Speed of convergence to an extreme value distribution for non-uniformly hyperbolic dynamical systems

M.P. Holland, M. Nicol

January 21, 2014

Abstract

Suppose $(f, \mathcal{X}, \nu)$ is a dynamical system and $\phi : \mathcal{X} \to \mathbb{R}$ is an observation with a unique maximum at a (generic) point in $\mathcal{X}$. We consider the time series of successive maxima $M_n(x) := \max\{\phi(x), \ldots, \phi \circ f^{n-1}(x)\}$. Recent works have focused on the distributional convergence of such maxima (under suitable normalization) to an extreme value distribution. In this article, for certain dynamical systems, we establish convergence rates to the limiting distribution. In contrast to the case of i.i.d random variables, the convergence rates depend on the rate of mixing and the recurrence time statistics. For a range of applications, including uniformly expanding maps, quadratic maps, and intermittent maps, we establish corresponding convergence rates. We also establish convergence rates for certain hyperbolic systems such Anosov systems, and discuss convergence rates for non-uniformly hyperbolic systems, such as Hénon maps.

Key Words: Ergodic Theory, Dynamical Systems, Extreme Statistics.

1 Introduction and background on extremes in dynamical systems

Consider a dynamical system $(f, \mathcal{X}, \nu)$, where $\mathcal{X} \subset \mathbb{R}$, $f : \mathcal{X} \to \mathcal{X}$ a measurable transformation, and ν is an f-invariant probability measure supported on $\mathcal{X}$. Given an observable $\phi : \mathcal{X} \to \mathbb{R}$ 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

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

Almost surely, $M_n \to \max \phi$. We consider instead the distributional behavior of M_n and in particular the existence of sequences $a_n, b_n \in \mathbb{R}$ such that

$$\nu \{x \in \mathcal{X} : a_n(M_n - b_n) \leq u\} \to G(u), \tag{3}$$

for some non-degenerate distribution function $G(u)$, $-\infty < u < \infty$. As in the i.i.d case, the sequences $u_n := u/a_n + b_n$ will be chosen so that

$$\lim_{n\to\infty} n\nu\{x \in \mathcal{X} : \phi(x) > u_n\} \to \tau(u), \tag{4}$$

for some non-degenerate function $\tau(u)$. For the stochastic process defined in (1), much recent work has focused on the computation of the function $G(u)$, and showing that it agrees, at least for sufficiently hyperbolic systems and for regular enough observations ϕ maximized at generic $\tilde{x}$, with that which would hold if $\{X_i\}$ were independent identically distributed (i.i.d.) random variables. If $\tilde{x}$ is periodic we expect different behavior (for details see [9, 4, 16, 20]). The case of i.i.d. random variables has been widely studied, see [11, 21, 25], and if the limit function $G(u)$ is a non-degenerate distribution function then the limit can only be of three following types:

Type I (Gumbel):

$$G(u) = \exp\left(-\exp\left[-\frac{u-b}{a}\right]\right), \quad -\infty < u < \infty;$$

Type II (Fréchet):

$$G(u) = \begin{cases} 0, & u \leq b, \\ \exp\left(-\left[\dfrac{u-b}{a}\right]^{-\alpha}\right), & u > b; \end{cases}$$

Type III (Weibull):

$$G(u) = \begin{cases} \exp\left(-\left[-\dfrac{u-b}{a}\right]^{\alpha}\right), & u < b; \\ 1, & u \geq b, \end{cases}$$

for some parameters $a > 0$, b and $\alpha > 0$. The functional form of $G(u)$ in fact depends on $\tau(u)$, see [21]. For example, in the case of i.i.d. random variables defined by the unit exponential probability distribution P, we have that $\tau(u) = e^{-u}$, and $P(M_n \leq u + \log n) \to \exp(-e^{-u})$. Type II/III distributions arise in the case where $\tau(u)$ has power law behaviour. Given a cumulative probability distribution G, we say that G follows an *Extreme Value Distribution* (EVD) if G is any of the three distributions above.

In this article we focus on the speed of convergence of $\nu\{M_n \leq u_n\}$ to $G(u)$. In the case of i.i.d random variables $\hat{X}_i$, with probability distribution function $F(u) := P(\hat{X}_i \leq u)$ the rate of convergence depends on the normalization sequences a_n, b_n, and the functional form of $F(u) := P(\hat{X}_i \leq u)$ as $u \to u_F := \max \hat{X}_i$. In the theory of domains of attraction [21], necessary and sufficient conditions for convergence to, for example, Type I are that the distribution function F satisfies:

$$\lim_{u \to u_F} \frac{1 - F(u + \ell\gamma(u))}{1 - F(u)} = e^{-\ell}, \quad \text{some } \gamma(u) > 0. \tag{5}$$

We will focus on this case primarily, with analogous (regular variation) conditions specified for convergence to Types II/III, see [21]. For uniform estimates on rates of convergence, we consider two examples, namely the exponential distribution and the Gaussian distribution. For the exponential distribution $F(u) = 1 - e^{-\lambda u}$, it is shown in [15] that:

$$\sup_u \left| P\left\{M_n \leq u + \frac{\log n}{\lambda}\right\} - e^{-e^{-u}} \right| = \sup_u \left| \left(1 - \frac{e^{-u}}{n}\right)^n - e^{-e^{-u}} \right| \leq \frac{1}{n}\left(1 + \frac{2}{n}\right) e^{-2}. \tag{6}$$

However, in general the convergence rate can be quite slow, and this is evident for the Gaussian distribution [14], where it is established that:

$$\frac{C_1}{\log n} \leq \sup_u \left| \{\Phi(u/a_n + b_n)\}^n - e^{-e^{-u}} \right| \leq \frac{C_2}{\log n}. \tag{7}$$

Here $\Phi(u)$ is the standard Gaussian distribution function, C_1, C_2 are uniform constants, and a_n, b_n satisfy:

$$a_n = b_n, \ 2\pi b_n^2 e^{b_n^2} = n^2.$$

In this case, the choice of constants is optimal for the error rate.

In the setting of dynamical systems, we consider the corresponding quantity $\nu\{x : \phi(x) < u\}$ (which corresponds to $F(u)$), and study the behaviour of this measure as $u \to \max\phi$. We will focus primarily on the specific observable $\phi(x) = -\log(\mathrm{dist}(x, \tilde{x}))$ for generic (in particular non-periodic) $\tilde{x} \in \mathcal{X}$. The analysis can be extended to study other observable types, such as $\phi(x) = \psi(\mathrm{dist}(x, \tilde{x}))$ where $\psi : \mathbb{R}^+ \to \mathbb{R}$ is regularly varying at 0. The extreme value statistics of observables maximized at periodic points is expected to be different to observables maximized at generic points, this has been proven rigorously for certain one-dimensional maps [9, 20, 16, 4].

We take the scaling sequence $u_n = u + \log n$ (as we would in the i.i.d case), and study the errors involved in approximating the Gumbel distribution by $\nu\{M_n \le u_n\}$. For a given sequence u_n, we define functions $\tau_n(u)$ and $G_n(u)$ by

$$\tau_n(u) = n\nu\{\phi(x) \ge u_n\}, \quad G_n(u) = \left(1 - \frac{\tau_n(u)}{n}\right)^n.$$

The function $\tau_n(u) \to \tau(u)$ uniformly for all u lying in a compact subset of $\mathbb{R}$.

For weakly dependent stochastic processes satisfying equation (4), it was shown in [21] that convergence to a EVD is still valid (with the same distribution type as in the i.i.d. case) provided two probabilistic conditions $D(u_n)$ and $D'(u_n)$ are shown to hold. In the dynamical systems setting, much (recent) work has been devoted to finding conditions analogous to $D(u_n)$ and $D'(u_n)$ that ensure that (2) converges to a EVD. To establish convergence rates, namely to estimate $|\nu\{M_n \le u_n\} - G(u)|$ (as $n \to \infty$, for u lying in compact subsets of $\mathbb{R}$), we cannot immediately express $\nu\{M_n \le u_n\}$ in terms of $G_n(u)$. However, if we appeal directly to the blocking arguments used in [3, 17] (which form the basis of checking conditions $D'(u_n)$,$D(u_n)$), we show that for certain non-uniformly expanding systems $\nu\{M_n \le u_n\}$ can be approximated by $G_{n^\beta}(u)$ (for some $\beta \in (0, 1]$), up to an error of order $O(n^{-1/2}(\log n)^{1+\epsilon})$, for all $\epsilon > 0$. Here

$$G_{n^\beta}(u) = \left(1 - \frac{\tau_n(u)}{n^\beta}\right)^{n^\beta},$$

and the optimal choice of β turns out to be 1/2 for most applications. The rate of convergence of $G_{n^\beta}(u)$ to $G(u)$ is a straightforward estimate. We recently learnt of the recent work [10] on rates of convergence, with somewhat related techniques but different applications.

