RENEWAL THEORY IN ANALYSIS OF TRIES AND STRINGS

SVANTE JANSON

To my colleague and friend Allan Gut on the occasion of his retirement

Abstract. We give a survey of a number of simple applications of renewal theory to problems on random strings and tries: insertion depth, size, insertion mode and imbalance of tries; variations for b-tries and Patricia tries; Khodak and Tunstall codes.

1. Introduction

Although it long has been realized that renewal theory is a useful tool in the study of random strings and related structures, it has not always been used to its full potential. The purpose of the present paper is to give a survey presenting in a unified way some simple applications of renewal theory to a number of problems involving random strings, in particular several problems on tries, which are tree structures constructed from strings. (Other applications of renewal theory to problems on random trees are given in, e.g., [4], [8], [9], [17], [22], [32], [33].)

Since our purpose is to illustrate a method rather than to prove new results, we present a number of problems in a simple form without trying to be as general as possible. In particular, for simplicity we exclusively consider random strings in the alphabet $\{0, 1\}$, and assume that the "letters" (bits) ξ_i in the strings are i.i.d. (i.e., memoryless sources). Note, however, that the methods below are much more widely applicable and extend in a straightforward way to larger alphabets. The methods also, extend to, for example, Markov sources where ξ_i is a Markov chain; see e.g. Savari and Gallager [39] for a pioneering study of Tunstall codes for Markov sources using renewal theory, and Savari [38] for some related coding problems. (See further e.g. Szpankowski [41, Section 2.1] and Clément, Flajolet and Vallée [7] for various interesting probability models of random strings. Renewal theory for Markov chains is treated for example by Kesten [25] and Athreya, McDonald and Ney [2].) Indeed, one of the purposes of this paper is to make propaganda for the use of renewal theory to study e.g. Markov models, even if we do not do this in the present paper.

Date: 11 December, 2009; revised 21 September, 2011.
2000 *Mathematics Subject Classification.* 68P05; 60K10.

The results below are (mostly) not new; they have earlier been proved by other methods, in particular Mellin transforms. (We try to give proper references for the theorems, but we do not attempt to cover the large literature on random tries and strings in any completeness.) Indeed, such methods often provide sharper results, with better error bounds or higher order terms, and these methods too certainly are important. Nevertheless, we believe that renewal theory often is a valuable method that yields the leading terms in a simple and intuitive way, and that it ought to be more widely used for this type of problems. Moreover, as said above, this method may be easier to extend to other situations. (Further, it gives one explanation for the oscillatory terms that often appear, as an instance of the arithmetic case in renewal theory. Note that oscillatory terms become much less common for larger alphabets, except when all letters are equiprobable, because it is more difficult to be arithmetic, see Appendix A.)

We treat a number of problems on random tries in Sections 3–5 and 8 (insertion depth, imbalance, size, insertion mode). We consider b-tries in Section 6 and Patricia tries in Section 7. Tunstall and Khodak codes are studied in Section 9. A random walk in a region bounded by two crossing lines is studied in Section 10, where we give a (partial) extension of a result by Drmota and Szpankowski [12]. The standard results from renewal theory that we use are for convenience collected in Appendix A.

Notation. We use $\xrightarrow{\mathrm{p}}$ and $\xrightarrow{\mathrm{d}}$ for convergence in probability and in distribution, respectively.

If Z_n is a sequence of random variables and μ_n and σ_n^2 are sequences of real numbers with $\sigma_n^2 > 0$ (for large n, at least), then $Z_n \sim \mathrm{AsN}(\mu_n, \sigma_n^2)$ means that $(Z_n - \mu_n)/\sigma_n \xrightarrow{\mathrm{d}} N(0,1)$.

We denote the fractional part of a real number x by $\{x\} := x - \lfloor x \rfloor$.

Acknowledgement. I thank Allan Gut and Wojciech Szpankowski for inspiration and helpful discussions.

2. Preliminaries

Suppose that $\Xi^{(1)}, \Xi^{(2)}, \dots$ is an i.i.d. sequence of random infinite strings $\Xi^{(n)} = \xi_1^{(n)} \xi_2^{(n)} \cdots$, with letters $\xi_i^{(n)}$ in an alphabet $\mathcal{A}$. (When the superscript n does not matter we drop it; we thus write $\Xi = \xi_1 \xi_2 \cdots$ for a generic string in the sequence.) For simplicity, we consider only the case $\mathcal{A} = \{0, 1\}$, and further assume that the individual letters ξ_i are i.i.d. with $\xi_i \sim \mathrm{Be}(p)$ for some fixed $p \in (0,1)$, i.e., $\mathbb{P}(\xi_i = 1) = p$ and $\mathbb{P}(\xi_i = 0) = q := 1 - p$.

Given a finite string $\alpha_1 \cdots \alpha_n \in \mathcal{A}^n$, let $P(\alpha_1 \cdots \alpha_n)$ be the probability that the random string Ξ begins with $\alpha_1 \cdots \alpha_n$. In particular, for a single letter, $P(0) = q$ and $P(1) = p$, and in general

$$P(\alpha_1 \cdots \alpha_n) = \prod_{i=1}^{n} P(\alpha_i) = \prod_{i=1}^{n} p^{\alpha_i} q^{1-\alpha_i}. \tag{2.1}$$

Given a random string $\xi_1\xi_2\cdots$, we define

$$X_i := -\ln P(\xi_i) = -\ln\bigl(p^{\xi_i}q^{1-\xi_i}\bigr) = \begin{cases} -\ln q, & \xi_i = 0, \\ -\ln p, & \xi_i = 1. \end{cases} \tag{2.2}$$

Note that $X_1, X_2, \dots$ is an i.i.d. sequence of positive random variables with

$$\mathbb{E}\, X_i = H := -p\ln p - q\ln q, \tag{2.3}$$

the usual *entropy* of each letter ξ_i, and

$$\mathbb{E}\, X_i^2 = H_2 := p\ln^2 p + q\ln^2 q, \tag{2.4}$$

$$\operatorname{Var} X_i = H_2 - H^2 = pq(\ln p - \ln q)^2 = pq\ln^2(p/q). \tag{2.5}$$

The variance (2.5) is in data compression known as the *minimal coding variance*, see [27].

Note that the case $p = q = 1/2$ is special; in this case $X_i = \ln 2$ is deterministic and $\operatorname{Var} X_i = 0$; for all other $p \in (0,1)$, $0 < \operatorname{Var} X_i < \infty$.

By (2.2), X_i is supported on $\{\ln(1/p), \ln(1/q)\}$. It is well-known, both in renewal theory and in the analysis of tries, that one frequently has to distinguish between two cases: the *arithmetic* (or *lattice*) case when the support is a subset of $d\mathbb{Z}$ for some $d > 0$, and the *non-arithmetic* (or *non-lattice*) case when it is not, see further Appendix A. For X_i given by (2.2), this yields the following cases:

arithmetic: The ratio $\ln p/\ln q$ is rational. More precisely, X_i then is d-arithmetic, where d equals $\gcd(\ln p, \ln q)$, the largest positive real number such that $\ln p$ and $\ln q$ both are integer multiples of d. If $\ln p/\ln q = a/b$, where a and b are relatively prime positive integers, then

$$d = \gcd(\ln p, \ln q) = \frac{|\ln p|}{a} = \frac{|\ln q|}{b}. \tag{2.6}$$

non-arithmetic: The ratio $\ln p/\ln q$ is irrational.

We let S_n denote the partial sums of X_i: $S_n := \sum_{i=1}^n X_i$. Thus

$$P(\xi_1\cdots\xi_n) = \prod_{i=1}^n P(\xi_i) = \prod_{i=1}^n e^{-X_i} = e^{-S_n}. \tag{2.7}$$

(This is a random variable, since it depends on the random string $\xi_1\cdots\xi_n$; it can be interpreted as the probability that another random string $\Xi^{(j)}$ begins with the same n letters as observed.)

We introduce the standard renewal theory notations (see e.g. Gut [16, Chapter 2]), for $t \geq 0$ and $n \geq 1$,

$$\nu(t) := \min\{n : S_n > t\}, \tag{2.8}$$

$$F_n(t) := \mathbb{P}(S_n \leq t) = \mathbb{P}(\nu(t) > n), \tag{2.9}$$

$$U(t) := \mathbb{E}\,\nu(t) = \sum_{n=0}^\infty F_n(t). \tag{2.10}$$

Note that (2.10) means that, for any function $g \geq 0$,

$$\int_0^\infty g(t)\,\mathrm{d}U(t) = \sum_{n=0}^\infty \int_0^\infty g(t)\,\mathrm{d}F_n(t) = \sum_{n=0}^\infty \mathbb{E}\, g(S_n). \tag{2.11}$$

We also allow the summation to start with an initial random variable X_0, which is independent of $X_1, X_2, \dots$, but may have an arbitrary real-valued distribution. We then define

$$\widehat{S}_n := \sum_{i=0}^n X_i = X_0 + \sum_{i=1}^n X_i, \tag{2.12}$$

$$\widehat{\nu}(t) := \min\{n : \widehat{S}_n > t\}. \tag{2.13}$$

3. Insertion depth in a trie

A *trie* is a binary tree structure designed to store a set of strings. It is constructed from the strings by the following recursive procedure, see further e.g. Knuth [26, Section 6.3], Mahmoud [30, Chapter 5] or Szpankowski [41, Section 1.1]: If the set of strings is empty, then the trie is empty; if there is only one string, then the trie consists of a single node (the root), and the string is stored there; if there is more than one string, then the trie begins with a root, without any string stored, all strings that begin with 0 are passed to the left subtree of the root, and all strings that begin with 1 are passed to the right subtree. In the latter case, the subtrees are constructed recursively by the same procedure, with the only difference that at the kth level, the strings are partitioned according to the kth letter. We assume that the strings are distinct (in our random model, this holds with probability 1), and then the procedure terminates. Note that one string is stored in each leaf of the trie, and that no strings are stored in the remaining nodes. The leaves are also called *external nodes* and the remaining nodes are called *internal nodes*; note that every internal node has one or two children.

The trie is a finite subtree of the complete infinite binary tree $\mathcal{T}_\infty$, where the nodes can be labelled by finite strings $\boldsymbol{\alpha} = \alpha_1 \cdots \alpha_k \in \mathcal{A}^* := \bigcup_{k=0}^\infty \mathcal{A}^k$ (the root is the empty string). It is easily seen that a node $\alpha_1 \cdots \alpha_k$ in $\mathcal{T}_\infty$ is an internal node of the trie if and only if there are at least 2 strings (in the given set) that start with $\alpha_1 \cdots \alpha_k$, and (for $k \geq 1$) that $\alpha_1 \cdots \alpha_k$ is an external node if and only if there is exactly one such string, and there is at least one other string beginning with $\alpha_1 \cdots \alpha_{k-1}$.

Let D_n be the depth (= path length) of the node containing a given string, for example the first, in the trie constructed from n random strings $\Xi^{(1)}, \dots, \Xi^{(n)}$. (By symmetry, any of the n strings will have a depth with the same distribution.) Denoting the chosen string by $\Xi = \xi_1\xi_2\cdots$, the depth D_n is thus at most k if and only if no other of the strings begins with $\xi_1 \cdots \xi_k$. Conditioning on the string Ξ, each of the other strings has this beginning with probability $P(\xi_1 \cdots \xi_k)$, and thus by independence, recalling

(2.7),

$$\mathbb{P}(D_n \le k \mid \Xi) = \bigl(1 - P(\xi_1 \cdots \xi_k)\bigr)^{n-1} = \bigl(1 - e^{-S_k}\bigr)^{n-1}. \tag{3.1}$$

Let $X_0 = X_0^{(n)}$ be a random variable, independent of Ξ, with the distribution

$$\mathbb{P}(X_0^{(n)} > x) = \bigl(1 - e^x/n\bigr)_+^{n-1} = \bigl(1 - e^{x-\ln n}\bigr)_+^{n-1}, \qquad x \in (-\infty, \infty). \tag{3.2}$$

As $n \to \infty$, this converges to $\exp(-e^x)$, and thus $X_0^{(n)} \xrightarrow{\text{d}} X_0^*$, where $-X_0^*$ has the Gumbel distribution with $\mathbb{P}(-X_0^* \le x) = \exp(-\exp(-x))$.

Remark 3.1. It is easily seen that $X_0^{(n)} \overset{\text{d}}{=} \ln n - \max\{Z_1, \dots, Z_{n-1}\}$, where $Z_1, Z_2, \dots$ are i.i.d. Exp(1) random variables. Cf. Leadbetter, Lindgren and Rootzén [28, Example 1.7.2].

Using (3.2), we can rewrite (3.1) as

$$\mathbb{P}(D_n \le k \mid \Xi) = \mathbb{P}\bigl(X_0^{(n)} > \ln n - S_k \mid \Xi\bigr) \tag{3.3}$$

and thus, recalling (2.12) and (2.13),

$$\mathbb{P}(D_n \le k) = \mathbb{P}\bigl(X_0 > \ln n - S_k\bigr) = \mathbb{P}\bigl(\widehat{S}_k > \ln n\bigr) = \mathbb{P}\bigl(\widehat{\nu}(\ln n) \le k\bigr). \tag{3.4}$$

Since $k \ge 1$ is arbitrary, this shows that

$$D_n \overset{\text{d}}{=} \widehat{\nu}(\ln n). \tag{3.5}$$

In the case $p = 1/2$, $S_k = k \ln 2$ is non-random, and the only randomness in $\widehat{\nu}(\ln n)$ comes from X_0; in fact, it is easy to see that $\mathbb{P}(D_n \le k) \to \mathbb{P}(-X_0^* \le t)$ if $k \to \infty$ and $n \to \infty$ along sequences such that $k \ln 2 - \ln n \to t \in (-\infty, \infty)$, see [18], [35], [30, Theorem 5.7], [29]. This result can also be expressed as $d_{\text{TV}}(D_n, \lceil (\ln n - X_0^*)/\ln 2 \rceil) \to 0$ as $n \to \infty$, where d_{TV} denotes the total variation distance of the distributions, see [23, Example 4.5].

However, if $p \ne 1/2$, then each X_k is truly random, which leads to larger dispersion of D_n. We can apply standard renewal theory theorems, see Theorems A.1–A.4 in the appendix, and immediately obtain the following. For other, earlier proofs see Knuth [26, Sections 6.3 and 5.2], Pittel [34, 35] and Mahmoud [30, Section 5.5]. The Markov case is treated by Jacquet and Szpankowski [21], ergodic strings by Pittel [34], and a class of general dynamical sources by Clément, Flajolet and Vallée [7].

