The Profile of Binary Search Trees [1]

Brigitte Chauvin (Université de Versailles), Michael Drmota (TU Wien)

and Jean Jabbour-Hattab (Université de Versailles)

Abstract

We characterize the limiting behaviour of the number of nodes in level k of binary search trees T_n in the central region $1.2 \log n \le k \le 2.8 \log n$. Especially we show that the width $\overline{V}_n$ (the maximal number of internal nodes at the same level) satisfies $\overline{V}_n \sim (n/\sqrt{4\pi \log n})$ as $n \to \infty$ a.s.

1 Introduction

A binary search tree is a binary tree in which each (internal) node is associated to a key, where the keys are drawn from some totally ordered set, say $\{1, 2, \ldots, n\}$. The first key is associated to the root. The next key is put to the left child of the root if it is smaller than the key of the root, and it is put to the right child of the root if it is larger than the key of the root. In this way we proceed further by inserting key by key. So starting from a permutation of $\{1, 2, \ldots, n\}$ we get a binary tree with n (internal) nodes such that the keys of the left subtree of any given node x are smaller than the key of x, and the keys of the right subtree are larger than the key of x.

[1] *AMS* 1991 *subject classifications.* 60F17, 68Q25, 05C05

Key words and phrases. Repartition of nodes for binary search trees, martingales, asymptotic series expansion, complex analysis.

Binary search trees are widely used to store (totally ordered) data, and many parameters have been discussed in the literature. (The monograph of Mahmoud [8] gives a very good overview of the state of the art.) Usually it is assumed that every permutation of $\{1, 2, \ldots, n\}$ is equally likely, and hence any parameter of binary search trees may be considered as a random variable.

An alternative way of looking at it is as a Markov chain of binary trees $(T_n)_{n\geq 0}$ describing the evolution of a binary search tree. T_n has n internal nodes and $n+1$ external nodes; especially T_0 has no internal nodes, i.e. it consists of exactly one external node which is the root, and T_1 has one internal node which is the root and two external nodes. Now T_2 is generated from T_1 by replacing one of the two external nodes by an additional internal one (with two external nodes as left and right children) with equal probability $1/2$. In that way we proceed further. T_{n+1} is generated from T_n by replacing one of the $n+1$ external nodes by an additional internal one (and two external nodes as left and right children) with equal probability $1/(n+1)$. It is an easy exercise to show that for any fixed n the probabilty distribution of T_n of this Markov chain $(T_n)_{n\geq 0}$ is exactly the same as the probability distribution induced by equally likely permuations of $\{1, 2, \ldots, n\}$ as above. However, in what follows we are mainly concerned with the Markov chain model.

It is clear that every parameter $Y = Y_T$ on binary trees T (e.g. the height, the total path length etc.) induces a sequence $(Y(n))_{n\in\mathbb{N}}$ of random variables, where $Y(n) = Y_{T_n}$.

In this paper we want to consider and denote the number of external nodes

U_k at level k, the number of internal nodes V_k at level k, and the total number of nodes $Z_k = U_k + V_k$ at this level.

Theorem 1 *We have a.s.*

$$\begin{aligned} \frac{U_k(n)}{n/\sqrt{4\pi \log n}} &= e^{-\frac{(k-2\log n)^2}{4\log n}} + \mathcal{O}\left(\frac{1}{\sqrt{\log n}}\right), \\ \frac{V_k(n)}{n/\sqrt{4\pi \log n}} &= e^{-\frac{(k-2\log n)^2}{4\log n}} + \mathcal{O}\left(\frac{1}{\sqrt{\log n}}\right), \quad \text{and} \\ \frac{Z_k(n)}{n/\sqrt{\pi \log n}} &= e^{-\frac{(k-2\log n)^2}{4\log n}} + \mathcal{O}\left(\frac{1}{\sqrt{\log n}}\right) \end{aligned}$$

as $n \to \infty$, where the error term $\mathcal{O}(1/\sqrt{\log n})$ is uniform for all $k \geq 0$.

From this theorem we directly obtain a result for the width.

Corollary 1 *Let $\overline{U}(n) = \max_{k\geq 0} U_k(n)$, $\overline{V}(n) = \max_{k\geq 0} V_k(n)$, and $\overline{Z}(n) = \max_{k\geq 0} Z_k(n)$. Then we have a.s.*

$$\begin{aligned} \frac{\overline{U}(n)}{n/\sqrt{4\pi \log n}} &= 1 + \mathcal{O}\left(\frac{1}{\sqrt{\log n}}\right), \\ \frac{\overline{V}(n)}{n/\sqrt{4\pi \log n}} &= 1 + \mathcal{O}\left(\frac{1}{\sqrt{\log n}}\right), \quad \text{and} \\ \frac{\overline{Z}(n)}{n/\sqrt{\pi \log n}} &= 1 + \mathcal{O}\left(\frac{1}{\sqrt{\log n}}\right) \end{aligned}$$

as $n \to \infty$.

It should be noted (see (3) and [3]) that in the range $k \in [(2-\sqrt{2}+\varepsilon)\log n, (2+\sqrt{2}-\varepsilon)\log n]$ we have (uniformly)

$$\mathbf{E}\, U_k(n) \sim \frac{n^{\alpha_{n,k}(1-\log(\alpha_{n,k}/2))-1}}{\sqrt{2\pi k}} = \frac{n}{\sqrt{4\pi \log n}} e^{-\frac{(k-2\log n)^2}{4\log n}} + \mathcal{O}\left(\frac{n}{\log n}\right) \tag{1}$$

and

$$\frac{(\mathbf{E}\, U_k(n))^2}{\mathbf{E}\, U_k(n)^2} \sim \frac{4\alpha_{n,k} - 2 - \alpha_{n,k}^2}{\alpha_{n,k}^2} \frac{\Gamma(2\alpha_{n,k}-1)}{\Gamma(\alpha_{n,k})^2}, \tag{2}$$

where $\alpha_{n,k} = k/\log n$. Similar estimates are true for $\mathbf{E}\, V_k(n)$ and $\mathbf{E}\, Z_k(n)$:

$$\mathbf{E}\, V_k(n) \sim \frac{\mathbf{E}\, U_k(n)}{\alpha_{n,k} - 1} \quad \text{and} \quad \mathbf{E}\, Z_k(n) \sim \frac{\alpha_{n,k}\mathbf{E}\, U_k(n)}{\alpha_{n,k} - 1}$$

as $n \to \infty$ if $k \in [(1+\varepsilon)\log n, (2+\sqrt{2}-\varepsilon)\log n]$

In view of (1) we can reformulate Theorem 1 (for $U_k(n)$) in a way that

$$\frac{U_k(n)}{\mathbf{E}\, U_k(n)} \sim 1 \quad a.s.$$

if $k = 2\log n + o(\sqrt{\log n})$ as $n \to \infty$. This concentration property is supported by the fact that in this range $(\mathbf{E}\, U_k(n))^2 \sim \mathbf{E}\, U_k(n)^2$. Since $(\mathbf{E}\, U_k(n))^2/\mathbf{E}\, U_k(n)^2 \not\to 1$ if $\alpha_{n,k} = k/\log n \to \alpha \neq 2$ we cannot expect a concentration property of this kind for $\alpha \neq 2$. Nevertheless Theorem 1 and (2) suggest that the ratio $U_k(n)/\mathbf{E}\, U_k(n)$ should behave *nicely*. In fact, we can prove the following theorem.

