

Computing pointwise fractal dimension by conditioning in multivariate distributions and time series

COLLEEN D. CUTLER

Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1. E-mail: cdcutler@math.uwaterloo.ca

It is natural in many contexts to employ conditioning arguments in order to deduce properties of a multivariate distribution P_n from properties of a lower-order distribution P_{n-1} . In this paper we show that the pointwise (fractal) dimension of P_n can be updated from the pointwise dimension of P_{n-1} via the one-step conditional distributions provided the latter satisfy certain Lipschitz-type properties. Specifically, we prove that pointwise dimension can be computed iteratively according to the conditional additivity rule

$$\alpha(x_1, \dots, x_n) = \alpha(x_1) + \alpha(x_2|x_1) + \alpha(x_3|x_1, x_2) + \dots + \alpha(x_n|x_1, \dots, x_{n-1}).$$

This approach is then used to analyse the behaviour of pointwise dimension for various stationary stochastic processes; the emphasis is on dynamical systems corrupted by noise. In particular, we show that for functionals of stochastic processes with discrete conditional distributions satisfying the necessary conditions (such as missing-data models of dynamical systems and randomly iterated function systems) pointwise dimension remains bounded over time just as in the strictly deterministic case. On the other hand, we prove that for stochastic dynamical systems with additive diffuse noise, pointwise dimension diverges to infinity over time. An example of a stationary dynamical system where the conditional additivity rule fails is also provided.

Keywords: determinism; fractal dimension; Grassberger–Procaccia algorithm; iterated function system; Markov process; missing data; pointwise dimension; stochastic dynamical system; time series

1. Introduction

In the following \mathcal{X} will always denote a complete separable metric space with metric ρ . If P is a probability distribution on the Borel sets of \mathcal{X} , the *pointwise dimension* of P at $x \in \mathcal{X}$ (sometimes also called the local Hölder exponent at x) is defined as

$$\alpha_P(x) = \lim_{r \rightarrow 0} \frac{\log P(B(x, r))}{\log r}, \quad (1.1)$$

provided this limit exists. Here $B(x, r)$ denotes the closed ball of radius r centred at x . If the limit does not exist, we may consider the upper and lower values $\alpha_P^+(x)$ and $\alpha_P^-(x)$ obtained by replacing ‘lim’ in (1.1) by ‘lim sup’ and ‘lim inf’, respectively. However, in all examples

considered in this paper, the limits in (1.1) will hold P -almost surely (a.s.) and, for simplicity, we will assume that this is the case throughout the following discussion.

Pointwise dimension and numerous related notions of scaling associated specifically with probability measures (such as information dimension and correlation dimension) have a relatively recent but intense history in physics, mathematics and probability theory. This great interest is largely the result of the application of fractal concepts to the study of nonlinear dynamical systems and related stochastic processes; see, for example, Billingsley (1960; 1961), Young (1982), Hentschel and Procaccia (1983), Farmer *et al.* (1983), Cutler (1990a, 1990b; 1991; 1993), Taylor (1986; 1992; 1995), Falconer (1990), Theiler (1990; 1991), Smith (1992), Barndorff-Nielsen *et al.* (1993), Ott *et al.* (1994), Serinko (1994), Pesin (1996), and Keller (1997). The mapping (1.1) obviously describes the local scaling behaviour of P as x varies over \mathcal{X} , but its behaviour is also connected to the manner in which P -mass is distributed globally over sets of differing dimensions; see Cutler (1990b; 1993) and Taylor (1992). In particular, we have the bounds

$$\text{ess inf } \alpha_P = \inf\{\dim(B) | P(B) > 0\} \leq \inf\{\dim(B) | P(B) = 1\} = \text{ess sup } \alpha_P, \quad (1.2)$$

where ‘dim’ denotes *Hausdorff dimension*, and ‘ess inf’ and ‘ess sup’ denote, respectively, the essential infimum and essential supremum of $\alpha_P(\cdot)$ over \mathcal{X} with respect to P . Thus the interval of numbers $[\text{ess inf } \alpha_P, \text{ess sup } \alpha_P]$ exactly bounds the effective *dimension range* of P . An important special case, currently of much interest in nonlinear dynamics, occurs when P is an ergodic invariant measure of a typical smooth chaotic dynamical system. Then (and in many other situations as well) there exists a constant α_0 such that $\alpha_P(x) = \alpha_0$ P -a.s. In this case the dimension range of P collapses to the single number α_0 and it is appropriate to speak of the *pointwise dimension* of P . (However, the P -null sets $\{x | \alpha_P(x) = \alpha\}$, $\alpha \neq \alpha_0$, are often significant in their own right and comprise the so-called *multifractal spectrum* of P ; see Halsey *et al.* (1986), Falconer (1990), Brown *et al.* (1992), Holley and Waymire (1992) and Taylor (1995).) The pointwise dimension of P is sometimes referred to as the *information dimension* by physicists because of its connection with entropy and the rate of information creation in an evolving dynamical system; see Farmer (1982) and Ott *et al.* (1994).

Experimental data from physical sources (usually obtained by observing some functional of an evolving system) typically take the form of a real- or vector-valued time series X_1, X_2, \dots , and an attempt is made to reconstruct the ‘hidden’ underlying state space via the method of *time-delay embeddings*; see Sauer *et al.* (1991) and Abarbanel *et al.* (1993). In the time-delay embedding method, observations are grouped into d -dimensional vectors of the form $(X_{m+1}, X_{m+2}, \dots, X_{m+d})$ and their behaviour is studied as the embedding dimension $d \rightarrow \infty$. This is equivalent to studying the behaviour of the joint distributions as the number n of components (X_1, \dots, X_n) increases. In particular, if dimension values approach a finite asymptote as $n \rightarrow \infty$, this is often interpreted as evidence that the underlying system has settled onto a finite-dimensional attractor and may be regarded as deterministic (this approach is called the *Grassberger–Procaccia algorithm*); see Ott *et al.* (1994) and Cutler (1993; 1994; 1997). Results in Cutler (1997) show that pointwise dimension does in fact remain bounded over time for any Lipschitz functional of a smooth finite-dimensional system. However, one goal of the present paper is to exhibit some simple natural *stochastic* processes which also possess bounded pointwise dimension over time –

see the related sequence of papers Osborne and Provenzale (1989), Theiler (1991), and Cutler (1994; 1997) on the theme of the relationship between determinism and finite fractal dimension in the context of correlation dimension; see also Wolff (1990). Here we are considering pointwise dimension, rather than correlation dimension, because the former is generally mathematically superior to the latter. We are able to derive a practical formula for computing pointwise dimension iteratively; moreover, pointwise dimension is directly linked to the dimension structure of the associated distribution via (1.2) while correlation dimension is sensitive to possible singularities in the measure and can take on arbitrarily small values; see examples and discussion in Cutler (1990a; 1991; 1993).

In the following we will consider time series X_1, X_2, \dots taking values in \mathcal{X} and having joint distributions $P_n, n = 1, 2, \dots$ that is, P_n is the distribution of (X_1, \dots, X_n) over the Borel sets of the product space $\mathcal{X}^n = \times_{m=1}^n \mathcal{X}$. The focus will be on the pointwise dimension map $\alpha_{P_n}(\cdot)$ of P_n and on developing a method of updating $\alpha_{P_n}(\cdot)$ from $\alpha_{P_{n-1}}(\cdot)$ using the one-step conditional distributions of the time series, provided these distributions satisfy certain Lipschitz-type properties. This updating method allows pointwise dimension to be computed iteratively, and is a particularly powerful theoretical tool when the conditional distributions of a time series are readily available (these are generally much more accessible and of much simpler structure than the joint distributions themselves). The later sections of the paper employ this procedure to compute pointwise dimension in a variety of examples. Section 3 considers time series which can be expressed as functionals of processes with discrete conditional distributions, with applications to missing-data in dynamical systems (where observations are randomly and unknowingly missed) and to iterated function systems. In Section 4 we establish results on the behaviour of pointwise dimension for stochastic dynamical systems with independent and state-dependent additive noise.

We also note that while we are focusing here on time series, the results (and technique) are equally valid when looking simply at multivariate distributions and computing pointwise dimension by conditioning on lower-order joint distributions.

