On the protected nodes in exponential recursive trees

Mehri Javanian\textsuperscript{1} \quad Rafik Aguech\textsuperscript{2}

\textsuperscript{1}Department of Statistics, university of Zanjan, Iran
\textsuperscript{2}Department of mathematics, University of Monastir, Tunisia

revisions 21\textsuperscript{st} Dec. 2022, 12\textsuperscript{th} June 2023, 18\textsuperscript{th} Aug. 2023, 30\textsuperscript{th} Aug. 2023; accepted 5\textsuperscript{th} Sep. 2023.

The exponential recursive trees model several kinds of networks. At each step of growing of these trees, each node independently attracts a new node with probability \( p \), or fails to do with probability \( 1 - p \). Here, we investigate the number of protected nodes, total path length of protected nodes, and a mean study of the protected node profile of such trees.

Keywords: exponential recursive trees, protected nodes, limiting distribution, Wasserstein metric, contraction method

1 Introduction

A social networking site is an Internet-based platform which people use to build social relationships with friends, family, colleagues, customers, or clients. Social networking can have a social purpose, a business purpose, or both, through sites like Facebook, Twitter, LinkedIn, Instagram, TikTok, Snapchat, Pinterest, Reddit, Tumblr, Telegram, WhatsApp and YouTube. These sites allow people and corporations to connect with one another so they can develop relationships very quickly and so they can share information, ideas, and messages. Social networking has become a significant base for marketers seeking to engage customers. In order to model certain aspects of fast-growing networks, the references Feng and Mahmoud (2018) and Mahmoud (2022) introduce exponential binary trees and exponential recursive trees, respectively.

A rooted tree grown on \( n + 1 \) nodes labeled distinctly with the numbers \( 1, 2, \ldots, n, n + 1 \); is a recursive tree of age \( n \), \( T_n \), that is built by attaching, at the \( n \)th step, the new node \( n + 1 \) to one node of a recursive tree of age \( n - 1 \), \( T_{n-1} \), according to some distribution on the set \( \{1, \ldots, n\} \) and independently of the structure of the tree \( T_{n-1} \). As an example, Figure 1 shows a sequence of recursive trees growing from a single node labeled with 1, \( T_0 \), into a tree of age 8, \( T_8 \), in 8 steps. See a survey of results of recursive trees in Smythe and Mahmoud (1995).

The recursive trees are slow-growing where one node is added at each step. So, the recursive tree models cannot be suitable for fast-growing phenomena (e.g., the Corona virus spreads very quickly from
In Mahmoud (2022), a fast-growing analogue of recursive trees has been defined as follows: Initially, $T_0$ is a tree of a single root node. For $n \geq 1$, at the $n$th step, a tree of age $n$, $T_n$, is constructed when every node of the tree of age $n-1$, $T_{n-1}$, independently attracts a child with probability $p \in (0, 1)$, or not to attract with probability $q := 1 - p$. After the $n$th step, the obtained tree is called exponential recursive tree (ERT) of index $p$. The trees in Figure 2(a) illustrate a sequence of ERT of index $p$ growing from $T_0$ into $T_7$, a tree of probability $p^{19}q^{14}$ in the 7th step.

By a protected node in a rooted tree, we mean a node that is not a leaf and not all of its children are leaves, for instance see Figure 2(b). For many types of random trees, protected nodes have been investigated, see e.g. Devroye and Janson (2014), Fuchs et al. (2016), Javanian et al. (2022), Mahmoud and Ward (2015). In this paper, we study the number of protected nodes in exponential recursive trees (ERT). Here, we present the asymptotic expectation, variance and characterizing of the limiting distribution of the number of protected nodes. Via contraction method, we also show the convergence in distribution for the sum of depths of all protected nodes, i.e., the total path length of all protected nodes in ERT of index $p$. Finally, we derive the expectation of protected node profile, i.e., the number of protected nodes at the same level.

2 Setting and Preliminary Lemmas

Let $T_n$ be an exponential recursive tree of age $n$ and index $p$. Define

- $X_n :=$ the number of protected nodes in $T_n$,
- $R_n :=$ the event that the root of $T_n$ is protected,
- $|T_n| :=$ the number of nodes in $T_n$ (the size of $T_n$).

At the first step, if the root of $T_0$ fails to attract a child, then $T_1$ will be a root and after $n - 1$ steps the root of $T_1$ will produce an exponential recursive tree $T_{n-1}$. Alternatively, if the root of $T_0$ attracts a child,
then, in \( n - 1 \) steps, the child and the root will independently develop two exponential recursive trees \( T_{n-1}' \) and \( T_{n-1}'' \), respectively. Let us indicate the number of protected nodes in \( T_{n-1}' \), \( T_{n-1}'' \) and \( T_{n-1}''' \) by \( X_{n-1}' \), \( X_{n-1}'' \) and \( X_{n-1}''' \), respectively. Therefore, \( X_{n-1}' \), \( X_{n-1}'' \) and \( X_{n-1}''' \) are independent copies of \( X_{n-1} \). We denote the events that the trees \( T_{n-1}' \), \( T_{n-1}'' \) and \( T_{n-1}''' \) have protected roots by \( R_{n-1}' \), \( R_{n-1}'' \) and \( R_{n-1}''' \), respectively.

If \( \mathbb{1}_A \) denotes the indicator function of an event \( A \), then we define

\[
\mathbb{I} := \mathbb{1}_{\{|T_1|=2\}}, \quad \mathbb{J}_n := \mathbb{1}_{\{|T_n'''|=1\}\cap\{|T_n''|\geq 2\}}, \quad \mathbb{G}_n := \mathbb{1}_{\{|T_n''|=1\}\cap R_{n-1}''}. \tag{1}
\]

The proof of our results requires to prove the following Lemmas.

