Journal of Discrete Mathematics and Its Applications 10 (4) (2025) 321–332

Journal of Discrete Mathematics and Its Applications

Available Online at: http://jdma.sru.ac.ir

Shahid Rajaee Teacher Training University

Research Paper

Path length of protected nodes in random binary search trees

Ramin Kazemi[*], **Sedigheh Zamani Mehreyan**

Department of Statistics, Faculty of Basic Sciences, Imam Khomeini International University, Qazvin, I. R. Iran

Academic Editor: Nader Jafari-Rad

Abstract. A protected node is a node that is not a leaf and none of its children is a leaf, and also a weakly protected node is not a leaf and at least one of its children is not a leaf. Let P_n and W_n be the path length of the protected and weakly protected nodes in a random binary search tree (BST) of size n, respectively. In this paper, we derive the exact mean and variance of these random variables and show that

$$\frac{15P_n}{11n\ln n} \to 1$$

and

$$\frac{15W_n}{14n\ln n} \to 1$$

in probability.

Keywords. Binary search trees, path length, protected node, weakly protected node, limiting rule.
Mathematics Subject Classification (2020): 05C05, 60F05.

1 Introduction

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. A node with no descendants is a leaf. In a rooted tree, a protected node is a node that is not a leaf and none of its children

[*]Corresponding author (Email address: r.kazemi@sci.ikiu.ac.ir).
Received 09 August 2025; Revised 20 August 2025; Accepted 28 November 2025
First Publish Date: 01 December 2025

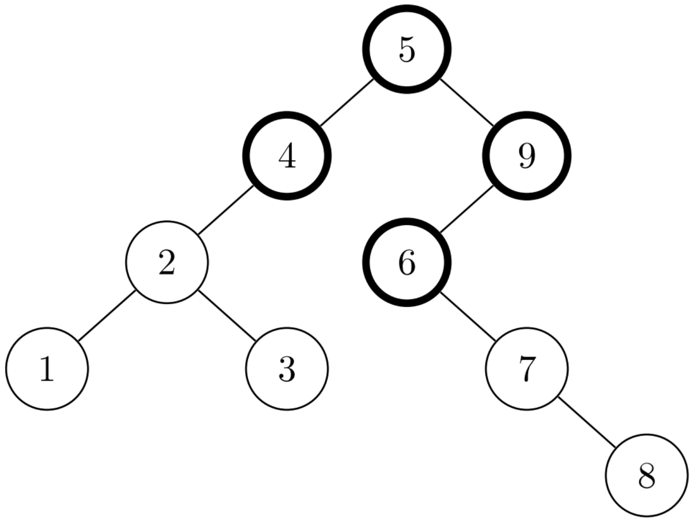

Figure 1. A BST corresponding to the permutation (5,9,6,4,7,2,3,1,8); protected nodes in bold where $P_9 = 4$ [8].

is a leaf. Also, a weakly protected node is a node that is not a leaf and at least one of its children is not a leaf [2, 4].

The binary search tree (BST) grows from a random permutation that induces a BST probability model, which is nonuniform. These models are of prime importance in computer science as it represents the backbone of some fundamental algorithms, such as Quicksort (see [5] or [6]), and are basic efficient data structures in their own right [7]. In a binary search tree (BST) grown from a uniform random permutation $(U_1, U_2, ..., U_n)$ of $\{1, 2, ..., n\}$, the first key U_1 is moved to the root node such that its left and right subtrees are distinct (which are empty so far). If the second key is smaller than the root key, it is moved to the left subtree (i.e., if $U_2 < U_1$), to the right subtree, where it is placed in a node and linked as a right child of the root. The subsequent keys go to the left or right subtrees, depending on whether they are smaller than the root key, where they are recursively inserted into the subtree by the same algorithm. If the permutations $\{1, 2, ..., n\}$ are equally likely, they lead to a non-uniform probability distribution over the BST shapes. We call such a distribution the BST probability model [8].