2. The conditional additivity rule for dimension

We now introduce some convenient notation and preliminaries. An element (x_1, \dots, x_n) of \mathcal{X}^n will sometimes be denoted by the subscripted bold italic \mathbf{x}_n ; similarly, the random vector (X_1, \dots, X_n) will sometimes be denoted by \mathbf{X}_n . The metric ρ on \mathcal{X}^n will always be the ‘sup’ metric derived from ρ on \mathcal{X} , that is,

$$\rho(\mathbf{x}_n, \mathbf{z}_n) = \max_{1 \leq j \leq n} \rho(x_j, z_j). \tag{2.1}$$

It follows from Theorem 2.12 of Cutler (1997) that pointwise dimension may be regarded as an increasing function in n since the inequality $\alpha_{P_{n-1}}(\mathbf{x}_{n-1}) \leq \alpha_{P_n}(\mathbf{x}_n)$ always holds at each point \mathbf{x}_n . Moreover, it is easy to show that if X_1, \dots, X_n are *independent* random variables, then

$$\alpha_{P_n}(\mathbf{X}_n) = \alpha_1(X_1) + \dots + \alpha_n(X_n) \text{ a.s.}, \tag{2.2}$$

where $\alpha_j(\cdot)$ is the pointwise dimension map of the marginal distribution of X_j . In this section we generalize this additivity property to the dependent case.

If (Z, Y) represents a random vector taking values in a complete separable metric space $\mathcal{Z} \times \mathcal{Y}$ and having joint distribution P , it follows from Theorem 10.2.2 of Dudley (1989) that there exists a unique family of *regular conditional distributions* $P(\cdot|Z = z)$ describing the distribution of Y given Z . For each fixed value $Z = z$ we let $\alpha(\cdot|z)$ denote the pointwise dimension map of the conditional distribution $P(\cdot|Z = z)$; that is,

$$\alpha(y|z) = \lim_{r \rightarrow 0} \frac{\log P(B(y, r)|Z = z)}{\log r} \tag{2.3}$$

as y varies over \mathcal{Y} . In the following we will be discussing the pointwise dimension maps associated with P_{n-1} , P_n , and the (one-step) conditional distributions of X_n given the past \mathbf{X}_{n-1} ; for convenience we will denote the conditional distribution of X_n given $\mathbf{X}_{n-1} = \mathbf{x}_{n-1}$ by $P(\cdot|\mathbf{x}_{n-1})$. To further simplify notation, we will usually drop the subscripts on the pointwise dimension maps, identifying the corresponding probability measure by the *domain* of $\alpha(\cdot)$. For example, $\alpha(\mathbf{x}_n) = \alpha(x_1, \dots, x_n)$ has domain \mathcal{X}^n and therefore represents the pointwise dimension map of P_n , while $\alpha(x_n|\mathbf{x}_{n-1})$ has domain \mathcal{X} conditional on $\mathbf{x}_{n-1} = (x_1, \dots, x_{n-1})$ and therefore represents the pointwise dimension map of the conditional distribution $P(\cdot|\mathbf{x}_{n-1})$.

The following is our main technical result.

Theorem 2.1. *Let $\mathbf{X}_n = (X_1, \dots, X_n)$ be a random vector in \mathcal{X}^n with distribution P_n , and let $P(\cdot|\mathbf{z}_{n-1})$ denote the regular conditional distribution of X_n given $\mathbf{X}_{n-1} = \mathbf{z}_{n-1}$. Suppose that $\mathbf{x}_n = (x_1, \dots, x_n)$ is a point for which there exists $a_1(r) > 0$, $a_2(r) > 0$, $b \geq 1$, $r_0 > 0$, and a Borel set $S \subseteq \mathcal{X}^{n-1}$ with $P_{n-1}(S) = 1$ such that, for all $0 < r < r_0$,*

$$P(B(x_n, r)|\mathbf{z}_{n-1}) \leq a_2(r) P(B(x_n, br)|\mathbf{x}_{n-1})$$

$$\text{for all } \mathbf{z}_{n-1} \in S \cap B(\mathbf{x}_{n-1}, r) \tag{2.4}$$

and

$$P(B(x_n, br)|\mathbf{z}_{n-1}) \geq a_1(r) P(B(x_n, r)|\mathbf{x}_{n-1})$$

$$\text{for all } \mathbf{z}_{n-1} \in S \cap B(\mathbf{x}_{n-1}, r), \tag{2.5}$$

where $\log a_j(r) = o(\log r)$ as $r \rightarrow 0$ for $j = 1, 2$.

Then, provided the limit (1.1) exists for each of the maps on the right-hand side of the following equation, we have

$$\alpha(\mathbf{x}_n) = \alpha(\mathbf{x}_{n-1}) + \alpha(x_n|\mathbf{x}_{n-1}), \tag{2.6}$$

where $\alpha(x_n|\mathbf{x}_{n-1})$ is the pointwise dimension of $P(\cdot|\mathbf{x}_{n-1})$ at x_n . More generally, even if the limit (1.1) does not exist for some of the maps, we obtain the bounds

$$\alpha^-(\mathbf{x}_n) \geq \alpha^-(\mathbf{x}_{n-1}) + \alpha^-(x_n|\mathbf{x}_{n-1}) \tag{2.7}$$

and

$$\alpha^+(\mathbf{x}_n) \leq \alpha^+(\mathbf{x}_{n-1}) + \alpha^+(x_n|\mathbf{x}_{n-1}). \tag{2.8}$$

Proof. Since $B(\mathbf{x}_n, r) = B(\mathbf{x}_{n-1}, r) \times B(x_n, r)$ under the sup metric, we can write

$$P_n(B(\mathbf{x}_n, r)) = \int_{B(\mathbf{x}_{n-1}, r)} P(B(x_n, r)|z_{n-1}) P_{n-1}(dz_{n-1}). \tag{2.9}$$

Applying (2.4) to (2.9), we obtain the upper bound

$$P_n(B(\mathbf{x}_n, r)) \leq a_2(r)P(B(x_n, br)|\mathbf{x}_{n-1})P_{n-1}(B(\mathbf{x}_{n-1}, r)). \tag{2.10}$$

Now replacing r by br everywhere in (2.9), it follows by reducing the region of integration to $B(\mathbf{x}_{n-1}, r)$ and applying (2.5) that we obtain the lower bound

$$P_n(B(\mathbf{x}_n, br)) \geq a_1(r)P(B(x_n, r)|\mathbf{x}_{n-1})P_{n-1}(B(\mathbf{x}_{n-1}, r)). \tag{2.11}$$

Taking logarithms, dividing by $\log r$, and letting $r \rightarrow 0$ while noting that $\log br / \log r \rightarrow 1$, we see that (2.10) yields (2.7) while (2.11) yields (2.8) as desired. \square

Remark 2.2. The above result shows that pointwise dimension is additive in a natural way provided the one-step conditional distributions satisfy a Lipschitz-type continuity in their arguments. If all limits (1.1) exist and equation (2.6) holds at the point \mathbf{x}_n , we will say that the pointwise dimension of P_n satisfies the *one-step conditional additivity rule* at \mathbf{x}_n . It is important to note that the Lipschitz-type conditions (2.4) and (2.5) are generally not expendable. In Example 3.5 we show that the conditional additivity rule can fail in the absence of these conditions.

Corollary 2.3 (Conditional additivity rule). *Suppose property (2.6) holds at each step $\mathbf{x}_k = (x_1, \dots, x_k)$, $k = 2, \dots, n$. Then inductively we obtain*

$$\alpha(x_1, \dots, x_n) = \alpha(x_1) + \alpha(x_2|x_1) + \alpha(x_3|x_1, x_2) + \dots + \alpha(x_n|x_1, \dots, x_{n-1}). \tag{2.12}$$

This extends the additivity equation for independent random variables to the general case.

Remark 2.4. We may also state Theorem 2.1 more generally in terms of the conditional distribution of X_j given the other $n - 1$ variables $X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_n$; that is, no particular index ordering is relevant to the result. (In fact, the result holds quite generally in product spaces admitting regular conditional distributions, but for the purpose of applications here we find it particularly useful to employ notation designed explicitly for the time-series framework.) Similarly a *k-step conditional additivity rule*,

$$\alpha(x_1, \dots, x_n) = \alpha(x_1, \dots, x_{n-k}) + \alpha(x_{n-k+1}, \dots, x_n|x_1, \dots, x_{n-k}), \tag{2.13}$$

holds under the assumption that the k -step conditional distributions satisfy analogous Lipschitz-type properties. In practice, however, the one-step conditional distributions are more likely to be available and therefore more useful.

3. Applications to time series, and functionals of time series, with discrete conditional distributions

The simplest scenario to which one might envision applying the conditional additivity rule is that of time series where, for sufficiently large n , the one-step conditional distributions are *discrete* (more generally, we may consider functionals of such processes). The pointwise dimension map of a discrete distribution is always 0 a.s., so the dimension behaviour of such conditional distributions is immediately known. Assuming (2.6) holds throughout, it follows from Corollary 2.3 that pointwise dimension in such a real- or vector-valued time series will level off at constant values for large n . Such behaviour is known and expected for *deterministic* time series (see Example 3.3 below) and, as noted earlier, more generally for any time series which can be represented as a Lipschitz functional of a smooth finite-dimensional system. However, in Sections 3.1 and 3.2 we construct some natural stationary stochastic processes where pointwise dimension remains bounded over time. These examples can be regarded as ‘totally stochastic’ in the sense that the randomness at each step is governed by random coin tosses (Bernoulli variables); hence they differ in nature from the example in Cutler (1997) where a stochastic element is introduced by a bad yet smooth projection of a finite-dimensional system.

