Discrete Mathematics and Theoretical Computer Science **6**, 2004, 409–424

The distribution of m-ary search trees generated by van der Corput sequences

Wolfgang Steiner[†]

TU Wien, Institut für Diskrete Mathematik und Geometrie, Wiedner Hauptstraße 8-10, 1040 Wien, Austria
Universität Wien, Institut für Mathematik, Strudlhofgasse 4, 1090 Wien, Austria
email: `steiner@dmg.tuwien.ac.at`

received Sep 18, 2003, *accepted Jan 4, 2004*.

We study the structure of m-ary search trees generated by the van der Corput sequences. The height of the tree is calculated and a generating function approach shows that the distribution of the depths of the nodes is asymptotically normal. Additionally a local limit theorem is derived.

Keywords: m-ary search trees, van der Corput sequence, tree height, central limit theorem, generating function

1 Introduction

In the last years, the height of binary search trees generated by sequences which are uniformly distributed modulo 1 has been studied. Devroye [3] has shown for the Weyl sequences $\{n\alpha\}$ that the height of the tree with its first N elements satisfies

$$H(N) \sim \frac{12}{\pi^2} \log N \log\log N$$

for almost all $\alpha \in (0,1)$. The minimal height is attained for the golden mean $\alpha = (\sqrt{5}+1)/2$, $H(N) \sim \log N / \log\alpha$, and the maximal height is almost as large as the theoretical maximum for binary search trees. More precisely, for every sequence $(c_N)_{N\geq 1}$ which decreases monotonically from 1 to 0, we have some α such that $H(N) \geq c_N N$ infinitely often (Devroye and Goudjil [5]).

For general uniformly distributed sequences modulo 1, Dekking and van der Wal [1] have shown

$$H(N) = o(N)$$

and that, for every $c \geq 1/\log 2$, we have sequences with $H(N) \sim c \log N$.

[†]This research was supported by the Austrian Science Foundation FWF, grant S8302-MAT.

Devroye and Neininger [6] studied random suffix search trees, which are binary search trees generated by the suffixes $S_n = 0.B_nB_{n+1}B_{n+2}\dots$ of independent identically distributed random q-ary digits $B_1, B_2, \dots$ for some $q \geq 2$. For these trees, the expected value of the depth of S_N is given by

$$\mathbb{E}\, d(S_N) = 2\log N + O\left(\log^2 \log N\right).$$

Note that the suffixes are uniformly distributed modulo 1 with probability 1.

For random binary search trees of size N, it was shown by several authors that the expected value of the depth of a node is again $2\log N + O(1)$ and we know from Mahmoud and Pittel [9] and Devroye [2] that the distribution of the depths is asymptotically normal with variance $2\log N$.

A natural generalization of binary search trees are m-ary search trees, which are constructed by placing the first $m-1$ keys in the root, sorted in increasing order from left to right, then guiding a subsequent key to the ℓth subtree of the root, $1 \leq \ell \leq m$, if that key is greater than exactly $\ell - 1$ of the root keys. In the ℓth subtree, the newcomer is subjected recursively to the same procedure until a node with less than $m-1$ keys is found.

Mahmoud and Pittel [10] showed that the distribution of the depths in random m-ary search trees is asymptotically normal with mean value $\frac{1}{\sum_{j=2}^m 1/j}\log N$ and variance $\frac{(\sum_{j=2}^m 1/j)^3}{\sum_{j=2}^m 1/j^2}\log N$. This and other limit laws for various kinds of trees can also be found in Devroye [4].

In this article, we consider m-ary search trees generated by particular uniformly distributed sequences modulo 1, the van der Corput sequences $(\phi_q(n))_{n\geq 1}$, where we omit $n = 0$ for convenience. Let

$$n = \sum_{j\geq 0} \varepsilon_j(n) q^j$$

be the (unique) q-ary digital expansion with digits $\varepsilon_j(n) \in \{0, 1, \dots, q-1\}$ for some integer $q \geq 2$. Then the van der Corput sequence to the base q is defined by the radical-inverse function

$$\phi_q(n) = \sum_{j\geq 0} \varepsilon_j(n) q^{-j-1}.$$

Let $d(n)$ denote the depth of the node containing the nth element of the sequence. Besides the height $H(N) = \max_{n\leq N} d(n)$, we will study the distribution of $d(n)$. To that end, we define a sequence of (discrete) random variables X_N by

$$\mathbb{P}\{X_N = k\} = \frac{a_{Nk}}{N} \quad \text{with} \quad a_{Nk} = |\{n \leq N : \ d(n) = k\}|,$$

i.e., X_N is the depth of a key randomly chosen among the first N keys inserted into the tree.

2 Results

Throughout the paper let $M = \lfloor \log_q m \rfloor$ be the integer part of the logarithm to the base q of m.

Theorem 1 *The height of the tree is given by*

$$H(N) = \frac{1}{M + h_{q,M}(\frac{m}{q^M})}\log_q N + O(1), \tag{1}$$

where $h_{q,M}(x)$ is determined by the sequence

$$\mu_j = \begin{cases} 0 & \text{if } x < q - \eta_j \\ 1 & \text{else} \end{cases} \quad \text{with } \eta_0 = 0,\ \eta_{j+1} = \begin{cases} \eta_j + m - 1 & \text{if } M + \mu_j = 0 \\ \left\lfloor \frac{xq}{q^{\mu_j}(q-\eta_j)} \right\rfloor - 1 & \text{else.} \end{cases}$$

Let J, p be the lengths of the preperiod and the period of μ_j, i.e., $\mu_{j+p} = \mu_j$ for all $j > J$. Then we have

$$h_{q,M}(x) = \frac{1}{p} \sum_{j=J+1}^{J+p} \mu_j, \tag{2}$$

The functions $h_{q,M} : [1,q) \to [0,1)$ are monotonically increasing functions.

Note that $h_{q,M} = h_{q,M'}$ for all $M, M' > 0$.

Theorem 2 *Expected value and variance of X_N are given by*

$$\mathbb{E}X_N = \frac{1}{N} \sum_{n \le N} d(n) = \mu \log_q N + O(1) \tag{3}$$

$$\mathbb{V}X_N = \frac{1}{N} \sum_{n \le N} (d(n) - \mathbb{E}X_N)^2 = \sigma^2 \log_q N + O(1) \tag{4}$$

with constants μ and σ given by (16) and (18). For $m = q^M$, $m = 2$ (binary search trees) and $q = 2$ (the binary van der Corput sequence), we have simple formulae for μ and σ:

$$m = q^M: \qquad \mu = \frac{1}{M}, \qquad \sigma = 0 \tag{5}$$

$$m = 2: \qquad \mu = (q-1)\left(\frac{1}{2} + \frac{1}{q}\right), \qquad \sigma^2 = \frac{(q-1)(q-2)(q^2+3q-6)}{12q^2} \tag{6}$$

$$q = 2: \qquad \mu = \frac{1}{M + \frac{m}{2^M} - 1}, \qquad \sigma^2 = \frac{\left(\frac{m}{2^M} - 1\right)\left(2 - \frac{m}{2^M}\right)}{\left(M + \frac{m}{2^M} - 1\right)^3} \tag{7}$$

For $m \neq q^M$, we have $\mu \in (\frac{1}{M+1}, \frac{1}{M})$ and $\sigma^2 > 0$.

