Extreme Value distribution for singular measures

Faranda, Davide

Department of Mathematics and Statistics, University of Reading;
Whiteknights, PO Box 220, Reading RG6 6AX, UK. d.faranda@pgr.reading.ac.uk

Lucarini, Valerio

Department of Meteorology, University of Reading;
Department of Mathematics and Statistics, University of Reading;
Whiteknights, PO Box 220, Reading RG6 6AX, UK. v.lucarini@reading.ac.uk

Turchetti, Giorgio

Department of Physics, University of Bologna.INFN-Bologna
Via Irnerio 46, Bologna, 40126, Italy. turchett@bo.infn.it

Vaienti, Sandro

UMR-6207, Centre de Physique Théorique, CNRS, Universités d'Aix-Marseille I,II,
Université du Sud Toulon-Var and FRUMAM
(Fédération de Recherche des Unités de Mathématiques de Marseille);
CPT, Luminy, Case 907, 13288 Marseille Cedex 09, France.
vaienti@cpt.univ-mrs.fr

Abstract

In this paper we perform an analytical and numerical study of Extreme Value distributions in discrete dynamical systems that have a singular measure. Using the block maxima approach described in Faranda et al. [2011] we show that, numerically, the Extreme Value distribution for these maps can be associated to the Generalised Extreme Value family where the parameters scale with the information dimension. The numerical analysis are performed on a few low dimensional maps. For the middle third Cantor set and the Sierpinskij triangle obtained using Iterated Function Systems, experimental parameters show a very good agreement with the theoretical values. For strange attractors like Lozi and Hènon maps a slower convergence to the Generalised Extreme Value distribution is observed. Even in presence of large statistics the observed convergence is slower if compared with the maps which have an absolute continuous invariant measure. Nevertheless and within the uncertainty computed range, the results are in good agreement with the theoretical estimates.

The existence of extreme value laws for dynamical systems preserving an absolutely continuous invariant measure or a singular continuous invariant measure has been recently proven if strong mixing properties or exponential hitting time statistics on balls are satisfied. In our previous work we have shown that there exists an algorithmic way

to study extrema by using a block-maxima approach for dynamical systems which possess an absolutely continuous invariant measure and satisfy certain mixing properties. In this work we test our algorithm for maps that do not have an absolutely continuous invariant measure showing that the cumulative distribution function of maxima is related to the scaling of the measure of a ball centered around generic points. The scaling exponent turns out to be the Hausdorff dimension of the measure (also known as information dimension). Even if we cannot estimate analytically the asymptotic behavior of the measure of the balls, the agreement with the numerical simulations we have carried out for different maps suggests the validity of our proposed scaling in terms of the information dimension. Our conjecture has been tested with numerical experiments on different low dimensional maps such as the middle third Cantor set, the Sierpinskij triangle, Iterated Function System (IFS) with non-uniform weights and strange attractors such as Lozi and Hénon. In all cases considered, there is a good agreement between the theoretical parameters and the experimental ones although, in case of strange attractors which exhibit multifractal structures, the convergence is slower. To perform the numerical simulations it has been used the L-moments procedure in order to overcome the difficulties of dealing with a singular continuous invariant measure.

1   Introduction

1.1   Classical Extreme Value Theory

Extreme Value Theory (EVT), developed for the study of stochastical series of independent and identical distributed variables by Fisher and Tippett [1928] and formalized by Gnedenko [1943], has been successfully applied to different scientific fields to understand and possibly forecast events that occur with very small probability but that can be extremely relevant from an economic or social point of view: extreme floods [Gumbel, 1941], [Sveinsson and Boes, 2002], [P. and Hense, 2007], amounts of large insurance losses [Brodin and Kluppelberg, 2006], [Cruz, 2002]; extreme earthquakes [Sornette et al., 1996], [Cornell, 1968], [Burton, 1979]; meteorological and climate events [Felici et al., 2007], [Vitolo et al., 2009b], [Altmann et al., 2006], [Nicholis, 1997], [Smith, 1989]. An extensive review of the techniques and applications related to the EVT is presented in Ghil et al. [2011].

Gnedenko [1943] studied the convergence of maxima of i.i.d. variables

$$X_0, X_1, ..X_{m-1}$$

with cumulative distribution (cdf) $F(x)$ of the form:

$$F(x) = P\{a_m(M_m - b_m) \leq x\}$$

Where a_m and b_m are normalizing sequences and $M_m = \max\{X_0, X_1, ..., X_{m-1}\}$. It may be rewritten as $F(u_m) = P\{M_m \leq u_m\}$ where $u_m = x/a_m + b_m$. Under general hypotesis on the nature of the parent distribution of data, Gnedenko [1943] show that the distribution of maxima, up to an affine change of variable, obeys to one of the following three laws:

- Type 1 (*Gumbel*).
$$E(x) = \exp(-e^{-x}), \quad -\infty < x < \infty \tag{1}$$

- Type 2 (*Fréchet*).
$$E(x) = \begin{cases} 0, & x \leq 0 \\ \exp(-x^{-\xi}), & \text{for some } \xi > 0, \; x > 0 \end{cases} \tag{2}$$

- Type 3 (*Weibull*).
$$E(x) = \begin{cases} \exp(-(-x)^{-\xi}), & \text{for some } \xi > 0, \; x \leq 0 \\ 1, & x > 0 \end{cases} \tag{3}$$

Let us define the right endpoint x_F of a distribution function $F(x)$ as:

$$x_F = \sup\{x : F(x) < 1\} \tag{4}$$

then, it is possible to compute normalizing sequences a_m and b_m using the following corollary of Gnedenko's theorem :

Corollary (Gnedenko): *The normalizing sequences a_m and b_m in the convergence of normalized maxima $P\{a_m(M_m - b_m) \leq x\} \to F(x)$ may be taken (in order of increasing complexity) as:*

- *Type 1:* $a_m = [G(\gamma_m)]^{-1}, \quad b_m = \gamma_m;$

- *Type 2:* $a_m = \gamma_m^{-1}, \quad b_m = 0;$

- *Type 3:* $a_m = (x_F - \gamma_m)^{-1}, \quad b_m = x_F;$

where

$$\gamma_m = F^{-1}(1 - 1/m) = \inf\{x; F(x) \geq 1 - 1/m\} \tag{5}$$

$$G(t) = \int_t^{x_F} \frac{1 - F(u)}{1 - F(t)} du, \quad t < x_F \tag{6}$$

In [Faranda et al., 2011] we have shown that this approach is equivalent to fit unnormalized data directly to a single family of generalized distribution called GEV distribution with cdf:

$$F_G(x; \mu, \sigma, \xi') = \exp\left\{ -\left[1 + \xi'\left(\frac{x - \mu}{\sigma}\right)\right]^{-1/\xi'} \right\} \tag{7}$$

which holds for $1 + \xi'(x - \mu)/\sigma > 0$, using $\mu \in \mathbb{R}$ (location parameter) and $\sigma > 0$ (scale parameter) as scaling constants in place of b_m, and a_m [Pickands III, 1968], in particular, in Faranda et al. [2011] we have shown that the following relations hold:

$$\mu = b_m \qquad \sigma = \frac{1}{a_m}.$$

$\xi' \in \mathbb{R}$ is the shape parameter also called the tail index: when $\xi' \to 0$, the distribution corresponds to a Gumbel type (Type 1 distribution). When the index is positive, it corresponds to a Fréchet (Type 2 distribution); when the index is negative, it corresponds to a Weibull (Type 3 distribution).

To analyze the extreme value distribution in a series of data two main approaches can be applied: the Peak-over-threshold approach and the Block-Maxima approach. The former consists in looking at exceedance over high thresholds [Todorovic and Zelenhasic, 1970] and a Generalized Pareto distribution is used for modeling data obtained as excesses over thresholds [Smith, 1984], [Davison, 1984], [Davison and Smith, 1990].
The so called block-maxima approach is widely used in climatological and financial applications since it represents a very natural way to look at extremes in fixed time intervals: it consists of dividing the data series of some observable into bins of equal length and selecting the maximum (or the minimum) value in each of them [Coles et al., 1999], [Felici et al., 2007], [Katz and Brown, 1992], [Katz, 1999], [Katz et al., 2005].

1.2   Extreme Value Theory for dynamical systems

As far as the classical EVT is concerned, we should restrict our domain of investigation to the output of stochastic processes. Obviously, it is of crucial relevance for both mathematical reason and for devising a framework to be used in applications, to understand under which circumstances the time series of observables of deterministic dynamical system can be treated using EVT.
Empirical studies show that in some cases a dynamical observable obeys to the extreme value statistics even if the convergence is highly dependent on the kind of observable we choose [Vannitsem, 2007], [Vitolo et al., 2009b], [Vitolo et al., 2009a]. For example, Balakrishnan et al. [1995] and more recently Nicolis et al. [2006] and Haiman [2003] have shown that for regular orbits of dynamical systems we don't expect to find convergence to EV distribution.
The first rigorous mathematical approach to extreme value theory in dynamical systems goes back to the pioneer paper by P. Collet in 2001 [Collet, 2001]. Important contributions have successively been given by Freitas and Freitas [2008], Freitas et al. [2009],Freitas et al. [2010a] and by Gupta et al. [2009]. The starting point of all these investigations was to associate to the stationary stochastic process given by the dynamical system, a new stationary independent sequence which enjoyed one of the classical three extreme value laws, and this laws could be pulled back to the original dynamical sequence. To be more precise we will consider a dynamical system $(\Omega, \mathcal{B}, \nu, f)$, where Ω is the invariant set in some manifold, usually $\mathbb{R}^d$, $\mathcal{B}$ is the Borel σ-algebra, $f : \Omega \to \Omega$ is a measurable map and ν a probability f-invariant Borel measure. The stationary stochastic process given by the dynamical system will be of the form $X_m = g \circ f^m$, for any $m \in \mathbb{N}$, where the observable g has values in $\mathbb{R} \cup \pm\infty$ and achieves a global maximum at the point $\zeta \in \Omega$. We therefore study the partial maximum $M_m = \max\{X_0, \ldots, X_{m-1}\}$, in particular we look for normalising real sequences $\{a_m\}, \{b_m\}, m \in \mathbb{R}^+$ for which $\nu\{x; a_m(M_m - b_m) \le t\} = \nu\{x; M_m \le u_m\}$ converge to a non-degenerate distribution function; here $u_m = \frac{t}{a_m} + b_m$ is such that $m\nu(X_0 > u_m) \to \tau$, for some positive τ depending eventually on t: we defer to the book [Leadbetter et al., 1983] for a clear and complete picture of this approach. We will associate to our process a new i.i.d.

4

sequence $\tilde{X}_0, \cdots, \tilde{X}_{m-1}$ whose distribution is the same as that of X_0 and with partial maximum: $\tilde{M}_m = \max\{\tilde{X}_0, \cdots, \tilde{X}_{m-1}\}$. Properly normalized the distribution of such a maximum converges to one of the three laws in equations 1-3 and this is the interesting content of the Extreme Value Theory. Equations 1-3 will be satisfied by our original process too, whenever we would be able to prove that

$$\lim_{m\to\infty} \nu(\tilde{M}_m \leq u_m) = \lim_{m\to\infty} \nu(M_m \leq u_m)$$

This can be achieved if one can prove two sufficient conditions called D_2 and D' and which we briefly quote and explain in the footnote: these conditions basically require a sort of independence of the stochastic dynamical sequence in terms of uniform mixing condition on the distribution functions. In particular condition D_2, introduced in its actual form by Freitas-Freitas Freitas and Freitas [2008], could be checked directly by estimating the rate of decay of correlations for Hölder observables [1]. Another interesting issue of of the previous works was the choice of the observables g's: it is chosen as a function $g(\mathrm{dist}(x,\zeta))$ of the distance with respect to a given point ζ, with the aim that g achieves a global maximum at almost all points $\zeta \in \Omega$; for example $g(x) = -\log x$. In particular the observable g was taken in one of three different classes g_1, g_2, g_3, see Sect. 2 below, each one being again a function of the distance with respect to a given point ζ. The choice of these particular forms for the g's is just to fit with the necessary and sufficient condition on the tail of the distribution function $F(u) = \nu\{x; X_0 \leq u\}$, in order to exist a non-degenerate limit distribution for the partial maxima [Freitas et al., 2009], [Holland et al., 2008]. We use here the fact that, thanks to conditions D_2 and D', the distributions of X_0 rules out the distribution of our non-independent process X_m as well. It is important to remind that the previous conditions will determine the exponent ξ in the types 2 and 3 for $E(x)$.

Another major step in this field was achieved by establishing a connection between the extreme value laws and the statistics of first return and hitting times, see the papers by Freitas et al. [2009] and Freitas et al. [2010b]. They showed in particular that for dynamical systems preserving an absolutely continuous invariant measure or a singular continuous invariant measure ν, the existence of an exponential hitting time statistics on balls around ν almost any point ζ implies the existence of extreme value laws for one of the observables of type $g_i, i = 1, 2, 3$ described above. The converse is also true, namely if we have an extreme value law which applies to the observables of type $g_i, i = 1, 2, 3$ achieving a maximum at ζ, then we have exponential hitting time statistics to balls with center ζ. Recently these results have been generalized to local returns around balls centered at periodic points [Freitas et al., 2010a].

