WEAK CONVERGENCE TO EXTREMAL PROCESSES AND RECORD EVENTS FOR NON-UNIFORMLY HYPERBOLIC DYNAMICAL SYSTEMS

MARK HOLLAND AND MIKE TODD

ABSTRACT. For a measure preserving dynamical system $(\mathcal{X}, f, \mu)$, we consider the time series of maxima $M_n = \max\{X_1, \ldots, X_n\}$ associated to the process $X_n = \phi(f^{n-1}(x))$ generated by the dynamical system for some observable $\phi : \mathcal{X} \to \mathbb{R}$. Using a point process approach we establish weak convergence of the process $Y_n(t) = a_n(M_{[nt]} - b_n)$ to an extremal process $Y(t)$ for suitable scaling constants $a_n, b_n \in \mathbb{R}$. Convergence here taking place in the Skorokhod space $\mathbb{D}(0, \infty)$ with the J_1 topology. We also establish distributional results for the record times and record values of the corresponding maxima process.

1. INTRODUCTION

Consider a measure preserving dynamical system $(\mathcal{X}, f, \mu)$, where $\mathcal{X} \subset \mathbb{R}^d$, $f : \mathcal{X} \to \mathcal{X}$ is a measurable transformation, and μ is an f-invariant probability measure supported on $\mathcal{X}$. Given a measurable (observable) function $\phi : \mathcal{X} \to \mathbb{R}$ we consider the stationary stochastic process $X_1, X_2, \ldots$ defined as

$$X_i = \phi \circ f^{i-1}, \quad i \geq 1, \tag{1}$$

and its associated maximum process M_n defined as

$$M_n = \max(X_1, \ldots, X_n). \tag{2}$$

Under appropriate assumptions on the system $(\mathcal{X}, f, \mu)$, there exist scaling constants $a_n > 0$ and $b_n \in \mathbb{R}$ and a non-degenerate limit function $G(u)$ for which

$$\lim_{n \to \infty} \mu\{x \in \mathcal{X} : a_n(M_n(x) - b_n) \leq u\} = G(u). \tag{3}$$

Beyond the distributional limit established in (3), here we consider the continuous time process $\{Y_n(t) : t \geq 0\}$ defined by

$$Y_n(t) = \begin{cases} a_n(M_{[nt]} - b_n) & t \geq 1/n; \\ a_n(X_1 - b_n) & 0 < t < 1/n. \end{cases} \tag{4}$$

For each $n \geq 1$, $Y_n(t)$ is a random graph with values in the Skorokhod space $\mathbb{D}(0, \infty)$. Under suitable hypotheses on $(\mathcal{X}, f, \mu)$ we prove existence of a non-degenerate limit process $Y(t)$ so that $Y_n(t) \xrightarrow{d} Y(t)$ in $\mathbb{D}(0, \infty)$ with respect to the Skorokhod J_1 metric. Here, $\xrightarrow{d}$ denotes weak convergence (or convergence in distribution). The limit process $Y(t)$ will be the so called *extremal process* which we now define.

1.1. **Extremal processes and weak convergence.** Consider a general probability space $(\Omega, \mathcal{B}, \mu)$, where $\mathcal{B}$ is the σ-algebra of sets in the sample space Ω. If $X : \Omega \to \mathbb{R}$ is a random variable, we let $F(u) := \mu\{X \leq u\}$, and define finite dimensional distributions:

$$F_{t_1, \ldots, t_k}(u_1, \ldots, u_k) = F^{t_1}\left(\bigwedge_{i=1}^{k}\{u_i\}\right) F^{t_2 - t_1}\left(\bigwedge_{i=2}^{k}\{u_i\}\right) \cdots F^{t_k - t_{k-1}}(u_k), \tag{5}$$

---

with $t_1 < t_2 < \cdots < t_k$, and $\wedge$ denoting the minimum operation. Suppose that $Y_F(t)$ is a stochastic process with these finite dimensional distributions, i.e.

$$\mu\{Y_F(t_1) \leq u_1, \ldots Y_F(t_k) \leq u_k\} = F_{t_1,\ldots,t_k}(u_1, \ldots, u_k). \tag{6}$$

By the Kolmogorov extension theorem such a process exists and is called an *extremal-F process*. A version can be taken in $\mathbb{D}(0, \infty)$, i.e. continuous to the right with left hand limits. It turns out that $Y_F(t)$ is a Markov jump process, see [EKM, R2, R3], and further properties include:

- For all $t, s > 0$, $\mu\{Y_F(t + s) \leq x \mid Y_F(s) = y\} = F^t(x) \cdot 1_{\{x \geq y\}}$.
- For all $t, s > 0$, $\mu\{Y_F(t + s) = Y_F(t) \mid Y_F(s) = y\} = F^t(y)$. Setting $Q(y) = -\log F(y)$ implies that the holding time in state y is given by an exponential distribution with parameter $Q(y)$.
- If $\{t_i\}$ denotes the sequence of jump times (or points of discontinuity) for $Y_F(t)$ then

$$\mu\{Y_F(t_{i+1}) \leq x \mid Y_F(t_i) = y\} = \begin{cases} 1 - Q(x)/Q(y) & x > y; \\ 0 & \text{if } x < y. \end{cases} \tag{7}$$

Our main result is to show that for certain chaotic dynamical systems the process $Y_n(t)$ in (4) converges (weakly) to an extremal-G process $Y_G(t)$. This is the first time extremal processes have been used in the dynamical systems context. The mode of convergence to the extremal process is in distribution on $\mathbb{D}(0, \infty)$ with respect to the Skorokhod J_1-topology. To be precise, let $\mu_n = \mu \circ Y_n^{-1}(t)$, so that for all measurable $A \subset \mathbb{R}$, $Y_n^{-1}(t)(A) = \{\omega \in \Omega : Y_n(t)(\omega) \in A\}$. Then $Y_n(t) \xrightarrow{d} Y(t)$ in $\mathbb{D}(0, \infty)$ if for all bounded continuous functions φ on $\mathbb{D}(0, \infty)$,

$$\int_{(0,\infty)} \varphi(x) d\mu_n \to \int_{(0,\infty)} \varphi(x) d\mu.$$

The space $\mathbb{D}(0, \infty)$ consists of right continuous functions, with existence of limits to the left (cadlag functions) [S]. To overview the construction of the Skorokhod J_1-topology consider first the space $\mathbb{D}[a, b]$, and let $\|\cdot\|$ denote the uniform norm on $[a, b]$, so that $\|\varphi\| = \sup_{a \leq t \leq b} |\varphi(t)|$. Then a metric on $\mathbb{D}[a, b]$ is given by:

$$d_{a,b}(\varphi_1, \varphi_2) := \inf_{h \in \Lambda} \left\{ \|\varphi_1 \circ h - \varphi_2\| \vee \|h - \mathrm{id}\| \right\},$$

where Λ is the set of strictly increasing, continuous functions $h : [a, b] \to [a, b]$ such that $h(a) = a$ and $h(b) = b$. The function id is the identity mapping. This metric is not complete, but an equivalent (complete) metric can be constructed by a homeomorphism, see [R3, S].

The construction carries over to $\mathbb{D}(0, \infty)$ by use of the following metric: let $r_{a,b}\varphi(x)$ denote the restriction of $\varphi(x)$ to the interval $[a, b]$, and define

$$d_{0,\infty}(\varphi_1, \varphi_2) := \int_0^1 \int_1^\infty e^{-t}(1 \wedge d_{s,t}(r_{s,t}\varphi_1, r_{s,t}\varphi_2)) \, dt \, ds.$$

Then convergence $\varphi_n \to \varphi$ in $\mathbb{D}(0, \infty)$ holds in the J_1 metric if $d_{0,\infty}(\varphi_n, \varphi) \to 0$ at each continuity point of φ. We remark that similar metrics can be constructed on $\mathbb{D}(-\infty, 0)$ and $\mathbb{D}(-\infty, \infty)$.

2. Main results: weak convergence to an extremal process.

As noted above, the heart of this paper is to prove that the process Y_n converges to a certain extremal process. In this section we give the main results in this direction after defining the relevant short-term and long-term mixing conditions that guarantee this convergence. In the next section these results will then be interpreted in terms of record times and record values via the Continuous Mapping Theorem.

We consider $(\mathcal{X}, f, \mu)$ a measure preserving system, and assume that μ is absolutely continuous with respect to Lebesgue m with density ρ (note that this condition could be removed in line with [FFT2]). Within this article, unless otherwise stated we consider observable functions of

the form $\phi(x) = \psi(\text{dist}(x, \tilde{x}))$. Here $\tilde{x}$ is a chosen point in $\mathcal{X}$, and $\psi : [0, \infty) \to \mathbb{R}$ is a measurable function with $\sup_{v \in [0, \infty)} \psi(v) = \psi(0)$.

2.1. Probabilistic mixing and recurrence conditions.

Let $\mathcal{S}$ denote the semi-ring of subsets whose elements are intervals of type $[a, b)$ for $a, b \in \mathbb{R}^+$, and let $\mathfrak{R}$ denote the ring generated by $\mathcal{S}$. So for every $A \in \mathfrak{R}$, there exists $k \in \mathbb{N}$, and intervals $I_1, \ldots, I_k \in \mathcal{S}$ such that $A = \cup_j I_j$. To fix notation, for any $\alpha \in \mathbb{R}$ and $I = [a, b) \in \mathcal{S}$ we have $\alpha I = [\alpha a, \alpha b)$, and $I + \alpha = [a + \alpha, b + \alpha)$. This notation extends in a natural way to any $A \in \mathfrak{R}$. For $A \in \mathfrak{R}$, we let

$$M(A) = \max\{X_i, i \in A \cap \mathbb{Z}\}.$$

In the case $A = [0, n)$, we have $M(A) = M_n$.

Given this setup, the probabilistic condition we define first concerns that of mixing and asymptotic independence of maxima in different blocks. We make this precise as follows. Given $0 < x_1 \leq x_2 \leq \cdots \leq x_r$, suppose that $u_n^{(1)} \geq u_n^{(2)} \geq \cdots \geq u_n^{(r)}$ are such that

$$n\mu\{X_1 > u_n^{(k)}\} \to x_k, \quad \forall k \leq r. \tag{8}$$

Definition 2.1. *We say that condition $\mathcal{D}_r(u_n^{(k)})$ holds, if for any disjoint collection of sets $\mathcal{A}_1, \ldots, \mathcal{A}_r \in \mathfrak{R}$:*

$$|\mu\left(\{X_1 > u_{n,0}\} \cap \{M(\mathcal{A}_1 + t) \leq u_{n,1}, \ldots, M(\mathcal{A}_r + t) \leq u_{n,r}\}\right)$$
$$- \mu\left(\{X_1 > u_{n,0}\}\right)\mu\left(\{M(\mathcal{A}_1) \leq u_{n,1}, \ldots, M(\mathcal{A}_r) \leq u_{n,r}\}\right)| \leq \gamma(n, t),$$

where $n\gamma(n, t) \to 0$ for some integer sequence $t_n = o(n)$, and for each i, $u_{n,i}$ denotes any one of the $u_n^{(k)}$, $(1 \leq k \leq r)$.

