Algorithmica (2006) 46: 367–407
DOI: 10.1007/s00453-006-0109-5

Algorithmica

Profiles of Random Trees: Limit Theorems for Random Recursive Trees and Binary Search Trees[1]

Michael Fuchs,[2] Hsien-Kuei Hwang,[3] and Ralph Neininger[4]

Abstract. We prove convergence in distribution for the profile (the number of nodes at each level), normalized by its mean, of random recursive trees when the limit ratio α of the level and the logarithm of tree size lies in $[0, e)$. Convergence of all moments is shown to hold only for $\alpha \in [0, 1]$ (with only convergence of finite moments when $\alpha \in (1, e)$). When the limit ratio is 0 or 1 for which the limit laws are both constant, we prove asymptotic normality for $\alpha = 0$ and a "quicksort type" limit law for $\alpha = 1$, the latter case having additionally a small range where there is no fixed limit law. Our tools are based on the contraction method and method of moments. Similar phenomena also hold for other classes of trees; we apply our tools to binary search trees and give a complete characterization of the profile. The profiles of these random trees represent concrete examples for which the range of convergence in distribution differs from that of convergence of all moments.

Key Words. Random recursive tree, Random binary search tree, Profile of trees, Probabilistic analysis of algorithms, Weak convergence.

1. Introduction. The profile or height profile of a tree is the sequence of numbers whose kth element enumerates the number of nodes at distance k from the root of the tree (or the number of descendants in the kth generation in branching process terms). Profiles of trees are fine shape characteristics encountered in diverse problems such as breadth-first search, data compression algorithms (Jacquet et al., 2001), random generation of trees (Devroye and Robson, 1995), and the levelwise analysis of quicksort (Chern and Hwang, 2001b; Evans and Dunbar, 1982). In addition to their interest in applications and connections to many other shape parameters, we show, through recursive trees and binary search trees, that profiles of random trees having roughly logarithmic height are a rich source of many intriguing phenomena. The high concentration of nodes at certain (log) levels results in the asymptotic bimodality for the variance, as already demonstrated in Drmota and Hwang (2005a); our purpose of this paper is to unveil and clarify the diverse phenomena exhibited by the limit distributions of the profiles of random recursive trees and binary search trees. The tools we use, as well as the results we derive, are of some generality.

---

[1] Michael Fuchs was partially supported by the National Science Council of ROC under Grant NSC-93-2119-M-009-003. Hsien-Kuei Hwang was partially supported by a research award of the Alexander von Humboldt Foundation and by the National Science Council under Grant NSC-92-2118-M-001-019. Ralph Neininger's research was supported by an Emmy Noether Fellowship of the DFG.
[2] Department of Applied Mathematics, National Chiao Tung University, Hsinchu 300, Taiwan. mfuchs@math.nctu.edu.tw.
[3] Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan. hkhwang@stat.sinica.edu.tw.
[4] Department of Mathematics, J.W. Goethe University, 60325 Frankfurt a.M., Germany. neiningr@math.uni-frankfurt.de.

Received January 13, 2005; revised January 20, 2006. Communicated by P. Jacquet, D. Panario, and W. Szpankowski. Online publication December 13, 2006.

Recursive trees. Recursive trees have been introduced as simple probability models for system generation (Na and Rapoport, 1970), spread of contamination of organisms (Meir and Moon, 1974), pyramid scheme (Gastwirth and Bhattacharya, 1984; Smythe and Mahmoud, 1995), stemma construction of philology (Najock and Heyde, 1982), Internet interface map (Janic et al., 2002), and stochastic growth of networks (Chan et al., 2003). They are related to some Internet models (van Mieghem et al., 2001; van der Hofstad et al., 2002; Devroye et al., 2002) and some physical models (Tetzlaff, 2002); they also appeared in Hopf algebra under the name "heap-ordered trees"; see Grossman and Larson (1989). The bijection between recursive trees and binary search trees not only makes the former a flexible representation of the latter but also provides a rich direction for further extensions; see for example Mahmoud and Smythe (1995).

A simple way of constructing a random recursive tree of n nodes is as follows. One starts from a root node with the label 1; at stage i ($i = 2, \dots, n$) a new node with label i is attached uniformly at random to one of the previous nodes $(1, \dots, i-1)$. The process stops after node n is inserted. By construction, the labels of the nodes along any path from the root to a node form an increasing sequence; see Figure 2 for a recursive tree of ten nodes. For a survey of probabilistic properties of recursive trees, see Smythe and Mahmoud (1995).

Known results for the profile of recursive trees. Let $X_{n,k}$ denote the number of nodes at level k in a random recursive tree of n nodes, where $X_{n,0} = 1$ (the root) for $n \geq 1$. Then $X_{n,k}$ satisfies (see van der Hofstad et al., 2002)

$$X_{n,k} \stackrel{d}{=} X_{I_n,k-1} + X^*_{n-I_n,k}, \tag{1}$$

for $n, k \geq 1$ with $X_{n,0} = 1 - \delta_{n,0}$ ($\delta_{n,0}$ being Kronecker's symbol), where $(X_{n,k})$, $(X^*_{n,k})$ and (I_n) are independent, $X_{n,k} \stackrel{d}{=} X^*_{n,k}$, and I_n is uniformly distributed over $\{1, \dots, n-1\}$.

Meir and Moon (1978) showed (implicitly) that

$$\mu_{n,k} := \mathbb{E}(X_{n,k}) = \frac{\mathsf{s}(n, k+1)}{(n-1)!} \qquad (0 \leq k < n), \tag{2}$$

where $\mathsf{s}(n, k)$ denotes the unsigned Stirling numbers of the first kind; see also Moon (1974) and Dondajewski and Szymański (1982). By the approximations given in Hwang (1995), we then have

$$\mu_{n,k} = \frac{\lambda_n^k}{\Gamma(1 + \alpha_{n,k}) k!}(1 + O(\lambda_n^{-1})), \tag{3}$$

uniformly for $1 \leq k \leq K\lambda_n$, for any $K > 1$, where, *here and throughout this paper*,

$$\lambda_n := \max\{\log n, 1\}, \qquad \alpha_{n,k} := k/\lambda_n,$$

and Γ denotes the Gamma function. This approximation implies, in particular, a local limit theorem for the depth (distance of a random node to the root); see Devroye (1998), Szymański (1990), Mahmoud (1991).

The second moment is also implicit in Meir and Moon (1978):

$$\mathbb{E}(X_{n,k}^2) = \sum_{0 \leq j \leq k} \binom{2j}{j} \frac{\mathsf{s}(n, k+j+1)}{(n-1)!};$$

see also van der Hofstad et al. (2002). Precise asymptotic approximations for the variance $\mathbb{V}(X_{n,k})$ were derived in Drmota and Hwang (2005a) for all ranges of k. In particular, the variance is asymptotically of the same order as $\mu_{n,k}^2$ when $\alpha \in (0, 2)$ except $k \sim \lambda_n$ (where the profile variance exhibits a bimodal behavior).

Limit distribution when $0 \leq \alpha < e$. From the asymptotic estimate (3), we have

$$\frac{\log \mu_{n,k}}{\lambda_n} \to \alpha - \alpha \log \alpha,$$

where *here and throughout this paper* $k = k(n)$ and $\alpha := \lim_{n\to\infty} k(n)/\lambda_n$. Thus $\mu_{n,k} \to \infty$ when $\alpha < e$. Note that the expected height (length of the longest path from the root) of random recursive trees is asymptotic to $e\lambda_n$; see Devroye (1987) or Pittel (1994).

Define a class of random variables $X(\alpha)$ by the fixed-point equation

$$X(\alpha) \stackrel{d}{=} \alpha U^{\alpha} X(\alpha) + (1-U)^{\alpha} X(\alpha)^*, \tag{4}$$

with $\mathbb{E}(X(\alpha)) = 1$, where $X(\alpha)$, $X(\alpha)^*$, and U are independent, $X(\alpha)^* \stackrel{d}{=} X(\alpha)$, and U is uniformly distributed in the unit interval; see Proposition 1 for the existence and properties of $X(\alpha)$. Define $X(0) = 1$.

THEOREM 1.

(i) *If* $0 \leq \alpha < e$, *then*

$$\frac{X_{n,k}}{\mu_{n,k}} \xrightarrow{d} X(\alpha), \tag{5}$$

where $\xrightarrow{d}$ *denotes convergence in distribution*.

(ii) *If* $0 \leq \alpha < m^{1/(m-1)}$, *where* $m \geq 2$, *then* $X_{n,k}/\mu_{n,k}$ *converges to* $X(\alpha)$ *with convergence of the first* m *moments but not the* $(m+1)$st *moment*.

In particular, convergence of the second moment holds for $0 \leq \alpha < 2$.

COROLLARY 1. *If* $0 \leq \alpha < 2$, *then*

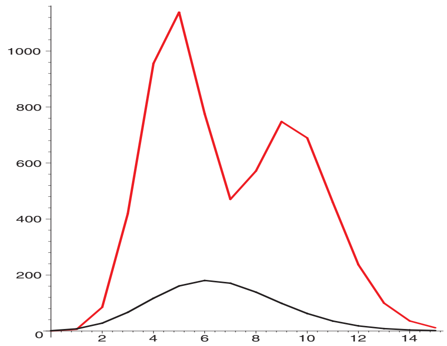
$$\mathbb{V}(X_{n,k}) \sim \left(\frac{\Gamma(\alpha+1)^2}{(1-\alpha/2)\Gamma(2\alpha+1)} - 1\right) \mu_{n,k}^2.$$

Note that the coefficient on the right-hand side becomes zero when $\alpha = 0$ and $\alpha = 1$, and the variance indeed exhibits a *bimodal behavior* when $\alpha = 1$; see Figure 1 for a plot and Drmota and Hwang (2005a) or below for more precise approximations to the variance.

Since $m^{1/(m-1)} \downarrow 1$, the unit interval is the only range where convergence of all moments holds.

COROLLARY 2. *If* $0 \leq \alpha \leq 1$, *then*

$$\frac{X_{n,k}}{\mu_{n,k}} \xrightarrow{\mathrm{M}} X(\alpha), \tag{6}$$



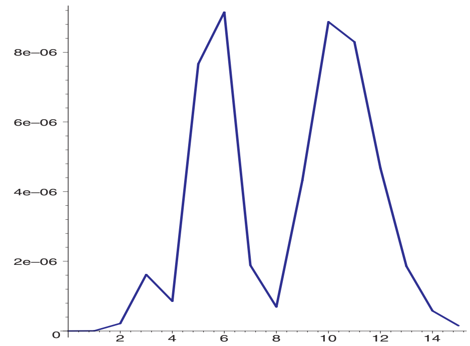


Fig. 1. A plot of $\mathbb{E}(X_{n,k})$ (the unimodal curve), $\mathbb{V}(X_{n,k})$ (the bimodal curve with higher valley), and $|\mathbb{E}(X_{n,k} - \mu_{n,k})^3|$ (right) of the number $X_{n,k}$ of nodes at level k in random recursive trees of $n = 1100$ nodes, all normalized by their maximum values. Note that the valley of $|\mathbb{E}(X_{1100,k} - \mu_{1100,k})^3|$ (when normalized by n^3) is deeper than that of $\mathbb{V}(X_{1100,k})$ (normalized by n^2); see Corollary 5 for the general description.

where $\stackrel{\mathrm{M}}{\longrightarrow}$ *denotes convergence of all moments. Convergence of all moments fails for* $1 < \alpha < e$.

Thus the profile of random recursive trees represents a concrete example for which *the range of convergence in distribution is different from that of convergence of all moments*. We will show that such a property also holds for random binary search trees; it is expected to hold for other trees like ordered (or plane) recursive trees and m-ary search trees, but the technicalities are expected to be much more complicated. We focus at this stage on new phenomena and their proofs, not on generality.

The proof of (5) relies on the contraction method developed in Neininger and Rüschendorf (2004) (see also the survey paper by Rösler and Rüschendorf (2001)), and the moment convergence $X_{n,k}/\mu_{n,k}$ uses the method of moments. Both methods are technically more involved because we deal with recurrences with two parameters. We will indeed prove a stronger approximation to (5) by deriving a rate under the Zolotarev metric (see Zolotarev, 1976).

However, why $m^{1/(m-1)}$? This is readily seen by the recurrence of the moments $\nu_m(\alpha) := \mathbb{E}(X(\alpha)^m)$ of $X(\alpha)$:

$$\nu_m(\alpha) = \frac{1}{m - \alpha^{m-1}} \sum_{1 \le h < m} \binom{m}{h} \nu_h(\alpha)\nu_{m-h}(\alpha)\alpha^{h-1} \times \frac{\Gamma(h\alpha + 1)\Gamma((m-h)\alpha + 1)}{\Gamma(m\alpha + 1)} \qquad (m \ge 2), \tag{7}$$

where $\nu_0(\alpha) = \nu_1(\alpha) = 1$. This recurrence is well defined for $\nu_m(\alpha)$ when $\alpha < m^{1/(m-1)}$. This explains the special sequence $m^{1/(m-1)}$.

Note that since $\mathbb{E}(X(\alpha)^m) = \infty$ for $\alpha \ge m^{1/(m-1)}$, we have $\mathbb{E}(X_{n,k}/\mu_{n,k})^m \to \infty$ in that range.

A "quicksort-type" limit distribution when $\alpha = 1$. Since $X(1) = 1$, we can refine the limit result (5) for $\alpha = 1$ as follows.

THEOREM 2.

(i) *If* $k = \lambda_n + t_{n,k}$, *where* $|t_{n,k}| \to \infty$ *and* $t_{n,k} = o(\lambda_n)$, *then*

$$\frac{X_{n,k} - \mu_{n,k}}{t_{n,k}\lambda_n^{k-1}/k!} \xrightarrow{\mathrm{M}} X'(1), \tag{8}$$

where $X'(1) := (d/d\alpha)X(\alpha)|_{\alpha=1}$ *satisfies*

$$X'(1) \stackrel{d}{=} UX'(1) + (1-U)X'(1)^* + U + U\log U + (1-U)\log(1-U),$$

with $X'(1)$, $X'(1)^*$, *and* U *independent and* $X'(1) \stackrel{d}{=} X'(1)^*$.

(ii) *If* $k = \lambda_n + O(1)$, *then the sequence of random variables* $(X_{n,k} - \mu_{n,k})/\sqrt{\mathbb{V}(X_{n,k})}$ *does not converge to a fixed law.*

Although (8) can also be proved by the contraction method, we prove both results of the theorem by the method of moments because the proof for the non-convergence part is readily modified from that for (8); see also Chern et al. (2002) for more examples having no convergence to fixed limit law. On the other hand, since the distribution of $X'(1)$ is uniquely characterized by its moment sequence (see (41)), we have the convergence in distribution as follows.

COROLLARY 3. *If* $k = \lambda_n + t_{n,k}$, *where* $|t_{n,k}| \to \infty$ *and* $t_{n,k} = o(\lambda_n)$, *then*

$$\frac{X_{n,k} - \mu_{n,k}}{t_{n,k}\lambda_n^{k-1}/k!} \xrightarrow{d} X'(1).$$

The same limit law $X'(1)$ also appeared in the total path length (which is $\sum_k kX_{n,k}$) of recursive trees (see Dobrow and Fill, 1999), or essentially the total left path length of random binary search trees, and the cost of an in-situ permutation algorithm; see Hwang and Neininger (2002).

The appearance of the same limit law as the total path length is not a coincidence. *Intuitively*, almost all nodes lie at the levels $k = \lambda_n + O(\sqrt{\lambda_n})$ (since $\mathbb{E}(X_{n,k}) \asymp n/\sqrt{\lambda_n}$ by (3)) and it is these nodes that contribute predominantly to the total path length; see also (9) below for an estimate of the variance. *Analytically*, a deeper connection between the profile and the total path length is seen through the level polynomials $\sum_k X_{n,k}z^k$ (properly normalized) for which we can derive, following Chauvin et al. (2001), an almost sure convergence to some (complex-valued) limit random variable. From such a uniform convergence, the profile is quickly linked to the total path length by taking the derivative of the normalized level polynomial with respect to z and substituting $z = 1$. Indeed, limit theorems for weighted path-lengths of the form $\sum_k k^m X_{n,k}$, as well as the width ($\max_k X_{n,k}$), can be obtained as by-products. These and finer results on correlations and expected width are discussed in Drmota and Hwang (2005b).

Asymptotics of the variance. As a consequence of our convergence of all moments, we have the following estimate for the variance.

COROLLARY 4. *If $k = \lambda_n + t_{n,k}$, where $t_{n,k} = o(\lambda_n)$, then the variance of $X_{n,k}$ satisfies*

$$\mathbb{V}(X_{n,k}) \sim p_2(t_{n,k}) \left(\frac{\lambda_n^{k-1}}{k!}\right)^2, \tag{9}$$

where $p_2(t_{n,k}) := c_2 t_{n,k}^2 + 2c_1 t_{n,k} + c_0$ with

$$\begin{aligned} c_2 &:= 2 - \frac{\pi^2}{6}, \qquad c_1 := c_2(1-\gamma) - \zeta(3) + 1, \\ c_0 &:= c_2(\gamma^2 - 2\gamma + 3) - 2(\zeta(3) - 1)(1-\gamma) - \frac{\pi^4}{360}. \end{aligned} \tag{10}$$

Here γ denotes Euler's constant and $\zeta(3) := \sum_{j\ge 1} j^{-3}$.

