Far Eastern Mathematical Journal. 2015. V. 15. № 1. P. 91–101

MSC2010 Primary 05C05, 60C05; secondary 60F05.

© Qunqiang Feng, Zhishui Hu[1]

Asymptotic normality of the Zagreb index of random b-ary recursive trees

The b-ary recursive trees model is one of simple families of increasing trees. In this work, the Zagreb index Z_n of a random b-ary recursive tree of size n is studied. As $n \to \infty$, the asymptotic normality of Z_n is established through the martingale central limit theorem, as well as the asymptotic expressions of the mean and variance of Z_n are given.

Keywords: *random tree, Zagreb index, martingale, asymptotic normality.*

1. Introduction

For any fixed integer $b \geq 2$, the b-ary recursive tree is a rooted, ordered, labeled tree where the out-degree is bounded by b, and the labels along each path beginning at the root increase. It was first obtained as an increasing tree by Bergeron *et al.* (1992), and is a special case of the general model of random trees in Broutin *et al.* (2008). Formally, a random b-ary recursive tree can be generated by the following recursive procedure. Consider the infinite complete b-ary rooted, ordered tree, and start with the root as the first internal node (labeled 1) and its b children as external nodes. Progressively, given the random b-ary recursive tree with $n-1$ ($n \geq 2$) internal nodes, the internal node labeled n is inserted as follows: An external node is chosen uniformly distributed on the set of all current external nodes, change it to the nth internal node, and add the b children of it to the set of external nodes. It is well-known that the binary search trees can be regards as special members of the class of random b-ary recursive trees with $b = 2$ (see, for example, Knuth(1998)). For more backgrounds for the random b-ary recursive trees, we refer the reader to Janson (2006), and Panholzer and Prodinger (2007).

As one of the well-known topological indices, the Zagreb index was introduced by the chemists Gutman and Trinajstić (1972). This index is an important molecular descriptor and has been closely correlated with many chemical properties. In chemistry, chemical graphs are generated from molecules by replacing atoms with vertices and bonds with edges, or represent only bare molecular skeletons, that is, molecular skeletons without hydrogen atoms. The Zagreb index of a graph G is defined as the sum of the squares

[1]Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei, 230026, China. E-mail: `fengqq@ustc.edu.cn, huzs@ustc.edu.cn`

of the degrees of all vertices in G. For a survey of the application of the Zagreb index in computational chemistry, we refer to Nikolić *et al.* (2003) and the references therein. Besides, the Zagreb index also attracts attention in the graph theory (see, for example, Andova *et al.* (2011)). Employing a similar method in Feng and Hu (2011), we shall study the the limit behavior of the Zagreb index of a random b-ary recursive tree of size n in this work.

The size of a b-ary recursive tree counts the number of internal nodes in it. In a random b-ary recursive tree of size n, let $D_{k,n}$ and $\bar{D}_{k,n}$ be respectively the degree and out-degree of the internal node labeled k, where the set of external nodes contributes nothing to the (out-)degree of any internal node. Note that

$$D_{1,n} = \bar{D}_{1,n}, \quad D_{k,n} = \bar{D}_{k,n} + 1, \qquad 2 \le k \le n. \tag{1}$$

Let $\mathcal{F}_n$ be the σ-field generated by the procedure of the first n internal nodes in random b-ary recursive trees. The Zagreb index of a b-ary recursive tree is defined as the sum of the squares of the degrees of all internal nodes in it. It is easier to work with a modified Zagreb index which is defined in exactly the same way as the standard index, except that the out-degrees of all internal nodes are instead of their degrees. Mathematically, for any integer $m \ge 2$, we can define the general Zagreb and modified Zagreb indices of a random b-ary recursive tree of size n as

$$Z_n^{(m)} = \sum_{k=1}^{n} D_{k,n}^m \quad \text{and} \quad \bar{Z}_n^{(m)} = \sum_{k=1}^{n} \bar{D}_{k,n}^m.$$

For the standard case $m = 2$, we will suppress the superscripts.

Throughout this work, all unspecified limits are taken to be $n \to \infty$. We will also use the following notation. For probabilistic convergence we use $\xrightarrow{\mathrm{P}}$ to denote convergence in probability, and $\xrightarrow{\mathrm{D}}$ to denote convergence in distribution. We denote the integer part of a real number x by $\lfloor x \rfloor$, and, for positive integers n and j,

$$s_n := n(b-1)+1, \quad c[n,j] := \frac{\Gamma\Big(n + \frac{1}{b-1}\Big)}{\Gamma\Big(n + \frac{1-j}{b-1}\Big)}.$$

For any fixed $j \ge 1$, it is easy to see that $c[n,j]$ is strictly increasing, and

$$c[n,j] = n^{\frac{j}{b-1}}(1 + O(n^{-1})). \tag{2}$$

The rest of this work is organized as follows. In Section 2, we give the first two moments of Z_n. In Section 3, we state and prove our main result that under suitable normalized, Z_n has the asymptotic normality property.

