On the variance of a class of inductive valuations of data structures for digital search

Werner Schachinger

Institut für Statistik, Operations Research und Computerverfahren, Universität Wien, Brünnerstrasse 72, A-1210 Wien, Austria

Abstract

Let an inductive valuation $L$ on the family of binary tries or Patricia tries or digital search trees be defined in the following way: $L(t) = L(t_\ell) + L(t_r) + R(t)$, where $t_\ell$ and $t_r$ denote the left and right subtrees of $t$ and $R$ depends only on the size (the number of records) $|t|$ of $t$. Let $L_N$ denote $L$ restricted to the trees of size $N$. In Theorem 1 we give sufficient conditions on the sequence $r(t) := R(t)$ for the variance $\text{Var } L_N$ to be of exact order $N$, if the family of tries (resp. Patricia tries, resp. digital search trees) is equipped with the Bernoulli model. For the symmetric Bernoulli model we prove the existence of a continuous periodic function $\delta$ with period 1, such that $\text{Var } L_N \sim \delta(\log_2 N) \cdot N$ holds.

1. Introduction and main result

Data structures designed for data having keys, that are sequences over a finite alphabet $\Sigma$ (for simplicity, take $\Sigma = \{0, 1\}$), are of great importance in computer science, as they occur in connection with dynamic hashing [17], radix exchange sort [6], multidimensional digital searching [26], conflict resolution algorithms [28], and parsing algorithms [21], which is not at all a complete list. There are 3 prominent data structures for digital search, namely the trie [6], the Patricia trie [24] and the digital search tree [1], which are all (in case of $\Sigma = \{0, 1\}$) binary trees, but follow different construction rules. See also [23] and [16]. Several characteristic valuations $L$ of these data structures, such as storage complexity, search costs or the number of minimal subtrees of fixed size can be inductively defined by $L(t) = L(t_\ell) + L(t_r) + R(t)$ together with some termination rule, where $t_\ell$ and $t_r$ denote the left and right subtrees of $t$ and $R$ is a “simple” valuation depending only on the size (the number of records) $|t|$ of $t$. These valuations $L$ have been studied from the average case point of view under the so-called Bernoulli model, where the tree is assumed to be built from a fixed number $N$ of keys that are infinite sequences of i.i.d. random variables with values in $\Sigma$. $L$ restricted to trees of size $N$ is now a random variable denoted by $L_N$. See Section 5 for definitions of some important valuations and references concerning their
expectations, variances and in two instances limiting distributions. Much of the work concerning expectations has been done by Knuth [16]. In recent years some results concerning variances have been achieved by very different methods, such as Mellin's inversion formula [25], Rice's method [13], the use of certain identities belonging to the theory of modular functions [11], singularity analysis of ordinary generating functions [7] and the use of difference equations on characteristic functions [19]. It is surprising, that the analyses of different valuations, which seem to demand their very specific methods, yield the same asymptotic order of the variance: \( \text{Var} \, L_N \asymp N \). The aim of this paper is to demonstrate that this behavior is characteristic for a considerably large class of inductively defined valuations.

We now enlarge on the concepts introduced so far.

1.1. Data structures for digital search

We start with a brief description of the data structures under consideration:

Let \( S := \{ s_1, \ldots, s_n \} \) be a finite set of infinite 0–1-strings. \( S_0 \) denotes the subset of elements of \( S \), whose first bit is 0, \( S_1 = S \setminus S_0 \) the subset of those, whose first bit is 1. By \( \theta \) we denote the string-operator, that drops the first bit and shifts the others one place to the left. \( \theta S \) is the same as \( \{ \theta s, s \in S \} \), and we define \( \theta_i, i = 0, 1, \) by \( \theta_i S = \theta(S_i) \). The subsets of \( S \) inherit the linear order \( < \) on \( S \) given by \( s_i < s_k : \Leftrightarrow i < k \). For \( R \subseteq S \), the smallest element of \( R \) in this order is denoted by \( \hat{R} \).

The trie \( \text{trie}(S) \) consisting of the keys in \( S \) can now be inductively defined by

\[
\begin{align*}
\text{trie}(\emptyset) &= \emptyset \\
\text{trie}(\{ s \}) &= \{ s \} \\
\text{trie}(S) &= \text{trie}(\theta_0 S) \cup \text{trie}(\theta_1 S)
\end{align*}
\]

It may happen, that one of \( \theta_0 S, \theta_1 S \) is empty, which results in a root with only one nonempty subtree. This possibility of one-way branching is eliminated by the Patricia trie \( \text{pat}(S) \), which is defined by

\[
\begin{align*}
\text{pat}(\emptyset) &= \emptyset \\
\text{pat}(\{ s \}) &= \{ s \} \\
\text{pat}(S) &= \text{pat}(\theta_0 \theta^k S) \cup \text{pat}(\theta_1 \theta^k S)
\end{align*}
\]

where \( k_S \) is the first \( k \geq 0 \), such that \( (\theta^k S)_0 \neq \emptyset \) and \( (\theta^k S)_1 \neq \emptyset \).

Finally, the digital search tree \( \text{dst}(S) \), that stores keys in the internal nodes, is defined by

\[
\begin{align*}
\text{dst}(\emptyset) &= \emptyset \\
\text{dst}(S) &= \text{dst}(\theta_0 (S \setminus \hat{S})) \cup \text{dst}(\theta_1 (S \setminus \hat{S}))
\end{align*}
\]

These procedures terminate, because the keys were assumed to be distinct. Note that the order of the keys is of importance only in the case of the dst.
Example. \( a = (0, 1, 0, \ldots), \ b = (1, 1, 0, \ldots), \ c = (1, 1, 1, \ldots), \ d = (0, 0, 1, \ldots), \ e = (0, 0, 0, \ldots). \) See Fig. 1.

1.2. Inductive valuations

A valuation on a family \( T \) of trees is simply a function \( X : T \to \mathbb{R} \). A very simple example of a valuation of a tree \( t \), that corresponds to a data structure, is its “size”, denoted by \(|t|\), that is the number of keys \( t \) consists of. Let now \( n_0 \in \mathbb{N} \) and a valuation \( R \) be given which is constant on each set \( T_n := \{ t \in T \mid |t| = n \} \) for \( n \geq n_0 \) (i.e. \( R(t) =: r_{|t|} \) for \(|t| \geq n_0\)), but not necessarily constant on the sets \( T_n \) for \( n < n_0 \). The inductive valuations we are dealing with are defined on the family of tries (resp. Patricia tries, resp. digital search trees) and can for some fixed \( n_0 > 2 \) be described by

\[
L(t) = \begin{cases} 
R(t), & |t| < n_0, \\
|t| + L(t), & |t| \geq n_0,
\end{cases}
\]

The restriction \( n_0 \geq 2 \) is justified by the fact, that there is only one \( t \) of size 0 and one of size 1.

1.3. The probability model

We now define \( T_N^T, T_N^P \) and \( T_N^D \) to be the families of tries, Patricia tries and digital search trees of size \( N \). Note that \( T_N^T \) and \( T_N^D \) are finite, but \( T_N^P \) is infinite for \( N \geq 2 \). The input model under which we will study random variables is the Bernoulli model: By assuming that the bits of each key are i.i.d. with \( P(0) = p, P(1) = 1 - p, \) and that different keys are independent as well, each \( T_N^\sigma, \sigma \in \{ T, P, D \} \), becomes a probability space. The probability \( P(t) \) of a specific \( t \) (\( P(t) \) is short-hand notation for \( P(t \mid |t|) \), i.e. the probability of \( t \) in its probability space \( T_N^\sigma \)) is given inductively by the use of the splitting probabilities

\[
p_{N,k} := P(|t_r| = k \mid |t| = N).
\]
Since there is exactly one tree of size 0 and one of size 1, we have $P(t) = 1$ for $|t| \leq 1$.

For $|t| > 1$, $t = \frac{1}{t_{\ell}} \frac{1}{t_{r}}$, we have

$$P(t) = p_{R_\ell t_{\ell}} p_{R_r t_r}.$$

We define matrices $P^T_p, P^P_p, P^D_p$ corresponding to the splitting probabilities of tries, Patricia tries and digital search trees respectively,

$$
(P^T_p)_{N,k} = \binom{N}{k} p^k (1 - p)^{N-k}, \\
(P^P_p)_{N,k} = \binom{N-1}{k} p^k (1 - p)^{N-1-k}, \\
(P^D_p)_{N,k} = \begin{cases} 
\binom{N}{k} p^k (1 - p)^{N-k} 
& 1 \leq k \leq N-1, \\
1 - p^N - (1 - p)^N 
& \text{else.}
\end{cases}
$$

A valuation $X$, restricted to a family $T_N$ equipped with the Bernoulli model, gives rise to a random variable $X_N = X \mid T_N$.

**Theorem 1.** Let $L_N$ be the random variable belonging to the inductive valuation $L$, defined by (1.1) on the family of tries (resp. Patricia tries, resp. digital search trees) built from $N$ records and equipped with the Bernoulli model. Let $\text{Var} L_N = \text{Var} R_N < \infty$ for $0 < N < n_0$. Then

(i)

$$\text{Var} L_N = \Omega(N)$$

if and only if there are no constants $c, d$, such that $L$ is of the simple form $L(t) = ct + d$ for $|t| \geq 1$.

