A Limit Theorem for Radix Sort and Tries with Markovian Input

Kevin Leckey and Ralph Neininger*
Institute for Mathematics
J.W. Goethe University Frankfurt
60054 Frankfurt am Main
Germany
{leckey, neiningr}@math.uni-frankfurt.de

Wojciech Szpankowski †‡
Department of Computer Science
Purdue University
W. Lafayette, IN 47907
U.S.A.
spa@cs.purdue.edu

May 27, 2015

Abstract

Tries are among the most versatile and widely used data structures on words. In particular, they are used in fundamental sorting algorithms such as radix sort which we study in this paper. While the performance of radix sort and tries under a realistic probabilistic model for the generation of words is of significant importance, its analysis, even for simplest memoryless sources, has proved difficult. In this paper we consider a more realistic model where words are generated by a Markov source. By a novel use of the contraction method combined with moment transfer techniques we prove a central limit theorem for the complexity of radix sort and for the external path length in a trie. This is the first application of the contraction method to the analysis of algorithms and data structures with Markovian inputs; it relies on the use of systems of stochastic recurrences combined with a product version of the Zolotarev metric.

1 Introduction

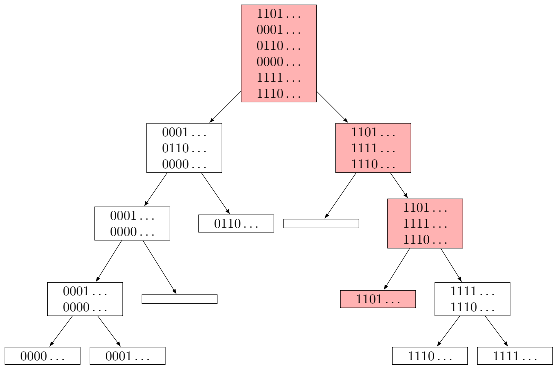
Tries are prototype data structures useful for many indexing and retrieval purposes. Tries were first proposed by de-la-Briandais in 1959 [4] for information processing. Fredkin in 1960 suggested the current name, part of the word re*trie*val [22, 25, 36]. They are pertinent to (internal) structure of (stored) words and several splitting procedures used in diverse contexts ranging from document taxonomy to IP addresses lookup, from data compression to dynamic hashing, from partial-match queries to speech recognition, from leader election algorithms to distributed hashing tables and graph compression.

Tries are trees whose nodes are vectors of characters or digits; they are a natural choice of data structure when the input records involve the notion of alphabets or digits. Given a sequence of n binary strings, we construct a trie as follows. If $n = 0$ then the trie is empty. If $n = 1$ then a single external node holding the word is allocated. If $n \geq 1$ then the trie consists of a root (i.e., internal) node directing strings to two subtrees according to the first symbol of each string, and strings directed to the same subtree recursively generate a trie among themselves, see Figure 1

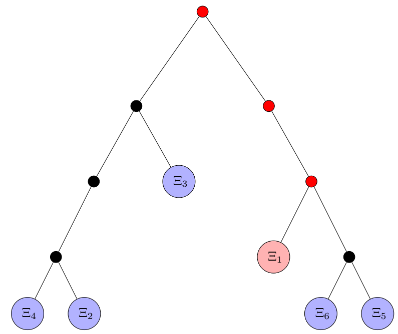
*This author's research was supported by DFG grant Ne 828/2-1.

†This author's work was partially done when visiting J.W. Goethe University Frankfurt a.M. on the Alexander von Humboldt research award. This work was also supported by NSF Center for Science of Information (CSoI) Grant CCF-0939370, and in addition by NSA Grant 130923, and NSF Grant DMS-0800568, NIH Grant 1U01CA198941-01, and the MNSW grant DEC-2013/09/B/ST6/02258.

‡Also with the Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, Poland.



(a) *Radix Sort*



(b) *Trie*

Figure 1: Radix sort and a trie applied to the strings: $\Xi_1 = 1101\ldots$, $\Xi_2 = 0001\ldots$, $\Xi_3 = 0110\ldots$, $\Xi_4 = 0000\ldots$, $\Xi_5 = 1111\ldots$, $\Xi_6 = 1110\ldots$ Note that radix sort places Ξ_1 into three sublists, also called buckets, (and has to read the first three symbols of Ξ_1) whereas the node storing Ξ_1 has depth three in the corresponding trie.

and Section 2 for a more formal definition. The *internal nodes* in tries are branching nodes, used merely to direct records to each subtrie; the record strings are all stored in *external nodes*, which are leaves of such tries.

Tries can be used in many fundamental algorithms, in particular for sorting known as radix sort or more precisely most significant digit radix sort [22]. In this cases, the n strings are binary representations of keys to be sorted. They are inserted in a trie as described above. A so-called depth-first traversal of the trie starting at the root node will visit each key in sorted order. In other words, keys that start with a 0 are moved to the left subtree also called a left bucket, while the other keys are stored in the right subtree or right bucket. In the sequel, we sort keys in the left and the right buckets using the second symbol, an so on as shown in Figure 1(a). A *recursive* description of the radix sort algorithm is presented in Section 2. In this paper, we shall use the trie and radix sort paradigms exchangeably. The complexity of such radix sort is equal to the external path length of the associated tries, that is, the sum of the lengths of the paths from the root to all external nodes.

We study the limit law of the radix sort complexity and the external path length of a trie built over n binary strings generated by a Markov source. More precisely, we assume that the input is a sequence of n independent and identically distributed random strings, each being composed of an infinite sequence of symbols such that the next symbol depends on the previous one and this dependence is governed by a given transition matrix (i.e., Markov model).

Digital trees, in particular, tries have been intensively studied for the last thirty years [3, 5, 14, 16, 18, 20, 6, 7, 22, 25, 36], mostly under Bernoulli (memoryless) model assumption. The typical depth under the Markov model was analyzed in [18], however, not the external path length. The external path length is more challenging due to stronger dependency, see [36]. In fact, this is already observed for tries under the Bernoulli model [36]. In this paper we establish a central limit theorem for the external path length in a trie built over a Markov model using a novel use of the *contraction method*.

The contraction method was introduced in 1991 by Uwe Rösler [31] for the distributional analysis of the complexity of the Quicksort algorithm. It was then developed independently by Rösler and by Rachev and Rüschendorf [30] in the early 1990's. Over the last 20 years this approach, which is based on exploiting an underlying contracting map on a space of probability distributions, has been developed as a fairly universal tool for the analysis of recursive algorithms and data structures. Here, randomness may come from a stochastic model for the input or from randomization within the algorithms itself (randomized algorithms). General developments of this method were presented in [32, 30, 33, 27, 28, 8, 19, 29] with numerous applications in computer science, information theory, and networking.

The contraction method has been used in the analysis of tries and other data structures only under the symmetric Bernoulli model (unbiased memoryless source) [27, Section 5.3.2], where limit laws for the size and the external path length of tries were re-derived. The application of the method there was heavily based on the fact that precise expansions of the expectations were available, in particular smoothness properties of periodic functions appearing in the linear terms as well as bounds on error terms which were $\mathrm{O}(1)$ for the size and $\mathrm{O}(\log n)$ for the path lengths. It should be observed that even in the asymmetric Bernoulli model such error terms seem to be out of reach for classical analytic methods; see the discussion in Flajolet, Roux, and Vallée [9]. Hence, for the more general Markov source model considered in the present paper we develop a novel use of the contraction method.

Furthermore, the contraction method applied to Markov sources hits another snag, namely, the Markov model is not preserved when decomposing the trie at its root into its left and right subtree. The initial distribution of the Markov source is changed when looking at these subtrees. To overcome these problems a couple of new ideas are used for setting up the contraction method: First of all, we will use a system of distributional recursive equations, one for each subtree. We then apply the contraction method to this system of recurrences capturing the subtree processes and prove asymptotic normality for the path lengths conditioned on the initial distribution. In fact, our approach avoids dealing with multivariate recurrences and instead we reduce the whole analysis to a system of one-dimensional equations. To come up with an appropriate contracting map we use a product version of the Zolotarev metric.

We also need asymptotic expansions of the mean and the variance for applying the contraction method. In contrast to very precise information on periodicities of linear terms for the symmetric Bernoulli model mentioned above and in view of the results in [9] mentioned above we cannot expect to obtain similarly precise expansions. In fact, our convergence proof does only require the leading order term together with a Lipschitz continuity property for the error term. The lack of a precise expansion is compensated by this Lipschitz continuity combined with a self-centering argument to obtain sufficiently tight control on error terms.

For the derivation of such an expansions of the mean (and the variance) we use moment transfer theorems. Such theorems were largely developed by H.-K. Hwang, see, e.g., [13, 10, 11, 1], for the control of moments related to one-dimensional recurrences. We extend such theorems to systems of recurrences as they occur for the analysis of our Markov model. For the expansion of the variance we also make use of a construction due to Schachinger [35].

This is the first application of the contraction method to the analysis of algorithms and data structures with Markovian inputs. Our results were announced in the extended abstract [24]. The methodology developed is general enough to cover related quantities and structures as well. Our approach also applies with minor adjustments at least to the path lengths of digital search trees and PATRICIA tries under the Markov source model, see the dissertation of the first mentioned author [23].

The Markov source model is more realistic and more flexible than the (memoryless) Bernoulli model. Even more general models have been analyzed in the context of tries. Vallée [37] introduced the dynamical source models which, in particular, cover the Markov model. The analysis of dynamical sources for tries started with the work of Clément, Flajolet and Vallée in [3], including the asymptotic of the expectation of several trie parameters such as height, size and the depth/external path length. There is a limit theorem for the depth in tries for special (so-called tame) dynamical sources, see [2], and a limit theorem for the depth in the (closely related) digital search tree for two types of general sources, see [12]. However, a limit theorem for the external path length in tries and the complexity of radix sort has not yet been derived for dynamical sources.

Notations: Throughout this paper we use the Bachmann–Landau symbols, in particular the big O notation. We declare $x \log x := 0$ for $x = 0$. By $B(n, p)$ with $n \in \mathbb{N}$ and $p \in [0, 1]$ the binomial distribution is denoted, by $B(p)$ the Bernoulli distribution with success probability p, by $\mathcal{N}(0, \sigma^2)$ the centered normal distribution with variance $\sigma^2 > 0$. We use C as a generic constant that may change from one occurrence to another.

2 Main Results

In this section we first describe succinctly the radix sort and his relation to tries. Then we present our probabilistic model, and the main result of this paper.

Radix sort. Given n keys represented by binary strings, we can sort them in the following way. We first split them according to the first bit: those string starting with a 0 go to the left bucket, while the others to the right bucket. In each bucket we sort remaining strings in the same manner using the second bit. And so on. At the end we read all keys from left to right and all n keys are sorted, see Figure 1. This is called a radix sort [22]. The number of inspected bits needed to sort such n keys (strings) is denoted by B_n and called it in short the number of bucket operations. It measures the complexity of radix sort. We study its limiting distribution in this paper.

It is easy to see that we can achieve the same result by building a trie from n strings and visit all external nodes in a tree traversal. Then B_n can be interpreted as the length of the external path length, that is, the sum of all paths from the root to all external nodes.

The Markov source: We now define the probabilistic model for string generation. We shall assume that binary data strings over the alphabet $\Sigma = \{0, 1\}$ are generated by a homogeneous Markov source. In general, a homogeneous Markov chain is given by its initial distribution $\mu = \mu_0\delta_0 + \mu_1\delta_1$ on Σ and the transition matrix $(p_{ij})_{i,j\in\Sigma}$. Here, δ_x denotes the Dirac measure in $x \in \mathbb{R}$. Hence, the initial state is 0 with probability μ_0 and 1 with probability μ_1. We have $\mu_0, \mu_1 \in [0, 1]$ and $\mu_0 + \mu_1 = 1$. A transition from state i to j happens with probability p_{ij}, $i, j \in \Sigma$. Now, a data string is generated as the sequence of states visited by the Markov chain. In the Markov source model assumed subsequently all data strings are independent and identically distributed according to the given Markov chain.

We always assume that $p_{ij} > 0$ for all $i, j \in \Sigma$. Hence, the Markov chain is ergodic and has a stationary distribution, denoted by $\pi = \pi_0\delta_0 + \pi_1\delta_1$. We have

$$\pi_0 = \frac{p_{10}}{p_{01} + p_{10}}, \qquad \pi_1 = \frac{p_{01}}{p_{01} + p_{10}}. \tag{1}$$

Note however, that our Markov source model does not require the Markov chain to start in its stationary distribution.

The case $p_{ij} = 1/2$ for all $i, j \in \Sigma$ is essentially the symmetric Bernoulli model (only the first bit may have a different initial distribution). The symmetric Bernoulli model has already been studied thoroughly also with respect to the external path length of tries; see [20, 27]. Hence, we exclude this case subsequently. For later reference, we summarize our conditions as:

$$p_{ij} \in (0, 1) \text{ for all } i, j \in \Sigma, \qquad p_{ij} \neq \frac{1}{2} \text{ for some } (i, j) \in \Sigma^2. \tag{2}$$

The entropy rate of the Markov chain plays an important role in the asymptotic behavior of the performance of radix sort. In particular, it determines the leading order constant of the average number of bucket operations (path length) performed by radix sort. The entropy rate for our Markov chain is given by

$$H := -\sum_{i,j\in\Sigma} \pi_i\, p_{ij} \log p_{ij} = \sum_{i\in\Sigma} \pi_i H_i, \tag{3}$$

where $H_i := -\sum_{j\in\Sigma} p_{ij} \log p_{ij}$ is the entropy of a transition from state i to the next state. Thus, H is obtained as weighted average of the entropies of all possible transitions with weights according to the stationary distribution π.

Our main result concerning the distribution of the number of bucket operations in radix sort or the path length in a trie is presented next. We will write B_n^μ for B_n to make its dependence on the initial distribution explicit.

Theorem 2.1. *The number B_n^μ of bucket operations under the Markov source model with conditions* (2) *satisfies, as $n \to \infty$,*

$$\mathbb{E}\left[B_n^\mu\right] = \frac{1}{H} n \log n + \mathrm{O}(n), \qquad \mathrm{Var}\left(B_n^\mu\right) = \sigma^2 n \log n + \mathrm{O}\left(n\sqrt{\log n}\right)$$

where the entropy rate H is defined in (3) and σ^2 is given by

$$\sigma^2 = \frac{\pi_0 p_{00} p_{01}}{H^3}\left(\log(p_{00}/p_{01}) + \frac{H_1 - H_0}{p_{01}+p_{10}}\right)^2 + \frac{\pi_1 p_{10} p_{11}}{H^3}\left(\log(p_{10}/p_{11}) + \frac{H_1 - H_0}{p_{01}+p_{10}}\right)^2.$$

Moreover, as $n \to \infty$,

$$\frac{B_n^\mu - \mathbb{E}[B_n^\mu]}{\sqrt{\mathrm{Var}(B_n^\mu)}} \xrightarrow{d} \mathcal{N}(0,1)$$

where $\mathcal{N}(0,1)$ denotes a random variable with the standard normal distribution.

