CentER

Discussion Paper

No. 2006–07

RARE EVENTS, TEMPORAL DEPENDENCE, AND THE EXTREMAL INDEX

By Johan Segers

February 2006

ISSN 0924-7815

Tilburg University

Applied Probability Trust (3 February 2006)

RARE EVENTS, TEMPORAL DEPENDENCE, AND THE EXTREMAL INDEX

JOHAN SEGERS,* *Tilburg University*

Abstract

Classical extreme-value theory for stationary sequences of random variables can up to a large extent be paraphrased as the study of exceedances over a high threshold. A special role within the description of the temporal dependence between such exceedances is played by the extremal index. Parts of this theory can be generalized not only to random variables on an arbitrary state space hitting certain failure sets but even to a triangular array of rare events on an abstract probability space. In the case of M4 processes, or maxima of multivariate moving maxima, the arguments take a simple and direct form.

JEL: C13, C14

Keywords: block maximum; exceedance; extremal index; failure set; mixing condition; M4 process; rare event; stationary sequence

AMS 2000 Subject Classification: Primary 60G70

Secondary 62G32

1. Introduction

Many applied sciences require handling events with low probability but large, often disastrous impact. Of particular interest is the way in which such rare events interact: an unusually stormy day at a particular site may well be followed by another one at the same or a neighboring site; a large drop in a stock index may trigger similar negative movements in the next time period for the same or other financial time series. Which, then, are the principles governing these dependencies?

The theory developed in this paper is inspired by a concept from classical extreme-value theory. A stationary sequence of random variables $\{X_n\}$ is said to have extremal

* Postal address: Tilburg University, Department of Econometrics and Operations Research, P.O. Box 90153, NL-5000 LE Tilburg, the Netherlands (E-mail: *jsegers@uvt.nl*).

index $\theta \in [0,1]$ if for every $0 < \tau < \infty$ there exists a sequence of thresholds $\{u_n\}$ such that $n \Pr(X_1 > u_n) \to \tau$ and $\Pr(\max_{i=1,\dots,n} X_i \le u_n) \to \exp(-\tau\theta)$ as $n \to \infty$ (Leadbetter, 1983). The extremal index θ quantifies the strength of dependence between threshold exceedances $\{X_i > u_n\}$, with $\theta = 1$ corresponding to asymptotic independence and $\theta \downarrow 0$ to an increasing propensity of large observations to occur in clusters. In the context of multivariate time series, the extremal index makes its appearance in the asymptotic distribution of the vector of component-wise maxima (Nandagopalan, 1994; Smith and Weissman, 1996; Perfekt, 1997; Beirlant *et al.*, 2004, chapter 10).

As hinted at already in Nandagopalan (1994), one can in fact start from a stationary process on an arbitrary state space in which a sequence of failure sets represents ever more extreme states for the process. The extremal index, which now also depends on the sequence of failure sets, describes the strength of temporal dependence between failure-set hits. Even a further abstraction is possible to a triangular array of events every row of which satisfies a certain stationarity condition.

For a single row of events, the following quantities are of interest: the probability that none of the events occurs; the probability that the occurrence of an event is not followed in the near future by another one; the mean number of events that occur given there occurs at least one; conditionally on the occurrence of an event, the time until the occurrence of the next one. The relations between these quantities can be described in terms of various inequalities. These complement the assessment of the accuracy of the compound Poisson approximation for the empirical point process of exceedances in Barbour, Xia and Novak (2002). Further, these inequalities lead to asymptotic results, which serve on the one hand to formulate in the framework of rare events known characterizations of the extremal index (Leadbetter, 1983; O'Brien, 1987; Ferro and Segers, 2003), and on the other hand to complement various Poisson limit results for triangular arrays (Hüsler, 1993; Hüsler and Schmidt, 1996). Point process results will not be pursued in this paper as the dependence restrictions in force will be weaker than in the aforementioned papers.

The exposition starts in Section 2 with an essay on the multivariate extremal index of M4 processes. In this relatively simple example, short and direct arguments suffice to illustrate the more general theory. By way of intermediate step, results for a stationary sequence in an arbitrary state space are formulated in Section 3. The highest level of abstraction is achieved in Sections 4–6. The set-up and the notations in force are detailed in Section 4. The core of the paper is Section 5, containing asymptotic theory for dependence within a triangular array of rare events. The theory is based on a meticulous analysis leading to sharp inequalities in Section 6. The appendices, finally, contain some technical arguments.

2. Maxima of multivariate moving maxima

M4 processes, or maxima of multivariate moving maxima, provide an instructive example of how phenomena in the context of extremes of univariate stationary processes carry over to a more general setting. For such processes, direct arguments suffice to reveal the connection between the extremal index and temporal dependence between exceedances over high multivariate thresholds.

2.1. M4 processes

A d-variate random sequence $X_i = (X_{i,1}, \ldots, X_{i,d})$, where $i \in \mathbb{Z}$, the set of integers, is called an *M4 process* if it admits the representation

$$X_{i,j} = \max_{l \geq 1} \max_{p \in \mathbb{Z}} a_{l,i-p,j} Z_{l,p}, \quad \text{for } i \in \mathbb{Z}; j = 1, \ldots, d; \tag{2.1}$$

the variables $Z_{l,p}$, where $l = 1, 2, \ldots$ and $p \in \mathbb{Z}$, are independent standard Fréchet random variables, that is, $\Pr(Z_{l,p} \leq x) = \exp(-1/x)$ for $0 < x < \infty$, while the $a_{l,k,j}$ are non-negative numbers such that

$$\sum_{l \geq 1} \sum_{k \in \mathbb{Z}} a_{l,k,j} = 1, \quad \text{for } j = 1, \ldots, d. \tag{2.2}$$

Note that the process $\{X_i : i \in \mathbb{Z}\}$ is constructed as the maximum of a sequence of multivariate moving maximum processes, whence the acronym 'M4'.

The M4 process (2.1) is strictly stationary, its marginal distributions being standard Fréchet. The distribution function G_m of the md-variate vector $(X_1, \ldots, X_m)$ is given by

$$\begin{aligned} G_m(x_1, \ldots, x_m) &= \exp\{-V_m(x_1, \ldots, x_m)\}, \\ V_m(x_1, \ldots, x_m) &= \sum_{l \geq 1} \sum_{p \in \mathbb{Z}} \max_{i=1,\ldots,m} \max_{j=1,\ldots,d} \frac{a_{l,i-p,j}}{x_{i,j}}, \end{aligned}$$

for $x_1, \dots, x_m \in (0, \infty]^d$. In particular, all finite-dimensional distributions of the process are simple max-stable, that is, $\{G_m(tx_1, \dots, tx_m)\}^t = G_m(x_1, \dots, x_m)$ for every $0 < t < \infty$. Such a process is called max-stable in de Haan (1984).

M4 processes were introduced in Smith and Weissman (1996) in order to provide examples for the multivariate extremal index, to be defined below. See Zhang (2002) for applications of M4 processes to the modelling of financial time series.

2.2. Temporal dependence between high-threshold exceedances

An observation X_i is said to exceed the threshold x if $X_i \not\leq x$, that is, if $X_{i,j} > x_j$ for some $j = 1, \dots, d$. For M4 processes, we will analyse the temporal dependence between exceedances over threshold sequences of the form nx with $x_j > 0$ for every $j = 1, \dots, d$.

For positive integer n and for $x \in (0, \infty]^d$, put

$$V_n(x) := V_n(x, \dots, x) = \sum_{l \geq 1} \sum_{p \in \mathbb{Z}} \max_{i=1,\dots,n} \max_{j=1,\dots,d} \frac{a_{l,i-p,j}}{x_j}; \tag{2.3}$$

also put $V_0 \equiv 0$. The following lemma is of great use in the study of the temporal dependence between extremes of an M4 process.

Lemma 2.1. *For $x \in (0, \infty]^d$, the functions V_n in (2.3) satisfy*

$$\lim_{n \to \infty} \{V_n(x) - V_{n-1}(x)\} = \lim_{n \to \infty} V_n(nx) = \sum_{l \geq 1} \max_{k \in \mathbb{Z}} \max_{j=1,\dots,d} \frac{a_{l,k,j}}{x_j} =: W(x).$$

Proof. For $l \geq 1$ and $k \in \mathbb{Z}$, put $b_{l,k} = \max_{j=1,\dots,d} a_{l,k,j}/x_j$. We have

$$V_n(x) - V_{n-1}(x) = \sum_{l \geq 1} \sum_{p \in \mathbb{Z}} \left(\max_{i=1,\dots,n} b_{l,i-p} - \max_{i=1,\dots,n-1} b_{l,i-p} \right).$$

Writing $\lambda_+ = \max(\lambda, 0)$ for $\lambda \in \mathbb{R}$, we get

$$\begin{aligned} V_n(x) - V_{n-1}(x) &= \sum_{l \geq 1} \sum_{p \in \mathbb{Z}} \left(b_{l,n-p} - \max_{i=1,\dots,n-1} b_{l,i-p} \right)_+ \\ &= \sum_{l \geq 1} \sum_{k \in \mathbb{Z}} \left(b_{l,k} - \max_{i=1,\dots,n-1} b_{l,i+k-n} \right)_+ \\ &= \sum_{l \geq 1} \sum_{k \in \mathbb{Z}} \left(b_{l,k} - \max_{i=1,\dots,n-1} b_{l,k-i} \right)_+ . \end{aligned}$$

By the dominated convergence theorem,

$$\lim_{n \to \infty} \{V_n(x) - V_{n-1}(x)\} = \sum_{l \geq 1} \sum_{k \in \mathbb{Z}} \left(b_{l,k} - \max_{r<k} b_{l,r} \right)_+ .$$

The identity

$$\sum_{k \in \mathbb{Z}} \left(b_{l,k} - \max_{r<k} b_{l,r} \right)_+ = \max_{k \in \mathbb{Z}} b_{l,k}$$

yields $\lim_{n\to\infty}\{V_n(x) - V_{n-1}(x)\} = W(x)$. Further,

$$V_n(nx) = \frac{1}{n} V_n(x) = \frac{1}{n} \sum_{k=1}^{n} \{V_k(x) - V_{k-1}(x)\}.$$

Since the Cesàro transform of a converging sequence converges to the same limit as the original sequence, also $\lim V_n(nx) = W(x)$. This concludes the proof of Lemma 2.1.

For $x \in (0, \infty]^d \setminus \{(\infty, \dots, \infty)\}$, put

$$\theta(x) = \frac{W(x)}{V_1(x)} = \frac{\sum_{l\geq 1} \max_{k\in\mathbb{Z}} \max_{j=1,\dots,d} a_{l,k,j}/x_j}{\sum_{l\geq 1}\sum_{k\in\mathbb{Z}} \max_{j=1,\dots,d} a_{l,k,j}/x_j}. \tag{2.4}$$

This θ is called the *(multivariate) extremal index (function)* of the M4 process (2.1). It inherits all the familiar properties of the extremal index of a univariate stationary process.

