



Extremal clustering in non-stationary random sequences

Graeme Auld¹ · Ioannis Papastathopoulos¹

Received: 25 October 2019 / Revised: 21 April 2021 / Accepted: 22 April 2021 /
Published online: 12 May 2021
© The Author(s) 2021

Abstract

It is well known that the distribution of extreme values of strictly stationary sequences differ from those of independent and identically distributed sequences in that extremal clustering may occur. Here we consider non-stationary but identically distributed sequences of random variables subject to suitable long range dependence restrictions. We find that the limiting distribution of appropriately normalized sample maxima depends on a parameter that measures the average extremal clustering of the sequence. Based on this new representation we derive the asymptotic distribution for the time between consecutive extreme observations and construct moment and likelihood based estimators for measures of extremal clustering. We specialize our results to random sequences with periodic dependence structure.

Keywords Clustering of extremes · Extremal index · Interexceedance times · Intervals estimator · Non-stationary sequences · Periodic processes

AMS 2000 Subject Classifications 60G70

1 Introduction

Extreme value theory for strictly stationary sequences has been extensively studied, initiated in the works of Watson (1954), Berman (1964), and Loynes (1965), and continued by Leadbetter (1974, 1983) and O'Brien (1987) amongst others. One of the key findings in this line of research is that unlike in independent and

✉ Graeme Auld
G.R.Auld@sms.ed.ac.uk

Ioannis Papastathopoulos
i.papastathopoulos@ed.ac.uk

¹ School of Mathematics and Maxwell Institute, University of Edinburgh,
Edinburgh, EH9 3FD, UK

identically distributed sequences where extreme values tend to occur in isolation, stationary sequences possess an intrinsic potential for clustering of extremes, i.e., several successive or close extreme values may be observed. Understanding the extremal clustering characteristics of a stochastic process is critical in many applications where a cluster of extreme values may have serious consequences. For example, if a sequence consists of daily temperatures at some fixed location then a cluster of extremes may correspond to a heatwave.

The extent to which extremal clustering may occur is naturally measured, for strictly stationary sequences, by a parameter known as the extremal index. Let $\{X_n\}_{n=1}^\infty$ be a sequence of random variables with common marginal distribution function F , and let $\bar{F} = 1 - F$ and $M_n = \max\{X_1, \dots, X_n\}$. Also, let $\{x_n\}_{n=1}^\infty$ be a sequence of real numbers that we may informally think of as thresholds or levels. In the special case that X_i and X_j are independent, $i \neq j$, then a necessary and sufficient condition for $\mathbb{P}(M_n \leq x_n)$ to converge to a limit in $(0, 1)$ as $n \rightarrow \infty$ is that $n\bar{F}(x_n) \rightarrow \tau > 0$, in which case $\mathbb{P}(M_n \leq x_n) \rightarrow e^{-\tau}$ (Leadbetter et al. 1983, Theorem 1.5.1). More generally, if $\{X_n\}_{n=1}^\infty$ is a strictly stationary sequence, then $n\bar{F}(x_n) \rightarrow \tau$ is not sufficient to ensure the convergence of $\mathbb{P}(M_n \leq x_n)$. However, in most cases of practical interest, provided that a suitable long range dependence restriction is satisfied, such as condition D of Leadbetter (1974), one has $\mathbb{P}(M_n \leq x_n) \rightarrow e^{-\theta\tau}$ where $\theta \in [0, 1]$ is the extremal index. Leadbetter (1983) showed that exceedances of the level x_n occur in clusters with the limiting mean cluster size being equal to θ^{-1} , and Hsing (1987) showed that distinct clusters may be considered independent in the limit.

Another characterization of θ that links it to the extremal clustering properties of a strictly stationary sequence can be found in O'Brien (1987). Defining $M_{j,k} = \max\{X_i : j + 1 \leq i \leq k\}$, it was shown that the distribution function of M_n satisfies

$$\mathbb{P}(M_n \leq x_n) - F(x_n)^{n\theta_n} \rightarrow 0, \quad \text{as } n \rightarrow \infty, \quad (1)$$

where

$$\theta_n = \mathbb{P}(M_{1,p_n} \leq x_n \mid X_1 > x_n), \quad (2)$$

for some $p_n = o(n)$, and provided the limit exists, $\theta_n \rightarrow \theta$ as $n \rightarrow \infty$. This result illustrates that smaller values of θ are indicative of a larger degree of extremal clustering, since the conditional probability in Eq. 2 is small when an exceedance of a large threshold is likely to soon be followed by another exceedance.

Early attempts at estimating θ were based on associating θ^{-1} with the limiting mean cluster size. Different methods for identifying clusters gave rise to different estimators, well known examples being the runs and blocks estimators (Smith and Weissman 1994). For the runs estimator, a cluster is identified as being initialized when a large threshold is exceeded and ends when a fixed number, known as the run length, of non-exceedances occur. The extremal index is then estimated by the ratio of the number of identified clusters to the total number of exceedances. A difficulty that arises when using this estimator is its sensitivity to the choice of run length (Hsing 1991).

The problem of cluster identification was studied by Ferro and Segers (2003) who considered the distribution of the time between two exceedances of a large threshold.

They found that the limiting distribution of appropriately normalized interexceedance times converges to a distribution that is indexed by θ . In particular, for a given threshold $u \in \mathbb{R}$, they define the random variable $T(u) = \min\{n \geq 1 : X_{n+1} > u \mid X_1 > u\}$, and found that as $n \rightarrow \infty$, $\bar{F}(x_n)T(x_n)$ converges in distribution to a mixture of a point mass at zero and an exponential distribution with mean θ^{-1} . Thus, by computing theoretical moments of this limiting distribution and comparing them with their empirical counterparts, they construct their so-called intervals estimator.

Motivated by the fact that many real world processes are non-stationary, in this paper we investigate the effect of non-stationarity on extremal clustering. Previous statistical works that consider extremal clustering in non-stationary sequences include (Süveges 2007), who used the likelihood function introduced by Ferro and Segers (2003) for the extremal index together with smoothing methods to capture non-stationarity in a time series of temperature measurements. In a similar application, Coles et al. (1994) used a Markov model together with simulation techniques to estimate the extremal index within different months.

An early work that developed extreme value theory for non-stationary sequences with a common marginal distribution is Hüsler (1983), which focused on the asymptotic distribution of the sample maxima but did not consider extremal clustering. Hüsler (1986) considered the more general case where the margins may differ and also discussed the difficulty of defining the extremal index for general non-stationary sequences.

Here, we consider a sequence of random variables $\{X_n\}_{n=1}^\infty$ with common marginal distribution function F , but do not assume stationarity in either the weak or strict sense. As we assume common margins, non-stationarity may arise through changes in the dependence structure. We show, under assumptions similar to O'Brien (1987), that

$$\mathbb{P}(M_n \leq x_n) - F(x_n)^{n\gamma_n} \rightarrow 0, \quad \text{as } n \rightarrow \infty, \tag{3}$$

where

$$\gamma_n = \frac{1}{n} \sum_{j=1}^n \mathbb{P}(M_{j,j+p_n} \leq x_n \mid X_j > x_n). \tag{4}$$

Thus, we find that the limiting distribution of the sample maximum at large thresholds is characterized by a parameter $\gamma = \lim_{n \rightarrow \infty} \gamma_n$, provided the limit exists, which by analogy with Eq. 2, may be regarded as the average of local extremal indices. In this paper we develop methods for estimating these local extremal indices by adapting the methods of Ferro and Segers (2003) for the extremal index to our non-stationary setting. In the special case that the sequence is stationary, so that all terms in the summation (4) are equal, the formula for γ_n reduces to θ_n in Eq. 2.

The structure of the paper is as follows. Section 2 defines the notation and assumed mixing condition used throughout the paper and states the main theoretical results regarding the asymptotic distribution of the sample maxima and normalized interexceedance times. Section 3 discusses approaches to parameter estimation using the result from Section 2 on the distribution of the interexceedance times. Section 4 considers the estimation problem for two simple non-stationary Markov sequences with

periodic dependence structures and Section 5 gives the proofs of the main theoretical results.

2 Theoretical results

2.1 Notation, definitions and preliminary results

Throughout the paper, when not explicitly stated otherwise, all limits should be interpreted as “as $n \rightarrow \infty$ ”. We assume that all random variables in the sequence $\{X_n\}_{n=1}^\infty$ have common marginal distribution F with upper endpoint $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$, though we do not assume stationarity. In addition to the definitions for M_n and $M_{j,k}$ given in the Section 1, we define $M(A) = \max\{X_i : i \in A\}$ where A is an arbitrary set of positive integers, and write $|A|$ for the number of elements in A . We also refer to a set of consecutive integers as an interval. If I_1 and I_2 are two intervals, we say that I_1 and I_2 are separated by q if $\min(I_2) - \max(I_1) = q + 1$ or $\min(I_1) - \max(I_2) = q + 1$, i.e., there are q intermediate values between I_1 and I_2 . The set $\{1, 2, 3, \dots\}$ is denoted by \mathbb{N} . Equality in distribution of two random variables X and Y is denoted by $X \stackrel{D}{=} Y$.

We assume that the sequence $\{X_n\}_{n=1}^\infty$ satisfies the asymptotic independence of maxima (AIM) mixing condition of O’Brien (1987) which restricts long range dependence.

Definition 1 The sequence $\{X_n\}_{n=1}^\infty$ is said to satisfy the asymptotic independence of maxima condition relative to the sequence x_n of real numbers, abbreviated to “ $\{X_n\}_{n=1}^\infty$ satisfies AIM(x_n)”, if there exists a sequence q_n of positive integers with $q_n = o(n)$ such that for any two intervals $I_1 = \{i_1, \dots, i_j\}$ and $I_2 = \{i_j + q_n + 1, \dots, i_j + q_n + k\}$ separated by q_n , we have

$$\alpha_n = \max | \mathbb{P}(M(I_1 \cup I_2) \leq x_n) - \mathbb{P}(M(I_1) \leq x_n)\mathbb{P}(M(I_2) \leq x_n) | \rightarrow 0, \quad (5)$$

where the maximum is taken over all positive integers i_1, i_j and k such that $|I_1| \geq q_n$, $|I_2| \geq q_n$ and $i_j + q_n + k \leq n$.

Definition 1 states a slightly weaker condition than the widely used $D(x_n)$ condition (Leadbetter 1983) in that only certain intervals I_1 and I_2 need to be considered in Eq. 5 rather than arbitrary sets of integers, so that all examples in the literature of sequences satisfying $D(x_n)$ also satisfy AIM(x_n). For example, stationary Gaussian sequences with autocorrelation function ρ_n satisfying Berman’s condition, $\rho_n \log n \rightarrow 0$ (Berman 1964), satisfy AIM(x_n) for any sequence x_n such that $n\bar{F}(x_n)$ is bounded and any $q_n = o(n)$ (Leadbetter et al. 1983, Lemma 4.4.1). The analogous result for non-stationary Gaussian sequences is given in Hüsler (1983), where Berman’s condition is replaced by $r_n \log n \rightarrow 0$ with $r_n = \sup\{|\rho(i, j)| : |i - j| \geq n\}$ and $\rho(i, j)$ the correlation between X_i and X_j .

O'Brien (1987) showed that if $\{X_n\}_{n=1}^\infty$ is a stationary positive Harris Markov sequence with separable state space S and $f : S \rightarrow \mathbb{R}$ is a measurable function then the sequence $Y_n = f(X_n)$ satisfies $\text{AIM}(x_n)$ for any x_n and $q_n = o(n)$ with $q_n \rightarrow \infty$.

We note that Definition 1 states a property of the dependence structure of the sequence $\{X_n\}_{n=1}^\infty$, with the specific marginal distributions playing essentially no role. In particular, if $\{X_n\}_{n=1}^\infty$ satisfies $\text{AIM}(x_n)$ and $g : \mathbb{R} \rightarrow \mathbb{R}$ is a monotone increasing function then $Y_n = g(X_n)$ satisfies $\text{AIM}(g(x_n))$ with the same q_n .

The assumption that $\{X_n\}_{n=1}^\infty$ satisfies $\text{AIM}(x_n)$ ensures the approximate independence of the block maxima of two sufficiently separated blocks. Lemma 1 below provides an upper bound for the degree of dependence of k block maxima for suitably separated blocks and will be useful in Section 2.2 when the limiting behaviour of $\mathbb{P}(M_n \leq x_n)$ is considered.

