Some variations on the extremal index

Buriticá, Gloria; Meyer, Nicolas Benjamin; Mikosch, Thomas Valentin; Wintenberger, Olivier

Published in:
Zapiski Nauchnykh Seminarov POMI

Publication date:
2022

Document version
Publisher's PDF, also known as Version of record

Document license:
Other

Citation for published version (APA):
Buriticá, G., Meyer, N. B., Mikosch, T. V., & Wintenberger, O. (2022). Some variations on the extremal index. Zapiski Nauchnykh Seminarov POMI, 501, 52–77.
G. Buriticá, N. Meyer, T. Mikosch, O. Wintenberger, Some variations on the extremal index, *Zap. Nauchn. Sem. POMI*, 2021, Volume 501, 52–77

Use of the all-Russian mathematical portal Math-Net.Ru implies that you have read and agreed to these terms of use
http://www.mathnet.ru/eng/agreement

Download details:
IP: 130.226.236.6
April 29, 2022, 12:28:28
SOME VARIATIONS ON THE EXTREMAL INDEX

Abstract. We re-consider 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 [2] 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, like 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.

Some personal words by Thomas Mikosch. During my PhD studies at the University of Leningrad 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 supported the Bernoulli Society actively, 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.

Key words and phrases: Extremal index, cluster Poisson process, extremal cluster, regularly varying time series, affine stochastic recurrence equation, autoregressive process.

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

$$M_n = \max_{t=1,\ldots,n} X_t, \quad n \geq 1,$$

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

$$\mathbb{P}(M_n \leq u_n) \approx \left[\mathbb{P}(X \leq u_n)^n\right]^\theta_X = ((F(u_n))^\theta_X, \quad n \to \infty,$$

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 with common distribution $F$ at the threshold $u_n$, the quantity $\theta_X$ describes the shrinking effect that the appearance of dependent extremes may have on the distribution of $M_n$ compared to $(F(u_n))^n$.

Leadbetter defined the extremal index $\theta_X$ 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 $\theta_X$ such that

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

If such a number $\theta_X$ 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 \((\tilde{X}_t)\) of \(X_1, \tilde{M}_n = \max(\tilde{X}_1, \ldots, \tilde{X}_n)\), there exist constants \(c_n > 0, d_n \in \mathbb{R}\) such that \(c_n^{-1}(\tilde{M}_n - d_n) \overset{d}{\to} \xi\) as \(n \to \infty\) and \(\xi\) has distribution \(H\). Then if \((X_t)\) has an extremal index \(\theta_X\) we have
\[
n \overline{F}(c_n x + d_n) \to -\log H(x), \quad n \to \infty, \quad x \in \text{supp} H,
\]
and
\[
P(c_n^{-1}(M_n - d_n) \leq x) \to H^{\theta_X}(x), \quad n \to \infty, \quad x \in \text{supp} H.
\]

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 \(P(M_n \leq u_n(\tau)) \to e^{-\tau}\). Hence \(\theta_X = 1\). The extremal index 1 is not exclusive to iid sequences. Indeed, in the book [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 \(\text{cov}(X_0, X_h) = o(1/\log h)\) 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 will give some justification for this interpretation.

2.1. The method of block maxima. The key is the definition of an extremal cluster in the sample \(X_1, \ldots, X_n\): split the sample into \(k_n = \lceil n/r_n \rceil\) blocks of equal length \(r_n\):
\[
\overbrace{X_1, \ldots, X_{r_n}}^{\text{Block 1}}, \overbrace{X_{r_n+1}, \ldots, X_{2r_n}}^{\text{Block 2}}, \ldots, \overbrace{X_{(k_n-1)r_n+1}, \ldots, X_{k_nr_n}}^{\text{Block } k_n},
\]
we ignore the last block of length less than \(r_n\), and we simply call a block an extremal cluster relative to a high threshold \(u = u_n(\tau)\) (this means that \(u_n \uparrow u_F\) as \(n \to \infty\)) if there is at least one exceedance of this threshold in this block. For an asymptotic theory it will be 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

\[
E \left[ \sum_{t=1}^{r_n} I(X_t > u_n) \mid M_{r_n} > u_n \right] = \sum_{t=1}^{r_n} \frac{P(X_t > u_n, M_{r_n} > u_n)}{P(M_{r_n} > u_n)}
\]

\[
= \sum_{t=1}^{r_n} \frac{P(X_t > u_n)}{P(M_{r_n} > u_n)}
\]

\[
= r_n \frac{P(X > u_n)}{P(M_{r_n} > u_n)} =: \frac{1}{\theta_n}.
\]

Obviously, \(\theta_n\) 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 \(\theta_X\); see Theorem 2.3 below. For this reason, the extremal index \(\theta_X\) is often referred to as the reciprocal of the expected extremal cluster size above high thresholds.

\[
\text{Figure 2.1. Visualization of the max-moving average}
\]

\[
X_t = \max(Z_t, Z_{t+1}, Z_{t+2}), \, t = 1, \ldots, 100,
\]

(blue) for iid student noise \(Z_t, \, t = 1, \ldots, 102\), 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 \(\theta_{|X|} = 1/3\).

### 2.2. Approximation of \(\theta_X\) by \(\theta_n\).

The following result can be found in slightly different forms in [9], proof of Lemma 2.8, [2, 34].

**Theorem 2.3.** 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:

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

Here \(M_{a,b} = \max_{i=a,\ldots,b} X_i\) for \(a \leq b\) such that \(M_b = M_{a,b}\) with \(a = 1\).

