RAIRO-Theor. Inf. Appl. 57 (2023) 5
https://doi.org/10.1051/ita/2023003

RAIRO - Theoretical Informatics and Applications
www.rairo-ita.org

NOTE ON THE EXPONENTIAL RECURSIVE k-ARY TREES

Mina Ghasemi[1], Mehri Javanian[2,*] and Ramin Imany Nabiyyi[1]

Abstract. In the present paper, we consider exponential recursive trees with no node of outdegree greater than k, called exponential recursive k-ary trees ($k \geq 2$). At each step of growing of these trees, every external node (insertion position) is changed into a leaf with probability p, or fails to do so with probability $1 - p$. We investigate limiting behavior of fundamental parameters such as size, leaves and distances in exponential recursive k-ary trees.

Mathematics Subject Classification. 05C05, 60C05, 60F05, 60G42.

Received April 6, 2022. Accepted February 17, 2022.

1. Introduction

A social network site is a web-based platform which people use to build social relationships with family, friends, colleagues, customers, or clients. Social networking can have a social purpose, a business purpose, or both, through sites like Instagram, WhatsApp, Facebook, Telegram, YouTube, and Twitter. These sites allow people and corporations to connect with one another so they can develop relationships very quickly. Certain aspects of fast-growing social networks can be modeled by exponential recursive trees, see [1, 4, 7].

By a *recursive tree*, we mean a labeled rooted tree where the nodes on a path from the root to any other node form an increasing sequence. Recursive trees grow dynamically, by attaching every new node to one of the already present nodes. The recursive trees are slow-growing where one node is added at each step (see Fig. 1). So, the recursive tree models cannot be suitable for fast-growing phenomena (*e.g.*, the Corona virus spreads very quickly from person to person). See a survey of results of recursive trees in [10].

A *recursive k-ary tree* is a slow version of recursive trees in which every node has at most k children. The nodes that already exist in the tree are called *internal*. In order to identify the locations of possible insertions of a new node (internal node), a recursive k-ary tree is extended by having $k - d$ *external* nodes attached as children of a node (internal node) of outdegree d (*e.g.*, for $k = 3$, in Fig. 2, see growing a recursive 3-ary tree of size 3). In an extended recursive k-ary tree of n nodes, the number of all external nodes is $(k - 1)n + 1$, see [5].

An alternative fast-growing model, the exponential recursive k-ary tree ($k \geq 2$), grows by having every external node either converted into a leaf (an internal node of outdegree 0) with probability p or stay unchanged with probability $q := 1 - p$.

We define the *random exponential recursive k-ary tree of index p and age n* grown as follows. Initially (at age 0), there is an external node. For $n \geq 1$, at the nth step (at age n), every external node independently is

Keywords and phrases: Recursive k-ary trees, contraction method, almost sure convergence, martingale.

[1] Department of Statistics, Faculty of Mathematical Sciences, University of Tabriz, Tabriz, Iran.
[2] Department of Statistics, Faculty of Sciences, University of Zanjan, Zanjan, Iran.

[*] Corresponding author: javanian@znu.ac.ir

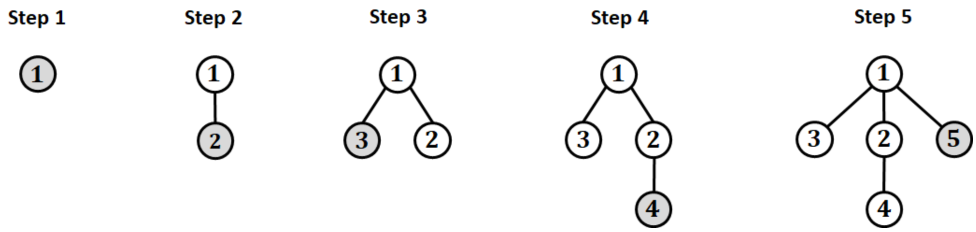

FIGURE 1. The steps of evolution of a recursive tree of size 5.

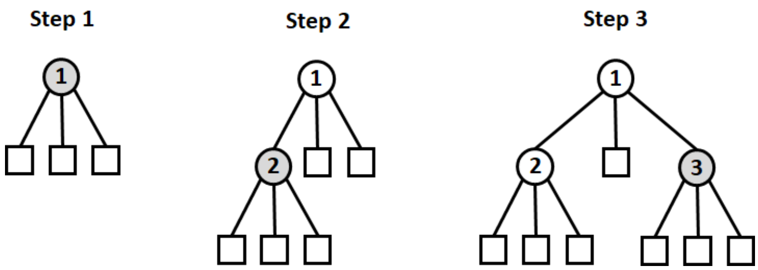

FIGURE 2. The steps of evolution of a recursive ternary tree of size 3, where internal nodes are shown as circles; external nodes are shown as squares.

changed into a leaf with probability $p \in (0,1)$, or not to be changed with probability $q = 1 - p$. Figure 3 shows all exponential recursive ternary trees of ages 0, 1 and 2 and their probabilities. In the figure, external nodes and internal nodes are represented by squares and circles, respectively. The probability of getting each tree from the previous one appears above the arrow leading to it. For example, the leftmost tree in the third row occurs with probability p^2q^2.

2. Size

Let S_n be the size of a random exponential recursive k-ary tree of index p and age n ($k \geq 2$), and let Y_n be the number of external nodes in the extension. Thus, we have

$$Y_n = (k-1)S_n + 1. \tag{2.1}$$

Let the notation $\mathrm{Bin}(m, p)$ stands for a binomial random variable that counts the number of successes in m independent identically distributed trials with success probability p in each trial. Each external node that succeeds to be converted into a leaf adds $k-1$ external nodes; we have

$$Y_n = Y_{n-1} + (k-1)\mathrm{Bin}(Y_{n-1}, p).$$

Let $\mathcal{F}_n$ be the sigma field generated by the first n steps of evolution. Then,

$$\mathbb{E}[Y_n|\mathcal{F}_{n-1}] = Y_{n-1} + (k-1)pY_{n-1} = ((k-1)p+1)Y_{n-1}. \tag{2.2}$$
$$\mathbb{E}[Y_n^2|\mathcal{F}_{n-1}] = ((k-1)p+1)^2Y_{n-1}^2 + (k-1)^2p(1-p)Y_{n-1}.$$

