Clustering indices and decay of correlations in non-Markovian models

Miguel Abadi\textsuperscript{1,4}, Ana Cristina Moreira Freitas\textsuperscript{2,5} and Jorge Milhazes Freitas\textsuperscript{3,6}

\textsuperscript{1} Instituto de Matemática e Estatística, Universidade de São Paulo, Rua do Matão 1010, Cid. Universitaria, 05508-090—São Paulo, SP, Brasil
\textsuperscript{2} Centro de Matemática & Faculdade de Economia da Universidade do Porto, Rua Dr. Roberto Frias, 4200-464 Porto, Portugal
\textsuperscript{3} Centro de Matemática & Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre 687, 4169-007 Porto, Portugal

E-mail: leugim@ime.usp.br, amoreira@fep.up.pt and jmfreita@fc.up.pt

Received 20 December 2018, revised 19 June 2019
Accepted for publication 1 August 2019
Published 29 October 2019

Abstract

When there is no independence, abnormal observations may have a tendency to appear in clusters instead of being scattered along the time frame. Identifying clusters and estimating their size are important problems arising in the statistics of extremes or in the study of quantitative recurrence for dynamical systems. In the classical literature, the extremal index appears to be linked to the cluster size and, in fact, it is usually interpreted as the reciprocal of the mean cluster size. This quantity involves a passage to the limit and in some special cases this interpretation fails due to an escape of mass when computing limiting point processes. Smith (1988 \textit{Adv. Appl. Probab.} 20 681–3), introduced a regenerative process exhibiting such disagreement. Very recently, in Abadi (2018 (arXiv:1808.02970)) the authors used a dynamical mechanism to emulate the same inadequacy of the usual interpretation of the extremal index. Here, we consider a general regenerative process that includes Smith’s model, show that it is important to consider finite time quantities instead of asymptotic ones and compare their different behaviours in relation to the cluster size. We consider other indicators, such as what we call the sojourn time, which corresponds to the size of groups of abnormal observations, when there is some uncertainty regarding where the cluster containing that group

\textsuperscript{4} http://miguelabadi.wixsite.com/miguel-abadi
\textsuperscript{5} www.fep.up.pt/docentes/amoreira/
\textsuperscript{6} www.fc.up.pt/pessoas/jmfreita/
was actually initiated. We also study the decay of correlations of the non-Markovian models considered.

Keywords: extremal index, clustering, cluster size distribution, sojourn time
Mathematics Subject Classification numbers: 60G70, 37A50, 37B20, 37A25

1. Introduction

In extreme value theory, the convergence of the maxima of a sequence of i.i.d. random variables is a very well studied subject, and the book [15] is a major reference. The starting point is that for an \((X_n)_{n \in \mathbb{N}}\) sequence of i.i.d. random variables over a probability space \((\Omega, \mathcal{F}, \mathbb{P})\) with a cumulative distribution function \(F\), it is straightforward to see that for any real positive \(\tau\) and a real sequence \((u_n)_{n \geq 0}\) one has
\[
\lim_{n \to \infty} \mathbb{P}(\max\{X_1, \ldots, X_n\} \leq u_n) = e^{-\tau} \quad \text{if and only if} \quad \lim_{n \to \infty} n(1 - F(u_n)) = \tau.
\]
The three possible classical limits for the maximum follow.

The independent case is far from being able to model the real world, and a major effort to extend this result to dependent processes has been made in the last few decades. The principal ingredient is the appearance of the extremal index \(\theta\) verifying
\[
\lim_{n \to \infty} \mathbb{P}(\max\{X_1, \ldots, X_n\} \leq u_n) = e^{-\theta \tau} \quad \text{whenever} \quad \lim_{n \to \infty} n(1 - F(u_n)) = \tau.
\]
This new factor describes the capacity of a given maximum to produce subsequent ones, due to the correlation of the random variables, which is an ingredient that is absent in the i.i.d. case. It follows by the above property that the extremal index is such that \(\theta \in [0, 1]\), which is strictly smaller than one where abnormally high observations tend to appear in clusters rather than isolated, as in the i.i.d. case, where it is equal to one. However, this way of introducing the extremal index—as a limiting value through the above properties—gives rise to certain difficulties in its calculation or even estimation.

It is the purpose of this paper to show the relevance of considering the extremal index, not just as an asymptotic limit but rather as a quantity at a finite time, as well as to show the different behaviours that both cases may present.

There is a natural interpretation of the extremal index as a reciprocal of the mean cluster size distribution, and the reason is heuristically clear. Suppose that one wants to observe a cluster of exceedances at least \(k\) in size and of the level \(u_n\). That is
\[
\mathbb{P}(N \geq k) = \mathbb{P}(\cap_{i=1}^{k} \{X_i > u_n\}).
\]
Here, \(N\) stands for the number of consecutive observations corresponding to exceedances of the threshold \(u_n\). Given the events \(A_1, \ldots, A_k\), the universal formula
\[
\mathbb{P}(\cap_{i=1}^{k} A_i) = \prod_{i=1}^{k} \mathbb{P}(A_i | \cap_{j=1}^{i-1} A_j)
\]
says that if dependence with respect to the remote past is small and only the close past matters, the factors on the right-hand side in the above equality should be all about the same value. If the meaning of "close past" is quantified by looking back up to a distance \(q\), then (1) suggests that
and then $N$ has a limiting geometric distribution with the success probability
\[
P(X_{q+1} \leq u_n | \cap_{i=1}^{q} \{X_i > u_n\}),
\]
which concludes the intuition. This heuristic argument was proved to hold under suitable conditions in [1]. On the other hand, Aytaç et al in [8] constructed several examples where both (limiting) parameters coincide even when the cluster size distribution has nothing to do with a geometric one. Smith in [18] proposed an example where this interpretation for the extremal index fails.

In the present paper we have two main purposes. Firstly, we want to show that one should consider not only asymptotic limits but also look at the behaviour for finite $n$ in order to get a full picture of the situation. This is because in the real world we only observe finite $n$, and also because things may behave differently at finite size and in the limit. For instance, we will show that even if in the limit the extremal index and the reciprocal of the cluster size are different, as in Smith’s example, they coincide for finite time. Also, we show that the above heuristic argument may not work, and a subtly different quantity could be more appropriate to consider. In the above argument, $N$ was considered assuming the existence of a cluster, i.e. assuming that we started with an exceedance, but without guaranteeing whether that exceedance initiated a cluster, which means that it could correspond to an exceedance within a cluster that was actually initiated earlier. But we can consider the case where that exceedance actually begins the cluster. This would make no difference if the far past is irrelevant. In the models we consider below, the extremal index will be related to the second case and not the first one.