Theorem 2 *There exists a random analytic function $M(z)$ for $|z-1| < (\sqrt{2})^{-1}$ with $M(1) = 1$ such that for any given $\varepsilon > 0$ we have a.s.*

$$\begin{aligned}
\frac{U_k(n)}{\mathbf{E}\, U_k(n)} - M\left(\frac{k}{2\log n}\right) &\to 0, \\
\frac{V_k(n)}{\mathbf{E}\, V_k(n)} - M\left(\frac{k}{2\log n}\right) &\to 0, \quad \textit{and} \\
\frac{Z_k(n)}{\mathbf{E}\, Z_k(n)} - M\left(\frac{k}{2\log n}\right) &\to 0
\end{aligned}$$

as $n \to \infty$, uniformly for all k with $1.2\log n \le k \le 2.8\log n$.

Remark Please note that in all three cases of Theorem 2 the random analytic function $M(z)$ is in fact the same, e.g. we have a.s.

$$\frac{U_k(n)}{\mathbf{E}\, U_k(n)} - \frac{V_k(n)}{\mathbf{E}\, V_k(n)} \to 0.$$

The reason is that the second and third relation follow from the first one in view of Lemma 1 and the fact that $M(z)$ is analytic.

It is very likely that the constants 1.2 and 2.8 can be replaced by $2-\sqrt{2}+\varepsilon$ (resp. $1+\varepsilon$) and $2+\sqrt{2}-\varepsilon$, compare with Theorem 4. There are only *technical reasons* that we cannot do more.

However, for $k < (2-\sqrt{2}-\varepsilon)\log n$ and for $k > (2+\sqrt{2}+\varepsilon)\log n$ we have (see [3])

$$\frac{(\mathbf{E}\, U_k(n))^2}{\mathbf{E}\, U_k(n)^2} = \mathcal{O}(n^{-\delta})$$

(for some $\delta > 0$) which indicates that Theorem 2 need not hold in a range larger than $2-\sqrt{2} < k/\log n < 2+\sqrt{2}$.

The paper is organized in the following way. In section 2 we collect some basic facts. In section 3 convergence properties and estimates for a martingale are provided, which will be the essential tools for the proofs of Theorems 1 and 2. In section 4 the proof of Theorem 1 is presented. In fact, a more precise version (Theorem 5) is provided indicating that there is an asymptotic series expansion for $U_k(n)$. Finally, the proof of Theorem 2 is given in section 5.

2 Preliminaries

Let us start with relations between U_k, V_k, and Z_k.

Lemma 1 *The following relations hold:*

1. $Z_{k+1} = 2V_k$.

2. $Z_{k+1} - Z_k = V_k - U_k$.

3. $Z_k = \sum_{j \ge k} 2^{k-j} U_j$

The *Proof* is obvious by induction.

The main tool for the proofs of Theorems 1 and 2 is the random power series (polynomial)

$$W_n(z) = \sum_{k \ge 0} U_k(n) z^k.$$

The first observation is the following one, see [5].

Lemma 2 *The expected value of $W_n(z)$ is given by*

$$\mathbf{E}\, W_n(z) = \prod_{j=0}^{n-1} \frac{j + 2z}{j+1} = (-1)^n \binom{-2z}{n}.$$

From this representation we can read off an explicit representation for

$$\mathbf{E}\, U_k(n) = \frac{2^k}{n!} s_{n,k},$$

where $s_{n,k}$ are the (absolute) Stirling number of the first kind, in other words the number of permutations σ of n elements such that the canonical cyclic representation of σ has exactly k cycles. (It seems that this explicit formula was first observed by Lynch [7], compare also with [8]). By well known asymptotics for Stirling numbers (see [9]) we derive (for $k = \mathcal{O}(\log n)$)

$$\mathbf{E}\, U_k(n) \sim \frac{2^k (\log n)^k}{k!\, n\, \Gamma(\alpha_{n,k})} \sim \frac{n^{\alpha_{n,k}(1-\log(\alpha_{n,k}/2))-1}}{\sqrt{2\pi k}}, \tag{3}$$

where $\alpha_{n,k} = k/\log n$, as above. (In [3] an alternate approach is provided and it is shown how one can derive asymptotic expansions for second moments $\mathbf{E}\, U_k(n)^2$, too.)

Lemma 3 *For any compact set C in the complex plane $\mathbb{C}$ we have*

$$\mathbf{E}\, W_n(z) = \frac{n^{2z-1}}{\Gamma(2z)} + \mathcal{O}(n^{2\Re z-2}) \tag{4}$$

uniformly for $z \in C$ as $n \to \infty$.

Proof. By Lemma 2, $\mathbf{E}\, W_n(z)$ is just a binomial coefficient $(-1)^n\binom{-2z}{n}$. So it is clear that for any fixed z we have (4), compare with [4]. In order to show uniformity we repeat (more or less) the proof of this asymptotic formula presented in [4].

For convenience set $\alpha = 2z$. Then $\mathbf{E}\, W_n(z)$ is exactly the n-th coefficient of the binomial series $(1-x)^{-\alpha}$, resp.

$$\mathbf{E}\, W_n(z) = \frac{1}{2\pi i}\int_c (1-x)^{-\alpha} x^{-n-1}\, dx,$$

where c is a closed curve in the unit circle with winding number 1 around 0. Note that $x = 1$ is a singularity of the analytic function $f(x) = (1-x)^{-\alpha}$ and that there is an analytic continuation of $f(x)$ to $\mathbb{C} \setminus \{x \in \mathbb{R} : x \geq 0\}$. So we can replace the contour of integration c by $\tilde{c} = c_1 \cup c_2 \cup c_3 \cup c_4$, where

$$\begin{aligned}
c_1 &= \left\{ x = 1 - \frac{i + \sqrt{n} - t}{n} : 0 \leq t \leq \sqrt{n} \right\}, \\
c_2 &= \left\{ x = 1 + \frac{i + t}{n} : 0 \leq t \leq \sqrt{n} \right\}, \\
c_3 &= \left\{ x = 1 - \frac{1}{n} e^{-it} : -\frac{\pi}{2} \leq t \leq \frac{\pi}{2} \right\}, \\
c_4 &= \left\{ |x| = \left| 1 + n^{-\frac{1}{2}} + \frac{i}{n} \right| : |\arg x| \geq \arg\left(1 + n^{-\frac{1}{2}} + \frac{i}{n} \right) \right\}.
\end{aligned}$$

The easiest part is to estimate the integral over c_4:

$$\left| \frac{1}{2\pi i} \int_{c_4} (1-x)^{-\alpha} x^{-n-1}\, dx \right| \leq (1 + n^{-\frac{1}{2}})^{-n} \max\left(n^{\frac{1}{2}\Re\alpha}, (2 + n^{-\frac{1}{2}})^{-\Re\alpha} \right) e^{2\pi|\alpha|}.$$

On the remaining part $c_1 \cup c_2 \cup c_3$ we use the substitution $x = 1 + t/n$, where t varies on a corresponding curve $\gamma_1 \cup \gamma_2 \cup \gamma_3$. Furthermore we approximate x^{-n-1} by $e^{-t}(1 + \mathcal{O}(t^2/n))$. Now the integral over $c_1 \cup c_2 \cup c_3$ is given by

$$\begin{aligned}\frac{1}{2\pi i}\int_{c_1\cup c_2\cup c_3}(1-x)^{-\alpha}x^{-n-1}\,dx &= \frac{n^{\alpha-1}}{2\pi i}\int_{\gamma_1\cup\gamma_2\cup\gamma_3}(-t)^{-\alpha}e^{-t}\,dt \\ &+ \frac{n^{\alpha-2}}{2\pi i}\int_{\gamma_1\cup\gamma_2\cup\gamma_3}(-t)^{-\alpha}e^{-t}\cdot\mathcal{O}(t^2)\,dt \\ &= n^{\alpha-1}I_1 + n^{\alpha-2}I_2.\end{aligned}$$