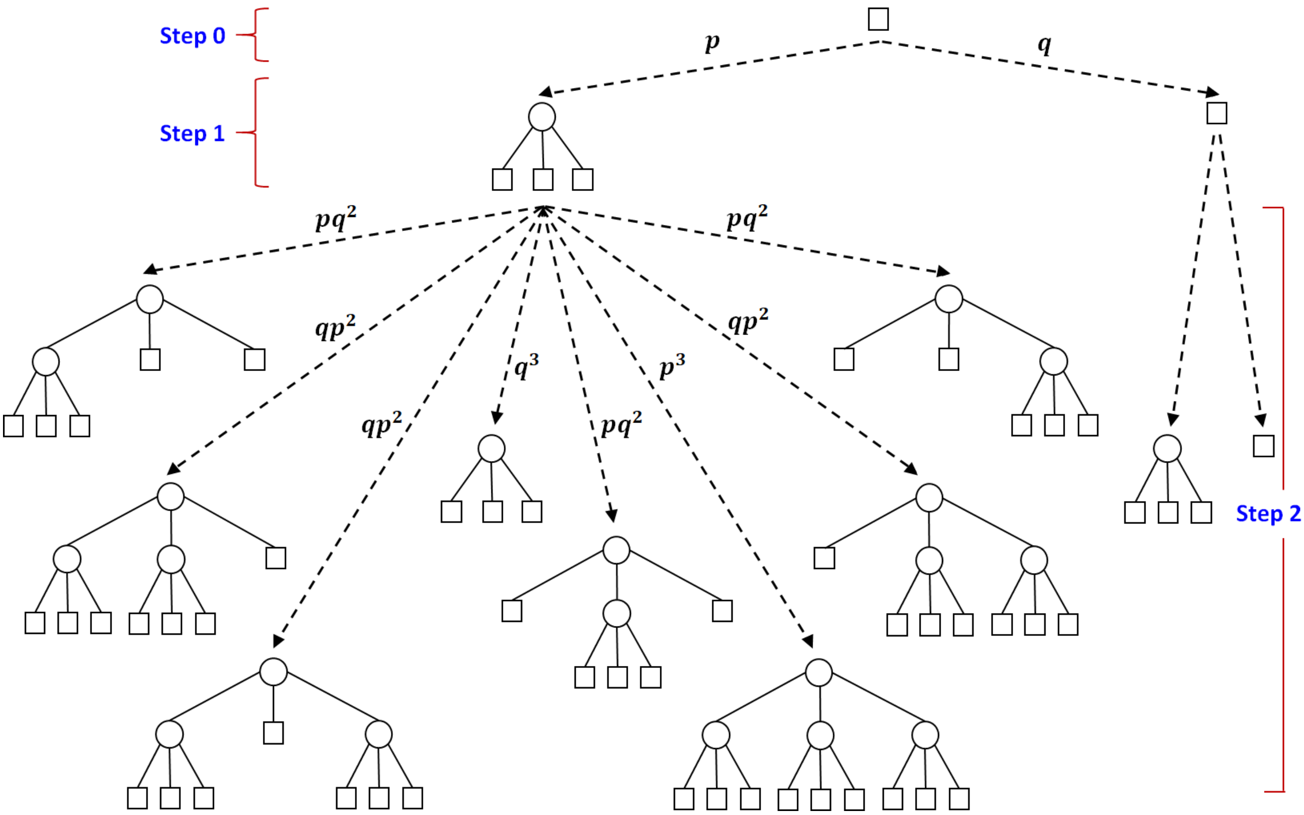

FIGURE 3. The first two steps of evolution of the exponential recursive ternary tree of index p. The probability of getting each tree from the previous one appears above the arrow leading to it.

It follows that $Y_n/((k-1)p+1)^n$ is a martingale with bounded moments

$$\mathbb{E}\left[\frac{Y_n}{((k-1)p+1)^n}\right] = \mathbb{E}[Y_0] = 1, \qquad \mathbb{E}\left[\frac{Y_n^2}{((k-1)p+1)^{2n}}\right] = \frac{(k-1)p\big((k-1)p+1\big)+k}{\big((k-1)p+1\big)^2} < \infty.$$

That is, $\mathbb{E}[Y_n] = ((k-1)p+1)^n$ and, by (2.1),

$$\mathbb{E}[S_n] = \frac{((k-1)p+1)^n - 1}{k-1}. \tag{2.3}$$

According to the martingale convergence theorem (Thm. 6.6.9 in [2]), we have

$$\frac{Y_n}{((k-1)p+1)^n} \xrightarrow{L^2 \text{ \& a.s.}} Y_*, \tag{2.4}$$

where Y_* is an integrable limit random variable. Since $\frac{1}{((k-1)p+1)^n} \to 0$, we have

$$\frac{S_n}{((k-1)p+1)^n} \xrightarrow{\text{a.s.}} S_* := \frac{Y_*}{k-1}. \tag{2.5}$$

for a limiting random variable S_*.

We now characterize the limiting distribution of $S_n/((k-1)p+1)^n$ by inductively-constructed moments. The calculation techniques were taken from [1].

Theorem 2.1. *Let S_n be the number of nodes in an exponential recursive k-ary tree of age n and index p ($k \ge 2$). As $n \to \infty$, we have almost sure convergence*

$$\frac{S_n}{((k-1)p+1)^n} \xrightarrow{\text{a.s.}} S_*,$$

where the limiting random variable S_* has moments $a_m := \mathbb{E}[S_*^m]$ defined inductively by

$$a_m = \frac{p}{((k-1)p+1)^m - ((k-1)p+1)} \sum_{\substack{i_1,\dots,i_k \ge 0 \\ i_1 \ne m,\dots,i_k \ne m \\ i_1+\cdots+i_k=m}} \binom{m}{i_1,\dots,i_k} a_{i_1}\cdots a_{i_k},$$

for $m \ge 2$, with $a_1 = 1/(k-1)$.

Proof. After one step of evolution, the initial external node either succeeds to recruit a leaf (event $\mathcal{R}$) or not. Let $\mathbb{I}$ be the indicator of event $\mathcal{R}$, that assumes the value 1, if $\mathcal{R}$ occurs, and is 0 otherwise. If $\mathcal{R}$ occurs, the extended exponential recursive k-ary tree has an internal node with k external nodes. By time n, these k external nodes produce k independent exponential recursive k-ary subtrees of sizes $S_{n-1}^{(1)}, S_{n-1}^{(2)}, \dots, S_{n-1}^{(k)}$. These k random variables are independent and each is distributed like S_{n-1}. Alternatively, if $\mathcal{R}$ does not occur, the size of the exponential recursive k-ary tree at step n is the same as S_{n-1}. Hence, the size satisfies the distributional equation

$$S_n \overset{d}{=} (1-\mathbb{I})S_{n-1} + \mathbb{I}\big(S_{n-1}^{(1)} + S_{n-1}^{(2)} + \cdots + S_{n-1}^{(k)} + 1\big), \tag{2.6}$$

where the symbol $\overset{d}{=}$ denotes the equality in distribution.