**Lemma 2.1** By the above setting, for \( n \geq 2 \), the probability of \( R_n \) is

\[
\mathbb{P}(R_n) = \sum_{k=1}^{n-1} pq^k (1 - q^k) \prod_{i=k+1}^{n-1} (1 - pq^i). \tag{2}
\]

**Proof:** We can observe that

\[
R_n = \left[ \{|T_{n-1}''| \geq 2\} \cap R_{n-1}'' \right] \cup \left[ \{ |T_1| = 2 \} \right] \cup \left[ \{ |T_{n-1}'''| = 1 \} \cap \{ |T_1| = 1 \} \right]
\]

\[
\cup \left[ \{|T_{n-1}'''| \geq 2\} \cap \{ |T_{n-1}'''| = 1 \} \cap \{ |T_1| = 2 \} \right].
\]
Fig. 3: Graphs of $\mathbb{P}(R_n)$ versus $p$, for $n = 2, \ldots, 20$ and $n = 200$. For each $p$, $\mathbb{P}(R_n)$ is increasing in $n$. E.g., the bottom and top curves are the graphs of $\mathbb{P}(R_2)$ and $\mathbb{P}(R_{20})$.

is credible for $n \geq 2$. Since $\mathbb{P}(|T'''_n| \geq 2) = 1 - \mathbb{P}(|T''_n| = 1) = 1 - q^n$ and the trees $T''_{n-1}$ and $T'''_{n-1}$ are developed independently, then

$$\mathbb{P}(R_n) = p(1 - q^{n-1})\mathbb{P}(R''_{n-1}) + q\mathbb{P}(R'_{n-1}) + pq^{n-1}(1 - q^{n-1})$$

Iterating the recurrence we obtain the claim. See Figure 3 as a plot for $\mathbb{P}(R_n)$.

Lemma 2.2 By the above setting, we have the distributional equation

$$X_n \overset{d}{=} X'_{n-1}(1 - \mathbb{I}) + (X''_{n-1} + X'''_{n-1} + \mathbb{J}_{n-1} - \mathbb{G}_{n-1})\mathbb{I},$$

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

Proof: In order to construct an exponential recursive tree $T_n$, a tree distributed like $T_n$, we can attach the root of the tree $T'''_{n-1}$ to the root of the tree $T''_{n-1}$ by adding an edge. Consequently, if $|T''_{n-1}| \geq 2$ and $T'''_{n-1}$ is a leaf, then the number of protected nodes in $T_n$ is increased by 1; as the attaching causes the root of $T_{n-1}$ to be protected. From the other point of view, if $|T''_{n-1}| = 1$ and the root of $T'''_{n-1}$ is protected, then this protection is lost after the attaching. Hence we obtain the assertion by \cite{1}. $\square$
Lemma 2.3 For \( m \geq 1 \), we have the following recurrences for \( \mathbb{E}[X^m_n] \) and \( \mathbb{E}[X^m_n \mathbb{1}_{\mathcal{R}_n}] \):

\[
\mathbb{E}[X^m_n] = (p + 1)\mathbb{E}[X^m_{n-1}] + p \sum_{k=1}^{m-1} \binom{m}{k} \mathbb{E}[X^k_{n-1}] \mathbb{E}[X^{m-k}_{n-1}] + pq^{n-1} \sum_{k=1}^{m-1} \binom{m}{k} \mathbb{E}[X^k_{n-1}]
\]

\[
+ pq^{n-1} \sum_{k=1}^{m-1} \binom{m}{k} (-1)^{m-k} \mathbb{E}[X^k_{n-1} \mathbb{1}_{\mathcal{R}_{n-1}}] + pq^{n-1}(1 - q^{n-1} + (-1)^m \mathbb{E}(\mathcal{R}_{n-1}))
\]

\[
\mathbb{E}[X^m_n \mathbb{1}_{\mathcal{R}_n}] = (1 - pq^{n-1}) \mathbb{E}[X^m_{n-1} \mathbb{1}_{\mathcal{R}_{n-1}}] + p \sum_{k=1}^{m} \binom{m}{k} \mathbb{E}[X^k_{n-1}] \mathbb{E}[X^{m-k}_{n-1} \mathbb{1}_{\mathcal{R}_{n-1}}]
\]

\[
+ pq^{n-1} \sum_{k=1}^{m} \binom{m}{k} \mathbb{E}[X^k_{n-1}] + pq^{n-1}(1 - q^{n-1}).
\]

Proof: Raise both sides of (3) to the \( m \)th power. So we get

\[
X^m_n \overset{d}{=} (X'_{n-1})^m (1 - \mathbb{1}) + \sum_{i,j,k,l \geq 0} \binom{m}{i,j,k,l} (X''_{n-1})^i (X'''_{n-1})^j \mathbb{1}_n^{k} (1 - \mathcal{G}_{n-1})^l \mathbb{1}.
\]

Now, separate the second term with \((k = 0, l \geq 1), (k \geq 1, l = 0)\) and \((k = 0, l = 0)\). Then

\[
X^m_n \overset{d}{=} (X'_{n-1})^m (1 - \mathbb{1}) + \sum_{i,j,l \geq 0, i \geq 1} \binom{m}{i,j,l} (X''_{n-1})^i (X'''_{n-1})^j \mathbb{1}_{n}^{l} (1 - \mathcal{G}_{n-1})^i \mathbb{1} \mathbb{1} + \sum_{i,j \geq 0} \binom{m}{i,j} (X''_{n-1})^i (X'''_{n-1})^j \mathbb{1} \mathbb{1} \mathbb{1}.
\]

