

## ON THE OCCUPANCY PROBLEM FOR A REGIME-SWITCHING MODEL

MICHAEL GRABCHAK,\* *University of North Carolina Charlotte*  
MARK KELBERT,\*\* AND  
QUENTIN PARIS,\*\* *National Research University Higher School of Economics*

### Abstract

This article studies the expected occupancy probabilities on an alphabet. Unlike the standard situation, where observations are assumed to be independent and identically distributed, we assume that they follow a regime-switching Markov chain. For this model, we (1) give finite sample bounds on the expected occupancy probabilities, and (2) provide detailed asymptotics in the case where the underlying distribution is regularly varying. We find that in the regularly varying case the finite sample bounds are rate optimal and have, up to a constant, the same rate of decay as the asymptotic result.

*Keywords:* occupancy problem; regime switching; Markov chain; regular variation

2010 Mathematics Subject Classification: Primary 60J10; Secondary 60J20

### 1. Introduction

Let  $A$  be a finite or countably infinite set and let  $X = (X_n)_{n \geq 1}$  be a discrete-time  $A$ -valued stochastic process defined on some probability space  $(\Omega, \mathcal{F}, P)$ . We refer to set  $A$  as the alphabet and to elements of  $A$  as letters. These letters may represent different things in the context of different applications. For instance, in linguistics they may represent words in some language, while in ecology they may represent species in an ecosystem. From a general point of view, the occupancy problem (or urn scheme) is to describe the repartition of the process  $(X_n)_{n \geq 1}$  over the set  $A$ . In this context two quantities of interest are

$$L_n = \sum_{i=1}^n \mathbf{1}\{X_i = X_{n+1}\} \quad \text{and} \quad M_{n,r} = P\{L_n = r \mid X_1, \dots, X_n\}.$$

These quantities are related by the fact that

$$P\{L_n = r\} = E\{M_{n,r}\}.$$

In words,  $L_n$  is the number of times that the letter observed at time  $n + 1$  had previously been observed, and  $M_{n,r}$  is the probability that, given the observations up to time  $n$ , the letter observed at time  $n + 1$  will have already been seen  $r$  times. We refer to the quantities  $M_{n,r}$

---

Received 7 January 2019; revision received 18 September 2019; accepted 3 October 2019.

\* Postal address: Department of Mathematics and Statistics, Charlotte, NC, USA. Email address: [mgrabcha@uncc.edu](mailto:mgrabcha@uncc.edu)

\*\* Postal address: National Research University Higher School of Economics (HSE), Faculty of Economics, Department of Statistics and Data Analysis, Moscow, Russia. Email address: [mkelbert@hse.ru](mailto:mkelbert@hse.ru)

\*\*\* Postal address: National Research University Higher School of Economics (HSE), Faculty of Computer Science, School of Data Analysis and Artificial Intelligence & HDI Lab, Moscow, Russia. Email address: [qparis@hse.ru](mailto:qparis@hse.ru)

as the occupancy probabilities. The quantity  $M_{n,0}$  is also sometimes called the missing mass. It corresponds to the probability of seeing a new letter at time  $n + 1$ . In certain ecological contexts, it represents the probability of discovering a new species. While properties of  $L_n$  and  $M_{n,r}$  have been thoroughly studied in the context where  $X_1, X_2, \dots$  are independent and identically distributed (i.i.d.) random variables, we have seen no work in the literature relating to the case where they follow a more general stochastic process. In this paper we give such results for a class of Markov chains that form a regime-switching model. This model expands the scope of potential applications. Moreover, it is our hope that this paper will stimulate interest in studying this problem in the context of other, more general, processes.

**1.1. Related work**

In the i.i.d. setting, the literature on the behavior of  $L_n, M_{n,r}$ , and related quantities is vast; see, for instance, the classic textbook [15], the survey [11], or the recent contributions in [2, 7]. Applications include fields such as ecology [5, 9, 12, 13], genomics [17], language processing [6], authorship attribution [8, 23, 25], information theory [1, 19], computer science [24], and machine learning [4, 14].

We now briefly sketch several key results for the case where the random variables  $X_1, X_2, \dots$  are i.i.d. with common distribution  $P = (p_a)_{a \in A}$  on  $A$ . In this case, it is readily shown that

$$P\{L_n = r\} = E\{M_{n,r}\} = \binom{n}{r} \sum_{a \in A} p_a^{1+r} (1 - p_a)^{n-r}.$$

This expression allows for a precise asymptotic analysis. Following [16], it is understood that the main ingredients for this analysis are given by the counting measure  $\nu_P$  and the counting function  $\nu$ . These are defined, respectively, by

$$\nu_P(du) = \sum_{a \in A} \delta_{p_a}(du) \tag{1}$$

and

$$\nu(\varepsilon) = \nu_P([\varepsilon, 1]) = \sum_{a \in A} \mathbf{1}\{p_a \geq \varepsilon\}, \quad 0 \leq \varepsilon \leq 1. \tag{2}$$

Next, recall that a function  $\ell : (0, +\infty) \rightarrow \mathbb{R}$  is said to be slowly varying at  $+\infty$  if, for any  $c > 0$ ,

$$\lim_{x \rightarrow +\infty} \frac{\ell(cx)}{\ell(x)} = 1.$$

In this case we write  $\ell \in \mathcal{SV}$ . With this notation, if  $\nu(\varepsilon) = \nu_P([\varepsilon, 1]) = \varepsilon^{-\alpha} \ell(1/\varepsilon)$  for some  $\alpha \in (0, 1)$  and some  $\ell \in \mathcal{SV}$ , then, for  $r \geq 0$ ,

$$E\{M_{n,r}\} \sim \frac{\alpha \Gamma(1 + r - \alpha)}{r!} n^{-(1-\alpha)} \ell(n). \tag{3}$$

This result is discussed, in greater detail, in the appendix. Non-asymptotic results are given in [7]. The main result of that paper is as follows.

**Lemma 1.1.** (Theorem 2.1 of [7].) Let  $P = (p_a)_{a \in A}$  be a probability measure on  $A$  with counting function  $v$ . For all  $n \geq 1$ , all  $0 \leq r \leq n - 1$ , and all  $0 \leq \varepsilon \leq 1$ ,

$$\begin{aligned} P\{L_n = r\} &= E\{M_{n,r}\} = \binom{n}{r} \sum_{a \in A} p_a^{1+r} (1 - p_a)^{n-r} \\ &\leq \frac{c(r)v(\varepsilon)}{n} + 2^{1+r} \binom{n}{r} \int_0^\varepsilon v\left(\frac{u}{2}\right) u^r \left(1 - \frac{u}{2}\right)^{n-r} du, \end{aligned}$$

where

$$c(r) = \begin{cases} e^{-1} & \text{if } r = 0, \\ e(1+r)/\sqrt{\pi} & \text{if } r \geq 1. \end{cases} \tag{4}$$

### 1.2. Regime-switching models

A natural extension of the i.i.d. case is to a class of regime-switching Markov chains or regime-switching models. In this context the elements in  $A$  no longer represent letters, but entire alphabets. Each  $a \in A$  represents an alphabet, which we denote by  $\{a\} \times \mathbb{N}_+$ , where  $\mathbb{N}_+ = \{1, 2, \dots\}$ . This alphabet has its own distribution  $P_a = (p_{a,k})_{k \geq 1}$ , and we assume that observations from each alphabet are i.i.d. with distribution  $P_a$ . However, we randomly perform transitions between alphabets following a Markov chain with transition operator  $Q$ . Formally, we consider a Markov chain  $Z = (Z_n)_{n \geq 1} = (X_n, K_n)_{n \geq 1}$  on the product space  $\mathcal{A} := A \times \mathbb{N}_+$  with transition operator  $Q$  defined by

$$Q((a', k'), (a, k)) = Q(a', a)p_{a,k}, \quad a, a' \in A, \quad k, k' \in \mathbb{N}_+. \tag{5}$$

We refer to  $(Z_n)_{n \geq 1}$  as the regime-switching model and to  $(X_n)_{n \geq 1}$  as the underlying process.

One important situation is when  $|A| < \infty$ . In this case we can use many properties of finite Markov chains, even though the process of interest is defined on the countably infinite state space  $\mathcal{A}$ . Some of our more detailed results are proved under this additional assumption. For other results, in the interests of generality we not only allow  $|A| = \infty$ , but we remove the assumption that the underlying process is a Markov chain. Nevertheless, our motivation comes from the case where the transitions are Markovian and  $|A| < \infty$ . Such models can be used to describe a variety of situations, such as:

1. (Classics) A researcher reads documents in an antique library. The documents are written in a variety of languages (e.g. Latin, Greek, Hebrew, etc.). Assume that transitions between documents written in different languages follow a Markov chain. Here, the regime-switching Markov chain  $(Z_n)_{n \geq 1}$  represents the sequence of ordered pairs comprised of the word that the researcher is currently reading and the language that the current document is written in. In this context, the missing mass represents the probability that the next word that the researcher encounters will be one that this researcher has not previously seen and will thus need to look up.
2. (Ecology) An ecologist is observing the animals that are found in a certain plot of a forest. However, the forest has several states (e.g. time of day, weather, etc.), with transitions between these following a Markov chain. To understand the differences in the distribution of species found under different states, the ecologist keeps track of both the species of the observed animal and the state of the forest.

3. (Computer science) A server periodically enters a state where there is a serious hacking attempt. Assume that transitions into and out of this state follow a Markov chain. To understand the effect of a serious hacking attempt on the number of packets that arrive, a researcher keeps track of the number of packets that arrive in time increments of, say, five minutes, along with the state of the server in that time period.
4. (Economics) An economy can be in one of several states, e.g. growth, recession, inflation, etc. One can model transitions between these states using a Markov chain. To understand the effect of the state of the economy on some economic indicator (e.g. the number of bank failures in a week), an economist keeps track of both the indicator and the state of the economy.

### 1.3. Organization

The main goal of this paper is to extend the results given in (3) and Lemma 1.1 from the i.i.d. case to the regime-switching model. We begin by giving results for a simple class of Markov chains, which will drive this model. Toward this end we introduce a useful technical result in Section 2, and then in Section 3 we consider the case of an ergodic Markov chain on a finite state space. In Section 4 we formally define the regime-switching model and give extensions of Lemma 1.1. In the interests of generality, most results in this section do not assume that transitions between alphabets are Markovian. However, this assumption is needed for the more detailed results. Then, in Section 5 we extend (3) to the case of the regime-switching model. Proofs are postponed to Section 6. A brief review of basic properties of regularly varying distributions on an alphabet is given in the appendix.

### 1.4. Notation

Before proceeding, we set up some notation. We write  $\mathbf{1}\{\cdot\}$  to denote the indicator function of event  $\{\cdot\}$ . For a set  $A$ , we write  $|A|$  to denote the cardinality of  $A$ . For real numbers  $a, b \in \mathbb{R}$ , we write  $a \vee b$  or  $\max\{a, b\}$  to denote the maximum of  $a$  and  $b$ , and we write  $a \wedge b$  or  $\min\{a, b\}$  to denote the minimum of  $a$  and  $b$ . For two sequences  $g(n)$  and  $h(n)$  we write  $g(n) \sim h(n)$  to mean  $g(n)/h(n) \rightarrow 1$  as  $n \rightarrow \infty$ . We write  $\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du$  for  $x > 0$  to denote the gamma function.

