Combinatorics, Probability and Computing (2010) 19, 391–424. © Cambridge University Press 2010
doi:10.1017/S096354830999068X

Random Records and Cuttings in Binary Search Trees

CECILIA HOLMGREN

Uppsala Universitet, Matematiska Institutionen, Box 480 751 06 Uppsala, Sweden
(e-mail: cecilia.holmgren@math.uu.se)

Received 9 November 2007; revised 22 November 2009

We study the number of random records in a binary search tree with n vertices (or equivalently, the number of cuttings required to eliminate the tree). We show that a classical limit theorem for convergence of sums of triangular arrays to infinitely divisible distributions can be used to determine the distribution of this number. The asymptotic distribution of the (normalized) number of records or cuts is found to be weakly 1-stable.

1. Introduction

1.1. Preliminaries

We study the number of records in a random binary search tree. As shown by Janson [13], this number is equivalent in distribution to the number of cuts needed to eliminate this type of tree. Given a rooted tree T with n nodes, let each vertex v have a random value λ_v attached to it, and assume that these values are i.i.d. with a continuous distribution. We say that the value λ_v is a *record* if it is the smallest value in the path from the root to v. Let $X_v(T)$ denote the (random) number of records. Only the order relations of the λ_v's are important, so the distribution of λ_v does not matter, *i.e.*, one can choose any continuous distribution for λ_v, or as another choice let the values λ_v be a random permutation of $\{1, 2, \ldots, n\}$; see [12, 13]. Alternatively, one may attach random variables to the edges and let $X_e(T)$ denote the number of edges with record values.

The same random variable appears when we consider *cuttings* of the tree T (*i.e.*, cuttings and records give random variables with the same distribution), as introduced by Meir and Moon [19] with the following definition. Make a random cut by choosing one vertex (respectively edge) at random. Delete this vertex (respectively edge) so that the tree separates into two parts, and keep only the part containing the root. Continue recursively until the root is cut (respectively only the root is left). Then the total (random) number of cuts made is $X_v(T)$ (respectively $X_e(T)$). The proof of this equivalence uses a natural coupling argument: see [12, 13].

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

A *binary search tree* is a binary tree in which each node is associated to a key, where the keys are drawn from a totally ordered finite set, say $\{1, 2, \dots, n\}$. The keys are drawn in a random order from the set until the set is exhausted. Each possible draw order corresponds to a permutation of the elements in the set (then each permutation is equally likely). The first key is added to the root. The next key is added to the left child of the root if it is smaller than the key of the root, and added to the right child of the root if it is larger than the key of the root. In this way we proceed progressively, inserting key by key. When all keys are drawn, one gets a binary tree on n vertices, such that the keys of the left subtree of any given node v are smaller than the key of v, and the keys of the right subtree are larger than the key of v. For further details, see, *e.g.*, [1, 4].

In [13] the asymptotic distributions for $X_v(T)$ (*i.e.*, the number of records or cuts) are found for random trees that can be constructed as conditioned Galton–Watson trees, *e.g.*, labelled trees and random binary trees. There the proof relies on the fact that the method of moments could be used. In [12] the asymptotic distribution for $X_v(T)$ is found for the deterministic (non-random) complete binary tree. For the complete binary tree it is, however, not possible to use the method of moments. To deal with this, Janson [12] introduced another strategy, which is to approximate $X_v(T)$ by a sum of independent random variables derived from λ_v and then apply a classical limit theorem (see, *e.g.*, [15, Theorem 15.28]) for triangular arrays. From the proof it is easy to see that the method also works for all complete d-ary trees.

In this paper *we consider the case of binary search trees*. These belong to the family of the so-called $\log n$ trees, that is, trees with height $\mathcal{O}(\log n)$. The binary search trees have similar properties to the deterministic complete binary tree considered in [12]. The height in a complete binary tree is $\lfloor \log_2 n \rfloor$, whereas the height in the binary search tree is $\sim C \log n$ (*i.e.*, the first-order asymptotic is $C \log n$), where C is a constant. In the complete binary tree most nodes are close to $\lfloor \log_2 n \rfloor$, *i.e.*, the height of the tree. In binary search trees, on the other hand, most nodes are close to $2 \ln n$ (see, *e.g.*, [4]). Here we extend the methods used by Janson in [12] to show that it is possible to apply a limit theorem (see, *e.g.*, [15, Theorem 15.28]) for convergence of sums of triangular arrays to infinitely divisible distributions to determine the asymptotic distribution of $X_v(T)$.

1.2. Main theorem

Theorem 1.1. *Suppose that $n \to \infty$. Then*

$$\left(X_v(T_n) - \frac{n}{2 \ln n} - \frac{n \ln \ln n}{2 \ln^2 n} \right) \Big/ \frac{n}{4 \ln^2 n} \xrightarrow{d} -W, \tag{1.1}$$

where W has an infinitely divisible distribution with characteristic function

$$\mathbf{E} e^{itW} = \exp\bigl(-\pi |t| + it(-1 - 2 \ln 2 - 2 \ln |t|)\bigr). \tag{1.2}$$

The same result holds for $X_e(T_n)$.

The class of stable distributions is included in the larger class of infinitely divisible distributions. The general formula for the characteristic function of an infinitely divisible

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

distribution is

$$\exp\left(itb - \frac{a^2t}{2} + \int_{-\infty}^{\infty}(e^{itx} - 1 - itx\mathbf{1}[|x| < 1])dv(x)\right), \tag{1.3}$$

for constants $a \geqslant 0$, $b \in \Re$, and v is the so-called Lévy measure. That a distribution of a random variable Z is α-stable for $\alpha \in (0, 2]$ means that, for a sequence of i.i.d. random variables Z_k, $k \geqslant 1$ such that $Z_k \stackrel{d}{=} Z$, we have

$$\sum_{k=1}^{n} Z_k \stackrel{d}{=} n^{\frac{1}{\alpha}}Z + c_n, \quad \forall n,$$

where $c_n \in \Re$ is a constant. The distribution is strictly stable if, for all n, $c_n = 0$, and weakly stable otherwise. If the Lévy measure v in (1.3) satisfies $\frac{dv}{dx} = \frac{c_\pm}{|x|^{\alpha+1}}$ on $\Re_\pm$, for $\alpha \in (0, 2)$ and constants $c_\pm$ the corresponding infinitely divisible distribution is weakly α-stable. The random variable W in (1.1) has support on $(-\infty, \infty)$, and has a heavy tailed distribution. As for other random variables with α-stable distributions, where $\alpha < 2$, the variance of W is infinite. Also, since $\alpha \leqslant 1$, the expected value of W is not defined. The characteristic function in (1.3) of a 1-stable distribution can be simplified to

$$\exp\left(idt - c|t|\left(1 + i\beta\frac{2}{\pi}\operatorname{sign}(t)\ln|t|\right)\right),$$

for constants $c > 0$, $\beta \in [-1, 1]$ and $d \in \Re$. The most well-known 1-stable distribution is the Cauchy distribution. However, in contrast to the distribution of W in (1.1), the Cauchy distribution is strictly 1-stable and symmetric. For further information about stable distributions, see, *e.g.*, [7, Section XVII.3].

Remarks. **(1)** In the proof of (1.1) given below, we get

$$\mathbf{E}e^{itW} = \exp\left(i(-3 - 2\ln 2 + 2\gamma)t + \int_{0}^{\infty}(e^{itx} - 1 - itx\mathbf{1}[x < 1])dv(x)\right), \tag{1.4}$$

where ($\gamma = 0.577\cdots$ is the Euler constant) and the Lévy measure v is supported on $(0, \infty)$ and has density

$$\frac{dv}{dx} = \frac{2}{x^2}.$$

Thus, we see that W has a weakly 1-stable distribution. We can simplify the expression in (1.4) to get (1.2) above.

(2) We note, as in [12], that even if most records occur at height close to $\sim 2\ln n$, where most of the vertices are, from Lemma 2.4 and the proof of Theorem 2.5, it follows that most of the random fluctuations of $X_v(T_n)$ can be explained by the values at heights close to $\ln\ln n$. As explained in [12], a few values at these heights are so small that they reduce a large number of records among their descendants. Vertices of smaller height are too few, and so there is usually no sufficiently small value among them. Vertices of larger height affect only a small subtree.

(3) For all rooted trees T_n (see [12, 13]), $\mathbf{E}X_v(T_n) = \sum_v \frac{1}{h(v)+1}$, and for the edge version $\mathbf{E}X_e(T_n) = \sum_{v\neq\sigma} \frac{1}{h(v)}$ (where σ is the root). For random trees, $\mathbf{E}(X_v(T_n) \mid T_n)$ (respectively

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

$\mathbf{E}(X_e(T_n) \mid T_n))$ is the conditional expected value given the tree T_n. Hence, for random trees,

$$\mathbf{E}(X_v(T_n)) = \mathbf{E}\left(\sum_v \frac{1}{h(v)+1}\right) \quad \text{and} \quad \mathbf{E}(X_e(T_n)) = \mathbf{E}\left(\sum_{v\neq\sigma} \frac{1}{h(v)}\right).$$

Thus, for the binary search tree (compare to [12, Remark 1.3]),

$$\mathbf{E}(X_e(T_n)) - \mathbf{E}(X_v(T_n)) = \mathbf{E}\left(\sum_{v\neq\sigma} \frac{1}{h(v)(h(v)+1)}\right) \sim C\frac{n}{\log^2 n},$$

for some constant $C > 0$, while there is no similar difference in the limit distribution; see Theorem 1.1. As noted for the complete binary tree [12], this is explained by the fact that the expected value is affected by the unlikely event that a node close to the root has a very small value, which reduces the number of records among its descendants significantly. This behaviour suggests it is impossible to use the method of moments to find the record distribution for the binary search tree as one could do for the Galton–Watson trees. Instead we intend to use methods similar to those used for the complete binary tree [12].

(4) The method used here should also work for most other random trees of logarithmic height; hence, the limiting distribution for the number of records or cuts in these trees should also be infinitely divisible and most likely also weakly 1-stable. Thus, in a recent study [9, 10] we have successfully extended the method of the present study to show that for split trees in general (a large class of random trees of logarithmic height for which the binary search tree is one example), the limiting distribution for the number of records (or cuts) is weakly 1-stable. In addition, although using completely different methods, Iksanov and co-workers [5, 11] showed that, for the random recursive tree (another tree of logarithmic height) also, we have

$$\left(X_e(T_n) - \frac{n}{\ln n} - \frac{n\ln\ln n}{\ln^2 n}\right) \Big/ \frac{n}{\ln^2 n} \xrightarrow{d} Y,$$

where Y has a weakly 1-stable distribution with characteristic function

$$\mathbf{E}e^{itY} = \exp\left(-\frac{\pi}{2}|t| + it\ln|t|\right).$$

1.3. Useful facts about binary search trees

In this subsection some of the most important facts and results for the binary search tree are collected for later use.

1.3.1. Properties. We say that an event A_n happens asymptotically almost surely (a.a.s.) if $\mathbf{P}(A_n) \to 1$ as $n \to \infty$; see [14].

The height is a.a.s. $c_1 \ln n + \mathcal{O}(\ln\ln n)$ (where $c_1 = 4.31107\cdots$). Most nodes are at height $2\ln n + \mathcal{O}(\sqrt{\ln n})$ and the peak around the 'belly' is Gaussian. Also, a.a.s. up to $c_2 \ln n$ (where $c_2 = 0.3711\cdots$), all levels are full. For more information the reader is referred to [2], [4] and [1], for example.

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

1.3.2. Quicksort. *Quicksort* is one of the most used sorting algorithms because of its speed and simplicity. It is the standard sorting algorithm in Unix systems, and is used for sorting an array of n distinct numbers. The *Quicksort* algorithm can be described as follows.

If $n = 0$ or $n = 1$, there is nothing to do (the elements are sorted). If $n \geqslant 2$, then pick a number uniformly at random from the list of n numbers. We can think of this as the key that is associated to the root in a binary search tree. Compare the other numbers to it (smaller or larger) to get two subsets of numbers. Next apply the same procedure to each of these two sets if they contain more than one element. The two chosen numbers correspond to the keys of the left and right child of the root in a binary search tree. By proceeding in the same manner we end up with a sorted list of the original numbers. Thus, the random tree that one gets from *Quicksort* is a binary search tree with n nodes, and this is one reason why these trees are very interesting: see, *e.g.*, [4, 6].

Let L_n denote the number of comparisons required. Then it is easy to see that L_n satisfies

$$L_n \stackrel{d}{=} L_{U_n-1} + L^*_{n-U_n} + n - 1, \quad n \geqslant 1,$$

where U_k is uniformly distributed on the set $\{1, 2, \ldots, k\}$, $\stackrel{d}{=}$ denotes equality in distribution and $L^*_j \stackrel{d}{=} L_j$ and $U_n; L_0, L_1, \ldots L_{n-1}; L^*_0, L^*_1, \ldots L^*_{n-1}$ are all independent. This random variable L_n is equal to the so-called *total path length*, which equals the sum of the heights (distances to the root) for all n nodes in the tree. This random variable has been extensively studied: see, *e.g.*, [6], [17] and [18]. Two easily established facts are the expected value of L_n (see, *e.g.*, [16]) and the variance of L_n (see [18, Exercise 6.2.2-8]):

$$\mathbf{E}L_n = 2(n+1)H_n - 4n = 2n\ln n + n(2\gamma - 4) + 2\ln n + 2\gamma + 1 + \mathcal{O}\left(\frac{1}{n}\right),$$

$$\mathbf{Var}L_n = \sigma^2 n^2 - 2n\ln n + \mathcal{O}(n),$$

where $H_n = \sum_{k=1}^{n} \frac{1}{k}$ (the nth harmonic number), $\sigma^2 = 7 - \frac{2}{3}\pi^2 = 0.42026\cdots$, and $\gamma = 0.5772\cdots$ (Euler's constant). In this study we write $\Upsilon(T)$ for the total path length of the binary search tree T.

1.3.3. The subtree method. There is a third way to generate the binary search tree, which is very useful in many applications. We use this representation frequently in the proofs below. The method can be explained briefly as follows.

Associate with each node in the binary search tree T_n (where T_n denotes a binary search tree with n vertices) the size of the subtree rooted at that node. This means that the root of T_n has value n. Since the rank of the root element is equally likely to be $\{1, \ldots, n\}$, the random number N of nodes in the left subtree of the root of T is uniformly distributed on $\{0, \ldots, n-1\}$. It follows that $N \stackrel{d}{=} \lfloor nU \rfloor$, where U is uniformly distributed on [0,1]. The size of the right subtree is $n - 1 - N$, which is distributed as $\lfloor n(1-U) \rfloor$. Now all subtree sizes in the tree can be explained similarly, by introducing independent uniform [0,1] (U) random variables. For further description, see, *e.g.*, [3]. This means that there is an independent uniform random variable $\mathcal{U}_v$ associated with each node v, such that if we let the left and right child of v be v_l, respectively v_r, and suppose that the subtree rooted at v has V nodes, then v_l and v_r have subtree size $\lfloor V\mathcal{U}_v \rfloor$, respectively $\lfloor V(1-\mathcal{U}_v) \rfloor$. This

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

representation of uniform random variables gives that the subtree size for any node v at height j is distributed as

$$\lfloor \cdots \lfloor \lfloor nU_1 \rfloor U_2 \rfloor \cdots U_j \rfloor, \tag{1.5}$$

where $U_r, r \in \{1, \ldots, j\}$, are i.i.d. uniform random variables given by the nodes in the unique path from v to the root. Thus, the value n and the uniform random variables associated to the nodes give all subtree sizes of the tree. For later use, to see that a representation of uniform random variables as in (1.5) is a representation of the subtree size rooted at v, we give a vertex index to these i.i.d. uniform random variables by letting $U_{r,v} := U_r$. Hence, the subtree size of v at height j is

$$\lfloor \cdots \lfloor \lfloor nU_{1,v} \rfloor U_{2,v} \rfloor \cdots U_{j,v} \rfloor.$$

This means, *e.g.*, for the left child v_l (at height $j+1$) of v at height j, $U_{j+1,v_l} = \mathcal{U}_v$, and for the right child v_r, $U_{j+1,v_r} = 1 - \mathcal{U}_v$.

2. Proofs

2.1. Notation

In this subsection some of the notation that we use in the present study is collected.

We use the usual notation lg for the 2-logarithm and ln for the e-logarithm, and let $\{x\} = x - \lfloor x \rfloor$ denote the fractional part of a real number x.

We treat the case $X_v(T_n)$ in Theorem 1.1 in detail and then indicate why the same result holds for $X_e(T_n)$ too. From now on, since it is clear that we consider the vertex model, we just write $X(T_n)$. First let $X(T_n)_y$ be $X(T_n) - 1$ conditioned on the root label $\lambda_\sigma = y$, *i.e.*, the number of records in T_n given the root value $\lambda_\sigma = y$ except from λ_σ, which is always a record.