Theorem 2.1. *Let $\{X_n\}$ be the M4 process of (2.1). For $x \in (0, \infty]^d \setminus \{(\infty, \dots, \infty)\}$,*

$$\begin{aligned} \Pr(\forall i = 1, \dots, n : X_i \leq nx) &= \{\Pr(X_1 \leq nx)\}^{n\theta(x)} + o(1) \\ &\to \exp\{-W(x)\}. \end{aligned} \tag{2.5}$$

If m_n is a positive integer sequence such that $m_n \to \infty$ and $m_n = o(n)$, then

$$\mathrm{E}\left[\sum_{i=1}^{m_n} \mathbf{1}(X_i \not\leq nx) \,\middle|\, \exists i = 1, \dots, m_n : X_i \not\leq nx \right] \to \frac{1}{\theta(x)}. \tag{2.6}$$

If s_n is a positive integer sequence such that $s_n \to \infty$ and $s_n/n \to \lambda \in [0, \infty]$, then

$$\Pr(\forall i = 2, \dots, s_n : X_i \leq nx \mid X_1 \not\leq nx) \to \theta(x) \exp\{-\lambda V_1(x)\theta(x)\}. \tag{2.7}$$

Proof. The proof relies on Lemma 2.1. First, $\Pr(\forall i = 1, \dots, n : X_i \leq nx) = \exp\{-V_n(nx)\} \to \exp\{-W(x)\}$, as well as $\{\Pr(X_1 \leq nx)\}^n = \exp\{-nV_1(nx)\} = \exp\{-V_1(x)\}$. Secondly,

$$\begin{aligned} &\mathrm{E}[\textstyle\sum_{i=1}^{m_n} \mathbf{1}(X_i \not\leq nx) \mid \exists i = 1, \dots, m_n : X_i \not\leq nx] \\ &= \frac{m_n \Pr(X_1 \not\leq nx)}{\Pr(\exists i = 1, \dots, m_n : X_i \not\leq nx)} \\ &= \frac{m_n[1 - \exp\{-V_1(nx)\}]}{1 - \exp\{-V_{m_n}(nx)\}} \\ &= \frac{n[1 - \exp\{-(1/n)V_1(x)\}]}{(n/m_n)[1 - \exp\{-(m_n/n)V_{m_n}(m_n x)\}]} \to \frac{V_1(x)}{W(x)}. \end{aligned}$$

Finally,

$$
\begin{aligned}
&\Pr(\forall i = 2, \ldots, s_n : X_i \le nx \mid X_1 \not\le nx) \\
&= \frac{\Pr(\forall i = 2, \ldots, s_n : X_i \le nx) - \Pr(\forall i = 1, \ldots, s_n : X_i \le nx)}{1 - \Pr(X_1 \le nx)} \\
&= \frac{\exp\{-V_{s_n - 1}(nx)\} - \exp\{-V_{s_n}(nx)\}}{1 - \exp\{-V_1(nx)\}} \\
&= \exp\{-V_{s_n}(nx)\} \frac{n[\exp\{V_{s_n}(nx) - V_{s_n-1}(nx)\} - 1]}{n[1 - \exp\{-V_1(nx)\}]} \\
&= \exp\left(-\frac{s_n}{n} V_{s_n}(s_n x)\right) \frac{n[\exp\{(1/n)(V_{s_n}(x) - V_{s_n-1}(x))\} - 1]}{n[1 - \exp\{-(1/n)V_1(x)\}]} \\
&\to \exp\{-\lambda W(x)\} \frac{W(x)}{V_1(x)}.
\end{aligned}
$$

This concludes the proof of Theorem 2.1.

Equation (2.5), due to Smith and Weissman (1996), states that the role played by the extremal index in the asymptotic distribution of the component-wise sample maximum is exactly similar as in the original definition for univariate sequences in Leadbetter (1983). Take x such that all but its jth coordinate are equal to infinity to arrive at the result that the extremal index of the jth coordinate process $\{X_{n,j} : n \in \mathbb{Z}\}$ is equal to $\theta_j = \sum_{l \ge 1} \max_{k \in \mathbb{Z}} a_{l,k,j}$.

By equation (2.6), the expected number of exceedances over a high threshold in a block with at least one exceedance converges to the reciprocal of the extremal index. For univariate stationary processes, this characterization is due to Leadbetter (1983).

Finally, equation (2.7) admits two interpretations. The case $s_n/n \to 0$ states that the probability that the exceedance $X_1 \not\le nx$ is followed by run of s_n non-exceedances converges to $\theta(x)$, a property originally discovered in O'Brien (1987). The case $s_n/n \to \lambda > 0$ can be reformulated as follows: denoting $T_x = \min\{i \ge 1 : X_{i+1} \not\le x\}$,

$$
\lim_{n\to\infty} \Pr[\{V_1(x)/n\} T_{nx} \ge \lambda \mid X_1 \not\le nx] = \theta(x) \exp\{-\lambda\theta(x)\}, \qquad \lambda > 0.
$$

In words, the normalized inter-arrival time $\{V_1(x)/n\}T_{nx}$ converges to the mixture distribution $\{1-\theta(x)\}\varepsilon_0 + \theta(x)\mathrm{Exp}(\theta(x))$, where ε_0 is a point mass at zero and $\mathrm{Exp}(\nu)$ is an exponential distribution with mean $1/\nu$. For univariate sequences, a similar property was exploited in Ferro and Segers (2003) to construct an estimator for the extremal index; see also chapter 10 of Beirlant *et al.* (2004).

3. Variables in general state space

3.1. Setting

Let $\{X_n : n \geq 1\}$ be a stationary sequence of random elements of a measurable space $(S, \mathcal{S})$ and let $B \in \mathcal{S}$. Think of the random elements X_n as representing the evolution of some system or process over time and of the set B as a failure set for which the events $\{X_i \in B\}$ have small probability but large repercussions if occurring. The archetypical situation is the one where the state space S is the real line and the failure set B is the open half-line (u, ∞), the event $\{X_i \in B\}$ corresponding to the threshold exceedance $\{X_i > u\}$. In the example of M4 processes in Section 2, the state space is $\mathbb{R}^d$ and the failure set is of the form $\{y \in \mathbb{R}^d : y \not\leq x\}$.

For $B \in \mathcal{S}$ and integer $m \geq 1$, consider the following probabilities related to the occurrence of the events $\{X_i \in B\}$:

$$\begin{aligned} p(B) &= \Pr(X_1 \in B), \\ p_m(B) &= \Pr(\exists i = 1, \ldots, m : X_i \in B), \\ q_m(B) &= 1 - p_m(B) = \Pr(\forall i = 1, \ldots, m : X_i \notin B). \end{aligned}$$

To avoid trivialities, assume $0 < p(B) < 1$. We will be interested in asymptotics arising from a sequence of failure sets $B_n \in \mathcal{S}$ such that the probability of a hit tends to zero, $p(B_n) \to 0$.

3.2. Quantities of interest

From the above probabilities we can derive a number of quantities all of which describe in a different way the dependence between failure-set hits $\{X_i \in B\}$. If these events are independent, then simply $q_m(B) = \{q_1(B)\}^m$. In general however, $q_m(B) = \{q_1(B)\}^{m\theta}$ for some $\theta = \theta_m^{\mathrm{M}}(B) \geq 0$, or explicitly

$$\theta_m^{\mathrm{M}}(B) = \frac{\log q_m(B)}{m \log q_1(B)}.$$

If $p_m(B)$ is small, then $\theta_m^{\mathrm{M}}(B)$ is approximately equal to

$$\theta_m^{\mathrm{B}}(B) = \frac{p_m(B)}{mp(B)}.$$

Note that $\theta_m^{\mathrm{B}}(B)$ is equal to the reciprocal of the expected number of hits in the block

$X_1, \ldots, X_m$ given that there is at least one hit, $\mathrm{E}[\sum_{i=1}^m \mathbf{1}(X_i \in B) \mid \bigcup_{i=1}^m \{X_i \in B\}] = mp(B)/p_m(B) = 1/\theta_m^{\mathrm{B}}$.

The conditional probability that a hit $\{X_1 \in B\}$ is followed by a run of non-hits is

$$\theta_m^{\mathrm{R}}(B) = \Pr\left(\forall i = 2, \ldots, m : X_i \notin B \mid X_1 \in B\right) = \frac{p_m(B) - p_{m-1}(B)}{p(B)}.$$

Conditionally on the process starting with a hit, $\{X_1 \in B\}$, the waiting time until the next one is

$$T_B = \min\{i \geq 1 : X_{i+1} \in B\}.$$

Its distribution is determined by

$$\Pr(T_B \geq m \mid X_1 \in B) = \theta_m^{\mathrm{R}}(B).$$

3.3. Long-range dependence

As our notation suggests, the quantities above turn out to be related – that is, provided the amount of long-range dependence is not too strong. To control the latter, we impose conditions on a kind of mixing coefficients measuring the force of dependence in a sample of size n between blocks of variables of size at least l and separated by a gap of precisely s,

$$\alpha_{n,s,l}(B) = \max_{u,v,w} \left| \Pr\left(\bigcap_{u<i\leq v} \{X_i \notin B\} \cap \bigcap_{v<j\leq w} \{X_{j+s} \notin B\} \right) - q_{v-u}(B) q_{w-v}(B) \right|, \tag{3.1}$$

the maximum ranging over all integer u, v, w such that $u \geq 0$, $v \geq u + l$, $w \geq v + l$ and $w + s \leq n$; here l and s are positive integers such that $2l + s \leq n$. Abbreviate $\alpha_{n,l}(B) = \alpha_{n,l,l}(B)$ and $\bar{\alpha}_{n,l}(B) = \sup\{\alpha_{n,s,l}(B) : l \leq s \leq n - 2l\}$.

3.4. Characterization theorem

Let $B_n \in \mathcal{S}$ be such that $0 < p(B_n) < 1$. Theorem 3.1 states the relations between the quantities $\theta_m^{\mathrm{M}}(B_n)$, $\theta_m^{\mathrm{B}}(B_n)$, and $\theta_m^{\mathrm{R}}(B_n)$. It is an immediate corollary to the theorems in Section 5 applied to the events $A_{i,n} = \{X_i \in B_n\}$.

Theorem 3.1. *Assume there exists an integer sequence* $1 \leq l_n \leq n$ *such that* $l_n = o(n)$ *and* $\alpha_{n,l_n}(B_n) \to 0$.

(i) If $l_n \le m_n \le n$ *is an integer sequence such that* $l_n = o(m_n)$ *and* $\alpha_{n,l_n} = o[\max\{m_n/n, p_{m_n}(B_n)\}]$*, then*

$$q_n(B_n) = \{q_{m_n}(B_n)\}^{n/m_n} + o(1).$$

In particular, $\liminf q_n(B_n) \ge \exp\{-\limsup np(B_n)\}$.