Note that, for \( i, j, k, l \geq 1 \),

\[
(\mathcal{G}_{n-1})^i = \mathcal{G}_{n-1} = \mathbb{1}_{|T''_{n-1}|=1} \mathbb{1}_{\mathcal{R}_{n-1}}, \quad (X''_{n-1})^i \mathbb{1}_{|T''_{n-1}|=1} = 0
\]

\[
(\mathbb{1}_{n})^k = \mathbb{1}_{n} = \mathbb{1}_{|T''_{n-1}| \geq 2} \mathbb{1}_{|T'''_{n-1}|=1}, \quad (X'''_{n-1})^j \mathbb{1}_{|T'''_{n-1}|=1} = 0.
\]

Therefore, the equation (4) can be simplified as

\[
X^m_n \overset{d}{=} (X'_{n-1})^m (1 - \mathbb{1}) + \sum_{j=0}^{m-1} \binom{m}{j} (X''_{n-1})^j (-1)^{m-j} \mathbb{1}_{|T''_{n-1}|=1} \mathbb{1}_{\mathcal{R}_{n-1}} \mathbb{1} \mathbb{1} \mathbb{1} + \sum_{i=0}^{m-1} \binom{m}{i} (X''_{n-1})^i (X'''_{n-1})^{m-i} \mathbb{1} \mathbb{1} \mathbb{1}.
\]
In this section, we obtain the expectation and variance of $X$. Taking the expectation of this equation, it implies the second recurrence of the Lemma. Let $T$ be $X$.

This yields the first recurrence of the Lemma, by the identical distribution in the subtrees and $T$.

By the definitions of the subtrees $T', T'', T'''$, we observe that

Taking the expectation of this equation, it implies the second recurrence of the Lemma.

3 The Expectation and Variance

In this section, we obtain the expectation and variance of $X_n$, using Lemma 2.1 and Lemma 2.3.

**Theorem 3.1** Let $X_n$ be the number of protected nodes in an exponential recursive tree of age $n$ and index $p$. For $q := 1 - p$,

$$E[X_n] = \frac{p}{q} \left( \sum_{j=2}^{n} (p+1)^{-j} q^j \left( 1 - q^{j-1} - \sum_{k=1}^{j-2} pq^k (1 - q^k) \prod_{i=k+1}^{j-2} (1 - pq^i) \right) \right) (p+1)^n$$

$$\mu_{n,p} = (p+1)^n, \quad (\mu_{n,p} < 1 \text{ for } n \geq 1 \text{ and } p \in (0,1)).$$

See Figure 4 for some graphs of $\mu_{n,p}$. For $n \geq 1, \mu_{n,p} < \mu_{n+1,p}$. So $\mu_p := \lim_{n \to \infty} \mu_{n,p}$ exists.

**Proof:** In Lemma 2.3, set $m = 1$ to obtain the recurrence

$$E[X_n] = (p+1)E[X_{n-1}] + pq^{n-1}(1 - q^{n-1}) - pq^{n-1} P(R_{n-1}),$$

with initial condition $E[X_1] = 0$. Iterating this formula, we find

$$E[X_n] = \sum_{j=0}^{n-2} (p+1)^j pq^{n-j-1} (1 - q^{n-j-1} - P(R_{n-j-1}))$$

$$= \frac{p}{q} \left( \sum_{j=2}^{n} (p+1)^{-j} q^j \left( 1 - q^{j-1} - P(R_{j-1}) \right) \right) (p+1)^n \quad (7)$$
Substituting $\mathbb{P}(R_{j-1})$ from (2) into (7), we get the exact solution in the Theorem.

In the second moment recurrence, we need $E[X_n \mathbb{1}_{R_n}]$ that is given in the following lemma.

**Lemma 3.1** Let $Y_n := X_n \mathbb{1}_{R_n}$, then

$$E[Y_n] = \left( \sum_{k=1}^{n-1} \mu_{n-k,p}(p+1)^{-k}(q^{n-k} + \mathbb{P}(R_{n-k})) \prod_{i=n-k+1}^{n-1} (1 - pq^i) \right) (p+1)^n + \mathbb{P}(R_n)$$

$$=: \alpha_n(p+1)^n + \mathbb{P}(R_n) = \alpha_p(p+1)^n + O(1), \quad (8)$$

where $\alpha_p := \lim_{n \to \infty} \alpha_{n,p}$ (see Figure 5) and the functions $\mathbb{P}(R_n)$ and $\mu_{n,p}$ are obtained in Lemma 2.1 and Theorem 3.1 respectively.

