta(r,b)z}\right)^i\sum_{j_1+\cdots+j_i=p,j_q\geq 1}\frac{1}{\prod_{m=1}^{i}(kj_m+1)}\binom{p}{j_1,\ldots,j_i}. \end{aligned}$$

From (2.1),

$$\begin{aligned} [y^p]S_k(z,1+y) &= \frac{1}{p!}\sum_{i=1}^{p}\frac{(\frac{z^k}{k}\beta(r,b))^p}{i}\left(\frac{\beta(r,b)z}{1-\beta(r,b)z}\right)^i \\ &\times \sum_{j_1+\cdots+j_i=p,j_q\geq 1}\frac{1}{\prod_{m=1}^{i}(kj_m+1)}\binom{p}{j_1,\ldots,j_i} \\ &+ \frac{z^k}{k}\beta(r,b)I(p=1). \end{aligned} \tag{2.5}$$

The formula (2.5) immediately gives

$$\begin{aligned}
\mu_{n,k,1} &= n\beta(r,n,b)^{-1}[z^n]\Big(\frac{z^k}{k}\frac{\beta(r,b)}{k+1}\frac{\beta(r,b)z}{1-\beta(r,b)z}+\frac{z^k}{k}\beta(r,b)\Big)\\
&= n\beta(r,n,b)^{-1}\left(\frac{\beta(r,b)^{n-k+1}}{k(k+1)}I(n\geq k+1)+\frac{\beta(r,b)}{k}I(n=k)\right)\\
&= \frac{n}{k(k+1)}\frac{\beta(r,b)^{n-k+1}}{\beta(r,n,b)}I(n\geq k+1)+\frac{\beta(r,b)}{\beta(r,n,b)}I(n=k).
\end{aligned}$$

We have

$$[z^n](1-z)^m=\binom{n+m-1}{n}$$

and $[z^n]f(qz)=q^n[z^n]f(z)$. Then for $p\geq 2$,

$$\begin{aligned}
\mathbb{E}(S^{\ell}_{n,k}) &= \beta(r,n,b)^{-1}n\ell![z^n y^\ell]S_k(z,1+y)\\
&= n\beta(r,n,b)^{-1}[z^n]\sum_{i=1}^{\ell}\frac{\beta(r,b)^{\ell+i}}{i\;k^\ell}\frac{z^{k\ell+i}}{(1-\beta(r,b)z)^i}\\
\times & \sum_{j_1+\cdots+j_i=\ell,j_q\geq 1}\frac{1}{\prod_{m=1}^{i}(kj_m+1)}\binom{\ell}{j_1,...,j_i}\\
&= n\beta(r,n,b)^{-1}\sum_{i=1}^{\ell}\frac{\beta(r,b)^{n+\ell-k\ell}}{i\;k^\ell}\binom{n-k\ell-1}{i-1}I(n\geq k\ell+1)\\
\times & \sum_{j_1+\cdots+j_i=\ell,j_q\geq 1}\frac{1}{\prod_{m=1}^{i}(kj_m+1)}\binom{\ell}{j_1,...,j_i}.
\end{aligned}$$

The proof is completed since

$$\mu_{n,k,p}=\mathbb{E}(S^{p}_{n,k})=\sum_{\ell=1}^{p}s(p,\ell)\mathbb{E}(S^{\ell}_{n,k}).$$

□

We use the notation $\xrightarrow{D}$ to denote convergence in distribution. The standard random variable Poi(λ) appear for the Poisson distributed with parameter $\lambda>0$. These random variables appear in the results as limiting random variables.

Theorem 2.3. *Let $S_{n,k}$ be the subtree size-k profile in size-n bucket recursive trees with variable capacities of buckets. Then*

$$S_{n,k}\xrightarrow{D}0,\ \text{as } k\sqrt{\frac{\beta(r,n,b)}{n\beta(r,b)^{n-k+1}}}\to\infty.$$

Proof. Suppose $\frac{n\beta(r,b)^{n-k+1}}{k^2\beta(r,n,b)}\to\lambda$ for some $\lambda>0$. Thus from Theorem 2.2, $\mu_{n,k,1}\to\lambda$. It is obvious that

$$\sum_{j_1+\cdots+j_i=p,j_q\geq 1}\binom{p}{j_1,...,j_i}=p![z^p](e^z-1)^i\leq p![z^p]e^{iz}=i^p. \tag{2.6}$$