Figure 1 shows a BST of size 9 grown from the permutation (5,9,6,4,7,2,3,1,8). The nodes represented by bold circles are protected. Also, Figure 2 shows a BST of size 9 grown from the permutation (5,2,6,1,8,7,4,9,3). There, the nodes represented by bold circles are weakly protected [1, 11].

We denote the number of protected and weakly protected nodes in a random binary search tree of size n by X_n and Y_n, respectively. Mahmoud and Ward [8] showed that

$$\mathbb{E}(X_n) = \frac{11}{30}n - \frac{19}{30}, \quad \text{for } n \geq 4, \tag{1}$$

$$\mathbb{E}(X_n^2) = \frac{121}{900}n^2 - \frac{151}{450}n + \frac{53}{100}, \quad \text{for } n \geq 8. \tag{2}$$

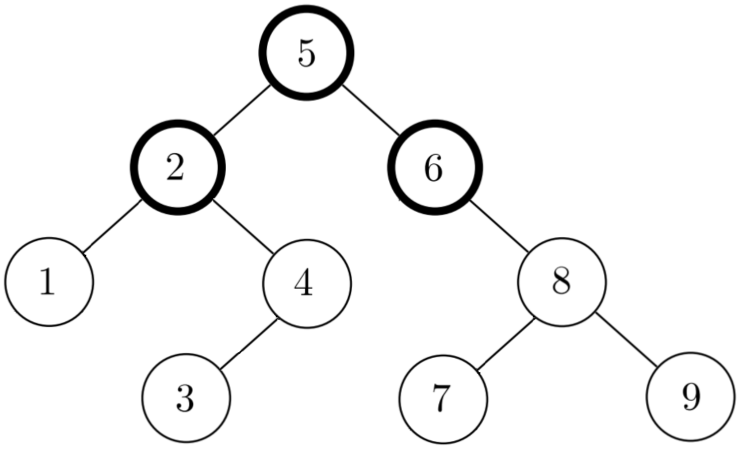

Figure 2. A BST corresponding to the permutation (5,2,6,1,8,7,4,9,3); weakly protected nodes in bold where $W_9 = 2$ [9].

Recently, Mohammad Nezhad et. al [9] showed that

$$\mathbb{E}(Y_n) = \frac{7}{15}n - \frac{8}{15}, \quad \text{for } n \geq 4, \tag{3}$$

$$\mathbb{E}(Y_n^2) = \frac{49}{225}n^2 - \frac{1357}{3150}n + \frac{123}{350}, \quad \text{for } n \geq 10. \tag{4}$$

Let D_i be the number of edges from the root to the ith node in a random BST (the depth of node i). The (internal) path length is defined as $K_n = \sum_{i=1}^{n} D_i$. Then [6],

$$\mathbb{E}(K_n) = 2(n+1)H_n - 4n,$$
$$\mathbb{V}(K_n) = O(n^2), \quad n \to \infty,$$

which $\mathbb{E}(K_n)$ is asymptotically equivalent to $2n\ln n$ and $H_n = \sum_{j=1}^{n} \frac{1}{j}$ is the n-th harmonic number.

Let P_n and W_n be the path length of the protected nodes and weakly protected nodes in a random BST of size n, respectively (see Figures 1 and 2). We derive the mean and variance of these random variables and show that

$$\frac{15P_n}{11n\ln n} \to 1$$

and

$$\frac{15W_n}{14n\ln n} \to 1$$

in probability.

