RANDOM RECORDS AND CUTTINGS IN COMPLETE BINARY TREES

SVANTE JANSON

Abstract. We study the number of records in a complete binary tree with randomly labeled vertices or edges. Equivalently, we may study the number of random cuttings required to eliminate a complete binary tree.

The distribution is, after normalization, asymptotically a periodic function of $\lg n - \lg\lg n$; thus there is no true asymptotic distribution but a family of limits of different subsequences; these limits are similar to a 1-stable distribution but have some periodic fluctuations.

1. Introduction

Let each vertex v in a rooted tree T have a random value λ_v attached to it, and assume that these values are i.i.d. with a continuous distribution (so that a.s. there are no ties). Say that a 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. Note that this generalizes the classical record problem (which is the case when T is a path), see for example [9].

Alternatively, we may attach random values to the edges, and let $X_e(T)$ denote the number of edges with record values (along the path from the root).

It is obvious that the choice of common distribution of the labels does not affect the result, and that we as well can count the values that are largest. We can also let the labels be a random permutation of $\{1, \dots, n\}$.

The same random variables appears when we consider random cuttings of the tree T defined as follows, see [6]. Make a random cut by choosing one vertex [edge] at random. Delete this vertex [edge] so that the tree separates into two parts, and keep only the part containing the root. Continue recusively until the root is cut [only the root is left]. Then the (random) number of cuts made is $X_v(T)$ [$X_e(T)$]. (More precisely, these random variables have the same distribution.) This equivalence is shown in [5], where the asymptotic distributions are found for the random trees that can be constructed as conditioned Galton–Watson trees, for example random labelled trees and and random binary trees. See also [6, 1, 7, 8] for earlier results.

We will in this paper study the case of a complete binary tree.

Date: April 7, 2004.

The *complete binary tree* with n vertices has height $m = \lfloor \lg n \rfloor$; it has 2^k vertices of height k, $0 \le k < m$, and $n - 2^m + 1$ vertices of height m, moreover, the vertices of height m have the leftmost positions among the 2^m possible ones, see e.g. [4, page 401]. We denote this rooted tree by T_n, and denote its root by o.

Let $\{x\} := x - \lfloor x \rfloor$ denote the fractional part of a real number x. Further, for a vertex v in a rooted tree, let $h(v)$ be its height (also known as depth), with the root having height 0.

Theorem 1.1. *Suppose that $n \to \infty$ such that $\{\lg n - \lg\lg n\} \to \gamma \in [0,1]$. Then*

$$\Bigl(X_v(T_n) - \frac{n}{\lg n} - \frac{n \lg\lg n}{\lg^2 n}\Bigr) \Bigm/ \frac{n}{\lg^2 n} \xrightarrow{\mathrm{d}} -W_\gamma \tag{1.1}$$

where W_γ has an infinitely divisible distribution with characteristic function

$$\mathbb{E}\, e^{\mathrm{i}tW_\gamma} = \exp\Bigl(\mathrm{i} f(\gamma)t + \int_0^\infty \bigl(e^{\mathrm{i}tx} - 1 - \mathrm{i}tx\mathbf{1}[x<1]\bigr)\, d\nu_\gamma(x)\Bigr), \tag{1.2}$$

where $f(\gamma) := 2^\gamma - 1 - \gamma$ and the Lévy measure ν_γ is supported on $(0,\infty)$ and has density

$$\frac{d\nu_\gamma}{dx} = 2^{\{\lg x + \gamma\}} x^{-2}.$$

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

We prove Theorem 1.1 in Section 2. The strategy of the proof is to approximate $X_v(T_n)$ and $X_e(T_n)$ by a sum of independent random variables derived from $\{\lambda_v\}$, see Lemma 2.4. We will then apply a classical limit theorem for triangular arrays.

Remark 1.2. Let $\widetilde{X}_n$ denote the left hand side of (1.1). Instead of stating the result for suitable subsequences, we may say that $\widetilde{X}_n$ has approximatively the same distribution as $-W_{\{\lg n - \lg\lg n\}}$ for large n; more precisely, the distance between the two distributions (in for example the Lévy metric) tends to 0 as $n \to \infty$.

Remark 1.3. Most records occur at height close to the maximum $m \approx \lg n$, simply because almost all vertices are there. On the other hand, it follows from the proof below, see Lemma 2.4 and the proof of Lemma 2.5, that most of the random fluctuations of $X_v(T_n)$ or $X_e(T_n)$ can be explained by the values at heights close to $\lg\lg n$. The explanation is that a few values λ_v at these heights will be so small that they significantly reduce the number of records among their descendants. Vertices of smaller height are too few, and there will usually not be any sufficiently small value among them, while vertices of larger height affect only a small proportion of the tree each, and the random effect caused by their values will be wiped out by the law of large numbers.

Remark 1.4. It is easy to see [5] that $\mathbb{E}\, X_v(T_n) = \sum_v 1/(h(v)+1)$ and $\mathbb{E}\, X_e(T_n) = \sum_{v \ne o} 1/h(v)$; both sums are easily evaluated as $n/m + O(n/m^2) =$

$n/\lg n + O(n/\lg^2 n)$. We see thus from Theorem 1.1 that $X_v(T_n)$ and $X_e(T_n)$ are concentrated well above their means (at a distance of about $n \lg\lg n/\lg^2 n$), so that e.g. $\mathbb{P}\big(X_v(T_n) \le \mathbb{E}\,X_v(T_n)\big) \to 0$. This is connected to the fact that the limit W_γ has infinite mean.

Note also that $\mathbb{E}\,X_e(T_n) - \mathbb{E}\,X_v(T_n) = \sum_{v\neq o}\big(h(v)(h(v)+1)\big)^{-1} - 1 \sim n/\lg^2 n$, while there is no similar difference in the limit distribution in Theorem 1.1.

An explanation of these facts is that the mean is affected by the unlikely event that a vertex close to the root has an extremely small value λ_v, which would reduce the number of records by a large amount.

