Extreme values statistics for Markov chains via the (pseudo-) regenerative method

Patrice Bertail, Stéphan Clémençon, Jessica Tressou

Extreme values statistics for Markov chains via the (pseudo-) regenerative method

Patrice Bertail
CREST (INSEE) - Laboratoire de Statistiques & MODALX - Université Paris X

Stéphan Clémençon*
LTCI - UMR Institut Telecom/CNRS No. 5141

Jessica Tressou
Unité Mét@risk - INRA & Hong Kong University of Science and Technology

May 9, 2008

Abstract

This paper is devoted to the study of specific statistical methods for extremal events in the markovian setup, based on the regenerative method and the Nummelin technique. Exploiting ideas developed in Rootzén (1988), the principle underlying our methodology consists of first generating a random number l of approximate pseudo-renewal times $\tau_1, \tau_2, \ldots, \tau_l$ for a sample path $X_1, \ldots, X_n$ drawn from a Harris chain X with state space E, from the parameters of a *minorization condition* fulfilled by its transition kernel, and then computing submaxima over the *approximate cycles* thus obtained: $\max_{1+\tau_1 \leq i \leq \tau_2} f(X_i), \ldots, \max_{1+\tau_{l-1} \leq i \leq \tau_l} f(X_i)$ for any measurable function $f : E \to \mathbb{R}$. Estimators of tail features of the sample maximum $\max_{1 \leq i \leq n} f(X_i)$ are then constructed by applying standard statistical methods, tailored for the i.i.d. setting, to the submaxima as if they were independent and identically distributed. In particular, the asymptotic properties of extensions of popular inference procedures based on the conditional maximum likelihood theory, such as Hill's method for the index of regular variation, are thoroughly investigated. Using the same approach, we also consider the problem of estimating the extremal index of the sequence $\{f(X_n)\}_{n \in \mathbb{N}}$ under suitable assumptions. Eventually, practical issues related to the application of the methodology we propose are discussed and preliminary simulation results are displayed.

Keywords and phrases: regenerative Markov chain, Nummelin splitting technique, extreme value statistics, cycle submaximum, Hill estimator, extremal index.
AMS 2000 Mathematics Subject Classification: 60G70, 60J10, 60K20.

*Address of corresponding author: Stéphan Clémençon - Telecom Paristech - 46, rue Barrault - 75634 Paris Cedex 13 - Email: Stephan.Clemencon@telecom-paristech.fr - Tel: +33 1 45 81 78 07 - Fax: +33 1 45 81 71 58

1 Introduction

In [10, 11], a statistical methodology based on approximating the pseudo-regeneration properties of general Harris Markov chains has been introduced for tackling various estimation problems in the markovian setup: mean and variance estimation, confidence intervals, $\mathcal{U}$-statistics, bootstrap and robust functional estimation. It has been proved to lead to asymptotically valid inference procedures with both theoretical and practical advantages over so called *blocking-techniques* for data exhibiting this specific pattern of dependence. The purpose of this paper is to further develop this approach, in order to propose novel inference methods for estimating some specific features related to the extremal behavior of certain functionals of Markov chains.

Motivated by various applications including statistical analysis of financial or insurance data, and queuing or inventory models in operations research, the challenging problem of extending estimation methods for extremal events elaborated for the i.i.d. setup to weakly dependent data has indeed received increasing attention in the statistical literature over the past several years, see [35, 36, 37, 52, 57] for instance. As shown by numerous former studies, generalization is not straightforward in many cases and dependency cannot be ignored. Most methods for statistical analysis of extremal events in such a dependent setting rely on *blocking-techniques*, which consist roughly of dividing an observed data series into (non overlapping) blocks of fixed length and then examining how extreme values occur over these data segments, in order to capture the dependency structure and determine its role in the extremal behavior. Indeed, whereas extreme values naturally occur in an isolated fashion in the i.i.d. setup, they generally tend to come in small clusters for weakly dependent sequences. The notion of *extremal index* accounts for this phenomenon, and the reader can refer to [42] for an account of extreme value theory for stochastic processes.

Given the ubiquity of the Markov assumption in time-series modeling and applied probability models, here we propose an alternative to those statistical methods, specifically tailored for the markovian framework. To be precise, this paper looks at statistical inference for extremal events from the renewal theory angle. As first observed in [56], see also [2, 3, 31, 32], certain extremal behavior features of Harris Markov chains may be also expressed in terms of *regeneration cycles*, namely data segments between consecutive regeneration times τ_1, τ_2, ..., *i.e.* random times at which the chain forgets its past. Working on this approach, the methodology proposed in this paper consists of splitting up the observed sample path into regeneration data blocks, or into data blocks drawn from a distribution approximating the regeneration cycle's distribution in the general case when regeneration times cannot be observed. It then analyzes the sequence of maxima over the resulting data segments, as if they were i.i.d., via standard statistical methods. In order to illustrate the advantages of this technique, we concentrate on several important inference problems. We focus on estimating the sample maximum's tail directly, the *extremal index* and the *regular variation index*, by means of the (pseudo-) regenerative method and the rigorous formulation of asymptotic results for these problems. Owing to space limitations, other possible estimation problems are not discussed here. Some simulation studies are

furthermore presented, with the aim of investigating the performance of the techniques introduced in the present paper from an empirical viewpoint and comparing them in this respect to other methods, standing as natural candidates in the markovian context.

The rest of the paper is structured as follows. In section 2, notations are first set out and basics about the regenerative properties of Markov chains or of suitable theoretical extensions of the latter are briefly recalled, together with the *plug-in* technique, first introduced in [10], aimed at generating approximate regeneration times for general Harris chains $\hat{\tau}_1$, $\hat{\tau}_2$, ... As a preamble, section 3 highlights the connection between the (pseudo-) regenerative properties of a Harris chain $X = (X_n)_{n\in\mathbb{N}}$ and the extremal behavior of sequences of type $f(X) = \{f(X_n)\}_{n\in\mathbb{N}}$, which constitute the main principle underlying the estimation methods considered in this paper. Preliminary statistical results are also stated. In section 4, as a first attempt, the regeneration-based approach is applied to the problem of directly estimating the tail of $\max_{1\leq i\leq n} f(X_i)$. Next, an estimation of the extremal index of the sequence $f(X)$ is tackled. It also shows how, by simply considering the submaxima $\max_{1+\tau_1\leq i\leq\tau_2} f(X_i)$, ..., or approximations of the latter $\max_{1+\hat{\tau}_1\leq i\leq\hat{\tau}_2} f(X_i)$, ..., one may extend straightforwardly statistical methods for extremal events proved consistent in the i.i.d. setup (under "maximum domain of attraction" assumptions) to markovian data, taking the popular Hill's procedure as an illustrative example in the Fréchet case. Consequent empirical results are displayed, together with a short discussion of practical issues related to the implementation of the pseudo-regeneration based approach. Section 6 yields the technical proofs.

2 On the (pseudo-) regenerative approach for markovian data

Here and throughout $X = (X_n)_{n\in\mathbb{N}}$ denotes a ψ-irreducible time-homogeneous Markov chain, valued in a countably generated measurable space $(E, \mathcal{E})$ with transition probability $\Pi(x, dy)$ and initial distribution ν (refer to [54] for basic concepts of the Markov chain theory). In what follows, $\mathbb{P}_\nu$ (respectively, $\mathbb{P}_x$ for x in E) denotes the probability measure on the underlying space such that $X_0 \sim \nu$ (resp., conditioned upon $X_0 = x$), $\mathbb{E}_\nu[.]$ the $\mathbb{P}_\nu$-expectation (resp. $\mathbb{E}_x[.]$ the $\mathbb{P}_x(.)$-expectation) and $\mathbb{I}\{\mathcal{A}\}$ the indicator function of any event $\mathcal{A}$. We also use the notations: $\forall (a, b)^2$, $a \vee b = \min(a, b)$ and $a \wedge b = \max(a, b)$. We assume furthermore that X is positive recurrent and denote by μ its unique invariant probability distribution.

2.1 Markov chains with regeneration times

A Markov chain X is said to be*regenerative* when it possesses an accessible atom, *i.e.* a measurable set A such that $\psi(A) > 0$ and $\Pi(x, .) = \Pi(y, .)$ for all x, y in A. Denote then by $\tau_A = \tau_A(1) = \inf\{n \geq 1,\ X_n \in A\}$ the hitting time on A, by $\tau_A(j) = \inf\{n > \tau_A(j-1),\ X_n \in A\}$ for $j \geq 2$ the successive return times to A and by $\mathbb{E}_A[.]$ the expectation conditioned on $X_0 \in A$. When the chain is Harris recurrent, the probability of returning infinitely often to atom A is equal to one, whatever the starting state. It follows

from the *strong Markov property* that, for any initial distribution ν, the sample paths of the chain may be divided into i.i.d. blocks of random length corresponding to consecutive visits to A, generally termed *regeneration cycles*:

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

taking their values in the torus $\mathbb{T} = \cup_{n=1}^{\infty} E^n$. The renewal sequence $\{\tau_A(j)\}_{j\geq 1}$ defines successive times at which the chain forgets its past, called *regeneration times*. We point out that the class of atomic Markov chains contains not only chains with a countable state space (for the latter, any recurrent state is an accessible atom), but also many specific Markov models arising from the field of operational research. Refer to [4] for regenerative models involved in queuing theory, see also §5.1 below. When an accessible atom exists, the *stochastic stability* properties of the chain reduce to properties concerning the speed of return time to the atom only. In this framework, one may show for instance that the chain X is positive recurrent if and only if $\mathbb{E}_A[\tau_A] < \infty$, see Theorem 10.2.2 in [47]. The unique invariant probability distribution μ is then the Pitman's occupation measure given by

$$\mu(B) = \frac{1}{\mathbb{E}_A[\tau_A]} \mathbb{E}_A[\sum_{i=1}^{\tau_A} \mathbb{I}\{X_i \in B\}], \text{ for all } B \in \mathcal{E}. \tag{1}$$

For atomic chains, limit theorems can be derived from the application of the corresponding results to the i.i.d. blocks $(\mathcal{B}_n)_{n\geq 1}$, see [59] and the references therein. One may refer for example to [47] for the LLN, CLT, LIL, [15] for the Berry-Esseen theorem, [44], and [45, 46, 10] for other refinements of the CLT. The same technique can also be applied to establish moment and probability inequalities, which are not asymptotic results, see [19, 12]. As mentioned above, these results are established from moment assumptions related to the distribution of the $\mathcal{B}_n$'s such as the ones stated below.

Moment assumptions. Let A be an accessible atom and $\kappa \geq 1$. Consider the following assumptions. Notice that they are independent from the atom A chosen, see §1.1 in Chapter 14 of [47] for instance.

$$\begin{array}{lcl} \mathcal{H}(\kappa) & : & \mathbb{E}_A[\tau_A^\kappa] < \infty, \\ \mathcal{H}(\nu, \kappa) & : & \mathbb{E}_\nu[\tau_A^\kappa] < \infty. \end{array}$$

Observe that, in the positive recurrent case, these assumptions are not independent when $\nu = \mu$: from basic renewal theory, one has $\mathbb{P}_\mu(\tau_A = k) = (\mathbb{E}_A[\tau_A])^{-1}\mathbb{P}_A(\tau_A \geq k)$ for all $k \geq 1$. Hence, conditions $\mathcal{H}(\mu, \kappa)$ and $\mathcal{H}(\kappa + 1)$ are equivalent.

2.2 Regenerative extensions of general Harris chains.

We now recall the *splitting technique* introduced in [48] for extending the probabilistic structure of the chain with the aim to construct an artificial regeneration set in the general

Harris case. It relies crucially on the notion of *small set*.

Minorization condition. Recall that a set $S \in \mathcal{E}$ is said to be *small* for X if there exist $m \in \mathbb{N}^*$, $\delta > 0$ and a probability measure Φ supported by S such that, for all $x \in S$, $B \in \mathcal{E}$,

$$\Pi^m(x, B) \geq \delta \Phi(B), \tag{2}$$

denoting by Π^m the m-th iterate of the transition kernel Π. In the sequel, (2) is referred to as the *minorization condition* $\mathcal{M}(m, S, \delta, \Phi)$. Recall that accessible small sets always exist for ψ-irreducible chains: any set $B \in \mathcal{E}$ such that $\psi(B) > 0$ contains such a set (*cf* [38]).

The Nummelin technique. We now explain how to construct the atomic chain onto which the initial chain X is embedded. Suppose that X satisfies $\mathcal{M} = \mathcal{M}(m, S, \delta, \Gamma)$ for $S \in \mathcal{E}$ such that $\psi(S) > 0$. Rather than replacing the initial chain X by the chain $\{(X_{nm}, ..., X_{n(m+1)-1})\}_{n\in\mathbb{N}}$, 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 δ by defining the joint distribution $\mathbb{P}_{\nu,\mathcal{M}}$ whose construction relies on the following randomization of the transition probability Π each time the chain hits S. Note that it occurs with probability one since the chain is Harris recurrent and $\psi(S) > 0$. If $X_n \in S$, and

- if $Y_n = 1$ (occurs with probability $\delta \in]0, 1[$), then $X_{n+1} \sim \Phi$,
- if $Y_n = 0$, then $X_{n+1} \sim (1 - \delta)^{-1}(\Pi(X_n, .) - \delta\Phi(.))$.