In the context of singular measures, the EVT has been developed in the recent paper by Freitas et al. [2010b]. The main goal of their paper was to establish a connection with hitting and return time statistics; for that purpose they considered returns in balls and also into cylinders. We are particularly interested in their Theorem 1, about balls, since it covers the class of observables considered in this paper; in particular we use here one direction of the theorem which allows us

---

[1] We briefly state here the two conditions, we defer to the next section for more details about the quantities introduced. If $X_m, m \geq 0$ is our stochastic process, we can define $M_{j,l} \equiv \{X_j, X_{j+1}, \cdots, X_{j+l}\}$ and we put $M_{0,m} = M_m$. The condition $D_2(u_m)$ holds for the sequence X_m if for any integer l, t, m we have $|\nu(X_0 > u_m, M_{t,l} \leq u_m) - \nu(X_0 > u_m)\nu(M_{t,l} \leq u_m)| \leq \gamma(m,t)$, where $\gamma(m,t)$ is non-increasing in t for each m and $m\gamma(m,t_m) \to 0$ as $m \to \infty$ for some sequence $t_m = o(m)$, $t_m \to \infty$.
We say condition $D'(u_m)$ holds for the sequence X_m if $\lim_{l\to\infty} \limsup_m m \sum_{j=1}^{[m/l]} \nu(X_0 > u_m, X_j > u_m) = 0$.

to get the extreme value distributions if the exponential return time statistics has been previously established for balls centered around almost all points and with respect to the probability invariant measure (in this manner we do not need to check conditions D_2 and D').

1.3 This work

In our previous work Faranda et al. [2011] we have shown that there exists an algorithmic way to study EVT by using a block-maxima approach for dynamical systems which possess an absolutely continuous invariant measure and satisfy the mixing properties given by conditions D_2 and D'. We have established the best conditions to observe convergence to the analytical results highlighting deviations from theoretical expected behavior depending on the number of maxima and number of block-observation. Furthermore, we have verified that the normalising process of variables can be applied a posteriori and a fit of unnormalised data produce a distribution that belongs to the Generalised Extreme Value (GEV) distributions family.

In this work we test our algorithm for maps that do not have an absolutely continuous invariant measure. We remind that in the context of dynamical system, the invariant measure plays the role of the probability measure on the space of events; in this respect the general theory of extremes will continue to apply no matter such a probability is absolutely continuous or singular with respect to Lebesgue. The interesting point is that for the choice of observables we did (the functions g_i), the cumulative distribution function F will be related to the scaling of the measure $\nu(B_r(z))$ of a ball $B_r(z)$ of radius r and centered at the point z, and such a scaling exponent turns out to be the Hausdorff dimension of the measure (also known as information dimension), when the point z is generically chosen. The experimental and accessible parameters of the GEV distributions will be explicitly expressed in terms of such a dimension.

In order to get the values of ξ and of a_m and b_m for finite m one should know how the measure of the ball $B_r(z)$ behaves as a function of r and of z and for measures which are not absolutely continuous. We notice that for absolutely continuous measure that approach works, at least in a few cases, and we quote our previous paper for that. Instead for singular continuous measures like those supported on Cantor sets, we are not aware of any analytic result allowing to get the few orders expansion of $\nu(B_r(z))$. This will prevent us to compute rigorously the normalising constant a_m for type 1 observables g_1; instead we will get the the limiting values of b_m for type 1 and the limiting values of a_m for type 2 and 3. Moreover we could not compute rigorously the exponent ξ. The values proposed for those non-rigorous constants are obtained by simply approximating $\nu(B_r(z))$ with r^D. The agreement with the numerical simulations suggests that there were good choices and suggests also a direct proof of the EVT for our observables and with the normalising constants indicated by our heuristic analysis.

As explained in the previous subsection we can either check the conditions D_2 and D' or the existance of an exponential return time statistics. The latter is the case of iterated function systems considered in section 3.2: these are in fact given by expanding maps (since they verify the so-called *open set condition*) and the exponential return time statistics for balls could be proved, for instance, using the technique in Bessis et al. [1987]. It will be also the case for the Hénon attractor with the parameters studied by Benedicks and Carlesson: for those parameters the attractor exists and carries an SRB measure; moreover very recently Chazottes and Collet established the Poissonian statistics for the number of visits in balls around generic points w.r.t. the SRB measure. Our

numerical computation will concern instead the *usual* Hénon attractor. Finally, we will consider the Lozi attractor, and in this case we will quote the result by Gupta et al. [2009], which proves the existence of the extreme value distributions for the observables constructed with the functions g_i and for balls around almost any point w.r.t. the SRB measure. As a final remark, we stress that the results by Freitas-Freitas and Todd have been proved under the assumption that $\nu(B_r(\zeta))$ is a continuous function of r: this is surely true for all the previous examples and such a condition will play a major role in our next considerations too.

This work is organized as follows: in Section 2 we present the analytical results for the EVT in maps with singular measures deriving the asymptotic behavior of normalising sequences and parameters. In Section 3 we present the numerical procedure used for the statistical inference of the GEV distribution and the numerical experiments that we have carried out for both singular measures generated with Iterated Function Systems and maps with a less trivial measures such as the Baker transformation, Hènon and Lozi maps. Eventually, in Section 4 we present our conclusion and proposal for future work.

2 Extreme Value Theory for maps with singular measures

2.1 Definitions and Remarks

Let us consider a dynamical systems $(\Omega, \mathcal{B}, \nu, f)$, where Ω is the invariant set in some manifold, usually $\mathbb{R}^d$, $\mathcal{B}$ is the Borel σ-algebra, $f : \Omega \to \Omega$ is a measurable map and ν a probability f-invariant Borel measure.

As we said in the Introduction and in order to adapt the extreme value theory to dynamical systems, we will consider the stationary stochastic process $X_0, X_1, ...$ given by:

$$X_m(x) = g(\text{dist}(f^m(x), \zeta)) \qquad \forall m \in \mathbb{N} \tag{8}$$

where 'dist' is a distance on the ambient space Ω, ζ is a given point and g is an observable function, and whose partial maximum is defined as:

$$M_m = \max\{X_0, ..., X_{m-1}\} \tag{9}$$

The probability measure will be here the invariant measure ν for the dynamical system. We will also suppose that our systems which verify the condition D_2 and D' which will allow us to use the EVT for i.i.d. sequences. As we said above, we will use three types of observables $g_i, i = 1, 2, 3$, suitable to obtain one of the three types of EV distribution for normalized maxima:

$$g_1(x) = -\log(\text{dist}(x, \zeta)) \tag{10}$$

$$g_2(x) = \text{dist}(x, \zeta)^{-1/\alpha} \tag{11}$$

$$g_3(x) = C - \text{dist}(x, \zeta)^{1/\alpha} \tag{12}$$

where C is a constant and $\alpha > 0 \in \mathbb{R}$.

These three type of functions are representative of broader classes which are defined, for instance,

throughout equations (1.11) to (1.13) in Freitas et al. [2009]; we now explain the reasons and the meaning of these choices. First of all these functions have in common the following properties: (i) they are defined on the positive semi-axis $[0, \infty]$ with values into $\mathbb{R} \cup \{+\infty\}$; (ii) 0 is a global maximum, possibly equal to $+\infty$; (iii) they are a strictly decreasing bijection in a neighborhood V of 0 with image W. Then we consider three types of behavior which generalize the previous specific choices:

Type 1: there is a strictly positive function $p : W \to \mathbb{R}$ such that $\forall y \in \mathbb{R}$ we have

$$\lim_{s \to g_1(0)} \frac{g_1^{-1}(s + yp(s))}{g_1^{-1}(s)} = e^{-y}$$

Type 2: $g_2(0) = +\infty$ and there exists $\beta > 0$ such that $\forall y > 0$ we have

$$\lim_{s \to \infty} \frac{g_2^{-1}(sy)}{g_2^{-1}(s)} = y^{-\beta}$$

Type 3: $g_3(0) = D < +\infty$ and there exists $\gamma > 0$ such that $\forall y > 0$ we have

$$\lim_{s \to 0} \frac{g_3^{-1}(D - sy)}{g_3^{-1}(D - s)} = y^{\gamma}$$

The Gnedenko corollary says that the different kinds of extreme value laws are determined by the distribution of

$$F(u) = \nu(X_0 \le u) \tag{13}$$

and by the right endpoint of F, x_F.

We need to compute and to control the measure $\nu(B_r(\zeta))$ of a ball of radius r around the point ζ. At this regard we will invoke, and assume, the existence of the following limit

$$\lim_{r \to 0} \frac{\log \nu(B_r(\zeta))}{\log r}, \text{ for } \zeta \text{ chosen } \nu - \text{a.e.} \tag{14}$$

Moreover we will assume that $\nu(B_r(\zeta))$ is a continuous function of r (see Freitas et al. [2009] for a discussion of this condition which shows that all the examples considered in our paper will fit it). When the limit (14) exists on a metric space equipped with the Borel σ-algebra and a probability measure ν, it gives the *Hausdorff dimension of the measure* or *information dimension*, defined as the infimum of the Hausdorff dimension taken over all the set of ν measure 1 [Young, 1982]. This limit could be proved to exist for a large class of dynamical systems and whenever ν is an invariant measure: let us indicate it with Δ without written explicitly its dependence on ν. For example, for a very general class of one-dimensional maps with positive metric entropy, Δ is equal to the ratio between the metric entropy and the (positive) Lyapunov exponent of ν [Ledrappier, 1981]. For two dimensional hyperbolic diffeomorphisms, Δ is equal to the product of the metric entropy times the difference of the reciprocal of the positive and of the negative Lyapunov exponents [Young, 1982]. The information dimension is a lower bound of the Hausdorff dimension of the support of the measure ν and it is an upper bound of the correlation dimension [Yakov, 1998], [Hentschel and Procaccia, 1983], [Grassberger, 1983], [Bessis et al., 1988], [Bessis et al., 1987], [Cutler and Dawson, 1989].

2.2 Limiting behavior of the Extreme Value Theory parameters

We summarize the three basic assumptions for the next considerations:

- Assumption 1: our dynamical system verifies conditions (D_2) and D'.

- Assumption 2: the measure of a ball is a continuous function of the radius for almost all the center points; moreover such a measure has no atoms.

- Assumption 3: the limit (14) exists (and its value is called Δ) at almost all points ζ.

Equipped with these conditions it is now possible to compute rigorously a few of the expected parameters for the three types of observables.

Case 1: $g_1(x) = -\log(\text{dist}(x,\zeta))$. Substituting equation 8 into equation 13 we obtain that:

$$
\begin{aligned}
1 - F(u) &= 1 - \nu(g(\text{dist}(x,\zeta)) \le u) \\
&= 1 - \nu(-\log(\text{dist}(x,\zeta)) \le u) \\
&= \nu(\text{dist}(x,\zeta) < e^{-u}) = \nu(B_{e^{-u}}(\zeta))
\end{aligned}
\tag{15}
$$

$$
x_F = \sup\{u; F(u) < 1\}
$$

To use Gnedenko corollary it is necessary to calculate x_F; in this case $x_F = +\infty$ as we will explain in the proof below.

According to Corollary 1.6.3 in Leadbetter et al. [1983] for type 1 $a_m = [G(\gamma_m)]^{-1}$ and $b_m = \gamma_m = F^{-1}(1 - \frac{1}{m})$. We now show how to get the limiting value of γ_m; a similar proof will hold for type II and III.

Proposition 1. *Let us suppose that our system verifies Assumptions 1,2,3 above and let us consider the observable g_1; then:*

$$
\lim_{m \to \infty} \frac{\log m}{\gamma_m} = \Delta
$$

Proof
By our choice of the observable we have: $1 - F(\gamma_m) = \nu(B_{e^{-\gamma_m}}(\zeta)) = \frac{1}{m}$; since the measure is not atomic and it varies continuously with the radius, we have necessarily that $\gamma_m \to \infty$ when $m \to \infty$. Now we fix $\delta > 0$ and small enough; there will be $m_{\delta,\zeta}$ depending on δ and on ζ, such that for any $m \ge m_{\delta,\zeta}$ we have

$$
-\delta\gamma_m \le \log \nu(B_{e^{-\gamma_m}}(\zeta)) + \Delta\gamma_m \le \delta\gamma_m
\tag{16}
$$

Since $\log m - \Delta\gamma_m = -[\log \nu(B_{e^{-\gamma_m}}(\zeta)) + \Delta\gamma_m]$ and by using the bounds (16) we immediately have

$$
-\delta\gamma_m \le \log m - \Delta\gamma_m \le \delta\gamma_m
$$

which proves the Proposition.

It should be clear that the previous proposition will not give us the value of γ_m and of b_m, which is equal to γ_m for type I observables. We have instead a rigourous limiting behavior and we will pose in the following:

$$\gamma_m = b_m \sim \frac{1}{\Delta} \log m$$

The values for finite m could be obtained if one would dispose of the functional dependence of $\nu(B_r(\zeta))$ on the radius r and the center ζ: this has been achieved for non-trivial absolutely continuous invariant measure in our previous paper [Faranda et al., 2011]. The same reason prevent us to get a rigorous limiting behavior for $a_m = [G(\gamma_m)]^{-1}$. The only rigorous statement we can do is that $G(\gamma_m) = o(\gamma_m)$; this follows by adapting the previous proof of the proposition to another result (see Leadbetter et al. [1983]) which says that for type I observables one has $\lim_{m\to\infty} n(1 - F\{\gamma_m + xG(\gamma_m)\}) = e^{-x}$, for all real x: choosing $x = 1$ gives us the previous domination result. In the following and again for numerical purposes we will take

$$a_m = [G(\gamma_m)]^{-1} \sim \frac{1}{\Delta}$$

This follows easily by replacing in formula (6) $\nu(B_r(\zeta)) \sim r^\Delta$ for r small. We finish this part by stressing that for our observable we expect $\xi = \xi' = 0$.