**Proof:** In Lemma 2.3 set $m = 1$ to find the recurrence

$$E[Y_n] = (1 - pq^{n-1})E[Y_{n-1}] + p(q^{n-1} + \mathbb{P}(R_{n-1}))E[X_{n-1}] + pq^{n-1}(1 - q^{n-1})$$

with initial condition $E[Y_1] = 0$. Iterating this formula, we obtain

$$E[Y_n] = \sum_{k=1}^{n-1} \left( p(q^k + \mathbb{P}(R_k))E[X_k] + pq^k(1 - q^k) \right) \prod_{i=k+1}^{n-1} (1 - pq^i)$$

$$= p \sum_{k=1}^{n-1} \mu_{k,p}(p+1)^k(q^k + \mathbb{P}(R_k)) \prod_{i=k+1}^{n-1} (1 - pq^i) + \mathbb{P}(R_n).$$

This implies the claim.\qed
Theorem 3.2 By the above definitions for $\mathbb{P}(\mathcal{R}_n)$, $\mu_{n,p}$ and $\alpha_{n,p}$, we have

$$E[X_n^2] = \sum_{j=1}^{n-1} \left( 2p\mu_{j,p}^2(p+1)^{-n+j-1} + 2pq^j(\mu_{j,p} - \alpha_{j,p})(p+1)^{-n-1} ight.$$
\[+ pq^j(1 - q^j - \mathbb{P}(\mathcal{R}_j))(p+1)^{-n-j-1} \left. \right)(p+1)^{2n} =: \beta_{n,p}(p+1)^{2n}. \]

See Figure 5 for some graphs of $\beta_{n,p}$ with $\beta_p := \lim_{n \to \infty} \beta_{n,p}$.

Proof: In Lemma 2.3 set $m = 2$ to find the recurrence

$$E[X_n^2] = (p+1)E[X_{n-1}^2] + 2pE[X_{n-1}^2] + 2pq^{n-1}E[X_{n-1}]$$
\[- 2pq^{n-1}E[X_{n-1} + p(1 - q^{n-1} + \mathbb{P}(\mathcal{R}_{n-1})), \]

with boundary condition $E[X_1^2] = 0$. This standard linear recurrence has the solution

$$E[X_n^2] = \sum_{j=0}^{n-2} (p+1)^j \left( 2pE^2[X_{n-j-1}] + 2pq^{n-j-1}E[X_{n-j-1}] \right.$$
\[+ 2pq^{n-j-1}E[X_{n-j-1}] + p(1 - q^{n-j-1} - \mathbb{P}(\mathcal{R}_{n-j-1})) \left. \right). \] (9)

After substituting (2), (6), (8) in (9), we have

$$E[X_n^2] = \sum_{j=0}^{n-2} \left( 2p\mu_{n-j-1,p}^2(p+1)^{-n-j-2} + 2pq^{n-j-1}(\mu_{n-j-1,p} - \alpha_{n-j-1,p})(p+1)^{-n-1} \right.$$
\[+ pq^{n-j-1}(p+1)^{-2n+j}(1 - q^{n-j-1} - \mathbb{P}(\mathcal{R}_{n-j-1})) \left. \right)(p+1)^{2n}. \]
protected nodes of exponential trees

This yields the claimed result by changing \( n - j - 1 \to j \) in the range of the above sum.

**Corollary 3.1** By the above definitions for \( \mu_{n,p} \) and \( \beta_{n,p} \), we have

\[
\text{Var}[X_n] = (\beta_{n,p} - \mu_{n,p}^2)(p + 1)^{2n} =: \sigma_{n,p}(p + 1)^{2n}.
\]

See Figure 6 for some graphs of \( \sigma_{n,p} \) with \( \lim_{n \to \infty} \sigma_{n,p} =: \sigma_p = \beta_p - \mu_p^2 \).

**Proof:** Using \( \text{Var}[X_n] = \mathbb{E}[X_n^2] - \mathbb{E}^2[X_n] \) and Theorem 3.2, the proof is straightforward.

## 4 Convergence in Distribution

Here, we characterize the limiting distribution of \( \frac{X_n}{(p + 1)^n} \), i.e., a scaled version of \( X_n \), the number of protected nodes in an ERT of age \( n \) and index \( p \), by its moments.

**Theorem 4.1** Let \( X_n \) be the number of protected nodes in an exponential recursive tree of age \( n \) and index \( p \). We have the convergence in distribution

\[
\frac{X_n}{(p + 1)^n} \xrightarrow{D} X_*,
\]

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

\[
b_m = \frac{p}{(p + 1)^m - (p + 1)} \sum_{i=1}^{m-1} \binom{m}{i} b_i b_{m-i}, \quad m \geq 2,
\]

with \( b_1 = \mathbb{E}[X_*] = \mu_p = \lim_{n \to \infty} \mu_{n,p} \) at the basis of the induction.
According to the two recurrences (10) and (11) with finite limits for

\[
\lim_{n \to \infty} \mathbb{E}\left[\frac{X_n}{(p+1)^n}\right] = \mu_p, \quad \text{and} \quad \lim_{n \to \infty} \mathbb{E}\left[\frac{X_n}{(p+1)^n} \mathbb{1}_{R_n}\right] = \alpha_p,
\]

we see that \( b_m := \lim_{n \to \infty} \mathbb{E}\left[\frac{X_n}{(p+1)^n}\right]^m \) exists, by induction on \( m \), and we get

\[(p+1)^m b_m = (p+1) b_m + \sum_{i=1}^{m-1} \binom{m}{i} b_i b_{m-i}, \quad m \geq 2.\]

Using this recurrence and \( b_1 = \mu_p < 1 \), we can conclude that \( \frac{b_m}{b_{m-1}} < 1 \) by induction on \( m \) (this induction is shown in page 7 of Mahmoud (2022) and in page 12 of Aguech et al. (2022), as well). Subsequently, for \( |z| < 1 \), the series \( \sum_{m=0}^{\infty} \frac{b_m}{b_{m-1}} z^m \) converges. Therefore, by Theorem 30.1 in Billingsley (2012), \( \frac{X_n}{(p+1)^n} \) converges in distribution to a unique limit \( X_* \).