The analysis of B_n^μ is based on a system of recursive distributional equations discussed in the next section. Section 4 contains some moment-transfer theorems that are used in the analysis of mean and variance. These theorems are applied to the analysis of the mean in section 5 in order to derive the asymptotic expansion in Theorem 2.1 as well as a more detailed study of the remaining term $f_\mu(n) := \mathbb{E}[B_n^\mu] - n\log n/H$ which is necessary to obtain the limit law in section 7.

The first order asymptotic of $\mathrm{Var}(B_n^\mu)$ with uniform error term is derived in section 6. It is based on the moment-transfer theorems from section 4 but requires some additional ideas such as a splitting of B_n^μ into a suitable sum and a poissonization argument.

Finally, the limit theorem is establish in section 7. The proof is based on the contraction method. In fact, the asymptotic analysis of the moments enables us to apply this technique. It is possible to obtain a more detailed asymptotic expansion of the mean by analytical techniques however, without the analysis of the increments in proposition 5.2 the analysis in section 7 would require an asymptotic expansion up to the order of $o(\sqrt{n\log n})$. It should be pointed out that analytic techniques allows asymptotics of the mean and the variance up to $o(n)$ [36].

3 Recursive Distributional Equations

We formulate in this section a system of distributional recurrences to capture the distribution of the number of bucket operations. Our subsequent analysis is entirely based on these equations. In the sequel, we phrase our discussion in terms of the radix sort algorithm.

We denote by B_n^μ the number of bucket operations (i.e., number of bits inspected by radix sort) performed sorting n data under the Markov source model with initial distribution μ using the radix sorting algorithm. We have $B_0^\mu = B_1^\mu = 0$ for all initial distributions μ. The transition matrix is given in advance and suppressed in the notation. We abbreviate $B_n^i := B_n^{\rho_i}$ for $i \in \Sigma$ and $\rho_i = p_{i0}\delta_0 + p_{i1}\delta_1$. We will study B_n^0 and B_n^1. From the asymptotic behavior of these two sequences we can then directly obtain corresponding results for B_n^μ for an arbitrary initial distribution $\mu = \mu_0\delta_0 + \mu_1\delta_1$ as follows: We denote by K_n the number of data among our n that start with bit 0. Then K_n has the binomial $B(n, \mu_0)$ distribution. In the Markov source model the distribution of the second bit of every data string that starts with bit 0 is ρ_0. In particular, for any data string $\Xi = \xi_1\xi_2\ldots$ in the left bucket (i.e. $\xi_1 = 0$) the remaining suffix $\xi_2\xi_2\ldots$ is generated by a Markov source model with initial distribution ρ_0 and the same transition matrix as the original source. Similarly, the remaining suffixes in the right bucket are generated by a Markov source model with initial distribution ρ_1 and the same transition matrix. Moreover, by the independence of data strings within the Markov source model, the number of bucket operations in the left bucket and the number of bucket operations in the right bucket are independent conditionally on K_n. This leads to the following stochastic recurrence:

$$B_n^{\mu} \stackrel{d}{=} B^0_{K_n} + B^1_{n-K_n} + n, \qquad n \geq 2, \tag{4}$$

where $(B_0^0, \ldots, B_n^0)$, $(B_0^1, \ldots, B_n^1)$ and K_n are independent and $\stackrel{d}{=}$ denotes that left and right hand side have identical distributions. We will see later that we can directly transfer asymptotic results for B_n^0 and B_n^1 to general B_n^{μ} via (4), see, e.g., the proof of Theorem 7.1.

In particular, (4) implies for $\mu = \rho_0$ that

$$B_n^0 \stackrel{d}{=} B^0_{I_n} + B^1_{n-I_n} + n, \qquad n \geq 2, \tag{5}$$

with $(B_0^0, \ldots, B_n^0)$, $(B_0^1, \ldots, B_n^1)$ and I_n independent binomially $B(n, p_{00})$ distributed. A similar argument yields a recurrence for B_n^1. Denoting by J_n a binomially $B(n, p_{10})$ distributed random variable, we have

$$B_n^1 \stackrel{d}{=} B^0_{J_n} + B^1_{n-J_n} + n, \qquad n \geq 2, \tag{6}$$

with $(B_0^0, \ldots, B_n^0)$, $(B_0^1, \ldots, B_n^1)$ and J_n independent. Our asymptotic analysis of B_n^{μ} is based on the distributional recurrence system (5)–(6) as well as (4).

For further references, we abbreviate (5) and (6) by

$$B_n^i \stackrel{d}{=} B^0_{I_n^i} + B^1_{n-I_n^i} + n, \qquad n \geq 2,\ i \in \Sigma, \tag{7}$$

with $(B_0^0, \ldots, B_n^0)$, $(B_0^1, \ldots, B_n^1)$ and I_n^i independent, I_n^i binomial $B(n, p_{i0})$ distributed.

4 Transfer Theorems for Mean and Variance

Throughout this section, let $(a_i(n))_{n\in\mathbb{N}_0}$ and $(\varepsilon_i(n))_{n\in\mathbb{N}_0}$ be real valued sequences for $i \in \{0,1\}$. Furthermore, let I_n^i follow the binomial distribution $B(n, p_{i0})$ for $i \in \{0,1\}$. Suppose that these sequences either satisfy

$$a_i(n) = \mathbb{E}[a_0(I_n^i)] + \mathbb{E}[a_1(n - I_n^i)] + \varepsilon_i(n), \qquad i \in \{0,1\},\ n \in \mathbb{N}, \tag{8}$$

which is the case for, e.g., $a_i(n) = \mathbb{E}[B_n^i]$ and $\varepsilon_i(n) = n\mathbf{1}_{[2,\infty)}(n)$, or satisfy

$$a_i(n) = p_{i0}\mathbb{E}[a_0(I_n^i)] + p_{i1}\mathbb{E}[a_1(n - I_n^i)] + \varepsilon_i(n), \qquad i \in \{0,1\},\ n \in \mathbb{N}, \tag{9}$$

which is the case for, e.g., $a_i(n) = f_i(n+1) - f_i(n)$ where $f_i(n) = \mathbb{E}[B_n^i] - \frac{1}{H} n \log n$ and $\varepsilon_i(n) = 1$.

Upper bounds on $\varepsilon_i(n)$ may be transferred to bounds on $a_i(n)$ by the following lemma:

Lemma 4.1. *Assume that* (8) *holds. Then,* $\varepsilon_i(n) = \mathrm{O}(n^{\alpha})$ *for an* $\alpha \in \mathbb{R}$ *and both* $i \in \{0,1\}$ *implies, as* $n \to \infty$,

$$a_i(n) = \begin{cases} \mathrm{O}(n), & \text{if } \alpha < 1, \\ \mathrm{O}(n^{\alpha}), & \text{if } \alpha > 1, \\ \mathrm{O}(n \log n), & \text{if } \alpha = 1. \end{cases}$$

More precisely, the first order asymptotic of linear $\varepsilon_i(n)$ terms yield the following first order asymptotic of $a_i(n)$:

Lemma 4.2. *Assume that* (8) *holds. Then,* $\varepsilon_i(n) = c_i n + \mathrm{O}(n^{\alpha})$ *for* $c_0, c_1 \in \mathbb{R}$ *and* $\alpha < 1$ *and both* $i \in \{0,1\}$ *implies that, as* $n \to \infty$,

$$a_i(n) = \frac{\pi_0 c_0 + \pi_1 c_1}{H} n \log n + \mathrm{O}(n)$$

with constants π_0, π_1 *and* H *given in* (1) *and* (3).

Similarly, there are the following results on transfers for (9):

Lemma 4.3. *Assume that* (9) *holds. Then,* $\varepsilon_i(n) = \mathrm{O}(n^\alpha)$ *for an* $\alpha \in \mathbb{R}$ *and both* $i \in \{0,1\}$ *implies that, as* $n \to \infty$,

$$a_i(n) = \begin{cases} \mathrm{O}(1) & \text{if } \alpha < 0, \\ \mathrm{O}(n^\alpha) & \text{if } \alpha > 0, \\ \mathrm{O}(\log n) & \text{if } \alpha = 0. \end{cases}$$

Lemma 4.4. *Assume that* (9) *holds. Then,* $\varepsilon_i(n) = c_i + \mathrm{O}(n^{-\alpha})$ *for* $c_i \in \mathbb{R}$, $\alpha > 0$ *and both* $i \in \{0,1\}$ *implies, as* $n \to \infty$,

$$a_i(n) = \frac{\pi_0 c_0 + \pi_1 c_1}{H} \log n + \mathrm{O}(1), \quad i \in \Sigma,$$

with constants π_0, π_1 *and* H *given in* (1) *and* (3).

Proof of lemma 4.1. The proof relies on the fact that I_n^0 and I_n^1 are concentrated around their means $p_{00}n$ and $p_{10}n$. This leads to a geometric decay in the size of the toll term when iterating (8) on the right hand side. It is more convenient to work with the monotone sequences given by

$$C_i(n) := \sup\{|a_i(k)| : 0 \le k \le n\}, \qquad C(n) := \max\{C_0(n), C_1(n)\}, \quad i \in \Sigma, n \in \mathbb{N}_0.$$

Due to the upper bound $|a_i(n)| \le C(n)$ for both $i \in \{0,1\}$, an upper bound on $C(n)$ is sufficient to prove the assertion. To this end, let $\max_{i,j\in\{0,1\}}\{p_{ij}\} < \delta < 1$ be a constant (the exact value of δ does not matter) and decompose (8) into

$$|a_i(n)| \le \mathbb{E}[(C(I_n^i) + C(n - I_n^i))\mathbf{1}_{\{I_n^i \in [(1-\delta)n, \delta n]\}}] + C(n)\mathbb{P}(I_n^i \notin [(1-\delta)n, \delta n]) + |\varepsilon_i(n)|. \quad (10)$$

Note that at least one of the following three equalities needs to hold by definition:

$$C(n) = |a_0(n)| \quad \text{or} \quad C(n) = |a_1(n)| \quad \text{or} \quad C(n) = C(n-1).$$

Thus, the assumption on $\varepsilon_i(n)$ implies that there exists a constant $L > 0$ such that *at least one* of the following two bounds holds

$$\begin{aligned} \beta(n)C(n) &\le \max_{i\in\Sigma} \left\{\mathbb{E}[(C(I_n^i) + C(n - I_n^i))\mathbf{1}_{\{I_n^i \in [(1-\delta)n, \delta n]\}}] + Ln^\alpha\right\} \\ C(n) &\le C(n-1), \end{aligned} \quad (11)$$

where $\beta(n) := 1 - 2\max_{i\in\Sigma}\{\mathbb{P}(I_n^i \notin [(1-\delta)n, \delta n])\}$ converges to 1 by a Chernoff bound on the binomial distribution (or the central limit theorem). Now (11) implies for any $\varepsilon > 0$ by induction on n that

$$C(n) \le Dn^{\max\left\{-\frac{\log 4}{\log \delta}, 2\alpha\right\}}(1+\varepsilon)^n$$

where $D = D(\varepsilon) > 0$ is a sufficiently large constant. This yields for any $K > 1$ the rough upper bound $C(n) = \mathrm{O}(K^n)$.

To refine this bound, note that a standard Chernoff bound on the binomial distribution implies the existence of a constant $c > 0$ such that for all $n \ge 0$

$$|\beta(n) - 1| \le 4e^{-cn}$$

which together with $C(n) = \mathrm{O}(K^n)$ for $1 < K < e^c$ yields a constant $L' > 0$ such that

$$|\beta(n) - 1|C(n) \le L'n^\alpha, \quad n \in \mathbb{N}.$$

Combined with (10), this bound implies by induction on n that

$$C(n) \le \tilde{L}n \sum_{j=0}^{\lfloor -\log n/\log \delta \rfloor} (\delta^{1-\alpha})^j$$

where $\tilde{L} = \max\left\{C(d+1), (L+L')\max\{\delta^{\alpha-1}, 1\}\right\}$. Thus, the assertion holds by the asymptotic of the geometric sum. ∎

Proof of lemma 4.2. An easy calculation reveals that the sequences

$$\tilde{a}_i(n) := a_i(n) - \frac{\pi_0 c_0 + \pi_1 c_1}{H} n \log n + \frac{c_{1-i} H_i}{(p_{10} + p_{01})H} n, \quad n \in \mathbb{N},\, i \in \{0,1\},$$

satisfy

$$\tilde{a}_i(n) = \mathbb{E}[\tilde{a}_0(I_n^i)] + \mathbb{E}[\tilde{a}_1(n - I_n^i)] + \mathrm{O}\left(n^{\max\{\alpha,1/3\}}\right).$$

Thus, lemma 4.1 yields $\widetilde{a}_i(n) = \mathrm{O}(n)$ and the assertion follows. More precisely, note that the transformed sequences satisfy for all $n \in \mathbb{N}$ and $i \in \{0,1\}$

$$\tilde{a}_i(n) = \mathbb{E}[\tilde{a}_0(I_n^i)] + \mathbb{E}[\tilde{a}_1(n - I_n^i)] + \tilde{\varepsilon}_i(n)$$

with, for $h(x) := x \log x$,

$$\begin{aligned}\tilde{\varepsilon}_i(n) &= \varepsilon_i(n) - c\left(h(n) - \mathbb{E}[h(I_n^i) + h(n - I_n^i)]\right) \\ &\quad + \frac{c_{1-i} H_i}{(p_{10} + p_{01})H} n - \frac{c_1 H_0}{(p_{10} + p_{01})H} n p_{i0} - \frac{c_0 H_1}{(p_{10} + p_{01})H} n p_{i1}.\end{aligned}$$

Thus, it only remains to show $\tilde{\varepsilon}_i(n) = \mathrm{O}\left(n^{\max\{\alpha,1/3\}}\right)$. To this end, note that

$$\begin{aligned}&h(n) - \mathbb{E}[h(I_n^i) + h(n - I_n^i)] \\ &= -\mathbb{E}[n h(I_n^i/n) + n h(1 - I_n^i/n)] \\ &= H_i n - n\mathbb{E}[h(I_n^i/n) - h(p_{i0}) + h(1 - I_n^i/n) - h(p_{i1})] \\ &= H_i n + \mathrm{O}\left(n^{1/3}\right)\end{aligned}$$

where the last equality holds by the concentration of the binomial distribution and the asymptotic of $\log(1+x)$ as $x \to 0$ (note that $\log(I_n^i/n) - \log(p_{i0}) = \log(1 + (I_n^i - np_{i0})/(np_{i0}))$). Details can be found in the appendix, equation (60). Therefore, an easy calculation yields $\tilde{\varepsilon}_i(n) = \mathrm{O}\left(n^{\max\{\alpha,1/3\}}\right)$ and the assertion follows. ☐

Proof of lemma 4.3. The idea is essentially the same as in the proof of lemma 4.1: Once again, it is more convenient to work with the monotone sequences $(C_i(n))_{n \geq 0}$ and $(C(n))_{n \geq 0}$ given by

$$C_i(n) := \sup\{|a_i(k)| : 0 \leq k \leq n\}, \qquad C(n) := \max\{C_0(n), C_1(n)\}, \quad n \in \mathbb{N}_0, i \in \Sigma.$$

With $\max_{i,j \in \{0,1\}}\{p_{ij}\} < \delta < 1$ equation (9) may be decomposed into

$$|a_i(n)| \leq \mathbb{E}[(p_{i0} C_0(I_n^i) + p_{i1} C_1(n - I_n^i))\mathbf{1}_{\{I_n^i \in [(1-\delta)n, \delta n]\}}] + C(n)\mathbb{P}(I_n^i \notin [(1-\delta)n, \delta n]) + |\varepsilon_i(n)|$$

As in the proof of 4.1 this implies $C(n) = \mathrm{O}(K^n)$ for any constant $K > 1$ and, by a standard Chernoff bound on the binomial distribution

$$|a_i(n)| \leq \mathbb{E}[(p_{i0} C_0(I_n^i) + p_{i1} C_1(n - I_n^i))\mathbf{1}_{\{I_n^i \in [(1-\delta)n, \delta n]\}}] + \mathrm{O}(n^\alpha).$$

One obtains by induction on n that

$$C(n) \leq \tilde{L} \sum_{k=0}^{\lfloor -\log n / \log \delta \rfloor} \delta^{-\alpha j}$$

and the assertion follows by the asymptotic behavior of the geometric sum. ☐

Proof of lemma 4.4. An easy calculation reveals that the sequences

$$\tilde{a}_i(n) := a_i(n) - L g(n) + \frac{c_{1-i} H_i}{(p_{01} + p_{10})H}, \quad i \in \{0,1\},\, n \in \mathbb{N}$$

with $L = (\pi_0 c_0 + \pi_1 c_1)/H$ satisfy

$$\tilde{a}_i(n) = p_{i0}\mathbb{E}[\tilde{a}_0(I_n^i)] + p_{i1}\mathbb{E}[\tilde{a}_1(n - I_n^i)] + \mathrm{O}\left(n^{-\min\{\alpha,1/2\}}\right).$$

Thus, lemma 4.3 implies the assertion. ☐

5 Analysis of the Mean

First we study the asymptotic behavior of the expected number of Bucket operations with a precise error term needed to derive a limit law in Section 7.