(ii) Let $\varepsilon > 0$. Each of the following conditions

(a) $r_N = O(N^{1/2 - \varepsilon})$,

(b) $\Delta r_N = O(N^{-\varepsilon})$, or

(c) $p = \frac{1}{2}$ and $\Delta^2 r_N = O(N^{-1/2 - \varepsilon})$

is sufficient for

$$\text{Var} L_N = O(N).$$

$\Delta$ is the difference operator ($\Delta r_k = r_{k+1} - r_k$).

(iii) In the symmetric case ($p = \frac{1}{2}$) each of (a)-(c) implies

$$\text{Var} L_N = \delta \log_2 N \cdot N + O(N^{1-\varepsilon}),$$

with a continuous periodic function $\delta$ with period 1.

**Remark.** Our method of proof is not capable of yielding Fourier coefficients of $\delta$ (as done by several authors for some special valuations, cf. the references given in Section 5), but as we have good reason to guess that $\delta$ need not be continuously differentiable.
(an example should be able to be constructed from a modification of the valuation given in (4.1)), a Fourier series expansion of $\delta$ need not exist in general.

The paper is organized as follows: In Section 2 we set up some notation, derive systems of linear equations for $\mathbb{E}L_N$ and $\text{Var} L_N$ and list elementary facts about the matrices we use. Section 3 contains the proof of Theorem 1, which is preceded by four lemmas, the first of them being of its own interest. Section 4 contains remarks on the possibility of weakening the conditions (a)-(c) of Theorem 1. In Section 5, Theorem 1 is applied to several valuations of practical importance. We compare our results with the results so far existing and, by the way, derive new asymptotics for asymmetric tries in Sections 5.2 and 5.4.

2. Preliminaries

2.1. Basic equations for $\mathbb{E}L_N$ and $\text{Var} L_N$

Let

$$\ell_N := \mathbb{E}L_N, \quad v_N := \text{Var} L_N.$$  

First of all we observe, that $\ell_N < \infty$ and $v_N < \infty$ for $N \geq 0$. This is trivial for the Patricia trie (case-P), and the digital search tree (case-D), because $T^p_N$ and $T^p_N$ are finite for each $N$. In the case of the trie we define sets $T^{(k)}_N \subset T^p_N$ by

$$T^{(0)}_N := \{ t \in T^p_N | t \not\in T^p_N \land t \not\in T^p_N \},$$

$$T^{(k+1)}_N := \{ t \in T^p_N | t \in T^{(k)}_N \lor t \in T^{(k)}_N \}, \quad k \geq 0.$$  

These sets are easily seen to satisfy $\bigcup_{k \geq 0} T^{(k)}_N = T^p_N$ and $T^{(k)}_N \cap T^{(j)}_N = \emptyset$ for $k \neq j$. Obviously $\ell_N = \sum_{k \geq 0} \sum_{t \in T^{(k)}_N} P(t) L(t)$, and this series converges absolutely: By Definition 1.1 of $L$ and the use of the splitting probabilities $p_{N,k} = (P^p_N)_{N,k}$ we have for $N \geq n_0$, $k \geq 1$,

$$\sum_{t \in T^{(0)}_N} P(t) L(t) = (p^N + q^N)^N \sum_{t \in T^{(0)}_N} P(t) L(t) + k(p^N + q^N)^k(1 - p^N - q^N)(r_N + \ell_0).$$  

Now

$$\sum_{t \in T^{(0)}_N} P(t) L(t) = \sum_{k=1}^{N-1} (P^p_N)_{N,k}(\ell_k + \ell_{N-k} + r_N)$$

depends only on $\ell_k$ for $k \leq N - 1$, so we can proceed by induction, since $\ell_N < \infty$ for $N < n_0$. In the same manner we can show $\mathbb{E}L^2_N < \infty$.

We consider now (case-T), where we have the following system of equations:

$$\ell_N = \mathbb{E}R_N =: r_N, \quad 0 \leq N < n_0,$$

$$\ell_N = \sum_{k=0}^{N} (P^T_N)_{N,k}(\ell_k + \ell_{N-k}) + r_N, \quad N \geq n_0.$$  

(2.1)
Now $\ell_N$ can be computed recursively, since $1 - (P^T_N)_{0,0} - (P^T_N)_{N,N} > 0$ for $N \geq 2$. More cumbersome though elementary manipulations yield

$$v_N = \text{Var } R_N =: s_N, \quad 0 \leq N < n_0,$$

$$v_N = \sum_{k=0}^{N} (P^T_N)_{N,k}(\ell_k + v_{N-k}) + s_N, \quad N \geq n_0,$$

with

$$s_N = \sum_{k=0}^{N} (P^T_N)_{N,k}(\ell_k + \ell_{N-k} + r_N - \ell_N)^2 \quad \text{for } N \geq n_0. \tag{2.3}$$

The equations for (case-P) are the same as (2.1)–(2.3) with $P^p$ in place of $P^T$.

Digital search trees (case-D) behave somewhat differently due to the fact, that one key is stored in the root. The equations are the same as (2.1)–(2.3) with $P^D_p$ in place of $P^T_p$, but each $\ell_{N-k}$ (resp. $v_{N-k}$) has to be changed for a $\ell_{N-1-k}$ (resp. $v_{N-1-k}$).

Let $q := 1 - p$. Then

$$(P^p)_{N,k} = (P^p_q)_{N,N-k}, \quad \sigma \in \{T, P\}, \quad (P^D_p)_{N,k} = (P^D_q)_{N,N-1-k},$$

and our equations read

$$(I - M_p - M_q)l = r, \tag{2.4}$$

$$(I - M_p - M_q)v = s, \tag{2.5}$$

with $r_N = \mathbb{E} R_N$ and $s_N = \text{Var } R_N$ for $0 < N < n_0$ and

$$s_N = \left\{ \begin{array}{ll}
\sum_{k=0}^{N} (M_p)_{N,k}(\ell_k + \ell_{N-k} - \sum_{k=0}^{N} (M_p)_{N,k}(\ell_k + \ell_{N-k}))^2, & \text{(case-T), (case-P),} \\
\sum_{k=0}^{N} (M_p)_{N,k}(\ell_k + \ell_{N-1-k} - \sum_{k=0}^{N} (M_p)_{N,k}(\ell_k + \ell_{N-1-k}))^2, & \text{(case-D),} 
\end{array} \right. \tag{2.6}$$

for $N \geq n_0$. $M_p = M_p(n_0, \sigma)$ is defined by

$$(M_p)_{N,k} = \left\{ \begin{array}{ll}
0 & \text{for } N < n_0, \\
(P^p_\sigma)_{N,k} & \text{for } N \geq n_0, 
\end{array} \right. \tag{2.7}$$

where $\sigma \in \{T, P, D\}$.

### 2.2. Properties of the splitting probability matrices

We want to list some elementary facts about the matrices $P^p_\sigma$ and $B_p$, which is defined by

$$(B_p)_{N,k} = \frac{(Np)^k}{k!} e^{-Np}. \tag{2.8}$$

$B_p$ can be thought of depending on $N \in \mathbb{R}$. Given a sequence $x$, we can define $X(z) = (B_1 x)_z = e^{-z} \sum_{k=0}^{\infty} (z^k/k!)x_k$, the Poisson generating function of $x$, which is
a well suited and by now standard device to obtain asymptotic results for the
Bernoulli model (cf. [8, 23]).

For \( P \in \{ P^T_p, P^p_p, P^D_p, B_p \} \) we have

\[
P_{N,k} \geq 0, \quad \sum_{k=0}^{\infty} P_{N,k} = 1, \quad N \geq 2, \quad \sum_{k=0}^{N} P_{N,k}(k - Np) = O(1).
\]  

(2.9)

For each \( \gamma > 0 \) there is a constant \( c_{p,\gamma} \), such that

\[
\sum_{|k - Np| \geq 2^3} P_{N,k} \leq c_{p,\gamma} N^{-\gamma},
\]  

(2.10)

and for \( \alpha > 0 \) we have

\[
\sum_{k=0}^{\infty} P_{N,k} |k - pN|^2 = Np(1 - p) + O(1)
\]  

(2.11)

Only for \( P \in \{ P^T_p, P^p_p, P^D_p \} \)

\[
\sum_{k=0}^{\infty} P_{N,k} |k - pN|^\alpha = O(N^{\alpha/2}).
\]  

(2.12)

holds. The case \( P = P^T_p \) of (2.10)--(2.12) can be found in [22] and the proofs there can
easily be adapted to the other cases.