Case 2: $g_2(x) = \text{dist}(x,\zeta)^{-1/\alpha}$.   In this case we have

$$
\begin{aligned}
1 - F(u) &= 1 - \nu(\text{dist}(x,\zeta)^{-1/\alpha} \le u) \\
&= 1 - \nu(\text{dist}(x,\zeta) \ge u^{-\alpha}) \\
&= \nu(B_{u^{-\alpha}}(\zeta))
\end{aligned}
\tag{17}
$$

and $x_F = +\infty$. Since $b_m = 0$ we have only to compute a_m which is the reciprocal of γ_m which is in turn defined by $\gamma_m = F^{-1}(1 - 1/m)$. By adapting Proposition 1 we immediately get that

$$\lim_{m\to\infty} \frac{\log m}{\log \gamma_m} = \alpha\Delta$$

which we allow us to use the approximation $a_m \sim \frac{1}{m^{\frac{1}{\alpha\Delta}}}$ The exponent ξ for Type II observables is given by the following limit (see [Leadbetter et al., 1983], Th. 1.6.2)

$$\lim_{t\to\infty} (1 - F(tx))/(1 - F(t)) = x^{-\xi}, \ \xi > 0, \ x > 0$$

The crude approximation $\nu(B_r(\zeta)) \sim r^\Delta$ for r small, will give immediately that $\xi \sim \alpha\Delta$ and this value will appear in the exponent $\xi' = 1/\xi$ in the distribution function given by the GEV.

Case 3: $g_3(x) = C - \text{dist}(x,\zeta)^{1/\alpha}$.   We have first of all:

$$
\begin{aligned}
1 - F(u) &= 1 - \nu(C - \text{dist}(x,\zeta)^{1/\alpha} \le u) \\
&= \nu(B_{(C-u)^\alpha}(\zeta))
\end{aligned}
\tag{18}
$$

In this case $x_F = C < \infty$ and $a_m = (C - \gamma_m)^{-1}; b_m = C$. The previous proposition immediately shows that $\lim_{n \to \infty} \frac{\log m}{-\alpha \log(C - \gamma_m)} = \Delta$ which gives the asymptotic scaling $\gamma_m \sim C - \frac{1}{m^{\frac{1}{\alpha\Delta}}}; a_m \sim m^{\frac{1}{\alpha\Delta}};$ $b_m = C$. Finally the exponent ξ is given again by Th. 1.6.2 in Leadbetter et al. [1983] by the formula

$$\lim_{h \to 0}(1 - F(C - hx))/(1 - F(C - h)) = x^\xi, \ \xi > 0, \ x > 0$$

which with our usual approximation furnishes $\xi \sim \alpha\Delta$.

3 Numerical Experiments

3.1 Procedure for statistical inference

For the numerical experiments we have used a wide class of maps that have singular measure, also considering the case of strange attractors such as the ones observed by iterating Lozi Map or Hènon map. The algorithm used is the same described in Faranda et al. [2011]: for each map we run a long simulation up to k iterations starting from a given initial condition. From the trajectory we compute the sequence of observables g_1, g_2, g_3 dividing it into n bins each containing $m = k/n$ observations and eventually obtaining the empirical cdf of maxima.

In Faranda et al. [2011] we have used a Maximum Likelihood Estimation (MLE) procedure working both on pdf and cdf (cumulative distribution function), since our distributions were absolutely continuous and the minimization procedure was well defined. In this case, we don't have anymore the pdf and consequently the fitting procedure via MLE could give us wrong results. To avoid these problems we have used an L-moments estimation as detailed in Hosking [1990]. This procedure is completely discrete and can be used both for absolutely continuous or singular continuous cdf. The L-moments are summary statistics for probability distributions and data samples. They are analogous to ordinary moments which meant that they provide measures of location, dispersion, skewness, kurtosis, but are computed from linear combinations of the data values, arranged in increasing order (hence the prefix L). Asymptotic approximations to sampling distributions are better for L-moments than for ordinary moments [Hosking [1990], Figure 4]. The relationship between the moments and the parameters of the GEV distribution are described in Hosking [1990], while the 95% confidence intervals has been derived using a bootstrap procedure. As comparison we have checked that the results presented in Faranda et al. [2011] are comparable with L-moments methods. We have found that both the methods give similar results even if L-moments has an uncertainty on the estimation of parameters generally slightly bigger.

The empirical cdf contains plateaux which correspond to non accessible distances in correspondence of the holes of the Cantor set. The discrete nature of L-moments allows to overcome difficulties that may arise in singular continuous cdf: the normalization procedure carried out with this method consists in dividing each quantity computed via L-moments by a function of the total number of data will prevent us from obtain a unnormalised distribution. The last issue we want to address is the choice of a suitable model for our data: in principle, using L-moments procedure we can fit the data to any kind of known cdf. To validate the use of the GEV model we proceeded in the following way:

-*A priori* the choice of a GEV model arises naturally if the assumptions presented in the section 3 are satisfied. In this set up we can directly compute the parameters of GEV distribution using

L-moments as described in Hosking [1990].

-*A posteriori* we can verify the goodness of fit to GEV family if we apply some parametric or non-parametric tests commonly used in statistical inference procedures. For this purpose we have fitted our experimental data to a wide class of well known continuous distributions. Using Kolmogorov Smirnov test (see Lilliefors [1967] for a description of the test) we have measured the deviation between the empirical cdf and the fitted cdf, finding that using the GEV distribution we effectively achieve a minimization of the deviation parameter.

We summarize below the results we expect from numerical experiments in respect to n according to the conjecture described in the previous section. Since we keep the length of the series $k = n \cdot m$ fixed, the following relationships can be obtained simply replacing $m = k/n$ in the equations derived in the previous section.

For g_1 type observable:

$$\sigma = \frac{1}{\Delta} \qquad \mu \sim \frac{1}{\Delta} \ln(k/n) \qquad \xi' = 0 \tag{19}$$

For g_2 type observable, we can either choose $b_m = 0$ or $b_m = c \cdot m^{-\xi'}$ where $c \in \mathbb{R}$ is positive constant, as detailed in Beirlant [2004]. A priori, we do not know which asymptotic sequences will correspond to the parameters μ in the experimental set up. The experimental procedure we use automatically select $b_m = c \cdot m^{-\xi'}$, therefore the following results are presented taking into account this asymptotic sequence:

$$\sigma \sim n^{-1/(\alpha\Delta)} \qquad \mu \sim n^{-1/(\alpha\Delta)} \qquad \xi' = \frac{1}{\alpha\Delta} \tag{20}$$

For g_3 type observable:

$$\sigma \sim n^{1/(\alpha\Delta)} \qquad \mu = C \qquad \xi' = \frac{1}{\alpha\Delta} \tag{21}$$

3.2   IFS for Cantor Sets

A Cantor set can be obtained as an attractor of some Iterated Function Systems (IFS). An IFS is a finite family of contractive maps $\{f_1, f_2, ..., f_s\}$ acting on a compact normed space Ω with norm $|\cdot|$ and possessing a unique compact limit set (the attractor) $K \in \Omega$ which is non-empty and invariant by the IFS, namely:

$$K = \bigcup_{i=1}^{s} f_i(K).$$

We will put a few restrictions on the IFS in order to see it as the *inverse* of a genuine dynamical system; we will explain in a moment why this change of perspective will help us to compute observables on fractal sets. We defer to the fundamental paper by Barnsley and Demko [1985], for all the material we are going to use.

First of all we will consider the f_i as strict contractions, namely there will be a number $0 < \lambda < 1$ such that for all $i = 1, \cdots, s$ we have $|f_i(x) - f_i(y)| < \lambda|x - y|$, for all $x, y \in X$.

Then we will suppose that each f_i is one-to-one on the attractor K and moreover $\forall i = 1, \cdots, s$

we have $f_i(K) \cap f_j(K) = \emptyset$, $i \neq j$ (*open set condition*). This will allow us to define a measurable map $T : K \to K$ by $T(x) = f_i^{-1}(x)$ for $x \in f_i(K)$: the attractor K will be the invariant set for the transformation T which will play therefore the role of a usual dynamical system. A complete statistical description of a dynamical system is given by endowing it with an invariant probability measure; in particular we ask that this measure by ergodic if we want to compute the maxima of the sequence of events constructed with the observables g_i; we remind that these events are nothing but the evaluations along the forward orbit of an initial point chosen according to the measure. If, as always happens, the attractor K has a fractal structure and zero Lebesgue measure, we could not get such an initial point for numerical purposes. We overcome this situation for attractors of global diffeomorphisms by taking the initial point in the basin of attraction and by iterating it: the orbit will be distributed according to the SRB measure (we will return later on this measure). For our actual attractors generated by non-invertible maps, the iteration of any point in the complement of the attractor will push the point far from it: it would be better to call repellers our invariant sets instead of attractors. We have therefore to proceed in a different manner. The measures supported on the attractor K will give the solution. First of all let us associated to each map f_i a positive weight p_i in such a way that $\sum_{i=1}^{s} p_i = 1$. Then it is possible to prove the existence of a unique measure ν (called *balanced*) which enjoy the following properties:

- The measure ν is supported on the attractor K and it will be invariant for the map T associated to our IFS (see above).

- For any measurable set B in X we have

$$\nu(B) = \sum_{i=1}^{s} p_i \nu(f_i^{-1}(B))$$

- Let us put $(Sg)(x) = \sum_{i=1}^{s} p_i g(f_i(x))$, for a continuous functions g on Ω and for *any* point $x \in \Omega$, then we have
$$\lim_{n \to \infty} S^n g(x) = \int_{\Omega} g d\nu$$

This last item is very important for us; first of all it holds also for the characteristic function of a set provided the boundary of this set has ν measure zero; therefore it is a sort of ergodic theorem because it states that the backward orbit constructed by applying to any point in X the maps f_i with weights p_i will distribute on the attractor K as the forward orbit (namely the orbit generated by the transformation T associated to the IFS) of a point $y \in K$ and chosen almost everywhere according to ν.

Let us give our first example.

3.2.1 Uniform weights and the Sierpisnkij triangle

We consider the middle one third Cantor set that is the attractor of the IFS $\{f_1, f_2\}$ defined as:

$$\begin{cases} f_1(x) = x/3 \text{ with weight } p_1 \\ f_2(x) = (x+2)/3 \text{ with weight } p_2 \end{cases} \tag{22}$$

where $x \in [0, 1]$ and we set $p_1 = p_2 = 1/2$ so that, at each time step, we have the same probability to iterate $f_1(x)$ or $f_2(x)$:

Equivalently the previous IFS can be written as:

$$x_{t+1} = (x_t + b)/3 \tag{23}$$

where, at each time step, we extract randomly with equal probability b to be 0 or 2.

We will consider also the so called Sierpisnkij triangle, defined by

$$\begin{cases} x_{t+1} = (x_t + v_{p,1})/2 \\ y_{t+1} = (y_t + v_{p,2})/2 \end{cases} \tag{24}$$

We extract randomly at each time step and with equal probability the number p to be the integer 1,2 or 3. Then we iterate the map 24 substituting the elements $v_{p,1}$ and $v_{p,2}$ of the following matrix:

$$v = \begin{vmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \end{vmatrix}.$$

For these attracting sets the information dimensions are well known, they are $\Delta = \log(2)/\log(3)$ for the Cantor set and $\Delta = \log(3)/\log(2)$ for the Sierpinskij triangle [Sprott, 2003].

In the following experiments In order to choose the centers ζ of our balls we proceed by using again the backward preimages of any point in Ω, namely we take a point $x \in \Omega$ and we consider ζ as one of the preimages $f^{-t}(x)$ with t much larger than the sequence of observed events; by what we said above, that preimage will be closer and closer to the invariant Cantor set and also it will approaches a generic point with respect to the balanced measure ν.

First of all we have analised the empirical cdf $F(u)$ of the extrema for g_1 observable. An example is shown for the IFS in eq. 22 in Figure 1. The histogram is obtained iterating the map in equation 22 for $5 \cdot 10^7$ iterations, $\zeta \simeq 0.775$, $\alpha = 4$, $C = 10$. Once computed the functions g_1 the series of maxima for each observable is computed taking each of them in bins containing 5000 values of g_1 for a total of 1000 maxima. As claimed in the previous section, the cdf is a singular continuous function and this is due to the structure of the Cantor set. The results are similar for the other observables and other initial conditions.
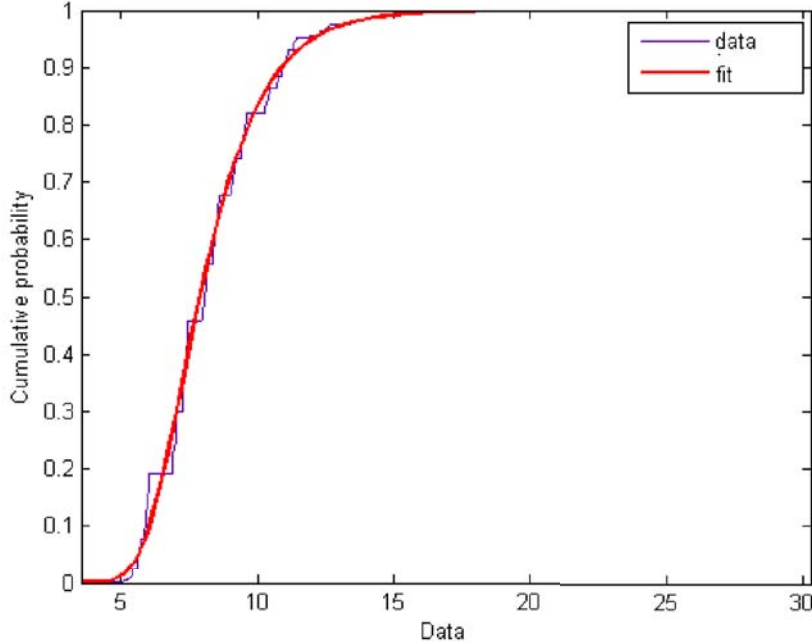
Figure 1: Empirical (blue) and fitted (red) cdf for IFS that generates a Cantor Set, $\zeta = 0.775$, g_1 observable function.