Theorem 5.1. *For the number B_n^μ of Bucket operations under the Markov source model with conditions (2) we have*

$$\mathbb{E}[B_n^\mu] = \frac{1}{H} n \log n + \mathrm{O}(n), \qquad (n \to \infty),$$

with the entropy rate H of the Markov chain given in (3). The $\mathrm{O}(n)$ error term is uniform in the initial distribution μ.

Our proof of Theorem 5.1 as well as the corresponding limit law in Theorem 7.1 depend on refined properties of the $\mathrm{O}(n)$ error term that are first obtained for the initial distributions $\rho_0 = p_{00}\delta_0 + p_{01}\delta_1$ and $\rho_1 = p_{10}\delta_0 + p_{11}\delta_1$ and then generalized to arbitrary initial distribution via (4). For those initial distributions we denote the error term for all $n \in \mathbb{N}_0$ and $i \in \Sigma$ by

$$f_i(n) := \mathbb{E}[B_n^i] - \frac{1}{H} n \log n. \tag{12}$$

The following Lipschitz continuity of f_0 and f_1 is crucial for our further analysis:

Proposition 5.2. *There exists a constant $C > 0$ such that for both $i \in \Sigma$ and all $m, n \in \mathbb{N}_0$*

$$|f_i(m) - f_i(n)| \leq C|m - n|.$$

In order to prove the Lipschitz continuity of the error terms f_0 and f_1 (proposition 5.2) we will analyze the increments of $(f_0(n))_{n\geq 0}$ and $(f_1(n))_{n\geq 0}$ and apply Lemma 4.4. We use the following notation for the increments:

For a sequence $x = (x(n))_{n\geq 0}$ in $\mathbb{R}$ we denote its (finite forward) difference sequence by $(\Delta x(n))_{n\geq 0}$, where

$$\Delta x(n) := (\Delta x)(n) := x(n+1) - x(n), \qquad n \in \mathbb{N}.$$

Note that the order of operation is first applying the Δ-operator to the sequence then evaluating the difference sequence at n. In particular, for any sequence $(m_n)_{n\in\mathbb{N}}$ in $\mathbb{N}_0$ we have

$$\Delta x(m_n) = x(m_n + 1) - x(m_n), \qquad n \in \mathbb{N}_0$$

(and in general $\Delta x(m_n) \neq x(m_{n+1}) - x(m_n)$).

In the analysis of $(\Delta f_i(n))_{n\geq 0}$, $i \in \Sigma$ we use the following Lemma which is a special case of Lemma 2 in Schachinger [35].

Lemma 5.3. *For any real sequence $(a(n))_{n\geq 0}$ and binomially $B(n,p)$ distributed X_n with $p \in (0,1)$ we have*

$$\Delta \mathbb{E}[a(X_n)] = p\mathbb{E}[\Delta a(X_n)], \quad n \in \mathbb{N}.$$

Proof. Note that $X_{n+1} \stackrel{d}{=} X_n + B$ in which B and X_n are independent and $\mathbb{P}(B = 1) = p = 1 - \mathbb{P}(B = 0)$. This yields

$$\Delta \mathbb{E}[a(X_n)] = \mathbb{E}[a(X_n + B) - a(X_n)] = p\mathbb{E}[\Delta a(X_n)]$$

which is the assertion. □

Proof of proposition 5.2. Note that (7) implies

$$f_i(n) = \mathbb{E}[f_0(I_n^i)] + \mathbb{E}[f_1(n - I_n^i)] + \varepsilon_i(n)$$

with the toll function

$$\varepsilon_i(n) = n - \frac{1}{H}(n \log n - \mathbb{E}[I_n^i \log I_n^i] - \mathbb{E}[(n - I_n^i)\log(n - I_n^i)].$$

Thus, lemma 5.3 yields for the increments $a_i(n) := \Delta f_i(n)$

$$a_i(n) = p_{i0}\mathbb{E}[a_0(I_n^i)] + p_{i1}\mathbb{E}[a_1(n - I_n^i)] + \Delta\varepsilon_i(n).$$

Moreover, another application of lemma 5.3 yields

$$\Delta\varepsilon_i(n) = 1 - \frac{1}{H}(\Delta h(n) - p_{i0}\mathbb{E}[\Delta h(I_n^i)] - p_{i1}\mathbb{E}[\Delta h(n - I_n^i)])$$

where $h(x) := x \log x$. Since $\Delta h(n) = \log(n+1) + n\log(1+1/n) = \log(n+1) + 1 + \mathrm{O}(1/n)$, one obtains

$$\begin{aligned}\Delta\varepsilon_i(n) &= 1 - \frac{1}{H}(\log(n+1) - p_{i0}\mathbb{E}[\log(I_n^i + 1)] - p_{i1}\mathbb{E}[\log(n - I_n^i + 1)] + \mathrm{O}(1/n) \\ &= 1 - \frac{1}{H}(-p_{i0}\log p_{i0} - p_{i1}\log p_{i1}) + \mathrm{O}(n^{-1/2}).\end{aligned}$$

The last equation is based on the fact that $\mathbb{E}[\log((I_n^i + 1)/(n+1))] = \log(p_{i0}) + \mathrm{O}(n^{-1/2})$ for any binomially $B(n, p_{i0})$ distributed I_n^i (details are given in the appendix, equation (58)). Therefore, lemma 4.4 implies $\Delta f_i(n) = L \log n + \mathrm{O}(1)$ with a constant

$$L = \frac{1}{H}\left(\pi_0\left(1 - \frac{1}{H}(-p_{00}\log p_{00} - p_{01}\log p_{01})\right) + \pi_1\left(1 - \frac{1}{H}(-p_{10}\log p_{10} - p_{11}\log p_{11})\right)\right) = 0.$$

Thus, $\Delta f_i(n)$ is bounded and the assertion follows. □

Proof of theorem 5.1. For $\mu = p_{i0}\delta_0 + p_{i1}\delta_1$, $i \in \{0,1\}$ theorem 5.1 is an immediate consequence of proposition 5.2. For the general case let $\nu_i(n) := \mathbb{E}[B_n^i]$. Then, the distributional recursion 4 yields

$$\mathbb{E}[B_n^\mu] = \mathbb{E}[\nu_0(K_n)] + \mathbb{E}[\nu_1(n - K_n)] + n.$$

Thus, $\nu_i(n) = n\log n/H + \mathrm{O}(n)$ implies

$$\mathbb{E}[B_n^\mu] = \frac{1}{H} n \log n + \frac{n}{H}\mathbb{E}[h(I_n^i/n)] + \mathbb{E}[h(1 - I_n^i/n)] + \mathrm{O}(n)$$

where $h(x) = x\log x$. Since h is uniformly bounded on $(0,1]$, the assertion follows. □

6 Analysis of the Variance

In this section we establish precise growth of the variance with a uniform bound. We prove the following theorem.

Theorem 6.1. *For the number B_n^μ of Bucket operations under the Markov source model with conditions (2) we have, as $n \to \infty$,*

$$\mathrm{Var}(B_n^\mu) = \sigma^2 n \log n + \mathrm{O}\left(n\sqrt{\log n}\right), \tag{13}$$

where $\sigma^2 > 0$ is independent of the initial distribution μ and given by

$$\sigma^2 = \frac{\pi_0 p_{00} p_{01}}{H^3}\left(\log(p_{00}/p_{01}) + \frac{H_1 - H_0}{p_{01} + p_{10}}\right)^2 + \frac{\pi_1 p_{10} p_{11}}{H^3}\left(\log(p_{10}/p_{11}) + \frac{H_1 - H_0}{p_{01} + p_{10}}\right)^2. \tag{14}$$

In order to derive the first order asymptotics of the variance without studying the mean in detail, we extend an idea of Schachinger in [35] to Markov Sources. The main ingredient is to split the number of Bucket operations into a sum of two random variables in which mean and variance of the first random variable is easy to derive and the variance of the second random variable is small (i.e. $\mathrm{O}(n)$).

Once again, for $i \in \Sigma$ and $n \in \mathbb{N}_0$ let I_n^i be a Binomial $B(n, p_{i0})$ distributed random variable. Now let $(X_n^0, Z_n^0)_{n\in\mathbb{N}_0}$, $(X_n^1, Z_n^1)_{n\in\mathbb{N}_0}$ and $(I_n^0, I_n^1)_{n\in\mathbb{N}_0}$ be independent sequences of random variables with finite second moments that satisfy the initial conditions

$$X_n^i = Z_n^i = 0, \quad i \in \Sigma, n \le 1$$

and, for all $n \ge 2$ and $i \in \Sigma$

$$\begin{pmatrix} X_n^i \\ Z_n^i \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} X_{I_n^i}^0 \\ Z_{I_n^i}^0 \end{pmatrix} + \begin{pmatrix} X_{n-I_n^i}^1 \\ Z_{n-I_n^i}^1 \end{pmatrix} + \begin{pmatrix} \eta_n^{i,1} \\ \eta_n^{i,2} \end{pmatrix}, \tag{15}$$

where the toll terms are given by $\eta_n^{i,1} = \eta_n^{i,2} = 0$ for $n \le 1$ and

$$\begin{aligned} \eta_n^{i,1} &:= \frac{1}{H}\left(n\log(n) - \mathbb{E}\left[I_n^i \log\left(I_n^i\right) + (n - I_n^i)\log\left(n - I_n^i\right)\right]\right) \\ &\quad + \pi_{1-i}\frac{H_{1-i} - H_i}{H} n + \frac{H_1 - H_0}{(p_{01} + p_{10})H} p_{i0} p_{i1}^{n-1} n, \qquad n \ge 2, \\ \eta_n^{i,2} &:= n - \eta_n^{i,1}. \end{aligned} \tag{16}$$

Since we have $\eta_n^{i,1} + \eta_n^{i,2} = n$, note that the sum $S_n^i := X_n^i + Z_n^i$ satisfies the same initial conditions and the same stochastic recurrence as B_n^i, i.e. equation (7) and $S_n^i = 0 = B_n^i$ for $n \le 1$. In particular, this implies that S_n^i and B_n^i have the same mean and variance. A discussion on the existence of a splitting satisfying (15) and the equality of the moments of S_n^i and B_n^i is given in section 6.1.

Remark. The choice of $\eta_n^{i,1}$ is motivated as follows: Since Z_n^i should be small ($\mathbb{E}[Z_n^i] = \mathrm{O}(n)$, $\mathrm{Var}(Z_n^i) = \mathrm{O}(n)$), X_n^i should satisfy $\mathbb{E}[X_n^i] \sim \frac{1}{H} n\log(n)$ which is the reason for the choice of the first summand in (16). The linear term is chosen to obtain $\eta_n^{i,1} \sim n$ and therefore $\eta_n^{i,2} = o(n)$ which implies a small variance for Z_n^i. The last summand is chosen for some technical reasons to compensate the second one in the calculation of $\mathbb{E}[X_n^i]$.

The proof of theorem 6.1 works as follows: first we study the asymptotics of $\mathrm{Var}(X_n^i)$ and $\mathrm{Var}(Z_n^i)$ and then deduce the asymptotics of $\mathrm{Var}(B_n^i)$ by the following Lemma:

Lemma 6.2. *For any random variables X, Y with finite second moments we have*

$$\left(\sqrt{\mathrm{Var}(X)} - \sqrt{\mathrm{Var}(Y)}\right)^2 \le \mathrm{Var}(X+Y) \le \left(\sqrt{\mathrm{Var}(X)} + \sqrt{\mathrm{Var}(Y)}\right)^2. \tag{17}$$

In particular, if sequences $(X_n)_{n\ge0}, (Y_n)_{n\ge0}$ with finite second moments satisfy $\mathrm{Var}(Y_n) = o(\mathrm{Var}(X_n))$ *then we have*

$$\mathrm{Var}(X_n + Y_n) = \mathrm{Var}(X_n) + \mathrm{O}\left(\sqrt{\mathrm{Var}(X_n)\mathrm{Var}(Y_n)}\right). \tag{18}$$

Proof. By the Cauchy-Schwarz inequality we have

$$|\mathrm{Cov}(X,Y)| \le \sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}$$

which together with $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)$ implies (17). Moreover, (17) obviously implies (18). □

The analysis of $\mathrm{Var}(X_n^i)$ is done with lemma 4.2. This requires a detailed asymptotic expansion of $\mathbb{E}[X_n^i]$. The choice of $\eta_n^{i,1}$ leads to the following representation of the mean:

Lemma 6.3. *Let $(X_n^i)_{n\in\mathbb{N}_0, i\in\Sigma}$ be as in (15). Then we have for all $n \in \mathbb{N}_0$*

$$\mathbb{E}[X_n^0] = \frac{1}{H} n \log n + \frac{H_1 - H_0}{(p_{01}+p_{10})H} n \mathbf{1}_{\{n\geq 2\}}, \qquad \mathbb{E}[X_n^1] = \frac{1}{H} n \log n. \tag{19}$$

Proof. Let $\nu_X^i : \mathbb{N}_0 \to \mathbb{R}$ be given by $\nu_X^i(n) := \mathbb{E}[X_n^i]$, $i \in \{0,1\}$. Note that ν_X^i is uniquely determined by its initial conditions $\nu_X^i(n) = 0$ for $n \leq 1$ and the recursion

$$\nu_X^i(n) = \mathbb{E}[\nu_X^0(I_n^i)] + \mathbb{E}[\nu_X^1(n - I_n^i)] + \eta_n^{i,1}, \quad i \in \Sigma,\, n \geq 2,$$

which arises from the recursion (15). Thus, it only remains to check that the choice given in (19) satisfies these conditions which is an easy calculation. Details are left to the reader. □

These expressions and lemma 4.2 lead to the following asymptotics of $\mathrm{Var}(X_n^i)$:

Lemma 6.4. *We have for both $i \in \Sigma$ as $n \to \infty$*

$$\mathrm{Var}\left(X_n^i\right) = \sigma^2 n \log n + \mathrm{O}(n)$$

where σ^2 is given by (14).

