Towards More Realistic Probabilistic Models for Data Structures: The External Path Length in Tries under the Markov Model

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

September 18, 2012

Abstract

Tries are among the most versatile and widely used data structures on words. They are pertinent to the (internal) structure of (stored) words and several splitting procedures used in diverse contexts ranging from document taxonomy to IP addresses lookup, from data compression (i.e., Lempel-Ziv'77 scheme) to dynamic hashing, from partial-match queries to speech recognition, from leader election algorithms to distributed hashing tables and graph compression. While the performance of tries under a realistic probabilistic model is of significant importance, its analysis, even for simplest memoryless sources, has proved difficult. Rigorous findings about inherently complex parameters were rarely analyzed (with a few notable exceptions) under more realistic models of string generations. In this paper we meet these challenges: By a novel use of the contraction method combined with analytic techniques we prove a central limit theorem for the external path length of a trie under a general Markov source. In particular, our results apply to the Lempel-Ziv'77 code. We envision that the methods described here will have further applications to other trie parameters and data structures.

1 Introduction

We study 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 [2, 5, 6, 7, 8, 9, 15, 16, 18, 20, 22, 23, 24, 26, 27, 40], mostly under Bernoulli (memoryless) model assumption. The typical depth under Markovian model was analyzed in [16, 20]. Size, external path length and height under more general *dynamical sources* were studied in the seminal paper of Clément, Flajolet, and Vallée [2], where in particular asymptotic expressions for expectations are identified as well as the asymptotic distributional behavior of the height, see also [3]. For further analysis of tries for probabilistic models beyond Bernoulli (memoryless) sources see Devroye [6, 7].

With respect to Markovian models, to the best of our knowledge, no asymptotic distributions for the external path length have been derived so far. It is well known [40] that the external path length is more challenging due to stronger dependency. In fact, this is already observed for tries under Bernoulli model [40]. In this paper we establish the central limit theorem for the external path length in a trie built over a Markov model using a novel use of the *contraction method*.

Let us first briefly review the contraction method. It was introduced in 1991 by Uwe Rösler [34] for the distributional analysis of the complexity of the Quicksort algorithm. Over the last 20 years this approach, which

*This author's contribution was made while visiting J.W. Goethe University Frankfurt a.M. with an Alexander von Humboldt research award. This author was also supported by the NSF Science and Technology Center on Science of Information Grant CCF-0939370, NSF Grants DMS-0800568 and CCF-0830140, AFOSR Grant FA8655- 11-1-3076, NSA Grant H98230-08-1-0092, and MNSW grant N206 369739. He is also Visiting Professor at ETI, Gdansk University of Technology, Poland.

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 [35, 32, 36, 29, 30, 11, 10, 21, 31] with numerous applications in Theoretical Computer Science.

The contraction method has been used in the analysis of tries and other digital trees only under the symmetric Bernoulli model (unbiased memoryless source) [29, 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. Let us observe 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 [12]. 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 into its left and right subtree of the root. 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 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. A comparison of a multivariate approach and our new version with systems of recurrences is drawn in Section 7.

We also need asymptotic expansions of the mean and the variance for applying the contraction method. However, in contrast to very precise information on periodicities of linear terms for the symmetric Bernoulli model mentioned above our convergence proof does only require the leading order term together with a Lipschitz continuity property for the error term.

In this extended abstract we develop the use of systems of recursive distributional equations in the context of the contraction method for the external path length of tries under a general Markov source model. In particular, we prove the central limit theorem for the external path length, a result that had been wanting since Lempel-Ziv'77 code was devised in 1977. The methodology used is general enough to cover related quantities and structures as well. We are confident that our approach also applies with minor adjustments at least to the size of tries, the path lengths of digital search trees and PATRICIA tries under the Markov source model as well as other more complex data structures on words such as suffix trees.

Notations: Throughout this paper we use the Bachmann-Landau symbols, in particular the big O notation. We declare $x \log x := 0$ for $x = 0$, where $\log x$ denotes the natural logarithm. 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 Tries and the Markov source model

The Markov source: We assume binary data strings over the alphabet $\Sigma = \{0,1\}$ generated by a homogeneous Markov chain. 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

$$\pi_0 = \frac{p_{10}}{p_{01}+p_{10}}, \qquad \pi_1 = \frac{p_{01}}{p_{01}+p_{10}}. \tag{2.1}$$

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 [14, 23, 29]. It behaves differently compared to the asymmetric Bernoulli model and the other Markov source models, as the variance of the external path length is linear with a periodic prefactor in the symmetric Bernoulli model. In our cases we will find a larger variance of the order $n \log n$ in Theorem 5.1 below. We exclude the symmetric Bernoulli model case subsequently. For later reference, we summarize our conditions as:

$$p_{ij} \in (0,1) \text{ for all } i,j \in \Sigma, \qquad p_{ij} \neq \frac{1}{2} \text{ for some } (i,j) \in \Sigma^2. \tag{2.2}$$

The entropy rate of the Markov chain plays an important role in the asymptotic behavior of tries. In particular, it determines leading order constants of parameters of tries that are related to depths of leaves and its external path length. The entropy rate for our Markov chain is given by

$$H := -\sum_{i,j\in\Sigma} \pi_i \, p_{ij} \log p_{ij} = \sum_{i\in\Sigma} \pi_i H_i, \tag{2.3}$$

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.

Tries: For a given set of data strings over the alphabet $\Sigma = \{0,1\}$ with each data string a unique infinite path in the infinite complete rooted binary tree is associated by identifying left branches with bit 0 and right branches with bit 1. Each string is stored in the unique node on its infinite path that is closest to the root and does not belong to any other data path, cf. Figure 1. It is the minimal prefix of a string that distinguishes this string from all others; for details see the monographs of Knuth [26], Mahmoud [27] or Szpankowski [40].