If $x_n \in \mathcal{X}$ is an atom of the conditional distribution $P(\cdot | \mathbf{x}_{n-1})$ then we automatically have $\alpha(x_n | \mathbf{x}_{n-1}) = 0$. Moreover, since the general inequality $\alpha(\mathbf{x}_n) \geq \alpha(\mathbf{x}_{n-1})$ is always true (the increasing property of pointwise dimension) it follows in this case that one direction $\alpha(\mathbf{x}_n) \geq \alpha(\mathbf{x}_{n-1}) + \alpha(x_n | \mathbf{x}_{n-1})$ of (2.6) immediately holds. Thus, in order to obtain (2.6), we need only prove the reverse inequality $\alpha(\mathbf{x}_n) \leq \alpha(\mathbf{x}_{n-1}) + \alpha(x_n | \mathbf{x}_{n-1})$, and for this it is sufficient to verify (2.5). Thus we have the following:

Theorem 3.1. *Let $\mathbf{X}_n = (X_1, \dots, X_n)$ be a random vector in \mathcal{X}^n with distribution P_n , and let $P(\cdot | \mathbf{x}_{n-1})$ denote the regular conditional distribution of X_n given $\mathbf{X}_{n-1} = \mathbf{x}_{n-1}$. Suppose $x_n \in \mathcal{X}$ is an atom of $P(\cdot | \mathbf{x}_{n-1})$. If the pointwise limit $\alpha(\mathbf{x}_{n-1})$ exists and the Lipschitz-type condition (2.5) is satisfied at $\mathbf{x}_n = (x_1, \dots, x_n)$, then (2.6) holds at \mathbf{x}_n and*

$$\alpha(\mathbf{x}_n) = \alpha(\mathbf{x}_{n-1}) + \underbrace{\alpha(x_n | \mathbf{x}_{n-1})}_{=0} = \alpha(\mathbf{x}_{n-1}). \tag{3.1}$$

Thus pointwise dimension remains constant in moving from \mathbf{x}_{n-1} to \mathbf{x}_n . More generally, even if the limit $\alpha(\mathbf{x}_{n-1})$ does not exist, we obtain

$$\alpha^-(\mathbf{x}_{n-1}) = \alpha^-(\mathbf{x}_n) \leq \alpha^+(\mathbf{x}_n) = \alpha^+(\mathbf{x}_{n-1}), \tag{3.2}$$

so the upper and lower limits remain constant.

In fact, proof of the left-hand equality in (3.2) requires a version of (2.8) for $\alpha^-(\mathbf{x}_n)$. Such an extension of (2.8) can be shown to hold assuming existence of the one limit $\alpha(x_n | \mathbf{x}_{n-1})$ and the Lipschitz-type condition (2.5). We will not provide a proof of this here.

There is a simple sufficient condition for (2.5) that is often useful in the case of discrete

conditional distributions. Suppose at the point \mathbf{x}_n there exist constants $b \geq 1$ and $c > 0$ and a Borel set S with $P_{n-1}(S) = 1$ such that $P(\cdot | \mathbf{x}_{n-1})$ satisfies the following for all $r > 0$:

$$P(B(x_n, br) | \mathbf{z}_{n-1}) \geq c \text{ for all } \mathbf{z}_{n-1} \in S \cap B(\mathbf{x}_{n-1}, r). \tag{3.3}$$

Note then that (2.5) holds by (non-optimally) taking $a_1(r) = c$. Thus we have the following:

Corollary 3.2. *Suppose $P(\cdot | \mathbf{x}_{n-1})$ is discrete for P_{n-1} -almost all \mathbf{x}_{n-1} and that for each such \mathbf{x}_{n-1} condition (3.3) holds at every atom of $P(\cdot | \mathbf{x}_{n-1})$ (the parameters b and c in (3.3) may vary from atom to atom). Then, assuming the limits $\alpha(\mathbf{x}_{n-1})$ exist P_{n-1} -a.s., we have $\alpha(\mathbf{x}_n) = \alpha(\mathbf{x}_{n-1})$ P_n -a.s.*

Example 3.3 *Deterministic time series.* Cutler (1997) defines a strictly stationary time series X_1, X_2, \dots to be *deterministic* if there exists an integer $n_p \geq 1$ and mapping $T: \mathcal{X}^{n_p} \rightarrow \mathcal{X}$ such that $X_{n_p+1} = T(X_1, \dots, X_{n_p})$ a.s. We assume that n_p is the smallest such integer and refer to it as the *predictive dimension* of the time series, while T is called the *predictor function*. (Such a time series can always be expressed as a smooth functional of a dynamical system evolving on the finite product space \mathcal{X}^{n_p} .) It follows from Theorem 2.18 of Cutler (1997) that $\alpha(\mathbf{x}_n) = \alpha(\mathbf{x}_{n_p})$ P_n -a.s. for all $n \geq n_p$ provided T is a Lipschitz mapping. However, this result also follows from Corollary 3.2 of this paper. For $n > n_p$ the one-step conditional distributions of the time series are discrete, in fact are Dirac point masses

$$P(\cdot | \mathbf{x}_{n-1}) = \delta_{x_n}(\cdot), \quad \text{where } x_n = T(x_{n-n_p}, \dots, x_{n-1}) \tag{3.4}$$

and $\delta_{x_n}(B) = 0$ unless $x_n \in B$. So, given an arbitrary point \mathbf{x}_{n-1} , we need only consider the single atom x_n given by (3.4). If T is a Lipschitz mapping, that is, there exists $K \geq 1$ such that $\rho(T(\mathbf{w}_{n_p}), T(\mathbf{z}_{n_p})) \leq K \rho(\mathbf{w}_{n_p}, \mathbf{z}_{n_p})$, then (3.3) holds at $\mathbf{x}_n = (x_1, \dots, x_n)$ with $c = 1$ and $b = K$. This follows because $\rho(\mathbf{z}_{n-1}, \mathbf{x}_{n-1}) \leq r \Rightarrow \rho(z_n, x_n) \leq Kr$ (where $z_n = T(z_{n-n_p}, \dots, z_{n-1})$) and so $z_n \in B(x_n, Kr)$. This gives $P(B(x_n, Kr) | \mathbf{z}_{n-1}) = 1$. It now follows from Corollary 3.2 that $\alpha(\mathbf{x}_n) = \alpha(\mathbf{x}_{n-1})$ P_n -a.s.

Remark 3.4. We should point out that stationarity of the time series is not actually required for the result stated in the previous example; the conditional distributions are Dirac point masses in any case and Corollary 3.2 applies generally. So, for example, if we selected a deterministic start (x_1, \dots, x_{n_p}) then we would have discrete P_n and $\alpha(\mathbf{x}_n) = 0$ for all $n \geq 1$. However, from the point of view of *observed* scaling behaviour in the time series, this result is not likely to be practically applicable in the absence of stationarity. In particular, if the process is moving towards stationarity (and this is often the case in the dynamical systems context) then the long-run scaling behaviour of the observed time series will tend to mimic that of the stationary distribution (at least at larger spatial scales) and not that of the initial distribution of (X_1, \dots, X_{n_p}) , even though the ‘true’ dimension is that of the latter. (We also assume ergodicity of the stationary distribution here.) This ‘inconsistency’ result is actually a blessing in disguise, since it is the stationary distribution which is usually of interest.

Example 3.5 *Deterministic Counterexample.* In Cutler (1990b) we construct a continuous but non-Lipschitz function $T: [0, 1] \rightarrow [0, 1]$ with ergodic invariant distribution μ having the

property that μ has positive mass on two sets of different Hausdorff dimensions. Specifically, there exist two values $0 < \gamma < \eta < 1$ and Borel sets D_γ and D_η such that $\mu(D_\gamma) = \mu(D_\eta) = 1/2$ and $\alpha_\mu(x) = \gamma$ for $x \in D_\gamma$ and $\alpha_\mu(x) = \eta$ for $x \in D_\eta$. (This construction, in which T can be regarded as isomorphic to the shift map on an underlying periodic Markov chain with three states, is sufficiently lengthy and non-trivial that we refer the interested reader to the original paper in lieu of reproducing the details here.) The effect of T is to shift back and forth between the two sets D_γ and D_η , that is, $T(D_\gamma) = D_\eta$ and $T(D_\eta) = D_\gamma$. Consequently the discrete conditional additivity rule (3.1) fails at the first step: for $x \in D_\gamma$ we have $\alpha(x, T(x)) = \eta > \gamma + 0 = \alpha(x)$. This shows that, in the absence of (2.5), a sudden jump in the dimension can occur even though, conditionally, we appear to be adding 0.