In the next definition, we consider the frequency of exceedances of the X_j (in a probabilistic sense) over a threshold sequence $\{u_n\}$.

Definition 2.2. *We say that condition $\mathcal{D}'_r(u_n^{(k)})$ holds for a sequence $\{u_n^{(k)}\}$ if*

$$\lim_{k \to \infty} \limsup_{n \to \infty} n \sum_{j=2}^{n/k} \mu\left(\left\{X_1 > u_n^{(k)}, X_j > u_n^{(k)}\right\}\right) = 0. \tag{9}$$

In the case where it is known that (3) holds for given sequences $\{a_n\}, \{b_n\}$ and non-degenerate distribution function $G(u)$ we can usually take

$$u_n^{(k)}(x_k) = a_n^{-1}G^{-1}(e^{-x_k}) + b_n \tag{10}$$

in the above definitions. This is certainly true in the i.i.d case, and for a wide class of dynamical systems. In fact when this scaling rule applies it is more natural to re-write (8) and assume the representation

$$n\mu\{X_1 > u/a_n + b_n\} \to -\log G(u), \tag{11}$$

i.e. to replace x_k by $-\log G(u)$ for some $u \in \mathbb{R}$. The limit relation (3) becomes a consequence of (11) (rather than the converse) and is known as the *Poisson approximation*, see [EKM].

We briefly compare our conditions with previous ones. In the classical probability literature there were conditions D and D', see [LLR], but Freitas and Freitas [FF, FFT1], inspired by Collet [C] converted these to conditions more straightforwardly checkable in a dynamical context. In conjunction with (11) these conditions are then used to imply the limit law in (3). Condition $\mathcal{D}_r(u_n^{(k)})$ is very similar to $D_3(u_n)$ used in [FFT1, Section 4] (itself similar to $D(u_n)$ in that paper): the only difference being the multiple thresholds $u_n^{(k)}$ for each given n leading to a slightly stronger condition on the mixing. Also $\mathcal{D}'_r(u_n^{(k)})$ is nearly identical to $D'(u_n)$ in [FFT1], but in this case the different thresholds do not add any real strength to the condition since they are each checked independently. The similarities between the conditions here and in [FFT1] are strong enough that by examining the proofs (most importantly of the D condition), any

3

dynamical system to which the old conditions have been shown to apply can quite easily be seen to satisfy our new conditions.

Theorem 2.3. *Suppose that $(\mathcal{X}, f, \mu)$ is a measure preserving dynamical system, and suppose that there exist sequences $\{a_n\}, \{b_n\}$ and a non-degenerate function $G(u)$ such that (11) holds. Suppose that for any $r \geq 1$, and any sequence $\{x_i, i \leq r\}$, condition $\mathcal{D}_r(u_n^{(k)})$ holds together with $\mathcal{D}'_r(u_n^{(k)})$. Then the process $Y_n(t) = a_n(M_{[nt]} - b_n)$ converges weakly to an extremal-G process $Y_G(t)$ in $\mathbb{D}(0, \infty)$ (endowed with the Skorokhod J_1 topology).*

We make several remarks about Theorem 2.3. First note that the hypotheses on $(\mathcal{X}, f, \mu)$ as stated are quite weak. However to check conditions $\mathcal{D}_r(u_n^{(k)})$ and $\mathcal{D}'_r(u_n^{(k)})$ in specific applications we usually require absolute continuity of μ and estimates on the recurrence time statistics (e.g. ergodicity and/or decay of correlation estimate). We discuss this in Section 4. A further remark is that conditions $\mathcal{D}_r(u_n^{(k)})$ and $\mathcal{D}'_r(u_n^{(k)})$ are sufficient conditions to ensure convergence to $Y_G(t)$. In the i.i.d case, existence of the limit (3) alone is enough to ensure convergence, see [R3]. Thus in certain situations we might expect the conclusion of Theorem 2.3 to hold under weaker hypotheses. We discuss this further in Section 6.

Equation (11) in itself imposes regularity conditions on the invariant density of μ and on the form of the observable $\phi(x) = \psi(\text{dist}(x, \tilde{x}))$. For linear scaling sequences $\{a_n\}, \{b_n\}$ there turn out to be three non-degenerate types for the distribution function G, namely Gumbel, Fréchet and Weibull [EKM]. We make the following remark detailing how these types can arise, and for the precise computations see e.g. [HNT1].

Remark 2.4. *Suppose that $(\mathcal{X}, f, \mu)$ is a measure preserving system, for given $\tilde{x} \in \mathcal{X}$ the invariant density $\rho(\tilde{x})$ lies in $(0, \infty)$ and for any $r \geq 1$, and any sequence $\{x_i, i \leq r\}$ condition $\mathcal{D}_r(u_n^{(k)})$ holds together with $\mathcal{D}'_r(u_n^{(k)})$. We have the following cases.*

  (i) *If $X_n = -\log(\text{dist}(f^{n-1}(x), \tilde{x}))$, then $Y_n(t) = M_{[nt]} - \log n$ converges weakly to the extremal-G process $Y_G(t)$ in $\mathbb{D}(0, \infty)$. In this case $G(u) = e^{-2\rho(\tilde{x})e^{-u}}$.*
  (ii) *If $X_n = \text{dist}(f^{n-1}(x), \tilde{x})^{-\alpha}$ for $\alpha > 0$, then $Y_n(t) = n^{-\alpha}M_{[nt]}$ converges weakly to the extremal-G process $Y_G(t)$ in $\mathbb{D}(0, \infty)$. In this case $G(u) = e^{-2\rho(\tilde{x})u^{-\alpha}}$.*
  (iii) *If $X_n = C - \text{dist}(f^{n-1}x, \tilde{x})^{\alpha}$ for $\alpha > 0$, then the process $Y_n(t) = n^{\alpha}(M_{[nt]} - C)$ converges weakly to the extremal-G process $Y_G(t)$ in $\mathbb{D}(0, \infty)$. In this case $G(u) = e^{-2\rho(\tilde{x})(-u)^{\alpha}}$.*

In each case, $\mathbb{D}(0, \infty)$ is endowed with the Skorokhod J_1 topology.

Given the extremal process $Y_G(t)$ its (path) inverse is defined by:

$$Y_G^{\leftarrow}(t) = \inf\{x : Y(x) > t\},$$

where the domain of $Y_G^{\leftarrow}$ is the left and right end-points of G. Given $Y_n(t) = a_n(M_{[nt]} - b_n)$, denote the inverse path process by $Y_n^{\leftarrow}(t)$. We have the following result.

Theorem 2.5. *Suppose that $(\mathcal{X}, f, \mu)$ is a measure preserving dynamical system, and suppose that there exist sequences $\{a_n\}, \{b_n\}$ and a non-degenerate function $G(u)$ such that (11) holds. Suppose that for any $r \geq 1$, and any sequence $\{x_i, i \leq r\}$ condition $\mathcal{D}_r(u_n^{(k)})$ holds together with $\mathcal{D}'_r(u_n^{(k)})$. Then the process $Y_n^{\leftarrow}(x)$ converges weakly to the inverse extremal-G process $Y_G^{\leftarrow}(t)$ in $\mathbb{D}(E)$, (endowed with the Skorokhod J_1 topology), where $E \subset \mathbb{R}$ is the domain of definition of G.*

For example, in Theorem 2.5 we take $E = (-\infty, \infty)$ in the case of the Gumbel distribution. For the Fréchet and Weibull distributions we take $E = (0, \infty)$ and $E = (-\infty, 0)$ respectively.

3. The distribution of record times and record values

Given the processes $Y_G(t)$ and $Y_G^{\leftarrow}(t)$, we describe next the distribution of their jump values, i.e. the locations of their discontinuities. This has natural application to the theory of record

times and record values which we describe as follows. Consider the original processes $\{X_n\}$, $\{M_n\}$, and let $\tau_1 = 1$. Define a strictly increasing sequence $\{\tau_k\}$ via:

$$\tau_k = \inf\{j > \tau_{k-1} : X_j > M_{j-1}\}. \tag{12}$$

Then this sequence $\{\tau_k\}$ forms the *record times* associated to the process M_n, namely the times where M_n jumps. The corresponding record values are given by the $X_{\tau_k} = M_{\tau_k}$. For the process $Y_n(t) = a_n(M_{[nt]} - b_n)$, we see that the jumps of $Y_n(t)$ occur precisely at the times $t_k = \tau_k/n$ where τ_k is a record time. The jump values $Y_n(t_k)$ are then the (normalised) record values $a_n(X_{\tau_k} - b_n)$. We shall use point process theory to describe the distributional behaviour of these jump times and jump values.

3.1. **Overview of point process theory.** To study extremal processes and their corresponding jump processes we use a point process approach. We recall some general properties of point processes, see [R3, Chapter 3]. Let $(\Omega, \mathcal{B}, \mu)$ be a probability space, where $\mathcal{B}$ is a σ-algebra of subsets of Ω. Let $E \subset \mathbb{R}^d$ be a state space, with Borel σ-algebra $\mathcal{E}$. The set E is the region in which points will be defined. Given a sequence of (vector-valued) random variables $X_i : \Omega \to E$, consider the quantity $\xi = \sum_{i=1}^{\infty} \delta_{X_i}$, where δ_x is the Dirac measure at x. Then $\xi : (\Omega, \mathcal{B}, \mu) \to (M_p(E), \mathcal{M}_p(E))$ defines a point process. The set $M_p(E)$ is the collection of point (counting) measures m on E, with $m(A) < \infty$ if $A \subset E$ is compact. The set $\mathcal{M}_p(E)$ is the corresponding σ-algebra of subsets of $M_p(E)$. Thus given $A \subset E$, $\xi(A) \in \mathbb{N}$ is itself a random variable associated to the sequence $\{X_i\}$, i.e., given $\omega \in \Omega$, and $A \subset E$,

$$\xi := \xi(A, \omega) = \sum_{i=1}^{\infty} \delta_{X_i(\omega)}(A) = \sharp\{i : X_i(\omega) \in A\}.$$

Of interest to us are the special class of point processes known as *Poisson random measures* (PRM). We say that a point process ξ is a PRM with mean measure λ if the following hold:

(1) For all $A \in \mathcal{E}$,
$$\mu(\xi(A, \cdot) = k) = \frac{\lambda(A)^k e^{-\lambda(A)}}{k!}.$$

(2) If $A_i \cap A_j = \emptyset$ for all $i, j \in \{1, \ldots, m\}$, and all $m \geq 1$ then $\xi(A_1), \ldots, \xi(A_m)$ are independent random variables.