Now I_1 approximates $1/\Gamma(\alpha)$ (by Hankel's integral representation) in the following way:

$$\begin{aligned}I_1 &= \frac{1}{\Gamma(\alpha)} + \mathcal{O}\left(\int_{\sqrt{n}}^{\infty} e^{2\pi|\alpha|}(1+t^2)^{-\frac{1}{2}\Re\alpha}e^{-t}\,dt\right) \\ &= \frac{1}{\Gamma(\alpha)} + \mathcal{O}\left(e^{2\pi|\alpha|}(1+n^2)^{-\frac{1}{2}\Re\alpha}e^{-\sqrt{n}}\right).\end{aligned}$$

Finally, I_2 can be estimated by

$$I_2 \ll \int_0^{\infty} e^{2\pi|\alpha|}(1+t^2)^{1-\frac{1}{2}\Re\alpha}\,dt + \mathcal{O}(1).$$

Thus, for any compact set C in $\mathbb{C}$ we have

$$\mathbf{E}\,W_n(z) = \frac{n^{\alpha-1}}{\Gamma(\alpha)} + \mathcal{O}(n^{\alpha-2})$$

uniformly for $z = \alpha/2 \in K$ as $n \to \infty$. □

3 Study of a Martingale

3.1 Definition and Main Result

It was shown in [5] that the ratio

$$M_n(z) = \frac{W_n(z)}{\mathbf{E}\,W_n(z)}$$

is a martingale with respect to the natural filtration $(\mathcal{F}_n)$ associated to the sequence of trees $(T_n)_{n\geq 0}$ (described in the introduction). Hence, for positive values of z, the martingale converges to an almost sure limit $M(z)$. It was proved to be bounded in L^p for $z \in \left(1-1/\sqrt{2}, 1+1/\sqrt{2}\right)$ and the limit was shown to be positive in this case. But no convergence result was established for complex values and no uniformity has been proved for the convergence over $z \in \left(1-1/\sqrt{2}, 1+1/\sqrt{2}\right)$. Now our main result concerning $M_n(z)$ reads as follows.

Proposition 1 *For any compact set $C \subseteq \{z \in \mathbb{C} : |z-1| < 1/\sqrt{2}\}$ the martingale $M_n(z)$ converges a.s. uniformly to its limit $M(z)$ (which is again an analytic function).*

We note that $M(z)$ is exactly the random analytic function appearing in Theorem 2. We also note that $M_n(1) = 1$. So there is no *probability* at $z = 1$.

In the next subsection (see Corollary 2) we will determine exactly the complex set $\mathcal{U} = B_{\mathbb{C}}(1, 1/\sqrt{2}) = \{z \in \mathbb{C} : |z-1| < 1/\sqrt{2}\}$ of L^2-convergence for this martingale and prove the regularity of the covariance function of its limit.

That will permit us to prove uniform convergence of $M_n(z)$ over the compact subsets of $\mathcal{U}$ (in subsection 3.3). The proof will follow the same path as Joffe-Lecam-Neveu in [6].

3.2 L^2-study

We start by establishing an explicit formula for the covariance function of $(W_n(z_1), W_n(z_2))$ which is valid for all $z_1, z_2 \in \mathbb{C}$ and which will be useful for section 4, too.

Lemma 4 *For all* $z_1, z_2 \in \mathbb{C}$*:*

$$\mathbf{E}\left(W_{n+1}(z_1)W_{n+1}(z_2)\right) = \sum_{j=0}^{n}\left(\tilde{\beta}_j(z_1,z_2)\prod_{k=j+1}^{n}\widetilde{\alpha}_k(z_1,z_2)\right) + \prod_{j=0}^{n}\widetilde{\alpha}_j(z_1,z_2)$$

where

$$\tilde{\alpha}_k(z_1,z_2) = 1 + \frac{2(z_1+z_2-1)}{k+1} \tag{5}$$

and

$$\tilde{\beta}_k(z_1,z_2) = (2z_1-1)(2z_2-1)\frac{\mathbf{E}\left(W_k(z_1z_2)\right)}{k+1}. \tag{6}$$

Proof. Denote by $\tilde{\Gamma}_n$ the covariance function of W_n:

$$\tilde{\Gamma}_n(z_1,z_2) = \mathbf{E}\left(W_n(z_1)W_n(z_2)\right).$$

We establish a linear recursion for $\tilde{\Gamma}_n(z_1,z_2)$. First, we recall

$$W_{n+1}(z) = W_n(z) + (2z-1)z^{k_n}.$$

Thus,

$$\tilde{\Gamma}_{n+1}(z_1,z_2) = \mathbf{E}\Bigg[\mathbf{E}\Big[\Big(W_n(z_1)+z_1^{k_n}(2z_1-1)\Big)\Big(W_n(z_2)+z_2^{k_n}(2z_2-1)\Big)\,|\mathcal{F}_n\Big]\Bigg], \tag{7}$$

so that

$$\tilde{\Gamma}_{n+1}(z_1,z_2) = \mathbf{E}\left[\sum_{k=0}^{+\infty}\frac{U_k(n)}{n+1}\Big(W_n(z_1)W_n(z_2) + W_n(z_1)z_2^k(2z_2-1) + W_n(z_2)z_1^k(2z_1-1) + (z_1z_2)^k(2z_1-1)(2z_2-1)\Big)\right]. \tag{8}$$

Hence

$$\tilde{\Gamma}_{n+1}(z_1,z_2) = \mathbf{E}\left[W_n(z_1)W_n(z_2) + W_n(z_1)\frac{(2z_2-1)W_n(z_2)}{(n+1)}\right.$$
$$\left. +W_n(z_2)\frac{(2z_1-1)W_n(z_1)}{(n+1)} + (2z_1-1)(2z_2-1)\frac{W_n(z_1z_2)}{n+1}\right], \tag{9}$$

which yields

$$\tilde{\Gamma}_{n+1}(z_1,z_2) = \tilde{\alpha}_n(z_1,z_2)\tilde{\Gamma}_n(z_1,z_2) + \tilde{\beta}_n(z_1,z_2) \tag{10}$$

for $\tilde{\alpha}$ and $\tilde{\beta}$ defined by (5) and (6).

Now, the explicit formula for $\tilde{\Gamma}_n$ follows from (10) (and $\tilde{\Gamma}_0(z_1,z_2)=1$). □

With help of Lemma 4 we can establish regularity of the covariance function of M over $\mathcal{U}^2$.

Corollary 2 *$(M_n(z))_{n\in\mathbb{N}}$ is bounded in L^2 if and only if $|z-1|<1/\sqrt{2}$. Hence, there exists a random variable $M(z)\in L^2$ such that $M_n(z) \underset{n\to\infty}{\longrightarrow} M(z)$ almost surely and in L^2 for $z\in\mathcal{U}=B_{\mathbb{C}}(1,1/\sqrt{2})$. Furthermore,*

$$\Gamma(z_1,z_2) \ := \mathbf{E}\left(M(z_1)M(z_2)\right)$$

is holomorphic over $\mathcal{U}^2\subseteq\mathbb{C}^2$.

