# BRANCHES IN BUCKET RECURSIVE TREES WITH VARIABLE CAPACITIES OF BUCKETS

Ramin KAZEMI[1]

*Bucket recursive trees with variable capacities of buckets (BRT-VCB) introduced by Kazemi (2012). In this paper, we study the random variable which counts the number of branches of size a attached to the bucket containing label j in a BRT-VCB of size n (the number of subtrees of size a rooted at the children of bucket containing label j).*

**Keywords:** BRT-VCB, branches, limiting distribution, joint distribution.
**MSC2010:** 05C05, 60F05

## 1. Introduction

Trees are defined as connected graphs without cycles, and their properties are basics of graph theory. A tree on $n$ nodes labeled 1, 2,..., $n$ is a recursive tree if the node labeled 1 is distinguished as the root, and for each $2 \leq k \leq n$, the labels of the nodes in the unique path from the root to the node labeled $k$ form an increasing sequence [6]. Bucket trees are a generalization of the ordinary trees where buckets (or nodes) can hold up to $b \geq 1$ labels. Mahmoud and Smythe [5] introduced bucket recursive trees as a generalization of ordinary recursive trees. In this model the capacity of buckets is fixed. They applied a probabilistic analysis for studying the height and depth of the largest label in these trees and Kuba and Panholzer [4] analyzed these trees as a special instance of bucket increasing trees which is a family of some combinatorial objects. Kazemi [2] introduced a new version of bucket recursive trees where the nodes are buckets with variable capacities labelled with integers $1, 2, ..., n$. In fact, the capacity of buckets is a random variable in these models. He studied the depth quantity and the first Zagreb index in these models [3]. *A bucket recursive tree with variable capacities of buckets (BRT-VCB)* starts with the root labelled by 1 that has $r \geq 0$ descendants each of them making a subtree. The nodes in the subtrees have capacities $c < b$ or $c = b$. The nodes with capacities $c < b$ are connected together with 1 edge and the nodes with capacities $b$ have descendants $\geq 0$ again each of them making a subtree such that the labels within these nodes are arranged in increasing order. The tree is completed when the label $n$ is inserted in the tree. Figure 1 illustrates such a tree of size 19 with $b = 3$. For constructing a tree of size $n+1$ (attracting label $n+1$ to a tree of size $n$), if a leaf $v$ has the capacity $c < b$, then we add the label $n+1$ to this node and make a node with capacity $c+1$ or produce a node $n+1$. But for a node with capacity $b$, we only produce a new node $n+1$. The last nodes with $c \leq b$ labels at the end of subtrees are called *leaves* and other nodes are called *non-leaves*. The probability $p$, which gives the probability that label $n+1$ is attracted by node $v$ in the model is $\frac{c(v)}{n-|\gamma|}$, where

$$\gamma = \{v \in T;\ c = c(v) = k < b,\ \text{and } v \text{ is a non-leaf}\}.$$

This model can be considered as a generalization of random recursive trees [1]. A sequence

[1]Assistant Professor, Department of Statistics, Imam Khomeini International University, Qazvin, Iran, e-mail: kazemi@ikiu.ac.ir

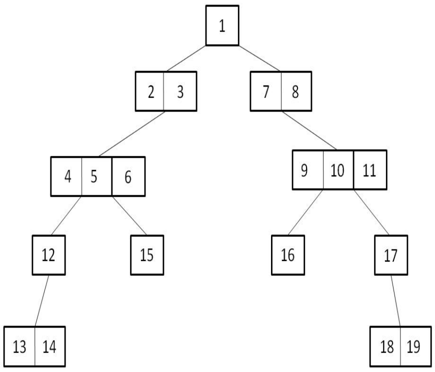


FIGURE 1. A BRT-VCB of size 19 with $b = 3$.