It is useful to note that if a time series arises as a Lipschitz functional of another time series, then pointwise dimension of the functional process remains bounded above by that of the original underlying time series. We make this statement precise in the following theorem; see Lemma 4.4 of Cutler (1997) for proof. For convenience, we will generally assume in this and the following sections that the pointwise limits $\alpha(x_n)$ exist.

Theorem 3.6. *Let Y_1, Y_2, \dots be a time series taking values in a complete separable metric space \mathcal{Y} and having joint distributions Q_n . If $h: \mathcal{Y} \rightarrow \mathcal{X}$ is a Lipschitz mapping, then the functional time series X_1, X_2, \dots given by $X_k = h(Y_k)$ and having joint distributions P_n satisfies $\alpha_{P_n}(x_n) \leq \inf \alpha_{Q_n}(y_n)$, where the infimum is taken over all $y_n = (y_1, \dots, y_n)$ satisfying $x_n = (h(y_1), \dots, h(y_n))$.*

In particular, if for sufficiently large n the underlying time series Y_1, Y_2, \dots has discrete conditional distributions satisfying the hypotheses of Corollary 3.2, thereby having pointwise dimension levelling off at constant values bounded above by some α_0 (this maximum value α_0 is necessarily finite if the Y s take values in \mathbb{R}^d), then the pointwise dimension of the functional time series X_1, X_2, \dots is also almost surely bounded above by α_0 . (However, it is worth noting that the conditional distributions of X_1, X_2, \dots will not necessarily be discrete.)

3.1. Missing-data models of dynamical systems

The occurrence of randomly missing observations during the recording of the evolution of a dynamical system will automatically turn the recorded time series into a stochastic process (the next recorded observation cannot be predicted with certainty based on the past, and the time series fails to meet the determinism criterion of Example 3.3). In our first and simplest missing-data model we consider the situation of randomly and independently missing observations in a time series generated by iterating a mapping φ . Specifically, let \mathcal{Y} be a complete separable metric space (with metric ρ) and $\varphi: \mathcal{Y} \rightarrow \mathcal{Y}$ a Lipschitz mapping with invariant distribution Q , that is, $Q\varphi^{-1} = Q$. The intended observed time series is the sequence Y_1, Y_2, \dots where $Y_n = \varphi^{n-1}(Y)$ is the $(n-1)$ th iterate of a random initial condition Y under φ . The initial condition Y is assumed to have distribution Q . The time series Y_1, Y_2, \dots is then strictly stationary and deterministic (as in Example 3.3) with

predictive dimension $n_p = 1$ and predictor function $T = \varphi$. Letting Q_n denote the distribution of (Y_1, \dots, Y_n) and noting that $Q_1 = Q$, it follows as a special case of Example 3.3 that, for all $n \geq 1$, $\alpha_{Q_n}(y_n) = \alpha_Q(y_1)$ Q_n -a.s.

Now suppose that observations are missing randomly and independently (each with probability p) from the deterministic series Y_1, Y_2, \dots , so that what we actually record is the stationary Markov process W_1, W_2, \dots with one-step discrete transition probabilities given by

$$P(W_n = \varphi^m(w) | W_{n-1} = w) = p^{m-1}(1 - p), \quad \text{for } m = 1, 2, \dots \tag{3.5}$$

The stationary marginal distribution of this Markov process is the invariant distribution Q of φ . In this simple model, each iterate Y_n in the original time series has probability p of going missing, and the independence assumption leads to the geometric probabilities in (3.5). We assume here that we do not know the time locations of the missing observations. In the case of one-dimensional maps it would probably be easy to determine the time locations, at least for small p , by plotting W_{n+1} versus W_n ; however, finding the time points becomes much more difficult in higher dimensions, as is certainly the case for the second missing-data model we consider later in this section.

Dickson (1995) studied the first missing-data model in the context of estimating the dominant Lyapunov exponent of the mapping φ using the popular Wolf algorithm (Wolf *et al.* 1985). He found that, even for very small p , estimates of the Lyapunov exponent were seriously corrupted by the ‘noise’ due to missing observations. This can be understood by noting that the Lyapunov exponent is determined by estimating derivatives of the (unknown) mapping φ , and a missing observation produces a sharp discontinuity in the numerator of the derivative expression, while the denominator (determined by nearest neighbours) tends to 0. This problem is only exacerbated by increasing the sample size, so that estimates of the Lyapunov exponent actually diverge to ∞ as the number of observations tends to ∞ . (The core of Dickson’s thesis was an algorithm that searched for, and removed, such discontinuity points from the calculation of the Lyapunov exponent.) While Lu and Smith (1997, p. 136) point out that one advantage of Lyapunov exponents over dimension is that the former have a direct interpretation in terms of the dynamics of the system, the following theorem shows that, in return, dimension is less sensitive to occasional disruptions, such as missing observations, in the dynamics. The pointwise dimension of the underlying deterministic system is preserved exactly. However, it should be expected that a substantial number of missing observations would have some effect on actual numerical estimates of dimension, likely slowing convergence.

Theorem 3.7 (First missing-data model). *Let φ be a Lipschitz mapping with invariant distribution Q , and let W_1, W_2, \dots be the stationary Markov process of the missing-data model described above, having one-step transition probabilities given by (3.5). Let \tilde{Q}_n denote the joint distribution of (W_1, \dots, W_n) . Then the pointwise dimension maps of the \tilde{Q}_n s are constant in n and coincide with the pointwise dimension map of Q ; that is, $\alpha_{\tilde{Q}_n}(\mathbf{w}_n) = \alpha_Q(w_1) = \alpha_Q(y_1)$ \tilde{Q}_n -a.s. for all $n \geq 1$, where y_1 is the initial condition of φ . In particular, if there exists α_0 such that $\alpha_Q(y) = \alpha_0$ Q -a.s., then also $\alpha_{\tilde{Q}_n}(\mathbf{w}_n) = \alpha_0$ \tilde{Q}_n -a.s.*

Proof. Note that $\tilde{Q}_1 = Q$, and let $K \geq 1$ be the Lipschitz constant of φ . To show that $\alpha_{\tilde{Q}_n}(\mathbf{w}_n) = \alpha_Q(w_1)$ \tilde{Q}_n -a.s. it suffices, by Corollary 3.2, to prove that the transition probabilities (3.5) satisfy (3.3) at each atom. Since $P(\cdot | W_{n-1} = w)$ has atoms $w_m = \varphi^m(w)$, $m = 1, 2, \dots$, consider the atom w_m . Then $\rho(z, w) \leq r \Rightarrow \rho(\varphi^m(z), w_m) \leq K^m r$, and so

$$P(W_n \in B(w_m, K^m r) | W_{n-1} = z) \geq P(W_n = \varphi^m(z) | W_{n-1} = z) = p^{m-1}(1 - p).$$

Thus (3.3) holds at (w, w_m) with $c = p^{m-1}(1 - p)$ and $b = K^m$; note that the values of c and b change with m and that the radius of the enlarged neighbourhood $B(w_m, br)$ expands to ∞ as $m \rightarrow \infty$. The equality $\alpha_Q(w_1) = \alpha_Q(y_1)$ follows from the fact that $w_1 = \varphi^k(y_1)$ for some $k \geq 0$ and $\alpha_Q(\cdot)$ is an invariant function (that is, $\alpha_Q(\varphi(y)) = \alpha_Q(y)$) when φ is Lipschitz; see Cutler (1990b). Finally, if we additionally have $\alpha_Q(y) = \alpha_0$ Q -a.s. then we also obtain $\alpha_{\tilde{Q}_n}(\mathbf{w}_n) = \alpha_0$ \tilde{Q}_n -a.s. since we can write $Q = \tilde{Q}_n \pi_n^{-1}$ where $\pi_n(\mathbf{w}_n) = w_1$. \square

