JOURNAL OF COMBINATORIAL THEORY, Series A 55, 292–312 (1990)

How Many Random Questions Are Necessary to Identify n Distinct Objects?

B. PITTEL

Department of Mathematics, The Ohio State University,
Columbus, Ohio 43210

AND

H. RUBIN

Department of Statistics, Purdue University,
West Lafayette, Indiana 47907

Communicated by the Managing Editors

Received August 15, 1987

Suppose that X and A are two finite sets of the same cardinality $n \geqslant 2$. Assume that there is a bijective mapping $\phi: X \to A$ which is unknown to us, and we must determine it. We are allowed to ask a sequence of questions each posed as follows. For a given $B \subset A$ what is $\phi^{-1}(B)$? In this paper we study a case when the subsets B are chosen uniformly at random. The main result is: if each subset has to split all the atoms of a field generated by the previous subsets, then the total number of questions (needed to determine the mapping completely) is $\log_2 n + (1 + o_p(1))(2 \log_2 n)^{1/2}$. Here $o_p(1)$ stands for a random term approaching 0 in probability as $n \to \infty$. © 1990 Academic Press, Inc.

1. INTRODUCTION

Suppose that X and A are two finite sets of the same cardinality $n \geqslant 2$. Assume that there is a bijective mapping $\phi: X \to A$ which is unknown to us, and our task is to determine it. We are allowed to ask a sequence of questions each formulated as follows. For a given $B \subset A$, what is $\phi^{-1}(B)$? (An example: X is a group of students, and A is the list of their names. To get acquainted with the audience, a teacher reads the names sequentially, i.e., $|B| \equiv 1$, and asks the students to identify themselves.) A finite sequence of questions, i.e., subsets B, generates a field in A, while the answers produce a corresponding field in X. The function ϕ is determined once the atoms of field in A, hence in X, have all become singletons. Obviously, the teacher's strategy requires $n-1$ questions. The smallest number of questions is

0097-3165/90 $3.00

$\lceil \log_2 n \rceil$; it is achieved if at every step, each atom of the current field in A, which has $s \geqslant 2$ elements, delegates $\lceil s/2 \rceil$ (or $\lfloor s/2 \rfloor$) of its members for inclusion in a subset B for the next question.

What happens if the subsets B are chosen at random? The models of randomness we are interested in represent in a certain sense two opposite ends of a broad spectrum of possibilities. Model 1: At each step all 2^n subsets B are admissible, and B is chosen according to the uniform distribution, independently of all past choices. Model 2: B is admissible provided that it splits *every* nontrivial atom of the current field; further, conditioned on this field, B is chosen, among the presently admissible subsets, according to the uniform distribution. (It is worth noting that for both the teacher's strategy and the "halving" strategy all the subsets satisfy this "splitting" condition.)

Let $H_n^{(1)}$ and $H_n^{(2)}$ be respectively the random number of necessary questions for Model 1 and for Model 2. Obviously, $H_n^{(1)}$, $H_n^{(2)}$ both exceed $\lceil \log_2 n \rceil$, and the problem is to determine the probable behavior of $H_n^{(1)} - \log_2 n$ and $H_n^{(2)} - \log_2 n$ for large n.

In his lectures at Michigan State University (summer of 1960), Rényi discussed a series of problems related to random subsets, see [7]. Among them there was the problem of identifying, via unconstrained random questions, a single individual in a group of size n. In our terminology, one has to determine $\phi(x)$ for a fixed $x \in X$ by asking questions at random as defined in Model 1. Denote the random number of necessary questions by $h_n^{(1)}$; let $h_n^{(2)}$ designate the corresponding number in the case when each subset B has to split the current atom in A which contains $\phi(x)$. A little reflection shows that $h_n^{(2)}$ can be viewed as the number of questions in the $h_n^{(1)}$-long sequence of questions for the Model 1 which happen to be splitting. Rényi proved that if $y = O(1)$ and $\log_2 n + y$ is an integer then

$$P(h_n^{(1)} \leqslant \log_2 n + y) - \exp(-2^{-y}) \to 0, \qquad n \to \infty,$$

so that

$$h_n^{(1)} = \log_2 n + O_p(1), \qquad n \to \infty,$$

where $O_p(1)$ stands for a random variable bounded in probability as $n \to \infty$. Since $h_n^{(1)}$, $h_n^{(2)}$ are defined on the same probability space, we can introduce $\Delta h_n = h_n^{(1)} - h_n^{(2)}$, which is the decrease in the number of questions caused by not allowing trivial questions. How large is Δh_n? We prove a rather surprising result.

THEOREM 1. *The distribution of Δh_n does not depend on n. More precisely,*

$$P(\Delta h_n = j) = 2^{-(j+1)}, \qquad j \geqslant 0,$$

so that Δh_n *is geometrically distributed with parameter* $\frac{1}{2}$. Consequently, $E(h_n^{(2)}) = E(h_n^{(1)}) - 1$. *Thus, on average, we reduce the number of questions by* 1 *if we ask only the meaningful questions.*

The corresponding reduction in the number of questions is far more significant in the case when we have to identify all the individuals. Namely, we prove

THEOREM 2. (i) *If* $y = O(1)$ *and* $2\log_2 n + y$ *is an integer, then*

$$P(H_n^{(1)} \leqslant 2\log_2 n + y) - \exp[-2^{-(y+1)}] \to 0, \qquad n \to \infty, \tag{1.2}$$

so that

$$H_n^{(1)} = 2\log_2 n + O_p(1), \qquad n \to \infty.$$

(ii) *For every* $\varepsilon > 0$,

$$P(|(H_n^{(2)} - \log_2 n)/(2\log_2 n)^{1/2} - 1| \leqslant \varepsilon) \to 1, \qquad n \to \infty,$$

i.e.,

$$H_n^{(2)} = \log_2 n + (1 + o_p(1))(2\log_2 n)^{1/2}, \qquad n \to \infty, \tag{1.3}$$

where $o_p(1)$ *stands for a random variable which approaches* 0 *in probability as* $n \to \infty$.

Remark. Thus, selecting questions more judiciously, we can cut, with high probability (w.h.p.), their total number by half. Still, $H_n^{(2)}$ exceeds the optimal number $\lceil \log_2 n \rceil$ w.h.p. by a weighty random quantity of order $(\log n)^{1/2}$. We had expected $H_n^{(2)} - \log_2 n$ to be unbounded in probability, but firmly believed that the difference would be considerably smaller, something of magnitude $\log\log n$.