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 $T$. The weight $w(v)$ of a bucket $v$ is given as follows:

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

Let $\mathcal{L}(T)$ 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 [1]

$$T_{n,b}(z) \quad := \quad \sum_{n=1}^{\infty} T_{n,b} \frac{z^n}{n!},$$

where $T_{n,b} := \sum_{|T|=n} w(T) \cdot L(T)$ is the total weights. Kazemi [2] showed

$$\begin{aligned} T_{n,b} &= \frac{(n-1)!(b!)^{n(1-\sum_{i=1}^{r} |\mathcal{P}_{n_i}|)}}{b}, \ n \geq 1, \\ T_{n,b}(z) &= -\frac{1}{b} \log\left(1 - b!^{1-\sum_{i=1}^{r} |\mathcal{P}_{n_i}|} z\right), \\ \varphi(T_{n,b}(z)) &= \frac{(b-1)!}{1 - b!^{1-\sum_{i=1}^{r} |\mathcal{P}_{n_i}|} z}, \end{aligned} \tag{1}$$

where $|\mathcal{P}_{n_i}|$ denotes the size of the set of all trees of size $n_i (i = 1, 2, ..., r)$. In the equation (1), if $i$-th subtree starts with a bucket with capacity $c = 1$, then we set $|\mathcal{P}_{n_i}| = 0$.

The motivation for studying the bucket recursive trees with variable capacities of buckets is multifold. For example, if $n$ atoms in a branching molecular structure (such as dendrimer) 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.

## 2. Number of Branches of Size $a$

Let $S_{n,j,a}$ counts the number of branches of size $a$ attached to the bucket containing label $j$ in our model of size $n$ (or in other words, the number of subtrees of size $a$ rooted at the children of node containing label $j$, in a random grown tree of size $n$). We use a combinatorial approach to find the differential equation corresponds to the random variable $S_{n,j,a}$ and give a closed formula for the probability distribution and the factorial moments of $S_{n,j,a}$. Furthermore limiting distribution result of $S_{n,j,a}$ is given, not only for $j$ fixed, but a full characterization dependent on the growth $j = j(n)$ compared to $n$ is presented. Moreover, the joint distribution of $S_{n,1,1}$, $S_{n,1,2}$, ..., $S_{n,1,n-j}$ is computed for all grown trees.

Let $S_{n,1,a}$ be the number of branches of size $a$ attached to the root node (label 1) in our model of size $n$. For encoding the behavior of this quantity we introduce the bivariate generating function

$$M(z,v) = \sum_{n\geq 1}\sum_{m\geq 0} \mathbb{P}(S_{n,1,a} = m) T_{n,b} \frac{z^n}{n!} v^m.$$

By definition of the model one gets the following explicit result for the probabilities $\mathbb{P}(S_{n,1,a} = m)$:

$$\mathbb{P}(S_{n,1,a} = m) \quad = \quad \sum_{r\geq m} \alpha_r \binom{r}{m} \sum_{n_1+\cdots+n_r=n-1} \frac{T^*_{n_1,b}\cdots T^*_{n_r,b}}{T_{n,b}} \binom{n-1}{n_1,...,n_r}, \qquad (2)$$

where for $1 \leq i \leq m$, $n_i = a$ and for $m+1 \leq j \leq r$, $n_j \neq a$. Also $T^*_{n_i,b}$ is the total weights of the $i$th subtree. Since $T^*_{n_1,b}\cdots T^*_{n_r,b} = b!^{-\sum_{i=1}^r |\mathcal{P}_{n_i}|} T_{n_1,b}\cdots T_{n_r,b}$, (c.f [2]) by multiplying with $T_{n,b} z^{n-1} v^m/(n-1)!$ and summing up over $n \geq 1, m \geq 0$, the equation (2) yields to an explicit formula for $\frac{\partial}{\partial z} M(z,v)$, which is given below:

$$\begin{aligned}\frac{\partial}{\partial z} M(z,v) \quad &= \quad b!^{-\sum_{i=1}^r |\mathcal{P}_{n_i}|} \left\{\varphi(T_{n,b}(z) + \frac{T_{a,b}}{a!} z^a (v-1))\right\} \\ &= \quad \frac{(b-1)!}{1 - b!^{1-\sum_{i=1}^r |\mathcal{P}_{n_i}|} z} \exp\left\{\frac{bT_{a,b}}{a!} z^a (v-1)\right\}. \qquad (3)\end{aligned}$$

For describing the behavior of arbitrary label $j > 1$ we introduce the trivariate generating function

$$N(z,u,v) = \sum_{k\geq 0}\sum_{j\geq 1}\sum_{m\geq 0} \mathbb{P}(S_{k+j,j,a} = m) T_{k+j,b} \frac{z^{j-1}}{(j-1)!} \frac{u^k}{k!} v^m.$$