Throughout we fix the following notations. For general sequences x_n, y_n, we say that $x_n \sim y_n$ if $x_n/y_n \to 1$ as $n \to \infty$. We say $x_n \approx y_n$ if there are real constants c_1, c_2 such that $c_1 \le x_n/y_n \le c_c$. For positive sequences we write $x_n = \mathcal{O}(y_n)$ if there is a constant $C > 0$ such that $x_n \le Cy_n$, and we write $x_n = o(y_n)$ if $x_n/y_n \to 0$.

2 Statement of the main results

We take $(f, \mathcal{X}, \nu)$ to be an ergodic dynamical system.

Let $\tilde{g} : \mathbb{N} \to \mathbb{R}$ be a monotonically increasing function and let E_n be a sequence of sets defined by:

$$E_n := \left\{x \in \mathcal{X} : \mathrm{dist}(x, f^j(x)) \le \frac{1}{n}, \text{ for some } j \in [1, \tilde{g}(n)]\right\}. \tag{8}$$

In [18], the sets E_n are referred to as *recurrence sets*. If the time scale $\tilde{g}(n) = o(n)$, these sets contain points which have what we call 'fast' returns. We will see that the distributional rate of convergence to an EVD is partly determined by the rate at which $\nu(E_n)$ converges to zero.

Assumptions on the invariant measure ν.

We will assume that the measure ν is absolutely continuous with respect to Lebesgue m. Some of our results require also that ν admits a density $\rho \in L^{1+\delta}(m)$ for some $\delta > 0$. These assumptions are mild and are satisfied by the non-uniformly expanding one-dimensional maps which form the bulk of our applications. For non-uniformly hyperbolic systems in dimension at least two, we require further assumptions on the regularity of ν. We will discuss this further in Section 4.

Dynamical assumptions on $(f, \mathcal{X}, \nu)$.

We make the following assumptions. The function $\tilde{g}(n)$ and the sets E_n are from equation (8).

(H1) **(Decay of correlations).** There exists a monotonically decreasing sequence $\Theta(j) \to 0$ such that for all φ_1 Lipschitz continuous and all $\varphi_2 \in L^\infty$:

$$\left| \int \varphi_1 \cdot \varphi_2 \circ f^j d\nu - \int \varphi_1 d\nu \int \varphi_2 d\nu \right| \leq \Theta(j) \|\phi_1\|_{\text{Lip}} \|\phi_2\|_{L^\infty},$$

where $\|\cdot\|_{\text{Lip}}$ denotes the Lipschitz norm.

(H2a) **(Strong quantitative recurrence rates).** There exist numbers $\gamma, \alpha > 0$ such that

$$\tilde{g}(n) \sim n^\gamma \quad \implies \quad \nu(E_n) \leq \frac{C}{n^\alpha}. \tag{9}$$

Condition (H1) is an assumption on the rate of mixing in a suitable Banach space of functions. For our purposes, it is sufficient to work with the Lipschitz class of functions. In the statement of the results will make precise asymptotic statements about the rate of decay of $\Theta(j)$.

Condition (H2a) is a quantitative control on the recurrence statistics. For general non-uniformly expanding systems, checking this condition requires careful analysis. See for example [12, 17]. In Section 3 we show how to check this condition for uniformly expanding Markov maps and certain Markov intermittency maps. For systems having sub-exponential decay of correlations we specifically require (H2a) to hold in order to derive error rates. It is conjectured in [18] that (H2a) holds for a broad class of non-uniformly hyperbolic systems, and this is observed numerically (when analytic estimates are not available). For systems with exponential decay of correlations we can work with a weaker version of (H2a), which we label as (H2b). This is stated as follows:

(H2b) **(Weak quantitative recurrence rates).** For some $\gamma' > 1$, $\alpha > 0$:

$$\tilde{g}(n) \sim (\log n)^{\gamma'} \quad \implies \quad \nu(E_n) \leq \frac{C}{n^\alpha}. \tag{10}$$

This latter condition only requires control of the recurrence up to a slow time scale $\tilde{g}(n) \sim (\log n)^{\gamma'}$ and is easier to check analytically relative to (H2a), see for example [3, 13, 18]. To obtain convergence to an EVD (and to find corresponding error rates), it is sufficient to check this condition provided the system has exponential decay of correlations. We will use this condition to calculate error rates for the quadratic family (in Section 3). To establish convergence rates for certain non-uniformly hyperbolic systems we will again assume condition (H2b), see Section 4.

We now determine the rate of convergence of $\nu\{M_n < u_n\}$ to $G(u)$ for a system $(f, \mathcal{X}, \nu)$ satisfying the assumptions above.

In the first of the results we state below, namely Theorem 2.1, we give a quantitative estimate on the difference between $\nu\{M_n < u_n\}$ and $G_{\sqrt{n}}(u)$, where we recall that

$$G_{\sqrt{n}}(u) = \left(1 - \frac{\tau_n(u)}{\sqrt{n}}\right)^{\sqrt{n}}, \quad \tau_n(u) = n\nu\left\{x : \phi(x) > \frac{u}{a_n} + b_n\right\}.$$

In the above, we could replace $\sqrt{n}$ by some other power n^β for $\beta \in (0, 1]$, but the choice $\beta = 1/2$ turns out to be optimal in most applications. Determining the rate of convergence of $G_{\sqrt{n}}(u)$ to $G(u)$ then becomes equivalent to the i.i.d. case, and the convergence rate depends specifically on the function form of ϕ and the regularity of ν in the vicinity of the point $\tilde{x} \in \mathcal{X}$ where ϕ is maximized. We will study the convergence rates for particular examples of ν and ϕ.

Theorem 2.1. *Suppose that $f : \mathcal{X} \to \mathcal{X}$ is a map with ergodic measure in $L^{1+\delta}(m)$ for some $\delta > 0$, and ν absolutely continuous with respect to m. We have the following cases.*

1. *Suppose that $\Theta(n) = \mathcal{O}(\theta_0^n)$ for some $\theta_0 < 1$ and (H1), (H2b) hold. Then for all $\epsilon > 0$ and ν-a.e. $\tilde{x} \in \mathcal{X}$ we have that*

$$\left|\nu\{M_n \leq u_n\} - G_{\sqrt{n}}(u)\right| \leq C\frac{(\log n)^{1+\epsilon}}{\sqrt{n}}, \tag{11}$$

   where $C(\tilde{x}) > 0$ is a constant independent of n, but dependent on $\tilde{x}$.

2. *Suppose that $\Theta(n) = \mathcal{O}(n^{-\zeta})$ for some $\zeta > 0$ and (H1), (H2a) hold. Then for all $\epsilon > 0$ and ν-a.e. $\tilde{x} \in \mathcal{X}$ we have that*

$$\left|\nu\{M_n \leq u_n\} - G_{\sqrt{n}}(u)\right| \leq Cn^{-\frac{1}{2}+\kappa}, \quad \text{with } \kappa = \epsilon + \frac{2(1+2\delta)}{\zeta\delta}. \tag{12}$$

   where $C(\tilde{x}) > 0$ is a constant independent of n, but dependent on $\tilde{x}$.

