

## Research Article

# Normal Limiting Distribution of the Size of Binary Interval Trees

Jie Liu<sup>1</sup> and Yang Yang<sup>2,3</sup>

<sup>1</sup>Department of Statistics and Finance, University of Science and Technology of China, Hefei 230026, China

<sup>2</sup>Department of Statistics, Nanjing Audit University, Nanjing 211815, China

<sup>3</sup>The Key Lab of Financial Engineering of Jiangsu Province, Nanjing Audit University, Nanjing 211815, China

Correspondence should be addressed to Yang Yang; [yyangmath@gmail.com](mailto:yyangmath@gmail.com)

Received 1 August 2015; Revised 3 September 2015; Accepted 9 September 2015

Academic Editor: Xinguang Zhang

Copyright © 2015 J. Liu and Y. Yang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The limiting distribution of the size of binary interval tree is investigated. Our illustration is based on the contraction method, and it is quite different from the case in one-sided binary interval tree. First, we build a distributional recursive equation of the size. Then, we draw the expectation, the variance, and some high order moments. Finally, it is shown that the size (with suitable standardization) approaches the standard normal random variable in the Zolotarev metric space.

## 1. Introduction

Random trees are usually generated based on combinatorics and occur also in the context of algorithms from computer science. There are many kinds of random trees with different structures, such as recursive trees, search trees, binary trees, and interval trees. The asymptotic probability behavior of random variables in random trees has attracted more scholars' attention and has become a popular research area. Drmota [1] introduced some labelled and unlabelled random trees in his book. Devroye and Janson [2] studied the protected nodes in several random trees. Feng and Hu [3] researched the phase changes of scale-free trees. The limiting law for the height, size, and subtree of binary search trees was also considered (see [4–6]). There were also some researchers investigating the Zagreb index and nodes of random recursive trees (see [7–9]).

Binary interval tree is a random structure that underlies the process of random division of a line interval and parking problems. It has recently been a popular subject. Sibuya and Itoh [10] showed that the number of internal and external nodes in different levels of binary interval tree is asymptotically normal, from which the asymptotic normality of the size of the tree could not be achieved directly. Prodinger [11] looked into various parameters of the incomplete trie,

a one-sided version of a random tree with a digital flavor. Fill et al. [12] followed with a study of the nonexistence of limit distribution for the height of the incomplete trie. Itoh and Mahmoud [13] considered five incomplete one-sided variants of binary interval trees and proved that their sizes all approach some normal random variables. Janson [14] drew the same result for a larger scale of one-sided interval trees by the renewal theory, and one kind of fragmentation trees was discussed by Janson and Neininger [15]. Javanian et al. [16] investigated the paths in  $m$ -ary interval trees. Su et al. [17] studied the complete binary interval trees and got the Law of Large Numbers. In addition, Pan et al. [18] considered the construction algorithm about binary interval trees.

The binary interval tree is a tree associated with repeated divisions of a line interval of length  $x$ . The process of divisions is as follows. If  $x < 1$ , there is no division in effect; the associated interval tree consists only of one terminal node. Supposing that  $x \geq 1$ , we begin with the interval  $(0, x)$ . Divide the interval  $(0, x)$  into two subintervals by choosing  $U_x$ , a point uniformly distributed over the interval  $(0, x)$ . Then, we get two intervals,  $(0, U_x)$  and  $(U_x, x)$ . Each of the two subintervals is further divided at a uniform point of its length, and two smaller subintervals are got as before. If the length of the subinterval is less than 1, we stop the division. Repeat this

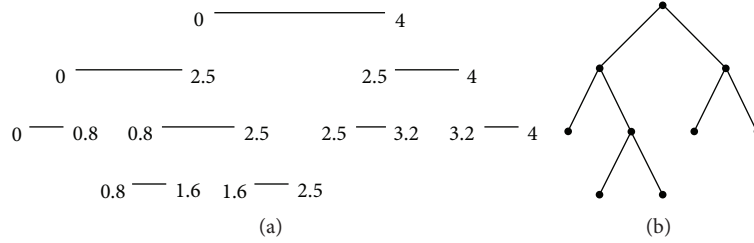


FIGURE 1: (a) The division process. (b) The binary interval tree.

process until the length of every interval (or subinterval) is less than 1.

We take  $x = 4$ , for instance. Figures 1(a) and 1(b) show how the above random division process of interval generates a binary interval tree.

If some different conditions are added and those intervals satisfying the conditions are not allowed to be divided (see [13, 14]), then we can get different incomplete interval trees. In particular, if we only divide one subinterval of every interval, then the interval tree we get is the so-called one-sided interval tree (see [13]).

It is obvious that interval tree could embody many properties of random division, so it can elicit lots of valuable subjects related to probability. For example, for  $x > 0$ , the height of the interval tree is the greatest level of all subintervals after the divisions, denoted by  $H_x$ ; the total number of nodes of an interval tree is the total number of intervals that were got from the random division process, and so on. Let  $S_x$  be the size of the interval trees, that is, the total number of nodes of the binary interval trees. Our intention is to investigate the random variable  $S_x$ , the size of binary interval trees.

In this paper, the central limit theorem of the size of binary interval trees is investigated. In view of the difficulty to calculate the moment generating function of  $S_x$ , the method we used is completely different from that in the case of one-sided interval trees. In Section 2, we build a distributional recursive equation of  $S_x$  and give the expectation, the variance, and some high order moments of  $S_x$ . In Section 3, via the contraction method, the limit law of  $S_x$  is shown to approach the unique solution of a fixed-point distributional equation in the Zolotarev metric space. Finally, we demonstrate that  $S_x$ , with suitable standardization, converges to a normal limiting random variable, as  $x \rightarrow \infty$ .