2. The mean and variance

Let the random variables Z_n and $\bar{Z}_n$ be respectively the standard and modified Zagreb index of a random b-ary recursive tree of size n. By (1), it is obvious that

$Z_1 = \bar{Z}_1 = 0$, and for $n \geq 2$,

$$Z_n = \sum_{k=1}^{n} D_{k,n}^2 = \sum_{k=1}^{n} \bar{D}_{k,n}^2 + 2\sum_{k=1}^{n} \bar{D}_{k,n} + (n-1) = \bar{Z}_n - 2\bar{D}_{1,n} + 3(n-1), \qquad (3)$$

where we use the fact that the sum of the out-degrees of all nodes in a rooted tree of size n is $n-1$.

The right-hand side of (3) involves the out-degree of the root $\bar{D}_{1,n}$. It has been known that

$$\mathbb{P}(\bar{D}_{1,n} = b) = 1 + O\left(\frac{\log n}{n^{\frac{1}{b-1}}}\right), \qquad (4)$$

and this probability is strictly increasing in n, and converges to 1 as $n \to \infty$ (see Kuba and Panholzer (2007), where the terminology "d-ary increasing tree"is used instead of "b-ary recursive tree"here). This, together with (3), shows that the standard Zagreb index Z_n is very close to a linear function of the modified one $\bar{Z}_n$. So we shall usually consider $\bar{Z}_n$ instead of Z_n itself, and then give the corresponding results of Z_n through (3) and (4).

Considering the insertion of the internal node labeled n to a random b-ary recursive tree of size $n-1$, we have

$$\bar{Z}_n = \bar{Z}_{n-1} + (\bar{D}_{V_n,n-1} + 1)^2 - \bar{D}_{V_n,n-1}^2 = \bar{Z}_{n-1} + 2\bar{D}_{V_n,n-1} + 1, \qquad (5)$$

where V_n denotes the parent of the internal node labeled n. Note that there are s_n external nodes in any random b-ary recursive tree of size n. It is easy to see that V_n has the following conditional distribution law

$$\mathbb{P}(V_n = k|\mathcal{F}_{n-1}) = \frac{b - \bar{D}_{k,n-1}}{s_{n-1}}, \qquad 1 \leq k \leq n-1. \qquad (6)$$

The relations (5) and (6) are useful for the computation of the moments of $\bar{Z}_n$. We will illustrate the procedure of the mean and variance. For the mean $\mathbb{E}[\bar{Z}_n]$, we proceed with

$$\mathbb{E}[\bar{Z}_n|\mathcal{F}_{n-1}] = \mathbb{E}[\bar{Z}_n|\bar{D}_{k,n-1},\, 1 \leq k \leq n-1] =$$

$$= \bar{Z}_{n-1} + 2\sum_{k=1}^{n-1} \bar{D}_{k,n-1}\mathbb{P}(V_n = k|\mathcal{F}_{n-1}) + 1 = \bar{Z}_{n-1} + \frac{2}{s_{n-1}}\Big(b\sum_{k=1}^{n-1} \bar{D}_{k,n-1} - \sum_{k=1}^{n-1} \bar{D}_{k,n-1}^2\Big) + 1 =$$

$$= \Big(1 - \frac{2}{s_{n-1}}\Big)\bar{Z}_{n-1} + \frac{2(n-2)b}{s_{n-1}} + 1 = \frac{c[n-1,2]}{c[n,2]}\bar{Z}_{n-1} + \alpha_{n-1}, \qquad n \geq 2, \qquad (7)$$

where

$$\alpha_n := \frac{2b(n-1)}{s_n} + 1.$$

Taking the expectation of both sides of (7), we have

$$\mathbb{E}[\bar{Z}_n] = \frac{c[n-1,2]}{c[n,2]}\mathbb{E}[\bar{Z}_{n-1}] + \alpha_{n-1}, \qquad (8)$$

which, with the initial value $\mathbb{E}[\bar{Z}_1] = 0$, implies that

$$\mathbb{E}[\bar{Z}_n] = \frac{1}{c[n,2]} \sum_{k=1}^{n-1} c[k+1,2]\alpha_k, \qquad n \geq 2.$$