The mean (or *intensity*) measure λ satisfies $\lambda(A) = \mathbb{E}(\xi(A))$ for all $A \in \mathcal{E}$. If we can write $\lambda = \int_A \gamma(t)\, dt$, then we call $\gamma(t)$ the intensity of the process ξ.

We now consider convergence of a sequence of point processes. A sequence $\{\xi_n\}$ of point processes converges in distribution (or converges weakly) to a point process ξ in $M_p(E)$ if for any finite collection $B_1, \ldots, B_m$ of bounded Borel sets in E with $\mu(\xi(\partial B_i) = 0) = 1$,

$$(\xi_n(B_1), \xi_n(B_2), \ldots, \xi_n(B_m)) \to (\xi(B_1), \xi(B_2), \ldots, \xi(B_m)).$$

To prove such convergence for sequences of (simple) point processes it suffices to check the following criteria due to Kallenberg [K]:

(a) $\mathbb{E}(\xi_n(B)) \to \mathbb{E}(\xi(B))$, where B is a semi-closed rectangle in E.
(b) $\mu\{\xi_n(B) = 0\} \to \mu\{\xi(B) = 0\}$ for all finite unions of semi-closed rectangles $B \subset E$.

In the definition above, we take semi-closed (and disjoint unions) of intervals of the form $(a, b]$ for subsets of $\mathbb{R}$, while in $\mathbb{R}^2$ we take semi-closed (and disjoint unions of) rectangles of the form $(a, b] \times (c, d]$.

Next we discuss vague convergence of measures, and convergence of certain transformations of measures. These concepts will be used in Section 5. We say that a sequence of measures $\{\mu_n\}$ convergences vaguely to μ (written $\mu_n \xrightarrow{v} \mu$), if $\int g\, d\mu_n \to \int g\, d\mu$ for all continunuous functions g that are compactly supported. In general vague convergence does not imply weak convergence, e.g. $\mu_n = \delta_n$ (the Dirac mass at $n \in \mathbb{N}$) converges vaguely to $\mu = 0$, but not weakly. For point processes, there exists a vague metric which makes $M_p(E)$ a complete metric space.

To study weak convergence of certain functionals of point process we will make use of the *Continuous Mapping Theorem* (CMT). Consider metric spaces M and M' and let μ be a probability measure. A function $h : M \to M'$ is a.s. continuous if the set of discontinuities of h has μ-measure zero. Suppose that $\xi_n \xrightarrow{d} \xi$ in M, then the Continuous Mapping Theorem asserts that $h(\xi_n) \xrightarrow{d} h(\xi)$ in M' provided h is a.s. continuous.

3.2. **The record time and record value point processes.** We now state distributional results for the record-time and record-value jump processes. For the process $Y_n(t) = a_n(M_{[nt]} - b_n)$, recall that $t_k = \tau_k/n$ are the jump times. i.e. where $X_{\tau_k} > M_{\tau_k - 1}$. The jump values are given by $Y(t_k)$. We consider the following two point processes (defined on subsets of $\mathbb{R}$):

$$\mathcal{R}_n := \sum_{j=1}^{\infty} \delta_{\frac{j}{n}} \cdot 1_{\{X_j > M_{j-1}\}}, \quad \mathcal{V}_n := \sum_{\tau_k} \delta_{Y_n(\tau_k/n)}, \tag{13}$$

the former is the *record time process* and the latter is the *record value process*.

Theorem 3.1. *Suppose that $(\mathcal{X}, f, \mu)$ is a measure preserving dynamical system, and suppose that there exist sequences $\{a_n\}, \{b_n\}$ and a non-degenerate function $G(u)$ such that (11) holds. Suppose that for any $r \geq 1$ and any sequence $\{x_i, i \leq r\}$ condition $\mathcal{D}_r(u_n^{(k)})$ holds together with $\mathcal{D}_r'(u_n^{(k)})$. We have the following cases:*

(1) *The point process $\mathcal{R}_n$ converges weakly to the point process $\mathcal{R}$ on state space $(0, \infty)$. The process $\mathcal{R}$ is a PRM with intensity $\gamma(t) = 1/t$. i.e. for any $0 < a < b < \infty$:*

$$\lim_{n \to \infty} \mu\{\mathcal{R}_n(a, b) = k\} = \frac{(\log(b/a))^k}{k!} \cdot \frac{a}{b}.$$

(2) *The point process $\mathcal{V}_n = \sum_{\tau_k} \delta_{Y(\tau_k/n)}$ converges weakly to the point process $\mathcal{V}$ on state space $E \subset \mathbb{R}$ (the domain of G), where $\mathcal{V}$ is a PRM with intensity measure λ_V given by $\lambda_V([a, b]) = -\log(-\log G(b)) + \log(-\log G(a))$.*

Hence Theorem 3.1 implies that the process $\mathcal{R}_n$ converges to a PRM $\mathcal{R}$ on $E = (0, \infty)$ with intensity $\gamma(t) = 1/t$ irrespective of the underlying distribution function G. We remark that if item (1) of the Theorem 3.1 holds for a particular observable $\phi : \mathcal{X} \to \mathbb{R}$, then $\mathcal{R}_n \xrightarrow{d} \mathcal{R}$ holds for any injective and monotone increasing transformation of X_1. By contrast, the limit process $\mathcal{V}_n$ does depend on G, and hence on the form of ϕ. We do not consider all possibilities, but remark that in the case of G being the Gumbel distribution,

$$\lim_{n \to \infty} \mu\{\mathcal{V}_n(a, b) = k\} = \frac{(b - a)^k e^{-(b-a)}}{k!}.$$

This would apply to the dynamical process $X_n = -\log \text{dist}(f^{n-1}(x), \tilde{x})$. We remark further that the process $\mathcal{V}_n$ determines the jump *times* for inverse process $Y_G^{\leftarrow}(t)$.

4. APPLICATION OF RESULTS

In this section we give an overview of dynamical system models that fit within our assumptions, namely $\mathcal{D}_r(u_n^{(k)})$ and $\mathcal{D}_r'(u_n^{(k)})$. As mentioned above, conditions of this type have been considered in many recent papers. In most applications it is the short-range condition which is hardest to prove, and as noted above, once the usual short range condition is proved, $\mathcal{D}_r'(u_n^{(k)})$ follows immediately. Checking that the long-range mixing condition $\mathcal{D}_r(u_n^{(k)})$ holds follows almost exactly as usual, namely via decay of correlations:

We say that $(\mathcal{X}, f, \nu)$ has decay of correlations in Banach spaces $\mathcal{B}_1$ versus $\mathcal{B}_2$, (DC), if there exists a monotonically decreasing sequence $\Theta(j) \to 0$ such that for all $\varphi_1 \in \mathcal{B}_1$ and all $\varphi_2 \in \mathcal{B}_2$:

$$\left| \int \varphi_1 \cdot \varphi_2 \circ f^j d\nu - \int \varphi_1 d\nu \int \varphi_2 d\nu \right| \leq \Theta(j) \|\phi_1\|_{\mathcal{B}_1} \|\phi_2\|_{\mathcal{B}_2},$$

where $\| \cdot \|_{\mathcal{B}}$ denotes the norm in space $\mathcal{B}$.

If this condition holds and $\mathcal{B}_2$ contains indicator functions on balls with norm of order the measure of the ball, and $\Theta(j)$ decays fast enough then it is easy to prove $\mathcal{D}_r(u_n^{(k)})$. When this doesn't happen, there are approximation arguments (eg. [Gu]) to derive the same result.

In checking condition $\mathcal{D}'_r(u_n^{(k)})$ the form of the observable is of significance, and in most applications observables take the form $\phi(x) = \psi(\mathrm{dist}(x, \tilde{x}))$ with $\mathrm{dist}(\cdot, \cdot)$ the Euclidean metric. Checking $\mathcal{D}'_r(u_n^{(k)})$ is then reduced to understanding the recurrence statistics for asymptotically shrinking balls. For other types of observables with general level set geometries, checking $\mathcal{D}'_r(u_n^{(k)})$ (and its variant) becomes a much harder problem.

We will list a set of dynamical systems to which our results apply, where the details of how to check our mixing conditions are essentially the same as in the references given.

4.1. **Expanding interval maps.** The simplest example is the tent map $f(x) = 1 - |1 - 2x|$, defined for $x \in [0, 1]$. In this case $\nu = \mathrm{Leb}$, and the system has exponentially decaying $\Theta(j)$ as defined in condition (DC) for $\mathcal{B}_1 = \mathrm{Lip}$ and $\mathcal{B}_2 = L^\infty$. The much more general case of Rychlik maps was considered in [FFT1]: here $\mathcal{B}_1 = BV$ and $\mathcal{B}_2 = L^1$.

4.2. **Non-uniformly expanding maps: rapid mixing.** Consider the quadratic map $f(x) = ax(1 - x)$, defined for $x \in [0, 1]$ and parameter set $a \in [0, 4]$. For a positive measure set of parameters it is known that ν is absolutely continuous with respect to Lebesgue measure and condition (DC) holds for $\Theta(n) = \theta_0^n$, (some $\theta_0 < 1$). Our conditions can be shown to hold using ideas from [C].

4.3. **Intermittent maps with polynomial decay of correlations.** Consider the class of intermittent type maps $f : [0, 1] \to [0, 1]$ which take the form

$$f(x) = \begin{cases} x(1 + 2^\alpha x^\alpha) & \text{if } 0 \leqslant x < \frac{1}{2}; \\ 2x - 1 & \text{if } \frac{1}{2} \leqslant x \leqslant 1, \end{cases}$$

with $\alpha \in (0, 1)$. This system admits an absolutely continuous invariant measure ν and condition (DC) applies, with $\Theta(n) = O\left(n^{1 - \frac{1}{\alpha}}\right)$, see [Y2]. Our conditions hold as in [HNT1], namely there is an explicit value $\alpha_0 \in (0, 1)$ such that conditions $\mathcal{D}_r(u_n^{(k)})$ and $\mathcal{D}'_r(u_n^{(k)})$ hold for all $\alpha \in [0, \alpha_0)$. We remark that it is the methodology used in the actual checking of the conditions that leads to the permissible range of α, and the bound α_0. It is a conjectural on whether we can take $\alpha_0 = 1$, see Section 6.