Raise both sides of the equation (2.6), to the mth power. Since $(1-\mathbb{I})^i\mathbb{I}^{m-i} = 0$, for $i > 0$ and $i < m$, we have

$$\begin{aligned} S_n^m &\overset{d}{=} \sum_{i=0}^{m} \binom{m}{i} (1-\mathbb{I})^i S_{n-1}^i \times \mathbb{I}^{m-i}\big(S_{n-1}^{(1)} + S_{n-1}^{(2)} + \cdots + S_{n-1}^{(k)} + 1\big)^{m-i} \\ &= (1-\mathbb{I})^m S_{n-1}^m + \mathbb{I}^m\big(S_{n-1}^{(1)} + S_{n-1}^{(2)} + \cdots + S_{n-1}^{(k)} + 1\big)^m \\ &\overset{d}{=} (1-\mathbb{I})S_{n-1}^m + \mathbb{I}\big(S_{n-1}^{(1)} + S_{n-1}^{(2)} + \cdots + S_{n-1}^{(k)} + 1\big)^m. \end{aligned} \tag{2.7}$$

Take expectations both sides of the equation (2.7). By the independence relations and identical distribution in the k subtrees, we get

$$\begin{aligned} \mathbb{E}[S_n^m] &= \mathbb{E}[1-\mathbb{I}]\mathbb{E}[S_{n-1}^m] + \mathbb{E}[\mathbb{I}] \sum_{\substack{i_1,\dots,i_k,j \ge 0 \\ i_1+\cdots+i_k+j=m}} \binom{m}{i_1,\dots,i_k,j} \mathbb{E}\big[\big(S_{n-1}^{(1)}\big)^{i_1}\big]\cdots\mathbb{E}\big[\big(S_{n-1}^{(k)}\big)^{i_k}\big] \\ &= (1-p)\mathbb{E}[S_{n-1}^m] + p \sum_{\substack{i_1,\dots,i_k,j \ge 0 \\ i_1+\cdots+i_k+j=m}} \binom{m}{i_1,\dots,i_k,j} \mathbb{E}\big[\big(S_{n-1}^{(1)}\big)^{i_1}\big]\cdots\mathbb{E}\big[\big(S_{n-1}^{(k)}\big)^{i_k}\big]. \end{aligned}$$

Scale the above recurrence by $((k-1)p+1)^{nm}$ to get

$$\mathbb{E}\Big[\Big(\frac{S_n}{((k-1)p+1)^n}\Big)^m\Big] = \frac{1-p}{((k-1)p+1)^m}\,\mathbb{E}\Big[\Big(\frac{S_{n-1}}{((k-1)p+1)^{n-1}}\Big)^m\Big] + \frac{p}{((k-1)p+1)^m}\sum_{\substack{i_1,\ldots,i_k,j\ge 0\\ i_1+\cdots+i_k+j=m}}\binom{m}{i_1,\ldots,i_k,j}\frac{1}{((k-1)p+1)^{j(n-1)}} \times \mathbb{E}\Big[\Big(\frac{S^{(1)}_{n-1}}{((k-1)p+1)^{n-1}}\Big)^{i_1}\Big]\cdots\mathbb{E}\Big[\Big(\frac{S^{(k)}_{n-1}}{((k-1)p+1)^{n-1}}\Big)^{i_k}\Big],$$

for $m\ge 2$. According to (2.3), $a_1 := \lim_{n\to\infty}\mathbb{E}[S_n]/((k-1)p+1)^n = 1/(k-1)$ exists. So, by induction on m, $a_m := \lim_{n\to\infty}\mathbb{E}[S_n^m]/((k-1)p+1)^{mn}$ exists. Taking limits, as $n\to\infty$, we get

$$\big((k-1)p+1\big)^m a_m = (1-p)a_m + p\sum_{\substack{i_1,\ldots,i_k\ge 0\\ i_1+\cdots+i_k=m}}\binom{m}{i_1,\ldots,i_k}a_{i_1}\cdots a_{i_k}.$$

Isolating the cases $i_j=m$, for $j=1,2,\ldots,k$, and rearranging, we get the claimed recurrence.

Now, by induction on m, we show that $a_m/m!\le 1/(k-1)$. This assertion is true for $m=1$. For $m\ge 2$ and $1\le i\le m-1$, we assume $a_i/i!\le 1/(k-1)$. From this assumption, we have

$$\begin{aligned}
\frac{a_m}{m!} &= \frac{p}{\big((k-1)p+1\big)^m-\big((k-1)p+1\big)}\sum_{\substack{i_1,\ldots,i_k\ge 0\\ i_1\ne m,\ldots,i_k\ne m\\ i_1+\cdots+i_k=m}}\frac{a_{i_1}}{i_1!}\cdots\frac{a_{i_k}}{i_k!}\\
&\le \frac{p}{\big(\sum_{j=0}^m\binom{m}{j}(k-1)^jp^j\big)-\big((k-1)p+1\big)}\sum_{\substack{i_1,\ldots,i_k\ge 0\\ i_1\ne m,\ldots,i_k\ne m\\ i_1+\cdots+i_k=m}}\frac{1}{k-1}\cdots\frac{1}{k-1}\\
&= \frac{\binom{m+k-1}{k-1}p-kp}{(m-1)(k-1)p+\sum_{j=2}^m\binom{m}{j}(k-1)^jp^j}\cdot\frac{1}{(k-1)^m}\\
&= \frac{(m+k-1)(m+k-2)\cdots(m+k-(m-1))k-km!}{(m-1)(k-1)+\sum_{j=2}^m\binom{m}{j}(k-1)^jp^{j-1}}\cdot\frac{1}{m!(k-1)^m}\\
&\le \frac{(m+k-1)(m+k-2)\cdots(m+k-(m-1))k-km!}{m!(m-1)(k-1)^m}\cdot\frac{1}{k-1}\\
&\le \frac{1}{k-1},
\end{aligned}\tag{2.8}$$

which completes the induction. Subsequently, we have

$$\Big|\sum_{m=0}^\infty\frac{a_m}{m!}z^m\Big|\le\sum_{m=0}^\infty\frac{a_m}{m!}|z|^m\le\frac{1}{k-1}\sum_{m=0}^\infty|z|^m\le\frac{1}{1-|z|}<\infty,$$

for all complex z, with $|z|<1$. Therefore, by Theorem 30.1 in [3], we conclude the uniqueness of the distribution with the moments a_m, $m\ge 1$. This verifies the convergence in distribution $S_n/((k-1)p+1)^n \xrightarrow{\mathcal{D}} \mathbf{S}$, for some random variable $\mathbf{S}$ with the moments $a_m=\mathbb{E}[\mathbf{S}^m]$, $m\ge 1$. Since $S_n/((k-1)p+1)^n \xrightarrow{\text{a.s.}} S_*$, so it converges in distribution to S_*. That is $\mathbf{S}\overset{d}{=}S_*$. Thus, S_* has the moments given by the stated recurrence. $\square$