3 Recursive Distributional Equations

For the Markov source model a challenge is to set the right framework under which data structures to analyze. We formulate in this section a system of distributional recurrences to capture the distribution of the external path length of tries. Our subsequent analysis is entirely based on these equations.

We denote by L_n^μ the external path length of a trie under the Markov source model with initial distribution μ holding n data. We have $L_0^\mu = L_1^\mu = 0$ for all initial distributions μ. The transition matrix is given in advance and suppressed in the notation. We abbreviate $L_n^i := L_n^{\delta_i}$ for $i \in \Sigma$. Hence, L_n^i refers to n independent strings all starting with bit i and then following the

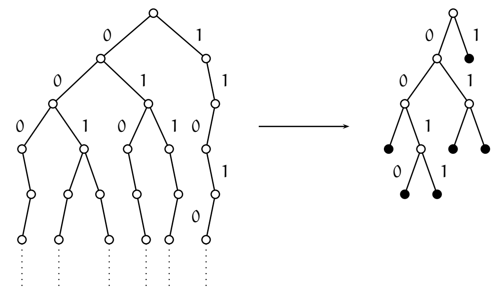

Figure 1: The infinite rooted binary tree contains the infinite paths of six strings (left). The corresponding trie is obtained by cutting each path at the closest node to the root that does not belong to any other path.

Markov chain. We will study L_n^0 and L_n^1. From the asymptotic behavior of these two sequences we can then directly obtain corresponding results for L_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 strings which start with bit 0. Then K_n has the binomial $B(n, \mu_0)$ distribution. The contributions of the two subtrees of the trie to its external path length can be represented by the following stochastic recurrence

$$L_n^\mu \stackrel{d}{=} L_{K_n}^0 + L_{n-K_n}^1, \qquad n \geq 2, \tag{3.4}$$

where $\stackrel{d}{=}$ denotes that left and right hand side have identical distributions and we have that $(L_0^0, \dots, L_n^0)$, $(L_0^1, \dots, L_n^1)$ and K_n are independent. We will see later that we can directly transfer asymptotic results for L_n^0 and L_n^1 to general L_n^μ via (3.4), see, e.g., the proof of Theorem 6.1.

For a recursive decomposition of L_n^0 note that we have initial distribution δ_0, thus all data strings start with bit 0 and are inserted into the left subtree of the root. We denote the root of this left subtree by w. At node w the data strings are split according to their second bit. We denote by I_n the number of data strings having 0 as their second bit, i.e., the number of strings being inserted into the left subtree of w. The Markov source model implies that I_n is binomial $B(n, p_{00})$ distributed. The right subtree of node w then holds the remaining $n - I_n$ data strings. Consider the left subtree of w together with its root w. Conditioned on its number I_n of data strings inserted it is generated by the same Markov source model as the original trie. However, the right subtree of w together with its root w conditioned on its number $n - I_n$ of data strings is generated by a Markov source model with the same transition matrix but another initial distribution, namely δ_1. Moreover, by the independence of data strings within the Markov source model, these two subtrees are independent conditionally on I_n. Phrased in a recursive distributional equation we have

$$L_n^0 \stackrel{d}{=} L_{I_n}^0 + L_{n-I_n}^1 + n, \qquad n \geq 2, \tag{3.5}$$

with $(L_0^0, \ldots, L_n^0)$, $(L_0^1, \ldots, L_n^1)$ and I_n independent. A similar arguments yields a recurrence for L_n^1. Denoting by J_n a binomial $B(n, p_{11})$ distributed random variable, we have

$$L_n^1 \stackrel{d}{=} L_{n-J_n}^0 + L_{J_n}^1 + n, \qquad n \geq 2, \tag{3.6}$$

with $(L_0^0, \ldots, L_n^0)$, $(L_0^1, \ldots, L_n^1)$ and J_n independent. Our asymptotic analysis of L_n^μ is based on the distributional recurrence system (3.5)–(3.6) as well as (3.4).

4 Analysis of the Mean

First we study the asymptotic behavior of the expectation of the external path length with a precise error term needed to derive a limit law in Section 6. The leading order term in Theorem 4.1 below has already been derived (even for more general models) in Clément, Flajolet and Vallée [2].

THEOREM 4.1. *For the external path length L_n^μ of a binary trie under the Markov source model with conditions (2.2) we have*

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

with the entropy rate H of the Markov chain given in (2.3). The $\mathrm{O}(n)$ error term is uniform in the initial distribution μ.

Our proof of Theorem 4.1 as well as the corresponding limit law in Theorem 6.1 depend on refined properties of the $\mathrm{O}(n)$ error term that are first obtained for the initial distributions $\mu = \delta_0$ and $\mu = \delta_1$ and then generalized to arbitrary initial distribution via (3.4). For $\mu = \delta_0$ and $\mu = \delta_1$ we denote this error term for all $n \in \mathbb{N}_0$ and $i \in \Sigma$ by

$$f_i(n) := \mathbb{E}[L_n^i] - \frac{1}{H} n \log n. \tag{4.7}$$

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

PROPOSITION 4.2. *There exists a constant $C > 0$ such that for both $i \in \Sigma$ and all $m, n \in \mathbb{N}_0$*

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

The proof of Proposition 4.2 is based on a refined analysis of transfers from growth of toll functions in systems of recursive equations to the growth of the quantities itself. The heart of the proof of Proposition 4.2 and hence Theorem 4.1 is the following transfer result. The proof is technical and provided in the full paper version of this extended abstract.