We see that this behaviour makes it impossible to use the method of moments to find the asymptotic distribution in Theorem 1.1, as we did for other trees in [5].

Remark 1.5. Recall that the Lévy measure $cx^{-2}\,dx$ gives a (weakly) 1-stable distribution, see e.g. [2, XVII.3]; the measure ν_γ is a version of this with periodic fluctuations, so the distribution of W_γ is roughly similar to a 1-stable distribution. More precisely, we have that if W_γ and W'_γ are independent with the same distribution, then $W_\gamma + W'_\gamma \overset{\mathrm{d}}{=} 2W_\gamma + 2$, as is easily checked from (1.2), but the corresponding statement for a sum of three copies of W_γ is false.

If we write (1.2) as $\mathbb{E}\,e^{\mathrm{i}tW_\gamma} = e^{\psi_\gamma(t)}$, it is possible to compute the Fourier coefficients of $\psi_\gamma(t)$ as a function of γ by integrations, using Fubini and some Gamma integrals, and obtain

$$\psi_\gamma(t) = -\frac{\pi}{2}|t| - (\gamma^* - \tfrac12)\mathrm{i}t \\ - \mathrm{i}t\sum_{n\neq 0}\frac{\Gamma(2\pi\mathrm{i}n/\ln 2 - 1)}{\ln 2 - 2\pi \mathrm{i}n} e^{-\pi^2 n \operatorname{sign} t/\ln 2}|t|^{-2\pi\mathrm{i}n/\ln 2} e^{2\pi n\mathrm{i}\gamma},$$

where γ^* is Euler's constant. We omit the details. This, again, shows the affinity with stable distributions.

The complete binary tree T_n has minimal height among all binary trees with n vertices, but among binary trees with this height, it is maximally unbalanced. The other extreme is the *balanced binary tree* T_n^*, where at each vertex, the two subtrees emanating from it differ in size by at most 1. This tree too has height $m = \lfloor \lg n\rfloor$, and the same number of vertices at each level as T_n. As a companion to Theorem 1.1, we give a similar theorem for T_n^*; note that the results are similar but not identical, which shows that the details of the structure of the tree are important. (If we consider only n of the form $2^k - 1$, $T_n = T$ is a *full binary tree*. Indeed, Theorems 1.1 and 1.6 yield the same result in this case.) In contrast, note that the means of X_v and X_e are the same for T_n and T_n^*, see Remark 1.4.

Theorem 1.6. *Suppose that $n \to \infty$ such that $\{\lg \lg n\} \to \beta \in [0,1]$. Then*

$$\Bigl(X_v(T_n^*) - \frac{n}{\lg n} - \frac{n \lg \lg n}{\lg^2 n}\Bigr) \Big/ \frac{n}{\lg^2 n} \xrightarrow{\mathrm{d}} -W_{1-\beta}$$

where $W_{1-\beta}$ is as in Theorem 1.1. The same result holds for $X_e(T_n^*)$.

The method used below applies also to other binary trees with minimal height, but we leave the details to the reader. Presumably, the method can be used also for a larger class of binary trees, but we have not explored this. In particular, we do not know whether our methods can be used to solve the following problem.

Problem 1.7. *What is the asymptotic distribution of X_v and X_e for a (random) binary search tree?*

2. Proofs

We first treat the case X_v of Theorem 1.1 in detail, and then indicate the small modifications needed for X_e and for T_n^*.

Let $X_n := X_v(T_n)$, and let, for $y > 0$, $X_{n,y}$ be $X_v(T_n) - 1$ conditioned on the root label $\lambda_o = y$, i.e. the number of records in the rest of the tree if we fix the root label (which always is a record).

We will use the notations $m := \lfloor \lg n \rfloor$ (as above) and $l := \lfloor \lg \lg n \rfloor$; we also let $L := \lfloor \frac{3}{2} \lg \lg n \rfloor \approx 3l/2$. We assume that n is so large that $0 < l < L < m$. If a_n are positive numbers and Z_n random variables such that $Z_n/a_n \xrightarrow{\mathrm{p}} 0$ as $n \to \infty$, we write $Z_n = o_p(a_n)$.

In the sequel, we will 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 be the number of vertices in T_v.

For later use we note that if we fix $j < m$ and consider the 2^j vertices of height j, labelling them $v_1, \dots, v_{2^j}$ from left to right, then, with $q_j := \lfloor (n - 2^m + 1)/2^{m-j} \rfloor$,

$$n_{v_i} = \begin{cases} 2^{m+1-j} - 1, & 1 \le i \le q_j, \\ 2^{m-j} - 1 + 2^{m-j}\{(n - 2^m + 1)/2^{m-j}\}, & i = q_j + 1, \\ 2^{m-j} - 1, & q_j + 1 < i \le 2^j. \end{cases} \tag{2.1}$$

We will further assume that the labels λ_v have an exponential distribution $\mathrm{Exp}(1)$ with mean 1; as remarked above, this does not affect the distribution of X_n.

Lemma 2.1. *We have*

$$\mathbb{E}\, X_{n,y} = \frac{n - 2^m + 1}{m}\bigl(1 - e^{-my}\bigr) + \sum_{k=1}^{m-1} \frac{2^{m-k}}{m-k}\bigl(1 - e^{-(m-k)y}\bigr) \tag{2.2}$$

and, uniformly in n and $y > 0$,

$$\operatorname{Var} X_{n,y} = O(m^{-3} n^2).$$

Proof. Fix $y > 0$, and let, for each vertex $v \in T$, I_v be the indicator that λ_v is a minimum, given that $\lambda_o = y$. Thus, $X_{n,y} = \sum_{v \neq o} I_v$. If $h(v) = j$, let $o, v_1, \dots, v_j = v$ be the vertices on the path from the root o to v. Then $I_v = 1$ if and only if $\lambda_{v_j} < y$ and $\lambda_{v_i} > \lambda_{v_j}$ for $i = 1, \dots, j-1$. Hence, since $\lambda_{v_i} \sim \mathrm{Exp}(1)$ are independent,

$$\mathbb{E}\, I_v = \int_0^y \prod_{i=1}^{j-1} \mathbb{P}(\lambda_{v_i} > x) e^{-x}\, dx = \int_0^y e^{-jx}\, dx = \frac{1 - e^{-jy}}{j}. \tag{2.3}$$

Consequently,

$$\mathbb{E}\, X_{n,y} = \sum_{j=1}^{m-1} 2^j \frac{1 - e^{-jy}}{j} + (n - 2^m + 1) \frac{1 - e^{-my}}{m},$$

proving (2.2) by letting $j = m - k$.