Lemma 1 *Let $\{X_n\}_{n=1}^\infty$ satisfy $\text{AIM}(x_n)$ and let I_1, I_2, \dots, I_k be distinct subintervals of $\{1, 2, \dots, n\}$ where $k \geq 2$ and $|I_i| \geq q_n, 1 \leq i \leq k$. Suppose that I_i and I_{i+1} are separated by q_n for $1 \leq i \leq k - 1$. Then*

$$\left| \mathbb{P}(M(\cup_{i=1}^k I_i) \leq x_n) - \prod_{i=1}^k \mathbb{P}(M(I_i) \leq x_n) \right| \leq (k - 1)\alpha_n + 2(k - 2)q_n \bar{F}(x_n). \tag{6}$$

2.2 Asymptotic distribution of M_n

In this section we investigate the limiting behaviour of $\mathbb{P}(M_n \leq x_n)$, with the main result being Theorem 1. In addition to assuming that $\{X_n\}_{n=1}^\infty$ satisfies $\text{AIM}(x_n)$, we will assume that the rate of growth of the sequence x_n is controlled via

$$n \bar{F}(x_n) \rightarrow \tau > 0. \tag{7}$$

In the case of continuous marginal distributions, Eq. 7 is immediately satisfied by $x_n = F^{-1}(1 - \tau/n)$. More generally, Theorem 1.7.13 of Leadbetter et al. (1983) guarantees the existence of a sequence x_n satisfying (7) when F is in the domain of attraction of any of the three classical extreme value distributions (Haan and Ferreira 2006, Section 1.2).

We use the standard technique of block-clipping, see for example Section 10.2.1 in Beirlant et al. (2004), to split the interval $\{1, 2, \dots, n\}$ into subintervals, or blocks, of alternating large and small lengths. Specifically, for sequences p_n and q_n such that $q_n = o(p_n)$ and $p_n = o(n)$ we define

$$\begin{aligned} A_i &= \{(i - 1)(p_n + q_n) + 1, \dots, ip_n + (i - 1)q_n\} \\ A_i^* &= \{ip_n + (i - 1)q_n + 1, \dots, i(p_n + q_n)\}, \end{aligned} \tag{8}$$

for $i = 1, 2, \dots, r_n$, where $r_n = \lfloor n/(p_n + q_n) \rfloor$.

If we take the sequence q_n appearing in the construction of the blocks A_i and A_i^* to be the same as that in Definition 1, then Lemma 1 bounds the degree of dependence of the collection of random variables $\{M(A_i)\}_{i=1}^{r_n}$, and this allows us to prove Lemma 2 below which modifies Lemma 3.1 from O'Brien (1987) to allow for non-stationarity.

Lemma 2 Let $\{X_n\}_{n=1}^\infty$ satisfy $AIM(x_n)$ and let the sequence p_n be such that

$$p_n = o(n), \quad n\alpha_n = o(p_n) \quad \text{and} \quad q_n = o(p_n). \tag{9}$$

Then if Eq. 7 holds, we have

$$\mathbb{P}(M_n \leq x_n) - \prod_{i=1}^{r_n} \mathbb{P}(M(A_i) \leq x_n) \rightarrow 0, \tag{10}$$

where the intervals $\{A_i\}_{i=1}^{r_n}$ are as in Eq. 8.

Remarks Equation 10 follows easily from Eq. 6 by making the identification $k = r_n$ and using Eqs. 7 and 9. Additionally, if $\{X_n\}_{n=1}^\infty$ satisfies $AIM(x_n)$ then we can always find a sequence p_n such that Eq. 9 holds, for example, by taking $p_n = \lfloor \{n \max(q_n, n\alpha_n)\}^{1/2} \rfloor$. Thus the only assumption in Lemma 2 beyond common margins is that $\{X_n\}_{n=1}^\infty$ satisfies $AIM(x_n)$ for a sequence x_n satisfying (7).

We can now state our main theorem.

Theorem 1 Under the same assumptions as in Lemma 2, we have

$$\mathbb{P}(M_n \leq x_n) - \exp\left\{-\sum_{j=1}^n \mathbb{P}(X_j > x_n, M_{j,j+p_n} \leq x_n)\right\} \rightarrow 0, \tag{11}$$

and consequently

$$\mathbb{P}(M_n \leq x_n) - F(x_n)^{n\gamma_n} \rightarrow 0, \tag{12}$$

where

$$\gamma_n = \frac{1}{n} \sum_{j=1}^n \mathbb{P}(M_{j,j+p_n} \leq x_n \mid X_j > x_n). \tag{13}$$

As it was noted in Section 1, for independent sequences (7) implies that $\mathbb{P}(M_n \leq x_n) \rightarrow e^{-\tau}$. For a random sequence satisfying the conditions of Lemma 2, the following result gives a necessary and sufficient condition for the convergence of $\mathbb{P}(M_n \leq x_n)$.

Corollary 1 Under the same assumptions as in Lemma 2, $\mathbb{P}(M_n \leq x_n)$ converges if and only if $\lim_{n \rightarrow \infty} \gamma_n$ exists, where γ_n is as in Eq. 13, in which case $\mathbb{P}(M_n \leq x_n) \rightarrow e^{-\tau\gamma}$ with $\gamma = \lim_{n \rightarrow \infty} \gamma_n \in [0, 1]$.

Corollary 1 follows from Eq. 12 since $n\bar{F}(x_n) \rightarrow \tau$ if and only $F^n(x_n) \rightarrow e^{-\tau}$ which is easily seen by taking logs in the latter expression and using $\log(1 - t) = -t + o(t)$ as $t \rightarrow 0$.

A basic question regarding the constant γ appearing in Corollary 1 is whether it is independent of the particular value of τ in Eq. 7, i.e., do we obtain the same limiting value of γ_n regardless of the specific sequence x_n and τ used in Eq. 7? We will see in Section 2.3 that for sequences with periodic dependence this is indeed the case, and Theorem 2 gives sufficient conditions for this to hold more generally.

We now turn our attention to the conditional probabilities appearing in the summation (13), which contain local information regarding the strength of extremal clustering in the sequence $\{X_n\}_{n=1}^\infty$.

Definition 2 Under the same assumptions as in Lemma 2, let $\{f_n\}_{n=1}^\infty$ be the sequence of functions defined on \mathbb{N} by

$$f_n(i) = \theta_{i,n} = \mathbb{P}(M_{i,i+p_n} \leq x_n \mid X_i > x_n), \quad i \in \mathbb{N}. \tag{14}$$

We define the extremal clustering function of $\{X_n\}_{n=1}^\infty$ to be the function $\theta : \mathbb{N} \rightarrow [0, 1]$ given by

$$\theta_i = \lim_{n \rightarrow \infty} f_n(i) \tag{15}$$

provided the limit exists.

In the special case that the sequence $\{X_n\}_{n=1}^\infty$ is stationary, the extremal clustering function is simply a constant function equal to the extremal index of the sequence. In the general case, if we think of the index i in X_i as denoting time, then we may regard θ_i as the extremal index at time i . The definition of θ_i entails pointwise convergence of the sequence of approximations $\{f_n\}_{n=1}^\infty$ in Eq. 14. When there is a uniformity in this convergence and the extremal clustering function is Cesàro summable we obtain the following result.

Theorem 2 Suppose $\{X_n\}_{n=1}^\infty$ satisfies AIM(x_n) with $n\bar{F}(x_n) \rightarrow \tau > 0$. Assume that $\{\theta_i\}_{i=1}^\infty$ is Cesàro summable and

$$\max_{1 \leq i \leq n} |\theta_i - \theta_{i,n}| \rightarrow 0 \quad \text{as } n \rightarrow \infty, \tag{16}$$

where $\theta_{i,n} = f_n(i)$ is as in Eq. 14. Then $\mathbb{P}(M_n \leq x_n) \rightarrow e^{-\tau\gamma}$ where

$$\gamma = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \theta_i. \tag{17}$$

Moreover, if $\{y_n\}_{n=1}^\infty$ is a sequence of real numbers such that $n\bar{F}(y_n) \rightarrow \tau'$ with $\tau' \leq \tau$ then $\mathbb{P}(M_n \leq y_n) \rightarrow e^{-\tau'\gamma}$ with γ as in Eq. 17.

As with the constant γ in Corollary 1, we may inquire as to whether the extremal clustering function is independent of the value of τ and sequence x_n used in Eq. 7. Although we do not attempt to answer this in full generality, we note that, as with the conditional probability formulation of the extremal index, for most sequences that are of practical interest, the formula defining θ_i may be reduced to a form that makes no explicit reference to the sequences x_n and p_n . For example, under the additional assumption due to Smith (1992) which requires that for any x_n in Theorem 1 we have

$$\lim_{p \rightarrow \infty} \lim_{n \rightarrow \infty} \sum_{k=p}^{p_n} \mathbb{P}(X_{i+k} > x_n \mid X_i > x_n) = 0, \tag{18}$$

for each i , then Eq. 15 reduces to

$$\theta_i = \lim_{p \rightarrow \infty} \lim_{x \rightarrow x_F} \mathbb{P}(M_{i,i+p} \leq x \mid X_i > x). \tag{19}$$

Another common assumption for statistical applications is the $D^{(k)}(x_n)$ condition of Chernick et al. (1991) which we define below in a slightly modified form for our non-stationary setting.

Definition 3 A sequence $\{X_n\}_{n=1}^\infty$ as in Theorem 1 is said to satisfy the $D^{(k)}(x_n)$ condition, where $k \in \mathbb{N}$, if

$$n \mathbb{P}(X_i > x_n, M_{i,i+k-1} \leq x_n, M_{i+k-1,i+p_n} > x_n) \rightarrow 0 \quad \text{as } n \rightarrow \infty \tag{20}$$

for each $i \in \mathbb{N}$. For the case $k = 1$, we define $M_{i,i} = -\infty$.

Note that it is assumed in Definition 3 that $\{X_n\}_{n=1}^\infty$ satisfies $\text{AIM}(x_n)$ in conjunction with Eq. 20. Whereas Eq. 5 limits the degree of long range dependence in the sequence, Eq. 20 is a local mixing condition that ensures that the probability of again exceeding the threshold x_n in a block of p_n observations, after dropping below it for $k - 1$ consecutive observations falls to zero sufficiently rapidly as $n \rightarrow \infty$. The case where $k = 1$ implies that in the limit, any exceedances of a high threshold occur in isolation and is implied in the stationary case by the $D'(x_n)$ condition of Leadbetter et al. (1983), Chapter 3. One might expect that a more natural condition in our non-stationary setting would be to replace the constant k in Eq. 20 by k_i to reflect possible variations in the strength of local dependence. However, when Eq. 20 holds for some particular k , then it also holds for any other k' with $k' > k$, and so provided that the sequence $\{k_i\}_{i=1}^\infty$ is bounded we may set $k = \max\{k_i : i \in \mathbb{N}\}$ and obtain (20) for each i . Thus the assumption of a single value of k in Definition 3 allows for variations in the strength of local dependence while at the same time restricting it to not persist too strongly to an arbitrary number of lags. If whenever x_n is a sequence as in Theorem 1 and the $D^{(k)}(x_n)$ condition holds then Eq. 15 reduces to

$$\theta_i = \lim_{x \rightarrow x_F} \mathbb{P}(M_{i,i+k-1} \leq x \mid X_i > x). \tag{21}$$

We will assume without further comment for the rest of the paper that the sequence $\{X_n\}_{n=1}^\infty$ has a well-defined extremal clustering function as may arise from assumptions (18) or (20).

2.3 Periodic dependence

In this section we assume that the sequence $\{X_n\}_{n=1}^\infty$ has a more refined structure than in the previous sections, namely that of periodic dependence, under which the results of Section 2.2 may be simplified considerably.

Definition 4 A sequence $\{X_n\}_{n=1}^\infty$ with common marginal distributions is said to have periodic dependence if there exists $d \in \mathbb{N}$ such that $(X_{t_1}, \dots, X_{t_k}) \stackrel{D}{=} (X_{t_1+d}, \dots, X_{t_k+d})$ for all $t_1, \dots, t_k \in \mathbb{N}$. The smallest d with this property is called the fundamental period.