For our model of size $n$ with root-degree $r$ and subtrees with sizes $n_1, ..., n_r$, enumerated from left to right, where the bucket containing label $j$ lies in the leftmost subtree and is the $i$-th bucket in this subtree, we can reduce the computation of the probabilities $\mathbb{P}(S_{n,j,a} = m)$ to the probabilities $\mathbb{P}(S_{n_1,i,a} = m)$, when the parameter does only depend on the subtree of bucket containing label $j$. We get as factor the total weight of the $r$ subtrees and the root node $\alpha_r b!^{-\sum_{i=1}^r |\mathcal{P}_{n_i}|} T_{n_1,b}\cdots T_{n_r,b}$, divided by the total weight $T_{n,b}$ of trees of size $n$ and multiplied by the number of order preserving relabellings of the $r$ subtrees, which are given here by

$$\binom{j-2}{i-1}\binom{n-j}{n_1-i}\binom{n-1-n_1}{n_2,n_3,...,n_r}.$$

Due to symmetry arguments we obtain a factor $r$, if the bucket containing label $j$ is the $i$-th bucket in the second, third, ..., $r$-th subtree. Summing up over all choices for the rank $i$ of bucket containing label $j$ in its subtree, the subtree sizes $n_1, ..., n_r$, and the degree $r$ of the root node gives for $n \geq j \geq 2$ the following recurrence:

$$\begin{aligned}\mathbb{P}(S_{n,j,a} = m) &= \sum_{r\geq 1} r\alpha_r \sum_{n_1+\cdots+n_r=n-1} \frac{T^*_{n_1,b}\cdots T^*_{n_r,b}}{T_{n,b}} \\ \times \quad &\sum_{i=1}^{\min\{n_1,j-1\}} \mathbb{P}(S_{n_1,i,a} = m)\binom{j-2}{i-1}\binom{n-j}{n_1-i}\binom{n-1-n_1}{n_2,n_3,...,n_r}. \qquad (4)\end{aligned}$$

With the same method of [2],

$$\frac{\partial}{\partial z}N(z,u,v) = b!^{-\sum_{i=1}^{r}|\mathcal{P}_{n_i}|}\varphi'(T_{n,b}(z+u))N(z,u,v)$$

with the initial condition

$$N(0,u,v) = \sum_{k\geq 0}\sum_{m\geq 0}\mathbb{P}(S_{k+1,1,a}=m)T_{k+1,b}\frac{u^k}{k!}v^m = \frac{\partial}{\partial u}M(u,v).$$

Thus

$$N(z,u,v) = b!^{-\sum_{i=1}^{r}|\mathcal{P}_{n_i}|}\frac{\varphi(T_{n,b}(z+u))\frac{\partial}{\partial u}M(u,v)}{\varphi(T_{n,b}(u))}.$$

## 3. The Main Results

In the following theorem we show that the marginal probabilities $\mathbb{P}(S_{n,j,a}=m)$ are independent of $b$.

**Theorem 3.1.** *The probability that there are $m$ branches of size $a$ attached to bucket containing label $j$ is given as follows:*

$$\mathbb{P}(S_{n,j,a}=m) = \frac{1}{a^m a!\binom{n-1}{j-1}}\sum_{\ell=0}^{\lfloor\frac{n-j-am}{a}\rfloor}\frac{(-1)^\ell}{a^\ell \ell!}\binom{n-1-a(m+\ell)}{j-1}. \tag{5}$$

*Proof.* Let $[z^n]f(z)$ denote the operation of extracting the coefficient of $z^n$ in the formal power series $f(z)=\sum f_n z^n$. Thus