We write $\stackrel{d}{=}$ for equality in distribution.

We say that $Y_n = o_p(a_n)$ if a_n is a positive number and Y_n is a random variable such that $Y_n/a_n \xrightarrow{p} 0$ as $n \to \infty$.

We say that $Y_n = \mathcal{O}_{L^p}(a_n)$ if a_n is a positive number and Y_n is a random variable such that $(\mathbf{E}(Y_n{}^p))^{\frac{1}{p}} \leqslant Ca_n$ for some constant C.

We sometimes use the notation $m = 2 \ln n$. For simplicity in the proofs below we write $\ln n_i$ when we mean $\max\{1, \ln n_i\}$.

From now on we write T instead of T_n. For a vertex $v \in T$, we let T_v be the subtree of T rooted at v, and let n_v denote the size (number of nodes) of T_v.

We write $\mathrm{Exp}(\theta)$ for an an exponential distribution with parameter θ, *i.e.*, the density function $f(x) = \frac{e^{-x/\theta}}{\theta}$. We can without loss of generality assume that the labels λ_v have an Exp(1) distribution. As mentioned above, this does not affect the distribution of $X(T_n)$. We write U for a uniform random variable distributed on [0,1], *i.e.*, with the density function $f(x) = 1$. We write $\Gamma(k, 1)$ for a gamma random variable with parameters k and 1, *i.e.*, with the density function $f(x) = \frac{e^{-x}x^{k-1}}{(k-1)!}$.

For each vertex v, where $U_{r,v}$ are the uniform random variables defined in Section 1.3.3, let $Y_{k,v} := -\sum_{r=1}^{k} \ln U_{r,v}$. Since the $Y_{k,v}$'s are identically distributed for vertices v on the

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

same level k, we sometimes skip the vertex index and just write Y_k. Using characteristic functions (or from standard results) it follows that $Y_k \stackrel{d}{=} \Gamma(k,1)$.

We write $h(v)$ for the height of v (distance to the root).

We write $P(v) = \sigma, \dots, v$, for the path between the root σ of T to v. Let Λ_{v_i} be the minimum of λ_v along the path $P(v_i) = \sigma, \dots, v_i$, where v_i, $1 \leqslant i \leqslant 2^L$, are the vertices at height $L = \lfloor \beta \lg \ln n \rfloor$ for some constant β. (Below we choose $\beta \in (1/(\lg 3 - 1), 2)$.) Thus, the definition of Λ_{v_i} and the assumption $\lambda_v \stackrel{d}{=} \mathrm{Exp}(1)$ give $\Lambda_{v_i} \stackrel{d}{=} \mathrm{Exp}(1/(L+1))$. For simplicity we write $T_i := T_{v_i}$, $n_i := n_{v_i}$ and $\Lambda_i := \Lambda_{v_i}$. We write $h_i(v) := h(v) - L$ (*i.e.*, this is the height in the subtree T_i of a vertex $v \in T_i$).

We write $\varphi(T_v) := \mathbf{E}(X(T_v) \mid T_v)$. We define for vertices v_i, $1 \leqslant i \leqslant 2^L$, the conditional expected value given the tree T_i and Λ_i as $\varphi(T_i, \Lambda_i) := \mathbf{E}(X(T_i)_{\Lambda_i} \mid T_i, \Lambda_i)$. (We can think of $X(T_i)_{\Lambda_i}$ as $X(T_i) - 1$ conditioned on the root label $\lambda_{v_i} = \Lambda_i$.) Similarly, we define the conditional variance $\psi(T_i, \Lambda_i) := \mathbf{Var}(X(T_i)_{\Lambda_i} \mid T_i, \Lambda_i)$.

The conditional expected value of a random variable Z given the subtree size n_i of T_i is denoted by $\mathbf{E}_{n_i}(Z) := \mathbf{E}(Z \mid n_i)$, and similarly we let $\mathbf{Var}_{n_i}(Z) := \mathbf{Var}(Z \mid n_i)$.

We also define $\mathscr{G}_j$ for $j = 1, 2, \dots$, as the σ-field generated by $\{U_{r,v},\ r \in \{1, 2, \dots, j\}\}$ for all vertices v with $h(v) = j$, or equivalently this is the σ-field generated by $\{\mathcal{U}_v,\ h(v) < j\}$.

Finally, we write $\{\xi_v := \frac{mn_v}{n} e^{-m\lambda_v}\}$; recall that $m = 2 \ln n$. This notation is used in the later part of the proof when we consider triangular arrays.

2.2. Expressing the normalized number of records as a sum of triangular arrays

Recall from Section 2.1 that we define $\varphi(T_i, \Lambda_i) := \mathbf{E}(X(T_i)_{\Lambda_i} \mid T_i, \Lambda_i)$, where T_i is the subtree rooted at v_i at height L, and Λ_i is the minimum of λ_v in the path from v_i to the root σ of T.

Lemma 2.1. *For all subtrees T_i rooted at v_i with $h(v_i) = L$, conditioning on the subtree size n_i gives*

$$\varphi(T_i, \Lambda_i) = \frac{n_i}{2 \ln n_i}(1 - e^{-(2 \ln n_i)\Lambda_i}) + \frac{n_i}{4 \ln^2 n_i} - \frac{\Upsilon(T_i) - 2n_i \ln n_i}{4 \ln^2 n_i} + \mathcal{O}_{L^1}\left(\frac{n_i}{\ln^{2.2} n_i}\right), \tag{2.1}$$

where $\Upsilon(T_i)$ is the total path length of the subtree T_i.

This conditional expected value appears in Lemma 2.3 below. (This is the result corresponding to [12, Lemmas 2.1 and 2.2].)

Proof of Lemma 2.1. Let for each vertex $v \in T_i$, I_v be the indicator that λ_v is the minimum value (the smallest of all the values attached to the nodes in the unique path from v to the root v_i) given T_i and Λ_i. We get that $\varphi(T_i, \Lambda_i) = \sum_{v \neq v_i} \mathbf{E} I_v$. If $h_i(v) = j$, let $v_i, v_{i1}, \dots, v_{ij} = v$ be the vertices in the path from the root v_i to v. Then $I_v = 1$, if and only if $\lambda_{v_{ij}} < \Lambda_i$ and $\lambda_{v_{ik}} > \lambda_{v_{ij}}$ for $k \in \{1, \dots, j-1\}$. Since the λ_v's (for all vertices v in T_i) are independent Exp(1) random variables, we have

$$\mathbf{E}(I_v) = \int_0^{\Lambda_i} \prod_{k=1}^{j-1} \mathbf{P}(\lambda_{v_i k} > x) e^{-x} dx = \int_0^{\Lambda_i} e^{-jx} dx = \frac{1 - e^{-j\Lambda_i}}{j}. \tag{2.2}$$

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

Thus,

$$\varphi(T_i, \Lambda_i) = \sum_{v \neq v_i} \frac{1 - e^{-h_i(v)\Lambda_i}}{h_i(v)}. \tag{2.3}$$

We remember from Section 1.3.1 above that the binary search tree of size n has most nodes at height $2\ln n + \mathcal{O}(\sqrt{\ln n})$.

Roughly, the idea of the proof is that we first find the Taylor expansion of $\sum_{v \neq v_i} \frac{1}{h_i(v)}$ about $2\ln n_i$ up to the error term $\mathcal{O}_{L^1}\left(\frac{n_i}{\ln^{2.2} n_i}\right)$. Then, by using that many nodes are close to $2\ln n_i$ we show that if we multiply this expansion by $(1 - e^{-(2\ln n_i)\Lambda_i})$, from (2.3) we get (2.1).

In [1, pp. 1043–1044] there is a sharp profile estimate of the binary search tree. Uniformly for all $k \in [(1+\epsilon)\ln n, (2+\sqrt{2}-\epsilon)\ln n]$, the number of nodes at height k in the binary search tree of size n (this number is denoted by $V_k(n)$ in [1]) is

$$\frac{n}{\sqrt{4\pi \ln n}} \cdot e^{-\frac{(k-2\ln n)^2}{4\ln n}} + \mathcal{O}_{L^1}\left(\frac{n}{\ln n}\right). \tag{2.4}$$

(The profile result in (2.4) is obtained from a generating function and asymptotics for Stirling numbers of the first kind.)

We say that a vertex $v \in T_i$ is 'good' if

$$2\ln n_i - \ln^{0.6} n_i \leqslant h_i(v) \leqslant 2\ln n_i + \ln^{0.6} n_i,$$

and 'bad' otherwise. As was mentioned above, 'most' nodes are in the strip of the good vertices. One deduces easily from [1, equation (3)] that the number of 'bad' nodes in T_i is bounded by $\mathcal{O}_{L^1}\left(n_i^{1-\epsilon}\right)$ for some $\epsilon > 0$. Thus, since the number of bad nodes is bounded by $\mathcal{O}_{L^1}\left(\frac{n_i}{\ln^{2.2} n_i}\right)$, they can be ignored.

If we expand $\frac{1}{h_i(v)}$ for arbitrary good $v \in T_i$,

$$\frac{1}{h_i(v)} = \frac{1}{2\ln n_i} - \frac{h_i(v) - 2\ln n_i}{4\ln^2 n_i} + \frac{(h_i(v) - 2\ln n_i)^2}{8\ln^3 n_i} + \mathcal{O}\left(\frac{|(h_i(v) - 2\ln n_i)^3|}{\ln^4 n_i}\right).$$

This implies, by summing over all nodes, that the total sum is

$$\sum_{v \neq v_i} \frac{1}{h_i(v)} = \frac{n_i}{2\ln n_i} - \frac{\Upsilon(T_i) - 2n_i \ln n_i}{4\ln^2 n_i} + \frac{\sum_v (h_i(v) - 2\ln n_i)^2}{8\ln^3 n_i} + \mathcal{O}_{L^1}\left(\frac{n_i}{\ln^{2.2} n_i}\right). \tag{2.5}$$

For the third term in the expansion in (2.5) we use the profile estimate in (2.4). We show that the error term $\mathcal{O}_{L^1}\left(\frac{n}{\ln n}\right)$ in the profile estimate in (2.4) is good enough for this term. The distance for the good vertices to $2\ln n_i$ is less than $\ln^{0.6} n_i$, and there are at most $2 \cdot \ln^{0.6} n_i$ levels. Thus, the contribution of the error term in (2.4) to the third term in the expansion in (2.5) is bounded by

$$\mathcal{O}_{L^1}\left(\frac{n_i}{\ln n_i}\right) \cdot \frac{(\ln^{0.6} n_i)^2 \cdot 2\ln^{0.6} n_i}{8\ln^3 n_i} = \mathcal{O}_{L^1}\left(\frac{n_i}{\ln^{2.2} n_i}\right).$$

Hence, the third term in the Taylor estimate in (2.5) is

$$\sum_{k=-\infty}^{\infty} (k - 2\ln n_i)^2 e^{\frac{(k-2\ln n_i)^2}{4\ln n_i}} \cdot \frac{n_i/\sqrt{4\pi \ln n_i}}{8\ln^3 n_i} + \mathcal{O}_{L^1}\left(\frac{n_i}{\ln^{2.2} n_i}\right). \tag{2.6}$$

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

Note that we can take the sum over all k in (2.6), since the sums over the tails for $k \leqslant 2\ln n_i - \ln^{0.6} n_i$ and $k \geqslant 2\ln n_i + \ln^{0.6} n_i$ in (2.6) are dominated by the error term $\mathcal{O}_{L^1}\big(\frac{n_i}{\ln^{2.2} n_i}\big)$.

To calculate the sum

$$\sum_{k=-\infty}^{\infty} (k - 2\ln n_i)^2 e^{\frac{(k-2\ln n_i)^2}{4\ln n_i}},$$

we use integral approximation, with an error term that is easily calculated to be bounded by $\mathcal{O}\big(\frac{n_i}{\ln^{2.2} n_i}\big)$, and get

$$\int_{-\infty}^{\infty} (x - 2\ln n_i)^2 e^{\frac{(x-2\ln n_i)^2}{4\ln n_i}} dx = 2\sqrt{4\pi}\ln n_i \sqrt{\ln n_i}.$$

Thus, (2.5) implies that

$$\sum_{v \neq v_i} \frac{1}{h_i(v)} = \frac{n_i}{2\ln n_i} - \frac{\Upsilon(T_i) - 2n_i \ln n_i}{4\ln^2 n_i} + \frac{n_i}{4\ln^2 n_i} + \mathcal{O}_{L^1}\left(\frac{n_i}{\ln^{2.2} n_i}\right). \tag{2.7}$$

Remark. We cannot use the profile approximation in (2.4) to calculate the second term in (2.5) since the error term $\mathcal{O}_{L^1}\big(\frac{n_i}{\ln n_i}\big)$ is too large. The total path length random variable $\Upsilon(T_i)$ (given n_i) has a variance of order $\mathcal{O}(n_i^2)$: see Section 1.3.2. Thus, the error estimate is not small enough.

We now show that

$$\sum_{v \neq v_i} \frac{1}{h_i(v)} (e^{-h_i(v)\Lambda_i} - e^{-2\Lambda_i \ln n_i}) = \mathcal{O}_{L^1}\left(\frac{n_i}{\ln^{2.2} n_i}\right), \tag{2.8}$$

which by (2.3) obviously implies that

$$\varphi(T_i, \Lambda_i) = \sum_{v \neq v_i} \frac{1}{h_i(v)} (1 - e^{-(2\ln n_i)\Lambda_i}) + \mathcal{O}_{L^1}\left(\frac{n_i}{\ln^{2.2} n_i}\right). \tag{2.9}$$

For simpler calculations below we consider the term $e^{-\Lambda_i \lfloor 2\ln n_i \rfloor}$ instead of $e^{-(2\ln n_i)\Lambda_i}$ in (2.8) and (2.9). That one can do this follows from the fact that multiplying the Taylor estimate in (2.7) by $e^{-\Lambda_i \lfloor 2\ln n_i \rfloor}$ gives the same expression, up to the error term $\mathcal{O}_{L^1}\big(\frac{n_i}{\ln^{2.2} n_i}\big)$, as multiplying by $e^{-2\Lambda_i \ln n_i}$.

For $j > 0$,

$$e^{(-\lfloor 2\ln n_i \rfloor + j)\Lambda_i} = e^{-\Lambda_i \lfloor 2\ln n_i \rfloor} + e^{(-\lfloor 2\ln n_i \rfloor + j)\Lambda_i} \cdot (1 - e^{-j\Lambda_i})$$

and

$$e^{(-\lfloor 2\ln n_i \rfloor - j)\Lambda_i} = e^{-\Lambda_i \lfloor 2\ln n_i \rfloor} + e^{(-\lfloor 2\ln n_i \rfloor + j)\Lambda_i} \cdot (e^{-2j\Lambda_i} - e^{-j\Lambda_i}).$$

We use the profile estimate in (2.4); in particular, we need to show that

$$\sum_{j=1}^{\ln^{0.6} n_i} \left(\left(\frac{n_i}{\sqrt{4\pi \ln n_i}} e^{\frac{-(j+\{2\ln n_i\})^2}{4\ln n_i}} + \mathcal{O}_{L^1}\left(\frac{n_i}{\ln n_i}\right) \right) \cdot e^{(-\lfloor 2\ln n_i \rfloor + j)\Lambda_i} (1 - e^{-j\Lambda_i}) \frac{1}{\lfloor 2\ln n_i \rfloor - j} \right) = \mathcal{O}_{L^1}\left(\frac{n_i}{\ln^{2.2} n_i}\right) \tag{2.10}$$

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

and

$$\sum_{j=1}^{\ln^{0.6} n_i} \left(\left(\frac{n_i}{\sqrt{4\pi \ln n_i}} \cdot e^{\frac{-(j-\{2\ln n_i\})^2}{4\ln n_i}} + \mathcal{O}_{L^1}\left(\frac{n_i}{\ln n_i}\right)\right) \right.$$
$$\left. \cdot\, e^{(-\lfloor 2\ln n_i\rfloor + j)\Lambda_i} (e^{-2j\Lambda_i} - e^{-j\Lambda_i}) \frac{1}{\lfloor 2\ln n_i\rfloor + j} \right) = \mathcal{O}_{L^1}\left(\frac{n_i}{\ln^{2.2} n_i}\right), \tag{2.11}$$

where $\{2\ln n_i\}$ is the fractional part of $2\ln n_i$.