To estimate the variance, assume that v and w are two vertices in T of heights $j = h(v)$ and $k = h(w)$, and with their last common ancestor u at height i.

Suppose first $i < j$ and $i < k$. Let $u_0 = o, u_1, \dots, u_i = u$ be the vertices on the path from o to u, and let $Z := \min\{\lambda_{u_s} : 1 \le s \le i\}$. Conditioned on Z, I_v and I_w are independent. Further, since v has height $j - i$ above u, (2.3) yields

$$\mathbb{E}(I_v \mid Z) = \frac{1 - e^{-(j-i)(Z \wedge y)}}{j - i},$$

and similarly for I_w. Consequently, since $Z \sim \mathrm{Exp}(i^{-1})$, being the minimum of i independent $\mathrm{Exp}(1)$ variables,

$$\begin{aligned}
\mathbb{E}(I_v I_w) &= \mathbb{E}\Big(\frac{1 - e^{-(j-i)(Z\wedge y)}}{j-i} \cdot \frac{1 - e^{-(k-i)(Z\wedge y)}}{k-i} \Big) \\
&= \frac{1}{j-i} \frac{1}{k-i} \bigg(\int_0^y \big(1 - e^{-(j-i)z}\big)\big(1 - e^{-(k-i)z}\big) i e^{-iz}\, dz \\
&\qquad\qquad + e^{-iy}\big(1 - e^{-(j-i)y}\big)\big(1 - e^{-(k-i)y}\big) \bigg) \\
&= \frac{1}{j-i} \frac{1}{k-i} \Big(1 - e^{-iy} - \frac{i}{j}\big(1 - e^{-jy}\big) - \frac{i}{k}\big(1 - e^{-ky}\big) \\
&\quad + \frac{i}{j+k-i}\big(1 - e^{-(j+k-i)y}\big) + e^{-iy} - e^{-jy} - e^{-ky} + e^{-(j+k-i)y} \Big).
\end{aligned} \tag{2.4}$$

Say that the pair (v,w) is *good* if $i \le m/3$ and $j,k \ge 2m/3$, and *bad* otherwise. For a good pair (v,w) we have, by (2.4) and (2.3),

$$\begin{aligned}
\operatorname{Cov}(I_v,I_w) &= \mathbb{E}\, I_v I_w - \mathbb{E}\, I_v\, \mathbb{E}\, I_w \\
&= \frac{1+O(i/m)}{jk}\Bigl(1-e^{-jy}-e^{-ky}+e^{-(j+k-i)y}+O(i/m)\Bigr) \\
&\qquad\qquad - \frac{1}{jk}\bigl(1-e^{-jy}\bigr)\bigl(1-e^{-ky}\bigr) \\
&= \frac{1}{jk}e^{-(j+k-i)y}\bigl(1-e^{-iy}\bigr)+O(i/m^3) \\
&= O(m^{-2}e^{-my}iy)+O(i/m^3)=O(i/m^3). \qquad (2.5)
\end{aligned}$$

For given i,j,k, there are at most 2^i choices of u and then at most 2^{j-i} choices of v and 2^{k-i} of w; thus the total number of such pairs is at most 2^{j+k-i}. Hence (2.5) yields

$$\sum_{\text{good }(v,w)} \operatorname{Cov}(I_v,I_w) = O\Bigl(\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m} 2^{j+k-i} i m^{-3}\Bigr) = O\bigl(2^{2m}m^{-3}\bigr). \qquad (2.6)$$

The total number of bad pairs is at most

$$\sum_{i>m/3,\ j,k\le m} 2^{j+k-i} + 2\sum_{i\ge 0,\ j<2m/3,\ k\le m} 2^{j+k-i} = O\bigl(2^{2m-m/3}\bigr). \qquad (2.7)$$

For the bad pairs we simply use $\operatorname{Cov}(I_v,I_w) \le \mathbb{E}\, I_v I_w \le 1$, and obtain from (2.6) and (2.7)

$$\operatorname{Var} X_{n,y} = \sum_{v,w} \operatorname{Cov}(I_v,I_w) = O\bigl(2^{2m}m^{-3}\bigr) = O\bigl(m^{-3}n^2\bigr). \qquad \square$$

Let $\varphi(n,y) := \mathbb{E}\, X_{n,y}$, given by (2.2). In the next lemma, we find it useful to be slightly more general than simply requiring $\bar m = m$.