While Theorem 3.7 is attractive, it will have limited value in practice because it is only in exceptional cases that we observe a dynamical system evolving in its natural phase space \mathcal{Y} (with or without missing observations). More typically we can measure only a functional $h: \mathcal{Y} \rightarrow \mathbb{R}^d$ of the system, so that the data take the form of a real- or vector-valued time series X_1, X_2, \dots , where $X_k = h(Y_k)$. If h is a typical ‘good’ functional – in the sense of *generic* as used by Takens (1981) or in the sense of *prevalent* as used by Sauer *et al.* (1991) and \mathcal{Y} is a finite-dimensional manifold, then the time series X_1, X_2, \dots has finite predictive dimension n_p and in fact an embedding Φ exists between the points Y_k and the vectors $(X_k, \dots, X_{k+n_p-1})$. In this case it is possible to recover many parameters (including the fractal dimensions) of the original underlying system. In the second missing-data model (below) we show that, provided h is at least Lipschitz continuous, pointwise dimension remains appropriately bounded above when observations are randomly missing from the functional time series X_1, X_2, \dots , and further that the exact dimension will be recovered asymptotically provided h is a good functional. Again we assume that the probability any observation X_j goes missing is p , and that observations go missing independently of one another. Let Z_1, Z_2, \dots denote the stationary vector-valued stochastic process we actually observe; it is worth noting that the Z process, unlike the W process of the first missing-data model, need not (in fact generally will not) be Markov. Moreover, the one-step conditional distributions of the Z process need not be discrete, even for large n . The key to the first part of our result below is to realize that the Z process can be represented as a functional of the W process; that is, although it is actually certain observations $X_k = h(Y_k)$ that are missing, it is probabilistically equivalent to regard the corresponding underlying Y_k s as missing and write $Z_k = h(W_k)$. We obtain the following:

Theorem 3.8 (Second missing-data model). *Let φ be a Lipschitz mapping on \mathcal{Y} with invariant distribution Q , and let Y_1, Y_2, \dots be the deterministic time series generated by iterating φ . Suppose $h: \mathcal{Y} \rightarrow \mathbb{R}^d$ is a Lipschitz functional and X_1, X_2, \dots is the functional time series $X_k = h(Y_k)$ with joint distributions P_n , $n = 1, 2, \dots$. Let Z_1, Z_2, \dots be the process generated by the second missing-data model with \tilde{P}_n the joint distribution of (Z_1, \dots, Z_n) . Then we have the inequality $\alpha_{\tilde{P}_n}(z_n) \leq \alpha_Q(y_1)$ \tilde{P}_n -a.s. (where y_1 is the initial condition of φ). If $\alpha_Q(y) \leq \alpha_0$ Q -a.s. then $\alpha_{\tilde{P}_n}(z_n) \leq \alpha_0$ \tilde{P}_n -a.s. Moreover, if $\alpha_Q(y) = \alpha_0$ Q -a.s. and h is a good functional with embedding dimension n_p , then $\lim_{n \rightarrow \infty} \alpha_{\tilde{P}_n}(z_n) = \alpha_0$ with probability one.*

Proof. As noted earlier, we may write $Z_k = h(W_k)$, where W represents the first missing-data model. Applying Theorem 3.6 followed by Theorem 3.7, we obtain

$$\alpha_{\tilde{P}_n}(h(w_1), \dots, h(w_n)) \leq \alpha_{\tilde{Q}_n}(w_1, \dots, w_n) = \alpha_Q(w_1) = \alpha_Q(y_1) \tilde{Q}_n \text{ a.s.},$$

which gives $\alpha_{\tilde{P}_n}(z_n) \leq \alpha_Q(y_1) \tilde{P}_n$ -a.s. since $\tilde{P}_n = \tilde{Q}_n h_n^{-1}$ where $h_n(w_n) = (h(w_1), \dots, h(w_n))$. Now suppose h is a good functional with embedding dimension n_p so that an embedding Φ is achieved between the points Y_k and the vectors $(X_k, \dots, X_{k+n_p-1})$. The embedding Φ is bi-Lipschitz with $\Phi(y_k) = (x_k, \dots, x_{k+n_p-1})$. As a consequence the X series is deterministic (see Example 3.3) with predictive dimension no greater than n_p . It follows that $\alpha_{P_n}(x_n) = \alpha_{P_{n_p}}(x_{n_p}) = \alpha_Q(y_1) P_n$ -a.s. for all $n \geq n_p$ (the last equality is a consequence of the bi-Lipschitz property of Φ). Now consider the Z process. By the increasing property of pointwise dimension, we know that $\lim_{n \rightarrow \infty} \alpha_{\tilde{P}_n}(z_1, \dots, z_n)$ exists for each sequence z_1, z_2, \dots ; moreover, that limit is bounded above almost surely by α_0 when $\alpha_Q(y) = \alpha_0$ Q -a.s. We now want to show that the limit in fact equals α_0 with probability one. First, note that by the monotone convergence theorem we have $\lim_{n \rightarrow \infty} E(\alpha_{\tilde{P}_n}(Z_1, \dots, Z_n)) \leq \alpha_0$. Then note that, with probability one, each z_1, z_2, \dots sequence contains a string $z_{r+1}, z_{r+2}, \dots, z_{r+n_p}$ consisting of n_p consecutive x observations where none have gone missing; that is, there exist $r \geq 0$ and $s \geq 0$ such that $z_{r+m} = x_{s+m}$ for each $m = 1, \dots, n_p$. Now condition on the event $I_r = \{(Z_{r+1}, \dots, Z_{r+n_p}) \text{ is the first unbroken } X\text{-string}\}$. It follows that the conditional distribution of $(Z_{r+1}, \dots, Z_{r+n_p})$ given I_r coincides with P_{n_p} ; hence

$$E(\alpha_{\tilde{P}_n}(Z_1, \dots, Z_n) | I_r) = \alpha_0 \text{ for } n \geq r + n_p.$$

Then

$$E(\alpha_{\tilde{P}_n}(Z_1, \dots, Z_n)) \geq \sum_{r=0}^{n-n_p} E(\alpha_{\tilde{P}_n}(Z_1, \dots, Z_n) | I_r) P(I_r) = \alpha_0 P(\cup_{r=0}^{n-n_p} I_r) \rightarrow \alpha_0$$

since

$$\lim_{n \rightarrow \infty} P(\cup_{r=0}^n I_r) = 1.$$

We thus conclude that $\lim_{n \rightarrow \infty} E(\alpha_{\tilde{P}_n}(Z_1, \dots, Z_n)) = \alpha_0$. Combining this with the fact that $\alpha_{\tilde{P}_n}(z_1, \dots, z_n)$ increases pointwise to a limit bounded above by α_0 , we get $\lim_{n \rightarrow \infty} \alpha_{\tilde{P}_n}(z_1, \dots, z_n) = \alpha_0$ with probability one, as claimed. \square

3.2. Random iterations of dynamical systems and iterated function systems

Here we consider the situation where we have two mappings φ and ψ with common domain and range \mathcal{Y} , and a randomly chosen initial condition Y_1 . At each iteration one of the mappings is randomly selected and applied to the current state, leading to a stochastic process Y_1, Y_2, \dots . We will assume the simplest possible model here, where the choice of mapping at the n th iteration is determined by the toss of a weighted coin, resulting in the Markov process

$$Y_{n+1} = \begin{cases} \varphi(Y_n) & \text{with probability } \theta \\ \psi(Y_n) & \text{with probability } 1 - \theta \end{cases} \tag{3.6}$$

with one-step conditional distributions given by the weighted average of two Dirac measures

$$P(\cdot|y) = \theta \delta_{\varphi(y)}(\cdot) + (1 - \theta) \delta_{\psi(y)}(\cdot). \tag{3.7}$$

Such processes are studied, for example, by Bhattacharya and Rao (1993); they consider the random iteration of two quadratic or logistic maps on the unit interval. It is clear that we could employ a more complicated randomization rule for selecting the mapping at each stage, or select from among a larger number of mappings, without greatly altering the flavour of the problem. In the case that the mappings are also contractive (that is, $\rho(\varphi(x), \varphi(y)) \leq c\rho(x, y)$, where $c < 1$), we have a *hyperbolic iterated function system* (IFS); these form the foundation for fractal image compression techniques (see Barnsley and Elton 1988). We obtain the following:

Theorem 3.9. *Let Y_1, Y_2, \dots be the Markov process given in (3.6) with joint distributions $P_n, n = 1, 2, \dots$. Provided the mappings φ and ψ are both Lipschitz continuous, we have $\alpha(y_n) = \alpha(y_{n-1}) P_n$ -a.s. for $n \geq 2$, regardless of the distribution of the initial condition Y_1 .*

Proof. Let $K_1 \geq 1, K_2 \geq 1$ denote the Lipschitz constants of φ and ψ respectively, and set $b = \max(K_1, K_2)$. Let $y \in \mathcal{Y}$. Then for $z \in B(y, r)$ we have

$$P(B(w, br)|z) = \theta \delta_{\varphi(z)}(B(w, br)) + (1 - \theta) \delta_{\psi(z)}(B(w, br)) \geq \min(\theta, 1 - \theta)$$

for either of the possible atoms $w = \varphi(y)$ and $w = \psi(y)$. Taking $c = \min(\theta, 1 - \theta)$, the result now follows from Corollary 3.2. □

We end this section by noting that while the above theorem is true no matter how the initial condition is chosen, we are typically interested in the stationary case, as noted in Remark 3.2. Bhattacharya and Rao (1993) discuss the existence and nature of unique invariant measures for the Markov process (3.6) generated by logistic maps, while Elton (1987) and Barnsley and Elton (1988) obtain results on unique invariant measures for certain types of IFS. Note that Theorem 3.9 helps explain why stochastic algorithms for iterating IFS nonetheless produce finite-dimensional attractors and finite-dimensional invariant measures in the limit.