Whereas for a strictly stationary sequence an arbitrary shift in time leaves the finite-dimensional distributions unchanged, for a sequence with periodic dependence only time shifts that are a multiple of the fundamental period leave finite-dimensional distributions unchanged. In particular, $M_{a,a+b} \stackrel{D}{=} M_{c,c+b}$ when $a \equiv c \pmod{d}$. Such sequences often mimic the dependence structure of certain environmental time series where we might expect a fundamental period of one year.

The following result concerning the convergence of $\mathbb{P}(M_n \leq x_n)$ shows that Theorem 3.7.1 of Leadbetter et al. (1983) for stationary sequences also holds for non-stationary sequences with periodic dependence.

Theorem 3 *Let $\{X_n\}_{n=1}^\infty$ have periodic dependence and satisfy the conditions of Lemma 2, with x_n satisfying (7) for some $\tau > 0$. Suppose that $y_n = \underline{y}_n(\tau')$ is a sequence of real numbers defined for each τ' with $0 < \tau' \leq \tau$ so that $n\bar{F}(y_n) \rightarrow \tau'$. Then there exist constants γ and γ' with $0 \leq \gamma \leq \gamma' \leq 1$ such that*

$$\begin{aligned} \limsup \mathbb{P}\{M_n \leq y_n(\tau')\} &= e^{-\tau'\gamma} \\ \liminf \mathbb{P}\{M_n \leq y_n(\tau')\} &= e^{-\tau'\gamma'} \end{aligned}$$

for all $0 < \tau' \leq \tau$. Hence if $\mathbb{P}\{M_n \leq y_n(\tau')\}$ converges for some τ' with $0 < \tau' \leq \tau$, then $\gamma = \gamma'$ and $\mathbb{P}\{M_n \leq y_n(\tau')\} \rightarrow e^{-\tau'\gamma}$ for all such τ' .

Although Theorem 3 makes no reference to the extremal clustering function, when $\mathbb{P}(M_n \leq x_n)$ converges, the constant γ in Theorem 3 is identified by Corollary 1 as $\gamma = \lim_{n \rightarrow \infty} \gamma_n$ with γ_n as in Eq. 13. Due to periodicity we obtain the simplified formula $\gamma = d^{-1} \sum_{i=1}^d \theta_i$, and the extremal clustering function is determined by the d values $\{\theta_i\}_{i=1}^d$ which repeat cyclically. Moreover, for sequences with periodic dependence, the convergence statement (16) can be strengthened to uniform convergence since $\sup_{i \in \mathbb{N}} |\theta_i - \theta_{i,n}| = \max_{1 \leq i \leq d} |\theta_i - \theta_{i,n}|$.

The following result is an immediate consequence of Theorem 3.

Corollary 2 *Let $\{X_n\}_{n=1}^\infty$ have periodic dependence with common marginal distribution function F . For each $\tau > 0$, let $x_n(\tau)$ be a sequence such that $n\bar{F}(x_n(\tau)) \rightarrow \tau$ and suppose that $\{X_n\}_{n=1}^\infty$ satisfies $AIM(x_n(\tau))$ for each such τ . If $\mathbb{P}\{M_n \leq x_n(\tau)\}$ converges for a single $\tau > 0$ then it converges for all $\tau > 0$, and in particular $\mathbb{P}\{M_n \leq x_n(\tau)\} \rightarrow e^{-\tau\gamma}$ for some $\gamma \in [0, 1]$.*

2.4 Interexceedance times

Ferro and Segers (2003) provided a method for estimating the extremal index of a stationary sequence without the need for identifying independent clusters of extremes. This was achieved by considering the distribution of the time between two exceedances of a threshold u , i.e.,

$$T(u) = \min\{n \geq 1 : X_{n+1} > u \mid X_1 > u\}, \tag{22}$$

as u approaches x_F . In particular, it was shown that the normalized interexceedance time $\bar{F}(x_n)T(x_n)$ converges in distribution as $n \rightarrow \infty$ to a mixture of a point mass

at zero, with probability $1 - \theta$, and an exponential random variable with mean θ^{-1} , with probability θ . The mixture arises from the fact that the interexceedance times can be classified in to two categories: within cluster and between cluster times. The mass at zero stems from the fact that the within cluster times, which tend to be small relative to the between cluster times, are dominated by the factor $\bar{F}(x_n)$.

In the stationary case, conditioning on the event $X_1 > u$ in Eq. 22 may be replaced with $X_i > u$, and X_{n+1} replaced by X_{n+i} , for any $i \in \mathbb{N}$, without affecting the distribution of $T(u)$. In the non-stationary case we consider for each $i \in \mathbb{N}$ and threshold u , the random variable $T_i(u)$ defined by

$$T_i(u) = \min\{n \geq 1 : X_{n+i} > u \mid X_i > u\}, \tag{23}$$

whose distribution in general depends on i . We find that the distribution of $\bar{F}(x_n)T_i(x_n)$ converges as $n \rightarrow \infty$ to a mixture of a mass at zero, with probability $1 - \theta_i$, and an exponential random variable with mean γ^{-1} , with probability θ_i . As in Ferro and Segers (2003), a slightly stronger mixing condition is required to derive this result than was needed for Theorem 1. We denote by $\mathcal{F}_{j_1, j_2}(u)$, the σ -algebra generated by the events $\{X_i > u : j_1 \leq i \leq j_2\}$, $j_1, j_2 \in \mathbb{N}$, and we define the mixing coefficients

$$\alpha_{n,q}^*(u) = \max_{1 \leq l \leq n-q} \sup |\mathbb{P}(E_2 \mid E_1) - \mathbb{P}(E_2)|, \tag{24}$$

where the supremum is over all $E_1 \in \mathcal{F}_{1,l}(u)$ with $\mathbb{P}(E_1) > 0$ and $E_2 \in \mathcal{F}_{l+q,n}(u)$. We will assume the existence of a sequence $q_n = o(n)$ such that $\alpha_{cn, q_n}^*(x_n) \rightarrow 0$ for all $c > 0$. This implies that $\{X_n\}_{n=1}^\infty$ satisfies AIM(x_n) with the same choice of q_n and so we may find a sequence p_n so that (9) is satisfied. We define $\theta_{i,n}$ to be as in Eq. 14 and assume a slightly stronger form of convergence than in Eq. 16 but weaker than uniform convergence $\sup_{i \in \mathbb{N}} |\theta_i - \theta_{i,n}| \rightarrow 0$.

The limiting distribution of the normalized interexceedance times is given in Theorem 4.

Theorem 4 *Let $\{X_n\}_{n=1}^\infty$ be a sequence of random variables with common marginal distribution F and $\{x_n\}_{n=1}^\infty$ a sequence of real numbers such that $n\bar{F}(x_n) \rightarrow \tau > 0$. Suppose that there is a sequence of positive integers $q_n = o(n)$ such that $\alpha_{cn, q_n}^*(x_n) \rightarrow 0$ and $\max_{1 \leq i \leq cn} |\theta_i - \theta_{i,n}| \rightarrow 0$ for all $c > 0$. Then, if $\{\theta_i\}_{i=1}^\infty$ is Cesàro summable we have, for each fixed $i \in \mathbb{N}$ and $t > 0$*

$$\mathbb{P}(\bar{F}(x_n)T_i(x_n) > t) \rightarrow \theta_i \exp(-\gamma t). \tag{25}$$

3 Estimation with a focus on periodic sequences

In this section we consider moment and maximum likelihood estimators for θ_i and γ based on the limiting distribution of normalized interexceedance times given in Theorem 4. We first show that the intervals estimator of Ferro and Segers (2003) may be used to estimate θ_i and then consider likelihood based estimation along the lines of Süveges (2007). For simplicity, we focus our discussion on the case of periodic dependence as in Definition 4. Such an assumption reduces estimation

of the extremal clustering function to estimating the vector $\theta = (\theta_1, \dots, \theta_d)$ with $\gamma = d^{-1} \sum_{i=1}^d \theta_i$ where d is the fundamental period which, for simplicity, we assume to be known a-priori. Such an assumption is important for the moment based estimators of Section 3.1 where one needs replications of interexceedance times in order to use the estimators, but can easily be relaxed for likelihood based inference.

3.1 Moment based estimators

Theorem 4 implies that the first two moments of $\bar{F}(u)T_i(u)$ satisfy $\mathbb{E}\{\bar{F}(u)T_i(u)\} = \theta_i/\gamma + o(1)$ and $\mathbb{E}\{\{\bar{F}(u)T_i(u)\}^2\} = 2\theta_i/\gamma^2 + o(1)$ as $u \rightarrow x_F$. Assuming the threshold is chosen to be suitably large so that the $o(1)$ terms can be neglected, these two equations can be solved with respect to the unknown parameters to give

$$\gamma = \frac{2\mathbb{E}(\bar{F}(u)T_i(u))}{\mathbb{E}(\{\bar{F}(u)T_i(u)\}^2)} \text{ and } \theta_i = \frac{2\{\mathbb{E}(\bar{F}(u)T_i(u))\}^2}{\mathbb{E}(\{\bar{F}(u)T_i(u)\}^2)} = \frac{2\{\mathbb{E}(T_i(u))\}^2}{\mathbb{E}(\{T_i(u)\}^2)}, \quad 1 \leq i \leq d. \tag{26}$$

A complication that arises in the non-stationary setting is that, since θ_i is defined via a conditional probability given the event $X_i > u$, if X_i does not exceed the threshold u then there are no interexceedance times to estimate θ_i . This problem doesn't arise in the stationary case where every interexceedance time may be used to estimate the extremal index θ .

In order to estimate θ_i then, it is natural to assume that the extremal clustering function is structured in some way, e.g., periodic or piecewise constant. Making such an assumption allows us to use multiple interexceedance times to estimate θ_i . Focusing on the case where $\{X_n\}_{n=1}^\infty$ has periodic dependence with fundamental period d , all exceedances of the threshold u occurring at points that are separated by a multiple of d give rise to interexceedance times that may be used to estimate the same value of the extremal clustering function. More precisely, suppose that X_1, \dots, X_n is a sample of size n of the process with exceedance times $E = \{1 \leq i \leq n : X_i > u\}$, and corresponding interexceedance times $I = \{T_i(u) : i \in E \setminus \{\max(E)\}\}$, with $T_i(u)$ as in Eq. 23. The set of interexceedance times that may be used for estimating θ_i is the subset $I_i \subseteq I$ defined by $I_i = \{T_j(u) \in I : j \equiv i \pmod{d}\}$. If $|I_i| = N_i$, then we may relabel the elements of I_i as $I_i = \{T_i^{(j)}\}_{j=1}^{N_i}$ where now the subscript remains fixed. Making further, more refined assumptions regarding the nature of the periodicity of the process under consideration may give rise to different sets I_i . For example, in an environmental time series setting it may be reasonable to assume that the extremal clustering function is piecewise constant within months or seasons, so that all interexceedance times that correspond to exceedances within the same calendar month or season belong to the same I_i .

