Bulletin of the Iranian Mathematical Society Vol. 38 *No.* 2 (*2012*), *pp 323-331.*

BRANCHES IN RANDOM RECURSIVE k-ARY TREES

M. JAVANIAN AND M. Q. VAHIDI-ASL*

Communicated by Ahmad Reza Soltani

Abstract. Using the generalized Pólya urn models, we find the expected value of the size of a branch in recursive k-ary trees. We also find the expectation of the number of nodes of a given outdegree in a branch of such trees.

1. **Introduction**

A *tree* is a connected graph which has no cycles (see [3] for basic properties).

A tree with n nodes labelled $1, 2, \ldots, n$ is a *recursive tree* if for each i such that $2 \leq i \leq n$, the labels of the nodes in the unique path from the *root* (the node with label 1) to the ith node, form an increasing sequence (see the survey by Smythe and Mahmoud [6] and also Bergeron, et al. [1] for a wide class of results). Figure 1 shows all the recursive trees of order 4.

The nodes with no descendants are the *leaves* of the tree. The number of edges incident to a node of the tree is the *degree* of that node. When the edges of a tree are directed with orientation from a node to its immediate descendants, the *outdegree* of a node is the number of its immediate descendants.

MSC(2010): Primary: 05C05; Secondary: 60G99.

Keywords: Trees, random recursive trees, generalized Pólya urn models.

Received: 30 May 2010, Accepted: 4 December 2010.

*Corresponding author

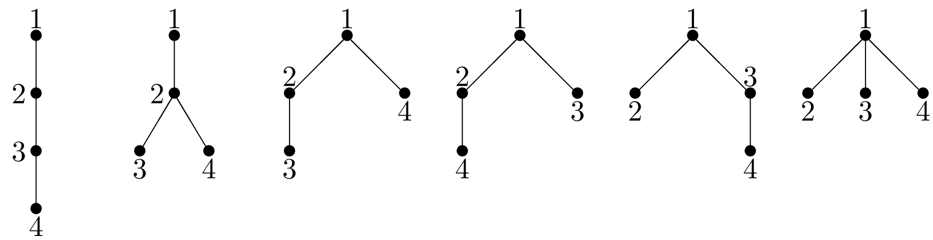

FIGURE 1. Recursive trees of order 4.

Note that there is no restriction on the outdegrees of the nodes of a recursive tree. A recursive tree in which outdegrees are equal to k, is called a *recursive k-ary tree.* In the following, the term *tree* without qualification will refer to a recursive k-ary tree.

The subtree rooted at the ith node in a tree is called the ith *branch* of the tree. The number of nodes in a tree is its *size.* we assume that if a tree has only one node, then that node is not a leaf.

Here, we work with some extensions of a tree, a representation in which a different type of nodes called *external* is added at each possible insertion position.

The nth node can be adjoined at any one of the insertion positions in a tree of size $n-1$. The probability that the nth node is joined to a given node of outdegree d, is then $k-d$, the number of remaining external nodes for that node, divided by $(k-1)(n-1)+1$, the number of all external nodes. A *random recursive k-ary tree* of size n, T_n, is a tree which is obtained by random choosing of a node as a parent in T_{n-1} and adjoining a node labelled n to it.

A *chain letter scheme with discount* can be considered as an application of random recursive k-ary trees. This model has been proposed as a model for chain letters where a company is founded to spread a particular item (lottery tickets, good luck charm, etc.). The initial recruiter looks for a willing participant to buy a copy of the letter. The recruiter and the new letter holder compete with offering discounts to attract new participants in proportion to the number of possible new participants that they are able to attract. The process proceeds in this way, where at each stage a participant (a node with outdegree d) who sells $d<k$ copies of the initial letter to d participants, offers discounts to attract new participants in proportion to $k-d$, the number of possible new participants that he or she is able to attract. This scheme increases

the chance of attracting new participants, for a participant who has had less customers.

Gastwirth and Bhattacharya [2] derived limiting distribution of the size of a branch in random recursive trees, to be a geometric distribution. Szymański [7] found the mean and variance of the size of random recursive trees. Also, the number of nodes of a given outdegree in a branch of recursive trees was investigated in Szymański [7].

In Section 2, we find the expectation of the size of a branch of a random recursive k-ary tree using the Pólya-Eggenberger urn scheme, a special case of a general Pólya urn scheme (see [7]). In sections 3 and 4, we compute the expected value of the number of nodes of a given outdegree in a branch.