Lemma 2.2. *If* $2^{\bar m}-1 \le n \le 2^{\bar m+1}-1$, *then*

$$\varphi(n,y) = \frac{n}{\bar m}\bigl(1-e^{-\bar m y}\bigr) + \frac{2^{\bar m+1}}{\bar m^2} + O\bigl(\bar m^{-2}e^{-\bar m y/4}n + \bar m^{-3}n\bigr).$$

Proof. Let

$$a_k = \frac{2^{m-k}}{m-k}\bigl(1-e^{-(m-k)y}\bigr) = \frac{2^{m-k}}{m}\Bigl(1-\frac{k}{m}\Bigr)^{-1}\Bigl(1-e^{-my}-e^{(k-m)y}\bigl(1-e^{-ky}\bigr)\Bigr).$$

For $m/2<k<m$ we use $a_k = O(2^{m/2})$, and for $k \le m/2$

$$a_k = \frac{2^{m-k}}{m}\Bigl(1+\frac{k}{m}+O\Bigl(\frac{k^2}{m^2}\Bigr)\Bigr)\Bigl(1-e^{-my}+O\bigl(kye^{(k-m)y}\bigr)\Bigr).$$

Summing over k, we see that (2.2) yields, using $mye^{-my/4} = O(1)$,

$$\begin{aligned}\varphi(n,y) &= \frac{n-2^m}{m}\bigl(1-e^{-my}\bigr) + \frac{2^m}{m}\Bigl(1+\frac{2}{m}+O(m^{-2})\Bigr)\bigl(1-e^{-my}\bigr)\\ &\qquad\qquad + O\Bigl(m2^{m/2}+\frac{2^m}{m}ye^{-my/2}\Bigr)\\ &= \frac{n}{m}\bigl(1-e^{-my}\bigr)+\frac{2^{m+1}}{m^2}+O(m^{-3}n)+O\bigl(m^{-2}ne^{-my/4}\bigr).\end{aligned}$$

This proves the result when $\bar{m} = m$. The only remaining case is $n = 2^{\bar{m}} - 1$ and $m = \bar{m} - 1$; the result follows easily in this case too, for example by adding a vertex v at height $\bar{m}$, using the case just considered, and subtracting $\mathbb{E}\, I_v = (1 - e^{-\bar{m}y})/\bar{m}$ from (2.3). □

Recall that $L = \lfloor \frac{3}{2} \lg\lg n \rfloor \approx \frac{3}{2} l \approx \frac{3}{2} \lg m$. Let v_i, $1 \le i \le 2^L$, be the 2^L vertices of height L, and let $n_i := n_{v_i}$. Note that $n_i = \Theta(n/2^L)$. Further, let Y_i be the minimum of λ_v along the path $P(v_i) = o \dots v_i$ from the root to v_i.

Lemma 2.3. *With notations as above,*

$$X_n = \sum_{i=1}^{2^L} \varphi(n_i, Y_i) + o_p(m^{-2}n). \tag{2.8}$$

Proof. We write the number of records X_n as $V^* + V_1 + \cdots + V_{2^L}$, where V^* is the number of records with height $\le L$ and V_i are the number of records in $T_{v_i} \setminus \{v_i\}$.

If we condition on $\{\lambda_v : h(v) \le L\}$, then V^* and all Y_i become fixed, while V_i, $1 \le i \le 2^L$, become independent random variables with $V_i \overset{\mathrm{d}}{=} X_{n_i,Y_i}$.

Let $\mathcal{F}_L$ be the σ-field generated by $\{\lambda_v : h(v) \le L\}$. Then, by the comments just made, $\mathbb{E}(V_i \mid \mathcal{F}_L) = \mathbb{E}(X_{n_i,Y_i} \mid Y_i) = \varphi(n_i, Y_i)$ and, with $m_i := \lfloor \lg n_i \rfloor = \lg n - L + O(1) \sim m$,

$$\begin{aligned}\mathbb{E}\Bigl(\Bigl(X_n - V^* - \sum_{i=1}^{2^L}\varphi(n_i,Y_i)\Bigr)^2 \Bigm| \mathcal{F}_L\Bigr) &= \mathbb{E}\Bigl(\Bigl(\sum_{i=1}^{2^L}\bigl(V_i-\varphi(n_i,Y_i)\bigr)\Bigr)^2 \Bigm| \mathcal{F}_L\Bigr)\\ = \sum_{i=1}^{2^L}\mathbb{E}\Bigl(\bigl(V_i-\varphi(n_i,Y_i)\bigr)^2 \mid \mathcal{F}_L\Bigr) &= \sum_{i=1}^{2^L}\operatorname{Var}\bigl(X_{n_i,Y_i} \mid Y_i\bigr)\\ = \sum_{i=1}^{2^L} O(m_i^{-3}n_i^2) = O\bigl(2^L m^{-3}2^{-2L}n^2\bigr) &= O\bigl(m^{-3}2^{-L}n^2\bigr)\\ = O\bigl(m^{-9/2}n^2\bigr).\end{aligned}$$

Taking the expectation, we find

$$\mathbb{E}\Bigl(X_n - V^* - \sum_{i=1}^{2^L}\varphi(n_i,Y_i)\Bigr)^2 = O\bigl(m^{-9/2}n^2\bigr) = o\bigl(m^{-4}n^2\bigr)$$

and thus

$$X_n - V^* - \sum_{i=1}^{2^L} \varphi(n_i, Y_i) = o_p\bigl(m^{-2}n\bigr).$$

The result follows because also

$$0 \le V^* < 2^{L+1} = O(m^{3/2}) = o\bigl(m^{-2}n\bigr). \qquad \square$$

Next, let $\bar{m} := m - L \sim m$. By (2.1) (with $j = L$), we can apply Lemma 2.2 to each n_i and $\bar{m}$; this yields

$$\varphi(n_i, Y_i) = \frac{n_i}{\bar{m}}\bigl(1 - e^{-\bar{m}Y_i}\bigr) + \frac{2^{\bar{m}+1}}{\bar{m}^2} + O\bigl(m^{-2}e^{-\bar{m}Y_i/4}n_i + m^{-3}n_i\bigr). \tag{2.9}$$

Since $Y_i \sim \operatorname{Exp}(1/(L+1))$, for every $a > 0$,

$$\mathbb{E}\, e^{-aY_i} = \int_0^\infty (L+1)e^{-ay-(L+1)y}\,dy = \frac{L+1}{L+1+a} = O\Bigl(\frac{L}{a}\Bigr). \tag{2.10}$$