Proof. By (5) we have

$$\prod_{k=j+1}^{n}\tilde{\alpha}_k(z_1,z_2) = \left(\frac{n}{j}\right)^{2(z_1+z_2-1)}\left(1+\mathcal{O}\left(\frac{1}{j}\right)\right)$$

and consequently (by Lemma 3)

$$\begin{aligned}\tilde{\Gamma}_n(z_1,z_2) &= (2z_1-1)(2z_2-1)\sum_{j=0}^{n}\frac{\mathbf{E}\,W_j(z_1z_2)}{j+1}\prod_{j=k+1}^{n-1}\tilde{\alpha}_k(z_1,z_2)+\prod_{j=0}^{n}\tilde{\alpha}_k(z_1,z_2)\\ &\ll \sum_{j=0}^{n} j^{2\Re(z_1z_2)-2}\left(\frac{n}{j}\right)^{2\Re(z_1+z_2-1)}+n^{2\Re(z_1+z_2-1)}\\ &\ll n^{2\Re(z_1+z_2-1)}\sum_{j=0}^{n} j^{-2\Re(z_1+z_2-z_1z_2)},\end{aligned}$$

where the notation $A \ll B$ means that there is a constant $c > 0$ such that $A \leq cB$. Thus,

$$\begin{aligned}\Gamma_n(z_1,z_2) &:= \mathbf{E}\,(M_{n+1}(z_1)M_{n+1}(z_2))\\ &= \frac{\mathbf{E}\,(W_{n+1}(z_1)W_{n+1}(z_2))}{\mathbf{E}\,W_{n+1}(z_1)\cdot\mathbf{E}\,W_{n+1}(z_2)}\\ &\ll \sum_{j=0}^{n} j^{-2\Re(z_1+z_2-z_1z_2)}.\end{aligned}$$

Obviously, we have the same lower bound. Hence, $(M_n(z))_{n\in\mathbb{N}}$ is bounded in L^2 if and only if $4\Re z - 2|z|^2 > 1$, respectively if and only if $z \in \mathcal{U}$.

Now, if $4\Re z_1 - 2|z_1|^2 > 1$ and $4\Re z_2 - 2|z_2|^2 > 1$ then we also have $2\Re(z_1 + z_2 - z_1z_2) > 1$. Thus, $\Gamma_n(z_1,z_2) \longrightarrow \Gamma(z_1,z_2)$ uniformly over the compact sets of $\mathcal{U}^2$. Since, for any n, Γ_n is holomorphic over $\left(\mathbb{C}\backslash\frac{1}{2}\mathbb{Z}^-\right)^2$, we conclude that Γ is holomorphic over $\mathcal{U}^2$. □

3.3 Proof of Proposition 1

The holomorphy of Γ proved in the previous section will give us (with help of the Kolmogoroff criterion) continuity of $M(z)$ over any parametered arc $\gamma \subseteq \mathcal{U}$. However, Kolmogoroff's criterion is not sufficient to establish directly continuity of M as a complex function.

Proposition 2 *Set* $I' := (1-\sqrt{2}/2, 1+\sqrt{2}/2)$. *Then* $(M(t))_{t\in I'}$ *has a continuous modification* $\widetilde{M}$ *such that, for any compact interval* $C \subseteq I'$

$$\mathbf{E}\left(\sup_{t\in C} |\tilde{M}(t)|^2\right) < +\infty.$$

More generally, if $\gamma : \mathbb{R} \to \mathcal{U}$ *is continuously differentiable, then there is a modification* $\tilde{M}_\gamma$ *of* $(M_n(\gamma(t)))_{t\in\mathbb{R}}$ *such that, for any compact set* C *of* $\mathbb{R}$

$$\mathbf{E}\left(\sup_{t\in C} |\tilde{M}_\gamma(t)|^2\right) < +\infty.$$

Proof. Observe that, as $M_n(z)$ is a real rational fraction, $\overline{M_n(z)} = M_n(\overline{z})$. Thus for all $z_1, z_2 \in \mathcal{U}$

$$\mathbf{E}\left(|M(z_1) - M(z_2)|^2\right) = \Gamma(z_1, \bar{z_1}) + \Gamma(z_2, \bar{z_2}) - 2\Re\left(\Gamma(z_1, \bar{z_2})\right). \tag{11}$$

Let C be a compact set of $\mathcal{U}$; since Γ is holomorphic, a local expansion of Γ up to order 2 yields

$$\Gamma(z_1, \bar{z_1}) + \Gamma(z_2, \bar{z_2}) - 2\Re\left(\Gamma(z_1, \bar{z_2})\right) \le K|z_1 - z_2|^2 \tag{12}$$

for some constant $K > 0$ and for all $z_1, z_2 \in C$. Hence, by (11) and (12)

$$\mathbf{E}\left(|M(z_1) - M(z_2)|^2\right) \le K|z_1 - z_2|^2 \tag{13}$$

for all $z_1, z_2 \in C$. Hence by Kolmogoroff's criterion (cf. [11, p. 25]), a continuous modification $\widetilde{M}$ exists and

$$\mathbf{E}\left[\left(\sup_{s,t\in C} \frac{|\widetilde{M}_t - \widetilde{M}_s|}{|t-s|^\alpha}\right)^2\right] < +\infty$$

for all $\alpha \in (0, \frac{1}{2})$. Consequently, for all compact set $C \subseteq (1 - 1/\sqrt{2}, 1 + 1/\sqrt{2})$, we have

$$\mathbf{E}\left(\sup_{t\in C} |\tilde{M}(t)|^2\right) < +\infty.$$

Now let $\gamma : \mathbb{R} \to \mathcal{U}$ be continuously differentiable. We can do the same as before with the martingales $(M_{n,\gamma}(t))_{t\in\mathbb{R}}$ for $M_{n,\gamma}(t) = M_n(\gamma(t))$. Equation (13) becomes

$$\mathbf{E}\left(|M_\gamma(t_1) - M_\gamma(t_2)|^2\right) \le K|\gamma(t_1) - \gamma(t_2)|^2 \le K'|t_1 - t_2|^2$$

for some constant $K' > 0$ depending on the compact interval $C \subseteq \mathbb{R}$ under consideration. Thus, $(M_\gamma(t))_{t\in\mathbb{R}}$ has a continuous modification $\widetilde{M}_\gamma$ such that $\mathbf{E}\left(\sup_{t\in C} |\widetilde{M}_\gamma(t)|^2\right) < +\infty$ for all compact set $C \subseteq \mathbb{R}$. □

Now uniform convergence of (M_n) follows from a theorem of vectorial martingales. (We proceed as in Joffe-Lecam-Neveu [6].)

Theorem 3 *For any compact set $C \subseteq (1 - 1/\sqrt{2}, 1 + 1/\sqrt{2})$, we have a.s.*

$$M_n \to M \qquad \textit{uniformly over } C$$

and

$$\mathbf{E}\left(\sup_{t\in C} |M_n(t) - M(t)|^2\right) \underset{n\to\infty}{\longrightarrow} 0.$$

More generally, let $\gamma : \mathbb{R} \to \mathcal{U}$ be continuously differentiable and set $M_{n,\gamma}(t) := M_n(\gamma(t))$ and $M_\gamma(t) = M(\gamma(t))$. Then the same result holds for $(M_{n,\gamma})$.

Proof. Let $[a, b] \subseteq (1 - 1/\sqrt{2}, 1 + 1/\sqrt{2})$. The modification $\widetilde{M}$ of the previous proposition is a random variable taking its values in the separable Banach space $E = \mathcal{C}([a, b], \mathbb{C})$. Let $\mathcal{E}$ be the borelian σ-field of E and $\mathcal{F}_\infty = \sigma(\mathcal{F}_n, n \ge 1)$, $\widetilde{M}$ is $\mathcal{E}|\mathcal{F}_\infty$-mesurable and is in $L^2_E = L^2(\Omega, E)$.

We will show that $\mathbf{E}(\widetilde{M}|\mathcal{F}_n)$ can be identified as $M_{n|[a,b]}$ if $M_{n|[a,b]}$ is understood as a random variable taking its values in E.

Observe that

$$\begin{aligned}\phi_t \ : \mathcal{C}([a,b],\mathbb{C}) &\rightarrow \mathbb{C}\\ X &\mapsto X(t)\end{aligned}$$

is a continuous linear form over $\mathcal{C}([a,b],\mathbb{C})$, hence $\mathbf{E}\,(\phi_t(\widetilde{M})|\mathcal{F}_n) = \phi_t(\mathbf{E}\,(\widetilde{M}|\mathcal{F}_n))$ almost surely. Saying that $\widetilde{M}$ is a modification of M means that for all $t \in [a,b]$, $\phi_t(\widetilde{M}) = M(t)$ (a.s.). Hence it follows that $M_n(t) = \mathbf{E}\,(\widetilde{M}|\mathcal{F}_n)(t)$ a.s., so that $M_n = \mathbf{E}\,(\widetilde{M}|\mathcal{F}_n)$ a.s.