## 2. Preliminaries

In this section we introduce a technical result that will be useful later. Toward this end, fix a finite or countably infinite set  $A$ , a Markov transition operator  $Q$  on  $A$ , and a probability measure  $\mu$  on  $A$ . Let  $X = (X_n)_{n \geq 1}$  be an  $A$ -valued random process defined on some probability space  $(\Omega, \mathcal{F}, P_\mu)$  such that  $X$  is a  $Q$ -Markov chain with initial distribution  $\mu$ . We write  $E_\mu$  to denote the expectation under  $P_\mu$ . We write  $Q^t$  to denote the  $t$ -step transition operator of the Markov chain. For all integers  $n \geq 1$  and  $a \in A$ , we set

$$L_n(a) := \sum_{i=1}^n \mathbf{1}\{X_i = a\}$$

to be the local time of Markov chain  $X$  in state  $a$ , and we set

$$L_n := \sum_{i=1}^n \mathbf{1}\{X_i = X_{n+1}\} = L_n(X_{n+1})$$

to denote the number of times that the state visited at time  $n + 1$  had been visited up to time  $n$ .

We now give a result that connects the distribution of  $L_n$  with that of the local times of the reversed chain. We assume that the  $Q$ -Markov chain  $(X_n)_{n \geq 1}$  is irreducible, aperiodic, positive recurrent, and has stationary distribution  $\pi = (\pi_a)_{a \in A}$ . We denote by  $\hat{X} = (\hat{X}_n)_{n \geq 1}$  the associated reversed chain, i.e. an  $A$ -valued Markov chain with transition operator  $\hat{Q}$  defined by

$$\hat{Q}(x, y) := \frac{\pi(y)Q(y, x)}{\pi(x)}.$$

It is easy to check that  $\pi$  is also the stationary distribution of  $\hat{X}$  and that the  $t$ -step transition operator of the reversed chain is given by

$$\hat{Q}^t(x, y) = \frac{\pi(y)Q^t(y, x)}{\pi(x)}. \tag{6}$$

We say that the chain  $X$  is reversible when  $\pi(x)Q(x, y) = \pi(y)Q(y, x)$ . In this case,  $\hat{Q} = Q$  and the chains  $X$  and  $\hat{X}$  have the same distribution so long as they both start at the stationary distribution. We write  $\hat{L}_n(a)$  to denote the local time of the reversed chain at  $a$ , i.e.

$$\hat{L}_n(a) := \sum_{i=1}^n \mathbf{1}\{\hat{X}_i = a\}.$$

**Lemma 2.1.** *Let  $A$  be a finite or countably infinite set. Suppose  $X = (X_n)_{n \geq 1}$  is an irreducible, aperiodic, and positive recurrent Markov chain on  $A$  with stationary distribution  $\pi$  and reversed chain  $\hat{X}$ . Let  $\mu$  and  $\eta$  be arbitrary distributions on  $A$ . Then, for any positive function  $f$  and all integers  $n \geq 1$ ,*

$$E_\mu \left[ \frac{\eta(X_{n+1})}{\pi(X_{n+1})} f(L_n) \right] = E_\eta \left[ \frac{\mu(\hat{X}_{n+1})}{\pi(\hat{X}_{n+1})} f(\hat{L}_{n+1}(\hat{X}_1) - 1) \right],$$

where, on the right-hand side, it is understood that  $\eta$  is taken as the initial distribution of the reversed chain, i.e. it is the distribution of  $\hat{X}_1$ .

**Remark 2.1.** Note, in particular, that taking  $\mu = \eta = \pi$  in the above formula, and supposing the chain to be reversible, we get that for any positive  $f$ ,

$$E_\pi \{f(L_n)\} = E_\pi \{f(L_{n+1}(X_1) - 1)\},$$

so that, under  $P_\pi$ ,  $L_n$  has the same distribution as  $L_{n+1}(X_1) - 1$ .

### 3. Finite Markov chains

In this section we provide a bound on  $P_\mu\{L_n = r\}$  in the context of an ergodic Markov chain on a finite state space. This result is interesting in itself, and it will be important in the following because such models will drive our regime-switching model. Let  $X = (X_n)_{n \geq 1}$  be an irreducible and aperiodic Markov chain with finite state space  $A$ , transition matrix  $Q$ , and stationary distribution  $\pi = (\pi_a)_{a \in A}$ . This implies that there exists an integer  $t_0 \geq 1$  such that

$$Q^{t_0}(a, b) > 0 \quad \text{for all } a, b \in A. \tag{7}$$

From (6), it follows that  $\hat{Q}^{t_0}(a, b) > 0$  for all  $a, b \in A$ . Let

$$\ell = \min_{a,b} Q^{t_0}(a, b), \quad \hat{\ell} = \min_{a,b} \hat{Q}^{t_0}(a, b), \quad \text{and } \lambda = |A| \min\{\ell, \hat{\ell}\}, \tag{8}$$

where  $|A|$  is the cardinality of  $A$ . Note that  $0 < \lambda \leq 1$ , and, for each  $a \in A$ ,

$$Q^{t_0}(a, \cdot) \geq \lambda u(\cdot) \quad \text{and} \quad \hat{Q}^{t_0}(a, \cdot) \geq \lambda u(\cdot), \tag{9}$$

where  $u$  is the uniform distribution on  $A$ . By Theorem 8 of [22], this implies that, for every  $a \in A$ ,

$$\max_{BCA} |Q^n(a, B) - \pi(B)| \leq (1 - \lambda)^{n/t_0 - 1}, \quad n = 1, 2, \dots$$

This result continues to hold if  $Q$  is replaced by  $\hat{Q}$ . In this context, Theorem 2 of [10] gives the following concentration inequality for  $L_n(a)$ .

**Lemma 3.1.** *If  $\lambda > 0$  and  $t_0$  are such that (9) holds, then, for any  $a \in A$ , any  $\gamma > 0$ , and any initial distribution  $\mu$ , we have*

$$P_\mu \{L_n(a) - E_\pi[L_n(a)] \geq n\gamma\} \vee P_\mu \{L_n(a) - E_\pi[L_n(a)] \leq -n\gamma\} \leq \exp\left(-\frac{n}{2}\left(\frac{\lambda\gamma}{t_0} - \frac{2}{n}\right)^2\right)$$

for  $n > \frac{2t_0}{\lambda\gamma}$ .

Clearly, the above holds for both the chain  $X$  and the reversed chain  $\hat{X}$ . Similar concentration inequalities can be obtained by applying Corollary 2.10 and Remark 2.11 of [20]. Combining this with Lemma 2.1 gives the following.

**Proposition 3.1.** *For  $n > \frac{2t_0 + r\lambda + \lambda(1 - \pi_\wedge)}{\lambda\pi_\wedge}$  and any initial distribution  $\mu = (\mu_a)_{a \in A}$ , we have*

$$P_\mu \{L_n = r\} \leq P_\mu \{L_n \leq r\} \leq C \exp\left(-\frac{n}{2}\left(\frac{\lambda\pi_\wedge}{t_0} - \frac{2 + (r + 1)\lambda/t_0}{n}\right)^2\right),$$

where  $t_0$  and  $\lambda$  are as above,  $\pi_\wedge = \min_{a \in A} \pi_a$ , and  $C = |A| \wedge \max_{a \in A} (\mu_a/\pi_a)$ .

In particular, note that when  $\mu = \pi$  the constant  $C = 1$ . It is straightforward to check that the asymptotic behavior of the upper bound is given by

$$C \exp\left(-\frac{n}{2}\left(\frac{\lambda\pi_\wedge}{t_0} - \frac{2 + (r + 1)\lambda/t_0}{n}\right)^2\right) \sim C' \exp\left(-n \frac{\lambda^2 \pi_\wedge^2}{2t_0^2}\right) \quad \text{as } n \rightarrow \infty,$$

where  $C' = C \exp(t_0^{-2} \lambda \pi_\wedge (2t_0 + (r + 1)\lambda))$ .

**Remark 3.1.** It may be interesting to note that Proposition 3.1 gives a bound with exponential decay. This holds, in particular, for the special case where  $X_1, X_2, \dots$  are i.i.d. random variables. In comparison, Corollary 2.1 of [7] focuses on the i.i.d. case and only gives the bound

$$P_\mu \{L_n = r\} \leq c(r) \frac{|A|}{n}, \quad 0 \leq r \leq n - 1,$$

where  $c(r)$  is given by (4).

The proof of Proposition 3.1 depends heavily on the assumption of a finite alphabet. While concentration inequalities for the local times of Markov chains in the case of infinite alphabets are well known and can be found in, e.g., [10, 20], there does not appear to be a simple way to transform these into bounds on  $P_\mu \{L_n = r\}$ . The issue comes from the fact that we need  $\pi_\wedge > 0$ , but it is always zero when  $A$  is an infinite set. An interesting situation, where we are able to deal with infinite alphabets, is the regime-switching model. This is the focus of the remainder of this paper.

### 4. The regime-switching model

This section formally introduces the regime-switching model and extends the finite-sample bounds given in Lemma 1.1 to this case. While we are primarily interested in the case where transitions between alphabets follow an ergodic Markov chain on a finite state space, our presentation is given in more generality. Let  $A$  be a finite or countably infinite set. For each  $a \in A$ , let  $P_a = (p_{a,k})_{k \geq 1}$  be a probability distribution on  $\mathbb{N}_+$ . The finite-dimensional distributions of any discrete-time stochastic process  $(Y_n)_{n \geq 1}$  on  $A$  can be described by a family of conditional distributions  $R = (R_n)_{n \geq 1}$ , where  $R_1(a) = P(Y_1 = a)$  and, for  $n \geq 2$ ,

$$R_n(a_n | a_1, a_2, \dots, a_{n-1}) = P(Y_n = a_n | Y_1 = a_1, Y_2 = a_2, \dots, Y_{n-1} = a_{n-1}).$$

In the case where  $P(Y_1 = a_1, Y_2 = a_2, \dots, Y_{n-1} = a_{n-1}) = 0$ , we will take  $R_n(\cdot | a_1, a_2, \dots, a_{n-1})$  to be an arbitrary probability measure on  $A$ .

We now introduce a process on the state space  $\mathcal{A} := A \times \mathbb{N}_+$  defined by the family of conditional distributions given by  $\mathcal{R} = (\mathcal{R}_n)_{n \geq 1}$ , where  $\mathcal{R}_1$  satisfies  $\mathcal{R}_1((a, k)) = R_1(a)p_{a,k}$  and, for  $n \geq 2$ ,

$$\mathcal{R}_n((a_n, k_n) | (a_1, k_1), (a_2, k_2), \dots, (a_{n-1}, k_{n-1})) = R_n(a_n | a_1, a_2, \dots, a_{n-1})p_{a_n, k_n}.$$

Now, let  $Z = (Z_n)_{n \geq 1}$  be an  $\mathcal{A}$ -valued stochastic process governed by  $(\mathcal{R}_n)_{n \geq 1}$ , and let  $X = (X_n)_{n \geq 1}$  and  $K = (K_n)_{n \geq 1}$  denote the first and second coordinate processes of  $Z$ , i.e.

$$Z_n = (X_n, K_n), \quad n \geq 1.$$

We will refer to the process  $X$  as the underlying process. Note, in particular, that  $X$  is  $A$ -valued, while  $K$  takes values in  $\mathbb{N}_+$ . The next result gives a more explicit description of the dynamics of the processes  $X$  and  $K$ .

**Lemma 4.1.** *In the above context, the following statements hold:*

- (1) *The finite-dimensional distributions of the process  $(X_n)_{n \geq 1}$  are determined by  $(R_n)_{n \geq 1}$ .*
- (2) *For all  $n \geq 1$  and for all  $k \geq 1$ , with probability 1,*

$$P\{K_n = k | X_1, \dots, X_n\} = p_{X_n, k},$$

where  $p_{X_n, k}$  is the random variable equal to  $p_{a, k}$  on the event  $\{X_n = a\}$ .

