FASCICULI MATHEMATICI

Nr 29 1999

JERZY SZYMAŃSKI *

BRANCHES IN RECURSIVE TREES

ABSTRACT. In the paper various properties of subtrees of a random recursive tree are studied. In particular we derive the probability distribution of the size of a branch, the first two moments of the number of leaves and the number of root free paths.

KEY WORDS. recursive tree, branch of a tree, subtree, path, root free path.

1. INTRODUCTION

A *tree* is a connected graph which has no cycles (see [1] for definitions not given here). A tree R with n vertices labeled $1,2,\dots,n$ is a *recursive tree* if for each k such that $2 \le k \le n$, labels of vertices in the unique path from the first vertex to the k-th vertex form an increasing subsequence of $\{1,2,\dots,n\}$. Such a tree can be also defined as a result of successively joining of the i-th vertex to one of the first $i-1$ vertices. Figure 1 shows all recursive trees with four vertices.

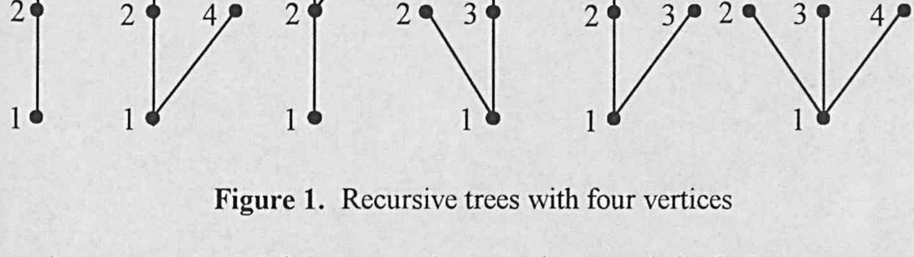

Figure 1. Recursive trees with four vertices

A vertex with label 1 is a *root* of a recursive tree. A *leaf* of a tree is a vertex of degree one (we assume that the root of a tree is not a leaf even if it has degree one). Leaves are also called *endvertices* or *external vertices* while the vertices that are not leaves are called *internal* or *inner vertices*. A *branch* (or a *subtree*) rooted at point j of a recursive tree R is a recursive tree which consist of all vertices of the tree R such that the path from the root to the vertices of the branch contains vertex j.

A *random recursive tree* with n vertices is a tree picked at random from the family $\mathcal{R}_n$ of all $(n-1)!$ recursive trees with n vertices. We assume that all $(n-1)!$ possible choices of a tree are equiprobable (see e.g. [5] for not-equiprobable models of a random recursive tree).

Our main object here is to find some properties of subtrees of a random recursive tree.

For a tree R with n vertices let $\alpha_i(R)$ denote a tree obtained from the tree R by adding new vertex labeled $n+1$ and joining it with vertex i. Of course $\alpha_i(R) \in \mathcal{R}_{n+1}$.

For a random variable X let $E[X]$, $E_k[X]$ and $\mathrm{Var}[X]$ denote the expected value, k-th factorial moment and the variance of X, respectively.

2. THE SIZE OF A BRANCH

Let $G_{ni} = G_{ni}(R)$ denote the number of vertices in the i-th subtree of a random recursive tree R with n vertices.

Theorem 1. For $1 \le i \le n$

$$E[G_{ni}] = \frac{n}{i}$$

and

$$\mathrm{Var}[G_{ni}] = \frac{n(n-i)(i-1)}{i^2(i+1)}.$$

Moreover, for $1 \le k \le n-i+1$

$$\mathrm{Prob}(G_{ni} = k) = \binom{n-i}{k-1} \frac{\left(\frac{1}{i}\right)^{[k-1]} \left(1-\frac{1}{i}\right)^{[n-i-k+1]}}{(n-i)!},$$

where $(n)^{[k]} = n(n+1)\cdots(n+k-1)$ *is a factorial power.*

Proof. Let us consider the process of adding new vertices to the recursive tree with respect to the number of vertices in the i-th branch. It is a particular case of a general Pólya urn scheme. We outline the scheme here, for more complete description see for example [3].