Expression (9) explains the valley for the variance in Figure 1. Note that $\mathbb{V}(X_{n,k})/\mu_{n,k}^2 = O(t_{n,k}^2/\lambda_n^2)$ when $t_{n,k} = o(\lambda_n)$.

Our proof indeed yields the following extremal orders of $|\mathbb{E}(X_{n,k} - \mu_{n,k})^m|$ for $m \ge 2$.

COROLLARY 5. *The absolute value of the* mth *central moment satisfies*

$$\begin{aligned} \max_{0\le k<n} |\mathbb{E}(X_{n,k} - \mu_{n,k})^m| &\asymp \lambda_n^{-m} n^m, \\ \min_{|k-\lambda_n| = O(\sqrt{\lambda_n})} |\mathbb{E}(X_{n,k} - \mu_{n,k})^m| &\asymp \lambda_n^{-3m/2} n^m, \end{aligned}$$

where the maximum is achieved at $k = \lambda_n \pm \sqrt{\lambda_n}(1 + o(1))$ and the minimum at $k = \lambda_n + O(1)$.

More refined results can be derived as in Drmota and Hwang (2005a). For example, by (40) below, we have

$$\max_{0\le k<n} \left|\mathbb{E}(X_{n,k} - \mu_{n,k})^m\right| \sim |\mathbb{E}(X'(1)^m)| e^{-m/2} \left(\frac{n}{\sqrt{2\pi}\lambda_n}\right)^m,$$

for $m \ge 2$, where $\mathbb{E}(X'(1)^m)$ can be computed recursively; see (41).

Asymptotic normality when $\alpha = 0$. The profile $X_{n,k}$ in the remaining range $1 \le k = o(\lambda_n)$ will be shown to be asymptotically normally distributed. It is known (see Bergeron et al., 1992) that the out-degree of the root $X_{n,1}$ satisfies

$$\mathbb{P}(X_{n,1} = j) = \frac{\mathsf{s}(n-1, j)}{(n-1)!} \qquad (1 \le j < n);$$

thus $X_{n,1}$ is asymptotically normal with the mean and variance both asymptotic to λ_n. Equivalently, $X_{n,1}$ is the number of nodes on the rightmost branch (the path starting from the root and always going right until reaching an external node) in a random binary search tree of $n-1$ nodes; see the transformation below for more information.

Let $\Phi(x) := (2\pi)^{-1/2}\int_{-\infty}^{x} e^{-t^2/2}\,\mathrm{d}t$ denote the distribution function of the standard normal distribution.

THEOREM 3. *The distribution of the profile $X_{n,k}$ satisfies*

$$\sup_x \left|\mathbb{P}\left(\frac{X_{n,k}-\lambda_n^k/k!}{\lambda_n^{k-1/2}/\sqrt{(k-1)!^2\,(2k-1)}} < x\right) - \Phi(x)\right| = O\left(\sqrt{\frac{k}{\lambda_n}}\right), \tag{11}$$

uniformly for $1 \le k = o(\lambda_n)$, with the mean and variance asymptotic to

$$\begin{cases} \mathbb{E}(X_{n,k}) \sim \dfrac{\lambda_n^k}{k!}, \\ \mathbb{V}(X_{n,k}) \sim \dfrac{\lambda_n^{2k-1}}{(k-1)!^2\,(2k-1)}. \end{cases}$$

In particular, $X_{n,2}$ is asymptotically normally distributed with mean asymptotic to $\frac{1}{2}\lambda_n^2$ and variance to $\frac{1}{3}\lambda_n^3$. A similar central limit theorem appeared in the logarithmic order of a random element in symmetric groups; see Erdős and Turán (1967).

Unlike previous cases, the proof of this result is based on a polynomial decomposition of the associated generating functions using characteristic functions and singularity analysis (see Flajolet and Odlyzko, 1990), the reasons being (i) this method leads to the optimal Berry–Esseen bound (11), which is not obvious by the method of moments, (ii) it is of independent methodological interests, and (iii) it can also be applied to give an alternative proof of (6).

The asymptotic normality of $X_{n,k}$ when $\alpha = 0$ indicates that nodes are generated in a very regular way in recursive trees, at least for the first $o(\lambda_n)$ levels. The rough picture here is that each node at these levels "attracts" about λ_n/k new-coming nodes, as is obvious from (3); see also Drmota and Hwang (2005b) for an asymptotic independence property for the number of nodes at two different levels, both being $o(\lambda_n)$ away from the root.

Profiles of random binary search trees. Binary search trees are one of the most studied fundamental data structures in Computer Algorithms. They have also been introduced in other fields under different forms; see Drmota and Hwang (2005a) for more references.

This tree model is characterized by a recursive splitting process in which $n \ge 2$ distinct labels are split into a root and two subtrees formed recursively by the same procedure (one may be empty) of sizes J_n and $n-1-J_n$, where J_n is uniformly distributed in $\{0, 1, \ldots, n-1\}$. Such a model is isomorphic to *binary increasing trees* in which a sequence of $n \ge 2$ continuous random variables (independent and identically distributed) is split into a root with the smallest label and two subtrees formed recursively by the same splitting process corresponding to the subsequences to the left and right respectively of the smallest label. Note that when given a random permutation of n elements the size of the left subtree of the binary increasing tree constructed from the permutation equals

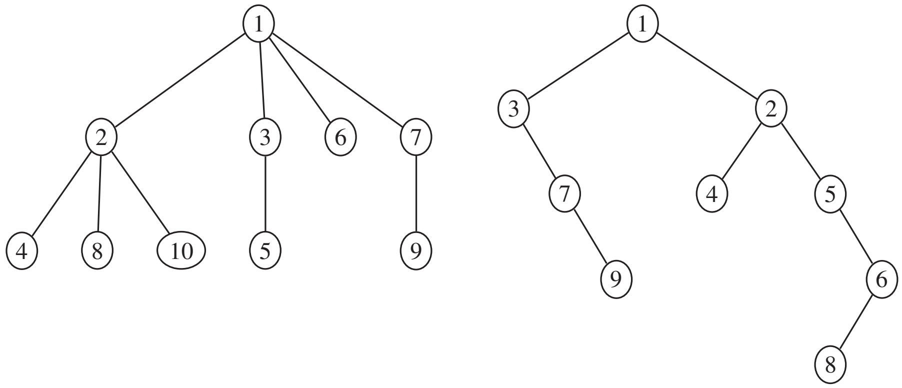


Fig. 2. A recursive tree of ten nodes and its corresponding transformed binary increasing tree of nine nodes.

j, $0 \le j \le n-1$, with equal probability $1/n$, the same as in random binary search trees.

A recursive tree can be transformed into a binary increasing tree by the well-known procedure (referred to as the *natural correspondence* in Kunth (1997) and the *rotation correspondence* by others): drop first the root and arrange all subtrees from left to right in increasing order of their root labels; sibling relations are transformed into right branches (of the leftmost node in that generation) and the leftmost branches remain unchanged; a final relabeling (using labels from 1 to $n-1$) of nodes then yields a binary increasing tree of $n-1$ nodes. Such a transformation is invertible; see Figure 2.

Under this transformation, the profile $X_{n,k}$ in recursive trees becomes essentially the number of nodes in random binary search trees of $n-1$ nodes with left-distance $k-1$ ($k \ge 1$), the *left-distance* of a node being the number of left-branches needed to traverse from the root to that node. This also explains the recurrence (1).

Known and new results for profiles of random binary search trees. We distinguish two types of nodes for binary search trees: external nodes $Y_{n,k}$ (virtual nodes completed so that all nodes are of out-degree either 0 or 2) and internal nodes $Z_{n,k}$ (nodes holding labels). Chauvin et al. (2001) established *almost sure convergence* for $Y_{n,k}/\mathbb{E}(Y_{n,k})$ and $Z_{n,k}/\mathbb{E}(Z_{n,k})$ when $1.2 \le \alpha \le 2.8$, and recently Chauvin et al. (2005) extended the range for $Y_{n,k}/\mathbb{E}(Y_{n,k})$ to the optimal range $\alpha_- < \alpha < \alpha_+$, the two numbers $\alpha_- \approx 0.37$, $\alpha_+ \approx 4.31$ being the fill-up and height constants (of binary search trees), namely, $0 < \alpha_- < 1 < \alpha_+$ solving the equation $e^{(z-1)/z} = z/2$; see also Chauvin and Rouault (2004). For other known results on the profiles $Y_{n,k}$, see Drmota and Hwang (2005a) and the references therein.

Our tools for recursive trees also apply to binary search trees. Briefly, we derive convergence in distribution for $Y_{n,k}/\mathbb{E}(Y_{n,k})$ and $Z_{n,k}/\mathbb{E}(Z_{n,k})$ in the range $\alpha \in (\alpha_-, \alpha_+)$ and convergence of all moments for $\alpha \in [1, 2]$, the degenerate cases $\alpha = 1, 2$ being further refined by more explicit limit laws; see Section 7 for details.

While it is expected that the profiles for both types of nodes have similar behaviors to $X_{n,k}$, we derive finer results showing more delicate structural difference between internal nodes and external nodes.

Organization of the paper. Since most of our asymptotic approximations are based on the solution (exact or asymptotic) of the underlying double-indexed recurrence (in n and k), we start by solving the recurrence in the next section. The proof of the convergence in distribution (5) of $X_{n,k}/\mu_{n,k}$ when $0 < \alpha < e$ by the contraction method is given in Section 3. Then we prove the moment convergence part of Theorem 1 in Section 4 and Theorem 2 in Section 5. The asymptotic normality when $\alpha = 0$ is proved in Section 6, where an alternative proof of (6) is also indicated. Our methods of proof can be easily amended for binary search trees, and the results are given in Section 7. We conclude this paper with a few questions.

Notations. Throughout this paper $\lambda_n := \max\{\log n, 1\}$, $\alpha_{n,k} := k/\lambda_n$ and $\alpha := \lim_{n\to\infty} \alpha_{n,k}$ when the limit exists. The symbol $[z^n]f(z)$ stands for the coefficient of z^n in the Taylor expansion of $f(z)$. The generic symbols ε and K always represent sufficiently small and large, respectively, positive constants whose values may vary from one occurrence to another. Finally, U represents a uniform [0, 1] random variable.

2. The Double-Indexed Recurrence and Asymptotic Transfer. Since all moments (centered or not) satisfy the same recurrence, we derive in this section the exact solution and study a simple type of asymptotic transfer (relating the asymptotics of the recurrence to that of the non-homogeneous part) for such a recurrence.

By (1), we have the recurrence for the probability generating functions $P_{n,k}(y) := \mathbb{E}(y^{X_{n,k}})$

$$P_{n,k}(y) = \frac{1}{n-1} \sum_{1\le j<n} P_{j,k-1}(y) P_{n-j,k}(y) \qquad (n \ge 2;\, k \ge 1), \tag{12}$$

with $P_{n,0}(y) = y$ for $n \ge 1$ and $P_{0,k}(y) = 1$.

Recurrence of factorial moments. Let

$$A_{n,k}^{(m)} := \mathbb{E}(X_{n,k}(X_{n,k}-1)\cdots(X_{n,k}-m+1)) = P_{n,k}^{(m)}(1).$$

Then $A_{n,k}^{(0)} = 1$ for $n, k \ge 0$. By (12), we have the recurrence

$$A_{n,k}^{(m)} = \frac{1}{n-1} \sum_{1\le j<n} \left(A_{j,k-1}^{(m)} + A_{j,k}^{(m)}\right) + B_{n,k}^{(m)} \qquad (n \ge 2;\, k, m \ge 1),$$

where

$$B_{n,k}^{(m)} = \sum_{1\le h<m} \binom{m}{h} \frac{1}{n-1} \sum_{1\le j<n} A_{j,k-1}^{(h)} A_{n-j,k}^{(m-h)}, \tag{13}$$

with the boundary conditions $A_{n,0}^{(1)} = 1$ for $n \ge 1$ and $A_{n,0}^{(m)}(0) = 0$ for $m \ge 2$ and $n \ge 1$.

Exact solution of the recurrence. Consider a recurrence of the form

$$a_{n,k} = \frac{1}{n-1} \sum_{1\le j<n} \left(a_{j,k} + a_{j,k-1}\right) + b_{n,k} \qquad (n \ge 2;\, k \ge 1), \tag{14}$$

with $a_{1,k}$ and $b_{n,k}$ given. We assume, without loss of generality, that $a_{0,k} = 0$ (otherwise, we need only to modify the values of $a_{1,k}$ and $b_{n,k}$).

LEMMA 1. *For $n \geq 1$ and $k \geq 0$,*

$$a_{n,k} = b_{n,k} + \sum_{1\leq j<n} \sum_{0\leq r\leq k} \frac{b_{j,k-r}}{j} [u^r](u+1) \prod_{j<\ell<n} \left(1+\frac{u}{\ell}\right), \tag{15}$$

where $b_{1,k} := a_{1,k}$.

PROOF. Let $a_n(u) := \sum_k a_{n+1,k} u^k$ and $b_n(u) := \sum_k b_{n+1,k} u^k$. Then $a_n(u)$ satisfies the recurrence

$$a_n(u) = \frac{1+u}{n} \sum_{0\leq j<n} a_j(u) + b_n(u) \qquad (n \geq 1),$$

with the initial condition $a_0(u) = \sum_k a_{1,k} u^k$. By taking the difference $na_n(u) - (n-1)a_{n-1}(u)$, we obtain

$$a_n(u) = \left(1+\frac{u}{n}\right) a_{n-1}(u) + b_n(u) - \frac{n-1}{n} b_{n-1}(u) \qquad (n \geq 2).$$

Solving this linear recurrence yields

$$a_n(u) = b_n(u) + (1+u) \sum_{0\leq j<n} \frac{b_j(u)}{j+1} \prod_{j+2\leq \ell\leq n} \left(1+\frac{u}{\ell}\right) \qquad (n \geq 1)$$

(since $b_0(u) := a_0(u)$). Taking the coefficient of u^k on both sides leads to (15). □

Mean value. Applying (15) with $b_{n,k} = \delta_{n,1}\delta_{0,k}$, we obtain, for $n \geq 1$ and $k \geq 0$,

$$\begin{aligned} \mu_{n,k} &= [u^k] \prod_{1\leq \ell<n} \left(1+\frac{u}{\ell}\right) \\ &= \frac{\mathsf{s}(n,k+1)}{(n-1)!}. \end{aligned} \tag{16}$$

This rederives (2).

A uniform estimate for the expected profile. For later use, we derive a uniform bound for $\mu_{n,k}$.

LEMMA 2. *The mean satisfies*

$$\mu_{n,k} = O((v\lambda_n)^{-1/2} v^{-k} n^v), \tag{17}$$

uniformly for $1 \leq k < n$, where $0 < v = O(1)$.

PROOF. Note that by (16), we have the obvious inequality

$$\mu_{n,k} v^k \le \prod_{1\le \ell < n} \left(1 + \frac{v}{\ell}\right) \qquad (v > 0),$$

which leads to $\mu_{n,k} = O(v^{-k} n^v)$ for $1 \le k < n$. However, this is too crude for our purpose.

By Cauchy's integral formula,

$$\begin{aligned} \mu_{n,k} &\le \frac{v^{-k}}{2\pi} \int_{-\pi}^{\pi} \prod_{1\le \ell\le n} \left|1 + \frac{ve^{it}}{\ell}\right| \, \mathrm{d}t \\ &\le \frac{v^{-k}}{2\pi} \int_{-\pi}^{\pi} \exp\left(v(\cos t) \sum_{1\le\ell\le n} \frac{1}{\ell} + O(1)\right) \mathrm{d}t \\ &= O((v\lambda_n)^{-1/2} v^{-k} n^v), \end{aligned}$$

proving (17). □

Note that when $k = O(\lambda_n)$, then the right-hand side of (17) is optimal if we take $v = k/\lambda_n$ and (17) becomes $\mu_{n,k} = O(\lambda_n^k / k!)$. Thus (17) is tight when $k = O(\lambda_n)$. This also explains why we write $(v\lambda_n)^{-1/2}$ instead of $\lambda_n^{-1/2}$ (to keep uniformity when $k = o(\lambda_n)$ and we choose $v = k/\lambda_n$).

On the other hand, leaving v unspecified in (17) and in many other estimates in this paper considerably simplifies the analysis.

A simple asymptotic transfer. We will need the following result when applying the contraction method. It roughly says that when the non-homogeneous part $b_{n,k}$ of (14) is of order $\mu_{n,k}^w$, where $w > 1$, then $a_{n,k}$ is also of the same order for a certain range of α.

LEMMA 3. *If $b_{n,k} = O(((v\lambda_n)^{-1/2} v^{-k} n^v)^w)$ for all $1 \le k \le n$, where $w > 1$ and $0 < v < v_0$, then*

$$a_{n,k} = O\left(\frac{1}{w - v^{w-1}} \left((v\lambda_n)^{-1/2} v^{-k} n^v\right)^w\right),$$

uniformly for $1 \le k \le n$, provided that $0 < v < \min\{w^{1/(w-1)}, v_0\}$. Similarly, replacing O by o in the estimate for $b_{n,k}$ yields an o-estimate for $a_{n,k}$.