(ii) If additionally $0 < \liminf np(B_n) \le \limsup np(B_n) < \infty$*, then* $\limsup \theta_n^{\mathrm{M}}(B_n) \le 1$ *and*

$$\lim_{n\to\infty} \sup_{m_n \le i \le j \le n} \left|\theta_i^{\mathrm{M}}(B_n) - \theta_j^{\mathrm{M}}(B_n)\right| = 0.$$

(iii) If additionally $m_n = o(n)$*, then*

$$\theta_n^{\mathrm{M}}(B_n) = \theta_{m_n}^{\mathrm{B}}(B_n) + o(1) = \theta_{m_n}^{\mathrm{R}}(B_n) + o(1).$$

(iv) If additionally $\bar{\alpha}_{\lambda n,l_n}(B_n) = o(1)$ *for every* $\lambda > 0$*, then for any sequence* θ_n *such that* $\theta_n = \theta_n^{\mathrm{M}}(B_n) + o(1)$*,*

$$\theta^{\mathrm{R}}_{\lceil x/p(B_n)\rceil}(B_n) = \Pr\{p(B_n)T_{B_n} \ge x \mid X_1 \in B_n\} = \theta_n \exp(-x\theta_n) + o(1)$$

locally uniformly in $0 < x < \infty$.

Remark 3.1. The condition that the process $\{X_n\}$ is stationary can be slightly weakened. It is sufficient that for all positive integers m and n the probabilities $\Pr(\forall i = 1, \ldots, m : X_{i+j} \in B_n)$ do not depend on j; see also Definition 4.1 below.

Example 3.1. Without additional assumptions, M4 processes (2.1) satisfy a kind of mixing condition for rare events making Theorem 3.1 available for many other failure sets than those of the form $\{y \in \mathbb{R}^d : y \not\le x\}$. For $x \in \mathbb{R}^d$ and $\lambda \in \mathbb{R}$, put, in obvious notation, $\max(x, \lambda) = (\max(x_1, \lambda), \ldots, \max(x_d, \lambda))$. For $-\infty < r \le s < \infty$ and $\lambda \in \mathbb{R}$, let $\sigma(r, s; \lambda)$ be the σ-field generated by the random vectors $\{\max(X_i, \lambda) : i \in \mathbb{Z} \cap [r, s]\}$. With these notations, every M4 process satisfies

$$\max_{s=1,\ldots,\nu n - l_n} \sup_{\substack{A \in \sigma(1,s;n\varepsilon) \\ B \in \sigma(s+l_n,\nu n;n\varepsilon)}} |\Pr(A \cap B) - \Pr(A)\Pr(B)| \to 0 \tag{3.2}$$

for every $0 < \nu < \infty$, every positive integer sequence $l_n = 1, \ldots, \nu n$ tending to infinity and every $0 < \varepsilon < \infty$. The proof of (3.2) is given in Appendix B. It is even possible to replace ε by a positive sequence ε_n tending to zero sufficiently slowly.

Note that for a finite set I of integers and for $u \in \mathbb{R}^d$, the event $\bigcap_{i \in I} \{X_i \leq u\}$ is contained in the σ-field $\sigma(r, s; \lambda)$ as soon as $I \subset \mathbb{Z} \cap [r, s]$ and $u_j \geq \lambda$ for every $j = 1, \ldots, d$. In particular, by (3.2) all M4 processes satisfy the multivariate version of Leadbetter's $D(u_n)$ condition for every multivariate threshold sequence u_n such that $\liminf u_{n,j}/n > 0$ for every $j = 1, \ldots, d$.

4. Rare events: assumptions and notations

Theorem 3.1 can be formulated completely in terms of the events $A_{i,n} = \{X_i \in B_n\}$; no reference needs to be made to the state space, the failure sets, or the random process. All we need is a triangular array $\{A_{i,n} : n \geq 1, 1 \leq i \leq n\}$ of events together with a notion of stationarity and restrictions on the amount of long-range dependence. The principal aim of this paper is to develop a theory of temporal dependence between rare events on this abstract level. In this section, we gather the ingredients that will appear in such a theory. The main results are stated in Section 5.

4.1. Block-stationarity

Throughout, we will work with the following notion of stationarity for a vector of events $A_1, \ldots, A_r$.

Definition 4.1. Events $A_1, \ldots, A_r$ on a common probability space are called *block-stationary* if $\Pr(\bigcup_{i=1}^m A_{i+j}) = \Pr(\bigcup_{i=1}^m A_i)$ for $m = 1, \ldots, r-1$ and $j = 1, \ldots, r-m$.

The probability that at least one of m consecutive events occurs is equal to

$$p_m = \Pr\left(\bigcup_{i=1}^m A_{i+j}\right), \qquad m = 1, \ldots, r, \quad j = 0, \ldots, r-m. \tag{4.1}$$

The probability, then, that none of m consecutive events occurs is

$$q_m = 1 - p_m = \Pr\left(\bigcap_{i=1}^m A_{i+j}^c\right), \qquad m = 1, \ldots, r, \quad j = 0, \ldots, r-m. \tag{4.2}$$

For simplicity, write $p = p_1$. To avoid trivialities, we assume henceforth $0 < p < 1$. For positive integers i and j with $i + j \leq r$,

$$p_i \leq p_{i+j} \leq p_i + p_j \quad \text{and} \quad q_{i+j} \leq q_i \leq q_{i+j} + p_j.$$

Remark 4.1. If $r \geq 4$, then the property that events $A_1, \ldots, A_r$ are block-stationary does not imply that the vector of indicator variables $\mathbf{1}(A_1), \ldots, \mathbf{1}(A_r)$ is stationary. See Examples 4.1 and 4.2 for some counterexamples. For the special case of $r = 3$ events, block-stationarity is the same as stationarity of the indicator variables, as kindly pointed out by a referee.

Example 4.1. Consider the discrete probability space $\Omega = \{1, 2, \ldots, 16\}$ with uniform probabilities, and put

$$\begin{aligned} A_1 &= \{1, \ldots, 8\}, \\ A_2 &= \{1, \ldots, 4\} \cup \{9, \ldots, 12\}, \\ A_3 &= \{1, 2\} \cup \{5, 6\} \cup \{9, 10\} \cup \{13, 14\}, \\ A_4 &= \{1, 2\} \cup \{9, 10\} \cup \{3, 7, 11, 15\}. \end{aligned}$$

Then $\Pr(A_i) = 1/2$ for $i = 1, \ldots, 4$, $\Pr(A_i \cup A_{i+1}) = 3/4$ for $i = 1, 2, 3$ and $\Pr(A_i \cup A_{i+1} \cup A_{i+2}) = 7/8$ for $i = 1, 2$. Hence the events $A_1, \ldots, A_4$ are block-stationary. However, $\Pr(A_1 \cup A_3) = 3/4$ while $\Pr(A_2 \cup A_4) = 5/8$, so the vector of corresponding indicator variables is not stationary.

Example 4.2. Let Y_n, where $n \in \mathbb{Z}$, be independent standard Fréchet random variables, $\Pr(Y_n \leq y) = \exp(-1/y)$ for $0 < y < \infty$. Further, let a_i, where $i \geq 0$, be non-negative numbers such that $a_i \geq a_{i+1}$ for $i \geq 0$ and $\sum_{i \geq 0} a_i = 1$. For positive integer n, put $\xi_n = \max\{a_i Y_{n-i} : i \geq 0\}$. The moving-maximum process $\{\xi_n\}$ is stationary and $\Pr\left(\max_{i=1,\ldots,n} \xi_i \leq x\right) = \exp[-\{(n-1)a_0 + 1\}/x]$ for $0 < x < \infty$.

Now let $\{\xi'_n\}$ be another such moving-maximum process, independent of $\{\xi_n\}$, and with parameters a'_i, $i \geq 0$, where again $a'_i \geq a'_{i+1} \geq 0$, $i \geq 0$, and $\sum_{i \geq 0} a'_i = 1$. Define a new process by intercalating $\{\xi_n\}$ and $\{\xi'_n\}$ through $(X_1, X_2, X_3, X_4, \ldots) = (\xi_1, \xi'_1, \xi_2, \xi'_2, \ldots)$. If $a_0 = a'_0$ but $a_i \neq a'_i$ for some $i \geq 1$, then the process $\{X_n\}$ is non-stationary. Nevertheless, the distribution of $\max\{X_{i+j} : i = 1, \ldots, m\}$ does not depend on j: for each real x, the events $A_i = \{X_i > x\}$ are block-stationary.

4.2. Quantities of interest

Let $A_1, \ldots, A_r$ be a row of block-stationary events (Definition 4.1). Recall p_m and q_m in (4.1) and (4.2). If the events are independent, then simply $q_m = q_1^m$ for all integer $1 \le m \le r$. In general, however, $q_m = q_1^{m\theta}$ for some $\theta = \theta_m^{\mathrm{M}} \ge 0$, or explicitly,

$$\theta_m^{\mathrm{M}} = \frac{\log(q_m)}{m \log(q_1)}, \qquad m = 1, \ldots, r. \tag{4.3}$$

If p_m is small, then $-\log(q_m)$ and $-\log(q_1)$ are approximately equal to p_m and p, respectively. Substituting these approximations into (4.3) yields

$$\theta_m^{\mathrm{B}} = \frac{p_m}{mp}, \qquad m = 1, \ldots, r. \tag{4.4}$$

Note that $0 < \theta_m^{\mathrm{B}} \le 1$. The interpretation is that $1/\theta_m^{\mathrm{B}} = mp/p_m$ is equal to the expected number of events that occur in a block of size m given there occurs at least one, $\mathrm{E}[\sum_{i=1}^m \mathbf{1}(A_i) \mid \bigcup_{i=1}^m A_i] = 1/\theta_m^{\mathrm{B}}$.

Conditionally on an event occurring, the probability that it is followed by a run of non-occurring events is equal to

$$\theta_m^{\mathrm{R}} = \Pr\left(\bigcap_{i=2}^m A_i^c \mid A_1\right) = \frac{p_m - p_{m-1}}{p}, \qquad m = 1, \ldots, r. \tag{4.5}$$

where $p_0 := 0$. By symmetry, θ_m^{R} is also equal to the probability that an extreme event is not preceded by another one for a certain time, $\theta_m^{\mathrm{R}} = \Pr(\bigcap_{i=1}^{m-1} A_i^c \mid A_m)$.

Finally, if the first event actually occurs, $\omega \in A_1$, then the time to wait until the next occurring event is equal to

$$T(\omega) = \min\{j \ge 1 : \omega \in A_{j+1}\};$$

the minimum of the empty set is set to infinity by convention. The distribution of the inter-arrival time T can be expressed as

$$\Pr(T \ge t \mid A_1) = \theta_t^{\mathrm{R}}, \qquad t = 1, \ldots, r. \tag{4.6}$$

The quantities θ_m^{M}, θ_m^{B}, and θ_m^{R} are ordered in the following way.