We show that (2.10) and (2.11) hold. Taylor expansion gives that the sum in (2.10) is equal to

$$\sum_{j=1}^{\ln^{0.6} n_i} \left(\frac{n_i}{\sqrt{4\pi \ln n_i}} e^{\frac{-(j+\{2\ln n_i\})^2}{4\ln n_i}} + \mathcal{O}_{L^1}\left(\frac{n_i}{\ln n_i}\right)\right) e^{(-\lfloor 2\ln n_i\rfloor + j)\Lambda_i} \mathcal{O}\left(\frac{j\Lambda_i}{\ln n_i}\right).$$

We calculate

$$\int_{j=0}^{\infty} j e^{\frac{-(j+\{2\ln n_i\})^2}{4\ln n_i}} dj = \mathcal{O}(\ln n_i).$$

Since Λ_i is an $\mathrm{Exp}\left(\frac{1}{L+1}\right)$ random variable, we have that, for all $j \leqslant \ln^{0.6} n_i$,

$$\begin{aligned} \mathbf{E}\Lambda_i e^{(-\lfloor 2\ln n_i\rfloor + j)\Lambda_i} &= \int_0^{\infty} (L+1) y e^{(-\lfloor 2\ln n_i\rfloor + j)y} e^{-y(L+1)} dy \\ &= \frac{L+1}{(\lfloor 2\ln n_i\rfloor - j + L + 1)^2}. \end{aligned} \tag{2.12}$$

Thus,

$$\sum_{j=1}^{\ln^{0.6} n_i} \frac{n_i}{\sqrt{4\pi \ln n_i}} e^{\frac{-(j+\{2\ln n_i\})^2}{4\ln n_i}} e^{(-\lfloor 2\ln n_i\rfloor + j)\Lambda_i} \mathcal{O}\left(\frac{j\Lambda_i}{\ln n_i}\right) = \mathcal{O}_{L^1}\left(\frac{n_i}{\ln^{2.2} n_i}\right).$$

We easily check that the error term in the profile estimate in (2.4) is small enough, *i.e.*,

$$\sum_{j=1}^{\ln^{0.6} n_i} \mathcal{O}_{L^1}\left(\frac{n_i}{\ln n_i}\right) e^{(-\lfloor 2\ln n_i\rfloor + j)\Lambda_i} \mathcal{O}\left(\frac{j\Lambda_i}{\ln n_i}\right) = \mathcal{O}_{L^1}\left(\frac{n_i}{\ln^{2.2} n_i}\right).$$

Hence, (2.10) holds; by similar calculations one shows that also (2.11) holds. Thus, it follows that (2.9) is satisfied. We have $e^{-(2\ln n_i)\Lambda_i} = \mathcal{O}_{L^1}\left(\frac{L}{\ln n_i}\right)$. Hence,

$$e^{-(2\ln n_i)\Lambda_i} \cdot \mathcal{O}\left(\frac{n_i}{\ln^2 n_i}\right) = \mathcal{O}_{L^1}\left(\frac{n_i}{\ln^{2.2} n_i}\right).$$

Recall that we write $\mathbf{E}_{n_i}(.) := \mathbf{E}(.|n_i)$, *i.e.*, this is the conditional expected value given the subtree size n_i. Since also $\mathbf{E}(e^{-(2\ln n_i)\Lambda_i})^2 = \mathcal{O}\left(\frac{L}{\ln n_i}\right)$ and (see Section 1.3.2)

$$\mathbf{E}_{n_i}\left(\frac{\Upsilon(T_i) - 2n_i \ln n_i}{4\ln^2 n_i}\right)^2 = \mathcal{O}\left(\frac{{n_i}^2}{\ln^4 n_i}\right),$$

we obtain

$$\frac{(\Upsilon(T_i) - 2n_i \ln n_i) e^{-(2\ln n_i)\Lambda_i}}{4\ln^2 n_i} = \mathcal{O}_{L^2}\left(\frac{n_i}{\ln^{2.2} n_i}\right) = \mathcal{O}_{L^1}\left(\frac{n_i}{\ln^{2.2} n_i}\right)$$

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

(where the last equality is obtained by using the well-known Lyapunov inequality, *e.g.*, [8, p. 129]), and thus Lemma 2.1 follows from (2.7) and (2.9). □

Recall from Section 2.1 that we define $\psi(T_i, \Lambda_i) := \mathbf{Var}(X(T_i)_{\Lambda_i} \mid T_i, \Lambda_i)$, and that we write $\mathbf{E}_{n_i}(.) := \mathbf{E}(.|n_i)$ for the conditional expected value given n_i.

Lemma 2.2. *For all vertices v_i with $h(v_i) = L$, conditioned on n_i,*

$$\mathbf{E}_{n_i}(\psi(T_i, \Lambda_i)) = \mathcal{O}\left(\frac{n_i^2}{\ln^3 n_i}\right).$$

(This is the result corresponding to [12, Lemma 2.1 (2nd part)].) This estimate is used in Lemma 2.3 below.

Proof of Lemma 2.2. For all vertices $v \in T_i$, let I_v be the same indicator as in the proof of Lemma 2.1 above, *i.e.*, the indicator that λ_v is the minimum value of the λ_v values from v to the root v_i given T_i and Λ_i.

Suppose that v and w are two vertices in T_i at height $h_i(v) = j, h_i(w) = k$, with their last common ancestor at height $h_i(u) = h$. Suppose first that $h < j, h < k$. (In fact we show below that other choices of vertices can be ignored.) Let $\{v_i, u_1, \ldots, u_h = u\}$ be the vertices in the path from v_i to u and let $Z = \min\{\lambda_{u_s} : 1 \leqslant s \leqslant h\}$. Conditioned on Z, I_v and I_w are independent. Let $Z \wedge \Lambda_i$ denote the minimum of Z and Λ_i. Since v has height $j - h$ above u, (2.2) yields

$$\mathbf{E}(I_v \mid Z) = \frac{1 - e^{-(j-h)(Z \wedge \Lambda_i)}}{j - h},$$

and similarly for I_w; see [12, p. 244]. As in [12, equation (6), p. 245], we get

$$\mathbf{E}(I_v I_w) = \frac{1}{j-h}\frac{1}{k-h}\left(1 - e^{-h\Lambda_i} - \frac{h}{j}(1 - e^{-j\Lambda_i}) - \frac{h}{k}(1 - e^{-k\Lambda_i}) + \frac{h}{j+k-h}(1 - e^{-(j+k-h)\Lambda_i}) + e^{-h\Lambda_i} - e^{-j\Lambda_i} - e^{-k\Lambda_i} + e^{-(j+k-h)\Lambda_i}\right). \quad (2.13)$$

The covariance of I_v and I_w is

$$\mathbf{Cov}(I_v, I_w) = \mathbf{E}(I_v I_w) - \mathbf{E}(I_v)\mathbf{E}(I_w).$$

Remark. The fewer common vertices in v's and w's paths to the root, the smaller the covariance of I_v and I_w. Consequently, to minimize the covariance, we want $h_i(v) = j$ and $h_i(w) = k$ to be large and $h_i(u) = h$ to be small.

We say that a pair (v, w) is 'good' if j and k satisfy

$$2\ln n_i - \ln^{0.6} n_i \leqslant j, k \leqslant 2\ln n_i + \ln^{0.6} n_i$$

and $h \leqslant \alpha \ln n_i$ for some α that we choose small enough, and 'bad' otherwise. If α is chosen to be small it is obvious that the covariance estimate for the good pairs of vertices in the binary search tree is no worse than the order of the estimate for the complete binary tree

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

in [12]. Thus (from [12, equation (7), p. 245]), by (2.2) and (2.13) above, for a good pair

$$\begin{aligned}\mathbf{Cov}(I_v, I_w) &= \mathbf{E}(I_v I_w) - \mathbf{E}(I_v)\mathbf{E}(I_w)\\ &= \frac{1}{jk} e^{-(j+k-h)\Lambda_i}(1 - e^{-h\Lambda_i}) + \mathcal{O}\left(\frac{h}{\ln^3 n_i}\right)\\ &= \mathcal{O}\left(\frac{1}{\ln^2 n_i} e^{-(j+k-h)\Lambda_i} h\Lambda_i\right) + \mathcal{O}\left(\frac{h}{\ln^3 n_i}\right)\\ &= \mathcal{O}_{L^1}\left(\frac{hL}{\ln^4 n_i}\right) + \mathcal{O}\left(\frac{h}{\ln^3 n_i}\right) = \mathcal{O}_{L^1}\left(\frac{h}{\ln^3 n_i}\right), \end{aligned} \tag{2.14}$$

where, in the fourth equality, we use that $e^{-(j+k-h)\Lambda_i} h\Lambda_i = \mathcal{O}_{L^1}\left(\frac{hL}{\ln^2 n_i}\right)$ (which is obtained by calculations similar to those in (2.12) above).

Recall the result of the profile estimate we discussed in the proof of Lemma 2.1 that there are $\mathcal{O}_{L^1}(n_i^{1-\epsilon})$ nodes that are not in the strip

$$2\ln n_i - \ln^{0.6} n_i \leqslant h_i(v) \leqslant 2\ln n_i + \ln^{0.6} n_i.$$

Thus, it follows that if v and w are not both in this strip, there are $\mathcal{O}_{L^1}(n_i^{2-\epsilon})$ ways to choose the pair v, w for some $\epsilon > 0$.

Let n_{hk}, $1 \leqslant k \leqslant 2^h$, be the subtree sizes rooted at height h (they are random numbers) in T_i. For a given height h there are at most $\sum_{k=1}^{2^h} n_{hk}^2$ choices of v and w. Now we use the subtree method above (see Section 1.3.3) to calculate the expected value of n_{hk}^2, $1 \leqslant k \leqslant 2^h$. Note that they all have the same expected value since they are on the same level (height) from the root v_i. It is easy to see that

$$\mathbf{E}_{n_i}(n_{hk}) \leqslant \left(\prod_{r=1}^{h} \mathbf{E}U_r\right) n_i = \frac{n_i}{2^h} \quad \text{and} \quad \mathbf{E}_{n_i}(n_{hk}^2) \leqslant \left(\prod_{r=1}^{h} \mathbf{E}U_r^2\right) n_i^2 = \frac{n_i^2}{3^h}, \tag{2.15}$$

since the U_r's are independent uniform random variables. Consequently, the expected number of pairs of vertices with $h > \alpha \ln n_i$ is

$$\mathbf{E}_{n_i}\left(\sum_{h>\alpha\ln n_i} \sum_{k=1}^{2^h} n_{hk}^2\right) = \mathcal{O}\left(\sum_{h>\alpha\ln n_i} \frac{n_i^2 \cdot 2^h}{3^h}\right) = \mathcal{O}\left(\frac{n_i^2}{\ln^3 n_i}\right).$$

Thus, the number of 'bad' pairs of vertices is $\mathcal{O}_{L^1}\left(\frac{n_i^2}{\ln^3 n_i}\right)$.

For the random variable ψ depending on the random tree T_i and Λ_i, we have

$$\psi(T_i, \Lambda_i) = \sum_{v,w \in T_i} \mathbf{Cov}(I_v, I_w).$$

Since the number of 'bad' pairs of vertices is $\mathcal{O}_{L^1}\left(\frac{n_i^2}{\ln^3 n_i}\right)$ and there is the obvious upper bound $\mathbf{Cov}(I_v, I_w) \leqslant 1$, we get

$$\mathbf{E}_{n_i}(\psi(T_i, \Lambda_i)) = \mathbf{E}_{n_i}\left(\sum_{\text{good } (v,w)\in T_i} \mathbf{Cov}(I_v, I_w)\right) + \mathcal{O}\left(\frac{n_i^2}{\ln^3 n_i}\right).$$

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

By using $\mathbf{E}_{n_i}(n_{hk}^2) \leqslant \frac{n_i^2}{3^h}$ (see equation (2.15)) and equation (2.14), we deduce

$$\mathbf{E}_{n_i}(\psi(T_i,\Lambda_i)) = \mathcal{O}_{L^1}\left(\mathbf{E}_{n_i}\left(\sum_{\text{good } (v,w)\in T_i} \frac{h}{\ln^3 n_i}\right)\right) + \mathcal{O}\left(\frac{n_i^2}{\ln^3 n_i}\right) \tag{2.16}$$

$$= \mathcal{O}\left(\sum_h \frac{n_i^2 \cdot 2^h \cdot h}{3^h \ln^3 n_i}\right) + \mathcal{O}\left(\frac{n_i^2}{\ln^3 n_i}\right) = \mathcal{O}\left(\frac{n_i^2}{\ln^3 n_i}\right). \qquad \square$$

The estimate in (2.16) is used in the proof of the following result.

Lemma 2.3. *For a binary search tree with n vertices, let $v_i, 1 \leqslant i \leqslant 2^L$, be the vertices at height $L = \lfloor \beta \lg\ln n \rfloor$ for $\beta > \frac{1}{\lg 3 - 1}$. Then*

$$X(T_n) = \sum_{i=1}^{2^L} \varphi(T_i,\Lambda_i) + o_p\left(\frac{n}{\ln^2 n}\right).$$

(This is the result corresponding to [12, Lemma 2.3, equation (10)].)

Proof of Lemma 2.3. We write the number of records as $\{V^* + V_1 + \ldots + V_{2^L}\}$, where V^* is the number of records with height less than L and V_i is the number of records in the subtree T_i that is rooted at v_i at height L except for λ_{v_i}.

Let $\mathcal{F}_L$ be the σ-field generated by $\{\lambda_v : h(v) \leqslant L\}$ and let $\mathcal{F}_L^*$ be the σ- field generated by T_n and $\mathcal{F}_L$. Recall that Λ_i is the minimum of λ_v in the path from v_i, $h(v_i) = L$, to the root σ of T. Conditioned on $\mathcal{F}_L^*$, V^* and Λ_i become fixed and $\mathbf{E}(V_i \mid \mathcal{F}_L^*) = \varphi(T_i,\Lambda_i)$. Conditioned on $\mathcal{F}_L^*$, the variables V_i and $V_j, i \neq j$, are independent, and thus

$$\mathbf{E}\left(\left(X(T_n) - V^* - \sum_{i=1}^{2^L} \varphi(T_i,\Lambda_i)\right)^2 \,\middle|\, \mathcal{F}_L^*\right)$$

$$= \mathbf{E}\left(\left(\sum_{i=1}^{2^L} V_i - \varphi(T_i,\Lambda_i)\right)^2 \,\middle|\, \mathcal{F}_L^*\right)$$

$$= \sum_{i=1}^{2^L} \mathbf{E}\left(\left(V_i - \varphi(T_i,\Lambda_i)\right)^2 \mid \mathcal{F}_L^*\right) = \sum_{i=1}^{2^L} \psi(T_i,\Lambda_i). \tag{2.17}$$

Now, taking expectations in (2.17) yields

$$\mathbf{E}\left(\left(X(T_n) - V^* - \sum_{i=1}^{2^L} \varphi(T_i,\Lambda_i)\right)^2\right) = \mathbf{E}\left(\sum_{i=1}^{2^L} \psi(T_i,\Lambda_i)\right) = \sum_{i=1}^{2^L} \mathbf{E}\psi(T_i,\Lambda_i). \tag{2.18}$$

We observe the obvious fact that the sum of those n_i, $i \in \{1,\ldots,2^L\}$, that are less than $\frac{n}{2^{kL}}$ for k large enough, is bounded by

$$2^L \cdot \frac{n}{2^{kL}} = \mathcal{O}\left(\frac{n}{\ln^3 n}\right). \tag{2.19}$$

(Note that by choosing k large enough in (2.19), the power of the logarithm can be taken to be arbitrarily large.)