Further properties of \( P^T_p \) and \( B_p \) are

\[
P^T_p P^T_q = p_{pq}, \quad B_p P^T_q = B_{pq}, \quad (B_p)_{N,k} = (B_1)_{pN,k}.
\]  

(2.13)

We will regard \( P \in \{ P^T_p, P^p_p, P^D_p, B_p \} \) as a sequence transformation matrix. For real
a

\[ s \in S_a \Rightarrow s_k = \begin{cases} O(N^a), & |k - Np| = O(N^{2/3}) \\ O(1 + N^a), & 0 \leq k \leq N \end{cases} \Rightarrow Ps \in S_a, \]  

(2.14)

\[ \Delta s \in S_a \Rightarrow (Ps)_N = s_{[Np]} + O(N^{a+1/2}), \]  

(2.15)

\[ \Delta^2 s \in S_a \Rightarrow (Ps)_N = s_{[Np]} + O(N^{a+2}). \]  

(2.16)
3. Proof of Theorem 1

First we need four lemmas. Lemma 1 is our tool to investigate the rate of growth of the solution $I$ of (2.4) and $v$ of (2.5) in dependence of the rate of growth of $r$ and $s$. However, a more general equation is treated in Lemma 1 in view of applications given in Section 5. With the aid of Lemma 2 we are able also to give estimates of the rate of growth of the first and second differences of $I$, which in turn are needed to obtain Taylor-like expansions of $\varepsilon_k$ around $k = Np$, that will be used in a way indicated in (2.15) and (2.16). Lemma 3 is concerned with estimating the sequence $s$ of (2.6). Lemma 4 is needed to estimate the solution of a certain functional equation satisfied by functions $V_1, V_2$, which are closely related to the Poisson generating function $V(z) = (B_1v)_z = e^{-z} \sum_{k=0}^{\infty} (z^k/k!) v_k$ of the sequence $v$.

**Lemma 1.** Let $a, b > 0, 0 < p \leq q < 1$, (not necessarily $p + q = 1$) and let $\alpha$ denote the unique real zero of $f(s) = 1 - ap^s - bq^s$. For fixed $n_0 \in \mathbb{N}$, $n_0 > \alpha$ and fixed $\sigma$ let $M_p = M_p(n_0, \sigma)$ and $M_q = M_q(n_0, \sigma)$ be defined by (2.7). Let a sequence $x := (x_k)_{k \geq 0}$ be given and $y := (y_k)_{k \geq 0}$ be defined by

$$(I - aM_p - bM_q)y = x.$$ 

(Note that $I - aM_p - bM_q$ is lower triangular with positive diagonal elements and is therefore invertible.)

Then the following statements are true:

(a) $x_k = O(k^{\alpha - \epsilon})$ for some $\epsilon \in \mathbb{R}$ implies

$$y_N = \begin{cases} O(N^{\alpha}(1 + 1/\epsilon)) & \text{if } \epsilon > 0, \\ O(N^{\alpha} \log N) & \text{if } \epsilon = 0, \\ O(N^{\alpha - \epsilon}(1 + 1/\epsilon)) & \text{if } \epsilon < 0. \end{cases}$$

(b) $x_k \geq 0$, $k \geq 0$ and $x_{k'} \neq 0$ for some $k' \geq 0$ ($k' \geq 1$ in (case-P)) implies

$$y_N = \Omega(N^\alpha).$$

(c) Let $1 \leq k_0 \in \mathbb{N}$. If $x_k = k^\alpha$, $k \geq k_0$, then

$$y_N = \frac{N^{\alpha} \log N}{ap^\alpha \log (1/p) + bq^\alpha \log (1/q)} + y'_N,$$

where $y'_N = O(N^\alpha)$. If $x_k \geq ck^\alpha$, $k \geq k_0$, $c > 0$, then $y_N = \Omega(N^{\alpha} \log N)$.

**Proof.** (a) It is easily seen by induction that $(I - aM_p - bM_q)^{-1} \geq 0$ holds for $N, k \geq 0$. We can therefore assume $x_k \geq 0$ for $k \geq 0$, which implies $y_k \geq 0$ for $k \geq 0$.

Let $C_0 := y_0$. For $N \geq 1$ we set $C_N = y_N/N^{\alpha}$ and for $N \geq 0$ we set $C_N = \max_{0 \leq n \leq N} C_n$, which yields

$$y_n \leq C_N n^{\alpha} \leq C_N (1 + N^{\alpha}), \quad 1 \leq n \leq N. \quad (3.1)$$
From
\[ y = (aM_x + bM_y)y + x, \]  
we deduce by using (2.10) and (3.1) and the abbreviations \( N' = N + N^{2/3} \) and \( N'' = N + N^{2/3} \sgn(x) \) the following inequality, valid for \( N \geq n_0 \):
\[
C_NN^a = y_N \leq a\bar{C}_{[0,N']} \left( pN'' \right)^a + b\bar{C}_{[a,N']} \left( qN'' \right)^a + (c_{p,q} + c_{q,p})N^{-\gamma}\bar{C}_N(N^2 + 2) + x_N.
\]
Let \( q' = (1 + q)/2 \). There is a number \( n_1 > n_0 \) such that \( \left\lfloor qN'' \right\rfloor \leq \lfloor q'N \rfloor \) and
\[
\left( \frac{zN''}{N} \right)^a \leq z^a(1 + 2|\alpha|N^{-1/3})
\]
are satisfied for \( N \geq n_1 \) and both \( z = p \) and \( z = q \). Now let \( \gamma = \frac{1}{2} + \max(-\alpha,0) \). We thus achieve for \( N \geq n_1 \)
\[
C_N \leq (ap^a + bq^a)\bar{C}_{[q,N']}(1 + 2|\alpha|N^{-1/3}) + (D - 2|\alpha|)N^{-1/3}\bar{C}_N + EN^{-\varepsilon},
\]
\( D > 2|\alpha|, E > 0 \) being constants, and by the definition of \( \alpha \)
\[
C_N - \bar{C}_{[q,N']} \leq D\bar{C}_NN^{-1/3} + EN^{-\varepsilon}. \tag{3.3}
\]
Let \( \mathcal{N} \subset \mathbb{N} \) be the set of numbers indexing the jumps of \( (\bar{C}_N)_{N \geq 1} \), i.e. \( N \in \mathcal{N} \iff \bar{C}_{N-1} < C_N \). If \( \mathcal{N} \) is finite, then \( (C_N)_{N \geq 0} \) is bounded and the proof is complete. Else we define an infinite strictly increasing sequence \( (N_i)_{i \geq 0} \) by
\[
N_0 = 0,
\]
\[
N_{i+1} = \max\{N \in \mathcal{N} : \bar{C}_{\lfloor q'N \rfloor} \leq C_{N_i} \},
\]
which satisfies \( \lfloor q'N_{i+2} \rfloor > N_i \) (otherwise \( \lfloor q'N_{i+2} \rfloor \leq N_i \) implies \( C_{\lfloor q'N_{i+2} \rfloor} \leq C_{N_i} \), which contradicts the maximality of \( N_{i+1} \)) and therefore
\[
N_{i+2} > \frac{1}{q'}N_i. \tag{3.4}
\]
Inequality (3.3) now reads
\[
\bar{C}_{N_i} - \bar{C}_{N_{i-1}} \leq \bar{C}_{N_i} - C_{\lfloor q'N_{i+2} \rfloor} \leq D\bar{C}_{N_i}N_i^{-1/3} + EN_i^{-\varepsilon}
\]
and yields by telescoping
\[
\bar{C}_{N_m} - \bar{C}_{N_{k-1}} \leq D\bar{C}_{N_m} \sum_{i=k}^{m} N_i^{-1/3} + E \sum_{i=k}^{m} N_i^{-\varepsilon}. \tag{3.5}
\]
From (3.4) we have \( N_i = \Omega((q')^{-i/2}) \) and \( N_{m-i} = N_m \cdot O((q')^{i/2}) \), from which it is evident that there exists a number \( k_1 \) with
\[
D \sum_{i=k_1}^{\infty} N_i^{-1/3} < \frac{1}{2},
\]
and that

\[ \sum_{i=k_1}^{m} N_i^{-\epsilon} = \begin{cases} O(1 + 1/\epsilon), & \epsilon > 0, \\ O(\log N_m), & \epsilon = 0, \\ O(N_m^{-\epsilon}(1 + 1/\epsilon)), & \epsilon < 0. \end{cases} \]

holds. Now for any \( N \) there is a \( N_m \leq N \), such that \( C_N \leq \tilde{C}_{N_m} \), and extracting \( \tilde{C}_{N_m} \) from (3.5) with \( k_1 \) in place of \( k \) proves (a).

(b) We first observe, that \( (M_p)_N,k > 0 \) for \( N \geq n_0 \) and \( 0 \leq k < N \) (resp. \( 0 < k < N \) in \( \text{case-P} \)). By using (3.2), \( y_N \) can be computed recursively in terms of \( y_i, i < N \), and we easily see, that our assumptions on \( (x_k)_{k \geq 0} \) imply \( y_N > 0 \) for \( N \geq n_2 := \max(n_0, k') \).