Noting the fact that

$$\alpha_n = \frac{3b-1}{b-1}(1 + O(n^{-1})),$$

we thus have

$$\mathbb{E}[\bar{Z}_n] = \frac{3b-1}{b+1} n + O(1). \tag{9}$$

We now compute the variance of $\bar{Z}_n$ in the following. By (5), one can get

$$\mathbb{E}(\bar{Z}_n - \bar{Z}_{n-1} - 1)^2 = 4\mathbb{E}\Big[\bar{D}^2_{V_n,n-1}\Big] = \frac{4}{s_{n-1}} \sum_{k=1}^{n-1} \mathbb{E}\Big[\bar{D}^2_{k,n-1}(b - \bar{D}_{k,n-1})\Big] =$$

$$= \frac{4}{s_{n-1}}\Big(b\mathbb{E}[\bar{Z}_{n-1}] - \mathbb{E}[\bar{Z}^{(3)}_{n-1}]\Big). \tag{10}$$

It follows from (7) that the sequence $\{c[n,2](\bar{Z}_n - \mathbb{E}[\bar{Z}_n]), \mathcal{F}_n, n \geq 1\}$ is a martingale. Then

$$\mathbb{E}(\bar{Z}_n - \bar{Z}_{n-1} - 1)^2 = \mathbb{E}(\bar{Z}_n - \mathbb{E}[\bar{Z}_n] - \bar{Z}_{n-1} + \mathbb{E}[\bar{Z}_{n-1}])^2 + (\mathbb{E}[\bar{Z}_n] - \mathbb{E}[\bar{Z}_{n-1}] - 1)^2 =$$

$$= \mathrm{Var}[\bar{Z}_n] + \Big(1 - \frac{2c[n-1,2]}{c[n,2]}\Big)\mathrm{Var}[\bar{Z}_{n-1}] + \frac{4}{s^2_{n-1}}\Big((n-2)b - \mathbb{E}[\bar{Z}_{n-1}]\Big)^2, \tag{11}$$

which, together with (10), implies that

$$\mathrm{Var}[\bar{Z}_n] = \Big(\frac{2c[n-1,2]}{c[n,2]} - 1\Big)\mathrm{Var}[\bar{Z}_{n-1}] + \beta_{n-1} = \frac{c[n-1,4]}{c[n,4]}\mathrm{Var}[\bar{Z}_{n-1}] + \beta_{n-1}, \tag{12}$$

where

$$\beta_n := \frac{4}{s_n}\Big(b\mathbb{E}[\bar{Z}_n] - \mathbb{E}[\bar{Z}^{(3)}_n]\Big) - \frac{4}{s^2_n}\Big((n-1)b - \mathbb{E}[\bar{Z}_n]\Big)^2. \tag{13}$$

With the initial value $\mathrm{Var}[\bar{Z}_1] = 0$, recurrence (12) gives that

$$\mathrm{Var}[\bar{Z}_n] = \frac{1}{c[n,4]} \sum_{k=1}^{n-1} c[k+1,4]\beta_k, \qquad n \geq 2. \tag{14}$$

To obtain the order of the variance of $\bar{Z}_n$, we should estimate β_n which involves the term $\mathbb{E}[\bar{Z}^{(3)}_n]$. For random variable $\bar{Z}^{(3)}_n$, an analogous technique to (5) yields

$$\bar{Z}^{(3)}_n = \bar{Z}^{(3)}_{n-1} + (\bar{D}_{V_n,n-1} + 1)^3 - \bar{D}^3_{V_n,n-1} = \bar{Z}^{(3)}_{n-1} + 3\bar{D}^2_{V_n,n-1} + 3\bar{D}_{V_n,n-1} + 1.$$

Thus,

$$\begin{aligned}\mathbb{E}[\bar{Z}_n^{(3)}] &= \mathbb{E}[\bar{Z}_{n-1}^{(3)}] + \frac{3}{s_{n-1}}\sum_{k=1}^{n-1}\mathbb{E}[(\bar{D}_{k,n-1}^2 + \bar{D}_{k,n-1})(b - \bar{D}_{k,n-1})] + 1 = \\ &= \Big(1 - \frac{3}{s_{n-1}}\Big)\mathbb{E}[\bar{Z}_{n-1}^{(3)}] + \frac{3}{s_{n-1}}\Big((b-1)\mathbb{E}[\bar{Z}_{n-1}] + (n-2)b\Big) + 1 = \\ &= \frac{c[n-1,3]}{c[n,3]}\mathbb{E}[\bar{Z}_{n-1}^{(3)}] + \gamma_{n-1},\end{aligned} \tag{15}$$

where

$$\gamma_n := \frac{3}{s_n}((b-1)\mathbb{E}[\bar{Z}_n] + (n-1)b) + 1.$$

Similarly, with the initial value $\mathbb{E}[\bar{Z}_1^{(3)}] = 0$, the solution of (15) is

$$\mathbb{E}[\bar{Z}_n^{(3)}] = \frac{1}{c[n,3]}\sum_{k=1}^{n-1} c[k+1,3]\gamma_k, \qquad n \geq 2. \tag{16}$$