Theorem 3.2. *For every $p \in (0,1)$,*

$$\frac{D_n}{\ln n} \xrightarrow{\text{p}} \frac{1}{H}, \tag{3.6}$$

with H the entropy given by (2.3)*. Moreover, the convergence holds in every L^r, $r < \infty$, too. Hence, all moments converge in* (3.6) *and*

$$\mathbb{E}\, D_n^r \sim H^{-r} (\ln n)^r, \qquad 0 < r < \infty. \tag{3.7}$$

Theorem 3.3. *More precisely:*

(i) *If* $\ln p/\ln q$ *is irrational, then, as* $n\to\infty$,

$$\mathbb{E}\, D_n = \frac{\ln n}{H} + \frac{H_2}{2H^2} + \frac{\gamma}{H} + o(1). \tag{3.8}$$

(ii) *If* $\ln p/\ln q$ *is rational, then, as* $n\to\infty$,

$$\mathbb{E}\, D_n = \frac{\ln n}{H} + \frac{H_2}{2H^2} + \frac{\gamma}{H} + \psi_1(\ln n) + o(1), \tag{3.9}$$

where $\psi_1(t)$ *is a small continuous function, with period* $d = \gcd(\ln p, \ln q)$ *in* t, *given by*

$$\psi_1(t) := -\frac{1}{H}\sum_{k\neq 0}\Gamma(-2\pi \mathrm{i}k/d)e^{2\pi \mathrm{i}kt/d}. \tag{3.10}$$

Proof. The non-arithmetic case (3.8) follows directly from (3.5) and (A.4); we can replace $X_0^{(n)}$ by the limit X_0^*, and since the Gumbel variable $-X_0^*$ has characteristic function $\mathbb{E}\, e^{-\mathrm{i}tX_0^*} = \Gamma(1-\mathrm{i}t)$, we have $\mathbb{E}\, X_0^* = \Gamma'(1) = -\gamma$.

In the arithmetic case, we use (A.6), together with Lemma A.5 which yields

$$\mathbb{E}\Bigl\{\frac{t}{d} - \frac{X_0^*}{d}\Bigr\} = \frac12 - \sum_{k\neq 0}\frac{\Gamma(1-2\pi\mathrm{i}k/d)}{2\pi k\mathrm{i}}e^{2\pi\mathrm{i}kt/d} = \frac12 + \frac1d\sum_{k\neq 0}\Gamma(-2\pi\mathrm{i}k/d)e^{2\pi\mathrm{i}kt/d}.$$

∎

Theorem 3.4. *Suppose that* $p\in(0,1)$. *Then, as* $n\to\infty$,

$$\frac{D_n - H^{-1}\ln n}{\sqrt{\ln n}} \overset{\mathrm{d}}{\longrightarrow} N\Bigl(0, \frac{\sigma^2}{H^3}\Bigr), \tag{3.11}$$

with $\sigma^2 = H_2 - H^2 = pq(\ln p - \ln q)^2$. *If* $p\neq 1/2$, *then* $\sigma^2>0$ *and this can be written as*

$$D_n \sim \mathrm{AsN}\bigl(H^{-1}\ln n, H^{-3}\sigma^2\ln n\bigr). \tag{3.12}$$

Moreover,

$$\operatorname{Var} D_n = \frac{\sigma^2}{H^3}\ln n + o(\ln n). \tag{3.13}$$

Remark 3.5. In the argument above, X_0 depends on n. This is a nuisance, although no real problem, see Theorem A.4. An alternative that avoids this problem is to Poissonize by considering a random number of strings. In this case it is simplest to consider $1+\mathrm{Po}(\lambda)$ strings, so that a selected string Ξ is compared to a Poisson number $\mathrm{Po}(\lambda)$ of other strings, for a parameter $\lambda\to\infty$. Conditioned on Ξ, the number of other strings beginning with $\xi_1\cdots\xi_k$ then has the Poisson distribution $\mathrm{Po}(\lambda P(\xi_1\cdots\xi_k))$. Thus we obtain instead of (3.3)–(3.4), now denoting the depth by $\widetilde{D}_\lambda$,

$$\begin{aligned}\mathbb{P}(\widetilde{D}_\lambda\le k\mid\Xi) &= e^{-\lambda P(\xi_1\cdots\xi_k)} = e^{-\lambda e^{-S_k}} = e^{-e^{-(S_k-\ln\lambda)}}\\ &= \mathbb{P}(-X_0^* < S_k - \ln\lambda\mid\Xi) = \mathbb{P}(S_k + X_0^* > \ln\lambda\mid\Xi)\end{aligned}$$

and

$$\mathbb{P}(\widetilde{D}_\lambda \le k) = \mathbb{P}(S_k + X_0^* > \ln\lambda) = \mathbb{P}\bigl(\widehat{\nu}(\ln\lambda) \le k\bigr),$$

where $X_0 := X_0^*$ now is independent of n, and consequently $\widetilde{D}_\lambda \overset{\mathrm{d}}{=} \widehat{\nu}(\ln\lambda)$. We obtain the same asymptotics as for D_n above, directly from Theorems A.1–A.3. It is in this case easy to depoissonize, by noting that D_n is stochastically monotone in n, and derive the results for D_n from the results for $\widetilde{D}_\lambda$ by choosing $\lambda = n \pm n^{2/3}$. More precisely, we find first, by Theorem A.3,

$$\frac{\widetilde{D}_\lambda - H^{-1}\ln\lambda}{\sqrt{\ln\lambda}} \overset{\mathrm{d}}{\longrightarrow} N\Bigl(0, \frac{\sigma^2}{H^3}\Bigr),$$

and thus taking $\lambda = n \pm n^{2/3}$, since $\ln\lambda - \ln n = O(n^{-1/3})$,

$$\frac{\widetilde{D}_{n\pm n^{2/3}} - H^{-1}\ln n}{\sqrt{\ln n}} \overset{\mathrm{d}}{\longrightarrow} N\Bigl(0, \frac{\sigma^2}{H^3}\Bigr).$$

Since $\mathbb{P}\bigl(\mathrm{Po}(n+n^{2/3}) > n\bigr) \to 1$ and $\mathbb{P}\bigl(\mathrm{Po}(n-n^{2/3}) < n\bigr) \to 1$, we can couple the variables such that with probability $1 - o(1)$,

$$\widetilde{D}_{n-n^{2/3}} \le D_n \le \widetilde{D}_{n+n^{2/3}}, \tag{3.14}$$

and (3.11) follows. For (3.13) it then suffices to show uniform square integrability of $(D_n - H^{-1}\ln n)/\sqrt{\ln n}$, which easily follows from the corresponding result for $\widetilde{D}_{n\pm n^{2/3}}$ by conditioning on $\mathrm{Po}(n+n^{2/3}) > n$ and $\mathrm{Po}(n - n^{2/3}) < n$ and sandwiching as in (3.14).

4. Imbalance in tries

Mahmoud [31] studied the imbalance factor of a string in a trie, defined as the number of steps to the right minus the number of steps to the left in the path from the root to the leaf where the string is stored. We define

$$Y_i := 2\xi_i - 1 = \begin{cases} -1, & \xi_1 = 0, \\ +1, & \xi_1 = 1, \end{cases}$$

and denote the corresponding partial sums by $V_k := \sum_{i=1}^k Y_i$. Thus the imbalance factor Δ_n of the string Ξ in a random trie with n strings is V_{D_n}, with D_n as in Section 3 the depth of the string.

It follows immediately from (3.3) that (3.4) holds also conditioned on the sequence $(Y_1, Y_2, \dots)$. As a consequence, for any k and v,

$$\mathbb{P}(D_n = k \mid V_k = v) = \mathbb{P}\bigl(\widehat{\nu}(\ln n) = k \mid V_k = v\bigr),$$

which shows that

$$(D_n, \Delta_n) = (D_n, V_{D_n}) \overset{\mathrm{d}}{=} \bigl(\widehat{\nu}(\ln n), V_{\widehat{\nu}(\ln n)}\bigr).$$

In particular,

$$\Delta_n \overset{\mathrm{d}}{=} V_{\widehat{\nu}(\ln n)}.$$

We may apply Theorem A.8 (and Remark A.9). A simple calculation yields $\operatorname{Var}(\mu_X Y_1 - \mu_Y X_1) = pq(\ln p + \ln q)^2 = pq \ln^2(pq)$, and we obtain the central limit theorem by Mahmoud [30]:

Theorem 4.1. *As $n \to \infty$,*

$$\Delta_n \sim \operatorname{AsN}\left(\frac{p-q}{H} \ln n, \frac{pq\ln^2(pq)}{H^3} \ln n\right).$$

5. The expected size of a trie

A trie built of n strings as in Section 3 has n external nodes, since each external node contains exactly one string. However, the number of internal nodes, W_n, say, is random. We will study its expectation. For simplicity we Poissonize directly and consider a trie constructed from $\operatorname{Po}(\lambda)$ strings; we let $\widetilde{W}_\lambda$ be the number of internal nodes. Results for a given number of strings then follow by comparison as in Remark 3.5. The results below have previously been found by other methods, in particular, more precise asymptotics have been found using Mellin transforms; see Knuth [26], Mahmoud [30], Fayolle, Flajolet, Hofri and Jacquet [13], and, in particular, Jacquet and Régnier [19, 20]. The Markov case is studied by Régnier [37] and dynamical sources by Clément, Flajolet and Vallée [7].

If $\boldsymbol{\alpha} = \alpha_1 \cdots \alpha_k$ is a finite string, let $I(\boldsymbol{\alpha})$ be the indicator of the event that $\boldsymbol{\alpha}$ is an internal node in the trie. We found above that this event occurs if and only if there are at least two strings beginning with $\boldsymbol{\alpha}$. In our Poisson model, the number of strings beginning with $\boldsymbol{\alpha}$ has a Poisson distribution $\operatorname{Po}(\lambda P(\boldsymbol{\alpha}))$, and thus

$$\mathbb{E}\,\widetilde{W}_\lambda = \sum_{\boldsymbol{\alpha}\in\mathcal{A}^*} \mathbb{E}\, I(\boldsymbol{\alpha}) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}^*} \mathbb{P}\big(\operatorname{Po}(\lambda P(\boldsymbol{\alpha})) \ge 2\big) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}^*} f(\lambda P(\boldsymbol{\alpha})), \quad (5.1)$$

where

$$f(x) := \mathbb{P}\big(\operatorname{Po}(x) \ge 2\big) = 1 - (1+x)e^{-x}. \quad (5.2)$$

Sums of the type in (5.1) are often studied using Mellin transform inversion and residue calculus. Renewal theory presents an alternative. As said in the introduction, this opens the way to straightforward generalizations, e.g. to Markov sources, although we consider only memoryless sources in the present paper.

Theorem 5.1. *Suppose that f is a non-negative function on $(0,\infty)$, and that $F(\lambda) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}^*} f(\lambda P(\boldsymbol{\alpha}))$, with $P(\boldsymbol{\alpha})$ given by (2.1). Assume further that f is a.e. continuous and satisfies the estimates*

$$f(x) = O(x^2), \quad 0 < x < 1, \qquad \text{and} \qquad f(x) = O(1), \quad 1 < x < \infty. \quad (5.3)$$

Let $g(t) := e^t f(e^{-t})$.

(i) *If $\ln p / \ln q$ is irrational, then, as $\lambda \to \infty$,*

$$\frac{F(\lambda)}{\lambda} \to \frac{1}{H} \int_{-\infty}^{\infty} g(t)\,\mathrm{d}t = \frac{1}{H} \int_0^\infty f(x) x^{-2}\,\mathrm{d}x. \quad (5.4)$$

(ii) *If* $\ln p/\ln q$ *is rational, then, as* $\lambda \to \infty$,

$$\frac{F(\lambda)}{\lambda} = \frac{1}{H}\psi(\ln\lambda) + o(1), \tag{5.5}$$

where, with $d := \gcd(\ln p, \ln q)$ *given by* (2.6), ψ *is a bounded* d*-periodic function having the Fourier series*

$$\psi(t) \sim \sum_{m=-\infty}^{\infty} \widehat{\psi}(m) e^{2\pi i m t/d} \tag{5.6}$$

with

$$\widehat{\psi}(m) = \widehat{g}(-2\pi m/d) = \int_{-\infty}^{\infty} e^{2\pi i m t/d} g(t)\,\mathrm{d}t = \int_0^{\infty} f(x) x^{-2-2\pi i m/d}\,\mathrm{d}x. \tag{5.7}$$

Furthermore,

$$\psi(t) = d\sum_{k=-\infty}^{\infty} g(kd - t). \tag{5.8}$$

If f *is continuous, then* ψ *is too.*

Proof. If $f_0(\boldsymbol{\alpha})$ is any non-negative function on $\mathcal{A}^*$, then, using (2.7), for each $k \geq 0$,

$$\begin{aligned}\sum_{\alpha_1,\dots,\alpha_k} f_0(\alpha_1\cdots\alpha_k) &= \sum_{\alpha_1,\dots,\alpha_k} \frac{f_0(\alpha_1\cdots\alpha_k)}{P(\alpha_1\cdots\alpha_k)} P(\alpha_1\cdots\alpha_k)\\ &= \mathbb{E}\,\frac{f_0(\xi_1\cdots\xi_k)}{P(\xi_1\cdots\xi_k)} = \mathbb{E}\big(e^{S_k} f_0(\xi_1\cdots\xi_k)\big),\end{aligned}$$

and thus,

$$\sum_{\boldsymbol{\alpha}\in\mathcal{A}^*} f_0(\boldsymbol{\alpha}) = \sum_{k=0}^{\infty} \mathbb{E}\big(e^{S_k} f_0(\xi_1\cdots\xi_k)\big). \tag{5.9}$$

With $f_0(\boldsymbol{\alpha}) = f(\lambda P(\boldsymbol{\alpha}))$, we have $f_0(\xi_1\cdots\xi_k) = f(\lambda e^{-S_k})$ and thus (5.9) yields, recalling (2.10),

$$F(\lambda) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}^*} f(\lambda P(\boldsymbol{\alpha})) = \sum_{k=0}^{\infty} \mathbb{E}\big(e^{S_k} f(\lambda e^{-S_k})\big) = \int_0^{\infty} f(\lambda e^{-x}) e^x\,\mathrm{d}U(x).$$

Define further $f_1(x) := f(x)/x$; thus $g(t) = f_1(e^{-t})$. Then,

$$F(\lambda) = \int_0^{\infty} \lambda f_1(\lambda e^{-x})\,\mathrm{d}U(x) = \lambda \int_0^{\infty} g(x - \ln\lambda)\,\mathrm{d}U(x). \tag{5.10}$$

We can now apply the key renewal theorem, Theorem A.7. The function g is a.e. continuous and it follows from (5.3) that $g(t) \leq Ce^{-|t|}$ for some C; hence g is directly Riemann integrable on $(-\infty,\infty)$ by Lemma A.6. In the non-arithmetic case (i) we obtain (5.4) from (5.10) and (A.11), since $\mu = \mathbb{E}\,X_i = H$ by (2.3) and, with $x = e^{-t}$,

$$\int_{-\infty}^{\infty} g(t)\,\mathrm{d}t = \int_{-\infty}^{\infty} e^t f(e^{-t})\,\mathrm{d}t = \int_0^{\infty} f(x) x^{-2}\,\mathrm{d}x. \tag{5.11}$$

Similarly, the aritmetic case (ii) follows from (A.13) and (A.15)–(A.17) together with the calculation, generalizing (5.11),

$$\widehat{g}(s) = \int_{-\infty}^{\infty} e^{-\mathrm{i}st} g(t)\,\mathrm{d}t = \int_{-\infty}^{\infty} e^{(1-\mathrm{i}s)t} f(e^{-t})\,\mathrm{d}t = \int_{0}^{\infty} f(x) x^{-2+\mathrm{i}s}\,\mathrm{d}x.$$

(This equals the Mellin transform $\widetilde{f}(-1+\mathrm{i}s)$.) □

Remark 5.2. The assumptions on f may be weakened (with the same proof); it suffices that $f(x) = O(x^{1-\delta})$ and $f(x) = O(x^{1+\delta})$ for $x \in (0,\infty)$ and some $\delta > 0$. If f is continuous, it is obviously sufficient that these estimates hold for small and large x, respectively.