2. **Size of a branch**

The Pólya urn models has been used in the context of size of a subtree of some classes of trees (see the survey in [5] for the wide applications of the Pólya urn models to random trees). We use a Pólya urn scheme to prove our first result for random recursive k-ary trees. Let $S_{n,i}$ denote the number of nodes in the ith branch of a random tree of size n.

Theorem 2.1. *For* $1 \leq i \leq n$,

$$\mathbf{E}[S_{n,i}] = \frac{kn - i + 1}{(k-1)i + 1}.$$

Proof. The process of joining new external nodes to an extended tree, with respect to the number of external nodes in the ith branch, is a Pólya-Eggenberger urn scheme. Initially, we have $(k-1)i + 1$ balls, $(k-1)(i-1)$ blue (not in ith branch) and k white (belonging to the ith branch). After each random drawing, $k-1$ balls of the same color as the ball withdrawn, are added to the urn. After $n - i$ drawings we have $(k-1)n + 1$ balls in the urn and the number of white balls equals $(k-1)S_{n,i} + 1$. The expectation of $(k-1)S_{n,i} + 1$ is an immediate consequence of results in [4]. □

3. **Number of leaves in a branch**

Let $L_{n,i}$ denote the number of leaves in the ith branch of a random tree T_n of size n. In the following, we determine $\mathbf{E}[L_{n,i}]$, the expectation of this random quantity. Some of our results involve the quantity

$$u_n = \prod_{j=1}^{n} \frac{(k-1)j}{(k-1)j+1},$$

where the product is interpreted as 1 when the range of j is empty.

Theorem 3.1. *For* $1 \le i \le n$,

$$\mathbf{E}[L_{n,i}] = \frac{k}{2k-1}\left(\frac{(k-1)n+1}{(k-1)i+1} - \frac{(i-1)u_{n-1}}{(n-1)u_{i-1}}\right). \tag{3.1}$$

If $n \longrightarrow \infty$, $i = o(n)$, *then*

$$\mathbf{E}[L_{n,i}] \sim \frac{k}{2k-1} \cdot \frac{n}{i}. \tag{3.2}$$

Proof. Adding the nth node to the random tree T_{n-1}, with $n-1$ nodes, the number of leaves in the ith branch either increases by 1, or stays the same. Then we have

$$\mathbf{P}\big(L_{n,i} = L_{n-1,i} + 1 \,|\, T_{n-1}\big) = \frac{(k-1)S_{n-1,i} + 1 - kL_{n-1,i}}{(k-1)(n-1)+1},$$

and

$$\mathbf{P}\big(L_{n,i} = L_{n-1,i} \,|\, T_{n-1}\big) = 1 - \frac{(k-1)S_{n-1,i} + 1 - kL_{n-1,i}}{(k-1)(n-1)+1}.$$

So,

$$\mathbf{E}[L_{n,i}|T_{n-1}] = (L_{n-1,i} + 1) \times \frac{(k-1)S_{n-1,i} + 1 - kL_{n-1,i}}{(k-1)(n-1)+1}$$
$$+ L_{n-1,i} \times \left(1 - \frac{(k-1)S_{n-1,i} + 1 - kL_{n-1,i}}{(k-1)(n-1)+1}\right).$$

Taking expectations, we obtain the following recurrence relation

$$\mathbf{E}[L_{n,i}] = \frac{(k-1)(n-2)}{(k-1)(n-1)+1}\mathbf{E}[L_{n-1,i}] + \frac{k}{(k-1)i+1},$$

with the boundary condition $\mathbf{E}[L_{i,i}] = 0$. Using standard methods for solving such recurrence relations, we get

$$\mathbf{E}[L_{n,i}] = \frac{k}{(k-1)i+1} \sum_{j=i}^{n-1} \frac{j}{n-1} \prod_{l=j+1}^{n-1} \frac{(k-1)l}{(k-1)l+1}.$$

This expression can be written in terms of the numbers u_j as

$$\begin{aligned}\mathbf{E}[L_{n,i}] &= \frac{k}{(k-1)i+1}\sum_{j=i}^{n-1}\frac{ju_{n-1}}{(n-1)u_j}\\ &= \frac{ku_{n-1}}{(n-1)[(k-1)i+1]}\left(\sum_{j=1}^{n-1}\frac{j}{u_j}-\sum_{j=1}^{i-1}\frac{j}{u_j}\right).\end{aligned}$$

The identity

$$\sum_{j=1}^{n-1}\frac{j}{u_j}=\frac{(n-1)[(k-1)n+1]}{(2k-1)u_{n-1}}$$

may be verified by induction and the result follows. □

4. Number of nodes of a given outdegree in a branch

Let $X_{n,d,i}$ denote the number of nodes of outdegree d in the ith branch of a random tree of size n. Of course, $L_{n,i} = X_{n,0,i}$.