5. Generating extremal and jump processes from point processes

Our approach to proving Theorems 2.3, 2.5 and 3.1 is to consider weak convergence of point processes ξ_n (defined on subsets of the plane) and apply the CMT. As before we suppose that there exist constants a_n, b_n and a non-degenerate function G such that $n\mu\{X_1 > u/a_n + b_n\} \to -\log G(u)$, and therefore have in mind a dynamical system $(\mathcal{X}, f, \mu)$ with absolutely continuous invariant measure μ, and observable function $\phi(x) = \psi(\mathrm{dist}(x, \tilde{x}))$, with $\psi : \mathbb{R}^+ \to \mathbb{R}$ regularly (or slowly) varying. The key tool is the following theorem.

Theorem 5.1. *Suppose that $(\mathcal{X}, f, \mu)$ is a measure preserving dynamical system, and suppose that there exist sequences $\{a_n\}, \{b_n\}$ and a non-degenerate function $G(u)$ such that (11) holds. Suppose that for any $r \geq 1$, and any sequence $\{x_i, i \leq r\}$ condition $\mathcal{D}_r(u_n^{(k)})$ holds together with $\mathcal{D}'_r(u_n^{(k)})$. Then the point process $\xi_n = \sum_{i=1}^n \delta_{z(i,n)}$ with $z(i, n) = (\frac{i}{n}, a_n(X_i - b_n))$ converges weakly to a PRM in state space $E = [0, \infty) \times \mathbb{R}$ whose intensity measure $\lambda = \mathrm{Leb} \times \lambda_G$ with $\lambda_G([a, b]) = \log G(b) - \log G(a)$.*

We shall postpone the proof of this theorem to Section 7 as it relies heavily on conditions $\mathcal{D}_r(u_n^{(k)})$ and $\mathcal{D}'_r(u_n^{(k)})$. We now show how the main theorems follow from Theorem 5.1 and the CMT. Our approach follows that of [R2, R3] and an outline is as follows. Let M and M' be two metric spaces with associated metrics d and d' (resp.). A main difficulty is showing when a given map $h : M \to M'$ is a.s. continuous. An element $\xi \in M$ is a discontinuity of h if there exists a sequence $\{\xi_n\}$ with $d(\xi_n, \xi) \to 0$ but $d'(h(\xi_n), h(\xi)) \not\to 0$. The map is a.s. continuous if the set of discontinuities $\xi \in M$ has μ-measure 0. This agrees with the usual notion of (dis)continuity, although we must keep track of the underlying metrics being used. For definiteness, consider the case where $M = M_p(E)$ for some $E \subset \mathbb{R}^m$, $(m \geq 1)$, and $M' = \mathbb{D}(0, \infty)$. We take d to be the vague metric and take d' to be the Skorokhod J_1 metric. To show a.s. continuity of h, we must show that $d'(h(\xi_n), h(\xi)) \to 0$ for all sequences ξ_n with $\xi_n \xrightarrow{v} \xi$, and for μ-a.e. $\xi \in M$. Applying the CMT will then imply that $h(\xi_n) \xrightarrow{d} h(\xi)$ for all such sequences $\xi_n \xrightarrow{d} \xi$.

Proof of Theorem 2.3. Given the planar point process $\xi := \sum_{i=1}^{\infty} \delta_{(t_i, y_i)}$, let $H_1 : M_p((0, \infty) \times \mathbb{R}) \to \mathbb{D}(0, \infty)$ be the real valued function defined by $H_1(\xi)(t) = \sup\{y_i : t_i \leq t\}$. Then H_1 maps point processes to $\mathbb{D}(0, \infty)$. If ξ is a PRM then as is shown in [R3, Chapter 4] the map H_1 is a.s. continuous with respect to ξ. This is achieved by taking any sequence $\{\xi_n\}$ with $\xi_n \xrightarrow{v} \xi$ (i.e. converging vaguely) and showing that $d_{a,b}(h(\xi_n)(t), h(\xi)(t)) \to 0$ for any $0 < a < b < \infty$. Checking the latter is sufficient to prove convergence in $\mathbb{D}(0, \infty)$.

In the case where ξ is a PRM with intensity measure $\lambda = \text{Leb} \times \lambda_G$, (where $\lambda_G([a, b]) = \log G(b) - \log G(a)$), then $H_1(\xi)(t)$ has finite dimensional distributions which coincide with that of an extremal-G process $Y_G(t)$. Hence for the process ξ_n defined in Theorem 5.1, the CMT asserts that $H_1(\xi_n)(t) \xrightarrow{d} H_1(\xi)(t)$, and so $Y_n(t)$ converges weakly to an extremal-G process $Y_G(t)$. □

Proof of Theorem 2.5. Given a planar point process $\xi := \sum_{i=1}^{\infty} \delta_{(t_i, y_i)}$ consider the function H_2 defined by $H_2(\xi)(t) = \inf\{t_i : y_i > t\}$. This function is again a.s. continuous with respect to a PRM ξ [R3, Chapter 4]. If in particular ξ is a PRM with intensity measure $\lambda = \text{Leb} \times \lambda_G$, (and $\lambda_G([a, b]) = \log G(b) - \log G(a)$), then $H_2(\xi)(t)$ has finite dimensional distributions equivalent to those of $Y_G^{\leftarrow}(t)$ (the inverse of $Y_G(t)$). Hence for the process ξ_n defined in Theorem 5.1, the CMT asserts that $H_2(\xi_n)(t) \xrightarrow{d} H_2(\xi)(t)$, and so $Y_n^{\leftarrow}(t)$ converges weakly to $Y_G^{\leftarrow}(t)$.

We remark that it is tempting to apply H_1 and then a mapping $\tilde{H}$ with $\tilde{H}(y) = y^{\leftarrow}$ for $y \in \mathbb{D}(0, \infty)$. However this latter map is not continuous on $\mathbb{D}(0, \infty)$. □

Proof of Theorem 3.1. We first consider item (1), and the process $\mathcal{R}_n$. Consider the subset $\widetilde{\mathbb{D}}(0, \infty)$ of $\mathbb{D}(0, \infty)$ consisting of functions which are constant between isolated jumps (i.e., the jumps do not accumulate anywhere in $(0, \infty)$). For an element $Y(t) \in \widetilde{\mathbb{D}}(0, \infty)$, let $H_3 : \widetilde{\mathbb{D}}(0, \infty) \to M_p(0, \infty)$ be the counting function: $H_3(Y(t)) = \sum_i \delta_{t_i}(0, t)$, where t_i are jump times for $Y(t)$. As shown in [R3] the function H_3 is a.s. continuous when restricted to functions on $\widetilde{\mathbb{D}}(0, \infty)$. This is achieved by taking a sequence $y_n \in \widetilde{\mathbb{D}}(0, \infty)$ converging to $y \in \widetilde{\mathbb{D}}(0, \infty)$ (with respect to J_1 metric), and then showing that $H_3(y_n) \to H_3(y)$ in the vague metric on $M_p(0, \infty)$. The basic observation here is that closeness of the graphs of y_n and y (in the J_1 sense) implies that their discontinuities are close. Hence the corresponding point masses of $H_3(y_n)$ and $H(y)$ are close (in the vague metric sense).

If $Y_G(t)$ is an extremal-G process, then $H_3(Y(t))$ is a PRM on $(0, \infty)$ with intensity $\gamma(t) = 1/t$. Hence to get the required convergence result for $\mathcal{R}_n$ we apply the composition $H_3 \circ H_1$ to the planar point process and then the CMT.

Now consider item (2), and the process $\mathcal{V}_n$. As in the proof of item (1), we again consider the function H_3, but this time apply it to elements of $Y_G^{\leftarrow}(t) \in \widetilde{\mathbb{D}}(-\infty, \infty)$. The corresponding process $H_3(Y_G^{\leftarrow})(t) = \sum_i \delta_{Y(t_i)}$ is a PRM with mean-measure $\lambda([a, b]) = -\log(-\log G(b)) + \log(-\log G(a))$. Hence to get the required convergence result for $\mathcal{V}_n$, we apply the composition

$H_3 \circ H_2$ to the planar point process sequence $\{\xi_n\}$. This composition is a.s. continuous with respect to the PRM ξ, and hence we apply the CMT. $\qquad\square$

6. Discussion

In this article we have developed a general approach to prove convergence to extremal processes for chaotic dynamical systems. We have also established consequential results such as determining the statistics of record events. We now make several remarks about the wider applicability of our results and scope for future investigations. Firstly our results apply to dynamical systems that satisfy the $\mathcal{D}_r(u_n^{(k)})$ and $\mathcal{D}'_r(u_n^{(k)})$ conditions. For a wide class of non-uniformly expanding dynamical systems, such as those considered within [C, FFT1, HNT1] our results apply. We note that these conditions are sufficient for our results. If these conditions fail to hold then it is possible that the conclusions of our results still hold. This might be the case for certain non-mixing systems, such as suspended flows considered in [HNT1]. Furthermore, we might ask on whether it is possible to by-pass the checking of conditions $\mathcal{D}_r(u_n^{(k)})$ and $\mathcal{D}'_r(u_n^{(k)})$ to obtain our convergence results. An assumption on the existence of an extreme distribution such as (3) is still expected to be required. However, our results would then apply to a broader class of systems where a link between extremes and return time statistics is known, see [FFT1].

For hyperbolic systems with attractors (i.e. those that support Sinai-Ruelle-Bowen (SRB) measures) then we expect similar conclusions to hold on the existence of an extremal process. For related results on extremes for hyperbolic systems, see [CC, GHN, Z]. To study extremal processes and records for these systems further work is required. Particular issues include the actual checking of the conditions $\mathcal{D}_r(u_n^{(k)})$ and $\mathcal{D}'_r(u_n^{(k)})$, and controlling the regularity of the SRB measure μ (or regularity of the observable function) to ensure existence of a non-degenerate limit function G.

It is possible to investigate further situations where $\mathcal{D}'_r(u_n^{(k)})$ fails, for example in the case of observable functions maximised at periodic points. For such observables the limit function G in (3) incorporates an extra parameter known as an *extremal index*, and for dynamical systems this has been recently studied, for example in [FFT3]. To ensure convergence to a corresponding extremal process, alternative conditions to $\mathcal{D}'_r(u_n^{(k)})$ would need to be formulated.

The results we have stated about extremal processes and records are not exhaustive. Combining Theorem 5.1 with the CMT we can obtain results about the distribution of inter-record times (i.e. $t_{i+1} - t_i$), the jump sizes $(Y(t_{i+1}) - Y(t_i))$, and the distributions governing the maximum inter-record times/jump sizes, see [R2, R1, R3]. For systems that satisfy conditions $\mathcal{D}_r(u_n^{(k)})$ and $\mathcal{D}'_r(u_n^{(k)})$ the joint asymptotic distribution of maxima can also be derived, see [LLR, Section 5.6]. In other directions beyond the scope of this work, we should mention that planar convergence of point processes to a PRM (i.e. conclusions similar to that of Theorem 5.1) have been used in the study of convergence to Levy processes for certain dynamical systems, see [T].