Returning to $\widetilde{W}_\lambda$, we obtain the following for the expected number of internal nodes in the Poisson trie.

Theorem 5.3. (i) *If* $\ln p/\ln q$ *is irrational, then, as* $\lambda \to \infty$,

$$\frac{\mathbb{E}\,\widetilde{W}_\lambda}{\lambda} \to \frac{1}{H}. \tag{5.12}$$

(ii) *If* $\ln p/\ln q$ *is rational, then, as* $\lambda \to \infty$,

$$\frac{\mathbb{E}\,\widetilde{W}_\lambda}{\lambda} = \frac{1}{H} + \frac{1}{H}\psi_2(\ln\lambda) + o(1), \tag{5.13}$$

where, with $d = \gcd(\ln p, \ln q)$, ψ_2 *is a continuous* d*-periodic function with average* 0 *and Fourier expansion*

$$\psi_2(t) = \sum_{k\neq 0} \frac{\Gamma(1-2\pi\mathrm{i}k/d)}{1+2\pi\mathrm{i}k/d} e^{2\pi\mathrm{i}kt/d} = \sum_{k\neq 0} \frac{2\pi\mathrm{i}k}{d}\Gamma\Bigl(-1-\frac{2\pi\mathrm{i}k}{d}\Bigr) e^{2\pi\mathrm{i}kt/d}.$$

Proof. We apply Theorem 5.1 to (5.1). It follows from (5.2) that $f'(x) = xe^{-x}$. Thus, by an integration by parts, since $f(x)/x \to 0$ as $x \to 0$ and $x \to \infty$,

$$\int_0^\infty f(x)x^{-2}\,\mathrm{d}x = \int_0^\infty f'(x)x^{-1}\,\mathrm{d}x = \int_0^\infty e^{-x}\,\mathrm{d}x = 1. \tag{5.14}$$

Consequently, (5.12) follows from (5.4).

Similarly, (5.13) follows from (5.5), and the calculation, generalizing (5.14),

$$\begin{aligned}\widehat{g}(s) &= \int_0^\infty f(x)x^{-2+\mathrm{i}s}\,\mathrm{d}x = (1-\mathrm{i}s)^{-1}\int_0^\infty f'(x)x^{-1+\mathrm{i}s}\,\mathrm{d}x \\ &= \frac{\Gamma(1+\mathrm{i}s)}{1-\mathrm{i}s} = -\mathrm{i}s\Gamma(-1+\mathrm{i}s).\end{aligned}$$ □

The case of a fixed number n of strings is easily handled by comparison, and (5.12) and (5.13) imply the corresponding results for W_n:

Theorem 5.4. (i) *If* $\ln p/\ln q$ *is irrational, then, as* $n \to \infty$,

$$\frac{\mathbb{E}\,W_n}{n} \to \frac{1}{H}.$$

(ii) *If* $\ln p/\ln q$ *is rational, then, as* $n\to\infty$*, with* ψ_2 *as in Theorem 5.3,*

$$\frac{\mathbb{E}\,W_n}{n}=\frac{1}{H}+\frac{1}{H}\psi_2(\ln n)+o(1).$$

Proof. $\mathbb{E}\,W_n$ is increasing in n. Thus, first, because $\mathbb{P}(\mathrm{Po}(2n)\geq n)\geq 1/2$, $\mathbb{E}\,\widetilde{W}_{2n}\geq\frac{1}{2}\,\mathbb{E}\,W_n$, and thus $\mathbb{E}\,W_n\leq 2\,\mathbb{E}\,\widetilde{W}_{2n}=O(n)$. Secondly, using this estimate, the standard Chernoff concentration bounds for the Poisson distribution easily implies, with $\lambda_\pm=n\pm n^{2/3}$, say, $\mathbb{E}\,\widetilde{W}_{\lambda_-}+o(n)\leq\mathbb{E}\,W_n\leq\mathbb{E}\,\widetilde{W}_{\lambda_+}+o(n)$. The results then follow from Theorem 5.3. □

Remark 5.5. It is well-known that the periodic function ψ_2 above, as in many similar results, fluctuates very little from its mean. In fact, the largest d is obtained for $p=q=1/2$, when $d=\ln 2$. Since $\Gamma(1+\mathrm{i}s)$ decreases rapidly as $s\to\pm\infty$, the Fourier coefficients of $\psi_2(t)$ are very small; the largest (in absolute value) are $|\widehat{\psi_2}(\pm1)|=|\Gamma(1+2\pi\mathrm{i}/\ln 2)|/|1-2\pi\mathrm{i}/\ln 2|\approx 0.542\cdot 10^{-6}$, so $|\psi_2(\ln n)|$ is at most about 10^{-6}, and the oscillations $\psi_2(\ln n)/H$ of $\mathbb{E}\,W_n/n$ are bounded by $1.6\cdot10^{-6}$. (See for example [30, pp. 23–28].) Other choices of p yield even smaller oscillations.

6. b-TRIES

As a variation, consider a b-trie, where each node can store b strings, for some fixed integer $b\geq 1$; as before, the internal nodes do not contain any string. A finite string $\boldsymbol{\alpha}$ now is an internal node if and only if at least $b+1$ of the strings start with $\boldsymbol{\alpha}$. In the argument above we only have to replace (5.2) by

$$f(x):=\mathbb{P}\big(\mathrm{Po}(x)\geq b+1\big); \tag{6.1}$$

thus $f'(x)=\mathbb{P}\big(\mathrm{Po}(x)=b\big)=x^be^{-x}/b!$ and (5.11) yields, with an integration by parts as in (5.14), $\int_{-\infty}^{\infty}g(t)\,\mathrm{d}t=1/b$. Hence, in the non-arithmetic case when $\ln p/\ln q$ is irrational, the expected number of internal nodes is $\mathbb{E}\,\widetilde{W}^{(b)}_\lambda\sim\lambda/(Hb)$, as found by Jacquet and Régnier [19, 20]. In the arithmetic case, we obtain a periodic function ψ, now with Fourier expansion

$$\psi(t)=\frac{1}{b!}\sum_{k\neq0}\frac{\Gamma(b-2\pi\mathrm{i}k/d)}{1+2\pi\mathrm{i}k/d}e^{2\pi\mathrm{i}kt/d}.$$

We can also analyze the external nodes. Let Z_j be the number of nodes where exactly j strings are stored, $j=1,\dots,b$. A finite string $\boldsymbol{\alpha}$ is one of these nodes if exactly j of the stored strings begin with $\boldsymbol{\alpha}$, and at least $b-j+1$ other strings begin with $\boldsymbol{\alpha}'$, the sibling of $\boldsymbol{\alpha}$ obtained by flipping the last letter. (We assume that there are at least b strings, so we can ignore the root.)

Consider again the Poisson model. In the case when $\boldsymbol{\alpha}$ ends with 1, i.e., $\boldsymbol{\alpha}=\boldsymbol{\beta}1$ for some $\boldsymbol{\beta}$, the probability of this event is, with $x=\lambda P(\boldsymbol{\beta})$, by independence in the Poisson model, $\mathbb{P}\big(\mathrm{Po}(px)=j\big)\,\mathbb{P}\big(\mathrm{Po}(qx)>b-j\big)$. If $\boldsymbol{\alpha}=\boldsymbol{\beta}0$, we similarly have the probability $\mathbb{P}\big(\mathrm{Po}(qx)=j\big)\,\mathbb{P}\big(\mathrm{Po}(px)>b-j\big)$.

Summing over $\beta \in \mathcal{A}^*$, we thus obtain a sum of the type in Theorem 5.1 with f replaced by

$$\begin{aligned}
f_j(x) &= \mathbb{P}\bigl(\mathrm{Po}(px) = j\bigr)\,\mathbb{P}\bigl(\mathrm{Po}(qx) > b-j\bigr) + \mathbb{P}\bigl(\mathrm{Po}(qx) = j\bigr)\,\mathbb{P}\bigl(\mathrm{Po}(px) > b-j\bigr) \\
&= \frac{p^j x^j}{j!}e^{-px}\left(1 - \sum_{k=0}^{b-j}\frac{q^k x^k}{k!}e^{-qx}\right) + \frac{q^j x^j}{j!}e^{-qx}\left(1 - \sum_{k=0}^{b-j}\frac{p^k x^k}{k!}e^{-px}\right) \\
&= \frac{p^j x^j}{j!}e^{-px} + \frac{q^j x^j}{j!}e^{-qx} - \sum_{k=0}^{b-j}\frac{(p^j q^k + q^j p^k)x^{j+k}}{j!\,k!}e^{-x}.
\end{aligned}$$

We let $g_j(t) := e^t f_j(e^{-t})$ and argue as above. The crucial constant in (5.4) and (5.11) is

$$c_j := \int_{-\infty}^{\infty} g_j(t)\,\mathrm{d}t = \int_0^{\infty} f_j(x)x^{-2}\,\mathrm{d}x. \tag{6.2}$$

For $2 \le j \le b$, this gives

$$\begin{aligned}
c_j &= \frac{p^j}{j!}\int_0^\infty x^{j-2}e^{-px}\,\mathrm{d}x + \frac{q^j}{j!}\int_0^\infty x^{j-2}e^{-qx}\,\mathrm{d}x - \sum_{k=0}^{b-j}\frac{p^j q^k + q^j p^k}{j!\,k!}\int_0^\infty x^{j+k-2}e^{-x}\,\mathrm{d}x \\
&= \frac{p}{j!}(j-2)! + \frac{q}{j!}(j-2)! - \sum_{k=0}^{b-j}\frac{(j+k-2)!}{j!\,k!}(p^j q^k + q^j p^k).
\end{aligned} \tag{6.3}$$

For $j = 1$, we treat the terms with $k \ge 1$ in the same way, but the term with $k = 0$ is combined with the first two terms to

$$\begin{aligned}
&\int_0^\infty \bigl(pxe^{-px} + qxe^{-qx} - (p+q)xe^{-x}\bigr)x^{-2}\,\mathrm{d}x \\
&\qquad = p\int_0^\infty x^{-1}\bigl(e^{-px} - e^{-x}\bigr)\,\mathrm{d}x + q\int_0^\infty x^{-1}\bigl(e^{-qx} - e^{-x}\bigr)\,\mathrm{d}x
\end{aligned} \tag{6.4}$$

To evaluate these integrals, we note that if $\Re z > 0$, then

$$\int_0^\infty x^{z-1}\bigl(e^{-px} - e^{-x}\bigr)\,\mathrm{d}x = \bigl(p^{-z} - 1\bigr)\Gamma(z). \tag{6.5}$$

Since $|e^{-px} - e^{-x}| \le |1-p|x$, the left-hand side converges and is an analytic function of z for complex z with $\Re z > -1$; hence (6.5) holds for $\Re z > -1$ by analytic continuation. In particular, taking the limit as $z \to 0$,

$$\int_0^\infty x^{-1}\bigl(e^{-px} - e^{-x}\bigr)\,\mathrm{d}x = \ln(1/p). \tag{6.6}$$

Consequently, we obtain, by (6.3)–(6.6),

$$c_j = \begin{cases} p\ln(1/p) + q\ln(1/q) - \sum_{k=1}^{b-1}\frac{1}{k}(pq^k + qp^k), & j = 1, \\ \frac{1}{j(j-1)} - \sum_{k=0}^{b-j}\frac{(j+k-2)!}{j!\,k!}(p^j q^k + q^j p^k), & 2 \le j \le b. \end{cases} \tag{6.7}$$

Alternatively, using

$$f_j(x) = \frac{p^j x^j}{j!} e^{-px} \sum_{k=b-j+1}^{\infty} \frac{q^k x^k}{k!} e^{-qx} + \frac{q^j x^j}{j!} e^{-qx} \sum_{k=b-j+1}^{\infty} \frac{p^k x^k}{k!} e^{-px}$$
$$= \sum_{k=b-j+1}^{\infty} \frac{(p^j q^k + q^j p^k) x^{j+k}}{j!\,k!} e^{-x},$$

we find

$$c_j = \sum_{k=b-j+1}^{\infty} \frac{(j+k-2)!}{j!\,k!} (p^j q^k + q^j p^k), \qquad 1 \le j \le b. \tag{6.8}$$

More generally (except when $(j,s) = (1,0)$),

$$\widehat{g}_j(s) = \int_0^\infty f_j(x) x^{-2+\mathrm{i}s}\,\mathrm{d}x$$
$$= \frac{\Gamma(j-1+\mathrm{i}s)}{j!}(p^{1-\mathrm{i}s} + q^{1-\mathrm{i}s}) - \sum_{k=0}^{b-j} \frac{\Gamma(j+k-1+\mathrm{i}s)}{j!\,k!}(p^j q^k + q^j p^k). \tag{6.9}$$

If we use the notation Z_{jn} for the trie with a fixed number n of strings and $\widetilde{Z}_{j\lambda}$ for the Poisson model with $\mathrm{Po}(\lambda)$ strings, we obtain as above the following result for the number of external nodes that store j strings.

Theorem 6.1. (i) *If* $\ln p/\ln q$ *is irrational, then, as* $n \to \infty$, *for* $j = 1, \dots, b$,

$$\frac{\mathbb{E}\, Z_{jn}}{n} \to \pi_j := \frac{c_j}{H},$$

with c_j *given by* (6.7)–(6.8).

(ii) *If* $\ln p/\ln q$ *is rational, then, as* $n \to \infty$, *for* $j = 1, \dots, b$,

$$\frac{\mathbb{E}\, Z_{jn}}{n} = \psi_{bj}(\ln n) + o(1),$$

where ψ_{bj} *is a continuous* d*-periodic function, with* d *as in Theorem 5.3;* ψ_{bj} *has average* π_j *and Fourier expansion*

$$\psi_{bj}(t) = H^{-1} \sum_{k=-\infty}^{\infty} \widehat{g}_j(-2\pi\mathrm{i}k/d) e^{2\pi\mathrm{i}kt/d} = \pi_j + H^{-1} \sum_{k\neq 0} \widehat{g}_j(-2\pi\mathrm{i}k/d) e^{2\pi\mathrm{i}kt/d},$$

with $\widehat{g}_j$ *given by* (6.9). *The same results (with* n *replaced by* λ*) hold for* $\widetilde{Z}_{j\lambda}$ *in the Poisson model.*

Proof. As just said, the Poisson case follows from Theorem 5.1, and it remains only to depoissonize. To do this, choose $\lambda = n$, and let $N \sim \mathrm{Po}(n)$ be the number of strings in the Poisson model. We couple the trie with n strings and the Poisson trie with N strings by starting with $\min(n, N)$ common strings. If we add a new string to the trie, it is either stored in an existing leaf or it converts a leaf to an internal node and adds two new leafs (and possibly a chain of further internal nodes). Thus at most 3 leaves are affected, and each Z_j changes by at most 3. Since we add $\max(n,N)-\min(n,N)=|N-n|$ new strings, we have $|\widetilde{Z}_{j\lambda}-Z_{jn}|\le 3|N-n|$ for each j, and thus $|\mathbb{E}\,\widetilde{Z}_{j\lambda}-\mathbb{E}\,Z_{jn}|\le 3\,\mathbb{E}\,|N-n|=O(\sqrt{n})$. □

For example, for $b=2,3,4$ we have the following limits in the non-aritmetic case, and up to small oscillations also in the aritmetic case:

b	π_1	π_2	π_3	π_4
2	$1-\frac{2}{H}pq$	$\frac{1}{H}pq$		
3	$1-\frac{5}{2H}pq$	$\frac{1}{2H}pq$	$\frac{1}{2H}pq$	
4	$1-\frac{17}{6H}pq+\frac{2}{3H}(pq)^2$	$\frac{1}{2H}pq-\frac{1}{H}(pq)^2$	$\frac{1}{6H}pq+\frac{2}{3H}(pq)^2$	$\frac{1}{3H}pq-\frac{1}{6H}(pq)^2$

Note that $\sum_1^b j\pi_j=1$, or equivalently $\sum_1^b jc_j=H$, since the total number of strings in the leaves is n; this can also be verified from (6.7).