Moreover, to emphasize the importance of looking at finite and not just limiting statistics, we present another model where the extremal index does not exist since its asymptotics fluctuate. The same happens with the distribution of the cluster size. However, both can still be identified as the reciprocal of each other for finite observations.

In our case study we also consider the following application.

Application: hitting times. Parallel to the extreme value theory and in a totally independent context, the theory of hitting times in Poincaré recurrence theory has been deeply studied. The review papers [5, 11, 14] bring a major panorama of classical results, with hitting times of ball and cylinder sets particularly considered. To be more concrete, consider a sequence $\alpha_{n-1}^0$ and define the hitting time
\[
\tau_n = \inf \{t \geq 1 | X_t + n - 1 = a_{n-1}^0\}.
\]
Now, if an infinite sequence $\alpha = (a_\infty^0)$ is fixed, then one can consider the number of (consecutive) letters of $\alpha$ that can be read in the process at any time $t$. Namely
\[
Y_t = \max \{k \geq 0 | X_t + k - 1 = a_{k-1}^0\}.
\]
The usual abuse of notation $X_t = a_{n-1}^0$ means that no letter $a$ can be read. Thus, one gets
\[
\{\tau_n > t\} = \left\{\max_{1 \leq j \leq t} Y_j < n\right\}.
\]
In other words, the hitting time problem can be translated in terms of a maximum problem. It is well known that under suitable mixing conditions the hitting time converges to an
exponential law. The most general result to date is from Abadi and Saussol (see [6]) and says that for $\alpha$-mixing systems and every $a$

$$\lim_{n \to \infty} P \left( \tau_n > \frac{t}{\theta_q \mu(a_0^{-1})} \right) = e^{-t},$$

for some $q = q(a, \alpha)$, which in general is as large as the memory of the process. Therefore, the problem is how to compute $\theta_q$, for large $q$ (which may also include determining the appropriate $q$). Under certain conditions ($\phi$-mixing, renewal process), in the papers [1, 2], the authors have shown that $q$ can be replaced by the periodicity of the observed set, which in general is short and makes $\theta_q$ easier to handle. In our case, these mixing conditions are not verified if the alphabet is infinite. However, we show that actually the periodicity of the observed set can still be used to calculate $\theta_q$.

The structure of the paper is as follows. In section 2, we introduce the general form of the regenerative processes we consider and some basic properties are derived. Section 3 is dedicated to the decay of correlations of the model. In section 4, we compute the parameters for the different cases we consider. The first one exhibits different values for the finite and limiting extremal index. The second one exhibits the geometric distribution, where the finite and infinite case coincide. The third one shows a case where the limit of the cluster size is actually a sub-distribution and the limiting extremal index does not exist. Finally, the fourth case shows a cluster size distribution which fluctuates cyclically and, thus, the limiting extremal index does not exist either. In all of them the finite extremal index coincides with the inverse of the finite mean cluster size. Only in the geometric case is the cluster size equal to the sojourn size, a case which we consider at the beginning of section 4.

2. Regenerative models

We consider a general construction of regenerative processes, of which Smith’s model, introduced in [18], is a particular case. These are discrete time models over a finite or countable alphabet. To simplify, from now on we consider that the alphabet is a set of positive integers $\mathbb{N}$.

First let $(Z_n)_{n \in \mathbb{Z}}$ be an i.i.d sequence of random variables taking positive integer values, with a common distribution $p_a = P(Z_n = a), a \in \mathbb{N}$ and finite mean $\mathbb{E}(Z_n) < \infty$. To each $a \in \mathbb{N}$ we also associate a distribution $q_a = (q_a(k))_{k \in \mathbb{N}}$ taking values on the positive integers. The process $(X_n)_{n \in \mathbb{Z}}$ that we are going to consider can be described informally in the following way: take $Z_n$, choose a random number $\xi_n$ with a distribution $q_{Z_n}$ independent of everything, and repeat the symbol $Z_n$ a number of times $\xi_n$. The blocks of size $\xi_n$ (filled up with the symbol $Z_n$) are concatenated to create the process $(X_n)_{n \in \mathbb{Z}}$. A suitable initial condition turns it into a stationary process if we assume that the mean regeneration time is finite. To formalize, define the sequence $(X_n)_{n \in \mathbb{Z}}$ as follows. Assume first that

$$\nu = \sum_{k \geq 1} p_k \sum_{j \geq 1} q_k(j) < \infty. \quad (4)$$

Now, let the random variable $\zeta_i$ have the distribution

$$P(\zeta_i = a) = \frac{1}{\nu} \sum_{k \geq 1} p_k \sum_{j > a} q_k(j). \quad (5)$$

This will be used to make the process stationary. To this end, for every $n \geq 2$ and each index $i$ such that
\[ \zeta_1 + \sum_{j=1}^{n-1} \xi_j \leq i < \zeta_1 + \sum_{j=1}^{n} \xi_j =: \zeta_n, \]

set \( X_i = Z_n \). (By convention the sum over an empty set of indices equals zero.) This defines \( X_i \) for all \( i \geq \zeta_1 \).

Secondly, we define the process for the remaining indices in a similar way. That is, for every \( n \geq 0 \) and each index \( i \) such that

\[ \zeta_1 - n \sum_{j=0}^{n-1} \xi_{j-1} \leq i < \zeta_1 - n \sum_{j=0}^{n-1} \xi_{j-1}, \]

set \( X_i = Z_{-n} \). Briefly, \((\zeta_n)_{n \in \mathbb{Z}}\) is the sequence of regenerations and \((\xi_n)_{n \in \mathbb{Z}}\) are the distances between consecutive regenerations (in other words, the length of the regenerative blocks). In particular, \( \zeta_1 \) is the time shift needed to make it stationary. The process is positive recurrent with the stationary measure \( \mu \) (and then ergodic) if and only if the regeneration time has a finite mean. In our case, this mean is denoted by \( \nu \), which we assume to be finite. By Kac’s lemma one has that the invariant measure of a regeneration is equal to the reciprocal of the above display. The regenerations are useful for computing the invariant measure of a measurable set \( A \), which will follow from the conditioning on the last regeneration time of \((X_n)_{n \in \mathbb{Z}}\).