PROOF. By the exact expression for $a_{n,k}$, we have, for $0 < v < v_0$,

$$(18) \qquad a_{n,k} - b_{n,k} = O\left(\sum_{1\le j<n} \sum_{0\le r\le k} \frac{1}{j} ((v\lambda_j)^{-1/2} v^{-k+r} j^v)^w [u^r](1+u) \prod_{j<\ell<n} \left(1 + \frac{u}{\ell}\right)\right).$$

The inner sum over r can be simplified as follows:

$$\sum_{0\le r\le k} v^{-(k-r)w}[u^r](1+u)\prod_{j<\ell<n}\left(1+\frac{u}{\ell}\right) \le v^{-kw}\sum_{r\ge 0} v^{rw}[u^r](1+u)\prod_{j<\ell<n}\left(1+\frac{v^w t}{\ell}\right) = v^{-kw}(1+v^w)\prod_{j<\ell<n}\left(1+\frac{v^w}{\ell}\right) = O\left(v^{-kw}\left(\frac{n}{j}\right)^{v^w}\right), \tag{19}$$

uniformly in j. Substituting this estimate into (18), we obtain

$$\begin{aligned} a_{n,k} &= O\left(\left((v\lambda_n)^{-1/2}v^{-k}n^v\right)^w + v^{-kw}n^{v^w}\sum_{1\le j<n}(v\lambda_j)^{-w/2}j^{wv-v^w-1}\right) \\ &= O\left(\frac{1}{w-v^{w-1}}\left((v\lambda_n)^{-1/2}v^{-k}n^v\right)^w\right), \end{aligned}$$

uniformly for $1 \le k \le n$, where $0 < v < w^{1/(w-1)}$. The o-estimate is similarly proved. This completes the proof of Lemma 3. □

3. Convergence in Distribution when $0 < \alpha < e$. We prove the first part of Theorem 1 (excepting $\alpha = 0$) in this section by contraction method based on the framework developed in Neininger and Rüschendorf (2004). The new difficulty arising here is the asymptotics of the double-indexed recurrence (14) (instead of the single-indexed ones previously encountered).

The underlying idea. The idea used here is roughly as follows.

Define $\bar{X}_{n,k} := X_{n,k}/\mu_{n,k}$. Then, by (1), $\bar{X}_{n,k}$ satisfies the recurrence

$$\bar{X}_{n,k} \stackrel{d}{=} \frac{\mu_{I_n,k-1}}{\mu_{n,k}}\bar{X}_{I_n,k-1} + \frac{\mu_{n-I_n,k}}{\mu_{n,k}}\bar{X}^*_{n-I_n,k}, \tag{20}$$

with independence conditions as in (1). By the estimates (3) and the relation $I_n = \lceil (n-1)U\rceil$, we expect that

$$\frac{\mu_{I_n,k-1}}{\mu_{n,k}} \approx \frac{k}{\lambda_n}\left(\frac{\lambda_n+\log U}{\lambda_n}\right)^{k-1} \to \alpha U^\alpha,$$

with a suitable meaning for the convergence; similarly,

$$\frac{\mu_{n-I_n,k}}{\mu_{n,k}} \to (1-U)^\alpha.$$

Thus if we expect that $\bar{X}_{n,k} \to X(\alpha)$, then $X(\alpha)$ satisfies the fixed-point equation (4).

To justify these steps, we apply the contraction method.

Contraction method. The fixed-point equation (4) has a few special properties not enjoyed by single-indexed recursions encountered in the literature for which the typical fixed-point equation has the form

$$X \stackrel{d}{=} \sum_{1\le j\le h} C_j X^{(j)} + b, \tag{21}$$

with $X^{(1)}, \dots, X^{(h)}, (C_1, \dots, C_h, b)$ independent, $X^{(j)} \stackrel{d}{=} X$, and $0 \le C_j \le 1$ almost surely for all $1 \le j \le h$. Here, h may be deterministic or integer-valued random variables. The special range [0, 1] for the coefficients $C_1, \dots, C_j$ is roughly due to the relation

$$\frac{\sigma(I_j^{(n)})}{\sigma(n)} \to C_j,$$

where, in various applications (see Neininger and Rüschendorf, 2004), σ is the leading term in the expansion of the standard deviation of the underlying random variable and $0 \le I_j^{(n)} \le n$ are the sizes of the subproblems. Typically, σ is a monotonically increasing function, hence we obtain $0 \le C_j \le 1$.

In general, the Lipschitz constant of the map of probability measures associated with (21) under the Zolotarev metric ζ_w is assessed by $\sum_j \mathbb{E}(C_j^w)$. This term is monotonically decreasing as w increases. Thus, in typical applications for which one expects a contraction, the sum $\sum_j \mathbb{E}(C_j^w)$ has to satisfy $\sum_j \mathbb{E}(C_j^w) < 1$, and for that purpose, one has to choose w sufficiently large; see Neininger and Rüschendorf (2004) for implications of this condition on the moments required.

For the bi-indexed recursion of $X_{n,k}$, we are led to the fixed-point equation (4), where the coefficient αU^α may have values larger than one for $\alpha > 1$. This implies that the corresponding estimate $\mathbb{E}(\alpha U)^w + \mathbb{E}(1-U)^w$ for the Lipschitz constant is not decreasing in w. When $\alpha < e$ increases, the range where we have contraction becomes smaller and vanishes in the boundary case $\alpha = e$.

Notations. We denote by $\mathcal{M}$ the space of univariate probability measures, by $\mathcal{M}_w \subset \mathcal{M}$ the space of probability measures with finite absolute wth moment, and by $\mathcal{M}_w(1) \subset \mathcal{M}_w$ the subspace of probability measures with unit mean, where $1 < w \le 2$. Zolotarev (1976) introduced a family of metrics ζ_w, which, for $1 < w \le 2$, are given by

$$\zeta_w(\nu_1, \nu_2) = \sup_{f\in\mathcal{F}_w} |\mathbb{E}(f(X) - f(Y))| \qquad (\nu_1, \nu_2 \in \mathcal{M}_w(1)),$$

where X and Y have the distributions $\mathcal{L}(X) = \nu_1$, $\mathcal{L}(Y) = \nu_2$.

We have

$$\mathcal{F}_w := \{f \in C^1(\mathbb{R}, \mathbb{R}) : |f'(x) - f'(y)| \le |x - y|^{w-1}\},$$

with $C^1(\mathbb{R}, \mathbb{R})$ the space of continuously differentiable functions on $\mathbb{R}$. We use the property that convergence in ζ_w implies weak convergence and that ζ_w is ideal of order w, i.e. we have, for W independent of (X, Y) and $c \ne 0$,

$$\zeta_w(X + W, Y + W) \le \zeta_w(X, Y), \qquad \zeta_w(cX, cY) = |c|^w \zeta_w(X, Y).$$

For general reference and properties of ζ_w, see Zolotarev (1977) and Rachev (1991).

We also use the minimal L_p metrics ℓ_p, defined for $1 < p \le 2$ by

$$\ell_p(\nu_1, \nu_2) = \inf\{\|X - Y\|_p : \mathcal{L}(X) = \nu_1, \mathcal{L}(Y) = \nu_2\} \qquad (\nu_1, \nu_2 \in \mathcal{M}_p),$$

where $\|X\|_p$ denotes the L_p-norm of a random variable X. For simplicity, we use the abbreviation $\zeta_w(X, Y) := \zeta_w(\mathcal{L}(X), \mathcal{L}(Y))$ for ζ_w as well as for the other metrics appearing subsequently.

In addition, we assume that

$$R(n) := |k - \alpha\lambda_n| = |\alpha_{n,k} - \alpha|\lambda_n = o(\lambda_n),$$

where $0 < \alpha < e$, and fix a constant s as follows. If $2 \le \alpha < e$, then $1 < s < \rho$ with $\rho \in (1, 2]$ the unique solution of $\rho = \alpha^{\rho-1}$, and $s := 2$ if $0 < \alpha < 2$. The bound ρ also identifies the best possible order for the existence of absolute moment of $X(\alpha)$. Note that s satisfies $s - \alpha^{s-1} > 0$, which is the continuous version of $m - \alpha^{m-1} > 0$ appearing in (7).

Properties of $X(\alpha)$. Define the map

$$T : \mathcal{M} \to \mathcal{M}, \quad \nu \mapsto \mathcal{L}(\alpha U^\alpha Z + (1 - U)^\alpha Z^*),$$

where Z, Z^*, U are independent, $\mathcal{L}(Z) = \mathcal{L}(Z^*) = \nu$.

PROPOSITION 1. *For $0 < \alpha < e$, the restriction of T to $\mathcal{M}_s(1)$ has a unique fixed point $\mathcal{L}(X(\alpha))$. Furthermore, $\mathbb{E}|X(\alpha)|^\rho = \infty$ for $2 \le \alpha < e$.*

PROOF. By Lemma 3.1 in Neininger and Rüschendorf (2004), T is a Lipschitz map in ζ_s with Lipschitz constant bounded above by

$$\operatorname{lip}(T) \le \frac{\alpha^s + 1}{\alpha s + 1}.$$

Thus $\operatorname{lip}(T) < 1$ by our choice of s. Also T has a unique fixed point in the subspace $\mathcal{M}_s(1)$ by Lemma 3.3 in Neininger and Rüschendorf (2004).

When $2 \le \alpha < e$, we assume $\mathbb{E}|X(\alpha)|^\rho < \infty$ and prove a contradiction. First we have $\mathbb{E}|X(\alpha)|^\rho = \mathbb{E}|\alpha U^\alpha X(\alpha) + (1 - U)^\alpha X(\alpha)^*|^\rho$, where $X(\alpha)$, $X(\alpha)^*$, and U are independent with $\mathcal{L}(X(\alpha)) = \mathcal{L}(X(\alpha)^*)$. Note that $X(\alpha) \ge 0$ almost surely. Furthermore, $\mathbb{E}(X(\alpha)) = 1$ implies that there is a set with positive probability in which we have $X(\alpha) > 0$ and $X(\alpha)^* > 0$. It follows that

$$\begin{aligned}\mathbb{E}|X(\alpha)|^\rho = \mathbb{E}(X(\alpha)^\rho) &= \mathbb{E}(\alpha U^\alpha X(\alpha) + (1 - U)^\alpha X(\alpha)^*)^\rho \\ &> \mathbb{E}\left(\alpha^\rho U^{\alpha\rho} X(\alpha)^\rho + (1 - U)^{\alpha\rho}(X(\alpha)^*)^\rho\right) \\ &= \frac{\alpha^\rho + 1}{\alpha\rho + 1}\mathbb{E}(X(\alpha)^\rho) \\ &= \mathbb{E}(X(\alpha)^\rho),\end{aligned}$$

by the definition of ρ and the inequality $(a + b)^\rho > a^\rho + b^\rho$ for $a, b > 0$ and $\rho > 1$. This is a contradiction, hence, we have $\mathbb{E}|X(\alpha)|^\rho = \infty$. □

Zolotarev distance between $X_{n,k}/\mu_{n,k}$ and $X(\alpha)$

THEOREM 4. *If $0 < \alpha < 2$, then*

$$\zeta_2\left(\frac{X_{n,k}}{\mu_{n,k}}, X(\alpha)\right) = O\left(\frac{R(n)+1}{\lambda_n}\right).$$

If $2 \le \alpha < e$, then

$$\zeta_s\left(\frac{X_{n,k}}{\mu_{n,k}}, X(\alpha)\right) \to 0,$$

where s is specified as above.

In particular, this theorem implies the convergence in distribution of $X_{n,k}/\mu_{n,k}$ for $0 < \alpha < e$ and proves the first part of Theorem 1.

Convergence rate of the factors in (20)

LEMMA 4. *With s and $R(n)$ specified as above, we have*

$$\left\|\frac{\mu_{I_n,k-1}}{\mu_{n,k}} - \alpha U^\alpha\right\|_s + \left\|\frac{\mu_{n-I_n,k}}{\mu_{n,k}} - (1-U)^\alpha\right\|_s = O\left(\frac{R(n)+1}{\lambda_n}\right).$$

PROOF. We consider only the L_s-norm of $\mu_{I_n,k-1}/\mu_{n,k} - \alpha U^\alpha$, the other part being similar. By (3), we have

$$\mu_{n,k} = \frac{\mathsf{s}(n,k+1)}{(n-1)!} = \frac{\lambda_n^k}{k!}H(n,k),$$

where

$$H(n,k) = \frac{1}{\Gamma(1+\alpha_{n,k})} + O\left(\frac{1}{\lambda_n}\right), \tag{22}$$

the O-term holding uniformly for $1 \le k \le K\lambda_n$. Then we decompose the ratio $\mu_{I_n,k-1}/\mu_{n,k}$ into three parts:

$$\frac{\mu_{I_n,k-1}}{\mu_{n,k}} = \frac{k}{\lambda_n}\left(\frac{\log I_n}{\lambda_n}\right)^{k-1}\frac{H(I_n,k-1)}{H(n,k)} =: F_n^{[1]}F_n^{[2]}F_n^{[3]}. \tag{23}$$

We first show that

$$|F_n^{[1]} - \alpha| + \|F_n^{[2]} - U^\alpha\|_{4s} + \|F_n^{[3]} - 1\|_{4s} = O\left(\frac{R(n)+1}{\lambda_n}\right).$$

These estimates imply that $\|F_n^{[2]}\|_{4s}, \|F_n^{[3]}\|_{4s} = O(1)$. Then Hölder's inequality gives

$$\left\|\frac{\mu_{I_n,k-1}}{\mu_{n,k}} - \alpha U^\alpha\right\|_s = O\left(\frac{R(n)+1}{\lambda_n}\right).$$

First, we introduce the set $\mathcal{A} := \{I_n \le n^{\alpha/6}\}$. Note that $\mu_{n,k} = O(1)$ for $k \ge 3\lambda_n$. On the set $\mathcal{A}$, we have $k - 1 = \alpha\lambda_n + R(n) - 1 \ge (\alpha/2)\lambda_n \ge (\alpha/2)\log I_n^{6/\alpha} = 3\log I_n$, for sufficiently large n; thus $\mu_{I_n,k-1} = O(1)$. On the other hand, since $\alpha < e$, the mean satisfies $\mu_{n,k} = \Omega(1)$; thus

$$\int_{\mathcal{A}} \left| \frac{\mu_{I_n,k-1}}{\mu_{n,k}} - \alpha U^\alpha \right|^{4s} d\mathbb{P} = O(\mathbb{P}(\mathcal{A})) = O(\mathbb{P}(I_n \le \sqrt{n})) = O(1/\sqrt{n}) = O(\lambda_n^{-4s}).$$

Thus we need only consider the complement set $\mathcal{A}^c$.

Obviously, $F_n^{[1]} = k/\lambda_n = \alpha + O(R(n)/\lambda_n)$.

For $F_n^{[2]}$, we observe that for $x \le 0$ the expansion $(1 + x/m)^m = e^x + O(e^{\vartheta x}/m)$ holds uniformly with $\vartheta < 1$. Thus, we obtain

$$\begin{aligned} F_n^{[2]} = \left(\frac{\log I_n}{\lambda_n}\right)^{k-1} &= \left(\frac{I_n}{n} + O\left(\frac{(I_n/n)^\vartheta}{\lambda_n}\right)\right)^{\alpha + (R(n)-1)/\lambda_n} \\ &= U^\alpha + O\left(\frac{R(n)(U^\alpha + U^{\alpha+\vartheta-1})\log U + U^{\alpha+\vartheta-1}}{\lambda_n}\right). \end{aligned}$$

Here we may choose ϑ with $1-\alpha < \vartheta < 1$. Then $(U^\alpha + U^{\alpha+\vartheta-1})\log U$ and $U^{\alpha+\vartheta-1}$ are both L_{4s}-integrable and the O-term in the last display is bounded above by $O((R(n) + 1)/\lambda_n)$ in L_{4s}.

For the third factor in (23), we have

$$H(n,k) = \frac{1}{\Gamma(1 + \alpha + R(n)/\lambda_n)} + O\left(\frac{1}{\lambda_n}\right) = \frac{1}{\Gamma(1+\alpha)} + O\left(\frac{R(n)+1}{\lambda_n}\right).$$

For $H(I_n, k-1)$, we restrict to the set $\mathcal{A}^c$. On $\mathcal{A}^c$, for n sufficiently large, we have $k - 1 \le 12\log I_n$, so the error in the expansion of $H(I_n, k-1)$ implied by (22) is uniformly $O(1/\log I_n) = O(1/\lambda_n)$. Thus we have

$$\begin{aligned} H(I_n, k-1) &= \frac{1}{\Gamma\left(1 + \alpha + (\alpha\log(n/I_n) + R(n) - 1)/\log I_n\right)} + O\left(\frac{1}{\log I_n}\right) \\ &= \frac{1}{\Gamma(1+\alpha)} + O\left(\frac{\log(n/I_n) + R(n)}{\lambda_n}\right). \end{aligned}$$

Since $\|\log(n/I_n)\|_{4s} \to \|\log U\|_{4s} < \infty$, the last error term is of order $O((R(n)+1)/\lambda_n)$ in L_{4s}. Collecting all estimates, we obtain $\|F_n^{[3]} - 1\|_{4s} = O((R(n)+1)/\lambda_n)$. □

Asymptotic transfer of the double-indexed recurrence (14). Consider the recurrence (14) with suitable initial conditions.