\section{The Total path length}

Let \( I_n \) be the total path length of protected nodes in an exponential recursive tree \( T_n \) (the tree of size \( n \), at time \( n-1 \)), which is the sum of the depths of all protected nodes in \( T_n \). There are two scenarios at the first step: Either the root recruits a node (let us call it \( v \)), or it does not. In the first scenario, by time
protected nodes of exponential trees

$n-1$, node $v$ has acquired a subtree (call it $T_{n-1}^{(1)}$) of total path length $I_{n-1}^{(1)}$ (measured from $v$), with $I_{n-1}^{(1)}$ distributed like $I_{n-1}$. Measured from the root of $T_n$, each protected node in $T_{n-1}^{(1)}$ is at depth $1$ plus its depth in $T_{n-1}^{(1)}$, there is a total number of protected nodes in $T_{n-1}^{(1)}$ distributed like $X_{n-1}$. Therefore, the contribution of $T_{n-1}^{(1)}$ to the total path length of protected nodes in $T_n$ is distributed like $I_{n-1}^{(1)} + X_{n-1}$.

In the meantime, the root is actively recruiting. Nodes not in $T_{n-1}^{(1)}$ (including) the root of $T_n$ form a tree of total path length $I_{n-1}^{(2)}$, with $I_{n-1}^{(2)}$ distributed like $I_{n-1}$. In this scenario, $T_n$ has a total path length of protected nodes distributed like $I_{n-1}^{(1)} + I_{n-1}^{(2)} + X_{n-1}$.

In the second scenario (failure to recruit in the first step), by time $n-1$, the tree $T_n$ has a total path length of protected nodes distributed like $\hat{I}_{n-1}^{(1)}$.

**Lemma 5.1**

$$I_n \overset{d}{=} I_n^{(1)} + I_n^{(2)} + X_{n-1}$$

where $I_n^{(1)}$, $I_n^{(2)}$, $X_n$, $\hat{I}_{n-1}^{(1)}$ and $\mathbb{I}$ are independent, $I_n^{(i)} \overset{d}{=} I_n$ for $i = 1, 2$, $\mathbb{I}$ is a Bernoulli random variable with success probability $p$, and $X_n$ is the number of protected nodes of an exponential recursive tree at age $n$.

**Remark 1** The internal path length can be obtained by another method, if we denote for all $k$ by

$N_k = \{\text{at step } k \text{ the root recruits a new child} \}$,

then, $I_n$ satisfies, almost surely,

$$I_n = \sum_{k=1}^{n} \mathbb{I}_{N_k} I_{n-k} + X_n - \mathbb{I}_{\mathcal{R}_n},$$

where $I_{n-k}$ and $N_k$ are independent.

**5.1 Mean of $I_n$**

Define, for all integer $n$, $x_n := \mathbb{E}[X_n]$ and $i_n := \mathbb{E}[I_n]$. From the Lemma 5.1, we deduce that

$$i_n = p \left(2i_{n-1} + x_{n-1}\right) + q \cdot i_{n-1} = (1 + p)i_{n-1} + px_{n-1}.$$  

(14)

**Proposition 5.1** The mean $i_n$ of $I_n$ is given by

$$\mathbb{E}[I_n] = p \left(1 + p\right)^n \sum_{k=1}^{n} \mu_{k,p},$$

and

$$\lim_{n \to \infty} \mathbb{E} \left[ \frac{I_n}{n \left(1 + p\right)^n} \right] = p \mu_p,$$

where $\mu_{k,p}$ and $\mu_p$ are given respectively in Theorem 3.7.
Proof: By iteration, we conclude that the solution of the recurrence (14) is
\[ i_n = p \sum_{k=1}^{n-1} (1 + p)^{n-k} x_k. \]
Since, for all \( k, x_k = \mu_k, p(1 + p)^k \) and \( \lim_n \mu_n, p := \mu_p \), the claim of the lemma follows. \( \square \)

5.2 Second moment of \( I_n \)

To obtain the variance of \( I_n \), we need to compute, at first, \( b_n := \mathbb{E}[I_n X_n] \).

Lemma 5.2 Using the notations \([1]\), we have
\[ b_n = (1 + p) b_{n-1} + 2 p x_{n-1} i_{n-1} + p q^{n-1} (i_{n-1} - \mathbb{E} [I_{n-1} \mathbb{I}_{n-1}]) . \]

Proof: The sequence \( X_n \) satisfies
\[ X_n = \mathbb{I} \left( X_{n-1}^{(1)} + X_{n-1}^{(2)} + J_{n-1} - G_{n-1} \right) + (1 - \mathbb{I}) \tilde{X}_{n-1}^{(1)}, \tag{15} \]
where
\[ J_n = \mathbb{I}_{\{ |T_n^{(1)}| = 1 \}} \mathbb{I}_{\{ |T_n^{(2)}| \geq 2 \}}, \quad G_n = \mathbb{I}_{\{ |T_n^{(2)}| = 1 \}} \mathbb{I}_{R_n^{(1)}}. \]

Multiplying (12) by (15) and recall that \( \mathbb{I} := \mathbb{I}_{\{ |T_1| \geq 2 \}} \), we deduce,
\[
X_n I_n = \mathbb{I} \left( X_{n-1}^{(1)} I_{n-1}^{(1)} + X_{n-1}^{(2)} I_{n-1}^{(1)} + J_{n-1} I_{n-1}^{(1)} - G_{n-1} I_{n-1}^{(1)} \right) \\
+ \mathbb{I} \left( X_{n-1}^{(1)} I_{n-1}^{(2)} + X_{n-1}^{(2)} I_{n-1}^{(2)} + J_{n-1} I_{n-1}^{(2)} - G_{n-1} I_{n-1}^{(2)} \right) \\
+ \mathbb{I} \left( X_{n-1}^{(1)} X_{n-1}^{(2)} + X_{n-1}^{(2)} X_{n-1}^{(2)} + J_{n-1} X_{n-1}^{(2)} - G_{n-1} X_{n-1}^{(2)} \right) \\
+ (1 - \mathbb{I}) \tilde{X}_{n-1}^{(1)} \tilde{J}_{n-1}^{(1)}.
\]