Equation 26 suggests the estimator

$$\hat{\theta}_i = \frac{2(\sum_{j=1}^{N_i} T_i^{(j)})^2}{N_i \sum_{j=1}^{N_i} (T_i^{(j)})^2}, \tag{27}$$

whose bias we now investigate. From Eq. 25 we have that for $n \in \mathbb{N}$

$$\mathbb{P}(T_i(x_n) > n) = \theta_i F(x_n)^{n\gamma} + o(1),$$

which motivates consideration of the positive integer valued random variable T defined by

$$\mathbb{P}(T > n) = \theta_i p^{n\gamma}, \quad \text{for } n \geq 1,$$

where $p \in (0, 1)$ and $\theta_i, \gamma \in (0, 1]$ and we may identify p with $F(x_n)$. In a similar manner to Ferro and Segers (2003), we find that $\mathbb{E}(T) = 1 + \theta_i p^\gamma (1 - p^\gamma)^{-1}$ and $\mathbb{E}(T^2) = 1 + \theta_i p^\gamma (1 - p^\gamma)^{-1} + 2\theta_i p^\gamma (1 - p^\gamma)^{-2}$, so that upon simplification we find that

$$\frac{2\{\mathbb{E}(T)\}^2}{\mathbb{E}(T^2)} = \frac{2(1 - p^\gamma + \theta_i p^\gamma)^2}{(1 - p^\gamma)^2 + \theta_i p^\gamma (1 - p^\gamma) + 2\theta_i p^\gamma}. \tag{28}$$

A Taylor expansion of the right hand side of Eq. 28 around $p = 1$ gives

$$\frac{2\{\mathbb{E}(T)\}^2}{\mathbb{E}(T^2)} = \theta_i + \gamma(2 - 3\theta_i/2)(1 - p) + O\{(1 - p)^2\}, \quad \text{as } p \rightarrow 1,$$

so that the first order bias of $\hat{\theta}_i$ is $\gamma(2 - 3\theta_i/2)\bar{F}(x_n)$. On the other hand, since

$$\theta_i = \frac{2\{\mathbb{E}(T - 1)\}^2}{\mathbb{E}\{(T - 1)(T - 2)\}},$$

this motivates the estimator

$$\tilde{\theta}_i = \frac{2 \sum_{j=1}^{N_i} (T_i^{(j)} - 1)^2}{N_i \sum_{j=1}^{N_i} (T_i^{(j)} - 1)(T_i^{(j)} - 2)}, \tag{29}$$

whose first order bias is zero. This estimator forms the key component of the intervals estimator of Ferro and Segers (2003), which we can use to estimate θ_i . We note that $\tilde{\theta}_i$ may take values greater than 1 and is not defined if $\max(I_i) \leq 2$ as then the denominator in Eq. 29 is zero. In order to deal with these cases, the intervals estimator θ_i^* of θ_i is defined as

$$\theta_i^* = \begin{cases} \min\{1, \hat{\theta}_i\} & \text{if } \max(I_i) \leq 2, \\ \min\{1, \tilde{\theta}_i\} & \text{if } \max(I_i) > 2. \end{cases}$$

While Eq. 26 also suggests an estimator for γ , this is based only on the interexceedances relevant to estimating θ_i and also requires an estimate of $\bar{F}(u)$. One possibility is to obtain d such estimates and take the mean of these as the estimate of γ . However, this estimator need not respect the relation $\gamma = d^{-1} \sum_{i=1}^d \theta_i$, a consequence of the fact that we dropped the $o(1)$ terms when solving the first two moment equations. In the examples that we consider in Section 4, we estimate γ using the mean of the estimates for the θ_i values.

3.2 Maximum likelihood estimation

Theorem 4 also allows for the construction of the likelihood function for the vector of unknown parameters. This is an attractive approach due to the modelling possibilities that become available, however, as discussed in Ferro and Segers (2003) in the stationary case, problems arise with maximum likelihood estimation due to uncertainty in how to assign interexceedance times to the components of the limiting mixture

distribution. Since the asymptotically valid likelihood is used as an approximation at some subasymptotic threshold u , all observed normalized interexceedance times are strictly positive. Assigning all interexceedance times to the exponential part of the limiting mixture means that they are all being classified as between cluster times. This is tantamount to exceedances of a large threshold occurring in isolation, and so the maximum likelihood estimator based on this, typically misspecified, likelihood converges in probability to 1 regardless of the true underlying value of θ .

This problem was addressed in Süveges (2007) for sequences satisfying the $D^{(2)}(x_n)$ condition, i.e., the case $k = 2$ in Eq. 20. For such sequences, in the limit as $n \rightarrow \infty$, exceedances above x_n cluster into independent groups of consecutive exceedances, so that all observed interexceedance times equal to one are assigned to the zero component of the mixture likelihood. On the other hand, all interexceedance times greater than one are assigned to the exponential component of the likelihood. It was found that, when the $D^{(2)}(x_n)$ condition is satisfied, maximum likelihood estimation outperforms the intervals estimator in terms of lower root mean squared error. The consecutive exceedances model of clusters implied by $D^{(2)}(x_n)$ is in contrast to the general situation where within clusters, exceedances may be separated by observations that fall below the threshold.

If we were to make the $D^{(2)}(x_n)$ assumption in our non-stationary setting, so that the consecutive exceedances model for clusters is accurate, then with $I_i = \{T_i^{(j)}\}_{j=1}^{N_i}$ the interexceedance times relevant for estimating θ_i as in Section 3.1, we obtain the likelihood function as

$$L(\theta; I) = \prod_{i=1}^d L_i(\theta; I_i)$$

where $I = \cup_{i=1}^d I_i$ is the set of all interexceedance times and

$$L_i(\theta; I_i) = \prod_{j=1}^{N_i} (1 - \theta_i)^{\mathbb{1}[T_i^{(j)}=1]} \{ \theta_i \gamma \exp(-\gamma \bar{F}(x_n) T_i^{(j)}) \}^{\mathbb{1}[T_i^{(j)}>1]}.$$

The full log-likelihood is then

$$l(\theta; I) = \sum_{i=1}^d (N_i - n_i) \log(1 - \theta_i) + \sum_{i=1}^d n_i \log(\theta_i) + \left(\sum_{i=1}^d n_i \right) \log(\gamma) - \gamma \bar{F}(x_n) \sum_{i=1}^d \sum_{j=1}^{N_i} (T_i^{(j)} - 1) - \gamma \bar{F}(x_n) \sum_{i=1}^d n_i, \tag{30}$$

where $\gamma = d^{-1} \sum_{i=1}^d \theta_i$, $n_i = \sum_{j=1}^{N_i} \mathbb{1}[T_i^{(j)} > 1]$, and in practice $\bar{F}(x_n)$ must be replaced with an estimate. Unlike in the stationary case, the likelihood equations don't have a closed form solution, essentially due to the dependence of γ on all the θ_i . Equation 30, however, is easily optimized numerically provided d is not too large. If d is large, it is more natural to parameterize θ_i in terms of a small number of parameters which we may estimate by maximum likelihood or consider non-parametric estimation along the lines of Einmahl et al. (2016).

We may generalise this idea and assign all interexceedance times less than or equal to some value k to the zero component of the likelihood, so that the corresponding expression for L_i becomes

$$L_i(\theta; I_i) = \prod_{j=1}^{N_i} (1 - \theta_i)^{\mathbb{1}[T_i^{(j)} \leq k]} \{ \theta_i \gamma \exp(-\gamma \bar{F}(x_n) T_i^{(j)}) \}^{\mathbb{1}[T_i^{(j)} > k]}. \quad (31)$$

This may be justified by the assumption that the sequence satisfies the $D^{(k+1)}(x_n)$ condition. Selection of an appropriate value of k is equivalent to the selection of the run length for the runs estimator, and this problem is considered in the stationary case in Süveges and Davison (2010) and Juan Cai (2019). However, in a non-stationary setting, where the clustering characteristics of the sequence may change in time, the appropriate value of k may also be time varying, so that k may be replaced with k_i in Eq. 31. Although, as was discussed in Section 3.1, we may take a constant value of k in the definition of $D^{(k)}(x_n)$, for the purposes of estimating θ_i , one wants to select for each i , the smallest $k = k_i$ such that Eq. 20 is satisfied (Hsing 1993). If too small a value is selected for k_i then some of the interexceedance times may be wrongly assigned to the exponential component of the likelihood leading to an overestimate of θ_i whereas if k_i is selected to be too large then we tend to underestimate θ_i .

4 Examples

In this section we consider two simple examples of non-stationary Markov sequences with a periodic dependence structure and common marginal distributions. The first example we consider is the Gaussian autoregressive model

$$X_{n+1} = \rho_n X_n + \epsilon_n, \quad n \geq 1, \quad (32)$$

where $\epsilon_n \sim N(0, 1 - \rho_n^2)$, $|\rho_n| < 1$ and $X_1 \sim N(0, 1)$. In our second example, we consider a model where (X_n, X_{n+1}) follow a bivariate logistic distribution with joint distribution function

$$F(x_n, x_{n+1}) = \exp \left\{ - (x_n^{-1/\alpha_n} + x_{n+1}^{-1/\alpha_n})^{\alpha_n} \right\}, \quad n \geq 1, \quad (33)$$

$\alpha_n \in (0, 1]$ and $X_1 \sim \text{Fréchet}(1)$ so that $\mathbb{P}(X_1 \leq x) = e^{-1/x}$, $x \geq 0$. For the Gaussian model, no limiting extremal clustering occurs at any point in the sequence, so that $\theta_i = 1$ for each i , in contrast to the logistic model where $\theta_i < 1$ for each i .

For sufficiently well behaved stationary Markov sequences, mixing conditions much stronger than those considered in Section 2 hold. For example, for the stationary Gaussian autoregressive sequence, with $\rho_n = \rho$ in Eq. 32 for all $n \geq 1$, Theorems 1 and 2 from Athreya and Pantula (1986) give that the mixing conditions of Theorem 1 and Theorem 4 hold for any sequence q_n such that $q_n \rightarrow \infty$, $q_n = o(n)$, for any x_n . Analogous results also hold for the non-stationary models that we consider in this section, see for example Bradley (2005) Theorem 3.3 and Davydov (1973) Theorem 4.

4.1 Gaussian autoregressive model

Stationary sequences $\{X_n\}_{n=1}^\infty$ where each X_i is a standard Gaussian random variable, are extensively studied in Chapter 4 of Leadbetter et al. (1983). It is shown there that if the lag n autocorrelation $\rho(n)$ satisfies $\rho(n) \log n \rightarrow 0$, then the extremal index θ of the sequence equals one and so no limiting extremal clustering occurs. Thus, the stationary autoregressive sequence with $\rho_n = \rho$ in Eq. 32 for all $n \geq 1$ has extremal index one, provided $\rho < 1$. This is a special case of the more general result that a stationary asymptotically independent Markov sequence has an extremal index of one (Smith 1992). We say that the stationary sequence $\{X_n\}_{n=1}^\infty$ is asymptotically independent at lag k if $\chi(k) = 0$ where

$$\chi(k) = \lim_{u \rightarrow x_F} \mathbb{P}(X_{n+k} > u \mid X_n > u), \quad k \geq 1,$$

and asymptotically independent if $\chi(k) = 0$ for all k (Ledford and Tawn 2003).

Here, we consider the non-stationary autoregressive model (32) and specify a periodic lag one correlation function $\rho_{n+1} = 0.5 + 0.25 \sin(2\pi n/7)$ for $n \geq 0$. Applying Theorem 6.3.4 of Leadbetter et al. (1983), and comparing the non-stationary sequence to an independent standard Gaussian sequence, we deduce that $\mathbb{P}(M_n \leq x_n) - \Phi(x_n)^n \rightarrow 0$ as $n \rightarrow \infty$ where Φ is the standard Gaussian distribution function, and thus conclude that $\gamma = 1$ and $\theta_i = 1$ for $i = 1, \dots, 7$. The same conclusion may also be drawn by applying Theorem 4.1 of Hüsler (1983), which shows that if x_n satisfies (7) then $\mathbb{P}(M_n \leq x_n) \rightarrow e^{-\tau}$.

We simulated 1000 realizations of this sequence of length 10^4 and, for each realization, estimated $\theta_1, \dots, \theta_7$ and γ for a range of high thresholds, using both the intervals estimator and maximum likelihood with k in Eq. 31 equal to zero and one. We then repeated this procedure for sequences of length 10^5 and 10^6 . We found that the maximum likelihood estimator with $k = 0$ gave by far the best performance as measured by root mean squared error in γ . In fact, in this case the 0.025 and 0.975 empirical quantiles of the estimated values of γ were both 1 to two decimal places in all simulations. This is not surprising since selecting $k = 0$ ensures that all interexceedance times have the correct asymptotic classification as between cluster times. However, in a real data example such a level of prior knowledge regarding asymptotic independence is not realistic and would render estimation redundant. Although maximum likelihood estimation with $k = 1$ performed slightly poorer than the intervals estimator, both methods produced broadly similar results.

Table 1 shows the 0.025 and 0.975 empirical quantiles of the parameter estimates obtained using the intervals estimator. In the table, $u = q_p$ corresponds to the threshold that there is probability p of exceeding at each time point i.e., $\mathbb{P}(X_i > q_p) = p$. Although the true value of each θ_i is 1, so that no extremal clustering occurs in the limit as $u \rightarrow \infty$, clustering may occur at subasymptotic levels. Moreover, there will tend to be more subasymptotic clustering in the sequence at points with a larger lag one autocorrelation, i.e., larger ρ_i . This point has been thoroughly discussed in the context of stationary sequences and estimation of the extremal index (Ancona-Navarrete and Tawn 2000; Eastoe and Tawn 2012) and leads to the notion of a subasymptotic or threshold based extremal index.