Set $Ber_\delta(\beta) = \delta\beta + (1 - \delta)(1 - \beta)$ for $\beta \in \{0, 1\}$. We have thus constructed the *split chain* $\{(X_n, Y_n)\}_{n\in\mathbb{N}}$, valued in $E \times \{0, 1\}$ with transition kernel $\Pi_{\mathcal{M}}$ defined by

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

$$\Pi_{\mathcal{M}}\left((x, \beta), B \times \left\{\beta'\right\}\right) = \Pi(x, B) \times Ber_\delta(\beta'),$$

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

$$\begin{cases} \Pi_{\mathcal{M}}((x, 1), B \times \{\beta'\}) &= \Phi(B) \times Ber_\delta(\beta'), \\ \Pi_{\mathcal{M}}((x, 0), B \times \{\beta'\}) &= (1 - \delta)^{-1}(\Pi(x, B) - \delta\Phi(B)) \times Ber_\delta(\beta'). \end{cases}$$

The key point of the construction relies on the fact that $A_S = S \times \{1\}$ is an atom for the bivariate Markov chain (X, Y), which inherits all its communication and stochastic stability properties from X. In particular, when any assumption $\tilde{\mathcal{H}}$ among the following is satisfied by X for a certain accessible small set S, it also holds for any other accessible small set and the analogue assumption $\mathcal{H}$ in the atomic case is then automatically fulfilled by the split chain (X, Y), refer to Chapter 14 in [47].

Moment assumptions. Let $\kappa \geq 1$. Consider the assumptions stated below.

$$\begin{array}{lcl} \tilde{\mathcal{H}}(\kappa) & : & \sup_{x \in S} \mathbb{E}_x[\tau_S^\kappa] < \infty, \\ \tilde{\mathcal{H}}(\nu, \kappa) & : & \mathbb{E}_\nu[\tau_S^\kappa] < \infty. \end{array}$$

Plug-in approximation of the Nummelin extension. Here we assume further that the conditional distributions $\{\Pi(x, dy)\}_{x \in E}$ and the initial distribution ν are dominated by a σ-finite measure λ of reference, so that $\nu(dy) = f(y)\lambda(dy)$ and $\Pi(x, dy) = \pi(x, y)\lambda(dy)$ for all $x \in E$. For simplicity, we suppose that condition $\mathcal{M}$ is fulfilled with $m = 1$. Hence, Φ is absolutely continuous with respect to λ too, and, setting $\Phi(dy) = \phi(y).\lambda(dy)$,

$$\forall x \in S,\ \pi(x, y) \geq \delta\phi(y),\ \lambda(dy)\text{-almost surely .} \tag{3}$$

If we were able to generate binary random variables $Y_1, ..., Y_n$, so that $((X_1, Y_1), ..., (X_n, Y_n))$ be a realization of the split chain described above, then we could divide the sample path $X^{(n)} = (X_1, ..., X_n)$ into regeneration blocks. Therefore, knowledge of π over S^2 is required to draw $Y_1, ..., Y_n$ this way. The distribution $\mathcal{L}^{(n)}(\pi, S, \delta, \phi, x^{(n+1)})$ of $Y^{(n)} = (Y_1, ..., Y_n)$ conditioned on $X^{(n+1)} = (x_1, ..., x_{n+1})$ is the tensor product of Bernoulli distributions given by: $\forall \beta^{(n)} = (\beta_1, ..., \beta_n) \in \{0, 1\}^n$, $\forall x^{(n+1)} = (x_1, ..., x_{n+1}) \in E^{n+1}$,

$$\mathbb{P}_\nu(Y^{(n)} = \beta^{(n)} \mid X^{(n+1)} = x^{(n+1)}) = \prod_{i=1}^{n} \mathbb{P}_\nu(Y_i = \beta_i \mid X_i = x_i,\ X_{i+1} = x_{i+1})$$

with for $1 \leq i \leq n$: if $x_i \notin S$,

$$\mathbb{P}_\nu(Y_i = \beta_i \mid X_i = x_i,\ X_{i+1} = x_{i+1}) = Ber_\delta(\beta_i),$$

and if $x_i \in S$,

$$\left\{ \begin{array}{lcl} \mathbb{P}_\nu(Y_i = 1 \mid X_i = x_i,\ X_{i+1} = x_{i+1}) & = & \delta\phi(x_{i+1})/\pi(x_i, x_{i+1}), \\ \mathbb{P}_\nu(Y_i = 0 \mid X_i = x_i,\ X_{i+1} = x_{i+1}) & = & 1 - \delta\phi(x_{i+1})/\pi(x_i, x_{i+1}). \end{array} \right. \tag{4}$$

In short, given the sample path $X^{(n+1)}$, the Y_i's are Bernoulli r.v.'s with parameter δ, unless X has hit the small set S at time i: in this case Y_i is drawn from the Bernoulli distribution with parameter $\delta\phi(X_{i+1})/\pi(X_i, X_{i+1})$. Our proposition for constructing data blocks relies on approximating this construction by computing first an estimate $\hat{\pi}_n(x, y)$ of the transition density $\pi(x, y)$ from data $X_1, ..., X_{n+1}$, and then drawing a random vector $(\hat{Y}_1, ..., \hat{Y}_n)$ from the distribution $\mathcal{L}^{(n)}(\hat{\pi}_n, S, \delta, \phi, X^{(n+1)})$, obtained by simply plugging $\hat{\pi}_n$ into (4), assuming that the estimate $\hat{\pi}_n(x, y)$ is picked so that $\hat{\pi}_n(x, y) \geq \delta\phi(y)$, $\lambda(dy)$ a.s., and $\hat{\pi}_n(X_i, X_{i+1}) > 0$, $1 \leq i \leq n$.

From a practical viewpoint, it suffices to draw the $\widehat{Y}_i$'s only at times i when the chain hits the small set S, $\widehat{Y}_i$ indicating whether the trajectory should be divided at time point

i or not. This way, setting $\widehat{l}_n = \sum_{1\leq k\leq n} \mathbb{I}\{(X_k, \widehat{Y}_k) \in S \times \{1\}\}$ one gets a sequence of *approximate renewal times*,

$$\widehat{\tau}_{A_S}(j+1) = \inf\{n \geq 1 + \widehat{\tau}_{A_S}(j) / (X_n, \widehat{Y}_n) \in S \times \{1\}\}, \text{ for } 1 \leq j \leq \widehat{l}_n - 1, \tag{5}$$

with $\widehat{\tau}_{A_S}(0) = 0$ by convention and forms the *approximate regeneration blocks* $\widehat{\mathcal{B}}_1, ..., \widehat{\mathcal{B}}_{\widehat{l}_n - 1}$.

The question of accuracy of this approximation has been tackled in [11] using a *coupling approach*. Precisely, the authors established a sharp bound for the deviation between the distribution of $((X_i, Y_i))_{1\leq i\leq n}$ and that of $((X_i, \widehat{Y}_i))_{1\leq i\leq n}$ in Mallows distance, see Theorem 4.1 therein. This essentially depends on the rate of the *mean squared error* (MSE)

$$\mathcal{R}_n(\widehat{\pi}_n,\ \pi) = \mathbb{E}[(\sup_{(x,y)\in S^2} |\widehat{\pi}_n(x,y) - \pi(x,y)|)^2], \tag{6}$$

with the sup norm over $S \times S$ as a loss function, under the following conditions:

A1. the parameters S and ϕ in (3) are chosen so that $\inf_{x\in S} \phi(x) > 0$,

A2. $\sup_{(x,y)\in S^2} \pi(x,y) < \infty$ and $\mathbb{P}_\nu$-almost surely $\sup_{n\in\mathbb{N}} \sup_{(x,y)\in S^2} \widehat{\pi}_n(x,y) < \infty$.

Before showing how the renewal properties of (the Nummelin extension of) Harris Markov chains may be practically exploited for statistical analysis of extremal events, a few remarks are in order.

Remark 1 (On estimating the transition density $\pi(x,y)$ over $S\times S$) Within the time-series asymptotic framework, the problem of estimating the transition density of a Harris recurrent Markov chain has received much attention in the statistical literature and estimation rates for the different estimators proposed have been established under suitable ergodicity conditions (including the null recurrent case, see [39]) and various smoothness assumptions on the marginal densities $\int_{z\in E} \mu(dz)\pi(z,x)$ and $\int_{z\in E} \mu(dz)\pi(z,x)\cdot\pi(x,y)$, see [14, 23, 6, 18]. For instance, under standard Hölder constraints of order s, the typical rate for the MSE (6) is of order $n^{-s/(s+1)}$.

Remark 2 (Data-driven choice of the $\mathcal{M}$-parameters) We point out that, so far, knowledge of the parameters $(S,\ \delta,\ \phi)$ of condition (3) is required for constructing the pseudo-cycles. However, in Section 5 of [9] an entirely data-driven method for picking those tuning parameters has been proposed, which aims at solving a trade-off for maximizing (an approximation of) the expected number of pseudo-cycles generated. The principle underlying this selection procedure is briefly recalled in § 5.2, where it is applied to some simulation datasets.

3 Preliminary results

Here we begin by briefly recalling the connection between the (pseudo-) regeneration properties of a Harris chain X and the extremal behavior of sequences of type $f(X) = \{f(X_n)\}_{n\in\mathbb{N}}$,

firstly pointed out in the seminal contribution of [56], see also [3] and [32]. The case of sequences of type $\{F(X_n, \ldots, X_{n-k})\}_{n \geq k}$ may be investigated in a similar manner. We also gather preliminary remarks, in order to give an insight into the principle underlying the statistical methods studied later on.

Cycle submaxima. We first consider the case when X possesses a known accessible atom A. For $j \geq 1$, we define the *submaximum* over the j-th cycle of the sample path:

$$\zeta_j(f) = \max_{1+\tau_A(j) \leq i \leq \tau_A(j+1)} f(X_i). \tag{7}$$

By virtue of the strong Markov property, the $\zeta_j(f)$'s are i.i.d. random variables with common df $G_f(x) = \mathbb{P}_A(\max_{1 \leq i \leq \tau_A} f(X_i) \leq x)$. The following result established in [56] shows that the limiting distribution of the sample maximum of X is entirely determined by the tail behavior of the df G_f and relies on the crucial observation that the maximum value $M_n(f) = \max_{1 \leq i \leq n} f(X_i)$ taken by the sequence $f(X)$ over a trajectory of length n, may be naturally expressed in terms of *submaxima* over regeneration cycles as follows

$$M_n(f) = \max\{\zeta_0(f), \max_{1 \leq j \leq l_n - 1} \zeta_j(f), \zeta_{l_n}^{(n)}(f)\}, \tag{8}$$

where $l_n = \sum_{i=1}^{n} \mathbb{I}\{X_i \in A\}$ denotes the number of visits of X to the regeneration set A until time n, $\zeta_0(f) = \max_{1 \leq i \leq \tau_A} f(X_i)$ and $\zeta_{l_n}^{(n)}(f) = \max_{1+\tau_A(l_n) \leq i \leq n} f(X_i)$ denote the maxima over the nonregenerative data blocks, and with the usual convention that maximum over an empty set equals to $-\infty$.

We stress that the number l_n of cycle submaxima over a trajectory of finite length n is random, and that the cycle submaxima are generally not independent similar to the blocks of which lengths sum up to n.

Proposition 1 *(Rootzén, 1988) Let $\alpha = \mathbb{E}_A[\tau_A]$ be the mean return time to the atom A. Under the assumption that the first (nonregenerative) block does not affect the extremal behavior, that is to say that*

$$\mathbb{P}_\nu(\zeta_0(f) > \max_{1 \leq k \leq l} \zeta_k(f)) \to 0 \text{ as } l \to \infty, \tag{9}$$

we then have

$$\sup_{x \in \mathbb{R}} |\mathbb{P}_\nu(M_n(f) \leq x) - G_f(x)^{n/\alpha}| \to 0 \text{ as } n \to \infty. \tag{10}$$

Remark 3 We point out that, under the further assumption that conditions **A1** and **A2** are fulfilled by the chain X, it has been proved in [31] that the rate at which convergence (10) takes place is actually of order $O(n^{-1/2}\log^{3/2}(n))$ as $n \to \infty$.

As shown by the result stated above, as soon as condition (9) is fulfilled, the asymptotic behavior of the sample maximum is entirely determined by the tail properties of the df

$G_f(dx)$. In particular, the limiting distribution of $M_n(f)$ for a suitable normalization is the extreme df $H_\xi(dx)$ of shape parameter $\xi \in \mathbb{R}$ (with $H_\xi(x) = \exp(-(1+\xi x)^{-1/\xi})\mathbb{I}\{1+\xi x > 0\}$ when $\xi \neq 0$ and $H_0(x) = \exp(-\exp(-x))$) if and only if G_f belongs to the maximum domain of attraction say $MDA(H_\xi)$ of the latter df (refer to [51] for basics in extreme value theory). Thus, when $G_f \in MDA(H_\xi)$, there are sequences of normalizing constants a_n and b_n such that $G_f(a_n x + b_n)^n \to H_\xi(x)$ as $n \to \infty$, we then have $\mathbb{P}_\nu(M_n(f) \leq a'_n x + b'_n) \to H_\xi(x)$ as $n \to \infty$, with $a'_n = a_{\lfloor n/\alpha \rfloor}$ and $b'_n = b_{\lfloor n/\alpha \rfloor}$.