Lemma 4.1. *For $d \geq 1$ and $n \geq i + d$,*

$$\mathbf{E}[X_{n,d,i}] = \frac{k-d+1}{(k-1)(n-1)+d}\sum_{j=i+d-1}^{n-1} C_{j,d}^{(n-1)}\mathbf{E}[X_{j,d-1,i}],$$

where

$$C_{j,d}^{(n)} = \prod_{l=j}^{n}\frac{(k-1)l+d}{(k-1)l+1}.$$

Proof. Adding the nth node to the random tree T_{n-1} with $n-1$ nodes, the number of nodes of outdegree $d \geq 1$ in the ith branch either increases by 1, decreases by 1, or stays the same. So,

$$\begin{aligned}\mathbf{E}[X_{n,d,i}|T_{n-1}] &= (X_{n-1,d,i}+1)\times\frac{(k-d+1)X_{n-1,d-1,i}}{(k-1)(n-1)+1}\\ &\quad +(X_{n-1,d,i}-1)\times\frac{(k-d)X_{n-1,d,i}}{(k-1)(n-1)+1}+X_{n-1,d,i}\\ &\quad \times\left(1-\frac{(k-d+1)X_{n-1,d-1,i}+(k-d)X_{n-1,d,i}}{(k-1)(n-1)+1}\right).\end{aligned}$$

Simplifying and taking expectations, we obtain

$$\begin{aligned}\mathbf{E}[X_{n,d,i}] &= \frac{(k-1)(n-2)+d}{(k-1)(n-1)+1}\mathbf{E}[X_{n-1,d,i}] \\ &\quad + \frac{k-d+1}{(k-1)(n-1)+1}\mathbf{E}[X_{n-1,d-1,i}],\end{aligned} \tag{4.1}$$

with boundary condition $\mathbf{E}[X_{i+d-1,d,i}] = 0$. By induction on n, we obtain the result. □

Corollary 4.2. *For* $1 \le i \le n$,

$$\begin{aligned}\mathbf{E}[X_{n,1,i}] &= \frac{k^2}{2(k-1)(2k-1)}\bigg(\frac{(k-1)n+1}{(k-1)i+1} - \frac{(k-1)(i-1)+1}{(k-1)(n-1)+1} \\ &\quad - \frac{2(k-1)(i-1)}{[(k-1)(n-1)+1]u_{i-1}}\sum_{j=i}^{n-1}\frac{u_{j-1}}{j-1}\bigg).\end{aligned} \tag{4.2}$$

If $n \longrightarrow \infty$, $i = o(n)$, *then*

$$\mathbf{E}[X_{n,1,i}] \sim \frac{k^2}{2(k-1)(2k-1)} \cdot \frac{n}{i}. \tag{4.3}$$

Proof. By the special case of Lemma 4.1, for $d = 1$, we have

$$\begin{aligned}\mathbf{E}[X_{n,1,i}] &= \frac{k}{(k-1)(n-1)+1}\sum_{j=i}^{n-1}\mathbf{E}[L_{j,i}] \\ &= \frac{k^2(n-i)[(k-1)(n+i-1)+2]}{2(2k-1)[(k-1)i+1][(k-1)(n-1)+1]} \\ &\quad - \frac{k^2(i-1)}{(2k-1)[(k-1)(n-1)+1]u_{i-1}}\sum_{j=i}^{n-1}\frac{u_{j-1}}{j-1} \\ &= \frac{k^2}{2(k-1)(2k-1)} \cdot \frac{(k-1)n+1}{(k-1)i+1} \\ &\quad - \frac{k^2}{2(k-1)(2k-1)} \cdot \frac{(k-1)(i-1)+1}{(k-1)(n-1)+1} \\ &\quad - \frac{k^2(i-1)}{(2k-1)[(k-1)(n-1)+1]u_{i-1}}\sum_{j=i}^{n-1}\frac{u_{j-1}}{j-1}.\end{aligned}$$