LEMMA 4.3. *Let $(a_i(n))_{n\geq 0}$ and $(\eta_i(n))_{n\geq 0}$ be real sequences and $(X_{i,n})_{n\geq 2}$ sequences of binomial $B(n, p_i)$ distributed random with $p_i \in (0,1)$ for $i \in \Sigma$. Assume that for constants $c_0, c_1, d_0, d_1 \in (0,1)$ with $c_0 + d_1 = c_1 + d_0 = 1$ we have for all $n \geq 2$ and $i \in \Sigma$*

$$\begin{aligned} a_i(n) = c_i \mathbb{E}[a_i(X_{i,n})] + d_i \mathbb{E}[a_{1-i}(n - X_{i,n})] \\ + \eta_i(n). \end{aligned} \tag{4.8}$$

If furthermore $\eta_i(n) = \mathrm{O}(n^{-\alpha})$ for an $\alpha > 0$ and both $i \in \Sigma$, then, as $n \to \infty$,

$$a_i(n) = \mathrm{O}(1), \quad i \in \Sigma.$$

5 Analysis of the Variance

To formulate an asymptotic expansion of the variance of the external path length we denote by $\lambda(s)$ the largest eigenvalue of the matrix $P(s) := (p_{ij}^{-s})_{i,j\in\Sigma}$. Note that λ as a function of s is smooth. We denote its first and second derivative by $\dot{\lambda}$ and $\ddot{\lambda}$ respectively. Then we have:

THEOREM 5.1. *For the external path length L_n^μ of a binary trie under the Markov source model with conditions (2.2) we have, as $n \to \infty$,*

$$\mathrm{Var}(L_n^\mu) = \sigma^2 n \log n + o(n \log n), \tag{5.9}$$

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

$$\sigma^2 = \frac{\ddot{\lambda}(-1) - \dot{\lambda}^2(-1)}{\dot{\lambda}^3(-1)}. \tag{5.10}$$

With H_0 and H_1 defined in (2.3) we have

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

We start with the analysis of the Poisson variance of the external path length, i.e. $\tilde{v}_i(\lambda) := \mathrm{Var}(L_{N_\lambda}^i)$, $i \in \Sigma$, where N_λ has the Poisson(λ) distribution and is independent of $(L_n^i)_{n\geq 0}$. In the second part we use depoissonization techniques of [19] to obtain the asymptotic behavior of $\mathrm{Var}(L_n^i)$.

The reason why we consider a Poisson number of strings is that for N_λ i.i.d. strings with initial distribution δ_i the number $N_{\lambda p_{i0}}$ of strings whose second bit equals 0 and the number $M_{\lambda p_{i1}}$ of strings whose second bit equals 1 are independent and remain Poisson distributed. Hence, in the Poisson case we obtain similarly to (3.5) and (3.6) that for $i \in \Sigma$

$$L^i_{N_\lambda} \overset{d}{=} L^0_{N_{\lambda p_{i0}}} + L^1_{M_{\lambda p_{i1}}} + N_{\lambda p_{i0}} + M_{\lambda p_{i1}} - \mathbf{1}_{\{N_{\lambda p_{i0}}+M_{\lambda p_{i1}}=1\}} \tag{5.11}$$

where $(L^0_n)_{n\geq 0}$, $(L^1_n)_{n\geq 0}$, $N_{\lambda p_{i0}}$ and $M_{\lambda p_{i1}}$ are independent, $N_{\lambda p_{i0}}$ has Poisson(λp_{i0}) distribution and $M_{\lambda p_{i1}}$ has Poisson(λp_{i1}) distribution. Note that $\mathbf{1}_{\{N_{\lambda p_{i0}}+M_{\lambda p_{i1}}=1\}}$ is necessary in order that (5.11) holds when $\{N_\lambda = 1\}$.

We denote by $\tilde{\nu}_i(\lambda) := \mathbb{E}[L^i_{N_\lambda}]$, $i \in \Sigma$, the Poisson expectation of the external path length which is

$$\tilde{\nu}_i(\lambda) = \sum_{n=0}^{\infty} e^{-\lambda}\frac{\lambda^n}{n!}\mathbb{E}[L^i_n].$$

Note that (5.11) implies

$$\tilde{\nu}_i(\lambda) = \tilde{\nu}_0(\lambda p_{i0}) + \tilde{\nu}_1(\lambda p_{i1}) + \lambda(1-e^{-\lambda}). \tag{5.12}$$

We need precise information about the mean (second order term) to derive the leading term of the variance. We shall use analytic techniques, namely the Mellin transform as surveyed in [40] that we discuss next. A Mellin transform $f^*(s)$ of a real function $f(x)$ is defined as

$$f^*(s) = \int_0^\infty f(x)x^{s-1}dx.$$

Let $\nu_i^*(s)$ be the Mellin transform of $\tilde{\nu}_i(\lambda)$. Then, by known properties of the Mellin transform [40], the functional equation (5.12) becomes an algebraic equation for $i \in \Sigma$

$$\nu_i^*(s) = \Gamma(s+1) + p_{i0}^{-s}\nu_0^*(s) + p_{i1}^{-s}\nu_1^*(s).$$

Define the column vector $\boldsymbol{\nu}^*(s) := (\nu_0^*(s), \nu_1^*(s))$ and the column vector $\boldsymbol{\gamma}(s) := (\Gamma(s), \Gamma(s))$. Then we can write the latter equations as the matrix equation $\boldsymbol{\nu}^*(s) = \boldsymbol{\gamma}(s+1) + P(s)\boldsymbol{\nu}^*(s)$ that we write as

$$\boldsymbol{\nu}^*(s) = (I - P(s))^{-1}\boldsymbol{\gamma}(s+1). \tag{5.13}$$

Then the Mellin transform $\nu^*(s)$ of the mean external path length $\mathbb{E}[L^\mu_{N_\lambda}]$ under the Poisson model satisfies

$$\nu^*(s) = \Gamma(s+1) + \boldsymbol{\mu}(s)\boldsymbol{\nu}^*(s) \tag{5.14}$$

where $\boldsymbol{\mu}(s) := (\mu_0^{-s}, \mu_1^{-s})$.