By (9), it is straightforward to get

$$\gamma_n = \frac{13b^2 - 9b + 2}{(b-1)(b+1)} + O(n^{-1}),$$

which implies that, by (16),

$$\mathbb{E}[\bar{Z}_n^{(3)}] = \frac{13b^2 - 9b + 2}{(b+1)(b+2)}n + O(1).$$

Also in a similar way, by the definition of β_n in (13) we have

$$\beta_n = \frac{4b(3b^2 - 2b - 4)}{(b+1)^2(b+2)} + O(n^{-1}). \tag{17}$$

From the explicit expression of the variance of $\bar{Z}_n$ given in (14), we can now proceed with the computation

$$\mathrm{Var}[\bar{Z}_n] = \frac{4b(b-1)(3b^2 - 2b - 4)}{(b+1)^2(b+2)(b+3)}n + O(1). \tag{18}$$

Since $\bar{D}_{1,n}$ is bounded by b for each $n \geq 1$, it follows from (4) that

$$\mathbb{E}[\bar{D}_{1,n}] \to b, \quad \mathrm{Var}[\bar{D}_{1,n}] \to 0,$$

which, by (3), implies that $\mathbb{E}[Z_n] = \mathbb{E}[\bar{Z}_n] + 3n + O(1)$, and the variance of Z_n is asymptotically equivalent to $\mathrm{Var}[\bar{Z}_n]$. Collecting the above conclusions in this section, we have thus proved the following result.

Proposition 1. *Let Z_n be the Zagreb index of a random b-ary recursive tree of size n. Then*

$$\mathbb{E}[Z_n] = \frac{6b+2}{b+1}n + O(1), \qquad \mathrm{Var}[Z_n] = \sigma_b^2 n + O(1),$$

where the constant

$$\sigma_b^2 = \frac{4b(b-1)(3b^2-2b-4)}{(b+1)^2(b+2)(b+3)}.$$

An immediate consequence of Proposition 1 is as follows.

Proposition 2. *Let Z_n be the Zagreb index of a random b-ary recursive tree of size n. Then*

$$\frac{Z_n}{n} \xrightarrow{\mathrm{P}} \frac{6b+2}{b+1}.$$

Proof. It follows directly by Chebyshev's inequality and the fact $\mathrm{Var}[Z_n] = o(\mathbb{E}^2[Z_n])$.

3. Asymptotic normality

For the asymptotic behavior of the Zagreb index of random b-ary recursive trees, we shall state our main result in Theorem 1, and then prove it based on the fact that the sequence $\{c[n,2](\bar{Z}_n - \mathbb{E}[\bar{Z}_n]), \mathcal{F}_n, n \geq 1)\}$ is a martingale.

Theorem 1. *Let Z_n be the Zagreb index of a random b-ary recursive tree of size n. Then we have*

$$\frac{Z_n - \frac{6b+2}{b+1}n}{\sqrt{n}} \xrightarrow{\mathrm{D}} N(0, \sigma_b^2),$$

where the constant σ_b^2 is given in Proposition 1.

In particular, for the binary search tree of size n as a special case $b = 2$, we have

$$\frac{Z_n - \frac{14}{3}n}{\sqrt{n}} \xrightarrow{\mathrm{D}} N\Big(0, \frac{8}{45}\Big).$$

In order to prove the above theorem, we need an auxiliary lemma as follows.

Lemma 1. *For the general modified Zagreb index with $m = 3$ of a random b-ary recursive tree of size n, we have*

$$n^{-\frac{b+3}{b-1}} \sum_{k=2}^{n} \frac{c[k,2]^2}{s_{k-1}} (\bar{Z}_{k-1} - \mathbb{E}[\bar{Z}_{k-1}]) \xrightarrow{\mathrm{P}} 0, \tag{19}$$

and

$$n^{-\frac{b+3}{b-1}} \sum_{k=2}^{n} \frac{c[k,2]^2}{s_{k-1}} \Big(\bar{Z}_{k-1}^{(3)} - \mathbb{E}[\bar{Z}_{k-1}^{(3)}]\Big) \xrightarrow{\mathrm{P}} 0. \tag{20}$$

