



# Tail Index Estimation for Discrete Heavy-Tailed Distributions with Application to Statistical Inference for Regular Markov Chains

Patrice Bertail, Stephan Cléménçon, Carlos Fernández

## ► To cite this version:

Patrice Bertail, Stephan Cléménçon, Carlos Fernández. Tail Index Estimation for Discrete Heavy-Tailed Distributions with Application to Statistical Inference for Regular Markov Chains. 2023. hal-04299756

**HAL Id: hal-04299756**

**<https://hal.science/hal-04299756>**

Preprint submitted on 22 Nov 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Tail Index Estimation for Discrete Heavy-Tailed Distributions with Application to Statistical Inference for Regular Markov Chains

Patrice Bertail<sup>1†</sup>, Stephan Cléménçon<sup>2†</sup>, Carlos Fernández<sup>1,2\*†</sup>

<sup>1</sup>MODAL'X, UMR CNRS 9023, Université Paris Nanterre, 200 Avenue de la République, Nanterre, 92000, Ile-de-France, France .

<sup>2\*</sup>LTCI, Telecom Paris, Institut Polytechnique de Paris, 19 place Marguerite Perey, Palaiseau, 91123, Ile-de-France, France .

\*Corresponding author(s). E-mail(s): [fernandez@telecom-paris.fr](mailto:fernandez@telecom-paris.fr);  
Contributing authors: [patrice.bertail@parisnanterre.fr](mailto:patrice.bertail@parisnanterre.fr);  
[stephan.clemencon@telecom-paris.fr](mailto:stephan.clemencon@telecom-paris.fr);

<sup>†</sup>These authors contributed equally to this work.

## Abstract

It is the purpose of this paper to investigate the issue of estimating the regularity index  $\beta > 0$  of a discrete heavy-tailed r.v.  $\mathbf{S}$ , *i.e.* a r.v.  $\mathbf{S}$  valued in  $\mathbb{N}^*$  such that  $\mathbb{P}(\mathbf{S} > \mathbf{n}) = \mathbf{L}(\mathbf{n}) \cdot \mathbf{n}^{-\beta}$  for all  $\mathbf{n} \geq 1$ , where  $\mathbf{L} : \mathbb{R}_+^* \rightarrow \mathbb{R}_+$  is a slowly varying function. Such discrete probability laws, referred to as generalized Zipf's laws sometimes, are commonly used to model rank-size distributions after a preliminary range segmentation in a wide variety of areas such as *e.g.* quantitative linguistics, social sciences or information theory. As a first go, we consider the situation where inference is based on independent copies  $\mathbf{S}_1, \dots, \mathbf{S}_n$  of the generic variable  $\mathbf{S}$ . Just like the popular Hill estimator in the continuous heavy-tail situation, the estimator  $\hat{\beta}$  we propose can be derived by means of a suitable reformulation of the regularly varying condition, replacing  $\mathbf{S}$ 's survivor function by its empirical counterpart. Under mild assumptions, a non-asymptotic bound for the deviation between  $\hat{\beta}$  and  $\beta$  is established, as well as limit results (consistency and asymptotic normality). Beyond the i.i.d. case, the inference method proposed is extended to the estimation of the regularity index of a regenerative  $\beta$ -null recurrent Markov chain. Since the parameter  $\beta$  can be then viewed as the tail index of the (regularly varying) distribution of the return time of the chain  $\mathbf{X}$  to any (pseudo-) regenerative set, in this case, the estimator is constructed from the successive regeneration times. Because the durations between consecutive regeneration times are asymptotically

independent, we can prove that the consistency of the estimator promoted is preserved. In addition to the theoretical analysis carried out, simulation results provide empirical evidence of the relevance of the inference technique proposed.

**Keywords:** generalized discrete Pareto distribution, nonparametric estimation, null recurrent Markov chain, regularity index, Zipf's law

**MSC Classification:** 60K35

## 1 Introduction

This article is devoted to the study of the problem of estimating the regularity index  $\beta > 0$  of a generalized discrete Pareto distribution, namely the probability distribution of a random variable  $S$  defined on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , taking its values in  $\mathbb{N}^*$  and such that:

$$\mathbb{P}(S > n) = n^{-\beta} L(n) \text{ for all } n \geq 1, \quad (1)$$

where  $L : \mathbb{R}_+ \rightarrow \mathbb{R}$  is a slowly varying function, *i.e.* such that  $L(\lambda z)/L(z) \rightarrow +1$  as  $z \rightarrow +\infty$  for any  $\lambda > 0$ , see [Bingham, Goldie, and Teugels \(1987\)](#). Such discrete power law probability distributions, also referred to as generalized Zipf's laws sometimes, are often used to model the distribution of discrete data exhibiting a specific rank-frequency relationship, namely when the logarithm of the frequency and that of the rank order are nearly proportional. Such a phenomenon has been empirically observed in many ranking systems: in quantitative linguistics (*i.e.* when analysing word frequency law in natural language, see *e.g.* [Manning and Scütze \(1999\)](#)) in the first place, as well as in a very wide variety of situations, too numerous to be exhaustively listed here. One may refer to [Sidra, Shougeng, and Nadeem \(2018\)](#), [Lazzardi et al. \(2021\)](#) or [Zanette \(2006\)](#) among many others. In this paper, we first consider the issue of estimating the parameter  $\beta$  involved in (1) (supposedly unknown, like the function  $L$ ) in the classic (asymptotic) i.i.d. statistical setting, *i.e.* based on an increasing number  $n \geq 1$  of independent copies  $S_1, \dots, S_n$  of the generic r.v.  $S$ . Statistical inference for discrete heavy-tailed distributions has not received much attention in the literature. Most of the very few dedicated methods documented either deal with very specific cases as in *e.g.* [Goldstein, Morris, and Yen \(2004\)](#), [Matsui, Mikosch, and Tafakori \(2013\)](#) and [Clauset, Shalizi, and Newman \(2009\)](#) or else consist in applying techniques originally designed for continuous heavy-tailed distributions to the discrete data after a preliminary addition of an independent uniform noise ([Voitalov, van der Hoorn, van der Hofstad, & Krioukov 2019](#)). The vast majority of the regular variation index estimators proposed in the literature, Hill or Pickand estimators in particular ([Hill 1975](#); [Pickands 1975](#)), are based on order statistics, which causes obvious difficulties in the discrete case because of the possible occurrence of many ties. In contrast, the estimator under study here is based on the analysis of the probability of exponentially separated tail events. It simply rests on the fact that, as can be immediately deduced from (1), we have  $\ln(p_k) - \ln(p_{k+1}) = \beta + \ln(L(e^k)/L(e^{k+1}))$ , where  $\ln(x)$  denotes the natural logarithm of any real number  $x > 0$  and  $p_l = \mathbb{P}(S > e^l)$  for all  $l \in \mathbb{N}$ , and

that  $L(e^{k+1})/L(e^k)$  is expected to be very close to 1 for  $k \in \mathbb{N}$  chosen sufficiently large. A natural (plug-in) inference technique can be then devised by replacing the tail probabilities  $p_l$  with their empirical versions  $\hat{p}_l^{(n)} = (1/n) \sum_{i=1}^n \mathbb{I}\{S_i > e^l\}$  for  $l \in \mathbb{N}$ , where  $\mathbb{I}\{\mathcal{A}\}$  denotes the indicator function of an event  $\mathcal{A}$ . This yields the estimator

$$\hat{\beta}_n(k) = \ln\left(\hat{p}_k^{(n)}\right) - \ln\left(\hat{p}_{k+1}^{(n)}\right), \quad (2)$$

provided that  $\hat{p}_{k+1}^{(n)} > 0$  (as shall be seen, this occurs with large probability if  $n$  is sufficiently large). By convention, we set  $\hat{\beta}_n(k) = 0$  when  $\hat{p}_{k+1}^{(n)} = 0$ . We point out that it has exactly the same form as that proposed and analysed in [Carpentier and Kim \(2015\)](#) in a different context, that of (continuous) *approximately Pareto distributions*<sup>1</sup> namely. In the discrete generalized Pareto framework, we prove that for an appropriate choice of the hyper-parameter  $k = k_n$  (typically chosen of order  $\ln(n)$ ), the estimator (2) is strongly consistent and asymptotically normal as  $n \rightarrow +\infty$ . Non-asymptotic upper confidence bounds for the absolute deviations between  $\hat{\beta}_n(k)$  and  $\beta$  are also established here.

Although estimation of the parameter  $\beta > 0$  in (1) in the discrete i.i.d. setting is an important issue in itself, the present paper also finds its motivation in the problem of recovering statistically the regularity index  $\beta \in (0, 1)$  of a regenerative *regular* null-recurrent Markov chain  $X = (X_n)_{n \in \mathbb{N}}$ , based on the observation of a finite sample path  $X_1, \dots, X_n$  with  $n \geq 1$ . As explained in [Chen \(1999\)](#) (see also [Chen \(2000\)](#)), for regular Markov chains, the regularity index  $\beta \in (0, 1]$  controls the (sublinear) rate at which the number of visits to any given Harris set increases with observation time  $n$ , no matter the initial distribution. In the *regenerative* case (*i.e.* when the chain  $X$  possesses an *accessible atom*, a Harris set on which the transition probability is constant), the distribution of the regenerative time, the return time to the atom, is a discrete generalized Pareto (1) and the parameter  $\beta$  is its tail index. Due to the non-standard behaviour of traditional estimators in this context, statistical inference for null-recurrent Markov chains is very poorly documented in the literature ([Gao, Tjøstheim, & Yin 2013](#); [Karlsen, Myklebust, & Tjøstheim 2010](#); [Karlsen & Tjøstheim 2001](#); [Myklebust, Karlsen, & Tjøstheim 2012](#)) and, to the best of our knowledge, estimation of the key quantity  $\beta$  has not received much attention. It is also the goal of this article to extend the use of the estimator (2) to the case where the  $S_i$ 's are the successive durations between the consecutive regeneration times up to time  $n$ . The main difficulty naturally arises from the fact that the number  $1 + N_n \geq 0$  of regeneration times (and thus the number of durations) is now random, and the variables  $S_1, \dots, S_{N_n}$  are not independent any more when  $N_n \geq 1$  (in particular, their sum is less than  $n$  by construction). We show that the limit properties (consistency and asymptotic normality) of the estimator are however preserved. For illustration purposes, numerical experiments have been carried out, providing empirical evidence of the relevance of the estimation method promoted. Extension to the general case of (pseudo-regenerative) null-recurrent chains is also discussed, the difficulties inherent in applying the methodology originally proposed in

---

<sup>1</sup>The distribution of a real-valued r.v.  $X$  is said to be *approximately Pareto* with tail index  $\beta > 0$  if its survivor function is of the form:  $\forall x > 0, \mathbb{P}(X > x) = L(x)x^{-\beta}$ , where  $L$  is asymptotically constant at infinity, *i.e.* there exists  $C \in (0, \infty)$  s.t.  $L(x) \rightarrow C$  as  $x \rightarrow +\infty$ .

Bertail and Cl  men  on (2006b) in the positive recurrent case to mimic regenerative Nummelin extensions (Nummelin 1984) being explained at length.

The paper is organized as follows. A thorough analysis of the behaviour of the estimator (2) in the i.i.d. case, illustrated by numerical experiments, is first carried out in section 2. The asymptotic results thus established are next extended in section 3 to the regenerative regular Markovian setup, when the estimator is computed based on a single finite-length trajectory of the atomic chain. Experimental results are also displayed and the main barrier to the extension of the methodology promoted to general (*i.e.* pseudo-regenerative) regular null-recurrent chains is also discussed therein. Technical proofs are deferred to the Appendix section.

## 2 Tail Index Estimation - The Discrete Heavy-Tailed i.i.d. Case

Throughout this section,  $S_1, \dots, S_n$  are independent copies of a generic discrete generalized Pareto r.v.  $S$ , *i.e.* a random variable  $S$  with survivor function of type (1), where the parameter  $\beta > 0$  and the slowly varying function  $L$  are supposedly unknown. As a first go, we start to investigate the (asymptotic) behaviour of the estimator (2) in this basic general framework and next develop the analysis in particular situations, *i.e.* when the function  $L$  has a specific form.

### 2.1 Main Results - Confidence Bounds and Limit Theorems

As explained in the Introduction section, the estimator (2) can be viewed as an empirical counterpart of the quantity

$$\beta(k) = \ln(p_k) - \ln(p_{k+1}) = \beta + \ln \left( \frac{L(e^k)}{L(e^{k+1})} \right), \quad (3)$$

see (1), which tends to  $\beta$  as  $k \rightarrow \infty$  by virtue of the slow variation property of  $L$ . As previously emphasized, unless the function  $L$  is supposed to be asymptotically constant (*i.e.* there exists  $C > 0$  s.t.  $L(x) \rightarrow C$  as  $x \rightarrow +\infty$ ), the discrete generalized Pareto model (1) is not a discrete version of the (continuous) approximately  $\beta$ -Pareto model considered in Carpentier and Kim (2015) and, consequently, the validity framework established therein does not apply directly here. The proposition below provides an upper confidence bound for the absolute deviations between (2) and  $\beta$  (respectively, between (2) and  $\beta(k)$ ).

**Proposition 2.1.** *Let  $\delta \in (0, 1/2)$  and set  $u_n(\delta) = \ln(2/\delta)/n$  for all  $n \geq 1$ . If  $k \geq 1$  is such that  $p_{k+1} \geq 16u_n(\delta)$ , then, with probability at least  $1 - 2\delta$ , we have:*

$$\left| \widehat{\beta}_n(k) - \beta \right| \leq 6\sqrt{\frac{u_n(\delta)}{p_{k+1}}} + \left| \ln \left( \frac{L(e^k)}{L(e^{k+1})} \right) \right|. \quad (4)$$

Refer to the Appendix section for the technical proof. The bound (4) reveals some sort of 'bias-variance' trade-off, ruled by the hyperparameter  $k > 0$ . The second term

on the right-hand side can be viewed as the bias of the inference method, insofar as the estimator (2) can be seen as an empirical version of the approximation (3). It decays to 0 as  $k$  increases towards infinity, while the first term, whose presence is due to the random nature of the estimator, tends to  $+\infty$ . We point out that *second-order slow variation conditions* (Goldie & Smith 1987) are required to bound the (vanishing) bias term in (4), as shall be explained in subsection 2.2. The following result reveals that for an appropriate choice of  $k = k_n$ , the estimator (2) is strongly consistent.

**Theorem 2.2** (Strong consistency). *Suppose that, as  $n \rightarrow +\infty$ , we have  $k_n \rightarrow +\infty$  so that  $(\ln n) \exp(k_n \beta)/n = o(L(\exp(k_n)))$ . Then, we have:*

$$\hat{\beta}_n(k_n) \rightarrow \beta \text{ almost surely, as } n \rightarrow +\infty.$$

In particular, as stated below, strong consistency is guaranteed when  $k_n$  is of logarithmic order.

**Corollary 2.1.** *Let  $0 < A < 1/\beta$ . Then, we have:*

$$\hat{\beta}_n(A \ln n) \rightarrow \beta \text{ almost surely, as } n \rightarrow +\infty.$$

Now, the following results establish the asymptotic normality of the deviation between (2) and  $\beta(k_n)$ , when appropriately normalized.

**Theorem 2.3** (Asymptotic normality). *Suppose that  $k_n$  satisfies the conditions of Theorem 2.2 and  $k_n = o(n)$  as  $n \rightarrow +\infty$ .*

(i) *Then, as  $n \rightarrow +\infty$ , we have the convergence in distribution:*

$$\sqrt{np_{k_n}} \left( \hat{\beta}_n(k_n) - \beta(k_n) \right) \Rightarrow \mathcal{N}(0, e^\beta - 1).$$

(ii) *In addition, asymptotic normality holds true for the 'standardized' deviation:*

$$\frac{\sqrt{n\hat{p}_{k_n}^{(n)}} \left( \hat{\beta}_n(k_n) - \beta(k_n) \right)}{\sqrt{e^{\hat{\beta}_n(k_n)} - 1}} \Rightarrow \mathcal{N}(0, 1), \text{ as } n \rightarrow +\infty.$$

The asymptotic normality results above can be extended to the deviation between (2) and  $\beta$ , provided that the bias term  $\beta(k_n) - \beta$  vanishes at an appropriate rate, as stated below.

