J. Appl. Math. & Data. Anal.

Available online at `https://jamda.abru.ac.ir/`
Journal of Applied Mathematics & Data Analytics doi:
Vol. 1, No. 3, 2025, 76-86
Research Article

The Sackin Index of Random Non-uniform Recursive Trees

Ramin Kazemi*
Department of Statistics, Imam Khomeini International University, Qazvin, Iran

Abstract

Abstract: A non-uniform recursive trees is a recursive tree with ordered sets of descendants. A random non-uniform recursive tree of order n is one chosen with equal probability from the space of all such trees. The Sackin index of a tree which summarizes the shape of a tree is defined as the sum of the depths of its leaves. The mean and variance of this index in random non-uniform recursive trees are given. Also, two inequalities related to this index are given.

Keywords: Non-uniform recursive tree, Sackin index, Total path length.
2020 Mathematics Subject Classification: 05C05; 60F05

1 Introduction

A social networking site is an internet-based platform. In this platform, social relationships are established with friends, family, colleagues, customers or clients. Social and business goals in social networks can be pursued through well-known sites such as Facebook, Twitter, TikTok, LinkedIn, Instagram, Telegram, Snapchat, WhatsApp and also YouTube. These sites develop relationships very quickly and share information, messages and ideas. Therefore, social networks are an important platform for marketers whose ultimate goal is to attract customers. A special and important type of these platforms are tree-like structures. Trees are defined as connected graphs without cycles, and their properties are basics of graph theory. A rooted tree is a tree with a countable number of nodes, in which a particular node is distinguished from the others and called the root node.

Let G be a graph. The vertex (node) and edge sets of a G are denoted by $V(G)$ and $E(G)$, respectively. The number of tail ends adjacent to a node v is called its outdegree and is denoted by $d^+(v)$. A recursive tree with n nodes is an unordered rooted tree, where the nodes are labelled by distinct integers from $\{1, 2, 3, ..., n\}$ in such a way that the sequence of labels lying on the unique path from the root node to any node in the tree are always forming an increasing sequence [5]. A non-uniform recursive tree is a recursive tree with ordered sets of descendants. In recursive trees, the ordering of the immediate descendants of a given node does not matter, as all ordering represent

---

*Corresponding author: r.kazemi@sci.ikiu.ac.ir

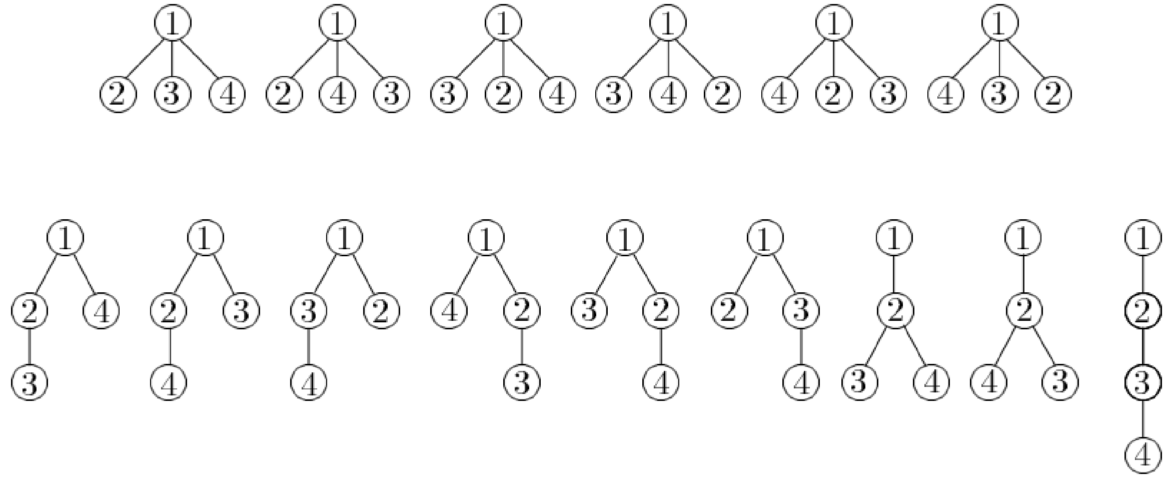

Figure 1: All 15 non-uniform recursive trees of order 4 [8]

