DOI 10.1007/s10958-023-06533-8
Journal of Mathematical Sciences, Vol. 273, No. 5, July, 2023

SOME VARIATIONS ON THE EXTREMAL INDEX

G. Buriticá,[*] N. Meyer,[†] T. Mikosch,[‡] and O. Wintenberger[**] UDC 519.2

We reconsider Leadbetter's extremal index for stationary sequences. It has interpretation as reciprocal of the expected size of an extremal cluster above high thresholds. We focus on heavy-tailed time series, in particular, on regularly varying stationary sequences, and discuss recent research in extreme value theory for these models. A regularly varying time series has multivariate regularly varying finite-dimensional distributions. Thanks to results by Basrak and Segers (2009), we have explicit representations of the limiting cluster structure of extremes, leading to explicit expressions of the limiting point process of exceedances and the extremal index as a summary measure of extremal clustering. The extremal index appears in various situations, which do not seem to be directly related, such as the convergence of maxima and point processes. We consider different representations of the extremal index which arise from the considered context. We discuss the theory and apply it to a regularly varying $AR(1)$ process and the solution to an affine stochastic recurrence equation. Bibliography: 38 titles.

Some personal words by Thomas Mikosch. During my PhD studies at the University of Leningrad in 1981–1984 I met Yasha Nikitin at seminars and workshops at LOMI (now POMI) and the University. I remember him as a professor who had a lot of humor and a balanced personality. Later, since the 1990s, Yasha Nikitin became a representative of Russian Probability Theory and Mathematical Statistics with a high international reputation. I appreciated his cosmopolitan attitude. Whenever one needed constructive advice as regards some international scientific event (such as the European Meeting of Statisticians) or editorial issues, he would help. He actively supported the Bernoulli Society as an organization which embraces all European probabilists and statisticians.

I met Yasha at the Vilnius Conference in 2018 at the last time and, as always, I enjoyed his warm-hearted personality. He went from us too early. His achievements for Probability Theory and Statistics at the University of St.Petersburg and worldwide remain alive.

1. Leadbetter's approach to modeling the extremes of a stationary sequence

The paper by Leadbetter [22] and the book of Leadbetter, Lindgren, and Rootzén [23] provided the first systematic approach to the extreme value theory of dependent stationary sequences. In particular, Leadbetter introduced mixing and anti-clustering conditions, the conditions D and D', which are tailored for the analysis of dependent extremal events. Moreover, [23] propagated the use of the *extremal index* as a measure for extremal clustering.

The idea of an extremal index originates from [24, 25, 27] who discovered that the maxima

$$M_n = \max_{t=1,\dots,n} X_t, \qquad n \ge 1,$$

of numerous examples of dependent stationary sequences (X_t) with common distribution F share the property that

$$\mathbb{P}(M_n \le u_n) \approx \big[\mathbb{P}(X \le u_n)\big]^{n\,\theta_X} = \big((F(u_n))^n\big)^{\theta_X}, \qquad n \to \infty,$$

[*]Research Center for Statistics, University of Geneva, Switzerland, e-mail: gloria.buriticaborda@unige.ch.

[†]Université Montpellier, CNRS, IMAG, Inria, France, e-mail: nicolas.meyer@umontpellier.fr.

[‡]Department of Mathematics, University of Copenhagen, Denmark, e-mail: mikosch@math.ku.dk.

[**]LPSM, Sorbonne Universités UPMC Université Paris 06, France, e-mail: olivier.wintenberger@upmc.fr.

Published in *Zapiski Nauchnykh Seminarov POMI*, Vol. 501, 2021, pp. 52–77. Original article submitted May 27, 2021.

for some number $\theta_X \in [0,1]$, provided (u_n) is a sequence of high thresholds converging sufficiently fast to the right endpoint x_F of F. Leadbetter [22] made this notion precise as the *expected size of an extremal cluster of exceedances above high-level thresholds.* Since $(F(u_n))^n$ is the distribution function of the maximum of n iid random variables with common distribution F at the threshold u_n, the quantity θ_X describes the shrinking effect that the appearance of dependent extremes can have on the distribution of M_n compared to $(F(u_n))^n$.

Leadbetter defined the extremal index θ_X as follows: assume that for every $\tau \in (0,\infty)$ there exists a sequence $(u_n(\tau))$ such that

$$n\,\overline{F}(u_n(\tau)) = n\,(1 - F(u_n(\tau))) \to \tau,$$

and there exists a number θ_X such that

$$\mathbb{P}(M_n \le u_n(\tau)) \to \mathrm{e}^{-\tau\,\theta_X}, \qquad n \to \infty.$$

If such a number θ_X exists, it belongs to the interval $[0,1]$ and is independent of the choice of the sequences (u_n).

An immediate application is to the convergence in distribution of the sequence (M_n). Assume that (X_t) belongs to the maximum domain of attraction of an extreme value distribution H, i.e., for iid copies $(\widetilde{X}_t)$ of X_1, $\widetilde{M}_n = \max(\widetilde{X}_1,\dots,\widetilde{X}_n)$, there exist constants $c_n > 0$, $d_n \in \mathbb{R}$ such that $c_n^{-1}(\widetilde{M}_n - d_n) \stackrel{d}{\to} \xi$ as $n \to \infty$ and ξ has distribution H. Then if (X_t) has an extremal index θ_X, we have

$$n\,\overline{F}(\underbrace{c_n\,x + d_n}_{=:u_n(\tau)}) \to \underbrace{-\log H(x)}_{=:\tau}, \qquad n \to \infty, \qquad x \in \mathrm{supp} H,$$

and

$$\mathbb{P}\big(c_n^{-1}(M_n - d_n) \le x\big) \to H^{\theta_X}(x)\,, \qquad n \to \infty\,, \qquad x \in \mathrm{supp} H.$$

In the case of an iid sequence, it is easily seen that $n\,\overline{F}(u_n(\tau)) \to \tau$ holds if and only if $\mathbb{P}(M_n \le u_n(\tau)) \to \mathrm{e}^{-\tau}$. Hence, $\theta_X = 1$. The extremal index 1 is not exclusive to iid sequences. Indeed, in [23], various examples of strictly stationary sequences are considered, for which $\theta_X = 1$. For example, if (X_t) is a Gaussian stationary sequence whose autocovariance function satisfies $\mathrm{cov}(X_0, X_h) = o(1/\log h)$ as $h \to \infty$, then $\theta_X = 1$.

2. Sufficient conditions for the existence of the extremal index

The extremal index is often interpreted as *the reciprocal of the expected size of an extremal cluster* for a stationary sequence (X_t). We give a justification for this interpretation.

2.1. The method of block maxima. The key is the definition of an *extremal cluster in the sample* $X_1,\dots,X_n$: split the sample into $k_n = [n/r_n]$ blocks of equal length r_n:

$$\underbrace{X_1,\dots,X_{r_n}}_{\text{Block 1}},\underbrace{X_{r_n+1},\dots,X_{2\,r_n}}_{\text{Block 2}},\dots,\underbrace{X_{(k_n-1)\,r_n+1},\dots,X_{k_n\,r_n}}_{\text{Block } k_n},$$

we ignore the last block of length less than r_n, and we simply call the block an *extremal cluster* relative to a high threshold $u = u_n$ (this means that $u_n \uparrow x_F$ as $n \to \infty$) if there is at least one exceedance of this threshold in this block. For an asymptotic theory, it is important that $r = r_n \to \infty$ such that r_n is small compared to n, i.e., $k_n \to \infty$.

In view of the stationarity of (X_t), the *expected cluster size of a block* is given by

$$\mathbb{E}\Big[\sum_{t=1}^{r_n} \mathbb{1}(X_t > u_n) \,\Big|\, M_{r_n} > u_n\Big] = \sum_{t=1}^{r_n} \frac{\mathbb{P}(X_t > u_n\,, M_{r_n} > u_n)}{\mathbb{P}(M_{r_n} > u_n)}$$
$$= \sum_{t=1}^{r_n} \frac{\mathbb{P}(X_t > u_n)}{\mathbb{P}(M_{r_n} > u_n)}$$
$$= \frac{r_n\,\mathbb{P}(X > u_n)}{\mathbb{P}(M_{r_n} > u_n)} =: \frac{1}{\theta_n}.$$

Obviously, θ_n is a number in $[0,1]$. Under mild regularity conditions, the limit $\theta = \lim_{n\to\infty} \theta_n$ exists, assumes values in $[0,1]$ and coincides with Leadbetter's extremal index θ_X (see Theorem 2.1 below). For this reason, the extremal index θ_X is often referred to as *the reciprocal of the expected extremal cluster size above high thresholds.*

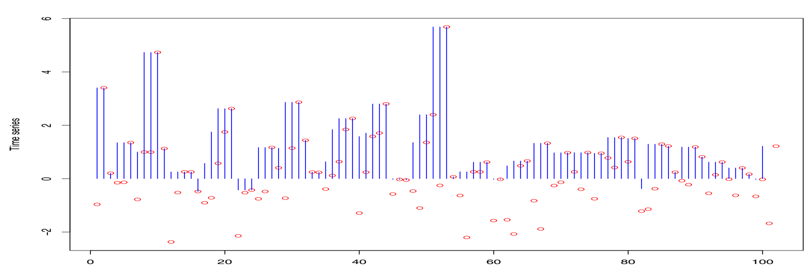

Fig. 2.1. Visualization of the max-moving average

$$X_t = \max(Z_t, Z_{t+1}, Z_{t+2}), t = 1, \dots, 100,$$

(blue) for iid student noise Z_t, $t = 1, \dots, 102$, with $\alpha = 4$ degrees of freedom (red dots). The values of X_t typically appear in clusters of size 3. The process $(|X_t|)$ has extremal index $\theta_{|X|} = 1/3$.