**Corollary 2.2.** *Suppose that the conditions of Theorem 2.3 are fulfilled. In addition, assume that  $k_n$  is such that*

$$\sqrt{np_{k_n}} \left( 1 - \frac{L(e^{k_n})}{L(e^{k_n+1})} \right) \rightarrow 0, \text{ as } n \rightarrow +\infty. \quad (5)$$

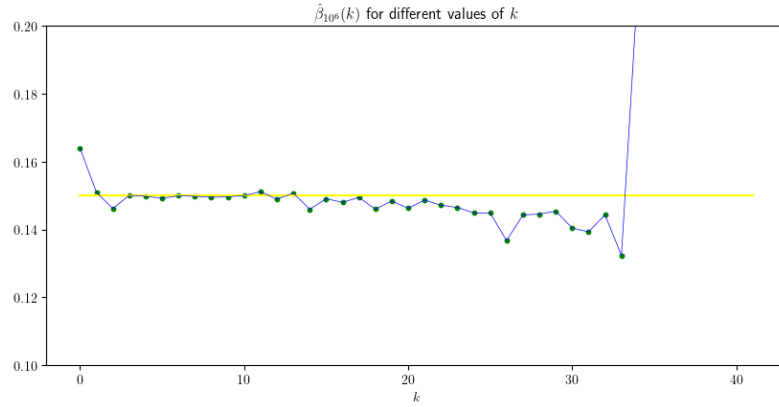
(i) *Then, we have the convergence in distribution*

$$\sqrt{np_{k_n}} \left( \hat{\beta}_n(k_n) - \beta \right) \Rightarrow \mathcal{N}(0, e^\beta - 1) \text{ as } n \rightarrow +\infty.$$

(ii) In addition, the “studentized” version is asymptotically normal:

$$\frac{\sqrt{n\hat{p}_{k_n}^{(n)}}(\hat{\beta}_n(k_n) - \beta)}{\sqrt{e^{\hat{\beta}_n(k_n)} - 1}} \Rightarrow \mathcal{N}(0, 1) \text{ as } n \rightarrow +\infty.$$

Of course, the condition (5) on  $k_n$  can be hardly checked in practice. This is a classic issue in tail estimation and in the statistical analysis of extreme values more generally. The choice of the hyperparameter  $k$  somehow rules the (asymptotic) bias-variance trade-off: the estimator (2) is expected to be of large variance when  $k$  is large and to have a large bias if  $k$  is too small. As depicted in Fig. 1, to choose  $k$ , one may use the same approach as that originally proposed for the Hill estimator (see *e.g.* Resnick (2007)), which consists in plotting the values of (2) for a range of values of  $k$  and choosing  $k$  in a region where a certain degree of stability is exhibited.



**Fig. 1:** Behaviour of  $\hat{\beta}_n(k)$  for different values of  $k$ , to estimate the parameter  $\beta = 0.15$  based on a dataset of  $10^6$  independent realizations of a Zeta distribution<sup>2</sup> with parameter  $\alpha = \beta - 1$ .

**Remark 1.** (AVERAGED VERSIONS) *From a practical point of view, rather than picking a single value for  $k$ , another natural approach would consist in averaging the estimators (2) over a range of values for the hyperparameter. Let  $k$  and  $m$  be such that  $k > m$  and define*

$$\beta(k, m) = \frac{1}{2m+1} \sum_{j=-m}^m \beta(k+j), \quad \hat{\beta}_n(k, m) = \frac{1}{2m+1} \sum_{j=-m}^m \hat{\beta}_n(k+j).$$

<sup>2</sup>A discrete r.v.  $W$  follows a Zeta distribution with parameter  $\alpha$  if  $\mathbb{P}(W = k) = (k^\alpha \zeta(\alpha))^{-1}$  where  $\zeta$  is the Riemann zeta function. The cdf of a Zeta distribution satisfies  $\mathbb{P}(W \geq k) \sim k^{\alpha+1}/(\zeta(\alpha)(\alpha+1))$ . This distribution is also known as Zipf’s distribution due to its relationship with Zipf’s law.

One may easily check that

$$\beta(k, m) = \beta + \frac{1}{2m+1} \left| \ln \left( \frac{L(e^{k-m})}{L(e^{k+m+1})} \right) \right|. \quad (6)$$

The non-asymptotic bound in Proposition 2.1 can be extended to the averaged version, as revealed by the analysis carried out in A.7 in the Appendix section, as well as the strong consistency and asymptotic normality results. However, the asymptotic variance of the averaged version is shown to increase with  $m$ .

In the next subsection, we discuss further how the behaviour of the slowly varying function  $L$  impacts the 'bias-variance' contributions revealed by the bound (4).

## 2.2 Refined 'Bias vs Variance' Analysis - Examples

We now consider several specific cases of distributions of type (1) (i.e. several instances of the slowly varying functions  $L$ ) to explicit the asymptotic order of magnitude of the terms  $1/\sqrt{np_{k+1}}$  and  $|\ln(L(e^{k_n})/L(e^{k_n+1}))|$  involved in the bound (4), when  $k_n$  is picked as in Corollary 2.1:  $k_n = A \ln n$  with  $0 < A < 1/\beta$ .

- **The logarithmic case:** Suppose that  $L(n) = C \ln n$ , where  $C > 0$ . In this situation, we have  $|\ln(L(e^{k_n})/L(e^{k_n+1}))| \sim 1/(A \ln n)$  as  $n \rightarrow +\infty$ , whereas  $1/\sqrt{np_{k+1}} = O(1/\sqrt{n^{1-A\beta} \ln n})$ .
- **The inversely logarithmic case:** Consider now the situation where  $L(n) = C/\ln n$  with  $C > 0$ . Then, we still have  $|\ln(L(e^{k_n})/L(e^{k_n+1}))| \sim 1/(A \ln n)$ , while  $1/\sqrt{np_{k+1}} = O(\sqrt{(\ln n)/n^{1-A\beta}})$  as  $n \rightarrow +\infty$ .

We point out that, in the two examples above, the conditions of Corollary 2.2 are not met, the bias being too big to get asymptotic normality (centered at  $\beta$ ).

- **The asymptotically constant case:** Suppose that  $L(n) = e^{C_0}(1 + \epsilon(n))$  where  $C_0 > 0$  and  $\epsilon(n) \rightarrow 0$  as  $n \rightarrow +\infty$ . In this case,  $|\ln(L(e^{k_n})/L(e^{k_n+1}))| = O(\epsilon(n^A))$  and  $1/\sqrt{np_{k+1}} = O(1/\sqrt{n^{1-A\beta}})$ . Hence, if  $|\epsilon(n^A)| = O(n^{-\lambda})$  for some  $\lambda > 0$ , then the conditions of Corollary 2.2 are satisfied if we take  $k_n = A \ln n$  such that  $\max\{(1-2\lambda)/\beta, 0\} < A < 1/\beta$ .
- **Slow variation with a remainder (SR2):** Consider the case where the slowly varying function satisfies the condition SR2 introduced in Bingham et al. (1987): there exist two real-valued functions  $k$  and  $g$  defined on  $\mathbb{R}_+$  such that, for all  $\lambda > 0$ ,

$$\frac{L(\lambda x)}{L(x)} - 1 \sim \kappa(\lambda)g(x), \text{ as } x \rightarrow +\infty, \quad (7)$$

where  $\kappa(\lambda) = c \int_1^\lambda \theta^{\rho-1} d\theta$ ,  $c > 0$  and  $g$  is regularly varying with index  $\rho \leq 0$ , i.e.  $g(x) = x^\rho U(x)$  where  $U$  is a slowly varying function. Under the additional assumption that  $g$  has positive decrease, Corollary 3.12.3 in Bingham et al. (1987) gives the following representation:

$$L(x) = C (1 - c|\rho|^{-1}g(x) + o(g(x))), \text{ as } x \rightarrow +\infty, \quad (8)$$



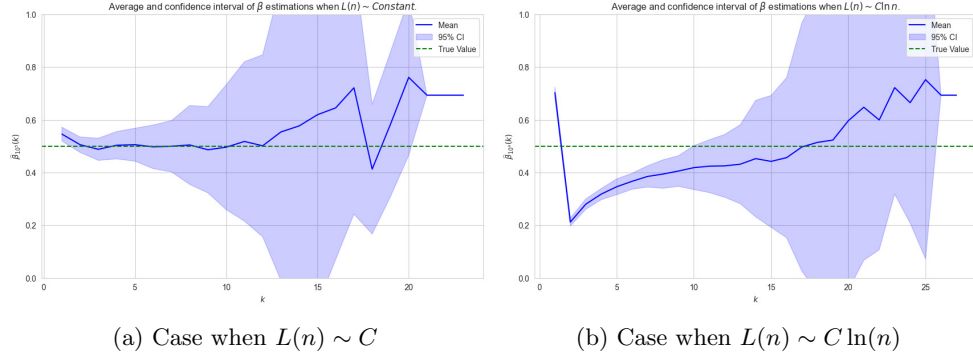
where  $C$  is a finite constant. The result below provides precise control of the bias of the estimation method in this case.

**Lemma 2.1.** *Suppose that conditions (7) and (8) are fulfilled. Then, as  $n \rightarrow +\infty$ , we have:*

$$\ln \left( \frac{L(n^A)}{L(en^A)} \right) = -c|\rho|^{-1}n^{-A|\rho|} \left( U(n^A) - e^{-|\rho|}U(en^A) \right) + o \left( n^{-A|\rho|}U(n^A) \right).$$

In this situation, the bias of the method is thus of order  $O(n^{-A|\rho|})$ , while  $1/\sqrt{np_{k+1}}$  is of order  $O(n^{-(1-A\beta)/2})$ . Hence, if  $1/(\beta + 2|\rho|) \leq A < 1/\beta$ , the conditions of Corollary 2.2 are satisfied with  $k_n = A \ln n$ .

To illustrate this trade-off, we present the following Monte-Carlo experiment: We generate  $10^4$  samples of a heavy-tailed distribution and calculate  $\hat{\beta}_{10^4}(k)$  for all admissible values of  $k$ , we repeat this experiment 100 times, and then we calculate the mean and the 95% confidence interval of  $\hat{\beta}_{10^4}(k)$  for each value of  $k$ . The results of these simulations, for the cases where  $L(n)$  is asymptotically constant and  $L(n)$  is logarithmic, are presented in Figures 2a and 2b. As expected, the behaviour of the estimator is way better in the former case than in the latter.



**Fig. 2:** Monte-Carlo average and 95% confidence interval for  $\hat{\beta}_{10^4}(k)$ , as  $k$  is varying. The true value of  $\beta$  in both cases is 0.5.

### 3 Regular Null-Recurrent Chains - Regularity Index Estimation

We start by setting out the notations used throughout this section, now standard in the Markov chain literature, and listing first the properties supposedly satisfied by the class of Markov chains under study. One may refer to [Meyn, Tweedie, and Glynn \(2009\)](#) for an excellent account of the Markov chain theory. The concept of  $\beta$ -regularity

for describing how fast a Harris chain returns to Harris sets is then recalled, together with related asymptotic properties, invoked in the subsequent statistical analysis, for clarity's sake. Then, the main results of this paper, related to the inference of the parameter  $\beta$  and the extended use of the estimator (2) in the (regenerative) Markovian case, are established and discussed. Here,  $X = (X_n)_{n \in \mathbb{N}}$  denotes a time-homogeneous Markov chain, with state space  $E$ , equipped with a countably generated  $\sigma$ -field  $\mathcal{E}$ , and transition probability  $\Pi(x, dy)$ . For any probability distribution  $\nu$  on  $E$ , we denote by  $\mathbb{P}_\nu$  the probability distribution on the underlying space such that  $X_0 \sim \nu(dx)$  and by  $\mathbb{E}_\nu[\cdot]$  the corresponding expectation. For notational convenience, we shall write  $\mathbb{P}_x$  and  $\mathbb{E}_x[\cdot]$  when  $\nu$  is the Dirac mass at  $x \in E$ . In the following, we also denote by  $t \in \mathbb{R} \mapsto \lfloor t \rfloor$  the floor function and by  $\Gamma(z) = \int_{t \geq 0} t^{z-1} e^{-t} dt$  the Gamma function.

### 3.1 Background and Preliminaries

Throughout the section, we suppose that the chain  $X$  is  $\psi$ -irreducible, meaning that there exists some  $\sigma$ -finite measure  $\psi$  on  $(E, \mathcal{E})$  such that any measurable set  $B \subset E$ , weighted by  $\psi$ , can be reached by the chain with positive probability in a finite number of steps, *i.e.*  $\sum_{n \geq 1} \Pi_n(x, B) > 0$ , no matter the starting point  $x \in E$ , denoting by  $\Pi_n(x, dy)$  the  $n$ -th iterate of the transition probability  $\Pi(x, dy)$ . Recall that an irreducibility measure is said to be *maximal* if it dominates any other irreducibility measure. We also assume that  $X$  is aperiodic (rather than replacing  $\Pi$  by an iterate) and Harris recurrent, *i.e.* that, with probability one, it visits an infinite number of times any measurable subset  $B \subset E$ , weighted by maximal irreducibility measures, whatever the initial state:  $\forall x \in E, \mathbb{P}_x(\sum_{n=1}^{\infty} \mathbb{I}\{X_n \in B\} = \infty) = 1$ . When Harris recurrent, a transition kernel  $\Pi(x, dy)$  has a non zero invariant (positive) measure  $\mu(dx)$  (*i.e.* such that  $\int_{x \in E} \mu(dx) \Pi(x, dy) = \mu(dy)$ ), that is unique up to a multiplicative factor (notice incidentally that  $\mu(dx)$  is a maximal irreducibility measure). Measurable sets weighted by  $\mu$  are said to be Harris. For Harris recurrent chains, recall that the following strong ratio limit theorem holds. We have indeed, as  $n \rightarrow \infty$ ,

$$\frac{\sum_{i=1}^n \mathbb{I}\{X_i \in B\}}{\sum_{i=1}^n \mathbb{I}\{X_i \in C\}} \rightarrow \frac{\mu(B)}{\mu(C)} \quad \mathbb{P}_\nu\text{-almost-surely}, \quad (9)$$

for any initial distribution  $\nu$  and any measurable sets  $B$  and  $C$  s.t.  $\mu(C) > 0$ . When the measure  $\mu(dx)$  is finite, the chain is said to be *positive recurrent* and, by convention, rather than considering  $\mu(dx)/\mu(E)$ , by  $\mu(dx)$  we mean the stationary probability measure in this case.

**Regular chains.** For a wide class of Harris Markov chains, the *regularity index* describes how fast the *occupation time* related to a Harris set  $B$  (*i.e.* the number of visits to  $B$ )  $\Sigma_n(B) = \sum_{i=1}^n \mathbb{I}\{X_i \in B\}$  increases with time  $n$ . When  $X$  is positive recurrent, it follows from the Strong Law of Large Numbers that occupation times of Harris sets grow in a linear fashion with the observation time: as  $n \rightarrow \infty$ ,  $\Sigma_n(B) \sim \mu(B)n$   $\mathbb{P}_\nu$ -almost surely. In the general Harris case, some technical assumptions are required in order to be able to specify the growing rate. In order to formulate them rigorously, further concepts are required. Recall that a *special set* (also referred to as a *D-set* sometimes (Chen 1999)) for the chain  $X$  is any Harris set  $D$  such that

$\mu(D) < \infty$  and  $\sup_{x \in E} \mathbb{E}_x[\sum_{i=1}^{\tau_B} \mathbb{I}\{X_i \in D\}] < \infty$ , for any Harris set  $B \subset E$ , denoting by  $\tau_B = \inf\{n \geq 1 : X_n \in B\}$  the hitting time to  $B$ . We recall that special sets not only exist but there are many of them: actually, any Harris set contains a special set at least, see Proposition 5.13 in Nummelin (1984). For any special set  $D$  and initial distribution  $\nu$ , consider the so-termed *truncated Green function*:  $G_{\nu,D}(t) = (1/\mu(D)) \sum_{n=1}^{\lfloor t \rfloor} \nu \Pi_n(D)$ , where  $\nu \Pi_n(B) = \int_{x \in E} \nu(dx) \Pi_n(x, B) = \mathbb{P}_\nu(X_n \in B)$  for any  $B \in \mathcal{E}$ . Observe that Harris recurrence entails that  $G_{\nu,D}(t) \rightarrow +\infty$  as  $t \rightarrow \infty$ . In the following, we restrict our attention to a specific class of Harris chains for which the rate at which  $G_{\nu,D}(t)$  grows to infinity as  $t \rightarrow \infty$  can be characterized. Notice that, in such cases, the rate would be independent from the pair  $(\nu, D)$ . Indeed, by virtue of Theorem 7.3 in Nummelin (1984), we have  $G_{\nu_1,D_1}(t)/G_{\nu_2,D_2}(t) \rightarrow 1$  as  $t$  goes to infinity, for any distributions  $\nu_1$  and  $\nu_2$  and any special sets  $D_1$  and  $D_2$ . One may thus give the following definition, see Chen (1999, 2000).