(3) A mixing condition holds:

\[
\mathbb{P}(M_n \leq u_n) - \left(\mathbb{P}(M_{r_n} \leq u_n)\right)^{k_n} \to 0, \quad n \to \infty,
\] (2.2)

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

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

Then the following statements hold:

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

\[
\lim_{k \to \infty} \limsup_{n \to \infty} \left| \theta_n - \mathbb{P}(M_n \leq u_n \mid X_0 > u_n) \right| = 0.
\] (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 \( \theta \).

Condition (2.2) is satisfied for strongly mixing \((X_t)\) with mixing rate \((\alpha_h)\) if one can find integer sequences \((\ell_n)\) 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 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
\[
\theta_X = \lim_{n \to \infty} P(M_{\ell_n} \leq u_n \mid X_0 > u_n) \tag{2.4}
\]
for a sequence \((\ell_n)\) with \(\ell_n/n \to 0\), thresholds \(u_n \uparrow x_F\) such that \(n P(u_n) \to 1\) as \(n \to \infty\), provided a mixing condition holds. O'Brien's condition (2.4) has the advantage that it avoids the definition of an extremal cluster.

Remark 2.4. Relation (2.3) provides a constructive way of calculating \(\theta_X\): if we know that the limits \(f(k) := \lim_{n \to \infty} P(M_k \leq u_n \mid X_0 > u_n)\) exist for every \(k \geq 1\) then we can try to derive \(\theta_X = \lim_{k \to \infty} f(k)\). In Section 3 we will follow this approach in the case of a regularly varying sequence.

§3. Regularly varying sequences

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 \(\alpha\) degrees of freedom we have \(\theta_{|X|} = 1 - |\varphi|^{\alpha}\); see Example 3.4 below. Thus, the smaller \(\alpha\) (the heavier the tail) for given \(\varphi\) 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 will 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 \( (\Theta_t) \) and a Pareto(\( \alpha \)) distributed \( Y \), i.e., \( \mathbb{P}(Y > y) = y^{-\alpha}, \ y > 1 \), such that \( (\Theta_t) \) and \( Y \) are independent and, for all \( h \geq 0 \),

\[
\mathbb{P}(x^{-1}(X_t)|_{t \in \mathbb{Z}} \in \cdot \mid |X_0| > x) \xrightarrow{\mathbb{W}} \mathbb{P}(Y(\Theta_t)|_{t \in \mathbb{Z}} \in \cdot), \quad x \to \infty.
\]

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 \( (\Theta_t) \) 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 \( \alpha \) (e.g. Pareto(\( \alpha \)) or student(\( \alpha \))). This means that a generic element \( Z \) satisfies \( \lim_{x \to \infty} \frac{\mathbb{P}(\pm Z > x)}{\mathbb{P}(|Z| > x)} = p_{\pm} \) for non-negative 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]):

\[
\mathbb{P}(\pm X > x) \mathbb{P}(|Z| > x) \sim \sum_{j=0}^{\infty} \left[ p_+ (\varphi^j)^{\alpha} + p_- (\varphi^j)^{\alpha} \right] = \mathbb{P}(\Theta_0 = \pm 1)(1 - |\varphi|^\alpha).
\]

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

\[
\Theta_t = \Theta_Z \text{sign}(\varphi^{J+t}) |\varphi|^\alpha \mathbb{I}(J + t \geq 0) = \Theta_0 \varphi^t \mathbb{I}(J + t \geq 0), \quad t \in \mathbb{Z},
\]

(3.1)

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

\[
\mathbb{P}(J = j) = (1 - |\varphi|^\alpha) |\varphi|^j, \quad j \geq 0.
\]
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 \( \alpha \). Under these conditions for a generic element \( X \), there exists a positive constant \( c_+ \) such that
\[
\mathbb{P}(X > x) \sim c_+ x^{-\alpha}, \quad x \to \infty.
\]
The forward spectral process is then given by
\[
(\Theta_t)_{t \geq 0} = (\Pi_t)_{t \geq 0}, \quad \text{where} \quad \Pi_t = A_1 \cdots A_t,
\]
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 2.4, we will derive the extremal index \( \theta_X \) of a stationary non-negative 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 \( k \geq 1 \),
\[
\mathbb{P}(a_n^{-1}M_k \leq 1 \mid X_0 > a_n) \to \mathbb{P}(Y \max_{1 \leq t \leq k} \Theta_t \leq 1) = \mathbb{P}(\max_{1 \leq t \leq k} \Theta_t^\alpha \leq Y^{-\alpha})
= \mathbb{E}[(1 - \max_{1 \leq t \leq k} \Theta_t^\alpha)^+] = \mathbb{E}[(1 - \max_{0 \leq t \leq k} \Theta_t^\alpha - \max_{1 \leq t \leq k} \Theta_t^\alpha)^+].
\]
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
\[
\lim_{n \to \infty} \theta_n = \lim_{k \to \infty} \lim_{n \to \infty} \mathbb{P}(a_n^{-1}M_k \leq 1 \mid X_0 > a_n)
= \lim_{k \to \infty} \mathbb{E}[(1 - \max_{1 \leq t \leq k} \Theta_t^\alpha - \max_{0 \leq t \leq k} \Theta_t^\alpha)]
= \mathbb{E}[(1 - \max_{t \geq 1} \Theta_t^\alpha)^+].
\]
From Theorem 2.3 we obtain the following result in [2].

**Corollary 3.3.** Consider a non-negative 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
   \[
   \theta = \mathbb{P}(Y \sup_{t \geq 1} \Theta_t \leq 1) = \mathbb{E}\left( (1 - \sup_{t \geq 1} \Theta_t^\alpha)_{+} \right) = \mathbb{E}\left[ (\sup_{t \geq 0} \Theta_t^\alpha - \sup_{t \geq 1} \Theta_t^\alpha)_{+} \right].
   \] (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 \(\theta_X\) exists and coincides with \(\theta\).