The main result concerns the distribution properties of X_N. We prove asymptotic normality in the weak sense and provide a local limit law.

Theorem 3 *If $m \neq q^M$, then we have, for every $\delta > 0$,*

$$\frac{1}{N} |\{n \le N : d(n) < \mathbb{E}X_N + x\mathbb{V}X_N\}| = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt + O\left((\log N)^{-1/2+\delta}\right) \tag{8}$$

uniformly for all real x as $N \to \infty$ and

$$|\{n \le N : d(n) = k\}| = \frac{N}{\sqrt{2\pi \mathbb{V}X_N}} \left(\exp\left(-\frac{(k - \mathbb{E}X_N)^2}{2\mathbb{V}X_N} \right) + O\left((\log N)^{-1/2+\delta}\right) \right) \tag{9}$$

uniformly for all nonnegative integers k as $N \to \infty$.

The (easy) case $m = q^M$ is treated in Section 3. The crucial part for all other cases is contained in Section 4, where the structure of the tree is analyzed and its generating function is calculated. Section 5 is devoted to the height of the tree, i.e., Theorem 1. Formulae for mean value and variance are derived in Section 6. The two parts of Theorem 3 are proved in Sections 7 and 8. These proofs are adapted from Drmota and Gajdosik [7]. Finally, the values of μ and σ for binary search trees and the binary van der Corput sequence are calculated in Section 9.

3 $m = q^M$

For $m = q^M$, the root contains the elements $\{1,\dots,q^M-1\}$ and its keys are just the q^M-1 possibilities for $\sum_{j=0}^{M-1} c_j q^{-j-1}$, $c_j \in \{0,\dots,q-1\}$ with $(c_0,\dots,c_{M-1}) \neq (0,\dots,0)$. Let us call prefix of the nth key a prefix of the corresponding digit word $\varepsilon_0(n)\varepsilon_1(n)\dots$. Then the ℓth subtree contains all keys with prefix $\varepsilon_0(\ell)\dots\varepsilon_{M-1}(\ell)$ and in the root of the ℓth subtree we have the q^M-1 keys $\sum_{j=0}^{M-1}\varepsilon_j(\ell)q^{-j-1} + \sum_{j=M}^{2M-1} c_j q^{-j-1}$ with $c_j \in \{0,\dots,q-1\}$ and $(c_M,\dots,c_{2M-1}) \neq (0,\dots,0)$. Thus, the depth of the nth element is k, i.e., $d(n) = k$, if and only if $q^{kM} \le n < q^{(k+1)M}$. With $L = \lfloor \log_{q^M} N \rfloor = \lfloor (\log_q N)/M \rfloor$, expected value and variance are given by

$$\frac{1}{N}\sum_{n<N} d(n) = \frac{1}{N}\left(\sum_{k=0}^{L-1}(q^M-1)q^{kM}k + (N-q^{LM})L\right) = L - \frac{q^M}{q^M-1}\frac{q^{LM}-1}{N} = \frac{1}{M}\log_q N + O(1),$$

$$\begin{aligned}
&\frac{1}{N}\sum_{n<N}\left(d(n) - \frac{1}{N}\sum_{n<N} d(n)\right)^2 \\
&= \frac{1}{N}\left(\sum_{k=0}^{L-1}(q^M-1)q^{kM}\left(k-L+\frac{q^M(q^{LM}-1)}{(q^M-1)N}\right)^2 + (N-q^{LM})\left(\frac{q^M(q^{LM}-1)}{(q^M-1)N}\right)^2\right) \\
&= \frac{1}{N}\left(\frac{(q^{LM}-1)q^M(q^M+1)}{(q^M-1)^2} - \frac{(q^{LM}-1)^2q^{2M}}{(q^M-1)^2}\left(\frac{1}{N}+\frac{1}{N^2}\right) - L^2 - \frac{2Lq^M}{q^M-1}\left(1-\frac{q^{LM}-1}{N}\right)\right) = O(1),
\end{aligned}$$

and the results for $m = q^M$ are proved.

4 Generating function

Define the bivariate generating function of the tree by

$$B(z,u) = \sum_{j\ge 0} b_j(z)u^j$$

with

$$b_j(z) = \sum_{k\ge 0} b_{jk} z^k, \quad b_{jk} = a_{q^{j+1}k} - a_{q^j k} = |\{n:\ q^j \le n < q^{j+1}, d(n) = k\}|,$$

i.e., z counts the depth of the elements and u the time of their insertion in the tree.

Lemma 1 *We have*

$$B(z,u) = \frac{Q(z,u)}{1-P(z,u)}$$

for some polynomials $Q(z,u)$ and $P(z,u)$ determined by (13) and

$$B(z,u) = \frac{G(z,u)}{1-F(z,u)}$$

for some analytic functions (in the domain $D_\rho = \{(z,u) : |z| < 1+\rho, |u| < 1/q+\rho\}$ for some $\rho > 0$)

$$F(z,u) = \sum_{j=M}^{\infty}\sum_{k=1}^{\infty} f_{jk} z^k u^j, \qquad G(z,u) = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} g_{jk} z^k u^j$$

with $f_{jk} \geq 0$, $g_{jk} \geq 0$ for all j,k. Assume, w.l.o.g, $\gcd(1-P(z,u), Q(z,u)) = 1$.

Proof. For $m = q^M$, the considerations of the previous section give

$$B(z,u) = \sum_{j=0}^{M-1}(q-1)q^j u^j + q^M z u^M B(z,u) = \frac{\sum_{j=0}^{M-1}(q-1)q^j u^j}{1-q^M z u^M}.$$

For $m > q^M$, the minimal key in the root is $\phi_q(q^M) = .0^M 1 = q^{-M-1}$. The leftmost subtree contains therefore all keys $\phi_q(n) < .0^M 1$, i.e., those with prefix 0^{M+1}. As in the case $m = q^M$, this subtree has the same shape as the whole tree, and its generating function is $zu^{M+1}B(z,u)$.

If $m-1 \geq 2q^M$ and $q > 2$, then the second smallest key in the root is $\phi_q(2q^M) = .0^M 2$. In this case, the second subtree contains all keys with prefix $0^M 1$ (except the key $.0^M 1$) and its generating function is again $zu^{M+1}B(z,u)$.