Proof. We prove the latter convergence (20) first. Analogously to (15), one can get

$$\mathbb{E}[\bar{Z}_n^{(3)}|\mathcal{F}_{n-1}] = \Big(1 - \frac{3}{s_{n-1}}\Big)\bar{Z}_{n-1}^{(3)} + \frac{3}{s_{n-1}}((b-1)\bar{Z}_{n-1} + (n-2)b) + 1,$$

which, together with (7), implies that

$$\mathbb{E}[\bar{Z}_n^{(3)} - 3(b-1)\bar{Z}_n|\mathcal{F}_{n-1}] = \frac{c[n-1,3]}{c[n,3]}(\bar{Z}_{n-1}^{(3)} - 3(b-1)\bar{Z}_{n-1}) - \frac{3(n-2)b(2b-3)}{s_{n-1}} - 3b+4.$$

Then it follows that the process $\{c[n,3]\bar{Z}_n^*, \mathcal{F}_n, n \geq 1\}$ is also a martingale, where

$$\bar{Z}_n^* := \bar{Z}_n^{(3)} - 3(b-1)\bar{Z}_n - \mathbb{E}[\bar{Z}_n^{(3)} - 3(b-1)\bar{Z}_n].$$

Similarly to the computation of the variance of $\bar{Z}_n$ in Section 2, one can obtain that the variance of $\bar{Z}_n^*$ is also in the form of $c_b^* n + O(1)$, where c_b^* is a constant independent of n. It follows from (2) that there exists an absolute constant c_0, independent of n, such that $c[n,3] < c_0 c[n,2]^{3/2}$. By Doob's inequality, we thus have

$$\begin{aligned}\mathbb{E}\Big[\max_{2\leq k\leq n}\Big\{c[k-1,3]\Big(\bar{Z}_{k-1}^{(3)} - \mathbb{E}[\bar{Z}_{k-1}^{(3)}]\Big)\Big\}^2\Big] &\leq 2\mathbb{E}\Big[\max_{2\leq k\leq n}\{c[k-1,3]\bar{Z}_{k-1}^*\}^2\Big] + \\ &+18(b-1)^2c_0^2c[n,2]\mathbb{E}\Big[\max_{2\leq k\leq n}\{c[k-1,2](\bar{Z}_{k-1} - \mathbb{E}[\bar{Z}_{k-1}])\}^2\Big] \leq \\ &\leq 8c[n,3]^2\mathrm{Var}[\bar{Z}_n^*] + 72(b-1)^2c_0^2c[n,2]^3\mathrm{Var}[\bar{Z}_n] = O(1)n^{\frac{b+5}{b-1}}. \end{aligned} \tag{21}$$

Note that

$$\lim n^{-\frac{1}{b-1}} \sum_{k=2}^{n} \frac{c[k,2]^2}{c[k-1,3]s_{k-1}} < \infty.$$

Therefore, by Chebyshev's inequality and (21), for any $\varepsilon > 0$,

$$\begin{aligned}&\mathbb{P}\left(\left|\sum_{k=2}^{n}\frac{c[k,2]^2}{s_{k-1}}\Big(\bar{Z}_{k-1}^{(3)} - \mathbb{E}[\bar{Z}_{k-1}^{(3)}]\Big)\right| > \varepsilon n^{\frac{b+3}{b-1}}\right) \leq \\ &\leq \frac{1}{\varepsilon^2 n^{\frac{2(b+3)}{b-1}}}\mathbb{E}\left(\sum_{k=2}^{n}\frac{c[k,2]^2}{s_{k-1}}\Big(\bar{Z}_{k-1}^{(3)} - \mathbb{E}\Big[\bar{Z}_{k-1}^{(3)}\Big]\Big)\right)^2 \leq \\ &\leq \frac{1}{\varepsilon^2 n^{\frac{2(b+3)}{b-1}}}\left(\sum_{k=2}^{n}\frac{c[k,2]^2}{c[k-1,3]s_{k-1}}\right)^2\mathbb{E}\left[\max_{2\leq k\leq n}\Big\{c[k-1,3]\Big(\bar{Z}_{k-1}^{(3)} - \mathbb{E}\Big[\bar{Z}_{k-1}^{(3)}\Big]\Big)\Big\}^2\right] = O(n^{-1}),\end{aligned}$$

which implies (20).