Hence

$$\begin{aligned} \mathbb{E}\bigl|e^{-\bar{m}Y_i} - e^{-mY_i}\bigr| &= \mathbb{E}\, e^{-\bar{m}Y_i} - \mathbb{E}\, e^{-mY_i} = \frac{L+1}{L+1+\bar{m}} - \frac{L+1}{L+1+m} \\ &= O\Bigl(L\frac{m-\bar{m}}{m^2}\Bigr) = O\Bigl(\frac{L^2}{m^2}\Bigr). \end{aligned} \tag{2.11}$$

It follows easily from (2.9), (2.10) and (2.11) that

$$\mathbb{E}\Bigl|\varphi(n_i, Y_i) - \frac{n_i}{\bar{m}} + \frac{n_i}{m}e^{-mY_i} - \frac{2^{\bar{m}+1}}{\bar{m}^2}\Bigr| = O\bigl(L^2 m^{-3} n_i\bigr) = o(m^{-2}n_i).$$

Summing over i we find, using Lemma 2.3, since $\sum_i n_i = n - (2^L - 1) = n - O(m^{3/2})$,

$$\begin{aligned} X_n &= \sum_{i=1}^{2^L}\Bigl(\frac{n_i}{\bar{m}} - \frac{n_i}{m}e^{-mY_i} + \frac{2^{\bar{m}+1}}{\bar{m}^2}\Bigr) + o_p(m^{-2}n) \\ &= \frac{n}{m-L} - \frac{1}{m}\sum_{i=1}^{2^L} n_i e^{-mY_i} + 2^L\frac{2^{m-L+1}}{(m-L)^2} + o_p(m^{-2}n) \\ &= \frac{n}{m} + L\frac{n}{m^2} - \frac{1}{m}\sum_{i=1}^{2^L} n_i e^{-mY_i} + \frac{2^{m+1}}{m^2} + o_p(m^{-2}n). \end{aligned} \tag{2.12}$$

We transform this once more.

Lemma 2.4.

$$X_n = \frac{n}{m} + L\frac{n}{m^2} - \frac{1}{m}\sum_{h(v)\le L} n_v e^{-m\lambda_v} + \frac{2^{m+1}}{m^2} + o_p(m^{-2}n). \tag{2.13}$$

Proof. We recall that each Y_i is the minimum of the $L+1$ independent variables λ_v, $v \in P(v_i)$; thus e^{-mY_i} is the maximum of the corresponding $e^{-m\lambda_v}$. Let $a = 2\ln m/m$. The probability that at least two λ_v, $v \in P(v_i)$, are less than a is $O(L^2a^2) = O(\ln^4 m/m^2)$; hence the probability that this happens for some i is $O(2^L \ln^4 m/m^2) = o(1)$. With probability tending to 1, there is thus at most one λ_v less than a in each $P(v_i)$, and in this case,

$$0 \le \sum_{v \in P(v_i)} e^{-m\lambda_v} - e^{-mY_i} \le Le^{-ma} = L/m^2,$$

and thus,

$$\begin{aligned}
\sum_{i=1}^{2^L} n_i e^{-mY_i} &= \sum_{i=1}^{2^L} n_i \sum_{v \in P(v_i)} e^{-m\lambda_v} + O(nL/m^2) \\
&= \sum_{h(v)\le L} e^{-m\lambda_v} \sum_{i: v \in P(v_i)} n_i + O(nL/m^2) \\
&= \sum_{h(v)\le L} e^{-m\lambda_v} n_v + O(nL/m^2),
\end{aligned}$$

because $n_v - 2^L \le \sum_{i:v\in P(v_i)} n_i \le n_v$. Hence,

$$\sum_{i=1}^{2^L} n_i e^{-mY_i} = \sum_{h(v)\le L} e^{-m\lambda_v} n_v + o_p(n/m),$$

and the result follows from (2.12). □

The sum in (2.13) is a sum of independent random variables. The proof will be completed by a classical result on convergence of such sums for triangular arrays to infinitely divisible distributions, see e.g. [3, Theorem 15.28].

We write, for convenience, $\xi_v := \frac{mn_v}{n} e^{-m\lambda_v}$. We further write $\alpha_n := \{\lg n\}$ and $\beta_n = \{\lg \lg n\}$; Thus $\lg n = m + \alpha_n$ and $\lg m = \lg\lg n + o(1) = l + \beta + o(1)$. We then have, by Lemma 2.4,

$$\begin{aligned}
&\frac{m^2}{n}\Big(X_n - \frac{n}{\lg n} - \frac{n \lg\lg n}{\lg^2 n}\Big) \\
&\quad = m^2\Big(\frac{1}{m} - \frac{1}{\lg n}\Big) + L - m \sum_{h(v)\le L} \frac{n_v}{n} e^{-m\lambda_v} + \frac{2^{m+1}}{n} - \lg\lg n + o_p(1) \\
&\quad = \alpha_n + L - l - \beta_n + 2^{1-\alpha_n} - \sum_{h(v)\le L} \xi_v + o_p(1). \qquad (2.14)
\end{aligned}$$

Since $m/\lg n \to 1$, it is thus enough to show that this converges in distribution to $-W_\gamma$ as $n \to \infty$ with $\{\lg n - \lg\lg n\} \to \gamma$.