## 2. The Moments of $S_x$

Compared with the one-sided interval trees, the properties of binary interval trees are much more complex. There are a lot of difficulties when it comes to obtaining the moment generating function of  $S_x$ . Therefore, the method used in the case of one-sided interval trees (see [13]) is no longer applicable. Here, we build a distributional recursive equation of  $S_x$ . We can calculate the expectation and the variance of  $S_x$ . Furthermore, we find that the order of the fourth central moment of  $S_x$  is  $O(x^2)$  as  $x$  goes to infinity.

From the definition of binary interval tree, it is easy to see that  $S_1 = 3$  and  $S_x = 1$ , for  $x < 1$ . For our purpose to investigate the case of  $x \geq 1$ , let  $U_x$  denote the point chosen uniformly from interval  $(0, x)$ ; hence,  $U_x \sim U(0, x)$ . For any fixed real number  $0 < u < x$ , if  $U_x = u$ , we denote  $S_u^{(1)}$  to be the size of the left subtree associated with the interval  $(0, u)$ . Correspondingly,  $S_{x-u}^{(2)}$  denotes the size of the right subtree associated with the interval  $(u, x)$ . According to the rule of division, we can see that  $S_u^{(1)}$  and  $S_{x-u}^{(2)}$  are mutually independent. Thus, we have

$$S_x |_{U=u} \stackrel{d}{=} 1 + S_u^{(1)} + S_{x-u}^{(2)}, \quad \forall 0 < u < x. \quad (1)$$

This formula implies that if  $U_x = u$  is given,  $S_x$  has the same distribution as  $1 + S_u^{(1)} + S_{x-u}^{(2)}$ . Obviously, we can rewrite the above formula as

$$S_x \stackrel{d}{=} 1 + S_{U_x}^{(1)} + S_{x-U_x}^{(2)}. \quad (2)$$

Define

$$\begin{aligned} m_1(x) &:= \mathbf{E}S_x; \\ m_2(x) &:= \mathbf{E}(S_x)^2. \end{aligned} \quad (3)$$

It is easy to see that

$$\begin{aligned} m_1(1) &= \mathbf{E}S_1 = 3; \\ m_2(1) &= 9; \\ m_1(x) &= m_2(x) = 1, \quad 0 < x < 1. \end{aligned} \quad (4)$$

From the distributional recursive equation (2) and the above boundary conditions, Su et al. [17] calculated the expectation  $\mathbf{E}S_x$  and the variance  $\mathbf{Var} S_x$ , for any  $x \geq 0$ .

**Lemma 1.** Let  $S_x$  be the size of a binary interval tree. Then

$$\mathbf{E}S_x = m_1(x) = 4x - 1, \quad x \geq 1. \quad (5)$$

**Lemma 2.** Let  $S_x$  be the size of a binary interval tree. Then

$$\mathbf{Var} S_x = \begin{cases} 32x \ln x - 16x^2 + 8x + 8, & 1 \leq x \leq 2, \\ (32 \ln 2 - 20)x, & x \geq 2. \end{cases} \quad (6)$$

In order to prove that the asymptotic distribution of  $S_x$  is normal, we also need the order of  $\mathbf{E}(S_x - \mathbf{E}S_x)^4$  as  $x \rightarrow \infty$ . The following proposition shows the fourth central moment of  $S_x$ .

**Proposition 3.** Let  $S_x$  be the size of a binary interval tree. Then

$$\mathbf{E}(S_x - \mathbf{E}[S_x])^4 = O(x^2), \quad x \rightarrow \infty. \quad (7)$$

*Proof.* See the appendix.  $\square$

### 3. The CLT for $S_x$

In this section, we will prove the asymptotic normality of  $S_x$  as  $x \rightarrow \infty$ . The main method is the contraction method and some metrics are needed especially the Zolotarev metrics (see [19]).

First we introduce the Zolotarev metrics. Denote the distribution of the random variable  $X$  by  $\mathcal{L}(X)$ . Let  $\mathcal{D}$  be the set of the distributions of all real random variables, and define

$$\begin{aligned} \mathcal{D}^* &= \left\{ F : F \in \mathcal{D}, \int_{\mathcal{R}} x dF(x) = 0, \int_{\mathcal{R}} x^2 dF(x) \right. \\ &= 1, \left. \int_{\mathcal{R}} |x|^3 dF(x) < \infty \right\}. \end{aligned} \quad (8)$$

It can be verified that random variable  $Z$  with  $\mathcal{L}(Z) = \mathcal{N}(0, \sigma^2)$  satisfies the following formula. For any  $u \in [0, 1]$ ,

$$Z \stackrel{d}{=} Z\sqrt{u} + \bar{Z}\sqrt{1-u}, \quad (9)$$

and more generally, we have the following lemma.

**Lemma 4.** If  $Z$  and  $\bar{Z}$  are standard normal random variables,  $U$  is uniformly distributed over interval  $[0, 1]$ , and  $(U, Z, \bar{Z})$  are mutually independent and then one has

$$Z \stackrel{d}{=} Z\sqrt{U} + \bar{Z}\sqrt{1-U}. \quad (10)$$

*Proof.* In fact, for any  $u \in [0, 1]$ , we have

$$\begin{aligned} &\mathbf{E} \exp \left\{ it \left( \sqrt{u}Z + \sqrt{1-u}\bar{Z} \right) \right\} \\ &= \mathbf{E} \exp \left\{ it \left( \sqrt{u}Z + \sqrt{1-u}\bar{Z} \right) \right\} \\ &= \mathbf{E} e^{i(t\sqrt{u})Z} \mathbf{E} e^{i(t\sqrt{1-u})\bar{Z}} \\ &= \exp \left\{ -\frac{ut^2}{2} \right\} \exp \left\{ -\frac{(1-u)t^2}{2} \right\} = \exp \left\{ -\frac{t^2}{2} \right\}. \end{aligned} \quad (11)$$