To recover the mean external path length under the Poisson model we need to apply the singularity analysis to (5.14). For matrix $P(s)$, we define the principal left eigenvector $\boldsymbol{\pi}(s)$, the principal right eigenvector $\boldsymbol{\psi}(s)$ associated with the largest eigenvalue $\lambda(s)$ such that $\langle \boldsymbol{\pi}(s), \boldsymbol{\psi}(s)\rangle = 1$ where we write $\langle \mathbf{x}, \mathbf{y}\rangle$ for the inner product of vectors $\mathbf{x}$ and $\mathbf{y}$. Then by the *spectral representation* [40] of $P(s)$ we find

$$\nu^*(s) = \frac{\Gamma(s)\boldsymbol{\psi}(s)}{1-\lambda(s)} + o(1/(1-\lambda(s)))$$

that leads to the following asymptotic expansion around $s = -1$

$$\nu^*(s) = \frac{-1}{\dot\lambda(-1)}\frac{1}{(s+1)^2} + \frac{1}{s+1}\left(\frac{\gamma}{\dot\lambda(-1)} + \frac{\ddot\lambda(-1)}{2\dot\lambda(-1)}\right) + \frac{1}{s+1}\left(-\frac{\langle\dot{\boldsymbol{\mu}}(-1)\dot{\boldsymbol{\psi}}(-1)\rangle}{\dot\lambda(-1)} + 1\right) + \mathrm{O}(1) \tag{5.15}$$

where $\dot{\mathbf{x}}(t)$ and $\ddot{\mathbf{x}}(t)$ denote the first and second derivatives of the vector $\mathbf{x}(t)$ at t.

Using (5.15), inverse Mellin transform, and the residue theorem of Cauchy, as well as analytic depoissonization of Jacquet and Szpankowski [19] we finally obtain

$$\mathbb{E}[L^\mu_n] = \frac{1}{H}n\log n + n\left(\frac{\gamma}{\dot\lambda(-1)} + \frac{\ddot\lambda(-1)}{2\dot\lambda(-1)}\right) + n\left(-\frac{\langle\dot{\boldsymbol{\mu}}(-1)\dot{\boldsymbol{\psi}}(-1)\rangle}{\dot\lambda(-1)} + 1 + \Phi(\log n)\right) + o(n) \tag{5.16}$$

where $\Phi(x)$ is a periodic function of small amplitude under certain rationality condition (and zero otherwise); see [20] for details.

The asymptotic analysis of the variance follows the same pattern, however, it is more involved. Our analysis of the Poisson variance $\tilde{v}_i(\lambda) = \mathrm{Var}(L^i_{N_\lambda})$ is based on the following decomposition:

Lemma 5.2. *For any $\lambda > 0$ and $i \in \Sigma$ we have*

$$\begin{aligned}\tilde{v}_i(\lambda) &= \tilde{v}_0(\lambda p_{i0}) + \tilde{v}_1(\lambda p_{i1}) + 2\lambda p_{i0}\tilde{\nu}_0'(\lambda p_{i0}) \\ &\quad + 2\lambda p_{i1}\tilde{\nu}_1'(\lambda p_{i1}) + 2\lambda e^{-\lambda}(\tilde{\nu}_0(\lambda p_{i0}) + \tilde{\nu}_1(\lambda p_{i1})) \\ &\quad + \lambda(1-e^{-\lambda}) + \lambda^2 e^{-\lambda}(2-e^{-\lambda})\end{aligned} \tag{5.17}$$

where $\tilde{\nu}_i', i \in \Sigma$, denotes the derivative of ν_i, i.e. for $z > 0$

$$\tilde{\nu}_i'(z) = \sum_{n=1}^{\infty} e^{-z}\frac{z^{n-1}}{(n-1)!}\mathbb{E}[L^i_n] - \tilde{\nu}_i(z).$$

The Mellin transform $v_i^*(s)$ of $\tilde{v}_i(\lambda)$ is

$$\begin{aligned} v_i^*(s) &= p_{i0}^{-s}v_0^*(s) + p_{i1}^{-s}v_1^*(s) - 2sp_{i0}^{-s}\nu_0^*(s) \\ &\quad - 2sp_{i1}^{-s}\nu_1^*(s) - \Gamma(s+1) + F_i^*(s) \end{aligned}$$

with $F_i^*(s)$ the Mellin transform of $e^{-\lambda}(\tilde{\nu}_0'(\lambda p_{i0}) + 2\lambda p_{i1}\tilde{\nu}_1'(\lambda p_{i1}) + \lambda^2(2-e^{-\lambda}))$. Thus, the column vector $\mathbf{v}^*(s) := (v_0^*(s), v_1^*(s))$ satisfies the following algebraic equation

$$\begin{aligned} \mathbf{v}^*(s) =& P(s)\mathbf{v}^*(s) - 2sP(s-1)\boldsymbol{\nu}^*(s) \\ & - \boldsymbol{\gamma}(s+1) + \mathbf{F}^*(s) \end{aligned}$$

where $\mathbf{F}^*(s) := (F_0^*(s), F_1^*(s))$. Then, as we did before for the mean analysis, we obtain

$$\begin{aligned} \mathbf{v}(s) = &-\frac{2s\Gamma(s+1)\langle \boldsymbol{\pi}(s), P(s-1)\boldsymbol{\psi}(s)\rangle \boldsymbol{\psi}(s)}{(1-\lambda(s))^2} \\ &+ \mathrm{O}(1/(1-\lambda(s)). \end{aligned}$$