Table 1 0.025 and 0.975 empirical quantiles of the estimates of $\theta_1, \dots, \theta_7, \gamma$, in the Gaussian autoregressive model using the intervals estimator

n	u	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	γ
10^4	$q_{0.10}$.53, .84	.39, .66	.34, .62	.46, .77	.62, .93	.69, 1.0	.66, 1.0	.62, .74
10^4	$q_{0.05}$.53, .96	.39, .78	.35, .74	.48, .91	.61, 1.0	.70, 1.0	.64, 1.0	.66, .82
10^4	$q_{0.01}$.54, 1.0	.41, 1.0	.39, 1.0	.51, 1.0	.61, 1.0	.62, 1.0	.60, 1.0	.75, .97
10^5	$q_{0.10}$.62, .73	.46, .56	.42, .51	.55, .65	.71, .81	.78, .90	.77, .87	.65, .69
10^5	$q_{0.05}$.66, .80	.50, .63	.46, .59	.60, .73	.75, .90	.81, .97	.79, .94	.70, .75
10^5	$q_{0.01}$.69, 1.0	.55, .84	.52, .81	.65, .96	.76, 1.0	.80, 1.0	.78, 1.0	.78, .88
10^6	$q_{0.05}$.71, .75	.54, .58	.51, .55	.64, .69	.80, .85	.87, .91	.84, .89	.71, .73
10^6	$q_{0.01}$.78, .89	.63, .73	.60, .70	.74, .84	.87, .97	.90, 1.0	.88, .99	.81, .84
10^6	$q_{0.001}$.77, 1.0	.65, .98	.65, .94	.75, 1.0	.82, 1.0	.83, 1.0	.83, 1.0	.86, .95

These are based on 1000 realizations of the process for different sample sizes n and thresholds u

4.2 Bivariate logistic dependence

The stationary logistic model, that is, Eq. 33 with $\alpha_n = \alpha$ for all $n \geq 1$, has been thoroughly studied (Smith et al. 1997; Ledford and Tawn 2003; Süveges 2007). The parameter α controls the strength of dependence between adjacent terms in the sequence, with $\alpha = 1$ corresponding to independence and $\alpha \rightarrow 0$ giving complete dependence. Such a sequence exhibits asymptotic dependence provided $\alpha < 1$, in particular, $\lim_{u \rightarrow \infty} \mathbb{P}(X_{n+1} > u \mid X_n > u) = 2 - 2^\alpha$. By exploiting the Markov structure of the sequence, precise calculation of θ can be achieved using the numerical methods described in Smith (1992), where it is found for example that the sequence with $\alpha = 1/2$ has $\theta = 0.328$, and moreover, Eq. 18 is shown to hold for all $\alpha \in (0, 1]$. The case of $\alpha = 1/2$ is also considered in Süveges (2007) where, based on diagnostic plots, it is concluded that the $D^{(2)}(x_n)$ condition is not satisfied for this sequence, and moreover, the maximum likelihood estimator for θ based on a run length of $k = 1$ has bias of around 20%. Süveges and Davison (2010) find that a more suitable run length is $k = 5$, and in this case the maximum likelihood estimator for θ has lower root mean squared error than the intervals estimator. Smaller values of α will tend to be associated with larger values of the run length k , though the precise nature of this relation is unclear.

We consider the non-stationary logistic model (33) with $\alpha_{n+1} = 0.5 + 0.25 \sin(2\pi n/7)$ for $n \geq 0$. Note that although we have specified the same parametric form for the dependence parameters α as in the previous example for ρ , the two parameters are not directly comparable. We simulated 1000 realizations of this process, of lengths 10^4 and 10^5 , and estimated $\theta_1, \dots, \theta_7$ using maximum likelihood with $k = 5$, at a range of different thresholds. Table 2 shows, for the different sample sizes and thresholds considered, the 0.025 and 0.975 empirical quantiles of the parameter estimates obtained from this simulation. Although the exact values of the parameters are unknown, making evaluation of any estimators perfor-

Table 2 0.025 and 0.975 empirical quantiles of the estimates of $\theta_1, \dots, \theta_7, \gamma$, in the logistic time series model using maximum likelihood with $k = 5$

n	u	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	γ
10^4	$q_{0.10}$.19, .39	.38, .58	.44, .65	.30, .52	.12, .29	.05, .19	.08, .23	.28, .35
10^4	$q_{0.05}$.21, .46	.41, .67	.46, .72	.30, .55	.11, .34	.05, .22	.08, .27	.29, .39
10^4	$q_{0.01}$.11, .65	.30, .85	.37, .92	.19, .74	.03, .50	.00, .34	.00, .41	.27, .48
10^5	$q_{0.10}$.22, .35	.42, .54	.48, .62	.32, .49	.16, .24	.08, .17	.11, .20	.29, .33
10^5	$q_{0.05}$.27, .42	.46, .61	.53, .66	.36, .51	.17, .29	.08, .19	.10, .21	.32, .37
10^5	$q_{0.01}$.27, .46	.48, .67	.53, .73	.35, .55	.15, .32	.07, .20	.11, .26	.33, .40

These are based on 1000 realizations of the process for different sample sizes n and thresholds u

mance impossible, an upper bound for θ_i is easily obtained as $\lim_{u \rightarrow \infty} \mathbb{P}(X_{i+1} \leq u \mid X_i > u) = 2^{\alpha_i} - 1$. In our case this gives the bounds $(\theta_1, \dots, \theta_7) \leq (0.41, 0.62, 0.67, 0.52, 0.31, 0.19, 0.24)$ and $\gamma \leq 0.42$ where the relation \leq is interpreted componentwise. It is conceivable that the methods in Smith (1992) could be adapted to the non-stationary case to allow exact computation of θ_i though we do not pursue this direction here.

We also considered estimation of θ_i using the intervals estimator and obtained similar results to the maximum likelihood estimates. The median value of the 1000 estimates for each parameter using the different methods of estimation are shown in Fig. 1 for the sample size of 10^5 and threshold $q_{0.05}$. The estimators clearly recover the periodicity in the dependence structure of the sequence and, on average, respect the upper bound for θ_i of $2^{\alpha_i} - 1$.

5 Proofs

5.1 Auxiliary results

In this section we state and prove some Lemmas that are required in the proof of Theorem 1.

Lemma 3 *Let $\{t_n\}_{n=1}^\infty$ be a sequence of positive integers and $a_{i,n}, i, n \in \mathbb{N}$, an array of non-negative real numbers such that $t_n \rightarrow \infty$ and $A_n = \max_{1 \leq i \leq t_n} a_{i,n} \rightarrow 0$ as $n \rightarrow \infty$. Then,*

$$\sum_{i=1}^{t_n} a_{i,n} \rightarrow \tau \geq 0 \tag{34}$$

if and only if

$$\prod_{i=1}^{t_n} (1 - a_{i,n}) \rightarrow e^{-\tau}. \tag{35}$$

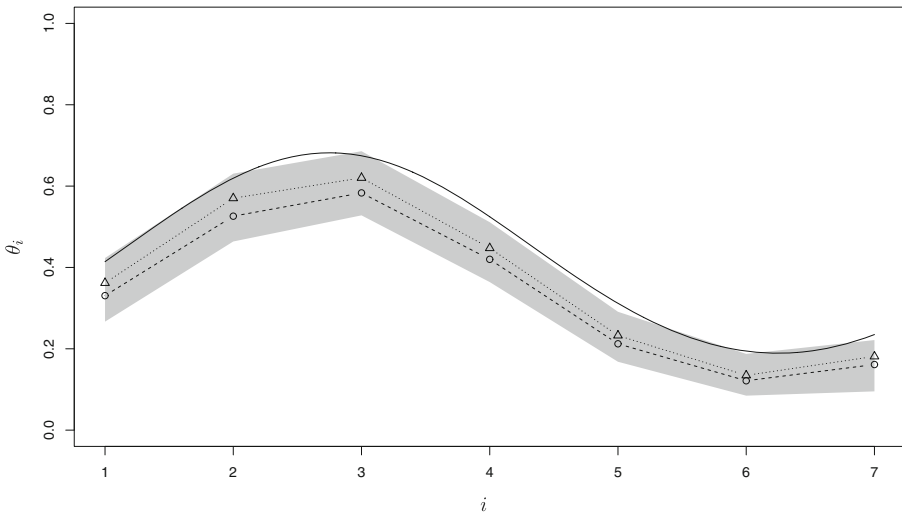


Fig. 1 Illustration of estimators (triangles: intervals estimator, and circles: maximum likelihood with $k = 5$) obtained from 10^3 realizations from the non-stationary logistic model of length 10^5 and threshold $q_{0,05}$. The marked points correspond to the median estimate from 10^3 realizations of the model. The grey region contains the 0.025 and 0.975 empirical quantiles of the parameter estimates using both the intervals and maximum likelihood estimators. It is constructed by taking the pointwise maxima of the 0.975 quantiles (upper boundary) and pointwise minima of the 0.025 quantiles (lower boundary). The solid black curve shows the upper bound for θ_i of $2^{\alpha_i} - 1$

Proof Using the fact that $\log(1 - x) = -x + R(x)$ where $|R(x)| < Cx^2$, for sufficiently small $x > 0$, for some $C > 0$, we have $\sum_{i=1}^{t_n} \log(1 - a_{i,n}) = -\sum_{i=1}^{t_n} a_{i,n} + \sum_{i=1}^{t_n} R(a_{i,n})$ and

$$\left| \sum_{i=1}^{t_n} R(a_{i,n}) \right| \leq C \sum_{i=1}^{t_n} a_{i,n}^2 \leq CA_n \sum_{i=1}^{t_n} a_{i,n}, \tag{36}$$

so that $\sum_{i=1}^{t_n} \log(1 - a_{i,n}) = -(\sum_{i=1}^{t_n} a_{i,n})(1 + o(1))$ or equivalently

$$\log \prod_{i=1}^{t_n} (1 - a_{i,n}) = -\left(\sum_{i=1}^{t_n} a_{i,n}\right)(1 + o(1)), \tag{37}$$

from which the result follows. □

Lemma 4 Let $\{t_n\}_{n=1}^\infty$ be a sequence of positive integers and $a_{i,n}$, $i, n \in \mathbb{N}$, an array of non-negative real numbers such that $t_n \rightarrow \infty$, $A_n = \max_{1 \leq i \leq t_n} a_{i,n} \rightarrow 0$ and $\sum_{i=1}^{t_n} a_{i,n}$ is bounded above as $n \rightarrow \infty$. Then,

$$\prod_{i=1}^{t_n} (1 - a_{i,n}) - \exp\left(-\sum_{i=1}^{t_n} a_{i,n}\right) \rightarrow 0. \tag{38}$$

Proof This follows from Lemma 3 by considering subsequences along which $\sum_{i=1}^{t_n} a_{i,n}$ converges. \square

Lemma 5 *Let $g : \mathbb{R} \rightarrow \mathbb{R}$ be a bounded function. If $f(x) = A(x)g(x)$ and $\lim_{x \rightarrow \infty} A(x) = 1$, then $f(x) = g(x) + o(1)$ as $x \rightarrow \infty$.*

Proof As g is bounded, $\exists M > 0$ such that $|g(x)| < M, \forall x \in \mathbb{R}$. Now let $\epsilon > 0$. As $\lim_{x \rightarrow \infty} A(x) = 1, \exists x_0$ such that

$$|A(x) - 1| < \epsilon/M \quad \text{for } x > x_0.$$

Then for $x > x_0$

$$|f(x) - g(x)| = |g(x)||A(x) - 1| < M\epsilon/M = \epsilon. \quad \square$$

Lemma 6 *Let $\{X_n\}_{n=1}^\infty, x_n, p_n$ and q_n be as in Lemma 2 such that Eq. 7 holds and assume $\mathbb{P}(M_n \leq x_n) \rightarrow L \in (0, 1)$. Let s_n be such that $p_n = o(s_n), s_n = o(n)$ and $t_n = \lfloor n/(s_n + q_n) \rfloor$. Then*

$$\sum_{i=1}^{t_n} \mathbb{P}(M_{s_n-p_n, s_n}^i > x_n) = o\left(\sum_{i=1}^{t_n} \mathbb{P}(M_{0, s_n-p_n}^i > x_n, M_{s_n-p_n, s_n}^i \leq x_n)\right) \quad (39)$$

where $M_{j,k}^i = \max\{X_{j+1}^i, \dots, X_k^i\}$ and $X_j^i = X_{(i-1)(s_n+q_n)+j}$.

