Philipp Godland

Markov renewal theory in the analysis of random strings and iterated function systems

2020

Mathematik

Markov renewal theory in the analysis of random strings and iterated function systems

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften im Fachbereich
Mathematik und Informatik
der Mathematisch-Naturwissenschaftlichen Fakultät
der Westfälischen Wilhelms-Universität Münster

vorgelegt von
Philipp Godland
aus Dortmund
- 2020 -

Dekan:	Prof. Dr. Xiaoyi Jiang
Erster Gutachter:	Prof. Dr. Gerold Alsmeyer
Zweiter Gutachter:	Prof. Dr. Sebastian Mentemeier
Tag der mündlichen Prüfung:	15.05.2020
Tag der Promotion:	15.05.2020

Danksagung

Summary

This thesis is dedicated to the promotion of the use of renewal theory in especially two fields in (applied) probability theory: The analysis of random strings and related structures, such as digital trees, and the study of (random) iterated function systems. In both areas, we are going to study models involving Markov modulation by some positive recurrent driving Markov chain on a (at most) countable state space $\mathcal{S}$. Discrete Markov modulation allows for an elegant usage of cyclic decomposition to lead the Markov model back to a corresponding independent model.

In Part I, we examine tree structures that appear in the analysis of algorithms and which are constructed from a set of random strings. These strings are assumed to form a Markov chain. As a consequence, an auxiliary process, that was introduced by Janson in an easier setting, is a Markov-modulated sequence, and thus admits the use of Markov renewal theory. We employ this model as an example of how the use of Markov renewal theory may provide an intuitive approach to problems in which it has been sparsely used so far. Therefore, in the first half of this part, we re-derive several limit theorems for characteristic trie parameters such as the depth, but also find new results, e.g. for the imbalance factor. In the second half, we develop a device for the average-case analysis of further characteristic parameters and provide a probabilistic proof.

In Part II, we investigate iterated function systems of Markov-modulated Lipschitz maps. We explain how our model fits in the stationary model of Elton and use our extra structural knowledge to prove polynomial and geometric convergence rates of the system in two regimes of moment conditions. Again, we apply Markov renewal theory. In particular, we use cyclic decomposition and controlling of the error that we make, when considering the occurring subsequences instead of the original process.

Contents

I. Renewal theory in the analysis of random digital trees 1

1. Introduction 3
1.1. What has been done? 5
1.2. Structure 8

2. Preliminaries 9
2.1. Tries and trie-like structures 9
2.2. Input models 10
2.3. Parameters 12
2.4. Type of main results 13
2.5. Markov renewal theory and notation 13
2.6. Central Markov-modulated sequence 15
2.7. Distributional identity 16
2.8. Lattice 17
2.9. Null-homology 22

3. Asymptotic analysis of depth and imbalance factor 25
3.1. Results for D_n 25
3.2. Results for Δ_n 27
3.3. Proofs 28
3.3.1. Weak law of large numbers for D_n 29
3.3.2. L^p-law of large numbers for D_n 31
3.3.3. Asymptotic expansion of the mean of D_n 33
3.3.4. Central limit theorem for D_n 40
3.3.5. Weak law of large numbers for Δ_n 48
3.3.6. L^p-law of large numbers for Δ_n 49
3.3.7. Central limit theorem for Δ_n 50

4. Average-case analysis of further characteristic parameters 55
4.1. Main results 56
4.2. Poisson model 59
4.3. Markov renewal theorems 60
4.4. Proofs of main results 64
4.5. Applications 69
4.5.1. Expected size of a trie 70
4.5.2. Expected size of a b-trie 72
4.5.3. Expected size of a PATRICIA-trie 76
4.5.4. Expected depth in a PATRICIA-trie 81
4.5.5. Expected external path length of a trie 86
4.5.6. Expected external path length of a PATRICIA-trie 87

A. Appendix 91
A.1. Properties of $(X_0^{(n)})_{n\geq 2}$ . . . 91
A.2. Depoissonization for stochastically monotone parameters . . . 92
A.3. Auxiliary convergence result . . . 94

II. Convergence rates of iterated function systems of Markov-modulated Lipschitz maps by regenerative methods 95

5. Introduction 97
5.1. Iterated function systems . . . 97
5.2. Structure . . . 99

6. Preliminaries 101
6.1. Measurability and remarks . . . 101
6.2. Dual chain and distributional identity . . . 102
6.3. Elton's theorem . . . 104
6.4. Markov renewal theory and conditions . . . 106

7. Convergence rates in Elton's theorem 111
7.1. Main results . . . 111
7.2. Proof of Theorem 7.1 by cyclic decomposition . . . 112
7.3. Convergence rate under polynomial-type moment conditions . . . 120
7.4. Convergence rate under geometric-type moment conditions . . . 131

B. Appendix 139
B.1. Kingman's subadditive ergodic theorem . . . 139
B.2. Prokhorov metric . . . 139
B.3. Some results from the i.i.d. case . . . 140
B.4. Moment results in Situations (1) and (2) . . . 142

References 145

Acronyms 153

List of symbols 153

Part I.

Renewal theory in the analysis of random digital trees

1. Introduction

Renewal theory and the use of regenerative processes have by now become standard tools in probability theory. But especially in the probabilistic analysis of digital trees, it is underrepresented, at least in our opinion, as it allows for easy, intuitive and wholly probabilistic approaches to problems which are usually treated with combinatorial and analytical methods.

Analysis of digital trees. Digital trees play an important role in the analysis of algorithms especially on words in theoretical computer science. They bear complexity characteristics of the corresponding algorithms which use them. Hence, the study of these trees enables us to compare the performance of different algorithms, designed for the same purpose, via the chosen characteristic, when we have a large amount of underlying data. In the beginnings of the study of algorithms, a lot of work was done to analyze the worst-case performance. However, this often leads to dealing with atypical inputs, that in a realistic setting would rarely appear. This nuisance motivated the so-called probabilistic analysis (including average-case analysis or distributional analysis) of algorithms, in which the input is randomized and modeled according to some suitable (*typical*) probabilistic model, and the (random) characteristic parameter is sought to be described as precisely as possible, e.g. its expectation is studied, of course accompanied by classical limit theorems such as central limit theorems (CLTs). The Quicksort algorithm, which sorts n numbers, is one famous example in which the worst-case number of comparisons needed is of order n^2 but this rarely appears, whereas the average-case number is of order $n \log n$. Presumably the first to conduct early instances of a structured probabilistic analysis of algorithms was Donald Knuth in [Knu63].

Digital trees. The most commonly known digital trees are tries (from retrieval), PATRICIA-tries (PAT-tries, short for *Practical Algorithm to Retrieve Information Coded in Alphanumeric*), digital search trees (DST) and suffix trees. All of the previously mentioned are recursively built tree structures, that store a set of strings which can be thought of as words or sequences of words. Examples of these trees are shown in Figure 1.1. As already mentioned, tries and the related structures form a visualization of algorithms, in particular searching algorithms and sorting algorithms. E.g. the non-comparative sorting algorithm Radix Sort is also called Triesort, since its mechanism is similar to the construction of a corresponding trie. Thus, the performance of Radix Sort is in total correspondence to certain characteristic parameters of the trie. Digital trees also appear in data compression as parsing trees for codes, such as Tunstall or Khodak codes, and in the analysis of the well-known Lempel-Ziv Parsing Scheme among others.

Sources/Input models. There are several input models which aim at describing a *typical* input string. The easiest and best understood, yet not always sensible, is the memoryless source, emitting *letters* independently. Further dependency was allowed and, among others, Markov sources, mixing sources, stationary ergodic sources, sources with density and the very general so-called dynamical sources were studied.

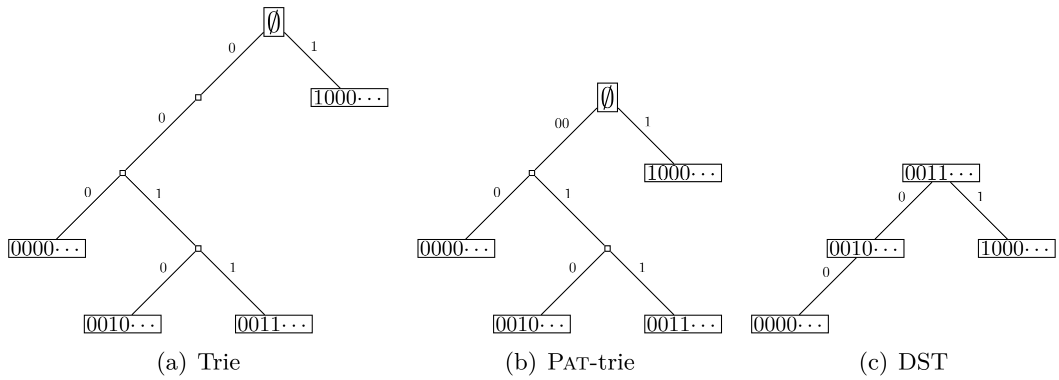

Figure 1.1.: Trie, PAT-trie and DST constructed from strings 0011···, 1000···, 0010··· and 0000···.

Usually, an algorithm processes a set of strings from some set $\mathcal{M}$. Frequently, two models appear in literature: In both models, different strings are produced independently and identically distributed (i.i.d.) by the chosen source. Yet, in the standard model (sometimes called *Bernoulli model*), the number of strings is some fixed integer number n, whereas in the second, the *Poisson model*, this number is random, viz. Poisson distributed. Often, asymptotics are derived for the Poisson model, which can be easier due to the various helpful properties of Poisson processes, and the corresponding results for the standard model are deduced from these afterwards. In the first chapter, we will not need the Poisson model, however the main device in the second chapter is built for the use in the Poisson model. Of course, there are other models allowing for dependence between the input strings such as in the suffix tree, where a trie is built from the successive prefixes of a given string. We will not deal with those models in this thesis.

Characteristic parameters. There are plenty of interesting characteristic parameters in digital trees. We take a selection from [Szp01] with minor supplements:

- The *depth* of a given string, i.e. the length of the path from the root to the (external) node containing that string.
- The *insertion depth*, i.e. the depth of the last inserted string.
- The *typical depth*, i.e. the depth of a randomly selected string.
- The *average depth*, i.e. the arithmetic mean of all depths.
- The *external path length*, i.e. the sum of the depths of all strings.
- The *height*, i.e. the largest depth.
- The *size*, i.e. the number of (internal) nodes.
- The *profile*, i.e. the number of nodes at each level.
- The *imbalance factor* of a given binary string, i.e. the number of steps to the right minus the number of steps to the left in the path from the root to the (external) node containing that string.

Obviously, most of the stated parameters measure complexity of the corresponding tree or the underlying algorithm in some way. E.g. the depth of a given string can be seen as the searching cost for that string, the external path length is the construction cost of the structure, in particular, the external path length of a trie is the number of so-called bucket operations that are performed by the corresponding Radix Sort algorithm.

Main goal. We will not deal with all of these digital trees, input models and parameters in this thesis, indeed in the spirit of Janson in [Jan12a], we will try to give a consistent collection of applications of Markov renewal theory to random strings and tries (and trie-like structures), and to convey the possibility of easily obtaining Markov model analogues for results (not only concerning digital trees) that have been established by standard renewal theory.

In his great survey, Janson showed how standard renewal theory can be used in a variety of applications, e.g. to derive asymptotics for parameters of random strings or tries while restricting himself to the basic situation that only strings in the alphabet $\{0,1\}$ are considered, where the *letters* or *bits* ξ_i in the strings are i.i.d. or so-called *memoryless sources*.

One purpose of this part is to follow that approach in the analysis of tries and to broaden it in an intuitive way by dealing with essentially the same (but not all) applications and additionally allowing the ξ_i to form a Markov chain, i.e. being generated by a *Markov source*. This requires an *upgrade* of the methods in use from standard renewal theory to Markov renewal theory. Nevertheless, it has long been known, cf. [Als14] for a survey of important results, that Markov renewal theory with a driving Markov chain on a discrete state space can be nicely related to standard renewal theory by cyclic decomposition. Hence, the analysis of tries, where we usually deal with a finite alphabet, appears to be an ideal playground for illustrating once more a generalization of results involving renewal theory for i.i.d. variables to those allowing for some dependence. Moreover, the transfer to Markov renewal theory conserves the relevance of arithmeticity in the average-case analysis for possible oscillatory terms and a probabilistic explanation for them.

Furthermore, we want to show that Markov renewal theory may provide intuitive probabilistic approaches to fields in which the usual methods are analytical. The analysis of algorithms is such a field. In the next section, we mention earlier publications and the methods that are commonly used in these.

1.1. What has been done?

Renewal theory. In fact, there are a few examples, where some instances of Markov renewal theory have already been used in the analysis of digital trees. In [SG97], [Sav99] and [Sav00] Savari and Gallager studied generalized variable-to-fixed length codes such as a generalized version of the Tunstall code for so-called unifilar Markov sources with the use of renewal theory. The simple forms of those variable-to-fixed length codes such as the Khodak and Tunstall codes have been studied by Janson as an illustration in [Jan12a] for the case of memoryless sources.

Generally, the use of renewal theory in this area is seldom found outside of [Jan12a]. Janson himself gives a list of examples that use standard renewal theory: [BH12], [DG07], [DJ11], [Hol12], [Jan04], [MR10] and [MR05]. There is also the unfinished work [Jan12b] extending [Jan12a].

Due to our extensive use of Markov renewal theory, we state the main sources of our methodology: The papers [AMN78] and [Kes74] by Athreya, McDonald and Ney and Kesten are classical for driving chains with a general state space, for the discrete driving chain setting we also suggest

[Çin69] by Çinlar, [Asm03] by Asmussen and, e.g. [Als14] by Alsmeyer. Of course, this list is by no means a complete list of important results in Markov renewal theory.

We also refer to the powerful [MT93] for the comprehensive treatment of Markov chains and to [Gut09] for a range of helpful results on stopped random walks.

Methods in the literature. With our renewal-theoretic approach we will derive results, most of them have already been found (sometimes more precisely) by other techniques. Most frequently used techniques are analytic because they "are extremely powerful and when they apply, they often yield estimates of unparalleled precision", as Szpankowski cites Andrew Odlyzko in the preface of his book [Szp01]. Therein, he gives an overview of both some probabilistic and combinatorial techniques such as *Inclusion-Exclusion Principle* and *Moment Methods*, and various analytic techniques such as *Generating Functions*, *Singularity Analysis*, *Saddle Point Method*, *Mellin Transforms* and *Poissonization* and *Depoissonization*. A huge treatment of most of the latter methods and lots more from *Analytical Combinatorics* is done by Flajolet and Sedgewick in the eponymous book [FS09], a technique which the authors specifically developed for the analysis of algorithms.

There are more probabilistic methods to add here. One that was frequently used, is the *Contraction Method*, which was developed by Rösler in [Rös91] to characterize the limit distribution of the standardized complexity of Quicksort. There have been numerous extensions e.g. by Rösler [Rös92], Neininger and Rüschendorf [NR04] and Neininger and Sulzbach [NS15] just to name a few. More references can be found in the PhD thesis of Leckey [Lec15], which itself uses a generalized Contraction Method. Also, as already mentioned, renewal theory belongs to the list of probabilistic methods but has not been used as extensively as the others.

Results on random digital trees. Here we give a short (and almost surely incomplete) overview of some results for parameters that will appear in the following chapters. We focus on results concerning tries, PAT-tries and b-tries, which are tries with a storage capacity of $b \geq 1$ strings in a node ($b = 1$ matches normal trie). We cover the four models that have been studied most.

The easiest scheme for modeling sources is the **memoryless source model**. The so-called unbiased model, where the letters are equiprobable, is the first that was investigated, in fact a precise asymptotic expansion of the expected depth is already found in [Knu73]. An asymptotic expansion of the variance in this setting can first be found in [KP88]. The authors also deal with the expected depth in PAT-tries and DSTs as well as with the variances. In [KP86], the same is done for the external path length. For general memoryless sources, results on the asymptotic distribution of the depth and the height were first developed independently by Pittel [Pit86] and Jacquet and Régnier [JR86b]. At that time, Pittel had already derived weak and strong laws of large numbers for depth and height, respectively, in the fairly more general setup of a stationary ergodic source in [Pit85]. In [JR86a], Jacquet and Régnier added a limit law for the size to [JR86b] and in [JR88] they extended their results towards a limit law for the size and path length of b-tries. To derive an asymptotic expansion of the variance of the depth, Szpankowski provided a formula for all factorial moments of the depth in [Szp88]. These results include b-tries. Precise asymptotics of the variance of the size of tries and b-tries was given by Régnier and Jacquet in [RJ89]. The depth of PAT-tries is studied by Szpankowski in [Szp90] and by Rais, Jacquet and Szpankowski in [RJS93]. The former provides, in particular, an asymptotic expansion for the variance, the latter establishes a limit law.

Of course (as it initiated our work), several parameters such as the depth and the size of a trie and PAT-trie, as well as the size of a b-trie and the path length of a trie, were re-derived by

Janson in [Jan12a] and [Jan12b] to motivate the usage of renewal theory in this kind of analysis. Therein, he also gave an asymptotic expansion for the number of external nodes in a b-trie, in which exactly $1 \leq l \leq b$ strings are stored, and a limit law for the imbalance factor in a binary trie. The imbalance factor was first analyzed by Mahmoud [Mah08], and he already derived an asymptotic expansion of the mean and the limit law. There are several other interesting parameters concerning tries, so we briefly state two more and some corresponding references: The profile is studied in [Par+08] or [Par+09], and protected nodes are considered in [Jan12b], [Gai+12], [GW13] and [FLY16]. We refer to [Knu98], [Mah92], [Szp01] and [Drm09] for a survey of these and other parameters and corresponding results.

The **Markov source model** appears to be studied to a much lesser extent than the memoryless model. Probably the first was Régnier [Rég88] with her result about asymptotical linearity of the expected size in a trie and b-trie. Mainly aiming at a corresponding result for suffix trees, Jacquet and Szpankowski [JS16] sketched a re-derivation (and improvement) of Régnier's result. Concerning the depth, Jacquet and Szpankowski [JS91] derived precise asymptotics for expectation and variance such as a limit law. They require the source to be stationary and base their analysis on the inclusion-exclusion rule. In the same setting, Szpankowski [Szp91] derived a weak law of large numbers (WLLN) for the height. Finally, Leckey, Neininger and Szpankowski [LNS15] provided a CLT for the external path length in a binary trie using a generalized Contraction Method and moment transfer techniques. The proof also applies to the external path length in binary PAT-tries and DSTs, cf. [Lec15].

Results for the **density model** are due to Devroye. He studied random tries (and also PAT-tries and DST in [Dev92]) built from the binary expansion of i.i.d. numbers in $[0, 1]$ with density f. In [Dev82], he derives the first order asymptotics for the expectation of depth and average depth (and thus of the external path length), which is complemented by the first order asymptotics of the expected height, the asymptotic distribution of the height and an asymptotic upper bound for the expected size in [Dev84]. In [Dev92], he improves his results once more by providing limit laws for both the depth and the height, as well as a strong convergence result for both. He also gives a large deviation result and several laws of the iterated logarithm for the height, and he investigates the improvement when using a PAT-trie instead of trie. There are some notes concerning DSTs, too.

The most general model containing all of those focused upon here is the **dynamical source model**. The model was introduced in [Val01] and [Clé00] by Vallée and her PhD student Clément. Together with Flajolet [CFV01], they analyzed the expected size and path length and performed a distributional analysis of the height. The average-case analysis of size and path length was complemented by Bourdon [Bou01] with corresponding results for PAT-tries. However, as Cesaratto and Vallée stated in [CV15], the former results and analyses in [Bou01] and [CFV01] have not been done entirely precisely and needed supplementary results, e.g. from [FRV10] for completion. Cesaratto and Vallée, among others, developed the notion of *tameness*, re-derived asymptotic expansions for expectation and variance of the depth for *simple sources* (i.e. memoryless, and irreducible and aperiodic Markov) in their tameness setup, and derived a CLT for a class of dynamical sources notably *not* containing any simple source. This was complemented by Hun and Vallée in [HV14] and [Hun14], where they proved a CLT for all tame sources not too similar to the unbiased memoryless source. Besides, it should be mentioned that not every (irreducible) Markov source is simple, and not every simple source is tame in the sense of the above-mentioned paper.

We ignored almost all results for one of the close relatives of the trie, the DST. Although there exist corresponding results, their discussion would go beyond the scope of this thesis.

1.2. Structure

We start with some preliminaries in Chapter 2 to set up the general situation and to provide the usual notation and basic results from Markov renewal theory that are used throughout the work. In addition, we introduce the depth D_n and the imbalance factor Δ_n of a random string in a trie which, as in [Jan12a], will be in the center of our analysis up to (and excluding) Chapter 4. We prove a central distributional identity in Section 2.7, that connects the depth and the imbalance factor with renewal-theoretic objects. A recurring issue will be the lattice type of the considered processes, to which we dedicate Section 2.8. Section 2.9 gives a short overview of null-homology of Markov random walks and how it enters in our results.

In Chapter 3, we state the main results concerning the asymptotics of the depth and imbalance factor and give the proofs in Section 3.3. The results cover laws of large numbers, CLTs and asymptotic expansions of the expectation (especially for the depth). At the end of Section 3.1, we provide references and comparisons to existing results concerning the depth.

In Chapter 4, we derive a general device for the average-case analysis of several further trie (and b-trie and PAT-trie) parameters, to subsequently give multiple applications including the size and the external path length in Section 4.5. We also provide references there.

Some auxiliary results are collected in Appendix A.

2. Preliminaries

Throughout this first part, we examine random infinite strings

$$\Xi = \xi_1 \xi_2 \cdots = (\xi_n)_{n \geq 1},$$

i.e. random variables on $(\Omega, \mathcal{A}, \mathbb{P})$ taking values in $\mathcal{S}^{\mathbb{N}}$, where $\mathcal{S}$ is a finite set with $\#\mathcal{S} = m \geq 2$ called *alphabet*. Sometimes we use $\mathcal{S} = \{0, 1\}$ for simplicity. Denote by $\mathcal{S}^*$ the complete infinite tree

$$\mathcal{S}^* = \bigcup_{n \geq 0} \mathcal{S}^n = \{\emptyset\} \cup \bigcup_{n \geq 1} \mathcal{S}^n$$

with nodes labeled by finite strings.

2.1. Tries and trie-like structures

We give a brief definition of tries, Pat-tries and b-tries, the most prevalent in this part. Our definition is basically adopted from [Jan12a]. For futher details see [Knu98, Section 6.3], [Mah92, Chapter 5] or [Szp01, Section 1.1]. The notion of a trie makes sense in the non-random situation, so it is defined ω-wise for $\omega \in \Omega$.

Definition 2.1. Let $\mathcal{M}$ be a finite set of (almost surely (a.s.)) pairwise distinct random infinite strings. Then the trie is constructed ω-wise recursively:

(1) If $\mathcal{M} = \emptyset$, then the trie is empty.

(2) If $\#\mathcal{M} = 1$, then the trie only consists of the root and the string is saved at that place.

(3) If $\#\mathcal{M} > 1$, then the trie starts with the root and all strings starting with $i \in \mathcal{S}$ are given to the i-th position on the next level (and create a connection to the root). They form the basis for (4). In this way we obtain m (possibly empty) subtrees.

(4) The subtrees from (3) are treated by cases (1)-(3) recursively, comparing the strings at the next spot next.

(5) The procedure stops as $\mathcal{M}$ is finite.

The resulting tree structure is denoted by $\text{Trie}(\mathcal{M})(\omega)$ for $\omega \in \Omega$. It can be regarded as a finite subtree of $\mathcal{S}^*$. We call the nodes which have no children *leaves* or *external nodes* and the remaining nodes *internal nodes*.

At this place, we give some observations concerning the trie's structure. Following [Jan12b], we write $\beta \succ \alpha$ for two strings (which may be finite) if β starts with α. Given strings $\Xi^{(1)}, \Xi^{(2)}, \ldots$, we write $\mathcal{M}_n := \{\Xi^{(1)}, \ldots, \Xi^{(n)}\}$.

Observation 2.2. The following holds for $\mathrm{Trie}(\mathcal{M}_n)$, $n \geq 0$:

(a) Every external node stores exactly one string.

(b) There are exactly n external nodes.

(c) The internal nodes do not store any string.

(d) Each internal node has one to m children.

(e) A node $\alpha \in \mathcal{S}^*$ is an internal node iff there exist at least two strings starting with α.

(f) Let $k \geq 1$. Then $\alpha_1 \cdots \alpha_k \in \mathcal{S}^*$ is an external node iff there is precisely one string starting with $\alpha_1 \cdots \alpha_k$ and at least one more string starting with $\alpha_1 \cdots \alpha_{k-1}$.

Although, we primarily deal with the ordinary trie in the first chapter, we introduce two easy variations, that have already been mentioned before. An obvious generalization of the trie from Definition 2.1, which only stores one string in each leaf, is a similar tree which stores up to $b \geq 1$, $b \in \mathbb{N}$, strings in each leaf (and no strings in the internal nodes). It is called *b-trie*.

Definition 2.3. Let $\mathcal{M}$ be a finite set of (a.s.) pairwise distinct random infinite strings. Then the b-trie is constructed ω-wise recursively:

(1) If $\mathcal{M} = \emptyset$, then the b-trie is empty.

(2) If $\#\mathcal{M} \leq b$, then the b-trie only consists of the root and all strings are saved at this place.

(3) If $\#\mathcal{M} > b$, then the b-trie starts with the root and all strings starting with $i \in \mathcal{S}$ are given to the i-th position on the next level (and create a connection to the root). They form the basis for (4). In this way we obtain m (possibly empty) subtrees.

(4) The subtrees from (3) are treated by cases (1)-(3) recursively, comparing the strings at the next spot next.

(5) The procedure stops as $\mathcal{M}$ is finite.

The resulting tree structure is denoted by $\mathrm{Trie}^{(b)}(\mathcal{M})(\omega)$ for $\omega \in \Omega$. It can also be regarded as a finite subtree of $\mathcal{S}^*$ with nodes labeled by finite strings.

Both, the original trie from Definition 2.1 as well as the b-trie from Definition 2.3 can contain long chains of internal nodes if two or more ($b+1$ or more) strings differ late for the first time. This makes the trie unnecessarily large. PAT-tries eliminate such nodes to compress data, see Figure 1.1 for a realisation of a PAT-trie.

Definition 2.4. Given $\mathrm{Trie}(\mathcal{M})$ from Definition 2.1. Then $\mathrm{Trie}^P(\mathcal{M})$ is the random tree resulting from $\mathrm{Trie}(\mathcal{M})$ after eliminating all nodes with exactly one child.

Remark 2.5. Note that there can be a node α whose parent node has a length not equal to the length of α minus 1. Obviously, the external nodes in $\mathrm{Trie}^P(\mathcal{M})$ and $\mathrm{Trie}(\mathcal{M})$ coincide and only the internal nodes differ.

2.2. Input models

We briefly assemble the predominant input sources, just to give an impression of how the input sources relate. We will later on focus on the Markov source, but we want to compare our results to the more general dynamical source model.

Memoryless source. This source emits i.i.d. letters ξ_i with

$$\mathbb{P}(\xi_1 = j) = p_j \in (0, 1),$$

for $j \in \mathcal{S}$. Letters j with $p_j = 0$ could otherwise be excluded from $\mathcal{S}$ and in case of $p_j = 1$ the source is not worth studying. Memoryless sources are called *symmetric* or *unbiased* if $p_j = 1/m$ for all $j \in \mathcal{S}$, and *asymmetric* or *biased* otherwise. To built a trie and study its asymptotics, a sequence of i.i.d. copies of Ξ is considered. We have seen, that the construction of a trie requires the strings to be distinct, in this model they obviously are $\mathbb{P}$-a.s. The Markov analogue of the following parameter runs like a common thread through the first part of this work: The *entropy*

$$H = -\sum_{j \in \mathcal{S}} p_j \log p_j$$

of the letters is a measure of balance for the corresponding trie and in fact, $H = \mathbb{E}X_1$, if we define $X_1 = -\log p_{\xi_1}$, cf. [Jan12a], a finding that we will generalize and apply later on.

Markov source. In the Markov setting, we need an initial distribution so we introduce ξ_0 as the initial random variable for the Markov chain $\Xi_0 := (\xi_n)_{n\geq 0}$ with transition matrix $P = (p_{i,j})_{i,j\in\mathcal{S}}$. We assume that Ξ_0 is time-homogeneous, irreducible (and thus positive recurrent since $\mathcal{S}$ is finite) and has no atoms. In particular, Ξ_0 has a unique stationary distribution π. We also exclude the case that $p_{i,j} = 1/|\mathcal{S}|$ for all $i, j \in \mathcal{S}$, which is contained in the i.i.d. setup and already extensively studied. Besides, we will see in Remark 2.11 (and also in Section 2.9), that a central object of our study is trivial in this case.

Remark 2.6. Every positive recurrent discrete Markov chain satisfies $\pi_i > 0$ for all $i \in \mathcal{S}$.

Remark 2.7. Throughout the text, $p_{i,j} = 0$ and $p_{i,j} = 1$ will be allowed, unless irreducibility is violated or the chain has atoms, which in particular excludes absorbing states (i.e. $p_{i,i} = 1$ for some $i \in \mathcal{S}$) and purely deterministic chains. These requirements ensure that two or more i.i.d. copies of the same string are distinguishable (cf. Remark 2.8) and hence the construction of a trie is possible. Thus, there is at least one $i \in \mathcal{S}$ such that $p_{i,j} < 1$ for all $j \in \mathcal{S}$.

We set $\mathbb{P}_i := \mathbb{P}(\cdot \mid \xi_0 = i)$ for all $i \in \mathcal{S}$ and

$$\mathbb{P}_\lambda := \sum_{i\in\mathcal{S}} \lambda_i \mathbb{P}_i$$

for an arbitrary distribution λ on $\mathcal{S}$. Then $\mathbb{P}_\lambda^{\xi_0} = \lambda$. We further define

$$P_i(\alpha_1 \cdots \alpha_n) := p_{i,\alpha_1} p_{\alpha_1,\alpha_2} \cdots p_{\alpha_{n-1},\alpha_n} = \mathbb{P}_i(\xi_1 = \alpha_1, \ldots, \xi_n = \alpha_n)$$

as a function on $\mathcal{S}^* := \bigcup_{k=0}^{\infty} \mathcal{S}^k$ with $P_i(\emptyset) := 1$ for all $i \in \mathcal{S}$. Note that $P_i(\alpha) = 0$ for a finite string α is allowed. Throughout the work, Ξ_0 or Ξ serves as the generic random infinite string. If not stated otherwise, $\Xi_0 = (\xi_n)_{n\geq 0}$ and $\Xi = (\xi_n)_{n\geq 1}$ behave in the above manner.

In the next step, we consider a sequence $\Xi (= \Xi^{(1)}), \Xi^{(2)}, \ldots$ of i.i.d. (w.r.t. every $\mathbb{P}_i$) random infinite strings which means that under each $\mathbb{P}_i$ the Markov chains $\Xi^{(2)}, \Xi^{(3)}, \ldots$ are distributed like the generic string Ξ, have the same initial state $\xi_0 = i$ and all strings or chains are independent. This model will be in force starting from Section 2.3.

Remark 2.8. Let $n \in \mathbb{N}$, then $\Xi^{(1)}, \ldots, \Xi^{(n)}$ are a.s. pairwise distinct since $\Xi^{(1)}$ has no atoms w.r.t. every $\mathbb{P}_i$.

Dynamical source. The rather abstract dynamical source model essentially consists of two defining components, cf. [CFV01] for further details: A density f on $[0,1]$ and a mechanism M that assigns $x \in [0,1]$ to a string $(\xi_n)_{n\geq 0} \in \mathcal{S}^{\mathbb{N}}$. E.g. if this mechanism is the binary expansion of x, then this is the **density model**, where the value x and thus the corresponding binary string is drawn according to the density f. The name *dynamical source* originates from the underlying dynamical system

$$M(x) = (\sigma x, \sigma T x, \sigma T^2 x, \ldots),$$

where T is a shift map and σ decodes a real value to a symbol. Every Markov source (and thus every memoryless source) can be associated to such a dynamical source, cf. [CFV01, Section 1.2] or [CV15, Section 4.2]. The input strings are then i.i.d. copies of $M(X)$ where X has density f.

2.3. Parameters

There are two of the above-mentioned parameters, that we focus on in the first chapter: The depth and the imbalance factor. Both will be formally introduced below. The reason why we choose these two is that, as in [Jan12a], the depth is, roughly speaking, distributionally equal to the first passage time of some underlying (Markov) renewal process. Furthermore, the imbalance factor is connected to the depth in a very easy way. Hence, all kinds of distributional limit theorems and asymptotics can be derived directly from Markov renewal theory. We will give a more precise version of the stated distributional identity in Section 2.7.

(Insertion) Depth. We now introduce the *depth* D_n of the generic string $\Xi = \Xi^{(1)}$ in Trie($\mathcal{M}_n$) as the depth of the node which stores Ξ, i.e. the path length from the root to that node. In other words D_n is the index where the chosen string differs from all other $n-1$ strings for the first time:

$$D_n := \min\{k \geq 0 \mid (\xi_1, \ldots \xi_k) \neq (\xi_1^{(j)}, \ldots \xi_k^{(j)}) \text{ for } j = 2, \ldots, n\}.$$

This definition is only interesting for $n \geq 2$, and trivially $D_1 = 0$. Obviously, D_n is increasing.

Remark 2.9. The definition of D_n as the depth of the first string is not arbitrary as, due to the formerly mentioned independence and identical distribution properties, every string has a depth with the same distribution (w.r.t. $\mathbb{P}_i$). However, the depths of two or more strings in the same trie are dependent. Concerning distributional properties, the depth of the first string, say, the insertion depth and the typical depth coincide, since for $k \geq 0$

$$\mathbb{P}_i(D_n = k) = \mathbb{P}_i(D_{n,n} = k) = \frac{1}{n}\sum_{l=1}^{n} \mathbb{P}_i(D_{n,l} = k),$$

where $D_{n,l}$ is the depth of the l-th string in Trie($\mathcal{M}_n$).

Imbalance factor. Considering a binary Trie($\mathcal{M}_n$), we define the *imbalance factor* Δ_n of the string $\Xi = \Xi^{(1)}$ which quantifies how much the trie hangs to the left or right. Formally, define for $n \in \mathbb{N}$

$$Y_n := 2\xi_n - 1 = \begin{cases} -1, & \text{if } \xi_n = 0, \\ +1, & \text{if } \xi_n = 1, \end{cases}$$

and $V_k := \sum_{n=1}^{k} Y_n$ with $V_0 := 0$. Then obviously, $\Delta_n := V_{D_n}$ is the right definition. Set

$$\mu_Y := \mathbb{E}_\pi Y_1 = \mathbb{E}_\pi(2\xi_1 - 1) = 2\pi_1 - 1$$

with

$$\pi_0 = \frac{p_{1,0}}{p_{1,0} + p_{0,1}} \quad \text{and} \quad \pi_1 = \frac{p_{0,1}}{p_{1,0} + p_{0,1}}$$

denoting the weights of the binary stationary distribution. In 2004, Donald Knuth proposed the study of imbalance measures for random binary trees, which was then performed by [Mah08] in the i.i.d. setting with the use of poissonization and Mellin transform techniques.

Further parameters. For the sake of completeness, we denote by $L_n := \sum_{i=1}^{n} D_{n,i}$ the *external path length* and by $H_n := \max_{1\le i\le n} D_{n,i}$ the *height*. The *size* W_n is simply the number of internal nodes which will be further specified in Chapter 4.

2.4. Type of main results

As already mentioned before, we will provide an extensive distributional asymptotic analysis for D_n and Δ_n, and an average-case analysis of further trie-related parameters such as the size and the external path length. We only deal with Markov chains of order 1, but similar results should also easily be verifiable for chains depending on the last k symbols, e.g. as then $(\xi_n, \xi_{n+1}, \ldots, \xi_{n+k-1})_{n\ge 0}$ is still a (rather artificial) finite state Markov chain of order 1. We utilize the finiteness of $\mathcal{S}$ a lot in our investigation. Also, many results should carry over to a countable $\mathcal{S}$, as Markov-renewal-theoretic results also exist in this setting. However, some annoying technical issues will almost surely appear in the analysis which is why we restrict ourselves to the yet relevant finite state case.

Arithmeticity plays a big role in the asymptotic expansion of the mean of all parameters under investigation in this part. We learn that, on the one hand, an arithmetic case appears, and on the other hand the non-arithmetic case is not at all *nice*, since the underlying random walk structure can never be spread-out (cf. Section 2.8). Thus, the analysis of tries constitutes an application example where the usual restriction to the spread-out case is never possible.

2.5. Markov renewal theory and notation

Markov renewal theory generalizes standard renewal theory. The objects of investigation in standard renewal theory are random walks (with i.i.d. increments). As a usual step, one tries to allow dependence. One way to achieve this is via a modulating Markov chain that influences the additive component, i.e. the random walk, in a sensible way. This leads to the notion of a *Markov-modulated sequence* (MMS):

Let $(\mathcal{S}, \mathfrak{S})$ be measurable with countably generated σ-field $\mathfrak{S}$ and let further $(\xi_n, X_n)_{n\ge 0}$ be a time-homogeneous Markov chain on $(\mathcal{S} \times \mathbb{R}, \mathfrak{S} \otimes \mathcal{B}(\mathbb{R}))$. Then $(\xi_n, X_n)_{n\ge 0}$ is called MMS if it admits a transition kernel $Q : \mathcal{S} \times (\mathfrak{S} \otimes \mathcal{B}(\mathbb{R})) \to [0, 1]$, i.e. for all $n \ge 0$

$$\mathbb{P}((\xi_{n+1}, X_{n+1}) \in \cdot|\xi_n, X_n) = \mathbb{P}((\xi_{n+1}, X_{n+1}) \in \cdot|\xi_n) = Q(\xi_n, \cdot) \quad \mathbb{P}\text{-a.s.}$$

Equivalently, $\Xi_0 = (\xi_n)_{n\ge 0}$ is time-homogeneous, $X_0, X_1, \ldots$ are conditionally independent given Ξ_0 with $\mathbb{P}(X_0 \in \cdot|\Xi_0) = \mathbb{P}(X_0 \in \cdot|\xi_0)$ $\mathbb{P}$-a.s. and

$$\mathbb{P}(X_n \in \cdot|\Xi_0) = \mathbb{P}(X_n \in \cdot|\xi_{n-1}, \xi_n) = K(\xi_{n-1}, \xi_n, \cdot) \quad \mathbb{P}\text{-a.s.}$$

for $n \geq 1$ and some stochastic kernel $K : \mathcal{S}^2 \otimes \mathcal{B}(\mathbb{R}) \to [0,1]$. We call Ξ_0 the *driving chain.* The corresponding *Markov random walk* (MRW) is (ξ_n, S_n) with $S_n := \sum_{i=0}^n X_i$. It constitutes a natural generalization of an ordinary random walk.

Basic methods and results. In particular, if the driving chain is positive recurrent and discrete (as it will be in our analysis, cf. Section 2.2), then one key method in Markov renewal theory is the cyclic decomposition of the MRW. For that reason, let $(\sigma_n(i))_{n\geq 0}$, $\sigma_0(i) := 0$, be the sequence of successive recurrence times (of the driving chain Ξ_0) of state i, i.e.

$$\sigma_n(i) := \inf\{k > \sigma_{n-1}(i) : \xi_k = i\},$$

and let $(\tau_n(i))_{n\geq 1}$ be the corresponding cycle lengths, i.e. $\tau_n(i) := \sigma_n(i) - \sigma_{n-1}(i)$. Then $(\xi_n, X_{n+1})_{n\geq 0}$ decomposes into i.i.d.

$$Z_n := (\tau_{n+1}(i), (\xi_k, X_{k+1})_{\sigma_n(i)\leq k<\sigma_{n+1}(i)}), \quad n \geq 0,$$

with distribution $\mathbb{P}_i(Z_0 \in \cdot)$ for $n \geq 1$ under every $\mathbb{P}_\lambda$ and for $n \geq 0$ under $\mathbb{P}_i$. Furthermore, $m_{ii} := \mathbb{E}_i\sigma_1(i) < \infty$ for all $i \in \mathcal{S}$ due to positive recurrence. As a consequence, $(\sigma_n(i))_{n\geq 0}$ forms a *standard renewal process* (SRP) under $\mathbb{P}_i$ and $(\sigma_n(i))_{n\geq 1}$ forms a *(delayed) renewal process* (RP) under $\mathbb{P}_j$, $j \neq i$. Equivalently, the $(\tau_n(i))_{n\geq 1}$ are independent under all $\mathbb{P}_j$, but identically distributed only for $n \geq 2$, unless $j = i$. In the latter case, all $\tau_n(i)$ are i.i.d.

We will usually work under the assumption $S_0 := 0 =: X_0$. Then we can make a similar statement concerning the additive component: $(S_{\sigma_n(i)})_{n\geq 0}$ is a *standard random walk* (SRW) under $\mathbb{P}_i$ and, for $i \neq j$, $(S_{\sigma_n(j)})_{n\geq 1}$ is a *(delayed) random walk* (RW) under $\mathbb{P}_i$.

As the driving chain is positive recurrent, its stationary distribution π is unique and has an occupation measure representation: For $j \in \mathcal{S}$

$$\pi_j := \pi_j^{(i)} = \frac{1}{m_{ii}}\mathbb{E}_i\left(\sum_{n=0}^{\sigma_1(i)-1} \mathbb{1}_{\{\xi_n=j\}}\right) = \frac{1}{m_{ii}}\mathbb{E}_i\left(\sum_{n=1}^{\sigma_1(i)} \mathbb{1}_{\{\xi_n=j\}}\right).$$

If for some function f the stationary expectation $\mathbb{E}_\pi f(\xi_1, X_1)$ exists, then by the kernel structure of the MMS, we can infer

$$\mathbb{E}_\pi f(\xi_1, X_1) = \frac{1}{m_{ii}}\mathbb{E}_i\left(\sum_{n=1}^{\sigma_1(i)} f(\xi_n, X_n)\right), \tag{2.1}$$

cf. [Als14] or [NN87]. For $f(x,y) = y$, we obtain the *Wald-type formula* for MRWs

$$\mathbb{E}_i S_{\sigma_1(i)} = m_{ii}\mathbb{E}_\pi X_1 = m_{ii}\mu,$$

where we set $\mu := \mathbb{E}_\pi X_1$ as the stationary mean or *drift* of X_1. Thus, the knowledge of $\mu > 0$ (as will be the case in the first part of this work) yields $\mathbb{E}_i S_{\sigma_1(i)} > 0$. Then $(S_{\sigma_n(i)})_{n\geq 0}$ is a SRW under $\mathbb{P}_i$ with positive mean, a structure that is well-investigated in standard renewal theory.

We denote by $\mathbb{U}^{ii}$ the ordinary *renewal measure* of $(S_{\sigma_n(i)})_{n\geq 0}$ under $\mathbb{P}_i$ as well as by $\mathbb{U}^{ij}$, $i \neq j$, the ordinary renewal measure of $(S_{\sigma_n(j)})_{n\geq 1}$ under $\mathbb{P}_i$, i.e.

$$\mathbb{U}^{ii} = \sum_{n\geq 0} F_{ii}^{*(n)} \quad \textit{and} \quad \mathbb{U}^{ij} = F_{ij} * \mathbb{U}^{jj}, \quad i \neq j,$$

for $F_{ii} := \mathbb{P}_i{}^{S_{\sigma_1(i)}}$ and $F_{ij} := \mathbb{P}_i{}^{S_{\sigma_1(j)}}$. The corresponding *renewal function* is, say, $\mathbb{U}^{ii}(t) := \mathbb{U}^{ii}((-\infty, t])$, $t \in \mathbb{R}$. Let further

$$\mathbb{U}_i(\cdot) = \mathbb{E}_i\left(\sum_{n\geq 0} \mathbb{1}_{\cdot}(\xi_n, S_n)\right) = \sum_{n\geq 0} \mathbb{P}_i\left((\xi_n, S_n) \in \cdot\right)$$

denote the *Markov renewal measure* of the MRW $(\xi_n, S_n)_{n\geq 0}$ and $\mathbf{U}(i,\cdot) := \mathbb{U}_i(\cdot)$ the corresponding *Markov renewal kernel*. A key identity is

$$\mathbb{U}^{ij} = \mathbb{U}_i(\{j\} \times \cdot)$$

for all $i, j \in \mathcal{S}$. It connects the Markov renewal measure with the ordinary renewal measure of the embedded RW. With these definitions we set

$$\mathbf{U} * g(i,t) := \int_{\mathcal{S}\times\mathbb{R}} g(s, t-x)\, \mathbb{U}_i(\mathrm{d}s, \mathrm{d}x) = \mathbb{E}_i\bigg(\sum_{n\geq 0} g(\xi_n, t - S_n)\bigg)$$

for the convolution of some measurable function $g : \mathcal{S} \times \mathbb{R} \to \mathbb{R}$ with the kernel $\mathbf{U}$. It is the starting point for the use of the key renewal theorem in the Markov case.

First passage times. First passage times $\nu(t) := \inf\{n \geq 0 : S_n > t\}$ are key objects in renewal theory: To give one hint of why this is the case, consider a SRW $(S_n)_{n\geq 0}$ with non-negative increments X_k. Then the renewal function is

$$\mathbb{U}(t) = \sum_{n\geq 0} \mathbb{P}(S_n \leq t) = \sum_{n\geq 0} \mathbb{P}(\nu(t) > n) = \mathbb{E}\nu(t)$$

and knowledge about ν induces knowledge about $\mathbb{U}$.

We can adopt the same definition of $\nu(t)$ for a MRW $(S_n)_{n\geq 0}$. We use the following notation throughout the first two chapters: For $t \geq 0$, the first passage time is $\nu(t) := \inf\{n \geq 0 : S_n > t\}$ and allowing an initial value or random variable we set $\nu(x,t) := \inf\{n \geq 0 : x + S_n > t\}$ for $x \in \mathbb{R}$. Thus, we have $\nu(t) = \nu(0,t)$ and $\nu(x,t) = \nu(t-x)\mathbb{1}_{\{t-x\geq 0\}}$, whenever $S_0 = 0$. Finally, denote by $\nu^i(t) = \inf\{n \geq 0 : S_{\sigma_n(i)} > t\}$ the first passage time of $(S_{\sigma_n(i)})_{n\geq 0}$ (or $(S_{\sigma_n(i)})_{n\geq 1}$).

2.6. Central Markov-modulated sequence

The following structure serves as the main auxiliary process in the renewal-theoretic study of trie parameters and is strongly inspired by [Jan12a]. Again, we consider the generic string $\Xi_0 = \xi_0\xi_1\cdots$ and introduce the sequence $(X_n)_{n\geq 1}$ of real-valued random variables via

$$X_n := -\log p_{\xi_{n-1},\xi_n}.$$

With $H_i := \mathbb{E}_i X_1 = -\sum_{j\in\mathcal{S}} p_{i,j} \log p_{i,j}$, the conditional entropy of ξ_1 given $\xi_0 = i$, we can calculate the drift

$$\mu = \mathbb{E}_\pi X_1 = \sum_{i\in\mathcal{S}} \pi_i \left(-\sum_{j\in\mathcal{S}} p_{i,j}\log p_{i,j}\right) = \sum_{i\in\mathcal{S}} \pi_i H_i \tag{2.2}$$

which is positive since all $\pi_j > 0$ and at least one $H_i > 0$ (corresponding to the row with $p_{i,j} < 1$ for all $j \in \mathcal{S}$). We set $X_0 := 0 =: S_0$, then the sequences $(X_n)_{n\geq 0}$, $(\xi_n, X_n)_{n\geq 0}$ and $(\xi_n, S_n)_{n\geq 0}$, respectively, with $S_n := \sum_{i=1}^n X_i$, form the basis for all further applications. These sequences involve Markov modulation by the driving chain $\Xi_0 = (\xi_n)_{n\geq 0}$, more precisely:

Lemma 2.10. $(\xi_n, X_n)_{n\geq 0}$ *is a MMS and* $(\xi_n, S_n)_{n\geq 0}$ *is a MRW with non-negative and a.s. bounded increments.*

Proof. The first assertion is obvious. Since $p_{i,j} \leq 1$ for all $i, j \in \mathcal{S}$, $X_n = -\log p_{\xi_{n-1},\xi_n} \geq 0$. Let $\mathbf{0}_i := \{j \in \mathcal{S} : p_{i,j} = 0\}$ and $C := \min\{p_{i,j} : i \in \mathcal{S}, j \notin \mathbf{0}_i\} \in (0,1)$. Then $-\log C \in (0,\infty)$. In this notation

$$\mathbb{P}_i(X_n \leq -\log C) = \mathbb{P}_i(-\log p_{\xi_{n-1},\xi_n} \leq -\log C) = \mathbb{P}_i(p_{\xi_{n-1},\xi_n} \geq C) = 1$$

for every $i \in \mathcal{S}$, $n \geq 1$. □

Analogously to the definition of $\mathbf{0}_i$ from the previous proof, we introduce

$$\mathbf{1} := \{j \in \mathcal{S} : \exists k \in \mathcal{S} \text{ s.t. } p_{j,k} = 1\}.$$

Note that $|\mathbf{1}| < |\mathcal{S}|$ because we forbid the chain to be purely deterministic and $|\mathbf{0}_i| < |S|$ since $(p_{i,j})_{j\in\mathcal{S}}$ is a probability distribution. With the P_i defined in Section 2.2, we can now build a connection between the string and the additive process S_n by

$$\begin{aligned} P_i(\xi_1\cdots\xi_n) &= p_{i,\xi_1}p_{\xi_1,\xi_2}\cdots p_{\xi_{n-1},\xi_n} = e^{\log p_{i,\xi_1}+\log p_{\xi_1,\xi_2}+\cdots+\log p_{\xi_{n-1},\xi_n}} \\ &= e^{-\sum_{i=1}^n X_i} = e^{-S_n} \quad \mathbb{P}_i\text{-a.s.} \end{aligned}$$

This forms the starting point of our analysis as it does in [Jan12a].

Remark 2.11. If $p_{i,j} = 1/|\mathcal{S}|$ for all $i, j \in \mathcal{S}$, then $X_n \equiv \log|\mathcal{S}|$ for all $n \geq 1$. As mentioned before, we exclude this case from our consideration (see Section 2.9 for a further discussion).

2.7. Distributional identity

Essentially, all results for D_n and Δ_n follow from the fact that D_n distributionally equals a first passage time of the formerly mentioned MRW, and the subsequent application of results concerning first passage times from Markov renewal theory. In order to establish this relation, we need to introduce a family $(X_0^{(n)})_{n\geq 2}$ of initial variables or delay variables with distribution

$$\mathbb{P}\left(X_0^{(n)} > x\right) = \left(1 - \frac{e^x}{n}\right)_+^{n-1} = \left(1 - e^{x-\log n}\right)_+^{n-1} \tag{2.3}$$

and which is independent of $\xi_0, \Xi, \Xi^{(2)}, \Xi^{(3)}, \ldots$ Here and whenever $X_0^{(n)}$ is involved, $\mathbb{P}$ means $\mathbb{P}_i$ for arbitrary $i \in \mathcal{S}$. We will use this notation to indicate, that the considered probability does not depend on i. With this definition we can show:

Lemma 2.12. $D_n \overset{\mathrm{d}}{=} \nu(X_0^{(n)}, \log n)$ *for* $n \geq 2$ *under every* $\mathbb{P}_i$.

Proof. Cf. [Jan12a, pp. 4, 5]. To give the basic steps, we first state that $D_n \leq k$, $k \geq 1$, iff none of the other strings starts with the same k letters as Ξ. Thus, we have

$$\begin{aligned} \mathbb{P}_i(D_n \leq k \mid \xi_1,\ldots,\xi_k) &= (1 - P_i(\xi_1\cdots\xi_k))^{n-1} = \left(1 - e^{-S_k}\right)^{n-1} = \left(1 - e^{(\log n - S_k)-\log n}\right)_+^{n-1} \\ &= \mathbb{P}_i(X_0^{(n)} > \log n - S_k \mid \xi_1,\ldots,\xi_k) \quad \mathbb{P}_i\text{-a.s.} \end{aligned}$$

as $S_k \geq 0$ a.s. This implies

$$\mathbb{P}_i(D_n \leq k) = \mathbb{P}_i(X_0^{(n)} > \log n - S_k) = \mathbb{P}_i(X_0^{(n)} + S_k > \log n) = \mathbb{P}_i(\nu(X_0^{(n)}, \log n) \leq k).$$

Also, $\mathbb{P}_i(D_n = 0) = 0 = \mathbb{P}_i(X_0^{(n)} > \log n) = \mathbb{P}_i(\nu(X_0^{(n)}, \log n) = 0)$ for $n \geq 2$. □

This leaves us with the study of the asymptotics of $\nu(X_0^{(n)}, \log n)$ which should be somehow similar to results for $\nu(0, \log n)$. Both will be derived in Section 3.3.

For simplicity the results are formulated for functionals of the form $\nu(X_0^{(\log n)}, \log n)$, but this is just a notational issue which we can ignore by identifying $X_0^{(\log n)}$ with $X_0^{(n)}$.

Remark 2.13. Let $D_n^{(b)}$ be the depth of the first string in $\mathrm{Trie}^{(b)}(\mathcal{M}_n)$. Then $D_n^{(2)} \le k$, $k \ge 1$, iff either none of the other $n-1$ strings starts with the same k letters as Ξ, or one string starts with the same k letters and the other $n-2$ do not. So

$$\begin{aligned}
\mathbb{P}_i(D_n^{(2)} \le k \mid \xi_1, \dots, \xi_k) &= (1 - P_i(\xi_1 \cdots \xi_k))^{n-1} + (n-1)\,(1 - P_i(\xi_1 \cdots \xi_k))^{n-2}\, P_i(\xi_1 \cdots \xi_k) \\
&= (1 - P_i(\xi_1 \cdots \xi_k))^{n-2}\,(1 + (n-2)P_i(\xi_1 \cdots \xi_k)) \\
&= \left(1 - e^{-S_k}\right)^{n-2} \left(1 + (n-2)e^{-S_k}\right) \\
&= \left(1 - e^{(\log n - S_k) - \log n}\right)^{n-2} \left(1 + (n-2)e^{(\log n - S_k) - \log n}\right) \\
&= \mathbb{P}_i(Y_0^{(n)} > \log n - S_k \mid \xi_1, \dots, \xi_k) \quad \mathbb{P}_i\text{-a.s.}
\end{aligned}$$

if we define

$$\mathbb{P}\left(Y_0^{(n)} > x\right) = \left(1 - \frac{e^x}{n}\right)^{n-2} \left(1 + (n-2)\frac{e^x}{n}\right) \mathbb{1}_{\{x \le \log n\}}$$

for $n \ge 3$. Hence, $D_n^{(2)} \stackrel{\mathrm{d}}{=} \nu(Y_0^{(n)}, \log n)$ for $n \ge 3$, and we can start a similar analysis as that for D_n. The same is possible for arbitrary $b \ge 1$ with respective

$$\mathbb{P}\left(Y_0^{(n)} > x\right) = \left(1 - \frac{e^x}{n}\right)^{n-b} \left(\sum_{m=0}^{b-1} \binom{n-b+m-1}{m} \left(\frac{e^x}{n}\right)^m\right) \mathbb{1}_{\{x \le \log n\}} \tag{2.4}$$

for $n \ge b+1$.

We remark at this point that the families $(X_0^{(n)})_{n\ge 2}$ and $(Y_0^{(n)})_{n \ge b+1}$ have nice properties. We will return to this point, when deriving the limiting results.

Considering Lemma 2.12 and $\Delta_n = V_{D_n}$, we expect the following identity to hold in the binary setting of the imbalance factor:

Lemma 2.14. *W.r.t. every* $\mathbb{P}_i$*,*

$$(D_n, \Delta_n) \stackrel{\mathrm{d}}{=} \left(\nu(X_0^{(n)}, \log n), V_{\nu(X_0^{(n)}, \log n)}\right),$$

in particular,

$$\Delta_n \stackrel{\mathrm{d}}{=} V_{\nu(X_0^{(n)}, \log n)}.$$

Proof. The proof is essentially the same as in [Jan12a, Section 4], using $\Delta_n = V_{D_n}$, $\sigma(V_k) \subset \sigma(\Xi) = \sigma(Y_1, Y_2, \dots)$ for every k and Lemma 2.12. We omit the details. □

2.8. Lattice

In [Jan12a], Janson used renewal theory to derive his results, so in combination with Lemma 2.10 it seems promising to use Markov renewal theory in this slightly more complicated setting. In standard renewal theory, one has to consider (at least) two different lattice cases where the X_n are supported on $d\mathbb{Z}$ for some $d > 0$ (*arithmetic*) or where they are not (*non-arithmetic*).

This concept does not carry over to Markov renewal theory without an adjustment. It goes back to Shurenkov [Shu85] that we need to introduce a *shift function* which accounts for the change of states of the driving chain Ξ_0. We call a MRW $(\xi_n, S_n)_{n\geq 0}$ with recurrent discrete driving chain Ξ_0 *d-arithmetic with shift function* β if

$$\mathbb{P}_j(X_1 \in \beta(\xi_1) - \beta(\xi_0) + d\mathbb{Z}) = 1 \tag{2.5}$$

for all $j \in \mathcal{S}$ and

$$d = \sup\{c > 0 : \exists \beta' : \mathcal{S} \to [0, c) : \forall j \in \mathcal{S} : \mathbb{P}_j(X_1 \in \beta'(\xi_1) - \beta'(\xi_0) + c\mathbb{Z}) = 1\} \tag{2.6}$$

hold for $d > 0$ and a function $\beta : \mathcal{S} \to [0, d)$. Otherwise, if $d = 0$ in (2.6), then $(\xi_n, S_n)_{n\geq 0}$ is called *non-arithmetic*. Note that (2.5) implies that S_n is concentrated on the affine lattice $\beta(\xi_n) - \beta(i) + d\mathbb{Z}$ w.r.t. every $\mathbb{P}_j$.

To give more details we recall that $(S_{\sigma_n(i)})_{n\geq 0}$ is a SRW w.r.t. $\mathbb{P}_i$ so we can talk about its standard lattice span $d(i)$. It is known that in case of a MRW $(\xi_n, S_n)_{n\geq 0}$ this $d(i)$ does not depend on i which means that the span of $\mathbb{P}_i(S_{\sigma_1(i)} \in \cdot)$ is a universal d for every $i \in \mathcal{S}$.

If $\mathcal{S}_{\sigma_1(i)}$ is d-arithmetic, then Shurenkov showed that there indeed exists $\beta : \mathcal{S} \to [0, d)$, such that the MRW $(\xi_n, S_n)_{n\geq 0}$ is d-arithmetic with shift function β. Also if $\mathcal{S}_{\sigma_1(i)}$ is non-arithmetic, then (2.6) holds and $(\xi_n, S_n)_{n\geq 0}$ is non-arithmetic. It is obvious that d-arithmeticity or non-arithmeticity of MRWs implies the corresponding property for $\mathcal{S}_{\sigma_1(i)}$.

For more details see e.g. [Als94], [Als97], [Als14] or [Shu85].

Necessity of lattice examination. Since it is not obvious whether the d-arithmetic occurs in our model, we briefly discuss the situation in the binary case $\mathcal{S} = \{0, 1\}$ with $p_{i,j} \in (0, 1)$, $i, j \in \mathcal{S}$. As explained above we want to find the lattice span of $S_{\sigma_1(0)}$ under $\mathbb{P}_0$ with

$$\sigma_1(0) = \min\{n \geq 1 \mid \xi_n = 0\}.$$

Obviously, $S_{\sigma_1(0)}$ is discrete and positive. Let $a > 0$. Then

$$\begin{aligned}
\mathbb{P}_0(S_{\sigma_1(0)} = a) &= \sum_{k=1}^{\infty} \mathbb{P}_0(\sigma_1(0) = k, S_k = a) \\
&= \mathbb{P}_0(\xi_1 = 0, -\log p_{0,0} = a) + \sum_{k=2}^{\infty} \mathbb{P}_0(\sigma_1(0) = k, S_{k-1} - \log p_{\xi_{k-1},0} = a) \\
&= p_{0,0} \cdot \delta_{-\log p_{0,0}}(a) + \sum_{k=2}^{\infty} \mathbb{P}_0(\xi_1 = \ldots = \xi_{k-1} = 1, \xi_k = 0, S_{k-1} - \log p_{\xi_{k-1},0} = a) \\
&= p_{0,0} \cdot \delta_{-\log p_{0,0}}(a) + \sum_{k=2}^{\infty} p_{0,1} p_{1,1}^{k-2} p_{1,0} \cdot \delta_{-\log p_{0,1} - (k-2)\log p_{1,1} - \log p_{1,0}}(a).
\end{aligned}$$

As $p_{0,0}$ und $p_{0,1}p_{1,1}^{k-2}p_{1,0}$ are positive for every $k \geq 2$, $S_{\sigma_1(0)}$ has support

$$\begin{aligned}
\mathcal{T}_0^{\text{bin}} &= \{-\log p_{0,0}\} \cup \{-\log p_{0,1} - \log p_{1,0} - k \log p_{1,1} \mid k \geq 0\} \\
&= \{-\log p_{0,0}\} \cup \Big[\{-\log(p_{0,1}p_{1,0})\} - \log p_{1,1} \cdot \mathbb{N}_0\Big].
\end{aligned}$$

For $\mathcal{T}_0^{\text{bin}}$ to be a subset of some $d_{(0)}\mathbb{Z}$, it can be easily shown (by considering $k = 0$ and $k = 1$) that

$$\frac{\log(p_{0,1}p_{1,0})}{\log p_{1,1}} \in \mathbb{Q} \quad \text{and} \quad \frac{\log(p_{0,1}p_{1,0})}{\log p_{0,0}} \in \mathbb{Q} \tag{2.7}$$

must hold. Conversely, (2.7) implies (via Euclidean algorithm) that $\mathcal{T}_0^{\text{bin}} \subset d_{(0)}\mathbb{Z}$ for $d_{(0)}$ defined as $d_{(0)} = \gcd(\log(p_{0,1}p_{1,0}), \log p_{0,0}, \log p_{1,1})$ in the sense that $d_{(0)}$ is the greatest positive real such that $\log(p_{0,1}p_{1,0})$, $\log p_{0,0}$ and $\log p_{1,1}$ are integer multiples of $d_{(0)}$. This $d := d_{(0)} = d_{(1)} > 0$ does not depend on 0 or 1 as was stated above. This can also be seen directly in (2.7) which is symmetric in 0 and 1.

Considering this condition, the question arises whether, with $p_{i,j}$ being transition probabilities, this can be non-trivially solved. The trivial case is $p_{i,j} = \frac{1}{2}$ for all $i, j \in \{0, 1\}$ with both ratios equal to 2, but we already excluded this case.

In order to give a positive answer to this question, we reformulate (2.7) under direct consideration of the $p_{i,j}$ being transition probabilities:

$$\frac{\log(1-p_{0,0})}{\log p_{1,1}} + \frac{\log(1-p_{1,1})}{\log p_{1,1}} \in \mathbb{Q} \quad \text{and} \quad \frac{\log(1-p_{0,0})}{\log p_{0,0}} + \frac{\log(1-p_{1,1})}{\log p_{0,0}} \in \mathbb{Q}. \tag{2.8}$$

Proposition 2.15. (2.8) *can be solved non-trivially.*

Proof. Define the function $f : (0,1)^2 \to \mathbb{R}^2$ by

$$f(x,y) = \left(\frac{\log(1-x)}{\log y} + \frac{\log(1-y)}{\log y}, \frac{\log(1-x)}{\log x} + \frac{\log(1-y)}{\log x}\right).$$

We show that f is a local diffeomorphism by computing the Jacobi matrix

$$\mathbf{D}f\,(x,y) = \begin{pmatrix} -\frac{1}{(1-x)\log y} & -\frac{\log(1-y)+\log(1-x)}{y\log^2 y} - \frac{1}{(1-y)\log y} \\ -\frac{\log(1-x)+\log(1-y)}{x\log^2 x} - \frac{1}{(1-x)\log x} & -\frac{1}{(1-y)\log x} \end{pmatrix}$$

and noting that $|\mathbf{D}f\,(1/2,1/2)| = -32\log^{-2} 2 \neq 0$. Thus, the inverse function theorem gives us open neighborhoods $U \subset (0,1)^2$ of $(1/2,1/2)$ and $V \in \mathbb{R}^2$ of $f\,(1/2,1/2) = (2,2)$, such that $f : U \to V$ is a diffeomorphism and as $\mathbb{Q}^2$ is dense in $\mathbb{R}^2$, we can find a rational $(q_1, q_2) \neq (2,2)$ in V. ☐

Shift function. In our binary example one can directly compute a shift function in the following way. Since $X_1 = -\log p_{\xi_0,\xi_1}$ is fully determined by ξ_0 and ξ_1 and all $p_{i,j}$ are positive, (2.5) is equivalent to the existence of $k_{ij} \in \mathbb{Z}$ with

$$\begin{aligned} -\log p_{0,0} &= k_{00}d, \\ -\log p_{1,1} &= k_{11}d, \\ -\log p_{0,1} &= \beta(1) - \beta(0) + k_{01}d =: \beta_{01} + k_{01}d, \\ -\log p_{1,0} &= \beta(0) - \beta(1) + k_{10}d =: -\beta_{01} + k_{10}d. \end{aligned} \tag{2.9}$$

We remark that we likewise obtain the first two equations with (2.7) and the definition of d.

Now we know that $\beta_{01} \in (-d, d)$, so $\left(-\frac{\log p_{0,1}}{d}\right) - k_{01} \in (-1, 1)$ and

$$k_{01} \in \left\{\left\lfloor -\frac{\log p_{0,1}}{d}\right\rfloor, \left\lceil -\frac{\log p_{0,1}}{d}\right\rceil\right\}.$$

Similarly, we get a more detailed description of $k_{10} \in \left\{\left\lfloor -\frac{\log p_{1,0}}{d}\right\rfloor, \left\lceil -\frac{\log p_{1,0}}{d}\right\rceil\right\}$. So equations three and four imply the connection

$$\beta(1) = \frac{\log p_{1,0} - \log p_{0,1} + (k_{10} - k_{01})d}{2} + \beta(0),$$

so one possible choice if k_{01} and k_{10} each are the bigger alternative is

$$\beta(0) = \frac{\log p_{0,1} + k_{01}d}{2} \in [0, d) \quad \text{and} \quad \beta(1) = \frac{\log p_{1,0} + k_{10}d}{2} \in [0, d),$$

and if they each are the smaller alternative, we may choose

$$\beta(0) = \frac{-\log p_{1,0} - k_{10}d}{2} \in [0, d) \quad \text{and} \quad \beta(1) = \frac{-\log p_{0,1} - k_{01}d}{2} \in [0, d).$$

If $k_{10} = \left\lceil -\frac{\log p_{1,0}}{d} \right\rceil$ and $k_{01} = \left\lfloor -\frac{\log p_{0,1}}{d} \right\rfloor$, then

$$0 \le a := -\log p_{0,1} - \left\lfloor -\frac{\log p_{0,1}}{d} \right\rfloor d < -\log p_{0,1} - \Big(-\frac{\log p_{0,1}}{d} - 1 \Big) d = d$$

and

$$0 \le b := \left\lceil -\frac{\log p_{1,0}}{d} \right\rceil d + \log p_{1,0} < \Big(-\frac{\log p_{1,0}}{d} + 1 \Big) d + \log p_{1,0} = d.$$

Hence, we may choose

$$\beta(0) = 0 \in [0, d) \quad \text{and} \quad \beta(1) = \frac{a+b}{2} \in [0, d).$$

The case $k_{10} = \left\lfloor -\frac{\log p_{1,0}}{d} \right\rfloor$ and $k_{01} = \left\lceil -\frac{\log p_{0,1}}{d} \right\rceil$ is similar.

General condition. Concerning a general condition for the Markov source to be arithmetic, the procedure from the binary case suggests a condition similar to (2.7) to hold here. By a cycle $(i, i_1, \ldots, i_k)$ we mean an element in $\mathcal{S}^{k+1}$ with pairwise distinct components $i, i_1, \ldots, i_k$ depicting a path from $i \in \mathcal{S}$ back to i that does not contain any other state twice. The probability of this cycle (w.r.t. the underlying Markov chain) is thus $P_i(i_1 \cdots i_k i)$, and we call the cycle *possible* if this probability is positive. Now the condition may be stated as follows:

$$\frac{\log P_i(i_1 \cdots i_k i)}{\log P_j(j_1 \cdots j_l j)} \in \mathbb{Q} \text{ for all } 0 \le k, l < |\mathcal{S}| \text{ and possible cycles } (i, i_1, \ldots, i_k), (j, j_1, \ldots, j_l). \tag{2.10}$$

Numerator and denominator are the values of the additive component (up to a $-$) conditioned on the driving chain returning to its starting point along the cycles $(i, i_1 \ldots i_k)$ and $(j, j_1 \ldots j_l)$, respectively.

We now give a hint at why this is a characterizing condition and, for simplicity, restrict ourselves to the case $p_{i,j} \in (0, 1)$ for all $i, j \in \mathcal{S}$: In the same way as in the binary setting, we can determine the support $\mathcal{T}_i$ of $S_{\sigma_1(i)}$ under $\mathbb{P}_i$ (with i being a reference point of $\mathcal{S}$). One representation of $\mathcal{T}_i$ is

$$\mathcal{T}_i = \bigcup_{k=0}^{\infty} \bigcup_{i_1 \ldots, i_k \neq i} \{-\log P_i(i_1 \cdots i_k i)\}.$$

This suggests a condition like

$$\frac{\log P_i(i_1 \cdots i_k i)}{\log P_i(j_1 \cdots j_l i)} \in \mathbb{Q} \quad \text{for all } k, l \ge 0 \text{ and } i_1, \ldots, i_k, j_1, \ldots, j_l \neq i \tag{2.11}$$

to be sufficient for $\mathcal{T}_i \subset d\mathbb{Z}$ with some $d > 0$. However, this is not clear since (2.11) is made up of infinitely many subconditions. Also, at first glance, this condition does not seem to be equivalent to the same condition with i replaced by some $j \neq i$, as it should be. For an easier reasoning, we thus introduce the symmetrized condition

$$\frac{\log P_i(i_1 \cdots i_k i)}{\log P_j(j_1 \cdots j_l j)} \in \mathbb{Q} \quad \text{for all } k, l \geq 0, i, j \in \mathcal{S} \text{ and } i_1, \ldots, i_k \neq i, j_1, \ldots, j_l \neq j. \tag{2.12}$$

Using the scheme

$$\mathcal{T}_i \subset d\mathbb{Z} \Rightarrow (2.11) \Rightarrow (2.12) \Rightarrow (2.10) \Rightarrow \mathcal{T}_i \subset d\mathbb{Z}$$

we can now show that (2.10) is indeed a characterizing condition for the MRW to be d-arithmetic with some shift function β, d being the gcd of all appearing numerators (or denominators). The first and third implication are of course trivial. It is also not hard to see that (2.11) implies (2.12), e.g., for $j \neq i$,

$$\frac{\log P_j(j)}{\log P_i(ji)} = \frac{\log P_i(jji)}{\log P_i(ji)} - \frac{\log P_i(ji)}{\log P_i(ji)} \in \mathbb{Q}$$

and thus $\log P_j(j)$ has rational ratios w.r.t. all $\log P_i(i_1 \cdots i_k i)$. Similarly, one can show that the other ratios are also rational. It is crucial for the fourth implication to note that the corresponding conditions for cycles longer than $|\mathcal{S}|$ can be composed of the conditions of two or more cycles with length smaller or equal to $|\mathcal{S}|$. Thus, we have overcome the problem of too many conditions in (2.11) and narrowed the number of conditions down to a finite number of essential ones. Also, the numerators (or denominators) in (2.10) span the support $\mathcal{T}_i$ (up to a $-$).

Note that (2.10) also applies if certain transitions are impossible or deterministic. In case of deterministic transitions, the problem is reduced to a $|\mathcal{S}| - |\mathbf{1}|$-dimensional problem. For instance, in the binary model with $p_{0,0}, p_{0,1} \in (0,1)$ and $p_{1,0} = 1 = 1 - p_{1,1}$, there are only two possible ways for the driving chain to return to 0, and therefore only one condition, namely

$$\frac{\log p_{0,0}}{\log p_{0,1} p_{1,0}} = \frac{\log p_{0,0}}{\log p_{0,1}} \in \mathbb{Q}$$

remains. This can also be easily checked in the way of the beginning of this section. This is the same condition as in the i.i.d. case with $p = p_{0,0}$ and $q = p_{0,1}$, so the problem was reduced by one dimension.

Condition (2.10) appears in [JST01] and [CV15, Section 3.3.] where it characterizes the set of poles to the vertical line $\operatorname{Re} s = 1$ of the Dirichlet series (cf. [CV15, Section 1.3.]) corresponding to the source. One of the few thorough discussions of periodicity in connection with the source probabilities was done in [FRV10] for the i.i.d. setting, where, for the first time, the error bounds in the average-case analysis were made explicit subject to the source probabilities. This paved the way for the analysis in [CV15], [HV14] or [Hun14]. Just like arithmeticity is the source of oscillations and periodicity in our analysis, it is in the afore-mentioned articles, too, while being justified by different methodology.

Although the problem appears to be a very discrete one with, e.g., the finite alphabet and the finite range of X_i, for most choices of P the conditions (2.7) (or (2.8)) and (2.10) are violated. So most of the time we are in the non-arithmetic case which is not as controllable as the arithmetic case, since generally it can contain all kinds of mass concentration from shifted lattices to $\lambda\!\!\lambda$-continuity. Hence, frequently in the non-arithmetic case one requires the process to be *spread-out* which means in our situation, that some convolution power of $\mathbb{P}_\pi(X_1 \in \cdot)$ has a $\lambda\!\!\lambda$-continuous component. This is often a sensible condition, however our problem is still discrete enough to make it impossible for our MRW to be spread-out (and thus to refuse us some powerful techniques): No convolution power of $\mathbb{P}_\pi(X_1 \in \cdot)$ can have a $\lambda\!\!\lambda$-continuous component, because X_1 has only a finite range under $\mathbb{P}_\pi$.

Tameness. Although we will not go into detail, it seems necessary for the classification of existing results to give a short survey about *tameness of dynamical sources* as it is first introduced in [Val+09] and e.g. used in [CV15], [HV14] and [Hun14]. We follow the brief overview in [CV15]: If the Dirichlet series associated to the source admits a pole on $\{\operatorname{Re} s = 1, s \neq 1\}$, then it already admits (periodically spaced) infinitely many poles. The source is therefore called *periodic*. Its *tameness region*, i.e. the region strictly containing $\{\operatorname{Re} s \geq 1\}$ where the series is analytic and of polynomial growth for $|s| \to \infty$, can thus shown to be $\{\operatorname{Re} s > 1 - \alpha, s \neq \text{poles}\}$ for some $\alpha > 0$, and the source is also called *P-tame*. This property corresponds to arithmeticity in our context (it has the same characteristic condition, cf. [CV15, Section 3.3.]).

Otherwise, the only pole on $\{\operatorname{Re} s = 1\}$ is $s = 1$ and the source is called *aperiodic*. We will not use this denotation again as it might confuse with aperiodicity of Markov chains. When $|\operatorname{Im}(s)| \to \infty$, then the poles may approach $\{\operatorname{Re} s = 1\}$ from the left and thus the source is called *tame* if that does not happen too fast: If the tameness region has a hyperbolic shape, then it is called *H-tame*, and if it is even a vertical strip, then it is called *S-tame*.

Moreover, [CV15] defines a *Good Class* of dynamical sources whose elements have certain good techniqual properties similar to those of irreducible and aperiodic Markov sources (and including these sources). A subclass is called the *UNI Class*, its elements, as said in [CV15], "strongly differ from sources with affine branches", i.e. are very different from simple sources, and can be shown to be S-tame. A different subclass is the *DIOP Class* (more precisely the *DIOP2 Class* and the *DIOP3 Class*) of diophantine dynamical sources, where (in our Markov case setting) ratios as in (2.10) are not only irrational but even diophantine, i.e. "not to well approximable by rational numbers" (cf. [CV15] and [Hun14]). Those sources can be shown to be H-tame.

It has to be noted that not all sources are either P-, H- or S-tame. Furthermore, memoryless sources and Markov sources are never S-tame. In particular, not all memoryless sources and Markov sources are P- or H-tame, and notably the notion of tameness only encompasses aperiodic Markov chains.

2.9. Null-homology

To classify RWs concerning their divergence type, one has to exclude the trivial case where the increments are 0 a.s. The corresponding *trivial* subclass of MRWs is the class of *null-homologous* MRWs, the definition of which goes back to [Lal86]. The following definition and lemma are taken from [AB17b, Section 5]. For further details, we refer the reader to the afore-mentioned publications.

A MRW $(\xi_n, S_n)_{n\geq 0}$ is called *null-homologous* if there exists a function $g : \mathcal{S} \to \mathbb{R}$ such that

$$X_n = g(\xi_n) - g(\xi_{n-1}) \quad \mathbb{P}_\pi\text{-a.s.} \tag{2.13}$$

or, equivalently,

$$S_n = g(\xi_n) - g(\xi_0) \quad \mathbb{P}_\pi\text{-a.s.} \tag{2.14}$$

for all $n \geq 1$.

Lemma 2.16. *Given a MRW $(\xi_n, S_n)_{n\geq 0}$, the following assertions are equivalent:*

(a) $(\xi_n, S_n)_{n\geq 0}$ is null-homologous.

(b) $(S_{\sigma_n(i)})_{n\geq 0}$ has zero increments under $\mathbb{P}_i$ for some $i \in \mathcal{S}$.

(c) $(S_{\sigma_n(i)})_{n\geq 0}$ has zero increments under $\mathbb{P}_i$ for all $i \in \mathcal{S}$.

Null-homology concerning the depth. In the formulation of Theorem 3.6 and in its proof in Subsection 3.3.4, we will encounter the condition $\mathbb{V}\mathrm{ar}_i(S_{\sigma_1(i)} - \mu\sigma_1(i)) > 0$. With regard to Remark 2.11, the question arises whether this condition is violated for all $i \in \mathcal{S}$, i.e. $(\xi_n, S_n - \mu n)_{n\geq 0}$ is null-homologous, if and only if $p_{i,j} = 1/|\mathcal{S}|$ for all $i, j \in \mathcal{S}$. In the binary case, we can give an answer to this question:

Proposition 2.17. *In the binary setting,*

$$\begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}, \quad \begin{pmatrix} 2(1+\sqrt{5})^{-1} & 4(1+\sqrt{5})^{-2} \\ 1 & 0 \end{pmatrix} \quad \textit{and} \quad \begin{pmatrix} 0 & 1 \\ 4(1+\sqrt{5})^{-2} & 2(1+\sqrt{5})^{-1} \end{pmatrix}$$

are the only irreducible transition matrices without atoms such that $(\xi_n, S_n - \mu n)_{n\geq 0}$ is null-homologous.