After further computations we find that the Poisson variance $\tilde{v}(\lambda) = \mathrm{Var}(L^{\mu}_{N_\lambda})$ is

$$\begin{aligned} \tilde{v}(\lambda) = &\frac{1}{\dot{\lambda}^2(-1)}\lambda\log^2\lambda + \left(\frac{\ddot{\lambda}(-1)}{2\dot{\lambda}^3(-1)} + \frac{A}{\dot{\lambda}^2(-1)}\right)\lambda\log\lambda \\ &+ \mathrm{O}(\lambda) \end{aligned}$$

for some explicitly computable constant A. Finally, with depoissonization, cf. [40], we obtain

$$\begin{aligned} \mathrm{Var}(L_n^{\mu}) &= \tilde{v}(n) - n[\tilde{\nu}'(n)]^2 \\ &= \frac{\ddot{\lambda}(-1) - \dot{\lambda}^2(-1)}{\dot{\lambda}^3(-1)} n\log n + \mathrm{O}(n) \end{aligned}$$

proving Theorem 5.1.

6 Asymptotic Normality

Our main result is the asymptotic normality of the external path length:

Theorem 6.1. *For the external path length L_n^{μ} of a binary trie under the Markov source model with conditions (2.2) we have*

$$\frac{L_n^{\mu} - \mathbb{E}[L_n^{\mu}]}{\sqrt{n\log n}} \xrightarrow{d} \mathcal{N}(0,\sigma^2), \qquad (n\to\infty), \tag{6.18}$$

where $\sigma^2 > 0$ is independent of the initial distribution μ and given by (5.10).

As in the analysis of the mean, we first derive limit laws for L_n^0 and L_n^1 and then transfer these to a limit law for L_n^{μ} via (3.4). We abbreviate for $i\in\Sigma$ and $n\in\mathbb{N}_0$

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

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

$$Y_n^i := \frac{L_n^i - \mathbb{E}[L_n^i]}{\sigma_i(n)}, \qquad i\in\Sigma, n\geq 2, \tag{6.19}$$

and $Y_0^i := Y_1^i := 0$. Then we have:

Proposition 6.2. *For both sequences $(Y_n^i)_{n\geq 0}$, $i\in\Sigma$, we have convergence in distribution:*

$$Y_n^i \xrightarrow{d} \mathcal{N}(0,1) \qquad (n\to\infty). \tag{6.20}$$

We now present a brief streamlined road map of the proof.

Step 1. Normalization. From the system (3.5)–(3.6), where we denote there $I_n^0 := I_n$ and $I_n^1 := J_n$, and the normalization (6.19) we obtain for all $n\geq 2$,

$$Y_n^i \stackrel{d}{=} \frac{\sigma_i(I_n^i)}{\sigma_i(n)} Y^i_{I_n^i} + \frac{\sigma_{1-i}(n-I_n^i)}{\sigma_i(n)} Y^{1-i}_{n-I_n^i} + b_i(n), \tag{6.21}$$

where

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

and in (6.21) 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. It can be shown by our expansions of the means $\nu_i(n)$ and the Lipschitz property from Proposition 4.2 that we have $b_i(n)\to 0$ as $n\to\infty$ for both $i\in\Sigma$, e.g., in the L_3-norm which below will be technically sufficient. Furthermore, the asymptotic of the variance from Theorem 5.1 implies together with the strong law of large numbers that the coefficients in (6.21) converge:

$$\frac{\sigma_i(I_n^i)}{\sigma_i(n)} \to \sqrt{p_{ii}}, \qquad \frac{\sigma_{1-i}(n-I_n^i)}{\sigma_i(n)} \to \sqrt{1-p_{ii}},$$

where we recall that $\sigma_i(I_n^i)$ is the standard deviation of $L^i_{I_n^i}$ conditioned on I_n^i, hence, in particular a random variable.

Step 2. System of limit equations. The convergence of the coefficients in (6.21) suggests, by passing formally with $n\to\infty$, that limits Y^0 and Y^1 of Y_n^0 and Y_n^1, if they exist, should satisfy the system of recursive distributional equations

$$Y^0 \stackrel{d}{=} \sqrt{p_{00}}Y^0 + \sqrt{1-p_{00}}Y^1, \tag{6.22}$$

$$Y^1 \stackrel{d}{=} \sqrt{1-p_{11}}Y^0 + \sqrt{p_{11}}Y^1, \tag{6.23}$$

where Y^0 and Y^1 are being independent on the right hand sides. Clearly, centered normally distributed

Y^0 and Y^1 with identical variances solve the system (6.22)–(6.23). The task now is to show that Y_n^0 and Y_n^1 converge in distribution towards these solutions Y^0 and Y^1 respectively.

Step 3. The operator of distributions. Our approach is based on the system (6.22)–(6.23) of limit equations together with an associated contracting operator (map) on the space of probability distributions as follows: We denote by $\mathcal{M}_s(0,1)$ the space of all probability distributions on the real line with mean 0, variance 1 and finite absolute moment of order s. Later $2 < s \leq 3$ will be an appropriate choice for us. With the abbreviation $\mathcal{M}^2 := \mathcal{M}_s(0,1) \times \mathcal{M}_s(0,1)$ we define the map

$$\begin{aligned} T &: \mathcal{M}^2 \to \mathcal{M}^2 \\ (\tau_0, \tau_1) &\mapsto \Big(\mathcal{L}\left(\sqrt{p_{00}}W^0 + \sqrt{1-p_{00}}W^1\right), \\ &\qquad \mathcal{L}\left(\sqrt{1-p_{11}}W^0 + \sqrt{p_{11}}W^1\right)\Big), \end{aligned}$$

where W^0, W^1 are independent with distributions $\mathcal{L}(W^i) = \tau_i$ for both $i \in \Sigma$.