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

Lemma 2.2 and (2.19) give that

$$\sum_{i=1}^{2^L} \mathbf{E}_{n_i}(\psi(T_i,\Lambda_i)) = \mathcal{O}\Big(\sum_{i=1}^{2^L} \frac{n_i^2}{\ln^3 n_i}\Big) = \mathcal{O}\Big(\sum_{i=1}^{2^L} \frac{n_i^2}{\ln^3 n}\Big) + o\Big(\frac{n^2}{\ln^4 n}\Big). \tag{2.20}$$

The expected value of this is the expected value in (2.18). Thus, from (2.15) and assuming that the constant $\beta > \frac{1}{\lg 3-1}$ in $L = \beta \lg\ln n$, we get

$$\mathbf{E}\Big(\Big(X(T_n) - V^* - \sum_{i=1}^{2^L} \varphi(T_i,\Lambda_i)\Big)^2\Big) = \mathcal{O}\Big(\frac{2^L n^2}{3^L \ln^3 n}\Big) = o\Big(\frac{n^2}{\ln^4 n}\Big).$$

Thus, the well-known Markov inequality (*e.g.*, [8, p. 120]) gives

$$X(T_n) - V^* - \sum_{i=1}^{2^L} \varphi(T_i,\Lambda_i) = o_p\Big(\frac{n}{\ln^2 n}\Big).$$

The result follows because also

$$V^* < 2^{L+1} = \mathcal{O}(\ln^\beta n) = o\Big(\frac{n}{\ln^2 n}\Big). \qquad \square$$

Using the bound in (2.19) for the sum of the subtree sizes with $n_i < \frac{n}{2^{kL}}$ (for large enough k), Taylor expansion gives

$$\sum_{i=1}^{2^L} \frac{n_i}{\ln^{2.2} n_i} = \frac{n}{\ln^{2.2} n} + \mathcal{O}\Big(\frac{n\ln\ln n}{\ln^3 n}\Big). \tag{2.21}$$

Using this and the fact that $\mathcal{O}_{L^1}\big(\frac{n}{\ln^{2.2} n}\big) = o_p\big(\frac{n}{\ln^2 n}\big)$ (from the Markov inequality), Lemmas 2.1 and 2.3 give, for $\beta > \frac{1}{\lg 3-1}$,

$$X(T_n) = \sum_{i=1}^{2^L}\Big(-\frac{n_i e^{-(2\ln n_i)\Lambda_i}}{2\ln n_i} + \frac{n_i}{4\ln^2 n_i} + \frac{n_i}{\ln n_i} - \frac{\Upsilon(T_i)}{4\ln^2 n_i}\Big) + o_p\Big(\frac{n}{\ln^2 n}\Big).$$

As in (2.21), we obtain

$$\sum_{i=1}^{2^L} \frac{n_i}{4\ln^2 n_i} = \frac{n}{4\ln^2 n} + \mathcal{O}\Big(\frac{n\ln\ln n}{\ln^3 n}\Big).$$

Comparing with [12, equations (12, 13), p. 247] for $n_i \geqslant \frac{n}{2^{kL}}$, we get

$$\mathbf{E}_{n_i}\big(|e^{-(2\ln n_i)\Lambda_i} - e^{-(2\ln n)\Lambda_i}|\big) = \frac{L+1}{L+1+2\ln n_i} - \frac{L+1}{L+1+2\ln n} = \mathcal{O}\Big(\frac{L^2}{\ln^2 n}\Big),$$

and it follows that

$$\mathbf{E}\Big(\Big|\frac{n_i}{2\ln n_i} e^{-(2\ln n_i)\Lambda_i} - \frac{n_i}{2\ln n} e^{-(2\ln n)\Lambda_i}\Big|\Big) = \mathcal{O}\Big(\frac{L^2 n}{2^L \ln^3 n}\Big).$$

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

Again we use the bound in (2.19) for the sum of those n_i with $n_i < \frac{n}{2^{kL}}$ (for large enough k). Thus, for $\beta > \frac{1}{\lg 3-1}$ with another application of the Markov inequality,

$$X(T_n) = \sum_{i=1}^{2^L} \frac{n_i}{\ln n_i} - \sum_{i=1}^{2^L} \frac{\Upsilon(T_i)}{4\ln^2 n_i} - \frac{1}{2\ln n} \sum_{i=1}^{2^L} n_i e^{-(2\ln n)\Lambda_i} + \frac{n}{4\ln^2 n} + o_p\left(\frac{n}{\ln^2 n}\right). \tag{2.22}$$

Lemma 2.4. *If the constant β in $L = \lfloor \beta \lg \ln n \rfloor$ is less than 2, then*

$$\sum_{i=1}^{2^L} n_i e^{-(2\ln n)\Lambda_i} = \sum_{h(v)\leqslant L} n_v e^{-(2\ln n)\lambda_v} + o_p\left(\frac{n}{\ln n}\right).$$

Thus, choosing $\beta \in \left(\frac{1}{\lg 3-1}, 2\right)$ from (2.22),

$$X(T_n) = \sum_{i=1}^{2^L} \frac{n_i}{\ln n_i} - \sum_{i=1}^{2^L} \frac{\Upsilon(T_i)}{4\ln^2 n_i} - \frac{1}{2\ln n} \sum_{h(v)\leqslant L} n_v e^{-(2\ln n)\lambda_v} + \frac{n}{4\ln^2 n} + o_p\left(\frac{n}{\ln^2 n}\right). \tag{2.23}$$

(Compare with [12, Lemma 2.4].)

Proof of Lemma 2.4. The proof follows by the same calculations as in [12, Lemma 2.4]. Recall that $P(v_i) = \sigma, \ldots, v_i$, is the path between the root σ of T to v_i, $1 \leqslant i \leqslant 2^L$, where v_i are the vertices at height L. Choosing $a = \frac{2\ln m}{m}$ where $m = 2\ln n$, we see as in [12, p. 248] that the probability for some i that at least two λ_v, $v \in P(v_i)$, are less than a is

$$\mathcal{O}\left(\frac{2^L \ln^4 m}{m^2}\right) = o(1). \tag{2.24}$$

Hence, with probability tending to 1 there is at most one λ_v less than a in each $P(v_i)$. This gives that

$$0 \leqslant \sum_{v\in P(v_i)} e^{-m\lambda_v} - e^{-m\Lambda_i} \leqslant Le^{-ma} = \frac{L}{m^2}. \tag{2.25}$$

From (2.25) it is rather easy to get the result. (For further details, see [12, p. 248].) □

Remark. From (2.24) we see the reason for choosing the constant β in L to be less than 2.

We recall that $\mathscr{G}_j$, defined in Section 2.1, is the σ-field generated by $\{U_{r,v},\ r \in \{1,2,\ldots,j\}\}$ for all vertices v with $h(v) = j$. In particular we get that $\mathscr{G}_L$ is the σ-field generated by $\{U_{r,v},\ r \in \{1,2,\ldots,L\}\}$ for all vertices v with $h(v) = L$. Note in particular that the σ-field generated by $\{n_v,\ h(v) \leqslant L\}$ is a sub-σ-field of $\mathscr{G}_L$.

Conditioned on $\mathscr{G}_L$, the summands $\Upsilon(T_i)$ in (2.23) are independent. Thus, from Section 1.3.2,

$$\mathbf{E}\left(\sum_{i=1}^{2^L} \frac{\Upsilon(T_i)}{4\ln^2 n_i} \,\middle|\, \mathscr{G}_L\right) = \sum_{i=1}^{2^L} \frac{n_i}{2\ln n_i} + \sum_{i=1}^{2^L} \frac{n_i(2\gamma-4)}{4\ln^2 n_i} + \sum_{i=1}^{2^L} \mathcal{O}\left(\frac{1}{\ln n_i}\right),$$

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

and

$$\mathbf{Var}\left(\sum_{i=1}^{2^L}\frac{\Upsilon(T_i)}{4\ln^2 n_i}\,\Big|\,\mathscr{G}_L\right)=\mathcal{O}\left(\sum_{i=1}^{2^L}\frac{n_i^2}{\ln^4 n_i}\right).$$

By using (2.19), in analogy with (2.20),

$$\sum_{i=1}^{2^L}\frac{n_i(2\gamma-4)}{4\ln^2 n_i}=\sum_{i=1}^{2^L}\frac{n_i(2\gamma-4)}{4\ln^2 n}+o\left(\frac{n}{\ln^2 n}\right).$$

Thus, the Markov inequality gives

$$\sum_{i=1}^{2^L}\frac{\Upsilon(T_i)}{4\ln^2 n_i}=\sum_{i=1}^{2^L}\frac{n_i}{2\ln n_i}+\frac{n(2\gamma-4)}{4\ln^2 n}+o_p\left(\frac{n}{\ln^2 n}\right).$$

For simplicity (in the calculations below) we change the notation n_i, $1\leqslant i\leqslant 2^L$, to n_v, $h(v)=L$.

Thus, from (2.23), choosing the constant $\beta\in\left(\frac{1}{\lg 3-1},2\right)$,

$$X(T_n)=\sum_{h(v)=L}\frac{n_v}{2\ln n_v}-\frac{1}{2\ln n}\sum_{h(v)\leqslant L}n_v e^{-(2\ln n)\lambda_v}+\frac{n(5-2\gamma)}{4\ln^2 n}+o_p\left(\frac{n}{\ln^2 n}\right).\tag{2.26}$$

Recall from Section 2.1 that

$$\left\{\xi_v:=\frac{mn_v}{n}e^{-m\lambda_v}\right\}.\tag{2.27}$$

The expression of $X(T_n)$ in (2.26) and the notation of ξ_v in (2.27) give

$$\begin{aligned}&\frac{4\ln^2 n}{n}\left(X(T_n)-\frac{n}{2\ln n}-\frac{n\ln\ln n}{2\ln^2 n}\right)\\&\quad=-\sum_{h(v)\leqslant L}\xi_v+\frac{4\ln^2 n}{n}\sum_{h(v)=L}\frac{n_v}{2\ln n_v}-2\ln\ln n-2\ln n+5-2\gamma+o_p(1).\end{aligned}$$

Let

$$D:=\frac{4\ln^2 n}{n}\sum_{h(v)=L}\frac{n_v}{2\ln n_v}-2\ln\ln n-2\ln n+5-2\gamma\quad\text{and}\quad\xi_i':=\frac{-D}{n}.\tag{2.28}$$

Thus,

$$\frac{4\ln^2 n}{n}\left(X(T_n)-\frac{n}{2\ln n}-\frac{n\ln\ln n}{2\ln^2 n}\right)=-\sum_{h(v)\leqslant L}\xi_v-\sum_{i=1}^{n}\xi_i'+o_p(1).\tag{2.29}$$

Since the n_v's in the sums in (2.26) are not independent (although they are less dependent for vertices that are far from each other), $\{\xi_v\}\bigcup\{\xi_i'\}$ is not a triangular array. However, conditioned on $\mathscr{G}_L$, $\{\xi_v\}\bigcup\{\xi_i'\}$ is a triangular array with ξ_i' conditioned on $\mathscr{G}_L$ deterministic.

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

2.3. Applying a limit theorem for sums of triangular arrays

2.3.1. Theorem 2.5, which proves Theorem 1.1. As in [12], the proof of Theorem 1.1 will be completed by a classical theorem for convergence of sums of triangular arrays to infinitely divisible distributions: see, *e.g.*, [15, Theorem 15.28]. For the sake of independence we intend to condition on the n_v's in (2.29), and show that the normalized $X(T_n)$ converges in distribution to W with an infinitely divisible distribution, which does not depend on the particular n_v's we conditioned on. Then it follows that, also unconditioned, the normalized $X(T_n)$ converges in distribution to W.

The main Theorem 1.1 is proved by Theorem 2.5 below.

Theorem 2.5. *Suppose that $n \to \infty$ and choose any constant $c > 0$ (γ is the Euler constant). Conditioning on the σ-field $\mathscr{G}_L$, where $L = \lfloor \beta \lg \ln n \rfloor$, $\beta \in \left(\frac{1}{\lg 3 - 1}, 2\right)$, the following hold:*

(i) $$\sup_{v,\ h(v) \leqslant L} \mathbf{P}\left(\xi_v > x \,|\, \mathscr{G}_L\right) \to 0 \quad \textit{for every } x > 0,$$

(ii) $$\sum_{h(v) \leqslant L} \mathbf{P}\left(\xi_v > x \,|\, \mathscr{G}_L\right) \xrightarrow{p} v(x, \infty) = \frac{2}{x} \quad \textit{for every } x > 0,$$

(iii) $$\left(\sum_{h(v) \leqslant L} \mathbf{E}\left(\xi_v \mathbf{1}[\xi_v \leqslant c] \,|\, \mathscr{G}_L\right) - \frac{4 \ln^2 n}{n} \sum_{h(v)=L} \frac{n_v}{2 \ln n_v} + 2 \ln \ln n \right.$$
$$\left. + 2 \ln n - 5 + 2\gamma \right) \xrightarrow{p} -3 - 2 \ln 2 + 2\gamma + 2 \ln c,$$

(iv) $$\sum_{h(v) \leqslant L} \mathbf{Var}\left(\xi_v \mathbf{1}[\xi_v \leqslant c] \,|\, \mathscr{G}_L\right) \xrightarrow{p} 2c.$$

(This is the result corresponding to [12, Lemma 2.5] for the complete binary tree.)

Before proving Theorem 2.5 we will show how it proves Theorem 1.1. Recall from (2.28) that

$$D := \frac{4 \ln^2 n}{n} \sum_{h(v)=L} \frac{n_v}{2 \ln n_v} - 2 \ln \ln n - 2 \ln n + 5 - 2\gamma.$$

We apply [15, Theorem 15.28], with $a = 0$ and $b = -3 - 2 \ln 2 + 2\gamma$, to $\sum_{h(v) \leqslant L} \xi_v + \sum_{i=1}^{n} \xi_i'$ conditioned on $\mathscr{G}_L$ with $\xi_i' = \frac{-D}{n}$ deterministic. The constants a and b are the constants that occur in the general formula of the characteristic function for infinitely divisible distributions in (1.3). Note that $\frac{D}{n} \to 0$; thus, because of (i), conditioned on $\mathscr{G}_L$, $\{\xi_v\} \bigcup \{\xi_i'\}$ is a null array. We have $\frac{dv}{dx} = \frac{2}{x^2}$, and hence

$$\int_0^c x^2 dv(x) = \int_0^c 2dx = 2c, \quad \text{and} \quad \int_c^1 x dv(x) = \int_c^1 \frac{2}{x} dx = -2 \ln c.$$

Thus, the right-hand sides of (iii) and (iv) of Theorem 2.5 are $b - \int_c^1 x dv(x)$, respectively $\int_0^c x^2 dv(x)$.

The convergence in Theorem 2.5 is, however, in probability, while [15, Theorem 15.28] requires usual convergence, *i.e.*, standard pointwise convergence of sequences with no

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

probability involved. For the complete binary tree Janson applies [15, Theorem 15.28] by using the result in [12, Lemma 2.5]. In [12, Lemma 2.5], the terms in (i), (ii), (iii) and (iv) are deterministic, due to the fact that the n_v's are not random for the complete binary tree, so in that lemma there is no need for conditioning on $\mathscr{G}_L$. Thus, the convergence in [12, Lemma 2.5] is in the usual sense, which is the type of convergence in [15, Theorem 15.28], rather than the probabilistic sense, which is the case in Theorem 2.5 above. However, if the convergence instead was a.s. in Theorem 2.5, then it would have been easy to see from this theorem that, conditioned on $\mathscr{G}_L$, the conditions of [15, Theorem 15.28] are fulfilled for $\sum_{h(v)\leqslant L}\xi_v+\sum_{i=1}^n\xi_i'$. We write

$$S(n):=\sum_{h(v)\leqslant L}\xi_v+\sum_{i=1}^n\xi_i'.$$

Thus, assuming a.s. convergence in Theorem 2.5, [15, Theorem 15.28] implies that, conditioned on $\mathscr{G}_L$, $S(n)\xrightarrow{d}W$. It follows that conditioning on $\mathscr{G}_L$ has no influence on the distributional convergence of $S(n)$ (unconditioned), since, for any continuous bounded function $g:\mathfrak{R}\to\mathfrak{R}$,

$$\mathbf{E}g(S(n)\,|\,\mathscr{G}_L)=\int g dF(S(n)\,|\,\mathscr{G}_L)\to\mathbf{E}g(W).$$

Thus, taking expectation by dominated convergence,

$$\mathbf{E}g(S(n))\to\mathbf{E}g(W).$$

This shows that, if also unconditioned, $S(n)\xrightarrow{d}W$.

It remains to show that convergence in probability, which is the type of convergence in Theorem 2.5, actually is sufficient for $S(n)\xrightarrow{d}W$ to hold. We do this by showing that every subsequence $\{\zeta_k\}$ of the positive integers contains yet another subsequence $\{\zeta_{k_j}\}$, so that $S(\zeta_{k_j})$ converges in distribution to the same limit W as $j\to\infty$. Then it follows easily (by the well-known fact that a sequence of reals converges to a constant a if and only if every subsequence contains a subsequence that converges to a; see, *e.g.*, [8, Section 5.7]) that also $S(n)\xrightarrow{d}W$.

To show this we use the fact that if a random variable $X_n\xrightarrow{p}X$ as $n\to\infty$, then there is an increasing subsequence of the integers $\{\zeta_k,\ k\geqslant 1\}$ such that $X_{\zeta_k}\xrightarrow{a.s}X$ as $n\to\infty$: see, *e.g.*, [8, Section 5.3]. Using this fact together with the famous diagonal method (see, *e.g.*, [8, Section 5.8]), we can for all subsequences $\{\zeta_k\}$ pick a subsequence $\{\eta_j\}$ such that (ii) of Theorem 2.5 above holds a.s. for all $x\in Q$ (where Q is the set of rational numbers). Since Q is dense in $\mathfrak{R}$, there are for all x arbitrarily close rational numbers r and s, $r<x<s$, so that $\sum_{h(v)\leqslant L}\mathbf{P}(\xi_v>s)\leqslant\sum_{h(v)\leqslant L}\mathbf{P}(\xi_v>x)\leqslant\sum_{h(v)\leqslant L}\mathbf{P}(\xi_v>r)$. Hence, it follows easily that (ii) holds (for the sub-subsequence we picked) for all $x\in\mathfrak{R}$. Now we can pick yet another subsequence from this sub-subsequence such that (iii) holds a.s., and from this yet another subsequence such that (iv) holds a.s. This means that for all subsequences $\{\zeta_k,\ k\geqslant 1\}$, we find a subsequence $\{\zeta_{k_j}\}$ such that $S(\zeta_{k_j})\xrightarrow{d}W$ and thus $S(n)\xrightarrow{d}W$ holds.