To check that effectively the parameters of GEV distribution obtained by L-moments estimation are related to the fractal dimension of the attracting Cantor Set and the Sierpinskij triangle, we have considered an ensemble of 10^4 different realizations of the eq. 22 and eq. 24, starting from the same initial conditions. To check the behavior we have varied n and m keeping fixed the length of the series $k = 10^7$. In Faranda et al. [2011] we have shown that a good convergence is observed when $n, m > 1000$, therefore we will make all the considerations for (n, m) pairs that satisfy this condition.

In figures 2-4 the results of the computation for the IFS that generates the Cantor Set (plots on the left) and the Sierpisnkij triangle (plots on the right) are presented. In all the cases considered the behavior well reproduce the theoretical expected trend described in equations 19-21. The initial condition here shown is $\zeta = 1/3$ for Cantor Set and $\zeta \simeq (0.02, 0.40)$ for Sierpinskij triangle, but similar results hold for different initial conditions if chosen on the attractive sets. The black line is the mean value over different realizations of the map, while the black dotted lines represent one standard deviation.

For g_1 observable function, according to equation 19, we expect to find $\xi = 0$ in both cases and this is verified by experimental data shown in Figure 2a). For the scale parameter a similar agreement is achieved in respect to the theoretical parameters $\Delta = \frac{1}{\sigma(g_1)} = \log(2)/\log(3) \simeq 0.6309$

$\Delta = 1/(\alpha|\kappa|)$	Cantor	Sierpinskij
Theoretical	$\log(2)/\log(3) \simeq 0.6309$	$\log(3)/\log(2) \simeq 1.5850$
$\mu(g_2)$	0.636 ± 0.006	$|\kappa| = 1.592 \pm 0.007$
$\sigma(g_2)$	0.634 ± 0.007	$|\kappa| = 1.56 \pm 0.02$
$\sigma(g_3)$	0.64 ± 0.01	$|\kappa| = 1.62 \pm 0.01$

Table 1: Information dimension Δ computed taking the logarithm of equations 20-21 and computing the angular coefficient κ of a linear fit of data; IFS with uniform weights and Sierpinskij triangle, $\alpha = 4$, $C = 10$.

for the Cantor Set and $\Delta = \frac{1}{\sigma(g_1)} = \log(3)/\log(2) \simeq 1.5850$ for Sierpinskij shown in figure 2b) with a green line. Eventually, the location parameter μ shows a logarithm decay with n as expected from equation 19. A linear fit of μ in respect to $\log(n)$ is shown with a red line in figure 2c). The linear fit angular coefficient κ of equation 19 satisfies $\Delta = 1/\kappa$ and the dimension computed from data using this relation is $\Delta = 0.64 \pm 0.01$ for Cantor and $\Delta = 1.59 \pm 0.01$ in the Sierpinskij triangle.

The agreement between the conjecture and the results are confirmed even for g_2 type and g_3 type observable functions shown in figures 3 and 4 respectively. In this case we have experienced some problems in the convergence for the Cantor map has using $\alpha = 2$ and $\alpha = 3$. The problem is possibly due to the fact that L-moments method works better if $\xi' \in [-0.5, 0.5]$ while for $\alpha \leq 3$ the shape parameter $|\xi'| > 0.5$ for the Cantor map. For this reason results are shown using $\alpha = 4$ both for Sierpinskij triangle and Cantor IFS, for all the experiments presented the constant value in g_3 will be $C = 10$. For both observables g_2, g_3 there is strong agreement between the experimental and theoretical ξ' values. In figures 3b), 3c), 4b) a log-log scale is used to highlight the behavior described by eq. 20 and 21. We can check again the value of the dimension using the angular coefficient κ which satisfies $\Delta = 1/(\alpha|\kappa|)$. In Table 1 we compare these results with the theoretical values. The error is here represented as one standard deviation of the ensemble of realizations and we find a good agreement between theoretical and experimental parameters within two standard deviations. Eventually, computing g_3 observable function we expect to find a constant value for $\mu(g_3) = C = 10$ while $\sigma(g_3)$ has to grow with a power law in respect to n as expected comparing with equation 21.

Other tests have been done computing the statistics using parameter $\alpha = 5, 6, 7, 8$ for g_2 and g_3 observables. Also in this cases, no deviation from the behavior described in eq. 19-21 has been found.
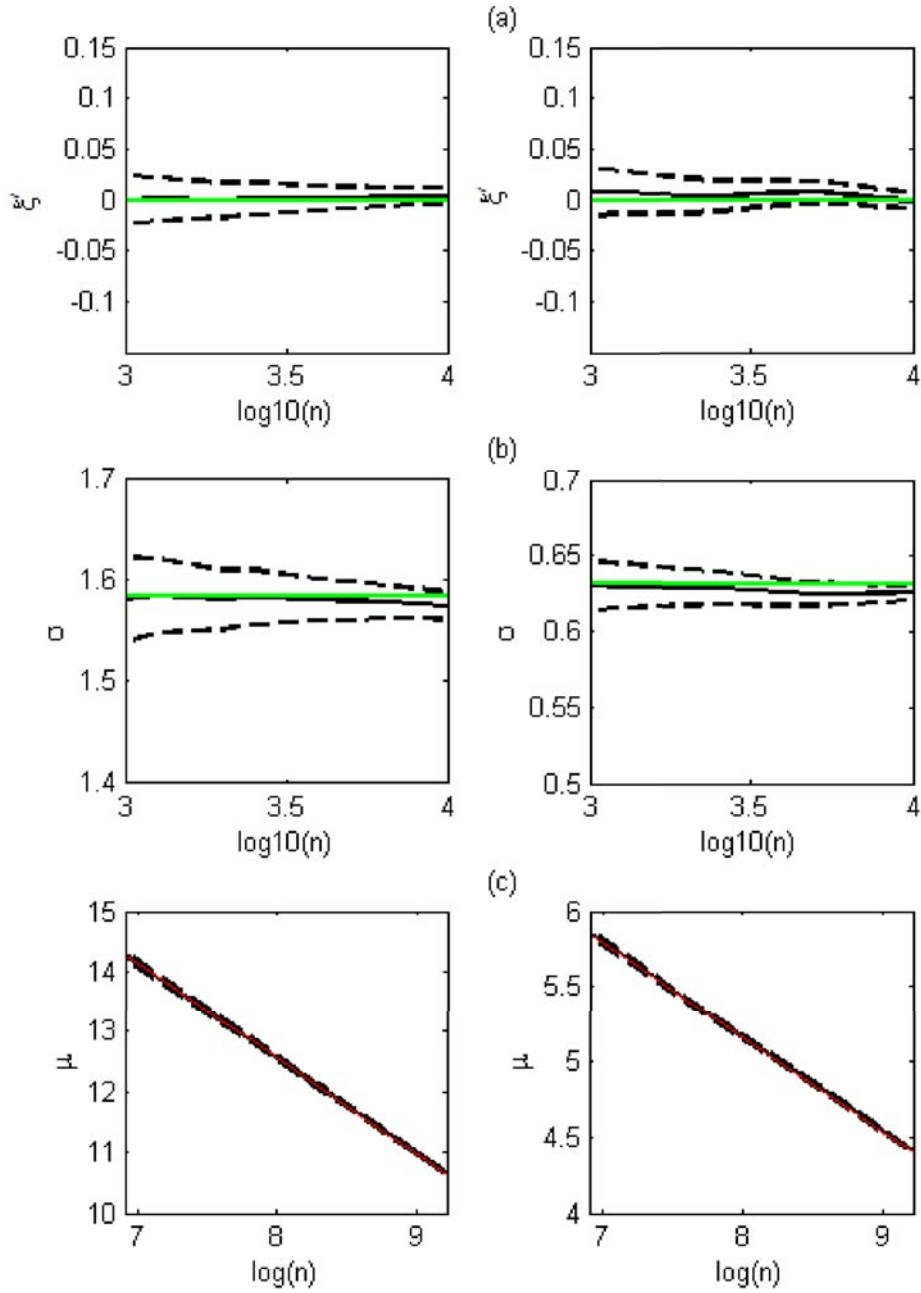
Figure 2: g_1 observable. **a)** ξ' VS $\log_{10}(n)$; **b)** σ VS $\log_{10}(n)$; **c)** μ VS $\log(n)$. Cantor set, Right: Sierpinskij triangle. Dotted lines represent one standard deviation, red lines represent a linear fit, green lines are theoretical values.
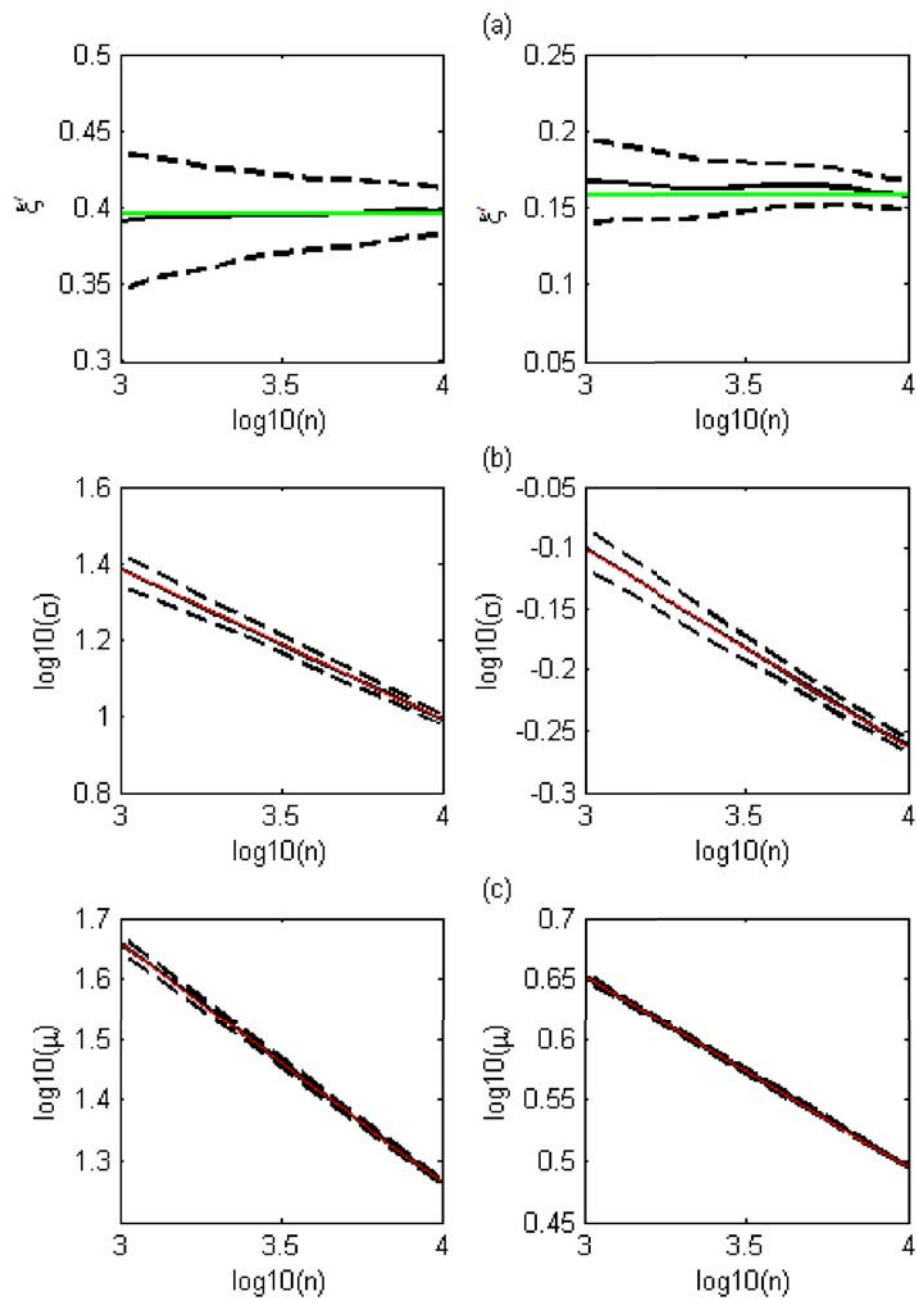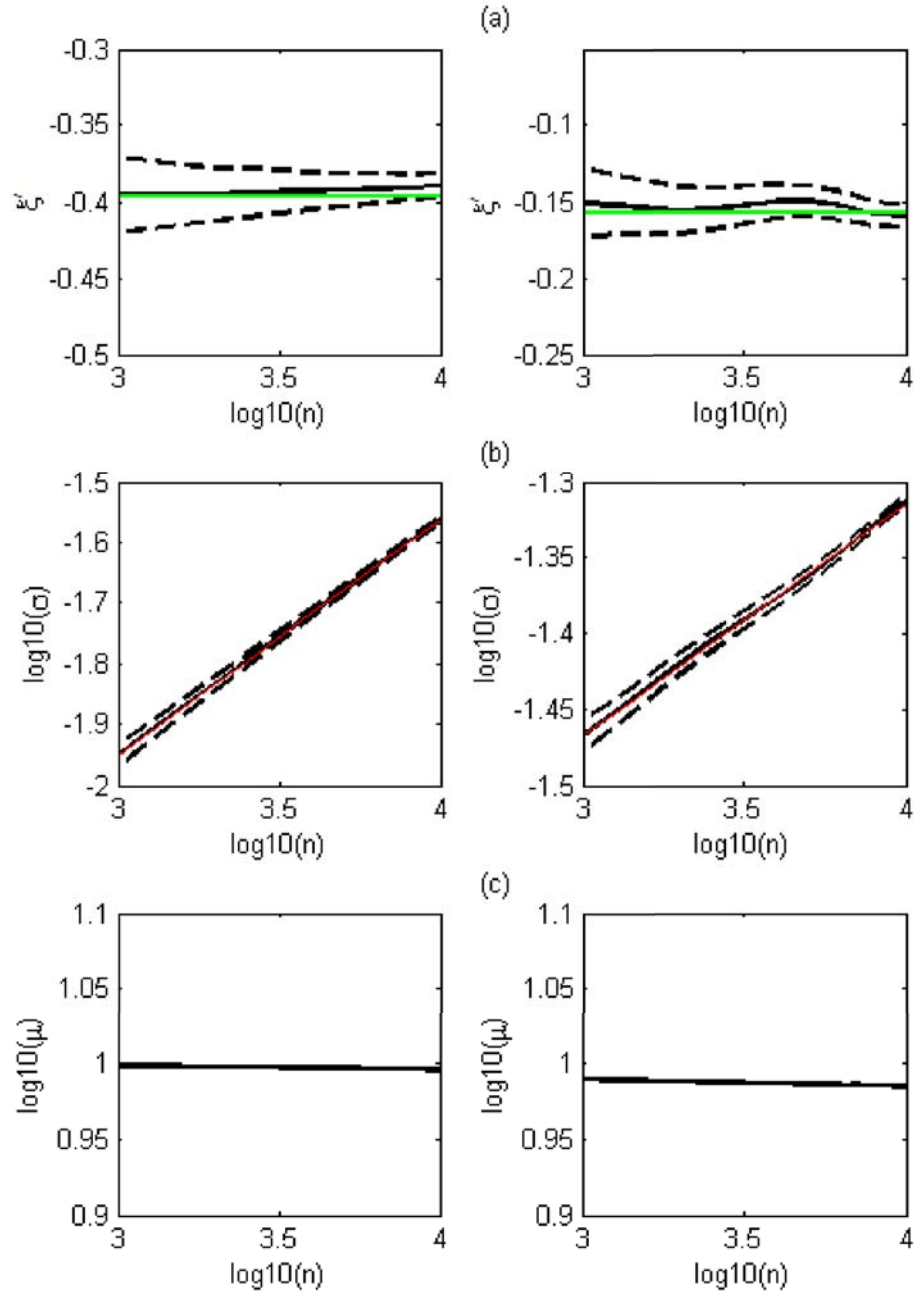
17