To simplify the notation, for every \( j \in \mathbb{Z} \), we define the events

\[ R_j \equiv \{ \exists \ i \in \mathbb{Z} \mid \zeta_i = j \} \quad \text{and} \quad W_j = R_j \cap \bigcap_{i=j+1}^{0} R_i, \quad \text{for} \ j \leq 0, \]

corresponding, respectively, to the occurrence of a regeneration at time \( j \), and that no other regeneration occurs until time 0 (also by convention \( W_0 = R_0 \)). The invariant measure of any measurable set \( U \) can be computed partitioning the past according to the \( W_{-j} \)

\[ \mu(U) = \sum_{i=0}^{\infty} \mu(U|W_{-i})\mu(W_{-i}). \quad (6) \]

In particular, \( U_a = \{ X_0 > a \} \) gives the tail distribution and we put

\[ g_a := \mathbb{P}(X_0 > a) = \sum_{j=a+1}^{\infty} \mu(j). \quad (7) \]

These formulae will be used later on with the ad hoc properties of each specific model considered. Finally, note that by construction, the process is reversible, and as a consequence we have the conditional independence of consecutive blocks, i.e.

\[ \mathbb{P}_{R_i}(X_0 \in A, X_1 \in B) = \mathbb{P}_{R_i}(X_0 \in A)\mathbb{P}_{R_i}(X_1 \in B), \]

for \( A, B \subseteq \mathbb{N} \).

We introduce now particular examples corresponding to different cases, which we are going to study in order to illustrate the finite and limiting behaviour of the extremal index.

2.1. Basic example: i.i.d

As a first basic example, notice that a sequence of i.i.d. random variables is included in this family of processes with \( q_a(k) = \delta_1(k) \) for all \( k \) and every \( a \).

We now come to the models we consider in this paper.
2.2. Smith’s model

The model considered by Smith in [18], to show that the limiting extremal index and the reciprocal of the limiting mean cluster size may be different, is defined by setting

\[ q_a(k) = \begin{cases} \frac{a-1}{a} & \text{for } k = 1 \\ \frac{1}{a} & \text{for } k = a + 1 \\ 0 & \text{otherwise} \end{cases} \]

Thus \( E(q_a) = 2 \), for all \( a \in \mathbb{N} \), and hence \( \nu = 2 \). Moreover, and in particular for the set \( \{ X_0 = a \} \), by (6) we get

\[
\mu(a) = \mathbb{P}(X_0 = a, W_0) + \sum_{j=1}^{a} \mathbb{P}(X_0 = a, W_{-j}) = \frac{1}{2} p_a + \sum_{j=1}^{a} \mathbb{P}(X_0 = a, R_{-j}, \gamma^0_{-j}, R^c_i)
\]

\[
= \frac{1}{2} p_a + \sum_{j=1}^{a} \mathbb{P}(R_{-j}) \mathbb{P}(X_0 = a, R^c_{-j}, \gamma^0_{-j}(R_i)) = \frac{1}{2} p_a + \sum_{j=1}^{a} \frac{1}{2} p_a = pa.
\]

(8)

2.3. The block model

This model is constructed to present a case where the limiting extremal index does not exist, since the exit probability fluctuates as the level \( u_n \) diverges. The same occurs for the limiting cluster size distribution, however, the finite parameters are equal. The model is constructed using any distribution \( (p_a)_{a \in \mathbb{N}} \) but with deterministic distributions \( (q_a)_{a \in \mathbb{N}} \). Set

\[ q_a(k) = \begin{cases} 1 & \text{for } k = a \\ 0 & \text{otherwise} \end{cases} \]

Hence \( E(q_a) = a \) for all \( a \in \mathbb{N} \), and we get \( \nu = \sum_{a=1}^{\infty} a p_a \), which we assume to be finite in order to have stationarity. In a similar way to Smith’s model, it follows by (6) that \( \mu(a) = ap_a/\nu \).

3. Decay of correlations

A general argument shows that a mixing regenerative process is weak Bernoulli (see for instance the book [17]). Thus, the models presented in this paper are all weak Bernoulli. In some cases, the stronger decay of correlations can be computed explicitly. As an illustration, we are going to compute here the probability of having a regeneration after \( n \)-steps, given that another one was observed in the present time. Namely,

\[ c_n = \mathbb{P}(R_{n+1} | R_0), \]

in the case of a two-symbol process.

3.1. Morse code and fibonacci numbers

Consider the following case as a basic example of the block model. Suppose \( p_2 = 1 - p_1 \) (and \( p_a = 0 \) for \( a \geq 3 \)); that is, the process only takes values 1 and 2. Further, for \( a = 1 \) and \( a = 2 \) consider \( q_a = \delta_a \); namely, when 1 is chosen it is written once, and when 2 is chosen it is written twice. This model represents the messages that can be written with Morse code where only
points and traces are allowed. Thus $c_n$ can be regarded as the probability of writing a message of exactly length $n$. Make $x$ the total number of 1s and similarly $y$ the total number of 2s in this message. We get

$$c_n = \sum_{y=0}^{\lfloor n/2 \rfloor} \binom{x+y}{y} p_1^x p_2^y. \quad (9)$$

The condition $x + 2y = n$ allows the above display to be rewritten as

$$c_n = p_1^n \sum_{y=0}^{\lfloor n/2 \rfloor} \binom{n-y}{y} \left( \frac{p_2}{p_1} \right)^y.$$

At the moment, notice that for the golden ratio $p_1 = (-1 + \sqrt{5})/2 = \phi$, one gets

$$c_n = \phi^n F_n,$$

where $F_n$ is the $n$th Fibonacci number. Pascal recurrence leads us to rewrite the above formula as

$$c_n = p_1 c_{n-1} + p_2 c_{n-2}.$$

Recursive formula conditioning on the previous regeneration could also be invoked to obtain this recursion. It is classical to obtain the solution of this recursion via the roots of its characteristic polynomial

$$x^2 - p_1 x - p_2,$$

which, since the different roots are $r_1 = 1$ and $r_2 = p_1 - 1$, takes the form

$$c_n = K_1 1^n + K_2 (p_1 - 1)^n.$$

With the initial condition $c_0 = 1, c_1 = p_1$, the constants become

$$K_1 = \frac{1}{2 - p_1}; \quad K_2 = \frac{1 - p_1}{2 - p_1}.$$

Thus, notice that $K_1 = \mathbb{P}(R_0)$ and since $c_n$ converges to $K_1$, we get the (exponential) decay of correlations. Now, it follows easily that the process is $\psi$-mixing with an exponential rate function $\phi(n) = (1 - p_1)^n$. For easy reference we remind the reader that $\psi$ is defined as

$$\psi(n) = \sup_{A \in \mathcal{C}^u, B \in f^{-n+u}\mathcal{C}^u, u \in \mathbb{N}} \left| \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)\mathbb{P}(B)} - 1 \right|,$$

where $f$ is the shift operator, $\mathcal{C}$ is the alphabet, and for $u \in \mathbb{N}$, we denote by $\mathcal{C}^u$ its $u$-fold Cartesian product.