3. Leaves

Let L_n be the number of leaves in an exponential k-ary tree of age n and index p. After one step of evolution, if the initial external node succeeds to recruit a leaf, then the extended exponential recursive k-ary tree has an internal node with k external nodes. In this case, for $i = 1, \dots, k$, let $\mathbb{J}_{n-1}^{(i)}$ be an indicator of the event that the ith external node does not progress into a tree of age $n-1$. By the nth step, these k external nodes produce k independent exponential recursive k-ary subtrees of age $n-1$. Let the number of leaves arising from these k subtrees be respectively $L_{n-1}^{(1)}$, $L_{n-1}^{(2)}$, $\dots$, $L_{n-1}^{(k)}$. These k random variables are independent and each is distributed like L_{n-1}. After n steps of evolution, the root of the tree is a leaf if and only if $\mathbb{J}_{n-1}^{(1)} \cdots \mathbb{J}_{n-1}^{(k)} = 1$. Alternatively, if the initial external node does not succeed to recruit a leaf, then the number of leaves in the exponential recursive k-ary tree at step n is the same as L_{n-1}. Hence, the number of leaves satisfies the distributional equation

$$L_n \overset{d}{=} (1-\mathbb{I})L_{n-1} + \mathbb{I}\big(L_{n-1}^{(1)} + L_{n-1}^{(2)} + \cdots + L_{n-1}^{(k)} + \mathbb{J}_{n-1}^{(1)} \cdots \mathbb{J}_{n-1}^{(k)}\big). \tag{3.1}$$

Note that, for each $n \ge 1$, the vectors $\big(L_{n-1}^{(i)}, \mathbb{J}_{n-1}^{(i)}\big)$, for $i = 1, \dots, k$, are independent. So, for $m \ge 1$, by raising both sides of the distributional equation (3.1) to the mth power, taking expectations, and observing independence and identical distribution in the subtrees, we have

$$\begin{aligned}
\mathbb{E}[L_n^m] &= (1-p)\mathbb{E}[L_{n-1}^m] + p \sum_{\substack{i_1,\dots,i_k,j\ge 0\\ i_1+\cdots+i_k+j=m}} \binom{m}{i_1,\dots,i_k,j} \mathbb{E}\big[\big(L_{n-1}^{(1)}\big)^{i_1}\big(\mathbb{J}_{n-1}^{(1)}\big)^{j}\big] \cdots \mathbb{E}\big[\big(L_{n-1}^{(k)}\big)^{i_k}\big(\mathbb{J}_{n-1}^{(k)}\big)^{j}\big]\\
&= (1-p)\mathbb{E}[L_{n-1}^m] + p \sum_{\substack{i_1,\dots,i_k\ge 0\\ i_1+\cdots+i_k=m}} \binom{m}{i_1,\dots,i_k} \mathbb{E}\big[\big(L_{n-1}^{(1)}\big)^{i_1}\big] \cdots \mathbb{E}\big[\big(L_{n-1}^{(k)}\big)^{i_k}\big] + p\mathbb{E}\big[\mathbb{J}_{n-1}^{(1)}\big]\cdots\mathbb{E}\big[\mathbb{J}_{n-1}^{(k)}\big]\\
&= (1-p)\mathbb{E}[L_{n-1}^m] + p \sum_{\substack{i_1,\dots,i_k\ge 0\\ i_1+\cdots+i_k=m}} \binom{m}{i_1,\dots,i_k} \mathbb{E}\big[\big(L_{n-1}^{(1)}\big)^{i_1}\big] \cdots \mathbb{E}\big[\big(L_{n-1}^{(k)}\big)^{i_k}\big] + p(1-p)^{k(n-1)}.
\end{aligned} \tag{3.2}$$

From (3.2), in particular, we can calculate the exact mean, and find its asymptotic equivalent.

Proposition 3.1. *Let L_n be the number of leaves in an exponential k-ary tree of age n and index p. Then we have*

$$\mathbb{E}[L_n] = \frac{p\,((k-1)p+1)^n - p(1-p)^{kn}}{(k-1)p+1-(1-p)^k} \sim \frac{p((k-1)p+1)^n}{(k-1)p+1-(1-p)^k}, \qquad \text{as } n\to\infty.$$

Proof. Take the expectation of (3.1), to write

$$\begin{aligned}
\mathbb{E}[L_n] &= (1-p)\mathbb{E}[L_{n-1}] + p\mathbb{E}\big[L_{n-1}^{(1)}\big] + \cdots + p\mathbb{E}\big[L_{n-1}^{(k)}\big] + p(1-p)^{k(n-1)}\\
&= ((k-1)p+1)\mathbb{E}[L_{n-1}] + p(1-p)^{k(n-1)}.
\end{aligned}$$

The solution to the recurrence of the mean is

$$\begin{aligned}
\mathbb{E}[L_n] &= p\sum_{j=0}^{n-1}(1-p)^{kj}((k-1)p+1)^{n-j-1}\\
&= p((k-1)p+1)^{n-1}\sum_{j=0}^{n-1}\left(\frac{(1-p)^k}{(k-1)p+1}\right)^j.
\end{aligned}$$

The sum is a geometric series, yielding

$$\mathbb{E}[L_n] = p((k-1)p+1)^{n-1} \cdot \frac{1-\left(\frac{(1-p)^k}{(k-1)p+1}\right)^n}{1-\frac{(1-p)^k}{(k-1)p+1}}.$$

This implies the assertion. □

The calculation techniques were taken from [1].

Theorem 3.2. *Let L_n be the number of leaves in an exponential recursive k-ary tree of age n and index p. We have the convergence in distribution*

$$\frac{L_n}{((k-1)p+1)^n} \xrightarrow{\mathcal{D}} L_*,$$

where the limiting random variable L_* has moments $b_m := \mathbb{E}[L_*^m]$ defined inductively by

$$b_m = \frac{p}{((k-1)p+1)^m - ((k-1)p+1)} \sum_{\substack{i_1,\dots,i_k \ge 0 \\ i_1 \ne m,\dots,i_k \ne m \\ i_1+\dots+i_k=m}} \binom{m}{i_1,\dots,i_k} b_{i_1}\cdots b_{i_k},$$

for $m \ge 2$ with $b_1 = p/((k-1)p+1-(1-p)^k)$.