**Definition 1** ( $\beta$ -REGULAR MARKOV CHAIN). *Let  $\beta \in [0, 1]$ . A Harris chain  $X$  is said to be  $\beta$ -regular if there exists a special set  $D$  and a distribution  $\nu$  such that the function  $G_{\nu,D}$  is  $\beta$ -regularly varying:  $\forall t > 0$ ,*

$$\lim_{\lambda \rightarrow \infty} \frac{G_{\nu,D}(\lambda t)}{G_{\nu,D}(\lambda)} = t^\beta. \quad (10)$$

We point out that property (10) can be rephrased as follows: there exists a slowly varying function  $L_{\nu,D}(t)$  such that  $G_{\nu,D}(t) = L_{\nu,D}(t)t^\beta$ . Notice incidentally that “ $\beta$ -regularity” is called “ $\beta$ -null recurrence” in Karlsen and Tjøstheim (2001) when  $\beta < 1$ , while  $\beta = 1$  corresponds to the positive recurrent case. The parameter  $\beta$  thus rules the ‘frequency’ at which a (supposedly regular) Harris chain  $X$  recurs (Chen 1999) and it is the purpose of this section to investigate the issue of estimating it with asymptotic guarantees, based on the observation of a single path  $X_1, \dots, X_n$  of size  $n \rightarrow +\infty$ . In particular, we shall focus in subsection 3.3 on the case of *regenerative* chains, for which an extension of the estimator (2) can be used with statistical guarantees.

**Regenerative regular chains.** Recall that a Markov chain is *regenerative* when it possesses an accessible atom, *i.e.* a measurable set  $A$  such that  $\psi(A) > 0$  and  $\Pi(x, \cdot) = \Pi(y, \cdot)$  for all  $(x, y) \in A^2$ . By  $\tau_A = \tau_A(1) = \inf\{n \geq 1, X_n \in A\}$  is meant the hitting time to  $A$  and we denote by  $\tau_A(j) = \inf\{n > \tau_A(j-1), X_n \in A\}$ , for  $j \geq 2$ , the successive return times to  $A$ , by  $\mathbb{P}_A$  the probability measure on the underlying space such that  $X_0 \in A$  and by  $\mathbb{E}_A[\cdot]$  the  $\mathbb{P}_A$ -expectation. In the regenerative case, it results from the *strong Markov property* that the blocks of observations in between consecutive visits to the atom

$$\mathcal{B}_1 = (X_{\tau_A(1)+1}, \dots, X_{\tau_A(2)}), \dots, \mathcal{B}_j = (X_{\tau_A(j)+1}, \dots, X_{\tau_A(j+1)}), \dots \quad (11)$$

form a collection of i.i.d. random variables, taking their values in the torus  $\mathbb{T} = \bigcup_{n=1}^{\infty} E^n$ , and the sequence  $\{\tau_A(j)\}_{j \geq 1}$ , corresponding to successive times at which the chain forgets its past is a (possibly delayed) renewal process. Incidentally, we point out that the class of regenerative chains is not that restrictive. Indeed, it includes all chains with a countable state space (any recurrent state is then an accessible atom), as well as

numerous Markov models used in the field of operations research, refer to *e.g.* [Asmussen \(2010\)](#). Examples 1 and 3 below also provide examples of (regular) regenerative chains. In the regenerative setting, all stochastic stability properties may be expressed in terms of speed of return to the atom. For instance, when  $X$  is *Harris recurrent*, see Theorem 10.0.1 in [Meyn et al. \(2009\)](#), the invariant distribution is equal to the occupation measure between two consecutive visits to the atom (up to a multiplicative factor):  $\forall B \in \mathcal{E}, \mu(B) \sim \mathbb{E}_A [\sum_{i=1}^{\tau_A} \mathbb{I}\{X_i \in B\}]$ . For instance, the chain is positive recurrent if and only if the expected return time to the atom is finite<sup>3</sup>, *i.e.*  $\mathbb{E}_A[\tau_A] < \infty$ , see Theorem 10.2.2 in [Meyn et al. \(2009\)](#). More generally, the  $\beta$ -regularity property can be characterized by the heaviness of the tail of the probability distribution of the regeneration times in the atomic case, as the following result shows.

**Proposition 3.1.** ([Karlsen and Tjøstheim \(2001\)](#), Theorem 3.1) *Suppose that  $X$  is regenerative Harris recurrent. Let  $A$  be an atom for  $X$  and  $\beta \in [0, 1]$ . The following assertions are equivalent.*

- (i) *The chain  $X$  is  $\beta$ -regular.*
- (ii) *There exists a slowly varying function  $L_A : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  such that:  $\forall n \geq 1$ ,*

$$\mathbb{P}_A(\tau_A \geq n) = L_A(n) \cdot n^{-\beta}. \quad (12)$$

As a direct consequence of Proposition 3.1, we have that if  $X$  is regenerative and  $\beta$ -regular, then  $\beta = \sup\{\theta \in [0, 1] : \mathbb{E}_A[\tau_A^\theta] < +\infty\}$ . Based on the decomposition (11) of the whole sample path, limit theorems for regenerative Markov chains can be derived from the application of their i.i.d. counterparts to the sequence of blocks  $(\mathcal{B}_k)_{k \geq 1}$ , see *e.g.* [Meyn et al. \(2009\)](#). This approach is usually referred to as the *regenerative method* and is extensively used to establish the asymptotic results stated in subsection 3.2 in the atomic case. Notice however that the regenerative blocks  $\mathcal{B}_1, \dots, \mathcal{B}_{N_n}$ , where  $\Sigma_n(A) = N_n - 1$  denotes the (random) number of regenerations before time  $n$ , forming the truncated trajectory up to time  $n$  are not independent (the sum of their length being less than  $n$  in particular), which causes technical difficulties when establishing higher-order or non-asymptotic results, see *e.g.* [Bertail and Cl  men  on \(2006b\)](#) or [Bertail and Cl  men  on \(2004\)](#) and the references therein. The inference technique for the regularity index  $\beta$  of the chain  $X$  developed in subsection 3.3 is based on characterization (ii): the parameter  $\beta$  is the tail index of a discrete generalized Pareto r.v., the regeneration time namely, *i.e.* the conditional survivor function of  $\tau_A$  given  $X_0 \in A$ . Incidentally, notice that the parameter  $\beta$  does not depend on the atom  $A$  considered (in contrast to the estimator analysed in subsection 3.3). Based on a (random) number  $N_n$  of (dependent) realizations of the regenerative time, namely

$$S_j = \tau_A(j+1) - \tau_A(j) \text{ for } j = 1, \dots, N_n,$$

one may naturally compute the estimator (2). As will be shown, in spite of the dependence structure between the  $S_j$ 's, the consistency property is preserved in the Markovian framework.

---

<sup>3</sup>Its (unique) invariant probability distribution  $\mu$  is then given by  $\mu(B) = (1/\mathbb{E}_A[\tau_A])\mathbb{E}_A[\sum_{i=1}^{\tau_A} \mathbb{I}\{X_i \in B\}]$ , for all  $B \in \mathcal{E}$ .

**Pseudo-regeneration.** Harris chains are not necessarily regenerative, of course. However, the construction proposed in Nummelin (1984), referred to as the *Nummelin splitting technique*, permits to build a regenerative extension of any Harris chain. We briefly recall it for clarity. It relies crucially on the notion of *small set*: a set  $K \in \mathcal{E}$  is said to be *small* if there exist  $m \in \mathbb{N}^*$ ,  $\delta > 0$  and a probability measure  $\Phi$  supported by  $K$  such that

$$\forall (x, B) \in K \times \mathbb{E}, \quad \Pi_m(x, B) \geq \delta \Phi(B). \quad (13)$$

We refer to (13) as the minorization condition  $\mathcal{M}(m, K, \delta, \Phi)$ . Recall that accessible small sets always exist for  $\psi$ -irreducible chains: any set  $B \in \mathcal{E}$  such that  $\psi(B) > 0$  contains such a set, see Jain and Jamison (1967). Suppose that  $X$  satisfies  $\mathcal{M} = \mathcal{M}(m, K, \delta, \Psi)$  for  $K \in \mathcal{E}$  s.t.  $\psi(K) > 0$ . Rather than replacing the initial chain  $X$  by the chain  $\{(X_{nm}, \dots, X_{n(m+1)-1})\}_{n \in \mathbb{N}}$ , we suppose  $m = 1$ . The sample space is expanded so as to define a sequence  $(Y_n)_{n \in \mathbb{N}}$  of independent Bernoulli r.v.'s with parameter  $\delta$  by defining the joint distribution,  $\mathbb{P}_{\nu, \mathcal{M}}$  whose construction relies on the following randomization of the transition probability  $\Pi$  each time the chain hits  $K$ . Note that it occurs with probability one, since the chain is Harris recurrent and  $\psi(K) > 0$ . If  $X_n \in K$ , and if  $Y_n = 1$  (this occurs with probability  $\delta \in ]0, 1[$ ), then  $X_{n+1} \sim \Phi$ , while, if  $Y_n = 0$ , we have  $X_{n+1} \sim (1 - \delta)^{-1}(\Pi(X_n, \cdot) - \delta \Phi(\cdot))$ . Let  $\text{Ber}_\delta$  be the Bernoulli distribution with parameter  $\delta$ . The *split chain*  $\{(X_n, Y_n)\}_{n \in \mathbb{N}}$  is valued in  $E \times \{0, 1\}$  and has transition kernel  $\Pi_{\mathcal{M}}$

- for any  $x \notin K$ ,  $B \in \mathcal{E}$ ,  $b$  and  $b'$  in  $\{0, 1\}$ ,

$$\Pi_{\mathcal{M}}((x, b), B \times \{b'\}) = \Pi(x, B) \times \text{Ber}_\delta(b'),$$

- for any  $x \in K$ ,  $B \in \mathcal{E}$ ,  $b'$  in  $\{0, 1\}$ ,

$$\begin{cases} \Pi_{\mathcal{M}}((x, 1), B \times \{b'\}) = \Phi(B) \times \text{Ber}_\delta(b'), \\ \Pi_{\mathcal{M}}((x, 0), B \times \{b'\}) = (1 - \delta)^{-1}(\Pi(x, B) - \delta \Phi(B)) \times \text{Ber}_\delta(b'). \end{cases}$$

The key point of the construction relies on the fact that  $A_K = K \times \{1\}$  is an atom for the bivariate chain  $X^{\mathcal{M}} = (X, Y)$ , which inherits all its communication and stochastic stability properties from  $X$ .

Recall also that conditions of type (13) can be replaced by Foster-Lyapunov drift conditions that are much more tractable in practice, see *e.g.* Chapter 11 in Meyn et al. (2009). The construction above permits the extension of probabilistic results established for regenerative chains to general recurrent Harris chains. In particular, we have the following result, presented in (Chen 1999, pp 19).

**Proposition 3.2.** *Let  $X$  be a  $\beta$ -regular chain, with  $\beta \in [0, 1]$ . Suppose that condition  $\mathcal{M} = \mathcal{M}(1, K, \delta, \Psi)$  is fulfilled, then, the split chain  $X^{\mathcal{M}}$  is  $\beta$ -regular.*

Hence,  $X$ 's regularity index  $\beta$  is the regular variation index of the conditional survivor function of the hitting time  $\tau_{A_K} = \inf\{n \geq 1 : (X_n, Y_n) \in A_K\}$  given  $(X_0, Y_0) \in A_K$ . However, the return times to  $A_K$  are not observable, just like the sample path of the Nummelin extension, and cannot be straightforwardly exploited from a statistical perspective. We shall explain in subsection 3.4 how estimators tailored to the regenerative case can be nevertheless extended to the pseudo-regenerative case

in practice by means of the *plug-in* approximation procedure originally proposed in Bertail and Cl  men  on (2006b) in the positive recurrent case.

In the next section, we explain why the (asymptotic) behaviour of the occupation times  $\Sigma_k(B)$  with  $k \leq n$  and  $B \in \mathcal{E}$ , though ruled by the parameter  $\beta$ , can be hardly directly used for inference.

Before that, we illustrate the fact that the class of (regenerative) regular Markov chains includes many stochastic processes used in probabilistic modelling by a few examples.

**Example 1.** (BESSEL RANDOM WALKS) *Recall that a Bessel random walk with drift  $\delta \in [-1, +\infty)$  is a Markov chain with  $\mathbb{N}$  as state space, jumps in  $\{-1, +1\}$ , reflecting at 0 and with transition probabilities of the form:  $\Pi(0, +1) = +1$  and, for all  $k \geq 1$ ,  $1 - \Pi(k, k-1) = \Pi(k, k+1) = (1 + h(k) - \delta/(2k))/2$ , where  $h(k) \in (-1 + \delta/(2k), 1 + \delta/(2k))$  and  $h(k) = o(1/k)$  as  $k \rightarrow +\infty$ . It is recurrent when  $\delta > -1$ , positive recurrent when  $\delta > 1$  and transient when  $\delta = -1$ . For  $\delta = 1$ , it is either recurrent or else transient, depending on the function  $h(x)$ . In the null recurrent case, the chain is  $\beta$ -regular with  $\beta = (1 + \delta)/2$ , see Theorem 2.1 in Alexander (2011). Of course, when  $\delta = 0$  and  $h \equiv 0$ , this chain corresponds to a simple reflected random walk with  $p = 1/2$ . As shown by the next example, this random walk is not the sole regular chain with an index equal to  $1/2$ .*

**Example 2.** (THRESHOLD AUTOREGRESSIVE MODEL) *Let  $\alpha_1 \in \mathbb{R}$  and  $K$  be a compact subset of  $\mathbb{R}$  (typically,  $K = (-\infty, \tau]$  with  $\tau \in \mathbb{R}$ ). Consider the parametric threshold autoregressive (TAR) model:  $X_0 = 0$  and  $\forall n \geq 1$ ,  $X_n = \alpha_1 X_{n-1} \mathbb{I}\{X_{n-1} \in K\} + X_{n-1} \mathbb{I}\{X_{n-1} \notin K\} + \varepsilon_n$ , where  $\{\varepsilon_n\}_{n \geq 1}$  is an i.i.d sequence of centred random variables such that  $\mathbb{E}[\varepsilon_n^4] < \infty$ ,  $\varepsilon_n$  is independent of  $X_k$  for all  $k < n$ . In addition, the distribution of the  $\varepsilon_n$ 's is absolutely continuous with density  $f_0$  supposed to be bounded away from zero on any compact set. It is proved in Gao et al. (2013) (see Lemma 2.1 therein) that  $(X_n)$  is a regular Markov chain with index  $1/2$ .*

**Example 3.** (NULL RECURRENT, NOT NECESSARILY REGULAR, CHAINS) *By means of the model below, originally presented in Myklebust et al. (2012), one can generate  $\beta$ -null recurrent chains for any  $\beta > 0$ , as well as null recurrent chains that are not regular. Let  $\{\eta_n\}_{n \in \mathbb{N}}$  be a sequence of i.i.d. real-valued random variables. Consider the chain defined by:  $\forall n \geq 1$ ,  $X_n = (X_{n-1} - 1) \mathbb{I}\{X_{n-1} > 1\} + \eta_n \mathbb{I}\{X_{n-1} \in [0, 1]\}$ . The chain  $X$  is regenerative, with the interval  $[0, 1]$  as atom. In addition, we have  $\mathbb{P}_x(\tau_{[0,1]} > n) = \mathbb{P}(\lfloor \eta_1 \rfloor > n)$ . Hence,  $X$  is null recurrent iff  $\mathbb{E}[\eta_1] = \infty$  and  $\beta$ -regular with  $\beta \in (0, 1]$  iff the r.v.  $\lfloor \eta_1 \rfloor$  has generalized discrete Pareto distribution with tail index  $\beta$ .*

### 3.2 Limit Theorems for Regular Markov Chains

We now recall the limit results related to the behaviour of the random occupation times  $\Sigma_n(\cdot)$  for regular Markov chains and discuss their limitations regarding their possible use to infer the regularity index  $\beta$  with (asymptotic) guarantees. The latter essentially reveals that the empirical occupation measures  $\Sigma_n(B)$  of Harris sets  $B$  grow at the sublinear rate  $n^\beta$  (up to a slowly varying factor). As shall be seen, however, due to the great dispersion of their (asymptotic) distribution, the empirical occupation measures can hardly be used directly to estimate the key parameter  $\beta$ . The following result corresponds to Theorem 2.4 in Chen (1999) (see equation (4.2) therein) when

specialized to  $\beta$ -regular chains (assertion (i)) and to Lemma 3.4 (assertion (ii)) in [Karlsen and Tjøstheim \(2001\)](#).