By considering subsequences, we may assume that the limits $\alpha := \lim \alpha_n$ and $\beta := \lim \beta_n$ exist. Thus $\lg n = m + \alpha + o(1)$ and $\lg m = \lg\lg n + o(1) =$

$l+\beta+o(1)$. Note that $\lg n-\lg\lg n=m-l+\alpha-\beta+o(1)$; thus $\gamma\equiv\alpha-\beta \pmod 1$ and more precisely,

$$\gamma=\begin{cases}\alpha-\beta & \text{if } \alpha>\beta;\\ \alpha-\beta+1 & \text{if } \alpha<\beta;\\ 0 \text{ or } 1 & \text{if } \alpha=\beta.\end{cases} \tag{2.15}$$

Lemma 2.5. *Suppose that $n\to\infty$ such that $\alpha_n\to\alpha$ and $\beta_n\to\beta$ for some α and β in $[0,1]$, and let $h:=2^{\beta-\alpha}$. Then*

(i) $\sup_v \mathbb{P}(\xi_v>x)\to 0$ *for every $x>0$. (I.e., $\{\xi_v\}$ form a null array.)*
(ii) $\sum_{h(v)\le L}\mathbb{P}(\xi_v>x)\to\nu_\gamma(x,\infty)$ *for every $x>0$.*
(iii) $\sum_{h(v)\le L}\mathbb{E}\big(\xi_v\mathbf{1}[\xi_v\le h]\big)-(L-l+2^{1-\alpha}+\alpha-\beta)\to\beta-\alpha.$
(iv) $\sum_{h(v)\le L}\operatorname{Var}\big(\xi_v\mathbf{1}[\xi_v\le h]\big)\to 3h/2.$

Before proving this lemma, we show how it implies Theorem 1.1. Let $C:=L-l+2^{1-\alpha}+\alpha-\beta$. We apply [3, Theorem 15.28] with $a=0$ and $b=f(\gamma)$ to $\sum_{h(v)\le L}\xi_v+\sum_{i=1}^n\xi_i'$, with $\xi_i'=-C/n$ deterministic. (Note that $C/n\to 0$; thus $\{\xi_v\}\cup\{\xi_i'\}$ is a null array.) We have $d\nu_\gamma/dx=2^{\{\lg x+\alpha-\beta\}}x^{-2}=2^{-i+\alpha-\beta}x^{-1}$ when $2^ih<x<2^{i+1}h$, and thus

$$\int_0^h x^2\,d\nu_\gamma(x)=\sum_{i=-\infty}^{-1}\int_{2^ih}^{2^{i+1}h}2^{-i+\alpha-\beta}x\,dx=\sum_{i=-\infty}^{-1}\tfrac32 2^ih=\frac{3h}{2}.$$

Similarly, if $\beta\le\alpha$ so $1/2\le h\le 1$, then

$$\int_h^1 x\,d\nu_\gamma(x)=\int_h^1 2^{\alpha-\beta}\,dx=2^{\alpha-\beta}-1,$$

while if $\beta\ge\alpha$ so $1\le h\le 2$, then

$$\int_h^1 x\,d\nu_\gamma(x)=-\int_1^h x\,d\nu_\gamma(x)=2^{1+\alpha-\beta}(1-h)=2(2^{\alpha-\beta}-1).$$

It follows, using (2.15) and $f(0)=f(1)$, that in both cases

$$f(\gamma)-\int_h^1 x\,d\nu_\gamma(x)=2^\gamma-1-\gamma-\int_h^1 x\,d\nu_\gamma(x)=\beta-\alpha.$$

It is now easy to see from Lemma 2.5 that the conditions of [3, Theorem 15.28] are satisfied, and consequently

$$\sum_{h(v)\le L}\xi_v-(L-l+2^{1-\alpha}+\alpha-\beta)=\sum_{h(v)\le L}\xi_v+\sum_{i=1}^n\xi_i'\xrightarrow{\text{d}}W_\gamma.$$

Theorem 1.1 now follows by (2.14).

Proof of Lemma 2.5. For any $x>0$,

$$\begin{aligned}\mathbb{P}(\xi_v>x)&=\mathbb{P}\Big(e^{-m\lambda_v}>\frac{nx}{mn_v}\Big)=\mathbb{P}\Big(m\lambda_v<\ln\frac{mn_v}{nx}\Big)\\&=1-\exp\Big(-\frac1m\ln_+\frac{mn_v}{nx}\Big).\end{aligned} \tag{2.16}$$

This shows first that for every $x > 0$,

$$\mathbb{P}(\xi_v > x) < \frac{1}{m} \ln_+ \frac{mn_v}{nx} \le \frac{1}{m} \ln_+ \frac{m}{x} \to 0, \tag{2.17}$$

which proves (i).

On a given level $j < m$ there are, by (2.1), $q_j = n2^{j-m} - 2^j + O(1) = (2^{\alpha_n} - 1)2^j + O(1)$ vertices with $n_v = 2^{m+1-j} - 1$, and $2^j - q_j - 1 = (2 - 2^{\alpha_n})2^j + O(1)$ vertices with $n_v = 2^{m-j} - 1$. There is one additional vertex with an intermediate n_v (which could coincide with one of the two main values); for convenience we call such a vertex *bad*. We also call a vertex v with $n_v \ge 2^{m-l/2}$ (which requires $j \le l/2$) bad. All other vertices v with $h(v) \le L$ are *good*. The good vertices thus have $n_v = 2^{m-k} - 1$ for some k with $l/2 \le k \le L$. For $l/2 \le k < L$, there are $(2 - 2^{\alpha_n})2^k + O(1)$ such vertices with $h(v) = k$ and $(2^{\alpha_n} - 1)2^{k+1} + O(1)$ with $h(v) = k + 1$; thus together $2^{k+\alpha_n} + O(1)$. For $k = L$, there are only $(2 - 2^{\alpha_n})2^k + O(1)$ such vertices, since we require $h(v) \le L$. In other words,

$$\#\{v \text{ good} : n_v = 2^{m-k} - 1\} = \begin{cases} 2^{k+\alpha_n} + O(1), & l/2 \le k < L, \\ (2 - 2^{\alpha_n})2^L + O(1), & k = L. \end{cases} \tag{2.18}$$

The number of bad vertices is $O\bigl(L + 2^{l/2}\bigr) = O\bigl(m^{1/2}\bigr)$. By (2.17), $\mathbb{P}(\xi_v > x) = O(\ln m/m)$ for every fixed $x > 0$. Hence the sum over bad vertices in (ii) is $O(m^{-1/2} \ln m) = o(1)$.