2 Means

Assume that U_1 is stored in the root of a binary search tree of size n. Then $U_1 - 1$ and $n - U_1$ are the sizes of the left and right subtrees of the root, respectively. It is obvious that

U_1 is uniformly distributed over $\{1,...,n\}$. We denote by (P_{U_1}, X_{U_1}) and (P'_{n-U_1}, X'_{n-U_1}) the pairs of the path length and the protected nodes in the left and right subtrees of the root, respectively. Thus by direct enumeration we obtain the recurrence

$$P_n \stackrel{D}{=} P_{U_1-1} + P'_{n-U_1} + X_{U_1-1} + X'_{n-U_1}, \quad \text{for } n \geq 1, \tag{5}$$

where $\stackrel{D}{=}$ denotes equality in distribution. Also, let (W_{U_1}, Y_{U_1}) and (W'_{n-U_1}, Y'_{n-U_1}) be the pairs of the path length and the weakly protected nodes in the left and right subtrees of the root, respectively. Also, by direct enumeration we obtain the recurrence

$$W_n \stackrel{D}{=} W_{U_1-1} + W'_{n-U_1} + Y_{U_1-1} + Y'_{n-U_1}, \quad \text{for } \ n \geq 1. \tag{6}$$

Theorem 2.1. *Let P_n and W_n denote the path length of protected nodes and weakly protected nodes in a random binary search tree of size n. Then*

$$\mathbb{E}(P_n) = \frac{11}{15}(n+1)H_n - \frac{63}{30}n + \frac{19}{30}, \quad \textit{for } n \geq 4,$$

which $\mathbb{E}(P_n)$ is asymptotically equivalent to $\frac{11}{15} n \ln n$ and

$$\mathbb{E}(W_n) = \frac{14}{15}(n+1)H_n - \frac{72}{30}n + \frac{16}{30}, \quad \textit{for } n \geq 4,$$

which $\mathbb{E}(W_n)$ is asymptotically equivalent to $\frac{14}{15} n \ln n$.

Proof. Conditioned on U_1, the left and right subtrees have the distributions of random binary search trees of sizes $U_1 - 1$ and $n - U_1$, respectively, and are (stochastically) independent of each other. This implies that P_n, P'_n, X_n, X'_n, and U_1 are independent and we obtain the distributional recurrence

$$P_n \stackrel{D}{=} P_{U_1-1} + P'_{n-U_1} + X_{U_1-1} + X'_{n-U_1}, \quad \text{for } n \geq 1. \tag{7}$$

Note that $P_0 = P_1 = P_2 = 0$. The depth of the root node is zero. Whether the root node is protected or not has no effect on the value of $\mathbb{E}(P_n)$. Taking the expectation of (7), we have

$$\mathbb{E}(P_n) = \mathbb{E}(P_{U_1-1}) + \mathbb{E}(P'_{n-U_1}) + \mathbb{E}(X_{U_1-1}) + \mathbb{E}(X'_{n-U_1}).$$

Since $P_{U_1-1} \stackrel{D}{=} P'_{n-U_1}$ and $X_{U_1-1} \stackrel{D}{=} X'_{n-U_1}$,

$$\begin{aligned}
\mathbb{E}(P_n) &= 2\mathbb{E}(P_{U_1-1}) + 2\mathbb{E}(X_{U_1-1}) \\
&= 2\sum_{u=1}^{n} \mathbb{E}(P_{U_1-1} | U_1 = u) P(U_1 = u) \\
&+ 2\sum_{u=1}^{n} \mathbb{E}(X_{U_1-1} | U_1 = u) P(U_1 = u) \\
&= \frac{2}{n}\sum_{u=1}^{n} \mathbb{E}(P_{u-1}) + \frac{2}{n}\sum_{u=1}^{n} \mathbb{E}(X_{u-1}).
\end{aligned}$$

Thus

$$n\mathbb{E}(P_n) = 2\sum_{u=1}^{n}\mathbb{E}(P_{u-1}) + 2\sum_{u=1}^{n}\mathbb{E}(X_{u-1}).$$

If $Q_n := n\mathbb{E}(P_n)$, then

$$Q_n - Q_{n-1} = 2\mathbb{E}(P_{n-1}) + 2\mathbb{E}(X_{n-1}).$$

Hence

$$n\mathbb{E}(P_n) = (n+1)\mathbb{E}(P_{n-1}) + 2\mathbb{E}(X_{n-1}),$$

and then

$$\mathbb{E}(P_n) = \frac{n+1}{n}\mathbb{E}(P_{n-1}) + \frac{2}{n}\mathbb{E}(X_{n-1}).$$