The representations of \(\theta\) 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 2.3. We conclude from Corollary 3.3 and the form of the spectral tail process given in (3.1) that
\[
\theta_{|X|} = \mathbb{E}\left( (1 - \max_{t \geq 1} \Theta_t^\alpha)_{+} \right) = 1 - \max_{t \geq 1} |\phi|^\alpha t = 1 - |\phi|^\alpha.
\]

This formula was already achieved in [10] in the 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 2.3; see [6]. We conclude from this result that \((X_t)\) has extremal index
\[
\theta_X = \mathbb{E}\left( (1 - \max_{t \geq 1} \Pi_t)_{+} \right) = \mathbb{E}\left[ (1 - \exp(\max_{t \geq 1} S_t))_{+} \right],
\]

where \(S_t = \sum_{i=1}^{t} \log A_i, \ t \geq 1\), is a random walk with a negative drift. This value of \(\theta_X\) was derived in [16]. In that paper a Monte Carlo simulation procedure for the evaluation of \(\theta_X\) was proposed. Direct calculation of \(\theta_X\) 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 non-negative 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

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

In particular, \(\Theta_t \to 0\) a.s. as \(|t| \to \infty\), and the time \(T^*\) of the largest record of \((\Theta_t)\) 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(\(\alpha\)) variable \(Y\) and the spectral tail process \((\Theta_t)\) are independent. We start by showing

\[ Y_t \xrightarrow{a.s.} 0, \quad t \to \infty. \quad (3.3) \]

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

\[
\lim_{h \to \infty} \lim_{n \to \infty} \mathbb{P}(M_{k,k+h} > x a_n \mid X_0 > a_n) = \lim_{h \to \infty} \mathbb{P}(\max_{k \leq t \leq k+h} Y_t > x) \\
= \mathbb{P}(\max_{t \geq k} Y_t > x).
\]

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

\[
\lim_{n \to \infty} \mathbb{P}(M_{k,k+h} > x a_n \mid X_0 > a_n) \\
\leq \limsup_{n \to \infty} \frac{\mathbb{P}(M_{k,r} > x a_n \mid X_0 > x a_n) \mathbb{P}(X > x a_n)}{\mathbb{P}(X > a_n)} \\
= x^{-\alpha} \limsup_{n \to \infty} \mathbb{P}(M_{k,r} > x a_n \mid X_0 > x a_n) \\
= x^{-\alpha} \varepsilon_k,
\]

and the right-hand side term \(\varepsilon_k\) vanishes for large \(k\). Hence, letting \(h \to \infty\),

we obtain for all \(x > 0,\)

\[ \mathbb{P}(\max_{t \geq k} Y_t > x) \leq x^{-\alpha} \varepsilon_k, \]

and therefore

\[ \lim_{k \to \infty} \mathbb{P}(\max_{t \geq k} Y_t > x) \leq \lim_{k \to \infty} x^{-\alpha} \varepsilon_k = 0, \]
implying \( \max_{t \geq k} Y_t \overset{p}{\to} 0 \) as \( k \to \infty \). Since \( (Y_t) = Y(\Theta_t) \) a.s. and \( Y > 0 \) is independent of \( (\Theta_t) \) this is only possible if \( \max_{t \geq k} \Theta_t \overset{p}{\to} 0 \) as \( k \to \infty \) but the latter relation is equivalent to \( \Theta_t \overset{a.s.}{\to} 0 \) as \( t \to \infty \), implying (3.3).

Next we show that
\[
Y_t \overset{a.s.}{\to} 0, \quad t \to \infty.
\]
Since \( Y_t \overset{a.s.}{\to} 0 \) as \( t \to \infty \) and \( Y_0 > 1 \) a.s. the following relation holds
\[
P\left( \bigcup_{t \geq 0} \{Y_t \geq 1 > \max_{t \geq i} Y_t\} \right) = \sum_{i \geq 0} P(Y_i \geq 1 > \max_{t \geq i} Y_t) = 1.
\]
Suppose that \( P\left( \sum_{j \leq 0} 1(Y_j > \epsilon) = \infty \right) > 0 \) for some \( \epsilon > 0 \). Then there exists some \( i \geq 0 \) such that
\[
P\left( \sum_{j \leq 0} 1(Y_j > \epsilon) = \infty, Y_i \geq 1 > \max_{t \geq i} Y_t \right) > 0.
\]