**Theorem 3.3** (Strong boundedness). *Suppose that the chain  $X$  is regular with index  $\beta \in [0, 1]$  and has initial distribution  $\nu$ . The following assertions hold true.*

- (i) *Let  $D$  be any special set. Then, there exists a positive constant  $\bar{c}$  such that, we  $\mathbb{P}_\nu$ -almost surely have:*

$$\limsup \frac{\Sigma_n(B)}{n^\beta K_{\nu,D}(n)} = \bar{c}\mu(B), \quad (14)$$

*for all Harris set  $B$  with a finite and strictly positive  $\mu$ -measure, where  $K_{\nu,D}(n) = L_{\nu,D}(n / \ln \ln G_{\nu,D}) (\ln \ln G_{\nu,D})^{1-\beta}$ .*

- (ii) *If  $\beta \in (0, 1)$ , for any  $\epsilon > 0$ , we  $\mathbb{P}_\nu$ -almost surely have:*

$$\lim_{n \rightarrow +\infty} \frac{\Sigma_n(B)}{n^{\beta+\epsilon}} = 0 \quad \text{and} \quad \lim_{n \rightarrow +\infty} \frac{n^{\beta-\epsilon}}{\Sigma_n(B)} = 0. \quad (15)$$

Incidentally, it is worth noticing that deriving a similar result to (14) for the limit inferior is not straightforward. The established result for the limit superior, which is presented in a more general form involving random integrals of measurable functions  $f : E \rightarrow \mathbb{R}$  with respect to  $(1/n) \sum_{i \leq n} \delta_{X_i}$  in Theorem 2.3 of [Chen \(1999\)](#), only applies to positive functions  $f(x)$ . A key part of its proof relies on Borel–Cantelli’s first lemma and the fact that the distribution of  $\Sigma_n(B)/(n^\beta K_{\nu,D}(n))$  satisfies the Large Deviation Principle, as detailed on pages 16-17 of [Chen \(1999\)](#).

The result stated below claims that the logarithm of the occupation time of any Harris set provides a strongly consistent estimator of the regularity index  $\beta$  when appropriately normalized. It corresponds to the comment in Remark 3.7 of [Karlsen and Tjøstheim \(2001\)](#) and follows immediately from (15).

**Corollary 3.1.** *Suppose that the chain  $X$  is  $\beta$ -regular with  $\beta \in (0, 1)$  and  $B$  is a Harris set. Let  $\nu$  be its initial probability distribution. Then, we have:*

$$\tilde{\beta}_n(B) \rightarrow \beta \quad \mathbb{P}_\nu - \text{a.s.}, \text{ as } n \rightarrow +\infty, \quad (16)$$

where

$$\tilde{\beta}_n(B) = \ln(\Sigma_n(B)) / \ln n. \quad (17)$$

It was pointed out in Remark 3.7 of [Karlsen and Tjøstheim \(2001\)](#) that this estimator is of limited practical use due to its slow rate of convergence, although no specific rate was given therein. Equation (14) may suggest that  $|\tilde{\beta}_n(B) - \beta|$  is almost-surely  $O(\ln K_{\nu,D}(n) / \ln n)$  as  $n \rightarrow +\infty$ . However, this has not been proven to the best of our knowledge, due to the lack of an equivalent of (14) for the limit inferior, as recalled above. The description of the limit distribution of the strongly consistent estimator  $\tilde{\beta}_n(B)$  relies on the asymptotic distribution of the (appropriately standardized) occupation measure, which is given in the following result, see ([Chen 1999](#), Theorem 2.3).



**Theorem 3.4** (Limit distributions). *Let  $\beta \in [0, 1)$  and  $\nu$  be any probability distribution on  $E$ . Suppose that the chain  $X$  is  $\beta$ -regular and let  $B$  be a Harris set with finite and strictly positive  $\mu$ -measure. We have the following convergences in  $\mathbb{P}_\nu$ -distribution.*

(i) *If  $\beta = 0$ , we then have, as  $n \rightarrow \infty$ ,*

$$\frac{1}{L_{\nu,D}(n)} \Sigma_n(B) \Rightarrow \mathcal{E}(1/\mu(B)) \text{ in } \mathbb{P}_\nu\text{-distribution,}$$

*where  $\mathcal{E}(\lambda)$  denotes the exponential distribution with mean  $1/\lambda > 0$ .*

(ii) *If  $\beta \in (0, 1)$ , we have, as  $n \rightarrow \infty$ ,*

$$\frac{1}{n^\beta L_{\nu,D}(n)} \Sigma_n(B) \Rightarrow \mu(B)/(Z_\beta)^\beta \text{ in } \mathbb{P}_\nu\text{-distribution,}$$

*where  $Z_\beta$  is a stable random variable with Laplace transform*

$$\psi_\beta(t) = \exp(-t^\beta/\Gamma(\beta+1)), \quad t \geq 0.$$

When  $\beta \in (0, 1)$ , a direct application of the Continuous Mapping Theorem provides the limit distribution of the estimator  $\tilde{\beta}_n(B)$  under the additional assumption that  $L_{\nu,D}$  has a finite non-zero limit. In this case, this estimator has a logarithmic rate of convergence, as revealed by the result stated below.

**Corollary 3.2.** *Suppose that the chain  $X$  is  $\beta$ -regular with  $\beta \in (0, 1)$ ,  $B$  is a Harris set, and  $\nu$  is its initial probability distribution. Then, as  $n \rightarrow +\infty$ ,*

$$\ln(n) \left( \tilde{\beta}_n(B) - \beta - \frac{\ln L_{\nu,D}}{\ln n} \right) \Rightarrow \ln(\mu(B)/(Z_\beta)^\beta) \text{ in } \mathbb{P}_\nu\text{-distribution.}$$

*In addition, if  $\lim_{n \rightarrow +\infty} L_{\nu,D}(n)$  exists and is not 0, then, there exists a constant  $\kappa > 0$  such that*

$$\ln(n) \left( \tilde{\beta}_n(B) - \beta \right) \Rightarrow \ln(\kappa/(Z_\beta)^\beta) \text{ in } \mathbb{P}_\nu\text{-distribution.}$$

We also establish a functional version of the preceding theorem. Let  $n \geq 1$ , and define the step function

$$\sigma_n(B) : t \geq 0 \mapsto \frac{\Sigma_{\lfloor nt \rfloor}(B)}{n^\beta L_{\nu,D}(n)}. \quad (18)$$

The result stated below is new and describes the asymptotic distribution of the process  $\sigma_n(B)$ . See Theorem 17.4.4 in [Meyn et al. \(2009\)](#) for an analogous result in the positive recurrent case. We denote by  $M_\beta = (M_\beta(t))_{t \geq 0}$  the Mittag-Leffler process with parameter  $\beta \in (0, 1)$ , defined by:

$$\begin{aligned} \mathbb{E}[(M_\beta(1))^m] &= \frac{m!}{\Gamma(1+m\beta)}, \quad \text{for all } m \geq 0, \\ M_\beta(t) &\stackrel{d}{=} t^\beta M_\beta(1), \quad \text{for all } t \geq 0. \end{aligned}$$



The characteristic functions describing the marginal distributions are given by (see 3.39 in [Karlsen and Tjøstheim \(2001\)](#))

$$\mathbb{E} \left[ e^{i\zeta M_\beta(t)} \right] = \sum_{k=0}^{+\infty} \frac{(i\zeta t^\beta)^k}{\Gamma(1+k\beta)}, \quad \zeta \in \mathbb{R}, \quad t \geq 0. \quad (19)$$

**Theorem 3.5** (Functional Limit Theorem). *Let  $\beta \in (0, 1)$  and  $\nu$  be any probability distribution. Suppose that the chain  $X$  is  $\beta$ -regular and  $B$  is a Harris set. Then, as  $n \rightarrow \infty$ , we have:*

$$\sigma_n(B) \Rightarrow \mu(B)\Gamma(1+\beta)M_\beta \text{ in } \mathbb{P}_\nu\text{-distribution,}$$

*in the sense of Skorokhod topology.*

**Remark 2.** Let  $Y = 1/Z_\beta^\beta$  where  $Z_\beta$  is as in Theorem 3.4. By virtue of equation (8.3) in page 453 of [Feller \(1971\)](#), the Laplace transform of  $Y$  is

$$\mathbb{E} [e^{-sY}] = \sum_{k=0}^{+\infty} \frac{(-\Gamma(1+\beta)s)^k}{\Gamma(1+k\beta)},$$

*which equals the Laplace transform of  $\Gamma(1+\beta)M_\beta(1)$ , cf (19).*

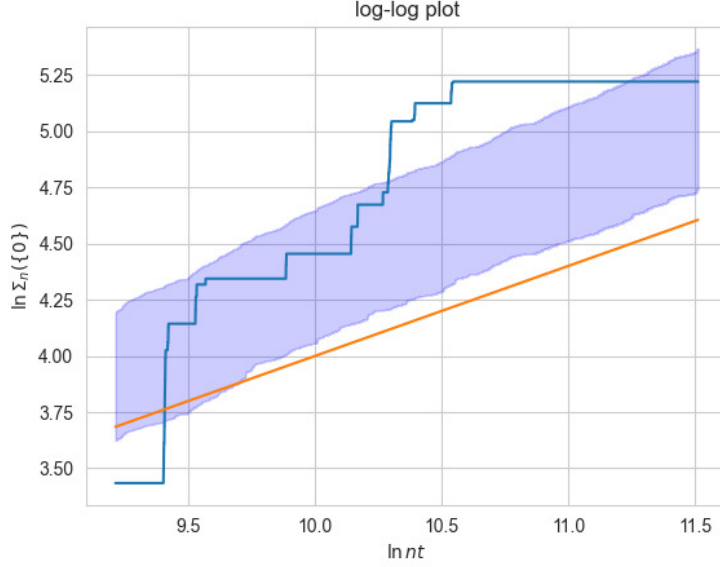
The almost sure convergence of  $\tilde{\beta}_n(B)$  towards  $\beta$  suggests that, for  $n$  large enough,  $\ln \Sigma_n(B) \approx \beta \ln(n)$  and that the log-log plot of  $\Sigma_n(B)$  and  $n$  should look like a linear function with slope  $\beta$ , which could be possibly used to infer the value of  $\beta$ . Unfortunately, the dispersion of such a plot (and that of the process  $\sigma_n(B)$ , asymptotically described by Theorem 3.5) is way too large in practice. To illustrate this, we simulated a Simple Symmetric Random Walk ( $\beta = 0.5$ ) with  $n = 10^5$  points, and we computed  $\ln \Sigma_{\lfloor nt \rfloor}(B)$  for  $0.1 \leq t \leq 1$  (choosing  $B = \{0\}$ , which is an atom for this regenerative regular chain). The outcomes of this simulation are depicted in Fig. 3.

This simulation illustrates in particular the slow convergence of  $\tilde{\beta}_n(B)$  described in Corollary 3.2. Hence, in the regenerative case, it is more suitable to exploit the tail behaviour of the regenerative times (cf Proposition 3.1) to estimate the regularity index  $\beta$ , as shall be investigated in the next subsection.

### 3.3 Regularity Index of a Regular Chain - Statistical Inference

Assume that  $X$  is a regenerative regular chain with atom  $A$  and unknown regularity index  $\beta \in (0, 1)$ , and suppose that a sample path  $X_1, \dots, X_n$  of length  $n \geq 1$  is observed. Because the chain is Harris recurrent, the number of observed regeneration times  $\Sigma_n(A)$  almost-surely tends to  $+\infty$  as  $n \rightarrow +\infty$ . Hence, with probability 1, we have  $N_n \geq 1$  for  $n$  large enough and one can define

$$\hat{p}_l^{(N_n)} = \frac{1}{N_n} \sum_{i=1}^{N_n} \mathbb{I}\{S_i > e^l\} \text{ for any } l \in \mathbb{R} \quad (20)$$



**Fig. 3:** Log-log plot of  $\ln \Sigma_{[nt]}(\{0\})$  in dark blue, the orange line representing the plot of the linear function  $x \mapsto 0.5 \times x$ , while the blue area represents the 95% confidence interval for  $\ln \Sigma_{[nt]}(\{0\})$  calculated from 100 independent trajectories.

and form the statistic

$$\hat{\beta}_{N_n}(k) = \ln \left( \hat{p}_k^{(N_n)} / \hat{p}_{k+1}^{(N_n)} \right), \quad (21)$$

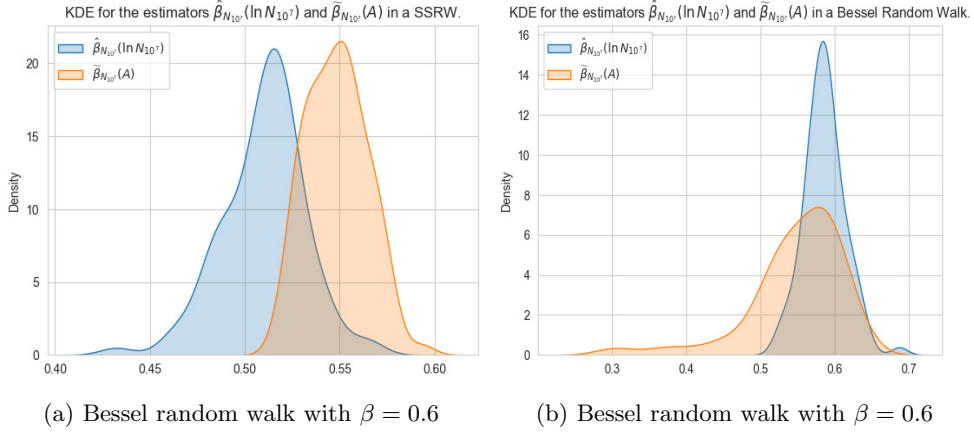
provided that  $\hat{p}_{k+1}^{(N_n)} > 0$ . We point out that, due to the randomness of  $N_n$ , (20) is a biased (strongly consistent) estimator of  $p_l = \mathbb{P}_A(\tau_A > e^l)$  for any  $l \in \mathbb{R}$ . The estimator (21) is of the same form as (2) except that the number  $N_n$  of observations is random, and the observations are not independent anymore (in particular  $S_1 + \dots + S_{N_n} \leq n$ ). **Theorem 3.6.** *Suppose that the atomic chain  $X$  is  $\beta$ -regular with  $\beta \in (0, 1)$ . Let  $\nu$  be its initial probability distribution. If, as  $n \rightarrow +\infty$ , we  $\mathbb{P}_\nu$  almost-surely have  $k_n \rightarrow +\infty$  so that  $(\ln n) \exp(k_n \beta) / n = o(L_A(\exp(k_n)))$ , the estimator (21) is strongly consistent:*

$$\hat{\beta}_{N_n}(k_{N_n}) \rightarrow \beta \quad \mathbb{P}_\nu - a.s. \text{ as } n \rightarrow +\infty. \quad (22)$$

In particular, strong consistency holds for  $\hat{\beta}_{N_n}(\ln N_n)$ .

In order to show the finite sample behaviour of our estimator in the Markovian scenario, we have simulated 100 independent trajectories of  $10^7$  points for two Markov chains (a Simple Symmetric Random Walk where  $\beta = 0.5$  and a Bessel Random Walk with  $\beta = 0.6$ ), and we have computed  $\hat{\beta}_{N_{10^7}}(\ln N_{10^7})$  for each one of the trajectories. As a comparison, we also include the estimations using the statistic  $\tilde{\beta}_{10^7}$  described in (17). Figure 4 shows the KDE for both estimators.