Lemma 4.1. *For integer* $1 \le m \le r$,

$$\theta_m^{\mathrm{R}} \le \theta_m^{\mathrm{B}} \le \theta_m^{\mathrm{M}} \le \theta_m^{\mathrm{B}}/q_m.$$

Proof. Since θ_i^{R} is decreasing in i,

$$p_m = \sum_{i=1}^m (p_i - p_{i-1}) = \sum_{i=1}^m p\theta_i^{\mathrm{R}} \ge mp\theta_m^{\mathrm{R}},$$

whence $\theta_m^{\mathrm{R}} \le \theta_m^{\mathrm{B}}$.

Next, the function $x \mapsto -x^{-1}\log(1-x) = \int_0^1 (1-xy)^{-1}\mathrm{d}y$ is increasing in $x < 1$. Since $p_m \ge p$, we get $-p_m^{-1}\log(q_m) \ge -p^{-1}\log(q_1)$ and thus $\log(q_m)/\log(q_1) \ge p_m/p$, whence $\theta_m^{\mathrm{M}} \ge \theta_m^{\mathrm{B}}$.

Finally, as $x \le -\log(1-x) \le x/(1-x)$ for $0 \le x < 1$, we have $-\log(q_m) \le p_m/q_m$ and $-\log(q_1) \ge p$, whence $\theta_m^{\mathrm{M}} \le (p_m/q_m)/(mp) = \theta_m^{\mathrm{B}}/q_m$, completing the proof.

4.3. Weak long-range dependence

The amount of long-range dependence will be controlled by putting bounds on the coefficients

$$\alpha_{s,l} := \max_{v=l,\dots,r-s-l} \max_{\substack{u=0,\dots,v-l\\ w=v+l,\dots,r-s}} \left| \Pr\left(\bigcap_{u<i\le v} A_i^c \cap \bigcap_{v<j\le w} A_{j+s}^c \right) - q_{v-u}q_{w-v} \right| \tag{4.7}$$

where $s = 0,\dots,r-2$ and $l = 1,\dots,\lfloor (r-s)/2 \rfloor$. The coefficient $\alpha_{s,l}$ describes the force of dependence between two blocks of length at least l and separated by a gap of size precisely s. Abbreviate $\alpha_l = \alpha_{l,l}$ and $\bar{\alpha}_l = \max\{\alpha_{s,l} : s = l,\dots,r-2l\}$.

The coefficients $\alpha_{s,l}$ were introduced by O'Brien (1987) in the classical setting of threshold exceedances $A_{i,n} = \{X_i > u_n\}$ in a stationary sequence $\{X_n\}$. More popular in this situation is Leadbetter's (1974) condition $D(u_n)$, which, in our notation, is based on the coefficients

$$\alpha_s^D := \max_{j=1,\dots,r-s-1} \max_{I,J} \left| \Pr\left(\textstyle\bigcap_{i\in I\cup J} A_i^c\right) - \Pr\left(\textstyle\bigcap_{i\in I} A_i^c\right) \Pr\left(\textstyle\bigcap_{i\in J} A_i^c\right) \right|$$

($s = 0,\dots,r-2$), the maximum being over all non-empty subsets $I \subset \{1,\dots,j\}$ and $J \subset \{j+s+1,\dots,r\}$. Clearly $\max\{\alpha_{t,l} : t = s,\dots,r-2l\} \le \alpha_s^D$, $s = 0,\dots,r-2$, so that dependence restrictions based on $\alpha_{s,l}$ are milder than the corresponding ones based on α_s^D. This improvement is useful for example for certain periodic Markov chains (O'Brien, 1987, p. 287).

Observe that α_s^D is in turn smaller than

$$\alpha_s^\Delta = \max_{j=1,\dots,r-s-1} \max_{E,F} |\Pr(E\cap F) - \Pr(E)\Pr(F)|, \qquad s = 0,\dots,r-2,$$

the maximum being over all $E \in \sigma(A_1,\dots,A_j)$ and $F \in \sigma(A_{j+s+1},\dots,A_r)$. Bounds on α_s^Δ are typically needed to establish convergence of empirical point processes of exceedances to a compound Poisson process (Hsing, Hüsler and Leadbetter, 1988; Barbour, Novak and Xia, 2002; Novak, 2002).

4.4. Triangular array of rare events

The set-up for asymptotic results will be a triangular array $A_{i,n}$, $n = 1, 2, \ldots$ and $i = 1, \ldots, r_n$, for which every row $A_{1,n}, \ldots, A_{r_n,n}$ consists of block-stationary events on a common probability space, which may vary with n. The probabilities of interest are $p_{m,n} = \Pr(\bigcup_{i=1}^m A_{i+j,n})$, $m = 1, \ldots, r_n$ and $j = 0, \ldots, r_n - m$, together with $q_{m,n} = 1 - p_{m,n}$ and $p_n = p_{1,n}$. The mixing coefficient (4.7) for the nth row is $\alpha_{s,l,n}$, and we write $\alpha_{l,n} = \alpha_{l,l,n}$ and $\bar{\alpha}_{l,n} = \max\{\alpha_{s,l,n} : s = l, \ldots, r_n - 2l\}$. Assume $0 < p_n < 1$ for all n, and for $m = 1, \ldots, r_n$, put

$$\theta^{\mathrm{M}}_{m,n} = \frac{\log(q_{m,n})}{m \log(q_{1,n})}, \qquad \theta^{\mathrm{B}}_{m,n} = \frac{p_{m,n}}{m p_n} \quad \text{and} \quad \theta^{\mathrm{R}}_{m,n} = \frac{p_{m,n} - p_{m-1,n}}{p_n}, \tag{4.8}$$

where $p_{0,n} := 0$. The distribution of the inter-arrival time between the first event and the next one is

$$\Pr(T_n \geq t \mid A_{1,n}) = \theta^{\mathrm{R}}_{t,n}, \qquad t = 1, \ldots, r_n. \tag{4.9}$$

Finally, all asymptotic statements are to be understood as $n \to \infty$.

5. Main results

The case of M4 processes in Section 2 suggests that properties of the extremal index of a univariate stationary sequence carry over to more general contexts. In this section, proper reformulations will be shown to remain true in the general setting of a triangular array $A_{1,n}, \ldots, A_{r_n,n}$, $n \geq 1$, of row-wise block-stationary events as in Section 4.4. The proofs of the results in this section depend on the results in section 6 and are deferred to Appendix A.

5.1. Big and small blocks

For independent and identically distributed random variables $\{X_n\}$, the distribution of the sample maximum $M_n = \max(X_1, \ldots, X_n)$ is given by $\Pr(M_n \leq x) = \{\Pr(X_1 \leq x)\}^n$. In case the sequence is stationary, certain mixing conditions still guarantee that $\Pr(M_r \leq x)$ is close to $\{\Pr(M_s \leq x)\}^{r/s}$ provided r and s are large enough. As a consequence, for such sequences, the only non-degenerate weak limits of affinely normalized sample maxima are the extreme-value distributions (Leadbetter, 1974). The argument can be extended to the multivariate case (Hsing, 1989; Hüsler, 1990). In the general setting, then, a natural question is how closely the probability $q_{r_n,n}$ of no extreme event in a row is approximated by the probability $q_{s_n,n}^{r_n/s_n}$ of no extreme event in r_n/s_n independent smaller blocks of size s_n.

Theorem 5.1. *Assume there exists an integer sequence* $1 \le l_n \le r_n$ *such that* $l_n = o(r_n)$ *and* $\alpha_{l_n,n} = o(1)$*. For every integer sequence* $l_n \le s_n \le r_n$ *such that* $l_n = o(s_n)$ *and* $\alpha_{l_n,n} = o\{\max(s_n/r_n, p_{s_n,n})\}$*,*

$$q_{r_n,n} = q_{s_n,n}^{r_n/s_n} + o(1).$$

Theorem 5.1 applies to any integer sequence s_n such that $l_n \le s_n \le r_n$ and $\liminf s_n/r_n > 0$, and even to some integer sequences s_n such that $s_n = o(r_n)$: let for instance s_n be the integer part of $\max\{(l_n r_n)^{1/2}, \alpha_{l_n,n}^{1/2} r_n\}$.

5.2. Extremal index

For univariate stationary sequences, the extremal index, whenever it exists, is defined through the relation $\Pr(M_n \le u_n) = \{\Pr(X_1 \le u_n)\}^{n\theta} + o(1)$ for threshold sequences u_n such that $0 < \liminf n \Pr(X_1 > u_n) \le \limsup n \Pr(X_1 > u_n) < \infty$. The extremal index typically arises as the reciprocal of the limit of the expected number of exceedances in a cluster of exceedances (Leadbetter, 1983) and also as the limit probability that an exceedance is followed by a run of non-exceedances (O'Brien, 1987). These characterizations carry over to the general set-up of a triangular array of rare events. Recall the quantities $\theta_{m,n}^{\mathrm{M}}$, $\theta_{m,n}^{\mathrm{B}}$ and $\theta_{m,n}^{\mathrm{R}}$ in (4.8).

Theorem 5.2. *Assume there exists an integer sequence* $1 \le l_n \le r_n$ *such that* $l_n = o(r_n)$ *and* $\alpha_{l_n,n} = o(1)$*.*

(i) If $\tau = \limsup r_n p_n < \infty$*, then* $\liminf q_{r_n,n} \ge \exp(-\tau)$ *and* $\limsup \theta_{r_n,n}^{\mathrm{M}} \le 1$*.*

(ii) If moreover $\liminf r_n p_n > 0$*, then for every integer sequence* $l_n \le m_n \le r_n$ *such that* $l_n = o(m_n)$ *and* $\alpha_{l_n,n} = o(m_n/r_n)$*,*

$$\lim_{n\to\infty} \sup_{m_n \le i \le j \le r_n} \left|\theta_{i,n}^{\mathrm{M}} - \theta_{j,n}^{\mathrm{M}}\right| = 0.$$

If $p_{m_n,n} = o(1)$, then $\theta_{m_n,n}^{\mathrm{M}} \sim \theta_{m_n,n}^{\mathrm{B}}$ by Lemma 4.1. The following theorem relates these to $\theta_{m_n,n}^{\mathrm{R}}$.

Theorem 5.3. *Assume there exists an integer sequence $1 \le l_n \le r_n$ such that $l_n = o(r_n)$ and $\alpha_{l_n,n} = o(1)$.*

(i) For every integer sequence $l_n \le m_n \le (r_n - l_n)/2$ such that $l_n = o(m_n)$, $p_{m_n,n} = o(1)$ and $\alpha_{l_n,n} = o(m_n p_n)$,

$$\theta^{\mathrm{R}}_{m_n,n} = \theta^{\mathrm{B}}_{m_n,n} + o(1) = \theta^{\mathrm{M}}_{m_n,n} + o(1).$$

(ii) If also $\alpha_{l_n,n} = o(p_{m_n,n})$, then

$$\theta^{\mathrm{R}}_{m_n,n} \sim \theta^{\mathrm{B}}_{m_n,n} \sim \theta^{\mathrm{M}}_{m_n,n}.$$

By definition, $q_{r_n,n} = q_{1,n}^{r_n\theta_n}$ with $\theta_n = \theta^{\mathrm{M}}_{r_n,n}$. The following theorem states conditions guaranteeing $q_{r_n,n} = q_{1,n}^{r_n\theta_n} + o(1)$ for other choices of θ_n.