### 3.2. The finite block model

The argument on the example above can be easily continued (except for interpretation (9)) to prove that the process $(X_n)$ considered in this paper over a finite alphabet $\mathcal{C}$ is exponentially $\psi$-mixing.

It is worth noticing that the above methodology captures the eigenvalues of the Perron–Frobenious operator, identifying the largest one with a modulus equal to 1, the remainder with a smaller modulus, and also the rate of mixing given by the spectral gap.
3.3. The infinite case

The infinite case must be considered with more attention and may not be $\psi$-mixing. Consider for instance a probability $(p_a)_{a \in \mathbb{N}}$ with no-null entries $p_a$. We treat the block model first. So, suppose further that for each positive integer $a$, one has the conditional probability $q_a = \delta_a$. That is, each time $a$ is chosen in a regeneration, it is repeated deterministically $a$ times. Thus

$$P(X_{a-1} = a | R_0, X_0 = a) = 1.$$ 

Since $a$ can be as large as we want, the process cannot be $\psi$-mixing.

Now consider Smith’s model. Fix $n \in \mathbb{N}$, which will stand for the size of the gap in the mixing condition, and for $a > n$ take $A = \{X_{-1} \neq a, X_0 = a\}$ and $B = \{X_{a+1} = a\}$. Thus $P(A \cap B)/P(A) \geq 1/a$ while $P(B) = p_a$, since there is a probability $1/a$ of $B$ occurring, in case the process repeats $a$ because $a > n$ (note that there are other ways of $B$ occurring). Thus the ratio of the last two probabilities cannot be close to one and the process is not $\psi$-mixing.

4. Extremal index and clumping expectations

4.1. Definitions

In this section we present definitions for the family of parameters we are going to consider. As one of the main purposes of this paper, we present them for finite observations and then consider their asymptotics. We begin with the extremal index. We introduce first some notation to simplify the expressions. For a size $q \in \mathbb{N}$ and a level $a > 0$, let us define the sets $U_a = \{X_0 > a\}$ and $A_a^{(q)} = \{X_0 > a, X_1 \leq a, \ldots, X_q \leq a\}$.

**Definition 1.** The (finite, $a$-level) extremal index (up to time $q$) is defined by

$$\theta_q(a) = P(A_a^{(q)} | U_a).$$

This is the probability of not observing another exceedance (of level $a$) up to time $q$ given that we begin with the observation of an exceedance at time 0. This formula was used firstly by O’Brien and then by other authors for the extremal index (see for instance [10, 12, 13, 16]). The value of $q$ is determined by the event $U_a$ and the decay of correlation properties of the process. See [9, equation (15)] and the discussion preceding it regarding adequate choices of $q$. In general, this becomes more difficult to compute the larger $q$ becomes. In the context of hitting times, it was shown that under fast mixing conditions ([1, 2]) $q$ can be taken as the (minimum) periodicity of the observable (also called the shortest possible return time, or shortest possible distance between two observations of $U_a$). It is given by the positive integer defined as follows

$$p(U_a) = \inf \{k \geq 1 | P(X_0 > a, X_k > a) > 0\}.$$ 

We call the $\theta_q$ the *escape probability* (see [4]) when taking $q = p(U_a)$. Namely, the escape probability is $\theta_{p(U_a)}(a) = P(A_a^{(p(U_a))} | U_a)$.

The clustering of abnormal observations creates an extremal index smaller than one. The size of this cluster, being random, has a distribution with an expectation related to the reciprocal of the extremal index; one must be careful in defining the size of this cluster. Two different cases are considered here. The first case, defined formally in definition 2, is due to the heuristic argument described in the introduction. It considers the process starting from the observable state of interest and counts how long it stays in the same state. The geometric behaviour of this quantity, called the *sojourn time* (under suitable conditions), was proved by Abadi and Vergne, in [7]. The sojourn time arises naturally in physical problems and computational
simulation where an initial condition must be imposed. It also corresponds to the case where some automatic mechanism detects the occurrence of $U_a$, but where failures on the mechanism or in the sample itself do not allow guarantees that the cluster actually started at this point. To formalize, let

$$N_a = \sup\{k \geq 0 \mid X_{np(U_a)} > a, \forall 0 \leq j \leq k\} + 1,$$

the number of consecutive observations of the exceedance of $a$. The $+1$ at the end corresponds to counting the occurrence of the exceedance at time zero, namely $X_0 > a$. (We set $N_a = 0$ if $X_0 \leq a$.)

We will use the classical notation $E_V(X)$ for the conditional expectation of the random variable $X$ given the event $V$.

**Definition 2.** The expected sojourn is defined by

$$E_{U_a}(N_a).$$

The second case, defined formally in definition 3, is due to a natural interpretation of the process as a time series evolution and then considering the beginning of a cluster; that is, when the process enters the observable state of interest. Stationarity lets us fix this entrance at any position in the time scale.

**Definition 3.** Let $E_a = \{X_{-p(U_a)} \leq a, X_0 > a\}$ denote the event corresponding to the entrance into an exceedance of $a$. We define the mean cluster size to be the conditional expectation of $N_a$

$$E_{E_a}(N_a).$$

Note that in the first one we know that at time 0 we have an exceedance, but do not know if a cluster of such exceedance could have been initiated earlier, while in the second one, we know that the occurrence at time 0 was the beginning of the cluster.

### 4.2. Values of the parameters

In this section we compute the clustering parameters defined in the previous section in order to illustrate the different behaviours already mentioned.

#### 4.2.1. Smith’s model

First, we are going to consider the case of an exceedance of level $a$ and then the case of hitting a cylinder of at least size $n$.

**Exceedances.** Consider an exceedance of level $a$ and let us compute the clustering parameters corresponding to this event. We begin with the escape probability. Since in our examples two exceedances can occur immediately one after the other, we get $p(U_a) = 1$. Thus we compute

$$\theta_1(a) = \frac{P(X_0 > a, X_1 \leq a)}{P(X_0 > a)}.$$
\[ P(X_0 > a, X_1 \leq a) = P(R_1)P(X_0 > a|R_1)P(X_1 \leq a|R_1). \]

Since the distribution of \( X_0 \), conditioned to a regeneration at the origin, is the distribution of \( Z_0 \) we get
\[ P(X_0 > a|R_1) = e_a := \sum_{j > a} p_j, \quad \text{and} \quad P(X_1 \leq a|R_1) = 1 - e_a. \]

Notice that in this model \( g_a = e_a \). We obtain \( \theta_1(a) = (1 - e_a)/2 \), and the limiting extremal index equals 1/2 as stated in Smith’s work in [18].