The other possibility for the second smallest key ($m > 2$) is $\phi_q(q^{M-1}) = .0^{M-1}1$. Then the second subtree contains all keys with prefix 0^M satisfying $\phi_q(n) > .0^M 1$. For $q = 2$, this is the same tree as in the latter case and its generating function is $zu^{M+1}B(z,u)$. For $q > 2$, this means (if we omit the prefix 0^M since it does not change the structure) that we start with $n = 2$ and consider just those n with $\varepsilon_0(n) \geq 1$. Call this tree T_1. In general, let T_i, $0 \leq i < q-1$, be the tree generated by the van der Corput sequence starting with $n = i+1$ and omitting the n's with digit $\varepsilon_0(n) < i$, i.e., just take the keys $\phi_q(n) > .i = i/q$. Denote its generating function by $B_i(z,u)$. Then the second subtree contributes $zu^M B_1(u,z)$ to $B(z,u)$. Furthermore note that T_0 is the whole tree and $B_0(z,u) = B(z,u)$.

The other subtrees have a similar structure. If $n + q^M \leq m-1$ or $\varepsilon_M(n) = q-1$, then the contribution to the generating function is $zu^{M+1}B(z,u)$. In the other cases, the tree is of type $T_{\varepsilon_M(n)}$ and the contribution is $zu^M B_{\varepsilon_M(n)}(z,u)$.

We have thus, for $m < (q-1)q^M$,

$$\begin{aligned} B(z,u) = B_0(z,u) = &\sum_{j=0}^{M-1}(q-1)q^j u^j + (m-q^M)u^M + (m-q^M)zu^{M+1}B_0(z,u) \\ &+ (m-\varepsilon_M(m)q^M)zu^M B_{\varepsilon_M(m)}(z,u) + ((\varepsilon_M(m)+1)q^M - m)zu^M B_{\varepsilon_M(m)-1}(z,u) \end{aligned}$$

and, for $\varepsilon_M(m) = q-1$,

$$B_0(z,u) = \sum_{j=0}^{M-1}(q-1)q^j u^j + (m-q^M)u^M + (2m-q^{M+1})zu^{M+1}B_0(z,u) + (q^{M+1}-m)zu^M B_{q-2}(z,u)$$

The sequence constituting T_i is $(i+1, i+2, \dots, q-1, q+i, q+i+1, \dots, 2q-1, 2q+i, \dots)$. Denote its mth element by m_i and set $M_i = \lfloor \log_q m_i \rfloor$. Note that the $c(q-i)q^k$th element of the sequence is $cq^{k+1}+i$ $(c<q)$. Hence we have $M_i = \lfloor \log_q \frac{mq}{q-i} \rfloor \in \{M, M+1\}$ and $\varepsilon_{M_i}(m_i) = \lfloor \frac{m_i}{q^{M_i}} \rfloor = \lfloor \frac{mq}{q^{M_i}(q-i)} \rfloor$ for $M_i > 0$, $\varepsilon_{M_i}(m_i) = m_i = m+i$ for $M_i = 0$.

The generating function of T_i is, for $M_i > 0$ and $\varepsilon_{M_i}(m_i) < q-1$,

$$\begin{aligned} B_i(z,u) = {} & \sum_{j=0}^{M_i-1} (1-i/q)(q-1)q^j u^j - i/q + (m-(1-i/q)q^{M_i})u^{M_i} + (m-(1-i/q)q^{M_i})zu^{M_i+1}B_0(z,u) \\ & + (m-\varepsilon_{M_i}(m_i)(1-i/q)q^{M_i})zu^{M_i}B_{\varepsilon_{M_i}(m_i)}(z,u) + ((\varepsilon_{M_i}(m_i)+1)(1-i/q)q^{M_i}-m)zu^{M_i}B_{\varepsilon_{M_i}(m_i)-1}(z,u), \end{aligned} \tag{10}$$

for $\varepsilon_{M_i}(m_i) = q-1$,

$$\begin{aligned} B_i(z,u) = {} & \sum_{j=0}^{M_i-1} (1-i/q)(q-1)q^j u^j - i/q + (m-(1-i/q)q^{M_i})u^{M_i} \\ & + (2m-(1-i/q)q^{M_i+1})zu^{M_i+1}B_0(z,u) + ((1-i/q)q^{M_i+1}-m)zu^{M_i}B_{q-2}(z,u) \end{aligned} \tag{11}$$

and finally for $M_i = 0$,

$$B_i(z,u) = m-1+(m-1)zuB_0(z,u) + zB_{i+m-1}(z,u). \tag{12}$$

Hence $B_i(z,u) = \sum_{j=0}^{q-2} zP_{ij}(u) + Q_i(u)$ for some polynomials $P_{ij}(u)$ and $Q_i(u)$, i.e.,

$$\begin{pmatrix} B_0(z,u) \\ \vdots \\ B_{q-2}(z,u) \end{pmatrix} = zA(u) \begin{pmatrix} B_0(z,u) \\ \vdots \\ B_{q-2}(z,u) \end{pmatrix} + \begin{pmatrix} Q_0(u) \\ \vdots \\ Q_{q-2}(u) \end{pmatrix}$$

with $A(u) = (P_{ij}(u))_{0 \le i,j \le q-2}$ and $B(z,u) = B_0(z,u)$ is given by

$$B(z,u) = (1,0,\dots,0)(I_{q-1} - zA(u))^{-1} \begin{pmatrix} Q_0(u) \\ \vdots \\ Q_{q-2}(u) \end{pmatrix} = \sum_{k=0}^{\infty} z^k (1,0,\dots,0) A(u)^k \begin{pmatrix} Q_0(u) \\ \vdots \\ Q_{q-2}(u) \end{pmatrix} \tag{13}$$

where I_{q-1} denotes the $(q-1)$-dimensional identity matrix, and the first equation determines $P(z,u), Q(z,u)$.

The functions $F(z,u), G(z,u)$ are obtained by recurrently replacing the $B_i(z,u)$'s, $i>0$, in the equation for $B_0(z,u)$ by their expressions given in (10)–(12),

$$\begin{aligned} B_0(z,u) &= Q_0(u) + \sum_{i=0}^{q-2} zP_{0i}(u)B_i(z,u) \\ &= Q_0(u) + zP_{00}(u)B_0(z,u) + \sum_{i=1}^{q-2} zP_{0i}(u) \left(Q_i(u) + P_{i0}(u)B_0(z,u) + \sum_{j=1}^{q-2} P_{ij}(u)B_j(z,u) \right) = \cdots \end{aligned}$$