In the Pólya scheme a single urn initialy contains w white balls and b black balls. A ball is drawn at random and then replaced, together with s balls of the same color. The procedure is repeated n times. Let X be the random variable representing the number of times a black ball is drawn. Then X has a Pólya-Eggenberger distribution, that is

$$\mathrm{Prob}(X=k)=\binom{n}{k}\frac{\alpha^{[k]}\beta^{[n-k]}}{(\alpha+\beta)^{[n]}},$$

where $\alpha=\frac{b}{s}$, $\beta=\frac{w}{s}$ and $x^{[n]}=x(x+1)(x+2)\cdots(x+k-1)$ is a factorial power. Moreover the expectation and the variance of the random variable X are equal to

$$E[X]=\frac{n\alpha}{\alpha+\beta}$$

and

$$Var[X]=\frac{n\alpha}{\alpha+\beta}\left(\frac{(n-1)(\alpha+1)}{\alpha+\beta+1}+1-\frac{n\alpha}{\alpha+\beta}\right).$$

Returning to recursive trees, initially we have i balls, $i-1$ white (not in the i-th subtree) and one black (belonging to the branch). After each drawing, the chosen ball is returned together with one additional ball of the same color (i.e. $s=1$). After $n-i$ drawings we have n balls in the urn and the number of black balls is equivalent to the size of i-th branch. Now, the random variable G_{ni} has a Pólya-Eggenberger distribution and the Theorem 1 is an immediate consequence of a general Pólya urn model theory. ■

Let us notice that for $i=1$ the whole tree is a branch, and the result is obvious. Similarly, for $i=n$ the size of the subtree is one.

3. THE NUMBER OF LEAVES IN A SUBTREE

Let $L_{ni}=L_{ni}(R)$ denote the number of endvertices in the i-th branch of a random recursive tree R with n vertices. It is known (see [2] or [4]) that $E[L_{n,1}]=n/2$. Here we derive the expected value and secound factorial moment of L_{ni} in a general case.

Theorem 2. For $1\le i\le n$

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

and

$$E_2[L_{ni}] = \frac{n(n+1)}{2i(i+1)} - \frac{2n}{3i} + \frac{(i-1)(i-2)}{6(n-1)(n-2)}.$$

Proof. From the way a recursive tree with $n+1$ vertices is obtained from a tree with n vertices we get

(1) $$E\left[L_{n+1,i}\right] = \frac{1}{n!} \sum_{R \in \mathfrak{R}_n} \sum_{j=1}^{n} L_{n+1,i}\left(\alpha_j(R)\right).$$

Fix a tree R with n vertices. Adding $(n+1)$-st vertex to this tree one can obtain n recursive trees with $n+1$ vertices $\alpha_1(R)$, $\alpha_2(R)$, ..., $\alpha_n(R)$. Then the number of leaves in the i-th branch can be increased by one (if we join $(n+1)$-st vertex to an inner vertex of a subtree) or be the same. So,

$$\begin{aligned}&\sum_{j=1}^{n} L_{n+1,i}\left(\alpha_j(R)\right) \\ &= \left(G_{ni}(R) - L_{ni}(R)\right)\left(L_{ni}(R)+1\right) + \left(n - G_{ni}(R) + L_{ni}(R)\right)L_{ni}(R) \\ &= G_{ni}(R) + (n-1)L_{ni}(R).\end{aligned}$$

Therefore, using (1) and Theorem 1 we get

$$E\left[L_{n+1,i}\right] = \frac{1}{n} E\left[G_{ni}\right] + \frac{n-1}{n} E\left[L_{ni}\right]$$

and

$$E\left[L_{n+1,i}\right] = \frac{n-1}{n} E\left[L_{ni}\right] + \frac{1}{i}.$$

Solving this linear recurrence equation with initial condition $E\left[L_{ii}\right] = 0$ one can get the required formula for $E\left[L_{ni}\right]$.

Similarly

(2) $$E_2\left[L_{n+1,i}\right] = \frac{1}{n!} \sum_{R \in \mathfrak{R}_n} \sum_{j=1}^{n} \left(L_{n+1,i}\left(\alpha_j(R)\right)\right)_2,$$

but

$$\sum_{j=1}^{n}\left(L_{n+1,i}\left(\alpha_j(R)\right)\right)_2$$
$$= \left(G_{ni}(R)-L_{ni}(R)\right)\left(L_{ni}(R)+1\right)_2 + \left(n-G_{ni}(R)+L_{ni}(R)\right)\left(L_{ni}(R)\right)_2$$
$$= n\left(L_{ni}(R)\right)_2 + 2L_{ni}(R)G_{ni}(R) - 2\left(L_{ni}(R)\right)^2$$
$$= (n-2)\left(L_{ni}(R)\right)_2 + 2L_{ni}(R)G_{ni}(R) - 2L_{ni}(R)$$

because $(x+1)_2 = (x)_2 + 2x$ and $x^2 = (x)_2 + x$. Putting it into formula (2) we get

$$E_2\left[L_{n+1,i}\right] = \frac{n-2}{n}E_2\left[L_{ni}\right] + \frac{2}{n}E\left[L_{ni}G_{ni}\right] - \frac{2}{n}E\left[L_{ni}\right]$$

and finally

(3) $$E_2\left[L_{n+1,i}\right] = \frac{n-2}{n}E_2\left[L_{ni}\right] + \frac{2}{n}E\left[L_{ni}G_{ni}\right] - \frac{2}{i}.$$