- (3) *Conditionally on the variables  $X_1, \dots, X_n$ , the variables  $K_1, \dots, K_n$  are independent. In particular, for all  $i = 1, \dots, n$  and all  $k \geq 1$ , with probability 1,*

$$P\{K_i = K_{n+1} | X_1, \dots, X_{n+1}, K_{n+1}\} = p_{X_i, K_{n+1}},$$

where  $p_{X_i, K_{n+1}}$  is the random variable equal to  $p_{a, k}$  on the event  $\{X_i = a, K_{n+1} = k\}$ .

**Remark 4.1.** We are motivated by the case where  $R = (R_n)_{n \geq 1}$  represents the conditional distributions of a Markov chain with transition operator  $Q$  and initial distribution  $\eta$ . In this case, we have  $R_1 = \eta$  and, for  $n \geq 2$ ,

$$R_n(a_n | a_1, a_2, \dots, a_{n-1}) = R_2(a_n | a_{n-1}) = Q(a_{n-1}, a_n).$$

It follows that, in this case,  $\mathcal{R}_1((a_1, k_1)) = \eta(a_1)p_{a_1, k_1}$  and, for  $n \geq 2$ ,

$$\mathcal{R}_n((a_n, k_n) | (a_1, k_1), (a_2, k_2), \dots, (a_{n-1}, k_{n-1})) = Q(a_{n-1}, a_n)p_{a, k},$$

which is the Markov operator denoted by  $\mathcal{Q}$  in (5). In this case, to emphasize the dependence on the initial distribution we will write  $P_\eta$  for  $P$  and  $E_\eta$  for  $E$ . It should be noted that the subscript refers to the initial distribution of the underlying process  $X$  and not of  $Z$ .

Our next result establishes a link between the quantities

$$\mathcal{L}_n = \sum_{i=1}^n \mathbf{1}\{Z_i = Z_{n+1}\} \quad \text{and} \quad L_n = \sum_{i=1}^n \mathbf{1}\{X_i = X_{n+1}\}.$$

**Lemma 4.2.** *For all  $n \geq 1$  and all  $0 \leq r \leq n$ ,*

$$P\{\mathcal{L}_n = r\} = E \left[ \binom{L_n}{r} \sum_{k=1}^{+\infty} p_{X_{n+1},k}^{1+r} (1 - p_{X_{n+1},k})^{L_n-r} \right],$$

where we take  $\binom{L_n}{r} = 0$  when  $L_n < r$ .

A slight modification of Lemma 4.2 brings us to the main result of this section, which extends Lemma 1.1 from the i.i.d. case to the regime-switching case. First, we introduce some notation. For all  $a \in A$ , we write  $v(a, \cdot)$  to denote the counting function of  $P_a = (p_{a,k})_{k \in \mathbb{N}_+}$ , which is defined, for all  $0 \leq \varepsilon \leq 1$ , by

$$v(a, \varepsilon) = \sum_{k \geq 1} \mathbf{1}\{p_{a,k} \geq \varepsilon\}. \tag{10}$$

**Theorem 4.1.** *For any  $n \geq 1$  and any  $0 \leq r \leq n - 1$ , we have*

$$P\{\mathcal{L}_n = r\} \leq P\{L_n = r\} \sup_{a \in A} \sum_{k=1}^{+\infty} p_{a,k}^{1+r} + \inf_{0 \leq \varepsilon \leq 1} \{a^{n,r}(\varepsilon) + b^{n,r}(\varepsilon)\}, \tag{11}$$

where

$$a^{n,r}(\varepsilon) = c(r) E \left[ \mathbf{1}\{L_n > r\} \frac{v(X_{n+1}, \varepsilon)}{L_n} \right],$$

$$b^{n,r}(\varepsilon) = 2^{1+r} \int_0^\varepsilon u^r E \left[ \mathbf{1}\{L_n > r\} v \left( X_{n+1}, \frac{u}{2} \right) \binom{L_n}{r} \left( 1 - \frac{u}{2} \right)^{L_n-r} \right] du,$$

and where  $c(r)$  is as in (4).

Since the formulation of Theorem 4.1 is quite general, an explicit evaluation of the coefficients  $a^{n,r}(\varepsilon)$  and  $b^{n,r}(\varepsilon)$  can require cumbersome computations. More tractable formulas can be provided in a number of situations. We give several examples.

**Example 4.1.** Consider the situation where all distributions  $P_a = (p_{a,k})_{k \geq 1}$  are equal to the same distribution  $P = (p_k)_{k \geq 1}$ , and therefore all counting functions  $v(a, \cdot)$  are equal to the counting function  $v$  of  $P$ . In this scenario, an elementary reordering of the terms in (11) yields that, for any  $\varepsilon \in [0, 1]$ ,

$$P\{\mathcal{L}_n = r\} \leq \sum_{m=r}^n C_{r,m}(\varepsilon) P\{L_n = m\},$$

where

$$C_{r,r}(\varepsilon) = C_{r,r} = \sum_{k=1}^{+\infty} p_k^{1+r}$$



and, for  $1 + r \leq m \leq n$ ,

$$C_{r,m}(\varepsilon) = \frac{c(r)v(\varepsilon)}{m} + 2^{1+r} \binom{m}{r} \int_0^\varepsilon u^r v\left(\frac{u}{2}\right) \left(1 - \frac{u}{2}\right)^{m-r} du,$$

where  $c(r)$  is as in (4).

**Example 4.2.** Another favorable scenario corresponds to the case where all probabilities  $P_a = (p_{a,k})_{k \geq 1}$  have support contained in  $\{1, \dots, M\}$  for some  $M < +\infty$  independent of  $a \in A$ , i.e.

$$p_{a,k} = 0 \quad \text{for } a \in A \text{ and } k \geq M + 1.$$

In this case, taking  $\varepsilon = 0$  on the right-hand side of (11), and noticing that  $v(a, 0)$  corresponds to the size of the support of  $P_a$ , yields

$$P\{\mathcal{L}_n = r\} \leq \sum_{m=r}^n C'_{r,m} P\{L_n = m\},$$

where

$$C'_{r,r} = \sup_{a \in A} \sum_{k=1}^M p_{a,k}^{1+r} \quad \text{and} \quad C'_{r,m} = \frac{c(r)M}{m} \quad \text{for } 1 + r \leq m \leq n$$

and where  $c(r)$  is as in (4).

We now turn to the important situation where the distribution is regularly varying. In the i.i.d. case, the corresponding result is given in Corollary 2.2 of [7].

**Proposition 4.1.** Assume that, for some  $\alpha \in [0, 1]$  and some non-increasing function  $\ell \in \mathcal{SV}$ , we have

$$v(a, \varepsilon) \leq \varepsilon^{-\alpha} \ell(1/\varepsilon)$$

for all  $a \in A$  and all  $\varepsilon \in (0, 1]$ . In this case,

$$P\{\mathcal{L}_n = r\} \leq c_1(\alpha, r) E[\mathbf{1}\{L_n > r\} L_n^{-(1-\alpha)} \ell(L_n)] + c_2(\alpha, r) P\{L_n = r\},$$

where

$$c_1(\alpha, r) = c(r) + \frac{4^{1+r}}{r!} (1+r)^{1+r-\alpha} \gamma\left(1+r-\alpha, \frac{1}{2}\right),$$

$$c_2(\alpha, r) = \begin{cases} 1 & r = 0, \\ \min\{1, p_\vee^{r+1} r^\alpha \ell(r) + r^{-r}\} & r \geq 1, \end{cases}$$

$p_\vee = \sup\{p_{a,k}\}_{(a,k) \in A}$ , and  $\gamma(t, x) = \int_0^x u^{t-1} e^{-u} du$  is the incomplete gamma function.

**Remark 4.2.** Note that in the case  $\alpha = 1$  and  $r = 0$ , the bound in Proposition 4.1 is trivial since it involves  $\gamma\left(0, \frac{1}{2}\right) = +\infty$ . Even in the i.i.d. case, the bounds given in [7] are not able to deal with this case.

**Remark 4.3.** Note that Theorem 4.1 and Proposition 4.1 are quite general and hold no matter what the underlying process is. However, this generality has a cost. In particular, we still need to know quite a bit about the underlying process. In the case where the underlying process is a

finite state space ergodic Markov chain, we can use Proposition 3.1 and related results to get more explicit formulas.

**Corollary 4.1.** *Assume that  $|A| < \infty$  and that the underlying process is an aperiodic and irreducible Markov chain with transition operator  $Q$ , stationary distribution  $\pi = (\pi_a)_{a \in A}$ , and initial distribution  $\eta$ . Let  $\pi_\wedge = \min_{a \in A} \pi_a$ , let  $t_0$  be as in (7), and let  $\lambda$  be as in (8). Assume further that, for some  $\alpha \in [0, 1]$  and some non-increasing function  $\ell \in \mathcal{SV}$ , we have*

$$v(a, \epsilon) \leq \epsilon^{-\alpha} \ell(1/\epsilon), \quad a \in A, \quad \epsilon \in (0, 1].$$

For any  $\epsilon \in (0, \pi_\wedge)$ , if  $n > \frac{2t_0+r\lambda+\lambda(1-\pi_\wedge)}{\lambda\pi_\wedge} \vee \frac{2t_0+\lambda(1-\pi_\wedge)}{\lambda(\pi_\wedge-\epsilon)}$ , then

$$P_\eta\{\mathcal{L}_n = r\} \leq H(n, \epsilon),$$

where

$$H(n, \epsilon) = c_1(\alpha, r)(n\epsilon)^{-(1-\alpha)}\ell(n\epsilon) + c_2(\alpha, r)C \exp\left(-\frac{n}{2}\left(\frac{\lambda\pi_\wedge}{t_0} - \frac{2+(r+1)\lambda/t_0}{n}\right)^2\right) + c_3(\alpha, r)C \exp\left(-\frac{n}{2}\left(\frac{\lambda(\pi_\wedge - \epsilon)}{t_0} - \frac{2+\lambda/t_0}{n}\right)^2\right).$$

Here,  $C$  is as in Proposition 3.1,  $c_1(\alpha, r)$  and  $c_2(\alpha, r)$  are as in Proposition 4.1, and

$$c_3(\alpha, r) = c_1(\alpha, r)(r+1)^{-(1-\alpha)}\ell(r+1).$$

It may be interesting to note that for any  $\epsilon \in (0, \pi_\wedge)$  we have

$$H(n, \epsilon) \sim \epsilon^{-(1-\alpha)}c_1(\alpha, r)n^{-(1-\alpha)}\ell(n) \quad \text{as } n \rightarrow \infty.$$

### 5. Asymptotics for the regime-switching model

In this section we extend (3) from the i.i.d. case to the case of the regime-switching model, where the underlying process is an ergodic Markov chain on a finite state space. We first define regular variation of  $P = (p_{a,k})_{(a,k) \in A \times \mathbb{N}_+}$ . For a review of basic facts about regularly varying distributions on  $\mathbb{N}_+$  we refer the reader to Appendix A.

**Definition 5.1.** We say that  $P = (p_{a,k})_{(a,k) \in A \times \mathbb{N}_+}$  is regularly varying with index  $\alpha \in [0, 1]$  if there exists an  $\ell \in \mathcal{SV}$  and a function  $C : A \mapsto [0, \infty)$ , which is not identically zero, such that, for each  $a \in A$ ,

$$\lim_{\epsilon \rightarrow 0} \frac{v(a, \epsilon)}{\epsilon^{-\alpha} \ell(1/\epsilon)} = C(a),$$

where  $v$  is defined as in (10). In this case we write  $P \in RV_\alpha(C, \ell)$ .

