Brazilian Journal of Probability and Statistics
2015, Vol. 29, No. 4, 897–908
DOI: 10.1214/14-BJPS252

On the number of leaves in a random recursive tree

Yazhe Zhang

Capital Normal University

Abstract. This paper studies the asymptotic behavior of the number of leaves L_n in a random recursive tree T_n with n nodes. By utilizing the size-bias method, we derive an upper bound on the Wasserstein distance between the distribution of L_n and a standard normal distribution. Furthermore, we obtain a weak version of an Erdös–Rényi type law and a large deviation principle for L_n.

1 Introduction

Random recursive trees have been introduced as simple probability models arising from diverse investigations, such as for system generation (Na and Rapoport (1970)), for spread of contamination of organisms (Meir and Moon (1974)), for pyramid schemes (Gastwirth and Bhattacharya (1984)), for stemma construction of philology (Najock and Heyde (1982)), for internet interface maps (Janic et al. (2002)), and for stochastic growth of networks (Chan et al. (2003)) etc. Moreover, random recursive trees can also be used for some internet models and for physical evolving system models (Tetzlaff (2002)). They also appeared in the study of Hopf algebras under the name of "heap-ordered trees" (Grossman and Larson (1989)). We refer the reader to the survey in Mahmoud and Smythe (1995) (and references there) for more detailed descriptions.

In this paper, we study the asymptotic behavior of random recursive trees. We are concerned with a rooted non-plane size-n tree T_n, $n \in \mathbb{N}$, labeled with distinct integers $1, 2, \dots, n$, where the node labeled 1 is distinguished as the root, and for $2 \le k \le n$, the labels of the nodes on the unique path from the root to node k form an increasing sequence. Such a tree T_n can be constructed uniquely by attaching a node with label n to one of the $n-1$ nodes in T_{n-1}. This immediately shows that, for $n \ge 1$, the number of recursive trees of size n is given by $(n-1)!$. Throughout this paper, all these $(n-1)!$ recursive trees are considered to appear equally likely. Furthermore, we generate a random recursive tree of size n by the following process. Step 1, we take the node labeled 1 as the root. At step $k+1$, we attach the node labeled $k+1$ to any previous nodes i ($i = 1, \dots, k$) with probability $1/k$. The process stops after node n is inserted. See Figure 1 for all recursive trees of size 4.

Key words and phrases. Random recursive tree, normal approximation, Stein's method, size-bias method, Wasserstein distance, Kolmogorov distance, random permutation, concentration inequality.
Received March 2014; accepted June 2014.

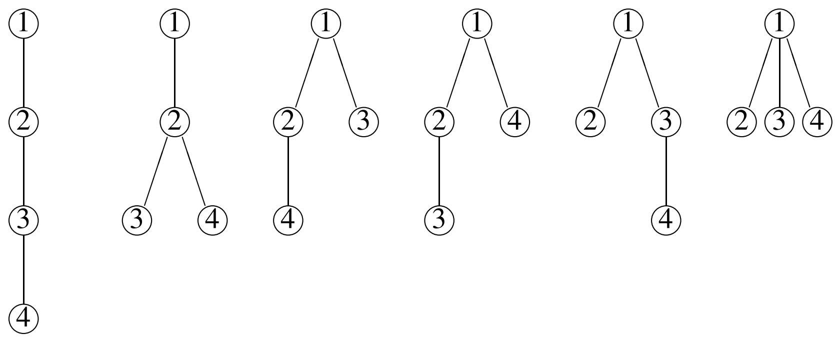

Figure 1 *All recursive trees of size* 4.

Leaves of a random recursive tree are the childless nodes. The number of leaves L_n in a random recursive tree T_n is a random variable of significant interest. There are a number of papers devoted to the study of L_n. We would like to mention the following ones. Na and Rapoport (1970) investigated the average number of leaves in a random recursive tree; Najock and Heyde (1982) found that the exact limiting distribution of the number of leaves in a random recursive tree approximately has a normal distribution by digital analysis, and also see Bergeron et al. (1992). They showed that the number of leaves tends to a Gaussian limit by analyzing properties of generating functions. But no bounds on the convergence to this Gaussian limits have been derived before. In this paper, we develop an explicit upper bound on the Wasserstein distance between the distribution of L_n and a standard normal distribution by using the size-bias method.