Estimation of the cycle submaximum cdf. In the atomic case, the cdf G_f of the cycle submaxima, $\zeta_j(f)$ with $j \geq 1$, may be naturally estimated by computing its empirical counterpart from the observation of a random number $l_n - 1$ of regenerative cycles, namely

$$G_{f,n}(x) = \frac{1}{l_n - 1} \sum_{j=1}^{l_n - 1} \mathbb{I}\{\zeta_j(f) \leq x\}, \tag{11}$$

with $G_{f,n} \equiv 0$ by convention when $l_n \leq 1$. Since, by Harris recurrence, $l_n \sim n/\alpha$ $\mathbb{P}_\nu$-almost surely as $n \to \infty$, it immediately follows from Glivenko-Cantelli's theorem that

$$\sup_{x \in \mathbb{R}} |G_{f,n}(x) - G_f(x)| \to 0, \ \mathbb{P}_\nu\text{-almost surely.} \tag{12}$$

Furthermore, by the law of iterated logarithm (LIL), we also have $\sup_{x \in \mathbb{R}} |G_{f,n}(x) - G_f(x)| = O_{\mathbb{P}_\nu}(\sqrt{\log\log(n)/n})$.

As previously noticed, cycles submaxima of the split chain are generally not observable in the general Harris case. However, regeneration-based statistical procedures may be directly extended by considering the submaxima over the approximate regeneration cycles

$$\hat{\zeta}_j(f) = \max_{1 + \hat{\tau}_{A_S}(j) \leq i \leq \hat{\tau}_{A_S}(j+1)} f(X_i), \tag{13}$$

for $i = 1, \ldots, \hat{l}_n - 1$, and computing the empirical counterpart as if they were the 'true' cycle submaxima

$$\hat{G}_{f,n}(x) = \frac{1}{\hat{l}_n - 1} \sum_{j=1}^{\hat{l}_n - 1} \mathbb{I}\{\hat{\zeta}_j(f) \leq x\}, \tag{14}$$

with, by convention, $\hat{G}_{f,n} \equiv 0$ if $\hat{l}_n \leq 1$. As shown by the next theorem, using the approximate cycle submaxima instead of the true ones in the average (11) does not affect the convergence, provided that $\hat{\pi}_n(x, y)$ is consistent in the MSE sense over S^2. Following in the footsteps of [9], the proof essentially relies on a *coupling argument*. Technical details are postponed to §6.1.

Theorem 2 *Let $f : (E, \mathcal{E}) \to \mathbb{R}$ be a measurable function. Suppose that conditions (3), **A1** and **A2** are fulfilled by the chain X. Assume further that $\mathcal{R}_n(\hat{\pi}_n, \pi) \to 0$ as $n \to \infty$.*

Then, $\widehat{G}_{f,n}(x)$ *is a consistent estimator of* $G_f(x) = \mathbb{P}_{A_S}(\max_{1 \leq i \leq \tau_{A_S}} f(X_i) \leq x)$*, uniformly over* $\mathbb{R}$*: as* $n \to \infty$*,*

$$\sup_{x \in \mathbb{R}} |\widehat{G}_{f,n}(x) - G_f(x)| = O_{\mathbb{P}_\nu}(\mathcal{R}_n(\hat{\pi}_n,\ \pi)^{1/2} \vee \sqrt{\log\log(n)/n}). \tag{15}$$

As shown by Eq. (15) in the above theorem, the loss resulting from the approximation step vanishes as the MSE rate gets closer to the parametric rate.

4 Regeneration-based statistical methods for extremal events

As argued in Proposition 2, the underlying renewal structure of the Harris chain X plays a key role in the analysis of the extremal behavior of the sequence $f(X) = \{f(X_n)\}_{n \in \mathbb{N}}$. The leitmotiv of the present paper is to show that, in the regenerative setup, consistent statistical procedures for extremal events may be derived from the application of standard inference methods introduced in the i.i.d. setting to the cycles submaxima observed over a finite trajectory on the one hand, and on the other hand that, when regeneration cycles cannot be determined by simple examination of the data, *i.e.* in the general Harris case, the latter can be extended straightforwardly by replacing the unknown theoretical submaxima by their approximate versions. We point out that the estimation principle exposed in this paper is by no means restricted to the sole markovian setup, but indeed applies to any possibly continuous-time process for which a regenerative extension can be constructed and simulated from available data, see Chapter 10 in [60]. Throughout this section, f denotes a fixed real-valued measurable function defined on the state space E. In order to lighten notation, we omit to index by the subscript f the distributions we consider.

4.1 Tail estimation based on (approximate) submaxima

In the case when assumption (9) holds, one may derive straightforwardly from (10) estimates of $H^{(n)}(x) = \mathbb{P}_\nu(M_n(f) \leq x)$ as $n \to \infty$ based on the observation of (a random number of) submaxima $\zeta_j(f)$ over a sample path of length N, as proposed in [27, 62]:

$$H_{N,\,l}(x) = (G_N(x))^l, \tag{16}$$

with $l \geq 1$. The next limit result establishes the asymptotic validity of estimator (16) for adequate choices of N and l, extending this way Proposition 3.6 of [27], of which the restrictive asymptotic framework stipulates the observation of a deterministic number of regeneration cycles. Furthermore, it also shows that, under certain conditions, the procedure remains consistent, even if computations are carried out from the approximate regeneration data blocks and one considers estimates of the form $\widehat{H}_{N,\,l}(x) = (\widehat{G}_N(x))^l$.

Proposition 3 *Suppose that assumption (9) holds. Let* $(u_n)_{n \in \mathbb{N}}$ *be a deterministic sequence of real numbers such that* $n(1 - G(u_n))/\alpha \to \eta < \infty$ *as* $n \to \infty$*. Then, we have*

$$H^{(n)}(u_n) \to \exp(-\eta) \ \textit{as}\ n \to \infty. \tag{17}$$

(i) In the regenerative setup, suppose furthermore that $\mathcal{H}(\nu, 1)$ *is fulfilled. Let* $N = N(n)$ *and* $l = l(n)$ *be picked such that* $n = o(\sqrt{N(n)/\log\log N(n)})$ *and* $l(n) \sim l_n$ *as* $n \to \infty$. *Then,*

$$H_{N(n),\, l(n)}(u_n)/H^{(n)}(u_n) \to 1 \text{ in } \mathbb{P}_\nu\text{- probability, as } n \to \infty. \tag{18}$$

(ii) In the general Harris recurrent framework, suppose that ***A1****,* ***A2*** *and* $\tilde{\mathcal{H}}(\nu, 1)$ *hold and* $\mathcal{R}_N(\hat{\pi}_N,\ \pi) = O(N^{-1+\epsilon})$ *as* $N \to \infty$ *for some* $\epsilon \in]0, 1[$. *If* $(N(n),\ l(n))$ *is chosen so that, as* $n \to \infty$, $l(n) \sim \hat{l}_n$ *and* $n = o(N(n)^{(1-\epsilon)/2})$, *then*

$$\hat{H}_{N(n),\, l(n)}(u_n)/H^{(n)}(u_n) \to 1 \text{ in } \mathbb{P}_\nu\text{- probability, as } n \to \infty. \tag{19}$$

Remark 4 (On estimating the sequence of "high threshold levels" $(u_n)_{n\in\mathbb{N}}$) Of course, the sequence $(u_n)_{n\in\mathbb{N}}$ of thresholds such that

$$n(1 - G(u_n))/\alpha \to \eta < \infty, \text{ as } n \to \infty, \tag{20}$$

must be estimated in practice. Indeed, the random levels may be picked, empirically from a sample path of length $N(n)$, as follows: $u_n = G^{-1}_{N(n)}(1 - \eta/l(n))$ in the regenerative setting and $u_n = \hat{G}^{-1}_{N(n)}(1 - \eta/\hat{l}(n))$ in the general case. Then, one may easily derive from the argument in §6.2 that, under the assumptions of the first part (respectively, of the second part) of Proposition 3, $H_{N(n),\, l(n)}(u_n)$ (respectively, $\hat{H}_{N(n),\, l(n)}(u_n)$) still converges to $\exp(-\eta)$ as $n \to \infty$.

This result indicates that, in the most favorable case, observation of a trajectory of length $N(n)$, with $n^2 = o(N(n)/\log\log N(n))$ as $n \to \infty$, is required for estimating consistently the extremal behavior of $f(X)$ over a trajectory of length n in this general setting. As shown below, it is nevertheless possible to estimate tail features of the sample maximum $M_n(f)$ from the observation of a sample path of length n only, when assuming some specific type of behavior for the latter, namely under a maximum domain of attraction hypothesis. As a matter of fact, if one assumes that $G \in \mathcal{MDA}(H_\xi)$ for some $\xi \in \mathbb{R}$, of which sign is *a priori* known, one may implement classical inference procedures (refer to § 6.4 in [24] for instance) from the observed submaxima $\zeta_1(f), \ldots, \zeta_{l_n-1}(f)$ for estimating the shape parameter ξ of the extremal distribution, as well as the norming constants a_n and b_n. We illustrate this point in the Fréchet case, *i.e.* when $\xi > 0$, through the example of the Hill's inference method in § 4.3.

4.2 The extremal index

When the regenerative chain X is positive recurrent, with limiting probability distribution μ given by (1), there always exists some index $\theta = \theta(f)$, namely the *extremal index* of the sequence $\{f(X_n)\}_{n\in\mathbb{N}}$ (see [49, 42] for instance), such that

$$\mathbb{P}_\mu(M_n(f) \leq u_n) \underset{n\to\infty}{\sim} F(u_n)^{n\theta}, \tag{21}$$

for any sequence $u_n = u_n(\eta)$ such that (20) holds, denoting by $F(x) = \alpha^{-1}\mathbb{E}_A[\sum_{i=1}^{\tau_A} \mathbb{I}\{f(X_i) \leq x\}]$ the cdf of $f(X_1)$ in steady-state, *i.e.* under $\mathbb{P}_\mu$. Indeed, any positive recurrent Markov chain is *strongly mixing*, see Theorem A in [7] for instance. Hence, it *a fortiori* fulfills *Leadbetter's mixing condition* $D(u_n)$, which guarantees the existence of the extremal index, see [41]. In this case, as remarked in [56], we deduce from Proposition 1 and (21) that

$$\theta = \lim_{n\to\infty} \frac{\log(\mathbb{P}_A(\max_{1\leq i\leq \tau_A} f(X_i) \leq u_n)/\alpha}{\log(\mathbb{E}_A[\sum_{i=1}^{\tau_A} \mathbb{I}\{f(X_i) \leq u_n\}]/\alpha)} \tag{22}$$

$$= \lim_{n\to\infty} \frac{\mathbb{P}_A(\max_{1\leq i\leq \tau_A} f(X_i) > u_n)}{\mathbb{E}_A[\sum_{i=1}^{\tau_A} \mathbb{I}\{f(X_i) > u_n\}]}, \tag{23}$$

provided that $\theta > 0$, which we assume from now on. In this respect, we mention that it has been established in [55] that the extremal index of a Markov chain cannot be zero, as soon as the chain is geometrically ergodic and fulfills an additional technical condition, see Theorem 4.1 therein. Hence, the probability of exceeding a sufficiently high threshold within a regenerative cycle is proportional to the mean time spent above the latter between consecutive regeneration times, with the index θ as proportionality constant. Notice that, in the i.i.d. setup, by taking the whole state space as an atom ($A = \mathcal{X}$, so that $\tau_A \equiv 1$), one immediately rediscovers that $\theta = 1$. Then, Proposition 1 combined with (21) also entails that for all ξ in $\mathbb{R}$,

$$G \in MDA(H_\xi) \Leftrightarrow F \in MDA(H_\xi). \tag{24}$$

The "blocks method" with (pseudo-) regeneration blocks. For regenerative chains, we may propose a natural estimate of the extremal index θ based on Eq. (23) from the observation of a trajectory of length n,

$$\theta_n(u) = \frac{\sum_{j=1}^{l_n - 1} \mathbb{I}\{\zeta_j(f) > u\}}{\sum_{i=1}^{n} \mathbb{I}\{f(X_i) > u\}}, \tag{25}$$

with the convention that $\theta_n(u) = 0$ if $M_n(f) < u$ and taking $F_n(x) = n^{-1}\sum_{1\leq i\leq n} \mathbb{I}\{f(X_i) \leq x\}$ as a natural empirical estimate of its (strong) limit $F(x)$. For general Harris chains, one naturally considers the counterpart computed from the approximate regeneration blocks

$$\hat{\theta}_n(u) = \frac{\sum_{j=1}^{\hat{l}_n - 1} \mathbb{I}\{\hat{\zeta}_j(f) > u\}}{\sum_{i=1}^{n} \mathbb{I}\{f(X_i) > u\}}, \tag{26}$$

with $\hat{\theta}_n(u) = 1$ by convention when $M_n(f) < u$. Beyond the consistency property of the estimators thus produced, stated in the next result, this method has an important advantage that makes it attractive from a practical perspective: blocks are here entirely determined by the data (up to the approximation step, see §5.2.1), in contrast to standard blocking-techniques of which performance crucially depends on the deterministic length arbitrarily chosen at hand for the blocks and is very sensitive to changes in block lengths.

Proposition 4 *Suppose that* $\theta > 0$. *Let* $(r_n)_{n\in\mathbb{N}}$ *increase to infinity in a way that* $r_n = o(\sqrt{n/\log\log n})$ *as* $n \to \infty$. *Consider* $(v_n)_{n\in\mathbb{N}}$ *such that* $r_n(1 - G_f(v_n))/\alpha \to \eta < \infty$ *as* $n \to \infty$.