Theorem 5.4. *Assume there exists an integer sequence $1 \le l_n \le r_n$ such that $l_n = o(r_n)$ and $\alpha_{l_n,n} = o(1)$.*

(i) For every integer sequence $l_n \le m_n \le r_n$ such that $l_n = o(m_n)$, $m_n = o(r_n)$, and $\alpha_{l_n,n} = o\{\max(m_n/r_n, p_{m_n,n})\}$,

$$q_{r_n,n} = q_{1,n}^{r_n\theta_n} + o(1) = \exp(-r_n p_n \theta_n) + o(1) \tag{5.1}$$

for $\theta_n \in \{\theta^{\mathrm{M}}_{m_n,n}, \theta^{\mathrm{B}}_{m_n,n}\}$.

(ii) If additionally $p_{m_n,n} \to 0$, then the above equation remains true with $\theta_n = \theta^{\mathrm{R}}_{m_n,n}$.

Remark 5.1. Without the extra condition $p_{m_n,n} \to 0$, part (ii) of Theorem 5.4 is not true. Consider for example independent events with $p_n \to 0$, $r_n \sim p_n^{-3}$, and $m_n \sim p_n^{-2}$: on the one hand $q_{r_n,n} = (1-p_n)^{r_n} \to 0$, while on the other hand

$$r_n \theta^{\mathrm{R}}_{m_n,n} = p_n^{-3}(1-p_n)^{m_n-1} = p_n^{-3}\exp[-p_n^{-1}\{1+o(1)\}] \to 0.$$

The condition $p_{m_n,n} \to 0$ is implied by each of the following ones: (i) $m_n p_n \to 0$, (ii) $\limsup_{n\to\infty} r_n p_n < \infty$, and (iii) $\liminf_{n\to\infty} q_{r_n,n} > 0$. Regarding (i), just observe that $p_{m_n,n} \le m_n p_n$. Since $m_n = o(r_n)$, (ii) implies (i). And since $q_{r_n,n} = (1-p_{m_n,n})^{r_n/m_n} + o(1)$ by Theorem 5.1, also condition (iii) is sufficient.

5.3. Inter-arrival times

Next, we focus on the inter-arrival time T_n between the occurrence of the event $A_{1,n}$ and the occurrence of the first subsequent event, conditionally on $A_{1,n}$, see (4.9). Since the probability of a single event is p_n, the average inter-arrival time should be $1/p_n$, regardless of the dependence structure. Under certain conditions, the standardized inter-arrival time $p_n T_n$ converges weakly to a non-degenerate limit. Recall $\bar{\alpha}_{l,n} = \max\{\alpha_{s,l,n} : s = l, \ldots, r_n - 2l\}$, with $\alpha_{s,l,n}$ as in (4.7) for the row $A_{1,n}, \ldots, A_{r_n,n}$.

Theorem 5.5. *If* $0 < \liminf r_n p_n \leq \limsup r_n p_n < \infty$ *and if there exists an integer sequence* $1 \leq l_n \leq r_n$ *such that* $l_n = o(r_n)$ *and* $\bar{\alpha}_{l_n,n} = o(1)$*, then for every sequence* θ_n *such that* $\theta_n = \theta^{\mathrm{M}}_{r_n,n} + o(1)$*,*

$$\Pr(p_n T_n \geq x \mid A_{1,n}) = \theta_n \exp(-x\theta_n) + o(1) \tag{5.2}$$

locally uniformly in $0 < x < \liminf r_n p_n$*.*

By (5.2), the normalized inter-arrival time $p_n T_n$ is approximately distributed according to the mixture distribution $(1-\theta_n)\varepsilon_0 + \theta_n \mathrm{Exp}(\theta_n)$, where ε_0 is the point mass at zero and $\mathrm{Exp}(\theta_n)$ is the exponential distribution with mean $1/\theta_n$. The point mass at zero describes the inter-arrival times between events *within* a cluster, while the exponential part describes the inter-arrival times between *different* clusters. This interpretation is in accordance with the compound Poisson limit (established under stronger mixing conditions) for the empirical point process of occurrence times of exceedances over a high threshold in a univariate stationary sequence (Hsing, Hüsler and Leadbetter, 1988). It is exploited in Ferro and Segers (2003) in the construction of an estimator for the extremal index.

6. Finite-sample inequalities

The key to the asymptotic results of Section 5 is a collection of sharp inequalities in the setting of a single row $A_1, \ldots, A_r$ of block-stationary events as in Definition 4.1. Throughout this section, employ the notations of Sections 4.1, 4.2 and 4.3.

6.1. Big and small blocks

The first lemma exploits an idea by Loynes (1965): a large block can be broken into approximately independent smaller blocks by clipping out an asymptotically negligible number of events between the smaller blocks and invoking the appropriate mixing coefficients. By convention the sum over the empty set is equal to zero and the product over the empty set is equal to one.

Lemma 6.1. *For integer $a_1, b_1, \ldots, a_k, b_k \in \{0, \ldots, r\}$ such that there exists a positive integer l such that $b_i - a_i \geq l$ for all $i = 1, \ldots, k$ and $a_{i+1} - b_i = l$ for all $i = 1, \ldots, k-1$,*

$$-(\alpha_l + p_l) \sum_{i=2}^{k} \prod_{j=i+1}^{k} q_{b_j - a_j} \leq q_{b_k - a_1} - \prod_{i=1}^{k} q_{b_i - a_i} \leq \alpha_l \sum_{i=2}^{k} \prod_{j=i+1}^{k} q_{b_j - a_j}.$$

Proof. We proceed by induction on k. For $k = 1$, there is nothing to prove. Let $k \geq 2$. We have

$$q_{b_k - a_1} = \Pr\left(\bigcap_{i=a_1+1}^{b_k} A_i^c\right)$$
$$\begin{cases} \leq & \Pr\left(\bigcap_{i=a_1+1}^{b_{k-1}} A_i^c \cap \bigcap_{i=a_k+1}^{b_k} A_i^c\right), \\ \geq & \Pr\left(\bigcap_{i=a_1+1}^{b_{k-1}} A_i^c \cap \bigcap_{i=a_k+1}^{b_k} A_i^c\right) - \Pr\left(\bigcup_{i=b_{k-1}+1}^{a_k} A_i\right). \end{cases}$$

Moreover,

$$\left|\Pr\left(\bigcap_{i=a_1+1}^{b_{k-1}} A_i^c \cap \bigcap_{i=a_k+1}^{b_k} A_i^c\right) - q_{b_{k-1}-a_1} q_{b_k - a_k}\right| \leq \alpha_l.$$

Together, we find

$$q_{b_{k-1}-a_1} q_{b_k - a_k} - \alpha_l - p_l \leq q_{b_k - a_1} \leq q_{b_{k-1}-a_1} q_{b_k - a_k} + \alpha_l.$$

Apply the induction hypothesis on $q_{b_{k-1}-a_1}$ to conclude the proof.

A useful special case of Lemma 6.1 is when the sizes $b_i - a_i$ of the smaller blocks are all the same. For a real number x, denote by $\lfloor x \rfloor$ the largest integer not larger than x and by $\lceil x \rceil$ the smallest integer not smaller than x.

Lemma 6.2. *For integer $1 \leq l \leq m \leq r$ and $1 \leq k \leq \lfloor (r+l)/(m+l) \rfloor$,*

$$q_r \leq q_m^k + \frac{\alpha_l}{\max(m/r, p_m)}. \tag{6.1}$$

If also $2l + m \leq r$, then for $k = \lceil (r+l)/(m+l) \rceil$,

$$q_r \geq q_m^k - \frac{\alpha_l + p_l}{\max(m/r, p_m)}. \tag{6.2}$$

Proof. Let $k = 1, \ldots, \lfloor (r+l)/(m+l) \rfloor$ and set $a_i = (i-1)(m+l)$ and $b_i = a_i + m$ for $i = 1, \ldots, k$. The integers $a_1, b_1, \ldots, a_k, b_k$ satisfy the conditions of Lemma 6.1; in particular $b_k = km + (k-1)l \leq r$. Hence

$$-(\alpha_l + p_l) \sum_{i=2}^{k} q_m^{k-i} \leq q_{km+(k-1)l} - q_m^k \leq \alpha_l \sum_{i=2}^{k} q_m^{k-i}.$$

Now $\sum_{i=2}^k q_m^{k-i} = (1-q_m^{k-1})/(1-q_m)$. Further, for $0 < x < 1$ and $a \geq 1$ or $a = 0$, we have $1-x^a \leq \min\{a(1-x), 1\}$ by the mean value theorem, and thus $(1-x^a)/(1-x) \leq \min\{a, 1/(1-x)\}$. Hence, for $k = 1, \ldots, \lfloor (r+l)/(m+l) \rfloor$,

$$-(\alpha_l + p_l)\min(k-1, 1/p_m) \leq q_{km+(k-1)l} - q_m^k \leq \alpha_l \min(k-1, 1/p_m). \tag{6.3}$$

Since $q_r \leq q_{km+(k-1)l}$, we get (6.1).

Next, suppose that $2l + m \leq r$. Apply Lemma 6.1 on $a_1 = 0$, $b_1 = m$, $a_2 = m+l$, and $b_2 = r$ to find

$$q_r \geq q_m q_{r-m-l} - (\alpha_l + p_l).$$

Let $k = \lceil (r+l)/(m+l) \rceil$. Since $r - m - l \leq (k-1)(m+l) - l \leq r$, by the left-hand inequality of (6.3),

$$q_{r-m-l} \geq q_{(k-1)(m+l)-l} \geq q_m^{k-1} - (\alpha_l + p_l)\min(k-2, 1/p_m).$$

Combine the previous two displays to get

$$q_r \geq q_m^k - (\alpha_l + p_l)\{q_m \min(k-2, 1/p_m) + 1\} \geq q_m^k - (\alpha_l + p_l)\min(k-1, 1/p_m),$$

whence (6.2). This completes the proof of the lemma.

Lemma 6.2 leads to inequalities for $q_r - q_m^{r/m}$ in case m is small compared to r.

Lemma 6.3. *For positive integer $l \leq m \leq r$,*

$$q_r \leq q_m^{r/m} + \frac{\alpha_l}{\max(m/r, p_m)} + \frac{l}{m} + \frac{m}{r}. \tag{6.4}$$

If also $2l + m \leq r$, then

$$q_r \geq q_m^{r/m} - \frac{\alpha_l + p_l}{\max(m/r, p_m)} - \frac{l}{m} - \frac{m}{r}. \tag{6.5}$$

Proof. By the mean value theorem,

$$|x^a - x^b| \leq \max(1 - a/b, 1 - b/a), \qquad 0 \leq x \leq 1; a > 0; b > 0. \tag{6.6}$$

Let $k = \lfloor (r+l)/(m+l) \rfloor$. Since $(r-m)/(m+l) \leq k \leq r/m$,

$$q_m^k - q_m^{r/m} \leq 1 - mk/r \leq l/m + m/r.$$

Combine this with Lemma 6.2, eq. (6.1), to arrive at (6.4).