**Remark 3.** (ON INVESTIGATING CONVERGENCE RATES) *Due to the impossibility of tightly controlling the sequence  $N_n$  by a deterministic quantity in probability and the*



**Fig. 4:** KDE for the estimators  $\hat{\beta}_{N_{10^7}}(\ln N_{10^7})$  and  $\tilde{\beta}_{N_{10^7}}(A)$  in different  $\beta$ -regular chains.

non-linearity of the estimator, we have not been able to extend to the Markovian case the asymptotic normality results of Theorem 2.3 and Corollary 2.2 via Anscombe's theorem (Gut 2013, Theorem 7.3.2). Heuristically, if in Theorem 2.3 we take  $k_n = \ln N_n$  and replace  $N_n$  by its approximate expectation  $n^\beta L_{\nu,A}(n)$  (Karlsen & Tjøstheim 2001, Lemma 3.3), we would get a convergence rate of order  $n^{-\beta(1-\beta)/2} L_1(n)$ , where  $L_1(n)$  is the slowly varying function given by  $\sqrt{L_{\nu,A}(n^\beta L_{\nu,A}(n)) / L_{\nu,A}(n)^{1-\beta}}$ . This suggests a convergence rate of order  $n^{-\beta(1-\beta)/2}$  when  $L_{\nu,A} \sim C > 0$ . However, we have not been able to prove this claim.

**Remark 4.** (TRAJECTORIES OF RANDOM LENGTH) Suppose that the trajectory is observed until the  $N$ -th regeneration, i.e.  $n = \tau_A(N)$ , with  $N \geq 2$ . In this case, we will obtain a sequence of  $N$  i.i.d blocks whose sizes follow the heavy-tailed distribution described in (1), and therefore, the results of Section 2.1 can be applied directly to this sequence. Notice that in this case, the total number of points observed in the chain (i.e. the amount of time we need to wait to collect the  $N$  blocks) is a random variable that, while finite with probability one, has infinite expectation.

**Remark 5.** (THE (ATOMIC) POSITIVE RECURRENT CASE) When the chain is positive recurrent (or equivalently 1-regular), the estimator (21) can be naturally used to estimate the tail index  $\beta' \geq 1$  of the regeneration time, when the latter has a regularly varying distribution. Dedicated theoretical results can be found in section A.8 of the Appendix.

### 3.4 Perspectives - Extension to the Pseudo-regenerative Case

Proposition 3.2 and the algorithmic construction described after Eq. (13) guarantee that if the chain satisfies the minorization condition  $\mathcal{M} = \mathcal{M}(1, K, \delta, \Psi)$ , and  $K, \delta$  and  $\Psi$  are known, then we can generate samples of the split chain, which is atomic and has the same  $\beta$  as the original chain. Assume the existence of a  $\sigma$ -finite measure  $\lambda$  of reference on  $(E, \mathcal{E})$  that dominates the conditional probability measures  $\Pi(x, dy)$ ,  $x \in E$ , and the initial distribution  $\nu$ :  $\Pi(., dy) = \pi(., y)\lambda(dy)$  and  $\nu(dy) = g(y)\lambda(dy)$ .

Notice incidentally that the measure  $\Psi$  involved in  $\mathcal{M}$  is then absolutely continuous w.r.t.  $\lambda$  as well:  $\Psi(dy) = \psi(y)\lambda(dy)$  and then  $\pi(x, y) \geq \delta\psi(y)$  for all  $(x, y) \in K^2$ . As shown in Section 3.2 in Bertail and Clémenton (2006b), given  $X^{(n+1)} = (X_1, \dots, X_{n+1})$ , samples from the distribution of  $Y^{(n)} = (Y_1, \dots, Y_n)$  can be obtained as follows. From  $i = 1$  to  $n$ , the r.v.  $Y_i$  is drawn from a Bernoulli distribution with parameter  $\delta$ , unless  $X$  hits the small set  $K$  at time  $i$ : in the latter case,  $Y_i$  is drawn from a Bernoulli distribution with parameter  $\delta\psi(X_{i+1})/\pi(X_i, X_{i+1})$ . Given that  $A_K = K \times \{1\}$  is an atom for the split chain, and the statistics under study in this paper only depends on the size of the regeneration blocks, sampling  $Y_i$  when  $X_i \in K$  is sufficient here. The accuracy of our estimator improves as the (random) number of samples (the number of regeneration blocks namely) increases. This number is influenced by the size of the chosen small set and how frequently the chain visits it in a finite-length trajectory. It is also affected by the sharpness of the lower bound in the minorization condition. Essentially, there is a trade-off that can be described as follows. Increasing the size of the small set  $K$  used for constructing the pseudo-blocks naturally increases the number of time points that could determine a block (or a cut in the trajectory). However, it also reduces the probability of cutting the trajectory, as the uniform lower bound for  $\pi(x, y)$  over  $K^2$  then decreases. This suggests a criterion for selecting the small set  $K$ : choose a small set that maximizes the maximum expected number of data blocks given the trajectory, that is

$$N_n(K) = \mathbb{E}_\nu \left[ \sum_{i=1}^n \mathbb{I}\{X_i \in K, Y_i = 1\} \mid X^{(n+1)} \right].$$

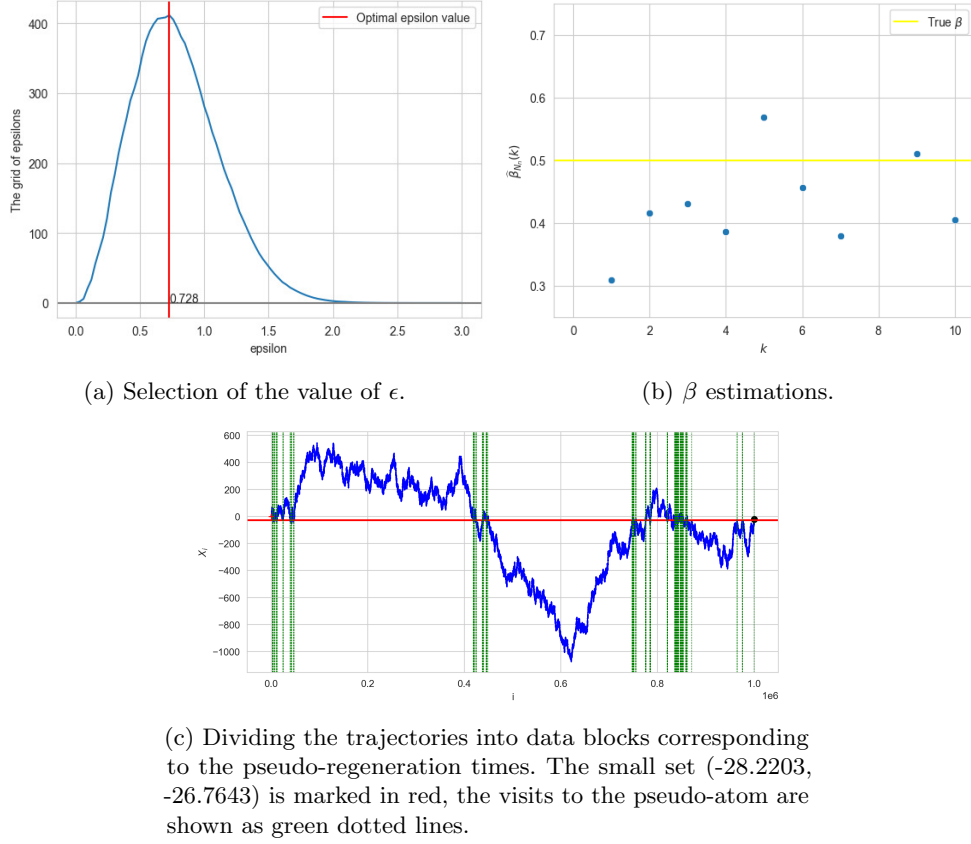
In Section 3.6 of Bertail and Clémenton (2006b), a data-driven approach to select the small set is proposed for the cases where the chain takes real values. The idea relies on the fact that, in many cases, for a well-chosen  $x_0$  and  $\epsilon$  small enough, certain intervals  $V_{x_0, \epsilon} = [x_0 - \epsilon, x_0 + \epsilon]$  are small sets, with the minorization measure  $\Psi$  being the Lebesgue measure on  $V_{x_0, \epsilon}$ . Given a point  $x_0$  (generally taken as the mean or the median of the  $X_i$ 's), the proposed algorithm finds the value of  $\epsilon$  that maximizes the expected number of regeneration blocks, that is

$$N_n(V_{x_0, \epsilon}) = \frac{\delta(V_{x_0, \epsilon})}{2\epsilon} \sum_{i=1}^n \frac{\mathbb{I}\{(X_i, X_{i+1}) \in V_{x_0, \epsilon}^2\}}{\pi(X_i, X_{i+1})},$$

where  $\delta(V_{x_0, \epsilon}) = 2\epsilon \inf_{(x, y) \in V_{x_0, \epsilon}^2} \pi(x, y)$ . Then, the samples of the split chain can be obtained by following the procedure described at the beginning of this subsection with  $K = V_{x_0, \epsilon}$ ,  $\delta = 2\epsilon \inf_{(x, y) \in V_{x_0, \epsilon}^2} \pi(x, y)$  and  $\psi(y) = 1/(2\epsilon)$ .

A practical example can be built by considering a random walk in  $\mathbb{R}$  defined by  $X_0 = 0$ ,  $X_{n+1} = X_n + Z_n$  for  $n \geq 1$ , where  $Z_n$  is a sequence i.i.d. standard normal random variables. The density of the kernel of this Markov chain is given by  $\pi(x, y) = f(x - y)$ , whereby  $f$  is meant the density of a standard normal random variable. Using the data-driven construction described above, we have generated the split chain for the random walk and then applied our estimator to it. The results are shown in Fig. 5. There, it can be seen that the estimations of  $\beta$  using the estimator

$\hat{\beta}_{N_n}(k)$  in the split chain are close to the true value ( $\beta = 1/2$ ) when  $k$  is chosen close  $\ln N_n$ .



**Fig. 5:** Application of the pseudo-regeneration technique to construct the split chain and estimate  $\beta$  using  $\hat{\beta}_{N_n}(k)$  in the random walk  $X_{n+1} = X_n + Z_n$  where  $Z_n$  is an i.i.d. sequence of standard normal random variables. The total number of observed points in the chain is  $10^6$  and the number of pseudo-blocks ( $N_n$ ) is 418.

When the kernel density is unknown, a procedure to approximate the split chain, based on an estimation  $\hat{\pi}_n$  of the kernel was presented as Algorithm 3 in Bertail and Cl  men  on (2006a). As indicated in Theorem 3.1 of Bertail and Cl  men  on (2006b), the accuracy of this construction depends on the rate at which  $\pi$  is estimated by  $\hat{\pi}_n$ . To our knowledge, in the null-recurrent case, the sole consistent estimator of the transition density documented in the literature is the Nadaraya-Watson estimator. The proof of the consistency can be found in section 5 of Karlsen and Tjostheim (2001). However, no results regarding its rate of convergence have been established so far. Moreover, the practical choice of the bandwidth parameter involved in this estimator is a difficult

and largely unresolved problem, as discussed on pp. 412 in [Karlsen and Tjøstheim \(2001\)](#). Hence, an ambitious line of further research consists in understanding how to implement practically the approximate regenerative block construction presented in [Bertail and Clémenton \(2006a\)](#) in order to extend the estimation methodology studied in the previous subsection to the regular pseudo-regenerative case with theoretical guarantees.

## Declarations

- Funding: This research has been conducted as part of the project Labex MME-DII (ANR11-LBX-0023-01).
- Competing Interests: The authors have no competing interests to declare that are relevant to the content of this article.
- Ethics approval: Not applicable
- Consent to participate: Not applicable
- Consent for publication: Not applicable
- Availability of data and materials: Not applicable
- Code availability: Not applicable
- Authors' contributions: Not applicable

## Appendix A Technical Proofs

This appendix contains the technical proofs of all the new results presented in the paper. Throughout this section we will use  $u_n(\delta)$  to denote  $\ln(2/\delta)/n$  for  $\delta > 0$  and  $n \in \mathbb{N}$ .

### A.1 Proof of Proposition 2.1

By the triangular inequality and equation (3), we have

$$\left| \widehat{\beta}_n(k) - \beta \right| \leq \left| \widehat{\beta}_n(k) - \beta(k) \right| + \left| \ln \left( \frac{L(e^k)}{L(e^{k+1})} \right) \right|.$$

Proposition 2.1 now follows by Lemma A.1.

**Lemma A.1.** *Let  $\delta > 0$  and  $k$  such that  $p_{k+1} \geq 16u_n(\delta)$ , then*

$$\left| \widehat{\beta}_n(k) - \beta(k) \right| \leq 6 \sqrt{\frac{u_n(\delta)}{p_{k+1}}}, \quad (\text{A1})$$

with probability larger than  $1 - 2\delta$ .

*Proof.* In order to prove this result, we need the following lemma, proved in the supplementary material of [Carpentier and Kim \(2015\)](#).

**Lemma A.2. Bernstein's inequality for Bernoulli random variables** *Let  $W_1, \dots, W_n$  be i.i.d. samples from a distribution  $F$ , and we define  $p_k = 1 - F(e^k)$ ,*

$\hat{p}_k^{(n)} = n^{-1} \sum_{i=1}^n \mathbb{I}\{W_i > e^k\}$ . Let  $\delta > 0$  and take  $n$  large enough so that  $p_k \geq 4u_n(\delta)$ . Then, with probability  $1 - \delta$ ,

$$\left| \hat{p}_k^{(n)} - p_k \right| \leq 2\sqrt{p_k u_n(\delta)}.$$

Because  $p_k \geq 16u_n(\delta)$  we can apply the previous lemma, then with probability greater than  $1 - \delta$  we have

$$\begin{aligned} -2\sqrt{p_k u_n(\delta)} &\leq \hat{p}_k^{(n)} - p_k \leq 2\sqrt{p_k u_n(\delta)} \\ p_k \left( 1 - 2\sqrt{\frac{u_n(\delta)}{p_k}} \right) &\leq \hat{p}_k^{(n)} \leq p_k \left( 1 + 2\sqrt{\frac{u_n(\delta)}{p_k}} \right), \end{aligned}$$

taking the log in the previous equation we get

$$\begin{aligned} \ln p_k + \ln \left( 1 - 2\sqrt{\frac{u_n(\delta)}{p_k}} \right) &\leq \ln \hat{p}_k^{(n)} \leq \ln p_k + \ln \left( 1 + 2\sqrt{\frac{u_n(\delta)}{p_k}} \right) \\ \ln \left( 1 - 2\sqrt{\frac{u_n(\delta)}{p_k}} \right) &\leq \ln \hat{p}_k^{(n)} - \ln p_k \leq \ln \left( 1 + 2\sqrt{\frac{u_n(\delta)}{p_k}} \right) \\ -3\sqrt{\frac{u_n(\delta)}{p_k}} &\leq \ln \hat{p}_k^{(n)} - \ln p_k \leq 2\sqrt{\frac{u_n(\delta)}{p_k}}, \end{aligned} \tag{A2}$$

where the last pair of inequalities is obtained by using  $\ln(1+x) \leq x$  and  $\ln(1-x) \geq -3x/2$  for  $x < 1/2$ . Inequality (A2) implies that

$$\left| \ln \hat{p}_k^{(n)} - \ln p_k \right| \leq 3\sqrt{\frac{u_n(\delta)}{p_k}} \tag{A3}$$

with probability bigger than  $1-\delta$ . Applying (A3) for  $k+1$  we get with probability bigger than  $1-\delta$  that

$$\left| \ln \hat{p}_{k+1}^{(n)} - \ln p_{k+1} \right| \leq 3\sqrt{\frac{u_n(\delta)}{p_{k+1}}}. \tag{A4}$$

Combining the triangular inequality with (A3) and (A4) completes the proof.  $\square$

## A.2 Proof of Theorem 2.2

The first element in the proof of Theorem 2.2 is the following simple lemma, that shows that the non-empirical version of  $\hat{\beta}_n(k)$  converges to  $\beta$ .