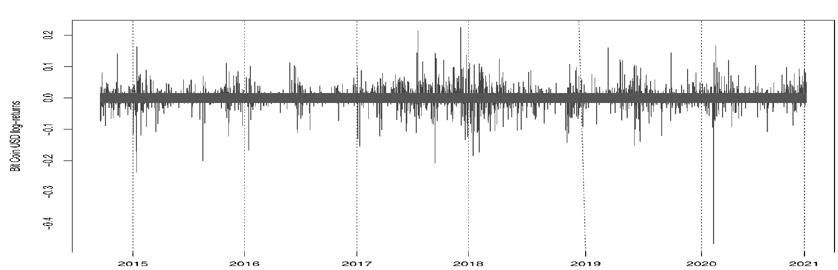

Fig. 2.2. The daily log-return series of the Bit Coin USD stock prices from September 17, 2014 until January 8, 2021. We only show the returns below -0.04 or above 0.04 which we interpret as extreme values. These limits roughly correspond to the 10% and 90% quantiles of the data. The extremes typically appear in clusters.

2.2. Approximation of θ_X by θ_n. The following result can be found in slightly different forms in [9], proof of Lemma 2.8 and [2, 34].

Theorem 2.1. *Consider the following conditions:*

(1) (X_t) *is a real-valued stationary sequence whose marginal distribution* F *does not have an atom at the right endpoint* x_F.

(2) *For a sequence* $u_n \uparrow x_F$ *and an integer sequence* $r = r_n \to \infty$ *such that* $k_n = [n/r_n] \to \infty$, *the following* anti-clustering condition *is satisfied*:

$$\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(M_{k,r_n} > u_n \mid X_0 > u_n\big) = 0. \tag{2.1}$$

Here $M_{a,b} = \max_{i=a,\ldots,b} X_i$ *for* $a \le b$ *such that* $M_b = M_{a,b}$ *with* $a = 1$.

(3) *A* mixing condition *holds:*

$$\mathbb{P}(M_n \le u_n) - \big(\mathbb{P}(M_{r_n} \le u_n)\big)^{k_n} \to 0, \qquad n \to \infty, \tag{2.2}$$

where (u_n), (k_n) *and* (r_n) *are as in* (2).

(4) *For all positive* τ *there exists a sequence* $(u_n) = (u_n(\tau))$ *such that* $n\,\overline{F}(u_n) \to \tau$ *and* (2), (3) *are satisfied for these sequences* (u_n).

Then the following assertions hold:

1. *If* (1) *and* (2) *are satisfied, then*

$$\lim_{k\to\infty} \limsup_{n\to\infty} \big|\theta_n - \mathbb{P}\big(M_k \le u_n \mid X_0 > u_n\big)\big| = 0, \tag{2.3}$$

and $\liminf_{n\to\infty} \theta_n > 0$.

2. *If* (1) *and* (4) *are satisfied and* $\theta = \lim_{n\to\infty} \theta_n$ *exists, then* $\theta_X \in (0, 1]$ *exists and coincides with* θ.

Condition (2.2) is satisfied for strongly mixing (X_t) with mixing rate (α_h) if one can find integer sequences (ℓ_n) and (r_n) such that $\ell_n/r_n \to 0$, $r_n/n \to 0$ and $k_n\alpha_{\ell_n} \to 0$ as $n \to \infty$. Anti-clustering conditions are common in extreme value theory, since Leadbetter introduced the D' condition which is much stronger than (2.1) but is also easily verified on examples. The goal of such a condition is to avoid that the stationary sequence stays above a high threshold for too long.

Relation (2.3) is in agreement with O 'Brien's [28] characterization of the extremal index of (X_t) as the limit

$$\theta_X = \lim_{n\to\infty} \mathbb{P}(M_{\ell_n} \le u_n \mid X_0 > u_n) \tag{2.4}$$

for a sequence (ℓ_n) with $\ell_n/n \to 0$, thresholds $u_n \uparrow x_F$ such that $n\overline{F}(u_n) \to 1$ as $n \to \infty$, provided a mixing condition holds. O'Brien's condition (2.4) has the advantage of avoiding the definition of an extremal cluster.

Remark 2.2. Relation (2.3) provides a constructive way of calculating θ_X: if we know that the limits $f(k) := \lim_{n\to\infty} \mathbb{P}\big(M_k \le u_n \mid X_0 > u_n\big)$ exist for every $k \ge 1$, then we can try to derive $\theta_X = \lim_{k\to\infty} f(k)$. In Sec. 3, we follow this approach in the case of a *regularly varying sequence.*

3.1. Definition and examples. As a matter of fact, clusters of extremes are more prominent in stationary sequences with heavy-tailed marginal distribution. To illustrate this fact, consider a stationary causal AR(1) process, which solves the difference equation $X_t = \varphi X_{t-1}+Z_t$, $t \in \mathbb{Z}$, for an iid noise sequence (Z_t). Necessarily, $\varphi \in (-1,1)$ and, if (Z_t) is iid standard normal, then $(|X_t|)$ has extremal index $\theta_{|X|} = 1$ (see [23]), while for iid student noise (Z_t) with α degrees of freedom, we have $\theta_{|X|} = 1 - |\varphi|^\alpha$ (see Example 3.4 below). Thus, the smaller α (the heavier the tail) for given φ the closer $\theta_{|X|}$ to zero.

An AR(1) process with student noise is an example of a *regularly varying time series*. This class of heavy-tailed processes has been studied rather extensively in the last 15 years; see [31] for some basics about multivariate regular variation, and [21] for a recent textbook treatment. This class was considered in full generality first by [9]: they required that the finite-dimensional distributions of the process satisfy a multivariate regular variation condition (see [30, 31] for the definition of this notion). It is an extension of power-law tail behavior from the univariate to the multivariate case defined via the vague convergence of tail measures with infinite limit measures which have the homogeneity property.

Here we follow an alternative approach by [2] tailored for stationary sequences, avoiding the vague convergence concept. They proved that a real-valued stationary sequence (X_t) is regularly varying with index $\alpha > 0$ in the sense of [9] if and only if there exists a sequence (Θ_t) and a Pareto(α) distributed Y, i.e., $\mathbb{P}(Y > y) = y^{-\alpha}$, $y > 1$, such that (Θ_t) and Y are independent and, for all $h \geq 0$,

$$\mathbb{P}\big(x^{-1}(X_t)_{|t|\leq h} \in \cdot \,\big|\, |X_0| > x\big) \xrightarrow{w} \mathbb{P}\big(Y\,(\Theta_t)_{|t|\leq h} \in \cdot\big), \quad x \to \infty.$$

In the latter relation, x can be replaced by any sequence (a_n) such that $n\,\mathbb{P}(|X| > a_n) \to 1$ as $n \to \infty$. Moreover, by definition, $|\Theta_0| = 1$ a.s. The sequence (Θ_t) is the *spectral tail process* of the regularly varying process (X_t); it describes the propagation of a value $|X_0| > x$ for large x through the stationary sequence (X_t) into its past and future.

Example 3.1. We consider a stationary AR(1) process given as the causal solution to the difference equation $X_t = \varphi X_{t-1} + Z_t$, $t \in \mathbb{Z}$, where (Z_t) is iid regularly varying with index α (e.g., Pareto(α) or student(α)). This means that a generic element Z satisfies $\lim_{x\to\infty} \mathbb{P}(\pm Z > x)/\mathbb{P}(|Z| > x) = p_\pm$ for nonnegative values $p_\pm$ such that $p_+ + p_- = 1$, and $\mathbb{P}(|Z| > x) = L(x)x^{-\alpha}$, $x > 0$, for some slowly varying function L. Then a generic element X inherits the regularly varying tail behavior from Z (see [10]):

$$\frac{\mathbb{P}(\pm X > x)}{\mathbb{P}(|Z| > x)} \sim \sum_{j=0}^{\infty} \big[p_\pm\,(\varphi^j)_\pm^\alpha + p_\mp\,(\varphi^j)_\mp^\alpha\big] = \mathbb{P}(\Theta_0 = \pm 1)(1 - |\varphi|^\alpha).$$

But even more is true: (X_t) is a regularly varying time series with spectral tail process

$$\Theta_t = \Theta_Z\,\mathrm{sgn}(\varphi^{J+t})\,|\varphi|^t\,\mathbb{1}(J + t \geq 0) = \Theta_0\,\varphi^t\,\mathbb{1}(J + t \geq 0), \quad t \in \mathbb{Z}, \tag{3.1}$$

where $\mathbb{P}(\Theta_Z = \pm 1) = p_\pm$, Θ_Z is independent of J which has distribution

$$\mathbb{P}(J = j) = (1 - |\varphi|^\alpha)\,|\varphi|^{j\,\alpha}, \quad j \geq 0.$$

In particular, the forward spectral tail process is given by $\Theta_t = \Theta_0\,\varphi^t$, $t \geq 0$.