the same tree. A random non-uniform recursive tree of order n is one chosen with equal probability from the space of all such trees. There is a simple growth rule for the class of non-uniform recursive trees. In this class, a random tree T_n, of order n, is obtained from T_{n-1}, a random tree of order $n-1$, by choosing a parent in T_{n-1} and adjoining a node labeled n to it. The node n can be adjoined at any of the insertion positions or gaps between the children of the chosen parent since insertion in each gap will give a different ordering. We can describe the non-uniform recursive trees evolution process which generates random trees (of arbitrary order n) of grown trees. This description is a consequence of the considerations made in:
Step 1: The process starts with the root labelled by 1.
Step $i+1$*:* At step $i+1$ the node with label $i+1$ is attached to any previous node v (with outdegree $d^{+}(v)$) of the already grown non-uniform recursive tree of order i with probability

$$p_{i+1}(v) = \frac{1 + d^{+}(v)}{2i-1}.$$

Non-uniform recursive trees were introduced in the literature under a few different names such as plane-oriented recursive trees, heap ordered trees, scale-free trees. All 15 non-uniform recursive trees of order 4 are shown in Figure 1.

The distance $D_{n,j}$ between the root and node j (the depth of j-th node) in a random non-uniform recursive tree of order n has been studied by Mahmoud [4]. He proved that the mean and variance of $D_{n,j}$ are given by

$$\begin{aligned}\mathbb{E}(D_{n,j}) &= H_{2j-2} - \frac{1}{2}H_{j-1},\\ \mathbb{V}ar(D_{n,j}) &= H_{2j-2} - \frac{1}{2}H_{j-1} - H^{(2)}_{2j-2} + \frac{1}{4}H^{(2)}_{j-1},\end{aligned}$$

where H_n, the n-th harmonic number and $H_n^{(2)}$ is the n-th harmonic number of order 2. Then $\mathbb{E}(D_{n,j}) = \mathbb{V}ar(D_{n,j}) = \frac{1}{2}\log j + \mathcal{O}(1)$. The total path length of a non-uniform recursive trees, namely,

$$I_n = \sum_{j=1}^{n} D_{n,j},$$

defined as the sum of all root-to-node distances. Linearity of expectation gives

$$\mathbb{E}(I_n) = \Big(H_{2n-3} - \frac{1}{2}H_{n-2}\Big)\Big(n - \frac{1}{2}\Big) - \frac{1}{2}(n-1),$$

which is asymptotically equivalent to $\frac{1}{2}n\log n$. A node with outdegree zero is known as leaf node. Originating from Sackin's idea to analyze the leaf depths of a tree, the Sackin index is one of the oldest and most widely applied measures of tree balance. It is defined for arbitrary rooted trees and it is an imbalance index, i.e., for a fixed $n \in \mathbb{N}_{\geq 1}$, it increases with decreasing balance of the tree. Its basic idea is to sum up the distances of all leaves to the root of the tree. Note that if this sum is small, it basically implies that all leaves are relatively close to the root, which makes the tree balanced [9, 10]. Let S_n be the Sackin index of a tree of order n and $\Gamma(\cdot)$ be the gamma function. Moradian et al. [6] have studied the Sackin index in random recursive trees. They showed that

$$\mathbb{E}(S_n) = \frac{n}{2}\Big(H_n - \frac{1}{2}\Big).$$

For d-ary increasing trees, Kazemi and Behtoei [3] showd that

$$\mathbb{E}(S_n) = c[n]\sum_{j=1}^{n-1}\frac{\alpha(j)}{c[j+1]},$$

where

$$c[n] = \begin{cases} \frac{\Gamma(n-1)}{\Gamma\left(n+\frac{1}{d-1}\right)}, & n \geq 2 \\ 0, & n = 1, \end{cases} \qquad \alpha(j) = \frac{(d-1)\mathbb{E}(I_j) + d\,j}{(d-1)j+1}, \qquad j \geq 1$$