The Wasserstein distance between any two probability measures μ and ν on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is defined as follows

$$d_W(\mu,\nu) = \sup_{h\in\mathcal{H}} \left| \int_{\mathbb{R}} h(x)\,\mathrm{d}\mu(x) - \int_{\mathbb{R}} h(x)\,\mathrm{d}\nu(x) \right|,$$

where $\mathcal{H} := \{h : \mathbb{R} \to \mathbb{R} : |h(x) - h(y)| \le |x - y|\}$.

The size-bias method for normal approximation provides a powerful tool for approximating probabilities by a normal distribution, which was first appeared in the context of Stein's method for normal approximation in Goldstein and Rinott (1996). It turns out that this method provides an efficient way to get an explicit upper bound on the probability metrics. The reader is referred to Arratia et al. (2013) and the references therein for a basic introduction to the size-biased coupling method.

Here let us give the definition of *size-based distributions*.

Definition 1.1. For a non-negative random variable X with $\mathbb{E}[X] < \infty$, we say a random variable X^s has the size-biased distribution with respect to X if

$$\mathbb{E}[Xf(X)] = \mathbb{E}[X]\mathbb{E}[f(X^s)],$$

for all $f : [0, \infty) \to \mathbb{R}$ such that $\mathbb{E}|Xf(X)| < \infty$.

Since we need to couple L_n to a size-biased version L_n^s in Section 3, it is necessary to develop a method for size-biased coupling. For a random variable $X = \sum_{i=1}^n X_i$, where $X_i \geq 0$ and $\mathbb{E}[X_i] = \mu_i$, we can construct a size-biased version of X as follows.

1. For each $i = 1, \dots, n$, let X_i^s have the size-biased distribution of X_i independent of $(X_j)_{j \neq i}$ and $(X_j^s)_{j \neq i}$. Define a vector $(X_j^{(i)})_{j \neq i}$ such that its conditional distribution given X_i^s coincides with that of $(X_j)_{j \neq i}$ given X_i.
2. Choose a summand X_I randomly, where the index I is chosen proportional to μ_i and independent of all others, i.e. $\mathbb{P}(I = i) = \mu_i/\mu$, where $\mu = \mathbb{E}[X]$.
3. $X^s = \sum_{j \neq I} X_j^{(I)} + X_I^s$.

Remark 1.1. Note that for an indicator random variable X, $\mathbb{P}(X^s = 1) = \mathbb{P}(X = 1)/\mathbb{E}[X] = 1$, which means $X^s = 1$. Thus, if $S = \sum_{i=1}^n X_i$, where $X_i, \dots, X_n$ are indicator random variables with $\mathbb{P}(X_i = 1) = p_i$ and for each $i = 1, \dots, n$, let $(X_j^{(i)})_{j \neq i}$ have the distribution of $(X_j)_{j \neq i}$ conditional on $X_i = 1$, and I is chosen independent of all else with $\mathbb{P}(I = i) = p_i/\mathbb{E}[S]$, then $S^s = \sum_{j \neq I} X_j^{(I)} + 1$ has the size-biased distribution of S.

The fundamental result we need in this paper is the following theorem.

Theorem 1.2 (Ross (2011)). *Let X be a non-negative random variable with mean $\mu < \infty$ and $\operatorname{Var}(X) = \sigma^2$. Let X^s have the size-biased distribution with respect to X. If $W = \frac{X-\mu}{\sigma}$ and $Z \sim \mathcal{N}(0, 1)$, then*

$$d_W(W, Z) \leq \frac{\mu}{\sigma^2}\sqrt{\frac{2}{\pi}}\sqrt{\operatorname{Var}\big(\mathbb{E}[X^s - X \mid X]\big)} + \frac{\mu}{\sigma^3}\mathbb{E}\big[(X^s - X)^2\big]. \tag{1.1}$$

The rest of the paper is organized as follows. In Section 2, we present our main results about the number of leaves in a size-n random recursive tree. Section 3 is devoted to the proofs of our results. We first use the size-bias method to prove Theorem 2.2 and then we get an Erdös–Rényi type law for L_n. Next, we manage to derive a concentration inequality via application of the same method so as to gain the large deviation principle for L_n in Theorem 2.4.

2 Main results

Many quantities of interest in the study of random trees can be naturally represented as a sum of indicator random variables. We will introduce a bijection between permutations and corresponding random recursive trees, which enables us to represent L_n as a sum of indicator random variables.

Let $\sigma = (\sigma_1, \dots, \sigma_{n-1})$ be a permutation on $\{2, \dots, n\}$. We can construct a recursive tree with nodes $1, 2, \dots, n$ by taking 1 the root, and attaching the node $i \geq 2$ to the rightmost element j of σ, which precedes i and is less than i. If there is no such an element j, then we define the root 1 to be the parent of i.