Proof. Let $V_X^i(n) := \mathrm{Var}(X_n^i)$ and $\nu_X^i(n) := \mathbb{E}[X_n^i]$ as in the previous proof. Then, the recursion (15) and the independence therein imply

$$V_X^i(n) = \mathbb{E}[V_X^0(I_n^i)] + \mathbb{E}[V_X^1(n - I_n^i)] + \mathrm{Var}(\nu_X^0(I_n^i) + \nu_X^1(n - I_n^i)). \tag{20}$$

It suffices to derive the first order asymptotic of $\mathrm{Var}(\nu_X^0(I_n^i) + \nu_X^1(n - I_n^i))$ to apply lemma 4.2. To this end, note that by lemma 6.3 with the notation $h(x) := x \log x$

$$\mathrm{Var}(\nu_X^0(I_n^i) + \nu_X^1(n - I_n^i)) = \mathrm{Var}\left(\frac{h(I_n^i) + h(n - I_n^i)}{H} + \frac{H_1 - H_0}{(p_{01}+p_{10})H} I_n^i + R_n^i\right) \tag{21}$$

where $R_n^i = -\frac{H_1-H_0}{(p_{01}+p_{10})H}\mathbf{1}_{\{I_n^i=1\}}$ and thus, $\mathrm{Var}(R_n^i) = o(1)$. Subtracting $n\log n$ in the variance on the right hand side of (21) yields

$$\mathrm{Var}(\nu_X^0(I_n^i) + \nu_X^1(n - I_n^i)) = \mathrm{Var}\left(\frac{1}{H}(I_n^i \log p_{i0} + (n - I_n^i)\log p_{i1}) + \frac{H_1 - H_0}{(p_{01}+p_{10})H} I_n^i + \widetilde{R}_n^i\right)$$

where $\widetilde{R}_n^i = R_n^i + \frac{1}{H}(I_n^i(\log(I_n^i/n) - \log p_{i0}) + (n - I_n^i)(\log(1 - I_n^i/n) - \log(p_{i1})))$. It is not hard to check that $\mathrm{Var}(\widetilde{R}_n^i) = \mathrm{O}(\log n)$, as formally proved below. Therefore, combined with lemma 6.2 and $\mathrm{Var}(I_n^i) = p_{i0}p_{i1}n$

$$\mathrm{Var}(\nu_X^0(I_n^i) + \nu_X^1(n - I_n^i)) = \left(\frac{1}{H}\left(\log p_{i0} - \log p_{i1} + \frac{H_1 - H_0}{(p_{01}+p_{10})}\right)\right)^2 p_{i0}p_{i1}n + \mathrm{O}(n^{2/3}).$$

Hence, the assertion follows by (20) and lemma 4.2.

To complete the proof we now establish that $\mathrm{Var}(\widetilde{R}_n^i) = O(\log n)$. Note that the function

$$\phi : [0,1] \to \mathbb{R}, \quad x \to x(\log x - \log p_{i0}) + (1-x)(\log(1-x) - \log(1-p_{i0}))$$

is bounded and that the derivative is given by $\phi'(x) = \log(x/p_{i0}) - \log((1-x)/(1-p_{i0}))$. In particular, there exists a constant $C > 0$ such that for all sufficiently large n

$$|\phi'(x)| \leq C\sqrt{\frac{\log n}{n}}, \qquad x \in \left[p_{i0} - \sqrt{(\log n)/n}\,,\, p_{i0} + \sqrt{(\log n)/n}\right].$$

One obtains

$$\mathrm{Var}(\phi(I_n^i/n)\mathbf{1}_{\{|I_n^i-np_{i0}|\geq\sqrt{n\log n}\}}) = \mathrm{O}(n^{-2})$$

by the boundedness of ϕ and a standard Chernoff bound and, by the previous observations, the mean value theorem and a self centering argument (let J_n^i be an independent copy of I_n^i)

$$\begin{aligned}
&\mathrm{Var}\left(\phi(I_n^i/n)\mathbf{1}_{\{|I_n^i-np_{i0}|<\sqrt{n\log n}\}}\right)\\
&=\frac{1}{2}\mathbb{E}\left[\left(\phi(I_n^i/n)\mathbf{1}_{\{I_n^i-np_{i0}|<\sqrt{n\log n}\}}-\phi(J_n^i/n)\mathbf{1}_{\{|J_n^i-np_{i0}|<\sqrt{n\log n}\}}\right)^2\right]\\
&=\frac{C^2}{2}\frac{\log n}{n}\mathbb{E}\left[\left(I_n^i/n-J_n^i/n\right)^2\right]+\mathrm{O}\left(n^{-2}\right)=\mathrm{O}\left(\frac{\log n}{n^2}\right).
\end{aligned}$$

The bound on $\mathrm{Var}(\widetilde{R}_n^i)$ follows by lemma 6.2 since $\widetilde{R}_n^i = R_n^i + n\phi(I_n^i/n)$ and $\mathrm{Var}(R_n^i) = o(1)$. □

In order to derive the asymptotics of $\mathrm{Var}(Z_n^i)$ we start with an upper bound on $\eta_n^{i,2}$:

Lemma 6.5. *For $\eta_n^{i,2}$ defined in (16) we have for both $i \in \Sigma$, as $n \to \infty$*

$$\eta_n^{i,2} = \mathrm{O}\left(\log n\right).$$

Proof. By the definition of $\eta_n^{i,2}$ in (16) one only needs to compute the asymptotic of

$$h(n) - \mathbb{E}[h(I_n^i)] - \mathbb{E}[h(n-I_n^i)], \qquad h(n) := n\log n.$$

Since $h(n) = \mathbb{E}[I_n^i \log n] + \mathbb{E}[(n-I_n^i)\log n]$, one obtains

$$h(n) - \mathbb{E}[h(I_n^i)] - \mathbb{E}[h(n-I_n^i)] = -n(\mathbb{E}[h(I_n^i/n)] + \mathbb{E}[h(1-I_n^i/n)]) = nH_i - n\mathbb{E}[\phi(I_n^i/n)]$$

where $H_i = -p_{i0}\log p_{i0} - p_{i1}\log p_{i1}$ and $\phi(x) = x(\log x - \log p_{i0}) + (1-x)(\log(1-x) - \log(1-p_{i0})$. With the same arguments as at the end of the previous proof one obtains $|\phi(x)| = \mathrm{O}((\log n)/n)$ uniformly for $x \in [p_{i0} - \sqrt{(\log n)/n}, p_{i0} + \sqrt{(\log n)/n}$ which implies by a standard Chernoff bound on the binomial distribution that $n\mathbb{E}[\phi(I_n^i/n)] = \mathrm{O}(\log n)$. Hence, $\eta_n^{i,1} = n + \mathrm{O}(\log n)$ since $H = \pi_0 H_0 + \pi_1 H_1$ and the assertion follows since $\eta_n^{i,2} = n - \eta_n^{i,1}$. □

Note that we have the following Lipschitz-continuity of the means:

Lemma 6.6. *For $i \in \Sigma$ let $\nu_Z^i : \mathbb{N}_0 \to \mathbb{R}$ be given by*

$$\nu_Z^i(n) = \mathbb{E}[Z_n^i],$$

where $(Z_n^i)_{n\in\mathbb{N}_0, i\in\Sigma}$ satisfies (15). Then, the functions ν_Z^0 and ν_Z^1 are Lipschitz continuous, i.e. there exists a constant $C > 0$ such that for $i \in \Sigma$ and $n, m \in \mathbb{N}_0$ we have

$$|\nu_Z^i(n) - \nu_Z^i(m)| \leq C|n-m|.$$

Proof. Since we have

$$\mathbb{E}[X_n^i + Z_n^i] = \mathbb{E}[B_n^i]$$

the assertion immediately follows from proposition 5.2 and lemma 6.3. □

The next step is to show that $\mathrm{Var}(Z_i) = O(n)$ which we present in lemma 6.9 below. However, to establish it we need another key ingredient, namely poissonization. In poissonization one replaces n by a Poisson $\Pi(\lambda)$ distributed random variable N to derive asymptotics as $\lambda \to \infty$. This turns out to be easier than the original problem owing to some nice properties of the Poisson process such as independence of the splitting processes. The transfer lemma used after poissonization is the following:

Lemma 6.7. *For $i \in \Sigma$ let $f_i : \mathbb{R}^+ \to \mathbb{R}$ be some function that is bounded on $(0, a]$ for all $a > 0$. Assume that there exist constants $p_0, p_1 \in (0,1)$ such that for all $x > 0$ and $i \in \Sigma$*

$$f_i(x) = f_i(xp_i) + f_{1-i}(x(1-p_i)) + \eta_i(x) \tag{22}$$

where $\eta_i : \mathbb{R}^+ \to \mathbb{R}$ is some function.

Then, as $x \to \infty$, $\eta_i(x) = \mathrm{O}(x^{1-\alpha})$ for some $\alpha > 0$ and both $i \in \Sigma$ implies

$$f_i(x) = \mathrm{O}(x), \quad i \in \Sigma.$$

Proof. Iterating (22), by induction on n we find that for a sufficiently large constant $C > 0$ and all $n \in \mathbb{N}$

$$|f_i(x)| \le Cx \sum_{j=0}^{\lfloor -\frac{\log x}{\log p_\vee} \rfloor} p_\vee^{\alpha j}, \qquad x \in [1, p_\vee^{-n}],\, i \in \{0,1\},$$

where $p_\vee := \max\{p_0, p_1, 1-p_0, 1-p_1\}$. The assertion follows since the sum converges as $x \to \infty$. Details on the induction are left to the reader. □

The crucial part after poissonization is to transfer the asymptotics as $\lambda \to \infty$ into asymptotics of the original problem. One way of doing this is the next lemma:

Lemma 6.8. *Let $(a(n))_{n \in \mathbb{N}_0}$ be a real valued sequence. Moreover, let N_λ be Poisson distributed with mean $\lambda > 0$. Then, as $n \to \infty$, $\Delta a(n) := a(n+1) - a(n) = \mathrm{O}(\sqrt{n})$ implies*

$$|a(n) - \mathbb{E}[a(N_n)]| = \mathrm{O}\left(n\right).$$

Proof. First note that $\Delta a(n) = \mathrm{O}(\sqrt{n})$ implies that there exists a constant $C > 0$ such that for all $n, m \in \mathbb{N}_0$

$$|a(n) - a(m)| = \left| \sum_{i=m \wedge n}^{m \vee n - 1} \Delta a(i) \right| \le C\sqrt{n+m}|n-m|.$$

Hence, we have that

$$|a(n) - \mathbb{E}[a(N_n)]| \le \mathbb{E}[|a(n) - a(N_n)|] \le C\mathbb{E}[\sqrt{n+N_n}|N_n - n|]$$

which implies the assertion by the Cauchy-Schwarz inequality. □

This finally leads to the following bounds on $\mathrm{Var}(Z_n^0)$ and $\mathrm{Var}(Z_n^1)$ which we present next.

Lemma 6.9. *We have for both $i \in \Sigma$, as $n \to \infty$*

$$\mathrm{Var}(Z_n^i) = \mathrm{O}(n).$$

Proof. Let $V_Z^i(n) := \mathrm{Var}(Z_n^i)$ and $\nu_Z^i(n) := \mathbb{E}[Z_n^i]$. First note that similar arguments to the ones given in the proof of lemma 6.4 reveal that

$$V_Z^i(n) = \mathbb{E}[V_Z^0(I_n^i)] + \mathbb{E}[V_Z^1(n - I_n^i)] + \mathrm{Var}(\nu_Z^0(I_n^i) + \nu_Z^1(n - I_n^i)). \tag{23}$$

Since ν_Z^0 and ν_Z^1 are Lipschitz-continuous, we have $\mathrm{Var}(\nu_Z^0(I_n^i) + \nu_Z^1(n - I_n^i)) = \mathrm{O}(n)$ which can be proven by a self centering argument similar to the one at the end of the proof of lemma 6.4. Thus, lemma 4.1 yields the rough upper bound

$$\mathrm{Var}(Z_n^i) = \mathrm{O}(n \log n). \tag{24}$$

In order to refine this bound, let N_λ be a Poisson distributed random variable with mean $\lambda > 0$ which is independent of $\{Z_n^i, I_n^i : n \ge 0, i \in \{0,1\}\}$. Then, (15) implies for both $i \in \Sigma$

$$Z_{N_\lambda}^i \stackrel{d}{=} Z_{N_{\lambda p_{i0}}}^0 + Z_{M_{\lambda p_{i1}}}^1 + \eta_{N_\lambda}^{i,2} \tag{25}$$

where $N_{\lambda p_{i0}} := I^i_{N_\lambda}$ and $M_{\lambda p_{i1}} := N_\lambda - I^i_{N_\lambda}$. It is a well known fact, e.g. from Poisson processes, that $N_{\lambda p_{i0}}$ and $M_{\lambda p_{i1}}$ are independent and Poisson distributed with means λp_{i0} and λp_{i1}.

Note that $V^i_Z(n) = \mathrm{O}(n \log n)$ and the Lipschitz continuity of ν^i_Z imply that, as $\lambda \to \infty$

$$\mathrm{Var}(Z^i_{N_\lambda}) = \mathbb{E}[V^i_Z(N_\lambda)] + \mathrm{Var}(\nu^i_Z(N_\lambda)) = \mathrm{O}(\lambda \log \lambda), \quad i \in \Sigma, \tag{26}$$

where $\mathbb{E}[N_\lambda \log(N_\lambda)] = \mathrm{O}(\lambda \log \lambda)$ is not hard to check (details are given in the appendix, lemma 7.4). Moreover, Lemma 6.5 implies, as $\lambda \to \infty$

$$\mathrm{Var}(\eta^{i,2}_{N_\lambda}) = \mathrm{O}\left(\mathbb{E}[(\log(N_\lambda + 1))^2\right) = \mathrm{O}\left(\sqrt{\lambda}\right), \tag{27}$$

where the second bound holds since $(\log(n+1))^2 = \mathrm{O}(\sqrt{n})$ and $\mathbb{E}[\sqrt{N_\lambda}] = \mathrm{O}(\sqrt{\lambda})$ as $\lambda \to \infty$ (details are given in the appendix, lemma 7.4). Hence, (25) implies for $\widetilde{V}_i(\lambda) := \mathrm{Var}(Z^i_{N_\lambda})$

$$\begin{aligned}
\widetilde{V}_i(\lambda) &= \mathrm{Var}(Z^0_{N_{\lambda p_{i0}}} + Z^1_{M_{\lambda p_{i1}}} + \eta^{i,2}_{N_\lambda}) \\
&= \mathrm{Var}(Z^0_{N_{\lambda p_{i0}}} + Z^1_{M_{\lambda p_{i1}}}) + \mathrm{O}\left(\lambda^{3/4}\sqrt{\log \lambda}\right) \\
&= \widetilde{V}_0(\lambda p_{i0}) + \widetilde{V}_1(\lambda p_{i1}) + \mathrm{O}\left(\lambda^{3/4}\sqrt{\log \lambda)}\right).
\end{aligned} \tag{28}$$

in which the second equality holds by (26), (27) and Lemma 6.2 and the last equality holds since $Z^0_{N_{\lambda p_{i0}}}$ and $Z^1_{M_{\lambda p_{i1}}}$ are independent (which is one of the reason for poissonization).