LEMMA 5. *If*

$$b_{n,k} = O\left(((v\lambda_n)^{-1/2} n^v v^{-k})^w \cdot \frac{R(n)+1}{\lambda_n}\right) \qquad (1 < w \le 2),$$

uniformly for $1 \le k < n$, *where* $0 < v < v_0$, *then*

$$a_{n,k} = O\left(\frac{1}{w - v^{w-1}}((v\lambda_n)^{-1/2} n^v v^{-k})^w \cdot \frac{R(n)+1}{\lambda_n}\right), \tag{24}$$

uniformly for $1 \le k < n$, *where* $0 < v < \min\{w^{1/(w-1)}, v_0\}$.

PROOF. The proof is similar to that for Lemma 3 but slightly more complicated. By the exact expression for $a_{n,k}$ and the estimate for $b_{n,k}$, we have, for $0 < v < v_0$,

$$a_{n,k} - b_{n,k} = O\left(v^{-wk-w/2} \sum_{1\le j<n} \sum_{0\le r\le k} |k - r - \alpha\lambda_j| \lambda_j^{-1-w/2} j^{wv-1} v^{wr} [u^r](1+u) \times \prod_{j<\ell<n} \left(1 + \frac{u}{\ell}\right)\right).$$

First, if $|k - \alpha\lambda_n| \ge \varepsilon\lambda_n$, then $|k - r - \alpha\lambda_j| = O(k + \lambda_n)$, so that (24) holds by the proof of Lemma 3. We assume now that $|k - \alpha\lambda_n| \le \varepsilon\lambda_n$. Split the sum in j into three parts:

$$a_{n,k} - b_{n,k} = O\left(v^{-wk-w/2}\left(\sum_{1\le j<\delta n} + \sum_{\delta n\le j\le(1-\delta)n} + \sum_{(1-\delta)n<j<n}\right) \times \sum_{0\le r\le k} |k - r - \alpha\lambda_j| \lambda_j^{-1-w/2} j^{wv-1} v^{wr} [u^r](1+u) \prod_{j<\ell<n}\left(1+\frac{u}{\ell}\right)\right),$$

where $\delta \in (0, 1)$ will be specified later. An analysis similar to the proof of Lemma 3 gives

$$a_{n,k} - b_{n,k} = O\left(\frac{(v\lambda_n)^{-w/2}}{w - v^{w-1}} v^{-wk} n^{wv} \left(\delta^{wv-v^w} + \frac{|k - \alpha\lambda_n| + 1}{\lambda_n} + \delta\right)\right),$$

where $0 < v < \min\{w^{1/(w-1)}, v_0\}$. Taking $\delta := ((R(n)+1)/\lambda_n)^{1/(wv-v^w)}$ yields (24). □

An inequality between ζ_s- *and* ℓ_s-*distances*

LEMMA 6. *For* $1 < w \le 2$ *and* $M > 0$, *there is a constant* $K > 0$ *such that*

$$\zeta_w(X, Y) \le K(\ell_w(X, Y) \vee \ell_w^{w-1}(X, Y)), \tag{25}$$

for all pairs $\mathcal{L}(X), \mathcal{L}(Y) \in \mathcal{M}_w(1)$ *with* $\|X\|_w, \|Y\|_w \le M$.

PROOF. We start from the inequality (see Theorem 3 in Zolotarev (1976)

$$\zeta_w(X, Y) \le \frac{1}{w}(2\beta_w(X, Y) + 2^{w-1}\beta_w^{w-1}(X, Y)(\|X\|_w^w \wedge \|Y\|_w^w)^{2-w}),$$

for $1 < w \le 2$, where β_w denotes the difference pseudo-moment

$$\beta_w(\nu_1, \nu_2) := \inf\{\mathbb{E}||X|^{w-1}X - |Y|^{w-1}Y| : \mathcal{L}(X) = \nu_1, \mathcal{L}(Y) = \nu_2\} \qquad (w > 1),$$

with $\nu_1, \nu_2 \in \mathcal{M}_w$. From $||x|^{w-1}x - |y|^{w-1}y| \le w(|x|^{w-1} \vee |y|^{w-1})|x-y|$ and Hölder's inequality, it follows that

$$\beta_w(X, Y) \le w(\mathbb{E}|X|^w + \mathbb{E}|Y|^w)^{(w-1)/w} \ell_w(X, Y),$$

which implies the desired inequality. □

PROOF OF THEOREM 4. We introduce a hybrid quantity

$$\Xi_n := \frac{\mu_{I_n,k-1}}{\mu_{n,k}} X(\alpha) + \frac{\mu_{n-I_n,k}}{\mu_{n,k}} X^*(\alpha),$$

where $X(\alpha)$, $X^*(\alpha)$, and I_n are independent and $X(\alpha)$ and $X^*(\alpha)$ are identically distributed. Since $\mathcal{L}(X(\alpha))$, $\mathcal{L}(\bar{X}_{n,k})$, $\mathcal{L}(\Xi_n) \in \mathcal{M}_s(1)$, the ζ_s-distances between these quantities are finite. For simplicity, write $h_{n,k} := \zeta_s(\bar{X}_{n,k}, X(\alpha))$. By triangle inequality

$$h_{n,k} \le \zeta_s(\bar{X}_{n,k}, \Xi_n) + \zeta_s(\Xi_n, X(\alpha)).$$

Note that ζ_s is ideal of order s. Thus

$$\begin{aligned}
\zeta_s(\bar{X}_{n,k}, \Xi_n) &= \zeta_s\left(\frac{\mu_{I_n,k-1}}{\mu_{n,k}} \bar{X}_{I_n,k-1} + \frac{\mu_{n-I_n,k}}{\mu_{n,k}} \bar{X}^*_{n-I_n,k}, \frac{\mu_{I_n,k-1}}{\mu_{n,k}} X(\alpha) + \frac{\mu_{n-I_n,k}}{\mu_{n,k}} X^*(\alpha)\right) \\
&\le \frac{1}{n-1} \sum_{1 \le j < n} \zeta_s \\
&\quad \times \left(\frac{\mu_{j,k-1}}{\mu_{n,k}} \bar{X}_{j,k-1} + \frac{\mu_{n-j,k}}{\mu_{n,k}} \bar{X}^*_{n-j,k}, \frac{\mu_{j,k-1}}{\mu_{n,k}} X(\alpha) + \frac{\mu_{n-j,k}}{\mu_{n,k}} X^*(\alpha)\right) \\
&\le \frac{1}{n-1} \sum_{1 \le j < n} \left(\left(\frac{\mu_{j,k-1}}{\mu_{n,k}}\right)^s h_{j,k-1} + \left(\frac{\mu_{n-j,k}}{\mu_{n,k}}\right)^s h_{n-j,k}\right).
\end{aligned}$$

We now show that

$$\zeta_s(\Xi_n, X(\alpha)) = O(D(n)^{s-1}), \tag{26}$$

where $D(n) := (R(n) + 1)/\lambda_n$.

First, by Lemma 4,

$$\begin{aligned}
\|\Xi_n\|_s &\le \left(\left\|\frac{\mu_{I_n,k-1}}{\mu_{n,k}}\right\|_s + \left\|\frac{\mu_{n-I_n,k}}{\mu_{n,k}}\right\|_s\right) \|X(\alpha)\|_s \\
&\to (\alpha\|U^\alpha\|_s + \|(1-U)^\alpha\|_s)\|X(\alpha)\|_s,
\end{aligned}$$

which implies that $\|\Xi_n\|_s$ is uniformly bounded for all n. Since $\mathcal{L}(X(\alpha)) \in \mathcal{M}_s(1)$, there is an $M > 0$ such that $\|X(\alpha)\|_s, \|\Xi_n\|_s \le M$ for all n. We apply Lemma 6 to bound the ζ_s-distance, which gives

$$\zeta_s(\Xi_n, X(\alpha)) \le K(\ell_s(\Xi_n, X(\alpha)) \vee \ell_s^{s-1}(\Xi_n, X(\alpha))).$$

By Lemma 4,

$$\begin{aligned}\ell_s(\Xi_n, X(\alpha)) &\le \left(\left\| \frac{\mu_{I_n,k-1}}{\mu_{n,k}} - \alpha U^\alpha \right\|_s + \left\| \frac{\mu_{n-I_n,k}}{\mu_{n,k}} - (1-U)^\alpha \right\|_s\right) \|X(\alpha)\|_s \\ &= O(D(n)).\end{aligned}$$

This proves (26).

Collecting the estimates, we obtain

$$h_{n,k} \le \frac{1}{n-1} \sum_{1\le j<n} \left(\left(\frac{\mu_{j,k-1}}{\mu_{n,k}}\right)^s h_{j,k-1} + \left(\frac{\mu_{n-j,k}}{\mu_{n,k}}\right)^s h_{n-j,k} \right) + O\left(D(n)^{s-1}\right).$$

Thus, $h_{n,k} = O(a_{n,k}\mu_{n,k}^{-s})$, where $a_{n,k}$ satisfies (14) with

$$b_{n,k} = O(\mu_{n,k}^s D(n)^{s-1}),$$

and suitable initial conditions. Theorem 4 then follows from applying the different types of asymptotic transfer given in Lemmas 3 and 5. □

REMARK. Note that the proof of Theorem 4 also yields a rate of convergence of order $O(((R(n)+1)/\lambda_n)^{s-1})$ for ζ_s for the range $2 \le \alpha < e$.

Recently, S. Janson (private communication) showed that Lemma 6 also holds with (25) there replaced by

$$\zeta_w(X, Y) \le K\ell_w(X, Y).$$

This inequality leads to an improvement of the error term in Theorem 4 for the range $2 \le \alpha < e$ to $O((R(n)+1)/\lambda_n)$.

4. Asymptotics of Moments. We prove in this section the moment estimate (6) whose proof is more involved than the asymptotic transfer in Lemma 3. The idea is first to derive a crude bound for higher moments of $X_{n,k}$, which holds uniformly for $1 \le k < n$. Then a more refined analysis leads to (6).

Note that the mth factorial moments of $X_{n,k}$ and the mth moments are asymptotically equivalent when $\mu_{n,k} \to \infty$, or roughly when $\alpha < e$.

A uniform estimate for higher moments. For convenience, define $\varphi_1(v) = 1$ and

$$\varphi_m(v) := \frac{1}{m - v^{m-1}} \qquad (m \ge 2).$$

We now prove by induction that

$$A_{n,k}^{(m)} = O(\varphi_m(v)((v\lambda_n)^{-1/2} v^{-k} n^v)^m) \qquad (m \ge 1), \tag{27}$$

uniformly for $1 \le k < n$, where $0 < v < m^{1/(m-1)}$.

Obviously, (27) holds for $m = 1$ by (17). By (13) and induction, we have, for $0 < v < (m-1)^{1/(m-2)}$,

$$
\begin{aligned}
(28)\quad B_{n,k}^{(m)} &= O\Bigg(\sum_{1\le h<m}\binom{m}{h}\varphi_h(v)\varphi_{m-h}(v) \\
&\qquad \times\ n^{-1}\sum_{1\le j<n}\left((v\lambda_j)^{-1/2}v^{-k+1}j^{v}\right)^h\left((v\lambda_{n-j})^{-1/2}v^{-k}(n-j)^v\right)^{m-h}\Bigg) \\
&= O\Bigg(\varphi_{m-1}(v)v^{-km}n^{-1}\sum_{\substack{1\le h<m\\1\le j<n}} j^{hv}(n-j)^{(m-h)v}(v\lambda_j)^{-h/2}(v\lambda_{n-j})^{-(m-h)/2}\Bigg) \\
&= O(\varphi_{m-1}(v)(v\lambda_n)^{-m/2}v^{-km}n^{mv}),
\end{aligned}
$$

uniformly for $1 \le k < n$.

By (15),

$$
(29)\qquad A_{n,k}^{(m)} = B_{n,k}^{(m)} + \sum_{1\le j<n}\sum_{0\le r\le k}\frac{B_{j,k-r}^{(m)}}{j}[u^r](u+1)\prod_{j<\ell<n}\left(1+\frac{u}{\ell}\right).
$$

Substituting the estimate (28) into (29) gives, for $0 < v < m^{1/(m-1)}$,

$$
\begin{aligned}
A_{n,k}^{(m)} &= O\Bigg(B_{n,k}^{(m)} + v^{-km}\sum_{1\le j<n}(v\lambda_j)^{-m/2}j^{mv-1}\sum_{0\le r\le k}v^{rm}[u^r](1+u)\prod_{j<\ell<n}\left(1+\frac{u}{\ell}\right)\Bigg) \\
&= O\left(B_{n,k}^{(m)} + \varphi_m(v)(v\lambda_n)^{-m/2}n^{mv}v^{-km}\right),
\end{aligned}
$$

similar to the proof of Lemma 3. This proves (27).

Note that when $\alpha \le m^{1/(m-1)} - \varepsilon$, the optimal choice of v in (27) minimizing $n^v v^{-k}$ is $v = \alpha_{n,k}$, which yields the estimate $A_{n,k}^{(m)} = O(\lambda_n^k/k!)$, uniformly in k. When $\alpha \ge m^{1/(m-1)} - \varepsilon$, the optimal choice is then $v = m^{1/(m-1)} - \varepsilon$. This says that the asymptotic behavior of $A_{n,k}^{(m)}$ when $\alpha < m^{1/(m-1)}$ is very different from that when $\alpha \ge m^{1/(m-1)}$. More precise estimates can be derived, but they are not needed here; see Drmota and Hwang (2005a) for asymptotic approximations to the variance (covering all ranges).

Asymptotics of $A_{n,k}^{(m)}$. Since the case $\alpha = 0$ is treated separately, we assume throughout this section that $\alpha > 0$. We refine the above inductive argument and show that

$$
(30)\qquad A_{n,k}^{(m)} \sim \nu_m(\alpha)\mu_{n,k}^m \sim \nu_m(\alpha)\left(\frac{\lambda_n^k}{\Gamma(1+\alpha)k!}\right)^m,
$$

for each $m \ge 1$ and $k/\lambda_n \to \alpha < m^{1/(m-1)}$, where $\nu_m(\alpha)$ denotes the moment sequence of $X(\alpha)$ given in (7). This will prove the moment convergence part of Theorem 1.

Note that by (3), (30) holds for $m = 1$ with $\nu_1(\alpha) = 1$. Assume that (30) holds for all $A_{n,k}^{(i)}$ with $i < m$. We split the right-hand side of (29) into three parts:

$$\begin{aligned} A_{n,k}^{(m)} &= B_{n,k}^{(m)} + \sum_{0\le r\le k}\left(\sum_{1\le j<\varepsilon n} + \sum_{\varepsilon n\le j\le(1-\varepsilon)n} + \sum_{(1-\varepsilon)n<j<n}\right)\frac{B_{j,k-r}^{(m)}}{j}[u^r](u+1) \\ &\quad\times\prod_{j<\ell<n}\left(1+\frac{u}{\ell}\right) \\ &=: B_{n,k}^{(m)} + A_{n,k}^{(m)}[1] + A_{n,k}^{(m)}[2] + A_{n,k}^{(m)}[3]. \end{aligned}$$

By the same proof used for Lemma 3, we have

$$\begin{aligned} A_{n,k}^{(m)}[1] &= O(\varepsilon^{mv-v^m}\varphi_m(v)\lambda_n^{-(m+1)/2}n^{mv}v^{-km}), \\ A_{n,k}^{(m)}[3] &= O(\varepsilon\varphi_m(v)\lambda_n^{-(m+1)/2}n^{mv}v^{-km}). \end{aligned}$$

Letting $\varepsilon \to 0$, we see that, by (27),

$$A_{n,k}^{(m)}[1] + A_{n,k}^{(m)}[3] = o(A_{n,k}^{(m)}).$$

Asymptotics of $A_{n,k}^{(m)}$*: the dominant terms.* We start by showing that, for $0 < \alpha < (m-1)^{1/(m-2)}$,

$$B_{n,k}^{(m)} \sim \nu_m^*(\alpha)\left(\frac{\lambda_n^k}{\Gamma(1+\alpha)k!}\right)^m \qquad (m\ge 2), \tag{31}$$

where

$$\nu_m^*(\alpha) := \sum_{1\le h<m}\binom{m}{h}\nu_h(\alpha)\nu_{m-h}(\alpha)\alpha^h\int_0^1 u^{h\alpha}(1-u)^{(m-h)\alpha}\,\mathrm{d}u.$$

By (13), induction and (30), we have, for $0 < \alpha < (m-1)^{1/(m-2)}$,

$$\begin{aligned} B_{n,k}^{(m)} &\sim \sum_{1\le h<m}\binom{m}{h}\nu_h(\alpha)\nu_{m-h}(\alpha)\frac{1}{n} \\ &\quad\times\sum_{\varepsilon n\le j\le(1-\varepsilon)n}\left(\frac{\lambda_j^{k-1}}{\Gamma(1+\alpha)(k-1)!}\right)^h\left(\frac{\lambda_{n-j}^k}{\Gamma(1+\alpha)k!}\right)^{m-h} \\ &\sim \left(\frac{\lambda_n^k}{\Gamma(1+\alpha)k!}\right)^m\sum_{1\le h<m}\binom{m}{h}\nu_h(\alpha)\nu_{m-h}(\alpha)\frac{1}{n} \\ &\quad\times\sum_{\varepsilon n\le j\le(1-\varepsilon)n}\alpha^h\left(\frac{j}{n}\right)^{kh/\lambda_n}\left(1-\frac{j}{n}\right)^{k(m-h)/\lambda_n}, \end{aligned}$$

which proves (31). The errors introduced for terms with $j < \varepsilon n$ and for $j \ge (1-\varepsilon)n$ can be easily bounded by using (27).