We can now apply the theorem of vectorial martingales (cf. [10], Proposition V-2-6, p104), which yields

$$M_n = \mathbf{E}\left(\widetilde{M}|\mathcal{F}_n\right) \longrightarrow \mathbf{E}\,\{\widetilde{M}|\mathcal{F}_\infty\} \quad \text{a.s. and in } L^2_E.$$

Since $\widetilde{M}$ is $\mathcal{F}_\infty$-measurable we get

$$\sup_{t\in[a,b]} |M_n(t) - \widetilde{M}(t)| \underset{n\to\infty}{\longrightarrow} 0 \quad \text{a.s.} \tag{14}$$

and

$$\mathbf{E}\left(\sup_{t\in[a,b]} |M_n(t) - \widetilde{M}(t)|^2\right) \underset{n\to\infty}{\longrightarrow} 0. \tag{15}$$

Hence, the first part of the theorem is proved since (14) implies that, almost surely, $M(t)$ exists for all $t \in [a,b]$ and is equal to $\widetilde{M}(t)$.

By the previous proposition we can proceed for $(M_{n,\gamma})$ along the same lines as for (M_n). This completes the proof of Theorem 3. □

Now, since M_n is holomorphic, for all n and any $\rho < 1/\sqrt{2}$, the uniform convergence of M_n over the arc $\gamma(t) = 1 + \rho e^{it}$ implies (via Cauchy's formula) uniform convergence of M_n and all its derivatives over compact subsets of $\mathcal{U} = B_{\mathbb{C}}(1, \rho)$. Thus, we can state the following strong corollary of the previous theorem:

Corollary 3 *$M_n(z)$ and all its derivatives converge uniformly over all the compact sets of $\mathcal{U}$.*

3.4 An Almost Sure Central Limit Theorem

We now show that Proposition 1 already implies a *global version* of Theorem 2, also indicating that the range $1.2 \log n \le k \le 2.8 \log n$ is surely not *natural.*

Theorem 4 *For every $\alpha \in (2-\sqrt{2}, 2+\sqrt{2})$ we have a.s.*

$$\frac{1}{\mathbf{E}\, W_n\left(\frac{\alpha}{2}\right)} \sum_{k \le \alpha \log n + x\sqrt{\alpha \log n}} U_k(n) \left(\frac{\alpha}{2}\right)^k \to \frac{M\left(\frac{\alpha}{2}\right)}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt,$$

uniformly for $x \in \mathbb{R}$ as $n \to \infty$.

Proof. Recall that $W_n(z) = (\mathbf{E}\, W_n(z))\, M_n(z)$ and notice that we have uniformly for t real, $t = O((\log n)^{-\frac{1}{2}})$,

$$\mathbf{E}\, W_n\left(\frac{\alpha}{2} e^{it}\right) \sim \mathbf{E}\, W_n\left(\frac{\alpha}{2}\right) e^{\alpha \log n \left(it - \frac{t^2}{2}\right)}$$

and, since $M(\alpha/2)$ is positive,

$$M_n\left(\frac{\alpha}{2} e^{it}\right) = M_n\left(\frac{\alpha}{2}\right) + O((\log n)^{-\frac{1}{2}}) \sim M\left(\frac{\alpha}{2}\right) \tag{16}$$

as $n \to \infty$. Thus we can apply Levy's theorem and directly obtain the wanted result. □

3.5 More Estimates for $W_n(z)$

In order to prove Theorems 1 and 2, which are local versions of Theorem 4 it is not sufficient to know the behaviour of $M_n(z)$ near the real axis, i.e. (16). We need more precise information about the limiting behaviour of $M_n(z)$ for $1 - 1/\sqrt{2} < |z| < 1 + 1/\sqrt{2}$, resp. of $W_n(z)$. The next proposition provides a.s. estimates for $|z| = 1$ which will be needed for the proof of Theorem 1.

Proposition 3 *For any $K > 0$ there exists $\delta > 0$ such that a.s.*

$$\sup_{|z|=1, |z-1| \geq 1/\sqrt{2}-\delta} |W_n(z)| = \mathcal{O}\left(\frac{n}{(\log n)^K}\right)$$

as $n \to \infty$.

The proof of Theorem 2 requires more precise estimates.

Proposition 4 *For any $K > 0$ there exists $\delta > 0$ such that a.s.*

$$W_n(z) = \mathcal{O}\left(\frac{n^{2|z|-1}}{(\log n)^K}\right)$$

uniformly for $z \in \mathbb{C}$ with $0.6 \leq |z| \leq 1.4$, $|z-1| \geq 1/\sqrt{2} - \delta$ as $n \to \infty$.

Corollary 4 *For any $K > 0$ and $\varepsilon > 0$ we have a.s. that there exists n_0 such that for all $n \geq n_0$*

$$|W_n(z)| \leq \frac{\mathbf{E}\, W_n(|z|)}{(\log n)^K}$$

for all $z \in \mathbb{C}$ with $0.6 \leq |z| \leq 1.4$ and $(\log n)^{-\frac{1}{2}+\varepsilon} \leq |\arg z| \leq \pi$ as $n \to \infty$.

Proof. By Proposition 4 this estimate is true for $z \in \mathbb{C}$ with $0.6 \leq |z| \leq 1.4$ and $|z-1| \geq 1/\sqrt{2} - \delta$. Moreover, for $z \in \mathbb{C}$ with $|z-1| \leq 1/\sqrt{2} - \delta$ we know that $M_n(z)$ is a.s. bounded. Furthermore, it follows from Lemma 3 that, uniformly in n and t for $(\log n)^\varepsilon/\sqrt{\log n} \le |t| \le \pi$

$$|\mathbf{E}\, W_n(z_0 e^{it})| \le \mathbf{E}\, W_n(z_0) e^{-ct^2 \log n}$$

for some constant $c > 0$ (depending continuously on z_0). A combination of these two estimates proves the corollary. □

Of course, Proposition 3 is contained in Proposition 4. However, we decided to state (and prove) them separately since the proof of Proposition 3 is much easier to follow. The proof of Proposition 4 does not contain new ideas but it is more involved.

We start with an estimate for $\mathbf{E}\,|W_n(z)|^2$.

Lemma 5 *For every $\delta > 0$ we uniformly have for z with $|z-1| \le 1/\sqrt{2} - \delta$*

$$\mathbf{E}\,|W_n(z)|^2 = \mathcal{O}(n^{4\Re z - 2})$$

and for z with $1/\sqrt{2} - \delta \le |z-1| \le 1/\sqrt{2}$

$$\mathbf{E}\,|W_n(z)|^2 = \mathcal{O}(n^{4\Re z - 2} \log n)$$

as $n \to \infty$. Furthermore, let C be a compact set in the complex plane such that $|z-1| \ge 1/\sqrt{2}$ for all $z \in C$. Then

$$\mathbf{E}\,|W_n(z)|^2 = \mathcal{O}(n^{2|z|^2 - 1} \log n)$$

as $n \to \infty$, uniformly for $z \in C$.

Proof. We recall that

$$\mathbf{E}\,|M_n(z)|^2 \ll n^{4\Re z - 2} \sum_{j=1}^{n} j^{-(4\Re z - 2|z|^2)}.$$

Furthermore we have $4\Re z - 2|z|^2 > 1$ for $|z-1| < 1/\sqrt{2}$ and $4\Re z - 2|z|^2 < 1$ for $|z-1| > 1/\sqrt{2}$. Thus, for $|z-1| \le 1/\sqrt{2} - \delta$, there exists $\delta' > 0$ such that

$$\mathbf{E}\,|W_n(z)|^2 \ll n^{4\Re z-2} \sum_{j=1}^{n} j^{-1-\delta'} \ll n^{4\Re z-2}$$

and for $1/\sqrt{2} - \delta \le |z-1| \le 1/\sqrt{2}$

$$\mathbf{E}\,|W_n(z)|^2 \ll n^{4\Re z-2} \sum_{j=1}^{n} j^{-1} \ll n^{4\Re z-2} \log n$$

which prove the first part of the lemma.