(i) In the regenerative case, suppose that $\mathcal{H}(v, 1)$ *and* $\mathcal{H}(2)$ *are fulfilled. Then,*

$$\theta_n(v_n) \to \theta \ \mathbb{P}_v\text{-almost surely, as } n \to \infty. \tag{27}$$

(ii) In the general case, assume that $\tilde{\mathcal{H}}(v, 1)$ *and* $\tilde{\mathcal{H}}(4)$ *are satisfied. Then,*

$$\hat{\theta}_n(v_n) \to \theta \text{ in } \mathbb{P}_v\text{-probability, as } n \to \infty. \tag{28}$$

Remark 5 We point out that, in practice, the levels v_n are generally picked as a function of the data. In this respect, it may be easily seen that results stated in Proposition 4 remain valid if these thresholds are chosen in a similar manner as in Remark 4, taking v_n equal to $G_n^{-1}(1 - \eta/r_n)$ or $\hat{G}_n^{-1}(1 - \eta/r_n)$. Beyond the appealing practical advantage of the "regeneration blocks method" previously mentioned, it is worth noticing that, to our knowledge, (27) is the sole strong consistency result related to the statistical estimation of the extremal index available in the literature until now.

Remark 6 (The extremal index θ seen as a limiting conditional probability) Exploiting the regeneration properties of the sequence $f(X)$, it has also been showed in [56] that

$$\theta = \lim_{n\to\infty} \mathbb{P}_A(\max_{2\leq i\leq \tau_A} f(X_i) \leq u_n \mid f(X_1) > u_n). \tag{29}$$

for any sequence u_n such that (20) holds. Based on Eq. (29), it is natural to propose

$$\theta'_n(u) = \frac{\sum_{j=1}^{l_n-1} \mathbb{I}\{f(X_{1+\tau_A(j)}) > u,\ \max_{2+\tau_A(j)\leq i\leq \tau_A(j+1)} f(X_i) \leq u\}}{\sum_{j=1}^{l_n-1} \mathbb{I}\{f(X_{1+\tau_A(j)}) > u\}}, \tag{30}$$

as an estimate of θ, for a properly chosen level $u > 0$. This may be seen as a "regenerative version" of the so-called *runs* estimator, see [37], in the sense that it measures the clustering tendency of high threshold exceedances within regeneration cycles only. Naturally, in the general Harris setting, one considers the estimate obtained by replacing the renewal times by their approximate versions in (30). Except for slight modifications, the same argument as the one to which Proposition 4's proof appeals could be used for investigating asymptotic properties of these estimators. Owing to space limitations, this question will be handled in a forthcoming paper.

4.3 The regeneration-based Hill estimator

The crucial equivalence (24) holds in particular in the Fréchet case, *i.e.* for $\xi > 0$. Recall that the assumption that a df F belongs to $MDA(H_\xi)$ then classically amounts to supposing it satisfies the tail regularity condition

$$1 - F(x) = L(x)x^{-a}, \tag{31}$$

where $\alpha = \xi^{-1}$ and L is a slowly varying function, *i.e.* a function L such that $L(tx)/L(x) \to 1$ as $x \to \infty$ for any $t > 0$, *cf* Theorem 8.13.2 in [13]. Since the seminal contribution of [33], numerous papers have been devoted to the development and study of statistical methods in the i.i.d. setting for estimating the tail index $\alpha > 0$ of a regularly varying df. Various inference methods, mainly based on an increasing sequence of upper order statistics, have been proposed for dealing with this estimation problem, among which the popular *Hill estimator*, relying on a conditional maximum likelihood approach. Precisely, based on i.i.d. observations $Z_1, ..., Z_n$ drawn from F, the Hill estimator is given by

$$H^Z_{k,n} = \left(k^{-1} \sum_{i=1}^{k} \log \frac{Z_{(i)}}{Z_{(k+1)}} \right)^{-1}, \tag{32}$$

with $1 \leq k < n$, and where $Z_{(i)}$ denotes the i-th largest order statistic of the sample $Z^{(n)} = (Z_1, ..., Z_n)$. In [22], strong consistency of this estimator has been established when $k = k_n \to \infty$ at a suitable rate, namely for $k_n = o(n)$ and $\log\log n = o(k_n)$ as $n \to \infty$, as well as asymptotic normality, see [29, 20]: under further conditions on F and k_n related to the slowly varying function L, $\sqrt{k_n}(H^Z_{k_n,n} - \alpha) \Rightarrow \mathcal{N}(0,\, \alpha^2)$.

Now let us define the *regeneration-based Hill estimator* from the observation of the $l_n - 1$ submaxima $\zeta_1(f), ..., \zeta_{l_n-1}(f)$, denoting by $\zeta_{(j)}(f)$ the j-th largest submaximum,

$$\alpha_{n,\,k} = \left(k^{-1} \sum_{i=1}^{k} \log \frac{\zeta_{(i)}(f)}{\zeta_{(k+1)}(f)} \right)^{-1}, \tag{33}$$

with $1 \leq k \leq l_n - 1$ when $l_n > 1$. Given that $l_n \to \infty$, $\mathbb{P}_\nu$- a.s. as $n \to \infty$, asymptotic results established in the case of i.i.d. observations extend straightforwardly to our setting, see part (i) of Proposition 5 below. Notice that in the i.i.d. setup, cycles only comprise a single observation, so that our regeneration-based estimator is exactly a Hill estimator.
In the general Harris case, the same estimator may be considered, except that, of course, approximate submaxima are used for computation:

$$\hat{\alpha}_{n,\,k} = \left(k^{-1} \sum_{i=1}^{k} \log \frac{\hat{\zeta}_{(i)}(f)}{\hat{\zeta}_{(k+1)}(f)} \right)^{-1}, \tag{34}$$

with $1 \leq k \leq \hat{l}_n - 1$ when $\hat{l}_n > 1$. As shown by the next result, consistency is not spoiled by the approximation step, provided that the latter is based on a sufficiently accurate estimator $\hat{\pi}$. In order to construct gaussian asymptotic confidence intervals, we also consider the estimate $\hat{\alpha}^{(N)}_{n,k}$, which is still given by Eq. (34) except that the transition estimate used in the approximation step is based on a trajectory of length N. Notice that $\hat{\alpha}_{n,k} = \hat{\alpha}^{(n)}_{n,k}$ with this notation.

In order to formulate the next result, we consider the following hypothesis.

VM Assumption. (Von Mises condition, [28]) Let $\gamma \geq 0$. Suppose $\bar{G}_f(x) = L(x)x^{-a}$,

$$\lim_{x\to\infty} \frac{\bar{G}_f(tx)/\bar{G}_f(x) - t^{-a}}{b(x)} = t^{-a}\frac{t^{-\gamma}-1}{\gamma}, \quad t > 0$$

where $b(x)$ is a measurable function of constant sign, and with, by convention, $(t^{-\gamma}-1)/\gamma = \log t$ when $\gamma = 0$.

Proposition 5 *Suppose that* $F_\mu \in \mathrm{MDA}(H_{a^{-1}})$ *with* $a > 0$. *Let* $\{k(n)\}$ *be an increasing sequence of integers such that:* $k(n) < n$, $k(n) = o(n)$ *and* $\log\log n = o(k(n))$ *as* $n \to \infty$.

(i) Then the regeneration-based Hill estimator is strongly consistent

$$a_{n,\,k(l_n)} \to a \ \mathbb{P}_\nu\text{- almost surely, as } n \to \infty. \tag{35}$$

Under the **VM assumption**, *and the further condition that*

$$\lim_{n\to\infty} \sqrt{k(n)}\, b(G_f^{-1}(1 - k(n)/n)) = 0, \tag{36}$$

it is also asymptotically normal in the sense that

$$\sqrt{k(l_n)}(a_{n,\,k(l_n)} - a) \Rightarrow \mathcal{N}(0,\ a^2) \text{ under } \mathbb{P}_\nu, \text{ as } n \to \infty. \tag{37}$$

(ii) In the general Harris case, if ***A1*** *and* ***A2*** *are furthermore fulfilled, and* $k = k(n)$ *is chosen such that* $\mathcal{R}_n(\hat{\pi}_n, \pi)^{1/2} n \log n = o(k(n))$, *then*

$$\hat{a}_{n,\,k(\hat{l}_n)} \to a \text{ in } \mathbb{P}_\nu\text{- probability, as } n \to \infty. \tag{38}$$

(iii) Suppose also that G_f *satisfies the* **VM assumption**, *and* $k(n)$ *is chosen accordingly, as in (i). Under* ***A1*** *and* ***A2***, *let* $(m_n)_{n\in\mathbb{N}}$ *be a sequence of integers increasing to infinity such that* $m_n \mathcal{R}_n(\hat{\pi}_n, \pi)^{1/2}/\sqrt{k(m_n)} \to 0$ *as* $n \to \infty$, *then*

$$\sqrt{k(\hat{l}_{m_n})}(\hat{a}^{(n)}_{m_n,\,k(\hat{l}_{m_n})} - a) \Rightarrow \mathcal{N}(0,\ a^2) \text{ under } \mathbb{P}_\nu, \text{ as } n \to \infty. \tag{39}$$

Before investigating the practical performance of the extreme-value regeneration-based statistics introduced in this paper on several examples, we gather a few remarks.

Remark 7 (Strong consistency of the regeneration-based Hill estimator) It is noteworthy that the tail index estimator (33) is proved strongly consistent under mild conditions in the regenerative setting. The alternative method proposed in [52] only establishes the consistency of the standard Hill estimator, though in a general linear time series framework.

Remark 8 (SUBSAMPLING) Notice also that condition "$\mathcal{R}_n(\hat{\pi}_n, \pi)^{1/2} n \log n = o(k(n))$" in (ii) may not hold for certain $k(n)$, in the typical case for instance where the slowly varying function appearing in the distribution of the submaximum is logarithmic. However, as shown by the next assertion, it is always possible to surmount this difficulty by taking a subsampling size m_n such that the conditions of (iii) holds, the problem of picking m_n in an optimal fashion in this setup nevertheless remains open.

Remark 9 (ON CHOOSING THE NUMBER k OF LARGEST (APPROXIMATE) SUBMAXIMA) Given the observed number $l > 2$ (l_n or $\hat{l}_n$) of (approximate) renewal times within the available data series, one may pick the parameter $k \in 1, \ldots, l-1$ according to standard methods in the i.i.d. setup. A standard approach consists of picking up the value of k that minimizes the estimated MSE

$$\widehat{MSE}(k) = \widehat{H}^2_{k,\,n}/k + (H_{k,\,n} - \widehat{H}_{k,\,n})^2,$$

where $\widehat{H}_{k,\,n}$ is a bias corrected version of the Hill estimator (32). Jackknife or analytical methods can be used in this aim, see [8, 26]. In our setting, one may proceed in a similar fashion, with the sole difference that here one works conditionally upon the random number of observed (approximate) submaxima.

5 Simulation studies

As an illustration, we now apply the inference methods previously described to some simulated markovian data sets. For comparison purposes from a nonparametric perspective, *i.e.* without exploiting the parametric form of the instrumental Markov model, numerical results obtained by implementing alternative procedures, standing as natural candidates for computing extreme value statistics in the weakly dependent setting are also displayed. Precisely, estimators of the extremal index are also computed using the *blocks method*, the *runs method* (see Eq. (1.4) and Eq. (1.7) in [58]) and the approach developed in [25] for the examples considered below and the standard Hill procedure is implemented for estimating the tail index, as proposed in [52, 53] for a certain class of autoregressive models.

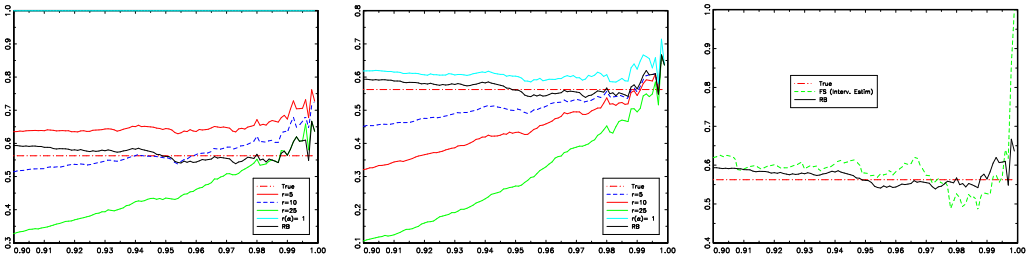
5.1 Regenerative case - Example

We start off with analyzing data simulated from a GI/G/1 queuing in absence of prior knowledge on the underlying model except the regenerative markovian structure. We focus on the sequence $W = (W_n)_{n\geq 1}$ of waiting times.. Classically, one has

$$W_{n+1} = (W_n + U_n - \Delta T_{n+1})_+, \tag{40}$$

where $x_+ = \max(x, 0)$ denotes the positive part of any $x \in \mathbb{R}$ and $(\Delta T_n)_{n\geq 1}$ and $(U_n)_{n\geq 1}$ the sequences of interarrival and service times, assumed i.i.d. and independent from each other. Suppose furthermore that the mean interarrival time $\mathbb{E}[\Delta T_n] = 1/\lambda$ and the mean

service time $\mathbb{E}[U_n] = 1/\mu$ are both finite and that the *load condition* "$\lambda/\mu < 1$" is fulfilled. The discrete-time process W is then a positive recurrent regenerative Markov chain with the "empty file" $A = \{0\}$ as a Harris recurrent atom, see §14.4.1 in [47].