Example 3.2. We consider the unique causal solution to the affine stochastic recurrence equation $X_t = A_t\,X_{t-1} + B_t$, $t \in \mathbb{Z}$, for an iid sequence $((A_t, B_t))_{t\in\mathbb{Z}}$ with generic element $(A, B) \in \mathbb{R}_+^2$. We assume that the distribution of (A, B) satisfies the conditions of the Kesten–Goldie theory (see [13, 20], cf. [6] for a textbook treatment). The most important condition in this context is the existence of a unique solution $\alpha > 0$ to the equation $\mathbb{E}[A^\alpha] = 1$ which yields the tail index α. Under these conditions, for a generic element X, there exists a positive constant c_+ such that

$$\mathbb{P}(X > x) \sim c_+ x^{-\alpha}, \quad x \to \infty.$$

Then the forward spectral process is given by

$$(\Theta_t)_{t\geq 0} = (\Pi_t)_{t\geq 0}, \text{ where } \Pi_t = A_1 \cdots A_t,$$

while the backward spectral tail process $(\Theta_t)_{t\leq -1}$ has a rather complicated structure.

Writing $S_t = \log \Pi_t = \sum_{i=1}^{t} \log A_i$, $t \geq 1$, we observe that (S_t) constitutes a random walk with a negative drift. Indeed, by Jensen's inequality, we have $\mathbb{E}[\log(A^\alpha)] < \log(\mathbb{E}[A^\alpha]) = 0$.

3.2. The extremal index. Following Remark 2.2, we derive the extremal index θ_X of a stationary nonnegative regularly varying sequence (X_t) in terms of its spectral tail process. First, we observe that by virtue of the continuous mapping theorem, as $n \to \infty$, for $k \geq 1$,

$$\begin{aligned}\mathbb{P}\big(a_n^{-1} M_k \leq 1 \,\big|\, X_0 > a_n\big) &\to \mathbb{P}\big(Y \max_{1\leq t\leq k} \Theta_t \leq 1\big) = \mathbb{P}\big(\max_{1\leq t\leq k} \Theta_t^\alpha \leq Y^{-\alpha}\big)\\ &= \mathbb{E}\big[\big(1 - \max_{1\leq t\leq k} \Theta_t^\alpha\big)_+\big] = \mathbb{E}\big[\max_{0\leq t\leq k} \Theta_t^\alpha - \max_{1\leq t\leq k} \Theta_t^\alpha\big].\end{aligned}$$

Here we used the fact that $Y^{-\alpha}$ is $U(0,1)$ uniformly distributed and $\Theta_0 = 1$ a.s. Using dominated convergence as $k \to \infty$, we proved under the anti-clustering condition (2.1) that

$$\begin{aligned}\lim_{n\to\infty} \theta_n = \lim_{k\to\infty} \lim_{n\to\infty} \mathbb{P}\big(a_n^{-1} M_k \leq 1 \big| X_0 > a_n\big) &= \lim_{k\to\infty} \mathbb{E}\big[\max_{0\leq t\leq k} \Theta_t^\alpha - \max_{1\leq t\leq k} \Theta_t^\alpha\big]\\ &= \mathbb{E}\big[\big(1 - \max_{t\geq 1} \Theta_t^\alpha\big)_+\big].\end{aligned}$$

From Theorem 2.1 we obtain the following result in [2].

Corollary 3.3. *Consider a nonnegative stationary regularly varying process (X_t) with index $\alpha > 0$. Then the following statements hold:*

1. *If the anti-clustering condition (2.1) holds for $(u_n) = (x\, a_n)$ and some $x > 0$, then the limit $\theta = \lim_{n\to\infty} \theta_n$ exists, is positive and has the representations*

$$\theta = \mathbb{P}\big(Y \sup_{t\geq 1} \Theta_t \leq 1\big) = \mathbb{E}\big[\big(1 - \sup_{t\geq 1} \Theta_t^\alpha\big)_+\big] = \mathbb{E}\big[\sup_{t\geq 0} \Theta_t^\alpha - \sup_{t\geq 1} \Theta_t^\alpha\big]. \tag{3.2}$$

2. *If (2.1) and the mixing condition (2.2) hold for $(u_n) = (x\, a_n)$ and all $x > 0$, then the extremal index θ_X exists and coincides with θ.*

The representations of θ given in (3.2) only depend on the forward spectral process $(\Theta_t)_{t\geq 0}$. In Proposition 3.10 below, we provide representations of the extremal index $\theta_{|X|}$ depending on the whole spectral tail process $(\Theta_t)_{t\in\mathbb{Z}}$.

Example 3.4. We consider the regularly varying AR(1) process from Example 3.1. It can be shown to satisfy the anti-clustering and mixing conditions of Theorem 2.1. We conclude from Corollary 3.3 and the form of the spectral tail process given in (3.1) that

$$\theta_{|X|} = \mathbb{E}\big[\big(1 - \max_{t\geq 1} \Theta_t^\alpha\big)_+\big] = 1 - \max_{t\geq 1} |\varphi|^{\alpha\, t} = 1 - |\varphi|^\alpha.$$

This formula was already achieved in [10] in a wider context of linear processes.

Example 3.5. We consider the regularly varying solution of an affine stochastic recurrence equation under the conditions and with the notation of Example 3.2. It can be shown to satisfy the anti-clustering and mixing conditions of Theorem 2.1 (see [6]). We conclude from this result that (X_t) has extremal index

$$\theta_X = \mathbb{E}\big[\big(1 - \max_{t\ge 1} \Pi_t\big)_+\big] = \mathbb{E}\big[\big(1 - \exp\big(\max_{t\ge 1} S_t\big)\big)_+\big],$$

where $S_t = \sum_{i=1}^{t} \log A_i$, $t \ge 1$, is a random walk with a negative drift. This value of θ_X was derived in [16]. In that paper, a Monte Carlo simulation procedure for the evaluation of θ_X was proposed. Direct calculation of θ_X is difficult (see Example 3.12 for an exception).

3.3. The extremal index and point process convergence toward a cluster Poisson process

3.3.1. *A useful auxiliary result.*

Lemma 3.6. *Consider a nonnegative stationary regularly varying sequence (X_t) with index $\alpha > 0$ and assume that (2.1) holds for $(u_n) = (x\, a_n)$ and all $x > 0$. Then*

$$\|\Theta\|_\alpha^\alpha := \sum_{j\in\mathbb{Z}} \Theta_j^\alpha < \infty \qquad \text{a.s.}$$

In particular, $\Theta_t \to 0$ a.s. as $|t| \to \infty$, and the time T^* of the largest record of (Θ_t) is finite, i.e., $|T^*|$ is the smallest integer such that

$$\Theta_{T^*} = \max_{t\in\mathbb{Z}} \Theta_t\,.$$

Proof. Write $(Y_t) = Y\,(\Theta_t)$, where the Pareto(α) variable Y and the spectral tail process (Θ_t) are independent. We start by showing

$$Y_t \stackrel{\text{a.s.}}{\to} 0, \quad t \to \infty. \tag{3.3}$$

Since (X_t) is regularly varying, we have for all $x > 0$ and integers $k \ge 1$,

$$\lim_{h\to\infty}\lim_{n\to\infty} \mathbb{P}\big(M_{k,k+h} > x\, a_n \;\mid\; X_0 > a_n\big) = \lim_{h\to\infty} \mathbb{P}\big(\max_{k\le t\le k+h} Y_t > x\big) = \mathbb{P}\big(\max_{t\ge k} Y_t > x\big).$$

On the other hand, using the anti-clustering condition (2.1) for all $x \in (0,1]$, we have for fixed $k, h \ge 1$,

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}\big(M_{k,k+h} > x\, a_n \;\mid\; X_0 > a_n\big) &\le \limsup_{n\to\infty} \mathbb{P}\big(M_{k,r_n} > x\, a_n \;\mid\; X_0 > x\, a_n\big) \frac{\mathbb{P}(X > x\, a_n)}{\mathbb{P}(X > a_n)}\\ &= x^{-\alpha} \limsup_{n\to\infty} \mathbb{P}\big(M_{k,r_n} > x\, a_n \;\mid\; X_0 > x\, a_n\big) = x^{-\alpha}\varepsilon_k,\end{aligned}$$

and the right-hand side term ε_k vanishes for large k. Hence, letting $h \to \infty$, we obtain for all $x > 0$,

$$\mathbb{P}\big(\max_{t\ge k} Y_t > x\big) \le x^{-\alpha}\varepsilon_k,$$

and, therefore,

$$\lim_{k\to\infty} \mathbb{P}\big(\max_{t\ge k} Y_t > x\big) \le \lim_{k\to\infty} x^{-\alpha}\varepsilon_k = 0,$$

implying $\max_{t\ge k} Y_t \stackrel{\mathbb{P}}{\to} 0$ as $k \to \infty$. Since $(Y_t) = Y\,(\Theta_t)$ a.s. and $Y > 0$ is independent of (Θ_t), this is only possible if $\max_{t\ge k} \Theta_t \stackrel{\mathbb{P}}{\to} 0$ as $k \to \infty$ but the latter relation is equivalent to $\Theta_t \stackrel{\text{a.s.}}{\to} 0$ as $t \to \infty$, implying (3.3).