and

$$\mathbb{V}ar(S_n) = t[n]\sum_{j=3}^{n-1}\frac{\beta(j)}{t[j+1]},$$

where

$$t[n] = \frac{\Gamma\Big(n - 2 - \frac{1}{d-1}\Big)}{\Gamma\Big(n + \frac{1}{d-1}\Big)}$$

and

$$\begin{aligned} \beta(i) &= 2\frac{d-1}{i(d-1)+1}\mathbb{C}ov(I_i, S_i) - \Big(\frac{-d\mathbb{E}(S_i)}{(d-1)i+1} + \alpha(i) - 1\Big)^2 \\ &+ \frac{1}{(d-1)i+1}\left(d\sum_{v\in V(T_i),d^+(v)\geq 1}\mathbb{E}(D_v^2) - \sum_{v\in V(T_i)}\mathbb{E}(D_v^2) + 2\mathbb{E}(I_i) - (i-1)\right). \end{aligned}$$

For more results on the Sackin index, see [4] and references therein. In the discourse we shall use the following notations:

$$Q_1[n,i] = \frac{\Gamma\Big(n-\frac{i}{2}\Big)}{\Gamma\Big(n-\frac{1}{2}\Big)},$$
$$Q_2[n,i] = \sum_{j=1}^{n} \frac{\alpha_i(j)}{Q_1[j+1,i]}, \quad i = 2,3$$

where $\alpha_i(j)$ for $i = 2,3$ are introduced in Theorems 2.2 and 2.6, respectively.

2 The Main Results

2.1 Mean

Lemma 2.1. *For each rooted tree T_n of order n, we have*

$$\begin{aligned} &a) \qquad \sum_{v\in V(T_n)} d^+(v)\,D(v) = I_n - (n-1), \\ &b) \qquad \sum_{v\in V(T_n)} D(v)^2 - \sum_{v\in V(T_n)} d^+(v)\,D^2(v) = 2I_n - (n-1), \end{aligned}$$

where $D(v)$ is the depth of node v.

Proof. Assume that $r = v_1$ is the root of T_n. Each vertex in $V(T_n)\setminus\{r\}$ is a child of another vertex and is counted in its outdegree. Hence

$$\bigcup_{v\in V(T)} \bigcup_{u\in N^+(v)} \{u\} = \bigcup_{v\in V(T)} N^+(v) = V(T)\setminus\{r\}, \tag{1}$$

where $N^+(v)$ is the set of children of v and

$$\sum_{v\in V(T_n)} d^+(v) = n-1.$$

Assume that $d^+(v_i) = t$ and its children are $v_{i_1}, v_{i_2}, ..., v_{i_t}$. Thus, for each $j \in \{1,2,...,t\}$, we have $D(v_{i_j}) = D(v_i) + 1$. This implies that

$$\sum_{j=1}^{t} D(v_{i_j}) = \sum_{j=1}^{t}(D(v_i)+1) = t\ (D(v_i)+1) = d^+(v_i)D(v_i) + d^+(v_i).$$

Since the depth of root is zero, we have [7]

$$\begin{aligned}
\sum_{v\in V(T_n)} (d^+(v)D(v)+d^+(v)) &= \sum_{v\in V(T_n)} \Big(\sum_{u\in N^+(v)} D(u)\Big)\\
&= \sum_{v\in V(T_n)} \sum_{u\in N^+(v)} D(u)\\
&= \sum_{x\in V(T_n)\setminus\{r\}} D(x)\\
&= \sum_{x\in V(T_n)} D(x)\\
&= I_n.
\end{aligned}$$

This implies that

$$\sum_{v\in V(T_n)} d^+(v)\,D(v) = I_n - \sum_{v\in V(T_n)} d^+(v) = I_n-(n-1).$$