Therefore,

$$\begin{aligned} &\mathbf{E} \exp \left\{ it \left( \sqrt{U}Z + \sqrt{1-U}\bar{Z} \right) \right\} \\ &= \int_0^1 \mathbf{E} \exp \left\{ it \left( \sqrt{u}Z + \sqrt{1-u}\bar{Z} \right) \right\} du \\ &= \int_0^1 \exp \left\{ -\frac{t^2}{2} \right\} du = \exp \left\{ -\frac{t^2}{2} \right\}. \end{aligned} \quad (12)$$

But, we can find that, in the set  $\mathcal{D}^*$ , there is only one distribution, the standard normal  $\mathcal{N}(0, 1)$ , satisfying (10).  $\square$

Suppose that  $m$  is a nonnegative integer. Denote  $\mathfrak{F}^{(m)}$  by the set of all real functions that are  $m$  times continuous and differentiable, defined on the real line. Let

$$\begin{aligned} &\mathfrak{F}_\alpha^{(m)} \\ &:= \left\{ f : f \in \mathfrak{F}^{(m)}, \left| f^{(m)}(x) - f^{(m)}(y) \right| \leq |x - y|^\alpha \right\}, \end{aligned} \quad (13)$$

where  $0 < \alpha \leq 1$  is a fixed real number. Let  $s = m + \alpha$  and

$$\begin{aligned} \zeta_s(X, Y) &:= \zeta_s(\mathcal{L}(X), \mathcal{L}(Y)) \\ &= \sup \left\{ \left| \mathbf{E}f(X) - \mathbf{E}f(Y) \right| : f \in \mathfrak{F}_\alpha^{(m)} \right\}, \end{aligned} \quad (14)$$

and then  $\zeta_s$  is the Zolotarev metrics with order  $s$  on the set  $\mathcal{D}$ . According to the properties of the Zolotarev metric, we know

$$\begin{aligned} \zeta_s(X, Y) < \infty \\ \iff \mathbf{E}|X|^s + \mathbf{E}|Y|^s < \infty, \end{aligned} \quad (15)$$

$$\mathbf{E}X^k = \mathbf{E}Y^k, \quad k = 1, \dots, m.$$

Therefore, we can choose  $\zeta_3$  as the metric we need on the subset  $\mathcal{D}^*$  (see [20, 21]); that is,  $m = 2, \alpha = 1$ . This is due to the fact that, for any  $\mathcal{L}(X) \in \mathcal{D}^*$  and  $\mathcal{L}(Y) \notin \mathcal{D}^*$ , we have  $\zeta_3(X, Y) = \infty$ , but if  $\mathcal{L}(X), \mathcal{L}(Y) \in \mathcal{D}^*$ , then  $\zeta_3(X, Y) < \infty$ .

The metric  $\zeta_s(X, Y)$  has several properties as follows (see [20]):

(1) For any constant  $c > 0$ ,

$$\zeta_s(cX, cY) = c^s \zeta_s(X, Y); \quad (16)$$

(2) if random variables  $Y$  and  $(X_1, X_2)$  are mutually independent, then

$$\zeta_s(X_1 + Y, X_2 + Y) \leq \zeta_s(X_1, X_2); \quad (17)$$

(3) for random variables  $X$  and  $\{X_n, n = 1, 2, 3, \dots\}$ ,

$$\zeta_s(X_n, X) \rightarrow 0 \implies X_n \xrightarrow{d} X. \quad (18)$$

Now, we begin to prove the main result in this paper.

**Theorem 5.** Let  $S_x$  be the size of a binary interval tree. Then, as  $x \rightarrow \infty$ ,

$$\frac{S_x - \mathbf{E}S_x}{\sqrt{\mathbf{Var}S_x}} \xrightarrow{d} \mathcal{N}(0, 1). \quad (19)$$

*Proof.* Denote

$$\begin{aligned} S_x^* &:= \frac{S_x - (4x - 1)}{\sqrt{(32 \ln 2 - 20)x}}, \quad x > 0, \\ h(x) &:= \sqrt{\frac{32x \ln x - 16x^2 + 8x + 8}{(32 \ln 2 - 20)x}}, \quad x > 0. \end{aligned} \quad (20)$$

Then from Lemmas 1 and 2, we know that

$$S_x^* = \begin{cases} \frac{S_x - \mathbf{E} S_x}{\sqrt{\mathbf{Var} S_x}}, & x \geq 2, \\ \frac{S_x - \mathbf{E} S_x}{\sqrt{\mathbf{Var} S_x}} \cdot h(x), & 1 \leq x < 2, \\ \frac{2 - 4x}{\sqrt{(32 \ln 2 - 20)x}}, & 0 < x < 1. \end{cases} \quad (21)$$

So, we have  $\mathcal{L}(S_x^*) \in \mathcal{D}^*$  for  $x \geq 2$  and  $\mathcal{L}(S_x^*) \notin \mathcal{D}^*$  for  $0 < x < 2$ .

According to the correlative inequality in [21], for any  $\mathcal{L}(X) \in \mathcal{D}^*$ ,  $\mathcal{L}(Y) \in \mathcal{D}^*$ ,

$$\zeta_3(X, Y) \leq \frac{\Gamma(2)}{\Gamma(4)} \int_{\mathcal{R}} |t|^3 d|P(X < t) - P(Y < t)|, \quad (22)$$

where  $\Gamma$  is the gamma function. Assume that the distribution of random variable  $Z$  is  $\mathcal{N}(0, 1)$ . It follows from Proposition 3 that

$$\sup_{x \geq 4} \mathbf{E} (S_x^*)^4 < \infty. \quad (23)$$

Therefore, there exists a constant  $C > 0$  such that

$$\sup_{x \geq 4} \zeta_3(S_x^*, Z) \leq C \left( \sup_{x \geq 4} \mathbf{E} |S_x^*|^3 + \mathbf{E} |Z|^3 \right) < \infty. \quad (24)$$

Denote

$$a_x := \zeta_3(\mathcal{L}(S_x^*), \Phi) = \zeta_3(S_x^*, Z), \quad (25)$$

where  $\Phi$  is standard normal distribution and  $Z$  is standard normal random variable; then we can see that

$$0 \leq b := \limsup_{x \rightarrow \infty} a_x < \infty. \quad (26)$$

Now, we just need to prove that  $b = 0$ ; then the theorem follows.