Remark 2.2. *For systems satisfying (H1) and (H2a) with superpolynomial decay of $\Theta(n)$ we see from equation (12) that the error rate is then of order $n^{-\frac{1}{2}+\epsilon}$, (for any $\epsilon > 0$). The error estimate is of little utility if either ζ or δ are close to zero.*

Remark 2.3. *We have explicitly stated our results for dimension 1. Following [17], we expect corresponding error rates to hold for non-uniformly expanding maps in dimension $d > 1$.*

We remark further that the constant C in the statement of the error rates depends on the regularity of the invariant density at $\tilde{x}$, and on the constants appearing in (H1), (H2a) and (H2b). The implied constant also depends on the recurrence properties associated to $\tilde{x}$ and hold only for generic $\tilde{x}$. In fact our results do not apply if $\tilde{x}$ is periodic. For observables maximized at periodic points, the corresponding EVDs may be expressed in terms of an additional parameter known as an *extremal index*, see [9] where this has been proven in a variety of settings. The recent work of [10] considers convergence rates in the presence of an extremal index.

The proof of the Theorem 2.1 relies on two propositions 2.4, 2.5 given in Section 3. We prove Theorem 2.1 in Section 3. The main examples that this theorem applies to are non-uniformly expanding maps such as those described in [3], intermittent maps [23, 30], and quadratic maps [6].

In Section 4 we make corresponding statements for certain non-uniformly hyperbolic dynamical systems admitting SRB measures, such as Lozi maps, hyperbolic billiards [13] and Hénon maps [2]. However, obtaining good control on the error rates also requires strong regularity constraints on the measure ν.

2.1 General estimates on the error rates

The following propositions give precise quantification of the error rates in terms of the assumptions on the correlation decay $\Theta(j)$, and the decay of $\nu(E_n)$. They will be used in the proof of Theorem 2.1. To state the propositions, we fix integers $p(n), q(n) > 0$ and let $n = pq + r$ with $0 \le r < p$ (by Euclid's division algorithm). The blocking argument consists of writing $n = p(n)q(n)$ and between each of the p gaps of length q we take a gap of length $t = g(n)$. The decay of correlations over the gap of length $t = g(n)$ allows us to consider successive blocks as approximately independent.

We suppose that $p, q \to \infty$ as $n \to \infty$. We let u_n be the sequence with the property that $n\nu\{X_1 > u_n\} \to \tau(u)$, for some function $\tau(u)$, and $u_n = u/a_n + b_n$.

Proposition 2.4. *Suppose that $f : \mathcal{X} \to \mathcal{X}$ is ergodic with respect to a measure ν which has a density $\rho \in L^{1+\delta}(m)$ for some $\delta > 0$. Suppose that (H1) holds. Then for ν-a.e. $x \in \mathcal{X}$, all p, q such that $n = pq + r$, and $t < p$, we have*

$$|\{M_n \le u_n\} - (1 - p\nu\{X_1 > u_n\})^q| \le \mathcal{E}_n, \tag{13}$$

where for any $\epsilon > 0$ and $\delta_1 = \delta/(1 + 2\delta)$:

$$\mathcal{E}_n = \max\{qt, p\}\nu\{X_1 \ge u_n\} + qp^2(\nu\{X_1 > u_n\})^2 + C_1 pq^2\Theta(t)^{\delta_1 - \epsilon} + p\sum_{j=2}^{t} \nu(X_1 > u_n, X_j > u_n). \tag{14}$$

Proposition 2.5. *Suppose that $f : \mathcal{X} \to \mathcal{X}$ is ergodic with respect to a measure ν with density $\rho \in L^{1+\delta}(m)$ for some $\delta > 0$. Suppose that (H2a) or (H2b) hold, and suppose for given $\epsilon > 0$ that $g(n) = \tilde{g}(n)^{1-\epsilon}$. Then for all $\rho_1 > 0$ and ν-a.e. $x \in \mathcal{X}$:*

$$\sum_{j=2}^{g(n)} \nu(X_1 > u_n, X_j > u_n) \le C(\rho_1, \tilde{x})\frac{1}{n^{\rho_1}}, \tag{15}$$

where the constant C depends on $\tilde{x}$ and ρ_1.

These propositions are proved in Section 5. Their proof involves an optimization of the dynamical blocking arguments and maximal function arguments developed in [3], and utilized in [17]. In these articles, the blocking and maximal functions arguments were used to prove convergence to EVD rather than to determine error rates for convergence.

2.2 On the convergence of $\nu\{M_n \le u_n\}$ to $G(u)$

In this subsection we describe the estimation procedure of the error involved in measuring $|G_{n^\beta}(u) - G(u)|$, for given $\beta \in (0, 1]$. As before we let $\tau_n(u) = n\nu\{\phi(x) > u_n\}$. We have the following elementary results, see [21].

Proposition 2.6. *For all $n \ge 2$:*

$$\Delta_{n^\beta} := \left|\left(1 - \frac{\tau_n(u)}{n^\beta}\right)^{n^\beta} - e^{-\tau_n(u)}\right| \le \frac{(\tau_n(u))^2 e^{-\tau_n(u)}}{2(n^\beta - 1)} \le \frac{0.3}{(n^\beta - 1)}. \tag{16}$$

Suppose that τ_n and τ satisfy $|\tau_n - \tau| \le \log 2$. Then for some $\theta \in (0, 1)$:

$$\Delta'_n := |e^{-\tau_n(u)} - e^{-\tau(u)}| \le e^{-\tau(u)}\{|\tau(u) - \tau_n(u)| + \theta(\tau(u) - \tau_n(u))^2\}. \tag{17}$$

We remark that the bound for Δ_{n^β} is uniform in u. This gives the corollary:

Corollary 2.7. *Suppose $(f, \mathcal{X}, \nu)$ is an ergodic dynamical system, then:*

$$|\nu\{M_n \leq u_n\} - G(u)| \leq |\nu\{M_n \leq u_n\} - G_{\sqrt{n}}(u)| + \Delta_{\sqrt{n}} + \Delta'_n. \tag{18}$$

Theorem 2.1 is used to estimate the first term on the right hand side of equation (18). The estimation of Δ_n and Δ'_n require knowledge of the explicit representation of $\phi(x)$. We will consider three explicit forms of $\phi(x)$, and then comment on the general cases. For these explicit forms, we will give explicit bounds on Δ_n and Δ'_n under the additional assumption that the density at $\tilde{x}$ is Lipschitz. For the class of non-uniformly expanding systems under consideration, this is a reasonable assumption. Consider the following representations $\phi = \phi_i : \mathcal{X} \to \mathbb{R}$ (for i=1,2 and 3) defined by:

$$\phi_1(x) = -\log(\mathrm{dist}(x, \tilde{x})),\ \phi_2(x) = \mathrm{dist}(x, \tilde{x})^{-\alpha},\ \phi_3(x) = C - \mathrm{dist}(x, \tilde{x})^{\alpha}, \tag{19}$$

for $\alpha > 0$ and $C \in \mathbb{R}$. In the cases of equation (19) the scaling sequences $u_n = u/a_n + b_n$ can be made explicit with $n\nu\{\phi > u_n\} \sim \tau(u)$, and we have:

$$\begin{aligned} \nu\{\phi_1(x) > u + \log n\} &\sim 2\rho(\tilde{x})e^{-u}/n, \\ \nu\{\phi_2(x) > u/(2n)^{1/\alpha}\} &\sim 2\rho(\tilde{x})u^{-1/\alpha}/n, \\ \nu\{\phi_3(x) \geq C - u/(2n)^{1/\alpha}\} &\sim 2\rho(\tilde{x})u^{1/\alpha}/n. \end{aligned}$$

Since the density is Lipschitz, the higher order terms in the above set of asymptotics are all $\mathcal{O}(1/n^2)$, where the implied constant depends on the Lipschitz norm. Hence by Proposition 2.6, $\Delta_{\sqrt{n}}$ is bounded by $\mathcal{O}(1/\sqrt{n})$ and Δ'_n is bounded by $\mathcal{O}(1/n)$. If instead the invariant density has Hölder exponent $\beta \in (0,1)$, the error Δ'_n is bounded by $\mathcal{O}(1/n^\beta)$. For observations that are general (regularly varying) functions of $\mathrm{dist}(x, \tilde{x})$, the scaling sequences a_n and b_n cannot always be made explicit and this leads to weaker estimates on the error bounds Δ_n and Δ'_n. Indeed for the Gaussian distribution, only bounds of order $1/(\log n)$ can be achieved, see [21].