Finally, for z with $|z-1| > 1/\sqrt{2}$ we obtain

$$\begin{aligned}
\mathbf{E}\,|W_n(z)|^2 &\ll n^{4\Re z-2} \frac{(n+1)^{1-4\Re z+2|z|^2} - 1}{1 - 4\Re z + 2|z|^2} \\
&\ll n^{2|z|^2-1} \frac{1 - e^{-(1-4\Re z+2|z|^2)\log(n+1)}}{1 - 4\Re z + 2|z|^2} \\
&\ll n^{2|z|^2-1} \log n.
\end{aligned}$$

This completes the proof of the lemma. □

We will also use an a.s. estimate for the derivative of $W_n(z)$.

Lemma 6 *We have for all* $z \ne 0$

$$|W_n'(z)| \le W_n'(|z|) \ll |z|^{-1} n^{2|z|-1} \log n \quad a.s.$$

Proof. Obviously, we have $|W_n'(z)| \le W_n'(|z|)$. Furthermore, it is known that $H_n \sim c \log n$ a.s., where $c = 4.31107\ldots$ is the solution greater than 2 of $c\log(2e/c) = 1$. Hence, a.s. there exists n_0 such that for $n \ge n_0$ we have $U_k(n) = 0$ if $k > (c+1)\log n$. This directly implies that a.s. we have for sufficiently large n

$$W_n'(|z|) = \sum_{k\ge 0} k U_n(k) |z|^{k-1} \le (c+1)\log n \sum_{k \ge 0} U_n(k)|z|^{k-1} = (c+1)\log n \frac{W_n(|z|)}{|z|}.$$

This proves the lemma. □

We now directly enter the *Proof of Proposition 3.* Firstly, Lemma 5 provides that for every $\varepsilon > 0$ there exists $\delta > 0$ such that for $|z| = 1$ and $|z-1| \ge 1/\sqrt{2}-\delta$

$$\mathbf{P}\left\{|W_n(z)| \ge n/(\log n)^K\right\} \le \frac{\mathbf{E}\,|W_n(z)|^2}{(n/(\log n)^K)^2} \ll \frac{(\log n)^{2K+1}}{n^{1-\varepsilon}}.$$

Now set $z_{n,j} := e^{it_{n,j}}$, where

$$t_{n,j} = \left(\arccos\frac{3}{4} - \delta\right) + \left(2\pi - 2\arccos\frac{3}{4} + 2\delta\right)\frac{j}{(\log n)^{K+1}}$$

and $j = 1, 2, \dots, [(\log N)^{K+1}]$. Thus we obtain

$$\mathbf{P}\left\{|W_n(z_{n,j})| \ge n/(\log n)^K \text{ for some } j\right\} \ll \frac{(\log n)^{3K+2}}{n^{1-\varepsilon}}.$$

Now observe that (for sufficiently small $\varepsilon > 0$) the series

$$\sum_{n\ge 0} \frac{(\log(n^2))^{3K+2}}{n^{2(1-\varepsilon)}}$$

converges. Hence, the Borel-Cantelli-lemma implies that a.s.

$$\sup_j |W_{n^2}(z_{n^2,j})| \le \frac{n^2}{(\log(n^2))^K}$$

for all but finitely many $n \ge 0$. Of course, by Lemma 6 we can interpolate between $z_{n^2,j}$. Suppose that for some z with $|z| = 1$ we have $t_{n^2,j} < \arg z < t_{n^2,j+1}$. Then we (uniformly) have

$$\begin{aligned} W_{n^2}(z) &= W_{n^2}(z_{n^2,j}) + \mathcal{O}(W'_{n^2}(1)(\log(n^2))^{-K-1}) \\ &\le \frac{n^2}{(\log(n^2))^K} + \mathcal{O}(n^2(\log(n^2))^{1-K-1}) \\ &\ll \frac{n^2}{(\log(n^2))^K}. \end{aligned}$$

So, in any case this implies that a.s.

$$\sup_{|z|=1,|z-1|\geq 1/\sqrt{2}-\delta} |W_{n^2}(z)| \leq \frac{n^2}{(\log(n^2))^K}$$

for all but finitely many $n \geq 0$. Finally we can use the relation $W_{n+1}(z) - W_n(z) = (2z-1)z^{k_n}$ to observe that

$$W_{n^2+k}(z) = W_{n^2}(z) + \mathcal{O}(k) \quad \text{for } 1 \leq k \leq (2n+1).$$

Thus, we get

$$W_{n^2+k}(z) \ll \frac{n^2}{(\log(n^2))^K} + \mathcal{O}(n) \ll \frac{n^2+k}{(\log(n^2+k))^K}$$

uniformly for $1 \leq k \leq (2n+1)$, which shows that we also have a.s.

$$\sup_{|z|=1,|z-1|\geq 1/\sqrt{2}-\delta} |W_n(z)| \ll \frac{n}{(\log n)^K}$$

for all but finitely many $n \geq 0$. □

As mentioned above the *Proof of Proposition 4* runs along the same lines as that of Proposition 3. By Lemma 5 we have for $0.6 \leq |z| \leq 1.4$ and $|z-1| \geq 1/\sqrt{2} - \delta$ (for some sufficiently small $\delta > 0$)

$$\mathbf{P}\left\{|W_n(z)| \geq n^{2|z|-1}/(\log n)^K\right\} \leq \frac{\mathbf{E}\,|W_n(z)|^2}{(n^{2|z|-1}/(\log n)^K)^2} \ll \frac{(\log n)^{2K+1}}{n^{4|z|-2|z|^2-1-0.001}}.$$

Firstly, let us consider the range $R_1 := \{z \in \mathbb{C} : |z-1| \geq 1/\sqrt{2} - \delta, 1 \leq |z| \leq 1.4\}$. We now use $\mathcal{O}((\log n)^{2K+2})$ points $z_{n,j}$ covering R_1 with maximal distance $(\log n)^{-K-1}$. Observe that for $z \in R_1$ the series

$$\sum_{n\geq 1} \frac{(\log[n^{\frac{3}{2}}])^{2K+1}}{[n^{\frac{3}{2}}]^{4|z|-2|z|^2-1-0.001}}$$

converges uniformly. Hence, by the Borel-Cantelli-lemma we have a.s.

$$\sup_j |W_{[n^{\frac{3}{2}}]}(z_{[n^{\frac{3}{2}}],j})| \le \frac{[n^{\frac{3}{2}}]^{2|z|-1}}{(\log[n^{\frac{3}{2}}])^K}$$

for all but finitely many n. By Lemma 6 we interpolate between $z_{[n^{\frac{3}{2}}],j}$ and obtain the same bound uniformly for all $z \in R_1$. Finally, we have to observe that a.s., uniformly for $n^{\frac{3}{2}} \le k \le (n+1)^{\frac{3}{2}}$

$$|W_k(z) - W_{[n^{\frac{3}{2}}]}(z)| \ll \frac{[n^{\frac{3}{2}}]^{2|z|-1}}{(\log[n^{\frac{3}{2}}])^K}. \tag{17}$$

Since

$$W_{n+k}(z) - W_n(z) = (2z-1)\sum_{l=n}^{n+k-1} z^{k_l}$$

we have to estimate z^{k_l}. We know that a.s. $H_n \le 4.3111 \log n$ (for sufficiently large n), compare with Devroye [2] or Biggins [1]. Hence it follows that a.s.

$$\max_{n^{\frac{3}{2}} \le l \le (n+1)^{\frac{3}{2}}} k_l \le 4.312 \cdot \log(n^{\frac{3}{2}}).$$

So we only have to check that

$$|2z-1|\left([(n+1)^{\frac{3}{2}}] - [n^{\frac{3}{2}}]\right)|z|^{4.312\log(n^{\frac{3}{2}})} \ll \frac{[n^{\frac{3}{2}}]^{2|z|-1}}{(\log[n^{\frac{3}{2}}])^K}.$$

Alternatively, it suffices to show that there exists $\eta > 0$ such that

$$\frac{1}{2} + \frac{3}{2}\cdot 4.312 \cdot \log|z| \le \frac{3}{2}(2|z|-1) - \eta$$

for $1 \le |z| \le 1.4$. A short inspection shows that this is true, e.g. for $\eta = 0.02$. Hence, (17) follows, which completes the proof for $z \in R_1$.