Proof We first note that Lemma 2 also holds with blocks of length s_n , i.e., with s_n in place of p_n in the definition of A_i in Eq. 8 and t_n in place of r_n . Thus from Eq. 10, with blocks of length s_n , we have that $\mathbb{P}(M_n \leq x_n) - \prod_{i=1}^{t_n} \mathbb{P}(M(A_i) \leq x_n) \rightarrow 0$ so that $\prod_{i=1}^{t_n} \mathbb{P}(M(A_i) \leq x_n) \rightarrow L \in (0, 1)$, or equivalently

$$\sum_{i=1}^{t_n} \log(1 - \mathbb{P}(M(A_i) > x_n)) \rightarrow \log(L). \quad (40)$$

Now we note that $\max_{1 \leq i \leq t_n} \mathbb{P}(M(A_i) > x_n) \rightarrow 0$ since $\mathbb{P}(M(A_i) > x_n) \leq s_n \bar{F}(x_n)$ and $s_n = o(n)$ and $\bar{F}(x_n) = \tau n^{-1} + o(n^{-1})$. Thus, using $\log(1 - t) = -t + o(t)$ as $t \rightarrow 0$, Eq. 40 may be written

$$-\sum_{i=1}^{t_n} \mathbb{P}(M(A_i) > x_n) + \sum_{i=1}^{t_n} o(\mathbb{P}(M(A_i) > x_n)) \rightarrow \log(L). \quad (41)$$

Now it is easily seen that the second sum in Eq. 41 converges to zero since

$$\begin{aligned} \sum_{i=1}^{t_n} o(\mathbb{P}(M(A_i) > x_n)) &= \sum_{i=1}^{t_n} o(s_n \bar{F}(x_n)) \leq t_n s_n \bar{F}(x_n) o(1) \\ &\leq \frac{s_n}{s_n + q_n} n \bar{F}(x_n) o(1) \rightarrow 0, \end{aligned}$$

and so Eq. 41 implies $\sum_{i=1}^{t_n} \mathbb{P}(M(A_i) > x_n) \rightarrow -\log(L)$. Now, decomposing the event $\{M(A_i) > x_n\}$ as a disjoint union we get

$$\sum_{i=1}^{t_n} \mathbb{P}(M_{0,s_n-p_n}^i > x_n, M_{s_n-p_n,s_n}^i \leq x_n) + \sum_{i=1}^{t_n} \mathbb{P}(M_{s_n-p_n,s_n}^i > x_n) \rightarrow -\log(L). \tag{42}$$

Again, the second sum in Eq. 42 goes to zero since it is bounded above by $t_n p_n \bar{F}(x_n) \leq \{p_n/(s_n + q_n)\}n\bar{F}(x_n) \rightarrow 0$, from which Eq. 39 follows. \square

5.2 Proof of Lemma 1

We use induction on the number of subintervals, k . The case $k = 2$ is just the fact that $\{X_n\}_{n=1}^\infty$ satisfies AIM(x_n). Assuming that the result is true for k such arbitrary subintervals, we will verify it also holds for the $k + 1$ intervals $I_1, I_2, \dots, I_k, I_{k+1}$. Let I_1^* be the interval separating I_1 and I_2 and let $J = I_1 \cup I_1^* \cup I_2$, and we note that J is an interval with $|J| > q_n$ and since $\{M(J \cup \cup_{i=3}^{k+1} I_i) \leq x_n\} \subseteq \{M(\cup_{i=1}^{k+1} I_i) \leq x_n\}$ we have,

$$0 \leq \mathbb{P}(M(\cup_{i=1}^{k+1} I_i) \leq x_n) - \mathbb{P}(M(J \cup \cup_{i=3}^{k+1} I_i) \leq x_n) \leq \mathbb{P}(M(I_1^*) > x_n) \leq q_n \bar{F}(x_n), \tag{43}$$

so we may write $\mathbb{P}(M(\cup_{i=1}^{k+1} I_i) \leq x_n) = \mathbb{P}(M(J \cup \cup_{i=3}^{k+1} I_i) \leq x_n) + R_{1,n}$ where the remainder $R_{1,n}$ satisfies $R_{1,n} \leq q_n \bar{F}(x_n)$. We then have

$$\left| \mathbb{P}(M(\cup_{i=1}^{k+1} I_i) \leq x_n) - \prod_{i=1}^{k+1} \mathbb{P}(M(I_i) \leq x_n) \right| \tag{44}$$

$$= \left| \mathbb{P}(M(J \cup \cup_{i=3}^{k+1} I_i) \leq x_n) - \prod_{i=1}^{k+1} \mathbb{P}(M(I_i) \leq x_n) + R_{1,n} \right| \tag{45}$$

$$= \left| \mathbb{P}(M(J) \leq x_n) \prod_{i=3}^{k+1} \mathbb{P}(M(I_i) \leq x_n) - \prod_{i=1}^{k+1} \mathbb{P}(M(I_i) \leq x_n) + R_{1,n} + R_{2,n} \right| \tag{46}$$

$$\leq \left| \mathbb{P}(M(J) \leq x_n) - \mathbb{P}(M(I_1) \leq x_n)\mathbb{P}(M(I_2) \leq x_n) \right| + |R_{1,n}| + |R_{2,n}| \tag{47}$$

where $|R_{2,n}| \leq (k - 1)\alpha_n + 2(k - 2)q_n \bar{F}(x_n)$ and we have used our induction hypothesis to get (46) since $J \cup \cup_{i=3}^{k+1} I_i$ is a union of k intervals with adjacent intervals separated by q_n . Now note that since $\{M(J) \leq x_n\} \subseteq \{M(I_1 \cup I_2) \leq x_n\}$ we have $0 \leq \mathbb{P}(M(I_1 \cup I_2) \leq x_n) - \mathbb{P}(M(J) \leq x_n) \leq q_n \bar{F}(x_n)$ and we may write $\mathbb{P}(M(J) \leq x_n) = \mathbb{P}(M(I_1 \cup I_2) \leq x_n) + R_{3,n}$ where $|R_{3,n}| \leq q_n \bar{F}(x_n)$. Thus from

Eqs. 44 and 47 we have

$$\begin{aligned}
 & \left| \mathbb{P}(M(\cup_{i=1}^{k+1} I_i) \leq x_n) - \prod_{i=1}^{k+1} \mathbb{P}(M(I_i) \leq x_n) \right| \\
 & \leq \left| \mathbb{P}(M(I_1 \cup I_2) \leq x_n) - \mathbb{P}(M(I_1) \leq x_n)\mathbb{P}(M(I_2) \leq x_n) + R_{3,n} \right| + |R_{1,n}| + |R_{2,n}| \\
 & \leq \left| \mathbb{P}(M(I_1 \cup I_2) \leq x_n) - \mathbb{P}(M(I_1) \leq x_n)\mathbb{P}(M(I_2) \leq x_n) \right| + |R_{1,n}| + |R_{2,n}| + |R_{3,n}| \\
 & \leq \alpha_n + q_n \bar{F}(x_n) + (k - 1)\alpha_n + 2(k - 2)q_n \bar{F}(x_n) + q_n \bar{F}(x_n) \\
 & = k\alpha_n + 2(k - 1)q_n \bar{F}(x_n)
 \end{aligned}$$

as required.

5.3 Proof of Lemma 2.

As $\{M_n \leq x_n\} \subseteq \cap_{i=1}^{r_n} \{M(A_i) \leq x_n\}$ we have

$$\begin{aligned}
 0 \leq \mathbb{P}(M(\cup_{i=1}^{r_n} A_i) \leq x_n) - \mathbb{P}(M_n \leq x_n) & \leq \mathbb{P}(M(\cup_{i=1}^{r_n} A_i^*) > x_n) \\
 & \leq r_n q_n \bar{F}(x_n) \\
 & \leq n(p_n + q_n)^{-1} q_n \{ \tau n^{-1} + o(n^{-1}) \} \rightarrow 0. \tag{48}
 \end{aligned}$$

Also, by Lemma 1 we have

$$\left| \mathbb{P}(M(\cup_{i=1}^{r_n} A_i) \leq x_n) - \prod_{i=1}^{r_n} \mathbb{P}(M(A_i) \leq x_n) \right| \leq (r_n - 1)\alpha_n + 2(r_n - 2)q_n \bar{F}(x_n) \rightarrow 0 \tag{49}$$

so that the triangle inequality and Eqs. 48 and 49 give the result.

5.4 Proof of Theorem 1.

In addition to the notation defined in Section 2.1, we also define

$$X_j^i = X_{(i-1)(p_n+q_n)+j}, \quad M_{j,k}^i = \max \{X_{j+1}^i, \dots, X_k^i\}. \tag{50}$$

Now, for $i = 1, \dots, r_n$, we have

$$\begin{aligned}
 \mathbb{P}(M(A_i) \leq x_n) & = 1 - \mathbb{P}(M(A_i) > x_n) \\
 & = 1 - \sum_{j=1}^{p_n} \mathbb{P}(X_j^i > x_n, M_{j,p_n}^i \leq x_n) \\
 & \leq 1 - \sum_{j=1}^{p_n} \mathbb{P}(X_j^i > x_n, M_{j,j+p_n}^i \leq x_n) \\
 & \leq \exp \left\{ - \sum_{j=1}^{p_n} \mathbb{P}(X_j^i > x_n, M_{j,j+p_n}^i \leq x_n) \right\}
 \end{aligned}$$

and so

$$\begin{aligned} \mathbb{P}(M_n \leq x_n) &= \prod_{i=1}^{r_n} \mathbb{P}(M(A_i) \leq x_n) + o(1) \\ &\leq \exp \left\{ - \sum_{i=1}^{r_n} \sum_{j=1}^{p_n} \mathbb{P}(X_j^i > x_n, M_{j,j+p_n}^i \leq x_n) \right\} + o(1). \end{aligned} \tag{51}$$

Now we note that

$$\sum_{j=1}^n \mathbb{P}(X_j > x_n, M_{j,j+p_n} \leq x_n) = \sum_{i=1}^{r_n} \sum_{j=1}^{p_n} \mathbb{P}(X_j^i > x_n, M_{j,j+p_n}^i \leq x_n) + o(1) \tag{52}$$

since the difference between the two sums is

$$\begin{aligned} \sum_{j=1}^n \mathbb{P}(X_j > x_n, M_{j,j+p_n} \leq x_n) &- \sum_{i=1}^{r_n} \sum_{j=1}^{p_n} \mathbb{P}(X_j^i > x_n, M_{j,j+p_n}^i \leq x_n) \\ &= \sum_{i=1}^{r_n} \sum_{j=p_n+1}^{p_n+q_n} \mathbb{P}(X_j^i > x_n, M_{j,j+p_n} \leq x_n) + o(1) \\ &\leq r_n q_n \bar{F}(x_n) + o(1) \\ &\leq \frac{q_n}{p_n + q_n} n \bar{F}(x_n) + o(1) \rightarrow 0 \end{aligned}$$

so that Eq. 51 gives

$$\mathbb{P}(M_n \leq x_n) \leq \exp \left\{ - \sum_{j=1}^n \mathbb{P}(X_j > x_n, M_{j,j+p_n} \leq x_n) \right\} + o(1). \tag{53}$$

Now we prove the reverse inequality of Eq. 53, i.e.,

$$\mathbb{P}(M_n \leq x_n) \geq \exp \left\{ - \sum_{j=1}^n \mathbb{P}(X_j > x_n, M_{j,j+p_n} \leq x_n) \right\} + o(1). \tag{54}$$