From the tree construction process above, L_n can be defined by

$$L_n = \sum_{i=1}^{n-2} I[\sigma_i > \sigma_{i+1}] + 1. \tag{2.1}$$

This is obtained by observing that every appearance of descents in σ means a leaf will be added to tree T_n. Moreover, the last element σ_{n-1} of σ is always a leaf.

Let L_n be defined as (2.1), then we can obtain $\mathbb{E}[L_n]$ and $\operatorname{Var}(L_n)$ by straightforward calculations.

Lemma 2.1. *Let L_n be the number of leaves in T_n, then*

$$\mathbb{E}[L_n] = \frac{n}{2}, \qquad \operatorname{Var}(L_n) = \frac{n}{12}. \tag{2.2}$$

Proof. $\mathbb{E}[L_n] = n/2$ follows by noticing the fact that $\mathbb{P}(I[\sigma_i > \sigma_{i+1}] = 1) = 1/2$. For the second term, we have

$$\begin{aligned}
\mathbb{E}[L_n^2] &= \mathbb{E}\left[\left(\sum_{i=1}^{n-2} I[\sigma_i > \sigma_{i+1}] + 1\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_{i=1}^{n-2} I[\sigma_i > \sigma_{i+1}]\right)^2 + 2\left(\sum_{i=1}^{n-2} I[\sigma_i > \sigma_{i+1}]\right) + 1\right] \\
&= \mathbb{E}\left[3\sum_{i=1}^{n-2} I[\sigma_i > \sigma_{i+1}] + \sum_{i \neq j} I[\sigma_i > \sigma_{i+1}] I[\sigma_j > \sigma_{j+1}] + 1\right] \\
&= \frac{3(n-2)}{2} + \frac{(n-3)(n-4)}{4} + \frac{2(n-3)}{6} + 1 \\
&= \frac{n^2}{4} + \frac{n}{12},
\end{aligned}$$

where we have used that

$$\mathbb{P}\big(I[\sigma_i > \sigma_{i+1}] I[\sigma_j > \sigma_{j+1}] = 1\big) = \begin{cases} 1/6 & \text{if } |i-j| = 1, \\ 1/4 & \text{if } |i-j| > 1 \end{cases}$$

in the penultimate equality.

Finally, since $\operatorname{Var}(L_n) = \mathbb{E}[L_n^2] - (\mathbb{E}[L_n])^2$, the desired result follows. □

Remark 2.1. Najock and Heyde (1982) have obtained the same result by using the property of Eulerian numbers. But here we have used a totally different approach.

We next state our main results about L_n, proofs are postponed until next section.

Theorem 2.2. *Let L_n be the number of leaves in a random recursive tree T_n. Define $W := \frac{(L_n - n/2)}{\sqrt{n/12}}$, then*

$$d_W(W, Z) \leq \frac{C_1\sqrt{n}}{n-2} + \frac{C_2}{\sqrt{n}},$$

where $C_1 = 3\sqrt{3}$, $C_2 = 12\sqrt{3}$, and $Z \sim \mathcal{N}(0, 1)$.

Using the result of Theorem 2.2, we can derive the following weak version of an Erdös–Rényi type law for L_n.

Theorem 2.3. *Let L_n be the number of leaves in a random recursive tree T_n, then*

$$\frac{L_n}{n/2} \xrightarrow{\text{P}} 1. \tag{2.3}$$

Furthermore, using the size-bias method for a concentration inequality, we can obtain the following large deviation principle for L_n.

Theorem 2.4. *Let L_n be the number of leaves in a random recursive tree T_n, for any $x > 0$, we have*

$$\mathbb{P}\left(\frac{L_n}{n} - \frac{1}{2} \geq x\right) \leq \exp\left\{-\frac{nx^2}{1+x}\right\}, \tag{2.4}$$

and

$$\mathbb{P}\left(\frac{L_n}{n} - \frac{1}{2} \leq -x\right) \leq \exp\{-x^2 n\}. \tag{2.5}$$

3 Proofs

Before proving Theorem 2.2, we need a lemma, which makes the computation or bounding of $\operatorname{Var}(\mathbb{E}[L_n^s - L_n \mid L_n])$ much easier, because in practice, it is quite challenging to give an explicit expression of the conditional expectation $\mathbb{E}[L_n^s - L_n \mid L_n]$.