3 Application of the main results

3.1 Proof of Theorem 2.1

Before considering specific dynamical systems, we show how Theorem 2.1 follows from Propositions 2.4 and 2.5. The proof requires optimizing the choice of constants q, p and the gap length $t < p$ which appear in the division algorithm $n \sim q(n)p(n)$ of the blocking argument, as well taking into account the decay of correlations and regularity of ν.

Given $\beta \in (0,1)$ we will suppose first that $p \sim n^{1-\beta}$ and so $q \sim n^\beta$. The constant r in $n = pq + r$ satisfies $r < p$ and hence $r = \mathcal{O}(n^{1-\beta})$. Assuming (H1) and either (H2a) or (H2b), we immediately have from Propositions 2.4 and 2.5 that:

$$\mathcal{E}_n \leq C_1 n^{\beta-1} g(n) + c_2 n^{-\beta} + C_3 n^{2-\beta}\{\Theta(g(n))\}^{\delta_1 - \epsilon} + C_4 \cdot o\left(\frac{1}{n}\right), \tag{20}$$

where $\delta_1 = \delta/(1+2\delta)$, and $\epsilon > 0$ arbitrary. In this estimate we taken $t = g(n)$ and used the fact that

$$\nu\{X_1 > u_n\} = \frac{\tau(u)}{n} + o\left(\frac{1}{n}\right), \tag{21}$$

Thus the constants C_1 and C_2 depend on the functional form of $\tau(u)$. In turn this behaves on the local behaviour of the invariant density at $\tilde{x}$ and on the functional form of the observable ϕ. The constant C_3 depends on δ (and hence the regularity of ν). The constant C_4 is from Proposition 2.5 and depends on the recurrence properties associated to $\tilde{x}$.

Let us now prove Theorem 2.1. In this case $\Theta(n) \leq \mathcal{O}(\theta_0^n)$, and under assumption (H2), we can take $t = (\log n)^\kappa$ for some $\kappa > 1$. For this choice it follows that $\Theta(t) \to 0$ at a super-polynomial rate. By assumption the measure ν has density in $L^{1+\delta}$ for some $\delta > 0$, and hence for any choice of $p, q = o(n)$ we have

$$C_3 n^{2-\beta}\{\Theta(t)\}^{\frac{\delta}{1+2\delta}-\epsilon} = o(1/n).$$

Inspecting the first two terms on the right hand side of equation (14), gives an optimal choice $p \approx q \approx \sqrt{n}$. For this choice of p and q, and noting that $n = pq + r$, we conclude the proof of part of Theorem 2.1 via the following sequence of estimates:

$$\begin{aligned}(1 - p\nu\{X_1 > u_n\})^q &= \left(1 - \frac{(n-r)}{\sqrt{n}}\nu\{X_1 > u_n\}\right)^{\sqrt{n}}, \\ &= \left(1 - \sqrt{n}\nu\{X_1 > u_n\}\right)^{\sqrt{n}} + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right),\end{aligned} \tag{22}$$

where in the last line we have used Proposition 2.6 in conjunction with the fact that $r\nu\{X_1 > u_n\} \leq \mathcal{O}(1/\sqrt{n})$. The convergence is pointwise in u.

To prove part 2 of Theorem 2.1, we see from (H2a) and Proposition 2.5 that in the estimation of $\mathcal{E}_n$ we can take $t = n^\kappa$ for any $\kappa < \gamma$. We inspect each term of the error $\mathcal{E}_n$ in equation (14). Again we take $\beta = 1/2$. The contribution of the first error term on the right hand side of (14) gives a contribution of order $n^{1/2+\kappa}$. The next significant error term is now the third right hand term of (14). Putting in $t = n^\kappa$ gives an error contribution:

$$C_3 n^{2-\beta}\{\Theta(n^\kappa)\}^{\delta_1} = C_3 n^{3/2} n^{-\zeta\kappa\delta_1}, \quad \text{with } \delta_1 = \frac{\delta}{1+2\delta} - \epsilon,$$

for any $\epsilon > 0$. In fact we require this error to be order $n^{-1/2}$, and hence this gives a bound on κ. We obtain for any $\epsilon_1 > 0$:

$$\kappa \geq \frac{2(1+2\delta)}{\delta\zeta} + \epsilon_1,$$

and thus we just take the minimal value of κ. This gives the required conclusion to part 2 of Theorem 2.1.

3.2 Uniformly expanding maps.

We derive explicit convergence rates for the tent map $f(x) = 1 - |1-2x|$ on $[0,1]$. Let $E_n^{(j)} := \{x : \mathrm{dist}(x, f^j(x)) < 1/n\}$, and suppose I is a monotonicity sub-interval of f^j and let $J = I \cap E_n^{(j)}$. Since $f^j(I) = [0,1]$ and f^j has slope 2^j, it follows easily that $|J| = O(2^{-j}/n)$. Hence, summing over all such J, we have $\nu(E_n^{(j)}) = O(1/n)$ and hence $\nu(E_n) = O(\tilde{g}(n)/n)$. To optimize $\mathcal{E}_n$, we can take a functional form $\tilde{g}(n) = (\log n)^{\gamma'}$ for any $\gamma' > 1$. By exponential decay of correlations, $\Theta(g(n))$ tends to zero at a superpolynomial rate. Here $g(n) = (\log n)^\kappa$ for some $1 < \kappa < \gamma'$. Thus conditions (H1) and (H2b) are valid. We summarize as follows:

Proposition 3.1. *Suppose* $(f, [0,1], \mathrm{Leb})$ *is the tent map. For the observation* $\phi(x) = -\log|x - \tilde{x}|$*, we have for* Leb*-a.e* $\tilde{x} \in [0,1]$ *and all* $\epsilon > 0$*:*

$$|\nu\{M_n \leq u + \log n\} - e^{-2e^{-u}}| \leq C\frac{(\log n)^{1+\epsilon}}{\sqrt{n}}, \tag{23}$$

where $C(\tilde{x}) > 0$ is a uniform constant depending on $\tilde{x}$.

Proof. Since part 1 of Theorem 2.1 applies, we have that

$$\left|\{M_n \le u_n\} - (1-\sqrt{n}\nu\{X_1 > u_n\})^{\sqrt{n}}\right| \le \mathcal{O}\left(\frac{(\log n)^{1+\epsilon}}{\sqrt{n}}\right). \tag{24}$$

To analyse Δ_n, Δ'_n from the function form $\phi(x) = -\log|x-\tilde{x}|$, we have in this case equality $\nu\{X_1 > u + \log n\} = 2e^{-u}/n$ (since ν is Lebesgue measure). Hence $\tau(u) = 2e^{-u}$, and Proposition 2.6 implies that:

$$\Delta_{\sqrt{n}} + \Delta'_n \le Cn^{-1/2},$$

and so the result follows. □

3.3 Non-uniformly expanding intermittency maps.

Consider the following interval map defined for $b > 0$:

$$f(x) = \begin{cases} x(1+(2x)^b) & \text{for } 0 \le x < \frac{1}{2}, \\ 2x-1 & \text{for } \frac{1}{2} \le x \le 1. \end{cases} \tag{25}$$

This map is non-uniformly expanding, and it has a neutral fixed point at $x = 0$. It was introduced in [23] as a simple model of intermittency and is sometimes called the Liverani-Saussol-Vaienti map. For $b \in (0,1)$, the map admits an absolutely continuous invariant measure, where the density lies in $L^{1+\delta}$ for any $\delta < 1/b - 1$. The system has polynomial decay of correlations: $\Theta(n) = O\left(n^{1-1/b}\right)$, see [30]. We have the following result:

Proposition 3.2. *Suppose $(f, [0,1], \nu)$ is the intermittent system (25) with $b < 1/20$ and consider the observable $\phi(x) = -\log|x-\tilde{x}|$. Then for all $\epsilon > 0$ and ν-a.e. $\tilde{x} \in \mathcal{X}$ we have that*

$$\left|\nu\{M_n \le u_n\} - e^{-2\rho(\tilde{x})e^{-u}}\right| \le Cn^{-\frac{1}{2}+\kappa}, \quad \text{with } \kappa = \epsilon + 5b. \tag{26}$$

where $C(\tilde{x}) > 0$ is a constant independent of n.