We recall the time-change formula from [2]:
\[
P((\Theta_{-h}, \ldots, \Theta_h) \in \cdot | \Theta_{-t} \neq 0) = E\left[ \frac{\Theta_t^\alpha}{E[\Theta_t^\alpha]} \mathbb{I}\left( \frac{(\Theta_{t-h}, \ldots, \Theta_{t+h})}{\Theta_t} \in \cdot \right) \right].
\]
In particular, \( P(\Theta_t \neq 0) = E[\Theta_t^\alpha] = 1 \) if and only if for all \( h \geq 0 \),
\[
P((\Theta_{-h}, \ldots, \Theta_h) \in \cdot) = E\left[ \frac{\Theta_t^\alpha}{E[\Theta_t^\alpha]} \mathbb{I}\left( \frac{(\Theta_{t-h}, \ldots, \Theta_{t+h})}{\Theta_t} \in \cdot \right) \right].
\]
Therefore
\[
\infty = E\left[ \sum_{j \leq 0} 1(Y_j > \epsilon) \mathbb{I}(Y_t \geq 1 > \max_{t \geq i} Y_t) \right]
\]
\[
= \sum_{j \leq 0} P(Y_j > \epsilon, Y_i \geq 1 > \max_{t \geq i} Y_t)
\]
\[
= \sum_{j \leq 0} \int_1^\infty E\left[ \mathbb{I}(y \Theta_j > \epsilon, y \Theta_t \geq 1 > y \max_{t \geq i} \Theta_t) \right] d(-y^{-\alpha})
\]
\[
= \sum_{j \leq 0} \int_1^\infty E\left[ \Theta_j^\alpha \mathbb{I}(y \epsilon \Theta_{-j}, y \frac{\Theta_{t-j}}{\Theta_{-j}} \geq 1 > y \max_{t \geq i} \Theta_{t-j}) \right] d(-y^{-\alpha})
\]
\[ \leq \varepsilon^{-\alpha} \sum_{j \leq 0} \mathbb{E} \left[ \int_{t>1} \mathbb{1}(z > 1, z \Theta_{1-j} \geq \varepsilon^{-1} > z \max_{t>i-j} \Theta_t) d(-\varepsilon^{-\alpha}) \right] \]
\[ = \varepsilon^{-\alpha} \sum_{j \leq 0} \mathbb{P}(Y_{i-j} \geq \varepsilon^{-1} > \max_{t>i-j} Y_t) \]
\[ = \varepsilon^{-\alpha} \sum_{k \geq 1} \mathbb{P}(Y_k \geq \varepsilon^{-1} > \max_{t>k} Y_t) \]
\[ \leq \varepsilon^{-\alpha}. \]

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

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

Now suppose that \(\mathbb{P}(\sum_{j \in \mathbb{Z}} \Theta_j = \infty) > 0\). Then there exists an \(i \in \mathbb{Z}\) such that
\[ \mathbb{P}(\sum_{j \in \mathbb{Z}} \Theta_j = \infty, T^* = i) > 0, \]
and an application of the time-change formula (3.4) yields
\[ \infty = \mathbb{E}\left[ \sum_{j \in \mathbb{Z}} \Theta_j \mathbb{1}(T^* = i) \right] = \sum_{j \in \mathbb{Z}} \mathbb{E}[\Theta_j \mathbb{1}(T^* = i)] \]
\[ = \sum_{j \in \mathbb{Z}} \mathbb{P}(T^* = i - j) = 1, \]
leading to a contradiction. Thus \(\sum_{j \in \mathbb{Z}} \Theta_j < \infty\) a.s. This proves the lemma.

\[ \square \]

3.3.2. **Point process convergence toward cluster Poisson processes.** The following point process result was proved in [9] and re-proved 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 \(E = \mathbb{R}_0 = \mathbb{R}\setminus\{0\}\)
has Laplace functional

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

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 \( N_n \mathop{\rightarrow}\limits^d N \) on \( E \) if and only if \( \Psi_{N_n} \rightarrow \Psi_N \) pointwise; see [30, 31].

**Mixing condition** \( A(a_n) \) Consider integer sequences \( r_n \rightarrow \infty \) and \( k_n = \lfloor n/r_n \rfloor \rightarrow \infty \) as \( n \rightarrow \infty \). The stationary regularly varying sequence \( (X_t) \) satisfies \( A(a_n) \) if there exist \( (r_n) \) and \( (k_n) \) such that

\[ \Psi_{N_n}(g) - \left( \Psi_{\tilde{N}_{r_n}}(g) \right)^{k_n} \rightarrow 0, \quad n \rightarrow \infty, \quad g \in \mathbb{C}_K^+. \]  

**Remark 3.7.** This condition is satisfied for a strongly mixing sequence \( (X_t) \) with mixing rate \( (\alpha_h) \) if one can find integer sequences \( (\ell_n) \) and \( (r_n) \) such that \( \ell_n/r_n \rightarrow 0, r_n/n \rightarrow 0 \) and \( k_n \alpha_{\ell_n} \rightarrow 0 \). This is a very mild condition indeed. Relation (3.5) ensures that, if \( N_n \mathop{\rightarrow}\limits^d N \) on the state space \( E = \mathbb{R}_0 \), then also \( \sum_{i=1}^{k_n} \tilde{N}_{r_n}^{(i)} \mathop{\rightarrow}\limits^d N \) where \( (\tilde{N}_{r_n}^{(i)})_{i=1,\ldots,k_n} \) are iid copies of \( \tilde{N}_{r_n} \). 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 \( A(a_n) \) for integer sequences \( r_n \rightarrow \infty \) such that \( k_n = \lfloor n/r_n \rfloor \rightarrow \infty \) as \( n \rightarrow \infty \).

2. The anti-clustering condition (2.2) for the same sequence \( (r_n) \).
Then we have the point process convergence on the state space $\mathbb{R}_0$

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

(3.6)

where

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

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

- $(\Gamma_i)$ are the points of a unit rate homogeneous Poisson process on $(0, \infty)$.
- $(\Gamma_i)$ and $(Q^{(i)})_{i=1, 2, \ldots}$ 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 $(\sum_{j \in \mathbb{Z}} \varepsilon_{Q_{ij}})$ are independent it is not difficult to calculate the Laplace functional of the limit process $N$:

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

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

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

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

$$\Psi_N(g) = \exp \left( - E|Q_{T^*}|^\alpha \right) \times \int_0^\infty E\left[ \frac{|Q_{T^*}|^\alpha}{E|Q_{T^*}|^\alpha} \left( 1 - e^{-\sum_{j \in \mathbb{Z}} g(zQ_j/|Q_{T^*}|)} \right) \right] d(-z^{-\alpha}).$$
According to Proposition 3.10 below, \( \theta_{|X|} = E[|Q_{T^*}|^\alpha] \). Now, changing the measure with the density \( |Q_{T^*}|^\alpha / E[|Q_{T^*}|^\alpha] \) and writing \( Q = (\tilde{Q}_j)_{j \in \mathbb{Z}} \) for the sequence \( Q/|Q_{T^*}| \) under the new measure, we arrive at

\[
\Psi_N(g) = \exp \left( - \int_0^{\infty} E \left[ 1 - e^{-\sum_{j \in \mathbb{Z}} g(\tilde{Q}_j)} \right] d \left( z/\theta_{|X|} \right) \right).
\]

However, this alternative expression of the Laplace functional \( \Psi_N \) corresponds to another representation of the point process \( N \):

\[
N = \sum_{i=1}^{\infty} \sum_{j=-\infty}^{\infty} \varepsilon_{(\Gamma_i/\theta_{|X|})^{-1/\alpha}\tilde{Q}_j},
\]

where the Poisson points \( (\Gamma_i^{-1/\alpha}) \) are independent of the sequence \( \left( \sum_{j \in \mathbb{Z}} \varepsilon_{\tilde{Q}_j} \right) \) of iid copies of \( \sum_{j \in \mathbb{Z}} \varepsilon_{\tilde{Q}_j} \). We observe that \( |\tilde{Q}_j| \leq 1 \) a.s. and \( |Q_{T^*}| = 1 \) 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 \( (\tilde{Q}_j)_{j \in \mathbb{Z}} \) 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:

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

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

**Remark 3.11.** We observe that

\[
E[|Q_{T^*}|^\alpha] = \mathbb{E} \left[ \max_{i \in \mathbb{Z}} |\Theta_i|^{\alpha} \right] = E_{\sum_{j \in \mathbb{Z}} |\Theta_j|^{\alpha}} = \theta_{|X|}.
\]
Since \( \theta_{|X|} = \mathbb{P}(T^* = 0) \) the extremal index \( \theta_{|X|} \) has the 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 \( E[A_t] = 1 \) 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 \( \theta_{|X|} \) in terms of the Riemann zeta function \( \zeta \); see Example 3.13. A first order approximation to this formula is given by

\[
\theta_{|X|} \approx \frac{1}{2} \exp\left(\frac{\zeta(0.5)}{\sqrt{2\pi}}\right) \approx \frac{1}{2} \exp(-0.5826) \approx 0.2792.
\]  

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

It has unit Fréchet marginals \( \mathbb{P}(X_t \leq x) = \Phi_1(x) = e^{-x^{-1}}, \quad x > 0 \). Any discretization \( X^{(d)} = (X^{(d)}_t)_{t \in \mathbb{Z}} \) for \( d > 0 \) is regularly varying with index 1 and spectral tail process \( \Theta^{(d)}_t = e^{\sqrt{2} B_t^{(d)} - d|t|}, t \in \mathbb{Z} \). Direct calculation of

\[
-x \log \mathbb{P}(n^{-1} \max_{1 \leq t \leq n} X^{(d)}_t \leq x), \quad x > 0,
\]

yields the extremal index of \( X^{(d)} \) as the limit

\[
\theta^{(d)}_{|X|} = \lim_{n \to \infty} n^{-1} E\left[ \sup_{0 \leq t \leq n} e^{\sqrt{2} B_t^{(d)} - d t} \right].
\]  

We use the expression \( \theta^{(d)}_{|X|} = \mathbb{P}(T^{*(d)} = 0) \) where \( T^{*(d)} \) is the first record time of \( (\Theta^{(d)}_t)_{t \in \mathbb{Z}} \); see (3.8). We consider the first ladder height epoch \( \tau_+(d) = \inf\{t \geq 1 : \sqrt{2} B_t^{(d)} + d t < 0\} \). Using the symmetry of the Gaussian distribution, \( (\Theta^{(d)}_t)_{t \geq 1} \overset{d}{=} (1/\Theta^{(d)}_{-t})_{t \geq 1} \), we obtain

\[
\theta^{(d)}_{|X|} = \mathbb{P}(T^{*(d)} = 0) = \mathbb{P}(\tau_+(d) = \infty)^2.
\]

Combining this with the classical identity \( \mathbb{P}(\tau_+(d) = \infty) = 1/E[\tau_-(d)] \) for \( \tau_-(d) = \inf\{t \geq 1 : \sqrt{2} B_t^{(d)} - d t \leq 0\} \), from random walk
theory (see [1]) we get
\[
\theta_X^{(\delta)} = \left( \frac{1}{\mathbb{E}[-\tau(\delta)]} \right)^2 = \left( \frac{\mathbb{E}[B_\delta - \delta]}{\mathbb{E}[\sqrt{2}B_{\tau(\delta)} - \tau(\delta)]} \right)^2 = \delta^2 (\mathbb{E}[\sqrt{2}B_{\tau(\delta)} + \tau(\delta)])^{-2},
\]
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
\[
\mathbb{E}[(\sqrt{2}B_\delta t + \tau(\delta))] = \sqrt{\delta} \exp \left( \frac{-\sqrt{\delta}}{2\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{\zeta(1/2 - n)}{n!(2n+1)} \left( -\frac{\delta}{4} \right)^n \right).
\]
This implies that
\[
\theta_X^{(\delta)} = \delta \exp \left( \sqrt{\frac{\delta}{\pi}} \sum_{n=0}^{\infty} \frac{\zeta(1/2 - n)}{n!(2n+1)} \left( -\frac{\delta}{4} \right)^n \right).
\]
We recover the Pickands constant of the Brown-Resnick process (see [29]) as the limit
\[
\mathcal{H}_X^{(0)} = \lim_{T \to \infty} \frac{1}{T} \mathbb{E} \left[ \sup_{0 \leq t \leq T} e^{\sqrt{2}B_{\tau(\delta)} t} \right] = 1.
\]