**Remark 5.1.** It is important to note that we allow  $C(a) = 0$  for some (but not all)  $a \in A$ . When  $C(a) = 0$ , it means that  $v(a, \epsilon)$  either does not approach infinity as  $\epsilon \rightarrow 0$ , or approaches it but at a rate that is slower than  $\epsilon^{-\alpha} \ell(1/\epsilon)$ . In particular, if  $P \in RV_\alpha(C, \ell)$  and  $v(a_1, \epsilon) = \epsilon^{-\alpha_1} \ell_1(1/\epsilon)$  for some  $a_1 \in A$ ,  $\alpha_1 \in [0, \alpha)$ , and  $\ell_1 \in \mathcal{SV}$ , then  $C(a_1) = 0$ .

When  $\alpha = 0$ , we additionally assume that there exists an  $\ell_0 \in \mathcal{SV}$  and a function  $D : A \mapsto [0, \infty)$ , which is not identically zero, such that, for each  $a \in A$ ,

$$\lim_{\epsilon \rightarrow 0} \frac{\sum_{k \geq 1} p_{a,k} \mathbf{1}\{p_{a,k} \leq \epsilon\}}{\epsilon \ell_0(1/\epsilon)} = D(a). \tag{12}$$

For simplicity of notation, for  $x > 0$  set

$$h_{\alpha,r}(x) = \begin{cases} x^{-1}\ell_0(x) & \alpha = 0; \\ \int_x^\infty u^{-1}\ell(u)du & \alpha = 1, r = 0; \\ x^{-(1-\alpha)}\ell(x) & \text{otherwise.} \end{cases}$$

Propositions A.1 and A.2 imply that if  $P \in RV_\alpha(C, \ell)$  then

$$\lim_{n \rightarrow \infty} \frac{\binom{n}{r} \sum_{k=1}^\infty p_{a,k}^{r+1} (1 - p_{a,k})^{n-r}}{h_{\alpha,r}(n)} = F(a, r), \tag{13}$$

where

$$F(a, r) = \begin{cases} D(a) & \alpha = 0; \\ C(a) & \alpha = 1, r = 0; \\ C(a) \frac{\alpha \Gamma(r+1-\alpha)}{r!} & \text{otherwise.} \end{cases} \tag{14}$$

Note that since  $|A| < \infty$  the convergence in (13) is uniform in  $a$ . We now give the main result for this section.

**Theorem 5.1.** *In the context of the regime-switching model, assume that  $|A| < \infty$  and that the underlying process is an aperiodic and irreducible Markov chain with stationary distribution  $\pi = (\pi_a)_{a \in A}$  and initial distribution  $\eta$ . Assume further that  $P \in RV_\alpha(C, \ell)$  with  $\alpha \in [0, 1]$  (when  $\alpha = 0$  additionally assume that (12) holds), and that  $\ell$  (or  $\ell_0$  when  $\alpha = 0$ ) is locally bounded away from 0 and  $\infty$  on  $[1, \infty)$ . In this case, for all  $r \geq 0$  we have*

$$\lim_{n \rightarrow \infty} \frac{P_\eta\{\mathcal{L}_n = r\}}{h_{\alpha,r}(n)} = \sum_{a \in A} \pi_a^\alpha F(a, r)$$

and

$$\lim_{n \rightarrow \infty} \frac{P_\eta\{\mathcal{L}_n = r\}}{E_\eta[\mathbf{1}\{L_n > r\}h_{\alpha,r}(L_n)]} = \frac{\sum_{a \in A} \pi_a^\alpha F(a, r)}{\sum_{a \in A} \pi_a^\alpha},$$

where  $F$  is given by (14).

This implies that for  $\alpha \in (0, 1)$  we have, up to a constant, the same asymptotics as for the upper bound in Corollary 4.1. It may be interesting to note that as part of the proof of the theorem we show that, for any  $r \geq 0$ ,

$$\lim_{n \rightarrow \infty} \frac{E_\eta[\mathbf{1}\{L_n > r\}h_{\alpha,r}(L_n)]}{h_{\alpha,r}(n)} = \sum_{a \in A} \pi_a^\alpha.$$

## 6. Proofs

### 6.1. Proofs for Sections 2 and 3

*Proof of Lemma 2.1.* Let us first prove that, for any distributions  $\mu$  and  $\eta$  on  $A$  and any bounded function  $g : A^{n+1} \rightarrow \mathbb{R}_+$ ,

$$E_\mu \left[ \frac{\eta(X_{n+1})}{\pi(X_{n+1})} g(X_1, \dots, X_{n+1}) \right] = E_\eta \left[ \frac{\mu(\hat{X}_{n+1})}{\pi(\hat{X}_{n+1})} g(\hat{X}_{n+1}, \dots, \hat{X}_1) \right], \tag{15}$$

where, on the right-hand side, it is understood that  $\eta$  is taken as the initial distribution of the reversed chain. From the definition of  $\hat{Q}$ , we obtain

$$\begin{aligned} & E_{\mu} \left[ \frac{\eta(X_{n+1})}{\pi(X_{n+1})} g(X_1, \dots, X_{n+1}) \right] \\ &= \sum_{x_1, \dots, x_{n+1}} \frac{\eta(x_{n+1})}{\pi(x_{n+1})} g(x_1, \dots, x_{n+1}) P_{\mu}(X_1 = x_1, \dots, X_{n+1} = x_{n+1}) \\ &= \sum_{x_1, \dots, x_{n+1}} \frac{\eta(x_{n+1})}{\pi(x_{n+1})} g(x_1, \dots, x_{n+1}) \mu(x_1) Q(x_1, x_2) \cdots Q(x_n, x_{n+1}) \\ &= \sum_{x_1, \dots, x_{n+1}} \frac{\eta(x_{n+1})}{\pi(x_{n+1})} g(x_1, \dots, x_{n+1}) \mu(x_1) \frac{\pi(x_2)}{\pi(x_1)} \hat{Q}(x_2, x_1) \cdots \frac{\pi(x_{n+1})}{\pi(x_n)} \hat{Q}(x_{n+1}, x_n) \\ &= \sum_{x_1, \dots, x_{n+1}} \frac{\mu(x_1)}{\pi(x_1)} g(x_1, \dots, x_{n+1}) \eta(x_{n+1}) \hat{Q}(x_{n+1}, x_n) \cdots \hat{Q}(x_2, x_1) \\ &= \sum_{x_1, \dots, x_{n+1}} \frac{\mu(x_1)}{\pi(x_1)} g(x_1, \dots, x_{n+1}) P_{\eta}(\hat{X}_1 = x_{n+1}, \dots, \hat{X}_{n+1} = x_1) \\ &= E_{\eta} \left[ \frac{\mu(\hat{X}_{n+1})}{\pi(\hat{X}_{n+1})} g(\hat{X}_{n+1}, \dots, \hat{X}_1) \right], \end{aligned}$$

which proves (15). Then, for any positive  $f$ ,

$$\begin{aligned} E_{\mu} \left[ \frac{\eta(X_{n+1})}{\pi(X_{n+1})} f(L_n) \right] &= E_{\mu} \left[ \frac{\eta(X_{n+1})}{\pi(X_{n+1})} f \left( \sum_{i=1}^n \mathbf{1}\{X_i = X_{n+1}\} \right) \right] \\ &= E_{\eta} \left[ \frac{\mu(\hat{X}_{n+1})}{\pi(\hat{X}_{n+1})} f \left( \sum_{i=2}^{n+1} \mathbf{1}\{\hat{X}_i = \hat{X}_1\} \right) \right] \\ &= E_{\eta} \left[ \frac{\mu(\hat{X}_{n+1})}{\pi(\hat{X}_{n+1})} f \left( \sum_{i=1}^{n+1} \mathbf{1}\{\hat{X}_i = \hat{X}_1\} - 1 \right) \right] \\ &= E_{\eta} \left[ \frac{\mu(\hat{X}_{n+1})}{\pi(\hat{X}_{n+1})} f(\hat{L}_{n+1}(\hat{X}_1) - 1) \right], \end{aligned}$$

where the second line follows by applying identity (15) with

$$g(x_1, \dots, x_{n+1}) = f \left( \sum_{i=1}^n \mathbf{1}\{x_i = x_{n+1}\} \right).$$

This completes the proof. □

*Proof of Proposition 3.1.* Fix  $r \geq 0$  and observe that the assumption on  $n$  implies that  $\pi_\wedge > \frac{r}{n}$ . As a result, since  $E_\pi L_n(a) = n\pi_a$ , we deduce from Lemma 3.1 that

$$\begin{aligned} P_\mu\{L_n(a) \leq r\} &= P_\mu\{L_n(a) - n\pi_a \leq -n(\pi_a - r/n)\} \\ &\leq P_\mu\{L_n(a) - n\pi_a \leq -n(\pi_\wedge - r/n)\} \\ &\leq \exp\left(-\frac{n}{2}\left(\frac{\lambda\pi_\wedge}{t_0} - \frac{2+r\lambda/t_0}{n}\right)^2\right) \end{aligned}$$

when  $n > 2t_0/(\lambda(\pi_\wedge - r/n))$ , which is equivalent to  $n > (2t_0 + r\lambda)/(\lambda\pi_\wedge)$ . From here, we provide two bounds on  $P_\mu\{L_n \leq r\}$ , which, when combined, give the desired result. First, note that

$$\begin{aligned} P_\mu\{L_n \leq r\} &= \sum_{a \in A} P_\mu\{X_{n+1} = a, L_n \leq r\} \\ &\leq \sum_{a \in A} P_\mu\{L_n(a) \leq r\} \\ &\leq |A| \exp\left(-\frac{n}{2}\left(\frac{\lambda\pi_\wedge}{t_0} - \frac{2+r\lambda/t_0}{n}\right)^2\right). \end{aligned} \tag{16}$$

Next, using Lemma 2.1 with  $f(u) = \mathbf{1}\{u \leq r\}$ , it follows that

$$\begin{aligned} P_\mu\{L_n \leq r\} &= E_\pi\left[\frac{\mu(\hat{X}_{n+1})}{\pi(\hat{X}_{n+1})} \mathbf{1}\{\hat{L}_{n+1}(\hat{X}_1) \leq r + 1\}\right] \\ &\leq \max_{b \in A} \frac{\mu(b)}{\pi(b)} P_\pi\{\hat{L}_{n+1}(\hat{X}_1) \leq r + 1\} \\ &= \max_{b \in A} \frac{\mu(b)}{\pi(b)} \sum_{a \in A} \pi(a) P_a\{\hat{L}_{n+1}(a) \leq r + 1\}, \end{aligned}$$

where  $P_a$  is the probability measure that corresponds to the case where the initial distribution is a point mass at  $a$ . Hence, once again using Lemma 3.1 and the fact that the stationary distribution of the reversed chain is the same as for the original chain, it follows that

$$P_\mu\{L_n \leq r\} \leq \max_{b \in A} \frac{\mu b}{\pi_b} \exp\left(-\frac{n+1}{2}\left(\frac{\lambda\pi_\wedge}{t_0} - \frac{2+(r+1)\lambda/t_0}{n+1}\right)^2\right), \tag{17}$$

provided  $n+1 > (2t_0 + (r+1)\lambda)/(\lambda\pi_\wedge)$ , or equivalently  $n > (2t_0 + r\lambda + \lambda(1 - \pi_\wedge))/(\lambda\pi_\wedge)$ . The desired result follows by combining (16) and (17).  $\square$