Besides being a natural extension of the original Rényi problem, Models 1 and 2 are closely related to digital search trees in computer science. To establish this connection, suppose that there are given n mutually independent *infinite* Bernoulli sequences $\omega_1 = \{\omega_{1j}\}_{j \geqslant 1}, ..., \omega_n = \{\omega_{nj}\}_{j \geqslant 1}$, with $\omega_{vj} \in \{0, 1\}$ and $P(\omega_{vj} = 1) = \frac{1}{2}$. We may interpret $\omega_1, ..., \omega_n$ as the nonterminating binary expansions of n independent numbers each distributed uniformly over $(0, 1]$. Given an n-tuple $\omega^n = (\omega_1, ..., \omega_n)$, we associate with it a finite binary tree $\mathcal{T}_n^{(1)} = \mathcal{T}_n^{(1)}(\omega^n)$ as follows. Introduce $\mathcal{T}$, the complete, infinite, plane, rooted, binary tree. Each sequence ω_v determines an infinite path $\mathcal{P}_v$ in $\mathcal{T}$; it starts at the root of $\mathcal{T}$ and is such that its jth arc (counting from the root) is directed left (right) if the jth digit ω_{vj} is 0(1). Clearly, the number of paths passing through a vertex k arcs away from the root is binomially distributed with parameters n, 2^{-k}. Thus, almost surely (a.s.) all the paths will eventually disengage. Cut each path $\mathscr{P}_v$ at the first vertex it does not share with any other path, and label this vertex ω_v. The n truncated paths define a finite (sub)tree $\mathscr{T}_n^{(1)} = \mathscr{T}_n^{(1)}(\omega^n)$ which has n end vertices labelled $\omega_1, ..., \omega_n$. This tree is interpreted as a dictionary-type arrangement of n numbers (records) in a computer's memory. Upon request, any one of ω_v can be located in the tree by starting the search at the root and using the consecutive digits of ω_v as the pointers showing where to move next. This algorithm is known as a "trie search," trie being taken from the word "retrieval" (Knuth [2, Sect. 6.3]). The height $\mathscr{H}_n^{(1)}$ of $\mathscr{T}_n^{(1)}$, i.e., the length of the longest path from the root to an end vertex, is therefore the largest number of digits to check in order to locate one of the end vertices $\omega_1, ..., \omega_n$.

Getting back to the questions-and-answers models, observe that for each model, we can also associate with every element a of A a path in $\mathscr{T}$ such that its jth arc is left (right) directed if the subset B for the jth question includes (does not include) the element a. Besides, in terms of the paths, we stop asking questions when all the paths have finally disengaged. Truncating each path at the first vertex which does not belong to any other path, we get a random binary (sub)tree, $T_n^{(1)}$ for the Model 1, and $T_n^{(2)}$ for Model 2. Obviously, the total number of questions for each of the models is equal to the height of its related tree.

For Model 1, at each stage all the questions are admissible and equally likely, independent of past questions. This implies that $T_n^{(1)}$ and $\mathscr{T}_n^{(1)}$ are equidistributed, or $T_n^{(1)} \overset{\mathscr{D}}{\equiv} \mathscr{T}_n^{(1)}$, in short. Consequently, $H_n^{(1)} \overset{\mathscr{D}}{\equiv} \mathscr{H}_n^{(1)}$.

For Model 2, unlike $T_n^{(1)}$ and $\mathscr{T}_n^{(1)}$, all the vertices of $T_n^{(2)}$, except the end vertices, must be of a branching type, i.e., have *two* direct descendants. We can get a random tree $\mathscr{T}_n^{(2)} \overset{\mathscr{D}}{\equiv} T_n^{(2)}$ from $\mathscr{T}_n^{(1)}$ by compressing the latter, that is, by removing each arc of $\mathscr{T}_n^{(1)}$ which starts at a non-branching-type vertex and patching the remaining pieces of $\mathscr{T}_n^{(1)}$ together. The resulting tree has the name "Patricia," the acronym for "practical algorithm to retrieve information coded in alphanumeric," (Knuth [3]). (More precisely, each nonend vertex of this tree is provided with an integer; it indicates the number of next irrelevant digits to be skipped over in the search for a desired ω_v.) Therefore, $H_n^{(2)} \overset{\mathscr{D}}{\equiv} \mathscr{H}_n^{(2)}$, where $\mathscr{H}_n^{(2)}$ is the height of $\mathscr{T}_n^{(2)}$. Needless to say, by compressing in the same way the tree $T_n^{(1)}$, we also get a random subtree of $T_n^{(1)}$ which is distributed like $T_n^{(2)}$. Hence, as in the case of a single individual, $H_n^{(1)}$ and $H_n^{(2)}$ can be thought of as defined on the same probability space, where $H_n^{(1)} \geqslant H_n^{(2)}$.

Note. In light of this discussion, we can reformulate Theorem 1 as follows. By compressing the tree $\mathscr{T}_n^{(1)}$ into the Patricia tree $\mathscr{T}_n^{(2)}$ we save each ω_v a random number of digits which is geometrically distributed with parameter $\frac{1}{2}$. This implies a well-known result (Knuth [3]): the *average* reduction is one digit.

This relation between Models 1, 2 and trees $\mathscr{T}_n^{(1)}$, $\mathscr{T}_n^{(2)}$ makes a proof of (1.2) unnecessary, because the corresponding estimate of $\mathscr{H}_n^{(1)}$, the height of the tree $\mathscr{T}_n^{(1)}$, is already known—see Mendelson [4], Devroye [2] (nonuniform density case), and Pittel [6] (Bernoulli sequences with a general $p=P(\omega_j=1)$). As for $H_n^{(2)}$, the situation is just the opposite, since a general result about $\mathscr{H}_n^{(2)}$ obtained by Pittel [5] implies only that $\mathscr{H}_n^{(2)}=(1+o_p(1))\log_2 n$, while (1.3)—with $\mathscr{H}_n^{(2)}$ replacing $H_n^{(2)}$—provides a considerably sharper estimate.

Actually, Mendelson [4] and Pittel [6] studied the height $\mathscr{H}_{nd}^{(1)}$ of a more general random tree $\mathscr{T}_{nd}^{(1)}$, $d\geqslant 1$; this tree is obtained via truncating every path $\mathscr{P}_v$ at the first vertex which belongs to at most d paths and labelling this end vertex by the corresponding subset of $\{\omega_1, ..., \omega_n\}$. The labelling end vertices and the subsets associated with them are interpreted as the pages of a dictionary and d as the capacity of a page. Clearly, $\mathscr{H}_{nd}^{(1)}$ and a similarly defined $\mathscr{H}_{nd}^{(2)}$ have the same distribution as $H_{nd}^{(1)}$ and $H_{nd}^{(2)}$, respectively, where $H_{nd}^{(1)}$ ($H_{nd}^{(2)}$) is the total number of questions in the Model 1 (2) needed to obtain a field with all atoms of size at most d. Using an asymptotic estimate for $\mathscr{H}_{nd}^{(1)}$, [4, 6], we obtain: if $y=O(1)$ and $d^{-1}(d+1)\log_2 n+y$ is an integer then (compare with (1.2))

$$P(H_{nd}^{(1)}\leqslant d^{-1}(d+1)\log_2 n+y)-\exp[-2^{-dy}/(d+1)!]\to 0,$$

as $n\to\infty$. As for $H_{nd}^{(2)}$, only a slight modification of the proof of Theorem 2(ii) is necessary to show that

$$H_{nd}^{(2)}=\log_2 n+(1+o_p(1))(2d^{-1}\log_2 n)^{1/2}, \qquad n\to\infty. \tag{1.4}$$

We should add that, in a parallel development, Aldous and Shields [1] studied another well-known digital tree. (For this tree, we cut each path $\mathscr{P}_v$ at the first vertex it does not share with any other path $\mathscr{P}_\mu$, $\mu\leqslant v-1$, and label this vertex ω_v.) Among other results, they proved that the height of this tree obeys the same asymptotic behavior as $\mathscr{H}_n^{(2)}$, the height of the Patricia tree. This is quite surprising since the two random trees do not have much in common.

2. Proof of Theorem 1

Let $x\in X$ be given.