For $z \in R_2 := \{z \in \mathbb{C} : |z-1| \ge 1/\sqrt{2} - \delta, 0.6 \le |z| \le 1\}$ we have to do almost the same. Again we use the subsequence $[n^{\frac{3}{2}}]$ to apply the Borel-Cantelli lemma. In order to estimate z^{k_l} we use the fact that a.s. $k_n \geq 0.373 \log n$ (see [1]) and finally have to check that there exists $\eta > 0$ such that

$$\frac{1}{2} + \frac{3}{2} \cdot 0.373 \cdot \log |z| \leq \frac{3}{2}(2|z| - 1) - \eta$$

for $0.6 \leq |z| \leq 1$, which is again true. □

4 Proof of Theorem 1

The idea of the proof is to use a.s. expansions of $W_n(e^{it})$ in order to obtain a.s. expansions for the $U_k(n)$ via saddle point approximations.

In fact, we can be much more precise. We can prove an a.s. asymptotic series expansion for $U_k(n)/(n+1)$ in terms of $1/\sqrt{2\log n}$, in which the coefficients depend on the derivatives $M_n(z)$ at $z = 1$. In order to demonstrate how full asymptotic series expansions can be obtained we present a complete proof for the following extended version of Theorem 1 (for $U_k(n)$). It will be then clear how to proceed further.

Theorem 5 *We have a.s.*

$$\frac{U_k(n)}{n/\sqrt{4\pi \log n}} = e^{-\frac{(k-2\log n)^2}{4\log n}} \left(1 - \frac{k - 2\log n}{4\log n} + \frac{(k-2\log n)^3}{24(\log n)^2} + \frac{k - 2\log n}{2\log n} M_n'(1)\right) + \mathcal{O}\left(\frac{1}{\log n}\right)$$

as $n \to \infty$, in which the error term $\mathcal{O}(1/\log n)$ is uniform for all $k \geq 0$.

Proof. First of all we use Cauchy's formula in order to extract $U_k(n)$ from

$W_n(z)$:

$$U_k(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} W_n(e^{it})e^{-kit}\,dt.$$

Then we split the integral into two parts:

$$I_1 := \frac{1}{2\pi}\int\limits_{|t|\le \arccos\frac{3}{4}-\delta} W_n(e^{it})e^{-kit}\,dt,$$

$$I_2 := \frac{1}{2\pi}\int\limits_{\arccos\frac{3}{4}-\delta\le t\le\pi} W_n(e^{it})e^{-kit}\,dt.$$

With help of Proposition 3 we can easily estimate I_2 from above. A.s. we have

$$\begin{aligned} |I_2| &\le \frac{1}{2\pi}\int\limits_{\arccos\frac{3}{4}-\delta\le t\le\pi} |W_n(e^{it})|\,dt \\ &\ll \frac{n}{(\log n)^K}. \end{aligned}$$

For $|t| \le \arccos\frac{3}{4} - \delta$, $M_n(e^{it})$ is uniformly bounded a.s. Hence, we have by Lemma 3

$$|W_n(e^{it})| \ll ne^{2(\cos t-1)\log n} \ll ne^{-c't^2\log n}$$

for some constant $c' > 0$. Now fix some (sufficiently small) $\eta > 0$. Then we have

$$\begin{aligned} \frac{1}{2\pi}\int\limits_{(\log n)^{-(1-\eta)/2}\le|t|\le\arccos\frac{3}{4}-\delta} |W_n(e^{it})|\,dt &\ll n\int\limits_{(\log n)^{(1-\eta)/2}}^{\infty} e^{-c't^2\log n}\,dt \\ &\ll ne^{-c'(\log n)^{\eta}}. \end{aligned}$$

So it remains so consider the integral

$$I_1' := \frac{1}{2\pi}\int\limits_{|t|\le(\log n)^{-(1-\eta)/2}} W_n(e^{it})e^{-kit}\,dt.$$

For $|t| \le (\log n)^{-(1-\eta)/2}$ we have by Lemma 3 a.s. and uniformly in k

$$\begin{aligned} W_n(e^{it})e^{-kit} &= (n+1)e^{it(2\log n - k) - t^2 \log n} \\ &\quad \times \left(1 + itM_n'(1) - i\frac{t^3}{3}\log n + \mathcal{O}(t^2 + t^4 \log n)\right). \end{aligned}$$

Since

$$\int_{-\infty}^{\infty} e^{-t^2 \log n}(t^2 + t^4 \log n)\, dt \ll (\log n)^{-\frac{3}{2}}$$

and

$$\int\limits_{|t| \ge (\log n)^{-(1-\eta)/2}} e^{-t^2 \log n}(1 + t + t^3 \log n) \ll e^{-(\log n)^\eta}$$

it follows that

$$\frac{I_1'}{n+1} = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{it(2\log n - k) - t^2 \log n}\left(1 + itM_n'(1) - i\frac{t^3}{3}\log n\right)\, dt + \mathcal{O}((\log n)^{-\frac{3}{2}}).$$

Set $\gamma_{n,k} = (k - 2\log n)/\sqrt{2\log n}$. Then

$$\begin{aligned} &\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{it(2\log n - k) - t^2 \log n}\left(1 + itM_n'(1) - i\frac{t^3}{3}\log n\right)\, dt = \\ &= \frac{1}{\sqrt{2\pi 2\log n}} e^{-\frac{1}{2}\gamma_{n,k}^2}\left(1 - \frac{3\gamma_{n,k} - \gamma_{n,k}^3}{6\sqrt{2\log n}} + \frac{\gamma_{n,k}}{\sqrt{2\log n}} M_n'(1)\right). \end{aligned}$$

Thus, we finally obtain the proposed result.

We now indicate how such (a.s.) uniform estimates for $U_k(n)$ imply corresponding estimates for $Z_k(n)$ and $V_k(n)$. Recall that by Lemma 1

$$Z_k(n) = \sum_{j \ge k} 2^{k-j} U_j(n).$$

Our aim is to show that (uniformly for all $k \ge 0$)

$$\sum_{j \ge 0} 2^{-j}\left(e^{-\frac{(k+j-2\log n)^2}{4\log n}} + \mathcal{O}\left(\frac{1}{\sqrt{\log n}}\right)\right) = 2e^{-\frac{(k+j-2\log n)^2}{4\log n}} + \mathcal{O}\left(\frac{1}{\sqrt{\log n}}\right). \tag{18}$$

Obviously, (18) implies the proposed relation (in Theorem 1) for $Z_k(n)$. Since $V_k(n) = \frac{1}{2}Z_{k+1}(n)$ we get the corresponding relation (in Theorem 1) for $V_k(n)$.