Now, we consider a fixed $p \geq 2$. Then

$$\begin{aligned}
\mathbb{E}(S_{n,k}^{\underline{p}}) &= n\beta(r,n,b)^{-1}\sum_{i=1}^{p}\frac{\beta(r,b)^{n-kp+p}}{ik^p}\binom{n-kp-1}{i-1} \\
&\times \sum_{j_1+\cdots+j_i=p, j_q\geq 1}\frac{1}{\prod_{k=1}^{i}(j_k+1)}\binom{p}{j_1,\ldots,j_i} \\
&\leq \frac{n\beta(r,b)^{n-kp+p}}{k^p\beta(r,n,b)}\sum_{i=1}^{p}\frac{\beta(r,b)^{n-kp+p}}{k^i\beta(r,n,b)}\frac{n^i}{(i-1)!}i^{p-1} \\
&\leq p^{p-1}\frac{n\beta(r,b)^{n-k+1}}{k^2\beta(r,n,b)}\sum_{i=0}^{\infty}\frac{\left(\frac{n\beta(r,b)^{n-k+1}}{k^2\beta(r,n,b)}\right)^i}{i!} \\
&= p^{p-1}\frac{n\beta(r,b)^{n-k+1}}{k^2\beta(r,n,b)}\exp\left(\frac{n\beta(r,b)^{n-k+1}}{k^2\beta(r,n,b)}\right),
\end{aligned}$$

since $p \geq i \geq 1$. By assumption $\frac{n\beta(r,b)^{n-k+1}}{k^2\beta(r,n,b)} \to 0$. Then for all $p \geq 1$, $\mathbb{E}(S_{n,k}^{\underline{p}}) \to 0$. i.e., the random variable $S_{n,k}$ convergent to a degenerate distribution at point 0. $\square$

Theorem 2.4. *Let $S_{n,k}$ be the subtree size-k profile in size-n bucket recursive trees with variable capacities of buckets. Then*

$$S_{n,k} \xrightarrow{D} \text{Poi}\left(\frac{1}{c^2}\right), \text{ as } k\sqrt{\frac{\beta(r,n,b)}{n\beta(r,b)^{n-k+1}}} \to c > 0,$$

and otherwise no limiting distribution exists for $S_{n,k}$.

Proof. Suppose $\frac{n\beta(r,b)^{n-k+1}}{k^2\beta(r,n,b)} \to \lambda$ for some $\lambda > 0$. From Theorem 2.2, $\mathbb{E}(S_{n,k}^{\underline{p}}) = A + B$, where

$$\begin{aligned}
A &= \sum_{i=1}^{p-1} n\beta(r,n,b)^{-1}\frac{\beta(r,b)^{n+p-kp}}{i\ k^p}\binom{n-kp-1}{i-1}I(n\geq kp+1) \\
&\times \sum_{j_1+\cdots+j_i=p, j_q\geq 1}\frac{1}{\prod_{m=1}^{i}(kj_m+1)}\binom{p}{j_1,\ldots,j_i} \\
&\leq \sum_{i=1}^{p-1}\frac{n^i i^k}{i!k^p k^i}\frac{\beta(r,b)^{n-k+1}}{\beta(r,n,b)} \\
&\leq \frac{p^{p-1}}{k}\sum_{i=1}^{p-1}\frac{1}{k^{p-i-1}}\left(\frac{n\beta(r,b)^{n-k+1}}{k^2\beta(r,n,b)}\right)^i\frac{1}{(i-1)!} \\
&\leq \frac{p^{p-1}}{k}\frac{n\beta(r,b)^{n-k+1}}{k^2\beta(r,n,b)}\sum_{i\geq 1}\frac{\left(\frac{n\beta(r,b)^{n-k+1}}{k^2\beta(r,n,b)}\right)^i}{i!} \\
&= \mathcal{O}\left(\sqrt{\frac{\beta(r,n,b)}{n\beta(r,b)^{n-k+1}}}\right)
\end{aligned}$$