Next we show that

$$Y_{-t} \stackrel{\text{a.s.}}{\to} 0, \qquad t \to \infty.$$

Since $Y_t \overset{\text{a.s.}}{\to} 0$ as $t \to \infty$ and $Y_0 > 1$ a.s., the following relation holds

$$\mathbb{P}\Big(\bigcup_{i\ge 0}\{Y_i \ge 1 > \max_{t>i} Y_t\}\Big) = \sum_{i\ge 0} \mathbb{P}\big(Y_i \ge 1 > \max_{t>i} Y_t\big) = 1.$$

Assume that $\mathbb{P}(\sum_{j\le 0} \mathbb{1}(Y_j > \varepsilon) = \infty) > 0$ for some $\varepsilon > 0$. Then there exists some $i \ge 0$ such that

$$\mathbb{P}\Big(\sum_{j\le 0} \mathbb{1}(Y_j > \varepsilon) = \infty, Y_i \ge 1 > \max_{t>i} Y_t\Big) > 0.$$

We recall the time-change formula from [2]:

$$\mathbb{P}((\Theta_{-h},\dots,\Theta_h) \in \cdot \mid \Theta_{-t} \neq 0) = \mathbb{E}\Big[\frac{\Theta_t^\alpha}{\mathbb{E}[\Theta_t^\alpha]} \mathbb{1}\Big(\frac{(\Theta_{t-h},\dots,\Theta_{t+h})}{\Theta_t} \in \cdot\Big)\Big]. \tag{3.4}$$

In particular, $\mathbb{P}(\Theta_t \neq 0) = \mathbb{E}[\Theta_t^\alpha] = 1$ if and only if for all $h \ge 0$,

$$\mathbb{P}((\Theta_{-h},\dots,\Theta_h) \in \cdot) = \mathbb{E}\Big[\frac{\Theta_t^\alpha}{\mathbb{E}[\Theta_t^\alpha]} \mathbb{1}\Big(\frac{(\Theta_{t-h},\dots,\Theta_{t+h})}{\Theta_t} \in \cdot\Big)\Big].$$

Therefore,

$$\begin{aligned}
\infty &= \mathbb{E}\Big[\sum_{j\le 0} \mathbb{1}(Y_j > \varepsilon)\, \mathbb{1}\big(Y_i \ge 1 > \max_{t>i} Y_t\big)\Big] = \sum_{j\le 0} \mathbb{P}\big(Y_j > \varepsilon, Y_i \ge 1 > \max_{t>i} Y_t\big) \\
&= \sum_{j\le 0} \int_1^\infty \mathbb{E}\big[\mathbb{1}\big(y\,\Theta_j > \varepsilon\,, y\,\Theta_i \ge 1 > y \max_{t>i} \Theta_t\big)\big] d\big(-y^{-\alpha}\big) \\
&= \sum_{j\le 0} \int_1^\infty \mathbb{E}\Big[\Theta_{-j}^\alpha\, \mathbb{1}\Big(y > \varepsilon\,\Theta_{-j}, y\,\frac{\Theta_{i-j}}{\Theta_{-j}} \ge 1 > y \max_{t>i-j} \frac{\Theta_t}{\Theta_{-j}}\Big)\Big] d\big(-y^{-\alpha}\big) \\
&\le \varepsilon^{-\alpha} \sum_{j\le 0} \mathbb{E}\Big[\int_1^\infty \mathbb{1}\big(z > 1\,, z\,\Theta_{i-j} \ge \varepsilon^{-1} > z \max_{t>i-j} \Theta_t\big) d\big(-z^{-\alpha}\big)\Big] \\
&= \varepsilon^{-\alpha} \sum_{j\le 0} \mathbb{P}\big(Y_{i-j} \ge \varepsilon^{-1} > \max_{t>i-j} Y_t\big) = \varepsilon^{-\alpha} \sum_{k\ge i} \mathbb{P}\big(Y_k \ge \varepsilon^{-1} > \max_{t>k} Y_t\big) \le \varepsilon^{-\alpha}.
\end{aligned}$$

In the last step, we used the fact that the events $\{Y_k \ge \varepsilon^{-1} > \max_{t>k} Y_t\}$, $k \ge i$, are disjoint. Thus, we got a contradiction. This proves that for all $\varepsilon > 0$ there exist only finitely many $j \le 0$ such that $Y_j > \varepsilon$, hence $Y_t \overset{\text{a.s.}}{\to} 0$ and also $\Theta_t \overset{\text{a.s.}}{\to} 0$ as $t \to -\infty$, as desired.

In particular, the time T^* of the largest record of the sequence (Θ_t) is finite a.s.

Now assume that $\mathbb{P}(\sum_{j\in\mathbb{Z}} \Theta_j^\alpha = \infty) > 0$. Then there exists an $i \in \mathbb{Z}$ such that

$$\mathbb{P}\Big(\sum_{j\in\mathbb{Z}} \Theta_j^\alpha = \infty, T^* = i\Big) > 0,$$

and an application of the time-change formula (3.4) yields

$$\infty = \mathbb{E}\Big[\sum_{j\in\mathbb{Z}} \Theta_j^\alpha \mathbb{1}(T^* = i)\Big] = \sum_{j\in\mathbb{Z}} \mathbb{E}\big[\Theta_j^\alpha \mathbb{1}(T^* = i)\big] = \sum_{j\in\mathbb{Z}} \mathbb{P}(T^* = i - j) = 1,$$

leading to a contradiction. Thus $\sum_{j\in\mathbb{Z}} \Theta_j^\alpha < \infty$ a.s. This proves the lemma. □

3.3.2. *Point process convergence toward cluster Poisson processes.* The following point process result was proved in [9] and reproved in [2] by using the terminology of the spectral tail process.

We adapt the mixing condition in [9] tailored for point process convergence. It is expressed in terms of the Laplace functionals of point processes. Recall that a point process N with state space $E = \mathbb{R}_0 = \mathbb{R}\backslash\{0\}$ has Laplace functional

$$\Psi_N(g) = \mathbb{E}\left[\exp\left(-\int_E g\,dN\right)\right] \qquad \text{for } g \in \mathbb{C}_K^+,$$

where the set $\mathbb{C}_K^+$ consists of the continuous functions on E with compact support. Since 0 is excluded from E, this means that $g \in \mathbb{C}_K^+$ vanishes in some neighborhood of the origin. Moreover, we have the weak convergence of point processes $N_n \stackrel{d}{\to} N$ on E if and only if $\Psi_{N_n} \to \Psi_N$ pointwise (see [30,31]).

Mixing condition $\mathcal{A}(a_n)$ Consider integer sequences $r_n \to \infty$ and $k_n = [n/r_n] \to \infty$ and the point processes with state space $E = \mathbb{R}_0$,

$$N_n = \sum_{i=1}^{n} \varepsilon_{a_n^{-1}X_i} \qquad \text{and} \qquad \widetilde{N}_{r_n} = \sum_{i=1}^{r_n} \varepsilon_{a_n^{-1}X_i}, \quad n \geq 1,$$

where ε_x denotes Dirac measure at x. The stationary regularly varying sequence (X_t) satisfies $\mathcal{A}(a_n)$ if there exist (r_n) and (k_n) such that

$$\Psi_{N_n}(g) - \big(\Psi_{\widetilde{N}_{r_n}}(g)\big)^{k_n} \to 0, \quad n \to \infty, \quad g \in \mathbb{C}_K^+. \tag{3.5}$$

Remark 3.7. This condition is satisfied for a strongly mixing sequence (X_t) with mixing rate (α_h) if one can find integer sequences (ℓ_n) and (r_n) such that $\ell_n/r_n \to 0$, $r_n/n \to 0$ and $k_n\alpha_{\ell_n} \to 0$. This is a very mild condition indeed. Relation (3.5) ensures that if $N_n \stackrel{d}{\to} N$ on the state space E, then also $\sum_{i=1}^{k_n} \widetilde{N}_{r_n}^{(i)} \stackrel{d}{\to} N$, where $(\widetilde{N}_{r_n}^{(i)})_{i=1,\dots,k_n}$ are iid copies of $\widetilde{N}_{r_n}$. This fact ensures that the limit processes considered are infinitely divisible (cf. [19]).

Theorem 3.8. *Consider a stationary regularly varying sequence (X_t) with index $\alpha > 0$. We assume the following conditions*:

(1) *The mixing condition $\mathcal{A}(a_n)$ for integer sequences $r_n \to \infty$ such that $k_n = [n/r_n] \to \infty$ as $n \to \infty$.*

(2) *The anti-clustering condition* (2.2) *for the same sequence (r_n).*

Then we have the point process convergence on the state space $\mathbb{R}_0$

$$N_n = \sum_{i=1}^{n} \varepsilon_{a_n^{-1}X_i} \stackrel{d}{\to} N = \sum_{i=1}^{\infty} \sum_{j=-\infty}^{\infty} \varepsilon_{\Gamma_i^{-1/\alpha} Q_{ij}}, \tag{3.6}$$

where

- $\sum_{j=-\infty}^{\infty} \varepsilon_{Q_{ij}}$, $i = 1, 2, \dots$, *is an iid sequence of point processes with state space $\mathbb{R}$. A generic element $Q = (Q_j)$ of the sequence $Q^{(i)} = (Q_{ij})_{j\in\mathbb{Z}}$, $i = 1, 2, \dots$, has the distribution of the* spectral cluster process

$$Q = \Big(\frac{\Theta_t}{\|\Theta\|_\alpha}\Big)_{t\in\mathbb{Z}}.$$

- (Γ_i) *are the points of a unit rate homogeneous Poisson process on $(0, \infty)$.*
- (Γ_i) *and $(Q^{(i)})_{i=1,2,\dots}$ are independent.*

Remark 3.9. In view of Lemma 3.6, we know that $\|\Theta\|_\alpha < \infty$ a.s. Hence, the spectral cluster process Q is well defined.