Proof. The following lemma, a version proved in [17] will be of use to us: Define $\mathcal{E}_n(\epsilon) := \{x : \mathrm{dist}(x, f^n(x)) < \epsilon\}$.

Lemma 3.3. *Suppose f is the interval map given by equation (25). There exists a uniform constant $C > 0$, such that $\forall n \ge 0$, and $\forall \epsilon > 0$ we have $m(\mathcal{E}_n(\epsilon)) \le C\sqrt{\epsilon}$.*

It follows that $m(E_n) \le C\tilde{g}(n)n^{-1/2}$, and hence by Hölder's inequality

$$\nu(E_n) \le C(\tilde{g}(n)n^{-1/2})^{1-b}.$$

Therefore if we take $\tilde{g}(n) = n^{\gamma}$ for any $\gamma \in (0, 1/2)$, $\nu(E_n)$ tends to zero, and so (H2a) applies. In order to control the error term $\mathcal{E}_n$ in Proposition 2.4 we require that $\Theta(g(n)) \to 0$ sufficiently fast, for some $g(n) = n^{\kappa} = o(\tilde{g}(n))$. The latter choice being made so that Proposition 2.5 applies. We can follow again the proof of part 2 of Theorem 2.1, and work out the minimal choice of κ. Taking $p \approx q \approx \sqrt{n}$ we will obtain an error of order $n^{\kappa+1/2}$ via the first right hand term in equation (14). To control all remaining right hand error terms of equation (14) (so that they are all of order $n^{-1/2}$) we need to take b sufficiently small. To work out the range of values of b, we need to show that

$$n^{3/2}\Theta(n^{\kappa})^{\delta_1} = \mathcal{O}\left(\frac{1}{\sqrt{n}}\right), \quad \text{with } \delta_1 = \frac{\delta}{1+2\delta} - \epsilon.$$

The constant $\epsilon > 0$ can be taken arbitrarily small. We recall that $\delta < 1/b - 1$, and $\Theta(n) = O\left(n^{1-1/b}\right)$. Hence this gives the bound on κ (valid for all $\epsilon_1 > 0$):

$$\kappa \geq \frac{2b(2-b)}{(1-b)^2} + \epsilon_1.$$

Notice for b close to one, this bound is of little utility since we require $\kappa < 1/2$. However for $b < 1/20$, then we can take $\kappa \geq 5b + \epsilon_1$, and the minimal choice leads to the required error estimate given in Proposition 3.2.

To analyse, $\Delta_{\sqrt{n}}$ and Δ'_n (recalling $q \sim \sqrt{n}$), we consider the functional form $\phi(x) = -\log|x - \tilde{x}|$ with scaling sequence is $u_n = u + \log n$. For $\tilde{x} \neq 0$, we can apply Lebesgue differentiation to deduce that

$$\nu\{\phi(x) > u + \log n\} \sim 2\rho(\tilde{x})e^{-u}/n.$$

However we need higher order regularity information on ν to deduce bounds on Δ_n, Δ'_n. In [23], the density of ρ is in fact locally Lipschitz ν-a.e. The fact that $\rho \in L^{1+\delta}$ is proved from an analysis of the singularity in $\rho(x)$ at $x = 0$. The density scales as $\mathcal{O}(x^{-b})$ as $x \to 0$. Hence for $\tilde{x} \neq 0$, we have

$$\nu\{\phi(x) > u + \log n\} = 2\rho(\tilde{x})e^{-u}/n + \mathcal{O}(n^{-2}).$$

Using Proposition 2.6 gives a combined error of order $n^{-1/2}$ for Δ_n, Δ'_n, and hence for all $b < \frac{1}{20}$ we have that:

$$|\nu\{M_n \leq u_n\} - e^{-2\rho(\tilde{x})e^{-u}}| \leq \mathcal{O}(1)n^{-\frac{1}{2}+\epsilon}, \tag{27}$$

□

3.4 Non-uniformly expanding quadratic maps.

Consider the quadratic family $f(x) = a - x^2$ for $x \in [-2, 2]$ and parameter $a \simeq 2$. For a positive measure set of parameter values, it is known that f admits an absolutely continuous invariant measure ν with density $\rho \in L^{1+\delta}$ for some $\delta > 0$. Moreover the system admits exponential decay of correlations, see [29]. Here E_n had explicit representation:

$$E_n = \{x \in [0,1] : \mathrm{dist}(f^j x, x) < \frac{1}{n}, \text{some } j \leq (\log n)^5\}. \tag{28}$$

It was shown in [3] that $\nu(E_n) \leq n^{-\alpha}$ for some $\alpha > 0$. Hence conditions (H1) and (H2b) are satisfied for this family, and we can take $g(n) = (\log n)^{1+\kappa}$ for some $\kappa > 0$. To get quantitative error estimates we require the critical orbit to satisfy a Misiurewicz condition, see [24]. Such a condition is required to ensure the invariant density is sufficiently regular. We have the following result:

Proposition 3.4. *Suppose $(f, [-2,2], \nu)$ is the quadratic family with a is a Misiurewicz parameter. For the observation $\phi(x) = -\log|x - \tilde{x}|$, we have for ν-a.e $\tilde{x} \in [0,1]$ and all $\epsilon > 0$:*

$$|\nu\{M_n \leq u + \log n\} - e^{-2\rho(\tilde{x})e^{-u}}| \leq C\frac{(\log n)^{1+\epsilon}}{\sqrt{n}}, \tag{29}$$

where $C(\tilde{x}) > 0$ is a uniform constant independent of n, but dependent on $\tilde{x}$.

Proof. Since part 1 of Theorem 2.1 applies, we have that

$$\left|\{M_n \leq u_n\} - (1-\sqrt{n}\nu\{X_1 > u_n\})^{\sqrt{n}}\right| \leq \mathcal{O}\left(\frac{(\log n)^{1+\epsilon}}{\sqrt{n}}\right), \tag{30}$$

To analyse Δ_n, Δ'_n from the function form $\phi(x) = -\log|x-\tilde{x}|$, we have in this case the asymptotic $\nu\{X_1 > u + \log n\} \sim 2\rho(\tilde{x})e^{-u}/n$ via the Lebesgue density theorem. Hence $\tau(u) = 2\rho(\tilde{x})e^{-u}$. However this asymptotic alone is not sufficient to get bounds on $\Delta_{\sqrt{n}}, \Delta'_n$ and we therefore require higher order regularity information on the density of $\rho(x)$. It is shown in [28] that $\rho(x) = \rho_1(x) + \rho_2(x)$, where $\rho_1(x)$ is of bounded variation, while

$$\rho_2(x) \leq \sum_{j=1}^{\infty} \frac{C(1.9)^{-j}}{|x - f^j(0)|}.$$

Hence if the parameter a is a Misiurewicz parameter, the critical orbit is pre-periodic, and thus by the bounded variation property the density $\rho(x)$ is locally Lipschitz ν-a.e. Proposition 2.6 therefore implies that:

$$\Delta_n + \Delta'_n \leq Cn^{-1/2},$$

and the result follows. ∎

4 Speed of convergence to EVD for non-uniformly hyperbolic systems

In this section we discuss corresponding convergence rate results for certain hyperbolic and non-uniformly hyperbolic dynamical systems. Examples under consideration include the Anosov Cat Map, The Lozi map, the Billiard map, and systems with rank one attractors such as the Hénon family. Extreme statistics for these systems have been studied for example in [2, 13, 19, 18] (amongst others). For these systems, we can again achieve a result analogous to part 1 of Theorem 2.1. However to establish rates of convergence (or even just convergence) to a limit $G(u)$ along the sequence $\nu\{M_n \leq u/a_n + b_n\}$ specific information on the regularity of ν is required, (such as continuity of the density). In general this is an open problem for non-uniformly hyperbolic systems.