This allows a measure theoretic reformulation of solutions of (6.22)–(6.23) that is convenient subsequently: Random variables (Y^0, Y^1) solve the system (6.22)–(6.23) if and only if their pair of distributions $(\mathcal{L}(Y^0), \mathcal{L}(Y^1))$ is a fixed point of T. Hence the identification of fixed-points and domains of attraction of such fixed-points plays an important role in the asymptotic behavior of our sequences $(Y_n^0)_{n\geq 0}$ and $(Y_n^1)_{n \geq 0}$ and is a core part of our proof.

Step 4. The Zolotarev metric. In accordance with the general idea of the contraction method we will endow the space $\mathcal{M}^2$ with a complete metric such that T becomes a contraction with respect to this metric. The issue of fixed-points is then reduced to the application of Banach's fixed-point theorem.

As building block we use the *Zolotarev metric* on $\mathcal{M}_s(0,1)$. It has been studied in the context of the contraction method systematically in [29]. We only need the following properties, see Zolotarev [41, 42]: For distributions $\mathcal{L}(X)$, $\mathcal{L}(Y)$ on $\mathbb{R}$ the Zolotarev distance ζ_s, $s > 0$, is defined by

$$\begin{aligned} \zeta_s(X,Y) &:= \zeta_s(\mathcal{L}(X), \mathcal{L}(Y)) \\ &:= \sup_{f \in \mathcal{F}_s} |\mathbb{E}[f(X) - f(Y)]| \end{aligned} \tag{6.24}$$

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

$$\mathcal{F}_s := \{f \in C^m : \|f^{(m)}(x) - f^{(m)}(y)\| \leq \|x-y\|^\alpha\},$$

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 $1, \ldots, m$ 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 $2 < s \leq 3$ 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. Convergence in ζ_s implies weak convergence on $\mathbb{R}$. Furthermore, ζ_s is $(s,+)$ ideal, i.e., we have

$$\begin{aligned} \zeta_s(X+Z, Y+Z) &\leq \zeta_s(X,Y), \\ \zeta_s(cX, cY) &= c^s \zeta_s(X,Y) \end{aligned}$$

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

Now, to measure distances on the product space $\mathcal{M}^2$ we define for $(\tau_0,\tau_1), (\varrho_0, \varrho_1) \in \mathcal{M}^2$ the distance

$$\zeta_s^\vee((\tau_0,\tau_1),(\varrho_0,\varrho_1)) := \zeta_s(\tau_0,\varrho_0) \vee \zeta_s(\tau_1, \varrho_1).$$

Here and later on, we use the symbols $\vee$ and $\wedge$ for max and min respectivly.

Step 5. The contraction property. We directly obtain that T is a contraction in $\zeta_s^\vee$ from the property that ζ_s is $(s,+)$ ideal: Denoting the components of T by T_0 and T_1 we have

$$\begin{aligned} &\zeta_s(T_0(\tau_0,\tau_1), T_0(\varrho_0,\varrho_1)) \\ &\leq p_{00}^{s/2}\zeta_s(\tau_0,\varrho_0) + (1-p_{00})^{s/2}\zeta_s(\tau_1,\varrho_1) \\ &\leq \left(p_{00}^{s/2} + (1-p_{00})^{s/2}\right)\zeta_s^\vee((\tau_0,\tau_1),(\varrho_0,\varrho_1)), \end{aligned}$$

and similary

$$\begin{aligned} &\zeta_s(T_1(\tau_0,\tau_1), T_1(\varrho_0,\varrho_1)) \\ &\leq (1-p_{11})^{s/2}\zeta_s(\tau_0,\varrho_0) + p_{11}^{s/2}\zeta_s(\tau_1,\varrho_1) \\ &\leq \left((1-p_{11})^{s/2} + p_{11}^{s/2}\right)\zeta_s^\vee((\tau_0,\tau_1),(\varrho_0,\varrho_1)). \end{aligned}$$

Hence together with $\xi := \max_{i \in \Sigma}(p_{ii}^{s/2} + (1-p_{ii})^{s/2})$ we obtain that

$$\zeta_s^\vee(T(\tau_0,\tau_1), T(\varrho_0,\varrho_1)) \leq \xi \zeta_s^\vee((\tau_0,\tau_1),(\varrho_0,\varrho_1)). \tag{6.25}$$

Since $p_{ii} \in (0,1)$ by assumption (2.2) we have $\xi < 1$ for all $s > 2$. On the other hand, it is known that one only obtains finiteness of ζ_s on $\mathcal{M}_s(0,1)$ for $s \leq 3$, hence (6.25) is only meaningful for $s \leq 3$. Thus, altogether, our choice of s is $2 < s \leq 3$. For these s we obtain that T is a contraction in $\zeta_s^\vee$.