For $M > 0$, we have $\sum_{i=1}^{q-2} P_{ij}(1/q) < 1$ for all $i > 0$. Thus $(1+\rho)\sum_{i=1}^{q-2} P_{ij}(1/q+\rho) \le 1$ for all $i > 0$ and some $\rho > 0$. Then, for $(z,u) \in D_\rho$, the coefficients of $B_i(z,u)$ tend to 0 in the above expression of $B_0(z,u)$ and we obtain

$$B_0(z,u) = G(z,u) + F(z,u)B_0(z,u) = \frac{G(z,u)}{1-F(z,u)}.$$

For $M = 0$, we have $\sum_{i=1}^{q-2} P_{ij}(1/q) < 1$ only for $i \ge q-m$ and $\sum_{i=1}^{q-2} P_{ij}(u) = P_{i,i+m-1}(u) = 1$ else. This means that we replace $B_i(z,u)$ by $Q_i(u) + zP_{i,i+m-1}(u)B_{i+m-1}(z,u)$ at most $\lceil \frac{q-m-1}{m-1} \rceil$ consecutive times before we have $\sum_{i=1}^{q-2} P_{ij}(1/q) < 1$. Hence choosing ρ such that $(1+\rho)^{\lceil \frac{q-m-1}{m-1} \rceil} \sum_{j=1}^{q-2} P_{ij}(1/q+\rho) \le 1$ for all $i \ge q-m$ gives the same result as for $M > 0$.

For the same reasons, $F(z,u)$ and $G(z,u)$ are analytic in D_ρ. The f_{jk} and g_{jk} are nonnegative because the coefficients of $Q_i(u)$ and $P_{ij}(u)$ are positive. □

5 Height

For every $k \ge 0$, we look for the minimal j such that $b_{jk} \ne 0$. Since all $Q_i(u)$ have a constant term, this is, by (13), the minimal exponent of u in the first row of $A(u)^k$. The ℓth element of this row is the sum of $P_{0s_1}(u)P_{s_1s_2}(u)\dots P_{s_{k-1}s_k}(u)$ over all sequences $s_1,\dots,s_k$ with $s_k = \ell$.

Recall from the last section that the ith row of $A(u)$ consists of terms with exponent M_i and (in the majority of the cases) $M_i + 1$ with $M_i \in \{M, M+1\}$. For $i < i'$, we have either $M_i < M_{i'}$ or $M_i = M_{i'}$, $\varepsilon_{M_i}(m_i) \le \varepsilon_{M_{i'}}(m_{i'})$. Thus the minimal exponent of u in the first row of $A(u)^k$ can be found by recursively choosing the minimal s_1 such that $P_{0s_1}(u)$ has a term u^{M_0}, the minimal s_2 such that $P_{s_1s_2}(u)$ has a term $u^{M_{s_1}}$ and so on. Hence $s_{i+1} = \varepsilon_{M_{s_i}}(m_{s_i}) - 1$ for all $i \ge 0$ if we set $s_0 = 0$. Furthermore, $\eta_i = s_i$ and $\mu_i = M_{s_i} - M$.

The minimal j such that $b_{jk} \ne 0$ is therefore

$$M_{\eta_0} + M_{\eta_1} + \dots + M_{\eta_{k-1}} = kM + \sum_{i=0}^{k-1} \mu_i = k\left(M + h_{q,M}\left(\frac{m}{q^M}\right)\right) + O(1)$$

and the height for $N = q^{j+1} - 1$ with this j is

$$H(q^{j+1}-1) = \max_{n<q^{j+1}} d(n) = k = \frac{j}{M + h_{q,M}(\frac{m}{q^M})} + O(1).$$

Clearly, $H(N)$ is monotonically increasing and thus (1) is proved. It is easy to see that, for all $k \ge 0$, $M_{\eta_0} + \dots + M_{\eta_{k-1}}$ does not decrease if $\frac{m}{q^M}$ increases and thus the $h_{q,M}$'s are monotonically increasing.

6 Expected value and variance

For $m = q^M$, expected value and variance have been calculated in Section 3. Thus we can restrict to $m \ne q^M$. The first step is to obtain proper information about $b_j(z)$.

Proposition 1 *For $m \ne q^M$, we have some $\nu > 0$ such that, as $j \to \infty$,*

$$b_j(z) = C(z)q(z)^j + O\left(q^{(1-\nu)j}\right) \tag{14}$$

uniformly in $|z| \le 1+\tilde\rho$ for some $\tilde\rho > 0$, where $q(z)$ is the (algebraic) function satisfying $q(1) = q$, $P(z, 1/q(z)) = 1$ and $C(z)$ is an analytic function in $|z| \le 1+\tilde\rho$ with $C(1) = q-1$.

Proof. First we study the poles of $B(z,u)$ with $|z| \le 1$, $|u| \le 1/q$. One pole is $(z,u) = (1,1/q)$ because of

$$B(1,u) = \sum_{j \ge 0} (q-1) q^j u^j = \frac{q-1}{1-qu}.$$

Hence $P(1,1/q) = 1$ and there exists an algebraic function $q(z)$ with $P(z,1/q(z)) = 1$ and $q(1) = q$. Clearly, $q(z)$ is a solution of $F(z,1/q(z)) = 1$ for $|z| \le 1+\rho$ too.

$F(z,u) = 1$ has no solutions with $|z| < 1$ or $|u| < 1/q$ since all f_{jk} are nonnegative. For a solution with $|z| = 1$ and $|u| = 1/q$, we need $z^k(uq)^j = 1$ for all j,k with $f_{jk} > 0$. We have $f_{M+1,1} > 0$ because of $m \ne q^M$ and thus $z(uq)^{M+1} = 1$. Now, let $\ell \ge 1$ be minimal such that $\mu_\ell = 0$. Then we have $f_{\ell(M+1)-1,\ell} > 0$ and thus $z^\ell (uq)^{\ell(M+1)-1} = 1$. Together with $z(uq)^{M+1} = 1$, this implies $1/(uq) = 1$, hence $u = 1/q$ and $z = 1$. Hence $(z,u) = (1,1/q)$ is the only pole of $B(z,u)$ with $|z| \le 1$, $|u| \le 1/q$.

Furthermore, $(1,1/q)$ is a simple zero of $1 - F(z,u)$ and $1 - P(z,u)$. Hence we have some $\tilde\rho, \hat\rho > 0$ such that we have no zeros with $|z| \le 1+\tilde\rho$, $|u| < 1/q + \hat\rho$ except $(z,1/q(z))$. Then

$$\frac{Q(z,u)}{1-P(z,u)} = \frac{Q(z,u)}{(1-q(z)u)\tilde P(z,u)} = \frac{1}{1-q(z)u}\left(\frac{Q(z,1/q(z))}{\tilde P(z,1/q(z))} + (1-q(z)u)R(z,u)\right) = \frac{C(z)}{1-q(z)u} + R(z,u)$$

for some algebraic function $\tilde P(z,u)$ and analytic functions $C(z), R(z,u)$ in $|z| \le 1+\tilde\rho$, $|u| < 1/q + \hat\rho$.