For each $v\in V(T_n)$ and $u\in N^+(v)$ we have $D(u)=D(v)+1$. Also, the depth of root is zero. These facts using the relation (1) imply that

$$\begin{aligned}
\sum_{x\in V(T_n)} D(x)^2 &= \sum_{x\in V(T_n)\setminus\{r\}} D(x)^2\\
&= \sum_{v\in V(T_n)} \sum_{u\in N^+(v)} D(u)^2\\
&= \sum_{v\in V(T_n)} \sum_{u\in N^+(v)} (D(v)+1)^2\\
&= \sum_{v\in V(T_n)} \left[d^+(v)\,(D(v)+1)^2\right]\\
&= \sum_{v\in V(T_n)} d^+(v)\,(D(v)^2+2D(v)+1)\\
&= (n-1)+2(I_n-(n-1))+\sum_{v\in V(T_n)} d^+(v)D(v)^2\\
&= 2I_n-(n-1)+\sum_{v\in V(T_n)} d^+(v)D(v)^2
\end{aligned}$$

and proof is completed [3]. ■

Set

$$\alpha_2(i)=\frac{2\mathbb{E}(I_i)+i}{2i-1}, \qquad i\geq 1.$$

Theorem 2.2. *Suppose $n\geq 3$. The mean of the Sackin index S_n of a random non-uniform recursive tree of order n is given by*

$$\mathbb{E}(S_n)=Q_1[n,2]Q_2[n-1,2].$$

Proof. Let $\mathcal{F}_n$ be the sigma-field generated by the first n stages of non-uniform recursive trees [1, 2]. For each $v \in V(T_n)$ define the indicator $I(v)$ as below:

$$I(v) = \begin{cases} 0, & d^+(v) = 0 \\ D(v), & d^+(v) \geq 1. \end{cases}$$

Assume that v_n is attached to a randomly chosen vertex U_{n-1} in T_{n-1}. Hence

$$S_n = S_{n-1} + I(U_{n-1}) + 1.$$

From Lemma 2.1 (a),

$$\begin{aligned} \mathbb{E}(S_n|\mathcal{F}_{n-1}) &= S_{n-1} + \mathbb{E}(I(U_{n-1})|\mathcal{F}_{n-1}) + 1 \\ &= S_{n-1} + \sum_{v \in V(T_{n-1})} p_{n-1}(v) I(v) + 1 \\ &= S_{n-1} + \sum_{v \in V(T_{n-1})} \frac{1 + d^+(v)}{2n-3} I(v) + 1 \\ &= S_{n-1} + \frac{1}{2n-3} \sum_{v \in V(T_{n-1})} (1 + d^+(v)) I(v) + 1 \\ &= S_{n-1} + \frac{1}{2n-3} \left[\sum_{v \in V(T_{n-1})} I(v) + \sum_{v \in V(T_{n-1})} d^+(v) I(v) \right] + 1 \\ &= S_{n-1} + \frac{1}{2n-3} \left[(I_{n-1} - S_{n-1}) + (I_{n-1} - (n-2) \right] + 1 \\ &= \frac{2n-4}{2n-3} S_{n-1} + \frac{2}{2n-3} I_{n-1} + \frac{n-1}{2n-3}. \end{aligned}$$

Since $\frac{2n-4}{2n-3} = \frac{Q_1[n,2]}{Q_1[n-1,2]}$, we have

$$\mathbb{E}(S_n) = \frac{Q_1[n,2]}{Q_1[n-1,2]} \mathbb{E}(S_{n-1}) + \alpha_2(n-1). \tag{2}$$

It is obvious that $\mathbb{E}(S_1) = 0$. The recurrence (2) gives

$$\mathbb{E}(S_n) = Q_1[n,2] \sum_{i=1}^{n-1} \frac{\alpha_2(i)}{Q_1[i+1,2]},$$

and proof is completed. ■

It can be shown that in terms of harmonic numbers [7]:

$$\mathbb{E}(S_n) = \left(n - \frac{1}{2}\right)\left(\frac{2}{3} H_{2n} - \frac{1}{3} H_n\right) - \frac{2}{3} n + \frac{1}{3}, \qquad n \geq 3.$$