Proof. Note that either $p_{i,j} \in (0,1)$ for all $i, j \in \{0,1\}$ or $p_{0,0} = p = 1 - p_{0,1} = 1 - q \in (0,1)$, $p_{1,0} = 1 = 1 - p_{1,1}$ (or $p_{0,0} = 0 = 1 - p_{0,1}$, $p_{1,0} = p = 1 - p_{1,1} = 1 - q \in (0,1)$) in order to form an irreducible transition matrix that does not allow atoms. In the first case, it can be extracted from (2.13) and (2.14) w.r.t. $(\xi_n, S_n - \mu n)_{n\geq 0}$ that

$$-\log p_{i,i} = \mu \quad \text{for all } i, j \in \{0,1\} \qquad \text{and} \qquad -\log(p_{i,j}p_{j,i}) = \mu \quad \text{for all } i \neq j \in \{0,1\}$$

which together with $p_{0,0} + p_{0,1} = 1$ and $p_{1,0} + p_{1,1} = 1$ easily yields

$$\mu = -\log(1 - e^{-\mu}).$$

The latter equation is uniquely solved by $\mu = \log 2$, thus $(p_{0,1}, p_{0,1}, p_{1,0}, p_{1,1}, \mu) = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \log 2)$ actually is the unique solution to the above system of equations and we obtain the asserted equivalence in the case $p_{i,j} \in (0,1)$ for all $i, j \in \{0,1\}$.

Suppose now, that $p_{0,0} = p = 1 - p_{0,1} = 1 - q \in (0,1)$, $p_{1,0} = 1 = 1 - p_{1,1}$. Then we obtain similarly that

$$-\log p = \mu \quad \text{and} \quad -\log q = 2\mu$$

which together with $p + q = 1$ yields $\mu = \log(\frac{1}{2}(1+\sqrt{5}))$. This agrees with

$$p = e^{-\mu} = \frac{2}{1+\sqrt{5}} \quad \text{and} \quad q = \frac{4}{(1+\sqrt{5})^2}$$

as can be checked with (2.2). Conversely, this choice of the transition matrix easily yields that $S_{\sigma_1(0)} - \mu\sigma_1(0) = 0$ $\mathbb{P}_i$-a.s. using the relations between p, q and μ from above. The case $p_{0,0} = 0 = 1 - p_{0,1}$, $p_{1,0} = p = 1 - p_{1,1} = 1 - q \in (0,1)$ is similar to the latter. $\square$

For $|\mathcal{S}| = 3$, the analogous system of equations is already quite complex. With the help of a computer algebra system, it can be verified that $p_{i,j} = \frac{1}{3} = 1/|\mathcal{S}|$ is indeed equivalent to $\mathbb{V}\mathrm{ar}_i(S_{\sigma_1(i)} - \mu\sigma_1(i)) = 0$, provided that $p_{i,j} \in (0,1)$ for all $i, j \in \mathcal{S}$. Under this condition, we expect equivalence of null-homology of $(\xi_n, S_n - \mu n)_{n\geq 0}$ and $p_{i,j} = 1/|\mathcal{S}|$ for all $i, j \in \mathcal{S}$ to hold for arbitrary finite $\mathcal{S}$. Without this condition, Proposition 2.17 gives reason to believe that the trivial case is generally not the only case in which null-homology occurs. However, we refrain from a further investigation.

Null-homology concerning the imbalance factor. In Theorem 3.11 and in Subsection 3.3.7, we will encounter a condition similar to $\mathbb{V}\mathrm{ar}_i(S_{\sigma_1(i)} - \mu\sigma_1(i)) > 0$, namely $\mathbb{V}\mathrm{ar}_i(\mu V_{\sigma_1(i)} - \mu_Y S_{\sigma_1(i)}) > 0$. Again, this conditions is related to null-homology, more precisely to that of $(\xi_n, \mu V_n - \mu_Y S_n)_{n\geq 0}$. Since we only consider the imbalance factor in the binary setting, we can entirely characterize null-homology for the related random walk:

Proposition 2.18. *There is no irreducible transition matrix without atoms such that $(\xi_n, \mu V_n - \mu_Y S_n)_{n\geq 0}$ is null-homologous.*

Proof. First we show that $p_{0,0} = p = 1 - p_{0,1} = 1 - q \in (0,1)$, $p_{1,0} = 1 = 1 - p_{1,1}$ can never entail null-homology of $(\xi_n, \mu V_n - \mu_Y S_n)_{n\geq 0}$. Then this is also true for $p_{0,0} = 0 = 1 - p_{0,1}$, $p_{1,0} = p = 1 - p_{1,1} = 1 - q \in (0,1)$. First of all, we calculate

$$\mu_Y = 2\pi_1 - 1 = \frac{2p_{0,1}}{p_{1,0} + p_{0,1}} - 1 = \frac{p_{0,1} - p_{1,0}}{p_{1,0} + p_{0,1}} = \frac{q-1}{1+q}.$$

Hence, (2.14) yields

$$\frac{q-1}{1+q} \log(q \cdot 1) = 0$$

which has the unique solution $q = 1$ and thus contradicts $q \in (0,1)$.

Let $p_{i,j} \in (0,1)$ for all $i, j \in \{0,1\}$ from now on. Suppose that $\mu_Y = 0$. Then (2.13) yields $\mu = 0$ which obviously is false. Thus, null-homology implies $\mu_Y \neq 0$. Now, (2.13) gives

$$-\mu + \mu_Y \log p_{0,0} = 0 \quad \text{and} \quad \mu + \mu_Y \log p_{1,1} = 0,$$

hence

$$\mu_Y \log(p_{0,0} p_{1,1}) = 0 \quad \text{or equivalently} \quad p_{0,0} p_{1,1} = 1. \tag{2.15}$$

Similarly, (2.14) gives

$$\mu_Y \log(p_{0,1} p_{1,0}) = 0 \quad \text{or equivalently} \quad p_{0,1} p_{1,0} = 1. \tag{2.16}$$

Together, (2.15) and (2.16) easily yield $p_{0,0} = p_{1,0} = 1 - p_{0,1} = 1 - p_{1,1}$. In other words, the ξ_n have to be independent. Set $p_{0,0} =: p$. We finally show that $\mu = -p \log p - (1-p)\log(1-p)$ does not agree with the only possible shape of the transition matrix that was derived before: Equation (2.13) yields

$$0 = -\mu + (1-2p)\log p = p\log p + (1-p)\log(1-p) + (1-2p)\log p = (1-p)\log p + (1-p)\log(1-p),$$

and this equation has no solution $p \in (0,1)$. This proves the assertion. $\square$

3. Asymptotic analysis of depth and imbalance factor

In this chapter, we will present many asymptotic results for D_n and Δ_n and prove them afterwards in Section 3.3. All results can also be formulated w.r.t. $\mathbb{P}_\lambda$ (sometimes with slight modifications) for a distribution λ on $\mathcal{S}$. As always, $i \in \mathcal{S}$ is an arbitrary starting state.

Let us mention at the outset that some of the results have been obtained earlier by different methods, and we give a more specific account when they are stated. However, we are not aware of any results concerning the imbalance factor in the Markov model.

3.1. Results for D_n

Theorem 3.1. *As $n \to \infty$, it holds that*

$$\frac{D_n}{\log n} \xrightarrow{\mathbb{P}_i} \frac{1}{\mu}. \tag{3.1}$$

Theorem 3.2. *Let $p > 0$. Then, as $n \to \infty$,*

$$\frac{D_n}{\log n} \xrightarrow{L^p} \frac{1}{\mu} \tag{3.2}$$

under $\mathbb{P}_i$, thus in particular, as $n \to \infty$,

$$\mathbb{E}_i \left(\frac{D_n}{\log n} \right)^p \to \frac{1}{\mu^p}. \tag{3.3}$$

Theorem 3.3. *(a) If $(\xi_n, S_n)_{n \geq 0}$ is non-arithmetic, then, as $n \to \infty$,*

$$\mathbb{E}_i D_n = \frac{\log n}{\mu} + \frac{1}{2\mu^2} \sum_{j \in \mathcal{S}} \pi_j^2 \mathbb{E}_j S_{\sigma_1(j)}^2 - \frac{1}{\mu} \sum_{j \neq i} \pi_j \mathbb{E}_i S_{\sigma_1(j)} + \frac{\gamma}{\mu} + o(1), \tag{3.4}$$

with the Euler constant $\gamma = 0.5772\ldots$

(b) If $(\xi_n, S_n)_{n \geq 0}$ is d-arithmetic with shift function 0, then, as $n \to \infty$,

$$\begin{aligned} \mathbb{E}_i D_n = &\frac{\log n}{\mu} + \frac{1}{2\mu^2} \sum_{j \in \mathcal{S}} \pi_j^2 \mathbb{E}_j S_{\sigma_1(j)}^2 - \frac{1}{\mu} \sum_{j \neq i} \pi_j \mathbb{E}_i S_{\sigma_1(j)} + \frac{\gamma}{\mu} \\ &+ \frac{1}{\mu} \psi_1(\log n) + o(1), \end{aligned} \tag{3.5}$$

where $\psi_1(t)$ is a continuous d-periodic function given by

$$\psi_1(t) := - \sum_{k \neq 0} \Gamma(-2\pi\,\mathrm{i}\,k/d) e^{2\pi\,\mathrm{i}\,kt/d}. \tag{3.6}$$

(c) If $(\xi_n, S_n)_{n\geq 0}$ is d-arithmetic with shift function β, then, as $n \to \infty$,

$$\begin{aligned}\mathbb{E}_i D_n = \frac{\log n}{\mu} + \frac{1}{2\mu^2}\sum_{j\in\mathcal{S}} \pi_j^2 \mathbb{E}_j S_{\sigma_1(j)}^2 - \frac{1}{\mu}\sum_{j\neq i} \pi_j \mathbb{E}_i S_{\sigma_1(j)} + \frac{\gamma}{\mu} \\ + \frac{1}{\mu}\sum_{j\in\mathcal{S}} \pi_j \psi_1(\log n - \beta_{ij}) + o(1)\end{aligned} \tag{3.7}$$

with $\beta_{ij} := \beta(j) - \beta(i) \in (-d, d)$.

Remark 3.4. In the d-arithmetic case, it could be possible to derive a better error term of the order $\mathcal{O}(n^{-r})$ with some $r \in (0, 1]$, cf. Remark 3.22 and the proof of Theorem 3.3. The bottleneck seems to be the verification of a corresponding convergence rate in (3.28). As opposed to the d-arithmetic case, the non-arithmetic case does not allow for a more precise version in general: The error term in (3.4) can be arbitrarily bad, cf. Remark 3.23.

Remark 3.5. As it turns out in the proof of the Poisson version of Theorem 4.28, ψ_1 from (3.6) also has a series representation of type (4.19) with $f(x) = f_{\mathrm{L}}(x) = x - xe^{-x}$, $a = 1$ and $g(t) = e^t f(e^{-t})$. See also Remark 4.34.

Theorem 3.6. *Let $\sigma^{(2)} := \frac{1}{m_{ii}} \mathbb{V}\mathrm{ar}_i \left(S_{\sigma_1(i)} - \mu\sigma_1(i)\right)$. Then $\sigma^{(2)} < \infty$ does not depend on i, and if $\sigma^{(2)} > 0$, then*

$$\frac{D_n - \frac{\log n}{\mu}}{\sqrt{\log n}} \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, \frac{\sigma^{(2)}}{\mu^3}\right), \tag{3.8}$$

as $n \to \infty$, w.r.t. $\mathbb{P}_i$. Furthermore, as $n \to \infty$,

$$\mathbb{V}\mathrm{ar}_i\, D_n = \frac{\sigma^{(2)}}{\mu^3}\log n + o(\log n) \quad \textit{and} \quad \mathbb{E}_i\left|\frac{D_n - \frac{\log n}{\mu}}{\sqrt{\log n}}\right|^p \to \mathbb{E}_i|N|^p \tag{3.9}$$

for every $p > 0$, with N having the distribution of the limit in (3.8). The convergence also holds without absolute value for $p \in \mathbb{N}$. If $\sigma^{(2)} = 0$, i.e. $S_{\sigma_1(j)} = \mu\sigma_1(j)$ $\mathbb{P}_j$-a.s. for some (and thus all) $j \in \mathcal{S}$, then, as $n \to \infty$,

$$\frac{D_n - \frac{\log n}{\mu}}{\log^{1/p} n} \xrightarrow{\mathbb{P}_i} 0, \tag{3.10}$$

and

$$\mathbb{V}\mathrm{ar}_i\, D_n = o(\log n) \quad \textit{and} \quad \mathbb{E}_i\left|D_n - \frac{\log n}{\mu}\right|^p = o(\log n) \tag{3.11}$$

for all $p > 0$.

Remark 3.7. Cf. Remark 3.32 for further details on $\sigma^{(2)}$. Also cf. Section 2.9 for a discussion about when $\sigma^{(2)} = 0$ actually occurs.

Remark 3.8. The above theorems in Section 3.1 and the following ones in Section 3.2 are formulated for tries. However, we can also derive corresponding results for b-tries. In fact, all theorems remain true replacing D_n by $D_n^{(b)}$ (also in V_{D_n}) except for Theorem 3.3, which describes a finer structure. More precisely (cf. Remark 3.14), there we need to replace γ by $\gamma - H_{b-1}$, where $H_{b-1} = \sum_{k=1}^{b-1} 1/k$ is the $(b-1)$-st harmonic number, in the non-arithmetic case. If it is possible to show (3.28) with $Y_0^{(n)}$ and Y_0^* instead of $X_0^{(n)}$ and X_0^*, then we also obtain an expansion in the d-arithmetic case with coefficients

$$-\frac{1}{(b-1)!}\prod_{l=1}^{b-1}(l-2\pi\,\mathrm{i}\,k/d)\Gamma(-2\pi\,\mathrm{i}\,k/d)$$

in ψ_1 instead of $-\Gamma(-2\pi\,\mathrm{i}\,k/d)$.

We remark that Pittel obtained a similar result as Theorem 3.1 in [Pit85] for stationary ergodic sources. Jacquet and Szpankowski [JS91] obtained Theorems 3.3 and 3.6 for stationary Markov sources, at first glance with better precision. However, there is no distinction between a *periodic* and *non-periodic* case in their analysis, and the actual shape of the fluctuating functions appearing in their expansions is unclear. In fact, Cesaratto and Vallée [CV15, Theorem 3.5] give a precise expansion of expectation and variance with asymptotical upper bounds for remainder terms in different cases. They use analytic methods and obtain a slightly more precise expansion of the mean in the periodic (or d-arithmetic) case and the H-tame case (contained in the non-arithmetic case). Indeed, they point out (and we can also justify this with our methods, cf. Remarks 3.4 and 3.23) that there are sources with arbitrarily *bad* remainder terms within $o(1)$ in Theorem 3.3, when the source is neither P- nor H-tame. Nevertheless, our probabilistic approach allows us to provide a rather explicit form of the periodic oscillatory term and also to deal with Markov sources where the chain is not aperiodic.

Moreover, Hun proves a CLT for tame sources in [Hun14], excluding only sources conjugated to an unbiased memoryless source, i.e. parts of the P-tame sources. Again, we remark that our result also applies to a more general class of Markov sources, namely those that are not tame, in particular those which are driven by a periodic chain.

3.2. Results for Δ_n

In the binary setting, we get the following results. Recall that

$$\pi_0 = \frac{p_{1,0}}{p_{1,0}+p_{0,1}} \quad \text{and} \quad \pi_1 = \frac{p_{0,1}}{p_{1,0}+p_{0,1}}$$

and $\mu_Y = \mathbb{E}_\pi Y_1 = 2\pi_1 - 1$.

Theorem 3.9. *As $n \to \infty$,*

$$\frac{\Delta_n}{\log n} \xrightarrow{\mathbb{P}_i} \frac{\mu_Y}{\mu}. \tag{3.12}$$

Theorem 3.10. *Let $p > 0$. Then, as $n \to \infty$,*

$$\frac{\Delta_n}{\log n} \xrightarrow{L^p} \frac{\mu_Y}{\mu} \tag{3.13}$$

under $\mathbb{P}_i$, thus in particular, as $n \to \infty$,

$$\mathbb{E}_i\left(\frac{|\Delta_n|}{\log n}\right)^p \to \left|\frac{\mu_Y}{\mu}\right|^p. \tag{3.14}$$

The convergence also holds without absolute value for $p \in \mathbb{N}$.

Theorem 3.11. *Let* $\gamma^{(2)} := \frac{1}{m_{ii}} \mathbb{V}\mathrm{ar}_i(\mu V_{\sigma_1(i)} - \mu_Y S_{\sigma_1(i)})$. $\gamma^{(2)} < \infty$ *does not depend on* i *and* $\gamma^{(2)} > 0$. *It holds that*

$$\frac{\Delta_n - \frac{\mu_Y}{\mu}\log n}{\sqrt{\log n}} \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, \frac{\gamma^{(2)}}{\mu^3}\right), \tag{3.15}$$

as $n \to \infty$, *w.r.t.* $\mathbb{P}_i$. *Furthermore, as* $n \to \infty$,

$$\mathbb{V}\mathrm{ar}_i\,\Delta_n = \frac{\gamma^{(2)}}{\mu^3}\log n + o(\log n) \quad \text{and} \quad \mathbb{E}_i\left|\frac{\Delta_n - \frac{\mu_Y}{\mu}\log n}{\sqrt{\log n}}\right|^p \to \mathbb{E}_i|N|^p \tag{3.16}$$

for every $p > 0$, *with* N *having the distribution of the limit in* (3.15). *The convergence also holds without absolute value for* $p \in \mathbb{N}$. *For all* $p > 0$

$$\mathbb{E}_i\Delta_n = \frac{\mu_Y}{\mu}\log n + o(\log^{1/p} n). \tag{3.17}$$

Remark 3.12. Cf. Remark 3.39 for further details on $\gamma^{(2)}$.

3.3. Proofs

In the view of Section 2.7, the results follow very quickly from Lemma 2.12 and the corresponding result from Markov renewal theory. Therefore, we derive several asymptotic results for $\nu(X_0^{(t)}, t)$ by using results for $\nu(t)$. Hence, the delay family needs to be controlled. In our auxiliary results, we usually do not specify the index set and only write $(X_0^{(t)})_t$ for some family of starting values, independent of the MRW. A priori, t can vary in a general index set $I \subset \mathbb{R}$ unbounded to the right, but typically (as in $\nu(X_0^{(t)}, t)$ with $X_0^{(t)}$ and t coupled) we implicitly assume $t \geq 0$, which is only a notational issue. For simplicity, we sometimes assume t to be a sequence, since in our application we only need $t = \log n$. Certainly, when we do such a simplification, the more general result should also hold.

The fact that the family $(X_0^{(n)})_{n\geq 2}$ from (2.3) is well-behaved is recorded in the following lemma.

Lemma 3.13. *Let* $(X_0^{(n)})_{n\geq 2}$ *be the family of real-valued random variables with laws defined by* (2.3). *Then*

(a) $X_0^{(n)} \xrightarrow{\mathrm{d}} X_0^*$, *as* $n \to \infty$, *with* $-X_0^* \sim \mathrm{Gumbel}(0,1)$. *In particular,* $(X_0^{(n)})_{n\geq 2}$ *is tight,* $-X_0^*$ *has Fourier transform* $\phi(t) = \Gamma(1 - \mathrm{i}\,t)$ *and* $\mathbb{E}X_0^* = -\gamma$, *where* γ *is the Euler constant.*

(b) There exists an $s > 0$ *such that* $(e^{s|X_0^{(n)}|})_{n\geq 2}$ *is uniformly integrable. In particular,* $(X_0^{(n)})_{n\geq 2}$ *and* $(|X_0^{(n)}|^p)_{n\geq 2}$ *are uniformly integrable for all* $p > 0$.

(c) $(X_0^{(n)})_{n\geq 2} \stackrel{\mathrm{d}}{=} (\log n - \max\{Z_1, \ldots, Z_{n-1}\})_{n\geq 2}$, *where* $Z_1, Z_2, \ldots$ *are i.i.d.* $\mathrm{Exp}(1)$ *distributed random variables.*

Proof. For (a), let $\log n > x$ for some $x \in \mathbb{R}$. Then, as $n \to \infty$,

$$\mathbb{P}(X_0^{(n)} > x) = \left(1 - \frac{e^x}{n}\right)^{n-1} \to e^{-e^x}.$$

For (b), it is easy to show that the tail probabilities of $X_0^{(n)}$ uniformly (in n) decrease exponentially (cf. Section A.1). It is certainly reasonable since the variables are close to Gumbel. (c) is easy. □

Remark 3.14. The family $(Y_0^{(n)})_{n\geq b+1}$ from (2.4) has corresponding properties to those in Lemma 3.13: The sequence converges in distribution to some Y_0^* with tail function

$$\mathbb{P}(Y_0^* > x) = e^{-e^x}\sum_{m=0}^{b-1}\frac{1}{m!}(e^x)^m, \quad x\in\mathbb{R},$$

in particular, $-Y_0^*$ has $\lambda\!\!\lambda$-density

$$f_{-Y_0^*}(x) = e^{-e^{-x}}\frac{1}{(b-1)!}e^{-bx},$$

Fourier transform

$$\phi_{-Y_0^*}(t) = \frac{1}{(b-1)!}\Gamma(b-\mathrm{i}\,t),$$

and thus, $\mathbb{E}(-Y_0^*) = -\Gamma'(b)/\Gamma(b) = -\Psi(b) = -(H_{b-1}-\gamma)$, where Ψ is the digamma function. The family obeys the same integrability conditions as $(X_0^{(n)})_{n\geq 2}$, but we will restrict ourselves to $(X_0^{(n)})_{n\geq 2}$ below.

Lemma 3.15. *$\sigma_1(i)$ has an exponential moment under all $\mathbb{P}_i$, thus it has polynomial moments of every order.*

Proof. Note that $\mathcal{S}$ is finite. Thus, the result is well-known. □

For the rest of this part, we call a MRW $(\xi_n, S_n)_{n\geq 0}$ *zero-delayed* if $S_0 = 0$ $\mathbb{P}_i$-a.s. for all $i\in\mathcal{S}$. For the investigation of the imbalance factor Δ_n, we will add a further additive component to our MRW $(\xi_n, S_n)_{n\geq 0}$, more precisely let $(\xi_n, X_n, Y_n)_{n\geq 0}$ be a MMS (on $\mathcal{S}\times(\mathbb{R}\times\mathbb{R})$) and denote by $(\xi_n, S_n, V_n)_{n\geq 0}$ the corresponding MRW. Then $(\xi_n, S_n, V_n)_{n\geq 0}$ is called *zero-delayed* if $S_0 = V_0 = 0$ $\mathbb{P}_i$-a.s. for all $i\in\mathcal{S}$. We also agree on $\mu = \mathbb{E}_\pi X_1$ and $\mu_Y = \mathbb{E}_\pi Y_1$ being the stationary drifts of $(S_n)_{n\geq 0}$ and $(V_n)_{n\geq 0}$, respectively. Obviously, we think of $(X_n)_{n\geq 0}$ from Section 2.6 and $(Y_n)_{n\geq 0}$ from Section 2.3.

We will now proceed as follows: We first derive a limit result for $\nu(t)$ (and $V_{\nu(t)}$ later on), and after choosing appropriate conditions for $(X_0^{(t)})_t$, the corresponding result for $\nu(X_0^{(t)}, t)$ comes almost for free: The proof of [Jan12a, Theorem A.4] applies since $(X_0^{(t)})_t$ is independent of the MRW. For a better understanding, we provide a full derivation for all of our results. In the results below, $\mathcal{S}$ may a priori be countable.

3.3.1. Weak law of large numbers for D_n

We provide a strong law of large numbers (SLLN) for $\nu(X_0^{(t)}, t)$ that translates into a WLLN for D_n. The following is a well-known result and is not restricted to a discrete state space $\mathcal{S}$.

Theorem 3.16 (SLLN)**.** *Let $(\xi_n, X_n)_{n\geq 0}$ be a MMS with positive recurrent discrete driving chain Ξ_0 which has a stationary distribution π. Suppose further that μ exists. Then, for all $i\in\mathcal{S}$,*

$$\lim_{n\to\infty}\frac{S_n}{n} = \mu \quad \mathbb{P}_i\text{-}a.s.$$

Proof. Follows easily from Birkhoff's ergodic theorem, see e.g. [AB17b, Theorem 10.1]. □

The idea of the proof of the following result goes back to Doob in [Doo48] who used it in the i.i.d. setting.

Lemma 3.17. *Let $(\xi_n, S_n)_{n\geq 0}$ be a zero-delayed MRW with positive recurrent and discrete driving chain Ξ_0 which has stationary distribution π. Let μ be positive. Then, for all $i \in \mathcal{S}$, as $t \to \infty$,*

$$\frac{\nu(t)}{t} \to \frac{1}{\mu} \quad \mathbb{P}_i\text{-a.s.}$$

Proof. We use Theorem 3.16 to infer $\nu(t) \uparrow \infty$, as $t \to \infty$ $\mathbb{P}_i$-a.s. and

$$\frac{S_{\nu(t)}}{\nu(t)} \to \mu \quad \text{and} \quad \frac{S_{\nu(t)-1}}{\nu(t)} \to \mu \quad \mathbb{P}_i\text{-a.s.},$$

as $t \to \infty$. Then the definition of $\nu(t)$ together with a sandwiching argument gives

$$\frac{S_{\nu(t)-1}}{\nu(t)} \leq \frac{t}{\nu(t)} \leq \frac{S_{\nu(t)}}{\nu(t)}$$

and $t/\nu(t) \to \mu$ $\mathbb{P}_i$-a.s., as $t \to \infty$. □

We recalled the previous proof mainly to convey the feeling that the statement of the next result is natural, even with initial variables varying over time (but of course obeying some regularity assumptions). In fact, as we require the family of initial variables to be independent of the MRW, the proof of the following result uses the same ideas as [Jan12a, Theorem A.4].

Lemma 3.18. *Given the situation of Lemma 3.17, let $(X_0^{(t)})_t$ be a tight family of real-valued random variables independent of $(\xi_n, S_n)_{n\geq 0}$. Then, for all $i \in \mathcal{S}$, as $t \to \infty$,*

$$\frac{\nu(X_0^{(t)}, t)}{t} \xrightarrow{\mathbb{P}_i} \frac{1}{\mu}.$$

Proof. The tightness of $(X_0^{(t)})_t$ implies

$$\frac{X_0^{(t)}}{t} \xrightarrow{\mathbb{P}_i} 0 \quad \text{and} \quad t - X_0^{(t)} \xrightarrow{\mathbb{P}_i} \infty,$$

as $t \to \infty$. Now, use $\nu(X_0^{(t)}, t) = \nu(t - X_0^{(t)})\mathbb{1}_{\{t - X_0^{(t)} \geq 0\}}$ and the asymptotics for $\nu(t)$. Due to the independence assumption, we can use Skorokhod's representation theorem to find independent copies of $(X_0^{(t)})_t$ and $(\nu(t))_t$ (for simplicity with the same names) such that, on the one hand, the above convergence even holds $\mathbb{P}_i$-a.s. and, on the other hand, we still have $\nu(t)/t \to 1/\mu$ $\mathbb{P}_i$-a.s. Then the result is immediate for the copies via

$$\frac{\nu(X_0^{(t)}, t)}{t} = \frac{\nu(t - X_0^{(t)})}{t}\mathbb{1}_{\{t - X_0^{(t)} \geq 0\}} = \frac{\nu(0)}{t}\mathbb{1}_{\{t - X_0^{(t)} = 0\}} + \frac{t - X_0^{(t)}}{t}\frac{\nu(t - X_0^{(t)})}{t - X_0^{(t)}}\mathbb{1}_{\{t - X_0^{(t)} > 0\}},$$

and because of distributional equality it also holds in distribution for the original variables. Since the limit is constant, the convergence also holds in probability. □

Proof of Theorem 3.1. $(\xi_n, S_n)_{n\geq 0}$ is a MRW fulfilling all conditions required in Lemma 3.18. Moreover, Lemma 3.13 ensures that $(X_0^{(n)})_{n\geq 2}$ is tight. Hence,

$$\frac{\nu(X_0^{(n)}, \log n)}{\log n} \xrightarrow{\mathbb{P}_i} \frac{1}{\mu},$$

and with Lemma 2.12 we obtain the assertion. □

3.3.2. L^p-law of large numbers for D_n

The procedure is again the same. We derive the result for $\nu(t)$ and then for $\nu(X_0^{(t)}, t)$.

Lemma 3.19. *Let $(\xi_n, S_n)_{n\geq 0}$ be a zero-delayed MRW with positive recurrent and discrete driving chain Ξ_0 which has stationary distribution π. Let μ be positive. Assume further that $\mathbb{E}_i\sigma_1(i)^p < \infty$ and $\mathbb{E}_i(S^-_{\sigma_1(i)})^p < \infty$ for some $p \geq 1$. Then $\{(\nu(t)/t)^p, t \geq 1\}$ is uniformly integrable w.r.t. $\mathbb{P}_i$ and thus*

$$\frac{\nu(t)}{t} \xrightarrow{L^p} \frac{1}{\mu}$$

under $\mathbb{P}_i$, as $t \to \infty$, in particular,

$$\mathbb{E}_i\left(\frac{\nu(t)}{t}\right)^q \to \frac{1}{\mu^q}.$$

for $0 < q \leq p$.

Proof. We try to reduce the situation to standard renewal theory. Recall the notation for recurrence times from Section 2.5. Obviously, $\nu^i(t)$ is a stopping time w.r.t.

$$\mathcal{F}_n := \sigma((\sigma_j(i), S_{\sigma_j(i)}) : j = 0, \ldots, n),$$

and $\sum_{k=1}^n \tau_k(i)$ has i.i.d. increments, is adapted to $\mathcal{F}_n$, and $\tau_n(i)$ is independent of $\mathcal{F}_{n-1}$. By [Gut09, Theorem 3.7.1], we have that

$$\left\{\left(\frac{\nu^i(t)}{t}\right)^p, t \geq 1\right\}$$

is uniformly integrable w.r.t. $\mathbb{P}_i$ since $\mathbb{E}_i(S^-_{\sigma_1(i)})^p < \infty$ and the SRW $S_{\sigma_n(i)}$ has positive drift. Hence, [Gut09, Theorem 1.6.1] yields the uniform integrability of

$$\left\{\left(\frac{\sum_{k=1}^{\nu^i(t)} \tau_k(i)}{t}\right)^p, t \geq 1\right\}$$

since $\mathbb{E}_i|\tau_1(i)|^p = \mathbb{E}_i\sigma_1(i)^p < \infty$. Finally, the estimate

$$\nu(t) \leq \sum_{k=1}^{\nu^i(t)} \tau_k(i)$$

yields the desired result. □

Lemma 3.20. *Given the situation of Lemma 3.19, let $(X_0^{(t)})_t$ be a tight and L^p-bounded family of real-valued random variables independent of $(\xi_n, S_n)_{n\geq 0}$. Then (along sequences)*

$$\frac{\nu(X_0^{(t)}, t)}{t} \xrightarrow{L^p} \frac{1}{\mu}$$

under $\mathbb{P}_i$, as $t \to \infty$, in particular,

$$\mathbb{E}_i\left(\frac{\nu(X_0^{(t)}, t)}{t}\right)^q \to \frac{1}{\mu^q} \tag{3.18}$$

for $0 < q \leq p$.

Proof. Since Lemma 3.19 provides the same kind of result as [Jan12a, Theorem 3.2] does, [Jan12a, Theorem A.4] applies again. Nevertheless, we recall the proof, since we will use the same procedure again. By Lemma 3.19, $\mathbb{E}_i\nu(t)^p = \mathcal{O}(t^p)$, $t \to \infty$. Hence, we can find $C > 0$ large enough such that for $t \geq 0$

$$\begin{aligned}\mathbb{E}_i\nu(X_0^{(t)}, t)^p &= \mathbb{E}_i\nu(t - X_0^{(t)})^p \mathbb{1}_{\{t-X_0^{(t)}\geq 0\}} = \int_{(-\infty,t]} \mathbb{E}_i\nu(t-s)^p\,\mathbb{P}^{X_0^{(t)}}(\mathrm{d}s)\\ &\leq C\int_{(-\infty,t]} 1 + (t-s)^p\,\mathbb{P}^{X_0^{(t)}}(\mathrm{d}s) \leq C + C\mathbb{E}(t + |X_0^{(t)}|)^p < \infty\end{aligned}$$

since $X_0^{(t)}$ is in L^p. For some $0 < q \leq p$, let

$$g_i(t) := \mathbb{E}_i\left(\frac{\nu(t)}{t\vee 1}\right)^q.$$

We know that $g_i(t) \to 1/\mu^q$, as $t \to \infty$, by Lemma 3.19. Since $\nu(t)$ is monotone, $\sup_{0\leq t<1} g_i(t) = \sup_{0\leq t\leq 1}\mathbb{E}_i\nu(t)^q = \mathbb{E}_i\nu(1)^q < \infty$. Furthermore, $\{(\nu(t)/t)^p, t \geq 1\}$ is uniformly integrable, hence also $\sup_{t\geq 1} g_i(t) < \infty$. Now, let $t \geq 1$ and consider Y_t defined by

$$\begin{aligned}Y_t &:= \mathbb{E}_i\left[\left(\frac{\nu(X_0^{(t)}, t)}{t}\right)^q \middle| X_0^{(t)}\right] = \mathbb{E}_i\left[\left(\frac{\nu(t - X_0^{(t)})}{t}\right)^q \mathbb{1}_{\{t-X_0^{(t)}\geq 0\}} \middle| X_0^{(t)}\right]\\ &= \mathbb{E}_i\left[\left(\frac{\nu(t-X_0^{(t)})}{(t-X_0^{(t)})\vee 1}\right)^q \middle| X_0^{(t)}\right]\left(\frac{(t-X_0^{(t)})\vee 1}{t}\right)^q \mathbb{1}_{\{t-X_0^{(t)}\geq 0\}}\\ &= g_i(t - X_0^{(t)})\left(\frac{(t-X_0^{(t)})\vee 1}{t}\right)^q \mathbb{1}_{\{t-X_0^{(t)}\geq 0\}},\end{aligned}$$

where we used independence of the MRW and $X_0^{(t)}$ in the last step. We aim to show

$$\mathbb{E}_i\left(\frac{\nu(X_0^{(t)}, t)}{t}\right)^q = \mathbb{E}_i Y_t \to \frac{1}{\mu^q},$$

as $t \to \infty$. W.l.o.g. (cf. the proof of Lemma 3.18), we assume that, as $t \to \infty$, $X_0^{(t)}/t \to 0$ and thus $t - X_0^{(t)} \to \infty$ $\mathbb{P}_i$-a.s. As a consequence, we obtain

$$Y_t\mathbb{1}_{\{|X_0^{(t)}|\leq t\}} = g_i(t - X_0^{(t)})\left(\frac{(t-X_0^{(t)})\vee 1}{t}\right)^q \mathbb{1}_{\{|X_0^{(t)}|\leq t\}} \to \frac{1}{\mu^q} \quad \mathbb{P}_i\text{-a.s.},$$

as $t \to \infty$, and it remains to find a majorant to apply the dominated convergence theorem. We find

$$\begin{aligned}Y_t\mathbb{1}_{\{|X_0^{(t)}|\leq t\}}\mathbb{1}_{\{t-X_0^{(t)}\geq 1\}} &= g_i(t - X_0^{(t)})\left(\frac{t-X_0^{(t)}}{t}\right)^q \mathbb{1}_{\{-t\leq X_0^{(t)}\leq t-1\}}\\ &\leq \sup_{s\geq 1} g_i(s)\cdot\left(\frac{t-X_0^{(t)}}{t}\right)^q \mathbb{1}_{\{-t\leq X_0^{(t)}\leq t-1\}} \leq C\cdot\left(\frac{2t}{t}\right)^q \mathbb{1}_{\{-t\leq X_0^{(t)}\leq t-1\}} \leq 2^q C < \infty,\end{aligned}$$

and since $t \geq 1$,

$$Y_t\mathbb{1}_{\{|X_0^{(t)}|\leq t\}}\mathbb{1}_{\{1>t-X_0^{(t)}\geq 0\}} = g_i(t - X_0^{(t)})\left(\frac{1}{t}\right)^q \mathbb{1}_{\{1>t-X_0^{(t)}\geq 0\}} \leq g_i(t - X_0^{(t)})\mathbb{1}_{\{1>t-X_0^{(t)}\geq 0\}}$$

$$\leq \sup_{0\leq s<1} g_i(s)\mathbb{1}_{\{1>t-X_0^{(t)}\geq 0\}} \leq \sup_{0\leq s<1} g_i(s) < \infty.$$

Hence, the dominated convergence theorem yields

$$\mathbb{E}_i\left(Y_t\mathbb{1}_{\{|X_0^{(t)}|\leq t\}}\right) \to \frac{1}{\mu^q}$$

and it suffices to remark

$$\begin{aligned}\mathbb{E}_i\left(Y_t\mathbb{1}_{\{|X_0^{(t)}|>t\}}\right) &= \mathbb{E}_i\left(g_i(t-X_0^{(t)})\left(\frac{(t-X_0^{(t)})\vee 1}{t}\right)^q \mathbb{1}_{\{X_0^{(t)}<-t\}}\right)\\ &\leq \frac{C}{t^q}\mathbb{E}\left(\left(2|X_0^{(t)}|\vee 1\right)^q \mathbb{1}_{\{X_0^{(t)}<-t\}}\right) \leq \frac{2^qC}{t^q}\sup_t \mathbb{E}|X_0^{(t)}|^q \to 0,\end{aligned}$$

as $t\to\infty$. So (3.18) follows and yields the L_i^p convergence at least along sequences. □

Proof of Theorem 3.2. We know from Lemma 3.13 that $(X_0^{(n)})_{n\geq 2}$ is tight and that it is also L^p-bounded for every $p\geq 1$. By Lemma 3.15, we also have $\mathbb{E}_i\sigma_1(i)^p<\infty$. Since $S_{\sigma_1(i)}$ is non-negative, Lemma 3.20 implies

$$\frac{\nu(X_0^{(n)},\log n)}{\log n} \xrightarrow{L^q} \frac{1}{\mu}$$

under $\mathbb{P}_i$ for every $0<q\leq p$, and with Lemma 2.12, we infer

$$\mathbb{E}_i\left|\frac{D_n}{\log n}-\frac{1}{\mu}\right|^p = \mathbb{E}_i\left|\frac{\nu(X_0^{(n)},\log n)}{\log n}-\frac{1}{\mu}\right|^p$$

for all $n\geq 2$, which completes the proof. □

3.3.3. Asymptotic expansion of the mean of D_n

In the following, the lattice type of the MRW matters. In Markov renewal theory, the important results in the d-arithmetic case are commonly stated in the easier case of a vanishing shift function, a case that we mostly cannot guarantee. Whenever the shift function does not vanish everywhere, we can consider the MRW

$$(\xi_n, S_n-\beta(\xi_n)+\beta(\xi_0))_{n\geq 0} = (\xi_n,\widetilde{S}_n)_{n\geq 0}$$

instead of $(\xi_n,S_n)_{n\geq 0}$. This process is again d-arithmetic and its shift function is 0. It actually has the same stationary drift as $(\xi_n,S_n)_{n\geq 0}$, which follows easily from stationarity. However, if we start with a MRW $(\xi_n,S_n)_{n\geq 0}$ with non-negative increments, then the transition to $(\xi_n,\widetilde{S}_n)_{n\geq 0}$ may be at the cost of the non-negativity, so in general the latter is only a MRW. We set $\beta_{ij}:=\beta(j)-\beta(i)\in(-d,d)$ and define $\{x\}:=x-\lfloor x\rfloor$ to be the fractional part of $x\in\mathbb{R}$.

In Section 2.5, we already mentioned that the mean of a first passage time connects in a nice way with renewal functions. We will show that the asymptotic behavior of certain renewal functions forms the basis for an asymptotic expansion of the mean of D_n.

Theorem 3.21. *Let $(\xi_n, S_n)_{n\geq 0}$ be a non-arithmetic ($d = 0$) or d-arithmetic zero-delayed MRW with positive recurrent and discrete driving chain Ξ_0 which has stationary distribution π. Let μ be positive and let $\mathbb{E}_\pi X_1^2 < \infty$. Then $\mathbb{U}^{ij}(t) < \infty$ for all $i, j \in \mathcal{S}$ and $t \in \mathbb{R}$. Furthermore, as $t \to \infty$,*

$$\mathbb{U}^{ij}(t) = \frac{\pi_j t}{\mu} + \frac{\pi_j^2 \mathbb{E}_j S_{\sigma_1(j)}^2}{2\mu^2} - \frac{\pi_j \mathbb{E}_i S_{\sigma_1(j)}}{\mu} \mathbb{1}_{\{j\neq i\}} + o(1) \tag{3.19}$$

in the non-arithmetic case. In the d-arithmetic case with shift function 0,

$$\mathbb{U}^{ij}(t) = \frac{\pi_j t}{\mu} + \frac{\pi_j d}{\mu}\left(\frac{1}{2} - \left\{\frac{t}{d}\right\}\right) + \frac{\pi_j^2 \mathbb{E}_j S_{\sigma_1(j)}^2}{2\mu^2} - \frac{\pi_j \mathbb{E}_i S_{\sigma_1(j)}}{\mu} \mathbb{1}_{\{j\neq i\}} + o(1). \tag{3.20}$$

If the shift function $\beta : \mathcal{S} \to [0, d)$ does not vanish, then, as $t \to \infty$,

$$\mathbb{U}^{ij}(t) = \frac{\pi_j t}{\mu} + \frac{\pi_j d}{\mu}\left(\frac{1}{2} - \left\{\frac{t-\beta_{ij}}{d}\right\}\right) + \frac{\pi_j^2 \mathbb{E}_j S_{\sigma_1(j)}^2}{2\mu^2} - \frac{\pi_j \mathbb{E}_i S_{\sigma_1(j)}}{\mu} \mathbb{1}_{\{j\neq i\}} + o(1). \tag{3.21}$$

Remark 3.22. If the MRW is d-arithmetic and we further suppose that $S^+_{\sigma_1(j)}$ has an exponential moment w.r.t. $\mathbb{P}_i$ and $\mathbb{P}_j$, then it can even be shown that there is an $r > 0$ such that the error terms in (3.20) and (3.21) are of order $o(e^{-rt})$ instead of $o(1)$. This follows from [Sto65] and the fact that every $S_{\sigma_1(j)}$ is d-arithmetic under $\mathbb{P}_j$.

Remark 3.23. We stated in Remark 3.4 that the error term in (3.4) can be arbitrarily bad. The reason for this is that the error term in (3.19) can be arbitrarily bad, cf. [Car83, Section 5]. There are more precise expansions of the renewal function, e.g. if the corresponding inter-arrival times are *strictly nonlattice* or *nonlattice of type p*, cf. [Car83]. Nevertheless, employing these results requires a more detailed investigation of issues like lattice type universality (as we deal with a MRW) from which we refrain.

Proof. The proof draws on results from standard renewal theory: We cite from [Als15b, Lemma 9.28, in Ger.] that $\mathbb{E}_\pi X_1^2 < \infty$ ensures $\mathbb{E}_i S^2_{\sigma_1(j)} < \infty$ for all $i, j \in \mathcal{S}$ and in particular, $\mathbb{E}_i S^2_{\sigma_1(i)} < \infty$ for all $i \in \mathcal{S}$. Recall that $\mathbb{U}^{ii}$ is the ordinary renewal measure of the SRW $(S_{\sigma_n(i)})_{n\geq 0}$ w.r.t. $\mathbb{P}_i$. Thus, $\mathbb{E}_i S^2_{\sigma_1(i)} < \infty$ is a sufficient condition for $\mathbb{U}^{ii}(t) < \infty$ for all $t \in \mathbb{R}$ and likewise, the $i = j$-versions of (3.19) and (3.20) follow immediately from i.i.d. renewal theory noting that $\mathbb{E}_i S_{\sigma_1(i)} = m_{ii}\mu = \frac{\mu}{\pi_i}$. If $i \neq j$, then we use

$$\mathbb{U}^{ij}(t) = \int_\mathbb{R} \mathbb{U}^{jj}(t-x) F_{ij}(\mathrm{d}x).$$

Together with $\sup_{t\geq 1} t^{-1}\mathbb{U}^{ii}(t) < \infty$ and $\mathbb{E}_i S_{\sigma_1(j)} < \infty$ this easily implies $\mathbb{U}^{ij}(t) < \infty$ for all $t \in \mathbb{R}$. Concerning the asymptotic expansion, we denote by $\varepsilon(t) = o(1)$, $t \to \infty$, the error term in the $i = j$-version of (3.19) and (3.20), respectively. Since $\mathbb{U}^{jj}(t) < \infty$ for all $t \in \mathbb{R}$, we also have $|\varepsilon(t)| < \infty$ for all $t \in \mathbb{R}$ and w.l.o.g. we can assume ε to be continuous. In the non-arithmetic case for $t \geq 0$

$$\begin{aligned}
&\left|\mathbb{U}^{ij}(t) - \left(\frac{\pi_j t}{\mu} + \frac{\pi_j^2 \mathbb{E}_j S_{\sigma_1(j)}^2}{2\mu^2} - \frac{\pi_j \mathbb{E}_i S_{\sigma_1(j)}}{\mu}\right)\right| \\
&= \left|\int_\mathbb{R} \mathbb{U}^{jj}(t-x) - \frac{\pi_j(t-x)}{\mu} - \frac{\pi_j^2 \mathbb{E}_j S_{\sigma_1(j)}^2}{2\mu^2} F_{ij}(\mathrm{d}x)\right|
\end{aligned}$$

$$\leq \int_{(-\infty,t]} \left| \mathbb{U}^{jj}(t-x) - \frac{\pi_j(t-x)}{\mu} - \frac{\pi_j^2 \mathbb{E}_j S_{\sigma_1(j)}^2}{2\mu^2} \right| F_{ij}(\mathrm{d}x)$$
$$+ \int_{(t,\infty)} \left(\mathbb{U}^{jj}(0) + \frac{\pi_j(t+x)}{\mu} + \frac{\pi_j^2 \mathbb{E}_j S_{\sigma_1(j)}^2}{2\mu^2} \right) F_{ij}(\mathrm{d}x).$$

The first term equals $\int_{(-\infty,t]} |\varepsilon(t-x)| F_{ij}(\mathrm{d}x)$ and vanishes, as $t \to \infty$, by the dominated convergence theorem, and the second term tends to zero since $\mathbb{E}_i S_{\sigma_1(j)} < \infty$ (and $\mathbb{U}^{jj}(0) < \infty$ as $\mathbb{E}_j S_{\sigma_1(j)} > 0$).

We now consider the d-arithmetic case with shift function 0. By definition of d as the lattice span of all F_{ii}, the measure $\mathbb{U}^{ii}$ is concentrated on $d\mathbb{Z}$. Since the shift is 0, X_1 and thus $S_{\sigma_1(j)}$ is also concentrated on $d\mathbb{Z}$ $\mathbb{P}_i$-a.s. Hence, as $t \to \infty$,

$$\left| \mathbb{U}^{ij}(t) - \left(\frac{\pi_j t}{\mu} + \frac{\pi_j d}{\mu}\left(\frac{1}{2} - \left\{\frac{t}{d}\right\}\right) + \frac{\pi_j^2 \mathbb{E}_j S_{\sigma_1(j)}^2}{2\mu^2} - \frac{\pi_j \mathbb{E}_i S_{\sigma_1(j)}}{\mu} \right) \right|$$
$$= \left| \int_{\mathbb{R}} \mathbb{U}^{jj}(t-x) - \frac{\pi_j(t-x)}{\mu} - \frac{\pi_j d}{\mu}\left(\frac{1}{2} - \left\{\frac{t-x}{d}\right\}\right) - \frac{\pi_j^2 \mathbb{E}_j S_{\sigma_1(j)}^2}{2\mu^2} F_{ij}(\mathrm{d}x) \right|$$
$$\leq \int_{(-\infty,t]} |\varepsilon(t-x)|\, F_{ij}(\mathrm{d}x) + \int_{(t,\infty)} \left(\mathbb{U}^{jj}(0) + \frac{\pi_j(t+x)}{\mu} + \frac{\pi_j d}{\mu}\left(\frac{1}{2} + 1\right) + \frac{\pi_j^2 \mathbb{E}_j S_{\sigma_1(j)}^2}{2\mu^2} \right) F_{ij}(\mathrm{d}x).$$

Both terms vanish, as $t \to \infty$. If β is not 0, then we consider the corresponding MRW with vanishing shift function via the connection

$$\mathbb{U}^{ij}(t) = \mathbb{E}_i \left(\sum_{n=0}^{\infty} \mathbb{1}_{\{j\}\times(-\infty,t]}(\xi_n, S_n) \right) = \mathbb{E}_i \left(\sum_{n=0}^{\infty} \mathbb{1}_{\{j\}\times(-\infty,t]}(\xi_n, \widetilde{S}_n + (\beta(\xi_n) - \beta(i))) \right)$$
$$= \mathbb{E}_i \left(\sum_{n=0}^{\infty} \mathbb{1}_{\{j\}\times(-\infty,t]}(\xi_n, \widetilde{S}_n + \beta_{ij}) \right) = \mathbb{E}_i \left(\sum_{n=0}^{\infty} \mathbb{1}_{\{j\}\times(-\infty,t-\beta_{ij}]}(\xi_n, \widetilde{S}_n) \right) = \widetilde{\mathbb{U}}^{ij}\left(t - \beta_{ij}\right),$$

where $\widetilde{\mathbb{U}}^{ij}$ is the $\mathbb{U}^{ij}$-analogue corresponding to $(\xi_n, \widetilde{S}_n)_{n\geq 0}$. Of course, π is also the stationary distribution of this driving chain, the transformed MRW has the same lattice span d, the stationary drift remains unchanged and

$$\mathbb{E}_\pi \widetilde{X}_1^2 = \mathbb{E}_\pi \left(X_1 - \beta(\xi_1) + \beta(\xi_0)\right)^2 = \mathbb{E}_\pi X_1^2 - 2\mathbb{E}_\pi X_1(\beta(\xi_1) - \beta(\xi_0)) + \mathbb{E}_\pi(\beta(\xi_1) - \beta(\xi_0))^2$$
$$\leq \mathbb{E}_\pi X_1^2 + 2d\mu + d^2 < \infty.$$

Altogether, (3.21) follows from (3.20) and from the observations

$$\widetilde{S}_{\sigma_1(j)}^2 = \left(S_{\sigma_1(j)} - \beta(\xi_{\sigma_1(j)}) + \beta(\xi_0)\right)^2 = \left(S_{\sigma_1(j)} - \beta(j) + \beta(j)\right)^2 = S_{\sigma_1(j)}^2 \quad \mathbb{P}_j\text{-a.s.}$$

and

$$\widetilde{S}_{\sigma_1(j)} = S_{\sigma_1(j)} - \beta(\xi_{\sigma_1(j)}) + \beta(\xi_0) = S_{\sigma_1(j)} - \beta(j) + \beta(i) = S_{\sigma_1(j)} - \beta_{ij} \quad \mathbb{P}_i\text{-a.s.}$$

Certainly, $\sigma_1(j) = \min\{n > 0 : \xi_n = j\}$ does not change during the transformation. That $\mathbb{U}^{ij}(t) < \infty$ also holds in this last case, follows from the afore-mentioned. □

Lemma 3.24. *Let $\mathcal{S}$ be finite and let $(\xi_n, S_n)_{n\geq 0}$ be a zero-delayed MRW with a.s. non-negative increments and positive recurrent and discrete driving chain Ξ_0 which has stationary distribution π. Let μ be positive and $\mathbb{E}_\pi X_1^2 < \infty$. Then $\mathbb{E}_i \nu(t) < \infty$ for all $i \in \mathcal{S}$ and $t \in \mathbb{R}$ and furthermore:*

(a) *If* $(\xi_n, S_n)_{n\geq 0}$ *is non-arithmetic, then, as* $t\to\infty$,

$$\mathbb{E}_i\nu(t) = \frac{t}{\mu} + \frac{1}{2\mu^2}\sum_{j\in\mathcal{S}}\pi_j^2\mathbb{E}_j S_{\sigma_1(j)}^2 - \frac{1}{\mu}\sum_{j\neq i}\pi_j\mathbb{E}_i S_{\sigma_1(j)} + o(1). \tag{3.22}$$

(b) *If* $(\xi_n, S_n)_{n\geq 0}$ *is d-arithmetic with shift function* $\beta\equiv 0$, *then, as* $t\to\infty$,

$$\mathbb{E}_i\nu(t) = \frac{t}{\mu} + \frac{d}{\mu}\left(\frac{1}{2} - \left\{\frac{t}{d}\right\}\right) + \frac{1}{2\mu^2}\sum_{j\in\mathcal{S}}\pi_j^2\mathbb{E}_j S_{\sigma_1(j)}^2 - \frac{1}{\mu}\sum_{j\neq i}\pi_j\mathbb{E}_i S_{\sigma_1(j)} + o(1). \tag{3.23}$$

(c) *If* $(\xi_n, S_n)_{n\geq 0}$ *is d-arithmetic with shift function* β, *then, as* $t\to\infty$,

$$\begin{aligned}\mathbb{E}_i\nu(t) = \frac{t}{\mu} - \frac{d}{\mu}\sum_{j\in\mathcal{S}}\pi_j\left\{\frac{t-\beta_{ij}}{d}\right\} + \frac{d}{2\mu} + \frac{1}{2\mu^2}\sum_{j\in\mathcal{S}}\pi_j^2\mathbb{E}_j S_{\sigma_1(j)}^2 \\ - \frac{1}{\mu}\sum_{j\neq i}\pi_j\mathbb{E}_i S_{\sigma_1(j)} + o(1).\end{aligned} \tag{3.24}$$

Remark 3.25. In view of Remark 3.22, we remark that there exists $r>0$ such that (3.23) and (3.24) hold with error term $o(e^{-rt})$ instead of $o(1)$ if the MRW is in the d-arithmetic case and $S_{\sigma_1(j)}$ has an exponential moment w.r.t. $\mathbb{P}_i$ and every $\mathbb{P}_j$.

Proof. Since S_n increases in n,

$$\begin{aligned}\mathbb{E}_i\nu(t) &= \sum_{n=0}^{\infty}\mathbb{P}_i(\nu(t)>n) = \sum_{n=0}^{\infty}\mathbb{P}_i(S_n\leq t) = \sum_{j\in\mathcal{S}}\sum_{n=0}^{\infty}\mathbb{P}_i(\xi_n=j, S_n\leq t) \\ &= \sum_{j\in\mathcal{S}}\mathbb{E}_i\left(\sum_{n=0}^{\infty}\mathbb{1}_{\{j\}\times(-\infty,t]}(\xi_n,S_n)\right) = \sum_{j\in\mathcal{S}}\mathbb{U}_i\left(\{j\}\times(-\infty,t]\right) = \sum_{j\in\mathcal{S}}\mathbb{U}^{ij}(t),\end{aligned}$$

which allows us to apply Theorem 3.21. Thus, $\mathbb{E}_i\nu(t)<\infty$ for all $i\in\mathcal{S}$ and $t\in\mathbb{R}$ in every case. In the non-arithmetic case we have, as $t\to\infty$,

$$\begin{aligned}\mathbb{E}_i\nu(t) = \sum_{j\in\mathcal{S}}\mathbb{U}^{ij}(t) &= \frac{t}{\mu}\sum_{j\in\mathcal{S}}\pi_j + \frac{1}{2\mu^2}\sum_{j\in\mathcal{S}}\pi_j^2\mathbb{E}_j S_{\sigma_1(j)}^2 + \frac{1}{\mu}\sum_{j\neq i}\pi_j\mathbb{E}_i S_{\sigma_1(j)} + o(1) \\ &= \frac{t}{\mu} + \frac{1}{2\mu^2}\sum_{j\in\mathcal{S}}\pi_j^2\mathbb{E}_j S_{\sigma_1(j)}^2 + \frac{1}{\mu}\sum_{j\neq i}\pi_j\mathbb{E}_i S_{\sigma_1(j)} + o(1).\end{aligned}$$

Analogously, in the d-arithmetic case with shift 0 we obtain

$$\mathbb{E}_i\nu(t) = \frac{t}{\mu} + \frac{d}{\mu}\left(\frac{1}{2} - \left\{\frac{t}{d}\right\}\right) + \frac{1}{2\mu^2}\sum_{j\in\mathcal{S}}\pi_j^2\mathbb{E}_j S_{\sigma_1(j)}^2 - \frac{1}{\mu}\sum_{j\neq i}\pi_j\mathbb{E}_i S_{\sigma_1(j)} + o(1)$$

and with general shift function β

$$\mathbb{E}_i\nu(t) = \frac{t}{\mu} - \frac{d}{\mu}\sum_{j\in\mathcal{S}}\pi_j\left\{\frac{t-\beta_{ij}}{d}\right\} + \frac{d}{2\mu} + \frac{1}{2\mu^2}\sum_{j\in\mathcal{S}}\pi_j^2\mathbb{E}_j S_{\sigma_1(j)}^2 - \frac{1}{\mu}\sum_{j\neq i}\pi_j\mathbb{E}_i S_{\sigma_1(j)} + o(1).$$

□

Lemma 3.26. *Given the situation of Lemma 3.24, let* $(X_0^{(t)})_t$ *be a uniformly integrable family of real-valued random variables, independent of* $(\xi_n, S_n)_{n\geq 0}$*. Then* $\mathbb{E}_i\nu(X_0^{(t)}, t) < \infty$ *for all* $i \in \mathcal{S}$ *and* $t \geq 0$ *and furthermore:*

(a) If $(\xi_n, S_n)_{n\geq 0}$ *is non-arithmetic, then, as* $t \to \infty$*,*

$$\mathbb{E}_i\nu(X_0^{(t)}, t) = \frac{t}{\mu} + \frac{1}{2\mu^2}\sum_{j\in\mathcal{S}} \pi_j^2 \mathbb{E}_j S_{\sigma_1(j)}^2 - \frac{1}{\mu}\sum_{j\neq i} \pi_j \mathbb{E}_i S_{\sigma_1(j)} - \frac{\mathbb{E}X_0^{(t)}}{\mu} + o(1). \tag{3.25}$$

(b) If $(\xi_n, S_n)_{n\geq 0}$ *is d-arithmetic with shift function* 0*, then, as* $t \to \infty$*,*

$$\begin{aligned}\mathbb{E}_i\nu(X_0^{(t)}, t) = \frac{t}{\mu} + \frac{d}{\mu}\left(\frac{1}{2} - \mathbb{E}\left\{\frac{t - X_0^{(t)}}{d}\right\}\right) + \frac{1}{2\mu^2}\sum_{j\in\mathcal{S}} \pi_j^2 \mathbb{E}_j S_{\sigma_1(j)}^2 \\ - \frac{1}{\mu}\sum_{j\neq i} \pi_j \mathbb{E}_i S_{\sigma_1(j)} - \frac{\mathbb{E}X_0^{(t)}}{\mu} + o(1).\end{aligned} \tag{3.26}$$

(c) If $(\xi_n, S_n)_{n\geq 0}$ *is d-arithmetisch with shift function* β*, then, as* $t \to \infty$*,*

$$\begin{aligned}\mathbb{E}_i\nu(X_0^{(t)}, t) = \frac{t}{\mu} - \frac{d}{\mu}\sum_{j\in\mathcal{S}} \pi_j \mathbb{E}\left\{\frac{t - X_0^{(t)} - \beta_{ij}}{d}\right\} + \frac{d}{2\mu} \\ + \frac{1}{2\mu^2}\sum_{j\in\mathcal{S}} \pi_j^2 \mathbb{E}_j S_{\sigma_1(j)}^2 - \frac{1}{\mu}\sum_{j\neq i} \pi_j \mathbb{E}_i S_{\sigma_1(j)} - \frac{\mathbb{E}X_0^{(t)}}{\mu} + o(1).\end{aligned} \tag{3.27}$$

In (3.25)*,* (3.26) *and* (3.27) *we can replace* $\mathbb{E}X_0^{(t)}$ *by* $\mathbb{E}X_0^*$ *(with error* $o(1)$*) if the family is distributionally convergent with limit* X_0^* *w.r.t.* $\mathbb{P}$*.*

Remark 3.27. In view of Remarks 3.22 and 3.25, we remark that there exists $r > 0$ such that (3.26) and (3.27) hold with error term $o(e^{-rt})$ instead of $o(1)$ if the MRW is in the d-arithmetic case, $S_{\sigma_1(j)}$ has an exponential moment w.r.t. $\mathbb{P}_i$ and every $\mathbb{P}_j$, and $(\exp(s|X_0^{(t)}|))_t$ is uniformly integrable for some $s > 0$.

Proof. We use the identity

$$\nu(X_0^{(t)}, t) = \nu(t - X_0^{(t)})\mathbb{1}_{\{t - X_0^{(t)} \geq 0\}}$$

and recall $\mathbb{E}_i\nu(X_0^{(t)}, t) < \infty$ for all $t \geq 0$ from the proof of Lemma 3.20, which can also be inferred from Lemma 3.24. We start with the non-arithmetic case. Let $t \geq 0$. Independence yields

$$\begin{aligned}&\left|\mathbb{E}_i\nu(X_0^{(t)}, t) - \left(\frac{t}{\mu} + \frac{1}{2\mu^2}\sum_{j\in\mathcal{S}} \pi_j^2 \mathbb{E}_j S_{\sigma_1(j)}^2 - \frac{1}{\mu}\sum_{j\neq i} \pi_j \mathbb{E}_i S_{\sigma_1(j)} - \frac{\mathbb{E}X_0^{(t)}}{\mu}\right)\right| \\ &= \left|\int_{(-\infty,t]} \mathbb{E}_i\nu(t-s)\mathbb{P}^{X_0^{(t)}}(\mathrm{d}s) - \frac{t}{\mu} - \frac{1}{2\mu^2}\sum_{j\in\mathcal{S}} \pi_j^2 \mathbb{E}_j S_{\sigma_1(j)}^2 + \frac{1}{\mu}\sum_{j\neq i} \pi_j \mathbb{E}_i S_{\sigma_1(j)} + \frac{\mathbb{E}X_0^{(t)}}{\mu}\right| \\ &\leq \int_{(-\infty,t]} \left|\mathbb{E}_i\nu(t-s) - \frac{t-s}{\mu} - \frac{1}{2\mu^2}\sum_{j\in\mathcal{S}} \pi_j^2 \mathbb{E}_j S_{\sigma_1(j)}^2 + \frac{1}{\mu}\sum_{j\neq i} \pi_j \mathbb{E}_i S_{\sigma_1(j)}\right| \mathbb{P}^{X_0^{(t)}}(\mathrm{d}s)\end{aligned}$$

$$+\left(\frac{t}{\mu}+\frac{1}{2\mu^2}\sum_{j\in\mathcal{S}}\pi_j^2\mathbb{E}_jS_{\sigma_1(j)}^2+\frac{1}{\mu}\sum_{j\neq i}\pi_j\mathbb{E}_iS_{\sigma_1(j)}\right)\mathbb{P}(X_0^{(t)}>t)+\mathbb{E}X_0^{(t)}\mathbb{1}_{(t,\infty)}(X_0^{(t)}).$$

In this estimation, the first term vanishes, as $t\to\infty$, because of (3.22), the tightness of $(X_0^{(t)})_t$ and the dominated convergence theorem, whereas the second and the third term tend to 0 since $(X_0^{(t)})_t$ is uniformly integrable. We treat the d-arithmetic case with shift 0 similarly:

$$\left|\mathbb{E}_i\nu(X_0^{(t)},t)-\left(\frac{t}{\mu}+\frac{d}{\mu}\left(\frac{1}{2}-\mathbb{E}\left\{\frac{t-X_0^{(t)}}{d}\right\}\right)+\frac{1}{2\mu^2}\sum_{j\in\mathcal{S}}\pi_j^2\mathbb{E}_jS_{\sigma_1(j)}^2-\frac{1}{\mu}\sum_{j\neq i}\pi_j\mathbb{E}_iS_{\sigma_1(j)}-\frac{\mathbb{E}X_0^{(t)}}{\mu}\right)\right|$$
$$\leq\int_{(-\infty,t]}\Bigg|\mathbb{E}_i\nu(t-s)-\frac{t-s}{\mu}-\frac{d}{\mu}\left(\frac{1}{2}-\left\{\frac{t-s}{d}\right\}\right)$$
$$-\frac{1}{2\mu^2}\sum_{j\in\mathcal{S}}\pi_j^2\mathbb{E}_jS_{\sigma_1(j)}^2+\frac{1}{\mu}\sum_{j\neq i}\pi_j\mathbb{E}_iS_{\sigma_1(j)}\Bigg|\,\mathbb{P}^{X_0^{(t)}}(\mathrm{d}s)$$
$$+\left(\frac{t}{\mu}+\frac{d}{2\mu}+\frac{1}{2\mu^2}\sum_{j\in\mathcal{S}}\pi_j^2\mathbb{E}_jS_{\sigma_1(j)}^2+\frac{1}{\mu}\sum_{j\neq i}\pi_j\mathbb{E}_iS_{\sigma_1(j)}\right)\mathbb{P}(X_0^{(t)}>t)$$
$$+\mathbb{E}X_0^{(t)}\mathbb{1}_{(t,\infty)}(X_0^{(t)})+\frac{d}{\mu}\mathbb{E}\left\{\frac{t-X_0^{(t)}}{d}\right\}\mathbb{1}_{(t,\infty)}(X_0^{(t)}).$$

Again, the upper bound converges to 0 as $t\to\infty$. Note that

$$\mathbb{E}\left\{\frac{t-X_0^{(t)}}{d}\right\}\leq 1$$

for every t. The case when β does not vanish is similar.

If the family is distributionally convergent with limit X_0^*, then we can replace $\mathbb{E}X_0^{(t)}$ by $\mathbb{E}X_0^*$ in all three cases since $(X_0^{(t)})_t$ is uniformly integrable. □

The next lemma from [Jan06] gives a series representation for the expectation of the fractional part in (3.26) and (3.27).

Lemma 3.28. *Suppose that X has a continuous distribution, $\mathbb{E}|X|<\infty$, and the Fourier transform $\phi(t):=\mathbb{E}e^{\mathrm{i}tX}$ satisfies $\phi(t)=\mathcal{O}(|t|^{-\delta})$ for some $\delta>0$. Then for all $u\in\mathbb{R}$*

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

Proof. Cf. [Jan06, Theorem 2.3]. □

Proof of Theorem 3.3. $(\xi_n,S_n)_{n\geq 0}$ is a MRW satisfying the requirements for Lemma 3.26. Since $X_1=-\log p_{\xi_0,\xi_1}$ is bounded, we have $\mathbb{E}_\pi X_1^2<\infty$. From Lemma 3.13 we extract that $X_0^{(n)}$ converges to X_0^* in distribution with $-X_0^*$ having a Gumbel(0, 1) distribution and that $(X_0^{(n)})_{n\geq 2}$ is uniformly integrable and tight. Thus, we apply Lemmas 3.26 and 2.12 which yield, as $n\to\infty$,

$$\mathbb{E}_iD_n=\mathbb{E}_i\nu(X_0^{(n)},\log n)=\frac{\log n}{\mu}+\frac{1}{2\mu^2}\sum_{j\in\mathcal{S}}\pi_j^2\mathbb{E}_jS_{\sigma_1(j)}^2-\frac{\mathbb{E}X_0^{(n)}}{\mu}-\frac{1}{\mu}\sum_{j\neq i}\pi_j\mathbb{E}_iS_{\sigma_1(j)}+o(1)$$

in the non-arithmetic case. Here we can substitute $\mathbb{E}X_0^{(n)}$ by $\mathbb{E}X_0^*$, and $-X_0^*$ is integrable with Fourier transform $\phi(t) = \Gamma(1 - \mathrm{i}\, t)$ and

$$\mathbb{E}(-X_0^*) = (-\mathrm{i})\phi'_{-X_0^*}(0) = (-i)^2\Gamma'(1) = -(-\gamma) = \gamma.$$

The error that emerges when considering $\mathbb{E}X_0^*$ instead of $\mathbb{E}X_0^{(n)}$ can even be given more precisely than $o(1)$. With an appeal to Lemma 3.13(c) and a well-known asymptotic expansion of the harmonic sum, we obtain

$$\begin{aligned}\mathbb{E}X_0^{(n)} &= \log n - \mathbb{E}\max\{Z_1, \dots, Z_{n-1}\} = \log n - \sum_{k=1}^{n-1}\frac{1}{k}\\ &= \log n - \log(n-1) - \gamma - \frac{1}{2(n-1)} + \mathcal{O}(n^{-2})\\ &= -\gamma + \log\left(1 + \frac{1}{n-1}\right) + \mathcal{O}(n^{-1}) = -\gamma + \mathcal{O}(n^{-1}),\end{aligned}$$

as $n \to \infty$, where $Z_1, Z_2, \dots$ are i.i.d. Exp(1) distributed random variables. In the d-arithmetic case with shift function 0, this reasoning leads to

$$\begin{aligned}\mathbb{E}_i D_n = \frac{\log n}{\mu} &+ \frac{d}{\mu}\left(\frac{1}{2} - \mathbb{E}\left\{\frac{\log n - X_0^{(n)}}{d}\right\}\right) + \frac{1}{2\mu^2}\sum_{j\in\mathcal{S}}\pi_j^2\mathbb{E}_j S_{\sigma_1(j)}^2\\ &- \frac{1}{\mu}\sum_{j\neq i}\pi_j\mathbb{E}_i S_{\sigma_1(j)} + \frac{\gamma}{\mu} + o(1),\end{aligned}$$

as $n \to \infty$ (we could replace $o(1)$ by $\mathcal{O}(n^{-(r\wedge 1)})$ since our situation meets all conditions required in Remark 3.27, cf. Lemmas 3.13 and 3.15). We would like to have X_0^* instead of $X_0^{(n)}$ in the expectation of the fractional part, so we need to estimate their asymptotical distance. First, as in [Jan12a], we use Lemma 3.28 ($|\Gamma(1 - \mathrm{i}\, s)| \sim \sqrt{2\pi}|s|^{1/2}e^{-\pi|s|/2}$) to compute

$$\mathbb{E}\left\{\frac{t - X_0^*}{d}\right\} = \frac{1}{2} - \sum_{k\neq 0}\frac{\Gamma(1 - 2\pi\,\mathrm{i}\, k/d)}{2\pi\,\mathrm{i}\, k}e^{2\pi\,\mathrm{i}\, kt/d}$$

and

$$d\left(\frac{1}{2} - \mathbb{E}\left\{\frac{t - X_0^*}{d}\right\}\right) = -\sum_{k\neq 0}\Gamma(-2\pi\,\mathrm{i}\, k/d)e^{2\pi\,\mathrm{i}\, kt/d} = \psi_1(t).$$

The latter series converges uniformly and is therefore continuous (as a function of t).

To compute the distance between the two expected fractional parts, let $M_n := \max\{Z_1, \dots, Z_n\}$ and $f_n(x) = ne^{-x}(1 - e^{-x})^{n-1}$ its $\lambda\!\!\lambda_{[0,\infty)}$-density. Furthermore, let $g(x) = ne^{-x}e^{-ne^{-x}}$ be the density of $\log n - X_0^*$. We will now show that

$$\begin{aligned}&\left|\mathbb{E}\left\{\frac{\log n - X_0^*}{d}\right\} - \mathbb{E}\left\{\frac{\log n - X_0^{(n)}}{d}\right\}\right| = \left|\mathbb{E}\left\{\frac{\log n - X_0^*}{d}\right\} - \mathbb{E}\left\{\frac{M_{n-1}}{d}\right\}\right|\\ &\leq \int_{-\infty}^{\infty}|g(t) - f_{n-1}(t)|\left\{\frac{t}{d}\right\}\mathrm{d}t = \int_{-\infty}^{\infty} ne^{-t}\left|e^{-ne^{-t}} - \frac{n-1}{n}(1 - e^{-t})^{n-2}\mathbb{1}_{\{t>0\}}\right|\left\{\frac{t}{d}\right\}\mathrm{d}t \quad (3.28)\\ &= \int_{-\infty}^{\infty} e^{-s}\left|e^{-e^{-s}} - \frac{n-1}{n}\Big(1 - \frac{e^{-s}}{n}\Big)^{n-2}\mathbb{1}_{\{s>-\log n\}}\right|\left\{\frac{s - \log n}{d}\right\}\mathrm{d}s = o(1),\end{aligned}$$

as $n \to \infty$, with the dominated convergence theorem. Note that the occurring fraction part is bounded by 1, thus we have to find an appropriate bound for the expression inside the absolute value. Obviously, it converges to 0 pointwise for all $s \in \mathbb{R}$. Indeed, we will show that

$$\left| e^{-e^{-s}} - \frac{n-1}{n}\Big(1 - \frac{e^{-s}}{n}\Big)^{n-2} \mathbb{1}_{\{s > -\log n\}} \right| \le C e^{-e^{-s}} \tag{3.29}$$

for all $n \ge 2$, $s \in \mathbb{R}$ and a positive constant $C > 0$, and thus

$$\int_{-\infty}^{\infty} e^{-s} \left| e^{-e^{-s}} - \frac{n-1}{n}\Big(1 - \frac{e^{-s}}{n}\Big)^{n-2} \mathbb{1}_{\{s > -\log n\}} \right| \left\{ \frac{s - \log n}{d} \right\} \mathrm{d}s \le C \int_{-\infty}^{\infty} e^{-s} e^{-e^{-s}} \mathrm{d}s = C < \infty$$

since the integrand is the density of the Gumbel distribution. In order to achieve the remaining estimate (3.29), let $x = e^{-s} > 0$ and note that

$$e^{-x} - \frac{n-1}{n}\Big(1 - \frac{x}{n}\Big)^{n-2} \mathbb{1}_{\{x < n\}} \le e^{-x}$$

for all $n \ge 2$. This leaves us with the verification of the lower bound

$$e^{-x} - \frac{n-1}{n}\Big(1 - \frac{x}{n}\Big)^{n-2} \mathbb{1}_{\{x < n\}} \ge -C e^{-x} \quad \text{or} \quad \frac{n-1}{n}\Big(1 - \frac{x}{n}\Big)^{n-2} \mathbb{1}_{\{x < n\}} \le (1 + C) e^{-x}$$

for some $C > 0$. Obviously, it is enough to show $(1 - \frac{x}{n})^{n-2} \mathbb{1}_{\{x<n\}} \le (1 + C)e^{-x}$ for all $n \ge 2$ and $x > 0$ (note that it is trivial for $x \ge n$) or equivalently $(n - 2)\log(1 - \frac{x}{n}) \le -x + \log(1 + C)$ for all $n \ge 2$ and $x \in (0, n)$. Using the series expansion of the logarithm, this amounts to

$$-(n-2) \sum_{k=1}^{\infty} \frac{(\frac{x}{n})^k}{k} \le -\frac{x}{n} \cdot n + \log(1 + C) \quad \text{or} \quad \frac{2x}{n} - \log(1 + C) \le (n-2) \sum_{k=2}^{\infty} \frac{(\frac{x}{n})^k}{k}$$

which is true for $C = e^2 - 1$ and all $n \ge 2$ and $x \in (0, n)$ since the left-hand side then is $\frac{2x}{n} - \log(1 + C) < 2 - \log(e^2) = 0$ and the right-hand side is non-negative. This proves (3.28).

In the case of a general β, we similarly get

$$\begin{aligned}
\mathbb{E}_i D_n = \mathbb{E}_i \nu(X_0^{(n)}, \log n) &= \frac{\log n}{\mu} + \frac{1}{2\mu^2} \sum_{j \in \mathcal{S}} \pi_j^2 \mathbb{E}_j S_{\sigma_1(j)}^2 - \frac{\mathbb{E} X_0^*}{\mu} \\
&\quad - \frac{1}{\mu} \sum_{j \ne i} \pi_j \mathbb{E}_i S_{\sigma_1(j)} + \frac{1}{\mu} \sum_{j \in \mathcal{S}} \pi_j d \left(\frac{1}{2} - \mathbb{E} \left\{ \frac{\log n - \beta_{ij} - X_0^*}{d} \right\} \right) + o(1) \\
&= \frac{\log n}{\mu} + \frac{1}{2\mu^2} \sum_{j \in \mathcal{S}} \pi_j^2 \mathbb{E}_j S_{\sigma_1(j)}^2 - \frac{1}{\mu} \sum_{j \ne i} \pi_j \mathbb{E}_i S_{\sigma_1(j)} + \frac{\gamma}{\mu} + \frac{1}{\mu} \sum_{j \in \mathcal{S}} \pi_j \psi_1(\log n - \beta_{ij}) + o(1).
\end{aligned}$$

□

3.3.4. Central limit theorem for D_n

We now prove a CLT for D_n by deriving results for $\nu(t)$ and then for $\nu(X_0^{(t)}, t)$.

Lemma 3.29. *Let $(\xi_n, S_n)_{n \ge 0}$ be a zero-delayed MRW with a.s. non-negative increments and positive recurrent and discrete driving chain Ξ_0 which has stationary distribution π. Let $\mu < \infty$ be positive. Assume further that $\mathbb{E}_i \sigma_1(i)^2 < \infty$, $\mathbb{E}_\pi X_1^2 < \infty$ and*

$$\sigma^{(2)} := \frac{1}{m_{ii}} \mathbb{V}\mathrm{ar}_i(S_{\sigma_1(i)} - \mu \sigma_1(i)) > 0.$$

Then $\sigma^{(2)} < \infty$ does not depend on i, and, as $t \to \infty$,

$$\frac{\nu(t) - \frac{t}{\mu}}{\sqrt{t}} \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, \frac{\sigma^{(2)}}{\mu^3}\right) \tag{3.30}$$

w.r.t. $\mathbb{P}_i$. If $\sigma^{(2)} = 0$, i.e. $S_{\sigma_1(j)} = \mu\sigma_1(j)$ $\mathbb{P}_j$-a.s. for some (and thus all) $j \in \mathcal{S}$, then the limit is 0 and the convergence holds $\mathbb{P}_i$-a.s. Assuming further $\mathbb{E}_i\sigma_1(i)^p < \infty$ and $\mathbb{E}_i S^p_{\sigma_1(i)} < \infty$ for some $p \geq 2$, the family $\{|(\nu(t) - \frac{t}{\mu})/\sqrt{t}|^p, t \geq 1\}$ is uniformly integrable w.r.t. $\mathbb{P}_i$ and

$$\mathbb{E}_i \left| \frac{\nu(t) - \frac{t}{\mu}}{\sqrt{t}} \right|^q \to \mathbb{E}_i |N|^q$$

for $0 < q \leq p$ in the case $\sigma^{(2)} > 0$. N has the distribution of the limit in (3.30). The convergence also holds without absolute value for $q \in \mathbb{N}$. If $\sigma^{(2)} = 0$, then

$$\frac{\nu(t) - \frac{t}{\mu}}{t^{1/p}} \to 0 \quad \mathbb{P}_i\text{-a.s.},$$

$\{|\nu(t) - \frac{t}{\mu}|^p/t, t \geq 1\}$ is uniformly integrable w.r.t. $\mathbb{P}_i$ and, in particular,

$$\mathbb{E}_i \left| \nu(t) - \frac{t}{\mu} \right|^p = o(t).$$

In both cases ($\sigma^{(2)} > 0, = 0$), as $t \to \infty$,

$$\mathbb{V}\mathrm{ar}_i(\nu(t)) = \frac{\sigma^{(2)}}{\mu^3} t + o(t) \quad \text{and} \quad \mathbb{E}_i \left(\nu(t) - \frac{t}{\mu}\right)^2 = \frac{\sigma^{(2)}}{\mu^3} t + o(t). \tag{3.31}$$

Remark 3.30. $\sigma^{(2)}$ from Lemma 3.29 does not depend on i. In fact, Meyn and Tweedie derive limit theorems for (in our terms) $S_n(g) := \sum_{k=1}^n g(\xi_k, X_k)$ for some function g, cf. [MT93, Chapter 17]. With $g = p_2$ being the projection on the second component, [MT93, Theorem 17.2.2] is a CLT for our auxiliary MRW S_n. The limiting variance there is $\sigma^{(2)}$, so by [MT93, Proposition 17.1.6] it is the limiting variance under every $\mathbb{P}_j$, and thus cannot depend on i.