(a) Since $\phi(\cdot)$ is a bijection from X to A, we can define $h_n^{(1)}$ and $h_n^{(2)}$ as follows. Let $\{C_j : j\geqslant 1\}$ be a sequence of independent random subsets of X such that each C_j is uniformly distributed on the set of all 2^n subsets of X. Introduce also a sequence $\{\mathscr{C}_j : j \geqslant 0\}$, where $\mathscr{C}_0 = X$ and, for $j \geqslant 1$, $\mathscr{C}_j = \mathscr{C}_j(x)$ is the atom of the field generated by $C_1, C_2, \ldots$ and C_j, which contains x, or formally

$$\mathscr{C}_j = \Big(\bigcap_{k \leqslant j : x \in C_k} C_k\Big) \cap \Big(\bigcap_{k \leqslant j : x \in C_k^c} C_k^c\Big)$$

Manifestly, for $j \geqslant 1$, the subset C_j does not satisfy the splitting condition iff $\mathscr{C}_j = \mathscr{C}_{j-1}$. Denote $|\mathscr{C}_j|$ by τ_j. Then

$$h_n^{(1)} = h_n^{(1)}(x) \underset{\text{def}}{=} \min\{j \geqslant 1 : \tau_j = 1\}, \tag{2.1}$$

and, introducing a sequence of events $D_j = \{h_n^{(1)} > j \text{ and } \mathscr{C}_j = \mathscr{C}_{j+1}\}$, we also have

$$h_n^{(2)} = h_n^{(2)}(x) \underset{\text{def}}{=} h_n^{(1)} - \Delta h_n, \qquad \Delta h_n = \sum_{j \geqslant 0} I_{D_j}; \tag{2.2}$$

here the summation is taken over the event indicators of D_j, $j \geqslant 0$.

(b) For $y \in X \backslash \{x\}$ and $j \geqslant 1$, we say that y is not separated from x by C_j if either $y, x \in C_j$ or $y, x \in C_j^c$. It is easy to see that every such event has probability $\frac{1}{2}$, and all these events are mutually independent. Since $\tau_j = t + 1$ iff there are exactly t elements $y \in X \backslash \{x\}$ which are not separated from x by any one of the subsets $C_1, \ldots, C_j$, we have subsequently

$$P(\tau_j = t + 1) = \binom{n-1}{t} (2^{-j})^t (1 - 2^{-j})^{n-1-t} \qquad (0 \leqslant t \leqslant n-1, j \geqslant 0). \tag{2.3}$$

Furthermore, $\{\tau_j : j \geqslant 0\}$ is a Markov chain such that

$$P(\tau_{j+1} = b \mid \tau_j = a) = \begin{cases} \binom{a-1}{b-1} 2^{-(a-1)}, & \text{if } \ 1 \leqslant b \leqslant a \leqslant n, \\ 0, & \text{if } \textit{otherwise}. \end{cases} \tag{2.4}$$

The relations (2.3) and (2.4) imply, in particular, that

$$P(h_n^{(1)} \leqslant m) = P(\tau_m = 1) = (1 - 2^{-m})^{n-1}, \qquad m \geqslant 1,$$

which directly leads to Rényi's result, see (1.1).

(c) We show next that, for every $k \geqslant 1$ and $0 \leqslant m_1 < m_2 \cdots < m_k$,

$$P\Big(\bigcap_{s=1}^{k} D_{m_s}\Big) = \Big(1 - \sum_{s=1}^{k} 2^{-(m_s+1)}\Big)^{n-1} - \Big(1 - \sum_{s=1}^{k-1} 2^{-(m_s+1)} - 2^{-m_k}\Big)^{n-1}. \tag{2.5}$$

Let $k=1$ and $m_1 \geqslant 0$. According to the definition of D_{m_1} and (2.3), (2.4),

$$\begin{aligned} P(D_{m_1}) &= P(\tau_{m_1+1} = \tau_{m_1} \geqslant 2) \\ &= \sum_{t=1}^{n-1} \binom{n-1}{t} (2^{-m_1})^t (1-2^{-m_1})^{n-1-t} 2^{-t} \\ &= (1-2^{-(m_1+1)})^{n-1} - (1-2^{-m_1})^{n-1}. \end{aligned}$$

Suppose that (2.5) is true for some $k \geqslant 1$ and all $n \geqslant 2$, $0 \leqslant m_1 < m_2 < \cdots < m_k$. Let $m_k < m_{k+1}$. Conditioning on τ_{m_1} again and using the Markov property of $\{\tau_j : j \geqslant 0\}$, by the induction hypothesis we have

$$\begin{aligned} P\left(\bigcap_{s=1}^{k+1} D_{m_s}\right) &= \sum_{t=1}^{n-1} \binom{n-1}{t} (2^{-m_1})^t (1-2^{-m_1})^{n-1-t} 2^{-t} \\ &\quad \cdot \left[\left(1 - \sum_{s=2}^{k+1} 2^{(m_s - m_1)}\right)^t \right. \\ &\quad \left. - \left(1 - \sum_{s=2}^{k} 2^{-(m_s - m_1)} - 2^{-(m_{k+1} - m_1 - 1)}\right)^t\right] \\ &= \left[1 - 2^{-m_1} + \left(2^{-(m_1+1)} - \sum_{s=2}^{k+1} 2^{-(m_s+1)}\right)\right]^{n-1} \\ &\quad - \left[1 - 2^{-m_1} + \left(2^{-(m_1+1)} - \sum_{s=2}^{k} 2^{-(m_s+1)} - 2^{-m_{k+1}}\right)\right]^{n-1}, \end{aligned}$$

or

$$\begin{aligned} P\left(\bigcap_{s=1}^{k+1} D_{m_s}\right) &= \left(1 - \sum_{s=1}^{k+1} 2^{-(m_s+1)}\right)^{n-1} \\ &\quad - \left(1 - \sum_{s=1}^{k} 2^{-(m_s+1)} - 2^{-m_{k+1}}\right)^{n-1}. \end{aligned}$$

(d) To prove that Δh_n is geometric with parameter $\frac{1}{2}$, it suffices to demonstrate that all its binomial moments are equal to 1. But, by (2.2),

$$E\left[\binom{\Delta h_n}{k}\right] = \sum_{0 \leqslant m_1 < \cdots < m_k} P\left(\bigcap_{s=1}^{k} D_{m_s}\right), \qquad k \geqslant 1.$$

Further, for fixed $0 \leqslant m_1 < \cdots < m_{k-1}$, according to (2.5), we have

$$\sum_{m_{k-1}<m_k} P\left(\bigcap_{s=1}^{k} D_{m_s}\right)=\sum_{m_{k-1}<m}\left[\left(1-\sum_{s=1}^{k-1} 2^{-(m_s+1)}-2^{-(m+1)}\right)^{n-1}\right.$$
$$\left.-\left(1-\sum_{s=1}^{k-1} 2^{-(m_s+1)}-2^{-m}\right)^{n-1}\right]$$
$$=\left(1-\sum_{s=1}^{k-1} 2^{-(m_s+1)}\right)^{n-1}$$
$$-\left(1-\sum_{s=1}^{k-1} 2^{-(m_s+1)}-2^{-(m_{k-1}+1)}\right)^{n-1}$$
$$=P\left(\bigcap_{s=1}^{k-1} D_{m_s}\right).$$

Therefore,

$$E\left[\binom{\Delta h_n}{k}\right]=\sum_{0\leqslant m_1<\cdots<m_{k-1}} P\left(\bigcap_{s=1}^{k-1} D_{m_s}\right)=E\left[\binom{\Delta h_n}{k-1}\right], \qquad k\geqslant 1,$$

and the proof is finished since $E[\binom{\Delta h_n}{0}]=1$.