(a) RB versus *blocks* estimators (b) RB versus *runs* estimator (c) RB versus FS *intervals* estimator

Figure 1: Extremal index estimation for waiting times of the M/M/1 queue with $\lambda = 0.2, \mu = 0.8, \theta = 0.56$ (Regenerative-based estimator (RB), *Blocks* method, *Runs* method, *Intervals* estimator (FS), $n = 10000$).

Estimating the extremal index. Explicit computations of the extremal index have been carried out for various Markov processes, see [61, 42, 50, 56]. In the case where interarrival and service times are both exponentially distributed, W is classically geometrically ergodic, so that moment assumptions stipulated in Proposition 4 are fulfilled, refer to §16.1.3 in [47]. Its extremal index is then simply $\theta = (1 - \lambda/\mu)^2$, see [34]. We simulated 1000 sample paths of length $n = 10000$ of such an M/M/1 process, with parameters $\lambda = 0.2, \mu = 0.8$. We have compared the RB estimator (25) of the extremal index (RB standing for *regeneration-based*) to the estimates obtained with the standard *blocks* method, the *runs* method, and the *intervals* estimator proposed in [25]. The *blocks* and *runs* methodologies, see [58] or [1], have been implemented using different choices for the deterministic block length $r + 1$: $r = 5$, 10 and 25, as well as $r = \lfloor n/l_n \rfloor$, the integer part of the estimated mean length of the regenerative blocks, see Figure 1(a) and 1(b). The *intervals estimator* does not require the choice of any tuning parameter similarly to our RB estimator, see Figure 1(c). All estimators have been computed for different values of the threshold level u, corresponding to high percentiles of the simulated data $\{W_i\}_{1 \le i \le n}$ in steady-state, namely from the 90th to the 99.9th. According to our experience, as illustrated in Figure 1, the RB estimator seems to be much less sensitive to the choice of the threshold u than its competitors, of which constructions require in contrast to select the block length except for the *intervals estimator*. Choosing $r = 10$ for the *blocks* method and $r = 5$ for the *runs* method enabled us to obtain the most accurate results. As illustrated by Figure 2, the RB estimator performs very well in this example: for any percentile between the 90th and the 99th, its mean squared error (MSE) is below 3×10^{-3}, whereas the *blocks*, *runs*, or *intervals* estimators

appear as much more unstable, even though, for certain high percentiles, the MSE of the *blocks* estimator may be a bit smaller, one still have to determine the block length r in this case. Here, mean squared errors are computed by averaging over $B = 1000$ replications of the process.

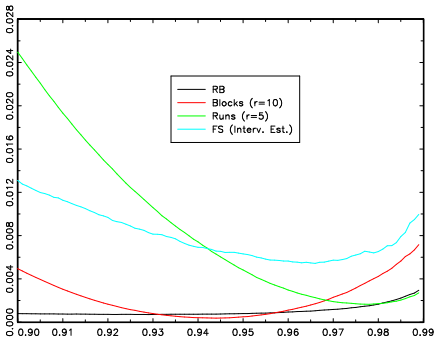


Figure 2: Extremal index estimation - MSE as a function of the threshold level u (waiting times of the M/M/1 queue with $\lambda = 0.2, \mu = 0.8, \theta = 0.56$, Regenerative-based estimator (RB), *Blocks* method ($r = 10$), *Runs* method ($r = 5$), *Intervals* estimator (FS), $n = 10000$, $M = 1000$).

Estimating the tail index. Consider now an M/G/1 queue with Pareto distributed service times. In this case, it is well-known that the right tail behavior of the waiting times in steady-state is governed by that of $U_n - \Delta T_{n+1}$, which is the same as that of U_n, see [3, 5]. Here, the ΔT_n's are exponentially distributed with mean $\lambda^{-1} = 2$ and $\mathbb{P}(U_n > x) = \mathbb{I}\{x > 1/\sqrt{2}\} \cdot x^{-a}$ with $a = 3$. For $n = 10000$, we obtain $l_n = 2589$ regenerative blocks in this simulation. The regeneration-based Hill estimator is plotted in Figure 5.1 as a function of the number k of submaxima used in the average involved in (33), together with a bias corrected version similar to the ones proposed in [8, 26] in the i.i.d. framework. As mentioned in Remark 9, this may serve as a guide for selecting the number k of extreme submaxima used in the computation of the estimator: through our simulations, we found considerable empirical evidence that the bias correction step is of crucial importance when estimating the tail index. However, the gain acquired from correcting the bias this way has proved much more significant for the regeneration based estimate than for the standard Hill estimate directly computed from the observed waiting times, as suggested in [53]: the RB Hill estimate is 0.311 for $1/a = 0.333$ while the standard Hill estimate is 0.252 with respectively $k_{\text{opt}} = 87$ and 276. For comparison purposes, we computed the MSE by averaging over 1000 replications of the process for both bias corrected estimators: the bias corrected regeneration-based Hill estimator enjoys an MSE of $1.13\ 10^{-2}$, while the standard Hill estimator behaves badly, with an MSE of $1.16\ 10^{-1}$. As shown in Table 1, although the RB estimate uses a lot less data (only one

Table 1: Bias-variance trade-off in the Hill estimation (RB Hill: Regeneration based Hill estimator, Standard Hill: Hill estimator suggested in [53], $M = 1000$)

	Squared Bias	Variance	MSE
RB Hill	0.0028	0.0085	0.0113
Standard Hill	0.0531	0.0625	0.1156

per regeneration cycle), potentially discarding other large observations within the cycle, the variance of our estimator is a lot smaller than that of the standard Hill estimator using all the large observations. The usual bias-variance trade-off appearing in tail estimation is as follows: when k is large, the variance of the Hill estimator is reduced but some of the observations used in the computation are not far enough along the tail of the distribution, which increases the bias. Our methodology yields a reduced variance for small values of k, and also maintains the bias at a reasonable level although some high observations are discarded because the selection of the extreme observations follows the dependence structure of the process.

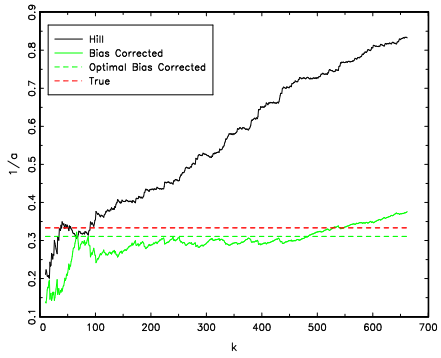


Figure 3: RB Hill estimate for the waiting times of the M/Pareto/1 model ($\lambda = 0.5, \alpha = 3$).

5.2 General Harris case

5.2.1 A data-driven method for splitting up the trajectory into blocks

In the general Harris case, the estimation methods described in Section 4 may be sensitive to the choice of the minorization condition parameters (S, δ, Φ) used in the approximation step. We now recall hints for optimally selecting these parameters in a data-driven fashion, first proposed in [10]. In order to get as many blocks as possible and thus compute meaningful statistics, the latter should ideally be picked, so as to maximize the mean

number of pseudo-regenerative blocks given the observed sample path, namely:

$$N_n(S) = \mathbb{E}_\nu[\sum_{i=1}^{n} \mathbb{I}\{X_i \in S, Y_i = 1\} \,|X^{(n+1)}]. \tag{41}$$

Heuristically, this task involves to finely adjust the size of the small set. Indeed, (41) depends on the frequency at which the chain visits S over a finite length trajectory and on the accuracy of the lower bound in (3) both at the same time, leading to consider the following trade-off: as the size of the small set S increases, the number of points of the path at which the trajectory may be possibly split naturally increases, but, since $\inf_{(x,y)\in S^2} p(x,y)$ then decreases, the probability of drawing $Y_i = 1$ also decreases, see Eq. (4). This gives insight into the fact that better numerical results may be expected in practice when choosing S so as to maximize the expected number of blocks given the data, namely $N_n(S) - 1$.

In absence of prior knowledge of the transition structure of the chain, a possible method for selecting the tuning parameters could be as follows. For simplicity, we suppose here that X takes on real values. Let $\mathcal{S}$ be a collection of compact intervals S and let $\mathcal{U}_S(dy) = \phi_S(y).\lambda(dy)$ denote the uniform distribution on S, where $\phi_S(y) = \mathbb{I}\{y \in S\}/\lambda(S)$ and λ is the Lebesgue measure on $\mathbb{R}$. Clearly, for any $S \in \mathcal{S}$, we have $p(x,y) \geq \delta(S)\phi_S(y)$ for all x, y in S, with $\delta(S) = \lambda(S).\inf_{(x,y)\in S^2} p(x,y)$. We point out that other approaches for practically determining small sets and establishing sharp minorization conditions have been considered, which do not involve uniform distributions, see [55] for instance. When $\delta(S) > 0$, the theoretical criterion (41), that we would ideally seek to maximize over $\mathcal{S}$, may be re-written as follows

$$N_n(S) = \inf_{(x,y)\in S^2} p(x,y) \times \sum_{i=1}^{n} \frac{\mathbb{I}\{(X_i, X_{i+1}) \in S^2\}}{p(X_i, X_{i+1})}. \tag{42}$$

An empirical counterpart $\widehat{N}_n(S)$ of (42) may be computed by simply replacing the unknown transition density $p(x,y)$ by an estimate $p_n(x,y)$ in (42). One is then led to maximize the practical empirical criterion over $\mathcal{S}$ in order to find

$$S^* = \arg\max_{S\in\mathcal{S}} \widehat{N}_n(S). \tag{43}$$

As recalled in Remark 1, numerous estimators of the transition density of Harris recurrent chains have been proposed and studied in the literature, among which the standard *Nadaraya-Watson estimator*

$$p_n(x,y) = \frac{\sum_{i=1}^{n} K(h^{-1}(x - X_i))K(h^{-1}(y - X_{i+1}))}{\sum_{i=1}^{n} K(h^{-1}(x - X_i))}, \tag{44}$$

computed from a Parzen-Rosenblatt kernel $K(x)$ and a bandwidth $h > 0$.

From a practical perspective, observe that, for most current examples of real-valued chains, any compact interval $V_{x_0}(\varepsilon) = [x_0 - \varepsilon, x_0 + \varepsilon]$ may be easily proved as small, for a properly chosen $x_0 \in \mathbb{R}$ and $\varepsilon > 0$ small enough, with ϕ as the density $\phi_{V_{x_0}(\varepsilon)}$ of the uniform distribution on $V_{x_0}(\varepsilon)$, see section §5.2.2 below. Considering a pre-selected grid $\mathcal{G} = \{(x_0(k), \varepsilon(l)),\ 1 \leq k \leq K,\ 1 \leq l \leq L\}$ such that $\inf_{(x,y)\in V_{x_0}(\varepsilon)^2} p_n(x,y) > 0$ for any $(x_0, \varepsilon) \in \mathcal{G}$, a numerically feasible selection rule could then consist of computing first, for all $(x_0, \varepsilon) \in \mathcal{G}$, the estimated expected number of approximate pseudo-regenerations

$$\widehat{N}_n(x_0, \varepsilon) = \frac{\delta_n(x_0, \varepsilon)}{2\varepsilon} \sum_{i=1}^{n} \frac{\mathbb{I}\{(X_i, X_{i+1}) \in V_{x_0}(\varepsilon)^2\}}{p_n(X_i, X_{i+1})}, \tag{45}$$

with $\delta_n(x_0, \varepsilon) = 2\varepsilon \cdot \inf_{(x,y)\in V_{x_0}(\varepsilon)^2} p_n(x,y)$, and then picking (x_0^*, ε^*) so as to maximize (45) over $\mathcal{G}$. Eventually, one gets the empirical minimizer $S^* = [x_0^* - \varepsilon^*,\ x_0^* + \varepsilon^*]$ and the corresponding minorization constant $\delta_n^* = \delta_n(x_0^*, \varepsilon^*)$. It then remains to construct the approximate pseudo-blocks using S^*, δ_n^* and p_n as described in § 2.2.

5.2.2 Examples - autoregressive models

Given the ubiquity of the markovian assumption in time-series modeling, here we consider the following (possibly nonlinear) autoregressive model

$$X_{n+1} = m(X_n) + \sigma(X_n)\epsilon_{n+1},\ n \in \mathbb{N}, \tag{46}$$

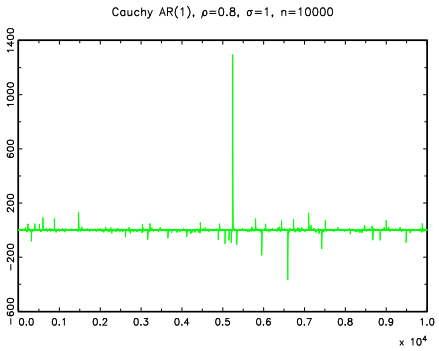
where $m(.)$ (respectively, $\sigma(.)$) is a continuous mapping from $\mathbb{R}$ to $\mathbb{R}$ (resp. to $\mathbb{R}_+^*$) and $(\epsilon_n)_{n\in\mathbb{N}}$ is a sequence of continuous i.i.d. r.v.'s with common density $h(x)$. The transition density with respect to the Lebesgue measure may then be written as

$$\pi(x,y) = \frac{1}{\sigma(x)} h\left(\frac{y - m(x)}{\sigma(x)}\right). \tag{47}$$

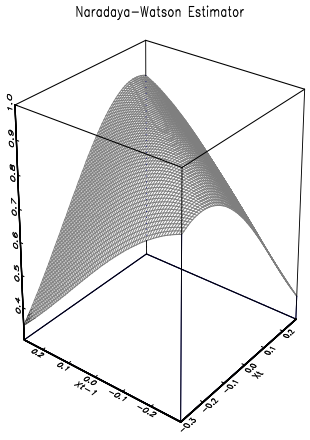
A linear AR(1) model. We start with a basic linear AR(1) model. In this case, we naturally have $m(x) = \rho \cdot x$ and $\sigma(x) \equiv \sigma > 0$.