$$\begin{aligned}
\mathbb{P}(S_{n,j,a}=m) &= \frac{(j-1)!(n-j)!}{T_{n,b}}[z^{j-1}u^{n-j}v^m]N(z,u,v)\\
&= \frac{bb!^{-\sum_{i=1}^{r}|\mathcal{P}_{n_i}|}}{b!^{n(1-\sum_{i=1}^{r}|\mathcal{P}_{n_i}|)}\binom{n-1}{j-1}}[z^{j-1}u^{n-j}v^m]\frac{\frac{\partial}{\partial u}M(u,v)}{1-\frac{b!^{1-\sum_{i=1}^{r}|\mathcal{P}_{n_i}|}}{1-b!^{1-\sum_{i=1}^{r}|\mathcal{P}_{n_i}|}u}z}\\
&= b!^{-\sum_{i=1}^{r}|\mathcal{P}_{n_i}|}\frac{b!^{1+(j-1)(1-\sum_{i=1}^{r}|\mathcal{P}_{n_i}|)}}{b!^{n(1-\sum_{i=1}^{r}|\mathcal{P}_{n_i}|)}\binom{n-1}{j-1}}[u^{n-j}v^m]\frac{e^{v\frac{bT_{a,b}}{a!}u^a}e^{-\frac{bT_{a,b}}{a!}u^a}}{(1-b!^{1-\sum_{i=1}^{r}|\mathcal{P}_{n_i}|}u)^j}\\
&= \frac{b!^{-(n-j-am)(1-\sum_{i=1}^{r}|\mathcal{P}_{n_i}|)}}{a^m m!\binom{n-1}{j-1}}[u^{n-j-am}]\frac{e^{-\frac{bT_{a,b}}{a!}u^a}}{(1-b!^{1-\sum_{i=1}^{r}|\mathcal{P}_{n_i}|}u)^j}\\
&= \frac{1}{a^m a!\binom{n-1}{j-1}}\sum_{\ell=0}^{\lfloor\frac{n-j-am}{a}\rfloor}\frac{(-1)^\ell}{a^\ell \ell!}\binom{n-1-a(m+\ell)}{j-1},
\end{aligned}$$

since$[z^n]f(qz)=q^n[z^n]f(z)$. □

**Theorem 3.2.** *Let $m^{\underline{s}}=m(m-1)\cdots(m-s+1)$. The factorial moments of the random variable $S_{n,j,a}$ in our model of size $n$ are given as follows:*

$$\mathbb{E}(S_{n,j,a}^{\underline{s}}) = \sum_{m\geq 0}m^{\underline{s}}\mathbb{P}(S_{n,j,a}=m) = \frac{1}{a^s}\frac{\binom{n-as-1}{j-1}}{\binom{n-1}{j-1}}. \tag{6}$$

*Proof.* Let $D_x$ be the differential operator with respect to $x$, and $E_x$ be the evaluation operator at $x=1$. Thus

$$\begin{aligned}
E_v D_v^s\frac{\partial}{\partial z}M(z,v) &= E_v D_v^s\frac{(b-1)!}{1-b!^{1-\sum_{i=1}^{r}|\mathcal{P}_{n_i}|}z}\exp\left\{\frac{bT_{a,b}}{a!}z^a(v-1)\right\}\\
&= \frac{(b-1)!}{1-b!^{1-\sum_{i=1}^{r}|\mathcal{P}_{n_i}|}z}\left(\frac{bT_{a,b}z^a}{a!}\right)^s.
\end{aligned}$$

Now

$$
\begin{aligned}
\mathbb{E}(S^s_{n,j,a}) &= \frac{(j-1)!(n-j)!}{T_{n,b}}[z^{j-1}u^{n-j}]E_v D_v^s N(z,u,v) \\
&= b!^{-\sum_{i=1}^{r}|\mathcal{P}_{n_i}|}\frac{bb!^{(j-1)(1-\sum_{i=1}^{r}|\mathcal{P}_{n_i}|)}}{b!^{n(1-\sum_{i=1}^{r}|\mathcal{P}_{n_i}|)}\binom{n-1}{j-1}}[u^{n-j}]\frac{\frac{(b-1)!\frac{1}{a^s}b!^{as(1-\sum_{i=1}^{r}|\mathcal{P}_{n_i}|)}u^{as}}{1-b!^{1-\sum_{i=1}^{r}|\mathcal{P}_{n_i}|}u}}{(1-b!^{1-\sum_{i=1}^{r}|\mathcal{P}_{n_i}|}u)^{j-1}} \\
&= \frac{b!^{1+(j-1-n+as)(1-\sum_{i=1}^{r}|\mathcal{P}_{n_i}|)+\sum_{i=1}^{r}|\mathcal{P}_{n_i}|}}{a^s\binom{n-1}{j-1}}[u^{n-j-as}]\frac{1}{(1-b!^{1-\sum_{i=1}^{r}|\mathcal{P}_{n_i}|}u)^j} \\
&= \frac{1}{a^s}\frac{\binom{n-as-1}{j-1}}{\binom{n-1}{j-1}}.
\end{aligned}
$$

□

We use notation $\xrightarrow{d}$ to denote convergence in distribution. The standard random variable Poi$(\lambda)$ appears in the following theorem for the Poisson distributed with parameter $\lambda > 0$.