3. Proof of Theorem 2

Recall (see the introduction) that the independent Bernoulli binary sequences ω_ν $(1\leqslant\nu\leqslant n)$ generate the random binary trees $\mathscr{T}_n^{(1)}$ and $\mathscr{T}_n^{(2)}$ as follows. We get $\mathscr{T}_n^{(1)}$ by tracing out the related paths $\mathscr{P}_\nu$ $(1\leqslant\nu\leqslant n)$ in the complete binary tree $\mathscr{T}$, and by cutting off each of the paths as its first vertex which does not belong to any other path. (Such a vertex for $\mathscr{P}_\nu$ is labelled ω_ν.) To get $\mathscr{T}_n^{(2)}$, we remove all the arcs of $\mathscr{T}_n^{(1)}$ which begin at a nonbranching-type vertex, and patch together the remaining pieces of $\mathscr{T}_n^{(2)}$.

As an illustration, Fig. 1 shows the trees $\mathscr{T}_5^{(1)}$, $\mathscr{T}_5^{(2)}$ for $\omega_1=(0, 1, 1, 1, 0, ...)$, $\omega_2=(0, 0, 1, 1, ...)$, $\omega_3=(0, 1, 1, 1, 1, ...)$, $\omega_4=(0, 0, 1, 0, ...)$, and $\omega_5=(0, 1, 1, 0, ...)$.

It was indicated in the Introduction that $H_n^{(i)}$ (the total number of questions for Model i) has the same distribution as the height $\mathscr{H}_n^{(i)}$ of the tree $\mathscr{T}_n^{(i)}$ $(i=1, 2)$. (In the examples above, $\mathscr{H}_5^{(1)}=5$ and $\mathscr{H}_5^{(2)}=3$.) In view of this and the existing results on $\mathscr{H}_n^{(1)}$ [2, 4, 6], we only have to prove that

$$\mathscr{H}_n^{(2)}=\log_2 n+(1+o_p(1))(2\log_2 n)^{1/2}, \qquad n\to\infty. \tag{3.1}$$

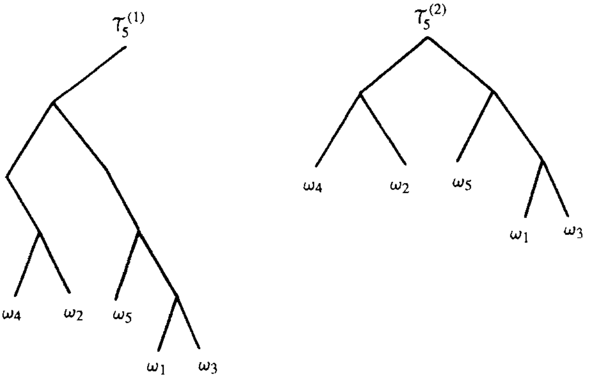

FIGURE 1

To this end, we need some auxiliary random variables $\{Y_{nk}:k\geqslant 0\}$ and $\{Z_{nk}:k\geqslant 1\}$. Call an end vertex ω_v of the tree $\mathscr{T}_n^{(1)}$ *stable* if the path leading to it from the root remains intact after compression, that is, if all the vertices along the path, except the end vertex, are branching. Y_{nk} is defined as the total number of the stable end vertices of $\mathscr{T}_n^{(1)}$ which are k arcs away from the root. As for Z_{nk}, it is equal to the total number of all the end vertices of $\mathscr{T}_n^{(1)}$ such that the corresponding path from the root has exactly k branching vertices. (In the example, $Y_{5k}=0$ for $k\geqslant 0$; $Z_{51}=0$, $Z_{52}=3$, $Z_{53}=2$, and $Z_{5k}=0$ for $k\geqslant 4$.) Evidently, Z_{nk} is also the number of the endvertices of $\mathscr{T}_n^{(2)}$ at distance k from the root.

The arguments below are organized as follows. In Section 3a, we derive the formulas for the exponential generating functions of $\{E(Y_{nk}):n\geqslant 1\}$ and $\{E(Y_{nk}(Y_{nk}-1)):n\geqslant 1\}$ $(k\geqslant 0)$. We prove then (Lemma 3.1) that, for $k_1=\log_2 n+(1-\varepsilon_n)(2\log_2 n)^{1/2}$ and $\varepsilon_n\in[a,b]\subset(0,1)$, $E(Y_{nk_1})\to\infty$. This makes plausible, but does not really prove, a conjecture that $Y_{nk_1}\to\infty$ in probability. That the conjecture is true is confirmed when we prove (Lemma 3.2) that $\operatorname{var}(Y_{nk_1})=o(E^2(Y_{nk_1}))$. This implies that w.h.p. the height $\mathscr{H}_n^{(2)}$ is *at least* k_1 when $n\to\infty$. We still have to demonstrate that w.h.p. $\mathscr{H}_n^{(2)}$ cannot be much larger. For this, in Section 3b we find a series which dominates the exponential generating function of $\{E(Z_{nk}):n\geqslant 1\}$ and using it show (Lemma 3.3) that if $k_2=\log_2 n+(1+\varepsilon)(2\log_2 n)^{1/2}$, $\varepsilon\geqslant a>0$, then $\sum_{k\geqslant k_2}E(Z_{nk})\to 0$. Consequently, w.h.p. $\mathscr{H}_n^{(2)}$ is *at most* k_2 when $n\to\infty$.

3a. *Behavior of the Moments $E(Y_{nk})$, $E^2(Y_{nk})$ and a Lower Estimate for $\mathscr{H}_n^{(2)}$*

Introduce a random variable $\gamma_n=|\{v:\omega_{v1}=0\}|$, i.e., the number of the paths $\mathscr{P}_v$ $(1\leqslant v\leqslant n)$ whose first arc is left oriented. Since $\omega_1,\ldots,\omega_n$ are independent Bernoulli, we have: (a) γ_n is binomial with parameters n, $\frac{1}{2}$, and (b) for $n \geqslant 2$, $k \geqslant 1$, the *conditional* distribution of Y_{nk}, given that $\gamma_n = j$ $(1 \leqslant j \leqslant n-1)$, is the same as the distribution of $Y_{j,k-1} + Y_{n-j,k-1}$, where $Y_{j,k-1}$ and $Y_{n-j,k-1}$ are independent; also, on $\{\gamma_n = 0 \text{ or } n\}$, $Y_{nk} = 0$. Denoting the generating function $E(\xi^{Y_{nk}})$ by $f_{nk}(\xi)$, we have

$$f_{nk}(\xi) = 2^{-n+1} + \sum_{j=1}^{n-1} 2^{-n} \binom{n}{j} f_{j,k-1}(\xi) f_{n-j,k-1}(\xi) \qquad (n \geqslant 2, k \geqslant 1). \quad (3.2)$$

Define

$$E_{nk}^{(1)} = E(Y_{nk}) = f'_{nk}(1), \qquad E_{nk}^{(2)} = E[Y_{nk}(Y_{nk}-1)] = f''_{nk}(1).$$

Differentiating both sides of (3.2) at $\xi = 1$, we obtain

$$E_{nk}^{(1)} = 2 \sum_{j=1}^{n-1} 2^{-n} \binom{n}{j} E_{j,k-1}^{(1)}, \qquad n \geqslant 2, k \geqslant 1. \quad (3.3)$$

Consider an exponential generating function of $\{E_{nk}^{(1)} : n \geqslant 1\}$, that is, $E_k^{(1)}(x) = \sum_{n \geqslant 1} E_{nk}^{(1)} x^n/n!$. Since $E_{1k}^{(1)} = \delta_{0k}$ $(k \geqslant 0)$, the recurrence relation (3.3) is equivalent to

$$\begin{aligned} E_k^{(1)}(x) &= 2 \sum_{n \geqslant 2} (x^n/n!) \sum_{j=1}^{n-1} 2^{-n} \binom{n}{j} E_{j,k-1}^{(1)} \\ &= 2 \sum_{j \geqslant 1} [E_{j,k-1}(x/2)^j/j!] \sum_{n > j} (x/2)^{n-j}/(n-j)!, \end{aligned}$$

or

$$E_k^{(1)}(x) = 2g(x/2)\, E_{k-1}^{(1)}(x/2), \qquad k \geqslant 1, \quad (3.4)$$

where

$$g(x) \underset{\text{def}}{=} e^x - 1.$$

As $E_0^{(1)}(x) = x$, it follows from (3.4) that

$$E_k^{(1)}(x) = xF_k(x), \qquad F_k(x) \underset{\text{def}}{=} \prod_{j=1}^{k} g(x/2^j). \quad (3.5)$$