We have

$$\frac{j}{(j+1)(j+2)} = \frac{2}{j+2} - \frac{1}{j+1} \quad \text{and} \quad \frac{1}{(j+1)(j+2)} = \frac{1}{j+1} - \frac{1}{j+2}. \tag{8}$$

For simplicity, we set

$$\alpha_j = \frac{2}{j+1}\mathbb{E}(X_j), \quad \text{for } j \geq 1.$$

From (1),

$$\alpha_j = \frac{11}{15}\frac{j}{j+1} - \frac{19}{15}\frac{1}{j+1}, \quad \text{for } j \geq 1.$$

Now, from (8) and by relation $H_{n+1} = H_n + \frac{1}{n+1}$,

$$\begin{aligned}
\sum_{j=1}^{n-1}\frac{\alpha_j}{j+2} &= \frac{11}{15}\sum_{j=1}^{n-1}\left(\frac{2}{j+2} - \frac{1}{j+1}\right) - \frac{19}{15}\sum_{j=1}^{n-1}\left(\frac{1}{j+1} - \frac{1}{j+2}\right)\\
&= \frac{11}{15}\left(2H_{n+1} - 3 - H_{n+1} + \frac{1}{n+1}\right) - \frac{19}{15}\left(H_{n+1} - \frac{1}{n+1} - H_{n+1} + \frac{1}{2}\right)\\
&= \frac{11}{15}\left(H_{n+1} - 2 + \frac{1}{n+1}\right) - \frac{19}{15}\left(\frac{1}{2} - \frac{1}{n+1}\right).
\end{aligned}$$

So,

$$\mathbb{E}(P_n) = \frac{n+1}{n}\mathbb{E}(P_{n-1}) + \alpha_{n-1}.$$

By iteration,

$$\mathbb{E}(P_n) = (n+1)\sum_{j=1}^{n-1}\frac{\alpha_j}{j+2},$$

since $P_1 = 0$. Hence,

$$\begin{aligned}
\mathbb{E}(P_n) &= (n+1)\left[\frac{11}{15}\left(H_{n+1} - 2 + \frac{1}{n+1}\right) - \frac{19}{15}\left(\frac{1}{2} - \frac{1}{n+1}\right)\right]\\
&= \frac{11}{15}(n+1)H_n - \frac{63}{30}n + \frac{19}{30},
\end{aligned}$$

and proof of the first part is completed. With the same method we can obtain the mean of W_n. The only difference is in relation (3). Since $\lim_{n\to\infty} H_n = \ln n$, asymptotic results are obtained. □

From Theorem 2.1,

$$\mathbb{E}(K_n) - \mathbb{E}(P_n) = \frac{19}{15}(n+1)H_n - \frac{57}{30}n - \frac{19}{30} \sim \frac{19}{15}n\ln n$$

and

$$\mathbb{E}(K_n) - \mathbb{E}(W_n) = \frac{16}{15}(n+1)H_n - \frac{48}{30}n - \frac{16}{30} \sim \frac{16}{15}n\ln n.$$

Also, asymptotically,

$$\mathbb{E}(P_n) \sim \frac{11}{14}\mathbb{E}(W_n).$$

This result is consistent with the definition of two types of nodes under investigation.

3 Variances

In this section, we use relations (7) and (6) to develop a recurrence for the second moments. If we raise both sides of the distribution equalities (7) and (6) to the power of two and take the expectation, expressions $\mathbb{E}(P_{U_1-1}X_{U_1-1})$ and $\mathbb{E}(W_{U_1-1}Y_{U_1-1})$ appear. Thus, to compute the variance of $\mathbb{V}(P_n)$ and $\mathbb{V}(W_n)$, we first need to evaluate $\mathbb{E}(P_nX_n)$ and $\mathbb{E}(W_nY_n)$. Let R_n be the event that the root node is not protected. Then

$$X_n \overset{D}{=} X_{U_1-1} + X'_{n-U_1} + 1 - \mathbf{I}_{R_n}, \quad \text{for } n \geq 1,$$

where $\mathbf{I}_{R_n} = 1$, if R_n occurs, and 0 otherwise [8]. When $n \geq 4$, R_n occurs only if either the left or right child of the root is a leaf.