We will show that Eq. 54 holds under the assumption that $\mathbb{P}(M_n \leq x_n)$ converges to some L in $[0, 1]$, with the more general case following by considering subsequences along which $\mathbb{P}(M_n \leq x_n)$ converges and repeating the following argument. By Lemma 2, specifically (10), and Lemma 4 with $a_{i,n} = \mathbb{P}(M(A_i) > x_n)$ we see that $L > 0$, and since (54) trivially holds when $L = 1$ we may assume $L \in (0, 1)$. Following (O'Brien 1987), introduce a new sequence $s_n = o(n)$ which plays the role of p_n such that $p_n = o(s_n)$ and let $t_n = \lfloor n/(s_n + q_n) \rfloor$ which now plays the role of r_n and note that $t_n = o(r_n)$ and the definitions in Eqs. 50 and 8 are modified by replacing p_n with s_n . Then for $i = 1, \dots, t_n$, we have

$$\mathbb{P}(M(A_i) > x_n) = \mathbb{P}(M_{0,s_n-p_n}^i > x_n, M_{s_n-p_n,s_n}^i \leq x_n) + \mathbb{P}(M_{s_n-p_n,s_n}^i > x_n)$$

and consequently, since Lemma 2 holds with s_n in place of p_n and t_n in place of r_n ,

$$\begin{aligned} \mathbb{P}(M_n \leq x_n) &= \prod_{i=1}^{t_n} \mathbb{P}(M(A_i) \leq x_n) + o(1) = \prod_{i=1}^{t_n} \left\{ 1 - \mathbb{P}(M(A_i) > x_n) \right\} + o(1) \\ &= \prod_{i=1}^{t_n} \left\{ 1 - \mathbb{P}(M_{0,s_n-p_n}^i > x_n, M_{s_n-p_n,s_n}^i \leq x_n) - \mathbb{P}(M_{s_n-p_n,s_n}^i > x_n) \right\} + o(1). \end{aligned} \tag{55}$$

Now, with $a_{i,n} = \mathbb{P}(M_{0,s_n-p_n}^i > x_n, M_{s_n-p_n,s_n}^i \leq x_n) + \mathbb{P}(M_{s_n-p_n,s_n}^i > x_n)$ we have, for all i , $a_{i,n} \leq s_n \bar{F}(x_n)$ and so $\max_{1 \leq i \leq t_n} a_{i,n} \leq s_n \bar{F}(x_n) \rightarrow 0$ as $s_n = o(n)$ and $\bar{F}(x_n) = \tau n^{-1} + o(n^{-1})$. Also, $\sum_{i=1}^{t_n} a_{i,n} \leq t_n \max_{1 \leq i \leq t_n} a_{i,n} \leq s_n (s_n + q_n)^{-1} n \bar{F}(x_n)$ which is bounded above as $n \rightarrow \infty$. Thus we may apply Lemma 4 to Eq. 55 to get

$$\begin{aligned} \mathbb{P}(M_n \leq x_n) &= \exp \left\{ - \sum_{i=1}^{t_n} \left(\mathbb{P}(M_{0,s_n-p_n}^i > x_n, M_{s_n-p_n,s_n}^i \leq x_n) + \mathbb{P}(M_{s_n-p_n,s_n}^i > x_n) \right) \right\} + o(1) \\ &= \exp \left\{ - \left(\sum_{i=1}^{t_n} \mathbb{P}(M_{0,s_n-p_n}^i > x_n, M_{s_n-p_n,s_n}^i \leq x_n) \right) (1 + o(1)) \right\} + o(1) \end{aligned} \tag{56}$$

with Eq. 56 following from Lemma 6. Now Lemma 5 reduces (56) to

$$\begin{aligned} \mathbb{P}(M_n \leq x_n) &= \exp \left\{ - \sum_{i=1}^{t_n} \mathbb{P}(M_{0,s_n-p_n}^i > x_n, M_{s_n-p_n,s_n}^i \leq x_n) \right\} + o(1) \tag{57} \\ &\geq \exp \left\{ - \sum_{i=1}^{t_n} \sum_{j=1}^{s_n} \mathbb{P}(X_j^i > x_n, M_{j,j+p_n}^i \leq x_n) \right\} + o(1) \end{aligned} \tag{58}$$

with Eq. 58 following from Eq. 57 by the inclusions $\{M_{0,s_n-p_n}^i > x_n, M_{s_n-p_n,s_n}^i \leq x_n\} \subseteq \bigcup_{j=1}^{s_n-p_n} \{X_j^i > x_n, M_{j,j+p_n}^i \leq x_n\} \subseteq \bigcup_{j=1}^{s_n} \{X_j^i > x_n, M_{j,j+p_n}^i \leq x_n\}$ and the union bound. A similar argument that was used to show (52) gives

$$\sum_{j=1}^n \mathbb{P}(X_j > x_n, M_{j,j+p_n} \leq x_n) = \sum_{i=1}^{t_n} \sum_{j=1}^{s_n} \mathbb{P}(X_j^i > x_n, M_{j,j+p_n}^i \leq x_n) + o(1) \tag{59}$$

so that Eq. 58 becomes

$$\mathbb{P}(M_n \leq x_n) \geq \exp \left\{ - \sum_{j=1}^n \mathbb{P}(X_j > x_n, M_{j,j+p_n} \leq x_n) \right\} + o(1) \tag{60}$$

and so Eqs. 53 and 60 together prove (11). Also, since

$$\exp \left\{ - \sum_{j=1}^n \mathbb{P}(X_j > x_n, M_{j,j+p_n} \leq x_n) \right\} = \left[\exp \left\{ - n \bar{F}(x_n) \right\} \right]^{\gamma_n}$$

with $\gamma_n = n^{-1} \sum_{j=1}^n \mathbb{P}(M_{j,j+p_n} \leq x_n \mid X_j > x_n)$, this also gives (12).

5.5 Proof of Theorem 2.

Throughout we let $\theta_{i,n} = \mathbb{P}(M_{i,i+p_n} \leq x_n \mid X_i > x_n)$. From Corollary (1) we know that $\mathbb{P}(M_n \leq x_n) \rightarrow e^{-\tau\gamma'}$ with $\gamma' = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \theta_{i,n}$ which is easily seen to converge to the same value as $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \theta_i$ since

$$\begin{aligned} |n^{-1} \sum_{i=1}^n \theta_i - n^{-1} \sum_{i=1}^n \theta_{i,n}| &\leq n^{-1} \sum_{i=1}^n |\theta_{i,n} - \theta_i| \\ &\leq \max_{1 \leq i \leq n} |\theta_{i,n} - \theta_i| \rightarrow 0. \end{aligned} \tag{61}$$

This establishes the first part of the Theorem.

To show that, $\mathbb{P}(M_n \leq y_n) \rightarrow e^{-\tau'\gamma}$, define $n' = \lfloor (\tau'/\tau)n \rfloor$ so that $n'\bar{F}(x_n) \rightarrow \tau'$ or equivalently $F(x_n)^{n'} \rightarrow e^{-\tau'}$. Then,

$$\begin{aligned} |\mathbb{P}(M_{n'} \leq x_n) - \mathbb{P}(M_{n'} \leq y_{n'})| &\leq n'|F(x_n) - F(y_{n'})| \\ &= n'|\bar{F}(x_n) - \bar{F}(y_{n'})| \rightarrow 0. \end{aligned} \tag{62}$$

Since $n' \leq n$ we have by Theorem 1, $\mathbb{P}(M_{n'} \leq x_n) = F(x_n)^{n'\gamma'_n}$ where $\gamma'_n = (n')^{-1} \sum_{i=1}^{n'} \theta_{i,n}$ and γ'_n also has limiting value $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \theta_i$ since $|(n')^{-1} \sum_{i=1}^{n'} \theta_{i,n} - (n')^{-1} \sum_{i=1}^{n'} \theta_i| \rightarrow 0$ as in Eq. 61. Then, since $F(x_n)^{n'} \rightarrow e^{-\tau'}$, we have $\mathbb{P}(M_{n'} \leq x_n) \rightarrow e^{-\tau'\gamma}$ and so from Eq. 62, $\mathbb{P}(M_n \leq y_n) \rightarrow e^{-\tau'\gamma}$ also with γ as in Eq. 17.

5.6 Proof of theorem 3.

The first step in the proof is to show that we have, for each integer $k \geq 1$,

$$\mathbb{P}(M_n \leq x_n) - \mathbb{P}^k(M_{n'} \leq x_n) \rightarrow 0 \tag{63}$$

where $n' = \lfloor n/k \rfloor$. To do this we define intervals I_i and I_i^* , $1 \leq i \leq k$, of alternating large and small lengths as follows,

$$I_i = \{(i-1)n' + 1, \dots, in' - q_n\}, \quad I_i^* = \{in' - q_n + 1, \dots, in'\}. \tag{64}$$

We show that

$$|\mathbb{P}(M_n \leq x_n) - \mathbb{P}(M(\cup_{i=1}^k I_i) \leq x_n)| \rightarrow 0, \tag{65}$$

$$|\mathbb{P}(M(\cup_{i=1}^k I_i) \leq x_n) - \prod_{i=1}^k \mathbb{P}(M(I_i) \leq x_n)| \rightarrow 0, \tag{66}$$

$$|\prod_{i=1}^k \mathbb{P}(M(I_i) \leq x_n) - \prod_{i=1}^k \mathbb{P}(M(I_i \cup I_i^*) \leq x_n)| \rightarrow 0, \tag{67}$$

and

$$|\prod_{i=1}^k \mathbb{P}(M(I_i \cup I_i^*) \leq x_n) - \mathbb{P}^k(M_{n'} \leq x_n)| \rightarrow 0, \tag{68}$$

from which Eq. 63 follows by the triangle inequality.

Equation 65 follows from $\{M_n \leq x_n\} \subseteq \{M(\cup_{i=1}^k I_i) \leq x_n\}$ and so $\mathbb{P}(M(\cup_{i=1}^k I_i) \leq x_n) - \mathbb{P}(M_n \leq x_n) \leq \mathbb{P}(\cup_{i=1}^k \{M(I_i^*) > x_n\}) \leq kq_n \bar{F}(x_n) \rightarrow 0$ since $q_n = o(n)$ and $\bar{F}(x_n) = \tau/n + o(n^{-1})$.

Equation 66 follows immediately from Lemma 1 as $I_j, 1 \leq j \leq k$, are distinct subintervals of $\{1, \dots, n\}$, and I_i and I_{i+1} are separated by q_n .

Equation 67 follows from $\{M(I_i \cup I_i^*) \leq x_n\} \subseteq \{M(I_i) \leq x_n\}$ and $0 \leq \mathbb{P}(M(I_i) \leq x_n) - \mathbb{P}(M(I_i \cup I_i^*) \leq x_n) \leq q_n \bar{F}(x_n) \rightarrow 0$ so that $\mathbb{P}(M(I_i) \leq x_n) = \mathbb{P}(M(I_i \cup I_i^*) \leq x_n) + o(1)$ for $1 \leq i \leq k$.

For Eq. 68, we first note that $\mathbb{P}(M(I_1 \cup I_1^*) \leq x_n) = \mathbb{P}(M'_n \leq x_n)$. Since $I_i \cup I_i^*$ is an interval of length n' , $1 \leq i \leq k$, we have by periodicity that $M(I_i \cup I_i^*) \stackrel{D}{=} M_{r,r+n'}$ for some $r \in \{0, 1, \dots, d-1\}$. Then for $2 \leq i \leq k$, we have

$$\begin{aligned} |\mathbb{P}(M(I_i \cup I_i^*) \leq x_n) - \mathbb{P}(M(I_1 \cup I_1^*) \leq x_n)| &= |\mathbb{P}(M_{r,r+n'} \leq x_n) - \mathbb{P}(M'_n \leq x_n)| \\ &\leq |\mathbb{P}(M_{r+n'} \leq x_n) - \mathbb{P}(M'_n \leq x_n)| + \\ &\quad |\mathbb{P}(M_{r,r+n'} \leq x_n) - \mathbb{P}(M_{r+n'} \leq x_n)| \\ &\leq r \bar{F}(x_n) + r \bar{F}(x_n) \\ &\leq 2d \bar{F}(x_n) \rightarrow 0. \end{aligned}$$

Hence $\prod_{i=1}^k \mathbb{P}(M(I_i \cup I_i^*) \leq x_n) = \prod_{i=1}^k \mathbb{P}(M(I_1 \cup I_1^*) \leq x_n) + o(1) = \mathbb{P}^k(M'_n \leq x_n) + o$, which establishes (63).

Now we note that since $\{X_n\}_{n=1}^\infty$ satisfies AIM(x_n) with $n\bar{F}(x_n) \rightarrow \tau$, it also satisfies AIM(y_n) whenever $n\bar{F}(y_n) \rightarrow \tau' \leq \tau$. This follows in the exact same way as for the $D(x_n)$ condition, see, e.g., Lemma 3.6.2. in Leadbetter et al. (1983). This fact together with Eq. 63 allows the proof to proceed in exactly the same manner as the proof of Theorem 3.7.1. in Leadbetter et al. (1983).