To evaluate $A^{(m)}_{n,k}[2]$, we first observe that

$$\prod_{j<\ell<n}\left(1+\frac{u}{\ell}\right) = \exp\left(u\sum_{j<\ell<n}\ell^{-1}+O\left(\frac{|u|^2}{j}\right)\right)$$
$$= \left(\frac{n}{j}\right)^u (1+O(|u|^2 j^{-1})),$$

uniformly for finite complex u and $j\to\infty$. It follows that

$$[u^r]\prod_{j<\ell<n}\left(1+\frac{u}{\ell}\right) = \frac{(\log(n/j))^r}{r!}\left(1+O\left(\frac{r^2}{j}\right)\right),$$

uniformly for $\varepsilon n \le j \le (1-\varepsilon)n$ and $0\le r\le k = o(\sqrt{j})$. Consequently, by (28) and (31),

$$A^{(m)}_{n,k}[2] \sim \nu^*_m(\alpha)\left(\frac{\lambda_n^k}{\Gamma(1+\alpha)k!}\right)^m \sum_{\varepsilon n\le j\le(1-\varepsilon)n} j^{-1}\left(\frac{j}{n}\right)^{m\alpha}$$
$$\times \sum_{r\ge 0}\alpha^{mr}\left(\frac{(\log(n/j))^{r-1}}{(r-1)!}+\frac{(\log(n/j))^r}{r!}\right)$$
$$\sim \nu^*_m(\alpha)(\alpha^m+1)\left(\frac{\lambda_n^k}{\Gamma(1+\alpha)k!}\right)^m\int_\varepsilon^{1-\varepsilon} x^{m\alpha-\alpha^m-1}\,\mathrm{d}x.$$

Letting $\varepsilon\to 0$, we then obtain, by (29), that

$$A^{(m)}_{n,k} \sim \nu^*_m(\alpha)\left(1+(\alpha^m+1)\int_0^1 x^{m\alpha-\alpha^m-1}\,\mathrm{d}x\right)\left(\frac{\lambda_n^k}{\Gamma(1+\alpha)k!}\right)^m$$
$$= \nu^*_m(\alpha)\frac{m\alpha+1}{m\alpha-\alpha^m}\left(\frac{\lambda_n^k}{\Gamma(1+\alpha)k!}\right)^m,$$

where

$$\nu^*_m(\alpha)\frac{m\alpha+1}{m\alpha-\alpha^m} = \frac{1}{m-\alpha^{m-1}}$$
$$\times\sum_{1\le h<m}\binom{m}{h}\nu_h(\alpha)\nu_{m-h}(\alpha)\alpha^{h-1}\frac{\Gamma(h\alpha+1)\Gamma((m-h)\alpha+1)}{\Gamma(m\alpha+1)}$$
$$= \nu_m(\alpha),$$

for $m\ge 2$, by (7). This completes the proof of (29) and thus Theorem 1(ii).

Moment convergence (6). Convergence of all moments implies convergence in distribution if the moment sequence (7) uniquely characterizes the distribution. By considering $\bar{\nu}_m(\alpha) := \nu_m(\alpha)\Gamma(m\alpha+1)/m!$, we easily obtain by induction that $\bar{\nu}_m(\alpha) = O(K^m)$ for $\alpha\in[0,1]$ (see Hwang and Neininger, 2002), and thus convergence in distribution of $X_{n,k}/\mu_{n,k}$ follows from (6) when $\alpha\in[0,1]$.

5. The Central Range $\alpha = 1$. We prove Theorem 2 in this section. The proof proceeds essentially along the same lines as we did above but with one major difference: we consider central moments instead of factorial moments. This minor step is crucial in dealing with the cancellations involved in the asymptotics of higher central moments. For simplicity, the case when $|t_{n,k}| \to \infty$ and $t_{n,k} = o(\lambda_n)$ is first analyzed; then the same method of proof is extended to the case when $t_{n,k} = O(1)$. Justifications of the error terms are similar to those for $A_{n,k}^{(m)}$ given above but become more complicated.

Recurrence of central moments. Consider $\bar{P}_{n,k}(y) := \mathbb{E}(e^{(X_{n,k}-\mu_{n,k})y}) = P_{n,k}(e^y)e^{-\mu_{n,k}y}$; see (12). Then we have the recurrence

$$\bar{P}_{n,k}(y) = \frac{1}{n-1}\sum_{1\le j<n} \bar{P}_{j,k-1}(y)\bar{P}_{n-j,k}(y)e^{\Delta_{n,k}(j)y} \qquad (n \ge 2;\, k \ge 1),$$

where

$$\Delta_{n,k}(j) := \mu_{j,k-1} + \mu_{n-j,k} - \mu_{n,k}$$

and $\bar{P}_{n,0}(y) = \bar{P}_{1,k}(y) = 1$ for $n, k \ge 1$.

Let now $P_{n,k}^{(m)} := \bar{P}_{n,k}^{(m)}(0)$ denote the mth central moment of $X_{n,k}$. Then $P_{n,k}^{(1)} \equiv 0$ and, for $m \ge 2$,

$$(32) \qquad P_{n,k}^{(m)} = \frac{1}{n-1}\sum_{1\le j<n}\left(P_{j,k-1}^{(m)} + P_{j,k}^{(m)}\right) + Q_{n,k}^{(m)} \qquad (n \ge 2;\, k \ge 1),$$

where

$$Q_{n,k}^{(m)} := \sum_{\substack{a+b+c=m\\ 0\le a,b<m\\ 0\le c\le m}} \binom{m}{a,\, b,\, c}\frac{1}{n-1}\sum_{1\le j<n} P_{j,k-1}^{(a)}P_{n-j,k}^{(b)}\Delta_{n,k}^{c}(j)$$

and $P_{n,0}^{(m)} = 0$ for $n, m \ge 1$.

Outline of the proof of Theorem 2. Similar to the proof of (30), we divide the proof of Theorem 2 into three main steps.

- We first derive a uniform estimate for $\Delta_{n,k}(j)$ for $1 \le j, k < n$, which then implies a uniform bound for $P_{n,k}^{(m)}$ for $1 \le k < n$. This bound is sufficient for our uses except when $|k - \lambda_n| = o(\sqrt{\lambda_n})$.
- We then derive a second estimate for $\Delta_{n,k}(j)$ uniformly valid for $k \sim \lambda_n$. This in turn implies a tight bound for $P_{n,k}^{(m)}$ when $k \sim \lambda_n$, and an asymptotic approximation to $P_{n,k}^{(m)}$ when $1 \ll |t_{n,k}| = o(\lambda_n)$.
- A finer estimate for $\Delta_{n,k}(j)$ is needed to deal with the case when $t_{n,k} = O(1)$.

An integral representation for $\Delta_{n,k}(j)$. By (2),

$$\mu_{n,k} = [u^k]\frac{n^u}{\Gamma(u+1)}(1 + O(n^{-1})).$$

Then

$$(33) \qquad \Delta_{n,k}(j) = \frac{1}{2\pi i}\oint_{|u|=v} u^{-k-1}n^u\varphi(u, j/n)(1 + O(j^{-1} + (n-j)^{-1}))\,\mathrm{d}u,$$

uniformly for $1 \le j < n$ (when j or $n-j$ is bounded, the O-term becoming $O(1)$ instead of $o(1)$), where

$$\varphi(u, x) := \frac{(1-x)^u + ux^u - 1}{\Gamma(u+1)}.$$

Here and throughout this section, we take $v = 1 + o(1)$ since $k \sim \lambda_n$.

A uniform estimate for $\Delta_{n,k}(j)$. Since $\varphi(1, x) = 0$, we have

$$|\varphi(u, x)| = O(|u-1|) \qquad (x \in [0, 1]).$$

Substituting this estimate into (33) gives

$$\begin{aligned} \Delta_{n,k}(j) \;&=\; O\left(v^{-k} n^v \int_{-\pi}^{\pi} \left|ve^{i\theta} - 1\right| n^{-v(1-\cos\theta)}\, \mathrm{d}\theta\right) \\ &=\; O((|v-1| + \lambda_n^{-1/2})\lambda_n^{-1/2} v^{-k} n^v), \end{aligned} \tag{34}$$

uniformly for $1 \le j, k < n$.

A uniform estimate for $P_{n,k}^{(m)}$. From the recurrence (32) and the estimate (34), we deduce, by an induction similar to that used for (27), that

$$Q_{n,k}^{(m)}, P_{n,k}^{(m)} = O((|v-1|^m + \lambda_n^{-m/2})(\lambda_n^{-1/2} v^{-k} n^v)^m) \qquad (m \ge 2), \tag{35}$$

uniformly for $1 \le k < n$. This bound is however not tight when $|k - \lambda_n| = o(\sqrt{\lambda_n})$, the reason being simply that v is not properly chosen to minimize the error term (the first $\lambda_n^{-1/2}$) in (34).

A finer estimate than (34). For a more precise estimate than (34), we use the two-term Taylor expansion

$$\varphi(u, x) = \varphi_u'(1, x)(u-1) + O(|u-1|^2),$$

where $\varphi_u'(1, x) = x + x\log x + (1-x)\log(1-x)$, which leads to

$$\begin{aligned} \Delta_{n,k}(j) \;&=\; \varphi_u'\left(1, \frac{j}{n}\right)(k-\lambda_n)\frac{\lambda_n^{k-1}}{k!}(1 + O(j^{-1} + (n-j)^{-1})) \\ &\quad + O((|v-1|^2 + \lambda_n^{-1})\lambda_n^{-1/2} v^{-k} n^v). \end{aligned} \tag{36}$$

Taking $v = k/\lambda_n$ gives

$$\Delta_{n,k}(j) = O\left((|k-\lambda_n| + 1)\frac{\lambda_n^{k-1}}{k!}\right). \tag{37}$$

This bound holds uniformly for $k \sim \lambda_n$ and $1 \le j < n$ since $\varphi_u'(1, x) = O(x|\log x|)$ as $x \to 0^+$.

A uniform bound for $P_{n,k}^{(m)}$ when $k \sim \lambda_n$. From (37), we deduce, again by induction, that

$$Q_{n,k}^{(m)}, P_{n,k}^{(m)} = O\left((|k-\lambda_n|^m + 1) \left(\frac{\lambda_n^{k-1}}{k!} \right)^m \right) \qquad (m \geq 2), \tag{38}$$

uniformly for $k \sim \lambda_n$. The proof differs slightly from that for (30) in that we split all sums of the form $\sum_{1\leq j<n}$ into three parts:

$$\sum_{1\leq j<n} = \sum_{1\leq j<n/\lambda_n^m} + \sum_{n/\lambda_n^m \leq j \leq n-n/\lambda_n^m} + \sum_{n-n/\lambda_n^m<j<n},$$

and then apply (38) and (37) to the middle sum, and (35) to the remaining two sums.

Asymptotics of $P_{n,k}^{(m)}$ when $|t_{n,k}| \to \infty$ and $t_{n,k} = o(\lambda_n)$. In this case the estimate (36) has the form

$$\Delta_{n,k}(j) \sim \varphi_u'\left(1, \frac{j}{n}\right) t_{n,k} \frac{\lambda_n^{k-1}}{k!}, \tag{39}$$

uniformly in k when $\varepsilon n \leq j \leq (1-\varepsilon)n$. Then we show that

$$P_{n,k}^{(m)} \sim g_m \left(t_{n,k} \frac{\lambda_n^{k-1}}{k!} \right)^m \qquad (m \geq 1), \tag{40}$$

where $g_0 = 1$, $g_1 = 0$ and, for $m \geq 2$,

$$g_m = \frac{m+1}{m-1} \sum_{\substack{a+b+c=m \\ 0\leq a,b<m \\ 0\leq c\leq m}} \binom{m}{a,b,c} g_a g_b \int_0^1 x^a (1-x)^b \varphi_u'(1,x)^c \, dx. \tag{41}$$

Equivalently, this can be written as

$$g_m = \sum_{\substack{a+b+c=m \\ 0\leq a,b,c\leq m}} \binom{m}{a,b,c} g_a g_b \int_0^1 x^a (1-x)^b \varphi_u'(1,x)^c \, dx.$$

In particular,

$$g_2 = 3 \int_0^1 \varphi_u'(1,x)^2 \, dx = 2 - \frac{\pi^2}{6}.$$

The inductive proof is almost the same as that for $A_{n,k}^{(m)}$, with the factor $(k-\lambda_n)^m$ handled by direct expansion and then estimated term by term. We also need to split sums of the form $\sum_{1\leq j<n}$ into five parts:

$$\sum_{1\leq j<n} = \sum_{1\leq j<n/\lambda_n^m} + \sum_{n/\lambda_n^m \leq j<\varepsilon n} + \sum_{\varepsilon n\leq j\leq (1-\varepsilon)n} + \sum_{(1-\varepsilon)n<j\leq n-n/\lambda_n^m} + \sum_{n-n/\lambda_n^m<j<n},$$

and then apply (40) to the middle sum, and the two estimates (35) and (38) to the other four sums.

The moment sequence (41) is easily checked to have the property of uniquely characterizing the distribution; see Hwang (2005) for similar details.

This proves the first part of Theorem 2.

The periodic case when $t_{n,k} = O(1)$. In this case we need a more precise expansion than (39) as follows:

$$\Delta_{n,k}(j) \sim \frac{\lambda_n^{k-1}}{k!}\left(\varphi_u'\left(1, \frac{j}{n}\right) t_{n,k} - \tfrac{1}{2}\varphi_{uu}''\left(1, \frac{j}{n}\right)\right), \tag{42}$$

uniformly for $j/n \in [\varepsilon, 1-\varepsilon]$ and $k \sim \lambda_n$, where

$$\varphi_{uu}''(1, x) = (x \log x + (1-x)\log(1-x))^2 - 2(1-\gamma)\varphi_u'(1, x).$$

This is proved by expanding more terms of $\varphi(u, x)$ at $u = 1$ and then estimating the error terms (see Hwang (1995) for similar details).

With the approximation (42), we first prove that, for $m \geq 0$,

$$\mathbb{E}(X_{n,k} - \mu_{n,k})^m = P_{n,k}^{(m)} \sim p_m(t_{n,k}) \left(\frac{\lambda_n^{k-1}}{k!}\right)^m, \tag{43}$$

where $p_m(t_{n,k})$ is a polynomial in $t_{n,k}$ of degree m with $p_0(t_{n,k}) = 1$ and $p_1(t_{n,k}) = 0$. This will imply that for $k = \lfloor \lambda_n \rfloor + \ell$, where $\ell \in \mathbb{Z}$,

$$\mathbb{E}\left(\frac{X_{n,k} - \mu_{n,k}}{\lambda_n^{k-1}/k!}\right)^m \sim p_m(\ell - \{\lambda_n\}),$$

for $m \geq 0$, where $\{\lambda_n\}$ denotes the fractional part of λ_n. Then we apply an argument based on the Frechet–Shohat moment convergence theorem similar to that used in Chern and Hwang (2001a) to prove that $(X_{n,k} - \mu_{n,k})/(\lambda_n^{k-1}/k!)$ does not converge to a fixed limit law. The proof for $(X_{n,k} - \mu_{n,k})/\sqrt{\mathbb{V}(X_{n,k})}$ is similar.

To prove (43), we again use induction. Assume $m \geq 2$. Then a similar analysis as above leads to

$$Q_{n,k}^{(m)} \sim q_m(t_{n,k}) \left(\frac{\lambda_n^{k-1}}{k!}\right)^m,$$

where $q_m(t)$ is a polynomial of degree m defined by

$$\begin{aligned} q_m(t_{n,k}) \; := \; & \sum_{\substack{a+b+c=m \\ 0 \leq a,b<m \\ 0 \leq c \leq m}} \binom{m}{a, b, c} \int_0^1 y^a (1-y)^b \\ & \times p_a(t_{n,k} - 1 - \log y) p_b(t_{n,k} - \log(1-y))(\varphi_u'(1, y) t_{n,k} - \tfrac{1}{2}\varphi_{uu}''(1, y))^c \, \mathrm{d}y. \end{aligned}$$

Then by (32), we deduce that, for $m \geq 2$,

$$\begin{aligned} P_{n,k}^{(m)} \left(\frac{\lambda_n^{k-1}}{k!}\right)^{-m} \; \sim \; & q_m(t_{n,k}) + \int_0^1 x^{m-1} \sum_{r \geq 0} \frac{\log(1/x)^r}{r!} \\ & \times (q_m(t_{n,k} - r - 1 - \log x) + q_m(t_{n,k} - r - \log x)) \, \mathrm{d}x, \end{aligned}$$

the infinite series on the right-hand side being convergent since q_m is a polynomial of degree m. This proves (43) and the second part of Theorem 2.