Twice differentiating both sides of (3.2) at $\xi = 1$, we also get

$$E_{nk}^{(2)} = 2 \sum_{j=1}^{n-1} 2^{-n} \binom{n}{j} [E_{j,k-1}^{(2)} + E_{j,k-1}^{(1)} E_{n-j,k-1}^{(1)}], \qquad n \geqslant 2, k \geqslant 1, \quad (3.6)$$

or, setting $E_k^{(2)}(x)=\sum_{n\geqslant 1} E_{nk}^{(2)} x^n/n!$ and using $E_{1k}^{(2)}\equiv 0$,

$$E_k^{(2)}(x)=2g(x/2)\,E_{k-1}^{(2)}(x/2)+2[E_{k-1}^{(1)}(x/2)]^2, \qquad k\geqslant 1. \tag{3.7}$$

Since $E_0^{(2)}(x)\equiv 0$ (and $E_0^{(1)}(x)=x$), the recurrence relation (3.7) and (3.5) imply that

$$\begin{aligned} E_k^{(2)}(x) &= \sum_{s=1}^{k} 2^s \prod_{j=1}^{s-1} g(x/2^j)[E_{k-s}^{(1)}(x/2^s)]^2 \\ &= \sum_{s=1}^{k} (x^2/2^s)\, G_s(x), \end{aligned} \tag{3.8}$$

where

$$G_s(x)=\prod_{j=1}^{s-1} g(x/2^j) \prod_{j'=s+1}^{k} g^2(x/2^{j'}). \tag{3.9}$$

With the help of (3.5), we can now prove

Lemma 3.1. *Suppose that $k=\log_2 n+r$, $r=r(n)\to\infty$, and $r=o(n^{1/2})$. Then*

$$E_{nk}^{(1)}=(1+o(1))\,e^{-n}nF_k(n)=n2^{-(1+o(1))r^2/2}, \qquad n\to\infty. \tag{3.10}$$

Consequently, $E_{nk}^{(1)}\to\infty$ if $k=k_1=\log_2 n+(1-\varepsilon_n)(2\log_2 n)^{1/2}$, where $\varepsilon_n\in[a,b]\subset(0,1)$, and a, b are fixed.

Proof. By (3.5) and the Cauchy integral formula

$$\begin{aligned} E_{nk}^{(1)} &= n!\,(2\pi i)^{-1}\int_{\mathscr{L}} x^{-n}F_k(x)\,dx \\ &= n!\,(2\pi i)^{-1}\int_{\mathscr{L}} \exp[\Psi_{kn}(x)](x^{-1}\,dx), \end{aligned} \tag{3.11}$$

where $\mathscr{L}$ is a circular contour around the origin in the complex plane x, and

$$\Psi_{kn}(x)=\sum_{j=1}^{k}\log g(x/2^j)-(n-1)\log x. \tag{3.12}$$

(Here and below log stands for the natural logarithm function). Choose the radius of the contour $\mathscr{L}$ equal to x_n which is the positive root of $\Psi'_{kn}(x)=0$, or implicitly

$$x\sum_{j=1}^{k} 2^{-j}+\sum_{j=1}^{k}(x/2^j)(e^{x/2^j}-1)^{-1}=n-1.$$

Since

$$u(e^u-1)^{-1}\in(0,1),\quad u>0, \tag{3.13}$$

we have

$$x_n=n+O(k),\qquad n\to\infty. \tag{3.14}$$

Set $x=x_ne^{i\theta}$, $\theta\in(-\pi,\pi]$, and partition $(-\pi,\pi]$ into $[-\theta_0,\theta_0]$ and $[-\theta_0,\theta_0]^c$, where $\theta_0=n^{-5/12}$. Then

$$\begin{aligned}(2\pi i)^{-1}\int_{\mathscr{L}}\exp[\Psi_{kn}(x)](x^{-1}\,dx)&=(2\pi)^{-1}\int_{|\theta|\leqslant\theta_0}\exp[\Psi_{kn}(x_ne^{i\theta})]\,d\theta\\&\quad+(2\pi)^{-1}\int_{|\theta|>\theta_0}\exp[\Psi_{kn}(x_ne^{i\theta})]\,d\theta\\&=(2\pi)^{-1}\int_1+(2\pi)^{-1}\int_2.\end{aligned}$$

(a) To estimate $\int_1$, observe that, according to (3.13), (3.14),

$$\begin{aligned}\Psi''_{kn}(x_n)&=x^{-2}(n-1)-\sum_{j=1}^{k}e^{x/2^j}(1/2^j)^2\,(e^{x/2^j}-1)^{-2}|_{x=x_n}\\&=(n/x_n^2)(1+O(k/n))=n^{-1}(1+o(1)),\end{aligned} \tag{3.15}$$

and, likewise but more tediously,

$$\Psi^{(3)}_{kn}(x)=O(n/x_n^3)=O(n^{-2}),\qquad -\pi/4\leqslant\theta\leqslant\pi/4.$$

Since $n\theta_0^2\to\infty$ and $n\theta_0^3\to 0$, we have then

$$\begin{aligned}(2\pi)^{-1}\int_1&=(2\pi)^{-1}\exp[\Psi_{kn}(x_n)]\\&\quad\times\int_{|\theta|\leqslant\theta_0}\exp[-\Psi''_{kn}(x_n)\,x_n^2\theta^2/2+O(n\theta_o^3)]\,d\theta\\&=(1+o(1))(2\pi n)^{-1/2}\exp[\Psi_{kn}(x_n)].\end{aligned} \tag{3.16}$$

(b) To estimate $\int_2$, we need the following inequaly [5, Appendix]: if $u=ve^{i\phi}$, $v\geqslant 0$, $\phi\in(-\pi,\pi]$, then

$$|e^u-1|\leqslant(e^v-1)\exp[(v/2)(\cos\phi-1)]. \tag{3.17}$$

In view of this, for all $\theta \in (-\pi, \pi]$ and some absolute constant $c > 0$,

$$\begin{aligned}|\exp[\Psi_{kn}(x_n e^{i\theta})]| &= \prod_{j=1}^{k} |\exp(x_n e^{i\theta}/2^j) - 1| \exp[-(n-1)\log x_n] \\ &\leqslant \exp[\Psi_{kn}(x_n)] \exp[(\cos\theta - 1)\, x_n/4] \\ &\leqslant \exp[\Psi_{kn}(x_n)] \exp(-cn\theta^2).\end{aligned}$$

Since $n\theta^2 \geqslant n^{1/6}$ for $|\theta| > \theta_o$, we conclude that

$$\int_2 = O(\exp[\Psi_{kn}(x_n)] \exp(-cn^{1/6})) = o\left(\int_1\right). \tag{3.18}$$

By (3.16) and (3.18),

$$(2\pi i)^{-1} \int_{\mathscr{L}} = (1 + o(1))(2\pi n)^{-1/2} \exp[\Psi_{kn}(x_n)]. \tag{3.19}$$