Remark 3.31. If $\mathbb{V}\mathrm{ar}_i(S_{\sigma_1(i)} - \mu\sigma_1(i)) = 0$ for some $i \in \mathcal{S}$, then the MRW $(\xi_n, S_n - \mu n)_{n\geq 0}$ is *null-homologous*, cf. Section 2.9. Thus, Lemma 2.16 is another way of seeing that $\mathbb{V}\mathrm{ar}_i(S_{\sigma_1(i)} - \mu\sigma_1(i)) = 0$ for all $i \in \mathcal{S}$ if it holds for some $i \in \mathcal{S}$.

Remark 3.32. Although one might expect $\sigma^{(2)}$ to be the stationary variance of X_1, it is not (except for the i.i.d. case), cf. [MT93, Sections 17.2.2 and 17.4.3]. We have the alternative form

$$\sigma^{(2)} = \frac{1}{m_{ii}} \mathbb{V}\mathrm{ar}_i(S_{\sigma_1(i)} - \mu\sigma_1(i)) = \frac{1}{m_{ii}} \mathbb{E}_i(S_{\sigma_1(i)} - \mu\sigma_1(i))^2 = \frac{1}{m_{ii}} \mathbb{E}_i \left(\sum_{k=1}^{\sigma_1(i)} (X_k - \mu)\right)^2,$$

whereas the stationary variance is

$$\frac{1}{m_{ii}} \mathbb{E}_i \left(\sum_{k=1}^{\sigma_1(i)} (X_k - \mu)^2\right).$$

We refer the read to [MT93, Section 17.4.3] for further details.

For the proof of Lemma 3.29, we use similar decomposition techniques as Meyn and Tweedie [MT93, e.g. Theorem 17.2.2] who established limit theorems for functions of Harris recurrent Markov chains in great generality. Our result and proof are very similar to those obtained by Alsmeyer and Gut [AG99], who derived limit theorems for stopped MRWs (with positive increments) with a Harris recurrent driving chain.

Proof of Lemma 3.29. The idea of the proof is a sandwiching argument together with the i.i.d. version of *Anscombe's Theorem* applied to the cyclically decomposed MRW along $(\sigma_n(i))_{n\geq 0}$.

First, with the definition of $\nu(t)$, we have

$$\frac{-\mu}{\sqrt{t}} + \frac{S_{\nu(t)-1} - (\nu(t)-1)\mu}{\sqrt{t}} \leq \frac{t - \nu(t)\mu}{\sqrt{t}} \leq \frac{S_{\nu(t)} - \nu(t)\mu}{\sqrt{t}}. \tag{3.32}$$

We want to show that, as $t \to \infty$, the left- and the right-hand side tend to $\mathrm{N}(0, \frac{\sigma^{(2)}}{\mu})$ under $\mathbb{P}_i$. Then by symmetry, as $t \to \infty$,

$$\frac{\nu(t)\mu - t}{\sqrt{t}} \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, \frac{\sigma^{(2)}}{\mu}\right)$$

w.r.t. $\mathbb{P}_i$, or equivalently (3.30). So our main goal will be to verify that the left- and the right-hand side have the desired limit. We start with

$$\frac{S_{\nu(t)} - \nu(t)\mu}{\sqrt{t}} = \frac{S_{\nu(t)} - S_{\sigma_{\nu^i(t)}(i)}}{\sqrt{t}} + \frac{S_{\sigma_{\nu^i(t)}(i)} - \mu\sigma_{\nu^i(t)}(i)}{\sqrt{t}} + \mu\frac{\sigma_{\nu^i(t)}(i) - \nu(t)}{\sqrt{t}} \tag{3.33}$$

and will show that the middle term tends to the right limit, whereas the first and last expression vanish $\mathbb{P}_i$-a.s. At first, we deal with the middle term. Since $(\sigma_{n+1}(i) - \sigma_n(i), S_{\sigma_{n+1}(i)} - S_{\sigma_n(i)})_{n\geq 0}$ is a sequence of i.i.d. random variables under $\mathbb{P}_i$, we see that $S_{\sigma_n(i)} - \mu\sigma_n(i)$ is a centered SRW, its increments having variance $\sigma_i^2 := \sigma^{(2)}m_{ii}$. We consider the case $\sigma^{(2)} > 0$ first. This, in particular, means $\sigma_j^2 > 0$ for all $j \in \mathcal{S}$, as $\sigma^{(2)}$ does not depend on i and $m_{jj} > 0$ for all $j \in \mathcal{S}$. The variance is also finite, since $\mathbb{E}_i\sigma_1(i)^2 < \infty$ and $\mathbb{E}_i S^2_{\sigma_1(i)} < \infty$, the latter being ensured by $\mathbb{E}_\pi X_1^2 < \infty$. Now, $\nu^i(t)/t \to 1/\mu m_{ii} \in (0,\infty)$ $\mathbb{P}_i$-a.s. allows us to apply the i.i.d. version of Anscombe's theorem [Gut09, Theorem 1.3.1] to obtain

$$\frac{S_{\sigma_{\nu^i(t)}(i)} - \mu\sigma_{\nu^i(t)}(i)}{\sigma_i\sqrt{\nu^i(t)}} \xrightarrow[t\to\infty]{\mathrm{d}} \mathrm{N}(0,1)$$

w.r.t. $\mathbb{P}_i$ and thus

$$\frac{S_{\sigma_{\nu^i(t)}(i)} - \mu\sigma_{\nu^i(t)}(i)}{\sqrt{t}} \xrightarrow[t\to\infty]{\mathrm{d}} \mathrm{N}\left(0, \frac{\sigma^{(2)}}{\mu}\right)$$

as desired. If $\sigma_j^2 = 0$ for some (and thus for all) $j \in \mathcal{S}$, then we know that

$$S_{\sigma_n(j)} - \mu\sigma_n(j) = \sum_{k=0}^{n-1}(S_{\sigma_{k+1}(j)} - S_{\sigma_k(j)}) - \mu(\sigma_{k+1}(j) - \sigma_k(j)) = 0 \quad \mathbb{P}_j\text{-a.s.}$$

or equivalently $S_{\sigma_n(j)} = \mu\sigma_n(j)$ $\mathbb{P}_j$-a.s. for every $n \geq 1$. It follows that the middle term in (3.33) is 0 $\mathbb{P}_j$-a.s.

To prove the vanishing of the remaining expressions in both cases $\sigma_i^2 > 0$ and $= 0$, we impose the stronger moment conditions $\mathbb{E}_i \sigma_1(i)^p < \infty$ and $\mathbb{E}_i S_{\sigma_1(i)}^p < \infty$ for $p \geq 2$ and start with

$$0 \leq \frac{\sigma_{\nu^i(t)}(i) - \nu(t)}{\nu^i(t)^{1/p}} \leq \frac{\sigma_{\nu^i(t)}(i) - \sigma_{\nu^i(t)-1}(i)}{\nu^i(t)^{1/p}}$$

by definition. The right-hand side converges to 0 by Lemma A.5 since $\nu^i(t) \to \infty$ $\mathbb{P}_i$-a.s., $\mathbb{E}_i \sigma_1(i)^p < \infty$ and $\sigma_n(i)$ is a SRW under $\mathbb{P}_i$. In particular, we get

$$\mu \frac{\sigma_{\nu^i(t)}(i) - \nu(t)}{t^{1/p}} \to 0 \quad \mathbb{P}_i\text{-a.s.} \tag{3.34}$$

as desired. This always covers the case $p = 2$. Similarly, as we require non-negative increments,

$$0 \leq \frac{S_{\sigma_{\nu^i(t)}(i)} - S_{\nu(t)}}{\nu^i(t)^{1/p}} \leq \frac{S_{\sigma_{\nu^i(t)}(i)} - S_{\sigma_{\nu^i(t)-1}(i)}}{\nu^i(t)^{1/p}}$$

and again the right-hand side tends to 0 since $\mathbb{E}_i S_{\sigma_1(i)}^p < \infty$ and thus

$$\frac{S_{\nu(t)} - S_{\sigma_{\nu^i(t)}(i)}}{t^{1/p}} \to 0 \quad \mathbb{P}_i\text{-a.s.} \tag{3.35}$$

This shows that the right-hand side in (3.32) converges to the right limit.

It remains to show that the left-hand side tends to the same limit. More precisely, we need to deal with

$$\begin{aligned} &\frac{S_{\nu(t)-1} - (\nu(t) - 1)\mu}{\sqrt{t}} \\ &= \frac{S_{\nu(t)-1} - S_{\sigma_{\nu^i(t)-1}(i)}}{\sqrt{t}} + \frac{S_{\sigma_{\nu^i(t)-1}(i)} - \mu\sigma_{\nu^i(t)-1}(i)}{\sqrt{t}} + \mu \frac{\sigma_{\nu^i(t)-1}(i) - (\nu(t) - 1)}{\sqrt{t}}. \end{aligned} \tag{3.36}$$

Now, since $\nu^i(t) - 1$ and $\nu^i(t)$ are asymptotically equivalent, as $t \to \infty$, the middle term tends to the right limit with the arguments from above, so that again we need to verify that the left and right expression vanish in the limit. To this end, we remark that

$$0 \leq (\nu(t) - 1) - \sigma_{\nu^i(t)-1}(i) \leq \sigma_{\nu^i(t)}(i) - \sigma_{\nu^i(t)-1}(i)$$

and

$$0 \leq S_{\nu(t)-1} - S_{\sigma_{\nu^i(t)-1}(i)} \leq S_{\sigma_{\nu^i(t)}(i)} - S_{\sigma_{\nu^i(t)-1}(i)},$$

and the left and right expression in (3.36) vanish as in (3.34) and (3.35). This completes the proof of the convergence in both cases.

To extend this result to moment convergence and an expansion of the variance in the case $\sigma_i^2 > 0$, we show that

$$\left\{ \left| \frac{\nu(t) - \frac{t}{\mu}}{\sqrt{t}} \right|^p, t \geq 1 \right\}$$

is uniformly integrable w.r.t. $\mathbb{P}_i$. For the first assertion in (3.31), we remark that

$$\mathbb{V}\mathrm{ar}_i\, \nu(t) = \mathbb{V}\mathrm{ar}_i \left(\nu(t) - \frac{t}{\mu} \right) = \mathbb{E}_i \left(\nu(t) - \frac{t}{\mu} \right)^2 - \left(\mathbb{E}_i \nu(t) - \frac{t}{\mu} \right)^2$$

$$= \mathbb{E}_i \left(\nu(t) - \frac{t}{\mu} \right)^2 + \mathcal{O}(1) = \frac{\sigma_i^2}{m_{ii}\mu^3} t + o(t),$$

where we used Lemma 3.24. Hence, it remains to verify the previously asserted uniform integrability. To this end, we extract from (3.32) that

$$|t - \nu(t)\mu| \leq |S_{\nu(t)} - \nu(t)\mu| + \mu + |S_{\nu(t)-1} - (\nu(t)-1)\mu| =: A_t + \mu + B_t,$$

and we show that $\{(A_t/\sqrt{t})^p, t \geq 1\}$ and $\{(B_t/\sqrt{t})^p, t \geq 1\}$ are uniformly integrable w.r.t. $\mathbb{P}_i$. We know that

$$A_t \leq |S_{\nu(t)} - S_{\sigma_{\nu^i(t)}(i)}| + |S_{\sigma_{\nu^i(t)}(i)} - \mu\sigma_{\nu^i(t)}(i)| + \mu(\sigma_{\nu^i(t)}(i) - \nu(t))$$

and apply the same procedure as before: The first expression satisfies

$$0 \leq |S_{\nu(t)} - S_{\sigma_{\nu^i(t)}(i)}| \leq S_{\nu(t)} - t + S_{\sigma_{\nu^i(t)}(i)} - t \leq 2(S_{\sigma_{\nu^i(t)}(i)} - t)$$

which we recognize (up to a multiplicative constant) as the overshoot of the SRP $(S_{\sigma_n(i)})_{n\geq 0}$. Since $\mathbb{E}_i(S^+_{\sigma_1(i)})^p = \mathbb{E}_i S^p_{\sigma_1(i)} < \infty$, [Gut09, Theorem 3.10.2] yields the uniform integrability of

$$\left\{ \frac{(S_{\sigma_{\nu^i(t)}(i)} - t)^p}{t}, t \geq 1 \right\} \quad \text{and} \quad \left\{ \frac{|S_{\nu(t)} - S_{\sigma_{\nu^i(t)}(i)}|^p}{t}, t \geq 1 \right\},$$

and particularly (recall $p \geq 2$) that of

$$\left\{ \left| \frac{S_{\nu(t)} - S_{\sigma_{\nu^i(t)}(i)}}{\sqrt{t}} \right|^p, t \geq 1 \right\}.$$

To deal with the middle term, we notice that the SRW $(S_{\sigma_n(i)} - \mu\sigma_n(i))_{n\geq 0}$ is centered and adapted to the filtration

$$\mathcal{F}_n := \sigma\big((\sigma_k(i), S_{\sigma_k(i)}) : 0 \leq k \leq n\big), \quad n \geq 0,$$

and that $\nu^i(t)$ is a stopping time w.r.t. this filtration. Furthermore, $\mathbb{E}_i|S_{\sigma_1(i)} - \mu\sigma_1(i)|^p < \infty$ and since $\{(\nu^i(t)/t)^{p/2}, t \geq 1\}$ is uniformly integrable w.r.t. $\mathbb{P}_i$ by [Gut09, Theorem 3.7.1], the uniform integrability of

$$\left\{ \left| \frac{S_{\sigma_{\nu^i(t)}(i)} - \mu\sigma_{\nu^i(t)}(i)}{\sqrt{t}} \right|^p, t \geq 1 \right\}$$

follows from [Gut09, Theorem 1.6.3]. The third and last expression satisfies

$$0 \leq \sigma_{\nu^i(t)}(i) - \nu(t) \leq \sigma_{\nu^i(t)}(i) - \sigma_{\nu^i(t)-1}(i)$$

which is the stopping summand of the i.i.d. sequence $(\sigma_n(i) - \sigma_{n-1}(i))_{n\geq 1}$ at the stopping time $\nu^i(t)$. Since $\mathbb{E}_i\sigma_1(i)^p < \infty$, we can apply [Gut09, Theorem 1.8.1] to conclude that

$$\left\{ \frac{(\sigma_{\nu^i(t)}(i) - \sigma_{\nu^i(t)-1}(i))^p}{t}, t \geq 1 \right\} \quad \text{and} \quad \left\{ \left(\frac{\sigma_{\nu^i(t)}(i) - \nu(t)}{\sqrt{t}} \right)^p, t \geq 1 \right\}$$

are uniformly integrable. This in particular, shows the uniform integrability of $\{(A_t/\sqrt{t})^p, t \geq 1\}$.

It remains to show that $\{(B_t/\sqrt{t})^p, t \geq 1\}$ is uniformly integrable. Note, that

$$\begin{aligned}|S_{\nu(t)-1} - (\nu(t)-1)\mu| &\leq |S_{\nu(t)-1} - S_{\nu(t)}| + |S_{\nu(t)} - \nu(t)\mu| + \mu \\ &= (S_{\nu(t)} - S_{\nu(t)-1}) + A_t + \mu,\end{aligned}$$

which is why we only care about the first summand which satisfies

$$0 \leq S_{\nu(t)} - S_{\nu(t)-1} \leq S_{\sigma_{\nu^i(t)}(i)} - S_{\sigma_{\nu^i(t)-1}(i)}.$$

This again is a stopping summand, now of $(S_{\sigma_n(i)} - S_{\sigma_{n-1}(i)})_{n\geq 1}$, at time $\nu^i(t)$, so by [Gut09, Theorem 1.8.1] the uniform integrability of

$$\left\{\frac{(S_{\sigma_{\nu^i(t)}(i)} - S_{\sigma_{\nu^i(t)-1}(i)})^p}{t}, t \geq 1\right\} \quad \text{and} \quad \left\{\left(\frac{S_{\nu(t)} - S_{\nu(t)-1}}{\sqrt{t}}\right)^p, t \geq 1\right\}$$

follows from $\mathbb{E}_i S_{\sigma_1(i)}^p < \infty$.

Concerning the case $\sigma_i^2 = 0$, we have shown a stronger uniform integrability statement in passing, namely that

$$\left\{\frac{|\nu(t) - \frac{t}{\mu}|^p}{t}, t \geq 1\right\}$$

is uniformly integrable under $\mathbb{P}_i$. Together with the already proven convergence, this completes the proof. □

Lemma 3.33. *Given the situation of Lemma 3.29, let $(X_0^{(t)})_t$ be a tight family of real-valued random variables independent of $(\xi_n, S_n)_{n\geq 0}$. If $\sigma^{(2)} > 0$, then*

$$\frac{\nu(X_0^{(t)}, t) - \frac{t}{\mu}}{\sqrt{t}} \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, \frac{\sigma^{(2)}}{\mu^3}\right), \tag{3.37}$$

as $t \to \infty$, w.r.t. $\mathbb{P}_i$. If $\sigma^{(2)} = 0$, then the limit is 0 and the convergence holds in probability w.r.t. $\mathbb{P}_i$. Assuming further the stronger moment conditions from Lemma 3.29 and that $(X_0^{(t)})_t$ is L^p-bounded for the same $p \geq 2$,

$$\mathbb{E}_i \left|\frac{\nu(X_0^{(t)}, t) - \frac{t}{\mu}}{\sqrt{t}}\right|^q \to \mathbb{E}_i |N|^q, \tag{3.38}$$

as $t \to \infty$, for $0 < q \leq p$ and N from Lemma 3.29 in the case $\sigma^{(2)} > 0$. The convergence also holds without absolute value (along sequences) for $q \in \mathbb{N}$. If $\sigma^{(2)} = 0$, then

$$\frac{\nu(X_0^{(t)}, t) - \frac{t}{\mu}}{t^{1/p}} \to 0 \quad \mathbb{P}_i\text{-a.s.}$$

and

$$\mathbb{E}_i \left|\nu(X_0^{(t)}, t) - \frac{t}{\mu}\right|^p = o(t). \tag{3.39}$$

In both cases ($\sigma^{(2)} > 0, = 0$), as $t \to \infty$,

$$\mathbb{V}\mathrm{ar}_i\, \nu(X_0^{(t)}, t) = \frac{\sigma^{(2)}}{\mu^3} t + o(t) \quad \text{and} \quad \mathbb{E}_i \left(\nu(X_0^{(t)}, t) - \frac{t}{\mu}\right)^2 = \frac{\sigma^{(2)}}{\mu^3} t + o(t).$$

Proof. Let $f : \mathbb{R} \to \mathbb{R}$ be a bounded and continuous function. We must show that

$$\mathbb{E}_i f\left(\frac{\nu(t - X_0^{(t)})\mathbb{1}_{\{t-X_0^{(t)}\geq 0\}} - \frac{t}{\mu}}{\sqrt{t}}\right) \to \mathbb{E}_i f(N),$$

as $t \to \infty$, for a random variable N following the distribution of the desired limit. W.l.o.g. (cf. the proof of Lemma 3.18) we assume that, as $t \to \infty$, both $X_0^{(t)}/\sqrt{t} \to 0$ (and thus $t - X_0^{(t)} \to \infty$) and $(\nu(t) - \frac{t}{\mu})/\sqrt{t} \to N$ $\mathbb{P}_i$-a.s. This is possible due to the tightness of $(X_0^{(t)})$, Lemma 3.29 and independence. We partition

$$\begin{aligned}
\frac{\nu(t - X_0^{(t)})\mathbb{1}_{\{t-X_0^{(t)}\geq 0\}} - \frac{t}{\mu}}{\sqrt{t}} &= \frac{\nu(t - X_0^{(t)})\mathbb{1}_{\{t-X_0^{(t)}\geq 0\}} - \frac{t-X_0^{(t)}}{\mu}}{\sqrt{t}} - \frac{X_0^{(t)}}{\mu\sqrt{t}} \\
&= \mathbb{1}_{\{t-X_0^{(t)}\geq 0\}} \frac{\nu(t - X_0^{(t)}) - \frac{t-X_0^{(t)}}{\mu}}{\sqrt{t}} - \frac{t - X_0^{(t)}}{\mu\sqrt{t}}\mathbb{1}_{\{t-X_0^{(t)}<0\}} - a_t \\
&= \mathbb{1}_{\{t-X_0^{(t)}>0\}} \frac{\nu(t - X_0^{(t)}) - \frac{t-X_0^{(t)}}{\mu}}{\sqrt{t - X_0^{(t)}}}\sqrt{\frac{t - X_0^{(t)}}{t}} + \frac{\nu(0)}{\sqrt{t}}\mathbb{1}_{\{t-X_0^{(t)}=0\}} - a_t - b_t \\
&= \mathbb{1}_{\{t-X_0^{(t)}>0\}} \frac{\nu(t - X_0^{(t)}) - \frac{t-X_0^{(t)}}{\mu}}{\sqrt{t - X_0^{(t)}}} \cdot d_t - a_t - b_t + c_t
\end{aligned}$$

where, as $t \to \infty$, $\mathbb{P}_i$-a.s.

$$a_t := \frac{X_0^{(t)}}{\mu\sqrt{t}} \to 0, \qquad b_t := \frac{t - X_0^{(t)}}{\mu\sqrt{t}}\mathbb{1}_{\{t-X_0^{(t)}<0\}} \to 0,$$

$$c_t := \frac{\nu(0)}{\sqrt{t}}\mathbb{1}_{\{t-X_0^{(t)}=0\}} \to 0 \qquad d_t := \sqrt{\frac{t - X_0^{(t)}}{t}} \to 1.$$

The remaining expression converges to N a.s., so we can conclude with the dominated convergence theorem (f is bounded and continuous) that

$$\mathbb{E}_i f\left(\frac{\nu(t - X_0^{(t)})\mathbb{1}_{\{t-X_0^{(t)}\geq 0\}} - \frac{t}{\mu}}{\sqrt{t}}\right) \to \mathbb{E}_i f(N).$$

The simple proof of the assertion when $\sigma^{(2)} = 0$ (both, with additional moment conditions and without) is similar to the proof of Lemma 3.18.

It remains to prove the moment convergence (3.38) and (3.39), which entails the two asymptotic expansions. For the variance, we note that

$$\begin{aligned}
\mathbb{V}\mathrm{ar}_i\, \nu(X_0^{(t)}, t) = \mathbb{V}\mathrm{ar}_i\left(\nu(X_0^{(t)}, t) - \frac{t}{\mu}\right) &= \mathbb{E}_i\left(\nu(X_0^{(t)}, t) - \frac{t}{\mu}\right)^2 - \left(\mathbb{E}_i\nu(X_0^{(t)}, t) - \frac{t}{\mu}\right)^2 \\
&= \mathbb{E}_i\left(\nu(X_0^{(t)}, t) - \frac{t}{\mu}\right)^2 + \mathcal{O}(1),
\end{aligned}$$

where we have used Lemma 3.26 ($\mathbb{E}_\pi X_1^2 < \infty$ and $(X_0^{(t)})_t$ is uniformly integrable). Furthermore, if we prove (3.38), then (3.39) follows along the same lines, and due to the already proven distributional convergence (3.37), we obtain uniform integrability of

$$\left\{ \left(\frac{\nu(X_0^{(t)}, t) - \frac{t}{\mu}}{\sqrt{t}} \right)^q, t \in \mathbb{T} \right\}$$

for a countable index set $\mathbb{T}$ unbounded to the right, and hence moment convergence without absolute value follows (at least along sequences).

To prove (3.38) we proceed as in the proof of Lemma 3.20. It is obvious from this lemma that $\mathbb{E}_i|(\nu(X_0^{(t)}, t) - \frac{t}{\mu})/\sqrt{t}|^p < \infty$ for $t \geq 1$. For some $0 < q \leq p$, let

$$g_i(t) := \mathbb{E}_i \left| \frac{\nu(t) - \frac{t}{\mu}}{\sqrt{t \vee 1}} \right|^q$$

which satisfies $g_i(t) \to \mathbb{E}_i|N|^q$, as $t \to \infty$, by Lemma 3.29. By the uniform integrability, $\sup_{t \geq 1} g_i(t) < \infty$, and by the monotonicity of ν,

$$\sup_{0 \leq t \leq 1} g_i(t) \leq \sup_{0 \leq t \leq 1} \mathbb{E}_i \left(\nu(t) + \frac{t}{\mu} \right)^q = \mathbb{E}_i \left(\nu(1) + \frac{1}{\mu} \right)^q < \infty.$$

Our main techniques are again conditioning on $X_0^{(t)}$ and using independence. Let $t \geq 1$ and

$$\begin{aligned} Y_t :=& \mathbb{E}_i \left[\left| \frac{\nu(X_0^{(t)}, t) - \frac{t - X_0^{(t)}}{\mu} \mathbb{1}_{\{t - X_0^{(t)} \geq 0\}}}{\sqrt{t}} \right|^q \,\middle|\, X_0^{(t)} \right] \\ =& \mathbb{E}_i \left[\left| \frac{\nu(t - X_0^{(t)}) - \frac{t - X_0^{(t)}}{\mu}}{\sqrt{(t - X_0^{(t)}) \vee 1}} \right|^q \,\middle|\, X_0^{(t)} \right] \left(\frac{(t - X_0^{(t)}) \vee 1}{t} \right)^{q/2} \mathbb{1}_{\{t - X_0^{(t)} \geq 0\}} \\ =& g_i(t - X_0^{(t)}) \left(\frac{(t - X_0^{(t)}) \vee 1}{t} \right)^{q/2} \mathbb{1}_{\{t - X_0^{(t)} \geq 0\}}. \end{aligned}$$

Then

$$\mathbb{E}_i \left| \frac{\nu(X_0^{(t)}, t) - \frac{t - X_0^{(t)}}{\mu} \mathbb{1}_{\{t - X_0^{(t)} \geq 0\}}}{\sqrt{t}} \right|^q = \mathbb{E}_i Y_t \to \mathbb{E}_i |N|^q$$

similar to the end of the proof of Lemma 3.20 with $q/2$ instead of q, and it suffices to verify

$$\left| \left|\left| \frac{\nu(X_0^{(t)}, t) - \frac{t - X_0^{(t)}}{\mu} \mathbb{1}_{\{t - X_0^{(t)} \geq 0\}}}{\sqrt{t}} \right|\right|_q - \left|\left| \frac{\nu(X_0^{(t)}, t) - \frac{t}{\mu}}{\sqrt{t}} \right|\right|_q \right| \to 0,$$

as $t \to \infty$. Here $||X||_q := \mathbb{E}|X|^q$ for $0 < q < 1$. We use the reverse triangle inequality that holds for all $0 < q < \infty$:

$$\left| \left|\left| \frac{\nu(X_0^{(t)}, t) - \frac{t - X_0^{(t)}}{\mu} \mathbb{1}_{\{t - X_0^{(t)} \geq 0\}}}{\sqrt{t}} \right|\right|_q - \left|\left| \frac{\nu(X_0^{(t)}, t) - \frac{t}{\mu}}{\sqrt{t}} \right|\right|_q \right|$$

$$\leq \left\| \frac{\frac{t-X_0^{(t)}}{\mu}\mathbb{1}_{\{t-X_0^{(t)}\geq 0\}} - \frac{t}{\mu}}{\sqrt{t}} \right\|_q = \left\| \frac{\frac{-X_0^{(t)}}{\mu}\mathbb{1}_{\{t-X_0^{(t)}\geq 0\}} - \frac{t}{\mu}\mathbb{1}_{\{t-X_0^{(t)}<0\}}}{\sqrt{t}} \right\|_q$$

$$\leq \frac{1}{(\mu\sqrt{t})^{q\wedge 1}} \left\| X_0^{(t)}\mathbb{1}_{\{t-X_0^{(t)}\geq 0\}} \right\|_q + \left(\frac{\sqrt{t}}{\mu}\right)^{q\wedge 1} \left\| \mathbb{1}_{\{t-X_0^{(t)}<0\}} \right\|_q .$$

If $q \geq 1$, then the latter equals

$$\frac{\sqrt{t}}{\mu}\mathbb{P}(X_0^{(t)} > t)^{1/q} \leq \frac{1}{\mu}\left(t^{p/2}\mathbb{P}(X_0^{(t)} > t)\right)^{1/q} \leq \frac{1}{\mu}\left(\mathbb{E}|X_0^{(t)}|^{p/2}\mathbb{1}_{\{X_0^{(t)}>t\}}\right)^{1/q}$$

since $t \geq 1$, and it vanishes, as $t \to \infty$, due to the uniform integrability of $\{|X_0^{(t)}|^{p/2}, t \geq 1\}$. The first term is

$$\frac{1}{\mu\sqrt{t}}\left(\mathbb{E}|X_0^{(t)}|^q\mathbb{1}_{\{t-X_0^{(t)}\geq 0\}}\right)^{1/q} \leq \frac{1}{\mu\sqrt{t}}\left(\sup_s \mathbb{E}|X_0^{(s)}|^q\right)^{1/q}$$

and vanishes, as $t \to \infty$, due to the L^q-boundedness of the family. The case $0 < q < 1$ is similar for both terms. This completes the proof. □

Proof of Theorem 3.6. $(\xi_n, S_n)_{n\geq 0}$ is a MRW with non-negative increments satisfying the assumptions of Lemma 3.33 since again the increments are bounded and $\mathbb{E}_i\sigma_1(i)^p < \infty$ for every $p \geq 2$ (and thus also $\mathbb{E}_i S^p_{\sigma_1(i)} < \infty$). Applying Lemma 2.12 and Lemma 3.33 yields the desired result. □

3.3.5. Weak law of large numbers for Δ_n

For the investigation of the imbalance factor Δ_n, we recall that $(\xi_n, S_n, V_n)_{n\geq 0}$ is the MRW corresponding to the MMS $(\xi_n, X_n, Y_n)_{n\geq 0}$, and that $\mu = \mathbb{E}_\pi X_1$ and $\mu_Y = \mathbb{E}_\pi Y_1$.

Lemma 3.34. *Let $(\xi_n, S_n, V_n)_{n\geq 0}$ be a zero-delayed MRW with positive recurrent discrete driving chain Ξ_0 which has stationary distribution π. Let $\mu < \infty$ be positive and let μ_Y exist. Then, for all $i \in \mathcal{S}$, as $t \to \infty$,*

$$\frac{V_{\nu(t)}}{t} \to \frac{\mu_Y}{\mu} \quad \mathbb{P}_i\text{-a.s.}$$

Proof. Since $\frac{V_{\nu(t)}}{t} = \frac{V_{\nu(t)}}{\nu(t)} \cdot \frac{\nu(t)}{t}\mathbb{1}_{\{\nu(t)>0\}}$, we can handle the factors separately. By Lemma 3.17, we have $\nu(t)/t \to 1/\mu$ $\mathbb{P}_i$-a.s. so we need to show that the left-hand factor converges to μ_Y $\mathbb{P}_i$-a.s. This is guaranteed by Theorem 3.16. □

Lemma 3.35. *Given the situation of Lemma 3.34, let further $(X_0^{(t)})_t$ be a tight family of real-valued random variables independent of $(\xi_n, X_n, Y_n)_{n\geq 0}$. Then, for all $i \in \mathcal{S}$, as $t \to \infty$,*

$$\frac{V_{\nu(X_0^{(t)},t)}}{t} \xrightarrow{\mathbb{P}_i} \frac{\mu_Y}{\mu}.$$

Proof. We apply the usual method with $\nu(X_0^{(t)}, t) = \nu(t - X_0^{(t)})\mathbb{1}_{\{t-X_0^{(t)}\geq 0\}}$ and use Lemma 3.34 in the same way as the corresponding result was used in the proof of Lemma 3.18. We omit the details. □

Proof of Theorem 3.9. By Lemma 3.35 we know that

$$\frac{V_{\nu(X_0^{(n)},\log n)}}{\log n} \xrightarrow{\mathbb{P}_i} \frac{\mu_Y}{\mu},$$

as $n \to \infty$. The integrability condition for Y_1 is trivial because it is bounded. Obviously, $(\xi_n, S_n, V_n)_{n\geq 0}$ forms a MRW. The limit is constant, so with Lemma 2.14 the same convergence holds for $\frac{\Delta_n}{\log n}$. □

3.3.6. L^p-law of large numbers for Δ_n

Lemma 3.36. *Let $(\xi_n, S_n, V_n)_{n\geq 0}$ be a zero-delayed MRW with a.s. non-negative increments in the second component, and with positive recurrent discrete driving chain Ξ_0 which has stationary distribution π. Let $\mu < \infty$ be positive and let $\mathbb{E}_i\big(\sum_{k=1}^{\sigma_1(i)} |Y_k|\big)^p < \infty$ for some $p \geq 1$ (and hence $\mu_Y < \infty$). Then $\{|V_{\nu(t)}/t|^p, t \geq 1\}$ is uniformly integrable w.r.t. $\mathbb{P}_i$ and thus*

$$\frac{V_{\nu(t)}}{t} \xrightarrow{L^p} \frac{\mu_Y}{\mu}$$

under $\mathbb{P}_i$, as $t \to \infty$, in particular,

$$\mathbb{E}_i\left|\frac{V_{\nu(t)}}{t}\right|^q \to \left|\frac{\mu_Y}{\mu}\right|^q$$

for $0 < q \leq p$. The convergence also holds without absolute value for $q \in \mathbb{N}$.

Proof. We only need to show the first assertion, then the rest follows by Lemma 3.34. We set $Z_n := \sum_{k=\sigma_{n-1}(i)+1}^{\sigma_n(i)} |Y_k|$ for $n \geq 1$, $Z_0 := 0$, and consider the filtration

$$\mathcal{F}_n = \sigma\big((\sigma_k(i), S_{\sigma_k(i)}), Z_k : 0 \leq k \leq n\big), \quad n \geq 0.$$

Note that $\nu^i(t)$ is a stopping time w.r.t. this filtration. We proceed with the estimate

$$0 \leq \left(\frac{|V_{\nu(t)}|}{t}\right)^p \leq \left(\frac{\sum_{k=1}^{\nu(t)} |Y_k|}{t}\right)^p \leq \left(\frac{\sum_{k=1}^{\sigma_{\nu^i(t)}(i)} |Y_k|}{t}\right)^p = \left(\frac{\sum_{k=1}^{\nu^i(t)} Z_k}{t}\right)^p$$

and since $\mathbb{E}_i Z_1^p < \infty$ and $\{(\nu^i(t)/t)^p, t \geq 1\}$ is uniformly integrable (still trivial by non-negativity), [Gut09, Theorem 1.6.1] guarantees that

$$\left\{\left(\frac{\sum_{k=1}^{\nu^i(t)} Z_k}{t}\right)^p, t \geq 1\right\}$$

is uniformly integrable. This completes the proof. □

Lemma 3.37. *Given the situation of Lemma 3.36, let $(X_0^{(t)})_t$ be a tight and L^p-bounded family of real-valued random variables independent of $(\xi_n, S_n, V_n)_{n\geq 0}$. Then (along sequences)*

$$\frac{V_{\nu(X_0^{(t)},t)}}{t} \xrightarrow{L^p} \frac{\mu_Y}{\mu}$$

under $\mathbb{P}_i$*, as* $t \to \infty$*, in particular,*

$$\mathbb{E}_i\left(\frac{\left|V_{\nu(X_0^{(t)},t)}\right|}{t}\right)^q \to \left|\frac{\mu_Y}{\mu}\right|^q$$

for $0 < q \le p$*. The convergence also holds without absolute value for* $q \in \mathbb{N}$*.*

Proof. Similar to the proof of Lemma 3.20, we get $\mathbb{E}_i|V_{\nu(X_0^{(t)},t)}|^p < \infty$ for all $t \ge 0$. We proceed in the same way and define

$$g_i(t) := \mathbb{E}_i\left(\frac{|V_{\nu(t)}|}{t \vee 1}\right)^q$$

for some $0 < q \le p$. Then by Lemma 3.36 we have $g_i(t) \to |\mu_Y/\mu|^q$, as $t \to \infty$. Note that $g_i(t)$ is bounded since, on the one hand, $\sup_{t\ge1} g_i(t) < \infty$ by uniform integrability and, on the other hand,

$$\begin{aligned}\sup_{0\le t\le 1} g_i(t) = \sup_{0\le t\le 1} \mathbb{E}_i|V_{\nu(t)}|^q \le \sup_{0\le t\le 1} \mathbb{E}_i\left(\sum_{k=1}^{\nu(t)}|Y_k|\right)^q \le \mathbb{E}_i\left(\sum_{k=1}^{\nu(1)}|Y_k|\right)^q \\ \le \mathbb{E}_i\left(\sum_{k=1}^{\sigma_{\nu^i(1)}(i)}|Y_k|\right)^q = \mathbb{E}_i\left(\sum_{k=1}^{\nu^i(1)} Z_k\right)^q \le B'_q \cdot \mathbb{E}_i Z_1^q \cdot \mathbb{E}_i \nu^i(1)^q < \infty\end{aligned}$$

by [Gut09, Theorem 1.5.2]. Here B'_q is a constant only depending on q and the Z_k are defined in the proof of the previous lemma. If we now define

$$Y_t = \mathbb{E}_i\left[\left(\frac{\left|V_{\nu(X_0^{(t)},t)}\right|}{t}\right)^q \middle| X_0^{(t)}\right] = g_i(t - X_0^{(t)})\left(\frac{(t - X_0^{(t)}) \vee 1}{t}\right)^q \mathbb{1}_{\{t - X_0^{(t)} \ge 0\}}$$

as in Lemma 3.20, then the rest follows exactly as in this lemma together with our preparations from above. □

Proof of Theorem 3.10. Applying Lemma 2.14, we see that for all n

$$\mathbb{E}_i\left(\frac{|\Delta_n|}{\log n}\right)^p = \mathbb{E}_i\left(\frac{\left|V_{\nu(X_0^{(n)},\log n)}\right|}{\log n}\right)^p.$$

Lemma 3.37 completes the proof using the tightness and L^p-boundedness of $(X_0^{(n)})_n$ as well as the fact that $\mathbb{E}_i\big(\sum_{k=1}^{\sigma_1(i)}|Y_k|\big)^p = \mathbb{E}_i\sigma_1(i)^p < \infty$ for every $p \ge 1$. □

3.3.7. Central limit theorem for Δ_n

The CLT for Δ_n is derived very similarly to the CLT for D_n. Again we can relate an auxiliary MRW to the situation of Meyn and Tweedie by setting $S_n(g) := \sum_{k=1}^n g(\xi_k, X_k, Y_k) = \mu V_n - \mu_Y S_n$ with $g = \mu p_3 - \mu_Y p_2$. Then [MT93, Theorem 17.2.2] is a CLT comprising our situation and the limiting variance is $\gamma^{(2)}$ (from Lemma 3.38), so $\gamma^{(2)}$ does not depend on i by the same reasoning as in Remark 3.30.

Lemma 3.38. *Let $(\xi_n, S_n, V_n)_{n\geq 0}$ be a zero-delayed MRW with a.s. non-negative increments in the second component, and with positive recurrent discrete driving chain Ξ_0 which has stationary distribution π. Let $\mu < \infty$ be positive. Assume further that $\mathbb{E}_\pi X_1^2 < \infty$, $\mathbb{E}_i\big(\sum_{k=1}^{\sigma_1(i)} |Y_k|\big)^2 < \infty$ (and hence $\mu_Y < \infty$) and*

$$\gamma^{(2)} := \frac{1}{m_{ii}}\gamma_i^2 := \frac{1}{m_{ii}}\,\mathbb{V}\mathrm{ar}_i(\mu V_{\sigma_1(i)} - \mu_Y S_{\sigma_1(i)}) > 0.$$

Then $\gamma^{(2)} < \infty$ does not depend on i and, as $t \to \infty$,

$$\frac{V_{\nu(t)} - \frac{\mu_Y}{\mu}t}{\sqrt{t}} \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, \frac{\gamma^{(2)}}{\mu^3}\right) \tag{3.40}$$

w.r.t. $\mathbb{P}_i$. Assuming further that $\mathbb{E}_i S_{\sigma_1(i)}^p < \infty$ and $\mathbb{E}_i\big(\sum_{k=1}^{\sigma_1(i)} |Y_k|\big)^p < \infty$ for some $p \geq 2$, the family $\{|(V_{\nu(t)} - \frac{\mu_Y}{\mu}t)/\sqrt{t}|^p, t \geq 1\}$ is uniformly integrable w.r.t. $\mathbb{P}_i$ and

$$\mathbb{E}_i\left|\frac{V_{\nu(t)} - \frac{\mu_Y}{\mu}t}{\sqrt{t}}\right|^q \to \mathbb{E}_i|N|^q$$

for $0 < q \leq p$. N has the distribution of the limit in (3.40). The convergence also holds without absolute value for $q \in \mathbb{N}$. Furthermore, as $t \to \infty$,

$$\mathbb{E}_i V_{\nu(t)} = \frac{\mu_Y}{\mu}t + o(t^{1/p}) \quad \textit{and} \quad \mathbb{V}\mathrm{ar}_i\, V_{\nu(t)} = \frac{\gamma^{(2)}}{\mu^3}t + o(t).$$

Remark 3.39. Remark 3.32 also applies to this modified setting in the sense that, in general, $\gamma^{(2)}$ is not equal to the stationary variance of $\mu Y_1 - \mu_Y X_1$.

Remark 3.40. There is of course a version of this lemma (and the following Lemma 3.41) in the case $\gamma^{(2)} = 0$, which is comparable to the corresponding case in Lemma 3.29. However, in the special situation of Theorem 3.11, $\gamma^{(2)} = 0$ never occurs, cf. Proposition 2.18. Thus, we refrain from a discussion of this case.

Proof. We are guided by the proof of Lemma 3.29 and [Gut09, Theorem 4.2.3], and thus we start by decomposing

$$\frac{V_{\nu(t)} - \frac{\mu_Y}{\mu}t}{\sqrt{t}} = \frac{V_{\nu(t)} - V_{\sigma_{\nu^i(t)}(i)}}{\sqrt{t}} + \frac{V_{\sigma_{\nu^i(t)}(i)} - \frac{\mu_Y}{\mu}S_{\sigma_{\nu^i(t)}(i)}}{\sqrt{t}} + \frac{\mu_Y}{\mu}\frac{S_{\sigma_{\nu^i(t)}(i)} - t}{\sqrt{t}}. \tag{3.41}$$

Again, we show that the middle term has the desired limit and that the remaining expressions vanish. Considering the middle term first, we notice that $(V_{\sigma_n(i)} - \frac{\mu_Y}{\mu}S_{\sigma_n(i)})_{n\geq 0}$ is a SRW w.r.t. $\mathbb{P}_i$. It is centered since $\mathbb{E}_i\big(V_{\sigma_1(i)} - \frac{\mu_Y}{\mu}S_{\sigma_1(i)}\big) = m_{ii}\mu_Y - \frac{\mu_Y}{\mu}m_{ii}\mu = 0$, so following the standard CLT we find that, as $n \to \infty$,

$$\frac{V_{\sigma_n(i)} - \frac{\mu_Y}{\mu}S_{\sigma_n(i)}}{\sqrt{n}} \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, \mathbb{V}\mathrm{ar}_i(V_{\sigma_1(i)} - \frac{\mu_Y}{\mu}S_{\sigma_1(i)})\right) = \mathrm{N}\left(0, \frac{\gamma_i^2}{\mu^2}\right)$$

whenever $\gamma^{(2)} > 0$. Using the i.i.d. version of Anscombe's theorem [Gut09, Theorem 1.3.1], we easily conclude that

$$\frac{V_{\sigma_{\nu^i(t)}(i)} - \frac{\mu_Y}{\mu}S_{\sigma_{\nu^i(t)}(i)}}{\sqrt{t}} \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, \frac{\gamma^{(2)}}{\mu^3}\right).$$

Concerning the last term in (3.41), we remark that the numerator is the overshoot $S_{\sigma_{\nu^i(t)}(i)} - t$ of the SRP $S_{\sigma_n(i)}$. Since $\mathbb{E}_i(S^+_{\sigma_1(i)})^p = \mathbb{E}_i S^p_{\sigma_1(i)} < \infty$ for some $p \geq 2$, [Gut09, Theorem 3.10.2] yields that

$$\frac{S_{\sigma_{\nu^i(t)}(i)} - t}{t^{1/p}} \to 0 \quad \mathbb{P}_i\text{-a.s.}$$

To deal with the first term in (3.41), we establish the upper bound

$$|V_{\nu(t)} - V_{\sigma_{\nu^i(t)}(i)}| = \left| \sum_{k=\nu(t)+1}^{\sigma_{\nu^i(t)}(i)} Y_k \right| \leq \sum_{k=\nu(t)+1}^{\sigma_{\nu^i(t)}(i)} |Y_k| \leq \sum_{k=\sigma_{\nu^i(t)-1}(i)+1}^{\sigma_{\nu^i(t)}(i)} |Y_k|$$

and as $(\sum_{k=1}^{\sigma_n(i)} |Y_k|)_{n\geq 0}$ is a SRW with L^p-increments under $\mathbb{P}_i$, we infer by Lemma A.5 that, as $n \to \infty$ or $t \to \infty$,

$$\frac{\sum_{k=\sigma_{n-1}(i)+1}^{\sigma_n(i)} |Y_k|}{n^{1/p}} \to 0 \quad \text{and} \quad \frac{\sum_{k=\sigma_{\nu^i(t)-1}(i)+1}^{\sigma_{\nu^i(t)}(i)} |Y_k|}{t^{1/p}} \to 0 \quad \mathbb{P}_i\text{-a.s.}$$

This completes the proof of the convergence. It remains to prove the uniform integrability. We have seen that

$$\begin{aligned} |V_{\nu(t)} - \frac{\mu_Y}{\mu} t| &\leq |V_{\nu(t)} - V_{\sigma_{\nu^i(t)}(i)}| + |V_{\sigma_{\nu^i(t)}(i)} - \frac{\mu_Y}{\mu} S_{\sigma_{\nu^i(t)}(i)}| + \frac{|\mu_Y|}{\mu}(S_{\sigma_{\nu^i(t)}(i)} - t) \\ &=: A_t + B_t + \frac{\mu_Y}{\mu} C_t \end{aligned}$$

and again it suffices to show that $\{(A_t/\sqrt{t})^p, t \geq 1\}$, $\{(B_t/\sqrt{t})^p, t \geq 1\}$ and $\{(C_t/\sqrt{t})^p, t \geq 1\}$ are uniformly integrable. For the first term, we use the Z_n, $n \geq 1$, from the proof of Lemma 3.36. They are non-negative i.i.d. random variables w.r.t. $\mathbb{P}_i$ and $\mathbb{E}_i Z_1^p < \infty$ by assumption. We know that $\nu^i(t)$ is a stopping time w.r.t. the filtration from the proof of Lemma 3.36 and

$$A_t \leq \sum_{k=\sigma_{\nu^i(t)-1}(i)+1}^{\sigma_{\nu^i(t)}(i)} |Y_k| = Z_{\nu^i(t)}$$

is bounded by a stopping summand. Therefore, the uniform integrability of $\{\nu^i(t)/t, t \geq 1\}$ together with [Gut09, Theorem 1.8.1] yields the uniform integrability of

$$\left\{ \frac{Z^p_{\nu^i(t)}}{t}, t \geq 1 \right\}, \left\{ \frac{A^p_t}{t}, t \geq 1 \right\} \quad \text{and} \quad \left\{ \left(\frac{A_t}{\sqrt{t}} \right)^p, t \geq 1 \right\}.$$

For B_t, we remark again that we deal with a centered SRW which lets us apply [Gut09, Theorem 1.6.3]. All required conditions are satisfied which was shown before or is very easy to show. Hence, the theorem yields the uniform integrability of

$$\left\{ \left(\frac{|V_{\sigma_{\nu^i(t)}(i)} - \frac{\mu_Y}{\mu} S_{\sigma_{\nu^i(t)}(i)}|}{\sqrt{t}} \right)^p, t \geq 1 \right\}.$$

To complete the proof, we remark that C_t is the overshoot of $(S_{\sigma_n(i)})_{n\geq 0}$ and, as all required conditions are satisfied, [Gut09, Theorem 3.10.2] yields that

$$\left\{\frac{(S_{\sigma_{\nu^i(t)}(i)}-t)^p}{t}, t\geq 1\right\} \quad \text{and} \quad \left\{\left(\frac{S_{\sigma_{\nu^i(t)}(i)}-t}{\sqrt{t}}\right)^p, t\geq 1\right\}$$

are uniformly integrable.

It is now easy to obtain asymptotic expansions from the moment convergence, but we only obtain $\mathbb{E}_i V_{\nu(t)} - \frac{\mu_Y}{\mu}t = o(\sqrt{t})$ directly. Nevertheless, the i.i.d. case theorem [Gut09, Theorem 4.2.4] enables us to improve the error by direct computation: The decomposition

$$\mathbb{E}_i V_{\nu(t)} = \mathbb{E}_i(V_{\nu(t)} - V_{\sigma_{\nu^i(t)}(i)}) + \mathbb{E}_i V_{\sigma_{\nu^i(t)}(i)} \tag{3.42}$$

lets us apply [Gut09, Theorem 4.2.4] to the second term which then is, as $t\to\infty$,

$$\mathbb{E}_i V_{\sigma_{\nu^i(t)}(i)} = \frac{\mathbb{E}_i V_{\sigma_1(i)}}{\mathbb{E}_i S_{\sigma_1(i)}}t + \mathcal{O}(1) = \frac{m_{ii}\mu_Y}{m_{ii}\mu}t + \mathcal{O}(1) = \frac{\mu_Y}{\mu}t + \mathcal{O}(1)$$

if $\mu_Y \neq 0$, and even equal to 0 otherwise. The first term in (3.42) can again be bounded by a stopping summand via

$$|\mathbb{E}_i(V_{\nu(t)} - V_{\sigma_{\nu^i(t)}(i)})| \leq \mathbb{E}_i\left(\sum_{k=\nu(t)+1}^{\sigma_{\nu^i(t)}(i)} |Y_k|\right) \leq \mathbb{E}_i\left(\sum_{k=\sigma_{\nu^i(t)-1}(i)+1}^{\sigma_{\nu^i(t)}(i)} |Y_k|\right) = o(t^{1/p})$$

by [Gut09, Theorem 1.8.1]. Concerning the variance, uniform integrability yields $\mathbb{E}_i\big(V_{\nu(t)} - \frac{\mu_Y}{\mu}t\big)^2 = \frac{\gamma^{(2)}}{\mu^3}t + o(t)$, so

$$\begin{aligned}\mathbb{V}\mathrm{ar}_i\, V_{\nu(t)} &= \mathbb{V}\mathrm{ar}_i\Big(V_{\nu(t)} - \frac{\mu_Y}{\mu}t\Big) = \mathbb{E}_i\Big(V_{\nu(t)} - \frac{\mu_Y}{\mu}t\Big)^2 - \Big(\mathbb{E}_i V_{\nu(t)} - \frac{\mu_Y}{\mu}t\Big)^2 \\ &= \frac{\gamma^{(2)}}{m_{ii}\mu^3}t + o(t) - o(t^{2/p}) = \frac{\gamma^{(2)}}{m_{ii}\mu^3}t + o(t).\end{aligned}$$

□

Lemma 3.41. *Given the situation of Lemma 3.38, let $(X_0^{(t)})_t$ be a tight family of real-valued random variables independent of $(\xi_n, S_n, V_n)_{n\geq 0}$. If $\gamma^{(2)} > 0$, then*

$$\frac{V_{\nu(X_0^{(t)},t)} - \frac{\mu_Y}{\mu}t}{\sqrt{t}} \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, \frac{\gamma^{(2)}}{\mu^3}\right),$$

as $t\to\infty$, w.r.t. $\mathbb{P}_i$. Assuming further the stronger moment conditions from Lemma 3.38 and that $(X_0^{(t)})_t$ is L^p-bounded for the same $p\geq 2$,

$$\mathbb{E}_i\left|\frac{V_{\nu(X_0^{(t)},t)} - \frac{\mu_Y}{\mu}t}{\sqrt{t}}\right|^q \to \mathbb{E}_i|N|^q,$$

as $t\to\infty$, for $0<q\leq p$ and N from Lemma 3.38 Furthermore, as $t\to\infty$,

$$\mathbb{E}_i V_{\nu(X_0^{(t)},t)} = \frac{\mu_Y}{\mu}t + o(t^{1/p}) \quad \textit{and} \quad \mathbb{V}\mathrm{ar}_i\, V_{\nu(X_0^{(t)},t)} = \frac{\gamma^{(2)}}{\mu^3}t + o(t).$$

Proof. We only remark that

$$V_{\nu(t-X_0^{(t)})\mathbb{1}_{\{t-X_0^{(t)}\geq 0\}}} = \mathbb{1}_{\{t-X_0^{(t)}\geq 0\}} V_{\nu(t-X_0^{(t)})}.$$

Then the proof is the same as for Lemma 3.33, except for the expansion of the expectation which can be proved as Lemma 3.26. □

Proof of Theorem 3.11. It is obvious that $(\xi_n, S_n, V_n)_{n\geq 0}$ forms a MRW, and that the integrability conditions for Lemma 3.41 are met, has already been argued in the proof of Theorem 3.10. Also, $\gamma^{(2)} > 0$ by Proposition 2.18. The result then follows from Lemma 3.41 and Lemma 2.14. □

4. Average-case analysis of further characteristic parameters

The power of Markov renewal theory has been demonstrated in the first part of this work. Based on the treatment of so-called harmonic sums (from Mellin transform theory, cf. [FGD95] or [FS09]) in the i.i.d. setting in [Jan12a], the following chapter is designated to a generalization of Janson's techniques so as to provide a simple device for finding asymptotics of various characteristic parameters of tries. As already mentioned, the asymptotic expansions in the applications are mostly not new. Nevertheless, our device enables us to immediately obtain at least the leading term in the asymptotic expansions of trie parameters whenever they have *additive* form which is described in the paragraph below. We provide comparisons with existing results in Section 4.5.

One of the most accessible parameters to motivate the following theorems with, is the expected size of a trie. It is easy to characterize and shows our device in the plainest form. We consider $\mathrm{Trie}(\mathcal{M}_n)$ which has a deterministic number of n leaves and contains a random number W_n of internal nodes. W_n is called the *size* of $\mathrm{Trie}(\mathcal{M}_n)$. By Observation 2.2, we have

$$\begin{aligned} W_n &= \#\{\alpha \in \mathcal{S}^* : \alpha \text{ is an internal node of } \mathrm{Trie}(\mathcal{M}_n)\} \\ &= \#\left\{\alpha \in \mathcal{S}^* : \sum_{k=1}^{n} \mathbb{1}_{\{\Xi^{(k)} \succ \alpha\}} \geq 2\right\} = \sum_{\alpha \in \mathcal{S}^*} \mathbb{1}_{\left\{\sum_{k=1}^{n} \mathbb{1}_{\{\Xi^{(k)} \succ \alpha\}} \geq 2\right\}}. \end{aligned}$$

We define $\widetilde{W}_\lambda := W_{\Pi(\lambda)}$ for a $\mathrm{Poi}(\lambda)$ distributed variable $\Pi(\lambda)$ and define $N_n(\alpha) := \sum_{k=1}^{n} \mathbb{1}_{\{\Xi^{(k)} \succ \alpha\}}$ as in [Jan12b], which counts the number of strings starting with α. The properties of a Poisson point process yield $\widetilde{N}_\lambda(\alpha) = N_{\Pi(\lambda)}(\alpha) \sim \Pi(\lambda P_i(\alpha))$ under $\mathbb{P}_i$ (cf. Section 4.2), and we obtain

$$\begin{aligned} \mathbb{E}_i \widetilde{W}_\lambda &= \sum_{\alpha \in \mathcal{S}^*} \mathbb{P}_i\left(\widetilde{N}_\lambda(\alpha) \geq 2\right) \mathbb{1}_{\{P_i(\alpha) > 0\}} = \sum_{\alpha \in \mathcal{S}^*} \mathbb{P}\left(\Pi(\lambda P_i(\alpha)) \geq 2\right) \mathbb{1}_{\{P_i(\alpha) > 0\}} \\ &= \sum_{\alpha \in \mathcal{S}^*} f(\lambda P_i(\alpha)) \mathbb{1}_{\{P_i(\alpha) > 0\}} \end{aligned} \tag{4.1}$$

for $f : (0, \infty) \to \mathbb{R}_{\geq 0}$ defined by $f(x) := \mathbb{P}(\Pi(x) \geq 2) = 1 - (1 + x)e^{-x}$. This is the same f as in [Jan12a, Section 5]. Hence, we need an analogue of [Jan12a, Theorem 5.1] to deal with sums of the form (4.1).

Additive parameters and search costs. Following [CFV01] or [Bou01], the above expression and the ones following in the applications are so-called *additive parameters* that can be expressed by summing over all nodes in the complete tree with weights called *search cost*. The latter indicates whether the considered node carries a certain property (e.g. having at least two children in the trie) or not.

Harmonic sums. As explained in [FGD95] (we follow their notation and sketch here), asymptotic expansions of *harmonic sums*

$$G(x) = \sum_k \lambda_k g(\mu_k x)$$

are usually found by means of Mellin transforms: The Mellin transform of G factorizes into a generalized Dirichlet series, containing λ_k and μ_k, and the Mellin transform of g. Then, roughly speaking, the Mellin inversion formula (which is similar to Fourier inversion) entails that we can find the asymptotics of G at 0 or ∞ by calculating residues. Of course, this is all subject to some regularity and growth conditions for the Dirichlet series and the Mellin transform of g. We will see that Mellin transforms appear in our statement as well (cf. Remark 4.4).

4.1. Main results

Several parameters of a trie can be expressed as a sum over all nodes α in $\mathcal{S}^*$ of some suitable function f at points $\lambda P_i(\alpha)$ or $\lambda P_i(\alpha j)$ as in (4.1). These expressions occur in the Poisson model. There we consider a random number of strings, from which we construct a trie to subsequently extract results for the standard model. We use the following theorems which reduce the problem of finding a limiting behavior of certain functionals to a calculation of integrals. Actually, the first theorem is a corollary of the second. With results of this kind at hand, the limiting behavior is clear up to $o(\lambda)$, as soon as the function f is known. These results provide a closed form for the periodic oscillatory error function in the d-arithmetic case as well as its Fourier series, which in all of our cases equals the function itself (cf. Remark 4.3). The theorem below applies to (4.1), hence we immediately obtain asymptotics for the expected size of a trie in the Poisson model.

Since for some parameters a similar approach as in (4.1) requires knowledge about the last letter in a string $\beta = \alpha j$, say, or about the letter following the considered string α, we prove Theorem 4.5 as a generalization of Theorem 4.1 first. In our applications we sometimes encounter a function f that does not satisfy (4.2). Therefore, we also need the generalizations Theorem 4.6 and Theorem 4.7. Theorems 4.1 and 4.7 are generalizations of [Jan12a, Theorem 5.1] and [Jan12b, Theorem 2.1], so we try to stay close to his notation.

In the formulation of the following theorems, we use the notation $f(x) = \mathcal{O}(g(x))$ for $a < x < b$, with positive g, meaning that there exists a positive constant $C > 0$ such that $|f(x)| \leq Cg(x)$ for all $a < x < b$. Also, we use $\sim$ here for the equivalence relation "$\lambda\!\!\lambda_0$-almost everywhere (a.e.) equal".

Theorem 4.1. *Let $f : (0, \infty) \to \mathbb{R}$ be a non-negative, $\lambda\!\!\lambda_0$-a.e. continuous function satisfying*

$$f(x) = \mathcal{O}(x^{1+\delta}) \text{ for } 0 < x < 1 \text{ and } f(x) = \mathcal{O}(x^{1-\delta}) \text{ for } 1 < x < \infty \tag{4.2}$$

for some $\delta > 0$, and $g(t) := e^t f(e^{-t})$. Then

$$F_i(\lambda) := \sum_{\alpha \in \mathcal{S}^*} f(\lambda P_i(\alpha)) \mathbb{1}_{\{P_i(\alpha) > 0\}} \tag{4.3}$$

converges for every $0 < \lambda < \infty$ and the following cases occur:

(a) If $(\xi_n, S_n)_{n \geq 0}$ is non-arithmetic, then, as $\lambda \to \infty$,

$$\frac{F_i(\lambda)}{\lambda} = \frac{1}{\mu} \int_{-\infty}^{\infty} g(t) \, \mathrm{d}t + o(1) = \frac{1}{\mu} \int_0^{\infty} f(x) x^{-2} \, \mathrm{d}x + o(1). \tag{4.4}$$

(b) If $(\xi_n, S_n)_{n \geq 0}$ is d-arithmetic with shift function β, then, as $\lambda \to \infty$,

$$\frac{F_i(\lambda)}{\lambda} = \frac{1}{\mu} \int_0^{\infty} f(x) x^{-2} \, \mathrm{d}x + \frac{1}{\mu} \sum_{j \in \mathcal{S}} \pi_j \psi(\log \lambda - \beta_{ij}) + o(1), \tag{4.5}$$

where $\psi(t)$ is a bounded d-periodic function given by

$$\psi(t) := d \sum_{n=-\infty}^{\infty} g(nd - t) - \widehat{\psi}(0) \tag{4.6}$$

with Fourier series

$$\psi(t) \sim \sum_{m \neq 0} \widehat{\psi}(m) e^{2\pi \mathrm{i}\, mt/d} \tag{4.7}$$

and Fourier coefficients

$$\widehat{\psi}(m) = \hat{g}(-2\pi m/d) = \int_{-\infty}^{\infty} g(t) e^{2\pi \mathrm{i}\, mt/d} \, \mathrm{d}t = \int_0^{\infty} f(x) x^{-2-2\pi \mathrm{i}\, m/d} \, \mathrm{d}x, \tag{4.8}$$

$m \in \mathbb{Z}$. Here $\hat{g}$ denotes the Fourier transform of non-periodic functions. If f is continuous, then the same holds for ψ. β_{ij} is defined just like in Theorem 3.3.

Remark 4.2. In all our applications, the occurring f is continuous. Thus, Condition (4.2) simplifies to

$$f(x) = \mathcal{O}(x^{1+\delta}), \text{ as } x \to 0, \text{ and } f(x) = \mathcal{O}(x^{1-\delta}), \text{ as } x \to \infty.$$

An analogous remark applies to Condition (4.12).

Remark 4.3. We cite [Jan12b, Remark 2.2] since our ψ is his ψ_0: Usually (and in all of our applications) f is continuous and $\hat{g}(s) = \mathcal{O}(s^{-2})$, $|s| \to \infty$, so the Fourier series in (4.7) converges absolutely, and thus its sum is continuous. Since ψ is continuous, as f is, the Fourier series converges to $\psi(t)$ for every t and we may replace $\sim$ by $=$ in (4.7).

Remark 4.4. As Janson points out in the proof of [Jan12a, Theorem 5.1], $\hat{g}(s)$ from Theorem 4.1 is the Mellin transform of f at $-1 + \mathrm{i}\, s$.

The following theorem is a variation of the previous theorem and actually more general:

Theorem 4.5. *Let $f : (0, \infty) \to \mathbb{R}$ be a non-negative, $\lambda\!\!\lambda_0$-a.e. continuous function satisfying (4.2) for some $\delta > 0$, and $g(t) := e^t f(e^{-t})$. Then, with $j \in \mathcal{S}$,*

$$F_i^j(\lambda) := \sum_{\alpha \in \mathcal{S}^*} f(\lambda P_i(\alpha j)) \mathbb{1}_{\{P_i(\alpha j) > 0\}} \tag{4.9}$$

converges for every $0 < \lambda < \infty$ and the following cases occur:

(a) If $(\xi_n, S_n)_{n \geq 0}$ is non-arithmetic, then, as $\lambda \to \infty$,

$$\frac{F_i^j(\lambda)}{\lambda} = \frac{\pi_j}{\mu} \int_{-\infty}^{\infty} g(t) \, \mathrm{d}t + o(1) = \frac{\pi_j}{\mu} \int_0^{\infty} f(x) x^{-2} \, \mathrm{d}x + o(1). \tag{4.10}$$

(b) If $(\xi_n, S_n)_{n \geq 0}$ is d-arithmetic with shift function β, then, as $\lambda \to \infty$,

$$\frac{F_i^j(\lambda)}{\lambda} = \frac{\pi_j}{\mu} \int_0^{\infty} f(x) x^{-2} \, \mathrm{d}x + \frac{\pi_j}{\mu} \psi(\log \lambda - \beta_{ij}) + o(1), \tag{4.11}$$

where $\psi(t)$ is the bounded d-periodic function from Theorem 4.1.

We can also prove a generalization of Theorem 4.5 which in turn provides a generalization of Theorem 4.1 as a consequence. The more general results describe the asymptotic behaviour for a broader range of functions f.

Theorem 4.6. *Let $f : (0, \infty) \to \mathbb{R}$ be a $\lambda\!\!\lambda_0$-a.e. continuous function satisfying*

$$\begin{aligned} f(x) &= \mathcal{O}(x^{1+\delta}) \text{ for } 0 < x < 1 \text{ and} \\ f(x) &= ax + \mathcal{O}(x^{1-\delta}) \text{ for } 1 < x < \infty \end{aligned} \tag{4.12}$$

for some $\delta > 0$ and $a \in \mathbb{R}$, and $g(t) := e^t f(e^{-t})$. Then, with $j \in \mathcal{S}$,

$$F_i^j(\lambda) := \sum_{\alpha \in \mathcal{S}^*} f(\lambda P_i(\alpha j)) \mathbb{1}_{\{P_i(\alpha j) > 0\}} \tag{4.13}$$

converges absolutely for every $0 < \lambda < \infty$ and the following cases occur:

(a) If $(\xi_n, S_n)_{n \geq 0}$ is non-arithmetic, then, as $\lambda \to \infty$,

$$F_i^j(\lambda) = \frac{a\pi_j}{\mu} \lambda \log \lambda + \frac{b_j}{\mu} \lambda - a\lambda \mathbb{1}_{\{j\}}(i) + o(\lambda), \tag{4.14}$$

where

$$b_j := \frac{a}{2\mu} \pi_j^2 \mathbb{E}_j S_{\sigma_1(j)}^2 - a\pi_j \mathbb{E}_i S_{\sigma_1(j)} \mathbb{1}_{\{j \neq i\}} + \pi_j \int_0^\infty (f(x) - ax\mathbb{1}_{\{x \geq 1\}}) x^{-2} \, \mathrm{d}x. \tag{4.15}$$

(b) If $(\xi_n, S_n)_{n \geq 0}$ is d-arithmetic with shift function β, then, as $\lambda \to \infty$,

$$F_i^j(\lambda) = \frac{a\pi_j}{\mu} \lambda \log \lambda + \frac{b_j}{\mu} \lambda + \frac{\pi_j}{\mu} \psi_0(\log \lambda - \beta_{ij}) \lambda - a\lambda \mathbb{1}_{\{j\}}(i) + o(\lambda), \tag{4.16}$$

with b_j from above and $\psi_0(t)$ is a bounded d-periodic function with Fourier series

$$\psi_0(t) \sim \sum_{m \neq 0} \widehat{\psi}_0(m) e^{2\pi \mathrm{i} mt/d} = \sum_{m \neq 0} \hat{g}(-2\pi m/d) e^{2\pi \mathrm{i} mt/d} \tag{4.17}$$

and Fourier coefficients $\widehat{\psi}_0(m) = \hat{g}(-2\pi m/d)$, where

$$\hat{g}(u) := \lim_{\varepsilon \searrow 0} \int_{-\infty}^\infty e^{-\mathrm{i} ut + \varepsilon t} g(t) \, \mathrm{d}t = \lim_{\varepsilon \searrow 0} \int_0^\infty f(x) x^{-2-\varepsilon+\mathrm{i} u} \, \mathrm{d}x \tag{4.18}$$

and furthermore

$$\psi_0(t) - \psi_0(0) = d \sum_{n=-\infty}^\infty (g(nd - t) - g(nd)) - at. \tag{4.19}$$

If f is continuous, then the same holds for ψ_0.

Note that $g(t) \to a$, as $t \to -\infty$, and $g(t) = \mathcal{O}(e^{-\delta t})$, as $t \to \infty$.

This result directly implies the following:

Theorem 4.7. *Let $f:(0,\infty)\to\mathbb{R}$ be a $\lambda\!\!\lambda_0$-a.e. continuous function satisfying* (4.12) *for some $\delta>0$ and $a\in\mathbb{R}$, and $g(t):=e^t f(e^{-t})$. Then*

$$F_i(\lambda) := \sum_{\alpha\in\mathcal{S}^*} f(\lambda P_i(\alpha))\mathbb{1}_{\{P_i(\alpha)>0\}} \tag{4.20}$$

converges absolutely for every $0<\lambda<\infty$ and the following cases occur:

(a) If $(\xi_n,S_n)_{n\geq 0}$ is non-arithmetic, then, as $\lambda\to\infty$,

$$F_i(\lambda) = \frac{a}{\mu}\lambda\log\lambda + \frac{b}{\mu}\lambda + o(\lambda), \tag{4.21}$$

where

$$b := \sum_{j\in\mathcal{S}} b_j = \frac{a}{2\mu}\sum_{j\in\mathcal{S}}\pi_j^2\mathbb{E}_j S^2_{\sigma_1(j)} - a\sum_{j\neq i}\pi_j\mathbb{E}_i S_{\sigma_1(j)} + \int_0^\infty (f(x) - ax\mathbb{1}_{\{x\geq 1\}})x^{-2}\,\mathrm{d}x. \tag{4.22}$$

(b) If $(\xi_n,S_n)_{n\geq 0}$ is d-arithmetic with shift function β, then, as $\lambda\to\infty$,

$$F_i(\lambda) = \frac{a}{\mu}\lambda\log\lambda + \frac{b}{\mu}\lambda + \frac{1}{\mu}\sum_{j\in\mathcal{S}}\pi_j\psi_0(\log\lambda - \beta_{ij})\lambda + o(\lambda), \tag{4.23}$$

with b from above and $\psi_0(t)$ is the bounded d-periodic function from Theorem 4.6.

4.2. Poisson model

To apply these results to functionals of the trie, we consider the Poisson model in the following sense: Let $(\Pi(\lambda))_{\lambda>0}$ be a family of independent $\mathrm{Poi}(\lambda)$ distributed variables. Let further $(\Pi(\lambda))_{\lambda>0},\xi_0,\Xi^{(1)},\Xi^{(2)},\dots$ be independent w.r.t. $\mathbb{P}_i$ for each $i\in\mathcal{S}$. Again, as with $X_0^{(n)}$, we will sometimes use $\mathbb{P}$ instead of $\mathbb{P}_i$ if the probability does not depend on i. Usually, we denote the first string by $\Xi^{(1)}=\Xi$, but Ξ is considered an additional string in the sequence $\Xi,\Xi^{(1)},\Xi^{(2)},\dots$ in Subsection 4.5.4. The set from which the trie is constructed is then $\mathcal{M}_n=\{\Xi^{(1)},\dots,\Xi^{(n)}\}$ or $\{\Xi,\Xi^{(1)},\dots,\Xi^{(n-1)}\}$ for $n\in\mathbb{N}$. In the Poisson model, we deal with the first $\Pi(\lambda)$ (or $1+\Pi(\lambda)$) strings. The definitions of the functionals rely on characterizations of specific nodes in $\mathrm{Trie}(\mathcal{M}_n)$. We recall, that with $|\alpha|$ denoting the length of α, the event

$$\{\Xi^{(k)}\succ\alpha\} := \{(\xi_1^{(k)},\dots,\xi_{|\alpha|}^{(k)})=\alpha\}$$

is the set on which α is a prefix of $\Xi^{(k)}$ or equivalently $\Xi^{(k)}$ starts with α.

An important tool is the fact that for $l\in\mathbb{N}$ and under $\mathbb{P}_i$

$$\sum_{k=1}^{\Pi(\lambda)}\mathbb{1}_{\{(\xi_1^{(k)},\dots,\xi_l^{(k)})\in\cdot\}} = \sum_{k=1}^{\Pi(\lambda)}\delta_{(\xi_1^{(k)},\dots,\xi_l^{(k)})}$$

is a Poisson point process on $\mathcal{S}^l$. Its intensity measure is $\zeta(\cdot)=\lambda P_i(\cdot)$ since the $(\xi_1^{(k)},\dots,\xi_l^{(k)})$, $k=1,2,\dots$, are independent with distribution $P_i(\cdot)$. Hence, for α with length l, we have

$$\sum_{k=1}^{\Pi(\lambda)}\mathbb{1}_{\{\Xi^{(k)}\succ\alpha\}} = \sum_{k=1}^{\Pi(\lambda)}\delta_{(\xi_1^{(k)},\dots,\xi_l^{(k)})}(\{\alpha\}) \overset{\mathbb{P}_i}{\sim} \Pi(\zeta(\alpha)) = \Pi(\lambda P_i(\alpha)).$$

4.3. Markov renewal theorems

The four main results in Section 4.1 are based on the two so-called *Markov renewal theorems*. These are Markov analogues of the well-known renewal theorem by Blackwell and its equivalent version which is known under the name of *key renewal theorem*. We state both, the Markov renewal theorem I and the Markov renewal theorem II as well as a refined version of the latter in the d-arithmetic case. As usual $\infty^{-1} := 0$ and we denote by $\lambda\!\!\lambda_0$ and $\lambda\!\!\lambda_d$ the Lebesgue measure on $(\mathbb{R}, \mathcal{B})$ and d-times the counting measure on $d\mathbb{Z}$, respectively. For the formulation of the next theorem, we point out that $d\text{-}\lim_{t\to\infty} f(t)$ means the usual limit in the case $d = 0$, and the limit $\lim_{n\to\infty} f(nd)$ in the case $d > 0$.

The *Markov renewal theorem I* is an analogue of *Blackwell's renewal theorem*:

Theorem 4.8 (Markov renewal theorem I)**.** *Let $(\xi_n, S_n)_{n\geq 0}$ be a non-arithmetic or d-arithmetic zero-delayed MRW with positive recurrent and discrete driving chain Ξ_0 which has stationary distribution π. Let μ be positive and let the shift function be 0 if $d > 0$. Then for all $i \in \mathcal{S}, A \subset \mathcal{S}$ and all bounded intervals I*

$$\underset{t\to\infty}{d\text{-}\lim}\, \mathbb{U}_i(A \times (t + I)) = \frac{1}{\mu}\pi(A)\lambda\!\!\lambda_d(I)$$

and

$$\underset{t\to-\infty}{d\text{-}\lim}\, \mathbb{U}_i(A \times (t + I)) = 0.$$

Proof. Cf. for example [Als14, Theorem 3.1] for the non-arithmetic case. □

The central theorems of Section 4.1 rely on the *Markov renewal theorem II* which is an analogue of the *key renewal theorem*:

Theorem 4.9 (Markov renewal theorem II, non-arithmetic or shift 0)**.** *Let $(\xi_n, S_n)_{n\geq 0}$ be a non-arithmetic or d-arithmetic zero-delayed MRW with positive recurrent and discrete driving chain Ξ_0 which has stationary distribution π. Let μ be positive and let the shift function be 0 if $d > 0$. Let further $g : \mathcal{S} \times \mathbb{R} \to \mathbb{R}$ be a measurable function which satisfies*

$$g(i, \cdot) \text{ is } \lambda\!\!\lambda_0\text{-a.e. continuous for all } i \in \mathcal{S}$$

and

$$\sum_{j\in\mathcal{S}} \pi_j \sum_{n\in\mathbb{Z}} \sup_{n\varepsilon<x\leq(n+1)\varepsilon} |g(j, x)| < \infty \quad \text{for some } \varepsilon > 0 \tag{4.24}$$

in the non-arithmetic case, and

$$\sum_{j\in\mathcal{S}} \pi_j \sum_{n\in d\mathbb{Z}} |g(j, n)| < \infty \tag{4.25}$$

in the d-arithmetic case. Then, for all $i \in \mathcal{S}$,

$$\underset{t\to\infty}{d\text{-}\lim}\, \mathbf{U} * g(i, t) = \frac{1}{\mu} \sum_{j\in\mathcal{S}} \pi_j \int_{\mathbb{R}} g(j, x)\, \lambda\!\!\lambda_d(\mathrm{d}x)$$

and

$$\underset{t\to-\infty}{d\text{-}\lim}\, \mathbf{U} * g(i, t) = 0.$$

Proof. Cf. for example [Als14, Theorem 3.2] for the non-arithmetic case. □

Remark 4.10. If $\mathcal{S}$ is finite, then (4.24) is equivalent to the direct Riemann integrability of $g(i,\cdot)$ for all $i \in \mathcal{S}$, and (4.25) is equivalent to the absolute summability of $(g(i,n))_{n\in d\mathbb{Z}}$, for all $i \in \mathcal{S}$.

The following easy lemma helps to verify direct Riemann integrability in the proof of the main results. It is taken from [Jan12a, Lemma A.6].

Lemma 4.11. *Let $f : \mathbb{R} \to \mathbb{R}$ be a non-negative, bounded and a.e. continuous function. Let further F be an integrable function satisfying $0 \le f \le F$, and let $A \ge 0$ such that F is increasing on $(-\infty, -A)$ and decreasing on (A, ∞). Then f is direct Riemann integrable.*

As mentioned before, the requirement of a vanishing shift function in the d-arithmetic case is a technical nuisance. Usually we then consider the MRW $(\xi_n, \widetilde{S}_n)_{n\ge 0}$ as defined in Subsection 3.3.3. If $\mathcal{S}$ is finite and we also assume direct Riemann integrability of the $g(i,\cdot)$ (instead of just summability), then Theorem 4.9 implies a more precise description of the asymptotics in the d-arithmetic case.