2.2 Variance

By Naderi et al. [7],

$$\mathbb{E}(I_i|\mathcal{F}_{i-1}) = \Delta_{i-1}I_{i-1} + \Lambda_{i-1},$$

where

$$\Delta_i = \frac{2i+1}{2i-1}, \quad \Lambda_i = \frac{i}{2i-1}, \qquad i \geq 1.$$

Lemma 2.3. *We have*

$$\mathbb{E}\left(I(U_i)\right) = \frac{1}{2i-1}\Big(2\mathbb{E}(I_i) - \mathbb{E}(S_i) - (i-1)\Big)$$

and

$$\mathbb{E}\left(I(U_i)D(U_i)\right) = \frac{2\sum\limits_{v\in V(T_i)} \mathbb{E}(D_v^2) - 2\mathbb{E}(I_i) - (i-1)}{2i-1}.$$

Proof. From Lemma 2.1 (a),

$$\begin{aligned}\mathbb{E}\left(I(U_i)\right) &= \frac{1}{2i-1}\mathbb{E}\left(\sum_{v\in V(T_i)} I(v)(1+d_v^+)\right)\\ &= \frac{1}{2i-1}\mathbb{E}\left((I_i - S_i) + \sum_{v\in V(T_i)} d_v^+ D_v\right)\\ &= \frac{1}{2i-1}\Big(2\mathbb{E}(I_i) - \mathbb{E}(S_i) - (i-1)\Big).\end{aligned}$$

From Lemma 2.1 (b),

$$\begin{aligned}\mathbb{E}\left(I(U_i)D(U_i)\right) &= \frac{1}{2i-1}\mathbb{E}\left(\sum_{v\in V(T_i)} (1+d_v^+)I(v)D_v\right)\\ &= \frac{2\sum\limits_{v\in V(T_i)} \mathbb{E}(D_v^2) - 2\mathbb{E}(I_i) - (i-1)}{2i-1}.\end{aligned}$$

■

In passing, we want to determine the variance of S_n. First we prove the following theorem. Let $\mathbb{C}ov(S_n, I_n)$ be the covariance between two random variables S_n and I_n.

Theorem 2.4. *We have*

$$\mathbb{C}ov\,(S_n, I_n) = \begin{cases} \sum\limits_{j=1}^{n-1}\left(d_j \prod\limits_{i=j+1}^{n-1} \Delta_i\right) - \mathbb{E}(S_n)\mathbb{E}(I_n), & n \geq 4 \\ 0, & n = 1, 2, 3 \end{cases}$$

where

$$d_i = \Lambda_i \mathbb{E}(S_i) + \mathbb{E}(I_{i+1}) + \frac{\Lambda_i - \Delta_i}{1-\Delta_i}\mathbb{E}(I(U_i)) - \frac{\Delta_i}{1-\Delta_i}\mathbb{E}(I(U_i)D(U_i)).$$

Proof. For $n = 1, 2, 3$, $\mathbb{C}ov\,(S_n, I_n) = 0$. Assume $n \geq 4$. By definition,

$$\mathbb{C}ov\,(S_n, I_n) = \mathbb{E}(I_n S_n) - a_n b_n.$$

Since $S_n = S_{n-1} + I(U_{n-1}) + 1$, we have

$$\begin{aligned}\mathbb{E}\,(S_n I_n) &= \mathbb{E}\Big(\big(S_{n-1} + I(U_{n-1}) + 1\big) I_n\Big) \\ &= \mathbb{E}\big(S_{n-1} I_n\big) + \mathbb{E}\,\big(I\big(U_{n-1}\big) I_n\big) + b_n.\end{aligned}$$