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

Alternatively, one can use Skorokhod's coupling theorem (*e.g.*, [15, Theorem 3.30]), *i.e.*, one can assume that (ii), (iii) and (iv) hold a.s., since one can couple the vector of

$$\Big(\sum_{h(v)\leqslant L}\mathbf{P}(\xi_v > x),\ \sum_{h(v)\leqslant L}\mathbf{E}(\xi_v\mathbf{1}[\xi_v\leqslant c]),\ \sum_{h(v)\leqslant L}\mathbf{Var}(\xi_v\mathbf{1}[\xi_v\leqslant c])\Big)$$

with a parallel vector that converges a.s. Thus, the proof of Theorem 1.1 for $X_v(T)$ is completed.

Now it follows easily that the result holds for $X_e(T)$ too. The only difference is that the root is ignored, so that $\Lambda_i \stackrel{d}{=} \mathrm{Exp}\big(\frac{1}{L}\big)$, instead of $\Lambda_i \stackrel{d}{=} \mathrm{Exp}\big(\frac{1}{L+1}\big)$ as for $X_v(T)$. However, it simply follows that this minor difference does not affect the calculations in the proofs. Alternatively, we could see $\widehat{T}$ (the rooted forest with two tree components) as the tree T with the root deleted. Then there is a natural 1–1 correspondence between edges of T and vertices of $\widehat{T}$, and this correspondence also preserves the record (and cutting) operations. Hence, $X_e(T) = X_v(\widehat{T})$. Thus, the proof of Theorem 1.1 is completed.

2.3.2. Proof of Theorem 2.5. Below we present three crucial lemmas by which we can then easily prove Theorem 2.5. First recall that $\mathscr{G}_j$ is the σ-field generated by $\{U_{r,v},\ r\in\{1,2,\dots,j\}\}$ for all vertices v with $h(v)=j$.

Lemma 2.6. *Suppose that $n\to\infty$ and choose any constant $c>0$. Then, for $L=\lfloor\beta\lg\ln n\rfloor$, $\beta\in\big(\frac{1}{\lg 3-1},2\big)$, the following hold:*

$$\mathbf{E}\Big(\sum_{h(v)\leqslant L}\mathbf{P}\big(\xi_v > x\,|\,\mathscr{G}_L\big)\Big) = \frac{2}{x}+o(1) = v(x,\infty)+o(1),\quad \textit{for every } x>0,$$

$$\mathbf{E}\Big(\sum_{h(v)\leqslant L}\mathbf{E}\big(\xi_v\mathbf{1}[\xi_v\leqslant c]\,|\,\mathscr{G}_L\big)\Big) = L+2-2\ln\ln n-2\ln 2+2\ln c+o(1),$$

$$\mathbf{E}\Big(\sum_{h(v)=L}\frac{n_v}{2\ln n_v}\Big) = \frac{n}{2\ln n}+\frac{nL}{4\ln^2 n}+o\Big(\frac{n}{\ln^2 n}\Big)+o(1),$$

$$\mathbf{E}\Big(\sum_{h(v)\leqslant L}\mathbf{Var}\big(\xi_v\mathbf{1}[\xi_v\leqslant c]\,|\,\mathscr{G}_L\big)\Big) = 2c+o(1).$$

Lemma 2.7. *Let $n\to\infty$. Then, for $L:=\lfloor\beta\lg\ln n\rfloor$, $\beta\in\big(\frac{1}{\lg 3-1},2\big)$ and $l:=\lfloor 100\ln\ln\ln n\rfloor$, the following hold:*

$$\mathbf{Var}\Big(\mathbf{E}\Big(\sum_{l\leqslant h(v)\leqslant L}\mathbf{P}(\xi_v > x\,|\,\mathscr{G}_L)\,\Big|\,\mathscr{G}_l\Big)\Big)\to 0, \tag{2.30}$$

$$\mathbf{Var}\Big(\mathbf{E}\Big(\sum_{l\leqslant h(v)\leqslant L}\mathbf{E}\big(\xi_v\mathbf{1}[\xi_v\leqslant c]\,|\,\mathscr{G}_L\big)-\frac{4\ln^2 n}{n}\sum_{h(v)=L}\frac{n_v}{2\ln n_v}\,\Big|\,\mathscr{G}_l\Big)\Big)\to 0, \tag{2.31}$$

$$\mathbf{Var}\Big(\mathbf{E}\Big(\sum_{l\leqslant h(v)\leqslant L}\mathbf{Var}\big(\xi_v\mathbf{1}[\xi_v\leqslant c]\,|\,\mathscr{G}_L\big)\,\Big|\,\mathscr{G}_l\Big)\Big)\to 0. \tag{2.32}$$

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

Lemma 2.8. *Let $n \to \infty$. Then, for $L := \lfloor \beta \lg \ln n \rfloor$, $\beta \in \left(\frac{1}{\lg 3-1}, 2\right)$ and $l := \lfloor 100 \ln \ln \ln n \rfloor$, the following hold:*

$$\mathbf{E}\left(\mathbf{Var}\left(\sum_{l \leqslant h(v) \leqslant L} \mathbf{P}\left(\xi_v > x \,|\, \mathscr{G}_L\right) \,\middle|\, \mathscr{G}_l\right)\right) \to 0, \tag{2.33}$$

$$\mathbf{E}\left(\mathbf{Var}\left(\sum_{l \leqslant h(v) \leqslant L} \mathbf{E}\left(\xi_v \mathbf{1}[\xi_v \leqslant c] \,|\, \mathscr{G}_L\right) - \frac{4 \ln^2 n}{n} \sum_{h(v)=L} \frac{n_v}{2 \ln n_v} \,\middle|\, \mathscr{G}_l\right)\right) \to 0, \tag{2.34}$$

$$\mathbf{E}\left(\mathbf{Var}\left(\sum_{l \leqslant h(v) \leqslant L} \mathbf{Var}\left(\xi_v \mathbf{1}[\xi_v \leqslant c] \,|\, \mathscr{G}_L\right) \,\middle|\, \mathscr{G}_l\right)\right) \to 0. \tag{2.35}$$

Before proving these lemmas (which we do in the final Section 2.3.3), we show how their use leads to the proof of Theorem 2.5.

Proof of Theorem 2.5. Recall that $m = 2 \ln n$, and that we write $\mathscr{G}_L$, where $L := \lfloor \beta \lg \ln n \rfloor$, $\beta \in \left(\frac{1}{\lg 3-1}, 2\right)$, for the σ-field generated by $\{U_{r,v},\ r \in \{1, 2, \ldots, L\}\}$ for all vertices with $h(v) = L$.

For any $x > 0$ and v with $h(v) \leqslant L$, we have

$$\begin{aligned} \mathbf{P}\left(\xi_v > x \,|\, \mathscr{G}_L\right) &= \mathbf{P}\left(e^{-m\lambda_v} > \frac{nx}{mn_v} \,\middle|\, \mathscr{G}_L\right) = \mathbf{P}\left(\lambda_v < \frac{1}{m} \ln \frac{mn_v}{nx} \,\middle|\, \mathscr{G}_L\right) \\ &= 1 - \exp\left(-\frac{1}{m} \ln_+ \frac{mn_v}{nx}\right). \end{aligned} \tag{2.36}$$

Thus, for every $x > 0$,

$$\mathbf{P}\left(\xi_v > x \,|\, \mathscr{G}_L\right) \leqslant \frac{1}{m} \ln_+ \frac{mn_v}{nx} \leqslant \frac{1}{m} \ln_+ \frac{m}{x} \to 0, \tag{2.37}$$

which proves (i).

Lemma 2.6 shows that the expected values of the left-hand sides in (ii), (iii) and (iv) of Theorem 2.5 converge to the right-hand sides.

We complete the proof of Theorem 2.5 by showing

$$\mathbf{Var}\left(\sum_{h(v) \leqslant L} \mathbf{P}\left(\xi_v > x \,|\, \mathscr{G}_L\right)\right) \to 0 \quad \text{for every } x > 0, \tag{2.38}$$

$$\mathbf{Var}\left(\sum_{h(v) \leqslant L} \mathbf{E}(\xi_v \mathbf{1}[\xi_v \leqslant c] \,|\, \mathscr{G}_L) - \frac{4 \ln^2 n}{n} \sum_{h(v)=L} \frac{n_v}{2 \ln n_v}\right) \to 0, \tag{2.39}$$

$$\mathbf{Var} \sum_{h(v) \leqslant L} \mathbf{Var}(\xi_v \mathbf{1}[\xi_v \leqslant c] \,|\, \mathscr{G}_L) \to 0. \tag{2.40}$$

Then, by Chebyshev's inequality, (ii), (iii) and (iv) of Theorem 2.5 follow. Thus, it remains to show how (2.38), (2.39) and (2.40) follow from Lemmas 2.7 and 2.8. Let $l := \lfloor 100 \ln \ln \ln n \rfloor$.

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

We note that by using (2.37) one easily obtains (see [12, equations (19), (21), (22)])

$$\sum_{h(v)\leqslant L} \mathbf{P}\big(\xi_v > x \,|\, \mathscr{G}_L\big) = \sum_{l\leqslant h(v)\leqslant L} \mathbf{P}\big(\xi_v > x \,|\, \mathscr{G}_L\big) + o(1), \tag{2.41}$$

$$\sum_{h(v)\leqslant L} \mathbf{E}(\xi_v \mathbf{1}[\xi_v \leqslant c] \,|\, \mathscr{G}_L) = \sum_{l\leqslant h(v)\leqslant L} \mathbf{E}\big(\xi_v \mathbf{1}[\xi_v \leqslant c] \,|\, \mathscr{G}_L\big) + o(1), \tag{2.42}$$

$$\sum_{h(v)\leqslant L} \mathbf{Var}\big(\xi_v \mathbf{1}[\xi_v \leqslant c] \,|\, \mathscr{G}_L\big) = \sum_{l\leqslant h(v)\leqslant L} \mathbf{Var}\big(\xi_v \mathbf{1}[\xi_v \leqslant c] \,|\, \mathscr{G}_L\big) + o(1). \tag{2.43}$$

(In fact one can choose a larger value for l in (2.41), (2.42) and (2.43), such as $\rho \ln\ln n$ for some constant ρ.)

To show (2.38), (2.39) and (2.40) we use a variance formula that is easy to establish (see, *e.g.*, [8, exercise 10.17-2]):

$$\mathbf{Var}(Y) = \mathbf{E}(\mathbf{Var}(Y \,|\, \mathscr{G})) + \mathbf{Var}(\mathbf{E}(Y \,|\, \mathscr{G})), \tag{2.44}$$

where Y is a random variable and $\mathscr{G}$ is a sub-σ-field. Let

$$S_1 := \sum_{l\leqslant h(v)\leqslant L} \mathbf{P}\big(\xi_v > x \,|\, \mathscr{G}_L\big),$$

$$S_2 := \sum_{l\leqslant h(v)\leqslant L} \mathbf{E}\big(\xi_v \mathbf{1}[\xi_v \leqslant c] \,|\, \mathscr{G}_L\big) - \frac{4\ln^2 n}{n} \sum_{h(v)=L} \frac{n_v}{2\ln n_v},$$

$$S_3 := \sum_{l\leqslant h(v)\leqslant L} \mathbf{Var}\big(\xi_v \mathbf{1}[\xi_v \leqslant c] \,|\, \mathscr{G}_L\big).$$

Recall that for $l := \lfloor 100 \ln\ln\ln n \rfloor$, $\mathscr{G}_l$ is the σ-field generated by $\{U_{r,v},\ r \in \{1,2,\dots,l\}\}$ for all vertices with $h(v) = l$. Consequently, from the variance formula in (2.44) we get

$$\mathbf{Var}(S_1) = \mathbf{E}(\mathbf{Var}(S_1 \,|\, \mathscr{G}_l)) + \mathbf{Var}(\mathbf{E}(S_1 \,|\, \mathscr{G}_l)), \tag{2.45}$$

$$\mathbf{Var}(S_2) = \mathbf{E}\big(\mathbf{Var}(S_2 \,|\, \mathscr{G}_l)\big) + \mathbf{Var}\big(\mathbf{E}(S_2 \,|\, \mathscr{G}_l)\big), \tag{2.46}$$

$$\mathbf{Var}(S_3) = \mathbf{E}\big(\mathbf{Var}(S_3 \,|\, \mathscr{G}_l)\big) + \mathbf{Var}(\mathbf{E}(S_3 \,|\, \mathscr{G}_l)). \tag{2.47}$$

Thus, by using (2.41), (2.42) and (2.43) we get that (2.38), (2.39) and (2.40) follow from Lemmas 2.7 and 2.8. □

We have proved Theorem 2.5, and thus also Theorem 1.1. □

2.3.3. Proofs of lemmas used in the proof of Theorem 2.5. Finally we present the proofs of Lemmas 2.6, 2.7 and 2.8.

Proof of Lemma 2.6. Recall from Section 2.1 that $Y_{k,v} := -\sum_{r=1}^{k} \ln U_{r,v}$, where $U_{r,v}$ are the uniform random variables defined in Section 1.3.3. Below we skip the vertex index and just write Y_k, since the $Y_{k,v}$'s are identically distributed for vertices v on the same level k. Noting that

$$n_v \leqslant n \prod_{r=1}^{k} U_{r,v} \leqslant n_v + k \tag{2.48}$$

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

motivates the following notation of

$$\widehat{n_v} := n\prod_{r=1}^{k} U_{r,v}. \tag{2.49}$$

We note that the σ-field $\mathscr{G}_j$ defined in Section 2.1 can equivalently be defined as the σ-field generated by $\{\widehat{n_v} : h(v) \leqslant j\}$.

Hence, by using (2.48), for large n,

$$\begin{aligned}\sum_{h(v)\leqslant L} \mathbf{P}\big(\xi_v > x \,|\, \mathscr{G}_L\big) &= \sum_{k=0}^{L}\sum_{h(v)=k} \frac{1}{m}\ln_+ \frac{mn_v}{nx}\left(1+\mathcal{O}\left(\frac{\ln m}{m}\right)\right)\\ &= \sum_{k=1}^{L}\sum_{h(v)=k} \frac{1}{m}\ln_+ \frac{m\widehat{n_v}}{nx}\left(1+\mathcal{O}\left(\frac{\ln m}{m}\right)\right) + o(1).\end{aligned} \tag{2.50}$$

Thus, the expected value of $\sum_{h(v)\leqslant L} \mathbf{P}\big(\xi_v > x \,|\, \mathscr{G}_L\big)$ in (2.50) is

$$\begin{aligned}&(1+o(1))\sum_{k=1}^{L} 2^k \mathbf{E}\frac{1}{m}\ln_+ \frac{m\prod_{i=1}^{k} U_i}{x} + o(1).\\ &\quad = (1+o(1))\sum_{k=1}^{L} 2^k \mathbf{E}\left(\frac{(\ln m - \ln x - Y_k)}{m} I\{Y_k \leqslant \ln m - \ln x\}\right) + o(1).\end{aligned} \tag{2.51}$$

Recall from Section 2.1 that $Y_k \overset{d}{=} \Gamma(k,1)$; hence this is equal to

$$\begin{aligned}&(1+o(1))\sum_{k=1}^{L}\int_0^{\ln m-\ln x} 2^k \frac{(\ln m - \ln x - y)}{m}\frac{e^{-y}y^{k-1}}{(k-1)!}dy + o(1)\\ &\quad = (1+o(1))\int_0^{\ln m-\ln x}\sum_{k=1}^{L} 2^k \frac{(\ln m - \ln x - y)}{m}\frac{e^{-y}y^{k-1}}{(k-1)!}dy + o(1).\end{aligned} \tag{2.52}$$

Recall that $m = 2\ln n$ and $L := \lfloor \beta \lg\ln n\rfloor$ for $\beta \in \left(\frac{1}{\lg 3-1}, 2\right)$. For $y \leqslant \ln m$, we use the ratio test on the sum

$$\sum_{k=L+1}^{\infty} \frac{2^k y^{k-1}}{(k-1)!}. \tag{2.53}$$

Note that if n is large then $k > 2\ln m > 2y$ for all $k > L$, and thus the sum in (2.53) is $\mathcal{O}(1)$. Hence,

$$\int_0^{\ln m-\ln x}\sum_{k=L+1}^{\infty} 2^k \frac{(\ln m - \ln x - y)}{m}\frac{e^{-y}y^{k-1}}{(k-1)!}dy = o(1). \tag{2.54}$$