Next, suppose $2l + m \le r$. Put $k = \lceil (r+l)/(m+l) \rceil$. By (6.6),

$$|q_m^k - q_m^{r/m}| \le \max\{1 - mk/r, 1 - r/(mk)\}.$$

Since $r/(m+l) \le k < r/m + 1$, we have $1 - mk/r \le l/m$ and $1 - r/(mk) \le m/r$, whence $\max\{1 - mk/r, 1 - r/(mk)\} \le l/m + m/r$. Combine this with Lemma 6.2, eq. (6.2), to arrive at (6.5). This completes the proof of the lemma.

6.2. The extremal index

The quantities $\theta_m^{\mathrm{M}} = \log(q_m)/\{m \log(q_1)\}$ of (4.3) are approximately constant over a wide range of m.

Lemma 6.4. *For integer $1 \le l \le m \le r$ such that $2l + m \le r$, denoting $\tau = rp$ and $\varepsilon = (r/m)\alpha_l + (1+\tau)l/m + m/r$,*

$$\left|\theta_r^{\mathrm{M}} - \theta_m^{\mathrm{M}}\right| \le \frac{\varepsilon}{\tau\{\exp(-\tau) - (\tau/2)(m/r) - \varepsilon\}_+}.$$

Proof. By Lemma 6.3, $|q_r - q_m^{r/m}| \le \varepsilon$. Now $q_r = \exp\{r \log(q_1)\theta_r^{\mathrm{M}}\}$ and $q_m^{r/m} = \exp\{r \log(q_1)\theta_m^{\mathrm{M}}\}$. By the mean value theorem,

$$r|\log(q_1)| \left|\theta_r^{\mathrm{M}} - \theta_m^{\mathrm{M}}\right| \min(q_r, q_m^{r/m}) \le \varepsilon.$$

Since $q_r \ge q_m^{r/m} - \varepsilon$ and $|\log(q_1)| \ge p$,

$$\left|\theta_r^{\mathrm{M}} - \theta_m^{\mathrm{M}}\right| \le \frac{\varepsilon}{\tau(q_m^{r/m} - \varepsilon)_+}.$$

As $\exp(-ax) - (1-x)^a \le x/2$ for $0 \le x \le 1$ and $a \ge 1$, we have $q_m^{r/m} = (1-p_m)^{r/m} \ge \exp\{-(r/m)p_m\} - p_m/2$. Apply the inequality $p_m \le mp$ to conclude the proof.

In Lemma 4.1 we already saw that $\theta_m^{\mathrm{R}} \le \theta_m^{\mathrm{B}}$. Here is a converse inequality.

Lemma 6.5. *For integer $1 \le l \le m \le r$ such that $2m + l \le r$,*

$$\theta_m^{\mathrm{R}} \ge \theta_m^{\mathrm{B}} - \frac{p_m^2}{mp} - \frac{\alpha_l + p_l}{mp}. \tag{6.7}$$

Proof. We have

$$p_m = \Pr\left(\bigcup_{i=1}^m A_i \cap \bigcap_{i=m+1}^{2m+l} A_i^c\right) + \Pr\left(\bigcup_{i=1}^m A_i \cap \bigcup_{i=m+1}^{2m+l} A_i\right).$$

On the one hand,

$$\Pr\left(\bigcup_{i=1}^m A_i \cap \bigcap_{i=m+1}^{2m+l} A_i^c\right) = \sum_{i=1}^m \Pr\left(A_i \cap \bigcap_{j=i+1}^{2m+l} A_j^c\right) = \sum_{i=1}^m p\theta^{\mathrm{R}}_{2m+l-i+1} \le mp\theta^{\mathrm{R}}_m,$$

while on the other hand,

$$\Pr\left(\bigcup_{i=1}^m A_i \cap \bigcup_{i=m+1}^{2m+l} A_i\right) \le \Pr\left(\bigcup_{i=1}^m A_i \cap \bigcup_{i=m+l+1}^{2m+l} A_i\right) + p_l \le p_m^2 + \alpha_l + p_l.$$

Combine the previous three displays to get

$$mp\theta^{\mathrm{R}}_m \ge p_m - p_m^2 - \alpha_l - p_l.$$

Dividing by mp yields (6.7). This completes the proof of the lemma.

By definition, $q_r = q_1^{r\theta}$ with $\theta = \theta^{\mathrm{M}}_r$. The following lemma gives bounds on the error induced by choosing θ equal to θ^{B}_m or θ^{R}_m. Note that $q_1^{r\theta} \le \exp(-rp\theta)$ for $\theta \ge 0$.

Lemma 6.6. *For integer* $1 \le l \le m \le r$ *and for* $\theta_m \in \{\theta^{\mathrm{B}}_m, \theta^{\mathrm{R}}_m\}$,

$$q_r \le q_1^{r\theta_m} + \frac{\alpha_l}{\max(m/r, p_m)} + \frac{l}{m} + \frac{m}{r}. \tag{6.8}$$

If additionally $2l + m \le r$, *then*

$$q_r \ge \exp(-rp\theta^{\mathrm{B}}_m) - \frac{\alpha_l + p_l}{\max(m/r, p_m)} - \frac{l}{m} - 2\frac{m}{r}. \tag{6.9}$$

If additionally $2m + l \le r$, *then*

$$q_r \ge \exp(-rp\theta^{\mathrm{R}}_m) - 3\frac{\alpha_l + p_l}{\max(m/r, p_m)} - \frac{l}{m} - 2\frac{m}{r} - 2p_m. \tag{6.10}$$

Proof. Note that $1 - ax \le (1-x)^a$ for $0 \le x \le 1$ and $a \ge 1$. As $m\theta^{\mathrm{B}}_m = p_m/p \ge 1$,

$$q_m^{r/m} = (1-p_m)^{r/m} = (1 - m\theta^{\mathrm{B}}_m p)^{r/m} \le (1-p)^{r\theta^{\mathrm{B}}_m}.$$

Since $\theta^{\mathrm{B}}_m \ge \theta^{\mathrm{R}}_m$ by Lemma 4.1, also $q_m^{r/m} \le (1-p)^{r\theta^{\mathrm{R}}_m}$. In combination with Lemma 6.3, eq. (6.4), this leads to (6.8).

For the proof of (6.9), we start from Lemma 6.3, eq. (6.5). We need to find suitable lower bounds for $q_m^{r/m}$. For $0 \le x \le 1$ and $a \ge 1$,

$$\begin{aligned} 0 \le \exp(-ax) - (1-x)^a &\le \{\exp(-x) - (1-x)\} a \exp\{-(a-1)x\} \\ &\le \frac{x^2}{2} a \exp(1-ax) = \frac{1}{a}\frac{\exp(1)}{2}(ax)^2\exp(-ax) \le \frac{1}{a}, \end{aligned}$$

since $\sup_{y \geq 0} y^2 \exp(-y) = 4 \exp(-2)$. Hence

$$q_m^{r/m} = (1 - p_m)^{r/m} \geq \exp\{-(r/m)p_m\} - \frac{m}{r} = \exp(-rp\theta_m^{\mathrm{B}}) - \frac{m}{r}, \tag{6.11}$$

which, in combination with Lemma 6.3, eq. (6.5), yields (6.9).

Finally, we will use Lemma 6.5 on the difference between θ_m^{B} and θ_m^{R} to convert the lower bound for q_r in terms of θ_m^{B} into a lower bound in terms of θ_m^{R}. Since $\exp(z) = \{\exp(z/2)\}^2 \geq (1 + z/2)^2$ for $z \geq 0$, we have for $0 \leq x \leq y$,

$$\begin{aligned} 0 \leq \exp(-x) - \exp(-y) &= \int_x^y \exp(-z) \mathrm{d}z \\ &\leq \int_x^y (1 + z/2)^{-2} \mathrm{d}z \leq \frac{y - x}{1 + y/2}. \end{aligned}$$

Hence, by Lemma 6.5, eq. (6.7),

$$\begin{aligned} \exp(-rp\theta_m^{\mathrm{R}}) - \exp(-rp\theta_m^{\mathrm{B}}) &\leq \frac{rp(\theta_m^{\mathrm{B}} - \theta_m^{\mathrm{R}})}{1 + rp\theta_m^{\mathrm{B}}/2} \\ &\leq \frac{rp\{p_m^2/(mp) + (\alpha_l + p_l)/(mp)\}}{1 + (r/m)p_m/2} \\ &= \frac{(r/m)p_m^2 + (\alpha_l + p_l)/(m/r)}{1 + (r/m)p_m/2}. \end{aligned}$$

If, on the one hand, $m/r \geq p_m$, then

$$\exp(-rp\theta_m^{\mathrm{R}}) - \exp(-rp\theta_m^{\mathrm{B}}) \leq p_m + \frac{\alpha_l + p_l}{m/r},$$

while if, on the other hand, $m/r < p_m$, then

$$\exp(-rp\theta_m^{\mathrm{R}}) - \exp(-rp\theta_m^{\mathrm{B}}) \leq 2 \left(p_m + \frac{\alpha_l + p_l}{p_m} \right).$$

All in all,

$$\exp(-rp\theta_m^{\mathrm{R}}) - \exp(-rp\theta_m^{\mathrm{B}}) \leq 2 \left(p_m + \frac{\alpha_l + p_l}{\max(m/r, p_m)} \right).$$

Combine this with (6.11) to get

$$q_m^{r/m} \geq \exp(-rp\theta_m^{\mathrm{R}}) - \frac{m}{r} - 2p_m - 2\frac{\alpha_l + p_l}{\max(m/r, p_m)}.$$

This inequality, in combination with Lemma 6.3, eq. (6.5), yields (6.10). The proof of the lemma is complete.

6.3. Inter-arrival times

Conditionally on A_1, the distribution of the time T until the next event is

$$\Pr(T \geq s \mid A_1) = \Pr\left(\bigcap_{i=2}^{s} A_i^c \,|\, A_1\right) = \theta_s^{\mathrm{R}}, \qquad s = 1, \ldots, r,$$

see (4.6). We break up the block $\bigcap_{i=2}^{s} A_i^c$ into an initial smaller block $\bigcap_{i=2}^{m} A_i^c$ and a subsequent larger block $\bigcap_{i=m+1}^{s} A_i^c$. The next lemma demonstrates how to control the dependence between A_1 and the initial block on the one hand and the subsequent block on the other hand. Recall $\bar{\alpha}_l = \max\{\alpha_{s,l} : s = l, \ldots, r-2l\}$, with $\alpha_{s,l}$ as in (4.7).