Figure 3: g_2 observable **a)** ξ' VS $\log_{10}(n)$; **b)** $\log_{10}(\sigma)$ VS $\log_{10}(n)$; **c)** $\log_{10}(\mu)$ VS $\log_{10}(n)$. Left: Cantor set, Right: Sierpinskij triangle. Dotted lines represent one standard deviation, red lines represent a linear fit, green lines are theoretical values.

Figure 4: g_3 observable. **a)** ξ' VS $\log_{10}(n)$; **b)** $\log_{10}(\sigma)$ VS $\log_{10}(n)$; **c)** $\log_{10}(\mu)$ VS $\log_{10}(n)$. Cantor set, Right: Sierpinskij triangle. Dotted lines represent one standard deviation, red lines represent a linear fit, green lines are theoretical values.   19

3.2.2 IFS with non-uniform weights

Let us now consider the case of an IFS with different weights:

$$f_k(x) = a_k + \lambda_k x \qquad x \in [0,1] \qquad k = 1, 2, ..., s \tag{25}$$

and each f_i is iterated with (different) probability w_i.

In this case it is possible to compute the information dimension as the ratio between the metric entropy and the Lyapunov exponent of the associated balanced measure [Barnsley, 2000]: we get the following expression:

$$\Delta = \frac{w_1 \log w1 + ... + w_s \log w_s}{w_1 \log \lambda_1 + ... + w_s \log \lambda_s} \tag{26}$$

In this analysis we have considered the following IFS:

$$\begin{cases} f_1(x) = x/3 & \text{with weight } w \\ f_2(x) = (x+2)/3 & \text{with weight } 1 - w \end{cases} \tag{27}$$

and we have changed the weight w between 0.35 and 0.65 with 0.01 step. For $w = 0.5$ we obtain the same results shown in the previous section, while for different weights we can check the expression 26.

The presence of different weights makes the convergence process sensible to the choice of the sample point ζ where our observable reaches its maximum. For that reason we took several different values of ζ in order to obtain a reliable estimations of the information that should be obtained, in this case, as an average property.

In Figure 5 we present the dimension Δ computed using relationship 19-21. In particular we can compute the dimension from eq. 19 as:

$$\Delta(\sigma(g_1)) = \frac{1}{<\sigma(g_1)>} \tag{28}$$

We can infer dimension also from eq. 20, eq. 21 as:

$$\Delta(\xi'(g_i)) = \frac{1}{\alpha| <\xi'(g_i)> |}, \quad i = 2, 3 \tag{29}$$

and in all expressions above the brackets $< . >$ indicate an average on different sample points ζ. For the rest of the numerical computations we set $\alpha = 5$. The parameters have been computed using 1000 different initial conditions on the support of the attractor, and for 30 realizations of each sample point ζ, the block-maxima approach is here used with $n = m = 1000$. The error bar are computed using the standard error propagation rules.

The agreement between the theoretical dimension and the experimental data is evident for all the weights and for all the observable considered. The uncertainty increases when w is much different from 0.5. This is due to the fact that as soon as we change the weight to be different from 0.5 the parameters spread increase to take in account the local properties of the attractor. The best agreement and less uncertainty is achieved considering the dimension as computed from $\sigma(g_1)$

observable. This is possibly due to the slower convergence for g_2 or g_3 observables to the respective theoretical distributions: in g_1 we modulate the distances with a logarithm function while in g_2 and in g_3 power laws are used. Nevertheless, all the data show the right trend.
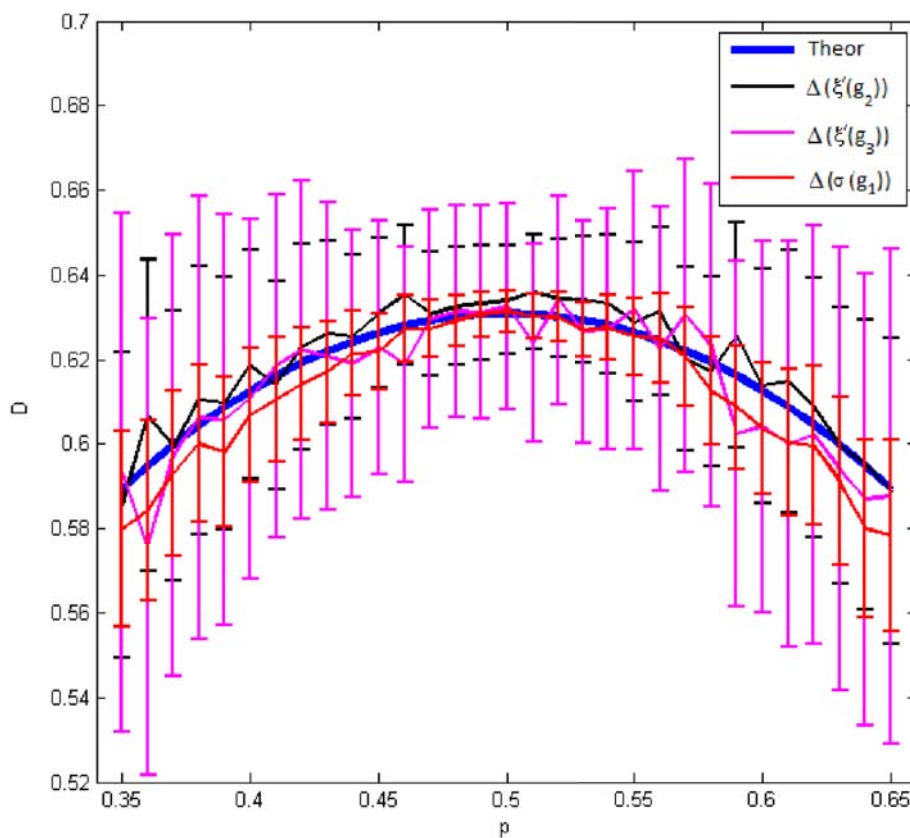

Figure 5: Different estimation of Δ dimension obtained using Extreme Value distribution for the IFS in equation 27

3.3   Non-trivial singular measures

In the previous subsection we have analysed the relatively simple cases of Cantor sets generated with IFS. In order to provide further support to our conjectures, we now present some application of our theory to the output of dynamical systems possessing a less trivial singular measures. We consider three relevant examples of two dimensional maps.

The Baker map

The Baker map is defined as follows:

$$x_{t+1} = \begin{cases} \gamma_a x_t \mod 1 & \text{if } y_t < \alpha \\ 1/2 + \gamma_b x_t \mod 1 & \text{if } y_t \geq \alpha \end{cases}$$

$$y_{t+1} = \begin{cases} \frac{y_t}{\alpha} \mod 1 & \text{if } y_t < \alpha \\ \frac{y_t - \alpha}{1-\alpha} \mod 1 & \text{if } y_t \geq \alpha \end{cases} \tag{30}$$

we consider the classical value for the parameter: $\alpha = 1/3, \gamma_a = 1/5$ and $\gamma_b = 1/4$.

Rigorous analytical results are available for the estimation of the information dimension [Kaplan and Yorke, 1979]. For our parameter values, the analytical expected value is $D \simeq 1.4357$.

We have performed the same analysis detailed in section 4.1, but with a difference. This map is invertible and its invariant set is an attractor given by the cartesian product of a segment, along the y-axis, and a one dimensional Cantor set along the x-axis. The system possess an invariant SRB measure, which can practically be constructed by taking ergodic sums for any point sitting on the basin of attraction. In order to compute the center of the balls on the attractor, we proceed in a similar manner as for repellers (see above), namely we take any point x in the basin of attraction and we iterate it t times with t much bigger than the sequence of observed events. Then we take ζ as the point $f^t(x)$: it will be closer and closer to the attractor and distributed according to the SRB measure. In our set up: $\alpha = 4$ for g_2 and g_3, $C = 10$. The results are shown in figures 6-8 the black continuous lines will represent the parameter average over different initial conditions and the black dotted lines the standard deviation of the distribution of the estimated parameters.

The expected theoretical ξ' values are within one standard deviation of the results of the fit for all three observable. The agreement seems to be better when we increase n even if this correspond to a decrease of m in our set up. This behavior is quite interesting since it seems that we obtain a much better convergence to theoretical values if $n \simeq 10^4$, while in all other example there is no such a difference between $n = 1000$ and $m = 10000$. A similar consideration can be made for $\sigma(g_1)$ shown in Figure 6b) that approaches the theoretical values value for bigger n values. The angular coefficient of the linear fit for $\mu(g_1)$ shown in the semilog plot in Figure 6c) allow us to estimate the dimension $\Delta = 1/|\kappa| = 1.48 \pm 0.03$ that is consistent with the theoretical values within two standard deviations .