But

$$\begin{aligned}\mathbb{E}\,\big(I\big(U_{n-1}\big) I_n\big) &= \mathbb{E}\,\big(I\big(U_{n-1}\big)\ \mathbb{E}\big(I_n|\mathcal{F}_{n-1}\big)\big) \\ &= \mathbb{E}\,\big(I\big(U_{n-1}\big)\ \big(\Delta_{n-1} I_{n-1} + \Lambda_{n-1}\big)\big) \\ &= \Delta_{n-1}\mathbb{E}\,\big(I\big(U_{n-1}\big)\big(I_n - D(U_{n-1}) - 1\big)\big) + \Lambda_{n-1}\mathbb{E}\,\big(I\big(U_{n-1}\big)\big) \\ &= \Delta_{n-1}\mathbb{E}\,\big(I\big(U_{n-1}\big) I_n\big) - \Delta_{n-1}\mathbb{E}\,\big(I\big(U_{n-1}\big) D(U_{n-1})\big) \\ &\quad -\Delta_{n-1}\mathbb{E}\,\big(I\big(U_{n-1}\big)\big) + \Lambda_{n-1}\mathbb{E}\,\big(I\big(U_{n-1}\big)\big).\end{aligned}$$

Hence,

$$\mathbb{E}\,\big(I\big(U_{n-1}\big) I_n\big) = \frac{\Lambda_{n-1} - \Delta_{n-1}}{1-\Delta_{n-1}}\mathbb{E}\,\big(I\big(U_{n-1}\big)\big) - \frac{\Delta_{n-1}}{1-\Delta_{n-1}}\mathbb{E}\,\big(I\big(U_{n-1}\big) D(U_{n-1})\big),$$

where the two terms on the right of the above equality were previously calculated in Lemma 2.3. For simplicity, set

$$\begin{aligned}a_i &= \mathbb{E}(S_i), \\ b_i &= \mathbb{E}(I_i), \\ c_i &= \frac{\Lambda_i - \Delta_i}{1-\Delta_i}\mathbb{E}(I(U_i)) - \frac{\Delta_i}{1-\Delta_i}\mathbb{E}(I(U_i)D(U_i)).\end{aligned}$$

Now,

$$\begin{aligned}\mathbb{E}\,(S_{n-1} I_n) &= \mathbb{E}\big(\mathbb{E}\,(S_{n-1} I_n|\mathcal{F}_{n-1})\big) \\ &= \mathbb{E}\big(S_{n-1}\ \mathbb{E}\,(I_n|\mathcal{F}_{n-1})\big) \\ &= \Delta_{n-1}\mathbb{E}\,(S_{n-1} I_{n-1}) + \Lambda_{n-1} a_{n-1}.\end{aligned}$$

Hence

$$\begin{aligned}\mathbb{E}\,(S_n I_n) &= \Delta_{n-1}\mathbb{E}\,(S_{n-1} I_{n-1}) + \Lambda_{n-1} a_{n-1} + c_{n-1} + b_n \\ &= \Delta_{n-1}\mathbb{E}\,(S_{n-1} I_{n-1}) + d_{n-1}.\end{aligned}$$

Since $S_1 I_1 = 0$, iterating this recurrence completes the proof. ■

Lemma 2.5. *We have*

$$\begin{aligned}\mathbb{C}ov(S_n, S_{n-1}) &= \frac{Q_1[n,2]}{Q_1[n-1,2]}\mathbb{V}ar(S_{n-1}) \\ &+ \frac{2}{2n-3}\Big(\sum_{j=1}^{n-1}\Big(d_j\prod_{i=j+1}^{n-2}\Delta_i\Big) - \mathbb{E}(S_{n-1})\mathbb{E}(I_{n-1})\Big), \qquad n \geq 4.\end{aligned}$$

Proof. From (2),

$$\begin{aligned}\mathbb{E}(S_n - \mathbb{E}(S_n)|\mathcal{F}_{n-1}) &= \mathbb{E}(S_n|\mathcal{F}_{n-1}) - \mathbb{E}(S_n) \\ &= \frac{Q_1[n,2]}{Q_1[n-1,2]}(S_{n-1} - \mathbb{E}(S_{n-1})) + \frac{2(I_{n-1} - \mathbb{E}(I_{n-1}))}{2n-3}.\end{aligned}$$