Step 6. Convergence of the $\mathbf{Y_n^i}$. An intuition why contraction properties of the map T lead to convergence of the Y_n^i towards the unique fixed-point

$(\mathcal{N}(0,1),\mathcal{N}(0,1))$ of T in $\mathcal{M}^2$ is as follows: The map T serves as a limit version of our recurrence system (6.21). Since in this recurrence system we could replace the $Y^i_{I^i_n}$ and $Y^{1-i}_{n-I^i_n}$ on the right hand side by the recurrence (6.21) itself, iterating these replacements leads approximatively to an iteration of the map T. However, by Banach's fixed-point theorem, the iteration of T applied to any starting point in $\mathcal{M}^2$ converges to the unique fixed-point of T in the metric $\zeta_s^\vee$.

Hence, the problem of proving the convergence of the Y^i_n to the standard normal distribution (the fixed-point) is reduced to the following technical task: Verify that not only the iterations of T itself convergence in the metric $\zeta_s^\vee$ to the fixed-point, but also that the iterations of the approximations of T that make the recurrence of the Y^i_n convergence within $\zeta_s^\vee$.

Once this is settled, we use that convergence in ζ_s is strong enough to imply weak convergence and $(\mathcal{N}(0,1),\mathcal{N}(0,1))$ is the unique fixed point of T. This finally yields Proposition 6.2. A detailed proof is given in the full paper version of this extended abstract.

Step 7. Transfer to arbitrary initial distributions. Finally, we prove Theorem 6.1. For this, we have to transfer the convergence of the Y^i_n from Proposition 6.2 to the convergence of the normalization of L^μ_n via (3.4). Recall that in (3.4), the K_n is a binomial $B(n,\mu_0)$ distributed random variable. We write

$$\frac{L^\mu_n - \mathbb{E}[L^\mu_n]}{\sqrt{n\log n}} = \frac{L^\mu_n - \nu_0(K_n) - \nu_1(n-K_n)}{\sqrt{n\log n}} + \frac{\nu_0(K_n) + \nu_1(n-K_n) - \mathbb{E}[L^\mu_n]}{\sqrt{n\log n}}.$$

By the Lemma of Slutsky, see, e.g. [1, Theorem 3.1], it is sufficient to show, as $n\to\infty$,

$$\frac{L^\mu_n - \nu_0(K_n) - \nu_1(n-K_n)}{\sqrt{n\log n}} \xrightarrow{d} \mathcal{N}(0,\sigma^2) \tag{6.26}$$

$$\frac{\nu_0(K_n) + \nu_1(n-K_n) - \mathbb{E}[L^\mu_n]}{\sqrt{n\log n}} \xrightarrow{\mathbb{P}} 0. \tag{6.27}$$

For showing (6.26) note that by Proposition 6.2 $(L^i_n - \mathbb{E}[L^i_n])/\sqrt{n\log n} \to \mathcal{N}(0,\sigma^2)$ 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 with $\kappa_{nj} := \mathbb{P}(K_n = j)$

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

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 $(L^0_j - \nu_0(j))/\sqrt{n\log n} \to \mathcal{N}(0,\mu_0\sigma^2)$ and $(L^1_{n-j} - \nu_1(n-j))/\sqrt{n\log n} \to \mathcal{N}(0,(1-\mu_0)\sigma^2)$ 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

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

where the latter convergence is justified by dominated convergence. This shows (6.26).

To establish the convergence in probability in (6.27) note that (3.4) implies

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

Hence, with the notation (4.7) and $g(x) := x\log x$ for $x\in[0,1]$ and $\|\cdot\|_1$ denoting the L_1-norm we have

$$\begin{aligned}
&\frac{1}{\sqrt{n\log n}}\|\nu_0(K_n) + \nu_1(n-K_n) - \mathbb{E}[L^\mu_n]\|_1\\
&\quad= \frac{1}{\sqrt{n\log n}}\|\nu_0(K_n) - \mathbb{E}[\nu_0(K_n)]\\
&\qquad\qquad + \nu_1(n-K_n) - \mathbb{E}[\nu_1(n-K_n)]\|_1\\
&\quad\le \frac{1}{H\sqrt{n\log n}}\|g(K_n) - \mathbb{E}[g(K_n)]\\
&\qquad\qquad + g(n-K_n) - \mathbb{E}[g(n-K_n)]\|_1\\
&\qquad + \frac{1}{\sqrt{n\log n}}\|f_0(K_n) - \mathbb{E}[f_0(K_n)]\|_1\\
&\qquad + \frac{1}{\sqrt{n\log n}}\|f_1(n-K_n) - \mathbb{E}[f_1(n-K_n)]\|_1.
\end{aligned}$$

With the concentration of the binomial distribution we obtain

$$\begin{aligned}
&\|g(K_n) - \mathbb{E}[g(K_n)] + g(n-K_n) - \mathbb{E}[g(n-K_n)]\|_1 \\
&\quad = n \left\| g\left(\frac{K_n}{n}\right) - \mathbb{E}\left[g\left(\frac{K_n}{n}\right)\right] \right. \\
&\qquad \left. + g\left(\frac{n-K_n}{n}\right) - \mathbb{E}\left[g\left(\frac{n-K_n}{n}\right)\right]\right\|_1 \\
&\quad = \mathrm{O}\left(n^{1/2}\right).
\end{aligned}$$