The proof procedure for (19) is similar and simpler. Note that

$$\sum_{k=2}^{n}\frac{c[k,2]^2}{c[k-1,2]s_{k-1}} = O(n^{\frac{2}{b-1}}),$$

and that in (21) we have used the following inequality

$$\mathbb{E}\Big[\max_{2\leq k\leq n}\{c[k-1,2](\bar{Z}_{k-1} - \mathbb{E}[\bar{Z}_{k-1}])\}^2\Big] \leq 4c[n,2]^2\mathrm{Var}[\bar{Z}_n] = O\Big(n^{\frac{b+3}{b-1}}\Big).$$

Also by Chebyshev's inequality, we have

$$\mathbb{P}\left(\left|\sum_{k=2}^{n}\frac{c[k,2]^2}{s_{k-1}}(\bar{Z}_{k-1}-\mathbb{E}[\bar{Z}_{k-1})\right|\geq \varepsilon n^{\frac{b+3}{b-1}}\right)\leq$$

$$\leq \frac{1}{\varepsilon^2 n^{\frac{2(b+3)}{b-1}}}\mathbb{E}\left(\sum_{k=2}^{n}\frac{c[k,2]^2}{s_{k-1}}\left(\bar{Z}_{k-1}-\mathbb{E}[\bar{Z}_{k-1}]\right)\right)^2\leq$$

$$\leq \frac{1}{\varepsilon^2 n^{\frac{2(b+3)}{b-1}}}\left(\sum_{k=2}^{n}\frac{c[k,2]^2}{c[k-1,2]s_{k-1}}\right)^2\mathbb{E}\left[\max_{2\leq k\leq n}\left\{c[k-1,2]\big(\bar{Z}_{k-1}-\mathbb{E}[\bar{Z}_{k-1}]\big)\right\}^2\right]=O\big(n^{-1}\big).$$

The proof of Lemma 1 is complete.

We now give the proof of our main theorem in the following.

Proof of Theorem 1. By (3), (4) and Proposition 1, we only need to prove that

$$\frac{\bar{Z}_n-\mathbb{E}[\bar{Z}_n]}{\sqrt{\mathrm{Var}[\bar{Z}_n]}}\xrightarrow{\mathrm{D}}N(0,1).$$

Recall that the sequence $\{c[n,2](\bar{Z}_n-\mathbb{E}[\bar{Z}_n]),\mathcal{F}_n,n\geq 1)\}$ is a martingale. Let

$$M_k:=c[k,2](\bar{Z}_k-\mathbb{E}[\bar{Z}_k])-c[k-1,2](\bar{Z}_{k-1}-\mathbb{E}[\bar{Z}_{k-1}]),\qquad k=2,3,\ldots,$$

with $M_1=0$. Then the process $\{M_k,k\geq 1\}$ is a martingale difference sequence. By Corollary 3.1 of Hall and Heyde (1980) and the expression of the variance $\mathrm{Var}[\bar{Z}_n]$ given in (18), it is sufficient to show that, for any $\varepsilon>0$,

$$\frac{1}{c[n,2]^2 n}\sum_{k=2}^{n}\mathbb{E}\left[M_k^2 I\Big(\Big|\frac{M_k}{c[n,2]\sqrt{n}}\Big|>\varepsilon\Big)\Big|\mathcal{F}_{k-1}\right]\xrightarrow{\mathrm{P}}0,\tag{22}$$

and

$$\frac{1}{c[n,2]^2\sigma_b^2 n}\sum_{k=2}^{n}\mathbb{E}\left[M_k^2|\mathcal{F}_{k-1}\right]\xrightarrow{\mathrm{P}}1.\tag{23}$$

We first prove (22). By (5) and (8), we can rewrite M_k as

$$M_k=c[k,2]\bar{Z}_k-c[k-1,2]\bar{Z}_{k-1}-c[k,2]\alpha_{k-1}=$$

$$=c[k,2]\left(\bar{Z}_{k-1}+2\bar{D}_{V_k,k-1}+1\right)-c[k-1,2]\bar{Z}_{k-1}-c[k,2]\alpha_{k-1}=$$

$$=c[k,2]\left(2\bar{D}_{V_k,k-1}+1+\frac{2\bar{Z}_{k-1}}{s_{k-1}}-\alpha_{k-1}\right).$$

Note that for any $n\geq 2$,

$$\alpha_n\leq\frac{3b-1}{b-1},\qquad \bar{D}_{V_n,n-1}\leq b,\qquad \frac{\bar{Z}_n}{s_n}\leq\frac{nb^2}{n(b-1)+1}\leq\frac{b^2}{b-1}.$$

Then there exists a positive constant c_b, which only depends on b, such that

$$\max_{2\leq k\leq n}|M_k|\leq c_b c[n,2]=o\left(c[n,2]\sqrt{n}\right),$$

which implies that (22) holds.