Suppose that  $x \geq 4$ ; by (A.1) and (21), we have

$$\begin{aligned} S_x^* |_{U_x=t} &= \frac{S_x - (4x - 1)}{\sqrt{(32 \ln 2 - 20)x}} \Big|_{U_x=t} \\ &= \begin{cases} \frac{S_t^{(1)} - \mathbf{E}[S_t^{(1)}]}{\sqrt{(32 \ln 2 - 20)x}} + \frac{S_{x-t}^{(2)} - \mathbf{E}[S_{x-t}^{(2)}]}{\sqrt{(32 \ln 2 - 20)x}}, & 2 \leq t \leq x-2, \\ \frac{S_t^{(1)} - \mathbf{E}[S_t^{(1)}]}{\sqrt{(32 \ln 2 - 20)x}} + \frac{S_{x-t}^{(2)} - \mathbf{E}[S_{x-t}^{(2)}]}{\sqrt{(32 \ln 2 - 20)x}}, & 1 \leq t < 2, \\ \frac{S_t^{(1)} - \mathbf{E}[S_t^{(1)}]}{\sqrt{(32 \ln 2 - 20)x}} + \frac{S_{x-t}^{(2)} - \mathbf{E}[S_{x-t}^{(2)}]}{\sqrt{(32 \ln 2 - 20)x}}, & x-2 < t \leq x-1, \\ \frac{S_{x-t}^{(2)} - \mathbf{E}[S_{x-t}^{(2)}] - (4t-2)}{\sqrt{(32 \ln 2 - 20)x}}, & 0 < t < 1 \\ \frac{S_t^{(1)} - \mathbf{E}[S_t^{(1)}] - [4(x-t)-2]}{\sqrt{(32 \ln 2 - 20)x}}, & x-1 < t < x \end{cases} \quad (27) \\ &= \begin{cases} S_t^* \sqrt{\frac{t}{x}} + \tilde{S}_{x-t}^* \sqrt{\frac{x-t}{x}}, & 2 \leq t \leq x-2; \\ S_t^* h(t) \sqrt{\frac{t}{x}} + \tilde{S}_{x-t}^* \sqrt{\frac{x-t}{x}}, & 1 \leq t < 2, \\ S_t^* \sqrt{\frac{t}{x}} + \tilde{S}_{x-t}^* h(x-t) \sqrt{\frac{x-t}{x}}, & x-2 < t \leq x-1, \\ \tilde{S}_{x-t}^* \sqrt{\frac{x-t}{x}} - \frac{4t-2}{\sqrt{(32 \ln 2 - 20)x}}, & 0 < t < 1; \\ S_t^* \sqrt{\frac{t}{x}} - \frac{4(x-t)-2}{\sqrt{(32 \ln 2 - 20)x}}, & x-1 < t < x, \end{cases} \end{aligned}$$

where  $U_x = t$  is the first point chosen from interval  $(0, x)$  and  $\{\tilde{S}_x^*, x > 0\}$  is an independent copy of  $\{S_x^*, x > 0\}$ .

If we denote  $U := U_x/x$ , then  $U \sim U(0, 1)$  and we can rewrite the above formula as

$$\begin{aligned} S_x^* |_{U=u} &= \frac{S_x - (4x - 1)}{\sqrt{(32 \ln 2 - 20)x}} \Big|_{U=u} \quad (28) \\ &= \begin{cases} S_{ux}^* \sqrt{u} + \tilde{S}_{(1-u)x}^* \sqrt{1-u}, & \frac{2}{x} \leq u \leq 1 - \frac{2}{x}; \\ S_{ux}^* h(ux) \sqrt{u} + \tilde{S}_{(1-u)x}^* \sqrt{1-u}, & \frac{1}{x} \leq u < \frac{2}{x}, \\ S_{ux}^* \sqrt{u} + \tilde{S}_{(1-u)x}^* h((1-u)x) \sqrt{1-u}, & 1 - \frac{2}{x} < u \leq 1 - \frac{1}{x}, \\ \tilde{S}_{(1-u)x}^* \sqrt{1-u} - \frac{4ux-2}{\sqrt{(32 \ln 2 - 20)x}}, & 0 < u < \frac{1}{x}; \\ S_{ux}^* \sqrt{u} - \frac{4(1-u)x-2}{\sqrt{(32 \ln 2 - 20)x}}, & 1 - \frac{1}{x} < u < 1. \end{cases} \quad (29) \end{aligned}$$

According to the definition of  $\mathcal{D}^*$  and  $\zeta_3$ , it could be found that  $(S_x^* |_{U_x=t}) \in \mathcal{D}^*$  for  $2 < t < x-2$  and  $S_x^* \in \mathcal{D}^*$ .

If we define  $S'_x := (S_x^* \mid U_x < 2 \text{ or } U_x > x - 2)$ , then we can also see that  $S'_x \in \mathcal{D}^*$ . Furthermore,  $\mathbf{E}((S'_x)^4) \leq C_1$  for some positive constant  $C_1$  by conditioning on  $U_x$  and using the similar calculation in the appendix. Hence,

$$\zeta_3(S'_x, Z) \leq \beta, \quad (30)$$

for some positive constant  $\beta$ .