Finally we note that our results are all on distributional convergence. In the i.i.d case almost sure convergence results for records is known. For an i.i.d process $\{X_n\}$, let τ_n denote the time to the n'th record (as in (12)), and W_n the number or records observed up to time n. Then almost surely we have $(\tau_n)^{1/n} \to e$ and $W_n/\log n \to 1$. For dynamical systems we conjecture that similar results hold. For the maximum process, recent work on almost sure convergence is established in [HNT2]. In the i.i.d case, almost sure convergence for records is proved by embedding the maximum process M_n into the extremal process $Y(t)$ for $t \in \mathbb{N}$. However, for dependent processes driven by dynamical systems new ideas are required.

7. Proof of Theorem 5.1

It remains to prove Theorem 5.1 and we do this as follows. In the first instance we show in Proposition 7.1 how conditions $\mathcal{D}_r(u_n^{(k)})$ and $D_r(u_n)$ can be used to obtain asymptotic independence of blocks of maxima on disjoint intervals. Using this result, we then apply the criteria of Kallenberg to prove the theorem via a thinning construction as used in [LLR].

7.1. Asymptotic independence of maxima on disjoint intervals.

We need to show that the behaviour of maxima in disjoint intervals is approximately independent. For dynamical systems, condition $D_r(u_n^{(k)})$ as used in [LLR] is not readily verifiable, and hence we propose condition $\mathcal{D}_r(u_n^{(k)})$ instead, inspired by [FFT1]. As mentioned in Section 4, such a condition can be easily checked for dynamical systems once information on decay of correlations is known. However we must show that this condition leads to the same conclusion in order to apply the thinning constructions of [LLR], which is the conclusion of the following result.

Proposition 7.1. *Suppose that $p \in \mathbb{N}$ and $A = \cap_{j=1}^p I_j$, where $I_j = [a_j, b_j)$. Let $x_k > 0$ for $k = 1, \ldots, p$ and for each $1 \le k \le p$, $(u_n^{(k)})_n$ be such that $n\mu(X_0 > u_n^{(k)}) \to x_k$. Assume that conditions $\mathcal{D}_p'(u_n^{(k)})$ and $\mathcal{D}_p(u_n^{(k)})$ hold. Then*

$$\mu\left(\bigcap_{j=1}^p \{M(nI_j) \le u_n^{(j)}\}\right) \to \prod_{j=1}^p e^{-x_j(b_j - a_j)}.$$

The proof of this proposition is technical and extends the ideas presented in [FFT1, Section 4]. We postpone the proof until Section 8.

7.2. Thinning constructions for point processes.

As a first step we consider the notion of an independent thinning of a Poisson point process ξ, see [LLR, Section 5.5]. We say that a process $\hat{\xi}$ is an independent thinning of ξ with parameter $p \in (0, 1)$, if for every point of ξ, there is a probability p that this point is retained in $\hat{\xi}$. An elementary argument shows that if ξ has intensity 1, then $\hat{\xi}$ has intensity p.

Suppose now that the sequence $x_1 < \cdots < x_r$ is defined, and $u_n^{(k)}$ is such that $n\mu\{X_1 > u_n^{(k)}\} \to x_k$. Then $u_n^{(1)} > \cdots > u_n^{(r)}$. Fix horizontal lines $L_1, \ldots, L_r$ in the plane, and define $\delta_{j/n}^{(k)}$ to be the Dirac mass concentrated at the point on L_k whose horizontal coordinate is j/n. The actual position of each L_k is not important provided their (vertical) order is preserved as described below. Given $k \le r$, we let $\xi_n^{(k)}$ denote the point process

$$\xi_n^{(k)} = \sum_{j=1}^n \delta_{j/n}^{(k)} \cdot 1_{\{X_j > u_n^{(k)}\}},$$

and let $\tilde{\xi}_n = \sum_{k=1}^r \xi_n^{(k)}$. Notice that for each k, if a $\tilde{\xi}_n^{(k)}$ has a point at location j/n, then each $\xi_n^{(l)}$ has a points at j/n too (for all $l \ge k$). This follows by the ordering of $u_n^{(k)}$.

We next define the point process $\tilde{\xi} = \sum_{k=1}^r \xi^{(k)}$, where $\{\xi^{(k)}\}$ is a sequence of (independently) thinned Poisson point processes, each having points on respective lines L_k, with $L_1 > L_2 > \cdots > L_r$ (ordered in vertical height), and with corresponding intensity parameter x_k. In particular, for each k, $\xi^{(k)}$ is an independent thinning of $\xi^{(k+1)}$ with probability parameter $p = x_k/x_{k+1}$. We have the following result:

Proposition 7.2. *Suppose that $(\mathcal{X}, f, \mu)$ is a measure preserving system. Suppose for any $r \ge 1$, and any sequence $\{x_i, i \le r\}$ condition $\mathcal{D}_r(u_n^{(k)})$ holds together with $\mathcal{D}_r'(u_n^{(k)})$. Then the point process $\tilde{\xi}_n$ defined on lines $\{L_k\}_{k=1}^r$ (as described above) converges weakly to the point process $\tilde{\xi}$ on $(0, 1] \times \mathbb{R}$.*

Remark 7.3. *For ease of exposition we prove convergence on $(0, 1] \times \mathbb{R}$. The method of proof extends to versions converging on state space $(0, \infty) \times \mathbb{R}$. In this case we require the $\mathcal{D}_r(u_n^{(k)})$ and $\mathcal{D}_r'(u_n^{(k)})$ conditions to hold for sequences of the form $\{u_n^{(k)}(mx_k)\}$, (for all $m \geq 1$). Here $\{u_n^{(k)}(x_k)\}$ is the sequence defined in (8).*

Proof of Proposition 7.2. Using the criteria of Kallenberg it is sufficient to check the following:

  (a) $\mathbb{E}(\tilde{\xi}_n(B)) \to \mathbb{E}(\tilde{\xi}(B))$, where B is of the form $(a, b] \times (c, d]$, for $c < d$, $0 < a < b$, and $b \leq 1$.
  (b) $\mu\{\tilde{\xi}_n = 0\} \to \mu\{\tilde{\xi} = 0\}$ for all $B = \cup_j(A_j \times C_j)$, where A_j, C_j are semi-closed intervals as described in (a) above, and with $A_j \times C_j$ disjoint.

Checking these conditions follows [LLR, Section 5.5] as applied to dependent processes. In particular checking part (a) is straightforward and is done as follows. We suppose that $B := (a, b] \times (c, d]$ intersects lines $L_s, \ldots, L_t$ for $1 \leq s \leq t \leq r$. Then $\tilde{\xi}_n(B) = \sum_{k=s}^t \xi_n^{(k)}((a, b])$ and $\tilde{\xi}(B) = \sum_{k=s}^t \xi^{(k)}((a, b])$. The expectation of the latter is $(b - a)\sum_{k=t}^s x_k$. Taking expectations of the former we obtain:

$$\mathbb{E}(\tilde{\xi}_n(B)) = \sum_{k=1}^r E\left(\sum_{j=1}^n \delta_{j/n}^{(k)}(B) \cdot 1_{\{X_j > u_n^{(k)}\}}\right) = ([nb] - [na])\sum_{k=s}^t \mu\{X_1 > u_n^{(k)}\}$$

$$\sim n(b - a)\sum_{k=s}^t \frac{x_k}{n} \to \mathbb{E}(\tilde{\xi}(B)).$$

We now check part (b) where Proposition 7.1 is used. Write $B = \cup_j(A_j \times C_j)$, with $A_j = (a_j, b_j]$, and $C_j = (c_j, d_j]$. By a rearrangement, we can express B as $\cup_j(a_j, b_j] \times D_j$, with $(a_j, b_j]$ disjoint, and D_j a finite union of semi-closed intervals. Hence

$$\{\tilde{\xi}_n(B) = 0\} = \bigcap_j\{\tilde{\xi}_n(E_j) = 0\},$$

where $E_j = (a_j, b_j] \times D_j$. For each j, we take L_{k_j} to be the lowest line intersecting E_j. By the thinning construction of $\tilde{\xi}_n$,

$$\{\tilde{\xi}_n(E_j) = 0\} = \{\xi_n^{(k_j)}((a_j, b_j]) = 0\}.$$

This corresponds to the set $\{M(([a_jn], [b_jn])) \leq u_{k_j}\}$. We can now immediately apply Proposition 7.1 to conclude that

$$\mu\{\tilde{\xi}_n(B) = 0\} \to \exp\left\{-\sum_j(b_j - a_j)x_{k_j}\right\} = \mu\{\tilde{\xi}(B) = 0\}.$$

$\square$

7.3. Concluding the proof of Theorem 5.1.

Given Proposition 7.2, we now show that the process $\xi_n = \sum_{i \geq 1} \delta_{z(i,n)}$ converges to a Poisson process on the plane with intensity measure $\lambda = \text{Leb} \times \lambda_G$, with $\lambda_G[a, b] = \log G(b) - \log G(a)$. The argument is purely probabilistic and follows [LLR, Section 5.7]. We give the main steps. It is convenient to work with the process $\hat{\xi}_n = \sum_{i \geq 1} \delta_{w(i,n)}$ with $w(i, n) = \left(\frac{i}{n}, u_n^{-1}(X_i)\right)$. Recall that the function $u_n(x)$ is determined via the limit relation $n\mu\{X_1 > u_n(x)\} \to x$ (from (8)), and hence if (10) applies then $w(i, n) = \left(\frac{i}{n}, -\log G(a_n(X_i - b_n))\right)$. We show that $\hat{\xi}_n$ converges to a Poisson process $\hat{\xi}$ in state space $(0, \infty) \times (0, \infty)$, with Lebesgue as the intensity measure. A simple change of measure argument then shows that ξ is a PRM with the corresponding intensity measure λ.