We compute now the mean cluster size and then the mean sojourn time. Let us consider the first case. Conditioning to \( E_a \) and without loss of generality, we can assume that \( \zeta_1 = 0 \). Thus we can write
\[ N_a = \sum_{i=1}^{\ell} \xi_i, \]
where \( \ell = \max\{i \geq 1 \mid Z_i \geq a\} \). We can use Wald’s formula to establish that
\[ E_{E_a}(N_a) = E_{E_a}(\xi_1)E_{E_a}(\ell). \]

We recall that by construction \( E_{E_a}(\xi_1) = 2 \). Furthermore, \( \ell \) is a geometric (conditioned to \( E_a \), starting at one) random variable with a success probability \( 1 - e_a \). Thus,
\[ E_{E_a}(N_a) = 2 \frac{1}{1 - e_a}, \]
which is the reciprocal of the finite escape probability. This holds even when the expectation of the limiting cluster size distribution is equal to one.

For the mean sojourn time we will use formula (A.1), proved in the appendix. To this end, we first need the second moment \( E_{E_a}(N_a^2) \). We write
\[ N_a^2 = \sum_{i=1}^{\ell} \xi_i^2 + 2 \sum_{1 \leq i < j \leq \ell} \xi_i \xi_j. \]

Again by Wald’s equation
\[ E_{E_a} \left( \sum_{i=1}^{\ell} \xi_i^2 \right) = E_{E_a}(\ell)E_{E_a}(\xi_1^2) \quad (10) \]
\[ E_{E_a} \left( \sum_{1 \leq i < j \leq \ell} \xi_i \xi_j \right) = E_{E_a}(\ell(\ell - 1))E_{E_a}^2(\xi_1). \quad (11) \]

Direct computations give that the two quantities above are equal, respectively, to
\[ \frac{1}{1 - e_a} \left( 3 + \sum_{j > a} \frac{p_j}{e_a} \right), \]
and
\[ \frac{2e_a}{(1 - e_a)^2} 4. \]

Collecting all this information and using formula (A.1), we conclude that
\[ \mathbb{E}U_a(N_a) = \frac{3 + \sum_{j>0} j p_j}{4} + \frac{4e_a}{1-e_a} + \frac{1}{2}. \]

Notice that \( \mathbb{E}U_a(N_a) \) differs from \( 1/\theta_1(a) \). In particular, the first one converges to \( 5/4 \) against the limit value of the second one which is 2, illustrating a considerable discrepancy between the mean sojourn time and the reciprocal of the extremal index.

**Hitting the cylinders.** Consider the infinite sequence \( a = (a, a, a, ...) \) consisting only of the symbol \( a \). For large \( n \in \mathbb{N} \), we are interested in observing more than \( n-1 \) consecutive \( a \). Thus, we are interested in exceedances of the level \( n-1 \) for the process \( Y_j = \sup\{ k \in \mathbb{N} \mid X_{j+k} = a^k \} \). The clustering of these exceedances corresponds to the clustering of observations of the set \( U_n = \{ Y_0 > n-1 \} = \{ X_0^{n-1} = a^n \} \). We are going to compute the extremal index \( \theta \) for \( U_n \) and then consider the asymptotics on \( n \). The period of \( U_n \) is 1. Thus, we are first going to obtain the escape probability \( \theta_1(n) \). Namely

\[ \theta_1(n) = \frac{\mathbb{P}(X_0^{n-1} = a^n, X_n \neq a)}{\mathbb{P}(X_0^{n-1} = a^n)}. \]

As in the previous case, we consider the numerator and the denominator separately. By reversibility and conditioning on the regeneration, the numerator is equal to

\[ \frac{1}{2} \mathbb{P}(X_0 \neq a | R_0) \mathbb{P}(X_0^{n-1} = a^n | R_0). \]

The first factor to compute is just \( 1 - p_a \). For the second one, put \( p = p_a/a \) and \( q = p_a(a-1)/a \). Since one either puts a block of length 1 or \( a + 1 \) immediately after the regeneration, a recursive equation can be constructed

\[ \mathbb{P}(X_0^{n-1} = a^n | R_0) = q \mathbb{P}(X_0^{n-2} = a | R_0) + p \mathbb{P}(X_0^{n-(a+2)} = a | R_0), \]

which has the characteristic polynomial

\[ x^{a+1} - qx^a - p = x^a(x-q) - p. \]

From the last expression it follows that for \( a \) odd, there are two roots \( r_1, r_2 \) positive and negative respectively with \( 0 \leq -r_2 < r_1 < 1 \) and thus the solution of the recursion takes the form

\[ \mathbb{P}(X_0^{n-1} = a^n | R_0) = K_1 r_1^n + K_2(n) r_2^n, \]

with \( K_1 \) a constant (independent of \( n \)) and \( K_2(n) \) a polynomial of degree \( a-1 \). For \( a \) even, there is only one root \( 0 < r_1 < 1 \) and so the solution of the recursion takes the form

\[ \mathbb{P}(X_0^{n-1} = a^n | R_0) = K_1 r_1^n. \]

In either case, the leading term is the first one. Now we compute the denominator, the stationary measure of \( a^n \). We decompose it with respect to the previous occurrence of a regeneration. Namely, it is equal to

\[ \frac{1}{2} \sum_{j=0}^{a} \mathbb{P}(W_j, X_j^{n-1} = a^{j+n} | W_{-j}). \]
The first term \( j = 0 \) has just been computed. For the remaining terms, since there is no regeneration at time 0, the first block (the one immediately after the regeneration at \(-j\)) has to have a size \( a + 1 \). Therefore, for \( 1 \leq j \leq a \)
\[
\mathbb{P}(W_{-j}, X_{-j}^{a-1} = a^{j+1}|W_{-j}) = p\mathbb{P}(X_{-j+a+1}^{a-1} = a^{j-a-1+n}|R_{-j+a+1})
\]
\[
= p\mathbb{P}(X_0^{a-j-a} = a^{j-a-1+n}|R_0).
\]

The second equality follows by stationarity; the last expression is the one already obtained. We conclude that

\[
\mu(a^n) = \frac{1}{2} \left[ \mathbb{P}(X_0^{a-1} = a^n|R_0) + p \sum_{j=1}^{a} \mathbb{P}(X_0^{a-1-j} = a^{j+n}|R_0) \right]
\]

\[
= \frac{1}{2} K_1 r_1^n \left[ 1 + \frac{p}{r_1} \sum_{j=0}^{a-1} r_1^j \right] + o(r_1^n). \quad (18)
\]