As we had pointed out before, the standard normal distribution is the only distribution satisfying (10) in the set  $\mathcal{D}^*$ . From (25), (14), and Lemma 4, for  $x > 4$ , we have

$$\begin{aligned} a_x &= \zeta_3(S_x^*, Z) \leq \zeta_3(S'_x, Z) \cdot \frac{4}{x} \\ &+ \int_{2/x}^{1-2/x} \zeta_3((S_x^* \mid U = u), Z) du \leq \frac{4\beta}{x} \\ &+ \int_{2/x}^{1-2/x} \zeta_3(S_{xu}^* \sqrt{u} + \bar{S}_{x(1-u)}^* \sqrt{1-u}, Z \sqrt{u} \\ &+ \bar{Z} \sqrt{1-u}) du \quad (\text{By (15) and (29)}) \\ &\leq \frac{4\beta}{x} + \int_{2/x}^{1-2/x} \zeta_3(S_{xu}^* \sqrt{u} + \bar{S}_{x(1-u)}^* \sqrt{1-u}, Z \sqrt{u} \\ &+ \bar{S}_{x(1-u)}^* \sqrt{1-u}) du + \int_{2/x}^{1-2/x} \zeta_3(Z \sqrt{u} \\ &+ \bar{S}_{x(1-u)}^* \sqrt{1-u}, Z \sqrt{u} + \bar{Z} \sqrt{1-u}) du \\ &\leq \frac{4\beta}{x} + \int_{2/x}^{1-2/x} \zeta_3(S_{xu}^* \sqrt{u}, Z \sqrt{u}) du \\ &+ \int_{2/x}^{1-2/x} \zeta_3(\bar{S}_{x(1-u)}^* \sqrt{1-u}, \bar{Z} \sqrt{1-u}) du \\ &= \frac{4\beta}{x} + 2 \int_{2/x}^{1-2/x} \zeta_3(S_{xu}^* \sqrt{u}, Z \sqrt{u}) du \\ &= \frac{4\beta}{x} + 2 \int_{2/x}^{1-2/x} u^{3/2} \zeta_3(S_{xu}^*, Z) du \\ &= \frac{4\beta}{x} + 2 \int_{2/x}^{1-2/x} u^{3/2} a_{xu} du. \end{aligned} \quad (31)$$

Given  $\varepsilon > 0$ , let  $\delta > 0$  be small enough such that  $\beta\delta^{5/2} < \varepsilon/8$ . For any fixed  $\delta > 0$ , when  $x$  is sufficiently large, then

$$\begin{aligned} \frac{4\beta}{x} &< \frac{\varepsilon}{10}, \\ \frac{2}{x} &< \delta, \\ \sup_{\delta \leq u \leq 1} a_{xu} &< b + \varepsilon. \end{aligned} \quad (32)$$

Thus,

$$\begin{aligned} 2 \int_{2/x}^{\delta} u^{3/2} a_{xu} du &\leq 2\beta \int_{2/x}^{\delta} u^{3/2} du \leq 2\beta \int_0^{\delta} u^{3/2} du \\ &= \frac{4\beta\delta^{5/2}}{5} < \frac{\varepsilon}{10}; \end{aligned} \quad (33)$$

$$\begin{aligned} 2 \int_{\delta}^{1-2/x} u^{3/2} a_{xu} du &\leq 2(b + \varepsilon) \int_{\delta}^{1-2/x} u^{3/2} du \\ &\leq 2(b + \varepsilon) \int_0^1 u^{3/2} du < \frac{4(b + \varepsilon)}{5}, \end{aligned}$$

where  $\beta$  is the constant as before and  $x$  is sufficiently large. It implies that

$$\begin{aligned} a_x &\leq \frac{4\beta}{x} + 2 \int_{2/x}^{\delta} u^{3/2} a_{xu} du + 2 \int_{\delta}^{1-2/x} u^{3/2} a_{xu} du \\ &< \frac{\varepsilon}{10} + \frac{\varepsilon}{10} + \frac{4(b + \varepsilon)}{5} < \frac{4b}{5} + \varepsilon, \end{aligned} \quad (34)$$

when  $x$  is sufficiently large. Therefore,

$$b := \limsup_{x \rightarrow \infty} a_x \leq \frac{4b}{5} + \varepsilon. \quad (35)$$

From this equation and the arbitrariness of  $\varepsilon > 0$ , we can conclude  $b = 0$  and

$$\lim_{x \rightarrow \infty} \zeta_3(S_x^*, Z) = \lim_{x \rightarrow \infty} a_x = 0 \quad (36)$$

immediately. By (18), the theorem holds.  $\square$

## Appendix

### Proof of Proposition 3

From the process of generating the binary interval trees, it is obvious that, for given  $U_x = t$ ,  $t \in (0, x)$ ,

$$\begin{aligned} (S_x - \mathbf{E}[S_x])|_{U_x=t} &= \begin{cases} (S_t^{(1)} - \mathbf{E}[S_t^{(1)}]) + (S_{x-t}^{(2)} - \mathbf{E}[S_{x-t}^{(2)}]), & 1 \leq t \leq x-1; \\ S_{x-t}^{(2)} - \mathbf{E}[S_{x-t}^{(2)}] - (4t-2), & 0 < t < 1; \\ S_t^{(1)} - \mathbf{E}[S_t^{(1)}] - [4(x-t)-2], & x-1 < t < x, \end{cases} \end{aligned} \quad (\text{A.1})$$

where  $U_x = t$  is the first point chosen from interval  $(0, x)$ . For  $x \geq 1$ , if we denote

$$T_x := S_x^{(1)} - \mathbf{E}[S_x^{(1)}], \quad (\text{A.2})$$

$$T_x^* := S_x^{(2)} - \mathbf{E}[S_x^{(2)}],$$

then we have

$$T_x|_{U_x=t} \stackrel{d}{=} \begin{cases} T_t + T_{x-t}^*, & 1 \leq t \leq x-1; \\ T_{x-t}^* - (4t-2), & 0 < t < 1; \\ T_t - [4(x-t)-2], & x-1 < t < x. \end{cases} \quad (\text{A.3})$$