11

Continuing with the proof, and by Kallenberg's criteria it is sufficient to check items (a), and (b) as specified in the proof of Proposition 7.2 (for a suitable collection of disjoint semi-closed rectangles). If $B = (a, b] \times [c, d)$, then

$$\mathbb{E}(\hat{\xi}_n(B)) = E\left(\sum_{j \geq 1} \delta_{w(j,n)}(B)\right) = ([nb] - [na])\mu\{c \leq u_n^{-1}(X_1) < d\}$$

$$\sim n(b-a)\mu\{u_n(c) \leq X_1 \leq u_n(d)\} \sim (b-a)(d-c) = \mathbb{E}(\hat{\xi}(B)).$$

To show (b), we consider the event $\{\hat{\xi}_n(B) = 0\}$, where (as before) $B = \cup_j(A_j \times C_j)$, with $A_j = (a_j, b_j]$, $C_j = [c_j, d_j)$, and $A_j \times C_j$ forming a disjoint collection. By a rearrangement, we can express B as $\cup_j(a_j, b_j] \times D_j$, with $(a_j, d_j]$ disjoint, and D_j a finite union of semi-closed intervals. Indeed we can also make the simplifying assumption that $B = \cup_j (A \times D_j)$ for $A = (a, b]$, since the proof essentially is the same otherwise. So writing $B = \bigcup_{j=1}^m (A \times D_j)$, with $D_j = [x_{2j-1}, x_{2j}]$, for all $j \leq m$, and some sequence $x_1 < \cdots < x_{2m}$,

$$\{\hat{\xi}_n(B) = 0\} = \bigcap_{j=1}^m \left\{\xi_n^{(2j-1)}(A) = \xi_n^{(2j)}(A)\right\}, \tag{14}$$

where $\xi_n^{(j)}(A) = \sharp\{\ell \leq n : \ell/n \in A, u_n(x_j) < X_\ell\}$, i.e. $\xi^{(j)}$ counts the number of times of an exceedance $\ell \leqslant n$ of $u_n(x_j)$ where $\ell/n \in A$. The decomposition in (14) corresponds to the fact that we cannot have any X_ℓ with $u_n(x_{2j}) < X_\ell < u_n(x_{2j-1})$, and therefore an exceedance of $u_n(x_{2j})$ implies an exceedance of $u_n(x_{2j-1})$.

The processes $\xi_n^{(j)}$ meet the criteria of the thinning processes used in Proposition 7.2, and so:

$$(\xi_n^{(1)}(A), \ldots, \xi_n^{(2m)}(A)) \to (\xi^{(1)}(A), \ldots, \xi^{(2m)}(A)),$$

where convergence is in distribution. The distributions $\xi^{(j)}$ $(1 \leq j \leq 2m)$ correspond to the successively thinned Poisson processes defined in Proposition 7.2, and we obtain

$$\mu\{\hat{\xi}_n(B) = 0\} \to \mu\left(\bigcap_{j=1}^m \{\xi^{(2j-1)}(A) = \xi^{2j}(A)\}\right). \tag{15}$$

To compute the right hand side of (15), we use the thinning properties of the $\xi^{(j)}$. Note first of all that

$$\mu\left(\xi^{(2m-1)}(A) = \xi^{(2m)}(A)\right) = \sum_{k=1}^\infty \frac{(x_{2m}(b-a))^k e^{-x_{2m}(b-a)}}{k!}\left(\frac{x_{2m-1}}{x_{2m}}\right)^k$$

$$= e^{-(x_{2m}-x_{2m-1})(b-a)},$$

and by successive thinning we obtain for $j_{m-1} \leq j_m$:

$$\mu\left\{\left(\xi^{(2m-1)}(A) = \xi^{(2m)}(A) = j_m\right) \cap \left(\xi^{(2m-3)}(A) = \xi^{(2m-3)}(A) = j_{m-1}\right)\right\} =$$

$$\frac{(x_{2m}(b-a))^{j_m} e^{-x_{2m}(b-a)}}{j_m!}\binom{j_m}{j_{m-1}}\left(\frac{x_{2m-3}}{x_{2m-1}}\right)^{j_{m-1}}\left(1 - \frac{x_{2m-2}}{x_{2m-1}}\right)^{j_m-j_{m-1}}$$

Summing over all $0 \leq j_{m-1} \leq j_m < \infty$, we obtain

$$\mu\left\{\left(\xi^{(2m-1)}(A) = \xi^{(2m)}(A)\right) \cap \xi^{(2m-3)}(A) = \xi^{(2m-3)}(A)\right\} =$$

$$\exp\{-(b-a)(x_{2m} - x_{2m-1} + x_{2m-2} - x_{2m-3})\}$$

We can clearly iterate this, and a formula for the general case is given by:

12

$$\mu\left\{\bigcap_{k=1}^{m}\left(\xi^{(2k-1)}(A)=\xi^{(2k)}(A)=j_k\right)\right\}=$$

$$\frac{(x_{2m-1}(b-a))^{j_m}e^{-x_{2m}(b-a)}}{k!}\prod_{k=2}^{m}\binom{j_k}{j_{k-1}}\left(\frac{x_{2k-3}}{x_{2k-1}}\right)^{j_{k-1}}\left(1-\frac{x_{2k-2}}{x_{2k-1}}\right)^{j_k-j_{k-1}}\quad(16)$$

The probability of the event in (15) is then obtained by summing together all probabilities in (16) with $0\leq j_1\leq j_2\leq\ldots\leq j_m$. By an iterative method we obtain

$$\mu\left(\bigcap_{j=1}^{m}\{\xi^{(2j-1)}(A)=\xi^{2j}(A)\}\right)=\exp\{-\sum_{j=1}^{m}(x_{2j}-x_{2j-1})(b-a)\}=\exp\{-\mathrm{Leb}(B)\}.$$

This completes the proof.

8. Proof of Proposition 7.1

We must first adapt [FFT1, Lemma 4.2]:

Lemma 8.1. (a) *Given sets $A_1,\ldots,A_w\subset[0,\infty)$ and $B_1\supset A_1$, let $\ell:=\#\{j\in\mathbb{N}:j\in B_1\setminus A_1\}$ and $(u_i)_{i=1}^{w}\subset\mathbb{R}$,*

$$\left|\mu\left(\bigcap_{i=1}^{w}\{M(A_i)\leqslant u_i\}\right)-\mu\left(\{M(B_1)\leqslant u_1\}\cap\bigcap_{i=2}^{w}\{M(A_i)\leqslant u_i\}\right)\right|$$
$$\leqslant\mu(M(A_1)\leqslant u_1)-\mu(M(B_1)\leqslant u_1)\leqslant\ell\mu(X>u_1),$$

(b) *For $w\in\mathbb{N}$, assuming $\min\{x:x\in A_i,i=1,\ldots,w\}\geqslant r+t$, for $A_0=[0,r+t)$, for $u'>0$,*

$$\left|\mu\left(\{M(A_0)\leqslant u'\}\cap\bigcap_{j=1}^{w}\{M(A_j)\leqslant u_j\}\right)-\mu\left(\bigcap_{j=1}^{w}\{M(A_j)\leqslant u_j\}\right)\right.$$
$$\left.+\sum_{i=0}^{r-1}\mu\left(\{X>u'\}\cap\bigcap_{j=1}^{w}\{M(A_j-i)\leqslant u_j\}\right)\right|$$
$$\leqslant 2r\sum_{i=1}^{r-1}\mu\left(\{X>u'\}\cap\{X_i>u'\}\right)+t\mu(X>u').$$

Proof. For the first part, the first equality is an elementary argument, while the final inequality follows from (4.1) in [FFT1, Lemma 4.2].

The second part follows as in (4.2) of [FFT1, Lemma 4.2], itself a minor adaptation of [FF, Lemma 3.2]. □

Proof of Proposition 7.1. We closely follow the proof of [FFT1, Proposition 1].

Let $h:=\inf_{j\in\{1,\ldots,p\}}\{b_j-a_j\}$ and $H:=\lceil\sup\{x:x\in A\}\rceil$. Take $k>2/h$ and n sufficiently large. Note this guarantees that if we partition $n[0,H]\cap\mathbb{Z}$ into blocks of length $r_n:=\lfloor n/k\rfloor$, $J_1=[Hn-r_n,Hn)$, $J_2=[Hn-2r_n,Hn-r_n),\ldots$, $J_{Hk}=[Hn-Hkr_n,n-(Hk-1)r_n)$, $J_{Hk+1}=[0,Hn-Hkr_n)$, then there is at least one of these blocks contained in nI_i. Let $S_\ell=S_\ell(k)$ be the number of blocks J_j contained in nI_ℓ minus 1, that is,

$$S_\ell:=\#\{j\in\{1,\ldots,Hk\}:J_j\subset nI_\ell\}-1.$$

13

So $S_\ell \geqslant 0 \; \forall \ell \in \{1, \ldots, p\}$. Set $i_\ell := \min\{j \in \{1, \ldots, k\} : J_j \subset nI_\ell\}$. Then $J_{i_\ell}, J_{i_\ell+1}, \ldots, J_{i_\ell+S_\ell} \subset nI_\ell$. Now, fix ℓ and for each $i \in \{i_{p-\ell+1}, \ldots, i_{p-\ell+1} + S_{p-\ell+1}\}$ let

$$B_i := \bigcup_{j=i_{p-\ell+1}}^{i} J_j, \quad J_i^* := [Hn - ir_n, Hn - (i-1)r_n - t_n) \;\; \text{and} \;\; J_i' := J_i - J_i^*$$

for $t_n = o(n)$ given in $\tilde{\mathcal{D}}_r(u_n)$. Note that $|J_i^*| = r_n - t_n$ and $|J_i'| = t_n$. See Figure 8 for more of an idea of the notation here.
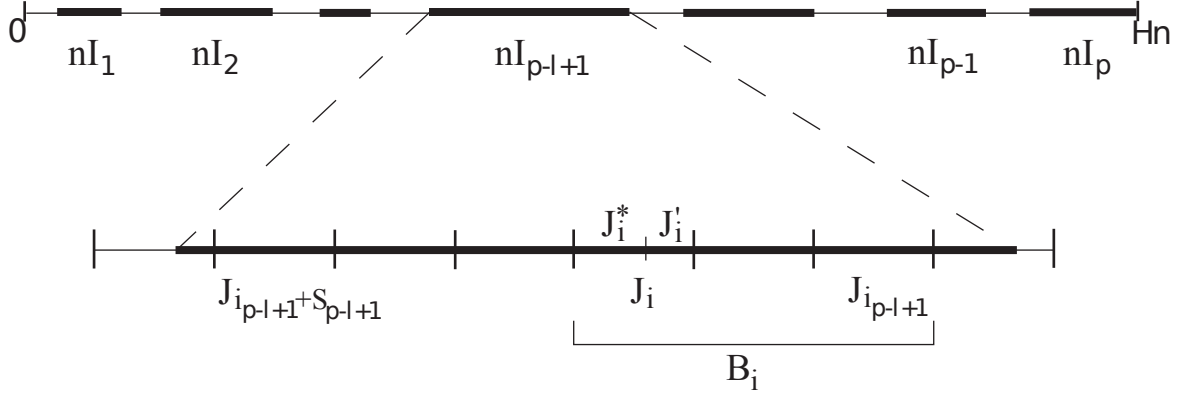