Furthermore, we know that $\Psi_{kn}(n) = F_k(n) - (n-1)\log n$, $\Psi'_{kn}(x_n) = 0$, and $\Psi''_{kn}(x) = O(n^{-1})$ uniformly for x between x_n and n. Also, $x_n - n = O(k)$ and $k = o(n^{1/2})$. As a result, in (3.19), for some x_n^* lying between n and x_n,

$$\begin{aligned}\Psi_{kn}(x_n) &= \Psi_{kn}(n) - \Psi''_{kn}(x_n^*)(n - x_n)^2/2 \\ &= \Psi_{kn}(n) + O(k^2/n) \\ &= F_k(n) - (n-1)\log n + o(1).\end{aligned} \tag{3.20}$$

Combining (3.11), (3.19), (3.20) and the Stirling formula for $n!$, we arrive at

$$E_{nk}^{(1)} = (1 + o(1))\, e^{-n} n F_k(n). \tag{3.21}$$

It remains to derive an approximate formula for $F_k(n)$ or, equivalently, for $L_{nk} = \log F_k(n)$ defined by

$$L_{nk} = \sum_{j=1}^{k} \log(e^{n/2^j} - 1). \tag{3.22}$$

To this end, set $j_n = \lfloor \log_2 n \rfloor$. Denote the partial sums for $1 \leqslant j \leqslant j_n$ and $j_n < j \leqslant k$, respectively, by L_{nk}^* and L_{nk}^{**}. In the first sum $n/2^j \leqslant 1$, so

$$\begin{aligned}L_{nk}^* &= n \sum_{j=1}^{j_n} 2^{-j} + \sum_{j=1}^{j_n} \log(1 - e^{-n/2^j}) \qquad (j = j_n - l) \\ &= n(1 - 2^{-j_n - 1}) + O\left(\left|\sum_{l \geqslant 0} \log(1 - \exp(-2^{l-1}))\right|\right) \\ &= n + O(1) + O\left(\sum_{l \geqslant 0} 2^{-l}\right) = n + O(1).\end{aligned}$$

In the second sum $n/2^j = O(1)$, so

$$\begin{aligned}\log(e^{n/2^j}-1) &= \log[(n/2^j)(e^{n/2^j}-1)/(n/2^j)]\\ &= \log n - j\log 2 + O(n/2^j).\end{aligned}$$

Thus, because $k = \log_2 n + r$, $r \to \infty$, $j_n = \log_2 n + O(1)$,

$$\begin{aligned}L_{nk}^{**} &= (k-j_n)\log n - 2^{-1}(k-j_n)(k+j_n+1)\log 2 + O(n/2^{j_n})\\ &= (k-j_n)[\log n - 2^{-1}\log 2(2\log_2 n + r + O(1))] + O(1)\\ &= -(\log 2)(r^2/2)(1+o(1)).\end{aligned}$$

Therefore,

$$L_{nk} = L_{nk}^{*} + L_{nk}^{**} = n - (\log 2)(r^2/2)(1+o(1)). \tag{3.23}$$

The relations (3.21) and (3.23) prove the lemma completely.

Next we prove

LEMMA 3.2. *If* $k = k_1$, *as defined in Lemma* 3.1, *then*

$$E_{nk}^{(2)} = (1+o(1))[E_{nk}^{(1)}]^2, \qquad n \to \infty. \tag{3.24}$$

Before proving (3.24), let us show how it implies—in conjunction with (3.10)—that for $k = k_1$

$$\lim P(\mathscr{H}_n^{(2)} \geqslant k) = 1. \tag{3.25}$$

Really, $E(Y_{nk}) = E_{nk}^{(1)} \to \infty$ and

$$\begin{aligned}\operatorname{var}(Y_{nk}) &= E(Y_{nk}(Y_{nk}-1)) + E(Y_{nk}) - E^2(Y_{nk})\\ &= E_{nk}^{(2)} + E_{nk}^{(1)} - [E_{nk}^{(1)}]^2 = o([E_{nk}^{(1)}]^2)\\ &= o(E^2(Y_{nk})).\end{aligned}$$

Hence, by Chebyshev's inequality,

$$P(Y_{nk} \geqslant E(Y_{nk})/2) \geqslant 1 - 4\operatorname{var}(Y_{nk})/E^2(Y_{nk}) \to 1, \qquad n \to \infty,$$

and (3.25) follows.

Proof of Lemma 3.2. According to (3.8) and (3.9), (cf. (3.5) and (3.11),

$$E_{nk}^{(2)} = \sum_{s=1}^{k} n!\,(2\pi i)^{-1}\,2^{-s}\int_{\mathscr{L}_s} \exp[\phi_{sn}(x)](x^{-1}\,dx), \tag{3.26}$$

where

$$\phi_{sn}(x)=\sum_{j=1}^{s-1}\log g(x/2^j)+2\sum_{j'=s+1}^{k}\log g(x/2^{j'})-(n-2)\log x \tag{3.27}$$

and $\mathscr{L}_s$ $(1\leqslant s\leqslant n)$ are circular contours around the origin. For each s, let the radius of $\mathscr{L}_s$ be equal to x_{ns}, the positive root of $\phi'_{sn}(x)=0$, i.e.,

$$x\left(\sum_{j=1}^{s-1}2^{-j}+2\sum_{j'=s+1}^{k}2^{-j'}\right)+\sum_{j=1}^{s-1}(x/2^j)(e^{x/2^j}-1)^{-1}$$
$$+2\sum_{j'=s+1}^{k}(x/2^{j'})(e^{x/2^{j'}}-1)^{-1}=n-2.$$

Since

$$\sum_{j=1}^{s-1}2^{-j}+2\sum_{j'=s+1}^{k}2^{-j'}=1-2^{-k+1},\qquad 1\leqslant s\leqslant k,$$

we have then, see (3.13),

$$x_{ns}=n+O(k) \tag{3.28}$$

uniformly for $1\leqslant s\leqslant k$. Consequently,

$$\phi''_{sn}(x_n)=n^{-1}(1+o(1)),\qquad \phi_{sn}(x_n)=\phi_{sn}(n)+o(1). \tag{3.29}$$

Using (3.28), (3.29) and arguing as in the proof of Lemma 3.1, we obtain (cf. (3.19))

$$(2\pi i)^{-1}\int_{\mathscr{L}_s}=(1+o(1))(2\pi n)^{-1/2}\exp[\phi_{sn}(n)],\qquad 1\leqslant s\leqslant k.$$

So, see (3.26), (3.27),

$$E_{nk}^{(2)}=(1+o(1))\sum_{s=1}^{k}b(s,n),$$

where

$$b(s,n)=e^{-n}(n^2/2^s)\,G_s(n), \tag{3.30}$$

and, see (3.9),

$$G_s(n)=\prod_{j=1}^{s-1}g(n/2^j)\prod_{j'=s+1}^{k}g^2(n/2^{j'})\qquad (g(u)=e^u-1). \tag{3.31}$$

We need to show that

$$\lim[E_{nk}^{(1)}]^{-2}\left(\sum_{s=1}^{k} b(s,n)\right)=1. \tag{3.32}$$

For this, note first that by the definition of $F_k(\cdot)$, Lemma 3.1, (3.30), and (3.31),

$$\begin{aligned} b(1,n)&=e^{-n}(n^2/2)\,F_k^2(n)(e^{n/2}-1)^{-2}\\ &=(1+o(1))[e^{-n}nF_k(n)]^2/2=(1+o(1))[E_{nk}^{(1)}]^2/2. \end{aligned} \tag{3.33}$$