By Cauchy's formula, we get

$$b_j(z) = \frac{1}{2\pi i} \int\limits_{|u| = \frac{1}{q} + \frac{\hat\rho}{2}} \frac{Q(z,u)}{1-P(z,u)} \frac{du}{u^{j+1}} = C(z) q(z)^j + \frac{1}{2\pi i} \int\limits_{|u| = \frac{1}{q} + \frac{\hat\rho}{2}} R(z,u) \frac{du}{u^{j+1}} = C(z) q(z)^j + O\left(q^{(1-\nu)j}\right)$$

with some $\nu > 0$. This completes the proof of the proposition. □

Now, we build the generating function of a_{Nk} for general N with the $b_j(z)$.

Lemma 2 *If we set $L = \lfloor \log_q N \rfloor$ and $d(0) = 0$, then*

$$\sum_{k \ge 0} a_{Nk} z^k = \sum_{j=0}^{L-1} b_j(z) \sum_{\ell=j+1}^{L} \sum_{c=0}^{\varepsilon_\ell(N)-1} z^{d(\sum_{s=\ell+1}^{L} \varepsilon_s(N) q^{s-j-1} + c q^{\ell-j-1}) + O(1)} + O\left(\log N \left(1 + z^{O(\log N)}\right)\right). \tag{15}$$

Proof. We have

$$\begin{aligned}
&\{0,1,2,\ldots,N-1\} \\
&= \{0,\ldots,\varepsilon_L(N)q^L - 1\} \cup \{\varepsilon_L(N)q^L, \ldots, \varepsilon_L(N)q^L + \varepsilon_{L-1}(N)q^{L-1} - 1\} \cup \cdots \cup \{\sum_{s=1}^{L} \varepsilon_s(N)q^s, \ldots, \sum_{s=0}^{L} \varepsilon_s(N) q^s - 1\} \\
&= \bigcup_{\ell=0}^{L} \bigcup_{c=0}^{\varepsilon_\ell(N)-1} \left(\sum_{s=\ell+1}^{L} \varepsilon_s(N) q^s + c q^\ell + \{0,\ldots,q^\ell - 1\} \right) \\
&= \bigcup_{j=0}^{L-1} \bigcup_{\ell=j+1}^{L} \bigcup_{c=0}^{\varepsilon_\ell(N)-1} \left(\sum_{s=\ell+1}^{L} \varepsilon_s(N) q^s + c q^\ell + \{q^j, \ldots, q^{j+1} - 1\} \right) \cup \bigcup_{\ell=0}^{L} \bigcup_{c=0}^{\varepsilon_\ell(N)-1} \{ \sum_{s=\ell+1}^{L} \varepsilon_s(N) q^s + c q^\ell \}
\end{aligned}$$

The element $\sum_{s=\ell+1}^{L} \varepsilon_s(N)q^s + cq^\ell + n$ with $q^j \le n < q^{j+1}$ and $j < \ell$ is located in a subtree under the node containing n. Its depth is therefore that of n plus some additional depth depending on the shape of the subtree (see the proof of Lemma 1), which can be bounded by $d(\sum_{s=\ell+1}^{L} \varepsilon_s(N)q^{s-j} + cq^{\ell-j}) + 2$.

The depth of the remaining $O(L)$ terms can be estimated by the height of the tree, $O(L)$. □

Now, we can calculate the mean value

$$\mathbb{E}X_N = \frac{d}{dz}\left(\frac{1}{N}\sum_{k\ge 0} a_{Nk}z^k\right)\Bigg|_{z=1} = \frac{1}{N}\sum_{j=0}^{L-1} b_j'(1)\sum_{\ell=j+1}^{L}\varepsilon_\ell(N) + \frac{1}{N}\sum_{j=0}^{L-1} b_j(1)\sum_{\ell=j+1}^{L}\varepsilon_\ell(N)O(L-\ell) + O(1)$$

$$= \frac{1}{N}\sum_{\ell=1}^{L}\varepsilon_\ell(N)\sum_{j=\ell-1}^{L-1} j(q-1)q^{j-1}q'(1) + \frac{1}{N}\sum_{\ell=1}^{L}\varepsilon_\ell(N)\sum_{j=\ell-1}^{L-1}(q-1)q^j O(L-\ell) + O(1) = L\frac{q'(1)}{q} + O(1)\,.$$

Thus (3) is proved and we have, since $F(z,u(z)) = 1$ for $u(z) = 1/q(z)$,

$$\frac{\partial F}{\partial z}(z,u(z)) + u'(z)\frac{\partial F}{\partial u}(z,u(z)) = 0, \text{ i.e., } u'(z) = -\frac{\frac{\partial F}{\partial z}(z,u(z))}{\frac{\partial F}{\partial u}(z,u(z))},$$

$$\mu = \frac{q'(1)}{q} = -\frac{u'(1)}{u(1)} = \frac{\frac{\partial F}{\partial z}(1,1/q)}{\frac{1}{q}\frac{\partial F}{\partial u}(1,1/q)} = \frac{\sum_{k=M}^{\infty}\sum_{j=1}^{\infty} kf_{jk}/q^j}{\sum_{k=M}^{\infty}\sum_{j=1}^{\infty} jf_{jk}/q^j}, \tag{16}$$

where F can be replaced by P.

For the variance, we have to be more careful. First we distinguish the elements by their place inside the node and the type of the node, in order to obtain

$$d(n + q^{j+1}\tilde{n}) = d(n) + d_\theta(\tilde{n})$$

for all n with $j = \lfloor \log_q n \rfloor$ at a position of type $\theta \in \Theta = \{1,\dots,m-1\}\times\{0,\dots,q-2\}$ and some functions d_θ with $d_\theta(\tilde{n}) = O(\log \tilde{n})$. With

$$b_{jk}^\theta = |\{n \in \{q^j,\dots,q^{j+1}-1\} :\ d(n) = k, \text{ the position of } n \text{ is of type } \theta\}|,$$

we have

$$b_j^\theta(z) = \sum_{k\ge 0} b_{jk}^\theta z^k = C_\theta(z)q(z)^j + O\left(q^{(1-\nu)j}\right)$$

for some analytic functions $C_\theta(z)$ (in $|z| \le 1+\tilde{\rho}$) because of

$$B^\theta(z,u) = \sum_{j\ge 0} b_j^\theta(z)u^j = \frac{G_\theta(z,u)}{1-F(z,u)}$$

for some analytic functions $G_\theta(z,u)$ (in D_ρ). This allows to refine (15) to

$$\sum_{k\ge 0} a_{Nk}z^k = \sum_{j=0}^{L-1}\sum_{\ell=j+1}^{L}\sum_{c=0}^{\varepsilon_\ell(N)-1}\sum_{\theta\in\Theta} b_j^\theta(z)z^{d_\theta(\sum_{s=\ell+1}^{L}\varepsilon_s(N)q^{s-j-1} + cq^{\ell-j-1})} + O\left(\log N\left(1 + z^{O(\log N)}\right)\right) \tag{17}$$