7. Patricia tries

Another version of the trie is the Patricia trie, where the trie is compressed by eliminating all internal nodes with only one child. (We use the notations above with a superscript P for the Patricia case.) Since each internal node in the Patricia trie thus has exactly 2 children, the number of internal nodes is one less than the number of external nodes, i.e. $W_n^P=n-1$ for a Patricia trie with n strings.

As another illustration of Theorem 5.1, we note that this trivial result, to the first order at least, also can be derived as above. The condition for a finite string $\boldsymbol{\alpha}$ to be an internal node of the Patricia trie is that there is at least one string beginning with $\boldsymbol{\alpha}0$ and at least one string beginning with $\boldsymbol{\alpha}1$. In the Poisson model, the number of strings with these beginnings are independent Poisson random variables with means $\lambda P(\boldsymbol{\alpha}0)=\lambda qP(\boldsymbol{\alpha})$ and $\lambda P(\boldsymbol{\alpha}1)=\lambda pP(\boldsymbol{\alpha})$, and we can argue as above with $f(x)=(1-e^{-px})(1-e^{-qx})$. In this case, $\int_{-\infty}^{\infty}g(t)\,\mathrm{d}t=\int_0^\infty f(x)x^{-2}=-p\ln p-q\ln q=H$, which implies $\mathbb{E}\,\widetilde{W}_\lambda^P\sim\lambda$ and $\mathbb{E}\,W_n^P\sim n$ in the non-arithmetic case. Moreover, we know that this holds in the arithmetic case too, without oscillations, which means that $\widehat{\psi}(m)=0$ for $m\neq 0$ in (5.6)–(5.7). Indeed, for example by integration by parts,

$$\begin{aligned}\widehat{g}(s)&=\int_0^\infty f(x)x^{-2+\mathrm{i}s}\,\mathrm{d}x=\int_0^\infty x^{-2+\mathrm{i}s}(1-e^{-px}-e^{-qx}+e^{-x})\,\mathrm{d}x\\&=\bigl(1-p^{1-\mathrm{i}s}-q^{1-\mathrm{i}s}\bigr)\Gamma(-1+\mathrm{i}s),\end{aligned}$$

and thus $\widehat{\psi}(m)=\widehat{g}(-2\pi m/d)=0$ for $m\neq 0$.

We can also consider a Patricia b-trie, and obtain the asymptotics of the expected number of internal nodes in a similar way, but it is simpler to use the result in Theorem 6.1 and the fact that the number of internal nodes is $\sum_{j=1}^{b} Z^P_{jn} - 1 = \sum_{j=1}^{b} Z_{jn} - 1$; in the non-arithmetic case this yields the asymptotics $\bigl(\sum_{j=1}^{b} \pi_j\bigr)n$.

The number of internal nodes in the Patricia trie is reduced to $n-1$ from about n/H in the trie (see Theorem 5.4, and ignore the small oscillations in the arithmetic case); this is a reduction by a factor H which is at most $\ln 2 \approx 0.693$, in other words a reduction with at least 30%. Nevertheless, the reduction in the path length to a given string is negligible. In fact, if we for simplicity, as in Section 3, consider $1 + \mathrm{Po}(\lambda)$ strings, with one selected string Ξ, then a string $\boldsymbol{\alpha}$ is an internal node on the path in the trie from the root to Ξ such that $\boldsymbol{\alpha}$ does not appear in the Patricia trie if and only if Ξ begins with $\boldsymbol{\alpha}$, and further, either Ξ begins with $\boldsymbol{\alpha}0$, there is at least one other such string, and there is no string beginning with $\boldsymbol{\alpha}1$, or, conversely, Ξ and at least one other string begins with $\boldsymbol{\alpha}1$ but no string begins with $\boldsymbol{\alpha}0$. The probability of this is $\lambda^{-1} f(x)$ with $x = \lambda P(\boldsymbol{\alpha})$ and

$$f(x) := xq(1 - e^{-qx})e^{-px} + xp(1 - e^{-px})e^{-qx}. \tag{7.1}$$

Hence, if $\Delta D_\lambda := D_\lambda - D^P_\lambda$ is difference between the path lengths to Ξ in the trie and in the Patricia trie, then $\mathbb{E}\,\Delta D_\lambda = \lambda^{-1} \sum_{\boldsymbol{\alpha}} f(\lambda P(\boldsymbol{\alpha}))$ and Theorem 5.1 yields

$$\begin{aligned}
\mathbb{E}\,\Delta D_\lambda &\to \frac{1}{H} \int_0^\infty f(x) x^{-2} \,\mathrm{d}x \\
&= \frac{q}{H} \int_0^\infty \frac{e^{-px} - e^{-x}}{x} \,\mathrm{d}x + \frac{p}{H} \int_0^\infty \frac{e^{-qx} - e^{-x}}{x} \,\mathrm{d}x \\
&= \frac{-q \ln p - p \ln q}{H}.
\end{aligned}$$

This holds also in the arithmetic case, since (7.1) and (6.5) show that the Fourier coefficients $\widehat{\psi}(m)$ in (5.7) are given by

$$\begin{aligned}
\widehat{\psi}(m) &= q \int_0^\infty x^{-1-2\pi \mathrm{i} m/d} (e^{-px} - e^{-x}) \,\mathrm{d}x + p \int_0^\infty x^{-1-2\pi \mathrm{i} m/d} (e^{-qx} - e^{-x}) \,\mathrm{d}x \\
&= q\bigl(p^{2\pi \mathrm{i} m/d} - 1\bigr)\Gamma(-2\pi \mathrm{i} m/d) + p\bigl(q^{2\pi \mathrm{i} m/d} - 1\bigr)\Gamma(-2\pi \mathrm{i} m/d) = 0,
\end{aligned}$$

for integer $m \neq 0$, since $p^{2\pi \mathrm{i} m/d} = e^{2\pi \mathrm{i} m \ln p/d} = 1 = q^{2\pi \mathrm{i} m/d}$ by (2.6). (This is an interesting example of cancellation in an arithmetic case where we would expect oscillations.) Hence the expected saving is 1 for $p = 1/2$, and $O(1)$ for any fixed p. (This is $o(\mathbb{E}\, D_\lambda)$ and thus asymptotically negligible.)

Again, we can depoissonize by considering $\lambda = n \pm n^{2/3}$, and we obtain the same result for a fixed number n of strings. Together with Theorem 3.3, we obtain the following, earlier found by Szpankowski [40], see also Knuth [26, Section 6.3] ($p = 1/2$) and Rais, Jacquet and Szpankowski [36]. (Dynamical sources are considered by Bourdon [5].)

Theorem 7.1. *For the expected depth $\mathbb{E}\, D^P_n$ in a Patricia trie:*

(i) *If* $\ln p/\ln q$ *is irrational, then, as* $n\to\infty$,

$$\mathbb{E}\, D_n^P = \frac{\ln n}{H} + \frac{H_2}{2H^2} + \frac{\gamma + q\ln p + p\ln q}{H} + o(1).$$

(ii) *If* $\ln p/\ln q$ *is rational, then, as* $n\to\infty$,

$$\mathbb{E}\, D_n^P = \frac{\ln n}{H} + \frac{H_2}{2H^2} + \frac{\gamma + q\ln p + p\ln q}{H} + \psi_1(\ln n) + o(1),$$

where $\psi_1(t)$ *is a small continuous function, with period* d *in* t*, given by* (3.10).

8. Insertion in a trie

When a new string is inserted in a trie, it becomes a new external node; it may also create one or several new internal nodes. Let $N\geq 0$ be the number of new internal nodes.

Theorem 8.1. *As* $n\to\infty$,

$$\mathbb{P}(N=0) = 1 - \frac{2pq}{H} - \psi_3(\ln n) + o(1),$$
$$\mathbb{P}(N=j) = \Bigl(\frac{2pq}{H} + \psi_3(\ln n)\Bigr) 2pq(1-2pq)^{j-1} + o(1), \qquad j\geq 1,$$

where $\psi_3=0$ *in the non-arithmetic case, while in the* d*-arithmetic case*

$$\psi_3(t) = \frac{2pq}{H}\sum_{k\neq 0}\Gamma\bigl(1-\frac{2\pi \mathrm{i}k}{d}\bigr)e^{2\pi \mathrm{i}kt/d}.$$

Further,

$$\mathbb{E}\, N = \frac{1}{H} + \frac{1}{2pq}\psi_3(\ln n) + o(1). \tag{8.1}$$

The same results hold in the Poisson case (with n *replaced by* λ*).*

Proof. Consider first the Poisson case, with insertion of Ξ in a trie with $\mathrm{Po}(\lambda)$ other strings.

Let K be the length of the longest prefix of Ξ that is shared with at least two strings already existing in the trie; this is the depth of the last internal node (in the existing trie) that the new string encounters while being inserted.

There is either no existing string with the same $K+1$ first letters as Ξ, or exactly one such string. In the first case, Ξ is inserted at depth $K+1$ without creating any new internal nodes, so $N=0$.

In the second case, we have reached an external node, which is converted into an internal node, and the string that was stored there is displaced and instead stored, together with the new string, at the end of a sequence of $N\geq 1$ new internal nodes, where N is the number of common letters, after the K first, in these two strings.

Thus, conditioned on $N\geq 1$, N has a geometric distribution:

$$\mathbb{P}(N=j) = \mathbb{P}(N\geq 1)(p^2+q^2)^{j-1}\cdot 2pq, \qquad j\geq 1. \tag{8.2}$$

Since further $\mathbb{P}(N=0)=1-\mathbb{P}(N\geq 1)$, it suffices to find $\mathbb{P}(N\geq 1)$.

For a given k, the event $N\geq 1$, $K=k$ and, say, $\xi_{K+1}=1$, happens if and only if $\xi_{k+1}=1$ and there is exactly one existing string beginning with $\xi_1\cdots\xi_k 1$ and at least one beginning with $\xi_1\cdots\xi_k 0$. The conditional probability of this given $\alpha:=\xi_1\cdots\xi_k$ is

$$\mathbb{P}(\xi_{k+1}=1)\,\mathbb{P}\big(\mathrm{Po}(\lambda P(\alpha)q)\geq 1\big)\,\mathbb{P}\big(\mathrm{Po}(\lambda P(\alpha)p)=1\big)=f_1\big(\lambda P(\alpha)\big),$$

with

$$f_1(x)=p(1-e^{qx})(pxe^{-px})=p^2xe^{-px}-p^2xe^{-x}.$$

Thus,

$$\begin{aligned}\mathbb{P}(N\geq 1,\,K=k\text{ and }\xi_{K+1}=1)&=\mathbb{E}\,f_1\big(\lambda\,\mathbb{P}(\xi_1\cdots\xi_k)\big)=\mathbb{E}\,f_1\big(\lambda e^{-S_k}\big)\\&=\mathbb{E}\,f_1\big(e^{-(S_k-\ln\lambda)}\big)\end{aligned}$$

and, summing over k and using (2.11),

$$\mathbb{P}(N\geq 1\text{ and }\xi_{K+1}=1)=\sum_{k=0}^{\infty}\mathbb{E}\,f_1\big(e^{-(S_k-\ln\lambda)}\big)=\int_0^\infty f_1\big(e^{-(x-\ln\lambda)}\big)\,\mathrm{d}U(x).$$

The function $g_1(x):=f_1(e^{-x})$ is directly Riemann integrable on $(-\infty,\infty)$ by Lemma A.6 (because $f_1(x)=O(x\wedge x^{-1})$), and thus the key renewal theorem Theorem A.7 yields

$$\mathbb{P}(N\geq 1\text{ and }\xi_{K+1}=1)=\frac{1}{H}\int_{-\infty}^{\infty}g_1(x)\,\mathrm{d}x+\psi_{31}(\ln\lambda)+o(1).\tag{8.3}$$

where $\psi_{31}(t)=0$ in the non-arithmetic case and

$$\psi_{31}(t)=\frac{1}{H}\sum_{m\neq 0}\widehat{g}_1(-2\pi m/d)e^{2\pi\mathrm{i}mt/d}\tag{8.4}$$

in the arithmetic case.

Routine integrations yield

$$\int_{-\infty}^{\infty}g_1(x)\,\mathrm{d}x=\int_0^\infty f_1(y)\frac{\mathrm{d}y}{y}=\int_0^\infty(p^2e^{-py}-p^2e^{-y})\,\mathrm{d}y=p-p^2=pq\tag{8.5}$$

and, more generally,

$$\widehat{g}_1(s)=\int_{-\infty}^{\infty}e^{-\mathrm{i}sx}g_1(x)\,\mathrm{d}x=\int_0^\infty f_1(y)y^{\mathrm{i}s-1}\,\mathrm{d}y=(p^{1-\mathrm{i}s}-p^2)\Gamma(1+\mathrm{i}s);$$

thus in the arithmetic case, since $p^{2\pi\mathrm{i}m/d}=1$ for integers m,

$$\widehat{g}_1(-2\pi m/d)=pq\Gamma(1-2\pi m\mathrm{i}/d).\tag{8.6}$$

By symmetry, (8.3) implies, for similarly defined g_0 and ψ_0,

$$\mathbb{P}(N\geq 1\text{ and }\xi_{K+1}=0)=\frac{1}{H}\int_{-\infty}^{\infty}g_0(x)\,\mathrm{d}x+\psi_{30}(\ln\lambda)+o(1),\tag{8.7}$$

where, noting that (8.5) and (8.6) are symmetric in p and q, $\int_{-\infty}^{\infty}g_0(x)\,\mathrm{d}x=pq$ and $\psi_{30}=\psi_{31}$.