So the result (4.2) is derived. Since $\sum_{j=i}^{n-1}\frac{u_{j-1}}{j-1} = O\big(\ln(\frac{n}{i})\big)$, we obtain (4.3). □

Note that

$$C_{j,2}^{(n)} = \prod_{l=j}^{n} \frac{(k-1)l+2}{(k-1)l+1} = \prod_{l=j}^{n} \Big(1 + \frac{1}{(k-1)l+1}\Big)$$

$$= 1 + \sum_{\left\{S_j \subseteq \{j,\dots,n\} | S_j \neq \phi\right\}} \prod_{r \in S_j} \frac{1}{(k-1)r+1}. \tag{4.4}$$

By Lemma 4.1, (4.2) and (4.4), if $n \longrightarrow \infty$, $i = o(n)$, then we obtain

$$\begin{aligned} \mathbf{E}[X_{n,2,i}] &= \frac{k^2(k-1)}{2(k-1)(2k-1)[(k-1)(n-1)+2]} \\ &\quad \times \sum_{j=i+1}^{n-1} \frac{(k-1)j+1}{(k-1)i+1} + o(1) \\ &= \frac{k^2(k-1)(n-i-1)[(k-1)(n+i)+2]}{2^2(k-1)(2k-1)[(k-1)i+1][(k-1)(n-1)+2]} + o(1) \\ &= \frac{k^2(k-1)}{2^2(k-1)^2(2k-1)} \cdot \frac{(k-1)n+1}{(k-1)i+1} + o(1). \qquad (4.5) \end{aligned}$$

The evidence accumulating from the relations (3.2), (3.2) and (4.5) suggests an asymptotic result for $d \geq 0$, as given below.

Theorem 4.3. *If $i, n \longrightarrow \infty$ such that $\frac{n}{i} \longrightarrow \infty$, then*

$$\mathbf{E}[X_{n,d,i}] \sim \frac{k \prod_{j=1}^{d}(k-j+1)}{2^d(k-1)^d(2k-1)} \cdot \frac{n}{i},$$

for $0 \leq d \leq k-1$, and

$$\mathbf{E}[X_{n,k,i}] \sim \left(\frac{k}{k-1} - \frac{k}{2k-1} \sum_{d=0}^{k-1} \frac{\prod_{j=1}^{d}(k-j+1)}{2^d(k-1)^d}\right) \frac{n}{i},$$

where the product $\prod_{j=1}^{d}(k-j+1)$ is interpreted as 1 when the range of j is empty.

Proof. By (3.2), the assertion is correct for $d = 0$. Let $\varepsilon_d = \frac{k \prod_{j=1}^{d}(k-j+1)}{2^d(k-1)^d(2k-1)}$ and $\alpha_{n,d,i} = \mathbf{E}[X_{n,d,i}] - \varepsilon_d \frac{(k-1)n+1}{(k-1)i+1}$. We shall show that

$$\frac{|\alpha_{n,d,i}|}{\varepsilon_d \cdot \frac{n}{i}} \overset{\frac{n}{i} \to \infty}{\longrightarrow} 0, \ \text{for } 1 \leq d \leq k-1. \tag{4.6}$$

Substitute $\alpha_{n,d,i}$ in (4.1) to get

$$\begin{aligned}\alpha_{n+1,d,i} &= \frac{(k-1)(n-1)+d}{(k-1)n+1}\alpha_{n,d,i}+\frac{k-d+1}{(k-1)n+1}\alpha_{n,d-1,i}\\ &\quad +\frac{(d-1)\varepsilon_d}{(k-1)i+1}. \qquad (4.7)\end{aligned}$$

Assume that for some $1 \le d \le k-1$, there exists a positive constant $C^{(j)}$ such that $|\alpha_{n,j,i}| \le C^{(j)}$, for all $n \ge i+j$, and for all $1 \le j \le d-1$ (by (4.2), this holds for $d = 2$). For this d in the assumption, it is sufficient to prove that there exists a positive constant $C^{(d)}$ such that $|\alpha_{n,d,i}| \le C^{(d)}$, for all $n \ge i+d$.
From (4.7),

$$\begin{aligned}|\alpha_{n+1,d,i}| &\le \frac{(k-1)(n-1)+d}{(k-1)n+1}|\alpha_{n,d,i}|+\frac{(k-d+1)C^{(d-1)}}{(k-1)n+1}\\ &\quad +\frac{(d-1)\varepsilon_d}{(k-1)i+1}. \qquad (4.8)\end{aligned}$$