**Theorem 3.3.** *A) For $n \to \infty, j = o(n)$ and a fixed: $S_{n,j,a} \xrightarrow{d} S_a \sim \text{Poi}\left(\frac{1}{a}\right)$.*

*B) For $n \to \infty, j = \rho n$ with $0 < \rho < 1$ and a fixed: $S_{n,j,a} \xrightarrow{d} S_{\rho,a} \sim \text{Poi}\left(\frac{1-\rho}{a}\right)$.*

*C) For $n \to \infty, n-j = o(n)$ and a fixed: $S_{n,j,a} \xrightarrow{d} S_a \sim \mathbb{P}(S_a = 0) = 1$.*

*Proof.* The proof is quite similar to increasing trees [1]. □

**Theorem 3.4.** *The joint distribution of $S_{n,1,1}, S_{n,1,2}, ..., S_{n,1,n-1}$ is given as follows:*

$$
\mathbb{P}(S_{n,1,1} = m_1, ..., S_{n,1,n-1} = m_{n-1}) = b\frac{\alpha_{\sum_{i=1}^{n-1} m_i}\left(\sum_{i=1}^{n-1} m_i\right)!}{b!^{n(1-\sum_{i=1}^{r}|\mathcal{P}_{n_i}|)}}\prod_{i=1}^{n-1}\frac{b!^{m_i(i(1-\sum_{i=1}^{r}|\mathcal{P}_{n_i}|))}}{(bi)^{m_i}m_i!}, \tag{7}
$$

*for all sequences of non-negative integers satisfying $\sum_{i=1}^{n-1} im_i = n-1$.*

*Proof.* We have the factor $\binom{n-1}{\underbrace{1,...,1}_{m_1},\underbrace{2,...,2}_{m_2},...,m_{n-1}}$ to the choices for the labels, the factor $\alpha_{\sum_{i=1}^{n-1} m_i}$ corresponds to the root degree and the factor $\binom{\sum_{i=1}^{n-1} m_i}{m_1,...,m_{n-1}}$ to the different positions of the subtrees. Thus

$$
\begin{aligned}
&T_{n,b}\mathbb{P}(S_{n,1,1} = m_1, ..., S_{n,1,n-1} = m_{n-1}) \\
&= \alpha_{\sum_{i=1}^{n-1} m_i}\binom{n-1}{1,...,1,2,...,2,...,m_{n-1}}\binom{\sum_{i=1}^{n-1} m_i}{m_1,...,m_{n-1}}\prod_{i=1}^{n-1}T_{i,b}^{m_i}.
\end{aligned}
$$

Since the total weights of BRT-VCB with $n$ vertices is [2]:

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

proof is completed. □

## 4. **Conclusion**

In this paper we studied the branches of size $a$ attached to the bucket containing label $j$ for investigating the effect of bucketing on random recursive trees. All results obtained for bucket recursive trees introduced by Mahmoud and Smythe are independent of $b$ since for these models $T_{n,b} = (n-1)!$ [4]. For bucket recursive trees with variable capacities of buckets, although $T_{n,b} = b^{-1}(n-1)!(b!)^{n(1-\sum_{i=1}^{r}|\mathcal{P}_{n_i}|)}$, but only the joint distribution of $S_{n,1,1}$, $S_{n,1,2}$, ..., $S_{n,1,n-1}$ is dependent on $b$.

## REFERENCES


[1] *F. Bergeron, P. Flajolet and B. Salvy*, Varieties of increasing trees, Lecture Notes in Computer Science, **581** (1992), 24-48.

[2] *R. Kazemi,* Depth in bucket recursive trees with variable capacities of buckets, Published online in Acta Mathematics Sinica, English Series, DOI: 10.1007/s10114-013-2115-2, 2013.

[3] *R. Kazemi,* Probabilistic analysis of the first Zagreb index. Trans. Comb., **2(2)** (2013), 35-40.

[4] *M. Kuba and A. Panholzer,* Combinatorial approach to the analysis of bucket recursive trees, *Theor. Comput. Sci.*, **411** (2010), Issues 34-36, 3253-3273.

[5] *H. Mahmoud and R. Smythe,* Probabilistic analysis of bucket recursive trees, *Theor. Comput. Sci.*, **144** (1995), 221-249.

[6] *A. Meir, and J. W. Moon,* On the altitude of nodes in random trees. Canadian J. Math, **80** (1978), 997-1015.