Proof. Scale the recurrence (3.2) by $\big((k-1)p+1\big)^{nm}$ to get

$$\begin{aligned}\mathbb{E}\bigg[\bigg(\frac{L_n}{((k-1)p+1)^n}\bigg)^m\bigg] &= \frac{1-p}{((k-1)p+1)^m}\,\mathbb{E}\bigg[\bigg(\frac{L_{n-1}}{((k-1)p+1)^{n-1}}\bigg)^m\bigg] \\ &\quad + \frac{p}{((k-1)p+1)^m} \sum_{\substack{i_1,\dots,i_k\ge 0\\ i_1+\dots+i_k=m}} \binom{m}{i_1,\dots,i_k}\mathbb{E}\bigg[\bigg(\frac{L_{n-1}^{(1)}}{((k-1)p+1)^{n-1}}\bigg)^{i_1}\bigg] \\ &\quad \times\cdots\times\mathbb{E}\bigg[\bigg(\frac{L_{n-1}^{(k)}}{((k-1)p+1)^{n-1}}\bigg)^{i_k}\bigg] + \frac{p(1-p)^{k(n-1)}}{((k-1)p+1)^{nm}}.\end{aligned}$$

Thus, inductively the sequence of the limit b_m satisfies the recurrence

$$((k-1)p+1)^m b_m = (1-p)b_m + pkb_m + p \sum_{\substack{i_1,\dots,i_k \ge 0 \\ i_1 \ne m,\dots,i_k \ne m \\ i_1+\dots+i_k=m}} \binom{m}{i_1,\dots,i_k} b_{i_1}\cdots b_{i_k}, \qquad \text{for } m \ge 2.$$

Using this recurrence and $b_1 < 1/(k-1)$, by (2.8), we can conclude that $\frac{b_m}{m!} < 1/(k-1)$ for $m \ge 1$. So, for $|z| < 1$, the series $\sum_{m=0}^{\infty} \frac{b_m}{m!} z^m$ converges. Therefore, by Theorem 30.1 in [3], $L_n/((k-1)p+1)^n$ converges in distribution to a unique limit L_*. □

4. External path length

Let W_n be the external path length, which is the sum of the root-to-external node distances (the sum of the depth of external nodes). We use T_n for an exponential recursive k-ary tree of age n and index p. Assume, we number the external nodes of T_n arbitrarily with the numbers 1, ..., Y_n. Let D_i be the depth of external node i, for $i = 1, \dots, Y_n$. Then the external path length of T_n is

$$W_n = \sum_{i=1}^{Y_n} D_i.$$

Similar to the equation (2.6), for $n \geq 1$, W_n and Y_n satisfy the distributional equations

$$W_n \stackrel{d}{=} (1-\mathbb{I})W'_{n-1} + \mathbb{I}\big(W^{(1)}_{n-1} + \cdots + W^{(k)}_{n-1} + Y^{(1)}_{n-1} + \cdots + Y^{(k)}_{n-1}\big), \tag{4.1}$$

$$Y_n \stackrel{d}{=} (1-\mathbb{I})Y'_{n-1} + \mathbb{I}\big(Y^{(1)}_{n-1} + \cdots + Y^{(k)}_{n-1}\big), \qquad (W_0 = 0, Y_0 = 1). \tag{4.2}$$

Here $W^{(i)}_{n-1}$, $i = 0, 1, \dots, k$, and W'_{n-1} are independent copies of W_{n-1} and independent of $\mathbb{I}$; $Y^{(i)}_{n-1}$, $i = 0, 1, \dots, k$, and Y'_{n-1} are independent copies of Y_{n-1} and independent of $\mathbb{I}$.

4.1. The expectation and variance

The pair of equations (4.1) and (4.2) is sufficient to determine the means and the quadratic order moments exactly.

Theorem 4.1. *Let W_n be the external path length in an exponential k-ary tree of age n and index p. We have*

$$\mathbb{E}[W_n] = kpn\left((k-1)p+1\right)^{n-1}, \tag{4.3}$$
$$\mathrm{Var}[W_n] \sim k^2(k-1)p^2(1-p)n^2\left((k-1)p+1\right)^{2n-3}, \qquad \textit{as } n \to \infty.$$

Proof. Raise both sides of (4.2) to the second power. So we get

$$Y_n^2 \stackrel{d}{=} (1-\mathbb{I})Y'^2_{n-1} + \mathbb{I}\left(\sum_{j=1}^{k}\big(Y^{(j)}_{n-1}\big)^2 + \sum_{\substack{1\leq j_1,j_2\leq k\\ j_1\neq j_2}} Y^{(j_1)}_{n-1}Y^{(j_2)}_{n-1}\right). \tag{4.4}$$

Take the expectation of (4.4) and observe the independence $Y^{(i)}_{n-1}$ and $Y^{(j)}_{n-1}$ for $i \neq j$, to write

$$\begin{aligned}\mathbb{E}[Y_n^2] &= (1-p)\mathbb{E}[Y^2_{n-1}] + p\Big(k\mathbb{E}[Y^2_{n-1}] + k(k-1)\big(\mathbb{E}[Y_{n-1}]\big)^2\Big)\\ &= ((k-1)p+1)\mathbb{E}[Y^2_{n-1}] + k(k-1)p((k-1)p+1)^{2n-2},\end{aligned}$$

with initial condition $\mathbb{E}[Y_0^2] = 1$. Iterating this formula, we find

$$\begin{aligned}\mathbb{E}[Y_n^2] &= ((k-1)p+1)^n + k(k-1)p\sum_{j=0}^{n-1}((k-1)p+1)^{2n-j-2}\\ &= \Big(k((k-1)p+1)^{2n-1}\Big)(1+o(1)), \qquad \text{as } n\to\infty.\end{aligned} \tag{4.5}$$

Taking the expectation of (4.1), we get

$$\begin{aligned}\mathbb{E}[W_n] &= (1-p)\mathbb{E}[W_{n-1}] + kp\mathbb{E}[W_{n-1}] + kp\mathbb{E}[Y_{n-1}]\\ &= ((k-1)p+1)\mathbb{E}[W_{n-1}] + kp((k-1)p+1)^{n-1}\end{aligned}$$

with initial condition $\mathbb{E}[W_0] = 0$. By iterating, the mean value follows.