**Proof of Proposition 3.10.** Consider the supremum of all points of the limit process \( N \) in Theorem 3.8:
\[
M = \sup_{i \geq 1} \Gamma_i^{-1/\alpha} \sup_{j \in \mathbb{Z}} |Q_{ij}|.
\]
The sequences \((\Gamma_i)\) and \((Q^{(i)})\) are independent and \( M = \sup_{i \geq 1} \Gamma_i^{-1/\alpha} V_i \) for the iid sequence \( V_i : = \sup_{j \in \mathbb{Z}} |Q_{ij}|, i = 1, 2, \ldots \), whose generic element \( V \) has the property \( \mathbb{E}[V^\alpha] < \infty \). Indeed, \( V \leq 1 \) 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 \( \mu((x, \infty) \times [0, y]) = x^{-\alpha} F_V(y), x > 0, y \geq 0 \), where \( F_V \) is the distribution function of \( V \). For \( x > 0 \) we consider \( B_x = \{(y, v) \in E : yv > x\} \). We observe that
\[
\mu(B_x) = \int_0^\infty \int_0^y \alpha y^{-\alpha-1} F_V(dv) = \int_0^\infty (x/v)^{-\alpha} F_V(dv) = x^{-\alpha} \mathbb{E}[V^\alpha].
\]
Therefore we have for $x > 0$,
\[
\begin{align*}
\mathbb{P}(M \leq x) &= \mathbb{P}(\Gamma_i^{-1/\alpha} V_i \leq x, i \geq 1) \\
&= P(N_r, V(B_x) = 0) \\
&= e^{-\mu(B_x)} = e^{-x^{-\alpha} E[V^\alpha]}. 
\end{align*}
\]
Thus $M$ is a scaled version of the standard Fréchet distribution, $\Phi_\alpha(x) = e^{-x^{-\alpha}}, x > 0$:
\[
\mathbb{P}(M \leq x) = \Phi_\alpha(x), \quad x > 0.
\]
On the other hand, Theorem 3.8 and an application of the continuous mapping theorem yield as $n \to \infty$,
\[
\begin{align*}
\mathbb{P}(a_n^{-1} M_n \leq x) &= \mathbb{P}(N_n(x, \infty) = 0) \\
&\to \mathbb{P}(N(x, \infty) = 0) \\
&= \mathbb{P}(M \leq x), \quad x > 0.
\end{align*}
\]
In view of the definition of the extremal index of the sequence $(|X_t|)$ we can identify
\[
E[V^\alpha] = E\left[\sup_{j \in \mathbb{Z}} |Q_j|^\alpha\right] = E[|Q_{T^*}|^\alpha].
\]
as the value $\theta_{|X|}$. This proves the first part of (3.8). The identity
\[
E[|Q_{T^*}|^\alpha] = \mathbb{P}(Y |Q_{T^*}| > 1) = \mathbb{P}(|Q_{T^*}|^\alpha > Y^{-\alpha})
\]
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
\[
\theta_{|X|} = E[|Q_{T^*}|^\alpha] = \sum_{k \in \mathbb{Z}} E\left[\sum_{j \in \mathbb{Z}} |\Theta_j|^\alpha \mathbb{I}(T^* = k)\right] = \sum_{k \in \mathbb{Z}} E\left[\sum_{j \in \mathbb{Z}} |\Theta_{j-k}|^{-\alpha} \mathbb{I}(T^* = 0)\right] = \mathbb{P}(T^* = 0).
\]
This proves the last identity in (3.8). \qed
§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 will consider some standard estimators of $\theta_X$. For the sake of argument we assume that $(X_t)$ is a non-negative 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 $\theta_X$ 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 $\hat{\theta}_{bl}$ is given by the ratio of the number $K_n(u)$ of such clusters and the total number of exceedences $N_n(u)$:

$$\hat{\theta}_{bl}(r) = \frac{K_n(u)}{N_n(u)} := \frac{\sum_{t=1}^{k_n} \mathbb{1}(M_{(t-1)r+1, tr} > u)}{\sum_{t=1}^{n} \mathbb{1}(X_t > u)}.$$  \hfill (4.11)

This method requires the choice of block length $r$ and threshold level $u$ satisfying $r_n F(u_n) \to 0$; if $r_n \to \infty$ does not hold at the prescribed rate $\hat{\theta}_{bl}$ 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]:

$$\hat{\theta}_{dbl} = \frac{\log(1 - K_n(u)/k_n)}{r \log(1 - N_n(u)/n)}.$$  

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

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

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

respectively, we obtain
\[ \theta_X \approx \frac{\log P(M_n \leq u_n)}{n \log F(u_n)} = \frac{\log(1 - P(M_n > u_n))}{n \log(1 - F(u_n))} \]
\[ \approx \frac{\log(1 - K_n(u)/k_n)}{r_n \log(1 - N_n(u)/n)} = \hat{\theta}^{dbl}. \]

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 \( \hat{\theta}^{dbl} \approx \hat{\theta}^{bl} \). [38] showed that \( \hat{\theta}^{dbl} \) has a smaller asymptotic variance than \( \hat{\theta}^{bl} \). [33] proposed a sliding blocks version of \( \hat{\theta}^{dbl} \) with an even smaller asymptotic variance.