Theorem 3.1. *For $n \geq 8$,*

$$\mathbf{E}(P_nX_n) = \begin{cases} \frac{121n^2-2225(n+1)H_n+4146n+330(n+1)(H^2_{n+1}-H^{(2)}_{n+1})}{450}, & \mathbf{I}_{R_n} = 0 \\ \frac{121n^2+1686n-665(n+1)H_n}{450}, & \mathbf{I}_{R_n} = 1 \end{cases}$$

and for $n \geq 10$,

$$\mathbb{E}(W_nY_n) = \frac{686n^2 - 9505(n+1)H_n + 16941n + 1245(n+1)(H^2_{n+1} - H^{(2)}_{n+1})}{1575},$$

where $H^{(2)}_n = \sum_{j=1}^{n}\frac{1}{j^2}$.

Proof. Since X_{U_1-1} and X'_{n-U_1} are conditionally independent (given U_1), from (5),

$$\begin{aligned} P_nX_n &= (P_{U_1-1} + P'_{n-U_1} + X_{U_1-1} + X'_{n-U_1})(X_{U_1-1} + X'_{n-U_1} + 1) \\ &= P_{U_1-1}X_{U_1-1} + P_{U_1-1}X'_{n-U_1} + P'_{n-U_1}X_{U_1-1} + P'_{n-U_1}X'_{n-U_1} \\ &+ X^2_{U_1-1} + 2X_{U_1-1}X'_{n-U_1} + X'^2_{n-U_1} + P_{U_1-1} + P'_{n-U_1} + X_{U_1-1} + X'_{n-U_1}. \end{aligned}$$

Hence

$$\begin{aligned}\mathbb{E}(P_nX_n) &= \frac{2}{n}\sum_{u=1}^{n}\mathbb{E}(P_{u-1}X_{u-1}) + \frac{2}{n}\sum_{u=1}^{n}\mathbb{E}(P_{u-1})\mathbb{E}(X_{n-u}) \\ &+ \frac{2}{n}\sum_{u=1}^{n}\mathbb{E}(X_{u-1})\mathbb{E}(X_{n-u}) + \frac{2}{n}\sum_{u=1}^{n}\mathbb{E}(X_{u-1}^2) \\ &+ \frac{2}{n}\sum_{u=1}^{n}\mathbb{E}(X_{u-1}) + \frac{2}{n}\sum_{u=1}^{n}\mathbb{E}(P_{u-1}).\end{aligned}$$

If $L_n := n\mathbb{E}(P_nX_n)$, then

$$L_n - L_{n-1} = 2\mathbb{E}(P_{n-1}X_{n-1}) + g_1(n),$$

where

$$g_1(n) = 2(\mathbb{E}(X_{n-1}^2) + \mathbb{E}(X_{n-1}) + \mathbb{E}(P_{n-1})).$$

Similar to the previous section,

$$n\mathbb{E}(P_nX_n) = (n+1)\mathbb{E}(P_{n-1}X_{n-1}) + g_1(n).$$

Then

$$\mathbb{E}(P_nX_n) = \frac{n+1}{n}\mathbb{E}(P_{n-1}X_{n-1}) + \frac{g_1(n)}{n}.$$