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

Taylor expansion of e^x implies that the right-hand side in (2.52) is given by

$$\begin{aligned}
&\int_0^{\ln m-\ln x} \frac{2e^y(\ln m-\ln x-y)}{m}dy+o(1)\\
&\quad=\int_0^{\ln m-\ln x} u\frac{2e^{\ln m-\ln x-u}}{m}du+o(1)=\frac{2}{x}\int_0^{\ln m-\ln x} e^{-u}du+o(1)\\
&\quad=\frac{2}{x}+o(1)=v(x,\infty)+o(1). \qquad (2.55)
\end{aligned}$$

Note that $v(x,\infty)$ is continuous and decreasing with $v(x,\infty)\to 0$ as $x\to\infty$. The derivative is

$$\frac{dv(x,\infty)}{dx}=\frac{-2}{x^2}.$$

For (iii) of Theorem 2.5 it is easy to calculate (compare with [12, p. 251]) that

$$\begin{aligned}
\sum_{h(v)\leqslant L}\mathbf{E}\big(\xi_v\mathbf{1}[\xi_v\leqslant c]\,|\,\mathscr{G}_L\big)&=\sum_{h(v)\leqslant L}\frac{mn_v}{n}\mathbf{E}\left(e^{-m\lambda_v}\mathbf{1}\left[e^{-m\lambda_v}\leqslant\frac{nc}{mn_v}\right]\,\Big|\,\mathscr{G}_L\right)\\
&=\sum_{h(v)\leqslant L}\frac{mn_v}{n(m+1)}e^{-\frac{m+1}{m}\ln_+(\frac{mn_v}{nc})}. \qquad (2.56)
\end{aligned}$$

By using (2.48) we obtain that this is equal to

$$\sum_{h(v)\leqslant L}\frac{m\widehat{n_v}}{n(m+1)}e^{-\frac{m+1}{m}\ln_+(\frac{m\widehat{n_v}}{nc})}+o(1). \qquad (2.57)$$

Thus, the expected value of $\sum_{h(v)\leqslant L}\mathbf{E}\big(\xi_v\mathbf{1}[\xi_v\leqslant c]\,|\,\mathscr{G}_L\big)$ in (2.56) is

$$\begin{aligned}
&\mathbf{E}\sum_{h(v)\leqslant L}\frac{m\prod_{r=1}^k U_{r,v}}{m+1}e^{-\frac{m+1}{m}\ln(\frac{m\prod_{r=1}^k U_{r,v}}{c})I\{\prod_{r=1}^k U_{r,v}\geqslant \ln c-\ln m\}}+o(1)\\
&\quad=\mathbf{E}\sum_{h(v)\leqslant L}\frac{m}{m+1}e^{-Y_k}e^{-\frac{m+1}{m}(\ln m-\ln c-Y_k)I\{Y_k\leqslant \ln m-\ln c\}}+o(1)\\
&\quad=\mathbf{E}\sum_{h(v)\leqslant L}\frac{m}{m+1}e^{-Y_k}e^{-\frac{m+1}{m}(\ln m-\ln c-Y_k)}I\{Y_k\leqslant\ln m-\ln c\}\\
&\qquad+\mathbf{E}\sum_{h(v)\leqslant L}\frac{m}{m+1}e^{-Y_k}I\{Y_k>\ln m-\ln c\}+o(1). \qquad (2.58)
\end{aligned}$$

Since $Y_k\overset{d}{=}\Gamma(k,1)$, we get that these equalities are equal to $E_1+E_2+o(1)$, where

$$\begin{aligned}
E_1&:=\sum_{k=1}^L 2^k\int_0^{\ln m-\ln c}\frac{m}{m+1}e^{-y}e^{-\frac{m+1}{m}(\ln m-\ln c-y)}\frac{e^{-y}y^{k-1}}{(k-1)!},\\
E_2&:=\mathbf{E}\sum_{k=1}^L 2^k\frac{m}{m+1}\prod_{r=1}^k U_{r,v}(1-I\{Y_k\leqslant\ln m-\ln c\}). \qquad (2.59)
\end{aligned}$$

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

We again use the ratio test on the sum in (2.53). By an argument similar to that used to show that (2.54) gives (2.55), from Taylor expansion of e^x, we obtain

$$\begin{aligned} E_1 &= \int_0^{\ln m-\ln c} \frac{m}{m+1} \sum_{k=1}^{L} 2^k e^{-2y} e^{-\frac{m+1}{m}(\ln m-\ln c-y)} \frac{y^{k-1}}{(k-1)!} dy + o(1) \\ &= \int_0^{\ln m-\ln c} \frac{2m}{m+1} e^{-\frac{m+1}{m}(\ln m-\ln c-y)} dy + o(1) \\ &= \left| \frac{2m^2}{(m+1)^2} e^{-\frac{m+1}{m}(\ln m-\ln c-y)} \right|_0^{\ln m-\ln c} + o(1) = 2 + o(1), \end{aligned} \tag{2.60}$$

and

$$\begin{aligned} E_2 &= \frac{m}{m+1} L - \int_0^{\ln m-\ln c} \frac{m}{m+1} \sum_{k=1}^{L} 2^k e^{-2y} \frac{y^{k-1}}{(k-1)!} dy + o(1) \\ &= L - \left| \frac{2my}{m+1} \right|_0^{\ln m-\ln c} + o(1) \\ &= L - 2\ln m + 2\ln c + o(1). \end{aligned} \tag{2.61}$$

Thus, the equalities in (2.58) are equal to

$$L + 2 - 2\ln m + 2\ln c + o(1) = L + 2 - 2\ln\ln n - 2\ln 2 + 2\ln c + o(1).$$

Using the bound in (2.19) for the sum of the subtree sizes with $n_v < \frac{n}{2^{kL}}$ (for k large enough), we get the expansion

$$\begin{aligned} \sum_{h(v)=L} \frac{n_v}{2\ln n_v} &= \sum_{h(v)=L} \frac{n_v}{2\ln n\left(1 + \frac{\ln(\frac{n_v}{n})}{\ln n}\right)} \\ &= \sum_{h(v)=L} \frac{n_v}{2\ln n} - \sum_{h(v)=L} \frac{n_v \ln(\frac{n_v}{n})}{2\ln^2 n} + o\left(\frac{n}{\ln^2 n}\right) \\ &= \frac{n}{2\ln n} - \sum_{h(v)=L} \frac{n_v \ln(\frac{n_v}{n})}{2\ln^2 n} + o\left(\frac{n}{\ln^2 n}\right). \end{aligned} \tag{2.62}$$

Using (2.48) again gives

$$\begin{aligned} \sum_{h(v)=L} \frac{n_v}{2\ln n_v} &= \frac{n}{2\ln n} - \sum_{h(v)=L} \frac{\widehat{n_v} \ln(\frac{\widehat{n_v}}{n})}{2\ln^2 n} + o\left(\frac{n}{\ln^2 n}\right) \\ &= \frac{n}{2\ln n} - \sum_{h(v)=L} \frac{n \prod_{r=1}^{L} U_{r,v} \sum_{r=1}^{L} \ln U_{r,v}}{2\ln^2 n} + o\left(\frac{n}{\ln^2 n}\right). \end{aligned} \tag{2.63}$$

Hence, the expected value of the sum in (2.63) is

$$\mathbf{E}\left(\sum_{h(v)=L} \frac{n_v}{2\ln n_v}\right) = \frac{n}{2\ln n} + \frac{nL}{4\ln^2 n} + o\left(\frac{n}{\ln^2 n}\right). \tag{2.64}$$

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

For (iv) of Theorem 2.5, by using the calculations in [12, pp. 251–252], we get

$$\sum_{h(v)\leqslant L}\mathbf{Var}\big(\xi_v\mathbf{1}[\xi_v\leqslant c]\,|\,\mathscr{G}_L\big)=\sum_{h(v)\leqslant L}\frac{m^2{n_v}^2}{n^2}\mathbf{Var}\left(e^{-m\lambda_v}\mathbf{1}\left[e^{-m\lambda_v}\leqslant\frac{nc}{mn_v}\right]\,\Big|\,\mathscr{G}_L\right)$$
$$=\sum_{h(v)\leqslant L}\frac{m^2{n_v}^2}{2mn^2}e^{-\frac{2m+1}{m}\ln_+(\frac{mn_v}{nc})}+o(1). \tag{2.65}$$

Using (2.48), we obtain that this is equal to

$$\sum_{h(v)\leqslant L}\frac{m^2\widehat{n}_v^{\,2}}{2mn^2}e^{-\frac{2m+1}{m}\ln_+(\frac{m\widehat{n}_v}{nc})}+o(1).$$

Since $Y_k\stackrel{d}{=}\Gamma(k,1)$, the expected value of $\sum_{h(v)\leqslant L}\mathbf{Var}\big(\xi_v\mathbf{1}[\xi_v\leqslant c]\,|\,\mathscr{G}_L\big)$ in (2.65) is

$$\sum_{k=1}^{L}\frac{2^k m}{2}\mathbf{E}\left(\prod_{r=1}^{k}U_{r,v}^2e^{-\frac{2m+1}{m}(\ln m-\ln c-Y_k)I\{Y_k\leqslant\ln m-\ln c\}}\right)+o(1)=V_1+V_2+o(1), \tag{2.66}$$

where

$$V_1:=\sum_{k=1}^{L}\frac{2^k m}{2}\int_0^{\ln m-\ln c}e^{-2y}e^{-\frac{2m+1}{m}(\ln m-\ln c-y)}\frac{e^{-y}y^{k-1}}{(k-1)!}dy,$$
$$V_2:=\mathbf{E}\left(\sum_{k=1}^{L}\frac{2^k m}{2}\prod_{r=1}^{k}U_{r,v}^2\big(1-I\{Y_k\leqslant\ln m-\ln c\}\big)\right). \tag{2.67}$$

Again we use the ratio test on the sum in (2.53) and Taylor expansion of e^x to get

$$V_1=\int_0^{\ln m-\ln c}me^{-\frac{2m+1}{m}(\ln m-\ln c)}e^{\frac{m+1}{m}y}dy+o(1)=\frac{m^2}{m+1}e^{(\ln c-\ln m)}+o(1)$$
$$=c+o(1), \tag{2.68}$$

and

$$V_2=\sum_{k=1}^{L}\frac{2^{k-1}m}{3^k}-\int_0^{\ln m-\ln c}\sum_{k=1}^{L}e^{-2y}\frac{e^{-y}2^{k-1}my^{k-1}}{(k-1)!}dy+o(1)$$
$$=m-\int_0^{\ln m-\ln c}me^{-y}dy+o(1)=c+o(1). \tag{2.69}$$

□

Proof of Lemma 2.7. All the conditional expected values below are, as always, defined a.s. even if we neglect to state this explicitly.

For a given vertex $v_i\in T$ with $h(v_i)=l$, there are at most 2^{j-l} choices of v at height j with ancestor v_i. Recall the notation of $\widehat{n}_v:=n\prod_{r=1}^{j}U_{r,v}$ in (2.49). From the notation of $Y_{k,v}$ in Section 2.1 we get $Y_{l,v_i}:=-\sum_{r=1}^{l}\ln U_{r,v_i}$. For all v with $h(v)=j$, we also write

$$Z_{j-l,v}:=Y_{j,v}-Y_{l,v_i}=-\sum_{r=l+1}^{j}\ln U_{r,v}. \tag{2.70}$$

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

As for $Y_{j,v}$, note that $Z_{j-l,v}$ are identically distributed for all vertices v at depth j. Since $Y_k \overset{d}{=} \Gamma(k,1)$, we get by calculations similar to those in (2.50)–(2.55) that

$$\begin{aligned}
&\mathbf{E}\Big(\sum_{l\leqslant h(v)\leqslant L}\mathbf{P}\big(\xi_v > x\,|\,\mathscr{G}_L\big)\,\Big|\,\mathscr{G}_l\Big)\\
&\quad=(1+o(1))\sum_{i=1}^{2^l}\int_0^{\ln m-\ln x-Y_{l,v_i}}\sum_{j=l+1}^{L}2^{j-l}\frac{(\ln m-\ln x-Y_{l,v_i}-z)}{m}\frac{e^{-z}z^{j-l-1}}{(j-l-1)!}dz+o(1)\\
&\quad=\sum_{i=1}^{2^l}\int_0^{\ln m-\ln x-Y_{l,v_i}}\frac{2e^z(\ln m-\ln x-Y_{l,v_i}-z)}{m}dz+o(1)\\
&\quad=\sum_{i=1}^{2^l}\int_0^{\ln m-\ln x-Y_{l,v_i}}\frac{2e^z}{m}dz+o(1)=\sum_{i=1}^{2^l}\prod_{r=1}^{l}U_{r,v_i}\frac{2}{x}+o(1)\\
&\quad=\frac{2}{x}+o(1).
\end{aligned}\tag{2.71}$$

Thus,

$$\mathbf{Var}\Big(\mathbf{E}\Big(\sum_{l\leqslant h(v)\leqslant L}\mathbf{P}\big(\xi_v > x\,|\,\mathscr{G}_L\big)\,\Big|\,\mathscr{G}_l\Big)\Big)\to 0,$$

which shows (2.30).

We show that (2.31) is true by calculations similar to those for showing (2.30). First, as before let v_i with $h(v_i)=l$ be a given vertex such that there are at most 2^{j-l} choices of v at height j with ancestor v_i. Recall the notation of $Z_{j-l,v}$ in (2.70), *i.e.*, $Y_{j,v}=Y_{l,v_i}+Z_{j-l,v}$.

By calculations similar to those in (2.71) (glancing at the calculations in (2.56)–(2.59)), we obtain

$$\begin{aligned}
&\mathbf{E}\Big(\sum_{l\leqslant h(v)\leqslant L}\mathbf{E}\big(\xi_v\mathbf{1}[\xi_v\leqslant c]\,|\,\mathscr{G}_L\big)\,\Big|\,\mathscr{G}_l\Big)\\
&\quad=\mathbf{E}\Big(\sum_{l\leqslant h(v)\leqslant L}\frac{m}{m+1}e^{-Y_{l,v_i}-Z_{j-l,v}}e^{-\frac{m+1}{m}(\ln m-\ln c-Y_{l,v_i}-Z_{j-l,v})}\\
&\qquad I\{Y_{l,v_i}+Z_{j-l,v}\leqslant \ln m-\ln c\}\,\Big|\,\mathscr{G}_l\Big)\\
&\qquad+\mathbf{E}\Big(\sum_{l\leqslant h(v)\leqslant L}\frac{m}{m+1}e^{-Y_{l,v_i}-Z_{j-l,v}}I\{Y_{l,v_i}+Z_{j-l,v}>\ln m-\ln c\}\,\Big|\,\mathscr{G}_l\Big)+o(1)\\
&\quad=F_1+F_2+o(1),
\end{aligned}$$

where

$$F_1:=\sum_{i=1}^{2^l}\sum_{j=l+1}^{L}2^{j-l}\int_0^{\ln m-\ln c-Y_{l,v_i}}\frac{m}{m+1}e^{-Y_{l,v_i}-z}e^{-\frac{m+1}{m}(\ln m-\ln c-Y_{l,v_i}-z)}\frac{e^{-z}z^{j-l-1}}{(j-l-1)!}dz,$$

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

and

$$F_2 := \sum_{i=1}^{2^l} \mathbf{E}\Bigg(\sum_{j=l+1}^{L} 2^{j-l} \frac{m}{m+1} \prod_{r=1}^{l} U_{r,v_i} \prod_{r=l+1}^{j} U_{r,v} \big(1 - I\{Y_{l,v_i} + Z_{j-l,v} \leqslant \ln m - \ln c\}\big) \,\Big|\, \mathscr{G}_l \Bigg).$$

Now, by using calculations similar to those in (2.60),

$$\begin{aligned} F_1 &= \sum_{i=1}^{2^l} \int_0^{\ln m - \ln c - Y_{l,v_i}} \frac{m}{m+1} \sum_{j=l+1}^{L} 2^{j-l} e^{-Y_{l,v_i} - 2z} e^{-\frac{m+1}{m}(\ln m - \ln c - Y_{l,v_i} - z)} \frac{z^{j-l-1}}{(j-l-1)!} dz \\ &= \sum_{i=1}^{2^l} \int_0^{\ln m - \ln c - Y_{l,v_i}} e^{-Y_{l,v_i}} \frac{2m}{m+1} e^{-\frac{m+1}{m}(\ln m - \ln c - Y_{l,v_i} - z)} dz + o(1) \\ &= \sum_{i=1}^{2^l} 2 \prod_{r=1}^{l} U_{r,v_i} + o(1) = 2 + o(1). \end{aligned}$$