Note that by induction

$$p_m(t) = q_m(t) + \int_0^1 x^m \left(p_m(t - 1 - \log x) + p_m(t - \log x)\right) \mathrm{d}x \qquad (m \geq 2).$$

Straightforward calculation of the integrals gives expression (10) for $p_2(t_{n,k})$.

Extrema of $|\mathbb{E}(X_{n,k}-\mu_{n,k})^m|$. To prove the maximum order of $\mathbb{E}(X_{n,k}-\mu_{n,k})^m$, we consider two cases. First, when $|k-\lambda_n| \le \lambda^{2/3}$, we apply (38), so that

$$\max_{|k-\lambda_n|\le\lambda_n^{2/3}} |P_{n,k}^{(m)}| = O\left(\lambda^{-3m/2}n^m \cdot \max_{|t_{n,k}|\le\lambda_n^{2/3}} \left(t_{n,k}^m+1\right) e^{-mt_{n,k}^2/(2\lambda_n)}\right)$$
$$= O(\lambda^{-m}n^m),$$

the maximum being reached when $t_{n,k} \sim \pm\sqrt{\lambda_n}$.

On the other hand, when $|k-\lambda_n| \ge \lambda^{2/3}$, we apply the estimate (35) and bound the maximum by the sum

$$\max_{|k-\lambda_n|\ge\lambda_n^{2/3}} |P_{n,k}^{(m)}| = O\left(|v-1|^m \lambda_n^{-m/2} n^{mv} \left(\sum_{k\le\lambda_n-\lambda^{2/3}} + \sum_{k\ge\lambda+\lambda^{2/3}}\right) v^{-mk}\right).$$

Taking $v = 1-\lambda_n^{-1/3}$ in the first sum and $v = 1+\lambda_n^{-1/3}$ in the second, we obtain

$$\max_{|k-\lambda_n|\ge\lambda_n^{2/3}} |P_{n,k}^{(m)}| = O(\lambda_n^{1/3-5m/6} n^m e^{-m\lambda_n^{1/3}/2}).$$

Thus

$$\max_{1\le k<n} |\mathbb{E}(X_{n,k}-\mu_{n,k})^m| = O(\lambda_n^{-m} n^m).$$

The proof for the minimum order is similar. This proves Corollary 5.

6. Asymptotic Normality when $\alpha = 0$. The approach we use in this section relies on manipulating the recurrences of two sequences of polynomials defined from the bivariate generating functions $P_k(z,y) := \sum_n \mathbb{E}(y^{X_{n,k}})z^n$. Not only can it be applied to prove Theorem 3 but it also gives an alternative proof of the moment convergence part of Theorem 1.

Main steps. Let

$$\sigma_{n,k} := \sqrt{\frac{\lambda_n^{2k-1}}{(k-1)!^2\,(2k-1)}},$$

$X_{n,k}^* := (X_{n,k}-\lambda_n^k/k!)/\sigma_{n,k}$, and $\Lambda := \lambda_n/k$. The proof of Theorem 3 uses the following estimates.

PROPOSITION 2. *The characteristic functions of $X_{n,k}^*$ satisfy the two estimates*:

(i)

$$(44) \qquad |\mathbb{E}(e^{X_{n,k}^* i\theta}) - e^{-\theta^2/2}| = O\left(e^{-\theta^2/2}\frac{|\theta|+|\theta|^3}{\sqrt{\Lambda}} + n^{-\varepsilon}\right),$$

uniformly for $|\theta| \le \varepsilon\Lambda^{1/6}$; *and*

(ii)

(45) $$\mathbb{E}(e^{X^*_{n,k} i\theta}) = O(e^{-\theta^2/4} + n^{-\varepsilon}),$$

uniformly for $\varepsilon\Lambda^{1/6} \le |\theta| \le \varepsilon\sqrt{\Lambda}$.

Theorem 3 then follows from applying the Berry–Esseen smoothing inequality (see Petrov, 1975).

These estimates are derived by singularity analysis (see Flajolet and Odlyzko, 1990), starting from Cauchy's integral representation

$$\mathbb{E}(e^{X_{n,k} i\theta/\sigma_{n,k}}) = \frac{1}{2\pi i}\oint_{|z|=\varepsilon} z^{-n-1} P_k(z, e^{i\theta/\sigma_{n,k}})\,\mathrm{d}z.$$

We then need estimates for the generating functions P_k, and for that purpose, we introduce two sequences of polynomials and derive approximations to P_k via those for the two polynomials.

Two sequences of polynomials. By (12), the generating function P_k satisfies

$$\begin{cases} P_0(z,y) = 1 + \dfrac{yz}{1-z}, \\ P_k(z,y) = 1 + z\exp\left(\displaystyle\int_0^z \frac{P_{k-1}(t,y)-1}{t}\,\mathrm{d}t\right) \qquad (k \ge 1). \end{cases}$$

It is more convenient to work with

$$Q_k(z,s) := \frac{P_k(z,e^s)-1}{z}.$$

Then

(46) $$\begin{cases} Q_0(z,s) = \dfrac{e^s}{1-z}, \\ Q_k(z,s) = \exp\left(\displaystyle\int_0^z Q_{k-1}(t,s)\,\mathrm{d}t\right) \qquad (k \ge 1). \end{cases}$$

Now, write $L(z) := -\log(1-z)$. We define two sequences of polynomials V and W as follows:

$$\begin{aligned} Q_k(z,s) &:= \exp\left(\sum_{m\ge 0} \frac{V_{k,m}(L(z))}{m!}\, s^m\right) \\ &:= \frac{1}{1-z}\sum_{m\ge 0} \frac{W_{k,m}(L(z))}{m!}\, s^m. \end{aligned}$$

LEMMA 7. *The two sequences of polynomials satisfy the recurrences*

(47) $$\begin{cases} V_{k,m}(x) = \displaystyle\int_0^x W_{k-1,m}(t)\,\mathrm{d}t \qquad (k \ge 2), \\ W_{k,m}(x) = \dfrac{1}{m}\displaystyle\sum_{1\le j\le m} \binom{m}{j} j V_{k,j}(x) W_{k,m-j}(x) \qquad (m \ge 1), \end{cases}$$

where $V_{1,m} = x$ *for* $m \ge 0$ *and* $W_{k,0}(x) = 1$ *for* $k \ge 1$.

PROOF. The first relation follows from (46) and the second from taking the derivative with respect to s and then collecting the coefficient of s^m on both sides. □

Mean value and variance. We first rederive the mean and variance by such a VW-polynomial approach.

By (47) with $m = 1$, we obtain

$$V_{k,1}(x) = W_{k,1}(x) = \frac{x^k}{k!} \qquad (k \geq 1). \tag{48}$$

Consequently, with $x = L(z)$,

$$\mu_{n,k} = [z^n]\frac{z}{1-z} \cdot \frac{L^k(z)}{k!} = \frac{\mathsf{s}(n,k+1)}{(n-1)!},$$

which rederives (2). The asymptotic behavior of $\mu_{n,k}$ when $k = o(\lambda_n)$ is derived as follows:

$$\begin{aligned}\mu_{n,k} &= [u^k]\frac{n^u}{\Gamma(1+u)}\left(1+O(n^{-1})\right)\\ &= \frac{\lambda_n^k}{k!}\sum_{0\leq j\leq k}\frac{k!}{(k-j)!\,\lambda_n^j}\cdot[u^j]\frac{1}{\Gamma(1+u)} + O\left(\frac{\lambda_n^k}{nk!}\right)\\ &\sim \frac{\lambda_n^k}{k!}.\end{aligned}$$

For $m = 2$, we have, again by (47),

$$\begin{aligned}V_{k,2}(x) &= \int_0^x W_{k-1,2}(t)\,\mathrm{d}t = \frac{x^{2k-1}}{(k-1)!^2\,(2k-1)} + \int_0^x V_{k-1,2}(t)\,\mathrm{d}t\\ &= \sum_{0\leq j<k}\binom{2j}{j}\frac{x^{k+j}}{(k+j)!};\end{aligned} \tag{49}$$

and then

$$W_{k,2}(x) = V_{k,2}(x) + V_{k,1}^2(x) = \sum_{0\leq j\leq k}\binom{2j}{j}\frac{x^{k+j}}{(k+j)!}.$$

Hence,

$$\begin{aligned}\mathbb{E}(X_{n,k}^2) &= [z^n]\frac{z}{1-z}\cdot\sum_{0\leq j\leq k}\binom{2j}{j}\frac{L^{k+j}(z)}{(k+j)!} = \sum_{0\leq j\leq k}\binom{2j}{j}\frac{\mathsf{s}(n,k+j+1)}{(n-1)!}\\ &= \sum_{0\leq j\leq k}\binom{2j}{j}[u^{k+j}]\frac{n^u}{\Gamma(1+u)}(1+O(n^{-1}));\end{aligned}$$

see Meir and Moon (1978) and van der Hofstad et al. (2002). Now, observe that, for $k = o(\lambda_n)$,

$$\binom{2k}{k}[u^{2k}]\frac{n^u}{\Gamma(1+u)} - \left([u^k]\frac{n^u}{\Gamma(1+u)}\right)^2 = O\left(\frac{k^2\lambda_n^{2k-2}}{k!^2}\right).$$

It follows that

$$\mathbb{V}(X_{n,k}) \sim \frac{\lambda_n^{2k-1}}{(k-1)!^2\,(2k-1)} \qquad (k = o(\lambda_n)),$$

which proves the variance estimate in Theorem 3.

This line of computation can be extended to higher moments. For example, a similar reasoning for $m = 3$ yields

$$\begin{aligned} V_{k,3}(x) &= \int_0^x V_{k-1,3}(t)\,\mathrm{d}t + \int_0^x \left(3V_{k-1,2}(t)V_{k-1,1}(t) + V_{k-1,1}^3(t)\right)\mathrm{d}t \\ &= 3\sum_{0\le \ell<k}\sum_{0\le j<\ell}\binom{2j}{j}\binom{j+2\ell}{\ell}\frac{x^{k+j+\ell}}{(k+j+\ell)!} + \sum_{0\le j<k}\binom{3j}{j,\,j,\,j}\frac{x^{k+2j}}{(k+2j)!}; \end{aligned}$$

and

$$W_{k,3}(x) = 3\sum_{0\le \ell\le k}\sum_{0\le j<\ell}\binom{2j}{j}\binom{j+2\ell}{\ell}\frac{x^{k+j+\ell}}{(k+j+\ell)!} + \sum_{0\le j\le k}\binom{3j}{j,\,j,\,j}\frac{x^{k+2j}}{(k+2j)!},$$

which was used to compute $\mathbb{E}(X_{n,k} - \mu_{n,k})^3$ in Figure 1. However, the resulting expressions soon become very involved. Thus we focus directly on the asymptotics of these polynomials and not on exact expressions.

Asymptotics of the V and W polynomials. First, by (48), we have

$$V_{k,1}(x) = W_{k,1}(x) \sim \frac{x^k}{k!} \qquad (x \in \mathbb{C}),$$

for $k = o(|x|)$.

Next, by (49), we have the following estimates, for $k = o(x)$:

$$\begin{aligned} V_{k,2}(x) &= \frac{x^{2k-1}}{(k-1)!^2\,(2k-1)}\left(1 + \sum_{1\le j\le k}\frac{2k-1}{2k-j}\prod_{1\le \ell<j}\left(\frac{k-\ell}{x}\cdot\frac{k-\ell}{2k-j-\ell}\right)\right) \\ &\sim \frac{x^{2k-1}}{(k-1)!^2\,(2k-1)} \end{aligned}$$

and

$$W_{k,2}(x) = V_{k,2}(x) + V_{k,1}^2(x) \sim \frac{x^{2k}}{k!^2}.$$

The general pattern is as follows.

LEMMA 8. *If $k = o(|x|)$, where $x \in \mathbb{C}$ is large, then*

$$\text{(50)} \qquad \begin{cases} V_{k,m}(x) \sim \dfrac{x^{m(k-1)+1}}{(k-1)!^m\,(m(k-1)+1)}, \\ W_{k,m}(x) \sim \dfrac{x^{mk}}{k!^m}. \end{cases}$$

PROOF. We use induction on m. We already proved (50) for $m = 1, 2$. Assume $m \geq 3$. By (47) and induction

$$\begin{aligned} V_{k,m}(x) &= \int_0^x W_{k-1,m}(t)\,\mathrm{d}t \\ &\sim \frac{1}{m}\sum_{1\le j<m}\binom{m}{j} j \int_0^x \frac{t^{j(k-2)+1}}{(k-2)!^j\,(j(k-2)+1)}\cdot\frac{t^{(k-1)(m-j)}}{(k-1)!^{m-j}}\,\mathrm{d}t \\ &\quad + \int_0^x V_{k-1,m}(t)\,\mathrm{d}t \\ &\sim \frac{x^{(k-1)m+1}}{(k-1)!^m\,((k-1)m+1)} + \int_0^x V_{k-1,m}(t)\,\mathrm{d}t. \end{aligned}$$

Hence, by iteration,

$$\begin{aligned} V_{k,m}(x) &\sim \sum_{0\le j<k}\frac{(mj)!}{j!^m}\cdot\frac{x^{k+j(m-1)}}{(k+j(m-1))!} \\ &\sim \frac{x^{(k-1)m+1}}{(k-1)!^m\,((k-1)m+1)}. \end{aligned}$$

Moreover, by applying (47) and induction again

$$\begin{aligned} W_{k,m}(x) &\sim \frac{1}{m}\sum_{1\le j\le m}\binom{m}{j} j \frac{x^{j(k-1)+1}}{(k-1)!^j\,(j(k-1)+1)}\cdot\frac{x^{k(m-j)}}{k!^{m-j}} \\ &\sim \frac{x^{mk}}{k!^m}. \end{aligned}$$

This proves (50). □

PROOF OF PROPOSITION 2. By Cauchy's formula, we have

$$\mathbb{E}(e^{X_{n,k}i\theta/\sigma_{n,k}}) = \frac{1}{2\pi i}\oint_{|z|=\varepsilon} z^{-n} Q_k(z, i\theta/\sigma_{n,k})\,\mathrm{d}z.$$

We then deform the integration circle onto the left contour shown in Figure 3, where $\delta_n = \lambda_n^2/n$. For the

larger circle, we have

$$\frac{1}{2\pi i}\int_{|z|=1+\delta_n/n} z^{-n} Q_k\left(\frac{z, i\theta}{\sigma_{n,k}}\right)\mathrm{d}z = O\left(e^{-\lambda_n^2}\sup_{|z|=1+\delta_n/n}\left|Q_k\left(\frac{z, i\theta}{\sigma_{n,k}}\right)\right|\right).$$

Now by the estimate

$$\sigma_{n,k} = O\left(\Lambda^{-1/2}\frac{\lambda_n^k}{k!}\right)$$

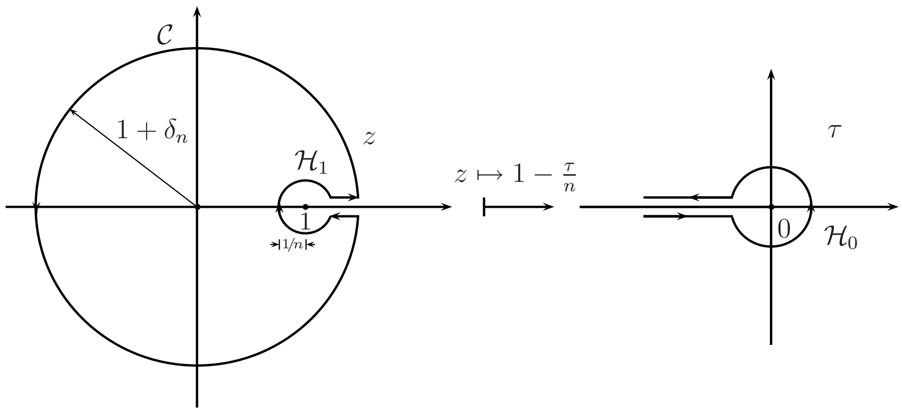


Fig. 3. The Hankel contours used to derive the asymptotics of the moments of $X_{n,k}$.

and (50), we have

$$V_{k,m}(\log(n/\omega_n))\sigma_{n,k}^{-m} = O(\Lambda^{-(m-2)/2}) \qquad (m \geq 1),$$

for any complex sequence ω_n satisfying $1 \ll |\omega_n| = O(\lambda_n^K)$. It follows that the contribution from the large circle is bounded above by

$$\begin{aligned}\frac{1}{2\pi i}\int_{|z|=1+\delta_n/n} z^{-n} Q_k\left(\frac{z, i\theta}{\sigma_{n,k}}\right) \mathrm{d}z \;&=\; O(n\lambda_n^{-2} e^{-\lambda_n^2+K\Lambda}),\\ &=\; O(n^{-\varepsilon}),\end{aligned}$$

uniformly for $|\theta| \leq \varepsilon\sqrt{\Lambda}$.