Theorem 4.12 (Markov renewal theorem II, d-arithmetic)**.** *Let $\mathcal{S}$ be finite and let $(\xi_n, S_n)_{n\ge 0}$ be a d-arithmetic zero-delayed MRW with positive recurrent and discrete driving chain Ξ_0 which has stationary distribution π. Suppose that μ is positive and that $\beta : \mathcal{S} \to [0, d)$ is the shift function. Let further $g : \mathcal{S} \times \mathbb{R} \to \mathbb{R}$ be a measurable function which satisfies*

$$g(i,\cdot) \text{ is } \lambda\!\!\lambda_0\text{-a.e. continuous for all } i \in \mathcal{S} \tag{4.26}$$

and

$$g(i,\cdot) \text{ is direct Riemann integrable for all } i \in \mathcal{S}. \tag{4.27}$$

Then, for all $i \in \mathcal{S}$, as $t \to \infty$,

$$\mathbf{U} * g(i,t) = \frac{1}{\mu} \sum_{j\in\mathcal{S}} \pi_j \psi(j, -t + \beta_{ij}) + o(1),$$

where $\psi(j,\cdot)$ is a bounded and d-periodic function for every $j \in \mathcal{S}$ defined by

$$\psi(j,t) := d \sum_{n=-\infty}^{\infty} g(j, nd - t) \tag{4.28}$$

which has the Fourier series (as a function in t)

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

The Fourier coefficients satisfy

$$\widehat{\psi}(j,m) = \hat{g}(j, -2\pi m/d) = \int_{-\infty}^{\infty} g(j,t) e^{2\pi \mathrm{i}\, mt/d}\, \mathrm{d}t,$$

where $\widehat{\psi}$ are the ordinary Fourier coefficients of $\psi(j,\cdot)$. The series in (4.28) converges uniformly on $[0,d]$, so if $g(j,\cdot)$ is continuous, then so is $\psi(j,\cdot)$. Furthermore, for all $i \in \mathcal{S}$, as $t \to -\infty$,

$$\mathbf{U} * g(i,t) = o(1).$$

Proof. We need to show that, as $t \to \infty$,

$$\mathbf{U} * g(i,t) - \frac{1}{\mu} \sum_{j \in \mathcal{S}} \pi_j \psi(j, -t + \beta_{ij}) = o(1).$$

First of all, $\widetilde{S}_n \in d\mathbb{Z}$ $\mathbb{P}_i$-a.s., since $\widetilde{S}_n$ has shift 0, so for $\tilde{g}_{ij}(s,x) := \mathbb{1}_{\{j\}}(s) g(s, x - \beta_{ij})$

$$\begin{aligned}
\mathbf{U} * g(i,t) &= \mathbb{E}_i \left(\sum_{n \geq 0} g(\xi_n, t - S_n) \right) = \mathbb{E}_i \left(\sum_{n \geq 0} g(\xi_n, t - \beta(\xi_n) + \beta(i) - \widetilde{S}_n) \right) \\
&= \sum_{j \in \mathcal{S}} \mathbb{E}_i \left(\sum_{n \geq 0} \mathbb{1}_{\{j\}}(\xi_n) g(\xi_n, t - \beta_{ij} - \widetilde{S}_n) \right) \\
&= \sum_{j \in \mathcal{S}} \mathbb{E}_i \left(\sum_{n \geq 0} \tilde{g}_{ij}(\xi_n, t - \widetilde{S}_n) \right) \\
&= \sum_{j \in \mathcal{S}} \sum_{k=-\infty}^{\infty} \sum_{n \geq 0} \tilde{g}_{ij}(j, t - kd) \cdot \mathbb{P}_i(\xi_n = j, \widetilde{S}_n = kd) \\
&= \sum_{j \in \mathcal{S}} \sum_{n=-\infty}^{\infty} \tilde{g}_{ij}(j, t - nd) \cdot \widetilde{\mathbb{U}}_i(\{j\} \times \{nd\}) \\
&= \sum_{j \in \mathcal{S}} \sum_{n=-\infty}^{\infty} \tilde{g}_{ij}(j, t - nd) \cdot \tilde{u}_{ij}(n)
\end{aligned}$$

with $\tilde{u}_{ij}(n) := \widetilde{\mathbb{U}}_i(\{j\} \times \{nd\})$, and as usual $\widetilde{\mathbb{U}}_i$ denotes the Markov renewal measure of $(\xi_n, \widetilde{S}_n)_{n \geq 0}$ with $\widetilde{\mathbb{U}}^{ij}(B) = \widetilde{\mathbb{U}}_i(\{j\} \times B)$ for all Borel sets B. Furthermore,

$$\begin{aligned}
\frac{1}{\mu} \sum_{j \in \mathcal{S}} \pi_j \psi(j, -t + \beta_{ij}) &= \sum_{j \in \mathcal{S}} \frac{d}{\mu} \pi_j \sum_{n=-\infty}^{\infty} g(j, nd + t - \beta_{ij}) = \sum_{j \in \mathcal{S}} \frac{d}{\mu} \pi_j \sum_{n=-\infty}^{\infty} g(j, -nd + t - \beta_{ij}) \\
&= \sum_{j \in \mathcal{S}} \sum_{n=-\infty}^{\infty} g(j, t - nd - \beta_{ij}) \cdot \frac{d}{\mu} \pi_j = \sum_{j \in \mathcal{S}} \sum_{n=-\infty}^{\infty} \tilde{g}_{ij}(j, t - nd) \cdot \frac{d}{\mu} \pi_j.
\end{aligned}$$

By Theorem 4.8 and the fact that the stationary drifts of $(\xi_n, \widetilde{S}_n)_{n \geq 0}$ and $(\xi_n, S_n)_{n \geq 0}$ coincide, $\tilde{u}_{ij}(n) = \widetilde{\mathbb{U}}_i(\{j\} \times \{nd\})$ tends to $\frac{d}{\mu} \pi_j$ for all $j \in \mathcal{S}$, as $n \to \infty$, and to 0, as $n \to -\infty$.

Now, let $\varepsilon > 0$ and choose $n_0 \in \mathbb{N}$ big enough such that $\max_{j \in \mathcal{S}} |\tilde{u}_{ij}(n) - \frac{d}{\mu} \pi_j| < \varepsilon$ for $n > n_0$, $\max_{j \in \mathcal{S}} |\tilde{u}_{ij}(n)| < \varepsilon$ for $n < -n_0$ and $\max_{j \in \mathcal{S}} \sum_{n < -n_0} |\tilde{g}_{ij}(j, t - nd)| < \varepsilon$ for $t \geq 0$. The existence of such an n_0 is guaranteed by the afore-mentioned renewal theorem together with the direct Riemann integrability of $\tilde{g}_{ij}(j, \cdot)$ and the finiteness of $\mathcal{S}$. The direct Riemann integrability of $\tilde{g}_{ij}(j, \cdot)$ follows from the direct Riemann integrability of $g(j, \cdot)$. Thus,

$$\begin{aligned}
&\left| \mathbf{U} * g(i,t) - \frac{1}{\mu} \sum_{j \in \mathcal{S}} \pi_j \psi(j, -t + \beta_{ij}) \right| = \left| \sum_{j \in \mathcal{S}} \sum_{n=-\infty}^{\infty} \tilde{g}_{ij}(j, t - nd) \left(\tilde{u}_{ij}(n) - \frac{d}{\mu} \pi_j \right) \right| \\
&\leq \sum_{j \in \mathcal{S}} \sum_{n < -n_0} |\tilde{g}_{ij}(j, t - nd)| \cdot \left(\varepsilon + \frac{d}{\mu} \pi_j \right) + \sum_{j \in \mathcal{S}} \sum_{n=-n_0}^{n_0} |\tilde{g}_{ij}(j, t - nd)| \cdot |\tilde{u}_{ij}(n) - \frac{d}{\mu} \pi_j| \\
&\quad + \sum_{j \in \mathcal{S}} \sum_{n > n_0} |\tilde{g}_{ij}(j, t - nd)| \cdot \varepsilon
\end{aligned}$$

$$\leq |\mathcal{S}| \cdot \varepsilon \cdot \Big(\varepsilon + \frac{d}{\mu}\Big) + \sum_{j\in\mathcal{S}} \sum_{n=-n_0}^{n_0} |\tilde{g}_{ij}(j, t-nd)| \cdot |\tilde{u}_{ij}(n) - \frac{d}{\mu}\pi_j| + |\mathcal{S}| \cdot C \cdot \varepsilon$$

for $t \geq 0$ and some constant $0 < C < \infty$ whose existence is again guaranteed by the direct Riemann integrability of $\tilde{g}_{ij}(j,\cdot)$ for every $j \in \mathcal{S}$. This property also implies $\lim_{|x|\to\infty} \tilde{g}_{ij}(j,x) = 0$ for all $j \in \mathcal{S}$ and hence

$$\limsup_{t\to\infty} \left| \mathbf{U} * g(i,t) - \frac{1}{\mu} \sum_{j\in\mathcal{S}} \pi_j \psi(j, -t+\beta_{ij}) \right| \leq |\mathcal{S}| \cdot \varepsilon \cdot \Big(\varepsilon + \frac{d}{\mu}\Big) + |\mathcal{S}| \cdot C \cdot \varepsilon.$$

As $\varepsilon > 0$ was arbitrary,

$$\limsup_{t\to\infty} \left| \mathbf{U} * g(i,t) - \frac{1}{\mu} \sum_{j\in\mathcal{S}} \pi_j \psi(j, -t+\beta_{ij}) \right| \leq 0.$$

It remains to note that $\psi(j,\cdot)$ is bounded, since g is direct Riemann integrable, and it is d-periodic by definition. Thus, the convergence in (4.28) is uniform on $[0,d]$ and hence $\psi(j,\cdot)$ is continuous if $g(j,\cdot)$ is continuous. The Fourier coefficients of $\psi(j,\cdot)$ are easily computed by

$$\begin{aligned}
\widehat{\psi}(j,\cdot)(m) &= \frac{1}{d}\int_0^d \psi(j,t) e^{-2\pi \mathrm{i} mt/d}\,\mathrm{d}t = \int_0^d \sum_{n=-\infty}^{\infty} g(j, nd-t) e^{-2\pi \mathrm{i} mt/d}\,\mathrm{d}t \\
&= \sum_{n=-\infty}^{\infty} \int_0^d g(j, nd-t) e^{-2\pi \mathrm{i} mt/d}\,\mathrm{d}t = \int_{-\infty}^{\infty} g(j,-t) e^{-2\pi \mathrm{i} mt/d}\,\mathrm{d}t \\
&= \int_{-\infty}^{\infty} g(j,t) e^{2\pi \mathrm{i} mt/d}\,\mathrm{d}t
\end{aligned}$$

using the uniform convergence on $[0,d]$. The convergence of the Fourier series holds in L^2 since $\psi(j,\cdot)$ is bounded, and hence $\psi(j,\cdot)$ and its Fourier series are equal $\lambda\!\!\lambda_0$-a.e.

For asymptotics at $-\infty$, we proceed in the same way. Choose $n_0 \in \mathbb{N}$ such that $\max_{j\in\mathcal{S}} |\tilde{u}_{ij}(n) - \frac{d}{\mu}\pi_j| < \varepsilon$ for $n > n_0$, $\max_{j\in\mathcal{S}} |\tilde{u}_{ij}(n)| < \varepsilon$ for $n < -n_0$, as well as

$$\max_{j\in\mathcal{S}} \sum_{n>n_0} |\tilde{g}_{ij}(j, t-nd)| < \varepsilon \quad \text{for } t \leq 0,$$

then, as $t \to -\infty$,

$$\begin{aligned}
|\mathbf{U} * g(i,t)| &\leq \sum_{j\in\mathcal{S}} \sum_{n=-\infty}^{\infty} |\tilde{g}_{ij}(j,t-nd)| \cdot |\tilde{u}_{ij}(n)| \\
&\leq \sum_{j\in\mathcal{S}} \sum_{n<-n_0} |\tilde{g}_{ij}(j,t-nd)| \cdot \varepsilon + \sum_{j\in\mathcal{S}} \sum_{n=-n_0}^{n_0} |\tilde{g}_{ij}(j,t-nd)| \cdot |\tilde{u}_{ij}(n)| \\
&\quad + \sum_{j\in\mathcal{S}} \sum_{n>n_0} |\tilde{g}_{ij}(j,t-nd)| \cdot \left| \tilde{u}_{ij}(n) - \frac{d}{\mu}\pi_j \right| + \sum_{j\in\mathcal{S}} \sum_{n>n_0} |\tilde{g}_{ij}(j,t-nd)| \cdot \frac{d}{\mu}\pi_j \\
&\leq |\mathcal{S}| \cdot C \cdot \varepsilon + \sum_{j\in\mathcal{S}} \sum_{n=-n_0}^{n_0} |\tilde{g}_{ij}(j,t-nd)| \cdot |\tilde{u}_{ij}(n)| + |\mathcal{S}| \cdot \varepsilon^2 + |\mathcal{S}| \cdot \varepsilon \cdot \frac{d}{\mu} = o(1)
\end{aligned}$$

with an analogous reasoning. □

4.4. Proofs of main results

We prove Theorem 4.5, then Theorem 4.1 follows immediately.

Proof of Theorem 4.5. Let $\bar{f}$ be a non-negative function on $\mathcal{S}^*$. For arbitrary $k \geq 0$, we have $P_i(\xi_1 \cdots \xi_k) = e^{-S_k}$ $\mathbb{P}_i$-a.s. and thus we get

$$\sum_{\alpha_1,\ldots,\alpha_k \in \mathcal{S}} \bar{f}(\alpha_1 \cdots \alpha_k j) \mathbb{1}_{\{P_i(\alpha_1 \cdots \alpha_k j) > 0\}} = \sum_{\alpha_1,\ldots,\alpha_k \in \mathcal{S}} \frac{\bar{f}(\alpha_1 \cdots \alpha_k j)}{P_i(\alpha_1 \cdots \alpha_k j)} P_i(\alpha_1 \cdots \alpha_k j) \mathbb{1}_{\{P_i(\alpha_1 \cdots \alpha_k j) > 0\}}$$
$$= \sum_{\alpha_1,\ldots,\alpha_k \in \mathcal{S}} \frac{\bar{f}(\alpha_1 \cdots \alpha_k j)}{P_i(\alpha_1 \cdots \alpha_k j)} \mathbb{1}_{\{P_i(\alpha_1 \cdots \alpha_k j) > 0\}} \mathbb{P}_i\left(\xi_1 = \alpha_1, \ldots, \xi_k = \alpha_k, \xi_{k+1} = j\right)$$
$$= \mathbb{E}_i\left(\frac{\bar{f}(\xi_1 \cdots \xi_{k+1})}{P_i(\xi_1 \cdots \xi_{k+1})} \mathbb{1}_{\{P_i(\xi_1 \cdots \xi_{k+1}) > 0\}} \mathbb{1}_{\{j\}}(\xi_{k+1})\right) = \mathbb{E}_i\left(e^{S_{k+1}} \bar{f}(\xi_1 \cdots \xi_{k+1}) \mathbb{1}_{\{j\}}(\xi_{k+1})\right),$$

since $\{P_i(\xi_1 \cdots \xi_{k+1}) = 0\}$ is a $\mathbb{P}_i$-null set. So

$$\sum_{\alpha \in \mathcal{S}^*} \bar{f}(\alpha j) \mathbb{1}_{\{P_i(\alpha j) > 0\}} = \sum_{k=0}^{\infty} \mathbb{E}_i\left(e^{S_{k+1}} \bar{f}(\xi_1 \cdots \xi_{k+1}) \mathbb{1}_{\{j\}}(\xi_{k+1})\right)$$
$$= \sum_{k=1}^{\infty} \mathbb{E}_i\left(e^{S_k} \bar{f}(\xi_1 \cdots \xi_k) \mathbb{1}_{\{j\}}(\xi_k)\right).$$

In particular, if $\bar{f}(\alpha) := f(\lambda P_i(\alpha))$ and thus $\bar{f}(\xi_1 \cdots \xi_k) = f(\lambda e^{-S_k})$, we have

$$F_i^j(\lambda) = \sum_{\alpha \in \mathcal{S}^*} f(\lambda P_i(\alpha j)) \mathbb{1}_{\{P_i(\alpha j) > 0\}} = \sum_{k=1}^{\infty} \mathbb{E}_i\left(e^{S_k} f(\lambda e^{-S_k}) \mathbb{1}_{\{j\}}(\xi_k)\right)$$
$$= \sum_{k=0}^{\infty} \mathbb{E}_i\left(e^{S_k} f(\lambda e^{-S_k}) \mathbb{1}_{\{j\}}(\xi_k)\right) - \mathbb{E}_i\left(e^0 f(\lambda e^{-0}) \mathbb{1}_{\{j\}}(\xi_0)\right)$$
$$= \mathbb{E}_i\left(\sum_{k=0}^{\infty} e^{S_k} f(\lambda e^{-S_k}) \mathbb{1}_{\{j\}}(\xi_k)\right) - f(\lambda) \mathbb{1}_{\{j\}}(i).$$

If we define $g(t) := e^t f(e^{-t})$ and $f_1(x) := \frac{f(x)}{x}$ so that $g(t) = e^t f(e^{-t}) = f_1(e^{-t})$, then we get

$$\mathbb{E}_i\left(\sum_{k=0}^{\infty} e^{S_k} f(\lambda e^{-S_k}) \mathbb{1}_{\{j\}}(\xi_k)\right) = \lambda \cdot \mathbb{E}_i\left(\sum_{k=0}^{\infty} f_1(\lambda e^{-S_k}) \mathbb{1}_{\{j\}}(\xi_k)\right)$$
$$= \lambda \cdot \mathbb{E}_i\left(\sum_{k=0}^{\infty} f_1(e^{-(S_k - \log \lambda)}) \mathbb{1}_{\{j\}}(\xi_k)\right) = \lambda \cdot \mathbb{E}_i\left(\sum_{k=0}^{\infty} \mathbb{1}_{\{j\}}(\xi_k) g(S_k - \log \lambda)\right)$$
$$= \lambda \cdot \mathbb{E}_i\left(\sum_{k=0}^{\infty} \tilde{g}_j(\xi_k, \log \lambda - S_k)\right) = \lambda \cdot \mathbf{U} * \tilde{g}_j(i, \log \lambda)$$

with $\tilde{g}_j(x,y) := \mathbb{1}_{\{j\}}(x) g(-y)$. Before applying the *Markov renewal theorem II* (Theorem 4.12) we note that $F_i^j(\lambda)$ is finite for every $0 < \lambda < \infty$. To see this, it suffices to argue, that $\mathbf{U} * g(i,t)$ is finite for every $t \in \mathbb{R}$ and g satisfying (4.26) and (4.27) (these conditions will be verified for $\tilde{g}_j$ below). However, this was shown in the proof of [Als15b, Theorem 9.18].

We want to apply *Markov renewal theorem II*, so we need to check conditions (4.26) and (4.27) for $\tilde{g}_j$. Condition (4.26) is satisfied since $\tilde{g}_j(j,y) = g(-y) = e^{-y} f(e^y)$ and f is $\lambda\!\!\lambda_0$-a.e. continuous, and $\tilde{g}_j(k,y)$ equals 0 for $k \neq j$. Also, $\tilde{g}_j(k,\cdot)$ is direct Riemann integrable for $k \in \mathcal{S}$. This is trivial for $k \neq j$ and reduces to the verification of direct Riemann integrability $g(-\cdot)$ or g in the case when $k = j$. The latter direct Riemann integrability was shown in the proof of [Jan12a, Theorem 5.1].

Applying Markov renewal theorem II (Theorem 4.9) and $\lim_{\lambda\to\infty} f(\lambda)/\lambda = 0$ in the non-arithmetic case, we get

$$\lim_{\lambda\to\infty}\left(\mathbf{U} * \tilde{g}_j(i,\log\lambda) - \frac{f(\lambda)}{\lambda}\mathbb{1}_{\{j\}}(i)\right) = \frac{1}{\mu}\sum_{k\in\mathcal{S}}\pi_k \int_{\mathbb{R}} \tilde{g}_j(k,y)\,\lambda\!\!\lambda_0(\mathrm{d}y)$$
$$= \frac{1}{\mu}\sum_{k\in\mathcal{S}}\pi_k \mathbb{1}_{\{j\}}(k)\int_{-\infty}^{\infty} g(y)\,\mathrm{d}y = \frac{\pi_j}{\mu}\int_{-\infty}^{\infty} g(y)\,\mathrm{d}y = \frac{\pi_j}{\mu}\int_{-\infty}^{\infty} e^{-t}e^{2t} f(e^{-t})\,\mathrm{d}t = \frac{\pi_j}{\mu}\int_0^{\infty} x^{-2} f(x)\,\mathrm{d}x,$$

where we substituted $x := e^{-t}$ in the last step. This proves the theorem in the non-arithmetic case.

In the d-arithmetic case with shift function β, the Markov renewal theorem II yields, as $\lambda \to \infty$,

$$\begin{aligned}
\mathbf{U} * \tilde{g}_j(i,\log\lambda) &= \frac{1}{\mu}\sum_{k\in\mathcal{S}} \pi_k \widetilde{\psi}_j(k, -\log\lambda + \beta_{ik}) + o(1) \\
&= \frac{1}{\mu}\sum_{k\in\mathcal{S}} \pi_k \cdot d \sum_{n=-\infty}^{\infty} \tilde{g}_j(k, nd - \beta_{ik} + \log\lambda) + o(1) \\
&= \frac{1}{\mu}\sum_{k\in\mathcal{S}} \pi_k \mathbb{1}_{\{j\}}(k) \cdot d \sum_{n=-\infty}^{\infty} g(-nd + \beta_{ik} - \log\lambda) + o(1) \\
&= \frac{\pi_j}{\mu}\widehat{\psi}(0) + \frac{\pi_j}{\mu}\cdot\left(d\sum_{n=-\infty}^{\infty} g(nd - (\log\lambda - \beta_{ij})) - \widehat{\psi}(0)\right) + o(1) \\
&= \frac{\pi_j}{\mu}\widehat{\psi}(0) + \frac{\pi_j}{\mu}\psi(\log\lambda - \beta_{ij}) + o(1)
\end{aligned}$$

with $\widetilde{\psi}_j$ from Theorem 4.12 corresponding to $\tilde{g}_j$, and $\widehat{\psi}(0) := \int_{-\infty}^{\infty} g(t)\mathrm{d}t$. Combining this with $\frac{f(\lambda)}{\lambda}\mathbb{1}_{\{j\}}(i) = o(1)$ as $\lambda \to \infty$, it gives us the desired result in the d-arithmetic case. All properties of the function ψ transfer directly from $\widetilde{\psi}_j$ in Theorem 4.12 (which is called ψ there). □

Proof of Theorem 4.1. We conclude the assertions of this theorem directly from Theorem 4.5 by noting that

$$F_i(\lambda) = \sum_{\alpha\in\mathcal{S}^*} f(\lambda P_i(\alpha))\mathbb{1}_{\{P_i(\alpha)>0\}} = f(\lambda P_i(\emptyset)) + \sum_{j\in\mathcal{S}}\left(\sum_{\alpha\in\mathcal{S}_j^*} f(\lambda P_i(\alpha))\mathbb{1}_{\{P_i(\alpha)>0\}}\right)$$

with $\mathcal{S}_j^* \subset \mathcal{S}^*$ being the subset of all nodes that end with j. So if we rearrange this expression, then we obtain the well-known quantity

$$F_i(\lambda) = f(\lambda) + \sum_{j\in\mathcal{S}}\left(\sum_{\alpha\in\mathcal{S}^*} f(\lambda P_i(\alpha j))\mathbb{1}_{\{P_i(\alpha j)>0\}}\right) = f(\lambda) + \sum_{j\in\mathcal{S}} F_i^j(\lambda).$$

Now $f(\lambda)/\lambda = o(1)$, as $\lambda \to \infty$, and thus an appeal to Theorem 4.5 completes the proof. Note that $F_i(\lambda)$ converges for every $\lambda > 0$ because $\mathcal{S}$ is finite and $F_i^j(\lambda)$ converges for every $0 < \lambda < \infty$. □

Proof of Theorem 4.6. We divide the proof into four steps, starting with non-negative f and $a = 0$. The generalization to real f is then straightforward. For the third step, we consider a special f with $a = 1$ and subsequently, we combine these results to show the full statement. This proof relies heavily on the proof of [Jan12b, Theorem 2.1].

Step 1: Let $f \geq 0$ and $a = 0$. Then the assertion is a reformulation of Theorem 4.5. We quickly check this below. Since we are in the exact situation of Theorem 4.5, the sum in (4.13) converges (absolutely) because it does in (4.9). If $(\xi_n, S_n)_{n\geq 0}$ is non-arithmetic, then obviously we get the same result. If $(\xi_n, S_n)_{n\geq 0}$ is d-arithmetic with shift function β, then we find that (4.16) is similar to (4.11) with $\psi_0(t) := \psi(t)$ and the notation from Theorem 4.5. We simply need to verify (4.17)-(4.19). First of all, by Theorem 4.5

$$\begin{aligned}\psi_0(t) - \psi_0(0) = \psi(t) - \psi(0) &= d \sum_{n=-\infty}^{\infty} g(nd - t) - d \sum_{n=-\infty}^{\infty} g(nd) \\ &= d \sum_{n=-\infty}^{\infty} (g(nd - t) - g(nd)) - 0 \cdot t,\end{aligned}$$

so (4.19) follows. Now, $\psi(t)$ has the Fourier series

$$\sum_{m\neq 0} \widehat{\psi}(m) e^{2\pi \mathrm{i}\, mt/d}$$

with

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

which coincides with $\widehat{\psi}_0(m)$ as Janson points out in [Jan12b, Proof of Theorem 2.1, Step 1]. Note that $\widehat{\psi}(0) = \int_0^\infty f(x) x^{-2} \,\mathrm{d}x$. If f is continuous, then so is ψ and consequently ψ_0 is, too. This completes the first step.

Step 2: Let f be real-valued and $a = 0$. We decompose $f = f^+ - f^-$ with non-negative f^+ and f^-. Since $\max\{\cdot, 0\}$ is a continuous function, the conditions in (4.12) are clearly fulfilled for f^+ and f^-, since f obeys them. With this partitioning we get

$$\sum_{\alpha\in\mathcal{S}^*} |f(\lambda P_i(\alpha j))| \mathbb{1}_{\{P_i(\alpha j)>0\}} \leq \sum_{\alpha\in\mathcal{S}^*} f^+(\lambda P_i(\alpha j)) \mathbb{1}_{\{P_i(\alpha j)>0\}} + \sum_{\alpha\in\mathcal{S}^*} f^-(\lambda P_i(\alpha j)) \mathbb{1}_{\{P_i(\alpha j)>0\}} < \infty$$

for every $0 < \lambda < \infty$ since both summands converge absolutely by our previous result. Step 1 also gives us the asymptotic expansions

$$F_i^{j,\pm}(\lambda) = \frac{\pi_j}{\mu} \int_0^\infty f^\pm(x) x^{-2} \,\mathrm{d}x \cdot \lambda + \frac{\pi_j}{\mu} \psi_0^\pm(\log\lambda - \beta_{ij})\lambda + o(\lambda)$$

with obviously defined $\psi_0^\pm$ which vanish in the non-arithmetic case. This yields the asymptotics

$$F_i^{j}(\lambda) = \frac{\pi_j}{\mu} \int_0^\infty f(x) x^{-2} \,\mathrm{d}x \cdot \lambda + \frac{\pi_j}{\mu} \psi_0(\log\lambda - \beta_{ij})\lambda + o(\lambda),$$

where $\psi_0(t) := \psi_0^+(t) - \psi_0^-(t)$ is bounded and d-periodic and satisfies (4.17)-(4.19). This completes the second step.

Step 3: Let $f(x) := x\mathbb{1}_{\{x\geq 1\}}$ be the $\lambda\!\!\lambda_0$-a.e. continuous function that trivially satisfies the conditions in (4.12) with $a = 1$ and $\delta = 1$. We calculate

$$g(t) = e^t f(e^{-t}) = \mathbb{1}_{\{e^{-t}\geq 1\}} = \mathbb{1}_{\{t\leq 0\}},$$

so $g(x-\log\lambda) = \mathbb{1}_{(-\infty,\log\lambda]}(x)$. Noting that $f \geq 0$, we proceed as in the proof of Theorem 4.5 and find

$$\begin{aligned} F_i^j(\lambda) &= \mathbb{E}_i\left(\sum_{k=0}^{\infty} e^{S_k} f(\lambda e^{-S_k})\mathbb{1}_{\{j\}}(\xi_k)\right) - \lambda\mathbb{1}_{\{\lambda\geq 1\}}\mathbb{1}_{\{j\}}(i) \\ &= \lambda\cdot\mathbb{E}_i\left(\sum_{k=0}^{\infty}\mathbb{1}_{\{j\}}(\xi_k)g(S_k-\log\lambda)\right) - \lambda\mathbb{1}_{\{\lambda\geq 1\}}\mathbb{1}_{\{j\}}(i) \\ &= \lambda\cdot\mathbb{U}_i(\{j\}\times[0,\log\lambda]) - \lambda\mathbb{1}_{\{\lambda\geq 1\}}\mathbb{1}_{\{j\}}(i) \\ &= \lambda\cdot\mathbb{U}^{ij}(\log\lambda) - \lambda\mathbb{1}_{\{\lambda\geq 1\}}\mathbb{1}_{\{j\}}(i) \end{aligned}$$

and the (absolute) convergence of $F_i^j(\lambda)$ for every $0<\lambda<\infty$ by Theorem 3.21. In the non-arithmetic case, this theorem further yields

$$\begin{aligned} F_i^j(\lambda) &= \lambda\cdot\left(\frac{\pi_j}{\mu}\log\lambda + \frac{\pi_j^2}{2\mu^2}\mathbb{E}_j S_{\sigma_1(j)}^2 - \frac{\pi_j}{\mu}\mathbb{E}_i S_{\sigma_1(j)}\mathbb{1}_{\{j\neq i\}} + o(1)\right) - \lambda\mathbb{1}_{\{\lambda\geq 1\}}\mathbb{1}_{\{j\}}(i) \\ &= \frac{\pi_j}{\mu}\lambda\log\lambda + \left(\frac{\pi_j^2}{2\mu}\mathbb{E}_j S_{\sigma_1(j)}^2 - \pi_j\mathbb{E}_i S_{\sigma_1(j)}\mathbb{1}_{\{j\neq i\}}\right)\frac{1}{\mu}\lambda - \lambda\mathbb{1}_{\{j\}}(i) + o(\lambda), \end{aligned}$$

as $\lambda\to\infty$, which is the desired result since the integral in b vanishes trivially. This leaves us with the study of the d-arithmetic case with shift function β. With $\psi_0(t) := d(\frac{1}{2}-\{\frac{t}{d}\})$ the theorem gives us

$$\begin{aligned} F_i^j(\lambda) &= \lambda\cdot\left(\frac{\pi_j}{\mu}\log\lambda + \frac{\pi_j}{\mu}\psi_0(\log\lambda-\beta_{ij}) + \frac{\pi_j^2}{2\mu^2}\mathbb{E}_j S_{\sigma_1(j)}^2 - \frac{\pi_j}{\mu}\mathbb{E}_i S_{\sigma_1(j)}\mathbb{1}_{\{j\neq i\}} + o(1)\right) \\ &\quad - \lambda\mathbb{1}_{\{\lambda\geq 1\}}\mathbb{1}_{\{j\}}(i) \\ &= \frac{\pi_j}{\mu}\lambda\log\lambda + \left(\frac{\pi_j^2}{2\mu}\mathbb{E}_j S_{\sigma_1(j)}^2 - \pi_j\mathbb{E}_i S_{\sigma_1(j)}\mathbb{1}_{\{j\neq i\}}\right)\frac{1}{\mu}\lambda + \frac{\pi_j}{\mu}\psi_0(\log\lambda-\beta_{ij})\cdot\lambda \\ &\quad - \lambda\mathbb{1}_{\{j\}}(i) + o(\lambda) \end{aligned}$$

which is again the desired result provided that ψ_0 satisfies the required conditions. From the definition of the fractional part, it is obvious that ψ_0 is bounded and d-periodic. It is easy to show (cf. [Jan12b, Proof of Theorem 2.1]) that

$$\widehat{\psi}_0(0) = \frac{1}{d}\int_0^d \psi_0(t)\,\mathrm{d}t = 0$$

and, for $m\neq 0$,

$$\widehat{\psi}_0(m) = \frac{1}{d}\int_0^d \psi_0(t)e^{-2\pi\,\mathrm{i}\,mt/d}\,\mathrm{d}t = \frac{d}{2\pi\,\mathrm{i}\,m}.$$

Additionally, we have for real $u\neq 0$ that

$$\hat{g}(u) = \lim_{\varepsilon\searrow 0}\int_{-\infty}^{\infty} e^{-\,\mathrm{i}\,ut+\varepsilon t}g(t)\,\mathrm{d}t = \lim_{\varepsilon\searrow 0}\int_{-\infty}^{0} e^{(\varepsilon-\mathrm{i}\,u)t}\,\mathrm{d}t = \frac{1}{-\,\mathrm{i}\,u}$$

which agrees with $\widehat{\psi}_0(m) = \hat{g}(-2\pi m/d)$ for $m\neq 0$. Together with

$$\sum_{n=-\infty}^{\infty}(g(nd-t)-g(nd)) = \sum_{n=-\infty}^{\infty}\left(\mathbb{1}_{\{n\leq\frac{t}{d}\}} - \mathbb{1}_{\{n\leq 0\}}\right) = \left\lfloor\frac{t}{d}\right\rfloor,$$

we find that

$$d \sum_{n=-\infty}^{\infty} (g(nd-t) - g(nd)) - 1 \cdot t = d \left\lfloor \frac{t}{d} \right\rfloor - t = -d \left\{ \frac{t}{d} \right\} = \psi_0(t) - \psi_0(0)$$

which gives us the validity of (4.19).

Step 4: In the last step, we consider a general f (with the properties from the theorem). We decompose f as $f(x) = f_1(x) + af_2(x)$ with

$$f_1(x) := f(x) - ax\mathbb{1}_{\{x \geq 1\}} \quad \text{and} \quad f_2(x) := x\mathbb{1}_{\{x \geq 1\}}.$$

Now, both parts of f_1 are of order $\mathcal{O}(x^{1+\delta})$ for $0 < x < 1$, hence the same holds for f_1. We extract from Step 3 that

$$f_1(x) = ax + \mathcal{O}(x^{1-\delta}) - (ax + \mathcal{O}(x^{1-\delta})) = \mathcal{O}(x^{1-\delta})$$

for $1 < x < \infty$, thus our corresponding constant a_1 equals 0, and Step 2 applies to f_1 (and Step 3 applies to f_2 as seen before). Denote the associated functionals by $F_i^{j,1}(\lambda)$ and $F_i^{j,2}(\lambda)$. Then each of them converges absolutely for every $0 < \lambda < \infty$, so for every λ

$$\sum_{\alpha \in \mathcal{S}^*} |f(\lambda P_i(\alpha j))| \mathbb{1}_{\{P_i(\alpha j) > 0\}} \leq \sum_{\alpha \in \mathcal{S}^*} |f_1(\lambda P_i(\alpha j))| \mathbb{1}_{\{P_i(\alpha j) > 0\}} + |a| F_i^{j,2}(\lambda) < \infty.$$

By the previous steps, we get

$$F_i^{j,1}(\lambda) = \frac{b_j^{(1)}}{\mu} \lambda + \frac{\pi_j}{\mu} \psi_0^{(1)}(\log \lambda - \beta_{ij}) \cdot \lambda + o(\lambda),$$

as $\lambda \to \infty$, where $b_j^{(1)} = \pi_j \int_0^\infty f_1(x) x^{-2} \, \mathrm{d}x$ and $\psi_0^{(1)} = 0$ in the non-arithmetic case and $\psi_0^{(1)}(t) - \psi_0^{(1)}(0) = d \sum_{n=-\infty}^{\infty} (g_1(nd-t) - g_1(nd))$ for the obvious g_1 in the d-arithmetic case with shift function β. We skip the consideration of the Fourier series at this point. Similarly, we get

$$\begin{aligned} F_i^{j,2}(\lambda) = {} & \frac{\pi_j}{\mu} \lambda \log \lambda + \left(\frac{\pi_j^2}{2\mu} \mathbb{E}_j S_{\sigma_1(j)}^2 - \pi_j \mathbb{E}_i S_{\sigma_1(j)} \mathbb{1}_{\{j \neq i\}} \right) \frac{1}{\mu} \lambda + \frac{\pi_j}{\mu} \psi_0(\log \lambda - \beta_{ij}) \cdot \lambda \\ & - \lambda \mathbb{1}_{\{j\}}(i) + o(\lambda) \end{aligned}$$

with the notation from Step 3 (which includes the vanishing of ψ_0 in the non-arithmetic case). A combination of these asymptotics yields

$$\begin{aligned} F_i^j(\lambda) = {} & \frac{a\pi_j}{\mu} \lambda \log \lambda + \frac{\pi_j}{\mu} \bar{\psi}_0(\log \lambda - \beta_{ij}) \cdot \lambda + \frac{1}{\mu} \left(\frac{a\pi_j^2}{2\mu} \mathbb{E}_j S_{\sigma_1(j)}^2 - a\pi_j \mathbb{E}_i S_{\sigma_1(j)} \mathbb{1}_{\{j \neq i\}} \right. \\ & \left. + \pi_j \int_0^\infty (f(x) - ax\mathbb{1}_{\{x \geq 1\}}) x^{-2} \, \mathrm{d}x \right) \lambda - \lambda \mathbb{1}_{\{j\}}(i) + o(\lambda) \end{aligned}$$

for $\bar{\psi}_0(t) := \psi_0^{(1)}(t) + a\psi_0(t)$. This coincides with the desired asymptotics if we can verify the conditions that we require for $\bar{\psi}_0$. To begin with, $\bar{\psi}_0$ is bounded and d-periodic. Furthermore, since $g_1(t) = g(t) - a\mathbb{1}_{\{t \leq 0\}}$,

$$\bar{\psi}_0(t) - \bar{\psi}_0(0) = (\psi_0^{(1)}(t) - \psi_0^{(1)}(0)) + a(\psi_0(t) - \psi_0(0)) = d \sum_{n=-\infty}^{\infty} (g_1(nd-t) - g_1(nd)) - ad \left\{ \frac{t}{d} \right\}$$

$$
\begin{aligned}
&= d \sum_{n=-\infty}^{\infty} (g(nd-t) - g(nd)) - ad \sum_{n=-\infty}^{\infty} (\mathbb{1}_{\{n \le \frac{t}{d}\}} - \mathbb{1}_{\{n \le 0\}}) - ad \left\{ \frac{t}{d} \right\} \\
&= d \sum_{n=-\infty}^{\infty} (g(nd-t) - g(nd)) - ad \left(\left\lfloor \frac{t}{d} \right\rfloor + \left\{ \frac{t}{d} \right\} \right) \\
&= d \sum_{n=-\infty}^{\infty} (g(nd-t) - g(nd)) - at
\end{aligned}
$$

as desired. The statement concerning the Fourier series follows from the fact that the Fourier series of $\bar{\psi}_0$ is the sum of the Fourier series of $\psi_0^{(1)}$ and $a\psi_0$ which we know well. The rest follows by linearity. If f is continuous, then g is continuous and $\bar{\psi}_0$ converges uniformly on every compact set, so $\bar{\psi}_0$ is continuous. □

Proof of Theorem 4.7. Again, we deduce the assertions of this theorem directly from Theorem 4.6 by noting that

$$
\begin{aligned}
F_i(\lambda) &= f(\lambda) + \sum_{j \in \mathcal{S}} F_i^j(\lambda) \\
&= f(\lambda) + \sum_{j \in \mathcal{S}} \left(\frac{a\pi_j}{\mu} \lambda \log \lambda + \frac{b_j}{\mu} \lambda + \frac{\pi_j}{\mu} \psi_0(\log \lambda - \beta_{ij}) \lambda - a\lambda \mathbb{1}_{\{j\}}(i) + o(\lambda) \right) \\
&= f(\lambda) - a\lambda + \frac{a}{\mu} \lambda \log \lambda + \frac{b}{\mu} \lambda + \sum_{j \in \mathcal{S}} \frac{\pi_j}{\mu} \psi_0(\log \lambda - \beta_{ij}) \lambda + o(\lambda) \\
&= \frac{a}{\mu} \lambda \log \lambda + \frac{b}{\mu} \lambda + \frac{1}{\mu} \sum_{j \in \mathcal{S}} \pi_j \psi_0(\log \lambda - \beta_{ij}) \lambda + o(\lambda)
\end{aligned}
$$

with vanishing ψ_0 in the non-arithmetic case. Note that $f(\lambda) - a\lambda = \mathcal{O}(\lambda^{1-\delta})$, as $\lambda \to \infty$. $F_i(\lambda)$ converges absolutely for every $\lambda > 0$ because $\mathcal{S}$ is finite and $F_i^j(\lambda)$ converges absolutely for every $0 < \lambda < \infty$. □

4.5. Applications

We apply our device from Section 4.1 to some selected parameters. The selection of parameters is taken from [Jan12a]. There are of course others, e.g. in [Jan12b], Janson examines protected nodes and fringe tries in this way in the i.i.d. model.

Before actually applying the device, we state once again that we intend to showcase a method. Nevertheless, we briefly compare our (following) results to the respective results from earlier publications in the Markov case: Part (4.29) of Theorem 4.13 is a replication of [Rég88, Theorem III.2] (as we understand it as a result for the average *size of a trie*), however the latter does not seem to be precise in the d-arithmetic case, as a periodic function should appear in the leading term. In [CFV01, Theorem 6 and Corollary 2], the authors derive a corresponding result in the more general setting of dynamical sources (also for the path length) and their corollary for the Markov case matches our result. However, we provide an explicit formula for the periodic oscillatory term. Also [JS16, Theorem 2] contains the sketch of a similar result to the one in [CFV01].

Régnier's result does also apply to *b-tries*, and we understand Theorem 4.18 as a more precise version of her [Rég88, Theorem III.2] in the context of b-tries. We are not aware of results for *external nodes of a b-trie* as in Theorem 4.19 in a different model than the memoryless model.

Theorem 4.22 contains the same result as [Bou01, Proposition 7] for the *size of a* PAT*-trie*. Bourdon also derived a corresponding result for dynamical sources first and obtained the Markov case as a corollary. Then again, we provide explicit formulas for the periodic oscillatory terms.

Although they did not formulate their results in terms of the depth but rather in terms of the external path length, Bourdon [Bou01, Proposition 7] and Leckey [Lec15, Corollary 2.2.2] (binary case) derived expansions corresponding to Theorem 4.37 for the expected *external path length of a* PAT*-trie* and to Theorem 4.32 for the expected *depth in a* PAT*-trie* (implicitly via $\mathbb{E}_i D_n^P = \frac{1}{n}\mathbb{E}_i \mathrm{L}_n^P$). Again, Bourdon also derived a corresponding result for dynamical sources first and obtained the Markov case as a corollary, but we provide explicit formulas for the periodic oscillatory terms. Leckey's expansion is not as precise as ours or the formerly mentioned of Bourdon, however this is owed to the objective to derive a CLT for the path length (and not the depth).

Regarding the *external path length of a trie*, Theorem 4.35 is again matched by [CFV01, Theorem 6 and Corollary 2], and still we provide an explicit formula for the periodic oscillatory term. Also, as for PAT-tries, [Lec15, Corollary 2.2.1] or [LNS15, Theorem 2.1] contain an expansion of the mean and the variance and a CLT in the binary case.

4.5.1. Expected size of a trie

We recall from the beginning of Chapter 4 that $\widetilde{W}_\lambda = W_{\Pi(\lambda)}$ satisfies

$$\mathbb{E}_i \widetilde{W}_\lambda = \sum_{\alpha \in \mathcal{S}^*} f(\lambda P_i(\alpha)) \mathbb{1}_{\{P_i(\alpha)>0\}}$$

for $f(x) = 1-(1+x)e^{-x}$. The main theorem is the following:

Theorem 4.13. *(a) If $(\xi_n, S_n)_{n\geq 0}$ is non-arithmetic, then, as $n\to\infty$,*

$$\frac{\mathbb{E}_i W_n}{n} \to \frac{1}{\mu}. \tag{4.29}$$

(b) If $(\xi_n, S_n)_{n\geq 0}$ is d-arithmetic with shift function β, then, as $n\to\infty$,

$$\frac{\mathbb{E}_i W_n}{n} = \frac{1}{\mu} + \frac{1}{\mu}\sum_{j\in\mathcal{S}} \pi_j \psi_2(\log n - \beta_{ij}) + o(1), \tag{4.30}$$

where β_{ij} are from Theorem 3.3 and the continuous and d-periodic function ψ_2 is from Theorem 4.1 with f from above and corresponding g. ψ_2 has Fourier series

$$\sum_{m\neq 0} \frac{\Gamma(1-2\pi\,\mathrm{i}\, m/d)}{1+2\pi\,\mathrm{i}\, m/d} e^{2\pi\,\mathrm{i}\, mt/d} = \sum_{m\neq 0} \frac{2\pi\,\mathrm{i}\, m}{d}\Gamma\left(-1-\frac{2\pi\,\mathrm{i}\, m}{d}\right) e^{2\pi\,\mathrm{i}\, mt/d}. \tag{4.31}$$

Remark 4.14. In the d-arithmetic case of the previous theorem we mention functions ψ_2, f and g. They are connected via (4.6), (4.8) (ψ_2 is called ψ there) and $g(t) = e^t f(e^{-t})$. Whenever such a triple of functions is mentioned in the following theorems, their connection is of that type (considering the right theorem out of Theorems 4.1, 4.5, 4.6 and 4.7).

Remark 4.15. We get asymptotics for the expected number of nodes in Trie($\mathcal{M}_n$) by $\mathbb{E}_i W_n + n$.

Remark 4.16. Appealing to Remark 4.3, the current $\hat{g}(s) = -\,\mathrm{i}\, s\Gamma(-1+\mathrm{i}\, s)$ is of order $\mathcal{O}(s^{-2})$, $|s|\to\infty$, since

$$|\Gamma(-1+\mathrm{i}\, s)| \sim \sqrt{2\pi}|s|^{-3/2}e^{-\pi|s|/2},$$

as $|s|\to\infty$.

First, we prove the Poisson version of Theorem 4.13. Now, Theorem 4.1 immediately yields:

Lemma 4.17. *(a) If $(\xi_n, S_n)_{n\geq 0}$ is non-arithmetic, then, as $\lambda \to \infty$,*

$$\frac{\mathbb{E}_i \widetilde{W}_\lambda}{\lambda} \to \frac{1}{\mu}.$$

(b) If $(\xi_n, S_n)_{n\geq 0}$ is d-arithmetic with shift function β, then, as $\lambda \to \infty$,

$$\frac{\mathbb{E}_i \widetilde{W}_\lambda}{\lambda} = \frac{1}{\mu} + \frac{1}{\mu}\sum_{j\in\mathcal{S}} \pi_j \psi_2(\log\lambda - \beta_{ij}) + o(1)$$

with β_{ij} and ψ_2 as in Theorem 4.13.

Proof. It is easily checked that f satisfies all conditions of Theorem 4.1. Therefore, the problem reduces to a calculation of integrals which was done in the proof of [Jan12a, Theorem 5.3]. Yet, we briefly recall the basic steps, as this provides a scheme for the further applications.

Note that f is continuous, of order $\mathcal{O}(1)$, as $x \to \infty$, and of order $\mathcal{O}(x^2)$, as $x \to 0$ (use l'Hôpital's rule), so $\delta = 1$ is a possible choice in Theorem 4.1. This implies $\int_0^\infty f(x)x^{-2}\,\mathrm{d}x = \int_0^\infty f'(x)x^{-1}\,\mathrm{d}x$, and as $f'(x) = xe^{-x}$, the latter integral can be easily calculated. Most of the time, integration by parts is indeed enough to calculate the occurring integrals. We obtain

$$\frac{\mathbb{E}_i \widetilde{W}_\lambda}{\lambda} = \frac{1}{\mu}\int_0^\infty f(x)x^{-2}\,\mathrm{d}x + o(1) = \frac{1}{\mu} + o(1)$$

and

$$\frac{\mathbb{E}_i \widetilde{W}_\lambda}{\lambda} = \frac{1}{\mu} + \frac{1}{\mu}\sum_{j\in\mathcal{S}} \pi_j \psi_2(\log\lambda - \beta_{ij}) + o(1)$$

in the non-arithmetic and the d-arithmetic case, respectively, as $\lambda \to \infty$, where in the latter case the Fourier transform of the corresponding g is

$$\hat{g}(s) = \int_0^\infty f(x)x^{-2+\mathrm{i}\,s}\,\mathrm{d}x = \frac{\Gamma(1+\mathrm{i}\,s)}{1-\mathrm{i}\,s} = -\mathrm{i}\,s\Gamma(-1+\mathrm{i}\,s).$$

Again, this follows from integration by parts, the definition of the Gamma function and its functional equation $\Gamma(z+1) = z\Gamma(z)$. Since f is continuous, ψ_2 is too. □

Proof of Theorem 4.13. We use the depoissonization result Lemma A.1 which essentially is a sandwiching argument with $\lambda = n \pm n^{2/3}$ and applying the results from the Poisson version of the theorem together with the fact that $\mathbb{E}_i W_n$ increases in n. The latter is obvious from the above characterization, in fact, W_n increases even pointwise.

Obviously, (A.2) applies with $C_1 = 0$, $C_2 = 1/\mu$ and either $\psi \equiv 1$ in the non-arithmetic case, or $\psi(t) = 1 + \sum_{j\in\mathcal{S}} \pi_j\psi_2(t - \beta_{ij})$ in the d-arithmetic case. ψ is bounded and uniformly continuous since all $\psi_2(\cdot - \beta_{ij})$ are. □

4.5.2. Expected size of a b-trie

We consider $\mathrm{Trie}^{(b)}(\mathcal{M}_n)$ which still contains a random number $W_n^{(b)}$ of internal nodes. However, the number of leaves is not deterministic anymore and 1 to b strings can be stored in each leaf. At first, we deal with the expected number of internal nodes which basically requires the same approach as in Subsection 4.5.1.

Internal nodes: Generalizing the characterization of internal nodes in the original trie, a node $\alpha \in \mathcal{S}^*$ is an internal node of $\mathrm{Trie}^{(b)}(\mathcal{M}_n)$ iff there are at least $b+1$ strings in $\mathcal{M}_n$ starting with α. We write

$$W_n^{(b)} = \sum_{\alpha \in \mathcal{S}^*} \mathbb{1}_{\{N_n(\alpha) \geq b+1\}}$$

and, as in the previous subsection, we consider $\widetilde{W}_\lambda^{(b)} := W_{\Pi(\lambda)}^{(b)}$, so

$$\begin{aligned}\mathbb{E}_i \widetilde{W}_\lambda^{(b)} &= \sum_{\alpha \in \mathcal{S}^*} \mathbb{P}_i\left(\widetilde{N}_\lambda(\alpha) \geq b+1\right) \mathbb{1}_{\{P_i(\alpha)>0\}} = \sum_{\alpha \in \mathcal{S}^*} \mathbb{P}(\Pi(\lambda P_i(\alpha)) \geq b+1) \mathbb{1}_{\{P_i(\alpha)>0\}} \\ &= \sum_{\alpha \in \mathcal{S}^*} f^{(b)}(\lambda P_i(\alpha)) \mathbb{1}_{\{P_i(\alpha)>0\}}\end{aligned}$$

for $f^{(b)}$ defined on $(0,\infty)$ by

$$f^{(b)}(x) := \mathbb{P}(\Pi(x) \geq b+1) = 1 - \mathbb{P}(\Pi(x) \leq b) = 1 - e^{-x} \sum_{k=0}^{b} \frac{x^k}{k!}.$$

This is the same function that is used in [Jan12a, Section 6]. Our main result is:

Theorem 4.18. *(a) If $(\xi_n, S_n)_{n\geq 0}$ is non-arithmetic, then, as $n \to \infty$,*

$$\frac{\mathbb{E}_i W_n^{(b)}}{n} \to \frac{1}{\mu b}. \tag{4.32}$$

(b) If $(\xi_n, S_n)_{n\geq 0}$ is d-arithmetic with shift function β, then, as $n \to \infty$,

$$\frac{\mathbb{E}_i W_n^{(b)}}{n} = \frac{1}{\mu b} + \frac{1}{\mu} \sum_{j \in \mathcal{S}} \pi_j \psi_2^{(b)}(\log n - \beta_{ij}) + o(1), \tag{4.33}$$

where β_{ij} are from Theorem 3.3 and the continuous and d-periodic function $\psi_2^{(b)}$ is from Theorem 4.1 with $f^{(b)}$ from above and corresponding $g^{(b)}$. $\psi_2^{(b)}$ has Fourier series

$$\frac{1}{b!} \sum_{m \neq 0} \frac{\Gamma(b - 2\pi\,\mathrm{i}\, m/d)}{1 + 2\pi\,\mathrm{i}\, m/d} e^{2\pi\,\mathrm{i}\, mt/d}. \tag{4.34}$$

The same holds in the Poisson model with $\widetilde{W}_\lambda^{(b)}$ and λ instead of $W_n^{(b)}$ and n.

Proof. Again, Theorem 4.1 is the key tool in the Poisson model. $f^{(b)}$ satisfies all conditions required for Theorem 4.1 with $\delta = 1$. Again following [Jan12a, Section 6], we note that

$$(f^{(b)})'(x) = e^{-x} \frac{x^b}{b!} = \mathbb{P}(\Pi(x) = b).$$

Integration by parts yields

$$\int_0^\infty f^{(b)}(x)x^{-2}\,\mathrm{d}x = \frac{1}{b}$$

and thus the asymptotics in the non-arithmetic case follow. The d-arithmetic case is treated similarly as in the case $b = 1$.

Obviously, $\mathbb{E}_i W_n^{(b)}$ increases in n, so the depoissonization follows from Lemma A.1, where (A.2) applies with $C_1 = 0$, $C_2 = 1/\mu$ and either $\psi \equiv 1/b$ (non-arithmetic case), or $\psi(t) = 1/b + \sum_{j\in\mathcal{S}} \pi_j \psi_2^{(b)}(t - \beta_{ij})$ (d-arithmetic case). □

External nodes: Let $Z_n^{(b)}$ be the (random) number of external nodes of $\mathrm{Trie}^{(b)}(\mathcal{M}_n)$. Now, $Z_n^{(b)} = \sum_{l=1}^b Z_{l,n}^{(b)}$ is composed of the numbers $Z_{l,n}^{(b)}$ of nodes which store exactly l strings. Let $k \geq 1$. Then $\alpha\alpha_k := \alpha_1 \cdots \alpha_{k-1}\alpha_k$ is an external node which stores exactly l strings iff l strings start with $\alpha\alpha_k$ and at least $b - l + 1$ further strings start with α but not with $\alpha\alpha_k$. The last part of the condition is necessary for the existence of $\alpha\alpha_k$ as a node in the b-trie in the first place. As the root is an external node iff $n \leq b$, we shall assume $n > b$ from now on. Then $\mathrm{Trie}^{(b)}(\mathcal{M}_n)$ is guaranteed to be bigger than just the root. Our characterization leads to

$$Z_{l,n}^{(b)} = \sum_{\alpha\in\mathcal{S}^*}\left(\sum_{a\in\mathcal{S}} \mathbb{1}_{\{N_n(\alpha a)=l,\ N_n(\{\alpha c, c\neq a\})\geq b-l+1\}}\right).$$

Recall that $\mathcal{S}_j^*$ is the set of nodes that end with j and set $\widetilde{Z}_{l,\lambda}^{(b)} := Z_{l,\Pi(\lambda)}^{(b)}$. Due to the independence properties of Poisson point processes and with $l > 0$ and $b - l + 1 > 0$, we get

$$\begin{aligned}
\mathbb{E}_i \widetilde{Z}_{l,\lambda}^{(b)} &= \sum_{\alpha\in\mathcal{S}^*} \mathbb{1}_{\{P_i(\alpha)>0\}}\left[\sum_{a\in\mathcal{S}} \mathbb{P}_i\left(\widetilde{N}_\lambda(\alpha a) = l\right)\mathbb{P}_i\left(\widetilde{N}_\lambda(\{\alpha c, c\neq a\}) \geq b - l + 1\right)\right]\\
&= \sum_{j\in\mathcal{S}}\sum_{\alpha\in\mathcal{S}_j^*} \mathbb{1}_{\{P_i(\alpha)>0\}}\left[\sum_{a\in\mathcal{S}} \mathbb{1}_{\{p_{j,a}\in(0,1)\}}\mathbb{P}\Big(\Pi(\lambda P_i(\alpha)p_{j,a}) = l\Big)\right.\\
&\quad \left.\cdot\,\mathbb{P}\Big(\Pi(\lambda P_i(\alpha)(1-p_{j,a})) \geq b - l + 1\Big)\right]\\
&\quad + \sum_{a\in\mathcal{S}} \mathbb{1}_{\{p_{i,a}\in(0,1)\}}\mathbb{P}\Big(\Pi(\lambda p_{i,a}) = l\Big)\cdot\mathbb{P}\Big(\Pi(\lambda(1-p_{i,a})) \geq b - l + 1\Big)\\
&= \sum_{j\in\mathcal{S}}\left(\sum_{\alpha\in\mathcal{S}_j^*} f_l^j(\lambda P_i(\alpha))\mathbb{1}_{\{P_i(\alpha)>0\}}\right) + f_l^i(\lambda),
\end{aligned}$$

where for $x > 0$

$$\begin{aligned}
f_l^j(x) &:= \sum_{a\in\mathcal{S}} \mathbb{1}_{\{p_{j,a}\in(0,1)\}}\mathbb{P}\left(\Pi(xp_{j,a}) = l\right)\cdot\mathbb{P}\left(\Pi(x(1-p_{j,a})) \geq b - l + 1\right)\\
&= \sum_{a\notin\mathbf{0}_j} \mathbb{1}_{\{p_{j,a}<1\}}\mathbb{P}\left(\Pi(xp_{j,a}) = l\right)\cdot\mathbb{P}\left(\Pi(x(1-p_{j,a})) \geq b - l + 1\right).
\end{aligned}$$

We keep in mind that f_l^j still depends on b and simplify the above expression to

$$\mathbb{E}_i \widetilde{Z}_{l,\lambda}^{(b)} = \sum_{j\in\mathcal{S}}\left(\sum_{\alpha\in\mathcal{S}^*} f_l^j(\lambda P_i(\alpha j))\mathbb{1}_{\{P_i(\alpha j)>0\}}\right) + f_l^i(\lambda).$$

We note substantial differences in the asymptotics of the harmonic sum depending on the transition probabilities in row j: If $j \in \mathbf{1}$, then $f_l^j \equiv 0$ and the corresponding harmonic sum vanishes. Otherwise, f_l^j satisfies (4.2) and we can use Theorem 4.5. Note that only summands with $p_{j,a} > 0$ contribute.

Theorem 4.19. *(a) If $(\xi_n, S_n)_{n\geq 0}$ is non-arithmetic, then for $l = 1, \dots, b$ and $n \to \infty$,*

$$\frac{\mathbb{E}_i Z_{l,n}^{(b)}}{n} \to \frac{\sum_{j \notin \mathbf{1}} \pi_j \cdot c_l^j}{\mu} =: \frac{\Pi_l^{(b)}}{\mu}, \tag{4.35}$$

where the c_l^j depend on b and are given by

$$c_l^j = \begin{cases} \sum_{a \notin \mathbf{0}_j} p_{j,a} \log(1/p_{j,a}) - \sum_{k=1}^{b-1} \frac{1}{k} \sum_{a \notin \mathbf{0}_j} p_{j,a}(1-p_{j,a})^k, & l = 1 \\ \frac{1}{l(l-1)} - \sum_{k=0}^{b-l} \left(\frac{(l+k-2)!}{l!k!} \sum_{a \notin \mathbf{0}_j} p_{j,a}^l (1-p_{j,a})^k \right), & 2 \leq l \leq b \end{cases} \tag{4.36}$$

for $j \in \mathcal{S}$ and $l = 1, \dots, b$.

(b) If $(\xi_n, S_n)_{n\geq 0}$ is d-arithmetic with shift function β, then for $l = 1, \dots, b$ and $n \to \infty$,

$$\frac{\mathbb{E}_i Z_{l,n}^{(b)}}{n} = \frac{\Pi_l^{(b)}}{\mu} + \frac{1}{\mu} \sum_{j \notin \mathbf{1}} \pi_j \psi_l^j (\log n - \beta_{ij}) + o(1), \tag{4.37}$$

where β_{ij} are from Theorem 3.3 and the continuous and d-periodic functions ψ_l^j are from Theorem 4.5 with f_l^j from above and corresponding g_l^j. The ψ_l^j depend on b and have Fourier series

$$\sum_{m \neq 0} \hat{g}_l^j(-2\pi m/d) e^{2\pi \mathrm{i}\, mt/d}, \tag{4.38}$$

respectively. $\hat{g}_l^j$ is defined in (4.39).

The same holds in the Poisson model with $\widetilde{Z}_{l,\lambda}^{(b)}$ and λ instead of $Z_{l,n}^{(b)}$ and n.

Proof. Here we use Theorem 4.5, for the first time, on each summand of the foregoing expression of $\mathbb{E}_i \widetilde{Z}_{l,\lambda}^{(b)}$.

We only consider $j \notin \mathbf{1}$. It is not hard to show that f_l^j satisfies the conditions for Theorem 4.5 with $\delta = 1$ for each combination of l and j (note that some summands may even vanish). An application of this theorem then yields in the non-arithmetic case, as $\lambda \to \infty$,

$$\frac{\sum_{\alpha \in \mathcal{S}^*} f_l^j(\lambda P_i(\alpha j)) \mathbb{1}_{\{P_i(\alpha j) > 0\}}}{\lambda} = \frac{\pi_j}{\mu} \int_0^\infty f_l^j(x) x^{-2}\, \mathrm{d}x + o(1) = \frac{\pi_j c_l^j}{\mu} + o(1).$$

The components of the last integral have been calculated in [Jan12a, Section 6] and amount to

$$\begin{aligned} c_l^j &= \sum_{a \notin \mathbf{0}_j} \int_0^\infty \mathbb{P}\left(\Pi(x p_{j,a}) = l\right) \mathbb{P}\left(\Pi(x(1-p_{j,a})) \geq b - l + 1\right) x^{-2}\, \mathrm{d}x \\ &= \begin{cases} \sum_{a \notin \mathbf{0}_j} p_{j,a} \log(1/p_{j,a}) - \sum_{k=1}^{b-1} \frac{1}{k} \sum_{a \notin \mathbf{0}_j} p_{j,a}(1-p_{j,a})^k, & l = 1 \\ \frac{1}{l(l-1)} - \sum_{k=0}^{b-l} \left(\frac{(l+k-2)!}{l!k!} \sum_{a \notin \mathbf{0}_j} p_{j,a}^l (1-p_{j,a})^k \right), & 2 \leq l \leq b. \end{cases} \end{aligned}$$

For $2 \le l \le b$, the calculation is straightforward using the explicit form of f_l^j and the aforementioned standard Gamma integral properties. The case $l = 1$ is slightly more involved and requires the knowledge of

$$\int_0^\infty (e^{-ax} - e^{-bx})x^{-1}\,\mathrm{d}x = \log(b/a)$$

for $a, b > 0$ which we cite from [Jan13, (25)].

The additional summand $f_l^i(\lambda)$, which originates from considering the root, is of order $o(\lambda)$, as $\lambda \to \infty$, since $f_l^i(\lambda) = \mathcal{O}(1)$, as $\lambda \to \infty$. So in the non-arithmetic case we obtain, as $\lambda \to \infty$,

$$\frac{\mathbb{E}_i \widetilde{Z}_{l,n}^{(b)}}{\lambda} \to \frac{\sum_{j \notin \mathbf{1}} \pi_j \cdot c_l^j}{\mu}.$$

In the d-arithmetic case, Theorem 4.5 yields

$$\frac{\sum_{\beta \in \mathcal{S}^*} f_l^j(\lambda P_i(\alpha j))}{\lambda} = \frac{\pi_j c_l^j}{\mu} + \frac{\pi_j}{\mu} \psi_l^j(\log \lambda - \beta_{ij}) + o(1),$$

as $\lambda \to \infty$, where as always $\psi_l^j(t)$ has Fourier series

$$\sum_{m \neq 0} \hat{g}_l^j(-2\pi m/d) e^{2\pi\,\mathrm{i}\, mt/d}.$$

The $\hat{g}_l^j(s)$ have been calculated in [Jan12a, Section 6] as well. For $(l, s) \neq (1, 0)$ they are

$$\hat{g}_l^j(s) = \frac{\Gamma(l - 1 + \mathrm{i}\, s)}{l!} \sum_{a \notin \mathbf{0}_j} p_{j,a}^{1 - \mathrm{i}\, s} - \sum_{k=0}^{b-l} \left(\frac{\Gamma(l + k - 1 + \mathrm{i}\, s)}{l!k!} \sum_{a \in \mathbf{0}_j} p_{j,a}^l (1 - p_{j,a})^k \right). \tag{4.39}$$

Now, depoissonization is needed for a characteristic parameter which is not increasing. The idea is to show that

$$|\mathbb{E}_i Z_{l,n}^{(b)} - \mathbb{E}_i \widetilde{Z}_{l,n}^{(b)}| = \mathcal{O}(\sqrt{n}),$$

as $n \to \infty$, by showing that

$$|Z_{l,n}^{(b)} - \widetilde{Z}_{l,n}^{(b)}| \le (\#\mathcal{S} + 1)|\,\Pi(n) - n|.$$

One can approach this by starting with a trie constructed from $\min\{n, \Pi(n)\}$ strings and successively adding

$$\max\{n, \Pi(n)\} - \min\{n, \Pi(n)\} = |\,\Pi(n) - n|$$

further strings and controlling for the changes in the external nodes which are caused by this addition. The proof is exactly the same as the proof of [Jan12a, Theorem 6.1], so we omit it here. □

Remark 4.20. Since $n = \sum_{l=1}^b l \cdot Z_{l,n}^{(b)}$ is the number of stored strings,

$$\sum_{l=1}^b l \cdot \Pi_l^{(b)} = \lim_{n\to\infty} \frac{\mu}{n} \sum_{l=1}^b l \cdot \mathbb{E}_i Z_{l,n}^{(b)} = \lim_{n\to\infty} \frac{\mu}{n} \mathbb{E}_i \left(\sum_{l=1}^b l \cdot Z_{l,n}^{(b)} \right) = \mu.$$

We conclude for the expected number of nodes in $\mathrm{Trie}^{(b)}(\mathcal{M}_n)$:

Remark 4.21. The expected number of nodes in $\mathrm{Trie}^{(b)}(\mathcal{M}_n)$ equals

$$\mathbb{E}_i W_n^{(b)} + \mathbb{E}_i Z_n^{(b)} = \mathbb{E}_i W_n^{(b)} + \sum_{l=1}^b \mathbb{E}_i Z_{l,n}^{(b)}$$

with asymptotics that follow from Theorems 4.18 and 4.19.

4.5.3. Expected size of a PATRICIA-trie

In [Jan12a, Section 7], Janson examined whether there is a noticeable asymptotic difference between the expected sizes of the trie and the PAT-trie in the i.i.d. setting. We proceed similarly in the Markov model: Since the leaves of both trees coincide, we only need to study the number of internal nodes of the PAT-trie to answer this question. First, we characterize these nodes:

A node α is an internal node of $\mathrm{Trie}^P(\mathcal{M}_n)$ iff there exist $b \neq c \in \mathcal{S}$ such that at least one string starts with αb and at least one string starts with αc. If we denote by W_n^P the number of those nodes, then

$$W_n^P = \sum_{\alpha \in \mathcal{S}^*} \mathbb{1}_{\bigcup_{b \neq c \in \mathcal{S}} \{N_n(\alpha b) \geq 1,\ N_n(\alpha c) \geq 1\}}.$$

In the Poisson model, taking complements yields

$$\begin{aligned}
\mathbb{E}_i \widetilde{W}_\lambda^P &= \sum_{\alpha \in \mathcal{S}^*} \mathbb{1}_{\{P_i(\alpha)>0\}} \left[1 - \mathbb{P}_i\left(\widetilde{N}_\lambda(\alpha) = 0\right) - \sum_{b \in \mathcal{S}} \mathbb{P}_i\left(\widetilde{N}_\lambda(\alpha b) \geq 1,\ \widetilde{N}_\lambda(\alpha c) = 0, c \neq b\right)\right] \\
&= \sum_{j \in \mathcal{S}} \sum_{\alpha \in \mathcal{S}_j^*} \mathbb{1}_{\{P_i(\alpha)>0\}} \Bigg[1 - \mathbb{P}\left(\Pi(\lambda P_i(\alpha)) = 0\right) \\
&\quad - \sum_{b \in \mathcal{S}} \mathbb{1}_{\{p_{j,b}>0\}} \mathbb{P}\left(\Pi(\lambda P_i(\alpha) p_{j,b}) \geq 1\right) \left[\mathbb{1}_{\{p_{j,b}<1\}} \mathbb{P}\left(\Pi(\lambda P_i(\alpha)(1-p_{j,b})) = 0\right) + \mathbb{1}_{\{p_{j,b}=1\}}\right]\Bigg] \\
&\quad + 1 - \mathbb{P}\left(\Pi(\lambda) = 0\right) \\
&\quad - \sum_{b \in \mathcal{S}} \mathbb{1}_{\{p_{i,b}>0\}} \mathbb{P}\left(\Pi(\lambda p_{i,b}) \geq 1\right) \left[\mathbb{1}_{\{p_{i,b}<1\}} \mathbb{P}\left(\Pi(\lambda(1-p_{i,b})) = 0\right) + \mathbb{1}_{\{p_{i,b}=1\}}\right] \\
&= \sum_{j \in \mathcal{S}} \left(\sum_{\alpha \in \mathcal{S}^*} f_j^P(\lambda P_i(\alpha j)) \mathbb{1}_{\{P_i(\alpha j)>0\}} \right) + f_i^P(\lambda)
\end{aligned}$$

for f_j^P defined on $(0, \infty)$ by

$$\begin{aligned}
f_j^P(x) &:= 1 - e^{-x} - \sum_{b \in \mathcal{S}} \mathbb{1}_{\{p_{j,b}>0\}} (1 - e^{-x p_{j,b}}) \cdot \left(\mathbb{1}_{\{p_{j,b}<1\}} e^{-x(1-p_{j,b})} + \mathbb{1}_{\{p_{j,b}=1\}} \right) \\
&= 1 - e^{-x} - \sum_{b \notin \mathbf{0}_j} (e^{-x(1-p_{j,b})} - e^{-x}).
\end{aligned}$$

Again, we observe different asymptotics of the corresponding harmonic sum $F_i^j(\lambda)$ (with $f = f_j^P$) depending on whether there is a b' with $p_{j,b'} = 1$ or not: If $j \in \mathbf{1}$, then

$$f_j^P(x) = 1 - e^{-x} - (1 - e^{-x}) = 0.$$

Those j obviously lead to a vanishing harmonic sum. Otherwise, if all $p_{j,b} < 1$, then f_j^P satisfies (4.2) and we can use Theorem 4.5.

Theorem 4.22. *(a) If $(\xi_n, S_n)_{n \geq 0}$ is non-arithmetic, then, as $n \to \infty$,*

$$\frac{\mathbb{E}_i W_n^P}{n} = \frac{1}{\mu} \sum_{j \notin \mathbf{1}} \pi_j \sum_{b \notin \mathbf{0}_j} -(1 - p_{j,b}) \log(1 - p_{j,b}) + o(1). \tag{4.40}$$

(b) If $(\xi_n, S_n)_{n\geq 0}$ is d-arithmetic with shift function β, then, as $n \to \infty$,

$$\frac{\mathbb{E}_i W_n^P}{n} = \frac{1}{\mu}\sum_{j\notin\mathbf{1}}\pi_j\sum_{b\notin\mathbf{0}_j} -(1-p_{j,b})\log(1-p_{j,b}) + \frac{1}{\mu}\sum_{j\notin\mathbf{1}}\pi_j\psi_j^P(\log n - \beta_{ij}) + o(1), \quad (4.41)$$

where β_{ij} are from Theorem 3.3 and the continuous and d-periodic functions ψ_j^P are from Theorem 4.5 with f_j^P from above and corresponding g_j^P. ψ_j^P has Fourier series

$$\sum_{m\neq 0}\widehat{\psi}_j^P(m)e^{2\pi\,\mathrm{i}\,mt/d} \quad (4.42)$$

with coefficients

$$\widehat{\psi}_j^P(m) = \Gamma(-1-2\pi\,\mathrm{i}\,m/d)\sum_{b\notin\mathbf{0}_j}\Big((1-p_{j,b}) - (1-p_{j,b})^{1+2\pi\,\mathrm{i}\,m/d}\Big). \quad (4.43)$$

The same holds in the Poisson model with $\widetilde{W}_\lambda^P$ and λ instead of W_n^P and n.

Proof. It is shown easily that f_j^P satisfies the conditions for Theorem 4.5 with $\delta = 1$ (and hence $f_j^P(\lambda) = o(\lambda)$ as $\lambda\to\infty$). We only consider $j\notin\mathbf{1}$. Via integration by parts,

$$\int_0^\infty f_j^P(x)x^{-2}\,\mathrm{d}x = \int_0^\infty (f_j^P)'(x)x^{-1}\,\mathrm{d}x$$

with

$$\begin{aligned}(f_j^P)'(x) &= e^{-x} + \sum_{b\notin\mathbf{0}_j}\Big((1-p_{j,b})e^{-x(1-p_{j,b})} - e^{-x}\Big)\\ &= \sum_{b\notin\mathbf{0}_j}\Big((1-p_{j,b})e^{-x(1-p_{j,b})} - e^{-x} + p_{j,b}e^{-x}\Big)\\ &= \sum_{b\notin\mathbf{0}_j}(1-p_{j,b})(e^{-x(1-p_{j,b})} - e^{-x}),\end{aligned}$$

such that

$$\begin{aligned}\int_0^\infty f_j^P(x)x^{-2}\,\mathrm{d}x &= \sum_{b\notin\mathbf{0}_j}(1-p_{j,b})\int_0^\infty(e^{-x(1-p_{j,b})} - e^{-x})x^{-1}\,\mathrm{d}x\\ &= \sum_{b\notin\mathbf{0}_j} -(1-p_{j,b})\log(1-p_{j,b}),\end{aligned}$$

cf. [Jan13, (25)]. In the non-arithmetic case, this yields the asymptotics

$$\frac{\mathbb{E}_i\widetilde{W}_\lambda^P}{\lambda} = \frac{1}{\mu}\sum_{j\notin\mathbf{1}}\pi_j\sum_{b\notin\mathbf{0}_j} -(1-p_{j,b})\log(1-p_{j,b}) + o(1).$$

In the d-arithmetic case, the Fourier coefficients are calculated via

$$\begin{aligned}\hat{g}_j^P(s) &= \int_0^\infty f_j^P(x)x^{-2+\mathrm{i}\,s}\,\mathrm{d}x = \frac{1}{1-\mathrm{i}\,s}\int_0^\infty(f_j^P)'(x)x^{-1+\mathrm{i}\,s}\,\mathrm{d}x\\ &= \frac{1}{\mathrm{i}\,s(\mathrm{i}\,s-1)}\int_0^\infty(f_j^P)''(x)x^{\mathrm{i}\,s}\,\mathrm{d}x\end{aligned}$$

for $s \neq 0$. Since

$$(f_j^P)''(x) = \sum_{b \notin \mathbf{0}_j} (1 - p_{j,b}) e^{-x} - (1 - p_{j,b})^2 e^{-x(1-p_{j,b})},$$

we arrive at

$$\begin{aligned}
\hat{g}_j^P(s) &= \frac{1}{\mathrm{i}\, s(\mathrm{i}\, s - 1)} \sum_{b \notin \mathbf{0}_j} \left((1 - p_{j,b}) \int_0^\infty x^{(1+\mathrm{i}\, s)-1} e^{-x} \,\mathrm{d}x - (1 - p_{j,b})^2 \int_0^\infty x^{(1+\mathrm{i}\, s)-1} e^{-x(1-p_{j,b})} \,\mathrm{d}x \right) \\
&= \frac{\Gamma(1 + \mathrm{i}\, s)}{\mathrm{i}\, s(\mathrm{i}\, s - 1)} \sum_{b \notin \mathbf{0}_j} \left((1 - p_{j,b}) - (1 - p_{j,b})^{1-\mathrm{i}\, s} \right) \\
&= \Gamma(-1 + \mathrm{i}\, s) \sum_{b \notin \mathbf{0}_j} \left((1 - p_{j,b}) - (1 - p_{j,b})^{1-\mathrm{i}\, s} \right)
\end{aligned}$$

with a simple substitution. Altogether, this yields the asymptotics

$$\frac{\mathbb{E}_i \widetilde{W}_\lambda^P}{\lambda} = \frac{1}{\mu} \sum_{j \notin \mathbf{1}} \pi_j \sum_{b \notin \mathbf{0}_j} -(1 - p_{j,b}) \log(1 - p_{j,b}) + \frac{1}{\mu} \sum_{j \notin \mathbf{1}} \pi_j \psi_j^P(\log \lambda - \beta_{ij}) + o(1),$$

$\lambda \to \infty$, in the d-arithmetic case. As usual, ψ_j^P is the continuous and d-periodic function from Theorem 4.5 with f_j^P from above and corresponding g_j^P. It has the Fourier series

$$\sum_{m \neq 0} \hat{g}_j^P(-2\pi m/d) e^{2\pi \mathrm{i}\, mt/d}.$$

To depoissonize, note that $\mathbb{E}_i W_n^P$ is monotone. The rest follows from Lemma A.1, where (A.2) applies with $C_1 = 0$, $C_2 = 1/\mu$ and either

$$\psi \equiv \sum_{j \notin \mathbf{1}} \pi_j \sum_{b \notin \mathbf{0}_j} -(1 - p_{j,b}) \log(1 - p_{j,b})$$

in the non-arithmetic case, or

$$\psi(t) = \sum_{j \notin \mathbf{1}} \pi_j \sum_{b \notin \mathbf{0}_j} -(1 - p_{j,b}) \log(1 - p_{j,b}) + \sum_{j \notin \mathbf{1}} \pi_j \psi_j^P(t - \beta_{ij})$$

in the d-arithmetic case. □

Remark 4.23. We get asymptotics for the expected number of nodes of $\mathrm{Trie}^P(\mathcal{M}_n)$ by $\mathbb{E} W_n^P + n$.