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 \( r_n F(u_n) \to 1 \), and \( l_n \leq r_n \).

Consider the exceedance times:
\[ S_0(u) = 0, \quad S_i(u) = \min\{t > S_{i-1}(u) : X_t > u_n\}, \quad i \geq 1, \]
with inter-exceedance times \( T_i(u) = S_i(u) - S_{i-1}(u), \quad i \geq 1 \). The sequence \( (T_i(u))_{i \geq 2} \) constitutes a stationary sequence. If \( r_n F(u_n) \to 1 \), [12] noticed that \( (n T_2(u)) \) converges in distribution to a limiting mixture given by
\[ (1 - \theta_X) 1_{x} + \theta_X (1 - e^{-\theta_X x}), \quad x \geq 0. \]
Calculation yields to the coefficient of variation \( \nu \) of \( T_2(u) \) whose square is given by
\[ \nu^2 = \text{var}(T_2(u))/[\text{E}[T_2(u)]]^2 = \text{E}[T_2^2(u)]/([\text{E}[T_2(u)]]^2 - 1 = 2/\theta_X - 1, \]
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

\[ \hat{\theta}^{\text{int}}(u) = 1 \wedge \frac{2 \left( \sum_{i=2}^{N_n(u)} (T_i(u)-1) \right)^2}{(N_n(u)-1) \sum_{i=2}^{N_n(u)} (T_i(u)-1)(T_i(u)-2)} . \] (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 \(x > 0\),

\[ \mathbb{P} \left( -r_n \log F(M_r) > x \right) = \mathbb{P} \left( F(M_r) \leq e^{-x/r} \right) = \mathbb{P} \left( \max_{i=1, \ldots, r_n} F(X_i) \leq e^{-x/r} \right) = (\mathbb{P}(F(X) \leq e^{-x/r}))^r = e^{-x} . \]

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

\[ \mathbb{P} \left( -r_n \log F(M_r) > x \right) = \mathbb{P} \left( \max_{i=1, \ldots, r_n} F(X_i) < e^{-x/r} \right) \rightarrow e^{-\theta_X x} , \quad x > 0 . \]

Thus the \((-r_n \log F(M_r))\) are asymptotically \(\text{Exp}(\theta_X)\) distributed. For iid \(\text{Exp}(\theta_X)\) the maximum likelihood estimator of \(\theta_X\) is given by the reciprocal of the sample mean. These ideas lead to Northrop’s estimators [26]. Mimicking the maximum likelihood estimator of iid \(\text{Exp}(\theta_X)\) 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 \(\theta_X\):

\[ \hat{\theta}^{\text{Nsl}}(r) = \left( \frac{1}{n-r+1} \sum_{t=1}^{n-r+1} (-r \log F_n(M_{t,t+r})) \right)^{-1} , \] (4.15)

\[ \hat{\theta}^{\text{Ndbl}}(r) = \left( \frac{1}{[n/r]} \sum_{i=1}^{[n/r]} (-r \log F_n(M_{r(i-1)+1,r i})) \right)^{-1} . \] (4.16)

Here \(F_n\) is the 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 non-negative regularly varying process \((X_t)\) with index \(\alpha > 0\), spectral tail process \((\Theta_t)\) 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 the spectral cluster process of \((X_t)\). We will construct an estimator based on this identity.

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

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

The following limit relation is proved in [7]:

\[
\lim_{n \to \infty} \mathbb{E}[M_{1,r}^{\alpha}/S_{1,r}^{(\alpha)} \mid S_{r}^{(\alpha)} > a_n^\alpha] = \mathbb{E}[Q_T^\alpha], \tag{4.17}
\]

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

\[
\hat{\theta}_v^{\text{sc}}(r) := \frac{\sum_{i=1}^{k_n} \frac{M_{i,r}^{\alpha}}{S_{i,r}^{(\alpha)}} \mathbb{I}(S_{i,r}^{(\alpha)} > v)}{\sum_{i=1}^{k_n} \mathbb{I}(S_{i,r}^{(\alpha)} > v)}.
\] \tag{4.18}

Here we choose \(v = S_{s(r),r}^{(\alpha)}\), the \(s\)th largest among \(S_{i,r}^{(\alpha)}\) 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 1000 simulated processes \((X_t)_{t=1,\ldots,5000}\) 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 \(\theta_{|X|} \approx 0.8\). 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.

- \(\hat{\theta}^{\text{bl}}\) and \(\hat{\theta}^{\text{uns}}\) are functions of the block and run lengths, respectively. \(u\) is the largest \(\lfloor n^{0.6} \rfloor\)th upper order statistic of the sample.
Figure 5.1. Boxplots based on 1000 simulations for the estimation of $\theta |X| = 0.8$ in the AR(1) model with $\varphi = 0.2$ and iid student(1) noise.

- $\hat{\theta}_{\text{slbl}}$ is a function of $r$. $u$ is the $r$th upper order statistic of the sample.
- $\hat{\theta}_{\text{int}}$ is a function of $x$. $u$ is the $[n/x]$th upper order statistic.
- $\hat{\theta}_{\text{Nsl}}, \hat{\theta}_{\text{scp}}$ are functions of $r$.
- For $\hat{\theta}_{\text{scp}}$ we choose $s = [n^{0.6}/r]$. The tail index $\alpha$ 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 $\hat{\theta}_{\text{Nsl}}$ outperforms the classical estimators (runs, blocks); it has smallest variance but it may be difficult to control its bias. Our experience with $\hat{\theta}_{\text{scp}}$ shows that it performs better than the other estimators as regards the bias, especially when $\theta_X$ is small. The intervals estimator $\hat{\theta}_{\text{int}}$ is preferred.
by practitioners because the choice of the hyperparameter $x$ is robust with respect to different values of $\theta_X$. This cannot be said about the other estimators with the exception of $\hat{\theta}_{scp}$. In our experiments with sample size $n = 5000$, the choices $x = 32$ and $r = 64$ work well for $\hat{\theta}_{int}$ and $\hat{\theta}_{scp}$, respectively. We did not fine-tune the hyperparameter $s$ in $\hat{\theta}_{scp}$ in our experiments.