and the variance is

$$\begin{aligned}
\mathbb{V}X_N &= \frac{d^2}{dz^2}\left(\frac{1}{N}\sum_{k\geq 0} a_{Nk} z^k\right)\bigg|_{z=1} + \mathbb{E}X_N - (\mathbb{E}X_N)^2 \\
&= \frac{1}{N}\sum_{j=0}^{L-1}\sum_{\ell=j+1}^{L}\sum_{c=0}^{\varepsilon_\ell(N)-1}\sum_{\theta\in\Theta}\Big(C_\theta(1) j q^{j-1} q''(1) + C_\theta(1) j(j-1) q^{j-2}(q'(1))^2 + 2C'_\theta(1) j q^{j-1} q'(1) \\
&\quad + 2C_\theta(1) j q^{j-1} q'(1) d_\theta\Big(\sum_{s=\ell+1}^{L}\varepsilon_s(N) q^{s-j-1} + c q^{\ell-j-1}\Big)\Big) + L\frac{q'(1)}{q} - \frac{1}{N}\sum_{j,\ell,c,\theta} C_\theta(1) j q^{j-1} q'(1) \\
&\quad \times \frac{1}{N}\sum_{\tilde{j},\tilde{\ell},\tilde{c},\tilde{\theta}}\Big(C_{\tilde{\theta}}(1)\tilde{j} q^{\tilde{j}-1} q'(1) + 2C'_{\tilde{\theta}}(1) q^{\tilde{j}} + 2C_{\tilde{\theta}}(1) q^{\tilde{j}} d_{\tilde{\theta}}\Big(\sum_{s=\tilde{\ell}+1}^{L}\varepsilon_s(N) q^{s-\tilde{j}-1} + c q^{\tilde{\ell}-\tilde{j}-1}\Big)\Big) + O(1) \\
&= L\frac{q''(1)}{q} - L\frac{(q'(1))^2}{q^2} + L\frac{q'(1)}{q} + O(1) = L\left(\frac{q''(1)}{q} + \mu - \mu^2\right) + O(1)
\end{aligned}$$

Thus (4) is proved with

$$\begin{aligned}
\sigma^2 &= \frac{q''(1)}{q} + \mu - \mu^2 = \frac{d}{dz}\left(-\frac{u'(z)}{u(z)}\right)\bigg|_{z=1} + \mu = \frac{d}{dz}\left(\frac{\frac{\partial F}{\partial z}(z,u(z))}{u(z)\frac{\partial F}{\partial u}(z,u(z))}\right)\bigg|_{z=1} + \mu \\
&= \frac{(\frac{\partial^2 F}{\partial z^2} + u'\frac{\partial^2 F}{\partial z\partial u})u\frac{\partial F}{\partial u} - (u'\frac{\partial F}{\partial u} + u\frac{\partial^2 F}{\partial z\partial u} + u'u\frac{\partial^2 F}{\partial u^2})\frac{\partial F}{\partial z} + u\frac{\partial F}{\partial z}\frac{\partial F}{\partial u}}{(u\frac{\partial F}{\partial u})^2}\Bigg|_{z=1} \\
&= \frac{\mu}{u\frac{\partial F}{\partial u}}\left(\frac{1}{\mu}\frac{\partial^2 F}{\partial z^2} - 2u\frac{\partial^2 F}{\partial z\partial u} + \frac{\partial F}{\partial z} + u^2\mu\frac{\partial^2 F}{\partial u^2} + u\frac{\partial F}{\partial u}\right)\bigg|_{z=1} \qquad (18) \\
&= \frac{q\mu}{\frac{\partial F}{\partial u}(1,1/q)}\sum_{j=M}^{\infty}\sum_{k=1}^{\infty}\frac{f_{jk}}{q^j}\left(\frac{k(k-1)}{\mu} - 2jk + j(j-1)\mu + j\right) \\
&= \frac{q\mu}{\frac{\partial F}{\partial u}(1,1/q)}\sum_{j=M}^{\infty}\sum_{k=1}^{\infty}\frac{f_{jk}}{q^j}\left(\frac{1}{\mu}(k-\mu j)^2 + \left(1-\frac{1}{\mu}\right)(k-\mu j)\right) = \frac{q}{\frac{\partial F}{\partial u}(1,1/q)}\sum_{j=M}^{\infty}\sum_{k=1}^{\infty}\frac{f_{jk}}{q^j}(k-\mu j)^2.
\end{aligned}$$

The last equation holds because of

$$\sum_{j=M}^{\infty}\sum_{k=1}^{\infty} f_{jk}(k-\mu j) = \frac{\partial F}{\partial z}(1,1/q) - \mu q\frac{\partial F}{\partial u}(1,1/q) = 0.$$

(16) shows $\mu \in (\frac{1}{M+1}, \frac{1}{M})$ for $m \neq q^M$ since we have $kM \leq j \leq k(M+1)$ for all j,k with $f_{jk} > 0$ and we have some j,k such that $kM < j$ and some j,k such that $j < k(M+1)$ (see the proof of Proposition 1). Furthermore, for $m \neq q^M$, j/k is not equal for all j,k with $f_{jk} > 0$ which implies $\sigma^2 > 0$.

7 Global limit law

Now, we prove the asymptotic normality of X_N. Observe that its characteristic function is

$$\frac{1}{N}\sum_{k\geq 0} a_{Nk} e^{ikt} = \mathbb{E}\, e^{itX_N}.$$

Proposition 2 *Suppose $m \ne q^M$ and set $\mu_N = \mathbb{E} X_N$, $\sigma_N^2 = \mathbb{V} X_N$. Then for every $\delta > 0$, we have uniformly for $|t| \le (\log N)^{1/2-\delta}$*

$$e^{-it\mu_N/\sigma_N} \frac{1}{N} \sum_{k\ge 0} a_{Nk} e^{ikt/\sigma_N} = e^{-t^2/2} + O\left((\log N)^{-1/2+\delta} \right). \tag{19}$$