We need to calculate  $\mathbf{E}T_x^3$  first before we get  $\mathbf{E}T_x^4$ . For  $x > 3$ , we have

$$\begin{aligned}
\mathbf{E}[T_x]^3 &= \mathbf{E}[\mathbf{E}(T_x^3 | U_x)] = \frac{1}{x} \int_0^1 \mathbf{E}[T_{x-t} \\
&\quad - (4t - 2)]^3 dt + \frac{1}{x} \int_{x-1}^x \mathbf{E}\{T_t \\
&\quad - [4(x-t) - 2]\}^3 dt + \frac{1}{x} \int_1^{x-1} \mathbf{E}[T_t + T_{x-t}^*]^3 dt \\
&= \frac{2}{x} \int_{x-1}^x \mathbf{E}\{T_t - [4(x-t) - 2]\}^3 dt + \frac{1}{x} \\
&\quad \cdot \int_1^{x-1} \mathbf{E}[T_t + T_{x-t}^*]^3 dt = \left(-\frac{2}{x} \right. \\
&\quad \cdot \int_{x-1}^x [4(x-t) - 2]^3 dt + \frac{6}{x} \\
&\quad \cdot \int_{x-1}^x [4(x-t) - 2]^2 \mathbf{E}[T_t] dt - \frac{6}{x} \\
&\quad \cdot \int_{x-1}^x [4(x-t) - 2] \mathbf{E}[T_t^2] dt + \frac{2}{x} \\
&\quad \cdot \int_{x-1}^x \mathbf{E}[T_t^3] dt) + \frac{1}{x} \int_1^{x-1} (\mathbf{E}[T_t]^3 \\
&\quad + 3\mathbf{E}[(T_t)^2 T_{x-t}^*] + 3\mathbf{E}[T_t (T_{x-t}^*)^2] \\
&\quad + \mathbf{E}[T_{x-t}^*]^3) dt.
\end{aligned} \tag{A.4}$$

In view of the independence between  $T_t$  and  $T_{x-t}^*$  and that  $\mathbf{E}[T_t] = \mathbf{E}[T_t^*] = 0$  holds for any  $1 \leq t \leq x-1$ , we have

$$\begin{aligned}
\mathbf{E}[T_x]^3 &= -\frac{2}{x} \int_{x-1}^x [4(x-t) - 2]^3 dt \\
&\quad - \frac{6}{x} \int_{x-1}^x [4(x-t) - 2] \mathbf{E}[T_t^2] dt \\
&\quad + \frac{2}{x} \int_{x-1}^x \mathbf{E}[T_t^3] dt + \frac{2}{x} \int_1^{x-1} \mathbf{E}[T_t^3] dt \\
&= -\frac{2}{x} \int_{x-1}^x [4(x-t) - 2]^3 dt \\
&\quad - \frac{6}{x} \int_{x-1}^x [4(x-t) - 2] \mathbf{E}[T_t^2] dt \\
&\quad + \frac{2}{x} \int_1^x \mathbf{E}[T_t^3] dt \\
&:= M_1 + M_2 + \frac{2}{x} \int_1^x \mathbf{E}[T_t^3] dt.
\end{aligned} \tag{A.5}$$

It is easy to see that

$$M_1 = -\frac{1}{2x} \int_{-2}^2 u^3 du = 0, \tag{A.6}$$

and when  $x > 3$ , for the part  $M_2$ , we have

$$\begin{aligned}
M_2 &= -\frac{6}{x} \int_{x-1}^x [4(x-t) - 2] \mathbf{Var}[S_t] dt \\
&= -\frac{6}{x} \int_{x-1}^x [4(x-t) - 2] [(32 \ln 2 - 20)t] dt \\
&= -\frac{6(32 \ln 2 - 20)}{x} \int_0^1 (4t - 2)(x-t) dt \\
&= \frac{2(32 \ln 2 - 20)}{x}.
\end{aligned} \tag{A.7}$$

Therefore,

$$\mathbf{E}[T_x^3] = \frac{2}{x} \int_1^x \mathbf{E}[T_t^3] dt + \frac{2(32 \ln 2 - 20)}{x}, \quad x > 3. \tag{A.8}$$

That is,

$$x\mathbf{E}[T_x^3] = 2 \int_1^x \mathbf{E}[T_t^3] dt + 2(32 \ln 2 - 20), \tag{A.9}$$

$x > 3.$

Via differentiation with respect to  $x$ , we get the differential equation:

$$(\mathbf{E}[T_x^3])' - \frac{1}{x}\mathbf{E}[T_x^3] = 0, \quad x > 3. \tag{A.10}$$

The solution to this differential equation is

$$\mathbf{E}[T_x^3] = k_0 x, \quad x > 3, \tag{A.11}$$

where  $k_0$  is a constant real number.

Similarly, for  $\mathbf{E}[T_x]^4$ , when  $x > 4$ , we have

$$\begin{aligned}
\mathbf{E}[T_x]^4 &= \frac{2}{x} \int_{x-1}^x \mathbf{E}\{T_t - [4(x-t) - 2]\}^4 dt \\
&\quad + \frac{1}{x} \int_1^{x-1} \mathbf{E}[T_t + T_{x-t}^*]^4 dt.
\end{aligned} \tag{A.12}$$