Lemma 3.1. *If X is a random variable and $\mathcal{F}, \mathcal{F}'$ are two σ-fields, satisfying $\mathcal{F}' \subseteq \mathcal{F}$, then*

$$\operatorname{Var}\big(\mathbb{E}[X \mid \mathcal{F}']\big) \leq \operatorname{Var}\big(\mathbb{E}[X \mid \mathcal{F}]\big).$$

Proof. In order to prove $\mathrm{Var}(\mathbb{E}[X \mid \mathcal{F}']) \le \mathrm{Var}(\mathbb{E}[X \mid \mathcal{F}])$, we need only to show that $\mathbb{E}[(\mathbb{E}[X \mid \mathcal{F}'])^2] \le \mathbb{E}[(\mathbb{E}[X \mid \mathcal{F}])^2]$ is true. For this, using Jensen's inequality and the convexity of $f(x) = x^2$, we have

$$\begin{aligned}(\mathbb{E}[X \mid \mathcal{F}'])^2 &= (\mathbb{E}[\mathbb{E}[X \mid \mathcal{F}] \mid \mathcal{F}'])^2 \\ &\le \mathbb{E}[(\mathbb{E}[X \mid \mathcal{F}])^2 \mid \mathcal{F}'].\end{aligned}$$

Taking expectation values on both side of this inequality, the desired result follows. □

Now, we are ready for the proof of Theorem 2.2.

Proof of Theorem 2.2. In order to apply Theorem 1.2, first we need to construct the size-biased coupling L_n^s with respect to L_n. Using the strategy mentioned in Remark 1.1, we choose an index I uniformly at random from the set $\{1, \dots, n-2\}$, then size-bias $I[\sigma_I > \sigma_{I+1}]$ by letting it equal to one, and take the remaining summands conditional on $I[\sigma_I > \sigma_{I+1}] = 1$. We can realize $I[\sigma_I > \sigma_{I+1}] = 1$ by adjusting the order of σ_I and σ_{I+1} such that $\sigma_I > \sigma_{I+1}$, and L_n^s denotes the number of descents in σ after adjusting the order of σ_I and σ_{I+1}.

Next, we need to compute $\mathrm{Var}(\mathbb{E}[L_n^s - L_n \mid L_n])$ and $\mathbb{E}[(L_n^s - L_n)^2]$. It is clear that if the index $2 \le I \le n-3$, then

$$\begin{aligned}L_n^s - L_n &= \big(I[\sigma_{I-1} > \sigma_{I+1}] + 1 + I[\sigma_I > \sigma_{I+2}] \\ &\qquad - I[\sigma_{I-1} > \sigma_I] - I[\sigma_{I+1} > \sigma_{I+2}]\big) I[\sigma_I < \sigma_{I+1}] \\ &= I[\sigma_I < \sigma_{I+1}] - \big(I[\sigma_I < \sigma_{I-1} < \sigma_{I+1}] + I[\sigma_I < \sigma_{I+2} < \sigma_{I+1}]\big).\end{aligned} \tag{3.1}$$

If $I = 1$, then

$$\begin{aligned}L_n^s - L_n &= \big(1 + I[\sigma_1 > \sigma_3] - I[\sigma_2 > \sigma_3]\big) I[\sigma_1 < \sigma_2] \\ &= I[\sigma_1 < \sigma_2] - I[\sigma_1 < \sigma_3 < \sigma_2].\end{aligned} \tag{3.2}$$

If $I = n-2$, then

$$\begin{aligned}L_n^s - L_n &= \big(1 + I[\sigma_{n-3} > \sigma_{n-1}] - I[\sigma_{n-3} > \sigma_{n-2}]\big) I[\sigma_{n-2} < \sigma_{n-1}] \\ &= I[\sigma_{n-2} < \sigma_{n-1}] - I[\sigma_{n-2} < \sigma_{n-3} < \sigma_{n-1}].\end{aligned} \tag{3.3}$$

Set $\mathcal{C} := \sigma(I[\sigma_1 > \sigma_2], \dots, I[\sigma_{n-2} > \sigma_{n-1}])$, then $\sigma(L_n) \subseteq \mathcal{C}$.