Proof. We have

$$q(e^{it}) = q e^{i\mu t - \sigma^2 t^2/2 + O\left(t^3\right)} \tag{20}$$

and, by using Proposition 1,

$$b_j(e^{it}) = (q-1) q^j e^{j(i\mu t - \sigma^2 t^2/2)} e^{O\left(t + jt^3\right)} + O\left(q^{(1-\nu)j}\right)$$

in an open (real) neighbourhood of $t = 0$. By Lemma 2, we obtain

$$\begin{aligned}\sum_{k\ge 0} a_{Nk} e^{ikt} &= \sum_{j=0}^{L-1} b_j(e^{it}) \sum_{\ell=j+1}^{L} \sum_{c=0}^{\varepsilon_\ell(N)-1} e^{it\,O(L-j)} + O(\log N) \\ &= \sum_{j=L-\lfloor L^\delta \rfloor}^{L-1} (q-1) q^j e^{ij\mu t - j\sigma^2 t^2/2} e^{O\left(t+jt^3\right)} \sum_{\ell=j+1}^{L} \varepsilon_\ell(N) e^{O\left(L^\delta t\right)} + O\left(q^{L-L^\delta}\right)\end{aligned}$$

for (small) $\delta > 0$. Now observe that $\mu_N = \mu L + O(1)$ and $1/\sigma_N = 1/(\sigma\sqrt{L})(1 + O\left(L^{-1}\right))$. Hence

$$\begin{aligned}&\mathbb{E}\, e^{it(X_N-\mu_N)/\sigma_N} = e^{-it\mu_N/\sigma_N} \frac{1}{N} \sum_{k\ge 0} a_{Nk} e^{ikt/\sigma_N} \\ &= e^{-t^2/2} \sum_{\ell=L-\lfloor L^{2\delta/3}\rfloor+1}^{L} \frac{\varepsilon_\ell(N)(q^\ell - q^{L-\lfloor L^{2\delta/3}\rfloor})}{N} e^{it(\mu/\sigma\sqrt{L})O(L-\ell) + t^2 O(L-\ell)/L} e^{O\left(tL^{-1/2} + t^3 L^{-1/2}\right)} + O\left(q^{-L^{2\delta/3}}\right) \\ &= e^{-t^2/2} e^{O\left(tL^{2\delta/3-1/2} + t^2 L^{2\delta/3-1} + t^3 L^{-1/2}\right)} + O\left(q^{-L^{2\delta/3}}\right),\end{aligned}$$

which implies (19) directly for $|t| \le (\log N)^{\delta/3}$. For $|t| > (\log N)^{\delta/3}$, we have

$$e^{-t^2/2} e^{O\left(t^3 L^{-1/2}\right)} = e^{-t^2(1/2 + O\left(tL^{-1/2}\right))} \le e^{-cL^{2\delta/3}} = O\left((\log N)^{-1/2+\delta} \right)$$

for some $c > 0$, which again implies (19). □

We can now prove the first part of Theorem 3. Set

$$\Delta_N(t) = e^{-t^2/2} - \mathbb{E}\, e^{it(X_N-\mu_N)/\sigma_N}.$$

Then, by Esseen's inequality [8, p. 32], we have

$$\frac{1}{N} |\{ n \le N :\ d(n) < \mathbb{E} X_N + x \mathbb{V} X_N \}| = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt + O\left(\frac{1}{T} + \int_{-T}^{T} \left| \frac{\Delta_N(t)}{t} \right| dt \right).$$

Choosing $T = (\log N)^{1/2-\delta}$, we directly obtain from Proposition 2 and by applying the estimate

$$e^{-it\mu_N/\sigma_N} \frac{1}{N} \sum_{k \geq 0} a_{Nk} e^{ikt/\sigma_N} = 1 + O\left(t^2\right)$$

for $|t| \leq (\log N)^{-1}$ that

$$\int_{-T}^{T} \left| \frac{\Delta_N(t)}{t} \right| dt = O\Big((\log N)^{-1/2+\delta} (\log\log N)\Big)$$

for every $\delta > 0$. Hence (8) follows.

8 Local limit law

For the local limit law, we have to study the b_{jk}^{θ}, which have the same asymptotic behavior as the b_{jk}. We use Proposition 1 and saddle point approximations.

Proposition 3 *We have*

$$b_{jk} = \frac{(q-1)q^j}{\sqrt{2\pi j \sigma^2}} \left(\exp\left(-\frac{(k-j\mu)^2}{2j\sigma^2} \right) + O\left(j^{-1/2}\right) \right)$$

uniformly for all $j,k \geq 0$.

Proof. We use Cauchy's formula

$$b_{jk} = \frac{1}{2\pi} \int_{-\pi}^{\pi} b_j(e^{it}) e^{-ikt}\, dt.$$

Since $q(z)$ is an algebraic function with $q(e^{it}) < q$ for $0 < t < 2\pi$ and $C(e^{it})$ is bounded, we have, by Propostion 1, some $\nu > 0$ and some $\tau > 0$ such that

$$b_j(e^{it}) = O\left(q^{(1-\nu)j}\right)$$

for $\tau \leq |t| \leq \pi$, which implies

$$\int_{\tau \leq |t| \leq \pi} |b_j(e^{it})|\, dt = O\left(q^{(1-\nu)j}\right) = O\left(q^j/j\right).$$

It remains to evaluate

$$I = \frac{1}{2\pi} \int_{|t| \leq j^{-\delta}} b_j(e^{it}) e^{-ikt}\, dt + \frac{1}{2\pi} \int_{j^{-\delta} \leq |t| \leq \tau} b_j(e^{it}) e^{-ikt}\, dt = I_1 + I_2$$

with $0 < \delta < \frac{1}{6}$. From (20), it follows that there exists a constant $c > 0$ such that $|q(e^{it})| \leq e^{-ct^2}$ for $|t| \leq \tau$. Hence

$$I_2 \leq \frac{1}{\pi} \int_{j^{-\delta}}^{\infty} e^{-cjt^2}\, dt + O\left(q^{(1-\nu)j}\right) = O\left(e^{-cj^{1-2\delta}}\right) + O\left(q^{(1-\nu)j}\right) = O\left(q^j/j\right)$$

Finally,

$$
\begin{aligned}
I_1 &= \frac{1}{2\pi}\int_{|t|\le j^{-\delta}} (q-1)q^j e^{it(j\mu-k)-j\sigma^2t^2/2}\left(1+O\left(t+jt^3\right)\right)dt + O\left(q^{(1-\nu)j}\right)\\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}(q-1)q^j e^{it(j\mu-k)-j\sigma^2t^2/2}\,dt + O\left(\int_{|t|>j^{-\delta}}(q-1)q^j e^{-j\sigma^2t^2/2}\,dt\right)\\
&\quad + O\left(\int_{|t|\le j^{-\delta}}(q-1)q^j e^{-j\sigma^2t^2/2}(|t|+j|t|^3)\,dt\right) + O\left(q^{(1-\nu)j}\right)\\
&= \frac{(q-1)q^j}{\sqrt{2\pi j\sigma^2}}\exp\left(-\frac{(k-j\mu)^2}{2j\sigma^2}\right) + O\left(q^j/j\right)
\end{aligned}
$$

and Proposition 3 is proved. □

Proposition 3 and (17) are used to prove (9). We have

$$
\begin{aligned}
a_{Nk} &= \sum_{j=0}^{L-1}\sum_{\ell=j+1}^{L}\sum_{c=0}^{\varepsilon_\ell(N)-1}\sum_{\theta\in T} b^{\theta}_{j,k-d_\theta(\sum_{s=\ell+1}^{L}\varepsilon_s(N)q^{s-j}+cq^{\ell-j})} + O(L)\\
&= \sum_{j=0}^{L-1}\sum_{\ell=j+1}^{L}\sum_{c=0}^{\varepsilon_\ell(N)-1}\frac{(q-1)q^j}{\sqrt{2\pi j\sigma^2}}\exp\left(-\frac{(k-O(L-\ell)-j\mu)^2}{2j\sigma^2}\right) + O\left(q^j/j\right)
\end{aligned}
$$

since the b^{θ}_{jk} have the same as asymptotics as the b_{jk}, with constants which sum up to $q-1$.