Because  $T_t$  is independent of  $T_{x-t}^*$ , and  $\mathbf{E}T_t = 0$  holds for any  $1 \leq t \leq x-1$ , we get

$$\begin{aligned}
\mathbf{E}[T_x]^4 &= \frac{2}{x} \int_{x-1}^x [4(x-t) - 2]^4 dt \\
&\quad + \frac{12}{x} \int_{x-1}^x [4(x-t) - 2]^2 \mathbf{E}[T_t^2] dt \\
&\quad - \frac{8}{x} \int_{x-1}^x [4(x-t) - 2] \mathbf{E}[T_t^3] dt
\end{aligned}$$

$$\begin{aligned}
& + \frac{2}{x} \int_{x-1}^x \mathbf{E} [T_t^4] dt \\
& + \frac{1}{x} \int_1^{x-1} (\mathbf{E} [T_t^4] + \mathbf{E} [T_{x-t}^*]^4 + 6\mathbf{E} [T_t T_{x-t}^*]^2) dt \\
& = \frac{2}{x} \int_{x-1}^x [4(x-t) - 2]^4 dt \\
& + \frac{12}{x} \int_{x-1}^x [4(x-t) - 2]^2 \mathbf{E} [T_t^2] dt \\
& - \frac{8}{x} \int_{x-1}^x [4(x-t) - 2] \mathbf{E} [T_t^3] dt \\
& + \frac{2}{x} \int_1^x \mathbf{E} [T_t]^4 dt + \frac{6}{x} \int_1^{x-1} \mathbf{E} [T_t T_{x-t}^*]^2 dt := I_1 \\
& + I_2 + I_3 + I_4 + I_5.
\end{aligned} \tag{A.13}$$

In particular, for the part  $I_1$ , we have

$$I_1 = \frac{32}{5x}. \tag{A.14}$$

When  $x > 3$ , for the part  $I_2$ , we have

$$\begin{aligned}
I_2 & = \frac{12}{x} \int_{x-1}^x [4(x-t) - 2]^2 \mathbf{E} [T_t^2] dt \\
& = \frac{12}{x} \int_{x-1}^x [4(x-t) - 2]^2 \mathbf{Var} [S_t] dt \\
& = \frac{12}{x} \int_{x-1}^x [4(x-t) - 2]^2 [(32 \ln 2 - 20)t] dt \\
& = \frac{12(32 \ln 2 - 20)}{x} \int_0^1 (4t - 2)^2 (x-t) dt \\
& = \frac{48}{3} (32 \ln 2 - 20) - \frac{24(32 \ln 2 - 20)}{3x} \\
& := 12k_1 - \frac{6k_1}{x},
\end{aligned} \tag{A.15}$$

where  $k_1 := 4(32 \ln 2 - 20)/3$  is a constant.

When  $x > 4$ , for the part  $I_3$ , we have

$$\begin{aligned}
I_3 & = -\frac{8}{x} \int_{x-1}^x [4(x-t) - 2] \mathbf{E} [T_t^3] dt \\
& = -\frac{8}{x} \int_{x-1}^x [4(x-t) - 2] k_0 t dt \\
& = -\frac{8k_0}{x} \int_0^1 (4t - 2)(x-t) dt = \frac{8k_0}{3x},
\end{aligned} \tag{A.16}$$

where  $k_0$  is the same as that in (A.11).

When  $x > 4$ , for the part  $I_5$ , we have

$$\begin{aligned}
I_5 & = \frac{6}{x} \int_1^{x-1} \mathbf{E} [T_t T_{x-t}^*]^2 dt \\
& = \frac{6}{x} \int_1^2 \mathbf{E} [T_t^2] \mathbf{E} [T_{x-t}^*]^2 dt \\
& + \frac{6}{x} \int_{x-2}^{x-1} \mathbf{E} [T_t^2] \mathbf{E} [T_{x-t}^*]^2 dt \\
& + \frac{6}{x} \int_2^{x-2} \mathbf{E} [T_t^2] \mathbf{E} [T_{x-t}^*]^2 dt \\
& = \frac{12}{x} \int_1^2 \mathbf{E} [T_t^2] \mathbf{E} [T_{x-t}^*]^2 dt \\
& + \frac{6}{x} \int_2^{x-2} \mathbf{E} [T_t^2] \mathbf{E} [T_{x-t}^*]^2 dt.
\end{aligned} \tag{A.17}$$

Noting that  $\mathbf{E} [T_t^2] = \mathbf{Var} [S_t]$  and (6), we can see that

$$\begin{aligned}
& \int_1^x \mathbf{E} [T_t^2] \mathbf{E} [T_{x-t}^*]^2 dt \\
& = 2 \int_1^2 (32t \ln t - 16t^2 + 8t + 8) \\
& \cdot ((32 \ln 2 - 20)(x-t)) dt \\
& + \int_2^{x-2} ((32 \ln 2 - 20)t) \\
& \cdot ((32 \ln 2 - 20)(x-t)) dt = \frac{1}{6} (32 \ln 2 - 20)^2 \\
& \cdot x^3 - \frac{4}{3} (32 \ln 2 - 20)^2 x^2 + \frac{1}{3} (32 \ln 2 - 20) \\
& \cdot (256 \ln 2 - 168)x + \frac{16}{9} (32 \ln 2 - 20) := a_3 x^3 \\
& + a_2 x^2 + a_1 x + a_0,
\end{aligned} \tag{A.18}$$

where

$$\begin{aligned}
a_0 & = \frac{16}{9} (32 \ln 2 - 20); \\
a_1 & = \frac{1}{3} (32 \ln 2 - 20) (256 \ln 2 - 168); \\
a_2 & = -\frac{4}{3} (32 \ln 2 - 20)^2; \\
a_3 & = \frac{1}{6} (32 \ln 2 - 20)^2.
\end{aligned} \tag{A.19}$$

Therefore,

$$\mathbf{E}[T_x]^4 = \frac{2}{x} \int_1^x \mathbf{E}[T_t]^4 dt + I_1 + I_2 + I_3 + I_5, \quad (\text{A.20})$$

$$x > 4.$$

That is,

$$x\mathbf{E}[T_x^4] = 2 \int_1^x \mathbf{E}[T_t]^4 dt + (I_1 + I_2 + I_3 + I_5)x, \quad (\text{A.21})$$

$$x > 4.$$

Via differentiation with respect to  $x$ , we get the differential equation:

$$\left(\mathbf{E}T_x^4\right)' - \frac{1}{x}\mathbf{E}T_x^4 = 12\frac{k_1}{x} + 6\left(3a_3x + 2a_2 + a_1\frac{1}{x}\right), \quad (\text{A.22})$$

$$x > 4.$$

The solution to this differential equation is

$$\mathbf{E}T_x^4 = 18a_3x^2 + 12a_2x \ln x + cx - 6a_1 - 12k_1, \quad (\text{A.23})$$

$$x > 4,$$

where  $c$  is a constant and the constants  $k_1, a_1, a_2, a_3$  are real numbers as defined before. From this equation, Proposition 3 follows.