The binary setup allows for an even stronger result on the limiting behavior.

Theorem 4.24. *In the binary setup, we have*

$$W_n^P = n - 1, \tag{4.44}$$

so the number of all nodes in $\mathrm{Trie}^P(\mathcal{M}_n)$ *is*

$$W_n^P + n = 2n - 1. \tag{4.45}$$

Proof. The internal strings in the Pat-trie are those nodes which have exactly two children since the nodes with only one child are eliminated. Additionally we know that $\mathrm{Trie}^P(\mathcal{M}_n)$ has n external nodes, so $\mathrm{Trie}^P(\mathcal{M}_n)$ is a tree with n leaves where every internal node has exactly two children. Hence, it must have $n - 1$ internal nodes. □

In [Jan12a, Section 7], Janson derived the latter result up to the first order with his theorem, by proving that the oscillatory terms cancel out in the arithmetic case. This must remain true in the Markov setting not to violate Theorem 4.24. Indeed, Expressions (4.40) and (4.41) from the general-alphabet setting in Theorem 4.22 simplify a lot and the fluctuating function appearing in the d-arithmetic case vanishes. We prove this in the following result which can be seen as a corollary of Theorem 4.22 (and is of course weaker than Theorem 4.24).

Corollary 4.25. *In the binary setup, as $n \to \infty$,*

$$\frac{\mathbb{E}_i W_n^P}{n} \to 1. \tag{4.46}$$

Proof. Note that for $j \notin \mathbf{1}$

$$\sum_{b \notin \mathbf{0}_j} -(1-p_{j,b}) \log(1-p_{j,b}) = \sum_{b \in \{0,1\}} -(1-p_{j,b}) \log(1-p_{j,b}) = H_j$$

and $H_j = 0$ for $j \in \mathbf{1}$, so

$$\frac{1}{\mu} \sum_{j \notin \mathbf{1}} \pi_j \sum_{b \notin \mathbf{0}_j} -(1-p_{j,b}) \log(1-p_{j,b}) = \frac{1}{\mu} \sum_{j \in \{0,1\}} \pi_j H_j = \frac{1}{\mu} \mathbb{E}_\pi X_1 = 1.$$

In the d-arithmetic case, we note for $j \notin \mathbf{1}$ that

$$\begin{aligned}\sum_{b \notin \mathbf{0}_j} \left((1-p_{j,b}) - (1-p_{j,b})^{1+2\pi\,\mathrm{i}\, m/d}\right) &= \sum_{b \in \{0,1\}} \left(p_{j,b} - p_{j,b}^{1+2\pi\,\mathrm{i}\, m/d}\right) \\ &= 1 - p_{j,0}^{1+2\pi\,\mathrm{i}\, m/d} - p_{j,1}^{1+2\pi\,\mathrm{i}\, m/d}.\end{aligned}$$

It suffices to show that

$$\sum_{j \notin \mathbf{1}} \pi_j \sum_{m \neq 0} \widehat{\psi}_j^P(m) e^{2\pi\,\mathrm{i}\, m(t-\beta_{ij})/d} = \sum_{m \neq 0} \sum_{j \notin \mathbf{1}} \pi_j \widehat{\psi}_j^P(m) e^{2\pi\,\mathrm{i}\, m(t-\beta_{ij})/d} = 0$$

and hence to show that the inner sum vanishes for every $i \in \mathcal{S}$. Now, $|\mathbf{1}| = 1$, $p_{1,0} = 1$, say, is the easy case. It is easy to show analogously to Section 2.8 that in this case the MRW is d-arithmetic iff $\log p_{0,0} / \log p_{0,1} \in \mathbb{Q}$, and then both $\log p_{0,0}$ and $\log p_{0,1}$ are integer multiples of d. Hence,

$$1 - p_{0,0}^{1+2\pi\,\mathrm{i}\, m/d} - p_{0,1}^{1+2\pi\,\mathrm{i}\, m/d} = 1 - p_{0,0} - p_{1,1} = 0.$$

If all $p_{i,j} \in (0,1)$, then consider $i = 0$ and recall (2.9). The conclusion from above is no longer valid, since

$$1 - p_{0,0}^{1+2\pi\,\mathrm{i}\, m/d} - p_{0,1}^{1+2\pi\,\mathrm{i}\, m/d} = 1 - p_{0,0} - p_{0,1} e^{-2\pi\,\mathrm{i}\, m\beta_{01}/d} = p_{0,1}(1 - e^{-2\pi\,\mathrm{i}\, m\beta_{01}/d})$$

and

$$1 - p_{1,0}^{1+2\pi\,\mathrm{i}\, m/d} - p_{1,1}^{1+2\pi\,\mathrm{i}\, m/d} = p_{1,0}(1 - e^{2\pi\,\mathrm{i}\, m\beta_{01}/d})$$

do not vanish alone. However, using $\pi_1 p_{1,0} = \pi_0 - \pi_0 p_{0,0}$ by stationarity,

$$\begin{aligned}&\sum_{j \in \{0,1\}} \pi_j \widehat{\psi}_j^P(m) e^{2\pi\,\mathrm{i}\, m(t-\beta_{0j})/d} \\ &= \pi_0 \Gamma(-1 - 2\pi\,\mathrm{i}\, m/d) p_{0,1} (1 - e^{-2\pi\,\mathrm{i}\, m\beta_{01}/d}) e^{2\pi\,\mathrm{i}\, mt/d}\end{aligned}$$

$$
\begin{aligned}
&+\pi_1\Gamma(-1-2\pi\,\mathrm{i}\,m/d)p_{1,0}(1-e^{2\pi\,\mathrm{i}\,m\beta_{01}/d})e^{2\pi\,\mathrm{i}\,m(t-\beta_{01})/d}\\
&=(\pi_0 p_{0,1}(1-e^{-2\pi\,\mathrm{i}\,m\beta_{01}/d})+\pi_0(e^{-2\pi\,\mathrm{i}\,m\beta_{01}/d}-1)-\pi_0 p_{0,0}(e^{-2\pi\,\mathrm{i}\,m\beta_{01}/d}-1))\\
&\quad\cdot\Gamma(-1-2\pi\,\mathrm{i}\,m/d)e^{2\pi\,\mathrm{i}\,mt/d}\\
&=(\pi_0(1-e^{-2\pi\,\mathrm{i}\,m\beta_{01}/d})+\pi_0(e^{-2\pi\,\mathrm{i}\,m\beta_{01}/d}-1)\cdot\Gamma(-1-2\pi\,\mathrm{i}\,m/d)e^{2\pi\,\mathrm{i}\,mt/d}=0.
\end{aligned}
$$

The case $i = 1$ is similar. □

Remark 4.26. As Janson points out for the binary i.i.d. case, the number of internal nodes in the Pat-trie compared to the trie is reduced by a factor H which is the (i.i.d.) source entropy and is at most $\log 2$, so the reduction is $\geq 1 - \log 2 \approx 0.307$. In the binary Markov case with non-vanishing transitions this factor becomes $\mu = \pi_0 H_0 + \pi_1 H_1$ which, as a function of $p_{0,1}$ and $p_{1,0}$, is again maximized at $(\frac{1}{2}, \frac{1}{2})$, so $\mu \leq \log 2$, and this is as specific as we can generally get.

However, if the source is partly degenerate, say $p_{0,0}, p_{0,1} \in (0,1)$ and $p_{1,0} = 1 = 1 - p_{1,1}$, then the reduction is $1 - C$ with

$$C := \frac{1}{1+p_{0,1}}\left(-p_{0,1}\log p_{0,1} - (1-p_{0,1})\log(1-p_{0,1})\right)$$

and approximately $C \leq 0.481$, maximized by $p_{0,1} = \frac{3}{2} - \frac{\sqrt{5}}{2}$. So the reduction is at least 0.519.

In general the reduction is $1 - C$ with

$$C := \sum_{j\notin\mathbf{1}} \pi_j \sum_{b\notin\mathbf{0}_j} -(1-p_{j,b})\log(1-p_{j,b}) \leq \sum_{j\notin\mathbf{1}} \pi_j(|\mathbf{0}_j^c|-1)\left(-\log\left(\frac{|\mathbf{0}_j^c|-1}{|\mathbf{0}_j^c|}\right)\right),$$

so it depends on the imbalance within each row j.

Let $\mathrm{Trie}^{(b),P}(\mathcal{M}_n)$ denote the Pat-b-trie which is formed by elimination of all nodes that have exactly one child in the b-trie. In the binary setup, it is easy to state an analogue of Theorem 4.24 for Pat-b-Tries.

Theorem 4.27. *In the binary setup, we have* $W_n^{(b),P} = \sum_{l=1}^b Z_{l,n}^{(b)} - 1$ *with* $Z_{l,n}^{(b)}$ *from Subsection 4.5.2. Then the following cases occur:*

(a) If $(\xi_n, S_n)_{n\geq 0}$ *is non-arithmetic, then, as* $n \to \infty$,

$$\frac{\mathbb{E}_i W_n^{(b),P}}{n} \to \frac{1}{\mu}\sum_{l=1}^b \Pi_l^{(b)} \tag{4.47}$$

with $\Pi_l^{(b)}$ *from Theorem 4.19.*

(b) If $(\xi_n, S_n)_{n\geq 0}$ *is d-arithmetic with shift function* β, *then, as* $n \to \infty$,

$$\frac{\mathbb{E}_i W_n^{(b),P}}{n} = \frac{1}{\mu}\sum_{l=1}^b \Pi_l^{(b)} + \frac{1}{\mu}\sum_{l=1}^b\sum_{j\notin\mathbf{1}} \pi_j \psi_l^j(\log n - \beta_{ij}) + o(1), \tag{4.48}$$

where β_{ij} *are from Theorem 3.3 and the continuous and d-periodic (and b-dependent) functions* ψ_l^j *are from Theorem 4.19.*

Proof. It is obvious that the Pat-b-trie and the b-trie have the same leaves, so $Z_{l,n}^{(b),P} = Z_{l,n}^{(b)}$ with the obvious definition of $Z_{l,n}^{(b),P}$. So $\mathrm{Trie}^{(b),P}(\mathcal{M}_n)$ has $\sum_{l=1}^b Z_{l,n}^{(b)}$ external nodes and each internal node has exactly two children. We conclude that $W_n^{(b),P} = \sum_{l=1}^b Z_{l,n}^{(b)} - 1$. Everything else follows immediately from Theorem 4.19. □

4.5.4. Expected depth in a PATRICIA-trie

In this subsection, we give an asymptotic expansion of the expected number of nodes eliminated in the path from the root to the node in which a selected string Ξ is stored and, in analogy to Theorem 3.3, an asymptotic expansion of the expected depth of this string in the Pat-trie. Here, too, we wonder whether there is a noticeable asymptotic effect of the compression mechanism.

We select the generic string Ξ as the first of the sequence $\Xi, \Xi^{(1)}, \Xi^{(2)}, \ldots$. Let D_n be the depth of Ξ from Section 2.3 (in the trie constructed from $\Xi, \Xi^{(1)}, \ldots, \Xi^{(n-1)}$) and let D_n^P be analogously defined as the path length from the root to the node in which Ξ is stored in the Pat-trie. We set $\Delta D_n := D_n - D_n^P \geq 0$ as the number of nodes on the path to Ξ which are eliminated during the transition to the Pat-trie.

These nodes are exactly the internal nodes α which have exactly one child (and lie on the path to Ξ). This occurs iff Ξ starts with αb for a $b \in \mathcal{S}$, at least one additional string starts with αb, too, and no other string starts with αc for $c \neq b$. Thus, we have

$$\Delta D_n = \sum_{\alpha \in \mathcal{S}^*} \left(\sum_{b \in \mathcal{S}} \mathbb{1}_{\{\Xi \succ \alpha b,\ N_{n-1}(\alpha b) \geq 1,\ N_{n-1}(\{\alpha c, c \neq b\}) = 0\}} \right).$$

Note that still $N_n(\alpha) := \sum_{k=1}^n \mathbb{1}_{\{\Xi^{(k)} \succ \alpha\}}$, ignoring Ξ. We consider $\widetilde{D}_\lambda := D_{1+\Pi(\lambda)}$, $\widetilde{D}_\lambda^P := D_{1+\Pi(\lambda)}^P$ and $\Delta\widetilde{D}_\lambda := \widetilde{D}_\lambda - \widetilde{D}_\lambda^P \geq 0$. This leads to

$$\begin{aligned}
\mathbb{E}_i \Delta\widetilde{D}_\lambda &= \sum_{\alpha \in \mathcal{S}^*} \mathbb{1}_{\{P_i(\alpha) > 0\}} \left[\sum_{b \in \mathcal{S}} \mathbb{P}_i\left(\Xi \succ \alpha b\right) \cdot \mathbb{P}_i\left(\widetilde{N}_\lambda(\alpha b) \geq 1\right) \cdot \mathbb{P}_i\left(\widetilde{N}_\lambda(\{\alpha c, c \neq b\}) = 0\right) \right] \\
&= \sum_{j \in \mathcal{S}} \sum_{\alpha \in \mathcal{S}_j^*} \mathbb{1}_{\{P_i(\alpha) > 0\}} \left[\sum_{b \in \mathcal{S}} \mathbb{1}_{\{p_{j,b} > 0\}} P_i(\alpha) p_{j,b} \cdot \mathbb{P}\Big(\Pi(\lambda P_i(\alpha) p_{j,b}) \geq 1 \Big) \right. \\
&\quad \left. \cdot \left(\mathbb{1}_{\{p_{j,b} < 1\}} \mathbb{P}\Big(\Pi(\lambda P_i(\alpha)(1 - p_{j,b})) = 0 \Big) + \mathbb{1}_{\{p_{j,b} = 1\}} \right) \right] \\
&\quad + \sum_{b \in \mathcal{S}} \mathbb{1}_{\{p_{i,b} > 0\}} p_{i,b} \cdot \mathbb{P}\Big(\Pi(\lambda p_{i,b}) \geq 1 \Big) \cdot \Big(\mathbb{1}_{\{p_{i,b} < 1\}} \mathbb{P}\Big(\Pi(\lambda(1 - p_{i,b})) = 0 \Big) + \mathbb{1}_{\{p_{i,b} = 1\}} \Big) \\
&= \sum_{j \in \mathcal{S}} \left(\frac{\sum_{\alpha \in \mathcal{S}^*} f_j^\Delta(\lambda P_i(\alpha j)) \mathbb{1}_{\{P_i(\alpha j) > 0\}}}{\lambda} \right) + \frac{f_i^\Delta(\lambda)}{\lambda},
\end{aligned}$$

for f_j^Δ defined on $(0, \infty)$ by

$$\begin{aligned}
f_j^\Delta(x) &:= \sum_{b \in \mathcal{S}} \mathbb{1}_{\{p_{j,b} > 0\}} x p_{j,b} \cdot \mathbb{P}(\Pi(x p_{j,b}) \geq 1) \cdot \Big(\mathbb{1}_{\{p_{j,b} < 1\}} \mathbb{P}(\Pi(x(1 - p_{j,b}) = 0) + \mathbb{1}_{\{p_{j,b} = 1\}} \Big) \\
&= \sum_{b \notin \mathbf{0}_j} x p_{j,b} \cdot (1 - e^{-x p_{j,b}}) \cdot e^{-x(1 - p_{j,b})}.
\end{aligned}$$

Here, the difference in the asymptotic behaviour of the f_j^Δ is slightly more involved than before: If $j \notin \mathbf{1}$, then f_j^Δ satisfies (4.2) and we can use Theorem 4.5, whereas if $j \in \mathbf{1}$, this condition is violated. Instead, (4.12) applies and we must use Theorem 4.6. We provide some intuition in Remark 4.30. The two main results of this subsection are stated below:

Theorem 4.28. *(a) If $(\xi_n, S_n)_{n\geq 0}$ is non-arithmetic, then, as $n \to \infty$,*

$$\begin{aligned}\mathbb{E}_i \Delta D_n = {} & \frac{\sum_{j\in\mathbf{1}} \pi_j}{\mu} \log n + \sum_{j\in\mathbf{1}} \left(\frac{\pi_j^2}{2\mu^2} \mathbb{E}_j S_{\sigma_1(j)}^2 - \frac{\pi_j}{\mu} \mathbb{E}_i S_{\sigma_1(j)} \mathbb{1}_{\{j\neq i\}} + \frac{\pi_j}{\mu}\gamma \right) \\ & + \frac{1}{\mu} \sum_{j\notin\mathbf{1}} \pi_j \sum_{b\notin\mathbf{0}_j} -p_{j,b} \log(1-p_{j,b}) + o(1)\end{aligned} \tag{4.49}$$

with the Euler constant γ.

(b) If $(\xi_n, S_n)_{n\geq 0}$ is d-arithmetic with shift function β, then, as $n \to \infty$,

$$\begin{aligned}\mathbb{E}_i \Delta D_n = {} & \frac{\sum_{j\in\mathbf{1}} \pi_j}{\mu} \log n + \sum_{j\in\mathbf{1}} \left(\frac{\pi_j^2}{2\mu^2} \mathbb{E}_j S_{\sigma_1(j)}^2 - \frac{\pi_j}{\mu} \mathbb{E}_i S_{\sigma_1(j)} \mathbb{1}_{\{j\neq i\}} + \frac{\pi_j}{\mu}\gamma \right) \\ & + \frac{1}{\mu} \sum_{j\notin\mathbf{1}} \pi_j \sum_{b\notin\mathbf{0}_j} -p_{j,b} \log(1-p_{j,b}) + \frac{1}{\mu} \sum_{j\in\mathbf{1}} \pi_j \psi_1(\log n - \beta_{ij}) \\ & + \frac{1}{\mu} \sum_{j\notin\mathbf{1}} \pi_j \psi_j^{\Delta}(\log n - \beta_{ij}) + o(1)\end{aligned} \tag{4.50}$$

where β_{ij} and ψ_1 are from Theorem 3.3 and the continuous and d-periodic functions ψ_j^{Δ} are from Theorem 4.5 with f_j^{Δ} from above and corresponding g_j^{Δ}. The ψ_j^{Δ} have Fourier series

$$\sum_{m\neq 0} \hat{g}_j^{\Delta}(-2\pi m/d) e^{2\pi\,\mathrm{i}\, mt/d} \tag{4.51}$$

with $\hat{g}_j^{\Delta}$ given in (4.54).

The same holds in the Poisson model with $\Delta\widetilde{D}_\lambda$ and λ instead of ΔD_n and n.

Remark 4.29. Equations (4.49) and (4.50) (and (4.52) and (4.53)) are unnecessarily complicated if all $p_{i,j} < 1$. Then $\mathbf{1} = \emptyset$ and many summands vanish.

Remark 4.30. Theorem 4.28 says that the reduction in depth of a string during transition to the Pat-trie is asymptotically negligible if no deterministic transitions happen at all, since $\mathbb{E}_i \Delta D_n = \mathcal{O}(1)$ for $n \to \infty$ (the ψ_j^{Δ} are bounded), whereas $\mathbb{E}_i D_n = \mathcal{O}(\log n)$. So, as $n \to \infty$,

$$\frac{\mathbb{E}_i \Delta D_n}{\mathbb{E}_i D_n} \to 0.$$

Then again, if $|\mathbf{1}| \geq 1$, we find

$$\frac{\mathbb{E}_i \Delta D_n}{\mathbb{E}_i D_n} \to \sum_{j\in\mathbf{1}} \pi_j,$$

arising from the fact, that every node following the letter $j \in \mathbf{1}$ is always eliminated. This differs from the i.i.d. setup, where the reduction in depth is always negligible (unless the source is wholly trivial), cf. [Jan12a, Section 7].

Remark 4.31. As opposed to the binary i.i.d. setup, where non-arithmetic and d-arithmetic asymptotic behaviour coincide and we observe no oscillations, this is not always the case in the binary Markov setup (even if all $p_{i,j} \in (0,1)$). Calculations similar to those in the proof of Corollary 4.25 lead to the result that $\frac{1}{\mu}\sum_{j\in\mathcal{S}} \pi_j \psi_j^{\Delta}(t - \beta_{ij})$ vanishes iff $p_{1,0} = p_{0,1}$.

Theorem 4.32. *(a) If $(\xi_n, S_n)_{n\geq 0}$ is non-arithmetic, then, as $n \to \infty$,*

$$\begin{aligned}\mathbb{E}_i D_n^P = \frac{\sum_{j\notin \mathbf{1}} \pi_j}{\mu} \log n + \sum_{j\notin\mathbf{1}} \left(\frac{\pi_j^2}{2\mu^2}\mathbb{E}_j S_{\sigma_1(j)}^2 - \frac{\pi_j}{\mu}\mathbb{E}_i S_{\sigma_1(j)} \mathbb{1}_{\{j\neq i\}} + \frac{\pi_j \gamma}{\mu}\right) \\ - \frac{1}{\mu}\sum_{j\notin\mathbf{1}} \pi_j \sum_{b\notin \mathbf{0}_j} -p_{j,b}\log(1-p_{j,b}) + o(1),\end{aligned} \tag{4.52}$$

with the Euler constant γ.

(b) If $(\xi_n, S_n)_{n\geq 0}$ is d-arithmetic with shift function β, then, as $n \to \infty$,

$$\begin{aligned}\mathbb{E}_i D_n^P = \frac{\sum_{j\notin \mathbf{1}} \pi_j}{\mu} \log n + \sum_{j\notin\mathbf{1}} \left(\frac{\pi_j^2}{2\mu^2}\mathbb{E}_j S_{\sigma_1(j)}^2 - \frac{\pi_j}{\mu}\mathbb{E}_i S_{\sigma_1(j)} \mathbb{1}_{\{j\neq i\}} + \frac{\pi_j \gamma}{\mu}\right) \\ - \frac{1}{\mu}\sum_{j\notin\mathbf{1}} \pi_j \sum_{b\notin \mathbf{0}_j} -p_{j,b}\log(1-p_{j,b}) + \frac{1}{\mu}\sum_{j\notin\mathbf{1}}\pi_j\psi_1(\log n - \beta_{ij}) \\ - \frac{1}{\mu}\sum_{j\notin\mathbf{1}}\pi_j\psi_j^\Delta(\log n - \beta_{ij}) + o(1),\end{aligned} \tag{4.53}$$

with continuous and d-periodic ψ_1 and ψ_j^Δ from Theorem 3.3 and Theorem 4.28, respectively (and β_{ij} from Theorem 3.3).

The same holds in the Poisson model with $\widetilde{D}_\lambda^P$ and λ instead of D_n^P and n.

The scheme for proving the two results is the following: We have seen, that it is simplest to depoissonize monotone parameters. ΔD_n is not monotone, but can be decomposed via $\Delta D_n := D_n - D_n^P$ into a difference of two monotone parts. Thus we derive results for all three parameters in the Poisson model, depoissonize the two monotone ones, and thus get the result for ΔD_n for free.

Proof of the Poisson version of Theorem 4.28. With the above explanation of the two types of asymptotic behaviour in mind, we assume $j \notin \mathbf{1}$ first. Then each f_j^Δ satisfies the conditions for Theorem 4.5 with $\delta = 1$. In the non-arithmetic case, Theorem 4.5 yields, as $\lambda \to \infty$,

$$\frac{\sum_{\alpha\in\mathcal{S}^*} f_j^\Delta(\lambda P_i(\alpha j))\mathbb{1}_{\{P_i(\alpha j)>0\}}}{\lambda} = \frac{\pi_j}{\mu}\int_0^\infty f_j^\Delta(x)x^{-2}\,\mathrm{d}x + o(1) = \frac{\pi_j}{\mu}\sum_{b\notin\mathbf{0}_j} -p_{j,b}\log(1-p_{j,b}) + o(1)$$

again using integrals of type [Jan13, (25)] as was done in [Jan12a, Section 7]. In the d-arithmetic case, we find that

$$\frac{\sum_{\alpha\in\mathcal{S}^*} f_j^\Delta(\lambda P_i(\alpha j))\mathbb{1}_{\{P_i(\alpha j)>0\}}}{\lambda} = \frac{\pi_j}{\mu}\sum_{b\notin\mathbf{0}_j} -p_{j,b}\log(1-p_{j,b}) + \frac{\pi_j}{\mu}\psi_j^\Delta(\log\lambda - \beta_{ij}) + o(1),$$

as $\lambda \to \infty$, where $\psi_j^\Delta(t)$ has Fourier series

$$\sum_{m\neq 0} \hat{g}_j^\Delta(-2\pi m/d)e^{2\pi\,\mathrm{i}\,mt/d}.$$

The coefficients $\hat{g}_j^{\Delta}(s)$ have been computed in [Jan12a, Section 7] using integrals of type [Jan13, (24)]:

$$\hat{g}_j^{\Delta}(s) = \sum_{b \notin \mathbf{0}_j} p_{j,b}((1-p_{j,b})^{-\mathrm{i}\, s} - 1)\Gamma(\mathrm{i}\, s). \tag{4.54}$$

It remains to perform an analogous procedure if $j \in \mathbf{1}$. Then $f_j^{\Delta}(x) = f_{\mathrm{L}}(x) = x - xe^{-x} = x + \mathcal{O}(1)$, $x \to \infty$, which will also occur in the analysis of the external path length in Subsection 4.5.5. This function satisfies (4.12) with $a = 1$ and $\delta = 1$ such that we have to apply Theorem 4.6 and not Theorem 4.5. As in [Jan12b, Theorem 3.1], we can show that

$$b_j = \frac{\pi_j^2}{2\mu}\mathbb{E}_j S_{\sigma_1(j)}^2 - \pi_j \mathbb{E}_i S_{\sigma_1(j)} \mathbb{1}_{\{j \neq i\}} + \pi_j \gamma$$

using $\int_0^{\infty}(e^{-x} - \mathbb{1}_{\{x<1\}})x^{-1}\,\mathrm{d}x = -\gamma$ from [Jan13, (16)], and for real u and $0 < \varepsilon < 1$

$$\int_0^{\infty} f_{\mathrm{L}}(x) x^{-2-\varepsilon+\mathrm{i}\, u}\,\mathrm{d}x = -\Gamma(-\varepsilon + \mathrm{i}\, u)$$

by [Jan13, (2)]. Hence, by (4.18) for $u \neq 0$ we have $\hat{g}_{\mathrm{L}}(u) = -\Gamma(\mathrm{i}\, u)$. This leads to

$$\frac{\sum_{\alpha \in \mathcal{S}^*} f_j^{\Delta}(\lambda P_i(\alpha j))\mathbb{1}_{\{P_i(\alpha j) > 0\}}}{\lambda} = \frac{\pi_j}{\mu}\log\lambda + \frac{b_j}{\mu} - \mathbb{1}_{\{j\}}(i) + o(1)$$

in the non-arithmetic case and

$$\frac{\sum_{\alpha \in \mathcal{S}^*} f_j^{\Delta}(\lambda P_i(\alpha j))\mathbb{1}_{\{P_i(\alpha j) > 0\}}}{\lambda} = \frac{\pi_j}{\mu}\log\lambda + \frac{b_j}{\mu} + \frac{\pi_j}{\mu}\psi_1(\log\lambda - \beta_{ij}) - \mathbb{1}_{\{j\}}(i) + o(1)$$

in the d-arithmetic case, where $\psi_1(t)$ is a continuous d-periodic function with Fourier series

$$-\sum_{k \neq 0} \Gamma(-2\pi\,\mathrm{i}\, k/d) e^{2\pi\,\mathrm{i}\, kt/d}$$

We remark that ψ_1 (or ψ_0 as it is called in the theorem) equals its Fourier series and the identically named ψ_1 from Theorem 3.3 (cf. Remark 4.3). It remains to remark that $f_i^{\Delta}(\lambda)/\lambda = \mathbb{1}_{\{i \in \mathbf{1}\}} + o(1)$.

Collecting all ingredients, we obtain, as $\lambda \to \infty$,

$$\begin{aligned}
\mathbb{E}_i \Delta \widetilde{D}_{\lambda} &= \sum_{j \notin \mathbf{1}} \left(-\frac{\pi_j}{\mu} \sum_{b \notin \mathbf{0}_j} p_{j,b} \log(1 - p_{j,b}) + o(1) \right) \\
&\quad + \sum_{j \in \mathbf{1}} \left(\frac{\pi_j}{\mu} \log\lambda + \frac{b_j}{\mu} - \mathbb{1}_{\{j=i\}} + o(1) \right) + \mathbb{1}_{\{i \in \mathbf{1}\}} + o(1) \\
&= \frac{\sum_{j \in \mathbf{1}} \pi_j}{\mu} \log\lambda + \sum_{j \in \mathbf{1}} \left(\frac{\pi_j^2}{2\mu^2}\mathbb{E}_j S_{\sigma_1(j)}^2 - \frac{\pi_j}{\mu}\mathbb{E}_i S_{\sigma_1(j)} \mathbb{1}_{\{j \neq i\}} + \frac{\pi_j}{\mu}\gamma \right) \\
&\quad + \frac{1}{\mu} \sum_{j \notin \mathbf{1}} \pi_j \sum_{b \notin \mathbf{0}_j} -p_{j,b} \log(1 - p_{j,b}) + o(1)
\end{aligned}$$

in the non-arithmetic case and

$$\mathbb{E}_i \Delta \widetilde{D}_{\lambda} = \frac{\sum_{j \in \mathbf{1}} \pi_j}{\mu} \log\lambda + \sum_{j \in \mathbf{1}} \left(\frac{\pi_j^2}{2\mu^2}\mathbb{E}_j S_{\sigma_1(j)}^2 - \frac{\pi_j}{\mu}\mathbb{E}_i S_{\sigma_1(j)} \mathbb{1}_{\{j \neq i\}} + \frac{\pi_j}{\mu}\gamma \right)$$

$$+\frac{1}{\mu}\sum_{j\notin\mathbf{1}}\pi_j\sum_{b\notin\mathbf{0}_j}-p_{j,b}\log(1-p_{j,b})+\frac{1}{\mu}\sum_{j\in\mathbf{1}}\pi_j\psi_1(\log\lambda-\beta_{ij})$$
$$+\frac{1}{\mu}\sum_{j\notin\mathbf{1}}\pi_j\psi_j^{\Delta}(\log\lambda-\beta_{ij})+o(1)$$

in the d-arithmetic case. □

The next lemma constitutes both a further step towards the proofs of the main theorems of this subsection and an alternative approach, suggested and demonstrated in [Jan12a, Remark 3.5], to the results from Chapter 3. In fact, as will be seen in the proof, poissonization spares us the nuisance of time-dependent starting variables, replacing them with their distributional limit. Nevertheless, we have to deal with depoissonization instead.

Lemma 4.33. *It holds for $\lambda\to\infty$ that*

$$\mathbb{E}_i\widetilde{D}_\lambda=\frac{\log\lambda}{\mu}+\frac{1}{2\mu^2}\sum_{j\in\mathcal{S}}\pi_j^2\mathbb{E}_jS_{\sigma_1(j)}^2-\frac{1}{\mu}\sum_{j\neq i}\pi_j\mathbb{E}_iS_{\sigma_1(j)}+\frac{\gamma}{\mu}+\frac{1}{\mu}\sum_{j\in\mathcal{S}}\pi_j\psi_{i,j}^*(\log\lambda)+o(1),$$

where $\psi_{i,j}^*\equiv 0$ in the non-arithmetic case and

$$\psi_{i,j}^*(t)=\psi_1(t-\beta_{ij})$$

in the d-arithmetic case with shift function β, corresponding β_{ij} and ψ_1 from Theorem 3.3. γ denotes the Euler constant.

Proof. The proof differs slightly from previous proofs in the Poisson model where we had characterizations for specific nodes. Here we describe the functional in the same way as in Chapter 3. However, we do not need a family of initial variables here, their distributional limit X_0^* suffices. Recall that $-X_0^*$ has a standard Gumbel distribution.

As in the i.i.d. setting of [Jan12a, Remark 3.5], using independence we get

$$\mathbb{P}_i(\widetilde{D}_\lambda\le k\mid\xi_1,\dots,\xi_k)=\mathbb{P}_i\left(\widetilde{N}_\lambda(\xi_1\cdots\xi_k)=0\mid\xi_1,\dots,\xi_k\right)=e^{-\lambda P_i(\xi_1\cdots\xi_k)}=e^{-\lambda e^{-S_k}}$$
$$=e^{-e^{-(S_k-\log\lambda)}}=\mathbb{P}_i(-X_0^*<S_k-\log\lambda\mid\xi_1,\dots,\xi_k)=\mathbb{P}_i(X_0^*+S_k>\log\lambda\mid\xi_1,\dots,\xi_k)\quad\mathbb{P}_i\text{-a.s.}$$

for $k\in\mathbb{N}$, so

$$\mathbb{P}_i(\widetilde{D}_\lambda\le k)=\mathbb{P}_i(X_0^*+S_k>\log\lambda)=\mathbb{P}_i(\nu(X_0^*,\log\lambda)\le k).$$

Thus, $\widetilde{D}_\lambda$ and $\nu(X_0^*,\log\lambda)$ have the same distribution w.r.t. $\mathbb{P}_i$ and the assertion follows from Lemma 3.26 (with a constant family $(X_0^*)_t$ of initial variables) and the subsequent calculation of the fractional parts in Theorem 3.3. □

Remark 4.34. There is even a further possibility to derive Lemma 4.33 in the spirit of Chapter 4: With the convention about $\Xi,\Xi^{(1)},\Xi^{(2)},\dots$ from the beginning of this subsection, we write

$$D_n=\sum_{\alpha\in\mathcal{S}^*}\mathbb{1}_{\{\alpha\text{ internal node}\}}\mathbb{1}_{\{\Xi\succ\alpha\}}=\sum_{\alpha\in\mathcal{S}^*}\mathbb{1}_{\{N_{n-1}(\alpha)\ge1\}}\mathbb{1}_{\{\Xi\succ\alpha\}}.$$

Thus, it is easy to show that

$$\mathbb{E}_i\widetilde{D}_\lambda=\frac{1}{\lambda}\sum_{\alpha\in\mathcal{S}^*}f_{\mathrm{L}}(\lambda P_i(\alpha))\mathbb{1}_{\{P_i(\alpha)>0\}}$$

with f_{L} from the proof of the Poisson version of Theorem 4.28 or from Subsection 4.5.5. Theorem 4.7 yields the assertion. Subsection 4.5.5 uses a similar representation of L_n.

Alternative proof of Theorem 3.3. As $\mathbb{E}_i D_n$ is monotone, we can depoissonize with Lemma A.1, where (A.1) applies with $C_1 = 1/\mu$, $C_2 = 1/\mu$ and

$$\psi(t) = \frac{1}{2\mu}\sum_{j\in\mathcal{S}} \pi_j^2 \mathbb{E}_j S_{\sigma_1(j)}^2 - \sum_{j\neq i} \pi_j \mathbb{E}_i S_{\sigma_1(j)} + \gamma + \sum_{j\in\mathcal{S}} \pi_j \psi_{i,j}^*(t).$$

□

Proof of the Poisson version of Theorem 4.32. The assertion follows from $\mathbb{E}_i\widetilde{D}_\lambda^P = \mathbb{E}_i\widetilde{D}_\lambda - \mathbb{E}_i\Delta\widetilde{D}_\lambda$ and applying the Poisson version of Theorem 4.28 together with Lemma 4.33. □

Proof of Theorem 4.32. Since $\mathbb{E}_i D_n^P$ is monotone, we can depoissonize with Lemma A.1, where (A.1) applies with $C_1 = \sum_{j\notin\mathbf{1}} \pi_j/\mu$, $C_2 = 1/\mu$ and

$$\psi \equiv \sum_{j\notin\mathbf{1}} \left(\frac{\pi_j^2}{2\mu}\mathbb{E}_j S_{\sigma_1(j)}^2 - \pi_j \mathbb{E}_i S_{\sigma_1(j)} \mathbb{1}_{\{j\neq i\}} + \pi_j\gamma\right) - \sum_{j\notin\mathbf{1}} \pi_j \sum_{b\notin\mathbf{0}_j} -p_{j,b}\log(1-p_{j,b})$$

in the non-arithmetic case, and the obvious choice of $\psi(t)$ in the d-arithmetic case. □

Proof of Theorem 4.28. Note that $\Delta D_n := D_n - D_n^P$. □

4.5.5. Expected external path length of a trie

We consider $\mathrm{Trie}(\mathcal{M}_n) = \mathrm{Trie}(\Xi^{(1)}, \dots, \Xi^{(n)})$ and its *external path length* L_n which is its construction cost. There are (at least) two ways of characterizing the external path length such that we can work with it. Therefore, let $D_{n,k}$ be the depth of $\Xi^{(k)}$ in $\mathrm{Trie}(\mathcal{M}_n)$ as defined in Section 2.3. Then we can define

$$\mathrm{L}_n := \sum_{k=1}^n D_{n,k}$$

and apply our results from Chapter 3. On the other hand, we reformulate

$$\mathrm{L}_n = \sum_{k=1}^n D_{n,k} = \sum_{k=1}^n \sum_{\alpha\in\mathcal{S}^*} \mathbb{1}_{\{\alpha \text{ internal node}\}} \mathbb{1}_{\{\Xi^{(k)}\succ\alpha\}} = \sum_{\alpha\in\mathcal{S}^*} \mathbb{1}_{\{N_n(\alpha)\geq 2\}} N_n(\alpha). \tag{4.55}$$

If we want to expand the expected external path length asymptotically, then it is easiest to use the first expression and the linearity of the expectation. We immediately get:

Theorem 4.35. *(a) If $(\xi_n, S_n)_{n\geq 0}$ is non-arithmetic, then, as $n\to\infty$,*

$$\mathbb{E}_i \mathrm{L}_n = \frac{n\log n}{\mu} + n\cdot\left(\frac{1}{2\mu^2}\sum_{j\in\mathcal{S}} \pi_j^2 \mathbb{E}_j S_{\sigma_1(j)}^2 - \frac{1}{\mu}\sum_{j\neq i} \pi_j \mathbb{E}_i S_{\sigma_1(j)} + \frac{\gamma}{\mu}\right) + o(n), \tag{4.56}$$

with the Euler constant *γ.*

(b) If $(\xi_n, S_n)_{n\geq 0}$ is d-arithmetic with shift function β, then, as $n\to\infty$,

$$\begin{aligned}\mathbb{E}_i \mathrm{L}_n = \frac{n\log n}{\mu} &+ n\cdot\left(\frac{1}{2\mu^2}\sum_{j\in\mathcal{S}} \pi_j^2 \mathbb{E}_j S_{\sigma_1(j)}^2 - \frac{1}{\mu}\sum_{j\neq i} \pi_j \mathbb{E}_i S_{\sigma_1(j)} + \frac{\gamma}{\mu}\right) \\ &+ \frac{n}{\mu}\cdot\sum_{j\in\mathcal{S}} \pi_j \psi_1(\log n - \beta_{ij}) + o(n)\end{aligned} \tag{4.57}$$

with β_{ij} and continuous and d-periodic function $\psi_1(t)$ from Theorem 3.3.

The same holds in the Poisson model with $\widetilde{\mathrm{L}}_\lambda$ and λ instead of L_n and n, where $\widetilde{\mathrm{L}}_\lambda$ is defined in an obvious way.

Proof. The result is immediate from Theorem 3.3 (and Lemma 4.33). □

Remark 4.36. In the same way, we can also obtain the analogous result for the external path length $\mathrm{L}_n^{(b)}$ of a b-trie with the adjustments mentioned in Remark 3.8.

Nevertheless, Theorem 4.7 presents an alternative approach to such an expansion. The standard method is thus to poissonize and consider $\widetilde{\mathrm{L}}_\lambda := \mathrm{L}_{\Pi(\lambda)}$ and

$$\mathbb{E}_i\widetilde{\mathrm{L}}_\lambda = \mathbb{E}_i\left(\sum_{\alpha\in\mathcal{S}^*}\mathbb{1}_{\{\widetilde{N}_\lambda(\alpha)\geq 2\}}\widetilde{N}_\lambda(\alpha)\right) = \sum_{\alpha\in\mathcal{S}^*}\mathbb{E}_i\left(\mathbb{1}_{\{\widetilde{N}_\lambda(\alpha)\geq 2\}}\widetilde{N}_\lambda(\alpha)\right) = \sum_{\alpha\in\mathcal{S}^*} f_{\mathrm{L}}(\lambda P_i(\alpha))\mathbb{1}_{\{P_i(\alpha)>0\}}$$

for $f_{\mathrm{L}} : (0,\infty)\to\mathbb{R}_{\geq 0}$ defined by

$$\begin{aligned} f_{\mathrm{L}}(x) &:= \mathbb{E}_i\left(\Pi(x)\mathbb{1}_{\{\Pi(x)\geq 2\}}\right) = \mathbb{E}\,\Pi(x) - \mathbb{E}\,\Pi(x)\mathbb{1}_{\{\Pi(x)=1\}} \\ &= x - \mathbb{P}(\Pi(x)=1) = x - xe^{-x} = x(1-e^{-x}). \end{aligned}$$

This is the same function as in [Jan12b, Section 3], and $f_{\mathrm{L}} = f_j^{\Delta}$ from Subsection 4.5.4 if $j\in\mathbf{1}$.

Alternative proof of Theorem 4.35. We have already encountered f_{L} in the proof of the Poisson version of Theorem 4.28. We recall that $f_{\mathrm{L}}(x) = x - xe^{-x} = x + \mathcal{O}(1)$, as $x\to\infty$, and that this is the reason why Theorem 4.1 does not apply here and we need to use Theorem 4.7. We find

$$b = \frac{1}{2\mu}\sum_{j\in\mathcal{S}}\pi_j^2\mathbb{E}_j S_{\sigma_1(j)}^2 - \sum_{j\neq i}\pi_j\mathbb{E}_i S_{\sigma_1(j)} + \gamma$$

and $\hat{g}_{\mathrm{L}}(u) = -\Gamma(\mathrm{i}\,u)$ for $u\neq 0$. In the d-arithmetic case, ψ_1 equals its Fourier series.

It is obvious from the definition of L_n that $\mathbb{E}_i\,\mathrm{L}_n$ increases in n, so we can depoissonize with Lemma A.1, where (A.2) applies with $C_1 = 1/\mu$, $C_2 = 1/\mu$ and

$$\psi \equiv \frac{1}{2\mu}\sum_{j\in\mathcal{S}}\pi_j^2\mathbb{E}_j S_{\sigma_1(j)}^2 - \sum_{j\neq i}\pi_j\mathbb{E}_i S_{\sigma_1(j)} + \gamma$$

in the non-arithmetic case, and the obvious choice of $\psi(t)$ in the d-arithmetic case. □

4.5.6. Expected external path length of a PATRICIA-trie

Just as in Subsection 4.5.5, we consider $\mathrm{Trie}^P(\mathcal{M}_n)$ and its *external path length* L_n^P. Again, we provide two ways of deducing asymptotic expansions of the expected external path length. Therefore, let $D_{n,k}^P$ be the depth of $\Xi^{(k)}$ in $\mathrm{Trie}^P(\mathcal{M}_n)$ as defined in Subsection 4.5.4 (for $D_{n,1}^P = D_n^P$ and $D_{n,k}^P$ defined similarly). Note that Remark 2.9 applies to this slightly different situation as well and all $D_{n,k}^P$ share the same distribution. Then

$$\mathrm{L}_n^P := \sum_{k=1}^n D_{n,k}^P$$

and Theorem 4.32 applies. On the other hand

$$\begin{aligned}\mathrm{L}_n^P &= \sum_{k=1}^{n} D_{n,k}^P = \sum_{k=1}^{n} \sum_{\alpha\in\mathcal{S}^*} \mathbb{1}_{\{\alpha \text{ internal node in Pat-trie}\}} \mathbb{1}_{\{\Xi^{(k)}\succ\alpha\}} \\ &= \sum_{\alpha\in\mathcal{S}^*} \mathbb{1}_{\bigcup_{b\neq c\in\mathcal{S}}\{N_n(\alpha b)\geq 1,\ N_n(\alpha c)\geq 1\}} N_n(\alpha)\end{aligned}$$

which requires once more the use of Theorem 4.6. We state the result first:

Theorem 4.37. *(a) If $(\xi_n, S_n)_{n\geq 0}$ is non-arithmetic, then, as $n\to\infty$,*

$$\begin{aligned}\mathbb{E}_i L_n^P = &\frac{\sum_{j\notin\mathbf{1}}\pi_j}{\mu} n\log n + n\cdot\sum_{j\notin\mathbf{1}}\left(\frac{\pi_j^2}{2\mu^2}\mathbb{E}_j S_{\sigma_1(j)}^2 - \frac{\pi_j}{\mu}\mathbb{E}_i S_{\sigma_1(j)}\mathbb{1}_{\{j\neq i\}} + \frac{\pi_j\gamma}{\mu}\right) \\ &- \frac{n}{\mu}\sum_{j\notin\mathbf{1}}\pi_j\sum_{b\notin\mathbf{0}_j} -p_{j,b}\log(1-p_{j,b}) + o(n),\end{aligned} \tag{4.58}$$

with the Euler constant γ.

(b) If $(\xi_n, S_n)_{n\geq 0}$ is d-arithmetic with shift function β, then, as $n\to\infty$,

$$\begin{aligned}\mathbb{E}_i L_n^P = &\frac{\sum_{j\notin\mathbf{1}}\pi_j}{\mu} n\log n + n\cdot\sum_{j\notin\mathbf{1}}\left(\frac{\pi_j^2}{2\mu^2}\mathbb{E}_j S_{\sigma_1(j)}^2 - \frac{\pi_j}{\mu}\mathbb{E}_i S_{\sigma_1(j)}\mathbb{1}_{\{j\neq i\}} + \frac{\pi_j\gamma}{\mu}\right) \\ &- \frac{n}{\mu}\sum_{j\notin\mathbf{1}}\pi_j\sum_{b\notin\mathbf{0}_j} -p_{j,b}\log(1-p_{j,b}) + \frac{n}{\mu}\sum_{j\notin\mathbf{1}}\pi_j\psi_1(\log n - \beta_{ij}) \\ &- \frac{n}{\mu}\sum_{j\notin\mathbf{1}}\pi_j\psi_j^{\Delta}(\log n - \beta_{ij}) + o(n),\end{aligned} \tag{4.59}$$

with continuous and d-periodic ψ_1 and ψ_j^{Δ} from Theorem 3.3 and Theorem 4.28, respectively (and β_{ij} from Theorem 3.3).

The same holds in the Poisson model with $\widetilde{\mathrm{L}}_\lambda^P$ and λ instead of L_n^P and n.

Proof. The result is immediate from Theorem 4.32. □

Remark 4.38. This theorem answers the question by Bourdon for the Markov case which he poses in [Bou01, Section 5.5], whether it is possible that a correcting term appears in the leading term of the asymptotic expansion. Indeed, in the Markov model, this happens iff $\mathbf{1}\neq\emptyset$. Then, the correcting term is $\sum_{j\in\mathbf{1}}\pi_j$.

Now, Theorem 4.6 presents an alternative approach to such an expansion. The standard method is thus to poissonize and consider $\widetilde{\mathrm{L}}_\lambda^P := \mathrm{L}_{\Pi(\lambda)}^P$ as well as

$$\begin{aligned}\mathbb{E}_i\widetilde{\mathrm{L}}_\lambda^P &= \sum_{\alpha\in\mathcal{S}^*}\mathbb{E}_i\left(\mathbb{1}_{\bigcup_{b\neq c\in\mathcal{S}}\{\widetilde{N}_\lambda(\alpha b)\geq 1,\ \widetilde{N}_\lambda(\alpha c)\geq 1\}}\widetilde{N}_\lambda(\alpha)\right) \\ &= \sum_{\alpha\in\mathcal{S}^*}\mathbb{1}_{\{P_i(\alpha)>0\}}\Big(\mathbb{E}_i\widetilde{N}_\lambda(\alpha) - \mathbb{E}_i\widetilde{N}_\lambda(\alpha)\mathbb{1}_{\{\widetilde{N}_\lambda(\alpha)=0\}} - \sum_{b\in\mathcal{S}}\mathbb{E}_i\widetilde{N}_\lambda(\alpha)\mathbb{1}_{\{\widetilde{N}_\lambda(\alpha b)\geq 1,\ \widetilde{N}_\lambda(\alpha c)=0, c\neq b\}}\Big).\end{aligned}$$

On $\{P_i(\alpha)>0\}$ the last summand simplifies by independence to

$$\sum_{b\in\mathcal{S}}\mathbb{E}_i\widetilde{N}_\lambda(\alpha)\mathbb{1}_{\{\widetilde{N}_\lambda(\alpha b)\geq 1,\ \widetilde{N}_\lambda(\alpha c)=0, c\neq b\}} = \sum_{b\in\mathcal{S}}\mathbb{E}_i\widetilde{N}_\lambda(\alpha b)\mathbb{1}_{\{\widetilde{N}_\lambda(\alpha b)\geq 1,\ \widetilde{N}_\lambda(\alpha c)=0, c\neq b\}}$$

$$= \sum_{b\in\mathcal{S}} \mathbb{E}_i \left[\widetilde{N}_\lambda(\alpha b)\mathbb{1}_{\{\widetilde{N}_\lambda(\alpha b)\geq 1\}}\right] \prod_{c\neq b} \mathbb{P}_i\left(\widetilde{N}_\lambda(\alpha c)=0\right)$$

and finally,

$$\begin{aligned}
\mathbb{E}_i\widetilde{\mathrm{L}}_\lambda^P &= \sum_{\alpha\in\mathcal{S}^*} \mathbb{1}_{\{P_i(\alpha)>0\}}\left(\mathbb{E}_i\widetilde{N}_\lambda(\alpha) - \sum_{b\in\mathcal{S}} \mathbb{E}_i\widetilde{N}_\lambda(\alpha)\mathbb{1}_{\{\widetilde{N}_\lambda(\alpha b)\geq 1,\ \widetilde{N}_\lambda(\alpha c)=0, c\neq b\}}\right) \\
&= \sum_{j\in\mathcal{S}}\sum_{\alpha\in\mathcal{S}_j^*} \mathbb{1}_{\{P_i(\alpha)>0\}}\Bigg(\mathbb{E}\left[\Pi(\lambda P_i(\alpha))\right] \\
&\qquad - \sum_{b\notin\mathbf{0}_j}\mathbb{E}\left[\Pi(\lambda P_i(\alpha b))\right]\left(\mathbb{1}_{\{p_{j,b}<1\}}\mathbb{P}(\Pi(\lambda P_i(\alpha)(1-p_{j,b}))=0)+\mathbb{1}_{\{p_{j,b}=1\}}\right)\Bigg) \\
&\qquad + \mathbb{E}\left[\Pi(\lambda)\right] - \sum_{b\notin\mathbf{0}_i}\mathbb{E}\left[\Pi(\lambda p_{i,b})\right]\left(\mathbb{1}_{\{p_{i,b}<1\}}\mathbb{P}(\Pi(\lambda(1-p_{i,b}))=0)+\mathbb{1}_{\{p_{i,b}=1\}}\right) \\
&= \sum_{j\in\mathcal{S}}\left(\sum_{\alpha\in\mathcal{S}^*} f_j^{\mathrm{L}^P}(\lambda P_i(\alpha j))\mathbb{1}_{\{P_i(\alpha j)>0\}}\right) + f_i^{\mathrm{L}^P}(\lambda)
\end{aligned}$$

for $f_j^{\mathrm{L}^P}:(0,\infty)\to\mathbb{R}_{\geq 0}$ defined by

$$\begin{aligned}
f_j^{\mathrm{L}^P}(x) &:= \mathbb{E}\left[\Pi(x)\right] - \sum_{b\notin\mathbf{0}_j}\mathbb{E}\left[\Pi(xp_{j,b})\right]\left(\mathbb{1}_{\{p_{j,b}<1\}}\mathbb{P}(\Pi(x(1-p_{j,b}))=0)+\mathbb{1}_{\{p_{j,b}=1\}}\right) \\
&= x - \sum_{b\notin\mathbf{0}_j} xp_{j,b}\left(\mathbb{1}_{\{p_{j,b}<1\}}e^{-x(1-p_{j,b})}+\mathbb{1}_{\{p_{j,b}=1\}}\right) \\
&= x\left(1-\sum_{b\notin\mathbf{0}_j} p_{j,b}e^{-x(1-p_{j,b})}\right).
\end{aligned}$$

As discussed in Subsection 4.5.3, $f_j^{\mathrm{L}^P}$ and the corresponding harmonic sum vanish if $j\in\mathbf{1}$. Otherwise, if all $p_{j,b}<1$, then $f_j^{\mathrm{L}^P}$ satisfies (4.12) and we can use Theorem 4.6.

Alternative proof of Theorem 4.37. Suppose $j\notin\mathbf{1}$. Obviously, $f_j^{\mathrm{L}^P}$ satisfies (4.12) with $a=1$ and $\delta=1$, use l'Hôpital's rule once for the first part. Hence, Theorem 4.6 applies and, as $\lambda\to\infty$,

$$\sum_{\alpha\in\mathcal{S}^*} f_j^{\mathrm{L}^P}(\lambda P_i(\alpha j))\mathbb{1}_{\{P_i(\alpha j)>0\}} = \frac{\pi_j}{\mu}\lambda\log\lambda + \frac{b_j}{\mu}\lambda - \lambda\mathbb{1}_{\{j\}}(i) + o(\lambda)$$

with

$$b_j = \frac{1}{2\mu}\pi_j^2\mathbb{E}_jS_{\sigma_1(j)}^2 - \pi_j\mathbb{E}_iS_{\sigma_1(j)}\mathbb{1}_{\{j\neq i\}} + \pi_j\int_0^\infty (f_j^{\mathrm{L}^P}(x) - x\mathbb{1}_{\{x\geq 1\}})x^{-2}\,\mathrm{d}x.$$

The integral is

$$\begin{aligned}
\int_0^\infty (f_j^{\mathrm{L}^P}(x) - x\mathbb{1}_{\{x\geq 1\}})x^{-2}\,\mathrm{d}x &= \int_0^\infty\left(\mathbb{1}_{\{x<1\}} - \sum_{b\notin\mathbf{0}_j} p_{j,b}e^{-x(1-p_{j,b})}\right)x^{-1}\,\mathrm{d}x \\
&= -\sum_{b\notin\mathbf{0}_j} p_{j,b}\int_0^\infty (e^{-x(1-p_{j,b})} - \mathbb{1}_{\{x<1\}})x^{-1}\,\mathrm{d}x
\end{aligned}$$

$$= -\sum_{b\notin\mathbf{0}_j} p_{j,b}\int_0^\infty (e^{-x(1-p_{j,b})} - e^{-x})x^{-1}\mathrm{d}x - \int_0^\infty (e^{-x} - \mathbb{1}_{\{x<1\}})x^{-1}\mathrm{d}x$$
$$= \sum_{b\notin\mathbf{0}_j} p_{j,b}\log(1-p_{j,b}) + \gamma,$$

cf. [Jan13, (16) and (25)]. Since $f_i^{\mathrm{L}^P}(\lambda) = (\lambda + o(\lambda))(1 - |\mathbf{1}_i|)$, as $\lambda \to \infty$, we obtain

$$\begin{aligned}\mathbb{E}_i\widetilde{L}_\lambda^P &= \frac{\sum_{j\notin\mathbf{1}}\pi_j}{\mu}\lambda\log\lambda + \lambda\cdot\sum_{j\notin\mathbf{1}}\left(\frac{\pi_j^2}{2\mu^2}\mathbb{E}_j S_{\sigma_1(j)}^2 - \frac{\pi_j}{\mu}\mathbb{E}_i S_{\sigma_1(j)}\mathbb{1}_{\{j\neq i\}} + \frac{\pi_j\gamma}{\mu}\right)\\
&\quad - \frac{\lambda}{\mu}\sum_{j\notin\mathbf{1}}\pi_j\sum_{b\notin\mathbf{0}_j} -p_{j,b}\log(1-p_{j,b}) - \lambda\sum_{j\notin\mathbf{1}}\mathbb{1}_{\{j=i\}} + (\lambda+o(\lambda))(1-|\mathbf{1}_i|) + o(\lambda)\\
&= \frac{\sum_{j\notin\mathbf{1}}\pi_j}{\mu}\lambda\log\lambda + \lambda\cdot\sum_{j\notin\mathbf{1}}\left(\frac{\pi_j^2}{2\mu^2}\mathbb{E}_j S_{\sigma_1(j)}^2 - \frac{\pi_j}{\mu}\mathbb{E}_i S_{\sigma_1(j)}\mathbb{1}_{\{j\neq i\}} + \frac{\pi_j\gamma}{\mu}\right)\\
&\quad - \frac{\lambda}{\mu}\sum_{j\notin\mathbf{1}}\pi_j\sum_{b\notin\mathbf{0}_j} -p_{j,b}\log(1-p_{j,b}) + o(\lambda)\end{aligned}$$

in the non-arithmetic case. Additionally, we obtain the summand

$$\frac{\lambda}{\mu}\sum_{j\notin\mathbf{1}}\pi_j\psi_j^{\mathrm{L}^P}(\log\lambda - \beta_{ij})$$

in the d-arithmetic case. As usual, the Fourier coefficients are

$$\hat{g}_j^{\mathrm{L}^P}(-2\pi m/d) = \lim_{\varepsilon\searrow 0}\int_0^\infty f_j^{\mathrm{L}^P}(x)x^{-2-\varepsilon-2\pi\,\mathrm{i}\,m/d}\,\mathrm{d}x$$

and we calculate for $u \in \mathbb{R}$ and $0 < \varepsilon < 1$ (and $j \notin \mathbf{1}$) that

$$\begin{aligned}&\int_0^\infty f_j^{\mathrm{L}^P}(x)x^{-2-\varepsilon+\mathrm{i}\,u}\,\mathrm{d}x = \int_0^\infty\left(1 - \sum_{b\notin\mathbf{0}_j}p_{j,b}e^{-x(1-p_{j,b})}\right)x^{-1-\varepsilon+\mathrm{i}\,u}\,\mathrm{d}x\\
&= \sum_{b\notin\mathbf{0}_j}p_{j,b}\int_0^\infty\left(1 - e^{-x(1-p_{j,b})}\right)x^{-1-\varepsilon+\mathrm{i}\,u}\,\mathrm{d}x\\
&= -\sum_{b\notin\mathbf{0}_j}p_{j,b}\left(\int_0^\infty(e^{-x}-1)\,x^{-1-\varepsilon+\mathrm{i}\,u}\,\mathrm{d}x + \int_0^\infty\left(e^{-x(1-p_{j,b})} - e^{-x}\right)x^{-1-\varepsilon+\mathrm{i}\,u}\,\mathrm{d}x\right)\\
&= -\sum_{b\notin\mathbf{0}_j}p_{j,b}\left(\Gamma(-\varepsilon+\mathrm{i}\,u) + \left((1-p_{j,b})^{\varepsilon-\mathrm{i}\,u} - 1\right)\Gamma(-\varepsilon+\mathrm{i}\,u)\right),\end{aligned}$$

where we used [Jan13, (2) and (24)]. Hence,

$$\hat{g}_j^{\mathrm{L}^P}(u) = -\Gamma(\mathrm{i}\,u) - \sum_{b\notin\mathbf{0}_j}p_{j,b}\left((1-p_{j,b})^{-\mathrm{i}\,u} - 1\right)\Gamma(\mathrm{i}\,u)$$

and we find that $\psi_j^{\mathrm{L}^P} = \psi_1 - \psi_j^{\Delta}$. This completes the proof in the Poisson model.

It is obvious from the definition of L_n^P that $\mathbb{E}_i\,\mathrm{L}_n^P$ increases in n, so we can depoissonize with Lemma A.1, where (A.2) applies with $C_1 = \sum_{j\notin\mathbf{1}}\pi_j/\mu$, $C_2 = 1/\mu$ and

$$\psi \equiv \sum_{j\notin\mathbf{1}}\left(\frac{\pi_j^2}{2\mu}\mathbb{E}_j S_{\sigma_1(j)}^2 - \pi_j\mathbb{E}_i S_{\sigma_1(j)}\mathbb{1}_{\{j\neq i\}} + \pi_j\gamma\right) - \sum_{j\notin\mathbf{1}}\pi_j\sum_{b\notin\mathbf{0}_j} -p_{j,b}\log(1-p_{j,b})$$

in the non-arithmetic case, and the obvious choice of $\psi(t)$ in the d-arithmetic case. □

A. Appendix

A.1. Properties of $(X_0^{(n)})_{n\geq 2}$

We have outsourced the following proof, since it is rather technical and the result seems natural due to the proximity of $X_0^{(n)}$ to the Gumbel distribution.

Proof of Lemma 3.13(b). We consider the tails: Let $n \geq 2$ and $x > 0$. Then, as $x \to \infty$,

$$\begin{aligned}
\mathbb{P}\left(X_0^{(n)} < -x\right) &= 1 - \mathbb{P}\left(X_0^{(n)} > -x\right) = 1 - \left(1 - \frac{e^{-x}}{n}\right)^{n-1} \\
&= 1 - \exp\left((n-1)\log\left(1 - \frac{e^{-x}}{n}\right)\right) = 1 - \exp\left(-(n-1)\sum_{k=1}^{\infty}\frac{1}{k}\left(\frac{e^{-x}}{n}\right)^k\right) \\
&= 1 - \exp\left(-\frac{(n-1)}{n}\sum_{k=1}^{\infty}\frac{e^{-kx}}{kn^{k-1}}\right) \leq 1 - \exp\left(-\sum_{k=1}^{\infty}\frac{e^{-kx}}{kn^{k-1}}\right) \leq 1 - \exp\left(-\sum_{k=1}^{\infty}(e^{-x})^k\right) \\
&= 1 - \exp\left(-\frac{e^{-x}}{1-e^{-x}}\right) \sim e^{-x},
\end{aligned}$$

where we used the Taylor expansion of $\log(1-x)$ for $|x| < 1$. Consequently,

$$\begin{aligned}
\sup_{n\geq 2}\mathbb{E}\left(e^{r(X_0^{(n)})^-} - 1\right) &= \sup_{n\geq 2}\int_0^\infty re^{rx}\mathbb{P}\left((X_0^{(n)})^- > x\right)\,\mathrm{d}x \\
&= \sup_{n\geq 2}\int_0^\infty re^{rx}\mathbb{P}\left(X_0^{(n)} < -x\right)\,\mathrm{d}x < \infty
\end{aligned}$$

for $r < 1$. Similarly, for $n \geq 2$ and $x > 0$

$$\begin{aligned}
\mathbb{P}\left(X_0^{(n)} > x\right) &= \left(1 - \frac{e^x}{n}\right)^{n-1}\mathbb{1}_{(0,\log n)}(x) = \exp\left((n-1)\log\left(1 - \frac{e^x}{n}\right)\right)\mathbb{1}_{(0,\log n)}(x) \\
&= \exp\left(-(n-1)\sum_{k=1}^{\infty}\frac{1}{k}\left(\frac{e^x}{n}\right)^k\right)\mathbb{1}_{(0,\log n)}(x) \\
&\leq \exp\left(-\frac{(n-1)}{n}e^x\right)\mathbb{1}_{(0,\log n)}(x) \\
&\leq \exp\left(-\frac{1}{2}e^x\right)\mathbb{1}_{(0,\log n)}(x) \leq \exp\left(-\frac{1}{2}e^x\right).
\end{aligned}$$

Hence, as above

$$\sup_{n\geq 2}\mathbb{E}\left(e^{r(X_0^{(n)})^+} - 1\right) = \sup_{n\geq 2}\int_0^\infty re^{rx}\mathbb{P}\left((X_0^{(n)})^+ > x\right)\,\mathrm{d}x = \sup_{n\geq 2}\int_0^\infty re^{rx}\mathbb{P}\left(X_0^{(n)} > x\right)\,\mathrm{d}x < \infty$$

for the same r. In particular,

$$\sup_{n\geq 2}\mathbb{E}e^{r|X_0^{(n)}|}\leq \sup_{n\geq 2}\mathbb{E}e^{r(X_0^{(n)})^+}+\sup_{n\geq 2}\mathbb{E}e^{r(X_0^{(n)})^-}<\infty,$$

and $(e^{s|X_0^{(n)}|})_{n\geq 2}$ is uniformly integrable for $s<r$. □

A.2. Depoissonization for stochastically monotone parameters

Depoissonization for most of our trie-related parameters is based on the fact that the expected parameter is monotone in the number of strings the tree is built from. Therefore, we state the following auxiliary result which applies to most of our depoissonization proofs. The following Lemma can be found in the proof of [Jan12a, Theorem 5.4] for one parameter, viz. the size W_n, indeed it does not require specific knowledge about the parameter except for a monotonicity assumption. It is not special to the Markov model.

Lemma A.1. *Let $(W_n)_{n\geq 0}$ be a sequence of $\mathbb{N}_0$-valued random variables, independent of $(\Pi(\lambda))_{\lambda>0}$ under $\mathbb{P}_i$, with $\mathbb{E}_iW_n\leq \mathbb{E}_iW_{n+1}$ for all $n\geq 0$. Suppose that $\widetilde{W}_\lambda := W_{\Pi(\lambda)}$ either admits an asymptotic expansion of the form ($\lambda\to\infty$)*

$$\mathbb{E}_i\widetilde{W}_\lambda = C_1\log\lambda + C_2\psi(\log\lambda)+o(1) \tag{A.1}$$

or

$$\mathbb{E}_i\widetilde{W}_\lambda = C_1\lambda\log\lambda + C_2\lambda\psi(\log\lambda)+o(\lambda) \tag{A.2}$$

for non-negative constants C_1, C_2, with $C_1\vee C_2>0$, and some bounded, uniformly continuous function ψ. Then, as $n\to\infty$,

$$\mathbb{E}_iW_n=\mathbb{E}_i\widetilde{W}_n.$$

The same holds for $\widetilde{W}_\lambda := W_{1+\Pi(\lambda)}$ instead of $\widetilde{W}_\lambda := W_{\Pi(\lambda)}$ if $\mathbb{E}_iW_n\leq\mathbb{E}_iW_{n+1}$ for all $n\geq 1$.

Remark A.2. Lemma A.1 applies to more asymptotic expansions than only to (A.1) and (A.2), which will be obvious from the proof. However, these are the only ones appearing in this work, so we restrict ourselves to this setting.

We use the following two lemmas for the proof of Lemma A.1:

Lemma A.3 (Chernoff bound for the Poisson distribution)**.** *It holds:*

(a) If $0<\lambda<x$, then

$$\mathbb{P}\left(\Pi(\lambda)\geq x\right)\leq e^{-\lambda}\left(\frac{e\cdot\lambda}{x}\right)^x.$$

(b) If $0<x<\lambda$, then

$$\mathbb{P}\left(\Pi(\lambda)\leq x\right)\leq e^{-\lambda}\left(\frac{e\cdot\lambda}{x}\right)^x.$$

Proof. Exponential Markov inequality and minimizing. □

Lemma A.4. *Let $\beta\in(\frac{1}{2},1)$. Then for $\alpha\in(0,\frac{1}{2})$, as $n\to\infty$,*

$$\mathbb{P}\left(\Pi(n+n^\beta)\leq n-1\right)\leq\mathbb{P}\left(\Pi(n+n^\beta)\leq n\right)=\mathcal{O}\left(\exp(-\alpha n^{-1+2\beta})\right)$$

as well as

$$\mathbb{P}\left(\Pi(n-n^\beta)\geq n\right)\leq\mathbb{P}\left(\Pi(n-n^\beta)\geq n-1\right)=\mathcal{O}\left(\exp(-\alpha n^{-1+2\beta})\right).$$

Proof. Chernoff bound and Taylor expansion. □

Proof of Lemma A.1. We use sandwiching with $\lambda = n \pm n^{2/3}$ to obtain the corresponding results for $\mathbb{E}_i W_n$. Let $n \geq 1$. As $\mathbb{P}(\Pi(2n) \geq n) \geq \frac{1}{2}$, we obtain

$$\begin{aligned}\mathbb{E}_i \widetilde{W}_{2n} &= \sum_{k=0}^{\infty} \mathbb{E}_i W_k \mathbb{P}_i(\Pi(2n) = k) \geq \sum_{k=n}^{\infty} \mathbb{E}_i W_n \mathbb{P}_i(\Pi(2n) = k) \\ &= \mathbb{E}_i W_n \mathbb{P}_i(\Pi(2n) \geq n) \geq \frac{1}{2} \mathbb{E}_i W_n\end{aligned}$$

using independence and the monotonicity of $\mathbb{E}_i W_n$. Thus, $\mathbb{E}_i W_n = \mathcal{O}(\mathbb{E}_i \widetilde{W}_{2n})$, as $n \to \infty$, which is at most $\mathcal{O}(n \log n)$, cf. (A.1) and (A.2) and note that ψ is bounded.

Similarly, with $\mathbb{P}(\Pi(n + n^{2/3}) < n)$ decreasing exponentially by the Chernoff bound from Lemma A.4, we get

$$\begin{aligned}\mathbb{E}_i W_n &= \mathbb{E}_i W_n \sum_{k=n}^{\infty} \mathbb{P}_i(\Pi(n + n^{2/3}) = k) + \mathbb{E}_i W_n \mathbb{P}_i(\Pi(n + n^{2/3}) < n) \\ &\leq \sum_{k=n}^{\infty} \mathbb{E}_i W_k \mathbb{P}_i(\Pi(n + n^{2/3}) = k) + \mathcal{O}(\mathbb{E}_i \widetilde{W}_{2n}) \cdot \mathcal{O}(e^{-\frac{1}{4} n^{1/3}}) \\ &= \mathbb{E}_i \widetilde{W}_{n+n^{2/3}} \mathbb{1}_{\{\Pi(n+n^{2/3}) \geq n\}} + o(1) \leq \mathbb{E}_i \widetilde{W}_{n+n^{2/3}} + o(1),\end{aligned}$$

as $n \to \infty$. For a lower bound, we obtain in the same way

$$\mathbb{E}_i \widetilde{W}_{n-n^{2/3}} \leq \mathbb{E}_i W_n + \mathbb{E}_i \widetilde{W}_{n-n^{2/3}} \mathbb{1}_{\{\Pi(n-n^{2/3}) \geq n\}},$$

where the second summand is of order $o(1)$, as $n \to \infty$, because for some $C, C' > 0$

$$\begin{aligned}\mathbb{E}_i \widetilde{W}_{n-n^{2/3}} \mathbb{1}_{\{\Pi(n-n^{2/3}) \geq n\}} &= \sum_{k=n}^{\infty} \mathbb{E}_i W_k \mathbb{1}_{\{\Pi(n-n^{2/3}) = k\}} \\ &\leq C \cdot \sum_{k=n}^{\infty} \mathbb{E}_i \widetilde{W}_{2k} \cdot \mathbb{P}(\Pi(n - n^{2/3}) = k) \\ &\leq C' \cdot \sum_{k=n}^{\infty} k \log k \cdot \mathbb{P}(\Pi(n - n^{2/3}) = k) \\ &\leq C' \cdot \sum_{k=n}^{\infty} k^2 \cdot \mathbb{P}(\Pi(n - n^{2/3}) = k) \\ &= C \cdot \mathbb{E}\left(\Pi^2(n - n^{2/3}) \mathbb{1}_{\{\Pi(n-n^{2/3}) \geq n\}}\right) \\ &= \mathcal{O}(n^2) \cdot \mathcal{O}(e^{-\frac{1}{8} n^{1/3}}) = o(1)\end{aligned}$$

by the Cauchy–Schwarz inequality. Hence, as $n \to \infty$, we have

$$\mathbb{E}_i \widetilde{W}_{n-n^{2/3}} + o(1) \leq \mathbb{E}_i W_n \leq \mathbb{E}_i \widetilde{W}_{n+n^{2/3}} + o(1).$$

Suppose we are in case (A.1). If we define

$$b_n := C_1 \log n + C_2 \psi(\log n),$$

then we need to show $|\mathbb{E}_i W_n - b_n| = o(1)$, as $n \to \infty$. We start with an upper bound: As $n \to \infty$,

$$\begin{aligned}\mathbb{E}_i W_n - b_n &\le \mathbb{E}_i \widetilde{W}_{n+n^{2/3}} - b_n + o(1)\\ &\le C_1 \left(\log(n+n^{2/3}) - \log n\right) + C_2 \left(\psi(\log(n+n^{2/3})) - \psi(\log n)\right)\\ &= o(1),\end{aligned}$$

since, on the one hand, $|\log(n+n^{2/3}) - \log n| = \mathcal{O}(n^{-1/3})$ and, on the other hand, ψ is uniformly continuous. So $\mathbb{E}_i W_n - b_n \le o(1)$, as $n \to \infty$, and the other inequality follows similarly. This proves case (A.1).

Suppose now that we are in case (A.2), then with

$$b_n := C_1 n \log n + C_2 n \psi(\log n)$$

and the use of the previous procedure

$$\begin{aligned}\mathbb{E}_i W_n - b_n &\le \mathbb{E}_i \widetilde{W}_{n+n^{2/3}} - b_n + o(1)\\ &\le C_1 \left((n+n^{2/3}) \log(n+n^{2/3}) - n \log n\right)\\ &\quad + C_2 \left((n+n^{2/3})\psi(\log(n+n^{2/3})) - n\psi(\log n)\right)\\ &\le n \cdot o(1) + C_1 n^{2/3} \log(n+n^{2/3}) + C_2 n^{2/3} \psi(\log(n+n^{2/3}))\\ &= o(n).\end{aligned}$$

Again, the other inequality follows similarly. For $\widetilde{W}_\lambda := W_{1+\Pi(\lambda)}$ the same proof applies with only minor modifications. This completes the proof. □

A.3. Auxiliary convergence result

Lemma A.5. *Let S_n be a SRW with L^p-increments X_n, $p > 0$. Then, as $n \to \infty$,*

$$\frac{S_n - S_{n-1}}{n^{1/p}} = \frac{X_n}{n^{1/p}} \to 0 \quad \mathbb{P}\text{-}a.s.$$

Proof. Well-known, included e.g. in the proof of [Gut09, Theorem 1.2.3(i)]. □

Part II.

Convergence rates of iterated function systems of Markov-modulated Lipschitz maps by regenerative methods

5. Introduction

Iterated function systems (IFSs) have long been studied in various fields of applied probability, e.g. in image encoding and the drawing of fractal images by [BE88] and [Bar93], or the construction of Markov chains and the simulation of distributions by the Propp-Wilson algorithm (cf. [PW96] and [PW98]).

5.1. Iterated function systems

An IFS is constructed by an iteration of functions on some space into itself. These functions are picked by some probability law and every iteration represents the next step of the thus constructed process. In practice, the random functions are often Lipschitz maps on some complete separable metric space and the space of Lipschitz maps can be equipped with a measurable structure. The easiest case of those systems is the IFS of i.i.d. Lipschitz maps, where the random functions are chosen independently and identically distributed. In the model of [AF01], one is given a starting state X_0, and then the Markov chain $(X_n)_{n\geq 0}$ is constructed recursively by

$$X_n = \Psi(\theta_n, X_{n-1})$$

for $n \geq 1$, where Ψ is a jointly measurable function that is Lipschitz continuous in the second component. Therefore, the i.i.d. $(\theta_n)_{n\geq 1}$ (and independent of X_0) represent the i.i.d. influence on the iterations. In fact, setting $\Psi_n(x) := \Psi(\theta_n, x)$, the $(\Psi_n)_{n\geq 1}$ are i.i.d. Lipschitz maps.

As one purpose of constructing IFSs is to simulate a (stationary) distribution, one key question is under which condition the chain converges to this stationary distribution (and how fast this convergence is). These questions have already been addressed by Dubins and Freedman [DF66] in 1966. In the great survey [DF99], the authors give a list of further publications including answers to these questions (and also several interesting examples): [Hut81], [BE88], [BEH89], [Elt90], [AC92] and [Duf97] are among these. Most notably, we will often allude to Elton's article [Elt90] in which he chose the more general setup of a stationary sequence $(\Psi_n)_{n\geq 1}$ of Lipschitz maps. Denoting by $L(\Psi)$ the Lipschitz constant of a Lipschitz map Ψ, Elton showed that whenever $\mathbb{E}\log^+ L(\Psi_1) < \infty$ and $\mathbb{E}\log^+ d(\Psi_1(x), x) < \infty$ for some x and also the Lyapunov exponent χ is negative a.s. (cf. Theorem 6.8 for a definition in our setting), then the postulated convergence holds, i.e. X_n converges in distribution to the stationary distribution, no matter what the initial value X_0 is.

Forward and backward iterations. A key ingredient, that (not only) Elton uses in his proof, and that is again easiest to be understood in the i.i.d. setting, is the notion of time reversal. More precisely, consider the *backward iterations* $\widehat{X}_n := \Psi_1 \circ \cdots \circ \Psi_n(X_0)$, $n \geq 1$, and note that X_n and $\widehat{X}_n$ have the same law. However, the backward iterations do actually converge a.s. In Elton's proof, time reversal enters via a stationary backward extension of the stationary sequence of Lipschitz maps. Some authors require the stronger but somewhat more intuitive assumption $\mathbb{E}\log L(\Psi_1) < 0$ instead of the a.s. negativity of the Lyapunov exponent, which expresses that the IFS converges if, as [DF99] call it, the functions are "contracting on average".

Convergence rates and central goal. Concerning convergence rates for the distributional convergence in the i.i.d. setting, several authors have contributed so far. Diaconis and Freedman give a condition in [DF99] for convergence at an exponential rate, while Alsmeyer and Fuh also give a condition for a polynomial rate in [AF01] and [AF02]. Denoting by P and π the transition kernel of the Markov chain $(X_n)_{n\geq 0}$ and the distributional limit of X_n, respectively, the two main results of the latter can be stated as follows: Suppose $\mathbb{E}\log L(\Psi_1) < 0$ and $\mathbb{E}\log^+ d(\Psi_1(x_0), x_0) < \infty$ for some x_0, and let d_{Pr} denote the Prokhorov metric of probability measures. Then there are two regimes:

Theorem 5.1 ([AF01], Theorem 2.2 (d))**.** *Let $p > 0$. If $\mathbb{E}\log^{p+1}(1 + L(\Psi_1)) < \infty$ and for some x_0 we have $\mathbb{E}\log^{p+1}(1 + d(\Psi_1(x_0), x_0)) < \infty$, then*

$$d_{\text{Pr}}(P^n(x, \cdot), \pi) \leq A_x(n+1)^{-p}$$

for all $n \geq 0$, all x and a positive constant A_x of the form $\max\{A, 2d(x, x_0)\}$, where A does neither depend on x nor on n.

Theorem 5.2 ([AF01], Theorem 2.3 (c))**.** *Let $p > 0$. If $\mathbb{E}L(\Psi_1)^p < \infty$ and $\mathbb{E}d(\Psi_1(x_0), x_0)^p < \infty$ for some x_0, then*

$$d_{\text{Pr}}(P^n(x, \cdot), \pi) \leq A_x r^n$$

for all $n \geq 0$, all x, some $r \in (0, 1)$ and a positive constant A_x of the form $\max\{A, d(x, x_0)\}$, where r and A do not depend on x nor on n.