Since the Poisson points $(\Gamma_i^{-1/\alpha})$ and the sequence of iid point processes $\big(\sum_{j\in\mathbb{Z}} \varepsilon_{Q_{ij}}\big)$ are independent, it is not difficult to calculate the Laplace functional of the limit process N:

$$\Psi_N(g) = \exp\Big(-\int_0^\infty \mathbb{E}\big[1 - \mathrm{e}^{-\sum_{j\in\mathbb{Z}} g(y\,Q_j)}\big]\,d(-y^{-\alpha})\Big)\,, \quad g \in \mathbb{C}_K^+.$$

Now we apply the change of variables $z = y\,|Q_{T^*}|$ in $\Psi_N(g)$, where

$$|Q_{T^*}| = \frac{|\Theta_{T^*}|}{\|\Theta\|_\alpha} = \frac{\max_{t\in\mathbb{Z}} |\Theta_t|}{\big(\sum_{j\in\mathbb{Z}} |\Theta_j|^\alpha\big)^{1/\alpha}}.$$

Then we obtain for $g \in \mathbb{C}_K^+$,

$$\Psi_N(g) = \exp\Big(-\mathbb{E}[|Q_{T^*}|^\alpha]\int_0^\infty \mathbb{E}\Big[\frac{|Q_{T^*}|^\alpha}{\mathbb{E}[|Q_{T^*}|^\alpha]}\big(1 - \mathrm{e}^{-\sum_{j\in\mathbb{Z}} g(z\,Q_j/|Q_{T^*}|)}\big)\Big]\,d(-z^{-\alpha})\Big).$$

According to Proposition 3.10 below, $\theta_{|X|} = \mathbb{E}[|Q_{T^*}|^\alpha]$. Now, changing the measure with the density $|Q_{T^*}|^\alpha/\mathbb{E}[|Q_{T^*}|^\alpha]$ and writing $\widetilde{Q} = (\widetilde{Q}_j)_{j\in\mathbb{Z}}$ for the sequence $Q/|Q_{T^*}|$ under the new measure, we arrive at

$$\Psi_N(g) = \exp\Big(-\int_0^\infty \mathbb{E}\big[\big(1 - \mathrm{e}^{-\sum_{j\in\mathbb{Z}} g(z\,\widetilde{Q}_j|)}\big)\big]\,d\big(-(z/\theta_{|X|}^{1/\alpha})^{-\alpha}\big)\Big).$$

However, this alternative expression of the Laplace functional Ψ_N corresponds to another representation of the point process N:

$$N = \sum_{i=1}^\infty \sum_{j=-\infty}^\infty \varepsilon_{(\Gamma_i/\theta_{|X|})^{-1/\alpha}\widetilde{Q}_{ij}}, \tag{3.7}$$

where the Poisson points $(\Gamma_i^{-1/\alpha})$ are independent of the sequence $\big(\sum_{j\in\mathbb{Z}} \varepsilon_{\widetilde{Q}_{ij}}\big)$ of iid copies of $\sum_{j\in\mathbb{Z}} \varepsilon_{\widetilde{Q}_j}$.

We observe that $|\widetilde{Q}_j| \le 1$ a.s. and $|\widetilde{Q}_{T^*}| = 1$ a.s. The extremal index $\theta_{|X|}$ plays an important role in representation (3.7). Each Poisson point $(\Gamma_i/\theta_{|X|})^{-1/\alpha}$ stands for the radius of a circle around the origin, and the points $(\widetilde{Q}_{ij})_{j\in\mathbb{Z}}$ are inside or on this circle. In this sense, each Poisson point $(\Gamma_i/\theta_{|X|})^{-1/\alpha}$ creates an extremal cluster. Therefore, we refer to the process N as a *cluster Poisson process.*

3.3.3. *Equivalent expressions for the extremal index.* Based on the results in the previous subsection, we can derive equivalent expressions of $\theta_{|X|}$ in terms of Q_{T^*} and T^*.

Proposition 3.10. *Assume the conditions of Theorem* 3.8. *Then the extremal index $\theta_{|X|}$ of $(|X_t|)$ coincides with the following quantities*:

$$\mathbb{E}[|Q_{T^*}|^\alpha] = \mathbb{P}(Y\,|Q_{T^*}| > 1) = \mathbb{P}(T^* = 0). \tag{3.8}$$

Here Y is a Pareto(α) independent of Q_{T^*} and T^* is the time of the largest record of $(|\Theta_t|)$.

Remark 3.11. We observe that

$$\mathbb{E}[|Q_{T^*}|^\alpha] = \mathbb{E}\Big[\frac{\max_{t\in\mathbb{Z}} |\Theta_t|^\alpha}{\sum_{j\in\mathbb{Z}} |\Theta_j|^\alpha}\Big] = \theta_{|X|}\,.$$

Since $\theta_{|X|} = \mathbb{P}(T^* = 0)$, the extremal index $\theta_{|X|}$ has an intuitive interpretation as the probability that $(|\Theta_t|)$ assumes its largest value at time zero.

Example 3.12. We consider the regularly varying solution of an affine stochastic recurrence equation under the conditions and with the notation of Example 3.2. An exception where the extremal index has an explicit solution is the case $\log A_t = N_t - 0.5$ for an iid standard normal sequence (N_t). Then $\mathbb{E}[A_t] = 1$ and the theory mentioned in Example 3.2 yields regular variation of (X_t) with index 1. Using the expression $\mathbb{P}(T^* = 0)$ and applying some random walk theory (such as the results in [8]), one obtains an exact expression for θ_X in terms of the Riemann zeta function ζ (see Example 3.13). A first order approximation to this formula is given by

$$\theta_X \approx \frac{1}{2}\exp\Big(\frac{\zeta(0.5)}{\sqrt{2\pi}}\Big) \approx \frac{1}{2}\exp(-0.5826) \approx 0.2792. \tag{3.9}$$

Example 3.13. Let $B^{(i)} = (B_t)_{t\in\mathbb{R}}$ be iid standard Brownian motions independent of $\Gamma_1 < \Gamma_2 < \cdots$, which are the points of a unit-rate Poisson process on $(0,\infty)$. We consider the stationary max-stable *Brown-Resnick* [4] *process*

$$X_t = \sup_{i\ge 1} \Gamma_i^{-1}\,\mathrm{e}^{\sqrt{2}\,B_t^{(i)} - |t|}, \quad t\in\mathbb{R}.$$

It has unit Fréchet marginals $\mathbb{P}(X_t \le x) = \Phi_1(x) = \mathrm{e}^{-x^{-1}}$, $x>0$. Any discretization $X^{(\delta)} = (X_{\delta t})_{t\in\mathbb{Z}}$ for $\delta>0$ is regularly varying with index 1 and spectral tail process $\Theta_t^{(\delta)} = \mathrm{e}^{\sqrt{2}B_{\delta t} - \delta|t|}$, $t\in\mathbb{Z}$. Direct calculation of $-x\,\log \mathbb{P}(n^{-1}\max_{1\le t\le n} X_{\delta t} \le x)$, $x>0$, yields the extremal index of $X^{(\delta)}$ as the limit

$$\theta_X^{(\delta)} = \lim_{n\to\infty} n^{-1}\mathbb{E}\Big[\sup_{0\le t\le n} \mathrm{e}^{\sqrt{2}B_{\delta t} - \delta t}\Big]. \tag{3.10}$$

We use the expression $\theta_X^{(\delta)} = \mathbb{P}(T^{*(\delta)} = 0)$, where $T^{*(\delta)}$ is the first record time of $(\Theta_t^{(\delta)})_{t\in\mathbb{Z}}$ (see (3.8)). We consider the first ladder height epoch $\tau_+(\delta) = \inf\{t\ge 1 : \sqrt{2}B_{\delta t} + \delta t < 0\}$. Using the symmetry of the Gaussian distribution, $(\Theta_t^{(\delta)})_{t\ge 1} \overset{d}{=} (1/\Theta_{-t}^{(\delta)})_{t\ge 1}$, we obtain $\theta_X^{(\delta)} = \mathbb{P}(T^{*(\delta)} = 0) = \mathbb{P}(\tau_+(\delta) = \infty)^2$. Combining this with the classical identity $\mathbb{P}(\tau_+(\delta) = \infty) = 1/\mathbb{E}[\tau_-(\delta)]$ for $\tau_-(\delta) = \inf\{t\ge 1 : \sqrt{2}\,B_{\delta t} - \delta t \le 0\}$, we get from random walk theory (see [1])

$$\theta_X^{(\delta)} = \Big(\frac{1}{\mathbb{E}[\tau_-(\delta)]}\Big)^2 = \Big(\frac{\mathbb{E}[B_\delta - \delta]}{\mathbb{E}[\sqrt{2}B_{\tau_-(\delta)} - \tau_-(\delta)]}\Big)^2 = \delta^2(\mathbb{E}[\sqrt{2}B_{\tau_+(\delta)} + \tau_+(\delta)])^{-2},$$

where we used Wald's lemma and the symmetry of the Gaussian distribution. To be able to apply Theorem 1.1 in [8], we standardize the increments of the random walk $\sqrt{2}B_{\delta t}$ dividing them by $\sqrt{2\delta}$, turning the drift into $\sqrt{\delta/2}$, and we get

$$\mathbb{E}[\sqrt{2}\,B_{\tau_+(\delta)} + \tau_+(\delta)] = \sqrt{\delta}\exp\Big(-\frac{\sqrt{\delta}}{2\sqrt{\pi}}\sum_{n=0}^\infty \frac{\zeta(1/2-n)}{n!(2n+1)}\Big(-\frac{\delta}{4}\Big)^n\Big).$$