4. Applications to stochastic dynamical systems with additive noise

In many cases of practical interest, it is both necessary and useful to modify the simple deterministic systems discussed in Example 3.3 to allow for routine perturbations due to external (typically diffuse) noise. The simplest (and one of the most useful) modifications is a *stochastic dynamical system with additive noise*:

$$X_n = T(X_{n-p}, \dots, X_{n-1}) + \varepsilon_n, \tag{4.1}$$

where T is the deterministic ‘skeleton’ of the system (playing the same role as in Example 3.3) and $\varepsilon_1, \varepsilon_2, \dots$ is a sequence of random variables representing the noise introduced into the system at each step. Such stochastic systems are discussed in detail in Kifer (1988), Tong (1995; 1997) and Berliner *et al.* (1997). The integer p in (4.1) is generally called the *order* of

the resulting stochastic time series X_1, X_2, \dots and in fact corresponds to the predictive dimension of the ‘noiseless’ system driven by T alone.

The situation here is quite different from that of systems corrupted by discrete noise as in Section 3, where the dimensions of the underlying deterministic systems are essentially preserved in spite of the noise. In the case of stochastic systems with additive diffuse noise, it is part of vague folklore that the fractal dimensions of the time series should increase without bound as $n \rightarrow \infty$, regardless of the behaviour of the deterministic skeleton T ; see the discussion in the context of correlation dimension in Tong (1995, p. 415). We provide a simple rigorous proof of this for pointwise dimension via our conditional additivity rule. Although the case of independent and identically distributed (i.i.d.) noise can be subsumed under the more general case of Theorem 4.2 below, it is of such importance (and its proof comparatively simpler) that we treat it separately first. Note that there are no stationarity requirements on the time series.

Theorem 4.1 (i.i.d. noise). *Let X_1, X_2, \dots be a time series taking values in \mathbb{R}^d of the form*

$$X_n = T(X_{n-p}, \dots, X_{n-1}) + \varepsilon_n, \tag{4.2}$$

where $T: (\mathbb{R}^d)^p \rightarrow \mathbb{R}^d$ is a Lipschitz mapping. Suppose that the sequence of noise random variables $\varepsilon_1, \varepsilon_2, \dots$ has the following property: there exists a distribution P_e in \mathbb{R}^d such that, for each $n > p$, the conditional distributions $\varepsilon_n / X_1, \dots, X_{n-1} \stackrel{\mathcal{L}}{\sim} P_e$. (These conditions are automatically met if $\varepsilon_1, \varepsilon_2, \dots$ is an i.i.d. noise sequence which is also independent of the initial distribution of the system.) Then for each $n > p$ and $\mathbf{x}_n \in (\mathbb{R}^d)^n$ the one-step conditional additivity rule holds for the joint distributions P_n of X_1, X_2, \dots ; that is

$$\alpha(\mathbf{x}_n) = \alpha(\mathbf{x}_{n-1}) + \alpha(x_n | \mathbf{x}_{n-1}). \tag{4.3}$$

If we further assume that there exists a constant $\alpha_0 > 0$ such that $\alpha_{P_e}(x) = \alpha_0$ P_e -a.s. (note that this condition is automatically met with $\alpha_0 = d$ if P_e is absolutely continuous with respect to Lebesgue measure in \mathbb{R}^d) then we have

$$\alpha(\mathbf{x}_{p+m}) = \alpha(\mathbf{x}_p) + m\alpha_0 \text{ } P_{p+m}\text{-a.s.}, \quad \text{for all } m \geq 0, \tag{4.4}$$

and so $\lim_{n \rightarrow \infty} \alpha(\mathbf{x}_n) = \infty$ with probability one.

Proof. Let $n > p$ and for convenience let $\mathbf{z}_{n-1}^* = (z_{n-p}, \dots, z_{n-1})$. Select any point $\mathbf{x}_n = (x_1, \dots, x_n)$ in $(\mathbb{R}^d)^n$. It follows from (4.2) and our hypotheses that for any $\mathbf{z}_{n-1} \in (\mathbb{R}^d)^{n-1}$ we have

$$P(B(x_n, r) | \mathbf{z}_{n-1}) = P_e(B(x_n - T(\mathbf{z}_{n-1}^*), r)). \tag{4.5}$$

Now by assumption T satisfies a Lipschitz condition, say with Lipschitz constant $K \geq 1$. It follows from this and the triangle inequality that if $\|\mathbf{z}_{n-1}^* - \mathbf{x}_{n-1}^*\| \leq r$ then

$$B(x_n - T(\mathbf{z}_{n-1}^*), r) \subseteq B(x_n - T(\mathbf{x}_{n-1}^*), (K + 1)r) \tag{4.6}$$

and similarly

$$B(x_n - T(\mathbf{x}_{n-1}^*), (K + 1)r) \supseteq B(x_n - T(\mathbf{z}_{n-1}^*), r). \tag{4.7}$$

Taking $b = K + 1$, from (4.5) and (4.6) we then have

$$\begin{aligned} P(B(x_n, r)|z_{n-1}) &\leq P_e(B(x_n - T(\mathbf{x}_{n-1}^*), br)) \\ &= P(B(x_n, br)|\mathbf{x}_{n-1}), \quad \text{for all } z_{n-1} \in B(\mathbf{x}_{n-1}, r), \end{aligned} \tag{4.8}$$

and similarly, from (4.5) and (4.7) we have

$$\begin{aligned} P(B(x_n, br)|z_{n-1}) &\geq P_e(B(x_n - T(\mathbf{x}_{n-1}^*), r)) \\ &= P(B(x_n, r)|\mathbf{x}_{n-1}), \quad \text{for all } z_{n-1} \in B(\mathbf{x}_{n-1}, r), \end{aligned} \tag{4.9}$$

verifying equations (2.4) and (2.5). This shows that the one-step conditional additivity rule (4.3) holds at every \mathbf{x}_n . Now for $n > p$ it also follows from (4.5) that

$$\alpha(x_n|\mathbf{x}_{n-1}) = \alpha_{P_e}(x_n - T(\mathbf{x}_{n-1}^*)). \tag{4.10}$$

Suppose now that there exists $\alpha_0 > 0$ such that $P_e(A) = 1$, where $A = \{x \in \mathbb{R}^d | \alpha_{P_e}(x) = \alpha_0\}$. We prove that (4.4) is true by induction on m . First, define $S_m = \{\mathbf{x}_{p+m} | \alpha(\mathbf{x}_{p+m}) = \alpha(\mathbf{x}_p) + m\alpha_0\}$ and $A_m = \{\mathbf{x}_{p+m} | \alpha(x_{p+m}|\mathbf{x}_{p+m-1}) = \alpha_0\} = \{\mathbf{x}_{p+m} | x_{p+m} - T(\mathbf{x}_{p+m-1}^*) \in A\}$. Note that (4.4) is obviously true for $m = 0$; that is, $P_p(S_0) = 1$. Now assume it is true for m . Letting P denote the probability measure on the underlying probability space of the time series, we have

$$\begin{aligned} P_{p+m+1}(A_{m+1}) &= P([X_{p+m+1} - T(\mathbf{X}_{p+m}^*) \in A]) \\ &= P([\varepsilon_{p+m+1} \in A]) \\ &= P_e(A) = 1. \end{aligned} \tag{4.11}$$

Now note that (4.3) implies $S_{m+1} \supseteq (S_m \times \mathbb{R}^d) \cap A_{m+1}$. Since the induction hypothesis gives $P_{p+m}(S_m) = 1$, it follows from this and (4.11) that $P_{p+m+1}(S_{m+1}) \geq P_{p+m+1}((S_m \times \mathbb{R}^d) \cap A_{m+1}) = 1$, which completes the proof. \square

In many cases it is more plausible to consider location-dependent noise; for example, we might expect the variance of the noise distribution to depend on the location in space. (Special types of state-dependence are given by ARCH/GARCH models; see Shephard 1996.) We address this more general situation in the following theorem. Let $\mathbf{x}_{n-1}^* = (x_{n-p}, \dots, x_{n-1})$ and $\mathbf{X}_{n-1}^* = (X_{n-p}, \dots, X_{n-1})$, following the shorthand notation of the preceding proof.