We define \( C_N = \min_{N_2 \leq n \leq N} C_n \), and, as before, find an inequality, now

\[ C_N \geq C_{(M_p)_N}(1 - FN^{-1/3}), \]

where \( F > 0 \) is a constant. If \( C_N \) has only finitely many jumps, we are done. Else we define a sequence of jump-indices of \( C_N \), also denoted \( (N_i)_{i \geq 0} \), by

\[ N_0 = n_2, \quad N_{i+1} = \max\{N \in \mathbb{N} : (C_N - i > C_N) \wedge (C_{(M_p)_N} \geq C_{N_i})\}, \]

which also satisfies (3.4). Iterating yields

\[ C_N = C_{N_m} \geq C_{N_{k_1}} \geq \prod_{i=k}^{m} (1 - FN_i^{-1/3}). \]

By (3.4), there exists a number \( k_2 \) such that \( N_{k_2} \geq n_2 \) and \( \prod_{i=k_2}^{m} (1 - FN_i^{-1/3}) > \frac{1}{2} \), thus \( C_N \geq \frac{1}{2} C_{N_{k_2}} \), therefore \( C_N = \Omega(1) \), which proves (b).

(c) We use the Taylor expansion of \( f(k) := k^s \log k \) at \( k = N_p \)

\[ f(k) = f(N_p) + f'(N_p)(k - N_p) + O((k - N_p)^2 N^{s-2} \log N), \]

valid for \( |k - N_p| \leq N^{2/3} \), and the crude estimate \( f(k) = O(1 + f(N)) \), \( 0 \leq k \leq N \), and employ (2.9)–(2.11), which results in

\[ \sum_{k=1}^{N} (M_p)_{N,k} k^s \log k = (Np)^s \log (Np) + O(N^{s-1} \log N), \]

which leads to

\[ (I - aM_p - bM_q)y' = x' \quad \text{with} \quad x'_N = O(k^{s-1} \log N). \]

Now part (a) can be invoked to give the estimate for \( y_N \). The proof of the second statement follows now from the observation that \( (I - aM_p - bM_q)^{-1} \) has nonnegative elements. This completes the proof of Lemma 1. \( \square \)

**Lemma 2.** (a) The differences of a sequence \( x \) and the differences of \( P_p^T x \) and \( P_p^D x \) are connected via

\[ \Delta P_p^T x = pP_p^T \Delta x, \quad (\Delta P_p^D x)_N = \begin{cases} x_0, & N = 0, \\ (pP_p^D \Delta x)_N, & N \geq 1. \end{cases} \]
and \( x \in S_a, a \in \mathbb{R} \), implies

\[
(\Delta P_p^T x)_N = (p P_p^T \Delta x)_N + O(N^a(p^N + q^N)).
\]

(b) For fixed \( p, 0 < p < 1 \) and \( x \in S_a, a \in \mathbb{R} \), we have for \( m \in \mathbb{N} \)

\[
\Delta^m P_p^T x \in S_{a-m/2}, \quad \Delta^m P_p^D x \in S_{a-m/2}, \quad \text{and} \quad \Delta^m P_p^p \in S_{a-m/2}.
\]

Proof. We will treat \( P_p^T \) first, the statements on \( P_p^D \) and \( P_p^p \) will then be easy consequences. Let \( b(k; N, p) = (\frac{1}{k})p^k(1-p)^{N-k}. \) Then

\[
(\Delta P_p^T x)_N = \sum_{k=0}^{N+1} b(k; N + 1, p)x_k - \sum_{k=0}^{N} b(k; N, p)x_k
\]

\[
= (1 - p) \sum_{k=0}^{N+1} b(k; N, p)x_k + p \sum_{k=0}^{N+1} b(k - 1; N, p)x_k - \sum_{k=0}^{N} b(k; N, p)x_k
\]

\[
= p \sum_{k=0}^{N} b(k; N, p)(x_{k+1} - x_k),
\]

which proves (a). The latter difference can be treated as

\[
|\Delta P_p^T x)_N| = \left| \sum_{k=0}^{N+1} b(k; N + 1, p) \frac{k - p(N + 1)}{(1 - p)(N + 1)} \right| x_k
\]

\[
\leq \frac{1}{(1 - p)(N + 1)} \sqrt{\sum_{k=0}^{N+1} b(k; N + 1, p) [k - p(N + 1)]^2 x_k^2}
\]

\[
= O(N^{a-1/2}).
\]

We used the Cauchy–Schwarz inequality, (2.11), the estimate \( x_k = O(N^a) \) for \( |k - Np| < pN^{2/3} \) and (2.10) for the sum over \( |k - Np| \geq pN^{2/3} \). This proves the case \( m = 1 \) of (b). For general \( m \geq 1 \) let \( \bar{p} \) denote the positive \( m \)th root of \( p \). Then by (2.13) and part (a) of this lemma,

\[
(\Delta^m P_p^T x)_N = (\Delta^m(P_p^T)^m x) = \bar{p}^{m(m-1)/2}((\Delta P_p^T)^m x)_N = O(N^{a-m/2}),
\]

which completes the proof of (b). The proof of the statements about \( P_p^D \) consists in replacing \( N \) by \( N - 1 \) and treating the case \( N = 0 \) separately. The statements about \( P_p^p \) are proved by making use of

\[
(P_p)_{N,k} = (P_p^T)_{N,k} + O(p^N + q^N). \tag{3.6}
\]

Lemma 3. Let \( 0 < \varepsilon < \frac{1}{2} \). Then under each of the hypotheses (a), (b) or (c) of Theorem 1 the sequence \( s \) defined by (2.6) satisfies

\[
s_N = \begin{cases} 
Npq(\Delta f'_{[Np]} - \Delta f'_{[Nq]})^2 + O(N^{1-\varepsilon}), & \text{cases (a), (b),} \\
O(N^{1-\varepsilon}), & \text{if } p = \frac{1}{2},
\end{cases}
\]

where \( l' \) is given by

\[
l = l' + r. \tag{3.7}
\]
Proof. We only treat (case-T) and (case-l'). The proof of (case-D) consists then just in repeating the arguments with every "N - k" changed for a "N - 1 - k". We distinguish in our proof the case, that \( r \) satisfies hypotheses (a) or (b), and the case, that \( r \) satisfies hypothesis (c).

Cases (a) and (b): We will make use of the fact, that \( l' \) is the solution of

\[
(I - M_p - M_q)l' = r',
\]

with

\[
r' = (M_p + M_q)r.
\]

\( r \) is the "small" term of the sum (3.7), for \( r \in \mathcal{S}_{1-\varepsilon} \), but only \( l' \in \mathcal{S}_1 \). On the other hand, \( l' \) is the "smooth" term of the sum (3.7). That is seen by Lemmas 1 and 2, since \( l' \) is the solution of (3.8) with "smooth" r.h.s. \( r' \).

According to (3.7) we have to inspect the quantities \( r_k + r_{N-k} \) and \( \ell_k + \ell_{N-k} \). We have

\[
r_k + r_{N-k} = \begin{cases} O(N^{1/2-\varepsilon}) & \text{in case (a)} \\ r_{[Np]} + r_{[Nq]} + O(1 + N^{-\varepsilon}|k - Np|) & \text{in case (b)} \end{cases}
\]

and \( r_k + r_{N-k} = O(N^{1-\varepsilon}) \), if \( 0 \leq k \leq N \), therefore by (2.14) and (2.15)

\[
\sum_{k=0}^{N} (M_p)_{N,k}(r_k + r_{N-k}) = \begin{cases} O(N^{1/2-\varepsilon}) & \text{in case (a)}, \\ r_{[Np]} + r_{[Nq]} + O(N^{1/2-\varepsilon}) & \text{in case (b)}. \end{cases}
\]

Now by (2.14) and Lemma 1 we have \( r \in \mathcal{S}_{1-\varepsilon} \Rightarrow r' \in \mathcal{S}_{1-\varepsilon} \Rightarrow l' \in \mathcal{S}_1 \) and applying Lemma 2 to (3.8) yields

\[
\begin{align*}
(I - pM_p - qM_q)\Delta l'_{N} - \Delta r'_{N} + O(N(p^N + q^N)), \\
(I - p^2M_p - q^2M_q)\Delta^2 l'_{N} = \Delta^2 r'_{N} + O(N(p^N + q^N)).
\end{align*}
\]

(3.10)

Lemma 2 also says that because of (3.9) the first and second differences of \( r' \) satisfy \( \Delta r'_k = O(k^{-\varepsilon}) \) and \( \Delta^2 r'_k = O(k^{-1/2-\varepsilon}) \), so by Lemma 1 the solutions of (3.10) satisfy

\[
\Delta \ell'_k = O(1), \quad \Delta^2 \ell'_k = O(k^{-1/2-\varepsilon})
\]