This implies that

$$\theta_X^{(\delta)} = \delta\exp\Big(\sqrt{\frac{\delta}{\pi}}\sum_{n=0}^\infty \frac{\zeta(1/2-n)}{n!(2n+1)}\Big(-\frac{\delta}{4}\Big)^n\Big).$$

We recover the *Pickands constant* of the Brown–Resnick process (see [29]) as $\lim_{\delta\downarrow 0}\delta^{-1}\,\theta_X^{(\delta)}$:

$$\mathcal{H}_X^{(0)} = \lim_{T\to\infty}\frac{1}{T}\mathbb{E}\Big[\sup_{0\le t\le T}\mathrm{e}^{\sqrt{2}B_t - t}\Big] = 1.$$

Proof of Proposition 3.10. Consider the supremum of all points of the limit process N in Theorem 3.8:

$$M = \sup_{i\ge 1}\Gamma_i^{-1/\alpha}\sup_{j\in\mathbb{Z}}|Q_{ij}|.$$

The sequences (Γ_i) and $(Q^{(i)})$ are independent and $M = \sup_{i\ge 1}\Gamma_i^{-1/\alpha}V_i$ for the iid sequence $V_i := \sup_{j\in\mathbb{Z}}|Q_{ij}|$, $i = 1,2,\dots$, whose generic element V has the property $\mathbb{E}[V^\alpha]<\infty$. Indeed, $V\le 1$ a.s. by construction. The points $(\Gamma_i^{-1/\alpha},V_i)$ constitute a marked Poisson process $N_{\Gamma,V}$ with state space $E = (0,\infty)\times[0,\infty)$ and mean measure given by $\mu((x,\infty)\times[0,y]) = x^{-\alpha}\,F_V(y)$, $x>0, y\ge 0$, where F_V is the distribution function of V. For $x>0$ we consider $B_x = \{(y,v)\in E : y\,v > x\}$. We observe that

$$\mu(B_x) = \int_{v=0}^{\infty}\int_{y=x/v}^{\infty}\alpha\,y^{-\alpha-1}\,F_V(dv) = \int_0^{\infty}(x/v)^{-\alpha}\,F_V(dv) = x^{-\alpha}\,\mathbb{E}[V^\alpha].$$

Therefore, we have for $x>0$,

$$\mathbb{P}(M\le x) = \mathbb{P}\big(\Gamma_i^{-1/\alpha}\,V_i\le x, i\ge 1\big) = P(N_{\Gamma,V}(B_x)=0) = \mathrm{e}^{-\mu(B_x)} = \mathrm{e}^{-x^{-\alpha}\,\mathbb{E}[V^\alpha]}.$$

Thus, M is a scaled version of the standard Fréchet distribution, $\Phi_\alpha(x) = \mathrm{e}^{-x^{-\alpha}}$, $x>0$:

$$\mathbb{P}(M\le x) = \Phi_\alpha^{\mathbb{E}[V^\alpha]}(x)\,,\quad x>0\,.$$

On the other hand, Theorem 3.8 and an application of the continuous mapping theorem yield as $n\to\infty$,

$$\mathbb{P}\big(a_n^{-1}M_n\le x\big) = \mathbb{P}\big(N_n(x,\infty)=0\big)\to\mathbb{P}\big(N(x,\infty)=0\big) = \mathbb{P}(M\le x),\quad x>0.$$

In view of the definition of the extremal index of the sequence $(|X_t|)$, we can identify

$$\mathbb{E}[V^\alpha] = \mathbb{E}\Big[\sup_{j\in\mathbb{Z}}|Q_j|^\alpha\Big] = \mathbb{E}[|Q_{T^*}|^\alpha]$$

as the value $\theta_{|X|}$. This proves the first part of (3.8). The identity

$$\mathbb{E}[|Q_{T^*}|^\alpha] = \mathbb{P}(Y\,|Q_{T^*}|>1) = \mathbb{P}(|Q_{T^*}|^\alpha > Y^{-\alpha})$$

is immediate, since Q and Y are independent, and $Y^{-\alpha}$ is $U(0,1)$ distributed.

Applying the time-change formula (3.4), shifting k to zero, we obtain

$$\theta_{|X|} = \mathbb{E}[|Q_{T^*}|^\alpha] = \sum_{k\in\mathbb{Z}}\mathbb{E}\Big[\frac{|\Theta_k|^\alpha}{\sum_{j\in\mathbb{Z}}|\Theta_j|^\alpha}\,\mathbb{1}(T^*=k)\Big] = \sum_{k\in\mathbb{Z}}\mathbb{E}\Big[\frac{|\Theta_{-k}|^\alpha}{\sum_{j\in\mathbb{Z}}|\Theta_{j-k}|^\alpha}\,\mathbb{1}(T^*=0)\Big] = \mathbb{P}(T^*=0).$$

This proves the last identity in (3.8). □

4. Estimation of the extremal index - a short review and a new estimator based on the spectral cluster process

First approaches to the estimation of the extremal index are due to [17, 38]. Estimators based on exceedences of a threshold were proposed in [12, 33, 35, 36]. A modern approach to the maxima method was started in [26]; improvements and asymptotic limit theory can be found in [3, 5].

We consider some standard estimators of θ_X. For the sake of argument we assume that (X_t) is a nonnegative stationary process with marginal distribution F, $k_n = n/r_n$ is an integer sequence such that $r_n \to \infty$, $k_n \to \infty$, and (u_n) is a threshold sequence satisfying $u_n \uparrow x_F$.

4.1. Blocks estimator. Recall that θ_X has interpretation as the reciprocal of the expected size of extremal clusters. This idea is the basis for inference procedures from the early 1990s (see [11, 37]). Clusters are identified as blocks of length $r = r_n$ with at least one exceedance of a high threshold $u = u_n$. A blocks estimator $\widehat{\theta}^{\mathrm{bl}}$ is given by the ratio of the number $K_n(u)$ of such clusters and the total number of exceedences $N_n(u)$:

$$\widehat{\theta}^{\mathrm{bl}}_u(r) = \frac{K_n(u)}{N_n(u)} := \frac{\sum_{t=1}^{k_n} \mathbb{1}(M_{(t-1)r+1,t\,r} > u)}{\sum_{t=1}^{n} \mathbb{1}(X_t > u)}. \tag{4.11}$$

This method requires the choice of block length r and threshold level u satisfying $r_n\overline{F}(u_n) \to 0$; if $r_n \to \infty$ does not hold at the prescribed rate, $\widehat{\theta}^{\mathrm{bl}}$ is biased. Estimators using clusters of extreme exceedences were also considered in [17].

A slight modification of the blocks estimator is the *disjoint blocks estimator* of [38]:

$$\widehat{\theta}^{\mathrm{dbl}} = \frac{\log(1 - K_n(u)/k_n)}{r\,\log(1 - N_n(u)/n)}.$$

Assuming some weak dependence condition on (X_t), the heuristic idea behind the estimator is the approximations

$$\big(\mathbb{P}(M_r \le u_n)\big)^{k_n} \approx \mathbb{P}(M_n \le u_n) \approx F^{\theta_X n}(u_n)$$

for a suitable sequence (u_n). Then, taking logarithms and replacing $\overline{F}(u_n)$ and $\mathbb{P}(M_n > u_n)$ by their empirical estimators $N_n(u)/n$ and $K_n(u)/k_n$, respectively, we obtain

$$\theta_X \approx \frac{\log \mathbb{P}(M_n \le u_n)}{n\,\log F(u_n)} = \frac{\log(1 - \mathbb{P}(M_n > u_n))}{n\,\log(1 - \overline{F}(u_n))} \approx \frac{\log(1 - K_n(u)/k_n)}{r_n\,\log(1 - N_n(u)/n)} = \widehat{\theta}^{\mathrm{dbl}}.$$

Assuming that both $K_n(u)/k_n$ and $N_n(u)/n$ converge to zero, a Taylor expansion of $\log(1+x) = x(1+o(1))$ as $x \to 0$ shows that $\widehat{\theta}^{\mathrm{bl}} \approx \widehat{\theta}^{\mathrm{dbl}}$. [38] showed that $\widehat{\theta}^{\mathrm{dbl}}$ has a smaller asymptotic variance than $\widehat{\theta}^{\mathrm{bl}}$. [33] proposed a sliding blocks version of $\widehat{\theta}^{\mathrm{dbl}}$ with an even smaller asymptotic variance

$$\widehat{\theta}^{\mathrm{slbl}}(u,r) = \frac{-\log\Big(\frac{1}{n-r+1}\sum_{t=1}^{n-r+1} \mathbb{1}(M_{t,t+r} \le u)\Big)}{N_n(u)/k_n}. \tag{4.12}$$

4.2. Runs and intervals estimator. [38] proposed an alternative *runs estimator*. It is based on the limit relation (2.4): the probability $\mathbb{P}(M_{\ell_n} \le u_n \mid X_0 > u_n)$ is replaced by a sample version for some sequence $l=l_n\to\infty$:

$$\widehat{\theta}_u^{\mathrm{runs}}(l)=\frac{1}{N_n(u)}\sum_{i=1}^{n-l}\mathbb{1}(X_i>u_n\,,M_{i+1,i+l}\le u_n). \tag{4.13}$$

Clusters are considered distinct if they are separated by at least l observations not exceeding u. In [12], a complete study of the runs estimator and the inter-exceedence times is given. The thresholds (u_n) need to satisfy $r_n\overline{F}(u_n)\to 1$, and $l_n\le r_n$.