Log-log plots of the parameters against n are shown in figures 7b), 7c) and 8b), and the value of the dimension Δ computed using the angular coefficients are reported in Table 2. The agreement with expected value is good enough for all the parameters and better for $\mu(g_2)$. Eventually, Figure 8c) shows that $\mu(g_3)$ approaches $C = 10$.
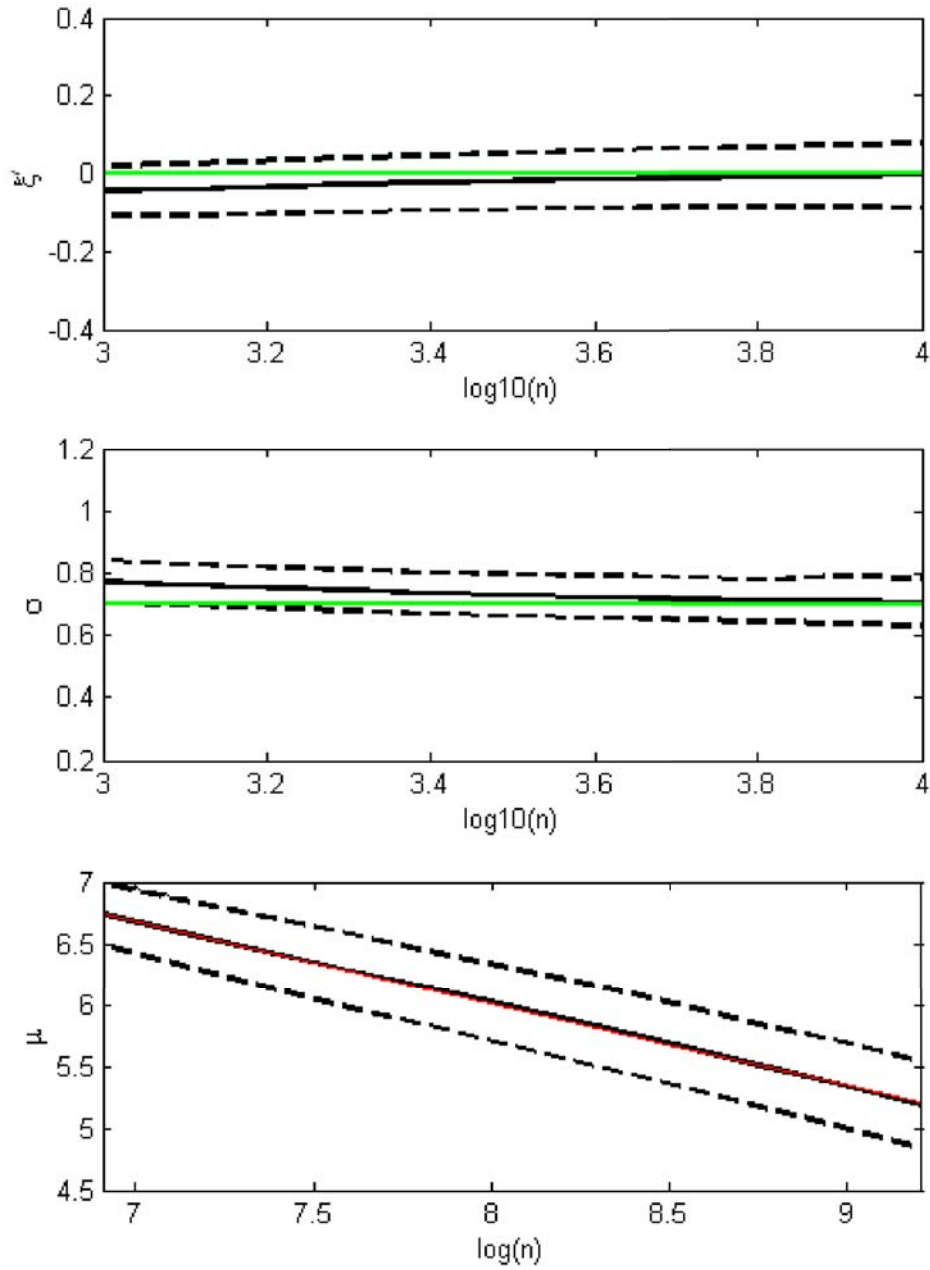
Figure 6: g_1 observable. **a)** ξ' VS $\log_{10}(n)$; **b)** σ VS $\log_{10}(n)$; **c)** μ VS $\log(n)$. Baker map. Dotted lines represent one standard deviation, red lines represent a linear fit, green lines are theoretical values.
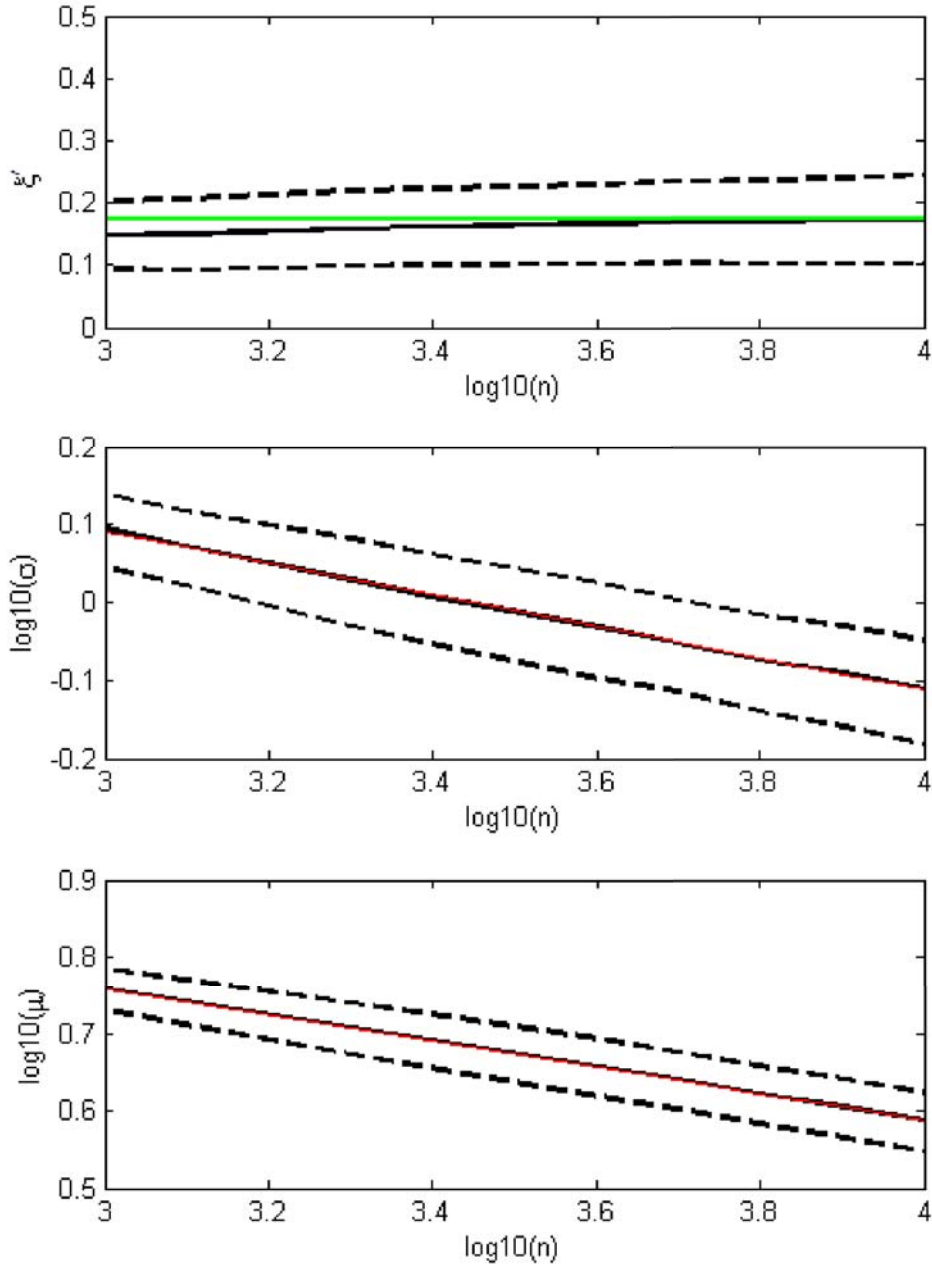
23

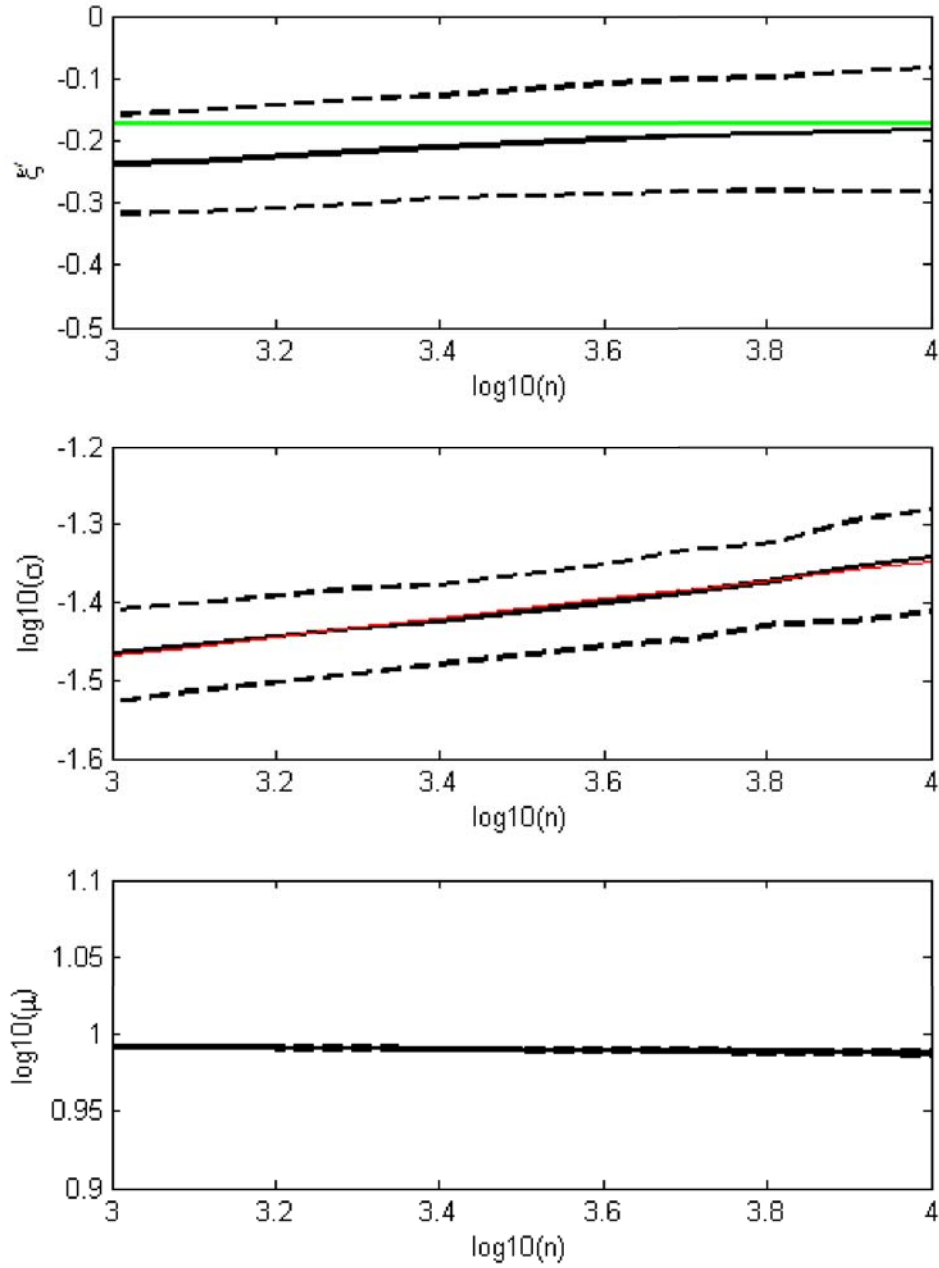Figure 7: g_2 observable **a)** ξ' VS $\log_{10}(n)$; **b)** $\log_{10}(\sigma)$ VS $\log_{10}(n)$; **c)** $\log_{10}(\mu)$ VS $\log_{10}(n)$. Baker map. Dotted lines represent one standard deviation,, red lines represent a linear fit, green lines are theoretical values.

Figure 8: g_3 observable. **a)** ξ' VS $\log_{10}(n)$; **b)** $\log_{10}(\sigma)$ VS $\log_{10}(n)$; **c)** $\log_{10}(\mu)$ VS $\log_{10}(n)$. Baker map. Dotted lines represent one standard deviation, red lines represent a linear fit, green lines are theoretical values.

25

The Hènon and Lozi maps

The Hènon map is defined as:

$$\begin{aligned} x_{t+1} &= y_t + 1 - ax_t^2 \\ y_{t+1} &= bx_t \end{aligned}$$

(31)

while in the Lozi map ax_t^2 is substituted with $a|x_t|$:

$$\begin{aligned} x_{t+1} &= y_t + 1 - a|x_t| \\ y_{t+1} &= bx_t \end{aligned}$$

(32)

We consider the classical set of parameter $a = 1.4$, $b = 0.3$ for the Hènon map and $a = 1.7$ and $b = 0.5$ for the Lozi map.

Young [1985] proved the existence of the SRB measure for the Lozi map, whereas for the Hènon map no such rigorous proof exists, even if convincing numerical results suggest its existence [Badii and Politi, 1987]. Note that Benedicks and Carleson [1991] proved the existence of an SRB measure for the Hènon map with a different set of parameters. Using the classical Young results which makes use of the Lyapunov exponents, we obtain an exact result for Δ for the Lozi attractor:

$$\Delta \simeq 1.40419$$

Instead, in the case of the Hènon attractor, we consider the numerical estimate provided by Grassberger [1983]:

$$\Delta = 1.25826 \pm 0.00006$$

As in the previous cases the GEV distribution is computed with L-moments methods varying n and m and averaging the distribution parameters over 1000 different sample points chosen as described before for the Baker map. Results are presented in figures 9-11, the plots on the left-hand side refer to the Hènon map, while on the right-hand side the results refer to the Lozi map.

When considering ξ', the numerical results are in agreement with the theoretical estimates. Nevertheless, the parameters distribution have a rather range spread which indicates a slower convergence towards the expected values in respect to what is observed for the IFS case. The experimental values of $\sigma(g_1)$, shown in Figure 9b , approach the theoretical values shown by a green line. The angular coefficient computed from the semilog plot of $\mu(g_1)$ represented in Figure 9c) gives us an estimate of the dimension $\Delta = 1/|\kappa| = 1.234 \pm 0.015$ for Hènon and $\Delta = 1/|\kappa| = 1.40 \pm 0.01$ for Lozi.

The other angular coefficients related to g_2 and g_3 observables for the plots shown in figures 10b), 10c), and 11b) are presented in Table 2. Within 2 standard deviation they are comparable with the theoretical ones and the best agreement is achieved considering $\mu(g_2)$. The constant $C = 10$ is approached quite well (Figure 11c)).

$\Delta = 1/(\alpha|\kappa|)$	Baker	Hènon	Lozi
Theor.	1.4357	1.2582	1.4042
$\mu(g_2)$	1.47 ± 0.02	1.238 ± 0.009	1.396 ± 0.008
$\sigma(g_2)$	1.39 ± 0.04	1.35 ± 0.07	1.38 ± 0.02
$\sigma(g_3)$	1.56 ± 0.08	1.15 ± 0.07	1.42 ± 0.01

Table 2: Information dimension Δ computed taking the logarithm of equations 20-21 and computing the angular coefficient κ of a linear fit of data; for Baker, Hènon and Lozi maps.