FIGURE 1. Notation

For the first part of the proof, we write $u_n^{(j)} = u_j$. Then for any $u' \in \mathbb{R}$

$$\left| \mu\left(\{M(B_i) \leqslant u'\} \cap \left(\bigcap_{i=1}^{\ell-1} \{M(nI_{p-i+1}) \leqslant u_{p-i+1}\} \right) \right) \right.$$

$$\left. - (1 - r_n\mu(X > u'))\mu\left(\{M(B_{i-1}) \leqslant u'\} \cap \left(\bigcap_{i=1}^{\ell-1} \{M(nI_{p-i+1}) \leqslant u_{p-i+1}\} \right) \right) \right|$$

$$\leqslant \left| \mu\left(\{M(B_i) \leqslant u'\} \cap \left(\bigcap_{i=1}^{\ell-1} \{M(nI_{p-i+1}) \leqslant u_{p-i+1}\} \right) \right) \right.$$

$$- \mu\left(\{M(B_{i-1}) \leqslant u'\} \cap \left(\bigcap_{i=1}^{\ell-1} \{M(nI_{p-i+1}) \leqslant u_{p-i+1}\} \right) \right)$$

$$\left. + r_n\mu(X > u')\mu\left(\{M(B_{i-1}) \leqslant u'\} \cap \left(\bigcap_{i=1}^{\ell-1} \{M(nI_{p-i+1}) \leqslant u_{p-i+1}\} \right) \right) \right|$$

$$\leqslant \left| \mu\left(\{M(B_i) \leqslant u'\} \cap \left(\bigcap_{i=1}^{\ell-1} \{M(nI_{p-i+1}) \leqslant u_{p-i+1}\} \right) \right) \right.$$

$$- \mu\left(\{M(B_{i-1}) \leqslant u'\} \cap \left(\bigcap_{i=1}^{\ell-1} \{M(nI_{p-i+1}) \leqslant u_{p-i+1}\} \right) \right)$$

$$\left. + \sum_{j=0}^{r_n-t_n-1} \mu\left(\{X_{j+Hn-ir_n} > u'\} \cap \{M(B_{i-1}) \leqslant u'\} \cap \left(\bigcap_{i=1}^{\ell-1} \{M(nI_{p-i+1}) \leqslant u_{p-i+1}\} \right) \right) \right|$$

$$+ \left| (r_n - t_n)\mu(X > u')\mu\left(\{M(B_{i-1}) \leqslant u'\} \cap \left(\bigcap_{i=1}^{\ell-1} \{M(nI_{p-i+1}) \leqslant u_{p-i+1}\} \right) \right) \right.$$

$$- \sum_{j=0}^{r_n - t_n - 1} \mu\left(\{X_{j + Hn - ir_n} > u'\} \cap \{M(B_{i-1}) \leqslant u'\} \cap \left(\bigcap_{i=1}^{\ell-1} \{M(nI_{p-i+1}) \leqslant u_{p-i+1}\} \right) \right) \Bigg|$$

$$+ t_n \mu(X > u').$$

We next apply Lemma 8.1(b) to the first sum in absolute value above. Since $B_i = J_i \cup B_{i-1} = J_i^* \cup J_i' \cup B_{i-1}$, we can translate the sets in the first two terms here back by $Hn - ir_n$ and with $J_i - (Hn - ir_n) = [0, r_n)$ taking the place of A_0 in our lemma, and correspondingly shifting the terms in the sum there by $d_j = (j + Hn - ir_n)$. Then (also shifting terms in the second absolute value by d_j and using the invariance of μ),

$$\left| \mu\left(\{M(B_i) \leqslant u'\} \cap \left(\bigcap_{i=1}^{\ell-1} \{M(nI_{p-i+1}) \leqslant u_{p-i+1}\} \right) \right) \right.$$

$$\left. - (1 - r_n \mu(X > u')) \mu\left(\{M(B_{i-1}) \leqslant u'\} \cap \left(\bigcap_{i=1}^{\ell-1} \{M(nI_{p-i+1}) \leqslant u_{p-i+1}\} \right) \right) \right|$$

$$\leqslant 2(r_n - t_n) \sum_{j=1}^{r_n - t_n - 1} \mu(\{X > u'\} \cap \{X_j > u'\}) + t_n \mu(X > u')$$

$$+ \sum_{j=0}^{r_n - t_n - 1} \left| \mu(X > u') \mu\left(\{M(B_{i-1}) \leqslant u'\} \cap \left(\bigcap_{i=1}^{\ell-1} \{M(nI_{p-i+1}) \leqslant u_{p-i+1}\} \right) \right) \right.$$

$$\left. - \mu\left(\{X > u'\} \cap \{M(B_{i-1} - d_j) \leqslant u'\} \cap \left(\bigcap_{i=1}^{\ell-1} \{M(nI_{p-i+1} - d_j) \leqslant u_{p-i+1}\} \right) \right) \right|$$

$$+ t_n \mu(X > u').$$

Now using condition $\tilde{\mathcal{D}}_r(u_n)$, we obtain

$$\left| \mu\left(\{M(B_i) \leqslant u'\} \cap \left(\bigcap_{i=1}^{\ell-1} \{M(nI_{p-i+1}) \leqslant u_{p-i+1}\} \right) \right) \right.$$

$$\left. - (1 - r_n \mu(X > u')) \mu\left(\{M(B_{i-1}) \leqslant u'\} \cap \left(\bigcap_{i=1}^{\ell-1} \{M(nI_{p-i+1}) \leqslant u_{p-i+1}\} \right) \right) \right|$$

$$\leq 2(r_n - t_n) \sum_{j=1}^{r_n - t_n - 1} \mu(\{X > u'\} \cap \{X_j > u'\}) + 2 t_n \mu(X > u') + (r_n - t_n) \gamma(n, t_n).$$

Set

$$\Upsilon_{k,n}(u') := 2(r_n - t_n) \sum_{j=1}^{r_n - t_n - 1} \mu(\{X > u'\} \cap \{X_j > u'\}) + 2 t_n \mu(X > u') + (r_n - t_n) \gamma(n, t_n).$$

By the definition of $u_j = u_{n,j}$, we may assume that n and k are sufficiently large so that $\frac{n}{k} \mu(X > u_j) < 2$ and $|1 - r_n \mu(X > u_j)| < 1$ which implies

$$\left| \mu\left(\{M(B_{S_{p-\ell+1}}) \leqslant u_{p-\ell+1}\} \cap \left(\bigcap_{i=1}^{\ell-1} \{M(nI_{p-i+1}) \leqslant u_{p-i+1}\} \right) \right) \right.$$

15

$$- (1 - r_n\mu(X > u_{p-\ell+1}))\mu\left(\{M(B_{S_{p-\ell+1}-1}) \leqslant u_{p-\ell+1}\} \cap \left(\bigcap_{i=1}^{\ell-1}\{M(nI_{p-i+1}) \leqslant u_{p-i+1}\}\right)\right)\Bigg|$$

$$\leqslant \Upsilon_{k,n}(u_{p-\ell+1}),$$

and

$$\left|\mu\left(\{M(B_{S_{p-\ell+1}}) \leqslant u_{p-\ell+1}\} \cap \left(\bigcap_{i=1}^{\ell-1}\{M(nI_{p-i+1}) \leqslant u_{p-i+1}\}\right)\right)\right.$$

$$\left. - (1 - r_n\mu(X > u_{p-\ell+1}))^2\mu\left(\{M(B_{S_{p-\ell+1}-2}) \leqslant u_{p-\ell+1}\} \cap \left(\bigcap_{i=1}^{\ell-1}\{M(nI_{p-i+1}) \leqslant u_{p-i+1}\}\right)\right)\right|$$

$$\leqslant \left|\mu\left(\{M(B_{S_{p-\ell+1}}) \leqslant u_{p-\ell+1}\} \cap \left(\bigcap_{i=1}^{\ell-1}\{M(nI_{p-i+1}) \leqslant u_{p-i+1}\}\right)\right)\right.$$

$$\left. - (1 - r_n\mu(X > u_{p-\ell+1}))\mu\left(\{M(B_{S_{p-\ell+1}-1}) \leqslant u_{p-\ell+1}\} \cap \left(\bigcap_{i=1}^{\ell-1}\{M(nI_{p-i+1}) \leqslant u_{p-i+1}\}\right)\right)\right|$$

$$+ |1 - r_n\mu(X > u_{p-\ell+1})|\left|\mu\left(\{M(B_{S_{p-\ell+1}-1}) \leqslant u_{p-\ell+1}\} \cap \left(\bigcap_{i=1}^{\ell-1}\{M(nI_{p-i+1}) \leqslant u_{p-i+1}\}\right)\right)\right.$$

$$\left. - (1 - r_n\mu(X > u_{p-\ell+1}))\mu\left(\{M(B_{S_{p-\ell+1}-2}) \leqslant u_{p-\ell+1}\} \cap \left(\bigcap_{i=1}^{\ell-1}\{M(nI_{p-i+1}) \leqslant u_{p-i+1}\}\right)\right)\right|$$

$$\leqslant 2\Upsilon_{k,n}(u_{p-\ell+1}),$$

Inductively, we obtain

$$\left|\mu\left(\{M(B_{S_{p-\ell+1}}) \leqslant u_{p-\ell+1}\} \cap \left(\bigcap_{i=1}^{\ell-1}\{M(nI_{p-i+1}) \leqslant u_{p-i+1}\}\right)\right)\right.$$

$$\left. - (1 - r_n\mu(X > u_{p-\ell+1}))^{S_{p-\ell+1}}\mu\left(\bigcap_{i=1}^{\ell-1}\{M(nI_{p-i+1}) \leqslant u_{p-i+1}\}\right)\right|$$

$$\leqslant S_{p-\ell+1}\Upsilon_{k,n}(u_{p-\ell+1}).$$

Using Lemma 8.1(a),

$$\left|\mu\left(\bigcap_{i=1}^{\ell}\{M(nI_{p-i+1}) \leqslant u_{p-i+1}\}\right)\right.$$

$$\left. - (1 - r_n\mu(X > u_{p-\ell+1}))^{S_{p-\ell+1}}\mu\left(\bigcap_{i=1}^{\ell-1}\{M(nI_{p-i+1}) \leqslant u_{p-i+1}\}\right)\right|$$

$$\leqslant \left|\mu\left(\bigcap_{i=1}^{\ell}\{M(nI_{p-i+1}) \leqslant u_{p-i+1}\}\right)\right.$$