Consider the *exceedance times*

$$S_0(u)=0,\quad S_i(u)=\min\{t>S_{i-1}(u):X_t>u_n\},\quad i\ge 1,$$

with *inter-exceedance times* $T_i(u)=S_i(u)-S_{i-1}(u)$, $i\ge 1$. The sequence $(T_i(u))_{i\ge 2}$ constitutes a stationary sequence. If $r_n\,\overline{F}(u_n)\to 1$, [12] noticed that $(n\,T_2(u))$ converges in distribution to a limiting mixture given by $(1-\theta_X)\mathbb{1}_0(x)+\theta_X\,(1-\mathrm{e}^{-\theta_X\,x})$, $x\ge 0$. Calculation yields the coefficient of variation ν of $T_2(u)$ whose square is given by

$$\nu^2=\mathrm{var}(T_2(u))/(\mathbb{E}[T_2(u)])^2=\mathbb{E}[T_2^2(u)]/(\mathbb{E}[T_2(u)])^2-1=2/\theta_X-1,$$

leading to overdispersion $\nu>0$ if and only if $\theta_X<1$. Replacing the moments on the left-hand side by sample versions and adjusting the empirical moments for bias, [12] arrived at the *intervals estimator*

$$\widehat{\theta}^{\mathrm{int}}(u)=1\wedge\frac{2\Big(\sum_{i=2}^{N_n(u)}(T_i(u)-1)\Big)^2}{(N_n(u)-1)\sum_{i=2}^{N_n(u)}(T_i(u)-1)(T_i(u)-2)}. \tag{4.14}$$

See also [35, 36].

4.3. Northrop's estimator. Assume for the moment that (X_i) is iid and F is continuous. Then $F(X)$ is uniform on $(0,1)$. Hence, for $r=r_n$ and $x>0$,

$$\begin{aligned}\mathbb{P}\big(-r_n\,\log F(M_r)>x\big)&=\mathbb{P}(F(M_r)\le\mathrm{e}^{-x/r})=\mathbb{P}(\max_{i=1,\ldots,r_n}F(X_i)\le\mathrm{e}^{-x/r})\\&=\big(\mathbb{P}(F(X)\le\mathrm{e}^{-x/r})\big)^r=\mathrm{e}^{-x}\,.\end{aligned}$$

For a weakly dependent sequence (X_i) with marginal distribution F, assume the existence of an extremal index for $(F(X_t))$ which, by monotonicity of F, coincides with θ_X:

$$\mathbb{P}\big(-r_n\,\log F(M_r)>x\big)=\mathbb{P}(\max_{i=1,\ldots,r_n}F(X_i)<\mathrm{e}^{-x/r})\to\mathrm{e}^{-\theta_X\,x},\quad x>0.$$

Thus, the $(-r_n\,\log F(M_r))$ are asymptotically $\mathrm{Exp}(\theta_X)$ distributed. For iid $\mathrm{Exp}(\theta_X)$, the maximum likelihood estimator of θ_X is given by the reciprocal of the sample mean. These ideas lead to Northrop's estimators [26]. Mimicking the maximum likelihood estimator of iid $\mathrm{Exp}(\theta_X)$ data for a stationary sequence (X_t), one considers the quantities $-r_n\,\log F(M_{t,t+r})$, $t=1,\ldots,n-r_n$, and constructs sliding or disjoint blocks estimators of θ_X:

$$\widehat{\theta}^{\mathrm{Nsl}}(r)=\Big(\frac{1}{n-r+1}\sum_{t=1}^{n-r+1}(-r\,\log F_n(M_{t,t+r}))\Big)^{-1}, \tag{4.15}$$

$$\widehat{\theta}^{\mathrm{Ndbl}}(r)=\Big(\frac{1}{[n/r]}\sum_{i=1}^{[n/r]}(-r\,\log F_n(M_{r\,(i-1)+1,r\,i}))\Big)^{-1}. \tag{4.16}$$

Here F_n is an empirical distribution function of the data. This particular choice of estimator of F depends on the whole sample, hence introduces additional dependence. This fact requires an optimal choice of block length r_n for implementation.

4.4. An estimator based on the spectral cluster process. In this subsection, we consider a stationary nonnegative regularly varying process (X_t) with index $\alpha > 0$, spectral tail process (Θ_t) and normalizing sequence (a_n) satisfying $n\,\mathbb{P}(X > a_n) \to 1$. Proposition 3.10 yields the alternative representation $\theta_X = \mathbb{E}[Q_{T^*}^\alpha]$, where (Q_t) is a spectral cluster process of (X_t). We construct an estimator based on this identity.

We consider sums and maxima over disjoint blocks of size $r = r_n = o(n)$:

$$S_{i,r}^{(\alpha)} := \sum_{t=(i-1)\,r+1}^{i\,r} X_t^\alpha\,, \quad M_{i,r} = \max_{t=(i-1)r+1,\dots,i\,r} X_t, \quad i = 1,\dots,k_n\,.$$

The following limit relation is proved in [7]:

$$\lim_{n\to\infty} \mathbb{E}\big[\,M_{1,r}^\alpha / S_{1,r}^{(\alpha)} \mid S_r^{(\alpha)} > a_n^\alpha\big] = \mathbb{E}[Q_{T^*}^\alpha], \tag{4.17}$$

which is based on large deviation results for regularly varying stationary sequences (see, for example, [7]). Now we build an estimator of θ_X from an empirical version of the left-hand expectation. Define the corresponding estimator by

$$\widehat{\theta}_v^{\rm scp}(r) := \frac{\sum_{i=1}^{k_n} \frac{M_{i,r}^\alpha}{S_{i,r}^{(\alpha)}}\, \mathbb{1}\big(S_{i,r}^{(\alpha)} > v\big)}{\sum_{i=1}^{k_n} \mathbb{1}\big(S_{i,r}^{(\alpha)} > v\big)}. \tag{4.18}$$

Here we choose $v = S_{(s),r}^{(\alpha)}$, the sth largest among $(S_{i,r}^{(\alpha)})_{i=1,\dots,k_n}$ for an integer sequence $s = s_n$ such that $s_n = o(k_n)$.

5. A Monte-Carlo study of the estimators

We run a short study based on 1 000 simulated processes $(X_t)_{t=1,\dots,5000}$ for comparing the performances of some of the aforementioned estimators. First, (X_t) is an AR(1) process with parameter $\varphi = 0.2$ and iid student(1) noise, resulting in a regularly varying process with index 1 and $\theta_{|X|} = 0.8$. Second, we consider the regularly varying solution of an affine stochastic recurrence equation with iid $\log A_t \sim N(-0.5, 1)$, $B_t \equiv 1$, and $\theta_X \approx 0.2792$ (see (3.9)).

Figures 5.1 and 5.2 show boxplots of the simulation study.

- $\widehat{\theta}^{\rm bl}$ and $\widehat{\theta}^{\rm runs}$ are functions of the block and run lengths, respectively. u is the largest $[n^{0.6}]$th upper order statistic of the sample.
- $\widehat{\theta}^{\rm slbl}$ is a function of r. u is the rth upper order statistic of the sample.
- $\widehat{\theta}^{\rm int}$ is a function of x. u is the $[n/x]$th upper order statistic.
- $\widehat{\theta}^{\rm Nsl}$, $\widehat{\theta}^{\rm scp}$ are functions of r.
- For $\widehat{\theta}^{\rm scp}$ we choose $s = [n^{0.6}/r]$. The tail index α is estimated by the Hill estimator from [14] based on $[n^{0.8}]$ upper order statistics of the sample.

According to the folklore in the literature, Northrop's estimator $\widehat{\theta}^{\rm Nsl}$ outperforms the classical estimators (runs, blocks); it has smallest variance but it can be difficult to control its bias. Our experience with $\widehat{\theta}^{\rm scp}$ shows that it performs better than the other estimators as regards the bias, especially when θ_X is small. The intervals estimator $\widehat{\theta}^{\rm int}$ is preferred by practitioners because the choice of the hyperparameter x is robust with respect to different values of θ_X. This cannot be said about the other estimators with the exception of $\widehat{\theta}^{\rm scp}$. In our experiments with sample size $n = 5000$, the choices $x = 32$ and $r = 64$ work well for $\widehat{\theta}^{\rm int}$ and $\widehat{\theta}^{\rm scp}$, respectively. We did not fine-tune the hyperparameter s in $\widehat{\theta}^{\rm scp}$ in our experiments.