### 6.2. Proofs for Section 4

For convenience, we sometimes denote  $Y_{1 \rightarrow m} = (Y_1, \dots, Y_m)$  for a given process  $(Y_n)_{n \geq 1}$ . *Proof of Lemma 4.1.*

- (1) The statement follows easily from the structure of  $\mathcal{R}$ . Let  $p_1$  and  $p_2$  be the functions defined, for  $(a, k) \in \mathcal{A}$ , by  $p_1(a, k) = a$  and  $p_2(a, k) = k$ . We have

$$P(X_1 = a) = \sum_{k \geq 1} P(Z_1 = (a, k)) = \sum_{k \geq 1} \mathcal{R}_1((a, k)) = \sum_{k \geq 1} R_1(a)p_{a,k} = R_1(a).$$

Further, for any  $n \geq 1$  and any bounded  $f : A \mapsto \mathbb{R}$ ,

$$E[f(X_{n+1}) | X_{1 \rightarrow n}] = E[E[f \circ p_1(Z_{n+1}) | Z_{1 \rightarrow n}] | X_{1 \rightarrow n}].$$

From here, the fact that

$$\begin{aligned} E[f \circ p_1(Z_{n+1}) | Z_{1 \rightarrow n}] &= \sum_{a \in A} \sum_{k \in \mathbb{N}_+} \mathcal{R}_{n+1}((a, k) | Z_1, Z_2, \dots, Z_n) f \circ p_1(a, k) \\ &= \sum_{a \in A} \sum_{k \geq 1} R_{n+1}(a | X_1, X_2, \dots, X_n) p_{a,k} f(a) \\ &= \sum_{a \in A} R_{n+1}(a | X_1, X_2, \dots, X_n) f(a) \end{aligned}$$

implies that

$$E[f(X_{n+1}) | X_{1 \rightarrow n}] = \sum_{a \in A} R_{n+1}(a | X_1, X_2, \dots, X_n) f(a).$$

In particular, taking  $f(a) = \mathbf{1}\{a = a'\}$  gives

$$P(X_{n+1} = a' | X_{1 \rightarrow n}) = R_{n+1}(a' | X_1, X_2, \dots, X_n),$$

which proves the claim.

(2) For all  $n \geq 2$ , all  $k_n \geq 1$ , and all  $a_1, \dots, a_n \in A$  satisfying  $P\{X_1 = a_1, \dots, X_n = a_n\} > 0$ ,

$$P\{K_n = k_n | X_1 = a_1, \dots, X_n = a_n\} = \sum_{k_1, \dots, k_{n-1}} \frac{P\{Z_n = (a_n, k_n), \dots, Z_1 = (a_1, k_1)\}}{P\{X_1 = a_1, \dots, X_n = a_n\}}. \quad (18)$$

Using point (1) it follows that

$$P\{X_1 = a_1, \dots, X_n = a_n\} = R_1(a_1)R_2(a_2 | a_1) \cdots R(a_n | a_1, a_2, \dots, a_{n-1}),$$

and that

$$\begin{aligned} P\{Z_n = (a_n, k_n), \dots, Z_1 = (a_1, k_1)\} &= \\ &R_1(a_1)R_2(a_2 | a_1) \cdots R(a_n | a_1, a_2, \dots, a_{n-1}) p_{a_1, k_1} p_{a_2, k_2} \cdots p_{a_n, k_n}. \end{aligned}$$

Combining these two identities with (18), we deduce that

$$\begin{aligned} P\{K_n = k_n | X_1 = a_1, \dots, X_n = a_n\} &= p_{a_n, k_n} \sum_{k_1, \dots, k_{n-1}} p_{a_1, k_1} p_{a_2, k_2} \cdots p_{a_{n-1}, k_{n-1}} \\ &= p_{a_n, k_n}, \end{aligned}$$

where the last identity follows from the fact that  $\sum_k p_{a,k} = 1$ . The case where  $n = 1$  is similar.

(3) For any  $k_1, \dots, k_n \in \mathbb{N}_+$  and any  $a_1, \dots, a_n \in A$  satisfying  $P\{X_1 = a_1, \dots, X_n = a_n\} > 0$ ,

$$P\{K_1 = k_1, \dots, K_n = k_n \mid X_1 = a_1, \dots, X_n = a_n\} = \prod_{i=1}^n p_{a_i, k_i} = \prod_{i=1}^n P\{K_i = k_i \mid X_1 = a_1, \dots, X_i = a_i\},$$

where the first identity follows by arguments similar to those used in the proof of point (2) and the second follows directly from point (2). Finally, the proof that, for  $i = 1, 2, \dots, n$ ,

$$P\{K_i = K_{n+1} \mid X_1, \dots, X_{n+1}, K_{n+1}\} = p_{X_i, K_{n+1}}$$

is very similar and is omitted for brevity. □

*Proof of Lemma 4.2.* Fix  $n \geq 1$  and  $0 \leq r \leq n$ . Since  $\{\mathcal{L}_n = r\} \subset \{L_n \geq r\}$ , we have

$$P\{\mathcal{L}_n = r\} = P\{L_n \geq r, \mathcal{L}_n = r\}.$$

Noticing that the variable  $L_n$  is  $\sigma(X_{1 \rightarrow n+1})$ -measurable by construction, we obtain

$$\begin{aligned} P\{\mathcal{L}_n = r\} &= P\left\{L_n \geq r, \sum_{i=1}^n \mathbf{1}\{X_i = X_{n+1}, K_i = K_{n+1}\} = r\right\} \\ &= E\left[\mathbf{1}\{L_n \geq r\} P\left(\sum_{i=1}^n \mathbf{1}\{X_i = X_{n+1}, K_i = K_{n+1}\} = r \mid K_{n+1}, X_{1 \rightarrow n+1}\right)\right]. \end{aligned}$$

Conditionally on  $K_{n+1}$  and  $X_{1 \rightarrow n+1}$  the variables  $K_1, \dots, K_n$  are, according to point (3) of Lemma 4.1, independent and satisfy

$$P\{K_i = K_{n+1} \mid X_{1 \rightarrow n+1}, K_{n+1}\} = p_{X_i, K_{n+1}}.$$

As a result, conditionally on  $K_{n+1}$  and  $X_{1 \rightarrow n+1}$ , the variable

$$\sum_{i=1}^n \mathbf{1}\{X_i = X_{n+1}, K_i = K_{n+1}\}$$

follows a binomial distribution with parameters  $L_n$  and  $p_{X_{n+1}, K_{n+1}}$ . Hence, we obtain

$$\begin{aligned} P\{\mathcal{L}_n = r\} &= E\left[\mathbf{1}\{L_n \geq r\} \binom{L_n}{r} p_{X_{n+1}, K_{n+1}}^r (1 - p_{X_{n+1}, K_{n+1}})^{L_n - r}\right] \\ &= E\left[\mathbf{1}\{L_n \geq r\} \binom{L_n}{r} E\left[p_{X_{n+1}, K_{n+1}}^r (1 - p_{X_{n+1}, K_{n+1}})^{L_n - r} \mid X_{1 \rightarrow n+1}\right]\right] \\ &= E\left[\mathbf{1}\{L_n \geq r\} \binom{L_n}{r} \sum_{k \geq 1} p_{X_{n+1}, k}^{1+r} (1 - p_{X_{n+1}, k})^{L_n - r}\right], \end{aligned}$$

where the last line follows from point (2) of Lemma 4.1. □

*Proof of Theorem 4.1.* From Lemma 4.2 it follows that

$$\begin{aligned} P\{\mathcal{L}_n = r\} &= E\left[\mathbf{1}\{L_n = r\} \binom{L_n}{r} \sum_{k \geq 1} p_{X_{n+1},k}^{1+r} (1 - p_{X_{n+1},k})^{L_n-r}\right] \\ &\quad + E\left[\mathbf{1}\{L_n > r\} \binom{L_n}{r} \sum_{k \geq 1} p_{X_{n+1},k}^{1+r} (1 - p_{X_{n+1},k})^{L_n-r}\right] \\ &=: A_1(n) + A_2(n). \end{aligned}$$

Note that

$$A_1(n) = E\left[\mathbf{1}\{L_n = r\} \sum_{k \geq 1} p_{X_{n+1},k}^{1+r}\right] \leq P\{L_n = r\} \sup_{a \in A} \sum_{k=1}^{+\infty} p_{a,k}^{1+r}.$$

Now, using Lemma 1.1 inside the expectation yields

$$A_2(n) \leq E\left[\mathbf{1}\{L_n > r\} \inf_{0 \leq \varepsilon \leq 1} \{\alpha^{n,r}(\varepsilon) + \beta^{n,r}(\varepsilon)\}\right],$$

where we have denoted

$$\begin{aligned} \alpha^{n,r}(\varepsilon) &= \frac{c(r)v(X_{n+1}, \varepsilon)}{L_n}, \\ \beta^{n,r}(\varepsilon) &= 2^{1+r} \binom{L_n}{r} \int_0^\varepsilon v\left(X_{n+1}, \frac{u}{2}\right) u^r \left(1 - \frac{u}{2}\right)^{L_n-r} du. \end{aligned}$$

Finally, observing the fact that

$$\begin{aligned} A_2(n) &\leq E\left[\inf_{0 \leq \varepsilon \leq 1} \{\mathbf{1}\{L_n > r\} \alpha^{n,r}(\varepsilon) + \mathbf{1}\{L_n > r\} \beta^{n,r}(\varepsilon)\}\right] \\ &\leq \inf_{0 \leq \varepsilon \leq 1} \{E[\mathbf{1}\{L_n > r\} \alpha^{n,r}(\varepsilon)] + E[\mathbf{1}\{L_n > r\} \beta^{n,r}(\varepsilon)]\} \\ &\leq \inf_{0 \leq \varepsilon \leq 1} \{a^{n,r}(\varepsilon) + b^{n,r}(\varepsilon)\} \end{aligned}$$

gives the result. □

*Proof of Proposition 4.1.* By Lemma 4.2, we have

$$\begin{aligned} P\{\mathcal{L}_n = r\} &= E\left[\mathbf{1}\{L_n > r\} \binom{L_n}{r} \sum_{k \geq 1} p_{X_{n+1},k}^{1+r} (1 - p_{X_{n+1},k})^{L_n-r}\right] \\ &\quad + E\left[\mathbf{1}\{L_n = r\} \sum_{k \geq 1} p_{X_{n+1},k}^{1+r}\right] \\ &=: E_1 + E_2. \end{aligned}$$

Corollary 2.2 from [7] implies that

$$E_1 \leq c_1(\alpha, r) E[\mathbf{1}\{L_n > r\} L_n^{-(\alpha-1)} \ell(L_n)].$$



From here, the results follows in the case where  $r = 0$  from the fact that  $\sum_{k \geq 1} p_{X_{n+1},k} = 1$ .