Then multiplying (4.1) and (4.2), we deduce that,

$$W_nY_n \stackrel{d}{=} (1-\mathbb{I})W'_{n-1}Y'_{n-1} + \mathbb{I}\Big(\sum_{j=1}^{k}\big(Y_{n-1}^{(j)}\big)^2 + \sum_{\substack{1\le j_1,j_2\le k\\ j_1\ne j_2}} Y_{n-1}^{(j_1)}Y_{n-1}^{(j_2)}\Big)$$
$$+\,\mathbb{I}\Big(\sum_{j=1}^{k} W_{n-1}^{(j)}Y_{n-1}^{(j)} + \sum_{\substack{1\le j_1,j_2\le k\\ j_1\ne j_2}} W_{n-1}^{(j_1)}Y_{n-1}^{(j_2)}\Big).$$

We next take the expectation and get the following recurrence for $\mathbb{E}[W_nY_n]$,

$$\begin{aligned}\mathbb{E}[W_nY_n] &= ((k-1)p+1)\mathbb{E}[W_{n-1}Y_{n-1}] + kp\mathbb{E}[Y_{n-1}^2] + k(k-1)p\Big(\mathbb{E}[W_{n-1}]\mathbb{E}[Y_{n-1}] + \big(\mathbb{E}[Y_{n-1}]\big)^2\Big)\\ &= ((k-1)p+1)\mathbb{E}[W_{n-1}Y_{n-1}] + kp((k-1)p+1)^{n-1} + k(k-1)p((k-1)p+1)^{2n-2}\\ &+ k^2(k-1)p^2(n-1)((k-1)p+1)^{2n-3} + k^2(k-1)p^2\sum_{i=0}^{n-2}((k-1)p+1)^{2n-i-4},\end{aligned}$$

with initial condition $\mathbb{E}[W_0Y_0] = 0$. This gives the solution

$$\begin{aligned}\mathbb{E}[W_nY_n] &= kpn((k-1)p+1)^{n-1} + k^2(k-1)p^2\sum_{j=0}^{n-1}(n-j-1)((k-1)p+1)^{2n-j-3}\\ &+ k(k-1)p\sum_{j=0}^{n-1}((k-1)p+1)^{2n-j-2} + k^2(k-1)p^2\sum_{j=0}^{n-1}\sum_{i=0}^{n-j-2}((k-1)p+1)^{2n-j-i-4}\\ &= \Big(k^2pn((k-1)p+1)^{2n-2}\Big)(1+o(1)), \qquad \text{as } n\to\infty. \end{aligned} \tag{4.6}$$

Now, we take the square of the equation (4.1). Then, by taking expectation, we have

$$\begin{aligned}\mathbb{E}[W_n^2] &= ((k-1)p+1)\mathbb{E}[W_{n-1}^2] + kp\mathbb{E}[Y_{n-1}^2] + k(k-1)p\big(\mathbb{E}[Y_{n-1}]\big)^2\\ &+ k(k-1)p\mathbb{E}[W_{n-1}]\mathbb{E}[Y_{n-1}] + k(k-1)p\big(\mathbb{E}[W_{n-1}]\big)^2 + kp\mathbb{E}[W_{n-1}Y_{n-1}].\end{aligned}$$

Again, by iterating, we obtain

$$\begin{aligned}\mathbb{E}[W_n^2] &= kpn((k-1)p+1)^{n-1} + k^2(k-1)p^2\sum_{l=0}^{n-1}\sum_{j=0}^{n-l-2}((k-1)p+1)^{2n-l-j-4}\\ &+ k(k-1)p\sum_{l=0}^{n-1}((k-1)p+1)^{2n-l-2} + k^2(k-1)p^2\sum_{l=0}^{n-1}(n-l-1)((k-1)p+1)^{2n-l-3}\\ &+ k^3(k-1)p^3\sum_{l=0}^{n-1}(n-l-1)^2((k-1)p+1)^{2n-l-4} + k^2p^2\sum_{l=0}^{n-1}(n-l-1)((k-1)p+1)^{n-l-2}\\ &+ k^2(k-1)p^2\sum_{l=0}^{n-1}\sum_{j=0}^{n-l-2}\Big(((k-1)p+1)^{2n-l-j-4} + kp\sum_{i=0}^{n-l-j-3}((k-1)p+1)^{2n-l-j-i-6}\Big)\end{aligned}$$

$$+ k^3(k-1)p^3 \sum_{l=0}^{n-1}\sum_{j=0}^{n-l-2}(n-l-j-2)((k-1)p+1)^{2n-l-j-5}$$
$$= \Big(k^3p^2n^2((k-1)p+1)^{2n-3}\Big)(1+o(1)), \qquad \text{as } n\to\infty. \tag{4.7}$$

This implies the claim for the variance of W_n. □

4.2. Characterizing the limiting distribution

Here, we use the contraction method by Roesler and Neininger to state the limiting distribution of $\hat{W}_n := W_n/kpn\,((k-1)p+1)^n$. Recently, this method has been applied to characterize the limiting distribution of the protected node path length in [6]. By equation (4.1), we have

$$\frac{W_n}{kpn\big((k-1)p+1\big)^{n-1}} \overset{d}{=} (1-\mathbb{I})\cdot\frac{n-1}{n}\cdot\frac{1}{(k-1)p+1}\cdot\frac{W'_{n-1}}{kp(n-1)\big((k-1)p+1\big)^{n-2}}$$
$$+\,\mathbb{I}\cdot\frac{n-1}{n}\cdot\frac{1}{(k-1)p+1}\cdot\left(\frac{W^{(1)}_{n-1}}{kp(n-1)\big((k-1)p+1\big)^{n-2}}+\cdots+\frac{W^{(k)}_{n-1}}{kp(n-1)\big((k-1)p+1\big)^{n-2}}\right)$$
$$+\,\mathbb{I}\cdot\frac{1}{kpn}\cdot\left(\frac{Y^{(1)}_{n-1}}{\big((k-1)p+1\big)^{n-1}}+\cdots+\frac{Y^{(k)}_{n-1}}{\big((k-1)p+1\big)^{n-1}}\right)$$

All the conditions of the contraction method are satisfied (Thm. 3 of [9], page 8), then

Theorem 4.2. *The scaled external path length $\hat{W}_n := W_n/kpn\,((k-1)p+1)^n$ converges in distribution to some random variable $\hat{W}$ with distribution, the unique solution of the following equation*

$$\hat{W} \overset{d}{=} \frac{1-\mathbb{I}}{(k-1)p+1}\hat{W}^{(0)} + \frac{\mathbb{I}}{(k-1)p+1}\big(\hat{W}^{(1)}+\cdots+\hat{W}^{(k)}\big), \tag{4.8}$$

where $\hat{W}^{(i)}$, $i=0,1,\ldots,k$, are independent copies of $\hat{W}$ and independent of $\mathbb{I}$.