Lemma 6.7 yields the refined upper bound

$$\widetilde{V}_i(\lambda) = \mathrm{O}(\lambda). \tag{29}$$

Finally, we need to deduce asymptotic results for $V^i_Z(n)$ out of (29). Since we have for both $i \in \Sigma$

$$\mathrm{Var}(Z^i_{N_\lambda}) = \mathbb{E}[V^i_Z(N_\lambda)] + \mathrm{Var}(\nu^i_Z(N_\lambda))$$

and, by the Lipschitz continuity of ν^i_Z that $\mathrm{Var}(\nu^i_Z(N_\lambda)) = \mathrm{O}(\lambda)$, we may conclude that, as $\lambda \to \infty$

$$\mathbb{E}[V^i_Z(N_\lambda)] = \mathrm{O}(\lambda). \tag{30}$$

In order to apply Lemma 6.8 we need to check that

$$\Delta V^i_Z(n) = \mathrm{O}(\sqrt{n}) \tag{31}$$

which may be done by the transfer theorem 4.3: First note that (23) and Lemma 5.3 imply for the differences

$$\Delta V^i_Z(n) = p_{i0}\mathbb{E}[\Delta V^0_Z(I^i_n)] + (1 - p_{i0})\mathbb{E}[\Delta V^1_Z(n - I^i_n)] + \varepsilon_i(n),$$

where ε_i is given by

$$\varepsilon_i(n) = \mathrm{Var}(\nu^0_Z(I^i_{n+1}) + \nu^1_Z(n + 1 - I^i_{n+1})) - \mathrm{Var}(\nu^0_Z(I^i_n) + \nu^1_Z(n - I^i_n)).$$

The Lipschitz-continuity of ν^i_Z yields $\mathrm{Var}(\nu^0_Z(I^i_n) + \nu^1_Z(n - I^i_n)) = \mathrm{O}(n)$. Moreover,

$$\mathrm{Var}(\nu^0_Z(I^i_{n+1}) + \nu^1_Z(n + 1 - I^i_{n+1})) = \mathrm{Var}\left(\nu^0_Z(I^i_n) + \nu^1_Z(n - I^i_n) + B\Delta\nu^0_Z(I^i_n) + (1 - B)\Delta\nu^1_Z(n - I^i_n)\right)$$

where B is independent of I^i_n and Bernoulli distributed with parameter p_{i0}. Since $\Delta\nu^0_Z$ and $\Delta\nu^1_Z$ are bounded, we may conclude by lemma 6.2 that

$$\mathrm{Var}(\nu^0_Z(I^i_{n+1}) + \nu^1_Z(n + 1 - I^i_{n+1})) = \mathrm{Var}(\nu^0_Z(I^i_n) + \nu^1_Z(n - I^i_n)) + \mathrm{O}(\sqrt{n})$$

which implies $\varepsilon_i(n) = \mathrm{O}(\sqrt{n})$ and therefore, $\Delta V^i_Z(n) = \mathrm{O}(\sqrt{n})$ by lemma 4.3. Hence, the depoissonization lemma 6.8 is applicable and the assertion follows. □

We finish the section with the proof of theorem 6.1:

Proof of theorem 6.1. Recall that for $n \in \mathbb{N}_0$, $i \in \Sigma$ we have

$$\rho_i := p_{i0}\delta_0 + p_{i1}\delta_1, \qquad B_n^i := B_n^{\rho_i}.$$

Moreover, we define for $n \in \mathbb{N}_0$ and $i \in \Sigma$

$$V_i(n) := \mathrm{Var}(B_n^i), \qquad \nu_i(n) := \mathbb{E}[B_n^i].$$

We start with the proof for the special cases $\mu = \rho_i$, $i \in \Sigma$. In these cases we have by definition of $(X_n^i, Z_n^i)_{n\geq 0, i\in\Sigma}$ that

$$V_i(n) = \mathrm{Var}(X_n^i + Z_n^i) = \sigma^2 n \log n + \mathrm{O}\left(n\sqrt{\log n}\right). \tag{32}$$

where the last equality holds by Lemma 6.4, 6.9 and 6.2.

In order to obtain the result for arbitrary initial distributions μ recall that, by (4),

$$B_n^\mu = B_{K_n^\mu}^0 + B_{n-K_n^\mu}^1 + n$$

where K_n^μ is binomial $B(n, \mu(0))$ distributed.

Hence, we have by the independence of $(B_n^0)_{n\geq 0}$, $(B_n^1)_{n\geq 0}$ and $(K_n^\mu)_{n\geq 0}$

$$\begin{aligned}\mathrm{Var}(B_n^\mu) &= \mathbb{E}[V_0(K_n^\mu)] + \mathbb{E}[V_1(n-K_n^\mu)] + \mathrm{Var}(\nu_0(K_n^\mu) + \nu_1(n-K_n^\mu)) \\ &= \sigma^2\mathbb{E}[K_n^\mu \log K_n^\mu + (n-K_n^\mu)\log(n-K_n^\mu)] + \mathrm{Var}(\nu_0(K_n^\mu) + \nu_1(n-K_n^\mu)) \\ &\quad + \mathrm{O}\left(n\sqrt{\log n}\right)\end{aligned}$$

where the second equality holds by (32). Therefore, it only remains to show that

$$\mathbb{E}[K_n^\mu \log K_n^\mu + (n-K_n^\mu)\log(n-K_n^\mu)] = n\log n + \mathrm{O}\left(n\sqrt{\log n}\right), \tag{33}$$

$$\mathrm{Var}(\nu_0(K_n^\mu) + \nu_1(n-K_n^\mu)) = \mathrm{O}\left(n\sqrt{\log n}\right). \tag{34}$$

For (33) note that $x \mapsto x\log x + (1-x)\log(1-x)$ is bounded on $[0,1]$ (with $0\log 0 := 0$). Therefore, we have

$$\begin{aligned}&\mathbb{E}[K_n^\mu \log K_n^\mu + (n-K_n^\mu)\log(n-K_n^\mu)] - n\log n \\ =& n\mathbb{E}[K_n^\mu/n \log(K_n^\mu/n) + (1-K_n^\mu/n)\log(1-K_n^\mu/n)] \\ =& \mathrm{O}(n)\end{aligned}$$

which implies (33). Note that by Proposition 5.2 we have for $i \in \Sigma$ and $n \in \mathbb{N}_0$

$$\nu_i(n) = \frac{1}{H} n\log n + f_i(n)$$

where f_0 and f_1 are Lipschitz continuous functions. Since the Lipschitz continuity implies $\mathrm{Var}(f_0(K_n^\mu) + f_1(n-K_n^\mu)) = \mathrm{O}(n)$, it only remains to show that

$$\mathrm{Var}(K_n^\mu \log K_n^\mu + (n-K_n^\mu)\log(n-K_n^\mu)) = \mathrm{O}(n),$$

which is an easy computation and essentially covered by the proof of lemma 6.4. Thus, we leave the details to the reader. $\square$

6.1 Existence of the Splitting

In the analysis of the variance we work with pairs $(X_n^i, Z_n^i)_{n\in\mathbb{N}_0}$, $i \in \Sigma$, that satisfy the initial conditions

$$X_n^i = Z_n^i = 0, \quad i \in \Sigma, n \leq 1, \tag{35}$$

as well as the stochastic recurrences

$$\begin{pmatrix} X_n^i \\ Z_n^i \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} X^0_{I_n^i} \\ Z^0_{I_n^i} \end{pmatrix} + \begin{pmatrix} X^1_{n-I_n^i} \\ Z^1_{n-I_n^i} \end{pmatrix} + \begin{pmatrix} \eta_n^{i,1} \\ \eta_n^{i,2} \end{pmatrix}, \quad n \geq 2, i \in \Sigma \tag{36}$$

where $(X_0^0, \ldots, X_n^0, Z_0^0, \ldots, Z_n^0)$, $(X_0^1, \ldots, X_n^1, Z_0^1, \ldots, Z_n^1)$ and I_n^i are independent, $\eta_n^{i,2} = n - \eta_n^{i,1}$ and $\eta_n^{i,1}$ is some constant satisfying

$$\eta_n^{i,1} = 0, \quad n \leq 1 \qquad \text{and} \qquad \eta_n^{i,1} = \mathrm{O}(n) \quad (n \to \infty).$$

We now discuss how to get $(X_n^i, Z_n^i)_{n\in\mathbb{N}_0, i\in\Sigma}$ with finite second moment that satisfy (35) and (36) as well as

$$\mathbb{E}[X_n^i + Z_n^i] = \mathbb{E}[B_n^i] \qquad \text{and} \qquad \mathrm{Var}(X_n^i + Z_n^i) = \mathrm{Var}(B_n^i), \quad n \in \mathbb{N}_0, \ i \in \Sigma. \tag{37}$$

By iterating (36) on the right hand side one expects

$$\begin{pmatrix} X_n^i \\ Z_n^i \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} \eta_n^{i,1} + \sum_{k=1}^\infty \sum_{I:=(i_1,\ldots,i_k)\in\{0,1\}^k} \eta^{i_k,1}_{J_i^I(n)} \\ \eta_n^{i,2} + \sum_{k=1}^\infty \sum_{I:=(i_1,\ldots,i_k)\in\{0,1\}^k} \eta^{i_k,2}_{J_i^I(n)} \end{pmatrix} \tag{38}$$

where $J_i^I(n)$ is some iteration of binomial distributed random variables that is generated as follows: For $n \in \mathbb{N}_0$ and $i \in \Sigma$ let $I_i(n) := \sum_{j=1}^n L_j^i$ where $(L_j^i)_{j\in\mathbb{N}}$ is a sequence of independent Bernoulli $B(p_{i0})$ distributed random variables. Moreover, for each $k \geq 1$, $i \in \Sigma$ and $I \in \{0,1\}^k$ let $(I^I_{i,0}(n), I^I_{i,1}(n))_{n\geq 0}$ be an independent copy of $(I_i(n), n - I_i(n))_{n\geq 0}$. Then we define for both $i \in \Sigma$

$$J_i^{(0)}(n) := I_i(n), \quad J_i^{(1)}(n) = n - I_i(n),$$

and, for $k \geq 2$ and $I = (i_1, \ldots, i_k) \in \{0,1\}^k$

$$J_i^I(n) := I^{(i_1,\ldots,i_{k-1})}_{i_{k-1},i_k}\left(J_i^{(i_1,\ldots,i_{k-1})}(n)\right).$$

In the context of radix sort $J_i^I(n)$ may be interpreted as the number of strings with prefix I among n i.i.d. strings generated by a Markov source.

Now let $\tau_i(n) := \min\{k \geq 1 : J_i^I(n) \leq 1 \text{ for all } I \in \{0,1\}^k\}$. Since $\eta_n^{i,1} = \eta_n^{i,2} = 0$ for $n \leq 1$ and $i \in \{0,1\}$, note that all summands for $k \geq \tau_i$ equal zero in (38). Hence, if we have $\tau_i(n) < \infty$ then the sum in (38) is finite.

We will now discuss that for every $n \in \mathbb{N}$ we have $\tau_i(n) < \infty$ almost surely and then use (38) to define (X_n^i, Z_n^i) and finally check that (36) and (37) holds. To this end note that

$$M_k^i(n) := \max\{J_i^I(n) : I \in \{0,1\}^k\}$$

is bounded by n, non-increasing in k and for $M_k(n) \geq 2$ the probability that $M_k(n)$ decreases by at least one (i.e. $M_{k+1}(n) \leq M_k(n) - 1$) is at least $(2p(1-p))^{n/2}$, $p := \max\{p_{ij} | i,j \in \Sigma\}$, which can be seen as follows: At each step k there are at most $n/2$ indices $I_1, \ldots, I_{n/2} \in \{0,1\}^k$ with $J_i^{I_j}(n) \geq 2$ since we have

$$\sum_{I\in\{0,1\}^k} J_i^I(n) = n.$$

For each of these indices $I_j = (i_1^j, \dots, i_k^j)$ the probability that the next binomial splitter decreases $\max\{J_i^{(i_1^j,\dots,i_k^j,0)}, J_i^{(i_1^j,\dots,i_k^j,1)}\}$ by at least one is at least $2p(1-p)$ since starting the underlying Bernoulli chain of $(I^{I_j}_{i_k^j,0}(m))_{m\geq 0}$ with 01 or 10 causes a decrease. By the independence of $(I^{I_1}_{i_k^1,0}(m))_{m\geq 0}, \dots (I^{I_{n/2}}_{i_k^{n/2},0}(m))_{m\geq 0}$ we obtain the upper bound $(2p(1-p))^{n/2}$.

This yields that $\tau_i(n)$ is stochastically dominated by a negative binomial $nB(n,(2p(1-p))^{n/2})$ distributed random variable. In particular, we have for all $n \in \mathbb{N}$

$$\mathbb{E}[\tau_i(n)] \leq \frac{n}{(2p(1-p))^{n/2}} < \infty \qquad \text{and} \qquad \mathrm{Var}(\tau_i) < \infty.$$

This implies that mean and variance of X_n^i and Z_n^i defined by (38) are finite since we have $|\eta_n^{i,1}| \leq Cn$ for some constant $C > 0$ which together with $\sum_{I\in\{0,1\}^k} J_i^I(n) = n$ yields

$$\mathbb{E}[|X_n^i|] \leq |\eta_n^{i,1}| + Cn\mathbb{E}[|\tau_i(n)|] < \infty, \qquad \mathrm{Var}(X_n^i) \leq \mathbb{E}[(|\eta_n^{i,1}| + Cn\tau_i(n))^2] < \infty$$

and similar bounds for Z_n^i since $\eta_n^{i,2} = \mathrm{O}(n)$.