Next we will prove (23). With similar calculations as in (10) and (11), we have

$$\mathbb{E}\left[(\bar{Z}_n - \bar{Z}_{n-1} - 1)^2 | \mathcal{F}_{n-1}\right] = \frac{4}{s_{n-1}}\left(b\bar{Z}_{n-1} - \bar{Z}_{n-1}^{(3)}\right),$$

and

$$\begin{aligned}\mathbb{E}\left[(\bar{Z}_n - \bar{Z}_{n-1} - 1)^2 | \mathcal{F}_{n-1}\right] &= \mathbb{E}\left[(\bar{Z}_n - \mathbb{E}[\bar{Z}_n])^2 | \mathcal{F}_{n-1}\right] + \\ &+ \left(1 - \frac{2c[n-1,2]}{c[n,2]}\right)\left(\bar{Z}_{n-1} - \mathbb{E}[\bar{Z}_{n-1}]\right)^2 + \\ +\frac{4}{s_{n-1}^2}\left((n-2)b - \mathbb{E}[\bar{Z}_{n-1}]\right)^2 &- \frac{8}{s_{n-1}^2}\left((n-2)b - \mathbb{E}[\bar{Z}_{n-1}])(\bar{Z}_{n-1} - \mathbb{E}[\bar{Z}_{n-1}]\right).\end{aligned}$$

Then

$$\begin{aligned}&\sum_{k=2}^{n} \mathbb{E}\left[M_k^2 | \mathcal{F}_{k-1}\right] = \\ &= \sum_{k=2}^{n}\left(c[k,2]^2 \mathbb{E}\left[\left(\bar{Z}_k - \mathbb{E}[\bar{Z}_k]\right)^2 | \mathcal{F}_{k-1}\right] - c[k-1,2]^2\left(\bar{Z}_{k-1} - \mathbb{E}[\bar{Z}_{k-1}]\right)^2\right) = \\ &= \sum_{k=2}^{n} \frac{4c[k,2]^2}{s_{k-1}}\left(b\bar{Z}_{k-1} - \bar{Z}_{k-1}^{(3)}\right) - \sum_{k=2}^{n} \frac{4c[k,2]^2}{s_{k-1}^2}\left(\bar{Z}_{k-1} - \mathbb{E}[\bar{Z}_{k-1}]\right)^2 - \\ &\quad - \sum_{k=2}^{n} \frac{4c[k,2]^2}{s_{k-1}^2}\left((k-2)b - \mathbb{E}[\bar{Z}_{k-1}]\right)^2 + \\ &\quad + \sum_{k=2}^{n} \frac{8c[k,2]^2}{s_{k-1}^2}\left((k-2)b - \mathbb{E}[\bar{Z}_{k-1}]\right)\left(\bar{Z}_{k-1} - \mathbb{E}[\bar{Z}_{k-1}]\right) = \\ &= \sum_{k=2}^{n} c[k,2]^2 \beta_{k-1} - \sum_{k=2}^{n} \frac{4c[k,2]^2}{s_{k-1}}\left(\bar{Z}_{k-1}^{(3)} - \mathbb{E}\left[\bar{Z}_{k-1}^{(3)}\right] - b\left(\bar{Z}_{k-1} - \mathbb{E}[\bar{Z}_{k-1}]\right)\right) - \\ &\quad - \sum_{k=2}^{n} \frac{4c[k,2]^2}{s_{k-1}^2}\left(\bar{Z}_{k-1} - \mathbb{E}[\bar{Z}_{k-1}]\right)^2 + \\ &+ \sum_{k=2}^{n} \frac{8c[k,2]^2}{s_{k-1}^2}\left((k-2)b - \mathbb{E}[\bar{Z}_{k-1}]\right)\left(\bar{Z}_{k-1} - \mathbb{E}[\bar{Z}_{k-1}]\right) =: J_1 - J_2 - J_3 + J_4,\end{aligned}$$

where β_n is defined in (13). By (17), it is easy to see that

$$\lim \beta_n = \frac{b+3}{b-1}\sigma_b^2,$$

from which one can directly get

$$\lim \frac{J_1}{c[n,2]^2 \sigma_b^2 n} = \lim \sigma_b^{-2} n^{-\frac{b+3}{b-1}} \sum_{k=2}^{n} c[k,2]^2 \beta_{k-1} = 1.$$