Similarly, using (2.17) again,

$$\mathbb{E}\bigl(\xi_v \mathbf{1}[\xi_v \le h]\bigr) \le \frac{1}{m} + h\,\mathbb{P}\Bigl(\xi_v > \frac{1}{m}\Bigr) \le \frac{1}{m} + h\frac{2\ln m}{m} = O\Bigl(\frac{\ln m}{m}\Bigr). \tag{2.19}$$

and

$$\operatorname{Var}\bigl(\xi_v \mathbf{1}[\xi_v \le h]\bigr) \le \mathbb{E}\bigl(\xi_v^2 \mathbf{1}[\xi_v \le h]\bigr) \le h\,\mathbb{E}\bigl(\xi_v \mathbf{1}[\xi_v \le h]\bigr) = O\Bigl(\frac{\ln m}{m}\Bigr). \tag{2.20}$$

Consequently, the sum over bad v is $o(1)$ in (ii), (iii) and (iv), so we may in the sequel ignore them and consider only good vertices.

Fix $x > 0$. Then, by (2.16) and (2.17),

$$\mathbb{P}(\xi_v > x) = \frac{1}{m} \ln_+ \Bigl(\frac{mn_v}{nx}\Bigr)\Bigl(1 + O\Bigl(\frac{\ln m}{m}\Bigr)\Bigr) \tag{2.21}$$

If $k \geq L$, then $m2^{m-k} \leq 2^{l+1+m-L} < nx$, provided n is large enough. Thus, for large n, by (2.18) and (2.21), with all $o(1)$ uniform in k for fixed x,

$$\begin{aligned}
\sum_{v \text{ good}} \mathbb{P}(\xi_v > x) &= \bigl(1+o(1)\bigr) \sum_{k=l/2}^{\infty} \bigl(2^{k+\alpha_n} + O(1)\bigr) \frac{1}{m} \ln_+\Bigl(\frac{(2^{m-k}-1)m}{nx}\Bigr) \\
&= \bigl(1+o(1)\bigr) \sum_{k \geq l/2} 2^{k+\alpha_n-l-\beta_n} \ln_+\bigl(2^{-k-\alpha_n+l+\beta_n+o(1)} x^{-1}\bigr) + o(1) \\
&= \bigl(1+o(1)\bigr) \sum_{i \leq l/2} 2^{-i+\alpha-\beta} \ln_+\bigl(2^{i-\alpha+\beta+o(1)} x^{-1}\bigr) + o(1) \\
&\to F(x) := \sum_{-\infty}^{\infty} 2^{-i+\alpha-\beta} \ln_+\bigl(2^{i-\alpha+\beta} x^{-1}\bigr).
\end{aligned}$$

Let $j := \lfloor \lg x + \alpha - \beta \rfloor$; thus $2^{j+\beta-\alpha} \leq x < 2^{k+\beta-\alpha+1}$ and

$$\begin{aligned}
F(x) &= \sum_{i=j+1}^{\infty} 2^{-i+\alpha-\beta} \ln\bigl(2^{i-\alpha+\beta} x^{-1}\bigr) \\
&= \sum_{k=1}^{\infty} 2^{-k-j+\alpha-\beta}\bigl(k \ln 2 + \ln\bigl(2^{j-\alpha+\beta} x^{-1}\bigr)\bigr) \\
&= 2^{-j+\alpha-\beta}\bigl(2 + \lg\bigl(2^{j-\alpha+\beta} x^{-1}\bigr)\bigr) \ln 2 \\
&= 2^{\alpha-\beta-\lfloor \lg x + \alpha - \beta \rfloor}\bigl(2 - \{\lg x + \alpha - \beta\}\bigr) \ln 2 \\
&= 2^{\gamma - \lfloor \lg x + \gamma \rfloor}\bigl(2 - \{\lg x + \gamma\}\bigr) \ln 2.
\end{aligned}$$

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

$$\frac{dF(x)}{dx} = -\frac{1}{x} 2^{\gamma - \lfloor \lg x + \gamma \rfloor} = -x^{-2} 2^{\{\lg x + \gamma\}}.$$

Thus $F(x) = \nu_\gamma(x, \infty)$, which proves (ii).

For (iii) and (iv) we calculate, for $s > 0$,

$$\begin{aligned}
\mathbb{E}\bigl(e^{-m\lambda_v} \mathbf{1}[e^{-m\lambda_v} \leq s^{-1}]\bigr) &= \int_{m^{-1}\ln_+ s}^{\infty} e^{-mx} e^{-x}\, dx \\
&= \frac{1}{m+1} e^{-(m+1)\frac{1}{m}\ln_+ s} = \frac{1}{m+1} 2^{-(1+1/m)\lg_+ s} \qquad (2.22)
\end{aligned}$$

and, similarly,

$$\mathbb{E}\bigl(e^{-2m\lambda_v} \mathbf{1}[e^{-m\lambda_v} \leq s^{-1}]\bigr) = \frac{1}{2m+1} 2^{-(2+1/m)\lg_+ s},$$

which gives

$$\operatorname{Var}\bigl(e^{-m\lambda_v} \mathbf{1}[e^{-m\lambda_v} \leq s^{-1}]\bigr) = \frac{1}{2m} 2^{-(2+1/m)\lg_+ s}\bigl(1 + O(m^{-1})\bigr). \qquad (2.23)$$

If v is a good vertex with $n_v = 2^{m-k} - 1 = 2^{m-k+o(1)}$, then

$$\frac{mn_v}{nh} = 2^{(l+\beta)+(m-k)-(m+\alpha)-(\beta-\alpha)+o(1)} = 2^{l-k+o(1)},$$

and thus, by (2.22),

$$\begin{aligned}\mathbb{E}\big(\xi_v \mathbf{1}[\xi_v \le h]\big) &= \frac{mn_v}{n}\,\mathbb{E}\Big(e^{-m\lambda_v}\mathbf{1}\Big[e^{-m\lambda_v} \le \frac{nh}{mn_v}\Big]\Big)\\ &= \frac{mn_v}{(m+1)n}2^{-(1+1/m)(l-k)_+ + o(1)} = 2^{-k-\alpha-(l-k)_+ + o(1)}.\end{aligned}$$