Now, assume that  $r \geq 1$ . Taking  $\epsilon = 1/r$  in (2.4) of [7] implies that

$$E_2 \leq \left( p_{\sqrt{r}}^{r+1} r^\alpha \ell(r) + r^{-r} \right) P\{L_n = r\}.$$

On the other hand, since  $\sum_{k \geq 1} p_{X_{n+1},k}^{1+r} \leq \sum_{k \geq 1} p_{X_{n+1},k} = 1$ , we also have

$$E_2 \leq P\{L_n = r\}.$$

This completes the proof. □

*Proof of Corollary 4.1.* Fix  $\epsilon \in (0, \pi_\wedge)$ , let

$$A(n) = \{r + 1 \leq L_n < n\epsilon\} \quad \text{and} \quad B(n) = \{L_n \geq (n\epsilon) \vee (r + 1)\},$$

and note that  $A(n) \cup B(n) = \{r + 1 \leq L_n\}$ . We can write

$$\begin{aligned} E_\eta \left[ \mathbf{1}\{L_n > r\} L_n^{-(1-\alpha)} \ell(L_n) \right] &= E_\eta \left[ \mathbf{1}_{A(n)} L_n^{-(1-\alpha)} \ell(L_n) \right] + E_\eta \left[ \mathbf{1}_{B(n)} L_n^{-(1-\alpha)} \ell(L_n) \right] \\ &= E_1 + E_2. \end{aligned}$$

Now note that

$$E_1 \leq (r + 1)^{-(1-\alpha)} \ell(r + 1) P_\eta\{r + 1 \leq L_n < n\epsilon\}$$

and

$$E_2 \leq (n\epsilon)^{-(1-\alpha)} \ell(n\epsilon) P_\eta\{L_n \geq n\epsilon\}.$$

Combining this with Proposition 4.1 gives

$$\begin{aligned} P_\eta\{\mathcal{L}_n = r\} \leq \inf_{\epsilon \in (0, \pi_\wedge)} \{ &c_1(\alpha, r)(n\epsilon)^{-(1-\alpha)} \ell(n\epsilon) P_\eta\{L_n \geq n\epsilon\} + c_2(\alpha, r) P_\eta\{L_n = r\} \\ &+ c_3(\alpha, r) P_\eta\{r + 1 \leq L_n < n\epsilon\}\}. \end{aligned}$$

From here, the result follows by applying Proposition 3.1. □

### 6.3. Proofs for Section 5

To prove Theorem 5.1, we begin with two technical results.

**Lemma 6.1.** *Let  $(X_n)_{n \geq 1}$  be an irreducible and aperiodic Markov chain on a finite state space  $A$  and with stationary distribution  $\pi = (\pi_a)_{a \in A}$ . Let  $\pi_\wedge = \min_{a \in A} \pi_a$  and let  $L_n = \sum_{k=1}^n \mathbf{1}\{X_k = X_{n+1}\}$ .*

1. *For any  $\beta \in \mathbb{R}$ , any  $\epsilon \in [0, \pi_\wedge)$ , any  $r > 0$ , and any initial distribution  $\eta$ , we have*

$$\lim_{n \rightarrow \infty} n^\beta P_\eta \left\{ \frac{L_n}{n} \leq \epsilon \right\} = 0$$

and

$$\lim_{n \rightarrow \infty} n^\beta P_\eta\{L_n = r\} = 0.$$

2. If  $\alpha \in [0, 1]$  and  $\ell \in SV$ , then, with probability 1,

$$\lim_{n \rightarrow \infty} \left( \frac{L_n^{-(1-\alpha)} \ell(L_n)}{n^{-(1-\alpha)} \ell(n)} - \pi_{X_{n+1}}^{-(1-\alpha)} \right) = 0 \tag{19}$$

and, for any  $r \geq 0$  and any initial distribution  $\eta$ ,

$$\lim_{n \rightarrow \infty} \frac{E_\eta[\mathbf{1}\{L_n > r\} L_n^{-(1-\alpha)} \ell(L_n)]}{n^{-(1-\alpha)} \ell(n)} = \sum_{a \in A} \pi_a^\alpha.$$

*Proof.* The first part follows immediately from the exponential bound in Proposition 3.1.

We now turn to the second part. For ease of notation, set  $h(x) = x^{-(1-\alpha)} \ell(x)$ . Since the Markov chain is irreducible and aperiodic on a finite state space, it is recurrent and hence  $\lim_{n \rightarrow \infty} L_n = \infty$  with probability 1. Further, it satisfies the strong law of large numbers, which means that for each  $a \in A$ , if  $L_n(a) = \sum_{k=1}^n \mathbf{1}\{X_k = a\}$  then  $\lim_{n \rightarrow \infty} L_n(a)/n = \pi_a$  with probability 1. Since  $A$  is a finite set, with probability 1, this convergence can be taken to be uniform in  $a$ . Let  $\Omega_0 \subset \Omega$  with  $P_\eta(\Omega_0) = 1$ , such that for any  $\omega \in \Omega_0$  we have  $\lim_{n \rightarrow \infty} L_n(\omega) = \infty$  and for any  $\epsilon > 0$  there exists an  $N'_\epsilon(\omega)$  such that if  $n \geq N'_\epsilon(\omega)$  then

$$\left| \frac{L_n(\omega)}{n} - \pi_{X_{n+1}(\omega)} \right| < \epsilon.$$

Now fix  $\epsilon > 0$  and  $\omega \in \Omega_0$ . There exists an  $N_\epsilon(\omega) > 0$  such that if  $n \geq N_\epsilon(\omega)$  then  $\left| \frac{L_n(\omega)}{n} - \pi_{X_{n+1}(\omega)} \right| < 0.5\pi_\wedge$  and

$$\left| \pi_{X_{n+1}(\omega)}^{-(1-\alpha)} - \left( \frac{L_n(\omega)}{n} \right)^{-(1-\alpha)} \right| < \epsilon/2.$$

Further, by the uniform convergence theorem for regularly varying functions, see e.g. Proposition 2.4 of [21], there is a  $T_\epsilon$  such that, for any  $x \in (0.5\pi_\wedge, 1]$  and any  $t \geq T_\epsilon$ ,

$$\left| \frac{h(xt)}{h(t)} - x^{-(1-\alpha)} \right| < \epsilon/2.$$

Since  $\frac{L_n}{n} \leq 1$ , it follows that, for  $n \geq \max\{N_\epsilon(\omega), T_\epsilon\}$ ,

$$\begin{aligned} \left| \frac{h\left(\frac{L_n(\omega)}{n} n\right)}{h(n)} - \pi_{X_{n+1}(\omega)}^{-(1-\alpha)} \right| &\leq \left| \frac{h\left(\frac{L_n(\omega)}{n} n\right)}{h(n)} - \left(\frac{L_n(\omega)}{n}\right)^{-(1-\alpha)} \right| \\ &\quad + \left| \pi_{X_{n+1}(\omega)}^{-(1-\alpha)} - \left(\frac{L_n(\omega)}{n}\right)^{-(1-\alpha)} \right| < \epsilon, \end{aligned}$$

which proves (19).

We now turn to the last part. Fix  $\epsilon \in (0, \pi_\wedge)$  and let

$$A(n) = \{r + 1 \leq L_n < n\epsilon\} \quad \text{and} \quad B(n) = \{L_n \geq (n\epsilon) \vee (r + 1)\}.$$

Note that  $A(n) \cup B(n) = \{r + 1 \leq L_n\}$ . We can write

$$\begin{aligned} & \frac{E_\eta[\mathbf{1}\{L_n > r\}L_n^{-(1-\alpha)}\ell(L_n)]}{n^{-(1-\alpha)}\ell(n)} \\ &= \frac{E_\eta[\mathbf{1}_{A(n)}L_n^{-(1-\alpha)}\ell(L_n)]}{n^{-(1-\alpha)}\ell(n)} + \frac{E_\eta[\mathbf{1}_{B(n)}L_n^{-(1-\alpha)}\ell(L_n)]}{n^{-(1-\alpha)}\ell(n)} \\ &=: E_A(n) + E_B(n). \end{aligned}$$

Fix  $\delta > 0$ ; by the Potter bounds (see, e.g., Theorem 1.5.6 of [3]), there exists a  $K_\delta > 0$  such that

$$E_A(n) \leq K_\delta E_\eta \left[ \mathbf{1}_{A(n)} \left( \frac{L_n}{n} \right)^{-(1-\alpha)-\delta} \right] \leq K_\delta P_\eta(A(n))n^{1-\alpha+\delta} \rightarrow 0,$$

where the convergence follows by the first part of this lemma. Similarly,

$$\frac{\mathbf{1}_{B(n)}L_n^{-(1-\alpha)}\ell(L_n)}{n^{-(1-\alpha)}\ell(n)} \leq K_\delta \mathbf{1}_{B(n)} \left( \frac{L_n}{n} \right)^{-(1-\alpha)-\delta} \leq K_\delta \epsilon^{-(1-\alpha)-\delta}.$$

Combining this with the fact that  $\pi_{X_{n+1}}^{-(1-\alpha)}$  is bounded means that we can use dominated convergence to get

$$\begin{aligned} \lim_{n \rightarrow \infty} E_B(n) &= \lim_{n \rightarrow \infty} (E_B(n) + E_\eta[\pi_{X_{n+1}}^{-(1-\alpha)}] - E_\eta[\pi_{X_{n+1}}^{-(1-\alpha)}]) \\ &= E_\eta \left[ \lim_{n \rightarrow \infty} \left( \frac{\mathbf{1}_{B(n)}L_n^{-(1-\alpha)}\ell(L_n)}{n^{-(1-\alpha)}\ell(n)} - \pi_{X_{n+1}}^{-(1-\alpha)} \right) \right] + \lim_{n \rightarrow \infty} E_\eta[\pi_{X_{n+1}}^{-(1-\alpha)}] \\ &= \lim_{n \rightarrow \infty} E_\eta[\pi_{X_{n+1}}^{-(1-\alpha)}] = E_\pi[\pi_{X_1}^{-(1-\alpha)}] = \sum_{a \in A} \pi_a^\alpha, \end{aligned}$$

where the third equality follows from (19) and the fact that, with probability 1, there exists a (random)  $N$  such that  $\mathbf{1}_{B(n)} = 1$  for all  $n \geq N$ , and the fourth equality follows by the fact that the distribution of  $X_n$  converges weakly to  $\pi$ , Skorokhod’s representation theorem, and dominated convergence. □

**Lemma 6.2.** *Let  $|A| < \infty$  and let  $P \in RV_\alpha(C, \ell)$ . When  $\alpha = 0$  assume, in addition, that (12) holds.*

1. Let  $(X_n)_{n \geq 1}$  be any sequence of  $A$ -valued random variables and let  $(N_n)_{n \geq 1}$  be a sequence of  $\mathbb{N}$ -valued random variables such that, with probability 1,  $N_n \rightarrow \infty$  as  $n \rightarrow \infty$ . With probability 1,

$$\lim_{n \rightarrow \infty} \left( \frac{\binom{N_n}{r} \sum_{k=1}^\infty P_{X_{n+1},k}^{r+1} (1 - P_{X_{n+1},k})^{N_n-r}}{h_{\alpha,r}(N_n)} - F(X_{n+1}, r) \right) = 0.$$

2. Let  $X = (X_k)_{k \geq 1}$  be an irreducible and aperiodic Markov chain with state space  $A$  and stationary distribution  $\pi = (\pi_a)_{a \in A}$ . If  $L_n = \sum_{k=1}^n \mathbf{1}\{X_k = X_{n+1}\}$ , then, with probability 1,

$$\lim_{n \rightarrow \infty} \left( \frac{\binom{L_n}{r} \sum_{k=1}^\infty P_{X_{n+1},k}^{1+r} (1 - P_{X_{n+1},k})^{L_n-r}}{h_{\alpha,r}(n)} - \pi_{X_{n+1}}^{-(1-\alpha)} F(X_{n+1}, r) \right) = 0.$$

Note that in the first part the sequences  $(X_n)$  and  $(N_n)$  may be dependent or independent.