Since $\mathbb{E}(P_1X_1) = 0$, by iteration,

$$\mathbb{E}(P_nX_n) = (n+1)\sum_{j=1}^{n}\frac{g_1(j)}{j(j+1)}.$$

From Thoerem 2.1, relations (1) and (2),

$$g_1(n) = \frac{1}{450}\left(121n^2 + 660nH_{n-1} - 2104n + 2460\right).$$

Thus

$$\mathbb{E}(P_nX_n) = \frac{n+1}{450}\left(\sum_{j=1}^{n}\frac{121j}{j+1} + \sum_{j=1}^{n}\frac{660H_{j-1}}{j+1} - \sum_{j=1}^{n}\frac{2104}{j+1} + \sum_{j=1}^{n}\frac{2460}{j(j+1)}\right).$$

As in Theorem 2.1, we try to write the mathematical expectation $\mathbb{E}(P_nX_n)$ in terms of the harmonic function. It is not difficult to show that

$$\begin{aligned}\sum_{j=1}^{n}\frac{j}{j+1} &= n+1-H_n-1/(n+1), \\ \sum_{j=1}^{n}\frac{1}{j+1} &= H_n+1/(n+1)-1, \\ \sum_{j=1}^{n}\frac{1}{j(j+1)} &= 1-1/(n+1).\end{aligned}$$

Also, we have [4]:

$$\sum_{j=1}^{n} \frac{H_j}{j+1} = \frac{1}{2}(H_{n+1}^2 - H_{n+1}^{(2)})$$

and thus

$$\begin{aligned}\sum_{j=1}^{n} \frac{H_{j-1}}{j+1} &= \sum_{j=1}^{n} \frac{H_j - \frac{1}{j}}{j+1} \\ &= \frac{1}{2}\left(H_{n+1}^2 - H_{n+1}^{(2)}\right) - 1 + 1/(n+1).\end{aligned}$$

Then

$$\mathbb{E}(P_n X_n) = \frac{1}{450}\left(121n^2 - 2225(n+1)H_n + 4146n + 330(n+1)(H_{n+1}^2 - H_{n+1}^{(2)})\right).$$

If R_n occurs, then

$$g_1(n) = 2\mathbb{E}(X_{n-1}^2) = \frac{1}{450}(121n^2 - 544n + 900).$$

Then

$$\mathbb{E}(P_n X_n) = \frac{1}{450}\left(121n^2 + 1686n - 665(n+1)H_n\right).$$

Also

$$Y_n \overset{D}{=} Y_{U_1-1} + Y'_{n-U_1} + 1 - \delta_{n,3}\mathbf{I}_{\{U_1=1\}}, \quad \text{for } n \geq 1,$$

where $\delta_{i,j}$ is the Kronecker delta [9]. With the same calculations:

$$\mathbb{E}(W_n Y_n) = (n+1)\sum_{j=1}^{n} \frac{g_2(j)}{j(j+1)},$$

where

$$g_2(n) = \frac{1}{1575}\left(686n^2 + 2490nH_{n-1} - 8819n + 9240\right).$$

Now, the mean of $W_n Y_n$ is obtained in the same way. □

Theorem 3.2. *We have*

$$\mathbb{V}(P_n) = (n+1)\sum_{j=1}^{n} \frac{f_1(j)}{j(j+1)} - (\mathbb{E}(P_n))^2, \quad \textit{for } n \geq 8,$$

where

$$f_1(n) = 2\mathbb{E}(X_{n-1}^2) + 4\mathbb{E}(P_{n-1}X_{n-1}),$$

and

$$\mathbb{V}(W_n) = (n+1)\sum_{j=1}^{n} \frac{f_2(j)}{j(j+1)} - (\mathbb{E}(W_n))^2, \quad \textit{for } n \geq 10,$$