(3.11)

and we have

\[
\ell'_k + \ell'_{N-k} = \ell'_{[Np]} + \ell'_{[Nq]} + (\Delta \ell'_{[Np]} - \Delta \ell'_{[Nq]})(k - Np) + O(N^{-1/2-\varepsilon}(k - Np)^2)
\]

for \( |k - Np| = O(N^{2/3}) \), and by (2.16)

\[
\sum_{k=0}^{N} (M_p)_{N,k}(\ell'_k + \ell'_{N-k}) = \ell'_{[Np]} + \ell'_{[Nq]} + O(N^{1/2-\varepsilon}).
\]
We summarize
\[
\ell_k + \ell_{N-k} - \sum_{k=0}^{N-1} (M_p)_{N,k}(\ell_k + \ell_{N-k})
= (\Delta'_{[Nq]} - \Delta'_{[Np]})(k - Np) + O(N^{1/2-\epsilon}) + O(N^{-1/2-\epsilon}(k - Np)^2).
\]
This has to be plugged into (2.6), and, using (2.10)-(2.12), we obtain
\[
s_N = Npq(\Delta'_{[Np]} - \Delta'_{[Nq]})^2 + O(N^{1-\epsilon}).
\]

Case (c). In case (c) \(\Delta^2 r_k = O(k^{-1/2-\epsilon})\) holds, therefore the last lines are true with \(l, r\) in place of \(l', r'\), particularly
\[
s_N = Npq(\Delta'_{[Np]} - \Delta'_{[Nq]})^2 + O(N^{1-\epsilon}).
\]
If \(p = \frac{1}{2}\), we have \([Np] = [Nq]\), therefore \(s_N = O(N^{1-\epsilon})\). \(\square\)

Lemma 4. Let \(f: \mathbb{R}^+ \rightarrow \mathbb{R}\) be continuous and satisfy
\[
f(x) = \begin{cases} O(x), & x \rightarrow 0 \\ O(x^{-\epsilon}), & x \rightarrow \infty \end{cases}
\]
for some \(\epsilon > 0\). Let \(0 < p, q < 1\) and \(p + q = 1\).

Then \(F(x) := \sum_{k=0}^{\infty} (i_{p}^k p^q)^{x^k} = O(1)\) for \(x \geq 0\).

Proof. We assume, without loss of generality, that
\[
f(x) = f_0(x) := \begin{cases} x, & 0 \leq x \leq 1 \\ x^{-\epsilon}, & x > 1 \end{cases},
\]
since any \(f\) in question is smaller in absolute value than a constant times \(f_0\). For \(x \leq 1\) we have by direct computation \(F(x) = x/(1 - p^2 - q^2) \leq 1/(1 - p^2 - q^2)\). The proof for \(x > 1\) calls for use of the Mellin inversion formula (see [23] for application-oriented examples of the Mellin transform and [2] for a detailed treatment.) The Mellin transform \(F^*\) of \(F\) is given by
\[
F^*(s) = \int_0^{\infty} x^{s-1} f(x) \, dx \sum_{\lambda, \ell = 0}^{\infty} (\lambda + \ell)^{(4^s+1)}(p^q q^s x) = O(1) \quad \text{for} \quad x \geq 0.
\]
and is analytic in the strip \(-1 < \Re s < 0\). By the Mellin inversion formula
\[
F(x) = \frac{1}{2\pi i} \int_{(0, s)} x^{-\epsilon} \frac{1 + \epsilon}{1 - p^{1-s} - q^{1-s} (s + 1)(\epsilon - s)} \, ds
\]
holds. The rest is residue calculus. From [3] we know, that the set \(\mathcal{Z}\) of zeros of \(\phi(s) := 1 - p^{1-s} - q^{1-s}\) is a uniformly discrete subset of \( \{s \in \mathbb{C} : \Re s \geq 0\} \) in the sense
that there exists \( c > 0 \) such that \( \forall s, s' \in \mathcal{L}, s \neq s' \Rightarrow |s - s'| > c \). The path of integration is now shifted to the right of \( \Re s = 0 \), say to \( \Re s = \epsilon/2 \) and deformed in a way that poles of \((1 - p^{-1-s} - q^{-1-s})^{-1}\) near \( \Re s = \epsilon/2 \) are driven round at small indentations to the left or right. This results in a path \( \mathcal{C} \) of integration, which is contained in the strip \( \epsilon/4 < \Re s < 3\epsilon/4 \). The integral along \( \mathcal{C} \) is \( O(x^{-\epsilon/4}) = O(1) \) for \( x > 1 \), and, since \( |\phi'(s)| \geq \min(\log(1/p), \log(1/q)) \) for \( s \in \mathcal{L} \) and \( f^*(s) = O(s^{-2}) \) for \( s \to \infty \), the residues from poles to the left of \( \mathcal{C} \) form an absolute convergent series, that converges uniformly for \( x \geq 1 \). This completes the proof of Lemma 4. [QED]

Now we are ready to prove Theorem 1 from Section 1.

**Proof of Theorem 1.** (i) We distinguish two cases. The first is \( \forall N \in \mathbb{N}: \text{Var } L_N = 0 \), the second is \( \exists N_0 \in \mathbb{N}, N_0 \geq 2: \text{Var } L_{N_0} > 0 \). \( (\text{Var } L_0 = \text{Var } L_1 = 0 \) is always fulfilled.)

The first case says, that each \( L_N \) is constant (we need not say almost surely, because \( T_N \) does not contain trees \( t \) with \( \mathbb{P}(t) = 0 \)). By (1.1) we have

\[
L_k + L_{N-k} = L_N - r_N =: c_N
\]

and

\[
L_{k+1} + L_{N-k} = L_{N+1} - r_{N+1} = c_{N+1}
\]

for \( N \geq n_0 \) and \( 0 \leq k \leq N \) in (case-T), (case-D) needs replacement of \( N - k \) by \( N - 1 - k \) and in (case-P) the above equations are valid only for \( 0 < k < N \) due to the fact that a nonempty Patricia trie has no empty subtrees. Subtracting yields for \( k \geq 1 \)

\[
L_{k-1} - L_k = c_{N+1} - c_N =: c
\]

and \( L_k = ck + d \) for some constant \( d \), from which we deduce, e.g., for (case-T)

\[
r_k = \begin{cases} 
ck + d, & 1 \leq k < n_0 \\
- d & k \geq n_0
\end{cases}
\quad (3.12)
\]

\( L_k = ck + d \) is even true for \( k = 0 \) in (case-T) and (case-D), \( L_k = ck + d + b\delta_{k,0} \) for some \( b \) is the analogue for (case-P).

The second case says \( s_{N_0} > 0 \), and since \( s_N \geq 0 \) for \( N \geq 0 \), Lemma 1(b) can be applied and gives \( \text{Var } L_N = \Omega(N) \), which completes the proof of the \( \Omega \)-part.

(ii) We can assume, without loss of generality, that \( \epsilon < 1/2 \), because if \( a < b \), then \( r_k = O(k^a) \) implies \( r_k = O(k^b) \). The only purpose of this assumption is, that it allows us to write things like \( O(N^{1-\epsilon} + N^{1/2}) = O(N^{1-\epsilon}) \).

Lemma 1 guarantees that if \( s_k = O(k^{1-\epsilon}) \) for fixed \( \epsilon > 0 \), then \( v_k = O(k) \). Later on we will derive \( s_k = \bar{s}_k + \hat{s}_k \), where \( \bar{s}_k = O(k^{1-\epsilon}) \) and \( \hat{s} := (I - M_p - M_q)^{-1}\bar{s} \) is given up to terms \( O(N^{1-\epsilon}) \) by an infinite series which will be shown to be \( O(N) \).

First of all we can restrict our attention to the case \( r_0 = r_1 = 0 \), for otherwise we could write \( r \) as \( r = r' + r'' \), where \( r'' = r''_0 = 0 \) and where \( r' \) is of the form (3.12), giving rise to a parameter \( L' \), which is constant \( cN + d \) on \( T_N \). On the other hand, if \( r \) satisfies (a), (b) or (c), so does \( r'' \). So if \( \text{Var } L_N \) exists, then \( \text{Var } L_N = \text{Var } (L'_N + L''_N) = \text{Var } L''_N \).
If \( p = \frac{1}{2} \), we have \( s_N = O(N^{1-\epsilon}) \) by Lemma 3, hence Lemma 1 yields \( v_N = O(N) \), and the \( O \)-part of the proof is complete in the symmetric case.

It remains to treat cases (a) and (b) in the asymmetric case. Let

\[
\bar{s}_N := Npq(\Delta s_{[np]} - \Delta s_{[nq]})^2
\]

for \( N \geq 2 \) and \( \bar{s}_0 = \bar{s}_1 = 0 \). Then by Lemma 3, \( \bar{s}_N := s_N - \bar{s}_N = O(N^{1-\epsilon}) \), and by Lemma 1, \( \bar{s}_N \) contributes a term \( \bar{v}_N = O(N) \) to \( v_N \). By (3.11) we have \( \bar{s}_N = O(N) \) and Lemma 1 yields

\[
\bar{v}_N = O(N \log N),
\]

where \( \bar{v}_N \) is the solution of

\[
(I - M_p - M_q)\bar{v} = \bar{s}.
\]

Since \( \Lambda \bar{s}_N = O(N^{1/2 - \epsilon}) \), we have

\[
\Delta \bar{s}_N = O(N^{1/2 - \epsilon}).
\]

We will transform (3.15), using \( B_1 \) defined in (2.8). Note that a slight modification of (2.15) applied to (3.16) yields for \( N \in \mathbb{N} \)

\[
(B_1 \bar{v})_N = \bar{v}_N + O(N^{1-\epsilon}),
\]

so that, in order to complete the proof, we only have to show \( (B_1 \bar{v})_N = O(N) \).