and

$$\begin{aligned} B &= \frac{n\beta(r,b)^{n+p-kp}}{pk^p\beta(r,n,b)}\binom{n-kp-1}{p-1}\frac{p!}{(k+1)^p} \\ &= \Big(\frac{n\beta(r,b)^{n-k+1}}{k^2\beta(r,n,b)}\Big)^p\left(1+\mathcal{O}\left(\sqrt{\frac{\beta(r,n,b)}{n\beta(r,b)^{n-k+1}}}\right)\right), \end{aligned}$$

for large n. Thus for every $p \geq 1$,

$$\mathbb{E}(S_{n,k}^{\underline{p}}) = \Big(\frac{n\beta(r,b)^{n-k+1}}{k^2\beta(r,n,b)}\Big)^p + \mathcal{O}\left(\sqrt{\frac{\beta(r,n,b)}{n\beta(r,b)^{n-k+1}}}\right) \to \lambda^p.$$

Now, if we use the substitution $c = \sqrt{\frac{1}{\lambda}}$, then $\sqrt{\frac{\beta(r,n,b)}{n\beta(r,b)^{n-k+1}}} \to c$ and proof is completed. $\square$

Corollary 2.5. *If $b = 1$, then all results reduce to the random recursive trees. We have*

$$\beta(r,1) = 1!^{-\Sigma_{i=1}^r|\mathcal{P}_{k_i}|} = 1, \ \beta(r,n,1) = 1^{-1}(1!)^{n(1-\sum_{i=1}^r|\mathcal{P}_{k_i}|)} = 1.$$

Thus

$$\mu_{n,k,1} = \frac{n}{k(k+1)}I(n \geq k+1) + I(n = k).$$

If $k = 1$, then $\mu_{n,1,1}$ is the number of leaves [12], *i.e.,*

$$\mu_{n,k,1} = \frac{n}{2}, \ \mathbb{V}ar(S_{n,1}) = \frac{n}{12}, \ n \geq 2.$$

References

1. Chang, J. and Fuchs, M.: Limit theorems for pattern in phylogenetic trees, *Journal of Mathematical Biology*, **60** (2010) 481–512.
2. Dennert, F. and Grübel, R.: On the subtree size profile of binary search trees, *Combinatorics, Probability and Computing*, **19** (2010) 561–578.
3. Devroye, L.: On the richness of the collection of subtrees in random binary search trees, *Information Processing Letters,* **65** (1998) 195–199.
4. Feng, Q., Mahmoud, H. and Panholzer, A.: Phase changes in subtree varieties in random recursive trees and binary search trees, *SIAM Journal on Discrete Mathematics*, **22** (2008) 160–184.
5. Feng, Q., Mahmoud, H. and Su, C.: *On the variety of subtrees in a random recursive tree*, Technical report the George Washington University, Washington, DC., 2007.
6. Feng, Q., Miao, B. and Su, C.: On the subtrees of binary search trees, *Chinese Journal of Applied Probability and Statistics*, **22** (2006) 304–310.
7. Flajolet, P., Gourdon, X. and Martinez, C.: Patterns in random binary search trees, *Random Structures and Algorithms*, **11** (1997) 223–244.
8. Kazemi, R.: Depth in bucket recursive trees with variable capacities of buckets, *Acta Mathematics Sinica, English Series*, **30(2)** (2014) 305–310.
9. Kazemi, R.: The subtree size profile of bucket recursive trees, *Iranian Journal of Mathematical Sciences and Informatics*, **11(1)** (2016) 1–11.
10. Mahmoud H., Smythe R.: Probabilistic analysis of bucket recursive trees. *Theor. Comput. Sci.* **144** (1995) 221–249.
11. Meir. A. and Moon, J. W.: Recursive trees with no nodes of out-degree one, *Congressus Numerantium*, **66** (1988) 49–62.

12. Zhang, Y.: On the number of leaves in a random recursive tree, *Braz. J. Probab. Stat.,* **29(4)** (2015) 897–908.

Ramin Kazemi: Department of Statistics, Imam Khomeini International University, Qazvin, Iran

E-mail address: r.kazemi@sci.ikiu.ac.ir

URL: http://ikiu.ac.ir/members/?id=231&lang=0