Furthermore,

$$\rho(s,n) \underset{\text{def}}{=} b(s,n)/b(s-1,n)=2^{-1}(e^{n/2^s}+1)(e^{n/2^s}-1)^{-1}, \tag{3.34}$$

so that

$$\lim \rho(s,n)=2^{-1}, \qquad s\geqslant 1. \tag{3.35}$$

We also obtain from (3.34) that, for $s_n=\lfloor \log_2(n/\log 4)\rfloor$,

$$\rho(s,n)\begin{cases} \leqslant \frac{5}{6}, & \text{if } 2\leqslant s\leqslant s_n,\\ =O(1), & \text{if } s=s_n+1,\\ \geqslant 1.5, & \text{if } s\geqslant s_n+2. \end{cases}$$

Then, (see (3.33), (3.35)), by the dominated convergence theorem,

$$\begin{aligned} \lim[E_{nk}^{(1)}]^{-2}\sum_{s=1}^{s_n} b(s,n)&=\lim\{b(1,n)/[E_{nk}^{(1)}]^2\}\,b^{-1}(1,n)\sum_{s=1}^{s_n} b(s,n)\\ &=2^{-1}\sum_{t\geqslant 0}2^{-t}=1. \end{aligned} \tag{3.36}$$

On the other hand,

$$\begin{aligned} \sum_{s=s_n+1}^{k} b(s,n)&=O[b(s_n,n)+b(k,n)]\\ &=O[(\tfrac{5}{6})^{s_n}\,b(1,n)+b(k,n)], \end{aligned} \tag{3.37}$$

where, according to (3.30), (3.31), Lemma 3.1, and the fact that $n/2^k=o(1)$,

$$\begin{aligned} b(k,n)&=e^{-n}(n^2/2^k)\prod_{j=1}^{k-1}(e^{n/2^j}-1)\\ &=(1+o(1))\,e^{-n}n\prod_{j=1}^{k}(e^{n/2^j}-1)=(1+o(1))\,E_{nk}^{(1)}. \end{aligned}$$

It follows from (3.36), (3.37), and the last estimate that

$$\lim\,[E_{nk}^{(1)}]^{-2} \sum_{s=s_n+1}^{k} b(s, n) = 0,$$

which together with (3.36) leads to (3.32).

The lemma is proven.

Finally,

3b. *Behavior of the Moments $E(Z_{nk})$ and an Upper Estimate for $\mathscr{H}_n^{(2)}$*

Unlike Y_{nk}, there does not seem to exist a tractable formula for the exponential generating function of $\{E(Z_{nk}): n \geqslant 1\}$. Fortunately, we only have to establish that $E(Z_{nk}) \to 0$ fast enough if $k \geqslant \log_2 \mathrm{n} + (1+\varepsilon)(2\log_2 n)^{1/2}$. For this, it suffices to find a reasonable analytic function whose Taylor coefficients are the upper estimates of $E(Z_{nk})/n!$.

Consider the set of all pairs (μ, ν), where $\mu = (\mu_1, ..., \mu_k)$ and $\nu = (\nu_1, ..., \nu_k)$ are k-tuples of integers such that

$$\begin{aligned} &\mu_1 \geqslant 0, \mu_2 \geqslant 1, ..., \mu_k \geqslant 1, \\ &\nu_1 \geqslant 1, ..., \nu_k \geqslant 1 \qquad \text{and} \qquad \nu_1 + \cdots + \nu_k = n-1. \end{aligned} \tag{3.38}$$

For a given (μ, ν), introduce $P_n(\mu, \nu)$, the probability that the path in the random tree $\mathscr{T}_n^{(1)}$, which leads to the endvertex ω_1, satisfies the following conditions. (i) It has exactly k branching vertices. (ii) The first of them is μ_1 arcs apart from the root, and the path segment between the $(s-1)$th and the sth branching vertices consists of μ_s arcs, $2 \leqslant s \leqslant k$. (iii) The number of the other paths which disengage from our path at the sth branching vertex equals ν_s, $1 \leqslant s \leqslant k$. (The ω_1-path in the figure is such that $k=3$, $\mu_1 = 1$, $\mu_2 = 2$, $\mu_3 = 1$ and $\nu_1 = 2$, $\nu_2 = 1$, $\nu_3 = 1$.) Once we evaluate this probability for all (μ, ν), the expected value $E(Z_{nk})$ can be determined via an obvious relation:

$$E(Z_{nk}) = n \sum_{(\mu,\nu)} P_n(\mu, \nu). \tag{3.39}$$

As for $P_n(\mu, \nu)$, it is given by

$$P_n(\mu, \nu) = \binom{n-1}{\nu_1, ..., \nu_k} \prod_{j=1}^{k} [2^{\mu_j} 2^{-\mu_j(1+\nu_j+\cdots+\nu_k)} 2^{-\nu_{j-1}}] \tag{3.40}$$

$(\nu_o = 0)$. Indeed, the multinomial coefficient is the number of ways to divide the group $\{\omega_2, ..., \omega_n\}$ into k distinct groups of respective sizes $\nu_1, ..., \nu_k$. Further, the first factor in the product is the probability that $\omega_1, ..., \omega_n$ all have a common initial segment of length at least μ_1; the second factor is the probability that the next μ_2 digits of ω_1 will be common for ωs from all the groups, except the first group, whose ωs disagree with ω_1 on the value of the (μ_1+1)th digit. The subsequent factors are interpreted similarly. Rearranging factors in (3.40), we simplify it to

$$P_n(\mu, \nu) = (n-1)! \prod_{j=1}^{k} (2^{-m_j})^{\nu_j}/\nu_j!, \tag{3.41}$$

where

$$m_j = 1 + \mu_1 + \cdots + \mu_j, \qquad 1 \leqslant j \leqslant k.$$

In view of (3.38), $0 < m_1 < \cdots < m_k$; in particular, $m_j \geqslant j$ for $1 \leqslant j \leqslant k$.

A combination of (3.39) and (3.41) yields

$$\begin{aligned}\sum_{n \geqslant 1} E(Z_{nk})\, x^n/n! &= x \sum_m \sum_\nu \prod_{j=1}^{k} (x/2^{m_j})^{\nu_j}/\nu_j! \\ &= x \sum_m \prod_{j=1}^{k} \left[\sum_{\nu_j \geqslant 1} (x/2^{m_j})^{\nu_j}/\nu_j!\right] \\ &= x \sum_m \prod_{j=1}^{k} g(x/2^{m_j}) \qquad (g(u) = e^u - 1).\end{aligned} \tag{3.42}$$

We cannot simplify the last expression any further since the condition $0 < m_1 < \cdots < m_k$ ties m_j together. To get around this obstacle, let us introduce a weaker condition $m_j \geqslant j$, $1 \leqslant j \leqslant k$. Since the Taylor coefficients of $g(u)$ are all nonnegative, the identity (3.42) leads to an inequality

$$\begin{aligned}E(Z_{nk})/n! \leqslant A_{nk} &\underset{\text{def}}{=} \operatorname{coeff}_{x^n}\left[x \sum_{m_1 \geqslant 1, \ldots, m_k \geqslant k} \prod_{j=1}^{k} g(x/2^{m_j})\right] \\ &= \operatorname{coeff}_{x^n}[xU_k(x)],\end{aligned} \tag{3.43}$$

where

$$U_k(x) = \prod_{j=1}^{k} u_j(x), \qquad u_j(x) = \sum_{m \geqslant j} g(x/2^m) \tag{3.44}$$

(compare with (3.5)).

Now we are ready to prove the last lemma.