We have to consider \( B_1 M_p \). Of course \( B_1 P_p^T = B_p \), as mentioned in (2.13). But \( M_p \) can differ from \( P_p^T \) in two respects: The first \( n_o \) rows of \( M_p(n_o, \sigma) \) are 0, and possibly \( \sigma \neq T \).

\[
M_p(n_o, \sigma) = M_p(0, T) + \underbrace{M_p(0, \sigma) - M_p(0, T)}_{M_p^{2,=}} + \underbrace{M_p(n_o, \sigma) - M_p(n_o, T)}_{M_p^{n_o,=}}.
\]

Only the first \( n_o \) rows of \( M_p^* \) contain nonzero elements, which means that \( (M_p^* \bar{v})_N = 0 \) for \( N \geq n_o \), hence

\[
(B_1 M_p^* \bar{v})_N = O(e^{-N^2} N^{n_o - 1}).
\]

If \( \sigma = P \), then \( (M_p^* \bar{v})_N = O(N^3(p^N + q^N)) \) by (3.6), hence after a short calculation

\[
(B_1 M_p^* \bar{v})_N = O(N^3(e^{-pN} + e^{-qN})).
\]

Finally, if \( \sigma = D \), for \( N \geq 1 \) we have

\[
(M_p^* \bar{v})_N = [(P_p^D - P_p^T) \bar{v}]_N = - (\Delta P_p^T \bar{v})_N = O(N^{1/2} \log N)
\]

by Lemma 2 and (3.14), hence

\[
(B_1 M_p^* \bar{v})_N = O(N^{1/2} \log N) = O(N^{1-\epsilon}).
\]

The promised transformation of (3.15) now reads

\[
(B_1 - B_p - B_q) \bar{v} = B_1 \bar{s} + s^1,
\]
with $s^1 = B_1(M_p - P^T_p + M_q - P^T_q)^\epsilon \in \mathcal{S}_{1-\epsilon}$. In the following the entire function $\tilde{V}$, defined by

$$
\tilde{V}(z) = (B_1\tilde{\epsilon})_z = e^{-z} \sum_{k=0}^\infty \frac{z^k}{k!} \tilde{\epsilon}_k
$$

will be used. $\tilde{V}(z)$ is obviously the Poisson generating function of the sequence $(\tilde{\epsilon}_k)_{k>0}$. Its purpose is indicated by $(B_1\tilde{\epsilon})_N = \tilde{V}(pN)$. We decompose the r.h.s. of (3.18)

$$
s^1 + (B_1\tilde{\epsilon})_x = \left[ s^1_x + (B_1\tilde{\epsilon})_x - xpq(B_p\Delta x' - B_q\Delta x'_2) \right] + xpq(B_p\Delta x' - B_q\Delta x')_x^2
$$

and set

$$
\tilde{V}(N) = V_1(N) + V_2(N),
$$

where $V_i$ corresponds to $f_i$:

$$
V_i(x) - V_i(px) - V_i(qx) = f_i(x), \quad i = 1, 2.
$$

The unique entire solutions of these functional equations can be found by iteration (cf. [3, 27]):

$$
V_i(x) = \sum_{\lambda, \ell = 0}^\infty \binom{\lambda + \ell}{\ell} f_i(p^\epsilon q^\lambda x).
$$

We now need Lemma 4 to estimate the sums in (3.21). The following estimates of $f_i$ are available:

- $f_1(x) = O(x^2)$ as $x \to 0$, since we assumed $r_0 = r_1 = 0$,
- $f_2(x) = O(x)$ as $x \to \infty$ follows trivially from (3.11), and
- $f_1(x) = O(x^{1-\epsilon}),$ as $x \to \infty$ is proved by approximating in the sense of (2.15).

Lemma 4 can therefore be applied to $f_1(x)/x$ and yields $V_1(x) = O(x)$, as $x \to \infty$. The same arguments do not apply to $V_2(x)$, since the weak estimate $f_2(x) = O(x)$ as $x \to \infty$ only allows to get the result (3.14). Fortunately, $V_2(x)$ can be computed explicitly in terms of

$$
A(x) := (B_1\Delta x')_x \quad \text{and} \quad \rho(x) := A(x) - pA(px) - qA(qx).
$$

We have

$$
A(x) = \begin{cases} O(x), & x \to 0 \\ O(1), & x \to \infty \end{cases} \quad \text{and} \quad \rho(x) = \begin{cases} O(x), & x \to 0 \\ O(x^{-\epsilon}), & x \to \infty \end{cases}.
$$

This follows partly from $r_0 = r_1 = 0$, from $\Delta x'_x = O(1)$ and from $\rho(x) = (B_1\Delta x')_x + r^1_1$, where $r^1_1$ is a small term (i.e. $r^1_1 = O(x^{-1/2})$ as $x \to \infty$) caused by $M^*, ..., M^*_q$, when transforming (3.10) with $B_1$ (cf. $s^1$ in (3.18)). The above equation defining $\rho$ solved for $A$ by iteration yields

$$
A(x) = \sum_{\lambda, \ell = 0}^\infty c_{\lambda, \ell} \rho(p^\epsilon q^\lambda x),
$$
where \( c_{\lambda, \ell} = (q^\ell)^p q^\lambda \). In our new notation (3.20) reads

\[
V_2(x) = V_2(px) - V_2(qx) = x pq (A(px) - A(qx))^2,
\]

with solution

\[
V_2(x) = x pq \sum_{\lambda, \ell = 0}^{\infty} c_{\lambda, \ell} (A(p^{\lambda+1} q^{\ell} x) - A(p^{\lambda+1} q^{\ell} x))^2.
\]

This is not better than anything before, if we know just \( A(x) = O(1) \) as \( x \to \infty \). But the last series can be rearranged:

\[
V_2(x) = 2x \sum_{\lambda, \ell = 0}^{\infty} c_{\lambda, \ell} \rho(p^{\lambda+1} q^{\ell} x) A(p^{\lambda+1} q^{\ell} x) - x A^2(x) - \sum_{\lambda, \ell = 0}^{\infty} c_{\lambda, \ell} \rho^2(p^{\lambda+1} q^{\ell} x),
\]

which is easily seen to be \( O(x) \), since both \( \rho(x) A(x) \) and \( \rho^2(x) \) are \( O(x^{-\epsilon}) \), as \( x \to \infty \) and therefore fit for Lemma 4. But from \( V_2(x) = O(x) \) it follows that \( \{ B_1, \tilde{v}_N \} = O(N) \), which was seen (cf. (3.17)) to imply \( u_N = O(N) \). This completes the proof of the \( O \)-part.

(iii) We start with \( (I - 2M_{1/2}) v = s \), where we know from Lemma 3, that \( s_k = O(k^{-\epsilon}) \) holds. Following (3.7), we define \( u' \) by \( u = u' + s \) and get

\[
0_2N - 20N = s_{2N} - 2s_{N} + 2(M_{1/2} u)_{2N} + 2[(M_{1/2} u')_{2N} - u_{N}] = O(N^{1-\epsilon}),
\]

where the term in brackets is estimated using (2.15) and \( \Delta u' \in \mathcal{F}_{1/2} \). We thus have

\[
\left| \frac{v_{2N}}{2N} - \frac{v_N}{N} \right| = O(N^{-\epsilon}). \tag{3.22}
\]

For \( N \in \mathbb{N} \) fixed, \( (v_{2^kN}/2^kN)_{k \geq 0} \) is therefore a Cauchy sequence converging to a limit \( a(N) \), which certainly fulfills \( a(N) = a(2N) \), and

\[
\left| a(N) - \frac{v_N}{N} \right| \leq \sum_{k \geq 0} \left| \frac{v_{2^{k+1}N}}{2^{k+1}N} - \frac{v_{2^kN}}{2^kN} \right| = O(N^{-\epsilon}) \tag{3.23}
\]

by (3.22). This is what was claimed about the existence of a periodicity and the error term. We now show, that the function \( b \), defined on the positive dyadic rationals by \( b(N2^{-k}) := a(N) \), is continuous. We need

\[
a(N+1) - a(N) = a(N+1) - \frac{v_{N+1}}{N+1} + \frac{v_{N+1}}{N+1} - \frac{v_N}{N} + \frac{v_N}{N} - a(N) = O(N^{-\epsilon}), \tag{3.24}
\]

which is clear from (3.23) and \( |(v_{N+1})/(N+1) - v_N/N| = O(N^{-\epsilon}) \), which follows from

\[
v_{N+1} - v_N = s_{N+1} - s_N + \Delta u'_{N} = O(N^{1-\epsilon}).
\]

We define piecewise linear continuous functions \( b_k : \mathbb{R}^+ \to \mathbb{R} \) on the sets \( \{ N2^{-k}, N \in \mathbb{N} \} \) by \( b_k(N2^{-k}) := a(N) \) and for \( x \in \mathbb{R}^+ \)

\[
b(x) := \lim_{k \to \infty} b_k(x) = b_0(x) + \sum_{k \geq 0} (b_{k+1}(x) - b_k(x)). \tag{3.25}
\]

\[
W. Schachinger / Theoretical Computer Science 144 (1995) 251–275 261
\]
The latter definition of \( b \) coincides with the former on the set of positive dyadic rationals. Since by (3.24)
\[
\sup_{x \in [1,2]} |b_{k+1}(x) - b_k(x)| \leq \max_{2^k < N < 2^{k+1}} |a(N+1) - a(N)| = O(2^{-ck}),
\]
the series in (3.25) converges uniformly and hence \( b(x) \) is continuous in \([1, 2]\) and by periodicity in \( \mathbb{R}^+ \). The function we were looking for is \( \delta(x) := b(2^x) \). The proof of Theorem 1 is now complete. \( \square \)

Adopting the very last proof, the following corollary is easily verified.