When $z \in \mathcal{H}_1$, we make the change of variables $z \mapsto 1-\tau/n$ and apply estimate (50), which gives

$$\begin{aligned}Q_k\left(1-\frac{\tau}{n}, \frac{i\theta}{\sigma_{n,k}}\right) \;=\; \frac{n}{\tau}\exp\Bigg\{&\frac{\lambda_n^k}{k!\,\sigma_{n,k}} i\theta\left(1+O\left(\frac{|\log\tau|}{\Lambda}\right)\right) - \frac{\theta^2}{2}\left(1+O\left(\frac{|\log\tau|}{\Lambda}\right)\right)\\ &+ O\left(\Lambda\sum_{m\geq 3}\frac{|\theta|^m}{m!\,\Lambda^{m/2}}\right)\Bigg\}.\end{aligned}$$

From this we deduce that if $|\theta| \leq \varepsilon\Lambda^{1/6}$, then

$$Q_k\left(1-\frac{\tau}{n}, \frac{i\theta}{\sigma_{n,k}}\right) = \frac{n}{\tau}\,\exp\left(\frac{\lambda_n^k}{k!\,\sigma_{n,k}} i\theta - \frac{\theta^2}{2}\right)\left(1+O\left((|\theta|+|\theta|^3)\frac{|\log\tau|}{\sqrt{\Lambda}}\right)\right);$$

and if $\varepsilon\Lambda^{1/6} \leq |\theta| \leq \varepsilon\Lambda^{1/2}$, then

$$\begin{aligned}Q_k\left(1-\frac{\tau}{n}, \frac{i\theta}{\sigma_{n,k}}\right) \;&=\; O\left(\frac{n}{|\tau|}\,|\tau|^{-\varepsilon} e^{-\theta^2/2+K|\theta|^3/\sqrt{\Lambda}}\right)\\ &=\; O\left(\frac{n}{|\tau|^{1-\varepsilon}}\,e^{-\theta^2/4}\right),\end{aligned}$$

for sufficiently small ε.

These estimates then yield

$$\begin{aligned}\mathbb{E}(e^{X_{n,k}^* i\theta}) &= \frac{e^{-\theta^2/2}}{2\pi i}\int_{\mathcal{H}_0}\frac{e^\tau}{\tau}\left(1+O\left((|\theta|+|\theta|^3)\frac{|\log\tau|}{\sqrt{\Lambda}}\right)\right)\\ &\quad\times\left(1+O\left(\frac{|\tau|^2}{n}\right)\right)\mathrm{d}\tau + O(n^{-\varepsilon})\\ &= e^{-\theta^2/2}\left(1+O\left(\frac{|\theta|+|\theta|^3}{\sqrt{\Lambda}}\right)\right)+O(n^{-\varepsilon}),\end{aligned}$$

uniformly for $|\theta| \le \varepsilon\Lambda^{1/6}$, where the contour $\mathcal{H}_0$ is shown in Figure 3, and similarly

$$\mathbb{E}(e^{X_{n,k}^* i\theta}) = O(e^{-\theta^2/4}+n^{-\varepsilon}),$$

uniformly for $\varepsilon\Lambda^{1/6} \le |\theta| \le \varepsilon\Lambda^{1/2}$. This completes the proof of Proposition 2. □

PROOF OF THEOREM 3. We now apply the Berry–Esseen smoothing inequality (see Petrov, 1975)

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(X_{n,k}^* < x\right)-\Phi(x)\right| = O\left(\frac{1}{\sqrt{\Lambda}}+J\right),$$

where

$$\begin{aligned}J &= \int_{-\varepsilon\sqrt{\Lambda}}^{\varepsilon\sqrt{\Lambda}}\left|\frac{\mathbb{E}\left(e^{X_{n,k}^* i\theta}\right)-e^{-\theta^2/2}}{\theta}\right|\mathrm{d}\theta\\ &= \left(\int_{|\theta|\le\Lambda^{-1/2}}+\int_{\Lambda^{-1/2}\le|\theta|\le\varepsilon\Lambda^{1/6}}+\int_{\varepsilon\Lambda^{1/6}\le|\theta|\le\varepsilon\Lambda^{1/2}}\right)\left|\frac{\mathbb{E}\left(e^{X_{n,k}^* i\theta}\right)-e^{-\theta^2/2}}{\theta}\right|\mathrm{d}\theta\\ &=: J_1+J_2+J_3.\end{aligned}$$

The integral J_1 is assessed as follows:

$$\begin{aligned}J_1 &\le \int_{|\theta|\le\Lambda^{-1/2}}\left|\frac{\mathbb{E}(e^{X_{n,k}^* i\theta})-1}{\theta}\right|\mathrm{d}\theta+\int_{|\theta|\le\Lambda^{-1/2}}\left|\frac{e^{-\theta^2/2}-1}{\theta}\right|\mathrm{d}\theta\\ &\le \mathbb{E}(X_{n,k}^{*2})\int_{|\theta|\le\Lambda^{-1/2}}|\theta|\,\mathrm{d}\theta+\int_{|\theta|\le\Lambda^{-1/2}}|\theta|\,\mathrm{d}\theta\\ &= O(\Lambda^{-1}).\end{aligned}$$

By (44), the integral J_2 satisfies

$$\begin{aligned}J_2 &= O\left(\Lambda^{-1/2}\int_{\Lambda^{-1/2}\le|\theta|\le\varepsilon\Lambda^{1/6}}(1+\theta^2)e^{-\theta^2/2}\,\mathrm{d}\theta+n^{-\varepsilon}\int_{\Lambda^{-1/2}\le|\theta|\le\varepsilon\Lambda^{1/6}}|\theta|^{-1}\,\mathrm{d}\theta\right)\\ &= O(\Lambda^{-1/2}+n^{-\varepsilon}\log\Lambda)\\ &= O(\Lambda^{-1/2}).\end{aligned}$$

The last integral J_3 is estimated by using (45)

$$\begin{aligned} J_3 &= O\left(\int_{\varepsilon\Lambda^{1/6}\le|\theta|\le\varepsilon\Lambda^{1/2}} \theta^{-1}e^{-\theta^2/4}\,\mathrm{d}\theta + n^{-\varepsilon}\log\Lambda\right)\\ &= O(\Lambda^{-1/2}). \end{aligned}$$

This proves Theorem 3. □

In particular, Theorem 3 implies and completes the case $\alpha = 0$ in Theorem 1.

An alternative proof of Theorem 1(ii). The above approach based on VW-polynomials can also be refined to give an alternative proof of Theorem 1. We outline the main steps.

First, by (47) and induction, we can prove that

$$\begin{cases} V_{k,m}(x) \sim \xi_m\left(\dfrac{k}{x}\right)\dfrac{(k/x)^{m-1}}{m}\cdot\dfrac{x^{mk}}{k!^m},\\ W_{k,m}(x) \sim \xi_m\left(\dfrac{k}{x}\right)\dfrac{x^{mk}}{k!^m}, \end{cases}$$

uniformly for $0 < k/|x| < m^{1/(m-1)}$ and large complex x, where $\xi_m(u)$ is defined recursively by

$$\xi_m(u) = \frac{1}{m-u^{m-1}}\sum_{1\le h<m}\binom{m}{h}\xi_h(u)\xi_{m-h}(u)u^{h-1} \qquad (m\ge 2),$$

with $\xi_1(u) = 1$.

Then when $k/\lambda_n \to \alpha$, $0 < \alpha < m^{1/(m-1)}$,

$$\begin{aligned} \mathbb{E}(X_{n,k}^m) &= [z^n]\frac{z}{1-z}W_{k,m}(L(z))\\ &\sim \frac{1}{2\pi i}\int_{\mathcal{H}} e^{\tau}\tau^{-1}W_{k,m}\left(\log\left(\frac{n}{\tau}\right)\right)\mathrm{d}\tau\\ &\sim \frac{\xi_m(\alpha)}{2\pi i}\int_{\mathcal{H}} e^{\tau}\tau^{-1}\frac{(\lambda_n-\log\tau)^{mk}}{k!^m}\,\mathrm{d}\tau\\ &\sim \xi_m(\alpha)\frac{\lambda_n^{mk}}{k!^m}\frac{1}{2\pi i}\int_{\mathcal{H}} e^{\tau}\tau^{-1-m\alpha}\,\mathrm{d}\tau\\ &\sim \frac{\xi_m(\alpha)}{\Gamma(1+m\alpha)}\cdot\frac{\lambda_n^{mk}}{k!^m}\\ &\sim \xi_m(\alpha)\frac{\Gamma(1+\alpha)^m}{\Gamma(1+m\alpha)}\mu_{n,k}^m, \end{aligned}$$

for a suitably chosen Hankel contour $\mathcal{H}$. It is straightforward to check, by (7), that

$$\xi_m(\alpha)\frac{\Gamma(1+\alpha)^m}{\Gamma(1+m\alpha)} = \nu_m(\alpha).$$

Note that this approach does not apply to profiles of binary search trees.

7. Profiles of Random Binary Search Trees. We consider briefly in this section random binary search trees whose profiles have been widely studied; see Drmota and Hwang (2005a) and the references therein. Our method of moments and contraction method apply. While the results for both trees are very similar, there is no range for binary search trees where the limit law of the profile is normal.

Let $Y_{n,k}$ denote the number of external nodes at distance k from the root and let $Z_{n,k}$ be the number of internal nodes at level k (root being at level 0) in a random binary search tree of n nodes (as constructed from a random permutation of n elements). Then, for $k, n \geq 1$,

$$Y_{n,k} \stackrel{d}{=} Y_{J_n,k-1} + Y^*_{n-1-J_n,k-1},$$

$$Z_{n,k} \stackrel{d}{=} Z_{J_n,k-1} + Z^*_{n-1-J_n,k-1},$$

with the initial conditions $Y_{n,0} = \delta_{n,0}$ and $Z_{n,0} = 1 - \delta_{n,0}$, where J_n is uniformly distributed over $\{0, \ldots, n-1\}$, the summands are independent and $Y_{n,k} \stackrel{d}{=} Y^*_{n,k}$, $Z_{n,k} \stackrel{d}{=} Z^*_{n,k}$. Note that $Z_{n,k} = \sum_{j>k} Y_{n,j} 2^{j-k}$.

Mean values. The expected value of $Y_{n,k}$ satisfies (see Drmota and Hwang (2005a) and the references therein)

$$\mathbb{E}(Y_{n,k}) = \frac{2^k}{n!}\,\mathsf{s}(n,k) = \frac{(2\lambda_n)^k}{\Gamma(\alpha_{n,k})k!\,n}\left(1 + O\left(\frac{1}{\lambda_n}\right)\right),$$

the O-term holding uniformly for $1 \leq k \leq K\lambda_n$.

For internal nodes, the asymptotic behavior is different:

$$\begin{aligned}\mathbb{E}(Z_{n,k}) &= \frac{2^k}{n!}\sum_{j>k}\mathsf{s}(n,j)\\ &\sim \begin{cases} 2^k - \dfrac{(2\lambda_n)^k}{(1-\alpha_{n,k})\Gamma(\alpha_{n,k})nk!}, & \text{if } \ 1 \leq k \leq \lambda_n - K\sqrt{\lambda_n};\\ 2^k\Phi(-x_{n,k}), & \text{if } \ x_{n,k} := (k-\lambda_n)/\sqrt{\lambda_n}\\ & \qquad = o((\lambda_n)^{1/6});\\ \dfrac{(2\lambda_n)^k}{(\alpha_{n,k}-1)\Gamma(\alpha_{n,k})nk!}, & \text{if } \ \lambda_n + K\sqrt{\lambda_n} \leq k \leq K\lambda_n,\end{cases}\end{aligned}$$

where the error terms in the first and the third approximations are of the form

$$O\left(\frac{(2\lambda_n)^k}{|k-\lambda_n|^2 nk!}\right),$$

and that of the middle is $O((1+|x_{n,k}|^3)/\sqrt{\lambda_n})$; see (51) below.

Note that

$$\frac{\log \mathbb{E}(Y_{n,k})}{\lambda_n} \to \alpha - 1 - \alpha\log\left(\frac{\alpha}{2}\right),$$

and the right-hand side is positive when $\alpha_- < \alpha < \alpha_+$, where $0 < \alpha_- < 1 < \alpha_+$ are the two real zeros of the equation $z - 1 - z\log(z/2)$ or $e^{(z-1)/z} = z/2$. These two constants are sometimes referred to as the *binary search tree constants* (or the fill-up level and height constants, respectively).

The limit law. Define the map

$$T\colon \mathcal{M} \to \mathcal{M}, \qquad \nu \mapsto \mathcal{L}\left(\frac{\alpha}{2}U^{\alpha-1}Z + \frac{\alpha}{2}(1-U)^{\alpha-1}Z^*\right),$$

where Z, Z^*, and U are independent and $\mathcal{L}(Z) = \mathcal{L}(Z^*) = \nu$.

The constant s is defined by $s := 2$ when $2-\sqrt{2} < \alpha < 2+\sqrt{2}$ and $1 < s < \varrho$ when $\alpha \in (\alpha_-, \alpha_+)\backslash(2-\sqrt{2}, 2+\sqrt{2})$, where $\varrho \in (1, 2]$ solves the equation $\varrho(\alpha-1)+1 = 2(\alpha/2)^{\varrho}$.

Similar to Proposition 1, we have the following properties.

PROPOSITION 3. *If $\alpha_- < \alpha < \alpha_+$, then the restriction of T to $\mathcal{M}_s(1)$ has a unique fixed point $Y(\alpha)$. In addition, $\mathbb{E}|Y(\alpha)|^{\varrho} = \infty$ for $\alpha \in (\alpha_-, \alpha_+)\backslash(2-\sqrt{2}, 2+\sqrt{2})$.*

Limit distribution when $\alpha_- < \alpha < \alpha_+$. The above estimates for the mean values of $Y_{n,k}$ and $Z_{n,k}$ say roughly that internal nodes are asymptotically full (of sizes 2^k) for the first $\lambda_n - K\sqrt{\lambda_n}$ levels, while external nodes are relatively sparse there. Observe that the second-order term of $\mathbb{E}(Z_{n,k})$ is asymptotically of the same order as $\mathbb{E}(Y_{n,k})$ when $\alpha < 1$. This suggests that we should consider

$$\bar{Z}_{n,k} := \begin{cases} 2^k - Z_{n,k}, & \text{if} \quad \alpha_- \le \alpha < 1, \\ Z_{n,k}, & \text{if} \quad 1 \le \alpha < \alpha_+. \end{cases}$$

THEOREM 5. *Let $Y(\alpha)$ and ϱ be defined as in Proposition 3. Assume that $k = \alpha\lambda_n + o(\lambda_n)$. Then, for $\alpha_- < \alpha < \alpha_+$,*

$$\frac{Y_{n,k}}{\mathbb{E}(Y_{n,k})}, \ \frac{\bar{Z}_{n,k}}{\mathbb{E}(\bar{Z}_{n,k})} \xrightarrow{d} Y(\alpha),$$

with convergence of all moments for $\alpha \in [1, 2]$ but not for α outside $[1, 2]$.

Chauvin et al. (2005) proved almost sure convergence for $Y_{n,k}/\mathbb{E}(Y_{n,k})$ when $\alpha_- < \alpha < \alpha_+$; their result is stronger than convergence in distribution but does not imply convergence of all moments.

As in Theorem 4, we can derive a convergence rate for the ζ_2-distance when $2-\sqrt{2} < \alpha < 2+\sqrt{2}$ and for ζ_s when $\alpha \in (\alpha_, \alpha_+)\backslash(2-\sqrt{2}, 2+\sqrt{2})$.

Moments of the limit law. The integral moments $\eta_m(\alpha)$ of $Y(\alpha)$ satisfy (when they exist) $\eta_0(\alpha) = \eta_1(\alpha) = 1$ and, for $m \ge 2$,

$$\begin{aligned}\eta_m(\alpha) \ &= \ \frac{(\alpha/2)^m}{m(\alpha-1)+1-2(\alpha/2)^m} \\ &\quad \times \sum_{1 \le h < m} \binom{m}{h} \eta_h(\alpha)\eta_{m-h}(\alpha) \frac{\Gamma(h(\alpha-1)+1)\Gamma((m-h)(\alpha-1)+1)}{\Gamma(m(\alpha-1)+1)}.\end{aligned}$$

Observe that the polynomial $m(z-1)+1-2(z/2)^m$ has two positive zeros z_m^- and z_m^+, where $z_m^- \in [2-\sqrt{2}, 1)$ and $z_m^+ \in (2, 2+\sqrt{2}]$ for $m \ge 2$. The two sequences of zeros

Table 1. Approximate numeric values of z_m^- and z_m^+ for $m = 2, \ldots, 11$.

	m				
	2	3	4	5	6
z_m^-	0.58578	0.69459	0.76045	0.80420	0.83509
z_m^+	3.41421	3.06417	2.86989	2.74376	2.65416
	m				
	7	8	9	10	11
z_m^-	0.85790	0.87533	0.88903	0.90006	0.90912
z_m^+	2.58668	2.53372	2.49085	2.45532	2.42531

for increasing m satisfy (see Table 1)

$$z_m^- \uparrow 1, \qquad z_m^+ \downarrow 2.$$

Thus the interval [1, 2] is the only range where convergence of all moments holds.

More precisely, $\eta_m(\alpha)$ is finite when $z_m^- < \alpha < z_m^+$ and we have convergence of the first mth moment (but not the $(m+1)$st moment) for $Y_{n,k}/\mathbb{E}(Y_{n,k})$ and $\bar{Z}_{n,k}/\mathbb{E}(\bar{Z}_{n,k})$ there. In particular, if $\alpha_- < \alpha \le 2-\sqrt{2}$ or $2+\sqrt{2} \le \alpha < \alpha_+$, then $Y(\alpha)$ has no second moment. This is consistent with the result in Drmota and Hwang (2005a).