By calculations similar to those in (2.59) and (2.61), we obtain

$$\begin{aligned} F_2 &= \sum_{i=1}^{2^l} \Bigg(\prod_{r=1}^{l} U_{r,v_i} (L-l) - \int_0^{\ln m - \ln c - Y_{l,v_i}} \Bigg(\frac{m}{m+1} \sum_{j=l+1}^{L} 2^{j-l} \\ &\qquad \cdot e^{-Y_{l,v_i} - 2z} \frac{z^{j-l-1}}{(j-l-1)!} \Bigg) dz \Bigg) + o(1) \\ &= \sum_{i=1}^{2^l} \prod_{r=1}^{l} U_{r,v_i} \left(L - l - \left| \frac{2mz}{m+1} \right|_0^{\ln m - \ln c - Y_{l,v_i}} \right) + o(1) \\ &= \sum_{i=1}^{2^l} \prod_{r=1}^{l} U_{r,v_i} \left(L - l - 2\ln m + 2\ln c - 2\sum_{r=1}^{l} \ln U_{r,v_i} \right) + o(1) \\ &= L - l - 2\ln m + 2\ln c - \sum_{i=1}^{2^l} 2 \prod_{r=1}^{l} U_{r,v_i} \sum_{r=1}^{l} \ln U_{r,v_i} + o(1). \end{aligned}$$

Thus,

$$\begin{aligned} &\mathbf{E}\Bigg(\sum_{l \leqslant h(v) \leqslant L} \mathbf{E}\big(\xi_v \mathbf{1}[\xi_v \leqslant c] \,|\, \mathscr{G}_L\big) \,\Big|\, \mathscr{G}_l \Bigg) \\ &\qquad = 2 + L - l - 2\ln m + 2\ln c - \sum_{i=1}^{2^l} 2 \prod_{r=1}^{l} U_{r,v_i} \sum_{r=1}^{l} \ln U_{r,v_i} + o(1). \end{aligned} \tag{2.72}$$

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

We now use the expansion of $\sum_{h(v)=L} \frac{n_v}{2\ln n_v}$ in (2.62) and (2.63). Let v_i be a vertex at height l and let v be a vertex at height L. Thus, using (2.48),

$$\begin{aligned}\mathbf{E}\Big(\sum_{h(v)=L}\frac{n_v}{2\ln n_v}\,\Big|\,\mathscr{G}_l\Big) &= \mathbf{E}\Big(\sum_{h(v)=L}\frac{\widehat{n_v}}{2\ln \widehat{n_v}}\,\Big|\,\mathscr{G}_l\Big)+o\Big(\frac{n}{\ln^2 n}\Big)\\ &= \mathbf{E}\Big(\frac{n}{2\ln n}-\sum_{i=1}^{2^l}\prod_{r=1}^{l}U_{r,v_i}\Big(2^{L-l}\Big(\frac{n\prod_{r=l+1}^{L}U_{r,v}\sum_{r=l+1}^{L}\ln U_{r,v}}{2\ln^2 n}\\ &\quad+\frac{n\prod_{r=l+1}^{L}U_{r,v}\sum_{r=1}^{l}\ln U_{r,v_i}}{2\ln^2 n}\Big)\Big)\,\Big|\,\mathscr{G}_l\Big)+o\Big(\frac{n}{\ln^2 n}\Big).\end{aligned}\tag{2.73}$$

Thus, the conditional expected value in (2.73) is (as in (2.64)):

$$\begin{aligned}&\mathbf{E}\Big(\sum_{h(v)=L}\frac{n_v}{2\ln n_v}\,\Big|\,\mathscr{G}_l\Big)\\ &\quad=\frac{n}{2\ln n}+\sum_{i=1}^{2^l}\prod_{r=1}^{l}U_{r,v_i}\frac{n(L-l)}{4\ln^2 n}-\sum_{i=1}^{2^l}\frac{n\prod_{r=1}^{l}U_{r,v_i}\sum_{r=1}^{l}\ln U_{r,v_i}}{2\ln^2 n}+o\Big(\frac{n}{\ln^2 n}\Big)\\ &\quad=\frac{n}{2\ln n}+\frac{n(L-l)}{4\ln^2 n}-\sum_{i=1}^{2^l}\frac{n\prod_{r=1}^{l}U_{r,v_i}\sum_{r=1}^{l}\ln U_{r,v_i}}{2\ln^2 n}+o\Big(\frac{n}{\ln^2 n}\Big).\end{aligned}\tag{2.74}$$

From (2.72) and (2.74) we obtain

$$\mathbf{Var}\Big(\mathbf{E}\Big(\sum_{l\leqslant h(v)\leqslant L}\mathbf{E}\big(\xi_v\mathbf{1}[\xi_v\leqslant c]\,|\,\mathscr{G}_L\big)-\frac{4\ln^2 n}{n}\sum_{h(v)=L}\frac{n_v}{2\ln n_v}\,\Big|\,\mathscr{G}_l\Big)\Big)\to 0,$$

which shows (2.31).

For (2.32) we proceed with the same method as for showing (2.30) and (2.31). By calculations similar to those in (2.66) and (2.67), we get

$$\begin{aligned}&\mathbf{E}\Big(\sum_{l\leqslant h(v)\leqslant L}\mathbf{Var}\big(\xi_v\mathbf{1}[\xi_v\leqslant c]\,|\,\mathscr{G}_L\big)\,\Big|\,\mathscr{G}_l\Big)\\ &\quad=\sum_{i=1}^{2^l}\sum_{j=l+1}^{L}\frac{2^{j-l}m}{2}\mathbf{E}\Big(\prod_{r=1}^{l}U_{r,v_i}^2\prod_{r=l+1}^{j}U_{r,v}^2\\ &\qquad\cdot e^{-\frac{2m+1}{m}(\ln m-\ln c-Y_{l,v_i}-Z_{j-l,v})I\{Y_{l,v_i}+Z_{j-l,v}\leqslant \ln m-\ln c\}}\,\Big|\,\mathscr{G}_l\Big)+o(1)\\ &\quad=I_1+I_2+o(1),\end{aligned}$$

where

$$I_1:=\sum_{i=1}^{2^l}\sum_{j=l+1}^{L}\frac{2^{j-l}m}{2}\int_0^{\ln m-\ln c-Y_{l,v_i}}\Big(e^{-2(Y_{l,v_i}+z)}e^{-\frac{2m+1}{m}(\ln m-\ln c-Y_{l,v_i}-z)}\frac{e^{-z}z^{j-l-1}}{(j-l-1)!}\Big)dz,$$

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

and

$$I_2 := \sum_{i=1}^{2^l} \mathbf{E}\Bigg(\sum_{j=l+1}^{L} \frac{2^{j-l}m}{2} \prod_{r=1}^{l} U_{r,v_i}^2 \prod_{r=l+1}^{j} U_{r,v}^2 \big(1 - I\{Y_{l,v_i} + Z_{j-l,v} \leqslant \ln m - \ln c\}\big) \,\Big|\, \mathscr{G}_l \Bigg).$$

We calculate (as in (2.68) and (2.69))

$$\begin{aligned}
I_1 &= \sum_{i=1}^{2^l} \int_0^{\ln m - \ln c - Y_{l,v_i}} m e^{-Y_{l,v_i}} e^{-\frac{2m+1}{m}(\ln m - \ln c)} e^{\frac{m+1}{m}(Y_{l,v_i}+z)} dz + o(1) \\
&= \sum_{i=1}^{2^l} \left| e^{-Y_{l,v_i}} \frac{m^2}{m+1} e^{-\frac{2m+1}{m}(\ln m - \ln c)} e^{\frac{m+1}{m}(Y_{l,v_i}+z)} \right|_0^{\ln m - \ln c - Y_{l,v_i}} + o(1) \\
&= \sum_{i=1}^{2^l} \prod_{r=1}^{l} U_{r,v_i} \frac{m^2}{m+1} e^{-(\ln m - \ln c)} + o(1) = \sum_{i=1}^{2^l} \prod_{r=1}^{l} U_{r,v_i} \frac{m^2 c}{m(m+1)} + o(1) \\
&= \sum_{i=1}^{2^l} \prod_{r=1}^{l} U_{r,v_i} c + o(1) = c + o(1),
\end{aligned}$$

and

$$\begin{aligned}
I_2 &= \sum_{i=1}^{2^l} \prod_{r=1}^{l} U_{r,v_i}^2 \sum_{j=l+1}^{L} \frac{2^{j-l}m}{2 \cdot 3^{j-l}} \\
&\quad - \sum_{i=1}^{2^l} \int_0^{\ln m - \ln c - Y_{l,v_i}} \sum_{j=l+1}^{L} 2^{j-l} m e^{-2(Y_{l,v_i}+z)} \frac{e^{-z} z^{j-l-1}}{(j-l-1)!} dz + o(1) \\
&= \sum_{i=1}^{2^l} \prod_{r=1}^{l} U_{r,v_i}^2 m - \sum_{i=1}^{2^l} \int_0^{\ln m - \ln c - Y_{l,v_i}} e^{-Y_{l,v_i}} m e^{-Y_{l,v_i} - z} dz + o(1) \\
&= \sum_{i=1}^{2^l} \prod_{r=1}^{l} U_{r,v_i} c + o(1) = c + o(1).
\end{aligned}$$

Thus,

$$\mathbf{Var}\Bigg(\mathbf{E}\Bigg(\sum_{l \leqslant h(v) \leqslant L} \mathbf{Var}\big(\xi_v \mathbf{1}[\xi_v \leqslant c] \,|\, \mathscr{G}_L\big) \,\Big|\, \mathscr{G}_l \Bigg) \Bigg) \to 0,$$

which shows (2.32). □

Proof of Lemma 2.8. To show (2.33), we first note that

$$\mathbf{Var}\Bigg(\sum_{l \leqslant h(v) \leqslant L} \mathbf{P}\big(\xi_v > x \,|\, \mathscr{G}_L\big) \,\Big|\, \mathscr{G}_l \Bigg) = \sum_{\substack{l \leqslant h(v) \leqslant L, \\ l \leqslant h(w) \leqslant L}} \mathbf{Cov}\Big(\mathbf{P}\big(\xi_v > x \,|\, \mathscr{G}_L\big), \mathbf{P}\big(\xi_w > x \,|\, \mathscr{G}_L\big) \,|\, \mathscr{G}_l \Big).$$

To calculate these conditional covariances we can suppose that the closest ancestor u for v at $h(v) \leqslant L$ and w at $h(w) \leqslant L$ is at height $h \geqslant l$, since the other terms are just 0 because of independence.

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

For $h \geqslant l$ we use

$$\mathbf{Cov}\Big(\mathbf{P}\big(\xi_v > x \,|\, \mathscr{G}_L\big), \mathbf{P}\big(\xi_w > x \,|\, \mathscr{G}_L\big) \,|\, \mathscr{G}_l\Big) \leqslant \mathbf{E}\Big(\mathbf{P}\big(\xi_v > x \,|\, \mathscr{G}_L\big)\mathbf{P}\big(\xi_w > x \,|\, \mathscr{G}_L\big) \,|\, \mathscr{G}_l\Big), \quad (2.75)$$

which implies that

$$\mathbf{E}\Big(\mathbf{Cov}\Big(\mathbf{P}\big(\xi_v > x \,|\, \mathscr{G}_L\big), \mathbf{P}\big(\xi_w > x \,|\, \mathscr{G}_L\big) \,|\, \mathscr{G}_l\Big)\Big) \leqslant \mathbf{E}\Big(\mathbf{P}\big(\xi_v > x \,|\, \mathscr{G}_L\big)\mathbf{P}\big(\xi_w > x \,|\, \mathscr{G}_L\big)\Big). \quad (2.76)$$

Let (v_u, w_u) denote a general pair of vertices with closest ancestor u. Then, by using (2.76), we get

$$\begin{aligned}\mathbf{E}\Big(\mathbf{Var}\Big(\sum_{l \leqslant h(v) \leqslant L} \mathbf{P}\big(\xi_v > x \,|\, \mathscr{G}_L\big) \,\Big|\, \mathscr{G}_l\Big)\Big) \\ \leqslant \sum_{h=l}^{L} \sum_{u:h(u)=h} \sum_{h\,(v_u,w_u)} \mathbf{E}\Big(\mathbf{P}\big(\xi_{v_u} > x \,|\, \mathscr{G}_L\big) \cdot \mathbf{P}\big(\xi_{w_u} > x \,|\, \mathscr{G}_L\big)\Big). \quad (2.77)\end{aligned}$$

Recall that $\mathscr{G}_{h+1}$ is the σ-field generated by $\{\widehat{n}_v : h(v) \leqslant h+1\}$. For the pair (v_u, w_u) with $h(u) = h$, we get that $\mathbf{E}\big(\mathbf{P}\big(\xi_{v_u} > x \,|\, \mathscr{G}_L\big) \,|\, \mathscr{G}_{h+1}\big)$ and $\mathbf{E}\big(\mathbf{P}\big(\xi_{w_u} > x \,|\, \mathscr{G}_L\big) \,|\, \mathscr{G}_{h+1}\big)$ are independent. Thus,

$$\begin{aligned}\mathbf{E}\Big(\mathbf{P}\big(\xi_{v_u} > x \,|\, \mathscr{G}_L\big)\mathbf{P}\big(\xi_{w_u} > x \,|\, \mathscr{G}_L\big) \,|\, \mathscr{G}_{h+1}\Big) \\ = \mathbf{E}\Big(\mathbf{P}\big(\xi_{v_u} > x \,|\, \mathscr{G}_L\big) \,|\, \mathscr{G}_{h+1}\Big) \cdot \mathbf{E}\Big(\mathbf{P}\big(\xi_{w_u} > x \,|\, \mathscr{G}_L\big) \,|\, \mathscr{G}_{h+1}\Big).\end{aligned}$$

Denote by $v, w : v \wedge w = u$ that the vertices v, w have closest ancestor u. Using (2.36) and (2.37), by calculations similar to those in (2.51) for v, w and u with $h(v) = j$, $h(w) = k$ and $h(u) = h$ respectively (where u is the closest ancestor to v and w), we get

$$\begin{aligned}&\sum_{h=l}^{L} \sum_{u:h(u)=h} \sum_{v,w \in T_u : v \wedge w = u} \mathbf{E}\Big(\mathbf{P}\big(\xi_v > x \,|\, \mathscr{G}_L\big) \cdot \mathbf{P}\big(\xi_w > x \,|\, \mathscr{G}_L\big) \,|\, \mathscr{G}_{h+1}\Big) \quad (2.78)\\
&= \sum_{h=l}^{L} \sum_{u:h(u)=h} \sum_{\substack{h \leqslant j \leqslant L,\\ h \leqslant k \leqslant L}} \sum_{\substack{v,w \in T_u : v \wedge w = u;\\ h(v)=j, h(w)=k}} \mathbf{E}\Big(\mathbf{P}\big(\xi_v > x \,|\, \mathscr{G}_L\big)\mathbf{P}\big(\xi_w > x \,|\, \mathscr{G}_L\big) \,|\, \mathscr{G}_{h+1}\Big)\\
&\leqslant \sum_{h=l}^{L} \sum_{u:h(u)=h} \sum_{\substack{h \leqslant j \leqslant L,\\ h \leqslant k \leqslant L}} \sum_{\substack{v,w \in T_u : v \wedge w = u;\\ h(v)=j, h(w)=k}}\\
&\qquad \mathbf{E}\Big(\frac{1}{m} \ln_+\Big(\frac{m \prod_{r=1}^{j} U_{r,v}}{x}\Big) \frac{1}{m} \ln_+\Big(\frac{m \prod_{r=1}^{k} U_{r,w}}{x}\Big) \,\Big|\, \mathscr{G}_{h+1}\Big)\\
&\leqslant \sum_{h=l}^{L} \sum_{u:h(u)=h} \sum_{j=h}^{L} \sum_{\substack{v \in T_u :\\ h(v)=j}} \mathbf{E}\Big(\frac{1}{m} \ln_+ \frac{m \prod_{r=1}^{j} U_{r,v}}{x} \,\Big|\, \mathscr{G}_{h+1}\Big)\\
&\qquad \cdot \sum_{h=l}^{L} \sum_{u:h(u)=h} \sum_{k=h}^{L} \sum_{\substack{w \in T_u :\\ h(w)=k}} \mathbf{E}\Big(\frac{1}{m} \ln_+ \frac{m \prod_{r=1}^{k} U_{r,w}}{x} \,\Big|\, \mathscr{G}_{h+1}\Big). \quad (2.79)\end{aligned}$$

Note that the expected value of the sum in (2.78) is equal to the right-hand side of the inequality in (2.77).