Hence, it only remains to show that the definition (38) implies (36) and (37). But (36) holds by construction and is not hard to check. For (37) note that (35) and (36) implies for the sum $S_n^i := X_n^i + Z_n^i$ in the case $d = 0$ that for both $i \in \Sigma$

$$S_n^i = 0, \quad n \leq 1 \qquad \text{and} \qquad S_n^i \stackrel{d}{=} S^0_{I_n^i} + S^1_{n-I_n^i} + n$$

which uniquely defines all moments of S_n^i that are finite. Since B_n^i satisfies the same conditions we obtain

$$\mathbb{E}[S_n^i] = \mathbb{E}[B_n^i] \qquad \text{and} \qquad \mathrm{Var}(S_n^i) = \mathrm{Var}(B_n^i).$$

7 Asymptotic Normality

Our main result is the asymptotic normality of the number of bucket operations:

Theorem 7.1. *For the number B_n^μ of bucket operations under the Markov source model with conditions (2) we have*

$$\frac{B_n^\mu - \mathbb{E}[B_n^\mu]}{\sqrt{n \log n}} \xrightarrow{d} \mathcal{N}(0, \sigma^2), \qquad (n \to \infty), \tag{39}$$

where $\sigma^2 > 0$ is independent of the initial distribution μ and given by (14).

As in the analysis of the mean, we first derive limit laws for B_n^0 and B_n^1 and then transfer these to a limit law for B_n^μ via (4). We abbreviate for $i \in \Sigma$ and $n \in \mathbb{N}_0$

$$\nu_i(n) := \mathbb{E}[B_n^i], \qquad \sigma_i(n) := \sqrt{\mathrm{Var}(B_n^i)}.$$

Note that we have $\nu_i(0) = \nu_i(1) = \sigma_i(0) = \sigma_i(1) = 0$ and $\sigma_i(n) > 0$ for all $n \geq 2$. We define the standardized variables by

$$Y_n^i := \frac{B_n^i - \mathbb{E}[B_n^i]}{\sigma_i(n)}, \qquad i \in \Sigma, n \geq 2, \tag{40}$$

and $Y_0^i := Y_1^i := 0$.

Our proof if based on an application of the contraction method to the recursive distributional system (5)–(6). The Zolotarev metric used here has been studied in the context of the contraction method systematically in [27]. We only need the following properties, see Zolotarev [38, 39]: For distributions $\mathcal{L}(X)$, $\mathcal{L}(Y)$ on $\mathbb{R}$ the Zolotarev distance ζ_s, $s > 0$, is defined by

$$\zeta_s(X,Y) := \zeta_s(\mathcal{L}(X),\mathcal{L}(Y)) := \sup_{f\in\mathcal{F}_s} |\mathbb{E}[f(X)-f(Y)]| \tag{41}$$

where $s = m+\alpha$ with $0<\alpha\le 1$, $m\in\mathbb{N}_0$, and

$$\mathcal{F}_s := \{f\in C^m(\mathbb{R},\mathbb{R}) : \|f^{(m)}(x)-f^{(m)}(y)\| \le \|x-y\|^\alpha\}, \tag{42}$$

the space of m times continuously differentiable functions from $\mathbb{R}$ to $\mathbb{R}$ such that the m-th derivative is Hölder continuous of order α with Hölder-constant 1. We have that $\zeta_s(X,Y)<\infty$, if all moments of orders $1,\dots,m$ of X and Y are equal and if the s-th absolute moments of X and Y are finite. Since later on only the case $2<s\le 3$ is used, for finiteness of $\zeta_s(X,Y)$ it is thus sufficient for these s that mean and variance of X and Y coincide and both have a finite absolute moment of order s.

Properties of ζ_s: (1) Convergence in ζ_s implies weak convergence on $\mathbb{R}$.
(2) ζ_s is $(s,+)$ ideal, i.e., we have

$$\zeta_s(X+Z,Y+Z)\le\zeta_s(X,Y), \quad \zeta_s(cX,cY)=c^s\zeta_s(X,Y)$$

for all Z being independent of (X,Y) and all $c>0$.

We will use an upper bound of ζ_s by the minimal L_p metric ℓ_p. For distributions $\mathcal{L}(X)$, $\mathcal{L}(Y)$ on $\mathbb{R}$ and $p>0$ we have

$$\ell_p(X,Y) := \ell_p(\mathcal{L}(X),\mathcal{L}(Y)) := \inf\left\{\|X'-Y'\|_p \;:\; X'\stackrel{d}{=}X, Y'\stackrel{d}{=}Y\right\},$$

where $\|X\|_p := (\mathbb{E}\|X\|^p)^{(1/p)\wedge 1}$ denotes the L_p norm. We have $\ell_p(X,Y)<\infty$ if $\|X\|_p,\|Y\|_p<\infty$. The bound used later for $2<s\le 3$ is, see Lemma 5.7 in [8],

$$\zeta_s(X,Y)\le\left((\mathbb{E}\|X\|^s)^{1-1/s}+(\mathbb{E}\|Y\|^s)^{1-1/s}\right)\ell_s(X,Y), \tag{43}$$

for all X and Y with joint mean and variance and finite absolute moments of order s.

Proposition 7.2. *For both sequences $(Y_n^i)_{n\ge0}$, $i\in\Sigma$, we have for all $2<s\le3$*

$$\zeta_s\left(Y_n^i,\mathcal{N}(0,1)\right)\to 0, \qquad (n\to\infty). \tag{44}$$

Proof. ¿From the recurrences (7) and the normalization (40) we obtain for $i\in\Sigma$

$$Y_n^i \stackrel{d}{=} \frac{\sigma_0(I_n^i)}{\sigma_i(n)}Y^0_{I_n^i} + \frac{\sigma_1(n-I_n^i)}{\sigma_i(n)}Y^1_{n-I_n^i} + b_i(n), \qquad n\ge 2, \tag{45}$$

where

$$b_i(n) = \frac{1}{\sigma_i(n)}\left(n+\nu_0(I_n^i)+\nu_1(n-I_n^i)-\nu_i(n)\right),$$

and in (45) we have that $(Y_0^0,\dots,Y_n^0)$, $(Y_0^1,\dots,Y_n^1)$ and (I_n^0,I_n^1) are independent.

For independent normal $\mathcal{N}(0,1)$ distributed random variables $\mathcal{N}_0,\mathcal{N}_1$ also independent of (I_n^0,I_n^1) we define

$$Q_n^i := \frac{\sigma_0(I_n^i)}{\sigma_i(n)}\mathcal{N}_0 + \frac{\sigma_1(n-I_n^i)}{\sigma_i(n)}\mathcal{N}_1 + b_i(n), \qquad n\ge 2. \tag{46}$$

Note that we have $\mathbb{E}[Q_n^i]=0$ and $\mathrm{Var}(Q_n^i)=1$ for all $n\ge2$. For the variance, this is seen by conditioning on I_n^i in (45) and (46) and using that Y_j^i and $\mathcal{N}_i$ have the same variance 1 for all $j\ge2$ and that for $j\in\{0,1\}$ the coefficients $\sigma_0(j)/\sigma_i(n)$ are zero, whereas for $j\in\{n-1,n\}$ the coefficients $\sigma_1(n-j)/\sigma_i(n)$ are zero. Hence, the Zolotarev distances $\zeta_s(Y_n^i,Q_n^i)$, $\zeta_s(Q_n^i,\mathcal{N}_i)$ and $\zeta_s(Y_n^i,\mathcal{N}_i)$ are finite for all $n\ge2$ and $i\in\Sigma$, where we have $2<s\le3$.

We denote by $\mathcal{N}$ another normal $\mathcal{N}(0,1)$ distributed random variable. Then we have

$$\zeta_s(Y_n^i, \mathcal{N}) \leq \zeta_s(Y_n^i, Q_n^i) + \zeta_s(Q_n^i, \mathcal{N}).$$

In the first step we show that $\zeta_s(Q_n^i, \mathcal{N}) \to 0$ as $n \to \infty$ for both $i \in \Sigma$. Note that $\|Q_n^i\|_s$ is uniformly bounded in $n \geq 2$ and $i \in \Sigma$. Hence, by (43) there exists a constant $C > 0$ such that $\zeta_s(Q_n^i, \mathcal{N}) \leq C\ell_s(Q_n^i, \mathcal{N})$. Thus, it is sufficient to show $\ell_s(Q_n^i, \mathcal{N}) \to 0$. With

$$\mathcal{N} \stackrel{d}{=} \sqrt{p_{i0}}\mathcal{N}_0 + \sqrt{1 - p_{i0}}\mathcal{N}_1$$

we obtain

$$\ell_s(Q_n^i, \mathcal{N}) \leq \left\| \left(\frac{\sigma_0(I_n^i)}{\sigma_i(n)} - \sqrt{p_{i0}} \right) \mathcal{N}_0 \right\|_s + \left\| \left(\frac{\sigma_1(n - I_n^i)}{\sigma_i(n)} - \sqrt{1 - p_{i0}} \right) \mathcal{N}_1 \right\|_s + \|b_i(n)\|_s. \qquad (47)$$

For the first summand in (47) we have, by the strong law of large numbers and the variance expansion (13) that $\sigma_0(I_n^i)/\sigma_i(n) \to \sqrt{p_{i0}}$ almost surely. Since $\mathcal{N}_0$ is independent from I_n^i and $\|\mathcal{N}_0\|_s < \infty$ we obtain from dominated convergence that this first summand tends to zero. By similar arguments we also have that the second summand in (47) tends to zero. The third summand $\|b_i(n)\|_s$ is bounded as follows: With the notation (12) and $h(x) = x \log x$ as in Lemma 7.3 of the Appendix, we have

$$\begin{aligned} b_i(n) = \frac{1}{\sigma_i(n)} \Big(\frac{n}{H} \left\{ h(I_n^i/n) - \mathbb{E}[h(I_n^i/n)] + h((n - I_n^i)/n) - \mathbb{E}[h((n - I_n^i)/n)] \right\} \\ + f_0(I_n^i) - \mathbb{E}[f_0(I_n^i)] + f_1(n - I_n^i) - \mathbb{E}[f_1(n - I_n^i)] \Big) \end{aligned}$$

With $\sigma_i(n) = \Omega(\sqrt{n \log n})$ and (59) the contributions of all summands involving the function h are $\mathrm{O}(1/\sqrt{\log n})$ in the L_s-norm, hence we have

$$\begin{aligned} \|b_i(n)\|_s \leq \|f_0(I_n^i) - \mathbb{E}[f_0(I_n^i)]\|_s + \|f_1(n - I_n^i) - \mathbb{E}[f_1(n - I_n^i)]\|_s \\ + \mathrm{O}(1/\sqrt{\log n}), \quad (n \to \infty). \end{aligned}$$

Furthermore, to bound $\|f_0(I_n^i) - \mathbb{E}[f_0(I_n^i)]\|_s$ we use an independent copy H_n^i of I_n^i. Then, by Jensen's inequality for conditional expectations and the Lipschitz property of f_i in Proposition 5.2 (with Lipschitz constant bounded by C)

$$\begin{aligned} \|f_0(I_n^i) - \mathbb{E}[f_0(I_n^i)]\|_s &= \|\mathbb{E}[f_0(I_n^i) - f_0(H_n^i) \,|\, I_n^i]\|_s \\ &\leq \|f_0(I_n^i) - f_0(H_n^i)\|_s \\ &\leq C\|I_n^i - H_n^i\|_s \\ &\leq 2C\|I_n^i - \mathbb{E}[I_n^i]\|_s \\ &= \mathrm{O}(\sqrt{n}). \end{aligned} \qquad (48)$$

Since $\|f_1(n - I_n^i) - \mathbb{E}[f_1(n - I_n^i)]\|_s$ is bounded analogously and $\sigma_i(n) = \Omega(\sqrt{n \log n})$ we obtain altogether as $n \to \infty$ and for both $i \in \Sigma$.

$$\|b_i(n)\|_s = \mathrm{O}\left(\frac{1}{\sqrt{\log n}} \right).$$

This completes the estimate for the first step $\zeta_s(Q_n^i, \mathcal{N}) \to 0$ as $n \to \infty$.

Now, we denote the distances $d_i(n) := \zeta_s(Y_n^i, \mathcal{N})$, for $n \geq 2$, and $d_i(0) := d_i(1) := 0$ for $i \in \Sigma$.

Conditioning on I_n^i and using that ζ_s is $(s,+)$ ideal we obtain for all $n \geq 2$

$$
\begin{aligned}
&d_i(n)\\
&\leq \zeta_s(Y_n^i, Q_n^i) + o(1)\\
&= \zeta_s\left(\frac{\sigma_0(I_n^i)}{\sigma_i(n)} Y^0_{I_n^i} + \frac{\sigma_1(n-I_n^i)}{\sigma_i(n)} Y^1_{n-I_n^i} + b_i(n), \frac{\sigma_0(I_n^i)}{\sigma_i(n)}\mathcal{N}_0 + \frac{\sigma_1(n-I_n^i)}{\sigma_i(n)}\mathcal{N}_1 + b_i(n)\right) + o(1)\\
&\leq \sum_{j=0}^n \binom{n}{j} p_{i0}^j (1-p_{i0})^{n-j} \zeta_s\left(\frac{\sigma_0(j)}{\sigma_i(n)} Y^0_j + \frac{\sigma_1(n-j)}{\sigma_i(n)} Y^1_{n-j} + \mathbf{1}_{\{I_n^i = j\}} b_i(n),\right.\\
&\qquad\qquad \left.\frac{\sigma_0(j)}{\sigma_i(n)}\mathcal{N}_0 + \frac{\sigma_1(n-j)}{\sigma_i(n)}\mathcal{N}_1 + \mathbf{1}_{\{I_n^i = j\}} b_i(n)\right) + o(1)\\
&\leq \sum_{j=2}^n \binom{n}{j} p_{i0}^j (1-p_{i0})^{n-j} \left\{\left(\frac{\sigma_0(j)}{\sigma_i(n)}\right)^s \zeta_s(Y_j^0, \mathcal{N}_0) + \left(\frac{\sigma_1(n-j)}{\sigma_i(n)}\right)^s \zeta_s(Y^1_{n-j}, \mathcal{N}_1)\right\} + o(1)\\
&= \mathbb{E}\left[\left(\frac{\sigma_0(I_n^i)}{\sigma_i(n)}\right)^s d_0(I_n^i) + \left(\frac{\sigma_1(n-I_n^i)}{\sigma_i(n)}\right)^s d_1(n-I_n^i)\right] + o(1). \qquad (49)
\end{aligned}
$$

With $d(n) := d_0(n) \vee d_1(n)$ we obtain for both $i \in \Sigma$ that

$$
\begin{aligned}
d_i(n) \leq\ & \mathbb{E}\left[\mathbf{1}_{\{1 \leq I_n^i \leq n-1\}} \left\{\left(\frac{\sigma_0(I_n^i)}{\sigma_i(n)}\right)^s + \left(\frac{\sigma_1(n-I_n^i)}{\sigma_i(n)}\right)^s\right\}\right] \sup_{1 \leq j \leq n-1} d(j) \qquad (50)\\
&+ ((1-p_{i0})^n + p_{i0}^n) d(n) + o(1).
\end{aligned}
$$

With

$$
\begin{aligned}
\xi(n) &:= \max_{i \in \Sigma} \mathbb{E}\left[\mathbf{1}_{\{1 \leq I_n^i \leq n-1\}} \left\{\left(\frac{\sigma_0(I_n^i)}{\sigma_i(n)}\right)^s + \left(\frac{\sigma_1(n-I_n^i)}{\sigma_i(n)}\right)^s\right\}\right],\\
\varepsilon(n) &:= \max_{i \in \Sigma} \{(1-p_{i0})^n + p_{i0}^n\}
\end{aligned}
$$

we obtain by taking the maximum of the right hand sides in (50)

$$
d(n) \leq \frac{\xi(n)}{1-\varepsilon(n)} \sup_{1 \leq j \leq n-1} d(j) + o(1). \qquad (51)
$$