Consequently, summing (8.3) and (8.7), with $\psi_3 := \psi_{30} + \psi_{31} = 2\psi_{31}$,

$$\mathbb{P}(N \geq 1) = \frac{2pq}{H} + \psi_3(\ln \lambda) + o(1). \tag{8.8}$$

The result in the Poisson case now follows from (8.2), (8.4), (8.6) and (8.8). For the mean we have by (8.2) and (8.8),

$$\mathbb{E}\, N = \sum_{j=0}^{\infty} j\, \mathbb{P}(N = j) = \frac{1}{2pq}\, \mathbb{P}(N \geq 1) = \frac{1}{H} + \frac{1}{2pq}\psi_3(\ln \lambda) + o(1).$$

To depoissonize, consider first adding Ξ to a trie with $\mathrm{Po}(n - n^{2/3})$ strings, and then increase the family by adding $\mathrm{Po}(n^{2/3})$ further strings; it is easily seen that with probability $1 - O(\lambda^{-1/3}) = 1 - o(1)$, this does not change the place where Ξ is inserted, and thus not N. The same holds for all intermediate tries, in particular for the one with exactly n strings if there is one, which there is w.h.p. because $\mathbb{P}\big(\mathrm{Po}(n - n^{2/3}) \leq n\big) \to 1$ and $\mathbb{P}\big(\mathrm{Po}(n + n^{2/3}) \geq n\big) \to 1$. Hence the variable N is w.h.p. the same for n strings and for $\mathrm{Po}(n)$ strings. □

It is easily verified that, at least if we ignore the error terms, the expected number of new internal nodes added for each new string given by (8.1) coincides with the derivative of $\mathbb{E}\, W_\lambda = \frac{\lambda}{H} + \frac{\lambda}{H}\psi_2(\ln \lambda) + o(\lambda)$ given by (5.13), as it should.

Remark 8.2. Christophi and Mahmoud [6] studied random climbing in random tries, taking (in one version) steps left or right with probabilities p and q; this is like inserting a new node but without moving any old one. The length of the climb is thus D_n when $N = 0$ or 1 but $D_n - (N - 1)$ when $N \geq 1$.

The average climb length found by Christophi and Mahmoud [6] for this version thus follows from Theorems 3.3 and 8.1.

9. Tunstall and Khodak codes

Tunstall and Khodak codes are variable-to-fixed length codes that are used in data compression. We give a brief description here. We refer to Savari and Gallager [39] for a treatment of Markov sources by similar methods. See [10], [11] and the survey [42] for more details and references, as well as for an analysis using Mellin transforms.

We recall first the general situation. The idea is that an infinite string can be parsed as a unique sequence of nonoverlapping *phrases* belonging to a certain (finite) *dictionary* $\mathcal{D}$. (For simplicity, we do not consider plurally parsable dictionaries, see Savari [38].) Each phrase in the dictionary then can be represented by a binary number of fixed length ℓ; if there are M phrases in the dictionary we take $\ell := \lceil \lg M \rceil$.

Note first that a set of phrases is a dictionary allowing a unique parsing in the way just described if and only if every infinite string has exactly one prefix in the dictionary. Equivalently, the phrases in the dictionary have to be the external nodes of a trie where every internal node has two children (so the Patricia trie is the same); this trie is the parsing tree.

By a random phrase we mean a phrase distributed as the unique initial phrase in a random infinite string Ξ. Thus a phrase $\boldsymbol{\alpha}$ in the dictionary $\mathcal{D}$ is chosen with probability $P(\boldsymbol{\alpha})$. We let the random variable L be the length of a random phrase.

If we parse an infinite i.i.d. string Ξ, the successive phrases will be independent with this distributions. Hence, if K_N is the (random) number of phrases required to code the N first letters $\xi_1 \cdots \xi_N$, then, see Appendix A and (2.8), $K_N = \nu(N-1)$ for a renewal process where the increments X_i are independent copies of L. Consequently, as $N \to \infty$, by Theorem A.1,

$$\frac{K_N}{N} \xrightarrow{\text{a.s.}} \frac{1}{\mathbb{E}\,L} \qquad \text{and} \qquad \frac{\mathbb{E}\,K_N}{N} \to \frac{1}{\mathbb{E}\,L}. \tag{9.1}$$

We obtain also convergence of higher moments and, by Theorem A.3, a central limit theorem for K_N. The expected number of bits required to code a string of length N is thus

$$\ell\,\mathbb{E}\,K_N \sim \frac{\ell N}{\mathbb{E}\,L} = \frac{\lceil \lg M \rceil}{\mathbb{E}\,L} N.$$

For simplicity, we consider the ratio $\kappa := \lg M / \mathbb{E}\,L$, and call it the *compression rate*. (One objective of the code is to make this ratio small.)

In Khodak's construction of such a dictionary, we fix a threshold $r \in (0,1)$ and construct a parsing tree as the subtree of the complete infinite binary tree such that the internal nodes are the strings $\boldsymbol{\alpha} = \alpha_1 \cdots \alpha_k$ with $P(\boldsymbol{\alpha}) \geq r$; the external nodes are thus the strings $\boldsymbol{\alpha}$ such that $P(\boldsymbol{\alpha}) < r$ but the parent, $\boldsymbol{\alpha}'$ say, has $P(\boldsymbol{\alpha}') \geq r$. The phrases in the Khodak code are the external nodes in this tree. For convenience, we let $R = 1/r > 1$. Let $M = M(R)$ be the number of phrases in the Khodak code.

In Tunstall's construction, we are instead given a number M. We start with the empty phrase and then iteratively $M-1$ times replace a phrase $\boldsymbol{\alpha}$ having maximal $P(\boldsymbol{\alpha})$ by its two children $\boldsymbol{\alpha}0$ and $\boldsymbol{\alpha}1$.

It is easily seen that Khodak's construction with some $r > 0$ gives the same result as Tunstall's with $M = M(R)$. Conversely, a Tunstall code is almost a Khodak code, with r chosen as the smallest $P(\boldsymbol{\alpha})$ for a proper prefix $\boldsymbol{\alpha}$ of a phrase; the difference is that Tunstall's construction handles ties more flexibly; there may be some phrases too with $P(\boldsymbol{\alpha}) = r$. Thus, Tunstall's construction may give any desired number M of phrases, while Khodak's does not. We will see that in the non-arithmetic case, this difference is asymptotically negligible, while it is important in the arithmetic case. (This is very obvious if $p = q = 1/2$, when Khodak's code always gives a dictionary size M that is a power of 2.)

Let us first consider the number of phrases, $M = M(R)$, in Khodak's construction with a threshold $r = 1/R$. This is a purely deterministic problem, but we may nevertheless apply our probabilistic renewal theory arguments.

In fact, M, the number of leafs in the parsing tree, equals 1 + the number of internal nodes. Thus, $M = 1 + \sum_{\boldsymbol{\alpha}} f(RP(\boldsymbol{\alpha}))$ with $f(x) := \mathbf{1}[x \geq 1]$, and we may apply Theorem 5.1.

Theorem 9.1. *Consider the Khodak code with threshold $r = 1/R$.*

(i) *If $\ln p / \ln q$ is irrational, then, as $R \to \infty$,*

$$\frac{M(R)}{R} \to \frac{1}{H}.$$

(ii) *If $\ln p / \ln q$ is rational, then, as $R \to \infty$,*

$$\frac{M(R)}{R} = \frac{1}{H} \cdot \frac{d}{1 - e^{-d}} e^{-d\{(\ln R)/d\}} + o(1).$$

Proof. The non-arithmetic case follows directly from Theorem 5.1(i), since $\int_0^\infty f(x) x^{-2} \,\mathrm{d}x = \int_1^\infty x^{-2} \,\mathrm{d}x = 1$.

In the arithmetic case, we use (5.8). Since $g(t) = e^t \mathbf{1}[t \leq 0]$, the sum in (5.8) is a geometric series that can be summed directly:

$$\psi(t) = d \sum_{kd \leq t} e^{kd - t} = \frac{d}{1 - e^{-d}} e^{d\lfloor t/d \rfloor - t} = \frac{d}{1 - e^{-d}} e^{-d\{t/d\}}. \qquad \square$$

Remark 9.2. In the arithmetic case (ii), $\ln P(\alpha)$ is a multiple of d for any string α. Hence $M(R)$ jumps only when $R \in \{e^{kd} : k \geq 0\}$, and it suffices to consider such R. For these R, the result can be written

$$M(R) \sim \frac{1}{H} \frac{d}{1 - e^{-d}} R, \qquad \ln R \in d\mathbb{Z}. \tag{9.2}$$

Next, consider the length L of a random phrase. We will use the notation L_M^{T} for a Tunstall code with M phrases and L_R^{K} for a Khodak code with threshold $r = 1/R$.

Consider first the Khodak code. By construction, given a random string $\Xi = \xi_1 \xi_2 \cdots$, the first phrase in it is $\xi_1 \cdots \xi_n$ where n is the smallest integer such that $P(\xi_1 \cdots \xi_n) = e^{-S_n} < r = e^{-\ln R}$. Hence, by (2.8),

$$L_R^{\mathsf{K}} = \nu(\ln R). \tag{9.3}$$

Hence, Theorems A.1–A.3 immediately yield the following (as well as convergence of higher moments).

Theorem 9.3. *For the Khodak code, the following holds as $R \to \infty$, with $\sigma^2 = H_2 - H^2 = pq \ln^2(p/q)$:*

$$\frac{L_R^{\mathsf{K}}}{\ln R} \xrightarrow{\text{a.s.}} \frac{1}{H}, \tag{9.4}$$

$$L_R^{\mathsf{K}} \sim \operatorname{AsN}\Big(\frac{\ln R}{H}, \frac{\sigma^2}{H^3} \ln R\Big), \tag{9.5}$$

$$\operatorname{Var} L_R^{\mathsf{K}} \sim \frac{\sigma^2}{H^3} \ln R. \tag{9.6}$$

If $\ln p/\ln q$ *is irrational, then*

$$\mathbb{E}\, L^{\mathsf{K}}_R = \frac{\ln R}{H} + \frac{H_2}{2H^2} + o(1). \tag{9.7}$$

If $\ln p/\ln q$ *is rational, then, with* $d := \gcd(\ln p, \ln q)$ *given by* (2.6)*,*

$$\mathbb{E}\, L^{\mathsf{K}}_R = \frac{\ln R}{H} + \frac{H_2}{2H^2} + \frac{d}{H}\Big(\frac{1}{2} - \Big\{\frac{\ln R}{d}\Big\}\Big) + o(1). \tag{9.8}$$

In the arithmetic case, as said in Remark 9.2, it suffices to consider thresholds such that $-\ln r = \ln R$ is a multiple of d; in this case (9.8) becomes

$$\mathbb{E}\, L^{\mathsf{K}}_R = \frac{\ln R}{H} + \frac{H_2}{2H^2} + \frac{d}{2H} + o(1). \tag{9.9}$$

We analyze the Tunstall code by comparing it to the Khodak code. Thus, suppose that M is given, and increase R (decrease r) until we find a Khodak code with $M(R) \geq M$ phrases. (By our definitions, $M(R)$ is right-continuous, so a smallest such R exists.) Let $M_+ := M(R) \geq M$ and $M_- := M(R-) < M$. Thus, there are $M_+ - 1$ strings α with $P(\alpha) \geq r = R^{-1}$, and $M_- - 1$ strings with $P(\alpha) > r$; consequently there are $M_+ - M_-$ strings with $P(\alpha) = r$. The strings with $P(\alpha) = r$ are not parsing phrases in the Khodak code (while all their children are), but we use some of them in the Tunstall code to achieve exactly M parsing phrases. Since each of these strings replaces two parsing phrases in the Khodak code, the total number of parsing phrases decreases by 1 for each used string with $P(\alpha) = r$, and thus the Tunstall code uses $M(R) - M = M_+ - M$ parsing phrases with $P(\alpha) = r$. The length L^{T}_M of a random phrase, realized as the first phrase in Ξ, equals L^{K}_R unless Ξ begins with one of the phrases α in the Tunstall code with $P(\alpha) = r$, in which case $L^{\mathsf{T}}_M = L^{\mathsf{K}}_R - 1$. The probability of the latter event is evidently $P(\alpha) = r$ for each such α, and is thus $(M(R)-M)r$. Consequently, with R as above,

$$L^{\mathsf{T}}_M = L^{\mathsf{K}}_R - \Delta_M, \tag{9.10}$$

where $\Delta_M \in \{0, 1\}$ and $\mathbb{P}(\Delta_M = 1) = (M(R) - M)/R$. We can now find the results for L^{T}_M:

Theorem 9.4. *For the Tunstall code, the following holds as* $M \to \infty$*, with* $\sigma^2 = H_2 - H^2 = pq \ln^2(p/q)$*:*

$$\frac{L^{\mathsf{T}}_M}{\ln M} \xrightarrow{\text{a.s.}} \frac{1}{H}, \tag{9.11}$$

$$L^{\mathsf{T}}_M \sim \operatorname{AsN}\Big(\frac{\ln M}{H}, \frac{\sigma^2}{H^3}\ln M\Big), \tag{9.12}$$

$$\operatorname{Var} L^{\mathsf{T}}_M \sim \frac{\sigma^2}{H^3}\ln M. \tag{9.13}$$

If $\ln p/\ln q$ *is irrational, then*

$$\mathbb{E}\, L^{\mathsf{T}}_M = \frac{\ln M}{H} + \frac{\ln H}{H} + \frac{H_2}{2H^2} + o(1). \tag{9.14}$$

If $\ln p/\ln q$ *is rational, then, with* $d := \gcd(\ln p, \ln q)$ *given by* (2.6)*,*

$$\mathbb{E}\, L^{\mathsf{T}}_M = \frac{\ln M}{H} + \frac{\ln H}{H} + \frac{H_2}{2H^2} + \frac{1}{H}\ln\frac{\sinh(d/2)}{d/2} + \frac{d}{H}\psi_4\Bigl(\Bigl\{\frac{\ln M + \ln(H(1-e^{-d})/d)}{d}\Bigr\}\Bigr) + o(1), \tag{9.15}$$

where

$$\psi_4(x) := \frac{e^{dx}-1}{e^d-1} - x. \tag{9.16}$$

Note that ψ_4 is continuous, with $\psi_4(0) = \psi_4(1) = 0$. ψ_4 is convex and thus $\psi_4 \le 0$ on [0,1]. In the symmetric case $p = q = 1/2$, $d = H = \ln 2$ and $\psi_4(x) = 2^x - 1 - x$, with a minimum $-0.086071\dots$.

Proof. Let as above R be the smallest number with $M(R) \ge M$; thus $M(R) \ge M > M(R-)$. By Theorem 9.1, $\ln R = \ln M + O(1)$, so (9.11)–(9.13) follow from (9.4)–(9.6) and the fact that $|L^{\mathsf{T}}_M - L^{\mathsf{K}}_R| \le 1$, see (9.10).

If $\ln p/\ln q$ irrational, Theorem 9.1 yields $M(R)/R \to 1/H$, and thus also $M(R-)/R \to 1/H$. Since $M(R) \ge M > M(R-)$, also

$$\frac{M}{R} \to \frac{1}{H}, \tag{9.17}$$

and further $M(R)/M \to 1$. Consequently,

$$\mathbb{E}\,\Delta_M = \frac{M(R)-M}{R} = \Bigl(\frac{M(R)}{M} - 1\Bigr)\frac{M}{R} \to 0,$$

and thus, by (9.10), $\mathbb{E}\, L^{\mathsf{T}}_M = \mathbb{E}\, L^{\mathsf{K}}_R - \mathbb{E}\,\Delta_M = \mathbb{E}\, L^{\mathsf{K}}_R + o(1)$. Since also, by (9.17) again, $\ln R = \ln M + \ln H + o(1)$, (9.14) follows from (9.7).