A direct calculation, using the fact that \( r_1 \) is the root of the characteristic polynomial, gives that the factor between brackets is equal to

\[
\frac{1 - pr_n}{1 - r_1}.
\]

Finally, we get

\[
\theta_{n}^{-1} \approx \frac{1}{1 - r_1}
\]

Now we compute the expectation of the cluster size. Set in this case

\[
E_n = \{X_{-1} \neq a, X_0^{a-1} = a^n\},
\]

for the set, which means entering the cylinder \( a^n \). Similarly to the exceedance case, the cluster size is given by the random variable

\[
N_n = \sup \{k \geq 1 \mid X_j^{a+n-1} = a^n \forall j \leq k\} + 1.
\]

It can be easily derived from the conditional measure of \( a^n \) derived in (13) and (14) that

\[
E_{E_n}(N_n) = \sum_{j=0}^{\infty} \frac{\mathbb{P}(X_0^{a-1-j} = a^n|X_0^{a-1} = a^n|R_0)}{\mathbb{P}(X_0^{a-1} = a^n|R_0)} = \sum_{j=0}^{\infty} r_1^j = \frac{1}{1 - r_1}.
\]

Observe that the sojourn distribution is geometric and this example shows how the cluster and sojourn size coincide in this case.

4.2.2. The block model. As before, we first consider the case of exceedances of the level \( a \) and then the case of hitting a sequence of at least size \( n \).

Exceedances. Consider an exceedance of level \( a \); we still have in this case \( p(U_a) = 1 \). As in Smith’s model, we compute \( \theta_1(a) = \mathbb{P}(X_0 > a, X_1 \leq a)/\mathbb{P}(X_0 > a) \). Similarly to this case,

\[
\mu(j) = \frac{j \mu_j}{\nu}, \quad \mathbb{P}(X_0 > a) = g_a = \frac{\sum_{j>a} \mu_j}{\nu} \quad \text{and} \quad \nu = \sum_{j \geq 1} \mu_j.
\]
For the numerator \( P(X_0 > a, X_1 \leq a) = \mu(R_1) P(R_1 > a, X_1 \leq a) \). It follows that
\[
P(X_0 > a, X_1 \leq a) = \frac{1}{\nu} e_a (1 - e_a),
\]
where we recall that \( e_a = \sum_{j > a} p_j \). Thus
\[
\theta_1(a) = \frac{e_a (1 - e_a)}{\nu \gamma_a},
\]
and the limiting extremal index is equal to zero.

Now, we compute the mean of the cluster size of exceedances of \( a \) using Wald’s formula as in Smith’s model.
\[
E_{E_a}(N_a) = E_{E_a}(\xi_1) E_{E_a}(\ell) = \frac{\nu \gamma_a}{e_a} \frac{1}{1 - e_a},
\]
which is the reciprocal of the escape probability. However, the distribution of \( N_a \), as \( a \) diverges, does not even converge to a limiting distribution. Actually, the cumulative distribution \( P_{E_a}(N_a \leq k) \) converges pointwise to zero for all \( k \), which means the reciprocal of the mean size does not even exist and one cannot compare it with the extremal index in the limit. Despite this, they are equal for finite time. The reason is clear: there is a mass escape in the distribution of the \( N_a \), they are not uniformly bounded by an integrable function and the dominated convergence theorem does not hold. The lack of tightness is at the core of the construction. Taking the limit of the expectation and not the expectation of the limit should be the recipe for relating it to the extremal index.

To compute the mean sojourn time, as in Smith’s model, we use formula (A.1) in the appendix. We first need to compute
\[
E_{E_a}(\xi_1^2) = \frac{1}{e_a} \sum_{j > a} j^2 p_j.
\]
Now, to compute \( E_{E_a}(N_a^2) \), we still use (10) and (11). We get respectively
\[
\frac{1}{1 - e_a} \frac{1}{e_a} \sum_{j > a} j^2 p_j,
\]
and
\[
\frac{2 e_a}{(1 - e_a)^2} \frac{\nu^2 \gamma_a^2}{e_a^2}.
\]
Collecting all this information, we conclude that
\[
E_{U_a}(N_a) = \frac{\sum_{j > a} j^2 p_j}{2 \nu \gamma_a} + \frac{2 \nu \gamma_a}{1 - e_a} + \frac{1}{2}.
\]
Notice that again \( E_{U_a}(N_a) \) differs from \( 1/\theta_1(a) \). In particular, the second term converges to zero as \( a \) diverges and
\[
E_{U_a}(N_a) \approx \frac{1}{2} \left[ \frac{\sum_{j > a} j^2 p_j}{\sum_{j > a} j p_j} + 1 \right].
\]
Thus, \( E_{U_a}(N_a) \) can take any limiting value, including infinity, depending on the \( p_j \).
Hitting cylinders. Take the infinite sequence $\mathbf{a} = (a, a, a, \ldots)$. Still in this model $p(U_n) = 1$. Let us compute the escape probability
\[
\theta_1(n) = \mathbb{P}(X_n \neq a|X_n^{-1} = a^n) = 1 - \frac{\mu(a^{n+1})}{\mu(a^n)}.
\]
This means that we are only left with the computation of $\mu(a^n)$. To do this we condition the last occurrence of a regeneration of the process before $X_0 = a$. Since the process repeats $a$ times the symbol $a$, this regeneration cannot go further than $a$ coordinates before 0. Thus
\[
\mu(a^n) = \mathbb{P}(X_0^{n-1} = a^n) = \sum_{j=0}^{a-1} \mathbb{P}(X_{-j}^{n-1} = a^{n+j}|R_{-j})\mathbb{P}(R_{-j}).
\]
Now, for $0 \leq j \leq a - 1$, write
\[
n + j = \left\lceil \frac{n + j}{a} \right\rceil a - s_{n+j}, \quad 0 \leq s_{n+j} \leq a - 1.
\]
By construction of the process
\[
\mathbb{P}(X_{-j}^{n-1} = a^{n+j}|R_{-j}) = \frac{\left\lceil \frac{s_j}{a} \right\rceil}{\nu} [s_n + 1 + (r_n - 1)p_a],
\]
We conclude that
\[
\mu(a^n) = \frac{1}{\nu} \sum_{j=0}^{a-1} \frac{\left\lceil \frac{s_j}{a} \right\rceil}{\nu} [s_n + 1 + (r_n - 1)p_a],
\]
where $r_n = a - s_n$. Therefore $\theta_1(n)$ is equal to
\[
1 - \frac{s_n + r_n p_a}{s_n + 1 + (r_n - 1)p_a} = \frac{1 - p_a}{s_n + 1 + (r_n - 1)p_a}.
\]
Thus, the extremal index as a limit does not exist since $s_n$ runs cyclically between 0 and $a - 1$.