So let us prove (18). First of all, note that we just have to consider j with $j \le (\log\log n)/\log 2$ since

$$\sum_{j\ge(\log\log n)/\log 2} 2^{-j} = \mathcal{O}\left(\frac{1}{\log n}\right).$$

Next, we can restrict ourselves to the range $|k-2\log n| \le \sqrt{2\log n\,\log\log n}$. Namely, if $|k-2\log n| > \sqrt{2\log n\,\log\log n}$ then

$$e^{-\frac{(k-2\log n)^2}{4\log n}} = \mathcal{O}\left(\frac{1}{\log n}\right).$$

So suppose that $j \le (\log\log n)/\log 2$ and $|k-2\log n| \le \sqrt{2\log n\,\log\log n}$. Then we have

$$\left| e^{-\frac{(k+j-2\log n)^2}{4\log n}} - e^{-\frac{(k-2\log n)^2}{4\log n}} \right| \ll e^{-\frac{(k-2\log n)^2}{4\log n}} \left(\frac{j^2}{\log n} + \frac{j|k-2\log n|}{\log n}\right).$$

Thus,

$$\begin{aligned}
&\sum_{j\le(\log\log n)/\log 2} 2^{-j} e^{-\frac{(k+j-2\log n)^2}{4\log n}} = \\
&= e^{-\frac{(k-2\log n)^2}{4\log n}} \sum_{j\le(\log\log n)/\log 2} 2^{-j} \\
&+ e^{-\frac{(k-2\log n)^2}{4\log n}} \mathcal{O}\left(\frac{1}{\log n}\sum_{j\ge 0} j^2 2^{-j} + \frac{|k-2\log n|}{\log n}\sum_{j\ge 0} j 2^{-j}\right) \\
&= 2e^{-\frac{(k-2\log n)^2}{4\log n}} + \mathcal{O}\left(\frac{1}{\log n}\right).
\end{aligned}$$

Note that this error term $\mathcal{O}(1/\log n)$ is better than the error term $\mathcal{O}(1/\sqrt{\log n})$ which is really needed. However, this is again an indication that asymptotic series expansions for $Z_k(n)$ directly follow from corresponding expansions for $U_k(n)$.

5 Proof of Theorem 2

The proof of Theorem 2 runs along similar lines. The only difference is that we now use a.s. expansions for $W_n(z_0 e^{it})$ (where $z_0 = k/(2\log n)$). For small t we use

$$M_n(z_0 e^{it}) = M_n(z_0) e^{itM_n'(z_0)/M_n(z_0) + \mathcal{O}(t^2)}$$

and $W_n(z_0e^{it}) = M_n(z_0e^{it})\mathbf{E}\, W_n(z_0e^{it})$ and for large t the estimate of Proposition 4 (resp. of its Corollary 4).

As above we have

$$U_k(n) = \frac{z_0^{-k}}{2\pi}\int_{-\pi}^{\pi} W_n(z_0e^{it})e^{-kit}\,dt.$$

Firstly, for any (sufficiently small) $\eta > 0$ we have by Corollary 4

$$|W_n(z_0e^{it})| \le \frac{\mathbf{E}\, W_n(z_0)}{\log n} \quad a.s.$$

for $0.6 \le z_0 \le 1.4$ and $(\log n)^{-(1-\eta)/2} \le |t| \le \pi$. Hence

$$\left| \frac{z_0^{-k}}{2\pi} \int\limits_{(\log n)^{-\frac{1-\eta}{2}} \le |t| \le \pi} W_n(z_0e^{it})e^{-kit}\,dt \right| \ll \frac{\mathbf{E}\, W_n(z_0)}{z_0^k \log n}.$$

Converserly, for t real with $|t| \le (\log n)^{-(1-\eta)/2}$ we have uniformly (by using Lemma 3 and $k = 2z_0\log n$)

$$\begin{aligned}
W_n(z_0e^{it})e^{-kit} &= M_n(z_0e^{it})\mathbf{E}\, W_n(z_0e^{it})e^{-kit} \\
&= M_n(z_0)\mathbf{E}\, W_n(z_0)e^{2z_0\log n(e^{it}-1)-kit}(1+\mathcal{O}(|t|)+\mathcal{O}(n^{-1})) \\
&= M_n(z_0)\mathbf{E}\, W_n(z_0)e^{-z_0\log n\, t^2}(1+\mathcal{O}((\log n)^{-\frac{1-3\eta}{2}})).
\end{aligned}$$

This implies that a.s.

$$\begin{aligned} U_k(n) &= \frac{z_0^{-k}}{2\pi} \int_{|t|\le(\log n)^{-(1-\eta)/2}} W_n(z_0e^{it})e^{-kit}\,dt + \mathcal{O}\left(z_0^{-k}\frac{\mathbf{E}\,W_n(z_0)}{\log n}\right) \\ &= M_n(z_0)z_0^{-k}\sqrt{2\pi k}\,\mathbf{E}\,W_n(z_0)(1+\mathcal{O}((\log n)^{-\frac{1-3\eta}{2}})). \end{aligned}$$

By combining (1) and Lemma 3 we have

$$z_0^{-k}\sqrt{2\pi k}\,\mathbf{E}\,W_n(z_0) \sim \mathbf{E}\,U_k(n)$$

uniformly for $0.6 \le z_0 \le 1.4$ as $n \to \infty$. Finally by Proposition 1 $M_n(z_0) \sim M(z_0)$ again uniformly. Thus, Theorem 2 follows for $U_k(n)$.

Finally, we can use this representation for $U_k(n)$ and Lemma 1 to derive the corresponding results for $Z_k(n)$ and $V_k(n)$.

Acknowledgement We want to thank M. Dekking for having suggested the L^2-method used in reference [6] of Joffe, Lecam and Neveu.

Addendum Let us mention that, due to Alain Rouault's help, the covariance result for martingale M_n in section 3.2 has been improved as follows :

$$\lim_n E\Big(M_{n+1}(z_1)M_{n+1}(z_2)\Big) = \frac{2z_1z_2+1}{2(z_1+z_2)-2z_1z_2-1} \times \frac{\Gamma(2z_1)\Gamma(2z_2)}{\Gamma\big(2(z_1+z_2)-1\big)}.$$

References

[1] J. D. Biggins. (1997). How fast does a general branching random walk spread?. Classical and modern branching processes. IMA Vol. Math. Appl. **84** 19–39.

[2] L. Devroye. (1987). Branching processes in the analysis of the height of trees. Acta Inf. **24** 277–298.

[3] M. Drmota and H. K. Hwang. Asymptotic expansions for second moments of the profile of binary search trees. manuscript.

[4] P. Flajolet and A. M. Odlyzko. (1990). *Singularity analysis of generating functions*. SIAM J. Discrete Math. **3** 216–240.

[5] J. Jabbour-Hattab. (1999). *Martingales and large deviations for the binary search trees*. submitted to Random Structure and Algorithms.

[6] A. Joffe, L. Lecam, J. Neveu. (1973). *Sur la loi des grands nombres pour des variables aléatoires attachées à un arbre dyadique*. C. R. Acad. Sci., Paris. Ser. A **277** 963–964.

[7] W. Lynch. (1965). *More combinatorial problems on certain trees.* The Computer J. **7** 299–302.

[8] H. M. Mahmoud. (1992). *Evolution of Random Search Trees*. John Wiley & Sons, New York.

[9] L. Moser and M. Wyman. (1958). *Asymptotic development of the Stirling numbers of the first kind*. J. London Math. Soc. **33** 133–146.

[10] J. Neveu. (1972). *Martingales à temps discret*. Masson & Cie.

[11] D. Revuz and M. Yor. (1990) *Continuous Martingales and Brownian Motion*. Springer.

LAMA - CNRS EP1755
Université de Versailles
78035 Versailles Cedex
France
E-mail: {jabbour,chauvin}@math.uvsq.fr

Department of Geometry
Technical University of Vienna
1040 Vienna
Austria
michael.drmota@tuwien.ac.at