**Corollary 1.** Let \( L_N \) be as in Theorem 1. Then in the symmetric case \( (p = \frac{1}{2}) \)
\[ r_k = O(k^{1-\varepsilon}) \]
implies
\[ E L_N = \tau(\log_2 N) \cdot N + O(N^{1-\varepsilon}), \]
with a continuous periodic function \( \tau \) with period 1.

**4. Remarks**

Can the conditions (a), (b) or (c) of Theorem 1 be weakened such that the conclusion is still true? The answer is yes, if you use a refined asymptotic scale. (a) could for example be replaced by
\[ (a') \quad r_N = O\left(\frac{\sqrt{N}}{\log^{1/2+\eta} N}\right) \]
for some \( \eta > 0 \),
since a refined version of Lemma 1(a) (whose proof does not differ in any respect from the original one) states:
\[ r_N = O\left(\frac{N^{\eta'}}{\log^{1+\eta'} N}\right) \Rightarrow \ell_N = O(N^{\eta'}), \]
if \( \eta' > 0 \).

However, if we stick to (a)–(c), as they are stated in Theorem 1, we cannot allow \( \varepsilon \leq 0 \) in either of them, as is seen by the following counterexample.

Let a parameter \( L \) be defined as in (1.1) with \( n_0 = 1 \) on the class of symmetric tries by fixing the values of the expectations:
\[
\ell_N = \begin{cases} 
0, & N = 0, 1, \\
\sqrt{N} \sin \sqrt{N}, & N \geq 2.
\end{cases}
\]

(4.1)

The corresponding sequence \( (r_N) \) can be computed using (2.4) and is seen not to satisfy either of conditions (a), (b) and (c) of Theorem 1, but being very close to it. We will
show $\text{Var } L_{N_i} = \Omega(N_i \log N_i)$ for infinitely many $N_i \in \mathbb{N}$. In order to compute $s$ we use the approximation

$$
\sqrt{k} \sin \sqrt{k} + \sqrt{N - k} \sin \sqrt{N - k}
$$

$$
= 2\sqrt{N/2} \sin \sqrt{N/2} \cos \left( \frac{1}{2} \sqrt{\frac{k - N/2}{N/4}} \right) + O\left( \frac{|k - N/2|}{\sqrt{N}} \right),
$$

which leads to

$$
s_N = (1 - e^{-1/8})^2 N \sin^2 \frac{\sqrt{N}}{2} + O(\sqrt{N}).
$$

Lemma 1(c) cannot directly be applied in order to show that $v_N = O(N)$ does not hold, because we do not have $s_N = \Omega(N)$. Since $s_0 = s_1 = 0$ by (4.1), $v$ is the solution of

$$
(I - 2P_{1/2}^T) v = s,
$$

satisfying $v_0 = v_1 = 0$. Let $C$ be the matrix with

$$
C_{N,k} = \begin{cases} 
(N + 1)^{-1}, & 0 \leq k \leq N, \\
0, & k > N,
\end{cases}
$$

Now $C$ commutes with $P_{1/2}^T$, which we use to derive $(I - 2P_{1/2}^T) CV = CS$. It is easily seen that $(CS)_N = \Omega(N)$ holds. By Lemma 1(c) we have $(CV)_N = \Omega(N \log N)$, which would be contradicted by $v_N = o(N \log N)$, since that implies $(CV)_N = o(N \log N)$.

Another counterexample is the external path length of an asymmetric trie considered in Section 5.3, whose corresponding sequence $(r_N)$, defined by $r_N = N - \delta_{N,1}$, violates (a) and (b) of Theorem 1, but satisfies $\Delta^2 r_N = 0$ for $N \geq 2$, so that we cannot drop $p = 1/2$ in condition (c).

5. Applications

Here we are going to demonstrate the power of Theorem 1 in applying it to four of the most important inductive valuations on data structures for digital search. We define these valuations only for tries. This is partly due to the fact, that the number of internal nodes defined in Section 5.1 is of interest only in the case of the trie. The results of Sections 5.2 and 5.3 can easily be carried over to Patricia tries and digital search trees, there being only unimportant differences in the definitions of the corresponding valuations, that do not affect the asymptotic growth of the sequence $(r_N)$. More care is needed in Section 5.4, since the approximation of $P_r^T P_q^T$ (resp. $P_r^D P_q^D$) by $P_r^q$ (resp. $P_r^D$) gives rise to additional terms which have to be estimated.
5.1. The number $L$ of internal nodes of a binary trie

In contrast to the case of Patricia trie and digital search tree, where the number of internal nodes is completely determined by the size (the number of records) of the tree, in the case of the trie there is no such connection. $L$ can be defined inductively by

$$L(t) = \begin{cases} 0, & |t| \leq 1, \\ 1 + L(t_L) + L(t_R), & |t| > 1, \end{cases}$$

where $t_L$ and $t_R$ are the left and right subtrees of $t$, respectively.

The corresponding sequence $(r_k)$ (compare (1.1) and (2.1)) is

$$r_k = \begin{cases} 0, & k \leq 1, \\ 1, & k > 1. \end{cases}$$

It satisfies condition (a) of Theorem 1. Now by Lemma 1 we have

$$\mathbb{E}L_N \asymp N$$

and by Theorem 1

$$\text{Var } L_N \asymp N.$$

Moreover, by Corollary 1 and Theorem 1 (iii) for the symmetric trie

$$\mathbb{E}L_N \sim \tau(\log_2 N) \cdot N, \quad \text{Var } L_N \sim \sigma(\log_2 N) \cdot N$$

holds with continuous periodic functions $\tau$ and $\sigma$ with period 1. The result on $\mathbb{E}L_N$ in the symmetric case is due to Knuth [16], who also computed the Fourier coefficients of $\tau$. The result on $\text{Var } L_N$ for the symmetric trie was found by Kirschenhofer and Prodinger [11] and independently by Régnier and Jacquet [25]. In [11] the authors used transformation formulae of Ramanujan to prove the vanishing of the coefficient at $N^2$ in their expression for $\text{Var } L_N$. In [25] the authors also considered the case of asymmetric tries. In [9] they proved asymptotic normality of $L_N$ using bivariate generating functions.

5.2. The number $L$ of external internal nodes of a binary trie

An internal node that is followed by two external nodes is called external internal node. $L$ can be defined inductively by

$$L(t) = \begin{cases} 0, & |t| \leq 1, \\ 1, & |t| = 2, \\ L(t_L) + L(t_R), & |t| > 2, \end{cases}$$

The corresponding sequence is $r_k = \delta_{k,2}$, which satisfies condition (a) of Theorem 1. By Lemma 1 we have $\mathbb{E}L_N \asymp N$ and by Theorem 1 $\text{Var } L_N \asymp N$. As in Section 5.1, better asymptotics are available for the symmetric case by Corollary 1 and Theorem 1(iii). Flajolet and Sedgewick [5] proved $\mathbb{E}L_N \sim \tau(\log_2 N) \cdot N$ with continuous
periodic function \( r \) with period 1 (not the same \( r \) as in Section 5.1), and computed the mean of \( r \). The corresponding result for \( \text{Var} \, L_N \) can be found in [11]. To our knowledge there are so far no results (not even a \( O \)-estimate) concerning \( \text{Var} \, L_N \) in the case of asymmetric tries.