Theorem 4.2 (Location-dependent noise). *Let X_1, X_2, \dots be a time series taking values in \mathbb{R}^d of the form*

$$X_n = T(X_{n-p}, \dots, X_{n-1}) + \varepsilon_n, \tag{4.12}$$

where $T: (\mathbb{R}^d)^p \rightarrow \mathbb{R}^d$ is a Lipschitz mapping. Suppose that the sequence of noise random variables $\varepsilon_1, \varepsilon_2, \dots$ has the following property: for each $n > p$ there exists a family of distributions $P_e(\cdot|\mathbf{x}_{n-1})$ in \mathbb{R}^d such that the conditional distributions $\varepsilon_n|\mathbf{X}_{n-1} = \mathbf{x}_{n-1} \stackrel{\mathcal{L}}{\sim} P_e(\cdot|\mathbf{x}_{n-1})$. Moreover, suppose that the noise distributions $P_e(\cdot|\mathbf{x}_{n-1})$ satisfy equations

(2.4) and (2.5) at each point \mathbf{x}_n . Then for each $n > p$ and $\mathbf{x}_n \in (\mathbb{R}^d)^n$ the one-step conditional additivity rule holds for the joint distributions P_n of X_1, X_2, \dots ; that is,

$$\alpha(\mathbf{x}_n) = \alpha(\mathbf{x}_{n-1}) + \alpha(x_n | \mathbf{x}_{n-1}). \tag{4.13}$$

If we further assume that there exists a constant $\alpha_0 > 0$ such that $\alpha_{P_e}(x | \mathbf{x}_{n-1}) = \alpha_0$ almost surely for each $n > p$ and $\mathbf{x}_{n-1} \in (\mathbb{R}^d)^{n-1}$, then we have

$$\alpha(\mathbf{x}_{p+m}) = \alpha(\mathbf{x}_p) + m\alpha_0 \text{ } P_{p+m}\text{-a.s.,} \quad \text{for all } m \geq 0, \tag{4.14}$$

and so $\lim_{n \rightarrow \infty} \alpha(\mathbf{x}_n) = \infty$ with probability one.

Proof. Analogously to (4.5) we have

$$P(B(x_n, r) | \mathbf{z}_{n-1}) = P_e(B(x_n - T(\mathbf{z}_{n-1}^*), r) | \mathbf{z}_{n-1}). \tag{4.15}$$

Now let $\mathbf{x}_n = (x_1, \dots, x_n)$ be given, and set $\tilde{\mathbf{x}}_n = (x_1, \dots, x_{n-1}, w_n) = (\mathbf{x}_{n-1}, w_n)$, where $w_n = x_n - T(\mathbf{x}_{n-1}^*)$. By assumption $P_e(\cdot | \mathbf{x}_{n-1})$ satisfies (2.4) and (2.5) at the point $\tilde{\mathbf{x}}_n$, for some parameters $a_1(r)$, $a_2(r)$, and $b_0 \geq 1$. (We reserve the symbol b for $b = K + 1$, where K is the Lipschitz constant of T , as used in the proof of Theorem 4.1.) Now (4.6) continues to hold and so for $\|\mathbf{z}_{n-1} - \mathbf{x}_{n-1}\| \leq r$ we obtain

$$\begin{aligned} P(B(x_n, r) | \mathbf{z}_{n-1}) &= P_e(B(x_n - T(\mathbf{z}_{n-1}^*), r) | \mathbf{z}_{n-1}) \\ &\leq P_e(B(x_n - T(\mathbf{x}_{n-1}^*), br) | \mathbf{z}_{n-1}) \\ &\leq a_2(r) P_e(B(x_n - T(\mathbf{x}_{n-1}^*), b_0 br) | \mathbf{x}_{n-1}) \\ &= a_2(r) P(B(x_n, b_0 br) | \mathbf{x}_{n-1}). \end{aligned} \tag{4.16}$$

Similarly (4.7) continues to hold, yielding

$$\begin{aligned} P(B(x_n, b_0 br) | \mathbf{z}_{n-1}) &= P_e(B(x_n - T(\mathbf{z}_{n-1}^*), b_0 br) | \mathbf{z}_{n-1}) \\ &\geq P_e(B(x_n - T(\mathbf{x}_{n-1}^*), b_0 r) | \mathbf{z}_{n-1}) \\ &\geq a_1(r) P_e(B(x_n - T(\mathbf{x}_{n-1}^*), r) | \mathbf{x}_{n-1}) \\ &= a_1(r) P(B(x_n, r) | \mathbf{x}_{n-1}). \end{aligned} \tag{4.17}$$

Together (4.16) and (4.17) give (4.13). Now set $A(\mathbf{x}_n) = \{\mathbf{x} \in \mathbb{R}^d | \alpha_{P_e}(x | \mathbf{x}_n) = \alpha_0\}$ and suppose that $P_e(A(\mathbf{x}_n) | \mathbf{x}_n) = 1$ for each $n \geq p$ and P_n -almost all \mathbf{x}_n . Setting $A_m = \{\mathbf{x}_{p+m} | \alpha(x_{p+m} | \mathbf{x}_{p+m-1}) = \alpha_0\}$ as before and noting that, by (4.15),

$$\alpha(x_n | \mathbf{x}_{n-1}) = \alpha_{P_e}(x_n - T(\mathbf{x}_{n-1}^*) | \mathbf{x}_{n-1}), \tag{4.18}$$

we obtain

$$\begin{aligned}
 P_{p+m+1}(A_{m+1}) &= P([X_{p+m+1} - T(\mathbf{X}_{p+m}^*) \in A(\mathbf{X}_{p+m})]) \\
 &= \int_{(\mathbb{R}^d)^{p+m}} P([X_{p+m+1} - T(\mathbf{X}_{p+m}^*) \in A(\mathbf{X}_{p+m}) | \mathbf{X}_{p+m} = \mathbf{x}_{p+m}) P_{p+m}(d\mathbf{x}_{p+m}) \\
 &= \int_{(\mathbb{R}^d)^{p+m}} P([\varepsilon_{p+m+1} \in A(\mathbf{x}_{p+m}) | \mathbf{X}_{p+m} = \mathbf{x}_{p+m}) P_{p+m}(d\mathbf{x}_{p+m}) \\
 &= \int_{(\mathbb{R}^d)^{p+m}} P_e(A(\mathbf{x}_{p+m}) | \mathbf{x}_{p+m}) P_{p+m}(d\mathbf{x}_{p+m}) \\
 &= 1,
 \end{aligned}
 \tag{4.19}$$

using the assumptions on the noise. The remainder of the proof follows by induction as in the proof of Theorem 4.1. □

Example 4.3. Consider the simple case where T is a Lipschitz mapping of \mathbb{R} into \mathbb{R} and the noise distributions are mean-zero Gaussian with location-dependent variances $\sigma^2(x) > 0$, $x \in \mathbb{R}$. We assume that the noise variable ε_n , given the current location $X_{n-1} = x$, is independent of the past of the process. Then

$$\phi(u|x) = \frac{\exp[-u^2/2\sigma^2(x)]}{(2\pi\sigma^2(x))^{1/2}}
 \tag{4.20}$$

is the density function of $P_e(\cdot|x)$ and we see that

$$\begin{aligned}
 P_e(B(w, r)|z) &= \int_{w-r}^{w+r} \phi(u|z) du \\
 &= \left(\frac{\sigma^2(x)}{\sigma^2(z)}\right)^{1/2} \int_{w-r}^{w+r} c(u, z, x) \phi(u|x) du,
 \end{aligned}
 \tag{4.21}$$

where $c(u, z, x) = \exp[u^2(\sigma^2(z) - \sigma^2(x))/(2\sigma^2(x)\sigma^2(z))]$. It is easy to see that provided $\sigma^2(z)$ is bounded away from 0 and away from ∞ over some r_0 -neighbourhood of x (as will certainly be the case if $\sigma^2(z)$ is continuous at x) then we have positive constants a_1 and a_2 (which generally depend on w and x) such that

$$a_1 P_e(B(w, r)|x) \leq P_e(B(w, r)|z) \leq a_2 P_e(B(w, r)|x)
 \tag{4.22}$$

for all $z \in B(x, r)$ and $0 < r < r_0$. Thus (2.4) and (2.5) hold for the noise distributions and so (4.13) is valid. Moreover, as existence of a density $\phi(u|x)$ on \mathbb{R} implies that $\alpha_{P_e}(w|x) = 1$ almost surely, all conclusions of Theorem 4.2 hold and the pointwise dimension of the stochastic system $X_n = T(X_{n-1}) + \varepsilon_n$ diverges to ∞ .