Now we will find $E\left[L_{ni}G_{ni}\right]$. For simplicity let us denote $\eta(n) = E\left[L_{ni}G_{ni}\right]$. It is easy to see, that due to the way of construction of a recursive tree one can get

$$\eta(n+1) = \frac{1}{n!}\sum_{R\in\Re_n}\sum_{j=1}^{n} L_{n+1,i}\left(\alpha_j(R)\right)G_{n+1,i}\left(\alpha_j(R)\right),$$

and by similar arguments we find

$$\sum_{j=1}^{n} L_{n+1,i}\left(\alpha_j(R)\right)G_{n+1,i}\left(\alpha_j(R)\right)$$
$$= \left(n-G_{ni}(R)\right)L_{ni}(R)G_{ni}(R) + L_{ni}^2(R)\left(G_{ni}(R)+1\right)$$
$$+ \left(G_{ni}(R)-L_{ni}(R)\right)\left(L_{ni}(R)+1\right)\left(G_{ni}(R)+1\right)$$
$$= nL_{ni}(R)G_{ni}(R) + G_{ni}^2(R) + G_{ni}(R) - L_{ni}(R).$$

So, we have

$$\eta(n+1) = \eta(n) + \frac{1}{n}E\left[G_{ni}^2\right] + \frac{1}{n}E\left[G_{ni}\right] - \frac{1}{n}E\left[L_{ni}\right]$$
$$= \eta(n) + 2\frac{n+1}{i(i+1)} - \frac{1}{2i} + \frac{i-1}{2n(n-1)}$$

with boundary condition $\eta(i) = 0$. Solving this recurrence relation we obtain

$$\eta(n) = \sum_{j=i}^{n-1}\left(2\frac{j+1}{i(i+1)} - \frac{1}{2i} + \frac{i-1}{2j(j-1)}\right)$$

and further, after elementary calculations

$$\eta(n)=\frac{n(n+1)}{i(i+1)}-\frac{n}{2i}-\frac{1}{2}+\frac{n-i}{2(n-1)}. \tag{4}$$

Putting this to the recurrence relation (3) we find that

$$E_2\left[L_{n+1,i}\right]=\frac{n-2}{n}E_2\left[L_{ni}\right]+f_n, \tag{5}$$

where $f_n=\frac{2(n+1)}{i(i+1)}-\frac{2}{i}$. Let us denote $g_n=(n-1)(n-2)E_2\left[L_{ni}\right]$. Then (5) can be rewrited in the form

$$g_{n+1}=g_n+n(n-1)f_n,$$

with initial condition $g_{i+1}=0$. Solving this recurrence we get

$$g_n=\frac{2}{i(i+1)}\sum_{j=i+1}^{n-1}(j)_3-\frac{2}{i}\sum_{j=i+1}^{n-1}(j)_2$$

and the required formula for $E_2\left[L_{ni}\right]$ follows. ■

Let us mention that due to the Theorem 1 the expected size of the i-th subtree is $\frac{n}{i}$, and for a recursive tree one half of its vertices are leaves, so one can expect that the number of leaves in the i-th branch approximates $\frac{n}{2i}$ in average. As a matter of fact, we have the following result.

Corollary 2.1. If $n\to\infty$ *and* i *is fixed then*

$$E[L_{ni}]\sim\frac{n}{2i}$$

and

$$\operatorname{Var}\left[L_{ni}\right]\sim\begin{cases}\frac{1}{12}n, & \text{if } i=1,\\ \frac{i-1}{4i^2(i+1)}n^2, & \text{if } i>1.\end{cases}$$ ■

Corollary 2.2. If $n\to\infty$ *and* $i\to\infty$ *but* $i=o(n)$ *then*

$$E[L_{ni}] \sim \frac{n}{2i}$$

and

$$\mathrm{Var}[L_{ni}] \sim \frac{n^2}{4i^2}. \qquad \blacksquare$$

Corollary 2.3. If $n \to \infty$ *and* $i = \alpha n$ *(where* α *is a constant such that* $0 < \alpha < 1$*) then*

$$E[L_{ni}] \sim \frac{(1+\alpha)(1-\alpha)}{2\alpha}$$

and

$$\mathrm{Var}[L_{ni}] \sim \frac{(1-\alpha)(\alpha^3 + 7\alpha^2 + \alpha + 3)}{12\alpha^2}. \qquad \blacksquare$$

3. ROOT FREE PATHS

A path of a recursive tree which does not contain the vertex with label one is called a *root free path*. Let us denote for a random recursive tree with n vertices:

- $F_{E,E}(n)$ the number of root free paths such that both their ends are endvertices,
- $F_{I,I}(n)$ the number of root free paths such that both their ends are not endvertices,
- $F_{I,E}(n)$ the number of root free paths such this an endvertex and the other is not,
- $F_T(n)$ the total number of root free paths.