• *Cauchy noise.* We first assume that the ϵ_n's are drawn from a standard Cauchy distribution, with characteristic function $\mathbb{E}[\exp(it\epsilon_1)] = \exp(-(1 - |\rho|)|t|)$, and we choose $\rho = 0.8$ and $\sigma = 1$. In this case, the extremal index θ of $\{X_n\}_{n\in\mathbb{N}}$ is known to be $1 - \rho$ (since $\rho = 0.8 > 0$, see section 2 in [17]) and the tail index of the stationary distribution is classically the same as the one of the residuals, namely 1 for a Cauchy distribution. It is also easy to see that the chain is geometrically ergodic, by checking a geometric drift condition of Foster-Lyapounov's type, with $V(x) = 1 + \sqrt{|x|}$ as test function, for instance, see §15.2.2 in [47]. Moment assumptions stipulated in Propositions 4 and 5 are thus clearly satisfied.

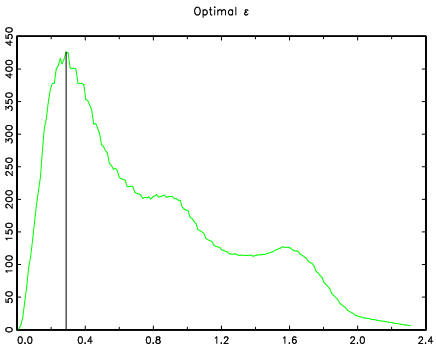
In this first non-atomic example, the steps of the pseudo-block construction described in §5.2.1 are summarized in Figure 4. Figure 4(a) displays a typical sample path of the



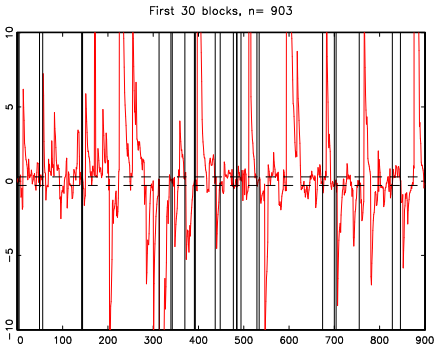
(a) A Cauchy AR(1) sample path



(b) Transition density estimate



(c) Small set optimization, $\varepsilon_{\text{opt}} = 0.287$

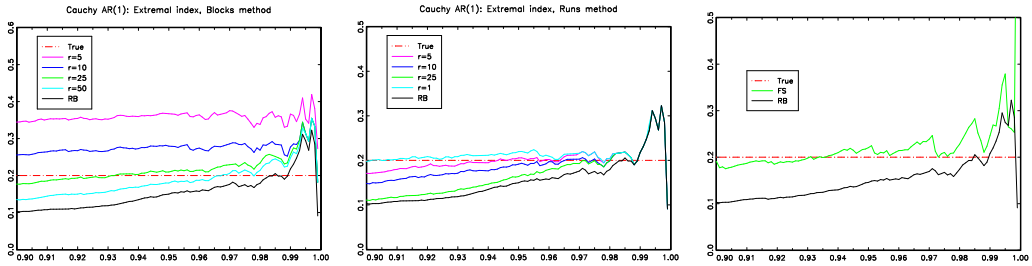


(d) A splitting example

Figure 4: Pseudo-block construction in the Cauchy AR(1) model ($\rho = 0.8, \sigma = 1$)

process with length $n = 10000$ and initial value $X_0 = 0$. Since the X_n's have median zero, we consider small sets candidates of the form $V_0(\epsilon) =]-\varepsilon, +\varepsilon[$. In Figure 4(b), a Nadaraya-Watson estimate (44) π_n of π is plotted, from which the approximate expected number of pseudo-blocks (45) associated to the small set $V_0(\epsilon)$ has been computed. The optimization step, consisting of maximizing the approximate expected number of pseudo-blocks with respect to ε, is described by Figure 4(c), yielding $\epsilon_{\mathbf{opt}} = 0.287$. Then, as explained in §2.2, each time an observation X_i falls into the optimal small set $V_0(\varepsilon_{\mathrm{opt}})$, a Bernoulli r.v. with parameter $\frac{\delta_n(0,\varepsilon_{\mathrm{opt}})}{2\varepsilon_{\mathrm{opt}}} \frac{\mathbb{I}\{(X_i,X_{i+1})\in V_0(\varepsilon_{\mathrm{opt}})^2\}}{p_n(X_i,X_{i+1})}$ is drawn in order to determine whether the path should be split up at time-point i or not. The trajectory is then divided into $\widehat{l}_n$ pseudo-blocks, as illustrated by Figure 4(d), that displays the first 30 pseudo-blocks. In the following, when computing the mean squared errors over M replications of the process, the same values of $\delta_n(0, \varepsilon_{\mathrm{opt}})$ and $\varepsilon_{\mathrm{opt}}$ are used for all replication to obtain a reasonable computing time even though, ideally, they should be computed again for each of them.

As in §5.1, the pseudo-regeneration based estimator (RB estimator, for short) of the extremal index θ is compared to estimators constructed using the *blocks* method, the *runs* method and the *intervals* estimator of [25] referred to as FS *intervals estimator* in the sequel, see the set of Figures 5 for an example (computation on one single chain). In addition to the practical advantage of the block lengths being entirely determined by the RB procedure, our estimator bears the comparison with its competitors, and behaves particularly well for a threshold u greater than the 98^{th} percentile of the X's. This result is confirmed by the computation of the MSE over $M = 500$ replications of the process as illustrated in Figure 6. For high percentiles, the RB estimator performs better than the FS *intervals* estimator and even slightly outperforms the *blocks* and *runs* estimators when the block length r is fixed to values providing the best results on the first simulation, namely $r = 50$ for the *blocks* method, and $r = 10$ for the *runs* method.



(a) RB versus *blocks estimators* (b) RB versus *runs estimators* (c) RB versus FS *intervals estimator*

Figure 5: Extremal index estimation in the Cauchy AR(1) model with $\rho = 0.8$, $\sigma = 1$, $\theta = 0.2$ (Regenerative-based estimator (RB), *Blocks* method ($r = 50$), *Runs* method ($r = 10$), *Intervals* estimator (FS), $n = 10000$).

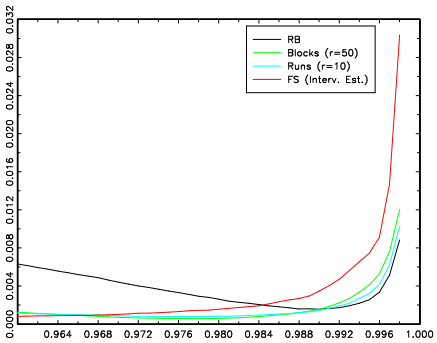


Figure 6: Extremal index estimation - MSE as a function of the threshold level u (Cauchy AR(1) model with $\rho = 0.8$, $\sigma = 1$, $\theta = 0.2$, Regenerative-based estimator (RB), *Blocks* method ($r = 50$), *Runs* method ($r = 10$), *Intervals* estimator (FS), $n = 10000$, $M = 500$).

Turning to tail index estimation, the bias corrected version of the RB Hill estimator is again compared to the standard Hill estimator suggested in [52] (illustration would be similar to Figure 3 and is omitted here). For example, looking at the chain of Figure 4(a), the resulting estimate, 0.915, is slightly closer to 1 than the standard Hill estimate based on the X-values, 0.901, with $k_{\text{opt}} = 51$ for the RB Hill and $k_{\text{opt}} = 783$ for the standard Hill estimate. Again, we include less observations than we would normally based on the standard Hill estimate but we still achieve a slightly better MSE due to the important reduction in the variance of our estimator, see Table 2, first two lines.

• *Pareto noise.* Still considering the problem of estimating the regular variation index, another example is the Pareto AR(1) model: the ϵ_n's are now drawn from a mean centered Pareto distribution with tail index $a = 3$. The parameters of the AR(1) model are here taken as $\rho = 0.9$ and $\sigma = 2$. For instance, on one specific chain, the regeneration-based Hill estimator is 0.321, while the standard Hill estimator takes the value 0.313, with $k_{\text{opt}} = 32$ for the RB Hill estimate and 276 for the standard Hill estimate. Here, the study of the MSE over $M = 500$ replications of the process shows similar variance but a greater bias for our estimator.

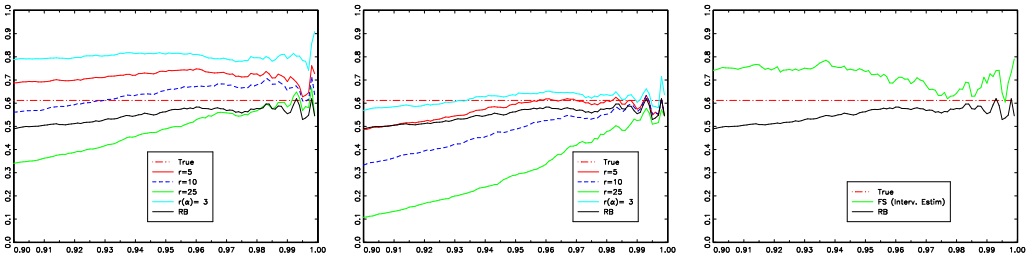
An ARCH(1) model. We now turn to a nonlinear time-series model, namely an ARCH(1) model and thus take $m(x) \equiv 0$, $\sigma^2(x) = \beta_0 + \beta_1 \cdot x^2$ as well as standardized and normally distributed residuals. In our simulation, we chose $\beta_0 = 1$ and $\beta_1 = 0.9$, the extremal index of such process was approximated in [21], see Table 3.2 therein, and is $\theta \approx 0.612$, see also [40]. Notice that the corresponding chain X is geometrically ergodic, see Theorem 1 in [16], and consequently fulfills the moment assumptions required in Propositions 4 and 5. The pseudo regenerative blocks are constructed using exactly the same procedure as for the AR(1) model. The estimation of the extremal index again gives satisfactory results, as illustrated in Figures 7(a)-7(c) on one chain. Note that in this example the mean length of the pseudo blocks is between 3 and 4, which corresponds to

Table 2: Tail index estimation - Comparison of the Mean Squared Errors ($M = 500$ for $n = 10000$, $M = 100$ for $n = 30000$)

Noise	Estimator	Squared Bias	Variance	MSE
Cauchy AR(1)	RB Hill	0.00519	0.0422	0.0474
$n = 10000$	Standard Hill	0.00303	0.0466	0.0496
Pareto AR(1)	RB Hill	0.00252	0.0106	0.0131
$n = 10000$	Standard Hill	0.00026	0.0101	0.0104
ARCH(1)	RB Hill	0.00066	0.0035	0.0042
$n = 10000$	Standard Hill	0.00013	0.0035	0.0036
ARCH(1)	RB Hill	0.00029	0.0020	0.0023
$n = 30000$	Standard Hill	0.00002	0.0018	0.0019

the length at which the *runs* estimator performs best. Figure 8 illustrates the performance of our estimator for $M = 500$ replications of the process. We observe again that the RB estimator performs better than the FS *intervals* estimator for large (greater the 97^{th}) percentiles of the X's and that it also bears the comparison with the *blocks* and *runs* method with respective fixed length of $r = 10$ and $r = 5$.

For the tail index estimation, various simulations were conducted and tend to show that the data size should be increased to obtain satisfactory results. In this ARCH(1) model, the tail index is $a = 2\kappa$, with κ being the solution of $\kappa\beta_1^{\kappa} = 1$ if residuals are standardized and normally distributed, that is $\kappa = 1.152$ when $\beta_1 = 0.9$, see [21]. The MSE of the RB Hill estimator is comparable to that of the standard Hill estimator suggested in [52], see the bottom of Table 2. The difference however slightly favors the standard Hill estimator in terms of bias but this is reduced with increased path length n.



(a) RB versus *blocks* estimators (b) RB versus *runs* estimators (c) RB versus FS *intervals* estimator

Figure 7: Estimation of the extremal index in the ARCH(1) model with $\beta_0 = 1, \beta_1 = 0.9$, $\theta \approx 0.612$ (Regenerative-based estimator (RB), *Blocks* method, *Runs* method, *Intervals* estimator (FS), $n = 10000$).

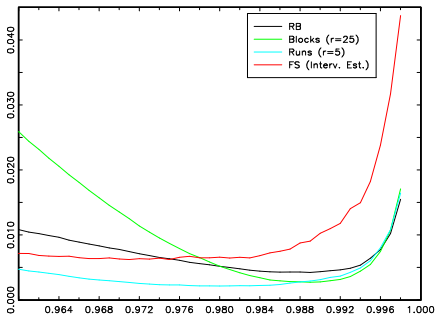


Figure 8: MSE for Extremal index estimation as a function of the threshold level u (ARCH(1) model with $\beta_0 = 1, \beta_1 = 0.9$, $\theta \approx 0.612$, Regenerative-based estimator (RB), *Blocks* method ($r = 25$), *Runs* method ($r = 5$), *Intervals* estimator (FS), $n = 10000$, $M = 500$).