Then

$$\begin{aligned}\mathbb{C}ov(S_n, S_{n-1}) &= \mathbb{E}(\mathbb{E}((S_n - \mathbb{E}(S_n))(S_{n-1} - \mathbb{E}(S_{n-1})))|\mathcal{F}_{n-1})) \\ &= \mathbb{E}((S_{n-1} - \mathbb{E}(S_{n-1}))\mathbb{E}(S_n - \mathbb{E}(S_n)|\mathcal{F}_{n-1})) \\ &= \frac{Q_1[n,2]}{Q_1[n-1,2]}\mathbb{V}ar(S_{n-1}) + \frac{2}{2n-3}\mathbb{C}ov(I_{n-1}, S_{n-1}).\end{aligned}$$

■

We can extend the above lemma to S_i and S_j for $i > j$, but it is not necessary for our goal. Set

$$\alpha_3(i) = 2\beta(i) + \frac{2}{2i-1}\Big(\sum_{j=1}^{i}\mathbb{E}(D_{i,j}^2) - \mathbb{E}(I_i)\Big) + \frac{i-1}{2i-1} - \Big(\frac{2i\mathbb{E}(S_i)}{2i-1} + \alpha_2(i)\Big)^2,$$

where

$$\beta(i) = \frac{2}{2i-1}\mathbb{C}ov(I_i, S_i), \quad i \geq 1.$$

Theorem 2.6. *Let $n \geq 3$. The variance of the Sackin index S_n of a random non-uniform recursive trees of order n is given by*

$$\mathbb{V}ar(S_n) = Q_1[n,3]Q_2[n-1,3].$$

Proof. From Lemma 2.1 (b),

$$\begin{aligned}\mathbb{E}(S_n - S_{n-1} - 1)^2 &= \mathbb{E}(I^2(U_{n-1})) \\ &= \mathbb{E}(\mathbb{E}(I^2(U_{n-1})|\mathcal{F}_{n-1})) \\ &= \mathbb{E}\Big(\sum_{j=1}^{n-1}\frac{1 + d^+(v_j)}{2n-3}D^2(v_j)\Big) \\ &= \frac{2}{2n-3}\Big(\sum_{j=1}^{n-1}\mathbb{E}(D_{n-1,j}^2) - \mathbb{E}(I_{n-1})\Big) + \frac{n-2}{2n-3}. \qquad (3)\end{aligned}$$

From Lemma 2.5,

$$
\begin{aligned}
\mathbb{E}(S_n - S_{n-1} - 1)^2 &= \mathbb{E}(S_n - \mathbb{E}(S_n) - S_{n-1} + \mathbb{E}(S_{n-1}) + \mathbb{E}(S_n) - \mathbb{E}(S_{n-1}) - 1)^2 \\
&= \mathbb{E}(S_n - \mathbb{E}(S_n) - S_{n-1} + \mathbb{E}(S_{n-1}))^2 \\
&+ \mathbb{E}(\mathbb{E}(S_n) - \mathbb{E}(S_{n-1}) - 1)^2 \\
&+ 2\mathbb{E}(S_n - \mathbb{E}(S_n) - S_{n-1} + \mathbb{E}((S_{n-1}))(\mathbb{E}(S_n) - \mathbb{E}(S_{n-1}) - 1)) \\
&= \mathbb{V}ar(S_n) + \Big(1 - 2\frac{Q_1[n,2]}{Q_1[n-1,2]}\Big)\mathbb{V}ar(S_{n-1}) - 2\beta(n-1) \\
&+ \Big(\frac{2n-2}{2n-3}\mathbb{E}(S_{n-1}) + \alpha_2(n-1)\Big)^2. \qquad (4)
\end{aligned}
$$

Now, from (3) and (4),

$$
\begin{aligned}
\mathbb{V}ar(S_n) &= \Big(2\frac{Q_1[n,2]}{Q_1[n-1,2]} - 1\Big)\mathbb{V}ar(S_{n-1}) + \alpha_3(n-1) \\
&= \frac{Q_1[n,3]}{Q_1[n-1,3]}\mathbb{V}ar(S_{n-1}) + \alpha_3(n-1).
\end{aligned}
$$