Proof. The proof is based on the Lemma 3.1 page 502 of the paper [8]. Let $\mathcal{M}_2$ be the space of all distributions with finite absolute second moment.

Consider the transformation $T:\mathcal{M}_2\to\mathcal{M}_2$ defined by

$$T(\mu) \overset{d}{=} \frac{1-\mathbb{I}}{(k-1)p+1}\hat{W}^{(0)} + \frac{\mathbb{I}}{(k-1)p+1}\big(\hat{W}^{(1)}+\cdots+\hat{W}^{(k)}\big), \tag{4.9}$$

where $\hat{W}^{(i)}$, $i=0,1,\ldots,k$ and $\mathbb{I}$ are independent and $\hat{W}^{(i)}$ have μ as distribution.

At a first step we have to prove that with respect to the metric $\mathcal{L}_2^2$ defined on $\mathcal{M}_2$ by

$$\mathcal{L}_2^2\,(\mu,\nu) := \min\left\{\mathbb{E}[|X-Y|^2],\ \text{where } X,Y \text{ have } \mu \text{ and } \nu \text{ as distribution, respectively}\right\},$$

the transformation have a unique fixed point. In fact let μ and ν be two measures of $\mathcal{M}_2$

$$T(\mu) \overset{d}{=} \frac{1-\mathbb{I}}{(k-1)p+1}\hat{W}^{(0)}_\mu + \frac{\mathbb{I}}{(k-1)p+1}\big(\hat{W}^{(1)}_\mu+\cdots+\hat{W}^{(k)}_\mu\big),$$
$$T(\nu) \overset{d}{=} \frac{1-\mathbb{I}}{(k-1)p+1}\hat{W}^{(0)}_\nu + \frac{\mathbb{I}}{(k-1)p+1}\big(\hat{W}^{(1)}_\nu+\cdots+\hat{W}^{(k)}_\nu\big),$$

$$\begin{aligned}\mathcal{L}_2^2\left(T(\mu),\, T(\nu)\right) &\le \mathbb{E}\left[\left|T(\mu)-T(\nu)\right|^2\right]\\ &= \frac{k\,\mathbb{E}[\mathbb{I}^2]}{\big((k-1)p+1\big)^2}\,\mathbb{E}\big[\hat W_\mu-\hat W_\nu\big]^2+\frac{\mathbb{E}\left[(1-\mathbb{I})^2\right]}{\big((k-1)p+1\big)^2}\,\mathbb{E}\big[\hat W_\mu-\hat W_\nu\big]^2\\ &= \frac{1}{(k-1)p+1}\,\mathbb{E}\big[\hat W_\mu-\hat W_\nu\big]^2=\frac{1}{(k-1)p+1}\,\mathcal{L}_2^2\left(\mu,\,\nu\right).\end{aligned}$$

Since $1/\big((k-1)p+1\big)<1$, then T is a contracted transform and it has only one fixed point.

Now, if $\hat W_n$ converges in distribution, the limit will be the unique fixed point of the transformation T. Namely, we have to prove that

$$\lim_{n\to\infty}\mathcal{L}_2^2\big(\hat W,\,\hat W_n\big)=0,$$

to conclude that $\hat W_n$ converges in distribution and in L_2 to $\hat W$. We have

$$\begin{aligned}\mathcal{L}_2^2\big(\hat W,\,\hat W_n\big) \le{}& \frac{p}{\big((k-1)p+1\big)^2}\left(\left|\hat W^{(1)}-\hat W^{(1)}_{n-1}\,\frac{n-1}{n}\right|_2^2+\cdots+\left|\hat W^{(k)}-\hat W^{(k)}_{n-1}\,\frac{n-1}{n}\right|_2^2\right)\\ &+\frac{1-p}{\big((k-1)p+1\big)^2}\left|\hat W^{(0)}-\hat W'_{n-1}\,\frac{n-1}{n}\right|_2^2\\ &+\frac{p}{k^2p^2n^2}\left|\frac{Y^{(1)}_{n-1}}{\big((k-1)p+1\big)^{n-1}}+\cdots+\frac{Y^{(k)}_{n-1}}{\big((k-1)p+1\big)^{n-1}}\right|_2^2\\ &+\frac{2}{k\big((k-1)p+1\big)n}\sum_{j=1}^{k}\sum_{i=1}^{k}\mathbb{E}\left[\frac{Y^{(i)}_{n-1}}{\big((k-1)p+1\big)^{n-1}}\left|\hat W^{(j)}-\hat W^{(j)}_{n-1}\,\frac{n-1}{n}\right|\right].\end{aligned}$$

Then we deduce

$$\lim_{n\to\infty}\mathcal{L}_2^2\big(\hat W,\,\hat W_n\big)\le\frac{1}{(k-1)p+1}\lim_{n\to\infty}\mathcal{L}_2^2\big(\hat W,\,\hat W_{n-1}\big)<\lim_{n\to\infty}\mathcal{L}_2^2\big(\hat W,\,\hat W_n\big).$$

This implies $\lim_{n\to\infty}\mathcal{L}_2^2\big(\hat W,\,\hat W_n\big)=0$. □

4.3. Almost sure convergence

In this subsection, the calculation techniques were taken from [1].

Theorem 4.3. *Let W_n be the external path length in an exponential k-ary tree of age n and index p. As $n\to\infty$, we have the almost sure and L^2 convergence,*

$$\frac{W_n}{n\big((k-1)p+1\big)^n}\xrightarrow{L^2\ \&\ \text{a.s.}}Y_*.$$

Proof. Let $\mathbb{B}_k$ be a sequence of independent Bernoulli (p) random variables, defined on the same probability space as the trees. Consider $\left\{\mathbb{B}_k\right\}_{k=1}^{\infty}$ to be independent of the structure of the tree at any age, too. If the ith external node recruits a leaf, it is converted into an internal node (a leaf node) at its own level and it adds k external nodes at the next level (as the children of the new leaf). The contribution of the ith external node of

T_{n-1} into W_n, is $D_i + (k(D_i+1) - D_i)\mathbb{B}_i$, and we write

$$W_n = \sum_{i=1}^{Y_{n-1}} \Big(D_i + \big(k(D_i+1) - D_i\big)\mathbb{B}_i\Big).$$