In the case when $\ln p/\ln q$ is rational, we argue similarly, but we have to be more careful. First, necessarily $R = e^{Nd}$ for some integer N, see Remark 9.2. Further, (9.2) applies. Let, for convenience,

$$\beta := H\,\frac{1-e^{-d}}{d} = H\,\frac{\sinh(d/2)}{d/2}e^{-d/2}; \tag{9.18}$$

thus (9.2) can be written $M(R) \sim \beta^{-1}R$ as $R \to \infty$. Let

$$x := \frac{1}{d}\ln(\beta M) - N + 1 = \frac{1}{d}\ln\frac{\beta M}{R} + 1. \tag{9.19}$$

Then, by these definitions and (9.2),

$$M = \beta^{-1}e^{d(N-1+x)}, \tag{9.20}$$

$$M(R) = \beta^{-1}R(1+o(1)) = \beta^{-1}e^{dN+o(1)}, \tag{9.21}$$

$$M(R-) = M(Re^{-d}) = \beta^{-1}(Re^{-d})(1+o(1)) = \beta^{-1}e^{d(N-1)+o(1)}. \tag{9.22}$$

Since $M(R-) < M \le M(R)$, we see that $o(1) \le x \le 1 + o(1)$. We define also, using (9.20),

$$x_0 := \Bigl\{\frac{\ln(\beta M)}{d}\Bigr\} = \Bigl\{\frac{\ln e^{d(N-1+x)}}{d}\Bigr\} = \{x\}. \tag{9.23}$$

Typically, $0 \le x < 1$, and then $x_0 = x$, but it may happen that x is slightly below 0 and $x_0 = x + 1$, or that x is slightly above 1 and then $x_0 = x - 1$.

By (9.19), $\ln R = \ln(\beta M) + d(1 - x)$, and thus (9.9) yields, using (9.18),

$$\begin{aligned}\mathbb{E}\, L_R^{\mathsf{K}} &= \frac{\ln(\beta M)}{H} + \frac{H_2}{2H^2} + \frac{d}{2H} + \frac{d}{H}(1-x) + o(1)\\ &= \frac{\ln M}{H} + \frac{\ln H}{H} + \frac{1}{H}\ln\frac{\sinh(d/2)}{d/2} + \frac{H_2}{2H^2} + \frac{d}{H}(1-x) + o(1).\end{aligned}$$

Furthermore, by $R = e^{dN}$, (9.20), (9.21) and (9.18),

$$\begin{aligned}\mathbb{E}\,\Delta_M &= \frac{M(R) - M}{R} = \beta^{-1}(1 - e^{d(x-1)}) + o(1)\\ &= \frac{d}{H}\,\frac{1 - e^{xd-d}}{1 - e^{-d}} + o(1) = \frac{d}{H}\Bigl(1 - \frac{e^{xd} - 1}{e^d - 1}\Bigr) + o(1).\end{aligned}$$

Combining these, we find by (9.10) and (9.16),

$$\begin{aligned}\mathbb{E}\, L_M^{\mathsf{T}} &= \mathbb{E}\, L_R^{\mathsf{K}} - \mathbb{E}\,\Delta_M\\ &= \frac{\ln M}{H} + \frac{\ln H}{H} + \frac{1}{H}\ln\frac{\sinh(d/2)}{d/2} + \frac{H_2}{2H^2} + \frac{d}{H}\psi_4(x) + o(1).\end{aligned}$$

This is almost (9.15), except that there $\psi_4(x)$ is replaced by $\psi_4(x_0) = \psi_4(\{\ln(\beta M)/d\})$, see (9.23). However, as noted above, $x \ne x_0$ can happen only when one of x and x_0 is $o(1)$ and the other is $1+o(1)$. Since the function ψ_4 is continuous and $\psi_4(0) = \psi_4(1)$, we see that in this case $\psi_4(x) - \psi_4(x_0) = \pm(\psi_4(1) - \psi_4(0)) + o(1) = o(1)$. Hence, $\psi_4(x) = \psi_4(x_0) + o(1)$ in all cases, and (9.15) follows. □

Remark 9.5. We have chosen to derive Theorem 9.4 from the corresponding result Theorem 9.3 for the Khodak code. An alternative is to note that in the Tunstall code, we obtain the random phrase length L_M^{T} by stopping Ξ at $M_+ - M$ of the $M_+ - M_-$ strings α with $P(\alpha) = r$, and all strings with smaller $P(\alpha)$. By symmetry, we obtain the same distribution of the length if we stop randomly with probability $(M_+ - M)/(M_+ - M_-)$ whenever $P(\alpha) = e^{-S_n} = r$; equivalently, we stop when $e^{-S_n - X_0} < r$, where X_0 is a random variable, independent of Ξ, with values 0 and ε, for some very small positive $\varepsilon = \varepsilon(M)$, and $\mathbb{P}(X_0 = \varepsilon) = (M_+ - M)/(M_+ - M_-)$. Consequently, we have $L_M^{\mathsf{T}} \overset{\mathrm{d}}{=} \widehat{\nu}(\ln R)$, with R and X_0 as above, and we can apply Theorems A.1–A.3 (and A.4) directly.

Corollary 9.6. *The compression rate for the Tunstall code is*

$$\kappa := \frac{\lg M}{\mathbb{E}\, L_M^{\mathsf{T}}} = \frac{H}{\ln 2}\left(1 - \frac{\ln H + H_2/2H + \delta}{\ln M} + o\bigl((\ln M)^{-1}\bigr)\right)$$

where $\delta = 0$ *when* $\ln p/\ln q$ *is irrational while when* $\ln p/\ln q$ *is rational,*

$$\delta := \ln\frac{\sinh(d/2)}{d/2} + d\psi_4\Bigl(\Bigl\{\frac{\ln M + \ln(H(1-e^{-d})/d)}{d}\Bigr\}\Bigr),$$

with d given by (2.6) and ψ_4 by (9.16).

For the Khodak code, the compression rate $\lg(M(R))/\mathbb{E}\,L_R^{\mathsf{K}}$ *is asymptotically given by the same formula, with* $\ln M$ *replaced by* $\ln R$*, except that the* ψ_4 *term does not appear in* δ*.*

The reason that the ψ_4 term does not appear for the Khodak code is that $L_R^{\mathsf{K}} = L_{M(R)}^{\mathsf{T}}$, and in the arithmetic case, we may assume that $R = e^{Nd}$, and then for $L_{M(R)}^{\mathsf{T}}$, the argument x_0 of ψ is $\{\ln(\beta M(R))/d\} = \{\ln(R)/d + o(1)\} = \{N + o(1)\}$ and thus close to 0 or 1, where ψ_4 vanishes.

10. A stopped random walk

Drmota and Szpankowski [12] consider (motivated by the study of Tunstall and Khodak codes) walks in a region in the first quadrant bounded by two crossing lines. Their first result, on the number of possible paths, seems to require a longer comment, and will not be considered here. Their second result is about a random walk in the plane taking only unit steps north or east, which is stopped when it exits the region; the probability of an east step is p each time. Coding steps east by 1 and north by 0, this is the same as taking our random string Ξ. Drmota and Szpankowski [12] study, in our notation, the exit time

$$D_{K,V} := \min\{n : n > K \text{ or } S_n > V \ln 2\}$$

for given numbers K and V, with K integer. We thus have

$$D_{K,V} = (K+1) \wedge \nu(V \ln 2). \tag{10.1}$$

We have here kept the notations K and V from [12], but for convenience we in the sequel write $V_2 := V \ln 2$. We assume $p \neq q$, since otherwise $D_{K,V} = (K \wedge \lfloor V \rfloor) + 1$ is deterministic.

We need a little more notation. Let as usual $\phi(x) := (2\pi)^{-1/2} e^{-x^2/2}$ and $\Phi(x) := \int_{-\infty}^{x} \phi(y)\,\mathrm{d}y$ be the density and distribution functions of the standard normal distribution. Further, let

$$\Psi(x) := \int_{-\infty}^{x} \Phi(y)\,\mathrm{d}y = x\Phi(x) + \phi(x). \tag{10.2}$$

This definition is motivated by the following lemma.

Lemma 10.1. *If $Z \sim N(0,1)$, then for every real t, $\mathbb{E}(Z \vee t) = \Psi(t)$ and $\mathbb{E}(Z \wedge t) = -\Psi(-t)$. Further, $\Psi(t) - \Psi(-t) = t$.*

Proof. Since $\mathbb{E}\,Z = 0$,

$$\mathbb{E}(Z \vee t) = \mathbb{E}(Z \vee t - Z) = \int_0^\infty \mathbb{P}(Z \vee t - Z > x)\,\mathrm{d}x = \int_0^\infty \Phi(t-x)\,\mathrm{d}x = \Psi(t).$$

Further, since $-Z \overset{\mathrm{d}}{=} Z$,

$$-\mathbb{E}(Z \wedge t) = \mathbb{E}\big((-Z) \vee (-t)\big) = \mathbb{E}\big(Z \vee (-t)\big) = \Psi(-t).$$

Finally, $\Psi(t) - \Psi(-t) = \mathbb{E}\big((Z \vee t) + (Z \wedge t)\big) = \mathbb{E}(Z + t) = t$. (This also follows from (10.2) and $\Phi(t) + \Phi(-t) = 1$, $\phi(-t) = \phi(t)$.) $\square$

We can now state our version of the result by Drmota and Szpankowski [12]. We do not obtain as sharp error estimates as they do (although our bounds easily can be improved when $|K - V_2/H|$ is large enough). On the other hand, our result is more general and includes the transition region when $V_2/H \approx K$ and both stopping conditions are important.

Theorem 10.2. *Suppose that $p \neq q$ and that $V, K \to \infty$. Let $V_2 := V \ln 2$ and $\widetilde{\sigma}^2 := (H_2 - H^2)/H^3 > 0$.*

(i) *If $(K - V_2/H)/\sqrt{V_2} \to +\infty$, then $D_{K,V}$ is asymptotically normal:*

$$D_{K,V} \sim \operatorname{AsN}\Big(\frac{V_2}{H}, \widetilde{\sigma}^2 V_2\Big). \tag{10.3}$$

Further, $\operatorname{Var}(D_{K,V}) \sim \widetilde{\sigma}^2 V_2$.

(ii) *If $(K - V_2/H)/\sqrt{V_2} \to -\infty$, then $D_{K,V}$ is asymptotically degenerate:*

$$\mathbb{P}(D_{K,V} = K + 1) \to 1. \tag{10.4}$$

Further, $\operatorname{Var} D = o(V_2)$.

(iii) *If $(K - V_2/H)/\sqrt{V_2} \to a \in (-\infty, +\infty)$, then $D_{K,V}$ is asymptotically truncated normal:*

$$V_2^{-1/2}(D_{K,V} - V_2/H) \xrightarrow{\mathrm{d}} (\widetilde{\sigma} Z) \wedge a = \widetilde{\sigma}\big(Z \wedge (a/\widetilde{\sigma})\big). \tag{10.5}$$

with $Z \sim N(0,1)$. Further,

$$\operatorname{Var}(D_{K,V}) \sim V_2 \operatorname{Var}(\widetilde{\sigma} Z \wedge a) = V_2 \widetilde{\sigma}^2 \operatorname{Var}(Z \wedge (a/\widetilde{\sigma})).$$

(iv) *In every case,*

$$\mathbb{E}\, D_{K,V} = \frac{V_2}{H} - \widetilde{\sigma}\sqrt{V_2}\,\Psi\Big(\frac{V_2/H - K}{\widetilde{\sigma}\sqrt{V_2}}\Big) + o(\sqrt{V_2}) \tag{10.6}$$

$$= K - \widetilde{\sigma}\sqrt{V_2}\,\Psi\Big(\frac{K - V_2/H}{\widetilde{\sigma}\sqrt{V_2}}\Big) + o(\sqrt{V_2}). \tag{10.7}$$

(v) *If $(K - V_2/H)/\sqrt{V_2} \geq \ln V_2$, then*

$$\mathbb{E}\, D_{K,V} = \frac{V_2}{H} + \frac{H_2}{2H^2} + \psi_5(V_2) + o(1), \tag{10.8}$$

where $\psi_5 = 0$ in the non-arithmetic case and $\psi_5(t) = \frac{d}{H}\big(1/2 - \{t/d\}\big)$ in the d-arithmetic case.

(vi) *If $(K - V_2/H)/\sqrt{V_2} \leq -\ln V_2$, then*

$$\mathbb{E}\, D_{K,V} = K + 1 + o(1). \tag{10.9}$$

Proof. Let

$$\widetilde{D} := \frac{D_{K,V} - V_2/H}{\sqrt{V_2}}, \qquad \widetilde{\nu} := \frac{\nu(V_2) - V_2/H}{\sqrt{V_2}},$$

$$\widetilde{K} := \frac{K - V_2/H}{\sqrt{V_2}}, \qquad \widetilde{K}_1 := \frac{K + 1 - V_2/H}{\sqrt{V_2}} = \widetilde{K} + o(1).$$

Thus, by (10.1), $\widetilde{D} = \widetilde{\nu} \wedge \widetilde{K}_1$. By Theorem A.3,

$$\widetilde{\nu} = \frac{\nu(V_2) - V_2/H}{\sqrt{V_2}} \xrightarrow{\mathrm{d}} N(0, \widetilde{\sigma}^2). \tag{10.10}$$

The results on convergence in distribution in (i)–(iii) follow immediately:

In (i), $\widetilde{K} \to \infty$ and $\widetilde{K}_1 \to \infty$ so (10.10) implies that w.h.p. $\widetilde{\nu} < \widetilde{K}_1$ and thus $\nu(V_2) < K+1$; hence w.h.p. $D_{K,V} = \nu(V_2)$ and (10.3) follows from (10.10).

In (ii), similarly w.h.p. $\widetilde{\nu} > \widetilde{K}_1$ and $\nu(V_2) > K+1$ so $D_{K,V} = K+1$.

In (iii), we have $\widetilde{K}_1 \to a$ and thus (10.10) implies $(\widetilde{\nu}, \widetilde{K}_1) \xrightarrow{\mathrm{d}} (\widehat{\sigma} Z, a)$; hence (10.5) follows by applying the continuous mapping theorem [3, Section 5] to $\wedge : \mathbb{R}^2 \to \mathbb{R}$.

For (iv), note first that the two expressions in (10.6) and (10.7) are the same by Lemma 10.1. We may by considering subsequences assume that one of the cases (i)–(iii) occurs.

Next, (A.9) can be written $\mathbb{E}(\widetilde{\nu}^2) \to \widetilde{\sigma}^2$, which together with (10.10) implies that $\widetilde{\nu}^2$ is uniformly integrable. (See e.g. [15, Theorem 5.5.9].) In case (iii), when $\widetilde{K}_1$ converges, this implies that $\widetilde{D}^2 = (\widetilde{\nu} \wedge \widetilde{K}_1)^2$ also is uniformly integrable, and thus the convergence in distribution already proved for (iii) implies $\mathbb{E}\,\widetilde{D} \to \mathbb{E}(\widetilde{\sigma}(Z \wedge (a/\widetilde{\sigma}))) = -\widetilde{\sigma}\Psi(-a/\widetilde{\sigma})$, which yields (10.6) when $K \to a \in \mathbb{R}$; further, the uniform square integrability of $\widetilde{D}^2$ implies $\operatorname{Var} \widetilde{D} \to \operatorname{Var}(\widetilde{\sigma} Z \wedge a)$ as asserted in (iii).