More precisely, we consider a non-uniformly hyperbolic dynamical system $(f, \mathcal{X}, \mu)$, where ν is a SRB measure supported on an f-invariant set $\mathcal{X} \subset \mathbb{R}^2$. We assume that the system is modelled by a Young tower, with recurrence time statistics that are at least polynomial, see [29] for details. In the case of applications, our examples will all have exponential decay of correlations, but in the statement of our results we allow for sub-exponential decay of correlations. We will assume that conditions (H1), (H2a) and/or (H2b), as defined in Section 2 continue to hold. For non-uniformly hyperbolic systems, we also need further control on the regularity of ν. Recall that the pointwise local dimension of ν is given by:

$$d := \lim_{r\to 0} \frac{\log \nu(B(x,r))}{\log r}, \tag{31}$$

whenever this limit exists. Here $B(x,r)$ is the ball of radius r centered at $x \in \mathcal{X}$. From [22] it is known that the local dimension of ν exists and is constant ν-a.e. for a non-uniformly hyperbolic system. However, we also need control on the regularity of ν on certain shrinking annuli. We state the following assumption (H3):

(H3) **(Regularity of ν on shrinking annuli).** For all $\delta > 1$ and ν-a.e.$x \in \mathcal{X}$, there exists $\kappa > 0$ such that

$$|\nu(B(x, r + r^{\delta})) - \nu(B(x, r))| \leq C r^{\kappa\delta}. \tag{32}$$

The constant C and κ depending on x (but not δ).

Condition (H3) (and versions thereof) have been stated and verified in [2, 13, 18]. We state the following result.

Theorem 4.1. *Suppose that $f : \mathcal{X} \to \mathcal{X}$ is a non-uniformly hyperbolic system with ergodic SRB measure ν. Suppose (H3) holds.*

1. *Suppose we have exponential decay of correlations, i.e. $\Theta(n) = \mathcal{O}(\theta_0^n)$ with $\theta_0 < 1$ and that (H1) and (H2b) hold. Then for all $\epsilon > 0$ and ν-a.e. $\tilde{x} \in \mathcal{X}$, we have that*

$$\left|\nu\{M_n \leq u_n\} - G_{\sqrt{n}}(u)\right| \leq C \frac{(\log n)^{1+\epsilon}}{\sqrt{n}}, \tag{33}$$

where $C(\tilde{x}) > 0$ is a constant independent of n.

2. *Suppose we have polynomial decay of correlations i.e. $\Theta(n) = \mathcal{O}(n^{-\zeta})$ for some $\zeta > 1$ and (H1) and (H2a) hold. Then for all $\epsilon > 0$ and ν-a.e. $\tilde{x} \in \mathcal{X}$ we have that*

$$\left|\nu\{M_n \leq u_n\} - G_{\sqrt{n}}(u)\right| \leq C_1 n^{-\frac{1}{2}+\kappa}, \quad \text{with } \kappa = \epsilon + \frac{C_\kappa}{\zeta}. \tag{34}$$

where C_1 is independent of n. The constant $C_\kappa > 0$ is independent of ζ, but depends on κ in (H3).

The proof of this theorem in large part follows from the proof Propositions 2.4 and 2.5, but it also utilizes the results [2, 13, 18]. These latter references use a direct blocking argument approach to verify $D'(u_n), D(u_n)$ for hyperbolic systems. We point out in Section 5, how the proofs of these propositions are modified in the hyperbolic setting. The range of examples that satisfy (H1), (H2a) and (H3) (with exponential decay of $\Theta(n)$) include Lozi maps and hyperbolic billiards [13] and the Hénon family [2]. For higher dimensional hyperbolic systems with polynomial decay of correlations, less is known about convergence to EVD (i.e. examples that satisfy (H1), (H2b) and (H3)), but some progress in this direction is made in [27].

For non-uniformly hyperbolic systems equations (33) and (34) are the best that we can achieve on the convergence rates, at least for regular observables $\phi : \mathcal{X} \to \mathbb{R}$. The main difficulty is in the control of the fluctuations of $\nu\{\phi(x) > u_n\}$ as $n \to \infty$. From the definition of local dimension the function $r \mapsto \nu\{B(x, r)\}$ need not be regularly varying as $r \to 0$, and hence even for smooth observables such as $\phi(x) = -\log(\text{dist}(x, \tilde{x}))$, the sequence $\tau_n(u) = n\nu\{\phi(x) > u + \log n\}$ may fluctuate wildly as $n \to \infty$. However, along other (non-linear) scalings $u_n(u)$ of u control on the the rate of convergence of $\tau_n(u)$ to $\tau(u)$ might be achievable. Alternatively, for observables $\phi : \mathcal{X} \to \mathbb{R}$ tailored to the measure ν, so that $\nu\{\phi(x) > u_n\}$ is regularly varying in u, convergence rates can again be achieved. Such observables are considered in [8].

An example of a (uniformly) hyperbolic system for which error rates can be achieved is in the case of the Arnold Cat Map $(f, \mathbb{T}^2, \nu)$, where $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ is the two-torus, ν is two-dimensional Lebesgue measure, and

$$f(x, y) = (2x + y, x + y). \tag{35}$$

we then have the following result:

Proposition 4.2. *Suppose* $(f, \mathbb{T}^2, \nu)$ *is the Arnold Cat Map defined by equation* (35). *For the observation* $\phi(x) = -\log(\mathrm{dist}(x, \tilde{x}))$, *we have for* ν*-a.e* $\tilde{x} \in [0,1]$ *and all* $\epsilon > 0$*:*

$$|\nu\{M_n \leq (u + \log n)/2\} - e^{-\pi e^{-u}}| \leq C\left(\frac{(\log n)^{1+\epsilon}}{\sqrt{n}}\right), \tag{36}$$

where C *is independent of* n*, but dependent on* $\tilde{x}$*.*

Proof. This map is uniformly hyperbolic, admitting a finite Markov partition, and having uniform expansion estimates on unstable manifolds. The methods of [13] immediately apply, and thus conditions (H1) and (H2b) hold, with $\Theta(j) \leq \mathcal{O}(\theta_0^j)$ for some $\theta_0 < 1$. Condition (H3) clearly holds since ν is Lebesgue measure. Hence (as with the tent map case)

$$\left|\{M_n \leq u_n\} - (1 - \sqrt{n}\nu\{X_1 > u_n\})^{\sqrt{n}}\right| \leq \mathcal{O}\left(\frac{(\log n)^{1+\epsilon}}{\sqrt{n}}\right), \tag{37}$$

To analyse $\Delta_{\sqrt{n}}, \Delta_n'$ from the functional form $\phi(x) = -\log \mathrm{dist}(x, \tilde{x})$, we have in this case equality

$$\nu\{X_1 > u + \log n\} = \nu\left\{x : \mathrm{dist}(x, \tilde{x}) \leq \frac{e^{-u/2}}{\sqrt{n}}\right\} = \frac{\pi e^{-u}}{n},$$

since ν is just Lebesgue measure. Hence $\tau(u) = \pi e^{-u}$, and Proposition 2.6 implies that:

$$\Delta_n + \Delta_n' \leq C n^{-1/2},$$

and so the result follows. □

5 Proofs of the main results

To prove Theorem 2.1 we work directly with the blocking methods developed in [3, 17]. If we just require convergence to EVD, then it suffices to check the conditions $D(u_n)$ and $D'(u_n)$ stated in [21] and/or the dynamical versions formulated in [5].