By iteration, proof is completed since $\mathbb{V}ar(S_1) = 0$. ■

2.3 Two Inequalities

A discrete-time submartingale is a sequence $X_1, X_2, ...$ of integrable random variables satisfying [1]:

$$\mathbb{E}(X_{n+1}|X_1, X_2, ..., X_n) \geq X_n.$$

Lemma 2.7. *The sequence* $(Z_n)_{n\geq 0}$ *with*

$$Z_n = S_n + \mathbb{E}(S_n),$$

is a submartingale.

Proof. Given the history of insertions $D_1, ..., D_{n-1}$, the values of $Z_1, ..., Z_{n-1}$ are completely determined. From Theorem 2.2,

$$
\begin{aligned}
\mathbb{E}(Z_n|Z_1, ..., Z_{n-1}) &= \mathbb{E}(Z_n|D_1, ..., D_{n-1}) \\
&\geq \mathbb{E}(S_{n-1} + 1|D_1, ..., D_{n-1}) + Q_1[n,2]Q_2[n-1,2] \\
&\geq S_{n-1} + Q_1[n-1,2]Q_2[n-2,2] \\
&= Z_{n-1}.
\end{aligned}
$$

Also $\mathbb{E}(|Z_n|) < \infty$ exists for each n and proof is completed. ■

Theorem 2.8. *Let* T *be a finite stopping time for* $\{Z_n\}_{n\geq 1}$. *Then*

$$\mathbb{E}(|Z_T|) \leq 2\sup_n \mathbb{E}(Z_n)$$

and for $\lambda \geq 0$,

$$P\Big(\sup_n Z_n > \lambda\Big) \leq \frac{\sup_n \mathbb{E}(Z_n)}{\lambda}.$$

Proof. By Lemma 2.7 and Doob's submartingale inequality, proof is completed [1]. ■

References

[1] Hall, P. & Heyde, C. C. (1980). *Martingale limit theory and its application, Probability and Mathematical Statistics,* Academic Press, Inc., Publishers, New York-London. doi:10.1016/C2013-0-10759-7

[2] Kazemi, R. (2014). The eccentric connectivity index of bucket recursive trees, *Iranian Journal of Mathematical Chemistry,* 5(2), 77–83. doi:10.22052/ijmc.2014.5293

[3] Kazemi, R. & Behtoei, A. (2021). The moments of the Sackin index of random d-ary increasing trees, *Matematicki Vesnik*, 73(1), 55–62. doi:10.5937/matves2101055K

[4] Mahmoud, H. M. (1992). Distances in random plane-oriented recursive trees, *Journal of Computational and Applied Mathematics*, 237–245. doi:10.1016/0377-0427(92)90285-2

[5] Meir, A. & Moon, J. W. (1988). Recursive trees with no nodes of out-degree one, *Congressus Numerantium,* 66, 49–62.

[6] Moradian, K., Kazemi, R. & Behzadi, M. H. (2017). The Sackin index of random recursive trees, *U.P.B. Sci. Bull., Series A*, 79(2), 125–130. doi:10.1016/j.cam.2006.08.029

[7] Naderi, S., Kazemi, R. & Behzadi, M. H. (2022). Limit laws for two distance-based indices in random recursive tree models, *Acta Universitatis Sapientiae, Informatica,* 14(1), 35–48. doi:10.2478/ausi-2022-0003

[8] Prodinger, H. (1996). Descendants in heap ordred trees or a triumph of computer algebra, *The Electronic Journal of Combinatorics*, 3, R29 doi:10.37236/1243

[9] Sackin, M. J. (1972). Good and bad phenograms, *Systematic Zoology,* 21, 225. doi:10.2307/2412292

[10] Shao, K. & Sokal, R. R. (1990). Tree balance, *Systematic Zoology,* 39, 266. doi:10.2307/2992186