On one hand, observe that, almost surely,
\[
J_{n-1} X_{n-1}^{(2)} = I_{n-1}^{(1)} J_{n-1}^{(1)} = G_{n-1} J_{n-1}^{(1)} = (I_{n-1}^{(2)} J_{n-1}^{(1)} |_{|T_{n-1}^{(2)}| = 1}) \mathbb{I}_{R_{n-1}^{(1)}} = 0,
\]
\[
G_{n-1} X_{n-1}^{(2)} = G_{n-1} J_{n-1}^{(2)} = (I_{n-1}^{(2)} J_{n-1}^{(1)} |_{|T_{n-1}^{(2)}| = 1}) \mathbb{I}_{R_{n-1}^{(1)}} = 0.
\]
We deduce, almost surely, that
\[
J_{n-1} I_{n-1}^{(1)} = G_{n-1} I_{n-1}^{(2)} = J_{n-1} X_{n-1}^{(2)} = G_{n-1} X_{n-1}^{(2)} = 0.
\]

On the other hand, from
\[
I_{n-1}^{(2)} J_{n-1} = (I_{n-1}^{(2)} J_{n-1}^{(1)} |_{|T_{n-1}^{(2)}| \geq 2}) \mathbb{I}_{|T_{n-1}^{(2)}| = 1} = I_{n-1}^{(2)} I_{|T_{n-1}^{(2)}| = 1},
\]
\[
I_{n-1}^{(1)} G_{n-1} = (I_{n-1}^{(1)} J_{n-1}^{(1)} \mathbb{I}_{R_{n-1}^{(1)}}) \mathbb{I}_{|T_{n-1}^{(2)}| = 1},
\]
we obtain
\[
\mathbb{E}[I_{n-1}^{(2)}] = \mathbb{E}[I_{n-1}^{(1)}] \mathbb{E}[I_{n-1}^{(1)} | T_{n-1} = 1] = i_{n-1} q^{n-1},
\]
\[
\mathbb{E}[I_{n-1}^{(1)} G_{n-1}] = \mathbb{E}[I_{n-1}^{(1)} I_{n-1}^{(2)}] = q^{n-1} \mathbb{E}[I_{n-1}^{(1)} I_{n-1}^{(2)}].
\]

Then
\[
b_n = p \left( b_{n-1} + x_{n-1} i_{n-1} + \mathbb{E}[I_{n-1}^{(1)} I_{n-1}^{(2)}] - \mathbb{E}[I_{n-1}^{(1)} G_{n-1}] \right)
+ p \left( x_{n-1} i_{n-1} + b_{n-1} + \mathbb{E}[I_{n-1}^{(2)} I_{n-1}^{(1)}] - \mathbb{E}[I_{n-1}^{(2)} G_{n-1}] \right)
+ x_{n-1}^2 + \mathbb{E}[I_{n-1}^{(2)} I_{n-1}^{(2)}] - \mathbb{E}[G_{n-1} X_{n-1}^{(2)}] + q b_{n-1}
= (1 + p) b_{n-1} + 2 p x_{n-1} i_{n-1} + p \mathbb{E}[I_{n-1}^{(2)} I_{n-1}^{(1)}]
+ p \left( \mathbb{E}[I_{n-1}^{(2)} I_{n-1}^{(1)}] - \mathbb{E}[G_{n-1} X_{n-1}^{(2)}] - \mathbb{E}[I_{n-1}^{(2)} G_{n-1}] - \mathbb{E}[I_{n-1}^{(1)} G_{n-1}] \right).
\]

So the assertion of the lemma follows.

To obtain a closed form of \( b_n \), we need to compute \( \mathbb{E}[I_n I_{\mathcal{R}_{n}}] \).

**Lemma 5.3**
\[
\mathbb{E}[I_n I_{\mathcal{R}_{n}}] = q \mathbb{E}[I_{n-1} I_{\mathcal{R}_{n-1}}] + p q^{n-1} (i_{n-1} + x_{n-1}) + p \mathbb{E}[I_{n-1}] \mathbb{P}(\mathcal{R}_{n-1})
+ p \left( 1 - q^{n-1} \right) \mathbb{E}[I_{n-1} I_{\mathcal{R}_{n-1}}] + p x_{n-1} \mathbb{P}(\mathcal{R}_{n-1}).
\]

\[
= (1 - p q^{n-1}) \mathbb{E}[I_{n-1} I_{\mathcal{R}_{n-1}}] + (p \mathbb{P}(\mathcal{R}_{n-1}) + p q^{n-1}) (i_{n-1} + x_{n-1})
= \sum_{k=0}^{n-1} \prod_{j=k+1}^{n} (1 - p q^{j-1}) \left( p \mathbb{P}(\mathcal{R}_k) + p q^k \right) (i_k + x_k).
\]