**Lemma A.3.**

$$\lim_{k \rightarrow +\infty} \beta(k) = \beta.$$

*Proof.* Because  $L$  is slowly varying,  $\lim_{x \rightarrow \infty} L(tx)/L(x) = 1$  (see 1.2.1 of [Bingham et al. \(1987\)](#)) for all  $t > 0$ , therefore  $L(e^k)/L(e^{k+1}) = L(e^k)/L(ee^k) \rightarrow 1$  and the result follows after taking limits in (3).  $\square$

Let  $\epsilon > 0$ . Because  $k_n \rightarrow \infty$ , Lemma A.3 implies that  $\beta(k_n) \rightarrow \beta$ , therefore we can find  $N_1 \in \mathbb{N}$  such that, for all  $n \geq N_1$

$$|\beta(k_n) - \beta| \leq \frac{\epsilon}{2}. \quad (\text{A5})$$

Take  $\delta = 2/n^2$ , then  $u_n(\delta) = 2 \ln n/n$ . Because  $L$  is slowly varying,  $L(e^{k_n+1}) \sim L(e^{k_n})$ , then  $e^{k_n \beta} \ln n/n = o(L(e^{k_n+1}))$  and we can find  $N_2 \in \mathbb{N}$  such that for all  $n \geq N_2$  we have  $p_{k_n+1} = L(e^{k_n+1})/e^{(k_n+1)\beta} \geq 32 \ln n/n = 16u_n(\delta)$ . Therefore, we can apply Lemma A.1, obtaining that, for all  $n \geq N_2$ ,

$$\mathbb{P}\left(\left|\widehat{\beta}_n(k_n) - \beta(k_n)\right| \leq 6\sqrt{\frac{2 \ln n}{np_{k_n+1}}}\right) \geq 1 - \frac{4}{n^2}. \quad (\text{A6})$$

Combining the triangular inequality with equations (A5) and (A6) we have that for all  $n \geq \max(N_1, N_2)$

$$\left|\widehat{\beta}_n(k_n) - \beta\right| \leq 6\sqrt{\frac{2 \ln n}{np_{k_n+1}}} + \frac{\epsilon}{2} \quad (\text{A7})$$

with probability bigger than  $1 - 4/n^2$ . Plugging  $p_{k_n+1} = L(e^{k_n+1})/e^{(k_n+1)\beta}$  in the first term of the right-hand side of (A7), we get

$$6\sqrt{\frac{2 \ln n}{np_{k_n+1}}} = 6\sqrt{\frac{2 \ln n}{n} \frac{e^{(k_n+1)\beta}}{L(e^{k_n+1})}} = 6\sqrt{2e^\beta} \sqrt{\frac{\ln n}{n} \frac{e^{(k_n)\beta}}{L(e^{k_n+1})}}.$$

The assumption that  $e^{k_n \beta} \ln n/n = o(L(e^{k_n+1}))$  implies that the above equality converges to 0, therefore we can find  $N_3 \in \mathbb{N}$  such that  $|6\sqrt{2 \ln n/(np_{k_n+1})}| \leq \epsilon/2$  for all  $n \geq N_3$ . Then, for all  $n \geq \max(N_1, N_2, N_3)$

$$\mathbb{P}\left(\left|\widehat{\beta}_n(k_n) - \beta(k_n)\right| \leq \epsilon\right) \geq 1 - \frac{4}{n^2},$$

which shows that  $\widehat{\beta}_n(k_n)$  converges in probability to  $\beta$ . Moreover, because the series  $\sum_n 4/n^2$  converges, Borell-Cantelli lemma implies that  $\widehat{\beta}_n(k_n) \rightarrow \beta$  almost surely.

### A.3 Proof of Corollary 2.1

If we take  $k_n = A \ln n$ , we have  $e^{\beta k_n} = n^{A\beta}$  then

$$\lim_n e^{k_n \beta} \frac{\ln n}{nL(e^{k_n})} = \lim_n \frac{\ln n}{n^{(1-A\beta)/2}} \frac{1}{n^{(1-A\beta)/2} L(n^A)} = 0.$$



For the last limit we have used that if  $L$  is slowly varying, then  $L(n^A)$  is also slowly varying and that  $\lim_n n^\gamma L(n) \rightarrow +\infty$  for  $\gamma > 0$  (Bingham et al. 1987, Proposition 1.3.6.v, pp. 16). Corollary 2.1 now follows by Theorem 2.2.

#### A.4 Proof of Theorem 2.3 and Corollary 2.2

Before starting with the proof of the asymptotic normality of the  $\widehat{\beta}_n(k_n)$  estimator, we need the following technical lemmas.

**Lemma A.4.** *Let  $W_n$  be a sequence of positive random variables and  $a_n$  and  $b_n$  two positive sequences such that  $a_n > 0$ ,  $b_n/a_n \rightarrow 0$ . If there exists a random variable  $W$  with continuous distribution function  $F$  such that  $(W_n - a_n)/b_n$  converges in distribution to  $W$ , then  $a_n(\ln W_n - \ln a_n)/b_n$  also converges in distribution to  $W$ .*

*Proof.* Let  $x \in \mathbb{R}$  be fixed. Because  $(W_n - a_n)/b_n \Rightarrow W$  in distribution, we have

$$\mathbb{P}(W_n \leq a_n + b_n x) \rightarrow F(x).$$

Using that  $a_n + b_n x = a_n(1 + a_n^{-1}b_n x)$  and taking logs we get

$$\mathbb{P}\left(\ln W_n \leq \ln a_n + \ln\left(1 + \frac{b_n}{a_n}x\right)\right) \rightarrow F(x).$$

The condition  $b_n/a_n \rightarrow 0$  implies that  $\ln(1 + a_n^{-1}b_n x) = a_n^{-1}b_n x + o(a_n^{-1}b_n)$ . Then,  $\mathbb{P}(a_n b_n^{-1}(\ln W_n - \ln a_n) \leq x + o(1))$  converges to  $F(x)$  and the Lemma follows by the continuity of  $F$ .  $\square$

**Lemma A.5.** *If  $k_n$  satisfies the hypothesis of Theorem 2.2, then,*

$$\frac{\widehat{p}_{k_n}^n}{\overline{F}(e^{k_n})} \rightarrow 1 \quad \text{almost surely.}$$

*Proof.* By Lemma A.2, for any  $\delta > 0$  such that  $p_k \geq 4u_n(\delta)$  we have that,

$$\mathbb{P}\left(\left|\frac{\widehat{p}_k^n}{p_k} - 1\right| \leq 2\sqrt{\frac{u_n(\delta)}{p_k}}\right) \geq 1 - \delta. \quad (\text{A8})$$

As in the proof of Theorem 2.2, let  $\delta = 2/n^2$ , so  $u_n(\delta) = 2 \ln n/n$ . The condition  $e^{k_n} \beta \ln n/n = o(L(e^{k_n}))$  implies that we can find  $N_1 \in \mathbb{N}$  such that  $p_{k_n} \geq 8 \ln n/n$  for all  $n \geq N_1$ , therefore, by equation (A8), we have, for all  $n \geq N_1$

$$\mathbb{P}\left(\left|\frac{\widehat{p}_{k_n}^n}{p_{k_n}} - 1\right| \leq 2\sqrt{\frac{2 \ln n}{np_{k_n}}}\right) \geq 1 - \frac{2}{n^2}.$$

Let  $\epsilon > 0$ . Notice that  $\ln n(np_{k_n})^{-1} = e^{k_n} \beta \ln n/(nL(e^{k_n}))$  and this goes to 0 as  $n$  goes to  $+\infty$ , therefore, we can find  $N_2$  such that  $2\sqrt{2 \ln n/(np_{k_n})} \leq \epsilon$  for all  $n \geq N_2$ ,

then, for all  $n \geq \max(N_1, N_2)$

$$\mathbb{P} \left( \left| \frac{\hat{p}_{k_n}^n}{p_{k_n}} - 1 \right| \leq \epsilon \right) \geq 1 - \frac{2}{n^2},$$

and the result follows by Borel-Cantelli's Lemma.  $\square$

The next result can be obtained using the same arguments of Example 11 on [Drees and Rootzén \(2010\)](#).

**Lemma A.6.** *Let  $W_n$  be a sequence of i.i.d. random variables with survival function (1),  $\phi_1$  and  $\phi_2$  bounded functions and  $u_n$  an increasing sequence of real numbers such that  $u_n \rightarrow +\infty$ . Define*

$$W_{n,i} = \frac{W_i}{u_n} \mathbb{I} \left\{ \frac{W_i}{u_n} > 1 \right\}, \quad v_n = \mathbb{P}(W_{n,i} \neq 0) \quad \text{and}$$

$$\tilde{Z}_n(\phi_k) = \frac{1}{\sqrt{nv_n}} \sum_{i=1}^n (\phi_k(W_{n,i}) - \mathbb{E}\phi_k(W_{n,i})).$$

*If there exists a sequence  $r_n$  such that*

A1  $r_n = o(n)$ .

A2  $r_n v_n \rightarrow 0$ .

A3  $nv_n \rightarrow +\infty$ .

A4  $\mathbb{E} \left[ \sum_{i=1}^{r_n} \phi_k(W_{n,i}) \right]^4 = O(r_n v_n) \quad k = 1, 2.$

A5  $\lim_n \frac{1}{r_n v_n} \sum_{i=1}^{r_n} \sum_{j=1}^{r_n} \mathbb{E}[\phi_k(W_{n,i}) \phi_l(W_{n,j})] = \sigma_{kl}.$

*Then  $(\tilde{Z}_n(\phi_k))_{1 \leq k \leq 2}$  converges weakly to a centred normal distribution with covariance matrix  $(\sigma_{kl})_{1 \leq k, l \leq 2}$ .*

Let  $k_n$  satisfy the conditions of Theorem 2.2, take  $u_n = e^{k_n}$ ,  $\phi_1(x) = \mathbb{I}\{x > 1\}$  and  $\phi_2(x) = \mathbb{I}\{x > e\}$ . With this notation:

$$\begin{aligned} \phi_1(W_{n,i}) &= \mathbb{I} \left\{ \left( \frac{W_i}{u_n} \mathbb{I} \left\{ \frac{W_i}{u_n} > 1 \right\} \right) > 1 \right\} = \mathbb{I} \left\{ \frac{W_i}{u_n} > 1 \right\}, \\ \phi_2(W_{n,i}) &= \mathbb{I} \left\{ \left( \frac{W_i}{u_n} \mathbb{I} \left\{ \frac{W_i}{u_n} > 1 \right\} \right) > e \right\} = \mathbb{I} \left\{ \frac{W_i}{u_n} > e \right\}, \\ \mathbb{E}[\phi_1(W_{n,i})] &= \mathbb{P} \left( \frac{W_i}{u_n} > 1 \right) = \bar{F}(u_n), \\ \mathbb{E}[\phi_2(W_{n,i})] &= \mathbb{P} \left( \frac{W_i}{u_n} > e \right) = \bar{F}(eu_n), \\ v_n &= \mathbb{P}(W_{n,i} \neq 0) = \mathbb{P} \left( \frac{W_i}{u_n} > 1 \right) = \bar{F}(u_n). \end{aligned}$$

Let  $w_n = \bar{F}(eu_n)$ ,  $\lambda_n = \bar{F}(u_n)/\bar{F}(eu_n) = v_n/w_n$  (notice that  $\lambda_n \rightarrow e^\beta$ ) and  $y_n = \sqrt{v_n/(nw_n^2)}$ , then,

$$\begin{aligned}\hat{\lambda}_n &= \frac{\sum_{i=1}^n \mathbb{I}\{W_i > u_n\}}{\sum_{i=1}^n \mathbb{I}\{W_i > eu_n\}} = \frac{n\mathbb{E}[\phi_1(W_{n,1})] + \sum_{i=1}^n \{\phi_1(W_{n,i}) - \mathbb{E}[\phi_1(W_{n,1})]\}}{n\mathbb{E}[\phi_2(W_{n,1})] + \sum_{i=1}^n \{\phi_2(W_{n,i}) - \mathbb{E}[\phi_2(W_{n,1})]\}} \\ &= \frac{\frac{\mathbb{E}[\phi_1(W_{n,1})]}{\mathbb{E}[\phi_2(W_{n,1})]} + \frac{\sum_{i=1}^n \{\phi_1(W_{n,i}) - \mathbb{E}[\phi_1(W_{n,1})]\}}{n\mathbb{E}[\phi_2(W_{n,1})]}}{1 + \frac{\sum_{i=1}^n \{\phi_2(W_{n,i}) - \mathbb{E}[\phi_2(W_{n,1})]\}}{n\mathbb{E}[\phi_2(W_{n,1})]}} = \frac{\frac{\mathbb{E}[\phi_1(W_{n,1})]}{\mathbb{E}[\phi_2(W_{n,1})]} + \tilde{Z}_n(\phi_1) \sqrt{\frac{v_n}{nw_n^2}}}{1 + \tilde{Z}_n(\phi_2) \sqrt{\frac{v_n}{nw_n^2}}} \\ &= \frac{\lambda_n + \tilde{Z}_n(\phi_1) y_n}{1 + \tilde{Z}_n(\phi_2) y_n}.\end{aligned}\tag{A9}$$

To apply Lemma A.6, take  $r_n = k_n$ . Condition A1 is satisfied by hypothesis. For condition A2 we have:

$$\lim_n r_n v_n = \lim_n k_n \bar{F}(e^{k_n}) = \lim_n \frac{k_n L(e^{k_n})}{e^{k_n \beta}} = \lim_n \frac{k_n}{e^{\frac{k_n \beta}{2}}} \frac{L(e^{k_n})}{e^{\frac{k_n \beta}{2}}} = 0.$$

Because  $e^{k_n \beta} \ln n / n = o(L(e^{k_n+1}))$  and  $L$  is slowly varying, we can write  $e^{k_n \beta} \ln n / n = L(e^{k_n}) \varepsilon(n)$  where  $\varepsilon(n) \rightarrow 0$  and this implies A3.

For A4, observe that  $\sum_{i=1}^{r_n} \phi_1(W_{n,i})$  follows a binomial distribution with parameters  $r_n$  and  $u_n$ , then,  $\mathbb{E}[\sum_{i=1}^{r_n} \phi_1(W_{n,i})]^4 = r_n v_n (1 - h(r_n, v_n))$ , where  $h$  is a two variables polynomial of degree 3 such that the degree of  $r_n$  on each monomial is always greater or equal than the degree of  $v_n$ . Condition A2 and  $v_n \rightarrow 0$  imply that  $\mathbb{E}[\sum_{i=1}^{r_n} \phi_1(W_{n,i})]^4 = O(r_n v_n)$ . With a similar argument but using that  $w_n/v_n \rightarrow e^{-\beta}$ , it can be shown that  $\mathbb{E}[\sum_{i=1}^{r_n} \phi_2(W_{n,i})]^4 = O(r_n v_n)$ . Hence, the condition A4 is fulfilled.

Finally, for A5, observe that

$$\begin{aligned}\mathbb{E}[\phi_1(W_{n,i}) \phi_1(W_{n,j})] &= \begin{cases} v_n, & i = j \\ v_n^2, & i \neq j \end{cases}, \\ \mathbb{E}[\phi_1(W_{n,i}) \phi_2(W_{n,j})] &= \begin{cases} w_n, & i = j \\ v_n w_n, & i \neq j \end{cases}, \\ \mathbb{E}[\phi_2(W_{n,i}) \phi_2(W_{n,j})] &= \begin{cases} w_n, & i = j \\ w_n^2, & i \neq j \end{cases},\end{aligned}$$

therefore,

$$\sigma_{11} = \lim_n \frac{1}{r_n v_n} \sum_{i=1}^{r_n} \sum_{j=1}^{r_n} \mathbb{E}[\phi_1(W_{n,i}) \phi_1(W_{n,j})] = \lim_n \frac{r_n v_n + r_n(r_n - 1)v_n^2}{r_n v_n} = 1,$$

$$\begin{aligned}\sigma_{12} &= \lim_n \frac{1}{r_n v_n} \sum_{i=1}^{r_n} \sum_{j=1}^{r_n} \mathbb{E} [\phi_1(W_{n,i}) \phi_2(W_{n,j})] = \lim_n \frac{r_n w_n + r_n (r_n - 1) v_n w_n}{r_n v_n} = e^{-\beta}, \\ \sigma_{22} &= \lim_n \frac{1}{r_n v_n} \sum_{i=1}^{r_n} \sum_{j=1}^{r_n} \mathbb{E} [\phi_2(W_{n,i}) \phi_2(W_{n,j})] = \lim_n \frac{r_n w_n + r_n (r_n - 1) w_n^2}{r_n v_n} = e^{-\beta}.\end{aligned}$$

By Lemma A.6,  $(\tilde{Z}_n(\phi_k))_{1 \leq k \leq 2}$  converges to a centred normal distribution with covariance matrix  $(\sigma_{kl})_{1 \leq k, l \leq 2}$ . Taking into account that  $y_n \sim e^\beta / \sqrt{nv_n}$ , it follows that

$$\begin{aligned}\hat{\lambda}_n &= \left( \lambda_n + \tilde{Z}_n(\phi_1) y_n \right) \left( 1 - \tilde{Z}_n(\phi_2) y_n + o_P \left( \frac{1}{\sqrt{nv_n}} \right) \right) \\ &= \lambda_n + y_n \left( \tilde{Z}_n(\phi_1) - \lambda_n \tilde{Z}_n(\phi_2) \right) + o_P \left( \frac{1}{\sqrt{nv_n}} \right).\end{aligned}\tag{A10}$$

Then,  $\sqrt{nv_n}(\hat{\lambda}_n - \lambda_n)$  converges weakly to a centred normal distribution with variance  $e^{2\beta}(\sigma_{11} + e^{2\beta}\sigma_{22} - 2e^\beta\sigma_{12}) = e^{2\beta}(e^\beta - 1)$ . This can be resumed in the following lemma.