A secondary goal of [AF01] was also to convey the effective use of renewal-theoretic/regenerative methods in the derivation of these convergence rates, an intention that we also pursue in this work.

In an attempt to generalize results on convergence rates, it seems natural to require $(\theta_n, \Psi_n)_{n\geq 0}$ to be a MMS. One big advantage of this setting is the following: If θ is a discrete Markov chain, then regeneration techniques, comparable to those used in [AF01], become available. Furthermore, starting θ stationary makes $(\Psi_n)_{n\geq 1}$ stationary and also makes it fit naturally into the framework of [Elt90]. Our main goal in this part will thus be, to derive suitable sufficient conditions for different rates of convergence in the stationary regime of a MMS of Lipschitz maps, making use of the regeneration techniques provided by the underlying structure.

Throughout this work, we are guided by [AF01] (and the corrigendum [AF02] to this publication) and [Als15a]. They, most notably, use cyclic decomposition of some RW of Lipschitz constants along ladder epochs, thereby gaining considerable knowledge about the size of these constants. We will instead use cyclic decomposition of a MRW along recurrence times and therefore lose this knowledge, while still maintaining independence at least. This forces us to come up with further theory at some points. We will address this issue at the beginning of Section 7.2.

Example: Affine functions. A well-studied example of IFSs of i.i.d. Lipschitz maps is the stochastic process resulting from $\Psi_n(x) := A_n x + B_n$, $x \in \mathbb{R}$, for i.i.d. $(A_n, B_n)_{n\geq 1}$, thus

$$X_n = A_n X_{n-1} + B_n$$

for $n \geq 1$, which is why this is called *random difference equation.* These processes often appear, among others, in time series analysis, e.g. as ARCH(1)-processes. In the case of Markov-modulated $(A_n, B_n)_{n\geq 1}$, i.e. equipped with a driving chain θ, related work has been done recently on the stability of these systems by Alsmeyer and Buckmann [AB17a] and by Buckmann [Buc16] in his

PhD thesis. They generalized results of Vervaat [Ver79] and Goldie and Maller [GM00] concerning the convergence of the forward and backward iterations. They used regeneration techniques and further results on fluctuation theory for MRWs from [AB17b]. For further references, including [Bra86] who examined stationary $(A_n, B_n)_{n\geq 1}$, we refer the reader to those listed in [AB17a].

5.2. Structure

In Chapter 6, we formalize what has been indicated in the introduction, including a definition of the model, measurability issues and duality as one key concept that we apply. In Section 6.3 we briefly recall Elton's theorem with a short proof and give a motivation for requiring *mean contractivity* in the subsequent Section 6.4. The latter contains a summary of Markov-renewal-theoretic tools to be used in the analysis, and also states two different sets of conditions for two different convergence-rate regimes.

After these preparations, Chapter 7 starts with the formulation of the main results of this part: First, Elton's theorem is re-derived for our special situation by regeneration techniques. Afterwards, given the conditions from the previous section, the corresponding convergence-rate results are established. Sections 7.2, 7.3 and 7.4 contain the respective proofs.

Some auxiliary results are collected in Appendix B.

6. Preliminaries

Let $(\mathbb{X}, d)$ be a complete separable metric space with Borel σ-field $\mathfrak{B}(\mathbb{X})$. Let $\mathrm{Lip}(\mathbb{X}, \mathbb{X})$ be the set of all Lipschitz continuous functions from $\mathbb{X}$ to $\mathbb{X}$. It is possible to equip this set with a measurable structure (cf. [DF99]) but we postpone the short discussion until after the specification of the central object of this part of the work. Throughout this part, we deal with the following situation: We define

$$X_n := \Psi_n(X_{n-1}) = \Psi_n \circ \cdots \circ \Psi_1(X_0) =: \Psi_n \cdots \Psi_1(X_0)$$

for $n \geq 1$ and call $(X_n)_{n\geq 0}$ or $(\theta_n, X_n)_{n\geq 0}$ *iterated function system of Markov-modulated Lipschitz maps* (MIFS) provided that:

(a) $(\theta, \Psi) := (\theta_n, \Psi_n)_{n\geq 0}$ is a MMS on $(\Omega, \mathcal{A}, \mathbb{P})$ with state space $\mathcal{S} \times \mathrm{Lip}(\mathbb{X}, \mathbb{X})$ and transition kernel

$$\mathbb{P}^{(\theta_{n+1},\Psi_{n+1})|\theta_n,\Psi_n} = Q(\theta_n, \cdot) \quad \mathbb{P}\text{-a.s.},$$

$n \geq 1$. It has driving chain θ which is assumed to be positive recurrent (and thus irreducible), time-homogeneous discrete Markov chain with (at most) countable state space $(\mathcal{S}, \mathfrak{S})$. Denote by $P = (p_{i,j})_{i,j\in\mathcal{S}}$ the transition matrix/kernel and by π the unique stationary distribution of θ, with $\pi_i > 0$ for all $i \in \mathcal{S}$, cf. Remark 2.6.

(b) X_0 is some $\mathbb{X}$-valued random variable on $(\Omega, \mathcal{A}, \mathbb{P})$ which is independent of (θ, Ψ).

Recall, that in Section 2.5 we introduced the concept of a MMS with a second component taking values in $\mathbb{R}$. This is not necessarily required, but every Borel space suffices. This is addressed in the next section.

6.1. Measurability and remarks

Measurability. We briefly discuss some measurability issues concerning random Lipschitz maps that we gather from [DF99, Section 5.1]. We refer the reader to their work for further details.

For a Lipschitz map f in $\mathrm{Lip}(\mathbb{X}, \mathbb{X})$, the mapping

$$L(f) := \sup_{x\neq y} \frac{d(f(x), f(y))}{d(x, y)} \geq 0$$

emits the corresponding Lipschitz constant. Let further $\mathbb{X}_0$ be a countable dense subset of $\mathbb{X}$ and let $\mathrm{Map}(\mathbb{X}_0, \mathbb{X})$ be the set of all maps from $\mathbb{X}_0$ to $\mathbb{X}$, endowed with the product topology and the product σ-field. Then, [DF99, Lemma 5.1] clears the way for our analysis: $\mathrm{Lip}(\mathbb{X}, \mathbb{X})$ is a Borel subset of $\mathrm{Map}(\mathbb{X}_0, \mathbb{X})$ with induced σ-field $\mathfrak{L}$, the map $\Psi \mapsto L(\Psi)$ is a Borel function on $(\mathrm{Lip}(\mathbb{X}, \mathbb{X}), \mathfrak{L})$, and the evaluation map $(\Psi, x) \mapsto \Psi(x)$ is a Borel function on $(\mathrm{Lip}(\mathbb{X}, \mathbb{X}) \times \mathbb{X}, \mathfrak{L} \otimes \mathfrak{B}(\mathbb{X}))$. We conclude with an easy finding:

Lemma 6.1. *Let* $\Psi, \Psi' \in \mathrm{Lip}(\mathbb{X}, \mathbb{X})$*, then* $L(\Psi \circ \Psi') \leq L(\Psi) \cdot L(\Psi')$.

Model remarks. We add some complementary remarks to the model description.

Remark 6.2. A direct consequence of (θ, Ψ) being a MMS is that $\Psi_0, \Psi_1, \dots$ are conditionally independent given θ with $\mathbb{P}(\Psi_0 \in \cdot|\theta) = \mathbb{P}(\Psi_0 \in \cdot|\theta_0)$ $\mathbb{P}$-a.s. and

$$\mathbb{P}(\Psi_n \in \cdot|\theta) = \mathbb{P}(\Psi_n \in \cdot|\theta_{n-1}, \theta_n) =: K(\theta_{n-1}, \theta_n, \cdot) \quad \mathbb{P}\text{-a.s.}$$

for all $n \geq 1$ and a stochastic kernel $K : \mathcal{S}^2 \times \mathfrak{L} \to [0,1]$. Moreover, we have (cf. [Als15b, Lemma 8.1, in Ger.])

$$Q(x, \mathrm{d}y \times \mathrm{d}z) = K(x, y, \mathrm{d}z)P(x, \mathrm{d}y),$$

more precisely, for $i, j \in \mathcal{S}$ and measurable B

$$Q(i, \{j\} \times B) = p_{i,j} \int_{\mathbb{X}} \mathbb{1}_B(z) K(i, j, \mathrm{d}z) = p_{i,j} K(i, j, B) =: p_{i,j} K_{i,j}(B).$$

Remark 6.3. If $\psi : (\mathcal{S} \times \mathbb{X}, \mathfrak{S} \otimes \mathfrak{B}(\mathbb{X})) \to (\mathbb{X}, \mathfrak{B}(\mathbb{X}))$ is a jointly measurable function which is Lipschitz continuous in the second component, then $\Psi_n := \psi(\theta_n, \cdot)$ is a random element in $\mathrm{Lip}(\mathbb{X}, \mathbb{X})$ and we are clearly in the above situation with $K_{i,j}(B) = \mathbb{1}_B(\psi(j, \cdot))$ for all $i \in \mathcal{S}$. Since $\mathcal{S}$ is countable, this model only allows for countably many different Lipschitz maps.

Remark 6.4. $(X_n)_{n\geq 0}$ itself is not a Markov chain since the next value depends on the current state of θ. However, $(\theta_n, X_n)_{n\geq 0}$ is a time-homogeneous Markov chain with state space $\mathcal{S} \times \mathbb{X}$ and transition kernel

$$\begin{aligned}
\bar{P}((i,x), \{j\} \times A) &= \mathbb{P}(\theta_n = j, X_n \in A|\theta_{n-1} = i, X_{n-1} = x) \\
&= \mathbb{P}(\theta_n = j, \Psi_n(x) \in A|\theta_{n-1} = i, X_{n-1} = x) \\
&= \mathbb{P}(\theta_n = j, \Psi_n(x) \in A|\theta_{n-1} = i) \\
&= p_{i,j}\mathbb{P}(\Psi_n(x) \in A|\theta_n = j, \theta_{n-1} = i) \\
&= p_{i,j}\mathbb{P}(\Psi_n \in p_x^{-1}(A)|\theta_n = j, \theta_{n-1} = i) = p_{i,j} K_{i,j}(p_x^{-1}(A))
\end{aligned}$$

for $i, j \in \mathcal{S}$, $x \in \mathbb{X}$ and $A \in \mathfrak{B}(\mathbb{X})$. Here $p_x : \Psi_n(\omega) \mapsto \Psi_n(\omega)(x)$ is measurable, because the evaluation map $(\Psi, x) \mapsto \Psi(x)$ from $\mathrm{Lip}(\mathbb{X}, \mathbb{X}) \times \mathbb{X}$ to $\mathbb{X}$ is measurable as seen in Section 6.1.

6.2. Dual chain and distributional identity

Dual chain. For our analysis, we introduce a MMS $({}^{\#}\theta, {}^{\#}\Psi) := ({}^{\#}\theta_n, {}^{\#}\Psi_n)_{n\geq 0}$ on $(\Omega, \mathcal{A}, \mathbb{P})$, with ${}^{\#}\theta_0 := \theta_0$, which is dual to (θ, Ψ), i.e. it has the dual transition kernel

$$ {}^{\#}Q(i, \{j\} \times B) = \frac{\pi_j p_{j,i}}{\pi_i} K(j, i, B) =: {}^{\#}p_{i,j} {}^{\#}K(i, j, B) =: {}^{\#}p_{i,j} {}^{\#}K_{i,j}(B) \tag{6.1}$$

for $i, j \in \mathcal{S}$ and measurable B, where ${}^{\#}P = ({}^{\#}p_{i,j}) = (\frac{\pi_j p_{j,i}}{\pi_i})$ denotes the transition matrix/kernel of ${}^{\#}\theta$. Set $\mathbb{P}_i := \mathbb{P}(\cdot|\theta_0 = {}^{\#}\theta_0 = i)$ and $\mathbb{P}_\lambda = \sum_{i\in\mathcal{S}} \lambda_i \mathbb{P}_i$ for a probability measure λ on $\mathcal{S}$. It is easy to show that the original and the dual chain are connected by the relations

$$\pi_{j_0}\mathbb{P}_{j_0}(\theta_1 = j_1, \dots, \theta_k = j_k) = \pi_{j_k}\mathbb{P}_{j_k}({}^{\#}\theta_1 = j_{k-1}, \dots, {}^{\#}\theta_k = j_0) \tag{6.2}$$

and

$$\mathbb{P}((\Psi_1, \dots, \Psi_k) \in \cdot|\theta_0 = j_0, \dots, \theta_k = j_k) = \mathbb{P}(({}^{\#}\Psi_k, \dots, {}^{\#}\Psi_1) \in \cdot|{}^{\#}\theta_0 = j_k, \dots, {}^{\#}\theta_k = j_0) \tag{6.3}$$

for every $k \geq 1$ and $j_0, \dots j_k \in \mathcal{S}$.

Remark 6.5. The dual driving chain ${}^{\#}\theta$ inherits all regularity assumptions from θ. It is time-homogeneous and irreducible, and also positive recurrent with the same stationary distribution π and mean recurrence time $\mathbb{E}_i {}^{\#}\sigma_1(i) = \mathbb{E}_i \sigma_1(i) =: m_{ii}$, where $\sigma_1(i) := \inf\{n \geq 1 : \theta_n = i\}$ and analogously ${}^{\#}\sigma_1(i) := \inf\{n \geq 1 : {}^{\#}\theta_n = i\}$.

We briefly record easy consequences of (6.2) and (6.3) that we will use frequently:

Lemma 6.6. *It holds that*

(a) ${}^{\#}\sigma_1(i) \overset{\mathrm{d}}{=} \sigma_1(i)$ *w.r.t.* $\mathbb{P}_i$ *for all* $i \in \mathcal{S}$.

(b) $(\theta_0, \ldots, \theta_n, \Psi_1, \ldots, \Psi_n) \overset{\mathrm{d}}{=} ({}^{\#}\theta_n, \ldots, {}^{\#}\theta_0, {}^{\#}\Psi_n, \ldots, {}^{\#}\Psi_1)$ *w.r.t.* $\mathbb{P}_\pi$ *for all* $n \geq 1$.

(c) $(\Psi_1, \ldots, \Psi_{\sigma_1(i)}) \overset{\mathrm{d}}{=} ({}^{\#}\Psi_{{}^{\#}\sigma_1(i)}, \ldots, {}^{\#}\Psi_1)$ *w.r.t.* $\mathbb{P}_i$ *for all* $i \in \mathcal{S}$.

Resulting from the above lemma, we obtain a lot of helpful identities linking the Lipschitz maps with their duals. In particular, we will make frequent use of the following: $\mathbb{E}_\pi f(\Psi_1) = \mathbb{E}_\pi f({}^{\#}\Psi_1)$ and $\mathbb{E}_\pi f(\Psi_n \cdots \Psi_1) = \mathbb{E}_\pi f({}^{\#}\Psi_1 \cdots {}^{\#}\Psi_n)$ for measurable f, whenever either expression is well-defined.

The dual chain accounts for the time-reversal idea in [Elt90]. In the i.i.d. setting, all arrangements of a certain number of Lipschitz maps have the same distribution, so it is not immediately apparent, that the natural perspective should be, to build not only the backward iteration, but the backward iteration of the time reversal. Part (b) of the above Lemma and the next paragraph justify this.

Central distributional identity. We return to the sequence $(\theta_n, X_n)_{n \geq 0}$ and additionally introduce the notation

$$X_n^x := \Psi_n \cdots \Psi_1(x), \quad X_0^x := x,$$

according to which $(\theta_n, X_n^x)_{n \geq 0}$ is a time-homogeneous Markov chain, too, with the same transition kernel. From now on, we will focus on the process with deterministic starting value $x \in \mathbb{X}$. Also, the dot $(\cdot)$ will denote the concatenation of functions throughout this work whenever it is the obvious interpretation. Referring to the approach in the i.i.d. setting, we introduce the four-part scheme (instead of two-part in the i.i.d. setting)

$$\begin{aligned} X_n^x &= \Psi_n \cdots \Psi_1(x) & \widehat{X}_n^x &:= \Psi_1 \cdots \Psi_n(x) \\ {}^{\#}X_n^x &:= {}^{\#}\Psi_n \cdots {}^{\#}\Psi_1(x) & {}^{\#}\widehat{X}_n^x &:= {}^{\#}\Psi_1 \cdots {}^{\#}\Psi_n(x). \end{aligned}$$

Obviously, the hashtag (${}^{\#}$) indicates that the dual chain is considered and the hat ($\widehat{\ }$) denotes the backward iteration corresponding to the (non-hatted) forward iteration of the respective chain. Denoting forward and backward iteration in this way is consistent with [AF01].

The following distributional identity is the main reason for introducing the dual chain: To obtain distributional results for forward iterations of the original MIFS, we can also examine the backward iterations of the "dual MIFS", which, as in the i.i.d. case, turn out to be much easier to analyze. It follows immediately from Lemma 6.6.

Lemma 6.7. *For all* $n \geq 0$ *and* $x \in \mathbb{X}$, $(\theta_0, \ldots, \theta_n, X_n^x) \overset{\mathrm{d}}{=} ({}^{\#}\theta_n, \ldots, {}^{\#}\theta_0, {}^{\#}\widehat{X}_n^x)$ *w.r.t.* $\mathbb{P}_\pi$. *In particular,* $X_n^x \overset{\mathrm{d}}{=} {}^{\#}\widehat{X}_n^x$ *w.r.t.* $\mathbb{P}_\pi$ *for all* $n \geq 0$ *and* $x \in \mathbb{X}$.

6.3. Elton's theorem

The next section contains parts of the original theorem by Elton from [Elt90, Theorem 3] in our setting with the more ore less original proof for completeness. A different proof is prepared in the next section and accomplished in the next chapter, where the actual work starts. As a slight motivation for the assumptions that we make, we state a version of the Furstenberg-Kesten theorem from [FK60], first. In our setting, the result is due to Elton [Elt90].

Theorem 6.8 (Furstenberg-Kesten, Elton)**.** *Let* $\mathbb{E}_\pi \log^+ L(\Psi_1) < \infty$. *Then*

$$\begin{aligned}&\lim_{n\to\infty} \frac{1}{n} \log L({}^{\#}\Psi_1 \cdots {}^{\#}\Psi_n) = \inf_{n\geq 1} n^{-1}\mathbb{E}_\pi \log L({}^{\#}\Psi_1 \cdots {}^{\#}\Psi_n)\\ &= \lim_{n\to\infty} \frac{1}{n} \log L(\Psi_n \cdots \Psi_1) = \inf_{n\geq 1} n^{-1}\mathbb{E}_\pi \log L(\Psi_n \cdots \Psi_1) =: \chi \quad \mathbb{P}_\pi\text{-a.s.}\end{aligned}$$

with $\chi \in \mathbb{R} \cup \{-\infty\}$. *If* $\chi \in \mathbb{R}$, *then the convergence also holds in* $L^1(\mathbb{P}_\pi)$. χ *is called* Lyapunov exponent.

Proof. We use a generalized version of *Kingman's subadditive ergodic theorem* by [Lig85] in the form of [Als15a], cf. Theorem B.1 in the appendix. The Theorem applies to the triangular schemes $Y_{k,n} := \log L({}^{\#}\Psi_{k+1} \cdots {}^{\#}\Psi_n)$ and $Y'_{k,n} := \log L(\Psi_n \cdots \Psi_{k+1})$, $n \geq 1$ and $0 \leq k \leq n$. It is easy to see that both schemes satisfy (SA-1)-(SA-4) and the sequences in (SA-2) are also ergodic under $\mathbb{P}_\pi$. Noting

$$\inf_{n\geq 1} n^{-1}\mathbb{E}_\pi \log L({}^{\#}\Psi_1 \cdots {}^{\#}\Psi_n) = \inf_{n\geq 1} n^{-1}\mathbb{E}_\pi \log L(\Psi_n \cdots \Psi_1)$$

completes the proof. □

The following theorem is part of what was proven by Elton in [Elt90] for general stationary $(\Psi_n)_{n\geq 1}$ with a.s. negative Lyapunov exponent. At this point we keep this condition, but starting with the next section, we will slightly strengthen it to $\mathbb{E}_\pi \log L(\Psi_1) < 0$ for our purposes. We justify why the latter causes the Lyapunov exponent to be a.s. negative at the beginning of Section 6.4. We recall the proof of Elton's result in our setting as it illustrates the influence of the Lyapunov exponent on the convergence of the backward iterations.

Theorem 6.9. *Given* $\mathbb{E}_\pi \log^+ L(\Psi_1) < \infty$ *and* $\mathbb{E}_\pi \log^+ d(x_0, \Psi_1(x_0)) < \infty$ *for some (and thus all)* $x_0 \in \mathbb{X}$, *let* $\chi < 0$ $\mathbb{P}_\pi$*-a.s. Then there exists a* ${}^{\#}\widehat{X}_\infty$ *with the following properties:*

(a) As $n \to \infty$

$${}^{\#}\widehat{X}_n^x \to {}^{\#}\widehat{X}_\infty \quad \mathbb{P}_\pi\text{-a.s. for every } x \in \mathbb{X}. \tag{6.4}$$

(b) As $n \to \infty$

$$(\theta_n, X_n^x) \xrightarrow{\mathrm{d}} (\theta_0, {}^{\#}\widehat{X}_\infty) \quad \text{w.r.t. } \mathbb{P}_\pi \text{ for all } x \in \mathbb{X}. \tag{6.5}$$

In particular, as $n \to \infty$,

$$X_n^x \xrightarrow{\mathrm{d}} {}^{\#}\widehat{X}_\infty \quad \text{w.r.t. } \mathbb{P}_\pi \text{ for all } x \in \mathbb{X}. \tag{6.6}$$

Remark 6.10. Up to now, many conditions are formulated in terms of $\log^+$. However, this function has not the good properties that one would desire. Hence, we will soon (already in the next remark) proceed to use $\log_*(x) := \log(1 + x)$ as a subadditive majorant of $\log^+$. It also satisfies $\log_*(xy) \leq \log_*(x) + \log_*(y)$ for all $x, y \geq 0$ and $\log_*(x) \leq 1 + \log^+(x)$ for $x \geq 0$.

Remark 6.11. Suppose that $\mathbb{E}_\pi \log^+ d(x_0, \Psi_1(x_0)) < \infty$ for some $x_0 \in \mathbb{X}$. Then for every other $x \in \mathbb{X}$, we have

$$\begin{aligned} d(x, \Psi_1(x)) &\le d(x, x_0) + d(x_0, \Psi_1(x_0)) + d(\Psi_1(x_0), \Psi_1(x)) \\ &\le (1 + L(\Psi_1)) \cdot d(x, x_0) + d(x_0, \Psi_1(x_0)). \end{aligned} \tag{6.7}$$

With the help of Remark 6.10, we get

$$\begin{aligned} \mathbb{E}_\pi \log^+ d(x, \Psi_1(x)) &\le \mathbb{E}_\pi \log_* d(x, \Psi_1(x)) \\ &\le \mathbb{E}_\pi \log_*(1 + L(\Psi_1)) + \log_* d(x, x_0) + \mathbb{E}_\pi \log_* d(x_0, \Psi_1(x_0)) \\ &\le \log_*(1) + \mathbb{E}_\pi \log_* L(\Psi_1) + \log_* d(x, x_0) + \mathbb{E}_\pi \log_* d(x_0, \Psi_1(x_0)) \end{aligned}$$

which is finite by assumption since $\log_* \le 1 + \log^+$.

Proof of Theorem 6.9. The Furstenberg-Kesten theorem implies

$$\lim_{n\to\infty} \frac{1}{n} \log L({}^\#\Psi_1 \cdots {}^\#\Psi_n) = \chi \quad \mathbb{P}_\pi\text{-a.s.} \tag{6.8}$$

Fix $x_0 \in \mathbb{X}$. Then

$$d({}^\#\widehat{X}^{x_0}_{n+1}, {}^\#\widehat{X}^{x_0}_n) \le L({}^\#\Psi_1 \cdots {}^\#\Psi_n) \cdot d(x_0, {}^\#\Psi_{n+1}(x_0)).$$

We show that $\sum_{n=1}^\infty d({}^\#\widehat{X}^{x_0}_{n+1}(\omega), {}^\#\widehat{X}^{x_0}_n(\omega)) < \infty$ for $\mathbb{P}_\pi$-almost all ω. Then $({}^\#\widehat{X}^{x_0}_n(\omega))_{n\ge 0}$ is a Cauchy sequence and thus converges ($\mathbb{X}$ is complete) to some ${}^\#\widehat{X}^{x_0}_\infty(\omega)$. To this end, fix $\omega \in \Omega_j \cap A_1$, where (6.8) holds on A_1 and $\mathbb{P}_\pi(A_1) = 1$. Then, there clearly is an $n_0(\omega) \in \mathbb{N}$ such that for all $n \ge n_0(\omega)$ the Lipschitz constants satisfy

$$L({}^\#\Psi_1 \cdots {}^\#\Psi_n)(\omega) < e^{-\frac{\alpha}{2}n}$$

for $\alpha = -\chi$ if $\chi > -\infty$ and some $\alpha > 0$ otherwise. Moreover,

$$\begin{aligned} \sum_{n\ge 1} \mathbb{P}_\pi\left(\log^+ d(x_0, {}^\#\Psi_{n+1}(x_0)) > \frac{\alpha}{4}n\right) &\le \mathbb{E}_\pi\left(\frac{4}{\alpha}\log^+ d(x_0, {}^\#\Psi_{n+1}(x_0))\right) \\ &= \mathbb{E}_\pi\left(\frac{4}{\alpha}\log^+ d(x_0, \Psi_1(x_0))\right) < \infty \end{aligned}$$

by stationarity, so the Borel-Cantelli lemma implies

$$\mathbb{P}_\pi(A_2) := \mathbb{P}_\pi\left(\log^+ d(x_0, {}^\#\Psi_{n+1}(x_0)) \le \frac{\alpha}{4}n \text{ for almost all } n \ge 1\right) = 1.$$

Hence, for $\omega \in A_2$ there is an $n_1(\omega)$ such that $\log^+ d(x_0, {}^\#\Psi_{n+1}(x_0))(\omega) \le \frac{\alpha}{4}n$ for all $n \ge n_1(\omega)$, and thus $d(x_0, {}^\#\Psi_{n+1}(x_0)) \le e^{\frac{\alpha}{4}n}$. For $\omega \in A_1 \cap A_2$ with $n_2(\omega) := \max\{n_0(\omega), n_1(\omega)\}$, we finally have that

$$d({}^\#\widehat{X}^{x_0}_{n+1}(\omega), {}^\#\widehat{X}^{x_0}_n(\omega)) \le L({}^\#\Psi_1 \cdots {}^\#\Psi_n)(\omega) \cdot d(x_0, {}^\#\Psi_{n+1}(x_0))(\omega) \le e^{-\frac{\alpha}{2}n} \cdot e^{\frac{\alpha}{4}n} = e^{-\frac{\alpha}{4}n}$$

for all $n \ge n_2(\omega)$. This is clearly summable and since $\mathbb{P}_\pi(A_1 \cap A_2) = 1$, this ends the first part of the proof. Given ${}^\#\widehat{X}^x_\infty$ for each $x \in \mathbb{X}$, we still need to verify (6.4) which means to check, whether ${}^\#\widehat{X}^x_\infty = {}^\#\widehat{X}^y_\infty$ $\mathbb{P}_\pi$-a.s. for all x, y. Therefore, we remark that

$$d({}^\#\widehat{X}^x_n, {}^\#\widehat{X}^y_n) \le L({}^\#\Psi_1 \cdots {}^\#\Psi_n) \cdot d(x, y) \to 0 \quad \mathbb{P}_\pi\text{-a.s.},$$

as $n \to \infty$. Thus all ${}^{\#}\widehat{X}_n^x$ converge $\mathbb{P}_\pi$-a.s. to the same random variable ${}^{\#}\widehat{X}_\infty := {}^{\#}\widehat{X}_\infty^{x_0}$, say. This proves (a).

To prove (b), let $\bar{\pi}(B) := \mathbb{P}_\pi((\theta_0, {}^{\#}\widehat{X}_\infty) \in B)$ and $f \in \mathcal{C}_b(\mathcal{S} \times \mathbb{X})$ be a bounded and continuous function on $\mathcal{S} \times \mathbb{X}$ with values in $\mathbb{R}$. Then

$$\begin{aligned}\int f \, \mathrm{d}\mathbb{P}_\pi^{(\theta_n, X_n^x)} &= \int f(\theta_n, \Psi_n \cdots \Psi_1(x)) \, \mathrm{d}\mathbb{P}_\pi = \int f({}^{\#}\theta_0, {}^{\#}\Psi_1 \cdots {}^{\#}\Psi_n(x)) \, \mathrm{d}\mathbb{P}_\pi \\ &= \int f({}^{\#}\theta_0, {}^{\#}\widehat{X}_n^x) \, \mathrm{d}\mathbb{P}_\pi \to \int f({}^{\#}\theta_0, {}^{\#}\widehat{X}_\infty) \, \mathrm{d}\mathbb{P}_\pi = \int f \, \mathrm{d}\bar{\pi},\end{aligned}$$

as $n \to \infty$, by duality, the dominated convergence theorem and ${}^{\#}\theta_0 = \theta_0$.

□

6.4. Markov renewal theory and conditions

One key requirement in Theorem 6.9 is the negative Lyapunov exponent χ. Slightly strengthening this by requiring the *mean contraction condition*

$$\mathbb{E}_\pi \log L(\Psi_1) < 0 \qquad \text{(MC1)}$$

to hold together with the *jump-size condition*

$$\mathbb{E}_\pi \log^+ d(x_0, \Psi_1(x_0)) < \infty \quad \text{for some (and thus all) } x_0 \in \mathbb{X}, \qquad \text{(MC2)}$$

allows us to easily reproduce (6.4) (and thus (6.6)) by cyclic decomposition. On the one hand, this illustrates a further intuitive approach to Elton's result in the spirit of [AF01], and on the other hand, this procedure founds the basis for our main goal, to find rates of convergence for (6.6) in two different regimes which will be introduced later on. From now on, both conditions given above are always in force.

It is easy to see that

$$L(\Psi_n \cdots \Psi_1) \le \prod_{k=1}^n L(\Psi_k) \quad \text{and} \quad L({}^{\#}\Psi_1 \cdots {}^{\#}\Psi_n) \le \prod_{k=1}^n L({}^{\#}\Psi_k)$$

for $n \ge 2$. Especially the latter inequality provides us with a leverage point for our analysis and also suggests a set of tools to use: Obviously, the sequence

$$\Big({}^{\#}\theta_n, \sum_{k=1}^n \log L({}^{\#}\Psi_k)\Big)_{n \ge 0}$$

is a MRW. We set $S_n := \sum_{k=1}^n \log L({}^{\#}\Psi_k)$, $n \ge 1$, and $S_0 := 0$. Since π is also the unique stationary distribution of ${}^{\#}\theta$, the MRW has negative stationary drift

$$\mu := \mathbb{E}_\pi \log L({}^{\#}\Psi_1) = \mathbb{E}_\pi \log L(\Psi_1) < 0.$$

Throughout our work, we will replace the increments of the additive component by $\log_* L({}^{\#}\Psi_k)$ or $\log_* d(x_0, {}^{\#}\Psi_k(x_0))$, say. These processes are also MRWs. Finally, we remark that mean contractivity does indeed imply an a.s. negative Lyapunov exponent since (MC1) and Theorem 3.16 together yield $\chi \le \mathbb{E}_\pi \log L(\Psi_1)$ $\mathbb{P}_\pi$-a.s.

$$\begin{aligned}
X_{\sigma_n(i)}\colon &\quad \left({}^{\#}\Psi_{\sigma_n(i)}\cdots{}^{\#}\Psi_{\sigma_{n-1}(i)+1}\right)\left({}^{\#}\Psi_{\sigma_{n-1}(i)}\cdots{}^{\#}\Psi_{\sigma_{n-2}(i)+1}\right)\cdots\left({}^{\#}\Psi_{\sigma_1(i)}\cdots{}^{\#}\Psi_1\right)\\
Y_n\colon &\quad \left({}^{\#}\Psi_{\sigma_{n-1}(i)+1}\cdots{}^{\#}\Psi_{\sigma_n(i)}\right)\left({}^{\#}\Psi_{\sigma_{n-2}(i)+1}\cdots{}^{\#}\Psi_{\sigma_{n-1}(i)}\right)\cdots\left({}^{\#}\Psi_1\cdots{}^{\#}\Psi_{\sigma_1(i)}\right)\\
\widehat{Y}_n=\widehat{X}_{\sigma_n(i)}\colon &\quad \left({}^{\#}\Psi_1\cdots{}^{\#}\Psi_{\sigma_1(i)}\right)\cdots\left({}^{\#}\Psi_{\sigma_{n-2}(i)+1}\cdots{}^{\#}\Psi_{\sigma_{n-1}(i)}\right)\left({}^{\#}\Psi_{\sigma_{n-1}(i)+1}\cdots{}^{\#}\Psi_{\sigma_n(i)}\right)
\end{aligned}$$

Figure 6.1.: Comparison of $X_{\sigma_n(i)}$, Y_n and $\widehat{Y}_n=\widehat{X}_{\sigma_n(i)}$.

Cyclic decomposition of the MIFS. Let $(\sigma_n(i))_{n\ge 0}$ and $({}^{\#}\sigma_n(i))_{n\ge 0}$ denote the successive recurrence times of the driving chain θ and ${}^{\#}\theta$, respectively, as defined in Section 2.5. We already set $m_{ii}:=\mathbb{E}_i{}^{\#}\sigma_1(i)$. Now, $(S_{{}^{\#}\sigma_n(i)})_{n\ge 0}$ forms a SRW under $\mathbb{P}_i$ with negative drift

$$\mathbb{E}_i S_{{}^{\#}\sigma_1(i)}=\mu\cdot\mathbb{E}_i{}^{\#}\sigma_1(i)=\mu\cdot m_{ii}<0.$$

Whenever we only talk about the dual chain (as will be the case most of the time), we abuse notation and suppress the $^{\#}$ in the name for simplicity. Which chain is used, will be clear from the context. We refer to Section 2.5 for more details.

The following lemma is essential for our further approach as it relates our MIFS to an embedded IFS of i.i.d. Lipschitz maps for which we know a lot of results. It most importantly states that we can identify an i.i.d. IFS (Y_n^x) in our model whose backward iterations equal ${}^{\#}\widehat{X}^x_{\sigma_n(i)}$. We illustrate how Y_n^x relates to $X_{\sigma_n(i)}$ in Figure 6.1. This figure is analogous to [Als15a, Figure 3.2]. Lemma 6.12 is very similar to [Als15a, Lemma 3.26] except for the fact that we consider recurrence times instead of ladder epochs. We will frequently have this kind of analogy to [AF01] and [Als15a] in our lemmas, and still we also need to deal with starting according to the stationary measure π.

Here and in the rest of this work, we will often use $i\in\mathcal{S}$ as a reference state.

Lemma 6.12. *Given a MIFS satisfying* (MC1) *and* (MC2)*, the embedded sequence* $({}^{\#}X^x_{\sigma_n(i)})_{n\ge 0}$ *forms a mean contractive IFS of i.i.d. Lipschitz maps under* $\mathbb{P}_i$ *satisfying the jump-size condition*

$$\mathbb{E}_i\log^+ d(x_0,{}^{\#}\Psi_1^i(x_0))<\infty$$

for some (and thus all) $x_0\in\mathbb{X}$*, with corresponding Lipschitz maps*

$${}^{\#}\Psi_n^i:={}^{\#}\Psi_{\sigma_n(i)}\cdots{}^{\#}\Psi_{\sigma_{n-1}(i)+1},\quad n\ge 1.$$

The same holds for the sequence $(Y_n^x)_{n\ge 0}$*, defined by* $Y_0^x:=x$ *and*

$$Y_n^x:={}^{\#}\vec{\Psi}_n^i\cdots{}^{\#}\vec{\Psi}_1^i(x)$$

with corresponding Lipschitz maps

$${}^{\#}\vec{\Psi}_n^i:={}^{\#}\Psi_{\sigma_{n-1}(i)+1}\cdots{}^{\#}\Psi_{\sigma_n(i)},\quad n\ge 1.$$

Proof. We note that the ${}^{\#}\Psi_n^i$ are i.i.d. under $\mathbb{P}_i$ and also ${}^{\#}X_{\sigma_n(i)}={}^{\#}\Psi_n^i({}^{\#}X_{\sigma_{n-1}(i)})$. The same holds for the ${}^{\#}\vec{\Psi}_n^i$ which are the corresponding Lipschitz maps of $(Y_n^x)_{n\ge 1}$. First, we verify the jump-size condition. For simplicity, we put $S_n^*:=\sum_{k=1}^n\log_* L({}^{\#}\Psi_k)\ge 0$, $n\ge 1$, $S_0^*:=0$. Then

$$\log d(x_0,{}^{\#}\vec{\Psi}_1^i(x_0))=\log d(x_0,{}^{\#}\Psi_1\cdots{}^{\#}\Psi_{\sigma_1(i)}(x_0))$$

$$
\begin{aligned}
&\leq \log\left(d(x_0, {}^{\#}\Psi_1(x_0)) + \sum_{n=2}^{\sigma_1(i)} d({}^{\#}\Psi_1 \cdots {}^{\#}\Psi_{n-1}(x_0), {}^{\#}\Psi_1 \cdots {}^{\#}\Psi_n(x_0))\right)\\
&\leq \log\left(d(x_0, {}^{\#}\Psi_1(x_0)) + \sum_{n=2}^{\sigma_1(i)} \prod_{k=1}^{n-1} L({}^{\#}\Psi_k) \cdot d(x_0, {}^{\#}\Psi_n(x_0))\right) \qquad (6.9)\\
&\leq \log\left(d(x_0, {}^{\#}\Psi_1(x_0)) + \sum_{n=2}^{\sigma_1(i)} e^{S^*_{n-1}} d(x_0, {}^{\#}\Psi_n(x_0))\right) = \log\left(\sum_{n=1}^{\sigma_1(i)} e^{S^*_{n-1}} d(x_0, {}^{\#}\Psi_n(x_0))\right)\\
&\leq \log\left(e^{S^*_{\sigma_1(i)}} \cdot \sum_{n=1}^{\sigma_1(i)} d(x_0, {}^{\#}\Psi_n(x_0))\right) \leq S^*_{\sigma_1(i)} + \sum_{n=1}^{\sigma_1(i)} \log_* d(x_0, {}^{\#}\Psi_n(x_0))
\end{aligned}
$$

where the subadditivity of $\log_*$ was used. We infer from (2.1) that

$$
\begin{aligned}
\mathbb{E}_i \log^+ d(x_0, {}^{\#}\vec{\Psi}_1^i(x_0)) &\leq \mathbb{E}_i S^*_{\sigma_1(i)} + \mathbb{E}_i \left(\sum_{n=1}^{\sigma_1(i)} \log_* d(x_0, {}^{\#}\Psi_n(x_0))\right)\\
&= m_{ii} \cdot \Big(\mathbb{E}_\pi \log_* L({}^{\#}\Psi_1) + \mathbb{E}_\pi \log_* d(x_0, {}^{\#}\Psi_1(x_0))\Big).
\end{aligned}
$$

All objects on the right side are finite by assumption. A similar procedure leads to the estimate

$$
\log d(x_0, {}^{\#}\Psi_1^i(x_0)) \leq S^*_{\sigma_1(i)} + \sum_{n=1}^{\sigma_1(i)} \log_* d(x_0, {}^{\#}\Psi_n(x_0)).
$$

If we replace x_0 by some other $x \in \mathbb{X}$ in both cases, then we conclude in the same way that the expectation is finite since $\mathbb{E}_\pi \log_* d(x, \Psi_1(x)) < \infty$ for all $x \in \mathbb{X}$. The mean contractivity in both cases follows from $\mathbb{E}_i \log L({}^{\#}\Psi_1^i) \leq \mathbb{E}_i S_{\sigma_1(i)}$ and $\mathbb{E}_i \log L({}^{\#}\vec{\Psi}_1^i) \leq \mathbb{E}_i S_{\sigma_1(i)}$. □

Conditions for convergence rates. For the results concerning convergence rates, we will impose two sets of moment conditions. These are gathered in this paragraph and are influenced by the previous lemma and the corresponding conditions in the i.i.d. setting, cf. [AF01, (1.8) and (1.9)]. Essentially, our conditions guarantee the latter two conditions for the embedded i.i.d. IFS $(Y_n^x)_{n\geq 0}$ and add a regularity constraint for the driving Markov chain ${}^{\#}\theta$. The main advantage of the below conditions is that they only involve quantities which are given by the model. However, we will mostly work with the dual chain and hence need conditions involving only the dual chain. Indeed, analogous conditions hold for the dual counterpart by duality. We record this fact in Remark 6.15.

The first set of conditions is tailored to result in polynomial convergence of the MIFS: We say that a MIFS satisfying (MC1) and (MC2) obeys the *polynomial-type moment conditions of order* $p > 0$ if the following conditions

$$
\mathbb{E}_i \sigma_1(i)^{p+1} < \infty, \qquad \text{(A1)}
$$

$$
\mathbb{E}_i \left(\sum_{n=1}^{\sigma_1(i)} \log_* L(\Psi_n)\right)^{p+1} < \infty, \qquad \text{(A2)}
$$

and

$$
\mathbb{E}_i \left(\sum_{n=1}^{\sigma_1(i)} \log_* d(x_0, \Psi_n(x_0))\right)^{p+1} < \infty \qquad \text{(A3)}
$$

hold for some $i \in \mathcal{S}$ and some (and thus all) $x_0 \in \mathbb{X}$. Note that $\sigma_1(i)$ is the first recurrence time of state i of the original chain θ.

Remark 6.13. If (A1), (A2) hold and (A3) holds for some $x_0 \in \mathbb{X}$, then (A3) holds for all $x \in \mathbb{X}$. This can be seen as follows. Analogously to (6.7), we obtain

$$d(x, \Psi_n(x)) \leq (1 + L(\Psi_n))d(x, x_0) + d(x_0, \Psi_n(x_0))$$

and

$$\log_* d(x, \Psi_n(x)) \leq \log_*(1) + \log_* d(x, x_0) + \log_* L(\Psi_n) + \log_* d(x_0, \Psi_n(x_0)),$$

and hence

$$\begin{aligned}\mathbb{E}_i \left(\sum_{n=1}^{\sigma_1(i)} \log_* d(x, \Psi_n(x)) \right)^{p+1} &\leq \mathbb{E}_i \left(\sum_{n=1}^{\sigma_1(i)} \log_* d(x_0, \Psi_n(x_0)) + \sum_{n=1}^{\sigma_1(i)} \log_* L(\Psi_n) + C\sigma_1(i) \right)^{p+1} \\ &\leq C'\mathbb{E}_i \left(\sum_{n=1}^{\sigma_1(i)} \log_* d(x_0, \Psi_n(x_0)) \right)^{p+1} + C'\mathbb{E}_i \left(\sum_{n=1}^{\sigma_1(i)} \log_* L(\Psi_n) \right)^{p+1} \\ &\quad + C'\mathbb{E}_i \sigma_1(i)^{p+1} < \infty\end{aligned}$$

with positive constants C and C'.

Remark 6.14. It is expected that solidarity in $i \in \mathcal{S}$ holds in (A1)-(A3) but this is not important for the following analysis. Concerning (A1), this is well-known.

Remark 6.15. Conditions (A1), (A2) and (A3) also hold with $\sigma_1(i)$ and Ψ_n replaced by ${}^{\#}\sigma_1(i)$ and ${}^{\#}\Psi_n$. This follows easily from Lemma 6.6.

The second set of conditions is tailored to result in geometric convergence of the MIFS. We say that a MIFS satisfying (MC1) and (MC2) obeys the *geometric-type moment conditions* if the following conditions

$$\mathbb{E}_i e^{\beta_i^* \sigma_1(i)} < \infty \text{ for some } \beta_i^* > 0, \tag{B1}$$

$$\mathbb{E}_i \left(\max_{1 \leq n \leq \sigma_1(i)} L(\Psi_n \cdots \Psi_1)^p + \max_{1 \leq n \leq \sigma_1(i)} L(\Psi_{\sigma_1(i)} \cdots \Psi_n)^p \right) < \infty, \tag{B2}$$

$$\mathbb{E}_i \left(\sum_{n=1}^{\sigma_1(i)} L(\Psi_n) \right)^p < \infty, \tag{B3}$$

and

$$\mathbb{E}_i \left(\sum_{n=1}^{\sigma_1(i)} d(x_0, \Psi_n(x_0)) \right)^p < \infty \tag{B4}$$

hold for some $i \in \mathcal{S}$, some $p > 0$ and some (and thus all) $x_0 \in \mathbb{X}$.

Remark 6.16. If (B1), (B3) hold and (B4) holds for some x_0, then the latter holds for every $x \in \mathbb{X}$. Again,

$$d(x, \Psi_n(x)) \leq d(x, x_0)(1 + L(\Psi_n)) + d(x_0, \Psi_n(x_0))$$

implies

$$\mathbb{E}_i \left(\sum_{n=1}^{\sigma_1(i)} d(x, \Psi_n(x)) \right)^p \leq C\mathbb{E}_i \sigma_1(i)^p + C\mathbb{E}_i \left(\sum_{n=1}^{\sigma_1(i)} L(\Psi_n) \right)^p + C\mathbb{E}_i \left(\sum_{n=1}^{\sigma_1(i)} d(x_0, \Psi_n(x_0)) \right)^p < \infty$$

for a constant $C > 0$.

Remark 6.17. See Remark 6.14 for a remark concerning solidarity in (B1)-(B4).

Remark 6.18. Conditions (B1), (B3) and (B4) (and (B2*) below) also hold with $\sigma_1(i)$ and Ψ_n replaced by ${}^{\#}\sigma_1(i)$ and ${}^{\#}\Psi_n$. This follows easily from Lemma 6.6. The lemma further yields that (B2) is equivalent to $\mathbb{E}_i(G_1^p + \bar{G}_1^p) < \infty$, with G_1 and $\bar{G}_1$ as in (7.12) and (B.2), respectively.

Remark 6.19. Condition (B2) and Remark 6.18 ensure

$$\mathbb{E}_i L({}^{\#}\vec{\Psi}_1^i)^p \leq \mathbb{E}_i G_1^p < \infty$$

and since the embedded IFS (Y_n^x) of i.i.d. Lipschitz maps is mean contractive, we can find a $q \leq p$ such that it is even strongly mean contractive of order q, i.e. $\mathbb{E}_i L({}^{\#}\vec{\Psi}_1^i)^q < 1$.

Remark 6.20. It is easy to show similarly to (7.19) that

$$\mathbb{E}_i \left(\prod_{n=1}^{\sigma_1(i)} (1 + L(\Psi_n)) \right)^p < \infty, \tag{B2*}$$

implies (B2) (and also (B3) which is easily verified using the subadditivity of $\log_*$). However, (B2*) seems to be a bit to restrictive.

Remark 6.21. For a geometric MIFS, we require the existence of $p > 0$ such that (B3) and (B4) hold. Taking $p \leq 1$ w.l.o.g. and using subadditivity of $x \mapsto x^p$ yields that $\mathbb{E}_\pi L(\Psi_1)^p < \infty$ and $\mathbb{E}_\pi d(x_0, \Psi_1(x_0))^p < \infty$ are sufficient for (B3) and (B4), respectively.

Lower bound γ_i^* for the rate of exponential convergence. In the following analysis, an object will frequently appear that was originally used in [AF01, Section 2] as a "lower bound for the rate of exponential convergence in the results" they proved in the i.i.d. setting. There, it is denoted by γ^* and defined by $\log \gamma^* := \inf_{\gamma \in (0,1)} \frac{\log \gamma}{\mu(\gamma)}$ with $\mu(\gamma) := \mathbb{E}\sigma_1(\gamma)$ and

$$\sigma_1(\gamma) := \inf\{n \geq 1 : \sum_{k=1}^{n} \log L_k \leq \log \gamma\}, \quad \gamma \in (0,1),$$

where the L_k are the i.i.d. Lipschitz constants in the i.i.d. setting, cf. [AF01] for further details. Note that this definition of $\sigma_1(\gamma)$ is only used in this paragraph to avoid confusion with $\sigma_1(i)$. Lemma B.5 in the appendix gathers useful properties and the two different forms of γ^* depending on whether $\mathbb{E}|\log L_1|$ is finite or not.

As a MIFS satisfying (MC1) and (MC2) contains an embedded i.i.d. IFS $(Y_n^x)_{n\geq 0}$ by Lemma 6.12, we denote by γ_i^* the corresponding γ^* of $(Y_n^x)_{n\geq 0}$ (recall that Y_n^x depends on i). By Lemma B.5, either $\mathbb{E}_i|\log L({}^{\#}\vec{\Psi}_1^i)| = \infty$ and then $\gamma_i^* = 0$ or $\mathbb{E}_i|\log L({}^{\#}\vec{\Psi}_1^i)| < \infty$ which implies $\gamma_i^* = e^{\mathbb{E}_i \log L({}^{\#}\vec{\Psi}_1^i)}$ and $e^{\mathbb{E}_i \log L({}^{\#}\vec{\Psi}_1^i)} \leq e^{\mathbb{E}_i S_{\sigma_1(i)}} = e^{\mu m_{ii}}$. So $e^{\mu m_{ii}} \geq \gamma_i^*$ and, in particular,

$$(e^{\mu m_{ii}}, 1) \subset (\gamma_i^*, 1) \quad \text{and} \quad (e^{\mu}, 1) \subset ((\gamma_i^*)^{\frac{1}{m_{ii}}}, 1).$$

These intervals represent validity ranges in the following theorems. Notably, the range $(e^\mu, 1)$ does *not* depend on i and e^μ constitutes the analogue to γ^* from the i.i.d. setting.

7. Convergence rates in Elton's theorem

7.1. Main results

For the formulation of our main result, we introduce the Prokhorov metric of probability measures which we denote (and already denoted) by d_{Pr}. Since $\mathbb{X}$ is separable, the Prokhorov metric is a metrization of the topology of weak convergence on the space of probability measures on $\mathbb{X}$. For a precise definition and a useful characterization of the Prokhorov metric, we refer the reader to Section B.2 in the appendix. With these preparations made, we can state our main results.

Theorem 7.1. *Given a MIFS $(\theta_n, X_n)_{n\geq 0}$ satisfying* (MC1) *and* (MC2)*, there exists a random variable ${}^{\#}\widehat{X}_\infty$ such that the following assertions hold, as $n \to \infty$.*

(a)

$$ {}^{\#}\widehat{X}_n^x \to {}^{\#}\widehat{X}_\infty \quad \mathbb{P}_\pi\text{-a.s. for every } x \in \mathbb{X}. \tag{7.1}$$

(b) For every $\gamma \in (e^\mu, 1)$

$$\gamma^{-n} d({}^{\#}\widehat{X}_n^x, {}^{\#}\widehat{X}_\infty) \to 0 \quad \mathbb{P}_\pi\text{-a.s. for every } x \in \mathbb{X}. \tag{7.2}$$

(c) For every $\gamma \in (e^\mu, 1)$

$$\mathbb{P}_\pi(d({}^{\#}\widehat{X}_n^x, {}^{\#}\widehat{X}_\infty) > \gamma^n) \to 0 \quad \text{for every } x \in \mathbb{X}. \tag{7.3}$$

Theorem 7.2. *Given a MIFS $(\theta_n, X_n)_{n\geq 0}$ satisfying the polynomial-type moment conditions of order $p > 0$ for $i \in \mathcal{S}$.*

(a) For every $\gamma \in ((\gamma_i^*)^{\frac{1}{m_{ii}}}, 1) \supseteq (e^\mu, 1)$

$$\sum_{n\geq 1} n^{p-1}\mathbb{P}_i(d({}^{\#}\widehat{X}_n^x, {}^{\#}\widehat{X}_\infty) > \gamma^n) \leq C_\gamma(1 + \log_*^p d(x, x_0)) \tag{7.4}$$

and

$$\lim_{n\to\infty} n^p \mathbb{P}_i(d({}^{\#}\widehat{X}_n^x, {}^{\#}\widehat{X}_\infty) > \gamma^n) = 0 \tag{7.5}$$

for every $x \in \mathbb{X}$ and some constant $C_\gamma \in (0, \infty)$.

(b) Assertion (a) holds with $\mathbb{P}_\pi$ instead of $\mathbb{P}_i$ (and possibly different C_γ).

(c) For all $n \geq 0$ and $x \in \mathbb{X}$

$$d_{\mathrm{Pr}}(\mathbb{P}_\pi^{X_n^x}, \mathbb{P}_\pi^{{}^{\#}\widehat{X}_\infty}) \leq A_x(n+1)^{-p}, \tag{7.6}$$

where $A_x = A_1 + A_2 \log_*^p d(x, x_0)$, and A_1 and A_2 are a positive constants not depending on n. A_1 and A_2 do moreover not depend on x.

(d) For every $x_0 \in \mathbb{X}$

$$\mathbb{E}_i \log_*^p d(x_0, {}^{\#}\widehat{X}_\infty) < \infty. \tag{7.7}$$

(e) Assertion (d) also holds with $\mathbb{E}_\pi$ instead of $\mathbb{E}_i$.

Theorem 7.3. *Given a MIFS $(\theta_n, X_n)_{n\geq 0}$ satisfying the geometric-type moment conditions for $i \in \mathcal{S}$.*

(a) For every $\gamma \in ((\gamma_i^*)^{\frac{1}{m_{ii}}}, 1) \supset (e^\mu, 1)$ there exists an $\alpha_{i,\gamma} > 1$ such that

$$\lim_{n\to\infty} \alpha_{i,\gamma}^n \mathbb{P}_i(d({}^{\#}\widehat{X}_n^x, {}^{\#}\widehat{X}_\infty) > \gamma^n) = 0 \tag{7.8}$$

for every $x \in \mathbb{X}$.

(b) Assertion (a) holds with $\mathbb{P}_\pi$ instead of $\mathbb{P}_i$ (and possibly different $\alpha_{i,\gamma}$).

(c) There exists an $r \in (0,1)$ such that

$$d_{\mathrm{Pr}}(\mathbb{P}_\pi^{X_n^x}, \mathbb{P}_\pi^{{}^{\#}\widehat{X}_\infty}) \leq A_x r^n, \tag{7.9}$$

for all $n \geq 0$, $x \in \mathbb{X}$, where $A_x = A_1 + d(x, x_0)^{A_2}$, and A_1 and A_2 are positive constants not depending on n. Moreover, r, A_1 and A_2 do not depend on x (and n).

(d) There exists an $\eta > 0$ such that for every $x_0 \in \mathbb{X}$

$$\mathbb{E}_i d(x_0, {}^{\#}\widehat{X}_\infty)^\eta < \infty. \tag{7.10}$$

(e) If additionally (B2*) *holds, then Assertion (d) also holds with $\mathbb{E}_\pi$ instead of $\mathbb{E}_i$.*

Remark 7.4. Condition (B2*) in (e) is certainly not optimal but required for technical reasons.

7.2. Proof of Theorem 7.1 by cyclic decomposition

We recall that the considered MIFS satisfies (MC1) and (MC2) throughout the rest of this work. This feature will therefore not be mentioned in the requirements of the following lemmas. Our first step towards a regenerative proof of Elton's theorem in the Markov setting is the following easy observation.

Lemma 7.5. *For every $i \in \mathcal{S}$, there exists a ${}^{\#}\widehat{X}_\infty^{(i)}$ such that for $n \to \infty$ and for each $x \in \mathbb{X}$*

$$ {}^{\#}\widehat{X}_{\sigma_n(i)}^x \to {}^{\#}\widehat{X}_\infty^{(i)} \quad \mathbb{P}_i\text{-a.s.}$$

Proof. We know from Lemma 6.12 that $(Y_n^x)_{n\geq 0}$ is an IFS of i.i.d. Lipschitz maps under $\mathbb{P}_i$ satisfying the jump-size condition and $\mathbb{E}_i \log L({}^{\#}\vec{\Psi}_1^i) \leq \mathbb{E}_i S_{\sigma_1(i)} < 0$, hence Elton's theorem for i.i.d. Lipschitz maps (e.g. [AF01, Theorem 2.1]) yields the existence of a ${}^{\#}\widehat{X}_\infty^{(i)}$ (which does not depend on x) such that

$$\widehat{Y}_n^x \to {}^{\#}\widehat{X}_\infty^{(i)} \quad \mathbb{P}_i\text{-a.s.},$$

as $n \to \infty$. Noting that $\widehat{Y}_n^x = {}^{\#}\widehat{X}_{\sigma_n(i)}^x$ completes the proof, choosing ${}^{\#}\widehat{X}_\infty^{(i)} := {}^{\#}\widehat{X}_\infty^{(i),x_0}$ as the limit of $\widehat{Y}_n^{x_0}$, where $x_0 \in \mathbb{X}$ is some reference point. The above applies to every $i \in \mathcal{S}$. □

Obviously, this ${}^{\#}\widehat{X}_{\infty}^{(i)}$ is a sensible candidate for a $\mathbb{P}_i$-a.s. limit of ${}^{\#}\widehat{X}_n^x$ and as we are primarily interested in a $\mathbb{P}_\pi$-a.s. limit, we *glue* these variables together to get the representation

$$ {}^{\#}\widehat{X}_{\infty} = \sum_{i \in \mathcal{S}} {}^{\#}\widehat{X}_{\infty}^{(i)} \mathbb{1}_{\{{}^{\#}\theta_0 = i\}} $$

of the anticipated limit. Certainly, it will turn out to be the right one.

In the following, we want to show that $d({}^{\#}\widehat{X}_n^x, {}^{\#}\widehat{X}_{\infty}) \to 0$ $\mathbb{P}_\pi$-a.s., thus we need to get a good estimate for $d({}^{\#}\widehat{X}_n^x, {}^{\#}\widehat{X}_{\infty})$. Therefore, we introduce

$$ \tau(n) := \inf\{k \geq 0 \mid \sigma_k(i) \geq n\}, $$

$n \geq 0$, the first time that $\sigma_k(i)$ is greater than or equal to n, and we keep in mind that $\tau(n)$ depends on i, but drop it here. Of course, $\tau(n)$ also makes sense for non-integer $n \geq 0$. Then the triangle inequality provides the estimate

$$ d({}^{\#}\widehat{X}_n^x, {}^{\#}\widehat{X}_{\infty}) \leq d({}^{\#}\widehat{X}_n^x, {}^{\#}\widehat{X}_n^{x_0}) + d({}^{\#}\widehat{X}_n^{x_0}, {}^{\#}\widehat{X}_{\sigma_{\tau(n)}(i)}^{x_0}) + d({}^{\#}\widehat{X}_{\sigma_{\tau(n)}(i)}^{x_0}, {}^{\#}\widehat{X}_{\infty}), $$

where the last summand clearly vanishes $\mathbb{P}_i$-a.s. due to the previous lemma, as $\tau(n) \to \infty$ $\mathbb{P}_i$-a.s. This estimate extracts the whole influence of x into the first summand, where we compare the iteration started in x to the iteration started in some fixed reference point x_0. Now, our main goal is to deal with the first and the second term on the right-hand side. To deal with the second summand, i.e. to control the error in distance when considering the subsequence $\sigma_{\tau(n)}(i)$ instead of n, we set for $n \geq 1$

$$ \begin{aligned} C_n := {} & d(x_0, {}^{\#}\Psi_{\sigma_{n-1}(i)+1} \cdots {}^{\#}\Psi_{\sigma_n(i)}(x_0)) \\ & \vee \max_{\sigma_{n-1}(i) < k < \sigma_n(i)} \{d({}^{\#}\Psi_{\sigma_{n-1}(i)+1} \cdots {}^{\#}\Psi_k(x_0), {}^{\#}\Psi_{\sigma_{n-1}(i)+1} \cdots {}^{\#}\Psi_{\sigma_n(i)}(x_0))\} \end{aligned} \tag{7.11} $$

as the maximal distance of ${}^{\#}\vec{\Psi}_n^i(x_0)$ from the set

$$ \{x_0, {}^{\#}\Psi_{\sigma_{n-1}(i)+1}(x_0), \ldots, {}^{\#}\Psi_{\sigma_{n-1}(i)+1} \cdots {}^{\#}\Psi_{\sigma_n(i)-1}(x_0)\}. $$

The C_n are i.i.d. under $\mathbb{P}_i$ and $C_{\tau(n)}$ converges in distribution (w.r.t. $\mathbb{P}_i$). For the sake of completeness, we also introduce for $n \geq 1$

$$ G_n := \max\{L({}^{\#}\Psi_{\sigma_{n-1}(i)+1}), \ldots, L({}^{\#}\Psi_{\sigma_{n-1}(i)+1} \cdots {}^{\#}\Psi_{\sigma_n(i)})\} \tag{7.12} $$

and for $n \geq 0$

$$ D_n := \sum_{j=1}^{\infty} \prod_{k=n+1}^{n+j-1} L({}^{\#}\vec{\Psi}_k^i) \; d({}^{\#}\vec{\Psi}_{n+j}^i(x_0), x_0), \tag{7.13} $$

where the empty product is set to 1. Obviously, the G_n are also i.i.d. We will need the D_n and G_n at a later point of this work. They are introduced to deal with the third and the first summand of the above inequality, respectively, and of course the names C_n and D_n are intentional. All three sequences have counterparts in the i.i.d. setting in [AF01, Section 3]. For instance, D_n is a generalization of the autoregressive

$$ D_n = \sum_{j=1}^{\infty} \gamma^{j-1} d(\Psi_{\sigma_{n+j}+1} \cdots \Psi_{\sigma_{n+j+1}}(x_0), x_0), \tag{7.14} $$

where the σ_n are ladder epochs and thus play a comparable role to the $\sigma_n(i)$ here, but entail more knowledge about the Lipschitz constants. In fact, we can bound the latter from above by γ by definition of σ_n. The loss of the deterministic part makes our object more complicated, however, we recognize it as a perpetuity corresponding to the sequence

$$\Big(L(^\#\vec{\Psi}^i_{n+k}), d(^\#\vec{\Psi}^i_{n+k}(x_0), x_0)\Big)_{k\geq 1}.$$

Keeping in mind that D_n (as well as C_n and G_n) depends on $i \in \mathcal{S}$, the following Lemma ensures that, under our general assumptions, the expression of D_n makes sense.

Lemma 7.6. *The D_n, $n \geq 0$, are $\mathbb{P}_i$-a.s. and $\mathbb{P}_\pi$-a.s. finite.*

Proof. We show that $D_0 < \infty$ $\mathbb{P}_i$-a.s. which yields $D_n < \infty$ $\mathbb{P}_i$-a.s. for all $n \geq 0$ since D_n is a stationary sequence. *Case 1*: If $\mathbb{P}_i(d(^\#\vec{\Psi}^i_1(x_0), x_0) = 0) = 1$, then $^\#\vec{\Psi}^i_k(x_0) = x_0$ for every $k \geq 1$ and thus $D_0 = 0$ $\mathbb{P}_i$-a.s. *Case 2*: If $\mathbb{P}_i(d(^\#\vec{\Psi}^i_1(x_0), x_0) = 0) < 1$ and

$$\mathbb{E}_i \log L(^\#\vec{\Psi}^i_1) \in (-\infty, 0)$$

(by Lemma 6.12 it is < 0 but we do not allow $= -\infty$ here) and thus $\mathbb{P}_i(L(^\#\vec{\Psi}^i_1) = 0) = 0$, then we are in the situation of [AB17a, Theorem 1.1] (which is a slightly stronger version of the original version [GM00, Theorem 2.1] by Goldie and Maller). Since by Lemma 6.12 $\mathbb{E}_i \log^+ d(^\#\vec{\Psi}^i_1(x_0), x_0) < \infty$ (and thus $\mathbb{E}_i J(\log^+ d(^\#\vec{\Psi}^i_1(x_0), x_0)) < \infty$ (cf. [AB17a] or [GM00] for more details on J, which is denoted differently in the latter)), the theorem yields $D_0 < \infty$ $\mathbb{P}_i$-a.s. *Case 3*: If $\mathbb{P}_i(d(^\#\vec{\Psi}^i_1(x_0), x_0) = 0) < 1$ and $\mathbb{E}_i \log L(^\#\vec{\Psi}^i_1) = -\infty$, then we can choose $c \in (0,1)$ from Lemma B.5 such that

$$\mathbb{E}_i \log(L(^\#\vec{\Psi}^i_1) \vee c) \in (-\infty, 0).$$

Now, the previous case applies and we find that $D_0 < \infty$ $\mathbb{P}_i$-a.s. by the estimate

$$D_0 := \sum_{j=1}^{\infty} \prod_{k=1}^{j-1} L(^\#\vec{\Psi}^i_k)\ d(^\#\vec{\Psi}^i_j(x_0), x_0) \leq \sum_{j=1}^{\infty} \prod_{k=1}^{j-1} (L(^\#\vec{\Psi}^i_k) \vee c)\ d(^\#\vec{\Psi}^i_j(x_0), x_0).$$

For the second part of the assertion, note that the distribution of $(^\#\vec{\Psi}^i_{n+k})_{k\geq 1}$ under $\mathbb{P}_\pi$ for every $n \geq 1$ is the same as that of $(^\#\vec{\Psi}^i_k)_{k\geq 1}$ under $\mathbb{P}_i$. Hence, D_n, $n \geq 1$, is also $\mathbb{P}_\pi$-a.s. finite and we only need to care about

$$\begin{aligned} D_0 &= d(^\#\vec{\Psi}^i_1(x_0), x_0) + \sum_{j=2}^{\infty} \prod_{k=1}^{j-1} L(^\#\vec{\Psi}^i_k) d(^\#\vec{\Psi}^i_j(x_0), x_0) \\ &= d(^\#\vec{\Psi}^i_1(x_0), x_0) + L(^\#\vec{\Psi}^i_1) \sum_{j=1}^{\infty} \prod_{k=1}^{j-1} L(^\#\vec{\Psi}^i_{k+1}) d(^\#\vec{\Psi}^i_{j+1}(x_0), x_0). \end{aligned}$$

The $\mathbb{P}_\pi$-a.s. finiteness of the latter perpetuity follows from the $\mathbb{P}_i$-a.s. finiteness of D_0, and the remaining objects $d(^\#\vec{\Psi}^i_1(x_0), x_0)$ and $L(^\#\vec{\Psi}^i_1)$ are obviously $\mathbb{P}_\pi$-a.s. finite. □

In the following Lemma we collect the crucial estimates for this analysis. It constitutes an analogue of [AF01, Lemma 3.2].

Lemma 7.7. *For every $n \geq 0$*

$$d(^\#\widehat{X}^x_n, ^\#\widehat{X}_\infty) \leq d(^\#\widehat{X}^x_n, ^\#\widehat{X}^{x_0}_n) + d(^\#\widehat{X}^{x_0}_n, ^\#\widehat{X}^{x_0}_{\sigma_{\tau(n)}(i)}) + d(^\#\widehat{X}^{x_0}_{\sigma_{\tau(n)}(i)}, ^\#\widehat{X}_\infty).$$

Furthermore, the following estimates hold for $n \geq 0$:

(a) $d({}^{\#}\widehat{X}_n^x, {}^{\#}\widehat{X}_n^{x_0}) \leq L({}^{\#}\Psi_1 \cdots {}^{\#}\Psi_n) \cdot d(x, x_0) \leq \prod_{k=1}^n L({}^{\#}\Psi_k) \cdot d(x, x_0)$.

(b) $d({}^{\#}\widehat{X}_n^x, {}^{\#}\widehat{X}_n^{x_0}) \leq \prod_{k=1}^{\tau(n)-1} L({}^{\#}\vec{\Psi}_k^i) \cdot G_{\tau(n)} \cdot d(x, x_0)$ *with* $G_0 := 1$.

(c) $d({}^{\#}\widehat{X}_n^{x_0}, {}^{\#}\widehat{X}_{\sigma_{\tau(n)}(i)}^{x_0}) \leq \prod_{k=1}^{\tau(n)-1} L({}^{\#}\vec{\Psi}_k^i) \cdot C_{\tau(n)}$ *with* $C_0 := 0$.

(d) $d({}^{\#}\widehat{X}_{\sigma_{\tau(n)}(i)}^{x_0}, {}^{\#}\widehat{X}_\infty) \leq \prod_{k=1}^{\tau(n)} L({}^{\#}\vec{\Psi}_k^i) \cdot D_{\tau(n)}$ $\mathbb{P}_i$*-a.s.*

Remark 7.8. As (7.13) generalizes (7.14) from the i.i.d. setting of [AF01] and thereby introduces additional randomness, the same happens in Parts (a)-(d) of the above Lemma 7.7. Here we obtain $L({}^{\#}\Psi_1 \cdots {}^{\#}\Psi_n)$, $\prod_{k=1}^{\tau(n)-1} L({}^{\#}\vec{\Psi}_k^i)$ and $\prod_{k=1}^{\tau(n)} L({}^{\#}\vec{\Psi}_k^i)$ instead of γ^n, $\gamma^{\tau(n)-1}$ and $\gamma^{\tau(n)}$, respectively.

Proof of Lemma 7.7. The first statement is just a reminder. We use the convention that empty products equal 1, i.e. $\prod_{k=1}^0 := 1$ and $\prod_{k=1}^{-1} := 1$. Now, (a) is immediate by definition of the Lipschitz constants. For (b), we remark that

$$L({}^{\#}\Psi_1 \cdots {}^{\#}\Psi_n) \leq \prod_{k=1}^{\tau(n)-1} L({}^{\#}\vec{\Psi}_k^i) \cdot L({}^{\#}\Psi_{\sigma_{\tau(n)-1}(i)+1} \cdots {}^{\#}\Psi_n) \leq \prod_{k=1}^{\tau(n)-1} L({}^{\#}\vec{\Psi}_k^i) \cdot G_{\tau(n)}$$

for $n \geq 1$ and, with our conventions, equality holds trivially if $n = 0$. To prove (c), we note that for $n \geq 1$

$$\begin{aligned}
d({}^{\#}\widehat{X}_n^{x_0}, {}^{\#}\widehat{X}_{\sigma_{\tau(n)}(i)}^{x_0}) &= d({}^{\#}\Psi_1 \cdots {}^{\#}\Psi_{\sigma_{\tau(n)-1}(i)} \cdots {}^{\#}\Psi_n(x_0), \\
&\qquad {}^{\#}\Psi_1 \cdots {}^{\#}\Psi_{\sigma_{\tau(n)-1}(i)} \cdots {}^{\#}\Psi_n \cdots {}^{\#}\Psi_{\sigma_{\tau(n)}(i)}(x_0)) \\
&\leq \prod_{k=1}^{\tau(n)-1} L({}^{\#}\Psi_{\sigma_{k-1}(i)+1} \cdots {}^{\#}\Psi_{\sigma_k(i)}) \\
&\quad \cdot d({}^{\#}\Psi_{\sigma_{\tau(n)-1}(i)+1} \cdots {}^{\#}\Psi_n(x_0), {}^{\#}\Psi_{\sigma_{\tau(n)-1}(i)+1} \cdots {}^{\#}\Psi_{\sigma_{\tau(n)}(i)}(x_0)) \\
&\leq \prod_{k=1}^{\tau(n)-1} L({}^{\#}\vec{\Psi}_k^i) \cdot C_{\tau(n)}
\end{aligned}$$

on $\{\tau(n) > 1\}$ and

$$d({}^{\#}\widehat{X}_n^{x_0}, {}^{\#}\widehat{X}_{\sigma_{\tau(n)}(i)}^{x_0}) = d({}^{\#}\widehat{X}_n^{x_0}, {}^{\#}\widehat{X}_{\sigma_1(i)}^{x_0}) \leq C_1$$

on $\{\tau(n) = 1\}$, hence

$$d({}^{\#}\widehat{X}_n^{x_0}, {}^{\#}\widehat{X}_{\sigma_{\tau(n)}(i)}^{x_0}) \leq \prod_{k=1}^{\tau(n)-1} L({}^{\#}\vec{\Psi}_k^i) \cdot C_{\tau(n)}.$$

Equality holds for $n = 0$ if we set $C_0 := 0$. For (d) let $n \geq 0$ and $m \geq 1$. Then

$$\begin{aligned}
d({}^{\#}\widehat{X}_{\sigma_{n+m}(i)}^{x_0}, {}^{\#}\widehat{X}_{\sigma_n(i)}^{x_0}) &\leq L({}^{\#}\vec{\Psi}_1^i \cdots {}^{\#}\vec{\Psi}_n^i) \cdot d({}^{\#}\vec{\Psi}_{n+1}^i \cdots {}^{\#}\vec{\Psi}_{n+m}^i(x_0), x_0) \\
&\leq \prod_{k=1}^n L({}^{\#}\vec{\Psi}_k^i) \cdot d({}^{\#}\vec{\Psi}_{n+1}^i \cdots {}^{\#}\vec{\Psi}_{n+m}^i(x_0), x_0) \\
&\leq \prod_{k=1}^n L({}^{\#}\vec{\Psi}_k^i) \cdot \Big(d({}^{\#}\vec{\Psi}_{n+1}^i(x_0), x_0)
\end{aligned}$$

$$+\sum_{j=2}^{m} d({}^{\#}\vec{\Psi}^{i}_{n+1}\cdots{}^{\#}\vec{\Psi}^{i}_{n+j}(x_0),{}^{\#}\vec{\Psi}^{i}_{n+1}\cdots{}^{\#}\vec{\Psi}^{i}_{n+j-1}(x_0))\Big)$$

$$\le \prod_{k=1}^{n} L({}^{\#}\vec{\Psi}^{i}_{k})\cdot\sum_{j=1}^{m}\prod_{k=n+1}^{n+j-1} L({}^{\#}\vec{\Psi}^{i}_{k})\; d({}^{\#}\vec{\Psi}^{i}_{n+j}(x_0),x_0).$$

The occurring sum increases to D_n, as $m\to\infty$, so

$$d({}^{\#}\widehat{X}^{x_0}_{\sigma_{n+m}(i)},{}^{\#}\widehat{X}^{x_0}_{\sigma_n(i)})\le\prod_{k=1}^{n} L({}^{\#}\vec{\Psi}^{i}_{k})\cdot D_n$$

for every $m\ge 1$, and since ${}^{\#}\widehat{X}^{x_0}_{\sigma_{n+m}(i)}$ converges to ${}^{\#}\widehat{X}_{\infty}$ $\mathbb{P}_i$-a.s. by Lemma 7.5, as $m\to\infty$, the assertion follows. □

We now collect some more properties of C_n which correspond to [Als15a, Lemma 3.27] in the i.i.d. case. Recall that C_n depends on $i\in\mathcal{S}$.

Lemma 7.9. *For all $i\in\mathcal{S}$, $\mathbb{E}_i\log^+ C_1<\infty$, hence $\frac{1}{n}\log^+ C_n\to 0$ and $e^{-\varepsilon n}C_n\to 0$ $\mathbb{P}_i$-a.s. for every $\varepsilon>0$.*

Proof. The proof uses the estimate

$$C_1\le d(x_0,{}^{\#}\Psi_1(x_0))+\sum_{n=2}^{\sigma_1(i)} d({}^{\#}\Psi_1\cdots{}^{\#}\Psi_{n-1}(x_0),{}^{\#}\Psi_1\cdots{}^{\#}\Psi_n(x_0)) \tag{7.15}$$

which holds since every argument in C_1 is smaller than or equal to this upper bound. Thus,

$$\log^+ C_1\le\log_*\left(e^{S^*_{\sigma_1(i)}}\sum_{n=1}^{\sigma_1(i)} d(x_0,{}^{\#}\Psi_n(x_0))\right)\le 1+S^*_{\sigma_1(i)}+\sum_{n=1}^{\sigma_1(i)}\log_* d(x_0,{}^{\#}\Psi_n(x_0))$$

as in (6.9). The rest follows using the SLLN. It is the same as the proof of [Als15a, Lemma 3.27] with recurrence times instead of ladder epochs, so we omit the details. □

Lemma 7.10. *The following assertions hold, as $n\to\infty$:*

(a)

$$\prod_{k=1}^{\tau(n)-1} L({}^{\#}\vec{\Psi}^{i}_{k})\cdot C_{\tau(n)}\to 0\quad\mathbb{P}_i\text{-a.s.}$$

(b) If $\gamma\in(\gamma_i^*,1)$, then

$$\gamma^{-(\tau(n)-1)}\prod_{k=1}^{\tau(n)-1} L({}^{\#}\vec{\Psi}^{i}_{k})\cdot C_{\tau(n)}\to 0\quad\mathbb{P}_i\text{-a.s.}$$

(c) If $\gamma\in((\gamma_i^*)^{\frac{1}{m_{ii}}},1)$, then

$$\gamma^{-(n-1)}\prod_{k=1}^{\tau(n)-1} L({}^{\#}\vec{\Psi}^{i}_{k})\cdot C_{\tau(n)}\to 0\quad\mathbb{P}_i\text{-a.s.}$$

Proof. The first assertion is a direct consequence of (b). To prove (b), we choose $n \geq 1$ and intend to show

$$\gamma^{-(n-1)} \prod_{k=1}^{n-1} L({}^{\#}\vec{\Psi}_k^i) \cdot C_n \to 0 \quad \mathbb{P}_i\text{-a.s.}$$

because the assertion then follows from $\tau(n) \to \infty$ $\mathbb{P}_i$-a.s. We write

$$\gamma^{-(n-1)} \prod_{k=1}^{n-1} L({}^{\#}\vec{\Psi}_k^i) = \exp\left(\sum_{k=1}^{n-1} \log L({}^{\#}\vec{\Psi}_k^i) - (n-1)\log\gamma\right) =: \exp\left(-\frac{Z'_{n-1}}{n} n\right),$$

where $Z'_n := \sum_{k=1}^{n}[-\log L({}^{\#}\vec{\Psi}_k^i) + \log\gamma]$ has i.i.d. increments and

$$\mathbb{E}_i Z'_1 = -\mathbb{E}_i \log L({}^{\#}\vec{\Psi}_1^i) + \log\gamma = -\log\gamma_i^* + \log\gamma = \log\left(\frac{\gamma}{\gamma_i^*}\right) \in (0,\infty]$$

by assumption.

If $\mathbb{E}_i \log L({}^{\#}\vec{\Psi}_1^i) > -\infty$ and thus $\mathbb{E}_i Z'_1 \in (0,\infty)$, then $\frac{Z'_{n-1}}{n} \to \log(\gamma/\gamma_i^*)$ $\mathbb{P}_i$-a.s. Let A be the intersection of the event that the latter convergence holds and the event that the last convergence in Lemma 7.9 holds. For $\omega \in A$, there exists an $n(\omega)$ such that $\frac{Z'_{n-1}(\omega)}{n} > \frac{\log(\gamma/\gamma_i^*)}{2}$ for all $n \geq n(\omega)$. Hence, we see

$$\gamma^{-(n-1)} \prod_{k=1}^{n-1} L({}^{\#}\vec{\Psi}_k^i)(\omega) \cdot C_n(\omega) \leq e^{-\frac{\log(\gamma/\gamma_i^*)}{2} n} C_n(\omega)$$

for $n \geq n(\omega)$, and the right-hand side converges to 0 by the choice of ω. Since $\mathbb{P}_i(A) = 1$, this proves (b) in this case. If $\mathbb{E}_i \log L({}^{\#}\vec{\Psi}_1^i) = -\infty$ and thus $\mathbb{E}_i Z'_1 = \infty$, pick $\varepsilon > 0$. Then $\frac{Z'_{n-1}(\omega)}{n} > \varepsilon$ holds for almost all ω and all $n \geq n(\omega)$ big enough. The rest of (b) follows as above.