Acknowledgements

I am grateful to a referee for pointing out a connection between the results of Section 4 and ARCH/GARCH models, and to my colleague Jiahua Chen for some useful discussions. This research has been supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

References

- Abarbanel, H.D.I., Brown, R., Sidorowich, J.J. and Tsimring, L.S. (1993) The analysis of observed chaotic data in physical systems. *Rev. Modern Phys.*, **65**, 1331–1392.
- Barndorff-Nielsen, O.E., Jensen, J.L. and Kendall, W.S. (eds) (1993) *Networks and Chaos – Statistical and Probabilistic Aspects*. London: Chapman & Hall.
- Barnsley, M.F. and Elton, J.H. (1988) A new class of Markov processes for image encoding. *Adv. Appl. Probab.*, **20**, 14–32.
- Berliner, L.M., MacEachern, S.N. and Forbes, C.S. (1997) Ergodic distributions of random dynamical systems. In C.D. Cutler and D.T. Kaplan (eds), *Nonlinear Dynamics and Time Series: Building a Bridge between the Natural and Statistical Sciences*, Fields Inst. Commun. 11, pp. 171–185. Providence, RI: American Mathematical Society.
- Bhattacharya, R.N. and Rao, B.V. (1993) Random iteration of two quadratic maps. In S. Cambanis, J.K. Gosh, R.L. Karandikar, and P.K. Sen (eds), *Stochastic Processes: A Festschrift in Honour of Gopinath Kallianpur*, pp. 13–22. New York: Springer-Verlag.
- Billingsley, P. (1960) Hausdorff dimension in probability theory. *Illinois J. Math.*, **4**, 187–209.
- Billingsley, P. (1961) Hausdorff dimension in probability theory II. *Illinois J. Math.*, **5**, 291–298.
- Brown, G., Michon, G. and Peyrière, J. (1992) On the multifractal analysis of measures. *J. Statist. Phys.*, **66**, 775–790.
- Cutler, C.D. (1990a) A dynamical system with integer information dimension and fractal correlation exponent. *Comm. Math. Phys.*, **129**, 621–629.
- Cutler, C.D. (1990b) Connecting ergodicity and dimension in dynamical systems. *Ergodic Theory Dynam. Systems*, **10**, 451–462.
- Cutler, C.D. (1991) Some results on the behaviour and estimation of the fractal dimensions of distributions on attractors. *J. Statist. Phys.*, **62**, 651–708. Errata and addenda: **65**, 417–418.
- Cutler, C.D. (1993) A review of the theory and estimation of fractal dimension. In H. Tong (ed.), *Dimension Estimation and Models*, pp. 1–107. Singapore: World Scientific.
- Cutler, C.D. (1994) A theory of correlation dimension for stationary time series. *Philos. Trans. Roy. Soc. London Ser. A*, **348**, 343–355. Expanded version with proofs in H. Tong (ed.) *Chaos and Forecasting*, pp. 31–56. Singapore: World Scientific, (1995).
- Cutler, C.D. (1997) A general approach to predictive and fractal scaling dimensions in discrete-index time series. In C.D. Cutler and D.T. Kaplan (eds), *Nonlinear Dynamics and Time Series: Building a Bridge between the Natural and Statistical Sciences*, Fields Inst. Commun. 11, pp. 29–48. Providence, RI: American Mathematical Society.
- Dickson, R.J. (1995) The estimation of Lyapunov exponents for systems undergoing occasional random shocks. M. Math. thesis, University of Waterloo.
- Dudley, R.M. (1989) *Real Analysis and Probability*. Pacific Grove, CA: Wadsworth and Brooks/Cole.
- Elton, J.H. (1987) An ergodic theorem for iterated maps. *Ergodic Theory Dynam. Systems*, **7**, 481–488.

- Falconer, K.J. (1990) *Fractal Geometry: Mathematical Foundations and Applications*. New York: Wiley.
- Farmer, J.D. (1982) Information dimension and probabilistic structure of chaos. *Z. Naturforsch.* **37A**, 1304–1325.
- Farmer, J.D., Ott, E. and Yorke, J.A. (1983) The dimension of chaotic attractors. *Physica D*, **7**, 153–180.
- Halsey, T.C., Jensen, M.H., Kadanoff, L.P., Procaccia, I. and Shraiman, B.I. (1986) Fractal measures and their singularities: the characterization of strange sets. *Phys. Rev. A*, **33**, 1141–1151.
- Hentschel, H.G.E. and Procaccia, I. (1983). The infinite number of generalized dimensions of fractals and strange attractors. *Physica D*, **8**, 435–444.
- Holley, R. and Waymire, E.C. (1992) Multifractal dimensions and scaling exponents for strongly bounded random cascades. *Ann. Appl. Probab.*, **2**, 819–845.
- Keller, G. (1997) A new estimator for information dimension with standard errors and confidence intervals. *Stochastic Process. Appl.*, **71**, 187–206.
- Kifer, Y. (1988) *Random Perturbations of Dynamical Systems*. Boston: Birkhäuser.
- Lu, Z.-Q. and Smith, R.L. (1997) Estimating local Lyapunov exponents. In C.D. Cutler and D.T. Kaplan (eds), *Nonlinear Dynamics and Time Series: Building a Bridge between the Natural and Statistical Sciences*, Fields Inst. Commun. 11, pp. 135–151. Providence, RI: American Mathematical Society.
- Osborne, A.R. and Provenzale, A. (1989) Finite correlation dimension for stochastic systems with power-law spectra. *Physica D*, **35**, 357–381.
- Ott, E., Sauer, T. and Yorke, J.A. (eds) (1994) *Coping with Chaos*. New York: Wiley.
- Pesin, Ya.B. (1996) *Dimension Theory in Dynamical Systems*. Chicago: University of Chicago Press.
- Sauer, T., Yorke, J.A. and Casdagli, M. (1991) Embedology. *J. Statist. Phys.*, **65**, 579–616.
- Serinko, R.J. (1994) A consistent approach to least squares estimation of correlation dimension in weak Bernoulli dynamical systems. *Ann. Appl. Probab.*, **4**, 1234–1254.
- Shephard, N. (1996) Statistical aspects of ARCH and stochastic volatility. In D.R. Cox, D.V. Hinkley, and O.E. Barndorff-Nielsen (eds), *Time Series Models: In Econometrics, Finance and Other Fields*, London: Chapman & Hall.
- Smith, R.L. (1992) Estimating dimension in noisy chaotic time series. *J. Roy. Statist. Soc. Ser. B*, **54**, 329–351.
- Takens, F. (1981) Detecting strange attractors in turbulence. In D.A. Rand and L.-S. Young (eds), *Dynamical Systems and Turbulence, Warwick 1980*, Lecture Notes in Math. 898, pp. 366–381. New York: Springer-Verlag.
- Taylor, S.J. (1986) The measure theory of random fractals. *Math. Proc. Cambridge Philos. Soc.*, **100**, 383–406.
- Taylor, S.J. (1992) A measure theory definition of fractals. *Rend. Circ. Mat. Palermo (2) Suppl.*, **28**, 371–378.
- Taylor, S.J. (1995) The fractal analysis of Borel measures in \mathbb{R}^d . *J. Fourier Anal. Appl.*, Special Issue, 553–568.
- Theiler, J. (1990) Estimating the fractal dimension of chaotic time series. *Lincoln Laboratory J.*, **1**, 63–86.
- Theiler, J. (1991) Some comments on the correlation dimension of $1/f^\alpha$ noise. *Phys. Lett. A*, **155**, 480–493.
- Tong, H. (1995) A personal overview of non-linear time series analysis from a chaos perspective. *Scand. J. Statist.*, **22**, 399–421.
- Tong, H. (1997) Some comments on nonlinear time series analysis. In C.D. Cutler and D.T. Kaplan

(eds), *Nonlinear Dynamics and Time Series: Building a Bridge between the Natural and Statistical Sciences*, Fields Inst. Commun. 11, pp. 17–27. Providence, RI: American Mathematical Society.

Wolf, A., Swift, J.B., Swinney, H.L. and Vastano, J.A. (1985) Determining Lyapunov exponents from a time series. *Physica D*, **16**, 285–317.

Wolff, R.C.L. (1990) A note on the behaviour of the correlation integral in the presence of a time series. *Biometrika*, **77**, 689–697.

Young, L.-S. (1982) Dimension, entropy, and Lyapunov exponents. *Ergodic Theory Dynam. Systems*, **2**, 109–124.

Received December 1997 and revised February 1999