The slower convergence for these maps may be related to the difficulties experienced computing the dimension with all box-counting methods, as shown in Grassberger [1983],Badii and Politi [1987]. In that case it has been proven that the number of points that are required to cover a fixed fraction of the attractor support diverges faster than the number of boxes itself for this kind of non uniform attractor. In our case the situation is similar since we consider balls around the initial condition ζ. As pointed out, the best result for the dimension is achieved using the parameters provided by g_1 observable since the logarithm modulation of the distance exalts proper extrema while weights less possible outliers.
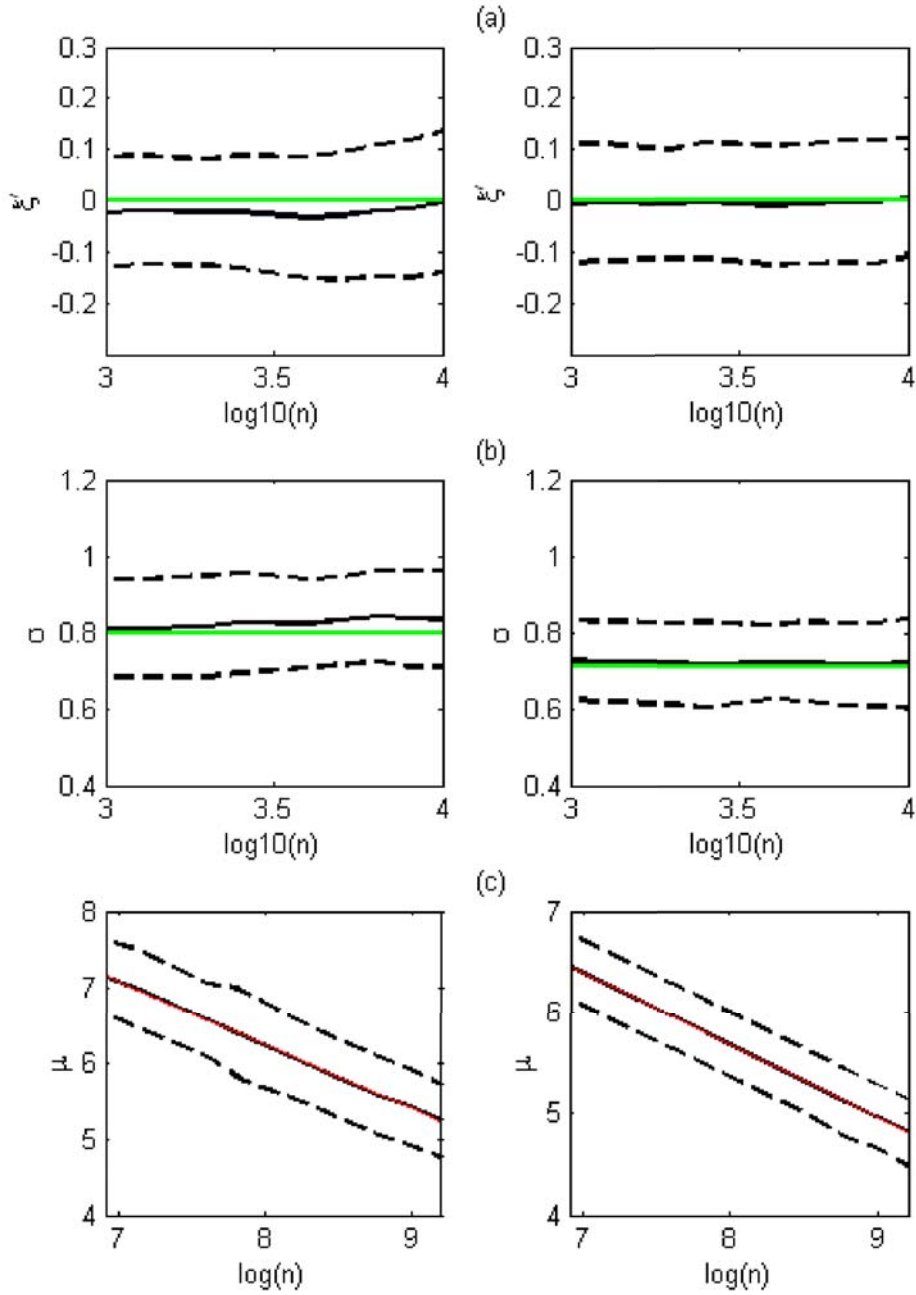
Figure 9: g_1 observable. **a)** ξ' VS $\log_{10}(n)$; **b)** σ VS $\log_{10}(n)$; **c)** μ VS $\log(n)$. Left: Hènon map, Right: Lozi map. Dotted lines represent one standard deviation, red lines represent a linear fit, green lines are theoretical values.
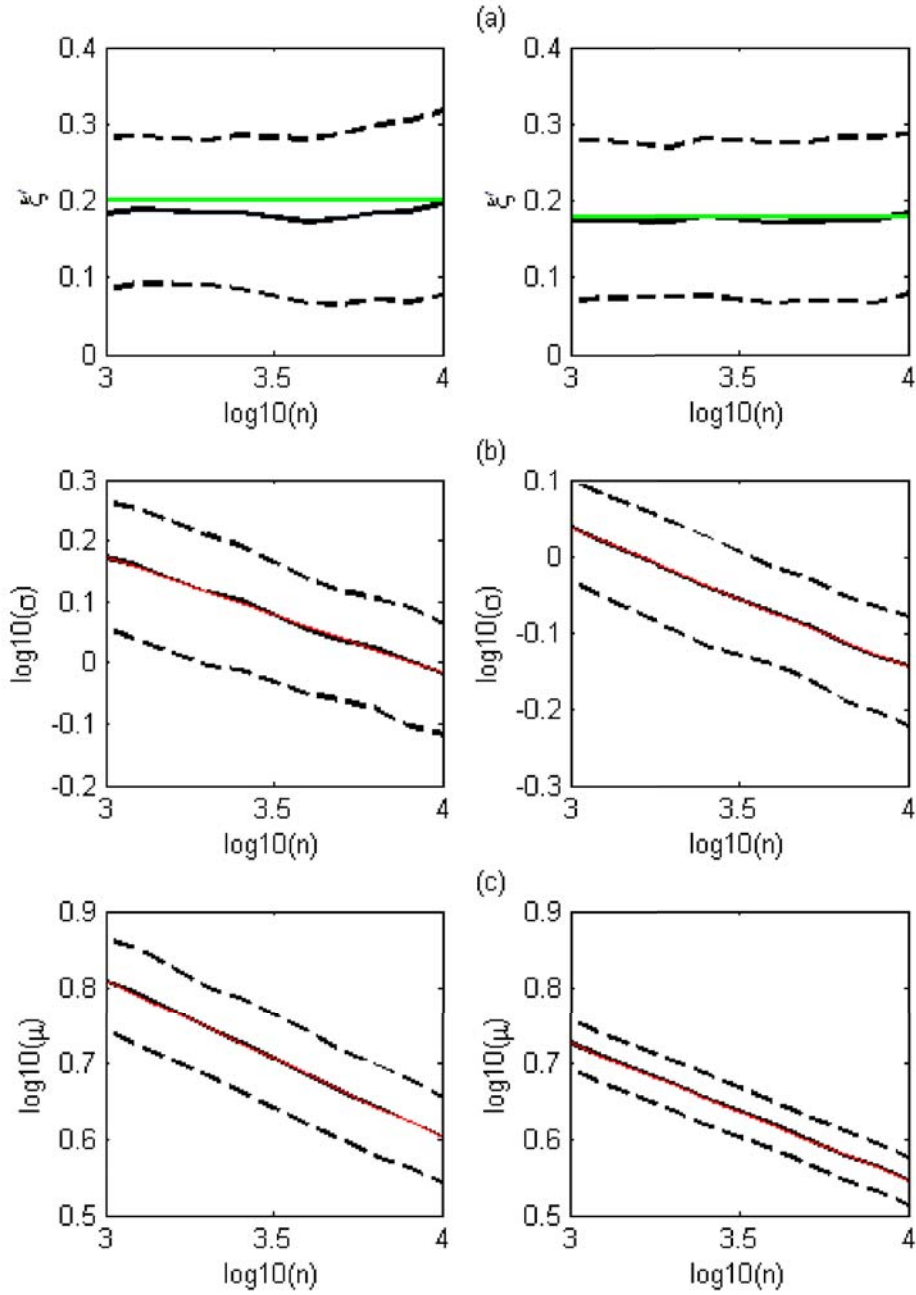
28

Figure 10: g_2 observable **a)** ξ' VS $\log_{10}(n)$; **b)** $\log_{10}(\sigma)$ VS $\log_{10}(n)$; fc) $\log_{10}(\mu)$ VS $\log_{10}(n)$. Left: Hènon map, Right: Lozi map. Dotted lines represent one standard deviation,, red lines represent a linear fit, green lines are theoretical values.   29
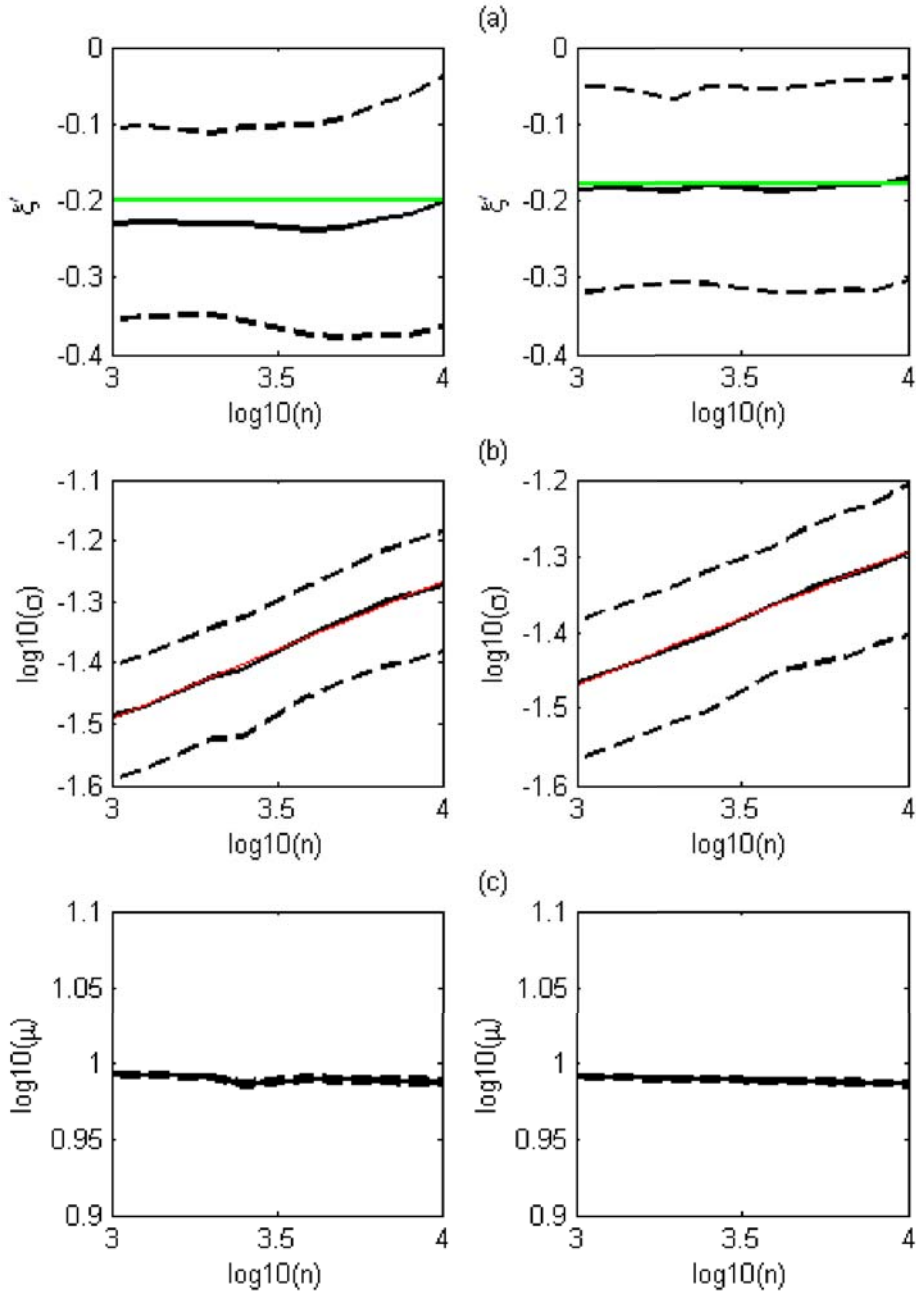
Figure 11: g_3 observable. **a)** ξ' VS $\log_{10}(n)$; **b)** $\log_{10}(\sigma)$ VS $\log_{10}(n)$; **c)** $\log_{10}(\mu)$ VS $\log_{10}(n)$. Left: Hènon map, Right: Lozi map. Dotted lines represent one standard deviation, red lines represent a linear fit, green lines are theoretical values.   30

4   Final Remarks

Extreme Value Theory is attracting a lot of interest both in terms of extending pure mathematical results and in terms of applications to many fields of social and natural science. As an example, in geophysical applications is crucial to have a tool to understand and forecast climatic extrema and events such as strong earthquakes and floods.

Whereas the classical Extreme Value Theory deals with stochastic processes, many applications demanded to understand whereas it could rigorously be used to study the outputs of deterministic dynamical systems. The mathematical models used to study them present a rich structure and their attracting sets are very often strange attractors. In such sense is extremely important to develop an extreme value theory for dynamical system with singular measures. Recently, The existence of extreme value laws for dynamical systems preserving an absolutely continuous invariant measure or a singular continuous invariant measure has been proven if strong mixing properties or exponential hitting time statistics on balls are satisfied.