Taking an expectation conditioned of $\mathcal{F}_{n-1}$, we obtain

$$\mathbb{E}[W_n|\mathcal{F}_{n-1}] = \sum_{i=1}^{Y_{n-1}} D_i + p\sum_{i=1}^{Y_{n-1}} \big(k(D_i+1) - D_i\big) = \big((k-1)p+1\big)W_{n-1} + kpY_{n-1}. \tag{4.10}$$

Let $M_n := \alpha_n W_n + \beta_n Y_n$, for specified factors α_n and β_n that transformed M_n a martingale. So, by (2.2), (4.10) and a martingalization procedure, we have

$$\begin{aligned}\mathbb{E}[M_n|\mathcal{F}_{n-1}] &= \alpha_n\big((k-1)p+1\big)W_{n-1} + kp\alpha_n Y_{n-1} + \beta_n\big((k-1)p+1\big)Y_{n-1}\\ &= M_{n-1} = \alpha_{n-1}W_{n-1} + \beta_{n-1}Y_{n-1}.\end{aligned}$$

This is possible, if we choose

$$\alpha_{n-1} = \alpha_n\big((k-1)p+1\big), \qquad \beta_{n-1} = kp\alpha_n + \beta_n\big((k-1)p+1\big).$$

From these solvable recurrences, we obtain

$$\alpha_n = \frac{\alpha_0}{((k-1)p+1)^n}, \qquad \beta_n = \frac{\beta_0}{((k-1)p+1)^n} - \frac{kp\alpha_0 n}{((k-1)p+1)^{n+1}}.$$

For simplicity, we take $\alpha_0 = 1$ and $\beta_0 = 0$. Thus, by (4.3), (4.5), (4.6) and (4.7),

$$M_n = \frac{1}{((k-1)p+1)^n}\,W_n - \frac{kpn}{((k-1)p+1)^{n+1}}\,Y_n,$$

is a martingale with bounded moments

$$\begin{aligned}\mathbb{E}[M_n] &= \frac{1}{((k-1)p+1)^n}\,\mathbb{E}[W_n] - \frac{kpn}{((k-1)p+1)^{n+1}}\,\mathbb{E}[Y_n]\\ &= \frac{kpn\,((k-1)p+1)^{n-1}}{((k-1)p+1)^n} - \frac{kpn\,((k-1)p+1)^n}{((k-1)p+1)^{n+1}} = \frac{kpn}{(k-1)p+1} - \frac{kpn}{(k-1)p+1} = 0 < \infty,\\ \mathbb{E}[M_n^2] &= \frac{1}{((k-1)p+1)^{2n}}\,\mathbb{E}[W_n^2] + \frac{k^2p^2n^2}{((k-1)p+1)^{2n+2}}\,\mathbb{E}[Y_n^2] - 2\frac{kpn}{((k-1)p+1)^{2n+1}}\,\mathbb{E}[W_nY_n]\\ &= k^3p^2n^2\bigg(\frac{((k-1)p+1)^{2n-3}}{((k-1)p+1)^{2n}} + \frac{((k-1)p+1)^{2n-1}}{((k-1)p+1)^{2n+2}} - 2\frac{((k-1)p+1)^{2n-2}}{((k-1)p+1)^{2n+1}}\bigg) + o(1)\\ &= \bigg(\frac{k^3p^2n^2}{((k-1)p+1)^3} + \frac{k^3p^2n^2}{((k-1)p+1)^3} - 2\frac{k^3p^2n^2}{((k-1)p+1)^3}\bigg) + o(1) = 0 + o(1) < \infty.\end{aligned}$$

So, by the martingale convergence theorem (Thm. 6.6.9 in [2]), M_n converges to some M_* almost surely and in L^2. Consequently,

$$\frac{M_n}{n} = \frac{1}{n\left((k-1)p+1\right)^n} W_n - \frac{kp}{\left((k-1)p+1\right)^{n+1}} Y_n \xrightarrow{L^2 \text{ \& a.s.}} 0.$$

Hence, by (2.4), the result follows. □

Corollary 4.4. *Let D_n^* be the depth of a randomly chosen external nodes in an exponential k-ary tree of age n and index p. Then we have*

$$\frac{\mathbb{E}[D_n^*]}{n} \longrightarrow \frac{kp}{(k-1)p+1}.$$

Proof. In T_n, we have

$$\mathbb{E}[D_n^*|T_n] = \frac{D_1 + D_2 + \cdots + D_{Y_n}}{Y_n} = \frac{W_n}{Y_n}.$$

Therefore, by Theorem (4.3), we have

$$\frac{\mathbb{E}[D_n^*|T_n]}{n} = \frac{W_n/\left(n((k-1)p+1)^n\right)}{Y_n/\left(((k-1)p+1)^n\right)} \xrightarrow{\text{a.s.}} \frac{kp}{(k-1)p+1}.$$

Since $W_n \le nY_n$, then $\mathbb{E}[D_n^*|T_n]/n \le 1$. Consequently,

$$\lim_{n\to\infty} \frac{\mathbb{E}[D_n^*]}{n} = \lim_{n\to\infty} \mathbb{E}\left[\frac{\mathbb{E}[D_n^*|T_n]}{n}\right] = \mathbb{E}\left[\lim_{n\to\infty} \frac{\mathbb{E}[D_n^*|T_n]}{n}\right]$$

yields the result. □

References

[1] R. Aguech, S. Bose, H.M. Mahmoud and Y. Zhang, Some properties of exponential trees. *Methodol. Comput. Appl. Probab.* **7** (2021) 16–32.

[2] R.B. Ash, Probability and Measure Theory, Second Edition. Academic Press, New York (1999).

[3] P. Billingsley, Probability and Measure, Anniversary Ed. Wiley, Hoboken, New Jersey (2012).

[4] Y. Feng and H.M. Mahmoud, Profile of random exponential binary trees. *Methodol Comput. Appl. Probab.* **20** (2018) 575–587.

[5] M. Javanian and M.Q. Vahidi-Asl, Depth of nodes in random recursive k-ary trees. *Inf. Process. Lett.* **98** (2006) 115–118.

[6] M. Javanian and R. Aguech, On the protected nodes in exponential recursive trees. *Discr. Math. Theor. Comput. Sci.* (2023) submitted.

[7] H.M. Mahmoud, Profile of random exponential recursive trees. *Methodol. Comput. Appl. Probab.* (2021). DOI: 10.1007/s11009-020-09831-9.

[8] R. Neininger, On a multivariate contraction method for random recursive structures with applications to quicksort. *Random Struct. Algor.* **19** (2001) 498–524.

[9] U. Roesler and L. Rueschendorf, The contraction method for recursive algorithms. *Algorithmica* **29** (2001) 3–33.

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