**Lemma A.7.** *Let  $W_n$  and  $u_n$  be as in Lemma A.6, if  $k_n$  satisfies the conditions of Theorem 2.2 and  $k_n = o(n)$ , then*

$$\sqrt{n\bar{F}(e^{k_n})} \left( \frac{\sum_{i=1}^n \mathbb{I}\{W_i > e^{k_n}\}}{\sum_{i=1}^n \mathbb{I}\{W_i > e^{k_n+1}\}} - \frac{\bar{F}(e^{k_n})}{\bar{F}(e^{k_n+1})} \right)$$

converges in distribution to a centred normal distribution with variance  $e^{2\beta}(e^\beta - 1)$ .

Lemmas A.3, A.4 and A.7 combined with equation (3) imply the first part of Theorem 2.3, the second part follows from Lemma A.5 and Slutsky's Theorem. Corollary 2.2 follows immediately.

## A.5 Proof of Theorem 3.5

Let  $L_s(n)$  be such that  $L(n) = (\Gamma(1 - \beta) L_s(n))^{-1}$  with  $L$  as defined in (1). Observe that,

$$G_{\nu,D}(n) = \frac{1}{\mu(D)} \sum_{k=1}^n \mathbb{P}_\nu(X_k \in D) = \frac{1}{\mu(D)} \mathbb{E}_\nu \left[ \sum_{k=1}^n \mathbb{I}\{X_k \in D\} \right].$$

By Lemma 3.1 and Definition 3.2 of [Karlsen and Tjøstheim \(2001\)](#)  $\mathbb{E}_\nu \left[ \sum_{k=1}^n \mathbb{I}\{X_k \in D\} \right]$  is asymptotically equivalent to  $\Gamma(1 + \beta)^{-1} n^\beta \mu(D) L_s(n)$ , therefore,

$$G_{\nu,D}(n) \sim \frac{1}{\Gamma(1 + \beta)} n^\beta L_s(n).\tag{A11}$$

Let  $u(n) = n^\beta L_s(n)$ , and define the process

$$\mathbf{T}_{n,B} = \left\{ \frac{S_{\lfloor nt \rfloor}(B)}{u(n)} \right\}_{t \geq 0} = \frac{G_{\nu,D}(n) \mu(B)}{u(n)} \mathbf{S}_{n,B}.$$

By Lemma 3.6 and Theorem 3.2 of [Karlsen and Tjøstheim \(2001\)](#),  $\mathbf{T}_{n,B}$  converges weakly on the Skorokhod topology to the process  $\mu(B) M_\beta$  and by [\(A11\)](#),  $G_{\nu,D}(n)/u(n) \rightarrow 1/\Gamma(1+\beta)$  which completes the proof.

## A.6 Proof of Lemma 2.1

The representation is a direct application of Lemma [A.8](#) and the fact that  $g$  is a regularly varying function of index  $\rho$ .

**Lemma A.8.** *Assume that  $L$  satisfies SR2, has positive decrease and  $x$  is big enough such that representation the [\(8\)](#) holds, then*

$$\ln \left( \frac{L(x)}{L(\lambda x)} \right) = -c|\rho|^{-1} (g(x) - g(\lambda x)) + o(g(x)). \quad (\text{A12})$$

*Proof.* Denote  $A(x) = c\rho^{-1}g(x) + o(g(x))$ . By [\(8\)](#) we have

$$\begin{aligned} \ln \left( \frac{L(x)}{L(\lambda x)} \right) &= \ln \left( \frac{C(1+A(x))}{C(1+A(\lambda x))} \right) = \ln \left( \frac{1+A(x)}{1+A(\lambda x)} \right) \\ &= \ln(1+A(x)) - \ln(1+A(\lambda x)). \end{aligned} \quad (\text{A13})$$

Using the first order expansion for  $\ln(1+A(x))$  we have that

$$\begin{aligned} \ln(1+A(x)) &= c\rho^{-1}g(x) + o(g(x)) + \underbrace{o(c\rho^{-1}g(x) + o(g(x)))}_{o(g(x))} \\ &= c\rho^{-1}g(x) + o(g(x)). \end{aligned} \quad (\text{A14})$$

Applying [\(A14\)](#) to  $\lambda x$  we get

$$\ln(1+A(\lambda x)) = c\rho^{-1}g(\lambda x) + o(g(x)), \quad (\text{A15})$$

where we have used that if  $g$  is regularly varying then  $o(g(\lambda x)) = o(g(x))$ . The result now follows by plugging [\(A14\)](#) and [\(A15\)](#) into [\(A13\)](#).  $\square$

## A.7 Averaged Estimators

Here we collect some remarks and results related to the averaged estimator  $\widehat{\beta}_n(k, m)$ . First, we detail how to get the expression [\(6\)](#) from [\(3\)](#). Let  $k > 0$  be fixed, for each  $j$  we have:

$$\beta(k+j) = \beta + \ln \left( \frac{L(e^{k+j})}{L(e^{k+j+1})} \right),$$

then,

$$\begin{aligned}
\frac{1}{2m+1} \sum_{j=-m}^m \beta(k+j) &= \frac{1}{2m+1} \sum_{j=-m}^m \beta + \frac{1}{2m+1} \sum_{j=-m}^m \ln \left( \frac{L(e^{k+j})}{L(e^{k+j+1})} \right) \\
&= \beta + \frac{1}{2m+1} \ln \left( \prod_{j=-m}^m \frac{L(e^{k+j})}{L(e^{k+j+1})} \right) \\
&= \beta + \frac{1}{2m+1} \ln \left( \frac{L(e^{k-m})}{L(e^{k+m+1})} \right).
\end{aligned}$$

Our first result in this regard is the concentration inequality equivalent to (4) but for the averaged version of the estimator.

**Proposition A.1.** *Let  $k$  and  $m$  such that  $k > m$  and let  $\delta \in (0, 1/(2(1+2m)))$ . Then, as soon as  $p_{k+m+1} \geq 16u_n(\delta)$ , we have with probability larger than  $1 - 2\delta(1+2m)$ :*

$$\left| \widehat{\beta}_n(k, m) - \beta \right| \leq 6 \sqrt{\frac{u_n(\delta)}{p_{k+m+1}}} + \frac{1}{2m+1} \left| \ln \left( \frac{L(e^{k-m})}{L(e^{k+m+1})} \right) \right|. \quad (\text{A16})$$

The following results show that, for well-chosen  $k_n$  and  $m_n$ , the estimator  $\widehat{\beta}_n(k_n, m_n)$  is strongly consistent.

**Theorem A.2** (Strong consistency). *Let  $k_n$  and  $m_n$  such that, as  $n \rightarrow \infty$*

- i)  $k_n - m_n \rightarrow +\infty$ ,
- ii)  $\sum_{n=1}^{+\infty} \frac{4}{n^2} (1 - 2m_n)$  is convergent,
- iii)  $e^{(k_n+m_n)\beta} \frac{\ln n}{n} = o(L(e^{k_n+m_n}))$ ;

*then,  $\widehat{\beta}_n(k_n, m_n)$  converges almost surely to  $\beta$ .*

**Corollary A.1.** *Let  $A, l$  be a positive numbers such that  $l > 1$  and  $l > A\beta/(1 - A\beta)$  then*

$$\widehat{\beta}_n \left( A \ln n, \frac{A \ln n}{l} \right) \rightarrow \beta \quad \text{a.s.}$$

### A.7.1 Proof of Proposition A.1

Similarly to the proof of Theorem 2.1, Theorem A.1 follows by triangular inequality, the definition of  $\beta(k, m)$  and the following Lemma A.9, which provides us a bound for the difference between  $\widehat{\beta}(k, m) - \beta(k, m)$ .

**Lemma A.9.** *Let  $\delta > 0$  and  $k$  and  $m$  such that  $p_{k+m+1} \geq 16u_n(\delta)$ , then*

$$\left| \widehat{\beta}_n(k, m) - \beta(k, m) \right| \leq 6 \sqrt{\frac{u_n(\delta)}{p_{k+m+1}}}, \quad (\text{A17})$$

*with probability larger than  $1 - 2\delta(1 - 2m)$ .*

*Proof.* For the left-hand side of equation (A17) we have

$$\begin{aligned} \left| \widehat{\beta}_n(k, m) - \beta(k, m) \right| &= \frac{1}{2m+1} \left| \sum_{j=-m}^m \left( \widehat{\beta}_n(k+j) - \beta(k+j) \right) \right| \\ &= \frac{1}{2m+1} \left| \sum_{j=0}^{2m} \left( \widehat{\beta}_n(k-m+j) - \beta(k-m+j) \right) \right|, \end{aligned}$$

then,

$$\left| \widehat{\beta}_n(k, m) - \beta(k, m) \right| \leq \frac{1}{2m+1} \sum_{j=0}^{2m} \left| \widehat{\beta}_n(k-m+j) - \beta(k-m+j) \right|. \quad (\text{A18})$$

Given  $p_{k+m+1} \geq 16u_n(\delta)$ , it follows that  $p_{k-m+j+1} \geq 16u_n(\delta)$  for all  $j$  between 0 and  $2m$ , allowing the application of Lemma A.1. This yields, for each  $j$ ,

$$\left| \widehat{\beta}_n(k-m+j) - \beta(k-m+j) \right| \leq 6 \sqrt{\frac{u_n(\delta)}{p_{k-m+j+1}}}, \quad (\text{A19})$$

with probability bigger than  $1 - 2\delta$ .

The joint probability of inequality (A19) holding for all  $j$  between 0 and  $2m$  is greater than  $(2m+1)(1-2\delta) - 2m = 1 - 2\delta(1-2m)$ . Hence, with at least this probability, equation (A18) simplifies to,

$$\left| \widehat{\beta}_n(k, m) - \beta(k, m) \right| \leq \frac{1}{2m+1} \sum_{j=0}^{2m} 6 \sqrt{\frac{u_n(\delta)}{p_{k-m+j+1}}} \leq 6 \sqrt{\frac{u_n(\delta)}{p_{k+m+1}}}.$$

□

### A.7.2 Proof of Theorem A.2

The following Lemma A.10 shows that if  $k_n - m_n \rightarrow +\infty$ , then  $\beta(k_n, m_n) \rightarrow \beta$ . Theorem A.2 now follows by the same argument used to prove Theorem 2.2, using the convergence of  $\beta(k_n, m_n)$  instead of Lemma A.3 and Lemma A.9 instead of Lemma A.1.

**Lemma A.10.** *Let  $\alpha_n$ ,  $k_n$  and  $b_n$  be sequences such that,  $\alpha_n \rightarrow \alpha$ ,  $k_n \rightarrow +\infty$  and  $k_n - n_n \rightarrow +\infty$ . Then,*

$$\frac{1}{2b_n+1} \sum_{j=-b_n}^{b_n} \alpha_{k_n+j} \rightarrow \alpha.$$

*Proof.* Let  $\rho_n = \frac{1}{2b_n+1} \sum_{j=-b_n}^{b_n} \alpha_{k_n+j}$ , then

$$\rho_n - \alpha = \frac{1}{2b_n+1} \sum_{j=-b_n}^{b_n} (\alpha_{k_n+j} - \alpha) = \frac{1}{2b_n+1} \sum_{j=0}^{2b_n} (\alpha_{k_n-b_n+j} - \alpha).$$

Fix  $\epsilon > 0$ . The convergence of  $\alpha_n$  ensures the existence of  $N_1$  such that  $|\alpha_n - \alpha| < \epsilon$  for all  $n \geq N_1$ . Given that  $k_n - b_n \rightarrow +\infty$ , there exists  $N_2$  satisfying  $k_n - b_n \geq N_1$ , for all  $n \geq N_2$ , which yields  $|\alpha_{k_n-b_n+j} - \alpha| \leq \epsilon$  for all  $n \geq N_2$  and  $j \in \mathbb{N}$ . Consequently, for  $n \geq N_2$ ,

$$|\rho_n - \alpha| \leq \frac{1}{2b_n+1} \sum_{j=0}^{2b_n} |\alpha_{k_n-b_n+j} - \alpha| \leq \frac{1}{2b_n+1} \sum_{j=0}^{2b_n} \epsilon = \epsilon.$$

□

### A.7.3 Proof of Corollary A.1

We just need to show that sequences  $k_n = A \ln n$  and  $m_n = A \ln n/l$  satisfy conditions (i), (ii) and (iii) of Theorem A.2. The first two are trivially satisfied, for the third one, notice that

$$\lim_n \frac{e^{(A \ln n + \frac{A \ln n}{l})^\beta}}{L\left(e^{A \ln n + \frac{A \ln n}{l}}\right)} \frac{\ln n}{n} = \lim_n \frac{1}{n^{\frac{1-(1+\frac{1}{l})A\beta}{2}} L\left(n^{(1+\frac{1}{l})A}\right)} \frac{\ln n}{n^{\frac{1-(1+\frac{1}{l})A\beta}{2}}}.$$

The condition  $l > A\beta/(1 - A\beta)$  implies that  $1 - (1 + 1/l)A\beta > 0$ , therefore,

$$\lim_n \frac{1}{n^{\frac{1-(1+\frac{1}{l})A\beta}{2}} L\left(n^{(1+\frac{1}{l})A}\right)} = 0 \quad \text{and} \quad \lim_n \frac{\ln n}{n^{\frac{1-(1+\frac{1}{l})A\beta}{2}}} = 0,$$

which shows that  $k_n$  and  $m_n$  satisfy condition (iii) in Theorem A.2.

## A.8 Tail index estimation in the positive recurrent case

In this section, we prove the asymptotic normality of our estimator in the positive recurrent case. The main result is the following

**Theorem A.3** (Asymptotic normality in the positive recurrent case). *Suppose  $X$  is a regenerative positive recurrent Markov chain such that the distribution of its regeneration time satisfies equation (1). Assume that  $k_n$  satisfies the hypothesis of Theorem 2.3 as well as the following extra assumption:*



A6 There exists  $D' > 0$  such that for any  $0 < D < D'$

$$\sum_{j=n+1}^{n(1+D)} \bar{F}(e^{k_j}) = nD\bar{F}(e^{k_n}) + o(\bar{F}(e^{k_n})). \quad (\text{A20})$$

Then,

(i) as  $n \rightarrow +\infty$ , we have the convergence in distribution:

$$\sqrt{N_n p_{k_{N_n}}} \left( \hat{\beta}_{N_n}(k_{N_n}) - \beta(k_{N_n}) \right) \Rightarrow \mathcal{N}(0, e^\beta - 1).$$

(ii) In addition, asymptotic normality holds true for the 'standardized' deviation:

$$\frac{\sqrt{N_n \hat{p}_{k_{N_n}}^{(N_n)}} \left( \hat{\beta}_{N_n}(k_{N_n}) - \beta(k_{N_n}) \right)}{\sqrt{e^{\hat{\beta}_{N_n}(k_{N_n})} - 1}} \Rightarrow \mathcal{N}(0, 1), \text{ as } n \rightarrow +\infty.$$

**Remark 6.** Assumption A6 can be replaced by the following slightly more restrictive, but easier to verify, assumption:

A6' There exists  $D' > 0$  such that for any  $0 < D < D'$

$$\bar{F}(e^{k_{n(1+D)}}) = \bar{F}(e^{k_n}) + o\left(\frac{\bar{F}(e^{k_n})}{n}\right).$$

As in the case of Theorem 2.3, the main result follows directly from the following lemma, which is an extension of Lemma A.7 for the case where the number of i.i.d samples we have is random.

**Lemma A.11.** Let  $W_n$  be a sequence of i.i.d. random variables with survival function (1), and  $k_n$  be a sequence such that satisfies the hypothesis of Theorem A.3. Suppose that  $T_n$  is a sequence of positive, integer-valued random variables such that  $\frac{T_n}{n}$  converges in probability to some positive number  $\theta$ . Then

$$\sqrt{T_n \bar{F}(e^{k_{T_n}})} \left( \frac{\sum_{i=1}^{T_n} \mathbb{I}\{W_i > e^{k_{T_n}}\}}{\sum_{i=1}^{T_n} \mathbb{I}\{W_i > e^{k_{T_n}+1}\}} - \frac{\bar{F}(e^{k_{T_n}})}{\bar{F}(e^{k_{T_n}+1})} \right).$$

converges weakly to a centred normal distribution with variance  $e^{2\beta}(e^\beta - 1)$ .