Choose $C^{(d)} > 2\ max\{(k-d+1)C^{(d-1)}, 1+(d+1)\varepsilon_d\}$. Then,

$$\begin{aligned}&\frac{[(k-1)(n-1)+d]C^{(d)}}{(k-1)n+1}+\frac{(k-d+1)C^{(d-1)}}{(k-1)n+1}+\frac{(d-1)\varepsilon_d}{(k-1)i+1}\\ &\quad < C^{(d)}+\frac{(d-k)C^{(d)}}{(k-1)n+1}+\frac{C^{(d)}/2}{(k-1)n+1}+\frac{C^{(d)}/2}{(k-1)i+1}\\ &\quad < C^{(d)}+\frac{(d-k+1)C^{(d)}}{(k-1)i+1} \le C^{(d)}, \qquad (4.9)\end{aligned}$$

for all $n \ge i+d$, (by the induction hypothesis on d, $1 \le d \le k-1$). By definition of $\alpha_{n,d,i}$,

$$\alpha_{i+d,d,i} = \prod_{j=1}^{d}\frac{(k-j+1)}{(k-1)(i+j-1)+1}-\varepsilon_d\frac{(k-1)(i+d)+1}{(k-1)i+1},$$

and so $|\alpha_{i+d,d,i}| < 1+(d+1)\varepsilon_d < C^{(d)}$. Using this, (4.9), and taking $n = i+d$ in (4.8), we get

$$\begin{aligned}|\alpha_{i+d+1,d,i}| &\le \frac{[(k-1)(i+d-1)+d]C^{(d)}}{(k-1)(i+d)+1}+\frac{(k-d+1)C^{(d-1)}}{(k-1)(i+d)+1}\\ &\quad +\frac{(d-1)\varepsilon_d}{(k-1)i+1} < C^{(d)},\end{aligned}$$

and an induction on n gives $|\alpha_{n,d,i}| < C^{(d)}$, for all $n \geq i + d$ (for the d of the induction hypothesis on d). So, an induction on d gives $|\alpha_{n,d,i}| < C^{(d)}$, for all $n \geq i + d$ and all $1 \leq d \leq k - 1$. Now, choose $C = \min\{C^{(1)}, C^{(2)}, \ldots, C^{(k-1)}\}$. Then, $|\alpha_{n,d,i}| < C$, for all $n \geq i + d$ and all $1 \leq d \leq k-1$. This implies (4.6). For the case of $d = k$, since $\mathbf{E}[X_{n,k,i}] = \mathbf{E}[S_{n,i}] - \sum_{d=0}^{k-1} \mathbf{E}[X_{n,d,i}]$ and by Theorem 2.1, $\mathbf{E}[S_{n,i}] \sim \frac{k}{k-1} \cdot \frac{n}{i}$, and the proof is complete. $\square$

References

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

[2] J. Gastwirth and P. Bhattacharya, Two probability models of pyramids or chain letter schemes demonstrating that their promotional claims are unreliable, *Operations Research* **32** (1984), no. 3, 527–536.

[3] F. Harary, Graphs Theory, Addison-Wesley, Reading, MA, 1969.

[4] N. Johnson and S. Kotz, Urn Models and Their Applications, Wiley, New York, 1977.

[5] H. M. Mahmoud, Urn models and connections to random trees: A review, *J. Iran. Stat. Soc.* **2** (2003), 53–114.

[6] R. Smythe and H. Mahmoud, A survey of recursive trees, *Theory Probab. Math. Statist.* **51** (1995), 1–27.

[7] J. Szymański, Branches in recursive trees, *Fasc. Math.* **29** (1999), 139–147.

Mehri Javanian

Department of Statistics, Faculty of Sciences, Zanjan University, Zanjan, Iran

Email: javanian_m@yahoo.com

Mohammad Q. Vahidi-Asl

Department of Statistics, Shahid Beheshti University, P.O. Box 19835-389, Tehran, Iran

Email: m_vahidi@sbu.ac.ir