**Proof:** Since
\[
\mathcal{R}_n = \left[ \{ |T_1| = 1 \} \cap \tilde{\mathcal{R}}_{n-1}^{(1)} \right] \cup \left[ \{ |T_1| = 2 \} \cap \mathcal{R}_{n-1}^{(1)} \cap \left( \{ T_{n-1} = 1 \} \cup \left( \{ T_{n-1} = 2 \} \cap \mathcal{R}_{n-1}^{(2)} \right) \right) \right],
\]
we have
\[
I_{\mathcal{R}_n} = (1 - \mathbb{I}) I_{\tilde{\mathcal{R}}_{n-1}^{(1)}} + \mathbb{I}_{\{ |T_{n-1}| = 1 \}} \mathbb{I}_{\{ |T_{n-1}^{(2)}| = 2 \}} I_{\tilde{\mathcal{R}}_{n-1}^{(1)}} + \mathbb{I}_{\{ |T_{n-1}^{(2)}| = 2 \}} I_{\mathcal{R}_{n-1}^{(1)}} I_{\{ |T_{n-1}^{(2)}| = 2 \}}.
\]

Multiplying by \( I_n \), we obtain
\[
I_n I_{\mathcal{R}_n} = (1 - \mathbb{I}) I_{n-1}^{(1)} I_{\tilde{\mathcal{R}}_{n-1}^{(1)}} + \mathbb{I}_{\{ |T_{n-1}| = 1 \}} \mathbb{I}_{\{ |T_{n-1}^{(2)}| = 2 \}} I_{n-1}^{(2)} + X_{n-1}^{(2)} \mathbb{I}_{\mathcal{R}_{n-1}^{(1)}} I_{\{ |T_{n-1}^{(2)}| = 2 \}} \mathbb{I}_{\mathcal{R}_{n-1}^{(1)}}.
\]

The result of the lemma is obtained by taking the expectation of this equation.

Using Lemmas 5.2 and 5.3 we deduce
Proposition 5.2 The second moment of $I_n$ satisfies the following recursion

$$E[I_{n}^2] = (p + 1) E[I_{n-1}^2] + 2p (i_{n-1}^2 + i_{n-1} x_{n-1} + b_{n-1}) + p E[X_{n-1}^2],$$

with as general solution

$$E[I_{n}^2] = p \sum_{k=1}^{n-1} (p + 1)^{n-k} (2(i_k^2 + i_k x_k + b_k) + E[X_k^2]).$$

5.3 Convergence in Distribution of $I_n$

The aim of this section is to use the contraction method in [Roesler and Rueschendorf (2001)] or the multivariate contraction method in [Neininger (2001)] to state the limiting distribution of $I_n := I_n/n (1 + p)^n$.

To apply the contraction method, we set up some notation as follows: For a random variable $X$, we write $X \sim \lambda$ if the distribution of $X$ is $\lambda$, i.e. the law $\mathcal{L}(X)$ of $X$ is $\lambda$. The symbol $\|X\|_2 := (E[|X|^2])^{1/2}$ denotes the usual $L_2$-norm of $X$. The Wasserstein-metric $\ell_2$ is defined on the space of probability distributions with existing second moments by

$$\ell_2(X, Y) := \ell_2(\mathcal{L}(X), \mathcal{L}(Y)) := \ell_2(\lambda, \nu) := \inf\{\|X - Y\|_2 : X \sim \lambda, Y \sim \nu\}.$$  

By $\mathcal{M}_2$ the space of all probability distributions $\lambda$ with mean $p \mu_\lambda$ (as in Proposition 5.1) and finite second moment is denoted. The metric space $(\mathcal{M}_2, \ell_2)$ is complete and convergence in $\ell_2$ is equivalent to convergence in distribution plus convergence of the second moments.

By equation (12) we have

$$\frac{I_n}{n(1 + p)^n} = \frac{\text{I}}{1 + p} \cdot \frac{n - 1}{n} \cdot \left(\frac{I_{n-1}^{(1)}}{(n-1)(1+p)^{n-1}} + \frac{I_{n-1}^{(2)}}{(n-1)(1+p)^{n-1}}\right) + \frac{1 - \text{I}}{1 + p} \cdot \frac{n - 1}{n} \cdot \frac{I_{n-1}^{(1)}}{(n-1)(1+p)^{n-1}} + \text{I} \cdot \frac{X_{n-1}}{n(1+p)^n}.  \tag{16}$$

Theorem 5.1 Let $I_n$ and $X_n$ be the total path length and the number of protected nodes of an exponential recursive tree at age $n$, respectively. The normalized total path length $\hat{I}_n := I_n/n (1 + p)^n$ satisfies the distributional recursion

$$\hat{I}_n \overset{d}{=} \frac{\text{I}}{1 + p} \cdot \frac{n - 1}{n} \cdot (\hat{I}_{n-1}^{(1)} + \hat{I}_{n-1}^{(2)}) + \frac{1 - \text{I}}{1 + p} \cdot \frac{n - 1}{n} \cdot \hat{I}_{n-1}^{(i)} + \text{I} \cdot \frac{X_{n-1}}{n(1+p)^n}, \tag{17}$$

where $\hat{I}_{n}^{(1)}$, $(\hat{I}_{n}^{(2)}, X_n)$, $\hat{I}_{n}^{(1)}$ and I are independent, $\hat{I}_{n}^{(i)} \overset{d}{=} \hat{I}_n$ for $i = 1, 2$, and I is a Bernoulli random variable with success probability $p$, then $\hat{I}_n \overset{D}{\to} \hat{I}$, as $n \to \infty$, where the random variable $\hat{I}$ is the unique distributional fixed-point of

$$\hat{I} \overset{d}{=} \frac{\text{I}}{1 + p} \cdot (\hat{I}^{(1)} + \hat{I}^{(2)}) + \frac{1 - \text{I}}{1 + p} \hat{I}^{(1)}, \tag{18}$$

with $\hat{I}^{(i)}$, $i = 1, 2$, are independent copies of $\hat{I}$ and independent of $\text{I}$. 