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

From the notation of $\widehat{n_v}$ in (2.49), for a triple (v,w,u) with $h(v)=j, h(w)=k, h(u)=h$, we have

$$\widehat{n_v} = n\prod_{r=1}^{j} U_{r,v} = n\prod_{r=1}^{h} U_{r,u}\widehat{\mathcal{U}_u}\prod_{r=h+2}^{j} U_{r,v},$$
$$\widehat{n_w} = n\prod_{r=1}^{k} U_{r,w} = n\prod_{r=1}^{h} U_{r,u}(1-\widehat{\mathcal{U}_u})\prod_{r=h+2}^{k} U_{r,w},$$

where $\widehat{\mathcal{U}_u}$ is either $\mathcal{U}_u$ or $1-\mathcal{U}_u$ depending on whether v is to the right or left of w. For given $h(u)=h\geqslant l, h(v)=j$ and $h(w)=k$ there are at most 2^h choices of u, and then at most 2^{j-h} choices of v and 2^{k-h} choices of w. (We can assume that $j>h+1$ and $k>h+1$, since it is easy to see that the other terms are few and the sum of them is small.) Let u_v be the child of u that is an ancestor of v; respectively, let u_w be the child of u that is an ancestor of w. Thus, for u_v with $h(u_v)=h+1$ we have $Y_{h+1,u_v} := -\sum_{r=1}^{h}\ln U_{r,u} - \ln\widehat{\mathcal{U}_u}$, and for u_w with $h(u_w)=h+1$ we have $Y_{h+1,u_w} := -\sum_{r=1}^{h}\ln U_{r,u} - \ln(1-\widehat{\mathcal{U}_u})$. For the vertex v with $h(v)=j$, from the notation of $Z_{j-l,v}$ in (2.70) we have that $Z_{j-h-1,v} := Y_{j,v} - Y_{h+1,u_v} = -\sum_{r=h+2}^{j}\ln U_{r,v}$ (and use the corresponding notation for w with $h(w)=k$).

By calculations similar to those in (2.52) and (2.55), the sum in (2.78) is given by

$$\begin{aligned}
&\frac{1}{m^2}O\Bigg(\sum_{h=l}^{L}\sum_{h(u)=h}\Bigg(\sum_{j=h+2}^{L}\int_0^{\ln m-\ln x-Y_{h+1,u_v}} \\
&\qquad \Big(2^{j-h-1}(\ln m-\ln x-Y_{h+1,u_v}-z)\frac{e^{-z}z^{j-h-2}}{(j-h-2)!}\Big)dz\Bigg) \\
&\qquad\cdot\sum_{k=h+2}^{L}\int_0^{\ln m-\ln x-Y_{h+1,u_w}} 2^{k-h-1}(\ln m-\ln x-Y_{h+1,u_w}-z)\frac{e^{-z}z^{k-h-2}}{(k-h-2)!}dz\Bigg) \\
&= O\Bigg(\sum_{h=l}^{L}\sum_{h(u)=h}\frac{1}{m^2}\Bigg(\int_0^{\ln m-\ln x-Y_h} e^z dz\Bigg)^2\Bigg) \\
&= O\Bigg(\sum_{h=l}^{L}\sum_{h(u)=h}\frac{1}{x^2}\prod_{r=1}^{h}U_r^2\Bigg).
\end{aligned}$$

The expected value of this is

$$O\Bigg(\mathbf{E}\Bigg(\sum_{h=l}^{L}\sum_{h(u)=h}\frac{1}{x^2}\prod_{r=1}^{h}U_r^2\Bigg)\Bigg) = O\Bigg(\sum_{h=l}^{L}\frac{2^h}{3^h}\frac{1}{x^2}\Bigg) = o(1),$$

and thus the right-hand side of the inequality in (2.77) is $o(1)$, which gives

$$\mathbf{E}\Bigg(\mathbf{Var}\Bigg(\sum_{l\leqslant h(v)\leqslant L}\mathbf{P}(\xi_v > x\,|\,\mathcal{G}_L)\,\Big|\,\mathcal{G}_l\Bigg)\Bigg)\to 0,$$

which shows (2.33).

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

We proceed by showing (2.34). First we consider

$$\mathbf{Var}\Big(\sum_{l\leqslant h(v)\leqslant L}\mathbf{E}\big(\xi_v\mathbf{1}[\xi_v\leqslant c]\,|\,\mathscr{G}_L\big)\,\Big|\,\mathscr{G}_l\Big)$$
$$=\sum_{\substack{l\leqslant h(v)\leqslant L,\\ l\leqslant h(w)\leqslant L}}\mathbf{Cov}\Big(\mathbf{E}\big(\xi_v\mathbf{1}[\xi_v\leqslant c]\,|\,\mathscr{G}_L\big),\mathbf{E}\big(\xi_w\mathbf{1}[\xi_w\leqslant c]\,|\,\mathscr{G}_L\big)\,|\,\mathscr{G}_l\Big). \tag{2.80}$$

As in (2.33), we can suppose that the closest ancestor u for v and w is at height $h\geqslant l$. As in (2.75) we have

$$\mathbf{Cov}\Big(\mathbf{E}\big(\xi_v\mathbf{1}[\xi_v\leqslant c]\,|\,\mathscr{G}_L\big),\mathbf{E}\big(\xi_w\mathbf{1}[\xi_w\leqslant c]\,|\,\mathscr{G}_L\big)\,|\,\mathscr{G}_l\Big)$$
$$\leqslant\mathbf{E}\Big(\mathbf{E}\big(\xi_v\mathbf{1}[\xi_v\leqslant c]\,|\,\mathscr{G}_L\big)\mathbf{E}\big(\xi_w\mathbf{1}[\xi_w\leqslant c]\,|\,\mathscr{G}_L\big)\,|\,\mathscr{G}_l\Big).$$

Thus,

$$\mathbf{E}\Big(\mathbf{Cov}\Big(\mathbf{E}\big(\xi_v\mathbf{1}[\xi_v\leqslant c]\,|\,\mathscr{G}_L\big),\mathbf{E}\big(\xi_w\mathbf{1}[\xi_w\leqslant c]\,|\,\mathscr{G}_L\big)\,|\,\mathscr{G}_l\Big)\Big)$$
$$\leqslant\mathbf{E}\Big(\mathbf{E}\big(\xi_v\mathbf{1}[\xi_v\leqslant c]\,|\,\mathscr{G}_L\big)\mathbf{E}\big(\xi_w\mathbf{1}[\xi_w\leqslant c]\,|\,\mathscr{G}_L\big)\Big).$$

We use calculations as in (2.56) to deduce that, for a vertex v with $h(v)=j$,

$$\mathbf{E}\big(\xi_v\mathbf{1}[\xi_v\leqslant c]\,|\,\mathscr{G}_L\big)=\frac{mn_v}{n(m+1)}e^{-\frac{m+1}{m}\ln_+(\frac{mn_v}{nc})}\leqslant\frac{n_v}{n}\leqslant\prod_{r=1}^{j}U_{r,v},$$

where the last inequality follows from (2.48).

Let (v_u,w_u) denote a pair of vertices with closest ancestor u as in (2.77). Consider one such pair (v_u,w_u) and let $h=h(u)$, $j=h(v)$ and $k=h(w)$. Since $\mathbf{E}(U_{r,v})=\frac{1}{2}$ and $\mathbf{E}(U_{h,u}(1-U_{h,u}))=\frac{1}{6}<\frac{1}{4}$, where u with $h(u)=h$ is the closest ancestor of v and w, it follows that

$$\mathbf{E}\Big(\mathbf{E}\big(\xi_{v_u}\mathbf{1}[\xi_{v_u}\leqslant c]\,|\,\mathscr{G}_L\big)\mathbf{E}\big(\xi_{w_u}\mathbf{1}[\xi_{w_u}\leqslant c]\,|\,\mathscr{G}_L\big)\Big)\leqslant\mathbf{E}\prod_{r=1}^{h}U_{r,u}^2\,2^{-(j-h)-(k-h)}$$
$$=\Big(\frac{1}{3}\Big)^h2^{-(j-h)-(k-h)}. \tag{2.81}$$

Thus, by using (2.80)–(2.81) and letting $v,w:v\wedge w=u$ denote that the vertices v,w have closest ancestor u as in (2.78) and (2.79), we get

$$\mathbf{E}\Big(\mathbf{Var}\Big(\sum_{l\leqslant h(v)\leqslant L}\mathbf{E}\big(\xi_v\mathbf{1}[\xi_v\leqslant c]\,|\,\mathscr{G}_L\big)\,\Big|\,\mathscr{G}_l\Big)\Big)$$
$$\leqslant\sum_{h=l}^{L}\sum_{u:h(u)=h}\sum_{v,w\in T_u:v\wedge w=u}\mathbf{E}\Big(\mathbf{E}\big(\xi_v\mathbf{1}[\xi_v\leqslant c]\,|\,\mathscr{G}_L\big)\mathbf{E}\big(\xi_w\mathbf{1}[\xi_w\leqslant c]\,|\,\mathscr{G}_L\big)\Big)$$

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

$$= \sum_{h=l}^{L} \sum_{u:h(u)=h} \sum_{\substack{h\leqslant j\leqslant L,\\ h\leqslant k\leqslant L}} \sum_{\substack{v,w\in T_u: v\wedge w=u\\ :h(v)=j, h(w)=k}} \mathbf{E}\Big(\mathbf{E}\big(\xi_v \mathbf{1}[\xi_v \leqslant c] \,|\, \mathscr{G}_L\big)\mathbf{E}\big(\xi_w \mathbf{1}[\xi_w \leqslant c] \,|\, \mathscr{G}_L\big)\Big)$$

$$\leqslant \sum_{h=l}^{L} \left(\frac{1}{3}\right)^h 2^h \sum_{j=h}^{L} 2^{j-h-(j-h)} \sum_{k=h}^{L} 2^{k-h-(k-h)} \leqslant 3L^2 \left(\frac{2}{3}\right)^l \to 0. \tag{2.82}$$

We now show that

$$\mathbf{E}\left(\mathbf{Var}\left(\frac{4\ln^2 n}{n} \sum_{h(v)=L} \frac{n_v}{2\ln n_v} \,\Big|\, \mathscr{G}_l\right)\right) \to 0. \tag{2.83}$$

Using the expansion of $\sum_{h(v)=L} \frac{n_v}{2\ln n_v}$ in (2.62) and (2.63), we only need to show that

$$\mathbf{E}\left(\mathbf{Var}\left(\sum_{v:h(v)=L} \prod_{r=1}^{L} U_{r,v} \sum_{r=1}^{L} \ln U_{r,v} \,\Big|\, \mathscr{G}_l\right)\right) \to 0.$$

Using (2.82), we obtain for each $s \leqslant L$

$$\mathbf{E}\left(\mathbf{Var}\left(\sum_{v:h(v)=L} \prod_{r=1}^{L} U_{r,v} \ln U_{s,v} \,\Big|\, \mathscr{G}_l\right)\right) = \mathcal{O}\left(L^2\left(\frac{2}{3}\right)^l\right).$$

Thus, the conditional Hölder inequality (see, *e.g.*, [8, p. 476]) gives that

$$\mathbf{E}\left(\mathbf{Var}\left(\sum_{v:h(v)=L} \prod_{r=1}^{L} U_{r,v} \sum_{r=1}^{L} \ln U_{r,v} \,\Big|\, \mathscr{G}_l\right)\right) = \mathcal{O}\left(L^4\left(\frac{2}{3}\right)^l\right).$$

Hence,

$$\mathbf{E}\left(\mathbf{Var}\left(\frac{4\ln^2 n}{n} \sum_{h(v)=L} \frac{n_v}{2\ln n_v} \,\Big|\, \mathscr{G}_l\right)\right) \to 0.$$

Again the conditional Hölder inequality gives that

$$\mathbf{Cov}\left(\sum_{l\leqslant h(v)\leqslant L} \mathbf{E}(\xi_v \mathbf{1}[\xi_v \leqslant c] \,|\, \mathscr{G}_L), \sum_{h(v)=L} \frac{n_v}{2\ln n_v} \,\Big|\, \mathscr{G}_l\right)$$
$$\leqslant \max\left(\mathbf{Var}\left(\sum_{l\leqslant h(v)\leqslant L} \mathbf{E}\big(\xi_v \mathbf{1}[\xi_v \leqslant c] \,|\, \mathscr{G}_L\big) \,\Big|\, \mathscr{G}_l\right), \frac{4\ln^2 n}{n} \mathbf{Var}\left(\sum_{h(v)=L} \frac{n_v}{2\ln n_v} \,\Big|\, \mathscr{G}_l\right)\right).$$

Thus, (2.82) and (2.83) imply that

$$\mathbf{E}\left(\mathbf{Var}\left(\sum_{l\leqslant h(v)\leqslant L} \mathbf{E}\big(\xi_v \mathbf{1}[\xi_v \leqslant c] \,|\, \mathscr{G}_L\big) - \frac{4\ln^2 n}{n} \sum_{h(v)=L} \frac{n_v}{2\ln n_v} \,\Big|\, \mathscr{G}_l\right)\right) \to 0,$$

which shows (2.34).

It remains to show that

$$\mathbf{E}\left(\mathbf{Var}\left(\sum_{l\leqslant h(v)\leqslant L} \mathbf{Var}\big(\xi_v \mathbf{1}[\xi_v \leqslant c] \,|\, \mathscr{G}_L\big) \,\Big|\, \mathscr{G}_l\right)\right) \to 0.$$

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press

To show this we observe that

$$\mathbf{Var}\big(\xi_v \mathbf{1}[\xi_v \leqslant c] \,|\, \mathscr{G}_L\big) \leqslant \mathbf{E}\big(\xi_v^2 \mathbf{1}[\xi_v \leqslant c] \,|\, \mathscr{G}_L\big) \leqslant c\mathbf{E}\big(\xi_v \mathbf{1}[\xi_v \leqslant c] \,|\, \mathscr{G}_L\big),$$

and thus (2.35) follows from (2.81) by calculations similar to those in (2.82). □

Acknowledgement

I gratefully acknowledge the help and support of Professor Svante Janson, who both presented me with the problem and gave invaluable guidance throughout the work.

References

[1] Chauvin, B., Drmota, M. and Jabbour Hattab, J. (2001) The profile of binary search trees. *Ann. Applied Probab.* **11** 1042–1062.
[2] Devroye, L. (1986) A note on the height of binary search trees. *J. Assoc. Comput. Mach.* **33** 489–498.
[3] Devroye, L. (1998) Branching processes and their applications in the analysis of tree structures and tree algorithms. In *Probabilistic Methods for Algorithmic Discrete Mathematics*, Vol. 16 (Habib *et al.*, eds), Springer, Berlin, pp. 249–314.
[4] Devroye, L. (2005) Applications of Stein's method in the analysis of random binary search trees. In *Stein's Method and Applications* (Chen and Barbour, eds), Vol. 5 of *Inst. for Math. Sci. Lect. Notes Ser.*, World Scientific, Singapore, pp. 47–297.
[5] Drmota, M., Iksanov, A., Moehle, M. and Roesler, U. (2009) A limiting distribution for the number of cuts needed to isolate the root of a random recursive tree. *Random Struct. Alg.* **34** 319–336.
[6] Fill, J. A. and Janson, S. (2002) Quicksort asymptotics. *J. Algorithms* **44** 4–28.
[7] Feller, W. (1971) *An Introduction to Probability Theory and Its Applications*, Vol. II, 2nd edn, Wiley, New York.
[8] Gut, A. (2005) *Probability: A Graduate Course*, Springer, New York.
[9] Holmgren, C. (2008) Random records and cuttings in split trees: Extended abstract. *Discrete Math. Theor. Comput. Sci.* (DMTCS Proceedings) **AI**, 269–282.
[10] Holmgren, C. (2009) A weakly 1-stable limiting distribution for the number of random records and cuttings in split trees. Submitted.
[11] Iksanov, A. and Moehle, M. (2007) A probabilistic proof of a weak limit law for the number of cuts needed to isolate the root of a random recursive tree. *Electron. Commun. Prob.* **12** 28–35.
[12] Janson, S. (2004) Random records and cuttings in complete binary trees. In *Mathematics and Computer Science III*, Birkhäuser, Basel, pp. 241–253.
[13] Janson, S. (2006) Random cuttings and records in deterministic and random trees. *Random Struct. Alg.* **29** 139–179.
[14] Janson, S., Łuczak, T. and Ruciński, A. (2000) *Random Graphs*, Wiley, New York.
[15] Kallenberg, O. (2002) *Foundations of Modern Probability*, 2nd edn, Springer.
[16] Knessl, C. and Szpankowski, W. (1999) Quicksort algorithm again revisited. *Discrete Math. Theor. Comput. Sci.* **no. 2** 43–64.
[17] Knuth, D. E. (1997) *The Art of Computer Programming I: Fundamental Algorithms*, 3rd edn, Addison-Wesley.
[18] Knuth, D. E. (2002) *The Art of Computer Programming III: Sorting and Searching*, Addison-Wesley.
[19] Meir, A. and Moon, J. W. (1970) Cutting down random trees. *J. Australian Math. Soc.* **11** 313–324.

https://doi.org/10.1017/S096354830999068X Published online by Cambridge University Press