Lemma 6.7. *For integer $1 \leq l \leq m \leq r$ such that $2m + l \leq r$ and for integer $m + l \leq s \leq r - m$,*

$$-\frac{\alpha_l + p_l}{mp} \leq \theta_s^{\mathrm{R}} - \theta_m^{\mathrm{B}} q_s \leq 2\frac{\bar{\alpha}_l}{mp} + p_m + p_l.$$

Proof. For integer $m + 1 \leq t \leq r$,

$$\Pr\left(\bigcup_{i=1}^{m} A_i \cap \bigcap_{i=m+1}^{t} A_i^c\right) = \sum_{k=1}^{m} \Pr\left(A_k \cap \bigcap_{i=k+1}^{t} A_i^c\right) = \sum_{k=1}^{m} p\theta_{t-k+1}^{\mathrm{R}},$$

so that

$$mp\theta_t^{\mathrm{R}} \leq \Pr\left(\bigcup_{i=1}^{m} A_i \cap \bigcap_{i=m+1}^{t} A_i^c\right) \leq mp\theta_{t-m}^{\mathrm{R}}.$$

Hence for integer $m + 1 \leq s \leq r - m$,

$$\Pr\left(\bigcup_{i=1}^{m} A_i \cap \bigcap_{i=m+1}^{s+m} A_i^c\right) \leq mp\theta_s^{\mathrm{R}} \leq \Pr\left(\bigcup_{i=1}^{m} A_i \cap \bigcap_{i=m+1}^{s} A_i^c\right).$$

Now

$$\begin{aligned} 0 &\leq \Pr\left(\bigcup_{i=1}^{m} A_i \cap \bigcap_{i=m+1}^{s} A_i^c\right) - \Pr\left(\bigcup_{i=1}^{m} A_i \cap \bigcap_{i=m+1}^{s+m} A_i^c\right) \\ &\leq \Pr\left(\bigcup_{i=1}^{m} A_i \cap \bigcup_{i=s+1}^{s+m} A_i\right) \leq p_m^2 + \alpha_{s-m,l}. \end{aligned}$$

Moreover,

$$\begin{aligned} 0 &\leq \Pr\left(\bigcup_{i=1}^{m} A_i \cap \bigcap_{i=m+l+1}^{s+m} A_i^c\right) - \Pr\left(\bigcup_{i=1}^{m} A_i \cap \bigcap_{i=m+1}^{s+m} A_i^c\right) \\ &\leq \Pr\left(\bigcup_{i=m+1}^{m+l} A_i\right) = p_l \end{aligned}$$

and, if $s \geq m + l$,

$$\left|\Pr\left(\bigcup_{i=1}^{m} A_i \cap \bigcap_{i=m+l+1}^{s+m} A_i^c\right) - p_m q_{s-l}\right| \leq \alpha_l.$$

Combine the four previous displays to get

$$p_m q_{s-l} - \alpha_l - p_l \le mp\theta_s^{\mathrm{R}} \le p_m q_{s-l} + \alpha_l + p_m^2 + \alpha_{s-m,l},$$

or, dividing by mp, as $p_m \le mp$,

$$\theta_m^{\mathrm{B}} q_{s-l} - \frac{\alpha_l + p_l}{mp} \le \theta_s^{\mathrm{R}} \le \theta_m^{\mathrm{B}} q_{s-l} + \frac{\alpha_l + \alpha_{s-m,l}}{mp} + p_m.$$

Use $q_s \le q_{s-l} \le q_s + p_l$ to conclude the proof.

Appendix A. Proofs for Section 5

Recall the setting of a triangular array of events as in Section 4.4. The probability $p_{l_n,n}$ of the blocks that are clipped away is dealt with in the next lemma.

Lemma A.1. *Let $1 \le l_n \le m_n \le r_n$ be integers with $l_n = o(m_n)$.*

(i) Let $0 < \lambda_n \to 0$. If $p_{m_n,n} = O(\lambda_n)$ and $\alpha_{l_n,n} = o(\lambda_n)$, then $p_{l_n,n} = o(\lambda_n)$.

(ii) If $0 < p_{m_n,n} \to 0$ and $\alpha_{l_n,n} = o(p_{m_n,n})$, then $p_{l_n,n} = o(p_{m_n,n})$.

Proof of Lemma A.1. (i) Let k be a positive integer. If n is large enough so that $(2k+1)l_n \le m_n$, then by Lemma 6.2, eq. (6.1), with the choices $l = l_n$, $m = l_n$, and $r = (2k+1)l_n$,

$$\begin{aligned} 1 - p_{m_n,n} &\le 1 - p_{(2k+1)l_n,n} \\ &\le (1 - p_{l_n,n})^k + (2k+1)\alpha_{l_n,n} \le \exp(-p_{l_n,n}k) + (2k+1)\alpha_{l_n,n}. \end{aligned}$$

If n is also large enough so that $p_{m_n,n} + (2k+1)\alpha_{l_n,n} < 1$, then, as $-\log(1-x) \le x/(1-x)$ for $x < 1$,

$$\begin{aligned} p_{l_n,n} &\le -\frac{1}{k}\log\{1 - p_{m_n,n} - (2k+1)\alpha_{l_n,n}\} \\ &\le \frac{1}{k}\frac{p_{m_n,n} + (2k+1)\alpha_{l_n,n}}{1 - p_{m_n,n} - (2k+1)\alpha_{l_n,n}}. \end{aligned}$$

Hence $\limsup p_{l_n,n}/\lambda_n \le k^{-1} \limsup p_{m_n,n}/\lambda_n$. Let $k \to \infty$ to see that $p_{l_n,n}/\lambda_n \to 0$.

(ii) Take $\lambda_n = p_{m_n,n}$ in (i).

Proof of Theorem 5.1. Without loss of generality, we can restrict n to a subsequence along which s_n/r_n converges to some limit $\lambda \in [0,1]$.

Suppose first $\lambda = 0$. By the first inequality of Lemma 6.3, $q_{r_n,n} \le q_{s_n,n}^{r_n/s_n} + o(1)$. Now consider a further subsequence along which $\mu_n := (r_n/s_n)p_{s_n,n}$ converges to some limit $\mu \in [0,\infty]$. If $\mu = \infty$, then $q_{s_n,n}^{r_n/s_n} = \{1-(s_n/r_n)\mu_n\}^{r_n/s_n} \to 0$ and hence also $q_{r_n,n} \to 0$ along this subsequence. If $\mu < \infty$, then, again along the subsequence, $p_{s_n,n} = O(s_n/r_n)$ and thus, by assumption, $\alpha_{l_n,n} = o(p_{s_n,n})$, whence, by the second inequality of Lemma 6.3 and by Lemma A.1(ii), also $q_{r_n,n} \ge q_{s_n,n}^{r_n/s_n} + o(1)$.

On the other hand, if $\lambda > 0$, then choose a positive integer sequence $l_n \le m_n \le s_n$ such that $l_n = o(m_n)$, $m_n = o(s_n)$ and $\alpha_{l_n,n} = o(m_n/s_n)$; take for instance m_n equal to the integer part of $\max\{(l_n s_n)^{1/2}, \alpha_{l_n,n}^{1/2} s_n\}$. By the case $\lambda = 0$, we have $q_{r_n,n} = q_{m_n,n}^{r_n/m_n} + o(1)$ and $q_{s_n,n} = q_{m_n,n}^{s_n/m_n} + o(1)$. As $r_n/s_n \sim 1/\lambda$, also $q_{s_n,n}^{r_n/s_n} = q_{s_n,n}^{1/\lambda} + o(1) = q_{m_n,n}^{(s_n/m_n)(1/\lambda)} + o(1) = q_{m_n,n}^{r_n/m_n} + o(1)$.

Proof of Theorem 5.2. (i) Let $l_n \le m_n \le r_n$ be an integer sequence such that $l_n = o(m_n)$, $m_n = o(r_n)$ and $\alpha_{l_n,n} = o(m_n/r_n)$; for instance, let m_n be the integer part of $\max\{(l_n r_n)^{1/2}, \alpha_{l_n,n}^{1/2} r_n\}$. By Theorem 5.1, $q_{r_n,n} = q_{m_n,n}^{r_n/m_n} + o(1)$. Since $r_n p_n = O(1)$ and $m_n = o(r_n)$, we have $p_{m_n,n} \le m_n p_n = o(1)$. Hence $q_{m_n,n}^{r_n/m_n} = \exp\{-(r_n/m_n)p_{m_n,n}\} + o(1) \ge \exp(-r_n p_n) + o(1)$.

Without loss of generality, suppose that $r_n p_n \to \tau \in [0,\infty)$. If $\tau = 0$, then $\theta_{r_n,n}^{\mathrm{M}} \le 1/q_{r_n,n} \to 1$ by Lemma 4.1. If $\tau > 0$, then $\theta_{r_n,n}^{\mathrm{M}} = \theta_{m_n,n}^{\mathrm{M}} + o(1)$ by Lemma 6.4 and $\theta_{m_n,n}^{\mathrm{M}} \le 1/q_{m_n,n} \to 1$ by Lemma 4.1.

(ii) Without loss of generality, suppose that $m_n = o(r_n)$; otherwise, apply a construction as in (i). We have to show that $\theta_{i_n,n}^{\mathrm{M}} = \theta_{j_n,n}^{\mathrm{M}} + o(1)$ for all positive integer sequences i_n and j_n such that $m_n \le i_n \le j_n \le r_n$. By restricting to a subsequence if necessary, we can assume that $i_n/r_n \to \lambda$ and $j_n/r_n \to \mu$ for some $0 \le \lambda \le \mu \le 1$. If $\lambda = 0$, then by Lemma 6.4, $\theta_{i_n,n}^{\mathrm{M}} = \theta_{r_n,n}^{\mathrm{M}} + o(1)$; similarly if $\mu = 0$. On the other hand, if $\lambda > 0$, then by Lemma 6.4, $\theta_{i_n,n}^{\mathrm{M}} = \theta_{m_n,n}^{\mathrm{M}} + o(1)$; similarly if $\mu > 0$. As moreover $\theta_{m_n,n}^{\mathrm{M}} = \theta_{r_n,n}^{\mathrm{M}} + o(1)$, we get $\theta_{i_n,n}^{\mathrm{M}} = \theta_{j_n,n}^{\mathrm{M}} + o(1)$ in all cases. The proof of the theorem is complete.

Proof of Theorem 5.3. (i) By Lemmas 4.1 and 6.5,

$$\theta^{\mathrm{B}}_{m_n,n} - \frac{p^2_{m_n,n}}{m_n p_n} - \frac{\alpha_{l_n,n} + p_{l_n,n}}{m_n p_n} \le \theta^{\mathrm{R}}_{m_n,n} \le \theta^{\mathrm{B}}_{m_n,n}.$$

Since $p_{m_n,n} \le m_n p_n$ and $p_{l_n,n} \le l_n p_n$, the conditions imply $\theta^{\mathrm{R}}_{m_n,n} = \theta^{\mathrm{B}}_{m_n,n} + o(1)$.