Note in particular that if $k \ge l+1$, then $\frac{mn_v}{nh} < 1$ for large n, and thus $\xi_v \le h$ and

$$\mathbb{E}\big(\xi_v \mathbf{1}[\xi_v \le h]\big) = \mathbb{E}\,\xi_v = \frac{mn_v}{(m+1)n} = 2^{-k-\alpha_n}\Big(1+O\Big(\frac{1}{m}\Big)\Big). \tag{2.24}$$

It follows from (2.18), (2.22) and (2.24) that, with o and O uniform in k,

$$\begin{aligned}&\sum_{v \text{ good}} \mathbb{E}\big(\xi_v \mathbf{1}[\xi_v \le h]\big)\\ &= \sum_{k=l/2}^{l} 2^{k+\alpha+o(1)}2^{-k-\alpha-(l-k)_+ + o(1)} + \sum_{k=l+1}^{L-1}\big(2^{k+\alpha_n}+O(1)\big)2^{-k-\alpha_n}\big(1+O(m^{-1})\big)\\ &\qquad + \big((2-2^{\alpha_n})2^L + O(1)\big)2^{-L-\alpha+o(1)}\\ &= \sum_{k=l/2}^{l} 2^{-(l-k)_+ + o(1)} + \sum_{k=l+1}^{L-1}\big(1+O(m^{-1})\big) + 2^{1-\alpha} - 1 + o(1)\\ &= 2 + L - 1 - l + 2^{1-\alpha} - 1 + o(1) = L - l + 2^{1-\alpha} + o(1).\end{aligned}$$

Similarly, using (2.23),

$$\begin{aligned}\operatorname{Var}\big(\xi_v \mathbf{1}[\xi_v \le h]\big) &= \frac{m^2 n_v^2}{n^2}\operatorname{Var}\Big(e^{-m\lambda_v}\mathbf{1}\Big[e^{-m\lambda_v} \le \frac{nh}{mn_v}\Big]\Big)\\ &= \frac{m^2 n_v^2}{2mn^2}2^{-(2+1/m)(l-k)_+ + o(1)} = 2^{l+\beta-1-2k-2\alpha-2(l-k)_+ + o(1)}\end{aligned}$$

and

$$\begin{aligned}\sum_{v \text{ good}} \operatorname{Var}\big(\xi_v \mathbf{1}[\xi_v \le h]\big) &= \sum_{k=l/2}^{L} 2^{k+\alpha+o(1)}2^{l+\beta-1-2k-2\alpha-2(l-k)_+ + o(1)} + o(1)\\ &= \sum_{k=-\infty}^{\infty} 2^{l-k-2(l-k)_+ + \beta-\alpha-1+o(1)} + o(1)\\ &= 3\cdot 2^{\beta-\alpha-1} + o(1) = 3h/2 + o(1).\end{aligned}$$

This completes the proof of Lemma 2.5. □

We have proved Theorem 1.1 for X_v. For X_e, the only difference is that λ_o is ignored, and thus $Y_i \sim \operatorname{Exp}(1/L)$. The estimates in (2.10) and (2.11) remain valid, and thus (2.13) and (2.14) still hold, summing over $v \ne o$ only. Since $\xi_0 = me^{-m\lambda_o} \xrightarrow{\text{p}} 0$ by Lemma 2.5(i), this makes no difference for the asymptotics of the distribution. (But note that $\mathbb{E}\,\xi_0 \to 1$, and that the means differ correspondingly, see Remark 1.4.)

For the completely balanced tree (Theorem 1.6), every vertex v with $h(v) = k$ has $2^{-k}n - 2 < n_v \le 2^{-k}n$. We call all vertices with $l/2 \le h(v) \le L$ good, and replace (2.18) by

$$\#\{v \in T_n^* \text{ good} : n_v = 2^{-k}n + O(1)\} = 2^k, \qquad l \le k \le L. \tag{2.25}$$

The remaining calculations hold as above, provided we replace α_n and α by 0 and thus γ by $1 - \beta$.

References

[1] P. Chassaing & R. Marchand. In preparation.

[2] W. Feller, *An Introduction to Probability Theory and Its Applications. Vol. II.* Second edition, Wiley, New York 1971.

[3] O. Kallenberg, *Foundations of Modern Probability.* 2nd ed., Springer-Verlag, New York, 2002.

[4] D.E. Knuth, *The Art of Computer Programming. Vol. 1: Fundamental Algorithms.* 3nd ed., Addison-Wesley, Reading, Mass., 1997.

[5] S. Janson, Random cutting and records in deterministic and random trees. Preprint, 2003. Available from `http://www.math.uu.se/~svante/papers`

[6] A. Meir & J.W. Moon, Cutting down random trees. *J. Australian Math. Soc.* **11** (1970), 313–324.

[7] A. Panholzer, Cutting down very simple trees. Preprint, 2003.

[8] A. Panholzer, Non-crossing trees revisited: cutting down and spanning subtrees. *Proceedings, Discrete Random Walks 2003*, Cyril Banderier and Christian Krattenthaler, Eds., *Discr. Math. Theor. Comput. Sci.* **AC** (2003), 265–276.

[9] A. Rényi, (1962). On the extreme elements of observations. *MTA III, Oszt. Közl.* **12** (1962) 105–121. Reprinted in *Collected Works*, Vol III, pp. 50-66, Akadémiai Kiadó, Budapest, 1976.

Department of Mathematics, Uppsala University, PO Box 480, S-751 06 Uppsala, Sweden

E-mail address: `svante.janson@math.uu.se`

URL: `http://www.math.uu.se/~svante/`