## Consent

Informed consent was obtained from all individual participants included in the study.

## Conflict of Interests

The authors declare that they have no conflict of interests.

## Acknowledgments

The authors are most grateful to the referee and the editor for their very thorough reading of the paper and valuable suggestions, which greatly improve the original results and presentation of this paper. Jie Liu's work was supported by the National Natural Science Foundation of China (nos. 11101394, 71471168, and 71520107002), China Postdoctoral Science Foundation Funded Project (nos. 201104312 and 20100480688), Fund for the Doctoral Program of Higher Education Foundation (no. 20113402120005), and the Fundamental Research Funds for the Central Universities of China (no. WK2040160008). Yang Yang's work was supported by the National Natural Science Foundation of China (no. 71471090), the Humanities and Social Sciences Foundation of the Ministry of Education of China (no. 14YJCZH182), China Postdoctoral Science Foundation (nos. 2014T70449 and 2012M520964), Natural Science Foundation of Jiangsu Province of China (no. BK20131339), the Major Research Plan of Natural Science Foundation of the Jiangsu Higher

Education Institutions of China (no. 15KJJA110001), Qing Lan Project, PAPD, Program of Excellent Science and Technology Innovation Team of the Jiangsu Higher Education Institutions of China, Project of Construction for Superior Subjects of Statistics of Jiangsu Higher Education Institutions, and Project of the Key Lab of Financial Engineering of Jiangsu Province.

## References

- [1] M. Drmota, *Random Trees*, Springer, New York, NY, USA, 2009.
- [2] L. Devroye and S. Janson, "Protected nodes and fringe subtrees in some random trees," *Electronic Communications in Probability*, vol. 19, article 6, 2014.
- [3] Q. Feng and Z. Hu, "Phase changes in the topological indices of scale-free trees," *Journal of Applied Probability*, vol. 50, no. 2, pp. 516–532, 2013.
- [4] N. Broutin and P. Flajolet, "The distribution of height and diameter in random non-plane binary trees," *Random Structures and Algorithms*, vol. 41, no. 2, pp. 215–252, 2012.
- [5] F. Dennert and R. Grübel, "On the subtree size profile of binary search trees," *Combinatorics, Probability and Computing*, vol. 19, no. 4, pp. 561–578, 2010.
- [6] H. M. Mahmoud, "One-sided variations on binary search trees," *Annals of the Institute of Statistical Mathematics*, vol. 55, no. 4, pp. 885–900, 2003.
- [7] Q. Feng and Z. Hu, "On the Zagreb index of random recursive trees," *Journal of Applied Probability*, vol. 48, no. 4, pp. 1189–1196, 2011.
- [8] R. Grübel and I. Michailow, "Random recursive trees: a boundary theory approach," *Electronic Journal of Probability*, vol. 20, article 37, 2015.
- [9] H. M. Mahmoud and M. D. Ward, "Asymptotic properties of protected nodes in random recursive trees," *Journal of Applied Probability*, vol. 52, no. 1, pp. 290–297, 2015.
- [10] M. Sibuya and Y. Itoh, "Random sequential bisection and its associated binary tree," *Annals of the Institute of Statistical Mathematics*, vol. 39, no. 1, pp. 69–84, 1987.
- [11] H. Prodinger, "How to select a loser," *Discrete Mathematics*, vol. 120, no. 1–3, pp. 149–159, 1993.
- [12] J. A. Fill, H. M. Mahmoud, and W. Szpankowski, "On the distribution for the duration of a randomized leader election algorithm," *The Annals of Applied Probability*, vol. 6, no. 4, pp. 1260–1283, 1996.
- [13] Y. Itoh and H. M. Mahmoud, "One-sided variations on interval trees," *Journal of Applied Probability*, vol. 40, no. 3, pp. 654–670, 2003.
- [14] S. Janson, "One-sided interval trees," *Journal of the Iranian Statistical Society*, vol. 3, pp. 149–164, 2004.
- [15] S. Janson and R. Neininger, "The size of random fragmentation trees," *Probability Theory and Related Fields*, vol. 142, no. 3–4, pp. 399–442, 2008.
- [16] M. Javani, H. M. Mahmoud, and M. Vahidi-Asl, "Paths in m-ary interval trees," *Discrete Mathematics*, vol. 287, no. 1–3, pp. 45–53, 2004.
- [17] C. Su, J. Liu, and Z. S. Hu, "The Laws of large numbers for the size of complete interval trees," *Advances in Mathematics*, vol. 36, no. 2, pp. 181–188, 2007.



- [18] W. Pan, S.-K. Li, X. Cai, and L. Zeng, "The construction algorithm of binary interval tree nodes based on lattice partition," *Journal of Computer-Aided Design & Computer Graphics*, vol. 23, pp. 1115–1122, 2011.
- [19] V. M. Zolotarev, *Modern Theory of Summation of Random Variables*, VSP, Utrecht, The Netherlands, 1997.
- [20] V. M. Zolotarev, "Approximation of distributions of sums of independent random variables with values in infinite-dimensional spaces," *Theory of Probability and Its Applications*, vol. 21, no. 4, pp. 721–737, 1976.
- [21] S. T. Rachev and L. Rüschendorf, "Probability metrics and recursive algorithms," *Advances in Applied Probability*, vol. 27, no. 3, pp. 770–799, 1995.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