6 Technical proofs

6.1 Proof of Theorem 2

The proof relies on a coupling argument, similar to the one used for proving Theorem 3.1 in [9]. Denote by $\tau_S = \tau_S(1) = \inf\{n \geq 1,\ X_n \in S\}$ and $\tau_S(j) = \inf\{n > \tau_S(j-1),\ X_n \in S\}$, $j \geq 2$, the (random) times of the successive visits to S and by $L_n = \sum_{i=1}^{n} \mathbb{I}\{X_i \in S\}$ the number of visits of X to S between times 1 and n. Following in the footsteps of [9], consider the joint distribution such that, conditioned upon the sample path $X^{(n+1)} = (X_1, ..., X_{\tau_S(1)}, ..., X_{\tau_S(L_n)}, ..., X_{n+1})$, the $(Y_i, \widehat{Y}_i)$'s are drawn independently for $1 \leq i \leq n$ so that

$$\left\{ \begin{array}{rcl} Y_{\tau_S(k)} & \sim & Ber\left(\delta\phi\left(X_{\tau_S(k)+1}\right)/\pi(X_{\tau_S(k)}, X_{\tau_S(k)+1})\right) \\ \widehat{Y}_{\tau_S(k)} & \sim & Ber\left(\delta\phi\left(X_{\tau_S(k)+1}\right)/\hat{\pi}_n(X_{\tau_S(k)}, X_{\tau_S(k)+1})\right) \end{array} \right. ,$$

and if $\pi(X_{\tau_S(k)}, X_{\tau_S(k)+1}) \leq \hat{\pi}_n(X_{\tau_S(k)}, X_{\tau_S(k)+1})$,

$$\begin{aligned} &\mathbb{P}(\widehat{Y}_{\tau_S(k)} = 1,\ Y_{\tau_S(k)} = 0 \mid X^{(n+1)}) = \hat{\pi}_n(X_{\tau_S(k)}, X_{\tau_S(k)+1}) - \pi(X_{\tau_S(k)}, X_{\tau_S(k)+1}), \\ &\mathbb{P}(\widehat{Y}_{\tau_S(k)} = 0,\ Y_{\tau_S(k)} = 1 \mid X^{(n+1)}) = 0, \end{aligned}$$

and if $\pi(X_{\tau_S(k)}, X_{\tau_S(k)+1}) \geq \hat{\pi}_n(X_{\tau_S(k)}, X_{\tau_S(k)+1})$,

$$\begin{aligned} &\mathbb{P}(\widehat{Y}_{\tau_S(k)} = 0,\ Y_{\tau_S(k)} = 1 \mid X^{(n+1)}) = \pi(X_{\tau_S(k)}, X_{\tau_S(k)+1}) - \hat{\pi}_n(X_{\tau_S(k)}, X_{\tau_S(k)+1}), \\ &\mathbb{P}(\widehat{Y}_{\tau_S(k)} = 1,\ Y_{\tau_S(k)} = 0 \mid X^{(n+1)}) = 0, \end{aligned}$$

for $k \in \{1, ...,\ L_n\}$, and that for all $i \in \{1, ...,\ n\} \backslash \{\tau_S(k),\ 1 \leq k \leq L_n\}$, $Y_i = \widehat{Y}_i \sim Ber(\delta)$. As a preliminary, notice first that we thus have

$$\mathbb{P}(\widehat{Y}_{\tau_S(k)} \neq Y_{\tau_S(k)} \mid X^{(n+1)}) = \left| \frac{\delta\phi\left(X_{\tau_S(k)+1}\right)}{\pi(X_{\tau_S(k)}, X_{\tau_S(k)+1})} - \frac{\delta\phi\left(X_{\tau_S(k)+1}\right)}{\hat{\pi}_n(X_{\tau_S(k)}, X_{\tau_S(k)+1})} \right| \ \mathbb{P}_\nu \text{ a.s.},$$

for $1 \leq k \leq L_n$. Therefore, using the fact that $\pi(x,y) \wedge \hat{\pi}_n(x,y) \geq \delta\phi(y) \geq \delta \inf_{z \in S} \phi(z)$ for all $(x,y) \in S^2$, we deduce that $\mathbb{P}_\nu$-almost surely,

$$\mathbb{P}(\widehat{Y}_{\tau_S(k)} \neq Y_{\tau_S(k)} \mid X^{(n+1)}) \leq (\delta \inf_{z \in S} \phi(z))^{-1} \sup_{(x,y) \in S^2} |\hat{\pi}_n(x,y) - \pi(x,y)|. \tag{48}$$

By triangular inequality, we also have for all $x \in \mathbb{R}$,

$$|\widehat{G}_{f,n}(x) - G_f(x)| \ \leq \ |\widehat{G}_{f,n}(x) - G_{f,n}(x)| + |G_{f,n}(x) - G_f(x)|. \tag{49}$$

As previously noticed, the term on the left hand-side of (49) tends to zero $\mathbb{P}_\nu$-almost surely as $n \to \infty$. Therefore, provided that both l_n and $\hat{l}_n$ are larger than 2 (which happens with probability $1 - O(n^{-1})$ under the moment conditions stipulated), the deviation in

the Kolmogorov-Smirnov's sense between the empirical cdf estimate computed from the approximate regeneration cycles and the one computed from the 'true' regeneration cycles may be bounded as follows. We have

$$|\widehat{G}_{f,n}(x) - G_{f,n}(x)| \leq \frac{1}{l_n - 1} \left| \sum_{i=1}^{l_n - 1} \mathbb{I}\{\zeta_j(f) \leq x\} - \sum_{i=1}^{\hat{l}_n - 1} \mathbb{I}\{\hat{\zeta}_j(f) \leq x\} \right| + \frac{|\hat{l}_n - l_n|}{\hat{l}_n - 1}. \tag{50}$$

From Lemma 6.3 in [9] combined with the fact that, by virtue of the SLLN,

$$l_n/n \to \mathbb{E}_{A_S}[\tau_{A_S}]^{-1}, \ \mathbb{P}_\nu - a.s. \text{ as } n \to \infty, \tag{51}$$

it follows that the second term on the left hand-side of (49) is $O_{\mathbb{P}_\nu}(\mathcal{R}_n(\hat{\pi}_n, \pi)^{1/2})$ as n tends to infinity. Furthermore, turning to the first term, we have the bound

$$\left| \sum_{i=1}^{l_n - 1} \mathbb{I}\{\zeta_j(f) \leq x\} - \sum_{i=1}^{\hat{l}_n - 1} \mathbb{I}\{\hat{\zeta}_j(f) \leq x\} \right| \leq \sum_{j=1}^{L_n} \mathbb{I}\{\widehat{Y}_{\tau_S(k)} \neq Y_{\tau_S(k)}\}.$$

From the bound (48), we deduce that its conditional expectation given $X^{(n+1)}$ is thus less than $n \sup_{(x,y)\in S^2} |\hat{\pi}_n(x,y) - \pi(x,y)|$ and, consequently, its $\mathbb{P}_\nu$-expectation is bounded by $n\mathcal{R}_n(\hat{\pi}_n, \pi)^{1/2}$. Clearly, the first term on the left hand-side of (49) is eventually also of order $O_{\mathbb{P}_\nu}(\mathcal{R}_n(\hat{\pi}_n, \pi)^{1/2})$ and thus

$$\sup_{x\in\mathbb{R}} |\widehat{G}_{f,n}(x) - G_{f,n}(x)| = O_{\mathbb{P}_\nu}(\mathcal{R}_n(\hat{\pi}_n, \pi)^{1/2}), \text{ as } n \to \infty. \tag{52}$$

This establishes (15).

6.2 Proof of Proposition 3

First, the convergence (17) follows straightforwardly from Proposition 1. Next, we show that $l(n)(1 - G_{N(n)}(u_n)) \to \eta$ in $\mathbb{P}_\nu$- *pr.* as $n \to \infty$. As $l(n)/n \to \alpha^{-1}$ $\mathbb{P}_\nu$- a.s. as $n \to \infty$ by the SLLN, it thus suffices to prove that

$$n(G(u_n) - G_{N(n)}(u_n)) \to 0 \text{ in } \mathbb{P}_\nu - pr. \text{ as } n \to \infty. \tag{53}$$

Therefore, using the standard LIL in the i.i.d. setup, we immediately get:

$$\sup_{x\in\mathbb{R}} |G(x) - G_{N(n)}(x)| = O_{\mathbb{P}_\nu}(\sqrt{\log\log N(n)/N(n)}) \text{ as } N(n) \to \infty. \tag{54}$$

Since $n^2 = o(N(n)/\log\log N(n))$ as $n \to \infty$, this immediately yields (53) and, consequently, (18).

Now, in the general Harris recurrent setup, using the coupling introduced in §6.1 again, combined with (52), we conclude that $n(\widehat{G}_{N(n)}(u_n) - G_{N(n)}(u_n)) = o_{\mathbb{P}_\nu}(1)$ as $n \to \infty$ since, by assumption, $n\mathcal{R}_{N(n)}(\hat{\pi}_{N(n)}, \pi)^{1/2} = o(1)$. Convergence (19) is thus established.

6.3 Proof of Proposition 4

By assumption, we have $r_n(1 - G(\nu_n))/\alpha \to \eta$ as $n \to \infty$ so that (21) implies that $r_n(1 - F(\nu_n)) \to \eta/\theta$. Now, as (53) holds under our assumptions, it is sufficient to show that $r_n(F_n(\nu_n) - F(\nu_n))$ also tends to zero as $n \to \infty$ for proving assertion (i). This follows from the fact that we have supposed $r_n = o(\sqrt{n/\log\log n})$ combined with the LIL for positive recurrent chains stated in the next lemma.

Lemma 6 *(*LIL FOR FUNCTIONALS OF POSITIVE CHAINS*) Let X be a positive recurrent Markov chain with state space $(E, \mathcal{E})$ and (unique) invariant probability distribution μ, satisfying assumptions $\mathcal{H}(\nu, 1)$ and $\mathcal{H}(2)$. Let $f : E \to \mathbb{R}$ be a measurable function. We have*

$$\limsup_{n\to\infty} \frac{\sup_{u\in\mathbb{R}} |F_n(u) - F(u)|}{\sqrt{(2\sigma_f^2 \log\log n)/n}} = +1 \ \ \mathbb{P}_\nu\text{-almost surely,} \tag{55}$$

with $\sigma_f^2 = \sup_{u\in\mathbb{R}} \sigma_f^2(u)$ and denoting by $\sigma_f^2(u)$, for all $u \in \mathbb{R}$, the limiting variance of $\sqrt{n}(F_n(u) - F(u))$ as $n \to \infty$.

PROOF. The proof requires a refinement of the argument of Theorem 17.5.3 in [47], which establishes a (scalar version of the) LIL for the sequence $(\mathbb{I}\{f(X_n \le u\})_{n\in\mathbb{N}}$ with fixed $u \in \mathbb{R}$. In order to prove an extension of the latter result to a Banach space valued sequence, we classically use the *regenerative method*. We place ourselves in the regenerative framework, where X possesses an accessible atom A. Recall that in such an atomic case, we classically have $\sigma_f^2(u) = \alpha^{-1}\mathbb{E}_A[(\sum_{i=1}^{\tau_A} \mathbb{I}\{f(X_i \le u\} - \alpha F(u))^2]$ (see Eq. (17.13) in [47] for instance) and consider the "block sums"

$$S_j(u) = \sum_{i=1+\tau_A(j)}^{\tau_A(j+1)} \{\mathbb{I}\{f(X_i \le u\} - F(u)\} \ \text{ for } j \ge 1.$$

Notice that, for all $u \in \mathbb{R}$, the $S_j(u)$'s are i.i.d. random variables with mean zero and variance $\alpha\sigma_f^2(u)$ by virtue of the strong Markov property. We decompose the deviation $F_n(u) - F(u)$ as follows:

$$F_n(u) - F(u) = \frac{1}{n}\sum_{i=1}^{\tau_A} \{\mathbb{I}\{f(X_i \le u\} - F(u)\} + \frac{1}{n}\sum_{j=1}^{l_n - 1} S_j(u) + \frac{1}{n}\sum_{i=1+\tau_A(l_n)}^{n} \{\mathbb{I}\{f(X_i \le u\} - F(u)\} \tag{56}$$

The first and last terms in (56) are clearly of order $O_{\mathbb{P}_\nu}(n^{-1})$, under the assumed moment conditions, while the extended version of Kolmogorov's LIL stated in Theorem 8.2 of [43] yields

$$\limsup_{l\to\infty} \frac{\sup_{u\in\mathbb{R}} |\sum_{j=1}^{l} S_j(u)|}{\sqrt{l\alpha\sigma_f^2 \log\log l}} = +1 \ \text{ almost surely.} \tag{57}$$

Given that $\alpha(l_n - 1) \sim n/\alpha$, and thus $\log(\log(l_n - 1)) \sim \log(\log n)$, as $n \to \infty$, (55) follows straightforwardly from (57).