where

$$f_2(n) = 2\mathbb{E}(Y_{n-1}^2) + 4\mathbb{E}(W_{n-1}Y_{n-1}).$$

Proof. From (7),

$$\begin{aligned}
\mathbb{E}(P_n^2) &= \mathbb{E}(P_{U_1-1} + P'_{n-U_1} + X_{U_1-1} + X'_{n-U_1})^2 \\
&= \mathbb{E}(P_{U_1-1}^2) + \mathbb{E}(P'^2_{n-U_1}) + \mathbb{E}(X_{U_1-1}^2) + \mathbb{E}(X'^2_{n-U_1}) \\
&+ 2\mathbb{E}(P_{U_1-1}P'_{n-U_1}) + 2\mathbb{E}(P_{U_1-1}X_{U_1-1}) + 2\mathbb{E}(P_{U_1-1}X'_{n-U_1}) \\
&+ 2\mathbb{E}(P'_{n-U_1}X_{U_1-1}) + 2\mathbb{E}(P'_{n-U_1}X'_{n-U_1}) + 2\mathbb{E}(X_{U_1-1}X'_{n-U_1}) \\
&= \frac{2}{n}\sum_{u=1}^{n}\mathbb{E}(P_{u-1}^2) + \frac{2}{n}\sum_{u=1}^{n}\mathbb{E}(X_{u-1}^2) + \frac{2}{n}\sum_{u=1}^{n}\mathbb{E}(P_{u-1})\mathbb{E}(P_{n-u}) \\
&+ \frac{4}{n}\sum_{u=1}^{n}\mathbb{E}(P_{u-1}X_{u-1}) + \frac{4}{n}\sum_{u=1}^{n}\mathbb{E}(P_{u-1})\mathbb{E}(X_{n-u}) + \frac{2}{n}\sum_{u=1}^{n}\mathbb{E}(X_{u-1})\mathbb{E}(X_{n-u}).
\end{aligned}$$

If $B_n := n\mathbb{E}(P_n^2)$, then

$$B_n - B_{n-1} = 2\mathbb{E}(P_{n-1}^2) + f_1(n),$$

where

$$\begin{aligned}
f_1(n) &= 2\mathbb{E}(X_{n-1}^2) + 4\mathbb{E}(P_{n-1}X_{n-1}) \\
&= \begin{cases} \frac{605n^2+1320n(H_n^2-H_n^{(2)})-8900nH_n+15072n-6300}{450}, & \mathbf{I}_{R_n}=0 \\ \frac{605n^2+5232n-2660nH_n-2700}{450}, & \mathbf{I}_{R_n}=1 \end{cases}.
\end{aligned}$$

Hence

$$n\mathbb{E}(P_n^2) = (n+1)\mathbb{E}(P_{n-1}^2) + f_1(n).$$

Then

$$\mathbb{E}(P_n^2) = \frac{n+1}{n}\mathbb{E}(P_{n-1}^2) + \frac{f_1(n)}{n}.$$

By iteration,

$$\mathbb{E}(P_n^2) = (n+1)\sum_{j=1}^{n}\frac{f_1(j)}{j(j+1)}.$$

From Theorem 3.1, for W_n^2,

$$\begin{aligned}
f_2(n) &= 2\mathbb{E}(Y_{n-1}^2) + 4\mathbb{E}(W_{n-1}Y_{n-1}) \\
&= \frac{1}{1575}\left(3430n^2 + 4980n(H_n^2 - H_n^{(2)}) - 38020nH_n + 59547n - 23850\right), \text{ for } n \geq 10.
\end{aligned}$$

The second relation is obtained in the same way. $\square$

Corollary 3.3. *We have the following upper bounds in terms of harmonic functions [3, 10]:*

$$\sum_{j=1}^{n} \frac{H_j}{j+1} \leq (n+1)H_n,$$

$$\sum_{j=1}^{n} \frac{H_j^2}{j+1} \leq \frac{n(n+1)}{2}H_n^2 + \frac{n(n-3)}{4},$$

$$\sum_{j=1}^{n} \frac{H_j^{(2)}}{j+1} \leq \frac{n(n+1)}{2}H_n^{(2)}.$$

Now, from Thoerem 3.2 and the order of harmonic numbers [6]:

$$\mathbb{V}(P_n) = O(n^2)$$

and

$$\mathbb{V}(W_n) = O(n^2).$$

4 Convergence in probability

Based on a standard argument of Chebychev's inequality, we can show the following theorem. The reason for using this approach is that the order of the functions appearing on the right side of the Chebyshev inequality is known. In fact, if the order of the variance of the variable under consideration is less than the order of the function in the denominator, convergence in probability will be achieved [6, Page 136].