We now estimate the mean of the distribution of consecutive observations of the target sequence $a^n$. We use again the unconventional Euclidean form (19) and get
\[
\mu_{E_n}(N_n \geq k) = \begin{cases} 
1 & \text{if } 1 \leq k \leq s_n + 1, \\
\frac{p_a^\ell}{\nu} & \text{if } \ell a + s_n + 1 < k \leq (\ell + 1)a + s_n + 1, \ell \geq 1.
\end{cases}
\]
We conclude that the distribution of $N_n$ does not converge to a limit distribution in $n$. Furthermore, since a new block of size $a$ is chosen with probability $p_a$, we can establish the following equation
\[
\mathbb{E}_{E_n}(N_n) = (\mathbb{E}_{E_n}(N_n) + a) p_a + (s_n + 1)(1 - p_a).
\]
It follows that
\[
\mathbb{E}_{E_n}(N_n) = s_n + 1 + \frac{ap_a}{1 - p_a}.
\]
Now, $s_n$ does not have a limit as $n$ diverges. Thus $\mathbb{E}_{E_n}(N_n)$ does not have a limit in $n$. However, it is easy to verify the identity
\[
\frac{1}{\mathbb{E}_{E_n}(N_n)} = \theta_1(n).
\]
Let us consider the mean sojourn time. In this case

\[ N_n = s_n + 1 + a\tilde{\ell}, \]

with \( \tilde{\ell} \) a random variable with geometric distribution starting at zero and success probability \( 1 - p_a \). Straightforward computations give

\[
E_{E_n}(N_n^2) = (s_n + 1)^2 + 2(s_n + 1)aE(\tilde{\ell}) + a^2E(\tilde{\ell}^2)
= (s_n + 1)^2 + 2(s_n + 1)a - \frac{p_a}{1 - p_a} + a^3p_a(p_a + 1) \frac{1}{(1 - p_a)^2}.
\]

Using (A.1), one can derive an expression for the mean sojourn time. Even though \( a \) is fixed and one considers the asymptotics in \( n \), it is interesting in particular to consider the case of large \( a \) for which \( apa \) is small. In this case, we have

\[ E_{E_n}(N_n) \approx s_n + 1 \quad \text{and} \quad E_{U_n}(N_n) \approx \frac{s_n}{2} + 1. \]

5. Extremal versus escape

O’Brien’s formula defines the extremal index as a function of a suitable number \( q = o(\mu(\lambda)^{-1}) \) and then putting \( \theta = \theta_q = p(A^{(q)}|U) \). This formula was also obtained independently for the exponential law for hitting/return times in [1, 2]. The precise value of \( q \) to be taken depends on the properties of the correlation decay of the process and on the observable itself. Generally, \( \theta_q \) becomes more difficult to compute the larger \( q \) becomes. The lemma below establishes that in the general model considered, and for any observable level \( a \) or \( n \)-cylinder set, any \( q = o(\mathbb{P}(U_a|R_0)^{-1}) \) can be taken, and thus the smallest one possible can be chosen, which is the period of the observable.

**Lemma 4.** Consider the regenerative process defined in section 2. Consider the level \( a \in \mathbb{N} \). The following inequality holds for all \( q \geq p(U_a) \)

\[ 1 - \frac{\theta_q}{\theta_{p(U_a)}} \leq q\mathbb{P}(U_a|R_0). \]

**Remark 5.** The monotonicity of \( \theta_q \) as a function of \( q \) and the lemma above establish that \( \theta_q \) are equivalent in ratio for all \( q = o(\mathbb{P}(U_a|R_0)^{-1}) \). In particular, this shows that the parameter \( \lambda_{U_a} \) in the exponential law of the hitting/return time of \( U_a \) can be replaced by \( \theta_{p(U_a)} \).

**Proof of lemma 4.** Consider

\[ \theta_{p(U_a)} = \frac{\mathbb{P}(X_0 > a, X_{p(U_a)} \leq a)}{\mathbb{P}(X_0 > a)} \quad \text{and} \quad \theta_q = \frac{\mathbb{P}(X_0 > a, \bigcap_{j=p(U_a)}^{q} X_j \leq a)}{\mathbb{P}(X_0 > a)}. \]

The difference of the probabilities in the numerators is equal to

\[ \mathbb{P}(X_0 > a, X_{p(U_a)} \leq a, \bigcup_{j=p(U_a)+1}^{q} X_j > a). \]

Making a disjoint partition of the union in the above probability as a function of the second exceedance, we get that it is equal to
\[ \mathbb{P}(X_0 > a, X_{p(U_a)} \leq a) \sum_{j=p(U_a)+1}^{q} \mathbb{P} \left( \bigcap_{i=p(U_a)+1}^{j-1} X_i \leq a, X_j > a, X_{p(U_j)} \leq a \right) \cdot \mathbb{P}(j-1) \bigcap_{i=p(U_i)}^{j} X_i \leq a, R_j | X_0 > a, X_{p(U_j)} \leq a). \]

Since there must be a regeneration at time \( j \), the leading term can be factorized as

\[
\mathbb{P}(\bigcap_{i=p(U_i)}^{j-1} X_i \leq a, R_j | X_0 > a, X_{p(U_j)} \leq a) \mathbb{P}(X_j > a | R_j). \tag{21}
\]

The left-most factor is bounded simply by one. The second one is equal to \( \mathbb{P}(X_0 > a | R_0) \), independently of \( j \), by stationarity. This concludes the proof of the lemma.

\[\square\]

**Exemple 6.** Consider first the case of exceedances of level \( a \) by the process. Then \( \mathbb{P}(X_0 > a | R_0) \) is equal to \( e_a \). For the case of cylinders \( a^n \), one has \( \mathbb{P}(X_0 > a | R_0) = O(K^n) \) for a constant \( 0 < K < 1 \). For a general cylinder \( a_0^{n-1} \), the same holds when changing \( \theta \) by \( \theta \left( a_0^{n-1} \right) \).

Sharpness. Instead of bounding the left-most factor in (21) by one, we can compute it exactly, at least in some cases. Suppose the \( X_i \) are independent random variables, so \( p(U_a) = 1 \). In this case, the left-most factor in (21) becomes equal to \( (1 - e_a)^{j-2} \). Summing up to \( q \) we obtain \( [1 - (1 - e_a)^{q-1}] / e_a \). The denominator cancels with \( e_a \) coming from \( \mu(X_j > a | R_j) \). For large \( a \), one has \( 1 - (1 - e_a)^{q-1} \approx 1 - \exp(-c(q-1)) \), which for moderate \( q \) is approximated by \( q e_a \). Thus, we obtain the order of magnitude of the upper bound given by the lemma. This means that the lemma cannot be improved; only a better constant may be obtained depending on the ad hoc properties of the process.