From Lemma 3.1, we have $\mathrm{Var}(\mathbb{E}[L_n^s - L_n \mid L_n]) \le \mathrm{Var}(\mathbb{E}[L_n^s - L_n \mid \mathcal{C}])$ and

$$\begin{aligned}&\mathrm{Var}(\mathbb{E}[L_n^s - L_n \mid \mathcal{C}]) \\ &\quad = \frac{1}{(n-2)^2} \mathrm{Var}\bigg\{ I[\sigma_1 < \sigma_2] - I[\sigma_1 < \sigma_3 < \sigma_2]\end{aligned}$$

$$
\begin{aligned}
&\qquad + \sum_{i=2}^{n-3}\big[I[\sigma_i < \sigma_{i+1}] \\
&\qquad\qquad - \big(I[\sigma_i < \sigma_{i-1} < \sigma_{i+1}] + I[\sigma_i < \sigma_{i+2} < \sigma_{i+1}]\big)\big] \\
&\qquad + I[\sigma_{n-2} < \sigma_{n-1}] - I[\sigma_{n-2} < \sigma_{n-3} < \sigma_{n-1}]\Bigg\} \\
&= \frac{1}{(n-2)^2}\operatorname{Var}\Bigg(\sum_{i=1}^{n-2} I[\sigma_i < \sigma_{i+1}] + \sum_{i=1}^{n-3} I[\sigma_i < \sigma_{i+2} < \sigma_{i+1}] \\
&\qquad + \sum_{i=2}^{n-2} I[\sigma_i < \sigma_{i-1} < \sigma_{i+1}]\Bigg) \\
&\le \frac{3}{(n-2)^2}\Bigg\{\operatorname{Var}\Bigg(\sum_{i=1}^{n-2} I[\sigma_i < \sigma_{i+1}]\Bigg) + \operatorname{Var}\Bigg(\sum_{i=1}^{n-3} I[\sigma_i < \sigma_{i+2} < \sigma_{i+1}]\Bigg) \\
&\qquad + \operatorname{Var}\Bigg(\sum_{i=2}^{n-2} I[\sigma_i < \sigma_{i-1} < \sigma_{i+1}]\Bigg)\Bigg\} \\
&=: \frac{3}{(n-2)^2}(S_1 + S_2 + S_3).
\end{aligned} \tag{3.4}
$$

Let us examine the three sums on the right-hand side of (3.4). First, we easily find that

$$S_1 = \frac{n}{12}, \tag{3.5}$$

since $\sum_{i=1}^{n-2} I[\sigma_i < \sigma_{i+1}] = n - 1 - L_n$.

For S_2, we first calculate

$$
\begin{aligned}
&\mathbb{E}\Bigg[\Bigg(\sum_{i=1}^{n-3} I[\sigma_i < \sigma_{i+2} < \sigma_{i+1}]\Bigg)^2\Bigg] \\
&\qquad = \mathbb{E}\Bigg[\sum_{i=1}^{n-3} I[\sigma_i < \sigma_{i+2} < \sigma_{i+1}] \\
&\qquad\qquad + \sum_{i \ne j} I[\sigma_i < \sigma_{i+2} < \sigma_{i+1}] I[\sigma_j < \sigma_{j+2} < \sigma_{j+1}]\Bigg] \\
&\qquad \le \frac{n-3}{6} + \frac{(n-4)(n-5)}{36} = \frac{n^2 - 3n + 2}{36},
\end{aligned}
$$

the inequality is obtained by the fact that

$$\mathbb{P}\big(I[\sigma_i < \sigma_{i+2} < \sigma_{i+1}]I[\sigma_j < \sigma_{j+2} < \sigma_{j+1}] = 1\big) \begin{cases} = 0 & \text{if } |i-j| = 1, \\ \leq \frac{1}{36} & \text{if } |i-j| > 1, \end{cases}$$

for all $i, j \in \{1, \dots, n-3\}$.

Since $\mathbb{E}[\sum_{i=1}^{n-3} I[\sigma_i < \sigma_{i+2} < \sigma_{i+1}]] = (n-3)/6$, we get

$$S_2 \leq \frac{3n-7}{36}. \tag{3.6}$$

We can derive a bound for S_3 in the same manner as for S_2 and we get that it is also bounded by $(3n-7)/36$, namely

$$S_3 \leq \frac{3n-7}{36}. \tag{3.7}$$

Next, inserting these three terms into (3.4), we then obtain that

$$\mathrm{Var}\big(\mathbb{E}\big[L_n^s - L_n \mid \sigma\big]\big) \leq \frac{9n-14}{12(n-2)^2}. \tag{3.8}$$

Finally, the last term we need to bound is $\mathbb{E}[(L_n^s - L_n)^2]$.