Theorem 4.1. *As $n \to \infty$,*

$$\frac{15P_n}{11n\ln n} \to 1$$

and

$$\frac{15W_n}{14n\ln n} \to 1$$

in probability.

Proof. For $\varepsilon > 0$,

$$\begin{aligned} P\left(\left|\frac{15P_n}{11n\ln n} - 1\right| \geq \varepsilon\right) &\leq \frac{\mathbb{E}\left(\frac{15P_n}{11n\ln n} - 1\right)^2}{\varepsilon^2} \\ &= \frac{\mathbb{V}(P_n)}{\varepsilon^2\left(\frac{11}{15}n\ln n\right)^2} \\ &= \frac{O(n^2)}{\varepsilon^2\left(\frac{11}{15}n\ln n\right)^2} \\ &\to 0, \quad \textit{as } n \to \infty. \end{aligned}$$

The second limit is obtained in a similar way. □

5 Conclusion

In this paper, we derived the mean and variance of path length of two types of protected nodes in random binary search trees and showed related convergences in probability.

Funding

This research received no external funding.

Data Availability Statement

No data used.

Conflicts of Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

[1] M. Bona, k-protected nodes in binary search trees. Adv. Appl. Math. 53 (2014) 1–11. https://doi.org/10.1016/j.aam.2013.09.003

[2] A. M. Frieze, M. Karonski, Introduction to Random Graphs, 1st Edition, Cambridge University Press, 2016. https://doi.org/10.1017/mag.2017.110

[3] R. Kazemi, On the Zagreb index of random m-oriented recursive trees, TOC. 12(2) (2023) 93–105. https://doi.org/10.22108/TOC.2022.132750.1964

[4] R. Kazemi, L. Meimondari, Degree distance and Gutman index of increasing trees, TOC. 5(2) (2016) 23–31. https://doi.org/10.22108/toc.2016.9915

[5] D. E. Knuth, The Art of Computer Programming, second ed., in: Sorting and Searching, vol. 3, Addison-Wesley, Reading, MA, 1998, Originally published in 1973. https://dl.acm.org/doi/book/10.5555/280635

[6] H. M. Mahmoud, Sorting: A Distribution Theory, Wiley, New York, NY, 2000. https://doi.org/10.1002/9781118032886

[7] H. M. Mahmoud, Evolution of Random Search Trees, Wiley, New York, NY, 1992. https://www. semanticscholar .org/ paper/ Evolution-of-random-search-trees- Mahmoud/ 965a46ba9d339bad3afde4b1468a60ed99c1725e

[8] H. M. Mahmoud, M. D. Ward, Asymptotic distribution of two-protected nodes in random binary search trees, Appl. Math. Lett. 25(12) (2012) 2218–2222. https://doi.org/10.1016/j.aml.2012.06.005

[9] E. Mohammad Nezhad, M. Javanian, R. Imany Nabiyyi, Weakly protected nodes in random binary search trees, RAIRO-Theor. Inf. Appl. 56(2) (2022) 1–8. https://doi.org/https://doi.org/10.1051/ita/2022002

[10] J. Spiess, Some identities involving harmonic numbers, Math. Comput. 55(192) (1990) 839–863. https://doi.org/10.2307/2008451

[11] L. Yang, S. L. Yang, Weakly protected points in ordered trees, Graphs Combin. 37(3) (2021) 775–788. https://doi.org/10.1007/s00373-021-02278-w

Citation: R. Kazemi, S. Zamani Mehreyan, Path length of protected nodes in random binary search trees, J. Disc. Math. Appl. 10(4) (2025) 321–332.

https://doi.org/10.22061/jdma.2025.12370.1151