(ii) By the above display,

$$\theta^{\mathrm{B}}_{m_n,n} \left(1 - p_{m_n,n} - \frac{\alpha_{l_n,n} + p_{l_n,n}}{p_{m_n,n}}\right) \le \theta^{\mathrm{R}}_{m_n,n} \le \theta^{\mathrm{B}}_{m_n,n}.$$

As $p_{l_n,n} = o(p_{m_n,n})$ by Lemma A.1(ii), $\theta^{\mathrm{R}}_{m_n,n} \sim \theta^{\mathrm{B}}_{m_n,n}$. Moreover, by Lemma 4.1, $\theta^{\mathrm{B}}_{m_n,n} \le \theta^{\mathrm{M}}_{m_n,n} \le \theta^{\mathrm{B}}_{m_n,n}/q_{m_n,n} \sim \theta^{\mathrm{B}}_{m_n,n}$.

Proof of Theorem 5.4. By Lemma 6.6, eq. (6.8),

$$q_{r_n,n} \le q_{1,n}^{r_n\theta_n} + o(1) \le \exp(-r_n p_n \theta_n) + o(1),$$

for $\theta_n \in \{\theta^{\mathrm{B}}_{m_n,n}, \theta^{\mathrm{R}}_{m_n,n}\}$. Without loss of generality, fix a subsequence along which $p_{m_n,n}$ converges to some $p \in [0,1]$.

In case $p > 0$, since $\theta^{\mathrm{B}}_{m_n,n} \le \theta^{\mathrm{M}}_{m_n,n}$ (see Lemma 4.1),

$$\exp(-r_n p_n \theta^{\mathrm{M}}_{m_n,n}) \le \exp(-r_n p_n \theta^{\mathrm{B}}_{m_n,n}) = \exp\{-(r_n/m_n) p_{m_n,n}\} = o(1),$$

so that $q_{r_n,n}$, $q_{1,n}^{r_n\theta_n}$ and $\exp(-r_n p_n \theta_n)$ are all $o(1)$ for $\theta_n \in \{\theta^{\mathrm{B}}_{m_n,n}, \theta^{\mathrm{M}}_{m_n,n}\}$.

In case $p = 0$, then $p_{l_n,n} = o\{\max(m_n/r_n, p_{m_n,n})\}$ by Lemma A.1(i); hence, by Lemma 6.6, eqs. (6.9) and (6.10),

$$q_{r_n,n} \ge \exp(-r_n p_n \theta_n) + o(1) \ge q_{1,n}^{r_n\theta_n} + o(1)$$

for $\theta_n \in \{\theta^{\mathrm{B}}_{m_n,n}, \theta^{\mathrm{R}}_{m_n,n}\}$. In combination with the first display of this proof, this yields

$$q_{r_n,n} = \exp(-r_n p_n \theta_n) + o(1) \ge q_{1,n}^{r_n\theta_n} + o(1)$$

for $\theta_n \in \{\theta^{\mathrm{B}}_{m_n,n}, \theta^{\mathrm{R}}_{m_n,n}\}$. As $p_{m_n,n} \to 0$ implies $\theta^{\mathrm{M}}_{m_n,n} \sim \theta^{\mathrm{B}}_{m_n,n}$ by Lemma 4.1, the above display remains valid for $\theta_n = \theta^{\mathrm{M}}_{m_n,n}$ by the fact that $a_n^{1+\varepsilon_n} = a_n + o(1)$ for any real sequences $0 \le a_n \le 1$ and $\varepsilon_n \to 0$.

Proof of Theorem 5.5. Let $l_n \leq m_n \leq r_n$ be an integer sequence such that $l_n = o(m_n)$, $m_n = o(r_n)$ and $\bar{\alpha}_{l_n,n} = o(m_n/r_n)$. By Lemma 6.7,

$$\max\{|\theta^{\mathrm{R}}_{s,n} - \theta^{\mathrm{B}}_{m_n,n} q_{s,n}| : s = m_n + l_n, \ldots, r_n - m_n\} \to 0.$$

Hence for any integer sequence $m_n + l_n \leq s_n \leq r_n - m_n$, we have $\theta^{\mathrm{R}}_{s_n,n} = \theta^{\mathrm{B}}_{m_n,n} q_{s_n,n} + o(1)$. By Theorem 5.4, also $q_{s_n,n} = \exp(-s_n p_n \theta^{\mathrm{B}}_{m_n,n}) + o(1)$. For $0 < x < \liminf r_n p_n$, the sequence $s_n = \lceil x/p_n \rceil$ falls in the required range, whence

$$\theta^{\mathrm{R}}_{\lceil x/p_n \rceil,n} = \theta^{\mathrm{B}}_{m_n,n} \exp(-x\theta^{\mathrm{B}}_{m_n,n}) + o(1),$$

locally uniformly in $0 < x < \liminf r_n p_n$.

Observe that $\theta^{\mathrm{B}}_{m_n,n} = \theta^{\mathrm{M}}_{m_n,n} + o(1) = \theta^{\mathrm{M}}_{r_n,n} + o(1)$ [Theorems 5.3(i) and 5.2(ii)] and, for nonnegative θ, θ' and x, that $|\theta \exp(-x\theta) - \theta' \exp(-x\theta')| \leq |\theta - \theta'|$ to complete the proof.

Appendix B. Proof of equation (3.2)

Fix a positive integer m. For $i \in \mathbb{Z}$ and $j = 1, \ldots, d$, put

$$\begin{aligned} X^{(m)}_{i,j} &= \max_{l \geq 1} \max_{|k| < m} a_{l,k,j} Z_{l,i-k}, \\ R^{(m)}_{i,j} &= \max_{l \geq 1} \max_{|k| \geq m} a_{l,k,j} Z_{l,i-k}. \end{aligned}$$

Observe that $X_{i,j} = \max(X^{(m)}_{i,j}, R^{(m)}_{i,j})$.

Put $b_{l,k} = \max_{j=1,\ldots,d} a_{l,k,j}$. For $0 < \varepsilon < \infty$,

$$\begin{aligned} &\Pr(\exists i = 1, \ldots, \nu n,\ j = 1, \ldots, d : R^{(m)}_{i,j} > n\varepsilon) \\ &\quad = \Pr(\exists i = 1, \ldots, \nu n,\ j = 1, \ldots, d,\ l \geq 1,\ |k| \geq m : a_{l,k,j} Z_{l,i-k} > n\varepsilon) \\ &\quad = \Pr(\exists i = 1, \ldots, \nu n,\ l \geq 1,\ |k| \geq m : b_{l,k} Z_{l,i-k} > n\varepsilon) \\ &\quad = \Pr\left(\exists l \geq 1,\ p \in \mathbb{Z} : \max_{|k| \geq m, 1 \leq k+p \leq \nu n} b_{l,k} Z_{l,p} > n\varepsilon\right) \\ &\quad \leq \frac{1}{n\varepsilon} \sum_{l \geq 1} \sum_{p \in \mathbb{Z}} \max_{|k| \geq m, 1 \leq k+p \leq \nu n} b_{l,k}. \end{aligned}$$

Replacing the last maximum by a summation and interchanging the summation over p with the resulting summation over k gives

$$\Pr(\exists i = 1, \ldots, \nu n,\ j = 1, \ldots, d : R^{(m)}_{i,j} > n\varepsilon) \leq \frac{\nu}{\varepsilon} \sum_{l \geq 1} \sum_{|k| \geq m} b_{l,k}. \tag{B.1}$$

Put $X_i^{(m)} = (X_{i,1}^{(m)}, \ldots, X_{i,d}^{(m)})$. We have

$$\begin{aligned} & \Pr\{\forall i = 1, \ldots, \nu n : \max(X_i^{(m)}, n\varepsilon) = \max(X_i, n\varepsilon)\} \\ & \quad \geq \quad 1 - \Pr(\exists i = 1, \ldots, \nu n,\ j = 1, \ldots, d : R_{i,j}^{(m)} > n\varepsilon), \end{aligned}$$

so that by (B.1),

$$\Pr\{\forall i = 1, \ldots, \nu n : \max(X_i^{(m_n)}, n\varepsilon) = \max(X_i, n\varepsilon)\} \to 1$$

for every positive integer sequence m_n tending to infinity. Equation (3.2) now follows from the fact that the process $\{X_i^{(m)} : i \in \mathbb{Z}\}$ is $2m$-dependent in the sense that $\{X_i^{(m)} : i \leq r - m\}$ and $\{X_i^{(m)} : i \geq r + m\}$ are independent for every integer r.

Acknowledgements

Gratefully acknowledged are encouraging comments by John Einmahl and Jürg Hüsler, as well as a meticulous reading combined with useful and constructive suggestions by an anonymous referee. The author's research was supported by a post-doctoral grant of the Fund for Scientific Research, Flanders, and a VENI grant of the Dutch Science Foundation (NWO).

References

Barbour, A.D., Novak, S.Y., and Xia, A. (2002). Compound Poisson approximation for the distribution of extremes. *Adv. Appl. Prob.* **24**, 223–240.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of extremes: Theory and Applications.* Wiley.

Ferro, C.A.T. and Segers, J. (2003). Inference for clusters of extremes. *J. R. Statist. Soc.* B **65**, 545–556.

de Haan, L. (1984). A spectral representation for max-stable processes. *Ann. Prob.* **12**, 1194–1204.

Hsing, T. (1989). Extreme value theory for multivariate stationary sequences. *J. Multivar. Anal.* **29**, 274–291.

Hsing, T., Hüsler, J., and Leadbetter, M.R. (1988). On the exceedance point process for a stationary sequence. *Prob. Theory Rel. Fields* **78**, 97–112.

Hüsler, J. (1990). Multivariate extreme values in stationary random sequences. *Stoch. Proc. Appl.* **35**, 99–108.

Hüsler, J. (1993). A note on exceedances and rare events of non-stationary sequences. *J. Appl. Prob.* **30**, 877–888.

Hüsler, J. and Schmidt, M. (1996). A note on the point processes of rare events. *J. Appl. Prob.* **33**, 654–663.

Leadbetter, M.R. (1974). On extreme values in stationary sequences. *Z. Wahrscheinlichkeitsth.* **28**, 289–303.

Leadbetter, M.R. (1983). Extremes and local dependence of stationary sequences. *Z. Wahrscheinlichkeitsth.* **65**, 291–306.

Loynes, R.M. (1965). Extreme values in uniformly mixing stationary stochastic processes. *Ann. Math. Statist.* **36**, 993–999.

Nandagopalan, S. (1994). On the multivariate extremal index. *J. Res. Natl. Inst. Stand. Technol.* **99**, 543–550.

Novak, S.Y. (2002). Multilevel clustering of extremes. *Stoch. Proc. Appl.* **97**, 59–75.

O'Brien, G.L. (1987). Extreme values for stationary and Markov sequences. *Ann. Prob.* **15**, 281–291.

Perfekt, R. (1997). Extreme value theory for a class of Markov chains with values in $\mathbb{R}^d$. *Adv. Appl. Prob.* **29**, 138–164.

Smith, R.L. and Weissman, I. (1996). *Characterization and Estimation of the Multivariate Extremal Index.* www.unc.edu/depts/statistics/postscript/rs/extremal.ps.

Zhang, Z. (2002). *Multivariate Extremes, Max-Stable Process Estimation and Dynamic Financial Modeling.* Ph.D. thesis, University of North Carolina at Chapel Hill.