If $L-\lfloor L^\delta\rfloor < j \le L$ and $k-\mu_N = O\left(\sqrt{L}\log L\right)$, then $O(L-\ell)=O\left(L^\delta\right)$,

$$
\begin{aligned}
\frac{(k-\mu_N)^2}{2\sigma_N^2} - \frac{(k-O(L-\ell)-j\mu)^2}{2j\sigma^2} &= \frac{(k-\mu_N)^2-\left(k-\mu_N+O\left(L^\delta\right)\right)^2}{2j\sigma^2} + \frac{(k-\mu_N)^2}{2}\left(\frac{1}{j\sigma^2}-\frac{1}{\sigma_N^2}\right)\\
&= O\left(L^{\delta-1/2}\log L\right) + O\left(L^{\delta-1}(\log L)^2\right)
\end{aligned}
$$

and

$$
a_{Nk} = \frac{N}{\sqrt{2\pi\sigma_N^2}}\exp\left(-\frac{(k-\mu_N)^2}{2\sigma_N^2}\right)\left(1+O\left(L^{\delta-1/2}\log L\right)\right) + O\left(\frac{N}{L}\right).
$$

If $|k-\mu_N| \ge \sqrt{L}\log L$, then we have, for $L-\lfloor L^\delta\rfloor < j \le L$,

$$
b_{jk} = O\left(q^j L^{-1/2}\exp\left(-\frac{(\log L)^2}{4\sigma^2}\right)\right) = O\left(q^j L^{-1}\right)
$$

and thus $a_{Nk} = O(N/L)$. This completes the proof of Theorem 3.

9 Binary search trees and the binary van der Corput sequence

For binary search trees ($m = 2$), we have

$$\begin{aligned}
B_0(z,u) &= 1 + zuB_0(z,u) + zB_1(z,u) \\
B_1(z,u) &= 1 + zuB_0(z,u) + zB_2(z,u) \\
&\vdots \\
B_{q-3}(z,u) &= 1 + zuB_0(z,u) + zB_{q-2}(z,u) \\
B_{q-2}(z,u) &= 1 + 2zuB_0(z,u)
\end{aligned}$$

and thus

$$B_0(z,u) = 1 + z + z^2 + \cdots + z^{q-1} + u(z + z^2 + \cdots + z^{q-2} + 2z^{q-1})B_0(z,u),$$
$$P(z,u) = F(z,u) = (z + z^2 + \cdots + z^{q-2} + 2z^{q-1})u.$$

Hence

$$q(z) = z + z^2 + \cdots + z^{q-2} + 2z^{q-1}$$

and

$$\mu = \frac{q'(1)}{q} = \frac{1}{q}(1 + 2 + \cdots + (q-2) + 2(q-1)) = (q-1)\left(\frac{1}{2} + \frac{1}{q}\right).$$

With

$$q''(1) = 2 + 6 + \cdots + (q-3)(q-2) + 2(q-2)(q-1) = (q-1)(q-2)\left(\frac{q}{3} + 1\right),$$

we get

$$\sigma^2 = \frac{q''(1)}{q} + \mu - \mu^2 = \frac{(q-1)(q-2)(q^2+3q-6)}{12q^2}.$$

For the binary van der Corput sequence ($q = 2$), we have

$$B_0(z,u) = \sum_{j=0}^{M-1} 2^j u^j + (m - 2^M)u^M + (2m - 2^{M+1})zu^{M+1}B_0(z,u) + (2^{M+1} - m)zu^M B_0(z,u),$$

thus

$$P(z,u) = F(z,u) = (2m - 2^{M+1})zu^{M+1} + (2^{M+1} - m)zu^M.$$

Using (16) and (18), we get

$$\begin{aligned}
\mu &= \frac{\frac{2m-2^{M+1}}{2^{M+1}} + \frac{2^{M+1}-m}{2^M}}{\frac{(M+1)(2m-2^{M+1})}{2^{M+1}} + \frac{M(2^{M+1}-m)}{2^M}} = \frac{1}{M + \frac{m}{2^M} - 1}, \\
\sigma^2 &= \mu\left(\frac{2m - 2^{M+1}}{2^{M+1}}(1 - (M+1)\mu)^2 + \frac{2^{M+1} - m}{2^M}(1 - M\mu)^2\right) \\
&= \mu^3\left(\frac{m}{2^M} - 1\right)\left(2 - \frac{m}{2^M}\right)\left(2 - \frac{m}{2^M} + \frac{m}{2^M} - 1\right) = \frac{\left(\frac{m}{2^M} - 1\right)\left(2 - \frac{m}{2^M}\right)}{\left(M + \frac{m}{2^M} - 1\right)^3}.
\end{aligned}$$

References

[1] F. M. DEKKING AND P. VAN DER WAL, Uniform distribution modulo one and binary search trees, *J. Theor. Nombres Bordx.* **14** (2002), 415–424.

[2] L. DEVROYE, Applications of the theory of records in the study of random trees, Acta Informatica **26** (1988), 123–130.

[3] L. DEVROYE, Binary search trees based on Weyl and Lehmer sequences, in: *Monte Carlo and Quasi-Monte Carlo Methods 1996 (H. Niederreiter, P. Hellekalek, G. Larcher and P. Zinterhof eds.)*, Springer, New York, 1998, 40–65.

[4] L. DEVROYE, Universal limit laws for depths in random trees, SIAM Journal on Computing **28** (1999), 409–432.

[5] L. DEVROYE AND A. GOUDJIL, A study of random Weyl trees, *Random Structures and Algorithms* **9** (1998), 271–295.

[6] L. DEVROYE AND R. NEININGER, Random suffix search trees, to appear in *Random Structures and Algorithms*.

[7] M. DRMOTA AND J. GAJDOSIK, The Distribution of the Sum-of-Digits Function, *J. Theor. Nombres Bordx.* **10** (1998), 17–32.

[8] C.-G. ESSEEN, Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law, *Acta Math.* **77** (1945), 1–125.

[9] H. MAHMOUD AND B. PITTEL, On the most probable shape of a search tree grown from random permutations, *SIAM J. Algebraic Discrete Methods* **5** (1984), 69–81.

[10] H. MAHMOUD AND B. PITTEL, On the joint distribution of the insertion path length and the number of comparisons in search trees, *Discrete Appl. Math.* **20** (1988), 243–251.