To prove (c) in the case $\mathbb{E}_i \log L({}^{\#}\vec{\Psi}_1^i) > -\infty$, we use the fact that

$$\beta^{-(\tau(n)-1)} \prod_{k=1}^{\tau(n)-1} L({}^{\#}\vec{\Psi}_k^i) \cdot C_{\tau(n)} \to 0 \quad \mathbb{P}_i\text{-a.s.}, \tag{7.16}$$

as $n \to \infty$, for $\beta \in (\gamma_i^*, 1)$ by (b), and that

$$\frac{\tau(n)}{n} \to \frac{1}{\mathbb{E}_i \sigma_1(i)} = \frac{1}{m_{ii}} > 0 \quad \mathbb{P}_i\text{-a.s.},$$

as $n \to \infty$. Let $\gamma \in ((\gamma_i^*)^{\frac{1}{m_{ii}}}, 1)$. Then there is a $0 < b < \frac{1}{m_{ii}}$ such that $\gamma = (\gamma_i^*)^b$. Pick $\varepsilon' > 0$ so small that $b + \varepsilon' < \frac{1}{m_{ii}}$ which means $0 < b < \frac{1}{m_{ii}} - \varepsilon'$ and thus particularly $\frac{1}{m_{ii}} - \varepsilon' > 0$. Now, choose $\varepsilon \in (0, m_{ii}\varepsilon'(\frac{1}{m_{ii}} - \varepsilon')^{-1})$ and note that the right boundary is indeed positive by the afore-mentioned. Then there is an n_0 (possibly depending on the realization) such that for $n \geq n_0$ we have $\frac{n}{\tau(n)} \leq m_{ii} + \varepsilon$ and hence

$$\begin{aligned}
\gamma^{\frac{n}{\tau(n)}} &= (\gamma_i^*)^{b \cdot \frac{n}{\tau(n)}} \geq (\gamma_i^*)^{b \cdot (m_{ii}+\varepsilon)} = (\gamma_i^*)^{b \cdot m_{ii} + b\varepsilon} \\
&> (\gamma_i^*)^{1 - m_{ii}\varepsilon' + \frac{1}{m_{ii}}\varepsilon - \varepsilon'\varepsilon} \\
&= (\gamma_i^*)^{1 - m_{ii}\varepsilon' + \varepsilon(\frac{1}{m_{ii}} - \varepsilon')} > (\gamma_i^*)^{1 - m_{ii}\varepsilon' + m_{ii}\varepsilon'} = \gamma_i^*.
\end{aligned}$$

The assertion follows from (7.16) with $\beta := (\gamma_i^*)^{1-m_{ii}\varepsilon'+\varepsilon(\frac{1}{m_{ii}}-\varepsilon')}$, since

$$\gamma^{-(n-1)} = \gamma^{-\frac{n}{\tau(n)}(\tau(n)-1)}\gamma^{1-\frac{n}{\tau(n)}} < \beta^{-(\tau(n)-1)}\gamma^{1-\frac{n}{\tau(n)}}$$

and $n/\tau(n)$ converges to a constant. If $\mathbb{E}_i \log L(^{\#}\vec{\Psi}_1^i) = -\infty$, then the proof is again easier because we only need to show that, for $\gamma \in (0,1)$ and large n, $\gamma^{\frac{n}{\tau(n)}}$ is bounded from below by some $\beta \in (0,1)$ in the same manner as before. This is obvious, since $n/\tau(n)$ converges to a positive constant. □

Lemma 7.11. *If $\gamma \in ((\gamma_i^*)^{\frac{1}{m_{ii}}}, 1)$, then, as $n \to \infty$,*

$$\gamma^{-n} d(^{\#}\widehat{X}^x_{\sigma_{\tau(n)}(i)}, {}^{\#}\widehat{X}_\infty) \to 0 \quad \mathbb{P}_i\text{-a.s.}$$

Proof. We know from Lemma B.6 that

$$\beta^{-n} d(^{\#}\widehat{X}^x_{\sigma_n(i)}, {}^{\#}\widehat{X}_\infty) \to 0 \quad \mathbb{P}_i\text{-a.s.}$$

for $\beta \in (\gamma_i^*, 1)$. Proceeding as in the proof of Lemma 7.10(c) yields the assertion. We omit the details. □

Lemma 7.12. *If $\gamma \in (e^\mu, 1)$, then, as $n \to \infty$,*

$$\gamma^{-n} \prod_{k=1}^n L(^{\#}\Psi_k) \to 0 \quad \mathbb{P}_\lambda\text{-a.s.}$$

for every probability measure λ on $\mathcal{S}$.

Proof. We use the fact that $\log \prod_{k=1}^n L(^{\#}\Psi_k) = \sum_{k=1}^n \log L(^{\#}\Psi_k)$ is a MRW with negative stationary drift μ. So in particular, as $n \to \infty$,

$$\frac{\log \prod_{k=1}^n L(^{\#}\Psi_k)}{n} \to \mu \quad \mathbb{P}_\lambda\text{-a.s.}$$

and

$$\frac{\log \prod_{k=1}^n L(^{\#}\Psi_k)}{n} - \log\gamma \to \mu - \log\gamma =: a < 0 \quad \mathbb{P}_\lambda\text{-a.s.}$$

by assumption. So with

$$\gamma^{-n} \prod_{k=1}^n L(^{\#}\Psi_k) = \exp\left(n\left(\frac{\log \prod_{k=1}^n L(^{\#}\Psi_k)}{n} - \log\gamma\right)\right),$$

we find an $n(\omega)$ for every ω from a $\mathbb{P}_\lambda$-a.s. set such that $\frac{\log \prod_{k=1}^n L(^{\#}\Psi_k)(\omega)}{n} - \log\gamma < \frac{a}{2}$ for all $n \geq n(\omega)$. This means that for these n

$$\gamma^{-n} \prod_{k=1}^n L(^{\#}\Psi_k)(\omega) \leq e^{n\frac{a}{2}} \to 0.$$

□

Proof of Theorem 7.1(a). As remarked earlier, the third summand in Lemma 7.7 vanishes $\mathbb{P}_i$-a.s. as does the second summand by Lemma 7.10 and the estimate Lemma 7.7(c). By Lemma 7.12 the first summand tends to 0 as well. Hence, ${}^{\#}\widehat{X}_n^x \to {}^{\#}\widehat{X}_\infty$ $\mathbb{P}_i$-a.s. Since in our whole construction $i \in \mathcal{S}$ was arbitrary, we obtain the latter assertion for all $i \in \mathcal{S}$, which is equivalent to the convergence $\mathbb{P}_\pi$-a.s. since $\pi_i > 0$ for all $i \in \mathcal{S}$. □

Proof of Theorem 7.1(b). Let i be an element of $\mathcal{S}$. We combine Lemma 7.7 with Lemmas 7.10, 7.11 and 7.12: The first provides

$$\gamma^{-n} d({}^{\#}\widehat{X}_n^x, {}^{\#}\widehat{X}_\infty) \leq \gamma^{-n} \prod_{k=1}^{n} L({}^{\#}\Psi_k) \cdot d(x, x_0) + \gamma^{-n} \prod_{k=1}^{\tau(n)-1} L({}^{\#}\vec{\Psi}_k^i) \cdot C_{\tau(n)} + \gamma^{-n} d({}^{\#}\widehat{X}_{\sigma_{\tau(n)}(i)}^{x_0}, {}^{\#}\widehat{X}_\infty).$$

Then Lemma 7.10 gives

$$\gamma^{-n} \prod_{k=1}^{\tau(n)-1} L({}^{\#}\vec{\Psi}_k^i) \cdot C_{\tau(n)} \to 0 \quad \mathbb{P}_i\text{-a.s.}$$

for all $\gamma \in (e^\mu, 1)$, as $n \to \infty$, while Lemma 7.11 gives

$$\gamma^{-n} d({}^{\#}\widehat{X}_{\sigma_{\tau(n)}(i)}^{x_0}, {}^{\#}\widehat{X}_\infty) \to 0 \quad \mathbb{P}_i\text{-a.s.}$$

for the same γ's, as $n \to \infty$. The last ingredient is provided by Lemma 7.12 which gives

$$\gamma^{-n} \prod_{k=1}^{n} L({}^{\#}\Psi_k) \to 0 \quad \mathbb{P}_i\text{-a.s.}$$

We recall that, as mentioned at the end of Section 6.4, $(e^\mu, 1)$ is always contained in the i-dependent ranges for γ. We conclude that

$$\lim_{n\to\infty} \gamma^{-n} d({}^{\#}\widehat{X}_n^x, {}^{\#}\widehat{X}_\infty) = 0 \quad \mathbb{P}_i\text{-a.s.}$$

for $\gamma \in (e^\mu, 1)$ and all $i \in \mathcal{S}$. This is equivalent to our assertion, since x was arbitrary. □

Proof of Theorem 7.1(c). This follows immediately from (b): For every $\gamma \in (e^\mu, 1)$ and as $n \to \infty$,

$$\gamma^{-n} d({}^{\#}\widehat{X}_n^x, {}^{\#}\widehat{X}_\infty) \to 0 \quad \mathbb{P}_\pi\text{-a.s.}$$

In particular, this convergence holds in probability and

$$\mathbb{P}_\pi(d({}^{\#}\widehat{X}_n^x, {}^{\#}\widehat{X}_\infty) > \gamma^n) = \mathbb{P}_\pi(\gamma^{-n} d({}^{\#}\widehat{X}_n^x, {}^{\#}\widehat{X}_\infty) > 1) \to 0.$$

□

Theorem 7.1 and the previous lemmas provide us with refined versions of Lemma 7.5 and Lemma 7.7(d).

Lemma 7.13. *The following assertions hold:*

(a) ${}^{\#}\widehat{X}_{\sigma_n(i)}^x \to {}^{\#}\widehat{X}_\infty^{(j)} = {}^{\#}\widehat{X}_\infty$ $\mathbb{P}_j$*-a.s., as* $n \to \infty$*, for all* $j \in \mathcal{S}$*.*

(b) $d({}^{\#}\widehat{X}_{\sigma_{\tau(n)}(i)}^{x_0}, {}^{\#}\widehat{X}_\infty) \leq \prod_{k=1}^{\tau(n)} L({}^{\#}\vec{\Psi}_k^i) \cdot D_{\tau(n)}$ $\mathbb{P}_j$*-a.s. for all* $j \in \mathcal{S}$*.*

Proof. For the first part, we recall that ${}^{\#}\widehat{X}_n^x \to {}^{\#}\widehat{X}_\infty$ $\mathbb{P}_j$-a.s. and $\sigma_n(i) \to \infty$ $\mathbb{P}_j$-a.s., as $n \to \infty$. For the second part, we look into the proof of Lemma 7.7(d), where we see

$$d({}^{\#}\widehat{X}^{x_0}_{\sigma_{n+m}(i)}, {}^{\#}\widehat{X}^{x_0}_{\sigma_n(i)}) \leq \prod_{k=1}^{n} L({}^{\#}\vec{\Psi}_k^i) \cdot D_n$$

for every $m \geq 1$. By (a), ${}^{\#}\widehat{X}^{x_0}_{\sigma_{n+m}(i)}$ converges to ${}^{\#}\widehat{X}_\infty$ $\mathbb{P}_j$-a.s., as $m \to \infty$. □

7.3. Convergence rate under polynomial-type moment conditions

Motivated by Theorem 7.1(c), we are interested in convergence rates. We need a number of lemmas as preparation and we are guided by the approach of [AF01]. The next result is an analogue of [AF01, Lemma 3.5].

Lemma 7.14. *If* (A1) *holds, then for all* $0 < a < \frac{1}{m_{ii}}$

$$\sum_{n\geq 1} n^{p-1}\mathbb{P}_\pi(\tau(n) \leq an) < \infty \quad \textit{and} \quad \lim_{n\to\infty} n^p\mathbb{P}_\pi(\tau(n) \leq an) = 0$$

and the same holds with $\mathbb{P}_i$ *instead of* $\mathbb{P}_\pi$*. If* (B1) *holds, then for all* $0 < a < \frac{1}{m_{ii}}$ *there exists an* $\alpha > 1$*, depending on* a *and* i*, such that*

$$\lim_{n\to\infty} \alpha^n \mathbb{P}_\pi(\tau(n) \leq an) = 0.$$

The same holds with $\mathbb{P}_i$ *instead of* $\mathbb{P}_\pi$ *with possibly different* α*.*

Proof. We remark that the $\mathbb{P}_i$ case for both parts of the lemma is [AF01, Lemma 3.5] with recurrence times instead of ladder epochs. Nevertheless, we need to adjust their result for the first part a little bit to prove the assertions under $\mathbb{P}_\pi$. By [CL75, Theorem 5] with $\alpha = 1$, we have

$$\mathbb{E}_i\Big(\sup_{k\geq 0}(\sigma_k(i) - k(m_{ii} + \varepsilon))\Big)^p < \infty \tag{7.17}$$

for all $\varepsilon > 0$ (the latter expression equals $m(\varepsilon, p+1, 1)$ as defined in [CL75]). As $0 < a < \frac{1}{m_{ii}}$, it is possible to choose $\varepsilon > 0$ such that $a(m_{ii} + \varepsilon) = 1 - \varepsilon$ and thus

$$\begin{aligned}
\mathbb{P}_i(\tau(n(1-\delta)) \leq an) &= \mathbb{P}_i(\sigma_{\lfloor an \rfloor}(i) - an(m_{ii} + \varepsilon) \geq n(1 - \delta - a(m_{ii} + \varepsilon))) \\
&\leq \mathbb{P}_i(\sigma_{\lfloor an \rfloor}(i) - \lfloor an \rfloor(m_{ii} + \varepsilon) \geq (\varepsilon - \delta)n) \\
&\leq \mathbb{P}_i\Big(\sup_{k\geq 0}(\sigma_k(i) - k(m_{ii} + \varepsilon)) \geq (\varepsilon - \delta)n\Big) =: c_{i,\varepsilon}((\varepsilon - \delta)n)
\end{aligned}$$

for $\delta \in [0, 1)$. We know from (7.17) that for every $\eta > 0$, as $n \to \infty$,

$$n^p c_{i,\varepsilon}(\eta n) \to 0,$$

and with the same means

$$\sum_{n\geq 1} n^{p-1} c_{i,\varepsilon}(\eta n) \leq C \cdot \mathbb{E}_i(\sup_{k\geq 0}(\sigma_k(i) - k(m_{ii} + \varepsilon)))^p < \infty$$

for some constant $C > 0$. Let $n \geq 1$. Starting according to the distribution π, we still have

$$\mathbb{P}_\pi(\tau(n) \leq an) \leq \mathbb{P}_\pi\Big(\sup_{k\geq 0}(\sigma_k(i) - k(m_{ii}+\varepsilon)) \geq \varepsilon n\Big).$$

The $\sigma_1(i), \sigma_2(i) - \sigma_1(i), \ldots$ are still independent w.r.t. $\mathbb{P}_\pi$ and furthermore, $\sigma_2(i) - \sigma_1(i), \sigma_3(i) - \sigma_2(i), \ldots$ are identically distributed under $\mathbb{P}_\pi$ with common distribution $\mathbb{P}_i(\sigma_1(i) \in \cdot)$. With $\delta \in (0,1)$, we decompose

$$\mathbb{P}_\pi(\tau(n) \leq an) \leq \mathbb{P}_\pi(\sigma_1(i) \geq n\delta) + \mathbb{P}_\pi(\tau(n) \leq an, \sigma_1(i) < n\delta)$$

and remark that

$$\begin{aligned}\tau(n) &= \inf\{k \geq 2 : \sigma_k(i) \geq n\} = 1 + \inf\{k \geq 1 : \sigma_{k+1}(i) - \sigma_1(i) \geq n - \sigma_1(i)\} \\ &\geq \inf\{k \geq 1 : \sigma_{k+1}(i) - \sigma_1(i) \geq n(1-\delta)\}\end{aligned}$$

on $\{\sigma_1(i) < n\delta\}$ to infer

$$\mathbb{P}_\pi(\tau(n) \leq an) \leq \mathbb{P}_\pi(\sigma_1(i) \geq n\delta) + \mathbb{P}_i(\tau(n(1-\delta)) \leq an) = \mathbb{P}_\pi(\sigma_1(i) \geq n\delta) + c_{i,\varepsilon}((\varepsilon-\delta)n).$$

Choosing $\delta \in (0, \varepsilon)$, the right-hand side vanishes, as $n \to \infty$, when multiplied by n^p, and is summable weighted with n^{p-1}, both if $\mathbb{E}_\pi \sigma_1(i)^p < \infty$. Lemma B.8 and Remark 6.15 show that this is the case under (A1)-(A3), in fact only $\mathbb{E}_i \sigma_1(i)^{p+1} < \infty$ is needed (cf. the proof).

For the second part, set $\phi^{(i)}(\beta) := \mathbb{E}_i e^{\beta\sigma_1(i)}$ and $\phi_a^{(i)}(\beta) := \mathbb{E}_i e^{\beta(\sigma_1(i) - \frac{1}{a})} = e^{-\frac{\beta}{a}}\phi^{(i)}(\beta)$. Then

$$(\phi_a^{(i)})'(0) = \mathbb{E}_i\left(\sigma_1(i) - \frac{1}{a}\right) = m_{ii} - \frac{1}{a} < 0$$

and $\phi_a^{(i)}(0) = 1$, thus there exists a $\beta_a \in (0, \beta_i^*]$ with $\phi_a^{(i)}(t) < 1$ for all $t \in (0, \beta_a]$. As in [AF01, Lemma 3.5], Markov's inequality yields

$$\mathbb{P}_i(\tau(n) \leq an) \leq e^{-\beta_a n}\phi^{(i)}(\beta_a)^{an} = \left(\phi_a^{(i)}(\beta_a)^a\right)^n =: \phi_{i,a}^n,$$

and we can pick every $\alpha \in (1, \phi_{i,a}^{-1})$. If we start according to π, then, with an adequate condition in mind, we refine the above procedure to

$$\mathbb{P}_\pi(\tau(n) \leq an) \leq e^{-tn}\mathbb{E}_\pi e^{t\sigma_{\lfloor an\rfloor}(i)}$$

for every $t > 0$. With the same properties of $\sigma_n(i)$ that we stated out in the first part of the proof, we find

$$\mathbb{E}_\pi e^{t\sigma_{\lfloor an\rfloor}(i)} = \mathbb{E}_\pi e^{t\sum_{k=2}^{\lfloor an\rfloor}\sigma_k(i) - \sigma_{k-1}(i)} e^{t\sigma_1(i)} = \phi^{(i)}(t)^{\lfloor an\rfloor - 1}\mathbb{E}_\pi e^{t\sigma_1(i)} \leq \phi^{(i)}(t)^{an}\mathbb{E}_\pi e^{t\sigma_1(i)},$$

so

$$\mathbb{P}_\pi(\tau(n) \leq an) \leq \phi_a^{(i)}(t)^{an}\mathbb{E}_\pi e^{t\sigma_1(i)} =: (\phi_{i,a,t})^n\,\mathbb{E}_\pi e^{t\sigma_1(i)}.$$

Then $\alpha \in (1, \phi_{i,a,t}^{-1})$ works here if $\mathbb{E}_\pi e^{t\sigma_1(i)} < \infty$ and $t < \beta_a$. Concerning this condition, we refer to Lemma B.8 again, which tells us that it is possible to find a $t' > 0$ with $\mathbb{E}_\pi e^{t'\sigma_1(i)} < \infty$, whenever $\mathbb{E}_i e^{\beta_i^*\sigma_1(i)} < \infty$. Hence, every t smaller or equal to $t' \wedge \beta_a$ suffices. □

The following Lemma is extracted from the proof of [AF01, Lemma 3.3]. We will also develop an analogue of [AF01, Lemma 3.3] which is split into Lemmas 7.19, 7.22 and 7.17.

Lemma 7.15. *Let* $F_n := f({}^{\#}\Psi_{\sigma_{n-1}(i)+1}, \dots, {}^{\#}\Psi_{\sigma_n(i)})$ *for some measurable function* f*. Then for every measurable function* $H : [0,\infty) \to [0,\infty)$*,* $t \geq 0$ *and* $n \geq 1$

$$\mathbb{P}_i(H(F_{\tau(n)}) > t) \leq \mathbb{E}_i \sigma_1(i) \mathbb{1}_{\{H(F_1)>t\}}$$

and

$$\mathbb{P}_\pi(H(F_{\tau(n)}) > t) \leq \mathbb{E}_i \sigma_1(i) \mathbb{1}_{\{H(F_1)>t\}} + \mathbb{P}_\pi(\sigma_1(i) > n-1, H(F_1) > t).$$

Remark 7.16. This situation applies to C_n and G_n, $n \geq 1$.

Proof. Using

$$\sum_{k=1}^{\infty} \mathbb{P}_\pi(\sigma_k(i) = l) \leq 1$$

for $l \geq 1$, a reasoning very similar to [AF01, (3.18)] (with recurrence times instead of ladder epochs, and under $\mathbb{P}_\pi$) yields

$$\begin{aligned}
\mathbb{P}_\pi(H(F_{\tau(n)}) > t) &= \sum_{l=1}^{n-1} \left(\sum_{k=2}^{l+1} \mathbb{P}_\pi(\sigma_{k-1}(i) = l) \right) \mathbb{P}_i(\sigma_1(i) \geq n-l, H(F_1) > t) \\
&\quad + \mathbb{P}_\pi(\sigma_1(i) \geq n, H(F_1) > t) \\
&\leq \sum_{l=0}^{n-2} \mathbb{P}_i(\sigma_1(i) > l, H(F_1) > t) + \mathbb{P}_\pi(\sigma_1(i) > n-1, H(F_1) > t)
\end{aligned}$$

for every $n \geq 1$. Hence, we conclude

$$\begin{aligned}
\mathbb{P}_\pi(H(F_{\tau(n)}) > t) &\leq \sum_{l \geq 0} \mathbb{P}_i(\sigma_1(i) > l, H(F_1) > t) + \mathbb{P}_\pi(\sigma_1(i) > n-1, H(F_1) > t) \\
&= \mathbb{E}_i \sigma_1(i) \mathbb{1}_{\{H(F_1)>t\}} + \mathbb{P}_\pi(\sigma_1(i) > n-1, H(F_1) > t).
\end{aligned}$$

When replacing π by i, the estimation simplifies to the i.i.d. situation in [AF01, Lemma 3.3] (with recurrence times instead of ladder epochs) and we obtain

$$\mathbb{P}_i(H(F_{\tau(n)}) > t) \leq \sum_{l \geq 0} \mathbb{P}_i(\sigma_1(i) > l, H(F_1) > t) = \mathbb{E}_i \sigma_1(i) \mathbb{1}_{\{H(F_1)>t\}}.$$

□

The following lemma is an analogue of the second part of [AF01, Lemma 3.3], and it also serves as the key tool for dealing with the additional randomness that originates from considering recurrence times instead of ladder epochs.

Lemma 7.17. *If* (A1)-(A3) *hold, then for all* $\gamma \in ((\gamma_i^*)^{\frac{1}{m_{ii}}}, 1)$

$$\sum_{n \geq 1} n^{p-1} \mathbb{P}_\pi \left(\prod_{k=1}^{\tau(n)-1} L({}^{\#}\vec{\Psi}_k^i) > \gamma^n \right) < \infty \quad \text{and} \quad \lim_{n\to\infty} n^p \mathbb{P}_\pi \left(\prod_{k=1}^{\tau(n)-1} L({}^{\#}\vec{\Psi}_k^i) > \gamma^n \right) = 0$$

and the same holds with $\mathbb{P}_i$ *instead of* $\mathbb{P}_\pi$*.*

Remark 7.18. Obviously, one can replace $\tau(n) - 1$ by $\tau(n)$ in the previous statement, the proof even becomes easier. The same holds for the corresponding Lemma 7.25 under (B1)-(B4).

Proof. Suppose first that $\mathbb{E}_i|\log L(^{\#}\vec{\Psi}_1^i)| < \infty$. Then $\gamma_i^* \in (0,1)$. Define

$$L_k^{*,i} := L(^{\#}\vec{\Psi}_k^i)/\gamma_i^*, \quad k \geq 1, \qquad S_n^{*,i} := \sum_{k=1}^n \log L_k^{*,i}, \quad n \geq 1, \quad S_0^{*,i} = 0.$$

The $L_k^{*,i}$ are i.i.d. under $\mathbb{P}_i$ with $\mathbb{E}_i \log L_1^{*,i} = 0$ and independent under $\mathbb{P}_\pi$ for $k \geq 1$, and they are identically distributed under $\mathbb{P}_\pi$ with common distribution $\mathbb{P}_i(L_1^{*,i} \in \cdot)$ for $k \geq 2$. Fix $n \geq 1$. We start with a $\gamma \in ((\gamma_i^*)^{\frac{1}{m_{ii}}}, 1)$. Thus, there is a $b \in (0, \frac{1}{m_{ii}})$ with $\gamma = (\gamma_i^*)^b$. Pick $a \in (b, \frac{1}{m_{ii}})$ and decompose

$$\mathbb{P}_\pi\left(\prod_{k=1}^{\tau(n)-1} L(^{\#}\vec{\Psi}_k^i) > \gamma^n\right) = \mathbb{P}_\pi\left(\prod_{k=1}^{\tau(n)-1} L_k^{*,i} > (\gamma_i^*)^{bn}\frac{1}{(\gamma_i^*)^{\tau(n)-1}}\right)$$
$$= \mathbb{P}_\pi\left(\prod_{k=1}^{\tau(n)-1} L_k^{*,i} > (\gamma_i^*)^{n\left(b-\frac{\tau(n)-1}{n}\right)}\right) = \mathbb{P}_\pi\left(\sum_{k=1}^{\tau(n)-1} \log L_k^{*,i} > n\log\left((\gamma_i^*)^{b-\frac{\tau(n)-1}{n}}\right)\right)$$

in the following: On the one hand, we consider

$$\mathbb{P}_\pi\left(\sum_{k=1}^{\tau(n)-1} \log L_k^{*,i} > n\log\left((\gamma_i^*)^{b-\frac{\tau(n)-1}{n}}\right), \tau(n) \leq an+1\right) \leq \mathbb{P}_\pi(\tau(n) \leq an+1)$$

which by Lemma 7.14 is summable weighted with n^{p-1}, and converges to 0, as $n \to \infty$, when multiplied by n^p. On the other hand, with $\varepsilon := \log((\gamma_i^*)^{b-a}) > 0$,

$$\begin{aligned}&\mathbb{P}_\pi\left(\sum_{k=1}^{\tau(n)-1} \log L_k^{*,i} > n\log\left((\gamma_i^*)^{b-\frac{\tau(n)-1}{n}}\right), \tau(n) > an+1\right)\\ &\leq \mathbb{P}_\pi\left(\sum_{k=1}^{\tau(n)-1} \log L_k^{*,i} > n\log\left((\gamma_i^*)^{b-a}\right)\right) \leq \mathbb{P}_\pi\left(\max_{0\leq k\leq n} S_k^{*,i} > n\varepsilon\right)\end{aligned} \tag{7.18}$$

since $\tau(n) \in \{1,\ldots,n\}$ for $n \geq 1$ and $S_0^{*,i} = 0$. Again, we try to relate this expression to $\mathbb{P}_i\left(\max_{0\leq k\leq n} S_k^{*,i} > n\varepsilon'\right)$ (and some arising toll term) for some $\varepsilon' > 0$. Taking such a relation for granted, [CL75, Remark on page 57], for $p' := p+1$ and $\alpha = r = 1$, gives us the desired one-sided tail estimate for the latter expression, i.e. the latter expression is summable weighted with $n^{p'\alpha-2} = n^{p-1}$ (recall that $\mathbb{E}_i|\log L_1^{*,i}| < \infty$ and $\mathbb{E}_i(\log^+ L_1^{*,i})^{p+1} \leq \mathbb{E}_i(\log_* L_1^{*,i})^{p+1} < \infty$ by assumption). We will show now, that the right-hand side in (7.18) is also of order $o(n^{-p})$, as $n \to \infty$. With the notation from [CL75], we set

$$T := T(\varepsilon', 1) := \sup\{n \geq 0 : S_n^{*,i} \geq \varepsilon' n\}$$

and

$$\widehat{T} := \widehat{T}(\varepsilon', 1) := \sup\{n \geq 0 : \max_{0\leq k\leq n} S_k^{*,i} \geq \varepsilon' n\}.$$

We already know that T is $\mathbb{P}_i$-a.s. finite since by [CL75, Theorem 5] it has a finite $p = (p+1)\cdot 1 - 1$-th moment, and by definition of T it is clear that $S_{T+l}^{*,i} < \varepsilon'(T+l)$ for all $l \geq 1$ and in particular $S_{T+l}^{*,i} < \varepsilon'(T+l) \leq \varepsilon'(T+m)$ for all $1 \leq l \leq m$. Hence, we find that

$$\widehat{T} = T + \sup\{m \geq 1 : \max_{1\leq k\leq T+m} S_k^{*,i} \geq \varepsilon'(T+m)\}$$

$$
\begin{aligned}
&= T + \sup\{m \geq 1 : \max_{1\leq k\leq T} S_k^{*,i} \geq \varepsilon'(T+m)\} \\
&= T + \sup\{m \geq 1 : \frac{1}{\varepsilon'} \max_{1\leq k\leq T} S_k^{*,i} - T \geq m\} \\
&= T + \lfloor \frac{1}{\varepsilon'} \max_{1\leq k\leq T} S_k^{*,i} \rfloor - T = \lfloor \frac{1}{\varepsilon'} \max_{1\leq k\leq T} S_k^{*,i} \rfloor \\
&\leq \frac{1}{\varepsilon'} \max_{1\leq k\leq T} S_k^{*,i}
\end{aligned}
$$

which has finite $p = ((p+1)\cdot 1 - 1)/1$-th moment by [CL75, Theorem 5]. Thus, $\widehat{T}$ has finite p-th moment and hence, as $n \to \infty$,

$$
\mathbb{P}_i\left(\max_{0\leq k\leq n} S_k^{*,i} > n\varepsilon'\right) \leq \mathbb{P}_i(\widehat{T} \geq n) = o(n^{-p}).
$$

We still need to establish the suggested relation. In order to do this, we decompose

$$
\mathbb{P}_\pi\left(\max_{0\leq k\leq n} S_k^{*,i} > n\varepsilon\right) \leq \mathbb{P}_\pi\Big(S_1^{*,i} > n\frac{\varepsilon}{2}\Big) + \mathbb{P}_\pi\Big(\max_{2\leq k\leq n} S_k^{*,i} > n\varepsilon, S_1^{*,i} \leq n\frac{\varepsilon}{2}\Big),
$$

and latter expression is bounded by

$$
\begin{aligned}
\mathbb{P}_\pi\Big(\max_{2\leq k\leq n} S_k^{*,i} > n\varepsilon, S_1^{*,i} \leq n\frac{\varepsilon}{2}\Big) &= \mathbb{P}_\pi\Big(\max_{2\leq k\leq n} (S_k^{*,i} - S_1^{*,i}) > n\varepsilon - S_1^{*,i}, S_1^{*,i} \leq n\frac{\varepsilon}{2}\Big) \\
&= \int_{(-\infty, n\frac{\varepsilon}{2}]} \mathbb{P}_i\Big(\max_{1\leq k\leq n-1} S_k^{*,i} > n\varepsilon - x\Big)\mathbb{P}_\pi(S_1^{*,i} \in \mathrm{d}x) \\
&\leq \mathbb{P}_i\Big(\max_{0\leq k\leq n} S_k^{*,i} > n\frac{\varepsilon}{2}\Big).
\end{aligned}
$$

Set $\varepsilon' := \frac{\varepsilon}{2}$, then we are in the situation that we talked about before, with toll term $\mathbb{P}_\pi(S_1^{*,i} > \frac{\varepsilon}{2}n)$. Hence, we require $\mathbb{E}_\pi[(S_1^{*,i})^+]^p < \infty$ to obtain the desired result in the first case. Again, Lemma B.8 guarantees that this holds under (A1)-(A3). When starting w.r.t. $\mathbb{P}_i$ instead of $\mathbb{P}_\pi$, then the tracing back procedure from before is not necessary.

Suppose $\mathbb{E}_i|\log L(^{\#}\vec{\Psi}_1^i)| = \infty$. Pick $\gamma \in (0,1)$ and c such that $-\infty < \mathbb{E}_i \log(L(^{\#}\vec{\Psi}_1^i) \vee c) < 0$ and $\gamma \in ((\gamma_{i,c}^*)^{\frac{1}{m_{ii}}}, 1)$. This is possible by Lemma B.5. If we set $L_{k,c}^{*,i} := (L(^{\#}\vec{\Psi}_k^i) \vee c)/\gamma_{i,c}^*$, then we are in the situation of the previous case with all necessary conditions fulfilled and we only need to remark that

$$
\mathbb{P}_\pi\left(\prod_{k=1}^{\tau(n)-1} L(^{\#}\vec{\Psi}_k^i) > \gamma^n\right) \leq \mathbb{P}_\pi\left(\prod_{k=1}^{\tau(n)-1} (L(^{\#}\vec{\Psi}_k^i) \vee c) > \gamma^n\right)
$$

and $\mathbb{E}_\pi\big(\log^+(L(^{\#}\vec{\Psi}_1^i) \vee c)\big)^p = \mathbb{E}_\pi\big(\log^+ L(^{\#}\vec{\Psi}_1^i)\big)^p < \infty$. □

A question that will arise in the crucial lemmas 7.20 and 7.21 is, when $\mathbb{E}_i \log_*^{p+1} C_1 < \infty$ and $\mathbb{E}_i \log_*^{p+1} G_1 < \infty$ hold for $p > 0$. As expected, this is the case under (A1)-(A3).

Lemma 7.19. *The following assertions hold under* (A1)-(A3)*:*

(a) $\mathbb{E}_i \log_*^{p+1} C_1 < \infty$.

(b) $\mathbb{E}_i \log_*^{p+1} G_1 < \infty$.

Proof. We found in Lemma 7.9 that

$$\log_* C_1 \le 1 + \sum_{n=1}^{\sigma_1(i)} \log_* L({}^{\#}\Psi_n) + \sum_{n=1}^{\sigma_1(i)} \log_* d(x_0, {}^{\#}\Psi_n(x_0)),$$

thus

$$\mathbb{E}_i \log_*^{p+1} C_1 \le C + C\mathbb{E}_i \left(\sum_{n=1}^{\sigma_1(i)} \log_* L({}^{\#}\Psi_n) \right)^{p+1} + C\mathbb{E}_i \left(\sum_{n=1}^{\sigma_1(i)} \log_* d(x_0, {}^{\#}\Psi_n(x_0)) \right)^{p+1} < \infty$$

for some $C > 0$ by (A2) and (A3). Using Remark 6.15 proves (a). To show (b), we recall

$$\begin{aligned} 1 + G_1 &= \max\{1 + L({}^{\#}\Psi_1), \dots, 1 + L({}^{\#}\Psi_1 \cdots {}^{\#}\Psi_{\sigma_1(i)})\} \\ &\le \max\{1 + L({}^{\#}\Psi_1), \dots, \prod_{n=1}^{\sigma_1(i)} (1 + L({}^{\#}\Psi_n))\} = \prod_{n=1}^{\sigma_1(i)} (1 + L({}^{\#}\Psi_n)) \end{aligned} \tag{7.19}$$

and hence

$$\log_*^{p+1} G_1 \le \left(\sum_{n=1}^{\sigma_1(i)} \log_* L({}^{\#}\Psi_n) \right)^{p+1}$$

which has finite expectation by (A2) and Remark 6.15. □

The next Lemma forms the first part of the proof of Theorem 7.2(a).

Lemma 7.20. *If* (A1)-(A3) *hold, then for all $\gamma \in ((\gamma_i^*)^{\frac{1}{m_{ii}}}, 1)$ and $c > 0$*

$$\sum_{n \ge 1} n^{p-1} \mathbb{P}_\pi \left(\prod_{k=1}^{\tau(n)-1} L({}^{\#}\vec{\Psi}_k^i) \cdot C_{\tau(n)} > \frac{\gamma^n}{c} \right) < \infty,$$

$$\lim_{n \to \infty} n^p \mathbb{P}_\pi \left(\prod_{k=1}^{\tau(n)-1} L({}^{\#}\vec{\Psi}_k^i) \cdot C_{\tau(n)} > \frac{\gamma^n}{c} \right) = 0$$

and the above statements also hold with $\mathbb{P}_i$ instead of $\mathbb{P}_\pi$.

Proof. Fix $n \ge 1$. We start with $\gamma \in ((\gamma_i^*)^{\frac{1}{m_{ii}}}, 1)$ and $\beta \in ((\gamma_i^*)^{\frac{1}{m_{ii}}}, \gamma)$. We decompose

$$\begin{aligned} \mathbb{P}_\pi \left(\prod_{k=1}^{\tau(n)-1} L({}^{\#}\vec{\Psi}_k^i) \cdot C_{\tau(n)} > \frac{\gamma^n}{c} \right) &= \mathbb{P}_\pi \left(\prod_{k=1}^{\tau(n)-1} L({}^{\#}\vec{\Psi}_k^i) \cdot C_{\tau(n)} > \frac{\gamma^n}{c}, 0 < \prod_{k=1}^{\tau(n)-1} L({}^{\#}\vec{\Psi}_k^i) \le \beta^n \right) \\ &\quad + \mathbb{P}_\pi \left(\prod_{k=1}^{\tau(n)-1} L({}^{\#}\vec{\Psi}_k^i) \cdot C_{\tau(n)} > \frac{\gamma^n}{c}, \prod_{k=1}^{\tau(n)-1} L({}^{\#}\vec{\Psi}_k^i) > \beta^n \right) \\ &\le \mathbb{P}_\pi \left(C_{\tau(n)} > \frac{1}{c} \left(\frac{\gamma}{\beta} \right)^n \right) + \mathbb{P}_\pi \left(\prod_{k=1}^{\tau(n)-1} L({}^{\#}\vec{\Psi}_k^i) > \beta^n \right). \end{aligned}$$

Set $\varepsilon := \log(\gamma/\beta) > 0$. From Lemma 7.15 we learn that

$$\mathbb{P}_\pi \left(C_{\tau(n)} > \frac{1}{c} \left(\frac{\gamma}{\beta} \right)^n \right) \le \mathbb{P}_\pi \left(\log_* c C_{\tau(n)} > \varepsilon n \right)$$

$$\leq \mathbb{E}_i\sigma_1(i)\mathbb{1}_{\{\log_* cC_1>\varepsilon n\}} + \mathbb{P}_\pi(\sigma_1(i) \geq n, \log_* cC_1 > \varepsilon n).$$

The latter summand is smaller than or equal to $\mathbb{P}_\pi(\sigma_1(i)) \geq n)$, so it is obviously summable over $n \geq 1$ weighted with n^{p-1}, and vanishes, as $n \to \infty$, when multiplied by n^p, both if $\mathbb{E}_\pi\sigma_1(i)^p < \infty$. Regarding the first summand, we remark that

$$\mathbb{E}_i\sigma_1(i)\log_*^p cC_1 \leq (\mathbb{E}_i\sigma_1(i)^{p+1})^{1/(p+1)}(\mathbb{E}_i\log_*^{p+1} cC_1)^{p/(p+1)} < \infty$$

and thus $n^p\mathbb{E}_i\sigma_1(i)\mathbb{1}_{\{\log_* cC_1>\varepsilon n\}}$ vanishes, as $n \to \infty$, as well as

$$\sum_{n\geq 1} n^{p-1}\mathbb{E}_i\sigma_1(i)\mathbb{1}_{\{\log_* cC_1>\varepsilon n\}} \leq C\mathbb{E}_i\sigma_1(i)\log_*^p cC_1 < \infty$$

if $\mathbb{E}_i\sigma_1(i)^{p+1} < \infty$ and $\mathbb{E}_i\log_*^{p+1} C_1 < \infty$. The finiteness follows from Lemma 7.19. We still need to deal with the second summand from our first estimate, but this was done in Lemma 7.17 for both cases $\mathbb{E}_i|\log L({}^\#\vec{\Psi}_1^i)| < \infty$ and $= \infty$ (note that $\beta \in ((\gamma_i^*)^{\frac{1}{m_{ii}}}, 1))$. This completes the proof, because the $\mathbb{P}_i$-case is easy. ☐

We still need another ingredient to prove the desired convergence-rate results. Therefore, having Lemma 7.7(b) in mind, we recall

$$G_n := \max\{L({}^\#\Psi_{\sigma_{n-1}(i)+1}), \ldots, L({}^\#\Psi_{\sigma_{n-1}(i)+1}\cdots{}^\#\Psi_{\sigma_n(i)})\},$$

for $n \geq 1$.

Lemma 7.21. *If* (A1)-(A3) *hold, then for all* $\gamma \in ((\gamma_i^*)^{\frac{1}{m_{ii}}}, 1)$ *and* $c > 0$

$$\sum_{n\geq 1} n^{p-1}\mathbb{P}_\pi\left(\prod_{k=1}^{\tau(n)-1} L({}^\#\vec{\Psi}_k^i)\cdot G_{\tau(n)} > \frac{\gamma^n}{c}\right) < \infty,$$

$$\lim_{n\to\infty} n^p\mathbb{P}_\pi\left(\prod_{k=1}^{\tau(n)-1} L({}^\#\vec{\Psi}_k^i)\cdot G_{\tau(n)} > \frac{\gamma^n}{c}\right) = 0$$

and the above statements also hold with $\mathbb{P}_i$ *instead of* $\mathbb{P}_\pi$.

Proof. We omit the proof since it is exactly the same as the proof Lemma 7.20 using Lemma 7.19(b) instead of (a). ☐

As for the last ingredient, recall that

$$D_n := \sum_{j=1}^{\infty}\prod_{k=n+1}^{n+j-1} L({}^\#\vec{\Psi}_k^i)\ d({}^\#\vec{\Psi}_{n+j}^i(x_0), x_0)$$

for $n \geq 0$ and that $\tau(n)$ is a stopping time w.r.t.

$$\mathcal{F}_n := \sigma((\sigma_k(i), {}^\#\vec{\Psi}_k^i) : k = 0, \ldots, n), \quad n \geq 0$$

and $(\sigma_{n+k}(i), {}^\#\vec{\Psi}_{n+k}^i)_{k\geq 1}$ is independent of $\mathcal{F}_n$ for each $n \geq 0$. The sequence $(D_{\tau(n)})_{n\geq 0}$ is stationary w.r.t. $\mathbb{P}_i$ because $({}^\#\vec{\Psi}_k^i)_{k\geq 1}$ is stationary and $({}^\#\vec{\Psi}_{\tau(n)+k}^i)_{k\geq 1}$ has the same distribution w.r.t. $\mathbb{P}_i$ as $({}^\#\vec{\Psi}_k^i)_{k\geq 1}$. Thus, we also get

$$D_{\tau(n)} \stackrel{\mathrm{d}}{=} D_0 \quad \text{w.r.t. } \mathbb{P}_i$$

for all $n \geq 0$. Furthermore,

$$\mathbb{P}_\pi(({}^{\#}\vec{\Psi}^i_{\tau(n)+k})_{k\geq 1} \in \cdot) = \mathbb{P}_\pi(({}^{\#}\vec{\Psi}^i_k)_{k\geq 2} \in \cdot) = \mathbb{P}_i(({}^{\#}\vec{\Psi}^i_k)_{k\geq 1} \in \cdot)$$

for $n \geq 1$ and thus

$$\mathbb{P}_\pi(D_{\tau(n)} \in \cdot) = \mathbb{P}_i(D_0 \in \cdot)$$

for all $n \geq 1$, we only have to exclude "$n = 0$". The following Lemma is needed in the proof of Lemma 7.23.

Lemma 7.22. *If* (A1)-(A3) *hold, then*

$$\mathbb{E}_i \log_*^p D_0 = \mathbb{E}_i \log_*^p D_{\tau(n)} < \infty$$

and equivalently

$$\mathbb{E}_i(\log^+ D_0)^p = \mathbb{E}_i(\log^+ D_{\tau(n)})^p < \infty.$$

Proof. It suffices to show $\mathbb{E}_i \log_*^p D_0 < \infty$. We are guided by Lemma 7.6. *Case 1*: If

$$\mathbb{P}_i(d({}^{\#}\vec{\Psi}^i_1(x_0), x_0) = 0) = 1,$$

then $\log_*^p D_0 = 0$ $\mathbb{P}_i$-a.s. *Case 2*: If $\mathbb{P}_i(d({}^{\#}\vec{\Psi}^i_1(x_0), x_0) = 0) < 1$ and $\mathbb{E}_i \log L({}^{\#}\vec{\Psi}^i_1) \in (-\infty, 0)$ ($\mathbb{E}_i \log^+ d({}^{\#}\vec{\Psi}^i_1(x_0), x_0) < \infty$ holds by assumption), then [Iks06, Theorem 2.1] (or [AI09, Theorem 1.2]) requires

$$\mathbb{E}_i \log^+ L({}^{\#}\vec{\Psi}^i_1)(\log^+ L({}^{\#}\vec{\Psi}^i_1))^p = \mathbb{E}_i(\log^+ L({}^{\#}\vec{\Psi}^i_1))^{p+1} < \infty$$

and

$$\mathbb{E}_i \log^+ d({}^{\#}\vec{\Psi}^i_1(x_0), x_0)(\log^+ d({}^{\#}\vec{\Psi}^i_1(x_0), x_0))^p = \mathbb{E}_i(\log^+ d({}^{\#}\vec{\Psi}^i_1(x_0), x_0))^{p+1} < \infty,$$

to ensure that

$$\mathbb{E}_i(\log^+ D_0)^p < \infty.$$

Both requirements are clearly met under (A1)-(A3). *Case 3*: If $\mathbb{P}_i(d({}^{\#}\vec{\Psi}^i_1(x_0), x_0) = 0) < 1$ and $\mathbb{E}_i \log L({}^{\#}\vec{\Psi}^i_1) = -\infty$. Then, we choose $c \in (0, 1)$ such that

$$\mathbb{E}_i \log(L({}^{\#}\vec{\Psi}^i_1) \vee c) \in (-\infty, 0).$$

Due to

$$(\log^+ D_0)^p \leq \left(\log^+ \sum_{j=1}^{\infty} \prod_{k=1}^{j-1} (L({}^{\#}\vec{\Psi}^i_k) \vee c)\ d({}^{\#}\vec{\Psi}^i_j(x_0), x_0)\right)^p =: (\log^+ D_0^c)^p,$$

where D_0^c has the obvious meaning, it follows from the previous case that

$$\mathbb{E}_i(\log^+(L({}^{\#}\vec{\Psi}^i_1) \vee c))^{p+1} < \infty$$

implies $\mathbb{E}_i(\log^+ D_0)^p \leq \mathbb{E}_i(\log^+ D_0^c)^p < \infty$. However, $c \in (0, 1)$ and thus

$$\mathbb{E}_i(\log^+(L({}^{\#}\vec{\Psi}^i_1) \vee c))^{p+1} = \mathbb{E}_i(\log^+ L({}^{\#}\vec{\Psi}^i_1))^{p+1} < \infty.$$

□

The next Lemma forms the last part of the proof of Theorem 7.2(a).

Lemma 7.23. *If* (A1)-(A3) *hold, then for all* $\gamma \in ((\gamma_i^*)^{\frac{1}{m_{ii}}}, 1)$ *and* $c > 0$

$$\sum_{n\geq 1} n^{p-1}\mathbb{P}_\pi\left(\prod_{k=1}^{\tau(n)} L({}^\#\vec{\Psi}_k^i)\cdot D_{\tau(n)} > \frac{\gamma^n}{c}\right) < \infty,$$

$$\lim_{n\to\infty} n^p\mathbb{P}_\pi\left(\prod_{k=1}^{\tau(n)} L({}^\#\vec{\Psi}_k^i)\cdot D_{\tau(n)} > \frac{\gamma^n}{c}\right) = 0$$

and the above statements also hold with $\mathbb{P}_i$ *instead of* $\mathbb{P}_\pi$*.*

Proof. Fix $n \geq 1$. We start with $\gamma \in ((\gamma_i^*)^{\frac{1}{m_{ii}}}, 1)$ and $\beta \in ((\gamma_i^*)^{\frac{1}{m_{ii}}}, \gamma)$. We decompose

$$\mathbb{P}_\pi\left(\prod_{k=1}^{\tau(n)} L({}^\#\vec{\Psi}_k^i)\cdot D_{\tau(n)} > \frac{\gamma^n}{c}\right) \leq \mathbb{P}_\pi\left(D_{\tau(n)} > \frac{1}{c}\left(\frac{\gamma}{\beta}\right)^n\right) + \mathbb{P}_\pi\left(\prod_{k=1}^{\tau(n)} L({}^\#\vec{\Psi}_k^i) > \beta^n\right). \tag{7.20}$$

Again, set $\varepsilon := \log(\gamma/\beta) > 0$. We have

$$\mathbb{P}_\pi\left(D_{\tau(n)} > \frac{1}{c}\left(\frac{\gamma}{\beta}\right)^n\right) = \mathbb{P}_i\left(D_0 > \frac{1}{c}\left(\frac{\gamma}{\beta}\right)^n\right) \leq \mathbb{P}_i\left(\log_* cD_0 > \varepsilon n\right)$$

which is obviously summable over $n \geq 1$ weighted with n^{p-1}, and vanishes, as $n \to \infty$, when multiplied by n^p, both if $\mathbb{E}_i \log_*^p D_0 < \infty$. The second summand from (7.20) was treated in Lemma 7.17 and Remark 7.18 (note that $\beta \in ((\gamma_i^*)^{\frac{1}{m_{ii}}}, 1)$). This completes the proof. □

Collecting everything together, we can prove Theorem 7.2(a) without further ado. The proof follows that of [AF01, Theorem 2.2(a)].

Proof of Theorem 7.2(a). Fix $n \geq 1$. We have

$$\begin{aligned}
&\mathbb{P}_i(d({}^\#\widehat{X}_n^x, {}^\#\widehat{X}_\infty) > \gamma^n)\\
&\leq \mathbb{P}_i\Bigg(\prod_{k=1}^{\tau(n)-1} L({}^\#\vec{\Psi}_k^i)\cdot C_{\tau(n)} + \prod_{k=1}^{\tau(n)} L({}^\#\vec{\Psi}_k^i)\cdot D_{\tau(n)}\\
&\qquad + d(x,x_0)\prod_{k=1}^{\tau(n)-1} L({}^\#\vec{\Psi}_k^i)\cdot G_{\tau(n)} > \gamma^n\Bigg)\\
&\leq \mathbb{P}_i\left(\prod_{k=1}^{\tau(n)-1} L({}^\#\vec{\Psi}_k^i)\cdot C_{\tau(n)} > \frac{\gamma^n}{3}\right) + \mathbb{P}_i\left(\prod_{k=1}^{\tau(n)} L({}^\#\vec{\Psi}_k^i)\cdot D_{\tau(n)} > \frac{\gamma^n}{3}\right)\\
&\quad + \mathbb{P}_i\left(d(x,x_0)\prod_{k=1}^{\tau(n)-1} L({}^\#\vec{\Psi}_k^i)\cdot G_{\tau(n)} > \frac{\gamma^n}{3}\right)
\end{aligned} \tag{7.21}$$

by Lemma 7.7. Now, Lemma 7.20 deals with the first term and we just examined the second term in Lemma 7.23. To extract the effect of x in the last term when $\mathbb{E}_i|\log L({}^\#\vec{\Psi}_1^i)| < \infty$, we write $\gamma = (\gamma_i^*)^b$ for $b \in (0, \frac{1}{m_{ii}})$, choose $a \in (b, \frac{1}{m_{ii}})$ and estimate

$$\begin{aligned}
&\mathbb{P}_i\left(d(x,x_0)\prod_{k=1}^{\tau(n)-1} L(^{\#}\vec{\Psi}_k^i)\cdot G_{\tau(n)} > \frac{\gamma^n}{3}\right)\\
&\le \mathbb{P}_i\left(d(x,x_0)\prod_{k=1}^{\tau(n)-1} L(^{\#}\vec{\Psi}_k^i)\cdot G_{\tau(n)} > \frac{\gamma^n}{3}, \prod_{k=1}^{\tau(n)-1} L(^{\#}\vec{\Psi}_k^i)\cdot G_{\tau(n)} > \frac{((\gamma_i^*)^a)^n}{3}\right)\\
&\quad + \mathbb{P}_i\left(d(x,x_0)\prod_{k=1}^{\tau(n)-1} L(^{\#}\vec{\Psi}_k^i)\cdot G_{\tau(n)} > \frac{\gamma^n}{3}, 0 < \prod_{k=1}^{\tau(n)-1} L(^{\#}\vec{\Psi}_k^i)\cdot G_{\tau(n)} \le \frac{((\gamma_i^*)^a)^n}{3}\right)\\
&\le \mathbb{P}_i\left(\prod_{k=1}^{\tau(n)-1} L(^{\#}\vec{\Psi}_k^i)\cdot G_{\tau(n)} > \frac{((\gamma_i^*)^a)^n}{3}\right) + \mathbb{P}_i\left(d(x,x_0) > (1/\gamma_i^*)^{(a-b)n}\right).
\end{aligned} \tag{7.22}$$

Lemma 7.21 handles the first term and the last term is deterministic and equals 1 if $n < \frac{\log d(x,x_0)}{(a-b)\log 1/\gamma_i^*} =: n_0$ and 0 otherwise. So the limit of the last term multiplied by n^p is trivially 0, and we complete the proof by noting

$$\sum_{n\ge 1} n^{p-1}\mathbb{P}_i\left(d(x,x_0) > (1/\gamma_i^*)^{(a-b)n}\right) = \sum_{n=1}^{\lceil n_0\rceil - 1} n^{p-1} \le n_0^p \le C_\gamma \cdot \log_*^p d(x,x_0)$$

for some constant $C_\gamma > 0$ depending only on γ and not on x. Otherwise, if $\mathbb{E}_i|\log L(^{\#}\vec{\Psi}_1^i)| = \infty$, then the proof is easier: Choose $\gamma \in (0,1)$, then also $\gamma^2 \in (0,1)$ and in the same way as above

$$\begin{aligned}
&\mathbb{P}_i\left(d(x,x_0)\prod_{k=1}^{\tau(n)-1} L(^{\#}\vec{\Psi}_k^i)\cdot G_{\tau(n)} > \frac{\gamma^n}{3}\right)\\
&\le \mathbb{P}_i\left(\prod_{k=1}^{\tau(n)-1} L(^{\#}\vec{\Psi}_k^i)\cdot G_{\tau(n)} > \frac{(\gamma^2)^n}{3}\right) + \mathbb{P}_i\left(d(x,x_0) > \gamma^{-n}\right),
\end{aligned}$$

so n_0 changes to $\frac{\log d(x,x_0)}{\log 1/\gamma}$, and the subsequent estimation is still valid with different C_γ. ☐

Proof of Theorem 7.2(b). We can copy the proof of Theorem 7.2(a) while replacing $\mathbb{P}_i$ by $\mathbb{P}_\pi$. The required lemmas always provide a statement in this setting. ☐

Proof of Theorem 7.2(c). Choose $\gamma \in ((\gamma_i^*)^{\frac{1}{m_{ii}}}, 1)$ and then A big enough such that $\gamma^n < A(n+1)^{-p}$ for all $n \ge 0$, and such that the first two summands on the right side of (7.21) and the first summand on the right side of (7.22) are $< \frac{A}{3}(n+1)^{-p}$ for all $n \ge 0$ (with $\mathbb{P}_\pi$ instead of $\mathbb{P}_i$). The latter is possible due to the lemmas mentioned in the proof of Theorem 7.2(a). With this choice, we get

$$\mathbb{P}_\pi\left(d(^{\#}\widehat{X}_n^x, {}^{\#}\widehat{X}_\infty) \ge \frac{A}{(n+1)^p}\right) \le \mathbb{P}_\pi(d(^{\#}\widehat{X}_n^x, {}^{\#}\widehat{X}_\infty) > \gamma^n) < \frac{A}{(n+1)^p} + \mathbb{1}_{\{n<n_0(x)\}}$$

with $n_0(x) = n_0$ from the proof of Theorem 7.2(a). We obtain

$$\frac{A}{(n+1)^p} + \mathbb{1}_{\{n<n_0(x)\}} = \frac{A + (n+1)^p\mathbb{1}_{\{n<n_0(x)\}}}{(n+1)^p} \le \frac{A + (n_0(x)+1)^p}{(n+1)^p}$$

$$\leq \frac{(A+2^p)+2^pC\log_*^p d(x,x_0)}{(n+1)^p} =: \frac{A_1+A_2\log_*^p d(x,x_0)}{(n+1)^p}$$

with some constant $C>0$ only depending on γ. Set $A_x := A_1 + A_2\log_*^p d(x,x_0)$, then we summarize

$$\mathbb{P}_\pi\left(d({}^{\#}\widehat{X}_n^x, {}^{\#}\widehat{X}_\infty) \geq \frac{A_x}{(n+1)^p}\right) \leq \mathbb{P}_\pi\left(d({}^{\#}\widehat{X}_n^x, {}^{\#}\widehat{X}_\infty) \geq \frac{A}{(n+1)^p}\right) < \frac{A_x}{(n+1)^p}$$

and this yields $d_{\mathrm{Pr}}(\mathbb{P}_\pi^{{}^{\#}\widehat{X}_n^x}, \mathbb{P}_\pi^{{}^{\#}\widehat{X}_\infty}) \leq \frac{A_x}{(n+1)^p}$ by Lemma B.4 and

$$d_{\mathrm{Pr}}(\mathbb{P}_\pi^{{}^{\#}\widehat{X}_n^x}, \mathbb{P}_\pi^{{}^{\#}\widehat{X}_\infty}) = d_{\mathrm{Pr}}(\mathbb{P}_\pi^{X_n^x}, \mathbb{P}_\pi^{{}^{\#}\widehat{X}_\infty})$$

because $\mathbb{P}_\pi^{{}^{\#}\widehat{X}_n^x} = \mathbb{P}_\pi^{X_n^x}$. □

Proof of Theorem 7.2(d). Denote by $\widehat{\pi}_i$ the distribution of ${}^{\#}\widehat{X}_\infty$ under $\mathbb{P}_i$. We know from Lemma 7.5 that ${}^{\#}\widehat{X}_\infty$ is the $\mathbb{P}_i$-a.s. limit of the backward iterations $\widehat{Y}_n^x = {}^{\#}\widehat{X}_{\sigma_n(i)}^x$ of the IFS of i.i.d. Lipschitz maps (Y_n^x). Hence, we can cite from [AF01] that

$$\mathbb{E}_i \log_*^p d(x_0, {}^{\#}\widehat{X}_\infty) = \int_{\mathbb{X}} \log_*^p d(x_0,x)\ \mathbb{P}_i^{{}^{\#}\widehat{X}_\infty}(\mathrm{d}x) = \int_{\mathbb{X}} \log_*^p d(x_0,x)\ \widehat{\pi}_i(\mathrm{d}x) < \infty$$

whenever the corresponding Lipschitz maps satisfy

$$\mathbb{E}_i \log_*^{p+1} L({}^{\#}\vec{\Psi}_1^i) < \infty \quad \text{and} \quad \mathbb{E}_i \log_*^{p+1} d(x_0, {}^{\#}\vec{\Psi}_1^i(x_0)) < \infty$$

for some (and thus all) $x_0 \in \mathbb{X}$. This is clearly the case under (A1)-(A3). □

Alternative proof of Theorem 7.2(d). From Lemma 7.7 we know, that

$$d(x_0, {}^{\#}\widehat{X}_\infty) \leq D_0 \quad \mathbb{P}_i\text{-a.s.}$$

which in $\log_*^p$ is $\mathbb{P}_i$-integrable by Lemma 7.22. □

Proof of Theorem 7.2(e). From Lemma 7.13 we know that even

$$d(x_0, {}^{\#}\widehat{X}_\infty) \leq D_0 \quad \mathbb{P}_\pi\text{-a.s.},$$

hence it suffices to verify $\mathbb{E}_\pi \log_*^p D_0 < \infty$. To finish the proof, we recall

$$D_0 = d({}^{\#}\vec{\Psi}_1^i(x_0), x_0) + L({}^{\#}\vec{\Psi}_1^i) D_1$$

from Lemma 7.6. This gives us the estimate

$$\begin{aligned}\mathbb{E}_\pi \log_*^p D_0 &\leq \mathbb{E}_\pi \left(\log_* d({}^{\#}\vec{\Psi}_1^i(x_0), x_0) + \log_* L({}^{\#}\vec{\Psi}_1^i) + \log_* D_1\right)^p \\ &\leq C\left(\mathbb{E}_\pi \log_*^p d({}^{\#}\vec{\Psi}_1^i(x_0), x_0) + \mathbb{E}_\pi \log_*^p L({}^{\#}\vec{\Psi}_1^i) + \mathbb{E}_i \log_*^p D_0\right)\end{aligned}$$

for some $C>0$. The last term is finite by Lemma 7.22 and the first two are finite by Lemma B.8. □

7.4. Convergence rate under geometric-type moment conditions

We will now derive distributional convergence rates under (B1)-(B4) along the path from Section 7.3. In contrast to the polynomial rate there, the rate under (B1)-(B4) turns out to be geometric, as one would expect from the i.i.d. analogue. In fact, Lemma 7.14 already incorporates auxiliary results under (B1)-(B4). To obtain an analogue of Lemma 7.17, we first need to establish an exponential version of (parts of) [CL75, Theorem 5]:

Lemma 7.24. *Let $(S_n)_{n\geq 0}$, $S_0 := 0$, be a centered SRW on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and let $T := \sup\{n \geq 0 : S_n \geq \varepsilon n\}$ for some $\varepsilon > 0$. Suppose $\mathbb{P}(X_1 \neq 0) > 0$ and the moment generating function $\varphi(\theta) := \mathbb{E}e^{\theta X_1}$ exists for at least one $\theta > 0$. Then $Z_T := \max_{1\leq k\leq T} S_k$ has an exponential moment, i.e. $\mathbb{E}e^{\beta Z_T} < \infty$ for some $\beta > 0$.*

Proof. We remark that $Z_T \geq 0$ a.s. by definition of T ($Z_T = 0$ if $T = 0$). Hence, a well-known integration formula yields

$$\mathbb{E}(e^{\beta Z_T} - 1) = \int_0^\infty \beta e^{\beta t}\mathbb{P}(Z_T > t)\ \mathrm{d}t \leq \beta \sum_{n\geq 0} e^{\beta(n+1)}\mathbb{P}(Z_T > n) = \beta e^{\beta} \sum_{n\geq 0} e^{\beta n}\mathbb{P}(Z_T > n),$$

which reduces the problem to verifying the finiteness of the latter series. Therefore, we partition

$$\begin{aligned}
\mathbb{P}(Z_T > n) &= \mathbb{P}\left(\max_{1\leq k\leq T} S_k > n, T > \delta n\right) + \mathbb{P}\left(\max_{1\leq k\leq T} S_k > n, T \leq \delta n\right) \\
&\leq \mathbb{P}\left(\bigcup_{k>\delta n} \{S_k \geq \varepsilon k\}\right) + \mathbb{P}\left(\max_{1\leq k\leq \delta n} S_k > n\right) \\
&\leq \sum_{k\geq \lceil \delta n\rceil} \mathbb{P}(S_k \geq \varepsilon k) + \sum_{k=1}^{\lfloor \delta n\rfloor} \mathbb{P}(S_k > n)
\end{aligned}$$

with a small $\delta > 0$ that will be appropriately chosen later. Now, *Cramér's theorem* from large deviations theory yields

$$\mathbb{P}(S_k \geq \varepsilon k) \leq e^{-kI(\varepsilon)}$$

for all k and some rate function $I(\varepsilon) > 0$ for all $\varepsilon > 0$. Hence,

$$\begin{aligned}
\sum_{n\geq 0} e^{\beta n} \sum_{k\geq \lceil \delta n\rceil} \mathbb{P}(S_k \geq \varepsilon k) &\leq \sum_{n\geq 0} e^{\beta n} \sum_{k\geq \lceil \delta n\rceil} e^{-kI(\varepsilon)} = \sum_{n\geq 0} e^{\beta n} \sum_{k\geq 0} \left(e^{-I(\varepsilon)}\right)^{k+\lceil \delta n\rceil} \\
&= \frac{1}{1-e^{-I(\varepsilon)}} \sum_{n\geq 0} e^{\beta n} \left(e^{-I(\varepsilon)}\right)^{\lceil \delta n\rceil} \\
&= \frac{1}{1-e^{-I(\varepsilon)}} \sum_{n\geq 0} e^{-n(I(\varepsilon)\frac{\lceil \delta n\rceil}{n} - \beta)} \\
&\leq \frac{1}{1-e^{-I(\varepsilon)}} \sum_{n\geq 0} e^{-n(I(\varepsilon)\delta - \beta)} < \infty,
\end{aligned}$$

whenever $I(\varepsilon)\delta - \beta > 0$, i.e. $\beta \in (0, I(\varepsilon)\delta)$. To deal with the second part, we estimate

$$\mathbb{P}(S_k > n) = \mathbb{P}(e^{\theta S_k} > e^{\theta n}) \leq e^{-\theta n}\varphi(\theta)^k$$

for $1 \leq k \leq \lfloor \delta n \rfloor$ and $\theta > 0$ with finite $\varphi(\theta)$. Thus,

$$\sum_{n\geq 0} e^{\beta n} \sum_{k=1}^{\lfloor \delta n \rfloor} \mathbb{P}(S_k > n) \leq \sum_{n\geq 0} e^{-n(\theta-\beta)} \sum_{k=1}^{\lfloor \delta n \rfloor} \varphi(\theta)^k \leq \frac{1}{1-\varphi(\theta)} \sum_{n\geq 0} e^{-n(\theta-\beta)} (1-\varphi(\theta)^{\lfloor \delta n \rfloor + 1})$$

since $\varphi(\theta) = \mathbb{E} e^{\theta X_1} > e^{\theta \mathbb{E} X_1} = 1$ for $\theta > 0$ by Jensen's inequality. Now,

$$\sum_{n\geq 0} e^{-n(\theta-\beta)} < \infty \quad \text{iff} \quad \theta - \beta > 0,$$

and

$$\sum_{n\geq 0} e^{-n(\theta-\beta)} \varphi(\theta)^{\lfloor \delta n \rfloor + 1} = \varphi(\theta) \sum_{n\geq 0} e^{-n(\theta-\beta-\frac{\lfloor \delta n \rfloor}{n}\log\varphi(\theta))} \leq \varphi(\theta) \sum_{n\geq 0} e^{-n(\theta-\beta-\delta\log\varphi(\theta))} < \infty,$$

whenever $\beta \in (0, \theta - \delta \log \varphi(\theta))$, where we note that $\log \varphi(\theta) > 0$. Collecting all ingredients, we select $\theta > 0$ such that $\varphi(\theta) < \infty$ and $\delta > 0$ such that $\theta - \delta \log \varphi(\theta) > 0$, i.e. $\delta \in (0, \frac{\theta}{\log \varphi(\theta)})$. Then $\beta \in (0, I(\varepsilon)\delta \wedge (\theta - \delta\varphi(\theta)))$ satisfies the desired properties. □

As previously announced, the following result is an analogue of Lemma 7.17.

Lemma 7.25. *If* (B1)-(B4) *hold, then for all* $\gamma \in ((\gamma_i^*)^{\frac{1}{m_{ii}}}, 1)$ *there exists an* $\alpha > 1$*, depending on* γ *and on* i*, such that*

$$\lim_{n\to\infty} \alpha^n \mathbb{P}_\pi \left(\prod_{k=1}^{\tau(n)-1} L(^\# \vec{\Psi}_k^i) > \gamma^n \right) = 0.$$

The same holds with $\mathbb{P}_i$ *instead of* $\mathbb{P}_\pi$ *with possibly different* α*.*

Proof. We proceed as in the proof of Lemma 7.17 and use its notation. Let $\mathbb{E}_i |\log L(^\# \vec{\Psi}_1^i)| < \infty$ and fix $n \geq 1$. We start with $\gamma \in ((\gamma_i^*)^{\frac{1}{m_{ii}}}, 1)$, $b \in (0, \frac{1}{m_{ii}})$ with $\gamma = (\gamma_i^*)^b$ and $a \in (b, \frac{1}{m_{ii}})$. Again, we consider

$$\mathbb{P}_\pi \left(\sum_{k=1}^{\tau(n)-1} \log L_k^{*,i} > n \log \left((\gamma_i^*)^{b - \frac{\tau(n)-1}{n}} \right), \tau(n) \leq an+1 \right) \leq \mathbb{P}_\pi(\tau(n) \leq an+1)$$

which converges to 0, as $n \to \infty$, by Lemma 7.14 when multiplied by some α_1^n, $\alpha_1 > 1$. On the other hand, with $\varepsilon := \log((\gamma_i^*)^{b-a}) > 0$,

$$\mathbb{P}_\pi \left(\sum_{k=1}^{\tau(n)-1} \log L_k^{*,i} > n \log \left((\gamma_i^*)^{b - \frac{\tau(n)-1}{n}} \right), \tau(n) > an+1 \right) \leq \mathbb{P}_\pi \left(\max_{0\leq k\leq n} S_k^{*,i} > n\varepsilon \right)$$

as in (7.18). Again, we relate this expression to $\mathbb{P}_i\big(\max_{0\leq k\leq n} S_k^{*,i} > n\varepsilon'\big)$ (and some arising toll term) for some $\varepsilon' > 0$. Taking such a relation for granted, we examine the $\mathbb{P}_i$ case first: Recall

$$T := T(\varepsilon', 1) := \sup\{n \geq 0 : S_n^{*,i} \geq \varepsilon' n\}$$

and

$$\widehat{T} := \widehat{T}(\varepsilon', 1) := \sup\{n \geq 0 : \max_{0\leq k\leq n} S_k^{*,i} \geq \varepsilon' n\}.$$

Since

$$\mathbb{P}_i\left(\max_{0\le k\le n} S_k^{*,i} > n\varepsilon'\right) \le \mathbb{P}_i(\widehat{T}\ge n)$$

and

$$\widehat{T} \le \frac{1}{\varepsilon'}\max_{1\le k\le T} S_k^{*,i}$$

as seen in the proof of Lemma 7.17, we aim at verifying the existence of exponential moments of the latter expression. This then entails the exponential decay of $\mathbb{P}_i\big(\max_{0\le k\le n} S_k^{*,i} > n\varepsilon'\big)$. For this purpose, we use Lemma 7.24 which gives us a criterion for the existence of an exponential moment of $\max_{1\le k\le T} S_k^{*,i}$. The RW is centered and recalling Remark 6.19 we note that

$$\mathbb{E}_i e^{\theta\log L_1^{*,i}} = e^{-\theta\log\gamma_i^*}\mathbb{E}_i e^{\theta\log L(^\#\vec{\Psi}_1^i)} = e^{-\theta\log\gamma_i^*}\mathbb{E}_i L(^\#\vec{\Psi}_1^i)^\theta < \infty$$

for $\theta\in(0,p]$. If additionally $\mathbb{P}_i(\log L_1^{*,i}\ne 0)>0$ holds, then Lemma 7.24 gives us the existence of $\beta_i>0$ such that $Z_T^{\varepsilon'} := \frac{1}{\varepsilon'}\max_{1\le k\le T} S_k^{*,i}$ satisfies $\mathbb{E}_i e^{\beta_i\varepsilon' Z_T^{\varepsilon'}}<\infty$. Hence, as $n\to\infty$,

$$\left(e^{\beta_i\varepsilon'}\right)^n\mathbb{P}_i(\widehat{T}\ge n) \le \left(e^{\beta_i\varepsilon'}\right)^n\mathbb{P}_i(Z_T^{\varepsilon'}\ge n) \le \mathbb{E}_i e^{\beta_i\varepsilon' Z_T^{\varepsilon'}}\mathbb{1}_{\{Z_T^{\varepsilon'}\ge n\}} \to 0.$$

We set $\alpha_2 := e^{\beta_i\varepsilon'}>1$ in this case.

If $\log L_1^{*,i}=0$ $\mathbb{P}_i$-a.s. and thus $L(^\#\vec{\Psi}_k^i)\equiv\gamma_i^*$ $\mathbb{P}_i$-a.s. for all k, Lemma 7.24 is not applicable but the desired conclusion then follows even easier: With the same choices as in the beginning of this proof, we write

$$\begin{aligned}\mathbb{P}_i\left(\prod_{k=1}^{\tau(n)-1} L(^\#\vec{\Psi}_k^i) > \gamma^n\right) &= \mathbb{P}_i\left((\gamma_i^*)^{\tau(n)-1} > \gamma^n\right) = \mathbb{P}_i\left((\gamma_i^*)^{bn-\tau(n)+1} < 1\right)\\ &= \mathbb{P}_i\left(bn-\tau(n)+1>0\right) \le \mathbb{P}_i\left(\tau(n)\le bn+1\right)\end{aligned}$$

and we obtain an alternative $\alpha_2>1$ in this case from Lemma 7.14.

The relation suggested earlier is

$$\mathbb{P}_\pi\left(\max_{0\le k\le n} S_k^{*,i} > n\varepsilon\right) \le \mathbb{P}_\pi(S_1^{*,i} > \frac{\varepsilon}{2}n) + \mathbb{P}_i\left(\max_{0\le k\le n} S_k^{*,i} > \frac{\varepsilon}{2}n\right)$$

from the proof of Lemma 7.17. Set $\varepsilon' := \frac{\varepsilon}{2}$. Then we are in the situation that we talked about before, with toll term $\mathbb{P}_\pi(S_1^{*,i} > \frac{\varepsilon}{2}n)$. It is decreasing exponentially if there is a β such that $\mathbb{E}_\pi e^{\beta\log^+ L(^\#\vec{\Psi}_1^i)}<\infty$, because then

$$\left(e^{\beta\frac{\varepsilon}{2}}\right)^n\mathbb{P}_\pi\left(\frac{2}{\varepsilon}(S_1^{*,i})^+ > n\right) \le \mathbb{E}_\pi\left(\mathbb{1}_{\{\frac{2}{\varepsilon}S_1^{*,i}>n\}}e^{\beta(S_1^{*,i})^+}\right)\to 0,$$

as $n\to\infty$, by the dominated convergence theorem. The finiteness of $\mathbb{E}_\pi e^{\beta\log^+ L(^\#\vec{\Psi}_1^i)}$ is guaranteed by Lemma B.8. Set $\alpha_3 := e^{\beta\frac{\varepsilon}{2}}>1$. Then any α from $(1,\alpha_1\wedge\alpha_2\wedge\alpha_3]$ is valid.

Now, let $\mathbb{E}_i|\log L(^\#\vec{\Psi}_1^i)|=\infty$. Pick $\gamma\in(0,1)$ and c such that $-\infty<\mathbb{E}\log(L(^\#\vec{\Psi}_1^i)\vee c)<0$ and $\gamma\in((\gamma_{i,c}^*)^{\frac{1}{m_{ii}}},1)$ which is possible by Lemma B.5. If we set $L_{k,c}^{*,i} := (L(^\#\vec{\Psi}_k^i)\vee c)/\gamma_{i,c}^*$, then we are in the previous situation with all necessary conditions fulfilled and we only need to remark that

$$\mathbb{P}_\pi\left(\prod_{k=1}^{\tau(n)-1} L(^\#\vec{\Psi}_k^i) > \gamma^n\right) \le \mathbb{P}_\pi\left(\prod_{k=1}^{\tau(n)-1}(L(^\#\vec{\Psi}_k^i)\vee c) > \gamma^n\right)$$

and $\mathbb{E}_\pi e^{\beta\log^+(L(^\#\vec{\Psi}_1^i)\vee c)} = \mathbb{E}_\pi e^{\beta\log^+ L(^\#\vec{\Psi}_1^i)}<\infty$. □

As Lemmas 7.19 and 7.22 contribute to Lemmas 7.20, 7.21 and 7.23, the following lemma contains moment results that contribute to Lemma 7.27 in a similar way.

Lemma 7.26. *The following assertions hold under* (B1)-(B4)*:*

(a) $\mathbb{E}_i\sigma_1(i)C_1^\eta < \infty$ *for* $\eta \le \frac{p}{4}$.

(b) There is a $q \in (0,p]$ *such that* $\mathbb{E}_i D_0^\eta = \mathbb{E}_i D_{\tau(n)}^\eta < \infty$ *for all* $\eta \in (0, q \wedge \frac{p}{2}]$.

(c) $\mathbb{E}_i\sigma_1(i)G_1^\eta < \infty$ *for* $\eta \le \frac{p}{2}$.

Proof. Part (a): The Cauchy-Schwarz inequality yields that $\mathbb{E}_i\sigma_1(i)C_1^\eta < \infty$ holds in particular if $\mathbb{E}_i C_1^{2\eta} < \infty$ ($\sigma_1(i)$ has an exponential moment). We recall that

$$C_1^{2\eta} \le \left(d(x_0, {}^{\#}\Psi_1(x_0)) + \sum_{n=2}^{\sigma_1(i)} d({}^{\#}\Psi_1 \cdots {}^{\#}\Psi_{n-1}(x_0), {}^{\#}\Psi_1 \cdots {}^{\#}\Psi_n(x_0)) \right)^{2\eta}$$
$$\le \left((1+G_1) \sum_{n=1}^{\sigma_1(i)} d(x_0, {}^{\#}\Psi_n(x_0)) \right)^{2\eta}$$

by (7.15), and once again we use the Cauchy-Schwarz inequality to get the desired result as long as $\eta \le \frac{p}{4}$ (consider (B1)-(B4) and Remark 6.18).