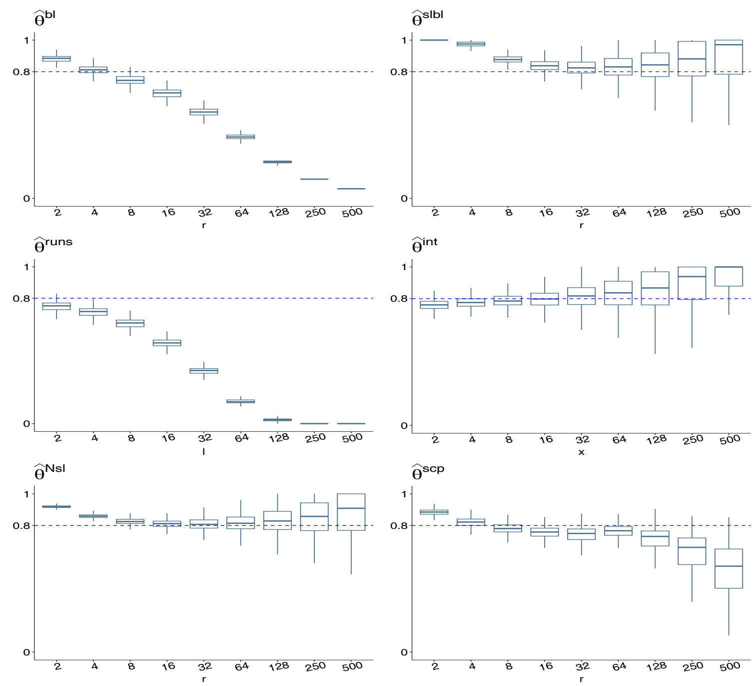

Fig. 5.1. Boxplots based on 1 000 simulations for the estimation of $\theta_{|X|} = 0.8$ in the $AR(1)$ model with $\varphi = 0.2$ and iid student(1) noise.

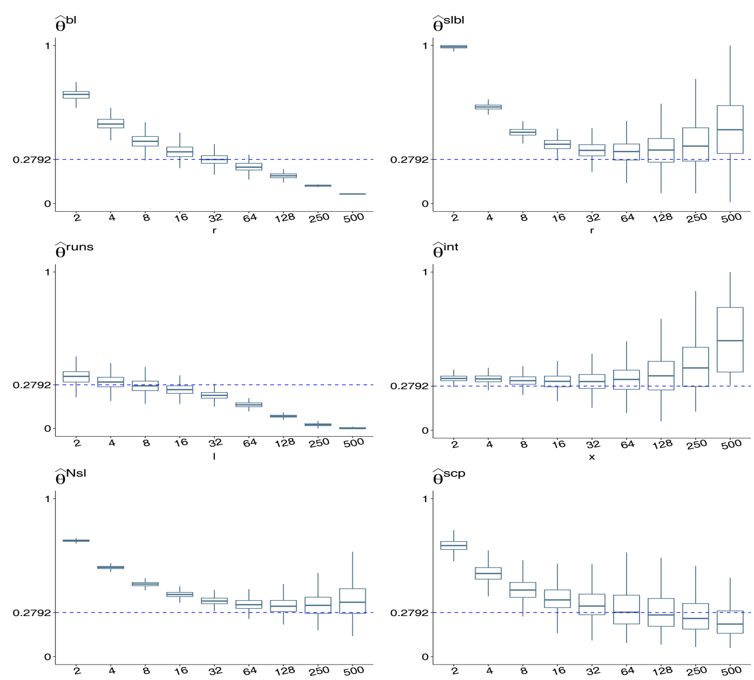

Fig. 5.2. Boxplots based on 1 000 simulations for the estimation of $\theta_X \approx 0.2792$ for the solution to a stochastic recurrence equation.

Thomas Mikosch's research is partially supported by Danmarks Frie Forskningsfond Grant No. 9040-00086B.

REFERENCES

1. S. Asmussen, *Applied Probability and Queues*, 2nd edition. Springer, New York, 2003.
2. B. Basrak and J. Segers, "Regularly varying multivariate time series," *Stoch. Proc. Appl.*, **119**, 1055–1080 (2009).
3. B. Berghaus and A. Bücher, "Weak convergence of a pseudo maximum likelihood estimator for the extremal index," *Ann. Stat.*, **46**, 2307–2335 (2018).
4. B. M. Brown and S. I. Resnick, "Extreme values of independent stochastic processes," *J. Appl. Probab.*, **14**, 732–739 (1977).
5. A. Bücher and T. Jennessen, "Method of moments estimators for the extremal index of a stationary time series," *Electr. J. Stat.*, **14** , 3103–3156 (2020).
6. D. Buraczewski, E. Damek, and T. Mikosch, *Stochastic Models with Power-Laws. The Equation $X = AX + B$*, Springer, New York, 2016.
7. G. Buriticá, T. Mikosch, and O. Wintenberger, "Large deviations of l^p-blocks of regularly varying time series and applications to cluster inference," Technical report, `arXiv citation: arXiv:2106.12822`, 2021.
8. J. T. Chang and Y. Peres, "Ladder heights, Gaussian random walks and the Riemann zero function," *Ann. Probab.*, **25**, 787–802 (1997).
9. R. A. Davis and T. Hsing, "Point process and partial sum convergence for weakly dependent random variables with infinite variance," *Ann. Probab.*, **23**, 879–917 (1995).
10. R. A. Davis and S. I. Resnick, "Limit theory for moving averages of random variables with regularly varying tail probabilities," *Ann. Probab.*, **13**, 179–195 (1985).
11. A. C. Davison and R. L. Smith, "Models for exceedances over high thresholds (with discussion)," *J. Royal Statist. Soc., Ser. B*, **52**, 393–442 (1990).
12. C. A. Ferro and J. Segers, "Inference for clusters of extreme values," *J. Royal Statist. Soc., Ser. B*, **65**, 545–556 (2003).
13. C. M. Goldie, "Implicit renewal theory and tails of solutions of random equations," *Ann. Appl. Probab.*, **1**, 126–166 (1991).
14. L. de Haan, C. Mercadier, and C. Zhou, "Adapting extreme value statistics to financial time series: dealing with bias and serial dependence," *Finance and Stochastics*, **20**, 321–354 (2016).
15. L. de Haan, "A spectral representation for max-stable processes," *Ann. Probab.*, **12**, 1194–1204 (1984).
16. L. de Haan, S. I. Resnick, H. Rootzén, and C.G. de Vries, "Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH processes," *Stoch. Proc. Appl.*, **32**, 213–224 (1989).
17. T. Hsing, "Extremal index estimation for a weakly dependent stationary sequence," *Ann. Stat.*, 2043–2071 (1993).
18. Z. Kabluchko, "Spectral representations of sum- and max-stable processes," *Extremes*, **12**, 401–424 (2009).
19. O. Kallenberg, *Random Measures, Theory and Applications*, Springer, Cham. 2017.
20. H. Kesten, "Random difference equations and renewal theory for products of random matrices," *Acta Math.*, **131**, 207–248 (1973).
21. R. Kulik and P. Soulier, *Heavy-Tailed Time Series*, Springer, New York, 2020.
22. M. R. Leadbetter, "Extremes and local dependence in stationary sequences," *Probab. Th. Relat. Fields*, **65**, 291–306 (1983).
23. M. R. Leadbetter, G. Lindgren, and H. Rootzén, *Extremes and Related Properties of Random Sequences and Processes*, Springer, Berlin, 1983.

24. R. M. Loynes, “Extreme values in uniformly mixing stationary stochastic processes,” *Ann. Math. Statist.*, **36**, 993–999 (1965).
25. G. F. Newell, “Asymptotic extremes for m-dependent random variables,” *Ann. Math. Statist.*, **35**, 1322–1325 (1964).
26. P. J. Northrop, “An efficient semiparametric maxima estimator of the extremal index,” *Extremes*, **18**, 585–603 (2015).
27. G. L. O’Brien, “Limit theorems for the maximum term of a stationary process,” *Ann. Probab.*, **2**, 540–545 (1974).
28. G. L. O’Brien, “Extreme values for stationary and Markov sequences,” *Ann. Probab.*, **15**, 281–291 (1987).
29. J. Pickands, “Asymptotic properties of the maximum in a stationary Gaussian process,” *Trans. Amer. Math. Soc.*, **145**, 75–86 (1969).
30. S. I. Resnick, *Extreme Values, Regular Variation, and Point Processes*, Reprint 2008. Springer, New York, 1987.
31. S. I. Resnick, *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Springer, New York, 2007.
32. C. Y. Robert, “Inference for the limiting cluster size distribution of extreme values,” *Ann. Stat.*, **37**, 271–310 (2009).
33. C. Y. Robert, C. A. Ferro, and J. Segers, “A sliding blocks estimator for the extremal index,” *Electr. J. Stat.*, **3**, 993–1020 (2009).
34. J. Segers, “Approximate distributions of clusters of extremes,” *Stat. Probab. Lett.*, **74**, 330–336 (2005).
35. M. Süveges, “Likelihood estimation of the extremal index,” *Extremes*, **10**, 41–55 (2007).
36. M. Süveges and A. C. Davison, “Model misspecification in peaks over threshold analysis,” *Ann. Appl. Stat.*, **4**, 203–221 (2010).
37. R. L. Smith, “Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone,” *Statist. Sci.*, **4**, 367–377 (1989).
38. R. L. Smith and I. Weissman, “Estimating the extremal index,” *J. Royal Statist. Soc., Ser. B*, **56**, 515–528 (1994).