5.7 Proof of Theorem 4

For $n, q \in \mathbb{N}$ and $u \in \mathbb{R}$, we define the mixing coefficients $\alpha_{n,q}(u)$ by

$$\alpha_{n,q}(u) = \max |\mathbb{P}(M(I_1 \cup I_2) \leq u) - \mathbb{P}(M(I_1) \leq u)\mathbb{P}(M(I_2) \leq u)|$$

where the maximum is taken over intervals I_1 and I_2 that are separated by q , with $\min\{|I_1|, |I_2|\} \geq q, \min\{\min(I_1), \min(I_2)\} \geq 1$ and $\max\{\max(I_1), \max(I_2)\} \leq n$.

We first note that since both $q_n = o(n)$ and $0 \leq \alpha_n = \alpha_{n,q_n}(x_n) \leq \alpha_{n,q_n}^*(x_n) \rightarrow 0$, we can find a sequence of positive integers $p_n = o(n)$ such that $n\alpha_n = o(p_n)$ and $q_n = o(p_n)$ so that the conditions of Theorem 1 are satisfied.

Let $t > 0$ and write $k_n = \lfloor t/\bar{F}(x_n) \rfloor \sim tn/\tau$ so that for sufficiently large n , $k_n > p_n + q_n$. Now, fix $i \in \mathbb{N}$. For sufficiently large n we have

$$\mathbb{P}(M_{i+p_n, i+p_n+q_n} > x_n \mid X_i > x_n) \leq q_n \bar{F}(x_n) + \alpha_{n,q_n}^*(x_n) \rightarrow 0$$

so that

$$\mathbb{P}(M_{i,i+k_n} \leq x_n \mid X_i > x_n) = \mathbb{P}(M_{i,i+p_n} \leq x_n, M_{i+p_n+q_n, i+k_n} \leq x_n \mid X_i > x_n) + o(1).$$

In a similar way, since $\{M_{i+k_n} \leq x_n\} \subseteq \{M_{i+p_n+q_n, i+k_n} \leq x_n\}$, we have

$$\begin{aligned} \mathbb{P}(M_{i+p_n+q_n, i+k_n} \leq x_n) - \mathbb{P}(M_{i+k_n} \leq x_n) &= \mathbb{P}(M_{i+p_n+q_n} > x_n, M_{i+p_n+q_n, i+k_n} \leq x_n) \\ &\leq (i + p_n + q_n) \bar{F}(x_n) \rightarrow 0, \end{aligned}$$

so that $\mathbb{P}(M_{i+p_n+q_n,i+k_n} \leq x_n) = \mathbb{P}(M_{i+k_n} \leq x_n) + o(1)$. Now we can derive the limiting distribution of $\bar{F}(x_n)T_i(x_n)$. We have

$$\begin{aligned} \mathbb{P}(\bar{F}(x_n)T_i(x_n) > t) &= \mathbb{P}(T_i(x_n) > k_n) = \mathbb{P}(X_{i+1} \leq x_n, \dots, X_{i+k_n} \leq x_n \mid X_i > x_n) \\ &= \mathbb{P}(M_{i,i+k_n} \leq x_n \mid X_i > x_n) \\ &= \mathbb{P}(M_{i,i+p_n} \leq x_n, M_{i+p_n+q_n,i+k_n} \leq x_n \mid X_i > x_n) + o(1) \\ &= \mathbb{P}(M_{i,i+p_n} \leq x_n \mid X_i > x_n)\mathbb{P}(M_{i+p_n+q_n,i+k_n} \leq x_n \mid X_i > x_n, M_{i,i+p_n} \leq x_n) + o(1) \\ &= \{\theta_i + o(1)\} \{\mathbb{P}(M_{i+k_n} \leq x_n) + o(1)\} + o(1). \end{aligned} \tag{69}$$

Now we focus on the term $\mathbb{P}(M_{i+k_n} \leq x_n)$ appearing in Eq. 69. Since $\mathbb{P}(M_{k_n} \leq x_n) - \mathbb{P}(M_{i+k_n} \leq x_n) \leq i\bar{F}(x_n)$ we have $\mathbb{P}(M_{i+k_n} \leq x_n) = \mathbb{P}(M_{k_n} \leq x_n) + o(1)$. Since $k_n = O(n)$ we have $\alpha_{k_n,q_n}^*(x_n) \rightarrow 0$ and so $\alpha_{k_n,q_n}(x_n) \rightarrow 0$ also. Thus we may find a sequence $p'_n = o(n)$ such that $k_n\alpha_{k_n,q_n} = o(p'_n)$ and $q_n = o(p'_n)$, e.g., we may take $p'_n = \lfloor (k_n \max\{q_n, k_n\alpha_{k_n,q_n}(x_n)\})^{1/2} \rfloor$. Then applying Theorem 1 to the first k_n terms we get $\mathbb{P}(M_{k_n} \leq x_n) = F(x_n)^{k_n\gamma'_n}$ where $\gamma'_n = k_n^{-1} \sum_{j=1}^{k_n} \theta'_{j,n}$ with $\theta'_{j,n} = \mathbb{P}(M_{j,j+p'_n} \leq x_n \mid X_j > x_n)$.

We now verify that γ'_n has limiting value $\gamma = \lim_{n \rightarrow \infty} n^{-1} \sum_{j=1}^n \theta_j$. Define sequences a_n, b_n and k'_n by $a_n = \max\{p_n, p'_n\}, b_n = \min\{p_n, p'_n\}$ and $k'_n = k_n + a_n$ and note that $k'_n = O(n)$. Then with the usual notation $\theta_{j,n} = \mathbb{P}(M_{j,j+p_n} \leq x_n \mid X_j > x_n)$, we have for all $1 \leq j \leq k_n$ that, for sufficiently large n , $|\theta'_{j,n} - \theta_{j,n}| \leq \mathbb{P}(M_{j+b_n,j+a_n} > x_n \mid X_j > x_n) \leq |p_n - p'_n| \bar{F}(x_n) + \alpha_{k'_n,q_n}^*(x_n)$ where we have used the fact that $b_n > q_n$ for sufficiently large n and $\alpha_{k'_n,q_n}^*(u)$ is non-decreasing in n for fixed q and u . Therefore, $|k_n^{-1} \sum_{j=1}^{k_n} \theta_{j,n} - k_n^{-1} \sum_{j=1}^{k_n} \theta'_{j,n}| \leq |p_n - p'_n| \bar{F}(x_n) + \alpha_{k'_n,q_n}^*(x_n) \rightarrow 0$ and so $k_n^{-1} \sum_{j=1}^{k_n} \theta_{j,n}$ and $k_n^{-1} \sum_{j=1}^{k_n} \theta'_{j,n}$ have the same limit which equals γ since $|k_n^{-1} \sum_{j=1}^{k_n} \theta_{j,n} - k_n^{-1} \sum_{j=1}^{k_n} \theta_j| \leq \max_{1 \leq j \leq k_n} |\theta_j - \theta_{j,n}| \rightarrow 0$.

Finally, we have $\mathbb{P}(M_{i+k_n} \leq x_n) = \mathbb{P}(M_{k_n} \leq x_n) + o(1) = F(x_n)^{k_n(\gamma+o(1))} = [e^{-t} + o(1)]^{\gamma+o(1)}$ since $n\bar{F}(x_n) \rightarrow \tau$ implies that $k_n\bar{F}(x_n) \rightarrow t$ which in turn implies that $F(x_n)^{k_n} \rightarrow e^{-t}$. Substituting $\mathbb{P}(M_{i+k_n} \leq x_n) = e^{-\gamma t} + o(1)$ in Eq. 69 then gives the result.

Acknowledgements The authors would like to express their gratitude to an anonymous reviewer and the Editor whose helpful comments have greatly improved this paper.

Data availability The datasets generated and analysed during the current study are available from the corresponding author upon request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as

you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ancona-Navarrete, M.A., Tawn, J.A.: A comparison of methods for estimating the extremal index. *Extremes* **3**, 5–38 (2000)
- Athreya, K.B., Pantula, S.G.: Mixing properties of Harris chains and autoregressive processes. *J. Appl. Probab.* **23**, 880–892 (1986)
- Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J.: *Statistics of extremes, theory and applications*. Wiley (2004)
- Berman, S.M.: Limit theorems for the maximum term in stationary sequences. *Ann. Math. Stat.* **35**, 502–516 (1964). <https://doi.org/10.1214/aoms/1177703551>
- Bradley, R.C.: Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2**, 107–144 (2005)
- Chernick, M., Hsing, T., McCormick, W.: Calculating the extremal index for a class of stationary sequences. *Adv. Appl. Probab.* **23**, 835–850 (1991)
- Coles, S.G., Tawn, J.A., Smith, R.L.: A seasonal Markov model for extremely low temperatures. *Environmetrics* **5**, 221–239 (1994)
- Davydov, Y.A.: Mixing conditions for Markov chains. *Theory Probab. Appl.* **18**, 312–328 (1973)
- Eastoe, E.F., Tawn, J.A.: Modelling the distribution of the cluster maxima of exceedances of sub-asymptotic thresholds. *Biometrika* **99**, 43–55 (2012)
- Einmahl, J.H.J., de Haan, L., Zhou, C.: Statistics of heteroscedastic extremes. *J. R. Stat. Soc. B* **78**, 31–51 (2016)
- Ferro, C.A.T., Segers, J.: Inference for clusters of extreme values. *J. R. Stat. Soc. B* **65**, 545–556 (2003)
- Haan, L., Ferreira, A.: *Extreme value theory: An introduction*. Springer Series in Operations Research and Statistics. Springer (2006)
- Hsing, T.: On the characterization of certain point processes. *Stoch. Process. Appl.* **26**, 297–316 (1987)
- Hsing, T.: Estimating the parameters of rare events. *Stoch. Process. Appl.* **37**, 117–139 (1991)
- Hsing, T.: Extremal index estimation for a weakly dependent stationary sequence. *Ann. Stat.* **21**, 2043–2071 (1993). <https://doi.org/10.1214/aos/1176349409>
- Hüsler, J.: Asymptotic approximation of crossing probabilities of random sequences. *Wahrscheinlichkeitsth. Gebiete* **63**, 257–270 (1983)
- Hüsler, J.: Extreme values of nonstationary random sequences. *J. Appl. Probab.* **23**, 937–950 (1986)
- Juan Cai, J.: A nonparametric estimator of the extremal index. arXiv:1911.06674 (2019)
- Leadbetter, M.R.: On extreme values in stationary sequences. *Z. Wahr-sch. Verw. Gebiete* **28**, 289–303 (1974)
- Leadbetter, M.R.: Extremes and local dependence in stationary sequences. *Z. Wahr-sch. Verw. Gebiete* **65**, 291–306 (1983)
- Leadbetter, M.R., Lindgren, G., Rootzén, H.: *Extremes and related properties of random sequences and series*. Springer, New York (1983)
- Ledford, A.W., Tawn, J.A.: Diagnostics for dependence within time series extremes. *J. R. Stat. Soc., B* **65**, 521–543 (2003)
- Loynes, R.M.: Extreme values in uniformly mixing stationary stochastic processes. *Ann. Math. Stat.* **36**, 993–999 (1965)
- O'Brien, G.L.: Extreme values for stationary and Markov sequences. *Ann. Probab.* **15**, 281–291 (1987)
- Smith, R.L.: The extremal index for a Markov chain. *J. Appl. Probab.* **29**, 37–45 (1992)
- Smith, R.L., Weissman, I.: Estimating the extremal index. *J. R. Stat. Soc. B* **56**, 515–528 (1994)
- Smith, R.L., Tawn, J.A., Coles, S.G.: Markov chain models for threshold exceedances. *Biometrika* **84**, 249–268 (1997)

- Süveges, M.: Likelihood estimation of the extremal index. *Extremes* **10**, 41–55 (2007)
- Süveges, M., Davison, A.C.: Model misspecification in peaks over threshold analysis. *Ann. Appl. Stat.* **4**, 203–221 (2010)
- Watson, G.S.: Extreme values in samples from m -dependent stationary stochastic processes. *Ann. Math. Stat.* **25**, 798–800 (1954)

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.