The terms $\|f_0(K_n) - \mathbb{E}[f_0(K_n)]\|_1$ and $\|f_1(n-K_n) - \mathbb{E}[f_1(n-K_n)]\|_1$ are also of the order $\mathrm{O}(n^{1/2})$ by a self-centering argument. Altogether we have

$$\frac{\|\nu_0(K_n) + \nu_1(n-K_n) - \mathbb{E}[L_n^\mu]\|_1}{\sqrt{n \log n}} = \mathrm{O}\left(\frac{1}{\sqrt{\log n}}\right),$$

which, by Markov's inequality, implies (6.27) as follows: For any $\varepsilon > 0$ we have

$$\begin{aligned}
&\mathbb{P}\left(\left|\frac{\nu_0(K_n) + \nu_1(n-K_n) - \mathbb{E}[L_n^\mu]}{\sqrt{n \log n}}\right| > \varepsilon\right) \\
&\leq \frac{1}{\varepsilon}\mathbb{E}\left[\left|\frac{\nu_0(K_n) + \nu_1(n-K_n) - \mathbb{E}[L_n^\mu]}{\sqrt{n \log n}}\right|\right] \\
&= \frac{1}{\varepsilon\sqrt{n \log n}}\|\nu_0(K_n) + \nu_1(n-K_n) - \mathbb{E}[L_n^\mu]\|_1 \\
&\to 0.
\end{aligned}$$

7 Comparison with a multivariate approach

We propose the use of systems of univariate recurrences in this extended abstract. Note however, that known limit theorems from the contraction method for multivariate recurrences can as well be applied to the bivariate random variable $Y_n := (Y_n^0, Y_n^1)$. (Technically easiest is to keep the components Y_n^0 and Y_n^1 independent by working with independent I_n^0 and I_n^1.) Applying such an approach as developed in [29], the system (6.22)–(6.23) is now replaced by the bivariate recursive distributional equation

$$Y \stackrel{d}{=} A_1 Y + A_2 \widehat{Y}, \tag{7.28}$$

where Y and $\widehat{Y}$ are independent and identically distributed bivariate random variables and the matrices A_1, A_2 are give by

$$A_1 := \begin{bmatrix} \sqrt{p_{00}} & 0 \\ 0 & \sqrt{p_{11}} \end{bmatrix}, \qquad A_2 := \begin{bmatrix} 0 & \sqrt{1-p_{00}} \\ \sqrt{1-p_{11}} & 0 \end{bmatrix}.$$

Any centered bivariate normal distribution solves the latter fixed-point equation (7.28). In particular Theorem 4.1 in [29] covers the arising bivariate recurrence, cf. also condition (38) in [29], which is satisfied for A_1, A_2 in (7.28)

However, for applying the contraction method in such a multivariate form, an underlying contraction is only implied for, see condition (25) in [29],

$$\|A_1\|_{\mathrm{op}}^3 + \|A_2\|_{\mathrm{op}}^3 < 1,$$

where $\|\cdot\|_{\mathrm{op}}$, here, is identical to the spectral radius of the matrix. This imposes the additional condition

$$(p_{00} \vee p_{11})^{3/2} + (1 - p_{00} \wedge p_{11})^{3/2} < 1 \tag{7.29}$$

to come up with a result similar to our Theorem 6.1.

Our new approach based on systems of univariate recursive equations given above does not require any further condition such as (7.29).

References

[1] Billingsley, P. (1999) Convergence of probability measures. Second edition. Wiley Series in Probability and Statistics: Probability and Statistics. A Wiley-Interscience Publication. *John Wiley & Sons, Inc., New York.*

[2] 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.

[3] Vallée, B., Clément, J., Fill, J.A. and Flajolet, P. (2009) The number of symbol comparisons in QuickSort and QuickSelect. *Automata, languages and programming. Part I*, 750–763, Lecture Notes in Comput. Sci., 5555, *Springer, Berlin.*

[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. (1982) A note on the average depth of tries. *Computing* **28**, 367–371.

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

[7] Devroye, L. (1992) A study of trie-like structures under the density model. *Ann. Appl. Probab.* **2**, 402–434.

[8] Devroye, L. (2002) Laws of large numbers and tail inequalities for random tries and Patricia trees. *J. Comput. Appl. Math.* **142**, 27–37.

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

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

[11] Fill, J.A. and Kapur, N. (2004) The Space Requirement of m-ary Search Trees: Distributional Asymptotics for $m \geq 27$. Invited paper, Proceedings of the 7th Iranian Statistical Conference, 2004. Available via `http://www.ams.jhu.edu/~fill/papers/periodic.pdf`

[12] 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.

[13] Gusfield, D. (1997) *Algorithms on Strings, Trees, and Sequences*, Cambridge University Press, Cambridge.

[14] Jacquet, Ph. and Régnier, M. (1988) Normal limiting distribution of the size and the external path length of tries. Technical Report RR-0827, INRIA-Rocquencourt.

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

[16] Jacquet, Ph. and Szpankowski, W. (1989) Analysis of Digital Tries with Markovian Dependency. *Computer Science Technical Reports.* Report 89-906, Purdue University. Available via `http://docs.lib.purdue.edu/cstech/772`

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

[18] 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.

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

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

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

[22] Kirschenhofer, P. and Prodinger, H. (1988) Further results on digital search trees. Thirteenth International Colloquium on Automata, Languages and Programming (Rennes, 1986). *Theoret. Comput. Sci.* **58**, 143–154.

[23] 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.

[24] Kirschenhofer, P., Prodinger, H. and Szpankowski, W. (1994) Digital search trees again revisited: the internal path length perspective. *SIAM J. Comput.* **23**, 598–616.

[25] 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.

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

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

[28] 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.

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

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

[31] Neininger, R. and Sulzbach, H. (2012) On a functional contraction method. Preprint available via `http://arxiv.org/abs/1202.1370`

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

[33] Rais, B., Jacquet, P. und Szpankowski, W. (1993) Limiting distribution for the depth in PATRICIA tries. *SIAM J. Discrete Math.* **6**, 197–213.

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

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

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

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

[38] 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.

[39] Szpankowski, W. (1991) A characterization of digital search trees from the successful search viewpoint. *Theoret. Comput. Sci.* **85**, 117–134.

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

[41] 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.

[42] 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.