### A.8.1 Proof of Lemma A.11

For this proof, we will reuse the notation we utilized in the proof of Lemma A.7 and will add the following definitions:  $q_n = nv_n$ ,  $y_n = \sqrt{v_n/(nw_n^2)}$  and

$$U_{n,k} = \sum_{i=1}^n (\phi_k(W_{n,i}) - \mathbb{E}[\phi_k(W_{n,i})]).$$

By equation (A10) and Lemma A.6

$$\sqrt{q_n} (\hat{\lambda}_n - \lambda_n) = y_n \sqrt{q_n} (\tilde{Z}_n(\phi_1) - \lambda_n \tilde{Z}_n(\phi_2)) + o_P(1),$$

and  $(\tilde{Z}_n(\phi_k))_{1 \leq k \leq 2}$  converges weakly to a centred normal distribution. Using that  $\lambda_n \rightarrow e^\beta$  and  $y_n \sim e^\beta / \sqrt{q_n}$ , this implies that

$$\frac{\sqrt{q_n} (\hat{\lambda}_n - \lambda_n)}{e^\beta} = \tilde{Z}_n(\phi_1) - e^\beta \tilde{Z}_n(\phi_2) + o_P(1). \quad (\text{A21})$$

Take  $n_0 = \lfloor \theta n \rfloor$ , and  $V_n = \sqrt{q_n} (\tilde{Z}_n(\phi_1) - e^\beta \tilde{Z}_n(\phi_2)) = U_{n,1} - e^\beta U_{n,2}$ , then

$$\frac{\sqrt{q_{T_n}}}{e^\beta} (\hat{\lambda}_{T_n} - \lambda_{T_n}) = \left( \frac{V_{n_0}}{\sqrt{q_{n_0}}} + \frac{V_{T_n} - V_{n_0}}{\sqrt{q_{n_0}}} \right) \sqrt{\frac{q_{n_0}}{q_{T_n}}} + o_P(1).$$

By Lemma A.7 and our assumption about the convergence in probability of  $T_n/n$ , we have that  $V_{n_0}/\sqrt{q_{n_0}}$  converges in distribution to a centred Normal random variable with variance  $e^{2\beta}(e^\beta - 1)$  and  $q_{n_0}/q_{T_n}$  converges in probability to 1. Therefore, if we show that

$$\frac{V_{T_n} - V_{n_0}}{\sqrt{q_{n_0}}} \rightarrow 0 \quad (\text{A22})$$

in probability then our lemma will be proved by two successive applications of Slutsky's theorem.

Notice that  $V_{T_n} - V_{n_0} = U_{T_n,1} - U_{n_0,1} - e^\beta (U_{T_n,2} - U_{n_0,2})$ , hence, if we show that  $(U_{T_n,1} - U_{n_0,1})/\sqrt{q_{n_0}}$  and  $(U_{T_n,2} - U_{n_0,2})/\sqrt{q_{n_0}}$  converge to 0 in probability, then (A22) will be proved. Given that the proofs of both convergences are analogous, we will only demonstrate the first one.

Let  $\epsilon > 0$  be fixed, and set  $n_1 = \lfloor n_0 (1 - \epsilon^3/32) \rfloor + 1$ ,  $n_2 = \lfloor n_0 (1 + \epsilon^3/32) \rfloor$ , then

$$\mathbb{P}(|U_{T_n,1} - U_{n_0,1}| > \epsilon \sqrt{q_{n_0}}) \leq I_{n,1} + I_{n,2}, \quad (\text{A23})$$

where,

$$\begin{aligned} I_{n,1} &= \mathbb{P}(T_n \notin [n_1, n_2]), \\ I_{n,2} &= \mathbb{P}(\{|U_{T_n,1} - U_{n_0,1}| > \epsilon \sqrt{q_{n_0}}\} \cap T_n \in [n_1, n_2]). \end{aligned}$$

The convergence in probability of  $T_n/n$  to  $\theta$  implies that there exists  $N_1$  such that  $I_{n,1} < \epsilon/2$  for all  $n \geq N_1$ , hence,

$$\forall n \geq N_1 \quad \mathbb{P}(|U_{T_n,1} - U_{n_0,1}| > \epsilon\sqrt{q_{n_0}}) \leq \frac{\epsilon}{2} + I_{n,2}. \quad (\text{A24})$$

To bound the second term on the right-hand side of the previous display, observe that  $I_{n,2}$  is smaller than

$$\mathbb{P}\left(\max_{n_1 \leq j \leq n_0} |U_{j,1} - U_{n_0,1}| > \epsilon\sqrt{q_{n_0}}\right) + \mathbb{P}\left(\max_{n_0 < j \leq n_2} |U_{j,1} - U_{n_0,1}| > \epsilon\sqrt{q_{n_0}}\right).$$

We just need to focus on the case  $n_0 < j \leq n_2$  because the other will be analogous. To ease the notation, for any  $a < b$ , we will write  $\bar{F}_a^b$  instead of  $\bar{F}(u_a) - \bar{F}(u_b)$  and we will use  $\bar{F}_a$  to denote  $\bar{F}(u_a)$ . Let  $\mathcal{C}_n$  be the set  $\{\max_{n_0 < j \leq n_2} |U_{j,1} - U_{n_0,1}| > \epsilon\sqrt{q_{n_0}}\}$ . We can write the difference  $U_{j,1} - U_{n_0,1}$  as

$$\begin{aligned} U_{j,1} - U_{n_0,1} &= \sum_{i=1}^j (\mathbb{I}\{W_i > u_j\} - \bar{F}_j) - \sum_{i=1}^{n_0} (\mathbb{I}\{W_i > u_{n_0}\} - \bar{F}_{n_0}) \\ &= \sum_{i=n_0+1}^j (\mathbb{I}\{W_i > u_{n_0}\} - \bar{F}_{n_0}) - \sum_{i=1}^j (\mathbb{I}\{W_i \in (u_{n_0}, u_j]\} - \bar{F}_{n_0}^j). \end{aligned} \quad (\text{A25})$$

Suppose for the moment that

$$\frac{\max_{n_0 < j \leq n_2} \left| \sum_{i=1}^j (\mathbb{I}\{W_i \in (u_{n_0}, u_j]\} - \bar{F}_{n_0}^j) \right|}{\sqrt{q_{n_0}}} \rightarrow 0 \quad (\text{A26})$$

in probability.

Let  $\mathcal{G}_n$  be the event  $\{\max_{n_0 < j \leq n_2} |\sum_{i=1}^j (\mathbb{I}\{W_i \in (u_{n_0}, u_j]\} - \bar{F}_{n_0}^j)| \leq \epsilon\sqrt{q_{n_0}}/2\}$ . By (A26), we can find  $N_2$  such that  $\mathbb{P}(\mathcal{G}_n) \geq 1 - \epsilon/8$  for all  $n \geq N_2$ . Therefore, for all  $n \geq N_2$  we have

$$\mathbb{P}(\mathcal{C}_n) \leq \mathbb{P}\left(\left\{\max_{n_0 < j \leq n_2} |U_{j,1} - U_{n_0,1}| > \epsilon\sqrt{q_{n_0}}\right\} \cap \mathcal{G}_n\right) + \frac{\epsilon}{8}.$$

Using (A25), we obtain that  $\{\max_{n_0 < j \leq n_2} |U_{j,1} - U_{n_0,1}| > \epsilon\sqrt{q_{n_0}}\} \cap \mathcal{G}_n$  is contained in the event

$$\left\{\max_{n_0 < j \leq n_2} \left| \sum_{i=n_0+1}^j (\mathbb{I}\{W_i > u_{n_0}\} - \bar{F}_{n_0}) \right| > \frac{\epsilon\sqrt{q_{n_0}}}{2}\right\}.$$

By Kolmogorov inequality,

$$\mathbb{P} \left( \max_{n_0 < j \leq n_2} \left| \sum_{i=n_0+1}^j (\mathbb{I}\{W_i > u_{n_0}\} - \bar{F}_{n_0}) \right| > \frac{\epsilon \sqrt{q_{n_0}}}{2} \right) \leq \frac{4(n_2 - n_0) \bar{F}_{n_0}}{\epsilon^2 n_0 \bar{F}_{n_0}}.$$

The right-hand side of the previous equation equals  $4(|n_0(1 + \epsilon^3/32)| - n_0)/(\epsilon^2 n_0)$ , and that is smaller than  $\epsilon/8$ . Hence,  $\mathbb{P}(\mathcal{C}_n) \leq \frac{\epsilon}{4}$  for all  $n \geq N_2$ . In a similar fashion, we can find  $N_3$  such that  $\mathbb{P}(\max_{n_1 \leq j \leq n_0} |U_{j,1} - U_{n_0,1}| > \epsilon \sqrt{q_{n_0}}) \leq \epsilon/4$  for  $n \geq N_3$ . This shows that  $I_{n,2} \leq \epsilon/2$  for  $n \geq \max(N_2, N_3)$ . Combining this with equation (A24), proofs (A22).

To finish, we proceed with the proof of (A26). Let  $\delta > 0$  be fixed. Without loss of generality, assume that  $1 + \epsilon^3/8 < 2$  and  $\epsilon^3/8 < D'$ . Denote by  $\mathcal{H}_{n,\delta}$  the event  $\{\max_{n_0 < j \leq n_2} |\sum_{i=1}^j (\mathbb{I}\{W_i \in (u_{n_0}, u_j]\} - \bar{F}_{n_0}^j)| > \delta \sqrt{q_{n_0}}\}$ , then,

$$\mathcal{H}_{n,\delta} \subseteq \bigcup_{j=n_0+1}^{n_2} \left\{ \left| \sum_{i=1}^j (\mathbb{I}\{W_i \in (u_{n_0}, u_j]\} - \bar{F}_{n_0}^j) \right| > \delta \sqrt{q_{n_0}} \right\}. \quad (\text{A27})$$

By Chebyshev's inequality,

$$\mathbb{P} \left( \left| \sum_{i=1}^j (\mathbb{I}\{W_i \in (u_{n_0}, u_j]\} - \bar{F}_{n_0}^j) \right| > \delta \sqrt{q_{n_0}} \right) \leq \frac{2\bar{F}_{n_0}^j}{\delta^2 \bar{F}_{n_0}} = \frac{2}{\delta^2} \left( 1 - \frac{\bar{F}_j}{\bar{F}_{n_0}} \right).$$

Combining this with (A27) and A6, we obtain

$$\begin{aligned} \mathbb{P}(\mathcal{H}_{n,\delta}) &\leq \frac{2}{\delta^2} \sum_{j=n_0+1}^{n_2} \left( 1 - \frac{\bar{F}_j}{\bar{F}_{n_0}} \right) \leq \frac{2}{\delta^2} \left( n_2 - n_0 - \frac{\sum_{j=n_0+1}^{n_2} \bar{F}_j}{\bar{F}_{n_0}} \right) \\ &\leq \frac{2}{\delta^2} \left( n_2 - n_0 - \frac{(n_2 - n_0) \bar{F}_{n_0} + o(\bar{F}_{n_0})}{\bar{F}_{n_0}} \right) = o(1), \end{aligned}$$

which completes the proof of (A26).

## References

- Alexander, K. (2011). Excursions and local limit theorems for bessel-like random walks. *Electronic Journal of Probability*, 16, , <https://doi.org/10.1214/ejp.v16-848>
- Asmussen, S. (2010). *Applied probability and queues* (2nd ed.). New York: Springer New York.
- Bertail, P., & Cl  men  on, S. (2004). Edgeworth expansions of suitably normalized sample mean statistics for atomic markov chains. *Probability theory and related fields*, 130(3), 388–414,

- Bertail, P., & Cl  men  on, S. (2006a). Dependence in probability and statistics. In P. Bertail, P. Soulier, & P. Doukhan (Eds.), *Dependence in probability and statistics* (pp. 3–54). New York: Springer New York.
- Bertail, P., & Cl  men  on, S. (2006b). Regenerative block bootstrap for markov chains. *Bernoulli*, 12(4), 689–712,
- Bingham, N.H., Goldie, C.M., Teugels, J.L. (1987). *Regular variation*. Cambridge: Cambridge University Press.
- Carpentier, A., & Kim, A.K.H. (2015). Adaptive and minimax optimal estimation of the tail coefficient. *Statistica Sinica*, 25(3), 1133–1144,
- Chen, X. (1999, 07). How often does a harris recurrent markov chain recur? *The Annals of Probability*, 27, , <https://doi.org/10.1214/aop/1022677449>
- Chen, X. (2000, 01). On the limit laws of the second order for additive functionals of harris recurrent markov chains. *Probability Theory and Related Fields*, 116, , <https://doi.org/10.1007/pl00008724>
- Clauset, A., Shalizi, C.R., Newman, M.E.J. (2009, 11 04). Power-law distributions in empirical data. *SIAM Review*, 51, , <https://doi.org/10.1137/070710111>
- Drees, H., & Rootz  n, H. (2010). Limit theorems for empirical processes of cluster functionals. *The Annals of Statistics*, 38(4), 2145 – 2186, <https://doi.org/10.1214/09-AOS788>
- Feller, W. (1971). *An introduction to probability theory and its applications* (2nd ed., Vol. 2). New York: Wiley.
- Gao, J., Tj  stheim, D., Yin, J. (2013). Estimation in threshold autoregressive models with a stationary and a unit root regime. *Journal of Econometrics*, 172(1), 1–13,
- Goldie, C.M., & Smith, R.L. (1987). Slow variation with remainder: theory and applications. *Quart. J. Math. Oxford*, 38(1), 45–71,

- Goldstein, M.L., Morris, S.A., Yen, G.G. (2004, 09). Problems with fitting to the power-law distribution. *The European Physical Journal B / Condensed Matter and Complex Systems*, 41, , <https://doi.org/10.1140/epjb/e2004-00316-5>
- Gut, A. (2013). *Probability: A graduate course* (2nd ed.). New York: Springer.
- Hill, B.M. (1975, 09). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3, , <https://doi.org/10.2307/2958370>
- Jain, J., & Jamison, B. (1967). Contributions to Doeblin's theory of Markov processes. *Z. Wahrsch. Verw. Geb.*, 8, 19–40,
- Karlsen, H., Myklebust, T., Tjøstheim, D. (2010). Nonparametric regression estimation in a null recurrent time series. *Journal of Statistical Planning and Inference*, 140, , <https://doi.org/10.1016/j.jspi.2010.04.029>
- Karlsen, H., & Tjøstheim, D. (2001, 04). Nonparametric estimation in null recurrent time series. *The Annals of Statistics*, 29, , <https://doi.org/10.1214/aos/1009210546>
- Lazzardi, S., Valle, F., Mazzolini, A., Scialdone, A., Caselle, M., Osella, M. (2021). Emergent statistical laws in single-cell transcriptomic data. *bioRxiv*, , <https://doi.org/10.1101/2021.06.16.448706>
- Manning, C., & Scütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, Massachusetts: The MIT Press.
- Matsui, M., Mikosch, T., Tafakori, L. (2013, feb 03). Estimation of the tail index for lattice-valued sequences. *Extremes*, 16, , <https://doi.org/10.1007/s10687-012-0167-9>
- Meyn, S., Tweedie, R., Glynn, P. (2009). *Markov chains and stochastic stability* (2nd ed.). Cambridge: Cambridge University Press.
- Myklebust, T., Karlsen, H., Tjøstheim, D. (2012, 2). Null recurrent unit root process. *Econometric Theory*, 28, , <https://doi.org/10.1017/S0266466611000119>
- Nummelin, E. (1984). *General irreducible markov chains and non-negative operators*. Cambridge: Cambridge University Press.

- Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics*, 3(1), 119 – 131, <https://doi.org/10.1214/aos/1176343003>
- Resnick, S. (2007). *Heavy-tail phenomena: probabilistic and statistical modeling*. New York: Springer Science & Business Media.
- Sidra, A., Shougeng, H., Nadeem, A.B. (2018). Zipf’s law and city size distribution: A survey of the literature and future research agenda. *Physica A: Statistical Mechanics and Its Applications*(492), 75–92,
- Voitalov, I., van der Hoorn, P., van der Hofstad, R., Krioukov, D. (2019). Scale-free networks well done. *Physical Review Research*, 1(3), 033034,
- Zanette, D. (2006). Zipf’s law and the creation of musical context. *Musicae Scientiae*, 10(1), 3-18, <https://doi.org/10.1177/102986490601000101>