References

1. S. Asmussen, *Applied Probability and Queues*, 2nd edition. Springer, New York, 2003.
2. B. Basrak, J. Segers, *Regularly varying multivariate time series*. — Stoch. Proc. Appl. 119 (2009), 1055–1080.
3. B. Berghaus, A. B"ucher, Weak convergence of a pseudo maximum likelihood estimator for the extremal index. — Ann. Stat. 46 (2018), 2307–2335.
4. B. M. Brown, S. I. Resnick, Extreme values of independent stochastic processes. — J. Appl. Probab. 14 (1977), 732–739.
5. A. B"ucher, T. Jennessen, Method of moments estimators for the extremal index of a stationary time series. — Electr. J. Stat. 14 (2020), 3103–3156.
6. D. Buraczewski, E. Damek, T. Mikosch, Stochastic Models with Power-Laws. The Equation $X = AX + B$, Springer, New York, 2016.
7. G. Buritica, T. Mikosch, O. Wintenberger. Threshold selection for cluster inference based on large deviation principles. Technical report, 2021.
8. J. T. Chang, Y. Peres, Ladder heights, Gaussian random walks and the Riemann zero function. — Ann. Probab. 25 (1997), 787–802.
9. R. A. Davis, T. Hsing, Point process and partial sum convergence for weakly dependent random variables with infinite variance. — Ann. Probab. 23 (1995), 879–917.
10. R. A. Davis, S. I. Resnick, Limit theory for moving averages of random variables with regularly varying tail probabilities. — Ann. Probab. 13 (1985), 179–195.
11. A. C. Davison, R. L. Smith, Models for exceedances over high thresholds (with discussion). — J. Royal Statist. Soc., Ser. B, 52 (1990), 393–442.
12. C. A. Ferro, J. Segers, Inference for clusters of extreme values. — J. Royal Statist. Soc., Ser. B, 65 (2003), 545–556.
13. C. M. Goldie, Implicit renewal theory and tails of solutions of random equations. — Ann. Appl. Probab. 1 (1991), 126–166.
14. L. de Haan, C. Mercadier, C. Zhou, Adapting extreme value statistics to financial time series: dealing with bias and serial dependence. — Finance and Stochastics 20 (2016), 321–354.
15. L. de Haan, A spectral representation for max-stable processes. — Ann. Probab. 12 (1984), 1194–1204.
16. L. de Haan, S. I. Resnick, H. Rootzén, C.G. de Vries, Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH processes. — Stoch. Proc. Appl. 32 (1989), 213–224.
17. T. Hsing, Extremal index estimation for a weakly dependent stationary sequence. — Ann. Stat. (1993), 2043–2071.
18. Z. Kabluchko, Spectral representations of sum- and max-stable processes. — Extremes 12 (2009), 401–424.
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 (1973), 267–248.
21. R. Kulik, 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 (1983), 291–306.
23. M. R. Leadbetter, G. Lindgren, 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 (1965), 993–999.
25. G. F. Newell, Asymptotic extremes for $m$-dependent random variables. — Ann. Math. Statist. 35 (1964), 1322–1325.
26. P. J. Northrop, *An efficient semiparametric maxima estimator of the extremal index*. — Extremes 18 (2015), 585–603.
27. G. L. O’Brien, *Limit theorems for the maximum term of a stationary process*. — Ann. Probab. 2 (1974), 540–545.
28. G. L. O’Brien, *Extreme values for stationary and Markov sequences*. — Ann. Probab. 15 (1987), 281–291.
29. J. Pickands, *Asymptotic properties of the maximum in a stationary Gaussian process*. — Trans. Amer. Math. Soc. 145 (1969), 75–86.
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 (2009), 271–310.
33. C. Y. Robert, C. A. Ferro, J. Segers, *A sliding blocks estimator for the extremal index*. — Electr. J. Stat. 3 (2009), 993–1020.
34. J. Segers, *Approximate distributions of clusters of extremes*. — Stat. Probab. Lett. 74 (2005), 330–336.
35. M. Ševčíková, *Likelihood estimation of the extremal index*. — Extremes 10 (2007), 41–55.
36. M. Ševčíková, A. C. Davison, *Model misspecification in peaks over threshold analysis*. — Ann. Appl. Stat. 4 (2010), 203–221.
37. R. L. Smith, *Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone*. — Statist. Sci. 4 (1989), 367–377.
38. R. L. Smith, I. Weissman, *Estimating the extremal index*. — J. Royal Statist. Soc., Ser. B 56 (1994), 515–528.

LPSM, Sorbonne Universités
UPMC Université Paris 06,
F-75005, Paris, France
E-mail: gloria.buritica@sorbonne-universite.fr

Department of Mathematics,
University of Copenhagen,
DK-2100 Copenhagen, Denmark
E-mail: meyer@math.ku.dk
E-mail: mikosch@math.ku.dk

LPSM, Sorbonne Universités
UPMC Université Paris 06,
F-75005, Paris, France
E-mail: olivier.wintenberger@upmc.fr