5.1 The blocking argument

We divide n successive observations $\{X_1, \ldots, X_n\}$ into q blocks of length $p + t$, with p, q, j dependent on n, and $n \sim p(n)q(n)$ as $n \to \infty$. The gap t will be large enough that successive p blocks are approximately independent but small enough so that $\nu(M_n \leq u_n)$ is approximately equal to $\nu(M_{q(p+t)} \leq u_n)$. Using approximate independence of p blocks it is shown in [3] that

$$|\nu(M_n \leq u_n) - \nu(M_{q(p+t)} \leq u_n)| \leq \max\{qt, p\}\nu(\phi > u_n), \tag{38}$$

$$|\nu(M_{l(p+t)} \leq u_n) - (1 - p\nu(\phi > u_n))\nu(M_{(l-1)(p+t)} \leq u_n)| \leq \tilde{\Gamma}_n, \quad (l \in [1, q]), \tag{39}$$

where

$$\tilde{\Gamma}_n = p\gamma(n,t) + 2p\sum_{i=1}^{p}\nu(X_1 > u_n, X_i > u_n),$$

and for $t = o(p(n))$, the quantity $\gamma(t,n)$ is given by:

$$\gamma(t,n) = \nu\{X_1 > u_n\}^{-(1+\eta)}\Theta(j) + \nu\{x_1 > u_n\}^{(1+\eta)\theta}, \tag{40}$$

where $\theta = \delta/(1+\delta)$. From this we deduce that

$$|\nu(M_n \leq u_n) - (1 - p\nu(X_1 > u_n))^q| \leq q\Gamma_n$$

where

$$\Gamma_n = \tilde{\Gamma}_n + t\nu(X_1 > u_n).$$

5.2 Proof of Proposition 2.5.

For a function $\varphi \in L^1(m)$ we define the Hardy–Littlewood maximal function

$$\mathcal{M}(x) := \sup_{a>0} \frac{1}{2a} \int_{x-a}^{x+a} \varphi(y) dm(y).$$

A theorem of Hardy and Littlewood [26], implies that

$$m(|\mathcal{M}(x)| > \lambda) \leq \frac{\|\varphi\|_1}{\lambda} \tag{41}$$

where $\|\cdot\|_1$ is the L^1 norm with respect to m. Recalling

$$E_n = \left\{ x : \operatorname{dist}(x, f^j(x)) \leq \frac{1}{n} \text{ for some } j \leq \tilde{g}(n) \right\},$$

let $\rho(x)$ denote the density of ν with respect to m and let $\mathcal{M}_n(x)$ denote the maximal function of $\varphi(x) := 1_{E_n}(x)\rho(x)$. For constants $a, b > 0$ to be fixed later consider sequences $\lambda_n = n^{-a}$ and $\alpha_n = \lfloor n^b \rfloor$. Inequality (41) gives

$$m(|\mathcal{M}_{\alpha_n}(x)| > \lambda_n) \leq \frac{\nu(E_{\alpha_n})}{\lambda_n} \leq \frac{1}{n^{\alpha b - a}}.$$

If $\alpha b - a > 1$ (first constraint required on a and b), then the First Borel–Cantelli Lemma implies for ν a.e. x there exists an $N := N(x)$ such that for all $n \geq N$ we have $|\mathcal{M}_{\alpha_n}(\tilde{x})| < \lambda_n$. Recall in the case of Theorem 2.1 that ν is absolutely continuous with respect to m. Hence, for all n sufficiently large

$$\begin{aligned} \nu(\{x : \operatorname{dist}(x, \tilde{x}) < \alpha_n^{-1}\} \cap E_{\alpha_n}) &\leq \int_{x-\alpha_n^{-1}}^{x+\alpha_n^{-1}} \varphi_{\alpha_n}(y) dm(y) \\ &\leq \alpha_n^{-1} \mathcal{M}_{\alpha_n}(\tilde{x}) \\ &\leq \alpha_n^{-1} \lambda_n = O(n^{-a-b}). \end{aligned} \tag{42}$$

Denote $A := \{X_1 > u_k, X_j > u_k\}$ with $2 \leq j \leq g(k)$, and $g(n) = \tilde{g}(n)^{(1-\epsilon)}$ for some $\epsilon > 0$. We assume that $\tilde{g}(n)$ has the representations given in either (H2a) or (H2b). Since $\psi^{-1}(u_k) \approx 1/k$, there exists a $v > 0$ such that

$$A \subset \left\{ x : \operatorname{dist}(\tilde{x}, x) \leq \frac{v}{k} \ \operatorname{dist}(\tilde{x}, f^j(x)) \leq \frac{v}{k} \text{ for some } j \leq g(k) \right\}.$$

Given the sequence α_n, let $k/(2v) \in [\alpha_n, \alpha_{n+1})$. Then (by monotonicity of $g(n)$),

$$A \subset \left\{ x : \operatorname{dist}(\tilde{x}, x) \leq \frac{1}{2\alpha_n}, \ \operatorname{dist}(\tilde{x}, f^j(x)) \leq \frac{1}{2\alpha_n} \text{ for some } j \leq g((2v)\alpha_{n+1}) \right\}.$$

Applying the triangle inequality $\operatorname{dist}(x, f^j(x)) \leq \operatorname{dist}(\tilde{x}, x) + \operatorname{dist}(\tilde{x}, f^j(x))$ gives

$$A \subset \left\{ x : \operatorname{dist}(\tilde{x}, x) \leq \frac{1}{\alpha_n}, \ \operatorname{dist}(x, f^j(x)) \leq \frac{1}{\alpha_n} \text{ for some } j \leq g((2v)\alpha_{n+1}) \right\}.$$

Since $\alpha_n = \lfloor n^b \rfloor$ we have that $\lim_{n\to\infty} \alpha_{n+1}/\alpha_n = 1$. By the growth properties of g and $\tilde{g}$ (as given in Proposition 2.5), there exists $\kappa_v > 0$ and a sequence $c_n \to 0$ such that for all sufficiently large α_n:

$$g((2v)\alpha_{n+1}) \leq g(2(2v)\alpha_n) \leq c_n \tilde{g}((2(2v)\alpha_n) \leq c_n \kappa_v \tilde{g}(\alpha_n).$$

Moreover, there exists N such that $\forall n \geq N$ we have $c_n\kappa_v < 1$, and hence

$$A \subset \left\{ x : \mathrm{dist}(\tilde{x}, x) \leq \frac{1}{\alpha_n},\ \mathrm{dist}(x, f^j(x)) \leq \frac{1}{\alpha_n} \text{ for some } j \leq \tilde{g}(\alpha_n) \right\}.$$

Applying inequality (42) gives

$$\nu(X_1 > u_k, X_j > u_k) = O(k^{-a-b}) \quad \text{for all } k > N,$$

so that

$$k \sum_{j=1}^{g(k)} \nu(X_1 > u_k, X_j > u_k) = O(k^{1-a-b} g(k)) \quad \text{for all } k > N.$$

By definition of α and $\tilde{g}(k)$ in condition (H2a) (or (H2b)), there exists $\beta_0 < 1$ such that for all $\beta > \beta_0$ we have $g(k) = o(k^{\beta})$. Hence the second constraint on a, b required is that $2 - a - b < 0$ so that $k^{1-a-b} g(k) = o(1)$. Simultaneous to $\alpha b - a > 1$, these constraints can be satisfied given any $\alpha \in (0, 1)$. Moreover the quantity $\rho_1 = a + b - 2 > 0$ can made arbitrarily large. Hence the first part of (15) is satisfied.

5.3 Proof of Proposition 2.4

We state and prove the following lemma. Combining this result with the blocking argument described earlier in Section 5.1 completes the proof of Proposition 2.4.

Lemma 5.1. *For any $g(n) < p$ we have that:*

$$\sum_{j=g(n)}^{p} \nu(X_0 > u_n, X_j > u_n) \leq p\mathcal{O}(1)(\Theta(g(n)))^{\delta_1} + p(\nu\{x_1 > u_n\})^2, \quad \text{with } \delta_1 = \frac{\delta}{(1+2\delta)} - \epsilon, \tag{43}$$

where $\epsilon > 0$ can be taken arbitrarily small, and the implied constant depends on δ.

Under assumptions (H2a) or (H2b), and Propostion 2.5 a condition on the choice of $g(n)$ is that $g(n) = o(\tilde{g}(n))$. We now use decay of correlations to prove the second part of (43). We will write $\phi(x) = \psi(\mathrm{dist}(x, \tilde{x})$, and recall that we work with the explicit observable $\psi(y) = -\log y$, for $y > 0$. As before we approximate the indicator function $\Phi := 1_{\{X_1 > u_n\}}$ by a Lipschitz continuous function Φ_B, which is set equal to 1 inside a ball centered at $\tilde{x}$ of radius $\ell'_n := \phi^{-1}(u_n) - [\psi^{-1}(u_n)]^{1+\eta}$, for some $\eta > 0$, and decaying to 0 at a linear rate so that Φ_B vanishes on the boundary of $\{X_1 > u_n\}$. The Lipschitz norm of Φ_B is bounded by $[\psi^{-1}(u_n)]^{-(1+\eta)}$.