Since $\mathbb{E}[(L_n^s - L_n)^2] = \mathbb{E}[\mathbb{E}[(L_n^s - L_n)^2 \mid \mathcal{C}]]$, and

$$\begin{aligned} &\mathbb{E}\big[\mathbb{E}\big[\big(L_n^s - L_n\big)^2 \mid \mathcal{C}\big]\big] \\ &\quad = \frac{1}{n-2}\mathbb{E}\Bigg[\big(I[\sigma_1 < \sigma_2] - I[\sigma_1 < \sigma_3 < \sigma_2]\big)^2 \\ &\qquad + \sum_{i=2}^{n-3}\big(I[\sigma_i < \sigma_{i+1}] \\ &\qquad\qquad - \big(I[\sigma_i < \sigma_{i-1} < \sigma_{i+1}] + I[\sigma_i < \sigma_{i+2} < \sigma_{i+1}]\big)\big)^2 \\ &\qquad + \big(I[\sigma_{n-2} < \sigma_{n-1}] - I[\sigma_{n-2} < \sigma_{n-3} < \sigma_{n-1}]\big)^2\Bigg] \\ &\quad \leq 1, \end{aligned} \tag{3.9}$$

where the last inequality follows from the fact that $|$(3.1)$|$, $|$(3.2)$|$ and $|$(3.3)$|$ are less than or equal to 1. Now combining (3.8), (3.9) with (1.1), we obtain the desired result. □

Before proving Theorem 2.3, let us first recall the concept of Kolmogorov distance between distribution functions. For random variables X and Y, the Kolmogorov distance between their distributions is defined as

$$d_K(X, Y) = \sup_x\big|\mathbb{P}(X \leq x) - \mathbb{P}(Y \leq x)\big|.$$

We know that if a random variable X has Lebesgue density bounded by C, then for any random variable Y, the Wasserstein distance and the Kolmogorov distance between X, Y satisfy the following relationship

$$d_K(X,Y) \le \sqrt{2C d_W(X,Y)}. \tag{3.10}$$

From Theorem 2.2 and the fact that $\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \le \frac{1}{\sqrt{2\pi}}$, $\forall x \in \mathbb{R}$, we have

$$d_K(W,Z) \le \sqrt{\sqrt{\frac{2}{\pi}}\left(\frac{C_1\sqrt{n}}{n-2} + \frac{C_2}{\sqrt{n}}\right)} =: D_k(n). \tag{3.11}$$

Now let us turn to prove Theorem 2.3. We use $\Phi(x)$ for the standard normal distribution function.

Proof of Theorem 2.3. Obviously, if $x \le 0$, then $\mathbb{P}(L_n \le \frac{nx}{2}) = 0$. Thus, it remains to deal with the case that $x > 0$. Since

$$\mathbb{P}\left(L_n \le \frac{nx}{2}\right) = \mathbb{P}\left(\frac{L_n - n/2}{\sqrt{n/12}} \le \frac{n(x-1)}{2\sqrt{n/12}}\right),$$

and by the definition of Kolmogorov distance between the distributions of two random variables and (3.11), we can conclude

$$\left|\mathbb{P}\left(W \le \frac{n(x-1)}{2\sqrt{n/12}}\right) - \Phi\left(\frac{n(x-1)}{2\sqrt{n/12}}\right)\right| \le D_k(n),$$

thus

$$\mathbb{P}\left(W \le \frac{n(x-1)}{2\sqrt{n/12}}\right) \le \Phi\left(\frac{n(x-1)}{2\sqrt{n/12}}\right) + D_k(n).$$

If $0 < x < 1$, then

$$\Phi\left(\frac{n(x-1)}{2\sqrt{n/12}}\right) \to 0 \qquad \text{as } n \to \infty.$$

If $x > 1$, we have

$$\Phi\left(\frac{n(x-1)}{2\sqrt{n/12}}\right) \to 1 \qquad \text{as } n \to \infty.$$

Noticing that $D_k(n) = O(n^{-1/4})$, thus we obtain

$$\lim_{n\to\infty} \mathbb{P}\left(L_n \le \frac{nx}{2}\right) = \begin{cases} 1, & x > 1, \\ 0, & x < 1, \end{cases}$$

which immediately leads to the desired result of Theorem 2.3. □

We need the following two lemmas about concentration inequalities for the proof of Theorem 2.4. Both of these two lemmas can be obtained by using the size-biased coupling method.