If instead $\widetilde{K}_1 \to +\infty$, case (i), we may assume $\widetilde{K}_1 > 0$; then $\widetilde{D}^2 = (\widetilde{\nu} \wedge \widetilde{K}_1)^2 \le \widetilde{\nu}^2$ and thus $\widetilde{D}^2$ is uniformly integrable in this case too. Hence (10.3) implies both $\operatorname{Var}(\widetilde{D}) \sim \widetilde{\sigma}^2$, or equivalently $\operatorname{Var} D_{K,V} \sim \widetilde{\sigma}^2 V_2$ as asserted in (i), and $\mathbb{E}\,\widetilde{D} \to 0$, which yields (10.6) in this case because $\Psi(-\widetilde{K}) \to 0$.

Finally, if $\widetilde{K}_1 \to -\infty$, case (ii), we may assume that $\widetilde{K}_1 < 0$; then $\widetilde{K}_1 - \widetilde{D} = (\widetilde{K}_1 - \widetilde{\nu})_+ \le |\widetilde{\nu}|$ is uniformly square integrable, and $\widetilde{K}_1 - \widetilde{D} \xrightarrow{\mathrm{p}} 0$ by (10.4). Hence $\widetilde{K}_1 - \mathbb{E}\,\widetilde{D} = \mathbb{E}(\widetilde{K}_1 - \widetilde{D}) \to 0$, and thus (10.7) holds, since $\Psi(\widetilde{K}) \to 0$ and $1 = o(\sqrt{V_2})$. Further, $\operatorname{Var} \widetilde{D} = \operatorname{Var}(\widetilde{K}_1 - \widetilde{D}) \to 0$, which yields $\operatorname{Var} D = o(V_2)$.

This completes the proof of (iv).

For (v), we have $D_{K,V} \le \nu(V_2)$ and thus, by the Cauchy–Schwarz inequality and Theorem A.1,

$$\begin{aligned} \mathbb{E}\,|D_{K,V} - \nu(V_2)| &\le \mathbb{E}\big(\nu(V_2)\mathbf{1}[D_{K,V} \ne \nu(V_2)]\big) \\ &\le \big(\mathbb{E}\,\nu(V_2)^2\big)^{1/2}\, \mathbb{P}\big(D_{K,V} \ne \nu(V_2)\big)^{1/2} \\ &= O(V_2)\, \mathbb{P}\big(D_{K,V} \ne \nu(V_2)\big)^{1/2}. \end{aligned} \tag{10.11}$$

For $\widetilde{K} \geq \ln V_2$, Chernoff's bound [24, Theorem 2.1] implies, because S_{K+1} is a linear transformation of a binomial $\mathrm{Bi}(K+1,p)$ random variable,

$$\begin{aligned}\mathbb{P}(D_{K,V} \neq \nu(V_2)) &= \mathbb{P}(\nu(V_2) > K+1) = \mathbb{P}(S_{K+1} \leq V_2) \\ &= \mathbb{P}(S_{K+1} - \mathbb{E}\,S_{K+1} \leq -H\widetilde{K}_1\sqrt{V_2}) \\ &\leq \exp\Bigl(-c_1 \frac{\widetilde{K}_1^2 V_2}{K+1+\widetilde{K}_1\sqrt{V_2}}\Bigr) \\ &\leq \exp(-c_2 \ln^2(V_2)).\end{aligned}$$

for some $c_1, c_2 > 0$ (depending on p); the last inequality is perhaps most easily seen by considering the case $K+1 \leq 2V_2/H$ (when $K+1 \asymp V_2$) and $K+1 > 2V_2/H$ (when $\widetilde{K}_1 \asymp K/\sqrt{V_2}$) separately. Hence, the right-hand side of (10.11) tends to 0, and thus $\mathbb{E}\,D_{K,V} = \mathbb{E}\,\nu(V_2) + o(1)$. Consequently, (v) follows from the formulas (A.3) and (A.5) for $\mathbb{E}\,\nu(V_2)$ provided by Theorem A.2.

The argument for (vi) is very similar. The Chernoff bound for S_K implies

$$\mathbb{P}(D_{K,V} \neq K+1) = \mathbb{P}(\nu(V_2) < K+1) = \mathbb{P}(S_K > V_2) \leq \exp(-c_3 \ln^2(V_2)),$$

and the Cauchy–Schwarz inequality then implies $\mathbb{E}\,|K+1-D_{K,V}| = o(1)$, proving (vi). ☐

Appendix A. Some renewal theory

For the readers' (and our own) convenience, we collect here a few standard results from renewal theory, sometimes in less standard versions. See e.g. Asmussen [1], Feller [14] or Gut [16] for further details.

We suppose that $X_1, X_2, \dots$ is an i.i.d. sequence of non-negative random variables with finite mean $\mu := \mathbb{E}\,X > 0$, and that $S_n := \sum_{i=1}^n X_i$. Moreover, we suppose that X_0 is independent of $X_1, X_2, \dots$ (but X_0 may have a different distribution, and is not necessarily positive) and define $\widehat{S}_n := \sum_{n=0}^\infty X_i = S_n + X_0$. We further define the first passage times $\nu(t)$ and $\widehat{\nu}(t)$ by (2.8) and (2.13) and the renewal function U by (2.10). (Recall that ν is a special case of $\widehat{\nu}$ with $X_0 = 0$. Hence the results stated below for $\widehat{\nu}$ hold for ν too.)

For some theorems, we have to distinguish between the arithmetic (lattice) and non-arithmetic (non-lattice) cases, in general defined as follows:

arithmetic (lattice): There is a positive real number d such that X_1/d always is an integer. We let d be the largest such number and say that X_1 is *d-arithmetic*. (This maximal d is called the *span* of the distribution.)

non-arithmetic (non-lattice): No such d exists. (Then X_1 is not supported on any proper closed subgroup of $\mathbb{R}$.)

Theorem A.1. *As $t \to \infty$,*

$$\frac{\widehat{\nu}(t)}{t} \xrightarrow{\text{a.s.}} \frac{1}{\mu}. \tag{A.1}$$

If further $0 < r < \infty$ and $\mathbb{E}\,|X_0|^r < \infty$, then $\widehat{\nu}(t)/t \to \mu^{-1}$ in L^r, i.e., $\mathbb{E}\,|\widehat{\nu}(t)/t - \mu^{-1}|^r \to 0$, and thus

$$\mathbb{E}\Bigl(\frac{\widehat{\nu}(t)}{t}\Bigr)^r \to \frac{1}{\mu^r}. \tag{A.2}$$

Proof. See e.g. Gut [16, Theorem 2.5.1] for the case $X_0 = 0$; the general case follows by essentially the same proof. □

Theorem A.2. *Suppose that $\mathbb{E}\,X_1^2 < \infty$ and $\mathbb{E}\,|X_0| < \infty$.*

(i) *If the distribution of X_1 is non-arithmetic, then, as $t \to \infty$,*

$$\mathbb{E}\,\nu(t) = \frac{t}{\mu} + \frac{\mathbb{E}\,X_1^2}{2\mu^2} + o(1) \tag{A.3}$$

and, more generally,

$$\mathbb{E}\,\widehat{\nu}(t) = \frac{t}{\mu} + \frac{\mathbb{E}\,X_1^2}{2\mu^2} - \frac{\mathbb{E}\,X_0}{\mu} + o(1). \tag{A.4}$$

(ii) *If the distribution of X_1 is d-arithmetic, then, as $t \to \infty$,*

$$\mathbb{E}\,\nu(t) = \frac{t}{\mu} + \frac{\mathbb{E}\,X_1^2}{2\mu^2} + \frac{d}{\mu}\Bigl(\frac{1}{2} - \Bigl\{\frac{t}{d}\Bigr\}\Bigr) + o(1). \tag{A.5}$$

and, more generally,

$$\mathbb{E}\,\widehat{\nu}(t) = \frac{t}{\mu} + \frac{\mathbb{E}\,X_1^2}{2\mu^2} + \frac{d}{\mu}\Bigl(\frac{1}{2} - \mathbb{E}\Bigl\{\frac{t-X_0}{d}\Bigr\}\Bigr) - \frac{\mathbb{E}\,X_0}{\mu} + o(1). \tag{A.6}$$

Proof. See e.g. Gut [16, Theorem 2.5.2] for the case $X_0 = 0$; the general case follows easily by conditioning on X_0. In the arithmetic case, note that $\widehat{\nu}(t) = \nu(t - X_0) = \nu\bigl(\lfloor (t-X_0)/d \rfloor d\bigr)$ and use $\mathbb{E}(\lfloor (t-X_0)/d \rfloor d) = t - \mathbb{E}\,X_0 - d\,\mathbb{E}\{(t-X_0)/d\}$. □

Theorem A.3. *Assume that $\sigma^2 := \operatorname{Var} X_1 < \infty$. Then, as $t \to \infty$,*

$$\frac{\widehat{\nu}(t) - t/\mu}{\sqrt{t}} \xrightarrow{\text{d}} N\Bigl(0, \frac{\sigma^2}{\mu^3}\Bigr). \tag{A.7}$$

If further $\sigma^2 > 0$, this can be written $\widehat{\nu} \sim \operatorname{AsN}(\mu^{-1}t, \sigma^2\mu^{-3}t)$.

Moreover, if also $\mathbb{E}\,X_0^2 < \infty$, then

$$\operatorname{Var}\bigl(\widehat{\nu}(t)\bigr) = \frac{\sigma^2}{\mu^3}t + o(t); \tag{A.8}$$

and

$$\mathbb{E}\bigl(\widehat{\nu}(t) - t/\mu\bigr)^2 = \frac{\sigma^2}{\mu^3}t + o(t). \tag{A.9}$$

Proof. See e.g. Gut [16, Theorem 2.5.2] for the case $X_0 = 0$, noting that (A.8) and (A.9) are equivalent because $\mathbb{E}\,\widehat{\nu}(t) - t/\mu = O(1)$ by Theorem A.2; again, the case with a general X_0 is similar, or follows by conditioning on X_0. The case $\sigma^2 = 0$ is trivial. □

In some arguments above, we let $X_0 = X_0^{(t)}$ to depend on t; this is no problem if we have some uniformity assumptions on $X_0^{(t)}$, for example the following. (The proof shows that weaker assumptions suffice.)

Theorem A.4. *We can allow $X_0 = X_0^{(t)}$ to depend on t in Theorems A.1–A.3 provided $\xrightarrow{\text{a.s.}}$ is weakened to $\xrightarrow{\text{p}}$ in* (A.1) *and we add the following assumptions: $X_0^{(t)}$ is tight; for L^r convergence and* (A.2) *we further assume that $\sup_t \mathbb{E}\,|X_0^{(t)}|^r < \infty$; for Theorem A.2 we assume that $X_0^{(t)}$ are uniformly integrable; for* (A.8) *and* (A.9) *we assume that $\sup_t \mathbb{E}\,|X_0^{(t)}|^2 < \infty$.*

Proof. This can be seen by conditioning on $X_0^{(t)}$, noting that $\widehat{\nu}(t) = \nu(t - X_0^{(t)})$. First, if $X_0^{(t)}$ is tight, then $X_0^{(t)}/t \xrightarrow{\text{p}} 0$ and $t - X_0^{(t)} \xrightarrow{\text{p}} \infty$ as $t \to \infty$, and thus by conditioning on $X_0^{(t)}$ we obtain from (A.1) for ν

$$\frac{\widehat{\nu}(t)}{t} = \frac{\nu(t - X_0^{(t)})}{t} = \frac{t - X_0^{(t)}}{t} \cdot \frac{\nu(t - X_0^{(t)})}{t - X_0^{(t)}} \xrightarrow{\text{p}} \frac{1}{\mu},$$

showing (A.1) with $X_0^{(t)}$. If $\operatorname{Var} X_1 < \infty$, then (A.7) follows similarly.

Similarly, if we define

$$Y_t := \mathbb{E}\Bigl(\Bigl(\frac{\widehat{\nu}(t)}{t}\Bigr)^r \Bigm| X_0^{(t)}\Bigr) = \mathbb{E}\Bigl(\Bigl(\frac{\nu(t - X_0^{(t)})}{t}\Bigr)^r \Bigm| X_0^{(t)}\Bigr),$$

then (A.2) and $(t - X_0^{(t)})/t \xrightarrow{\text{p}} 1$ yield $Y_t \xrightarrow{\text{p}} \mu^{-r}$. Moreover, (A.2) also implies $\mathbb{E}\,\nu(t)^r = O(1 + t^r)$, and thus, for $t \ge 1$,

$$Y_t = O\bigl((1 + |t - X_0^{(t)}|^r)/t^r\bigr) = O\bigl(1 + |X_0^{(t)}|^r t^{-r}\bigr).$$

Hence, $\mathbb{E}\bigl(Y_t \mathbf{1}[|X_0^{(t)}| \le t]\bigr) \to \mu^{-r}$ as $t \to \infty$ by dominated convergence, while, assuming $\sup_t \mathbb{E}\,|X_0^{(t)}|^r < \infty$, $\mathbb{E}\bigl(Y_t \mathbf{1}[|X_0^{(t)}| > t]\bigr) = O\bigl(\mathbb{E}\,|X_0^{(t)}|^r t^{-r}\bigr) \to 0$, so $\mathbb{E}(\widehat{\nu}(t)/t)^r = \mathbb{E}\,Y_t \to \mu^{-r}$, showing (A.2). $\mathbb{E}\,|\widehat{\nu}(t)/t - \mu^{-1}|^r \to 0$ follows similarly.

If we denote the error term in (A.3) or (A.5) by $r(t)$, then $r(t) \to 0$ as $t \to \infty$ and $r(t) = O(1 + t_-)$, where $t_- := \max(-t, 0)$; hence $\mathbb{E}\,r(t - X_0^{(t)}) \to 0$ by dominated convergence, and (A.4) and (A.6) follow.

For (A.8) and (A.9) we use the standard decomposition

$$\operatorname{Var}\bigl(\widehat{\nu}(t)\bigr) = \mathbb{E}\bigl(\operatorname{Var}\bigl(\widehat{\nu}(t) \bigm| X_0^{(t)}\bigr)\bigr) + \operatorname{Var}\bigl(\mathbb{E}\bigl(\widehat{\nu}(t) | X_0^{(t)}\bigr)\bigr). \tag{A.10}$$

Let $\operatorname{Var}(\nu(t)) = \sigma^2 \mu^{-3} t + r_2(t)$, where by (A.8) $r_2(t) = o(t)$ as $t \to \infty$, and thus $r_2(t) = O(1 + |t|)$. Then

$$\operatorname{Var}\bigl(\widehat{\nu}(t) \bigm| X_0^{(t)}\bigr) = \sigma^2 \mu^{-3} (t - X_0^{(t)}) + r_2(t - X_0^{(t)})$$

and dominated convergence yields

$$\mathbb{E}\big(t^{-1}\operatorname{Var}\big(\widehat{\nu}(t)\bigm| X_0^{(t)}\big)\big) \to \sigma^2\mu^{-3}.$$

For the last term in (A.10), we note that (A.4) and (A.6) show that

$$\mathbb{E}\big(\widehat{\nu}(t)|X_0^{(t)}\big) = \mathbb{E}\big(\nu(t-X_0^{(t)})|X_0^{(t)}\big) = \frac{t}{\mu} + O(|X_0^{(t)}|+1),$$

and thus the variance is $0(1) = o(t)$. Hence, (A.10) yields (A.8). Finally, (A.9) follows by (A.8) and (A.4) or (A.6). □

For the evaluation of (A.6) when X_0 is non-trivial, we note the following formula.