Limit distributions when $\alpha = 1$. Note that $Y(1) = Y(2) \equiv 1$.

The following theorem states that there is a delicate difference between the limit distribution of $Y_{n,k}$ and that of $Z_{n,k}$ (properly normalized) when $\alpha = 1 + O(1/\sqrt{\lambda_n})$.

THEOREM 6. *Assume $k = \lambda_n + t_{n,k}$, where $t_{n,k} = o(\lambda_n)$. If $|t_{n,k}| \to \infty$, then*

$$\frac{Y_{n,k} - \mathbb{E}(Y_{n,k})}{2t_{n,k}(2\lambda_n)^{k-1}/(nk!)} \xrightarrow{\text{M}} Y'(1);$$

if $t_{n,k} = O(1)$, then the sequence of random variables $(Y_{n,k} - \mathbb{E}(Y_{n,k}))/\sqrt{\mathbb{V}(Y_{n,k})}$ does not converge to a fixed limit law.

For internal nodes, uniformly for $t_{n,k} = o(\lambda_n)$,

$$\frac{Z_{n,k} - \mathbb{E}(Z_{n,k})}{(2\lambda_n)^k/(nk!)} \xrightarrow{\text{M}} Y'(1).$$

Thus periodicity does not play a special role for internal nodes when $\alpha = 1$. Note that the normalizing constants differ by the factor $\alpha_{n,k} - 1 = t_{n,k}/\lambda_n$.

The limit law $Y'(1)$ can also be defined as

$$Y'(1) \stackrel{d}{=} \tfrac{1}{2}Y'(1) + \tfrac{1}{2}Y'(1)^* + 1 + \tfrac{1}{2}\log U + \tfrac{1}{2}\log(1-U),$$

with independent summands and $Y'(1) \stackrel{d}{=} Y'(1)^*$. Note that the random variables $\sum_{j\ge 0} Z_{n,j}/2^j$ have mean equal to $\sum_{1\le j\le n} j^{-1}$ and converge to $Y'(1)$ (after centered and normalized).

Since the distribution of $Y'(1)$ is uniquely characterized by its moment sequence, the convergence in distribution is also implied by the Frechet–Shohat moment convergence theorem.

The quicksort limit law when $\alpha = 2$

THEOREM 7. *Assume* $\alpha_{n,k} = 2 + t_{n,k}/\lambda_n$, *where* $t_{n,k} = o(\lambda_n)$. *If* $|t_{n,k}| \to \infty$, *then*

$$\frac{Y_{n,k} - \mathbb{E}(Y_{n,k})}{2t_{n,k}(2\lambda_n)^{k-1}/(nk!)}, \frac{Z_{n,k} - \mathbb{E}(Z_{n,k})}{2t_{n,k}(2\lambda_n)^{k-1}/(nk!)} \xrightarrow{\mathrm{M}} Y'(2);$$

if $t_{n,k} = O(1)$, *then neither of the two sequences*

$$\left\{ \frac{Y_{n,k} - \mathbb{E}(Y_{n,k})}{\sqrt{\mathbb{V}(Y_{n,k})}}, \frac{Z_{n,k} - \mathbb{E}(Z_{n,k})}{\sqrt{\mathbb{V}(Z_{n,k})}} \right\}$$

converges to a fixed limit law.

The limit law $Y'(2)$ is essentially the quicksort limit law (see Hwang and Neininger, 2002):

$$Y'(2) \stackrel{d}{=} UY'(2) + (1-U)Y'(2)^* + \tfrac{1}{2} + U\log U + (1-U)\log(1-U),$$

with independent summands on the right-hand side and $Y'(2) \stackrel{d}{=} Y'(2)^*$.

Convergence in distribution in the case when $|t_{n,k}| \to \infty$ is also implied.

The approach given in this paper gives not only the bimodality of the variances $\mathbb{V}(Y_{n,k})$ and $\mathbb{V}(Z_{n,k})$ but also the extremal (reachable) orders of $|\mathbb{E}(Y_{n,k} - \mathbb{E}(Y_{n,k}))^m|$ and $|\mathbb{E}(Z_{n,k} - \mathbb{E}(Z_{n,k}))^m|$ for $m \geq 3$ when $\alpha = 2$.

Sketch of proofs. We sketch a few steps for internal nodes, external nodes being similar and simpler.

Starting from the recurrence for the probability generating function of $Z_{n,k}$,

$$P_{n,k}(y) = \frac{1}{n} \sum_{0 \le j < n} P_{j,k-1}(y) P_{n-1-j,k-1}(y) \qquad (n \geq 2;\, k \geq 1),$$

with $P_{0,0}(y) = 1$ and $P_{n,0}(y) = y$ for $n \geq 1$, we have the recurrence for the mean value

$$\mathbb{E}(Z_{n,k}) = \frac{2}{n} \sum_{0 \le j < n} \mathbb{E}(Z_{j,k-1}) \qquad (n \geq 2;\, k \geq 1).$$

LEMMA 9. *The solution to the recurrence*

$$a_{n,k} = \frac{2}{n} \sum_{0 \le j < n} a_{j,k-1} + b_{n,k}$$

is given explicitly by

$$a_{n,k} = b_{n,k} + \frac{2}{n} \sum_{0 \le j < n} \sum_{0 \le r < k} b_{j,k-1-r} [u^r] \prod_{j < \ell < n} \left(1 + \frac{2u}{\ell}\right),$$

where $b_{0,k} := a_{0,k}$.

Then we have, by applying the exact solution with $b_{n,0} = 1$ for $n \geq 1$ and $b_{n,k} = 0$ otherwise,

$$\begin{aligned} \mathbb{E}(Z_{n,k}) &= \frac{2}{n}[u^{k-1}] \sum_{1 \leq j < n} \prod_{j < \ell < n} \left(1 + \frac{2u}{\ell}\right) \\ &= 2^k [u^{k-1}] \frac{1}{u-1} \left(\frac{\Gamma(n+u)}{\Gamma(n+1)\Gamma(u+1)} - 1 \right) \\ &= \frac{2^k}{2\pi i} \oint_{|u| = \alpha_{n,k} > 1} u^{-k-1} \frac{1}{u-1} \binom{n+u-1}{n} \, \mathrm{d}u. \end{aligned} \tag{51}$$

Thus

$$\begin{aligned} &\mathbb{E}(Z_{j,k-1}) + \mathbb{E}(Z_{n-1-j,k-1}) - \mathbb{E}(Z_{n,k}) \\ &\quad = \frac{2^k}{2\pi i} \oint_{|u|=\alpha_{n,k}} u^{-k-1} n^{u-1} \varphi\left(u, \frac{j}{n}\right) (1 + O(j^{-1} + (n-j)^{-1})) \, \mathrm{d}u, \end{aligned}$$

where

$$\varphi(u, x) = \frac{u x^{u-1} + u(1-x)^{u-1} - 2}{2\Gamma(u)(u-1)}.$$

Note that, unlike recursive trees and external nodes of binary search trees, $\varphi(1, x)$ is not zero and $\varphi(1, x) = 1 + \frac{1}{2}\log x + \frac{1}{2}\log(1-x)$. This is why there is no periodic case for internal nodes when $\alpha = 1 + O(1/\sqrt{\lambda_n})$.

All estimates required for $\mathbb{E}(Z_{n,k})$ and for its difference $\mathbb{E}(Z_{j,k-1}) + \mathbb{E}(Z_{n-1-j,k-1}) - \mathbb{E}(Z_{n,k})$ can be derived as for recursive trees. For example, we have, uniformly for $\lambda_n + K\sqrt{\lambda_n} \leq k \leq K\lambda_n$,

$$\mathbb{E}(Z_{n,k}) \sim \frac{(2\lambda_n)^k}{(\alpha - 1)\Gamma(\alpha) k!\, n}.$$

8. Conclusions. Most random trees in discrete probability or data structures have a height of order either in $\sqrt{n}$ or in $\log n$; see Aldous (1991). While profiles and other related processes defined on random trees of $\sqrt{n}$-height have been thoroughly studied in the literature (see Aldous, 1991; Drmota and Gittenberger, 1997; Kersting, 1998; Pitman, 1999; and the references therein), profiles of trees with logarithmic height have received little attention (except for digital search trees; see Aldous and Shields (1988) and Jacquet et al. (2001)). This paper shows that the phenomena exhibited in such trees are drastically different yet highly attractive. A detailed study of more general random search trees (including m-ary search trees, quadtrees, fringe-balanced binary search trees, etc.) will be given elsewhere.

Many questions remain unclear at this stage. For example, are there more "humps" or valleys for higher central moments or cumulants in the central range? Are there interesting process approximations? How to simulate the limit laws appearing in this paper? What happens when $\alpha = e$ for recursive trees and $\alpha = \alpha_-, \alpha_+$ for binary search trees? Do we still have the same convergence in distribution for $X_{n,k}/\mu_{n,k}$ when $\mu_{n,k} \to \infty$? Note that for recursive trees, $\mathbb{E}(X_{n,k}) \to \infty$ for $k \le e\lambda_n - e_1 \log \lambda_n$, where $e_1 > \frac{1}{2}$, but $\mathbb{V}(X_{n,k}) \to \infty$ for $k \le (4/\log 4)\lambda_n - e_2 \log \lambda_n$, where $e_2 > 1/(2\log 4)$. Since $4/\log 4 \approx 2.88 > e$, there is still a small range in k where the mean goes to zero but the variance goes to infinity.

Acknowledgements. Part of the work of the second author was carried out while he was visiting the Institut für Stochastik und Mathematische Informatik, J. W. Goethe-Universität (Frankfurt) and the Institut für Diskrete Mathematik und Geometrie, Technische Universität Wien; he thanks both Institutes for their hospitality and support.

References

D. Aldous (1991). The continuum random tree. II. An overview. In *Stochastic Analysis* (*Durham*, 1990), pp. 23–70. London Mathematical Society Lecture Note Series, 167. Cambridge University Press, Cambridge.

D. Aldous and P. Shields (1988). A diffusion limit for a class of randomly-growing binary trees. *Probability Theory and Related Fields* **79**:509–542.

F. Bergeron, P. Flajolet and B. Salvy (1992). Varieties of increasing trees. In *CAAP* '92 (Rennes, 1992), pp. 24–48. Lecture Notes in Computer Science, 581. Springer, Berlin.

D. Y. C. Chan, B. D. Hughes, A. S. Leong and W. J. Reed (2003). Stochastically evolving networks. *Physical Review E* **68**:06624-1–06624-24.

B. Chauvin, M. Drmota and J. Jabbour-Hattab (2001). The profile of binary search trees. *Annals of Applied Probability* **11**:1042–1062.

B. Chauvin, T. Klein, J.-F. Marckert and A. Rouault (2005). Martingales and profile of binary search trees. *Electronic Journal of Probability* **10**:420–435.

B. Chauvin and A. Rouault (2004). Connecting Yule process, bisection and binary search tree via martingales. *Journal of the Iranian Statistical Society* **3**:88–116.

H.-H. Chern and H.-K. Hwang (2001a). Phase changes in random m-ary search trees and generalized quicksort. *Random Structures and Algorithms* **19**:316–358.

H.-H. Chern and H.-K. Hwang (2001b). Transitional behaviors of the average cost of quicksort with median-of-(2t+1). *Algorithmica* **29**:44–69.

H.-H. Chern, H.-K. Hwang and T.-H. Tsai (2002). An asymptotic theory for Cauchy–Euler differential equations with applications to the analysis of algorithms. *Journal of Algorithms* **44**:177–225.

L. Devroye (1987). Branching processes in the analysis of the heights of trees. *Acta Informatica* **24**:277–298.

L. Devroye (1998). Universal limit laws for depths in random trees. *SIAM Journal on Computing* **28**:409–432.

L. Devroye and J. M. Robson (1995). On the generation of random binary search trees. *SIAM Journal on Computing* **24**:1141–1156.

L. Devroye, C. McDiarmid and B. Reed (2002). Giant components for two expanding graph processes. In *Mathematics and Computer Science*, *II* (*Versailles*, 2002), pp. 161–173. Birkhäuser, Basel.

R. P. Dobrow and J. A. Fill (1999). Total path length for random recursive trees. *Combinatorics*, *Probability and Computing* **8**:317–333.

M. Dondajewski and J. Szymański (1982). On the distribution of vertex-degrees in a strata of a random recursive tree. *Bulletin de l'Académie Polonaise des Sciences*. *Série des Sciences Mathématiques* **30**:205–209.

M. Drmota and B. Gittenberger (1997). On the profile of random trees. *Random Structures and Algorithms* **10**:421–451.

M. Drmota and H.-K. Hwang (2005a). Bimodality and phase transitions in the profile variance of random binary search trees. *SIAM Journal on Discrete Mathematics* **19**:19–45.

M. Drmota and H.-K. Hwang (2005b). Profiles of random trees: correlation and width of random recursive trees and binary search trees. *Advances in Applied Probability* **37**:321–341.

P. Erdős and P. Turán (1967). On some problems of a statistical group-theory. III. *Acta Mathematica Academiae Scientiarum Hungaricae* **18**:309–320.

D. J. Evans and R. C. Dunbar (1982). The parallel quicksort algorithm. I. Run time analysis. *International Journal of Computer Mathematics* **12**:19–55.

P. Flajolet and A. M. Odlyzko (1990). Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics* **3**:216–240.

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

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

H.-K. Hwang (1995). Asymptotic expansions for the Stirling numbers of the first kind. *Journal of Combinatorial Theory*, *Series A* **71**:343–351.

H.-K. Hwang (2005). Profiles of random trees: plane-oriented recursive trees. Preprint submitted for publication.

H.-K. Hwang and R. Neininger (2002). Phase change of limit laws in the quicksort recurrence under varying toll functions. *SIAM Journal on Computing* **31**:1687–1722.

P. Jacquet, W. Szpankowski and J. Tang (2001). Average profile of the Lempel–Ziv parsing scheme for a Markovian source. *Algorithmica* **31**:318–360.

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

G. Kersting (1998). On the height profile of a conditioned Galton–Watson tree. Preprint.

D. E. Knuth (1997). *The Art of Computer Programming*, *Volume* 1, *Fundamental Algorithms*, second edition. Addison-Wesley, Reading, MA.

H. M. Mahmoud (1991). Limiting distributions for path lengths in recursive trees. *Probability in the Engineering and Informational Sciences* **5**:53–59.

H. M. Mahmoud and R. T. Smythe (1995). Probabilistic analysis of bucket recursive trees. *Theoretical Computer Science* **144**:221–249.

A. Meir and J. W. Moon (1974). Cutting down recursive trees. *Mathematical Biosciences* **21**:173–181.

A. Meir and J. W. Moon (1978). On the altitude of nodes in random trees. *Canadian Journal of Mathematics* **30**:997–1015.

J. W. Moon (1974). The distance between nodes in recursive trees. In *Combinatorics* (edited by T. P. McDonough and V. C. Marron), pp. 125–132. London Mathematical Society Lecture Note Series, 13. Cambridge University Press, Cambridge.

H. S. Na and A. Rapoport (1970). Distribution of nodes of a tree by degree. *Mathematical Biosciences* **6**:313–329.

D. Najock and C. C. Heyde (1982). On the number of terminal vertices in certain random trees with an application to stemma construction in philology. *Journal of Applied Probability* **19**:675–680.

R. Neininger and L. Rüschendorf (2004). A general limit theorem for recursive algorithms and combinatorial structures. *Annals of Applied Probability* **14**:378-418.

V. V. Petrov (1975). *Sums of Independent Random Variables*. Springer-Verlag, Heidelberg. Translated from the Russian by A. A. Brown.

J. Pitman (1999). The SDE solved by local times of a Brownian excursion or bridge derived from the height profile of a random tree or forest. *Annals of Probability* **27**:261–283.

B. Pittel (1994). Note on the heights of random recursive trees and random m-ary search trees. *Random Structures and Algorithms* **5**:337–347.

S. T. Rachev (1991). *Probability Metrics and the Stability of Stochastic Models*. Wiley, Chichester.

U. Rösler and L. Rüschendorf (2001). The contraction method for recursive algorithms. *Algorithmica* **29**:3–33.

R. T. Smythe and H. M. Mahmoud (1995). A survey of recursive trees. *Theory of Probability and Mathematical Statistics* **51**:1–27.

J. Szymański (1990). On the maximum degree and the height of a random recursive tree. In *Random Graphs* '87 (*Poznań*, 1987), pp. 313–324. Wiley, Chichester.

G. T. Tetzlaff (2002). Breakage and restoration in recursive trees. *Journal of Applied Probability* **39**:383–390.

P. van Mieghem, G. Hooghiemstra and R. van der Hofstad (2001). On the efficiency of multicast. *IEEE/ACM Transactions on Networking* **9**:719–732.

R. van der Hofstad, G. Hooghiemstra and P. van Mieghem (2002). On the covariance of the level sizes in random recursive trees. *Random Structures and Algorithms* **20**:519–539.

V. M. Zolotarev (1976). Approximation of the distributions of sums of independent random variables with values in infinite-dimensional spaces. *Theory of Probability and its Applications* **21**:721–737.

V. M. Zolotarev (1977). Ideal metrics in the problem of approximating distributions of sums of independent random variables. *Theory of Probability and its Applications* **22**:433–449.