*Proof.* We begin with the first part. Let  $\Omega_0 \in \mathcal{F}$  be a set with  $P(\Omega_0) = 1$  such that, for any  $\omega \in \Omega_0$ ,  $N_n(\omega) \rightarrow \infty$ . Fix  $\epsilon > 0$  and  $\omega \in \Omega_0$ . Since (13) holds uniformly in  $a$ , it follows that there is an  $M_\epsilon > 0$  such that for all  $m \geq M_\epsilon$  and all  $n \geq 1$  we have

$$\left| \frac{\binom{m}{r} \sum_{k=1}^{\infty} p_{X_{n+1}(\omega),k}^{r+1} (1 - p_{X_{n+1}(\omega),k})^{m-r}}{h_{\alpha,r}(m)} - F(X_{n+1}(\omega), r) \right| < \epsilon.$$

Now let  $M'_\epsilon(\omega) > 0$  be a number such that if  $n \geq M'_\epsilon(\omega)$  then  $N_n(\omega) \geq M_\epsilon$ . For all such  $n$ , the above holds with  $N_n(\omega)$  in place of  $m$ . From here, the first part follows.

For the second part, we have

$$\begin{aligned} & \frac{\binom{L_n}{r} \sum_{k=1}^{\infty} p_{X_{n+1},k}^{1+r} (1 - p_{X_{n+1},k})^{L_n-r}}{h_{\alpha,r}(n)} - \pi_{X_{n+1}}^{-(1-\alpha)} F(X_{n+1}, r) \\ &= \left( \frac{h_{\alpha,r}(L_n)}{h_{\alpha,r}(n)} - \pi_{X_{n+1}}^{-(1-\alpha)} \right) \left( \frac{\binom{L_n}{r} \sum_{k=1}^{\infty} p_{X_{n+1},k}^{1+r} (1 - p_{X_{n+1},k})^{L_n-r}}{h_{\alpha,r}(L_n)} - F(X_{n+1}, r) \right) \\ &+ \pi_{X_{n+1}}^{-(1-\alpha)} \left( \frac{\binom{L_n}{r} \sum_{k=1}^{\infty} p_{X_{n+1},k}^{1+r} (1 - p_{X_{n+1},k})^{L_n-r}}{h_{\alpha,r}(L_n)} - F(X_{n+1}, r) \right) \\ &+ F(X_{n+1}, r) \left( \frac{h_{\alpha,r}(L_n)}{h_{\alpha,r}(n)} - \pi_{X_{n+1}}^{-(1-\alpha)} \right). \end{aligned}$$

Since the Markov chain  $X$  is irreducible on a finite state space, all of its states are recurrent and hence  $\lim_{n \rightarrow \infty} L_n = \infty$  with probability 1. Thus, by the first part of this lemma, the fact that  $\max_{a \in A} F(a, r) < \infty$ , and the fact that  $\pi_{X_{n+1}}^{-(1-\alpha)} \leq (\min_a \pi_a)^{-(1-\alpha)}$ , it suffices to show that, with probability 1,

$$\lim_{n \rightarrow \infty} \left( \frac{h_{\alpha,r}(L_n)}{h_{\alpha,r}(n)} - \pi_{X_{n+1}}^{-(1-\alpha)} \right) = 0,$$

which holds by Lemma 6.1. □

*Proof of Theorem 5.1.* Note that, by Lemma 4.2,

$$\begin{aligned} P_\eta \{ \mathcal{L}_n = r \} &= E_\eta \left[ \mathbf{1} \{ L_n > r \} \binom{L_n}{r} \sum_{k \geq 1} p_{X_{n+1},k}^{1+r} (1 - p_{X_{n+1},k})^{L_n-r} \right] \\ &+ E_\eta \left[ \mathbf{1} \{ L_n = r \} \sum_{k \geq 1} p_{X_{n+1},k}^{1+r} \right] =: E_1 + E_2. \end{aligned}$$

We begin with  $E_2$ . Since  $\sum_{k \geq 1} p_{X_{n+1},k}^{1+r} \leq \sum_{k \geq 1} p_{X_{n+1},k} = 1$ , it follows that

$$\frac{E_2}{h_{\alpha,r}(n)} = \frac{E_\eta \left[ \mathbf{1} \{ L_n = r \} \sum_{k \geq 1} p_{X_{n+1},k}^{1+r} \right]}{h_{\alpha,r}(n)} \leq \frac{P_\eta \{ L_n = r \}}{h_{\alpha,r}(n)} \rightarrow 0,$$

where the convergence follows by Lemma 6.1. We next turn to  $E_1$ . Note that (13), the fact that  $|A| < \infty$ , and the fact that  $h_{\alpha,r}$  is locally bounded imply that there is a constant  $K' > 0$

depending only on  $r$  with

$$\begin{aligned} \frac{\binom{L_n}{r} \sum_{k \geq 1} p_{X_{n+1},k}^{1+r} (1 - p_{X_{n+1},k})^{L_n-r}}{h_{\alpha,r}(n)} &\leq K' \frac{h_{\alpha,r}(L_n)}{h_{\alpha,r}(n)} \\ &\leq H_\delta K' \left(\frac{L_n}{n}\right)^{-(1-\alpha)-\delta} \end{aligned}$$

for any  $\delta > 0$  and some  $H_\delta > 1$ . Here, the second inequality follows by the Potter bounds; see, e.g., Theorem 1.5.6 of [3]. For simplicity, set  $K_\delta = K'H_\delta$ . Fix  $\epsilon \in (0, \pi_\wedge)$  and let

$$A(n) = \{r + 1 \leq L_n < n\epsilon\} \quad \text{and} \quad B(n) = \{L_n \geq (n\epsilon) \vee (r + 1)\}.$$

Note that  $A(n) \cup B(n) = \{r + 1 \leq L_n\}$ . We can write

$$\begin{aligned} E_1 &= E_\eta \left[ \mathbf{1}_{A(n)} \binom{L_n}{r} \sum_{k \geq 1} p_{X_{n+1},k}^{1+r} (1 - p_{X_{n+1},k})^{L_n-r} \right] \\ &\quad + E_\eta \left[ \mathbf{1}_{B(n)} \binom{L_n}{r} \sum_{k \geq 1} p_{X_{n+1},k}^{1+r} (1 - p_{X_{n+1},k})^{L_n-r} \right] =: E_{1A} + E_{1B}. \end{aligned}$$

By Lemma 6.1, we have

$$\begin{aligned} \frac{E_{1A}}{h_{\alpha,r}(n)} &\leq K_\delta E_\eta \left[ \mathbf{1}_{A(n)} \left(\frac{L_n}{n}\right)^{-(1-\alpha)-\delta} \right] \\ &\leq K_\delta P_\eta(A(n)) n^{1-\alpha+\delta} \rightarrow 0. \end{aligned}$$

Similarly,

$$\begin{aligned} \frac{\mathbf{1}_{B(n)} \binom{L_n}{r} \sum_{k \geq 1} p_{X_{n+1},k}^{1+r} (1 - p_{X_{n+1},k})^{L_n-r}}{h_{\alpha,r}(n)} &\leq K_\delta \mathbf{1}_{B(n)} \left(\frac{L_n}{n}\right)^{-(1-\alpha)-\delta} \\ &\leq K_\delta \epsilon^{-(1-\alpha)-\delta}. \end{aligned}$$

Combining this with the fact that  $\pi_{X_{n+1}}^{-(1-\alpha)} F(X_{n+1}, r)$  is bounded for fixed  $r$  means that we can use dominated convergence to get

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{E_{1B}}{h_{\alpha,r}(n)} &= E_\eta \left[ \lim_{n \rightarrow \infty} \left( \mathbf{1}_{B(n)} \frac{\binom{L_n}{r} \sum_{k \geq 1} p_{X_{n+1},k}^{1+r} (1 - p_{X_{n+1},k})^{L_n-r}}{h_{\alpha,r}(n)} \right. \right. \\ &\quad \left. \left. - \pi_{X_{n+1}}^{-(1-\alpha)} F(X_{n+1}, r) \right) \right] \\ &\quad + \lim_{n \rightarrow \infty} E_\eta [\pi_{X_{n+1}}^{-(1-\alpha)} F(X_{n+1}, r)] \\ &= \lim_{n \rightarrow \infty} E_\eta [\pi_{X_{n+1}}^{-(1-\alpha)} F(X_{n+1}, r)] = E_\pi [\pi_{X_1}^{-(1-\alpha)} F(X_1, r)] = \sum_{a \in A} \pi_a^\alpha F(a, r), \end{aligned}$$

where the second equality follows from Lemma 6.2 and the fact that, with probability 1, there exists a (random)  $N$  such that  $\mathbf{1}_{B(n)} = 1$  for all  $n \geq N$ . The third equality follows from the fact that the distribution of  $X_n$  converges weakly to  $\pi$ , Skorokhod's representation theorem, and dominated convergence. This gives the first part of Theorem 5.1. The second part follows from the first and Lemma 6.1. □

### Appendix A. Regular variation

In this appendix we briefly review several basic facts about regularly varying distributions on  $\mathbb{N}_+$ . First, we recall that for a probability measure  $P = (p_k)_{k \in \mathbb{N}_+}$  on  $\mathbb{N}_+$ , the counting measure  $\nu_P$  is defined by (1) and the counting function  $\nu$  is defined by (2).

**Definition A.1.** A probability distribution  $P = (p_k)_{k \geq 1}$  with counting function  $\nu$  is said to be regularly varying, with exponent  $\alpha \in [0, 1]$ , if

$$\lim_{\varepsilon \rightarrow 0} \frac{\nu(\varepsilon)}{\varepsilon^{-\alpha} \ell(1/\varepsilon)} = 1$$

for some  $\ell \in \mathcal{SV}$ . In this case, we write  $P \in RV_\alpha(\ell)$ .

To motivate this definition, we recall the following fact from [11]. For  $\alpha \in (0, 1)$ , we have  $P \in RV_\alpha(\ell)$  if and only if  $p_k = k^{-1/\alpha} \ell^*(k)$  for some  $\ell^* \in \mathcal{SV}$ , which is, in general, different from  $\ell$ . When  $\alpha = 0$ , a sufficient condition for  $P \in RV_\alpha(\ell)$  is that there exists an  $\ell_0 \in \mathcal{SV}$  with

$$\int_{[0, \varepsilon]} x \nu_P(dx) = \sum_{k \geq 1} p_k \mathbf{1}\{p_k \leq \varepsilon\} \sim \varepsilon \ell_0(1/\varepsilon) \quad \text{as } \varepsilon \rightarrow 0. \quad (20)$$

In this case, we necessarily have

$$\ell(x) \sim \int_1^x u^{-1} \ell_0(u) du \quad \text{as } x \rightarrow \infty$$

and  $\ell_0(x)/\ell(x) \rightarrow 0$  as  $x \rightarrow \infty$ ; see Proposition 15 of [11]. We will generally assume that (20) holds in this case.

**Proposition A.1.** Let  $P = (p_k)_{k \geq 1} \in RV_\alpha(\ell)$ . If  $\alpha \in (0, 1)$  then, for all  $r \geq 0$ ,

$$\lim_{n \rightarrow \infty} \frac{\binom{n}{r} \sum_{k=1}^{\infty} p_k^{r+1} (1-p_k)^{n-r}}{n^{\alpha-1} \ell(n)} = \frac{\alpha \Gamma(r+1-\alpha)}{r!}. \quad (21)$$

If  $\alpha = 0$  and (20) holds then, for every  $r \geq 0$ ,

$$\lim_{n \rightarrow \infty} \frac{\binom{n}{r} \sum_{k=1}^{\infty} p_k^{r+1} (1-p_k)^{n-r}}{n^{-1} \ell_0(n)} = 1.$$

If  $\alpha = 1$  then for every  $r \geq 1$  the result in (21) holds. If  $\alpha = 1$  and  $r = 0$  then

$$\lim_{n \rightarrow \infty} \frac{\sum_{k=1}^{\infty} p_k (1-p_k)^n}{\ell_1(n)} = 1,$$

where  $\ell_1(x) = \int_x^\infty u^{-1} \ell(u) du$  for  $x > 1$  is a function with  $\ell_1 \in \mathcal{SV}$  and  $\ell(x)/\ell_1(x) \rightarrow 0$  as  $x \rightarrow \infty$ .