Lemma 3.2 (Ghosh and Goldstein (2011), Theorem 1.1). *Let Y be a nonnegative random variable with $\mathbb{E}[Y] = \mu < \infty$ and $\mathrm{Var}(Y) = \sigma^2 < \infty$. Suppose Y^s has the size-biased distribution with respect to Y which satisfies $|Y^s - Y| \le C$ for some constant $C > 0$. If $\mathbb{E}[e^{\theta Y}] < \infty$ for $\theta = 2/C$, then*

$$\mathbb{P}\left(\frac{Y-\mu}{\sigma} \ge t\right) \le \exp\left\{\frac{-t^2}{2(C\mu/\sigma^2 + Ct/(2\sigma))}\right\}, \tag{3.12}$$

for all $t > 0$.

Lemma 3.2 enables us to derive the upper bound of (2.4). In order to obtain an upper bound on $\mathbb{P}(\frac{Y-\mu}{\sigma} \le -t)$, which leads to the desired result of (2.5), we need to establish the following new lemma.

Lemma 3.3. *Let Y be a nonnegative random variable with finite mean μ and variance σ^2. Suppose Y^s has the size-biased distribution with respect to Y. If $|Y^s - Y| \le C$ for some constant $C > 0$, then*

$$\mathbb{P}\left(\frac{Y-\mu}{\sigma} \le -t\right) \le \exp\left\{\frac{-t^2}{2(C\mu/\sigma^2)}\right\}, \tag{3.13}$$

for all $t > 0$.

Proof. Let $m(\theta) := \mathbb{E}[e^{\theta Y}]$, then the definition of size-biased distribution implies

$$m'(\theta) = \mathbb{E}[Ye^{\theta Y}] = \mu\mathbb{E}[e^{\theta Y^s}]. \tag{3.14}$$

For $\theta < 0$, we can make use of Markov's inequality to obtain

$$\begin{aligned}\mathbb{P}\left(\frac{Y-\mu}{\sigma} \le -t\right) &= \mathbb{P}\left(e^{\theta(Y-\mu)/\sigma} \ge e^{-\theta t}\right) \le \frac{\mathbb{E}[e^{\theta(Y-\mu)/\sigma}]}{e^{-\theta t}} \\ &= \exp\left\{\log m\left(\frac{\theta}{\sigma}\right) - \frac{\mu\theta}{\sigma} + \theta t\right\}.\end{aligned} \tag{3.15}$$

Moreover, we need the following inequality, which follows by the convexity of the exponential function in the manner that for any different $x, y \in \mathbb{R}$ we have

$$\frac{e^x - e^y}{x - y} = \int_0^1 e^{ty+(1-t)x}\,\mathrm{d}t \le \int_0^1 te^y + (1-t)e^x\,\mathrm{d}t = \frac{e^x + e^y}{2}. \tag{3.16}$$

Using the fact that $\theta < 0$ and $Y^s \succeq_{\mathrm{st}} Y$ (here $\succeq_{\mathrm{st}}$ stands for the usual stochastic ordering), and putting $x = \theta Y$ and $y = \theta Y^s$ in (3.16) and then taking expectations, we get

$$\mathbb{E}[e^{\theta Y}] - \mathbb{E}[e^{\theta Y^s}] \le \frac{-C\theta}{2}\left(\mathbb{E}[e^{\theta Y}] + \mathbb{E}[e^{\theta Y^s}]\right). \tag{3.17}$$

Since $\theta < 0$, (3.17) can be rewritten as

$$\mathbb{E}[e^{\theta Y^s}] \geq \frac{1 + C\theta/2}{1 - C\theta/2} \mathbb{E}[e^{\theta Y}]. \tag{3.18}$$

From (3.14) and (3.18), we have

$$\frac{m'(\theta)}{m(\theta)} - \mu \geq \frac{C\theta\mu}{1 - C\theta/2},$$

or put otherwise

$$\begin{aligned} \log(m(\theta)) - \mu\theta &= -\int_\theta^0 \frac{m'(s)}{m(s)} - \mu \, \mathrm{d}s \leq -\int_\theta^0 \frac{C\mu s}{1 - Cs/2} \, \mathrm{d}s \\ &\leq -\int_\theta^0 C\mu s \, \mathrm{d}s = \frac{C\mu\theta^2}{2}, \end{aligned} \tag{3.19}$$

thus

$$\begin{aligned} \mathbb{E}[e^{\theta(Y-\mu)/\sigma}] &= \exp\left\{\log m\left(\frac{\theta}{\sigma}\right) - \frac{\mu\theta}{\sigma}\right\} \\ &\leq \exp\left\{\frac{C\mu\theta^2}{2\sigma^2}\right\}. \end{aligned} \tag{3.20}$$