Furthermore, the upper bound for the approximation of \( \theta_j \) by \( \theta_1 \) is almost trivial to compute. For the exceedances, in both models, \( \mathbb{P}(X_0 > a | R_0) = e_a \). In the case of hitting \( \{X_0^{n-1} = a^n\} \), one gets \( \mathbb{P}(X_0^{n-1} = a^n | R_0) = m_a \bigl( n/a \bigr) \) for the block model. For the Smith model, its exponential decay on \( n \) was already computed in (13) and (14).

**Acknowledgments**

All authors were partially supported by the joint project FAPESP (SP-Brazil) and FCT (Portugal) with reference FAPESP/19805/2014. ACMF and JMF were partially supported by FCT projects PTDC/MAT-CAL/3884/2014 and PTDC/MAT-PUR/28177/2017, with national funds, and by CMUP (UID/MAT/00144/2019), which is funded by FCT with national (MCTES) and European structural funds through the programs FEDER, under the partnership agreement PT2020. We thank Mike Todd and the anonymous referees for useful comments and suggestions that helped to improve the paper.

**Appendix**

The following lemma establishes a general tool for computing the expected sojourn time in terms of the first and second moments of the cluster size.

**Lemma A.1.** Suppose that \( p(U_a) = 1 \) in the exceedance case, or that \( p(U_a) = 1 \) in the cylinder case. The mean sojourn time verifies
\[
\mathbb{E}_{U_a}(N_a) \frac{\theta_1(a)}{2} \left( \mathbb{E}_{E_{a}}(N_a^2) + \mathbb{E}_{E_{a}}(N_a) \right).
\] (A.1)

**Proof.** Consider first the case of exceedances, with \( p(U_a) = 1 \). Put

\[
\mathbb{E}_{U_a}(N_a) = \frac{1}{P(U_a)} \sum_{k=1}^{\infty} k \mathbb{P}(\bigcap_{j=0}^{k-1} X_j > a, X_k \leq a).
\] (A.2)

A classical equality for a stationary measure establishes that for all \( k \)

\[
\mathbb{P}(\bigcap_{j=0}^{k-1} X_j > a, X_k \leq a) = \mathbb{P}(X_{-1} \leq a, \bigcap_{j=0}^{k-1} X_j > a).
\] (A.3)

Then the last sum can be stated as

\[
\sum_{k=1}^{\infty} k \mathbb{P}(X_{-1} \leq a, N_a \geq k).
\]

Now, since \( N_a \) takes positive integer values, one can express its second moment as

\[
\mathbb{E}_{E_{a}}(N_a^2) = \sum_{k=1}^{\infty} (2k - 1) \mathbb{P}_{E_{a}}(N_a \geq k).
\]

Thus it follows that the right-hand side of (A.2) is

\[
\frac{\mathbb{P}(E_{a}) \mathbb{E}_{E_{a}}(N_a^2) + \mathbb{E}_{E_{a}}(N_a)}{2}.
\]

By (A.3) with \( k = 1 \) one gets \( \mathbb{P}(E_{a}) = \mathbb{P}(X_0 > a, X_1 \leq a) \) and this ends the proof of this case. To prove the case of \( n \)-cylinders with \( p(U_n) = 1 \), it is enough to replace the sets \( \{ X_j > a \} \) with the sets \( \{ X_j^{j+n-1} = a^n \} \), and \( U_a, \theta_1(a), E_a, N_a \) by \( U_n, \theta_1(n), E_n, N_n \).

□

**References**

[1] Abadi M 2006 Hitting, returning and the short correlation function *Bull. Braz. Math. Soc.* 37 593–609

[2] Abadi M, Cardeño L and Gallo S 2015 Potential well spectrum and hitting time in renewal processes *J. Stat. Phys.* 159 1087–106

[3] Abadi M, Freitas A C M and Freitas J M 2018 Dynamical counterexamples regarding the extremal index and the mean of the limiting cluster size distribution (arXiv:1808.02970)

[4] Abadi M, Gallo S and Rada-Mora E A 2018 The shortest possible return time of \( \beta \)-mixing processes *IEEE Trans. Inform. Theory* 64 4895–906

[5] Abadi M and Galves A 2004 A version of Maurer’s conjecture for stationary \( \psi \)-mixing processes *Nonlinearity* 17 1357–66

[6] Abadi M and Saussol B 2011 Hitting and returning to rare events for all alpha-mixing processes *Stoch. Process. Appl.* 121 314–23

[7] Abadi M and Vergne N 2009 Sharp error terms for return time statistics under mixing conditions *J. Theor. Probab.* 22 18–37

[8] Aytaç H, Freitas J M and Vaienti S 2015 Laws of rare events for deterministic and random dynamical systems *Trans. Am. Math. Soc.* 367 8229–78
[9] Azevedo D, Freitas A C M, Freitas J M and Rodrigues F B 2016 Clustering of extreme events created by multiple correlated maxima Physica D 315 33–48
[10] Azevedo D, Freitas A C M, Freitas J M and Rodrigues F B 2017 Extreme value laws for dynamical systems with countable extremal sets J. Stat. Phys. 167 1244–61
[11] Coelho Z 2000 Asymptotic laws for symbolic dynamical systems Topics in Symbolic Dynamics and Applications (Temuco, 1997) (London Mathematical Society Lecture Note Series vol 279) (Cambridge: Cambridge University Press) pp 123–65
[12] Freitas A C M, Freitas J M and Todd M 2012 The extremal index, hitting time statistics and periodicity Adv. Math. 231 2626–65
[13] Freitas A C M, Freitas J M and Todd M 2015 Speed of convergence for laws of rare events and escape rates Stoch. Process. Appl. 125 1653–87
[14] Haydn N T A 2013 Entry and return times distribution Dyn. Syst. 28 333–53
[15] Leadbetter M R, Lindgren G and Rootzén H 1983 Extremes and Related Properties of Random Sequences and Processes (Springer Series in Statistics) (New York: Springer)
[16] O’Brien G L 1987 Extreme values for stationary and Markov sequences Ann. Probab. 15 281–91
[17] Shields P C 1996 The Ergodic Theory of Discrete Sample Paths (Graduate Studies in Mathematics vol 13) (Providence, RI: American Mathematical Society)
[18] Smith R L 1988 A counterexample concerning the extremal index Adv. Appl. Probab. 20 681–3