Part (b): We begin with the reminder that $\mathbb{E}_i L({}^{\#}\vec{\Psi}_1^i)^p < \infty$ implies the existence of a $q \in (0,p]$ such that $\mathbb{E}_i L({}^{\#}\vec{\Psi}_1^i)^\eta < 1$ for all $\eta \in (0,q]$. This was indicated in Remark 6.19 and is an easy consequence of the mean contractivity. From the proof of (a), we can further extract that $\mathbb{E}_i C_1^\eta < \infty$ for $\eta \le \frac{p}{2}$. Now, we consider two cases: If $1 \le \eta \le q \wedge \frac{p}{2}$, then

$$(\mathbb{E}_i D_0^\eta)^{\frac{1}{\eta}} \le \sum_{j\ge1} \left(\mathbb{E}_i \left[\left(\prod_{k=1}^{j-1} L({}^{\#}\vec{\Psi}_k^i) \right) C_j \right]^\eta \right)^{\frac{1}{\eta}} = \sum_{j\ge1} \left(\mathbb{E}_i \left(\prod_{k=1}^{j-1} L({}^{\#}\vec{\Psi}_k^i) \right)^\eta \right)^{\frac{1}{\eta}} \left(\mathbb{E}_i C_j^\eta \right)^{\frac{1}{\eta}}$$
$$= (\mathbb{E}_i C_1^\eta)^{\frac{1}{\eta}} \cdot \sum_{j\ge1} \left[\left(\mathbb{E}_i L({}^{\#}\vec{\Psi}_1^i)^\eta \right)^{\frac{1}{\eta}} \right]^{j-1},$$

where we used the infinite version of the Minkowski inequality and independence. The first factor is finite and the series is a convergent geometric series, so we get the desired result.

If $0 < \eta \le q \wedge \frac{p}{2} \wedge 1$, then we can use subadditivity and more easily get

$$\mathbb{E}_i D_0^\eta \le \sum_{j\ge1} \mathbb{E}_i \left[\left(\prod_{k=1}^{j-1} L({}^{\#}\vec{\Psi}_k^i) \right) C_j \right]^\eta = \sum_{j\ge1} \mathbb{E}_i \left(\prod_{k=1}^{j-1} L({}^{\#}\vec{\Psi}_k^i) \right)^\eta \mathbb{E}_i C_j^\eta$$
$$= \mathbb{E}_i C_1^\eta \cdot \sum_{j\ge1} \left[\mathbb{E}_i L({}^{\#}\vec{\Psi}_1^i)^\eta \right]^{j-1} < \infty.$$

In every case, we find that $\mathbb{E}_i D_0^\eta < \infty$ for all $\eta \in (0, q \wedge \frac{p}{2}]$.

Part (c): Again, $\mathbb{E}_i\sigma_1(i)G_1^\eta < \infty$ holds in particular if $\mathbb{E}_i G_1^{2\eta} < \infty$. This is clearly the case for $\eta \le \frac{p}{2}$, cf. Remark 6.18. □

As the last ingredient for the proof of Theorem 7.3 and as the counterpart of Lemmas 7.20, 7.21 and 7.23, we state the following result:

Lemma 7.27. *If* (B1)-(B4) *hold, then for all* $\gamma \in ((\gamma_i^*)^{\frac{1}{m_{ii}}}, 1)$ *there exists an* $\alpha > 1$*, depending on* γ *and* i*, such that for all* $c > 0$

$$\lim_{n\to\infty} \alpha^n \mathbb{P}_\pi \left(\prod_{k=1}^{\tau(n)-1} L({}^{\#}\vec{\Psi}_k^i) \cdot C_{\tau(n)} > \frac{\gamma^n}{c} \right) = 0, \tag{7.23}$$

$$\lim_{n\to\infty} \alpha^n \mathbb{P}_\pi \left(\prod_{k=1}^{\tau(n)-1} L({}^{\#}\vec{\Psi}_k^i) \cdot G_{\tau(n)} > \frac{\gamma^n}{c} \right) = 0, \tag{7.24}$$

and

$$\lim_{n\to\infty} \alpha^n \mathbb{P}_\pi \left(\prod_{k=1}^{\tau(n)} L({}^{\#}\vec{\Psi}_k^i) \cdot D_{\tau(n)} > \frac{\gamma^n}{c} \right) = 0. \tag{7.25}$$

The same holds with $\mathbb{P}_i$ *instead of* $\mathbb{P}_\pi$ *with possibly different* α*.*

Proof. Fix $n \geq 1$. We start with $\gamma \in ((\gamma_i^*)^{\frac{1}{m_{ii}}}, 1)$ and $\beta \in ((\gamma_i^*)^{\frac{1}{m_{ii}}}, \gamma)$ and decompose

$$\begin{aligned} &\mathbb{P}_\pi \left(\prod_{k=1}^{\tau(n)-1} L({}^{\#}\vec{\Psi}_k^i) \cdot C_{\tau(n)} > \frac{\gamma^n}{c} \right) \\ &\leq \mathbb{P}_\pi \left(C_{\tau(n)} > \frac{1}{c} \left(\frac{\gamma}{\beta}\right)^n \right) + \mathbb{P}_\pi \left(\prod_{k=1}^{\tau(n)-1} L({}^{\#}\vec{\Psi}_k^i) > \beta^n \right). \end{aligned} \tag{7.26}$$

Similar estimates hold for $G_{\tau(n)}$ and $D_{\tau(n)}$. Lemma B.8 yields the existence of $t > 0$ with $\mathbb{E}_\pi e^{t\sigma_1(i)} < \infty$. Choosing $\eta \leq \frac{p}{4} \wedge \frac{t}{\log \frac{\gamma}{\beta}}$, Lemma 7.15 gives

$$\begin{aligned} &\mathbb{P}_\pi \left(C_{\tau(n)} > \frac{1}{c} \left(\frac{\gamma}{\beta}\right)^n \right) \leq \mathbb{P}_\pi \left(c^\eta C_{\tau(n)}^\eta > \left(\frac{\gamma}{\beta}\right)^{\eta n} \right) \\ &\leq \mathbb{E}_i \sigma_1(i) \mathbb{1}_{\{c^\eta C_1^\eta > (\frac{\gamma}{\beta})^{\eta n}\}} + \mathbb{P}_\pi \left(\sigma_1(i) > n-1, c^\eta C_1^\eta > (\frac{\gamma}{\beta})^{\eta n} \right) \\ &\leq \mathbb{E}_i \sigma_1(i) \mathbb{1}_{\{c^\eta C_1^\eta > (\frac{\gamma}{\beta})^{\eta n}\}} + \mathbb{P}_\pi \left(\sigma_1(i) \geq n \right) \end{aligned}$$

and Lemma 7.26 yields

$$\begin{aligned} \left[\left(\frac{\gamma}{\beta}\right)^\eta\right]^n \mathbb{P}_\pi \left(C_{\tau(n)} > \frac{1}{c} \left(\frac{\gamma}{\beta}\right)^n \right) &\leq \left[\left(\frac{\gamma}{\beta}\right)^\eta\right]^n \mathbb{E}_i \sigma_1(i) \mathbb{1}_{\{c^\eta C_1^\eta > (\frac{\gamma}{\beta})^{\eta n}\}} + \left[\left(\frac{\gamma}{\beta}\right)^\eta\right]^n \mathbb{P}_\pi(\sigma_1(i) \geq n) \\ &\leq c^\eta \mathbb{E}_i \sigma_1(i) C_1^\eta \mathbb{1}_{\{c^\eta C_1^\eta > (\frac{\gamma}{\beta})^{\eta n}\}} + \mathbb{E}_\pi e^{(\eta \log \frac{\gamma}{\beta})\sigma_1(i)} \mathbb{1}_{\{\sigma_1(i) \geq n\}} \to 0, \end{aligned}$$

as $n \to \infty$. In particular, we have

$$\alpha_1^n \mathbb{P}_\pi \left(C_{\tau(n)} > \frac{1}{c} \left(\frac{\gamma}{\beta}\right)^n \right) \to 0$$

for $\alpha_1 \in (1, (\frac{\gamma}{\beta})^\eta]$. We still need to deal with the second summand from (7.26) but this was done in Lemma 7.25 for both cases $\mathbb{E}_i |\log L({}^{\#}\vec{\Psi}_1^i)| < \infty$ and $= \infty$. More precisely, it provides the existence of $\alpha_2 > 1$ such that

$$\lim_{n\to\infty} \alpha_2^n \mathbb{P}_\pi \left(\prod_{k=1}^{\tau(n)-1} L({}^{\#}\vec{\Psi}_k^i) > \gamma^n \right) = 0.$$

If we choose α from $(1, \alpha_1 \wedge \alpha_2]$, then we arrive at the desired conclusion for $C_{\tau(n)}$.

Concerning (7.24), we simply substitute $C_{\tau(n)}$ with $G_{\tau(n)}$ in the first estimate and use the corresponding result from Lemma 7.26 in the same way as before (keeping in mind that the restrictions on η by Lemma 7.26 are different).

We deal with the remaining (7.25) alike, but there it is easiest to use the integrability of D_0^η by Lemma 7.26 and

$$\mathbb{P}_\pi\left(D_{\tau(n)} > \frac{1}{c}\left(\frac{\gamma}{\beta}\right)^n\right) = \mathbb{P}_i\left(D_0 > \frac{1}{c}\left(\frac{\gamma}{\beta}\right)^n\right)$$

for $n \geq 1$. Ultimately, each procedure provides us with a respective α, so we just consider their minimum which obviously serves as the desired exponential rate for (7.23)-(7.25). Under $\mathbb{P}_i$, the proof is the same. □

Proof of Theorem 7.3(a). With the notation from the proof of Theorem 7.2(a) and in the setting when $\mathbb{E}_i|\log L(^\#\vec{\Psi}_1^i)| < \infty$, we have the estimate

$$\begin{aligned}
&\mathbb{P}_i(d(^\#\widehat{X}_n^x, {}^\#\widehat{X}_\infty) > \gamma^n) \\
&\leq \mathbb{P}_i\left(\prod_{k=1}^{\tau(n)-1} L(^\#\vec{\Psi}_k^i)\cdot C_{\tau(n)} > \frac{\gamma^n}{3}\right) + \mathbb{P}_i\left(\prod_{k=1}^{\tau(n)} L(^\#\vec{\Psi}_k^i)\cdot D_{\tau(n)} > \frac{\gamma^n}{3}\right) \\
&\quad + \mathbb{P}_i\left(\prod_{k=1}^{\tau(n)-1} L(^\#\vec{\Psi}_k^i)\cdot G_{\tau(n)} > \frac{((\gamma_i^*)^a)^n}{3}\right) + \mathbb{P}_i\left(d(x,x_0) > (1/\gamma_i^*)^{(a-b)n}\right).
\end{aligned} \tag{7.27}$$

Lemma 7.27 deals with the first three terms. The fourth is deterministic and vanishes for n large enough. If $\mathbb{E}_i|\log L(^\#\vec{\Psi}_1^i)| = \infty$, then we proceed as in the proof of Theorem 7.2(a). □

Proof of Theorem 7.3(b). We can copy the proof of Theorem 7.3(a) while replacing $\mathbb{P}_i$ by $\mathbb{P}_\pi$. The required lemmas always provide a statement in this setting. □

Proof of Theorem 7.3(c). Fix $n \geq 1$. Choose $\gamma \in ((\gamma_i^*)^{\frac{1}{m_{ii}}}, 1)$ and C big enough for the first three summands of the right side of (7.27) (with $\mathbb{P}_\pi$ instead of $\mathbb{P}_i$) to be $\leq \frac{C}{3}\alpha^{-n}$, $\alpha > 1$. Let further $\varepsilon > 0$ be so small that $\frac{1}{\alpha} + \varepsilon \in (0,1)$, and with this define $A_1 := \max\{C, 1+\varepsilon\}$ and $r := \max\{\gamma, \frac{1}{\alpha} + \varepsilon\} \in (0,1)$. Hence, we get

$$\mathbb{P}_\pi(d(^\#\widehat{X}_n^x, {}^\#\widehat{X}_\infty) \geq A_1 r^n) \leq \mathbb{P}_\pi(d(^\#\widehat{X}_n^x, {}^\#\widehat{X}_\infty) > \gamma^n) \leq C\left(\frac{1}{\alpha}\right)^n + \mathbb{1}_{\{n<n_0(x)\}}$$

with $n_0(x)$ from the proof of Theorem 7.2(a). We estimate

$$\begin{aligned}
C\left(\frac{1}{\alpha}\right)^n + \mathbb{1}_{\{n<n_0(x)\}} < A_1 r^n + \mathbb{1}_{\{n<n_0(x)\}} &= r^n\left(A_1 + \left(\frac{1}{r}\right)^n \mathbb{1}_{n<n_0(x)}\right) \\
&\leq r^n\left(A_1 + \left(\frac{1}{r}\right)^{n_0(x)}\right) =: r^n(A_1 + d(x,x_0)^{A_2}) = A_x r^n
\end{aligned}$$

for some positive constant A_2. This yields $d_{\mathrm{Pr}}(\mathbb{P}_\pi^{^\#\widehat{X}_n^x}, \mathbb{P}_\pi^{^\#\widehat{X}_\infty}) \leq A_x r^n$ by Lemma B.4 and

$$d_{\mathrm{Pr}}(\mathbb{P}_\pi^{^\#\widehat{X}_n^x}, \mathbb{P}_\pi^{^\#\widehat{X}_\infty}) = d_{\mathrm{Pr}}(\mathbb{P}_\pi^{X_n^x}, \mathbb{P}_\pi^{^\#\widehat{X}_\infty})$$

because $\mathbb{P}_\pi^{^\#\widehat{X}_n^x} = \mathbb{P}_\pi^{X_n^x}$. □

Proof of Theorem 7.3(d). Denote by $\widehat{\pi}_i$ the distribution of ${}^{\#}\widehat{X}_\infty$ under $\mathbb{P}_i$. We know from Lemma 7.5 that ${}^{\#}\widehat{X}_\infty$ is the $\mathbb{P}_i$-a.s. limit of the backward iterations $\widehat{Y}_n^x = {}^{\#}\widehat{X}^x_{\sigma_n(i)}$ of the IFS of i.i.d. Lipschitz maps (Y_n^x). Hence, we can cite from [AF01] that

$$\mathbb{E}_i d(x_0, {}^{\#}\widehat{X}_\infty)^\eta = \int_{\mathbb{X}} d(x_0, x)^\eta \; \widehat{\pi}_i(\mathrm{d}x) < \infty$$

whenever the corresponding Lipschitz maps satisfy $\mathbb{E}_i L({}^{\#}\vec{\Psi}_1^i)^\eta < \infty$ and $\mathbb{E}_i d(x_0, {}^{\#}\vec{\Psi}_1^i(x_0))^\eta < \infty$ for some (and thus all) $x_0 \in \mathbb{X}$. This is clearly the case under (B1)-(B4). Note for the latter that $d(x_0, {}^{\#}\vec{\Psi}_1^i(x_0)) \leq D_0$ and use Lemma 7.26 as below. □

Alternative proof of Theorem 7.3(d). From Lemma 7.7 we know that

$$d(x_0, {}^{\#}\widehat{X}_\infty) \leq D_0 \quad \mathbb{P}_i\text{-a.s.}$$

which to the power of $\eta \in (0, q \wedge \frac{p}{2}]$ is $\mathbb{P}_i$-integrable by Lemma 7.26 for some $q \in (0, p]$. □

Proof of Theorem 7.3(e). We know from Lemma 7.13 that even

$$d(x_0, {}^{\#}\widehat{X}_\infty) \leq D_0 \quad \mathbb{P}_\pi\text{-a.s.},$$

hence it suffices to verify $\mathbb{E}_\pi D_0^\eta < \infty$ for some $\eta > 0$. To finish the proof, we recall

$$D_0 = d({}^{\#}\vec{\Psi}_1^i(x_0), x_0) + L({}^{\#}\vec{\Psi}_1^i) D_1$$

from Lemma 7.6. This gives us the estimate

$$\begin{aligned} \mathbb{E}_\pi D_0^\eta &\leq C \left(\mathbb{E}_\pi d({}^{\#}\vec{\Psi}_1^i(x_0), x_0)^\eta + \mathbb{E}_\pi L({}^{\#}\vec{\Psi}_1^i)^\eta D_1^\eta \right) \\ &= C\big(\mathbb{E}_\pi d({}^{\#}\vec{\Psi}_1^i(x_0), x_0)^\eta + \mathbb{E}_\pi L({}^{\#}\vec{\Psi}_1^i)^\eta \; \mathbb{E}_i D_0^\eta\big) \end{aligned}$$

for some $C > 0$. As seen in (d), $\mathbb{E}_i D_0^\eta < \infty$ for $\eta \in (0, q \wedge \frac{p}{2}]$ and some $q \in (0, p]$. The remaining two terms are finite by Lemma B.8 (choosing $\eta \leq \tau$). Actually, (B2*) is required here for technical reasons. □

B. Appendix

B.1. Kingman's subadditive ergodic theorem

The following theorem is [Lig85, Theorem 1.10] or [Als15a, Theorem A.5].

Theorem B.1. *Let* $(Y_{k,n})_{n\geq 1}^{0\leq k\leq n}$ *be a family of real-valued random variables which satisfies the following conditions:*

(SA-1) $Y_{0,n} \leq Y_{0,k} + Y_{k,n}$ *a.s. for all* $0 \leq k < n$.

(SA-2) $(Y_{nk,(n+1)k})_{n\geq 1}$ *is a stationary sequence for each* $k \geq 1$.

(SA-3) The distribution of $(Y_{k,k+n})_{n\geq 1}$ *does not depend on* $k \geq 0$.

(SA-4) $\mathbb{E}Y_{0,1}^{+} < \infty$ *and* $\mu := \inf_{n\geq 1} n^{-1}\mathbb{E}Y_{0,n} > -\infty$.

Then

(a) $\lim_{n\to\infty} n^{-1}\mathbb{E}Y_{0,n} = \mu$.

(b) $n^{-1}Y_{0,n}$ *converges a.s. and in* L^1 *to a random variable* χ *with mean* μ.

(c) If all stationary sequences in (SA-2) are ergodic, then $\chi = \mu$ *a.s.*

(d) If $\mu = -\infty$ *in (SA-4), then* $n^{-1}Y_{0,n} \to -\infty$ *a.s.*

B.2. Prokhorov metric

We follow [DF99] and [AF01] for the definition of the Prokhorov metric.

Definition B.2. Let λ_1 and λ_2 be two probability measures on $\mathbb{X}$ and let

$$B^\delta := \{x \in \mathbb{X} : d(x,y) < \delta \text{ for some } y \in B\}$$

be the δ-neighborhood of $B \in \mathfrak{B}(\mathbb{X})$. We set

$$d_{\mathrm{Pr}}(\lambda_1, \lambda_2) := \inf\{\delta \geq 0 : \lambda_1(B) < \lambda_2(B^\delta) + \delta \text{ and } \lambda_2(B) < \lambda_1(B^\delta) + \delta \text{ for all } B \in \mathfrak{B}(\mathbb{X})\}.$$

Then d_{Pr} is a metric on the space of probability measures on $\mathbb{X}$.

We state two well-known facts about the Prokhorov metric:

Lemma B.3. *If* $(\mathbb{X}, d)$ *is separable, then convergence of probability measures in* d_{Pr} *is equivalent to distributional convergence of the corresponding random variables.*

Lemma B.4. *Let* X_1, X_2 *be two* $\mathbb{X}$*-valued random elements with distributions* λ_1, λ_2 *w.r.t.* $\mathbb{P}$. *Then* $\mathbb{P}(d(X_1, X_2) \geq \delta) < \delta$ *implies* $d_{\mathrm{Pr}}(\lambda_1, \lambda_2) \leq \delta$.

Proof. Cf. [DF99, Lemma 5.8]. □

B.3. Some results from the i.i.d. case

In this Section, we slightly improve the statement of [AF01, Theorem 2.2 (c)] or [Als15a, Theorem 3.24 (b)], providing the actual lower bound for the rate of exponential convergence. Therefore, we mix notation from both works: Let $(X_n)_{n\geq 0}$ be an IFS of i.i.d. Lipschitz maps $(\Psi_n)_{n\geq 1}$ with corresponding Lipschitz constants $(L_n)_{n\geq 1}$ satisfying the mean contraction condition $\mathbb{E}\log L_1 < 0$ and the jump-size condition $\mathbb{E}\log^+ d(\Psi_1(x_0), x_0) < \infty$ for some $x_0 \in \mathbb{X}$. Let further $\sigma_1(\gamma) := \inf\{n \geq 1 : \sum_{k=1}^n \log L_k \leq \log\gamma\}$, $\gamma \in (0,1)$, be a corresponding first passage time, $\mu(\gamma) := \mathbb{E}\sigma_1(\gamma)$ its mean, and $\log\gamma^* := \inf_{\gamma\in(0,1)} \frac{\log\gamma}{\mu(\gamma)}$. Lemma B.5 provides some more insight, and we refer to [AF01, Section 2] for more details.

Lemma B.5. *The following assertions hold:*

(a) If $\mathbb{E}|\log L_1| < \infty$, then $\log\gamma^* = \mathbb{E}\log L_1 < 0$ or $\gamma^* = e^{\mathbb{E}\log L_1} \in (0,1)$, equivalently.

(b) If $\mathbb{E}|\log L_1| = \infty$, then $\gamma^* = 0$.

(c) If $\mathbb{E}|\log L_1| = \infty$, then there exists a $c' \in (0,1)$ such that $-\infty < \mathbb{E}\log(L_1 \vee c) < 0$ for all $c \in (0, c']$.

(d) If $\mathbb{E}|\log L_1| = \infty$ and γ_c^* corresponds to the truncated Lipschitz constants $L_k \vee c$, $c \in (0,1]$, then $\gamma_c^* \downarrow \gamma^* = 0$ for $c \downarrow 0$.

Proof. Part (a) was proven in [AF01]. For Part (b), we remark that our mean contraction assumption implies $\mathbb{E}\log L_1 = -\infty$. By definition of γ^*, we get the estimate

$$\log\gamma^* = \inf_{\gamma\in(0,1)} \frac{\log\gamma}{\mu(\gamma)} \leq \lim_{\gamma\downarrow 0} \frac{\log\gamma}{\mu(\gamma)} = \lim_{\gamma\downarrow 0} \left(\frac{\mu(\gamma)}{\log\gamma}\right)^{-1} = -\infty$$

with an appeal to [Gut09, Theorem 3.6.1 and Remark 3.6.1]. This means $\gamma^* = 0$. For Part (c), we remark that $\mathbb{E}\log L_1 = -\infty$ as before, and $\mathbb{E}\log^+ L_1 < \infty$. It is obvious that $\mathbb{E}\log(L_1 \vee c') \geq \log c' > -\infty$ for every $c' \in (0,1]$. It remains to show that

$$\lim_{c\downarrow 0,\ c\in(0,1]} \mathbb{E}\log(L_1 \vee c) = -\infty$$

which is indeed true since $\log(L_1 \vee c)$ is decreasing, $\mathbb{E}\log(L_1 \vee 1) = \mathbb{E}\log^+ L_1 < \infty$ and the limit is $\mathbb{E}\log L_1 = -\infty$. Part (d) follows from the proof of (c) since the truncated Lipschitz constants satisfy (a) and thus

$$\lim_{c\downarrow 0} \gamma_c^* = e^{\lim_{c\downarrow 0} \mathbb{E}\log(L_1\vee c)} = 0 = \gamma^*.$$

□

The next lemma constitutes the afore-mentioned slight improvement. As it does in our work, $\widehat{X}_n^x$ denotes the backward iteration of X_n with $X_0 = x$. $\widehat{X}_\infty$ is the a.s. limit of $\widehat{X}_n^x$.

Lemma B.6. *For every $x \in \mathbb{X}$ and $\beta \in (\gamma^*, 1)$, we have*

$$\lim_{n\to\infty} \beta^{-n} d(\widehat{X}_n^x, \widehat{X}_\infty) = 0 \quad \mathbb{P}\text{-}a.s.$$

Proof. First of all, we note that $\sigma_1(\gamma)$ is an a.s. finite first passage time and its mean $\mu(\gamma)$ is finite and positive for all γ. Let $\beta \in (\gamma^*, 1)$. By definition, there exists a $\gamma \in (0,1)$ such that

$$\log \gamma^* < \frac{\log \gamma}{\mu(\gamma)} < \log \beta$$

and a $b \in (0, \frac{1}{\mu(\gamma)})$ such that $\beta = \gamma^b = e^{b \log \gamma}$. Hence, there exists an $\varepsilon' > 0$ such that $0 < b < \frac{1}{\mu(\gamma)} - \varepsilon'$. In particular, this applies to the case $\mathbb{E} \log L_1 = -\infty$, i.e. $\gamma^* = 0$. The essential estimate for the proof is the following from [Als15a, Lemma 3.28]:

$$\beta^{-n} d(\widehat{X}_n^x, \widehat{X}_\infty) \leq \beta^{-n} e^{\log \gamma (\tau(n)-1)} C_{\tau(n)}^x + ((\beta^{-1})^{\frac{n}{\tau(n)}})^{\tau(n)} d(\widehat{Y}_{\tau(n)}^x, \widehat{X}_\infty). \tag{B.1}$$

The occurring (Y_n^x) (dependent on γ) is the strongly contractive IFS of i.i.d. Lipschitz maps with $\log L(\phi_n^{\leftarrow}) \leq \log \gamma$ from [Als15a, Lemma 3.26] which satisfies the jump-size condition. Also, C_n^x is C_n with $X_0 = x$ from [Als15a, (3.25)]. [Als15a, Proposition 3.18 (b)] gives us

$$\lim_{n \to \infty} \hat{\gamma}^{\tau(n)} d(\widehat{Y}_{\tau(n)}^x, \widehat{X}_\infty) = 0 \quad \mathbb{P}\text{-a.s.}$$

for every $\hat{\gamma} \in (1, e^{-\log \gamma})$ since $\tau(n) \to \infty$ $\mathbb{P}$-a.s. In order to apply this result, we try to find a valid $\hat{\gamma}$ such that for n big enough

$$(\beta^{-1})^{\frac{n}{\tau(n)}} \leq \hat{\gamma} \in (1, e^{-\log \gamma}).$$

It is known that $\frac{n}{\tau(n)} \to (\frac{1}{\mu(\gamma)})^{-1} = \mu(\gamma) > 0$ $\mathbb{P}$-a.s., as $n \to \infty$. Choose $\varepsilon \in (0, \mu(\gamma)\varepsilon'(\frac{1}{\mu(\gamma)} - \varepsilon')^{-1})$. Then $\frac{n}{\tau(n)} \leq \mu(\gamma) + \varepsilon$ for n big enough, and hence

$$\begin{aligned}(\beta^{-1})^{\frac{n}{\tau(n)}} &= e^{-b \log \gamma \frac{n}{\tau(n)}} \leq e^{-\log(\gamma) b(\mu(\gamma)+\varepsilon)} < e^{-\log \gamma (1 + \frac{\varepsilon}{\mu(\gamma)} - \varepsilon' \mu(\gamma) - \varepsilon' \varepsilon)} \\ &= e^{-\log \gamma (1 + \varepsilon(\frac{1}{\mu(\gamma)} - \varepsilon') - \varepsilon' \mu(\gamma))} =: \hat{\gamma} < e^{-\log \gamma}\end{aligned}$$

because $\varepsilon(\frac{1}{\mu(\gamma)} - \varepsilon') < \varepsilon' \mu(\gamma)$. Hence, the second term in (B.1) vanishes.

It remains to verify the existence of $\varepsilon'' > 0$ such that for n big enough

$$\beta^{-n} e^{\log \gamma (\tau(n)-1)} < e^{-\varepsilon'' \tau(n)},$$

since $e^{-\varepsilon'' \tau(n)} C_{\tau(n)}^x \to 0$ $\mathbb{P}$-a.s., as $n \to \infty$, by [Als15a, Lemma 3.27]. Note that

$$1 - \frac{1}{\tau(n)} - \frac{bn}{\tau(n)} \to 1 - b\mu(\gamma) =: a_\gamma \in (0,1) \quad \mathbb{P}\text{-a.s.},$$

so choosing n_0 big enough, such that additionally $1 - \frac{1}{\tau(n)} - \frac{bn}{\tau(n)} > \frac{a_\gamma}{2} > 0$, implies

$$\begin{aligned}\beta^{-n} e^{\log \gamma (\tau(n)-1)} &= e^{-bn \log \gamma + \log \gamma (\tau(n)-1)} = e^{\log \gamma (1 - \frac{1}{\tau(n)} - \frac{bn}{\tau(n)}) \tau(n)} \\ &< e^{\log \gamma \frac{a_\gamma}{2} \tau(n)} =: e^{-\varepsilon'' \tau(n)}\end{aligned}$$

since $\log \gamma < 0$. So the first term in (B.1) vanishes and the proof is complete. □

B.4. Moment results in Situations (1) and (2)

To deal with conditions like $\mathbb{E}_\pi \log_*^p L({}^{\#}\vec{\Psi}_1^i) < \infty$ under (A1)-(A3) and with similar conditions under (B1)-(B4), we give a formula to compute those expressions in terms of expectations w.r.t. $\mathbb{P}_i$.

Lemma B.7. *Let $g(({}^{\#}\Psi_n)_{n\geq 1}, \sigma_1(i))$ be some measurable non-negative function of the dual Lipschitz maps and of $\sigma_1(i)$. Then*

$$\mathbb{E}_\pi g(({}^{\#}\Psi_n)_{n\geq 1}, \sigma_1(i)) = \frac{1}{m_{ii}} \mathbb{E}_i \left(\sum_{k=0}^{\sigma_1(i)-1} g(({}^{\#}\Psi_{k+n})_{n\geq 1}, \sigma_1(i) - k) \right).$$

Proof. Define $\sigma_1(i) \circ {}^{\#}\theta_k := \inf\{n \geq 1 : {}^{\#}\theta_{k+n} = i\}$. We use the well-known identity

$$\int f \, \mathrm{d}\pi = \frac{1}{m_{ii}} \mathbb{E}_i \left(\sum_{k=0}^{\sigma_1(i)-1} f({}^{\#}\theta_k) \right).$$

With $f : j \mapsto \mathbb{E}_j g(({}^{\#}\Psi_n)_{n\geq 1}, \sigma_1(i))$, we calculate

$$\begin{aligned}
&\mathbb{E}_\pi g(({}^{\#}\Psi_n)_{n\geq 1}, \sigma_1(i)) = \int f(j) \, \pi(\mathrm{d}j) = \frac{1}{m_{ii}} \mathbb{E}_i \left(\sum_{k=0}^{\sigma_1(i)-1} f({}^{\#}\theta_k) \right) \\
&= \frac{1}{m_{ii}} \mathbb{E}_i \left(\sum_{k=0}^{\sigma_1(i)-1} \mathbb{E}_{{}^{\#}\theta_k} g(({}^{\#}\Psi_n)_{n\geq 1}, \sigma_1(i)) \right) = \frac{1}{m_{ii}} \sum_{k\geq 0} \mathbb{E}_i \left(\mathbb{1}_{\{\sigma_1(i) > k\}} \mathbb{E}_{{}^{\#}\theta_k} g(({}^{\#}\Psi_n)_{n\geq 1}, \sigma_1(i)) \right) \\
&= \frac{1}{m_{ii}} \sum_{k\geq 0} \mathbb{E}_i \left(\mathbb{1}_{\{\sigma_1(i) > k\}} \mathbb{E}_i \left(g(({}^{\#}\Psi_{k+n})_{n\geq 1}, \sigma_1(i) \circ {}^{\#}\theta_k) | {}^{\#}\theta_0, \ldots, {}^{\#}\theta_k, {}^{\#}\Psi_0, \ldots, {}^{\#}\Psi_k \right) \right) \\
&= \frac{1}{m_{ii}} \sum_{k\geq 0} \mathbb{E}_i \left(\mathbb{1}_{\{\sigma_1(i) > k\}} g(({}^{\#}\Psi_{k+n})_{n\geq 1}, \sigma_1(i) \circ {}^{\#}\theta_k) \right) \\
&= \frac{1}{m_{ii}} \sum_{k\geq 0} \mathbb{E}_i \left(\mathbb{1}_{\{\sigma_1(i) > k\}} g(({}^{\#}\Psi_{k+n})_{n\geq 1}, \sigma_1(i) - k) \right) \\
&= \frac{1}{m_{ii}} \mathbb{E}_i \left(\sum_{k=0}^{\sigma_1(i)-1} g(({}^{\#}\Psi_{k+n})_{n\geq 1}, \sigma_1(i) - k) \right).
\end{aligned}$$

□

Lemma B.8. *Under* (A1)-(A3)*, the following assertions hold:*

(a) $\mathbb{E}_\pi \sigma_1(i)^p < \infty$.

(b) $\mathbb{E}_\pi \log_*^p L({}^{\#}\vec{\Psi}_1^i) < \infty$.

(c) $\mathbb{E}_\pi \log_*^p d({}^{\#}\vec{\Psi}_1^i(x_0), x_0) < \infty$.

Under (B1)-(B4)*, the following assertions hold:*

(d) $\mathbb{E}_\pi e^{t\sigma_1(i)} < \infty$ *for some* $t > 0$.

(e) $\mathbb{E}_\pi e^{\alpha \log^+ L({}^{\#}\vec{\Psi}_1^i)} < \infty$ *for some* $\alpha > 0$ *or equivalently* $\mathbb{E}_\pi L({}^{\#}\vec{\Psi}_1^i)^\tau < \infty$ *for some* $\tau > 0$.

Under (B1)*,* (B2*) *and* (B4)*, the following assertion holds:*

(f) $\mathbb{E}_\pi d({}^\#\vec{\Psi}_1^i(x_0), x_0)^\tau < \infty$ *for some* $\tau > 0$*.*

Proof. We apply Lemma B.7. For (a), we get

$$\mathbb{E}_\pi \sigma_1(i)^p = \frac{1}{m_{ii}}\mathbb{E}_i\left(\sum_{k=0}^{\sigma_1(i)-1}(\sigma_1(i)-k)^p\right) = \frac{1}{m_{ii}}\mathbb{E}_i\left(\sum_{k=1}^{\sigma_1(i)}k^p\right) \le \frac{1}{m_{ii}}\mathbb{E}_i\sigma_1(i)^{p+1} < \infty$$

under (A1). For (d), we estimate

$$\mathbb{E}_\pi e^{t\sigma_1(i)} = \frac{1}{m_{ii}}\mathbb{E}_i\left(\sum_{k=0}^{\sigma_1(i)-1}e^{t(\sigma_1(i)-k)}\right) = \frac{1}{m_{ii}}\mathbb{E}_i\left(\sum_{k=1}^{\sigma_1(i)}e^{tk}\right) \le \frac{1}{m_{ii}}\mathbb{E}_i\left(\sigma_1(i)e^{t\sigma_1(i)}\right) < \infty$$

for some $t < \beta_i^*$ due to Hölder's inequality, because $\mathbb{E}_i e^{\beta_i^*\sigma_1(i)} < \infty$ under (B1). We prepare (b) and (e) with the definition

$$\bar{G}_1 := \max\{L({}^\#\Psi_1\cdots{}^\#\Psi_{\sigma_1(i)}),\ldots,L({}^\#\Psi_{\sigma_1(i)})\} \tag{B.2}$$

which is closely related to G_1. We obtain

$$\begin{aligned}\mathbb{E}_\pi \log_*^p L({}^\#\vec{\Psi}_1^i) &= \frac{1}{m_{ii}}\mathbb{E}_i\left(\sum_{k=0}^{\sigma_1(i)-1}\log_*^p L({}^\#\Psi_{k+1}\cdots{}^\#\Psi_{\sigma_1(i)})\right) \le \frac{1}{m_{ii}}\mathbb{E}_i\left(\sigma_1(i)\log_*^p\bar{G}_1\right)\\ &\le \frac{1}{m_{ii}}\left(\mathbb{E}_i\sigma_1(i)^{p+1}\right)^{1/p+1}\left(\mathbb{E}_i\log_*^{p+1}\bar{G}_1\right)^{p/p+1} < \infty\end{aligned}$$

since similar to Lemma 7.19 we find that $1+\bar{G}_1 \le \prod_{n=1}^{\sigma_1(i)}(1+L({}^\#\Psi_n))$ and thus

$$\log_*^{p+1}\bar{G}_1 \le \left(\sum_{n=1}^{\sigma_1(i)}\log_* L({}^\#\Psi_n)\right)^{p+1} < \infty.$$

To prove (e), we use the Cauchy-Schwarz inequality to obtain

$$\begin{aligned}\mathbb{E}_\pi e^{\alpha\log^+ L({}^\#\vec{\Psi}_1^i)} &\le \frac{1}{m_{ii}}\mathbb{E}_i\left(\sum_{k=0}^{\sigma_1(i)-1}e^{\alpha\log_* L({}^\#\Psi_{k+1}\cdots{}^\#\Psi_{\sigma_1(i)})}\right) \le \frac{1}{m_{ii}}\mathbb{E}_i\left(\sigma_1(i)e^{\alpha\log_*\bar{G}_1}\right)\\ &= \frac{1}{m_{ii}}\mathbb{E}_i\sigma_1(i)(1+\bar{G}_1)^\alpha < \infty\end{aligned}$$

under (B1)-(B4) for $\alpha \le \frac{p}{2}$ ($\sigma_1(i)$ has an exponential moment w.r.t. $\mathbb{P}_i$). It is easily seen that the first and the second part of (e) are in fact equivalent with $\tau = \alpha$.

We still need to show (c) and (f). We write

$$\mathbb{E}_\pi\log_*^p d({}^\#\vec{\Psi}_1^i(x_0), x_0) = \frac{1}{m_{ii}}\mathbb{E}_i\left(\sum_{k=0}^{\sigma_1(i)-1}\log_*^p d(x_0, {}^\#\Psi_{k+1}\cdots{}^\#\Psi_{\sigma_1(i)}(x_0))\right).$$

The expression inside the sum is smaller than or equal to

$$\log_*^p\left(d(x_0,{}^\#\Psi_{k+1}(x_0)) + \sum_{n=2}^{\sigma_1(i)-k}d({}^\#\Psi_{k+1}\cdots{}^\#\Psi_{k+n-1}(x_0), {}^\#\Psi_{k+1}\cdots{}^\#\Psi_{k+n}(x_0))\right)$$

$$\leq \log_*^p \left(d(x_0, {}^{\#}\Psi_{k+1}(x_0)) + \sum_{n=2}^{\sigma_1(i)-k} \prod_{l=k+1}^{k+n-1} L({}^{\#}\Psi_l)\ d(x_0, {}^{\#}\Psi_{k+n}(x_0)) \right)$$
$$= \log_*^p \left(\sum_{n=1}^{\sigma_1(i)-k} \prod_{l=k+1}^{k+n-1} L({}^{\#}\Psi_l)\ d(x_0, {}^{\#}\Psi_{k+n}(x_0)) \right)$$
$$\leq \log_*^p \left(\sum_{n=1}^{\sigma_1(i)-k} e^{S^*_{\sigma_1(i)}}\ d(x_0, {}^{\#}\Psi_{k+n}(x_0)) \right) \leq \log_*^p \left(e^{S^*_{\sigma_1(i)}} \sum_{n=1}^{\sigma_1(i)} d(x_0, {}^{\#}\Psi_n(x_0)) \right),$$

with $\mathcal{S}_n^*$ from the proof of Lemma 6.12, hence we conclude

$$\mathbb{E}_\pi \log_*^p d({}^{\#}\vec{\Psi}_1^i(x_0), x_0) \leq \frac{1}{m_{ii}} \mathbb{E}_i \left(\sum_{k=0}^{\sigma_1(i)-1} \log_*^p \left(e^{S^*_{\sigma_1(i)}} \sum_{n=1}^{\sigma_1(i)} d(x_0, {}^{\#}\Psi_n(x_0)) \right) \right)$$
$$\leq \frac{C}{m_{ii}} \mathbb{E}_i \left(\sum_{k=0}^{\sigma_1(i)-1} (1 + \log^+(e^{S^*_{\sigma_1(i)}}))^p \right) + \frac{C}{m_{ii}} \mathbb{E}_i \left(\sum_{k=0}^{\sigma_1(i)-1} \left(\sum_{n=1}^{\sigma_1(i)} \log_* d(x_0, {}^{\#}\Psi_n(x_0)) \right)^p \right)$$

with some constant $C > 0$. All occurring terms are finite under (A1)-(A3):

$$\mathbb{E}_i \left(\sum_{k=0}^{\sigma_1(i)-1} (1 + \log^+(e^{S^*_{\sigma_1(i)}}))^p \right) \leq C\mathbb{E}_i \sigma_1(i) + C\mathbb{E}_i \sigma_1(i) (S^*_{\sigma_1(i)})^p < \infty$$

by Hölder's inequality. Furthermore, the finiteness of

$$\mathbb{E}_i \left(\sum_{k=0}^{\sigma_1(i)-1} \left(\sum_{n=1}^{\sigma_1(i)} \log_* d(x_0, {}^{\#}\Psi_n(x_0)) \right)^p \right) = \mathbb{E}_i \sigma_1(i) \left(\sum_{n=1}^{\sigma_1(i)} \log_* d(x_0, {}^{\#}\Psi_n(x_0)) \right)^p$$

follows from (A1) and (A3) by Hölder's inequality. The proof of (f) is similar:

$$\mathbb{E}_\pi d({}^{\#}\vec{\Psi}_1^i(x_0), x_0)^\tau = \frac{1}{m_{ii}} \mathbb{E}_i \left(\sum_{k=0}^{\sigma_1(i)-1} d(x_0, {}^{\#}\Psi_{k+1} \cdots {}^{\#}\Psi_{\sigma_1(i)}(x_0))^\tau \right)$$

and, as above, the part inside the sum satisfies

$$d(x_0, {}^{\#}\Psi_{k+1} \cdots {}^{\#}\Psi_{\sigma_1(i)}(x_0))^\tau \leq \left(e^{S^*_{\sigma_1(i)}} \sum_{n=1}^{\sigma_1(i)} d(x_0, {}^{\#}\Psi_n(x_0)) \right)^\tau$$

which is why

$$\mathbb{E}_i \left(\sum_{k=0}^{\sigma_1(i)-1} d(x_0, {}^{\#}\Psi_{k+1} \cdots {}^{\#}\Psi_{\sigma_1(i)}(x_0))^\tau \right) \leq \mathbb{E}_i \left(e^{\tau S^*_{\sigma_1(i)}} \sigma_1(i) \left(\sum_{n=1}^{\sigma_1(i)} d(x_0, {}^{\#}\Psi_n(x_0)) \right)^\tau \right) < \infty$$

under (B1), (B2*) and (B4). Note that the first and the third factor have finite expectation with $\tau \leq p$, and $\sigma_1(i)$ has an exponential moment. □

References

[Als94] G. Alsmeyer. "On the Markov renewal theorem". In: *Stochastic Processes and their Applications* 50.1 (Mar. 1994), pp. 37–56.

[Als97] G. Alsmeyer. "The Markov renewal theorem and related results". In: *Markov Processes and Related Fields* 3.1 (1997), pp. 103–127.

[Als14] G. Alsmeyer. "Quasistochastic matrices and Markov renewal theory". In: *Journal of Applied Probability* 51.A (Dec. 2014), pp. 359–376.

[Als15a] G. Alsmeyer. "Random Recursive Equations and Their Distributional Fixed Points". June 2015. URL: `https://www.uni-muenster.de/Stochastik/lehre/SS15/StRek/StochRekgl.pdf`.

[Als15b] G. Alsmeyer. "Stochastische Prozesse Band 2". German. Erneuerungs- und Fluktuationstheorie für Random Walks, Markov-Erneuerungstheorie. 2015.

[AB17a] G. Alsmeyer and F. Buckmann. "Stability of perpetuities in Markovian environment". In: *Journal of Difference Equations and Applications* 23.4 (Jan. 2017), pp. 699–740.

[AB17b] G. Alsmeyer and F. Buckmann. "Fluctuation Theory for Markov Random Walks". In: *Journal of Theoretical Probability* 31.4 (Aug. 2017), pp. 2266–2342.

[AF01] G. Alsmeyer and C.-D. Fuh. "Limit theorems for iterated random functions by regenerative methods". In: *Stochastic Processes and their Applications* 96.1 (Nov. 2001), pp. 123–142.

[AF02] G. Alsmeyer and C.-d. Fuh. "Corrigendum to: "Limit theorems for iterated random functions by regenerative methods"". In: *Stochastic Processes and their Applications* 97.2 (Feb. 2002), pp. 341–345.

[AG99] G. Alsmeyer and A. Gut. "Limit Theorems for Stopped Functionals of Markov Renewal Processes". In: *Annals of the Institute of Statistical Mathematics* 51.2 (June 1999), pp. 369–382.

[AI09] G. Alsmeyer and A. Iksanov. "A Log-Type Moment Result for Perpetuities and Its Application to Martingales in Supercritical Branching Random Walks". In: *Electronic Journal of Probability* 14.0 (2009), no. 10, 289–312.

[AC92] L. Arnold and H. Crauel. "Iterated Function Systems and Multiplicative Ergodic Theory". In: *Diffusion Processes and Related Problems in Analysis, Volume II.* Vol. 27. Progr. Probab. Birkhäuser Boston, 1992, pp. 283–305.

[Asm03] S. Asmussen. *Applied Probability and Queues.* Second. Vol. 51. Applications of Mathematics (New York). Stochastic Modelling and Applied Probability. Springer New York, May 15, 2003, pp. xii+438. 460 pp.

[AMN78] K. B. Athreya, D. McDonald, and P. Ney. "Limit Theorems for Semi-Markov Processes and Renewal Theory for Markov Chains". In: *The Annals of Probability* 6.5 (Oct. 1978), pp. 788–797.

[Bar93] M. F. Barnsley. *Fractals Everywhere*. 2nd. Academic Press, Aug. 1993.

[BE88] M. F. Barnsley and J. H. Elton. “A new class of Markov processes for image encoding”. In: *Advances in Applied Probability* 20.1 (Mar. 1988), pp. 14–32.

[BEH89] M. F. Barnsley, J. H. Elton, and D. P. Hardin. “Recurrent iterated function systems”. In: *Constructive Approximation* 5.1 (Dec. 1989). Fractal approximation, pp. 3–31.

[Bou01] J. Bourdon. “Size and path length of Patricia tries: Dynamical sources context”. In: *Random Structures and Algorithms* 19.3-4 (2001). Analysis of algorithms (Krynica Morska, 2000), pp. 289–315.

[Bra86] A. Brandt. “The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients”. In: *Advances in Applied Probability* 18.1 (Mar. 1986), pp. 211–220.

[BH12] N. Broutin and C. Holmgren. “The total path length of split trees”. In: *The Annals of Applied Probability* 22.5 (Oct. 2012), pp. 1745–1777.

[Buc16] F. Buckmann. “Fluctuation theory of markov random walks and markov modulated random difference equations”. PhD thesis. Westfälische Wilhelms-Universität Münster, 2016.

[Car83] H. Carlsson. “Remainder Term Estimates of the Renewal Function”. In: *The Annals of Probability* 11.1 (Feb. 1983), pp. 143–157.

[CV15] E. Cesaratto and B. Vallée. “Gaussian Distribution of Trie Depth for Strongly Tame Sources”. In: *Combinatorics, Probability and Computing* 24.1 (Dec. 2015), pp. 54–103.

[CL75] Y. S. Chow and T. L. Lai. “Some one-sided theorems on the tail distribution of sample sums with applications to the last time and largest excess of boundary crossings”. In: *Transactions of the American Mathematical Society* 208 (1975), pp. 51–72.

[Çin69] E. Çinlar. “Markov renewal theory”. In: *Advances in Applied Probability* 1.2 (1969), pp. 123–187.

[CFV01] J. Clément, P. Flajolet, and B. Vallée. “Dynamical sources in information theory: A general analysis of trie structures”. In: *Algorithmica* 29.1-2 (Feb. 2001), pp. 307–369.

[Clé00] J. Clément. “Arbres Digitaux et Sources Dynamiques”. PhD thesis. Université de Caen Basse-Normandie, 2000.

[DG07] F. Dennert and R. Grübel. “Renewals for exponentially increasing lifetimes, with an application to digital search trees”. In: *The Annals of Applied Probability* 17.2 (Apr. 2007), pp. 676–687.

[Dev82] L. Devroye. “A note on the average depth of tries”. In: *Computing* 28.4 (Dec. 1982), pp. 367–371.

[Dev84] L. Devroye. “A probabilistic analysis of the height of tries and of the complexity of triesort”. In: *Acta Informatica* 21.3 (Oct. 1984), pp. 229–237.

[Dev92] L. Devroye. “A Study of Trie-Like Structures Under the Density Model”. In: *The Annals of Applied Probability* 2.2 (May 1992), pp. 402–434.

[DJ11] L. Devroye and S. Janson. “Long and short paths in uniform random recursive dags”. In: *Arkiv för Matematik* 49.1 (Apr. 2011), pp. 61–77.

[DF99] P. Diaconis and D. Freedman. “Iterated Random Functions”. In: *SIAM Review* 41.1 (Jan. 1999), pp. 45–76.

[Doo48] J. L. Doob. "Renewal theory from the point of view of the theory of probability". In: *Transactions of the American Mathematical Society* 63.3 (Mar. 1948), pp. 422–438.

[Drm09] M. Drmota. *Random Trees*. Springer Vienna, 2009.

[DF66] L. E. Dubins and D. A. Freedman. "Invariant Probabilities for Certain Markov Processes". In: *The Annals of Mathematical Statistics* 37.4 (Aug. 1966), pp. 837–848.

[Duf97] M. Duflo. *Random Iterative Models*. Springer Berlin Heidelberg, 1997.

[Elt90] J. H. Elton. "A multiplicative ergodic theorem for lipschitz maps". In: *Stochastic Processes and their Applications* 34.1 (Feb. 1990), pp. 39–47.

[FGD95] P. Flajolet, X. Gourdon, and P. Dumas. "Mellin transforms and asymptotics: Harmonic sums". In: *Theoretical Computer Science* 144.1-2 (June 1995). Special volume on mathematical analysis of algorithms, pp. 3–58.

[FRV10] P. Flajolet, M. Roux, and B. Vallée. "Digital trees and memoryless sources: from arithmetics to analysis". In: *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, 2010, pp. 233–260.

[FS09] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009, pp. xiv+810.

[FLY16] M. Fuchs, C.-K. Lee, and G.-R. Yu. "On 2-protected nodes in random digital trees". In: *Theoretical Computer Science* 622 (Apr. 2016), pp. 111–122.

[FK60] H. Furstenberg and H. Kesten. "Products of Random Matrices". In: *The Annals of Mathematical Statistics* 31.2 (June 1960), pp. 457–469.

[Gai+12] J. Gaither, Y. Homma, M. Sellke, and M. D. Ward. "On the number of 2-protected nodes in tries and suffix trees". In: *23rd Intern. Meeting on Probabilistic, Combinatorial, and Asymptotic Methods for the Analysis of Algorithms (AofA'12)*. Discrete Math. Theor. Comput. Sci. Proc., AQ. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2012, pp. 381–397.

[GW13] J. Gaither and M. D. Ward. "The Variance of the Number of 2-Protected Nodes in a Trie". In: *2013 Proceedings of the Tenth Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*. Society for Industrial and Applied Mathematics, Jan. 2013, pp. 43–51.

[GM00] C. M. Goldie and R. A. Maller. "Stability of perpetuities". In: *The Annals of Probability* 28.3 (June 2000), pp. 1195–1218.

[Gut09] A. Gut. *Stopped Random Walks*. Second. Springer Series in Operations Research and Financial Engineering. Limit theorems and applications. Springer, New York, 2009, pp. xiv+263.

[Hol12] C. Holmgren. "Novel characteristics of split trees by use of renewal theory". In: *Electronic Journal of Probability* 17.0 (2012), no. 5, 27.

[Hun14] K. Hun. "Analysis of depth of digital trees built on general sources". PhD thesis. Université de Caen Basse-Normandie, 2014.

[HV14] K. Hun and B. Vallée. “Typical Depth of a Digital Search Tree built on a general source”. In: *2014 Proceedings of the Eleventh Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*. Society for Industrial and Applied Mathematics, Dec. 2014, pp. 1–15.

[Hut81] J. E. Hutchinson. “Fractals and Self Similarity”. In: *Indiana University Mathematics Journal* 30.5 (1981), pp. 713–747.

[Iks06] O. M. Iksanov. “On the rate of convergence of a regular martingale related to a branching random walk”. In: *Ukrainian Mathematical Journal* 58.3 (Mar. 2006), pp. 368–387.

[JST01] P. Jacquet, W. Szpankowski, and J. Tang. “Average Profile of the Lempel-Ziv Parsing Scheme for a Markovian Source”. In: *Algorithmica* 31.3 (Nov. 2001). Mathematical analysis of algorithms, pp. 318–360.

[JR86a] P. Jacquet and M. Régnier. *Limiting distributions for trie parameters*. Tech. rep. RR-0502. INRIA, Mar. 1986. URL: https://hal.inria.fr/inria-00076052.

[JR86b] P. Jacquet and M. Régnier. “Trie partitioning process: Limiting distributions”. In: *CAAP ’86*. Vol. 214. Lecture Notes in Comput. Sci. Springer, Berlin, 1986, pp. 196–210.

[JR88] P. Jacquet and M. Régnier. *Normal limiting distribution for the size and the external path length of tries*. Tech. rep. RR-0827. INRIA, Apr. 1988. URL: https://hal.inria.fr/inria-00075724.

[JS91] P. Jacquet and W. Szpankowski. “Analysis of digital tries with Markovian dependency”. In: *IEEE Transactions on Information Theory* 37.5 (1991), pp. 1470–1475.

[JS16] P. Jacquet and W. Szpankowski. “Average size of a suffix tree for Markov sources”. In: *Proceedings of the 27th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms—AofA’16*. Jagiellonian Univ., Dep. Theor. Comput. Sci., Kraków, 2016, p. 13.

[Jan04] S. Janson. “One-Sided Interval Trees”. In: *Journal of the Iranian Statistical Society* 3.2 (2004). URL: http://jirss.irstat.ir/article-1-109-en.html.

[Jan06] S. Janson. “Rounding of continuous random variables and oscillatory asymptotics”. In: *The Annals of Probability* 34.5 (Sept. 2006), pp. 1807–1826.

[Jan12a] S. Janson. “Renewal theory in the analysis of tries and strings”. In: *Theoretical Computer Science* 416 (Jan. 2012), pp. 33–54.

[Jan12b] S. Janson. “More on tries”. Sept. 2012.

[Jan13] S. Janson. “Some integrals related to the Gamma integral”. June 2013. URL: http://www2.math.uu.se/~svante/papers/sjN11.pdf.

[Kes74] H. Kesten. “Renewal Theory for Functionals of a Markov Chain with General State Space”. In: *The Annals of Probability* 2.3 (June 1974), pp. 355–386.

[KP86] P. Kirschenhofer and H. Prodinger. “Some further results on digital search trees”. In: *Automata, Languages and Programming*. Vol. 226. Lecture Notes in Comput. Sci. Springer Berlin Heidelberg, 1986, pp. 177–185.

[KP88] P. Kirschenhofer and H. Prodinger. “Further results on digital search trees”. In: *Theoretical Computer Science* 58.1-3 (June 1988). Thirteenth International Colloquium on Automata, Languages and Programming (Rennes, 1986), pp. 143–154.

[Knu63] D. E. Knuth. *Notes On "Open" Addressing*. CiteSeerX. Typeset by H. Prodinger. 1963.

[Knu73] D. E. Knuth. *The art of computer programming. Vol. 3*. Sorting and searching, Addison-Wesley, Reading, Mass.-London-Don Mills, Ont., 1973, xi+722 pp. (1 foldout).

[Knu98] D. E. Knuth. *The art of computer programming. Vol. 3*. Sorting and searching, Second edition. Addison-Wesley, Reading, MA, 1998, pp. xiv+780.

[Lal86] S. P. Lalley. "Renewal theorem for a class of stationary sequences". In: *Probability Theory and Related Fields* 72.2 (1986), pp. 195–213.

[Lec15] K. Leckey. "Probabilistic analysis of radix algorithms on Markov sources". PhD thesis. Johann Wolfgang Goethe-Universität Frankfurt am Main, 2015.

[LNS15] K. Leckey, R. Neininger, and W. Szpankowski. "A Limit Theorem for Radix Sort and Tries with Markovian Input". In: *arXiv.org* (May 27, 2015). URL: `https://arxiv.org/abs/1505.07321`.

[Lig85] T. M. Liggett. "An Improved Subadditive Ergodic Theorem". In: *The Annals of Probability* 13.4 (Nov. 1985), pp. 1279–1285.

[Mah92] H. M. Mahmoud. *Evolution of random search trees*. Wiley-Interscience Series in Discrete Mathematics and Optimization. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1992, pp. xii+324.

[Mah08] H. M. Mahmoud. "Imbalance in Random Digital Trees". In: *Methodology and Computing in Applied Probability* 11.2 (June 2008), pp. 231–247.

[MT93] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Communications and Control Engineering Series. Springer London, 1993, pp. xvi+ 548.

[MR05] H. Mohamed and P. Robert. "A probabilistic analysis of some tree algorithms". In: *The Annals of Applied Probability* 15.4 (Nov. 2005), pp. 2445–2471.

[MR10] H. Mohamed and P. Robert. "Dynamic tree algorithms". In: *The Annals of Applied Probability* 20.1 (Feb. 2010), pp. 26–51.

[NR04] R. Neininger and L. Rüschendorf. "On the contraction method with degenerate limit equation". In: *The Annals of Probability* 32.3B (July 2004), pp. 2838–2856.

[NS15] R. Neininger and H. Sulzbach. "On a functional contraction method". In: *The Annals of Probability* 43.4 (July 2015), pp. 1777–1822.

[NN87] P. Ney and E. Nummelin. "Markov Additive Processes I. Eigenvalue Properties and Limit Theorems". In: *The Annals of Probability* 15.2 (Apr. 1987), pp. 561–592.

[Par+08] G. Park, H.-K. Hwang, P. Nicodème, and W. Szpankowski. "Profile of Tries". In: *Lecture Notes in Computer Science*. Vol. 4957. Lecture Notes in Comput. Sci. Springer Berlin Heidelberg, 2008, pp. 1–11.

[Par+09] G. Park, H.-K. Hwang, P. Nicodème, and W. Szpankowski. "Profiles of Tries". In: *SIAM Journal on Computing* 38.5 (Jan. 2009), pp. 1821–1880.

[Pit85] B. Pittel. "Asymptotical Growth of a Class of Random Trees". In: *The Annals of Probability* 13.2 (May 1985), pp. 414–427.

[Pit86] B. Pittel. "Paths in a random digital tree: limiting distributions". In: *Advances in Applied Probability* 18.1 (Mar. 1986), pp. 139–155.

[PW96] J. G. Propp and D. B. Wilson. "Exact sampling with coupled Markov chains and applications to statistical mechanics". In: *Proceedings of the Seventh International Conference on Random Structures and Algorithms (Atlanta, GA, 1995)*. Vol. 9. 1-2. Wiley, Aug. 1996, pp. 223–252.

[PW98] J. G. Propp and D. B. Wilson. "How to Get a Perfectly Random Sample from a Generic Markov Chain and Generate a Random Spanning Tree of a Directed Graph". In: vol. 27. 2. 7th Annual ACM-SIAM Symposium on Discrete Algorithms (Atlanta, GA, 1996). Elsevier BV, May 1998, pp. 170–217.

[RJS93] B. Rais, P. Jacquet, and W. Szpankowski. "Limiting Distribution for the Depth in PATRICIA Tries". In: *SIAM Journal on Discrete Mathematics* 6.2 (May 1993), pp. 197–213.

[Rég88] M. Régnier. "Trie hashing analysis". In: *Proceedings. Fourth International Conference on Data Engineering.* IEEE Comput. Soc. Press, 1988.

[RJ89] M. Régnier and P. Jacquet. "New results on the size of tries". In: *IEEE Transactions on Information Theory* 35.1 (1989), pp. 203–205.

[Rös91] U. Rösler. "A limit theorem for "quicksort"". In: *RAIRO - Theoretical Informatics and Applications* 25.1 (1991), pp. 85–100.

[Rös92] U. Rösler. "A fixed point theorem for distributions". In: *Stochastic Processes and their Applications* 42.2 (Sept. 1992), pp. 195–214.

[Sav99] S. A. Savari. "Variable-to-fixed length codes and the conservation of entropy". In: *IEEE Transactions on Information Theory* 45.5 (July 1999), pp. 1612–1620.

[Sav00] S. A. Savari. "Renewal theory and source coding". In: *Proceedings of the IEEE* 88.11 (Nov. 2000), pp. 1692–1702.

[SG97] S. A. Savari and R. G. Gallager. "Generalized Tunstall codes for sources with memory". In: *IEEE Transactions on Information Theory* 43.2 (Mar. 1997), pp. 658–668.

[Shu85] V. M. Shurenkov. "On the Theory of Markov Renewal". In: *Theory of Probability & Its Applications* 29.2 (Jan. 1985), pp. 247–265.

[Sto65] C. Stone. "On Moment Generating Functions and Renewal Theory". In: *The Annals of Mathematical Statistics* 36.4 (Aug. 1965), pp. 1298–1301.

[Szp88] W. Szpankowski. "Some results on V-ary asymmetric tries". In: *Journal of Algorithms* 9.2 (June 1988), pp. 224–244.

[Szp90] W. Szpankowski. "Patricia tries again revisited". In: *Journal of the ACM* 37.4 (Oct. 1990), pp. 691–711.

[Szp91] W. Szpankowski. "On the height of digital trees and related problems". In: *Algorithmica* 6.2 (June 1991), pp. 256–277.

[Szp01] W. Szpankowski. *Average Case Analysis of Algorithms on Sequences.* Wiley-Interscience Series in Discrete Mathematics and Optimization. With a foreword by Philippe Flajolet. John Wiley & Sons, Apr. 2001, pp. xxii+551. 580 pp.

[Val01] B. Vallée. "Dynamical sources in information theory: Fundamental intervals and word prefixes". In: *Algorithmica* 29.1-2 (Feb. 2001). Average-case analysis of algorithms (Princeton, NJ, 1998), pp. 262–306.

[Val+09] B. Vallée, J. Clément, J. A. Fill, and P. Flajolet. “The Number of Symbol Comparisons in QuickSort and QuickSelect”. In: *Automata, Languages and Programming.* Vol. 5555. Lecture Notes in Comput. Sci. Springer Berlin Heidelberg, 2009, pp. 750–763.

[Ver79] W. Vervaat. “On a stochastic difference equation and a representation of non-negative infinitely divisible random variables”. In: *Advances in Applied Probability* 11.4 (Dec. 1979), pp. 750–783.

Acronyms

CLT	central limit theorem
DST	digital search tree
IFS	iterated function system (of i.i.d. Lipschitz maps)
MIFS	iterated function system of Markov-modulated Lipschitz maps
MMS	Markov-modulated sequence
MRW	Markov random walk
PAT	PATRICIA(-trie), short for *Practical Algorithm to Retrieve Information Coded in Alphanumeric*
RP	(delayed) renewal process
RW	(delayed) random walk
SLLN	strong law of large numbers
SRP	standard renewal process
SRW	standard random walk
WLLN	weak law of large numbers

List of symbols

$a \wedge b$	$= \min\{a, b\}$
$a \vee b$	$= \max\{a, b\}$
x^+	$= \max\{0, x\}$, positive part of x (rarely denoted by x_+ for simplicity)
x^-	$= \max\{0, -x\}$, negative part of x
$\{x\}$	$= x - \lfloor x \rfloor$, fractional part of x
$(\Omega, \mathcal{A}, \mathbb{P})$	underlying probability space
$\mathbb{P}^X$	$= \mathbb{P}(X \in \cdot)$, distribution of X w.r.t. the probability measure $\mathbb{P}$
$\sim$	asymptotically equivalent, distributed as, $\lambda\!\!\lambda_0$-almost everywhere equal

Part 1: Renewal theory in the analysis of random digital trees

$\mathcal{S}$	finite state space (= alphabet)
$\mathcal{S}^*$	complete infinite tree with nodes labeled by finite strings of letters in $\mathcal{S}$
$\mathcal{S}_j^*$	$\subset \mathcal{S}^*$, subset of all nodes that end with j
Ξ	$= (\xi_n)_{n \geq 1}$, random string with random letters $\xi_n \in \mathcal{S}$, Markov chain on $\mathcal{S}$
Ξ_0	$= (\xi_n)_{n \geq 0}$, Markov chain on $\mathcal{S}$, includes initial variable ξ_0
P	$= (p_{i,j})_{i,j \in \mathcal{S}}$, transition matrix of Ξ and Ξ_0
π	stationary distribution of Ξ and Ξ_0
$\mathbb{P}_i$	$= \mathbb{P}(\cdot \vert \xi_0 = i)$, $\mathbb{P}_\lambda = \sum_{i \in \mathcal{S}} \lambda_i \mathbb{P}_i$
$P_i(\alpha_1 \cdots \alpha_n)$	$= \mathbb{P}_i(\xi_1 = \alpha_1, \ldots, \xi_n = \alpha_n)$, function on $\mathcal{S}^*$
$(\Xi^{(k)})_k$	sequence of i.i.d. copies of Ξ

$\mathcal{M}$, $\mathcal{M}_n$	set of strings, $\mathcal{M}_n = \{\Xi^{(1)},\dots,\Xi^{(n)}\}$
$\mathrm{Trie}(\mathcal{M})$	trie constructed from the strings in $\mathcal{M}$, $\mathrm{Trie}^{(b)}(\mathcal{M})$, $\mathrm{Trie}^{P}(\mathcal{M})$, $\mathrm{Trie}^{(b),P}(\mathcal{M})$ analogously for b-, Pat-, and Pat-b-trie
$\beta \succ \alpha$	β starts with α, for two strings (which may be finite)
D_n, $D_n^{(b)}$, D_n^P	depth of Ξ in $\mathrm{Trie}(\mathcal{M}_n)$, $\mathrm{Trie}^{(b)}(\mathcal{M}_n)$ and $\mathrm{Trie}^{P}(\mathcal{M}_n)$, resp.
$D_{n,l}$, $D_{n,l}^P$	depth of $\Xi^{(l)}$ in $\mathrm{Trie}(\mathcal{M}_n)$ and $\mathrm{Trie}^{P}(\mathcal{M}_n)$, resp.
Y_k	$= 2\xi_k - 1$, $k \ge 1$, in the binary setting
$(V_n)_{n\ge0}$	$V_n = \sum_{k=1}^n Y_k$, $n \ge 1$, $V_0 = 0$
Δ_n	$= V_{D_n}$, imbalance factor of $\Xi = \Xi^{(1)}$ in $\mathrm{Trie}(\mathcal{M}_n)$
$(\sigma_n(i))_{n\ge0}$	sequence of successive recurrence times of the state i with underlying Markov chain Ξ_0, $\sigma_n(i) := \inf\{k > \sigma_{n-1}(i) : \xi_k = i\}$, $n \ge 1$, $\sigma_0(i) = 0$
$\tau_n(i)$	$= \sigma_n(i) - \sigma_{n-1}(i)$, $n \ge 1$, cycle lengths corresponding to $(\sigma_n(i))_{n\ge0}$
m_{ii}	$= \mathbb{E}_i\sigma_1(i)$
$(X_n)_{n\ge1}$	central Markov-modulated sequence driven by Ξ_0, $X_n = -\log p_{\xi_{n-1},\xi_n}$
$(S_n)_{n\ge0}$	$S_n = \sum_{k=1}^n X_k$, $n \ge 1$, $S_0 = 0$
H_i	$= \mathbb{E}_i X_1 = -\sum_{j\in\mathcal{S}} p_{i,j}\log p_{i,j}$
μ	$= \mathbb{E}_\pi X_1 = \sum_{i\in\mathcal{S}} \pi_i H_i$, stationary drift of $(S_n)_{n\ge0}$
μ_Y	$= \mathbb{E}_\pi Y_1$, stationary drift of $(V_n)_{n\ge0}$
F_{ii}, F_{ij}	$= \mathbb{P}_i^{S_{\sigma_1(i)}}$, $= \mathbb{P}_i^{S_{\sigma_1(j)}}$, resp.
$\mathbb{U}^{ii}$, $\mathbb{U}^{ij}$	$= \sum_{n\ge0} F_{ii}^{*(n)}$, $= F_{ij} * \mathbb{U}^{jj}$, resp., corresponding renewal measures w.r.t. $\mathbb{P}_i$
$\mathbb{U}^{ij}(t)$	$= \mathbb{U}^{ij}((-\infty,t])$, $t \in \mathbb{R}$, corresponding renewal function
$\mathbb{U}_i$	$= \sum_{n\ge0} \mathbb{P}_i((\xi_n, S_n) \in \cdot)$, Markov renewal measure of $(\xi_n, S_n)_{n\ge0}$
$\mathbf{U}(i,\cdot)$	$= \mathbb{U}_i(\cdot)$, corresponding Markov renewal kernel
$\mathbf{U} * g(i,t)$	$= \int_{\mathcal{S}\times\mathbb{R}} g(s, t-x)\mathbb{U}_i(\mathrm{d}s, \mathrm{d}x)$, for measurable $g : \mathcal{S}\times\mathbb{R} \to \mathbb{R}$
$\nu(t)$	$= \inf\{n \ge 0 : S_n > t\}$, first passage times of $(S_n)_{n\ge0}$
$\nu(x,t)$	$= \inf\{n \ge 0 : x + S_n > t\}$
$\nu^i(t)$	$= \inf\{n \ge 0 : S_n > t\}$, first passage times of $(S_{\sigma_n(i)})_{n\ge0}$
$\mathbf{0}_i$	$= \{j \in \mathcal{S} : p_{i,j} = 0\}$
$\mathbf{1}$	$= \{j \in \mathcal{S} : \exists k \in \mathcal{S} \text{ s.t. } p_{j,k} = 1\}$
β	shift function of the MRW $(\xi_n, S_n)_{n\ge0}$, $\beta : \mathcal{S} \to [0, d)$
β_{ij}	$= \beta(j) - \beta(i)$
$(\widetilde{S}_n)_{n\ge0}$	shifted MRW with $\widetilde{S}_n = S_n - \beta(\xi_n) + \beta(\xi_0)$, $n \ge 1$, $\widetilde{S}_0 = 0$
$d(i) = d$	lattice span of $S_{\sigma_1(i)}$ and $\widetilde{S}_n$ w.r.t. $\mathbb{P}_i$
$\widetilde{\mathbb{U}}^{ij}$, $\widetilde{\mathbb{U}}_i$	same objects as $\mathbb{U}^{ij}$ and $\mathbb{U}_i$ but with S_n replaced by $\widetilde{S}_n$
γ	$= 0.5772\ldots$, Euler constant
$(X_0^{(n)})_{n\ge2}$	family of initial variables with laws defined in (2.3)
X_0^*	distributional limit of the above sequence
$(Y_0^{(n)})_{n\ge b+1}$	family of initial variables with laws defined in (2.4)
Y_0^*	distributional limit of the above sequence
$\sigma^{(2)}$, $\gamma^{(2)}$	$= \frac{1}{m_{ii}}\mathbb{V}\mathrm{ar}_i(S_{\sigma_1(i)} - \mu\sigma_1(i))$, $= \frac{1}{m_{ii}}\mathbb{V}\mathrm{ar}_i(\mu V_{\sigma_1(i)} - \mu_Y S_{\sigma_1(i)})$, resp.
$F_i(\lambda)$	$= \sum_{\alpha\in\mathcal{S}^*} f(\lambda P_i(\alpha))\mathbb{1}_{\{P_i(\alpha)>0\}}$
$F_i^j(\lambda)$	$= \sum_{\alpha\in\mathcal{S}^*} f(\lambda P_i(\alpha j))\mathbb{1}_{\{P_i(\alpha j)>0\}}$
W_n, $W_n^{(b)}$	size of trie and b-trie, resp.
W_n^P, $W_n^{(b),P}$	size of Pat-trie and Pat-b-trie, resp.
$N_n(\alpha)$	$= \sum_{k=1}^n \mathbb{1}_{\{\Xi^{(k)}\succ\alpha\}}$, $\alpha \in \mathcal{S}^*$

$Z_n^{(b)}$	number of external nodes of $\mathrm{Trie}^{(b)}(\mathcal{M}_n)$
$Z_{l,n}^{(b)}$	number of external nodes of $\mathrm{Trie}^{(b)}(\mathcal{M}_n)$ containing exactly l strings
$Z_{l,n}^{(b),P}$	number of external nodes of $\mathrm{Trie}^{(b),P}(\mathcal{M}_n)$ containing exactly l strings
ΔD_n	$= D_n - D_n^P$
L_n	$= \sum_{l=1}^n D_{n,l}$, external path length of $\mathrm{Trie}(\mathcal{M}_n)$
L_n^P	$= \sum_{l=1}^n D_{n,l}^P$, external path length of $\mathrm{Trie}^P(\mathcal{M}_n)$
$(\Pi(\lambda))_{\lambda>0}$	family of $\mathrm{Poi}(\lambda)$ distributed random variables
$\widetilde{W}_\lambda$	$= W_{\Pi(\lambda)}$, and similarly $\widetilde{W}_\lambda^{(b)}, \widetilde{W}_\lambda^P, \dots$ (possibly $1+\Pi(\lambda)$ instead of $\Pi(\lambda)$)

Part 2: Convergence rates of MIFSs by regenerative methods

$(\mathbb{X}, d)$	complete separable metric space
$\mathcal{S}$	(at most) countable state space
$\mathrm{Lip}(\mathbb{X}, \mathbb{X})$	set of all Lipschitz continuous functions from $\mathbb{X}$ to $\mathbb{X}$
$L(f)$	Lipschitz constant of f
d_{Pr}	Prokhorov metric on the space of probability measures on $\mathbb{X}$
(θ, Ψ)	$= (\theta_n, \Psi_n)_{n\geq 0}$, central MMS with state space $\mathcal{S} \times \mathrm{Lip}(\mathbb{X}, \mathbb{X})$
P, Q	transition kernels of θ and (θ, Ψ), resp., $P = (p_{i,j})_{i,j\in\mathcal{S}}$
K_{ij}	$= P(\Psi_n \in \cdot \mid \theta_{n-1} = i, \theta_n = j)$ for $i, j \in \mathcal{S}$, $n \geq 1$
π	stationary distribution of θ (and ${}^{\#}\theta$)
$(X_n)_{n\geq 0}$	central MIFS with $X_n = \Psi_n \circ \cdots \circ \Psi_1(X_0)$
$({}^{\#}\theta, {}^{\#}\Psi)$	$= ({}^{\#}\theta_n, {}^{\#}\Psi_n)_{n\geq 0}$, dual MMS with ${}^{\#}\theta_0 = \theta_0$
${}^{\#}P$, ${}^{\#}Q$, ${}^{\#}K_{ij}$	corresponding kernels as defined in (6.1)
$\mathbb{P}_i$	$= \mathbb{P}(\cdot \mid {}^{\#}\theta_0 = \theta_0 = i)$, $\mathbb{P}_\lambda = \sum_{i\in\mathcal{S}} \lambda_i \mathbb{P}_i$
$(\sigma_n(i))_{n\geq 0}$	sequence of successive recurrence times of the state i with underlying Markov chain θ, defined as in Part 1
$({}^{\#}\sigma_n(i))_{n\geq 0}$	similar to $(\sigma_n(i))_{n\geq 0}$ but with ${}^{\#}\theta$ instead of θ, often the $^{\#}$ is dropped for simplicity
m_{ii}	$= \mathbb{E}_i \sigma_1(i) = \mathbb{E}_i {}^{\#}\sigma_1(i)$
$(X_n^x)_{n\geq 0}$	$= (X_n)_{n\geq 0}$ with $X_0 = x$
$({}^{\#}X_n^x)_{n\geq 0}$	$= (X_n^x)_{n\geq 0}$ with Lipschitz maps ${}^{\#}\Psi$ instead of Ψ
$\widehat{\ }$	denotes backward iterations of the corresponding MIFS or IFS
${}^{\#}\widehat{X}_\infty$	$\mathbb{P}_\pi$-a.s. limit of $({}^{\#}\widehat{X}_n^x)_{n\geq 0}$
χ	$= \lim_{n\to\infty} \frac{1}{n} \log L({}^{\#}\Psi_1 \cdots {}^{\#}\Psi_n)$, Lyapunov exponent
$\log_*(x)$	$= \log(1+x)$
$(S_n)_{n\geq 0}$	central MRW (with driving chain ${}^{\#}\theta$), $S_n = \sum_{k=1}^n \log L({}^{\#}\Psi_k)$, $n \geq 1$, $S_0 = 0$
μ	$= \mathbb{E}_\pi \log L(\Psi_1) = \mathbb{E}_\pi \log L({}^{\#}\Psi_1)$, stationary drift of $(S_n)_{n\geq 0}$
$(S_n^*)_{n\geq 0}$	similar to $(S_n)_{n\geq 0}$ but $S_n^* = \sum_{k=1}^n \log_* L({}^{\#}\Psi_k)$, $n \geq 1$
$({}^{\#}\Psi_n^i)_{n\geq 1}$	cyclic concatenation of Lipschitz maps, ${}^{\#}\Psi_n^i = {}^{\#}\Psi_{\sigma_n(i)} \cdots {}^{\#}\Psi_{\sigma_{n-1}(i)+1}$
$({}^{\#}\vec{\Psi}_n^i)_{n\geq 1}$	reversed cyclic concatenation of Lipschitz maps, ${}^{\#}\vec{\Psi}_n^i = {}^{\#}\Psi_{\sigma_{n-1}(i)+1} \cdots {}^{\#}\Psi_{\sigma_n(i)}$
$(Y_n^x)_{n\geq 0}$	IFS of i.i.d. Lipschitz maps $({}^{\#}\vec{\Psi}^i)_{n\geq 1}$ and $Y_0^x = x$
β_i^*	parameter with $\mathbb{E}_i e^{\beta_i^* \sigma_1(i)} < \infty$
γ_i^*	$= e^{\mathbb{E}_i \log L({}^{\#}\vec{\Psi}_1^i)}$ if $\mathbb{E}_i \log L({}^{\#}\vec{\Psi}_1^i) > -\infty$, and $= 0$ if $\mathbb{E}_i \log L({}^{\#}\vec{\Psi}_1^i) = -\infty$, lower bound for the rate of exponential convergence of $(Y_n^x)_{n\geq 0}$
$\gamma_{i,c}^*$	$= e^{\mathbb{E}_i \log(L({}^{\#}\vec{\Psi}_1^i) \vee c)}$

$L_k^{*,i}$	$= L(^{\#}\vec{\Psi}_k^i)/\gamma_i^*$, $k \geq 1$
$L_{k,c}^{*,i}$	$= (L(^{\#}\vec{\Psi}_k^i) \vee c)/\gamma_{i,c}^*$, $k \geq 1$
$(S_n^{*,i})_{n\geq 0}$	auxiliary SRW, $S_n^{*,i} = \sum_{k=1}^n \log L_k^{*,i}$, $n \geq 1$, $S_0^{*,i} = 0$
$\tau(n)$	$= \inf\{k \geq 0 \mid {}^{\#}\sigma_k(i) \geq n\}$
$(C_n)_{n\geq 1}$	as defined in (7.11)
$(G_n)_{n\geq 1}$	as defined in (7.12)
$(D_n)_{n\geq 0}$	as defined in (7.13)
$\bar{G}_1$	as defined in (B.2)