We now take $j \in [g(n), n]$. We have the following triangle inequality:

$$\begin{aligned} \left| \int \Phi(\Phi \circ f^j) d\nu - \left(\int \Phi d\nu \right)^2 \right| \quad &\leq \quad \left| \int \Phi_B(\Phi \circ f^j) d\nu - \int \Phi_B d\nu \int \Phi d\nu \right| \\ &\quad + \left| \int (\Phi_B - \Phi) \Phi \circ f^j d\nu - \int (\Phi - \Phi_B) d\nu \int \Phi d\nu \right|, \end{aligned}$$

and we estimate each term on the right hand side. By decay of correlations and for sufficiently large n:

$$\left| \int \Phi_B(\Phi \circ f^j) d\nu - \int \Phi_B d\nu \int \Phi d\nu \right| \leq \|\Phi_B\|_{\mathrm{Lip}} \|\Phi\|_{\infty} \Theta(g(n)) = O\left(n^{1+\eta} \Theta(g(n))\right),$$

and

$$\left|\int(\Phi_B-\Phi)\Phi\circ f^j d\nu - \int(\Phi-\Phi_B)d\nu\int\Phi d\nu\right| \leq 2\|\Phi\|_\infty \nu(x:\Phi_B(x)\neq\Phi(x)) = O(n^{-\theta(1+\eta)}),$$

where we can take any $\theta<\delta/(1+\delta)$. Hence for each $j>g(n)$ we obtain:

$$\nu(X_1>u_k,X_j>u_k)\leq C_1|\{X_1\geq u_n\}|^{-1-\eta}\Theta(j)+C_2|\{X_1\geq u_n\}|^{\theta(1+\eta)}+(\nu\{X_1>u_n\})^2, \quad (44)$$

where C_1, C_2 depend on the regularity of $\rho(x)$ at $\tilde{x}$ and on the Lipschitz norm. The constant η is arbitrary, and hence we can optimize the right hand side by varying η. An elementary calculus argument shows that

$$\nu(X_1>u_k,X_j>u_k)\leq \mathcal{O}(1)(\Theta(j))^{\delta_1}+(\nu\{X_1>u_n\})^2, \text{with } \delta_1=\frac{\delta}{(1+2\delta)}-\epsilon. \quad (45)$$

Here $\epsilon>0$ can be made arbitrarily small. The implied constant depends only δ. This gives the required result.

5.4 Proof of Theorem 4.1

The blocking argument described in Section 5.1 is purely probabilistic. For hyperbolic systems, the extra hypothesis required is (H3) on the regularity of the measure ν. For non-uniformly expanding systems, we used the fact that the density of ν belonged to $L^{1+\delta}$ for some $\delta>0$. Condition (H3) is the corresponding regularity constraint required to prove the result. We point out the main modifications required over and above the details presented in the proofs of Propositions 2.4 and 2.5 in order to prove Theorem 4.1.

For hyperbolic systems, and assuming condition (H2), the proof of Proposition 2.5 follows step by step but instead we use a modified maximal function

$$\mathcal{M}(x):=\sup_{a>0}\frac{1}{m_\gamma(B(x,r))}\int_{B(x,r)}\varphi(y)dm(y),$$

where m_γ is the Riemannian measure on the local unstable manifold γ centered at x, and $\varphi(x):=1_{E_n}(x)\rho_\gamma(x)$ where ρ_γ is the conditional density of ν on γ (which is absolutely continuous with respect to m_γ. With this modification all steps are as before. See [13, 18].

Under the assumption of (H1) and (H3), the modification of proof of Proposition 2.4 in the hyperbolic case again follows [13, 18]. In particular condition (H3) comes into the Lipschitz approximation of $1_{\{X_1>u_n\}}$, to estimate $\|\Phi(x)-\Phi_B(x)\|_1$. The result is that equation (45) is modified to

$$\nu(X_1>u_k,X_j>u_k)\leq \mathcal{O}(1)(\Theta(j))^{\delta'}+(\nu\{X_1>u_n\})^2, \quad (46)$$

where the implied constant and δ' both depend on κ (from (H3)). Theorem 4.1 now follows.

References

[1] M. Benedicks and L. Carleson. On iterations of $1-ax^2$ on $(-1,1)$, *Annals of Math.*, **122**, (1985), 1-25.

[2] 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.

[3] P. Collet. Statistics of closest return for some non-uniformly hyperbolic systems. *Erg. Th. Dyn. Syst.* **21** (2001), 401-420.

[4] A. Ferguson and M. Pollicott. Escape rates for Gibbs measures. *Erg. Th. Dyn. Syst.* **32** (2012), no. 3, 961988.

[5] 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.

[6] J. Freitas and A. Freitas. 'Extreme values for Benedicks Carleson maps'. *Erg. Th. Dyn. Sys.*, **28**, (2008), no. 4, 1117-1133.

[7] J. Freitas, A. Freitas and M. Todd. Hitting Times and Extreme Value Theory, *Probab. Theory Related Fields*,**147(3)**:675–710, 2010.

[8] J. Freitas, A. Freitas and M. Todd. Extreme value laws in dynamical systems for non-smooth observations. *J. Stat. Phys.* **142** (2011), no. 1, 108–126.

[9] A.C.M. Freitas, J.M. Freitas, and M. Todd. Extremal Index, Hitting Time Statistics and periodicity, *Adv. Math.*, **231**, no. 5, 2012, 2626-2665.

[10] A.C.M. Freitas, J.M. Freitas, and M. Todd. Speed of convergence for laws of rare events and escape rates. Preprint, 2014.

[11] J. Galambos. *The Asymptotic Theory of Extreme Order Statistics*, John Wiley and Sons, 1978.

[12] C. Gupta. Extreme-value distributions for some classes of non-uniformly partially hyperbolic dynamical systems. *Ergodic Theory and Dynamical Systems,* **30**, (3), (2011), 757-771.

[13] 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.

[14] P. Hall. On the Rate of Convergence of Normal Extremes. *Journal of Applied Probability,* **16**, no. 2 (1979), 433-439.

[15] W. Hall and J. Wellner The rate of convergence in law of the maximum of an exponential sample. *Statist. Neerlandica*, **33**, no. 3, (1979), 151154.

[16] M. Hirata. Poisson Limit Law for Axiom A diffeomorphisms. *Erg. Th. Dyn. Syst.* **13** (1993), no. 3, 533556.

[17] 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.

[18] M. P. Holland, P. Rabassa and A. E. Sterk. On the convergence to an extreme value distribution for non-uniformly hyperbolic dynamical systems. Preprint 2013.

[19] M. P. Holland, R. Vitolo, P. Rabassa, A. E. Sterk, H. Broer. Extreme value laws in dynamical systems under physical observables. *Physica D: Nonlinear Phenomena,* 241 (2012), 497-513.

[20] G. Keller. Rare events, exponential hitting times and extremal indices via spectral perturbation. *Dyn. Syst.* **27** (2012), no. 1, 1127.

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

[22] F. Ledrappier, *Dimension of invariant measures*, Proceedings of the conference on ergodic theory and related topics, II (Georgenthal, 1986), Teubner-Texte, Math. **94** (1987), 116–124

[23] C. Liverani, B. Saussol and S. Vaienti. A probabilistic approach to intermittency. *Ergodic Theory and Dynamical Systems.*, **19**, (1999), 671-685.

[24] W. de Melo and S. van Strien. *One-dimensional dynamics.* Springer-Verlag. 1993.

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

[26] W. Rudin. *Real and Complex Analysis*, Third Edition, 1987, McGraw Hill.

[27] P. Wasilewska. *Limiting distribution and error terms for the number of visits to balls in mixing dynamical systems.* PhD Thesis. 2013.

[28] L. S. Young. Decay of correlations for certain quadratic maps. *Comm. Math. Phys.*, **146**, (1992), 123-138.

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

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