Turning now to the general Harris case, it is sufficient to observe that, under the assumptions of assertion (ii), (52) holds and $n(\hat{F}_{N(n)}(u) - F_{N(n)}(u)) = o_{\mathbb{P}_\nu}(1)$. Indeed, the latter identity immediately follows from Lemma 6.2 in [10], which establishes more generally that

$$N(n) \times \sup_{u \in \mathbb{R}} |\hat{F}_{N(n)}(u) - F_{N(n)}(u)| = O_{\mathbb{P}_\nu}(\mathcal{R}_{N(n)}(\hat{\pi}_{N(n)}, \pi)^{1/2}) \text{ as } N \to \infty. \tag{58}$$

■

6.4 Proof of Proposition 5

Part (i) follows straightforwardly from the classical results available in the i.i.d. setting (see [22] and Corollary 4 in [30]). Details are omitted and we directly turn to the second part of the result. Again, we use the coupling defined in §6.1. In the same fashion as the Hill estimator may be expressed as an L-statistic (see [30] for instance), we write:

$$a_{n,k} = \frac{n}{k} \int_{y=\zeta_{(k)}(f)}^{\infty} \frac{1 - G_{f,n}(y)}{y} dy. \tag{59}$$

Similarly, we have for the approximate version

$$\hat{a}_{n,k} = \frac{n}{k} \int_{y=\hat{\zeta}_{(k)}(f)}^{\infty} \frac{1 - \hat{G}_{f,n}(y)}{y} dy. \tag{60}$$

Now let $1 \leq k \leq \hat{l}_n \wedge l_n - 1$, combining (59) and (60), one immediately gets that

$$\begin{aligned} |a_{n,k} - \hat{a}_{n,k}| &\leq \frac{n}{k} \int_{\zeta_{(k)}(f) \wedge \hat{\zeta}_{(k)}(f)}^{\zeta_{(1)}(f) \vee \hat{\zeta}_{(1)}(f)} \frac{|\hat{G}_{f,n}(y) - G_{f,n}(y)|}{y} dy \\ &\leq \frac{n}{k} \sup_{y \in \mathbb{R}} |\hat{G}_{f,n}(y) - G_{f,n}(y)| \times \log(M_n(f)). \end{aligned} \tag{61}$$

Under our assumptions, recall that $\sup_{y \in \mathbb{R}} |\hat{G}_{f,n}(y) - G_{f,n}(y)| = O_{\mathbb{P}_\nu}(\mathcal{R}_n(\hat{\pi}_n, \pi)^{1/2})$ as $n \to \infty$ (see Eq. (52)). Therefore, given the equivalence (24) and the tail assumption (31), we have $1 - G_f(x) \sim x^{-a}$, as $x \to \infty$. From (9), we deduce that, as $n \to \infty$, $\log(M_n(f))$ goes to infinity at the same rate as $\max_{1 \leq i \leq \lfloor n/\alpha \rfloor} \eta_i$, where $(\eta_n)_{n \in \mathbb{N}}$ is a sequence of exponential i.i.d. r.v.'s with common mean a^{-1}. Since $\max_{1 \leq i \leq \lfloor n/\alpha \rfloor} \eta_i$ is almost surely equivalent to $a^{-1} \log n$ as $n \to \infty$ (see Eq. (3.71) in [24] for instance), and given the choice made for $k = k(n)$, the term on the right hand side of (61) converges to zero in $\mathbb{P}_\nu$-probability. Combined with (i), this eventually establishes (38).

Now, for proving (39), it suffices to observe that, for $1 \leq m \leq n$ and $1 \leq k < \hat{l}_m$,

$$\sqrt{k}(\hat{a}^{(n)}_{k,m} - a) = \sqrt{k}(\hat{a}^{(n)}_{k,m} - a^{(n)}_{k,m}) + \sqrt{k}(a^{(n)}_{k,m} - a). \tag{62}$$

And choosing $m = m_n$ and $k = k(\hat{l}_{m_n})$ such that $m_n \mathcal{R}_n(\hat{\pi}_n, \pi)^{1/2} / \sqrt{k(m_n)} \to 0$ as $n \to \infty$, the argument above obviously shows that the approximation term in (62) degenerates. Now, it suffices to notice that part (i) applies to the term on the right hand side of (62), establishing (39).

References

[1] M.A. Ancona-Navarette and J.A. Tawn. A comparison of methods for estimating the extremal index. *Extremes*, 3(1):5–38, 2000.

[2] S. Asmussen. Extreme value theory for queues via cycle maxima. *Extremes*, 1(2):137–168, 1998.

[3] S. Asmussen. Subexponential asymptotics for stochastic processes: Extremal behavior, stationary distributions and first passage probabilities. *Adv. Appl. Probab.*, 8(2):354–374, 1998.

[4] S. Asmussen. *Applied Probability and Queues*. Springer-Verlag, New York, 2003.

[5] S. Asmussen and C. Klüppelberg. Stationary M/G/1 excursions in the presence of heavy tails. *J. Appl. Probab.*, 34:208–212, 1997.

[6] K.B. Athreya and G.S. Atuncar. Kernel estimation for real-valued Markov chains. *Sankhya*, 60(1):1–17, 1998.

[7] K.B. Athreya and S.G. Pantula. Mixing properties of Harris chains and autoregressive processes. *J. Appl. Prob.*, 23:880–892, 1986.

[8] J. Beirlant, G. Dierckx, Y. Goegebeur, and G. Matthys. Tail index estimation and an exponential regression model. *Extremes*, 2(2):177–200, 1999.

[9] P. Bertail and S. Clémençon. Edgeworth expansions for suitably normalized sample mean statistics of atomic Markov chains. *Prob. Th. Rel. Fields*, 130(3):388–414, 2004.

[10] P. Bertail and S. Clémençon. Regenerative-block bootstrap for Markov chains. *Bernoulli*, 12(4), 2005.

[11] P. Bertail and S. Clémençon. Regeneration-based statistics for Harris recurrent Markov chains. In P. Bertail, P. Doukhan, and P. Soulier, editors, *Probability and Statistics for dependent data*, volume 187 of *Lecture notes in Statistics*, pages 3–54. Springer, 2006.

[12] P. Bertail and S. Clémençon. Sharp bounds for the tails of functionals of Markov chains, 2007. Available at `http://hal.archives-ouvertes.fr/hal-00140591/`.

[13] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular Variation*. Encyclopedia of Mathematics and its applications. Cambridge Univ Press, Cambridge, 1987.

[14] L. Birgé. Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahrsch. Verw. Geb.*, 65:181–237, 1983.

[15] E. Bolthausen. The Berry-Esseen theorem for functionals of discrete Markov chains. *Z. Wahrsch. Verw. Geb.*, 54(1):59–73, 1980.

[16] M. Chen and G. Chen. Geometric ergodicity of nonlinear autoregressive models with changing conditional variances. *The Canadian Journal of Statistics*, 28(3):605–613, 2000.

[17] M.R. Chernick, T. Hsing, and W.P. McCormick. Calculating the extremal index for a class of stationary sequences. *Adv. Appl. Probab.*, 23(4):835–850, 1991.

[18] S. Clémençon. Adaptive estimation of the transition density of a regular Markov chain by wavelet methods. *Math. Meth. Statist.*, 9(4):323–357, 2000.

[19] S. Clémençon. Moment and probability inequalities for sums of bounded additive functionals of a regular Markov chains via the Nummelin splitting technique. *Statistics and Probability Letters*, 55:227–238, 2001.

[20] L. de Haan and S. Resnick. On asymptotic normality of the Hill estimator. *Stochastic Models*, 14:849–867, 1998.

[21] L. de Haan, H. Rootzén, S. Resnick, and C.G. de Vries. Extremal behaviour of solutions to a stochastic differential equation with application to ARCH processes. *Stoch. Proc. Appl.*, 32:213–224, 1989.

[22] P. Deheuvels, E. Häusler, and D.M. Mason. Almost sure convergence of the Hill estimator. *Math. Proc. Cambridge Philos. Soc.*, 104:371–381, 1988.

[23] P. Doukhan and M. Ghindès. Estimation de la transition de probabilité d'une chaîne de Markov Doeblin récurrente. *Stoch. Proc. Appl.*, 15:271–293, 1983.

[24] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Applications of Mathematics. Springer-Verlag, 1997.

[25] C.A.T. Ferro and J. Segers. Inference for clusters of extreme values. *J. R. Statist. Soc.*, 65(2):545–556, 2003.

[26] A. Feuerverger and P. Hall. Estimating a tail exponent by modelling departure from a Pareto Distribution. *Ann. Statist.*, 27:760–781, 1999.

[27] P.W. Glynn and A.J. Zeevi. Estimating tail probabilities in queues via extremal statistics. In D.R. McDonald and S.R. Turner, editors, *Analysis of Communication Networks: Call Centres, Traffic, and Performance*, pages 135–158, Providence, Rhode Island, 2000.

[28] C. M. Goldie and R. L. Smith. Slow variation with remainder: theory and applications. *Quart. J. Math. Oxford*, 38(1):45–71, 1987.

[29] C.M. Goldie. Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Probab.*, 1:126–166, 1991.

[30] E. Haeusler and J. L. Teugels. On asymptotic normality of Hill's estimator for the exponent of regular variation. *Ann. Statist.*, 13(2):743–756, 1985.

[31] G. Haiman, M. Kiki, and M.L. Puri. Extremes of Markov sequences. *Journal of Statistical Planning Inference*, 45:185–201, 1995.

[32] N.R. Hansen and A.T. Jensen. The extremal behaviour over regenerative cycles for Markov additive processes with heavy tails. *Stoch. Proc. Appl.*, 115:579–591, 2005.

[33] B. Hill. A simple approach to inference about the tail of a distribution. *Ann. Statist.*, 3:1163–1174, 1975.

[34] G. Hooghiemstra and L.E. Meester. Computing the extremal index of special Markov chains and queues. *Stoch. Proc. Appl.*, 65:171–185, 1995.

[35] T. Hsing. On the extreme order statistics for a stationary sequence. *Stoch. Proc. Appl.*, 29(1):155–169, 1988.

[36] T. Hsing. On tail estimation using dependent data. *Ann. Statist.*, 19:1547–1569, 1991.

[37] T. Hsing. Extremal index estimation for a weakly dependent stationary sequence. *Ann. Statist.*, 21(4):2043–2071, 1993.

[38] J. Jain and B. Jamison. Contributions to Doeblin's theory of Markov processes. *Z. Wahrsch. Verw. Geb.*, 8:19–40, 1967.

[39] H.A. Karlsen and D. Tjøstheim. Nonparametric estimation in null recurrent time series. *Ann. Statist.*, 29(2):372–416, 2001.

[40] F. Laurini and J.A. Tawn. New estimators for the extremal index and other cluster characteristics. *Extremes*, 6(3):189–211, 2003.

[41] M.R. Leadbetter. Extremes and local dependence in stationary sequences. *Z. Wahrscheinlichkeitsch*, 65:291–306, 1983.

[42] M.R. Leadbetter and H. Rootzén. Extremal theory for stochastic processes. *Ann. Probab.*, 16:431–478, 1988.

[43] M. Ledoux and M. Talagrand. *Probability in Banach spaces: isoperimetry and processes.* Springer-Verlag, 1991.

[44] V.K. Malinovskiĭ. On some asymptotic relations and identities for Harris recurrent Markov chains. In *Statistics and Control of Stochastic Processes*, pages 317–336, 1985.

[45] V.K. Malinovskiĭ. Limit theorems for Harris Markov chains I. *Theory Prob. Appl.*, 31:269–285, 1987.

[46] V.K. Malinovskiĭ. Limit theorems for Harris Markov chains II. *Theory Prob. Appl.*, 34:252–265, 1989.

[47] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability.* Springer-Verlag, 1996.

[48] E. Nummelin. A splitting technique for Harris recurrent chains. *Z. Wahrsch. Verw. Gebiete*, 43:309–318, 1978.

[49] G.L. O'Brien. Extreme values for stationary and Markov sequences. *Ann. Probab.*, 15:281–291, 1987.

[50] R. Perfekt. Extremal behaviour of stationary Markov chains with applications. *Ann. Appl. Probab.*, 4(2):529–548, 1994.

[51] S. Resnick. *Extreme Values, Point Processes and Regular Variation.* Springer-Verlag, New York, 1987.

[52] S. Resnick and C. Stărică. Consistency of Hill's estimator for dependent data. *J. Appl. Probab.*, 32:139–167, 1995.

[53] S. Resnick and C. Stărică. Tail index estimation for dependent data. *Ann. Appl. Probab.*, 8:1156–1183, 1998.

[54] D. Revuz. *Markov Chains.* 2nd edition, North-Holland, 1984.

[55] G.O. Roberts, J.S. Rosenthal, J. Segers, and B. Sousa. Extremal indices, geometric ergodicity of Markov chains, and MCMC. *Extremes*, 9:213–229, 2006.

[56] H. Rootzén. Maxima and exceedances of stationar Markov chains. *Adv. Appl. Probab.*, 20:371–390, 1988.

[57] H. Rootzén. Weak convergence of the tail empirical process for dependent sequences, 2006. Available at `http://www.math.chalmers.se/~rootzen/papers/tail_empirical060816.pdf`.

[58] R.L. Smith and I. Weissman. Estimating the extremal index. *J. R. Statist. Soc.*, 56:515–528, 1994.

[59] W. L. Smith. Regenerative stochastic processes. *Proc. Royal Stat. Soc.*, 232:6–31, 1955.

[60] H. Thorisson. *Coupling Stationarity and Regeneration.* Probability and its applications. Springer, 2000.

[61] I. Weissman and U. Cohen. The extremal index and clustering of high values for derived stationary sequences. *J. Appl. Probab.*, 32(4):972–981, 1995.

[62] A.J. Zeevi and P.W. Glynn. Estimating tail decay for stationary sequences via extreme values. *Adv. Appl. Probab.*, 36(1):198–226, 2004.