*Proof.* For  $\alpha \in (0, 1)$  this is Proposition 7 of [18]. For  $\alpha = 1$  the result follows by combining Proposition 18 of [11] with Lemma 2 of [14]. Similarly, for  $\alpha = 0$  the result follows by combining Proposition 19 of [11] with Lemma 2 of [14]. The facts about  $\ell_1$  are given in Proposition 14 of [11].  $\square$

**Proposition A.2.** Fix  $\alpha \in [0, 1]$  and  $\ell \in \mathcal{SV}$ . When  $\alpha \neq 0$  assume that

$$\lim_{\varepsilon \rightarrow 0} \frac{\nu(\varepsilon)}{\varepsilon^{-\alpha} \ell(1/\varepsilon)} = 0,$$

and when  $\alpha = 0$  assume that

$$\lim_{\varepsilon \rightarrow 0} \frac{\int_{[0, \varepsilon]} x \mathbf{v}_P(dx)}{\varepsilon \ell(1/\varepsilon)} = 0.$$

If  $\alpha \in [0, 1)$  then, for all  $r \geq 0$ ,

$$\lim_{n \rightarrow \infty} \frac{\binom{n}{r} \sum_{k=1}^{\infty} p_k^{r+1} (1-p_k)^{n-r}}{n^{\alpha-1} \ell(n)} = 0. \tag{22}$$

If  $\alpha = 1$  then for all  $r \geq 1$  the result in (22) holds. If  $\alpha = 1$  and  $r = 0$  then

$$\lim_{n \rightarrow \infty} \frac{\sum_{k=1}^{\infty} p_k (1-p_k)^n}{\ell_1(n)} = 0,$$

where  $\ell_1(x)$  is derived from  $\ell$  as in Proposition A.1.

*Proof.* Let  $q = r + 1$ , let  $\mathbf{v}_P^q(dx) = x^q \mathbf{v}_P(dx)$ , let  $\Phi_q(n) = \frac{n^q}{q!} \sum_{k=1}^{\infty} p_k^q e^{-np_k}$ , and note that

$$\Phi_q(n) = \frac{n^q}{q!} \int_0^1 e^{-nx} \mathbf{v}_P^q(dx).$$

A standard application of Fubini’s theorem gives

$$\mathbf{v}_P^q([0, s]) = \int_{[0, s]} x^q \mathbf{v}_P(dx) = q \int_0^s u^{q-1} \nu(u) du - s^q \mathbf{v}_P((s, 1]).$$

Fix  $\delta > 0$ . For  $\alpha \neq 0$ , the assumptions imply that for small enough  $s$  we have  $\nu(s) \leq \delta s^{-\alpha} \ell(1/s)$ . It follows that for  $\alpha \in (0, 1)$  or  $\alpha = 1$  and  $q \geq 2$  we have

$$\begin{aligned} \lim_{s \rightarrow 0^+} \frac{\mathbf{v}_P^q([0, s])}{s^{q-\alpha} \ell(1/s)} &= \lim_{s \rightarrow 0^+} \frac{q \int_0^s u^{q-1} \nu(u) du}{s^{q-\alpha} \ell(1/s)} \leq \delta \lim_{s \rightarrow 0^+} \frac{q \int_0^s u^{q-1-\alpha} \ell(1/u) du}{s^{q-\alpha} \ell(1/s)} \\ &= \delta \lim_{s \rightarrow 0^+} \frac{q \int_{1/s}^{\infty} u^{-(q+1-\alpha)} \ell(u) du}{s^{q-\alpha} \ell(1/s)} = \frac{q}{q-\alpha} \delta, \end{aligned}$$

where the last equality follows by Karamata’s theorem (Proposition 1.5.10 in [3]). Hence,

$$\lim_{s \rightarrow 0^+} \frac{\mathbf{v}_P^q([0, s])}{s^{q-\alpha} \ell(1/s)} = 0.$$

Similarly, when  $\alpha = 1$  and  $q = 1$  we have

$$\begin{aligned} \lim_{s \rightarrow 0^+} \frac{\mathbf{v}_P^1([0, s])}{\ell_1(1/s)} &= \lim_{s \rightarrow 0^+} \frac{\int_0^s \nu(u) du}{\ell_1(1/s)} \leq \delta \lim_{s \rightarrow 0^+} \frac{\int_0^s u^{-1} \ell(1/u) du}{\ell_1(1/s)} \\ &= \delta \lim_{s \rightarrow 0^+} \frac{\int_{1/s}^{\infty} u^{-1} \ell(u) du}{\ell_1(1/s)} = \delta \lim_{s \rightarrow 0^+} \frac{\ell_1(1/s)}{\ell_1(1/s)} = \delta, \end{aligned}$$

and hence

$$\lim_{s \rightarrow 0^+} \frac{\mathbf{v}_P^1([0, s])}{\ell_1(1/s)} = 0.$$

When  $\alpha = 0$  we have

$$\mathbf{v}_P^q([0, s]) = \int_{[0, s]} x^{q-1} \mathbf{v}_P^1(dx) \leq s^{q-1} \mathbf{v}^1([0, s]),$$

and hence

$$\lim_{s \rightarrow 0^+} \frac{\mathbf{v}_P^q([0, s])}{s^q \ell(1/s)} \leq \lim_{s \rightarrow 0^+} \frac{s^{q-1} \mathbf{v}_P^1([0, s])}{s^q \ell(1/s)} = \lim_{s \rightarrow 0^+} \frac{\mathbf{v}_P^1([0, s])}{s \ell(1/s)} = 0.$$

From here a version of Karamata's Tauberian theorem (Theorem 1.7.1' in [3]) implies that for  $\alpha \in [0, 1)$  or  $\alpha = 1$  and  $q \geq 2$ ,

$$\lim_{n \rightarrow \infty} \frac{\Phi_q(n)}{n^\alpha \ell(n)} = \frac{\int_0^1 e^{-nx} \mathbf{v}_P^q(dx)}{q! n^{\alpha-q} \ell(n)} = 0$$

and that the corresponding result holds for the case  $\alpha = 1$  and  $q = 1$ . From here, since  $(n+1)^{\alpha-1} \ell(n+1) \sim n^{\alpha-1} \ell(n)$  and  $\ell_1(n+1) \sim \ell_1(n)$ , we can use Lemma 2 of [14] to complete the result.  $\square$

### Acknowledgements

The work of M. Kelbert and Q. Paris has been funded by the Russian Academic Excellence Project 5-100. The work of M. Grabchak is supported, in part, by the Russian Science Foundation (project number 17-11-01098). The authors would like to thank the two anonymous referees, whose comments led to improvements in the presentation of this paper.

### References

- [1] BEN-HAMOU, A., BOUCHERON, S. AND GASSIAT, E. (2016) Pattern coding meets censoring: (Almost) adaptive coding on countable alphabets. [arXiv:1608.08367](https://arxiv.org/abs/1608.08367).
- [2] BEN-HAMOU, A., BOUCHERON, S. AND OHANNESSIAN, M. I. (2017) Concentration inequalities in the infinite urn scheme for occupancy counts and the missing mass, with applications. *Bernoulli* **23**, 249–287.
- [3] BINGHAM, N. H., GOLDIE, C. M. AND TEUGELS, J. L. (1987) *Regular Variation (Encyclopedia of Mathematics And Its Applications)*. Cambridge University Press.
- [4] BUBECK, S., ERNST, D. AND GARIVIER, A. (2013) Optimal discovery with probabilistic expert advice: Finite time analysis and macroscopic optimality. *J. Mach. Learn. Res.* **14**, 601–623.
- [5] CHAO, A. (1981) On estimating the probability of discovering a new species. *Ann. Statist.* **9**, 1339–1342.
- [6] CHEN, S. F. AND GOODMAN, J. (1999) An empirical study of smoothing techniques for language modeling. *Comput. Speech Lang.* **13**, 359–394.
- [7] DECROUEZ, G., GRABCHAK, M. AND PARIS, Q. (2018) Finite sample properties of the mean occupancy counts and probabilities. *Bernoulli* **24**, 1910–1941.
- [8] EFRON, B. AND THISTED, R. (1976) Estimating the number of unseen species: How many words did Shakespeare know? *Biometrika* **63**, 435–447.
- [9] GANDOLFI, A. AND SASTRI, C. C. A. (2004) Nonparametric estimations about species not observed in a random sample. *Milan J. Math.* **72**, 81–105.
- [10] GLYNN, P. W. AND ORMONEIT, D. (2002) Hoeffding's inequality for uniformly ergodic Markov chains. *Statist. Prob. Lett.* **56**, 143–146.
- [11] GNEDIN, A., HANSEN, B. AND PITMAN, J. (2007) Notes on the occupancy problem with infinitely many boxes: General asymptotics and power laws. *Prob. Surv.* **4**, 146–171.
- [12] GOOD, I. J. (1953) The population frequencies of species and the estimation of population parameters. *Biometrika* **40**, 237–264.
- [13] GOOD, I. J. AND TOULMIN, G. H. (1956) The number of new species, and the increase in population coverage, when a sample is increased. *Biometrika* **43**, 45–63.
- [14] GRABCHAK, M. AND ZHANG, Z. (2017) Asymptotic properties of Turing's formula in relative error. *Mach. Learn.* **106**, 1771–1785.
- [15] JOHNSON, N. L. AND KOTZ, S. (1977) *Urn Models and Their Application*. Wiley, New York.



- [16] KARLIN, S. (1967) Central limit theorems for certain infinite urn schemes. *J. Math. Mech.* **17**, 373–401.
- [17] MAO, C. X. AND LINDSAY, B. G. (2002) A Poisson model for the coverage problem with a genomic application. *Biometrika* **89**, 669–681.
- [18] OHANNESSIAN, M. I. AND DAHLEH, M. A. (2012) Rare probability estimation under regularly varying heavy tails. In *Proc. 25th Ann. Conf. on Learning Theory*, Vol. 23, pp. 21.1–21.24.
- [19] ORLITSKY, A., SANTHANAM, N. P. AND ZHANG, J. (2004) Universal compression of memoryless sources over unknown alphabets. *IEEE Trans. Inf. Theory* **50**, 1469–1481.
- [20] PAULIN, D. (2015) Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electron. J. Probab.* **20**, 1–32.
- [21] RESNICK, S. I. (2007) *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, New York.
- [22] ROBERTS, G. O. AND ROSENTHAL, J. S. (2004) General state space Markov chains and MCMC algorithms. *Prob. Surv.* **1**, 20–71.
- [23] THISTED, R. AND EFRON, B. (1987) Did Shakespeare write a newly discovered poem? *Biometrika* **74**, 445–455.
- [24] ZHANG, C. H. (2005) Estimation of sums of random variables: Examples and information bounds. *Ann. Statist.* **33**, 2022–2041.
- [25] ZHANG, Z. AND HUANG, H. (2007) Turing’s formula revisited. *J. Quant. Ling.* **14**, 222–241.