We have $\varepsilon(n) \to 0$ and, since $s > 2$ and $p_{ii} \in (0,1)$ for both $i \in \Sigma$,

$$
\xi := \lim_{n \to \infty} \xi(n) = \max_{i \in \Sigma} \left\{p_{i0}^{s/2} + (1-p_{i0})^{s/2}\right\} < 1. \qquad (52)
$$

With (51) this implies that $(d(n))_{n \geq 1}$ remains bounded. We denote $\varrho := \sup_{n \geq 0} d(n)$ and $\eta := \limsup_{n \to \infty} d(n)$. Hence, we have $\varrho, \eta < \infty$ and for any $\varepsilon > 0$ there exists an $n_0 \geq 2$ such that for all $n \geq n_0$ we have $d(n) \leq \eta + \varepsilon$. From (49) we obtain with (52) for both $i \in \Sigma$

$$
\begin{aligned}
d_i(n) &\leq \mathbb{E}\left[\mathbf{1}_{\{I_n^i < n_0\} \cup \{I_n^i > n - n_0\}} \left\{\left(\frac{\sigma_0(I_n^i)}{\sigma_i(n)}\right)^s + \left(\frac{\sigma_1(n-I_n^i)}{\sigma_i(n)}\right)^s\right\}\right] \varrho \qquad (53)\\
&\quad + \mathbb{E}\left[\mathbf{1}_{\{n_0 \leq I_n^i \leq n - n_0\}} \left\{\left(\frac{\sigma_0(I_n^i)}{\sigma_i(n)}\right)^s + \left(\frac{\sigma_1(n-I_n^i)}{\sigma_i(n)}\right)^s\right\}\right] (\eta + \varepsilon) + o(1) \qquad (54)\\
&\leq (\xi + o(1))(\eta + \varepsilon) + o(1) \qquad (55)
\end{aligned}
$$

with appropriate $o(1)$ terms. Maximizing over $i \in \Sigma$ this yields $d(n) \leq o(1) + (\xi + o(1))(\eta + \varepsilon)$ and with $n \to \infty$

$$
\eta \leq \xi(\eta + \varepsilon).
$$

Since $\varepsilon > 0$ can be chosen arbitrarily small we obtain $\eta = 0$, i.e. $\zeta_s(Y_n^i, \mathcal{N}) \to 0$ as $n \to \infty$ for both $i \in \Sigma$. □

Proof of Theorem 7.1. We write

$$\frac{B_n^\mu - \mathbb{E}[B_n^\mu]}{\sqrt{n\log n}} \stackrel{d}{=} \frac{B^0_{K_n} - \nu_0(K_n) + B^1_{n-K_n} - \nu_1(n-K_n)}{\sqrt{n\log n}} + \frac{\nu_0(K_n) + \nu_1(n-K_n) + n - \mathbb{E}[B_n^\mu]}{\sqrt{n\log n}}.$$

By the Lemma of Slutzky it is sufficient to show, as $n\to\infty$,

$$\frac{B^0_{K_n} - \nu_0(K_n) + B^1_{n-K_n} - \nu_1(n-K_n)}{\sqrt{n\log n}} \xrightarrow{d} \mathcal{N}(0,\sigma^2) \tag{56}$$

$$\frac{\nu_0(K_n) + \nu_1(n-K_n) + n - \mathbb{E}[B_n^\mu]}{\sqrt{n\log n}} \xrightarrow{\mathbb{P}} 0. \tag{57}$$

For showing (56) note that by Proposition 7.2 $(B_n^i - \mathbb{E}[B_n^i])/\sqrt{n\log n} \to \mathcal{N}(0,\sigma^2)$ in distribution for both $i\in\Sigma$. We set $A_n := [\mu_0 n - n^{2/3}, \mu_0 n + n^{2/3}] \cap \mathbb{N}_0$ and $A_n^c := \{0,\dots,n\}\setminus A_n$. Then by Chernoff's bound (or the central limit theorem) we have $\mathbb{P}(K_n\in A_n)\to 1$. For all $x\in\mathbb{R}$ we have

$$\begin{aligned}&\mathbb{P}\left(\frac{B^0_{K_n} - \nu_0(K_n) + B^1_{n-K_n} - \nu_1(n-K_n)}{\sqrt{n\log n}} \le x\right)\\ &= o(1) + \sum_{j\in A_n} \mathbb{P}(K_n=j)\mathbb{P}\left(\frac{B^0_j - \nu_0(j)}{\sqrt{n\log n}} + \frac{B^1_{n-j} - \nu_1(n-j)}{\sqrt{n\log n}} \le x\right).\end{aligned}$$

For $j\in A_n$ we have $\sqrt{j\log j}/\sqrt{n\log n} \to \sqrt{\mu_0}$ and $\sqrt{(n-j)\log(n-j)}/\sqrt{n\log n} \to \sqrt{1-\mu_0}$. Hence, we have $(B^0_j - \nu_0(j))/\sqrt{n\log n} \to \mathcal{N}(0,\mu_0\sigma^2)$ and $(B^1_{n-j} - \nu_1(n-j))/\sqrt{n\log n} \to \mathcal{N}(0,(1-\mu_0)\sigma^2)$ in distribution and the two summands are independent. Together, denoting by N_{0,σ^2} an $\mathcal{N}(0,\sigma^2)$ distributed random variable we obtain

$$\begin{aligned}\mathbb{P}\left(\frac{B^0_{K_n} - \nu_0(K_n) + B^1_{n-K_n} - \nu_1(n-K_n)}{\sqrt{n\log n}} \le x\right) &= o(1) + \sum_{j\in A_n}\mathbb{P}(K_n=j)(\mathbb{P}\left(N_{0,\sigma^2}\le x\right) + o(1))\\ &\to \mathbb{P}\left(N_{0,\sigma^2}\le x\right),\end{aligned}$$

where the latter convergence is justified by dominated convergence. This shows (56).

For (57) note that (4) implies

$$\mathbb{E}[B_n^\mu] = \mathbb{E}[\nu_0(K_n)] + \mathbb{E}[\nu_1(n-K_n)] + n.$$

Hence, with the notation (12) and $h(x) = x\log x$, $x\in[0,1]$, we have

$$\begin{aligned}&\frac{1}{\sqrt{n\log n}}\|\nu_0(K_n) + \nu_1(n-K_n) + n - \mathbb{E}[B_n^\mu]\|_3\\ &= \frac{1}{\sqrt{n\log n}}\|\nu_0(K_n) - \mathbb{E}[\nu_0(K_n)] + \nu_1(n-K_n) - \mathbb{E}[\nu_1(n-K_n)]\|_3\\ &\le \frac{1}{H\sqrt{n\log n}}\|h(K_n) - \mathbb{E}[h(K_n)] + h(n-K_n) - \mathbb{E}[h(n-K_n)]\|_3\\ &\quad + \frac{1}{\sqrt{n\log n}}\|f_0(K_n) - \mathbb{E}[f_0(K_n)]\|_3 + \frac{1}{\sqrt{n\log n}}\|f_1(n-K_n) - \mathbb{E}[f_1(n-K_n)]\|_3\end{aligned}$$

An easy calculation reveals (details are given in the appendix, equation (59))

$$\begin{aligned}&\|h(K_n) - \mathbb{E}[h(K_n)] + h(n-K_n) - \mathbb{E}[h(n-K_n)]\|_3\\ &= n\left\|h\left(\frac{K_n}{n}\right) - \mathbb{E}\left[h\left(\frac{K_n}{n}\right)\right] + h\left(\frac{n-K_n}{n}\right) - \mathbb{E}\left[h\left(\frac{n-K_n}{n}\right)\right]\right\|_3\\ &= \mathrm{O}\left(n^{1/2}\right).\end{aligned}$$

The terms $\|f_0(K_n) - \mathbb{E}[f_0(K_n)]\|_3$ and $\|f_1(n-K_n) - \mathbb{E}[f_1(n-K_n)]\|_3$ are also of the order $\mathrm{O}(n^{1/2})$ by the argument used in (48). Altogether we have

$$\frac{1}{\sqrt{n \log n}} \|\nu_0(K_n) + \nu_1(n-K_n) + n - \mathbb{E}[B_n^\mu]\|_3 = \mathrm{O}\left(\frac{1}{\sqrt{\log n}}\right),$$

which implies (57). □

References

[1] Bai, Z.-D., Chao, C.-C., Hwang, H.-K. and Liang, W.-Q. (1998) On the variance of the number of maxima in random vectors and its applications. *Ann. Appl. Probab.* **8**, 886–895.

[2] Cesaratto, E. and Vallée, B. (2015) Gaussian Distribution of Trie Depth for Strongly Tame Sources. *Combinatorics, Probability and Computing* **24**, 54–103.

[3] Clément, J., Flajolet, P. and Vallée, B. (2001) Dynamical sources in information theory: a general analysis of trie structures. Average-case analysis of algorithms (Princeton, NJ, 1998). *Algorithmica* **29**, 307–369.

[4] de la Briandais, R. (1959) File searching using variable length keys, in *Proceedings of the AFIPS Spring Joint Computer Conference*. AFIPS Press, Reston, Va., 295-298.

[5] Devroye, L. (1984) A probabilistic analysis of the height of tries and of the complexity of triesort. *Acta Inform.* **21**, 229–237.

[6] Devroye, L. (2002) Laws of large numbers and tail inequalities for random tries and PATRICIA trees. Probabilistic methods in combinatorics and combinatorial optimization. *J. Comput. Appl. Math.* **142**, 27–37.

[7] Devroye, L. (2005) Universal asymptotics for random tries and PATRICIA trees. *Algorithmica* **42**, 11–29.

[8] Drmota, M., Janson, S. and Neininger, R. (2008) A functional limit theorem for the profile of search trees. *Ann. Appl. Probab.* **18**, 288–333.

[9] Flajolet, Ph., Roux, M. and Vallée, B. (2010) Digital trees and memoryless sources: from arithmetics to analysis. *21st International Meeting on Probabilistic, Combinatorial, and Asymptotic Methods in the Analysis of Algorithms (AofA'10)*, Discrete Math. Theor. Comput. Sci. Proc., AM, Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 233–260.

[10] Fuchs, M., Hwang, H.-K. and Zacharovas, V. (2010) Asymptotic variance of random symmetric digital search trees. *Discrete Math. Theor. Comput. Sci.* **12**, 103–165.

[11] Fuchs, M., Hwang, H.-K. and Zacharovas, V. (2014) An analytic approach to the asymptotic variance of trie statistics and related structures. *Theoret. Comput. Sci.* **527**, 1–36.

[12] Hun, K. and Vallée, B. (2014) Typical depth of a digital search tree built on a general source. *ANALCO14—Meeting on Analytic Algorithmics and Combinatorics*, 1–15.

[13] Hwang, H.-K. (2003) Second phase changes in random m-ary search trees and generalized quicksort: convergence rates. *Ann. Probab.* **31**, 609–629.

[14] Jacquet, Ph. and Régnier, M. (1988) Normal limiting distribution of the size of tries. *Performance '87 (Brussels, 1987)*, 209–223, *North-Holland, Amsterdam.*

[15] Jacquet, Ph. and Szpankowski, W. (1991) Analysis of digital tries with Markovian dependency *IEEE Trans. Information Theory*, **37**, 1470–1475.

[16] Jacquet, Ph. and Szpankowski, W. (1995) Asymptotic behavior of the Lempel-Ziv parsing scheme and [in] digital search trees. Special volume on mathematical analysis of algorithms. *Theoret. Comput. Sci.* **144**, 161–197.

[17] Jacquet, Ph. and Szpankowski, W. (1998) Analytical Depoissonization and Its Applications, *Theoretical Computer Science*, 201, 1–62.

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

[19] Janson, S. and Neininger, R. (2008) The size of random fragmentation trees. *Probab. Theory Related Fields* **142**, 399–442.

[20] Kirschenhofer, P., Prodinger, H. and Szpankowski, W. (1989) On the variance of the external path length in a symmetric digital trie. Combinatorics and complexity (Chicago, IL, 1987). *Discrete Appl. Math.* **25**, 129–143.

[21] Kirschenhofer, P., Prodinger, H. and Szpankowski, W. (1996) Analysis of a Splitting Process Arising in Probabilistic Counting and Other Related Algorithms, *Random Structures & Algorithms*, 9, 379–401.

[22] Knuth, D.E. (1998) *The Art of Computer Programming, Volume III: Sorting and Searching*, Second edition, Addison Wesley, Reading, MA.

[23] Leckey. K. (2015) *Probabilistic Analysis of Radix Algorithms on Markov Sources.* Ph.D. dissertation, submitted at the Goethe University Frankfurt a.M. Available via `http://www.math.uni-frankfurt.de/~leckey/Dissertation.pdf`

[24] Leckey, K., Neininger, R. and Szpankowski, W. (2013) Towards More Realistic Probabilistic Models for Data Structures: The External Path Length in Tries under the Markov Model. *Proceedings ACM-SIAM Symp. Disc. Algo. (SODA)*, 877–886.

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

[26] Neininger, R. (2001) On a multivariate contraction method for random recursive structures with applications to Quicksort. Analysis of algorithms (Krynica Morska, 2000). *Random Structures Algorithms* **19**, 498–524.

[27] Neininger, R. and Rüschendorf, L. (2004) A general limit theorem for recursive algorithms and combinatorial structures. *Ann. Appl. Probab.* **14**, 378–418.

[28] Neininger, R. and Rüschendorf, L. (2004) On the contraction method with degenerate limit equation. *Ann. Probab.* **32**, 2838–2856.

[29] Neininger, R. and Sulzbach, H. (2015) On a functional contraction method. *Ann. Probab.*, to appear. Preprint available via `http://arxiv.org/abs/1202.1370`

[30] Rachev, S.T. and Rüschendorf, L. (1995) Probability metrics and recursive algorithms. *Adv. in Appl. Probab.* **27**, 770–799.

[31] Rösler, U. (1991) A limit theorem for "Quicksort". *RAIRO Inform. Théor. Appl.* **25**, 85–100.

[32] Rösler, U. (1992) A fixed point theorem for distributions. *Stochastic Process. Appl.* **42**, 195–214.

[33] Rösler, U. (1999) On the analysis of stochastic divide and conquer algorithms. Average-case analysis of algorithms (Princeton, NJ, 1998). *Algorithmica* **29**, 238–261.

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

[35] Schachinger, W. (1995) On the variance of a class of inductive valuations of data structures for digital search. *Theoret. Comput. Sci.* **144**, 251–275. Special volume on mathematical analysis of algorithms.

[36] Szpankowski, W. (2001) *Average Case Analysis of Algorithms on Sequences*, John Wiley, New York.

[37] Vallée, B. (2001) Dynamical sources in information theory: fundamental intervals and word prefixes, *Algorithmica* **1-2**, 262–306.