Hence, to prove (23), we only need to prove that, for $i = 2, 3, 4$,

$$\frac{J_i}{c[n,2]^2 n} \xrightarrow{\mathrm{P}} 0. \tag{24}$$

It follows by Lemma 1 that the convergence (24) is valid for $i = 2$. Applying Chebyshev's inequality, for any $\varepsilon > 0$, we have

$$\mathsf{P}(J_3 > \varepsilon n^{\frac{b+3}{b-1}}) \le \varepsilon^{-1} n^{-\frac{b+3}{b-1}} \mathbb{E}[J_3] = \varepsilon^{-1} n^{-\frac{b+3}{b-1}} \sum_{k=2}^{n} \frac{4c[k,2]^2}{s_{k-1}^2} \mathrm{Var}[\bar{Z}_{k-1}] \to 0,$$

which implies that (24) holds for $i = 3$. It is easy to check that

$$\lim n^{-\frac{b+3}{b-1}} \sum_{k=2}^{n} \frac{8c[k,2]^2}{s_{k-1}^2} ((k-2)b - \mathbb{E}[\bar{Z}_{k-1}])^2 = \frac{8(b-1)^3}{(b+1)^2(b+3)} =: l_b > 0.$$

We thus have that, for any $\varepsilon > 0$,

$$J_4 \le \frac{l_b}{\varepsilon} J_3 + \frac{\varepsilon}{l_b} \sum_{k=2}^{n} \frac{4c[k,2]^2}{s_{k-1}^2} \left((k-2)b - \mathbb{E}[\bar{Z}_{k-1}]\right)^2,$$

which implies that if n is sufficiently large,

$$\mathbb{P}\left(n^{-\frac{b+3}{b-1}} J_4 > \varepsilon\right) \le \mathbb{P}\left(\frac{l_b}{\varepsilon} n^{-\frac{b+3}{b-1}} J_3 + \frac{\varepsilon}{2} > \varepsilon\right) = \mathbb{P}\left(J_3 > \frac{\varepsilon^2}{2l_b} n^{\frac{b+3}{b-1}}\right) \to 0.$$

Then (24) also holds for $i = 4$. This completes the proof of the stated result.

References

[1] V. Andova, S. Bogoev, D. Dimitrov, M. Pilipczuk and R. Škrekovski, "On the Zagreb index inequality of graphs with prescribed vertex degrees", *Discrete Applied Mathematics*, **159** (2011), 852–858.

[2] F. Bergeron, P. Flajolet and B. Salvy, "Varieties of increasing trees, in: Raoult, J.C. (Ed.)", *Proc. 17th Coll. Trees in Algebra and Programming (Lecture Notes Comput. Sci.)*. V. 581, Springer, Berlin, 1992, 24–48.

[3] N. Broutin, L. Devroye, E. McLeish and M. de la Salle, "The height of increasing trees", *Random Structures and Algorithms*, **32** (2008), 494–518.

[4] Q. Feng and Z. Hu, "On the Zagreb index of random recursive trees", *Journal of Applied Probability*, **48** (2011), 1189–1196.

[5] I. Gutman and N. Trinajstić, "Graph theory and molecular orbitals. Total φ-electron energy of alternant hydrocarbons", *Chemical Physics Letters*, **17** (1972), 535–538.

[6] P. Hall and C. C. Heyde, *Martingale limit theory and its application*, Academic Press, New York, 1980.

[7] S. Janson, "Random cutting and records in deterministic and random trees", *Random Structures and Algorithms*, **29** (2006), 139–179.

[8] D. Knuth, *The Art of Computer Programming, Vol. 3: Sorting and Searchingn*, ed. 2nd, Addison-Wesley, Reading, Massachusetts, 1998.

[9] M. Kuba and A. Panholzer, "On the degree distribution of the nodes in increasing trees", *Journal of Combinatorial Theory, Series A*, **114** (2007), 597–618.

[10] S. Nikolić, G. Kovačević, A. Miličević and N. Trinajstić, "The Zagreb indices 30 years after", *Croatica Chemica ACTA*, **76** (2003), 113–124.

[11] A. Panholzer and H. Prodinger, "The level of nodes in increasing trees revisited", *Random Structures and Algorithms*, **31** (2007), 203–226.

Submitted 11 August 2014

The work of the first author is supported by NSFC(grant number 11101395).

Фэн К., Ху З. Асимптотическая нормальность загребского индекса случайных b-арных рекурсивных деревьев. Дальневосточный математический журнал. 2015. Том 15. № 1. С. 91–101.

АННОТАЦИЯ

Модель b-арных рекурсивных деревьев является одним из простых семейств растущих деревьев. В данной работе изучается загребский индекс Z_n случайного b-арного дерева размера n. При $n \to \infty$ асимптотическая нормальность Z_n устанавливается из центральной предельной теоремы о мартингалах. Вместе с этим получаются асимптотические выражения среднего значения и дисперсии Z_n.

Ключевые слова: *случайное дерево, загребский индекс, мартингал, асимптотическая нормальность.*