Lemma A.5. *Suppose that X has a continuous distribution with finite mean, and a characteristic function $\varphi(t) := Ee^{\mathrm{i}tX}$ that satisfies $\varphi(t) = O(|t|^{-\delta})$ for some $\delta > 0$. Then, for any real u,*

$$\mathbb{E}\{X+u\} = \frac{1}{2} - \sum_{n\neq 0} \frac{\varphi(2\pi n)}{2\pi n\mathrm{i}} e^{2\pi\mathrm{i}nu}.$$

Proof. Let $X_u := \lfloor X+u \rfloor - u + 1$. Then $\{X+u\} = X - X_u + 1$, and the result follows from the formula for $\mathbb{E}\,X_u$ in [23, Theorem 2.3]. □

For the next theorem (known as the *key renewal theorem*), we say that a function $f \geq 0$ on $(-\infty,\infty)$ is *directly Riemann integrable* if the upper and lower Riemann sums $\sum_{k=-\infty}^{\infty} h \sup_{[(k-1)h,kh)} f$ and $\sum_{k=-\infty}^{\infty} h \inf_{[(k-1)h,kh)} f$ are finite and converge to the same limit as $h \to 0$. (See further Feller [14, Section XI.1]; Feller considers functions on $[0,\infty)$, but this makes no difference.) For most purposes, the following sufficient condition suffices. (Usually, one can take $F = f$.)

Lemma A.6. *Suppose that f is a non-negative function on $(-\infty,\infty)$. If f is bounded and a.e. continuous, and there exists an integrable function F with $0 \leq f \leq F$ such that F is non-decreasing on $(-\infty,-A)$ and non-increasing on (A,∞) for some A, then f is directly Riemann integrable.*

Proof. Let $f_{h+}(x) := \sup_{[(k-1)h,kh)} f$ and $f_{h-}(x) := \inf_{[(k-1)h,kh)} f$ for $x \in [(k-1)h, kh)$; thus the upper and lower Riemann sums are $\int_{-\infty}^{\infty} f_{h+}(x)\,\mathrm{d}x$ and $\int_{-\infty}^{\infty} f_{h-}(x)\,\mathrm{d}x$. As $h \to 0$, $f_{h+}(x) \to f(x)$ and $f_{h-}(x) \to f(x)$ at every continuity point of f, and thus a.e. Moreover, if we define $g(x) := F(x+1)$ for $x < -A-1$, $g(x) := F(x-1)$ for $x > A+1$, and $g(x) := \sup f < \infty$ for $|x| \leq A+1$, then g is integrable and for $0 < h < 1$, $0 \leq f_{h-}(x) \leq f_{h+}(x) \leq g(x)$. Hence dominated convergence applies and shows $\int_{-\infty}^{\infty} f_{h+} \to \int_{-\infty}^{\infty} f$ and $\int_{-\infty}^{\infty} f_{h-} \to \int_{-\infty}^{\infty} f$ as $h \to 0$, as we wanted to show. □

Theorem A.7. *Let f be any non-negative directly Riemann integrable function on $(-\infty,\infty)$.*

(i) *If the distribution of X_1 is non-arithmetic, then, as $t \to \infty$,*

$$\int_0^\infty f(s-t)\,\mathrm{d}U(s) \to \frac{1}{\mu}\int_{-\infty}^\infty f(s)\,\mathrm{d}s, \tag{A.11}$$

$$\int_0^\infty f(t-s)\,\mathrm{d}U(s) \to \frac{1}{\mu}\int_{-\infty}^\infty f(s)\,\mathrm{d}s. \tag{A.12}$$

(ii) *If the distribution of X_1 is d-arithmetic, then, as $t \to \infty$,*

$$\int_0^\infty f(s-t)\,\mathrm{d}U(s) = \frac{1}{\mu}\psi(t) + o(1), \tag{A.13}$$

$$\int_0^\infty f(t-s)\,\mathrm{d}U(s) = \frac{1}{\mu}\psi(-t) + o(1), \tag{A.14}$$

where $\psi(t)$ is the bounded d-periodic function

$$\psi(t) := d\sum_{k=-\infty}^{\infty} f(kd-t); \tag{A.15}$$

ψ has the Fourier series

$$\psi(t) \sim \sum_{m=-\infty}^{\infty} \widehat{\psi}(m)e^{2\pi \mathrm{i} mt/d} \tag{A.16}$$

with

$$\widehat{\psi}(m) = \widehat{f}(-2\pi m/d) = \int_{-\infty}^\infty e^{2\pi \mathrm{i} mt/d} f(t)\,\mathrm{d}t. \tag{A.17}$$

In particular, the average of ψ is $\widehat{\psi}(0) = \int_{-\infty}^\infty f$. The series (A.15) converges uniformly on $[0,d]$; thus ψ is continuous if f is. Further, if f is sufficiently smooth (an integrable second derivative is enough), then the Fourier series (A.16) converges uniformly.

Proof. The two formulas (A.11) and (A.12) are equivalent by the substitution $f(x) \to f(-x)$. The theorem is usually stated in the form (A.12) for functions f supported on $[0,\infty)$; then the integral is $\int_0^t f(t-s)\,\mathrm{d}U(s)$. However, the proof in Feller [14, Section XI.1] applies to the more general form above as well. (The proof is based on approximations with step functions and the special case when $f(x)$ is an indicator fuction of an interval; the latter case is known as *Blackwell's renewal theorem.*) In fact, a substantially more general version of (A.12), where also the increments X_k may take negative values, is given in [2, Theorem 4.2].

Part (ii) follows similarly (and more easily) from the fact that the measure $\mathrm{d}U$ is concentrated on $\{kd : k \geq 0\}$, and thus $\int_0^\infty f(s-t)\,\mathrm{d}U(s) - \frac{1}{\mu}\psi(t) = \sum_{k=-\infty}^\infty f(kd-t)(\mathrm{d}U\{kd\} - d/\mu)$ together with the renewal theorem $\mathrm{d}U\{kd\} - d/\mu \to 0$ as $k \to \infty$. The Fourier coefficient calculation in (A.17) is straightforward and standard. $\square$

Finally, we consider a situation where we are given also another sequence $Y_1, Y_2, \dots$ of random variables such that the pairs (X_i, Y_i), $i \geq 1$, are i.i.d., while Y_i and X_i may be (and typically are) dependent on each other. (Y_i need not be positive.) We denote the means by $\mu_X := \mathbb{E}\, X_1$ and $\mu_Y := \mathbb{E}\, Y_1$; thus $\mu_X = \mu$ in the earlier notation, and we assume as above that $0 < \mu_X < \infty$. We also suppose that X_0 is independent of all (X_i, Y_i), $i \geq 1$. Let $V_n := \sum_{i=1}^n Y_i$.

Theorem A.8. *Suppose that* $\sigma_X^2 := \operatorname{Var} X_1 < \infty$ *and* $\sigma_Y^2 := \operatorname{Var} Y_1 < \infty$, *and let*

$$\widehat{\sigma}^2 := \operatorname{Var}(\mu_X Y_1 - \mu_Y X_1).$$

Then

$$\frac{V_{\widehat{\nu}(t)} - \frac{\mu_Y}{\mu_X} t}{\sqrt{t}} \xrightarrow{\mathrm{d}} N\Bigl(0, \frac{\widehat{\sigma}^2}{\mu_X^3}\Bigr).$$

If $\widehat{\sigma}^2 > 0$, *this can also be written as*

$$V_{\widehat{\nu}(t)} \sim \operatorname{AsN}\left(\frac{\mu_Y}{\mu_X} t, \frac{\widehat{\sigma}^2}{\mu_X^3} t\right).$$

Note that the special case $Y_i = 1$ yields (A.7).

Proof. For $X_0 = 0$, and thus $\widehat{\nu}(t) = \nu(t)$, this is Gut [16, Theorem 4.2.3]. The general case follows by the same proof, or by conditioning on X_0. □

Remark A.9. Again, we can allow $X_0 = X_0^{(n)}$ to depend on n, as long as the $X_0^{(n)}$ is tight.

References

[1] S. Asmussen, *Applied Probability and Queues.* Wiley, Chichester, 1987.

[2] K. B. Athreya, D. McDonald & P. Ney, Limit theorems for semi-Markov processes and renewal theory for Markov chains. *Ann. Probab.* **6** (1978), no. 5, 788–797.

[3] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, 1968.

[4] N. Broutin & C. Holmgren, The total path length of split trees. *Ann. Appl. Probab.*, to appear. `arXiv:1102.2541.`

[5] J. Bourdon, Size and path length of Patricia tries: dynamical sources context. *Random Structures Algorithms* **19** (2001), no. 3–4, 289–315.

[6] C. Christophi & H. Mahmoud, On climbing tries. *Probab. Engrg. Inform. Sci.* **22** (2008), no. 1, 133–149.

[7] J. Clément, P. Flajolet & B. Vallée, Dynamical sources in information theory: a general analysis of trie structures. *Algorithmica* **29** (2001), no. 1–2, 307–369.

[8] F. Dennert & R. Grübel, Renewals for exponentially increasing lifetimes, with an application to digital search trees. *Ann. Appl. Probab.* **17** (2007), no. 2, 676–687.

[9] L. Devroye and S. Janson, Long and short paths in uniform random recursive dags. *Ark. Mat.* **49** (2011), no. 1, 61–77.

[10] M. Drmota, Y. Reznik, S. Savari & W. Szpankowski, Precise asymptotic analysis of the Tunstall code. *Proc. 2006 International Symposium on Information Theory (Seattle, 2006)*, 2334–2337.

[11] M. Drmota, Y. A. Reznik & W. Szpankowski, Tunstall code, Khodak variations, and random walks. *IEEE Trans. Inform. Theory* **56** (2010), no. 6, 2928–2937.

[12] M. Drmota & W. Szpankowski, On the exit time of a random walk with positive drift. *Proceedings, 2007 Conference on Analysis of Algorithms, AofA 07 (Juan-les-Pins, 2007)*, Discrete Math. Theor. Comput. Sci. Proc. **AH** (2007), 291–302.

[13] G. Fayolle, P. Flajolet, M. Hofri & P. Jacquet, Analysis of a stack algorithm for random multiple-access communication. *IEEE Trans. Inform. Theory* **31** (1985), no. 2, 244–254.

[14] W. Feller, *An Introduction to Probability Theory and its Applications, Volume II.* 2nd ed., Wiley, New York, 1971.

[15] A. Gut, *Probability: A Graduate Course.* Springer, New York, 2005.

[16] A. Gut, *Stopped Random Walks.* 2nd ed., Springer, New York, 2009.

[17] C. Holmgren, Novel characteristics of split trees by use of renewal theory. Preprint, 2010. `arXiv:1005.4594.`

[18] P. Jacquet & M. Régnier, Trie partitioning process: limiting distributions. *CAAP '86 (Nice, 1986)*, 196–210, Lecture Notes in Comput. Sci., 214, Springer, Berlin, 1986.

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

[20] P. Jacquet & M. Régnier, New results on the size of tries. *IEEE Trans. Inform. Theory* **35** (1989), no. 1, 203–205.

[21] P. Jacquet & W. Szpankowski, Analysis of digital tries with Markovian dependency. *IEEE Trans. Inform. Theory*, **37** (1991), no. 5, 1470–1475.

[22] S. Janson, One-sided interval trees. *J. Iranian Statistical Society* **3** (2004), no. 2, 149–164.

[23] S. Janson, Rounding of continuous random variables and oscillatory asymptotics. *Ann. Probab.* **34** (2006), no. 5, 1807–1826.

[24] S. Janson, T. Łuczak & A. Ruciński, *Random Graphs.* Wiley, New York, 2000.

[25] H. Kesten, Renewal theory for functionals of a Markov chain with general state space. *Ann. Probab.* **2** (1974), 355–386.

[26] D.E. Knuth, *The Art of Computer Programming. Vol. 3: Sorting and Searching.* 2nd ed., Addison-Wesley, Reading, Mass., 1998.

[27] I. Kontoyiannis, Second-order noiseless source coding theorems. *IEEE Trans. Inform. Theory* **43** (1997), no. 4, 1339–1341.

[28] M. R. Leadbetter, G. Lindgren & H. Rootzén, *Extremes and Related Properties of Random Sequences and Processes.* Springer-Verlag, New

York, 1983.

[29] G. Louchard & H. Prodinger, Asymptotics of the moments of extreme-value related distribution functions. *Algorithmica* **46** (2006), no. 3-4, 431–467.

[30] H. Mahmoud, *Evolution of Random Search Trees*, Wiley, New York, 1992.

[31] H. Mahmoud, Imbalance in random digital trees. *Methodol. Comput. Appl. Probab.* **11** (2009), no. 2, 231–247.

[32] H. Mohamed & P. Robert, A probabilistic analysis of some tree algorithms. *Ann. Appl. Probab.* **15** (2005), no. 4, 2445–2471.

[33] H. Mohamed & P. Robert, Dynamic tree algorithms. *Ann. Appl. Probab.* **20** (2010), no. 1, 26–51.

[34] B. Pittel, Asymptotical growth of a class of random trees. *Ann. Probab.* **13** (1985), no. 2, 414–427.

[35] B. Pittel, Paths in a random digital tree: limiting distributions. *Adv. in Appl. Probab.* **18** (1986), no. 1, 139–155.

[36] B. Rais, P. Jacquet & W. Szpankowski, Limiting distribution for the depth in PATRICIA tries. *SIAM J. Discrete Math.* **6** (1993), no. 2, 197–213.

[37] M. Régnier, Trie hashing analysis, *Proc. Fourth Int. Conf. Data Engineering (Los Angeles, 1988)*, IEEE, 1988, pp. 377–387.

[38] S. A. Savari, Renewal theory and source coding. *Proc. IEEE*, **88** (2000), no. 11, 1692–1702.

[39] S. A. Savari and R. G. Gallager, Generalized Tunstall codes for sources with memory. *IEEE Trans. Inform. Theory*, **43** (1997), no. 2, 658–668.

[40] W. Szpankowski, Patricia tries again revisited. *J. Assoc. Comput. Mach.* **37** (1990), no. 4, 691–711.

[41] W. Szpankowski, *Average Case Analysis of Algorithms on Sequences.* Wiley, New York, 2001.

[42] W. Szpankowski, Average redundancy for known sources: ubiquitous trees in source coding. *Proceedings, Fifth Colloquium on Mathematics and Computer Science (Blaubeuren, 2008)*, Discrete Math. Theor. Comput. Sci. Proc. **AI** (2008), 19–58.

Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden

E-mail address: `svante.janson@math.uu.se`

URL: `http://www.math.uu.se/~svante/`