[38] Zolotarev, V. M. (1976) Approximation of the distributions of sums of independent random variables with values in infinite-dimensional spaces. (Russian.) *Teor. Veroyatnost. i Primenen.* **21**, 741–758. Erratum *ibid* **22** (1977), 901. English transl. *Theory Probab. Appl.* **21**, 721–737; *ibid.* **22**, 881.

[39] Zolotarev, V. M. (1977) Ideal metrics in the problem of approximating the distributions of sums of independent random variables. (Russian.) *Teor. Veroyatnost. i Primenen.* **22**, 449–465. English transl. *Theory Probab. Appl.* **22**, 433–449.

Appendix

Asymptotics of the Binomial and Poisson distribution

The appendix is meant to cover some elementary asymptotic moment calculations of the binomial and Poisson distribution. These calculations were made for the sake of completeness and may be removed in the published version of this paper.

The following approximations are immediate consequences of the concentration of the binomial distribution. Recall $x\log x = 0$ for $x = 0$.

Lemma 7.3. *Let $p \in (0,1)$, $h(x) := x\log x$ for $x \in [0,1]$ and $X_{n,p}$ be binomial $B(n,p)$ distributed for $n \in \mathbb{N}$. Then we have as $n \to \infty$*

$$\mathbb{E}\left[\log\left(\frac{X_{n,p}+1}{n+1}\right) - \log p\right] = \mathrm{O}\left(n^{-1/2}\right), \tag{58}$$

$$\|h(X_{n,p}/n) - \mathbb{E}\left[h(X_{n,p}/n)\right]\|_3 = \mathrm{O}\left(n^{-1/2}\right), \tag{59}$$

$$\mathbb{E}[h(X_{n,p}/n) - h(p)] = \mathrm{O}(n^{-2/3}). \tag{60}$$

Proof. Proof of (58): Note that we have for all $\varepsilon \in (0,1)$ by the mean value theorem

$$|\log(x) - \log(y)| \le \varepsilon^{-1}|x-y|, \quad x, y \in [\varepsilon, 1].$$

This yields

$$\begin{aligned}
&\left|\mathbb{E}\left[\log\left(\frac{X_{n,p}+1}{n+1}\right) - \log p\right]\right| \\
&\le \mathbb{E}\left[\left|\log\left(\frac{X_{n,p}+1}{n+1}\right) - \log p\right| \mathbf{1}_{\{X_{n,p} \ge np/2\}}\right] + \mathrm{O}\left(\log n \mathbb{P}(X_{n,p} < np/2)\right) \\
&\le \frac{2}{p}\mathbb{E}\left[\left|\frac{X_{n,p}+1-np-p}{n+1}\right|\right] + \mathrm{O}\left(\log n \mathbb{P}(X_{n,p} < np/2)\right).
\end{aligned}$$

The assertion follows since $\mathbb{E}[|(X_{n,p}-np)/\sqrt{np(1-p)}|]$ converges to the first absolut moment of the standard normal distribution and $\log n\mathbb{P}(X_{n,p} < np/2) = o(n^{-1/2})$ by Chernoff's bound.

Proof of (59): First note that h is bounded on $[0,1]$ and that we have for all $\varepsilon \in (0,1)$

$$|h'(x)| \le \log(1/\varepsilon) + 1, \quad x \in [\varepsilon, 1].$$

In particular, we obtain by the mean value theorem that

$$|h(x) - h(y)| \le (\log(1/\varepsilon)+1)|x-y|, \quad x, y \in [\varepsilon, 1]. \tag{61}$$

With an independent copy $\widetilde{X}_{n,p}$ of $X_{n,p}$ we obtain by Jensen's inequality and (61)

$$\begin{aligned}
&\|h(X_{n,p}/n) - \mathbb{E}\left[h(X_{n,p}/n)\right]\|_3^3 \\
&= \mathbb{E}[(\mathbb{E}[h(X_{n,p}/n) - h(\widetilde{X}_{n,p}/n)|X_{n,p}])^3] \\
&\le \mathbb{E}[(h(X_{n,p}/n) - h(\widetilde{X}_{n,p}/n))^3] \\
&= \mathbb{E}[(h(X_{n,p}/n) - h(\widetilde{X}_{n,p}/n))^3 \mathbf{1}_{\{X_{n,p}, \widetilde{X}_{n,p} \in [np/2, n]\}}] + \mathrm{O}(\mathbb{P}(X_{n,p} \le np/2)) \\
&\le (\log(2/p)+1)^3 \mathbb{E}[(X_{n,p}/n - \widetilde{X}_{n,p}/n)^3] + \mathrm{O}(\mathbb{P}(X_{n,p} \le np/2)) \\
&\le \left(\frac{\log(2/p)+1}{\sqrt{n}}\right)^3 (2\|X_{n,p}/\sqrt{n}\|_3)^3 + \mathrm{O}(\mathbb{P}(X_{n,p} \le np/2)).
\end{aligned}$$

The assertion follows by Chernoff's bound on $\mathbb{P}(X_{n,p} \le np/2)$ and $\|X_{n,p}/\sqrt{n}\|_3 \to \|N\|_3$ where N is $\mathcal{N}(0, p(1-p))$ distributed.

Proof of (60): It is sufficient to show that

1. $h(p) - p\mathbb{E}[\log(X_{n,p}/n)\mathbf{1}_{\{X_{n,p}\geq 1\}}] = \mathrm{O}(n^{-2/3})$,
2. $\mathbb{E}[h(X_{n,p}/n) - p\log(X_{n,p}/n)\mathbf{1}_{\{X_{n,p}\geq 1\}}] = \mathrm{O}(n^{-2/3})$.

For the first part note that we have

$$\begin{aligned}
&\left|h(p) - p\mathbb{E}[\log(X_{n,p}/n)\mathbf{1}_{\{X_{n,p}\geq 1\}}]\right| \\
&= p\left|\mathbb{E}\left[\log\left(\frac{X_{n,p}}{np}\right)\mathbf{1}_{\{X_{n,p}\geq 1\}}\right]\right| + \mathrm{O}\left((1-p)^n\right) \\
&= p\left|\mathbb{E}\left[\left(\log\left(1+\frac{X_{n,p}-np}{np}\right) - \frac{X_{n,p}-np}{np}\right)\mathbf{1}_{\{X_{n,p}\geq 1\}}\right]\right| + \mathrm{O}\left((1-p)^n\right) \\
&\leq p\left|\mathbb{E}\left[\left(\log\left(1+\frac{X_{n,p}-np}{np}\right) - \frac{X_{n,p}-np}{np}\right)\mathbf{1}_{\{|X_{n,p}-np|\leq n^{2/3}\}}\right]\right| \\
&\quad + (\log(np) + 1/p)\mathbb{P}(|X_{n,p}-np| > n^{2/3}) + \mathrm{O}\left((1-p)^n\right).
\end{aligned}$$

Since we have $\log(1+x) - x = \mathrm{O}(x^2)$ for $x \to 0$ and $\mathbb{P}(|X_{n,p}-np| > n^{2/3}) = o(n^{-1})$ by Chernoff's bound, we may conclude that

$$h(p) - p\mathbb{E}[\log(X_{n,p}/n)\mathbf{1}_{\{X_{n,p}\geq 1\}}] = \mathrm{O}(n^{-2/3}).$$

In order to obtain the second bound, note that

$$\begin{aligned}
&\mathbb{E}[h(X_{n,p}/n) - p\log(X_{n,p}/n)\mathbf{1}_{\{X_{n,p}\geq 1\}}] \\
&= \mathbb{E}\left[(h(X_{n,p}/n) - p\log(X_{n,p}/n))\mathbf{1}_{\{X_{n,p}\geq 1\}}\right] + \mathrm{O}\left((1-p)^n\right) \\
&= \frac{1}{\sqrt{n}}\mathbb{E}\left[\frac{X_{n,p}-np}{\sqrt{n}}\log\left(\frac{X_{n,p}}{n}\right)\mathbf{1}_{\{X_{n,p}\geq 1\}}\right] + \mathrm{O}\left((1-p)^n\right) \\
&= \frac{1}{\sqrt{n}}\mathbb{E}\left[\frac{X_{n,p}-np}{\sqrt{n}}\log\left(\frac{X_{n,p}}{n}\right)\mathbf{1}_{\{|X_{n,p}-np|\leq n^{2/3}\}}\right] + o\left(n^{-2/3}\right) \\
&= \frac{1}{\sqrt{n}}\mathbb{E}\left[\frac{X_{n,p}-np}{\sqrt{n}}\log(p)\mathbf{1}_{\{|X_{n,p}-np|\leq n^{2/3}\}}\right] \\
&\quad + \frac{1}{\sqrt{n}}\mathbb{E}\left[\frac{X_{n,p}-np}{\sqrt{n}}\log\left(1+\frac{X_{n,p}-np}{np}\right)\mathbf{1}_{\{|X_{n,p}-np|\leq n^{2/3}\}}\right] + o\left(n^{-2/3}\right).
\end{aligned}$$

Since $\log(1+x) = \mathrm{O}(x)$ as $x \to 0$ and $\mathbb{E}[|(X_{n,p}-np)/\sqrt{n}|]$ converges to the first absolute moment of the $\mathcal{N}(0, p(1-p))$ distribution, we obtain for the second summand

$$\frac{1}{\sqrt{n}}\mathbb{E}\left[\frac{X_{n,p}-np}{\sqrt{n}}\log\left(1+\frac{X_{n,p}-np}{np}\right)\mathbf{1}_{\{|X_{n,p}-np|\leq n^{2/3}\}}\right] = \mathrm{O}(n^{-5/6}).$$

For the first summand note that $\mathbb{E}[(X_{n,p}-np)/\sqrt{n}] = 0$ which implies

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\mathbb{E}\left[\frac{X_{n,p}-np}{\sqrt{n}}\log(p)\mathbf{1}_{\{|X_{n,p}-np|\leq n^{2/3}\}}\right] \\
&= -\frac{1}{\sqrt{n}}\mathbb{E}\left[\frac{X_{n,p}-np}{\sqrt{n}}\log(p)\mathbf{1}_{\{|X_{n,p}-np|> n^{2/3}\}}\right] \\
&= \mathrm{O}(\mathbb{P}(|X_{n,p}-np| > n^{2/3})) \\
&= o(n^{-2/3}).
\end{aligned}$$

Hence, we obtain $\mathbb{E}[h(X_{n,p}/n) - p\log(X_{n,p}/n)\mathbf{1}_{\{X_{n,p}\geq 1\}}] = \mathrm{O}(n^{-2/3})$ which combined with the first result yields the assertion. □

The next Lemma provides asymptotic results for the poisson distribution that are needed for the analysis of the variance:

Lemma 7.4. *For $\lambda > 0$ let N_λ be Poisson(λ) distributed. Then we have for all $\alpha, \beta > 0$ as $\lambda \to \infty$*

$$\mathbb{E}[N_\lambda^\alpha] = \mathrm{O}(\lambda^\alpha),$$
$$\mathbb{E}\left[N_\lambda^\alpha (\log N_\lambda)^\beta\right] = \mathrm{O}\left(\lambda^\alpha (\log \lambda)^\beta\right).$$

Proof. We start with the analysis of $\mathbb{E}[N_\lambda^\alpha]$: For $\alpha \in \mathbb{N}$ the assertion follows by induction and the fact that for every $n \in \mathbb{N}_0$ we have

$$\mathbb{E}\left[\prod_{i=0}^{n}(N_\lambda - i)\right] = \lambda^{n+1}.$$

For $\alpha \in (0,1)$ note that $x \mapsto x^\alpha$ is concave on $[0,\infty)$ and therefore, by Jensen's inequality

$$\mathbb{E}[N_\lambda^\alpha] \le (\mathbb{E}[N_\lambda])^\alpha = \lambda^\alpha.$$

Finally, for $\alpha \in (1,\infty) \cap \mathbb{N}^c$ we have that $x \mapsto x^{\alpha/\lceil \alpha \rceil}$ is concave on $[0,\infty)$ which yields

$$\mathbb{E}[N_\lambda^\alpha] \le (\mathbb{E}[N_\lambda^{\lceil \alpha \rceil}])^{\alpha/\lceil \alpha \rceil}$$

and the assertion follows by the results for $\alpha \in \mathbb{N}$.

For the second part of the proof we use the following decomposition

$$\begin{aligned}
\mathbb{E}\left[N_\lambda^\alpha (\log N_\lambda)^\beta\right] &= \mathbb{E}\left[N_\lambda^\alpha (\log N_\lambda)^\beta \mathbf{1}_{\{N_\lambda \le \lambda^{\alpha+1}\}}\right] + \mathbb{E}\left[N_\lambda^\alpha (\log N_\lambda)^\beta \mathbf{1}_{\{N_\lambda > \lambda^{\alpha+1}\}}\right] \\
&\le (\alpha+1)^\beta (\log \lambda)^\beta \mathbb{E}[N_\lambda^\alpha] + \mathbb{E}\left[N_\lambda^\alpha (\log N_\lambda)^\beta \mathbf{1}_{\{N_\lambda > \lambda^{\alpha+1}\}}\right] \\
&= \mathrm{O}(\lambda^\alpha (\log \lambda)^\beta) + \mathbb{E}\left[N_\lambda^\alpha (\log N_\lambda)^\beta \mathbf{1}_{\{N_\lambda > \lambda^{\alpha+1}\}}\right],
\end{aligned}$$

where the last step holds since $\mathbb{E}[N_\lambda^\alpha] = \mathrm{O}(\lambda^\alpha)$. Hence, it is sufficient to show that

$$\mathbb{E}\left[N_\lambda^\alpha (\log N_\lambda)^\beta \mathbf{1}_{\{N_\lambda > \lambda^{\alpha+1}\}}\right] = \mathrm{O}(\lambda^\alpha).$$

Since we have $n^\alpha (\log n)^\beta \le C_{\alpha\beta} n^{3\alpha/2}$ for a sufficiently large constant $C_{\alpha\beta}$ and all $n \in \mathbb{N}_0$, we obtain

$$\begin{aligned}
\mathbb{E}\left[N_\lambda^\alpha (\log N_\lambda)^\beta \mathbf{1}_{\{N_\lambda > \lambda^{\alpha+1}\}}\right] &\le C_{\alpha\beta} \mathbb{E}\left[N_\lambda^{3\alpha/2} \mathbf{1}_{\{N_\lambda > \lambda^{\alpha+1}\}}\right] \\
&\le C_{\alpha\beta} \sqrt{\mathbb{E}[N_\lambda^{3\alpha}] \mathbb{P}(N_\lambda > \lambda^{\alpha+1})}
\end{aligned}$$

where the last inequality holds by the Cauchy-Schwarz inequality. Together with the previous result $\mathbb{E}[N_\lambda^{3\alpha}] = \mathrm{O}(\lambda^{3\alpha})$ and Markov's inequality this yields

$$\mathbb{E}\left[N_\lambda^\alpha (\log N_\lambda)^\beta \mathbf{1}_{\{N_\lambda > \lambda^{\alpha+1}\}}\right] = \mathrm{O}(\lambda^\alpha)$$

and the assertion follows. □