Lemma 3.3. *Let* $k = \log_2 n + r$, $\log\log n = o(r)$. *Then*

$$E(Z_{nk}) = O(e^{-n} n F_k(n)) = O(n 2^{-(1+o(1)) r^2/2}), \qquad n \to \infty.$$

Therefore

$$\lim E\left(\sum_{k \geqslant k_2} Z_{nk}\right) = 0,$$

if $k_2 = \log_2 n + (1 + \varepsilon_n)(2 \log_2 n)^{1/2}$, $\varepsilon_n \geqslant a > 0$, *and* a *is fixed.*

Proof. Not too surprisingly, we are going to use contour integration again. Significantly—in view of a more complicated function $U_k(x)$—now we can afford some short cuts since we do not need an estimate as sharp as those in Lemmas 3.1 and 3.2.

To begin with, let us choose the circular contour $\mathscr{L}$ of radius *equal* to n (cf. (3.14), (3.28)), and write (see (3.43)),

$$\begin{aligned} A_{nk} &= (2\pi i)^{-1} \int_{\mathscr{L}} x^{-n} U_k(x)\, dx \\ &\leqslant (2\pi n^{n-1})^{-1} \int_{-\pi}^{\pi} |U_k(ne^{i\theta})|\, d\theta. \end{aligned} \tag{3.45}$$

Further, using (3.44) and the inequality (3.17), we have

$$|U_k(ne^{i\theta})| \leqslant \prod_{j=1}^{k} u_{nj}(\theta), \qquad \theta \in (-\pi, \pi], \tag{3.46}$$

where

$$u_{nj}(\theta) = \sum_{m \geqslant j} v_{nm}(\theta), \qquad v_{nm}(\theta) = g(n/2^m) \exp(-c\theta^2 n/2^m) \tag{3.47}$$

and c is a positive constant. A closer study demonstrates that, for each series $u_{nj}(\theta)$, the first term $v_{nj}(\theta)$ is dominant; more precisely, for some constant $\beta > 0$,

$$\prod_{j=1}^{k} u_{nj}(\theta) \leqslant e^{\beta r} \prod_{j=1}^{k} v_{nj}(\theta) \tag{3.48}$$

$(r = k - \log_2 n)$. Consequently, by the definition of $v_{nj}(\theta)$, $F_k(n)$, and (3.46),

$$|U_k(ne^{i\theta})| \leqslant e^{\beta r} F_k(n) \exp[-c\theta^2 n(1 - 2^{-k-1})].$$

So, using (3.45),

$$A_{nk} = O(e^{\beta r} n^{-n+1/2} F_k(n)), \qquad n \to \infty,$$

and, (see (3.23), (3.43)),

$$\begin{aligned} E(Z_{nk}) &= O(e^{\beta r} n!\, n^{-n+1/2} F_k(n)) = O(e^{\beta r} e^{-n} n F_k(n)) \\ &= O(n 2^{-(1+o(1)) r^2/2}), \qquad n \to \infty. \end{aligned}$$

Thus, it remains only to prove (3.48).

By (3.47) and the definition of $g(\cdot)$, the ratio of two consecutive terms in a series $u_{nj}(\theta)$ is given by

$$w_{nm}(\theta) = v_{nm}(\theta)/v_{n,m-1}(\theta) = \eta_{nm}^{c\theta^2}(\eta_{nm}+1)^{-1}, \qquad \eta_{nm} = e^{n/2^m}.$$

Simple calculus shows that $\eta^{\alpha}(\eta+1)^{-1}$ is a decreasing function of $\eta \in [0, 1)$, provided that $\alpha \leqslant \frac{1}{2}$. Since we may and do assume that $c\pi^2 \leqslant \frac{1}{2}$, $w_{nm}(\theta)$ increases with m, whence

$$w_{nm}(\theta) \leqslant \lim_{\mu \to \infty} w_{n\mu}(\theta) = \tfrac{1}{2}, \qquad \forall n, m, \theta. \tag{3.49}$$

To estimate $u_{nj}(\theta)$, consider separately $1 \leqslant j \leqslant m_n$ and $m_n < j \leqslant k$, where $m_n = \lfloor \log_2 n - \log_2 \log n \rfloor$.

(i) $1 \leqslant j \leqslant m_n$.

Since $n/2^{m_n} \geqslant \log n$, we have

$$w_{nm}(\theta) \leqslant w_{nm_n}(\theta) \leqslant n^{-(1-c\theta^2)} \leqslant n^{-1/2}, \qquad \text{if} \quad m \leqslant m_n,$$
$$w_{n,m_n+1}(\theta) \leqslant n^{-(1-c\theta^2)/2} \leqslant n^{-1/4}.$$

It follows that, see also (3.49),

$$\begin{aligned} u_{nj}(\theta) &= \sum_{m=j}^{m_n} v_{nm}(\theta) + \sum_{m>m_n} v_{nm}(\theta) \\ &= (1+O(n^{-1/2}))\, v_{nj}(\theta) + O(n^{-1/4} v_{nj}(\theta)) \\ &= (1+O(n^{-1/4}))\, v_{nj}(\theta). \end{aligned} \tag{3.50}$$

(ii) $m_n < j \leqslant k$.

Invoking again (3.49), we have trivially

$$u_{nj}(\theta) = O(v_{nj}(\theta)). \tag{3.51}$$

Putting together (3.50) and (3.51) and remembering that $k - m_n = O(\log\log n + r) = O(r)$, we arrive at

$$\begin{aligned} \prod_{j=1}^{k} u_{nj}(\theta) &\leqslant e^{\alpha r}(1+O(m_n n^{-1/4})) \prod_{j=1}^{k} v_{nj}(\theta) \\ &\leqslant e^{\beta r} \prod_{j=1}^{k} v_{nj}(\theta), \end{aligned}$$

where $\beta > \alpha$ are constants.

The lemma is completely proven.

Since $\mathscr{H}_n^{(2)} \geqslant k$ iff $\sum_{t \geqslant k} Z_{nt} > 0$, it immediately follows from this lemma that, for every $\varepsilon > 0$,

$$\lim P(\mathscr{H}_n^{(2)} \geqslant \log_2 n + (1+\varepsilon)(2 \log_2 n)^{1/2}) = 0.$$

Together with (3.25), the last relation completes the proof of the theorem.

Final Remark. The proof of a more general relation

$$\mathscr{H}_{nd}^{(2)} = \log_2 n + (1 + o_p(1))(2d^{-1} \log_2 n)^{1/2}, \qquad n \to \infty,$$

goes along the same lines, with $g(u) = e^u - 1$ replaced by $g_d(u) = e^u - \sum_{0 \leqslant j \leqslant d-1} u^j/j!$.

Acknowledgments

We are very grateful to a referee and to Joe Rosenblatt for helping us to improve the presentation of these results. We also thank the Managing Editors for consideration.

References

1. D. Aldous and P. Shields, A diffusion limit for a class of randomly-growing binary trees, *Probab. Th. Rel. Fields* **79** (1988), 509–542.
2. L. Devroye, A probabilistic analysis of the height of tries and of the complexity of trie sort, *Acta Inform.* **21** (1984), 229–237.
3. D. E. Knuth, "The Art of Computer Programming Vol. 3: Sorting and Searching," Addison–Wesley, Reading, MA, 1973.
4. H. Mendelson, Analysis of extendible hashing, *IEEE Trans. Software Engrg.* **8** (1982), 611–619.
5. B. Pittel, Asymptotical growth of a class of random trees, *Ann. Probab.* **13** (1985), 414–427.
6. B. Pittel, Paths in a random digital tree: Limiting distributions, *Adv. Appl. Prob.* **18** (1986), 139–155.
7. A. Rényi, On random subsets of a finite set, *Mathematica (Cluj)* **3** (1961), 355–362.