Of course $F_T(n) = F_{E,E}(n) + F_{I,I}(n) + F_{I,E}(n)$.

Notice, that the total number of paths in a tree is equal to the number of pairs of vertices (i.e. $\binom{n}{2}$).

Theorem 3. If $n \to \infty$ *then*

$$E[F_{E,E}(n)] \sim \frac{n^2}{16},$$

$$E[F_{I,I}(n)] \sim \frac{n^2}{16},$$

$$E[F_{I,E}(n)] \sim \frac{n^2}{8}$$

and

$$E[F_T(n)] \sim \frac{n^2}{16}.$$

Proof. We only prove the first relation, proofs of the others are similar.

Let $S_1 = S_1(R)$ denote the set of vertices incident to the root of a recursive tree R. A subtree rooted at a vertex from S_1 is called *main branch*.

Notice, that a path of a tree is root free if and only if both its ends are in the same main branch. So,

$$E[F_{E,E}(n)] = \frac{1}{2(n-1)!} \sum_{R \in \Re_n} \sum_{i \in S_1(R)} L_{ni}(R)(L_{ni}(R) - 1).$$

Let $\xi_i(R)$ be defined as follow

$$\xi_i(R) = \begin{cases} 0, & \text{if } i \notin S_1, \\ 1, & \text{if } i \in S_1. \end{cases}$$

Using this notation we get

$$E[F_{E,E}(n)] = \frac{1}{2(n-1)!} \sum_{R \in \Re_n} \sum_{i=2}^{n-2} \xi_i(R) L_{ni}(R)(L_{ni}(R) - 1) \tag{6}$$
$$= \frac{1}{2} \sum_{i=2}^{n-2} E[\xi_i L_{ni}(L_{ni} - 1)].$$

Let us fix i. From the definition of the expected value we have

$$E[\xi_i L_{ni}(L_{ni} - 1)] = \sum_j j \operatorname{Prob}(\xi_i L_{ni}(L_{ni} - 1) = j)$$
$$= \sum_j j \operatorname{Prob}(L_{ni}(L_{ni} - 1) = j \mid i \in S_1) \operatorname{Prob}(i \in S_1).$$

Clearly $\operatorname{Prob}(i \in S_1) = \frac{1}{i-1}$ and one can see that random events $i \in S_1$ and $L_{ni}(L_{ni} - 1) = j$ are independent. Therefore

$$E[\xi_i L_{ni}(L_{ni}-1)] = \frac{1}{i-1} E_2[L_{ni}].$$

Using (6) we get

$$E[F_{E,E}(n)] = \frac{1}{2}\sum_{i=2}^{n-2} \frac{E_2[L_{ni}]}{i-1},$$

and due to the Theorem 2 we obtain

$$\begin{aligned} &E[F_{E,E}(n)] \\ &= \frac{n(n+1)}{4}\sum_{i=2}^{n-2}\frac{1}{(i+1)i(i-1)} - \frac{n}{3}\sum_{i=2}^{n-2}\frac{1}{i(i-1)} + \frac{1}{6(n-1)(n-2)}\sum_{i=2}^{n-2}(i-2) \\ &= \frac{(n-3)(3n^3-13n^2+20n-16)}{48(n-1)(n-2)}. \end{aligned}$$

This implies that $E[F_{E,E}(n)] \sim \dfrac{n^2}{16}$. ■

Notice, that asymptotically one half of all the paths of a recursive tree are root free paths.

REFERENCES

[1] C. Berge, *Graphs and Hypergraphs*, North-Holland, 1973.

[2] M. Dondajewski, J. Szymański, On the distribution of vertex degrees in strata of a random recursive tree, *Bull. Acad. Polon. Sci. Ser. Sci. Math.*, XXX, 5–6(1982), 205–209.

[3] N.L. Johnson, S. Kotz, *Urn Models and Their Application*, John Wiley & Sons, 1985.

[4] H.S. Na, A. Rapoport, Distribution of nodes of a tree by degree, *Math. Biosciences*, 6(1970), 313–329.

[5] J. Szymański, On a nonuniform random recursive trees, *Annals of Discrete Math.* 33(1987), 297–306.

(*Department of Discrete Mathematics, Adam Mickiewicz University, 60–769 Poznań, Poland, *E-mail address*: jszymans@amu.edu.pl)

Received on 05.12.1997 and, in revised form, on 17.09.1998.

Research supported by grant KBN 2 P03A 023 09