$$\left. - \mu\left(\{M(B_{S_{p-\ell+1}}) \leqslant u_{p-\ell+1}\} \cap \left(\bigcap_{i=1}^{\ell-1}\{M(nI_{p-i+1}) \leqslant u_{p-i+1}\}\right)\right)\right|$$

$$+ \left|\mu\left(\{M(B_{S_{p-\ell+1}}) \leqslant u_{p-\ell+1}\} \cap \left(\bigcap_{i=1}^{\ell-1}\{M(nI_{p-i+1}) \leqslant u_{p-i+1}\}\right)\right)\right.$$

$$-\left(1 - r_n\mu(X > u_{p-\ell+1})\right)^{S_{p-\ell+1}}\mu\left(\bigcap_{i=1}^{\ell-1}\{M(nI_{p-i+1}) \leqslant u_{p-i+1}\}\right)\Bigg|$$

$$\leqslant \left|\mu\left(\bigcap_{i=1}^{\ell}\{M(nI_{p-i+1}) \leqslant u_{p-i+1}\}\right)\right.$$

$$-\mu\left(\left\{M\left(\cup_{i=i_\ell}^{S_{p-\ell+1}}J_i\right) \leqslant u_{p-\ell+1}\right\} \cap \left(\bigcap_{i=1}^{\ell-1}\{M(nI_{p-i+1}) \leqslant u_{p-i+1}\}\right)\right)\Bigg|$$

$$+ S_{p-\ell+1}\Upsilon_{k,n}(u_{p-\ell+1})$$

$$\leqslant 2r_n\mu(X > u_{p-\ell+1}) + S_{p-\ell+1}\Upsilon_{k,n}(u_{p-\ell+1}).$$

For the next step we estimate

$$\left|\mu\left(\bigcap_{i=1}^{\ell}\{M(nI_{p-i+1}) \leqslant u_{p-i+1}\}\right)\right.$$

$$-\left(1 - r_n\mu(X > u_{p-\ell+1})\right)^{S_{p-\ell+1}}\left(1 - r_n\mu(X > u_{p-\ell+2})\right)^{S_{p-\ell+2}}\mu\left(\bigcap_{i=1}^{\ell-2}\{M(nI_{p-i+1}) \leqslant u_{p-i+1}\}\right)\Bigg|$$

$$\leqslant \left|\mu\left(\bigcap_{i=1}^{\ell}\{M(nI_{p-i+1}) \leqslant u_{p-i+1}\}\right)\right.$$

$$-\left(1 - r_n\mu(X > u_{p-\ell+1})\right)^{S_{p-\ell+1}}\mu\left(\bigcap_{i=1}^{\ell-1}\{M(nI_{p-i+1}) \leqslant u_{p-i+1}\}\right)\Bigg|$$

$$+ |1 - r_n\mu(X > u_{p-\ell+2})|^{S_{p-\ell+2}}\left|\mu\left(\bigcap_{i=1}^{\ell-1}\{M(nI_{p-i+1}) \leqslant u_{p-i+1}\}\right)\right.$$

$$-\left(1 - r_n\mu(X > u_{p-\ell+1})\right)^{S_{p-\ell+2}}\mu\left(\bigcap_{i=1}^{\ell-2}\{M(nI_{p-i+1}) \leqslant u_{p-i+1}\}\right)\Bigg|$$

$$\leqslant 2r_n\left(\mu(X > u_{p-\ell+1}) + \mu(X > u_{p-\ell+2})\right) + S_{p-\ell+1}\Upsilon_{k,n}(u_{p-\ell+1}) + S_{p-\ell+2}\Upsilon_{k,n}(u_{p-\ell+2})$$

Therefore, by induction, we obtain

$$\left|\mu\left(\bigcap_{i=1}^{\ell}\{M(nI_{p-i+1}) \leqslant u_{p-i+1}\}\right) - \prod_{j=1}^{p}\left(1 - r_n\mu(X > u_j)\right)^{S_j}\right|$$

$$\leqslant 2r_n\sum_{j=1}^{p}\mu(X > u_j) + \sum_{j=1}^{p}S_j\Upsilon_{k,n}(u_j).$$

Now, it is easy to see that $S_j \sim k|I_j|$, for each $j \in \{1, \ldots, p\}$. Consequently, recalling $u_{n,j} = u_j$,

$$\lim_{k\to+\infty}\lim_{n\to+\infty}\prod_{j=1}^{p}\left(1 - r_n\mu(X > u_j)\right)^{S_j} = \lim_{k\to+\infty}\lim_{n\to+\infty}\prod_{j=1}^{p}\left(1 - \left\lfloor\frac{n}{k}\right\rfloor\mu(X > u_j)\right)^{S_j}$$

$$= \lim_{k\to+\infty}\lim_{n\to+\infty}\prod_{j=1}^{p}\left(1 - \left\lfloor\frac{n}{k}\right\rfloor\mu(X > u_j)\right)^{k|I_j|} = \lim_{k\to+\infty}\lim_{n\to+\infty}\prod_{j=1}^{p}\left(1 - \frac{x_j}{k}\right)^{k|I_j|} = \prod_{j=1}^{p}e^{-x_j|I_j|}$$

To conclude the proof it suffices to show that

$$\lim_{k\to+\infty}\lim_{n\to+\infty}\left(2r_n\sum_{j=1}^{p}\mu(X>u_{n,j})+\sum_{j=1}^{p}S_j\Upsilon_{k,n}(u_{n,j})\right)=0.$$

We start by noting that, since $n\mu(X>u_{n,j})\to x_j\geq 0$,

$$\lim_{k\to+\infty}\lim_{n\to+\infty}2r_n\mu(X>u_{n,j})=\lim_{k\to+\infty}\frac{2x_j}{k}=0.$$

Next we need to check that for each $j=1,\ldots,p$

$$\lim_{k\to+\infty}\lim_{n\to+\infty}2k(r_n-t_n)\sum_{j=1}^{r_n-t_n-1}\mu(\{X>u_{n,j}\}\cap\{X_j>u_{n,j}\})+2t_n\mu(X>u_{n,j})$$
$$+k(r_n-t_n)\gamma(n,t_n)=0.$$

Recall that $t_n=o(n)$ is given by $\mathcal{D}_p(u_n^{(k)})$. Now, observe that for each $j=1,\ldots,p$ and every $k\in\mathbb{N}$, we have $\lim_{n\to\infty}kt_n\mu(X>u_n^{(j)})=0$. Finally, we use $\mathcal{D}_p(u_n^{(k)})$ and $\mathcal{D}'_p(u_n^{(k)})$ to prove that the two remaining terms also go to 0. $\qquad\square$

References

[CC] J.-R. Chazottes and P. Collet. Poisson approximation for the number of visits to balls in non-uniformly hyperbolic dynamical systems. *Ergodic Theory Dynamical Systems.* **33** (2013), 49-80.

[C] P. Collet. Statistics of closest return for some non-uniformly hyperbolic systems. *Ergodic Theory Dynamical Systems,* **21** (2001), 401-420.

[EKM] P. Embrechts, C. Klüpperlberg, and T. Mikosch. *Modelling extremal events.* **33**, Springer, 1997.

[FF] A. C. M. Freitas and J. M. Freitas. On the link between dependence and independence in extreme value theory for dynamical systems. *Stat. Probab. Lett.,* **78**, (2008), 1088-1093.

[FFT1] J. Freitas, A. Freitas and M. Todd. Hitting times and extreme value theory. *Probab. Theory Related Fields,* **147(3)** (2010), 675–710

[FFT2] A.C.M. Freitas, J.M. Freitas, M. Todd, Extreme value laws in dynamical systems for non-smooth observations. *J. Stat. Phys.,* **42**, (1), (2011) 108–126.

[FFT3] A.C.M. Freitas, J.M. Freitas, M. Todd, Extremal index, hitting time statistics and periodicity. *Adv. Math.,* **231**, no. 5, (2012), 2626–2665.

[G] N. Glick. Breaking records and breaking boards. *Amer. Math. Monthly,* **85**, (1978), 2–26.

[GHN] C. Gupta, M. P. Holland and M. Nicol. Extreme value theory for dispersing billiards, Lozi maps and Lorenz maps. *Ergodic Theory and Dynamical Systems,* **31**, (5), (2011), 1363–1390.

[Gu] C. Gupta. Extreme-value distributions for some classes of non-uniformly partially hyperbolic dynamical systems. *Ergodic Theory Dynam. Systems,* **30** (2010), 757–771.

[HNT1] M. P. Holland and M. Nicol and A. Török. Extreme value distributions for non-uniformly expanding dynamical systems. *Trans. Amer. Math. Soc.,* **364**, (2012), 661–688.

[HNT2] M. P. Holland and M. Nicol and A. Török. Almost sure convergence of maxima for chaotic dynamical systems. *Stoch. Proc. and App.,* **126** (2016), 3145–3170.

[K] O. Kallenberg, *Random measures,* Academic Press Inc., New York, 1986.

[LLR] M.R. Leadbetter, G. Lindgren and H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes,* Springer-Verlag, (1980).

[R1] S. I. Resnick. Inverses of extremal processes. *Advances in Appplied Probability,* **6**, (1973), 392-406.

[R2] S. I. Resnick. Weak convergence to extremal processes. *Ann. Probab.,* **3**, (6), (1975), 951-960.

[R3] S. I. Resnick. *Extreme values, regular variation, and point processes*, Applied Probability Trust, **4**, Springer-Verlag, 1987.

[S] A. V. Skorokhod. Limit theorems for stochastic processes. *Theory Prob. Appl.*, **1**, (1957), 261-290.

[T] M. Tyran-Kamińska. Weak convergence to Levy stable processes in dynamical systems. *Stoch. Dyn.* **10**, (2010), 263-289.

[Y1] L.-S. Young. Statistical properties of dynamical systems with some hyperbolicity. *Ann. of Math.* **147**, (1998), 585–650.

[Y2] L.-S. Young. Recurrence times and rates of mixing. *Israel J. Math.* **110**, (1999), 153–188.

[Z] L. Zhang. Borel Cantelli lemmas and extreme value theory for geometric Lorenz models. *Nonlinearity*, **29**, (1), (2015), 232–256.

Mark Holland, Mathematics (CEMPS), Harrison Building (327), North Park Road, Exeter, EX4 4QF, UK

E-mail address: `M.P.Holland@exeter.ac.uk`

URL: `http://empslocal.ex.ac.uk/people/staff/mph204/`

Mike Todd, Mathematical Institute, University of St Andrews, North Haugh, St Andrews, KY16 9SS, Scotland

E-mail address: `m.todd@st-andrews.ac.uk`

URL: `http://www.mcs.st-and.ac.uk/~miket/`