5.3. The external path length \( L \) of a binary trie

\( L(t) \) denotes the sum of the distances of the external nodes to the root measured in edges. There is a close relationship between \( L(t) \) and the average number of bit inspections in a successful search in \( t \). \( L \) can be defined inductively by

\[
L(t) = \begin{cases} 
0, & |t| \leq 1 \\
|t| + L(t_1) + L(t_2), & |t| > 1, \quad t = t_1 \cup t_2 
\end{cases}
\]

The corresponding sequence is \( r_k = k - \delta_{k,1} \), which satisfies in the case of symmetric tries condition (c) of Theorem 1. By Lemma 1 we have \( \mathbb{E} L_N \sim N \log_2 N \) and by Theorem 1(iii) \( \text{Var} \, L_N \sim \delta (\log_2 N) \cdot N \). More precise asymptotics (including Fourier coefficients of the periodic functions) of \( \mathbb{E} L_N \) can be found in [16] and of \( \text{Var} \, L_N \) in [14]. See also [13]. In the asymmetric case the sequence \( r_k \) violates conditions (a) and (b) of Theorem 1, and, as we will see, \( \text{Var} \, L_N = O(N) \) is no longer true. In fact we can compute the main term of the asymptotics of \( \text{Var} \, L_N \) using only Lemma 1 and Theorem 1. From Lemma 1(c) we know

\[
\ell_N = \frac{N \log N}{h(p)} + \ell'_N = \ell''_N + \ell''_N,
\]

where \( h(p) = p \log (1/p) + q \log (1/q) \). To \( l' \) and \( l'' \) correspond in an obvious way sequences \( r' \) and \( r'' \), which define parameters \( L'_N \) and \( L''_N \). Using approximations as in Lemma 1(c), we get \( r'_N = O(1) \). Theorem 1 can be applied to the parameter \( L'_N \) and yields \( \text{Var} \, L'_N = O(N) \). The sequence \( s'' \) is in a similar manner seen to satisfy

\[
s''_N = \frac{\log^2(p/q)}{h^3(p)} pqN + O(1)
\]

from which we deduce by Lemma 1(c)

\[
\text{Var} \, L''_N = \frac{pq \log^2(p/q)}{h^3(p)} N \log N + O(N).
\]

Using now

\[
(\sqrt{\text{Var} \, X} - \sqrt{\text{Var} \, Y})^2 \leq \text{Var} \, (X + Y) \leq (\sqrt{\text{Var} \, X} + \sqrt{\text{Var} \, Y})^2
\]

we finally obtain

\[
\text{Var} \, L_N = \text{Var}(L'_N + L''_N) = \frac{pq \log^2(p/q)}{h^3(p)} N \log N + O(N \sqrt{\log N}).
\]
The variance of the internal pathlength (the sum of the distances of the internal nodes to the root measured in edges) of an asymmetric digital search tree was recently computed by Jacquet and Szpankowski [10], who not only have a better error term (O(N) instead of O(N √ log N)), but also could prove asymptotic normality. Limiting distributions of the bit inspections valuation in digital search trees can be found in [18] (symmetric case) and [20] (asymmetric case). For corresponding results on bit inspections in tries, cf. [8].

5.4. The number \( L \) of internal nodes visited during a partial match retrieval in a binary trie

In the design of database systems one has to deal with data whose keys have several components. A suitable datastructure is the so-called M-d trie (M-dimensional trie, cf. [4]). We consider the case \( M = 2 \): data, whose keys have two components (each an infinite string of i.i.d. 0–1 Bernoulli random variables, the two components being independent), can be stored in a trie by first producing new keys by “regular shuffling”. (Example: from the components \((a, b, c, ...), (z, y, x, ...)\) we get \((a, z, b, y, c, ...)\).) A binary trie is now built from the new keys. A particular problem called “partial match retrieval” consists in finding all data, whose keys match a certain search pattern, the first component being unspecified (denoted by \((\ast, \ast, \ast, \ast, \ldots)\)) and the second being the sequence \((0, 0, 0, \ldots)\), say. By \( L(t) = L_\omega(t) \) we denote the number of internal nodes visited during a partial match retrieval for \( \omega = (\ast, 0, \ast, 0, \ast, \ldots) \) in \( t \). At levels with even index (the root has level 0) we have to traverse both subtrees, at levels with odd index we traverse only the left subtree. With the use of an auxiliary parameter \( L^1 \) we define \( L = L^0 \) as follows:

\[
L^0(t) = \begin{cases} 0, & |t| \leq 1, \\ 1 + L^1(t_2) + L^1(t_1), & |t| > 1. \end{cases}
\]

\[
L^1(t) = \begin{cases} 0, & |t| \leq 1, \\ 1 + L^0(t_2), & |t| > 1. \end{cases}
\]

Let \( \ell^i = (\ell^i_N)_{N \geq 0} \) with \( \ell^i_N = \mathbb{E}L^i_N \) and let \( e \) be given by \( e_k = 0, k = 0, 1, e_k = 1, k > 1 \). Our definition of the parameters \( L^i \) above readily leads to

\[
\begin{pmatrix} l^0 \\ l^1 \end{pmatrix} = \begin{pmatrix} e \\ 1 + P^T_p + P^T_q \end{pmatrix} \begin{pmatrix} l^0 \\ l^1 \end{pmatrix},
\]

which (after one iteration) is equivalent to

\[
\begin{pmatrix} l^0 \\ l^1 \end{pmatrix} = \begin{pmatrix} 0 & P^T_p + P^T_q \\ 0 & P^T_p + P^T_q \end{pmatrix} \begin{pmatrix} l^0 \\ l^1 \end{pmatrix} \begin{pmatrix} e \\ 1 + P^T_p + P^T_q \end{pmatrix},
\]

which decomposes into two separate equations, that can be treated by Lemmas 1 and 2: Let \( \alpha \) be the unique real zero of \( f(s) = 1 - p^{2s} - (pq)^s = 0 \). Note that \( 0 < \alpha < 1 \). The above equations have unique solutions \( \ell^i \) that satisfy \( \ell^0_0 = \ell^1_0 = 0 \) and

\[
\ell^i_N \asymp N^\alpha, \quad \Delta \ell^i_N = O(N^{\alpha-1}), \quad \Delta^2 \ell^i_N = O(N^{\alpha-2}).
\]
Let \( v^i_N \) denote the variance of \( L^i_N \). The sequences \( v^i \) are now solutions of

\[
\begin{pmatrix}
v^0 \\
v^1
\end{pmatrix} = \begin{pmatrix}
s^1 \\
s^0
\end{pmatrix} + \begin{pmatrix}
P_p^T + P_q^T & 0 \\ 0 & P_p^T
\end{pmatrix} \begin{pmatrix}
v^0 \\
v^1
\end{pmatrix},
\]

where the \( s^i \) are given by

\[
\begin{align*}
s^1_N &= \sum_{k=0}^{N} (P_p^T)_{N,k} \left( \ell^1_k + \ell^1_{N-k} - \sum_{k=0}^{N} (P_p^T)_{N,k} (\ell^1_k + \ell^1_{N-k}) \right)^2, \\
s^0_N &= \sum_{k=0}^{N} (P_p^1)_{N,k} \left( \ell^0_k - \sum_{k=0}^{N} (P_p^1)_{N,k} \ell^0_k \right)^2.
\end{align*}
\]

By using the estimates of the differences of \( \ell^i \) it can be shown in a similar manner as in the proof of Theorem 1, that

\[
s^1_N = O(N^{2\alpha - 1}), \quad s^0_N = O(N^{2\alpha - 1})
\]

holds. Now we decompose the system of equations as in the case of the expectation, and since \( 2\alpha - 1 < \alpha \) another application of Lemma 1 yields the growth of the variance:

\[
\text{Var } L_N \asymp N^\alpha.
\]

It is easy to see, how this result could be extended to cases \( M > 2 \).

Flajolet and Puech [4] derived asymptotics of \( \mathbb{E}L_{\omega,N} \) under the symmetric Bernoulli model for \( M \geq 2 \) and periodic search patterns:

\[
\mathbb{E}L_{\omega,N} = \gamma \left( \frac{1}{M} \log_2 N \right) N^{1-s/M} + O(1),
\]

where \( \gamma \) is a periodic function with period 1, depending on \( \omega \), and \( s \) is the number of unspecified components in the search pattern. The analysis of [4] includes the computation of the Fourier coefficients of \( \gamma \). \( \mathbb{E}L_{\omega,N} \) under the asymmetric Bernoulli model for \( M \geq 2 \) and \( \text{Var } L_{\omega,N} \) under the symmetric Bernoulli model for \( M = 2 \) can be found in [15]. Our result, for the derivation of which we only used Lemmas 1 and 2, seems to be the first that concerns \( \text{Var } L_{\omega,N} \) under the asymmetric Bernoulli model.

6. Conclusion

We investigated inductive valuations on tries, Patricia tries and digital search trees in an unified framework under both the symmetric and asymmetric Bernoulli model with binary alphabet. A robust methodology is presented which provides precise informations about the asymptotic behavior of the variance of most of the important parameters on these digital trees. Some generalizations of our results seem straightforward: The essential properties of the splitting matrices used in our derivations are shared also by the splitting matrices (under the Bernoulli model) of other digital data.
structures such as \( b \)-tries and \( b \)-digital search trees with bucket capacity \( b > 1 \). Another input model, the so-called Poisson model, where the number \( N \) of records is a Poisson random variable, should also be tractable by our matrix approach (again \( B_p \) comes into play). Our results can also be extended to alphabets \( \Sigma \) of more than two symbols and probably to input models assuming keys with Markovian dependencies of symbols. We are sure that the existence of a normal limiting distribution of the sequence \( (L_N)_{N \geq 0} \) of Theorem 1 can be shown under like general conditions on \( r \) as in Theorem 1. A future paper will be devoted to this problem.

Acknowledgements

The author appreciated the referees' comments which led to considerable improvement of the presentation.

References