Proof: The equation (17) is an immediate consequence of the equation (16), where we define \( \hat{I}_n(i) := \frac{\hat{I}_n(i)}{n(1 + p)^n} \). Using Theorem 4.1 and Slutsky’s theorem, we have \( \frac{X_{n-1}}{n(1 + p)^n} \xrightarrow{\mathcal{D}} 0 \), and then convergence in probability \( \frac{X_{n-1}}{n(1 + p)^n} \xrightarrow{P} 0 \). That is,

\[
\hat{I}_n \xrightarrow{d} \frac{1}{1 + p} \cdot \frac{n - 1}{n} \cdot (\hat{I}_{n-1}^{(1)} + \hat{I}_{n-1}^{(2)}) + \frac{1 - \frac{1}{1 + p}}{n} \cdot n - 1 \hat{I}_{n-1}^{(1)} + o_p(1),
\]

where \( o_p(1) \) denotes a quantity tending to zero in probability. For purposes of convergence we can, and hence will, ignore the \( o_p(1) \) term.

Consider the well-defined transformation \( T : \mathcal{M}_2 \rightarrow \mathcal{M}_2 \),

\[
T(\lambda) := \mathcal{L} \left( \frac{1}{1 + p} \left( \hat{I}_\lambda^{(1)} + \hat{I}_\lambda^{(2)} \right) + \frac{1 - \frac{1}{1 + p}}{1 + p} \hat{I}_\lambda^{(1)} \right),
\]

where \( \hat{I}_\lambda^{(i)} := \hat{I}^{(i)} \), \( i = 1, 2 \), and \( \mathcal{L} \) are independent and \( \hat{I}^{(i)} \) have a distribution.

At a first step, we have to prove that the transformation \( T \) has a unique fixed point with respect to the \( \ell_2 \)-metric. Let \( \lambda, \nu \in \mathcal{M}_2 \) be given. By (20), we have

\[
T(\lambda) = \mathcal{L} \left( \frac{1}{1 + p} \left( \hat{I}_\lambda^{(1)} + \hat{I}_\lambda^{(2)} \right) + \frac{1 - \frac{1}{1 + p}}{1 + p} \hat{I}_\lambda^{(1)} \right),
\]

\[
\ell_2^2(T(\lambda), T(\nu)) \leq 2 \frac{\mathbb{E}[\mathcal{T}^2]}{(1 + p)^2} \mathbb{E}[|\hat{I}_\lambda - \hat{I}_\nu|^2] + \mathbb{E}[\mathcal{T}^2] \mathbb{E}[|\hat{I}_\lambda - \hat{I}_\nu|^2]
\]

\[
= \frac{1}{1 + p} \mathbb{E}[|\hat{I}_\lambda - \hat{I}_\nu|^2].
\]

Therefore, we have \( \ell_2^2(T(\lambda), T(\nu)) \leq \frac{1}{1 + p} \ell_2^2(\lambda, \nu) \). Since \( \frac{1}{1 + p} \leq 1 \), we deduce that \( T \) is a contraction with respect to the \( \ell_2 \)-metric. Thus, Banach’s fixed point theorem provides existence and uniqueness of solutions of the fixed point equation \( T(\lambda) = \lambda \). By (19), if \( \hat{I}_n \) converges in distribution to some random variable \( \hat{I} \), then \( \hat{I} \) satisfies (18). Consequently, the distribution of \( \hat{I} \) will be \( \lambda_0 \), the unique fixed point of \( T \), i.e. \( \mathcal{L}(\hat{I}) = T(\lambda_0) = \lambda_0 \). Therefore, we have to prove \( \lim_{n \to \infty} \ell_2(\hat{I}_n, \hat{I}) = 0 \) to conclude \( \hat{I}_n \xrightarrow{\mathcal{D}} \hat{I} \).

Since \( \hat{I}_n(i) \equiv \hat{I}_n(i) \equiv \hat{I} \), for \( i = 1, 2, 3 \), we deduce

\[
\lim_{n \to \infty} \ell_2^2(\hat{I}_n, \hat{I}) \leq \frac{p}{(1 + p)^2} \lim_{n \to \infty} \left( \left\| \frac{n - 1}{n} \hat{I}_{n-1}^{(1)} - \hat{I}^{(1)} \right\|^2 + \left\| \frac{n - 1}{n} \hat{I}_{n-1}^{(2)} - \hat{I}^{(2)} \right\|^2 \right)
\]

\[
+ \frac{1 - p}{(1 + p)^2} \lim_{n \to \infty} \left( \left\| \frac{n - 1}{n} \hat{I}_{n-1}^{(3)} - \hat{I}^{(3)} \right\|^2 \right)
\]

\[
\leq \frac{1}{1 + p} \lim_{n \to \infty} \left\| \hat{I}_n - \hat{I} \right\|^2.
\]

Therefore, we have

\[
\lim_{n \to \infty} \ell_2^2(\hat{I}_n, \hat{I}) \leq \frac{1}{1 + p} \lim_{n \to \infty} \ell_2^2(\hat{I}_n, \hat{I}).
\]

This is true only if \( \lim_{n \to \infty} \ell_2(\hat{I}_n, \hat{I}) = 0 \). \quad \Box
References