In this work we have extended the results presented in Faranda et al. [2011] to the case of dynamical systems with singular measures. Our main results is that there exist an extreme value distribution for this kind of systems that is related to the GEV distribution when observable functions of the distance between the iterated orbit and the initial conditions are chosen. The three extreme value type for the limit distribution laws for maxima and the generalized distribution family (GEV) are absolutely continuous function. We will recover the GEV using histograms on the frequency of maxima; in this way the cumulative distribution function which we got from such an histogram will have *plateaux* just in correspondence of the holes of the Cantor set, whenever this one is the invariant set. This could be easily explained by the very nature of our observables which measure the distance with respect to a given point: there will be distances which are not allowed when such distances are computed from points in the holes. It should be stressed that such a cumulative distribution function, which is a sort of devil staircase and therefore is a singular continuous function, in any way could converge to a GEV distribution. The latter as the three type extreme values laws are normalized laws which must be adjusted in order to give a non-degenerate asymptotic distribution. The strength of our approach, as we said above, relies in the possibility to infer the nature and the value of such normalizing constants by a fitting procedure on the unnormalized data, the histograms. This worked very well for probability measure which were absolutely continuous. We will see that it works also for singular measures (and the normalizing constants will be related to the information dimension), provided we remind that this time the fitting procedure will contain a sort of extrapolation to smooth out the gaps of the Cantor sets.

It is interesting to observe that on Cantor sets the notion of *generic point* is not so obvious as for smooth manifolds which support Lebesgue measure: in this case in fact one could suppose that each point accessible for numerical iterations is generic with respect to an invariant measure which is in turn absolutely continuous. This notion of genericity is restored on attractor by considering the SRB measure. Instead for Iterated Function Systems we can dispose of uncountably many measures, but we have a precise manner to identify them and this will be reflected in the different dimensions produced by the numerical computation of the parameters of the GEV. The possibility to discriminate among different singular measures having all the same topological support is another indication of the validity and of the efficiency of our approach.

We have also shown that the parameters of the distribution are intimately related to the information

dimension of the invariant set. We have tested our conjecture with numerical experiments on different low dimensional maps such as : the middle third Cantor set, the Sierpinskij triangle), Iterated Function System (IFS) with non-uniform weights, strange attractors such as Lozi and Hènon. In all cases considered there is agreement between the theoretical parameters and the experimental ones. The extimates of Δ are in agreement with the theoretical values in all cases considered. It is interesting to observe that the algorithm described with the selection of maxima acts like a magnifying glass on the neighborhood of the initial condition. In this way we have both a powerful tool to study and highlight the fine structure of the attractor, but, on the other hand we can obtain global properties averaging on different initial conditions. Even if we are dealing with very simple maps for which many properties are known it is clear from numerical experiments that is not so obvious to observe a good convergence to the GEV distribution. Even if we are able to compute very large statistics and the results are consistent with theoretical values, the error range is wide if compared to the experiment for maps with a.c.i.m. measures that we have carried out in Faranda et al. [2011]. This should be taken in consideration each time this statistics is applied in a predictive way to more complicated systems.

In the case of an experimental temporal series, for which the underlying dynamics is unknown, a classical problem is to obtain the dimensionality of the attractor of the dynamical systems which generated it. This can be achieved through the so called Ruelle-Takens delay embedding where, starting from the time series of an observable $O(n)$, we can construct the multivariate vectors in a Δ-dimensional space:

$$\phi(N) = [O(n), O(n+1), ..., O(n+d-1)]$$

and study the geometrical properties using the Recurrence Qualification Analysis [Marwan et al., 2007]. The minimum value of Δ needed to reconstruct the actual dimension d^* is given by $[2d] + 2$. Using the procedure described in this paper we could find an estimate for d^* thus determining the minimum value of Δ to be used in the Recurrence Quantification Analysis. We will test this strategy in a subsequent publication.

The theory and the algorithm presented in this work and in Faranda et al. [2011] allow to study in detail the recurrence of an orbit around a point: this is due to the particular choice of the observables that require to compute distances between initial and future states of the system. Understanding the behavior of a dynamical system in a neighborhood of a particular initial condition is of great interest in many applications. As an example, in weather forecast and climate it is important to study the recurrence of patterns (the so called analogues). In principle, applying the extreme value statistics to the output of meteorological models, will make possible to infer dynamical properties related to the closest return towards a certain weather pattern. EVT will give information not only about the probability distribution of the extrema but also about the scaling of the measure of a ball centered on the chosen initial condition providing an insight to the dynamical structure of the attractor.

5   Acknowledgments

References

E.G. Altmann, S. Hallerberg, and H. Kantz. Reactions to extreme events: Moving threshold model. *Physica A: Statistical Mechanics and its Applications*, 364:435–444, 2006.

R. Badii and A. Politi. Renyi dimensions from local expansion rates. *Physical Review A*, 35(3): 1288, 1987.

V. Balakrishnan, C. Nicolis, and G. Nicolis. Extreme value distributions in chaotic dynamics. *Journal of Statistical Physics*, 80(1):307–336, 1995. ISSN 0022-4715.

M.F. Barnsley. *Fractals everywhere*. Morgan Kaufmann Pub, 2000. ISBN 0120790696.

M.F. Barnsley and S. Demko. Iterated function systems and the global construction of fractals. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 399(1817): 243, 1985.

J. Beirlant. *Statistics of extremes: theory and applications*. John Wiley & Sons Inc, 2004. ISBN 0471976474.

M. Benedicks and L. Carleson. The dynamics of the hénon map. *The Annals of Mathematics*, 133 (1):73–169, 1991.

D. Bessis, JD Fournier, G. Servizi, G. Turchetti, and S. Vaienti. Mellin transforms of correlation integrals and generalized dimension of strange sets. *Physical Review A*, 36(2):920, 1987.

D. Bessis, G. Paladin, G. Turchetti, and S. Vaienti. Generalized dimensions, entropies, and liapunov exponents from the pressure function for strange sets. *Journal of statistical physics*, 51(1):109–134, 1988.

E. Brodin and C. Kluppelberg. Extreme Value Theory in Finance. *Submitted for publication: Center for Mathematical Sciences, Munich University of Technology*, 2006.

P.W. Burton. Seismic risk in southern Europe through to India examined using Gumbel's third distribution of extreme values. *Geophysical Journal of the Royal Astronomical Society*, 59(2): 249–280, 1979. ISSN 1365-246X.

S. Coles, J. Heffernan, and J. Tawn. Dependence measures for extreme value analyses. *Extremes*, 2 (4):339–365, 1999. ISSN 1386-1999.

P. Collet. Statistics of closest return for some non-uniformly hyperbolic systems. *Ergodic Theory and Dynamical Systems*, 21(02):401–420, 2001. ISSN 0143-3857.

C.A. Cornell. Engineering seismic risk analysis. *Bulletin of the Seismological Society of America*, 58(5):1583, 1968. ISSN 0037-1106.

M.G. Cruz. *Modeling, measuring and hedging operational risk*. John Wiley & Sons, 2002. ISBN 0471515604.

C.D. Cutler and D.A. Dawson. Estimation of dimension for spatially distributed data and related limit theorems. *Journal of multivariate analysis*, 28(1):115–148, 1989.

A.C. Davison. Modelling excesses over high thresholds, with an application. *Statistical Extremes and Applications*, pages 461–482, 1984.

AC Davison and R.L. Smith. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(3):393–442, 1990. ISSN 0035-9246.

D. Faranda, V. Lucarini, G. Turchetti, and S. Vaienti. Numerical convergence of the block-maxima approach to the Generalized Extreme Value distribution. *Accepted for publication in Journal of Statistical Physics*, 2011.

M. Felici, V. Lucarini, A. Speranza, and R. Vitolo. Extreme Value Statistics of the Total Energy in an Intermediate Complexity Model of the Mid-latitude Atmospheric Jet. Part I: Stationary case.(3337K, PDF). *Journal of Atmospheric Science*, 64:2137–2158, 2007.

RA Fisher and LHC Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Proceedings of the Cambridge philosophical society*, volume 24, page 180, 1928.

A.C.M. Freitas and J.M. Freitas. On the link between dependence and independence in extreme value theory for dynamical systems. *Statistics & Probability Letters*, 78(9):1088–1093, 2008. ISSN 0167-7152.

A.C.M. Freitas, J.M. Freitas, and M. Todd. Hitting time statistics and extreme value theory. *Probability Theory and Related Fields*, pages 1–36, 2009.

A.C.M. Freitas, J.M. Freitas, and M. Todd. Extremal Index, Hitting Time Statistics and periodicity. *Arxiv preprint arXiv:1008.1350*, 2010a.

A.C.M. Freitas, J.M. Freitas, M. Todd, B. Gardas, D. Drichel, M. Flohr, RT Thompson, SA Cummer, J. Frauendiener, A. Doliwa, et al. Extreme value laws in dynamical systems for non-smooth observations. *Arxiv preprint arXiv:1006.3276*, 2010b.

M. Ghil et al. Extreme events: Dynamics, statistics and prediction. *Non Linear process in Geophysics*, in press, 2011.

B. Gnedenko. Sur la distribution limite du terme maximum d'une série aléatoire. *The Annals of Mathematics*, 44(3):423–453, 1943.

P. Grassberger. Generalized dimensions of strange attractors. *Physics Letters A*, 97(6):227–230, 1983. ISSN 0375-9601.

EJ Gumbel. The return period of flood flows. *The Annals of Mathematical Statistics*, 12(2):163–190, 1941. ISSN 0003-4851.

C. Gupta, M. Holland, and M. Nicol. Extreme value theory for hyperbolic billiards. *Lozi-like maps, and Lorenz-like maps, preprint*, 2009.

G. Haiman. Extreme values of the tent map process. *Statistics & Probability Letters*, 65(4):451–456, 2003. ISSN 0167-7152.

HGE Hentschel and I. Procaccia. The infinite number of generalized dimensions of fractals and strange attractors. *Physica D: Nonlinear Phenomena*, 8(3):435–444, 1983.

M. Holland, M. Nicol, and A. Török. Extreme value distributions for non-uniformly hyperbolic dynamical systems. *preprint*, 2008.

J.R.M. Hosking. L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(1):105–124, 1990. ISSN 0035-9246.

J. Kaplan and J. Yorke. Chaotic behavior of multidimensional difference equations. *Functional Differential equations and approximation of fixed points*, pages 204–227, 1979.

RW Katz. Extreme value theory for precipitation: Sensitivity analysis for climate change. *Advances in Water Resources*, 23(2):133–139, 1999. ISSN 0309-1708.

R.W. Katz and B.G. Brown. Extreme events in a changing climate: variability is more important than averages. *Climatic change*, 21(3):289–302, 1992. ISSN 0165-0009.

R.W. Katz, G.S. Brush, and M.B. Parlange. Statistics of extremes: Modeling ecological disturbances. *Ecology*, 86(5):1124–1134, 2005. ISSN 0012-9658.

MR Leadbetter, G. Lindgren, and H. Rootzen. *Extremes and related properties of random sequences and processes*. Springer, New York, 1983.

F. Ledrappier. Some relations between dimension and lyapounov exponents. *Communications in Mathematical Physics*, 81(2):229–238, 1981.

H.W. Lilliefors. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62(318):399–402, 1967. ISSN 0162-1459.

N. Marwan, M. Carmen Romano, M. Thiel, and J. Kurths. Recurrence plots for the analysis of complex systems. *Physics Reports*, 438(5-6):237–329, 2007.

N. Nicholis. CLIVAR and IPCC interests in extreme events'. In *Workshop Proceedings on Indices and Indicators for Climate Extremes, Asheville, NC. Sponsors, CLIVAR, GCOS and WMO*, 1997.

C. Nicolis, V. Balakrishnan, and G. Nicolis. Extreme events in deterministic dynamical systems. *Physical review letters*, 97(21):210602, 2006. ISSN 1079-7114.

Friederichs P. and A. Hense. Statistical downscaling of extreme precipitation events using censored quantile regression. *Monthly weather review*, 135(6):2365–2378, 2007. ISSN 0027-0644.

J. Pickands III. Moment convergence of sample extremes. *The Annals of Mathematical Statistics*, 39(3):881–889, 1968.

R.L. Smith. Threshold methods for sample extremes. *Statistical extremes and applications*, 621:638, 1984.

R.L. Smith. Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statistical Science*, 4(4):367–377, 1989. ISSN 0883-4237.

D. Sornette, L. Knopoff, YY Kagan, and C. Vanneste. Rank-ordering statistics of extreme events: application to the distribution of large earthquakes. *Journal of Geophysical Research*, 101(B6): 13883, 1996. ISSN 0148-0227.

J.C. Sprott. *Chaos and time-series analysis.* Oxford Univ Pr, 2003. ISBN 0198508409.

O.G.B. Sveinsson and D.C. Boes. Regional frequency analysis of extreme precipitation in northeastern colorado and fort collins flood of 1997. *Journal of Hydrologic Engineering*, 7:49, 2002.

P. Todorovic and E. Zelenhasic. A stochastic model for flood analysis. *Water Resources Research*, 6(6):1641–1648, 1970. ISSN 0043-1397.

S. Vannitsem. Statistical properties of the temperature maxima in an intermediate order Quasi-Geostrophic model. *Tellus A*, 59(1):80–95, 2007. ISSN 1600-0870.

R. Vitolo, M.P. Holland, and C.A.T. Ferro. Robust extremes in chaotic deterministic systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 19:043127, 2009a.

R. Vitolo, PM Ruti, A. Dell'Aquila, M. Felici, V. Lucarini, and A. Speranza. Accessing extremes of mid-latitudinal wave activity: methodology and application. *Tellus A*, 61(1):35–49, 2009b. ISSN 1600-0870.

B. Yakov. Pesin. dimension theory in dynamical systems: Contemporary views and applications. chicago lectures in mathematics, 1998.

L.S. Young. Dimension, entropy and lyapunov exponents. *Ergodic theory and dynamical systems*, 2 (01):109–124, 1982.

L.S. Young. Bowen-ruelle measures for certain piecewise hyperbolic maps. *american mathematical society*, 287(1), 1985.