We can apply (3.15) to find that

$$\mathbb{P}\left(\frac{Y - \mu}{\sigma} \leq -t\right) \leq \exp\left\{\frac{C\mu\theta^2}{2\sigma^2} + \theta t\right\}. \tag{3.21}$$

The right-hand side of (3.21) is minimized at $\theta = \frac{-t\sigma^2}{C\mu}$, and substituting this value into (3.21) yields the desired bound. □

Proof of Theorem 2.4. First, we know that $|L_n - L_n^s| \leq 1$ from the expressions of (3.1), (3.2) and (3.3). Next, we represent

$$\mathbb{P}\left(\frac{L_n}{n} - \frac{1}{2} \geq x\right) = \mathbb{P}\left(\frac{L_n - n/2}{\sqrt{n/12}} \geq \frac{xn}{\sqrt{n/12}}\right).$$

Set $t = \frac{xn}{\sqrt{n/12}}$, and substitute this value into (3.12), we get (2.4); similarly,

$$\mathbb{P}\left(\frac{L_n}{n} - \frac{1}{2} \leq -x\right) = \mathbb{P}\left(\frac{L_n - n/2}{\sqrt{n/12}} \leq \frac{-xn}{\sqrt{n/12}}\right).$$

Set $t = \frac{xn}{\sqrt{n/12}}$, and substitute this value into (3.13), we obtain (2.5). □

Acknowledgments

The author would like to thank the anonymous reviewers for their insightful comments. Special thanks to professors Jiang-lun Wu and Xian-Yuan Wu for their helpful discussions and sincere advice. Research supported in part by Natural Science Foundation of China under Grant 10971143, 11271356.

References

Arratia, R., Goldstein, L. and Kochman, F. (2013). Size bias for one and all. Preprint. Available at arXiv:1308.2729.

Bergeron, F., Flajolet, P. and Salvy, B. (1992). Varieties of increasing trees. In *CAAP'92: 17th Colloquium on Trees in Algebra and Programming Rennes* (*France*, *February 26–28*) *Proceedings*. *Lecture Notes in Comput. Sci.* **581**, 24. Berlin: Springer. MR1251994

Chan, D. Y. C., Hughes, B. D., Leong, A. S. and Reed, W. J. (2003). Stochastically evolving networks. *Phys. Rev. E (3)* **68**, 1–24. MR2060980

Gastwirth, J. L. and Bhattacharya, P. K. (1984). Two probability models of pyramid or chain letter schemes demonstrating that their promotional claims are unreliable. *Oper. Res.* **32**, 527–536. MR0755999

Ghosh, S. and Goldstein, L. (2011). Concentration of measures via size-biased couplings. *Probab. Theory Related Fields* **149**, 271–278. MR2773032

Goldstein, L. and Rinott, Y. (1996). Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Probab.* **33**, 1–17. MR1371949

Grossman, R. and Larson, R. G. (1989). Hopf-algebraic structure of families of trees. *J. Algebra* **126**, 184–210. MR1023294

Janic, M., Kuipers, F., Zhou, X. and Van Mieghem, P. (2002). Implications for QoS provisioning based on traceroute measurements. In *Proceedings of the 3rd International Workshop on Quality of Future Internet Services, QofIS2002, Zurich, Switzerland*. *Lecture Notes in Computer Science* **2511**, 3–14. Berlin: Springer.

Meir, A. and Moon, J. W. (1974). Cutting down recursive trees. *Math. Biosci.* **21**, 173–181.

Na, H. S. and Rapoport, A. (1970). Distribution of nodes of a tree by degree. *Math. Biosci.* **6**, 313–329. MR0278985

Najock, D. and Heyde, C. C. (1982). On the number of terminal vertices in certain random trees with an application to stemma construction in philology. *J. Appl. Probab.* **19**, 675–680. MR0664852

Ross, N. (2011). Fundamentals of Stein's method. *Probab. Surv.* **8**, 210–293. MR2861132

Smythe, R. and Mahmoud, H. (1995). A survey of recursive trees. *Theory Probab. Math. Statist.* **51**, 1–27. MR1445048

Tetzlaff, G. T. (2002). Breakage and restoration in recursive trees. *J. Appl. Probab.* **39**, 383–390. MR1908952

School of Mathematical Sciences
Capital Normal University
105 North Xi San Huan
Beijing
China
E-mail: zhangyazhe@cnu.edu.cn